Tandem gene arrays: a challenge for functional genomics

Georg Jander¹ and Carina Barth²

¹Boyce Thompson Institute for Plant Research, Tower Road, Cornell University, Ithaca, NY 14853, USA
²Department of Biology, 53 Campus Drive, West Virginia University, Morgantown, WV 26506, USA

In sequenced plant genomes, 15% or more of the identified genes are members of tandem-arrayed gene families. Because mutating only one gene in a duplicated pair often produces no measurable phenotype, this poses a particular challenge for functional analysis. To generate phenotypic knockouts, it is necessary to create deletions that affect multiple genes, select for rare meiotic recombination between tightly linked loci, or perform sequential mutant screens in the same plant line. Successfully implemented strategies include PCR-based screening for fast neutron-induced deletions, selection for recombination between herbicide resistance markers, and localized transposon mutagenesis. Here, we review the relative merits of current genetic approaches and discuss the prospect of site-directed mutagenesis for generating elusive knockouts of tandem-arrayed gene families.

Tandem gene multiplications pose a challenge for functional analysis

Completed (Arabidopsis thaliana, rice and poplar), in-progress (tomato, Medicago truncatula and maize), and pending plant genome projects will provide an unprecedented amount of information about the protein repertoire of higher plants. However, with estimates of the number of protein-coding regions ranging from 27 000 in Arabidopsis to 59 000 in maize [1–3], functional genetic analysis will be one of the greatest challenges facing plant biologists in coming years. Even in Arabidopsis, currently the best-studied model plant, there is direct experimental evidence for the function of only about a third of the predicted genes [2].

One of the more conclusive approaches for assigning function to unknown plant genes is to study phenotypic changes associated with knockout mutations. In several model plant species, efficient reverse genetics approaches, including TILLING [4], transposon mutagenesis [5–14], and large collections of sequenced T-DNA insertions [15–19], enable the identification of knockout mutations in almost any gene of interest. However, functional analysis is complicated by the lack of measurable phenotypes associated with most knockout mutations [20]. A major contributing factor to this observation is the highly redundant nature of plant genomes. It is likely that all known plant species have polyploid origins [21,22]. In Arabidopsis, the most recent genome duplication event occurred ~40 million years ago (mya) whereas the diploid maize probably evolved from an allotetraploid origin within the past 5 my [23]. By contrast, the hexaploid wheat genome is the result of polyploidization events that occurred during the domestication process [24]. Another significant contribution to genome redundancy comes from tandem gene multiplications, where two or more homologous genes are directly adjacent to one another in the genome. Although enumeration of tandem arrays is influenced by the particular bioinformatic analyses that are employed, it is clear that they represent a significant proportion of all plant genes: 17% in Arabidopsis [1], 14% in rice [25], 16% in poplar [26], and 35% in maize [3]. Most tandem multiplications in Arabidopsis (75%) and rice (79%) contain only two genes, and gene arrays with more than three members are uncommon [27].

Because of genetic redundancy, it is often necessary to knock out more than one member of a plant gene family to obtain a mutant phenotype [28–31]. Creating such double mutants can be relatively simple if the genes of interest are genetically unlinked and non-essential: mutations are identified in two separate plant lines, the mutant lines are crossed (e.g. a/a B/B × A/A b/b), and 1/16 of the resulting F2 progeny are homozygous for both mutations. However, in the absence of reliable site-directed mutagenesis protocols for higher plants, finding double mutants of tightly linked genes is difficult. For instance, if two mutations are 5000 bp apart in the Arabidopsis genome, then genetic recombination would bring both mutations together in one out of every 8000 gametes (assuming 200 kbp/cM in Arabidopsis). Homozygous double mutants would be found as roughly one out of every 64 million F2 progeny or one out of every 32 000 F3 progeny from such a cross. Thus, except in rare cases where there is a simple screen or selection for the desired phenotype, finding such double mutants can be challenging using standard genetic methods. Here we describe alternate strategies (Table 1) that can be used to identify knockouts of two or more tandem-arrayed plant genes for functional analysis.

