CHAPTER 4

HIGH-THROUGHPUT SCREENING OF A BANANA cDNA LIBRARY USING DNA MICROARRAY ANALYSIS

Published as a component of the journal article:

High-throughput screening of suppression subtractive hybridization cDNA libraries using DNA microarray analysis. Noëlani van den Berg, Bridget G. Crampton, Ingo Hein, Paul R.J. Birch, and Dave K. Berger.

BioTechniques 37:818-824 (November 2004)

ABSTRACT

The efficient construction of a cDNA library enriched for differentially expressed transcripts is an important first step in many biological studies. This study presents a quantitative procedure for screening a banana cDNA library constructed by Suppression Subtractive Hybridisation (SSH). Following two-colour cyanin dye labelling and hybridisation of subtracted "tester" with either unsubtracted "driver" or unsubtracted "tester" cDNAs to the SSH library arrayed on glass slides, two values were calculated for each clone; an enrichment ratio 1 (ER1) and an enrichment ratio 2 (ER2). Graphical representation of ER1 and ER2 enabled the identification of clones that were likely to represent up-regulated transcripts. Normalization of each clone by the SSH process was determined from the ER2 values, thereby indicating whether clones represented rare or abundant transcripts. Differential expression of banana clones identified by this quantitative approach was verified by inverse Northern blot analysis.

INTRODUCTION

Genes that are differentially expressed in plants, and their pattern of expression, provide important information about processes such as cell differentiation, morphological and metabolic changes, and disease development. Various molecular techniques exist for studying differential gene expression. These include representational difference analysis (RDA) (Lisitsyn and Wigler, 1993), Suppression Subtractive Hybridization (SSH) (Diatchenko *et al*., 1996; 1999), differential display, differential hybridization (Liang and Pardee, 1992), subtractive library construction (Tedder *et al*., 1988), serial analysis of gene expression (SAGE) (Velculescu *et al*., 1995), and cDNA microarray analysis (Schummer *et al*., 1997). The technology of cDNA microarray hybridization has become a standard tool for genome-wide monitoring of gene expression in animal studies, and is now also starting to contribute to plant biology (Reymond, 2001). cDNA microarray analysis offers the possibility of providing a rapid, high throughput method for screening SSH cDNA libraries for differentially expressed genes (Yang *et al*., 1999).

Genes that were up-regulated in Cavendish bananas tolerant to *Fusarium oxysporum* Schechtend.:Fr. f.sp. *cubense* (E.F. Smith) Snyder & Hansen (*Foc*) were isolated by means of SSH and compiled in a cDNA library (Chapter 3). In this study, microarray analysis was used to screen the banana cDNA library to identify genes expressed in response to *Foc* infection and to obtain information about the relative abundance of these gene transcripts upon the induction of the plant defence response. The aim was to identify and discard genes that escaped the differential subtraction process and to select defence response-associated genes for sequencing and expression studies. Such a quantitative approach would allow me to determine the extent to which a transcript was induced upon pathogen infection, to determine whether a transcript was rare or abundant, and to assess the redundancy of clone sequences comprising the SSH library.

