Macroarray analysis of gene transcription during sucrose accumulation in sugar beet (*Beta vulgaris* L.) root: identification of developmental and metabolism related candidate genes

Inaugural –Dissertation

zur

Erlangung des Doktorgrades
der Mathematisch-Naturwissenschftlichen Fakultät
der Universität zu Köln

vorgelegt von

Diana Bellin

aus Verona

2005

Die vorliegende arbeit wurde am Max Planck Institut für Züchtungsforschung in Köln in der Abteilung für Pflanzenzüchtung und Ertragsphysiologie (Direktor: Prof. F.Salamini) angefertig

Berichterstatter: Prof . F. Salamini
Prof. UI Flugge

Tag der mündlichen Prüfung: 16.02.2004

ABBREVIATIONS

%	percentage
ADIS	automatic DNA isolation and Sequencing
AFLP	amplified fragment length polymorphism
AN	amino nitrogen content
A_x	absorbance at a given "x" wavelength
BKG	backgroung
bp	base pair
BY	beet yield
cDNA AFLP	complementary deoxyribonucleic acid amplified fragment polymorphism
cDNA	complementary deoxyribonucleic acid
CG	candidate gene
Ci	Curie
cM	centimorgan
Cpm	counts per minute
CSY	corrected sugar yield
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DD	differential display
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphates
DTT	1,4-dithiothreitol
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
GA	giberellic acid
GDD	growing degree day
gr	gram
h	hour

HEPES	N-2-hydroxyethylpiperazine-N`-2-ethane sulfonic acid
IB	ion balance
K	potassium
Kb	kilobasepair
LD	linkage disequilibrium
LUT	look-up table
Mbp	Megabasepair
MDC	molecular dynamics counts
min	minutes
MIP	membrane intrinsic protein
MIPS	Münich Information Center for Protein Sequences
mM	millimolar
MOPS	3-(N-Morpholino) propane sulfonic acid
MPIZ	Max Planck Institute for Plant Breeding
mRNA	messenger ribonucleic acid
NA	sodium
PCR	polymerase chain reaction
PIPES	1,4-Piperazine-N,N`-bis(2-ethane sulfonic acid)
PR	pathogenesis related
QTL	quantitative trait locus
RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RGA	resistance gene analogues
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RT	reverse transcriptase
RZPD	deutsches Resourcezentrum für Genomforschung, Berlin
SAGE	serial analysis of gene expression
SBSS	sugar beet sucrose synthase
SC	sugar content
SDS	sodium dodecyl sulfate
sec	second
SNP	single nucleotide polymorphism

SPS	sucrose phosphate synthase
SSCP	single Strand conformational polymorphism
SY	sucrose yield
TIP	tonoplast intrinsic protein
Tris	tris-(hydroxymethyl)-aminomethane
tRNA	transfer ribonucleic acid
x g	gravity constant

1 – INTRODUCTION

Sugar beet (*Beta vulgaris* L.) is the major sugar crop grown in temperate regions of the world. Only within the last 200 years it became a commercial crop and an object for breeding. In 1747, a German chemist, Andreas Sigismund Marggraf, performed an alcoholic extraction from the macerated roots of fodder beet and purified a substance that had the same properties as sugar from sugar cane when crystallized. Some 50 years later, a former student of Marggraf, Franz Carl Achard, discovered that it was possible to extract sugar from white skinned Silesian beets. These beets are believed to have contained about 6% sugar, from which Achard recovered about half. He began mass selection for increased sugar content in 1786. The first sugar factory was erected in Cunern in Silesia in 1802. It was demonstrated that sugar could be economically produced from beets. Since that time, repeated selection and breeding raised the sugar content of sugar beet to the present value of 18-20% per fresh weight.

Today, sugar beet is grown in 48 countries throughout the five continents of the world. It is used as crop to extract sugar, a carbohydrate source that contributes significantly to the flavour, aroma, texture, color and body of a variety of foods. In addition, sugar factories produce dry sugar beet pulp to feed cattle and sheep, and molasses, for production of yeast, chemicals, pharmaceuticals, as well as mixed cattle feed (Cooke and Scott, 1993). The major producers of sugar from sugar beet are Europe and USA, with Europe responsible for the 45-50% of the world production of sugar beet sugar. Concerning the world sucrose production sugar beet has a share of 30% whereas the other 70% are produced from sugar cane.

With this background it is clear that understanding the genetic factors underlying sucrose accumulation is not only of scientific interest, but also of economic importance to improve the amount of sugar that can be accumulated in the taproot of sugar beets.

Breeding strategies like mass selection have been extensively applied to this aim, but to study single genetic factors at the molecular level fine-mapping experiments should be performed which is obviously a highly time-consuming process. Together with this the complexity of the sugar beet genome of 758 Mbp make whole genome approaches

quite unfeasible und costly. Therefore to develop a targeted approach would be obviously a more attractive strategy

Today the availability of novel technologies in plant molecular biology provides a new chance to this aim. Genes putatively affecting a certain trait can be identified using molecular techniques and in a second step they can be validated by other approaches. This can significantly accelerate the discovery of underlying genetic factors responsible for the sucrose storage in sugar beet roots.

1.1 - SUGAR BEET

Sugar beet has evolved a special root morphology and physiology that allow the accumulation of sucrose. Therefore a description of the plant and in particular of the root morphology and physiology is given.

1.1.1 – The plant

Sugar beet (*Beta vulgaris* L.) belongs to the family of Chenopodiaceae. It is a highly variable species containing four main groups of agricultural significance: leaf beets (such as Swiss chard), garden beets (such as beetroot), fodder beets and sugar beets.

Sugar beet is a biennial species. The sugar beet plants develop a large thicken taproot in the first year, the vegetative season, and a seed stalk the second year, the reproductive season. Sugar beets are sown in spring and the beets are harvested in the autumn of the same year. For seed production, however, an over-wintering period of cold temperature of 4-7°C (vernalization) is required to initiate bolting of the shoot in the next season (Smith, 1987)

During the first year the sugar beet plant develops a rosette of glabrous, dark green, glossy leaves with prominent midribs and petioles. Leaf production continues throughout the first season, while the root is expanding and accumulating sucrose.

During the second year, the shoot produces a flowering stalk which is approximately 1.2-1.8 meters tall.

Sugar beet produces flowers consisting of a tricarpelate pistil, surrounded by five stamens and a perianth of five narrow sepals, but no petals. The flowers are small and sessile. They occur singly (monogerm beet seed) or in clusters (multigerm beet seed) (Cooke and Scott, 1993).
European sugar beet varieties are triploids, hybrids of diploid, male sterile plants used as female parents,, and tetraploid pollinators. Hybridis are usually more vigorous and therefore they are used in agriculture. For this reason in this study will be employed hybrid plants as well.

1.1.2 – Root morphology

The storage organ of the sugar beet plants, called beet, is only 90% root-derived with the upper 10% (the crown) contributed by the hypocotyl (Elliott and Weston, 1993). It undergoes extensive thickening caused by division, enlargement and differentiation of the derivatives of concentric secondary cambia. Artschwager (1926) provided a detailed description of the sugar beet root anatomy. The primary cambium appears very soon after germination and is followed by the successive initiation of 12 or more anomalous cambia, each external to the previous one (Figure 1.1). The bundles contain xylem towards the inside and phloem towards the outside. Each vascular zone is separated from the next by a zone of parenchymatous cells that is considered to be derived from proliferating phloem and ray parenchyma. While new cambia form centrifugally, cell division and differentiation of vascular elements and parenchymatic tissue occur in the already existing layers. In the horizontal plane, a mature beet is composed of a central core of xylem and phloem, surrounded by concentric rings of vascular tissue separated by broad bands of large celled parenchyma. Additional unexpanded rings are found near the periphery. In fact, a total of 12 or more cambial rings are formed, but expansion of the storage root involves significant contribution only from about half of the cambia. The largest expansion occurs in rings 1 and 2, while rings 3 to 8 show progressively less activity.
During vegetative growth, the sucrose content is increasing from around 4 up to 20% of root fresh weight (reviewed in Scott and Jaggard, 1993).

Sucrose enters the root via the phloem and is stored in the vacuoles of parenchyma cells both in the vascular zones and in the parenchymatous zone itself. Numerous lateral connections link adjacent rings of vascular tissue and allow the distribution of photosynthate from any leaf to any root zone. The greatest sucrose concentration was found in the cells of the vascular zone.

A gradient of sucrose concentration is found also along the beet. Sucrose concentration is highest in the centre of the section of the root with the largest diameter and it falls off above, below, and outside this area.

1.1.3 – Physiology of sucrose accumulation

Sucrose accumulation is dependent on the amount of assimilates generated by photosynthesis, on their efficient transport and storage, but also on processes competing for carbohydrates in the sink tissue.

Sucrose is synthesized in leaves. The triose phosphate products of the Calvin Cycle are used to produce either starch in the chloroplast or sucrose in the cytosol. Synthesis pathways for starch and sucrose have a number of steps in common, but usually isozymes are unique to the appropriate cellular compartment. The two compartments communicate with one another via the phosphate/triose phosphate translocator, a transport protein in the chloroplast envelope membrane that catalyzes the movement of orthophosphate and triose phosphate in opposite directions between the compartments. Low cytosolic orthophosphate concentration limits the export of triose phosphate through the translocator. In this condition triose phosphate is used for starch synthesis. Conversely an abundance of orthophosphate in the cytosol inhibits starch synthesis and promotes the export of triose phosphate into the cytosol, where it is converted to sucrose. Cytosolic sucrose biosynthesis is strongly regulated by fructose 2,6-bisphosphate in a mixture of feedforward and feedback mechanisms (Huber *et al*., 1986, Stitt and Quick, 1989) to allow a fine regulation of assimilate partitioning.

In sugar beet, a temporary storage of sucrose was also observed in leaves: 40% of leaf sucrose is in a storage pool and not being in transport (Fondy and Gaiger, 1982). So it

can be postulated that other activities rather than sucrose biosynthesis limit allocation (Elliott and Weston, 1993).

Assimilate transport involves three steps: lateral transport from the mesophyll to the conducting tissue, translocation in the sieve tubes, and lateral transport from the sieve tubes to the receiving cells. Lateral transport can take the symplastic or apoplastic route. In many plants sucrose is loaded into the phloem by a proton-sucrose symporter that links active transport to the proton motive force (PMF) across the plasma membrane of plant cells. Also in sugar beet a sucrose symporter, *BvSUT1,* was found expressed specifically in phloem companion cell (Chiou and Bush, 1996). Recent studies on this transporter in sugar beet provided the first evidence for a sucrose dependent signal transduction pathway, mediated at least in part by phosphorylation that regulates the sucrose symporter (Vaughn *et al*., 2002, Ransom-Hodgkins *et al*., 2003). From this observation Chiou and Bush (1998) hypothesized that the symporter could act as a sucrose sensor to modulate transport activity and assimilate partitioning. High sucrose levels in the phloem, which can result from decreased sink demand, would down-regulate symporter activity. As a result of decreased phloem loading, carbohydrates would accumulate in the surrounding mesopyll resulting in a concomitant down-regulation of sucrose biosynthesis. The active loading of sucrose into the phloem mediates the long-distance transport in the phloem cells of the vascular system by a positive hydrostatic pressure difference between the source and sink tissues that drives mass flow of solution. Positive pressure in the leaf phloem results from the hyperaccumulation of the osmotically active solute sucrose.

At the other end of the long distance transport, the dissaccharide is hydrolyzed either in the apoplast of taproot phloem elements or in the cytoplasm of adjacent cells. According to the model proposed by Fieuw and Willenbrink, (1990) and reported schematically in Figure 1.2, a cell wall bound invertase cleaves sucrose into glucose and fructose in the apoplast of the vascular beet tissue, preventing a backflow of sucrose into the sieve tubes. Glucose enters the cytoplasm of phloem parenchyma cells via a H^+/glucose symport mechanism. A second carrier system probably facilitates fructose influx, whereas a low affinity uptake system for sucrose may only be important at periods of high assimilate delivery.

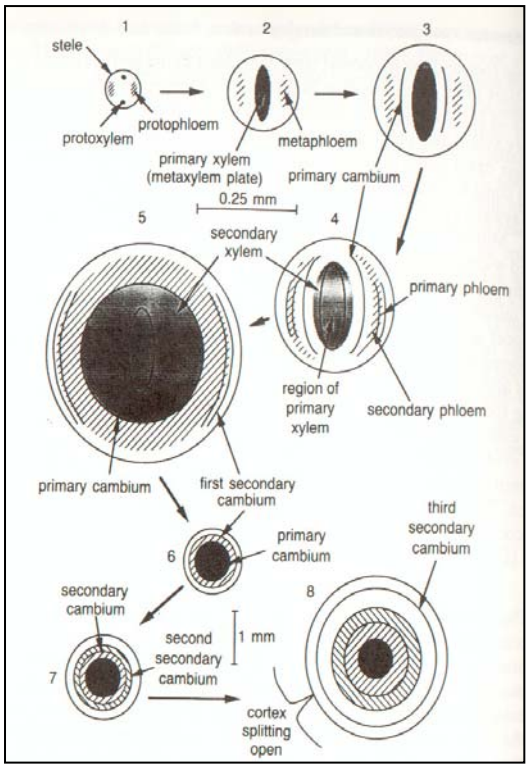


Figure 1.1 The sequence of development of the primary and secondary cambial rings in sugar beet during the early stages of growth. As the primary cambia develop it begins to enclose the secondary xylem around the primary xylem. Then new cambia differentiate centrifugally.

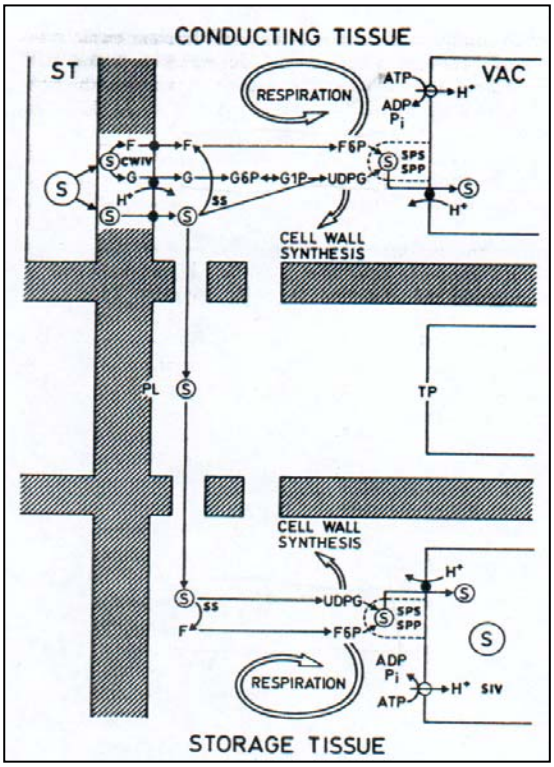


Figure 1.2 - Working hypothesis for the pathway of sucrose accumulation in sugar beet storage roots according to Fiew and Willenbrink, 1990. S=sucrose, F=fructose, G=glucose, PL=plasmalemma, TP=tonoplast, SPP=sucrose phosphate phosphatase, ST=sieve tubes, SIV=soluble acid invertase, CWIV=cell wall bound acid invertase

Inside the cell sucrose is cleaved by sucrose synthase, the activity of which is a prerequisite for both anabolic and catabolic pathways. This may be significant for storage cells which undergo considerable cell enlargement. According to estimates about 50% of the translocated carbohydrates are used for respiration and cell wall synthesis. The activity of sucrose synthase has been correlated with sugar import (Sung *et al*., 1989), with cell wall synthesis (Chourey *et al*., 1991) and with sink strength in storage systems (Zrenner *et al*., 1995). It was believed to be a soluble cytosolic enzyme, but recently partial activity was detected in close association with the plasmalemma (Amor *et al*., 1995), actin filaments (Winter *et al*., 1998), and more recently with the tonoplast (Etxeberria and Gonzalez, 2003). In the taproot of sugar beet, sucrose synthase activity may also contribute to the generation of ATP needed to maintain the membrane potential and the proton gradient at the tonoplast by the action of a vacuolar-H^+-ATPase. This in turn may then promote the translocation of sucrose into the vacuoles of the storage root cells by an H^+-sucrose antiport system. An

important enzyme for the resynthesis of sucrose from hexoses is sucrose phosphate synthase. This process occurs in conducting as well as in storage tissue. The importance of this enzyme for sucrose concentration was shown in fruits of various species and the activity of this enzyme is known to be highly regulated at post transcriptional level (Lunn and MacRae, 2003). The possibility of an association of this enzyme with the tonoplast is also considered on the basis of experimental data by Fieuw and Willenbrink (1993). Finally the sucrose is transferred to the vacuole by the above mentioned H^+-sucrose antiport system and accumulates in this subcellular compartment.

1.2 - THE CANDIDATE GENE APPROACH AS TOOL FOR A GENETIC ANALYSIS OF COMPLEX TRAITS

The sucrose accumulation process in sugar beet is a complex process involving a network of pathways each consisting of several highly regulated steps. According to the reported description only few of the enzymes involved in this process have been identified and many steps are just based on hypotheses and models. Additionally, numerous approaches taken by many laboratories to control carbon fluxes in other non-model plant species through modifying individual enzymatic steps have been largely unsuccessful (Herbers and Sonnewald, 1998). This, together with the available literature on source-sink regulation, indicates that plants may display an enormous and underestimated metabolic flexibility and crosstalk between different signal transduction pathways. The absence of a storage root in *Arabidopsis thaliana* does not allow to benefit from the molecular tools and the knowledge established for this model plant to clarify these steps. To analyze complex polygenic traits like sucrose accumulation, the genome of an organism can be scanned for regions, which affect the trait, in a QTL (quantitative trait locus) analysis (Gelderman, 1975). To focus such an analysis the candidate gene (CG) approach was developed. This strategy involves the proposal of CGs on the basis of their functions, but other criteria, like gene expression levels are also selected for identifying yet unknown genes playing a role in the sucrose storage process in sugar beet roots. Genetic mapping of CGs and QTL studies are the next steps to assess their importance for the complex trait.

1.2.1 – QTL analysis and the theory of the candidate gene approach

In a given population the allelic constitution at QTL loci and their interaction with the environment determine the phenotype of single plants.

QTLs can be identified by closely linked molecular markers in a population. However, depending on the density of the genetic map QTL positions can be quite imprecise because the associated confidence interval might cover several centimorgans which could contain hundred to several thousand genes. The candidate gene approach allows to select genes putatively involved in the trait on the basis of their function, their position on the genetic map or other factors like expression level of the genes. In this way the amount of genes putatively underlying the trait is reduced.

Candidate genes (CGs) are defined as genes with molecular polymorphisms genetically linked to a QTL (Pflieger *et al*., 2001). The working hypothesis assumes that a molecular polymorphism within the CG is related to the phenotypic variation.

The candidate gene approach consists of three chronological steps. First, CGs are proposed based on molecular and physiological studies (in this case they are functional CGs), based on linkage data of the locus being characterized (positional CGs) or based on other selection criteria like expression levels under a particular condition or at a specific developmental stage. Second, a molecular polymorphism must be revealed to localize the candidate gene on a genetic linkage map. Genotypic and phenotypic data of a segregating population are then compared in a QTL analyis to assess linkage between the CGs and the trait being characterized. The polymorphism can also be used to calculate the statistical correlation between the CG polymorphism and the phenotypic variation in a set of genealogically unrelated individuals (see also paragraph 1.2.5). This additional strategy contributes to confirm the candidate gene. Third, if co-segregation between CG and QTL and/or statistical correlation have been found complementary experiments must be conducted at the physiological and biochemical level to prove the involvement of the CG in the trait variation. This is the validation step (Pflieger *et al*., 2001).

The candidate gene approach has been successfully used in human and animal genetics (Rothschild and Soller, 1997), and since the 1990s in plant genetics. As an example, in maize this approach has been successful in assessing the role of genes

encoding key enzymes in carbohydrate metabolism during the early growth of the plant (Causse *et al*., 1995).

1.2.2 – Genetic maps and QTL studies in sugar beet

Molecular genetic maps based on anonymous restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP) markers are available for sugar beet (Barzen *et al*., 1995, Schumacher *et al*., 1997). These maps cover 700 cM of the 758-Mbp genome of the species (Arumuganathan and Earle, 1991) organized in nine linkage groups.

Recently, the first functional map of sugar beet was developed based on a new segregating population (population 618, Schneider *et al*., 1999). This included 75 expressed genes related to sugar metabolism and transport. Additional 18 known RFLP markers were included in this map to provide a link to previous maps, and 99 AFLPs were added to improve map density (Schneider *et al*., 2002).

Measurable traits related to sugar production are: sugar content (SC), beet yield (BY), content of amino nitrogen (AN), potassium (K) and sodium (NA). The last three can be used to calculate the ion balance (IB), a complex trait which influences the industrial purification process of sucrose and therefore contributes to sugar yield determination. The most important complex trait is corrected sugar yield (CSY). It is dependent on beet yield and sugar content.

With respect to sucrose content, a QTL analysis for yield data of sugar beet was produced by Weber *et al*. (1999, 2000) based on two segregating populations grown in different environments. Many QTLs were determined, they mapped to different chromosomal positions, but few were stably expressed in the same population across locations. This highlights the strong influence of the environment on sucrose yield in sugar beet.

These are the limits of a QTL analysis and they also affect the validation of CGs. Co-localization of CGs and QTLs validates CGs, but the opposite does not exclude a role for the CG in explaining a certain QTL that may be identified under different conditions.

On the same population used to develop the first functional map in sugar beet, a QTL study was also performed. This resulted in mapping of 21 significant QTLs for the traits listed above. No evidence for a correlation to the QTLs identified by Weber *et al*. (1999, 2000) was found. This was mainly explained by the different plant material used and the lack of common markers.

The functional genes on the map by Schneider *et al*. (2002), selected according to their functional involvement in the carbohydrate metabolism, represent the first candidate gene approach in sugar beet. Candidate genes were identified for all the QTLs and used in association studies (Schneider *et al*., 2001, 2002, pers.com.)

1.2.3 – Identification of candidate genes in sugar beet

The first step is the selection of the candidate genes. If the biochemical or physiological pathway related to the trait is well known and sequence information of the relevant genes is available, CGs may be chosen from the genes involved in this pathway. Of particular importance are those genes which encode functions that represent bottlenecks in the determination of the trait. The number of gene sequences available limits this step. When neither genomic or cDNA sequences are available for the species of interest, sequence information from other species may be used to deduce consensus motifs and to design degenerate primers. This possibility was initially exploited in sugar beet, and the described functional map was the result of this strategy. CGs for each of the reported QTLs were identified some of which were used in association studies (Schneider *et al*., 1999, pers. com.).

However, this approach excludes gene products playing a role in less conserved reactions which may be sugar beet specific. To identify new CGs involved in the sucrose storage process independently of DNA sequence information and to retrieve CGs responsible for biochemical or physiological steps in yet unknown pathways, expression levels of around 2000 to 10000 different genes were analysed with respect to their organ specificity and developmental regulation during the vegetative growth phase in the studies presented here. Expression levels here mean transcript levels.

The idea behind this approach is that a gene which is active in a particular process must be expressed. A further assumption is that regulation takes place at the

transcriptional level. This is only true for a fraction of genes because the activity may also be regulated at the posttranscriptional or the posttranslational level. However, the assessment of steady state transcript levels in high throughput format has become a recognized approach to filter genes participating in complex biological processes. Therefore this strategy was chosen here to identify genes which are expressed in roots with particular emphasis on the phase when the root is accumulating sucrose.
To cover those genes which are not regulated at the transcriptional level multiparallel protein profiles have to be generated, and furthermore protein activity needs to be determined for all putative candidates.

1.2.4 - Expression analysis systems

Prior to the advent of transcriptomics approaches, expression analysis was performed by Northern blotting, S1 nuclease activity and *in situ* hybridization. These technologies are characterized by high sensitivity, but they are all time consuming because just one gene can be analyzed at a time. Now high-throughput global transcript profiling technologies are available. These methods can be divided in (1) direct analysis, including procedures involving nucleotide sequencing and fragment sizing, and (2) indirect analysis involving nucleic acid hybridization of mRNA or cDNA fragments (reviewed in Donson *et al*., 2002).
In the first group the large scale expressed sequence tag (EST) analysis should be mentioned. This system generates partial sequences from cloned cDNA fragments. It allows the discovery of new genes as well as the assessment of the gene expression levels in the representative tissue or under a specified condition. The basis of the approach is that the level of an mRNA species in a specific tissue is reflected by the frequency of its corresponding EST in a cDNA library. The analysis of an EST collection provides a powerful system for candidate gene identification. However, EST sequencing is also used in combination with other systems like subtractive cDNA libraries (Jin *et al*., 1997) or the production of macroarray and microarray to be analyzed by hybridization (Schena *et al*., 1995). The mentioned strategies help to narrow down the number of CGs for specific processes.

To reduce costs of EST sequencing Velculescu *et al*. (1995) developed the SAGE (serial analysis of gene expression) technology. The number of individual clones to be sequenced is very much reduced by the concatenation of multiple sequence tags of 10-14 basepairs each prior to cloning. Concerning plants this system was so far applied to rice and *Arabidopsis thaliana*.

The fragment sizing-based methods involve the discrimination of mRNA by differential separation of representative cDNA fragments, and they are all PCR-based. The differential display (DD) (Lang and Pardee, 1992) approach is based on reverse transcription followed by amplification of the cDNAs with arbitrary primers. The amplified cDNAs are separated in acrylamide gels followed by sequencing. Due to low stringency primer annealing this procedure creates many artifacts and is scarcely reproducible.

To overcome this limit the cDNA-AFLP technology was developed (Vos *et al*., 1995; Bachem *et al*., 1996). In this system, cDNAs are restricted by endonucleases and ligated to adaptors. The use of specific primer sets enables stringent PCR conditions to amplify fragments of expressed transcripts. This technology was widely used in plant systems because of its sensitivity, reproducibility and good correlation with Northern analysis (Durrant *et al*., 2000, Jones et.al., 2000).

The indirect analysis of gene expression levels is based on hybridizations. The principle underlying hybridizations of complementary nucleotide sequences is based on the double helix structure of nucleic acids. With the availability of nucleotide sequences represented in libraries of thousands of clones, hybridization-based approaches now allow the simultaneous analysis of many thousand of genes. Two technologies were developed for this aim: one is based on spotting cDNA fragments, either on nylon filters or on glass slides, the other is based on the arrayed synthesis of oligonucleotides on glass slides.

This last technology is expensive. Oligonucleotide arrays are now available for all genes of the model plant *Arabidopsis thaliana* whose genome was the first to be sequenced completely (The Arabidopsis genome initiative, 2000).

For non-model systems the second technology coupled to EST sequencing is more affordable. To compare different physiological situations, cDNA arrays have become a powerful tool (DeRisi *et al*., 1997). Robotic instruments transfer PCR-amplified cDNAs onto nylon filters in case of macroarrays (Desprez *et al*., 1998) or glass slides

in case of microarrays (Aharoni and Vorst, 2002). Multiparallel transcription profiles are generated in hybridization experiments with complex cDNA probes. The statistical analysis of expression patterns obtained in hybridizations with different transcript populations allows to classify groups of genes according to their profiles as members of differently regulated or coordinate processes. Thus, analyses of array data contribute to a better understanding of complex gene expression patterns related to physiology and metabolism, unraveling networks or pathways previously unknown. In this respect, array data on transcripts encoding gene products with unknown function are considered a first step towards their characterization. Therefore the macroarray technology coupled to EST data analysis was selected in this study to identify new candidate genes for the sucrose accumulation process in sugar beet based on gene expression levels. First the sugar beet transcripts were classified with respect to their expression in three different organs, the root, the leaf and the inflorescence, and transcripts preferentially expressed in the root, the storage organ, were regarded as representatives of CGs. In a second experiment, transcripts at different developmental stages were identified. Correlation of expression profiles with the sucrose accumulation process during the development was the selection criterion for the identification of CGs in this kinetic study.

1.2.5 – Association studies

For every gene a number of different alleles exist. If a gene crucially affects a process the process will be affected by the type of allele present. This concept is employed for the validation of CGs in association studies. The molecular variation of a CGs can be analyzed in a set of genealogically unrelated lines to look for statistical associations between the CG polymorphism and phenotypic variation.

In animal and human medical research, several studies on developmental genetics and hereditary diseases used this concept of statistical association. One of the first studies concerned the characterization of the loci involved in hypertension, a multifactor disease. In this case, physiological studies provided criteria for selection of candidate genes. Molecular polymorphisms within the gene encoding angiotensinogen were associated with hypertension (Cambien *et al*., 1992)., the statistical correlation does

not demonstrate the causal relationship. In sugar beet the correspondence of alleles of expressed genes to haplotypes has been recently demonstrated (Schneider *et al*., 2001). That means that non-random association among polymorphisms at different linked sites exists. On average, three to five haplotypes were identified per gene.

The existence of haplotypes allows the mapping of QTLs by association or by the LD (linkage disequilibrium) approach. These methods were shown to provide higher mapping resolution compared to traditional linkage mapping.

Therefore the identification of candidate genes in sugar beet is considered a first step to identify targets for association studies and mapping of QTLs concerning traits like sucrose yield or sucrose content in sugar beet roots.

2 - MATERIALS AND METHODS

2.1 – MATERIALS

2.1.1 – Plant materials

2.1.1.1 - Growing season 2000

For the synthesis of hybridization probes and RT-PCR experiments, leaves, roots and inflorescences were harvested from field-grown plants of the hybrid sugar beet genotype KWS86203. Plants were grown at the Schaugarten of the Max Planck Institute for plant Breeding Research in Köln, Germany. Root samples were harvested 15 and 17 weeks after sowing, leaf samples 12 and 17 weeks. Inflorescence samples were taken in two consecutive years, 2000 and 2001. In all cases samples from three plants were pooled.

2.1.1.2 - Growing season 2001

For the synthesis of hybridization probes and RT-PCR experiments roots were harvested from field-grown plants of the hybrid sugar beet genotype KWS86203. Plants were grown at the Schaugarten of the Max Planck Institute for plant Breeding Research in Köln, Germany. Root samples were harvested at the indicated time-points pooling roots of at least three plants.

TIME-POINT	DATE
t1	15.06.01
t2	01.07.01
t3	17.07.01
t4	07.08.01
t5	21.08.01
t6	10.09.01
t7	17.10.01

2.1.1.3 - Growing season 2002

For the synthesis of hybridization probes and RT-PCR experiments roots were harvested from field-grown plants of the hybrid sugar beet genotype KWS86203. Plants were grown at the Schaugarten of the Max Planck Institute for plant Breeding Research in Köln, Germany. Root samples were harvested at the indicated time-points pooling roots of at least three plants.

TIME-POINT	DATE
t1	05.06.02
t2	25.06.02
t3	12.07.02
t4	02.08.02
t5	23.08.02
t6	17.09.02
t7	09.10.02

2.1.2 – ESTs libraries

2.1.2.1 - Library A006

For the construction of the cDNA library, leaves and roots of four week old pot-grown, diploid sugar beet plants (line KWS51102, KWS SAAT AG, Einbeck, Germany) were used. cDNA library was prepared by Stratagene (Amsterdam, The Netherlands) according to the procedure described in Bellin *et al*., (2002). Library was constituted of 3840 cDNA clones organized in 10 microtiter plates. Sequences were produced as well as described in the cited paper by ADIS unit at Max Planck Institute for Plant Breeding Research in Köln, Germany and stored in the local database Genagent (Bellin *et al*., 2002). Based on trace files of sequencing runs, high quality sequences were determined (Staden *et al*., 1998[30]). In a second step vector sequences were eliminated by the program CROSSMATCH (Ewing and Green, 1998[13]). Sequences were clustered using the program StackPack (Miller *et al*., 1999[23], Burke *et al*., 1999[7]). All sequences were compared against the non-

redundant protein database from NCBI using the BLAST algorithm (Altschul *et al*., 1990[3]). A domain analysis was performed with InterProScan (Zdobnov and Apweiler, 2001[38]) linked to the GO-database (The Gene Ontology Consortium, 2000[33]). For assignment of functional categories, the ontology for biological processes was selected and manually adapted to more plant specific terms.
Sugar beet EST sequences were submitted to GenBank and are available under the accession numbers BQ487526- BQ490673 and BQ654408-BQ654412.

2.1.2.2 - Library A024

cDNA libraries from sugar beet leaves, developing root, storage root and inflorescences were generated by Life Technologies Inc according to the procedure described in Herwig *et al*., *et al*. (2002). Leaf material and young developing roots were harvested in spring, inflorescences including buds, open flowers and developing fruits in summer and mature taproot in autumn. cDNA libraries were normalized using the oligofingerprinting strategy described in the cited paper. A subset of 11520 clones was organized in 30 microtiters plates. Sequences were produced as well as described in the cited paper by ADIS unit at Max Planck Institute for plant Breeding Research in Köln, Germany and are stored in the Sputnik database at MIPS (Münich Information Center for Protein Sequence). Sequences of this library are now available in GeneBank and at PD (Primary Database, GABI)**.** A domain analysis was performed on the first 10752 ESTs using the InterProScan program (Zdobnov and Apweiler, 2001).

2.1.3 – RZPD macroarrays

Macroarray for library A006 were generated by RZPD (Deutsches Resourcezentrum für Genomforschung, Berlin) from amplified PCR products of the 10 plates constituting library A006.
Parameters relative to the spotting are in table. Each spot was transferred in duplicate. To the 10 plates constituting the library two series of serial dilution of spiked control plus negative controls defined in paragraph 2.2.7.3 were added

PARAMETER:	VALUE:
spotting pattern	5x5
membrane size	7.3 x 11.5 cm
pin size	250 µm
number of spotting in same position	10 times
transferred volume x spot	0.15µL
transferred amount of sample	7.5 ng

2.1.4 – Enzymes

Enzymes were purchased mainly from Invitrogen (Groningen, The Netherlands) and New England Biolabs with the 10X buffer supplied. When different enzyme were employed it will be indicate in the following. *Taq* Polymerase was purchased from Invitrogen (Groningen, The Netherlands) or self-produced according to Pluthero, (1993). Adequate *Taq* Polymerase dilution was experimentally estimated.

2.1.6 – Control clones

Clones to be used as controls for the macroarray analysis were obtained by**:**

INSERT:	VECTOR:	FROM:	REFERENCE:
desmin cDNA (human)	pBluescriptII KS	Dr. Rubiera	Desprez *et al*., 1998
nebulin cDNA(human)	pBluescriptII KS	Dr. Rubiera	Desprez *et al*., 1998
*Inf1*cDNA (*Phytophthora infestans*)	pBluescriptII KS	Dr. Valkhamp	Kamoun *et al*., 1997a
*Inf2A*cDNA (*Phytophthora infestans*)	pBluescriptII KS	Dr. Valkhamp	Kamoun *et al*., 1997b
UidA gene (bacterial)	pBluescriptII KS	Dr.Smith-Espinoza, Dr. Santi	Schlaman *et al*., 1994
pAW109	pGEM-T Easy	H.D.Dr. Schneider	Applied Biosystem
pTA71 (ribosomal RNA gene, wheat and barley)	pAC184	H.D.Dr. Schneider	Gerlach and Bedbrook, 1979

2.1.5 - Vectors

In this study the vectors pBluescript II KS (+/-) (Stratagene Cloning System, Amsterdam, The Netherlands), pGEM-T Easy (Promega, USA) and PCMVSport6.0 (Life Technologies) were used.

2.1.7 – Oligonucleotides

Oligonucleotides were purchased from Invitrogen (Groningen, The Netherlands) or Metabion (Martinsried, Germany). They were all resuspended in sterile water. For some experiment (especially RT-PCR) primers already available in the lab either already published in Schneider et 1999(*) or unpublished and relative to clones with 100% identity at nucleotidic level to clones present in the libraries.

T7	GTAATACGACTCACTATAGGGC
T3	AATTAACCCTCACTAAAGGG
uniT	GCTAGCAGTTTTCCCAGTCACGAC
revT	AGCGGATAACAATTTCACACAGGA
uni	GTAAAACGACGGCCAGTGA
rev	GGAAACAGCTATGACCATG
GUST3gpaf	AATTAACCCTCACTAAAGGGAGGTGGACGATATCACC
GUSpolyAgpar	TTTTTTTTTTTTTCGAAGCGGGTAGATATCACACTC
T7-short	ACGACTCACTATAG
6-10N1-s	ACTGTTTAGTGTTTACCTCCC
6-10N1-as	ATGATGAGTACTGGTTTTCCG
6-6G6-s	AAGGAAAGTTAGCAGAAACCC
6-6G6-as	CGTTTTCCCTTAAACCTCCC
24-8F10-s	AGAAGAAGTAGAGTATTTGAGC
24-8F10-as	AATCAGTTCAGCAATGGGGC
24-14K9-s	AAAGACGAACTGTTGCTTACC
24-14K9-as	CTTCATCGATGATTGCACCC
24-12N2s	CTTAAACATTAGGAAAATGGCC
24-12N2as	GGGTAATGACCACCATGCC
24-16L24s	AGCACAACTTCTGCAGTTGC
24-16L24as	ACAATTGACCCTGCCTCGG
24-26I9s	TTCATCTCTCCTTCAAATGGC
24-26I9as	GAATTGTGGGCTTGGTCCC
24-12J3s	TTAAGAAATGGCAATGTCAAGG
24-12J3as	TCCAGTTGTTGTAGCAAGGG
6-50E24s	CCACCACAACCATCACAACA
6-50E24as	GATTCCATGTGGGTGTTGTG
6-3F9f	GCCGCAAAGAGCAATATTGC
6-3F9r2	GTTGAAGCCATGACTTGTGG
6-9F9f	GCAACAAGGTCTTGACATCG
6-9F9r	ATTAGATCAGCAGTGAACTGG
6-2K23f	CGATCCTTTCATTACCACCG
6-2K23r	CACCTCTCCAGTCCTTAGC
6-4P11f	CATCAGCTGAAGAATTCTTGG
6-4P11r	CCTTGATATCTGCCAAGTCC
6-7G18f	TAAAATGAGCAGCTCTAAGGC
6-7G18r	TGTCCACCTTGACTGTATCC
ss-s	CTCTGAACTGAATGTGGAGC(*)
ss-3as	GGAGCCTGAAGGATATCTAG(*)
6-3P11f	GAATGGCTTCTCTTGTAGCC
6-3P11r	TTGAGGTAGTTCTTCCCAGC

2.1.8 - Mapping populations

For mapping the already characterized populations 618 (Schneider *et al*., 1999) and K2 (Schäfer-Pregl *et al*., 1999) were used.

2.2 – METHODS

2.2.1 – Morphological analysis

At each harvesting time point the harvested plants were characterized. The following parameters were measured:

- weight: in grams
- number of rings: all the rings were counted, enlarged and not enlarged. Number of rings was counted on the cross section of maximum diameter
- root length: measured in centimetre from the region of maximum diameter of the root to the tip
- number of leaves: total number of leaves was counted at each time-point including leaves in a stage of advanced senescence.
- root thickness: the root maximum root diameter was measure in centimetres.

Average and standard deviations were calculated in Excel (commercial software).

2.2.2 – Microscope analysis

Root fresh transverse-sections were produced and stained with 0.05% toluidine-blue. Bright field optic Samples were observed at the Leica microscope in bright field and images were acquired with Diskus system**.**

2.2.3 – Sucrose concentration determination

Sucrose concentration was measured using the kit Sucrose/D-Glucose/D-Fructose, for the determination of sucrose in foodstuff (Boehringer Mannheim/R-

Biopharm). Test was performed according to manufacturer instructions. Calculations were performed using the Excel software (commercial software).

2.2.4 – Libraries replication

Libraries were replicated in collaboration with Tania Theis from ADIS service unit MPIZ employing the robotic system MicroGridI (Biorobotics) from master plates. For storage copy were produced in glycerol growth medium (100 mL of (360 mM K_2HPO_4, 132mM KH_2PO_4, 17 mM tri-sodium-citrate, 4mM $MgSO_4$, 6mM NH_4SO_4, 44% glycerol) added to 900 mL LB medium and Ampicilin100 µg/mL) and stored at -80°C. For library PCR amplification copy simple LB growth medium (1% tryptone, 0.5% yeast extract, 1% NaCl pH 7.5 plus Ampicilin 100 µg/mL final concentration) were produced.
In both cases replication was followed by overnight growth at 37°C.

2.2.5 – PCR amplification of cDNA inserts from libraries

2.2.5.1 - Library A006 amplification

The cDNA inserts of all 3840 clones of library A006 were PCR-amplified from 7µL of bacterial lysates (10µL added to 40µL water and heated at 100°C for 5 min) using the primers T7 and T3. In a final volume of 100 µL, concentrations of the PCR reagents were 0.4 µM of each primer, 2mM $MgCl_2$, 0.16mM dNTPs and 5U of *Taq* DNA Polymerase (Invitrogen, Groningen, The Netherlands) in the buffer system supplied by the manufacturer. The PCR program involved an initial denaturation step at 95°C for 5 min, followed by 35 cycles of 30 sec at 93°C, 1 min at 58° C, 1 min at 72°C and a final extension step at 72°C for 5 min. Amplification products were checked for concentration and purity on 1% agarose gels.
Problematic amplifications were repeated with the same procedure, but using as template 4 µL of purified plasmid.

2.2.5.2 - Library A024 amplification

The cDNA inserts of all 11520 clones were PCR-amplified from 7µL bacterial lysates using the primers uniT and revT. In a final volume of 100 µL, concentrations of the PCR reagents were 0.4 µM of each primer, 2mM $MgCl_2$, 0.16mM dNTPs and *Taq* DNA Polymerase (produced according Pluthero, (1993)) in buffer system as in the cited pubblication. Adequate Taq dilution was experimentally estimated. The PCR program involved an initial denaturation step at 94°C for 2 min, followed by 35 cycles of 30 sec at 94°C, 30 sec at 60° C, 3 min at 72°C and a final extension step at 72°C for 5 min. Amplification products were checked for concentration and purity on 1% agarose gels.

2.2.6 – Generation of macroarray

2.2.6.1 – Macroarrays for library A006

Macroarray for library A006 were generated at MPIZ (Max Planck Institute for plant Breeding Research in Köln, Germany) from amplified PCR products of the 10 plates constituting library A006.
Membrane Hybond N+ (Amersham Biosciences, Heidelberg, Germany) were first denatured in denaturation buffer (1.5 M NaCl; 0.5 M NaOH) 20 min. Spotting was performed with the robotic system MicroGridII (Biorobotics) using for protocol parameters as in table.

PARAMETER:	VALUE:
spotting pattern	4x4
membrane size	7.3x11.5 cm
pin size	400 µm
number of spotting in same position	8
transferred volume x spot	0.16 µL
transferred amount of sample	8 ng

After spotting membranes were neutralized in neutralization buffer (1.5 M NaCl; 0.5 M Tris-HCl pH 7.2) for 30 min and dried overnight on Whatman paper. UV crosslinking (0.012-0.12 J cm^{-2}) was performed with Stratalinker for 30 sec (Stratagene, Amsterdam, The Netherlands).

