Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 68 (2007) 2163–2175
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

On the processing of metabolic information through metabolite–gene communication networks: An approach for modelling causality

Jedrzej Szymanski [a], Monika Bielecka [a], Fernando Carrari [a,b], Alisdair R. Fernie [a], Rainer Hoefgen [a], Victoria J. Nikiforova [a,c,*]

[a] *Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany*
[b] *Instituto de Biotechnología, CICVyA, Instituto Nacional de Technología Agrícola Argentina, Buenos Aires, Argentina*
[c] *Timiryazev Institute of Plant Physiology, Russian Academy of Sciences, Moscow 127276, Russia*

Received 19 January 2007; received in revised form 5 April 2007
Available online 4 June 2007

Abstract

Gene-metabolite correlation networks of three independent biological systems were interrogated using an approach to define, and subsequently model, causality. The major goal of this work was to analyse how information from those metabolites, that displayed a rapid response to perturbation of the biological system, is processed through the response network to provide signal-specific adaptation of metabolism. For this purpose, comparison of network topologies was carried out on three different groups of system elements: transcription factors, other genes and metabolites, with special emphasis placed on those features which are possible sites of metabolic regulation or response propagation. The degree of connectivity in all three analysed gene-metabolite networks followed power-law and exponential functions, whilst a comparison of connectivities of the various cellular entities suggested, that metabolites are less involved in the regulation of the sulfur stress response than in the ripening of tomatoes (in which metabolites seem to have an even greater regulatory role than transcription factors). These findings reflect different degree of metabolic regulation for distinct biological processes. Implementing causality into the network allowed classification of metabolite-gene associations into those with causal directionality from gene to metabolite and from metabolite to gene. Several metabolites were positioned relatively early in the causal hierarchy and possessed many connections to the downstream elements. Such metabolites were considered to have higher regulatory potential. For the biological example of hypo-sulfur stress response in Arabidopsis, the highest regulatory potential scores were established for fructose and sucrose, isoleucine, methionine and sinapic acid. Further developments in profiling techniques will allow greater cross-systems comparisons, necessary for reliability and universality checks of inferred regulatory capacities of the particular metabolites.

Keywords: *Solanum lycopersicum*; Solanaceae; Tomato; *Arabidopsis thaliana*; Brassicaceae; Sulfur starvation; Fruit ripening; Correlation network; Causality; Metabolic regulation

1. Introduction

A living organism consists of many highly diverse molecular entities (such as genes, proteins, metabolites) organized in a functional dynamic system. This system must be capable of simultaneously maintaining homeostasis and reacting to changes within its environment. In order to perform both functions, the systemic response to perturbation constitutes a branched chain of consecutive changes of cellular entities. Information concerning these changes propagates along this chain, forming a dense network of interactions. Given that changes in states of downstream elements occur as a result of preceding upstream changes, it follows that infor-

* Corresponding author. Address: Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany. Tel.: +49 331 5678254; fax: +49 331 5678250.
E-mail address: nikiforova@mpimp-golm.mpg.de (V.J. Nikiforova).

0031-9422/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2007.04.017

mation processing induced by a perturbation (signal) is causally directed – starting from an exciter (a cause) and its perception through transduction and transformation to an endpoint response (an effect). For biologists who attempt to decipher these chains of response reactions and to elucidate regulatory points within them, the intrinsic causality of information processing networks represents a useful framework that, in combination with empirical biological knowledge, can be used in the analysis of biological reasoning. In order to attempt such evaluation two basic features are required within the dataset under consideration: (i) the connections between recorded changes reflecting their putative mutual dependencies must be determinable and (ii) some biological knowledge concerning temporal points of excitation (or final effect) must be available. For this reason kinetic experiments appear to provide the richest source of information for such studies. Time series of transcript and metabolite profiling data, allowing cross-correlation analysis and obtained in the experiments with plants, when the response was induced by a treating agent recognized as a response exciter, clearly satisfy both of the above mentioned prerequisites. The present study is focused on a particular aspect of the information processing in plants, which is the transmission of information from a metabolite to a gene. To access this question, we decipher causality in the integrated gene-metabolite correlation networks.

Whilst the **integrated analysis of metabolite and gene expression profiles** is recognized nowadays as a powerful tool for biotechnology (Fernie et al., 2005), and is increasingly useful in gene annotation studies (Tohge et al., 2005), relatively little attempt has been made to decipher information flow across and between these molecular entities. Initial experiments in which transcript and metabolic profiles were analysed in parallel revealed important co-regulatory behaviour between different molecular species (Askenazi et al., 2003; Urbanczyk-Wochniak et al., 2003). Subsequently from more detailed analysis of the systemic response of *Arabidopsis thaliana* to sulfur deficiency it was possible to identify clusters of co-regulated genes and metabolites, that were interpretable on the basis of a priori biological knowledge (Hirai et al., 2004, 2005). Similarly, several aspects of the regulation of metabolism were revealed following the interrogation of the integrated gene-metabolite network of tomato fruit ripening (Carrari et al., 2006) and folate responsiveness in mice (Ernest et al., 2006). When taken together, data from these examples (and several subsequent studies: Scheible et al., 2004; Osuna et al., 2007) suggest that there is often a considerable delay in the temporal response of metabolites with respect to transcripts. These time response lags have to be taken into account when biological reasoning in the informational exchange between genes and metabolites is accessed.

Approaches to infer **causality** have been developed and applied for gene co-expression networks. These are generally based on dynamic measurements of response which yield hierarchical information about causal relationships within networks. Such analysis has been performed to dissect gene networks following hormone and insulin signaling data (Kam, 2002), whilst information regarding the time lag between species at which the highest correlation was found has also been used as a method to infer causality (D'haeseleer et al., 2000). However, such reasoning cannot be directly extrapolated from pure gene co-expression networks to gene–metabolite correlation networks due to the multi-scale nature of metabolic and gene expression responses. To address to this problem, we previously elaborated a new approach for implementing causal directionality into gene-metabolite correlation networks with the use of the a priori knowledge on the molecule, which excites the systems response and can thus be considered as a 'cause' (Nikiforova et al., 2005b). In such networks, propagation of the information flow from the exciter to physiological endpoints through alterations in gene expressions and metabolite concentrations can be followed. Keeping the strict statistical regime about significance of observed patterns in correlation networks, from this analysis we can conclude biologically meaningful information, even if observed interactions are putative or indirect.

In the current work we apply this approach to extract from gene–metabolite correlation network causally directed associated pairs metabolite-to-gene and estimate the putative regulatory potential of the metabolites by means of comparative network topology analysis. Three exemplary gene–metabolite networks (sulfur stress response in *A. thaliana*, application of inhibitors of folate biosynthesis to *A. thaliana*, and the tomato fruit ripening process) are considered. In an attempt to identify putative regulators of the adaptive response, capable to influence inter alia gene expression, we distinguish those pairs of gene–metabolite associations, in which dependency is putatively directed from metabolite to gene. Three groups of biosystem elements with regulatory potential were considered: metabolites, transcription factors as known regulators of gene expression and other genes. Putative expressional regulation activity of metabolites, which we put in the focus of the current research, has been recently recognized (Tucker and Breaker, 2005; Grundy and Henkin, 2006; Ladurner, 2006) and is not yet well understood.

Data from the two successful implementations are discussed in the context of current models of metabolic regulation of gene expression. In addition, the reasons behind the inability to implement causality in the third dataset, as well as current limitations and future prospects of this approach, are discussed.