MATERIALS AND METHODS

cDNA Microarray Analysis of the SSH Library

Preparation of SSH clones and slide spotting

A banana SSH library that contains 736 clones, expressed in tolerant GCTCV-218 but not in susceptible Williams bananas upon infection with *Foc* (Chapter 3), was used for cDNA microarray analysis. Clones from the library were PCR-amplified using T7 (5'-ATTATGCTGAGTGATATCCC-3') and SP6 (5'-ATTTAGGTGACACTATAGAAT-3') primers. The amplification procedure was conducted in a 100 µl reaction volume containing the following reagents: 10 µl 10x NH_4 buffer, 5 µl $MgCl_2$, 8.0 µl dNTP mix (2.5 mM) (Roche Diagnostics, Mannheim, Germany), 3.0 µl of T7 and SP6 (10 µM) and 0.2 µl BIOTAQ™ DNA polymerase (5 U/µl) (Bioline Ltd., London), 1.0 µl of the bacterial colony grown in Luria-Bertani (LB) broth and 70.2 µl sterile distilled water (SDW). PCR conditions were as follows: Initial denaturation at 94ºC for 2 min, 30 cycles of 94ºC for 30 s, 50ºC for 30 s, 72ºC for 1 min and a final elongation step at 72ºC for 5 min. PCR products were separated by electrophoresis on a 1% agarose Electro-Fast® Stretch gel (Abgene, Epson, UK), and purified using Multiscreen® PCR Purification Plates (Millipore, Molsheim, France). They were then eluted in 45 µl SDW prior to being robotically printed onto silanized microarray slides (Amersham Biosciences, Little Chalfont, UK) using an Array Spotter Generation III (Molecular Dynamics, Sunnyvale, CA, USA). On average, 300 ηg/µl of each banana SSH fragment was spotted onto each slide, in duplicate. For global normilisation, four fragments, *uidA* (300 ρg), *luc* (30 ρg) and *bar* (3000 ρg) genes, and a fungal rDNA internal transcribed spaces (ITS) fragment (3000 ρg), were spotted to serve as controls.

Probe preparation

Probes were prepared from cDNA transcripts called the subtracted "tester" (ST), unsubtracted "tester" (UT) and unsubtracted "driver" (UD). The ST represents cDNA transcripts expressed in tolerant (GCTCV-218) but not in susceptible (Williams) bananas, while the UT and UD represent those transcripts expressed in GCTCV-218 and Williams, respectively (Chapter 3). Prior to labelling PCR products of ST, UT

and UD were purified with a Mini Elute™ PCR purification kit (Qiagen, Valencia, California, USA) and digested with *Rsa*I (Roche Diagnostics) to remove the adaptors. Samples were separated on a 1.5% low melting point (LMP) agarose gel and the desired products without the adaptors were cut out and cleaned using the QIAquick® Gel Extraction Kit (Qiagen), according to the manufacturer's instructions. Two hundred nanograms of each purified ST, UT and UD probe were transferred to two tubes. Probes in one set of tubes were then labelled with Cy™5, while the other set was labelled with Cy3 dUTP (Amersham Biosciences) using Klenow enzyme (USB, Cleveland, OH, USA) (Ramonell *et al.*, 2002). Each probe was thereafter spiked with Cy-labelled *uidA* (0.3 ηg), *luc* (0.03 ηg) and *bar* (3 ηg) genes, and a Cy-labelled fungal ITS fragment (3 ηg). After incubation at 37ºC for 20 hrs, the probes were purified using a Multiscreen® PCR Purification Plate (Millipore) and eluted in 45 µl SDW. The probes were then pooled and dried in a SPD111V vacuum centrifuge (Savant, Holbrook, New York, USA), before being resuspended in Hybridization buffer (Amersham Biosciences).

Hybridization and washing

The glass slide spotted with the SSH library was incubated in pre-treatment solution (3.5x Sodium Saline Citrate (SSC); 0.2% Sodium Dodecyl Sulfate (SDS); 1% bovine serum albumin) at 60ºC for 20 min. Microarray slides were then placed face-up in a HybUP hybridization chamber (NB Engineering, Pretoria, South Africa) and covered with 20 µl of MilliQ water (Millipore, Simplicity 185) (Millipore). Thirty- five µl of the probe was placed on each slide, allowed to spread evenly, and covered with a cover slip. The slides were then hybridised at 42ºC for 16 hrs. After hybridization, the slides were washed with 1x SSC/0.2% SDS, 0.1x SSC/0.2% SDS for 4 min at 42ºC (twice), followed by three washes in 0.1x SSC/0.2% SDS, 0.1x SSC/0.2% SDS (twice) and three washes in 0.1x SSC for 1 min at room temperature. After the final wash, the slides were briefly rinsed with distilled water and dried with high-pressure nitrogen, and then scanned with a Genepix™ 4000B scanner (Axon Instruments, Foster City, CA, USA).

