Charles University in Prague
Faculty of Science

Institute of Experimental Botany
Academy of Sciences of the Czech Republic

Jiří Libus

Transcriptional response of plants to genotoxic stress and water deficit

PhD. thesis

Supervisor: RNDr. Helena Štorchová, CSc.

Prague 2007

Acknowledgments

First, I am grateful to my supervisor, Dr. Helena Štorchová and the former director of our institute, Dr. Ivana Macháčková, who jointly enabled me to finish my PhD. studies. Dr. Imre E. Somssich and other staff of MPIZ, as well as Dr. Tomáš Mráček, who substantially facilitated the experimental part of my work. My thanks belong to my parents and grandparents for setting the course of my journey in the direction of natural sciences and to Lenka for supporting me in all aspects of our life. In addition, I owe a lot to many other persons, for making my life easier and more beautiful, also by helping me with my work.

Declaration

I hereby declare that the work presented in this thesis is my own and was carried out entirely with help of literature and aid cited in the manuscript.

Prague, Czech Republic

July 13, 2007

Table of contents

1 List of abbreviations

ABA	Abscisic acid
ANOVA	Analysis of variance
BER	Base excission repair
cDNA	Complementary DNA
CK	Cytokinin
Cp	Crossing point
DNA	Deoxyribonucleic acid
ds	Double-stranded
GO	Gene ontology
HR	Homologous recombination
ICT	Internal control transcript
LoTrEC	Local Trend based gene Expression data Clustering
MMR	Mismatch repair
MMS	methyl-methanesulfonate
mRNA	Messenger RNA
MS	Microsoft
NER	Nucleotide excission repair
NHEJ	Non-homologous end joining
PCD	Programmed cell death
PCR	Polymerase chain reaction
qPCR	Quantitative real-time PCR
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT	Reverse transcription
RWC	Relative water content
SCF	Skp1/cullin/F-box protein complex
SD	Standard deviation
ss	Single-stranded
STEM	Short Time-series Expression data Miner
TLS	Translesion synthesis
WT	Wild type (not transgenic)

2 Aims of the work

1) To study the transcriptional response of thale cress (*Arabidopsis thaliana*) suspension culture to MMS by means of low density cDNA array and quantitative RT-PCR.

   a) To design an algorithm for cluster analysis of the time-series expression data.

   b) To compare the expression profiles characteristic for treatments with high concentration of MMS, low MMS an high MMS with pretreatment.

   c) To confirm selected profiles by qRT-PCR.

2) To explore the transcript abundance changes in response of selected genes in tobacco (*Nicotiana tabacum*) subjected to drought stress.

   a) To compare the effect of drought on gene expression of wild type and *ZOG1*-expressing plants.

   b) To evaluate the effect of leaf position on drought-response.

   c) To compare suitability of different RT-priming strategies for the analysis.

3 Introduction

3.1 Transcriptional response of *Arabidopsis thaliana* to alkylation stress

3.1.1 DNA damage

The vital function of DNA as the principal carrier of genetic information is constantly threatened by various attacks against its integrity. In general, the causative factor can be physical (such as radiation – ultraviolet, ionizing) or chemical. In the aqueous environment inside the cell, hydrolytic damage to various parts of DNA molecule is common, resulting in loss of bases (mainly depurination) and base deamination. Reactive forms of oxygen produced in both mitochondrial and chloroplast (in autotrophic cells) metabolic processes also contribute to DNA alterations. In total, the frequency of spontaneous lesions was estimated to 20,000 per mammalian cell and day ().

In addition, many chemicals can react with either DNA bases or the backbone, creating diverse types of modification. Whole molecules or just reactive groups of atoms can be added to DNA bases (for example benzo(a)pyrene). If the adduct possesses another reactive part (it is bi-functional), it can react once more with the same or with another molecule, forming thereby a crosslink. Crosslinked DNA strands cannot be properly unwinded and therefore transcribed or replicated. Mitomycin C is a popular crosslinking drug employed both experimentally and for cancer therapy. Other compounds can catalyze production of reactive oxygen species - the most abundant representants are hydrogen peroxide, superoxide anion and hydroxyl radical. These rather short-lived agents result in base oxidation or DNA strand breaks, resembling thereby effects of ionizing radiation (e.g. anti-cancer drug bleomycin).

DNA damage has two more-or-less separate adverse effects on the cell. First, it is cytotoxic either directly – by preventing replication and transcription of the affected genes, in some cases by fragmenting chromosomes which leads to genome instability - or indirectly, by activation of PCD. Second, the DNA-repair is not always perfect in terms of fidelity and so it causes an increased rate of genetic information change – mutation. In addition, some base modifications directly mispair with one or more canonical counterparts (e.g. 8-oxo-guanine is complementary to any of A, C, G and T). The cytotoxicity of some alkylating, DNA strand breaks or interstrand crosslinks inducing compounds is the basis of function of some anti-cancer chemotherapeutics (temozolomide, bleomycin, mitomycin C). The increased frequency of mutations after application of mutagens has been widely used in basic as well as in applied research and breeding of plants.

3.1.2 DNA alkylation

One mechanism of chemical modification of DNA is its alkylation. The donors of the alkyl groups can be either endogenous (such as S-adenosylmethionine) or from the environment (e.g. naturally occurring organohalogens , tobacco-specific N-nitosamines from tobacco smoke and many artificial chemical pollutants). Reactions of DNA with monofunctional alkylation agents yield mainly bases modified at various positions – either on C or N and O atoms but also alkyl,bis(polynukleotidyl)-phosphotriesters resulting from DNA backbone modification ().

3.1.3 Mechanisms of repair of alkylated DNA

Several DNA-repair pathways safeguard the genome integrity, specifically removing distinct classes of the products. They include receptors that recognize the lesion, enzyme machinery to resolve the problem and also members that connect DNA-repair to *signaling* cascades of the cell, regulating among others cell cycle progression and programmed cell death (PCD).
Some of the alkyl-bases destabilize the glycosidic bond between deoxyribose and the base and spontaneously hydrolyze away resulting in an abasic site (e.g. N^7-methyladenine). Besides of direct reversal of alkylated bases by alkyltransferases and oxidative demethylases (AlkB-homologs; reviewed by), this type of lesions is predominantly removed by means of base excision repair (BER). The defective base hydrolysis is catalyzed by a family of specific alkyl-base-DNA-glycosylases. Apurinic/apyrimidinic site specific (AP-) endonuclease introduces a single strand break 5' of the abasic site and DNA polymerase β or λ replaces the abasic nucleotide with a complete one, matching the opposite strand. DNA-ligase then seals the DNA strand break.
Alternatively, nucleotide excision repair (NER) or mismatch repair (MMR) pathways can remove the alkylation damage. NER is accomplished by multisubunit complexes containing XP (*Xeroderma pigmentosum*) proteins. The damaged DNA strand is incised up- and downstream of the lesion by specific endonucleases and unwinded by a helicase. The resulting gap is then replaced with a stretch of DNA newly synthesized by a repair DNA-polymerase and sealed by a DNA-ligase. The actions performed by MMR are in general similar to NER. The main functional difference is that while NER detects distortion of the double-helix caused by DNA modifications, MMR scans the molecule for mispairing bases.
All of BER, NER and MMR produce single strand breaks as an intermediate of their action. If not repaired before the next round of replication, single strand breaks transform to double strand breaks (), one of the most toxic types of DNA damage. Faulty repair of double strand

breaks represents the mechanism of inducing chromosomal aberrations by simple alkylating agents.

Homologous recombination (HR) and nonhomologous end joining (NHEJ) are the two possible ways of repairing double-strand DNA breaks. Homologous recombination takes advantage of the other copy of the affected chromosome, using it as a template for break repair. However, any DNA molecule with long enough stretch of sequence homology can be employed. That is why HR is the mechanism utilized for gene-targeting methods. HR results in somatic sister chromatide exchanges but it is generally error-free. Double strand breaks are repaired predominantly by HR in budding yeast.

On the other hand, NHEJ joins the loose ends of DNA molecules at a site of microhomology, which often produces sequence changes (insertions or deletions). However, NHEJ is the preferred pathway in most higher eukaryotes, including plants. The reason for this could be the high proportion of repetitive sequences in most genomes, that could lead to illegitimate HR and thus chromosomal aberrations.

If modified bases or abasic sites persist in the DNA until replication, they block the progression of replicative DNA-polymerase complex. The stalled replication fork can be rescued by translesion synthesis (TLS) DNA-polymerases. These enzymes generally possess low fidelity (and thus produce a considerable number of mutations) but they can replicate a damaged template, thereby retrieving the cell cycle progression. There are several TLS DNA-polymerases, more or less specific to various DNA modifications.

The importance of capability to respond to all possible kinds of DNA damage is illustrated by the severity of syndromes resulting from impaired DNA-repair genes in all organisms. Although the pathways are partially redundant, loss of function of one of them usually causes developmental defects, susceptibility to DNA damage and in animals a high cancer rate. Human syndromes linked with DNA-repair insufficiency include *Xeroderma pigmentosum* (compromised function of NER or TLS), Cockayne syndrome (impaired transcription-coupled NER), Fanconi anemia (problems in response to DNA crosslinks), Nijmegen breakage syndrome (lack of efficient NHEJ), Hereditary non-polyposis colon cancer (MMR) or *Ataxia telangiectasia* (mutations in the gene coding for a proteinkinase activated after detection of DNA breaks).

3.1.4 Methyl methanesulfonate

Methyl methanesulfonate (MMS, Figure 1) has been used as the most frequent model DNA-damaging (genotoxic) compound for tens of years. It is a methylating agent of S_N2 type; that means it attacks predominantly positions N^7 of guanine and N^3 of adenine. As mutagenicity of

methylation agents correlates with their O-alkylation efficiency (producing O^6-methylguanine and O^4-methylthymine), MMS is not as mutagenic as S_N1 alkylating chemicals (e.g. other common experimental genotoxins, N-methyl-N-nitrosourea and N-methyl-N'-nitro-N-nitrosoguanidine). Ethyl methanesulfonate (EMS), used to mutagenize plants for experimental and breeding purposes is also a more potent mutagen because ethylating agents generally show higher tendency to modify oxygen positions than their methyl- counterparts ().
The DNA methylation produced by MMS has been shown to induce DNA strand breaks (as a result of excision repair processing) and subsequent chromosome aberrations (). MMS can also methylate RNA () and amino-acids cysteine and histidine *in vitro* (). Accordingly, various organisms possess mechanisms to repair or eliminate modified mRNA () and protein (). Mutations in the involved genes produce generally phenotypes sensitive to methylating chemicals.

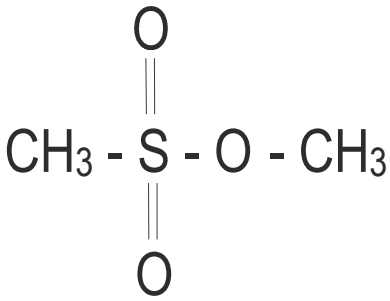


Figure 1. Structural formula of MMS.

3.1.5 Transcriptional response to MMS

The global changes in gene expression after treatment with MMS were studied in several organisms. First of them was yeast (*Saccharomyces cerevisiae*;). Among the genes induced by 1 hour in 0.1% (approximately 10 mM) MMS, several were known to be involved in resistance to DNA-damaging agents or directly acting in DNA repair. The treatment led to activation of a group of genes connected with general stress response, proteolysis and also primary metabolism. Among the repressed transcripts, the most prominent group was involved in ribosome biosynthesis.
Transcriptional responses of mouse and human cell cultures to MMS have been also studied (). The observed spectra of differential gene activities were significantly different between these organisms, although they were both mammalian leukemia cells. The reason for the fact could be that the mouse culture (L5178Y) possessed a nonfunctional allele of p53, a key modulator of cell-fate after DNA-damage. The human TK6 cells might thus represent a more realistic picture of normal mammalian response to genotoxins. This assumption is consistent

with a much broader transcriptional response of TK6 cells to MMS as well as to bleomycin, encompassing genes known to participate in DNA-repair, cell cycle regulation and apoptosis. The discrepancy could however also relate to the original species. In the mouse L5178Y cells, the set of MMS-induced transcripts did not at all overlap with those upregulated by bleomycin (with a single exception). Also in TK6 culture were most genes possessing differential activity treatment-specific. Anyway, there were 2 genes induced by both treatments after 4 hours and 15 common transcripts with higher abundancy after another 20 hours. Nine of the 15 common activated genes could be annotated as involved in *p53*, *TNF* (Tumor Necrosis Factor), *ERK* (Extracellular-signal-Regulated Kinase) or *JNK* (c-Jun N-terminal Kinase) pathways, thus connected to cell fate. As many of these demonstrated changes in expression also appeared after application of other genotoxins, it seems to be a general response to DNA damage that takes some time to start. Unfortunately, plants lack some of the key players in the mentioned pathways and so these data cannot be expected to apply directly to them. Expressions of various plant genes have been shown separately to be influenced by MMS (e.g.). There is, however, no comprehensive transcriptomic study concerning methylation stress in plants up to date. The only results of macro-/microarray experiments dealing with DNA damaging chemicals, that are either published or accessible in public data repositories, concern *Arabidopsis* plants treated with a combination of bleomycin and mitomycin C (, AtGenExpress at http://web.uni-frankfurt.de/fb15/botanik/mcb/AFGN/atgenex.htm). Although both of these compounds are also genotoxic, their main mode of action is very different. Bleomycin induces double-strand breaks while mitomycin C forms predominantly interstrand crosslinks. As the enzymatic machinery necessary to remove these lesions only partially overlaps with that for alkylation damage repair, the spectrum of transcripts induced and repressed by MMS treatment should be pressumably also different.

3.1.6 Adaptation to genotoxic stress

The fact that pretreatment with a moderate stress enhances survival of various organisms in conditions of high stress is generally accepted. In the field of DNA damage, studied the effect of preconditioning of *Escherichia coli* with low concentration of various mutagens on the level of mutagenesis caused by a higher dose. observed a reduction in the frequency of genotoxin-induced mutations and also chromosomal aberrations (clastogenic adaptation) in Chinese hamster cells, when such a preconditioning was applied. The phenomenon can be induced also by a pretreatment with another (but not any) DNA-damaging chemical.

3.1.7 DNA array technology

High throughput methods for assaying simultaneously expressions of many genes have become widely used. They give a global image of the cell processes affected by the particular treatment, mutation, developmental stage transition etc. The most common method to accomplish this task employs hybridization to an array of immobilized probes. Although whole-genome arrays are readily accessible for several species, in some cases it is still effective to use a smaller custom set of selected transcripts spotted on a lower-density array. The main advantages are lower price and no need for highly specialized equipment. Hybridization arrays allow simultaneous assaying of abundances of many transcripts. However, they suffer from a range of artefacts. One of them is hybridization of diverse but similar molecules to a particular probe. If the authentic target is rare, even a limited complementarity of a more abundant transcript can lead to substantial overestimation of its expression level. In addition, this can also mask expression changes (e.g. turning-on from effectively zero to moderate). Therefore it is advisable to test the results at least for a subset of differentially expressed genes using an independent method.

3.1.8 Cluster analysis of expression data

Any array experiment creates a big amount of data. To organize them and make them easier to comprehend, it is useful to find transcripts with similar expression changes in various conditions. Groups of such genes might have a common mode of regulation and/or function. Several possible similarity (or dissimilarity) measures exist that can be used to define the degree of concordance between expresion profiles. Most of them treat all the measured expression values as coordinates of a vector in a multidimensional space (with number of dimensions equal to the number of conditions tested). Some of them are based on Euclidean distance, i.e. absolute difference between the compared vectors – they take into account both the changes and absolute intensity of expression. Others use various kinds of correlation – they generally overlook the absolute values, only the shape of the expression profile is considered. Various algorithms then organize the gene expresion vectors either to separate clusters (K-means etc.) or to a form in which the groups of similar profiles can be discovered and picked. Examples of the latter strategy are hierarchical clustering (the size and number of the clusters is defined by setting a cut-off value), self-organizing maps and principal component analysis ().

3.1.9 Array data validation by real-time PCR

Recently, quantitative polymerase chain reaction coupled with reverse transcription (RT-qPCR) is usually the method of choice for the confirmation of DNA-array data (). In contrast to DNA-arrays, this method can be truly quantitative. In comparison with Northern blotting, also sometimes used to confirm array data, RT-qPCR does not depend on hybridization of a long probe and so it can be very specific – even alleles differing by single nucleotide polymorfisms may be distiguished. The key to this specificity is selection of primers and in some experimental setups also a relatively short labelled probe. Another advantage of qPCR-based methods is their outstanding sensitivity (down to several copies of the target) combined with enormous dynamic range of concentrations in which the target stays quantifiable. However, qRT-PCR reliability depends on, among other factors, appropriate normalization (for a review, see).

3.1.10 qPCR data normalization

Although individual qPCR reactions, if performed properly, generally suffer only from little variability, it is still necessary to compensate (or "normalize") for differences occuring both in the starting material and qPCR process themselves. Measured data normalization thus allows to directly compare individual samples to each other. The preferred method of quantitative RT-PCR normalization uses housekeeping genes with presumably invariant levels of expression as internal controls. Housekeeping gene-based normalization corrects the measured transcript levels for variable starting RNA amounts and for differences in RT efficiency. However, as there are no universally applicable genes with invariant expression, it is necessary to carefully evaluate the expression of candidate reference genes for every particular experimental system. Normalization with suboptimal house-keeping genes may result in different estimated values and lead to erroneous interpretations (). To avoid a bias caused by the expression fluctuation of a single reference gene, proposed computation of the correction factor from several internal controls. However, this approach may significantly increase the cost and laboriousness of experiments. Another approach derives the correction factor for each sample from the input RNA amount, based either on spectrophotometric (A_{260}), or on fluorometric estimation (), meaning that there is no need to select a proper reference gene and verify its expression. However, this method relies upon the reproducibility of the RT reaction, which has been shown to be a major source of quantitative RT-PCR variation (). Several other normalization strategies have been reviewed recently ().

3.2 Drought stress in tobacco

Water represents a major limiting factor for plant growth and development. Accordingly, plant responses to its shortage are complex (). Stress mitigation strategy involves fast changes, e.g. closure of stomata, as well as longer-term modification of plant metabolism and growth. The latter processes involve changes in expression of many genes, as it was shown for *Arabidopsis thaliana* using whole genome DNA arrays (, Genevestigator at https://www.genevestigator.ethz.ch/;). Distinct parts of the plant – leaves of different position on the stem – obviously differ in their stress response. To perform multiple analyses from each leaf separately, Arabidopsis appears to be a rather difficult model plant. On the contrary, tobacco plants are significantly bigger and thus allow to quantify various metabolites from a single leaf in addition to RNA isolation for assessing gene expression. This is important for a complex evaluation of metabolism in a precisely specified site of the plant. Tobacco thus represents a more suitable alternative than Arabidopsis. In addition, this model plant has been used for many physiological experiments including stress treatments.

The drought responses are orchestrated by hormone signaling. Phytohormones modulate both fast stress responses and transformation of the stress-related signals into specific changes of gene expression. Most studies on the role of plant hormones in stress physiology have focused on abscisic acid (ABA). Elevation of ABA concentration after initiation of various stresses has been numerously reported (). In response to drought, ABA decreases the transpiration rate by promoting stomatal closure. In addition, expression of many drought- and cold-inducible genes depends on ABA action either via bZIP transcription factors binding to ABRE (ABA response element) present in promoters of the target gene or via MYB and MYC proteins binding to their respective recognition sites MYBRS and MYCRS ().

3.2.1 Cytokinins

In order to achieve precise control in any system, both positive and negative regulators are required. Many physiological processes induced by drought, particularly those mediated by ABA, e.g. acceleration of leaf senescence and closing of stomata, may be counteracted by cytokinins (CKs; for review see).

CKs are a group of derivatives of adenine with regulatory activities in plant growth and development. They stimulate cell division, reproduction of plastids, chlorophyll synthesis and development of the shoot organs. The latter is connected with the release from apical dominance and inhibition of leaf senescence. CKs occur in several forms. Apart from free CK

bases and ribosides (the active forms), modifications such as phosphates and glucosides exist. In addition, cis and trans isomers differ in biological activity and also metabolic processing. Exogenous CKs were reported to increase the tolerance to mild stress and to speed up the subsequent recovery of the plant (). Applied CKs relieved negative effects of water stress on chlorophyll and carotenoid content and photochemical activity (e.g.) or improved recovery of stomatal conductance and net photosynthetic rate after rehydration of drought-stressed plants (). Moreover, transcription of many stress-induced genes can be stimulated by CK ().

3.2.2 CK metabolism manipulation

It should be noted that the composition and concentration of exogenous phytohormones at the site of action might substantially differ from the applied dose, not only because of the losses during their uptake and translocation, but also due to their metabolism and allocation (e.g.). These drawbacks may be overcome by altering endogenous CK levels by manipulation of genes coding for CK metabolic enzymes. Plants expressing ectopicaly the gene *IPT* (coding for isopentenyltransferase, the rate-limiting enzyme of CK synthesis) under the control of either constitutive (Cauliflower mosaic virus *35S* promoter) or inducible promoters (responsive to Cu^{2+} or heat-shock) were prepared (for review see). However, substantially elevated content of active CKs in transformants with increased CK biosynthesis usually causes severe changes in plant morphology, which may complicate evaluation of the potential CK role in response to stress conditions. To minimize this limitation, transgenic tobacco plants containing *IPT* gene driven by a senescence-specific *SAG12* promoter were developed. Activation of *IPT* transcription in these plants is triggered only in senescing leaves and the plants develop normally. *SAG12::IPT* plants were used for evaluation of CK effect on leaf senescence (), photosynthesis and nitrogen partitioning (), flooding tolerance (), seed germination and response to mild drought stress (). An alternative approach is over-expression of the gene coding for *trans*-zeatin O-glucosyltransferase (*ZOG1*). *ZOG1* cDNA was isolated from Lima bean (*Phaseolus lunatus;*) and transferred to tobacco plants, under the control of either constitutive (*35S*) or senescence inducible (*SAG12*) promoters (). The levels of physiologically active CKs (bases, ribosides) did not substantially differ between the transgenic and wild type plants, being still under the control of mechanisms regulating CK homeostasis. However, the total CK level (including glucosides) was highly elevated. CK-O-glucosides, responsible for the increase, represent a readily mobilisable CK reserve, capable to provide additional active CKs if a need arises. Conversion of CK O-glucosides to the active CK forms can be accomplished by □-glucosidases (e.g.). CK O-glucosides are resistant to cytokinin oxidase/dehydrogenase (*CKX*), the main CK degrading enzyme ().

3.2.3 Drought stress and cytokinin signaling related genes

NtERD10B, one of dehydrins, is a relatively small (169 amino acids), glycine-rich protein. It is a member of the *LEA* (Late Embryogenesis Abundant) family of mainly stress-inducible proteins. It is a group 2 *LEA* protein, as it contains a conservative element EKKGIMDKIKEKLPG at its C-terminus (amino acids 148 – 162) and an additional incomplete copy of the motif (114 - 122). It is presumed to protect the cell components against effects of osmotic stress, possibly as a molecular chaperone. *NtERD10B* was found as one of the *Arabidopsis thaliana ERD10* homologues in tobacco by . It was shown to be induced by drought and low temperature treatments and also in tobacco plants over-expressing *Arabidopsis DREB1A* gene. *DREB1A* is a transcriptional activator of stress-induced genes with 41 identified targets in *Arabidopsis* genome (). When ectopically expressed in tobacco, as well as in rice, wheat and *Arabidopsis*, it confers drought tolerance, probably through regulation of expression of the stress mitigating genes ().

cig1, Cytokinin Induced Gene 1, codes for a putative proline dehydrogenase/oxidase (493 amino acids). Proline dehydrogenases are key enzymes in conversion of proline to glutamate (). As proline is known to serve as an osmoprotectant during various conditions leading to loss of cellular water, its metabolism is directly connected to drought stress. Indeed, observed drought induced decrease in a proline dehydrogenase mRNA concentration in *Arabidopsis*. *cig1* mRNA was shown to accumulate after cytokinin treatment of tobacco suspension culture if the cells were grown in presence of auxin. In medium lacking auxin, *cig1* transcript abundance specifically increased in response to ABA (). *cig1* could therefore represent a functional link between metabolic stress responses and cytokinin signaling.

CRK1 (Cytokinin-Regulated Kinase) was predicted to be a transmembrane receptor Ser/Thr-proteinkinase of 794 amino acids. The abundance of its transcript is quickly (within 30 minutes) reduced after addition of cytokinins (benzyladenine) to tobacco suspension culture. However, this reduction seems to be transient. Proteosynthesis and an unidentified phosphatase activity but not a proteinkinase were shown to be necessary for the reaction (). The authors anticipate *CRK1* function in the early steps of cytokinin signaling. Its disappearance could confer a temporary desensitization of the cells towards its ligand. The ligand is however not yet known.

4 Material and Methods

4.1.1 Arabidopsis thaliana suspension culture

The *Arabidopsis thaliana* suspension culture was obtained from Dr. K. J. Angelis. It was cultivated in a dark orbital shaker at 21°C and 90 rpm in the following medium (per 1 liter): 30 g sucrose, 4.9 g Murashige & Skoog Medium Including Vitamins and MES buffer (Duchefa Biochemie, Haarlem, Netherlands), 1 ml vitamine cocktail and 1 ml of 0.1% (w/v) 2,4-dichlorophenoxyacetic acid in ethanol. After autoclaving, ampicilin was added to the final concentration of 100 mg/l. The vitamine cocktail comprised of 0.5 mg nicotinic acid, 0.5 mg pyridoxine-HCl, 0.1 mg thiamine-HCl and 0.1 mg myoinositol per ml.

4.1.2 MMS treatment

For the MMS treatment (5 mM and 0.5 mM), late logarithmic culture was prepared, 3 days after the last subculture. The cells were sifted out by a stainless steel tea strainer and gently drained away. Approx imately 1 g of material in 10 ml of medium in a separate 50 ml Erlenmeyer flask were used for each time point. To reduce the effect of medium exchange, appropriate amount of MMS was added to the original medium, swirled and immediately added to the cells. The samples were sequentially harvested after expiration of 15 min, 30 min, 1 h, 2 h, 3 h, 5 h and 8 h. All the incubations took place in the standard cultivation conditions. The Control sample (supplied with re-used medium only) was harvested after 2 hours (to experience comparable manipulation). Harvested and drained-away cells were aliquoted into 3 microcentrifuge tubes and flash-frozen in liquid nitrogen.
For the combined MMS treatment, 10 g of drained-away cells were resuspended in 100 ml of re-used medium containing 0.5 mM MMS in a 500 ml flask. After 1 hour, the medium was replaced with another 100 ml of medium, lacking MMS. 2 hours later, the culture was treated with 5 mM MMS in the same way as the non-preconditioned samples.

