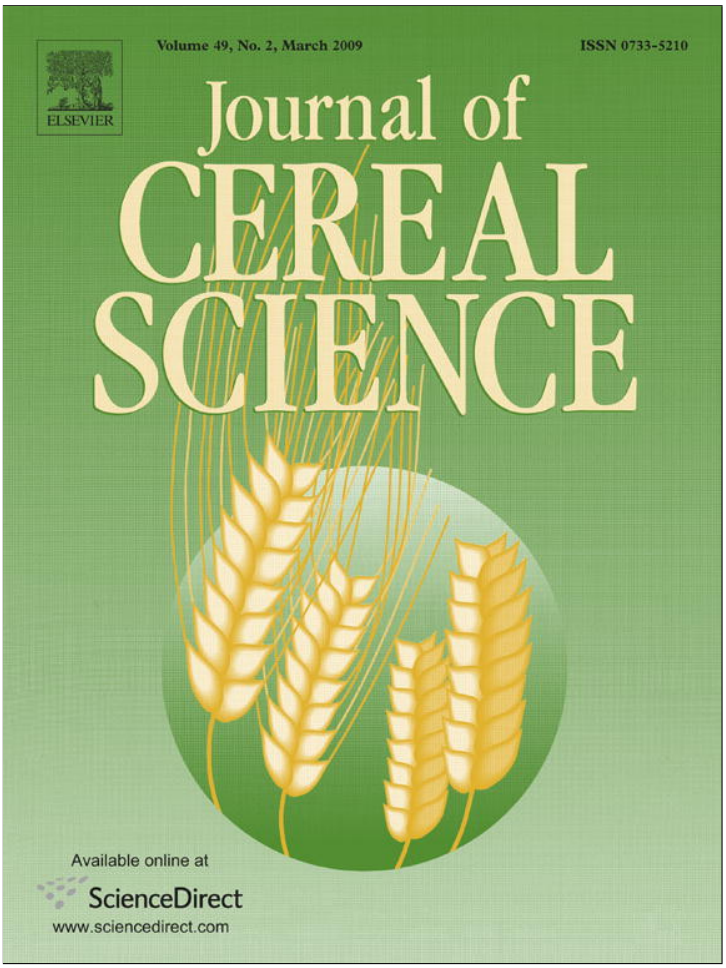
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Journal of Cereal Science 49 (2009) 316–318

Contents lists available at ScienceDirect

Journal of Cereal Science

journal homepage: www.elsevier.com/locate/jcs

Research Note

Identification of differentially expressed UniGenes in developing wheat seed using Digital Differential Display

Imad Eujayl [a,*], Craig Morris [b]

[a] USDA-ARS, Northwest Irrigation and Soils Research Laboratory, 3793 N. 3600 E., Kimberly, ID 83341, USA
[b] USDA-ARS, Western Wheat Quality Laboratory, E-202 Food Science and Human Nutrition Facility East, Washington State University, Pullman, WA 99164, USA

ARTICLE INFO

Article history:
Received 30 May 2008
Received in revised form
5 November 2008
Accepted 11 November 2008

Keywords:
Gene expression
Kernel hardness

ABSTRACT

The wheat UniGene sets, derived from over one million Expressed Sequence Tags (ESTs) in the NCBI GenBank, offer a platform for identifying differentially expressed genes in wheat seeds. This report illustrates a means to efficiently utilize this public database for gene expression (transcriptome) profiling of developing wheat seed. Using a data mining tool known as Digital Differential Display (DDD), thirteen pair-wise comparisons were performed on seven seed cDNA libraries from five varieties at various seed development stages. DDD identified 46 seed-specific UniGene sets, excluding the well characterized "housekeeping" and seed storage protein genes. Additionally, seed- and developmentally-specific UniGenes were identified. Some of these genes encode for proteins such as purothionins, serpins, α-amylase inhibitors, lipid transfer proteins, and other unknown but novel gene sequences. Specifically, the wheat serpin and β-purothionin precursor were found to be expressed at higher levels in hard varieties than soft varieties. This study supports the starting premise that by implementing *in-silico* analysis of the wheat UniGene database, it is possible to rapidly create transcriptional profiles of known and novel genes in developing seeds.

Published by Elsevier Ltd.