As controls for unspecific hybridization in spotting were added:

- empty vector as in paragraph 2.1.5
- the amplified insert of pAW109 (paragraph 2.2.7.3)
- the amplified control *Inf2a* as serial dilution as in paragraph 2.2.7.3

As positive controls in spotting were added:

- serial dilutions of positive controls as in paragraph 2.2.7.2

2.2.6.2 – Macroarrays for library A024

Macroarray for library A024 were generated at MPIZ (Max Planck Institute for plant Breeding Research in Köln, Germany) in collaboration with Diana Lehman at the ADIS service unit, from amplified PCR products of the 30 plates constituting library A024.

Spotting was performed as explained in paragraph 2.2.6.1 using for protocol parameters as in table.

PARAMETER:	VALUE:
spotting pattern	4x4
membrane size	22x22 cm
pin size	400 µm
number of spotting in same position	8
transferred volume x spot	0.16 µL
transferred amount of sample	8 ng

As controls for unspecific hybridization in spotting were added:

- empty vector sequences as in paragraph 2.1.5

amplified insert of pAW109 (paragraph 2.2.7.3)

- the amplified control *Inf2a* as serial dilution as in paragraph 2.2.7.3.

As positive controls in spotting were added:

- serial dilutions as in paragraph 2.2.7.2.

- because of the higher number of clones and the bigger size of the filters more controls were included in the second macroarray set. As the sensitivity and the linearity range were proved to be comparable to the first set of macroarrays the control *nebulin* at a concentration of 50 ng μL^{-1} was spotted in 96 duplicates for each of the 6 subset in which each filter could be subdivided for normalization purposes. The sub-filter average of this *nebulin* spots was used to normalize the relative sub-filter portion.

2.2.7 – Macroarray controls

2.2.7.1 - Controls amplification

The cDNA clones for the human *nebulin* and *desmin* genes and for the *Inf1* and *Inf2A* genes of *Phytophthora infestans* (paragraph 2.1.6) were amplified with primers T3/T7 in 100 µL final volumes. The genomic clone for the bacterial gene *uidA* (paragraph 2.1.6) was instead amplified with the specific primers GUST3gpaf/GUSpolyAgpar, to produce a PCR fragment having the T3 primer sequence in the beginning and a polyA tail in the end (according to the procedure described in Smith-Espinoza, 2001).

Concentrations of the PCR reagents were 0.4 µM of each primer, 2mM $MgCl_2$, 0.16mM dNTPs and homemade *Taq* DNA Polymerase in a final volume of 100 µL. For amplification 10 ng of plasmid purification were used as template. The PCR program involved an initial denaturation step at 94°C for 2 min, followed by 35 cycles of 45 sec at 94°C, 45 sec at 58° C, 1 min at 72°C and a final extension step at 72°C for 5 min.

The clone pAW109 (paragraph 2.1.6) was amplified with primers uni and rev in 100 µL final volume. Concentrations of the PCR reagents were 0.4 µM of each primer, 2mM $MgCl_2$, 0.16mM dNTPs and homemade *Taq* DNA Polymerase in a final volume of 100 µL. For amplification 5 ng of plasmid purification were used as template. The PCR program involved an initial denaturation step at 94°C for 2 min, followed by 35 cycles of 45 sec at 94°C, 45 sec at 60° C, 1 min at 72°C and a final extension step at 72°C for 5 min.

2.2.7.2 - Positive controls for spotting

After purification, the PCR products for the controls *uidA*, *nebulin*, *desmin*, *Inf1* (paragraph 2.1.6) were quantified by spectrophotometric analysis.
For each of these amplified controls samples with concentration 200, 150, 100, 75, 50, 25, 10, 5 ng μL^{-1} were prepared by serial dilution to be used in spotting.

2.2.7.3 - Negative controls for spotting

With the same procedure as explained in paragraph 2.2.7.2 samples with concentration 200, 150, 100, 75, 50, 25, 10, 5 ng μL^{-1} to be used in spotting were prepared for the control *Inf2a* (paragraph 2.1.6).
The empty vectors pBluescript II KS (+/-) (Stratagene Cloning System, Amsterdam, The Netherlands), pGEM-T Easy (Promega, USA) and PCMVSport6.0 (Life Technologies) were purified as explained in paragraph 2.2.16. After spectophotometric determination of their concentration sample at 50 ng μL^{-1} were prepared for spotting.

2.2.7.4 - *In vitro* transcription

The control clones containing the inserts *uidA*, *nebulin*, *desmin*, and *Inf1* genes were linearized and used for *in vitro* transcription. 1µg of DNA was used as template. The reaction mix included: 1M DTT, 1 mM of ATP,GTP,CTP, and UTP ribonucleotides, 10X supplied buffer for T3 RNA polymerase and 40 units of T3 RNA polymerase(Roche, Mannheim, Germany) in a final volume of 40µL. It was performed at 37°C for 1-2 h and was stopped by incubation at 65°C for 5 min. Residual DNA was removed with the kit DNA-Free-DNAase treatment and Removal Reagent (Ambion, Inc., Huntingdon, UK), following the instruction of the manufacturer.

2.2.7.5 - Spiked positive controls

The *in vitro*-transcripts for the controls *uidA*, *nebulin*, *desmin* and *Inf1* were added to the poly(A)$^+$RNA in amounts of 0.001 to 1% prior to probe preparation as follows:

CONTROL	PERCENTAGE ADDED FOR PROBE SYNTHESIS
desmin	0.001%
nebulin	0.01%
Inf1	0.10%
UidA	1%

2.2.8 – RNA extraction

Poly(A)$^+$RNA was isolated using the procedure reported in Bartels and Thompson (1983). Frozen plant material (5 gr leaves and inflorescences or 20 gr roots) was ground to a fine powder in liquid nitrogen. The powder was shaken in buffer I (0.1 M NaCl, 0.05 M Tris-HCl ph 9, 0.01 M EDTA, 2% SDS, ph9) and incubated at 37°C for 15 min. Equal volume of phenol-chloroform was added to the homogenate and shake for 10 min. The mix was centrifuged at 12000 X g for 10 min at RT. Surnatant was transferred to clean tubes and the phenol-chloroform extraction repeated. Then a chloroform extraction was performed and aqueous phase collected. To this 1/10 volumes of 4M NaCl was added and, after centrifugation, 0.1 gr of Oligo dT-cellulose were added for each 10 gr of tissue. The sample was mixed for 30 min at room temperature. Then the cellulose was spun down and washed three times in buffer II (0.01 M Tris-HCl ph 7.5, 0.4 M NaCl, 0.2% SDS) and subsequently three times in buffer III (0.01 M Tris-HCl ph 7.5, 0.1 M NaCl) until the eluate A_{260}=0.
The polyA$^+$ enriched RNA was eluted at 55°C with 2 mL of prewarmed buffer IV (0.01 M Tris-HCl ph 7.5). The nucleic acids precipitated by the addition of 1/10volumes of 4M NaCl and 2.5 volumes of absolute ethanol. The precipitated nucleic acids were pelleted by centrifugation at 12000 X g for 15 min at 4°C and then washed three times in 70% ethanol. Then, the pellet was dried and resuspended in DEPC treated water and stored at -80°C.

Alternatively the Poly(A)$^+$RNA was purified from total RNA using the mRNA purification kit from Amersham Biosciences, Heidelberg, Germany.
In this case instruction of the manufacturer for both total RNA extraction and mRNA purification were followed.

2.2.9 – Probe synthesis

2.2.9.1 - Complex probe

After DNAse treatment (Ambion, Inc., Huntingdon, UK), 0.6 µg of poly(A)$^+$RNA were used to synthesize cDNA. To the poly(A)$^+$RNA, 500 ng of Primer d(DT)$_{15}$ for cDNA synthesis (Roche, Mannheim, Germany) and the positive controls according to what reported in paragraph 2.2.7.5 were added to a final volum of 11 µL prior to incubation at 70 °C for 10 min. The sample was then equilibrated at 43°C. A mix containing 5X RTBuffer, 0.01 M DTT, 1mM of dATP, dTTP, and dGTP (final concentration), 5 µM dCTP (final concentration), 200 units of SuperscriptII Reverse Transcriptase (Invitrogen, Groningen, The Netherlands) and 30µCi of [^{33}P]-α-dCTP as radionucleotide was added. RNA was hydrolyzed by adding 1 µL of 1% SDS, 1 µL of 0.5M EDTA and 3 µLof 3M NaOH at 65 °C for 30 min.
Samples were incubated for 15 min at room temperature and neutralized with 10 µL of 1M Tris-HCl pH 5.3 and 3 µL of 2N HCl. Unincorporated nucleotides were removed by precipitation adding 5 µL of 3M Na-Acetate ph 5.3, 5 µL of yeast tRNA (Invitrogen, Groningen, The Netherlands) at the concentration of 10mg/ml, and 60 µL of Isopropanol. After incubation of 1 h at -20°C samples were centrifuged 30 min RT at maximal speed and the pellet was resuspended in 100 µL of water.
Alternatively purification through Sephadex G-50 (Amersham Biosciences, Heidelberg, Germany) columns was used to remove unincorporated nucleotides.
The labeled cDNA was denatured and added to the hybridization solution. Probes used for hybridization revealed at least 30% incorporation of [^{33}P] according to scintillation counting.

2.2.9.2 - Oligo labelling for oligohybridization

The T7-short (14-mer) part of the T7 oligonucleotide was labeled adding to 20 pmol of oligo 1 µL of 10X PNK Buffer, 10 µCi ^{33}P-γ-dATP and 10 units of T4 Polynucleotide Kinase (Roche, Mannhaim,Germany) in a final volume of 10 µL. The mix was incubated 1 h at 37°C and the reaction was stopped by incubation at 80 °C for 5 min.
The labeled oligonucleotide was purified using columns G-25 MicroSpin (Amersham Biosciences, Heidelberg, Germany) following the instruction of the manufacturer. Probes used for hybridization revealed at least 30% incorporation of [^{33}P] according to scintillation counting.

2.2.10 – Hybridization, washing and exposition

2.2.10.1 - Mock hybridization

Each filter was submitted to a mock hybridization without probe, and subsequent regeneration to clear the membrane of improperly immobilized DNA molecules. The procedure was as for complex hybridization but in absence of probe.

2.2.10.2 - Complex hybridization

For hybridization with complex probes (Hoheisel *et al*.,1994), nylon filters were initially prehybridized with 20 mL of Church buffer (0.5 M sodium phosphate buffer pH 7.2, 7% SDS; 1 mM EDTA) including 200 µL salmon sperm DNA (10 mg/ml) at 65°C for 2 h.
The hybridization was started with fresh Church buffer and the denatured probe at 65°C for 16 h. Filters were washed twice in wash buffer (40 mM sodium phosphate buffer pH 7.2, 0.1 % SDS) for 40 min at 65°C. After that filters were wrapped and exposed to phosphor screen.

2.2.10.3 - Oligohybridization

To estimate the amount of hybridization target spotted, a hybridization with a [^{33}P]-γ-ATP-labeled 14-mer, which is part of the T7 oligonucleotide used for the amplification, was performed.
Both prehybridization and oligonucleotide hybridization were performed in SSARC buffer (600 mM NaCl, 60 mM sodium citrate, 7.2 % sodium N-lauroylsarcosine salt), including 200 µL salmon sperm DNA (10 mg/ml) at 10°C. Filters were washed in precooled SSARC buffer at 10 °C for 10 min.

2.2.10.4 - Filter regeneration

Hybridization signals were removed by twice adding boiling regeneration buffer (5 mM sodium phosphate pH 7.2, 0.1 % SDS) to the filters and incubating for 30 min at 85°C. Filters were used in five consecutive hybridization experiments.

2.2.11 – Image acquisition

Filters were exposed to imaging plates for 16 h or 32 h. Signals were laser scanned by PhosphorImager (STORM 860, Molecular Dynamics) to quantify the activity of radioactive bands or spots on the nylon membranes.
Samples were exposed to phosphor screen (Kodak). They were sensitive to source of ionizing radiation. Once exposed to the storage phosphor screen the ionizing radiation induces a latent image formation that can be subsequently scanned.
The unit used to quantify the signal are given in PhosporImager Counts (or Molecular Dynamics Counts, MCD) which is an arbitrary unit that describes the intensity of photon emissions released from the storage phosphor screen during scanning.

2.2.12 – Computer software and procedures

2.2.12.1 - Macroarray quantification: ArrayVision

Hybridization signals were quantified by the program Arrayvision (Imaging Research Inc., Haverhill, UK) and corrected for the local background (values for hybridization with complex probes = $x_{ij\text{-}b}$; values for hybridization with oligonucleotide = $x_{ijo\text{-}b}$).

2.2.12.2 - Normalization

To normalize the values for signal intensities obtained in different hybridization experiments, spiking controls were used (Bernard *et al*., 1996). Among the four non-sugar beet transcripts added in different amounts for probe synthesis, the transcript *nebulin* present in 0,01% of the poly(A)$^+$RNA was selected as normalization standard.

For each hybridization experiment relative to library A006, the value for the signal intensity corresponding to the median of the serial dilution of the normalization standard (*nebulin* control) was calculated (for hybridization with complex probes = $ns_{p\text{-}b}$; for oligonucleotide hybridizations = $ns_{o\text{-}b}$).

For each experiment relative to library A024, the median of the 96 samples spotted for the control *nebulin* in each of the six subfilter of each filter, was estimated. This value was used as normalization standard for the relative subfilter (for hybridization with complex probes = $ns_{p\text{-}b}$; for oligonucleotide hybridizations = $ns_{o\text{-}b}$).

For all cDNA clones, the normalized values x_{ijn} were calculated according to the following formula: $x_{ijn} = (x_{ij\text{-}b} / ns_{p\text{-}b})/(x_{ijo\text{-}b}/ ns_{o\text{-}b})$.

For each hybridization experiment, eight replications of each data point were considered.

Different macros and routines were developed within the software Excel (commercial software) to perform the different normalization steps.

2.2.12.3 - Data processing: Array Stat

The significance of differences in signals obtained by hybridizing with probes from different plant organs was evaluated by the ArrayStat program package (Imaging Research Inc., Haverhill, UK).
The pooled-curve fit-based random error estimation method was used to exclude outliers analyzing the library A006. Transcripts were automatically defined as preferentially expressed in one organ if the values of normalized signal intensities for this organ were at least twice as high as those for the other organ(s).
Concerning library A024 the small sample procedure was selected to exclude outliers. The significance of the differential expression was evaluated applying the *F-test* with false positive rate set to $\alpha < 0.05$ and the correction procedure *Stepdown Bonferroni*.

2.2.12.4 - Clustering analysis: Genesis

The free downloadable software Genesis (Sturn, 2000) was used for clustering analysis of the expression profiles. Data were transformed from $\log_{10}$ to $\log_2$ prior analysis. To highlight differential expression during the time-course the "median center" function was applied.
Data were filtered for samples containing data points not measurable to avoid distortions. As distance measure the "Person correlation" was selected.
Finally the partition clustering algorithms *k*-means was used for clustering, selecting k value from 2 to 15. Evaluation of stability was performed exporting the obtained clusters in Excel (commercial software) and applying functions included in the program to evaluate stability of results.
For discussion about the choice of parameters see also chapter 5.

2.2.12.5 - DNA sequence analysis

For all basic sequence analysis the software package Genetics Computer Group (Madison, WI) version 9.0 was used.

2.2.13 – Northern blotting

2.2.13.1 - RNA Electroforesis

3 µg of PolyA$^+$RNA were mixed to 4 µL of the 10X RNA loading buffer (1X MOPS, 1,75 % formaldehyde, 0.5% deionized formamide, 0.4%(w/v) bromophenol blue) and incubated at 55°C for 15 min. Following the incubation, the denatured RNA samples were separated in a 1% (w/v) denaturing agarose gel containing 1X MOPS buffer and 2.2 M formaldehyde, using 1X MOPS running buffer.
Mops 5X was prepared with 0.2 M MOPS, 50 mM NaAc, 50 mM EDTA, pH 7.5 with NaOH and was autoclaved before use (Bartels *et al*., 1986).

2.2.13.2 - Northern blot transfer

The samples were blotted onto Hybond N+ nylon membranes (Amersham Biosciences, Heidelberg, Germany). A wick of Whatman 3MM paper was placed on a support over a reservoir of 20X SSC (3M NaCl and 300mM sodium citrate). After complete saturation of the wick, the gel was placed carefully on top to ensure that no air bubbles were preset between the gel and the wick. A sheet of Hybond N+ membrane, cut at the size as the gel, was wetted in 20X SSC and placed on top of the gel. Saran Wrap was placed around the edges of the gel and membrane to prevent "short-circuiting" of the blotting procedure. Six sheet of 3MM paper (same size as the gel, with the first wetted with 20X SSC) and a stack of paper towels were placed on top of the membrane. Following over-night transfer of the RNA, the filter was rinsed with 2X SSC and placed on two sheets of Whatman 3MM wetted with 2X SSC. The RNA was fixed to the membrane by UV irradiation at 0.012-0.12 J cm^{-2} for 30 sec using the UV Stratalinker (Stratagene, Amsterdam, The Netherlands).

2.2.13.3 - Random primed labelled probe for Northern analysis

Probes were prepared from agarose gel electrophoresis-separated DNA fragments using the Random Primed DNA Labelling method. The labelling was

carried out in 20 µL of the following reaction mix: 25 ng denatured DNA, 5 µL of OLB, 1 µL Klenow enzyme and 2 µL (10 µCi μL^{-1}) [α-^{32}P]dCTP.
For OLB solution three solutions were prepared: A: 1ml (1.25 M Tris-HCl pH 8, 0.125 M $MgCl_2$), 5 µL of 20 mM dATP, dATG and dTTP, 18 µL 2β-mercaptoethanol; B: 2M HEPES, pH6 with NaOH; C: Hexadeoxyribonucleotides resuspended un 3M Tris-HCl, 0.2 mM EDTA pH 7. Then 100µL of solution A were joined to 250µLof solution B and 150 µL of solution C to produce the OLB solution
The reaction was incubated at room temperature for 1 h. The unincorporated nucleotides were removed using Sephadex G-50 column (Amersham Biosciences, Heidelberg, Germany). The labelling reaction was increased in volume to 100 µL with 50mM Tris-HCl pH8. Column effluents were collected in Eppendorf tubes. The probe was denatured at 100°C for 5 min and immediately chilled in ice prior to use.

2.2.13.4 - Pre-hybridisation and hybridisation of RNA filters

Prehybridisation was for 0.5 to 4 h in 20mL hybridisation solution (5X SSC, 0.1 M PIPES, 1X Denhardt`s, 0.1% SDS, 50% deionized formamide) at 42 °C. Hybridisation was carried out overnight at 42 °C using the same solution with the addition of heat denatured [α-^{32}P]dCTP-labelled probe. Filters were washed at high stringency with 2X SSC, 0.1% SDS at 65°C for 30 min three times. Filters were sealed in a plastic bag and exposed phosporscreen. Signals were recorded by phosphorimaging.

2.2.14 – Semiquantitative RT-PCR

For quantitative RT-PCR experiments, first strand-cDNA was synthesized from 0.5 µg DNAse-treated (Ambion Inc., Huntingdon, UK) poly(A)$^+$RNA using the Superscript II Reverse Transcriptase (Invitrogen, Groningen, The Netherlands). To the poly(A)$^+$RNA, 500 ng of Primer $p(DT)_{15}$ for cDNA synthesis (Roche, Mannheim, Germany) were added prior to incubation at 70°C for 10 min followed by immediate chilling. To these the 5X RTBuffer, 2 µL of DTT and 1 µL of dNTPs mix (10 mM each) and 20 units of enzyme were added. Reaction was carried for 1 h at 43 °C.

Three µL of a 1:100 dilution of cDNA samples were used as templates for the PCR assays with sequence-specific primers. Concentrations of the PCR reagents were 0.4 µM of each primer, 2mM $MgCl_2$, 0.16mM dNTPs and homemade *Taq* DNA Polymerase in a final volume of 25 µL. The PCR program involved an initial denaturation step at 94°C for 2 min, followed by variable number of cycles of 45 sec at 94°C, 45 sec at the annealing temperature specific for each primer pair, 1 min at 72°C and a final extension step at 72°C for 5 min. For each transcript, the number of cycles providing the most dynamic range of the transcript profile was experimentally determined.

2.2.15 – SSCP mapping

PCR fragments were amplified with specific primers from the genomic DNA for each population. During the setting of the assay only the parents and few individual of the population were analysed.

Once set the assay was extended to the full population for segregation analysis. PCR was performed with the following reagent: 0.4 µM of each primer, 2mM $MgCl_2$, 0.16mM dNTPs and homemade *Taq* DNA Polymerase in a final volume of 25 µL. 50 ng of genomic DNA were used as template. The PCR program involved an initial denaturation step at 94°C for 2 min, followed by 35 cycles of 45 sec at 94°C, 45 sec at the annealing temperature specific for each primer pairs (normally 55°C), 1 min at 72°C and a final extension step at 72°C for 5 min.

PCR products were subjected to SSCP electrophoresis. Two µL of each PCR reaction were added to 9 µL denaturing solution (95% formamide, 0.01M NaOH, 0.05% xylene cyanol, 0.05% bromophenol blue), heated to 94°C for 2 min, then chilled. Sample of 3-5 µL were run on 0.5 x MDE gels using 0.6 x TBE (Sambrook *et al*., 1989) running buffer.

Gels were run at constant power 2 watts for 12-18 h at RT. After run gels were stained with silver according to the procedure described in Sanguinetti *et al*., (1994).

2.2.16 - Plasmid purification

A volume equal to 1.5 mL of overnight culture containing the required plasmid was pelleted in Eppendorf tube at 15000 x g for 5 min. The protocol described in Sambrook *et al*., (1989) was followed for plasmidic DNA purification.

3 – SUGAR-BEET MACROARRAYs

Macroarrays are defined as those arrays that rely on robotically spotted targets (bacterial colonies or amplified cDNAs/DNAs) that have been immobilized on membrane-based matrices like nylon-filters. These are then hybridized with radioactively labelled complex probes produced by reverse transcription of different poly(A)$^+$RNA samples, in order to identify differential expression. Macroarrays were developed to screen whole cDNA libraries by hybridization (Lennon and Lehrach, 1991) and were first applied in plants by Desprez *et al*. (1998). The power of this technology is that it allows to evaluate the expression of many genes simultaneously.
This chapter describes the establishment of the macroarray technology to select for candidate genes among sugar beet ESTs. Additionally, the technical parameters of the established high throughput analysis are evaluated.

3.1 - ESTABLISHMENT OF SUGAR-BEET MACROARRAYS

3.1.1 – Generation of cDNAs arrays using nylon filters

To produce nylon filter arrays the following steps are necessary:

1 - amplification of the inserts from a cDNA library by PCR
2 - transfer of PCR products and control samples in duplicates to nylon membranes by robot
3 - hybridization of filters with complex probes prepared from poly(A)$^+$RNA extracted from the tissue under investigation
4 - quantification of signal intensities from images and statistical evaluation of data with specific computer software

For each of these steps, protocols are reported in the Material and Methods section.
In the first experiment inserts of 3840 cDNA clones belonging to the sugar beet cDNA library A006 organized in 10 microtiter plates, were amplified using universal primers. The presence of a PCR product and a concentration at least 50 ng μL^{-1}were

verified by loading onto agarose gels and the sizes of the amplified fragments were estimated to range from 500 to 3000 bp, with an average size of more than 1 kb**.** An example is shown in Figure 3.1.

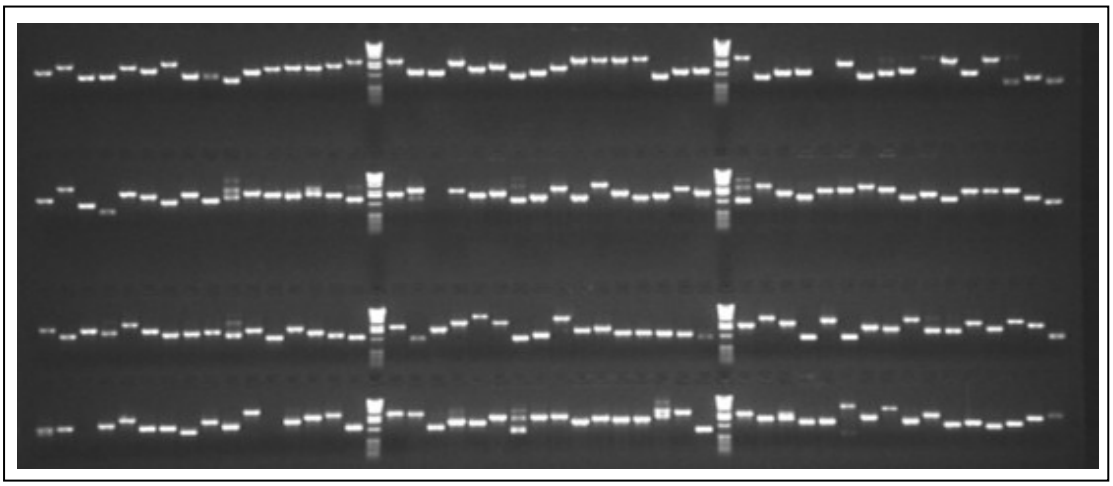


Figure 3.1 Amplification of inserts of 192 cDNA clones belonging to the library A006.

In case no PCR product was amplified or if the product concentration was below 50 ng μL^{-1} the amplification was repeated at least twice more and in selected cases plasmids were purified to repeat the amplification. In the end 89.9% of the clones belonging to the library were successfully amplified. This collection including controls for unspecific hybridization and normalization of the data (paragraph 2.2.4), was transferred to nylon membranes to produce the macroarrays (Figure3.2).

To assess and compare the quality of filter printing, one set of filters was commercially produced by RZPD (Berlin, Germany) and a second was generated in collaboration with the ADIS facility at MPIZ. Concerning the RZPD filters, spotting was performed in a 5 x 5 pattern; therefore the complete library could be printed in duplicate onto 7.3 x 11.5 cm nylon filters. As to the filters produced at the ADIS unit, a 4 x 4 pattern was chosen to accommodate the complete set of clones in duplicate onto two 7.3 x 11.5 cm filters.

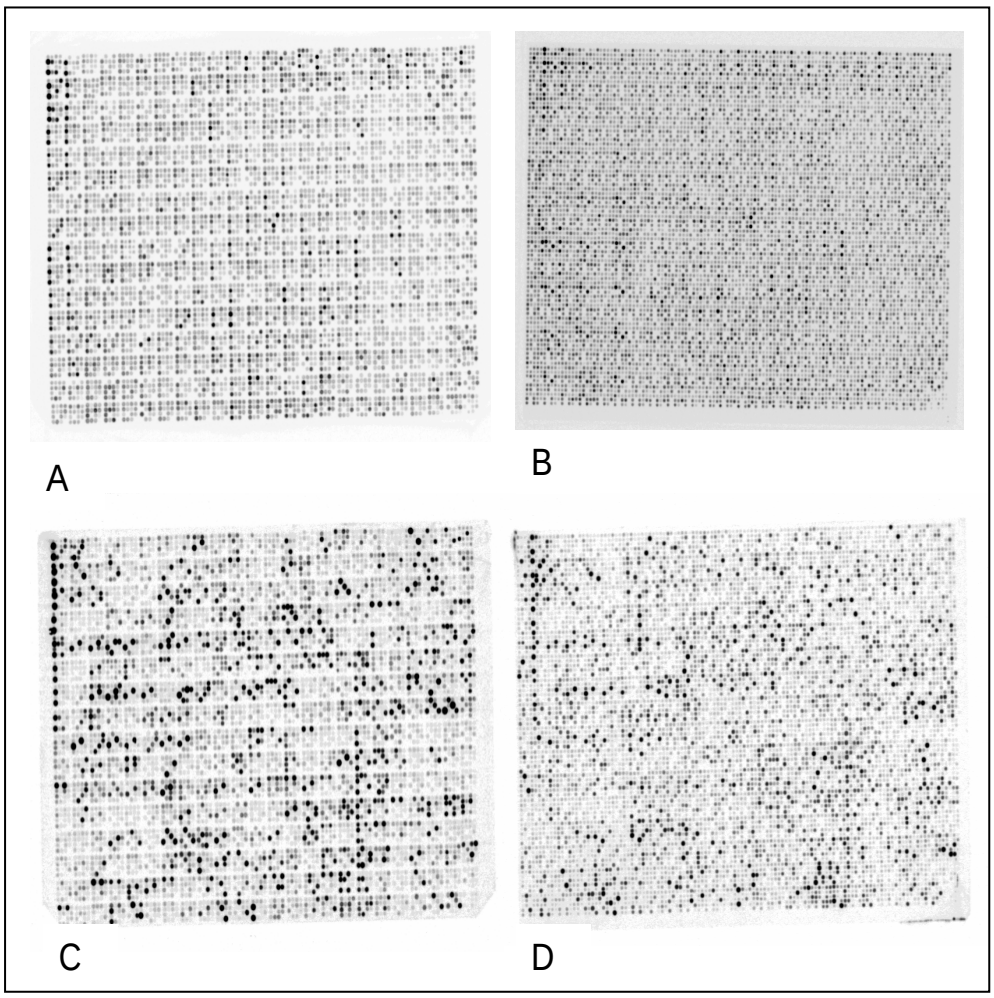


Figure 3.2 (**A, B**) Oligo-hybridization and (**C, D**) hybridization with complex probes (produced from leaf poly(A)$^+$RNA) of filters produced in collaboration with ADIS service unit at MPIZ (**A**, **C**) and by RZPD (**B**, **D**). Serial dilutions for controls are evident in the oligo-hybridization.

Hybridization results were found to be very consistent between both filter sets (see chapter 3.2.1). In following experiments the filters produced at MPIZ were used.

For every filter spotting efficiency was confirmed by performing an oligo-hybridization using an end-labelled short (14-mer) T7 universal primer, matching all PCR products spotted, as probe. Examples of oligo-hybridizations and complex hybridizations for both filter sets produced are shown in Figure 3.2.

3.1.2 - Experimental design

In the first series of experiments radioactively labelled complex probes were produced from different organs like root, leaves and inflorescences of sugar beet plants to study their expression. Because of the low efficiency of reverse transcription

when performed on total RNA extracted from root material, poly(A)$^+$RNA was preferred for the synthesis of complex probes. Hybridization experiments are influenced by many parameters. Therefore an important aspect for evaluation is the number of times hybridization should be repeated in order to check the reproducibility of the signals. The application of statistical procedures (Herwig *et al*., 2001) allows to distinguish between true changes in expression and errors in hybridization or data collection (see also chapter 3.2.2). Hornberg *et al*., (2002) demonstrate that using different membranes for one sample during the experiments ("reversing membranes") increases the statistical significance of the data because this allows to minimize the effect of quantitative differences in spotted PCR fragments. These aspects led to the introduction of both technical and biological replicates for the experiment (Fig. 3.3).

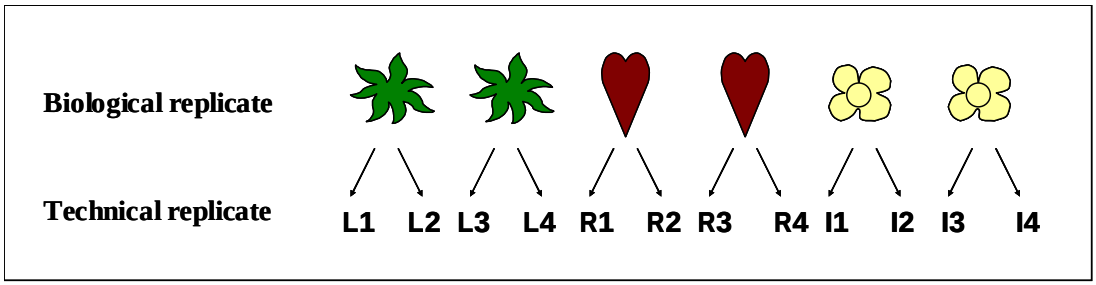


Figure 3.3 Definition of biological and technical replicates for the experiment in which the expression in three different organs was compared (see text). In this experiment two different probes (technical replicates) were produced for each biological replicate. The biological replicates were harvested independently from different plants on different days. L = leaf, R = root, I = inflorescence. The number defines the hybridization probe as four hybridizations were carried out.

"Technical replicates" refers to replications in which the poly(A)$^+$RNA used to synthesize complex probes derives from the same extraction. They generally involve a smaller degree of variation in measurements than "biological replicates", which refer to hybridizations in which the poly(A)$^+$RNA used to synthesize the probes was taken from different extractions and/or from different individuals. A measurement of the degree of variation related to each of these replication levels is given in section 3.2.2, in which technical parameter of these filters are evaluated.

The interpretation of the results is affected by the kind of replications introduced (Yang and Speed, 2002). Therefore "biological replicates" were introduced here as it was the aim to select for preferential expression in one tissue. Poly(A)$^+$RNA was then isolated from two independent samples of leaves, roots and inflorescences, respectively. Two independent [^{33}P]-labelled probes were prepared from each of these samples as "technical replicates". This way a total of four replica for each tissue were

produced according to the scheme reported in Figure 3.3. To complete the experiments, four filters were used for two to four hybridizations each. Probes synthesized from the same tissue were applied, where possible, to different filters (with one exception) to follow an experimental design involving "reversing of the membranes". More details about how hybridizations were performed are given in Figure 3.4.

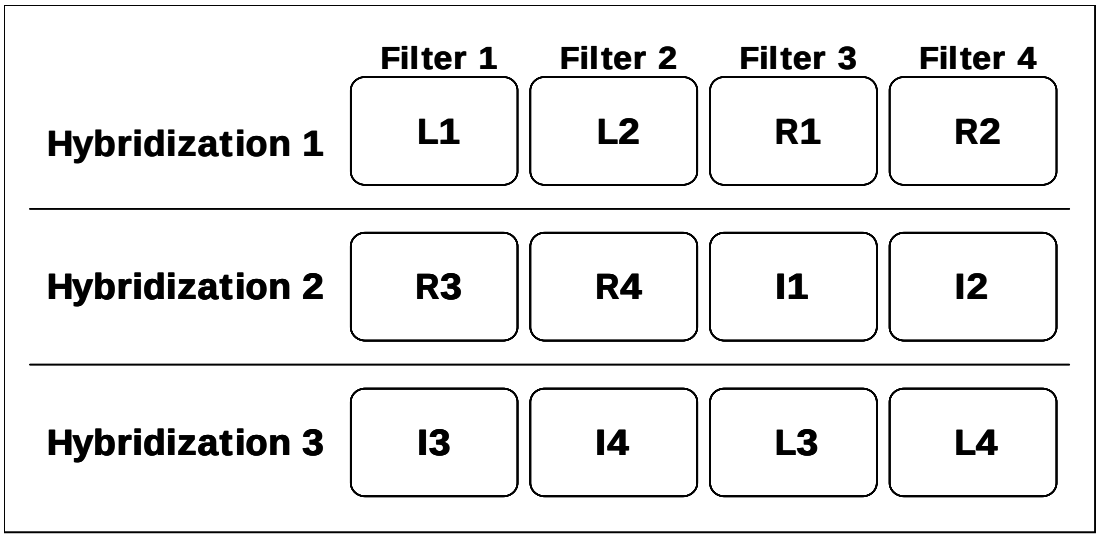


Figure 3.4 Scheme of the experimental design developed for this experiment. Due to decay of Filter 2 after the second hybridization, Hybridization 3 could not be performed as outlined here. Therefore the probe "I4" was hybridized to Filter 3 in a fourth hybridization,. L = leaf, R = root, I = inflorescence. Number refers to the probe number as defined in Figure 3.3.

3.1.3 - Image acquisition and quantification

Digital images of all hybridization results were produced by phosphorimaging. Therefore after hybridization and washing of the filters, these were exposed to a phosphor screen. The surface of this is excited by isotopes like ^{33}P. Following 24 hours of exposure, the imaging plate was developed by scanning with a laser of an appropriate wavelength, thus releasing the energy of the excited electrons. This release was detected by the phosphorimager, yielding a digital image of the radioactivity with respect to its location and intensity (Freeman *et al*. 2000).

The image analysis software ArrayVision (Imaging Research) was used to quantify the signal intensity of each spot in these images. Some information like spotting pattern, spot distances, etc is required from the user to create a grid superimposed on

the image. Automatic alignment of such a grid with the spots was performed, followed by visual inspection of the image to adjust the grid where necessary.

To accept the hybridization an empirical threshold was set using the "Gray/Color adjust" function in the Imagequant software for which a value higher than 1000 was required. A second requirement concerned the LUT (Look-up table) map in the ArrayVision image files which had to produce a X value smaller than 64000 if the Auto Contrast feature is applied. This value gives an indication of the gray levels required to map 99.9% of the image pixels, if 0 is black and 66635 is white.

3.1.4 - Data processing

Data processing includes background subtraction, normalization, detection of outliers and logarithmic transformation. An example of the effect of each of these steps on the data is shown in the Figure 3.5 and in Figure 3.7

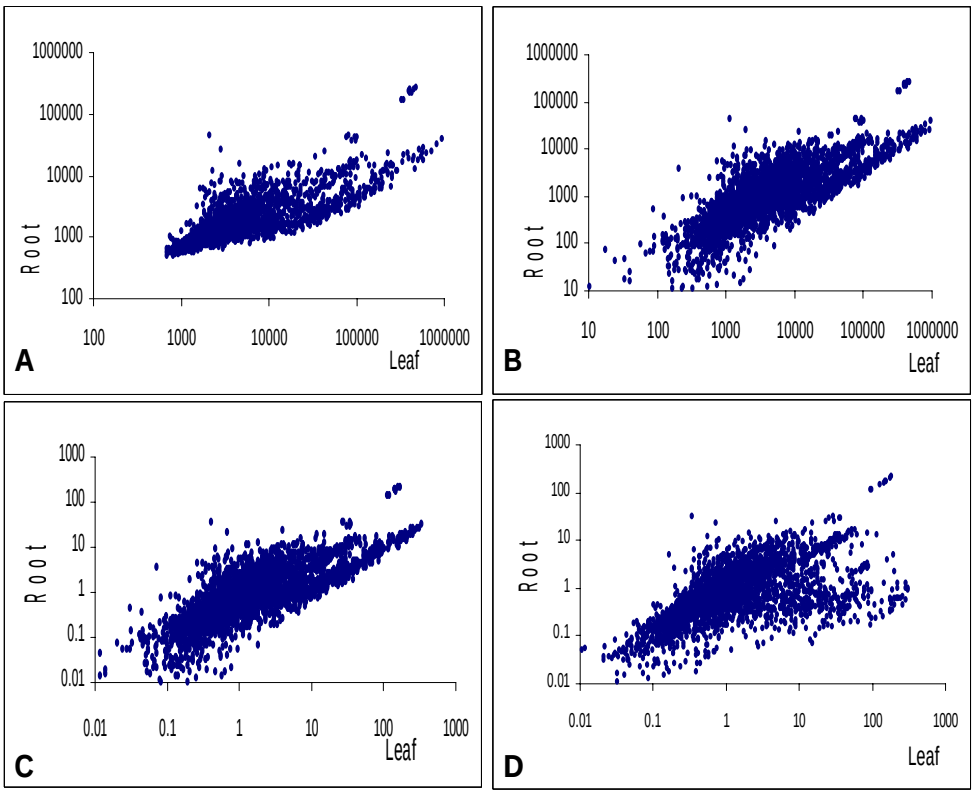


Figure 3.5 Effect of the different steps in data management. Double logarithmic scatter plots of the signal intensity for probes prepared from leaf and root poly(A)$^+$RNA at different analysis stages. In A) the raw data relative to a single hybridization experiment are plotted. In B) the same data are plotted after background subtraction and in C) after normalization. Finally in D) the effect due to replication can be assessed. In the last are plotted median data from 8 replica of hybridizations performed with leaf and root probes. .

3.1.4.1 - Background subtraction

The effects of background signals on the data are supposed to be additive (Beißbarth *et al*., 2000). Therefore the portion of the signal due to the background was measured on spotted water by the ArrayVision computer software and automatically subtracted from the spot intensity value. Background values were measured on water spots belonging to the same subgrid (option "local" in the software) and the median value of all the pixels in these spots was used to calculate its value.

Final values for signal intensities calculated by the Array Vision were expressed therefore as

subtracted volumes = volume - background,

where volume = density (average value of all the pixels in the spot) x area

and

the background = density (median value of the pixels in the empty spots) x area.

Their unit was MDC (Molecular Dynamics Counts), an arbitrary unit dependent on the scanning device. Tables produced by the Array Vision program were exported to the Excel program.

After background subtraction, the intensity values relative to the spotted negative controls listed in chapter 2.2.7.3 were checked. Empty spots showed on average no residual signal. In contrast, a residual signal was observed for the spotted empty vector pBluescript and for the negative control spotted to monitor the unspecific hybridization signal. They were automatically excluded at further stages in the data analysis and never appeared among the differentially expressed genes. Therefore to avoid complications in the further statistical analysis due to the application of a second BKG subtraction (Imaging Research, personal communication), no other correction was introduced.

3.1.4.2 - Normalization

In order to identify differentially expressed genes, hybridization results cannot be compared directly. They need to be normalized to compensate for differences due to varying efficiencies of reverse transcription, probe purification, hybridization, filter quality, etc (Eickhoff *et al*, 1999). Several strategies are used for normalization: a)

global normalization uses all genes on the array, b) normalization based on housekeeping genes uses constantly expressed housekeeping/invariant genes and c) normalization based on internal controls uses known constant amount of exogenous control genes added during hybridization ("spiked controls").

The first option, using all genes on the array, appears be appropriate for whole genome arrays or when the intensity of all signals is supposed to be similar in the samples analyzed. However, when working with a subset of genes or samples very different from each other like samples from different organs, this approach was shown to be inappropriate.

The use of housekeeping genes seems an attractive alternative, but, comparing the expression in different tissues, constancy of housekeeping gene expression cannot be assumed *a priori*. Reports about sugar beet housekeeping genes are limited and, furthermore, there are now several reports in literature describing housekeeping genes to be regulated (Lee, 2002).

Therefore the third indicated strategy was applied using spiked controls. This normalization procedure was already reported by Zhao *et al*. (1995) and Bernard *et al*. (1996). Exogenous cDNAs coding for the human genes *nebulin* and *desmin* and for the gene *Inf2A* isolated from *Phytophthora infestans*, as well as the genomic DNA coding for the bacterial gene *uidA* were *in-vitro* transcribed as explained in the chapter 2.2.7.4. After spectrophotometric quantification, the corresponding poly(A)$^+$RNAs were spiked in the poly(A)$^+$RNA used for reverse transcription as follows: *uidA* as 1%, *Inf2a* as 0.1%, *nebulin* as 0.01% and *desmin* as 0.001% in relation to the poly(A)$^+$RNA used for probe synthesis.

For normalization purposes, the same genes need to be present on the filters. Therefore PCR amplified fragments of the above named cDNA and genomic clones were as well quantified and transferred to the nylon membranes as serial dilutions (200ng/µl- 5ng/µl) in duplicates.

After hybridization with complex probes including the indicated amounts of spiked controls, signal intensities for the controls were quantified and plotted against the amounts of synthetic poly(A)$^+$RNA spiked (Figure 3.6-A). A linearity range was identified (see also chapter 3.2.1.1) and close to its lower boundary, the spiked control *nebulin* was selected to be used for normalization. The median value of the intensity

signals of all the duplicated spots belonging to its serial dilution was used in the calculations.

