SAGE of the developing wheat caryopsis

Shane McIntosh
Southern Cross University

Loraine Watson
Southern Cross University

Peter C. Bundock
Southern Cross University

Allison C. Crawford

Jessica F. White
Southern Cross University

See next page for additional authors

Publication details
SAGE of the developing wheat caryopsis

Shane McIntosh, Loraine Watson, Peter Bundock, Allison Crawford, Jessica White, Giovanni Cerdonio, Daniel Barbary, Lee Rooko and Robert Henry

Grain Foods CRC, Centre for Plant Conservation Genetics, Southern Cross University, PO Box 157, Lismore, NSW 2480 Australia

To understand the processes of gene expression the sequence data requires a tag-to-gene-matching process. BLASTN matching using EST clusters (TIGR, Unigene) showed: 70% of tags (counts ≥2) have a perfect match to one or more wheat ESTs; 38% of tags (counts <2) had a match to a wheat EST; combined results show that 48% of all unique tags had a match to a wheat EST.

This matching process provides information on the orientation and position of the tag. Alternative mRNA processing, heterogeneous polyadenylation and antisense transcripts have been described as a source of noncanonical tags. Examples of these processing events were evident in the abundant tags analysed and included 5 tags matching gliadins, glutenins, ns-LTPs and ribosomal proteins for which there was evidence for antisense orientation. The abundance of these transcripts being expressed may suggest a more global approach to regulatory processes in planta.

Annotation of the 500 most abundant tags spanning development highlights the array of functional groups being expressed and the functional division of the transcriptome (Figure 1). We have identified activities of: cellular proliferation/differentiation (histone H4, tubulin, cellulose synthase like protein, glycin-rich proteins, nsLTPs); the accumulation of storage proteins (gliadins, glaucins, grain softness protein, purindolines, globulins); starch energy biosynthesis (ADP-glucose pyrophosphorylase, sucrose synthase2, formate dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, GBSS, starch synthase IIa / III). The abundance of calcium-dependent protein kinases indicates their importance in signalling across development and includes Cdpk1, NDPK, PPDK and PSK2. Acquisition of a broad array of defence coincides with storage accumulation and is dominated by inhibitors of amylose activity and osmoprotection later in development.

Understanding the development of the cereal caryopsis holds the future for metabolic engineering in the interests of enhancing global food production. We have developed a Serial Analysis of Gene Expression (SAGE) data platform to investigate the developing wheat (Triticum aestivum) caryopsis. LongSAGE libraries have been constructed at five time points post anthesis (8dpa, 14dpa, 20dpa, 30dpa and 40dpa /mature dry) to coincide with key processes in caryopsis development. More than 90,000 LongSAGE tags have been sequenced generating 29,261 unique tag sequences across all five libraries. Tag abundance, generated from cumulative tag counts, provides insight into the redundancy and diversity of each library (Table 1). The abundance range of tags between libraries varies considerably reflecting the events that are occurring at that time point in development.

Table 1. Summary of LongSAGE library analysis including tag abundances across T. aestivum var. Banks caryopsis development.

<table>
<thead>
<tr>
<th>Developmental stage</th>
<th>8dpa</th>
<th>14dpa</th>
<th>20dpa</th>
<th>30dpa</th>
<th>40dpa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total tags sequenced</td>
<td>2017</td>
<td>19299</td>
<td>17706</td>
<td>17709</td>
<td>21710</td>
</tr>
<tr>
<td>Unique tags</td>
<td>10045</td>
<td>7792</td>
<td>6388</td>
<td>6444</td>
<td>10022</td>
</tr>
<tr>
<td>Abundance range #</td>
<td>1-227</td>
<td>1-718</td>
<td>1-712</td>
<td>1-637</td>
<td>1-1027</td>
</tr>
<tr>
<td>Abundance range %</td>
<td>0.0048-1.13</td>
<td>0.0052-3.72</td>
<td>0.0056-3.96</td>
<td>0.0056-3.59</td>
<td>0.0048-4.73</td>
</tr>
<tr>
<td>Tags occurring ≥10 times</td>
<td>28%</td>
<td>43%</td>
<td>50%</td>
<td>49%</td>
<td>36%</td>
</tr>
<tr>
<td>Gene count</td>
<td>236</td>
<td>157</td>
<td>145</td>
<td>156</td>
<td>191</td>
</tr>
<tr>
<td>Tags occurring once</td>
<td>38%</td>
<td>31%</td>
<td>26%</td>
<td>28%</td>
<td>36%</td>
</tr>
<tr>
<td>Gene count</td>
<td>7579</td>
<td>6658</td>
<td>4892</td>
<td>4985</td>
<td>7298</td>
</tr>
</tbody>
</table>

To understand the processes of gene expression the sequence data requires a tag-to-gene-matching process. BLASTN matching using EST clusters (TIGR, Unigene) showed: 70% of tags (counts ≥2) have a perfect match to one or more wheat ESTs; 38% of tags (counts <2) had a match to a wheat EST; combined results show that 48% of all unique tags had a match to a wheat EST.

This matching process provides information on the orientation and position of the tag. Alternative mRNA processing, heterogeneous polyadenylation and antisense transcripts have been described as a source of noncanonical tags. Examples of these processing events were evident in the abundant tags analysed and included 5 tags matching gliadins, glutenins, ns-LTPs and ribosomal proteins for which there was evidence for antisense orientation. The abundance of these transcripts being expressed may suggest a more global approach to regulatory processes in planta.

It is evident that a complex array of genes are being up- and down-regulated across development. Using the quantitative feature of the SAGE data it is possible to compare expression levels of individual genes at the different time points in development. The expression levels of the 500 most abundant tags spanning seed development were quantitatively analysed using GeneSpring® software. Generation of expression profiles allows individual genes to be examined or clustered according to similarities in their profiles. Clustering of expression patterns using Self Organising Maps (SOM) is shown in Figure 2. It is these expression profiles that highlight the individual genes involved in the cellular events of caryopsis development.

This SAGE platform has also provided a resource of novel sequence and expression information including the identification of potentially useful promoter activities. Further investigations into both the abundant and low expressing transcripts will provide greater insight into wheat caryopsis development and assist in wheat improvement programs.

Figure 1. Division of transcriptional activities by ontology.

Annotated sequences were grouped according to function and the sum of relative abundances for each grouping was calculated. Abbreviations for functional groups are as follows: 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.

Figure 2. Self Organising Maps of gene expression.

Using GeneSpring® software expression profiles for the top 500 most abundant tags were generated from normalized relative abundances and displayed using log10 scale (relative abundance versus time: days post anthesis). The generation of SOMs relied solely on algorithm determined grouping by the software.

Produced by: SCU Marketing & Media