Reverse genetics techniques: engineering loss and gain of gene function in plants

Erin Gilchrist and George Haughn

Advance Access publication date 16 January 2010

Abstract
Genetic analysis represents a powerful tool that establishes a direct link between the biochemical function of a gene product and its role in vivo. Genome sequencing projects have identified large numbers of plant genes for which no role has yet been defined. To address this problem a number of techniques have been developed, over the last 15 years, to enable researchers to identify plants with mutations in genes of known sequence. These reverse genetic approaches include RNAi and related technologies and screening of populations mutagenised by insertion (PCR), deletion (PCR) and point mutation (TILLING), each with its own strengths and weaknesses. The development of next-generation sequencing techniques now allows such screening to be done by sequencing. In the future, it is likely that the genomes of thousands of plants from mutagenised populations will be sequenced allowing for the identification of plants with mutations in specific genes to be done in silico.

Keywords: reverse genetics; RNAi; TILLING; mutagenesis; plants

INTRODUCTION
The number of sequenced genes whose function remains unknown has soared in recent years with the advent of affordable genome sequencing. Even in those species whose entire genome has not been sequenced, there are genes whose sequence is known but whose function is not yet well understood. Bioinformatics has been invaluable in investigating the function of these genetic components, and powerful in silico techniques have been developed for the analysis of genome sequence information, but the elucidation of gene function must always be verified 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 (Table 1).

VIRUS-INDUCED GENE SILENCING
Virus-induced gene silencing (VIGS) has emerged as one of the most adaptable techniques in recent years but its effects are transient in nature [1–3]. VIGS is performed by cloning a 200–1300 bp cDNA fragment from a plant gene of interest into a DNA copy of the genome of an RNA-virus and transfecting the plant with this construct using Agrobacterium. Double-stranded RNA from the viral genome, including the sequence from the gene of interest, is formed during viral replication. The double-stranded RNA molecules are degraded into small interfering RNA (siRNA) molecules by the plant Dicer-like enzymes, thus activating the siRNA silencing pathway. This results in the degradation of the target gene transcript leading to a knockout or knockdown phenotype for the gene of interest. The most common vectors currently used are based on the tobacco rattle virus (TRV)
but, recently, the apple latent spherical virus (ALSV) has been reported to be even broader in its host range and to have minimal side effects [4]. VIGS has been used in a wide variety of plants including both dicot and monocot species. The advantages of this technique compared to other reverse genetics techniques include the facts that it is quite rapid and inexpensive and, because the phenotype is transient, deleterious loss-of-function effects may sometimes be observed without concomitant lethality (Table 1). In addition, the function of several homologous genes may be affected with a single construct. Further, it is a technique limited only by the host range of the virus used, making it applicable to many species. However, the phenotypes induced using this technique are not heritable and so cannot be used for genetic engineering. In addition, the silencing level is variable depending on the construct and the growth conditions used and it is rare that gene(s) are completely silenced. A final disadvantage of this technique is that it has not yet been developed for all plant species.

RNA-MEDIATED INTERFERENCE

RNA-mediated interference (RNAi) is similar in action to VIGS, but is heritable in nature and so offers a different scope of investigation [5–7]. In this technique, a DNA construct that produces either single- or double-stranded RNA complementary to the gene of interest is introduced into a cell where it activates the RNA silencing pathway and degrades some or all of the transcripts from the gene of interest as described above. There are several techniques commonly used to activate the RNAi pathway in plants but the one growing fastest in popularity is the use of artificial microRNAs (amiRNA). In this procedure, a microRNA (miRNA) gene carrying a 21 bp insert

