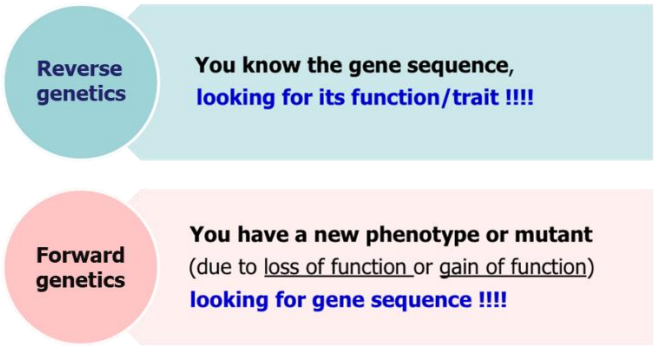
T-DNA tool for various applications in Plants

Understanding the role of Agrobacterium Ti plasmid genetic components (i.e., biology of T-DNA transfer process) in gene transfer from soil borne plant pathogenic bacteria to plant (across kingdom) is essential for executing a variety of applications in plants, such as:

1. Development of transgenic plants by transferiing candidate gene(s) to genome of crop of interest for
   - increasing production or yield of crops
   - improving the quality of plant products
   - tailoring plants for production of novel molecules (antibodies, industrial enzymes, functional foods etc.)
2. Understanding gene function (Gene discovery by **Forward genetics**) by development of
   - Loss of function mutatnts (by **T-DNA insertional muatgenesis**)
   - Gain of function mutants (by **T-DNA activation tagging**)
3. Targeted gene mutagenesis (Gene discovery by **Reverse genetics**)
   - Over-expression of a gene (of which sequence is known but trait/character controlled by it is unknown)
   - Suppression of a gene (of which sequence is known but trait/character controlled by it is unknown) by
     # Antisense approach
     # RNA interference (RNAi)
     # VIGS (Virus induced gene silencing)
4. Promoter trapping (discovery of plant promoters)
5. Enhancer trapping (discovery of new regulatory elements from plants)
6. Basic cellular studies in plants (such as protein localization and movement etc. through reporter tags)



Understanding gene function by Reverse Genetics

With the advent of affordable genome sequencing, genome sequence data from a wide variety of plants are now available and the number of sequenced genes whose function remains unknown has increased in recent years. The databases are filled with sequence information of genes with no known biological function, and while bioinformatics tools can help analyze genome sequences and predict gene structures, experimental approaches are needed to discover gene functions through verification *in vivo* using genetic analysis. Reverse genetics is a powerful tool that establishes a direct link between the biochemical function of a gene product and its role *in vivo*. Several approaches have been developed for plants, some of which are applicable to many species and each of which has advantages and limitations.

REVERSE GENETICS: ENGINEERING LOSS AND GAIN OF FUNCTION

Several plant genomes have been already sequenced and in the few next years, we will be in front of many others fully-sequenced genomes (which are now in progress). The functions of a small number of genes were assigned, but those of many of them remain unclear. The current and future challenge is to investigate the biological function of identified and annotated gene sequences.

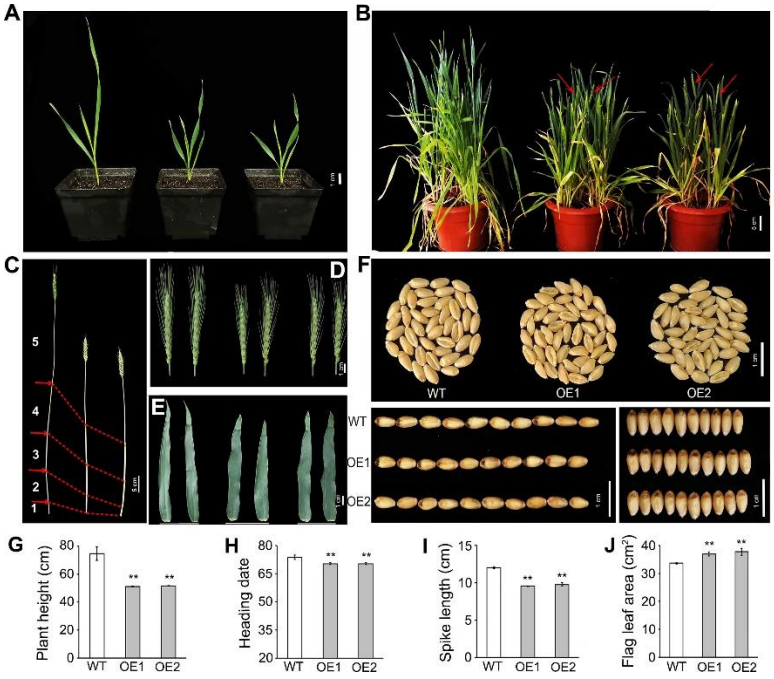
Classically, genes are transcribed into mRNA and translated into proteins determining the phenotypic traits. Each **disruption in the genome (DNA sequence) can affect indirectly physical traits via the alteration of protein synthesis or activity**. With the advent of recombinant DNA technology, gene engineering by **gain- or loss-of-function** approaches that **mutate or knock out the gene function** will switch on or off protein synthesis/activity **resulting in new phenotypes**. For example, changes in regulatory gene sequences or site-directed mutations targeting the open reading frame of gene-coding protein could identify amino residues for protein function.

Reverse genetics experiments usually start with a cDNA, often corresponding to a transcript with an interesting pattern of expression, and then attempt to ascribe a biological function associated with a phenotypic trait to the target gene. The function is generally investigated by **using the cDNA sequence to create a mutation** of the wild-type allele in a transgenic plant (**Ectopic expression using transgenics strategy**). This mutation can be made by insertion of cDNA in a specific site (**Insertional mutagenesis**) or by point mutation through induced lesion in the genome (**TILLING:** Target Induced Local Lesions in Genome), or by deletion (**Knock-out**). The phenotypic alteration will be further investigated. The target gene can be **down-regulated** by the expression of an antisense mRNA or RNA interference (**Gene silencing**) or virus-induced gene silencing (**VIGS**) which interferes with the gene transcript causing RNA degradation or inhibiting translation. These approaches are used to **deactivate transcripts of unknown genes** and then check the phenotypic changes. Another way to elucidate the gene function is to **exchange wild-type allele with an inactive mutant allele by homologous recombination (Gene targeting)**. These methods provide a way for linking genotype to phenotype to analyze the function of unknown genes for both fundamental and practically oriented studies. Such tools each with its own strengths and weaknesses, try to explore gene function by analyzing the phenotypic effects of specific engineered gene sequences and seek to find what phenotypes arise as a result of particular genetic sequences.

In most instances, genes of known sequence are not associated with a phenotype. This is particularly true in non-model species where forward genetics can be more challenging due to genetic redundancy. Reverse genetics is a powerful tool that can be used to identify the phenotype that results from disruption of a specific sequenced gene, even with no prior knowledge of its function. Several approaches have been developed in plants that have led to the production of resources including collections of T-DNA insertion mutants, RNAi-generated mutants, and **populations carrying point mutations that can be detected by TILLING, direct sequencing or high resolution melting analysis**. These reverse genetics resources allow for the identification of mutations in candidate genes and subsequent phenotypic analysis of these mutants.

1) Transgene-induced Ectopic Expression

Large number of plant genes continues to be identified every day with no defined functions. Although powerful in silico techniques can predict the function of many gene-encoding proteins, evidence of gene function have to be constantly verified in vivo using transgenic approach allowing analysis of phenotypic changes occurred by transgene integration and expression. The production of transgenic lines which ectopically express recombinant genes or those in which endogenous genes are knocked down remains a major bottleneck. Gene transfer technology appears to be an essential tool implemented for functional studies through gain- or loss-of-function approaches. Overexpression of a wild-type gene product, however, can cause mutant phenotypes, providing geneticists with an alternative yet powerful tool to identify pathway components that might remain undetected using loss-of-function screens. Gain-of-function is achieved by increasing gene expression level through the activation of endogenous genes by transcriptional enhancers or through ectopic expression of individual transgene via transformation. Heterologous expression has also been exploited to study gene functions across species barriers and transcript abundance is increased by cloning the full open reading frame (ORF) downstream of a **strong constitutive promoter** such as 35S promoter of cauliflower mosaic virus (CaMV35S), or using **chemical- or stress-inducible promoters** and **transcription factors** to control ectopic expression level.



Overexpression of *TaLBD16-4D* alters plant architecture and heading date in transgenic wheat

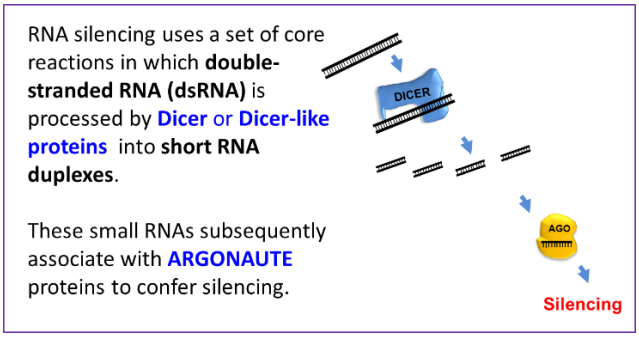
Comparisons of several important traits between the two *TaLBD16-4D*-overexpression transgenic plants (OE1 and OE2) and wild type (WT). **(A)** Photos of WT (left) and two OE lines (middle and right) taken when OE lines reached the tillering stage. **(B)** Phenotypic characteristic of WT (left) and two OE lines (middle and right) taken when OE lines reached the heading stage. Red arrow heads indicate spikes at the heading stage. **(C)** Main culms of WT (left) and two OE lines (middle and right). **(D)** Main spikes of WT (left) and two OE lines (middle and right). **(E)** Flag leaf of WT (left) and two OE lines (middle and right). **(F)** Morphology of grains in the WT and two OE lines. Comparison of plant height **(G)**, heading date **(H)**, spike length **(I)**, and flag leaf area **(J)** in WT and two OE lines, respectively.

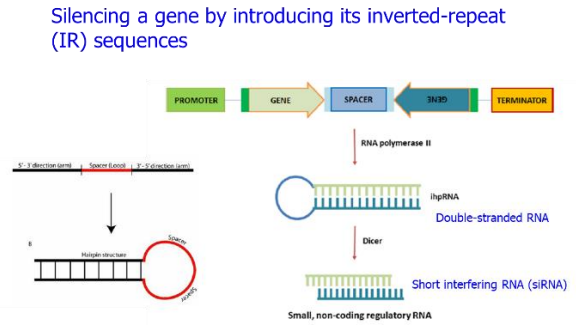
Several approaches were used for gene transfer to plants, depending on the fact of how the gene construct is expressed in a stable or in a transient manner. Over the last decades, hundreds of model and crop plant species have been genetically transformed essentially via Agrobacterium- or viral-mediated gene transfer technologies as well as by direct gene delivery using bombardment of DNA-coated particles or protoplast electroporation. Methods for effective DNA transfer into regenerable competent cells were extensively developed and optimized for stable transformation and generation of transgenic plants. However, transgenic approaches for reverse genetic studies are not

yet practical in several plants in which transformation method-ology is so far not efficient or not available. Heterologous expression approach provides a solution for the high-throughput characterization of gene functions in these plant species.

(2) RNA Silencing Approaches for Functional Gene Analysis in Plants

RNA-mediated Interference or RNA interference (RNAi) is induced by **double-stranded RNAs** (dsRNAs) leading to specific **post-transcriptional gene silencing** (PTGS) mechanism. This may be caused commonly by the expression of sequence homologous to an endogenous gene or by the expression of transgene that include a segment of gene sequence in an **inverted repeat orientation** that will generate dsRNAs. The amplification process of RNAi involves **RNA-dependent RNA polymerases** (RdRPs) that are required for the RNA silencing pathways. Besides its antiviral functions as part of the host defense response evolved to control plant virus replication, RNAi process fulfils fundamental regulatory roles through the activities of **microRNAs and small interfering RNAs**.





