

Review

Using transcriptomics to understand the wheat genome

Tristan E. Coram^{1,*}, Gina Brown-Guedira¹ and Xianming Chen²

Address: ¹ US Department of Agriculture – Agricultural Research Service, and Department of Crop Science, North Carolina State University, Campus Box 7620, Raleigh, NC 27695-7620, USA. ² US Department of Agriculture – Agricultural Research Service, and Department of Plant Pathology, Washington State University, Pullman, WA 99163, USA.

***Correspondence:** Tristan E. Coram. Fax: +1-919-513-2039. Email: tristan.coram@ars.usda.gov

Received: 4 September 2008

Accepted: 9 October 2008

doi: 10.1079/PAVSNNR20083083

The electronic version of this article is the definitive one. It is located here: <http://www.cababstractsplus.org/cabreviews>

© CAB International 2008 (Online ISSN 1749-8848)

Abstract

Wheat (*Triticum aestivum* L.) is one of the most important food crops in the world, and transcriptomics studies of this crop promise to reveal the expression dynamics of genes that control many agriculturally important traits. In this review of wheat transcriptomics research, the current status of transcriptome surveying technology is presented, with particular emphasis on breakthrough techniques that will promote rapid progress in understanding the wheat genome. Microarrays have now become routine in wheat research, and the 55K Affymetrix Wheat GeneChip has enabled the generation of numerous high-quality datasets. In fact, the broad range of gene expression datasets provides future opportunities for integrating these data in a systematic approach that may reveal gene coexpression networks that underlie important traits. Wheat microarrays have also recently been used in other valuable approaches, including simultaneous transcriptome and genome profiling through single-feature polymorphism markers, the mapping of translocation breakpoints, and surveying of antisense transcription. The future use of wheat microarrays for gene expression measurement may be challenged by new sequencing-based transcriptomics techniques. These new techniques are presented, and the application of sequencing-by-synthesis as a future area of wheat transcriptomics research is highlighted. However, the yet to be fully sequenced polyploid wheat genome poses problems for some of these technologies when attempting to annotate and assign short-sequence tags. For this reason, the Roche 454 technology is considered the best option for future progress because of its longer sequence reads that can be more easily annotated, as well as its unbiased potential for covering the entire wheat transcriptome.

Keywords: Wheat, Transcriptome, Microarray, Gene expression, Genomics, Sequencing

Review Methodology: We searched the following databases for wheat transcriptomics related publications: CAB Abstracts, Web of Science and PubMed. Additionally, we searched company websites and the Genome Technology website for the latest sequencing-by-synthesis technology information. Finally, we spoke to colleagues for information on the most recent research yet to be published.

Wheat (*Triticum aestivum* L.) is the most important food crop in the world. In 2007, approximately 607 million tonnes of wheat grain was produced globally from approximately 217 million hectares of land [1]. Behind wheat, which had the largest area used for crop production, were maize (~158 million hectares) and rice (~157 million hectares). Wheat feeds nearly half of the world's expanding population: thus improving wheat yield capacity is of great importance for global food security.

Improvements in yield can be made, for example, through engineering wheat cultivars that are adaptable to diverse environmental conditions and that are able to tolerate the many pests and diseases that can attack the crop. To make progress in achieving these goals, an understanding of how the wheat genome responds to certain biotic and abiotic stresses is required. The field of transcriptomics focuses on such studies, where experiments are designed to monitor and manipulate the dynamics of gene expression

events that occur during both tolerance and susceptibility to particular stresses. In combination with gene knock-down methods, the results of transcriptomics studies may enable the identification of gene expression signatures associated with tolerance to a particular stress, and allow the development of gene expression markers that can be used to predict stress tolerance. In this review, the latest developments of knowledge and technology in the area of wheat transcriptomics will be presented, with the aim of identifying valuable prospects for future research and development.

Wheat Transcriptomics Using Microarrays

Microarray technology has become routine in many research laboratories. The release of the commercial Wheat GeneChip by Affymetrix (Santa Clara, CA, USA) has allowed for easy access to a high-quality large-scale wheat microarray containing probes to over 55 000 wheat transcripts. Subsequently, several transcriptomics studies are now being performed and published utilizing this resource. These studies focus on a wide range of cellular processes and stresses, including characterization of the transcriptome during meiosis [2], grain development [3], salt tolerance [4], aluminium tolerance [5] and hessian fly resistance [6]. Additionally, three published studies have focused on gene expression events associated with resistance to rust fungi [7–9], as well as one yet to be published involving one author of this review (Gina Brown-Guedira). These rust fungi resistance studies are of particular interest as they offer the opportunity for alignment of results to discover central gene co-expression networks involved in resistance to rust pathogens.

