CellPress

Review

Haploid Strategies for Functional Validation of Plant Genes

Yaou Shen,[1,2] Guangtang Pan,[1] and Thomas Lübberstedt[2,*]

Increasing knowledge of plant genome sequences requires the development of more reliable and efficient genetic approaches for genotype–phenotype validation. Functional identification of plant genes is generally achieved by a combination of creating genetic modifications and observing the according phenotype, which begins with forward-genetic methods represented by random physical and chemical mutagenesis and move towards reverse-genetic tools as targeted genome editing. A major bottleneck is time need to produce modified homozygous genotypes that can actually be used for phenotypic validation. Herein, we comprehensively address and compare available experimental approaches for functional validation of plant genes, and propose haploid strategies to reduce the time needed and cost consumed for establishing gene function.

Genetic Modification-Based Functional Validation for Plant Genes

Recent advances in plant genomics and sequencing technology revealed numerous associations between phenotypes and candidate genes. However, definitive functional annotations after *in vivo* validation have been thoroughly established for only few of these genes [1]. Genetic approaches for further validation of gene functions aim to create **genetic modifications** (see Glossary) that cause phenotypes of interest [1], including physical or chemical mutagenesis, insertional mutagenesis, Targeting Induced Local Lesions in Genomes (TILLING), gene overexpression, and gene silencing [2–5]. Most recently, genome editing comprising Zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR/Cas) enables investigators to manipulate any sequence in plant genomes *in situ* for validating gene and motif functions [6,7].

Classical and Current Approaches to Establish Plant Gene Function

The classical approach for plant gene function establishment began with loss-of-function mutagenesis after treatment with mutagens such as radiation with X-rays or neutrons, or chemicals that introduce random small deletions or point mutations in plant genomes [5,8]. Chemical mutagens such as ethyl methanesulfonate (EMS) were more popular as they are less destructive, easier available, and have a higher efficiency than physical mutagens [9]. Theoretically, we can find an EMS mutation in any given gene by screening no more than 5000 plants from the mutagenized M_1 generation for the model plant *Arabidopsis* [3,8]. Conventional mutagenesis has been widely used in **forward-genetic** strategies that start with a phenotype of interest and address identification of genes affecting this phenotype [9–11].

As an alternative forward-genetic tool, insertional mutagenesis, including T-DNA (Transferred DNA) and **transposon tagging**, facilitate the identification of genes disrupted by these elements [5]. Currently, T-DNA-tagged lines have been generated in large numbers, becoming a popular resource for plant gene function [12,13]. Superior to T-DNA, mobilizable transposons can

Trends

Recent advances in plant genomics and sequencing technology revealed numerous associations between phenotypes and plant genes. However, definitive functional annotations after *in vivo* validation have been thoroughly established for few of these genes.

Genetic methods for establishing gene functions are achieved by creating genetic modifications that cause phenotypes of interest. Diploid plants are usually heterozygous for the modified region in the first generation. More additional generations are required to obtain homozygous genotypes that can actually be used for validation of their phenotypic effect.

Efforts to obtain haploid plants have been undertaken in many plant species of almost all families in the plant kingdom. We proposed the strategies based on haploid mutagenesis and haploid transformation, respectively, which contribute to accelerating functional validation of plant genes.

[1]Maize Research Institute, Sichuan Agricultural University, 211 Huimin Road, Wenjiang District, Chengdu, 611130, China
[2]Department of Agronomy, Iowa State University, 100 Osborn Drive, Ames, IA 50011, USA

*Correspondence:
shenyaou@gmail.com (Y. Shen),
thomasl@iastate.edu (T. Lübberstedt)

provide verification about mutational effects of insertions when they are remobilized from the insertion site to recover a potential phenotype [14]. Insertional mutagenesis is less practicable for species, for which a systematic transformation platform has not been established [15].

As the number of characterized plant genes increases, **reverse-genetic** methodologies play an increasingly important role in gene function validation [16]. TILLING is the first reverse-genetic tool, in which chemical or physical mutagenesis is followed by a high-throughput screening for point mutations [3,4,16]. TILLING is practicable for plant species with large-sized genomes and without transformation system because it is not different from traditional chemical or physical mutagenesis in creating mutations [12].