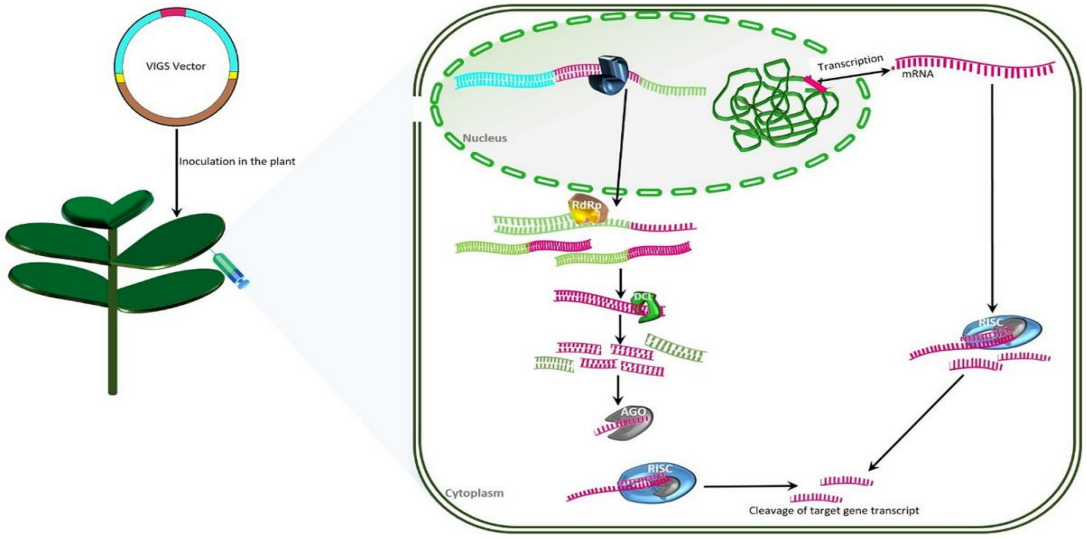
Previous studies on gene silencing have revealed two RNA-mediated epigenetic processes, **RNA-directed RNA degradation** and **RNA-directed DNA methylation**, providing new avenues for gene suppression technology in plants. There is increasing evidence that components of the RNAi machinery are associated also with the **formation of heterochromatin** and that **RNA-mediated chromatin modifications** play an important role in **epigenetic transcriptional gene silencing**. Since RNAi can specifically suppress the function of the targeted gene, the technique has been very useful in functional genomics studies. RNAi have been widely described in plants and this is currently one of the hottest areas of biological research in the new era of functional genomics since interfering RNA-mediated silencing has revolutionized the understanding of gene function for basic and applied studies.

Transgene-induced RNAi has been effective at silencing gone or more genes in a wide range of plants across the plant kingdom. The principle is simple: **a fragment of a gene is introduced into a cell as dsRNA** or as **DNA that will give rise to dsRNA** that would trigger silencing of endogenous genes in a homology dependent fashion. The dsRNA activates the **DICER/RISC process** so that the properties of the affected cell reflect a loss of function in the corresponding gene. In this way, several functional genomics projects attempted to generate lines that are deficient for the activity of a subset of genes, and check their phenotypes to characterize the function of the knocked down gene. A particularly useful property of RNA silencing is that it does not require neither full sequence of the target gene, nor complete sequence identity in the dsRNA and the target RNA.

Since **RNAi is highly sequence-specific**, it is possible to knockdown simultaneously multiple closely related genes by targeting their conserved sequences. For that reason, transgene-induced RNA silencing was implemented since years as a versatile reverse genetics tool and is well suited to the systematic analysis of gene function. RNAi technology has recently become a highly effective and powerful tool of functional genomics to explore plant genome and has been applied to elucidate gene function in various crop species like tomato, potato, maize, wheat and rice.

(3) Virus-induced Gene Silencing Strategy

Another alternative to knock-down endogenous plant genes may be achieved via viral vectors through RNA-mediated post-transcriptional gene silencing mechanism. This powerful reverse genetics approach, known as Virus-induced gene silencing (VIGS), is a virus vector technology that exploits RNA-mediated anti-viral defense mechanism. Production of dsRNA activates the RNA silencing pathway, resulting in down-regulation of the host gene transcript. VIGS offers an easy way to test the function of several genes in a short time, as it only requires a fragment (typically 300-800 bp) of the target plant gene inserted into a suitable viral vector to form a recombinant virus. Upon infection of a plant host usually via Agrobacterium tumefaciens, this recombinant viral vector induces PTGS targeting both the virus RNA and homologous endogenous plant RNA sequences for degradation. Besides the lack of suppressors of gene silencing needed in VIGS strategy, successful RNA virus-based VIGS requires simultaneous infiltration of both viral clones including genomic RNA1 and RNA2.



Nowadays, there are infectious clones of several plant viruses that have been used as VIGS vectors, most of which have RNA genomes. Other source of VIGS vectors used recently for silencing is the small subviral RNA satellite along with helper virus-dependent replication. While several plant viruses have been so far developed into VIGS vectors, the Tobacco Rattle Virus (TRV) established by David Baulcombe's group provides the most robust results in terms of efficiency, ease of application, and absence of disease symptoms. VIGS has been broadly regarded as the tool of choice for transient induction of silencing that occurs for only few weeks with further decrease resulting in plant recovery. However, recent evidences suggest that using some vectors under specific conditions, VIGS can persist for years or even transmitted to progeny behaving like stable transgenic plants. Besides many advantages of VIGS over other approaches including methodological simplicity, switching off genes specifically, rapid monitoring, robustness and speedy results, the system allows the study of genes whose functions are essential to plant viability. VIGS has been widely used in model plants as tool to assess function of candidate genes and to discover new genes required for diverse pathways. VIGS technology has been successfully used to validate and functionally analyze the contribution of candidate genes in many plant species. For instance, several stress-responsive-genes were studied using VIGS system in various model plants including tobacco, chili pepper, soybean, rose and wheat. Such strategy can be also very helpful to assess gene function, especially in species recalcitrant to transformation making the VIGS an attractive alternative instrument for high-throughput functional genomics.

(4) Insertional Mutagenesis (by T-DNA and Transposable Elements)

Insertional mutagenesis is an alternative means of disrupting gene function and is based on the insertion of foreign DNA into the gene of interest. With the development of high efficient *Agrobacterium*-mediated transformation of model plant Arabidopsis and crop plant rice, T-DNA mutagenesis has become a major method to generate a large collection of insertion mutants. Generally, T-DNA can be randomly and stably inserted into plant genome, which made it possible to generate a population saturated with insertions, i.e. having at least one insertion in each gene.

Insertional mutagenesis is a mutation caused by insertion of new genetic material (foreign DNA) into the target gene. Both T-DNA and transposon insertional mutants are being produced recently as an extremely valuable research tools for model plant systems to study gene function. Insertional mutagenesis has been developed within the last decade to generate specific mutations in organisms in which homologous recombination is a low frequency event. The most effective method for insertional mutagenesis is targeted gene disruption. Large populations of T-DNA-tagged lines or mutants with transposon activation having an insertion at unique site in the genome have been generated in algae and model plants. The approach relies on the generation of thousands of transformants followed by PCR-based screenings that allow for identification of lines harboring the introduced mutation within specific genes of interest. Defining the insertion site for each transformant has allowed for the establishment of sequence-indexed libraries of mutant plants. T-DNA or transposon insertion has been exploited to create disruptions in target genes of interest, introduce new genes, or activate endogenous genes in the plant genome. Transposons have several advantages over T-DNA including the ability to produce multiple independent insertion lines from individual starter lines, as well as producing revertants by remobilization.

A. T-DNA insertional mutagenesis

In Arabidopsis, this involves the use of either **T-DNA or transposable elements**. The foreign DNA not only disrupts the expression of the gene into which it is inserted but also acts as a marker for subsequent identification of the mutation. Because Arabidopsis introns are small, and because there is very little intergenic material, the insertion of a piece of T-DNA on the order of 5 to 25 kb in length generally produces a dramatic disruption of gene function. If a large enough population of T-DNA–transformed lines is available, one has a very good chance of finding a plant carrying a T-DNA insert within any gene of interest. Mutations that are homozygous lethal can be maintained in the population in the form of heterozygous plants. Loss-of-function mutations achieved by T-DNA insertion or MGE insertion provide practicable methods to identify the genes disrupted by these transfer elements because their insertion sites can be explored by PCR base methods such as tail-PCR, plasmid rescue, inverse PCR or adapter PCR. *Arabidopsis* is currently the only multi-cellular organism reported possible to achieve saturation of mutation with accurate index for each genes because *Arabidopsis* owns several crucial advantages than other species: genome with small size, easily to be transformed, self-pollination, short life cycle and bulk storage of seeds. Rice is the second better learned species that has been widely used for establishing systematic insertion mutant libraries because rice is one of the most important crops and easy to be transformed by *Agrobacterium* mediated transformation.

In *Arabidopsis*, at least 225,000 independent T-DNA insertion lines have been created that represent near saturation of the gene space; the precise locations were determined for more than 88,000 T-DNA insertions, which resulted in the identification of mutations in more than 21,700 of the approximately 29,454 predicted *Arabidopsis* genes.

Saturation mutagenesis by T-DNA insertion is one of the most successful approaches for analysis of systematic gene functions in the past decade years benefitting from the continuous improvement of transgenic techniques. The *Agrobacterium* vacuum infiltration method for *Arabidopsis* transformation was developed to avoid the complex transgenic process of tissue culture based method including introduction of DNA by particle bombardment or *Agrobacterium* and plant regeneration. This method was further modified and the transgenic *Arabidopsis* can be easy achieved by simply dip the floral tissue into a solution containing sucrose, surfactant Silwet L-77 and *Agrobacterium tumefaciens* carrying target genes. The basic scheme for the generation of T-DNA insertion mutants by flora dip method for *Arabidopsis* is shown in Figure 1. *Agrobacterium* harbouring the binary vector was used to transform *Arabidopsis* by floral dip method. The T-DNA region between left border (LB) and right border (RB) of the binary vector was randomly transformed into the genome of host plant. The transgenic plants can be screened

by selection marker depends on the T-DNA vector used (For example: *NPTII* is the selection marker of pROK2 T-DNA vector which is used to generate Salk insertion mutant lines). The flanking sequence at both sides of insertion can be further sequenced by PCR based method. Taken the advantage of flora dip method, a vast number of T-DNA insertion lines have been generated which represent almost saturated insertions into *Arabidopsis* genes in past ten years.