The benefits of high repeatability, low chip-to-chip variation and high probe density for the Wheat GeneChip also come with the disadvantage of relatively high cost, which has prohibited larger-scale studies involving many RNA samples. In these cases, researchers may prescreen for interesting transcripts using the Wheat GeneChip, and then construct a custom microarray based on the publicly available DNA sequences. For example, the authors of the present review used the results of two Wheat GeneChip stripe rust resistance studies [7, 8] to construct a custom long oligonucleotide microarray that is being used in an association-like gene expression study of differing stripe rust resistance sources in varying genetic backgrounds. The aim of this study is to identify the potential presence of common transcript patterns during successful stripe rust resistance, despite the resistance source (resistance gene) or genetic background of the wheat plant [10]. It is important to note that other high-density oligonucleotide microarrays for wheat exist and have been used successfully in transcriptomics studies. For example, a 19K microarray was constructed by Mohammadi *et al.* [11] that has been used to study the

dehydration response in wheat [11, 12]. Additionally, 22K and 11K custom oligonucleotide microarrays have been used to study salt-responsive gene expression events in wheat [13, 14], and a smaller (95 probe) custom microarray was constructed for the study of cold/heat tolerance [15].

The oligonucleotide-based microarrays, including the on-slide synthesized Wheat GeneChip and custom spotted long oligonucleotide arrays, require existing sequence knowledge to be constructed. As an alternative, spotted cDNA (or amplicon) arrays do not require sequence knowledge and can easily be constructed from polymerase chain reaction (PCR) products of a cDNA library. For example, Wilson *et al.* [16] developed an ~9K cDNA microarray from expressed sequence tags (ESTs) of 35 specific cDNA libraries, each representing specific stages in grain and plant development. The annotation for this array is available at <http://www.cerealsdb.uk.net>, and has recently been put to use for studying senescence in wheat flag leaves [17]. Because spotted cDNA arrays do not require sequence knowledge for construction, specific 'boutique' cDNA arrays can easily be made from libraries that have been enriched for potential target transcripts.

Limitations of Microarray-based Transcriptomics

The most widely recognized limitation of microarrays is the low reproducibility observed when using different array platforms [18]. This is usually the result of probe-specific effects such as oligonucleotide probes versus PCR product probes, and array-synthesis effects such as on-slide synthesis versus robotic spotting. Much argument exists over which is the better platform, but for wheat an important study was recently reported by Poole *et al.* [19] that compared the Affymetrix Wheat GeneChip array, a custom-spotted cDNA array, and quantitative reverse-transcription PCR (qRT-PCR). The authors postulated that the general weak reproducibility in cross-platform transcriptomics studies would be further weakened in wheat because it is an allohexaploid consisting of three closely related genomes (A, B and D), has about 25% of its gene motifs represented by at least two paralogous loci [20], and has a genome mainly consisting of repetitive sequences [21]. Briefly, the authors found that the genes having homologues and/or paralogues were more likely to show platform-specific differences in gene expression measurement as a result of cross-hybridization, and that the Affymetrix short oligonucleotide platform is less likely to experience such cross-hybridization. Globally, the two array platforms showed poor correlation, but this was improved when looking at only the differentially expressed transcripts. The results of this study are highly significant as they reveal that the polyploid nature of wheat may necessitate homologue-specific microarrays, otherwise the integration of cross-platform wheat

transcriptomics data may not be possible. The relatively deep coverage of EST resources for wheat has enabled the identification of probable homologues by alignment of sequences and the identification of polymorphisms that separate the homologues. Additionally, high-throughput sequencing of wheat aneuploid genetic stocks may also be used to identify homologues. Thus, a complete genome sequence for wheat will not be necessary for constructing homologue-specific microarrays.

Other limitations of microarray transcriptomics to be considered include within-platform variations that can arise between laboratories. However, standardized protocols can overcome most of this variation. In general, microarray platforms and experimental procedures once suffered from a lack of standardization that hindered the comparison of results between platforms and laboratories. However, efforts from the research community have addressed these problems. The Microarray Gene Expression Data (MGED) society has developed standards for microarray expression data known as Minimum Information About Microarray Experiments (MIAME) that include requirements for descriptions of biological samples, microarray platform and experimental procedures [22]. The MGED society has also developed MAGE-ML, which provides standards for microarray data exchange. To publish research involving microarrays, most journals require authors to follow these guidelines and deposit their data into a public repository that follows these guidelines. These include the Gene Expression Omnibus (GEO) [23], ArrayExpress [24] and PLEXdb [25], which is of particular importance as it specializes in plant and plant-pathogen microarrays.

Expression-level and Single-feature Polymorphism (SFP)

Wheat microarrays have also been exploited for novel purposes in addition to studying gene expression kinetics. A difference in gene expression between two genotypes (parents of a mapping population, for example) can be considered as an allelic marker, and is known as an expression level polymorphism (ELP). In plants, ELPs have been studied in mapping populations [26–29], and have been associated with various phenotypic traits, for example, seed development in wheat [30], qualitative pathogen resistance in *Arabidopsis thaliana* [31, 32] and domestication in maize [33]. ELPs are known to be caused by both *cis*-acting [34, 35] and *trans*-acting factor DNA sequence polymorphisms [36], as well as expression-altering deletions within a gene itself [34, 37]. ELPs between genotypes have been shown to correlate with underlying DNA sequence polymorphism in *A. thaliana* [26] at both a whole-genome level and a sub-chromosomal level, which indicated a significant impact of *cis*- or linked *trans*-acting, DNA polymorphism on gene expression.

