

TECHNICAL ADVANCE

MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes

Oliver Thimm¹, Oliver Bläsing¹, Yves Gibon¹, Axel Nagel², Svenja Meyer², Peter Krüger¹, Joachim Selbig¹, Lukas A. Müller³, Seung Y. Rhee³ and Mark Stitt^{1,*}

¹Max Planck Institute for Molecular Plant Physiology, Golm, Germany,

²RZPD German Resource Centre for Genome Research, Heubnerweg 6, D-14059 Berlin, Germany, and

³TAIR, The Arabidopsis Information Resource, Department of Plant Biology, Carnegie Institution of Washington, Stanford, Germany

Received 1 October 2003; revised 4 December 2003; accepted 9 December 2003.

*For correspondence (fax +49 331 567 8101; e-mail mstitt@mpimp-golm.mpg.de).

Summary

MAPMAN is a user-driven tool that displays large data sets onto diagrams of metabolic pathways or other processes. SCAVENGER modules assign the measured parameters to hierarchical categories (formed 'BINs', 'subBINs'). A first build of TRANSCRIPTSCAVENGER groups genes on the *Arabidopsis* Affymetrix 22K array into >200 hierarchical categories, providing a breakdown of central metabolism (for several pathways, down to the single enzyme level), and an overview of secondary metabolism and cellular processes. METABOLITE-SCAVENGER groups hundreds of metabolites into pathways or groups of structurally related compounds. An IMAGEANNOTATOR module uses these groupings to organise and display experimental data sets onto diagrams of the users' choice. A modular structure allows users to edit existing categories, add new categories and develop SCAVENGER modules for other sorts of data. MAPMAN is used to analyse two sets of 22K Affymetrix arrays that investigate the response of *Arabidopsis* rosettes to low sugar: one investigates the response to a 6-h extension of the night, and the other compares wild-type Columbia-0 (Col-0) and the starchless *pgm* mutant (plastid phosphoglucomutase) at the end of the night. There were qualitatively similar responses in both treatments. Many genes involved in photosynthesis, nutrient acquisition, amino acid, nucleotide, lipid and cell wall synthesis, cell wall modification, and RNA and protein synthesis were repressed. Many genes assigned to amino acid, nucleotide, lipid and cell wall breakdown were induced. Changed expression of genes for trehalose metabolism point to a role for trehalose-6-phosphate (Tre6P) as a starvation signal. Widespread changes in the expression of genes encoding receptor kinases, transcription factors, components of signalling pathways, proteins involved in post-translational modification and turnover, and proteins involved in the synthesis and sensing of cytokinins, abscisic acid (ABA) and ethylene revealing large-scale rewiring of the regulatory network is an early response to sugar depletion.

Keywords: data display, expression profiling, metabolite profiling, sugar sensing.

Introduction

Technologies like whole-genome expression arrays (Celis *et al.*, 2000; De Risi *et al.*, 1997; Michaut *et al.*, 2003; Wang *et al.*, 2003) and mass spectrometry (MS)-based metabolite profiling (Fiehn *et al.*, 2000; Stitt and Fernie, 2003) generate huge multiparameter data sets, which would have been unimaginable a few years ago. Their exploitation is limited

by our ability to interpret them. Many studies just use them to earmark candidate genes. To realise their potential to provide a comprehensive analysis of system responses, it will be necessary to combine them with a portfolio of interpretational tools. While many tools are available to analyse data sets by clustering and supervised machine

learning, relatively few allow the data to be organised and displayed by the user in the context of pre-existing biological knowledge.

Example of tools that allow data sets to be viewed in the context of biological pathways, gene regulatory networks or protein–protein interactions includes GENMAPP (<http://www.GenMAPP.org>), PATHWAYASSIST (<http://www.ariadnegenomics.com>), PATHWAY Processor (Grosu *et al.*, 2002) and BIOMINER (<http://voyager.bioinf.uni-sb.de/HPL/Projects/BioMiner>). Their usefulness for plant data sets is restricted. First, they were developed for microbial or animal systems, so irrelevant categories are imported and plant-specific pathways and processes are absent. In a first plant-specific application, a database that collects information about metabolic pathways in microbes and animals (<http://metacyc.org/>) was combined with the annotated *Arabidopsis* genome to generate a database of metabolic pathways (Müller *et al.*, 2003; <http://www.arabidopsis.org/tools/aracyc/>). AraCyc currently contains about 2000 gene annotations in 177 individual pathways, including manually entered 60 plant-specific pathways. The pathways are summarised figuratively on an overview map, many are available as detailed diagrams, and a tool is available to paint transcript expression onto the pre-defined overview. Second, their flexibility is restricted; for example, they often do not display family members individually. Plants have small to medium-sized families for enzymes in central metabolism and very large families for many classes of enzymes involved in biosynthetic and secondary metabolism (e.g. cytP450s, alcohol dehydrogenases, glycosyl transferases (Arabidopsis Genome Initiative, 2000)). Tools that do not resolve them will not realise the full potential of genome chips. Third, an incomplete knowledge base hampers approaches that depend on bottom-up reconstruction of pathways. Although an annotation is available for about half the *Arabidopsis* genes, a precise function has been defined for relatively few (Arabidopsis Genome Initiative, 2000; Müller *et al.*, 2003). Gene families exacerbate the problem, as we rarely know the precise location and function of individual members, even for relatively well-understood enzymes.

In a complementary approach, we have developed a tool called MAPMAN, which displays large data sets onto pictorial diagrams that symbolically depict areas of biological function. Each individual gene is represented by a discrete signal. Genes are initially organised in blocks rather than as pathways. This allows genes to be tentatively assigned, even when their function is only approximately known. By grouping genes, we also hoped to allow trends to be detected, which would be less apparent by inspection of a list of individual genes. The area of function can be a sector of metabolism, a particular cellular function (e.g. protein synthesis), a biological response (e.g. genes involved in metabolism and/or responses to a hormone) or, in

the case of the large families that encode classes of enzymes for which the function of most of the members is not well understood, a particular type of enzyme (e.g. cytochrome P450). By using hierarchical categories and diagrams with increasing detail, different functional areas can be analysed at different levels of resolution, depending on the question of interest and the amount of prior information available. A high priority was given to flexibility, as our aim was to produce a tool that allows each individual user to decide which data subsets to display and – if needed – to create new functional categories and diagrams as they learn more about the system they are studying. This paper presents the first build of MAPMAN and illustrates its application by analysing two sets of 22K Affymetrix expression arrays and a gas chromatography (GC)/MS metabolite profile that investigate the response to low sugar.

Results and discussion

Overall design of MAPMAN

MAPMAN consists of SCAVENGER modules that collect and classify the measured parameters into a set of hierarchical functional categories (BINs, subBINs ... individual enzymes), and an IMAGEANNOTATOR module imports the classifications and uses them to organise and display data at discrete locations on diagrams of the users' choice (Figure 1). The SCAVENGER and IMAGEANNOTATOR modules are separate, and no attempt is made to generate pathways or other schemes internally from within the system.