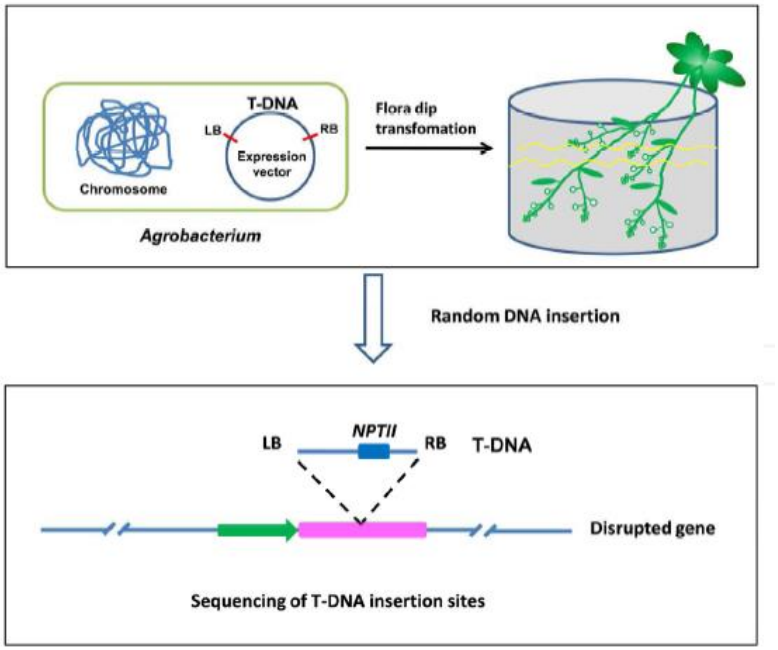


Fig. 1. A basic scheme for the T-DNA insertion method by flora dip in *Arabidopsis*

Saturation of the Arabidopsis genome with T-DNA insertions is an experimental goal that requires the actualization of specific quantitative considerations. To date, the quantitative exigencies associated with mutational saturation of the genome by T-DNA have not been fully satisfied. Nevertheless, we have recently established a population of 60,480 T-DNA–transformed lines as a significant step toward the production of genomewide mutations. Access to these lines is now available through the **Arabidopsis Knockout Facility at the University of Wisconsin** (http://www.biotech.wisc. edu/arabidopsis/default.htm). This facility will serve the research community by allowing users to screen the entire population of lines for the presence of a T-DNA insert within their gene of interest. The organization of this population of 60,480 lines, as well as the operation of the service facility.

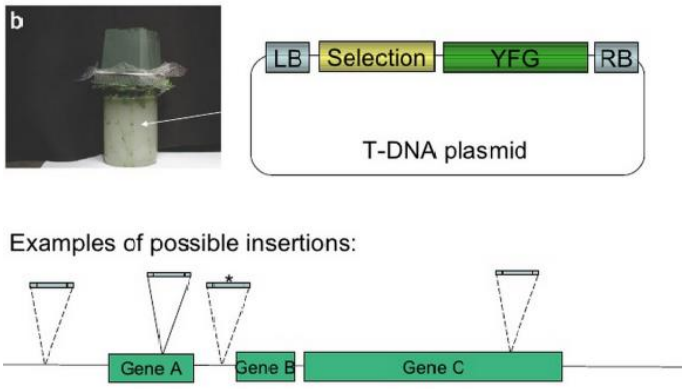


Fig. 2. T-DNA insertion at random locations in Arabdopsis plant genome to generate saturation mutations

In rice, several research groups have contributed to the generation of T-DNA insertion lines. For example, An's group has generated approximately 100,000 insertion lines. Around 42,000 T-DNA insertion lines have been generated by Zhang and Wu's group. Recently, Hsing et al. (2007) have reported the generation of 55,000 T-DNA insertion lines. Several other groups also independently produced T-DNA insertional mutant lines in rice. According to the previous reports, the average copy number of T-DNA inserts per line is 1.4-2.0. Thus, more than 450,000 T-DNA tags have now been generated in rice. Recent progresses on the generation of T-DNA insertion lines have been reviewed by several researchers. If there are only 30,000 or less protein-coding genes in rice genome, these populations are large enough to find a knockout in a given gene, assuming that T-DNA is randomly inserted into a chromosome. This suggestion was strengthened by the fact that T-DNA have been observed to insert preferentially in gene-rich regions.

After T-DNA insertion, various phenotypes have been observed including changed growth rates, different plant statues, pollen and seed fertility and so on. Those visible differences in phenotypes could significantly contribute to the identification of gene functions. Since the establishment of T-DNA insertion populations, at least 43 genes have been functionally characterized by T-DNA insertion mutants. For example, a knockout line of *OSMADS3* by T-DNA insertion shows homeotic transformation of stamens into lodicules and ectopic development of lodicules in the second whorl near the palea where lodicules do not form in the wild type but carpels develop almost normally. Their data show that this gene plays a crucial role in regulating stamen identity.

B. Insertional Mutagenesis by Mobile Genetic Elements (MGE), insertion through T-DNA

In addition to T-DNA insertion lines, mobile genetic elements (MGE) insertion is also a popular approach to generate large number of mutations. MGEs which can move around within the genome include several kinds of mobile DNA elements such as **transposon** or **retrotransposon**. Transposons describe the DNA which can be cut away from one site and paste to other place within the genome. Retrotransposon however, make themselves a copy and then paste to other position within the genome.

Transposons:

Several transposable elements identified in maize have been used to obtain large population of insertions in genes for functional genomics studies. For example, the maize transposable element *Activator* (*Ac*) first identified by McClintock (Mc 1950) is a kind of transposon widely used for creating MGE insertions. *Ac* element can insert themselves into genes and cause insertion mutations to create a recessive allele. The mutations caused this way are **unstable** because the ***Ac* element** can be excised from the inserted gene by the transposase which is coded by Ac element itself. *Dissociation* (*Ds*) element is usually stable because they are incapable of excising itself from the inserted gene unless with the help of *Ac* element.

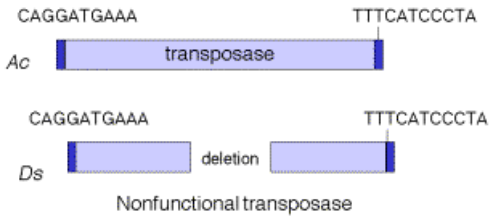


Fig. 3. Maize autonomous (*Ac*) and nonautonomous (*Ds*) transposable elements

The *Ac* element encodes a transposase that **binds to the terminal inverted repeat ends of both *Ac* and *Ds* elements**, catalyzing their transposition to new locations in the genome. *Ds* elements are most often derivatives of *Ac* that have lost the ability to produce a transposase but retain the terminal inverted repeats. The *Ac* transposase, when produced in *trans*, is able to recognize the ends of *Ds* elements and catalyze their movement to new chromosomal locations. The use of a two-component system allows for stable insertions to be generated, because the autonomous element can be segregated away from the insertion.

For obtaining **stable insertion lines**, researchers combine these two mobile elements and named it as the *Ac/Ds* system to generate mutant populations. In this system, the individual *Ds* parental lines and *Ac* parental lines are created by transformation of *Ac* element and *Ds* element independently into the host organism. One parental line contains *Ac* element, in which ***Ac* is immobilized and provides only *Ac* transposase** under the control of 35S promoter. **This Ac element lacks one of the Ac termini** and therefore cannot transpose, but it supplies transposase constitutively. Another parental line is transgenic *Ds* plant, in which *Ds* is also non-autonomous element and **provides only two wings of *Ds* element (5' *Ds* and 3' *Ds*)**. Then, two different transgenic parental lines are used for sexual crossing (Figure 4). Thus, in both *Ac* and *Ds* parental lines transposon *Ac* or *Ds* can't mobilize by themselves. However, after crossing between *Ac* and *Ds* plants, *Ds* element will be induced to translocate from one position to another position within the genome, and get inserted to different genome positions under the presence of *Ac* transposase. In the next generation, these lines containing only *Ds* element and without *Ac* transposase were selected. Therefore these *Ds* insertional lines will be stable since the plants contained no *Ac* transposase. Besides *Ac* and *Ds*, other transposons such as *En* and *Spm* were also used to generate transposon insertion mutants.

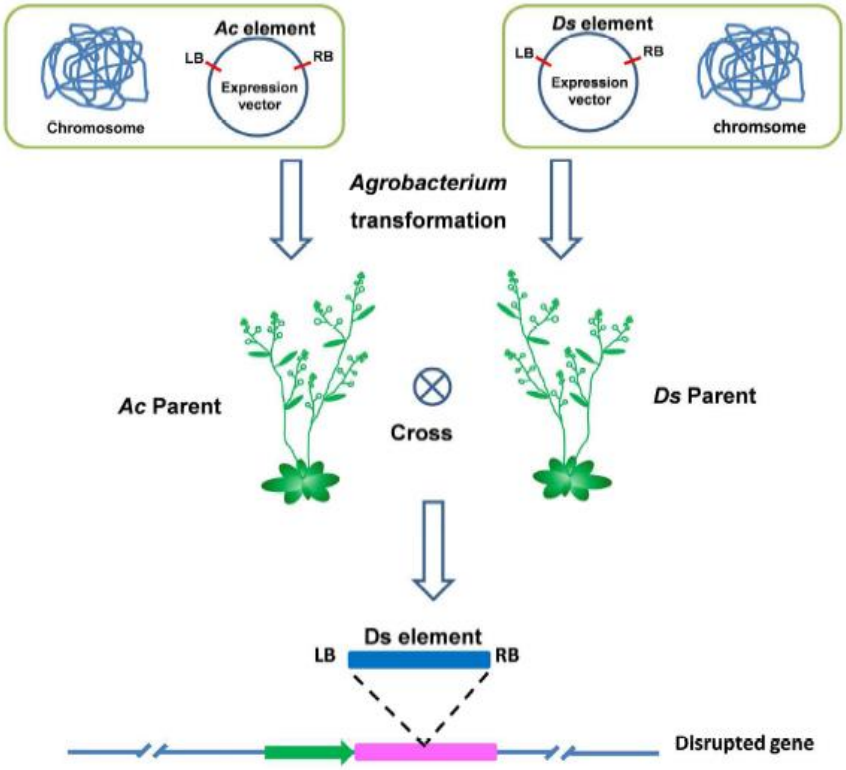


Fig. 4. Scheme for the generation of insertion mutant lines by *Ds* transposon

Selection schemes for transposed elements

For transposon tagging, transposase genes and non-autonomous elements are separately introduced into *Arabidopsis* by T-DNA transformation. Transformants are selected that have the transposon-bearing T-DNA integrated at a single locus. Mutagenesis is typically initiated by crossing homozygous plants that have non-autonomous elements with homozygous plants that have transposase genes. In the F1 progeny of these crosses, non-autonomous elements will transpose to new locations during development. Normally, CaMV35S promoter-Ac transposase fusion is used that generates high Ds excision frequencies in trans.

In order to recover stable germinal insertions, F1 progeny are self-pollinated, and F2 plants are selected that have inherited a transposed element, but have lost the transposase gene by segregation. Loss of the

transposase is important because unstable transposed elements will revert or transpose germinally causing confusion in progeny analysis. They will also revert or transpose somatically, which could obscure some mutant phenotypes caused by insertion. These problems can be avoided by selecting against the T-DNA that carries the transposase gene, typically by using a negative selectable marker on the T-DNA. There are several ways in which the F2 progeny of a mutagenesis cross can be enriched for those progeny that carry stable transposed elements.

For example: Mutagenesis is initiated with the ***Ds*** or *dSpm* **element located in the untranslated leader of an antibiotic resistance gene** (Figure 5). Chloroplast biogenesis in plants is sensitive to growth on medium supplemented with some antibiotics, resulting in bleached inviable seedlings. Excision of the transposon from the antibiotic resistance gene results in restoration of marker gene expression, so that green seedlings can be selected for antibiotic resistance. **By using a different antibiotic resistance gene within the transposon, seedlings that have the transposon integrated at a new site can be selected by including both antibiotics in the selection medium**. This type of selection scheme has been widely used in plants and has been generally successful.