Selection for deletions spanning multiple genes

Large chromosomal deletions, unlike point mutations and DNA insertions, readily knock out the function of multiple adjacent genes. Therefore, strategies for the efficient generation and detection of chromosomal deletions will facilitate the functional analysis of tandem-arrayed genes. Treatment with a 60 Gy fast neutron dose creates about ten deletions per genome, knocking out individual Arabidopsis
genes with a frequency of one in 2500 [32]. For detecting such fast neutron-generated deletions, PCR primer extension times can be adjusted such that the deletion allele of a given locus is selectively amplified (Figure 1a). Xin Li et al. [33] pioneered this amplification strategy by showing that a gene of interest can be detected by PCR strategies such as those described above for fast neutron-induced deletions. A community resource of Arabidopsis Ds-loxP T-DNA insertion sites is available for Arabidopsis and rice [5,35]. Finally, after a second cross to bring in the Cre recombinase gene, Cre-mediated site-specific recombination between the twoloxP sites causes the excision of the intervening DNA. Excision of a gene of interest can be detected by PCR strategies such as those described above for fast neutron-induced deletions. A community resource of Arabidopsis Ds-loxP T-DNA insertion sites is available for Arabidopsis and rice [5,35]. Finally, after a second cross to bring in the Cre recombinase gene, Cre-mediated site-specific recombination between the twoloxP sites causes the excision of the intervening DNA.

An alternate approach for generating chromosomal deletions makes use of localized transposition by the maize Dissociation (Ds) element and site-specific DNA excision by the bacteriophage Cre-loxP recombination system [35]. First, a Ds-loxP T-DNA construct, containingloxP sites in both the Ds element and the T-DNA backbone is transformed into the genome (Figure 1b). Next, transposition ofDs-loxP is induced by crossing in Activator (Ac). Typically, 10% of all reinsertions are within 10 kbp of theDs-loxP T-DNA insertion site [35]. Finally, after a second cross to bring in the Cre recombinase gene, Cre-mediated site-specific recombination between the twoloxP sites causes the excision of the intervening DNA. Excision of a gene of interest can be detected by PCR strategies such as those described above for fast neutron-induced deletions. A community resource of Arabidopsis Ds-loxP T-DNA insertion sites is available for Arabidopsis and rice [5,35]. Finally, after a second cross to bring in the Cre recombinase gene, Cre-mediated site-specific recombination between the twoloxP sites causes the excision of the intervening DNA.

Table 1. Advantages and disadvantages of approaches for generating mutations in tandem-arrayed genes

<table>
<thead>
<tr>
<th>Approach</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection for genomic deletions</td>
<td>• Can delete multiple adjacent genes</td>
<td>• Requires DNA sequence data for region adjacent to the genes of interest</td>
<td>[33,34]</td>
</tr>
<tr>
<td></td>
<td>• Mutagenesis is performed on dry seeds</td>
<td>• Can delete nearby genes, resulting in phenotypes not related to gene of interest</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No plant transformation is required</td>
<td>• Requires generation of a large mutant population</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Effective, particularly for targeting small genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fairly fast owing to multiplexed PCR screens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localized Ds transposition and excision by Cre-loxP</td>
<td>• Arabidopsis Ds-loxP T-DNA insertion resource is being developed</td>
<td>• Requires two crosses to generate a deletion</td>
<td>[5,35]</td>
</tr>
<tr>
<td></td>
<td>• Can delete multiple adjacent genes</td>
<td>• Resource is not available for most plant species</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Does not require full genome sequence</td>
<td>• Requires plant transformation to generate Ds-loxP T-DNA insertion population</td>
<td></td>
</tr>
<tr>
<td>RNA interference</td>
<td>• Allows downregulation of multiple genes using one construct</td>
<td>• Possible off-target effects</td>
<td>[38–44]</td>
</tr>
<tr>
<td></td>
<td>• Variable silencing efficiency is useful if null mutations are lethal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Does not require a full genome sequence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>miRNA-based gene silencing</td>
<td>• Allows silencing of multiple genes</td>
<td>• Degree of silencing can vary when targeting multiple genes</td>
<td>[45,46]</td>
</tr>
<tr>
<td></td>
<td>• miRNA design software is available</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Relatively fast and easy to implement</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Does not require full genome sequence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection for meiotic recombination</td>
<td>• Selection is based on herbicide resistance</td>
<td>• Depends on rare meiotic recombination</td>
<td>[51,53]</td>
</tr>
<tr>
<td></td>
<td>• Homozygous lethal double mutants can be identified in the heterozygous state</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Insertion mutant populations are available for Arabidopsis and rice</td>
<td>• Requires many crosses and PCR screens</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Can select complete gene knockouts</td>
<td>• Tiedious in the absence of insertion mutants</td>
<td></td>
</tr>
<tr>
<td>Sequential mutagenesis</td>
<td>• Heterozygous transposon insertions can be detected by PCR</td>
<td>• Gene knockout populations are not available for most plant species</td>
<td>[61]</td>
</tr>
<tr>
<td></td>
<td>• Does not require a full genome sequence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Transposon launch sites are available for several model plant species</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Can select complete gene knockouts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site-directed mutagenesis by zinc finger nucleases (ZFNs)</td>
<td>• Highly specific when using appropriately designed zinc finger domains</td>
<td>• Not yet a proven technology</td>
<td>[70–74]</td>
</tr>
<tr>
<td></td>
<td>• Might allow deletion of multiple genes</td>
<td>• Limited by the ability to engineer zinc finger domains that recognize specific targets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Software for identification of potential ZFN target sites is available</td>
<td>• Possible off-target effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Automated zinc finger design methods are being developed</td>
<td>• Potential cytotoxicity of ZFNs</td>
<td></td>
</tr>
</tbody>
</table>
least the first cross. Ac-Ds transposons are active in many plant species and, in theory, could be broadly applicable for generating deletions. A similar recombinase system, employing R/RS from the yeast Zygosaccharomyces rouxii in combination with Ac, has been shown to induce deletion formation in rice [36] and petunia [37].

