Turk J Agric For
29 (2005) 97-103

The Impact of Genomics and Genetics on Wheat Quality Improvement

John SNAPE[1,*], Lesley FISH[1], David LEADER[2], Robert BRADBURNE[1], Adrian TURNER[1]

[1]John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UJ - UK

[2] Syngenta, Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EY - UK

Received: 04.12.2003

Abstract: This paper discusses how genetic and genomic tools may be used to understand the genetical and molecular control of cereal quality and to develop tools for its manipulation by conventional and genetic engineering approaches. Comprehensive genetic maps are a first step in the genetical and molecular analysis of traits and these now allow the dissection of the genetical control of complex traits. As an example, data on the genetical control of grain protein content in UK winter wheats is presented. From a series of field trials, several potentially useful new QTLs were identified on chromosomes 2B, 5D, 6A, 6B, 7A which can be targets for marker-assisted selection. Expressed sequence tags (ESTs) and cDNA microarrays identify candidate genes involved in differences in quality and a means for understanding the molecular control of endosperm development. This paper gives data on the use of micro-arrays for identifying differential expression of genes during early endosperm development. Genetic engineering provides a complementary tool to conventional breeding for cereal quality modification, and the prospects and problems of applying the technology for wheat end-use quality improvement are discussed.

Key Words: genetics, genomics, wheat, quality, transcriptomics

Introduction

In this millennium, the production of new varieties by plant breeders, adapted for novel end-uses, will be achieved either by 'conventional' cross breeding, or by genetic modification, using isolated, cloned homologous or heterologous genes. Both methods will rely on a greater understanding of plant genetics, physiology, and metabolism, to identify the genes that mediate plant performance. This will allow a more targeted manipulation of phenotype and will rely heavily on the development of advanced genetic and genomics technologies. Emerging tools will enable the discovery and manipulation of new desirable genetic variation, which can be combined into the next generations of crop varieties, designed to meet specific challenges of the new economic and environmental constraints on farming, and the specific requirements of end-users.

Simplistically, the wheat grain can be divided into three constituents, the germ, bran and endosperm. All three are complex structures made up of different constituents. Eighty percent of the endosperm is made up of starch, and most of the remainder is protein. Generally, the endosperm composition that has received most attention with respect to the genetical analysis and manipulation of quality traits, since proportionally, this is, by far, the greatest component. However, the composition of the embryo and the grain coats, the pericarp and aleurone, are complex and contain a range of constituents whose biochemistry is partially known, but whose genetic control is very poorly understood. This paper describes how advances in genetics and genomics are contributing to our understanding of the genetic and molecular control of cereal quality and methods for its manipulation using conventional breeding and genetic engineering.

From phenotype to genotype: traits to genes

The first step towards understanding the whole genetics of individual crop plants is to develop comprehensive genetic maps. Good genetic maps, based on molecular marker technologies are now available for all major cereal species, including wheat. Presently, the

* Correspondence to: john.snape@bbsrc.ac.uk

major use of genetic maps is to locate genes of interest so that the maps can be fully annotated with the locations of genes, be it for quality, agronomic performance, disease resistance, adaptability, or any other trait, so that they can be manipulated in a directed manner by marker-assisted selection. Much of the variation for important quality traits in wheat is quantitative in nature and controlled by many genes of small effect acting together, so called QTL. QTL analysis in crop species with complex genomes is an important tool which allows the location of multiple loci such as those involved in quality differences. This analysis is complicated in wheat by the complexity of its polyploid genome, with the three genomes interacting in the regulation of one trait. In addition, low levels of polymorphism in molecular markers, especially in the "D" genome of wheat, make it very difficult to construct complete genetic maps. To produce complete maps, researchers have in the past resorted to studying very wide crosses such as that of the ITMI population to increase the chances of finding polymorphic markers. However, to study quality traits, crosses need to be made between much more closely related varieties to ensure the results are relevant to the modern market.