1. Introduction

The expression pattern of many genes is species-, tissue- or developmentally-specific. Differential expression of any given gene is determined by whether it is significantly under- or over-expressed relative to some other reference gene(s). The level of expression of a gene is commonly estimated using two analysis approaches referred to as 'analog' and 'digital' (Audic and Claverie, 1997). These methods vary considerably in their efficacy, genome coverage, data-points delivery, and cost-effectiveness. The analog methods are based on oligonucleotide probe hybridizations such as Northern blotting, mRNA differential display, and DNA microarrays. The digital methods are based on high throughput generation of gene transcripts, which vary in length from 300 to 500 bp as in the case of Expressed Sequence Tags (ESTs) or as short as 9 bp in the case of Serial Analysis of Gene Expression (SAGE) (Velculescu et al., 1995). *In-silico* gene expression profiling using EST database mining has a documented impact on gene discovery in mammalian systems (Huminiecki and Bicknell, 2000; Scheurle et al., 2000). The approach follows the logic that a given set of ESTs is a representation of the transcriptome of that species, tissue or developmental stage (Okubo et al., 1992). A key aspect of the *in-silico* approach is the efficient analysis of these very large EST data sets, and the identification of key features between/among different gene populations (i.e. genes of potential interest). A relatively recent tool for *in-silico* database mining is Digital Differential Display (DDD). Here, we illustrate the utility of DDD for the *in-silico* analysis of the wheat EST database for quantitative transcriptional profiling. A further objective of this study is to identify differentially expressed genes at different stages of wheat seed development and to analyze gene expression patterns related to kernel texture (hardness).

2. Materials and methods

2.1. Wheat cDNA libraries and varieties

The total number of wheat ESTs in the National Center for Biotechnology Information (NCBI) database (dbEST) is 1,051,300 (October, 2008). This database was used to analyze gene expression levels related to wheat seed development and kernel texture. The ESTs were computationally clustered into 41,289 UniGene sets as a derivative database provided as an additional resource by the

Abbreviations: bp, base pairs; dbEST, EST database; DDD, Digital Differential Display; dpa, days post anthesis; EST, Expressed Sequence Tag; NCBI, National Center for Biotechnology Information; Pina, Puroindoline a; Pinb, Puroindoline b; RT-PCR, Reverse-Transcriptase Polymerase Chain Reaction; SAGE, Serial Analysis of Gene Expression.

* Corresponding author. Tel.: +1 208 423 6544; fax: +1 208 423 6555.
E-mail address: imad.eujayl@ars.usda.gov (I. Eujayl).

0733-5210/$ – see front matter Published by Elsevier Ltd.
doi:10.1016/j.jcs.2008.11.003

Author's personal copy

NCBI. An individual UniGene set is defined as a group of transcript sequences that, based on sequence homology, originate from the same gene or expressed pseudogene (Pontius et al., 2002). In this study, the UniGene database was analyzed by an *in-silico* tool known as Digital Differential Display (DDD) (http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi). DDD is an algorithmic system for the identification of differentially expressed genes based on the relative abundance of ESTs from two or more contrasting cDNA libraries. To account for the unequal number of ESTs in each library, DDD uses the Fisher Exact Test (Siegel, 1956) to determine statistically significant differences ($P \leq 0.05$). The test is formulated for the analysis of randomly collected samples from populations with unequal size (in this case, number of ESTs) and uses the Bonferroni inequality test in declaring statistical differences.

DDD was used to analyze seven cDNA libraries from five bread wheat varieties. Based on database annotation, tissue samples were collected from seeds at various stages of development. The varieties and seed developmental stages were as follows: Chinese Spring (CS) (soft kernel texture), 10 and 30 days post anthesis (dpa); Glenlea (hard kernel texture), 5 and 15 dpa; Butte 86 (hard kernel texture), 3–44 dpa pooled; Cheyenne (hard kernel texture), 5–30 dpa pooled; and Wyuna (soft kernel texture), 8–12 dpa pooled. Additionally, three libraries from CS root, crown and shoot tissues from seedlings, and mature plants were included in the analysis.

2.2. Digital Differential Display analysis

DDD comparisons were performed to identify differentially expressed genes in endosperm and seed at various developmental stages. The first comparison was performed between CS seed libraries and non-seed libraries (vegetative and root) to identify UniGene sets that were seed-specific. Subsequently ten inter-varietal DDD comparisons were conducted to identify differentially expressed genes between individual or pooled cDNA libraries. cDNAs from different cultivars were compared based on the tissue, either endosperm or whole grain (seed) tissue. Also comparisons targeted soft vs. hard kernel contrasts. Two DDD comparisons were conducted to identify developmentally stage-specific differential gene expression within CS (10 vs. 30 dpa) and within Glenlea (5 vs. 15 dpa) seeds. The comparisons were systematically designed to achieve step-wise identification of "house-keeping genes", constitutive expression levels, developmentally-specific genes, and finally, expression levels associated with kernel texture class. Statistical analysis of DDD was based on the Fisher Exact Test. The results of DDD were tabulated numerically as a fraction of the pool of ESTs, and as a graphic dot intensity plot reflecting relative sequence abundance.