Table 1: Reverse genetics techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus-induced gene silencing</td>
<td>- No transformation necessary&lt;br&gt;- Rapid results&lt;br&gt;- Can obtain partial loss-of-function&lt;br&gt;- Low cost</td>
<td>- Transient&lt;br&gt;- Effects on non-target genes&lt;br&gt;- Silencing level is variable&lt;br&gt;- Not developed for all species&lt;br&gt;- No way of obtaining missense mutations</td>
<td>[1–4]</td>
</tr>
<tr>
<td>(VIGS)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA-mediated interference</td>
<td>- Heritable&lt;br&gt;- Rapid results&lt;br&gt;- Can get knockdown of multiple homologous genes&lt;br&gt;- Can obtain partial loss-of-function&lt;br&gt;- Can be temporal or tissue specific&lt;br&gt;- Low cost</td>
<td>- Effects on non-target genes&lt;br&gt;- Silencing is variable depending on construct and transformant&lt;br&gt;- Not developed for all species&lt;br&gt;- No way of obtaining missense mutations&lt;br&gt;- Requires transformation&lt;br&gt;- Some genes are resistant to silencing</td>
<td>[5–10]</td>
</tr>
<tr>
<td>(RNAi)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insertional mutagenesis</td>
<td>- Heritable&lt;br&gt;- Can obtain complete loss-of-function&lt;br&gt;- Can use enhancers/gene traps to obtain over-expression lines&lt;br&gt;- Over-expression mutants are useful for examining the function of functionally redundant genes&lt;br&gt;- Mutant libraries available for many species&lt;br&gt;- Site of mutation easily detectable by PCR&lt;br&gt;- Low cost</td>
<td>- May require transformation&lt;br&gt;- Not developed for all species&lt;br&gt;- Variable effects depending on site of integration&lt;br&gt;- No way of obtaining missense mutations</td>
<td>[11–16]</td>
</tr>
<tr>
<td>Fast-neutron mutagenesis</td>
<td>- Heritable&lt;br&gt;- Complete knockout&lt;br&gt;- Can delete more than one gene if linked&lt;br&gt;- No transformation required</td>
<td>- Low frequency of mutations&lt;br&gt;- Screening is laborious&lt;br&gt;- Can delete multiple genes at once&lt;br&gt;- Analysis can be confounded by complex genomic rearrangements&lt;br&gt;- Relatively expensive</td>
<td>[17–20]</td>
</tr>
<tr>
<td>Chemical mutagenesis and TILLING</td>
<td>- Heritable&lt;br&gt;- Can obtain complete loss-of-function&lt;br&gt;- Can obtain missense mutations (may cause dominant or neomorphic phenotypes)&lt;br&gt;- Multiple alleles per locus</td>
<td>- Requires a priori knowledge of sequence&lt;br&gt;- Need a large mutant population&lt;br&gt;- Relatively expensive</td>
<td>[21–28]</td>
</tr>
</tbody>
</table>
complementary to the target gene is transformed into the plant where it is processed by the Dicer pathway to produce the small silencing RNA molecules that target endogenous transcript(s) for degradation. A modified version of this techniques involves the production of interfering RNAs using promoters that are temporally or spatially specific or that are inducible by some exogenous factor [8].

The advantages of RNAi-based reverse genetics are summarised in Table 1 and include the facts that a partial loss of function can be achieved, silencing is directed against a specific gene(s) so the screening of large populations is not required, and transcripts of multiple genes from a family can be silenced by a single construct [9]. This feature is especially useful in plants given that many plants are polyploid or have experienced partial polyploidisation at some stage in their evolution. Other advantages of this technique are that the induced phenotypes are dominant, they can be observed in the T1 generation, and stable inheritance of the transgenic RNAi gene makes the technique suitable for genetic engineering. A disadvantage of the RNAi approach is that some genes are resistant to silencing by exogenous RNA, probably because of sequence or structural features of these genes. In addition, transcripts of genes that are similar in sequence to the target locus may be down-regulated inadvertently as well as the transcripts from the actual target gene. Such 'off-target' silencing may be difficult to interpret [7]. Like VIGS, the silencing level may also be variable and rarely are the genes completely silenced. completely silenced. Furthermore, because RNAi involves transgenic loci that can be silenced in future generations of transgenic lines, the stability of the RNAi effect can sometimes be problematic. This technique, or older versions of it, have, however, been used in many plants to improve productivity or increase disease resistance as well as for basic investigations of gene function [10].