An example of the application of genetic maps for new gene discovery for end-use quality can be illustrated by our work to identify novel loci and alleles regulating grain protein content in wheat, in order to find genes which could be used to boost protein levels in UK varieties without causing a subsequent loss in yield. In a major study, two parental lines, Avalon and Hobbit Sib, were studied, which represent parentages of UK varieties with hard textured grain, high protein and good bread-making quality (BMQ); and soft, low protein, suitable for biscuit making, respectively. Additionally, these varieties differ only in one high-molecular- weight (HMW) glutenin subunit (Hobbit Sib has a null allele at the *Glu-A1* locus and Avalon produces the subunit 1 band). This suggests that they can offer insight into other important factors involved in BMQ other than the major storage proteins. Avalon has also been shown consistently to yield better than would be expected for its high protein content, making it an ideal candidate for investigating yield independent protein alleles. A population of 97 recombinant inbred lines (RILs) was produced by single seed descent from the cross between the parents, being at F_7 for this study. The results from two years of field trials were gathered, analysed, and collated. In 2000, the population was planted out in spaced plant field trials consisting of one-metre row plots of 11 plants of each line, hand-dibbed, in a triplicated block experiment. In 2001, 1m x 6m drilled plots of each line were sown out in a triplicated experiment, alongside dibbed rows of all lines. Yield components were measured for all lines (spikelet number, seed weight, grain number per ear, tiller number) in the dibbed rows, and in addition, a yield assessment was made from the drilled plots in 2001. Following harvest, grain protein content was measured using Near Infra Red Reflectance spectroscopy (NIR) on a Bran and Luebbe Infra-alyser 2000. The protein measurements were calibrated according to calibration samples analysed by the Dumas method.

For genetic mapping simple sequence repeat (SSR) markers were used. DNA was extracted from the leaves of seedlings of individual lines as described in Magrath et al. (1994). SSRs were analysed as described in Roder et al. (1998), being run on 5% polyacrylamide gels and visualised by silver staining (Sourdille et al., 1998). Five different groups of microsatellites were used "gwm" (Roder et al., 1998), "gdm" (Pestsova et al., 2000), specifically designed to map to the D genome, "wmc" (Wheat Microsatellite Consortium), psr (Mike Gale, John Innes Centre, UK) and barc (http://www.scabusa.org). The genetic map was developed using JoinMap (Stam & Van Ooijen, 1995). and QTL analysis on the phenotypic data was carried out using QTL Café (http://web.bham.ac.uk/g.g.seaton/) and MapQTL software (Van Ooijen & Maliepaard, 1996).

In the RIL studies, single marker ANOVA detected a total of 35 markers across 13 chromosomes associated with significant ($P < 0.05$) differences in grain protein content. Of these, 12 markers showed consistent differences in both years of the study (Table 1). Increased grain protein content was consistently associated with markers on chromosomes 1A, 2B, 2D, 3B, 5BS/7BS and 6B with Avalon contributing the increasing allele, and markers on chromosomes 3A, 6A and 7A with Hobbit Sib contributing the increasing allele. QTL analysis by interval mapping of grain protein content showed significant QTL on chromosomes 2B and 6B (Avalon contributing increasing effect) and chromosomes 6A and 7A (Hobbit Sib contributing increasing effect). The QTL on chromosome 2B (Figure 1) was located near *Xgwm644* and explained about 15% of the variation, and had significant Lod scores in both years. The QTL on

Table 1. Marker means analysis of grain protein concentration on the Avalon x Hobbit Sib RIL population over two years. Only significant ($P < 0.05$) consistent year-on year differences are presented. Where the additive effect is positive, Avalon contributes the increasing allele; where the additive effect is negative, Hobbit Sib contributes the increasing allele.