4.1.3 RNA isolation

Total RNA was isolated using RNeasy Plant Mini kit (Qiagen, Carlsbad, USA) with starting amount of about 300 mg of the material per column. RNA quantity was measured as absorbance at 260 nm wavelength by spectrophotometer and its quality checked on an electrophoretic gel. To denature RNA structures, the RNA samples were heated at 75°C for 2 min in presence of 2,5 % SDS (w/v) prior to electrophoresis.

For development and primary demonstration of the RT yield measurement, total RNA was isolated from *Arabidopsis thaliana* suspension culture using acidic phenol-chloroform-isoamylalcohol extraction according to .

In both cases, the RNA was treated with 0.1 units RNase-free DNaseI (Roche Diagnostics, Mannheim, Germany) per µg RNA in 20 mM Tris-HCl (pH = 8.3), 50 mM KCl and 1 mM $MgCl_2$ for 30 minutes at room temperature. The enzyme was then inactivated by 10 minutes at 65°C after adding EDTA to a final concentration of 1 mM. Alternative treatment with DNA-free (Ambion, Austin, USA) or DNaseI on-column digestion (Qiagen) produced similar results.

4.1.4 Hybridization target labeling

To produce labeled hybridization target, 20 µg RNA of each sample was reverse transcribed in a 30-µl reaction by 400 u MMLV reverse transcriptase RNaseH$^-$ Point Mutant (Promega, Madison, USA) in the buffer recommended by manufacturer, supplied with 1 µg anchored oligo dT primer ($dT_{23}dV$), 0.5 mM dBTP (C, G and T), 0.05 mM dATP and 1 µCi α-^{32}P-dATP. RNA with the primer were heated to 65°C for 5 min and cooled on ice. After addition of all the remaining components, the reaction mixture was incubated at 42°C for 50 min, then denatured at 96°C for 5 min and again cooled on ice. The probe was ready for hybridization at this stage.

4.1.5 cDNA aray design and hybridization

The macroarrays carried PCR-amplified inserts from indicated cDNA library clones from the collection at Max Planck Institute for Plant Breeding, Cologne, Germany, spotted in duplicates by the institute Core Facility staff and equipment onto Hybond-N+ Nylon membrane (Amersham-Pharmacia-Biotech, Uppsala, Sweden). In addition, 34 cDNA fragments not present in the cDNA collection were cloned as described below.

Prior to hybridization, each membrane was prehybridized in 20 ml of 6x SSC supplied with 1x Denhardt solution, 0.5% (w/v) sodium laurylsulfate (SDS) and 100 mg/l sonication-sheered salmon sperm DNA (according to) at 65°C for 2 hours rolling in a hybridization oven.

The hybridization solution was prepared from 5 ml of new prewarmed prehybridization solution and a denatured probe and it immediately replaced the prehybridization solution. After overnight incubation at 65°C with the probe, the membranes were washed twice with prewarmed 2x SSC, 0.1% SDS and once with 0.5x SSC, 0.1% SDS, 10 min at 65°C each.

4.1.6 Expression data acquisition and processing

Still wet washed arrays were heat-sealed into plastic bags and submitted to BAS5000 scanner (Fuji Photo. Film Co., Ltd, Kanagawa, Japan). The densitometry of resulting images was performed using Quantity One software (Bio-Rad, Hercules, CA, USA), subtracting local background surrounding each spot.

Standardization of signal intensities among the arrays was performed by subtracting 2nd percentile (to remove "spot background" corresponding to empty spots) and dividing each value by the median signal of the corresponding array. Median of all standardized values refering to the same transcript-treatment-time point combination (1 to 5 arrays, in most cases 2 to 4; each array possessed 2 spots per transcript) was taken as the mean value for subsequent evaluation.

Visual Basic for Applications implemented in Microsoft Excel versions 2000, XP and 2003 served for programming LoTrEC, the macro presented in this work, aimed for automation of cluster formation from the measured data.

4.1.7 Cloning of cDNA fragments for the cDNA array

cDNA fragments for cloning were PCR amplified by AccuTaq LA (Sigma-Aldrich, St. Louis, USA) from cDNA prepared with SuperScript II reverse transcriptase (Invitrogen, Carlsbad, USA) according to protocols supplied by manufacturers. The list of genes and corresponding primer sequences can be found in Table 1. The PCR products were introduced to pUC57 (Fermentas, Vilnius, Lithuania) plasmid modified for TA cloning as follows. pUC57 was digested by EcoRV, precipitated by ethanol and 2',3'-dideoxy T was added to its 3' ends by 30 u terminal deoxynucleotide-transferase (Promega, Madison, USA) in the buffer recommended by manufacturer, supplied with 100 µM ddTTP and 10 µg linearized plasmid. After 30 min at 37°C, the product was precipitated once more and dissolved in water to a final concentration of 50 ng/µl. 10 µl ligation reaction contained 1 µl of the vector solution, 1 µl 10x ligase buffer, 0.5 µl T4 DNA-ligase (0.5 u; Fermentas, Vilnius, Lithuania) and 7.5 µl of the PCR reaction (containing the product of interest as a single band on a control electrophoresis). The ligation was incubated on ice let to melt at room temperature overnight and stopped by 65°C for 15 min. Electroporation of 1 µl of the reaction mixture usually yielded tens to hundreds of collonies. Clones carrying the desired insert were detected among the white ones on X-Gal/IPTG LB plates(; IPTG and X-gal were spread onto the agar medium just before plating the bacteria) by means of PCR. Part of the colony was transfered to 10 µl PCR reaction containing the primes used for amplification of the cloned fragment. The identity of

the cloned fragment was confirmed by sequencing with Autoread Sequencing Kit Auto (Fermentas, Vilnius, Lithuania) and modified M13-universal and/or M13-reverse primers (Table 1) using ALF express II sequencer (Amersham-Pharmacia-Biotech, Uppsala, Sweden).

4.1.8 Reverse transcription

Reverse transcription (RT) reaction for the quantitative polymerase chain reaction (qPCR) of the *Arabidopsis* samples contained 100 u Moloney murine leukemia virus (MoMLV) RNaseH$^-$ point mutant (Promega, Madison, USA) supplied with its recommended buffer, 0.5 mM each dNTP, 0.5 μg anchored oligo-dT ($dT_{23}dV$) or 2 μg random hexanucleotides (dN_6) and 1 μg RNA. The procedure followed otherwise the steps described in the hybridization target labelling protocol above with the exception of the final denaturation step. 5 min at 70°C was used instead.

For tobacco RNA, 1 μg of RNA was reverse transcribed by 10 units of Transcriptor Reverse Transcriptase (Roche Diagnostics, Mannheim, Germany) with random hexanucleotides (4 μg) or oligo dT primers (dT_{15}; 500 ng). After heating 5 minutes at 65°C, buffer, 0.5 μl of Protector RNase Inhibitor (Roche), 2 μl of 10 mM dNTPs and enzyme were added on ice. The cDNA synthesis proceeded at 55°C for 30 min. If random hexamers were used, a step of 25°C for 10 min was inserted directly after enzyme addition.

4.1.9 Quantitative PCR

The real-time quantitative PCR (qPCR) was performed using LightCycler 1.2, DNA Master Kit PLUS and glass reaction capillaries (Roche Diagnostics, Mannheim, Germany) or polycarbonate capillaries from Genaxxon (Biberach, Germany). contais the list of primers specific for assayed transcripts and their respective annealing temperatures and optimum concentrations. The reaction mixture consisted of 2 μl Master mix (component 1 of the kit), 2 μl of the investigated cDNA (20x or more diluted) and primers. A common mix was prepared for as many samples as possible, avoiding pipetting of volumes lower than 2 μl. The typical PCR program included 10 min at 95°C to activate the enzyme and 50 cycles of 10 s at 95°C, 6 s at primer-specific annealing temperature and 10 s at 72°C. After the PCR, melting curve analysis was performed (95°C, 60°C and heating at 0.1°C/s to 95°C). Annealing temperatures and concentrations of all primers were optimized to get maximum PCR efficiency without nonspecific products. Second derivative maximum (as implemented in LightCycler software v. 3.5) was used to determine crossing points (CP). PCR efficiency (E) was estimated from calibration curve of cDNA serial dillutions. Relative transcript abundance (RA) in sample Y

was calculated from CP in the following way: $RA_{SampleX} = E^{(CPCalibrator - CPSampleX)}$. The result was then standardized by dividing with RA of an internal control transcript (ICT) presumed to have invariable abundance in the experimental conditions. The ICTs used were *PDF2* (At1g13320, primers according to , Table 2) for *Arabidopsis* samples and *Tac9* (GenBank accession X69885),) and *18S* rRNA (AJ236016), in case of random primed cDNA) for tobacco.

4.1.10 Total cDNA yield

Total cDNA yield measurement employed Cary Eclipse fluorometer (Varian, Palo Alto, USA) equipped with 40 µl cuvette. MMLV reverse transcriptase RNaseH$^-$ Point Mutant (Promega, Madison, USA) or Transcriptor (Roche Diagnostics, Mannheim, Germany; similar data were obtained but the yield was generally lower) carried out the RT reaction with random hexanucleotides (dN_6) or anchored oligo-dT primer ($dT_{23}dV$) essentially according to the manufacturer's instructions.

A 5 µl aliquot from the RT reaction mixture was taken before the reaction started, it was supplied with 15 µl of 4/3x alkaline hydrolysis solution (133 mM NaOH and 1.33 mM EDTA) and incubated at 70°C for 20 min. The mixture was neutralized by adding 6 µl 0.5 M Tris-HCl (pH = 6.4). After the RT had finished, another 5 µl were taken and treated in the same way. 5.2 µl of each hydrolyzed sample (corresponding to 1 µl of the RT reaction) with 0.2 µl RiboGreen solution from RiboGreen RNA Quantitation Kit (Invitrogen, Carlsbad, USA) and 34.6 µl water were submitted to the fluorometer set to 500/520 nm excitation/emission wavelength. The net cDNA-RiboGreen fluorescence (signal from the sample taken after RT minus that of the pre-RT aliquot) divided by the net value of a standard (or calibrator) sample was designated relative cDNA fluorescence and it was proportional to total cDNA yield. The relative cDNA fluorescence was employed as an alternative standardization factor for qRT-PCR.

The techniques not described in detail followed essentially the standard laboratory procedures ().

4.1.11 Tobacco plants and drought treatment

SAG12::ZOG1 and *35S::ZOG1* tobacco lines were a gift from prof. D. Mok, Oregon State University, Corvallis, USA (for detailed information see). Wild-type tobacco (*Nicotiana tabacum* L. cv. Wisconsin 38) and transgenic plants were grown in soil in a growth chamber (SANYO MLR 350H, Osaka, Japan) at a 16/8 h photoperiod (light/dark) at 130 □mol m^{-2} s^{-1}

light intensity, day/night temperature of 25/23°C and relative humidity ca 80%. After 8.5 weeks (wild type and *35S::ZOG1*) or 7 weeks (*SAG12::ZOG1*), half of the plants were transferred into another growth chamber of the same type with the same light and temperature regime, but relative humidity decreased to 35%. Those plants were not watered for 7 days. Leaf samples were collected after 1, 4 and 7 daysand always after removal of the main vein and immediately frozen in liquid nitrogen. Relative water content (RWC) in the leaves was calculated as follows: RWC = (fresh weight – dry weight) / (water saturated weight – dry weight). Water saturation was achieved by letting the leaf absorb water from a beaker.

AGI code	Gene name	Primer sequence	Genbank	Citation
0	*UVH1*	ATC ggT CTg TTT ggT TAT TTg C	AF089003	
		CCT Tgg gAg TCT TTT gTT TCT TC		
0	*NTH1*	CCT gAA AAC Tgg gTC gAA gTg CTC	AJ272248	
		gCA ACC CAT TCT TCC TTT ggA AgC		
0	*BI-1*	gAT gCg TTC TCT TCC TTC TTC g	AB025927	
		CTC CTT TTC TTC TTC TTC TCT TC		
0	*UBC2*	gAC TCC AgC gAg gAA gAg	Y13031	
		Cgg CAg TCC AgC TTT gTT C		
0	*RAD23.*	TTT CgT gAA gAC TCT CAg Tgg	BAB09359	
		ggC TgT gCT AgA ggA TTg g		
0	*RAD50*	gAT CTg AAg AAA ggg gAg Cg	AF168748	
		CAA AgA TCT CTT ggg CCT Cg		
0	*Rad51*	CgT CgT TgA ggA AAg gAA gAg CA	U43528	
		CCT AAC TAC AAA AAC Cgg Tg		
0	*PR1-like*	gTg ggT gCT CTT gTT gTT C	AC005398	
		gCA TAg TTC CCC ggA gg		
0	*ARP*	TTA CCC ATT ggA CCA CAC C	Z49776	
		gTT CCC ATC CCT TCT CTC C		
0	*PARP*	gTC AgC ATg ATg gCA CAg C	Z48243	
		Agg ATC CTT AgT gCT TgT AgT TgA AT		
0	*AP-endon.*	gAg CTC CgC CAT ATA TgT C	X76912	
		ggC AAT gAT CAC TTC CAT Tg		
0	*PolL*	gTg TTC CCT CTA CTA gCg	AJ289628	
		gAg ATT CCT CTC gTg Tgg		
0	*MIM*	AAT gAC TCT gTC CCC TCC AAA Tg	AF120932	
		CTg ggT Cgg gTT CgA TTC TgA g		
0	*LIG4*	CTC TAg CAg Agg AAA ATg TgC	AF233527	
		CAT CCC ggT AgC TgA CTT TC		
0	*UBQ3*	CAA TCT CTC CCA AAg CCT AAA g	L05363	
		TCg ACT CCT TTT gAA TgT TgT Ag		
0	*MAG*	ATg AAA ACg CCg gCT CgT Cg	X76169	
		gTg TTC TTA ggA gCA CTT ATC C		
0	*MPK4*	gTC ggC ggA gAg TTg TTT C	AY040031	
		gAT TgA ACT TgA CTg TTT CAC gg		
0	*RAD17*	CAA gTT CCg TTC gTC TTC C	AB030250	
		CAC TCA TgC ACT gTg ACT CC		

0	*SNI1*	CCA CCg gAg CTA AAg ACA g	AF169596	
		CCT TgT gCA ATT CTA CCA AAC		
0	*NPR1*	gAT CTT gCC gAT gTC AAC C	U76707	
		gAA TCg TTT CCC gAg TTC C		
0	*MRE11*	CTC AgA TgT CTg gTT TAg CTC	AJ243822	
		gCT CCA TAg TTC CTT gTA ACC		
0	*PARP-1*	gAg gAT gAA AgT ggC AAT Cg	AJ131705	
		gCT TAC AAT gTC CAA Cgg gA		
0	*RAD23*	CCC AgT CTT CAC CTg TTC C	AY034912	
		CAT TTC TTg CTC ggg TTg		
0	*OGG*	gCg Agg ATT ACg AAg ATg g	AJ277400	
		TTC CAA ggT CAA gTT TCT gg		
0	*UBC13*	CCg CCA TgA ACT CAC AAg	U33758	
		CgC gAT TAA AAg ATg CAA g		
0	*UBC3*	TgA TgT ggg ATT TCA AgA gAC	L19352	
		TgC TCC ACg ATC TCg ATT AC		
0	*CPR5*	CgC TCT ACC ATA gAC ggA Tg	AY033229	
		ggC AgA TgA Agg TgT TgA Ag		
0	*zf-PARP*	ggA CAT Tgg Tgg ATT TTC Ag	NP_188107	
		TTg CCg CTT gTT CTT CC		
0	*PMS1*	AAg ggT TTC AgA ggA CAA TC	AY047228	
		Cgg gAg gAg CAC TTg g		
0	*UVH3*	TTg TTg gAT ggg gAT gAC	AF312711	
		TgA Cgg AgA CTg gCA AAg		
0	*ERCC1*	gCC AAA CgC ATC AAA CTg	AY050335	
		TgA AAg gCT CgT ggA ATg		
0	*DRT111*	AAT CAg CAC CTT CgC ATC	AY045859	
		CCA CAC TCg CCT CCT ACC		
0	*TEB*	CgT TAT gCT gCT ggT gAA g	CAA18591.1	
		CCg CCA AAg Tgg ATg TAg		
0	*LIG1*	gTT gTT gCT TTC gAC Agg gAg	X97924	
		TCA CTT gAg AgC gAA gAC Tgg		
Sequencing primers				
M13-forward		gTA AAA CgA Cgg CCA gTg		
M13-reverse		gCT ATg ACC ATg ATT ACg C		

Table 1. List of PCR primers used for cloning and sequencing cDNA fragments used on the macroarray. If the primer sequences were adopted from a published work, the respective citations are given. Some of the other primers were designed by K. J. Angelis.

AGI code	Gene name	Primer sequence	C [µM]	T_A [°C]	Genbank accession
Alkylation stress in *Arabidopsis*					
At1g13320	*PDF2*	TAA CgT ggC CAA AAT gA TgC	0.2	58	
		gTT CTC CAC AAC CgC TTg gT	0.2		
At4g05320	*UBQ10*	ggCCTTgTATAATCCCTgATgAATAAg	0.5	58	
		AAAgAgATAACAggAACggAAACATAgT	0.5		
At4g34160	*Cyclin*	TTT CgT TCg TAg ACC ACA TT	0.2	58	NM_119579
		gCT gCg gCAA CTA CTg AT	0.2		
At1g75800	*TL1*	ACg AgT TTA gAC ggC AgA	0.2	58	NM_106230
		gAT gAT TTT TgA CTg gTg TT	0.2		
At3g28910	*Myb30*	ggg AgT TCA AgA TCA TAA	0.2	58	NM_113812
		TCT TTT CTT CAT gTT CTg T	0.2		
At1g02930	*GST1*	CAA CCA CgA AAC Agg AgT	0.2	58	AY097392
		CTA TgA TCg CCA TgT CCT T	0.2		
At3g45640	*MPK3*	gAg gAT gCg AAA AgA TAC A	0.2	58	BT000007
		CgA ACT CAA AAg AgA ATg g	0.2		
Drought stress in tobacco					
	ZOG1	ggT CAC CTC AAC CAg T	0.3	58	AF101972
		CgT CTT CTg gAT Tgg g	0.3		
	cig1	CgT gAC CTC CgA TCATTT g	0.3	58	AB046419
		Tgg TCC gTT ggT ATT AAg gC	0.3		
	CRK1	TgT ggT Cgg ggA CTA ATg	0.3	58	AF302082
		TCg TTT gAg ATg gCg TTg	0.3		
	NtERD10B	ATg gAC AAg gCg gAA gAA g	0.3	58	AB049336
		gTT gTT gCA gTT gAA TgA gT	0.3		
	Tac9	CTA TTC TCC gCT TTg gAC TTg gCA	0.3	60	X69885
		Agg ACC TCA ggA CAA Cgg AAA Cg	0.3		
	18S rRNA	CCT CCA ATg gAT CCT CgT TA	0.3	64	AJ236016
		AAA Cgg CTA CCA CAT CCA Ag	0.3		
	L25	TgC AAT gAA gAA gAT TgA ggA CAA CA	0.3	60	L18908
		CCATTCAAgTgTATCTAgTAACTCAAATCCAAg	0.3		

Table 2. List of the PCR primers used for quantitative PCR. The *Tac9* and *L25* primer sequences were adopted from , those for PDF2 and UBQ10 from . Some primers were designed by Helena Štorchová.

5 Results

5.1 Methylation stress in *Arabidopsis thaliana* cells

To study the transcriptional response of thale cress (*Arabidopsis thaliana*) to methylation stress exerted by MMS, a suspension culture was employed. This system should be more homogeneous than a whole plant consisting of many different cell types. As the plan was to asses the transcript abundances at several time points, it was also advantageous that all the cells were exposed to the compound at the same time. On the other hand, the metabolism and cell regulations might differ between the suspension culture and real tissue cells.

Three different treatment regimes were chosen:

1. High concentration – 5 mM MMS, which was shown to cause a strong DNA damage () and which is finally lethal for *Arabidopsis* cells.
2. Low concentration – 0.5 mM MMS, DNA damage still detectable but much lower; if applied to *Arabidopsis* seedlings their development is normal, the plantlets are just slightly smaller ().
3. High concentration with pretreatment (designated as combined treatment) – 0.5 mM MMS for 1 hour, 2 hours of recovery in media without MMS and then 5 mM MMS – a regime based on experiments of that appeared to raise adaptation (and thereby higher resistance to DNA damage) in *Vicia faba* roots.

As the use of whole-genome microarrays for my experiments would be too expensive, a custom macroarray was designed, carrying 376 *Arabidopsis thaliana* cDNAs in duplicates. Most of them originated from the EST collection of Max-Planck-Institute for Plant Breeding, Cologne, Germany, 34 of them were cDNA fragments cloned for this reason. The cDNAs were selected based on their known or assumed functions (by sequence similarity) or on their expression profiles observed in published experiments. The following terms describe principal functional groups of genes that were sought in the collection:

- DNA repair
- Systemic acquired resistance, pathogen resistance
- Apoptosis, hypersensitive response
- Antioxidants
- Detoxification
- Cell cycle
- Signal cascades
- Transcription factors
- Proteolysis
- Secondary metabolism
- Published interesting expression profiles

For a complete list and other details, see the supplementary file SupplementA.doc

5.1.1 Gene expression data processing

Gene expression data sets are often analyzed for the presence of groups of transcripts that possess similar changes in abundance across different cell types, developmental stages, genotypes, treatments etc. Number of methods have been suggested to search for such groups, with many variants. To perform cluster analysis of the data obtained from the experiments with MMS-treated *Arabidopsis* cells, various algorithms were applied that are implemented in web-based tools (such as EPCLUST, http://www.bioinf.ebc.ee/EP/EP/EPCLUST/,) as well as in commercial software (for example Vector Expression; Informax, USA). The outcome was however not satisfying. First, the obtained results varied substantially among different algorithms and settings. Even after an extensive search there was no clear indication that any of the procedures would be more appropriate to analyze data that I had acquired than the other ones. Many authors seem to choose the method that gives results supporting best their theories. Second, the results were obviously strongly influenced by random noise – even visually very diverse expression profiles could under certain circumstances cluster together. At least the latter problem should be partially alleviated by taking into account the time-series nature of the data. Most of the well-established clustering algorithms treat the single values as independent measurements. This means a loss of information that could otherwise be used during the analysis. In addition, the whole expression profile is analyzed globally, making it relatively easy for a single outlier to substantially change the outcome for the gene. Methods for analyzing time-series data have been developed and successfully used in various areas of science and business (*e.g.*). However, most of them tend to be better suited for longer series – optimally in the order of tens of observations.

Therefore, there was a need for an alternative dealing with the above-mentioned issues. The aim was to design an algorithm that would comply with the following objectives:

1. Robust against random variations and outlying values.
2. Yield big clusters of transcripts sharing fundamental trends.

To cover the first point, a novel system for describing the shape of the expression profile has been devised, following the signal changes in time. A simple description using a scale consisting of only few levels is likely to be more stable against random noise than numerical values. In addition, the absolute numbers (or ratios of them) derived from the spot signal densitometry are often misleading due to nonlinearity of hybridization and/or cross-hybridization of similar transcripts (*e.g.*). Time-series character of the data allows using signal processing approach not applicable to sets of independent measurements. To deal with outliers, values most probably resulting from a technical problem that can significantly deteriorate the worth of the data, the concept of local trend has been introduced. The direction of the signal change along the profile is assessed for each segment of the curve separately (*i.e.*

between each pair of consecutive values). The set of directions, or local trends, is used as a part of the profile characteristics. Even if a part of the profile is heavily distorted by an outlier the rest of the curve can remain untouched and still qualify the gene in its original relationships.

To address the second point, the numerical values of parameters of the algorithm used to weight the similarities and dissimilarities between the expression profile descriptions were set in a way preserving the ability to differentiate among distinct shapes of the time-series but not hypersensitive to minor differences. On the top of that, the algorithm prefers bigger clusters. The whole procedure can be briefly summarized into following steps:

1) Time-series profiles of the signal (expression) are described in terms of:
   a) Relative values compared to Control – in this case a sample not treated with MMS
   b) Local trends of the curve.
2) The descriptions of all the profiles (*i.e.* for all genes) are compared with each other.
3) Big clusters of similar profile descriptions are formed.

To automate the data processing, an MS Excel macro was written. As it employs a concept of local trends (see bellow) it was named LoTrEC (standing for Local Trend based gene Expression data Clustering). It can be found on the CD submitted as a part of this thesis (LoTrEC.xls). Results obtained from the experiments comprising methyl-methanesulphonate treatment of *Arabidopsis* suspension culture, processed by LoTrEC, are also present on the CD: SupplementB1_high_MMS.xls, SupplementB2_low_MMS.xls and SupplementB3_combined_MMS.xls.

5.1.1.1 Data input

LoTrEC accepts as input a datasheet with one or more columns describing the measured features (*e.g.* genes) and two or more (up to about 100) columns of experimental values (one time-point each). First of the latter is expected to be the Control sample – beginning of the time series. First row serves as a header containing labels of the time points. Each of further rows of the table is considered to contain an expression profile. Theoretically, there could be over 65,000 profiles. However, gene counts in the order of thousands can produce rather long processing times (depending on the hardware), mainly due to steeply growing number of comparisons to be performed.

If repeated measurements are to be evaluated, which is generally preferable, they must be pre-processed first to produce a single value for each time-point and profile combination. The data presented in this work were standardized as described in Material and Methods chapter. To summarize the result, no adjustment of data distribution was applied, the distribution is thus left-skewed (approximately log-normal). Median value for each time-point is close to 1

and almost all the values are positive. For optimal performance of LoTrEC, the input data should follow the characteristics mentioned above. However, the macro could be easily adapted to fit other structures of data.

5.1.1.2 Data filtering

To reduce the influence of impulse noise (isolated outlying signals in the time series), low-pass filtering was applied with window size = 3 and double weight for the central value. That means each value was averaged with the two surrounding time points and it was itself counted twice. Arithmetic mean, usually used for low-pass filtering, is rather sensitive to outliers possessing high values that can occur quite often in this type of experiment. Therefore geometric mean was employed that is also not so much disturbed by non-normal distribution of the data. For most of the data, using the two means produced only a very slight difference but in case of sudden isolated peaks (probable impulse noise) the geometric mean behaved more conservatively (Figure 2A). Unfiltered values were only used for the control samples (which integrated highest numbers of arrays - 4 or 5 - to minimize the interference of random noise). The last time points were only averaged (with double weight) with their preceding neighbors. Similarly, the values from 15 minutes of combined treatment were only averaged with the consequent values (30 min), as the control sample in this case cannot be considered to be preceding (see bellow).