Structure of the TRANSCRIPTSCAVENGER module

Different SCAVENGER modules are used for different types of data, for example, expression arrays or metabolite profiles. Their development will be described in detail for the TRANSCRIPTSCAVENGER module, which was developed to sort the genes on the Affymetrix 22K array into a set of hierarchically organised BINs and subBINs. While the main emphasis was on genes involved in central metabolism, we undertook a general organisation of other areas of function. Assignments were based on gene annotations available in the public domain, with The Institute for Genomic Research (TIGR) release version 3.0 as the standard (<ftp://ftp.arabidopsis.org/home/tair/Microarrays/Affymetrix/>). The process involved an alternation between automatic recruitment and manual correction, and was guided by three general considerations. First, to assign as many genes as possible, in order to minimise loss of information, and because placing tentatively annotated genes in the immediate vicinity of *bona fide* family members will make it easier to assess the assignment in the light of experimental data. For this reason, genes with a 'putative' annotation were

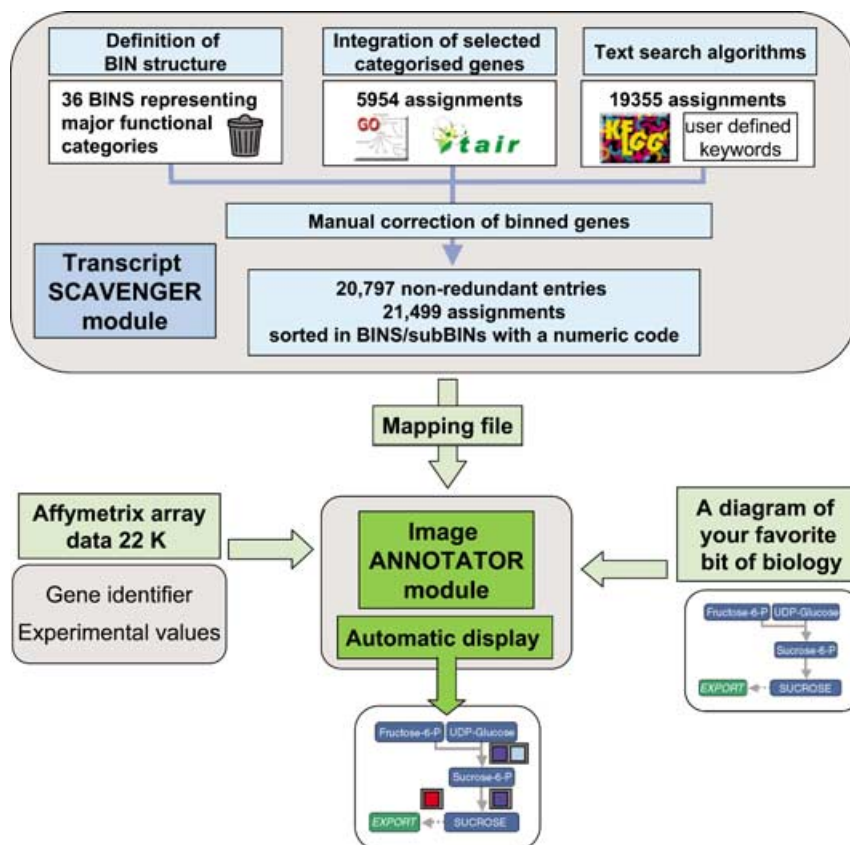


Figure 1. Overview of the structure of MAPMAN.

MAPMAN consists of SCAVENGER modules (illustrated here by the TRANSCRIPTSCAVENGER) that organises the measured parameters into functional categories, and an IMAGEANNOTATOR module that uses these categories to organise the experimental data and display it onto a scheme of the user's choice. The TRANSCRIPTSCAVENGER was developed by a combination of automatic downloads from the public domain and automatised searches, followed by manual checking. The current build organises almost 21 500 genes into 36 major functional categories or 'BINS', which are themselves split into about 160 subcategories ('subBINS'; see Table 1 and Supplementary Material for more details). The output of a SCAVENGER module is a 'mapping file', which contains a list of all the measured parameters in a particular profiling technology (e.g. all the genes represented on a 22K Affymetrix array), each with a unique identifier (e.g. the Affymetrix identifier) and allocated a numeric code corresponding to the BIN/subBIN to which it has been assigned. The IMAGEANNOTATOR imports the mapping file, plus experimental data sets (consisting, for each measured parameter of an experimental value and the unique identifier for that parameter) and images or 'maps'. The images can be custom-built, downloaded from other files or scanned in and are stored as BMP files. The sites at which the data relating to a particular BIN or subBIN are deposited on the image are selected by the user, as illustrated in Figure 2, and stored as an XML file that is associated with the corresponding BMP file. The IMAGEANNOTATOR can store a portfolio of mapping files, images and experimental data files. The mapping files are used to organise the experimental data files and to provide the code for defining what data are placed where on the image (see Figure 2).

tentatively assigned to the corresponding BIN. Second, when there was not enough information to assign most of the relevant genes to a particular BIN or subBIN, the BIN structure was modified or simplified. For example, no systematic attempt was made to assign genes to different subcellular compartments or cell types because this can only be performed with certainty for a relatively small number of genes at present. Third, as far as possible, genes were initially assigned to one BIN and, within a BIN, to one subBIN. Assignment of genes to multiple BINS (especially in a given hierarchical tree) would greatly decrease the usefulness of the data display. There would be no pressure to remove irrelevant or redundant categories, and genes whose function is least well understood would be represented at multiple sites. This criterion could be relaxed in the future.

Initially 25 BINS were manually defined or imported from the Gene Ontology Consortium (GOC; see Table 1). Each BIN was given a corresponding numerical code (e.g. 'photosynthesis' = 1), which can be extended in a hierarchical manner (e.g. the subBINS 'light reactions', 'photorespiration' and 'Calvin cycle' = 1.1, 1.2 and 1.3, respectively). In an initial download from public databases, 1476 pre-assigned genes were imported from The Arabidopsis Information Resource (TAIR) (<http://www.arabidopsis.org/tools/aracyc/>) and 2812 from the Gene Ontology Consortium (GOC) (<http://www.arabidopsis.org>; <http://www.geneontology.org>; see column 2 of Table 1). Another 3080 entries were recruited via a text search in the functionally categorised Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.ad.jp/kegg/kegg2>), and 6131 via a text search with manually pre-defined keywords of TIGR

