GENOMICS RESEARCH AT SASEX: PERSPECTIVES FROM A SMALL-SCALE PROGRAM

By

DEBORAH L. CARSON1,2, BARBARA I. HUCKETT1,2 and FREDERIK C. BOTHA2

1Biotechnology Department, South African Sugar Association Experiment Station (SASEX), Private Bag X02, Mount Edgecombe, 4300, South Africa
2Institute for Plant Biotechnology, University of Stellenbosch, Private Bag XI, Matieland, 7602, South Africa

Abstract

At SASEX, a genomics-based approach towards gene discovery and expression analysis that is suitable for application at a small-scale level has been developed. Expressed sequence tags (ESTs) have been identified from leaf roll and maturing culm total cDNA libraries as well as from immature and maturing culm subtracted cDNA libraries (reciprocal subtractive hybridisation). To identify differentially expressed genes, random cDNA clones from both total and subtracted cDNA libraries were screened by reverse Northern hybridisation analysis using filter-based macroarrays. Probing with mRNA from leaf, leaf roll, immature and maturing culm identified 85 differentially expressed clones from the total cDNA libraries (8.5% of total), none of which were tissue-specific. In contrast, 130 differentially expressed clones were identified from the subtracted cDNA libraries (33% of total) and comprised tissue-specific and tissue-preferential expressed sequences. Combining cDNA subtraction with macroarray screening is an effective strategy to identify and analyse genes suitable for small research programs.

Establishment of expressed sequence tags for leaf roll and maturing culm

The use of random, single-pass DNA sequencing of cDNA to obtain expressed sequence tags (ESTs) has made a powerful impact on genetic analysis in a wide range of organisms. ESTs were first developed by Adams et al. (1991) and since the creation in 1993 of the public dbEST database, more than 7 million cDNA sequences have been made available, with many more sequences contained in private databases. The popularity of EST analysis as a tool for gene identification is derived largely from the ease of data generation, the fact that it requires no prior knowledge of the genome and is independent of its complexity. A small-scale EST program was initiated at SASEX in 1995 to evaluate whether the identification of ESTs would provide helpful information about the types of genes being expressed in various sugarcane tissues. Furthermore, we wished to establish whether the ESTs would be a source of appropriate candidate genes for promoter isolation, genetic marker development or as potential transgenes.

The identification of ESTs associated with different developmental stages was chosen as a model system. Altogether 500 ESTs were identified, 250 each from leaf roll and maturing culm total cDNA libraries. The sequences of these ESTs were submitted to the GenBank dbEST database between 1996 and 1998 and represent the first sugarcane ESTs to be released into the public domain.

Results indicated that EST analysis was a simple and effective way of identifying expressed sequences in sugarcane but could only provide preliminary and limited information about the types of genes being expressed (Carson and Botha, 2000). For both these developmental stages, many of the ESTs were homologous to genes associated with a wide variety of general metabolic processes.

While small differences in the distribution of ESTs between the various functional categories were observed, the only major distinction between ESTs identified in the leaf roll and maturing culm was a 30% increase in stress-response ESTs in the latter. However, with a limited sample number of 500 ESTs it was not possible to interpret the differences as representative changes in the pattern of gene expression between the leaf roll and maturing culm.

To utilise putative identities of ESTs as a tool for gene expression profiling, it is necessary to generate comparative data sets of several thousand ESTs that are sufficiently representative of the metabolic state of each particular source tissue or organ. This approach, commonly referred to as ‘electronic northern’ analysis, has become a popular system for determining steady state mRNA levels in a variety of tissues. As SASEX does not have the resources to develop large representative EST databases, it was more appropriate that further research focussed on analysing the expression patterns of individual ESTs to identify those sequences that are differentially regulated.

Determining the expression patterns of individual ESTs

Techniques designed to monitor the expression of large numbers of genes have evolved in parallel with the expansion of gene sequence data. Hybridisation of

KEYWORDS: ESTs, Subtractive Hybridisation, cDNA Macroarrays, Differential Gene Expression.
transcripts to arrays of cDNA molecules immobilised on nylon filter supports was first reported in the early 1990s (Dunne et al., 1992).