Second, by taking advantage of the ability for short oligonucleotide probes to distinguish single base-pair differences during target hybridization, SFP genotyping markers have recently come of age. SFPs represent DNA sequence polymorphism between genotypes within an individual oligonucleotide probe that can be detected by hybridization affinity difference [38]. Importantly, SFPs occur within a transcribed gene that may reflect variation in phenotype. Recent studies have been successful in identifying DNA polymorphisms from RNA hybridized to Affymetrix GeneChips in addition to transcript abundance [38–42]. Use of RNA acts to reduce genome complexity, which overcomes the difficulties associated with hybridizing genomic DNA from organisms with large and complex genomes such as wheat. SFPs are often caused by single nucleotide polymorphisms (SNPs), as demonstrated by studies that have sequenced SFPs [38, 40]. However, SFPs are only predictive of a potential SNP within an oligonucleotide probe, so further work must be performed to actually identify the causative SNP.

In fact, expression-level polymorphisms (ELPs) [38] and SFPs [40] can be identified from single hybridization studies using short oligonucleotide microarrays, in particular the Affymetrix platform. While ELPs are assessed from whole probe sets of a transcript, SFPs are identified from individual 25-mer probes irrespective of the probe set. Both SFPs and ELPs can also be mapped as alleles in segregating populations ('genetical genomics'), and ELPs can be considered as e-traits to establish expression quantitative trait loci (eQTL) [29, 30]. In wheat, one eQTL study has been reported, which utilized the Affymetrix Wheat GeneChip to map ELPs for seed development in a population of 41 doubled-haploid lines [30]. By using wheat–rice synteny, the authors found that most eQTL of major effect were *cis*-acting, which has also been observed in eQTL studies of other organisms, including *A. thaliana* [29] and the mouse [43]. The authors did not identify SFPs in this report, but have recently indicated that this aspect of the study is near completion [44]. Also in wheat, one other recent study has been published concerning ELPs and SFPs, which assessed the extent and relationship of these two polymorphisms in near-isogenic genotypes that differed for stripe rust resistance [45]. The main findings of this study were that the proportion of the transcriptome estimated to be variable between isolines was 0.30% for the ELPs and 0.39% for the SFPs, which was highly similar to the theoretical genome difference between isolines of ~0.39%. Using wheat–rice synteny, the authors found that both ELPs and SFPs clustered together, and this strong physical correlation indicated that the ELPs may be regulated by *cis*-acting DNA polymorphisms. The results also demonstrated the usefulness of SFPs for developing PCR-based genetic markers for traits of interest. Overall, the use of ELPs and SFPs in wheat is very much in its infancy, but the techniques promise to greatly enhance transcriptomics in this crop by allowing for the detection of genome regions

that control gene expression responsible for certain traits.

Other Wheat Microarray Applications

The Affymetrix Wheat GeneChip has also been used to map translocation breakpoints in wheat. Bhat *et al.* [46] compared RNA from a normal hexaploid wheat with a ditelosomic aneuploid stock missing the short arm of chromosome 1B, as well as wheat-rye translocations that replace 1BS with rye 1RS. Under the assumption that transcripts detected by a probe set could come from any of the three wheat genomes and that absence/replacement of 1BS would result in transcript abundance patterns significantly different from normal hexaploid wheat, the authors assigned 257 probe sets to chromosome 1BS. With the assistance of wheat-rice synteny, the translocations further allowed for the placement of the probe sets into three groups with narrow translocation breakpoints. This study is important as it exploits the availability of wheat aneuploid stocks and the power of transcriptomics to map transcripts to physical bins on the wheat genome defined by deletions or translocation junctions. This technique could easily be applied to the entire wheat genome to allow for the cataloguing of GeneChip probe sets into chromosome segments, and also for identifying transcriptionally active gene-rich chromosomal regions.