Plotting the intensity signal against the serial dilution of control DNA spotted (Figure 3.6-B), the behavior of the signal intensity for increasing amounts of target could be also investigated. A linear increase of the signal with the amount of spotted DNA was observed. This was not obvious *a priori*, if considered that macroarray hybridizations are known to be performed under probe limiting conditions and target excess, but was already reported by Nguyen *et al*., (1995).

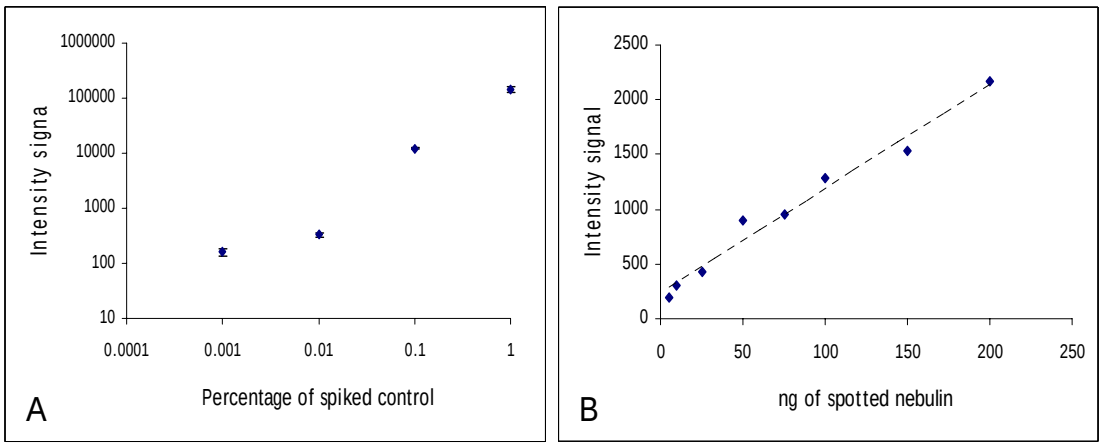


Figure 3.6 Plotting of the signal intensity depending on the spiked controls. In A) the signal intensity of the spiked controls is plotted against the percentage at which the control was spiked. This plot generated for a specific hybridization, exemplifies the general situation. Signal intensities are read for the control spotted at the concentration 50 ng/µl. B) An example of how the intensity signal changes depending on the spiked control *nebulin* for increasing concentrations of spotted *nebulin* (5-200 ng/µl) is reported.

This was the reason for which a further step was introduced in the normalization procedure: the signal intensity for each spot was divided by the relative baseline signal intensity for the labeled T7-short oligohybridization. In this way the effect of different concentrations in the spotted samples was corrected.

All the mentioned normalization steps were performed by using Excel (commercial software) macros developed for this purpose.

For microarrays, recently more complex normalization strategies based on function of intensity rather than a single factor, were developed (Quackenbush, 2002) to correct for microarray specific technical biases. Similar strategies have not been reported for nylon filter macroarrays.

3.1.4.3 - Outlier detection

The last step of the data processing was the detection of outliers. For this purpose an excel matrix with the normalized expression values was produced. For each clone spotted, eight normalized intensity values (double pattern times four hybridizations) were reported for each of the three tissues analysed. Clones were listed in rows, intensity values in columns.

Outliers are extreme values in the distribution of replicates and they are revealed only by the extreme deviance of their expression value as compared to other replicates. Undetected outliers bias the estimation of both the expression value and its associated random error, compromising tests for differential expression. Large sample sizes are needed to detect outliers accurately and precisely.

Outlier detection was performed applying the software ArrayStat. It requires a prior estimation of the random error. The "Pooled: Curve-fit" based random error estimation was used to calculate the random error. In this approach, estimates of the random error can be obtained by pooling error variance across probes locally according to expression intensities.

Based on this estimation, outliers were detected among replica when the deviance of a specific value from other values was exceeding the random error, and eliminated from further analysis.

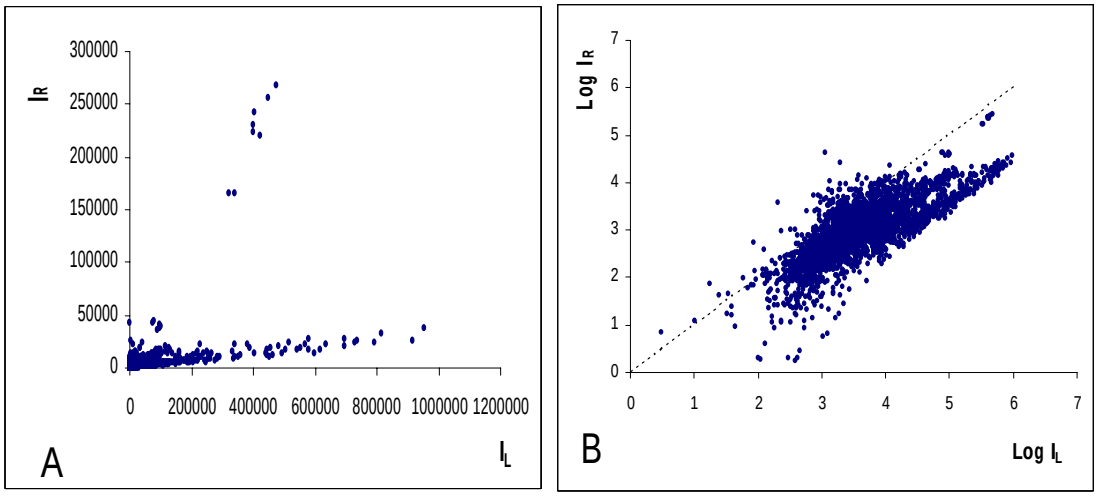


Figure 3.7 Scatter plots of the raw intensity data (A) and of the logarithmic transformed intensity (B) data for hybridizations with leaf and root probes.

3.1.4.4 - Logarithmic transformation

Finally a logarithmic transformation of the data was applied by the ArrayStat software. This implies two advantages: it reduces the proportional relationship between random error and signal intensity, and it generates a normal distribution of the replicated expression values. This effect is seen in Figure 3.7 where a scatter plot of raw data from a single comparison between tissues and a scatter plot of the same data after logarithmic transformation are reported.

3.1.5 - Data analysis

The general purpose of the statistical data analysis is to detect if there is a reliable, biologically relevant difference in expression levels of different samples. Difference is considered to be due to a biological component and to an error component, which is divided into systematic and random error. Systematic error is due to biases and is mainly corrected by background subtraction and normalization procedures. On the contrary, random error is a measure of the uncertainty in the measurement. It cannot be eliminated, but it can be estimated from observed data. The estimation applied here was performed as explained in the section 3.1.4 and it was already employed to detect outliers.

Differentially expressed genes were inferred by a fixed “threshold cut off method” (two-fold increase or decrease). An empirically set threshold always bears the risk of generating false positive or false negative results. To overcome this problem recently statistical tests have been introduced to evaluate the significance of the difference in expression. Here the *z*-test was applied to pairwise comparisons between tissues. This test cannot be applied to experiments with more than two conditions if the random error is estimated using the “Pooled: Curve fit” method. Applying this test, observed differences in expression between two tissues that exceed a threshold defined jointly by random error and by the probability of a false positive were considered “statistically significant”, a minimum requirement for biological significance. To set the false-positive rate for the statistical test, $\alpha < 0.05$ was chosen in advance. When large numbers of statistical tests are conducted, as in macroarray analysis, a procedure

to correct the false positives is recommended (Nadon and Shoemaker, 2002). The *Stepdown Bonferroni* correction procedure (Hochberg, 1988) was applied here for this purpose.

From the results obtained applying the described statistical test, a cut off value for the ratio value was estimated and then applied to the analysis performed on three tissues. A ratio for the expression values in two organs of at least two was showing comparable results to the statistical test applied in the pairwise comparison.

3.2 - EVALUATION OF THE PERFORMANCE OF SUGAR BEET MACROARRAYS

3.2.1 - Sensitivity of the macroarray hybridization

3.2.1.1 - Checking sensitivity on controls

The sensitivity of the macroarray system can be defined as the minimum detectable level of poly(A)$^+$RNA. It means that a poly(A)$^+$RNA species present at least at this level can be quantified.

First indications about the possibility to use the macroarray technology in a quantitative way were reported by Zhao *et al*. (1995). These authors were also the first to report an amount of mRNA equal to 0.01% of the mRNA used for probe synthesis as sensitivity limit for this technology. This value is 10 times lower than that previously reported by Gress *et al.* in 1992 and Sargent *et al*. (1987). Employing a similar system to that described in the cited publication by Zhao, the sensitivity limit of the two different batches of filters produced was estimated. Using the spiked controls and the intensity values derived from spots which were printed from a 50 ng μL^{-1} solution (see section 3.1.4.2), the control spiked at 0.01% of the poly(A)$^+$RNA used for probe synthesis, was found to produce a signal intensity exceeding the "unspecific hybridization signal" in all cases. In this context "unspecific hybridization signal" is defined as the highest intensity signal among signals from empty plasmids and signals from the amplified insert of the plasmid pAW109. This result was

confirmed with both filters batches for the four hybridizations using leaf samples as probes (Table 3.1).

Assuming that about 100.000 mRNA molecules are present in a cell (Bishop *et al.*, 1974) the results described above show that this system is able to detect transcripts present in as little as 10 copies per cell.

Table 3.1 Raw intensity signal for unspecific hybridization as compared to hybridization signal for *nebulin* (spiked as 0.1% of the poly(A)$^+$RNA used for probe synthesis). Data are derived from hybridizations of probes generated from leaf material (L1, L2, L3, L4 as in Figure 3.3) to both filters sets produced: RZPD filters, produced by RZPD in Berlin and MPIZ filters produced in collaboration with ADIS unit at MPIZ.

Filter set	Signal type	L1	L2	L3	L4
RZPD filters	Unspecific hybridization signal (MCD)	1045.71	567.55	436.19	714.63
	Signal of the control spiked as 0.01% (MDC)	2272.95	1053.24	584.65	2217.56
MPIZ filters	Unspecific hybridization signal (MDC)	1153.67	560.55	491.65	1452.90
	Signal of the control spiked as 0.01% (MDC)	3140.69	1361.48	2488.80	2787.56

Proportionality between signal intensity and amount of labelled transcripts is a prerequisite for quantifying expression levels in comparative analyses. Using spiked controls it was shown that transcripts present in 0.01 to 0.5% of the poly(A)$^+$RNA used for probe synthesis were leading to proportional intensity signals (for an example of data used to estimate this range see Figure 3.6-A). At higher levels in many cases a saturation of the signal was observed, therefore for poly (A)$^+$RNA represented at levels higher than 0.5%, the abundance could be underestimated. Also for transcripts represented at levels lower than 0.01% of the probe a levelling-off of the signal was observed, in this case it was mainly due to the contribution of the background. The signal intensities in the range in which proportionality was observed (referred to as proportionality region from now on) were covering between two and three orders of magnitude. The percentage of clones with expression levels in this range was estimated for all the hybridizations performed on filters produced by RZPD. A case study is given in Figure 3.8. The distribution of the signal intensity is reported and the "proportionality region" as defined above is indicated. It is evident that the majority of the clones fall within this range. Considering hybridizations with leaf and root probes, on average 85% of all clones are covered in the proportionality region. When

probes prepared from inflorescences were used only 75% of all clones produced a signal in this range. This was expected because the library A006 was produced only from leaf and root tissue, and did not include inflorescences.

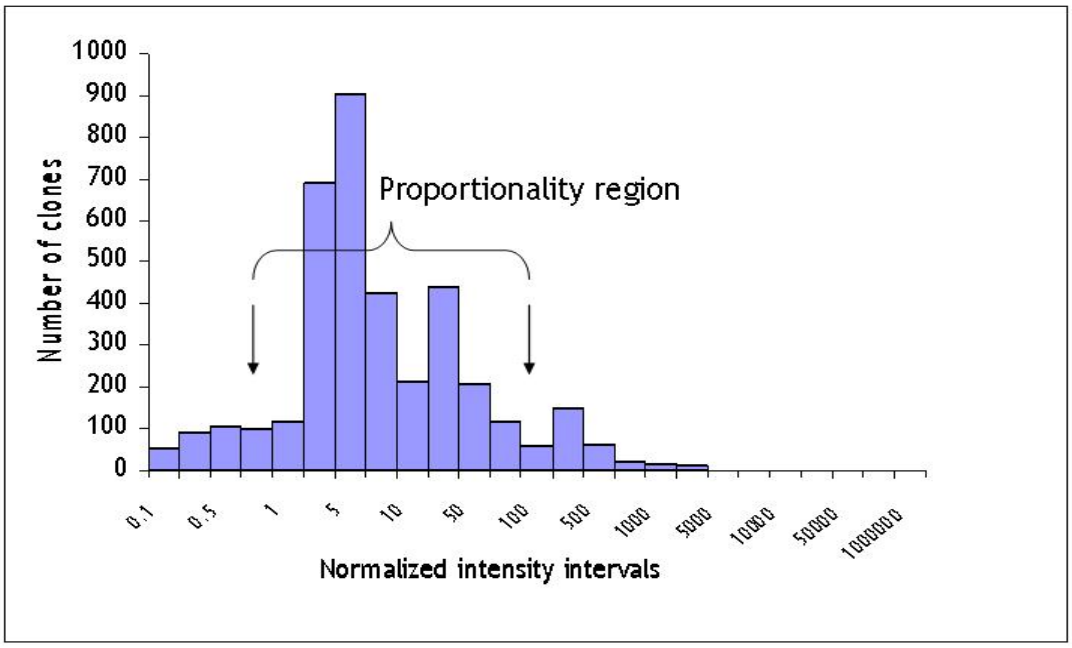


Figure 3.8 Distribution of the signal intensities for all clones hybridized with a probe produced from poly(A)$^+$RNA extracted from leaves. The signal intensity for which a proportionality between signal intensity and amount of mRNA was identified is indicated. This range was estimated on the signals obtained for the spiked controls.

3.2.1.2 - Checking sensitivity on cDNAs: expression analysis of selected RGAs (Resistance genes analogues)

Rare mRNA species are known to be expressed at an average of 10 copies per cell and they are estimated to share more than half of the total mRNA species (Bishop *et al*., 1974, Jendrisak *et al*., 1987). They represent an important part of the mRNA population present in a cell, but their intensity signals are close to the limits of sensitivity of this technology.

To determine the sensitivity of the system set up for sugar beet, intensity signals for a set of genes known to be transcribed at low levels were evaluated. For this purpose, EST sequences (see section 4.1) with similarity to known resistance genes (R-genes) or disease-resistance proteins were analyzed. For a number of proven resistance genes it is known that they are not highly transcribed (Dixon *et al*., 1996; Jones *et al*., 1994).

Table 3.2 Normalized expression values for 29 identified R ESTs revealed by macroarray experiments performed with RZPD filters and with MPIZ filters. If expression values were lower than the sensitivity level of the system no value is reported. Ratio values exceeding the threshold of two-fold increase or decrease are reported in bold. Additionally, ratios over the 1.5-fold increase or decrease threshold and confirmed independently in both experiments are considered in this case for differential expression. Preferential expression in root is highlighted in red and preferential expression in leaf is highlighted in green.

		RZPD filters			MPIZ filters		
Coordinates	**Annotation**	**Expression value in leaf**	**Expression value in root**	**R/L**	**Expression value in leaf**	**Expression value in root**	**R/L**
M - 2, 2	receptor-like protein kinase 1	-	-	-	-	-	-
D - 12, 1	serine/threonine protein kinase-like protein	-	-	-	-	-	-
L - 4, 6	PUTATIVE KINASE-LIKE PROTEIN TMKL1	-	-	-	-	-	-
B - 17, 7	leucine-rich receptor-like protein kinase	-	-	-	-	-	-
M - 20, 7	receptor protein kinase-like	-	-	-	-	-	-
L - 14, 1	receptor protein kinase-like	-	-	-	-	-	-
O - 14, 1	hypothetical protein	-	-	-	-	-	-
G - 24, 9	hypothetical protein	-	-	-	-	-	-
F - 9, 1	putative disease resistance response protein	-	-	-	-	-	-
H - 4, 8	NBS-LRR-like protein	2.61	2.92	1.12	-	-	-
A - 24, 7	Pto kinase interactor	-	-	-	13.03	12.38	0.95
D - 17, 1	protein kinase	-	-	-	-	2.87	-
E - 14, 3	protein kinase-like	-	3.34	-	2.82	5.57	**1.97**
H - 14, 7	putative receptor-like protein kinase	-	2.39	-	-	2.29	-
D - 12, 5	wall-associated kinase 1	10.97	35.54	**3.24**	17.22	91.89	**5.34**
M - 11, 7		4.59	6.90	1.50	3.79	6.89	1.82
L - 1, 7	protein kinase-like	2.50	3.37	1.35	2.47	3.94	1.60
F - 3, 4	hypothetical protein	2.80	3.54	1.27	3.97	4.22	1.06
M - 1, 8	hypothetical protein	3.28	4.14	1.26	3.74	5.12	1.37
G - 14, 2		6.00	6.94	1.16	9.15	7.59	0.83
E - 20, 5	putative receptor-like protein kinase	20.39	22.44	1.10	37.70	35.05	0.93
P - 9, 4	UVB-resistance protein-like	3.19	3.47	1.09	-	5.94	-
M - 7, 4	leucine-rich receptor-like protein kinase	3.94	4.04	1.03	4.06	5.80	1.43
C - 4, 5	unknown	4.78	4.60	0.96	10.14	5.20	0.51
C - 5, 8	putative protein kinase	3.47	3.31	0.95	4.89	3.57	0.73
A - 20, 10		3.64	3.45	0.95	4.16	4.80	1.16
L - 18, 4	protein kinase homolog	7.57	3.94	0.52	3.09	4.50	1.46
J - 14, 9	receptor-like protein kinase1	26.84	4.03	**0.15**	17.77	4.38	**0.25**
D - 6, 2	hypothetical protein	11.17	-	-	19.12	-	-

ESTs were classified R-ESTs if their corresponding amino acid sequence showed at least 38% similarity to the deduced protein sequence of genes associated with disease resistance in the aligned region. In the same way protein kinase related ESTs were considered only if they showed at least 38% similarity to *Pto* or *Xa21*, which are protein kinases involved in disease resistance. A total of 29 R ESTs were identified among the 2996 EST sequences. Details about their regions of similarity to resistance gene products are reported in Hunger *et al*. (2003).

Macroarray data for the expression levels of these 29 cDNAs in leaves and roots are reported in Table 3.2 and have already been published in Hunger *et al*. (2003). Nine cDNAs showed intensity signals below the sensitivity limit with both batches of filters considered. Therefore they were discarded from the analysis. For the filters printed by RZPD, two more cDNAs failed the sensitivity limit, but for the MPIZ filters only one additional cDNA had to be excluded for this reason.

For seven RGAs a differential expression concerning leaves and roots was detected. Three RGAs were confirmed to be preferentially expressed in the roots and two in the leaves in both experiments independently. For an additional clone, clone M-11,7, preferential expression in roots was assumed although the ratio was just above an 1.5-fold increase for both filter sets. The reason to accept preferential expression in this case was the high reproducibility of the results in the two independent experiments. A fifth clone, clone D-17,1, was considered preferentially expressed in roots based on data from MPIZ filters only.

3.2.2 – Reproducibility of the macroarray hybridization

Hybridization signals are influenced by many parameters, like the physical properties of the membrane, the immobilized targets or the composition of the probe. Replicates at different experimental levels can be used to quantify the variability due to each of these parameters. For this purpose the number of genes showing low reproducibility between the different replicates considered was quantified.

As already shown in Figures 3.5 and 3.7 a simple tool to visualize the comparison and the reproducibility of two hybridization experiments is the scatter plot. The intensity of every spot in experiment 1 is plotted against its intensity in experiment 2. The

larger the deviation of a gene from the diagonal (x=y) the more significant is the variation between the measurements in the two experiments.

The first level of repetition concerned the variability between duplicated spots on the same membrane defined as "intra-filter variability". Signal intensities obtained from double spots of the same clone in nearby positions of the same 5 x 5 subgrid were compared (representative example of one filter shown in Figure 3.9-A).

Based on a total of four hybridizations, the percentage of spots showing greater than two-fold variation was found to be 1.1±0.5%.

As second level of repetition, signal intensities from spots of the same clone on different filters were compared to define a "inter-filter variability". A typical comparison of one filter pair is illustrated in Figure 3.9-B. Evaluating the signals of the same clone on four different filter pairs resulted in a variation of 10.2±1.0%. This value is 10 times higher than the "intra-filter variability" and indicates the technical limits.

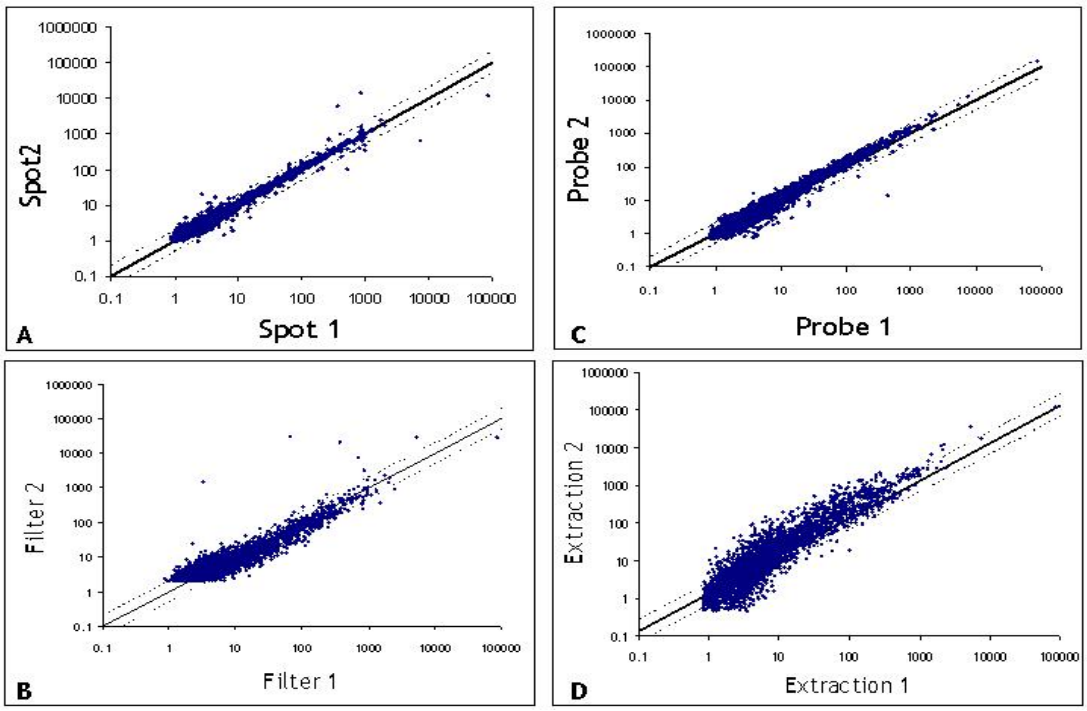


Figure 3.9 Scatter plots reveal normalized signal intensities obtained for the same cDNA clone (A) spotted on the same filter in duplicate, (B) spotted on two different filters, (C) after hybridizations with two different probes prepared from the same poly(A)$^+$RNA extraction, and (D) after hybridization with two probes prepared from two different samples. Hybridization probes were prepared from poly(A)$^+$RNA isolated from sugar beet leaves. Solid lanes represent diagonals (y=x) for coincidence of data, and dashed lines indicate a two-fold deviation. In each experiment all 3840 cDNA clones except those with signals below the sensitivity limit were assessed.

A further level of repetition is probe synthesis starting with the same RNA sample. The composition of the probe is dependent on the sample from which poly(A)$^+$RNA is extracted, on the extraction of poly(A)$^+$RNA and on the labelling reaction. The hybridization results obtained with two different probe preparations of the same extraction varied in 2.4±0.6% of all clones more than two-fold when four samples were considered (Figure 3.9-C).

Finally the use of two different samples of the same organ for extraction and probe preparation ("biological replication", see section 3.1.2) introduced at least two-fold variation in 30.3±3.7% of all clones based on the evaluation of eight such comparisons (example in Figure 3.9-D). This result is attributable to adaptive responses of the field-grown plants to the environmental conditions and changes in the plant developmental program between the two sampling dates. As it was the aim to retrieve candidates with more general importance in root morphology and physiology, only genes showing constitutive and relatively high transcription in the root were the targets. To reduce the bias for a particular sample, we considered the data for each organ when six replications passed the statistical requirements as outlined in materials and methods. As four values (including duplicated spots on membranes) were related to one sample and the next four to another independent sample (biological replica), this implies that expression values should show similarity across "biological replica" to be considered in further analysis.

3.2.3 – Data on cross-hybridization

In macroarray analysis, cross-hybridization is a potential source of error, especially if it is considered that 65% of the genes in *Arabidopsis thaliana* are members of gene families, and the percentage is expected to be higher for other plant organisms. Values for cross-hybridization have been estimated for microarrays (Girke *et al*., 2000). It was found that sequences only cross-hybridize under the given experimental conditions if their identity exceeds 70-80%.

Cross-hybridization has not been assessed in detail in the presented study, but there are hybridization data on the spotted amplified cDNA for the gene *Inf1* from *Phytophthora infestans* and *Inf2A* which shows 60% sequence identity to *Inf1*. A

residual signal intensity for *Inf1* of 5% of the signal intensity of *Inf2A* was detected when no poly(A)$^+$RNA for *Inf1* was added for the synthesis of the probe.

4 - SELECTION OF CANDIDATE GENES BASED ON PREFERENTIAL EXPRESSION IN THE BEET

This chapter reports how the established macroarray technology was used to analyse an EST collection of 3840 clones to identify candidate genes for the sucrose accumulation process preferentially expressed in root.

As sucrose accumulates in the root of sugar beet plants genes specifically involved in this process are assumed to be expressed in root at the same time.

Therefore the macroarray technology was applied to classify the sugar beet transcripts with respect to their expression in three different organs, the root, the leaf and the inflorescence. Seventy-six transcripts with preferential expression in root cells were identified and analysed in more detail with respect to their function.

Expression data were confirmed by two different experimental approaches.

4.1 – EXPRESSION ANALYSIS BASED ON EST DATA (*)

The cDNA library A006 analyzed in this experiment was generated from young shoots and roots of sugar beet plants and consists of 3840 cDNA clones (see section 2.1.2). The data for 2996 EST sequences were incorporated into the data processing pipeline of the integrated software package GenAgent (Bellin *et al*., 2002). Cluster analysis on the ESTs dataset was performed to asses the redundancy of each EST in the dataset.

All sequences which showed at least 96% identity in a fragment of at least 50 nucleotides were clustered using the software StackPack (Miller *et al*., 1999, Burke *et al*., 1999). Among 2996 sequences, 405 clusters with two to 40 members were identified, and 1643 sequences (54.8%) were classified singletons.

(*) The sequencing and clustering data to produce this part of the results have been obtain from the candidate from the database Genagent under the supervision of the DR. Martin Werber which is kindly acknowledged

Table 4.1 List of most redundant ESTs according to cluster analysis. For each of the 36 largest clusters, the table lists a representative annotation, the cluster size and the relative expression pattern of cluster members (preferential expression in l: leaf; r: root; i: inflorescences), n.p.: not preferentially expressed; n.a.: not analysable due to failure of statistical requirements of data in at least one tissue or due to inconsistencies of expression values among the members of the cluster.

Annotation	**Cluster size (number of ESTs)**	**Preferential expression**
Ribulose bisphosphate carboxylase, small chain 1, chloroplast precursor	40	l
Alpha-Tubulin, chain 2	32	r, i
S-adenosylmethionine synthetase1	25	n.p.
Jasmonate-induced protein homolog	27	l
Benzothiadiazole-induced protein	23	l
Elongation factor-1, alpha subunit	21	n.p.
Fructose-bisphosphate aldolase precursor, chloroplast	18	l
RNA helicase-like protein	15	l
Carbonic anhydrase, chloroplast	14	l
Rubisco activase, chloroplast	13	l
Glutamine synthetase GS2, chloroplast	11	l
Jacalin homolog	10	l
Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	10	n.p.
Chlorophyll a/b-binding protein, chloroplast	9	l
Malate dehydrogenase, cytosolic	9	n.p.
Reversibly glycosylated polypeptide	8	r
Chlorophyll a/b binding protein, chloroplast precursor	8	l
Xyloglucan endo-1,4-beta-D-glucanase, brassinosteroid-regulated protein	8	r, i
S-adenosyl-L-homocystein hydrolase	8	r, l
ATP synthase beta chain, mitochondrial precursor	8	l
Putative nematode-resistance protein	8	n.a.
Triose phosphate translocator, chloroplast precursor	7	l, f
PSI type III chlorophyll a/b-binding protein, chloroplast	7	l
Chlorophyll a/b binding protein, chloroplast precursor	7	l
Glyceraldehyde-3-phosphate dehydrogenase (NADP+) subunit B precursor, chloroplast	7	l
Sucrose synthase	7	r
Phosphoglycerate kinase precursor, chloroplast	7	l
Putative preprocysteine proteinase	7	i
ABC transporter homolog	7	n.p.
Choline monooxygenase, chloroplast	6	l
Chlorophyll a/b binding protein, chloroplast precursor	6	l
Probable peroxidase	6	n.a.
Oxygen-evolving enhancer protein 2, chloroplast	6	l
Eukaryotic initiation factor 4A-9	6	n.a.
Germin-like protein	6	l
DnaJ homologue	6	r

Taken together, the analysed ESTs specified 2048 unique expressed sequences under the conditions selected.

The 36 largest clusters with at least six members each are listed in Table 4.1. They consist of 423 cDNA clones representing 14.1% of all ESTs.

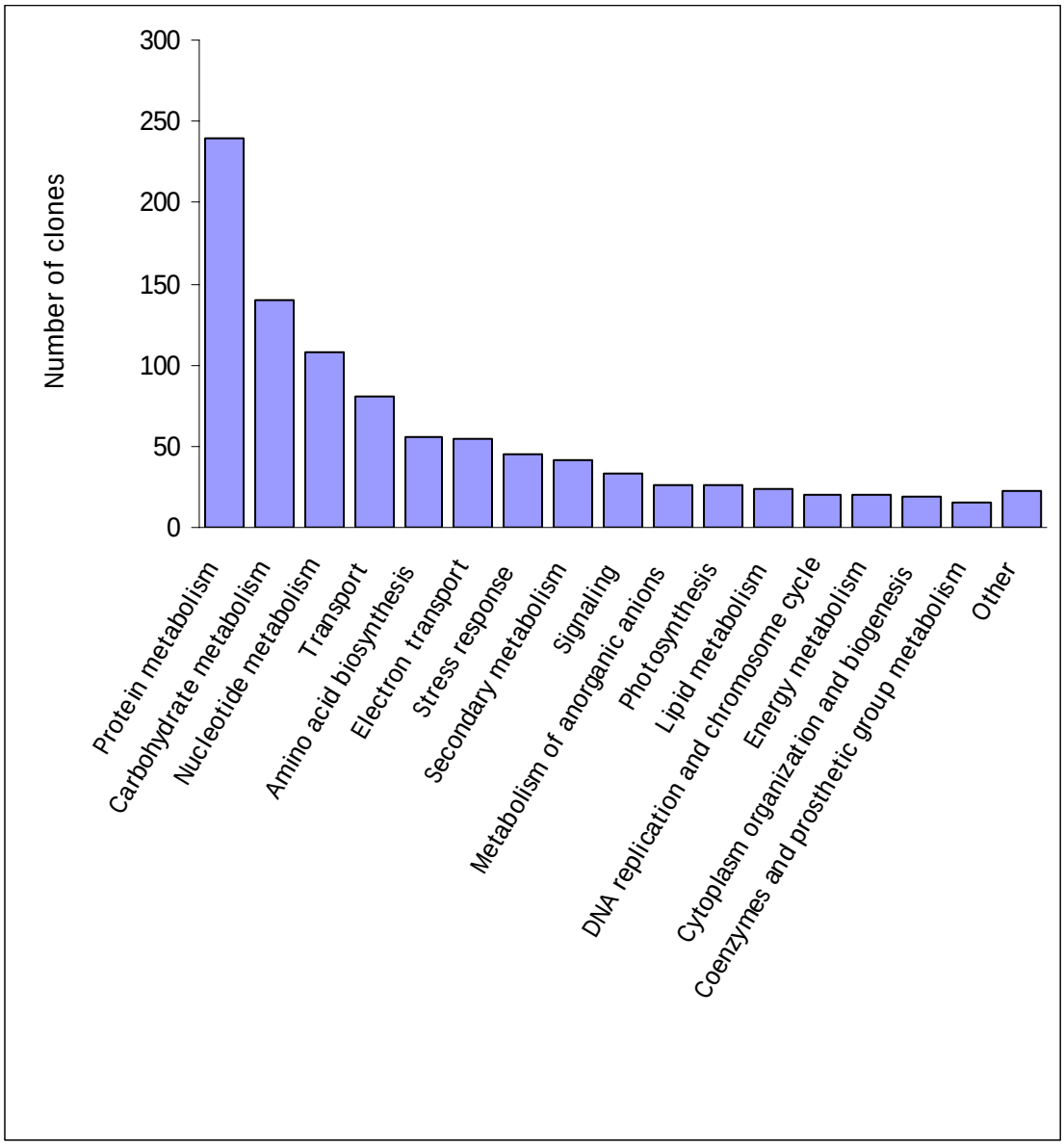


Figure 4.1 Functional categories of sugar beet gene products with predicted functions deduced from 973 non-redundant EST sequences. Column height corresponds to the number of members for each category as indicated on the y-axis. Another 1075 unique sequences could not be assigned.

Sequence data also served to evaluate the GC-content. Considering 2911 sugar beet 5'-EST sequences out of the 2996, an average GC-content of 42.7% was determined. The values for average exon GC-content in *Arabidopsis thaliana* and rice were 43.2 and 51.4%, respectively (The Arabidopsis Genome Initiative, 2000, Yu *et al*., 2002). Additionally, the contribution of the new EST sequences to the public databases was evaluated. According to a cluster analysis of previously published sugar beet ESTs in

dbEST NCBI (http://www.ncbi.nlm.nih.gov/dbEST/ as in June 2002) alone and together with this new data, the EST collection here reported contributed 1732 new unique sugar beet sequences.

The ESTs of the library A006 were classified according to their function. To relate the spectrum of presumed gene functions - as identified by BLASTX searches with expect values of at least e^{-10} (Altschul *et al.*, 1990) - to the comprehensive physiological activity of the cell, gene products with known functions were categorized according to biological processes (The Gene Ontology Consortium, 2000). After manual editing and unification of related subgroups, 973 gene products deduced from non-redundant EST sequences were assigned and fell into 17 different categories (see Figure 4.1), representing all major cellular activities like metabolism of proteins, carbohydrates and nucleic acids as well as transport and signalling processes. The size of categories representing activities in primary metabolism, e.g. for protein and carbohydrate synthesis, reflected especially the role of these processes in young developing sugar beet plants. For 1075 unassigned unique sequences either no analogous was found or analogous genes without known function were retrieved from the databases.

4.2 – MACROARRAY EXPRESSION ANALYSIS

4.2.1 – Macroarray analysis: experimental design and parameter selected for the data analysis

Macroarray analysis as described in chapter 3 was performed with the cDNA library A006 to classify the sugar beet transcripts with respect to their expression in three different organs: leaf, root and inflorescence. The results presented here were generated with the batch of filters printed by RZPD (Berlin, Germany).

4.2.2 - Differential gene expression in three different organs

Macroarray hybridization data of library A006 were integrated with the results of the cluster analysis and the variability concerning the results of the differential expression

analysis among cluster members was considered in order to evaluate the reproducibility of the results. Consistency in the expression patterns among cluster members was found for 90% of the clusters and the clusters not showing reproducible expression among members were not considered for further analysis.

Table 4.2 Results obtained from the comparison between different organs based on the statistical evaluation of the normalized expression data derived from eight replica for each tissue. For the clusters showing reproducibility among members, results were reported only once. Differential expression was estimated as explained in chapter 3.

	Number of clones (percentage)
Total unique sequences	2048 (100%)
Discarded for statistical reasons	323 (15.77%)
No differential expression	808 (39.45%)
Differential expression	917 (44.78%)

The distribution of clones that either failed the statistical requirements, showed or did not show differential expression are given in Table 4.2. As a result, a total of 917 unique sequences were identified as differentially expressed among the 2048 unique sequences considered in the analysis.

The preferential expression of these 917 unique sequences is illustrated in Figure 4.2. Unique sequences preferentially expressed in one of the three organs considered were identified by their ratio values.

Additionally, three more classes were introduced that contain unique sequences with at least two-fold higher expression in two organs with respect to the third (preferential expression in leaves and root, preferential expression in leaves and inflorescences, preferential expression in root and inflorescences).

In summary, 362 unique sequences showed preferential expression in leaves, 76 showed a preferential expression in roots and 23 in inflorescences. Concerning preferential expression in two organs with respect to the third one other 424 unique sequences were preferentially expressed in leaves and roots, 20 unique sequences were preferentially expressed in leaves and inflorescences and 12 in roots and inflorescences.

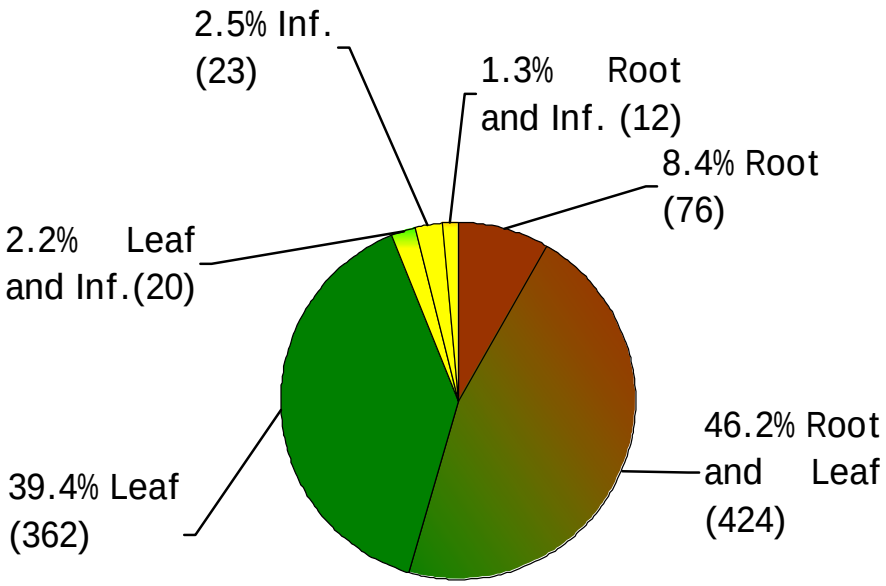


Figure 4.2 Distribution of the ESTs according to preferential expression in the different organs. In each class the expression values for the respective organ(s) were at least twice as high as those of the other organ(s). Percentages are indicated, and absolute numbers of clones are given in brackets.

Further details on expression ratios values for all analysed cDNA clones are available from http://www.mpiz-koeln.mpg.de/sugarbeet.

To obtain some general information on the functions of gene products preferentially expressed in leaf, root or inflorescence, the established categories for biological processes (The Gene Ontology Consortium, 2000) were assigned to 184 of 362 preferentially leaf-expressed ESTs, to 27 of 76 preferentially root-expressed ESTs and to 12 of 23 preferentially inflorescence-expressed ESTs**.** As the number of ESTs assigned per organ was uneven and relatively small, only major and obvious differences are mentioned. As expected, gene products involved in photosynthesis were exclusively found in leaves. Gene products with a putative function in electron transport were overrepresented in leaves with respect to roots, reflecting the complexity of electron transport in photosynthetic membranes of leaf cells. Among the preferentially root-expressed ESTs the categories for cytoplasm organization and biogenesis, nucleotide metabolism, stress response and transport showed an increased relative abundance.

4.2.3 – Sequence analysis and functional classification of the 76 preferentially root expressed cDNAs

As it was the aim to retrieve candidate genes for the sucrose accumulation process which is specific to the root of sugar beet plants, the 76 preferentially root-expressed ESTs identified in the macroarray analysis were considered in more detail.
The manual classification of these ESTs according to predicted functions was performed on annotations retrieved using expect values of maximally e^{-4} for sequence similarities based on the deduced amino acid sequence. This means that the threshold for the expect values was elevated e^{6}-fold as compared to the threshold used to retrieve the automatic annotations to group the ESTs according to the Gene Ontology functional categories. The reason for using expected values of maximally e^{-4} was to allow the detection of more subtle sequence similarities. Results with weak scores have to be considered with care, and even sugar beet analogues with strong similarities to known sequences may fulfill a different function in sugar beet.
In this way, analogous sequences were retrieved for 53 ESTs from the databases. Putative functions were predicted for 43 of them. A further 23 sugar beet ESTs did not show similarity to any other sequence under these conditions. Therefore they may represent sugar beet specific sequences.
A possible classification of these 43 preferentially root expressed ESTs in ten groups according to their putative function is presented in Table 4.3.
The first group summarizes six gene products with putative functions in carbohydrate metabolism. Sequences with similarities to sucrose synthase, for which cluster analysis identified four different types, showed preferential expression in roots. Interestingly, two qualitatively different sucrose synthase-expression patterns were observed with respect to transcription in inflorescences. The other two gene products encode an alcohol dehydrogenase analogous to a gene product identified in grape berries (Tesnière and Verriès, 2000) and a putative phosphoenolpyruvate carboxylase kinase, which was previously identified as root-expressed in sugar beet by Kloos *et al*. (2002).
In the second group five clones showed homologies to gene products involved in the transfer of sugar moieties. The activity of the two different glucosyltransferases is either required for the biosynthesis of cell wall polymers or for regulatory functions in

secondary metabolism like the transfer of UDP-glucose to aglycons including plant hormones and xenobiotics (Keegstra and Raikhel, 2001). A similar function in regulating substrate activity may be fulfilled by alpha-mannosidase which is involved in the catabolism and turnover of N-linked glycoproteins in the vacuoles. Two different xyloglucan endotransglycosylases identified in sugar beet roots are possibly involved in hydrolyzing xyloglucan chains in localized areas of the cell wall (Campbell and Braam, 1999).

The third group includes preferentially root-expressed gene products involved in the biosynthesis of the primary and secondary cell wall such as arabinogalactan proteins, extensins and reversibly glycosylated polypeptides. An association with strong vascular elements is also likely for the alanine and glutamic acid rich protein. Its transcript shows a much higher expression in roots with respect to the softer tissues of leaf and flower. The extremely high expression of this transcript in stems (Kloos *et al*., 2002) is in accordance with a possible role in the vasculature.

Among the preferentially root-expressed ESTs, a sequence with similarity to nodulins was identified. Nodulin-related gene products, putatively located in the cell wall, were also found among ripening-induced cDNA clones from grape (Davies and Robinson, 2000). Concerning the two different peroxidases, which were found preferentially expressed in sugar beet roots, they either play a role in the oxidation of phenolic compounds in the cell wall or function in other oxidative processes.

Predicted preferentially root-expressed gene products of the fifth group are associated with cytoskeletal reorganization, such as an alpha- and a beta-tubulin, an annexin and a DnaJ protein, which possibly interacts with the cytoskeleton. Ripening-associated tubulins have also been reported from strawberries (Aharoni *et al*., 2002) as has an annexin (Wilkinson *et al*., 1995).

