Review TRENDS in Plant Science Vol.11 No.2 February 2006

Harvesting the potential of induced biological diversity

Robbie Waugh[1], David J. Leader[1], Nicola McCallum[1] and David Caldwell[1,2]

[1]Scottish Crop Research Institute, Invergowrie, Dundee, UK DD2 5DA
[2]Current address: Michelmore Lab, UC Davis Genome Center, GBSF Building, Room 4302, 451 East Health Sciences Dr, Davis, CA 95616, USA

For most of the past century, chemical and physical mutagens have been used in plant genetic research to introduce novel genetic variation. In crop improvement, more than 2000 plant varieties that contain induced mutations have been released for cultivation having faced none of the regulatory restrictions imposed on genetically modified material. In plant science, mutational approaches have found extensive use in forward genetics and for enhancer and suppressor screens – particularly in model organisms where positional cloning is easily achieved. However, new approaches that combine mutagenesis with novel and sensitive methods to detect induced DNA sequence variation are establishing a new niche for mutagenesis in the expanding area of (crop) plant functional genomics and providing a bridge that links discovery in models to application in crops.

Mutagenesis: inducing biological diversity

Mutagenesis has remained popular over the past 70 years because it is simple, cheap to perform, applicable to all plant species and usable at a small or large scale [1]. By varying mutagen dose, the frequency of induced mutations can be regulated and saturation can be readily achieved [2,3]. Historically, mutation screens have assembled rich collections of phenotypic variants. However, five years ago a novel application for mutagenesis that avoids *a priori* screening for phenotypes was established. The approach, which couples mutagenesis with sequence-specific mutation detection, promises to provide a general and effective platform for 'reverse genetics' in a wide range of plants. Today, several variants of this basic strategy have been developed and acronyms such as TILLING (Targeted Induced Local Lesions IN Genomes), Deleteagene™ and DEALING (DEtecting Adduct Lesions IN Genomes) have been coined.

Reverse genetics is a generic term for approaches that attempt to determine the biological function of a gene or protein by analysing individuals in which that gene has been deliberately perturbed. In plants, reverse genetics has relied heavily on insertional mutagenesis using mobile genetic elements such as T-DNA and endogenous or introduced transposons. In *Arabidopsis*, rice and maize, large populations have been developed that contain insertions dispersed throughout the genome [4–6] (Table 1a). Gene-specific polymerase chain-reaction (PCR)-based screening for insertions has now given way to *in silico* screens of databases that contain the DNA sequences flanking all insertions in a given population [5] (Table 1a), and lines containing insertions in or adjacent to a gene of interest can be ordered directly online. Insertional mutagenesis is an immensely powerful tool but the drawbacks are: (i) insertion sites are non-random, with certain genes less susceptible to insertion than others; (ii) induced mutations generally produce a complete loss of function allele that might not provide much insight into the normal function of the gene; (iii) to date, the approach has been restricted to a few (model) species; (iv) because many insertional-mutagenesis strategies rely on genetic manipulation to develop the resource, the incorporation of promising experimental lines directly into breeding programmes is highly restricted.

Deletogens

Chemical and physical mutagens induce a high frequency of mutations at random locations across the genome (Box 1). However, translating the potential of this well-established feature into an effective reverse genetics strategy has remained elusive. In 1997, that changed. Gert Jansen and colleagues used the chemical diepoxybutane (DEB) to induce gene and sub-gene-sized deletions in the genomic DNA of *Caenorhabditis elegans* and, in parallel, developed a PCR approach [critical extension-PCR or CE-PCR (Figure 1a)] that simplified the identification of deletions in selected genes [7]. The strategy was as follows: a synchronized population of animals was mutagenized with DEB. DNA was then isolated from the bulked F2 progeny from groups of approximately 20 F1 individuals and arrayed in 48 microtitre plates. Each pooled sample of ~40 randomly mutagenized *C. elegans* genomes was then incorporated into microtitre plate superpools comprising ~4000 mutated genomes. The relatively short deletions induced by DEB in *C. elegans* (average 1400 bp; range 700–2900 bp) ensured that shorter than wild-type fragments were preferentially amplified by CE-PCR against a vast excess of wild-type fragments, which allowed lesions to be identified in 127 genes, representing a success rate of more than 96% of genes targeted [8], amply illustrating

Corresponding author: Waugh, R. (rwaugh@scri.sari.ac.uk).
Available online 9 January 2006

 1360-1385/$ - see front matter doi:10.1016/j.tplants.2005.12.007

Table 1. Web links

a	http://www.*Arabidopsis*.org/links/insertion.jsp	*Arabidopsis* insertion, knockout and mutation resources
b	http://www.affymetrix.com/products/arrays/index.affx	Affymetrix microarrays
c	https://www.fastlane.nsf.gov/servlet/showaward?award=0077737	Summary of the ATP NSF award
d	http://tilling.fhcrc.org:9366/files/Welcome_to_ATP.html	ATP home page. Several of the web-based utilities can be accessed directly from here as well as information on the history and development of TILLING
e	http://www.arcadiabio.com/	Arcadia Biosciences, Inc. homepage
f	http://www.WGIN.org.uk	British Wheat Genetic Improvement Network homepage – TILLING in wheat
g	http://www.botany.ubc.ca/can-till/	CAN-TILL homepage – TILLING in poplar
h	http://www.soybeantilling.org/index.jsp	Soybean TILLING homepage
i	http://nsf.gov/awardsearch/showAward.do?AwardNumber=0408768	Development of array screening to identify mutant alleles in mutagenized maize population
j	http://tilling.fhcrc.org:9366/arab/status.html	Summarizes the up-to-date progress of the ATP
k	http://www.licor.com/bio/Tilling/TillingMain.jsp	Describes the LI-COR platform and its specific use for TILLING
l	http://www.proweb.org/gelbuddy/download.html	Gelbuddy software for LI-COR gel analysis
m	http://www.proweb.org/coddle/	Portal to the CODDLE utility
n	http://blocks.fhcrc.org/sift/SIFT.html	Portal to the SIFT utility
o	http://www.proweb.org/parsesnp/	The PARSESNP utility
p	http://tgrc.ucdavis.edu/Monogenic-stocks-2002.pdf	R.T. Chetelat, revised list of tomato monogenic stocks at UC Davis, CA, USA
q	http://www.untamo.net/bgs/	Barley morphological mutants
r	http://zamir.sgn.cornell.edu/mutants	A database containing phenotypic descriptions and photographs of a wide range of tomato mutant phenotypes. 'The genes that make tomato'
s	http://www.lotusjaponicus.org/	Lotus phenotypic database
t	http://germinate.scri.sari.ac.uk/barley/mutants	Portal to a structured mutant population of barley providing phenotypic descriptions, photographs and access to a barley TILLING service
u	http://www.niob.knaw.nl/researchpages/cuppen/index.html	*Caenorhabditis elegans* and Rat TILLING in the Cuppen Group. Includes protocols and recipes