To demonstrate the procedure practically, let there be a set of values corresponding to a time series of signals:

Control – 15min – 30min – 1h – 2h – 3h – 5h – 8h

Low-pass filtering of the data:

$Filt(15min) = (15min^2 * Control * 30min)^{(1/4)}$

$Filt(30min) = (30min^2 * 15min * 1h)^{(1/4)}$

...

$Filt(8h) = (8h^2 * 5h)^{(1/3)}$

The description is then based on a new series of values:

Control – Filt(15min) - Filt(30min) - Filt(1h) - Filt(2h) - Filt(3h) - Filt(5h) - Filt(8h)

Figure 2B illustrates the effect of filtering on a smooth and a devious (possibly noisy) profiles.

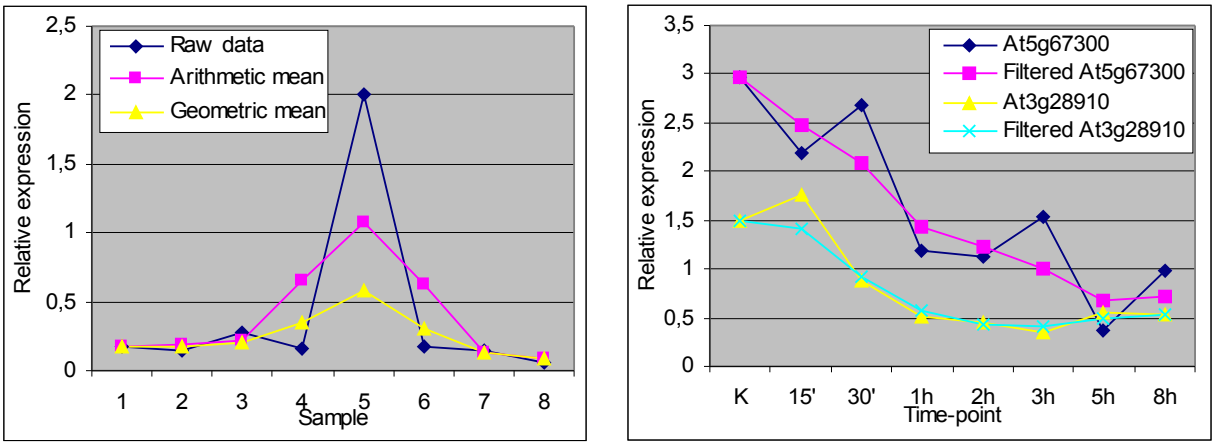


Figure 2. The effect of low-pass filtering applied in the course of describing expression profiles. A – Effect of a single outlying value on the filtered profile. Arithmetic and geometric means are compared. B – Comparison of different shapes of otherwise similar profiles. While relatively smooth profile of the gene *At3g28910* was only slightly affected, a zigzag one of *At5g67300* got straightened to reveal its fundamental trend.

5.1.1.3 Profile description

Each of the filtered values is compared with that of Control sample to obtain its relative position. If the higher of the two values being compared does not exceed a certain threshold (by default set to 0.4 *i.e.* 40% of median of values for all the genes) the result is always "no difference". In other cases, the outcome of the comparison depends on whether they differ at least by a set constant. The difference can be either absolute (relevant mainly for higher values) or by certain-fold. The possible relative positions are: "higher" (than Control), "lower" or "no difference".

Each of the filtered values is compared with the value preceding it in the time series to obtain its local trend ("rising", "falling" or "no trend"; using the procedure described above).

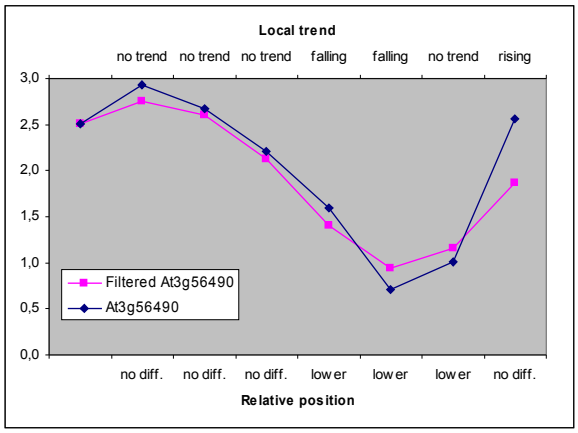
The set of thresholds has been devised empirically to describe visually similar profiles in a similar way.

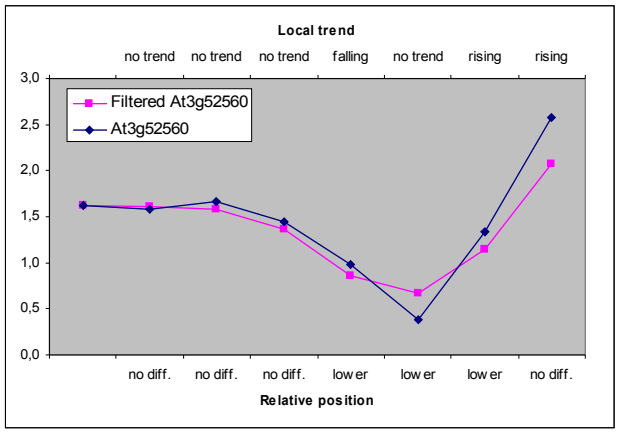
5.1.1.4 Comparing profile descriptions

The produced descriptions are compared pair-wise and their similarities rated. The rating depends on the number of segments (time points) at which the relative positions and/or local trends agree (positive rating) or disagree (negative rating) between the two profiles.

illustrates the process of comparing two described profiles.

All the mutual similarities among the profile descriptions are ordered to form a matrix of similarities.





At3g56490	Local trend:	no trend	no trend	no trend	falling	falling	no trend	rising	Sum
At3g52560	Local trend:	no trend	no trend	no trend	falling	no trend	rising	rising	
	Rating:	0	0	0	5	-7	-7	5	-4
At3g56490	Relative position:	no diff.	no diff.	no diff.	lower	lower	lower	no diff.	
At3g52560	Relative position:	no diff.	no diff.	no diff.	lower	lower	lower	no diff.	
	Rating:	0	0	0	3	3	3	0	9
							Final rating:		**5**

Figure 3. Comparison of expression profiles of two genes with attributed values of local trends and relative positions is performed at each segment separately and the results are rated (default parameter settings are used in this example). The sum of all the ratings then forms the final rating. As final rating for this example exceeds the threshold rating (by default set to 0), expression profiles of genes *At3g56490* and *At3g52560* are similar enough to cluster together (if either of them is the founder of the cluster).

5.1.1.5 Cluster formation

For each profile containing at least one segment with a local trend other than "no trend" a cluster of similar profiles is formed based on the ratings stored in the matrix of similarities. The degree of similarity between profile descriptions necessary to cluster them together depends on threshold rating (see legend of Figure 3). Many of the profiles reside in several clusters at this stage. As bigger clusters are preferable, the cluster with highest number of members is taken and all of its member profiles removed from the initial dataset. In the next round, the biggest cluster of the remaining profiles is taken. This process reiteratively removes clusters of profiles until only single-profile clusters remain. In case of coincident sizes of several clusters, the most compact of them (with highest average rating) is taken. The basic premise explaining why large clusters are preferable is that the expression profiles appear in clusters (or clouds) in the virtual space of similarities (where more similar profiles lie closer to each other). If a fixed diameter part of the space is taken for a cluster then clusters with more members are more likely to contain most of the cloud (or its denser part). In other

words, the profile founding the largest group of similar members is likely to lie near to the center of the cloud.

5.1.1.6 Output

The first sheet of the produced MS Excel workbook serves as an overview of the results. It contains the input data and lists local trends and relative positions for each profile. It is also possible to find there the cluster to which a particular profile belongs. This feature makes it easier to find any profile in the corresponding context. The values of analysis parameters can be found at the bottom of the sheet.

The individual clusters are visualized in separate MS Excel worksheets named by the content of the first column of input table for the founder profile. In a table form, they include all the supplied annotations and the original values for member profiles with color-coded profile descriptions (red or green background for “rising”/”falling” and blue or yellow font for “higher”/”lower”, respectively). For each profile, its similarity rating to the founder profile of the cluster is also given. As a visual impression usually helps to absorb information faster, graphical output in the form of line-charts was also implemented. It illustrates both the non-filtered shapes of the profiles and also the intensities of the signals on the left-side graph. The right-side chart presents ratios of the values to Control. The latter representation is often more transparent but in some cases it can mask absolute expression changes in abundant transcripts.

5.1.1.7 Comparison with a previously published method

There have been several attempts to address the problem of clustering short time-series data (). Recently, published their method based on creating a set of artificial “expression profiles” and distributing the time-series expression data being analyzed among them according to a correlation coefficient. The expected size of every cluster formed just by chance from the analyzed data around each artificial profile is estimated using a Monte Carlo method and compared to the real counts. If calculated probability of the obtained cluster size occurring at random is very low the authors suggest that the enrichment signals a biologically significant process. This approach seems to be more sensitive in discovering nonrandom accumulation of certain expression profiles while reducing false-positive detection of others (that are likely to occur randomly). In addition to publication of the method principle, its Java implementation (Short Time-series Expression data Miner - STEM) has been launched, making it easy to use for anybody. The application adds a statistical analysis of Gene Ontology (GO) categories distribution among the members of particular gene clusters (). STEM was employed as an alternative method of clustering data derived from experiments presented below to compare

the two methods. Unfortunately, the numbers of genes sharing most of the relevant (meaning not very general) annotations are too low to show a statistically significant enrichment.

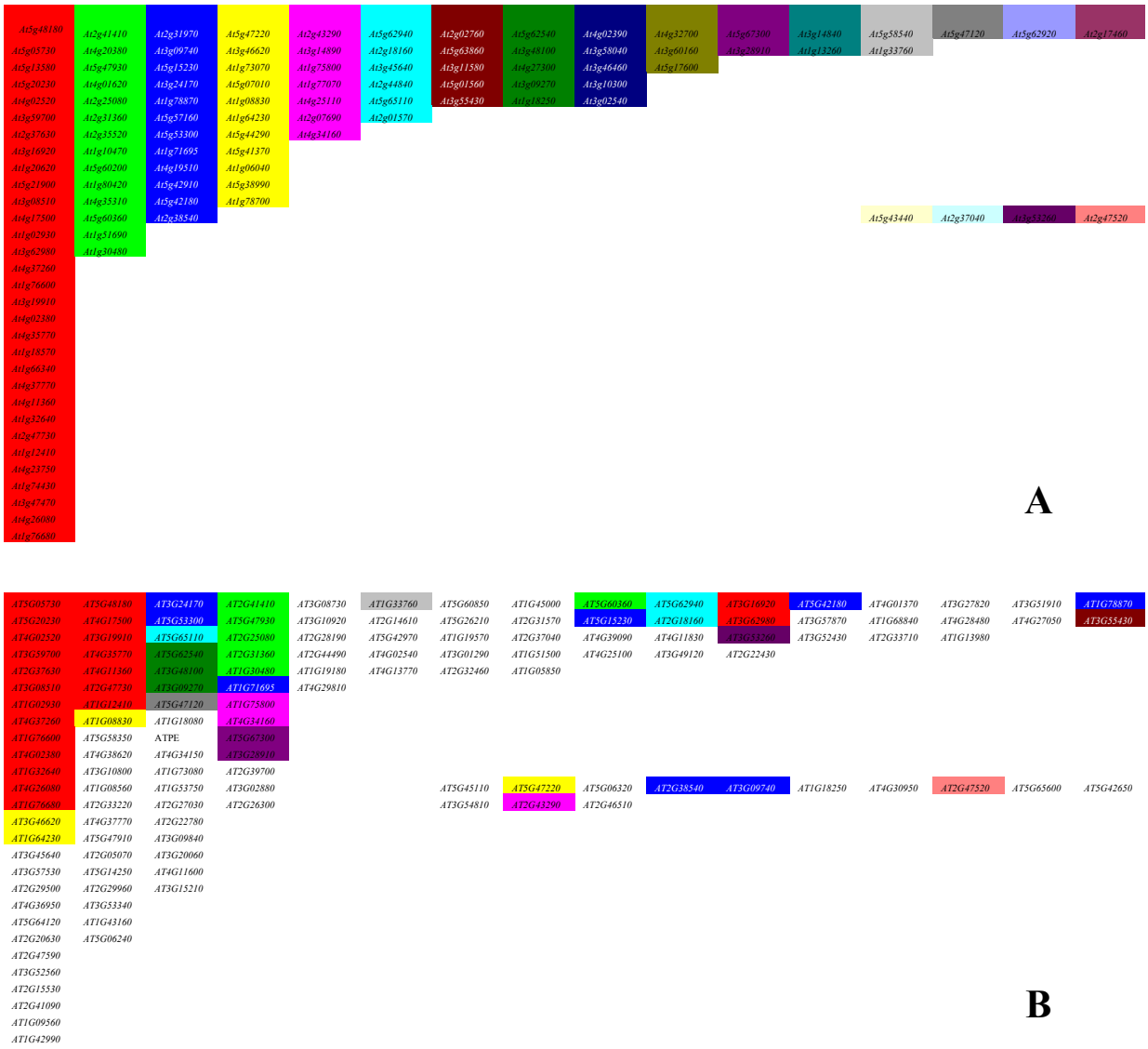


Table 3.Comparison of clusters of expression profiles after high MMS concentration treatment created by LoTrEC (A) and STEM (B). Color coding represents belonging of the profile (gene) to a particular cluster according to A. Among the most populated clusters in A, expression generally increases in clusters 1, 4, 6 and 9 while decreases in clusters 2, 3, 5, 7 and 8 (all from left; see SupplementB1_high_MMS.xls). For practical reasons, the smallest clusters form second rows. The table serves rather as an overview of the results. For detailed inspection, it is recommended to open it in MS Excel as the supplementary file SupplementC1_high_MMS.xls

To compare clustering results of the presented procedure to those obtained using STEM, genes were color-coded based on the clusters they belong to after LoTrEC analysis (Table 3, Table 4 and Table 5). The presence of groups of genes with the same color in clusters

produced by STEM means that both of the methods rank expression profiles of these genes as similar.

The results are therefore similar, although the two algorithms make use of different similarity measures (correlation coefficient in STEM versus local similarity of trends and relative positions). Some clusters appear split into two or more parts. However, closer look reveals that the sections of a particular color tend to group together with members of certain other clusters (represented by certain different colors) according to the STEM procedure. For an example, see red, light green and yellow sections in Table 4B. The corresponding clusters share the general trends. These are, however, too loosely specified to define a common cluster in any of the programs that each emphasize another set of differences among the profiles. The genes with white background in the Table 3, Table 4 and Table 5 were not detected as differentially expressed by LoTrEC. On the other hand, there are color-coded genes not found in the STEM selection.

The combined MMS treatment is specific in several aspects. There are two controls for it: the common Control not treated with any concentration of MMS and a pretreated control (C-Control) which was subjected to the low dose for 1 hour, let to recover for 2 hours (together with all the combined treatment samples) and harvested after another 2 hours without MMS. Thus it cannot be counted as the beginning of the time-series but it rather illustrates the state of gene-expression in the cells if only the preconditioning were applied. It is *de facto* a branching of the time-series. Therefore, it is not allowed to form a trend and is only used for relative position assessment (compared to Control). For the analysis the of combined MMS treatment results, LoTrEC was modified to handle the data in this way. Using STEM, the full series of samples (Control, C-Control and the samples treated with both the concentrations) was analyzed. However, it is not completely fair as STEM assumes pure time-series data. Therefore, just the time-series (without the controls) were also submitted. As expected, the full series clustering results are more similar to clusters created by LoTrEC which means less cluster fragmentation in Table 5C compared to B. However, both of the STEM options completely missed the third largest cluster (founded by *RAD51*, *At5g20850*; blue in Table 5A) of combined treatment expression profiles which is strongly enriched with DNA-repair associated genes (10 out of 22; only 7 of the 10 genes however were entitled with the corresponding GO-term). Most of the genes do not qualify for the clustering and the rest of them (5 profiles) span 2 (full data series) or 3 STEM clusters. The DNA-repair genes are often missed in gene expression analyses because their expression in not-induced cells tends to be rather low and their induction is typically temporally restrained. Another cluster with predominating DNA-repair genes, 11th from the left in combined treatment (founded by DNA polymerase kappa homolog, *At1g47980*; violet; 3 out of 4 genes) was also completely missed

by STEM with default settings – only a GPX gene (the only one without any known function in DNA-repair) came out through the initial selection of differentially expressed genes. Although under a specific parameter settings some of the missed profiles would be acceptable for STEM, this phenomenon illustrates the global tendency to neglect data of transcripts that are rare in some (or most) samples. Such data were included in the presented analysis although it must be noted that low expression values are more influenced by random fluctuations and various artifacts. However, there are whole groups of proteins that are not necessary in a big amount in the cells, or they even could be harmful (*e.g.* homologs of DNA repair gene AlkA can catalyze depurination of intact DNA –). Their transcripts are therefore only accessible for DNA-array analysis if we accept this kind of risk. Anyway, it is still possible to remove these "less reliable" profiles by custom setting of parameters of the procedure, increasing the minimum acceptable signal and/or minimum differences for detection of differential expression. Default parameters were used in both programs.

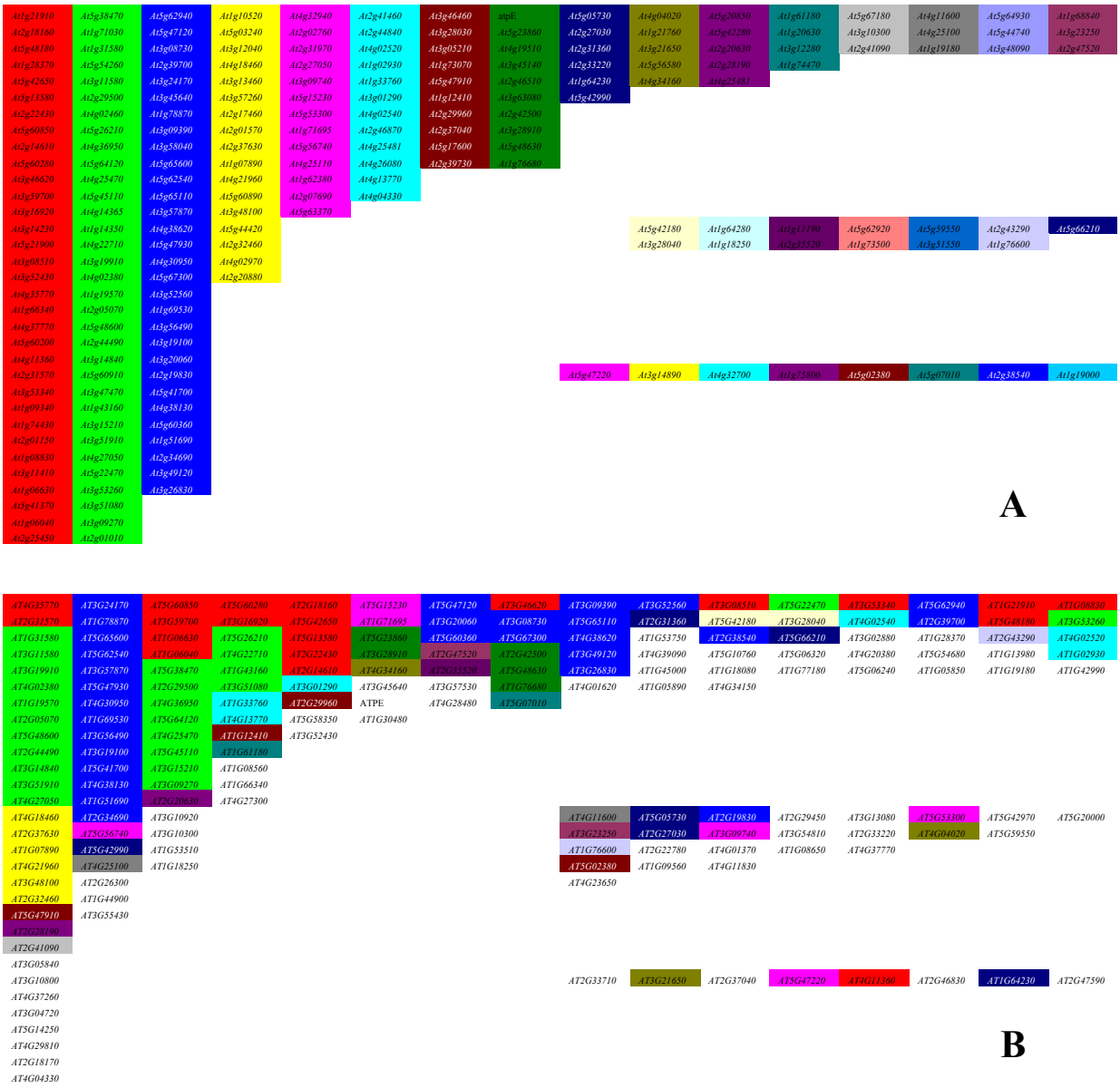
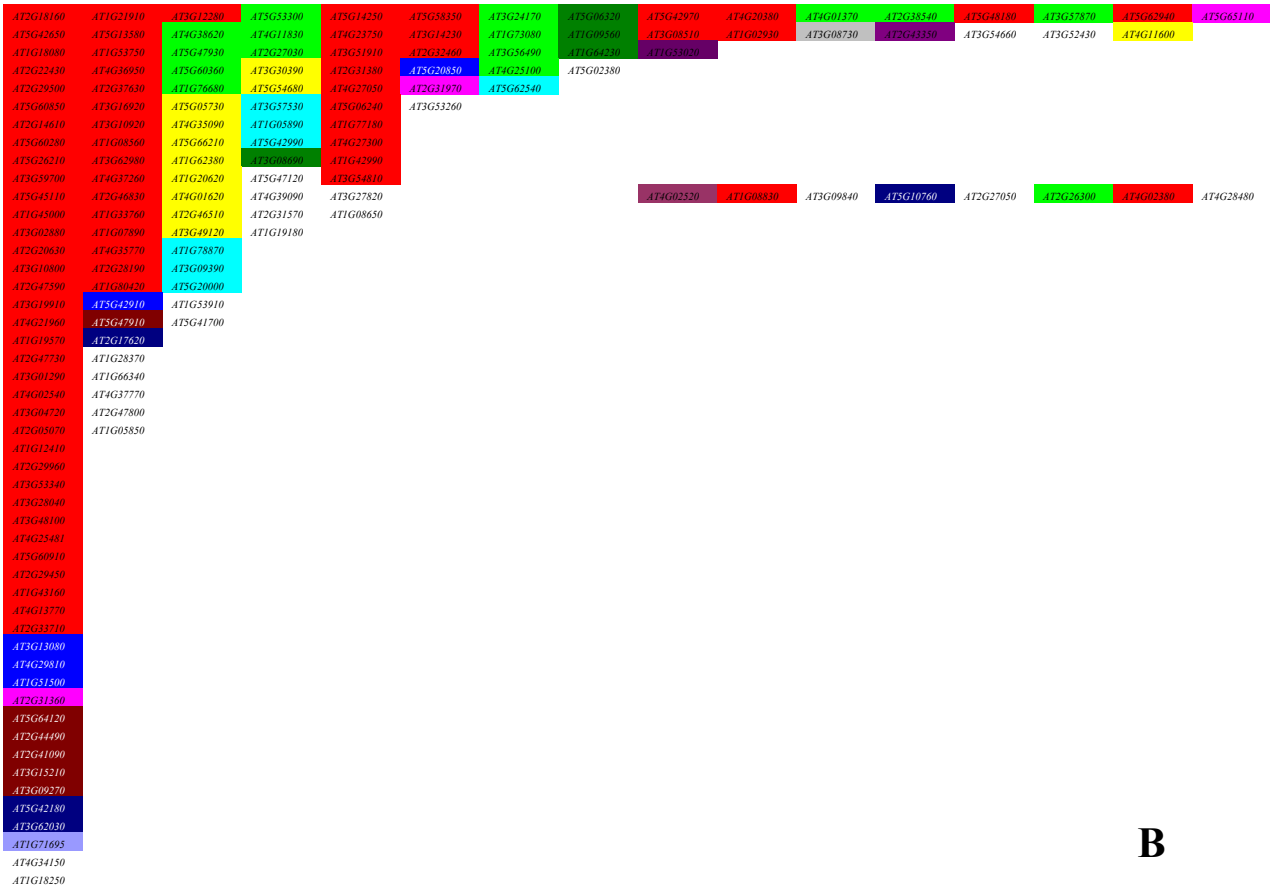
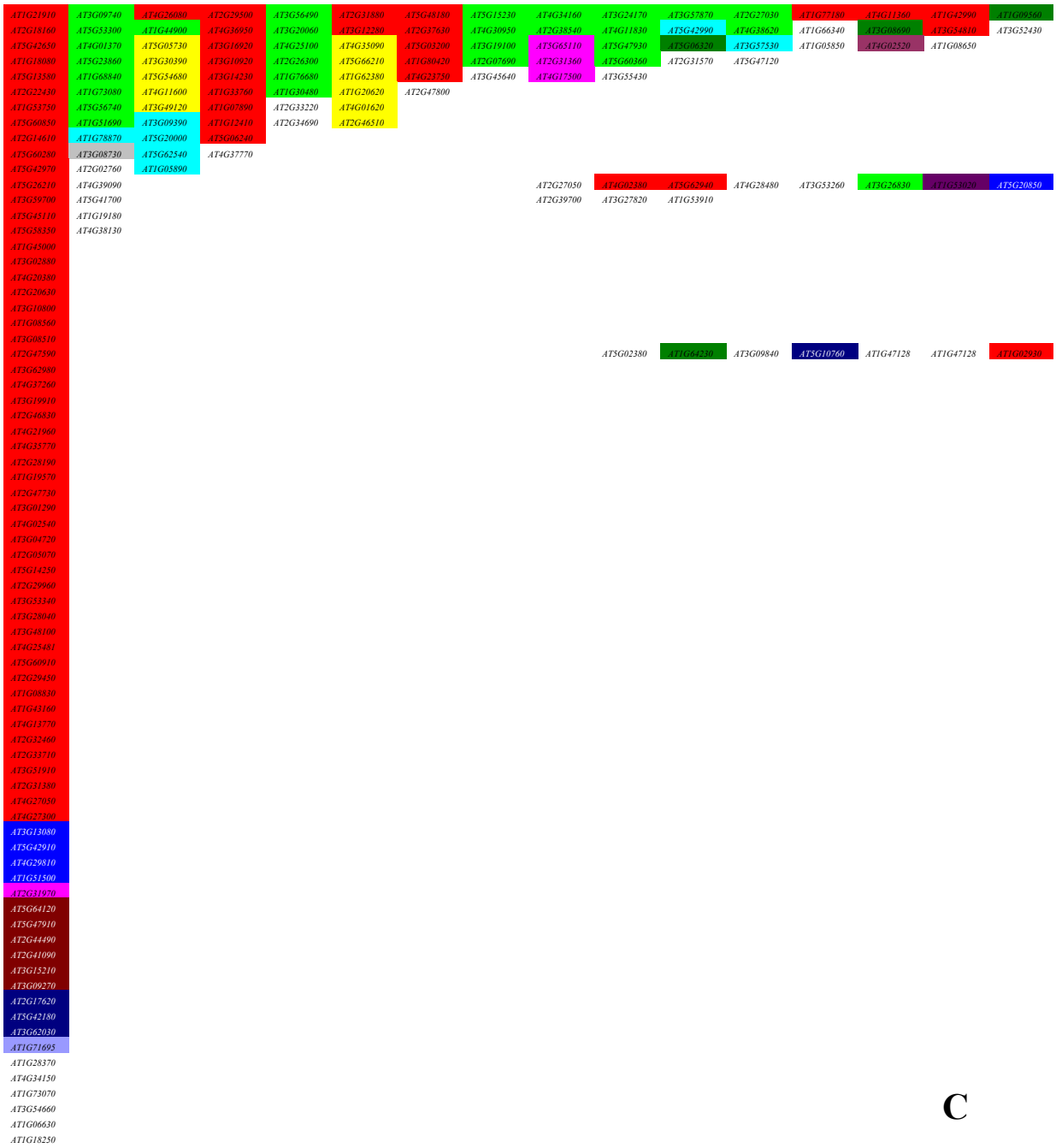