For five deduced root-expressed gene products functions in intra- and intercellular transport and transfer processes are predicted. Among these are two aquaporins, which regulate water homoeostasis and are indicated by the accession numbers BQ488238 and BQ488455 (Yamada *et al*., 1995). The group also comprises an amino acid transporter-like protein, a putative lipid transfer protein which can transfer phospholipids across membranes, and a potential ADP-ribosylation factor 1, which is thought to play a role in intracellular vesicle transport.

Table 4.3 List of preferentially root-expressed EST clones grouped according to putative functions. GeneBank accession numbers for (a) sugar beet ESTs, their annotations and (b) the accession numbers of the protein sequences with highest similarities including the respective e-values are given. In the last two columns, macroarray expression ratios for root/leaf and root/inflorescences are reported. Clones marked as not analysable (n.a.) with respect to the ratio values were manually included if expression in roots was consistently high, but very low expression in leaves and/or inflorescences did not allow an automatic calculation of expression ratio values.

Acc. N° ESTs Sequences [a]	Annotation	Acc. N° A.A. Sequences [b]	E-value	Ratio R/L	Ratio R/I
CARBOHYDRATE METABOLISM					
BQ490013	Sucrose synthase	CAA57881	1.00E-109	8.78	3.13
BQ489399	Sucrose synthase	Q42652	2.00E-75	4.87	na
BQ489472	Sucrose synthase	CAA57881	5.00E-54	9.39	3.61
BQ490130	Sucrose synthase - beet	S71493	3.00E-79	3.73	13.37
BQ489637	Alcohol dehydrogenase 2	AAG01382	2.00E-60	7.01	17.93
BQ488374	Putative phosphoenolpyruvate carboxylase kinase	CAC43293	1.00E-109	2.06	2.63
TRANSFER OF SUGAR MOIETIES					
BQ488698	Putative glucosyltransferase	NP_180375	6.00E-25	na	na
BQ490448	Glucosyltransferase-like protein	NP_197666	2.00E-64	2.27	4.12
BQ487564 - BQ487565	Xyloglucan endotransglycosylase XET2	AAF80591	6.00E-53	3.45	3.89
BQ654409	Xyloglucan endotransglycosylase	AAF80591.1	1.00E-92	2.03	3.5
BQ488084	Alpha-mannosidase	NP_201416	1.00E-65	3.38	5.88
CELL WALL ARCHITECTURE					
BQ489853	Reversibly glycosylated polypeptide	CAA77235	5.00E-86	3.37	2.09
BQ490217	Arabinogalactan protein	CAC16734	1.00E-05	5.41	2.55
BQ488391	Extensin	S20790	3.00E-04	3.47	6.05
BQ489156	Alanine and glutamic acid rich protein	CAC43296	1.00E-08	239	14.79
BQ489314	Nodulin-like protein	NP_565111	5.00E-54	2.44	na
OXIDATIVE PROCESSES					
BQ488951	Peroxidase	NP_201440	2.00E-77	3.11	2.73
BQ488942	Peroxidase	T10790	3.00E-52	4.53	4.02
ORGANIZATION OF CYTOSKELETON AND MEMBRANE ASSEMBLY					
BQ490546	Tubulin alpha-5 chain-like protein	NP_197478	2.00E-36	2.87	2.19
BQ489740	Tubulin beta-1 chain	Q9ZRB2	6.00E-70	3.1	3.44
BQ488829	Annexin	AAF01250	5.00E-61	2.04	4.7
BQ488834	DnaJ protein homolog	Q04960	3.00E-69	2.65	2.23
INTRA- AND INTERCELLULAR TRANSPORT AND TRANSFER PROCESSES					
BQ488238	Plasma membrane major intrinsic protein 2 - beet	T14600	4.00E-90	3.13	2.21
BQ488455	PM28B protein	CAB56217	1.00E-63	2.97	3.07
BQ489146	Amino acid transporter protein-like	NP_201400	1.00E-28	2.16	na
BQ489904	Lipid transfer protein, putative	NP_188456	2.00E-24	2.84	5.19
BQ489734	ADP-ribosylation factor 1	P51822	3.00E-73	3.22	2.01

ATP METABOLISM					
BQ489455	F1L3.21	AAF79467	4.00E-31	na	2.4
BQ488735	Vacuolar ATP synthase 16 Kd proteolipid subunit	Q39437	2.00E-23	na	na
RNA METABOLISM AND PROTEIN SYNTHESIS					
BQ488741	SOF1 protein-like protein	AAL32701	1.00E-38	2.07	2.03
BQ489160	S-like ribonuclease	AAF82615	4.00E-34	2.17	na
BQ489683	Putative ribonucleoprotein	NP_171845.1	6.00E-07	na	na
BQ490182	Ribosome-inactivating protein	BAB83507	4.00E-19	3.2	na
SIGNAL TRANSDUCTION					
BQ490595	Transcription factor	NP_564156	4.00E-18	2.56	na
BQ488907	GT-2 transcription factor	AAL65125	7.00E-23	na	na
BQ488580	Putative auxin-repressed protein	AAB88876	8.00E-10	2.87	4.94
BQ490464	Auxin-repressed protein like-protein	AAK25768	4.00E-19	6.66	3.76
BQ488855	Jasmonate-induced protein homolog	P42764	4.00E-15	2.2	4.09
BQ490059	Jasmonate-induced protein homolog	AAA86977	4.00E-07	5.82	2.33
BQ488352	Protein phosphatase-2c	T51101	1.00E-30	2.1	na
OTHERS					
BQ488661	Globulin-like protein	AAF64423	3.00E-09	2.2	10.13
BQ489047	Cytochrom B5	AAK73138	5.00E-35	2.79	2.88
BQ488897	Translationally controlled tumor protein homolog	Q9ZSW9	2.00E-64	2.61	3.79

Two more ESTs with preferential expression in the root encode gene products involved in energy metabolism, as they show similarity to a component of vacuolar ATP synthase and F1L3.21, a putative membrane spanning Ca^{2+}-ATPase.

A further group with gene products possibly playing a role in RNA metabolism and protein biosynthesis consists of a SOF1-like protein associated with preRNA processing, an S-like ribonuclease with a putative function in RNA degradation, a ribonucleoprotein and a ribosome-inactivating protein.

The last group with relatively high transcription in roots contains seven gene products with potential function in signal transduction. Among them are two putative transcription factors, four plant growth factor-regulated proteins and a protein phosphatase-2c, which may control protein activity.

Other unrelated preferentially root-expressed ESTs comprise a globulin-like protein, a cytochrom B5, which may be involved in electron transport, and a translationally controlled tumor protein homolog, which may play a role in the cell division process.

4.3 - VALIDATION OF PREFERENTIALLY ROOT-EXPRESSED GENES

For technical validation of macroarray results there are two possible approaches: *in silico* analysis and laboratory-based analysis (Chuaqui *et al*. 2002).

The *in silico* method compares array results with information available in the literature.

Among the ESTs preferentially expressed in beet, the ESTs BQ490130 and BQ489399 were showing 99% of identity at nucleotidic level with a sucrose synthase encoding cDNA isolated from sugar beet for which a predominant expression in tap roots was reported (Hesse and Willmitzer, 1996). The ESTs BQ488374 and BQ489156 annotated as putative phosphoenolpyruvate carboxylase kinase and as alanin and glutamic acid-rich protein, respectively, were 99 and 100% identical at nucleotide level to the two cDNAs AJ309171 and AJ309174 isolated from sugar beet tap root using Suppression Subtractive Hybridization (Kloos *et al*., 2002). The ESTs BQ490572 and BQ490059 with similarity to a putative protein and to a jasmonate-induced protein homolog (expected value of 2.00 e^{-72} and 4.00 e^{-07}) were also identified as preferentially expressed in root by matching cDNA-AFLP fragments (Schneider, pers. com.)

On the other side, experimental validation of the macroarray data provides independent verification of gene-expression data. For best comparisons it is performed on the same samples as used in the array experiments.

Commonly used techniques for this purpose are semi-quantitative reverse transcription PCR and Northern blot. However, it should be noted that differential expression can only be compared at the qualitative and not at the quantitative level because differential effects are often larger with RT-PCR and Northern analysis (Taniguchi *et al*, 2001, Wurmbach *et al*., 2001).

Additionally, some non-differentially expressed sequences with a ratio value close to one should be used for technical confirmation.

In this experiment macroarray data were evaluated both by Northern hybridization with amplified cDNAs as hybridization probes, and by quantitative RT-PCR using sequence-specific primers (Figure 4.3 and Figure 4.4, respectively).

Two cDNAs encoding cytosolic glyceraldehyde 3-phosphate dehydrogenase (Acc. N$^{o.}$ BQ488127) and the beta chain of a GTP-binding protein (Acc. N$^{o.}$ BQ490211),

respectively, with root/leaf and root/inflorescence macroarray expression ratios close to one were chosen to cross-reference between the three different techniques for expression detection.

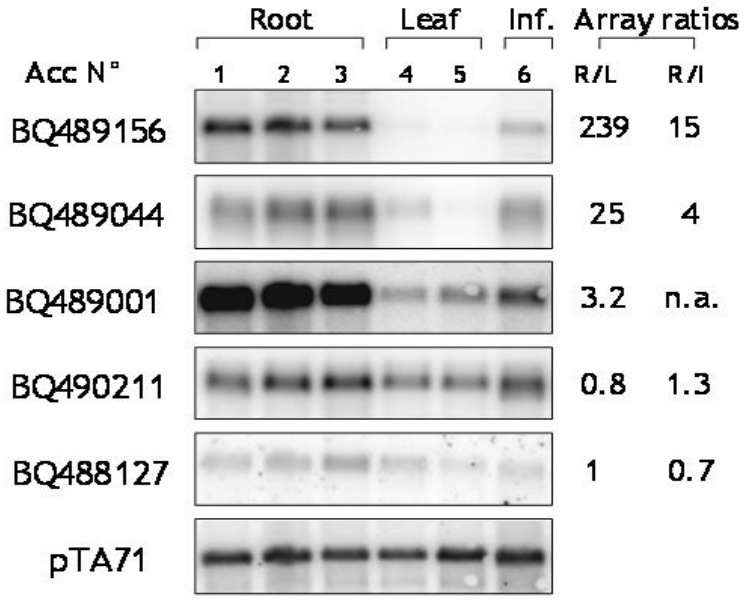


Figure 4.3 Northern analyses of selected EST clones. Inserts of cDNA clones with the indicated accession numbers were hybridized to Northern filters containing in each lane 3 µg of poly(A)$^+$RNA extracted from (1-3) three different root samples, (4, 5) two different leaf samples and (6) one inflorescence sample. In the last two columns the macroarray expression ratios root/leaf and root/inflorescence are given, for legend see Table 2.

The results of the Northern-blot experiment for these genes confirmed the expression ratio estimated by macroarray analysis and justified their use to equalize the cDNAs concentrations in the semi-quantitative RT-PCR experiment.

Northern analysis was used to validate the expression values obtained for other three genes with at least two-fold higher expression in roots than in the other tested organs, according to macroarray analysis.

The expression values of two of these and four more genes were verified by semi-quantitative RT-PCR. For RT-PCR, the template poly(A)$^+$RNA samples were the same as used for the preparation of complex probes in the array experiments. In each case, the number of amplification cycles was adjusted to obtain PCR products in a linear range. In all cases the relative expression patterns were confirmed.

Additionally, two genes with root/leaf array expression ratios of 1.4 and 1.6 were assessed by RT-PCR for differential transcription.

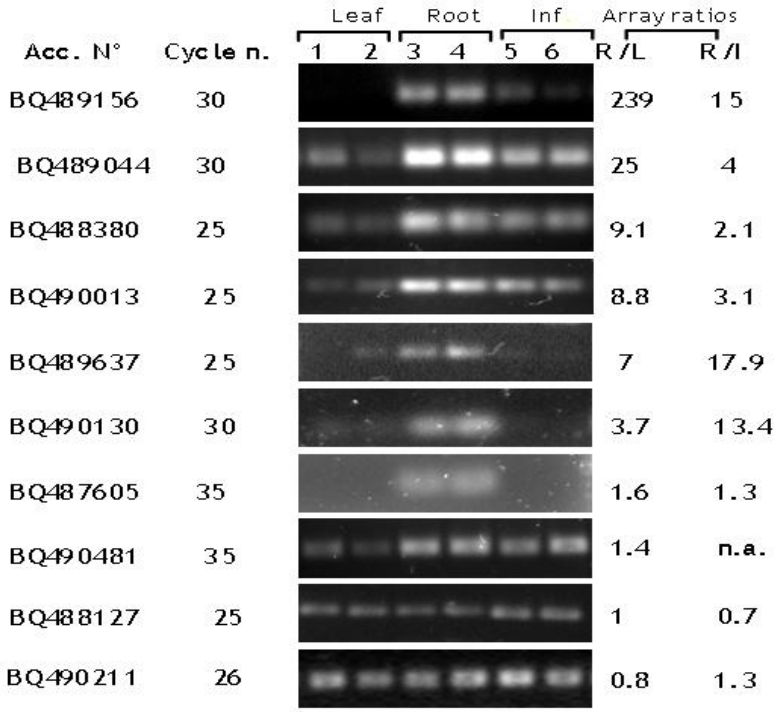


Figure 4.4 Semi-quantitative RT-PCR as validation tool for macroarray expression ratios. Semi-quantitative RT-PCR with EST-specific primers was performed based on cDNA synthesized from the same two different samples of leaf (1, 2), root (3, 4), and inflorescence (5, 6) as used for the macroarray hybridizations. Macroarray expression ratios root/leaf and root/inflorescence are indicated, for legend see Table 2.

Expression was generally low as indicated by the number of 35 cycles in RT-PCR required to visualize the transcript, but it was preferentially found in roots. This last finding supports both the sensitivity of the macroarray hybridizations and the stringency of the threshold, which was selected for automatic detection of preferential expression.

4.4 - GENETIC MAPPING OF PREFERENTIALLY ROOT- EXPRESSED GENES (*)

To validate the identified candidate genes at the genetic level segregation analysis was performed to assign the loci to chromosomal positions (see paragraph 1.2).

(*) Mapping experiment were started first by the candidate. As soon as molecular methods were established, help in mapping was obtained by the T.A. Susanne Schwarz which performed the remaining necessary experiments under the supervision of the H.D. Dr. Katharina Schneider.

Table 4.4 Results of segregation analysis and genetic mapping of the 76 candidate genes preferentially expressed in root . In case mapping data were deduced from previously available data it is indicated (*).

Coord.	Acc. N° ESTs Seequence[a]	Arabidopsis description	Mapped in pop. 618	Mapped in pop.K2	Chr
A-10-9	BQ489859	Putative protein	x		6
D-20-8	BQ489740	Tubulin beta-1 chain	x	x	3
D-24-6	BQ489146	Amino acid transporter protein-like	x		6
D-3-10	BQ490233	Putative protein	x		7
F-9-9	BQ490013	Sucrose synthase		x	8
E-11-5	BQ490613	Hypothetical protein		x	3
E-21-9	BQ489820		x		9
E-24-5	BQ490637		x		8
E-8-2	BQ488188		x		3
E-8-9	BQ489853	Reversibly glycosylated polypeptide	x		5
F-12-2	BQ488151		x		4
F-23-10	BQ490303		x		1
F-9-3	BQ488380			x	8
G-14-6	BQ488951	Peroxidase		x	6
G-18-7	BQ489294		x		6
G-22-10	BQ490217		x		8
G-22-9	BQ489904	Llipid transfer protein, putative		x	2
I-24-7	BQ489314	Nodulin-like protein		x	1
J-7-3	BQ488374	Putative phosphoenolpyruvate carboxylase kinase	x	x	4
K-19-2	BQ488084	Aalpha-mannosidase	x	x	6
K-20-5	BQ490562		x	x	9
O-19-5	BQ490546	Tubulin alpha-5 chain-like protein		x	3
O-19-6	BQ488897	Translationally controlled tumor protein homolog	x		4
O-6-3	BQ488519		x	x	4
P-11-3	BQ488391		x		4
P-11-4	BQ488829	Annexin	x		2
P-12-6	BQ489110		x		3
P-14-7	BQ489439			x	6
A-7-3	BQ488238	Plasma membrane major intrinsic protein 2 - beet	Homologous to *mip2* (e-value = 0)*		9
D-4-7	BQ489399	Sucrose synthase	Homologous to *ss* (e-value = 0)*		7
E-14-1	BQ487564	Xyloglucan endotransglycosylase XET2	Homologous to *extu110* (e-180)*		2
M-22-4	BQ488735	Vacuolar ATP synthase16 Kd proteolipid subunit	Homologous to *atp3* (9E-51)*		1
I-17-10	BQ490130	sucrose synthase - beet	Homologous to *ss* (e-value = 0)*		7
C-9-5	BQ490572	Putative protein	*		3
J-21-9	BQ490059		*		4
F-18-8	BQ489734	ADP-ribosylation Factor 1		x	3
O-16-8	BQ489595		x		5
N-8-5	BQ654410			x	9
O-7-6	BQ488855	Jasmonate-induced protein homolog		x	2
K-24-5	BQ654411	putative protein	x		4
A-3-7	BQ489160	S-like ribonuclease		x	6
P-12-8	BQ489719	Unknown protein		x	4

The EST sequences of the genes preferentially expressed in beet were used to develop specific primers.

The parents of two mapping populations (the population 618, described in Schneider *et al*., 1999 and the population K2 described in Schäfer-Pregl *et al*., 1999) were screened for polymorphisms by the SSCP technique.

Seventeen genes were polymorphic only in the population 618 and thirteen only in the population K2 (P952). In total 5 genes were polymorphic in both populations. Linkage group assignment and a summary of all data relative to the mapping of 35 candidate genes are reported in Table 4.4.

These data were also integrated with map position data relative to six more genes already mapped (Schneider *et al*., 1999 and Schneider unpublished) because of the homology of the identified candidates to them (the relative expected values are reported as well in Table 4.4.

By mapping, the sucrose synthase-like gene represented by the EST with the accession number BQ490013, for which a qualitatively different expression pattern respect to the expression for the already reported cDNA clone SBSS (Hesse and Willmitzer, 1995) was observed (represented here by the EST BQ490130, see also Figure 4.4 for confirmation of the different expression pattern), was in fact associated to a new genetic locus on chromosome VIII as shown in Table 4.4.

5 - SELECTION OF CANDIDATE GENES RELATED TO ROOT DEVELOPMENT AND TO THE SUCROSE ACCUMULATION PROCESS

This chapter reports how the established macroarray technology was used to analyze a new EST collection of 11520 unique cDNA clones from young and mature root leaves and inflorescences of sugar beet in a time-course experiment, with the aim to identify developmental and metabolism related candidate genes in sugar beet roots. Plants were characterized morphologically and metabolically with respect to sucrose content during the development in two different years. The profile of sucrose accumulation was correlated with the expression profile in both years. Candidate genes which showed differential expression during the development were identified. Candidate genes were classified with respect to their function, and as a technical validation the data were confirmed by semiquantitative RT-PCR.

5.1 - MORPHOLOGICAL CHARACTERIZATION OF THE SAMPLES USED FOR THE TIME-COURSE ANALYSIS

5.1.1 – Field data: description of plant growth parameters measured during the time-course analysis in the years 2001-2002

The development of sugar beet plants was analysed during the years 2001 and 2002. Plants were grown in a field plot at the Schaugarten of the Max Planck Institute for Plant Breeding Research in Köln, Germany. They were harvested in intervals of 15 days during the summer time (time points indicated in paragraphs 2.1.1.2 and 2.1.1.3) and morphological parameters were measured to characterize the samples. A photographic documentation of the development was produced for the year 2001 and is shown in Figure 5.1. At each harvesting time-point a representative plant of the hybrid genotype 8J6203 used for the molecular analysis was photographed.

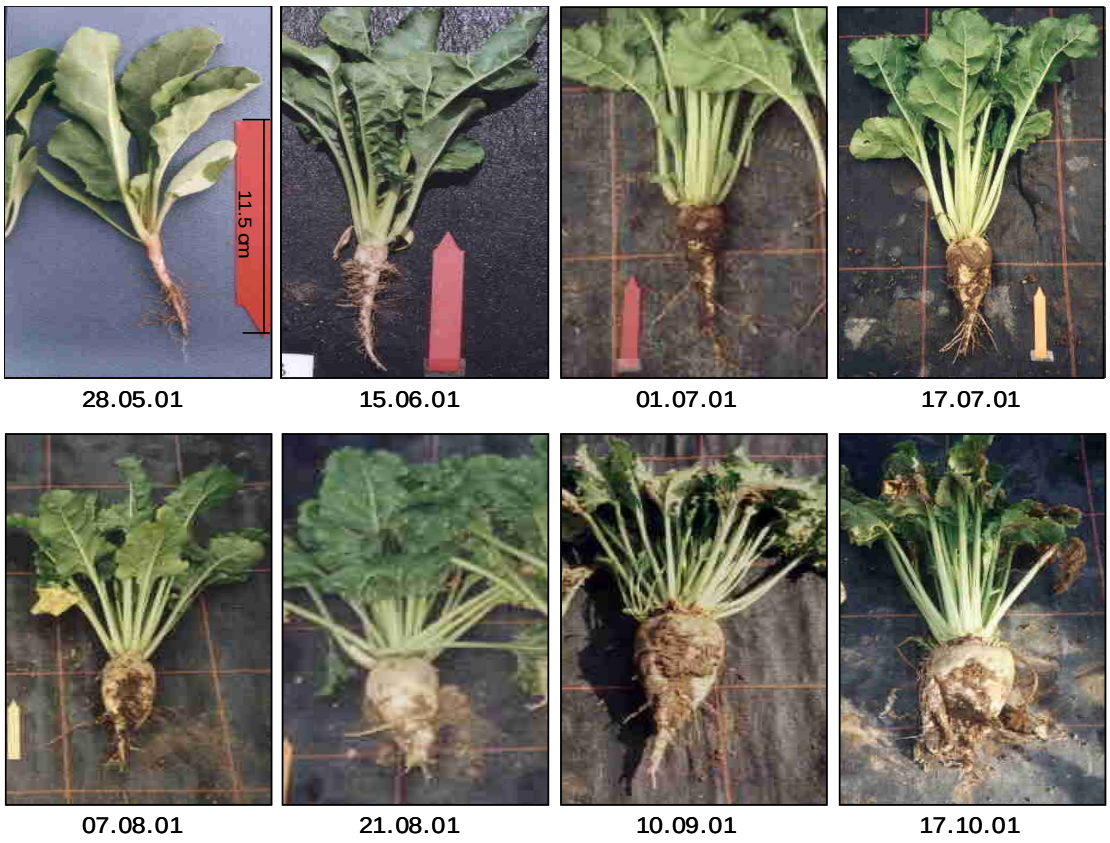


Figure 5.1 Phenotype of sugar beet plants (genotype 8J6203) at harvest dates during beet development in 2001. Length of colored label is 11.5 cm.

An additional time-point in the early development was also photographed, but no sample was collected at this time and no morphological characterization was performed (Figure 5.1).

The parameters selected for the morphological characterization of the sugar beet plants were: the "weight of root" in grams, the "number of rings" referring to the secondary cambium rings in the root, the "length of root" in centimetres, the "number of leaves" and the root diameter or "root thickness" in centimetres. List of all the data on these parameters measured on 3 to 6 plants in the 2001 for each time point and on 4 to 6 plants in the 2002 for each time point is reported in the Table 5.1. Concerning the number of leaves it is important to mention that in the second part of the development the total number of leaves was measured including those at the stage of advanced senescence.

At late stages in development many leaves are already dead, but still attached to the plant.

Table 5.1 Morphological data evaluated at the seven harvesting time-points in the year 2001(A) and 2002(B). Average values and standard deviations were calculated on the basis of 3-6 (2001) and 4-10 (2002) plants for each time point (n.a. not analyzed).

A)

HARVESTING DATE	PLANT N°	WEIGHT (gr)	N° OF RINGS	ROOT LENGTH (cm)	N° OF LEAVES	ROOT THICKNESS (cm)
15.06.01	1	n.a.	7	10	10	2
	2	n.a.	8	12	9	1
	3	n.a.	8	9	9	2
	4	n.a.	n.a.	12	9	1
	5	n.a.	n.a.	8	9	1
	6	n.a.	n.a.	14	11	2
	average	n.a.	8	11	10	2
	st. dev.	n.a.	±1	±2	±1	0
01.07.01	1	n.a.	8	18	16	6
	2	n.a.	9	22	16	5
	3	n.a.	n.a.	15	19	6
	4	n.a.	n.a.	19	14	5
	5	n.a.	8	23	20	6
	6	n.a.	9	13	24	7
	average	n.a.	9	18	18	6
	st. dev.	n.a.	±1	±4	±4	±1
17.07.01	1	155	9	21	20	7
	2	n.a.	10	19	20	7
	3	220	10	19	23	8
	average	188	10	20	21	7
	st. dev.	±46	±1	±1	±2	±1
07.08.01	1	518	11	12	37	9
	2	654	14	13	35	10
	3	415	9	15	27	8
	4	725	11	15	34	11
	average	578	11	14	33	9
	st. dev.	±139	±2	±1	±4	±1
21.08.01	1	970	14	30	38	11
	2	470	12	16	31	9
	3	776	10	20	37	10
	4	n.a.	11	13	37	9
	average	739	12	20	36	10
	st. dev.	±252	±2	±8	±3	±1
10.09.01	1	650	11	17	36	11
	2	725	12	23	55	11
	3	1133	12	27	45	11
	4	525	11	24	37	9
	average	758	12	23	43	10
	st. dev.	±263	±1	±4	±9	±1
17.10.01	1	850	14	21	37	12
	2	770	14	22	40	12
	3	1200	15	24	n.a.	12
	average	940	14	22	39	12
	st. dev.	±229	±1	±2	±2	0

B)

HARVESTING DATE	PLANTS	WEIGHT (gr)	N° OF RINGS	ROOT LENGTH (cm)	N° OF LEAVES	ROOT THICKNESS (cm)
05.06.02	1	11	6	10	13	2
	2	3	5	7	10	1
	3	4	5	10	11	1
	4	10	7	11	11	2
	5	3	5	10	9	1
	6	4	6	9	9	1
	7	3	5	9	9	1
	8	5	5	10	9	1
	9	8	7	n.a.	9	2
	10	6	6	n.a.	10	2
	average	6	6	9	10	1
	st. dev.	±3	±1	±1	±1	0
25.06.02	1	86	8	14	16	5
	2	135	9	15	25	6
	3	90	7	20	18	5
	4	102	8	15	18	5
	5	102	8	16	20	5
	6	46	8	15	16	4
	average	94	8	15	19	5
	st. dev.	±29	±1	±2	±3	±1
12.07.02	1	144	9	13	17	6
	2	255	11	20	22	7
	3	228	9	14	22	8
	4	373	11	18	29	9
	5	154	10	13	26	7
	6	210	11	19	26	7
	average	227	10	16	24	7
	st. dev.	±83	±1	±3	±4	±1
02.08.02	1	634	11	23	40	9
	2	515	11	20	39	9
	3	475	14	22	37	9
	4	287	12	24	32	8
	5	500	12	n.a.	29	10
	average	482	12	22	35	9
	st. dev.	±125	±1	±2	±5	±1
23.08.02	1	505	14	26	48	10
	2	609	13	22	36	10
	3	1176	14	22	53	12
	4	1048	14	27	48	13
	average	835	14	24	46	11
	st. dev.	±327	±1	±3	±7	±2
17.09.02	1	965	12	27	43	13
	2	1270	13	30	59	14
	3	1783	14	26	42	14
	4	863	11	23	49	12
	5	898	13	25	35	12
	6	969	12	27	48	13
	average	1125	13	26	46	13
	st. dev.	±353	±1	±2	±8	±1
9.10.02	1	n.a.	15	17	48	11
	2	694	12	29	35	9
	3	701	13	20	46	11
	4	707	13	26	41	9
	5	950	11	18	51	11
	average	763	13	22	44	10
	st. dev.	±125	±1	±5	±6	±1

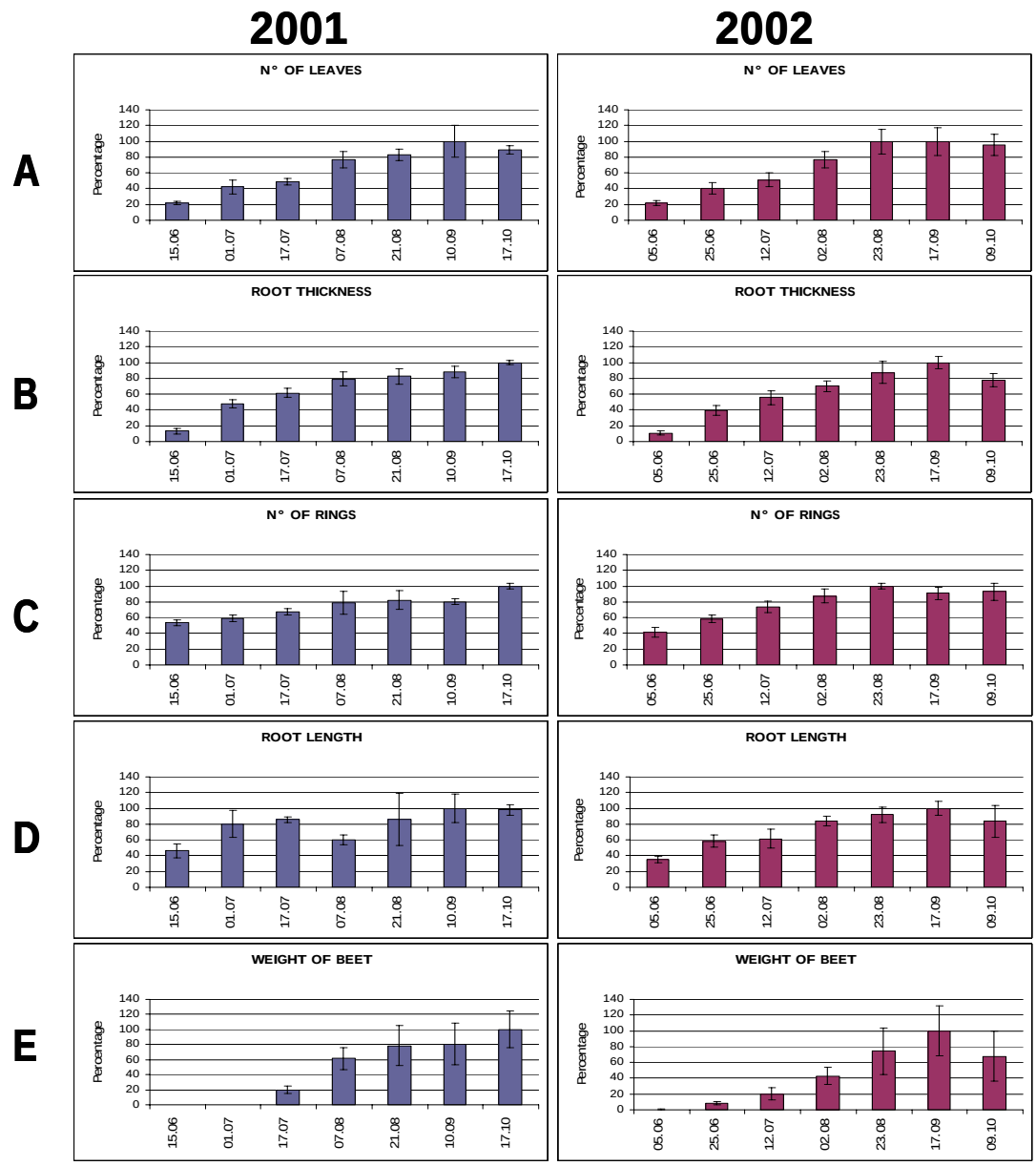


Figure 5.2 Averaged morphological data at each of the seven harvesting time-points in the years 2001 (blue) and 2002 (red). Error bars indicate the standard deviation. Values are plotted as percentage of the highest value in each season for each parameter. A - Number of leaves, B - Root thickness, C - Number of rings, D – Root length, E – Weight of beet.

To compare the graphs illustrating the different morphological parameters during the development, the values for the different parameters were expressed as percentage of the maximum reached in the season by the plants and plotted for the years 2001 and 2002 separately (Figure 5.2).

In both years the number of leaves (Figure 5.2–A) increased linearly along the first five time-points and entered a phase of saturation in late August or beginning of September. The graphs for root thickness (Figure 5.2–B) showed similar behaviour.

Concerning the number of secondary cambium rings (Figure 5.2–C) it is interesting to observe that already at the first time point, when the plant was one and half month old in the year 2001 and 2 months old in the year 2002, 50 or 40% of the rings were already laid down. To follow the ring initiation in the beginning of the development, transverse sections of young beets were analyzed under the microscope. Figure 5.3 shows that at 3.5 weeks (additional time point, not represented in the plots in Figure 5.2-C) the primary cambium was already complete and the first two secondary cambia were formed (Figure 5.3-b). At the same time point the situation was different in the tip of the beet, where only one secondary cambium ring was detected (Figure 5.3-a). The microscopic analysis revealed 6 cambium rings in the central root region in 7 week old sugar beet plantlets (Figure 5.3-c).

These data together with the data on root thickness illustrate how a beet develops: ring initiation begins early in the development and by 9 weeks after sowing 6 to 9 rings are already set (Elliott and Weston, 1993). In the second phase the emphasis is on ring enlargement and expansion. The maximum number of rings at harvest was found to be 13-14 rings.

Figure 5.2-D shows that the length of the root was determined quite early even if this parameter was producing unstable data. Finally Figure 5.2-E illustrates that the final weight of the root was set mainly in an advanced stage of the development, when less emphasis was on leaf development and the majority of the secondary cambium rings were already set.

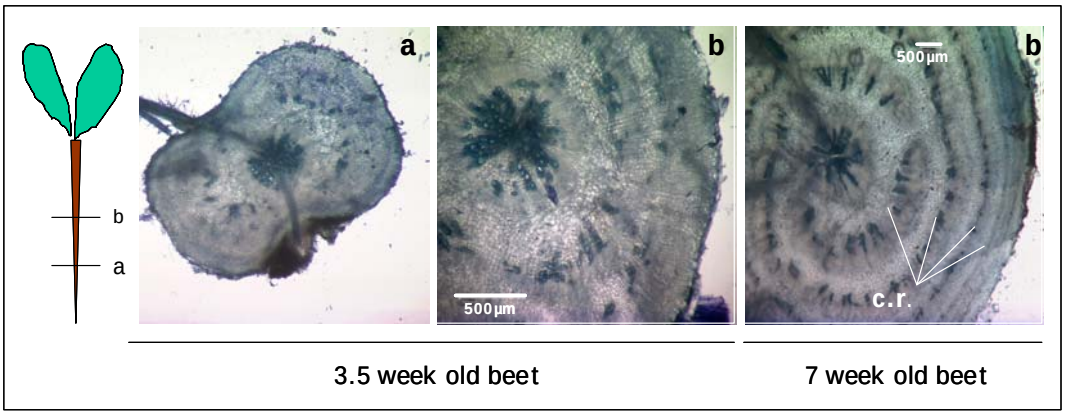


Figure 5.3 Toluidine Blue stained transverse sections of sugar beet plantlets (genotype 8J6203) of the indicated age. The planes for the transverse section for a and b are indicated in the scheme on the left side. cr: cambium rings.

5.1.2 – Biochemical data: sucrose content of the samples used for the time-course analysis in the years 2001-2002

As this study was aimed to follow the sucrose accumulation in the beet during the development at the morphological, biochemical and molecular level, morphological parameters were complemented by a measurement of the sucrose content of the beets. For each sample different plants were pooled, as explained in the paragraphs 2.1.1.2 and 2.1.1.3, therefore these data are representative of the specific harvesting time-points.

Percentages of sucrose per fresh weight for two measurements for each of the seven samples harvested during the years 2001 and 2002, respectively, are shown in the Table 5.2.

In both years relatively high sucrose content was observed at the first analysed time-point. In 2001, when plants were 9 weeks old the sucrose concentration was found to be 10.17% of the fresh weight, this value is equal to nearly 60% of the sucrose percentage at the end of the season. At the first time point in 2002, when the plants had the same age as the plants considered for 2001, the sucrose content was equal to 7.54% of the fresh weight, 44% of the sucrose percentage at harvest. Even if a certain difference between the two years is shown, in both cases it is evident that sucrose accumulation is a process that starts early in the development.

Table 5.2 Beets Sucrose content during development measured at 7 time-points during the years 2001 and 2002. The values are given as percentage of sucrose per FW. Average values and standard deviations are indicated as well. n.a.: not analyzed.

2001	**15.06.01**	**01.07.01**	**17.07.01**	**07.08.01**	**21.08.01**	**10.09.01**	**17.10.01**
I measurement (%sucrose/FW)	n.a.	10.18	13.77	13.85	16.73	16.24	17.39
II measurement (%sucrose/FW)	n.a.	10.15	13.77	13.58	17.74	16.19	17.39
Average (%sucrose/FW)	n.a.	10.17	13.77	13.72	17.24	16.21	17.39
Standard deviation	n.a.	0.02	0.00	0.19	0.72	0.04	0.00
2002	**05.06.02**	**26.06.02**	**12.07.02**	**02.08.02**	**23.08.02**	**17.09.02**	**09.10.02**
I measurement (%sucrose/FW)	7.46	11.88	12.60	15.42	15.56	16.91	16.43
II measurement (%sucrose/FW)	7.61	11.91	12.09	15.77	15.57	17.21	16.88
Average (%sucrose/FW)	7.54	11.89	12.35	15.59	15.56	17.06	16.66
Standard deviation	0.11	0.02	0.36	0.25	0.01	0.21	0.32

Maximum sucrose content was found to be 17.39% per fresh weight in 2001 and 17.06% per fresh weight in 2002.

To compare the graphs for sucrose accumulation in the two different years, the sucrose percentages per fresh weight were expressed as percentages of the maximum sucrose percentage per fresh weight in each season and plotted separately for the years 2001 and 2002 (Figure 5.4).

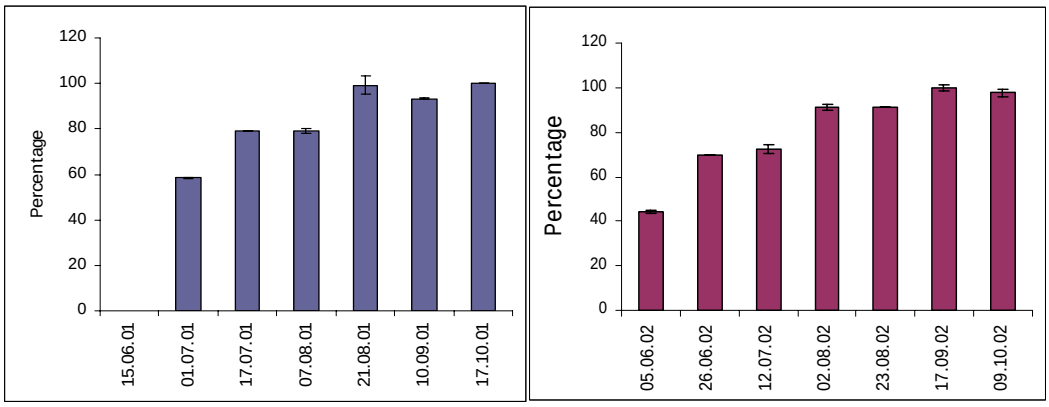


Figure 5.4 Percentage of averaged sucrose content per fresh weight at each time point during the seasons 2001 (blue) and 2002 (red). Values are calculated as percentage of the highest percentage of sucrose per FW for each season.

In both years 80% of the maximal sucrose content was already reached at the beginning of August. After this time point the sucrose content continues to grow, but much more slowly than in the first part of the development. This must be kept in mind, as the purpose here was to identify genes with an expression profile correlated to the sucrose accumulation process in the beet, and therefore putatively involved in such activity.

5.1.3 – Comparing the development in the two years 2001-2002

As already mentioned in chapter 4, the variability attributed to macroarray experiments is high, especially if field grown plants are considered. To overcome this variability technical and biological replicates are introduced in macroarray experiments (see paragraph 3.1.2). To produce real biological replication for a time-course experiment is not easy. As sucrose accumulation and beet development are

very generic processes in sugar beet the idea here was to consider two different years and focus on genes showing similar expression pattern in the two years. The rationale behind this was that sugar beet is growing and accumulating sucrose in both years, and therefore genes related to these processes should be expressed in both years when storage and growth take place.

A problem related to this approach is to define comparable stages between years. A common practice in agronomy is to use the "temperature sum rule" to compare developmental stages over years. This bases on the idea that plant development under field condition can be described and followed quantitatively using a unit called "cumulative growing-degree-days" (GDD) (Thornley and Johnson, 1991). This value can be considered as a measure of the amount of warmth that plants have experienced. GDDs were calculated according to Thornley and Johnson (1990), and they are reported in Table 5.3 together with the age of the plants for each of the time points considered in the two years. To identify comparable stages between years, GDD values for the year 2001 were compared to GDD values relative to the time points in the year 2002 in a plot (Figure 5.5-B). In a similar way a plot for the ages of the plants was produced (Figure 5.5-A). The GDD values were more consistent between the two years, as compared to the high differences in the ages of the plants. This observation about comparability of time

Table 5.3 For each time point the age of the plant is calculated based on the sowing day 02.05.01 for the year 2001 and on the sowing day 04.04.02 for the year 2002. The GDDs are calculated as reported in Thornley and Johnson (1990), using the daily temperatures measured in Köln Flughafen and considering T_0= 6°C.

2001				2002			
Time-point	Date	Days after sowing	GDD	Time-point	Date	Days after sowing	GDD
t1	15.06.01	45	452	t1	05.06.02	63	514
t2	01.07.01	61	657	t2	25.06.02	84	807
t3	17.07.01	77	877	t3	12.07.02	100	996
t4	07.08.01	98	1201	t4	02.08.02	121	1293
t5	21.08.01	112	1420	t5	23.08.02	142	1599
t6	10.09.01	132	1668	t6	17.09.02	168	1916
t7	17.10.01	169	2019	t7	09.10.02	190	2072

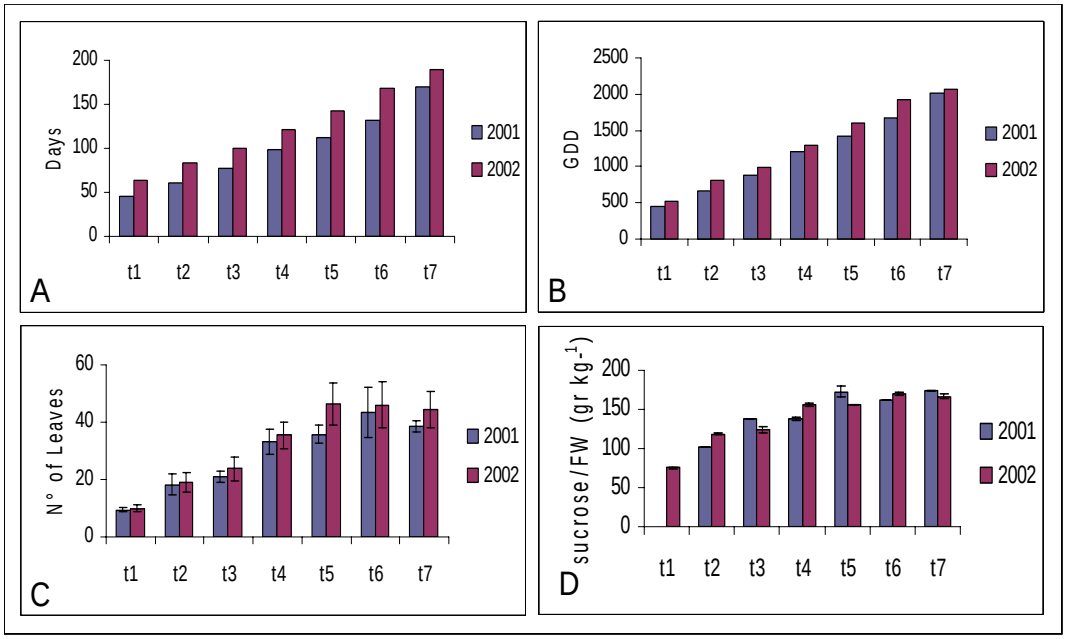


Figure 5.5 Evaluation of comparable developmental stages between the two years considered in this study based on the parameters plant age, GDDs, number of leaves (as absolute value) and sucrose percentage on FW (as absolute value).

points in the development between the two years was in agreement with the conclusions that could be inferred from data on "number of leaves" and "sucrose content" (Figures 5.5-C,D). However, the option to use the relative time points in the two years as "biological replicates" was not applicable because of the relevant differences in the two seasons. The strategy was then to perform the macroarray analysis independently for each year and introduce the critical requirement of reproducibility in the expression patterns (as defined later in paragraph 5.3.4) between the two years to select for candidate genes

5.2 – ANALYSIS OF EST DATA (*)

The time–course analysis was performed using macroarrays with collection A024 of 11520 unigene cDNAs described in paragraph 2.1.2.2. A sequence analysis concerning functional domains was performed with the first 10752 ESTs. Among these, 1393 sequences were excluded because of low quality.