Abbreviations: ATP, *Arabidopsis* TILLING Project; NSF, National Science Foundation.

the potential of a chemical mutagenesis-based reverse genetics strategy. Unfortunately, attempts to reproduce the induction of similar-sized deletions by DEB in plants have not been as successful. It appears that although DEB is an effective mutagen in rice [9], *Arabidopsis* and barley, it induces only small 1–2 bp deletions rather than the larger (>700 bp) deletions necessary for CE-PCR screens (R. Waugh *et al.*, unpublished), although the

Box 1. Mutagenic agents

Mutagens can be classified as forms of energy or chemical substances that significantly increase the frequency of mutations in the genomes of exposed organisms. Two general categories of biological effects result from exposure to mutagens: somatic and genetic. Genetic effects are heritable because they are present in the germ line of the affected individual. Induced germline mutations have been used as a source of novel variation in both crop plants and experimental organisms since the 1920s. The mode of action of many mutagenic agents is known.

Forms of energy

Examples include ultraviolet light, electro magnetic waves (X-rays, gamma rays and cosmic rays) and fast moving particles (α particles, β particles and neutrons). Some have low energy (e.g. UV light) and cause relatively little damage except after prolonged exposure, others have high energy and are extremely damaging to DNA (e.g. fast-neutrons). UV light is absorbed by pyrimidines in DNA, causing adjacent bases on the same DNA strand to bond covalently to form pyrimidine dimers that subsequently cause errors during DNA replication. X-rays, gamma rays and cosmic rays have much more energy. They ionize water and other molecules, forming radicals that break DNA strands and alter purine and pyrimidine bases. Fast-moving particles such as fast neutrons have sufficient energy to physically 'punch holes' in DNA.

Chemical substances

These generally work in one of three ways:

- Base analogues such as 2-amino purine (which resembles adenine) and 5-bromouracil (which resembles thymine) get incorporated into DNA during replication. However, they do not have the same hydrogen bonding properties. 5-Bromouracil incorporated instead of thymine will pair with guanine (instead of adenine), which results in the incorporation of cytosine into the daughter DNA strands during subsequent rounds of replication (resulting in A/T to G/C transitions).
- Intercalating agents such as ethidium bromide, proflavin and acridine orange are compounds that slip between adjacent base pairs in DNA, reducing the fidelity of DNA replication and cause insertions, deletions or additions that frequently induce frameshifts.
- Base-modifying agents include alkylating agents [e.g. ethyl methane sulfonate (EMS), *N*-ethyl-*N*-nitrosourea (ENU)], deaminating agents (e.g. nitrous acid and nitrosoguanidine) and hydroxylating agents (e.g. hydroxylamine). EMS adds alkyl groups to the hydrogen-bonding oxygen of guanine to produce *O*-6 alkylguanine, which pairs with T (instead of C) and causes G/C to A/T transitions. Nitrous acid converts cytosine to uracil (by oxidative deamination), which forms hydrogen bonds with adenine rather than with guanine and causes A/T to G/C transitions. Hydroxylamine reacts with cytosine, converting it to a modified base that pairs only with adenine and results in C/G to T/A transitions.

For other chemicals, such as diepoxybutane (DEB) and sodium azide, the precise mechanisms of action remain unknown.