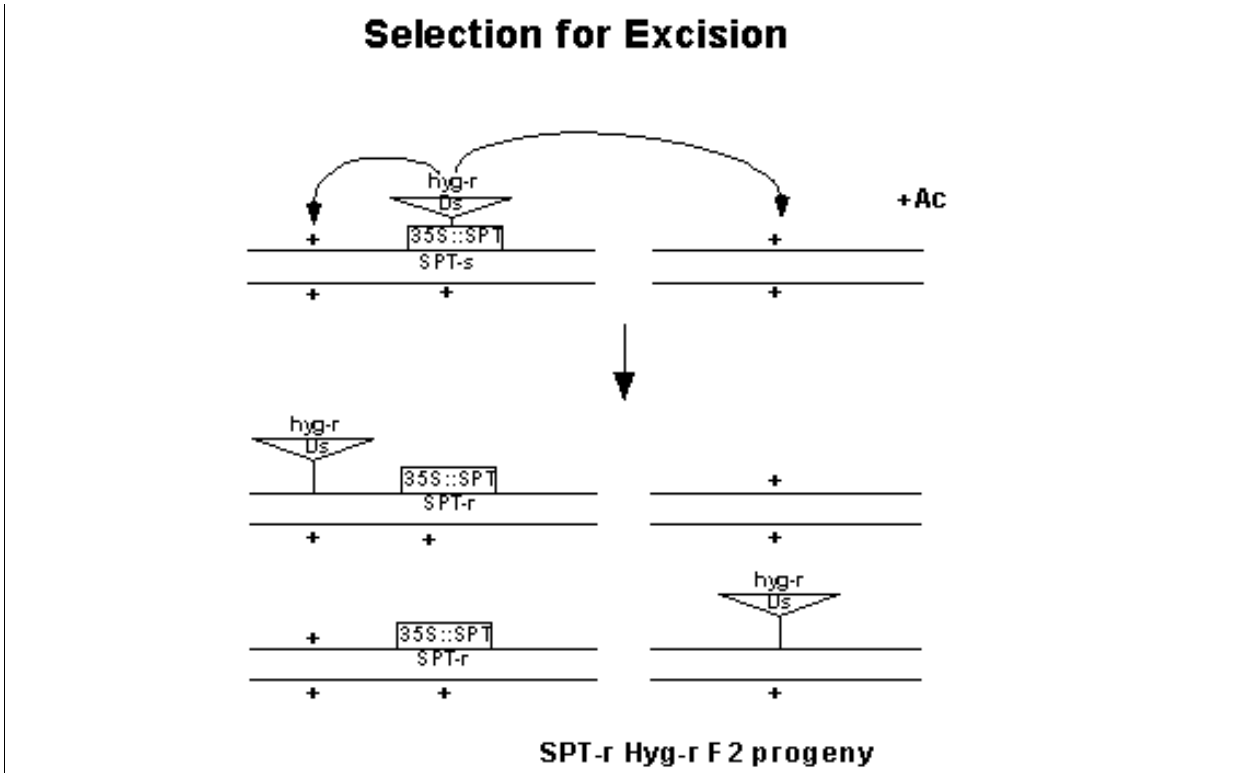


Fig 5. Selection schemes for transposed elements.

A. Selection for excision. Donor *Ds* element carries a hygromycin resistance gene and interrupts a streptomycin resistance gene. In the presence of *Ac*, the *Ds* excises, resulting in streptomycin resistance. F2 plants that are resistant to streptomycin and hygromycin have a *Ds* element that has excised from its original location. Both linked and unlinked transpositions are recovered.

With the *AC/Ds* or other similar MGE systems, several research groups have generated mutant resources with a high proportion of single-copy transposon insertions.

Retrotransposons:

Retrotransposons are one of the two major groups of transposable elements detected both in animal and plant genomes at the last decade and defined according to their mode of propagation. They differ from other transposons (class II elements that use DNA in movement such as Ac/Ds and En / Spm) by their ability to transpose via an RNA intermediate (also termed class I elements). Retrotransposons are mobile genetic elements that **replicate through reverse transcription of a messenger mRNA** intermediate forming a cDNA daughter copy.

Retrotransposons were first characterized in animal and yeast genomes, but evidence has accumulated in recent years to show that they are present in high copy number in many plant genomes and can constitute a very large part of some of them, especially those with large genomes. As one of the most fluids of genomic components, varying greatly in copy number over relatively short evolutionary timescale they represent one of the most important factors affecting the structural evolution of plant genomes, especially those of the higher plants. The basic parameters for these mobile elements varying during the evolution of higher plant genomes are the population structure (how many copies of which kinds of retrotransposons are present) and transpositional activity (what proportion of which transposons are active and what are their transposition rates).

Regulation of the activity of retrotransposons

The expression of retrotransposons in animal and yeast is under the control of hormonal and developmental factors. A general picture for the regulation of the expression of retrotransposons in plants is not yet fully established because of the absence of comparative studies in different plant tissues. Transcription of most of the active plant elements characterized to date is largely quiescent during normal development but can be induced by biotic and/or abiotic stresses, including cell culture, wounding, and pathogen attack.

(i) Developmental regulation

The developmental regulation of the expression have been showed for Tnt1, which is only expressed in roots at very low levels (69); for Tto1, Tos10 and Tos17, which are not expressed in leaf tissues and for mobile B5, Hopscotch, Stonor and Magelan elements, which are not expressed in plant tissues. Expression of the maize Prem-2 element has been detected only in early microspores. The expression of maize Opie, Huck and Cinful retroelements and barley BARE-1 has been observed in leaf tissues.

(ii) ***Activation of retrotransposons by stress factors***

Retrotransposons as other transposable elements are the major source of genetic variation that can ranges from gross chromosomal alterations up to very fine tuning of the expression of cellular genes. Evaluation the potential of activating retrotransposons through plant tissue culture to generate new insertion mutations into the genome is of a great importance for improving the present-day plant species. A common feature of the most retrotransposons is their activation by stress and environmental factors. The most well characterized plant LTR-retrotransposons (Ty1-copia subgroup) are particularly affected by protoplast isolation or in vitro cell or tissue culture.

The fact that some retrotransposons can be activated in their host species after protoplast isolation as well as cell and callus culture leads to the question as to whether their expression is linked to the activation of cell division programmes or to the activation of stress responses or to both. Protoplast isolation and cell or callus culture are the major inductors of modifications of the cell metabolism and gene expression. In the leaf-derived protoplasts, the former metabolic activity of the leaf cell is replaced by a new programme which is characterized by activation of growth- and stress-related genes (defense genes, which are activated after pathogen attack). Growth-related genes probably involved in the re-initiation of cell division and the activation of stress-related genes might be a consequence of the original wounding. Protoplast isolation included enzymatic degradation of cell wall from pathogenic compounds present in the extracts of phytopathogenic fungi. The activation of stress related genes as a consequence of protoplast isolation might be also a result from cell wall hydrolysis or from pathogenic compounds of fungal extracts. Growth- and defense-related genes are also expressed in callus and cell cultures, suggesting that callus tissue culture programmes are similar to those induced after wounding and during callus formation in the plant, involving both a stress response and cell division.

An **insertion of retrotransposons into coding sequences** after protoplast or cell tissue tobacco and rice cultures has been observed, indicating that they might take a **significant contribution to somaclonal variation**.

Tos17 is one kind of copia-like retrotransposons in rice (Hirochika et al. 1996), which can duplicate and paste to elsewhere in the genome. *Tos17* owns several special features that make it suitable for engineering large scale insertion mutagenesis:

a. The copy number of *tos17* is quiet low, ranging from one to five among rice cultivars. For example, the genome of cv. Nipponbare, the selected cultivar for the IRGSP (International Rice Genome Sequencing Project) (Sasaki and Burr 2000), contains only two native copies of *tos17*.

b. Transposition of tos17 is inactive under normal conditions but only activated in the callus by tissue culture and then becoming stable again in the regenerated plants.

c. The transposition site of *tos17* prefers gene-dense regions over centromeric heterochromatin regions with a three times higher insertion frequency in genic regions than in intergenic regions.

d. Its size is just a little bit over 4kb and its insertion sequence is clearly known for flanking sequencing.

Tos17 is stably present in genome during the normal life cycle of rice. By the tissue culture of the rice callus, the transcription of *tos17* is activated and the reverse transcript DNA fragments are integrated into new places in the genome which creates disruption of genes. The original *tos17* and its duplications become silence again in the regenerated insertion mutant plants (Figure 6). Taking these advantages of *tos17*, Large-scale *tos17* T-DNA mutant library of Nipponbare has been established by tissue culture and stably preserved by normal generation.

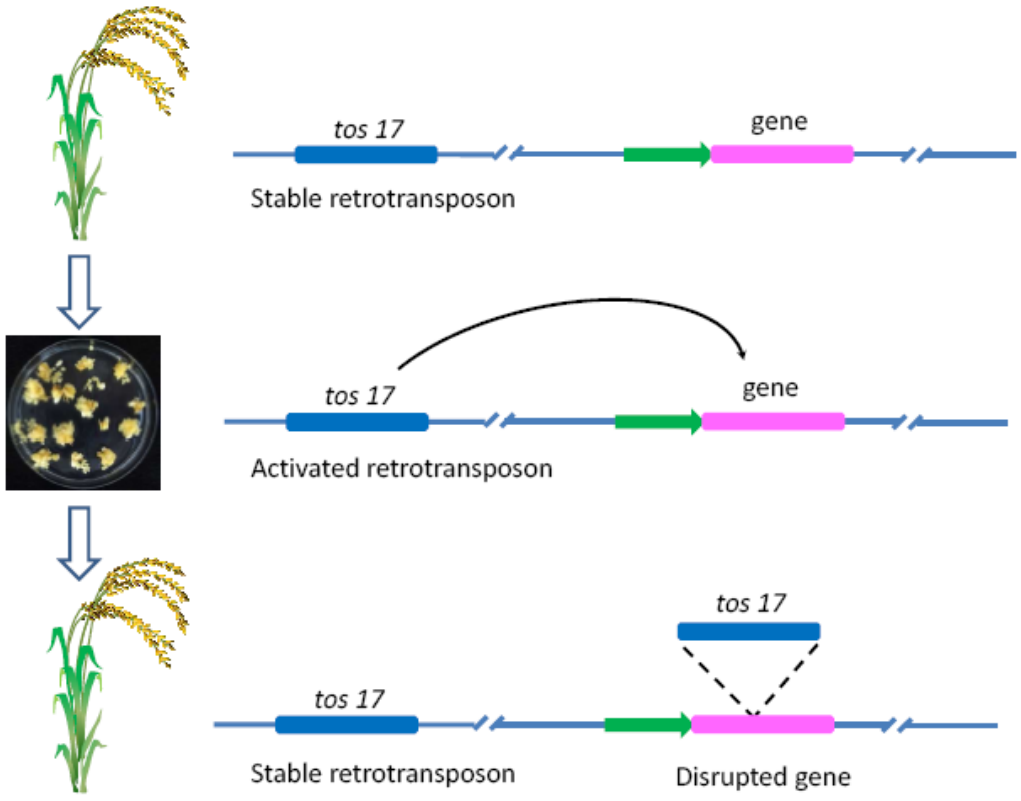


Fig. 6. Scheme for the generation of *tos17* insertion mutant lines by tissue culture.

Polymerase chain reaction (PCR) methods have been developed that allow one to easily isolate individual plants that carry a particular T-DNA or transposon mutation of interest (Figure 7).

Transposon mutagenesis is widely employed in the insertional disruption procedure in the unknown gene identification process typically composed of random mutagenesis with insertion of transposon, selection of a phenotype of interest, plasmid rescue of the inserted region, and DNA sequence analysis.