(*) The sequencing data and the funcat (functional category) for the 11520 ESTs were obtained by the database Sputnik at the MIP (Münich Information Protein Resource Center) with the collaboration of the Dr. Stephen Rudd which is kindly acknowledged

For 5253 sequences, no functional domain was identified (Table 5.4).

Table 5.4 Classification of ESTs with respect to the analysis of functional domains

ESTs	N° OF SEQUENCES
Total number of ESTs analysed	10752
Sequences of bad quality or not considered in this analysis	1393
Sequences containing no functional domain	5253
Sequences containing a functional domain	4106

The remaining 4106 sequences were grouped into 17 main categories as shown in Figure 5.6. The most abundant functional category was encoding proteins involved in primary metabolism. A large number of domains typical of proteins involved in transcription as well as in cell growth, cell division and DNA synthesis, translation, protein destination, and energy related processes were found.

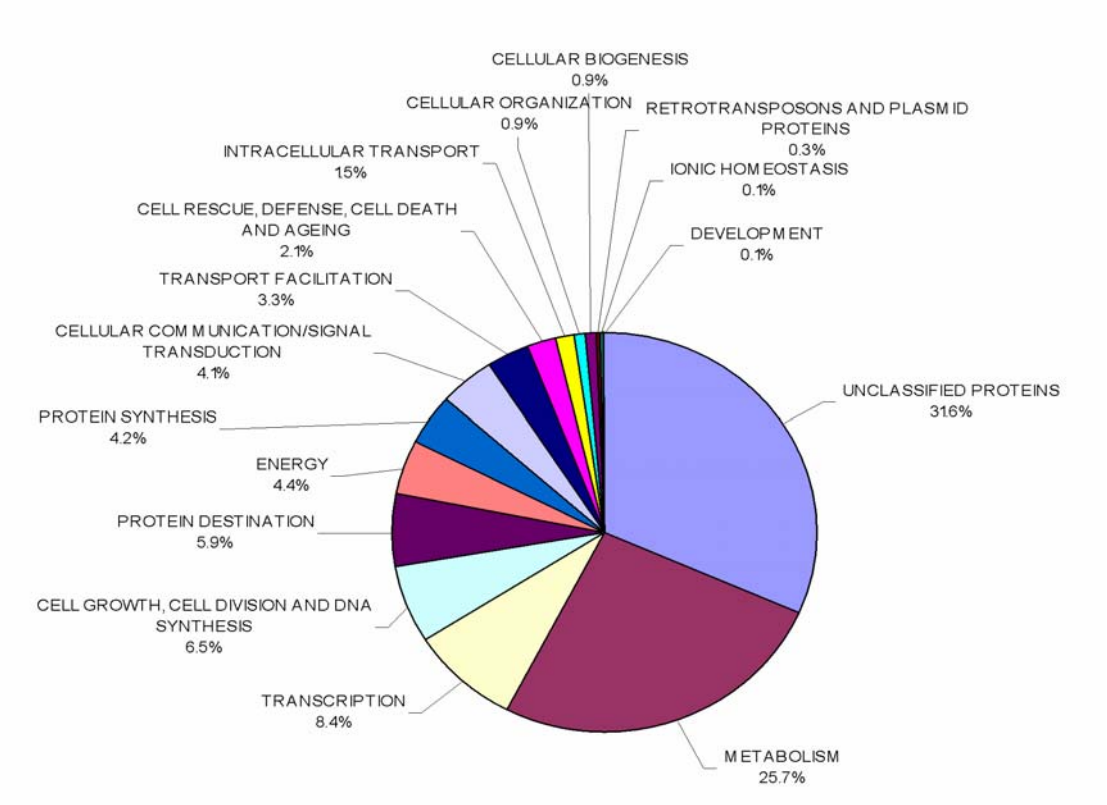


Figure 5.6 Classification of 4106 annotated gene products into functional categories (Gene Ontology, 2000) using the InterProScan program. For another 5253 sequences matching the quality parameters required by the program, no functional domain was identified. Additional 1392 of a total of 10752 sequences failed the quality criteria for categorization.

Interestingly, 1299 ESTs contained domains not functionally characterized. These, together with EST sequences for which no functional domain was identified, were accounting for a total of 6552 ESTs, 61% of the EST collection.

5.3 –EXPRESSION ANALYSIS OF THE TIME-COURSE EXPERIMENT

5.3.1 – Generation of macroarrays for cDNA library A024

Macroarrays of the cDNA library A024 were produced according to the same procedure as described before for library A006. A total of 11520 amplified cDNA inserts including controls and cDNAs for the 76 candidate genes identified in the macroarray analysis described in the chapter 4 were spotted in duplicates.
Macroarrays were produced in collaboration with the ADIS service unit at the MPIZ as described in the paragraph 2.2.6.2, and the full library could be accommodated on 22 x 22 cm nylon filters.

5.3.2 – Macroarray analysis: experimental design and parameters selected for the data analysis

Concerning the experimental design employed here, four independent [^{33}P]-labelled probes were prepared for each of the 14 samples used for the time course analysis, to produce technical replicates.
Probes produced from the same sample were always applied to different filters. In total 25 filters were subjected to two to four hybridizations plus a further hybridization using the radioactively labelled T7-short oligo as control.
Preliminary hybridization results for the 22 x 22 filters showed that a double amount of probe should be used to improve the quality of the signal. Additionally, the exposure time of the filters to phosphor screens was doubled. Image acquisition and analysis were performed as described for the previous filters, and the normalization was based on the median signal of all the spotted cDNA for the spiked control *nebulin.*

After normalization, two different Excel (commercial software) matrices with the normalized expression values were produced for the years 2001 and 2002. For each clone, eight normalized values were reported for each of the seven time points considered. For outlier detection a procedure to calculate the random error was applied which differed from the one described in the paragraph 3.1.4. The "small sample" strategy was used, in which random error is estimated based on the data from one gene, without considering the data for all the other genes on the array. This different approach was applied in order to allow the application of the statistical test *F-test* to identify clones differentially expressed during the development.

5.3.3 - Differential expression during the development in the years 2001-2002

Differential expression was assessed using the program Array Stat and the significance of the differential expression was estimated using the *F-test*. "False positive rate" was set to $\alpha < 0.05$ and the correction procedure *Stepdown Bonferroni* was applied. Results for the differential expression in 2001 and 2002 are reported in Table 5.4.

Table 5.4 Statistical evaluation of the normalized expression data derived from eight replica for each-time point. A statistical test was applied to asses the significance of the differential expression (ArrayStat software package, Imaging Research). The number of clones is given and the respective percentages are indicated in brackets.

Number of clones (percentage)		
	2001	2002
Total numbers of clones	11520 (100%)	11520 (100%)
Differential expression	3486 (30.26%)	2052 (17.81%)
No differential expression	7692 (66.77%)	9076 (78.78%)
Discarded for statistical reasons	342 (2.97%)	392 (3.41%)

Among the clones showing differential expression in 2001 and 2002, 1107 clones were identified for which the differential expression was confirmed in both years.

5.3.4 – Identification of clones showing similar expression profiles in the two years 2001-2002: cluster1 and cluster2

5.3.4.1 - Clustering analysis of the expression profiles of the clones showing differential expression in both the years

To discover similarity among expression profiles of many genes exceeds the ability of human assessment. Therefore clustering algorithms have been developed and employed for this purpose after plotting the expression of each gene over the various conditions.

The rationale behind this approach is that genes showing similarity in expression pattern may be functionally related, controlled by the same genetic elements or be part of similar pathways. Analysing data by clustering results in groups comprising both known and unknown genes, and allows associating putative functions to the unknown genes by employing the concept of "guilt by association" (Aharoni *et al*., 2002).

Although cluster analysis techniques are extremely powerful, great care must be taken in applying these procedures. Selecting different algorithms or different distance metrics, will place different objects in different clusters revealing unique aspects of the data (Leung and Cavalieri, 2003). Furthermore, clustering unrelated data will still produce clusters, although they may not be biologically meaningful. The challenge is therefore to select the metrics and to apply the algorithms appropriately so that the classification that arises partitions data sensibly.

In many cases a preliminary problem can be that data are dominated by the variables that have the largest value, obscuring other important differences. One way to circumvent this problem is to adjust or re-scale the data so that the average expression of each gene is zero. In this process, the basal expression level of a gene is subtracted from each experimental measurement. This has the effect of enhancing the variation of the expression pattern of each gene across experiments, without regard to whether the gene is primarily up-or down-regulated. It has been shown to be particularly useful for the analysis of time-course experiments, in which one might like to find genes that show similar variation around their basal expression values (Quackenbush, 2001).

The choice of the similarity measure is a delicate step, because it can produce different results. The “Pearson`s correlation” metric measures how similar the expression patterns are, irrespective of the amplitudes in the expression profile, whereas the “Euclidean measure” considers the absolute distance between two expression profiles. To identify genes with similar expression profile during the time-course experiment, the “Pearson`s correlation” similarity measure was used in this study.

Clustering techniques can be divided into divisive and agglomerative. A divisive method begins with all elements in one cluster that is gradually broken down into smaller and smaller clusters. Agglomerative techniques start with single-member clusters and these are gradually fused together. Additionally, clustering can be either supervised or unsupervised. Supervised methods use existing biological information about specific genes that are functionally related to guide the clustering algorithm. Finally clustering techniques can be divided into hierarchical and not hierarchical. In hierarchical clustering there is an increasing number of nested classes and the results resemble a phylogenetic tree. This system has the advantage that it is simple and the results can be easily visualized. It was the first clustering algorithm to be employed with gene-expression data by Eisen *et al*. (1998) and therefore it is widely used. One potential problem with this clustering system is that at a critical size of the growing clusters it loses accuracy (Sherlock, 2000). These problems can be avoided by first partitioning the data into reasonably homogeneous groups that can eventually be individually clustered in a second step. Many non hierarchical clustering techniques can be used for partitioning data, such as *k*-mean clustering, which simply partitions objects into different clusters without trying to specify the relationships between individual elements.

For the time-course experiment described here, partitioning of data was performed using the *k*-means clustering system in the software Genesis (Sturn, 2002). The *k*-means system requires an advanced knowledge about the number of clusters. To evaluate this number clustering was performed with different k values from 2 up to 15. As results for k-values higher than two were not reproducible it was concluded that there are two groups of genes with major differences between them. The analysis was performed in absence of the first time point for which the average values of expression were displaying the highest difference between the profiles of the two

years in both clusters identified. Therefore this data-point was excluded from the analysis. Prior to application of the clustering algorithm the expression data for the 1107 clones identified in the previous paragraph were log 10 to log 2 transformed, using a function in the program Genesis. In the next step they were filtered and only clones for which all six time points were showing a value were considered. Finally the "median center" function was employed to highlight induction/repression changes as already explained. As distance measure the "Pearson`s correlation" was selected for the *k*-means clustering algorithm. Average values of expression calculated for all members of the identified clusters in two years are reported in Figure 5.7.

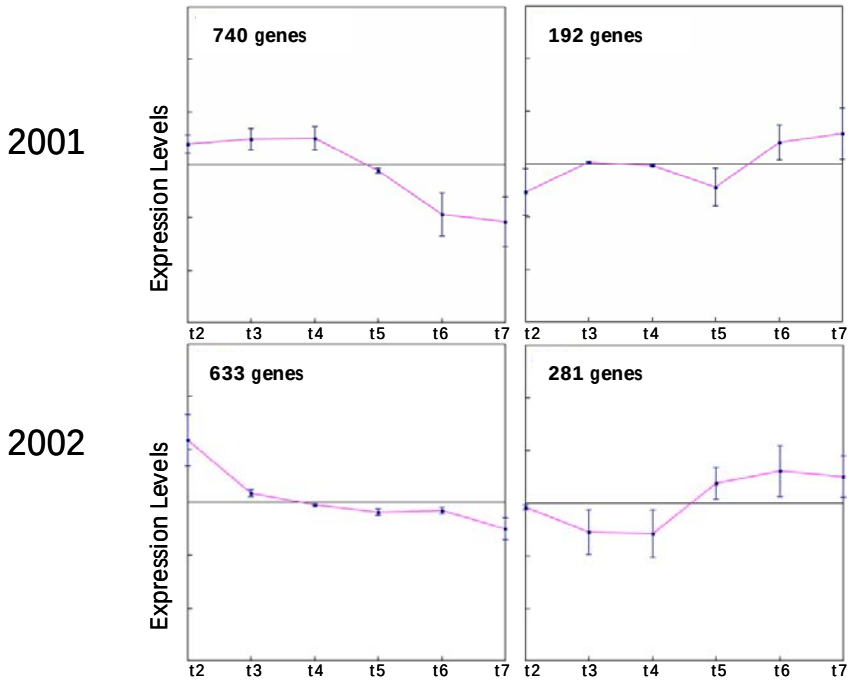


Figure 5.7 K-means clustering analysis of transcription profiles. The analysis is based on complete hybridization data sets of cDNAs confirmed to be differentially expressed in both years studied.

5.3.4.2 - Clones showing similar expression profiles during the development in 2001-2002: identification of two expression clusters

In a second step, the clones showing a similar expression pattern in the two years were selected. Therefore clones belonging to the first cluster in 2001 were compared to clones belonging to the first and to the second cluster in 2002. The same procedure was done for the clones belonging to the second cluster of the 2001. Results of this comparison are reported in Table 5.5.

Table 5.5 Numbers of cDNAs shared between the two clusters of each year. Total numbers of clones for each cluster is given in brackets

	Cluster 1-2001(740)	Cluster 2-2001(192)
Cluster 1-2002 (633)	599	8
Cluster 2-2002 (281)	98	175

In this way clones with comparable expression patterns during the development were identified. In summary, 80.9 % of the cDNAs belonging to the first cluster in 2001 were identified in the first cluster in the second year as well, and, vice versa, 95% of the clones belonging to the first cluster in 2002 were found in the first cluster of 2001. For the second cluster 91% of the clones identified in the first year were in common with 62% of the clones identified in the second year.

In conclusion, 599 putative genes were identified as preferentially expressed during the first part of the development in both years, and 175 putative genes were identified as related to maturity in both years.

The expression values for these 599 and 175 clones in the two different years were retrieved and plotted separately in Figure 5.8.

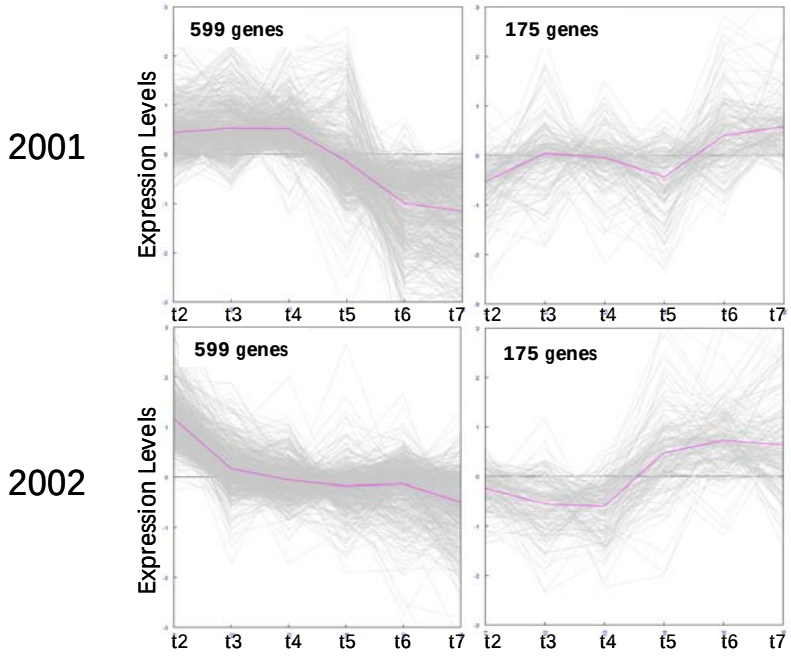


Figure 5.8 Median centered expression values for each of the 599 cDNAs belonging to cluster 1 in both years, and for each of the 175 ESTs belonging to cluster 2 in both years. Tendency lines are indicated in purple.

The graph reveals the variability within each cluster and the differences between the single clones and the average curves (purple) in the two years.

To visualize the results of the clustering procedures different representations are available. A commonly used approach relies on the creation of an expression matrix in which each column of the matrix represents a single time point and each row represents the expression pattern for a particular cDNA. Colouring each of the matrix elements according to its expression value provides a visual impression of gene-expression patterns along the time-course. The efficiency of the *k*-means partitioning procedure is illustrated in Figure 5.9.

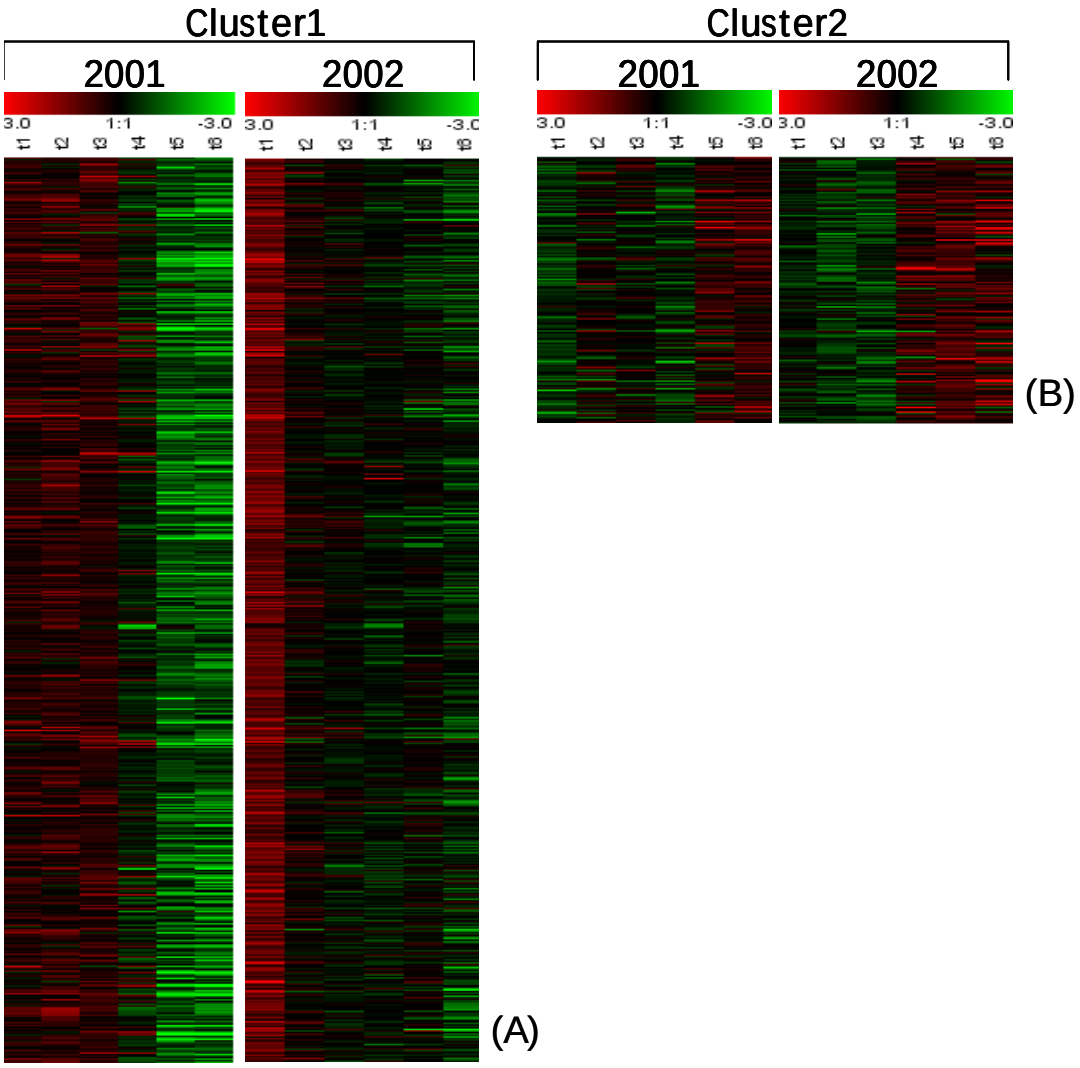


Figure 5.9 Transcription profiles of (A) the confirmed 599 cDNAs belonging to cluster 1, and of (B) the confirmed 175 cDNAs belonging to cluster 2 in the years 2001 and 2002. A scale for expression values is given on top of the figures, with red indicating induction and green color indicating repression of gene products. Black color symbolizes no differential expression. As t1 was eliminated t1refers to t2 and so on in this figure.

5.3.5 – Clones highly expressed and highly induced in early development and at maturity

To characterize the extremes of the two identified clusters, 40 clones with highest expression values and 40 clones showing the highest induction were selected from the clones present in both clusters in 2001 and 2002.

Annotations were retrieved by comparing deduced amino acid sequences to the MIPS (Munich Information Center for protein sequences) *Arabidopsis thaliana* database using an expect value of maximally e^{-10} for sequence similarities (Tables 5.6 and 5.7). Annotations retrieved for clones with similarity to other organisms were considered as well.

In case they were more informative than the annotations retrieved for the similar sequence in *Arabidopsis thaliana* they are indicated in the following.

As a general observation, the sequences of around half of the clones did not show any similarity to any other sequence applying the conditions described. Additionally, sequences showing similarity to unknown proteins were also highly present.

In the first cluster, high expression (Table 5.6-A) was observed for clones showing similarity to the gene product adenosylhomocysteinases, and to gene products involved in the cell growth related processes and biogenesis. Among these, especially tubulins and proteins involved in the biosynthesis and modification of cell walls were retrieved. Four clones showed similarity to gene products responsible for homeostatic equilibrium in the cell like an aquaporin, two water channel proteins (Yamada *et al*. 1995) and a MIP (plasma membrane intrinsic proteins). Interestingly, analogues of two nucleic acid binding proteins were found in this group of clones highly expressed.

Concerning the clones showing the highest induction in the first cluster, a list of potential analogues is reported in Table 5.6-B. Among these, homologies to gene products involved in transport processes like the mitochondrial dicarboxylate carrier, a TIP (tonoplast intrinsic protein, clone D-23,25), isolated in spinach when large vacuoles are formed (Karlsson *et al*., 2000) and an amino acid transport protein were retrieved. A total of six clones showed homology to gene products involved in cell wall metabolism. Two clones (B-12,27, E-8,27) revealed homology to *expansin 4* from *Cicer arietinum* and *expansin 2* from *Zinnia elegans*. According to Im *et al*. (2000), expression of expansin2 correlates with primary cell wall expansion and

secondary cell wall thickening. Four clones of which two encoding proteins with homology to xyloglucan endo 1.4-beta-D-glucanase, one an endo-polygalacturonase-I and another a putative prolin-rich protein are supposed to be associated with cell wall expansion (Fukuda *et al*, 1997) as well. Finally the nodulin-like gene product (clone J-15,9) is showing 98% of homology at nucleotidic level to the ESTs BQ489314 identified among the root preferentially expressed candidates of library A006. As mentioned in chapter 4 the analogous gene product has already been reported as located in cell wall (Davies and Robinson, 2000) during fruit ripening, even if the function is not yet clear. Three different clones showing homology to the phi-1-like phosphate-induced protein with unknown function (Sano et, 1999, Farrar *et al*, 2003) were found to be present in both tables.

In the second cluster some of the clones showing high expression (Table 5.7-A) were similar to gene products involved in protein synthesis and destination like ribosomal proteins and a translation factor from *Pinus pinaster* (clone H-18,5). Additionally, clones with similarity to proteins involved in signal transduction and pathogen defence as protease inhibitor II were identified. Three independent clones (C-22,3, I-20,18, K-9,5) were showing homology to gene products associated with dormancy in apple and *Pisum sativum* (Lee *et al*., 1993) and further three clones (I-17,18, H-9,18, K-8,18) were found similar to gene products involved in gibberellic acid signalling (Shi *et al*., 1992). Furthermore, a cDNA with high homology to the gamma-VPE (vacuolar processing enzyme) isolated from sugar beet roots by Kloos *et al*. (2002) and two clones with similarity to NADH dehydrogenase subunit 4 involved in energy production related processes, were identified.

In the second cluster, high induction (Table 5.7-B) was found for clones with homology to gene products encoding PR proteins (pathogenesis related) like glucanase, chitinase, and osmotin. Ethylene responsive transcriptional coactivators, also present in this group, are known to induce ripening related genes (Zegzouti *et al*., 1999) as well as PR genes (Fujimoto *et al*, 2000). A clone with similarity to the gene product gibberellin 20-oxidase (Xu *et al*., 1995) involved in GA biosynthesis and regulation, was also identified. Gibberellin 20-oxidase was correlated with cell elongation (Huang *et al*.1998) and, more recently, with the xylogenesis process (Israelsson *et al*. 2003).

Table 5.6 (A)List of most highly expressed genes belonging to the first cluster with expression and annotation data. The expression value is the maximal intensity among the early time-points in the two years. (B) List of most highly induced genes in the first cluster. The ratio is calculated between the highest value for the early time points in the two years and the lowest value for the late time-points in the two years.
(A)

Coordinates	Max. expression (signal intensity)	Accession of *Arabidopsis* homologue	*Arabidopsis* description	*Arabidopsis* e-value
P - 20, 28	59.09		no hit	
O - 2, 25	45.38	At1g50010	tubulin alpha-2/alpha-4 chain, putative	6.00E-63
O - 13, 28	33.27	At4g13940	adenosylhomocysteinase	1.00E-66
A - 13, 14	28.14	At5g17920	5-methyltetrahydropteroyltriglutamate homocysteine S-methyltransferase	8.00E-58
N - 6, 22	27.41	At1g68370	ARG1 protein (Altered Response to Gravity)	3.00E-56
P - 20, 1	21.63		no hit	
I - 10, 2	20.16		no hit	
L - 9, 8	20.13	At4g25810	xyloglucan endo-1,4-beta-D-glucanase (XTR-6)	5.00E-78
F - 5, 24	20.11	At4g01470	putative water channel protein	4.00E-67
K - 24, 20	20.11		no hit	
I - 19, 18	19.69	At2g16850	putative aquaporin (plasma membrane intrinsic protein)	2.00E-39
J - 24, 20	19.59		no hit	
B - 9, 5	18.92		no hit	
I - 1, 13	17.84		no hit	
K - 15, 23	17.32	At1g50010	tubulin alpha-2/alpha-4 chain, putative	6.00E-63
J - 1, 1	17.02	At4g01470	putative water channel protein	4.00E-67
E - 4, 4	16.30		no hit	
A - 22, 24	15.81		no hit	
P - 10, 17	15.62	At5g51550	putative protein	3.00E-56
P - 3, 25	15.09		no hit	
A - 8, 13	15.06		no hit	
I - 18, 15	14.71		no hit	
G - 9, 28	13.94	At5g07030	nucleoid DNA-binding-like protein	4.00E-42
A - 18, 27	13.07		no hit	
M - 2, 18	12.82	At4g13940	adenosylhomocysteinase	2.00E-94
O - 1, 25	11.69	At5g01650	light-inducible protein ATLS1	8.00E-50
N - 1, 25	11.60	At2g02990	putative ribonuclease, RNS1	6.00E-23
D - 15, 8	11.50		no hit	
O - 18, 27	11.35	At4g08950	putative phi-1-like phosphate-induced protein	2.00E-32
B - 12, 27	11.34	At2g03090	expansin like protein	5.00E-38
G - 9, 13	11.11	At1g78040	similar to phosphoglycerate mutase 1	4.00E-34
H - 12, 10	10.78		no hit	
M - 7, 11	10.52	At2g40140	putative CCCH-type zinc finger protein	2.00E-16
K - 9, 20	10.46	At4g08950	putative phi-1-like phosphate-induced protein	7.00E-44
D - 1, 25	10.17		no hit	
K - 17, 6	10.12	At4g00430	probable plasma membrane intrinsic protein 1c	1.00E-77
G - 2, 25	9.82	At1g70830	unknown protein (At1g70830)	3.00E-38
M - 1, 25	9.62		no hit	
J - 15, 9	9.55	At1g75500	nodulin-like protein	1.00E-65
M - 5, 26	9.49		no hit	

(B)

Coordinates	Ratio value	Accession of *Arabidopsis* homologue	Arabidopsis description	Arabidopsis e-value
F - 8, 30	143.90	At5g38700	MBB18 putative protein	1.00E-06
B - 9, 5	115.45		no hit	
O - 18, 27	69.30	At4g08950	putative phi-1-like phosphate-induced protein	2.00E-32
B - 12, 27	64.61	At2g03090	expansin like protein	5.00E-38
P - 14, 23	62.03	At1g70840	unknown protein	2.00E-36
E - 8, 26	56.48		no hit	
D - 14, 27	54.44		no hit	
P - 15, 21	38.94	At4g30270	xyloglucan endo-1,4-beta-D-glucanase precursor	2.00E-73
D - 12, 27	38.14	At4g08950	putative phi-1-like phosphate-induced protein	8.00E-45
O - 20, 23	33.61	At3g07270	GTP cyclohydrolase I	3.00E-62
O - 13, 22	31.98		no hit	
L - 6, 26	30.39	At5g09220	amino acid transport protein AAP2	2.00E-54
G - 6, 26	28.21	At1g73620	thaumatin-like protein	5.00E-58
H - 13, 27	27.94	At4g38400	putative pollen allergen	1.00E-36
N - 17, 21	27.56		no hit	
E - 9, 12	25.33	At2g30020	putative protein phosphatase 2C	6.00E-63
E - 8, 27	24.45	At2g40610	putative expansin	1.00E-56
J - 3, 12	23.77	At1g74670	GAST1-like protein	3.00E-28
B - 14, 8	23.43	At2g22500	putative mitochondrial dicarboxylate carrier protein	1.00E-27
O - 20, 8	22.44	At1g09070	unknown protein	1.00E-19
J - 19, 28	22.25	At3g60130	beta-glucosidase-like protein	6.00E-20
L - 9, 8	22.22	At4g25810	xyloglucan endo-1,4-beta-D-glucanase (XTR-6)	5.00E-78
E - 6, 14	22.09		no hit	
D - 4, 2	21.71		no hit	
F - 5, 29	21.49	At3g48950	RR_C_21 endo-polygalacturonase - like protein	2.00E-08
M - 1, 25	21.30		no hit	
J - 15, 9	21.11	At1g75500	nodulin-like protein	1.00E-65
C - 11, 21	20.94	At4g23500	putative protein	8.00E-76
O - 15, 21	20.68		no hit	
M - 18, 15	19.74	At5g64660	putative protein	3.00E-29
F - 20, 12	19.62	At3g18710	hypothetical protein	6.00E-24
N - 7, 8	19.51		no hit	
O - 18, 10	19.35	At1g09070	unknown protein	1.00E-19
A - 18, 27	19.29		no hit	
G - 17, 14	18.72		no hit	
C - 10, 22	18.64	At3g61490	putative protein	4.00E-23
D - 23, 25	18.61	At3g16240	delta tonoplast integral protein (delta-TIP)	3.00E-35
G - 3, 27	18.45		no hit	
E - 3, 25	18.36	At1g14890	unknown protein	2.00E-18
B - 9, 24	17.87	At2g45180	putative proline-rich protein	2.00E-21

Table 5.7 (A) List of most highly expressed genes belonging to the second cluster with expression and annotation data. The expression value is the maximal intensity among the late time-points in the two years. (B) List of most highly induced genes in the second cluster. The ratio is calculated between the highest value for the late time points in the two years and the lowest value for the early time-points in the two years.

(A)

Coordinates	Max. expression (signal intensity)	Accession of *Arabidopsis* homologue	Arabidopsis description	Arabidopsis e-value
G - 12, 18	62.61		no hit	
C - 22, 3	36.40	At2g33830	putative auxin-regulated protein	1.00E-15
G - 21, 19	35.29		no hit	
G - 21, 18	24.29	At2g02100	protease inhibitor II	2.00E-18
I - 20, 18	22.13	At1g56220	hypothetical protein	1.00E-15
A - 3, 1	21.69	At2g02100	protease inhibitor II	2.00E-18
I - 17, 18	21.35	At1g75750	putative protein	6.00E-25
L - 16, 6	20.43		no hit	
H - 12, 16	19.07	At4g32940	gamma-VPE (vacuolar processing enzyme)	8.00E-53
A - 4, 3	15.57	At1g60420	putative protein	2.00E-15
H - 11, 18	14.88	At1g25580	unknown protein	1.00E-49
I - 21, 29	14.37	At4g23100	SA_D_20 gamma-glutamylcysteine synthetase	3.00E-69
K - 9, 5	14.23	At1g56220	hypothetical protein	1.00E-15
L - 19, 20	13.38		no hit	
H - 14, 24	12.24	At4g25810	xyloglucan endo-1,4-beta-D-glucanase (XTR-6)	1.00E-66
H - 18, 5	10.20	At1g54290	putative protein	2.00E-54
B - 17, 13	9.05		no hit	
G - 11, 18	8.86		no hit	
H - 9, 18	8.64	At1g75750	putative protein	1.00E-29
N - 14, 18	7.35		no hit	
F - 4, 25	7.11	At4g12600	Ribosomal protein L7Ae -like	2.00E-48
O - 14, 17	6.64		no hit	
A - 18, 23	6.36	*nad4*	mitochondrial genome- NADH dehydrogenase subunit 4	2.00E-77
J - 7, 5	6.21	At4g38590	galactosidase like protein	4.00E-54
K - 15, 20	6.16		no hit	
K - 8, 18	6.08	At1g75750	putative protein	1.00E-29
K - 15, 6	6.07		no hit	
L - 13, 7	5.99	*nad4*	mitochondrial genome- NADH dehydrogenase subunit 4	2.00E-77
G - 10, 18	5.64		no hit	
G - 5, 25	5.52		no hit	
E - 22, 18	5.34	At5g05340	peroxidase	1.00E-95
A - 2, 1	5.29		no hit	
G - 9, 18	5.28		no hit	
L - 15, 6	5.26	At5g21090	leucine-rich repeat protein	2.00E-61
B - 3, 1	5.24		no hit	
O - 15, 11	5.23	At5g53330	proline-rich cell wall protein-like	1.00E-16
L - 14, 6	4.90		no hit	
I - 2, 29	4.89	At3g54160	gf_c_21 putative protein	0.06
L - 14, 19	4.83		no hit	
N - 6, 4	4.76	At5g15200	40S ribosomal protein - like	2.00E-46

(B)

Coordinates	Ratio value	Accession of *Arabidopsis* homologue	Arabidopsis description	Arabidopsis e-value
F - 21, 7	35.97	At4g11650	osmotin precursor	4.00E-28
E - 18, 16	33.74		no hit	
H - 14, 24	25.78	At4g25810	xyloglucan endo-1,4-beta-D-glucanase (XTR-6)	1.00E-66
K - 24, 15	23.30		no hit	
N - 12, 15	23.10	At3g12500	hypothetical protein	e-124
O - 20, 15	22.84		no hit	
N - 14, 18	22.23		no hit	
D - 18, 16	21.78	At3g09030	hypothetical protein	7.00E-53
L - 24, 16	20.85	At4g16260	beta-1,3-glucanase class I precursor	2.00E-42
L - 14, 19	20.68		no hit	
E - 22, 18	18.95	At5g05340	Peroxidase	1.00E-95
D - 10, 27	18.24		no hit	
M - 1, 3	17.11	At1g29930	putative protein	8.00E-65
C - 14, 16	16.95		no hit	
O - 19, 15	16.28	At4g31700	ribosomal protein S6 - like	2.00E-50
O - 15, 17	14.79	At5g67360	cucumisin-like serine protease (gb|AAC18851.1)	6.00E-39
O - 14, 17	13.61		no hit	
F - 22, 18	13.46	At4g19810	putative chitinase	9.00E-33
F - 16, 1	12.89	At3g12500	hypothetical protein	e-124
O - 9, 13	12.86		no hit	
F - 18, 16	12.61	At2g32700	unknown protein	7.00E-54
D - 5, 4	12.32		no hit	
D - 9, 24	10.82	At3g24500	ethylene-responsive transcriptional coactivator	6.00E-47
I - 2, 29	10.45	At3g54160	gf_c_21 putative protein	0.06
L - 24, 15	10.01		no hit	
I - 19, 13	9.53	At1g29930	putative protein	8.00E-65
M - 4, 22	9.09	At3g07310	unknown protein	2.00E-16
A - 8, 24	8.88	At3g24500	ethylene-responsive transcriptional coactivator	6.00E-47
A - 19, 15	8.66		no hit	
J - 5, 23	8.53		no hit	
M - 16, 20	8.32		no hit	
N - 15, 17	8.18	At4g28690	hypothetical protein	7.00E-12
N - 2, 12	8.04		no hit	
H - 12, 16	7.98	At4g32940	gamma-VPE (vacuolar processing enzyme)	8.00E-53
H - 16, 29	7.94	At5g23575	MQM1 unknown protein	2.00E-11
H - 11, 18	7.86	At1g25580	unknown protein	1.00E-49
E - 7, 14	7.46		no hit	
D - 18, 14	7.46	At4g21200	gibberellin 20-oxidase - like protein	1.00E-29
I - 21, 29	7.37	At4g23100	SA_D_20 gamma-glutamylcysteine synthetase	3.00E-69
E - 21, 14	7.34	At4g16560	hypothetical protein	5.00E-19

The clone H-14, 24 was present in both tables, 5.7-A and 5.7-B. At the protein level it showed homology to a xyloglucan endo-1,4-beta-D-glucanase deduced from the

Arabidopsis thaliana gene At4g25810. In the first two Tables 5.6-A and 5.6-B the clone,
L-9, 8, showed homology to the same deduced gene product. The hybridization values for the two clones were checked and proved reliable. Xyloglucans endo-1,4-beta-D-glucanase are known to be a gene family (Rose *et al*., 2002) and that other related xyloglucan glucanase gene families exist at least in *Arabidopsis thaliana* (Xu *et al*., 1996). Different members with different expression pattern, but presenting conserved domains could be the reason for this apparently contradictory result.

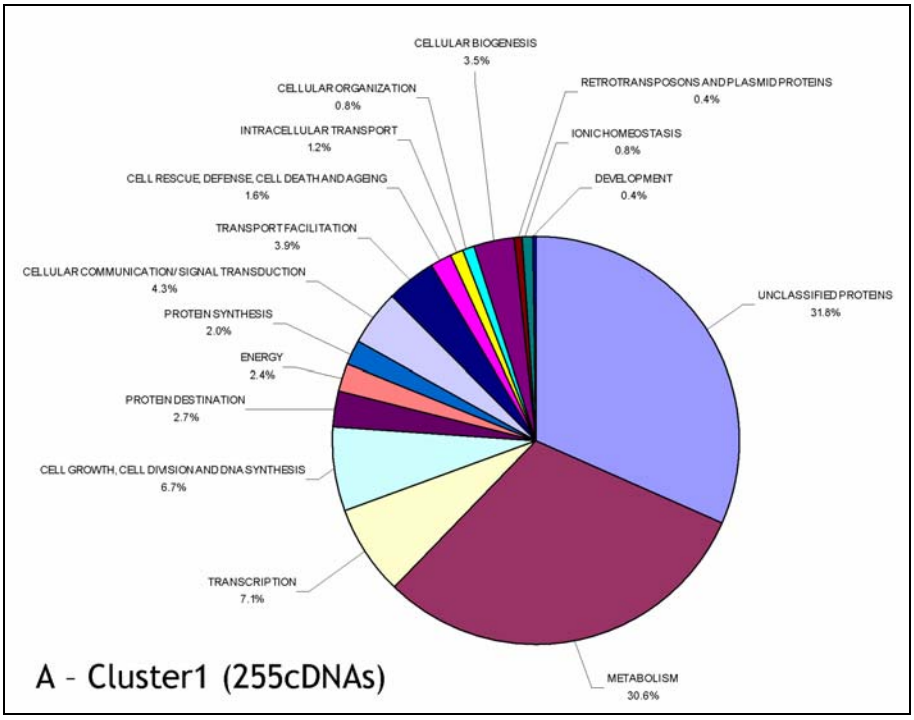
5.3.6 - Classification of the clones belonging to the two identified clusters in different functional categories

The clones belonging to cluster 1 and to cluster 2 were classified with respect to functional categories deduced by domain analysis of the corresponding EST sequences as described for the entire library.
Among 599 sequences derived from clones with preferential expression in the first part of the development, only for 255 a functional domain was identified. Concerning the cluster of 175 genes preferentially expressed in the second part of the development, only for 67 the presence of a functional domain was retrieved (Table 5.8).

Table 5.8 Classification of EST sequences of the cDNAs belonging to clusters 1 and 2 identified in the time-course experiment.