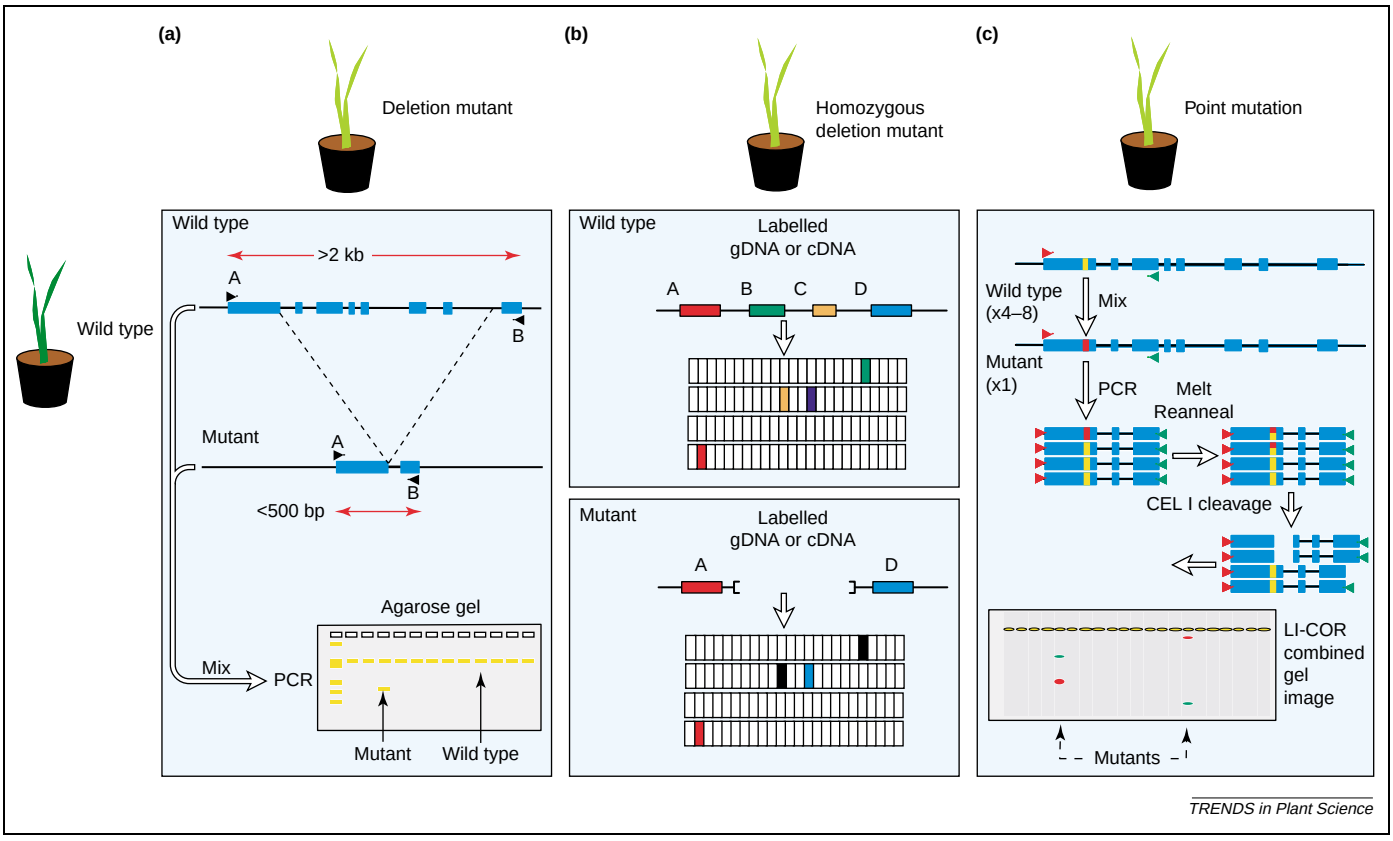

Figure 1. Detection of different mutant classes. **(a)** Deletions within candidate genes can be detected by a simple PCR method. Gene-specific primers (depicted by black arrow heads) are designed to amplify between distal regions of the genomic sequence. Deletions generated by chemical agents such as DEB or fast neutron irradiation can bring the primer binding sites into closer proximity. PCR conditions designed to amplify the shorter amplicons preferentially or selectively allow identification of a single mutant within a large pool of wild-type individuals (up to 4000) using simple agarose gels. **(b)** Large (>2 kb) deletions might also remove entire genes, resulting in a mutant phenotype in homozygous individuals. In some circumstances, it might be possible to identify directly the genes affected by these deletions using near whole genome or expression microarrays such as those produced by Affymetrix. Probe sets are identified that hybridize to wild-type, labelled cDNA or gDNA but fail to hybridize to a comparable probe from homozygous mutant individuals. These probe sets potentially represent genes not present in the mutant and delineate the deletion. **(c)** Reverse genetics screening for single point mutations (e.g. generated by EMS) can be achieved in four- to eightfold pools of DNA isolated from M2 individuals. Primers labelled with two different fluorophores (depicted by red and green arrowheads) are designed to amplify ~1 kb of genomic DNA containing conserved exons of the target gene. Pooled DNA samples are used for amplifications and the resulting amplicons heated and allowed to re-anneal. Pools that contain a point mutation in one of the component amplicons will produce heteroduplexes with the wild-type amplicons. These are a suitable substrate for mismatch cleavage by CEL I, producing DNA fragments of less than full length. These fragments are typically visualized using acrylamide slab gels (such as the LI-COR system). Lanes from mutant pools will contain both the full-length wild-type DNA labelled with both fluorophores as well as two smaller fragments labelled with each fluorophore. Crucially, the sum of the lengths of these fragments should be approximately the length of the wild-type fragment.

reasons for this are not entirely clear. DEB, trimethyl psoralen and UV irradiation are currently being evaluated as deletogens in diploid wheat as the basis of the DEALING methodology (S. Kianian, unpublished).

In contrast to DEB, a low dose of fast neutron irradiation has been used successfully to generate a large population of independently induced deletions in *Arabidopsis* in an approach coined 'Deleteagene™' [10,11]. Starting with a population of 51 840 lines, Xin Li and colleagues used a CE-PCR-based mutation detection method to identify lesions in 21 of 25 genes targeted, including a single deletion encompassing a three-gene tandem array. Finding a single deletion encompassing a three-gene tandem array is potentially attractive because it could help solve problems arising from 'phenotypic masking' caused by functionally redundant paralogous gene family members. The success rate that Li *et al.* obtained suggested that a population of 85 000 individual M2 lines of *Arabidopsis* would provide a >95% chance of detecting a deletion in any gene. Two significant attractions of the Deleteagene™ approach are (i) the depth of pooling possible (these authors had 2592 pooled DNAs in their primary 'megapool' screens and still detected deletions), and (ii) that the majority of the mutations result in complete loss of gene function. However, maintaining such large populations is a logistical challenge. Furthermore, the effectiveness of the Deleteagene™ approach in crop plant genomes that frequently have a paucity of long-range sequence information surrounding any target gene and have genomes more than 100 times the size of *Arabidopsis* remains to be demonstrated. The use of oligonucleotide arrays to identify genes conferring a phenotype arising from deletion mutagenesis is a promising new approach even in large-genome crop plants. DNA or RNA from wild-type and mutant plants is labelled and hybridized to high-density microarrays such as Affymetrix Genechips™ [12,13] (Table 1b). The differential hybridization of DNA or RNA from wild type (present call) versus mutant (absent call) to probes

or probe sets on the array is indicative of a mutation in the corresponding gene and can even delineate the deleted region (Figure 1b).