cDNA Microarray screening of the SSH library

The Array-Vision™ programme (Imaging Research Inc., Ontario, Canada) was used to localize and integrate the spots on the microarray slides. Green spots visible on the computer image represented transcripts that were more abundant in the probe labelled with Cy3, while red spots represented transcripts that were more abundant in the probe labelled with Cy™5. Yellow spots represented transcripts that were present in equal amounts in probes labelled with Cy3 and Cy™5.

Prior to mathematical quantification of the computer image of each clone, background signal intensities were subtracted, the signal intensities of duplicate spots averaged, and spots with a signal/noise ratio of less than 2 were rejected. Global normalization of data for the cyanin dye effect was performed using a control gene set to calculate normalization functions c and c' for each pair of dye swap slides (Yang *et al.*, 2002). Enrichment ratios of ST:UD (ER1) and ST:UT (ER2) were calculated from experiments performed in duplicate. To determine ER1, slides hybridized with ST and UD were used, and for ER2, slides hybridised with ST and UT were used. ER1 and ER2 were calculated using the formula of Yang *et al.* (2002):

$$\tfrac{1}{2}\,[\log_2 \text{Cy3 ST/Cy5 UD} - c - (\log_2 \text{Cy3 UD/Cy5 ST} - c')]$$

UT/UD values were calculated from the ER1 and ER2 values as follows:

Because ER1-ER2 ~ $\log_2$ ST/UD – $\log_2$ ST/UT = $\log_2$ UT/UD, therefore:

UT/UD = antilog of (ER1 – ER2) in the base 2.

Enrichment ratios ER1 and ER2 provided data on the abundance of transcripts. A positive ER1 and ER2 value indicated transcripts that were enriched during subtraction relative to their levels in UD or UT, respectively. Conversely, negative ER1 and ER2 values indicated transcripts that were reduced in abundance during SSH relative to their levels in UD or UT, respectively, due to normalization.

rRNA redundancy for the library was determined by hybridising the microarray slide with a banana rRNA probe. A clone was considered to have hybridised to the rRNA probe if fluorescence was more than two standard deviations above local background

fluorescence (Leung and Cavalieri, 2003). When artefacts were visible on spots, they were flagged and the data rejected.

Inverse Northern Dot Blots

Membrane preparation

A Hybond-N^+ nylon transfer membrane (Amersham Biosciences) was pre-wetted for 10 min in distilled water, fitted into a 96-well BioRad Biodot ™ manifold (Bio-Rad Laboratories Inc., Hercules, CA, USA) and flushed with 500 µl SDW. Three hundred nanograms of each SSH clone was made up to a 100 µl with SDW, denatured in a total volume of 200 µl containing 2 M NaCl and 0.8 M NaOH and incubated at 37°C for 15 min. One hundred µl DNA of each clone was then loaded into separate wells on the membrane. The clones were fixed to the membrane by applying a vacuum of 20 inches of Hg, followed by flushing the wells with a 100 µl solution consisting of 1 M NaCl and 0.4 M NaOH. The SSH clones were arrayed in duplicate on each membrane. After loading of the clones, the membranes were washed in 2x SSC and cross-linked on an Ultraviolet Cross linker CL-508 (UVI-tec, St John's Innovation Centre, Cambridge, England) at 0.167 J for 3 min and stored at –20°C.