3. Results and discussion

This study explored the utility of DDD as a means of *in-silico* digital transcriptome profiling to identify differentially expressed genes in developing wheat endosperm and whole grain. An intra-varietal comparison between UniGene sets representing endosperm and seed libraries and libraries from shoot, crown and root in CS identified a number of differentially expressed "house-keeping" genes and seed storage protein genes (various types of gliadins and glutenins). For example, gamma-gliadin (Ta.27702) transcripts were detected only in seed cDNA libraries supporting the validity of the DDD approach (data not shown). These UniGenes were excluded for further analysis. Of the remainder, forty-six UniGene clusters were present in the seed libraries but showed no expression (zero transcripts detected) in CS shoot, crown, and root libraries. These seed and endosperm-specific candidate genes were subjected to a second round of DDD comparisons between seed and endosperm libraries that varied for variety, kernel texture (puroindoline haplotype; Morris, 2002), and developmental stage. An advantage of DDD is that multiple cDNA libraries can be digitally 'pooled' or combined to construct various contrasts. These comparisons were aimed at minimizing any confounding by seed developmental stage, that is, comparisons included transcripts that were expressed at multiple stages of endosperm development. The DDD indicated a number of significant differences in the gene transcript frequency. Table 1 shows differentially expressed UniGenes between hard and soft cultivars presented in percentage and fold change. For example, Puroindoline a (Pina), α-1-purothionin precursor, and α-amylase inhibitor were all up-regulated in Cheyenne at 13.8 fold increase compared to Wyuna. On the other hand, in the same comparison the putative avenin-like precursor was found up-regulated in Wyuna at 5.6 fold higher than Cheyenne, but UniGene Ta.55453 transcripts were not detected in Wyuna (Table 1). Comparisons between CS 10 plus 30 dpa vs. Butte 86, and Wyuna (soft) vs. Glenlea 5 plus 15 dpa pools returned no significant differences for either Pin gene. When Cheyenne was compared with Glenlea 5 plus 15, no expression (0 ESTs) of Pina was detected in the Glenlea pool. This result is consistent with Glenlea being a Pina null genotype (data not shown). Interestingly, also no Pinb transcript was detected in the Glenlea pool, which might be explained by the relatively early stages (5 and 10 dpa) at which the tissue was collected for library construction. The comparison between soft and hard whole grain cDNA libraries of CS and Butte 86 revealed that there are three UniGenes; Ta.56011, Ta.54994, and Ta.54616 that were not transcribed in CS (Table 2).

Table 1
Differentially expressed wheat UniGenes detected by Digital Differential Display (DDD) comparisons of endosperm cDNA libraries between Wyuna (soft) and Cheyenne (hard) cultivars. Expression levels are presented in percentage of total transcripts (ESTs).

UniGene	UniGene description	Cheyenne, %	Wyuna, %	Fold change CNN vs. WY[a]
Ta.41965	Puroindoline-a	0.0069	0.0005	+13.8
Ta.54476	Putative avenin-like precursor	0.0018	0.0100	−5.6
Ta.91	α-1 purothionin	0.0191	0.0024	+7.9
Ta.28296	α-amylase inhibitor gene	0.0093	0.0008	+11.6
Ta.50490	Full inset of unknown-mRNA (wde2f.pk001.115:fis)	0.0087	0.0005	+17.4
Ta.54224	Transcribed locus strongly similar to unknown rice clone – NP_001046024.1	0.0107	0.0027	+3.9
Ta.55453	Transcribed locus weakly similar to unknown rice clone – NP_001059189.1	0.0067	0.0	–

– No UniGenes detected in Wyuna.
[a] CNN = Cheyenne, WY = Wyuna, (+) sign indicates fold increase and (−) indicates decrease in CNN.

Table 2
Differentially expressed wheat UniGenes detected by Digital Differential Display (DDD) comparisons of seed cDNA libraries pooled from various developing stages between Chinese Spring (soft) and Butte 86 (hard) cultivars. Expression levels are presented in percentage of total transcripts (ESTs).

UniGene	Description	Butte 86, %	Chinese Spring, %	Fold change Butte vs. CS[a]
Ta.54284	β-amylase (LOC542896)	0.0057	0.0013	+4.4
Ta.9226	Pathogenesis-related protein 4 (PR 4)	0.0030	0.0001	+30.0
Ta.1315	Monomeric α-amylase inhibitor (Ima1)	0.0042	0.0006	+7.0
Ta.817	Precursor protein (AA-25 to 143)	0.0081	0.0015	+5.4
Ta.2799	Metallothionin (class II) – EC protein	0.0045	0.0001	+45.0
Ta.56011	Thaumatin-like protein (LOC542887)	0.0021	0.0	–
Ta.54994	Transcribed protein moderately similar to rice locus – NP001050984.1	0.0024	0.0	–
Ta.54616	Transcribed protein weakly similar to rice locus – NP001695130.1	0.0015	0.0	–

– No UniGenes detected in CS.
[a] CS = Chinese spring. (+) sign indicates fold increase in Butte.