**INSERTIONAL MUTAGENESIS**

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 [11–14]. For insertional knockouts this technology has been used in both monocotyledonous and dicotyledonous plants and remains one of the most effective ways of performing reverse genetics in many species. A population of plants each having an insertion(s) at a unique site in the genome is generated either by transformation (T-DNA) or transposon activation. Such populations with genome-wide insertions have been generated for several plant species as a reverse genetic service. Plants carrying an insertion in a gene of interest can be identified by screening the population with PCR using one gene-specific primer and one insertion based primer [15]. Gene disruption studies using this technique typically result in a total loss of function and the insertion can be easily followed, using PCR, in a population of plants where it is segregating. Some of the disadvantages include the facts that phenotypes may not be obvious if the gene function is redundant and insertions in essential genes will typically result in lethality making these types of genes difficult to examine using this technique (Table 1). In addition, with T-DNA insertion libraries, any given plant will carry at most a few insertions, therefore, very large populations (hundreds of thousands of plants) are needed to achieve genome saturation. This problem is alleviated for species in which endogenous transposons can be activated [16], but for those that don’t have this advantage, generating and screening very large populations for T-DNA insertions is difficult.

**FAST NEURON MUTAGENESIS**

Fast neutron bombardment is used to generate deletions and chromosome rearrangements of various

<table>
<thead>
<tr>
<th>Organism</th>
<th>Population resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td><a href="http://www.arabidopsis.org/abrc/index.jsp">http://www.arabidopsis.org/abrc/index.jsp</a></td>
</tr>
<tr>
<td>Brachypodium distachyon</td>
<td><a href="http://www.brachypod.org/collection.htm">http://www.brachypod.org/collection.htm</a></td>
</tr>
<tr>
<td>Brassica napus</td>
<td><a href="http://www.botany.ubc.ca/can-till/Projects.html#napus">http://www.botany.ubc.ca/can-till/Projects.html#napus</a></td>
</tr>
<tr>
<td>Rice</td>
<td><a href="http://signal.salk.edu/cgi-bin/RiceGE">http://signal.salk.edu/cgi-bin/RiceGE</a>; <a href="http://signal.salk.edu/cgi-bin/RiceGE">http://signal.salk.edu/cgi-bin/RiceGE</a>; <a href="http://orygenesdb.cirad.fr">http://orygenesdb.cirad.fr</a>; <a href="http://irfcr.irri.org/cgi-bin/gbrowse/IR64_deletion_mutants/">http://irfcr.irri.org/cgi-bin/gbrowse/IR64_deletion_mutants/</a></td>
</tr>
<tr>
<td>Wheat</td>
<td><a href="http://www.k-state.edu/wgrc/Resources/resources_index.html">http://www.k-state.edu/wgrc/Resources/resources_index.html</a></td>
</tr>
</tbody>
</table>
sizes randomly in the genome. For reverse genetics the major fast neutron technique currently used in plants is Deletagene [17]. In this technique, seeds are mutagenised with fast neutron radiation and deletions are identified using PCR on pooled DNA using primer sequences that flank the gene of interest. Amplification time is restricted so as to preferentially allow amplification of the mutant DNA where a smaller PCR product (deletion) is generated. The Deletagene system has been used very effectively in both Arabidopsis and rice but is laborious because of the number of plants that must be screened and has limitations in terms of the sizes of deletions that can be recovered. A modified PCR detection strategy involving more steps than the Deletagene protocol has recently been tried on a Medicago truncatula fast neutron population [18]. This involves the use of two rounds of PCR using nested primers. In the first round of PCR amplification a ‘poison primer’ is included that has been targeted to a region of the amplicon that is deleted in the mutant. This poison primer causes aberrant amplification of a partial product from the wild-type but not the mutant DNA. In addition, the first PCR product is digested with a restriction endonuclease whose target cleavage site has been deleted in the mutant. Thus, during the second round of PCR, the mutant product should be better represented than it was initially, allowing detection of the mutation even in very large pools of samples (1:240,000). The authors have named this technique ‘De–TILLING’, although its connection to the original TILLING technique is unclear. Another alternative screening protocol that has been tested in rice and in clementine involves hybridizing genomic DNA to cDNA-based microarrays to detect fast neutron-induced deletions in these plants [19, 20]. Such array-based technology may be useful for this type of reverse genetics screen in other systems as well. For these projects, genomic DNA was labelled and hybridised to the oligonucleotide arrays in order to detect genes, or groups of genes, that had been deleted in the mutant lines. The authors were successful at detecting several deletions, but report that the limitations of this technique include the fact that very small deletions, and deletions in repetitive regions or more complex chromosomal rearrangements could not be detected. This mutation detection technique will also only be effective in species for which enough sequence information is available to design microarrays.