Probe preparation, hybridisation and detection

Banana "tester" (GCTCV-218) and "driver" (Williams) probes were prepared from 30 µg total RNA collected from roots 3 and 6 hrs after infection with *Foc*. cDNA was synthesised from total RNA and labelled with DIG-dUTP using a DIG DNA Labelling and detection Kit (Roche Diagnostics). Hybridizations (Southern, 1975) were then performed using 40 ηg/µl of either the DIG-labelled "tester" or "driver" cDNA. Membranes containing the SSH clones were first pre-hybridized in 15 ml DIG Easy Hyb (Roche Diagnostics) at 42°C for 3 to 5 hrs. For hybridization, the probes were denatured at 96°C for 10 min, placed on ice and then added to 15 ml new Dig Easy Hyb, pre-warmed to 42°C. Membranes were incubated overnight at 42°C in a Hyb Oven. They were then washed twice under low stringency conditions (2x SSC/0.1% SDS) for 15 min at room temperature, followed by two washes under high stringency (0.5x SSC/0.1% SDS) for 15 min at 68°C. Hybridization signals were detected using CDP Star as outlined by the manufacturer (Roche Diagnostics) and

exposed to X-ray film. X-ray images were captured on the VersaDoc BIO-RAD imaging system Model 3000 (Bio-Rad Laboratories, Inc.) using Quantity One 4.4.1 software (Bio-Rad Laboratories, Inc.), and exported to Array vision™ (Imaging Research Inc.). Array vision™ (Imaging Research Inc.) was then used to calculate signal density for each spot, following normalization, by comparing values of rDNA dots. Inverse Northern expression ratios were calculated by dividing normalised density measurements for each clone hybridised with "tester" probes, with measurements for the same clones hybridised with the "driver" probe.

RESULTS AND DISCUSSION

cDNA Microarray Screening of the SSH Library

Hybridized glass slides were of high quality, with very low background, high signal/noise ratios and very few artefacts. Microarray slides where the ST was labelled with Cy3 and the UD labelled with Cy™5 produced a computer image with more green spots compared to red spots (Fig.1). This indicated that the SSH library mostly contained transcripts that were up-regulated in response to *Foc*. It is, however, important to bear in mind that the computer image does not provide quantitative data and it, therefore, was necessary to present the computer images as numerical data.

Enrichment ratios ER1 and ER2 provided data on the abundance of transcripts. A positive ER1 and ER2 value indicated transcripts that were enriched during subtraction relative to their levels in UD or UT, respectively. Conversely, negative ER1 and ER2 values indicated transcripts that were reduced in abundance during SSH relative to their levels in UD or UT, respectively, due to normalization. Normalization equalizes the concentration of individual transcripts, which may be present at very different concentrations prior to normalization (Diatchenko *et al.*, 1996). This therefore has the advantage of enriching for rare transcripts in the ST samples and reducing levels of abundant transcripts.

A relative abundance of cDNAs in UT and UD was found when plotting ER1 against ER2 for individual banana clones (Fig. 2). Clones that were plotted on the diagonal

line were derived from transcripts of equal abundance in UT and UD. The majority of banana clones (63%) were plotted above the line (ER1>ER2), indicating a greater abundance of these transcripts in UT than in UD. This confirmed that SSH enriched for transcripts that are up-regulated in the "tester" in response to *Foc* infection. Clones above the diagonal line with positive ER2 values (86%) were likely to be derived from low abundance transcripts. Expression of some of these clones might be difficult to detect in Northern blots, and accurate comparisons of gene expression between "tester" and "driver" might require real-time reverse transcription PCR (RT-PCR). Several clones above the diagonal line showed negative ER2 values (14%), representing transcripts enriched relative to levels in UD and reduced in abundance during normalization relative to levels in UT. These clones represent abundant transcripts specifically up-regulated in the "tester" that should be more easily detected in Northern blot analyses.

The SSH library used in the current study was prepared from total RNA, since isolation of mRNA proved to be difficult and resulted in extremely low concentrations. When the SSH library was screened with rDNA probes, they hybridised to 28% of the banana clones, showing that a substantial number of rDNA clones escaped the subtraction process. In future, it is suggested that a technique be used to isolate sufficient amounts of mRNA from banana roots for cDNA synthesis in order to increase the selection for poly $(A)^+$-tail fragments.

