Digital gene expression analysis during seedling development in winter oilseed rape

Bertha Salazar, Aysha Kiran and Rod Snowdon
Department of Plant Breeding, Justus Liebig University, Heinrich-Buff-Ring 26-31, 35392 Giessen, Germany

Introduction

Seedling vigour is an important trait in winter oilseed rape (*Brassica napus*) due to its influence on seedling establishment before winter and the consequent effects on yield and yield stability. The aim of this work is to investigate global gene expression during seedling development by eQTL analysis based on digital gene expression (DGE) profiles. Illumina sequencing of short 3'-EST tag sequences is a powerful alternative to conventional microarray expression analysis, particularly for accurate quantification of low-abundance transcripts and potential identification of unknown genes. We are performing Illumina EST tag sequencing on seedling cDNA libraries from a DH population segregating for seedling developmental traits. Digital expression libraries are being generated from the cross parents 'Express 617' and 'V8', their F1 and 96 ExV8-DH lines that show maximal phenotypic diversity during seedling development. Seeds were sown under controlled conditions in Jacobsen germination vessels before mRNA extraction of whole seedlings at 8 and 12 days after sowing. EST tag libraries are generated using a modified LongSAGE protocol involving restriction of immobilised cDNA fragments with *DpnII* and subsequent ligation of Illumina sequencing oligos. Genotype barcoding enables 8-fold multiplexing for the tag sequencing. The results of the digital gene expression study will also be combined with detailed hormone profiles and developmental phenotype data in a systems-based approach to identify potential regulatory genes involved in seedling development.

The idea that variation of transcript levels could be considered as a quantitative trait in a mapping population was first introduced by Jansen and Nap (2001), who termed this new approach as expression quantitative trait locus (eQTL) analysis. Their strategy was to use the genetic variation between related individuals in a segregating population to genetically map genome-wide expression profiling data. If transcript levels are measured across a population of plants, this variation in mRNA transcript abundance for each gene can be treated as a heritable trait that can be subjected to statistical genetic analyses. This approach can generate substantial additional insights into the function of complex traits (e.g. development, heterosis, yield) by interrelation of gene products and gene function from any method of expression profiling based on RNA, proteins or metabolites. In this context, digital gene expression (DGE) analysis based on next-generation sequencing technologies is becoming an important tool for unraveling the genetic architecture of complex traits in plants.

DGE profiling (see Morrisy et al. 2010) achieves high sensitivity and reproducibility for transcriptome analysis, out-performing conventional microarray analysis in detecting low-abundant transcripts or unknown genes. Short, transcript-specific tags are extracted from the 3' end of mRNA molecules and sequenced on the Illumina platform. The quantification of sequence reads from each unique EST tag gives a complex representation of the transcriptome (Hanriot et al. 2008). The basic procedure involves capturing mRNA via their poly(A)+ tails, generating the DGE tags, sequencing them on the Illumina platform, and analyzing the resulting data. The aim of our study is to apply this method to perfom transcriptome analysis of winter oilseed rape (*Brassica napus*) during seedling development. The resulting eQTL mapping will help us to identify potential regulatory loci influencing complex seedling traits.

Material and Methods

Plant material
A doubled haploid (DH) population of 250 lines from the cross Express 617 x V8 (ExV8-DH) was used for the work. Extensive phenotype and QTL data are available for seedling development, plant hormones, seed quality, germination, heterosis, seed yield and various yield related traits in the ExV8-DH population (e.g. Basunanda et al. 2010). Beside the parental lines, a total of 47 ExV8-DH lines with the highest and 47 with the lowest shoot fresh weight at two weeks after sowing, respectively, were selected for parallel transcriptome and hormone analysis.

Growth conditions and sampling
The 94 ExV8-DH lines, 'Express 617', 'V8' and their F1 were grown under controlled conditions in a climate chamber with 16h/8h, 20°C/15°C, RH 55% day/nigh conditions. Two experimental replications were performed under identical growth climate chamber conditions. Seeds were sown in Jacobsen germination vessels and seedlings were harvested at two time points, 8 and 12 days after sowing. Harvesting of whole seedlings was realised within one hour to prevent circadian clock effects during transcriptome analysis. All samples were inmediately shock-frozen in liquid nitrogen and stored at -80°C until RNA extraction.