With the ongoing efforts of the International Wheat Genome Sequencing Consortium (IWGSC; <http://www.wheatgenome.org>) and the availability of high-throughput and low-cost next-generation DNA sequencing platforms, the once unattainable goal of completing the wheat genome sequence is now possible within the foreseeable future. The completion of this task would open many doors for the expansion of wheat transcriptomics, including the development of whole genome tiling microarrays (Whole genome amplifications(WGAs)). WGAs are unbiased and allow for the discovery of all transcriptional activity within the genome. The complete genome, including inter-genic regions is represented as relatively short (<100-mer) probes on an array, which can also be used in comparative genomic hybridization (CGH) to detect deletions and polymorphisms, as well as methylation profiling and the analysis of chromatin immunoprecipitation samples (ChIP-chip) to map regulatory DNA motifs. For an in-depth explanation of the current status of WGA research in plants, please refer to the excellent recent review by Gregory *et al.* [47]. Of course, specific chromosome segment tiling arrays for wheat could also be developed for those studies that wish to target-specific genome regions. The results of WGA studies in *A. thaliana* [48] and rice [49] have identified a large number of novel transcription sites that were not predicted by gene prediction algorithms or cDNA collections. Thus, the transcriptome of plants is likely to be much more complex

than previously thought. In fact, many of the newly identified transcripts were found to be antisense (expressed from the opposite DNA strand and in the reverse orientation).

Using microarrays to discover antisense transcription has also been a recent development in the transcriptomics field. Interestingly, the first study of this type took advantage of incorrectly orientated probes in the first version of the Affymetrix Mouse array, and showed that the sensitivity of the commercial Affymetrix arrays was sufficient to detect natural antisense transcripts (NATs) [50]. NATs exist on the opposite DNA strand to the sense strand either at the same locus (*cis*-encoded) or a different locus (*trans*-encoded). Because NATs affect the expression of sense transcripts during transcription, RNA processing/transport and translation, they can affect all cellular processes [51]. Recently, Ge *et al.* [52] demonstrated that the Affymetrix Human Genome Exon array could be used to detect antisense transcripts at the whole-genome level. This report presented a method for detecting both sense and antisense transcription from the same RNA sample by modifying the labelling procedures. Building on this thinking, the authors of the present review (Tristan Coram and Xianming Chen, manuscript in preparation) used the Affymetrix Wheat GeneChip to survey both sense and antisense transcription from tissue-specific wheat RNA samples. The preliminary results of this study indicate a significant level of antisense transcription in the wheat genome that is generally tissue and developmental-stage specific. This discovery will lead the way for further microarray-based profiling of antisense transcription that may be vital to the control of certain important traits.

Sequencing-based Transcriptomics

While microarrays can be considered a 'closed' system for transcriptomics, several 'open' and unbiased systems have recently been developed, including the tag-based approaches of Serial Analysis of Gene Expression (SAGE) and Massively Parallel Signature Sequencing (MPSS). The first tag-based technique for studying transcriptomics was the generation of ESTs from cloned cDNAs. ESTs are one of the most important resources for transcriptomics in wheat, as the majority of current genome sequence available is derived from ESTs (1 050 791 ESTs submitted at dbEST as of 8 August, 2008; http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). Large-scale EST sequencing has been used to calculate transcript abundance and infer gene expression. Studies by Ogihara *et al.* [53], Mochida *et al.* [54] and Houde *et al.* [55] demonstrate this technique, and were valuable for identifying transcripts associated with abiotic stress tolerance. However, EST sequencing is limited in detecting low-abundance transcripts, thus SAGE was developed as a short-tag sequencing method with an estimated 26 times greater

sensitivity than EST sequencing [56]. Briefly, SAGE involves the synthesis of double-stranded cDNA from mRNA, followed by cleavage with an anchoring enzyme at the 3'-end of the molecule. Adaptor ligation with a Type II or Type III restriction enzyme site (tagging enzyme) is carried out, which is used to cut the short (14–26 bp) 3' tags out of the cDNA [57–59]. Multiple tags are then ligated together before cloning and sequencing, which greatly enhances the ability for high-throughput and maximizes data generation per sequencing run. MPSS is similar to SAGE in producing short tags (17–21 bp), but uses a sequencing-by-synthesis approach that is capable of producing much deeper coverage of the transcriptome than SAGE [60]. SAGE and MPSS data can be used to calculate absolute gene expression by counting the number of tags for each transcript, sometimes called an 'electronic northern'. Several SAGE-derived methods such as LongSAGE [58] and SuperSAGE [57] have built on the original technique to produce longer tags that can more easily be annotated. In wheat, just one SAGE-based study has been published, which used LongSAGE to study the transcriptome of developing wheat caryopsis [61]. The major problems of SAGE and MPSS is the difficulty of matching the short tags to EST or cDNA sequences during annotation, which is further exacerbated by the presence of highly similar homologous and paralogous sequences in the hexaploid wheat genome. In the wheat study, only 48% of the LongSAGE tags could be matched to existing sequence, which may also reflect the low sampling depth of EST sequencing. Studies have shown that short tag-based studies are of most value when a reference genome sequence is available for annotation of the tags [62]. For this reason, it is perhaps no surprise that many wheat researchers have turned to massively parallel sequencing-by-synthesis, which promises to combine the high-throughput of SAGE and MPSS with the accuracy of EST sequencing.