2. Results and discussion

2.1. Suitable datasets for reconstruction of causally directed gene–metabolite correlation networks

External signals excite a chain of response reactions in biological systems, finally leading to an adaptation of the biological system to accommodate to its changing environ-

ment. The final physiological response and the signal which elicits it are generally linked by a complex chain, in which information is processed through consecutive propagation of information-bearing changes from upstream to downstream elements within the chain. A multitude of data collected at the metabolite and transcript levels provide correlative evidence that both types of molecular entities are involved in information processing networks. First, there are many instances in the literature documenting that both transcripts and metabolites change in response to external perturbation (see for example Nikiforova et al., 2004; Nikiforova et al., 2005a). Secondly, distinct alterations in transcript levels (which can be considered as an informational signal) may lead to multiple changes at the level of the metabolite (see for example Bohmert et al., 2000; Wrobel-Kwiatkowska et al., 2004; Bino et al., 2005; Busov et al., 2006) and vice versa, treatments with external metabolites lead to multiple changes in the abundance of transcripts (Kiddle et al., 2003; Hoth et al., 2002; Rashotte et al., 2003; Goda et al., 2004; Uppalapati et al., 2005). These data suggest that both transcripts and metabolites are involved in the transmission of information concerning external perturbation. They, furthermore, imply that informational exchange between genes and metabolites occurs in both directions. In the present study we focus on a single aspect of such interactions of informational exchange in the direction from metabolite to gene.

To identify putative metabolic regulators of response control in plant biosystems, we used an approach of correlation network analysis, which allows reconstruction of gene–metabolite interactions. Such analysis seems to be especially suitable for simultaneous studies of systems' responses at the levels of transcripts and of metabolites, as the nature of mutual influences between these system elements is not well understood, and in correlation networks connections do not directly represent a physical interaction between nodes, but rather indicate co-expression or co-regulation of the elements in the conditions under study. Items with similar patterns of co-regulation are usually considered to be more likely functionally associated, due to a variety of different biological reasons. These functional associations imply an exchange of information between items whilst the entire correlation network represents a sum of such associations.

For the production of the initial datasets on constitutive changes in transcript and metabolite levels in plants in response to external signal, the system in steady-state was perturbed and the response development in a time scale was registered. Then, transcript and metabolite profiles of control and perturbed samples were taken and analysed. After purification of the obtained dataset from noise, correlation matrices were calculated and gene–metabolite communication networks were reconstructed from all significantly correlating pairs gene–gene, gene–metabolite and metabolite–metabolite. For the question of possible metabolic regulation of response control through e.g., altered gene expression, the gene–metabolite correlating pairs are of particular interest, especially if causal relationships gene-to-metabolite and metabolite-to-gene can be distinguished.

To be suitable for such analysis, correlation networks have to (i) represent statistically reliable correlations, (ii) contain both transcripts and metabolites and (iii) be causally directed. Table 1 contains the descriptions of the datasets under analysis in the present study, with corresponding numbers of analytes used for network reconstruction. Three biological examples are considered: response to hypo-sulfur stress and to herbicide treatments in Arabidopsis and a developmental process of fruit ripening in tomato.

2.2. Topology analysis of integrated metabolite–gene correlation networks

2.2.1. Analysis of distribution of associated pairs of network elements by correlation coefficient

Correlation networks reconstructed from the above three datasets were further characterized regarding distribution density of correlation strength between their elements (reflected by calculated Pearson correlation coefficients, Fig. 1). In the case of the response development of Arabidopsis plants to sulfur deficiency and herbicide treatment (Fig. 1.1a and 1.2a) distribution densities were strongly biased toward positive correlations, comparing to normal distributions in shuffled datasets used as no-information-containing controls. Similar question on significance of the observed shift to positive correlations between performance and neuronal response in monkey was raised in the discussion between Roelfsema (2002) and Hasegawa et al. (2002). For our biological examples, it is conceivable that this observation may reflect a preference for feedforward rather than feedback control. Interestingly, no such bias was observed in the dataset in

Table 1
Initial datasets used for reconstruction of gene–metabolite correlation networks (tf, transcription factors)

#	Organism	Experiment	Series length	tf genes	Non-tf genes	Metabolites	Reference
1	Thale cress (*Arabidopsis thaliana* Col-0 G1)	Sulfur starvation	12 values	424 (7%)	5629	30	Nikiforova et al. (2003, 2004, 2005a, 2006)
2	Thale cress (*Arabidopsis thaliana* Col-0 G1)	Treatment with inhibitors of folate biosynthesis, Asulam and Methotrexate	6 values	425 (6.9%)	5737	78	Unpublished
3	Tomato (*Solanum lycopersicum*)	Fruit ripening	10 values	164 (4.9%)	3169	92	Carrari et al. (2006)

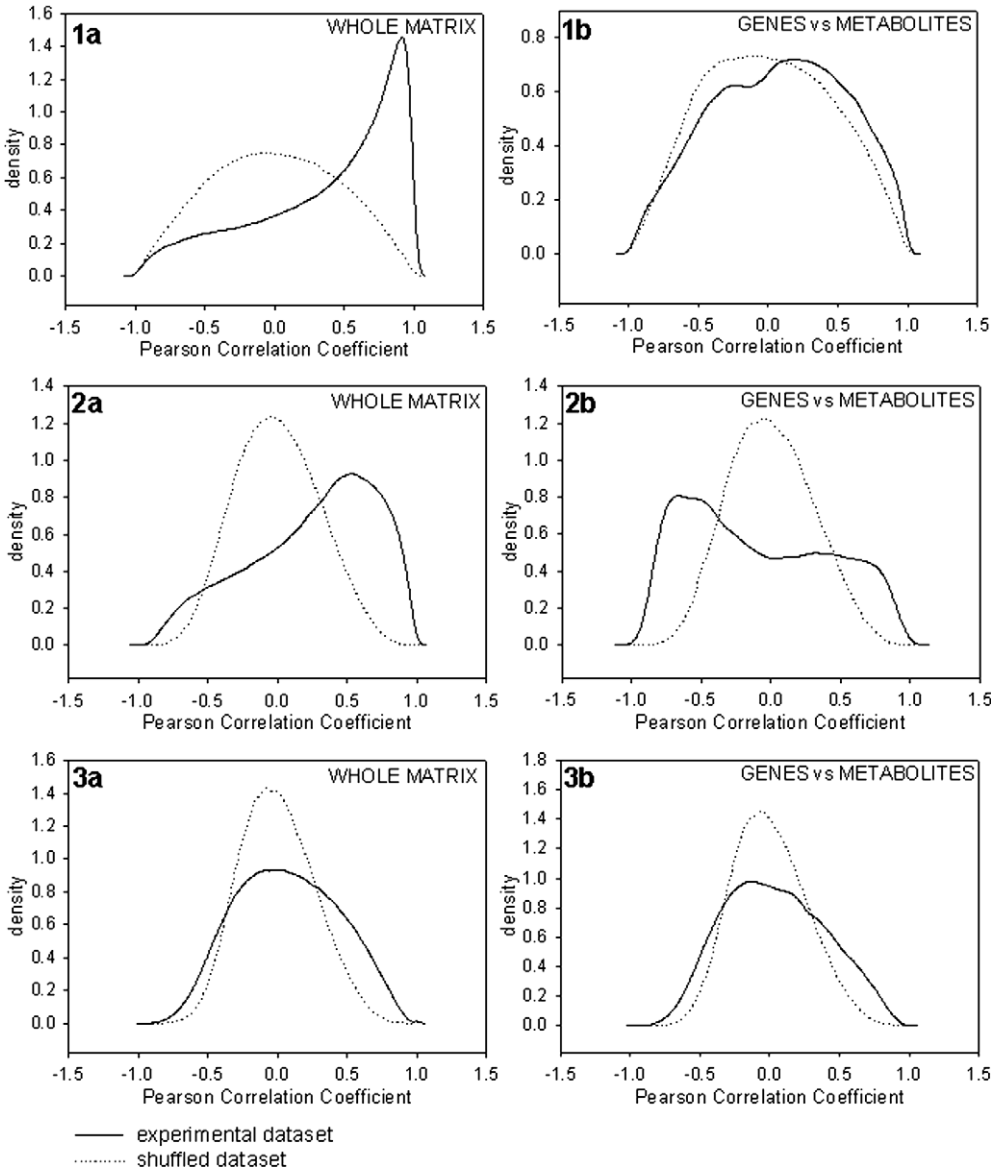

Fig. 1. The density of Pearson correlation coefficients for the whole distance matrices (a) and for the subset of mixed gene-metabolite associated pairs (b), in analysed datasets of Arabidopsis plants responding to either herbicide treatment (1) or sulfur deficiency (2), or of ripening tomato fruits (3). The density function describes the probability for an associated pair to exhibit a correlation coefficient within a defined interval; if the whole probability area equals 1, then the probability to find an association in any particular interval [a, b] is equal to the integral of the function *f* from a to b. Dotted lines describing probability densities for no-information controls (shuffled datasets), are given for comparison.

which metabolic and transcript changes in the process of tomato fruit ripening were monitored. Here, associations are more evenly distributed between the area of positive and negative correlations (Fig. 1.3a), demonstrating no preference in favor of either feed forward or feed back mutual influences.