Table 4. Comparison of clusters of expression profiles after low MMS concentration treatment created by LoTrEC (A) and STEM (B). Color coding represents belonging of the profile to a particular cluster according to A. Among the most populated clusters in A, expression generally increases in clusters 1, 2, 4, 6, 7, 9 and 11 while decreases in clusters 3, 5, 8 and 10 (all from left; see SupplementB2_low_MMS.xls). For practical reasons, the smallest clusters form second and third rows. The table serves rather as an overview of the results. For detailed inspection, it is recommended to open it in MS Excel as the supplementary file SupplementC2_low_MMS.xls


At2g31450 At2g46830 At4g04020 At5g41150 At5g05730 At2g31970 At1g78870 At5g64120 At3g08690 At5g10760 At1g31580 At1g21710 At3g46820 At3g08730 At4g08500 At1g71695
At5g62940 At1g33760 At3g09740 At5g38470 At4g35090 At5g56580 At3g09390 At5g47910 At5g06320 At2g17620 At5g58540 At2g43350 At1g19000
At1g49670 At1g07890 At5g15230 At5g20850 At5g66210 At1g12860 At3g57530 At2g44490 At5g19450 At5g42180 At5g47640 At1g47980
At5g63860 At4g21960 At3g24170 At2g14580 At3g46460 At3g61630 At5g20000 At2g41090 At1g09560 At3g62030 At1g16880 At5g22470
At1g21910 At4g35770 At3g45640 At4g02390 At3g30390 At5g47220 At5g62540 At3g15210 At1g64230
At2g18160 At2g28190 At1g69840 At2g41460 At1g62380 At5g65110 At3g63080 At3g09270
At5g48180 At1g19570 At5g53300 At1g10520 At1g63650 At3g61580 At1g05890 At1g18250
At2g44840 At5g03200 At4g01370 At5g61460 At5g54680 At5g26920 At4g00330
At1g28370 At5g03200 At5g23860 At4g18460 At1g20620 At2g31360 At5g42990
At5g42650 At2g28510 At1g68840 At4g02460 At4g01620 At4g17500
At1g18080 At5g07010 At1g73080 At3g13080 At2g46510 At5g48600
At5g13580 At1g66340 At4g11830 At3g05210 At4g11600 At3g61810
At1g71030 At4g37770 At2g27030 At5g42910 At3g49120
At2g22430 At5g60200 At5g56740 At4g25470
At3g11580 At1g80420 At1g75800 At2g46870
At4g34150 At4g11360 At3g57870 At4g29810
At2g29500 At2g47730 At4g25110 At1g31500
At1g53750 At3g01290 At4g38620 At5g43470
At5g60850 At4g02540 At1g72930 At5g44740
At2g14610 At3g04720 At5g47930 At2g39730
At3g21270 At2g05070 At4g30950 At2g01010
At5g60280 At3g60160 At2g25080
At3g28030 At3g12280 At5g67300
At5g42970 At5g14250 At2g33220
At5g26210 At1g12410 At3g56490
At3g57260 At2g29960 At3g19100
At2g31880 At3g53340 At2g07690
At4g36950 At3g28040 At3g20060
At2g17460 At3g48100 At4g34160
At3g59700 At4g23750 At3g20770
At2g01570 At1g72610 At3g16720
At5g45110 At1g09340 At3g51550
At5g58350 At4g25481 At2g38540
At1g45000 At4g25481 At4g25100
At3g02880 At3g10300 At2g26300
At4g20380 At5g60910 At1g13980
At1g14350 At2g29450 At1g19180
At1g14350 At3g47470 At5g48630
At2g37630 At4g26080 At5g60360
At2g20630 At1g08830 At1g44900
At3g16920 At1g43160 At1g76680
At3g10800 At4g13770 At1g51690
At3g10920 At2g32460 At1g30480
At1g08560 At2g33710 At3g26830
At4g22710 At3g51910
At3g14230 At2g31380
At5g21900 At4g27050
At5g01880 At5g17600
At3g08510 At1g05850
At2g47590 At5g06240
At1g02930 At1g77180
At3g62980 At2g20880
At4g37260 At3g11410
At3g10910 At4g27300
At4g02380 At1g42990
At1g73070 At3g54810
A


AT2G18160 AT1G21910 AT3G12280 AT5G53300 AT5G14250 AT5G58350 AT3G24170 AT5G06320 AT5G42970 AT4G20380 AT4G01370 AT2G38540 AT5G48180 AT3G57870 AT5G62940 AT5G65110
AT5G42650 AT5G13580 AT4G38620 AT4G11830 AT4G23750 AT3G14230 AT1G73080 AT1G09560 AT3G08510 AT1G02930 AT3G08730 AT2G43350 AT3G54660 AT3G52430 AT4G11600
AT1G18080 AT1G53750 AT5G47930 AT2G27030 AT3G51910 AT2G32460 AT3G56490 AT1G64230 AT1G53020
AT2G22430 AT4G36950 AT5G60360 AT3G30390 AT2G31380 AT5G20850 AT4G25100 AT5G02380
AT2G29500 AT2G37630 AT1G76680 AT5G54680 AT4G27050 AT2G31970 AT5G62540
AT5G60850 AT3G16920 AT5G05730 AT3G57530 AT5G06240 AT3G53260
AT2G14610 AT3G10920 AT4G35090 AT1G05890 AT1G77180
AT5G60280 AT1G08560 AT5G66210 AT5G42990 AT4G27300
AT5G26210 AT3G62980 AT1G62380 AT3G08690 AT1G42990
AT3G59700 AT4G37260 AT1G20620 AT5G47120 AT5G54810
AT5G45110 AT2G46830 AT4G01620 AT4G39090 AT3G27820 AT4G02520 AT1G08830 AT3G09840 AT5G10760 AT2G27050 AT2G26300 AT4G02380 AT4G28480
AT1G45000 AT1G33760 AT2G46510 AT2G31570 AT1G08650
AT3G02880 AT1G07890 AT3G49120 AT1G19180
AT2G20630 AT4G35770 AT1G78870
AT3G10800 AT2G28190 AT3G09390
AT2G47590 AT1G80420 AT5G20000
AT3G19910 AT5G42910 AT1G53910
AT4G21960 AT5G47910 AT5G41700
AT1G19570 AT2G17620
AT2G47730 AT1G28370
AT3G01290 AT1G66340
AT4G02540 AT4G37770
AT3G04720 AT2G47800
AT2G05070 AT1G05850
AT1G12410
AT2G29960
AT3G53340
AT3G28040
AT3G48100
AT4G25481
AT5G60910
AT2G29450
AT1G43160
AT4G13770
AT2G33710
AT3G13080
AT4G29810
AT1G51500
AT2G31360
AT5G64120
AT2G44490
AT2G41090
AT3G15210
AT3G09270
AT5G42180
AT3G62030
AT1G71695
AT4G34150
AT1G18250
B



C

Table 5. Comparison of clusters of expression profiles after combined MMS treatment created by LoTrEC (A) and STEM, using only the time-series without controls (B) or the full series (C). Color coding represents belonging of the profile to a particular cluster according to A. Among the most populated clusters in A, expression generally increases in clusters 1, 3, 4, and 7 while decreases in clusters 2, 5, 6 and 8 (all from left; see SupplementB1_c_MMS.xls). For practical reasons, the smallest clusters form second and third rows. In A, the first cluster occupies 2 columns. The table serves rather as an overview of the results. For detailed inspection, it is recommended to open it in MS Excel as the supplementary files SupplementC3_c_MMS1.xls (B) and SupplementC4_c_MMS2.xls (C).

5.1.2 Analysis of cDNA-array results

The algorithm described above selected the differentially expressed genes from the set present on the cDNA array and grouped them according to similarities of the time-course of their transcript abundances. The result comprised 20, 31 and 21 clusters (high concentration of MMS, low and combined doses, respectively). However, 9 – 15 of them (corresponding to 43 to 50% of the cluster numbers) contained less than 3 allocated profiles. For further analysis, it was thus useful to merge several clusters together to form larger groups possessing rather a common fundamental tendency and containing large enough numbers of genes.
The following pages will concentrate on groups of transcripts defined by GO annotations that appear to be more represented among either the up-regulated or the down-regulated lots. Some of these enrichments were statistically significant (based on tests performed by FatiGO+ accessible from Babelomics web-site,) but most of the smaller (more narrowly specified) groups were not due to insufficient number of representatives.
Functional groups of *Arabidopsis* genes affected by MMS treatment

5.1.2.1 DNA repair

The DNA repair genes induced or repressed by MMS belong to various repair pathways. These subgroups are also mentioned among the GO terms but the number of the transcripts corresponding to each of them is very low. There are genes involved in base excision repair (BER), nucleotide excision repair (NER), mismatch repair, non-homologous end-joining (NHEJ) and homologous recombination. Interestingly, the high MMS concentration only induced a single transcript and 4 were reduced over the period of 8 hours of experiment. A strikingly different regulation takes place in cells treated with low concentration of MMS and with combined dose. They increased abundance of 9 and 13 transcripts, respectively, compared to 1 or 2 decreased ones (Table 6).The expression profiles of majority of the transcripts start at very low values but rise in later time-points.
The only DNA-repair gene that was detected as induced by high MMS dose was one of the *Rad23* homologs present in *Arabidopsis* genome (*At3g02540*). *Rad23* genes generally have a part in NER, the pathway with broadest spectrum of substrates . On the contrary, the other MMS treatments (low and combined doses) resulted in activation of a wide range of DNA-repair genes representing various pathways. While *PMS1* (*At4g02460*) is involved in mismatch-repair, *RAD51* (*At5g20850*) is a key player in homologous recombination. Induction was observed of the transcripts of 3 and 5 genes playing a role in BER (low and combined doses, respectively). BER is the preferred pathway for removing methylated bases

and it is thus not surprising to find some of BER members to be affected by a DNA-methylating chemical. Three DNA-glycosylases removing modified bases were induced by the treatment – *3-methyladenine-DNA-glycosylase* (*MAG*, *At3g12040*) by low MMS and endonuclease-three homolog 1 (*NTH1*, *At2g31450*) and *8-oxoguanine-DNA-glycosylase* (*OGG*, *At1g21710*) by the combined dose. The repair *DNA-polymerase λ* (*At1g10520*) and nuclease *UVH3* (*At3g28030*) were induced by low and combined MMS treatment. *UVH3* is a homolog of *RAD2* (*Saccharomyces cerevisiae*) and mammalian *XPG* which are endonucleases involved in both BER and NER. *Arabidopsis uvh3* mutant is indeed hypersensitive to ultraviolet and ionizing radiations as well as to hydrogen peroxide. It also displays a premature senescence phenotype ().

Involvement of NER in repair of DNA-damage inflicted by the low and combined MMS treatments is documented by an increase in abundances of 3 and 4 transcripts. They include in both cases another homolog of *Rad23* (*At5g38470*), *ERCC1* (*At3g05210*, acting as an endonuclease in a complex with *UVH1-XPF-At5g41150* which was activated by the combined dose as well) and *UVH3*.

A TLS gene *AtRad30*, coding for (*At5g44740*), a putative homolog of *DNA-polymerase η* (eta) , yeast *Rad30* and *XP-V*, also designated "*DNA damage induced protein P*" was also induced by the two treatments mentioned above.

The only mRNA of a DNA-repair gene that was consistently repressed by MMS was *Rad50* (*At2g31970*), a subunit of DNA-end processing complex (together with *Mre11*) employed during NHEJ. Transcript abundance of *DRT111* (*DNA-damage-Repair-Toleration 111*), involved in resolving Holiday structures during homologous recombination (), decreased when high and combined doses of MMS were applied.

A homolog of *DNA-polymerase θ* (theta; *PolQ*, *At4g32700*) is one of the transcripts connected with DNA-repair that were cloned (as a fragment of cDNA) for the purpose of the array. It is homologous to *Drosophila melanogaster mus308* and mammalian *PolQ* that were demonstrated to be important for DNA repair and genome stability (*chaos1* mouse mutant, an abbreviation of *Chromosomal Aberrations Occurring Spontaneously 1*) – {Leonhardt 1993; Shima 2004}. Arabidopsi*s* plants mutant in the *PolQ*-homolog display increased sensitivity to genotoxins mitomycin C and MMS and also an abnormal cell division pattern in meristems. It was designated *tebichi* (*teb*; {Inagaki 2006}). The abundance of *teb* transcript gradually increases from early after the onset of low MMS dose treatment as well as in the later time-points of the treatment with high MMS concentration. However, its non-induced expression is relatively very low and the profile could be missed if a classical approach were used. This is what happened in published Affymetrix data (Genevestigator,

https://www.genevestigator.ethz.ch/at/;) – the transcript was flagged absent in many experiments and for ratio-based methods it is thus effectively lost. Although *teb* gene was shown to be connected with DNA-repair in *Arabidopsis* as well as in other organisms, the corresponding GO-term was not present in the annotation of the gene (as of December 2006). It might be caused by the fact that *teb* function was revealed first recently and the annotation still needs to be updated.

				At5g20850	*RAD51*
				At3g05210	*ERCC1*
				At1g21710	*OGG*
				At5g44740	*AtRad30*
		At5g20850	*RAD51*	*At3g28030*	*UVH3*
		At3g05210	*ERCC1*	*At1g08130*	*LIG1*
		At5g44740	*AtRad30*	*At5g41150*	*UVH1*
		At3g28030	*UVH3*	*At5g38470*	*RAD23*
		At5g54260	*MRE11*	*At4g02460*	*PMS1*
		At5g38470	*RAD23*	*At2g47590*	*Photolyase (PHR2)*
		At3g12040	*MAG*	*At2g31450*	*NTH1*
		At4g02460	*PMS1*	*At1g80420*	*XRCC1*
At3g02540	*RAD23*	*At1g10520*	*PolL*	*At1g10520*	*PolL*
At5g57160	*LIG4*	*At2g31970*	*RAD50*	*At1g30480*	*DRT111*
At1g80420	*XRCC1*			*At2g31970*	*RAD50*
At1g30480	*DRT111*				
At2g31970	*RAD50*				

Table 6. Comparison of sets of mRNAs GO-annotated "DNA repair", induced (over the black line) and repressed (under the black line) in treatments with high concentration of MMS (left), low concentration of MMS (middle) and the combined treatment (right).

Poly(ADP-ribose)-polymerase (*PARP*) activity is closely related to DNA-repair, as *PARP* proteins were shown to serve as sensors of some DNA-damage types, binding to exposed molecule ends. Concurrent poly-ADP-ribosylation of various proteins attracts the repair machinery. *PARP* interactions were documented for BER () and double strand break repair pathways (). From the 4 *PARP* genes present on the array, 2 were induced by MMS. *APP* (At4go2390) by the high concentration and *PARP3* (*At5g22470*) by low dose. Combined treatment elevated abundances of both the transcripts. The activation of *Arabidopsis PARP* genes after a genotoxic treatment is in agreement with previously published data of , who stressed the plants using ionizing radiation.

5.1.2.2 Programmed cell death

Programmed cell death (PCD) is another process presumably affected by MMS. It is a long-known fact that extensive or persistent unrepaired DNA-damage leads to PCD in various multicellular organisms. Also in plants, including *Arabidopsis*, genotoxic stress can lead to removal of some affected cells in a regulated and organized way (). Accordingly, MMS was expected to change expression levels of some involved genes. Indeed, several changes were observed (Table 7).

		At5g47910	*RbohD*		
At5g47120	*BI-1*	*At5g64930*	*CPR5*	*At5g47910*	*RbohD*
At3g28910	*Myb30*	*At3g28910*	*Myb30*	No genes	
At4g19510	*protein N-like*	*At4g19510*	*protein N-like*		
At2g35520	*DAD-2*	*At5g47120*	*BI-1*		

Table 7. Comparison of sets of mRNAs GO-annotated "programmed cell death", induced (over the black line) and repressed (under the black line) in treatments with high concentration of MMS (left), low concentration of MMS (middle) and the combined treatment (right).

High MMS dose strongly increased abundance of the transcript for *BI-1* (Bax-Inhibitor-1; *At5g47120*) – a known inhibitor of PCD () and decreased that of *Myb30* (*At3g28910*), a transcription factor important for cell death onset (). This set of changes obviously does not support the theory of cell death triggered by a high level of DNA damage. Although another PCD inhibitor, *DAD-2* (*Defender Against cell Death-2*; *At2g35520*), was repressed, 2 important regulatory transcripts behave in a way clearly contradicting PCD.
Using the low concentration, MMS similarly lowered the expression of *Myb30*, yet also of *BI-1*. This treatment induced expression of *RbohD* (*Respiratory Burst Oxidase Homolog D*; *At5g47910*) that generates reactive oxygen species, contributing thereby to the cell death program activation. After applying the combined dose of MMS, only *RbohD* was up-regulated. Neither of the latter two regimes seems to result in PCD as well.
Caspases are the executive proteases accomplishing PCD (apoptosis) in animal cells. The only plant sequences distantly similar to caspases were designated metacaspases (). However, although these proteases are suspected that they might play a role in plant-type PCD (), it has never been proven (). They are thus not annotated as involved in "programmed cell death". Anyway, transcript abundance of the single metacaspase present on the macroarray (*At4g25110*) decreased after application of any of the MMS treatments.

At1g32640	*RAP1*
At1g76680	*OPR1*
At4g37260	*Myb-related*
At2g37630	*Myb-putative*
At1g74430	*Myb95*
At5g20230	*Pistillata*
At1g66340	*ETR1*
At1g18570	*Myb51*
At5g05730	*ASA1*
At3g53260	*PAL2*
At2g37040	*PAL1*
At3g28910	*Myb30*
At5g67300	*Myb*

At5g60890	*ATR1*
At3g15210	*ERF-4*
At5g42650	*AOS*
At1g71030	*Myb-related*
At1g19570	*DHAR*
At2g37630	*Myb-putative*
At1g74430	*Myb95*
At5g44420	*AFP3*
At1g66340	*ETR1*
At5g05730	*ASA1*
At2g37040	*PAL1*
At3g53260	*PAL2*
At5g64930	*CPR5*
At3g23250	*Y19*
At5g67300	*Myb*
At1g76680	*OPR1*
At3g28910	*Myb30*
At3g45140	*Lox2*
At4g38620	*Myb4*
At4g38130	*HDAC*

At3g15210	*ERF-4*
At4g08500	*MEKK1*
At5g42650	*AOS*
At1g71030	*Myb-related*
At1g19570	*DHAR*
At4g37260	*Myb-related*
At2g37630	*Myb-putative*
At2g46830	*CCA1*
At1g66340	*ETR1*
At5g05730	*ASA1*
At5g67300	*Myb*
At1g76680	*OPR1*
At4g01370	*MPK4*
At4g38620	*Myb4*

Table 8. Comparison of sets of mRNAs GO-annotated "response to wounding", induced (over the black line) and repressed (under the black line) in treatments with high concentration of MMS (left), low concentration of MMS (middle) and the combined treatment (right).

5.1.2.3 Response to wounding

Induction of genes annotated to be involved in response to wounding documents the fact that general stress response also takes place in MMS treated cells (Table 8).

Among these, several enzymes metabolizing aromatic compounds can be found. These include *phenylalanine-ammonia-lyases 1* (*PAL1*, *At2g37040*) and *2* (*PAL2*, *At3g53260*) and *anthranilate-synthase alpha 1* (*ASA1*, *At5g05730*). Interestingly, *ASA1* transcript was detected in an increased amount in the low MMS dose treatment along with *ATR1* (*Altered Tryptophan Regulation 1*, *At5g60890*) – a Myb transcription factor that regulates expression of *ASA1* (). Wound signaling is known to employ jasmonic acid and ethylene (). Indeed, E*thylene-Response Protein 1* (*ETR1*, *At1g66340*) was induced by all the regimes and E*thylene Responsive element binding Factor-4* (*ERF-4*, *At3g15210*) by the low and combined treatments. Also an A*llene Oxide-Synthase* (*AOS*, *At5g42650*), a biosynthetic enzyme of jasmonic acid , exhibited an induction in low MMS and combined treatments. Transcription factors of the Myb-family play an important role in stress signaling including wound response. Of these, MMS induced the expression of *Myb91* (*At2g37630*), *Myb95* (*At1g74430*)

and two other Myb-related genes (*At1g71030* and *At4g37260*) in at least two treatments. *MYBR1* (*MYB-Related protein 1*, *At5g67300*) was repressed in all the samples, *Myb30* (*At3g28910*) and *Myb4* (*At4g38620*) in two of them.

5.1.2.4 Oxidative stress response

Among MMS affected transcripts sorted by enzymatic activity of their protein products (GO-molecular function), groups of glutathione-S-transferases (GSTs) and superoxide-dismutases (SODs) become prominent. There is a large family of GST genes in the *Arabidopsis* genome with a wide-spread spectrum of roles and activities in the cell. Some GSTs are quite specialized while others respond to a broad spectrum of stress conditions (). However, all members of this family that were detected to change their expression after MMS treatment are annotated as involved in "toxin catabolism". Each MMS treatment induced 2 to 4 GST transcripts out of the 5 present on the array: increase in transcript levels of *GST1* (*At1g02930*) and *GST2* (*At4g02520*) appeared in all MMS regimes, *GST6* (*At2g47730*) with high and combined doses and yet another *GST* (*At2g29450*) in the combined treatment only.
SOD proteins are quenchers of very dangerous superoxide anions formed either by a pathological activity of energy metabolism, or as a part of a defense reaction. Also some chemicals catalyze production of reactive oxygen species including superoxide. SODs contain a metal cation as a cofactor. There are three groups of SODs based on the preferred metal and also on their phylogenetic origin. *Arabidopsis* cells contain members of all three existing groups and so does the macroarray. There are two cytoplasmic *Cu,ZnSODs* (*At1g08830* and *At2g28190*), a mitochondrial *MnSOD At3g10920* and a plastid *FeSOD* (*At4g25100*). All of them are induced by MMS but the *FeSOD* which is reduced by either low or combined doses of MMS. High MMS concentration resulted only in *Cu,ZnSOD* (*At1g08830*) activation.

At3g45640	*MAPK3*
At5g20230	*Pistillata*
At1g08830	*Cu,ZnSOD*
At3g53260	*PAL2*
At2g37040	*PAL1*
At1g20620	*Catalase3-like*

At2g29500	*Small HSP*
At1g07890	*APX1*
At2g31570	*GPx*
At1g08830	*Cu,ZnSOD*
At5g64120	*ATP15a*
At5g42180	*PRXR4*
At2g37040	*PAL1*
At3g53260	*PAL2*
At4g21960	*PRXR1*

At2g29500	*Small HSP*
At1g07890	*APX1*
At4g35090	*Catalase2*
At1g20620	*Catalase3-like*
At2g43350	*GPX-putative*
At4g11600	*PHGPX*
At1g08830	*Cu,ZnSOD*
At5g64120	*ATP15a*
At5g42180	*PRXR4*
At4g21960	*PRXR1*

At1g71695	*ATP4a*
At5g42180	*PRXR4*
At2g25080	*ATCHLGPX*

At3g45640	*MAPK3*
At4g11600	*PHGPX*
At3g63080	*GPX-like*

At3g45640	*MPK3*
At2g25080	*ATCHLGPX*
At3g63080	*GPX-like*

At1g71695	*ATP4a*	*At1g71695*	*ATP4a*

Table 9. Comparison of sets of mRNAs GO-annotated "response to oxidative stress", induced (over the black line) and repressed (under the black line) in treatments with high concentration of MMS (left), low concentration of MMS (middle) and the combined treatment (right).

Although SOD proteins play a role in protection of the cells against oxidative stress, only one of them (*Cu,ZnSOD At1g08830*) reflects this fact in its GO annotation. The GO term "response to oxidative stress" is shared also by *catalases 2* (*At4g35090*, induced by combined MMS treatment) and *3* (*At1g20620*, activated by high and combined doses) that catalyze decomposition of hydrogen peroxide, *PAL1* and *PAL2* (up-regulated by high and low MMS) and a group of peroxidases (Table 9). Interestingly, all three peroxidases affected by high MMS concentration (*ATP4a - At1g71695*, *PRXR4 - At5g42180* and chloroplast *AtChlGPX - At5g42180*) were repressed. On the other hand, each of the other treatments induced 5 (including *APX1 - At1g07890*, *PRXR4* and *ATP15a - At5g64120*) and repressed 3 "peroxidase" annotated transcripts (*e.g. ATP4a*). *MPK3* (*Mitogen-activated Proteinkinase 3*, *At3g45640*) that is known to be activated by various stresses () follows an opposite scheme – abundance of its RNA is elevated by high MMS and decreases in the other treatments.