Year	Chromosome	Marker	Position (cM)	Additive effect (%)	P value
2001	1A	*Xbarc83*	87.9	0.36	0.014
2002	1A	*Xbarc83*	87.9	0.3	0.0086
2001	2B	*Xgwm410*	10.1	0.28	0.0066
2002	2B	*Xgwm410*	10.1	0.21	0.0062
2001	2B	*Xgwm644*	10.9	0.28	0.006
2002	2B	*Xgwm644*	10.9	0.21	0.0069
2001	2B	*Xwmc179.1*	14.7	0.29	0.006
2002	2B	*Xwmc179.1*	14.7	0.21	0.0063
2001	2D	*Xwmc441.3*	124.1	0.29	0.0045
2002	2D	*Xwmc441.3*	124.1	0.17	0.0251
2001	3A	*Xgwm32*	62.8	-0.22	0.0463
2002	3A	*Xgwm32*	62.8	-0.18	0.0216
2001	3B	*Xbarc229*	141.3	0.25	0.0141
2002	3B	*Xbarc229*	141.3	0.17	0.0288
2001	5BS/7BS	*Xgwm537*	90.1	0.27	0.014
2002	5BS/7BS	*Xgwm537*	90.1	0.22	0.0062
2001	6A	*Xgwm334*	140.1	-0.29	0.0056
2002	6A	*Xgwm334*	140.1	-0.25	0.0022
2001	6B	*Xbarc24*	89.6	0.23	0.0452
2002	6B	*Xbarc24*	89.6	0.23	0.0076
2001	7A	*Xwmc9*	114.3	-0.31	0.0029
2002	7A	*Xwmc9*	114.3	-0.2	0.012
2001	7A	*Xbarc108*	115.2	-0.44	0.00007
2002	7A	*Xbarc108*	115.2	-0.29	0.0037

chromosome 6A was located near *Xgwm334* and explained around 13% of the variation, and had a significant Lod score in 2002 only. The QTL on chromosome 6B was located near *Xbarc24* and explained around 18% of the variation, and had significant Lod scores in both years. The QTL on chromosome 7A was located near *Xbarc108* and explained between 8.8 and 10.3% of variation across the chromosome, and had significant Lod scores in both years. Additionally, a separate analysis using recombinant substitution lines identified a QTL associated with the *Ha* gene for grain hardness on the short arm of chromosome 5D, with Avalon contributing the allele for increased protein.

This study is the first to provide a complete analysis of grain protein content in UK wheat varieties, and it has shown that it is possible to dissect the genetics of a difficult-to-measure and highly-environmentally sensitive character using modern methods of genetic analysis. It has shown that the genetic control is complex and there are no underlying major genes as in the case of grain texture where most of the variation is controlled by the major gene, *Ha*, found on the short arm of chromosome 5D. Significant QTL were identified on chromosomes 5D, 2B and 6B (Avalon contributing increasing effect) and chromosomes 6A and 7A (Hobbit Sib contributing increasing effect) in data from both years. These effects were also relatively large, and, interestingly were dispersed between the parents. So, although Avalon is generally regarded as a high protein wheat, it carries alleles at certain loci for reduced grain protein relative to their homologues from Hobbit sib, and vice versa. Thus, in this cross, genes for higher levels of grain protein are dispersed between the parents and transgressive segregation for higher protein containing lines than Avalon is possible from this cross. Diagnostic markers for particular alleles could be sought/designed to enable this to be a tool by plant breeders for protein content selection.