	Cluster1	Cluster2
Total number of cDNAs belonging to the cluster	599	175
Sequences of bad quality or not considered in this analysis	137	31
Sequences containing no functional domain	207	77
Sequences containing a functional domain	255	67

The 255 and 67 sequences, for which a functional domain was identified, were grouped according to the different functional categories (Figure 5.10). List of the clones and relative annotation grouped according to the functional category for both clusters are



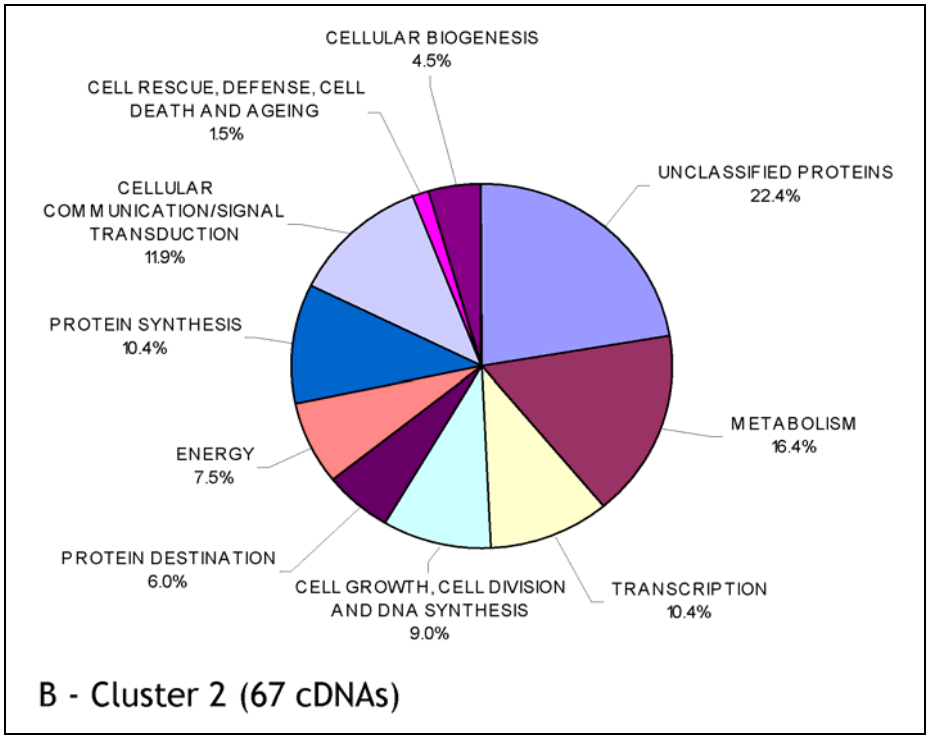


Figure 5.10 Classification of gene products containing a functional domain in cluster 1 (A) and in cluster 2 (B) into the different functional categories.

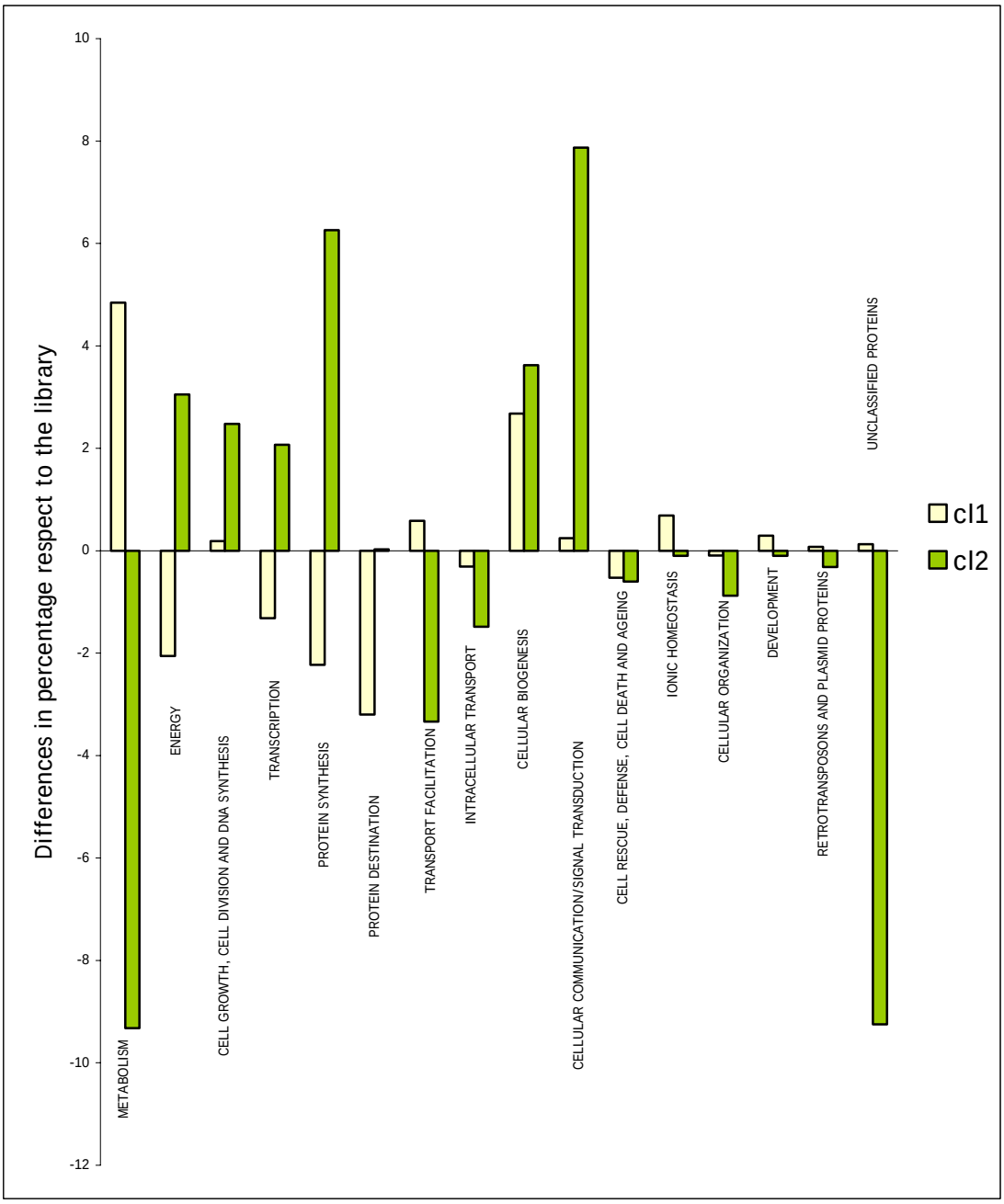


Figure 5.11 Representation of functional categories in cluster1 (cl1) and cluster2 (cl2) expressed as difference between the percentage values for the respective cluster and the entire library.

reported in the annex. Due to the small size, less functional categories were represented in the second cluster by the identified clones.

To visualize and to interpret these results, the percentage of sequences belonging to each functional class in the clusters was compared to the percentage of sequences belonging to the same class in the entire library. The difference between the two percentage values is indicated in Figure 5.11. This graphic presentation highlights over- or under-represented functional categories in each of the two clusters.

For the first cluster, represented in yellow in Figure 5.11, the highest up-regulated functional class was the metabolism, and the highest down-regulated class was protein destination. The categories “genes transcription” and “protein synthesis” were as well down-regulated. Up-regulation was observed for the functional category “cellular biogenesis”, as well as for “ionic homeostasis” and “transport facilitation”, all potentially important for sucrose accumulation. A slight up-regulation was observed in this cluster for the functional class “development”. Concerning the second cluster, the functional category “metabolism” showed the strongest down-regulation compared to the entire library. Therefore this functional class was the one showing the greatest difference between the two clusters when compared to the composition of the entire library. A strong up-regulation for the functional categories “signal transduction” and “cellular biogenesis” was observed. As opposed to the results for the first cluster, a larger fraction of sequences encoding domains involved in transcription and protein synthesis as well as cell growth and energy metabolism were identified in this second cluster.

5.3.7 - Results for the candidate genes preferentially expressed in beets

The 76 candidate genes identified as preferentially expressed in roots (chapter 4), were spotted on the same filter as the A024 cDNA library, but analyzed separately in the time-course experiment (Table 5.9).

Among the 76 clones preferentially expressed in roots, 14 show a differential expression during beet development confirmed in both years. Thirteen of them were preferentially expressed in the first part of the development, and one cDNA showed preferential expression in the second part of the development. Among the 14 identified clones for seven an analogue was retrieved as already explained in chapter 4. In particular should be noted that the EST BQ490013 encoding a different sucrose synthase respect to the already known sucrose synthase *sbss* of sugar beet (Hesse and Willmitzer, 1996, Schneider *et al*. 1999), was shown here to be preferentially expressed in the first part of the development.

Table 5.9 Preferential expression of the preferentially root-expressed candidate genes (chapter 4) in the time-course experiment. Accession numbers of the candidates and annotation are reported again together with data on the preferential expression. cl1=cluster 1, clones preferentially expressed in both years in cluster 1, cl2= cluster2, clones identified as preferentially expressed in cluster 2 in both years

Acc. N° ESTs Seq.	Annotation	Acc. N° A.A. Sequence	E-value	Preferential expression in dev.	Marker
BQ490013	Sucrose synthase	CAA57881	1.00E-109	cl1	9F09
BQ654408	Expressed protein			cl1	
BQ489314	Nodulin-like protein	NP_565111	5.00E-54	cl1	7I24
BQ488951	Peroxidase	NP_201440	2.00E-77	cl1	6G14
BQ488455	PM28B protein	CAB56217	1.00E-63	cl1	
BQ489160	S-like ribonuclease	AAF82615	4.00E-34	cl1	
BQ487564	Xyloglucan endotransglycosylase	AAF80591	6.00E-53	cl1	extu110
BQ488151				cl1	2F12
BQ489294				cl1	7G18
BQ489439				cl1	7P14
BQ488187				cl1	
BQ489728				cl1	
BQ488661				cl1	
BQ488580				cl2	

5.4 – BIOLOGICAL AND TECHNICAL VALIDATION OF DIFFERENTIAL EXPRESSION DURING THE DEVELOPMENT

As already explained in paragraph 4.3 macroarray data need independent validation by other techniques. In this experiment, macroarray data were confirmed by semi-quantitative RT-PCR using sequence-specific primers (Figure 5.12).

The template poly(A)$^+$RNA samples used for RT-PCR were the same as used for the preparation of complex probes in the array experiments, to minimize the variation introduced by the samples. In each case, the number of amplification cycles was adjusted to the abundance of the respective transcript to obtain PCR products in a linear range.

In Figure 5.12 the results for six candidate genes of cluster 1 identified in the time-course experiment are reported. For each type of expression pattern clones from the EST collection A024 and from the preferentially beet expressed clones derived from EST collection A006 were selected.

The clone J-3,12 showing homology at the aminoacidic level to GAST1-like protein (see also Table 5.6-B), the clone E-9,9 annotated as putative carboxyphosphonoenolpyruvate mutase and the clone I-9,26 (see annex) were selected to perform the RT_PCR experiments.

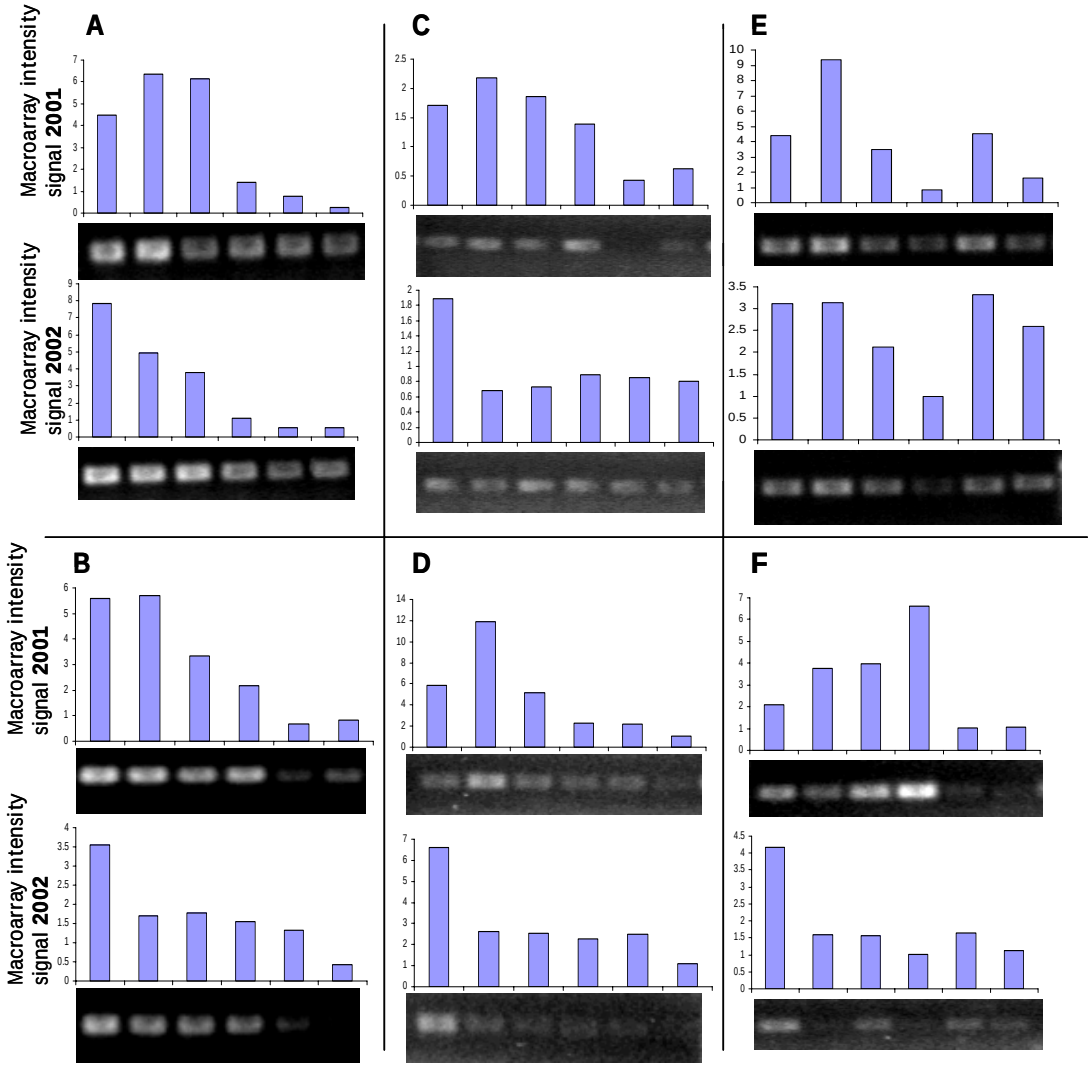


Figure 5.12 Semi-quantitative RT-PCR for clones of the first cluster as validation tool for macroarray time-course results . Semi-quantitative RT-PCR with EST-specific primers was performed based on cDNA synthesized from the same samples (2001 and 2002) as used for the time-course experiment. Number of cycles used in the RT-PCR is optimized for each clone as explained in the text. A=clone J-3,12 annotated as GAST1-like protein, B=clone E-9,9 annotated as putative carboxyphosphonoenolpyruvate mutase, C=clone I-9,26 annotated as ethylene-responsive element -like protein, D=clone encoding sucrose synthase belonging to the candidate identified in the first experiment (Acc N. BQ490013), E= clone encoding PM28B protein belonging to the candidate genes identified in the first experiment (Acc N. BQ488455), F= clone annotated as xyloglucan endotransglycosylase belonging to the candidate genes identified in the first experiment (Acc. N. BQ487564) with 100% identity at nucleotidic level to the clone L-9,8 belonging to library A024.

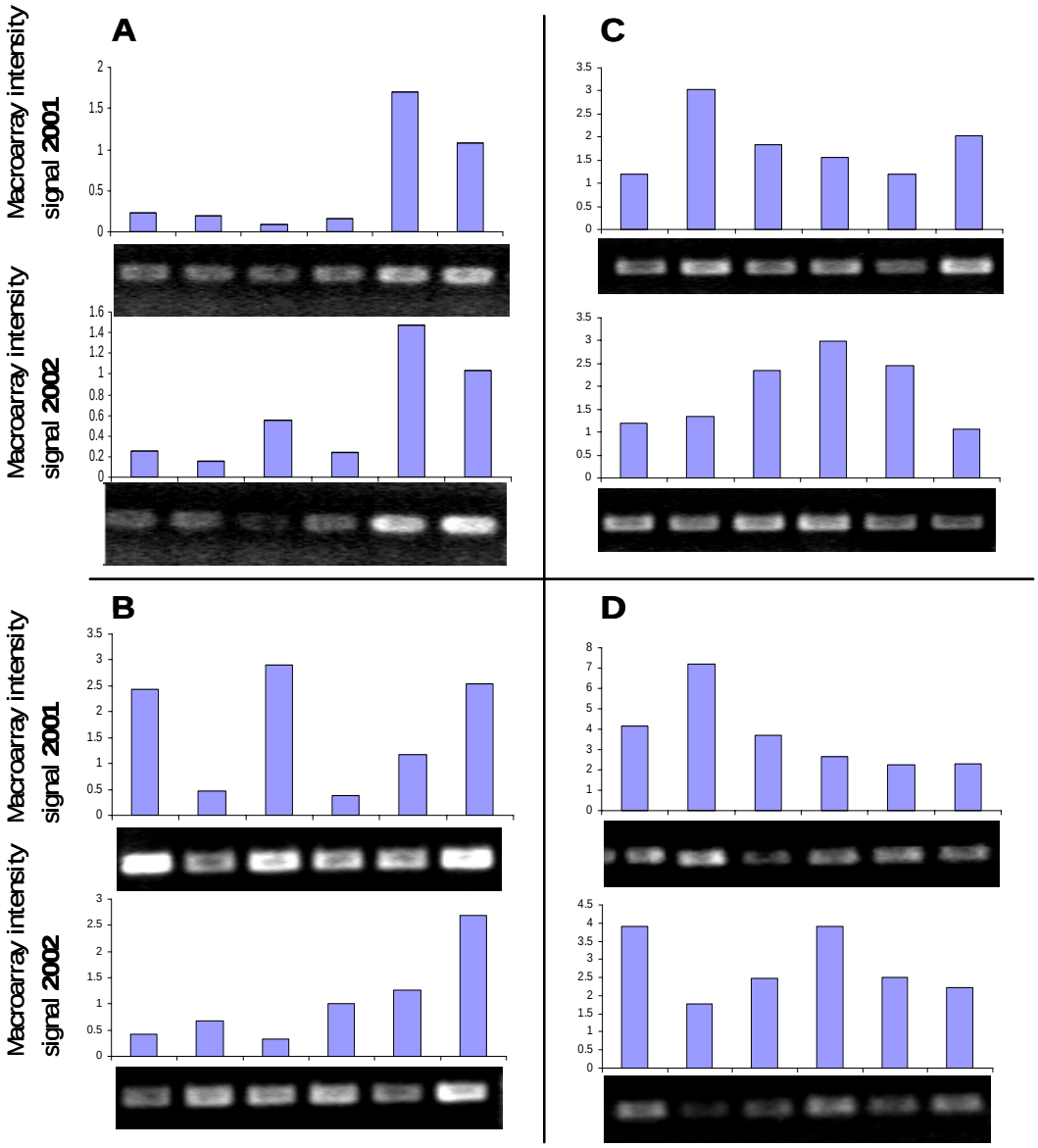


Figure 5.13 Semi-quantitative RT-PCR for clones belonging to the second cluster as validation tool for macroarray time-course results (A=clone L-24,16 annotated as beta-1,3-glucanase class I precursor and B=clone N-2,12 not annotated) and for two more interesting cases not belonging to the clusters. C=clone annotated as putative phosphoenolpyruvate carboxylase kinase (Acc.N. BQ488374) and belonging to the candidate genes identified for the library A006 and D=clone annotated as sucrose synthase-beet and belonging to the candidates identified in the library A006 with100% identity to the clone P-16,24 belonging to library A024.

The sequence relative to this last clone was showing similarity to an ethylene-responsive element - like protein. For the first two (Figure 5.12-A, B) good agreement between RT-PCR data and macroarray results was shown. The RT-PCR expression pattern of the third analysed clone (Fig. 5.12 C) was different from the macroarray result, especially in the year 2002. Three more clones belonging to the candidate

genes identified in chapter 4 and assigned to cluster 1 (paragraph 5.3.7) were used as well for validation. The clones encoding a new sugar beet sucrose synthase identified in chapter 4 and a gene product with homology to a PM28B protein, respectively, were selected. Additionally, the clone with the EST BQ487564 annotated as xyloglucan endotransglycosylase was chosen to perform RT-PCR. This clone showed 100% homology to the clone L-9,8, present in the library A024, at nucleotide level. Both clones belong to cluster 1 with maximum expression in the first part of the development (Table 5.6-A and B and Table5.9). Good reproducibility of the expression patterns during the development was shown for all three clones by the two employed techniques (Figure 5.12-D, E, F).

Concerning reproducibility of the expression patterns in the two years 2001 and 2002, the requirement, as already discussed for the time course experiment, was in general respected. A general delay in the first year was as well observed, in agreement with the data about the two seasons reported in Table 5.3.

Figures 5.13-A and B summarize RT-PCR results for candidate genes induced in the second part of the development (Table 5.7-B). One clone (L-24,16) encoded a gene product similar to a beta-1,3-glucanase and the other (clone N-2,12) showed no similarity to any protein under the conditions already explained in this chapter. In both cases macroarray results were very well reproduced by RT-PCR. Concerning the biological reproducibility comparable expression patterns were observed in the two years for the first clone. The second clone showed some differences in the expression pattern, confirmed by both techniques.

In this figure two more clones were considered which showed either differential, but not reproducible patterns in the two years (Fig. 5.13-C) or no differential expression according to the statistical analysis (Fig. 5.13-D).However, the latter clone with the EST BQ489399 encoding the already known sucrose synthase *sbss* of sugar beet (Hesse and Willmitzer, 1996), showed some differences in expression, not reproducible in the two years, when analyzed either with macroarrays or RT-PCR.. From the technical point of view, RT-PCR was confirming the results of the macroarray for both clones (Figure 5.13-C, D).

Figure 5.14 presents RT-PCR results for clones showing no differential expression during the development according to the statistical analysis of the macroarray data. One clone with similarity to an alanine and glutamic acid rich protein was selected

from the candidate genes identified in chapter 4. No significant differential expression was detected by the statistical analysis of the macroarray data, and this result was confirmed here by RT-PCR (Figure 5.14-A).

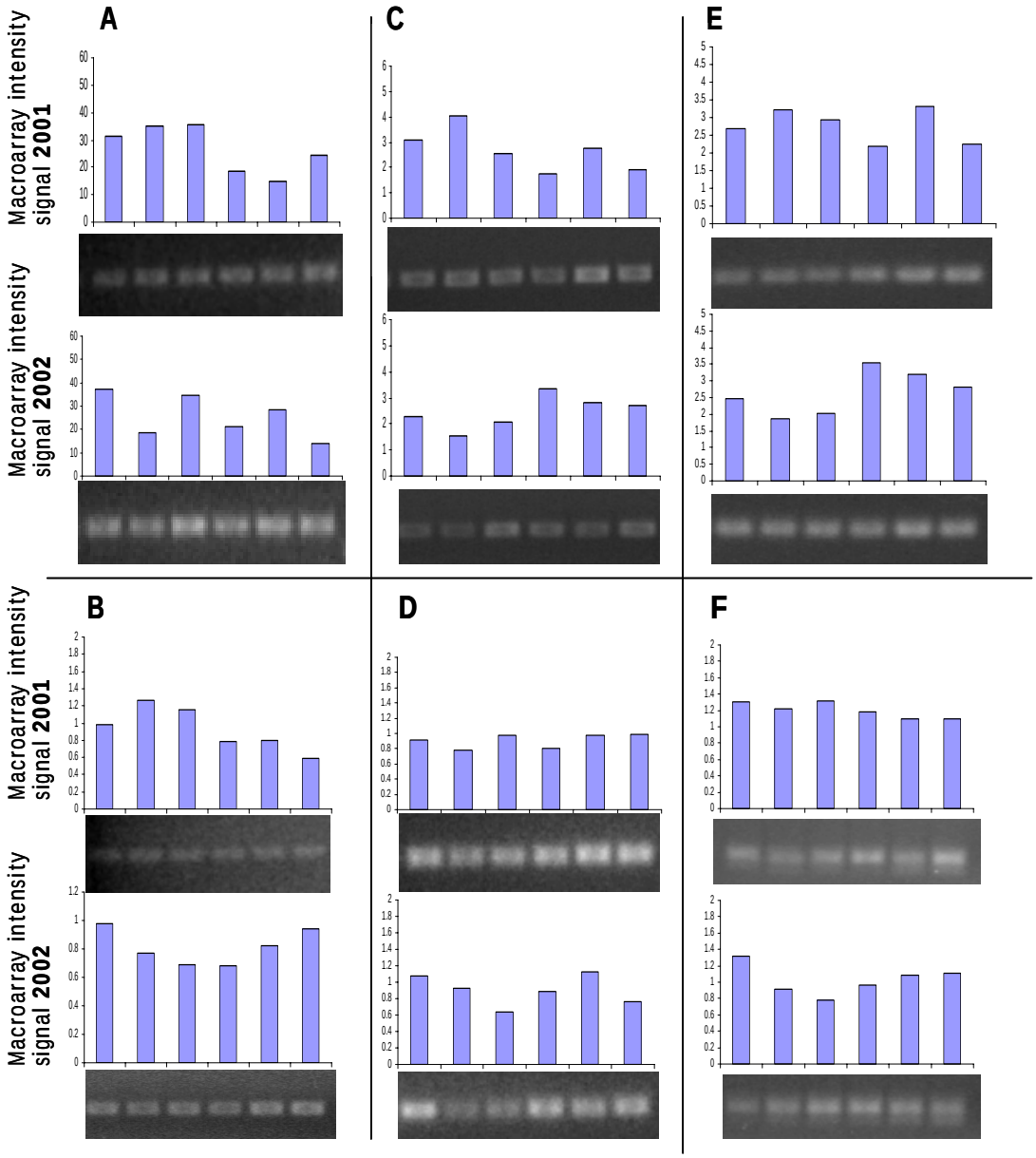


Figure 5.14 Semi-quantitative RT-PCR for clones showing no significant differential expression during the development in the two years 2001 and 2002, as validation tool for macroarray time-course results . A= clone annotated as alanine and glutamic acid-rich protein and belonging to the candidate genes identified for the library A006 (Acc. N. BQ489156), B= clone K-9,14 annotated as putative AAA-type ATPase, C= clone belonging to the library A006 annotated as GTP binding protein (Acc. N. BQ490211). D=clone belonging to library A006 (Acc. N. BQ490231) and encoding an aminoacylase , E= clone belonging to library A006 (Acc. N BQ488925) and encoding a 60S ribosomal protein F= clone F-10,8 not annotated.

Additionally, two clones derived from library A024 (clone K-9,14 and F-10,8 in Figure 5.14-B and F, respectively) were selected. Three more clones derived from the library A006 and represented by the ESTs with the accesssion numbers BQ490211, BQ490231, BQ488925, were included on the arrays of the EST collection A024 as putative “housekeeping genes” because they showed equal expression in the three different organs of sugar beet. No significant differential expression was reported for these clones during the development by macroarray analysis, and the RT-PCR results in Figures 5.14- C, D and E confirmed this result.

6 - DISCUSSION

6.1 - SUGAR BEET MACROARRAYS

The aim of the presented work is to identify genetic factors involved in the process of sucrose accumulation in sugar beet roots. The capacity to store as much as 20% sucrose per fresh weight is enabled by a special root morphology and physiology and by the sucrose transport from leaf mesophyll cells, where sucrose is produced, to the roots, where it is stored in the vacuoles of the parenchymatic cells (Elliott and Weston, 1993). To study this mechanism known genes involved in pathways related to sugar metabolism have previously been mapped and tested for their association with QTLs for sugar yield and quality (Schneider *et al*., 1999, 2002). However, this approach excludes gene products playing a role in less conserved reactions which may be sugar beet- specific. To identify new candidate genes involved in this storage process gene expression was analyzed in roots during beet development and sucrose accumulation. To analyze gene expression the macroarray technology was selected because it represents a powerful multiparallel approach without requiring any previous knowledge of the pathway.

The possibility to use high-density DNA arrays as tools to study gene expression was first demonstrated by Schena *et al*., (1995) at a small scale on *Arabidopsis thaliana* to monitor the expression of 45 Arabidopsis genes. De Risi *et al*., (1997) reported the first genome wide application of microarray to produce yeast expression profiles during the shift from fermentation to respiration. Since that time this technology was adapted and widely used to study changes in gene expression in the model plant *Arabidopsis thaliana* under different experimental conditions like in different tissues (Ruan *et al*. 1998), or as answer to the circadian clock (Harmer *et al*. 2000). Many reviews are available about all the steps involved in the technology (Aharoni and Vorst, 2001, Churchill 2002, Quackenbush 2002, Leung and Cavalieri, 2003 and others). At present it is the aim to develop standards of minimum information (MIAME, Minimum Information About Microarray Experiments) at least for microarray data (Brazma *et al*, 2001), in order to promote reproducibility of the data by other scientists if standardization is still not feasible. The flexible nature of the

fabrication and hybridization methods of cDNA micro- and macroarrays allows the application of the technology to non-model organisms as well. Aharoni *et al*. (2000), for example, employed the microarray technology to follow ripening in strawberry. In the work presented here multiparallel expression profiling using different cDNA macroarrays was established and applied to sugar beet (chapters 3, 4, and 5). A reliable procedure for macroarray analysis was developed. To assess the quality, controls were included on the arrays, and technical parameters of the produced macroarrays and hybridization experiments were evaluated and compared to available data in literature.

6.1.1 – Sugar Beet macroarray sensitivity

In Desprez *et al*. (1998) the first example of a macroarray study in plants is reported. In this study the expression profiles of light–grown and dark-grown *Arabidopsis thaliana* seedlings are compared. Using spiked controls the authors define the sensitivity limit for the employed system. They conclude that RNA species with an abundance of 0.01% of the poly(A)$^+$RNA used to produce the probe can be detected. This finding is in agreement with previous reports for macroarray in other organisms (Zhao *et al*., 1995). For the sugar beet macroarray generated here, the signal intensity which was produced by the control spiked as 0.01% of the poly(A)$^+$RNA used for probe synthesis, always exceeded the signals for unspecific hybridization (Table 3.1). Therefore, it is concluded that the detection limit for the sugar beet macroarray was at least comparable to the detection limit reported in previous studies. Thus, with the established system it was possible to detect transcripts present in as little as 10 copies per cell.

Additionally, a region in which the signal intensity was found to grow proportionally to the amount of poly(A)$^+$RNA added to the probe was defined. For 75% to 85% of the clones in the macroarray, the signal intensity values fell in this range. This allowed to infer information on their quantitative expression as well.

For microarrays, more sensitive results were reported including the detection of single-copy transcripts (Ruan *et al*., 1998). However, Bertucci *et al*. (1999) reported an interesting experiment in this context. Defining sensitivity independently of the

amount of poly(A)$^+$RNA used in the hybridization, they demonstrated that the two technologies, macroarray and microarray, show comparable performances. With additional experiments they showed that this is mainly due to the larger amount of target present on filters compared to the target present on glass slides. This dependence on the amount of target spotted was also demonstrated in the experiments shown in Figure 3.6-B. To correct for mistakes introduced by a different concentration of the spotted samples, the intensities of the hybridization signals were divided by the corresponding signal intensity produced by the hybridization with the labeled T7-short oligo complementary to all PCR products. This step was considered important here although other authors of macroarray analysis (Desprez *et al*., 1998) omit it.

A biological approach to assess the sensitivity of the macroarrays is to test them on genes known to be transcribed at low level, like R-genes (resistance genes). Using the established technology, for 20 of the 29 RGAs (resistance gene analogues) present on the macroarray of the A006 cDNA clone set, transcripts were identified in uninfected leaves and/or roots, and for seven of them a differential expression was shown (Table 3.2). This result illustrates the sensitivity of the method. The identified expression patterns coincide with what is known for proven R-genes. Basal transcript levels were detected in uninfected plants for genes like *Rps2* and *Cf-9* (Mindrinos *et al*., 1994, Jones *et al*., 1994). The biological explanation for this may be that the cost of constitutive low expression of this type of genes is compensated for by an immediate response in case of pathogen attack.

6.1.2 – Sugar beet macroarray reproducibility

The variability among filters, due mainly to technical limits of the robot employed for spotting, was quantified in the present study for four filter pairs and shown to involve 10.2±1.0% of the spots. To correct this effect an experimental hybridization design involving hybridizations of replica of the same sample to different membranes was employed. This problem is omitted when using the microarray technology because simultaneous hybridizations of differently labeled probes largely eliminate errors caused by the use of different arrays.

Further sources of variability, which are not related to differences immanent to the biological system, were probe synthesis and sampling. Both of them were quantified and variability related to sampling of field grown plants was shown to be the most relevant, affecting 30.3±3.7% of the cDNA clones when eight different comparisons were analyzed. To avoid false positives due to the environmental effects and subtle differences in the developmental program, two biological replicates of field-grown plants were harvested in a two weeks interval and included in the study. Only spots producing similar hybridization signals in the two biological replicates were considered in the analysis. Consequently, 15.8% of all unique sequences were eliminated as outliers, but this procedure was regarded as necessary to avoid misinterpretations.

6.1.3 – Validation of macroarray results by other technologies

Macroarray reliability is also assessed by comparing the results to those obtained for the same genes with other experimental approaches. Concerning the already known sucrose synthase *sbss* (Hesse and Willmitzer, 1996) the preferential expression in root revealed by Northern blot analysis was confirmed here by the macroarray procedure. Four clones, a putative phosphoenolpyruvate carboxylase kinase (BQ488374), an alanin and glutamic acid-rich protein (BQ489156), a putative protein (BQ490572) and a (BQ490059) jasmonate-induced protein homolog, respectively, showed high homology to gene products for which preferential expression in sugar beet roots was identified independently by SSH (suppression subtractive hybridization) and cDNA AFLP (Kloos *et al*., 2002, Schneider, pers. com.) .

As additional proof of the reliability of the macroarray analysis performed, the results for selected cDNA clones were validated by RT-PCR and Northern blot analysis (chapters 4 and 5).

Concerning the first macroarray experiment using the library A006, expression patterns were confirmed in all the eleven cases studied. In the macroarray experiment, cDNA clones were considered automatically preferentially expressed if they showed at least two-fold higher expression in one organ with respect to the other(s). The

stringency of the threshold of 2.0 selected for preferential expression was obvious when comparing expression ratio values and data from RT-PCR. ESTs with expression ratio values between 1.0 and 2.0 were shown to be as well preferentially expressed by RT-PCR. However, they were below the defined threshold and appear as false negatives in the automatic analysis. On the other side, only a very high stringency ascertains the avoidance of false positives and allows to focus on real candidates.

For the second macroarray, produced from library A024, the expression patterns of sixteen genes followed during the development in both years considered in the experiment, were reanalyzed by RT-PCR. With the exception of one case, the profiles were confirmed.

In conclusion, all these observations proved that the established macroarray procedure was reliable. Therefore, the results concerning differential expression generated in the two different experiments with macroarrays are considered real.

6.2 - IDENTIFICATION OF CANDIDATE GENES PREFERENTIALLY EXPRESSED IN ROOT

The macroarray technology was used to analyse the cDNA collection A006 with 3840 cDNA clones generated from young sugar beet plants. A preferential expression in root tissue of sugar beet was considered a selection criterion to identify candidate genes involved in sucrose accumulation and storage. As this process is localized in the parenchymatic cells of the taproot, the genes involved are assumed to be preferentially expressed in the root at the same time. In the performed experiment, clones were classified with respect to their expression in three different organs, the root, the leaf, and the inflorescence. The comparison to leaves was selected in other to eliminate ubiquitously expressed genes and the comparison to inflorescences to discriminate between beet-specific and sink-specific preferential expression. As a result seventy-six candidate genes preferentially expressed in sugar beet roots were identified. Further information on these clones was inferred from their sequences.

The combination of EST data mining and macroarray analysis revealed a powerful tool to identify candidate genes. Their putative function in relation to taproot formation and sucrose accumulation is discussed.

6.2.1 - Sequence analysis of 2996 sugar beet ESTs

A collection of 2996 EST sequences was generated from library A006 containing 3840 cDNA clones. Sequencing cDNAs from a random collection derived from a particular sample serves two purposes: the discovery of new genes and the assessment of their expression levels. This is possible because the expression level of an mRNA species in a specific tissue or at a defined stage is reflected by the frequency of its corresponding EST in a cDNA library. Therefore analysis of EST redundancy is an efficient method to obtain information on genes transcribed in a particular tissue or at a defined stage.

This approach is of special importance to organisms with a large genome, for which little molecular information is available, as recently shown for pine (Whetten *et al*., 2001). For this purpose, cluster analysis on the EST dataset was performed to assess the redundancy of each EST in the dataset.

In summary, 2048 unique sequences were identified, each possibly defining a different gene. However, as sugar beet is an outbreeding species and the used plant material is diploid, it is possible that two different unique sequences represent two alleles of the same gene. On the other side, a chance exists that paralogues with little sequence deviation were unified in one cluster. In both cases, only segregation analysis and genetic mapping can determine the exact number of genes contained in the presented EST collection.

The cDNAs encoding the small subunit of ribulose bisphosphate carboxylase represented 1.3% of all ESTs, and therewith the most abundantly expressed transcript in the source material of the library (Table 4.1). The importance of photosynthesis-related processes is furthermore illustrated by at least 14 other chloroplast–localized gene products deduced from the unique sequences of the 36 largest clusters. Each of these unique sequences accounts for at least 0.2% of all ESTs. Only three preferentially root-expressed clusters with six to maximally eight members were

identified. Sequences of clones in these clusters showed homology to the gene products sucrose synthase, DnaJ protein and to a reversibly glycosylated polypeptide putatively involved in biosynthesis of the primary and secondary cell wall (Bocca *et al*., 1999).

6.2.2 - Putative functions of preferentially root-expressed genes

The macroarray analysis identified eighty-six cDNAs representing 76 unique sequences showing preferential expression in the root (Figure 4.2, Table 4.3). Forty-three gene products with predicted functions were distributed in nine different classes, namely carbohydrate metabolism, transfer of sugar moieties, cell wall architecture, oxidative processes, organization of cytoskeleton and membrane assembly, intra- and intercellular transport and transfer processes, ATP metabolism, RNA metabolism and protein synthesis, and signal transduction. The first five classes are associated with processes such as sucrose accumulation as well as cell division and expansion during beet formation. Sucrose accumulation and growth are two concomitant processes in the beet because a separate ripening phase for the uptake of assimilates in the sink tissue does not exist (reviewed in Elliott and Weston, 1993). Interestingly, some preferentially beet-expressed genes have closely related analogues expressed in the ripening period of non climacteric fruits such as strawberries and grapes (Aharoni *et al*., 2002, Davies and Robinson, 2000). Parallels in the underlying processes can be found in vascular tissue development, the establishment of a stress response and possibly the regulation by the growth factor auxin in both systems.

Mesophyll cells of *Z. elegans* are the model system for *in vitro* tracheary element (TE) differentiation and gene expression in these cells has been subdivided into distinct phases (Fukuda, 1997, Milioni *et al*., 2001). The first phase involves changes in RNA and protein metabolism to adapt the gene expression, and in sugar beet roots this phase is represented by three ESTs. The second phase comprises the organization of cytoskeleton and membrane assembly. In our collection of preferentially root-expressed genes there were four ESTs with putative functions in these processes. Among them was an annexin with strong similarity to an annexin from ripening strawberry (Wilkinson *et al*., 1995). Annexin genes are members of multigene

families with a variety of functions, but annexin transcripts were not detected in roots of strawberry, indicating the existence of function- or process- rather than organ-related expression patterns across different genomes. A lipid transfer protein with a possible role in TE differentiation is also in common between sugar beet roots and ripening strawberries (Aharoni *et al*., 2002).

The third phase in TE differentiation concerns the biogenesis of primary and secondary cell walls, and five putative candidates were identified in sugar beet roots. The two root-expressed peroxidases could also assume a function in this process, but they could also be part of a stress response. In fact, the establishment of a stress response is another parallel between sugar beet roots and ripening fruits. The two preferentially root-expressed, jasmonate-induced gene products listed in the group of signal transduction are also likely to be part of a stress response. Potential reasons for their expression may be similar between sugar beet roots and ripening fruit because stress is caused by cell expansion and the increase in the osmotic potential due to sugar accumulation. A sucrose content of 20% related to the fresh weight of mature sugar beet roots is comparable to the sugar content of grapes (Davies and Robinson, 2000, and references therein). Alternatively the stress response is seen as a prophylactic program against pathogen attack. The occurrence of peroxidases with preferential expression in the roots of soil-grown *Arabidopsis thaliana* (Ruan *et al*., 1998), which are neither known to accumulate sugars or to expand to the degree sugar beet roots do, appears to support the last reason. The exposure to the spectrum of soil-borne pathogens may necessitate an elevated defence program in roots.

A further parallel between sugar beet roots and ripening fruit is the occurrence of auxin-repressed proteins, of which two representatives were identified by sequence similarity in our study. Their direct function remains unknown, but auxin-repressed gene products are known to be involved in processes like cell wall biosynthesis or the establishment of a stress response. Only more physiological experiments will reveal a complete catalogue of genes affected by plant growth factors in sugar beet roots. Interestingly our analysis also identified a root-expressed globulin-like gene, for which no direct function can be predicted, but a gene product with a similar homology was identified among auxin-independent and ripening related genes in strawberry (Aharoni *et al*., 2002).

6.3 - IDENTIFICATION OF CANDIDATE GENES RELATED TO THE DEVELOPMENT

In a second series of experiments, the macroarray technology was used to analyze cDNA collection A024 of 11520 clones generated from sugar beet leaves, young and mature roots as well as inflorescences. To identify candidate genes related to the development plants were characterized morphologically during two different years. Concerning gene expression, two main expression patterns were identified during the vegetative phase of growth. As a result, 599 clones showed confirmed preferential expression in the first part of the development in both years. Further 175 clones were expressed mainly at late stages of the development in both years. A search for functional domains in the corresponding EST sequences was the basis for classifying these ESTs into functional categories. As the number of genes preferentially expressed in each of the two groups is too high to allow a detailed analysis of every single gene (Annex), conclusions are drawn about developmentally regulated processes and the corresponding candidate genes by inspection of the functional categories present in the two clusters.

6.3.1 - Morphological characterization of sugar beet development

To follow plant development, six parameters were selected to be measured at each harvesting time-point. They were considered indicative of different agronomical traits. For crops, the number of leaves is correlated to the plant's productivity. This can be considered true also in sugar beet because sucrose is first synthesized in the leaves, from which it is then transported to the roots where it is stored. Additionally, tthe number of leaves was considered as parameter to follow the development of the aerial portion of the plant to allow the comparison to root development.

However, the focus of this study was on the storage organ, therefore beet growth was characterized using five parameters. Agronomically, the main goal of breeding sugar beet is to improve sugar yield, the product of beet yield and sucrose content. In different studies these two traits were negatively correlated (Elliott and Weston, 1993). Therefore the simultaneous optimization of the two component characters of

sugar yield is difficult to obtain and varieties have been developed to optimize one or the other of these traits.

Beet yield can be followed during the development measuring the thickness and the weight of the beet.

Concerning sucrose content, sucrose concentration per fresh weight can directly be measured, but also the number of rings. Ring density, for example, defined as number of rings per centimeter is used as selection character for breeding work, because it showed correlation with sugar concentration. This is obvious because sucrose enters the storage organ through the phloem. The more phloem layers there are as derivatives from secondary cambium rings, the more sucrose can be delivered.

As a last parameter, root length was measured. This parameter usually depends on the water availability in the soil, and is therefore scarcely related to sucrose yield.

According to the results in Figure 5.2-A and B, root thickness and the number of leaves grow simultaneously during the development. Early results in sugar beet research showed that shoot and root growth occupied separate phases of the development. They believed that the vegetative development of the plant was divided into three distinct phases: a phase of leaf formation from emergence until end of July, a phase of root formation during August, and a phase of sugar storage or ripening occupying the rest of the season. More recent work led to the current view, in agreement with findings presented here, that sugar beet does not show separate growth phases, neither does it exhibit a ripening phenomenon (Elliott and Weston, 1993).