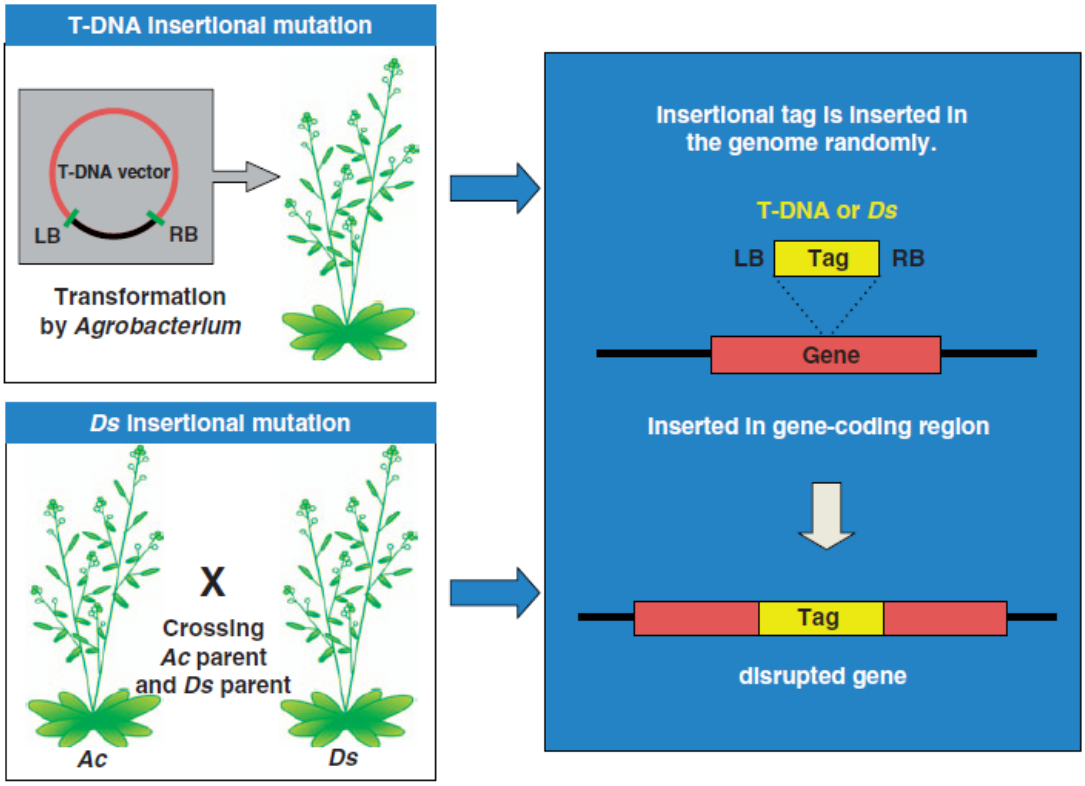


Fig. 7 Scheme for the generation of insertional mutants by T-DNA or a Ds transposon.

To construct T-DNA insertional mutant lines, plants are transformed by Agrobacterium harboring the T-DNA vector, which is inserted randomly into the genome. To construct Ds insertional mutant lines, Ds parental lines are crossed with Ac parental lines, hereupon the Ds transposon is translocated into the plant genome in the next generation. When the tag is inserted into a gene-coding region, that gene will be disrupted and lose its function.

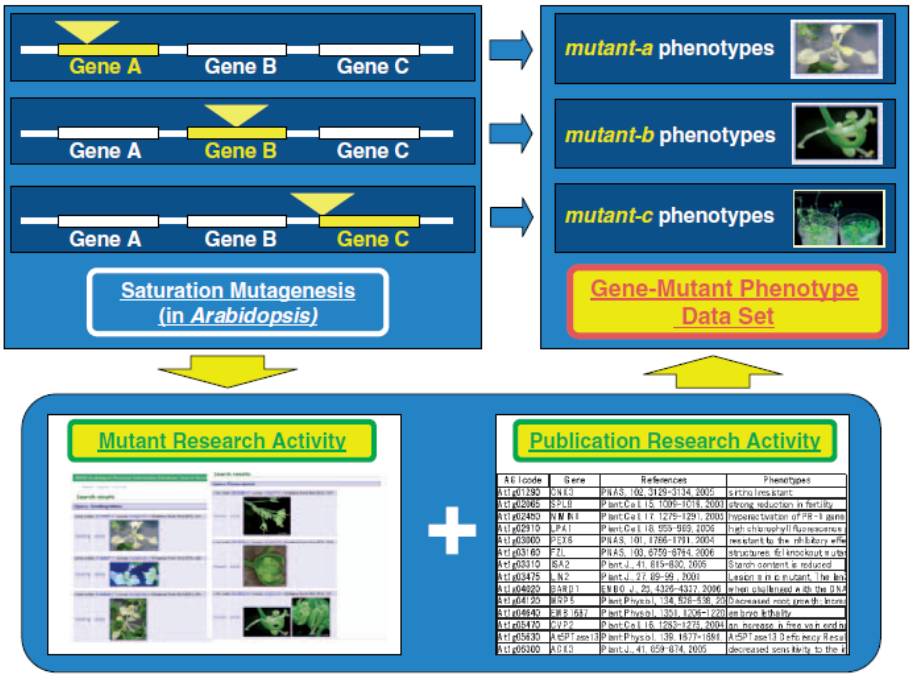


Fig. 2 The concept of gene-based phenome analysis and construction of a comprehensive gene–mutant phenotype data set.

One of the aims of phenome analysis is to identify the mutant phenotype for every gene and to establish a complete list of genes and mutant phenotypes. Phenotyping data of mutant resources and mutant phenotype information extracted from the published literature can be combined to make a comprehensive gene–mutant phenotype data set.

Advantages and Limitations of these Tools

An **advantage of using T-DNAs as the insertional mutagen, as opposed to transposons,** is that T-DNA insertions will not transpose subsequent to integration within the genome and are therefore chemically and physically stable through multiple generations. The mobility of transposons is not necessarily a bad thing, however. In situations in which multiple members of a gene family are arranged in tandem along a chromosome, the ability of transposons to "hop" to nearby locations provides a convenient method for creating mutations within all of the members of the gene family within a single plant.

Although transferred DNA **(T-DNA) tagging** or **transposon tagging** methods were developed to generate loss-of-function mutations because these tag sequences can be used to identify the genes disrupted by these elements. However, because many plant genes in *Arabidopsis*, rice, and other plants belong to gene families, the characterization of gene functions by single-gene mutagenesis is not always possible. Many mutants generated by single-gene disruption do not show clear phenotypes because of genetic redundancy.

A large proportion of genes in the *Drosophila melanogastor*, *Caenorhabditis elegans* and *Saccharomyces cerevisiae* genomes have no obvious loss-of-function phenotype. This phenomenon can be explained by two mechanisms of compensation. The first is the presence of duplicate genes, where loss-of-function in one copy can be compensated by the other copy or copies. The second mechanism of compensation stems from the existence of alternative metabolic pathways and regulatory networks. Therefore, eukaryotic organisms have a significant genetic robustness against null mutations in a large number of genes. A significant limitation of classical loss-of-function screens to dissect genetic pathways is that they rarely identify such genes. There is another tranche of genes that are also difficult to uncover by classical screens. These genes encode products that function at multiple stages during the life cycle of the organism, one of which is essential for early embryo or gametophyte development. Loss-

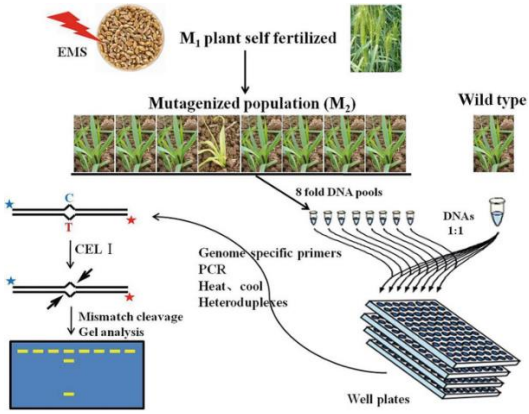
offunction mutations in these genes result in lethality. A number of approaches have emerged to help circumvent these potential problems. The enhanced expression of genes providing gain-of-function phenotypes has proved a productive strategy to identify gene function.

A particularly useful aspect of this technology is that it permits the identification of genes that would have been missed in conventional mutagenesis screens. Many genes have multiple roles, acting at early as well as late stages in development, so that their later functions are obscured in phenotypic screens but are revealed in screens for gene expression patterns. Second, if a gene is functionally redundant because of the presence of a second locus that can subsitute for the same function, inactivation of the gene will not result in any phenotype.

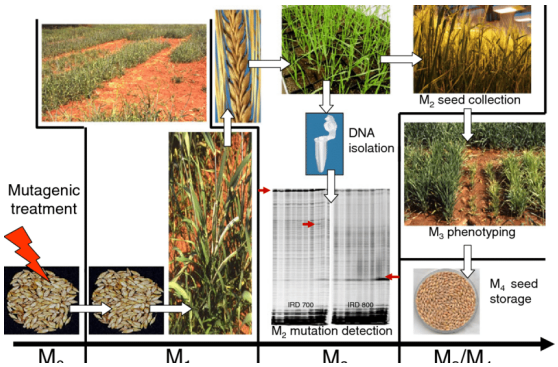
(5) TILLING

Targeting-induced local lesions in genomes (TILLING) is a strategy for the discovery and mapping of induced point mutations that was raised a decade ago as an alternative to insertional mutagenesis. This method was originally developed by the Henikoff laboratory to screen libraries of Ethylmethane sulfonate (EMS)-treated Arabidopsis for desired mutant alleles [169] and was subsequently adapted to others plant species. TILLING, which combines traditional mutagenesis with genome-wide high-throughput screening for point mutations in desired genes, is moving beyond functional genomics into crop improvement. This technique was developed as a high-throughput and low cost reverse genetic method that has been successfully applied to many plant species making the TILLING process broadly applicable. Besides model plants, the feasibility of TILLING has already been demonstrated for a large number of agronomically important crops, including rice, barley, wheat, maize, sorghum, soybean, rapeseed and tomato plants. Furthermore, this method does not require transformation procedures and thus it is suitable for recalcitrant species and recommended as non-GMO technology to avoid controversies. Recently, large-scale TILLING platforms for identifying mutants in plants offer great potential as a tool for functional genomics.

In general, a targeting DNA construct contains typically part of the gene to be targeted along with a reporter gene or/and a selectable marker. The transformed organism carrying a loss-of-function mutation can then be analyzed for its phenotype. Gene targeting is thus a powerful tool for directed 'knockout' of genes. Plant DNA-repair machinery predominantly uses non-homologous end-joining (NHEJ), making the homologous recombination (HR)-based methods, which have proved fruitful for gene targeting in non-plant systems, unsuitable for use in plants. Precise genome modification is an attractive method for understanding gene function. Gene targeting is known as one of the best method currently available to induce specific change of endogenous gene via HR. This strategy can be used to delete (knock-out) or substitute a gene (knock-in), remove exons, and introduce point mutations. Gene targeting can be applied to the whole organism or may be limited to a particular developmental stage and definite plant tissue. It can be used also for any gene, regardless of transcriptional activity or gene size.



Using TILLING Lines for Functional Genomics in Wheat



Development of a TILLING platform in barley

TILLING - IN THE ERA OF MODERN PLANT GENOMICS

Twinkle Kumari Bhagat*
Uttar Banga Krishi Vishwavidyala, Cooch Behar

Ajay Kumar Chandra
GBPUAT, Pantnagar

S. Pavithra
Uttar Banga Krishi Vishwavidyala, Cooch Behar

In general, the Targeting Induced Local Lesions IN Genomes (TILLING) is a powerful reverse genetic tool that employs non-transgenic techniques to combine chemical mutagenesis with modern PCR-based screening for detection of sequence variation in loci of interest. This technique enables the identification of SNPs and INDELS in a gene of interest that can help decipher the function of thousands of newly identified genes from a mutagenized population.