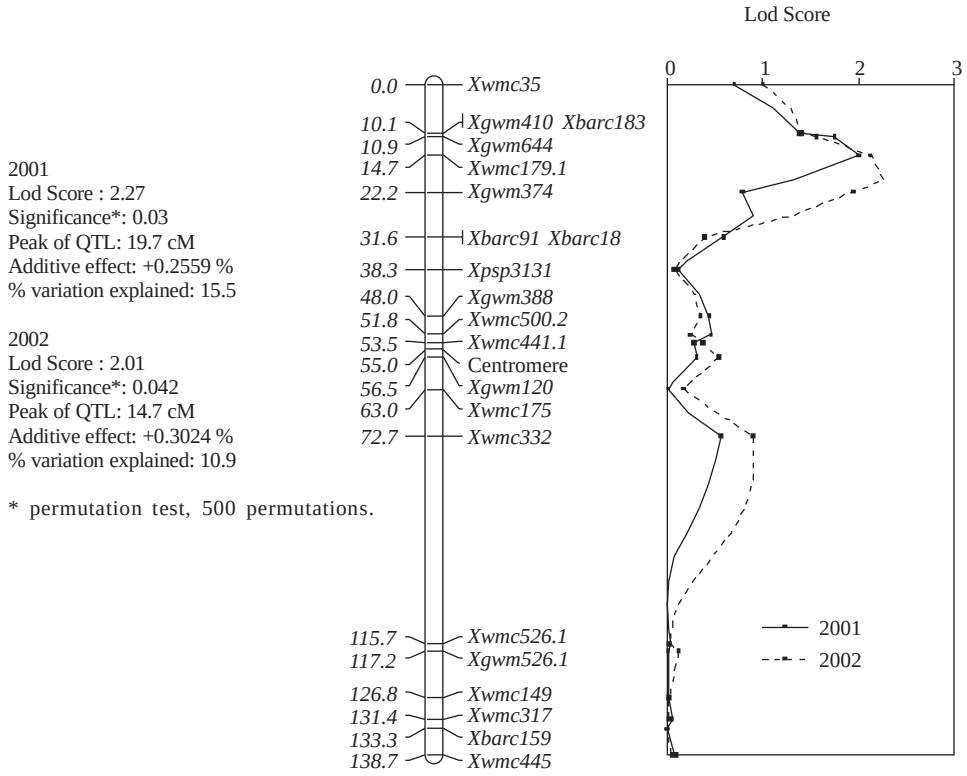

Figure 1. Interval mapping of chromosome 2B for grain protein content. One significant QTL detected (both years).

Functional genomics: from genes to traits

Following on from mapping, the new science of genomics is enabling the understanding of the relationship between gene structure and function. This is being addressed, firstly, by sequencing whole genomes, for example, Arabidopsis and rice, and secondly by the accumulation of libraries of 'Expressed Sequence Tags' - ESTs. The first crop species to be completely sequenced is rice, and the release of the first draft sequences by groups in China and Syngenta in 2002 suggests that cereals have between 40,00-50,000 genes. Relating these genes to homologues in wheat and barley will be a major goal for cereal geneticists. The full genomic sequencing of wheat, however, is presently not a realistic strategy in most cases because of the high amount of repetitive DNA (for example, 80% of the wheat genome is repetitive DNA). Thus, the alternative ways of finding genes and gene sequence information in crop plants is either to clone individual genes of interest, or to directly isolate the sequence of genes from their expression, where the sequence information in the messenger RNA can be 'captured' by making ESTs by reverse transcription from the RNA. The latter is easier, and ESTs can provide information on the structure and function of genes, and through bioinformatics, can be identified by homology with other species, be it plant, animal or bacterial. Also, isolating RNA from any particular specific tissue, for example, the endosperm, gives a description of what genes are expressed in that tissue. If RNA is isolated from many different tissue types, at several different times of the life cycle, then it is hoped, eventually, that all expressed genes can be captured by the RNA that they produce. However, as in whole genome sequencing, the challenge is then to ascribe function to these genes.

If we are to understand 'plant circuitry' and how to modify seeds for quality attributes, we need to know where and when individual genes and groups of genes are expressed. Another new genomic tool to do this, used in conjunction with ESTs, is DNA microarrays, also known as DNA chips. This is where glass slides each containing the individual signatures of thousands of genes, either identified from sequencing genomic or copy DNA, are produced. Challenging these arrays with fluorescently labelled RNA from any tissue or growth stage will indicate

which genes are expressed in that particular tissue, and thus give hints on which to manipulate. By following expression under different growth conditions, the spatial and temporal expression of groups of genes can be worked out, thereby indicating the ongoing metabolic processes and their control. Much of such work is being focused on the developing grain. Libraries are being produced from RNA extracted at different stages of development following anthesis, and expression patterns identified to discover the key metabolic enzymes involved in starch, lipid and protein deposition, and the cell cycle. Figure 2, for example, shows differences in gene expression for genes expressed early in endosperm development in studies carried out by Syngenta in a joint JIC-Syngenta collaboration looking at the control of grain development in wheat. This analysis has identified key genes in the pathways of grain development which can be targets for future genetic manipulation to increase grain size and hence yield.