Silencing gene families by RNA interference

RNA interference, which allows down-regulation of gene expression in transgenic plants, is an attractive method for functional analysis of gene families, including those where the genes of interest are located in tandem arrays. Gene silencing by RNA interference results from sequence-specific RNA degradation that follows the formation of double-stranded RNA that is homologous in sequence to the targeted gene[38–40]. By targeting shared sequences, it is possible to reduce expression of multiple members of a gene family simultaneously. For instance, seven members of the OsRac gene family in rice were silenced with varying efficiency by a single RNA interference construct [41]. In hexaploid wheat (Triticum aestivum), silencing of homologous genes with double-stranded RNA constructs induced phenotypes that would have been difficult or impossible to attain by standard genetic approaches [42].

Rather than targeting multigene families with a single shared RNA segment, it is possible to achieve effective suppression using chimeric silencing constructs made with pieces of several genes. For instance, a chimeric hairpin RNA construct was used to silence all members of the opium poppy (Papaver somniferum) codeinone reductase gene family [43]. Similarly, a chimeric construct employing partial sequences of several genes was used to suppress lignin biosynthetic enzymes in tobacco (Nicotiana tabacum) [44].

The use of synthetic microRNA constructs represents an alternate approach to using relatively large RNA fragments for gene silencing. Two recent studies demonstrated that down-regulation of gene families by endogenous and synthetic microRNAs mimics the phenotypes produced by previously described multiple mutations in the targeted genes [45,46]. Software for automated design of artificial microRNAs, based on the backbone of a natural microRNA and the input of individual or several related target sequences, is available (http://wmd.weigelworld.org).

The RNA-based gene silencing approaches described above can easily be applied to reduce the expression of duplicated genes. However, interpretation of the results can be complicated by the variable effectiveness of gene silencing by RNA interference. Silencing efficiency depends on the type of construct that is used, the site of integration, and the particular target genes [47–50]. Furthermore, because gene silencing is generally not 100% effective, mutant phenotypes that require the complete knockout of a gene family might not be detected.
using the described silencing approaches. For instance, insertional knockout of a single defense-related *Arabidopsis* myrosinase reduced total enzymatic activity by more than 95% (similar to what one might achieve by RNA interference), but had no measurable effect on insect resistance. By contrast, a tgg1 tgg2 double myrosinase knockout reduced enzymatic activity to undetectable levels and made the plants much more sensitive to lepidopteran herbivory [51].

**Selection for meiotic recombination between adjacent genes**

Most tandem arrays in plants contain only two genes [27]. Therefore, if viable mutants are available for both loci, an obvious approach is to cross them and identify homozygous recombinants in subsequent generations. However, as discussed above, in the absence of a strong screen or selection for the expected phenotype, finding meiotic recombination between tightly linked genes is difficult. Screening for recombinants among thousands of progeny from a cross is feasible only for mutations that cause easily scored visible phenotypes, for instance kernel color changes induced by bronze mutations [48] and intragenic recombinants at the r locus [52] in maize. Although it is theoretically possible to screen several thousand plants by PCR-based methods to identify double mutants, we are not aware of any published examples of this approach.