Point mutations

In contrast to 'deletogens', several mutagenic agents have been used to induce point mutations in the genomes of a diverse range of plants (Box 1) [1,14–16]. Of these, ethyl methane sulfonate (EMS) is emerging as the 'mutagen of choice', largely because of its well established mode of action, which generates G to A and C to T transitions [17], and its effectiveness in inducing a high frequency of point mutations in a wide range of organisms in the absence of gross chromosomal abnormalities (henceforth, for clarity, we will refer only to EMS when discussing point mutations, although the discussion is appropriate to any potent mutagen). The breakthrough in exploiting point mutations for reverse genetics came in 2000, when Claire McCallum and colleagues [18–21] in Seattle developed a general strategy that they christened 'Targeted Induced Local Lesions in Genomes' or TILLING. In short, they efficiently coupled saturating EMS mutagenesis with a sensitive method for detecting induced point mutations in pooled DNA samples (Figure 1c). Since then, the emergence of TILLING has been impressive. Proof-of-principle studies were followed rapidly by National Science Foundation funding (Table 1c) that established the *Arabidopsis* TILLING Project (ATP) (Table 1d) to identify and provide induced mutations in any gene to users in the research community. A patent application was filed and a company (Anawah, Inc., Seattle, WA, USA) spun out to exploit TILLING for commercial crop improvement. Anawah is now part of Arcadia Biosciences, Inc. (Table 1e). Its impact on the research community has been similarly significant. TILLING has been demonstrated to be effective in maize [22], wheat [23], *Lotus* [24] and barley [25]. Furthermore, several publicly funded TILLING projects are under development in crops including wheat and *Triticum monococcum* [Andy Phillips and Kim Hammond-Kosack (Rothamsted Research, UK), and Rob Koebner (John Innes Centre, Norwich, UK), personal communication] (Table 1f), poplar and brassicas (Table 1g), pea, soybean and *Medicago* (Table 1h). TILLING has become an accepted and commonly used acronym for most chemical mutagenesis-based reverse genetics in the plant research community and its potential for application in both basic and applied research has been widely recognized [26–29]. Array profiling methods described above can also be extended to identify chemically induced point mutations (in DNA or RNA) by single feature polymorphism (SFP) analysis [13,30–32] in spite of the probe capacity of currently available oligonucleotide arrays, which covers only a small proportion of the coding sequence. However, as the density of oligonucleotide arrays continues to increase this should become a useful approach that might eventually surpass standard TILLING methodologies for SNP (single nucleotide polymorphism) or small deletion detection. This platform is currently being developed for mutation detection in maize (Table 1i).

Creating structured mutant populations

For reverse genetics, structured populations are essential to track any discovered mutations back to the families from which they originated. Figure 2a describes the approach adopted in inbreeding *Arabidopsis*, *Lotus*, barley and wheat [18,23–25]. The best example of how the whole process works is the ATP: a population of ~3000 M2 plants with a mutation frequency averaging 1 lesion per 170 kb forms the basic biological platform [21]. As of May 2005, >460 fragments had been screened for mutations, detecting >5600 mutations that have been distributed back to the community as M3 seed (Table 1j).

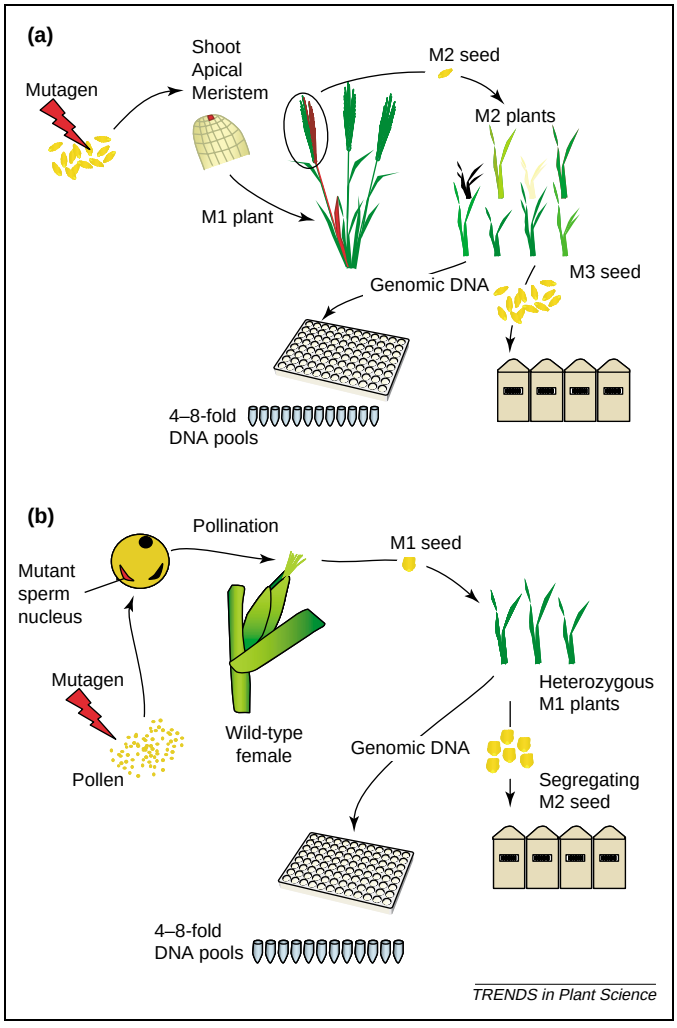

Figure 2. Generation of structured mutant populations. **(a)** Seed mutagenesis. Batches of M0 seed are exposed to an appropriate dose of the chosen mutagen. This generates random lesions in the genome of cells within the embryo shoot apical meristem. Subsequently, tissue sectors derived from this cell by mitosis will be heterozygous for the mutant allele. Some of these cells will develop into the reproductive organs. Pollination of egg cells by selfing the M1 effectively purifies a unique collection of mutations in each individual M2 seed. M2 plants will therefore contain all the mutations transmitted through the seed, and these will be maintained (and segregate) in the M3 and subsequent generations. DNA is prepared from tissue of each M2 plant. These are given a unique identity code or bar code to allow tracking, and the self-pollinated M3 seed is collected and stored. For mutation detection, multiple M2 DNA samples are pooled at varying depths according to the detection protocol. **(b)** Pollen mutagenesis. Pollen is mutagenized and used to fertilize wild-type embryos generating heterozygous M1 seed. All seed on a cob will contain a different collection of mutations. DNA is harvested from the M1 plants and pooled for mutation screening; the selfed M2 seed is archived as the biological resource for maintenance or phenotypic analysis.

US$1000 is the current levy to academic laboratories for TILLING of a ~1 kb fragment, which is a cost-effective approach. A similar service is currently being established for barley and other species.