Genetic modification and quality improvement

Genetic modification allows the introduction of isolated individual genes from any biological source, and offers a huge variety of opportunities for the improvement of quality in cereals or even for the production of entirely novel crops for industrial use. Technologies for the genetic modification of most, if not all, cereal crops are now available, and their application in agriculture has more to do with the economics and politics of Genetically Modified Organisms (GMO's) than biology or technology.

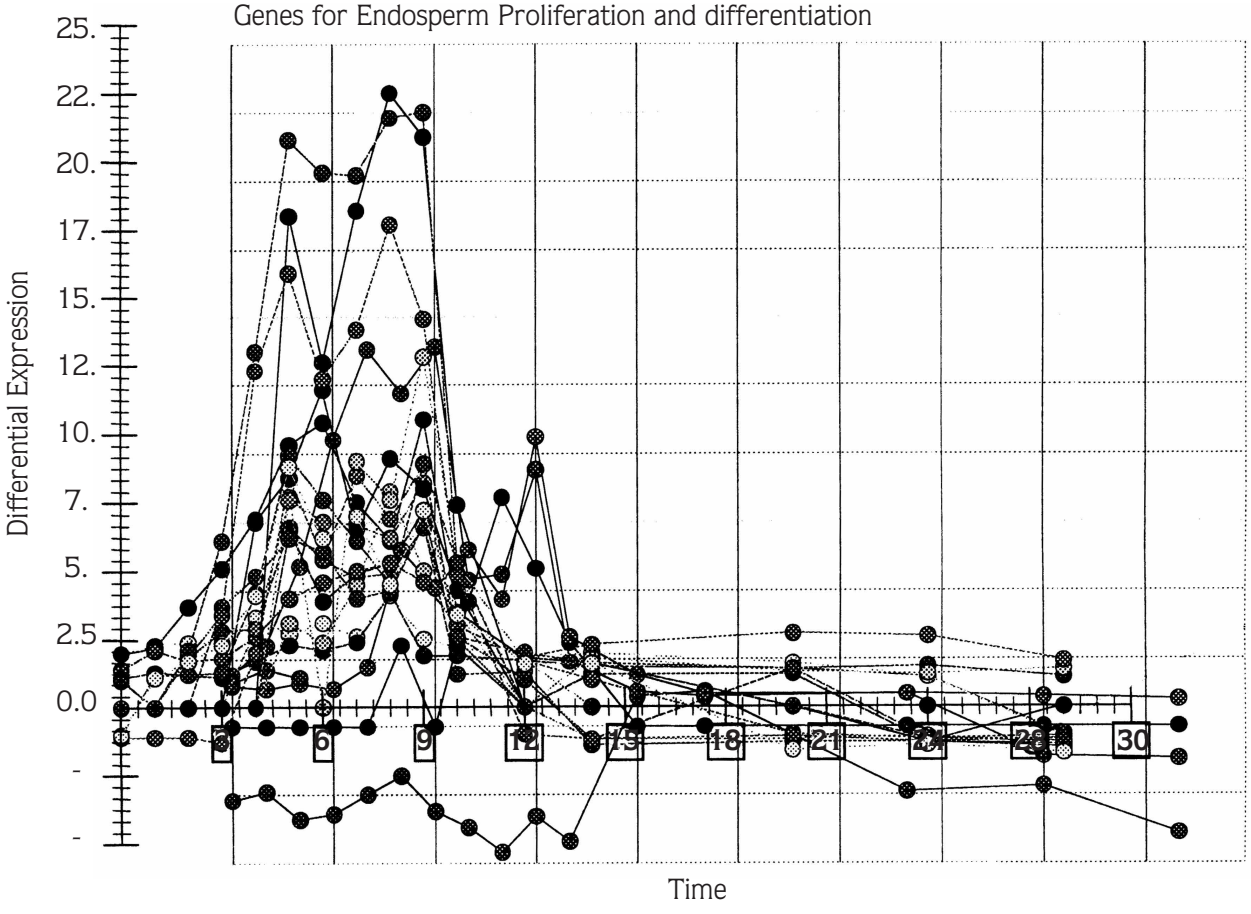

Figure 2. Differential expression of genes involved in endosperm development in wheat. Certain genes are up regulated between 4 and 12 days in development whilst other remain the same.

