



## **Genomics and its relevance to barley**

Elizabeth S. Dennis

CSIRO, Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia

**September 2001**

# **Introduction**

Genomics has led to a new paradigm for the study of biology. In genomics, the focus of research has shifted from studying individual genes to studying all the genes in an organism. In plants, genomics has been very dependent on the use of two model organisms, Arabidopsis as a general model and model for dicots and rice as model for the cereals.

There are different stages in the study of genomics. A landmark stage is obtaining the complete genome sequence of an organism. On the way to the genome sequence there is usually a physical map developed with as many DNA markers as possible. For example, in rice there are some 3,500 DNA markers (RFLP, AFLP or SSP markers) scattered over the genome. This physical map is integrated with the genetic map and is of great use for linking phenotype with genotype and QTLs. Concurrent with the production of a physical map, the whole genome DNA is broken up into small pieces of about 100 KB and each is cloned as a 'BAC' or bacterial artificial chromosome in bacteria. A library of bacteria each with one BAC containing one genome segment is produced. These BACS are ordered so that every region of the genome is covered by a BAC and each BAC is from a known position in the genome. BACS are then sequenced, leading to the complete genome sequence.

So far, in plants only the Arabidopsis genome has been sequenced completely. The sequence was published in Nature last December. The Arabidopsis sequencing has shown that there are about 120 million bases in the Arabidopsis genome. Computer programs predict about 22,000 different genes for the complete Arabidopsis complement. In fact, with more careful study we now know that a portion of the Arabidopsis genome is duplicated and there are approximately 17,000 unduplicated genes. So that to carry out all of the processes of a plant we need something of the order of 17,000 genes.

The rice genome sequencing is in progress. The rice genome is about four times larger than the Arabidopsis genome (but about 20 times smaller than barley). There are 427 million bases in the rice genome. So far about 25% of the rice genome has been

sequenced, mostly by scientists in Japan and the USA but Taiwan, China and France scientists are now making an impact. Indian and Korean scientists are starting. Australia is not participating. Two private companies, Monsanto and Syngenta (formerly Novartis), each have a private rice genome sequence. Monsanto will donate their sequence to the International Rice Genome Project but it cannot be accessed publicly until the sequence is completed and released.

So for both Arabidopsis and rice we have, or can anticipate having in the next year, the complete genome sequence. The rice genome sequence will be in the public domain after its release as is the Arabidopsis sequence. The question is how can we use these sequences for the good of the Australian agriculture industry?

Gene sequences themselves cannot be protected as intellectual property. There must be a use or function described in a patent. Where there is an opportunity for Australia is in identifying genes which have specific uses for our Agriculture and protecting them. In this way Australia may keep up with the multinationals.

How do you work out what the function of a gene is? There are three major ways. Firstly, you can use computer matching by comparing the DNA sequence of a gene against a data base of all the genes that have been sequenced in all organisms. This requires very sophisticated computer programs, many of which are in the public domain and powerful computer hardware. We can place approximately 50% of plant gene sequences into a general category of what each is doing, for example, a gene may be similar to genes with the biochemical function of a kinase without knowing exactly when and where it acts or what is its substrate.

The second approach is to ask where is the gene active? The basis of this approach being that if a gene is active in a particular tissue it may play a role in that tissue. For example, if a gene is active during flowering then it is probably has a role in flowering, if it is switched on during stress conditions it may play a role in response to that stress.

This method identifies genes active at a particular time or in a particular tissue. When a gene is active it makes RNA, this RNA can be converted to DNA and the sequence of the gene can be determined (EST sequencing) giving an idea of which genes are active in that tissue. EST sequencing has provided much useful information and is been the basis of cereal genome programs.

The new improved technology which moves a step further on from ESTs is microarrays. In microarrays we generate a microscope slide with one copy of each of the genes of the organism on the slide. The microscope slide can fit approximately 20,000 spots or genes. DNA from one gene is placed in one well in a microtitre dish. A robot takes a sample out of each of the microtitre wells and spots them on a microscope slide. Then we take messenger RNA from various plant tissues and we look to see where the gene is active. We can take RNA from flowers or from leaves and we can label each with a different colour dye, eg. flower RNA labelled red and leaf RNA labelled green. By using these labelled RNAs and reacting them with the DNA on the microscope slide, we can determine which genes are active in flowers and which genes are active in leaves by the ratio of the two colours for each gene. In

this way we can identify genes which are only expressed in flowers or during stress or any other condition.

The third way of determining gene function is to use what geneticists have always done, to make a mutation in a gene and examine whether it has any effect on the plant. If a gene is knocked out or altered in its activity then this should have an effect on the plant, the type of effect should give a clue as to the gene function.

The best way is to use a DNA tag of known sequence to make the mutation. This enables the tag to be used to determine which gene is mutated. A genome is mutagenised by randomly inserting pieces of DNA in many genes. Each of these mutated lines is grown up separately and a library of lines is obtained each with a piece of DNA in one gene. Then the tag is used to isolate the flanking sequence of the gene which identifies which gene the DNA was inserted into. The phenotype of the plant is examined. We then have a link between the DNA sequence of the mutated gene and the phenotype. In this way gene sequence can be linked to function.

These are the basic areas of genomics. The ultimate aim is to give a complete description of what is occurring in the cell. The ten-year goal of the Arabidopsis Genome Project is to understand every molecular interaction in every cell throughout a plant life cycle. In essence, to understand the function of every gene by 2010. Many other types of genomic techniques will need to be used to accomplish this goal. For example, all of the proteins that are made in a particular tissue or under particular conditions can be identified (proteomics). All of the metabolites that alter under specific conditions (metabolomics) can be identified.

## Applications to Barley

The main genome resource in barley so far is ESTs or Expressed Sequence Tags. Many ESTs have been generated from barley and these are available in the public sector. Also available are BACS for barley. The barley genome has been cloned into BACS and now those BACS are in the process of being ordered to make a physical map. There are also barley DNA markers available. Genome resources are listed at GrainGene <http://wheat.pw.usda.gov>.

Because barley is a diploid cereal it will also be the model for wheat. The main resource for barley genomics comes from the rice and Arabidopsis genome efforts. The cost of sequencing the genome of barley and/or wheat is very high (probably \$200 million at present) so it will not happen in the next few years.

Because of the conservation of genes and gene response across the plant kingdom, Arabidopsis and rice are currently important. Rice can be used as a model cereal and be used for barley because rice and barley show a good deal of synteny. That means the genes are in the same order on chromosome segments so that the order of markers and genes will be conserved over long distances. This is very useful in identifying the rice homologous of genes responsible for QTLs in barley. The rice gene sequences are similar to those of barley so that if a rice gene is isolated then it is relatively straight forward to isolate the same gene from barley. In general, gene sequences are

approximately 80% identical between rice and barley. In this way barley genomics can progress using rice as the model.

It will be important to use barley to determine the function of specific genes in barley which have been identified as candidate genes in another way. This can be done using the barley transformation systems. Candidate genes can be introduced into barley to determine the effect of increasing or decreasing the level of gene product. In this way findings from plant genomics can be utilized for improved barley production.

---