ELSEVIER

Journal of Cereal Science 41 (2005) 149–163

Journal of
CEREAL
SCIENCE

www.elsevier.com/locate/jnlabr/yjcrs

Review

Transcriptional analysis and functional genomics in wheat

David J. Leader

Crop Performance and Improvement Division, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

Received 22 July 2004; revised 25 October 2004; accepted 26 October 2004

Abstract

In the post-genomic era the development of tools such as transcription microarrays allows the investigation of the molecular basis and regulation of plant growth, development and physiology in ever increasing detail. Systems such as the 'model' plant species *Arabidopsis* offer the opportunity to develop generalised models for complex genetic regulatory networks. However, researchers will need to apply tools to both monitor and alter gene expression in crop plants such as wheat to produce sophisticated and predictive models of development and environmental interactions. Such models will allow crops to be developed which are increasingly adapted to the needs of agriculture in the 21st century. In this paper some of the tools that have been developed both in wheat and other plants for high throughput transcriptional analysis are described. The key strengths and weaknesses of application of these technologies to different aspects of wheat research are discussed. Potential approaches for the high throughput modification of gene expression for functional genomics in wheat are discussed and the specific challenges posed by wheat and what can be learnt from model species considered.

Keywords: Wheat; Functional genomics; Microarray; Expression

1. Introduction

The past decade has seen an unprecedented increase in our understanding of the genomes of a variety of different organisms. Among these, the publication of the first plant genome sequence from *Arabidopsis thaliana* (The Arabidopsis Genome Initiative (AGI), 2000) and more recently the publication of the genomic sequences of two rice subspecies (Goff et al., 2002; Yu, 2002) represent major landmarks in plant biology. As well as studies of model systems, various public and industrial EST sequencing efforts have provided a wealth of sequence data on ESTs (expressed sequence tags corresponding to mRNAs) from most of the world's major crops. For example, the sequences of well over half a million ESTs from wheat are now freely available in public databases. However, few of the predicted genes from the analysis of the rice genome or the ESTs from other crop plants have any assigned functions while many other sequences can only be assigned to broad functional categories within large groups of related genes. Thus, for the overwhelming majority of these sequences there is little information, which can be used to place the function of the gene in a precise developmental, cellular or metabolic context which would be required to establish its use in trait development.

Functional genomics promises to solve this problem and to define the role that each gene plays in the life of the complete organism. This is likely to be a stepwise process with a variety of approaches and genomics tools being used to develop and refine models for the function of each gene.

In many instances, for those working with crops, the functional characterisation of a gene will begin by the analysis of the orthologous gene in either *Arabidopsis* or rice, since many of the most powerful tools for functional analysis have been developed first in these species, taking advantage of the availability of complete genome sequences. However, it is clear that strategies will need to

Abbreviations: AFLP, amplified fragment-length polymorphism; dpa, days post-anthesis; EMS, ethylmethane sulphonate; EST, expressed sequence tags; FACS, fluorescence activated cell sorting; GEM, gene expression microarray; GUS, β-glucuronidase gene; MPSS, massively parallel signature sequencing; PCR, polymerase chain reaction; RNAi, RNA interference; VIGS, virus induced gene silencing; T-DNA, transposon DNA; TAIL-PCR, thermal asymmetric interlaced polymerase chain reaction; tpm, transcripts per million.

E-mail address: david.leader@bbsrc.ac.uk.

0733-5210/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jcs.2004.10.006

be developed for genes, which are unique to each crop plant or for situations, where a clear orthologue cannot be identified. Crop-specific tools also need to be developed to understand how the modification of gene function during evolution and selection contributes to the unique characteristics of different species and cultivars.

Functional genomics offers new opportunities for the rapid identification of the genes underlying traits of agronomic importance and provides novel strategies for the manipulation of these traits. Biologists look to functional genomics to deliver increasingly complex and dynamic molecular descriptions of biological systems which will facilitate the accurate prediction of how these systems can be manipulated to improve performance (Duyk, 2002).