Table 1 List of BINs, numbers of assigned genes, and information about subBINs

BIN	Imported from TAIR/GOC	Assigned in MAPMAN	BIN designation	Numbers of, and information about, the subBINs
1		152	Photosynthesis	3, See Figure 4
2		99	Major carbohydrates	6, See Figure 4, resolved to enzyme level (Figures 5 and 6)
3		78	Minor carbohydrates	7, See Figure 4
4		74	Glycolysis	Resolved to enzyme level (Figure 6)
5		10	Fermentation	–
6		11	Gluconeogenesis/glyoxylate cycle	Resolved to enzyme level (Figure 7)
7		25	OPP cycle	2, See Figure 4
8		78	TCA/organic acid transformations	3, See Figure 4 resolved to enzyme level (Figure 7)
9		111	Mitochondrial electron transport/ATP synthesis	Resolved to complex level (Figure 7)
10		479	Cell wall	8, See Figures 4 and 8
11		375	Lipid metabolism	11, See Figures 4 and 8
12		20	Nitrogen assimilation	3, See Figure 4
13		347	Amino acid metabolism	41, See Figure 4
14		12	S-assimilation	None
16		324	Secondary metabolism	7, See Figure 4
18		44	Cofactor and vitamin synthesis	–
19		38	Tetrapyrrole synthesis	–
22		12	Polyamine synthesis	–
23		125	Nucleotide metabolism	5, See Figures 4 and 8
25		24	C1-metabolism	–
	1467	2438	Subtotal for genes involved in metabolism	
26		876	Miscellaneous enzyme families	
34	397	683	Transport	9, See Figure 10
17	54	432	Hormones	23, See Figures 10 and 11
21	6	133	Redox	8, See Figures 10 and 11
30	183	948	Signalling	6, See Figure 11
15		78	Metal handling	9, See Figure 10
27	64	210	RNA	27.1, Processing
	18	76		27.2, Transcription
	518	1331		27.3, Regulation of transcription, see Figure 10
28		499	DNA	
	200	273		28.1, Synthesis/chromatin structure
		23		28.2, Repair, see Figure 10
29	34	71	Protein	29.1, Amino acid activation
	227	438		29.2, Synthesis
	47	145		29.3, Targeting
	658	820		29.4, Post-translational modification
	186	608		29.5, Degradation, see Figure 10
31	96	347	Cell	31.1, Organisation
	8	84		31.2, Division
	15	36		31.2, Cycle
		81		31.4, Vesicle transport, see Figure 10
33	107	347	Development	See Figure 10
20	59	378	Stress	20.1, Biotic
		267		20.2, Abiotic, see Figure 10
35		2027	Not assigned	35.1, No ontology
		9821		35.2, Hypothetical or unknown proteins, see Figure 10

release version 3.0. The file was then split into three for manual checking. The prior automatic sorting aided manual work, by allowing attention to be sequentially focused on different groups of genes. One subfile contained about 5800 genes, which had been automatically assigned to a single BIN/subBIN. The text of each annotation was checked to

identify obvious errors. A second subfile with >3900 entries contained genes that had been allocated to two or more BINs/subBINs. In most cases, one assignment was chosen, based on biological knowledge. Occasionally, a multiple assignment was retained, e.g. for enzymes that are known to be involved in more than one pathway (e.g. aldolase and

triose phosphate isomerase are involved in the Calvin cycle and glycolysis; enzymes involved in lipid β -oxidation are also involved in aliphatic amino acid catabolism). In some cases, multiple assignments were eradicated by redefining the interface between BINs to achieve a simpler structure. For example, extension of the pathway 'glycolysis' to include phosphoglucosmutase and UDP glucose pyrophosphorylase eradicated multiple assignments to six BINs and subBINs in starch, sucrose, minor carbohydrate and cell wall metabolism. The third subfile contained unassigned genes. Over half of these were manually assigned. The three subfiles were then combined, many BINs manually subdivided into subBINs (see Table 1), and the order of the genes in the mapping file was organised to reflect the order of enzymes in pathways, or to group genes that share a similar annotation.

Manual correction gave insights into reasons for incorrect/redundant/missed assignments, which will improve future automatic searches. Some trivial errors arose because of insufficiently stringent text searches. Numerous errors were generated by KEGG because it imports non-plant pathways and shows the 'function of pathways' graphically as a large network, rather than defining pathways individually. This resulted in numerous redundant assignments (e.g. phosphoglycerate mutase, enolase and pyruvate kinase were automatically assigned to seven pathways, including lipid synthesis and fermentation of various substrates as well as glycolysis). Import of existing categories from GOC led to similar problems because they usually provide an exhaustive rather than a specific description of function. Even AraCyc, which has clearly demarcated pathways, generated multiple assignments because the pathways are short and overlap at their interfaces.

The first build of TRANSCRIPTSCAVENGER has 11 638 entries in BINs with an ascribed function, including 689 multiple assignments. Table 1 summarises the BINs and numbers of entries and provides information about the sub-BIN structure, or provides a link to a later figure where this information can be obtained. A complete list is provided in Supplementary Material. The BIN structure can also be inspected by opening the mapping file in the downloadable MAPMAN package (see below). About 2500 entries are to specific metabolic pathways/processes, 1560 to large enzyme families, 683 to transport and 1513 to redox, hormones or signalling. Further, large groups are assigned to regulation of transcription, to protein synthesis, to modification and degradation, and to various aspects of cell organisation. In all cases, more genes have been assigned than in the original automatic recruitment from TAIR and GOC. Direct comparison underestimates the number of new recruits because >1800 multiple assignments were removed. The file still contains 2027 genes with an annotation in TIGR release version 3.0 that have not yet been assigned (BIN 35.1) and 9821 genes that are annotated as expressed or hypothetical proteins of unknown function (BIN 35.2).

Structure of the IMAGEANNOTATOR module

Three types of file are imported into the IMAGEANNOTATOR (Figure 1). (i) 'Mapping files' are imported from the SCAVENGER module. They are in EXCEL format and contain, for each measured parameter, a unique identifier (e.g. the list of the Affymetrix identifiers, a list of metabolites in a GC/MS profile in a specified vocabulary), a text annotation and a numeric code that mirrors the BIN/subBIN to which the measured parameter has been assigned by the SCAVENGER module. (ii) Experimental data sets are imported as EXCEL files and contain, for each measured parameter, the unique identifier and an experimental value (e.g. the change in expression between the treatment and a control sample, given on a \log_2 scale). The mapping file automatically organises the experimental data file into the BINs and subBINs that are defined in the SCAVENGER module. (iii) Diagrams, or 'maps', onto which the experimental data are to be displayed. These can be custom-made by the user, downloaded from websites or scanned in from textbooks. They are imported and stored as bitmap (BMP) files. Examples are shown later. The user defines what data are to be deposited at what site on the map by clicking at a chosen position to open a dialogue box, and typing in the numerical code of the BIN or subBIN whose data should be deposited at that location. The precise position can be adjusted by mouse-drag. This operation is repeated to allow different groups of data to be displayed at different locations. After completing this process (which takes only 10–20 min even for a large image), the data overlay can be stored as an associated XML file, which automatically shows up the coordinates of the displayed BIN when the corresponding image file is chosen from the map folder. With time, a library of prepared images can be built up.