This was followed by the development of the more sophisticated and sensitive microarray glass and silicon supports (Schena et al., 1995, Wodicka et al., 1997 and Ruan et al., 1998 for some of the earliest reports). Although microarrays are becoming an increasingly popular tool for the analysis of differential gene expression, the technology is extremely costly and not readily available to smaller programs. We therefore used filter-based cDNA macroarrays containing random sugarcane cDNA clones and examined whether serial hybridisation analysis of the macroarrays could reliably identify developmental-stage specific genes.

In an initial investigation, cDNA macroarrays were prepared with 1000 cDNA clones derived from leaf roll and maturing culm total cDNA libraries and the macroarrays hybridised to total cDNA probes prepared from leaf, leaf roll, immature culm and maturing culm. Discrimination between different levels of hybridisation signal intensity for individual clones was achieved and these differences confirmed as significant by signal quantification using densitometry.

Hybridisation analysis indicated that the majority of clones were constitutively expressed but 85 cDNA fragments were identified that exhibited differential levels and patterns of expression in the four tissue-types tested. This latter subset of clones exhibited sequence homology to a variety of genes including those associated with meristematic growth, various regulatory processes and stress responses. However, none of the differentially expressed cDNA clones were exclusively expressed in only one tissue-type. These results suggested that a limited selection of 1000 clones randomly picked from total cDNA libraries might not be sufficiently representative to include tissue-specific expressed sequences.

Identification of differentially expressed sequences is facilitated by cDNA subtraction

The use of subtracted cDNA libraries was evaluated as a suitable mechanism for improving the complement of differentially expressed genes in the sample set for the identification of culm developmental-stage specific expressed sequences. Reciprocal subtractive hybridisation between immature and maturing culm tissues generated two subtracted cDNA libraries.

Expression analysis to assess the success of the subtractions and to identify the differentially expressed clones was performed by probing a macroarray containing 400 randomly selected subtracted cDNA clones (200 from each library) with total cDNA probes prepared from immature and maturing culm.

Results indicated that 37% of immature culm and 31% of maturing culm clones tested exhibited tissue-preferential expression. As only a few (400) clones were analysed from the subtracted libraries, the high proportion of ESTs displaying tissue-preferential expression could not, however, be considered as representative. Nonetheless, this represents an almost fourfold increase over the proportion observed for the total cDNA libraries.

Northern blot analysis of selected clones confirmed the presence of culm-specific and culm-preferential expressed sequences (Table 1), indicating that subtractive hybridisation had successfully enriched the libraries. Putative identities assigned to the differentially expressed sequences indicated homology to genes predominantly associated with regulatory processes, stress responses, cell wall metabolism and carbohydrate metabolism. The putative identities revealed that, in all cases, the specific differentially expressed ESTs detected in the immature and maturing culm were distinct from each other.

<table>
<thead>
<tr>
<th>L, LR</th>
<th>lnt2</th>
<th>lnt7</th>
<th>E value</th>
<th>Putative identity</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 x 10^-8</td>
<td>Calmodulin-related protein 2, touch-induced</td>
<td>Int2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 x 10^-9</td>
<td>UDP-glucose 6-dehydrogenase</td>
<td>Int2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 x 10^-33</td>
<td>Putative senescence-associated protein</td>
<td>Int2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 x 10^-29</td>
<td>Latex-abundant protein</td>
<td>Int2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 x 10^-20</td>
<td>Jacalin</td>
<td>Int7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.004</td>
<td>GOS9 protein</td>
<td>Int7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion

The isolation of genetic sequences exhibiting specific expression patterns is an important first step towards the isolation of promoters or potential transgenes. This study demonstrates that differentially expressed genes can be identified using simple filter-based cDNA macroarrays as a screening tool although the number and source of cDNA molecules analysed influences the success of the approach. Using a limited selection of random clones from both total and subtracted cDNA libraries may not adequately capture a representation of differentially expressed sequences. Subtracted cDNA libraries however offered an enriched sample set that, combined with the efficiency of macroarray screening, presented a very practical way of identifying tissue-specific and tissue-preferential expressed sequences.

The implementation of this approach has resulted in a resource of candidate developmentally regulated genes for the sugarcane culm that will be characterised and evaluated for use in genetic engineering and marker programs. This procedure can also be applied to other gene regulation studies such as differential expression in response to biotic and abiotic challenges. Research of this nature has already been initiated at.
SASEX, one example being the identification of genes specifically expressed in root tips in response to an aluminium challenge.

Acknowledgments

The critical input of Derek Watt during manuscript preparation is gratefully acknowledged.

REFERENCES