It is known that most of the secondary meristematic rings are laid down very early in the development of sugar beet (Milford, 1973). This is confirmed in this study (Figure 5.2-C and 5.3) where it is shown that by nine weeks six to nine rings were visible. The cambium rings are known to develop simultaneously rather than sequentially (Elliott *et al*., 1984).

To elucidate the thickening process of sugar beet, Rapoport and Loomis (1986) reported a comparative study of thickening in sugar beet and chard, a member of the *cicla* group (foliage beet) of the species *Beta vulgaris,* with a much smaller root. They found that the differences in thickening cannot be accounted for by a difference in the number of cambia. The two root types appeared to differ mainly in aspects of cell expansion, especially concerning the cells of the inter-cambial parenchymatic region.

Therefore cell expansion rather than cell division seems to be relevant for the thickening of sugar beet.

As a further observation, it should be mentioned that lateral expansion involves only about half of the cambia. This was observed also in the study presented here although data on the degree of expansion per ring were not collected. This finding together with the observation that sucrose concentration is higher in proximity of the vascular regions, where cells are smaller, and lower in the expanded parenchymatic region, where cell size is larger (Milford, 1973), leads to the current view that increasing the number of vascular zones and shortening the diffusion path between phloem and storage vacuoles could improve the sucrose concentration in the storage root. This is the basis for the efforts to understand processes like ring initiation and cell expansion involved in beet development, the topic of this developmental study.

In conclusion, the morphological data presented here show good reproducibility between the two years. However, the seasons 2001 and 2002 are different, if GDD (growing degree days) are considered as parameters to compare the years. Therefore, even if similarity in the development was demonstrated in the two years a strict alignment of time points was avoided. Gene expression patterns reproducible during the development were considered and compared to the described developmental processes.

6.3.2 – Candidate genes for developmental processes

The time-course macroarray analysis concerning the development identified two main groups of clones expressed and confirmed in both years analyzed: one with maximum expression in the first part of the development and the other with highest transcript levels at maturity. A search for functional domains was performed on all EST sequences of the library. Therefore the clones belonging to the two clusters could be grouped according to the respective functional categories. Lists of the clones belonging to cluster 1 and cluster 2 grouped according to their functional categories are reported in Annex.

In the first cluster of expression, where genes involved in processes during the first part of the development are expressed, the most abundant functional category was

metabolism. This reflects the high metabolic activity during this period. As sucrose is accumulating at a higher rate the presence of genes involved in the sucrose pathway was investigated, and details about metabolism-related ESTs will be discussed in the next section 6.4.

In this first part of the development a weak over-representation was also found for the functional category "cell growth, cell division and DNA synthesis". The presence of high amounts of tubulins and actins as well as of a cyclin among the members of this class indicated that many cells were probably undergoing cell divisions at this time. A high rate of cell division is in good agreement with the observation that the cambium rings are laid down early in the development. The same functional category was found even more significantly over-represented in the second part of the development. However, the functional annotations retrieved for those clones did not allow a clear interpretation. Interestingly, only for two more functional classes over-representation in both first and second part of the development was found. In the following, the class "cellular biogenesis" is discussed. Among the nine members of this functional category in cluster 1, four analogues of expansin and two analogues of putative prolin-rich proteins were identified. All proteins are involved in cell wall modification and expansion. As already mentioned in paragraph 5.3.5, one of these expansins, the *expansin 2* from *Zinnia elegans,* was reported by Im *et al*. (2000) as expressed during cell elongation in differentiating xylem cells. Parallels between genes preferentially expressed in sugar beet root and genes expressed during vascular tissue development were already mentioned when discussing candidates identified as preferentially expressed in sugar beet roots. This second macroarray experiment confirmed those findings and refined the maximum expression to the early stages of beet development. Concerning the clones belonging to the functional category "biogenesis" in the second part of the development, sequence similarities were not very obvious. A particular member of cluster 2 is a clone with similarity to the gibberellin 20-oxidase like protein. This gene involved in the GA biosynthesis and regulation was correlated as well to expansion (Vidal *et al*., 2003). Additionally, in a recent experiment Israelsson *et al*. (2003) investigated the gene expression of a transgenic GA 20-oxidase over-expressing hybrid aspen by microarray. They mainly observed an overexpression of genes involved in cell wall formation, extension and xylogenesis. The involvement of the sugar beet GA 20-oxidase-like gene in induction of genes related to cell wall

formation, extension and xylogenesis remains to be demonstrated although clones with similarity to xyloglucan glucanases are co-expressed during this part of the development and thus hint at such a possibility. In their comparative study about sugar beet and chard thickening, Rapoport and Loomis (1986) reported evidence for a longer duration of cell expansion in sugar beet than in chard. They hypothesized that expansion in sugar beet could be composed of two different phases. In chard, in fact, after 12 weeks no more expansion is observed, but in sugar beet expansion continues from week 12 onwards even at a higher rate than before. The clones, which are members of the functional category "biogenesis" in cluster 1 and in cluster 2, can be considered putative candidates playing a role in these two phases of expansion.

In the first part of the development, functional domains involved in transport processes also appear to be over-represented. Analogous gene products retrieved by similarity were mainly aquaporins and water channel proteins that facilitate the permeation of water across membranes (Yamada *et al*., 1995), either associated with the plasmalemma or with the tonoplast. There were also over-represented genes involved in ionic homoeostasis. These two observations coincide with reports that regulation of turgor pressure in sugar beet root is essential (Elliott and Weston, 1993). The high solute concentration increases the turgor pressure that is known to act as an inhibitor of the tonoplast proton-dependent ATPase, responsible for the production of the proton gradient involved in the active sucrose import in the vacuole (Fieuw and Willenbrink, 1990). For this reason, a fine regulation of the turgor pressure is crucial for the sucrose accumulation. Aquaporins and water channel proteins could be involved in this process and therefore they are putative candidate genes. However, other processes like cell wall relaxation are as well supposed to play a role in this regulation. In fact, the numerous ESTs annotated as gene products involved in cell wall expansion and modification could be part of the cell wall relaxation.

Another member of the functional class "transport" in cluster 1 was an EST similar to the gene product vacuolar-typeH^+-ATPase subunit D. The putative involvement of this gene product in the sucrose accumulation process will be considered in more detail in paragraph 6.4.2.

Concerning the clones preferentially expressed at maturity and their classification in functional categories, the most abundant functional category was "cellular communication and signal transduction". Among the members of this group there was

a large number of unknown genes, but also a PR-protein like osmotin was present. The presence of a conserved domain in a less conserved sequence is the reason for the assignment of a gene to a functional category in the absence of an annotation. For R-genes, for example, the presence of conserved domains is typical, but the overall similarity is usually low (Hunger *et al*., 2003). Other clones with homology to PR-proteins like glucanase and chitinase were retrieved among the clones belonging to cluster 2. The establishment of a stress response was already identified as a parallel between sugar beet roots and ripening fruits from the results of the first macroarray analysis and possible reasons have already been discussed in chapter 6.2.2.

The over-representation of the functional classes “cell growth, division and DNA synthesis”, “transcription” and “protein synthesis” in cluster 2 was found remarkable. In particular, ESTs with similarity to gene products in the class “protein synthesis” were highly represented during this period of the development. Considering that sugar beet is a biennial plant, it can be argued that the plant possibly adjusts its protein composition for the winter phase. Confirmation of this hypothesis or the discovery of other underlying processes accounting for this large number of ESTs involved in protein synthesis requires more experimental work. The concomitant over-representation of ESTs belonging to the functional class “energy” is explained by the costs of protein synthesis.

In conclusion, the time course analysis provides a detailed understanding of the different processes connected with the different phases of the sugar beet development and the gene products involved. Thus, putative candidate genes for beet yield and sucrose content are now available as targets for validation by further experimental approaches.

6.4 - IDENTIFICATION OF CANDIDATE GENES RELATED TO METABOLISM

Candidate genes obtained for the time course analysis were reconsidered with respect to their importance for sucrose metabolism in the sugar beet root. To this end, the sucrose concentration in the beet was analyzed during the development and related to the gene expression. Here it is one aim to identify candidates for the sucrose

accumulation based on the correlation of their expression profiles to the sucrose accumulation process.

6.4.1 – Sucrose accumulation during the development

At each of the harvesting time-points, the sucrose concentration was estimated. It was measured as sucrose concentration per fresh weight, and the results are reported in Table 5.2. These data are in agreement with the current view (see 6.3.1) that a separate ripening phase for sucrose accumulation in sugar beet root does not exist. Milford (1973) reported that already six weeks after emergence a relative high concentration of sucrose amounting to 9% per fresh weight was found. In the study reported here, 10.17% of sucrose per fresh weight was measured nine weeks after sowing in the year 2001, and the concentration of sucrose per fresh weight was 7.54% at the first time point in 2002. Sucrose concentration in leaves was also measured and accounted for less than 2% of fresh weight (Schneider, pers.com.). From these observations it can be concluded that accumulation of sucrose is a process starting early in the development, and that from the beginning the cells of the root must get adapted to high osmotic stress levels. From week nine onwards the amount of sucrose in the root grew linearly with time as already reported by Milford (1973). However, if the sucrose concentration is considered (Figure 5.4) it can be seen that in the first part of the development the sucrose concentration is increasing linearly, but in the second part of the development saturation is reached. This may be due to the negative correlation between beet yield and sucrose concentration.

6.4.2 – Candidate genes for the sucrose accumulation

The sucrose concentration is increasing at a higher rate in the first part of the development. Therefore the macroarray time-course analysis can be employed to retrieve candidate genes showing a pattern of expression correlating to the sucrose accumulation. Applying the clustering strategy explained in paragraph 5.3.4, a group of 599 clones was preferentially expressed in the first part of the development in both

years. This expression pattern reveals correlation with the sucrose accumulation process.

Among this group, 81 ESTs with a domain involved in metabolism were identified. Thirty-three genes were involved in the carbohydrate metabolism. This represents 40.1% of the clones for which a functional domain related to metabolism was identified. This subgroup is over-represented in this cluster when compared to the proportion of carbohydrate metabolism-related genes in the entire A024 EST set. Therefore it is supposed that processes related to the carbohydrate metabolism are active in this period of the development. The majority of the corresponding gene products are involved in cell wall biosynthesis and modification like xyloglucan endo-1,4-beta-D-glucanase, pectinesterase-like protein, pectate lyase, putative glucosyltransferase and a polygalacturonase. According to the analysis of the functional domains, only one gene product with a role in the sugar metabolism was identified encoding the fructokinase of sugar beet (Chaubron *et al*., 1995). The low number of genes involved in the sucrose biosynthesis pathway confirms results reported for sugar cane. Casu *et al*. (2003) showed that among 7242 ESTs obtained from sucrose accumulating maturing stem of sugar cane, 2.4% sequences encoded genes with potential roles in the carbohydrate metabolism. For comparison, the authors produced a second EST library of 1082 ESTs from immature stem tissue of sugar cane. In this case a fraction of 2.1% of the ESTs encoded genes related to the carbohydrate metabolism. From this result they conclude that the metabolism associated with sucrose accumulation in the maturing stem is not primarily regulated at the transcriptional level.

There are many reports on postranscriptional regulation of the enzymes involved in the sucrose pathway. According to the model for the sucrose accumulation in sugar beet roots (Fieuw and Willenbrink, 1989), the main enzymes involved in this pathway are the sucrose synthase and the sucrose phosphate synthase (SPS).

This last enzyme, responsible for the synthesis of sucrose-6-phosphate from fructose-6-phosphate and UDP-Glucose, has been shown to be highly regulated in plants. For spinach SPS, three different phosphorylation sites have been identified, involved in different regulatory mechanisms (Toroser and Huber 1997, Toroser *et al*. 1998, Toroser *et al*. 1999). Additionally, SPS is allosterically activated by glucose 6-phosphate and inhibited by inorganic phosphate. In sugar beet, a sucrose phosphate

synthase was cloned (Hesse *et al*., 1995). A phosphorylation site corresponding to the main phosphorylation site in the spinach SPS was also identified for the gene from sugar beet. The authors describe a preferential expression in the sugar beet tap root, and as a conclusion from screening several libraries to identify further sugar beet SPS, they suggest the presence of only one SPS in sugar beet. A second sugar beet sequence named SPS2 with an amino acid identity of 74% compared to the known SPS, was found in the database EMBL. In the A024 EST collection, three different clones annotated as sucrose phosphate synthase were identified. For two of them the deduced amino acid sequence showed similarity to the SPS2 of sugar beet with an expected value of e^{-98}. The third revealed similarity to the sugar beet sucrose phosphate synthase SPS with an expected value of e^{-40} and no similarity to the other EST sequences present in the library at nucleotide level. The transcription patterns of these three clones did not show significant differential expression during the development and consequently no correlation with the sucrose accumulation process was noticed. It is suspected that other genes regulating the activity of SPS after transcription influence the activity of this enzyme.

Concerning sucrose synthase (SS), the other enzyme detected in sugar beet roots (Fieuw and Willenbrink, 1987), mechanisms of posttranscriptional regulation were reported as well. This enzyme catalyses the reversible conversion of sucrose and UDP to UDP-glucose and fructose. Two phosphorylation sites were identified in maize sucrose synthase, one involved in the activation of the cleavage reaction (Huber *et al*., 1996), and the second putatively involved in a mechanism of targeting the enzyme to proteasome-mediated degradation (Hardin *et al*., 2003). In sugar beet, the expression pattern of a sucrose synthase SBSS cDNA clone was first investigated by Hesse and Willmitzer (1996). They reported a high expression in sugar beet roots that was confirmed by the clone BQ490130 showing similarity at amino acid level to the SBSS with an expected value of zero (chapter 4). The preferential expression in roots of this gene, together with the preferential expression of a sucrose phosphate synthase in roots (Hesse *et al*., 1995) was considered a confirmation of the model for sucrose accumulation in sugar beet roots proposed by Fieuw and Willenbrink (1990). In the results presented in chapter 4 it was mentioned that the gene SBSS was mapped to chromosome VII of sugar beet (Schneider *et al*., 1999), and the preferential expression in beet was confirmed. In chapter 4 the identification of a second clone

BQ490013 is described which encodes a second sucrose synthase for sugar beet. It shows as well preferential expression in roots, but it is also expressed in inflorescences. This gene maps to a different genetic locus on chromosome VIII. This finding is in agreement with biochemical data reported by Klotz *et al*. (2003). By isoelectrofocusing these authors identified two sucrose synthase isoforms in sugar beet root (SusyI and SusyII). From their data, it was not possible to deduce whether the two isoforms were products of different genes or arose from different processing of a single gene product. Additionally, they observed differences in the activities of the two isoforms in response to pH conditions. In a previous study (Klotz *et al*., 2002) the expression of these two enzymes during the development of the root was investigated. The authors reported a differential expression pattern for the two isoforms: SusyI was found present throughout the first 16 weeks of development analysed, although a decrease in the expression was found in week 12 and 16, while SusyII was only expressed when plants were 16 weeks old.

The expression patterns of both the SS clones, BQ490130 and BQ4090013, were followed during the sugar beet root development in the time-course analysis (chapter 5). For the first clone, homologous to SBSS, no significant differential expression during the development was identified. In contrast to this, the second clone revealed differential expression, and the expression pattern was confirmed in both years. It belongs to cluster 1, for which a correlation to the sucrose accumulation process was shown. A correlation between the transcription pattern of this clone and the pattern of SusyI enzyme activity can be deduced if the sixteen weeks considered in the study performed by Klotz *et al*. (2002) are regarded as the first half of the developmental analysis in the time-course experiment presented here.

The model proposed by Fieuw and Willenbrink also involves a cell wall associated invertase and a vacuolar-H^+-ATPase to maintain the proton gradient responsible for the active import of sucrose into the vacuole.

No invertase with preferential expression in the first part of the development was identified in the time course analysis. However, a vacuolar H^+-ATPase subunit D was retrieved. In sugar beet, subunit A and C of this protein were already known, but subunit D was not previously identified. The preferential expression of this clone in the first part of the development (see Annex) indicates correlation with the sucrose accumulation process.

In conclusion, the time course analysis allowed the identification of putative candidate genes for sucrose accumulation. Assumptions based on the homology and the putative functional role of the respective gene products led to the identification of three putative candidate genes for the sucrose accumulation process. For other candidates without obvious sequence similarities, functions and activities remain to be elucidated by bioinformatic approaches like supervised hierarchical clustering or at the genetic level by mapping and QTL analysis.

6.5 - VALIDATION OF CANDIDATE GENES BY MAPPING

The importance of any of the identified ESTs for beet formation and sucrose accumulation as well as for related traits can only be inferred from genetic and biochemical studies. According to the candidate gene approach (reviewed in Pflieger *et al*., 2001) a gene becomes a candidate for a trait if its map position coincides with a significant QTL region. Genomic regions with effects on traits like sucrose content (SC) and beet yield (BY), have already been identified by analyses of quantitative trait loci in a genetic map including expressed sequences related to carbohydrate metabolism (Schneider *et al*., 2002). Additionally, genomic regions with effects on the complex trait corrected sugar yield (CSY), depending on beet yield and sucrose content, were identified. As a first step, genetic mapping of the potential candidate genes identified in the macroarray study provided information on chromosomal localizations for 35 clones. The map positions of selected clones and their location with respect to QTL regions are described and discussed in the following. It does not substitute for a complete QTL analysis comparing phenotypic and genotypic data of each plant of the population, but this interpretation will provide first clues as to the value of certain candidates.

The clones BQ488374 encoding a putative phosphoenolpyruvate carboxylase kinase, and the clones BQ488519, BQ654411, BQ489719, BQ654410, BQ489820 and BQ490562 mapping on chromosome IV or IX and for which no annotation was retrieved, appear to be promising candidates for CSY (corrected sugar yield). The last three were also candidates for a QTL for sucrose content, SC5, because of the partial overlap between this QTL and the QTL CSY3 on chromosome IX. An additional

clone, BQ490303, becomes candidate for the QTL SC1. With regard to the trait beet yield, for which a weak positive correlation to the sugar content trait was reported (Schneider *et al*., 2002) the clones BQ488374 (the already mentioned putative phosphoenolpyruvate carboxylase kinase), BQ488519, BQ654411 and BQ489719, mapping all to the same region on chromosome IV are proposed as candidates. With the mentioned exception, neither of these clones showed homology to any gene product in the database. As sucrose accumulation is not present in model systems without a root storage organ for sucrose, genes involved in this process are likely to be uncharacterized or even specific for sugar beet.

Mapping data are available for ten of the candidates genes from the library A024 belonging to cluster 1 or 2 of the time course analysis (Möhring, pers. com.). In the following, their map positions are correlated to QTL regions. The clone E-15,27 maps to the region of the QTL SC4 on chromosome VII. The deduced amino acid sequence of the EST showed homology to a pectate lyase. The clones E-9,9 encoding a putative carboxyphosphonoenolpyruvate mutase and the clone D-16,2 encoding a P-Protein - like protein can be proposed as candidates for the QTLs SC1 and SC2, respectively. For further validation of these candidates, the organization of expressed genes in conserved haplotype structures (Schneider *et al*., 2001) allows the application of association studies. It is then possible to identify alleles with positive effects on the trait of interest and to test them in association studies. The power of this approach has already been demonstrated (Prioul *et al*., 1999, Fridman *et al*., 2000).

7 – SUMMARY

7.1 - Summary

The presented work integrates molecular data on gene expression with anatomical and biochemical data to analyze the development and the sucrose accumulation process in sugar beet (*Beta vulgaris* L.) roots. Sugar beet is a biennial Chenopodiacean plant, and it is the major crop for sucrose production in temperate regions. A special root morphology and physiology allow the accumulation of sucrose up to 20% of the fresh weight of the mature root. Approaches to study this storage process at the molecular level have so far been limited to known genes involved in pathways related to sugar metabolism which were mapped and tested for their association with QTLs for sugar yield and quality (Schneider *et al*., 1999, 2002). In the study presented here, transcription levels in sugar beet roots were analyzed to select candidate genes for the sucrose accumulation process.

For this purpose macroarrays were generated from two cDNA collections. The first experiment was performed with 3840 redundant sugar beet cDNAs. A procedure for the analysis including control steps was developed. The performance of the macroarrays was evaluated and compared to commercially produced nylon filters. Both systems could detect transcripts present in as little as 10 copies per cell in agreement with reports by Desprez *et al*. (1998). Their capacity to analyse transcripts of low abundance was demonstrated in a case study using resistance gene analogues (RGAs). Within an interval of two-fold variation in signal intensities, reproducibility between spots on the same filter was determined to be 98.9%, between spots on different filters 89.8%, and reproducibility after hybridization with two probes synthesized from the same poly(A)$^+$RNA sample was 97.6%. Hybridizations with probes synthesized from different field grown samples of the same organ showed reproducibility for 69.7% of the spots on average. Some precautions were introduced to reduce the sampling effects caused by the variability of environmental conditions.

Expression profiles from roots, leaves and inflorescences were generated for 2048 unique cDNAs of the first cDNA clone set. Expression values for each organ were

determined by stringent statistical analysis based on eight replica for each clone. Differential expression among the three organs was shown for 917 unique cDNAs, and for 76 unique cDNAs, the amount of detected transcript in roots was at least twice as high as in other organs. For 40 of them a map position was identified and linkage to QTLs is discussed. Additionally, possible functions of preferentially root-expressed candidate genes in taproot morphology and physiology are proposed. As a technical validation, macroarray expression data were confirmed by Northern blot analysis and quantitative RT-PCR experiments.

The second set of macroarray experiments was performed with 11520 unique cDNA clones to identify candidate genes in sugar beet roots related to sucrose accumulation or development. For this purpose, a time-course experiment was repeated in two different years. Plants were characterized morphologically and metabolically with respect to their sucrose content during the development. Among the genes differentially expressed in the development, 599 clones with highest expression in the early stages of the first vegetation period were identified in both years. For additional 175 clones, a reproducible preferential expression in the last stages of the development was demonstrated. These candidate genes were classified with respect to their function, and their putative role during development and sucrose accumulation is discussed. Additionally, strategies to focus on the validation of candidates related to sucrose accumulation are discussed.

In conclusion, the macroarray technology as established here, together with the selection and characterization of appropriate physiological samples, proved to be a valuable tool to identify new candidate genes related to development and to the sucrose accumulation in the sugar beet root. This is of special importance to sugar beet research because the considered processes cannot be analyzed in model systems without a root storage organ for sucrose.

7.2 – Zusammenfassung

In der vorgelegten Arbeit werden molekulare Daten zur Genexpression mit anatomischen und biochemischen Daten integriert, um die Entwicklung und den Prozeß der Saccharoseakkumulation in der Wurzel der Zuckerrübe (*Beta vulgaris* L.) zu analysieren. Die Zuckerrübe ist eine zweijährige Pflanze, die zur Familie der Chenopodiaceen gehört. Sie ist eine wichtige Nutzpflanze zur Produktion von Saccharose in der gemäßigten Klimazone. Ihre spezielle Wurzelmorphologie und –physiologie erlauben die Akkumulation von bis zu 20% Saccharose bezogen auf das Frischgewicht der reifen Wurzel. Ansätze, diesen Einlagerungsprozeß auf molekularer Ebene zu analysieren, waren bis jetzt auf bekannte Gene des Zuckerstoffwechsels beschränkt. Solche Gene wurden bereits genetisch kartiert und auf Assoziation mit QTLs für Zuckergehalt und –qualität hin überprüft (Schneider *et al*., 1999, 2002). In der hier vorgelegten Arbeit wurden Transkriptionsprofile in den Wurzeln der Zuckerrüben untersucht, um Kandidatengene für die Saccharoseakkumulation auszuwählen.

Zu diesem Zweck wurden Makroarrays von zwei cDNA Kollektionen hergestellt. Das erste Experiment wurde mit 3840 redundanten Zuckerrüben cDNAs durchgeführt. Ein Protokoll für die Analyse einschließlich relevanter Kontrollen wurde erstellt. Die Qualität der Makroarrays wurde evaluiert und mit kommerziell produzierten Nylonfiltern verglichen. In beiden Systemen konnten Transkripte, die nur in 10 Kopien pro Zelle vorkamen, nachgewiesen werden. Dieses Ergebnis stimmt mit veröffentlichten Resultaten von Desprez *et al*. (1998) überein. Die Sensitivität bei der Analyse schwach exprimierter Transkripte wurde für Resistenzgenanaloga (RGAs) erfolgreich getestet. Innerhalb eines Intervalls von einer Abweichung der Signalintensitäten um den Faktor zwei betrug die Reproduzierbarkeit zwischen Signalen desselben Filters 98.9%, die Reproduzierbarkeit von Signalen verschiedener Filter 89.8%, und die Reproduzierbarkeit von Signalen nach der Hybridisierung mit zwei Proben, die von derselben poly(A)$^+$RNA synthetisiert wurden, lag bei 97.6%. Hybridisierungen, die mit Proben von verschiedenen im Feld angezogenen Pflanzen durchgeführt wurden, zeigten eine Reproduzierbarkeit von 69.7% im Durchschnitt. Es

wurden Parameter eingeführt, um Effekte, die durch Umweltvariabilität hervorgerufen werden, zu reduzieren.

Für 2048 nicht-redundante cDNA Klone des ersten cDNA Klonsets wurden Expressionsprofile von Wurzeln, Blättern und Infloreszenzen generiert. Expressionsdaten für jedes Organ wurden einer stringenten statistischen Analyse unterworfen, die auf acht Wiederholungen für jeden Klon basiert. Für 917 nicht-redundante cDNAs wurde differentielle Expression zwischen den drei Organen nachgewiesen. Davon zeigten 76 cDNAs eine wenigstens zweifach erhöhte Expressionsstärke in Wurzeln im Vergleich zu den anderen Organen. Für 40 dieser cDNAs wurde der korrespondierende Genort kartiert, und die Korrelation mit QTL Positionen wird diskutiert. Zusätzlich werden mögliche Funktionen für die präferentiell in der Wurzel exprimierten Transkripte vorgeschlagen. Zur technischen Überprüfung wurden die Makroarray Daten einiger Klone durch Northern Analyse und quantitative RT-PCR bestätigt.

Die zweite Serie von Makroarrayexperimenten wurde mit 11520 nicht-redundanten cDNA Klonen durchgeführt, um Kandidatengene, die bei der Akkumulation der Saccharose und der Rübenentwicklung eine Rolle spielen, zu identifizieren. Zu diesem Zweck wurde eine Kinetik zur Rübenentwicklung in zwei verschiedenen Jahren wiederholt. Die Pflanzen wurden morphologisch und in Bezug auf ihren Saccharosegehalt hin während der Entwicklung untersucht. Unter den differentiell exprimierten cDNAs waren 599, die die höchste Expression in den frühen Stadien der ersten Vegetationsperiode in beiden Jahren zeigten. Für weitere 175 Klone wurde eine reproduzierbare präferentielle Expression im Reifestadium gefunden. Die entsprechenden Kandidategene wurden im Hinblick auf ihre Funktion klassifiziert, und ihre mögliche Rolle während der Entwicklung und der Saccharoseakkumulation wird diskutiert. Zusätzlich werden Strategien zur Validierung von Kandidatengenen vorgestellt.

Zusammenfassend läßt sich sagen, daß die hier etablierte Makroarraytechnologie zusammen mit der Auswahl und Charakterisierung von physiologisch relevanten Proben ein wertvolles System sind, um neue Kandidatengene für die Rübenentwicklung und Saccharoseakkumulation in der Rübenwurzel zu identifizieren. Dies ist von besonderer Bedeutung für die Forschung an Zuckerrüben,

weil diese Prozesse nicht in Modellpflanzen ohne ein Wurzelspeicherorgan für die Saccharose analysiert werden können.

8 - REFERENCES

Aharoni, A., Keizer, LC.P., Van den Broeck, H.C., Blanco-Portales R., Munoz-Blanco, J., Bois, G., Smit, P., De Vos, R.C.H., O'Connell, A.P. (2002) Novel insight into vascular, stress, and auxin-dependent and –Independent gene expression programs in strawberry, a non-climacteric fruit. Plant Phys. 129, 1019-1031.

Aharoni, A., Vorst, O. (2002) DNA microarrays for functional plant genomics. Plant Mol. Biol. 48, 99-118.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Amor Y, Haigler CH, Johnson S, Wainscott M, Delmer DP. (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. Proc Natl Acad Sci U S A. Sep 26; 92(20):9353-7.

Artschwager E, (1926): Anatomy of the vegetative organs of the sugar beet. J. Agr. Res. 33, 143-176.

Arumuganathan and Earle, (1991) Nuclear DNA content of some important plant. Plant Mol Biol Rep 9: 208-218.

Bachem CW, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG. (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J. May;9(5):745-53

Bartels, D., Altosaar, I., Harberd, N.P., Barker, R.F., Thompson, R.D. (1986) Molecular analysis of gliadin gene families at the complex Gli-1 locus of bread wheat (*T. aestivum* L.). Theor. Appl. Genet. 72, 845 – 853.

Bartels, D., and Thompson, R.D. (1983) The characterization of cDNA clones coding for wheat storage proteins. Nucleic Acids Res. 11, 2961-2978.

Barzen E, Mechelke W, Ritter E, Schulte-Kappert E, Salamini F (1995)An expected map of the sugar beet genome containing RLFP RAPD loci. Theor Appl Genet 90:189-193.

Beissbarth T, Fellenberg K, Brors B, Arribas-Prat R, Boer J, Hauser NC, Scheideler M, Hoheisel JD, Schutz G, Poustka A, Vingron M.(2000): Processing and quality control of DNA array hybridization data. Bioinformatics. Nov;16(11):1014-22.

Bellin , M. Werber , T. Theis , B. Schulz , B. Weisshaar , K. Schneider EST Sequencing, Annotation and Macroarray Transcriptome Analysis Identify Preferentially Root-Expressed Genes in Sugar Beet (2002) Plant biol;6; 700-710.

Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A.M., Jordan, B.R., and Nguyen, C. (1996) Multiplex messenger assay: simultaneous, quantitative measurement of expression of many genes in the context of T-cell activation. Nucleic Acids Res. 24, 1435-1442.

Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR. (1999) Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. Hum Mol Genet. Sep;8(9):1715-22. Erratum in: Hum Mol Genet 1999 Oct;8(11):2129.

Bishop J O, Morton J G, Rosbash M and Richardson M (1974) Three abundance classes in HeLa cell messenger RNA. Nature 250, 199-203.

Bocca SN, Kissen R, Rojas-Beltran JA, Noel F, Gebhardt C, Moreno S, du Jardin P, Tandecarz JS. (1999) Molecular cloning and characterization of the enzyme UDP-glucose: protein transglucosylase from potato This paper is specially dedicated to the memory of Dr Juana S. Tandecarz, deceased on December 10, 1996. Plant Physiol Biochem. Nov;37(11):809-819.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. Dec;29(4):365-71.

Burke, J., Davison, D., Hide, W. (1999) D2_cluster: A Validated Method for Clustering EST and Full-length cDNA Sequences. Genome Res. 9, 1135-1142.

Cambien F, Poirier O, Lecerf L, Evans A, Cambou JP, Arveiler D, Luc G, Bard JM, Bara L, Ricard S, *et al*. (1992) Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. Nature. Oct 15;359(6396):641-4.

Campbell, P., Braam, J. (1998) *In vitro* activities of four xyloglucan endotransglycosylases from *Arabidopsis*. The Plant Journal 18, 371-382.

Casu RE, Grof CP, Rae AL, McIntyre CL, Dimmock CM, Manners JM. (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol. May;52(2):371-86.

Causse M, Rocher J P, Henry A M, Charcosset A, Prioul J L, de Vienne D (1995) Genetic dissection of the relationship between carbon metabolism and early growth in maize, with emphasis of key-enzyme loci. Mol Breed 1: 259-272.

Chaubron, F, Taylor, M.A., Ross, H.A., Ledoigt, G.,Davies, H.V. (1995): Isolation and characterisation of a cDNA clone of fructokinase from sugar beet (Beta vulgaris L.). PIR database

Chiou T.J., Bush D.R., (1996) Molecular cloning, immunochemical localization to the vacuole, and expression in transgenic yeast and tobacco of a putative sugar transporter from sugar beet. Plant Physiol. Feb;110(2):511-20

Chiou TJ, Bush DR.(1998) Sucrose is a signal molecule in assimilate partitioning Proc Natl Acad Sci U S A. Apr 14;95(8):4784-8

Chourey P S, Miller M E (1995) On the role of sucrose synthase in cellulose and callose biosynthesis in plants. In: Pontis H G, Salerno G L, Echeverria E eds. Sucrose metabolism, biochemistry. physiology and molecular biology, Vol 14. Rockville, MD: American society of plant physiologists, 80-87.

Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V, Emmert-Buck MR. (2002) Post-analysis follow-up and validation of microarray experiments Nat Genet. Dec;32 Suppl:509-14.

Churchill GA. (2002) Fundamentals of experimental design for cDNA microarrays.Nat Genet Dec;32 Suppl:490-5.

Cooke D.A., and Scott R. K., (1993). In: The sugar beet crop: science into practice (Cooke, D.A., Scott, R.K. eds), London, Chapman & Hall

Davies, C., Robinson, S.P. (2000) Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins. Plant Phys. 122, 803-812.

DeRisi, J.L., Vishwanath, R.I., Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680-686.

Desprez, T., Amselem, J., Caboche, M., Hofte H. (1998) Differential gene expression in Arabidopsis monitored using cDNA arrays. Plant J.14, 643-652.

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JD. (1996) The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell. Feb 9;84(3):451-9.

Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W.(2002) Comprehensive gene expression analysis by transcript profiling. Plant Mol Biol. Jan;48(1-2):75-97

Durrant WE, Rowland O, Piedras P, Hammond-Kosack KE, Jones JD.(2000) cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles. Plant Cell. Jun;12(6):963-77

Eickhoff B, Korn B, Schick M, Poustka A, van der Bosch J.(1999): Normalization of array hybridization experiments in differential gene expression analysis. Nucleic Acids Res. Nov 15;27(22):e33.

Eisen MB, Spellman PT, Brown PO, Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns.Proc Natl Acad Sci U S A. Dec 8;95(25):14863-8.

Elliott M.C,. Hosford D.J.,Lenton JR, Milford GFJ, Pocock, TO, Smith JE, Lawrence DK and Firby DJ (1984) Hormonal control of storage root growth. In Growth Regulators inRoot Development (eds MB Jackson and Stead) British Plant Regulator Group Wantage, Monograph no.10, pp 25-35.

Elliott, M.C., Weston, G.D. (1993) Biology and physiology of the sugar-beet plant. In: The sugar beet crop: science into practice (Cooke, D.A., Scott, R.K. eds), London, Chapman & Hall, 37-66.

Etxeberria E, Gonzalez P., (2003) Evidence for a tonoplast-associated form of sucrose synthase and its potential involvement in sucrose mobilization from the vacuole. J Exp Bot. May;54(386):1407-14.

Ewing, B., Green, P. (1998) Basecalling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. 8, 186-194.

Farrar K, Evans IM, Topping JF, Souter MA, Nielsen JE, Lindsey K. (2003) EXORDIUM-a gene expressed in proliferating cells and with a role in meristem function, identified by promoter trapping in Arabidopsis. Plant J. Jan;33(1):61-73.

Fieuw S and Willenbrink J (1987) Sucrose synthase and sucrose phosphate synthase in sugar beet plants (*Beta vulgaris* L ssp. *Altissima*) J. Plant Physiol. 107.

Fieuw, S., Willenbrink, J. (1990) Sugar transport and sugar–metabolizing enzymes in sugar storage roots (*Beta vulgaris* ssp. *altissima*). J. Plant Physiol. 137, 216-223.

Fondy and Gaiger, (1982) Diurnal pattern of translocation and carbohydrate metabolism in source leaves of *Beta vulgaris.* Plant physiol, 70, 671-676

Freeman W M, Robertson D J, Brebner K, Lynch W J, Nader S H, Nader M A, Gioia L, Daunias J B *et al*.(2000) Functional genomics of cocaine in non human primates and rats, Society for Neuroscience, Abstract.

Fridman E, Pleban T, Zamir D. (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci U S A. 2000 Apr 25;97(9):4718-23.

Fujimoto S Y, Ohta M, Usui A, Shinshi H, and Ohme-Takagi M (2000) Arabidopsis ethylene-responsive element binding factors act as transcriptional activatiors or repressors of GCC box-mediated gene expression. Plant Cell 12, 393-404.

Fukuda, H. (1997) Tracheary element differentiation. The Plant Cell 9, 1147-1156.

Geldermann G (1975) Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theor Appl Genet 46: 319-330.

Gerlach W.L. and Bedrook, J.R. (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley Nucl.Ac.Res. 7(7): 1869

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J.(2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol. Dec;124(4):1570-81.

Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H. (1992) Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. Mamm Genome.;3(11):609-19.

Hardin SC, Tang GQ, Scholz A, Holtgraewe D, Winter H, Huber SC. (2003) Phosphorylation of sucrose synthase at serine 170: occurrence and possible role as a signal for proteolysis. Plant J. Sep;35(5):588-603.

Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA. (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science. Dec 15;290(5499):2110-3.

Herbers K, Sonnewald U. Molecular determinants of sink strength. (1998) Curr Opin Plant Biol. Jun;1(3):207-16.

Herwig R, Aanstad P, Clark M, Lehrach H.(2001): Statistical evaluation of differential expression on cDNA nylon arrays with replicated experiments. Nucleic Acids Res. Dec 1;29(23):E117.

Herwig R, Schulz B, Weisshaar B, Hennig S, Steinfath M, Drungowski M, Stahl D, Wruck W, Menze A, O'Brien J, Lehrach H, Radelof U. (2002) Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes. Plant J. Dec;32(5):845-57.

Hesse H, Willmitzer L. Expression analysis of a sucrose synthase gene from sugar beet (Beta vulgaris L.) (1996). Plant Mol Biol. Mar;30(5):863-72.

Hesse, H., Sonnewald, U., Willmitzer, L. (1995) Cloning and expression analysis of sucrose-phosphate synthase from sugar beet (*Beta vulgaris* L.). Mol. Gen. Genet. 247, 515-520.

Hochberg Y (1988) A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75, 800-802.

Hoheisel, J.D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., Nizetic, D., Francis, F., and Lehrach, H. (1993) High-resolution cosmid and P1 maps spanning the 14-Mbp genome of the fission yeast *Schizosaccharomyces pombe*. Cell, 73: 109-120.

Hornberg JJ, de Haas RR, Dekker H, Lankelma J.(2002): Analysis of multiple gene expression array experiments after repetitive hybridizations on nylon membranes. Biotechniques. Jul;33(1):108, 110, 112-3,

Huang S, Raman AS, Ream JE, Fujiwara H, Cerny RE, Brown SM. (1998) Overexpression of 20-oxidase confers a gibberellin-overproduction phenotype in Arabidopsis. Plant Physiol. Nov;118(3):773-81.

Huber SC (1986) Fructose, 2,6-bisphosphate as a regulatory metabolite in plants. Annual Review of plant Physiology, 37, 233-46.

Huber SC, Huber JL, Liao PC, Gage DA, McMichael RW Jr, Chourey PS, Hannah LC, Koch K. (1996) Phosphorylation of serine-15 of maize leaf sucrose synthase. Occurrence in vivo and possible regulatory significance. Plant Physiol. Oct;112(2):793-802.

Hunger S, Di Gaspero G, Mohring S, Bellin D, Schafer-Pregl R, Borchardt DC, Durel CE, Werber M, Weisshaar B, Salamini F, Schneider K. (2003) Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (Beta vulgaris L.). Genome. Feb;46(1):70-82.

Im KH, Cosgrove DJ, Jones AM.(2000) Subcellular localization of expansin mRNA in xylem cells. Plant Physiol. Jun;123(2):463-70.

Israelsson M, Eriksson ME, Hertzberg M, Aspeborg H, Nilsson P, Moritz T. (2003) Changes in gene expression in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth. Plant Mol Biol. Jul;52(4):893-903.

Jendrisak J, Young RA, Engel JD. (1987) Cloning cDNA into lambda gt10 and lambda gt11. Methods Enzymol. 152:359-71.

Jin H, Cheng X, Diatchenko L, Siebert PD, Huang CC. (1997) Differential screening of a subtracted cDNA library: a method to search for genes preferentially expressed in multiple tissues. Biotechniques. Dec;23(6):1084-6

Jones CS, Davies HV, Taylor MA.(2000) Profiling of changes in gene expression during raspberry (Rubus idaeus) fruit ripening by application of RNA fingerprinting techniques. Planta. Oct;211(5):708-14.

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD. (1994) Isolation of the tomato Cf-9 gene for resistance to Cladosporium fulvum by transposon tagging.Science. Nov 4;266(5186):789-93.

Kamoun, S., Lindqvist, H., Govers, F. (1997b) A novel class of elicitin-like genes from *Phytophthora infestans*. Mol. Plant Microbe Interact. 10, 1028-1030.

Kamoun, S., van West, P., de Jong, A.J., de Groot, K.E., Vleeshouwers, V.G., Govers F. (1997a) A gene encoding a protein elicitor of *Phytophthora infestans* is down-regulated during infection of potato. Mol. Plant Microbe Interact. 10, 13-20

Karlsson M, Johansson I, Bush M, McCann MC, Maurel C, Larsson C, Kjellbom P. (2001) An abundant TIP expressed in mature highly vacuolated cells.Plant J. Jan;21(1):83-90.

Keegstra, K., Raikhel, N. (2001) Plant Glycosyltransferases. Current Opinion in Plant Biology 4, 219-224.

Kloos, D.U., Ottmanns, H., Dock, C., Stahl, D., and Hehl, R. (2002) Isolation and molecular analysis of six taproot expressed genes from sugar beet. Journal of Experimental Botany 53, 1533-1534**.**

Klotz K and Finger F (2002) Contribution of invertase and sucrose synthase isoforms to sucrose catabolism in developing sugar beet roots. Journal of sugar beet research. Vol 39 No 1-2.

Klotz Karen L., Fernando L. Finger, Weilin L. Shelver (2003) Characterization of two sucrose synthase isoforms in sugarbeet root. Plant Physiology and Biochemistry 41 107–115

Lee PD, Sladek R, Greenwood CM, Hudson TJ.: Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. Genome Res. (2002) Feb;12(2):292-7.

Lee S.A., Gardner R., Lay-Yee M.,(1993) An apple gene highly expressed in fruit Plant Physiol. 103:1017

Lennon GG, Lehrach H. (1991):Hybridization analyses of arrayed cDNA libraries. Trends Genet. Oct;7(10):314-7.

Leung YF, Cavalieri D. (2003) Fundamentals of cDNA microarray data analysis. Trends Genet. Nov;19(11):649-59.

Liang P and Pardee A B (1992) Differential of eukaryotic messenger RNA by means of PCR. Science 257: 967-971.

Lunn JE, MacRae E. (2003) New complexities in the synthesis of sucrose. Curr Opin Plant Biol. Jun;6(3):208-14.

Milford,GFJ (1973), The growth and development of the storage root of sugar beet., Annals of Applied Biology, 75: 427-38

Milioni, D., Sado, P.E., Stacey, N., Domingo, D., Roberts, K., McCann, M. (2001) Differential expression of cell-wall-related genes during the formation of tracheary elements in the Zinnia mesophyll cell system. Plant Mol. Biol. 47, 221-238.