When connectivities in the subset of mixed gene-metabolite associations were analysed, no shifts towards positive correlations were observed (Fig. 1.1b and 1.3b). Moreover, in the correlation network describing response to sulfur deficiency, correlations in the mixed gene–metabolite associated pairs were shifted to the negative area (Fig. 1.2b), pointing at possible feed-back mechanism of metabolic regulation. Thus, analysis of connectivity distribution densities pointed at possible preferential mechanism of the information processing in the examined biological examples.

2.2.2. Analysis of distribution of network elements by number of connections (connectivity degree)

To analyse whether the examined networks exhibit a typical power-law topology, we used non-linear least squares for distribution function parameters estimation and Akaike Information Criteria (AIC) for goodness-of-fit determination (Sakamoto et al., 1986). These analyses suggested that a power law topology was inherent in the correlation networks of Arabidopsis plants responding to sulfur deficiency (Fig. 2.1) and to herbicide treatment (Fig. 2.2). However, it is worthy to note that an exponen-

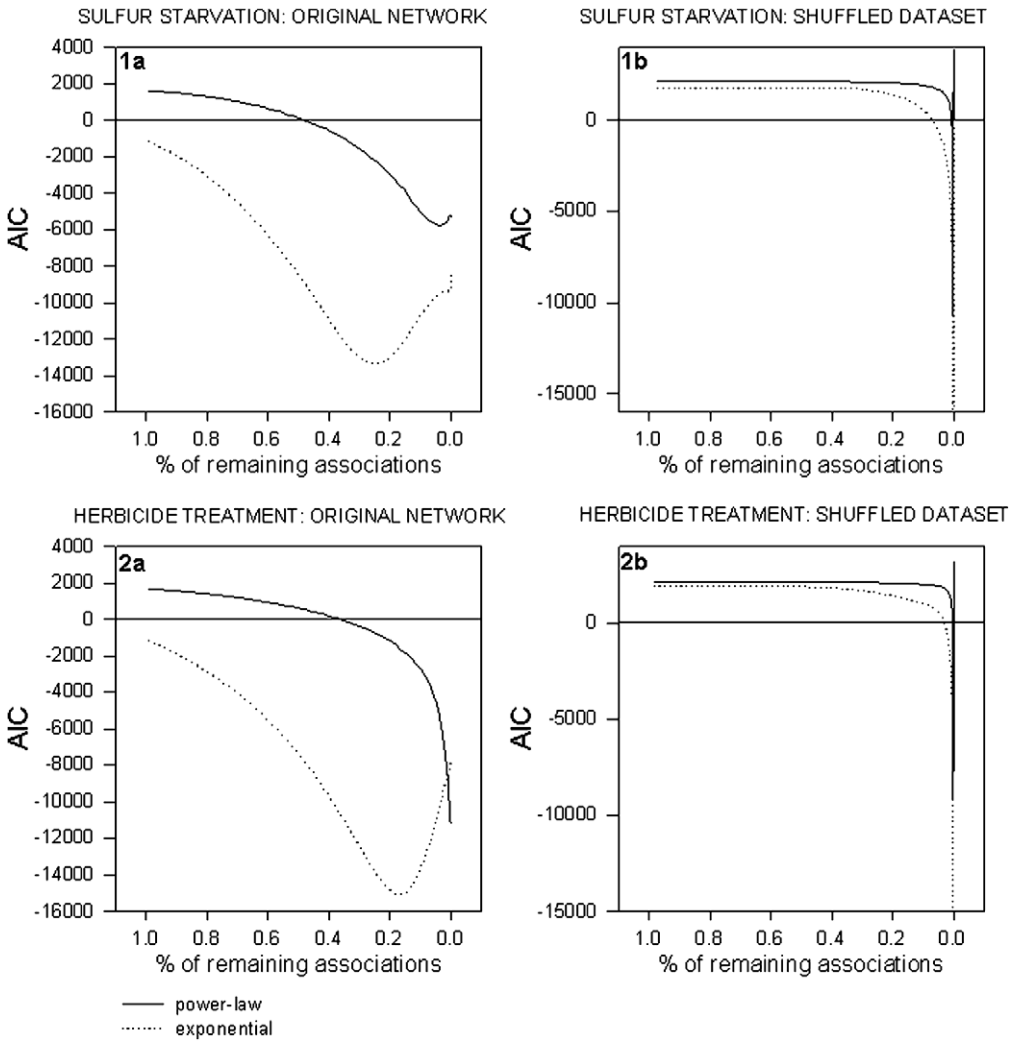

Fig. 2. The cumulative connectivity degree distribution of the networks analysed. The cumulative connectivity degree distribution was analysed in the context of goodness-of-fit to exponential and power-law distributions. Networks were iteratively filtered for the strongest correlations starting from 1 (for fully connected networks) and ending on 0 (for a graph displaying completely disconnected nodes, as described in Section 4). The fitness score is in inversely dependence with the AIC i.e., the lower the AIC, the better the fit. For corresponding network reconstructions, original (1a) and shuffled (1b) data on sulfur starvation and herbicide treatment data (original, 2a, and shuffled, 2b) were used.

tial distribution curve seems to better describe the data at lower significance thresholds. Comparison of AIC plots of experimental data with the shuffled sets of data revealed no significant fit neither to power-law nor to exponential distribution for shuffled dataset (unless only a couple of nodes remains in the network), implying that conclusions based on the experimental data are not merely artefactual.

Occurrence of a power-law and exponential degree distribution properties in correlation networks suggests a high structural organization of analyzed system. The same distribution characteristics were obtained for approximately the same percentage of remaining associations in both analysed networks. It indicates that despite different experimental background, reconstructed networks are structurally comparable. This fact may have a strong natural reason, as all the observed correlations come from similar underlying cellular systems.

Even though all our networks represent similar structural organisation, a deeper analysis of network structure compatibility has to be considered. Inclusion of more datasets, different organisms and more experimental conditions into analysis should let us statistically examine relationships between a particular biological system, source of variation and correlation network structure.