In maize, Bradley Till and colleagues adopted a modified strategy [22] (Figure 2b). They mutagenized pollen from the maize inbred B73 and applied it to the silks of wild-type B73 ears. As a result, each M1 seed was immediately heterozygous for each mutation present in its contributing male gamete. DNA was isolated from 750 M1 plants, and M2 seed archived for phenotypic analysis. By screening 11 PCR amplified gene fragments they were able to identify and verify 17 induced mutations, translating into a mutation frequency of slightly less than 1 mutation per 500 kb. A logical extension of pollen mutagenesis might be to exploit microspore culture when it is both routine and highly efficient. All non-lethal mutational events would be captured immediately in microspore-derived plants and these would collectively represent a population of fixed inbred lines, simplifying screening, phenotypic analysis, storage and maintenance (all be it at the cost of homozygous lethal or highly deleterious mutations).

A remaining challenge is the development of reverse genetics populations in highly heterozygous outbreeding or vegetatively propagated plants. Traditional mutagenesis is used extensively in, for example, floricultural plants and fruit trees. In such material, irradiation of rooted stem cuttings, detached leaves or dormant plants is common, although *in vitro* cultured or micropropagated plants, regenerable callus cultures, stolons, axillary or adventitious buds have also been used [1,14,33]. The major hurdle is overcoming the need for sex to purify mutations biologically (or plants will remain chimaeric). One option is to consider mutagenizing single-cells that can subsequently be regenerated into plants. Protoplasts or embryonic callus cultures are an immediate possibility, particularly where highly efficient regeneration systems already exist (e.g. for potato), although the frequency of chromosome abnormalities (e.g. spontaneous doubling) in a target organism should be established. A less technologically demanding approach might be to repeatedly subculture single-meristem-derived tissues such as single nodal buds or tubers over multiple cycles. Alternatively, a final (but intuitively messy) possibility is to perform seed mutagenesis and simply live with the consequences of segregation at all heterozygous loci in the developed population. Importantly, the procedures deployed for mutation detection will also potentially have to contend with heterozygosity at any locus. Although this might be achievable for point mutations through smart primer design or by modifications of the mutation detection assay, deletion-based approaches would be more amenable.

Mutation frequency and population size: getting the balance right

Optimizing mutation frequency is paramount and must be empirically determined: if it is too low, too many plants will be required to discover mutations in a target gene; if it is too high, viability and/or sterility is likely to be a problem. Traditionally, it has been common for mutation frequency to be estimated on the basis of phenotype, using screens for embryonic lethality (seed set), seedling lethality, chlorophyll deficiency or single-copy gene phenotypes as a measure [19,28,29,34–38]. Unfortunately, there are no direct measures to assess mutation frequency globally based on genotype. Amplified fragment length polymorphism (AFLP) has been used to estimate somaclonal mutation frequency induced by tissue culture [39]. AFLP exploits four- and six-cutter restriction endonuclease sites and extension nucleotides combined with PCR amplification to generate anonymous molecular profiles comprising 50–100 short amplified genomic DNA fragments [40]. Because each fragment is the product of a specific combination of nucleotides at each end [six for the restriction site plus three selective nucleotides (nt) on one end, and four plus three at the other], a 100-fragment fingerprint represents ~1.7 kb of DNA sequence per assay [100 fragments×(6+3 +4+3 nt)]. Thus, running 96 lanes allows ~160 kb of sequence to be screened in a single experiment. We used AFLP to assess EMS, DEB and gamma-irradiation-induced mutation frequency in barley by scoring mutations as the addition or loss of bands in a uniform background (R. Waugh *et al.*, unpublished). Because only homozygous loss, and homozygous and heterozygous addition of bands can be scored, AFLP scanning is quantitative, providing only an estimate of mutation frequency. The best approach is to perform a mutation screen on a few targets, confirm induced alleles by re-sequencing and directly calculate the effective mutation frequency. Although the logistics and cost efficiency of doing this for large quantities of plants can appear daunting, the information it provides will unequivocally determine the overall utility of a population.

In the literature, effective mutation frequencies range from 1 per 24 kb in hexaploid wheat [23] to less than 1 per 500 kb in maize [22] and barley [25]. At these frequencies, populations need not be excessively large. At one extreme, in hexaploid and tetraploid wheat, only 1920 plants were required to identify 246 induced mutations in the waxy locus (granule bound starch synthase I) [23]. This exceptionally high mutation frequency was attributed to the polyploid wheat genome buffering against deleterious mutations. Surprisingly, even with this mutational load, few morphological variants were observed. However, in polyploids, the ease with which mutations can be discovered is balanced by the crossing and selection required to assess their impact effectively because it is likely to be necessary to combine homozygous disruptions in all homoeologous alleles to test for phenotype. Similarly, the optimization of PCR conditions to allow homoeologous allele-specific amplification required for screening is far from trivial and needs careful consideration. In diploids, working populations comprise 3072 individuals (extended to 6912) in *Arabidopsis* [22], 3697 in *Lotus* [24], 8600 in barley [25] and are being extended to 10 000 in maize (quoted in [23]).

Mutation detection and verification

Table 2 summarizes several approaches that have been developed to assay SNPs and insertions or deletions. In

Table 2. Assays for single nucleotide polymorphism or insertion or deletion discovery