### CHEMICAL MUTAGENESIS AND TILLING

Chemical mutagenesis using ethylmethane sulphonate (EMS) or ethyl nitrosourea (ENU) induces point mutations in DNA in all species in which it has been tested. Generation of mutant populations using these or similar mutagens was an established practice long before the advent of gene sequencing and reverse genetics. Mutations induced using these mutagens are distributed in the genome randomly and because point mutations are less damaging than large rearrangements, a high degree of saturation can be achieved in a mutant population, more easily enabling the examination of gene function on a genomic level. In addition, while most reverse genetics techniques provide only loss-of-function alleles, chemical mutagenesis can result in either loss-of-function or gain-of-function mutations causing null, hypomorphic, neomorphic or hypermorphic phenotypes. While hypomorphic alleles may be most useful for determining wild-type gene function, neomorphic or hypermorphic alleles are more likely to be dominant and so are more likely to cause an observable phenotype in cases of either redundant or essential genes. The frequency of missense alleles is on average three times higher than that of non-sense alleles [21], but it is difficult to predict how many of these mutations will have an effect on gene function since many of them may not alter the gene product(s) significantly. Examples of dominant point mutations that do have an effect on gene function have, however, been well documented in several cases, for example, the ethylene response pathway [22], leaf polarity determination [23] and host-pathogen defense [24].

For forward genetics, mutagenised populations are screened for overt phenotypes, and then genetic and molecular analysis is used to discern the sequence of the gene responsible. The difficulty with using point mutations for reverse genetics is that there have not been many direct, cost-effective ways of screening for individuals from the mutagenised population that carry mutations in specific genes of interest. The advent of TILLING introduced such a method of screening at a reasonable cost within a reasonable time frame. Most TILLING operations use the technique described by Colbert et al. [25] that employs a mismatch-specific endonuclease found in celery for identifying SNPs in genes of interest. In this procedure, DNA from several
different plants is pooled and then used as a template for PCR amplification with fluorescently tagged, gene-specific primers. The PCR products (amplicons) are denatured and allowed to randomly re-anneal before being digested with a celery juice extract (CJE) [26]. Mismatches in the amplified DNA occur, following renaturation, if a pool of DNA includes at least one plant with a mutation in the amplified region. When a DNA strand from a mutant amplicon anneals to the complementary strand of the wild-type amplicon a heteroduplex is formed that presents a target for the CJE enzyme. When cleavage of some of the PCR products at the mismatch occurs, the novel fragments can be detected using a LI-COR DNA Analyser (LI-COR Biosciences, Lincoln, NE) or other equivalent DNA separation technology, thus identifying plants with mutations in the gene of interest. Modifications of this procedure, such as the use of an Arabidopsis endonuclease (ENDO-1) [27] instead of CJE, have been tested with similar results. TILLING is robust, and has worked well in a wide variety of plant species [28]. In addition, until recently, it has been the least expensive and one of the only techniques for identifying de novo SNPs or point mutations in genomes for which the entire sequence is not available. However, a mutagenised population arrayed for TILLING must be available and for many species the development of such a population is expensive and time consuming. Further, the technique itself is labour-intensive and it has proven more difficult to obtain large numbers of mutations in some genes because of the technical requirements for primer design and visualisation using CJE and the Li-cor gel apparatus. Finally, results must be verified by sequencing adding to the time and cost.