Author's personal copy

Table 3
Wheat UniGenes not detected at either 10 or 30 days post anthesis (dpa) in Chinese Spring as determined by Digital Differential Display (DDD) comparisons of two seed cDNA libraries (total 12,556 ESTs).

UniGene ID	Description	Expression detected, dpa	Fraction of ESTs in the cDNA library, %	No expression detected, dpa
Ta.115	Puroindoline-b	10	0.0017	30
Ta.51862	Ribulose-1,5 biphosphate carboxylase	10	0.0025	30
Ta.8157	Glucose-1-phosphate adenyltransferase	10	0.0041	30
Ta.28296	α-amylase inhibitor gene	10	0.0030	30
Ta.817	α-amylase tetrameric inhibitor-subunit CM3	10	0.0032	30
Ta.54456	Type 1 non-specific lipid transfer protein precursor (LTP9.1b gene)	30	0.0481	10
Ta.54469	Transcribed locus weakly similar to rice gene – NP_0010655831.1	30	0.0463	10
Ta.14501	Transcribed locus strongly similar to rice gene – NP_001056364.1	30	0.0027	10

Developmental processes in seeds are under temporal control. As expected, performing DDD contrasts within the same variety but using libraries constructed at different stages of seed development, we detected genes that showed no expression at certain developmental stages (complete up and down regulation). Table 3 lists several selected genes that showed no transcripts at either 10 or 30 dpa of CS seed development. The functions of these genes vary from cell wall related functions (nsLTP precursor, Ta.54456), amylase inhibitors (Ta.28296), and novel full-length mRNAs of unknown function (Ta.14501). Ribulose-1,5 bisphosphate carboxylase/oxygenase (RuBisCO) large subunit gene (Ta.51862) was found to be present in CS 10 dpa, but was down-regulated to an undetectable level at 30 dpa. The DDD detected no Pinb transcripts at 30 dpa in CS while the number of transcripts of Pina significantly declined at 30 dpa but were still present at a low frequency (data not shown).

Comparisons between the two Glenlea libraries (5 and 15 dpa) identified sucrose synthase (Ta.93) to be down-regulated at 15 dpa, whereas Serpin (Ta.1314) was significantly up-regulated with no transcripts detected at 5 dpa. Hejgaard (2001) reported that serpins are a well characterized super-family of cereal proteins with documented inter- and intra-cellular substrate binding abilities and involvement in plant defense systems.

4. Conclusions

This study demonstrated the utility of Digital Differential Display (DDD) as a tool to analyze the wheat transcriptome via the NCBI UniGene dbEST database. To our knowledge this is the first report illustrating the application of DDD in wheat. It also supports the starting premise that *in-silico* data mining can rapidly create a targeted list of candidate genes and generate transcriptional profiles of known and novel genes associated with seed development. In this study, DDD enabled the identification of numerical differences in transcript frequency between individual or pooled cDNA libraries from various seed development stages and genotypes. These differences are likely related to biological processes of interest to cereal chemists and other biologists. The challenge of identifying the precise functions of the deciphered differentially expressed genes still remains, but the fast pace of bioinformatics and advancement in gene annotation can offer insights that may be verified through further research. It should be noted that any DDD results should be validated experimentally using quantitative RT-PCR or other methods using a panel of genotypes and tissues.

References

Audic, S., Claverie, J.M., 1997. The significance of digital gene expression profiles. Genome Research 7, 986–995.
Hejgaard, J., 2001. Inhibitory serpins from rye grain with glutamine as P1 and P2 residues in the reactive center. Federation of European Biochemical Societies Letters 488, 149–153.
Huminiecki, L., Bicknell, R., 2000. In-silico cloning of novel endothelial-specific genes. Genome Research 10, 1796–1806.
Morris, C.F., 2002. Puroindolines, the molecular genetic basis of wheat grain hardness. Plant Molecular Biology 48, 633–647.
Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Koijama, Y., Matsuba, K., 1992. Large scale cDNA sequencing for analysis of quantitative aspects of gene expression. Nature Genetics 23, 173–179.
Pontius, J., Wagner, L., Schuler, G.D., 2002. UniGene, a unified view of the transcriptome. In: The NCBI Handbook National Center for Biotechnology Information, Bethesda, MD, USA.
Scheurle, D., DeYound, M.P., Binniger, D.M., Page, H., Jahanzeb, M., Narayann, R., 2000. Cancer gene discovery using Digital Differential Display. Cancer Research 60, 4037–4043.
Siegel, S., 1956. Nonparametric Statistics for the Behavioral Sciences. McGraw Hill Publ., New York.
Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science 270, 484–487.