5.1.2.5 Proteolysis

Protein turnover plays a fundamental role in many cell processes (). It is a way of removing inhibitory or unnecessary proteins during cell reprogramming. It also removes damaged and unrepairably denatured protein molecules during stress. For example reports imbalance in DNA repair pathways and consequent increase in killing and mutation rate caused by MMS and UV in yeast (*Saccharomyces cerevisiae*) with a defect in protein turnover. There are various proteases involved in the system. The most general machinery with broadest spectrum of substrates is the proteasome. It is a large protein complex processing polypeptides "marked" by covalent binding of ubiquitin. The specificity of the system is achieved at the level of ubiquitin conjugation to specific targets. The cascade consists of ubiquitin-activating proteins (E1) ubiquitin conjugating enzymes (E2) and ubiquitin ligases (E3) that are either single polypeptides or make up a part of the anaphase promoting complex or Skp1/cullin/F-box protein complex (APC/SCF). In *Arabidopsis*, there are 2 different E1s, over 35 E2s and many F-box proteins (). Some of the mentioned genes are represented on the macroarray, as well as several regulatory proteasome subunits and a group of ubiquitin-independent proteases.

While the search for GO-term "proteolysis" among differentially expressed genes after high concentration of MMS yields 5 transcripts induced and 5 repressed, the other treatments caused mainly down-regulation (Table 10). Some of the affected genes can be found in common, anyway. Majority of the affected genes represent known or putative ubiquitin conjugating enzymes (UBC, E2s). All the MMS treatments increased the abundance of *UBC13* (*At3g46460*) transcript and repressed those of *UBC3* (*At5g62540*) and *UBC10* (*At5g53300*). Most of the other UBC mRNA levels decreased in one or two of the MMS regimes, including *UBC2* (*At2g02760*, high and low MMS). *UBC2* is a homolog of yeast *Rad6*, an E2 enzyme involved in DNA-damage resistance and DNA-repair via translesion synthesis (). However, its function has not been confirmed in plants and *Arabidopsis* genome also lacks a gene coding for a homolog of *Rad18p* that forms a heterodimer with *Rad6p* and is vital for *Rad6* function in yeast.

At3g46460	*UBC13*
At3g58040	*?protein*
At1g64230	*UBC-like*
At3g62980	*TIR1*
At1g12410	*nClpP2*
At2g02760	*UBC2*
At5g60360	*AALP*
At5g62540	*UBC3*
At5g53300	*UBC10*
At4g25110	*Metacaspase*

At3g46460	*UBC13*
At1g64230	*UBC-like*
At1g12410	*nClpP2*
At5g42990	*UBC-like*
At3g58040	*?protein*
At5g60360	*AALP*
At4g25110	*Metacaspase*
At5g53300	*UBC10*
At2g02760	*UBC2 (RAD6)*
At5g62540	*UBC3*
At5g41700	*UBC8*
At3g20060	*UBC*
At4g32940	*VPE gamma*

At3g46460	*UBC13*
At3g62980	*TIR1*
At5g10760	*CND41-like*
At1g53750	*RPT1a*
At1g12410	*nClpP2*
At5g62540	*UBC3*
At1g64230	*UBC-like*
At3g08690	*UBC11*
At4g25110	*Metacaspase*
At5g20000	*RPT6a-putat.*
At5g60360	*AALP*
At3g20060	*UBC*
At5g53300	*UBC10*
At5g42990	*UBC-like*

Table 10. Comparison of sets of mRNAs GO-annotated "proteolysis", induced (over the black line) and repressed (under the black line) in treatments with high concentration of MMS (left), low concentration of MMS (middle) and the combined treatment (right).

Up-regulated proteolysis-involved genes (other than E2s) were represented *e.g.* by the senescence associated chloroplast protease *nClpP2* (*At1g12410*, by high and low MMS) and *TIR1* (transport inhibitor response 1, *At3g62980*, by high and combined MMS doses). *TIR1* is an F-box containing subunit of SCF-type ubiquitin-ligase complex (E3). SCF^{TIR} was shown to be necessary for proper auxin signaling (). *Metacaspase* (*At4g25110*) mRNA was down-regulated by all treatments while the transcript abundance of another potential senescence-

associated protease, cysteine protease *AALP* (*At5g60360*), was reduced following treatment with high and low concentration of MMS.

5.1.2.6 Cell cycle

Various organisms have been reported to respond to DNA damage by arresting cell cycle progression to prevent propagation of the damage to new cells and also increasing the inflicted injury by the cell division machinery (). The arrest is achieved by a signaling cascade involving proteins able to sense specific DNA lesions, translate the signal and finally members directly regulating cell proliferation. As MMS is a known DNA damaging agent, expression of cell cycle regulators was expected to be influenced by its administration. Cyclins and cyclin dependent kinases are the principal players in progression to the next cell cycle phase. Overexpression of *cyclin D3;1* (*At4g34160*) in *Arabidopsis* was shown to lead to hyperproliferation caused by induction of both replication and cell division in the affected cells (). *Cyclin D3;1* expression is repressed in all the treatments as well as that of *cyclin C-like* (*At5g48630*) in low and combined doses (Table 11). On the other hand, combined MMS treatment activates *cyclin B2;1* (identical to *cyclin 2a*, *At2g17620*) and a *pRB* (*Plant Retinoblastoma protein*, *At3g12280*). RB proteins of animals play a fundamental role in licensing cell cycle progression and thereby act as antioncogenes. They inhibit transition to S-phase until phosphorylated by a cyclin-dependent proteinkinase, usually complexed with a D- or E-cyclin. Arabidopsis and maize pRB proteins were demonstrated to interact with D-type cyclins and several components of the system are interchangeable with animal proteins (for a review see). Strong elevation of *pRB* transcript level connected with *cyclin D3;1* down-regulation is thus likely to signal a cell-cycle arrest inflicted to the cells by MMS treatment.

				At3g12280	*RB-related*
No genes		No genes		*At2g17620*	*Cyclin B2;1*
At4g34160	*Cyclin D3;1*	*At4g34160*	*Cyclin D3;1*	*At4g34160*	*Cyclin D3;1*
		At5g48630	*Cyclin C-like*	*At5g48630*	*Cyclin C-like*

Table 11. Comparison of sets of mRNAs GO-annotated "cell cycle", induced (over the black line) and repressed (under the black line) in treatments with high concentration of MMS (left), low concentration of MMS (middle) and the combined treatment (right).

Cyclin D3;1, rapidly down-regulated by MMS, was shown to be regulated by the cytokinin signaling pathway (). Other transcripts possessing GO annotation "response to cytokinin stimulus" comprise *Arabidopsis* Response Regulators *ARR4* (*ATRR1*, *At1g10470*, repressed in

high MMS), *ARR2* (*At3g48100*) and *ARR6* (*At5g62920*). The latter two transcripts were also reduced by the high concentration of MMS but induced in both the other treatments.

5.1.3 Macroarray data evaluation by RT-qPCR

As discussed in the Introduction, hybridization-based methods suffer from a range of potential problems that can compromise the obtained results. On the other hand, the throughput of the contemporary microarrays allows addressing most of the genes comprising an eukaryotic genome. It is thus advisable to combine the enormous data flow of array hybridization with the precision of another, low-throughput, technique for a subset of transcripts selected on the base of array data. Recently, it is often RT-qPCR, as it is the most quantitative method. RT-qPCR was also used to evaluate the results obtained from the cDNA-array experiment with MMS-treated *Arabidopsis* suspension culture.

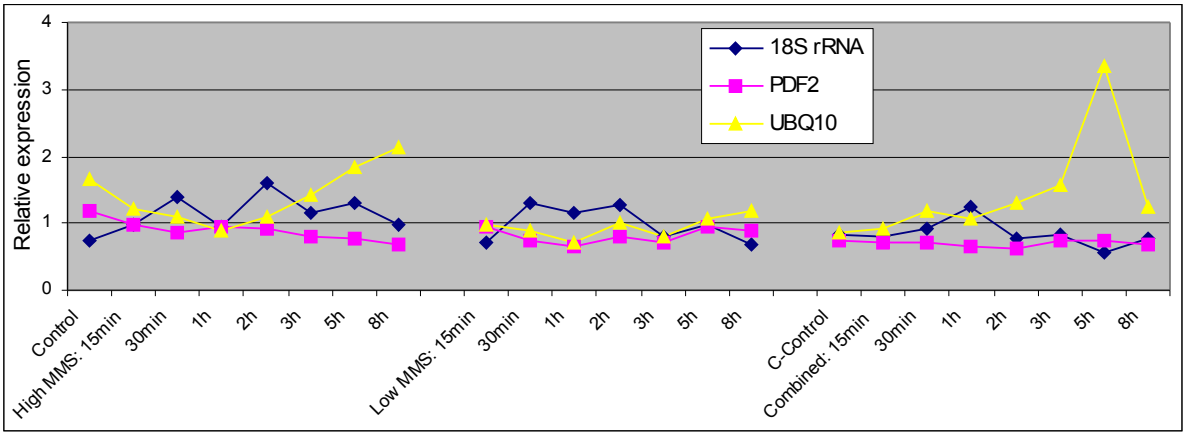


Figure 4. Comparison of three potential internal control transcripts. Under ideal conditions, an internal control should be represented by a horizontal line.

Three transcripts were tested as internal controls for normalization of the data: *18S* rRNA, protein phosphatase 2A regulatory subunit *PDF2* (*At1g13320*) and polyubiquitin 10 *UBQ10* (identical to *SEN3* – senescence-associated protein, *At4g05320*). As the RT reactions contained a constant input of 1µg RNA, a good internal control gene should display an invariable expression level, only disturbed by irregularities in RT. It follows from Figure 4 that *PDF2* had the lowest variability of the three. While *18S* rRNA fluctuated, *UBQ10* transcript was clearly elevated by high concentration of MMS. If pretreatment was used, the *UBQ10* induction appeared even stronger. However, at the low dose, the behavior of *UBQ10* mRNA appeared very similar to that of *PDF2*. Therefore *PDF2* served for normalization obtained for all the other transcripts.

Five transcripts presented in Figure 5 were quantified by RT-qPCR as described in Materials and Methods. The trends of expression profiles are generally similar, although there are also

discrepancies. *Cyclin D3;1* mRNA abundance dropped immediately (before 15 minutes of any treatment) and stayed low for the rest of the experiment, including the whole combined treatment. On the other hand, using array hybridization, the transcript amount decreased gradually for several hours. This might be indicative of another mRNA containing a similar sequence and beeing subject to slower removal after MMS tretment. However, the trend was retained and the interpretation thereby as well. *TL1* (*Thaumatin-Like protein 1*, *At1g75800*) demonstrated very similar expression profiles between the two compared methods. Progressive lowering of *TL1* mRNA abundance in all treatments was slightly slower in the low MMS dose. The expression of *Myb30* was shown to decrease in high and low MMS by both techniques. However, when the combined treatment was applied, the obtained trends strikingly differ. As results for the other two doses are congruent, it seams that the difference could have arisen from divergent behaviour of the cells subjected to combined MMS treatment. RT-qPCR employing primers specific for *GST1* and *MPK3* cDNAs confirmed induction of these genes by high MMS concentration. However, similar (but weaker) trend in qPCR results was observed for the combined treatment, while no obvious change in expression appeared in the array experiment. Low concentration of MMS additionally elevated mRNA level of *GST1* in the array- but not in the qPCR experiment.

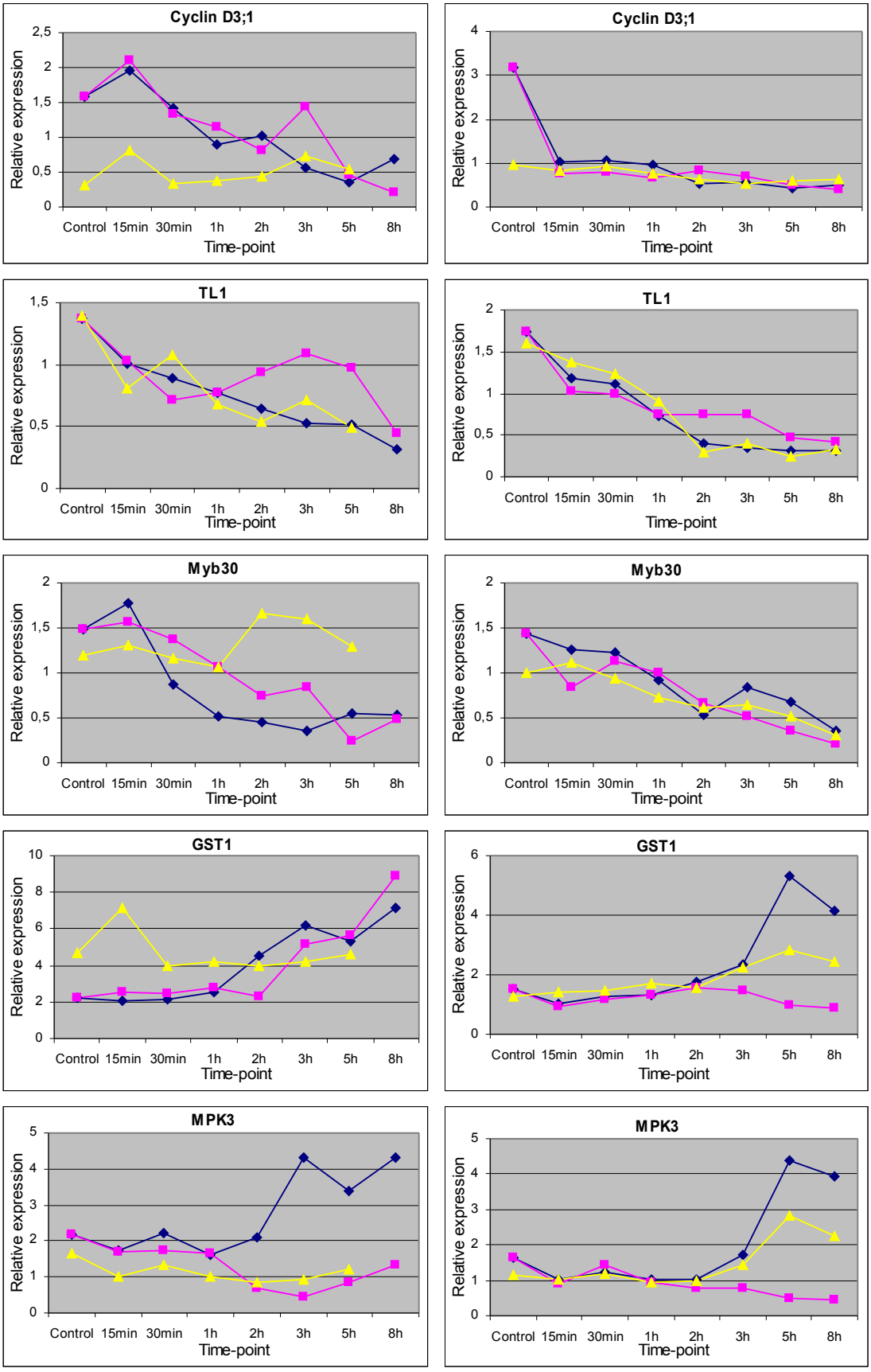


Figure 5. Comparison of expression profiles of 5 genes obtained from macroarray experiment (left) and RT-qPCR (right). The designations are: Cyclin D3;1 - *At4g34160*, TL1 - *At1g75800*, *Myb30* - *At3g28910*, GST1 - *At1g02930*, MPK3 - *At3g45640*. The curves represent high MMS concentration (♦), low MMS (■) and the combined treatment (▲). Please note that "Control" position of the combined treatment curve corresponds to the pretreated control, C-Control.

5.2 Reverse transcription yield for quantitative PCR normalization

The idea of measuring yield of RT reaction is not completely novel. used PicoGreen for quantifying RNA-DNA hybrid molecules after RT with polyA$^+$ RNA added as a template. PicoGreen is a fluorescent dye preferentially binding to double stranded (ds) nucleic acids, however, the fluorescence of RNA-PicoGreen complexes can reach up to 10% of dsDNA-PicoGreen fluorescence at the same concentration (). Their method cannot be thus applied to cDNA reverse transcribed from total RNA. Ribosomal RNA (rRNA), predominant in total RNA, contains many double stranded motifs that bind PicoGreen, and this RNA-borne signal outweighs the cDNA-associated fluorescence. Because total RNA is employed in RT-qPCR assays much more frequently than polyA$^+$ RNA, a modification of the method to overcome this limitation would make it more useful.

Removing RNA from the RT reaction mixture is thus a critical step in this procedure. The efficiency of RNA removal was assayed by RiboGreen RNA Quantitation Kit (Invitrogen) as a decrease in fluorescence. Using RNaseA (Qiagen, Carlsbad, USA) resulted in incomplete RNA hydrolysis, even after increasing the reaction temperature to 62°C to reduce secondary structures (according to). In addition, a component of the reaction partially inhibited PCR intended to prove that cDNA remained intact. As an alternative, alkaline hydrolysis of RNA was employed, using 0.1 M NaOH for 20 minutes at 70°C. The RNA-RiboGreen fluorescence dropped about 100 fold (Figure 6A), and neither higher hydroxide concentration nor increased temperature or extended time lowered the signal further. The cDNA was not affected as assayed by real-time PCR with primers specific for the gene *MPK3* (*At3g45640*), Figure 6B.

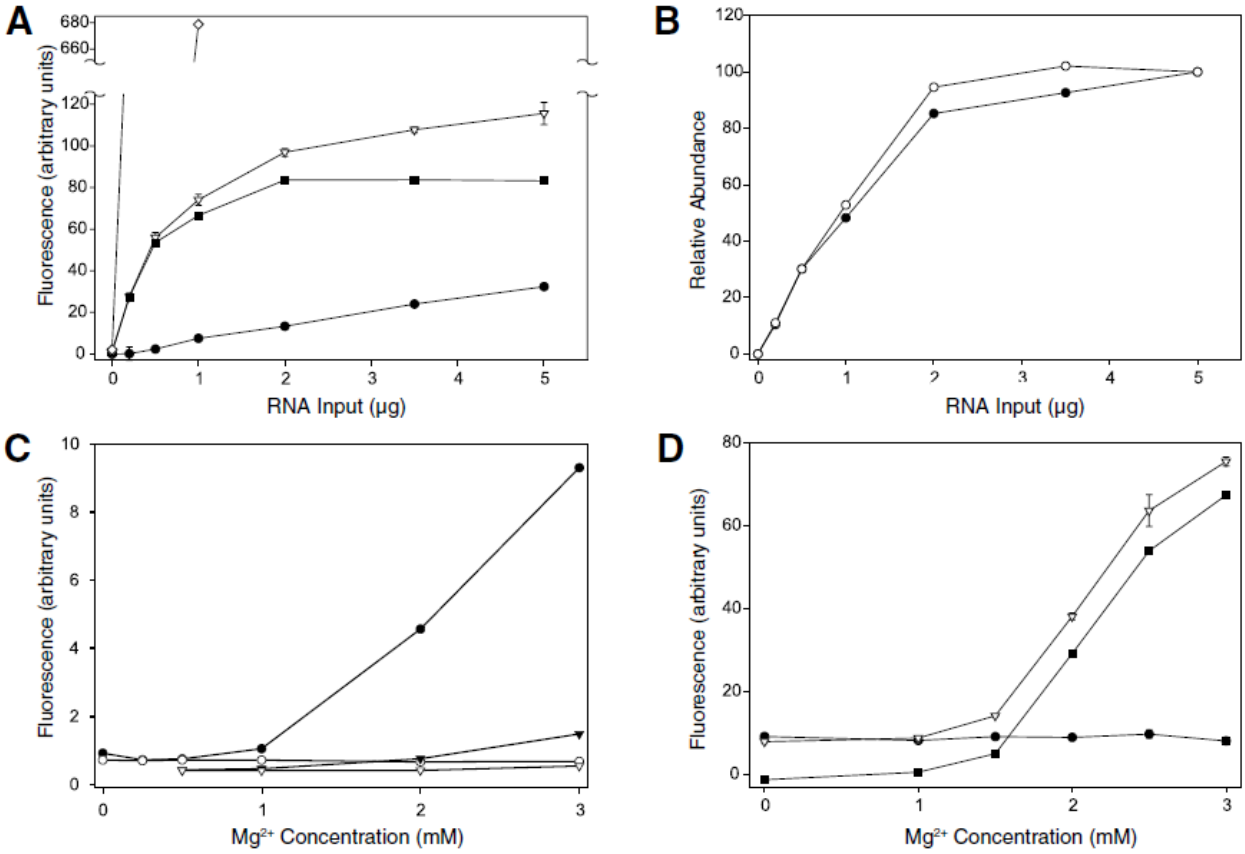


Figure 6. RiboGreen fluorescence as a function of first strand cDNA yield. (A) RiboGreen fluorescence of hydrolyzed samples after reverse transcription (RT;▽), background before RT (●), and the difference between these (■) from reactions at various RNA input. "Mock-hydrolyzed" RT samples in which NaOH was neutralized before adding to the sample (◇) illustrate the efficiency of RNA removal. (B) cDNA survives alkali treatment - relative *MPK3* cDNA abundance in the hydrolyzed (●) and mock-hydrolyzed (○) samples from panel A. (C) RiboGreen and PicoGreen staining of identical samples - RiboGreen fluorescence of RT (●) and background (○); PicoGreen fluorescence of RT (▼) and background (▽). (D) RiboGreen fluorescence of reaction mixtures at various Mg^{2+} concentrations - legend is the same as in panel A. In panels A and D, three aliquots of a single RT reaction were independently hydrolyzed, neutralized, and measured; error bars denote the standard deviation.

After the alkaline hydrolysis, obtaining single stranded (ss) cDNA could be expected. As this cannot be measured with PicoGreen, finding another dye was necessary. Although RiboGreen is a part of the RNA Quantitation Kit, it is capable of forming a fluorescent complex with any nucleic acid . Indeed, Figure 7 indicates that while PicoGreen and SYBR GreenI are specific to double-stranded (ds) nucleic acids (in this case DNA), RiboGreen is relatively nonspecific.Its fluorescence is however still about 2-fold higher when it binds to dsDNA than to ssDNA or RNA. For the specificity of the RiboGreen RNA Quantitation Kit is thus critical the use of RNase-free DNase.

Figure 6C documents comparison of the fluorescence of RNA samples hydrolyzed before and after RT with both PicoGreen and RiboGreen. Only RiboGreen fluorescence significantly increased after RT, which confirms single-stranded nature of the hydrolysis-resistant DNA.

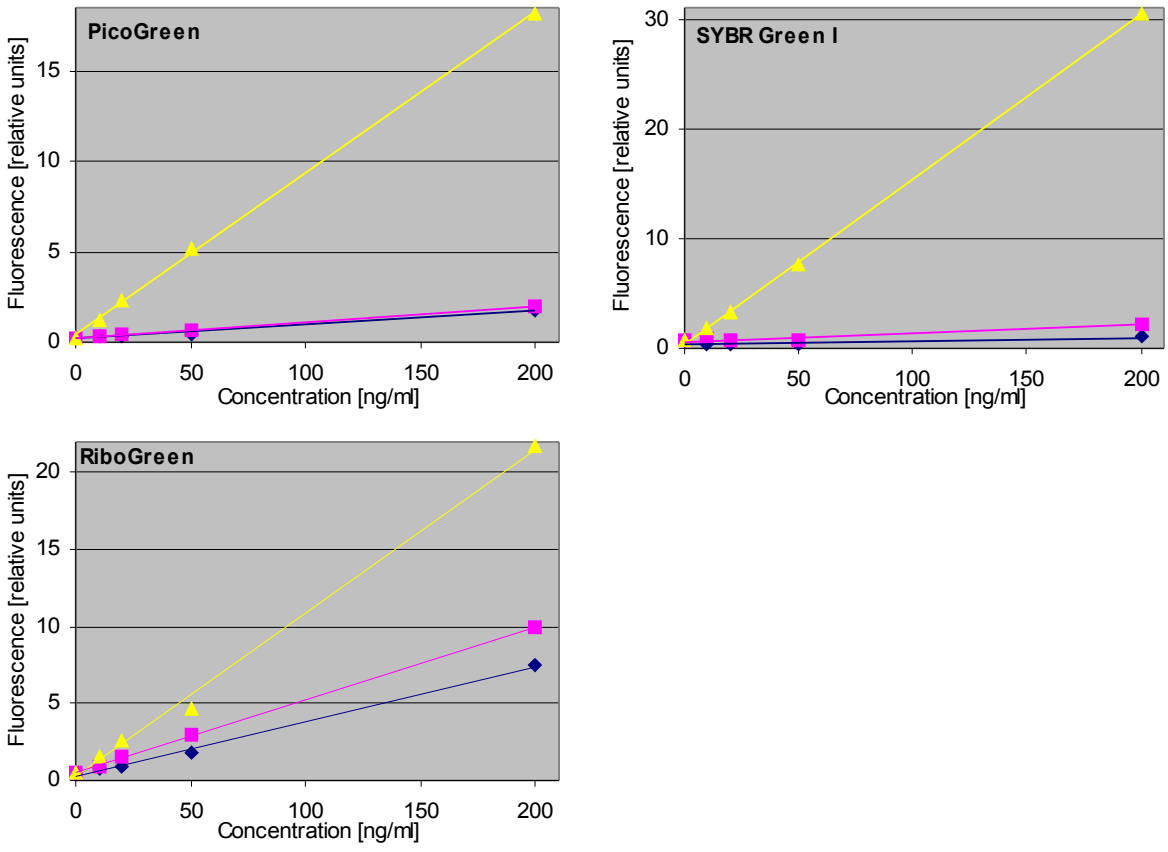


Figure 7. Comparison of PicoGreen, RiboGreen and SYBR GreenI fluorescence intensity when bound to various types of nucleic acids: dsDNA (▲), ssDNA (■) and total RNA (◆).

The PicoGreen/RiboGreen comparison was performed with a set of RT reactions containing a range of concentrations of free Mg^{2+}; since reverse transcriptase requires Mg^{2+} as a cofactor,

as the free Mg^{2+} concentration decreases (and thus RT efficiency declines), the RiboGreen signal is also expectedly reduced (see Figure 6C).

Figure 6A shows RiboGreen signal as measured with a range of different cDNA yields. As might be expected, the amounts of cDNA produced initially increase with the amount of RNA input, then level out as the reaction becomes saturated. In the experiment shown in Figure 6D, RNA input was held constant (1 µg), and the free magnesium ion concentration was decreased by adding various amounts of EDTA. The consequent reduction in RT efficiency was mirrored by lower RiboGreen fluorescence readings.