Inverse Northern Blot Analysis of Selected Transcripts

Results obtained by microarray analysis were confirmed by inverse Northern blot analyses. Clones with ER1>ER2 had inverse Northern expression ratios greater or equal to one (Table 1), confirming that these clones represented transcripts that were up-regulated in the "tester" compared to the "driver" (Fig. 2). Clones where ER1<ER2 showed inverse Northern expression ratios of less than one (Table 1), confirming that they represented transcripts that were more abundant in the "driver" than the "tester", and that these clones had escaped the subtraction process. In order to relate the inverse Northern and microarray data, a ratio of UT/UD was calculated from ER1 and ER2 ratios (Table 1). Inverse Northern ratios correlated with UT/UD

ratios (i.e., clones with inverse ratios >1 had UT/UD values ≥1, whereas clones with inverse Northern ratios <1 gave UT/UD values <1) (Table 1). This means that these transcripts were more abundant in the tolerant GCTCV-218 after *Foc* infection compared to the susceptible Williams.

The banana SSH library was screened by a high-throughput DNA microarray technique, to ensure that false positives were removed before the library was sequenced. The majority of clones showed ER1>ER2. The library yielded more clones that were reduced in abundance in ST relative to UT. Many of the clones that were enriched during the SSH were of equal or greater abundance in UD than UT and therefore many of the clones would be eliminated from further analysis. The experimental set-up of the SSH (challenging both "tester" and "driver" with *Foc*) led to many possible defence genes being eliminated from the subtracted material, as they were present in both the "tester" and "driver" samples. However, the scientific question was to investigate the unique difference in gene expression in the tolerant GCTCV-218 cultivar compared to the susceptible Williams upon infection with the Fusarium wilt pathogen.

In the past, SSH libraries were screened by reverse northern blot analysis or cDNA-AFLP to identify cDNA clones that were differentially expressed in plants (Birch *et al*., 1999; Mahalingam *et al*., 2003). However, these methods were time consuming, and did not allow for the quantitative analysis of the level of induction or abundance of a transcript upon treatment. cDNA microarrays proved to be highly effective for screening SSH libraries in a rapid, high-throughput manner. The advantages of this screening method are that it provides an objective and quantitative way to identify differentially expressed genes as well as to determine the relative abundance of transcripts in the original UT samples. In contrast to membrane-based methods where different membranes are compared with each other, hybridisation to glass slide cDNA microarrays can be performed with different fluorescent tags, which allows direct comparison of the relative abundance of transcripts in ST, UT and UD. Furthermore, hybridisations are performed on a small surface area, which reduces the amount of labelled probe needed. Finally, the computerised scanning of the array provides a

high-throughput quantitative method to choose which genes to sequence and study further using Northern blot analysis, real-time RT-PCR, or custom microarrays.

REFERENCES

Birch, P.R.J., Avrova, A., Lyon, G.D., Duncan, J.M. and Toth, R.L. 1999. Isolation of potato genes that are induced during an early stage of the hypersensitive response to *Phytophthora infestans*. *Molecular Plant-Microbe Interactions* 12: 356-361.

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences of the USA* 93: 6025-6030.

Diatchenko, L., Lukyanov, S., Lau, Y.F., and Siebert, P.D. 1999. Suppression subtractive hybridization: A versatile method for identifying differentially expressed genes. *Methods in Enzymology* 303: 349-380.

Leung, Y.F. and Cavalieri, D. 2003. Fundamentals of cDNA microarray data analysis. *Trends in Genetics* 19: 649-659.

Liang, P. and Pardee, A.B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257: 967-971.

Lisitsyn, N. and Wigler, M. 1993. Cloning the differences between two complete genomes. *Science* 259: 946-951.

Mahalingam, R., Comez-Buitrago, N., Eckardt, N., Shah, A., Guevara-Garcia, P., Day, R.R. and Federoff, N.V. 2003. Characterizing the stress/defence transcriptome of *Arabidopsis*. *Genome Biology* 4: R20.1-R20.14.