The user can choose between two modes of data display (Figure 2). (i) In the default mode, each gene is symbolised by a small box. The change of expression is displayed via a false colour code: genes whose expression does not change are coloured white, and an increasingly large increase or decrease is shown as an increasingly intense blue or red colour, respectively. In the scale used in the first build, colour increases exponentially with the magnitude of the increase or decrease. This was performed in order to allow small changes to be ignored. The scale can be selected in the 'option menu'. As a default, a scale (setting 3) is used in which a 60% change leads to faint coloration and the response saturates at an eightfold change. A mouse-over action can be used to reveal the identity of each individual gene. (ii) Alternatively, the genes in the selected BIN/subBIN can be treated as a population, and their collective response displayed as a frequency histogram. In this mode, all of the genes in the selected BIN/subBIN are sorted according to their change in expression, and the resulting groups are represented as bars sited

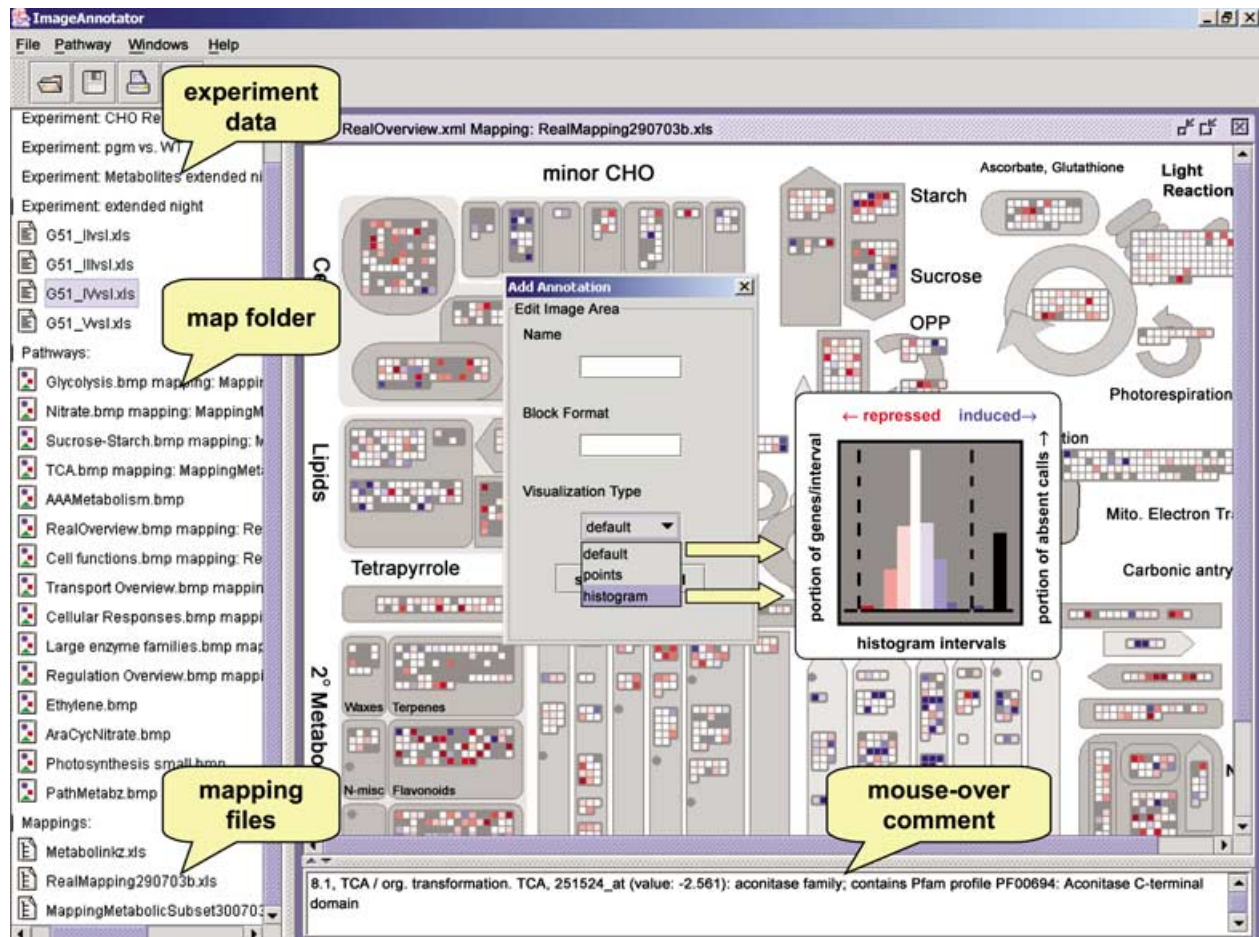


Figure 2. Use of the IMAGEANNOTATOR module.

The menu on the left-hand side contains a set of experimental files, a portfolio of maps/pathways, and a set of mapping files. A map/pathway is selected by mouse click. In this case, the image shows a custom-built schematic representation of metabolism. A further mouse click opens a dialogue box in which the mapping file is selected (not shown). The scale is selected by a mouse click on 'pathway', which opens a dialog box in which the scale is entered manually. The user then chooses which experimental data set is to be viewed and by mouse click displays it automatically onto the pathway map. The site of the original mouse click defines the initial position of the block, which can also be subsequently adjusted by a mouse-drag operation. The 'default presentation' is as a block, in which each gene is represented as a small square. The shape of the block is set via the entry 'blockformat', where it is defined how many squares are in a row in the 'x' and 'y' dimension. In this display mode, a mouse-over action on a particular gene allows the gene annotation, its BIN/subBIN, the experimental value and a brief annotation to be viewed in the window at the bottom of the screen. An alternative data display mode is also available in which the changes in expression of all the genes in a particular category is shown as a frequency histogram. Genes that change by less than a filter value (e.g. <0.33 and >-0.33 on a \log_2 scale) are grouped in the central white bar, genes that increase are grouped in a series of blue bars on the right-hand side (corresponding on this scale to changes between 0.33–0.99, 0.99–1.66, 1.66–2.33, 2.33–3.0 and >3.0 , respectively), and genes that decrease by a similar set of red bars on the left-hand side. Genes called 'not present' by the Affymetrix software are shown as a black bar, on the far right-hand side. The y-axis represents the number of genes in each class, and uses a relative scale in order to allow the plot to be always displayed in a square of a set size. When viewed via the MAPMAN software by a mouse-over action, it will reveal the total number of genes shown in the histogram and the number of genes collected in each class.

along the x-axis. Genes that change by less than a filter value (e.g. <0.33 and >-0.33 on a \log_2 scale) are grouped in the central white bar, genes that increase are grouped in a series of blue bars on the right-hand side (corresponding on the default scale to changes between 0.33–0.99, 0.99–1.66, 1.66–2.33, 2.33–3.0 and >3.0 , respectively), and genes that decrease are represented by a similar set of red bars on the left-hand side. Genes called 'not present' by the Affymetrix software are represented by a black bar on the far right-hand side. The y-axis gives the number of genes in each group. To allow the data to be displayed in a square with a

uniform size, the scale of the y-axis is relative. The number of genes in each class can be accessed by a mouse-over action when the data is viewed using the MAPMAN software. Visual inspection provides a quick impression of how transcript levels for genes in that functional area is responding. If expression of most of the genes is unaltered, there is a large white bar in the middle of the plot. Large blue and red bars show that the expression of many genes is changing. Skewing of the plot towards the blue or red columns, respectively, reveals that the genes in this functional area are being preferentially induced or repressed. A large black

bar on the far right-hand side reveals that many of the genes are not expressed, or are below detection on the Affymetrix chips.