2.3. Comparing connectivities in three groups of analytes: metabolites, transcription factors and other genes

To better characterize the functional role of metabolites in information processing and in order to estimate their capacity in regulation of adaptive responses, their preferential association with genes was compared with association frequencies of other combinations of molecular entities. For this purpose, besides metabolites, two groups of network elements were considered separately: transcription factors (tf) (which have a well-established regulatory function), and other genes. In the case of the sulfur-deficient Arabidopsis dataset, the highest enrichment was found in

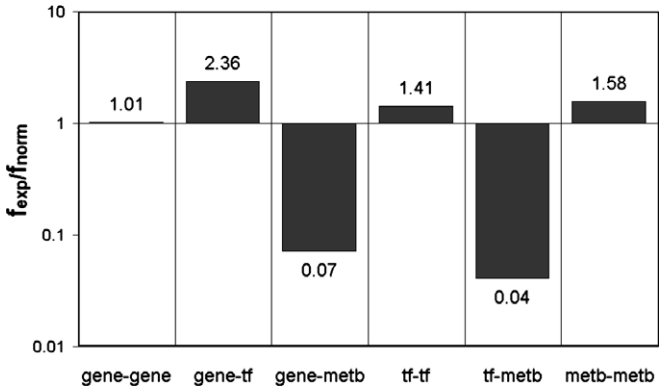

Fig. 3. Evaluation of true associations. The ratios of observed association frequency to the proportional distribution (associations returned on the basis of chance alone) in the correlation network of the sulfur deficiency stress response of Arabidopsis plants are depicted. Datasets were split into three groups: metabolites, transcription factors and other genes, denoted as ''genes''.

the frequencies of associations between transcription factors and other genes (Fig. 3). Homogeneous associations tf–tf and metabolite–metabolite were slightly enriched in comparison with random associations, whilst the number of associations of gene–metabolite and tf–metabolite were much less frequent than were associations made by chance indicating that metabolites contribute little to the regulatory events under these biological circumstances.

Comparison of metabolite connectivities with those of tf and non-tf genes for the process of fruit ripening in tomato and hypo-sulfur stress response in Arabidopsis is given in Table 2. Here, over-represented connectivities are indicated by connectivity enrichments higher than 1, and under-represented enrichments are less than 1. Interestingly, comparison between these events suggests that metabolites appear to play a greater role during tomato fruit ripening (in which they display a connectivity enrichment of 1296 which is even higher than that displayed by tfs, 0,896). This finding is particularly interesting given recent interest in understanding the metabolic shifts during fruit development and

Table 2
Comparative analysis of metabolite connectivities with those of transcription factors (tf) and non-tf genes in gene–metabolite correlation networks reconstructed from profiling data on ripening tomato fruits and sulfur-stressed *Arabidopsis thaliana* seedlings

	# Vertices	# Connections	Connectivity enrichment
Tomato, ripening:			
Total	3425	4080	
Non-tf genes	3169	3763	0,997
tf genes	164	175	0,896
Metabolites	92	142	1,296
Arabidopsis thaliana, *S response*:			
Total	2014	238410	
Non-tf genes	1843	216310	0,991
tf genes	141	20589	1,234
Metabolites	30	1511	0,425

Connectivity enrichment was calculated from average connectivities for metabolites, tf and non-tf genes divided by total average connectivities.

the hypothesis of the importance of metabolites in ripening should be directly tractable given the widespread availability of mutants and transgenics of ripening and metabolism (Giovannoni, 2004; Carrari and Fernie, 2006). The fact that metabolites seem to be more influential in tomato ripening than in the response of Arabidopsis to sulfur stress may well reflect that there is a different degree of metabolism-mediated regulation in distinct biological processes.

2.4. Approaching causality in correlation networks

Network reconstruction based on correlation matrices does not imply causal directionality, and the direction of informational exchange between elements is not defined in the reconstruction procedure, as can be illustrated with the correlation network reconstructed from the dataset 2 of herbicide-treated Arabidopsis plants.

However, as we demonstrated previously (Nikiforova et al., 2005b), in some correlation networks causal directionality can be implemented, when the initial exciter (a 'cause') of the response development is known and is present as one of the elements in the network. For the network of hypo-sulfur stress response, such exciter is represented by sulfur molecule itself, based on an assumption, that internal sulfur level is the first to change in response to altered external sulfur. As the values for decreasing internal sulfur in the time scale of starvation experiment were measured, it was possible to give the 'causal' hierarchical ordering to all elements in the correlation network reconstructed from transcript and metabolite profiles of sulfur stress response: those elements which strongly correlated to sulfur were put at the beginning of the putative information processing flow (correlators of the first order), those correlating with the first order correlators were assigned to the second position in the causal hierarchy, and so on, down to the very endpoint reaction elements. Following this principle, by measuring a distance from the exciter each element in the response network can be assigned to a numbered causal level. The step-by-step techniques of getting the systems elements assigned to the numbered classes of causal hierarchy, starting from sulfur as the response exciter (causal level 0), is described in Section 4.3 of the Experimental, ''Ordering network elements by causal hierarchy''.

For another network considered in the present study, which describes the process of tomato fruit ripening (dataset 3), such initial exciter evidently can not be identified, though here causal directionality was also implemented to the network. For this we utilised the reverse logics of gained knowledge on molecules, which are the last to accumulate in the ripening tomato fruits (i.e., neoxanthin, β-carotene and lycopene, Carrari et al., 2006) and thus can be considered as an 'effect' in establishing cause-to-effect relationships.

Now, as in the reconstructed response network we got, on the one hand, both genes and metabolites, and, on the other hand, general causal directionality, we could distinguish those metabolite-containing associations, in which

putative information flow is directed from a metabolite to a gene. Metabolites from such associated pairs were assumed as putative regulators of response development.

2.5. Identification of metabolite-to-gene relationships in directed correlation networks

In the analysed networks general number of causal connections was high, forming a dense network with fast exhaustion of the involved system elements as the distance from sulfur or from the ripening end-point increased. The high network densities are reflected in their relatively small diameter, which in turn depends on the character of connectivity distributions. Network diameter can be characterized by the number of hierarchical classes in the network, where each class corresponds to a causal level, determined by distance from the initial exciter in a correlation network. In the example of sulfur deficiency response the latest class in causal hierarchy was 12, and a major portion of connections was already exhausted before class 9 (Fig. 4). Together with the identified close fit to power-law connectivity distribution, this finding attributes the examined network to the graphs displaying 'small world' behaviour (Fell and Wagner, 2000), which may serve to minimize transition times between metabolic states (Wagner and Fell, 2001). The network diameter is generally considered as a measure of systems functionality (Dartnell et al., 2005) and appears to be intertwined to the level of the systems complexity. As was demonstrated by A.K. Seth (2005) in reconstruction of causal interactions in neural dynamics, the networks which describe rich adaptive behaviour show a higher density of causal interactions and a stronger input-to-output causal flow.

Table 3 compiles the data on number of connections to upstream and downstream elements for 30 different metabolites in the hypo-sulfur stress response network. With these data, we estimated the capacity for expressional regulation of the response process for the considered metabolites as a portion of connections to downstream ('regulated') elements, expressed as a % of the total number of connections for a particular metabolite. In this analysis, the network distance of a metabolite from sulfur had to be accounted for, since the probability to have many connections with downstream elements decreases with the increasing class number. Furthermore, the higher a metabolite is in the causal hierarchy the larger the portion of the network it can potentially influence. For this reason we provide class number for every metabolite in Table 3, and compare their connectivities primarily with those of other metabolites of their own class.

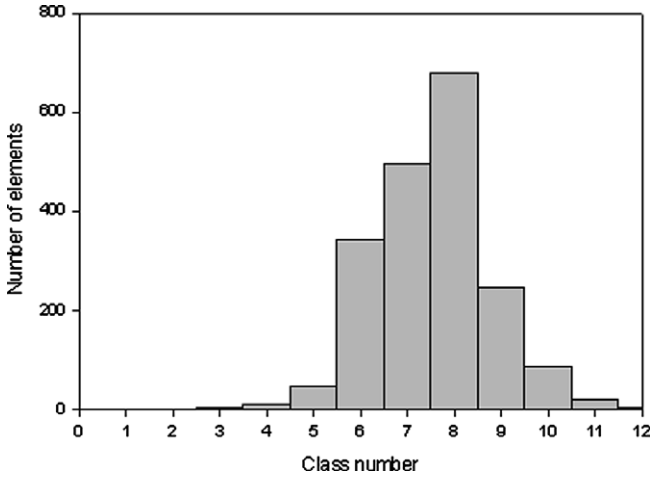

Fig. 4. Distribution of elements between network classes in the sulfur deficiency response correlation network. The class number indicates the distance of the element from the initial response exciter of the network as defined in Section 4.