Ramonell, K.M., Zhang, B., Ewing, R.M., Chen, Y., Xu, D., Stacey, G. and Somerville, S. 2002. Microarray analysis of chitin elicitation in *Arabidopsis thaliana*. *Molecular Plant Pathology* 3: 301-311.

Reymond, P. 2001. DNA microarrays and plant defence. *Plant Physiology and Biochemistry* 39: 313-321.

Schummer, M., Ng, W., Nelson, P.S., Bumgarner, R.E. and Hood, L. 1997. Inexpensive handheld device for the construction of high-density nucleic acid arrays. *BioTechniques* 23: 1087-1092.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98: 503-517.

Tedder, T.F., Strueli, M., Schlossman, S.F. and Saito, H. 1988. Isolation and structure of a cDNA encoding b1 (CD20) cell-surface antigen of human B lymphocytes. *Proceedings of the National Academy of Sciences of the USA* 85: 208-212.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. 1995. Serial analysis of gene expression. *Science* 270: 484-487.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* 30: e15.

Yang, G.P., Ross, D.T., Kuang, W.W., Brown, P.O. and Weigel, R.J. 1999. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Plant Physiology and Biochemistry* 27: 1517-1523.

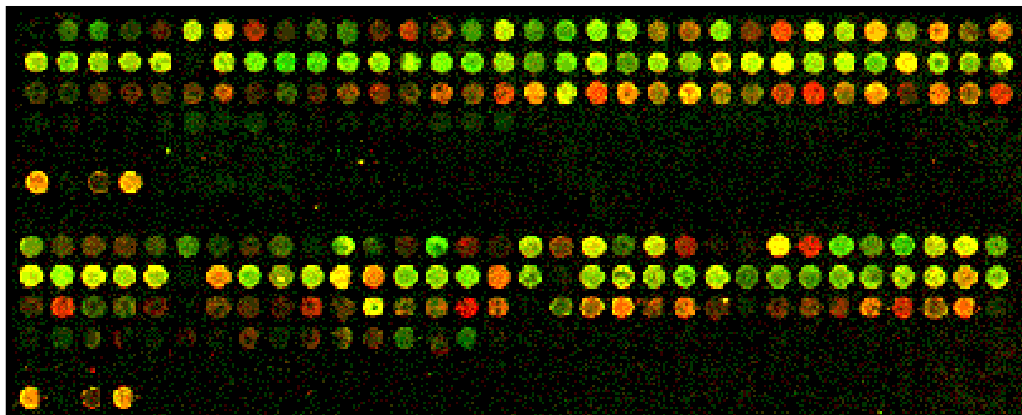

Figure 1. Microarray glass slide image of banana Suppression Subtractive Hybridisation clones spotted onto a glass slide and hybridised with subtracted "tester" (ST) (Cy3) and unsubtracted "driver" (UD) (Cy^{TM}5) probes. The Array-Vision™ programme (Imaging Research Inc., Ontario, Canada) was used to localize and integrate the spots on the microarray slides. Green spots represent transcripts that are more abundant in the ST, while red spots represent transcripts that are present in the UD. Yellow spots are present in equal amounts in both the ST and UD.