To demonstrate the utility of the proposed method in quantitative RT-PCR, primers specific for genes *MPK3* and *UBQ10* were used (see Table 2 in Materials and Methods). For comparison of samples in which RNA input differed, all values were expressed relative to the cDNA abundance measured in the quantitative RT-PCR based on 5 µg input RNA; in the reactions involving different magnesium concentrations, all values were expressed relative to the cDNA abundance measured in the quantitative PCR using the 3 mM Mg^{2+} RT reaction. The relative percent abundances as measured by quantitative PCR were then compared to the relative fluorescence measured by RiboGreen. For different RNA input amounts in RT reactions primed by the dN_6 primer, the relative measures are shown in Figure 8A and the correlation is presented in Figure 8B. Figure 8C shows that the degree of correspondence is equally high in experiments in which RT was performed using the $dT_{23}dV$ primer. Similarly, Figure 8D demonstrates a high level of correlation between relative fluorescence and relative abundance in reactions performed using dN_6-primed RT with various Mg^{2+} concentrations. The one-to-one correspondence that was observed in these experiments confirms the suitability of the RiboGreen fluorescence reading as a correction factor: if these samples were normalized according to the RiboGreen readings, the resulting abundance values would vary only slightly. They would thereby confirm that *MPK3* and *UBQ10* expression levels did not fluctuate under the conditions of the experiment (they only appeared to differ because of the differences in global RNA input or RT efficiency; this is not surprising as we always used the same RNA stock). It should be noted that some data points correspond to conditions in which the RT reaction is saturated. Saturation occurs if the RNA input to enzyme activity ratio is too high, due to either excess RNA (*e.g.*, inputs > 1 µg RNA/50 U MoMLV, dN_6 primed, see Figure 8A) or RT inhibition (EDTA chelation of Mg^{2+}). Under saturation, normalization to RNA input becomes misleading; however, the fluorescence intensity attributable to cDNA yield correlates with levels of specific cDNA estimated by quantitative PCR whether or not the reaction is saturated.

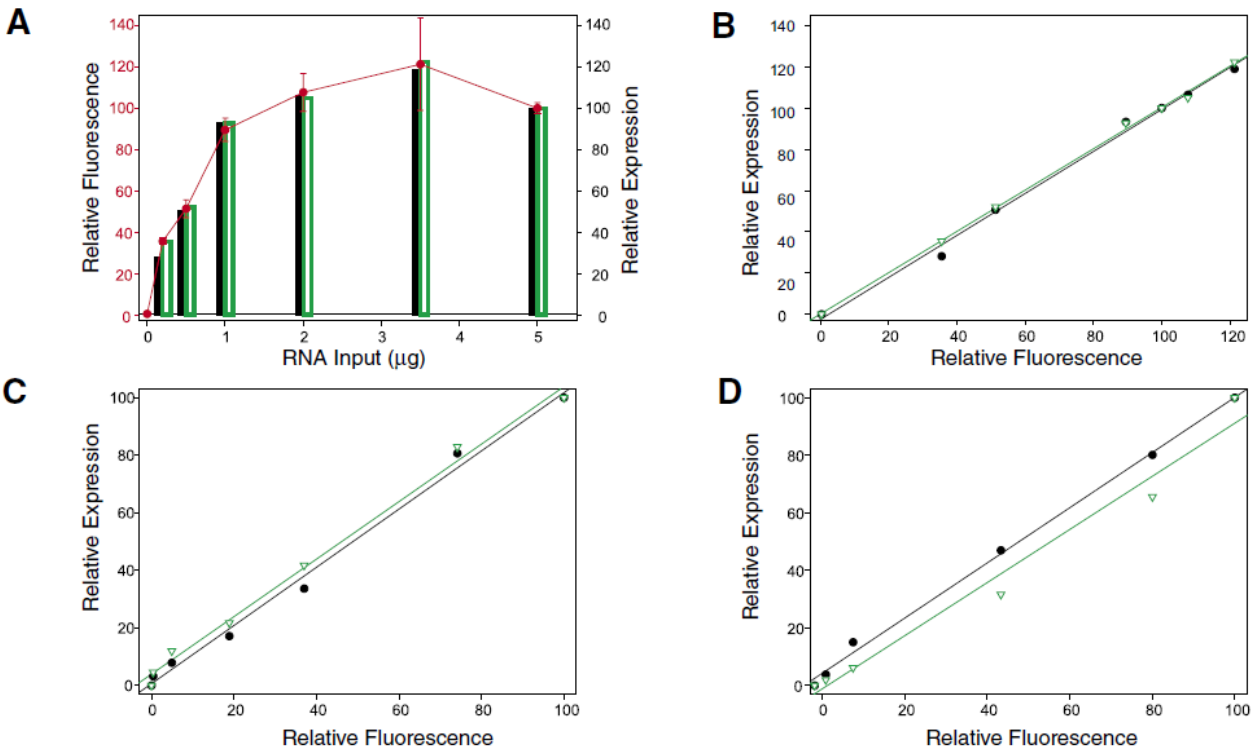


Figure 8. Correspondence of RiboGreen fluorescence to the specific cDNA abundance. (A) Relative abundance of *MPK3* (black bars) and *UBQ10* (green open bars) compared with relative RiboGreen fluorescence (red circles and line) in dN_6 primed RT reactions with a range of RNA inputs. The values of 5 µg sample were set to 100. For RiboGreen fluorescence readings, three aliquots of a single RT reaction were independently hydrolyzed, neutralized, and measured; errors bars represent standard deviations. (B) Correlation plot of the data shown in (A); ● = *MPK3* and ∇ = *UBQ10*. (C) Correlation of specific cDNA amount with RiboGreen fluorescence in $dT_{23}dV$ primed RT with a range of RNA inputs. (D) Correlation of specific cDNA amount with RiboGreen fluorescence in dN_6 primed RT with a range of free Mg^{2+} concentration (data from Figure 1D). The values of 3 mM Mg^{2+} sample were set to 100.

5.3 Response of tobacco to drought

To explore the influence of cytokinin storage pool size and/or of the stimulated cytokinin turn-over during drought progression, tobacco lines carrying constructs *35S::ZOG1* and *SAG12::ZOG1* were investigated along with wild type plants in our experiment. The two transgenic lines made possible an evaluation of the effect of generalized cytokinin content elevation as well as that of cytokinin overproduction restricted to senescing tissues on tobacco water deficit response.

My part in the project was to estimate the transcript abundances of the *ZOG1* transgene and of genes related to drought-response and its potential connection to cytokinin metabolism. Helena Štorchová and Jana Dobrá and Marie Havlová participated on the work as indicated bellow and in Materials and Methods.

5.3.1 Drought stress application

Drought stress was achieved by cessation of watering. To illustrate the reduction of water content, decrease of the soil moisture was estimated as a loss of pot weight. A major part of the final reduction of pot weight occurred during the first 3 days; further drought resulted in only marginal differences. This was probably mainly due to a high initial transpiration rate of plants with still fully open stomata. Figure 9A shows the values for *SAG12::ZOG1* plants. The reduction of water content in the plant body can be expressed as the relative water content (RWC). Figure 9B documents the RWC trend of *SAG12::ZOG1* plants after 1 day, 6 days and 11 days of drought treatment. After 6 days of water deficit, young leaves appeared to retain more water than old ones; however, after prolonged treatment, this difference vanished. In well-watered plants, the leaf number changed during 7 days of the experiment from 10 - 11 to 13-14. The stressed plants did not produce any new leaves; in addition they progressively lost 1-2 oldest leaves.

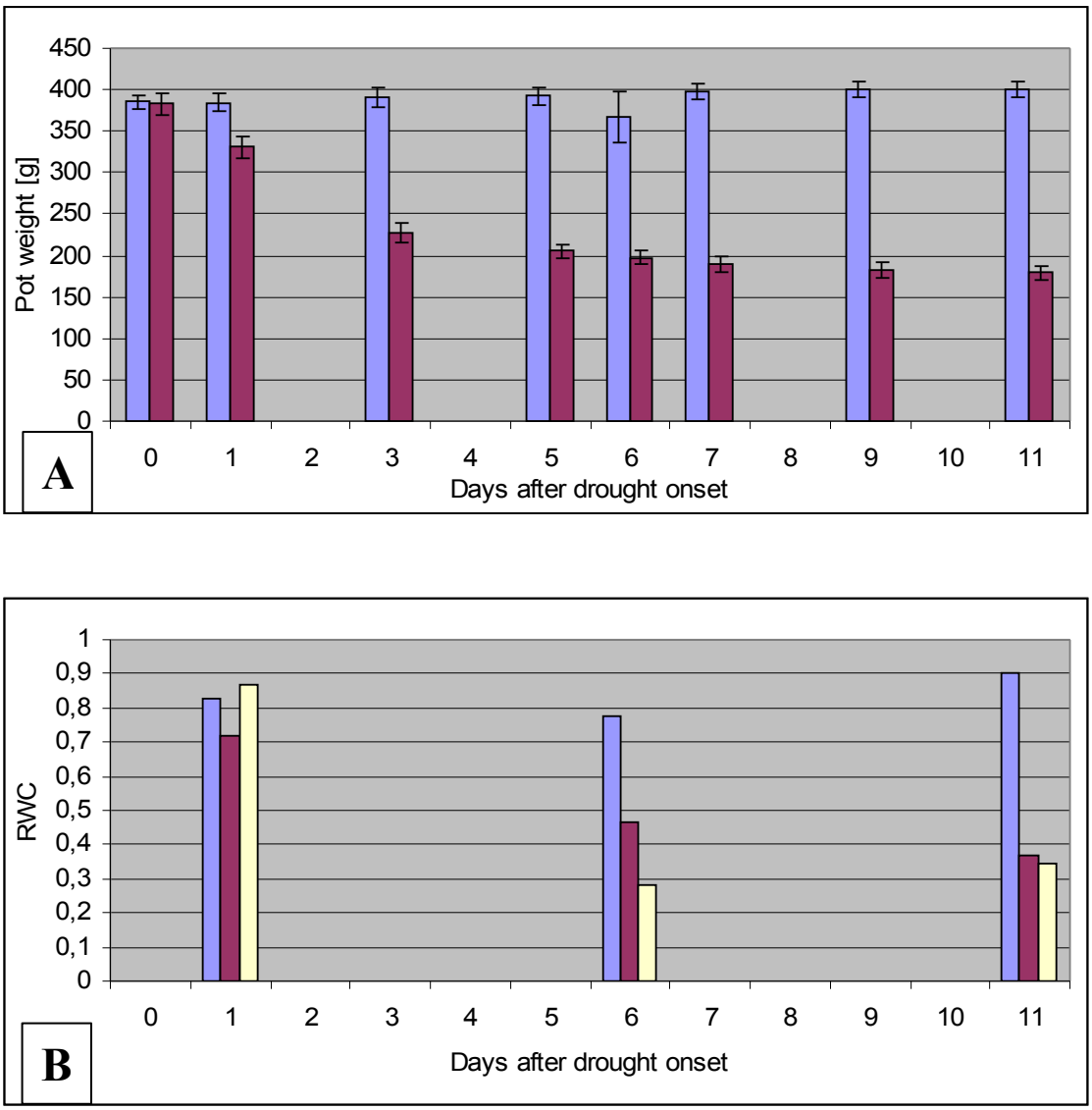


Figure 9. Time course of the water loss from *SAG12::ZOG1* plants.

(A) Pot weight of control (watered; blue columns) and drought-stressed (magenta columns) plants. (B) Relative water content in control (blue) and drought stressed plants – upper 2 leaves (magenta) and lowest 2 leaves (yellow) are depicted separately. Results from 4 – 9 plants are presented (mean value ± SD are shown). Data for Figure 9 were produced by Marie Havlová.

5.3.2 Quantification of transcript abundances

Quantitative RT-PCR was employed to measure transcript abundances of the genes *Cig1, CRK1, NtERD10B* and the *ZOG1* transgene in individual leaves of the control (watered) and drought treated plants. The expression levels were normalized to actin (*Tac9*) or to *18S* ribosomal RNA. As the normalization to these two transcripts led to similar results, both of them seem to be useful as reference genes. Data presented below were normalized to *Tac9*.

The third tested reference gene, coding for ribosomal protein *L25,* possessed higher variability and therefore was not suitable for data normalization (Figure 10).

ANOVA implemented in XLSTAT MS-Excel add-in was employed to test the influence of 4 factors on expression of the examined genes. These factors were: Genotype of the plant, growth conditions (drought or control), time of sampling (*i.e.* severity of the respective stress) and leaf position on the stem, numbered from the youngest unfolded one downwards. The P-values given below are results of the test, if not stated otherwise.

As the experiment performed with *SAG12::ZOG1* plants differed in some aspects (see Material and Methods), it would not be fair to directly compare its results to those of the other lines. However, as the basic scheme of the experiments remained the same, the data will be presented simultaneously.

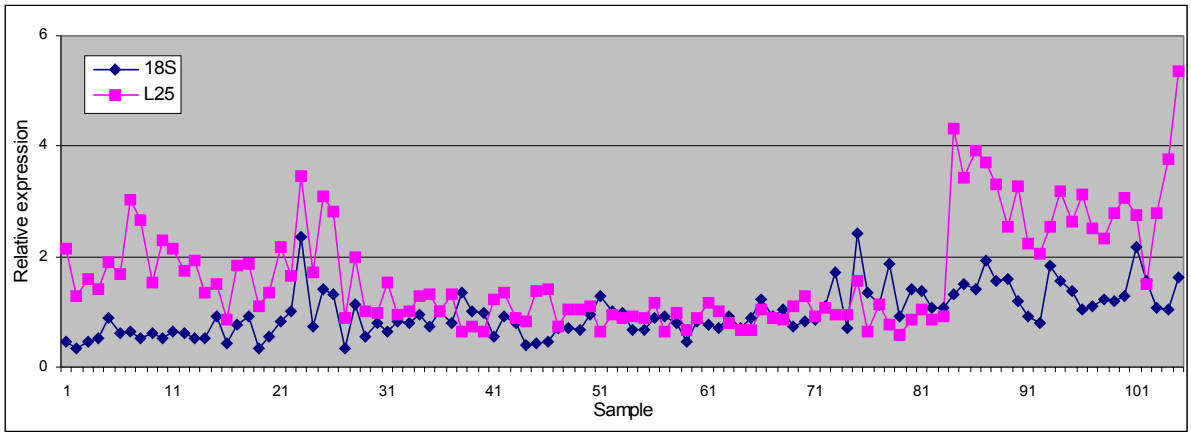


Figure 10 Comparison of calculated relative abundances of *18S* rRNA and *L25* transcripts normalized to *Tac9*. In this projection, a smooth horizontal line would represent a perfect correlation of expression levels of the respective transcript with those of *Tac9*.

5.3.2.1 ZOG1

The transcript abundance of *ZOG1* was assessed in the transgenic lines. *35S* promoter is generally considered to be constitutive under various conditions. However, analysis of individual leaves of two plants grown under moderate and severe stress and two control plants revealed that *ZOG1* transcript level was significantly higher (2.6-fold) in drought stressed plants than in the watered ones (Students t-test, $P < 0.0002$).

As *SAG12* promoter is switched on by senescence (), quantification of *SAG12::ZOG1* expression effectively visualized the onset of senescence in both stressed and control plants. The impact of drought stress on *SAG12* promoter activity is shown in Figure 11. Considerable stimulation of *ZOG1* expression started in the sixth leaf (from the top) of the drought stressed plants and in the eighth leaf of the control plants. It corresponded

approximately to the same position in an apex-base metabolic gradient because of difference in leaf numbers due to growth cessation under drought conditions. However, the induction of *ZOG1* expression was significantly higher in stressed plants (P = 0.02; Students t-test using leaves that reached the plateau of *ZOG1* expression - lowest 5 ones of control and 3 of stressed plants).

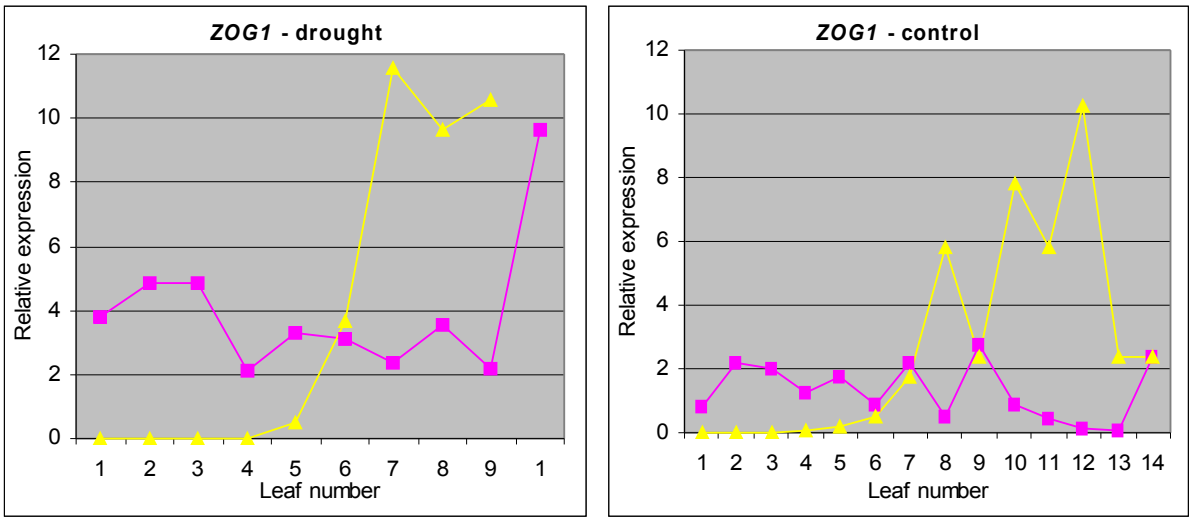


Figure 11. Expression of *ZOG1* in control and drought-stressed plants (*35S::ZOG1* - ■, *SAG12::ZOG1* - ▲). For *SAG12::ZOG1*, data from 6th day of drought are shown, for *35S::ZOG1* average values from 4th and 7th day of drought are shown. *ZOG1* expression levels were normalized to actin (*Tac9*).

5.3.2.2 NtERD10B

The abundance of *NtERD10B* transcript was generally induced by drought (P=0.01; Figure 12). *NtERD10B* was expressed at a stable level throughout the course of the experiment in the control plants. However, under stress conditions, its induction appeared already after one day and increased even further with the ongoing stress conditions. The mRNA was detected in all the leaves, with a slight ascending tendency towards older leaves. *35S::ZOG1* plants reached higher average *NtERD10B* expression levels than wild type and this comparison held also under stress conditions.

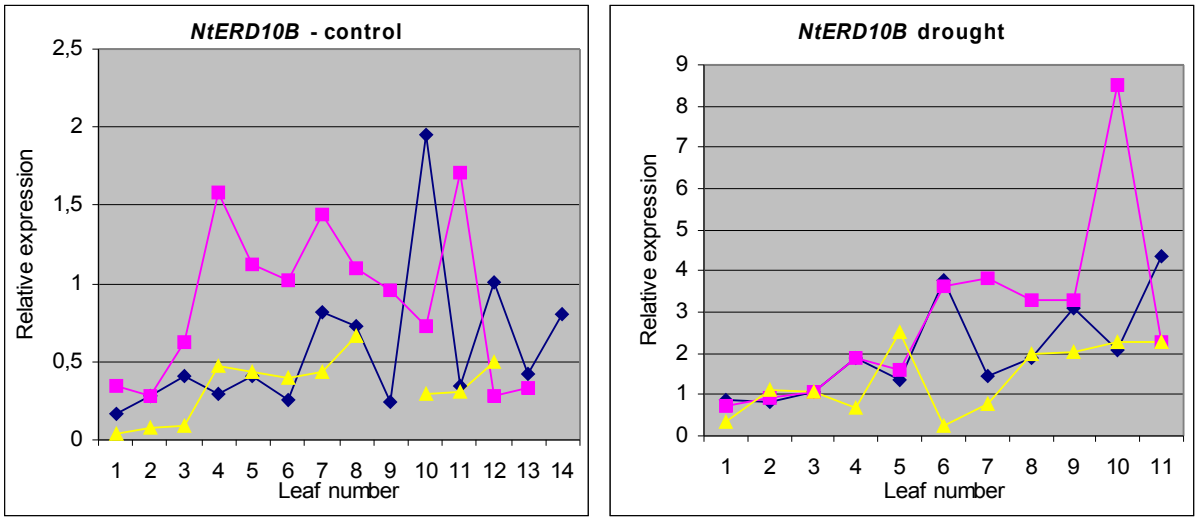


Figure 12. Expression of NtERD10B in control and drought-stressed plants (WT - ◆, *35S::ZOG1* - ■, *SAG12::ZOG1* - ▲). For *SAG12::ZOG1*, data from 6th day of drought are shown, for *35S::ZOG1* average values from 4th and 7th day of drought are shown. The expression levels were normalized to actin (*Tac9*). Please mention the different scale.

5.3.2.3 cig1

cig1 expression in WT plants followed a profile that rose progressively from low levels in youngest leaves to much higher (up to 20-fold) in the old ones (Figure 13). In stressed plants, the (otherwise low) expression sharply increased only in the 2 lowest leaves. While under control conditions *35S::ZOG1* possessed less *cig1* mRNA with the point of abundance increase shifted to older leaves, the expression profile under drought stress was very similar to that of WT. *SAG12::ZOG1* possessed increased *cig1* mRNA amounts already in relatively young leaves (from leaf number 3) with a peak in middle leaves (5 - 7). Further down along the plant axis, the expression decreased, although it always stayed higher than in the youngest 2 leaves. When drought was applied, *cig1* expression throughout the plant decreased to low values, even in the oldest leaves.

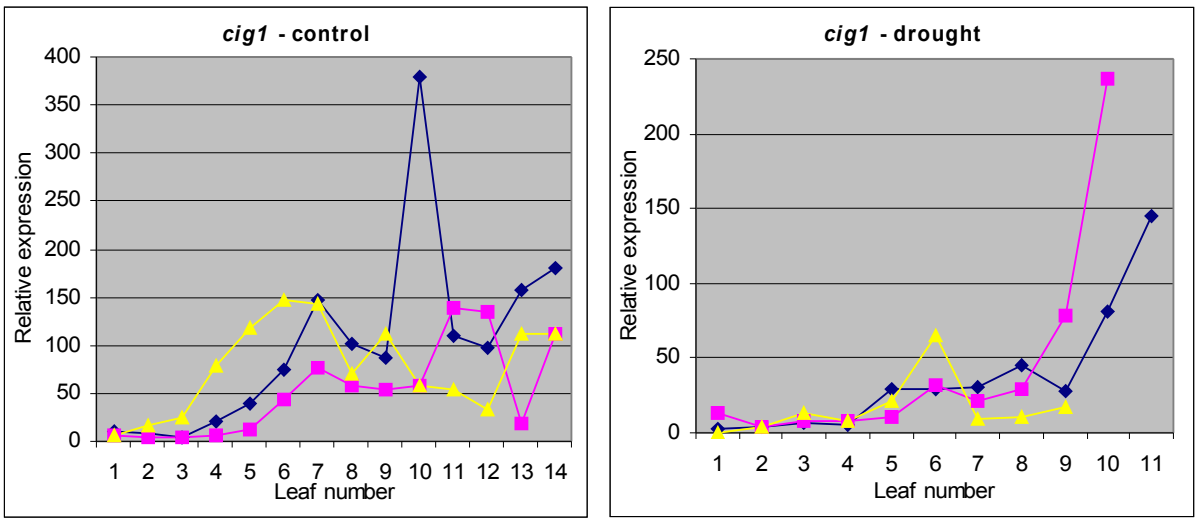


Figure 13. Expression of *cig1* in control and drought-stressed plants (WT - ◆, *35S::ZOG1* - ■, *SAG12::ZOG1* - ▲). For *SAG12::ZOG1*, data from 6th day of drought are shown, for *35S::ZOG1* average values from 4th and 7th day of drought are shown. The expression levels were normalized to actin (*Tac9*). Please mention the different scale.

5.3.2.4 CRK1

CRK1 transcript was generally significantly more abundant in well-watered plants than in the stressed ones under all circumstances (Figure 14). In agreement with previously published results (), it was found throughout the control WT leaves at a relatively low, rather uniform concentration with an indistinct peak around leaf number 5. On average, *35S::ZOG1* reached the highest expression levels of all tested genotypes. The detected mRNA abundances notably varied among individual leaves but also among whole single plants. The control in sampling 2 (pertaining to 4 days of drought) expressed the gene at higher level than any of the other experimental objects, several fold higher than the other *35S::ZOG1* controls. *SAG12::ZOG1* plants also appeared to possess a relatively high *CRK1* expression. However, on the contrary to the constitutive *ZOG1* expresser, *CRK1* transcript abundance followed a rather uniform distribution. Under drought conditions, *CRK1* expression in WT dropped approximately 2-fold and in *35S::ZOG1* about 20-fold (still being 2-fold higher than in WT). Also stressed *SAG12::ZOG1* plants possessed less *CRK1* mRNA then control. About 10-fold drop was observed in the young leaves, older ones however retained higher transcript levels.

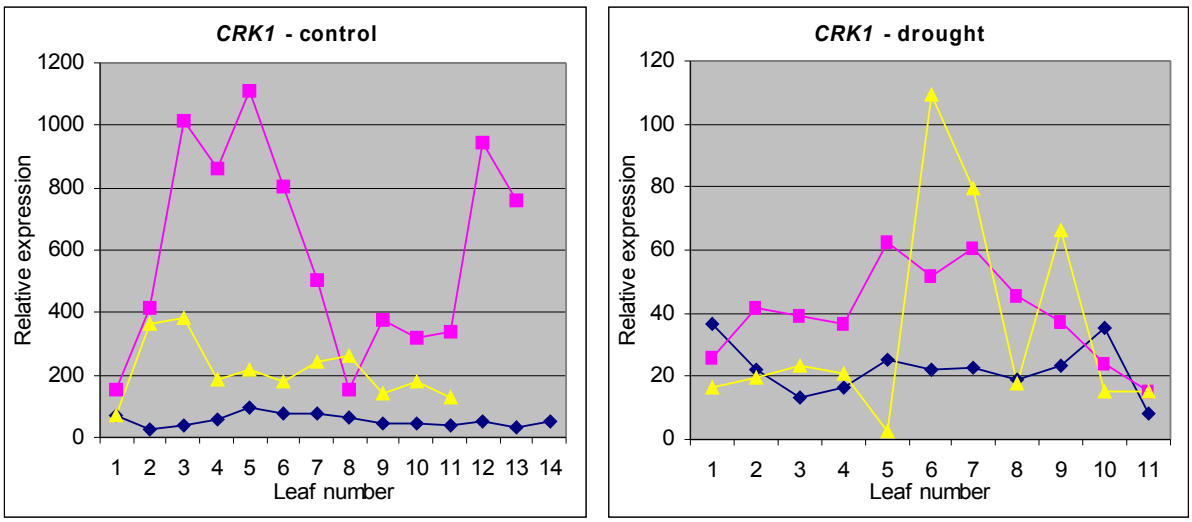


Figure 14. Expression of *CRK1* in control and drought-stressed plants (WT - ◆, *35S::ZOG1* - ■, *SAG12::ZOG1* - ▲). For *SAG12::ZOG1*, data from 6th day of drought are shown, for *35S::ZOG1* average values from 4th and 7th day of drought are shown. The expression levels were normalized to actin (*Tac9*). Please mention the different scale.