Despite the often negative public attitude to GMO's, particularly in Europe, 2001 saw a Worldwide increase in the growth of GM crops to 52.6 million hectares, an increase of 19% over 2000 (James, 2002). Globally the principal crops were soybean, maize, cotton and oil-seed rape, with an overwhelming predominance of herbicide (glyphosate, glufosinate) resistance and insect (Bt) resistance. The expectation is that this trend will continue, and it can be foreseen that despite current problems, GM wheat and barley crops will eventually be grown in the UK and other countries, although there are barriers to be overcome in terms of public acceptability. The main issues are first, technology limitations, particularly the need to remove unwanted DNA of herbicide or antibiotic resistance genes; secondly, concerns about food safety, namely the unpredictability of transgene expression; and, thirdly, environmental, the problem of transgene transfer into non-GMO crops. Some of these limitations are technological, and solutions, for example, 'clean-gene' technology are emerging where marker genes can be separated from the target gene leaving transgenic lines only containing the target genes. We have already demonstrated this in rice using an *Agrobacterium* system and are presently transferring the technology to wheat.

Other limitations require more information on the consequences of transgene insertion on target and non-target gene expression. This information is currently being produced in a number of studies. However, even if these limitations and problems are overcome, consumer choice has to be favourable, and industry has to produce more consumer orientated transgenic crops than those available at present. It would be better to see the introduction of traits which have more advantages for the consumer in terms of quality, environmental or health benefits, rather than just agronomic advantages to the producers. Some examples of possible targets in crop plants for quality traits in this respect are shown in Table 2.

Conclusions

In this millennium, genetic and genomics research on our cereal species have the potential to define the total extent of the genetic variation for end-use traits, whether exhibiting simple or complex inheritance. This will allow plant breeders and geneticists to synthesis novel gene combinations by marker-assisted-breeding, using high-throughput molecular marker systems, to produce 'designer' varieties. Also, there is the capacity to modify metabolic pathways by genetic engineering, leading to novel products and processes for industry. As well as leading to economic prosperity, this research can also make an important contribution to World food security through the development of varieties much more resistant to pest and diseases both in major crops such as wheat, and also in 'orphan' crops of the less developed world through comparative approaches. Clearly, we have only just started to see the fruits of this genomics revolution leading, hopefully, to the evolution of a new Green Revolution.

Table 2. Examples of targets for the genetic modification of wheat for quality attributes.

Character	Genes
Nutritional quality for animal feed	Phyt gene encoding phytase to increase phosphorous availability for animal feed
	DapA gene encoding di-hydro-di-picolinate synthase (DHDPS) to increase lysine content
Bread-making quality	High and low molecular-weight glutenin subunits
Industrial applications	Modified starches

References

James, C. 2002. Global review of commercialized transgenic crops: 2001. ISAAA Briefs 24.

Magrath, R., F. Bano, M. Morgner, I. Parkin, A. Sharpe, C. Lister, C. Dean, J. Turner, D. Lydiate and R. Mithen. 1994. Genetics of aliphatic glucosinolates .1. Side-chain elongation in *Brassica napus and Arabidopsis thaliana*. Heredity 72: 290-299.

Pestsova E., M. W. Ganal and M. Röder. 2000. Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43: 689-697.

Roder M.S., V. Korzun, K. Wendehake, J. Plaschke, M.H. Tixier, P. Leroy and M.W. Ganal. 1998. A microsatellite map of wheat. Genetics 149: 2007-2023.

Sourdille P., M.R. Perretant, G. Charmet, P. Leroy, M.F. Gautier, P. Joudrier, J.C. Nelson, M.E. Sorrells and M. Bernard. 1996. Linkage between RFLP markers and genes affecting kernel hardness in wheat Theor. Appl. Genet. 93: 580-586.

Stam, P. and J. Van Ooijen. 1995. JoinMap ™ version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen, Netherlands.

Van Ooijen, J., and C. Maliepaard. 1996. MapQTL ™ Version 3.0: software for the calculation of QTL positions on genetic maps. CPRO-DLO, Wageningen, Netherlands.