Amongst the parameters we examined it seems likely that the total number of connections in the correlation network may reflect the functional importance of a given element in information processing. To consider all three parameters – class number, total connectivity and connectivity to downstream elements in our evaluation, we implemented a combined regulatory potential (RP) score by multiplying out total connectivity and connectivity to downstream elements, divided by class number.

This analysis revealed several interesting features. When considering metabolites of class 5, isoleucine possesses the highest number of connections to the downstream elements in the causal hierarchy. In the case of metabolites from class 6, fructose stands out from the others in that the major portion of its connections are to downstream elements, whilst, valine, sucrose and glutamine also possess a larger number of connections to the downstream elements, than the other metabolites of class 6. In class 7, the majority of connections have been already exhausted in higher causal classes 1–6. Nevertheless, three metabolites from this class, glycine, putrescine and tryptophan, still have significant number of downstream connections. Metabolites of classes 9 and 11 can be assumed to have little regulatory potential in the biological scenario under study, as their involvement into regulation, if any, is too late to be significant.

The same metabolites, which are connected to the significant number of downstream elements, demonstrated also high RP scores. The highest RP score was registered for fructose, then isoleucine, sinapic acid, and methionine and sucrose with similarly high scores of combined regulatory potential.

Importantly, regulatory roles for several of the metabolites with high RP scores and significant numbers of the downstream connections have already been described in the literature. From some of the reported experiments, the possible mode of action of these metabolites on gene expression, either direct or indirect, can be elucidated. Among such metabolites, **methionine** in gram-positive bacteria is involved into the regulation of gene expression through the S-box transcription termination control system (Grundy and Henkin, 1998; Murphy et al., 2002). Expression of the S-box genes is induced in response to limitation for methionine and depends on S-adenosyl-L-methionine (SAM), concentration of which is sensed directly by the

Table 3
Distribution of metabolite connections between upstream and downstream elements in causally directed gene–metabolite correlation network (reconstructed from transcript and metabolite profiling data on the hypo-sulfur stress response of *Arabidopsis thaliana* plants; in the last column connectivity enrichment in tomato fruit ripening network is given for comparison)

Metabolite	Class #	Connectivity enrichment	# Connections to upstream elements (%)	# Connections to elements of the same class (%)	# Connections to downstream elements (%)	Regulatory potential score	Connectivity enrichment in tomato fruit ripening network
Sulfur	0	0.02	0	0	100		Not measured
Methionine	5	0.38	2	56	42	3.2	Disconnected
Isoleucine	5	0.53	2	37	62	6.6	Disconnected
c-sinapic acid	5	0.49	2	47	52	5.1	Not measured
Arginine	5	0.07	25	50	25	0.3	Disconnected
Valine	6	0.37	11	73	16	1.0	0.84
Threonine	6	0.58	30	67	3	0.3	1.68
Sucrose	6	1.35	50	36	14	3.2	Disconnected
Proline	6	1.13	37	54	8	1.5	Disconnected
Phenylalanine	6	0.54	39	61	0	0	10.91
Glycerol-1-P	6	0.47	36	61	4	0.3	Disconnected
Glycerol	6	0.36	51	49	0	0	0.84
Glutamine	6	0.22	12	77	12	0.4	Disconnected
Fructose	6	1.36	3	56	41	9.3	Disconnected
b-alanine	6	0.54	33	67	0	0	Disconnected
Aspartic acid	6	0.02	50	50	0	0	Disconnected
Tryptophan	7	0.18	52	38	10	0.2	12.59
Shikimic acid	7	0.16	16	79	5	0.1	0.84
Putrescine	7	0.08	33	44	22	0.2	0.84
Phosphoric acid	7	2.42	41	59	0	0.1	Disconnected
Malic acid	7	0.24	7	83	10	0.4	0.84
Glycine	7	0.13	27	40	33	0.6	Disconnected
Glyceric acid	7	0.35	5	93	2	0.1	Disconnected
Glutamic acid	7	0.21	12	84	4	0.1	Disconnected
Fumaric acid	7	0.39	15	78	7	0.4	Disconnected
Serine	9	0.10	17	42	42	0.5	0.84
g-aminobutyric acid	9	0.05	33	50	17	0.1	Disconnected
Raffinose	11	0.02	50	0	50	0.1	Not measured
Succinic acid				Disconnected			Disconnected
Myo-inositol				Disconnected			Disconnected

Class number was defined as network distance from sulfur, sulfur class number was assigned to 0.
Connectivity enrichment was calculated as number of connections divided by average connectivity of all network elements.

nascent transcript in the absence of a trans-acting factor (McDaniel et al., 2003). In eukaryotes disturbed in sulfur metabolism methionine may be also indirectly involved into the regulation of gene expression as an immediate precursor of SAM. SAM is the main methyl group donor in DNA methylation reactions, an important epigenetic factor in the regulation of gene expression. When in diets of experimental animals methionine content was manipulated, resulting imbalance in SAM was critical for the long-term regulation of gene expression (Rees, 2002). SAM, in, turn is a precursor of sulfur-containing nucleoside 5′-methylthioadenosine, which has been also shown to influence regulation of gene expression in mammalian cells (Avila et al., 2004).

Sucrose is a well-known modulator of gene expression in plants. When the effects of sucrose at a genome-wide level of *A. thaliana* wild type and mutant plants were studied, a significant amount of genes strongly responsive to sucrose were identified (Gonzali et al., 2006; Lloyd and Zakhleniuk, 2004; Osuna et al., 2007). Similar analysis in sucrose-starved rice revealed consensus cis-elements in the promoter regions of many sucrose starvation-upregulated genes, being the promising response activators (Wang et al., 2007). In potato, sucrose inhibited the promoter of glucosyltransferase gene (Korobczak et al., 2005) and induced the promoter activity of 14-3-3 protein gene (Szopa et al., 2003). In spite of the vast evidence for sucrose being a strong modulator of gene expression, the direct mechanisms of its action are still not fully understood.

Isoleucine, together with threonine, regulates the expression of the threonine operon enzymes of *Escherichia coli* through the attenuation response determined by the threonine and isoleucine codons in the leader transcript (Lynn et al., 1987; Vitreschak et al., 2004).

The action of **glutamine** on gene expression is studied for some time in mammalian cells. From these studies, it appears that glutamine may regulate gene expression by, at least, two different mechanisms: one through the gluta-

mine-induced cell swelling, and another through its intracellular metabolism (Quillard et al., 1997; Lavoinne et al., 1998).

Lyubetsky et al. (2006) proposed recently the model for attenuating mRNA regulation of gene expression via transcription termination, exemplified with **tryptophan**.

The **putrescine** observation is quite interesting since polyamines have been shown to be involved in regulation of gene expression (Koza and Herbst, 1992; Tabib and Bachrach, 1994; Morgan, 1999). Of particular note is the finding that they are feedback regulators of methionine synthase gene expression, involved in sulfate assimilation (Kenyon et al., 1996). It is worth mentioning here that within our sulfur stress response network, putrescine had two downstream connected genes as putative 'causal sinks'. One of these genes, At1g62180, is also a gene of sulfate assimilation, encoding putative adenosine-5′-phosphosulfate reductase; the other, At4g23990, encodes cellulose synthase catalytic subunit-like protein. All these findings indicate the indirect linkage of polyamines with the regulation of gene expression. The recent study by Lindemose et al. (2005) demonstrated that polyamines at submillimolar concentrations bind preferentially to bent adenine tracts in double-stranded DNA. These results provide the evidence for the sequence-specific binding of polyamines to DNA, and thereby suggest a mechanism by which the cellular effects of polyamines in terms of regulation of gene expression could, at least partly, be a direct consequence of sequence-specific interactions of polyamines with promoters at the DNA sequence level.