Figure 2 provides a screen shot of the user interface when experimental data is being viewed (see <http://gabi.rzpd.de/projects/MapMan/> for a detailed description). The user first selects a prepared map file and opens it by mouse click. A dialogue box prompts selection of the appropriate mapping file. The experimental file is then selected and uploaded by mouse click. A series of data sets can be called up one after. The first upload requires a few seconds, but once a file is open, it is possible to move back and forwards between data sets in a fraction of a second, allowing a time or treatment sequence to be viewed as a movie. The scale can be changed at any time via the header menu (by accessing the submenu 'options' in the menu 'pathway') if, for example, the user is only interested in genes that show especially large changes in expression, or wants to explore whether a large proportion of the genes in a particular category is showing a small but consistent trend that is below the threshold set by the normal setting. A mouse-over action or an individual signal calls up the precise numerical experimental value and text annotation for that particular gene in a field in the lower part of the screen. The user can click on a free site on the image to open a dialogue field, in which a request can be typed to view data for BINs or subBINs that were not set to appear automatically on the map. All displays can be exported, saved as individual JPG or PNG files and printed out.

Application to other kinds of genomics data sets

The open structure allows other types of data to be imported into the Annotator module. The steps are to develop an appropriate SCAVENGER module and suitable map images. A first build of a METABOLITSCAVENTER assigns >50 metabolites measured in a GC/MS profile to individual positions in central metabolism, and another 500 to general areas of metabolism or broad chemical groupings (see below).

Two different experimental systems to investigate the response to sugar depletion

To illustrate how MAPMAN aids interpretation of complex data sets, we have used it to analyse two treatments that investigate the short-term response of gene expression to carbohydrate depletion.

One treatment compared Columbia-0 (Col-0) wild-type rosettes at the end of the night and after an additional 6 h of darkness. Carbohydrates are low at the end of the night, and fall to very low levels during the next few hours (data not shown). One of two biologically replicated data sets is shown. The second treatment compared rosettes from wild-type Col-0 and the *pgm* mutant at the end of the normal

night. This mutant cannot synthesise starch because it lacks plastid phosphoglucomutase. Sugars accumulate during the day, but are rapidly depleted in the dark, falling to very low levels by the middle of the night (Caspar *et al.*, 1985, 1989; Schulze *et al.*, 1994; and data not shown). By the end of the night, *pgm* has therefore experienced several hours of acute carbohydrate depletion. Both treatments should lead to similar changes of expression for genes that are rapidly regulated in response to low sugar. Differences may occur for genes that are subject to circadian regulation, or that respond differently to a single carbohydrate deficiency and repeated alternation between high and low sugar.

The experiments were carried out with separately grown sets of plants. Tables of the original data are provided in Supplementary Material. Of the 22 000 genes on array, 57–64% were called present. The control samples (wild-type Col-0, harvested at the end of the night) gave similar results in both experiments (Figure 3a, $r = 0.973$), documenting the stability of the growth conditions and quality of the biological material. About 20 genes deviated conspicuously in this and another 6 independent pair-wise comparisons of biological replicates (data not shown). A list of these genes is given in Supplementary Material. Most are plastid-encoded. The scatter may reflect variation in the efficiency of extraction or labelling of plastid RNA. The results for the treatments (extended night, *pgm* mutant) were normalised on the respective control sample (the wild type at the end of the night) and plotted against each other (Figure 3b). Many genes responded in a qualitatively similar manner (see below for further discussion).

Use of overlay plots to compare the response to different treatments on a gene-to-gene basis

Figure 4 displays the response of genes that MAPMAN assigns to metabolism. Figure 4(a) summarises which BINs and subBINs are displayed at which sites on the map, and Figure 4(b) shows the changes in expression after an extension of the night for all the genes in these BINs and subBINs. Because of the large amount of data, it is impossible to identify each gene in the hard copy. To allow the data to be explored with the MAPMAN software, an IMAGE-ANNOTATOR package, including the experimental data files, mapping files and maps is freely available for downloading at <http://gabi.rzpd.de/projects/MapMan/>. The reader is strongly recommended to use this tool to view and explore the data sets, while reading the remainder of this article.

An analogous image for the changes of expression in the *pgm* mutant is included in Supplementary Material. Visual inspection revealed similarities between the changes in an extended night, and in the *pgm* mutant. To aid comparison of the response, thousands of individual genes, we used MAPMAN to generate an overlay plot (Figure 4c, see Experimental procedures), which highlights similarities

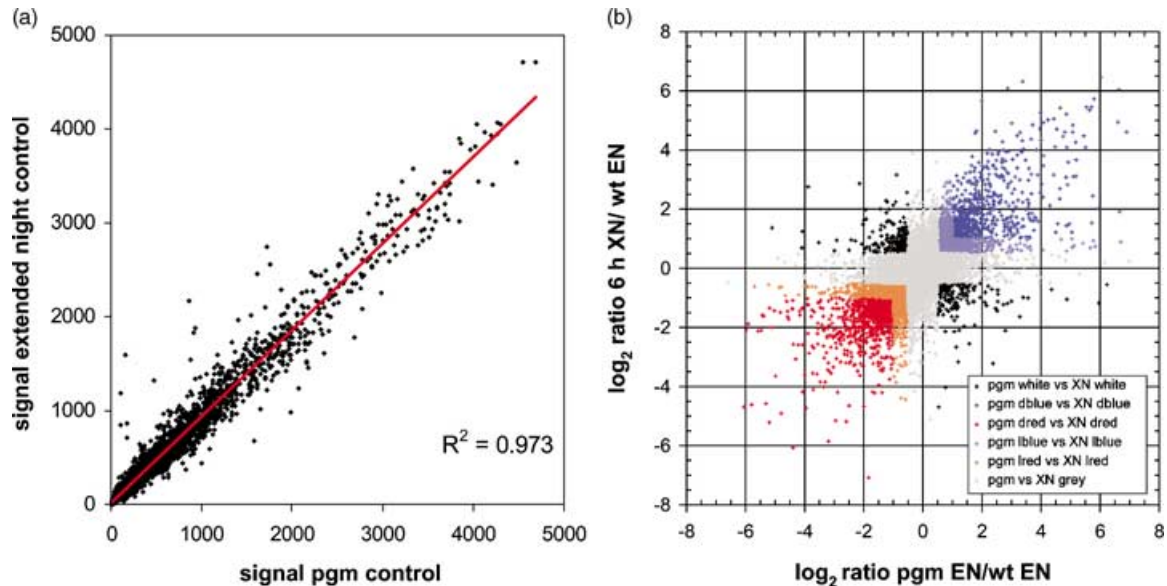


Figure 3. Comparison of two different treatments to generate an *in planta* low-sugar state.

In one treatment, Col-0 wild-type rosettes harvested at the end of the night (control treatment) were compared with Col-0 wild-type rosettes harvested after an additional 6 h dark. In the other treatment, rosettes from wild-type Col-0 harvested at the end of the night (control treatment) were compared with rosettes from starchless *pgm* harvested at the end of the night. The experiments were carried out with separately grown batches of plants. Each array was carried out with a sample containing the pooled rosettes from 15 individual plants.