An alternate approach for identifying rare meiotic recombination between closely linked plant genes relies on mutant alleles that confer herbicide resistance. The first implementation of this method involved selection for intragenic recombination between the *Arabidopsis* csr1-1 and csr1-2 mutations, which confer resistance to chlorsulfuron and imidazolinone herbicides, respectively [53]. First, the two mutants were crossed to create a double-resistant heterozygous line with the mutations in *trans*. Then, the heterozygote was used as the male parent in a second cross, and meiotic recombination bringing the mutations in *cis* was identified by transfer of both herbicide resistance genes in one pollen grain. This genetic strategy can be generalized to select for meiotic recombination between any two *Arabidopsis* insertional mutations that confer resistance to different herbicides (Figure 2). Meiotic recombination between T-DNA insertions conferring Basta and kanamycin resistance, respectively, in the tandem-duplicated TGG1 and TGG2 *Arabidopsis* myrosinase genes was found with a frequency of one in 3000 [51]. Use of a male sterile mutant as the pollen recipient facilitated the otherwise laborious number of crosses that were needed to find the desired recombinants.

An added advantage of the approach outlined in Figure 2 is that double mutants are selected in the heterozygous state. This reduces the number of plants that need to be screened to identify recombinants and also allows

---

**Figure 2.** Selection for recombination between T-DNA or transposon insertions [51]. Insertions conferring herbicide resistance (in this example kanamycin<sup>R</sup> and Basta<sup>R</sup>) are identified in two genes of interest. The first cross brings insertions in tandem-duplicated genes into a *trans* configuration. Progeny of the first cross are used as the pollen donors for a second cross, preferably to a male sterile recipient to reduce the amount of effort involved. Transfer of both herbicide resistance markers in the same pollen grain selects for meiotic recombination that brings the insertions into the *cis* configuration. Self-pollination of these plants allows screening for homozygous double mutants.

www.sciencedirect.com
identification of homozygous lethal double mutations. In this case, self-pollination of the doubly heterozygous individuals would lead to seed or seedling lethality in ∼25% of the progeny. Given the large collections of T-DNA and transposon insertion mutants that are available to Arabidopsis researchers [15–17,54,55], it is likely that the approach outlined in Figure 2 can be applied to many other tandem-duplicated gene pairs. Large collections of insertional mutants are also being developed for rice [18,19,56–60], making a similar approach feasible for this important crop plant.

**Sequential mutagenesis to create a double mutant**

Rather than identifying two mutations independently and bringing them together by genetic recombination, it is also possible to select for a second mutation in the genetic background of the first one. If the second mutation causes a known, easily scored phenotype, this could be fairly straightforward and would essentially represent a genetic enhancer screen. Otherwise, reverse genetics approaches such as TILLING [4] could be used to identify a second mutation without regard to the possible phenotype. However, the hurdles involved in setting up a large mutant population and using reverse genetics to find one particular mutated gene can be prohibitive. We are aware of no published reports where such approaches have been applied to generate double knockouts of tandem-duplicated plant genes.

Available collections of sequenced transposon launch sites in Arabidopsis, maize, rice and tomato [5–14] allow localized saturation mutagenesis of plant genomes. As was demonstrated recently with the Arabidopsis CYP79F1 and CYP79F2 genes [61], the propensity of Ac–Ds to both reinsert nearby in the genome and leave a ‘footprint’ in the DNA sequence at the original insertion site can be used to create double mutants of adjacent genes (Figure 3). First, a Ds transposon insertion in CYP79F2 was reactivated by induction of transposase expressed from a heat shock promoter. Next transposition into the adjacent CYP79F1 gene was identified by visually screening for a known cyp79F1 rosette phenotype in the second generation after transposase induction. Finally the transposon footprint in cyp79F2 and the reinsertion in cyp79F1 were verified by PCR amplification and DNA sequencing. In theory, this localized transposon mutagenesis could be used to identify double mutants of many other genes. However, not all such selections will benefit from easily identified phenotypes such as those associated with the cyp79F1. In most cases, it will be necessary to screen for the desired second insertion using PCR-based approaches. However, PCR-based screening has the advantage that heterozygous insertional mutations can be detected in the first generation after transposon mutagenesis. By contrast, visually identifying recessive cyp79F1 homozygotes required screening hundreds of mutant families in the second generation after mutagenesis [61]. Pooling strategies, where multiple DNA samples are combined for PCR and individual lines are analyzed only if a pool containing those lines produces a positive result [6], can greatly reduce the number of PCR reactions required to find transposon insertions in a particular gene.

**Prospects for site-directed mutagenesis in higher plants**

All the methods described above would be trivialized by the development of an effective site-directed mutagenesis protocol for plants. Although targeted mutagenesis is commonly used to study model organisms such as yeast and mice, there is as yet no working system for creating such mutations in higher plants. To date, high frequency homologous recombination has been achieved only in the moss Physcomitrella patens [62,63]. In flowering plants, oligonucleotide-directed gene targeting resulting in selectable phenotypes has been demonstrated in tobacco [64] and maize [65,66]. However, gene conversion occurred with a frequency of only 10⁻⁴ in these experiments [64,66]. Gene targeting by homologous recombination in Arabidopsis was enhanced more than tenfold by expression of the yeast (Saccharomyces cerevisiae) RAD54 gene [67], a member of the SWI2/SNF2 superfamily of chromatin remodeling genes [68].