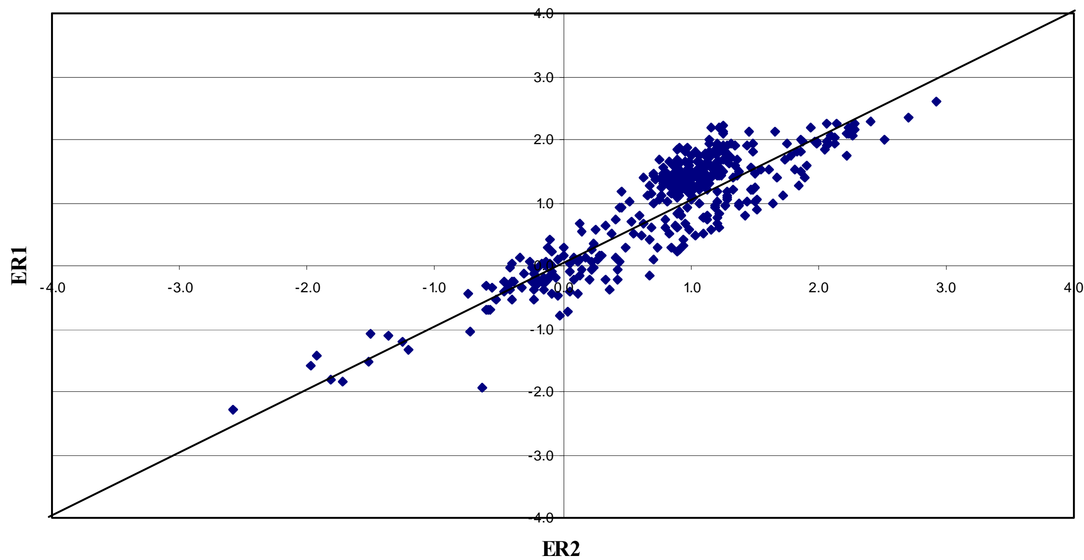

Figure 2. Suppression Subtractive Hybridisation (SSH) enrichment ratio 1 (ER1) is plotted against SSH enrichment ratio 2 (ER2). ER1 was calculated for each clone by log_2 transforming the value of the subtractive "tester" (ST) fluorescence divided by the unsubtracted "driver" fluorescence (UD). ER2 for each clone was calculated by log_2 transforming the value of the subtractive "tester" (ST) fluorescence divided by the unsubtracted "tester" (UT) fluorescence. The diagonal line indicates clones derived from transcripts of equal abundance in UD and UT (i.e. ER1=ER2). Clones that lie above the line indicate transcripts that are induced upon treatment (ER1>ER2), while those below the line indicate transcripts that have escaped the subtraction (ER1<ER2). Clones above the line with positive ER2 values represent rare transcripts, whereas clones above the line with a negative ER2 value are regarded as abundant and have been reduced in relative concentration during normalization.

Table 1. Validation of Microarray screening of selected banana Suppression Subtractive Hybridisation clones by Inverse Northern data.

Clone no.	ER1[a]	ER2[b]	UT/UD Ratio[c]	Inverse Northern Expression Ratio[d]
3-105	-0.3	-0.6	1.2	8.6
2-35	0.3	-0.1	1.3	6.9
3-169	1.0	0.7	1.2	3.1
2-28	0.6	0.3	1.2	2.1
1-77	0.8	0.6	1.2	2.1
1-158	0.7	0.6	1.1	2.0
3-7	0.6	0.3	1.2	1.7
2-21	1.8	0.9	1.9	1.5
1-149	0	-0.1	1.1	1.4
2-18	2.1	1.5	1.6	1.4
2-70	1.9	1.2	1.6	1.2
1-136	-0.1	-0.1	1.0	1.2
2-45	1.4	1.0	1.3	1.0
1-1	0.8	0.9	0.9	0.98
2-14	0.7	0.9	0.9	0.97
3-94	-0.1	0.2	0.8	0.94
2-134	1.5	1.6	0.9	0.1
3-156	0.6	0.8	0.9	0.1

SSH, suppression subtractive hybridisation; ST, subtracted "tester"; UD, unsubtracted "driver"; UT, unsubtracted "tester".

[a,b] Enrichment ratio 1 and 2 (ER1 and ER2) were calculated from microarray screening as $\log_2(ST/UD)$ and $\log_2(ST/UT)$, respectively.

[c] UT/UD = antilog of (ER1-ER2) in the base 2.

[d] The inverse Northern expression ratio was calculated as follows: density of "tester"/"driver" samples after normalization of the data using an rDNA clone.