Changes of gene expression levels may result in modified phenotypes, which can be another powerful approach for elucidating gene function [5]. Gene silencing or downregulation induced by RNAi can be achieved by expressing gene-specific double-stranded RNA (known as **siRNA**) or single-stranded RNA (known as **microRNA**) in plant cell, which in turn generates **loss-of-function mutations** [17,18]. RNAi is of great value for functional studies in polyploid plants because of its potential of silencing multigene families and homologous genes [19]. However, RNAi seldom leads to complete suppression of target gene expression, thus loss-of-function phenotypes cannot be observed by RNAi when the residual expression is still sufficient for gene function [20]. In contrast to RNAi, **overexpression** or **misexpression** of an inactive gene or a gene coding a limiting protein will lead to **gain-of-function mutations** in transformants [21]. In this way, even the phenotypes of individual members in a gene family are observable without interference from functionally redundant genes [5].

Targeted mutagenesis can be achieved by ZFNs, TALENs, or CRISPR/Cas in which custom DNA-binding motifs direct nonspecific nucleases to cleave a double strand in the genome at a specific site that further stimulates error-prone nonhomologous end joining or homology-directed repair at specific genomic locations [6,22–24]. More applicable and easier to manipulate than ZFNs and TALENs, the CRISPR/Cas system only requires a single short RNA to generate target specificity [6], which even allows the genome-wide functional identification [25]. For these reasons, CRISPR/Cas is becoming a popular technique for gene targeting [26–29]. Relevant bioinformatic tools for selecting optimal CRISPR/Cas target sites have been developed and are available online [24].

A comparison of different approaches for functional validation of plant genes is displayed in Table S1 in the supplementary material online. As recent developed techniques, ZFNs, TALENs, and CRISPR/Cas can generate **custom mutagenesis** effectively, resulting in both targeted **gene knockouts** and **gene knock-ins** [6,25], which cannot be achieved by any other traditional method. Current research on genome editing addresses increase in precision and efficiency of gene targeting [24,30]. Targeted mutagenesis still depends on plant transformation, by which T-DNA carrying chimeric enzymes and binding motifs are integrated into the plant genome and expressed for targeted double-strand breaks [27,28,31]. The methods independent of genetic transformation for genome editing, such as direct delivery of these reagents or transient expression of these enzymes in plant cells, will substantially simplify the process of gene editing, even in species with large genomes [32,33].

Time and Resources Required for Current Approaches

The approaches for validating gene function can be classified into mutagenesis- and transformation-based genetic modification (Figures 1 and 2). Independent of the approach, diploid plants are usually heterozygous for the modified region in the first generation (T_0 or M_1) [25,34]. Altered genes and sequence motifs are often recessive and, therefore, without phenotypic effect in mutagenized T_0 or M_1 plants [34]. One or more additional generations are thus required to

Glossary

Clonal propagation: an asexual process to reproduce plant cells by tissue culture.
Custom mutagenesis: the process to create any site-directed mutation by ZFNs, TALENs, or CRISPR/Cas.
Embryogenesis: the process by which the embryo forms and develops *in situ*.
Forward genetics: an approach of determining the genetic basis responsible for a phenotype, which was initially done by generating mutants by physical, chemical, or insertional mutagenesis and subsequently followed by isolation of mutant individuals and identification of functional gene.
Gain-of-function mutation: a mutation that leads to new or enhanced protein function.
Gene knock-in: a genetic engineering method that involves the insertion of a protein coding cDNA sequence at a particular locus in an organism's chromosome.
Gene knockout: a genetic technique where genes in an organism are made inoperative in functions.
Genetic modification: a process by which the genetic information of an organism is changed in a stable manner, resulting in a mutation.
Haploid inducer: a specific plant genotype used for the production of haploid plants by cross with a donor plant.
Haploid transformation: a process where haploid cells or tissues are genetically transformed.
Inducing medium: a type of medium containing hormones, which is used for dedifferentiation of plant cell or tissue.
Loss-of-function mutation: a mutation that results in reduced or abolished protein function.
microRNA: a class of single-stranded RNA molecules containing approximately 22 nt found in plants, animals, and some viruses, which function in RNA silencing by post-transcriptional regulation of gene expression.
Misexpression: expression of a gene in a cell type or developmental stage or condition where it normally is not expressed.
Overexpression: excessive expression of an endogenous gene in an organism by genetic transformation to enhance its