

*Centre for Plant Conservation Genetics*

*Centre for Plant Conservation Genetics Papers*

---

*Southern Cross University*

*Year 2007*

---

Understanding gene expression in the  
developing grain

Toni Pacey-Miller*	Lorraine Watson†	Jessica F. White‡
Shane R. McIntosh**	Allison C. Crawford††	Peter C. Bundock‡‡
Daniel Barbary§	Giovanni M. Cordeiro¶	Robert J. Henry

\*Southern Cross University

†Southern Cross University

‡Southern Cross University

\*\*Southern Cross University

††

‡‡Southern Cross University

§Southern Cross University

¶Southern Cross University

||Southern Cross University

This paper is posted at ePublications@SCU.

[http://epubs.scu.edu.au/cpcg\\_pubs/200](http://epubs.scu.edu.au/cpcg_pubs/200)

Pacey-Miller, T, Watson, L, White, JF, McIntosh, SR, Crawford, AC, Bundock, PC, Barbary, D, Cordeiro, GM & Henry, RJ 2007, 'Understanding gene expression in the developing grain', *57th Australian Cereal Chemistry Conference*, Melbourne VIC, Australia, 5-10 August.

## UNDERSTANDING GENE EXPRESSION IN THE DEVELOPING GRAIN

T. Pacey-Miller, S. McIntosh, L. Watson, J. White, A. Crawford, P. Bundock, D. Barbary, G. Cordeiro and R. Henry

Grain Foods CRC, Ltd., Centre for Plant Conservation Genetics, Southern Cross University, Lismore, 2480, Australia.

### INTRODUCTION

Serial Analysis of Gene Expression (SAGE) has been used to understand more comprehensively the gene expression involved in the growth of a developing wheat grain. SAGE is a technique that allows rapid, detailed analysis of thousands of transcripts in a cell. The abundance of a particular tag relates directly to the expression level of the gene from which it is derived. The analysis of gene expression of different physiological states provides both qualitative and quantitative information. Seed development is a complex process involving both the up and down regulation of many genes. Samples of the developing grain were taken at 8, 14, 20, 30 and 40 days post anthesis. From these developmental stages five SAGE libraries were constructed totalling approximately 100 000 tags, of which around 28 900 were unique. Differential expression profiles of abundant tags from each library reveal the coordinated expression of genes responsible for the cellular events constituting caryopsis development. Analysis has identified genes involved in activities of cellular proliferation, structure and differentiation, storage protein accumulation and starch biosynthesis. This SAGE platform has also provided a resource of novel sequence and expression information including promoter activities. Further investigations into the abundant, low expressing transcripts will hopefully provide greater insight into wheat caryopsis development and assist in wheat improvement programs.

### MATERIALS AND METHODS

#### *Plant material*

The Australian spring wheat variety *Triticum aestivum* cv. Banks was grown in a controlled plant growth cabinet (Phoenix Biosystems, Australia). Conditions of growth ranged from 10 to 16 hours of daylight and temperatures from 15 to 23°C. The conditions for growth were altered at three week intervals to mimic field conditions. Plants were monitored daily during florescence to identify the point of anthesis. Seed material was collected from the middle region of each ear and from the same tiller number of each plant at 8 days post anthesis (dpa), 14dpa, 20dpa, 30dpa and 40dpa (mature dried seed). To allow for biological variability pooled samples were derived from between 6 and 10 plants.

#### *RNA Extraction*

All seeds were surface sterilized prior to RNA extraction. For each developmental time point 300mg of seed was extracted firstly with TRIzol<sup>®</sup> Reagent (Invitrogen, Cat. No. 15596-026) as per manufacturers instructions, followed by ethanol precipitation and purified with the Qiagen RNeasy Midi Kit (Cat. No. 75142) as per manufactures protocol. The total RNA was assessed for quantity and quality by absorbance at 260nm and by gel electrophoresis.