Notably, from amongst a dozen metabolites known so far, for which direct regulation of gene expression through riboswitch mechanism has been reported, only glycine is present in our experimental set of metabolites, and its RP score is the highest for its class (7). (For reviews on riboswitch mechanism see e.g., Grundy and Henkin, 2006; Winkler, 2005. Data on riboswitch-related metabolites are collected in RibEx database [<http://132.248.32.45:8080/cgi-bin/ribex.cgi>], Abreu-Goodger and Merino, 2005; or as a special subset in RegRNA database [<http://bidlab.life.nctu.edu.tw/RegRNA2/website/>], Bengert and Dandekar, 2003.)

Whilst the retrieval of known regulatory metabolites from our data set is promising, that of previously unreported metabolites could ultimately prove more informative. However, it is likely that considerably more direct experimentation is required in order to corroborate these.

While accessing the question on how common/specific are the identified regulatory properties of the analysed metabolites, we compared their connectivity enrichment in the networks of sulfur stress response and tomato fruit ripening (Table 3). The conservation is very low between these networks, and little can be gleaned by way of comparative analysis since the overlap between metabolites detected in both systems which remain connected in both networks is low. It thus remains possible, that the role of metabolites in mediating response relay is rather conditional than universal, but from the limited number of studies to date it is too premature to reach such a conclusion.

3. Conclusions

The approach elaborated here allowed the modelling of causality in gene-metabolite correlation networks reconstructed from series of transcript and metabolite profiles. For the purposes of estimating regulatory potential (RP) of metabolites it was possible to identify associations between genes and metabolites that were causally connected in the direction from metabolite gene. We hypothesize that the metabolites possessing the highest RP score were involved in the regulation of the adaptive response. With the use of this approach metabolites known to regulate gene expression were retrieved from our data set. It should be borne in mind, that analyses such as those described here are in their infancy and as yet the number of testable datasets is still strictly limited (an observation that is reflected in the fact that we were unable to reconstruct a gene–metabolite correlation network with implemented causality for the herbicide treated Arabidopsis samples), and it is thus difficult to draw general conclusions on the influence of metabolites on network behaviour. Interestingly, however, the fact that the networks resulting from two distinctly different biological processes have some conserved but many different features suggests, that some aspects of regulation are likely conditional. Further problems come from the facts, that metabolite profiling methods in particular but also transcript profiling experiments in some species offer far from absolute coverage, and that available data generally does not allow for turnover of the molecular entities to be taken into account, when data integration is performed. That said despite these limitation the findings reported here provide a list of candidates of metabolites that are putatively important conduits of information transmission. The combination of enhanced metabolite coverage and increased data quality, that can be anticipated in the near future given advances in analytic techniques and the adoption of reporting standards by the metabolomics community, will likely allow us to test the generality of our findings in the not too distant future. However, irrespective of this the direct testing of these candidates by artificially modifying their content via metabolite feeding studies or by genetic means is likely to be of high importance in the validation of these hypotheses.

4. Experimental

4.1. Data collection

Dataset 1: the *A. thaliana* hypo-sulfur stress response data consisted of combined transcript and metabolite profiles, obtained and normalized as described previously (Nikiforova et al., 2003, 2004, 2005a, 2006). In brief,

A. thaliana Col-G1 plants that had been pre-grown on full-nutrition solid agarose or liquid medium, were transferred into sulfur-deficient medium and later sampled after 2, 6, 10 and 13 days of sulfur starvation. Transcript profiles were performed by means of array hybridization whilst metabolite profiles were mainly obtained following the application of established GC-MS and LC-MS protocols.

Dataset 2 of the *A. thaliana* herbicide treatment consisted of combined transcript and metabolite profiles. To obtain them, *A. thaliana* Col-G1 plants growing on standard solid agarose medium, were treated with two herbicides asulam and methotrexate (MTX), which are known inhibitors of folate biosynthesis. After pre-growth for one week, seedlings were transferred to the medium containing 5×106 M asulam or 5 μg/l MTX. Plants were harvested (in five replicates per treatment) three and six days after transfer. For both time points, samples of the plants transferred to herbicide-free media, were taken as a control. For transcript profiling, macroarray hybridisation and data normalization/analysis was performed exactly as described previously (Nikiforova et al., 2003). In order to obtain metabolite profiles, GC-MS analysis was performed, as described by Nikiforova et al. (2005a).

Dataset 3 of tomato fruit ripening was obtained as described by Carrari et al. (2006). In brief, *Solanum lycopersicum* plants were grown in a greenhouse and fruit samples were harvested at various time points *post* anthesis. Microarray hybridization was carried out as described in Urbanczyk-Wochniak et al. (2006), whilst GC-MS analysis was performed as described by Roessner-Tunali et al. (2003). Both methods were optimized for use with tomato fruits.

Constituents of all three datasets are summarized in Table 1.

4.2. Reconstruction of statistically reliable correlation networks

Datasets were cross-normalized by scaling each species profile into a range from 0 to 1 (a scaling that preserves all relative differences between observations). A previously described algorithm (Nikiforova et al., 2005a) was modified to allow precise statistics on correlation significance and was subsequently used for relevant network reconstruction.

1. MetaGeneAlyse software (Daub et al., 2003) was used for min–max data normalisation and data shuffling with an over-sampling rate of 500. For full details on how this was implemented see the User Manual on the MetaGeneAlyse web page (<http://metagenealyse.mpimp-golm.mpg.de/>).
2. A Pearson correlation coefficient-based distance matrix was subsequently calculated for each experimental dataset and its corresponding shuffled (randomized) dataset. The R 2.2.1 software environment was used for all computations and data handling (Gentleman and Ihaka, 1996).
3. Matrices were filtered using two significance criteria: *p*-value for $\alpha = 0.01$ and a 1% random threshold. For *p*-value filtering, the significance thresholds were set as two degrees of freedom, and α at 0.01. For 1% threshold filtering, an improved shuffling approach procedure was utilized. Following this approach, the correlation distributions of the original and shuffled (randomised) datasets were compared, and the significance threshold value for the Pearson correlation coefficient R was set as the level, at which number of pair associations from the shuffled dataset constituted not more than 1% to the number of associations from the original dataset. All associations correlating at a level lower than the set threshold were filtered out from the further network analysis.
4. Network graphs were built from the list of the selected associated pairs using a combination of R language for data analysis and graphics (Gentleman and Ihaka, 1996) and Pajek software (Batagelj and Mrvar, 2003) to facilitate display of all selected network elements as interconnected nodes.
5. As a control for network topology analysis two types of artificial networks were reconstructed. In the first type of reconstruction, the number of nodes in the network was conserved but the localization of edges (between nodes) was varied randomly. In the second type of reconstruction, an artificial scale free network with conserved λ, number of nodes and other structural parameters was generated.

4.3. Ordering network elements by causal hierarchy

Causal relationships between particular network elements were elucidated based on an original approach described by Nikiforova et al. (2005b). Pajek software for network analysis was used to assign sulfur in the sulfur stress response network to a hierarchical class '0'. Hierarchical layers of all the other network elements were subsequently determined by the program, depending on the distance from sulfur and downloaded in a form of a list of elements with the corresponding hierarchical class number. Within the associated pairs the nodes with the smaller class number were regarded 'regulating', and those with the larger class numbers as 'regulated' elements (so-called 'causal sources' and 'causal sinks', a term by Seth (2005)).