With the recent advancement of molecular biology, several new techniques have been developed to identify mutations to determine the function of genes. TILLING is an effective tool for functional genomic studies in several plants and animal species that can be mutagenized, despite its mating system, ploidy level, and genome size. This technique is well suited for plant species because they can be self-fertilized and their seeds can be stored for a longer duration. Many other reverse genetic tools have been used to bridge system biology with the resulting phenotype for gene knockout, RNAi, site-directed mutagenesis, and transposon tagging. But these approaches rely on the creation of transgenic which is not always feasible for many plants or animals. Thus, are less efficient with limited success and inefficient transgene expression in model species. Thus the Tilling has proven to be powerful and efficient tooling that ultimately gives rise to allelic series of induced mutations in a gene of interest.

TILLING: TOOLS AND TECHNIQUES

In essence, the tilling combines chemical mutagenesis with PCR-based screening to identify induced mutations that might alter the protein function. EMS is a widely used mutagenic agent, as it is stable and produces a high density of random mutations throughout the genome with very low aneuploidy and dominant lethality. Once a mutagenized population is created DNA samples are pooled together based on the ploidy level. Fluorescently labelled primers are being targeted for the gene of interest. SNP discovery methods used in TILLING include denaturing high-performance liquid chromatography (dHPLC), heteroduplex mismatch cleavage assay using CEL I endonuclease, detection of digested

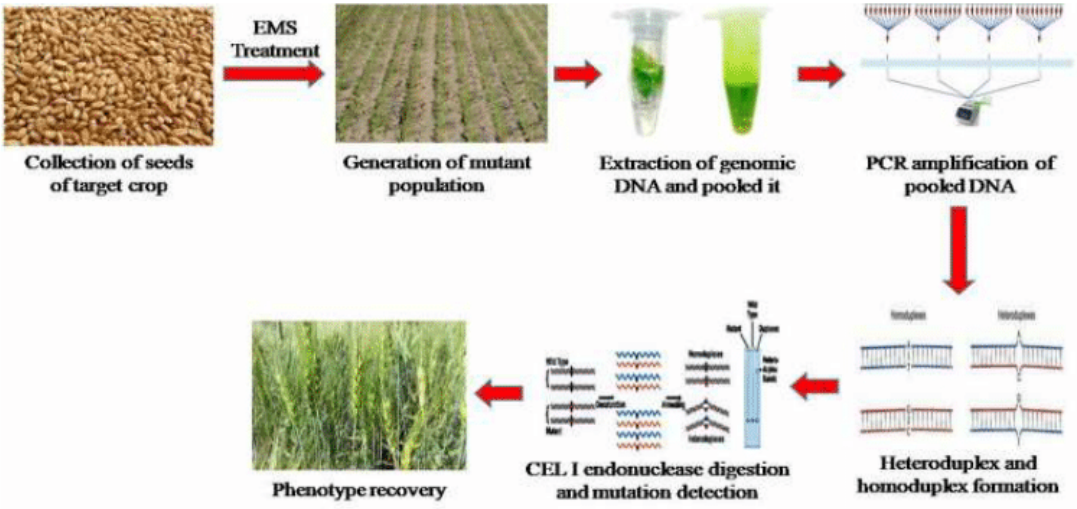
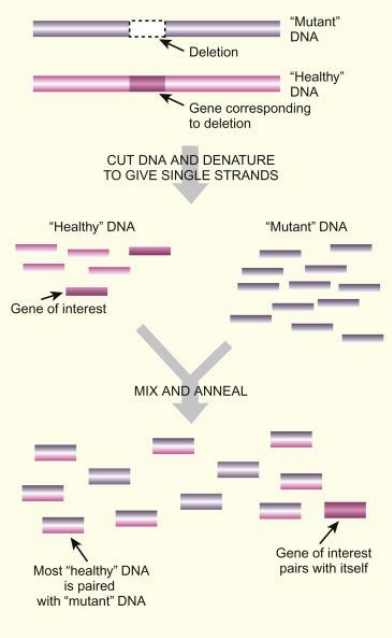


Figure 01: Systematic view of TILLING events for quality improvement in the model crop.

Subtractive Hybridization to detect differentially expressed genes

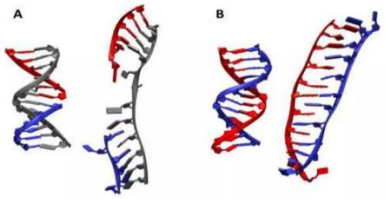
Several approaches are available for studying transcription profiling from early methods, such as Expressed Sequence Tags (EST) and Suppression Subtractive Hybridization (SSH), through to microarray technology and more recently Next Generation Sequencing (NGS) platforms. SSH is a powerful technique for the amplification of differentially expressed genes by simultaneous amplification and suppression of target and non-target sequences, respectively. Despite the advent of next generation sequencing, SSH is still a popular technique because of the relatively low cost, low amount of starting material and the relatively low rate of false positives. It is however quite laborious and costly if the number of sequenced clones is increased to several thousands.



Subtractive hybridization is a technique for identifying and characterizing differences between two populations of nucleic acids.

It detects differences between the RNA in different cells, tissues, organisms, or sexes under normal conditions, or during different growth phases, after various treatments (ie, hormone application, heat shock) or in diseased (or mutant) versus healthy (or wild-type) cells.

Subtractive hybridization also detects DNA differences between different genomes or between cell types where deletions or certain types of genomic rearrangements have occurred.



To validate the data obtained from the statistical analysis of the SSH libraries, a sqRT-PCR.

Subtractive hybridization is a technique used to isolate a DNA segment that is missing from one particular sample of DNA. Obviously, a second DNA sample that contains the fragment of interest is necessary. Suppose that a hereditary defect is due to the deletion of the DNA for a particular gene. A sample of DNA from the appropriate chromosome of the afflicted individual will lack this particular segment of DNA. To find the missing DNA, the corresponding chromosome from a healthy (wild-type) individual is isolated. For example, the ***dmd* gene**, for **Duchenne muscular dystrophy**, is located in the Xp21 band, close to the middle of the short or p-arm of the X-chromosome. Using light microscopy to analyze chromosomal banding patterns, a patient was found who had a deletion large enough that the Xp21 band was missing. Subtractive hybridization of the mutant chromosome with a normal chromosome allowed the *dmd* gene to be cloned.

Cloned genes can sometimes by found by a negative approach. Hybridization is used to remove genes shared by two organisms, leaving behind only those that are unique.

To do subtractive hybridization, both the mutant and wild-type DNA samples are cut into fragments of convenient size using a restriction enzyme. Then the two sets of fragments are hybridized together. This will give hybrid molecules for all regions of the DNA except the region of the deletion, which is present only in the wild-type chromosome.

Subtractive hybridization is a technique used to isolate a DNA segment that is missing from one particular sample of DNA. Obviously, a second DNA sample that contains the fragment of interest is necessary. Suppose that a hereditary defect is due to the deletion of the DNA for a particular gene. A sample of DNA from the appropriate chromosome of the afflicted individual will lack this particular segment of DNA. To find the missing DNA, the corresponding chromosome from a healthy (wild-type) individual is isolated. For example, the ***dmd* gene**, for **Duchenne muscular dystrophy**, is located in the Xp21 band, close to the middle of the short or p-arm of the X-chromosome. Using light microscopy to analyze chromosomal banding patterns, a patient was found who had a deletion large enough that the Xp21 band was missing. Subtractive hybridization of the mutant chromosome with a normal chromosome allowed the *dmd* gene to be cloned.

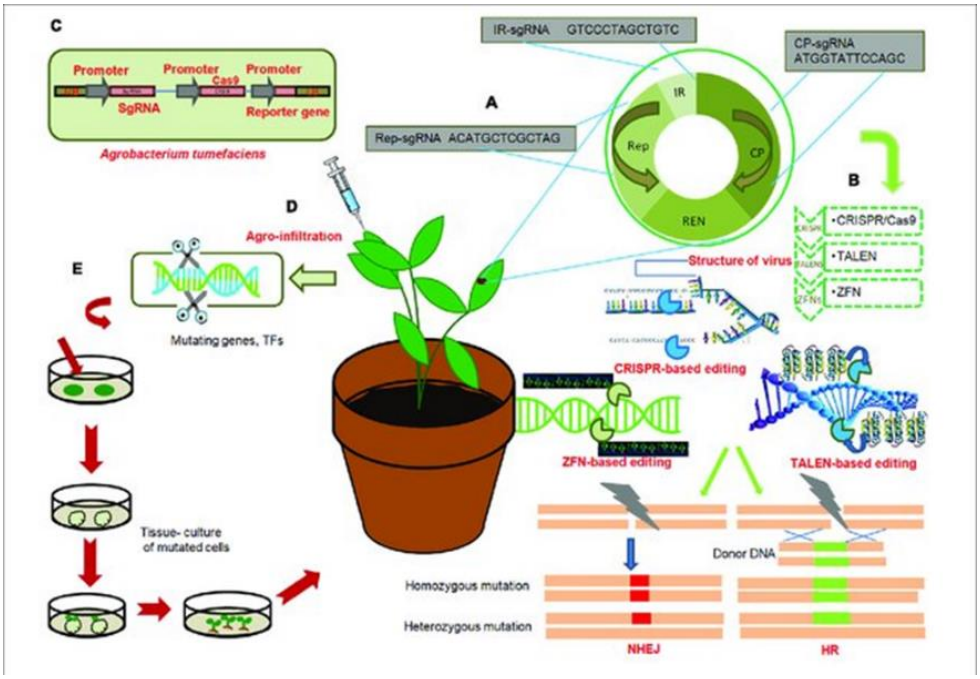
Cloned genes can sometimes by found by a negative approach. Hybridization is used to remove genes shared by two organisms, leaving behind only those that are unique.

To do subtractive hybridization, both the mutant and wild-type DNA samples are cut into fragments of convenient size using a restriction enzyme. Then the two sets of fragments are hybridized together. This will give hybrid molecules for all regions of the DNA except the region of the deletion, which is present only in the wild-type chromosome.

(6) Artificial Endonucleases-mediated Genome Editing

Targeted genome engineering (also known as genome editing) has emerged as an option to classical plant breeding and transgenic methods for crop improvement and it is currently the most attractive topic in plant molecular biology and genomics research. A key step in genome editing is the generation of a double-stranded DNA break that is specific to the target gene. This is achieved by engineered endonucleases, which enable site-directed mutagenesis via a NHEJ repair pathway and/or gene targeting via HR to occur efficiently at specific sites in the genome. Some technical advances described targeted mutagenesis and gene targeting by either NHEJ machinery using site-specific induction of double-strand breaks (DSBs), or by activation of a HR pathway through overexpression of a yeast DNA recombination gene in transgenic plants. Although gene targeting efficiency has been restricted to some plant species, its frequency was significantly enhanced by the recent use of programmable nucleases. Since that, many plant genes have been knocked out by this method. Over the past 15 years, tremendous efforts have been made and a variety of technologies have been developed to target mutations to a specific location in the genome using artificial endonucleases.