In particular, the recently developed Roche 454 pyrosequencing technology [63] may provide sufficient length of sequence information to allow accurate annotation without the need for a reference genome sequence. The technique firstly involves the nebulization of cDNA into fragments, which are linked to nanobeads that are captured in droplets of an emulsion where template amplification occurs. The nanobeads are then distributed into a high-density plate for sequencing-by-synthesis. As sequencing reagents are passed over the plate light is emitted and recorded. Currently, the Roche 454's FLX technology is capable of producing reads 200–300 bp in length and >100 Mb of sequence data from a single 7-h plate run (<http://www.454.com>), but improved technology known as 'GS FLX Titanium' allows for reads of >400 bp and a total of >500 Mb of sequence data per run and is now available to researchers (www.454.com). The technology has experienced problems in cDNA sequencing when it encounters long homopolymer runs such as poly-T stretches, but several successful cDNA sequencing projects have been reported in plants

[64–67]. Additionally, the new GS FLX Titanium technology also promises to eliminate this problem. Other next-generation sequencing-by-synthesis platforms are available and include Illumina's Solexa (<http://www.illumina.com>) and Applied Biosystems SOLiD (<http://solid.appliedbiosystems.com>). These platforms generate more data per run than Roche 454, but at the expense of shorter reads of <40 bp. For similar reasons to SAGE and MPSS, these short reads pose problems for wheat transcriptomics. Subsequently, wheat researchers should select the Roche 454 technology for transcriptome studies at least until a more comprehensive wheat genome sequence becomes available.

Several wheat research groups have in fact begun using the Roche 454 technology, although publication of these research findings is yet to appear. The authors of the present review are aware of one such study being performed at Washington State University (Pullman, WA, USA), where cDNA from the parents of two mapping population is being sequenced (Deven See, personal communication 2008). However, cDNA is being used for this study mainly to reduce genome complexity while the overall goal is to discover SNPs that can be used for genetic mapping. Also in wheat, Roche 454 sequencing has been used to discover novel microRNAs [68]. MicroRNAs are single-stranded non-coding RNAs approximately 20–22 nucleotides in length that regulate the expression of protein coding genes. MicroRNAs are an important aspect of transcriptome regulation in plants, and for more information please refer to the excellent review by Jones-Rhoades *et al.* [69].

Roche 454 sequencing has clear advantages over microarrays for transcriptomics research, because it is an open system that can potentially capture all the transcripts in a given sample and can also accurately measure gene expression by counting the number of identical transcripts detected. The main disadvantage of the technology at this stage is the cost, which restricts routine use of the technology in most laboratories. It will be interesting to see the future results of wheat transcriptomics studies using Roche 454 sequencing, as we expect the discovery of many new transcripts that are not represented on the current microarrays. Additionally, the results may provide guidance as to the size of the wheat transcriptome in relation to the genome, and greatly assist in identifying gene-rich regions.

Summary and Future Directions

Many tools are available for wheat transcriptomics research, which are briefly summarized in Table 1. The future ability of wheat transcriptomics to expand will depend greatly on the continuing efforts for sequencing the gene-space of this important crop. Currently, closed-end wheat microarrays are limited in their transcriptome coverage, while open-ended technologies including SAGE

Table 1 Current transcriptomics resources available in wheat

Tool/Resource	Description
Genome sequence	<ul style="list-style-type: none"> Coordinated effort underway by the IWGSC (http://www.wheatgenome.org) Recognized as a priority by research community (http://www.csrees.usda.gov/nea/plants/pdfs/wheat_conference_summary.pdf)
ESTs ¹	<ul style="list-style-type: none"> 1 050 791 entries
Oligonucleotide microarray	<ul style="list-style-type: none"> 55K Affymetrix Wheat GeneChip (http://www.affymetrix.com)
cDNA microarray	<ul style="list-style-type: none"> Multiple including ~9 K array [16]
Tiling microarray	<ul style="list-style-type: none"> Not currently available
Serial Analysis of Gene Expression (SAGE)	<ul style="list-style-type: none"> Applied for studying developing wheat caryopsis [61]
Massively Parallel Signature Sequencing (MPSS)	<ul style="list-style-type: none"> Not reported
Sequencing-by-synthesis	<ul style="list-style-type: none"> Roche 454 cDNA sequencing (Deven See, personal communication, 2008)
Deletion and aneuploid genetic stocks	<ul style="list-style-type: none"> Large collection available from the Wheat Genetic and Genomic Resources Center (http://www.k-state.edu/wgrc)
Transformation	<ul style="list-style-type: none"> Biolistic- and <i>Agrobacterium</i>-mediated DNA delivery systems [71]
Gene knockdown	<ul style="list-style-type: none"> RNA interference [72] Viral-induced gene silencing [73]
Databases/tools	<ul style="list-style-type: none"> Graingenes (http://wheat.pw.usda.gov) Gramene (http://www.gramene.org) TIGR Genome Database (http://www.tigr.org/tdb/e2k1/tae1) Wheat Genome Project (http://wheat.pw.usda.gov/NSF/htmlversion.html) INRA TriAnnot Project (http://urgi.versailles.inra.fr/projects/TriAnnot) wEST (http://wheat.pw.usda.gov/wEST) CerealsDB (http://www.cerealsdb.uk.net) HarvEST Wheat (http://harvest.ucr.edu) PLEXdb (http://plexdb.org) Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae) GrainSAGE (http://www.scu.edu.au/research/cpcg/igfp/index.php) Wheat SNP Project (http://wheat.pw.usda.gov/SNP/new/index.shtml)