Method	Mutations detected	Sensitivity	Platform	Basis	Pros	Cons
Heteroduplex analysis	SNP, small in/del	2–16 alleles	dHPLC	Detects different single-stranded DNA conformations	No post-PCR processing	Machines can be expensive Sensitivity depends on context of polymorphism and low sample throughput
Cel I, Surveyor	SNP, small in/del	8–32 alleles	dHPLC, DNA fragment analysis system	Detects changes in DNA size following cleavage at mismatched nucleotides	Can detect all mismatches in fragments up to 1 kb. Sensitive pooling allows for high-throughput mutation scanning	Sensitivity and cost varies with platform Infra-red dye primers expensive but cost can be reduced, see methods in Table 1r
SSCP	SNP, small in/del	2 alleles	Polyacrylamide gel electrophoresis	Detects different single-stranded DNA conformations	Inexpensive	Low reproducibility and sequence-dependent detection
TGCE	SNP, small in/del	2–16 alleles	Temperature gradient capillary electrophoresis	Detects different melting characteristics of heteroduplex and homoduplex DNA	Fast and requires no post-PCR processing	Equipment can be expensive Sensitivity dependent on location of polymorphism
CE-PCR	Large in/del	Up to 4000 alleles	Agarose gels	Selective amplification across regions containing deletions	Inexpensive Allows extremely deep pooling	Can only detect medium to large deletions
GIRAFF	SNP, small to medium in/del	Genome size dependent	DNA size fractionation Agarose gel Southern blots	Cell double-strand digestion of heteroduplex, size fractionated restriction digested genomic DNA followed by Southern blot analysis	No PCR Little specialist equipment Inexpensive detection in large DNA fragments	Inexact mapping of polymorphisms Uncertain relevance to large genomes
Microarray (e.g. Affymetrix)	SNP, small to large deletions	1–2 alleles	GeneChip™	Hybridization of genomic DNA or RNA to oligo chip. Changed hybridization pattern indicates polymorphism	Massively parallel and high-throughput	Not available for most organisms Complex specialist analysis Expensive Predetermined sequence coverage
Direct Sequencing	All	2 alleles	DNA fragment analysis system	Sequencing of PCR amplified DNA from different lines	Requires no downstream confirmation of polymorphism	Expensive equipment Small fragments Laborious Sequencing of both strands required to confirm

Abbreviations: In/del, insertion or deletion; SSCP, single strand conformation polymorphism; TGCE, temperature gradient capillary electrophoresis.

the original TILLING method, the authors used heteroduplex analysis by denaturing high-pressure liquid chromatography (dHPLC) to detect mutations in pooled DNA samples [18]. The development of a DNA mismatch cleavage assay using an endonuclease purified from celery (CEL I) that cleaves heteroduplex dsDNA at single base mismatches was a significant advance [41–43]. Currently, CEL I is distributed by Transgenomic (http://www.transgenomic.com) under the trademark Surveyor™ nuclease [44] – however, straightforward and effective protocols for its purification have been published [45]. TILLING combines CEL I cleavage with gel electrophoresis on LI-COR, fluorescence-based fragment analysis systems (Table 1k). The system is rapid, high throughput and relatively tolerant of PCR amplicon quality. In addition, software has been developed that facilitates the rapid analysis of the gel images [46] (Table 1l). By adding different fluorescent dye labels to each of the PCR primers, CEL I cleavage generates different sized and coloured fragments that together total the size of the uncleaved amplicon. This key piece of information allows potential false positives to be excluded from subsequent analyses and provides accurate information regarding the location of the mutation (Figure 1c). Routinely, equal quantities of the DNA of four to eight individuals are pooled and mutations detected in four steps: PCR amplification, heteroduplex formation, CEL I cleavage and detection of the cleaved products by fragment analysis. In addition to gel electrophoresis, fragment analysis by capillary-based DNA sequencers has also been assessed.

After de-convoluting pooled DNAs and re-sequencing alleles from individual plants, sequence comparison with the parental allele will confirm an induced mutation, define its nature and determine whether it is consistent with the known mode of action of the mutagen used. The archived M3 family seed can then be accessed to provide seed of the affected families. However, as populations can suffer from contamination at source or during development (e.g. from outcrossing) re-sequencing the same gene from a range of genetic materials (e.g. alternative cultivars) will separate induced alleles from natural variants. Low-level contamination within a population is

also indicated by a common pattern of CEL I cleaved fragments in multiple pools. The link between induced mutation and phenotype can then be assessed by sibling analysis of heterozygous M3 families directly or in selfed families of homozygous M3 plants backcrossed to the original parent. In practice, there is seldom need for extensive backcrossing or further population development to prove this link, particularly when multiple independent alleles have been induced. Furthermore, the need to remove background mutations is generally not required to assess potential phenotypic impact, although the development of isogenic lines might be desirable for other purposes.

One final method termed GIRAFF (Genomic Identity Review by Annealing of Fractioned Fragments) was originally developed to detect polymorphisms between different strains of microorganisms [47]. Restriction digested DNA is size fractionated, heat denatured and reannealed to allow heteroduplex formation. The DNA is then digested with CEL I to cleave heteroduplex DNA, Southern blotted and hybridized to a labelled cosmid clone. Mutant alleles are detected as bands running ahead of the main size fraction. This method has yet to be applied to large genome eukaryotic organisms, although genome size, repetitive DNA content and multigene family structure is likely to limit its application.

Mutation spectrum analysis

The ATP has provided the most comprehensive set of data available upon which to assess the spectrum of mutations induced by EMS. Analysing 1890 mutations in 192 *Arabidopsis* genes, Elizabeth Greene and colleagues reported an unbiased analysis of mutation statistics [48]. In >99% of cases, EMS generated G to A and C to T transition mutations as expected from its established mode of action and induced changes in genes that were randomly distributed across the *Arabidopsis* genome. Mutations in DNA can be categorized as silent, mis-sense or truncation depending upon how they affect the encoded protein. Across all 1890 mutations, the observed distribution of silent, mis-sense and truncation events (45.0:50.1:4.9) was close to the expected distribution (44.4:48.3:5.3), as was the overall heterozygous:homozygous ratio of 2:1. The one clear exception was that truncation events were detected 3.6 times as often in heterozygotes, suggesting that such severe lesions are frequently deleterious to plants in the homozygous or haploid condition. The authors detected a local compositional bias in the nucleotides flanking individual mutations but no evidence for individual hotspots for EMS-induced mutations. They concluded that 'EMS is a nearly perfect mutagen for inducing G to A and C to T mutations' and that all classes of mutations can be recovered at the expected frequencies. All other reports, although less comprehensive, generally support and extend this conclusion.

