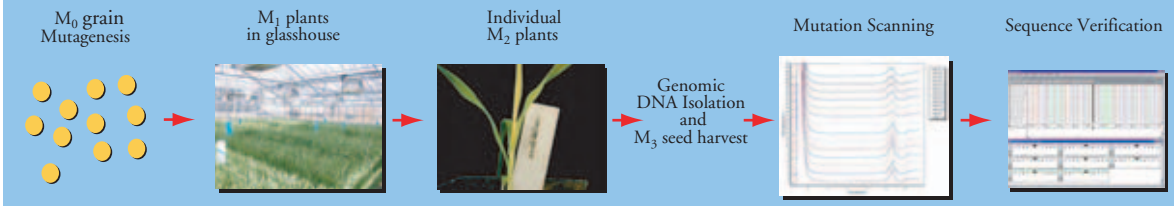
A physical / chemical mutation grid for barley functional genomics

D. Caldwell, N. McCallum, S. Mudie, P. Hedley, L. Ramsay, H. Liu, D. Marshall, R. Waugh

Progress in DNA sequencing has propelled a shift in genome analysis from structural to functional genomics (ie. the genome-driven systematic study of gene function). As a result, rapid methods for ascertaining the function of large numbers of genes are highly desired. With the development of sensitive methods for the detection of 'aberrant' DNA fragments in pooled samples, the use of chemical or physical mutagens to facilitate targeted gene inactivation has recently been shown to have significant potential in a reverse genetics program. Strategies deploying mutagens which induce random point or small deletion mutations coupled with PCR-based DNA mismatch screens therefore theoretically facilitate the detection of mutations in any specified target region. Some inherent attractions of this approach, over alternative biological ones such as insertional mutagenesis, include the ability to manipulate the mutagen and its dose (influencing mutation type and frequency) and to scale up or down easily. We are investigating the potential of chemical / physical mutagenesis to facilitate reverse genetics in barley. In a necessary parallel set of experiments, mutation detection protocols are being evaluated with the specific objective of being able to routinely identify mutant alleles in pooled DNA samples at various pooling depths. If we are successful in our objectives then the mutant populations that we develop will be available to all in the global research community to screen for mutations in their favourite gene. The process is outlined graphically in Figure 1.

Production and management of a large collection of mutagenised plants In pilot studies we have generated M_1 test populations for each of three mutagens (gamma-irradiation, DEB, EMS). Grain from the M_1 plants from four different dose rates of each mutagen (12 populations, 50 families from each = 600 families) has been carried forward to the M_2 generation (comprising c. 36,000 plants). Mutation frequency in the different M_2 populations has been calculated based on lethality, visible phenotypes and genome-wide mutagen-induced DNA sequence polymorphism (using AFLP). As expected, M_1 lethality increased with a corresponding increase in mutagen dose and ranged from <10% to over 90% kill. We observed little correlation between lethality in the M_1 and M_2 generations. Of the mutagens tested, EMS generated the highest frequency of visible mutations with up to 15% of the families tested showing obvious phenotypes (mainly chlorotic or albinos which tend to be seedling lethal). The AFLP analysis – while not quantitative – provided an overall 'guestimate' of the relative mutation frequency in the various populations at the level of mutations per base pair. Based on the appearance of new bands or disappearance of existing AFLP bands, frequencies of 1 mutation per 50,000 bp to 1 per 450,000 bp were estimated (i.e. 5,500 – 50,000 mutations per mutant genome) in the different populations. Correlation with the 'kill' and visible phenotype frequency was high. Taken together, the results indicated that in barley the overall approach outlined in Figure 1 will be an appropriate reverse



M_0 grain was mutagenised with EMS. The mutagenised M_0 grain gave rise to M_1 plants. The M_1 plants yielded M_2 grain. No more that two M_2 grain per one M_1 parent was taken forward, self polinated, and grown to maturity. Young leaf tissue was harvested for genomic DNA isolation and the M_3 grain was harvested. Individual M_2 genomic DNA was 1-Dimensionally pooled and stored in 96 well titer plates. Regions of interest will be amplified with gene specific primers and scanned for mutations. Individual M2 lines will be identified and mutations confirmed.

Figure 1 Overview of creating the EMS population.

genetics strategy. These results prompted the development of a large scale "EMS Mutant Population".

In this case, approximately 80,000 M_1 plants were grown in the glasshouses at SCRI in 2001 and tillers harvested. From these, 50,000 grains (no more than 2 grains from each M_1 plant) were selected and grown in the field. Leaf material was collected from each of the remaining 25,000 viable M_2 plants and DNA is being isolated from optimally sized pools using high throughput (HTP) protocols. As indicated above, because of the necessary population size, mutations in target sequences will require to be identified by the analysis of pooled samples. Individual mutants will then be identified from the de-convoluted pools.

Rapid and systematic identification of mutations in target sequences In parallel with the development of the populations, a critical aspect in implementing our reverse genetics approach is the choice of a mutation detection method which will allow us to screen PCR–amplified target gene sequences with diagnostic sensitivity, high specificity and low cost. It should preferably provide some information on the location of the mutation. Through a collaboration with Dr. Tony Yeung (Fox Chase Cancer Center), we have investigated the efficacy of enzymatic cleavage of heteroduplex DNAs using CEL I, a mismatch specific endonuclease. The CEL I system is a simple assay that requires PCR amplification of the target sequence, denaturation and annealing to allow formation of heteroduplexes between the wild type and the mutant allele, enzymatic mismatch cleavage, and analysis of the cleaved products by gel electrophoresis. In addition, we have investigated the use of denaturing HPLC which is advantageous over other mismatch detection systems as it requires no post amplification template modification, is not gel-based, and as a result is both inexpensive and HTP. Both systems work well. We are currently comparing and improving the detection methods in the framework of high-throughput procedures for plant functional genomics to identify the most robust (in terms of specificity and sensitivity) and cost-effective protocols for the targeting of candidate genes. This decision will largely be based on the number of samples which can be pooled while still allowing single base pair mutations to be detected at an optimal rate. At the moment, a pool depth of 1:24 alleles is routinely allowing us to identify single bp mutations by dHPLC.

Database development and the correlation between mutation data and phenotype The choice of target sequences and the type of mutations sought for further analysis will be dictated by the biological questions being asked by individual research groups who wish to screen the mutation grid. Those in turn will depend upon the availability of EST, cDNA and genomic sequences and is essentially case-specific. Embodied in our strategy is the recognised need that multiple steps in the process require informatics and Laboratory Information Management tools (LIMS) which monitor the tracking of M_2 plants, their DNA and their grains; the design of primers for PCR amplification of targeted regions, the interpretation and databasing of mutant alleles; and the integrated analysis of mutant alleles and phenotypic information. We are, therefore, in the process of developing a relational database to assist in the management of the populations and their progenies and to record, store and exploit all data generated within the project. Our objective is to have the whole system 'open for business' by the end of 2002.

Acknowledgements

This work was funded as part of the SEERAD/BBSRC jointly funded project "Cereal Community Resources for investigating gene function" as part of the IGF (Investigating Gene Function) Initiative.