Miller, R.T., Christoffels, A.G., Gopalakrishnan, C., Burke, J., Ptitsyn, A.A., Broveak, T.R., Hide, W.A. (1999) A comprehensive approach to clustering of expressed human gene sequence: The Sequence Tag Alignment and Consensus Knowledgebase. Genome Res. 9, 1143-1155.

Mindrinos M, Katagiri F, Yu GL, Ausubel FM. (1994) The A. thaliana disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats.Cell. Sep 23; 78(6):1089-99.

Nadon R, Shoemaker J. (2002) Statistical issues with microarrays: processing and analysis. Trends Genet. 2002 May; 18(5):265-71.

Nguyen C, Rocha D, granjeaud S, Baldit M, Berbard K, Naquet P and Jordan B R (1995) Differential gene expression in the Murine Thymus assayed by quantitative hybridization of arrayed cDNA clones. Genomics 29, 207-216.

Pflieger (2001) The candidate gene approach in plant genetics: a review. Molecular Breeding 7(4): 275-291; Jan.

Pluthero F.G. (1993) Rapid purification of high-activity Taq DNA polymerase. Nucl Ac Res 21(20): 4850-4851

Prioul**,** S Pelleschi, M Sene, C Theevenot, M Causse, D de Vienne, and A Leonardi (1999) From QTLs for enzyme activity to candidate genes in maize. J. Exp. Bot. 1999 50: 1281-1288.

Quackenbush J. (2002) Microarray data normalization and transformation. Nat Genet. Dec;32 Suppl:496-501.

Rafalski A. (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant BiolApr; 5(2):94-100.

Ransom-Hodgkins WD, Vaughn MW, Bush DR. (2003) Protein phosphorylation plays a key role in sucrose-mediated transcriptional regulation of a phloem-specific proton-sucrose symporter. Planta. Jul;217(3):483-9.

Rapoport H F and Loomis R S (1986) Structural aspects of root thickening in *Beta vulgaris* L.: Comparative thickening in sugarbeet and chard. Bot Gaz. 147(3): 270-277.

Rose J K, Braam J, Fry S C, Nishitani K (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. Plant Cell Physiol. 2002 Dec;43(12):1421-35.

Rothschild M F and SollerM (1997) Candidate gene analysis to detect genes controlling traits of economic importance in domestic live stock. Probe 8: 13-20.

Ruan, Y., Gilmore, J., Conner, T. (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. The Plant Journal 15, 821-833.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) "Molecular Cloning",Cold Spring Harbor Laboratory Press, 2nd edition , USA

Sanguinetti CJ, Dias Neto E, Simpson AJ. (1994) Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. Biotechniques. Nov; 17(5):914-21.

Sano T, Kuraya Y, Amino S, Nagata T. (1999) Phosphate as a limiting factor for the cell division of tobacco BY-2 cells. Plant Cell Physiol. Jan;40(1):1-8.

Sargent TD. (1987) Isolation of differentially expressed genes.Methods Enzymol. 1987;152:423-32.

Schäfer-Pregl R, Borchardt D C, Barzen E, Glass C, Mechelke W, Seitzer J F, Salamini F (1999) Localization of QTLs for tolerance to *Cercospora beticola* on sugar beet linkage groups. Theor. Appl Genet 99: 829-836.

Schena M, Shalon D, Davis RW, Brown PO. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. Oct 20;270(5235):467-70.

Schlaman, H.R., Risseeuw, E., Franke-van Dijk, M.E., Hooykaas, P.J. (1994) Nucleotide sequence corrections of the *uidA* open reading frame encoding beta-glucuronidase. Gene 138, 259-260.

Schneider, K., Borchardt, D.C., Schäfer-Pregl, R., Nagl, N., Glass, C., Jeppsson, A., Gebhardt, C., and Salamini, F. (1999) PCR-based cloning and segregation analysis of functional gene homologues in *Beta vulgaris*. Mol. Gen. Genet. 262, 515-524.

Schneider, K., Schäfer-Pregl, R., Borchardt, D.C., and Salamini, F. (2002) Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor. Appl. Genet. 104, 1107-1113.

Schneider, K., Weisshaar, B., Borchardt, D.C., and Salamini, F. (2001). SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Mol. Breed. 8, 63-74 .

Schumacher K, Schondelmaier J, Barzen E, Steinrücken G, Borchardt D, Weber W E, Jung C, Salamini F (1997) Conbining different linkage maps in sugar beet (*Beta vulgaris* L) to make one map. Plant Breed 116: 23-38

Scott and Jaggard, (1993) Biology and physiology of the sugar-beet plant. In: The sugar beet crop: science into practice (Cooke, D.A., Scott, R.K. eds), London, Chapman & Hall,

Scott, R.K. and Jaggard, K.W. (1993) Crop physiology and agronomy. In: The sugar beet crop: science into practice (Cooke, D.A. and Scott, R.K. eds), London, Chapman & Hall, 179-237.

Sherlock G. (2000) Analysis of large-scale gene expression data. Curr Opin Immunol. Apr;12(2):201-5.

Shi L, Gast RT, Gopalraj M, Olszewski NE. (1992) Characterization of a shoot-specific, GA3- and ABA-regulated gene from tomato. Plant J. Mar;2(2):153-9.

Smith, G., (1987). Sugar beet: Principles of cultivar development. Fehr, W.R. (ed.) MacMillan Publishing Company, pp. 577-625

Smith-Espinoza C.J., (2001) Analysis of ABA and Drought Stress Mediated Gene Expression in the desiccation Tolerant Resurrection Plant Craterostigma plantagineum, PhD Thesis

Staden, R., Beal, K.F., Bonfield, J.K. (1998) The Staden Package. Methods Mol. Biol. 2000, 132:115-130.

Stitt and Quick, (1989) Photosynthetic carbon partitioning: its regulation and possibilities for manipulation. Physiologia Plantarum, 77, 633-41.

Sturn A, Quackenbush J, Trajanoski Z. (2002) Genesis: cluster analysis of microarray data.Bioinformatics. Jan;18(1):207-8.

Sung S., Xu D. Black C., (1989) Identification of actively filling sucrose sinks Plant Physiol 891117-1121

Taniguchi Y, Morita I, Kubota T, Murota S, Aso T.(2001) Human uterine myometrial smooth muscle cell proliferation and vascular endothelial growth-factor production in response to platelet-derived growth factor. J Endocrinol. Apr;169(1):79-86.

Tesnière, C., Verriès, C. (2000) Molecular cloning and expression of cDNAs encoding alcohol dehydrogenases from *Vitis vinifera* L. during berry development. Plant Science 157, 77-88.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796-815.

The Gene Ontology (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25-29.

Thronley J H M and Jhonson I A (1990) Plant and crop modeling: A mathematical approach to plant and crop physiology. Oxford Science Publications.

Toroser D, Athwal GS, Huber SC. (1998) Site-specific regulatory interaction between spinach leaf sucrose-phosphate synthase and 14-3-3 proteins. FEBS Lett. Sep 11;435(1):110-4.

Toroser D, Huber SC. (1997) Protein phosphorylation as a mechanism for osmotic-stress activation of sucrose-phosphate synthase in spinach leaves. Plant Physiol. Jul;114(3):947-55

Toroser D, McMichael R Jr, Krause KP, Kurreck J, Sonnewald U, Stitt M, Huber SC. (1999) Site-directed mutagenesis of serine 158 demonstrates its role in spinach leaf sucrose-phosphate synthase modulation. Plant J. Feb;17(4):407-13.

Vaughn M.W., Harrington G.N., Bush D.R., Sucrose-mediated transcriptional regulation of sucrose symporter activity in the phloem. (2002) Proc Natl Acad Sci U S A. Aug 6;99 (16):10876-80

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. (1995) Serial analysis of gene expression. Science. Oct 20;270(5235):484-7

Vidal AM, Ben-Cheikh W, Talon M, Garcia-Martinez JL. (2003) Regulation of gibberellin 20-oxidase gene expression and gibberellin content in citrus by temperature and citrus exocortis viroid. Planta. Jul;217(3):442-8. Epub 2003 Feb 15.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, *et al*. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. Nov 11;23(21):4407-14

Weber W E, Borhardt D C, Koch G (1999) Combined linkage maps and QTLs in sugar beet (Beta vulgaris L) from different populations. Plant Breed 119, 193-204.

Weber W E, Borhardt D C, Koch G (2000) Marker analysis for quantitative traits in sugar beet. Plant Breed 119, 97-106.

Whetten, R., Sun, Y.H., Zhang, Y., Sederoff, R. (2001) Functional genomics and cell wall biosynthesis in loblolly pine. Plant Mol. Biol. 47, 275-291.

Wilkinson, J.Q., Lanahan, M.B., Conner, T.W., Klee, H.J. (1995) Identification of mRNAs with enhanced expression in ripening strawberry fruit using polymerase chain reaction differential display. Plant Mol. Biol. 27, 1097-1108.

Winter H, Huber JL, Huber SC. (1998) Identification of sucrose synthase as an actin-binding protein. FEBS Lett. Jul 3;430(3):205-8

Wurmbach E, Yuen T, Ebersole BJ, Sealfon SC. (2001) Gonadotropin-releasing hormone receptor-coupled gene network organization. J Biol Chem. Dec 14;276(50):47195-201. Epub 2001 Oct 01.

Xu W, Campbell P, Vargheese AK, Braam J. (1996) The Arabidopsis XET-related gene family: environmental and hormonal regulation of expression.Plant J. Jun;9(6):879-89.

Xu YL, Li L, Wu K, Peeters AJ, Gage DA, Zeevaart JA. (1995) The GA5 locus of Arabidopsis thaliana encodes a multifunctional gibberellin 20-oxidase: molecular cloning and functional expression. Proc Natl Acad Sci U S A. 1995 Jul 3;92(14):6640-4.

Yamada, S., Katsuhara, M., Kelly, W.B., Michalowski, C.B., Bohnert, H.J. (1995) A family of transcripts encoding water channel proteins: tissue-specific expression in the common ice plant. The Plant Cell 7, 1129-1142.

Yang Y H and Speed T (2002) Design issues for cDNA microarrys experiments. Nature reviews Vol 3 579-588.

Yu, J. *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296, 79-92.

Zdobnov, E.M. and Apweiler, R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847-848.

Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latch A, Pech JC, Bouzayen M. (1999) Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display Plant J. Jun;18(6):589-600.

Zhao N, Hashida H, Takahashi N, Misumu Y and Sakaki Y (1995) High-density DNA analysis: a novel approach for large scale, quantitative analysis of gene expression. Gene, 156, 207-213.

Zrenner R, Salanoubat M, Willmitzer L, Sonnewald U. (1995) Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum* L.). Plant J. Jan;7(1):97-107.

ERKLÄRUNG

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertig , die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen - , die anderen Werken in Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof F. Salamini betreut worden.

Teilpublikationen:

- Hunger S, Di Gaspero G, Mohring S, Bellin D, Schafer-Pregl R, Borchardt DC, Durel CE, Werber M, Weisshaar B, Salamini F, Schneider K. (2003) Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (Beta vulgaris L.). Genome. Feb;46(1):70-82
- Bellin , M. Werber , T. Theis , B. Schulz , B. Weisshaar , K. Schneider EST Sequencing, Annotation and Macroarray Transcriptome Analysis Identify Preferentially Root-Expressed Genes in Sugar Beet (2002) Plant biol;6; 700-710

Köln, den 15.12.2003

Diana Bellin

LEBENSLAUF

Name:	Diana Bellin
Geburtsdatum:	14.08.1973
Geburtsort:	Verona, Italien
Familienstand:	ledig
Staatsangehörigkeit:	italienisch
Adresse:	Carl-von-Linne`-Weg 10 50829 Köln 0221-5062485/404

Schulbildung:

1987 - 1992	Abitur in „Liceo scientifico statale Galileo Galilei“ Verona, Italien.

Berufliche Ausbildung:

1993 - 1998	Biotechnologiestudium an der Universität von Verona, Italien Diplomarbeit: “Effects of cold stress on light harvesting complexes”

Promotion:

2000 - 2003	Wissenschaftlicher Mitarbeiter am Max Planck Institut für Züchtungsforschung, Köln. Abteilung: Pflanzenzüchtung und Ertragsphysiologie (Prof. Salamini). Arbeitsgruppe: Genomanalyse der Zuckerrübe (H.D. Dr. K. Schneider)

ACKNOWLEDGEMENTS

This work was performed at the Max Planck Institute for Plant Breeding, in the department of Prof. Salamini.

I would like to acknowledge:

Prof. Salmini and Prof. Flügge for dedicating their time to this thesis work.

I would like then to thank my group leader H.D. Dr Katharina Schneider for all the help, support and for the things she teached me in these years, and I would like to thank Prof. Salamini for all his suggestions along these studies. Finally I would like to thank all the people that helped me in this period, and in particular:

Susanne Schwarz for her help in the lab, Dr. Martin Werber and Dr. Stephen Rudd for help in dealing with big amount of data and sequences, Prof. Bernd Weisshaar, Diana Lehman and Tania Theis from ADIS service unit at MPIZ for helping in dealing with robotic systems, KWS and Dr. Britta Schulz for support and plant material, and then Dr. Sirak Kifle, Dr Lars Valkhamp, Dr. Maria Piques, Dr. Claudia Smith Espinosa and all the macro/micro array community at the MPIZ for useful discussions.

I want also to thank all the beet girls: Antje, Silke, Sandra, Elena, Charlotte, Maria for the nice athmosphere in the lab and in particular for their kit for survivence for PhD students, it really helped me a lot in the last time…..

Again a particular thank to Silke and Ernesto for the time they dedicated to help me reading my thesis.

Thanks again to my grandmother, without her help I would have never mamaged, to my parents for their encouragement and to Ernesto to have been comprehensive and for his support.

ANNEX

List of genes belonging to cluster 1, for which a functional annotation was reported. The 255 genes are grouped according to the functional category to which they belong and ordered according decreasing induction

Coordinates	Ratio	Arabidopsis accession	Arabidopsis description	Arabidopsis e-value
			METABOLISM	
P - 15, 21	38.94	At4g30270	xyloglucan endo-1,4-beta-D-glucanase precursor	2.00E-73
O - 20, 23	33.61	At3g07270	GTP cyclohydrolase I	3.00E-62
E - 9, 12	25.33	At2g30020	putative protein phosphatase 2C	6.00E-63
O - 20, 8	22.44	At1g09070	unknown protein	1.00E-19
J - 19, 28	22.25	At3g60130	beta-glucosidase-like protein	6.00E-20
L - 9, 8	22.22	At4g25810	xyloglucan endo-1,4-beta-D-glucanase (XTR-6)	5.00E-78
C - 11, 21	20.94	At4g23500	putative protein	8.00E-76
O - 18, 10	19.35	At1g09070	unknown protein	1.00E-19
C - 10, 22	18.64	At3g61490	putative protein	4.00E-23
E - 3, 25	18.36	At1g14890	unknown protein	2.00E-18
B - 10, 21	14.80	At4g23500	putative protein	8.00E-76
L - 6, 12	13.83	At4g35000	L-ascorbate peroxidase	2.00E-62
O - 14, 8	13.08	At1g77060	putative carboxyphosphonoenolpyruvate mutase	9.00E-69
M - 17, 9	12.11	At3g07320	putative beta-1,3-glucanase precursor	1.00E-22
C - 11, 24	11.69	At4g07960	putative glucosyltransferase	2.00E-73
D - 5, 8	10.94	At2g16500	arginine decarboxylase	2.00E-33
P - 16, 26	10.57	At4g32940	gamma-VPE (vacuolar processing enzyme)	5.00E-75
C - 3, 25	9.97	At1g14890	unknown protein	2.00E-18
E - 9, 9	9.93	At1g77060	putative carboxyphosphonoenolpyruvate mutase	4.00E-34
L - 6, 11	9.50	At1g10360	putative glutathione S-transferase TSI-1	2.00E-56
K - 22, 28	9.23	At1g13440	putative protein	e-156
E - 4, 27	9.06	At5g24090	acidic endochitinase	4.00E-42
D - 2, 24	7.80	At4g02340		7.00E-42
H - 22, 24	7.54	At5g67460	putative protein	2.00E-13
K - 23, 18	7.15	At5g05340	peroxidase	1.00E-59
H - 19, 4	7.11	At1g77060	putative carboxyphosphonoenolpyruvate mutase	9.00E-69
K - 20, 23	7.08	At1g68560	alpha-xylosidase precursor	3.00E-69
E - 4, 8	6.69	At1g77060	putative carboxyphosphonoenolpyruvate mutase	4.00E-34
M - 17, 6	6.58	At2g07050	cycloartenol synthase	1.00E-77
F - 12, 25	6.49	At5g66920	pectinesterase like protein	4.00E-58
E - 22, 8	6.46	At1g58440	unknown protein	1.00E-66
C - 12, 8	6.18	At3g23820	NAD dependent epimerase, putative	2.00E-73
D - 5, 24	5.72	At5g11160	adenine phosphoribosyltransferase - like protein	3.00E-53
A - 10, 20	5.67	At2g38700	mevalonate diphosphate decarboxylase	4.00E-41
L - 5, 8	5.57	At4g30440	nucleotide sugar epimerase-like protein	5.00E-14
G - 9, 26	5.49	At5g54060	flavonol 3-O-glucosyltransferase-like	3.00E-39
L - 14, 13	5.46	At1g05010	1-aminocyclopropane-1-carboxylate oxidase	5.00E-66
O - 13, 28	5.23	At4g13940	adenosylhomocysteinase	1.00E-66
H - 24, 12	5.21	At1g77060	putative carboxyphosphonoenolpyruvate mutase	2.00E-44
E - 5, 26	5.20	At5g55180	beta-1,3-glucanase-like protein	2.00E-45
N - 7, 20	5.13	At3g55430	beta-1,3-glucanase - like protein	2.00E-35
P - 18, 24	5.01	At4g33360	putative protein	4.00E-50
M - 20, 23	5.01	At1g41830	pectinesterase, putative	7.00E-39
L - 24, 9	4.82	At2g15480	putative glucosyltransferase	2.00E-34
J - 13, 3	4.57	At1g48100	polygalacturonase PG1, putative	2.00E-27
G - 14, 25	4.50	At2g34190	putative membrane transporter	2.00E-58
H - 2, 12	4.47	At1g75680	endo-beta-1,4-glucanase, putative	1.00E-36
A - 13, 14	4.27	At5g17920	5-methyltetrahydropteroyltriglutamate	8.00E-58

K - 20, 1	3.85	At1g78570	Similar to dTDP-D-glucose 4,6-dehydratase	e-109
J - 24, 9	3.56	At1g78950	hypothetical protein	5.00E-61
H - 1, 13	3.53	At1g30370	hypothetical protein	1.00E-34
M - 2, 18	3.47	At4g13940	adenosylhomocysteinase	2.00E-94
H - 23, 4	3.46	At3g54690	sugar-phosphate isomerase - like protein	9.00E-35
G - 2, 28	3.29	At3g25860	dihydrolipoamide S-acetyltransferase	2.00E-33
L - 21, 13	3.25	At5g12210	Rab geranylgeranyltransferase, beta subunit	2.00E-61
C - 24, 4	3.22	At2g39770	GDP-mannose pyrophosphorylase	5.00E-55
E - 15, 27	3.14	At5g63180	pectate lyase	9.00E-53
O - 17, 6	3.11	At5g23960	(+)-delta-cadinene synthase (d-cadinene synthase)	2.00E-20
E - 12, 28	2.93	At3g23920	beta-amylase, putative	2.00E-83
E - 9, 5	2.91	At4g33230	pectinesterase - like protein	1.00E-30
I - 20, 1	2.76	At3g59480	fructokinase-like protein	1.00E-54
I - 2, 13	2.75	At2g26080	putative glycine dehydrogenase	4.00E-52
E - 13, 12	2.71	At2g27500	beta-1,3-glucanase like protein	1.00E-17
I - 13, 13	2.65	At3g27060	ribonucleotide reductase small subunit, putative	4.00E-99
M - 4, 27	2.61	At1g04690	putative K+ channel, beta subunit	7.00E-62
G - 7, 3	2.54	At2g41530	putative esterase D	e-104
H - 18, 3	2.53	At3g47520	chloroplast NAD-dependent malate dehydrogenase	4.00E-26
M - 2, 13	2.53	At4g38970	putative fructose-bisphosphate aldolase	5.00E-42
E - 24, 9	2.52	At3g06860	fatty acid multifunctional protein (AtMFP2)	8.00E-60
N - 17, 6	2.44	At3g49960	peroxidase ATP21a	1.00E-62
E - 13, 28	2.43	At3g23920	beta-amylase, putative	2.00E-83
D - 16, 2	2.34	At4g33010	P-Protein - like protein	4.00E-50
J - 18, 6	2.28	At3g24840	phosphatidylinositol transfer protein, putative	4.00E-43
F - 10, 12	2.19	At2g16570	amidophosphoribosyltransferase	1.00E-67
K - 14, 12	2.16	At4g24040	trehalase - like protein	6.00E-28
H - 14, 13	2.13	At3g13080	ABC transporter, putative	2.00E-30
L - 10, 6	2.05	At4g24000	putative protein	4.00E-29
J - 12, 3	2.00	At3g13390	L-ascorbate oxidase precursor, putative	4.00E-47
			ENERGY	
J - 24, 23	11.46	At2g02850	putative basic blue protein (plantacyanin)	3.00E-31
P - 15, 5	9.10	At1g20020	ferredoxin--NADP reductase precursor, putative	3.00E-35
M - 17, 5	2.88	At1g17740	putative protein	1.00E-52
A - 20, 8	2.61	At2g05710	cytoplasmic aconitate hydratase	1.00E-45
K - 10, 13	2.36	At1g20850	putative cysteine proteinase	5.00E-78
P - 10, 3	2.15	At1g23800	putative aldehyde dehydrogenase NAD+	8.00E-19
			CELL GROWTH, CELL DIVISION AND DNA SYNTHESIS	
O - 2, 25	14.68	At1g50010	tubulin alpha-2/alpha-4 chain, putative	6.00E-63
G - 6, 14	8.39	At5g26751	shaggy-like kinase alpha	1.00E-65
E - 6, 24	8.11	At4g37490	cyclin cyc1	2.00E-33
J - 20, 21	7.86	At1g13180	actin-like protein	1.00E-87
N - 6, 22	6.87	At1g68370	ARG1 protein (Altered Response to Gravity)	3.00E-56
K - 15, 23	5.53	At1g50010	tubulin alpha-2/alpha-4 chain, putative	6.00E-63
F - 12, 26	5.38	At3g01490	putative protein kinase	3.00E-83
F - 24, 20	4.45	At4g05050		3.00E-74
K - 14, 2	4.07	At1g50010	tubulin alpha-2/alpha-4 chain, putative	6.00E-63
G - 12, 2	3.41	At5g62700	tubulin beta-2/beta-3 chain	8.00E-77
J - 10, 21	3.36	At5g62700	tubulin beta-2/beta-3 chain	8.00E-77
F - 9, 17	2.74	At5g23860	beta tubulin	5.00E-29
I - 24, 9	2.74	At5g65270	GTP-binding protein	2.00E-46
J - 6, 24	2.67	At3g07720	unknown protein	9.00E-57
G - 8, 12	2.54	At5g23860	beta tubulin	e-143
N - 3, 17	2.45	At4g12400	stress-induced protein sti1 -like protein	9.00E-34
			TRANSCRIPTION	
A - 15, 24	13.06	At5g11590	transcription factor like protein	1.00E-37
P - 4, 27	10.69	At5g59970	histone H4 - like protein	3.00E-38
B - 18, 23	6.48	At2g41130	unknown protein	2.00E-15
N - 17, 22	5.21	At1g22490		2.00E-12

I - 9, 26	5.03	At5g25190	ethylene-responsive element - like protein	9.00E-43
C - 15, 1	4.88	At5g59970	histone H4 - like protein	7.00E-41
C - 16, 10	4.67	At5g02560	putative protein	2.00E-38
D - 17, 7	3.86	At3g54560	histone H2A.F/Z	5.00E-48
I - 15, 6	3.38	At2g36320	unknown protein	3.00E-15
N - 24, 6	3.35	At3g22320	RNA polymerase I, II and III 24.3 kDa subunit	1.00E-46
E - 16, 21	3.30	At5g02560	putative protein	2.00E-38
N - 10, 13	2.55	At4g25550	putative protein	7.00E-80
G - 8, 13	2.39	At1g12770	similar to ATP-dependent RNA helicase	3.00E-39
P - 14, 1	2.29	At4g21840	putative protein	7.00E-32
C - 17, 7	2.25	At5g54640	histone H2A	1.00E-52
E - 15, 11	2.05	At2g44830	protein kinase like protein	1.00E-77
C - 9, 10	1.95			
L - 14, 4	1.90	At1g53170	putative ethylene response element binding factor 4	6.00E-27
			PROTEIN SYNTHESIS	
D - 13, 27	11.97	At5g27700	ribosomal protein S21 - like	5.00E-31
P - 9, 8	3.53	At4g31180	aspartate--tRNA ligase - like protein	3.00E-49
J - 17, 13	2.05	At1g62750	elongation factor G, putative	6.00E-59
O - 14, 6	1.96	At3g63490	chloroplast ribosomal L1 - like protein	3.00E-71
			PROTEIN DESTINATION	
L - 16, 25	17.63	At4g33490	nucellin -like protein	8.00E-25
A - 17, 21	10.52	At1g28110	serine carboxypeptidase II	e-107
C - 12, 21	7.85	At4g39220	AtRer1A	1.00E-34
M - 9, 7	6.22	At5g06860	polygalacturonase inhibiting protein 1	2.00E-27
N - 12, 24	5.04	At5g67360	cucumisin-like serine protease	6.00E-39
K - 16, 2	3.67	At5g56040	receptor protein kinase-like protein	2.00E-21
A - 11, 12	2.06	At5g42080	dynamin-like protein	9.00E-55
			TRANSPORT FACILITATION	
B - 14, 8	23.43	At2g22500	putative mitochondrial dicarboxylate carrier protein	1.00E-27
D - 23, 25	18.61	At3g16240	delta tonoplast integral protein (delta-TIP)	3.00E-35
F - 5, 24	13.34	At4g01470	putative water channel protein	4.00E-67
J - 1, 1	11.91	At4g01470	putative water channel protein	4.00E-67
C - 17, 5	11.39	At3g54820	aquaporin/MIP - like protein	3.00E-47
G - 4, 25	4.58	At1g15460	putative protein	3.00E-54
I - 19, 18	2.83	At2g16850	putative aquaporin (plasma membrane intrinsic protein)	2.00E-39
K - 17, 6	2.71	At4g00430	probable plasma membrane intrinsic protein 1c	1.00E-77
A - 18, 9	2.68	At3g58730	vacuolar-type H+-ATPase (v-ATPase) subunit D	1.00E-23
F - 12, 2	2.59	At1g52190	putative peptide transporter	4.00E-40
			INTRACELLULAR TRANSPORT	
L - 22, 26	10.68	At2g13830	predicted GPI-anchored protein	6.00E-29
B - 22, 22	5.47	At3g18140	WD-repeat protein, putative	1.00E-34
F - 9, 26	3.98	At3g54300	synaptobrevin -like protein	6.00E-18
			CELLULAR BIOGENESIS (proteins are not localized to the corresponding organelle)	
B - 12, 27	64.61	At2g03090	expansin like protein	5.00E-38
H - 13, 27	27.94	At4g38400	putative pollen allergen	1.00E-36
E - 8, 27	24.45	At2g40610	putative expansin	1.00E-56
B - 9, 24	17.87	At2g45180	putative proline-rich protein	2.00E-21
A - 18, 22	11.72	At1g26770	expansin 10	1.00E-38
N - 22, 23	8.69	At3g45960	putative protein	1.00E-60
O - 1, 14	4.23	At2g45180	putative proline-rich protein	4.00E-30
G - 3, 4	4.15	At2g28950	putative expansin	9.00E-64
I - 14, 2	3.36	At1g67980	S-adenosyl-L-methionine	1.00E-14
			CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION	
G - 6, 26	28.21	At1g73620	thaumatin-like protein	5.00E-58
J - 3, 12	23.77	At1g74670	GAST1-like protein	3.00E-28
F - 20, 12	19.62	At3g18710	hypothetical protein	6.00E-24
C - 10, 26	14.05	At2g28790	putative thaumatin	3.00E-47
G - 23, 25	7.51	At1g74670	GAST1-like protein	2.00E-37

H - 20, 9	7.42	At1g74670	GAST1-like protein	3.00E-33
A - 18, 12	3.80	At3g11410	protein phosphatase 2C (PP2C)	2.00E-26
P - 3, 14	2.86	At1g16140		1.00E-29
J - 21, 12	2.66	At1g14000	putative protein kinase	9.00E-57
L - 12, 4	1.93	At5g54380	receptor-protein kinase-like protein	1.00E-61
K - 12, 3	1.87	At3g45440	receptor like protein kinase	3.00E-18
			CELL RESCUE, DEFENSE, CELL DEATH AND AGEING	
L - 1, 25	12.36	At4g04220	putative disease resistance protein	7.00E-16
G - 18, 2	8.74	At5g66390	peroxidase	9.00E-39
F - 8, 23	7.54	At5g59720	heat shock protein 18	3.00E-24
F - 4, 12	3.94	At2g28190	putative copper/zinc superoxide dismutase	7.00E-54
			IONIC HOMEOSTASIS	
I - 12, 2	3.19	At5g01600	ferritin 1 precursor	7.00E-13
G - 10, 2	2.60	At2g40300	putative ferritin	2.00E-41
			CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle)	
C - 8, 11	2.46	At3g63140	mRNA binding protein precursor - like	5.00E-37
H - 8, 13	2.44	At3g63410	putative chloroplast inner envelope protein	4.00E-32
			DEVELOPMENT	
J - 15, 9	21.11	At1g75500	nodulin-like protein	1.00E-65
			RETROTRANSPOSONS AND PLASMID PROTEINS	
E - 13, 25	3.92	At4g03810	putative retrotransposon protein	9.00E-11
			UNCLASSIFIED PROTEINS	
O - 18, 27	69.30	At4g08950	putative phi-1-like phosphate-induced protein	2.00E-32
P - 14, 23	62.03	At1g70840	unknown protein	2.00E-36
D - 12, 27	38.14	At4g08950	putative phi-1-like phosphate-induced protein	8.00E-45
D - 9, 13	16.36	At1g30200	unknown protein	6.00E-15
C - 2, 12	15.92	At4g08950	putative phi-1-like phosphate-induced protein	3.00E-46
B - 8, 26	14.53	At5g61660	structural protein - like	1.00E-16
G - 9, 28	12.56	At5g07030	nucleoid DNA-binding-like protein	4.00E-42
O - 1, 25	12.38	At5g01650	light-inducible protein ATLS1	8.00E-50
K - 20, 18	11.61	At4g12420	pollen-specific protein - like	2.00E-10
H - 6, 28	11.60	At5g07030	nucleoid DNA-binding-like protein	4.00E-42
B - 11, 18	10.52	At5g35570	axi 1 (auxin-independent growth promoter)-like protein	2.00E-20
M - 7, 11	10.28	At2g40140	putative CCCH-type zinc finger protein	2.00E-16
L - 22, 22	9.75	At5g23870	pectinacetylesterase	1.00E-57
G - 1, 2	9.26	At2g47710	Unknown protein	2.00E-61
K - 3, 11	9.17	At4g25030	putative protein	1.00E-85
B - 11, 13	8.96	At1g49660	unknown protein	2.00E-30
P - 10, 17	8.75	At5g51550	putative protein	3.00E-56
E - 2, 2	8.73	At1g29880	glycyl-tRNA synthetase	3.00E-64
G - 2, 25	8.52	At1g70830	unknown protein	3.00E-38
O - 6, 26	8.19	At3g49290	putative protein	2.00E-16
O - 20, 22	7.56	At5g28010	major latex protein homolog - like	2.00E-32
F - 9, 23	7.42	At5g07030	nucleoid DNA-binding-like protein	7.00E-30
C - 2, 28	6.96	At5g08560	WD-repeat protein-like	2.00E-52
G - 6, 25	6.52	At1g71180	putative dehydrogenase	2.00E-49
A - 9, 10	6.21	At3g62040	putative protein	4.00E-67
P - 12, 5	5.58	At2g32150	putative hydrolase	5.00E-61
N - 21, 23	5.49	At2g45750	hypothetical protein	6.00E-49
L - 3, 4	5.32	At2g41010	unknown protein	2.00E-14
D - 10, 25	5.22	At1g73010	hypothetical protein	5.00E-33
K - 16, 22	4.68	At3g62570	putative protein	3.00E-11
O - 17, 27	4.60	At5g06570	putative protein	2.00E-26
N - 10, 6	4.54	At4g00410	putative protein	1.00E-27
L - 15, 24	4.27	At5g65160	putative protein	2.00E-46
K - 9, 20	4.16	At4g08950	putative phi-1-like phosphate-induced protein	7.00E-44
L - 23, 13	4.15	At5g56170	predicted GPI-anchored protein	1.00E-33
B - 19, 21	4.11	At5g48930	anthranilate N-benzoyltransferase	2.00E-26

G - 9, 13	4.04	At1g78040	similar to phosphoglycerate mutase 1	4.00E-34
K - 22, 9	3.87	At1g55090	hypothetical protein	9.00E-56
K - 15, 2	3.86	At5g55480	predicted GPI-anchored protein	4.00E-51
J - 14, 2	3.85	At5g20060	putative protein	3.00E-68
P - 21, 7	3.83	At4g10840	putative protein	1.00E-57
H - 9, 2	3.71	At1g64110	unknown protein	8.00E-12
D - 15, 9	3.70	At1g70090	putative protein	6.00E-87
L - 13, 2	3.60	At4g38180	hypothetical protein	3.00E-26
O - 16, 24	3.56	At3g54800	putative protein	1.00E-19
G - 4, 8	3.47	At4g29950	putative protein	2.00E-18
F - 17, 8	3.40	At5g12010	putative protein	7.00E-65
P - 4, 18	3.31	At2g39130	unknown protein	6.00E-13
A - 19, 13	3.30	At4g31080	putative protein	3.00E-19
E - 13, 21	3.23	At5g11700	putative protein	4.00E-74
O - 21, 6	3.20	At4g29720	putative protein	2.00E-43
P - 7, 26	3.20	At1g46480	hypothetical protein	4.00E-42
P - 12, 9	3.11	At4g28300	predicted proline-rich protein	5.00E-27
O - 10, 28	3.08	At5g64030	ankyrin-like protein	9.00E-41
G - 21, 9	3.00	At5g62180	putative protein	3.00E-40
B - 12, 8	2.90	At1g48750	putative lipid transfer protein	2.00E-21
H - 8, 2	2.89	At3g49190	putative protein	1.00E-17
H - 17, 13	2.88	At4g17600	Lil3 protein	2.00E-64
P - 8, 26	2.81	At1g46480	hypothetical protein	4.00E-42
N - 3, 18	2.77	At3g63010	putative protein	4.00E-31
K - 14, 18	2.76	At4g31080	putative protein	3.00E-19
J - 1, 13	2.70	At5g41970	GAMM1 protein-like	2.00E-46
K - 14, 11	2.69	At1g28380	putative protein	3.00E-45
O - 24, 6	2.68	At1g60420	putative protein	3.00E-39
B - 7, 10	2.67	At1g76990	unknown	5.00E-38
E - 8, 13	2.67	At1g28510	putative protein	6.00E-51
C - 11, 9	2.65	At4g33640	putative protein	2.00E-23
B - 16, 7	2.53	At3g52610	putative protein	2.00E-57
P - 22, 7	2.45	At1g21070	unknown protein	2.00E-74
A - 17, 11	2.38	At2g17220	putative protein kinase	1.00E-17
E - 15, 7	2.36	At3g09320	unknown protein	2.00E-53
H - 11, 2	2.31	At1g30580	putative GTP-binding protein	1.00E-33
G - 9, 20	2.26	At4g38800	putative protein	2.00E-68
P - 19, 3	2.24	At3g09740	unknown protein	1.00E-58
J - 16, 11	2.20	At3g23600	unknown protein	7.00E-67
E - 16, 2	2.18	At3g16060	kinesin-like protein	7.00E-81
I - 17, 13	2.17	At3g50150	putative protein	2.00E-16
G - 8, 3	2.15	At5g51260	acid phosphatase	1.00E-24
E - 24, 13	2.13	At4g17600	Lil3 protein	2.00E-64
O - 17, 12	1.92	At1g27460	unknown protein	3.00E-30
L - 13, 4	1.89	At5g10730	putative protein	3.00E-37
J - 12, 23	1.84	At1g09920	unknown protein	8.00E-49
E - 13, 3	1.73	At5g46870	putative protein	7.00E-24

List of the genes belonging to cluster two for which a functional annotation was reported. The 67 genes are grouped according to the functional category to which they belong and ordered according to decreasing induction.

Coordinates	Ratio	Arabidopsis accession	Arabidopsis description	Arabidopsis e-value
			METABOLISM	
H - 14, 24	25.78	At4g25810	xyloglucan endo-1,4-beta-D-glucanase (XTR-6)	1.00E-66
L - 24, 16	20.85	At4g16260	beta-1,3-glucanase class I precursor	2.00E-42
E - 22, 18	18.95	At5g05340	peroxidase	1.00E-95
H - 12, 16	7.98	At4g32940	gamma-VPE (vacuolar processing enzyme)	8.00E-53
D - 18, 14	7.46	At4g21200	gibberellin 20-oxidase - like protein	1.00E-29
N - 6, 15	6.63	At1g30710	putative reticuline oxidase-like protein	9.00E-46
I - 5, 2	3.38	At1g60710	unknown protein	2.00E-86
A - 8, 16	2.86	At3g46970	starch phosphorylase H (cytosolic form) - like protein	3.00E-65
J - 7, 5	2.55	At4g38590	galactosidase like protein	4.00E-54
L - 11, 10	2.19	At2g45290	putative transketolase precursor	5.00E-69
B - 17, 8	1.95	At5g25110	serine/threonine protein kinase-like protein	2.00E-29
			ENERGY	
M - 1, 3	17.11	At1g29930	putative protein	8.00E-65
I - 19, 13	9.53	At1g29930	putative protein	8.00E-65
L - 16, 13	6.92	At4g10340	chlorophyll a/b-binding protein - like	2.00E-53
A - 18, 23	4.08	nad4	-mitochondrial genome- NADH dehydrogenase subunit 4	2.00E-77
L - 13, 7	3.32	nad4	-mitochondrial genome- NADH dehydrogenase subunit 4	2.00E-77
			CELL GROWTH, CELL DIVISION AND DNA SYNTHESIS	
L - 4, 19	4.18	At3g56070	peptidylprolyl isomerase	1.00E-64
D - 19, 1	3.92	At3g12580	putative protein	e-113
B - 6, 25	3.66	At5g02500	dnaK-type molecular chaperone hsc70.1	3.00E-43
P - 8, 20	3.06	At3g57150	putative protein	2.00E-62
J - 20, 5	2.89	At2g16700	actin depolymerizing factor 5	1.00E-59
E - 9, 26	2.33	At3g18480	unknown protein	1.00E-50
			TRANSCRIPTION	
F - 18, 16	12.61	At2g32700	unknown protein	7.00E-54
D - 9, 24	10.82	At3g24500	ethylene-responsive transcriptional coactivator, putative	6.00E-47
A - 8, 24	8.88	At3g24500	ethylene-responsive transcriptional coactivator, putative	6.00E-47
J - 4, 15	6.18	At4g29190	putative protein	4.00E-52
F - 4, 25	5.71	At4g12600	Ribosomal protein L7Ae -like	2.00E-48
B - 19, 25	3.86	At3g57290	eukaryotic initiation factor 3E subunit (TIF3E1, eIF3e)	1.00E-74
N - 17, 20	3.19	At4g32720	putative protein	3.00E-38
			PROTEIN SYNTHESIS	
O - 19, 15	16.28	At4g31700	ribosomal protein S6 - like	2.00E-50
J - 19, 2	6.52	At3g56150	PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR	1.00E-45
P - 19, 15	6.05	At1g77940	similar to ribosomal protein L30	1.00E-51
N - 6, 4	2.62	At5g15200	40S ribosomal protein - like	2.00E-46
K - 5, 4	2.32	At5g02960	putative protein	3.00E-36
H - 18, 5	2.25	At1g54290	putative protein	2.00E-54
H - 21, 9	2.18	At3g59540	60S RIBOSOMAL PROTEIN L38-like protein	5.00E-33
			PROTEIN DESTINATION	

O - 15, 17	14.79	At5g67360	cucumisin-like serine protease	6.00E-39
M - 7, 19	7.00	At1g62290	putative aspartic protease	3.00E-54
L - 15, 6	3.07	At5g21090	leucine-rich repeat protein	2.00E-61
E - 18, 1	2.90	At3g59510	putative protein	1.00E-42
CELLULAR BIOGENESIS (proteins are not localized to the corresponding organelle)				
N - 12, 15	23.10	At3g12500	hypothetical protein	e-124
F - 16, 1	12.89	At3g12500	hypothetical protein	e-124
J - 18, 14	3.35	At1g07360	Unknown protein	1.00E-80
CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION				
F - 21, 7	35.97	At4g11650	osmotin precursor	4.00E-28
M - 22, 9	6.44	At4g11650	osmotin precursor	1.00E-38
P - 15, 22	6.37	At1g78290	similar to protein kinase 1	1.00E-28
H - 9, 18	6.00	At1g75750	putative protein	1.00E-29
I - 17, 18	5.59	At1g75750	putative protein	6.00E-25
K - 8, 18	5.43	At1g75750	putative protein	1.00E-29
A - 10, 13	4.03	At4g08920	Arabidopsis thaliana flavin-type blue-light photoreceptor	1.00E-79
K - 16, 6	2.32	At5g59845	putative protein	1.00E-15
CELL RESCUE, DEFENSE, CELL DEATH AND AGEING				
F - 22, 18	13.46	At4g19810	putative chitinase	9.00E-33
UNCLASSIFIED PROTEINS				
N - 15, 17	8.18	At4g28690	hypothetical protein	7.00E-12
H - 11, 18	7.86	At1g25580	unknown protein	1.00E-49
E - 21, 14	7.34	At4g16560	hypothetical protein	5.00E-19
F - 8, 27	3.81	At3g62630	putative protein	8.00E-26
C - 19, 27	3.70	At1g27470	PWP2 like protein	3.00E-50
I - 7, 7	3.34	At3g61060	putative protein	1.00E-47
I - 23, 18	3.23	At5g19000	putative protein	3.00E-42
J - 16, 19	3.19	At3g47080	putative protein	4.00E-10
C - 1, 1	3.19	At1g65600	hypothetical protein	3.00E-51
P - 8, 11	2.76	At1g71230	c-Jun coactivator-like protein (AJH2)	2.00E-51
O - 2, 12	2.59	At4g09830	putative protein	6.00E-26
N - 14, 6	2.47	At1g47480	hypothetical protein	5.00E-28
L - 13, 6	2.04	At3g57790	putative protein	2.00E-38
A - 3, 3	2.01	At1g33810	unknown protein	3.00E-29
D - 24, 7	1.72	At3g48050	putative protein	1.00E-11