5.3.3 Comparison of oligo-dT and random hexanucleotide RT priming

As the RT-qPCR was carried out twice, once with anchored oligo-dT primer ($dT_{23}dV$) in the reverse transcription and once using random hexamers (dN_6), the impact of various priming strategies could be compared on quantification of the transcripts. Figure 15 presents the comparison of obtained crossing points (Cp) from the real-time PCR in a graphical way. While for *CRK1* and *cig1* dN_6 priming resulted in lower Cp values (on average by about 1 cycle; Figure 15 A,B), PCR specific for the other transcripts generated Cp higher by 2, 1.9 and 4.5 cycles (*Tac9*, *NtERD10B* and *ZOG1*, respectively; Figure 15 C, D and E) with a random-primed RT. Nevertheless, when observing the trends of gene expression profiles, the two methods provided congruent results.

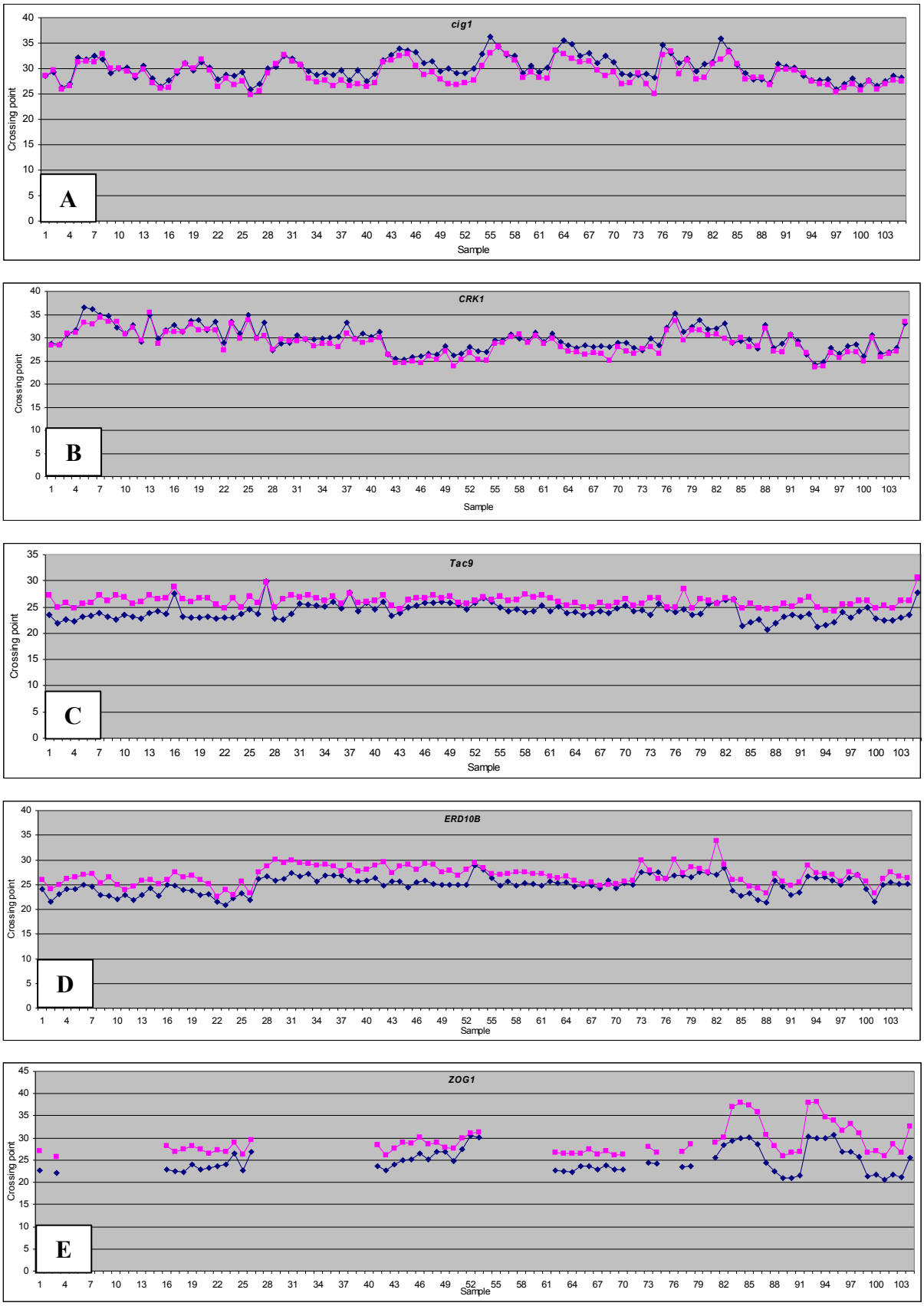


Figure 15. Comparison of crossing points obtained from real-time PCR of identical RNA samples, reverse transcribed using oligo-dT (◆) or random hexamer (■) primers. PCR primers were specific for cDNA of the following genes: *cig1* (A), *CRK1* (B), *Tac9* (C), *NtERD10B* (D) and *ZOG1* (E)

6 Discussion

6.1 Clustering of the expression data

For the purpose of analysis of the data obtained from experiments comprising MMS treatment of *Arabidopsis thaliana* suspension culture, a novel application named LoTrEC was developed. In contrast to most published procedures, LoTrEC is based on local trends of the expression profile curve combined with low-pass filtering. The low-pass filtering routine is in fact the principle of trend analysis used for various types of noisy or cyclic data. These features improve the resistance of LoTrEC to both random noise and outlying values. The results of LoTrEC processing of the above mentioned data were compared to those produced by STEM (), another algorithm dedicated to perform cluster analysis of short time-series gene-expression data.

Both the mentioned clustering algorithms (LoTrEC and STEM) share several characteristics. Either of them deals with time-series data and takes advantage of this fact in the analysis. The expression profiles are transformed to descriptions allowing for only a limited number of levels (states) at each time point. The description specifies two characteristics: the change from the previous value and its relative position compared to other samples (or to Control). Both of the algorithms also seek for large clusters.

However, there are significant differences between the procedures. STEM employs correlation coefficient of the whole time-series as a similarity measure. It was shown that correlation coefficient sometimes fails to recognize an obvious similarity (or dissimilarity) of curve shapes (). Outlying values can also severely interfere with the correlation. Indeed, when various similarity measures were tested, variants of correlation coefficient yielded sometimes very doubtful results. It happened mainly in cases where some outlying values occurred. Therefore a novel similarity measure was devised based on comparing corresponding segments of the profile descriptions.

The main difference in the process of cluster formation is that STEM generates theoretical profiles first and then assigns measured data to them. The cluster founders in LoTrEC algorithm are real expression profiles from the analyzed data set. Groups of similar members are formed around each profile that possesses a trend. Clusters representative of largest numbers of profiles are then chosen for further work. As the model profiles in STEM are randomly generated, this adds some variation to the results produced even by repeated analysis of identical data.

STEM offers statistical testing of likelihood that a particular cluster would have so many members by chance. Profile clusters possessing unlikely high counts of members are then suspected to represent biologically relevant pathways or co-regulated genes. This works well for large numbers of genes and for groups of co-regulated transcripts relatively rich in members (in the order of tens). However, for macroarrays possessing just hundreds of genes, only a portion of which are differentially expressed, it is quite rare to obtain a statistically significant enrichment in number of profiles attributed to a single cluster. Similarly, STEM performs testing of randomness of distribution of genes sharing same annotations (based on GO) among clusters. If a single cluster contains *e.g.* most of the pathogenesis-related genes considered in the study, it is likely to occur for some reason – they are probably co-regulated. Essentially only widely specified (and thus more populated) functional groups could turn out as statistically significantly enriched in some expression profile clusters, if the number of genes under study is not very high. However, such very general GO-terms ("regulation of cellular metabolism" etc.) are often not too informative about the processes occurring in the cell. If this kind of analysis is required for selected clusters (generated by any clustering algorithm) anyway, one can use either STEM (after setting its parameters to effectively skip the clustering routine) or a dedicated tool (such as FatiGO+,).

To keep real trends (which are usually plainer than a random zigzag) and to reduce random noise, LoTrEC applies low-pass filtering. This helps to detect subtler expression changes than in standard setups.

Applied cluster analysis differentiates expression profiles that vary not only in the basic trend (up- or down-regulation) but also in the temporal organization of the transcript abundance changes. It means that *e.g.* a profile possessing a "falling" trend at 30 min time-point and "rising" at time-points 2 h and 3 h is quite distinct from another one with "falling" at 2 h (opposite to the previous profile) and "rising" at 5 h and 8 h, although the latter curve might look just shifted to later time points.

The fact that LoTrEC discovered clusters possessing a statistically significant enrichment of a functional group of genes (according to FatiGO+) is indicative of the fact that the obtained clusters have a biological meaning. In addition, one of these groups (cluster number 3 in Table 5A) was not at all detected by the algorithm used for comparison (STEM). As the cluster contained profiles representing a substantial fraction of all the DNA-repair genes considered in the experiment and MMS is known to inflict DNA-damage, the biological relevance of their similar expression profiles appears quite realistic. In addition, there were other examples of GO-term enrichment that were however not significant due to a small size

of the corresponding clusters (*e.g.* cluster number 3 of low MMS treatment contained 5 of 10 ubiquitin conjugating enzymes present on the array).
LoTrEC allows to adjust the thresholds, sensitivity of trend detection and cluster size or *e.g.* to increase the influence of relative position on the clustering by setting the parameters used in the cluster analysis.

6.2 Analysis of transcriptomic data of MMS-treated *Arabidopsis* cells

Arabidopsis suspension culture was incubated in medium supplied with two different concentrations of MMS. To study the phenomenon of adaptation to methylation stress, the combined treatment was also applied, comprising pretreatment with a low concentration followed by the challenge dose. The cells were then harvested at a range of time-points to assess the dynamics of expression changes of selected genes. The transcriptional response differed significantly among the three treatment regimes.
High concentration of MMS that would be lethal for the cells after a prolonged incubation time clearly elicited a stress response. However, the reaction of the cells was not very specific. The predominantly activated functional groups of genes (based on GO annotation) comprised such as “response to wounding”,”response to toxin”,” response to pest, pathogen or parasite”. On the other hand, there was almost no effect (or even a negative one) on the DNA-repair machinery. As MMS is known to damage DNA, the finding is striking. To add more stones to the mosaic depicting state of the cell suffering from high dose of MMS, cell cycle-related transcripts were also investigated. The only observed effect was a fast down-regulation of cyclin D3;1 which plays an important role in the cell cycle progression. The number of differentially expressed proteases and other genes involved in protein catabolism was evenly split to groups of 5 up- and 5 down-regulated members. This contrasts to the preferential reduction of abundance of the transcripts of this type in the other treatments. To summarize it, 5 mM MMS seems to wreak havoc to the cells, as shown by on the level of DNA damage. The only reaction to the treatment that was detected in the presented experiment is a general stress response that does not target the specific lesions produced by MMS. It was previously shown that a massive DNA damage induces programmed cell death. In plants, presented this phenomenon in UV-C irradiated *Arabidopsis* seedlings. However, the gene expression changes observed with 5 mM MMS obviously contradict such a scenario. 0.5 mM was chosen as the low concentration of MMS because while it still produces a detectable amount of DNA damage, plant cells or even whole plants would survive the treatment. Low MMS produced expression changes in more transcripts than the high dose. In

contrast to the situation described above, 9 genes annotated as involved in DNA repair were activated while only one was repressed. It thus seems that DNA damage inflicted by MMS was properly detected in this case. However, some of the general stress response genes were induced simultaneously. As well as in high MMS concentration, the cell cycle obviously tended to be halted, as an additional cyclin transcript appeared to be down-regulated. This corresponds with the fact that all three affected genes implicated in "cell growth" were repressed. They comprised 2 expansins (*At2g39700* and *At1g69530*) and a peroxidase (*At3g49120*). Proteolysis was affected by low MMS in a pattern distinct from the high concentration effect. There are more transcripts of this category down-regulated than induced.

The response to low MMS concentration appears to be significantly more specific, aiming, among others, to remove the lesions produced in DNA. However, the pattern of differential expressions does not make PCD likely either in this experimental regime.

The combined MMS treatment continues the experiments of . They studied the effect of adaptation to DNA damaging conditions. Assessing the changes in gene expression represents the next step in the elucidation of the adaptation process. In general, the expression pattern obtained from cells subjected to the combined treatment resembles more closely the response to low than to high MMS dose (although the challenge concentration was also 5 mM). Some aspects of the reaction however seem to be even more pronounced in the combined treatment. For example 7 of the 9 DNA repair genes induced by the low dose are also activated in the combined treatment, along with additional 6 genes. The extent of the general stress response (represented here by transcripts annotated as "response to wounding" and "response to oxidative stress") remains approximately unchanged. Interestingly, the pattern of transcripts involved in proteolysis was different from the other treatments. There was only a single UBC transcript among the 5 induced proteolysis-related mRNAs (the others being three various proteases and an F-box E3-complex subunit) while 6 UBCs, 2 proteases and a 26S proteasome AAA-ATPase subunit *RPT6a* were repressed. *Metacaspase* is a protease that was down-regulated by all the MMS treatments. Although its function in PCD has been never confirmed, its expression profile is consistent with the fact that neither the combined MMS treatment seemed to induce apoptosis. However, reduced sensitivity to induction of PCD might be a trait advantageous for cell survival in a suspension culture.

From the presented data, it appears that preconditioning of the cells with a lower concentration of MMS indeed resulted in an adaptation. The pretreatment rendered the transcriptional response of the cells more similar to the low MMS concentration. The expression profiles of many genes indicated a more specific (and thereby presumably more

efficient) reaction to the stress, contrasting to the high MMS concentration applied alone. As MMS was shown to modify proteins, a possibility arises that molecules involved in a step of signaling could suffer methylation damage that would inactivate them. If the response had been prepared during the preconditioning it could fully develop even in the presence of otherwise prohibitive MMS concentration. This hypothesis would however need to be confirmed.

An alternative theory is that the extent of damage inflicted to an unprepared cell was as big as it disorganized the cell maintenance. The pretreatment induced detoxification and repair pathways in the cells, enabling them to prevent and mend the damage from the very beginning of the high dose application and thereby keep the level of sustained injuries relatively low.

The fact that so many transcript abundances were changed following MMS treatments can be attributed to two facts. First, although only a limited number of genes were assayed, their spectrum covered predominantly the ones assumed to have a differential expression. Second, the criteria for detection of the changes in transcript abundances were set relatively sensitive. However, because time-series data and low-pass filtering of the profiles is used, the threshold values for detection are in fact more stringent than they seem to be (*e.g.* default minimum fold-change of 1.4 would not effectively detect every expression change of 1.4 fold but rather ones significantly higher – depending on the context of the particular profile).

In addition, some genes are up- or down-regulated only in a certain period of time of the treatment. In experiments comprising one or two time-points, many differentially expressed transcripts of this kind could be missed. While not working with ratios (between expression at a particular time-point and in Control, for instance), also such cases can be included in which either of the compared values is very low (effectively zero) as long as the higher of them reaches a set threshold. This is useful to not discard genes that are virtually switched on or off (from or to a very low expression, respectively) and also those possessing a very low expression only eventually rising to reliably measurable levels. DNA-repair genes are a good example of the latter behavior.

6.3 Macroarray data evaluation by RT-qPCR

Biological systems are inherently variable in their forms and also in their reactions to the environment. The experiment addressing the influence of MMS treatment on gene expression of *Arabidopsis thaliana* suspension culture was therefore performed twice. The first experiment provided material for the cDNA-array hybridization. Part of the obtained results were then evaluated by means of RT-qPCR using cells treated with MMS in an independent

experiment following the same procedure. Five genes were selected that possessed clear trends of expression profiles. Both induced and repressed genes were represented. Most of the trends were confirmed, they appeared similar between the two methods. However, there were striking differences as well (Figure 5). These could be explained by either artifacts produced by the employed techniques or by biological variability of the material in conjunction with possible unintended slight differences in the procedure. These inconsistencies comprise *Myb30* and *MPK3* (combined treatment) and *GST1* (low and combined dose). Obviously, the deviation in all three of the transcripts appeared when combined MMS treatment was applied. One possible explanation is that the cells were in a different physiological state and specifically reacted in a distinct way to the combined treatment. However, it might have happened that the deviation resulted from a difference in handling the cells during the experiment. Indeed, the combined treatment comprised higher number of operations than the other two regimes and was therefore more likely to be performed differently. From the results obtained for *GST1* and *MPK3*, it seems that the pretreatment produced less adaptation to the effects of MMS in the second round of the experiment, rendering the combined treatment more similar to high MMS dose. Anyway, this does not apply to *Myb30* that possessed a unique expression profile, obviously distinct from those obtained after either high- or low MMS treatments.

6.4 Reverse transcription yield for quantitative PCR normalization

As discussed before, normalization of RT-qPCR results is vital for obtaining meaningful information. Internal control transcripts (ICTs) are the most common means of normalization. However, if not properly evaluated, using an ICT can produce misleading results. For example, one of the ICTs that was thoroughly evaluated in a wide range of conditions and treatments, *UBQ10* () appeared to be unreliable for MMS treatment. *UBQ10* mRNA level increased in the later time-points of the high MMS and combined but not low MMS regimes. Additionally, there are no tested ICTs yet for many organisms and establishing a reliable assay can thus be a complicated task.

Therefore, a novel method for RT-qPCR normalization was developed that relies on the total cDNA yield as a factor used for the normalization. The procedure takes advantage of differential susceptibility of DNA and RNA in regard of alkaline hydrolysis. While RNA is damaged to a great extent after 20 min incubation in 100 mM NaOH at 70°C, DNA appears not severely affected. The removal of RNA is a necessary step because there is no fast and easy detection method discriminating between RNA and DNA. Common nucleic acid quantification methods are based either on spectrophotometric (absorbance at 260 nm wavelength) or fluorimetric (using dyes binding to nucleic acids) detection. The latter techniques are generally more sensitive, allowing to detect amounts of cDNA produced in a typical RT reaction. However, RNA usually predominates over the newly synthesized cDNA and most of the measured signal therefore represents the amount of RNA. As the RNA-dependent fluorescence falls to approximately 1 % of the original value after alkaline hydrolysis, even a lower DNA-borne signal can be measured.

Single stranded nature of cDNA resulting from the hydrolysis reaction precludes employing double-stranded DNA specific dyes SYBR Green I and PicoGreen. On the other hand, RiboGreen proved useful for this task. If constant ionic strength and dye concentration are used, RiboGreen fluorescence correlates with RNA input amount and also with RT-efficiency. A range of RT-efficiencies was produced by manipulating concentration of free magnesium ions.

The presented method of direct measurement of cDNA amount makes possible an independent verification of invariant expression of candidate reference genes. Alternatively, it offers an alternative way to normalize mRNA expression levels in the absence of suitable housekeeping genes. As the RiboGreen nucleic acid detection is very sensitive () low yield RT reactions may be measured. However, in such cases, random noise influence can increase to unacceptable levels, depending mainly on the sensitivity of applied fluorimeter. As oligo-

dT primed reactions generally produce less cDNA, they tend to suffer from this fact more than random primed RTs. Therefore the gene expression data in this work, when normalized using cDNA yield, possessed substantially more variation than when normalized to our evaluated ICTs.

6.5 Response of tobacco to drought

As cytokinins play a role indrought stress response, it is tempting to speculate that altered cytokinin metabolism could lead toto more drought-resistant crop plants. To study the influence of total cytokinin pool on drought response of tobacco, with wild type plants were employed together with two transgenic lines. Both of them express *ZOG1*, an enzyme converting cytokinins to their inactive storage form, O-glucoside. As the plant compensates for *ZOG1* activity by increasing cytokinin production, the total cytokinin pool and turnover is elevated . The transgene was active either in the whole plant (*35S::ZOG1*) or exclusively in the senescing leaves (*SAG12::ZOG1*). All the plants were treated with a range of water deficit conditions. Transcript abundances were measured for three tobacco genes connected to drought stress and/or cytokinin signaling. Macroscopically, there was only a marginal difference between the transgenic and wild type plants, however the transcript abundances differed substantially.

To illustrate the expression profiles of *ZOG1* transgene, its mRNA level was measured by means of RT-qPCR. The transcript was detected throughout the *35S::ZOG1* plants, although in varying amounts. The fluctuations however did not seem to follow any distinctive pattern. Interestingly, the drought treatedplants possessed significantly enhanced *ZOG1* expression. However, in a recent experiment performed in our laboratory (Jana Dobrá, personal communication), the *ZOG1* expression appeared lower. Therefore, we cannot conclude that *35S* promoter driven expression is enhanced by drought but merely that it is influenced by water deficit.

	Stress			Recovery
Leaves	1 day	6 days	11 days	1 day
Upper	0.0006	0.0036	0.0025	0.0002
Middle	0.0261	0.0283	1.2087	0.0089
Lower	0.9792	1.7981	2.3478	0.0041

Table 12. *SAG12::ZOG1* relative expression following three intensities of drought stress and subsequent rehydration quantified by RT-qPCR. *ZOG1* expression levels were normalized to actin (*Tac9*). Results from two distinct experiments (each comprising triplicate determination) were combined. “Upper” means the youngest unfolded leaf (number 1), “middle” is leaf number 5 and “lower” is one but lowest of still viable leaves. Data presented in Table 12 were provided by Jana Dobrá.

In *SAG12::ZOG1*, only older leaves possessed high concentration of the mRNA. This finding confirms the published profile of *SAG12* promoter activity . It is also in agreement with the results of Jana Dobrá cited in Table 12 and below.

Comparison of *SAG12::ZOG1* expression in upper, middle and lower leaves during the drought stress progression and after 1-day recovery is given in . Apparently, the expression driven by the senescence specific promoter increased under the moderate stress in lower leaves. Indeed, the lowest leaves started to die shortly after the day 6. Under severe stress, at day 11, *SAG12* promoter appeared active also in the middle leaves. Interestingly, expression of the transgene was virtually undetectable throughout the whole plant after a single day of rehydration. The leaves that apparently had acquired the fate of senescence were reprogrammed to stay alive.

As expected, water deficit increased the abundance of the mRNA for dehydrin, *NtERD10B*. The effect of drought was stronger in wild type and *SAG12::ZOG1* plants, although *35S::ZOG1* possessed highest average expression of all the lines. This fact is a result of relatively high dehydrin mRNA level in control (well watered) *35S::ZOG1* plants. Detection of higher abundances of *NtERD10B* transcript in older leaves corresponds well with the fact that old leaves are usually most affected by the stress – they are first to die in conditions of persistent drought.

The observed induction ratio was much lower than some published data on various dehydrins (*e.g.*). However, *NtERD10B* transcript was relatively abundant also in the control plants in the presented experiment. Recent data from our laboratory suggest that under modified conditions *NtERD10B* expression can be induced substantially more.

Transcript abundance of *cig1* proline dehydrogenase appears to depend on the leaf age/rank. In standard growth conditions, the gene is active in all but youngest 2-3 leaves. The onset of *cig1* expression in *35S::ZOG1* is delayed compared to wild type. Regardless of genotype, *cig1* mRNA level is reduced by water deficiency. Only the oldest leaves retain a high *cig1* activity, with the exception of *SAG12::ZOG1*. As the overall *cig1* signal is lower in *35S::ZOG1* than in wild type, *ZOG1* activity (or high concentration of cytokinin-O-glucosides) appears to negatively influence *cig1* expression. In *SAG12::ZOG1* plants, possessing high expression of *ZOG1* just in the oldest leaves, this could preclude the normal *cig1* expression pattern. As other parts of the plant do not express the transgene, a compensatory feedback mechanism of amino acid metabolism might activate proline decomposition in younger leaves than in wild type.

The observed reaction of *cig1* transcript abundance to drought is in agreement with the osmoprotective function of proline (). The oldest leaves are obviously senescing (possessing macroscopic signs of ageing and *SAG12* promoter activation) and they are likely to die if the conditions do not ameliorate. Proline decomposition might then help nutrient recuperation from them by providing glutamic acid, the central point of amino acid metabolism making a connection to sugar metabolism. As proline serves to bind water in the cells, its removal by the dehydrogenase might enable the plant to redistribute water from the dying old leaves to other parts of the body. To support this theory, it would be helpful to know the proline concentration in the leaves. This kind of analysis will be performed soon.

Receptor protein kinases constitute an inevitable component of many signaling pathways in all eukaryotic organisms. Plants possess a large family of receptor-like protein kinases (RLKs). For example about 600 RLK genes were discovered in the genome of *Arabidopsis thaliana* (). Larger genomes can contain even more members of the family – *e.g.* rice has about 1100 RLKs (). *CRK1* is a RLK, shown to be negatively regulated by addition of an active cytokinin, benzyladenine. However, the reaction was only transient. On the contrary, the expression of *ZOG1* from either of the employed constructs increases the concentration of a putative storage form of cytokinins in a long-term manner. This appeared to strongly induce *CRK1* expression in both the transgenic lines. The induction seems not limited to the area of active *ZOG1* transcription but might be proportional to the total *ZOG1* enzymatic activity or the amount of cytokinin-O-glucoside. Therefore *CRK1* mRNA abundance in *SAG12::ZOG1* background was elevated throughout the whole shoot. Although water deficit significantly reduces *CRK1* expression levels in all lines, the transgenics still maintain notably elevated levels. *CRK1* therefore might play a role in systemic signaling, integrating clues from various sensory pathways. *CRK1* transcript abundance exhibited the highest degree of variation even among control plants, predominantly in *35S::ZOG1* genotype. The fact that the fluctuations appeared among whole plants (and not single samples) and that they were specific for *CRK1* mRNA is our evidence that it was not a technical error of transcript abundance measurement. The discrepancies rather reflected physiological differences of unknown nature present among the plants, although they lived in an apparently homogeneous environment.

The data presented here are a result of a pilot project, finding the optimum conditions for a larger screen beeing performed recently. That is the main reason why the conditions were changed in the course of the experiment. This made the interpretation of the data more complicated; on the other hand, it provided more experience and a solid fundament for

systematic testing of more genes and transgenic lines. The results of this follow-up study should be published soon.

6.5.1 Comparison of oligo-dT and random hexanucleotide RT priming

Reverse transcription output was studied under conditions of two different priming strategies. The results show that there is no best primer generaly usable for any transcript, as some mRNAs were more efficiently transcribed with oligo-dT, others with random hexanucleotides. This finding is in agreement with that of . The differences result in different relative expression values among different studied genes. However, as these are relative values, comparable only among measurements of a single mRNA and only if produced with the same reaction setup, the effect on the resulting trends is insignificant. Data derived from oligo-dT primed RT reactions were used for the presented analyses, as this strategy proved superior for the majority of the studied transcripts.