¹ESTs listed in the National Center for Biotechnology Information (NCBI) EST database (GenBank dbEST) (8 August 2008; http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

and MPSS are limited by difficulties in annotating the short-sequence tags. Perhaps the most promising technique available is the Roche 454 sequencing-by-synthesis, which allows the potential for deep transcriptome coverage and produces long reads that are more amenable to annotation and specific genome assignment in the hexaploid genome. As this technology becomes more cost-effective and accessible, it may replace the use of microarrays in studying gene expression for many researchers. However, tiling microarrays will be of great use to the wheat research community, but again can only come to realization with large-scale genome sequencing.

A major focus in the future will be the merging and comparative analysis of transcriptome data between researchers. In particular, there are large resources of wheat microarray data now available, which provide opportunities for gene coexpression network analysis. The merging of data derived from a standardized platform, such as the Affymetrix Wheat GeneChip, will be much easier than cross-platform merging. For example, the authors of this review have participated in four studies of the wheat transcriptional response to rust pathogens

using the Affymetrix Wheat GeneChip platform. The data from these four studies will be integrated to perform a gene coexpression network analysis that has the potential to identify key gene expression signatures involved in rust resistance. This type of analysis has been successful in associating abiotic stress phenotypes in *A. thaliana* with underlying common and unique gene expression pathways and marker genes [70]. The linking of gene expression data to genetic information is also of great importance to understand the interplay among genome, transcriptome and phenotype. Simultaneous surveying of the genome and transcriptome on a single platform through, for example, the ELPs and SFPs discussed earlier, is a valuable technique for uncovering genetic elements that influence the transcriptome. Overall, wheat transcriptomics is progressing rapidly and in the future we will see a complete systematic approach that integrates the transcriptome with genome, proteome and metabolome. Additionally, gene expression databases will expand and become more standardized, which will accelerate the discovery and characterization of all genes in the wheat genome.

Acknowledgements

This research was supported in part by the US Department of Agriculture (USDA), Agricultural Research Service (ARS) (Project No. 5348-22000-014-00D) and the USDA-ARS Postdoctoral Program.