### *SAGE Library Construction and Sequencing*

The five longSAGE libraries were all constructed using the I-SAGE™ Long Kit (Invitrogen, Cat. No. T5000-03) with modifications to the manufacturer's protocol as described by Crawford *et al.* (Crawford *et al.*, 2005), Gowda *et al.* (Gowda *et al.*, 2004). Cloned concatemers were sequenced using Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit and the M13 (-20) forward primer. Automated DNA sequencing was carried out with an ABI 3730xl 48 capillary DNA analyser (Applied Biosystems).

### *Tag to Gene Annotation*

Confirmation of tag identity was conducted using BLASTN (2.2.2) to locate matching tag sites in the NCBI (National Center for Biotechnology Information) database dbEST\_others. Unannotated ESTs were subject to BLASTN searches in either the NCBI nr (non-redundant) database or the plant gene indices data bases on TIGR (The Institute for Genomic Research).

### *GeneSpring analysis*

SAGE tag sequences and abundances from all five libraries were clustered to generate a *Triticum aestivum* var. Banks transcriptome using GeneSpring® GX version 7.2 software (SiliconGenetics; Agilent Technologies, www.agilent.com).

### *Functional Grouping/ Ontology*

Assignment of BLASTN match to functional grouping or ontology utilised the data base listed: the Gene Ontology ([www.geneontology.org/](http://www.geneontology.org/)); ExPASy- Proteomic Server of the Swiss Institute of Bioinformatics (<http://au.expasy.org/>); UniProt - ([www.pir.uniprot.org/](http://www.pir.uniprot.org/)); AmiGO- browser (<http://amigo.nbn.ac.za/cgi-bin/go.cgi>) for BLASTN queries against GOst; European Bioinformatics Institute (<http://www.ebi.ac.uk/>) and The Institute for Genomic Research (TIGR) - <http://www.tigr.org/>. For the majority, ontology assignments were reinforced with a review of publicly available literature.

## **RESULTS AND DISCUSSION**

A total of five longSAGE libraries were constructed. For each library approximately 20,000 tags were isolated and sequenced giving a total of 96,441 tag sequences across development (Table 1). Of these tags between 6,000 and 10,000 tags were unique to each library. A comparison of the total unique tags from each library revealed that 28,952 tags are unique across all libraries. Of the 31528 singletons, 20,238 are considered real and not artefacts of reverse transcription or sequencing as they appear in one or more of the other libraries. It is these low-level expressed transcripts which may provide regulatory information and represent a large number of possible novel genes previously undetected in other expression studies.