At the next step, from the associated pairs 'gene–gene', 'metabolite–metabolite' and 'gene–metabolite' the later were selected in order to search for potential metabolic regulators of gene expression. Within these pairs, the mutual order of class numbers for genes and metabolites was analysed. Here three situations are possible: the gene possessed a smaller class number than the associated metabolite, the class numbers of gene and metabolite are equal or the class number for the metabolite is smaller than that of the associated gene. Metabolites of the later case were considered as putative regulators for their associated genes.

For the network describing tomato fruit ripening, the final causal element (a causal sink) in the network was assigned to a hierarchical class '0'. The hierarchical classes for all the other elements in the network were subsequently determined by the program as −1 (significantly correlating and thus directly connected to the latest element of class '0'), −2 (directly connected to any of the elements from class '−1') and so on. For ease of comparison with the other networks these were then re-annotated in reverse order. By this procedure, elements which were the most distant from the final causal element were identified as being at the beginning of the causal hierarchy and thus were assigned the lowest class number.

For the network of the response to herbicides neither the initial exciter (a causal source) nor the final response element (a causal sink) could be identified in the independent experiments, and causal relationships in the reconstructed gene–metabolite correlation network could not be obtained.

4.4. Degree distribution analysis in study of network topologies

In order to evaluate, if the examined networks exhibit defined topology, the non-linear least square approach for distribution function parameters estimation and the Akaike Information Criteria (AIC) for goodness-of-fit determination (Sakamoto et al., 1986) were applied. The same procedure was performed for both linear and exponential distribution functions. For this purpose, datasets were iteratively filtered using increasingly stringent significance thresholds (starting with a completely connected network and ending with totally disconnected nodes). In each step an additional 1% of associations were filtered out, distribution function parameters were estimated and the AIC was computed in order to describe changes in degree distribution characteristics for all iterations.

Acknowledgement

This work was supported by the Max Planck Society.

References

Abreu-Goodger, C., Merino, E., 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. Nucleic Acids Res. 33 (Suppl S), W690–W692.

Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J.V., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C., Madden, K.T., 2003. Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. Nat. Biotechnol. 21, 150–156.

Avila, M.A., Garcia-Trevijano, E.R., Lu, S.C., Corrales, F.J., Mato, J.M., 2004. Methylthioadenosine. Int. J. Biochem. Cell Biol. 36, 2125–2130.

Batagelj, V., Mrvar, A., 2003. Pajek – Analysis and Visualization of Large Networks. In: Jünger, M., Mutzel, P. (Eds.), Graph Drawing Software. Springer, Berlin, pp. 77–103.

Bengert, P., Dandekar, T., 2003. A software tool-box for analysis of regulatory RNA elements. Nucleic Acids Res. 31, 3441–3445.

Bino, R.J., de Vos, C.H.R., Lieberman, M., Hall, R.D., Bovy, A., Jonker, H.H., Tikunov, Y., Lommen, A., Moco, S., Levin, I., 2005. The light-hyperresponsive high pigment-2(dg) mutation of tomato: alterations in the fruit metabolome. New Phytol. 166, 427–438.

Bohmert, K., Balbo, I., Kopka, J., Mittendorf, V., Nawrath, C., Poirier, Y., Tischendorf, G., Trethewey, R.N., Willmitzer, L., 2000. Transgenic Arabidopsis plants can accumulate polyhydroxybutyrate to up to 4% of their fresh weight. Planta 211, 841–845.

Busov, V., Meilan, R., Pearce, D.W., Rood, S.B., Ma, C.P., Tschaplinski, T.J., Strauss, S.H., 2006. Transgenic modification of gai or rgl1 causes dwarfing and alters gibberellins, root growth, and metabolite profiles in Populus. Planta 224, 288–299.

Carrari, F., Fernie, A.R., 2006. Metabolic regulation underlying tomato fruit development. J. Exp. Bot. 57, 1883–1897.

Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanor, M-I., Nunes-Nesi, A., Nikiforova, V., Centero, D., Ratzka, A., Pauly, M., Sweetlove, L., Fernie, A.R., 2006. Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behaviour. Plant Physiol. 142, 1380–1396.

Dartnell, L., Simeonidis, E., Hubank, M., Tsoka, S., Bogle, I.D.L., Papageorgiou, L.G., 2005. Robustness of the p53 network and biological hackers. FEBS Lett. 579, 3037–3042.

Daub, C.O., Kloska, S., Selbig, J., 2003. MetaGeneAlyse: analysis of integrated transcriptional and metabolite data. Bioinformatics 19, 2332–2333.

D'haeseleer, P., Liang, S., Somogyi, R., 2000. Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics 16, 707–726.

Ernest, S., Carter, M., Shao, H.F., Hosack, A., Lerner, N., Colmenares, C., Rosenblatt, D.S., Pao, Y.H., Ross, M.E., Nadeau, J.H., 2006. Parallel changes in metabolite and expression profiles in crooked-tail mutant and folate-reduced wild-type mice. Hum. Mol. Genet. 15, 3387–3393.

Fell, D.A., Wagner, A., 2000. The small world of metabolism. Nat. Biotechnol. 18, 1121.

Fernie, A.R., Urbanczyk-Wochniak, E., Willmitzer, L., 2005. Parallel metabolite and transcript profiling – Hypothesis generation for biotechnology. In: Metabolome Analyses: Strategies for Systems Biology. Springer Science + Business Media, Inc., New York, pp. 291–306.

Gentleman, R., Ihaka, R., 1996. R: A language for data analysis and graphics. J. Comput. Graph. Stat. 5, 299–314.

Giovannoni, J.J., 2004. Genetic regulation of fruit development and ripening. Plant Cell 16 (Suppl S), S170–S180.

Goda, H., Sawa, S., Asami, T., Fujioka, S., Shimada, Y., Yoshida, S., 2004. Comprehensive comparison brassinosteroid-regulated of auxin-regulated and brassinosteroid-regulated genes in arabidopsis. Plant Physiol. 134, 1555–1573.

Gonzali, S., Loreti, E., Solfanelli, C., Novi, G., Alpi, A., Perata, P., 2006. Identification of sugar-modulated genes and evidence for in vivo sugar sensing in Arabidopsis. J. Plant Res. 119, 115–123.

Grundy, F.J., Henkin, T.M., 1998. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. Mol. Microbiol. 30, 737–749.

Grundy, F.J., Henkin, T.M., 2006. From ribosome to riboswitch: Control of gene expression in bacteria by RNA structural rearrangements. Crit. Rev. Biochem. Mol. Biol. 41, 329–338.

Hasegawa, R.P., Blitz, A.M., Geller, N.L., Goldberg, M.E., 2002. Do neurons predict the future? Response. Science 295, 227a.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., Saito, K., 2004. Integration

of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA 101, 10205–10210.

Hirai, M.Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D.B., Yamazaki, Y., Kanaya, S., Nakamura, Y., Kitayama, M., Suzuki, H., 2005. Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem 280, 25590–25595.

Hoth, S., Morgante, M., Sanchez, J.P., Hanafey, M.K., Tingey, S.V., Chua, N.H., 2002. Genome-wide gene expression profiling in Arabidopsis thaliana reveals new targets of abscisic acid and largely impaired gene regulation in the abi1-1 mutant. J Cell Sci. 115, 4891–4900.

Kam, Z., 2002. Generalized analysis of experimental data for interrelated biological measurements. Bull. Math. Biol. 64, 133–145.

Kenyon, S.H., Nicolaou, A., Ast, T., Gibbons, W.A., 1996. Stimulation in vitro of vitamin B-12-dependent methionine synthase by polyamines. Biochem. J. 316, 661–665.

Kiddle, G., Pastori, G.M., Bernard, S., Pignocchi, C., Antoniw, J., Verrier, P.J., Foyer, C.H., 2003. Effects of leaf ascorbate content on defense and photosynthesis gene expression in *Arabidopsis thaliana*. Antioxid. Redox Signal. 5, 23–32.