References

1. FAOSTAT. FAO Statistics Division 2008. Available from: URL: <http://www.faostat.fao.org> (accessed 18 August 2008).
2. Crismani W, Baumann U, Sutton T, Shirley N, Webster T, Spangenberg G, *et al.* Microarray expression analysis of meiosis and microsporogenesis in hexaploid bread wheat. *BMC Genomics* 2006;7:267.
3. Wan YF, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, *et al.* Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics* 2008;9:121.
4. Mott IW, Wang RRC. Comparative transcriptome analysis of salt-tolerant wheat germplasm lines using wheat genome arrays. *Plant Science* 2007;173(3):327–39.
5. Houde M, Diallo AO. Identification of genes and pathways associated with aluminum stress and tolerance using transcriptome profiling of wheat near-isogenic lines. *BMC Genomics* 2008;9:400.
6. Xuming L, Jianfa B, Li H, Lieceng Z, Xiang L, Nanyan W, *et al.* Gene expression of different wheat genotypes during attack by virulent and avirulent hessian fly (*Mayetiola destructor*) larvae. *Journal of Chemical Ecology* 2007;33(12):2171–94.
7. Coram TE, Settles ML, Chen XM. Transcriptome analysis of high-temperature adult-plant resistance conditioned by Yr39 during the wheat-*Puccinia striiformis* f. sp. *tritici* interaction. *Molecular Plant Pathology* 2008;9(4):479–93.
8. Coram TE, Wang MN, Chen XM. Transcriptome analysis of the wheat – *Puccinia striiformis* f. sp. *tritici* interaction. *Molecular Plant Pathology* 2008;9(2):157–69.
9. Hulbert SH, Bai J, Fellers JP, Pacheco MG, Bowden RL. Gene expression patterns in near isogenic lines for wheat rust resistance gene *Lr34/Yr18*. *Phytopathology* 2007;97(9):1083–93.
10. Coram TE, Huang X, Settles ML, Zhan G, Chen X. Comparative analysis of transcripts associated to all-stage resistance and high-temperature adult-plant resistance to stripe rust in wheat. *Phytopathology* 2008;98(6):S41.
11. Mohammadi M, Kav N, Deyholos M. Transcriptional profiling of hexaploid wheat (*Triticum aestivum* L.) roots identifies novel, dehydration-responsive genes. *Plant Cell and Environment* 2007;30:630–45.
12. Mohammadi M, Kav H, Deyholos M. Transcript expression profile of water-limited roots of hexaploid wheat (*Triticum aestivum* 'Opata'). *Genome* 2008;51:357–67.
13. Kawaura K, Mochida K, Yamazaki Y, Ogihara Y. Transcriptome analysis of salinity stress responses in common wheat using a 22k oligo-DNA microarray. *Functional and Integrative Genomics* 2006;6(2):132–42.
14. Kawaura K, Mochida K, Ogihara Y. Genome-wide analysis for identification of salt-responsive genes in common wheat. *Functional and Integrative Genomics* 2008;8:277–86.
15. Skinner DZ, Okubara PA, Baek K-H, Call DR. Long oligonucleotide microarrays in wheat: evaluation of hybridization signal amplification and an oligonucleotide-design computer script. *Functional and Integrative Genomics* 2005;5(2):70–9.
16. Wilson ID, Barker GLA, Beswick RW, Shepherd SK, Lu CG, Coghill JA, *et al.* A transcriptomics resource for wheat functional genomics. *Plant Biotechnology Journal* 2004;2(6):495–506.
17. Gregersen PL, Holm PB. Transcriptome analysis of senescence in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal* 2007;5(1):192–206.
18. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 2002;18(3):405–12.
19. Poole R, Barker G, Wilson ID, Coghill JA, Edwards KJ. Measuring global gene expression in polyploidy; a cautionary note from allohexaploid wheat. *Functional and Integrative Genomics* 2007;7(3):207–19.
20. Akhunov ED, Goodyear AW, Geng S, Qi LL, Echaliér B, Gill BS, *et al.* The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Research* 2003;13(5):753–63.
21. Mitra R, Bhatia CR. Repeated and non-repeated nucleotide sequences in diploid and polyploidy wheat species. *Heredity* 1973;31:251–62.
22. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, *et al.* Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nature Genetics* 2001;29(4):365–71.
23. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, *et al.* NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Research* 2005;33:D562–6.
24. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, *et al.* ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Research* 2003;31(1):68–71.
25. Wise RP, Caldo RA, Hong L, Shen L, Cannon E, Dickerson JA. BarleyBase/PLEXdb – a unified expression profiling database for plants and plant pathogens. *Methods in Molecular Biology* 2007;406:347–63.
26. Kliebenstein D, West M, van Leeuwen H, Kim K, Doerge R, Michelmore R, *et al.* Genomic survey of gene expression diversity in *Arabidopsis thaliana*. *Genetics* 2006;172:1179–89.
27. Kliebenstein D, West M, van Leeuwen H, Loudet O, Doerge R, St. Clair D. Identification of QTLs controlling gene expression networks defined *a priori*. *BMC Bioinformatics* 2006;7:308.
28. van Leeuwen H, Kliebenstein D, West M, Kim K, van Poecke R, Katagiri F, *et al.* Natural variation among *Arabidopsis thaliana* accessions for transcriptome response to exogenous salicylic acid. *The Plant Cell* 2007;19:2099–110.
29. West M, Kim K, Kliebenstein D, van Leeuwen H, Michelmore R, Doerge R, *et al.* Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics* 2007;175:1441–50.
30. Jordan MC, Somers DJ, Banks TW. Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. *Plant Biotechnology Journal* 2007;5(3):442–53.