Stacking the odds

Having concluded that induced mutants would be an appropriate biological resource to explore the function of a gene or to generate a novel allele for practical application, it makes sense to stack the odds in favour of identifying functional mutations before performing a screen, and then to prioritize identified mutants for further (functional) analysis. Thankfully, the Seattle groups of Steve Henikoff, Luca Comai and Elizabeth Greene have developed computational tools that simplify this process and are generally available via the ATP website (Table 1d). Prior to screening, the web-based utility CODDLE (Codons Optimized to Discover Deleterious LEsions) largely automates the identification of regions that are likely to be important for protein function (Table 1m) and suggests suitable primers (via primer3) once a defined region has been selected. Effective primer design is imperative, although sometimes challenging in large genome and/or polyploid species, and should be tested empirically. After induced alleles have been characterized, two further web-based utilities, SIFT (Sorting Intolerant from Tolerant) [49] (Table 1n) and PARSESNP (Project Aligned Related Sequences and Evaluate SNPs) [50] (Table 1o) predict the effect that a particular amino acid substitution might have on protein function (amongst other things) and are effective filters for prioritizing mutant alleles for further study. The same utilities are also of value when analysing natural biological diversity.

To stack the odds even more, Jillian Perry and colleagues at the John Innes Centre (Norwich, UK) assembled phenotypically selected subpopulations of plants that exhibit a range of morphological, symbiotic and metabolic defects. Using nodulation after inoculation with *Mesorhizobium loti* as an example they pre-selected a subpopulation of 288 non-nodulating mutants and screened for mutations in *SYMRK*, a gene that confers a non-nodulating phenotype when defective [24]. Using pools of three plants, they identified 15 homozygous mis-sense mutants (six independent alleles) and a splice acceptor site mutation in this enriched population. They argue that – depending on the purpose – phenotypic pre-screening for a trait of interest before TILLING might be a useful enrichment strategy. Phenotypic enrichment can be envisaged for a wide range of phenotypes (e.g. developmental abnormalities [51] and disease resistance [52]); one general and topical example is when addressing the function of putative orthologues of functionally characterized genes from model species in non-model plants. Of course, a major disadvantage of this approach is that the minor or weak alleles that are important for investigating structure and/or function studies might be missed completely.

Linking phenotypes to underlying genes

Decades of research on natural and induced genetic variation has assembled extensive collections of well-characterized genetic material that are a fundamental resource for understanding plant development and 'how plants work' [15] (Table 1o, p and q). Forward genetics has been particularly successful in linking genes to phenotypes (whether induced or naturally occurring). With the rapid advances being made in plant genome sequencing and functional genomics outside of model organisms, the value of mutant collections is understandably being reappraised. For example, Dani Zamir's group [3] has

developed a comprehensive collection of 13 000 M2 families of tomato. They have described, photographed, catalogued and made available over the web descriptions of 3417 mutations, including >1000 that have never been described before, under the banner 'The genes that make tomatoes' (Table 1r). The tomato genome is currently being sequenced by an international consortium. Once it is complete, the value of these resources for understanding tomato biology will be paramount. Similar web-based resources are being developed for other plants [e.g. *Lotus* (Table 1s) and barley (Table 1q and t)]. Given the prominent role that forward genetics has already played in plant science in general, as genomics tools and resources in crop plants improve, it seems likely that crop plant researchers will increasingly embrace forward genetics approaches to identify genes controlling biologically important phenotypes.

Perspective

The marriage between mutagenesis and sensitive mutation detection assays in pooled DNA samples has invigorated plant functional genomics. The TILLING methodology in particular is having a direct impact on many plant species and on other areas of biology including *Drosophila* [53], *C. elegans*, (Table 1u), zebrafish [54,55] and rat [56]. Furthermore, the concept of Eco-TILLING, an inexpensive approach for discovering natural polymorphisms in the sequences of genes from diverse genetic materials has been proposed [57] and is being used in several species. Within the plant research community there is considerable interest in using induced mutations as a bridge between the wealth of fundamental knowledge on genes and systems in models such as *Arabidopsis* and making discoveries and applications in crops. Furthermore, the plant research community is waking up to the realization that it is now feasible and realistic to discover stable and heritable genetic lesions in effectively any gene of interest and to study the impact of these lesions on the process under study directly in the crop. Some of this induced variation will have significant commercial or agronomic potential, as has already been demonstrated by modifying the properties of cereal starches [23] and by generating *de novo* resistance to an important agricultural pathogen [58]. These examples emphasize how this rapid, effective and uncontroversial route to the generation of biological diversity promises much for both basic and applied research. As a result, it is imperative that ongoing projects in numerous crop species survive past the pilot phase and mature into resource-based services that are widely accessible and thus serve to enrich the entire research community (like the ATP, Lotus and Barley Distilling projects). Only then will we reap the true value of induced biological diversity.

Acknowledgements

We thank Brian Forster and Jerry Franckowiac for their contribution to the phenotypic characterization of our 'barley mutation grid' at the Scottish Crop Research Institute and for valuable comments on this manuscript. R.W., D.J.L., N.M. and D.C. are funded by the Biotechnology and Biological Sciences Research Council and the Scottish Executive Environment and Rural Affairs Department.