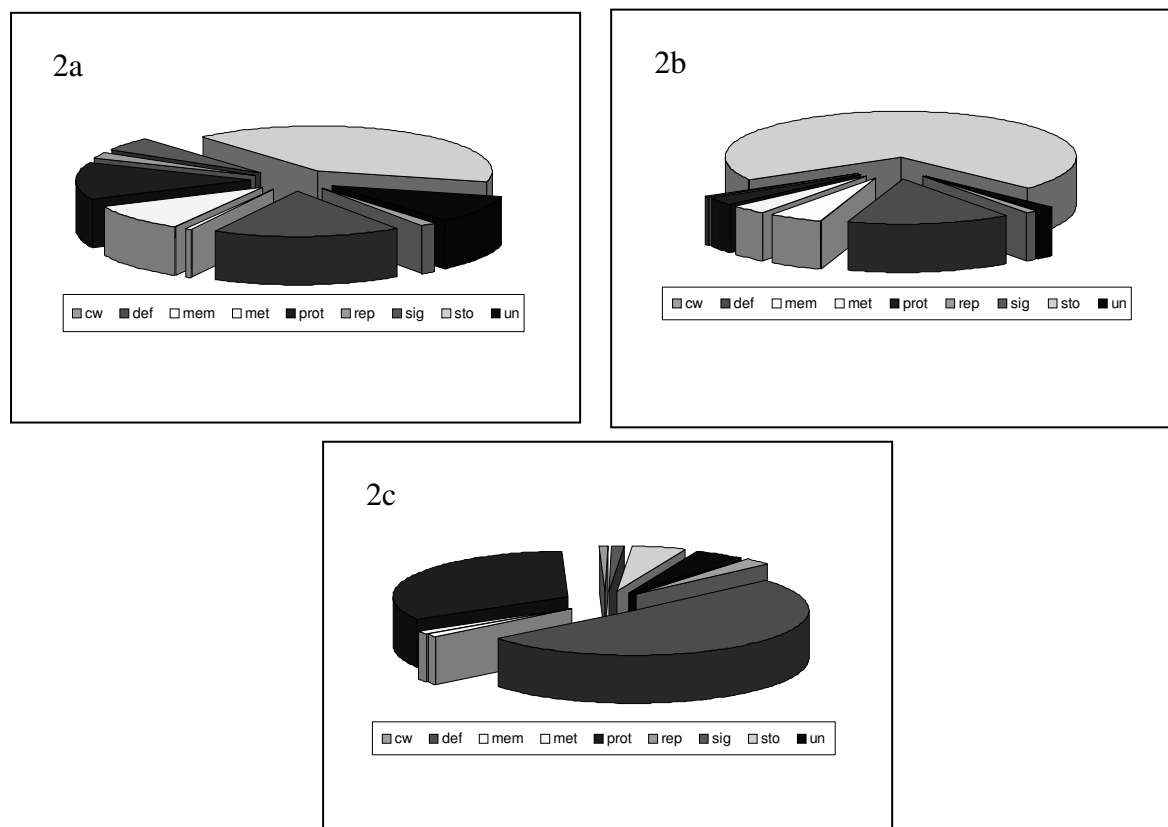
The main trends that are evident in the graphs (Figure 1) includes the dominance of storage proteins at 20dpa and the increase in transcripts for defense and protein synthesis in the 40dpa tissue. 8dpa is reflective of the broader array of metabolic and cellular events occurring at this early stage in development. Early stages in development focuses on building and sculpting of cellular architecture for the onset of seed fill, incorporating processes of cellular division, elongation and differentiation.

**Table 1.** Summary of transcripts

<i>Developmental stage</i>	8dpa	14dpa	20dpa	30dpa	40dpa
<i>Days Post Anthesis (dpa)</i>					
<i>Total tags sequenced</i>	20017	19299	17706	17709	21710
<i>Unique tags</i>	10045	7782	6388	6444	10022
<i>Abundance range #</i>	1-227	1-718	1-702	1-637	1-1027
<i>Abundance range %</i>	0.0049-1.13	0.0052-3.72	0.0056-3.96	0.0056-3.59	0.0046-4.73
<i>Tags occurring to <math>\geq 10</math> times</i>	28%	43%	50%	49%	36%
<i>Gene count</i>	238	157	145	156	191
<i>Tags occurring once</i>	38%	31%	28%	28%	36%
<i>Gene count</i>	7579	6058	4980	4985	7926

**Figure 1a-1c.** Division of relative abundances of the 500 most abundant expressed transcripts by ontology.

2a-8dpa; 2b-20dpa 2c-40dpa. Abbr. cw- cell wall; def- defence; mem- transport and membrane proteins; met- metabolism (enzymes); prot- protein synthesis and degradation, ribosome components; rep- cell cycle, reproduction, cytoskeleton, RNA/DNA binding; sig- signalling; sto- storage proteins; un- unknown and hypothetical.



The construction of five longSAGE libraries using wheat variety 'Banks' has generated a broad platform of information representing the global gene expression of a developing wheat caryopsis. We have examined the genes and how they relate to the developmental process. It is evident that SAGE is a powerful tool for identifying varying levels of transcriptional and translational control. A greater understanding of these cellular events occurring during development and the underlying mechanism of control will contribute to future crop improvement programs.

#### REFERENCES

Crawford, A., White, J., Bundock, P., Cordeiro, G., McIntosh, S., Pacey-Miller, T., Rooke, L., and Henry, R. (2005). *Plant Molecular Biology Reporter* 23, 139-143.

Gowda, M., Jantasuriyarat, C., Dean, R., and Wang, G. (2004). Robust-LongSAGE (RL-SAGE): *Plant Physiology* 134, 890-897.