One of the earliest nuclease technologies involved mega-nucleases that were very difficult to engineer and requiring along and costly process. The DNA-binding do-mains of Zinc finger transcription factors (ZFNs) were thefirst to be used as genome editing tools and have proven easier to manipulate and have been used in tobacco, Arabidopsis and maize. More recently, transcription activator-like effector nucleases (TALENs), as well as the clustered regularly interspaced short palindromic repeats/Cas (CRISPR/Cas9) system using an RNA-guided bacterial immune system complex were re-ported. First studies confirmed that CRISPR provides protection against invading viruses when combined with Cas9 genes involving an RNA-mediated DNA targeting. Thereafter, CRISPR/Cas9 system is revolutionizing the field of genomic editing and have proved very useful and functional providing scientists with a powerful tool able to change any gene, in any cell in a highly targeted manner and without introducing foreign DNA. So far, several reports demonstrated the immense versatility of the technology in the field of plant biology by using a range of transformation platforms (PEG-protoplast transfection, agroinfiltration, virus transfection and generation of stable transgenic plants. This CRISPR/Cas9 allows for specific genome disruption and replacement in a flexible, robust and simple system resulting in high specificity and low cell toxicity, by targeting both endogenous genes and transgenes and by exploiting NHEJ and HR to generate small deletions, targeted insertions and multiplex genome modifications. This method has just been applied to a number of species including Arabidopsis, tobacco, tomato, sweet orange, wheat and rice. For all instances, genome editing was shown to work in both model and crop plants, as well as in a variety of other organisms using engineered nucleases as powerful tools to target specific DNA sequences to edit genes precisely in the plant genome.



General work-flow of gene editing technologies to engineer disease resistance in crops

(A) General genome organization of viruses; Target sgRNAs from each region of viral genome; replication associated protein (Rep), Intergenic region (IR), viral capsid protein (CP), with hypothetical sequences are shown in red. Multiplex genome editing strategy based on multiplex sgRNA targeting IR, CP and Rep of different viruses can be achieved by CRISPR/Cas9. (B) Illustration of three genome editing techniques conferring immunity of plants against virus: CRISPR/Cas9, TALENS, ZFNs. These technologies target different regions of viral genome and induce precise breaks at target sequences. Endogenous machinery of cells repair the breaks by non-homologous end joining (NHEJ) or homologous recombination (HR) thereby inducing genomic mutations at target locations. Induced mutagenesis in the viral or bacterial genome renders them ineffective. (C) T-DNA of Agrobacterium tumefaciens expressing sgRNA under CaMV-promoter, Cas9 protein under CaMV-promoter and reporter gene (GFP) under CaMV promoter. (D) Agroinfiltration of plant cells; injecting Agrobacterium containing engineered virus expressing sgRNA of target virus into Cas9-expressing plant. (E) Genome editing of genes or transcription factors, negatively regulating resistance against bacterial, viral or fungal pathogens, by deleting certain base pairs, in plants and subsequent raising of resistant plant by tissue culture techniques.

Gain of Function Mutagenesis

(Gene discovery by Activation Tagging using T-DNA as a tool)

Screen for loss-off-function mutations is a primary tool for dissecting a genetic pathway. However, because many genes belong to gene families, loss-of-function screens are not always possible to identify genes that act redundantly. In addition, some genes are critically required for the survival of plants. The homozygote mutants of these genes will not be available for entire function research because of embryonic or gametophytic lethality. As another option, gain-of-function technologies were developed to compensate the limitations of loss-of-function approaches or confer new function in transgenic plants, which is achieved through **activation expression of endogenous genes by transcription enhancer** which is randomly introduced into the genome or through ectopic gene over-expression driven by constitutive promoter.

The first directed method for undertaking **gain-of-function genetics** in plants exploited the enhancer element from the cauliflower mosaic virus (CaMV) 35S gene. A T-DNA vector was constructed containing four copies of this element. The resulting tetrameric CaMV 35S enhancer could then mediate transcriptional activation of nearby genes. This procedure became known as activation tagging. **Activation tagging** is a gain-of-function method that generates transgenic plants by T-DNA vectors with tetrameric cauliflower mosaic virus (CaMV) 35S enhancers which can lead to an enhancement expression of adjacent genes in the distance ranging between 0.4 to 3.6 kb from the insertion site (Weigel et al. 2000). Differently from the action of the complete CaMV 35S promoter, CaMV 35S enhancers can activate both the upstream and downstream gene transcription. In addition, it has been reported that in at least one case, rather than led to constitutive ectopic expression, CaMV 35S enhancers elevate transcriptional activity based on the native gene expression pattern.

Activation tagging technique was firstly developed by Walden and colleagues in decades years ago. Since then, several large scale activation tagging mutant resources have been generated and activation tagging method was widely used to isolate new genes. As an early example, the activation-tagging technique was used in tissue culture to identify cytokinin-independent mutants in *Arabidopsis* and *CKI1* gene whose overexpression can bypass the requirement for cytokinin in the regeneration of shoots was identified. Based on the original activation tagging vectors, Weigel and colleagues (2000) developed new generation of vectors possessing resistance to the antibiotic kanamycin or herbicide glufosinate which is low toxic to humans or easy to select transgenic plants in soil in large scale. By screening a set of the transgenic lines, they identified 11 dominant mutants with obviously morphological phenotypes and 9 of them were confirmed due to the activation of adjunct genes by reproducing the phenotype in a new set of transgenic lines through overexpression of the adjunct candidate genes on both sides of insertion.

To accelerate the recapitulation process of phenotype resulted from the enhancer of T-DNA insertion, a new activation-tagging method has been developed using a pair of plasmids including pEnLOX and pCre. pEnLOX contains multimerized CaMV 35S transcriptional enhancers flanked by two *lox P* sites on both sides while pCre includes the *cre* gene which can remove the DNA sequence between two *lox P* sites. The activation-tagging lines containing the pEnLOX were named the E-lines, and the helper lines containing pCre was named the C-lines. By crossing the E-lines with the C-lines, the CaMV 35S enhancers can be removed from the chromosome coming from E-lines and thus the reversion from mutant phenotypes to the wild-type phenotype may be detected in the next generation.

Two vectors, pEnLox and pCre constructed to obtain and effectively identify genes that are not expressed in wild-type plants under normal environmental conditions. Vector pEnLox is intended to induce insertion mutations and vector pCre to obtain transgenic lines of plants with expressed Cre recombinase genes. The enhancer in the structure of TDNA in the plasmid vector pEnLox is flanked by loxP-sites (Cre recombinase recognition sites). When the insertion mutant being study is obtained through transformation with pEnLox and is crossed with another transgenic line of A. thaliana carrying the cre gene, the enhancer sequence appears to be deleted. The system of vectors developed here makes it possible to induce insertion mutations and to determine their possible influence on the mutant phenotype. These insertions may lead either to insertion-induced inhibition of the gene function or to overexpression of genes immediately adjacent to the insertion due to the presence of an enhancer sequence in the structure of the insertion

Activation tagging has also been applied to generate rice activation-tagging lines. Based on the basic activation tagging technology, a dual function T-DNA vectors have been developed for both promoter trapping and CaMV 35S enhancers activation tagging. By analysis of the gene expression in these rice activation tagging lines, the authors reported the activated the expression of genes located up to 10.7 kb from insertion site of the enhancers. The activation tagging vector pSK1015 is used as an example in Figure 4. The T-DNA contains *BAR* gene as transgenic plant selection marker and 4×35 enhancers as activation element. The activation tagging T-DNA is integrated into the genome of host plants by *Agrobacterium* mediated transformation. The expressions of both side genes around the inserted T-DNA enhancer are elevated.

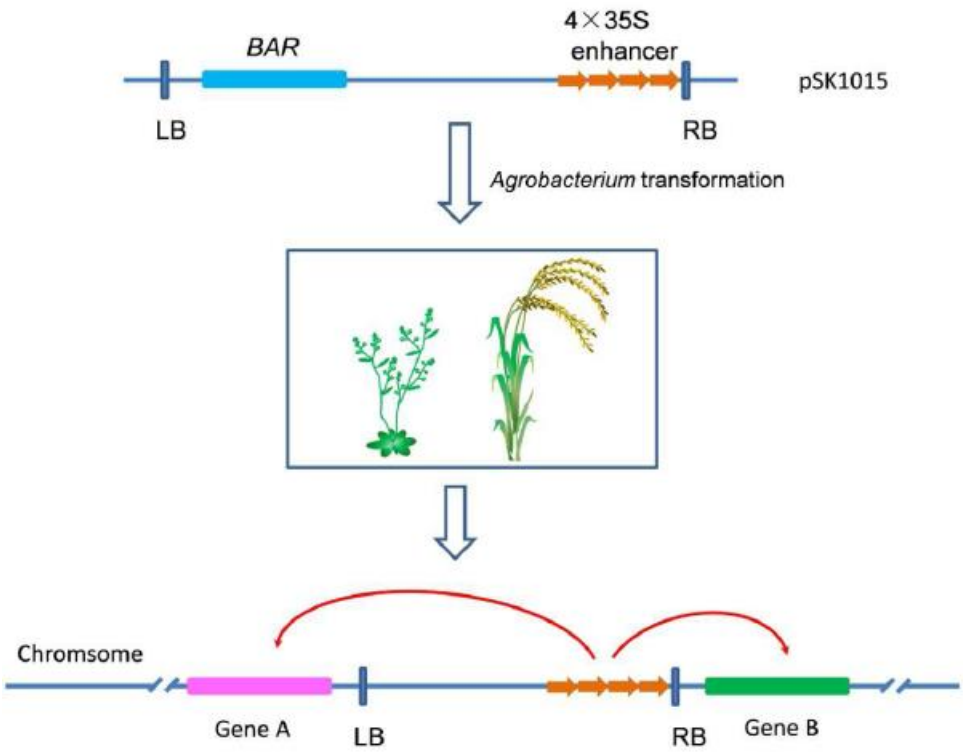


Fig.4. Scheme for the generation of activation tagging transgenic plants.

A similar strategy was developed for the **activation of genes** in Arabidopsis plants using a **nonautonomous maize Ds element** modified by the addition of the CaMV 35S promoter sequence, which could drive the expression of endogenous genes adjacent to the site of transposition. Contemporary transformation procedures for Arabidopsis now circumvent the requirement for tissue culture. Large populations composed of independent transgenic plants can therefore be generated relatively easily. Employing these methods, large sets of transgenic plants have been generated in a number of laboratories by transformation with modified activation tagging vectors. The replacement of the kanamycin resistance gene with one conveying resistance against the herbicide glufosinate facilitated selection of transformed plants on soil rather than in culture. Lines identified from these initial populations have lead to the discovery of a number of novel alleles and genes which undertake important functions in plant development, metabolism and environmental interactions.

FORWARD GENETICS: Understanding gene function

Mutation study is an important strategy to dissect gene functions. **Forward genetics** begins with a mutant phenotype and asks the question "What is the genotype?" that is, what is the sequence of the mutant gene causing the altered phenotype? **Reverse genetics** begins with a mutant gene sequence and asks the question "What is the resulting change in phenotype?" These two approaches are fundamentally different, and whereas forward genetics has been in operation for more than a century, the recent avalanche of complete genome sequences has only now created the opportunity for pursuing reverse genetics in an exhaustive and complete manner.

The most conventional approach to the analysis of gene function is **loss-of-function mutagenesis** (or Insertional mutagenesis) by chemicals such as ethyl methannesulfonate (EMS) or by high energy irradiation such as fast neutrons that introduce random mutations or deletions in the genome. The **classical mutagenesis** is one of the most powerful screen approaches to uncover genes involved in certain genetic pathways.

1. Natural Mutagenesis and Map-based Cloning

Natural mutants were generated during species evolution. Generally, the ratio of natural mutation is very low at only 10^{-5}-10^{-8} in higher plants. However, a large collection is still available during long evolutionary history. Some of such mutants were harmful or neutral and might be lost during evolution. Others might exhibit higher resistance to various abiotic / biotic stresses or have some specific agricultural traits, which were valuable germplasm resources for rice breeding. One example is the utilization of dwarf germplasm Dee-geo-woo-gen from China and release of rice variety IR8, which was developed from the dwarf line. Another example is the application of cytoplasmic male sterile (CMS) and photoperiod-sensitive genic male sterile rice lines, which are widely utilized to develop hybrid rice seeds for commercial release.