Zinc finger nucleases (ZFNs) represent an exciting new approach for site-directed mutagenesis in plants. These chimeric protein constructs, consisting of sequence-specific...
Zinc finger DNA binding domains and a FokI DNA nuclease, stimulate site-specific recombination and repair processes by generating double-strand breaks at specific target sites in genomic DNA [69]. Because double-strand breaks stimulate DNA repair and recombination in essentially all organisms, targeted cleavage with ZFNs should be broadly applicable for site-directed mutagenesis. Two proof-of-concept publications [70,71] show that ZFNs can induce chromosomal mutations in Arabidopsis (Figure 4). Alan Lloyd et al. [70] combined a ZFN gene driven by a heat-shock promoter with a transgenic target in the Arabidopsis genome. ZFN cleavage at the specific genomic site stimulated imperfect repair mutations with a frequency of ~10%. In an alternate approach, David Wright et al. [71] demonstrated that chromosome breaks created by ZFNs greatly enhance the frequency of homologous recombination. An artificial target gene containing a translational fusion between β-glucuronidase (GUS) and neomycin phosphotransferase (NPTII) was inactivated by a deletion in the GUS active site and the insertion of a ZFN recognition site. ZFN cleavage and restoration of a functional chromosomal GUS:NPTII fusion by homologous recombination with transfected DNA occurred with a frequency of ~20%.

Although targeted mutagenesis using ZFNs has great potential, this approach is currently limited by the difficulty of engineering new zinc finger domains, each of which recognizes a specific nucleotide triplet. Modular assembly of individual zinc finger domains has allowed the design of proteins that recognize 18 bp DNA sequences with high specificity and affinity [72]. Given an 18-bp target sequence, it should be possible to identify unique zinc finger DNA binding sites in most plant genes. Not surprisingly, several groups are developing semi-automated methods for designing zinc finger DNA binding proteins that target specific DNA sequences [72–74]. In one such approach [74], the procedure involves the identification of potential ZFN target sites using ZiFiT software (http://bindr.gdcb.iastate.edu/ZiFiT/), assembly of multi-finger arrays, testing the DNA-binding activity of a three-finger domain using a bacterial two-hybrid reporter assay, and construction of ZFN expression vectors. All the required steps take about four weeks. However, the use of ZFNs for homologous recombination in planta still requires optimization to avoid off-target DNA breaks and potential cytotoxicity.

With further improvements in the design of ZFNs that target specific genomic sequences, generating knockouts of tandem-arrayed plant genes might become routine. For instance cleavage with a ZFN could be used to induce a second mutation in a tandem-duplicated gene pair in a line where one of the genes already has a T-DNA insertion. Alternatively, if a single ZFN creates double-strand breaks in two members of a gene array, it might be possible to delete the intervening sequence and thereby inactivate multiple genes in one step. In either case, one will be able to use knockout mutations of tandem gene arrays as a tool for their functional analysis.

Although site-directed mutagenesis protocols for higher plants are still being optimized, the other genetic strategies described in this review (summarized in Table 1) are available for making tightly linked double mutations. Together, the described methods represent an effective genetic toolkit for plant researchers. The particular species being investigated and the available mutant alleles will dictate which approach is the best choice for the functional analysis of tandem-arrayed plant genes.

Acknowledgements
We thank Dan Voytas and Patrick Krysan for providing access to their unpublished manuscripts, and Hong Gu Kang, Tom Brutnell and Martin De Vos for helpful discussions. We apologize to the authors of relevant papers that we were not able to cite because of space constraints.

References


9 Enns, L.C. et al. (2005) Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant Mol. Biol.* 58, 333–349


Plant Science Conferences in 2007

Gordon Research Conference – Epigenetics
5–10 August 2007
Holderness School, Plymouth, New Hampshire, USA

10th International Colloquium on Endocytobiology and Symbiosis
10–13 September 2007
Gmunden, Upper Austria
http://www.endocytobiology.org/

Fourth International Symposium on Dynamics of Physiological Processes in Roots of Woody Plants
16–19 September 2007
Bangor, UK

16th Biennial Australasian Plant Pathology Society Conference
24–27 September 2007
Adelaide, Australia