NEXT-GENERATION SEQUENCING FOR REVERSE GENETICS

With the advent of next-generation sequencing (NGS) technology it is now conceivable to Sequence Candidate Amplicons in Multiple Parallel Reactions (SCAMPR) for mutations in any gene at costs comparable to current TILLING procedures. Direct sequencing identifies the position and nucleotide change for each mutation at the time of detection and also provides haplotype information (which is currently not available from TILLING data). The question of which NGS technology would be the simplest and most cost effective for screening populations by SCAMPRing still needs to be addressed. In addition, the pooling procedures needed for practical application of this technology to mutation/SNP detection must be determined. There are two major classes of NGS machines for which a reasonable amount of data is already available: short read (25–75 bp; for example the Illumina Genome Analyser or ABI SOLiD) and longer read (400–500 bp; like the Roche454) [29]. The 454 technology is more expensive than the short-read technologies at this time, but should require less depth of sequencing to distinguish between bona fide mutations and background sequencing errors within a population. Both 454 and Illumina technologies have now been tested on TILLING populations, but not directly compared with other TILLING techniques [30, 31]. Keygene NV (Wageningen, The Netherlands) has tested the Roche 454 GS FLX using 5-fold pooled DNA for TILLING. They sequenced DNA from 15 000 M2 tomato plants (representing 3008 M1 lines) in 28 pools on a full picotitre plate using a three-dimensional labelling strategy with 12X, 8Y and 8Z column, row and plate pools. They amplified a region of 287 bp from a single locus and obtained total of 580 471 sequencing reads which they analysed using in-house software designed for this purpose (not publicly available). On the basis of this trademarked protocol (named KeyPoint technology), they identified two mutations that were confirmed by Sanger sequencing.

The Comai lab at the University of California at Davis Genome Center has also been testing TILLING by NGS but using the Illumia GA2 machine to sequence PCR amplicons [31]. They offer a TILLING service in both rice and tomato as well as workshops to train other laboratories in this novel technique. This group is using a two-dimensional pooling strategy of 96-fold row pools and 64-fold column pools. One of the most difficult problems with using NGS for mutation detection is separating the true mutations from the sequencing errors generated by chance in the thousands of reads that are produced in each run. The UC Davis TILLING lab reports a variable success rate for this that is dependent on the species and pooling depth, but averages <0.7. No publications have resulted from this TILLING service at this time, but all data and protocols are being made publically available.