Map-based cloning is a widely-used method to isolate genes using such mutants. Genetic analysis is the first step to use these mutants for identifying gene functions, by which we know that how many genetic loci control the mutated phenotype. The next step is to finely map these loci to rice genome and then to clone the mutated gene confirmed by genetic complementation experiments. In fact, many genes have been isolated and functionally characterized by using natural mutant lines. For example, the rice *Xa21* and *Xa27* gene, which confers resistance to *Xanthomonas oryzae* pv. *oryzae* race 6, was isolated by map-based cloning from a natural mutated rice variety. Another such example is the isolation of rice semidwarf gene *sd-1*, which encodes a gibberellin 20-oxidase. Up to now, at least 67 rice genes have been isolated and functionally characterized by the map-based cloning.

2. Physical Mutagenesis and Deleteagene Detecting System

In 1930, Muller observed that mutation could be induced by X-rays. Subsequent researches found that the most efficient mutagenesis was mediated by fast neutron bombardment. A short deletion of DNA fragment was usually observed following the bombardment. Thus, a truncated gene might be detected by genomic subtraction and its functions could be identified by corresponding mutant phenotype. One example is the isolation and characterization of *Arabidopsis ga1-3* gene.

These methods are easy to generate large scaled mutagenesis, but the mutation site identifications by traditional forward genetic method need tedious works to hunt down the gene corresponding to a mutant allele by **map-base cloning** make an intrinsic limitation of this approach. Although popular to individual research groups who focus on certain aspect of their interested field, such kinds of methods are not convenient for systematic research of gene functions on a genome wide scale.

Currently, a new reverse genetics method has been developed to identify and isolate such mutants. This method was named as **Deleteagene**. In this system, DNA samples were extracted from the fast neutron-treated plants and were used to screen for deletion mutants by polymerase chain reaction (PCR) using specific primers flanking the targeted genes. Li et al. (2001) has generated an *Arabidopsis* population of 51,840 lines by fast neutron mutagenesis. This library was then used for screening deletion mutants of 25 gene loci, among which deletion mutants were obtained for 21 (84%) gene loci. Similarly, they also generated a rice fast neutron mutant pool with 24,660 lines and similar method was successfully used for identification and isolation of targeted genes. Evidence showed that this method can be efficiently used for the identification of small genes or tandemly arrayed genes. Wu

et al. (2005) reported the generation of around 10,000 rice mutant lines by the fast neutron bombardment and around 20,000 lines by γ-ray. Since the establishment of the method, many genes have been isolated and functionally characterized including the phytochrome family gene PHYC and phytochrome-interacting transcription factor *PIF3*, 3 genes encoding TGA transcription factors TGA2, TGA5, TGA6 and so on.

3. Chemical Mutagenesis and Tilling Detecting System

Chemical mutagenesis is mediated by certain chemical reagents. One of the most frequently used reagents is ethyl methane sulfonic acid (EMS). This alkylating agent can efficiently induce chemical modification of nucleotides, which results in various point mutations including nonsense, missense and silent mutations, among which silent mutations can not generate any modification in phenotype and thus can not be used for mutagenesis. In *Arabidopsis*, EMS mainly induces C to T changes resulting in C/G to T/A substitutions and at a low frequency, EMS generates G/C to C/G or G/C to T/A transversions by 7-ethylguanine hydrolysis or A/T to G/C transition by 3-ethyladenine pairing errors. Based on codon usage in *Arabidopsis*, the frequency of EMS-induced stop codon and missense mutations has been calculated to be ~5% and ~65%, respectively.

In *Arabidopsis*, at least 125,000 M1 lines should be generated in order to achieve saturation of EMS mutagenesis. However, it is not difficult to produce such a population since viable seeds can be used for EMS treatment. The difficulty is how to detect single-nucleotide polymorphisms or substitutions in these mutation lines in a large scale. Based on the technology in detecting single-nucleotide polymorphisms, McCallum et al (2000) established a new detecting method named as TILLING (Targeting Induced Local Lessions In Genomes) complemented with denaturing high-performance liquid chromatography (DHPLC). These technologies allow chemically induced mutant pools to be used for reverse genetics. With help of automation, robust and rapid detection makes it possible to screen a wide range of mutant pools in a short time and to avoid the laborious process of forward genetic screening. Now the technology has been used in various species including animals and plants and some improved methods were also provided.

In rice, around 18,000 and 9,000 mutants were generated from diepoxybutane and EMS mutagenesis, respectively. Total of 10 genes were screened using TILLING and independent mutations were detected in two genes: *pp2A4* encoding serine/threonine protein phosphatase catalytic subunit and *cal7* encoding callose synthase, suggesting the feasibility of this screen method in chemical mutagenesis. In another report, they screened 10 genes including *Os1433* (LOC_Os02g36974), *OsBZIP* (LOC_Os01g64000), *OsCALS8R* (LOC_Os01g55040), *OsDREB* (LOC_Os01g07120), *OsEXTE* (LOC_Os10g33970), *OsMAPK* (LOC_Os07g38530) *OsPITA* (LOC_Os12g18360), *OsR1A* (LOC_Os05g41290), *OsRPLD1* (LOC_Os01g07760) and *OsTPS1* (LOC_Os02g44230). Independent mutants were detected for all 10 genes. They also found that multiple nucleotide changes can be detected in each gene, suggesting that they have developed a useful method for more reliable and exact functional identification of a gene.

There are many ways to implement **targeted mutagenesis** so as to compromise specific genes. In mice, knockout mutations are now routinely obtained by promoting the homologous recombination of null gene constructs with the genomic wild-type sequence in embryonic stem cells. Provided that the given mutation is not embryonic lethal, "**knockout mice**" can then be developed in utero by injecting such stem cells into blastocysts. **Yeast and *Escherichia coli*** are other organisms in which homologous recombination are the preferred means for reverse genetics. Although there has been a report of homologous recombination with intact Arabidopsis plants (Kempin et al., 1997), the frequency of this event may be so low as to preclude its use for generating knockout mutations in each of the 25,000 genes that comprise the 120-Mb genome.

PROMOTER TRAP USING T-DNA AS A TOOL

With the ever increasing reports of creation of transgenic in various crop plants, the search for novel genes and variety of **regulatory elements of DNA required for controlled expression of the introduced genes** is gaining more importance. Regulatory elements that impart a tissue specific, stage-specific and/or environmental-stimuli specific expression to the transgene are being identified and cloned using T-DNA based vectors that have been designed to identify and clone such regulatory sequences.

Besides to disrupt the gene functions by direct insertion for large-scale discovery of gene function, T-DNA can also be utilized to identify novel regulatory elements. The general principle behind this approach is to **integrate a reporter gene that either lacks a promoter** (gene/promoter trap) or **carries only a minimal promoter (enhancer trap)**, at random sites in the wild type genome. A reporter gene cassette containing a minimal promoter (enhancer trap) close to the end of the insertion element can be *cis* activated when inserted close to a transcriptional enhancer that will drive the expression of the reporter gene (Figure 5). The gus (*uid* A) reporter gene is the most commonly used reporter gene system in plants, because of the absence of endogenous β-glucuronidase (GUS) activity in most plants and the opportunity to visualize the presence of the enzyme by sensitive histochemical techniques (Jefferson et al. 1987).

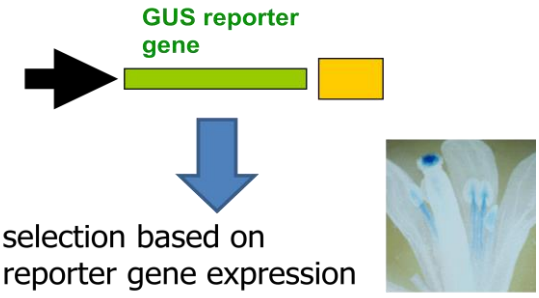


Fig. 5 Scheme for the promoter trap in plants

Mutagenised populations can be screened for lines expressing reporter gene in specific cell types or in specific environmental conditions. Genes with interesting expression patterns and their promoters can be isolated from such lines. The T-DNA tagged lines generated are being used for identifying novel genes and tissue specific promoters. Apart from the ease of identifying redundant genes this method helps detection of insertions in UTR's. An insertion in the 3' UTR's of genes generally will not lead to suppression of gene expression as the coding region remains intact, but can lead to reporter gene expression enabling detection of the otherwise hidden mutation. Repeated observations of a significant bias towards T-DNA integration in promoters and UTR's justify a high frequency of GUS expressing lines in the mutant population. Promoter trap lines can also be exploited to mark certain cell types for developmental studies. Enhancer traps allow an ingenious way of identifying genomic sequences expressed in precise developmental patterns. In the case of GUS as the reporter gene, the presence of GUS activity in a particular organ or cell at a particular developmental stage will identify sequences expressed at this place and time. Indeed, GUS enhancer traps have proved successful in detecting novel genes in *Arabidopsis*. A disadvantage, however, is that it may not be easy to identify and locate the enhancer causing the specific expression pattern, since the enhancer element could be in either side of the T-DNA and could also be at a far off distance.

THE ENHANCER TRAP SYSTEM

As a further development of T-DNA random insertion strategy, the enhancer trap system is established by random integration of a report gene cassette as the T-DNA into the genome. This **reporter cassette includes a minimum or truncated promoter** which is not able to drive the expression of the report gene unless activated by the endogenous regulatory sequence close to the integration site (Figure 6). Therefore, the **expression of reporter gene implies the existence of enhancer element close to the insertion site**.

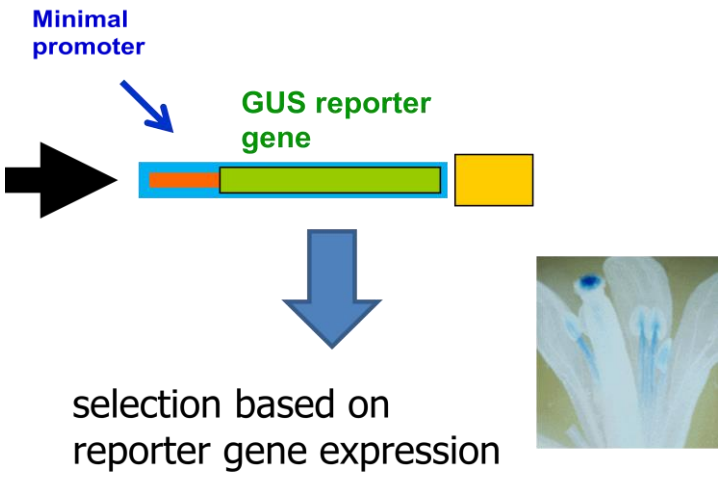


Fig. 6 Scheme for the enhancer trap in plants

Because the T-DNA cassette of enhancer trap system can both generate gene mutations and report the presence of enhance element around the insertion site, the enhancer trap system was widely applied in the bacterium, Drosophila, *Arabidopsis*, moss and rice to unveil gene functions and identify regulator elements. As a successful example, the Rice Mutant Database (RMD) is established with the enhancer trap system and maintained by National Center of Plant Gene Research (Wuhan) at Huazhong Agricultural University.

Reference:

1. Krysan et al (1999) T-DNA as an Insertional Mutagen in Arabidopsis. The Plant Cell, Vol. 11, 2283–2290.

2. Tani et al (2004) Activation tagging in plants: a tool for gene discovery. Funct Integr Genomics, 1–9

3. Meng et al. Transgenic Plants as Gene-Discovery Tools, Chapter-14 in Transgenic Plants – Advances and Limitations (www.intechopen.com)