8 Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources

31. Grant M, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, *et al.* Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* 1995;269:843–6.
32. Gassmann W, Hinsch ME, Staskawicz BJ. The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant Journal* 1999;20(3):265–77.
33. Wang R, Stec A, Hey J, Lukens L, Doebley J. The limits of selection during maize domestication. *Nature* 1999;398:236–9.
34. Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein D, Gershenzon J. The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. *The Plant Cell* 2001;13:2793–807.
35. Cong B, Liu JP, Tanksley SD. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99(21):13606–11.
36. Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101(44):15670–5.
37. Kliebenstein D, Lambrix V, Reichelt M, Gershenzon J, Mitchell-Olds T. Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. *The Plant Cell* 2001;13:681–93.
38. Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, *et al.* Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research* 2003;13(3):513–23.
39. Cui X, Xu J, Asghar R, Condamine P, Svensson J, Wanamaker S, *et al.* Detecting single-feature polymorphisms using oligonucleotide arrays and robustified projection pursuit. *Bioinformatics* 2005;21:3852–8.
40. Rostoks N, Borevitz J, Hedley P, Russell J, Mudie S, Morris J, *et al.* Single-feature polymorphism discovery in the barley transcriptome. *Genome Biology* 2005;6:R54.
41. Walia H, Wilson C, Condamine P, Ismail A, Xu J, Cui X, *et al.* Array-based genotyping and expression analysis of barley cv. Maythorpe and Golden Promise. *BMC Genomics* 2007;8:87.
42. West MAL, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA, *et al.* High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. *Genome Research* 2006;16(6):787–95.
43. Schadt E, Monks S, Drake T, Lusis A, Che N, Colinayo V, *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* 2003;422:297–302.
44. Somers DJ, Jordan MC, Banks TW. Single Feature Polymorphism Discovery Using The Affymetrix Wheat Genechip. *Plant and Animal Genomes XVI Conference*; San Diego, CA; 2008. p. 272.
45. Coram T, Settles M, Wang M, Chen X. Surveying expression level polymorphism and single-feature polymorphism in near-isogenic wheat lines differing for the Yr5 stripe rust resistance locus. *Theoretical and Applied Genetics* 2008;117:401–11.
46. Bhat PR, Lukaszewski A, Cui XP, Xu J, Svensson JT, Wanamaker S, *et al.* Mapping translocation breakpoints using a wheat microarray. *Nucleic Acids Research* 2007;35(9):2936–43.
47. Gregory BD, Yazaki J, Ecker JR. Utilizing tiling microarrays for whole-genome analysis in plants. *Plant Journal* 2008;53(4):636–44.
48. Yamada K, Lim J, Dale JM, Chen HM, Shinn P, Palm CJ, *et al.* Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 2003;302(5646):842–6.
49. Li L, Wang XF, Stolc V, Li XY, Zhang DF, Su N, *et al.* Genome-wide transcription analyses in rice using tiling microarrays. *Nature Genetics* 2006;38(1):124–9.
50. Werner A, Schmutzler G, Carlile M, Miles CG, Peters H. Expression profiling of antisense transcripts on DNA arrays. *Physiological Genomics* 2007;28:294–300.
51. Terry N, Rouze P. The sense of naturally transcribed antisense RNAs in plants. *Trends in Plant Science* 2000;5(9):394–6.
52. Ge XJ, Rubinstein WS, Jung YC, Wu QF. Genome-wide analysis of antisense transcription with Affymetrix exon array. *BMC Genomics* 2008;9:27.
53. Ogiwara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin-I T, *et al.* Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant Journal* 2003;33:1001–11.
54. Mochida K, Kawaura K, Shimosaka E, Kawakami N, Shin-I T, Kohara Y, *et al.* Tissue expression map of a large number of expressed sequence tags and its application to *in silico* screening of stress response genes in common wheat. *Molecular Genetics and Genomics* 2006;276:304–12.
55. Houde M, Belcaid M, Ouellet F, Danyluk J, Monroy A, Dryanova A, *et al.* Wheat EST resources for functional genomics of abiotic stress. *BMC Genomics* 2006;7:149.
56. Sun M, Zhou GL, Lee S, Chen JJ, Shi RZ, Wang SM. SAGE is far more sensitive than EST for detecting low-abundance transcripts. *BMC Genomics* 2004;5:1.
57. Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, *et al.* Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100(26):15718–23.
58. Gowda M, Jantasuriyarat C, Dean RA, Wang GL. Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. *Plant Physiology* 2004;134(3):890–7.
59. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene-expression. *Science* 1995;270(5235):484–7.
60. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 2000;18(6):630–4.
61. McIntosh S, Watson L, Bundock P, Crawford A, White J, Cordeiro G, *et al.* SAGE of the developing wheat caryopsis. *Plant Biotechnology Journal* 2007;5(1):69–83.

62. Robinson SJ, Cram DJ, Lewis CT, Parkin IAP. Maximizing the efficacy of SAGE analysis identifies novel transcripts in *arabidopsis*(1[w]). *Plant Physiology* 2004;136(2):3223–33.
63. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437(7057): 376–80.
64. Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB. Sampling the *arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiology* 2007; 144(1):32–42.
65. Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou RL, Chen TL, *et al.* Global gene expression analysis of the shoot apical meristem of maize (*Zea mays* L.). *Plant Journal* 2007;52(3):391–404.
66. Cheung F, Haas BJ, Goldberg SMD, May GD, Xiao YL, Town CD. Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. *BMC Genomics* 2006;7:10.
67. Emrich SJ, Barbazuk WB, Li L, Schnable PS. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research* 2007;17(1):69–73.
68. Yao YY, Guo GG, Ni ZF, Sunkar R, Du JK, Zhu JK, *et al.* Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). *Genome Biology* 2007;8(6):13.
69. Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAs and their regulatory roles in plants. *Annual Review Plant Biology* 2006;57:19–53.
70. Weston DJ, Gunter LE, Rogers A, Wulfschleger SD. Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. *BMC Systems Biology* 2008;2:17.
71. Jones HD. Wheat transformation: current technology and applications to grain development and composition. *Journal of Cereal Science* 2005;41(2):137–47.
72. Fu DL, Uauy C, Blechl A, Dubcovsky J. RNA interference for wheat functional gene analysis. *Transgenic Research* 2007;16(6):689–701.
73. Scofield SR, Huang L, Brandt AS, Gill BS. Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the *Lr21*-mediated leaf rust resistance pathway. *Plant Physiology* 2005;138(4): 2165–73.