(a) Comparison of the control samples in the two experiments. The small number of genes that deviate conspicuously also showed deviant behaviour in six other pair-wise comparison of biological replicates, and are almost all plastid-encoded genes (see text).

(b) Comparison of the response in the two treatments. For each gene on the array, the value in the treatment was divided by the value in the control in that experiment and set to a log₂ scale. The resulting ratios were plotted against each other. Each gene is represented by an individual point. The points are colour coded to allow direct comparison with the data display in Figure 4(c) and following panels in Figures 5–12, in which the data in the extended night treatment and the *pgm* treatment are superimposed: deep blue = the signal increases by more than 1.0 (=twofold on a log scale) in both treatments, light blue = the signal increases by ≥ 0.5 in both treatments, but by < 1.0 in at least one treatment, light red = the signal decreases by ≤ -0.5 in both treatments, but by > -1.0 in at least one treatment, deep red = the signal decreases by more than -1 (=a 50% decrease) in both treatments, light grey = the signal changes by less than $0.5/-0.5$ in both treatments, and black = the signal changes by more than 0.5 and in opposite changes in both directions. Note that in Figure 4(c), the genes that show opposing responses are shown in white and not black.

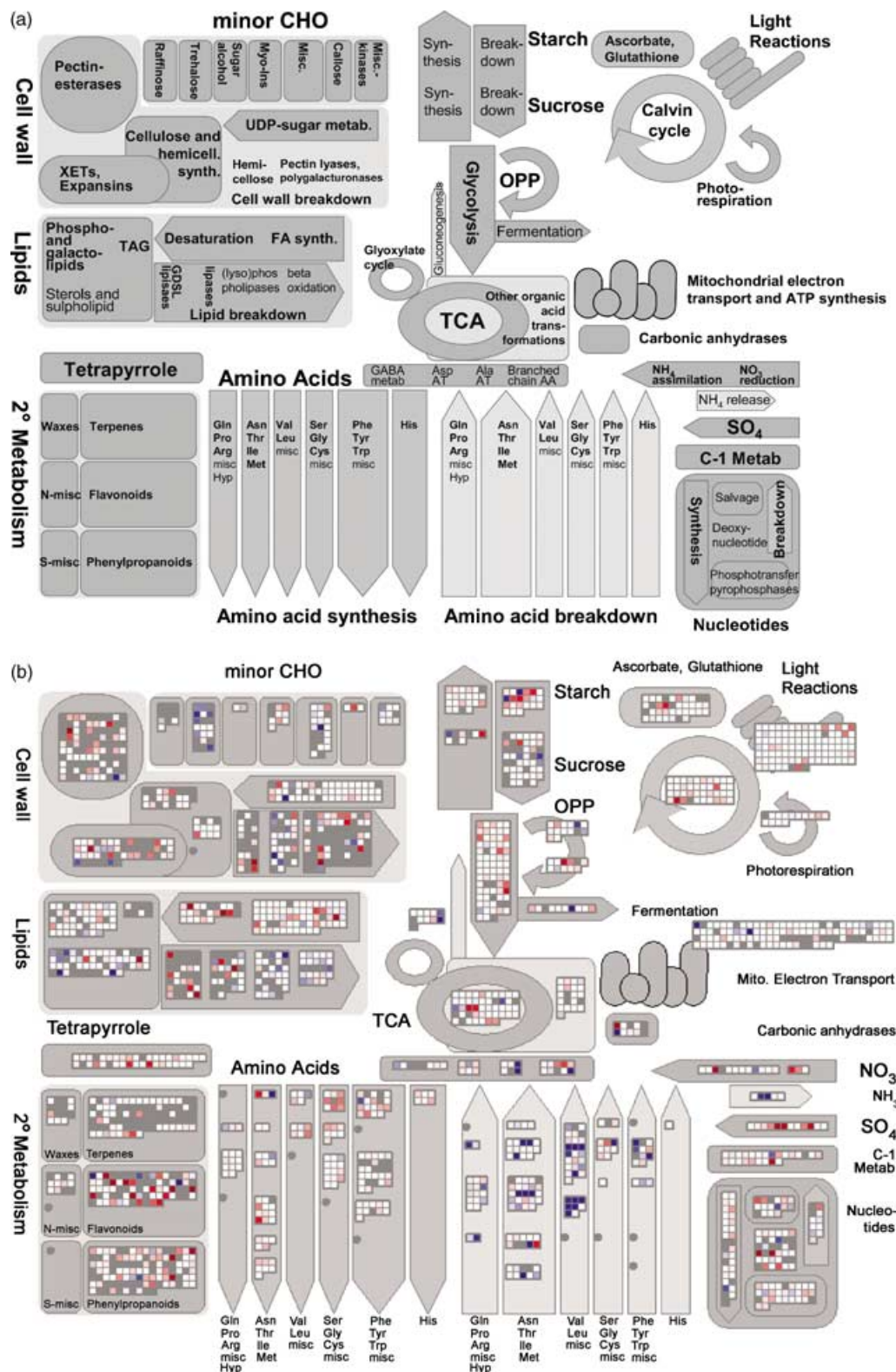
and differences on an area-to-area and a gene-to-gene basis. Genes that increase ≥ 1.0 (on a log₂ scale) in both treatments are shown as blue, genes that increase by ≥ 0.5 in both treatments but by < 1.0 in at least one treatment are pale blue, genes that decrease by ≤ -0.5 in both treatments but by > -1.0 in at least one treatment are pale red and genes that decrease ≤ -1.0 in both treatments are red. Genes that show an opposite response are shown in white. Genes called not present or that change by < 0.5 are shaded grey, and cannot be distinguished from the background (note: the same colour scale is used in Figures 3(b) and 4(c), except that genes with an opposite response are black in Figure 3 and white in Figure 4(c). The lower cut-off was empirically selected to maximise recruitment of genes that have a shared response, while minimising the recruitment of false positives (see Experimental procedures and Supplementary Material). The overlay plot (Figure 4c) reveals that the majority of genes have a qualitatively similar response in both treatments.

The following sections explore these two sets of Affymetrix arrays in detail, to provide new insights into the

wide-ranging impact of low sugar on the transcriptional regulation of metabolism and cellular responses. The first sections will be primarily of value to readers with an interest in metabolism or sugar sensing. Other readers are encouraged to move skip through to the later sections, which cover sets of genes in which they may have more interest. As in all interpretations of expression profile data, the conclusions are tentative. However, several features of MAPMAN and some aspects of our experimental design strengthen the conclusions. First, not all of the biologically relevant events will have been identified. The grouping of many genes with a related function, however, allows more sensitive detection of trends than inspection of a list of unsorted individual genes. Second, the reliability of the conclusions obviously depends on the quality of the available gene annotations, and the correctness of the assignments. This affects conclusions about individual genes more strongly than conclusions drawn from the response of a group of genes, where incorrect assignment will lead to 'noise' but will not necessarily lead to incorrect interpretations. Third, some of the changes may have no

biological relevance; for example, changes of transcripts do not necessarily affect protein levels or activities. We will provide some data from metabolite profiling to support our general conclusions. For brevity, we refer to

increases in transcript levels as induction, and to decreases as repression, but emphasise that the data do not distinguish between changes in transcription rate and transcript turnover.



