Transcriptional profiling in *Brassica napus* seed using Serial Analysis of Gene Expression (SAGE)

C. OBERMEIER, B. HOSSEINI and R. SNOWDON

Serial analysis of gene expression (SAGE) is a high-throughput sequencing-based genomic technique that allows identification and quantification of tissue-specific gene expression based on identification of short (10-26 bp) tags derived from expressed poly A+ transcripts. The procedure involves PCR-based amplification and ligation of transcript-derived tags to form high molecular weight concatemers for cost-effective high-throughput cloning, sequencing and whole transcriptome data analysis. SAGE is commonly used in animal genomics, but to date has been used little in plant genome analysis. The present study aims to adapt the SAGE technique for analysis of global gene expression in the complex polyploid *Brassica napus* genome. Of particular interest is the identification of genes that are differentially expressed during seed development of *B. napus* and might be associated with synthesis of commercially valuable seed compounds. Time points in seed development were identified in the cultivar 'Express' where contrasting patterns of fatty acid metabolite patterns were present. Total RNAs extracted from seeds at two developmental stages were used for SAGE library production based on the Robust-LongSAGE protocol, a modification of the original protocol that generates 20-21 bp tags. This protocol was applied to enable efficient cloning and transcript identification from the large *Brassica napus* genome. To date 53,319 and 14,978 tags were obtained from two libraries produced from seeds harvested at 23 and 35 days after pollination (DAP), respectively. Within these two libraries 77% and 81% of tags were found to be singletons, indicating that the *B. napus* seed transcriptome is larger than the current sampling size. To eliminate potential sequencing errors only tags detected two or more times were considered reliable and included in further analysis. From 7,866 different tags sampled twice or more 268 (4%) were differentially expressed in seeds between 23 and 35 DAP at $P<0.01$. From these differentially expressed tags 87% were successfully matched to expressed sequence tags (ESTs) from public data bases and 75% were matched to ESTs produced from Brassica seeds, suggesting a good coverage of the Brassica seed transcriptome in public databases. However, a limited number of these Brassica seed-specific EST hits were found to be linked to well annotated genes, resulting in only 27% of differentially expressed tags matching to plant genes from the EMBL database. This indicates that the availability and annotation quality of Brassica EST and genomic data is currently a major factor limiting the efficiency of tag-to-gene matching in Brassica. However, tag-to-gene efficiency ratios using EST and genomic data also indicates that SAGE provides a tool for quantification and identification of previously undescribed Brassica genes. Brassica genes up-regulated at 35 DAP and identified from the EMBL database include genes involved in storage protein accumulation, fatty acid and protein metabolism, photosynthesis, development and secondary compound metabolism. This is consistent with previous seed expression profiling studies using EST sequencing and microarray hybridization technology.

Autoren: Dr. Christian OBERMEIER, Bashir HOSSEINI and Dr. Rod SNOWDON, Justus Liebig University, Department of Plant Breeding, Research Centre for BioSystems, Land Use and Nutrition, Heinrich-Buff-Ring 26-32, D-35392 GIESSEN