Korobczak, A., Aksamit, A., Lukaszewicz, M., Lorenc, K., Rorat, T., Szopa, J., 2005. The potato glucosyltransferase gene promoter is environmentally regulated. Plant Sci. 168, 339–348.

Koza, R.A., Herbst, E.J., 1992. Deficiencies in DNA replication and cell-cycle progression in polyamine-depleted HeLa cells. Biochem. J. 281, 87–93.

Ladurner, A.G., 2006. Rheostat control of gene review expression by metabolites. Mol. Cell 24, 1–11.

Lavoinne, A., Meisse, D., Quillard, M., Husson, A., Renouf, S., Yassad, A., 1998. Glutamine and regulation of gene expression in rat hepatocytes: The role of cell swelling. Biochimie 80, 807–811.

Lindemose, S., Nielsen, P.E., Mollegaard, N.E., 2005. Polyamines preferentially interact with bent adenine tracts in double-stranded DNA. Nucleic Acids Res. 33, 1790–1803.

Lloyd, J.C., Zakhleniuk, O.V., 2004. Responses of primary and secondary metabolism to sugar accumulation revealed by microarray expression analysis of the Arabidopsis mutant, pho3. J. Exp. Bot. 55, 1221–1230.

Lynn, S.P., Burton, W.S., Donohue, T.J., Gould, R.M., Gumport, R.I., Gardner, J.F., 1987. Specificity of the attenuation response of the threonine operon of *Escherichia coli* is determined by the threonine and isoleucine codons in the leader transcript. J. Mol. Biol. 194, 59–69.

Lyubetsky, V.A., Rubanov, L.I., Seliverstov, A.V., Pirogov, S.A., 2006. Model of gene expression regulation in bacteria via formation of RNA secondary structures. Mol. Biol. 40, 440–453.

McDaniel, B.A.M., Grundy, F.J., Artsimovitch, I., Henkin, T.M., 2003. Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. Proc. Natl. Acad. Sci. USA 100, 3083–3088.

Morgan, D.M., 1999. Polyamines. An overview. Mol. Biotechnol. 11, 229–250.

Murphy, B.A., Grundy, F.J., Henkin, T.M., 2002. Prediction of gene function in methylthioadenosine recycling from regulatory signals. J. Bacteriol. 184, 2314–2318.

Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H., Hoefgen, R., 2003. Transcriptome analysis of sulfur depletion in Arabidopsis thaliana: Interlacing of biosynthetic pathways provides response specificity. Plant J. 33, 633–650.

Nikiforova, V., Gakière, B., Kempa, S., Adamik, M., Willmitzer, L., Hesse, H., Hoefgen, R., 2004. Towards dissecting nutrient metabolism in plants: a systems biology case study on sulfur metabolism. J. Exp. Bot. 55, 1861–1870.

Nikiforova, V.J., Kopka, J., Tolstikov, V., Fiehn, O., Hopkins, L., Hawkesford, M.J., Hesse, H., Hoefgen, R., 2005a. Systems rebalancing of metabolism in response to sulfur deprivation, as revealed by metabolome analysis of Arabidopsis plants. Plant Physiol. 138, 304–318.

Nikiforova, V.J., Daub, C.O., Hesse, H., Willmitzer, L., Hoefgen, R., 2005b. Integrative gene-metabolite network with implemented causality deciphers informational fluxes of sulphur stress response. J. Exp. Bot. 56, 1887–1896.

Nikiforova, V.J., Bielecka, M., Gakière, B., Krueger, S., Rinder, J., Kempa, S.R., Morcuende, R., Scheible, W.-R., Hesse, H., Hoefgen, R., 2006. Effect of sulfur availability on the integrity of amino acid biosynthesis in plants. Amino Acids 30, 173–183.

Osuna, D., Usadel, B., Morcuende, R., Gibon, Y., Blasing, O.E., Hohne, M., Gunter, M., Kamlage, B., Trethewey, R., Scheible, W.R., Stitt, M., 2007. Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. Plant J. 49, 463–491.

Quillard, M., Renouf, S., Husson, A., Meisse, D., Lavoinne, A., 1997. Glutamine and regulation of gene expression in mammalian cells – Special reference to phosphoenolpyruvate carboxykinase (PEPCK). Biochimie 79, 125–128.

Rashotte, A.M., Carson, S.D.B., To, J.P.C., Kieber, J.J., 2003. Expression profiling of cytokinin action in arabidopsis. Plant Physiol. 132, 1998–2011.

Rees, W.D., 2002. Manipulating the sulfur amino acid content of the early diet and its implications for long-term health. Proc. Nutr. Soc. 61, 71–77.

Roelfsema, P.R., 2002. Do neurons predict the future? Science 295, 227a.

Roessner-Tunali, U., Hegemann, B., Lytovchenko, A., Carrari, F., Bruedigam, C., Granot, D., Fernie, A.R., 2003. Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. Plant Physiol. 133, 84–99.

Sakamoto, T., Ishiguro, M., Kitagawa, G., 1986. Akaike Information Criterion Statistics. Kluwer Academic Publishers, Norwell, MA, USA.

Scheible, W.R., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, N., Schindelasch, D., Thimm, O., Udvardi, M.K., Stitt, M., 2004. Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. Plant Physiol. 136, 2483–2499.

Seth, A.K., 2005. Causal connectivity of evolved neural networks during behavior. Network-Comp. Neural 16, 35–54.

Szopa, J., Lukaszewicz, M., Aksamit, A., Korobczak, A., Kwiatkowska, D., 2003. Structural organisation, expression, and promoter analysis of a 16R isoform of 14-3-3 protein gene from potato. Plant Physiol. Bioch. 41, 417–423.

Tabib, A., Bachrach, U., 1994. Activation of the proto-oncogene c-*myc* and c-*fos* by c-*ras*: involvement of polyamines. Biochem. Biophys. Res. Commun. 202, 720–727.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M., Saito, K., 2005. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J. 42, 218–235.

Tucker, B.J., Breaker, R.R., 2005. Riboswitches as versatile gene control elements. Curr. Opin. Struct. Biol. 15, 342–348.

Uppalapati, S.R., Ayoubi, P., Wenig, H., Palmer, D.A., Mitchell, R.E., Jones, W., Bender, C.L., 2005. The phytotoxin coronatine and methyl jasmonate impact multiple phytohormone pathways in tomato. Plant J. 42, 201–217.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., Fernie, A.R., 2003. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Rep. 4, 989–993.

Urbanczyk-Wochniak, E., Usadel, B., Thimm, O., Nunes-Nesi, A., Carrari, F., Davy, M., Blasing, O., Kowalczyk, M., Weicht, D., Polinceusz, A., Meyer, S., Stitt, M., Fernie, A.R., 2006. Conversion of

MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. Plant Mol. Biol. 60, 773–792.

Vitreschak, A.G., Lyubetskaya, E.V., Shirshin, M.A., Gelfand, M.S., Lyubetsky, V.A., 2004. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. FEMS Microbiol. Lett. 234, 357–370.

Wagner, A., Fell, D.A., 2001. The small world inside large metabolic networks. Proc. Roy. Soc. Lond. B: Bio. Sci. 268, 1803–1810.

Wang, H.J., Wan, A.R., Hsu, C.M., Lee, K.W., Yu, S.M., Jauh, G.Y., 2007. Transcriptomic adaptations in rice suspension cells under sucrose starvation. Plant Mol. Biol. 63, 441–463.

Winkler, W.C., 2005. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. Curr. Opin. Chem. Biol. 9, 594–602.

Wrobel-Kwiatkowska, M., Lorenc-Kukula, K., Starzycki, M., Oszmianski, J., Kepczynska, E., Szopa, J., 2004. Expression of beta-1,3-glucanase in flax causes increased resistance to fungi. Physiol. Mol. Plant Pathol. 65, 245–256.