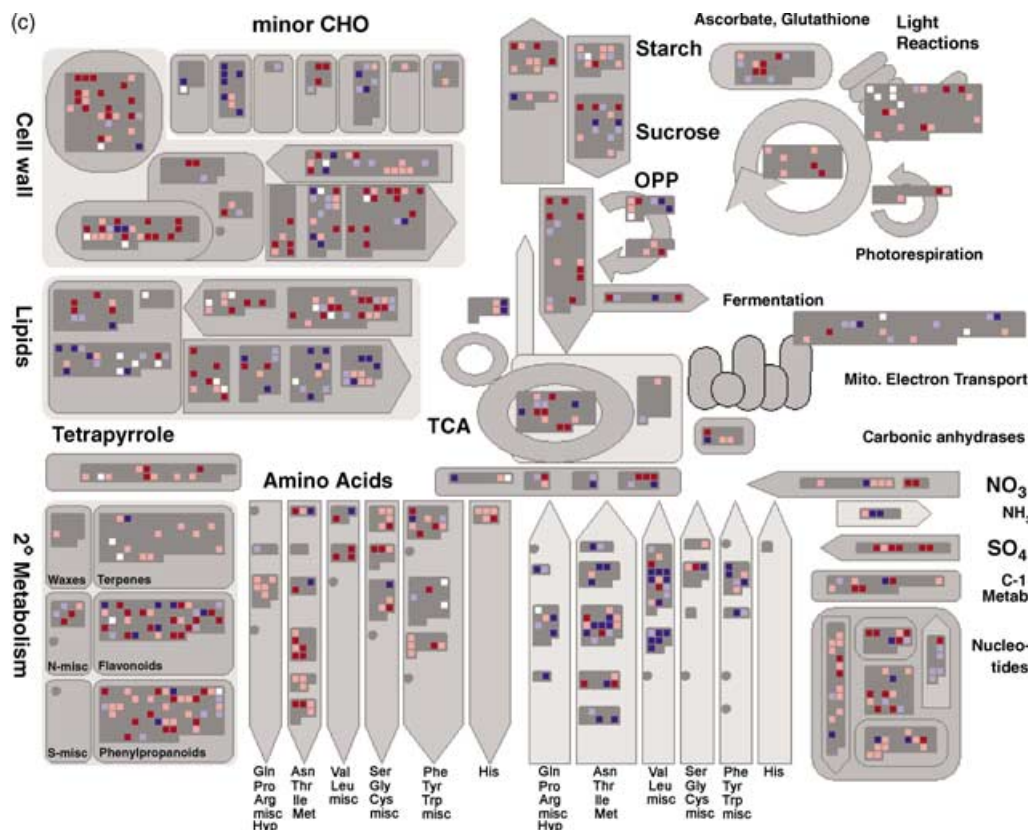


Figure 4. continued

Repression of genes involved in photosynthesis, and sucrose and starch synthesis

Many genes assigned to chlorophyll synthesis (BIN 18), light harvesting, photosynthetic electron transport and ATP synthesis (BIN 1.1, termed 'thylakoid reactions' for brevity), the Calvin cycle (BIN 1.3) and photorespiration (BIN 1.2) were repressed in the *pgm* mutant (see Supplementary Material). A similar result was found after an extended night, except that some genes assigned to the 'thylakoid reactions' showed an upward trend (Figure 4b). This was confirmed in a biological replicate (not shown). Five genes showed relatively large divergent changes, resulting in a

conspicuous cluster of white signals in BIN 1.1 (Figure 4c). For others, the increase was below the threshold selected for Figure 4(b,c), but can be seen by exploring the files at the website with the mouse-over option, or by altering the colour scale.

Figure 5 summarises the response of these groups of genes as a histogram frequency plot. A fuller description of this display mode is given in the legend to Figure 2 and the accompanying text. The sensitivity of the colour scale has been slightly increased, compared to Figure 4. In the *pgm* mutant, there is a coordinated repression of many genes in all four functional areas. This is visualised by the strong skew towards the red-coloured bars on the left side of all

Figure 4. Overview display of genes assigned to metabolism.

(a) Explanation of the functional categories deposited at each location on the scheme.

(b) Change of transcript levels in wild-type Col-0 after a 6-h extension of the night, compared to the end of the night. The sister set of data for transcript levels in the starchless *pgm* mutant at the end of the night compared to wild-type Col-0 at the end of the night is given in Supplementary Material. Both sets of material were harvested at the end of the night. In panels (b,c), red and blue represent a decrease and increase of expression, respectively, relative to the control value. A colour scale was used, which in a slight coloration represents 60% change, and the response saturates at an eightfold change.

(c) Direct comparison of the response in the two treatments. Genes with a value of ≥ 1.0 in both treatments are shown as blue, genes with a value ≤ -1.0 in both treatments are red, genes with a value ≥ 0.5 in both treatments but at least one is < 1.0 are pale blue, genes with a value ≤ -0.5 in both treatments but at least one > -1.0 are pale red, and genes that show an opposite response are shown in white. Genes whose value lay between 0.5 and -0.5 in both treatments and genes that were called 'not present' are shaded grey and cannot be distinguished from the background. The same scale is used in panel (b), in Figure 9 and in the overlay plots in Supplementary Material. In all cases, the complete rosette was harvested from 15 plants, and the plants were pooled before RNA extraction. This figure is best viewed, and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

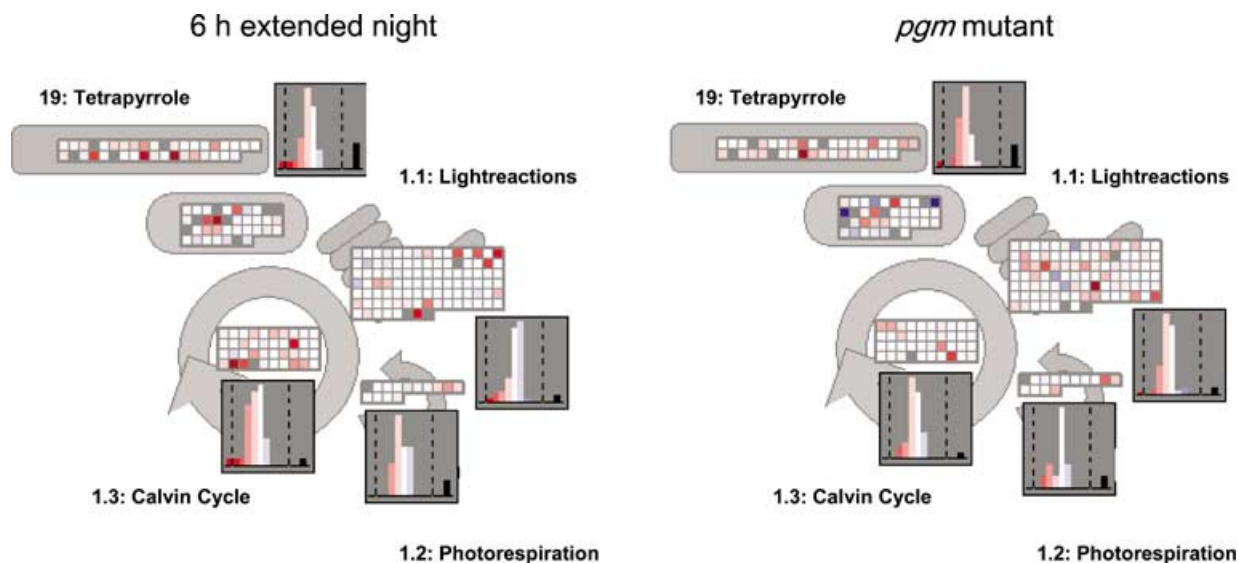


Figure 5. Histogram display to compare the response of genes involved in photosynthesis to an extended night and in the starchless *pgm* mutant.