This paper summarises the technologies that have been developed for transcriptional analysis in plants and other organisms and gives examples of how these can be used to identify candidate genes associated with key traits and processes in wheat. The latter are based on studies carried out as part of my own research programme at the Syngenta Wheat Improvement Centre in Norwich, UK, using custom arrays prepared specifically for the project. Similar arrays are also freely available in the public sector including the UK transcriptomic resource described by Wilson et al. (2004) (see also http://www.cerealsdb.uk.net).

Transcriptome profiling can be used to identify genes with putative roles in plant development and metabolism but the functions of these need to be confirmed using other approaches. The review therefore also briefly discusses how this can be achieved, using transgenic and non-transgenic approaches for gene disruption. These also require the development of efficient transformation systems for wheat, which are discussed in detail elsewhere in this issue (Jones, in press).

2. Genome scale analysis of gene expression

One of the most easily determined indicators of the function of a gene is its spatial and temporal expression pattern. A widely held assumption is that genes that are involved in the same process will be co-expressed and consequently a number of different approaches have been used to examine gene expression at the genomic scale in plants. These methodologies broadly fall into two categories; open systems which do not rely on previous sequence knowledge and closed systems such as microarrays in which large predefined sets of genes are analysed simultaneously.

2.1. Closed systems

Two main classes of microarrays have been developed for global transcriptional analysis; spotted arrays and arrays which use photolithograpy to synthesize oligonuclotide probes in situ (Affymetrix arrays) (Fodor et al., 1993; Lipshutz et al., 1999). Spotted arrays use robotics to deposit small volumes of a cDNA or oligonucleotide onto a glass or other support. They are highly flexible and can be produced by individual groups or small consortia using off the shelf equipment. These arrays can be made on a variety of different scales ranging from the near genomic scale arrays containing many thousands of individual elements to more limited 'boutique' arrays containing a subset of genes of particular interest to the scientist (Duggan et al., 1999). Importantly, cDNA arrays do not require a full genome sequence and they can be generated relatively easily from pre-sequenced or anonymous cDNA clones (reviewed by Zhao and Bruce, 2003). With appropriate quality control parameters, the data generated by cDNA microarrays approaches the sensitivity and reproducibility of Affymetrix arrays which are potentially less flexible and more costly (Hennig et al., 2003; Yue et al., 2001; Zhu, 2003).

Fig. 1 shows data from a typical microarray experiment performed by my team in Syngenta using a wheat gene expression microarray (GEM) developed in collaboration with Incyte. This array contained approximately 10,000 cDNAs isolated from a variety of different tissues, the cDNAs were selected to each represent a single EST gene assembly and are thus largely non-redundant. Poly-(A)+ RNA was extracted from developing grain tissues (test RNA), labelled with Cy3 and hybridised to gene expression microarrays (GEMs) along with a standard RNA sample extracted from whole pre-anthesis spikes (control RNA) and labelled with Cy5. In this experiment, test material was grains without pericarp harvested at 12 days post-anthesis (DPA). This representation can only show general trends in gene expression, for instance the outlined group of genes showing much higher expression in grains than the control material contains a large number of genes involved in grain filling including starch biosynthetic genes and storage protein genes. A series of hybridisations were performed using the same test material with RNA from grains harvested at various days post-anthesis. The use of a common control sample allows the experiments to be linked together to show how gene expression changes with time after anthesis as shown for a group of genes with different function in Fig. 1(b). Analysis of the full dataset identified 2225 genes, almost a quarter of the total, which showed significant differential expression in the 28 days following anthesis. Clearly this group will contain many candidates, which could be used to usefully modify the development or composition of the wheat grain; however, with such a large number additional criteria are required to prioritise genes for low throughput techniques such as mutagenesis or transgenic analysis. One approach is to identify genes which have known functions in regulating gene expression for instance transcription factors. These genes offer the opportunity to alter the expression of many genes simultaneously resulting in changes in whole biosynthetic or developmental pathways. In the experiments described here 66 genes were identified which showed homology to