In our laboratory, a *B. napus* TILLING population of ~1800 M2 plants has been constructed and screened for mutations in five different genes using CJE/LI-COR TILLING. Using a standard TILLING protocol with an amplicon of 1500 bp, we have identified an average of 20 mutations per gene for each of the five genes. This population provides an ideal basis from which to compare results, timing and cost of our standard TILLING protocol with different NGS technologies for mutation detection. We have developed a three-dimensional pooling strategy to re-screen this population using both the Roche 454 FLX Titanium and the Illumina GA2. The pooling scheme used for SCAMPRing is key because the success of this technology for reverse genetics depends on whether or not it is cost-effective. A considerable number of samples must be sequenced at once to be cost effective but, on the other hand, if too many samples are pooled the rare mutations in the population will be indistinguishable from the background sequencing errors that are inherent in these technologies. In addition, the pooling depth is limited by the ability to distinguish one sample from another in a single run. Samples are distinguished using multiplex indexing (MID) primer tags which are 10 bp unique sequences that are recognised by the analysis software and used to distinguish between the products amplified from different pools. Our pooling strategy involves amplification using DNA pooled from the 96 plants in each plate, 144 plants from a given row and 96 plants from a given column. PCR amplification will be done on 4-fold pools for the five different genes, using six primer sets each to cover ~2.5 kb per gene (completely overlapping so that 2× coverage is obtained for each locus). These samples will then be sequenced using one half of a picotitre plate on the Roche 454 GS FLX Titanium. This would provide ~5.5× sequencing coverage, but because of the overlapping amplicons this coverage doubles to 11×. In reality, coverage of original genomic sequence will be three times this number because three biological replicates of each line will be sequenced (column, row and plate pools) resulting in a total of 33× coverage of the original sequence. The estimated coverage required to minimise false positive and false negative errors using the old 454 technology and one-dimensional sequencing was calculated to be ~34-fold [32] so our test run should provide almost this depth. Further, we will be using the longer read Titanium technology, sequencing in three-dimensions providing greater statistical power for data analysis and eliminating the number of false negatives and false positives that have plagued other amplicon sequencing studies. In parallel with the 454 sequencing SCAMPR testing, we will also test the Illumina GA2 using the same pooling strategy. If the GA2 delivers 10 Gb per run as advertised, this would amount to >350× coverage and the coverage needed to ensure <10% error rate of false positive and negative errors using the short-read GA technology has been calculated to be 110-fold [32], indicating that we should be well within the margin of error using this strategy. The results from these tests should allow us to directly compare SCAMPRing to traditional TILLING technique in terms of cost, labour and reliability since mutations identified in our previous TILLING screen can be compared to the ones identified by SCAMPRing.

CONCLUSION
There are many different tools and resources available for reverse genetics studies in plants. All have their advantages and disadvantages depending on the species and researcher and the questions being asked. The low cost of VIGS and RNAi make those techniques attractive for studying genes of unknown function in some species where they have been developed, and the public availability of T-DNA and transposon insertion lines make those resources attractive as well. Physical mutagenesis followed by reverse genetic analysis is more costly than the above, but is useful in cases where other techniques have not been successful. Chemical mutagenesis and TILLING is also costly, but may provide more mutant information than is available using any other technique. Finally, with the continuous development of new technologies, the most cost-effective technique for examining gene function in plants in the future may involve direct sequencing of part or complete genomes or some other genome analysis technique that is yet to be developed.

FUNDING
Genome Canada/Genome Alberta genomics project: Designing Oilseeds for Tomorrow’s Market.
Key Points

- There are advantages and disadvantages to all of the reverse genetics techniques currently in use (Table 1).
- Both reverse and forward genetics have been used to generate populations of mutant plants, several of which are available to the public. A subset of these is summarised in Table 2.
- While it is clear that many of the reverse genetics techniques currently used in plants will continue to be useful for some time to come, the advent of new technologies, particularly next-generation sequencing, is likely to revolutionise the state of reverse genetics.
- We predict that, in the foreseeable future, the most cost-effective technique for examining gene function in plants will be the sequencing of the entire genomes of a population of thousands of individuals.
- One of the main difficulties with the use of next-generation sequencing technologies for reverse genetics is the inability to differentiate between actual mutations or SNPs in the DNA and sequencing errors (or false positives). This problem should become easier as more and more data is generated and programmers are able to refine their algorithms. Thus, the construction of mutant populations and development of programmes and infrastructure for in silico screening and storage of DNA sequences will be key to the future of reverse genetics in plants as well as other organisms.

Acknowledgement
The authors are grateful to Andrew Sharpe and the NRC/Plant Biotechnology Institute for discussion and future collaboration with SCAMPRing, Tanya Machlines for technical assistance in constructing our B. napus population and TILLING, and David Kaplan for greenhouse support.

References