RNA extraction and DGE-DpnII library preparation
Total RNA was isolated from pooled seedlings from the two experimental replications by using TRIzol reagent (Invitrogen) according to manufacturer's protocol. The quality of the total RNA was evaluated on 1% agarose gels and concentration was estimated with a Nanodrop 1000 spectrophotometer. DGE tag library was constructed following the protocol of Morrissy et al. (2010) for digital gene expression by tag sequencing on the Illumina Genome Analyzer (reviewed in Figure 1). Total RNA (~100 µg) was incubated with oligo-dT Dynal magnetic beads (Invitrogen) to capture the polyadenylated RNA fraction. mRNA bound to the beads was then used as a template for first and second strand cDNA synthesis. Inmobilised double stranded cDNA was digested with *DpnII* to retain cDNA fragments from the most 3' GATC to the poly(A)-tail. Subsequently, the barcoded modified Illumina GEX adapter 1 was ligated to the free 5' end of the ds-cDNA. This modification in adapter 1 included four nucleotide "barcodes" for 8-fold multiplexing of sequencing reactions. ESTs were digested with *Mme*I, which cuts 17 bp downstream from its recognition site, detaching the fragments from the beads. After dephosphorilation and phenol extraction, the GEX adapter 2 was ligated to the 3' end of the tags. PCR amplification with only 13 cycles using Phusion polymerase (Finnzymes) was performed with primers complementary to the adapter sequences to enrich the samples for desired fragments.

The resulting 98 bp fragments were purified by excision from a 12 % polyacrylamide TAE gel. Quantification and quality assessment of the libraries was performed with an Agilent 2100 Bioanalyzer. Normalised libraries were pooled into 8-plex samples and sent for Illumina GAII sequencing by Service XS, Wageningen, Netherlands.
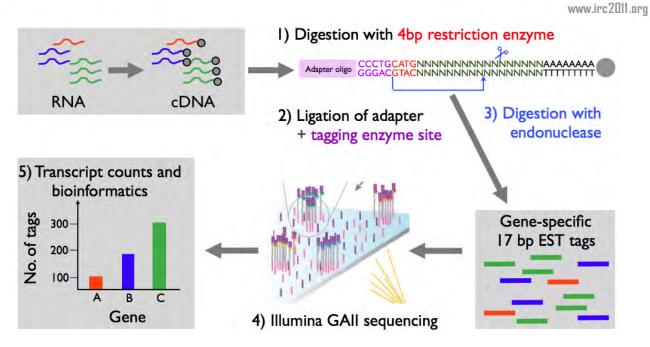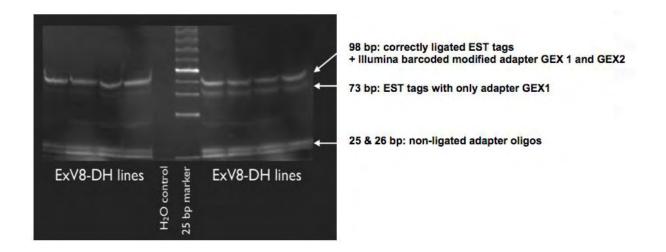
Figure 1. Overview of Illumina digital gene expression profiling

Figure 2. DGE-*DpnII* libraries loaded on 12% Polyacrilamide gels (PAGE). Staining was done by using SybrGreen DNA I (Lonza). Correctly sized fragments (98 bp) were excised carefully and incubated into elution buffer.

Results and Discussion

Sequencing results from the first sequenced libraries show the power of this method for quantification of global gene expression. On average more than 3 million EST tags per sample can be obtained (30 million sequence reads from 8 samples), giving the potential for exact quantification of expression levels based on tag counts.

Sequencing and data analysis is currently in progress for all remaining libraries from the two time points with the 96 lines of the ExV8-DH population. The data will be used for eQTL analysis of seedling development in ExV8-DH. Furthermore the expression data will also be compared with other trait data including seedling developmental data and mass spectrometry quantification of seedling hormones from the same samples. A systems analysis using weighted gene co-expression network analysis will be applied by other partners in the project to identify gene networks involved in regulation of seedling development and vigour.

Acknowledgements

References

Basunanda P., M. Radoev, W. Ecke, W. Friedt, H. C. Becker and R. J. Snowdon. 2010. Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). Theor Appl Genet 120:271-281.

Hanriot L., C. Keime, N. Gay, C. Faure, C. Dossat, P. Wincker, C. Scoté-Blachon, C. Peyron and O. Gandrillon. 2008. A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. BMC Genomics 9: 418

Jansen R. C. and Nap J. 2001. Genetical genomics: the added value from segregation. TRENDS in Genetics. Vol 17. No.7

Morrissy S., Y. Zhao, A. Delaney, J. Asano, N. Dhalla, I. Li, H. McDonald, P. Pandoh, A. Prabhu, A. Tam, M. Hirst, and M. Marra. 2010. Digital Gene Expression by Tag Sequencing on the Illumina Genome Analyzer. *Current Protocols in Human Genetics* 11.11.1-11.11.36.