7 Conclusions

1) cDNA array was prepared and used to assay transcript abundances of 376 selected *Arabidopsis* transcripts following various treatments with MMS.

   a) LoTrEC, clustering algorithm based on local trends of expression profiles, was designed and applied to the data. It succeeded to discover functionally rellevant clusters of expression profiles.

   b) Transcriptional responses to various MMS treatment regimes were investigated. While high MMS concentration seemed only to induce nonspecific stress reaction, the low and combined MMS resulted in a set of more specific expression changes.

   c) Expression levels of five transcripts were estimated by qRT-PCR. Trends of most of the profiles were confirmed.

2) Expression of 3 genes related to drought stress and/or response to cytokinin were measured by qRT-PCR in wild type and *ZOG1* transgenic plants. Transcript levels of all the genes were altered by water deficit.

   a) Although there are no significant macroscopic differences between wild type and *ZOG1* transgenic plants, the mRNA abundances appeared to be influenced by the genotype.

   b) Leaf position (age) significantly influenced the expression of *cig1* and *ZOG1* driven by *SAG12* promoter.

   c) RT primed with oligo-dT appeared more eficient than random hexanucleotide-primed reaction for 3 out of 5 mRNAs (*ZOG1,Tac9* and *NtERD10B*). The other two genes (*cig1* and *CRK1*) showed marginally more efficiency in random-primed RT.

8 List of cited literature

Al Shahrour F, Diaz-Uriarte R, Dopazo J. 2004 Mar. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics 20(4):578-580.

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J. 2002 Feb. MAP kinase signalling cascade in *Arabidopsis* innate immunity. Nature 415(6875):977-983.

Azuaje F. 2003 Mar. Clustering-based approaches to discovering and visualising microarray data patterns. Brief Bioinform 4(1):31-42.

Babiychuk E, Cottrill PB, Storozhenko S, Fuangthong M, Chen Y, O'Farrell MK, Van Montagu M, Inze D, Kushnir S. 1998 Sep. Higher plants possess two structurally different poly(ADP-ribose) polymerases. Plant J 15(5):635-645.

Babiychuk E, Kushnir S, Belles-Boix E, Van Montagu M, Inze D. 1995 Nov. *Arabidopsis thaliana* NADPH Oxidoreductase Homologs Confer Tolerance of Yeasts Toward the Thiol-oxidizing Drug Diamide. J Biol Chem 270(44):26224-26231.

Babiychuk E, Kushnir S, Van Montagu M, Inze D. 1994 Apr. The *Arabidopsis thaliana* apurinic endonuclease Arp reduces human transcription factors Fos and Jun. Proc Natl Acad Sci U S A 91(8):3299-3303.

Bachmair A, Potuschak T, Becker F, Nejinskaia V. 1994. Ubiquitin-dependent proteolysis in plants - a key metabolic pathway influencing plant-pathogen interaction. Advances in Molecular Genetics of Plant-Microbe Interaction 3:375-379.

Bailly V, Lauder S, Prakash S, Prakash L. 1997 Sepa. Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities. J Biol Chem 272(37):23360-23365.

Bailly V, Prakash S, Prakash L. 1997 Augb. Domains required for dimerization of yeast Rad6 ubiquitin-conjugating enzyme and Rad18 DNA binding protein. Mol Cell Biol 17(8):4536-4543.

Ballschmiter K. 2003 Jul. Pattern and sources of naturally produced organohalogens in the marine environment: biogenic formation of organohalogens. Chemosphere 52(2):313-324.

Bender J, Fink GR. 1998 May. A Myb homologue, ATR1, activates tryptophan gene expression in *Arabidopsis*. Proc Natl Acad Sci U S A 95(10):5655-5660.

Boffa LC, Bolognesi C. 1985 Oct. Nuclear proteins damage by alkylating agents with different degrees of carcinogenicity. Chem Biol Interact 55(1-2):235-245.

Bray CM, West CE. 2005 Dec. DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. New Phytologist 168(3):511-528.

Bustin SA. 2000 Oct. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25(2):169-193.

Callis J, Vierstra RD. 2000 Oct. Protein degradation in signaling. Current Opinion in Plant Biology 3(5):381-386.

Campos N, Bako L, Feldwisch J, Schell J, Palme K. 1992. A Protein from Maize Labeled with Azido-Iaa Has Novel Beta-Glucosidase Activity. Plant Journal 2(5):675-684.

Chen IP, Haehnel U, Altschmied L, Schubert I, Puchta H. 2003 Sep. The transcriptional response of *Arabidopsis* to genotoxic stress - a high-density colony array study (HDCA). Plant J 35(6):771-786.

Chomczynski P, Sacchi N. 1987 Apr. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162(1):156-159.

Cowan AK, Freeman M, Bjorkman PO, Nicander B, Sitbon F, Tillberg E. 2005. Effects of senescence-induced alteration in cytokinin metabolism on source-sink relationships and ontogenic and stress-induced transitions in tobacco. Planta 221(6):801-814.

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005 Sep. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiology 139(1):5-17.

Daniel X, Lacomme C, Morel JB, Roby D. 1999 Oct. A novel myb oncogene homologue in *Arabidopsis thaliana* related to hypersensitive cell death. Plant Journal 20(1):57-66.

Danon A, Gallois P. 1998 Oct. UV-C radiation induces apoptotic-like changes in *Arabidopsis thaliana*. Febs Letters 437(1-2):131-136.

Davies WJ, Kudoyarova G, Hartung W. 2005. Long-distance ABA signaling and its relation to other signaling pathways in the detection of soil drying and the mediation of the plant's response to drought. Journal of Plant Growth Regulation 24(4):285-295.

de Hoon MJL, Imoto S, Miyano S. 2002 Nov. Statistical analysis of a small set of time-ordered gene expression data using linear splines. Bioinformatics 18(11):1477-1485.

Despres C, DeLong C, Glaze S, Liu E, Fobert PR. 2000 Feb. The *Arabidopsis* NPR1/NIM1 Protein Enhances the DNA Binding Activity of a Subgroup of the TGA Family of bZIP Transcription Factors. Plant Cell 12(2):279-290.

Dewitte W, Riou-Khamlichi C, Scofield S, Healy JMS, Jacqmard A, Kilby NJ, Murray JAH. 2003 Jan. Altered cell cycle distribution, hyperplasia, and inhibited differentiation in *Arabidopsis* caused by the D-type cyclin CYCD3. Plant Cell 15(1):79-92.

Dheda K, Huggett JF, Chang JS, Kim LU, Bustin SA, Johnson MA, Rook GAW, Zumla A. 2005 Sep. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Anal Biochem 344(1):141-143.

Doucet-Chabeaud G, Godon C, Brutesco C, de Murcia G, Kazmaier M. 2001 Aug. Ionising radiation induces the expression of PARP-1 and PARP-2 genes in *Arabidopsis*. Mol Genet Genomics 265(6):954-963.

Drablos F, Feyzi E, Aas PA, Vaagbo CB, Kavli B, Bratlie MS, Pena-Diaz J, Otterlei M, Slupphaug G, Krokan HE. 2004 Nov. Alkylation damage in DNA and RNA - repair mechanisms and medical significance. Dna Repair 3(11):1389-1407.

Ernst J, Bar-Joseph Z. 2006 Apr. STEM: a tool for the analysis of short time series gene expression data. Bmc Bioinformatics 7.

Ernst J, Nau GJ, Bar-Joseph Z. 2005 Jun. Clustering short time series gene expression data. Bioinformatics 21:I159-I168.

Fariduddin Q, Ahmad A, Hayat S. 2004. Responses of Vigna radiata to foliar application of 28-homobrassinolide and kinetin. Biologia Plantarum 48(3):465-468.

Fowler S, Thomashow MF. 2002 Aug. *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14(8):1675-1690.

Frova C. 2003 Dec. The plant glutathione transferase gene family: genomic structure, functions, expression and evolution. Physiologia Plantarum 119(4):469-479.

Furukawa T, Kimura S, Ishibashi T, Mori Y, Hashimoto J, Sakaguchi K. 2003 Jan. OsSEND-1: a new RAD2 nuclease family member in higher plants. Plant Molecular Biology 51(1):59-70.

Gallego F, Fleck O, Li A, Wyrzykowska J, Tinland B. 2000 Mar. AtRAD1, a plant homologue of human and yeast nucleotide excision repair endonucleases, is involved in dark repair of UV damages and recombination. Plant Journal 21(6):507-518.

Gan S, Amasino RM. 1995 Dec. Inhibition of leaf senescence by autoregulated production of cytokinin. Science 270(5244):1986-1988.

Gray WM, del Pozo JC, Walker L, Hobbie L, Risseeuw E, Banks T, Crosby WL, Yang M, Ma H, Estelle M. 1999 Jul. Identification of an SCF ubiquitin-ligase complex required for auxin response in *Arabidopsis thaliana*. Genes Dev 13(13):1678-1691.

Hamilton JTG, McRoberts WC, Keppler F, Kalin RM, Harper DB. 2003 Jul. Chloride methylation by plant pectin: An efficient environmentally significant process. Science 301(5630):206-209.

Hanway D, Chin JK, Xia G, Oshiro G, Winzeler EA, Romesberg FE. 2002 Aug. Previously uncharacterized genes in the UV- and MMS-induced DNA damage response in yeast. Proceedings of the National Academy of Sciences of the United States of America 99(16):10605-10610.

Hare PD, van Staden J. 1997. The molecular basis of cytokinin action. Plant Growth Regulation 23(1-2):41-78.

Hecht SS. 1999 Mar. DNA adduct formation from tobacco-specific N-nitrosamines. Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis 424(1-2):127-142.

Heindorff K, Rieger R, Schubert I, Michaelis A, Aurich O. 1987 Nov. Clastogenic adaptation of plant cells - reduction of the yield of clastogen-induced chromatid aberrations by various pretreatment procedures. Mutat Res 181:157-171.

Henrie MS, Kurimasa A, Burma S, Menissier-de Murcia J, de Murcia G, Li GC, Chen DJ. 2003 Feb. Lethality in PARP-1/Ku80 double mutant mice reveals physiological synergy during early embryogenesis. Dna Repair 2(2):151-158.

Hoeberichts FA, ten Have A, Woltering EJ. 2003 Jul. A tomato metacaspase gene is upregulated during programmed cell death in Botrytis cinerea-infected leaves. Planta 217(3):517-522.

Hoffmann GR. 1980 Jan. Genetic effects of dimethyl sulfate, diethyl sulfate, and related compounds. Mutat Res 75(1):63-129.

Huggett J, Dheda K, Bustin S, Zumla A. 2005 Jun. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun 6(4):279-284.

Huynh LN, VanToai T, Streeter J, Banowetz G. 2005. Regulation of flooding tolerance of SAG12 : ipt *Arabidopsis* plants by cytokinin. J Exp Botany 56(415):1397-1407.

Islaih M, Halstead BW, Kadura IA, Li B, Reid-Hubbard JL, Flick L, Altizer JL, Thom DJ, Monteith DK, Newton RK, Watson DE. 2005 Oct. Relationships between genomic, cell cycle, and mutagenic responses of TK6 cells exposed to DNA damaging chemicals. Mutat Res 578(1-2):100-116.

Islaih M, Li B, Kadura IA, Reid-Hubbard JL, Deahl JT, Altizer JL, Watson DE, Newton RK. 2004. Comparison of gene expression changes induced in mouse and human cells treated with direct-acting mutagens. Environ Mol Mutagen 44(5):401-419.

Itai C, Benzioni A, Munz S. 1978. Heat-Stress - Effects of Abscisic-Acid and Kinetin on Response and Recovery of Tobacco-Leaves. Plant and Cell Physiology 19(3):453-460.

Jelinsky SA, Samson LD. 1999 Feb. Global response of Saccharomyces cerevisiae to an alkylating agent. Proc Natl Acad Sci U S A 96(4):1486-1491.

Jones LJ, Yue ST, Cheung CY, Singer VL. 1998 Dec. RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization. Anal Biochem 265(2):368-374.

Jordi W, Schapendonk A, Davelaar E, Stoopen GM, Pot CS, De Visser R, Van Rhijn JA, Gan S, Amasino RM. 2000. Increased cytokinin levels in transgenic P-SAG12-IPT tobacco plants have large direct and indirect effects on leaf senescence, photosynthesis and N partitioning. Plant Cell and Environment 23(3):279-289.

Kaina B. 1983 Nov. Cross-resistance studies with V79 Chinese hamster cells adapted to the mutagenic or clastogenic effect of N-methyl-N'-nitro-N-nitrosoguanidine. Mutat Res 111(3):341-352.

Kaina B. 1998 Aug. Critical steps in alkylation-induced aberration formation. Mutat Res 404(1-2):119-124.

Kaminek M, Motyka V, Vankova R. 1997. Regulation of cytokinin content in plant cells. Physiologia Plantarum 101(4):689-700.

Kapushesky M, Kemmeren P, Culhane AC, Durinck S, Ihmels J, Korner C, Kull M, Torrente A, Sarkans U, Vilo J, Brazma A. 2004 Jul. Expression Profiler: next generation--an online platform for analysis of microarray data. Nucleic Acids Res 32(Web Server issue):W465-W470.

Kasuga M, Miura S, Shinozaki K, Yamaguchi-Shinozaki K. 2004 Mar. A combination of the *Arabidopsis* DREB1A gene and stress-inducible rd29A promoter improved drought-

and low-temperature stress tolerance in tobacco by gene transfer. Plant and Cell Physiology 45(3):346-350.

Kawai M, Pan L, Reed JC, Uchimiya H. 1999 Dec. Evolutionally conserved plant homologue of the Bax inhibitor-1 (BI-1) gene capable of suppressing Bax-induced cell death in yeast(1). FEBS Lett 464(3):143-147.

Kimura T, Nakano T, Taki N, Ishikawa M, Asami T, Yoshida S. 2001 Jun. Cytokinin-induced gene expression in cultured green cells of Nicotiana tabacum identified by fluorescent differential display. Biosci Biotechnol Biochem 65(6):1275-1283.

King RW, Deshaies RJ, Peters JM, Kirschner MW. 1996 Dec. How Proteolysis Drives the Cell Cycle. Science 274(5293):1652-1659.

Kiyosue T, Yoshiba Y, Yamaguchi-Shinozaki K, Shinozaki K. 1996 Aug. A nuclear gene encoding mitochondrial proline dehydrogenase, an enzyme involved in proline metabolism, is upregulated by proline but downregulated by dehydration in *Arabidopsis*. Plant Cell 8(8):1323-1335.

Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF. 2002 Dec. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. Plant Physiol 130(4):2129-2141.

Krokan HE, Kavli B, Slupphaug G. 2004 May. Novel aspects of macromolecular repair and relationship to human disease. Journal of Molecular Medicine-Jmm 82(5):280-297.

Leon J, Rojo E, Sanchez-Serrano JJ. 2001 Jan. Wound signalling in plants. J Exp Botany 52(354):1-9.

Lindahl T. 1993 Apr. Instability and Decay of the Primary Structure of Dna. Nature 362(6422):709-715.

Lindahl T, Barnes DE, Klungland A, Mackenney VJ, Schar P. 1997. Repair and processing events at DNA ends. Ciba Found Symp 211:198-205.

Liu Z, Hall JD, Mount DW. 2001 May. *Arabidopsis* UVH3 gene is a homolog of the Saccharomyces cerevisiae RAD2 and human XPG DNA repair genes. Plant J 26(3):329-338.

Malek JA, Shatsman SY, Akinretoye BA, Gill JE. 2000 Aug. Degradation of persistent RNA in RNase-containing, high-throughput alkaline lysis DNA preparations. Biotechniques 29(2):250-252.

Mansfield TA, McAinsh MR. 1995. Hormones as regulators of water balance. In: Plant hormones. Physiology, Biochemistry and Molecular Biology. Davies PJ, editor, Kluwer Acad. Publ., Dordrecht. 598-616.

Martin RC, Mok DWS, Smets R, Van Onckelen HA, Mok MC. 2001. Development of transgenic tobacco harboring a zeatin O-glucosyltransferase gene from *Phaseolus*. In Vitro Cellular & Developmental Biology-Plant 37(3):354-360.

Martin RC, Mok MC, Mok DWS. 1999. Isolation of a cytokinin gene, ZOG1, encoding zeatin O-glucosyltransferase from *Phaseolus lunatus*. Proceedings of the National Academy of Sciences of the United States of America 96(1):284-289.

Mcgaw BA, Horgan R. 1983. Cytokinin Oxidase from *Zea mays* Kernels and *Vinca rosea* Crown-Gall Tissue. Planta 159(1):30-37.

Mengiste T, Revenkova E, Bechtold N, Paszkowski J. 1999 Aug. An SMC-like protein is required for efficient homologous recombination in *Arabidopsis*. Embo Journal 18(16):4505-4512.

Menke M, Chen IP, Angelis KJ, Schubert I. 2001 Jun. DNA damage and repair in *Arabidopsis thaliana* as measured by the comet assay after treatment with different classes of genotoxins. Mutation Research-Genetic Toxicology and Environmental Mutagenesis 493(1-2):87-93.

Metwally A, Tsonev T, Zeinalov Y. 1997. Effect of cytokinins on the photosynthetic apparatus in water-stressed and rehydrated bean plants. Photosynthetica 34(4):563-567.

Miles GP, Samuel MA, Zhang Y, Ellis BE. 2005 Nov. RNA interference-based (RNAi) suppression of AtMPK6, an *Arabidopsis* mitogen-activated protein kinase, results in hypersensitivity to ozone and misregulation of AtMPK3. Environ Pollut 138(2):230-237.

Mironov V, De Veylder L, Van Montagu M, Inze D. 1999 Apr. Cyclin-dependent kinases and cell division in plants - The nexus. Plant Cell 11(4):509-521.

Mohamed MF, Kang DW, Aneja VP. 2002 Jun. Volatile organic compounds in some urban locations in United States. Chemosphere 47(8):863-882.

Oono Y, Seki M, Nanjo T, Narusaka M, Fujita M, Satoh R, Satou M, Sakurai T, Ishida J, Akiyama K, Iida K, Maruyama K, Satoh S, Yamaguchi-Shinozaki K, Shinozaki K. 2003 Jun. Monitoring expression profiles of *Arabidopsis* gene expression during rehydration process after dehydration using ca 7000 full-length cDNA microarray. Plant J 34(6):868-887.

Pang QS, Hays JB, Rajagopal I. 1993 Apr. Two cDNAs from the plant *Arabidopsis thaliana* that partially restore recombination proficiency and DNA-damage resistance to Escherichia coli mutants lacking recombination-intermediate-resolution activities. Nucleic Acids Research 21(7):1647-1653.

Peddada SD, Lobenhofer EK, Li LP, Afshari CA, Weinberg CR, Umbach DM. 2003 May. Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference. Bioinformatics 19(7):834-841.

Petersen M, Brodersen P, Naested H, Andreasson E, Lindhart U, Johansen B, Nielsen HB, Lacy M, Austin MJ, Parker JE, Sharma SB, Klessig DF, Martienssen R, Mattsson O, Jensen AB, Mundy J. 2000 Dec. *Arabidopsis* map kinase 4 negatively regulates systemic acquired resistance. Cell 103(7):1111-1120.

Pospisilova J, Dodd IC. 2005. Role of plant growth regulators in stomatal limitation to photosynthesis during water stress. In: Handbook of Photosynthesis. Pessarakli M, editor, Taylor and Francis, New York. 811-825.

Pospisilova J, Synkova H, Rulcova J. 2000. Cytokinins and water stress. Biologia Plantarum 43(3):321-328.

Rajeevan MS, Vernon SD, Taysavang N, Unger ER. 2001 Feb. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. Journal of Molecular Diagnostics 3(1):26-31.

Revenkova E, Masson J, Koncz C, Afsar K, Jakovleva L, Paszkowski J. 1999 Jan. Involvement of *Arabidopsis thaliana* ribosomal protein S27 in mRNA degradation triggered by genotoxic stress. Embo Journal 18(2):490-499.

Riccardi F, Gazeau P, Jacquemot MP, Vincent D, Zivy M. 2004 Dec. Deciphering genetic variations of proteome responses to water deficit in maize leaves. Plant Physiol Biochem 42(12):1003-1011.

Ries G, Heller W, Puchta H, Sandermann H, Seidlitz HK, Hohn B. 2000 Jul. Elevated UV-B radiation reduces genome stability in plants. Nature 406(6791):98-101.

Rizhsky L, Liang HJ, Mittler R. 2002 Nov. The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiology 130(3):1143-1151.

Ruegger M, Dewey E, Gray WM, Hobbie L, Turner J, Estelle M. 1998 Jan. The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast grr1p. Genes Dev 12(2):198-207.

Rulcova J, Pospisilova J. 2001. Effect of benzylaminopurine on rehydration of bean plants after water stress. Biologia Plantarum 44(1):75-81.

Rydberg B, Lindahl T. 1982. Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. EMBO J 1(2):211-216.

Sambrook J, Russel I. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Sanchez P, de Torres ZM, Grant M. 2000 Feb. AtBI-1, a plant homologue of Bax inhibitor-1, suppresses Bax-induced cell death in yeast and is rapidly upregulated during wounding and pathogen challenge. Plant J 21(4):393-399.

Santerre A, Britt AB. 1994 Mar. Cloning of a 3-methyladenine-DNA glycosylase from *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 91(6):2240-2244.

Schafer S, Schmulling T. 2002 Sep. The CRK1 receptor-like kinase gene of tobacco is negatively regulated by cytokinin. Plant Molecular Biology 50(2):155-166.

Schaller F. 2001 Jan. Enzymes of the biosynthesis of octadecanoid-derived signalling molecules. J Exp Bot 52(354):11-23.

Schendel PF, Defais M, Jeggo P, Samson L, Cairns J. 1978 Aug. Pathways of mutagenesis and repair in Escherichia coli exposed to low levels of simple alkylating agents. J Bacteriol 135(2):466-475.

Schliep A, Schonhuth A, Steinhoff C. 2003. Using hidden Markov models to analyze gene expression time course data. Bioinformatics 19 Suppl 1:i255-i263.

Schnittger A, Schobinger U, Bouyer D, Weinl C, Stierhof YD, Hulskamp M. 2002 Apr. Ectopic D-type cyclin expression induces not only DNA replication but also cell division in *Arabidopsis* trichomes. Proceedings of the National Academy of Sciences of the United States of America 99(9):6410-6415.

Schwartz JL. 1989 Apr. Monofunctional alkylating agent-induced S-phase-dependent DNA damage. Mutat Res 216(2):111-118.

Shinozaki K, Yamaguchi-Shinozaki K. 1996 Apr. Molecular responses to drought and cold stress. Curr Opin Biotechnol 7(2):161-167.

Shiu SH, Bleecker AB. 2001 Sep. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci U S A 98(19):10763-10768.

Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH. 2004 May. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. Plant Cell 16(5):1220-1234.

Shooter KV, Howse R, Shah SA, Lawley PD. 1974 Feb. The molecular basis for biological inactivation of nucleic acids. The action of methylating agents on the ribonucleic acid-containing bacteriophage R17. Biochem J 137(2):303-312.

Singer B. 1985. In vivo Formation and Persistence of Modified Nucleosides Resulting from Alkylating-Agents. Environmental Health Perspectives 62(OCT):41-48.

Singer VL, Jones LJ, Yue ST, Haugland RP. 1997 Jul. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Anal Biochem 249(2):228-238.

Soni R, Carmichael JP, Shah ZH, Murray JAH. 1995 Jan. A Family of Cyclin D-Homologs from Plants Differentially Controlled by Growth-Regulators and Containing the Conserved Retinoblastoma Protein-Interaction Motif. Plant Cell 7(1):85-103.

Stahlberg A, Hakansson J, Xian X, Semb H, Kubista M. 2004 Mar. Properties of the reverse transcription reaction in mRNA quantification. Clin Chem 50(3):509-515.

Suarez MF, Filonova LH, Smertenko A, Savenkov EI, Clapham DH, von AS, Zhivotovsky B, Bozhkov PV. 2004 May. Metacaspase-dependent programmed cell death is essential for plant embryogenesis. Curr Biol 14(9):R339-R340.

Taylor RM, Hamer MJ, Rosamond J, Bray CM. 1998 Apr. Molecular cloning and functional analysis of the *Arabidopsis thaliana* DNA ligase I homologue. Plant Journal 14(1):75-81.

Trobacher CP, Senatore A, Greenwood JS. 2006 Apr. Masterminds or minions? Cysteine proteinases in plant programmed cell death. Canadian Journal of Botany-Revue Canadienne de Botanique 84(4):651-667.

Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K. 2006 Apr. Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Curr Opin Biotechnol 17(2):113-122.

Uren AG, O'Rourke K, Aravind LA, Pisabarro MT, Seshagiri S, Koonin EV, Dixit VM. 2000 Oct. Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. Mol Cell 6(4):961-967.

Vailleau F, Daniel X, Tronchet M, Montillet JL, Triantaphylides C, Roby D. 2002 Jul. A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. Proceedings of the National Academy of Sciences of the United States of America 99(15):10179-10184.

van Attikum H, Bundock P, Overmeer RM, Lee LY, Gelvin SB, Hooykaas PJJ. 2003 Jul. The *Arabidopsis* AtLIG4 gene is required for the repair of DNA damage, but not for the integration of Agrobacterium T-DNA. Nucleic Acids Research 31(14):4247-4255.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002 Jun. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3(7):Research 0034.

Vierstra RD. 1996 Oct. Proteolysis in plants: mechanisms and functions. Plant Mol Biol 32(1-2):275-302.

Volkov RA, Panchuk II, Schoffl F. 2003 Oct. Heat-stress-dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT-PCR. J Exp Botany 54(391):2343-2349.

Watanabe N, Lam E. 2005 Apr. Two *Arabidopsis* metacaspases AtMCP1b and AtMCP2b are arginine/lysine-specific cysteine proteases and activate apoptosis-like cell death in yeast. J Biol Chem 280(15):14691-14699.

Wong ML, Medrano JF. 2005 Jul. Real-time PCR for mRNA quantitation. Biotechniques 39(1):75-85.

Yoshiba Y, Kiyosue T, Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. 1997 Oct. Regulation of levels of proline as an osmolyte in plants under water stress. Plant Cell Physiol 38(10):1095-1102.

Yu YB, Yang SF. 1980. Biosynthesis of Wound Ethylene. Plant Physiology 66(2):281-285.

Zhang J, Van Toai T, Huynh L, Preiszner J. 2000. Development of flooding-tolerant *Arabidopsis thaliana* by autoregulated cytokinin production. MOLECULAR BREEDING 6(2):135-144.

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. 2004 Sep. GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. Plant Physiol 136(1):2621-2632.