References

1 Siddiqui, B.A. and Khan, S., eds (1999) *Breeding in Crop Plants: Mutations and In Vitro Mutation Breeding*, Kalyani Publishers
2 Jander, G. *et al*. (2003) Ethylmethanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiol.* 131, 139–146
3 Menda, N. *et al*. (2004) *In silico* screening of a saturated mutation library of tomato. *Plant J.* 38, 861–872
4 Stepanova, A. *et al*. (2003) Genome wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657
5 An, S. *et al*. (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. *Plant Physiol.* 133, 2040–2047
6 Lawrence, C.J. *et al*. (2005) The maize genetics and genomics database. The community resource for access to diverse maize data. *Plant Physiol.* 138, 55–58
7 Jansen, G. *et al*. (1997) Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat. Genet.* 17, 119–121
8 Liu, L.X. *et al*. (1999) High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res.* 9, 859–867
9 Wang, G.L. *et al*. (2004) Isolation and characterization of rice mutants compromised in Xa21-mediated resistance to *X. oryzae* pv. *oryzae*. *Theor. Appl. Genet.* 108, 379–384
10 Li, X. *et al*. (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J.* 27, 235–242
11 Li, X. *et al*. (2002) Deleteagene: a fast neutron deletion mutagenesis-based gene knockout system for plants. *Comp. Funct. Genomics* 3, 158–160
12 Werner, J.D. *et al*. (2005) quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2460–2465
13 Hazen, S.P. *et al*. (2005) Rapid array mapping of circadian clock and developmental mutations in *Arabidopsis*. *Plant Physiol.* 138, 990–997
14 Ahloowalia, B.S. *et al*. (2004) Global impact of mutation derived varieties. *Euphytica* 135, 187–204
15 Koornneef, M. *et al*. (1982) EMS-induced and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat. Res.* 93, 109–123
16 Davis, M.P. *et al*. (1996) *Special Issue Barley Genetics Newsletter* 26, 1–533
17 Ashburner, M. (1990) *Drosophila: A Laboratory Handbook*, Cold Spring Harbor Laboratory Press
18 McCallum, C.M. *et al*. (2000) Targeted screening for induced mutations. *Nat. Biotechnol.* 18, 455–457
19 McCallum, C.M. *et al*. (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123, 439–442
20 Colbert, T. *et al*. (2001) High-throughput screening for induced point mutations. *Plant Physiol.* 126, 480–484
21 Till, B.J. *et al*. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.* 13, 524–530
22 Till, B. *et al*. (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol.* 4, 12
23 Slade, A.J. *et al*. (2005) A reverse genetic, non-transgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.* 23, 75–81
24 Perry, J.A. *et al*. (2004) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol.* 131, 866–871
25 Caldwell, D.G. *et al*. (2004) A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). *Plant J.* 40, 143–150
26 Henikoff, S. and Comai, L. (2003) Single-nucleotide mutations for plant functional genomics. *Annu. Rev. Plant Biol.* 54, 375–401
27 Henikoff, S. *et al*. (2004) TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol.* 135, 630–663
28 New strategies to improve grain legumes for food and feed (acronym GRAIN LEGUMES) EC Contract FP6-2002-FOOD-1-506223
29 Gilchrist, E.J. and Haughn, G.W. (2005) TILLING without a plough: a new method with applications for reverse genetics. *Curr. Opin. Plant Biol.* 8, 211–215
30 Borevitz, J.O. *et al*. (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* 13, 513–523

31 Rostocks, N. *et al*. (2005) Single-feature polymorphism discovery in the barley transcriptome. *Genome Biol.* 6, R54
32 Mitra, R.M. *et al*. (2004) A Ca^{2+}/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4701–4705
33 Maluszynski, M. *et al*. (2000) Officially released mutant varieties – the FAO/IAEA Database. *Mutat. Breed. Rev.* 12, 1–84
34 Stemple, D.L. (2004) TILLING – a high-throughput harvest for functional genomics. *Nat. Rev. Genet.* 5, 145–150
35 Emery, G. (1960) Biological effects of a chemical mutagen, Diepoxybutane, on tomato. *Science* 131, 1732–1733
36 Li, S.L. and Redei, G.P. (1969) Estimation of mutation rate in autogamous diploids. *Radiat. Bot.* 9, 125
37 Kleinhofs, A. *et al*. (1978) Induction and selection of specific gene mutations in *Hordeum* and *Pisum*. *Mutat. Res.* 51, 29–35
38 Prina, A.R. and Favret, E.A. (1983) Parabolic effect is sodium azide mutagenesis in barley. *Hereditas* 98, 89–94
39 Matthes, M. *et al*. (2001) Variation in oil palm (*Elaeis guineensis* Jacq.) tissue culture-derived regenerants revealed by AFLPs with methylation-sensitive enzymes. *Theor. Appl. Genet.* 102, 971–979
40 Vos, P. *et al*. (1995) AFLP – a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414
41 Oleykowski, C.A. *et al*. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res.* 26, 4597–4602
42 Kulinski, J. *et al*. (2000) CEL I enzymatic mutation detection assay. *Biotechniques* 29, 44–48
43 Yeung, A.T. *et al*. (2005) Enzymatic mutation detection technologies. *Biotechniques* 38, 749–758
44 Qiu, P. *et al*. (2004) Mutation detection using Surveyor™ nuclease. *Biotechniques* 36, 702–707
45 Yang, B. *et al*. (2000) Purification, cloning, and characterization of the CEL I nuclease. *Biochemistry* 39, 3533–3541
46 Zerr, T. and Henikoff, S. (2005) Automated band mapping in electrophoretic gel images using background information. *Nucleic Acids Res.* 33, 2806–2812
47 Sokurenko, E.V. *et al*. (2001) Detection of simple mutations and polymorphisms in large genomic regions. *Nucleic Acids Res.* 29, e111
48 Greene, E.A. *et al*. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* 164, 731–740
49 Ng, P.C. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814
50 Taylor, N.E. and Greene, E.A. (2003) PARSESNP: a tool for the analysis of nucleotide polymorphisms. *Nucleic Acids Res.* 31, 3808–3811
51 Parker, J.S. *et al*. (2000) Genetic interactions during root hair morphogenesis in *Arabidopsis*. *Plant Cell* 12, 1961–1974
52 Büschges, R. *et al*. (1997) The barley Mlo gene: a novel control element of plant pathogen resistance. *Cell* 88, 695–705
53 Winkler, S. *et al*. (2005) Target selected mutant screen by TILLING in Drosophila. *Genome Res.* 15, 718–723
54 Wienholds, E. *et al*. (2002) Target-selected inactivation of the zebrafish *rag1* gene. *Science* 297, 99–102
55 Wienholds, E. *et al*. (2003) Efficient target-selected mutagenesis in zebrafish. *Genome Res.* 13, 2700–2707
56 Smits, B.M.G. *et al*. (2004) Target-selected mutagenesis of the rat. *Genomics* 83, 332–334
57 Comai, L. *et al*. (2004) Efficient discovery of DNA polymorphisms in natural populations by ecotilling. *Plant J.* 37, 778–786
58 Kanyuka, K. *et al*. (2005) The recessive resistance to bymoviruses correlates with specific amino acid changes in the barley translation initiation factor 4E. *Mol. Plant Pathol.* 6, 449–458