This figure shows excerpts from screen shots, in which the response of genes in BINs 1.1, 1.2, 1.3 and 18 (light reactions, photorespiration, Calvin cycle and tetrapyrrole synthesis) are displayed as a frequency histogram. The data are identical to that in Figure 4 and the accompanying Supplementary Material, except that a slightly more sensitive scale is used. Genes that change (all values given on a \log_2 scale) by less 0.27 are grouped in the central white bar, genes that increase a series are grouped into a series of blue bars on the right-hand side (corresponding to genes that changes between 0.27–0.83, 0.83–1.38, 1.38–1.94, 1.94–2.5 and >2.5, respectively, genes that decrease are by a similar set of red bars on the left-hand side, and genes called 'not present' by the Affymetrix software are shown as a black bar on the right-hand side. This figure is best viewed and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures). To view these data using the MAPMAN software, first open the map 'overview' and display the data sets in it, and then use the dialog box to display the subBINs (see figure for the numbers) in the histogram mode (to obtain the identical scaling, open the menu 'Pathway' and then the submenu 'Options', and then on the scale 'adjust scaling', select the value 2.5. A mouse-over action will give the total number of genes in the category, and for each group displayed as a bar in the histogram, gives the limits and the number of genes in the group. A biological replicate for the extended night treatment gave close agreement (see Figure 3a, and data not shown).

four histogram plots. In the extended night treatment, genes assigned to chlorophyll synthesis, the Calvin cycle and photorespiration show a downward trend, but many of the genes assigned to the 'thylakoid reactions' show a slight upward trend, which is visualised as a large light-blue column, just to the right side of the central white column. A plausible explanation for the different response in the two treatments is that many genes involved in light harvesting are subject to circadian control (Schaffer *et al.*, 2001). This may antagonise the effect of low sugar in the extended night treatment.

Fixed carbon is converted to starch in the plastid (where it acts as a transient store and is degraded the following night), or exported as triose phosphate to the cytosol, converted to sucrose and exported. There was a clear trend to repression of genes required for starch synthesis in an extended night (BIN 2.2.1, Figure 4b) and the *pgm* mutant, allowing a group of genes to be identified that are repressed in both conditions (see the overlay plot, Figure 4c). A more differentiated response was found for sucrose synthesis. Triose phosphates are converted to hexose phosphates via the cytosolic fructose-1,6-bisphosphatase (cFBP). cFBP was strongly induced in both treatments (Figure 4b,c, see below for further discussion of this unexpected result). The final committed reactions of sucrose synthesis are catalysed by sucrose phosphate synthase and sucrose phosphatase. In

both treatments, two genes that encode major members of the families for these enzymes were repressed. Details of the changes in expression of genes involved in sucrose and starch synthesis can be viewed in Supplementary Material.

Inhibition of nitrate and sulphate assimilation and amino acid biosynthesis, and induction of amino acid breakdown

Some of the fixed carbon, ATP and reducing equivalents formed in photosynthesis are used to convert nitrate and ammonium into amino acids. Low sugar repressed many genes assigned to these processes (Figure 4, BIN 12.1), in particular nitrate (*NIA1*, *NIA2*) and nitrite (*NII*) reductase (Figure 4, BIN 12.1; see also Klein *et al.*, 2000). Ammonium assimilation was less affected; indeed, one putative GS was weakly induced (BIN 12.2).

Enzymes in central amino acid metabolism (BIN 13.1) like aspartate aminotransferases induced and others repressed, indicating that some are involved in synthesis and others in the degradation of amino acids. A similar picture emerged for the branched amino acid aminotransferases, which are involved in the metabolism of different groups of aliphatic amino acids.

Many of the genes assigned to amino acid biosynthesis were repressed (Figure 4b,c), indicating that there is a

broad transcriptional inhibition of amino acid biosynthesis. However, a few of these genes were induced. There are two different explanations for this kind of minority response. One is that they are indeed induced by low sugar or some other factor in our treatments, and have a special function in these conditions. For example, one of the genes that is induced in both treatments (asparagine synthetase 1 (*ASN1*)) is well known (Lam *et al.*, 1994) to be strongly induced by low sugar (see below for further discussion of its role). An alternative explanation is that the genes are wrongly annotated or assigned. For example, one of three genes annotated as the small subunit of acetolactate and some putative aromatic aminotransferases, which were tentatively assigned to amino acid synthesis, are induced by both treatments. This illustrates how MAPMAN can be used to pinpoint genes that have an unexpected expression pattern, putting them on a short list for a focused re-assessment of their annotation and assignment.

Many genes assigned to amino acid breakdown were induced (Figure 4b,c), including proline oxidases, proline dehydrogenases, members of the glyoxalases I and II families, all three members of the alpha ketoacid dehydrogenase family, 3-methylcrotonyl-CoA dehydrogenase, isovaleryl-CoA dehydrogenase, L-allo threonine aldolase, 4-hydroxyphenylpyruvate dioxygenase and homogentisate 1,2-dioxygenase. The only genes assigned to amino acid degradation that were repressed were an S-adenosyl-L-homocysteinase, some individual members of the enoyl-CoA hydratase and 3-hydroxyisobutyryl hydrolase families, and several genes assigned to glycine degradation (which remained unaltered or increased, BIN 13.5.2.2). The latter can be understood because the same pathway is involved in photorespiration.

NADH-glutamate dehydrogenase (*GDH*) catalyses the reversible conversion of glutamate to oxoglutarate and ammonium. It has long been debated if it is involved in ammonium assimilation or release (Mifflin and Habash, 2002). An early step in the catabolism of many amino acids is a transamination in which the amino group is transferred onto oxoglutarate, leading to formation of glutamate. NADH-GDH is encoded by three genes. Two of them were strongly induced in both treatments (Figure 4b,c; BIN 12.3). This is consistent with their role being to recycle glutamate back to 2-oxoglutarate during amino acid catabolism. In agreement, *gdh1* mutants are compromised in their ability to grow on nitrogenous sources alone (Melo-Oliveira *et al.*, 1996). The third member of the GDH family and both members of the NADP-GDH family were not induced (expression of the former even decreased in *pgm*), indicating they have a different function. The array results prompt the hypothesis that the solitary induced GS (see above) and *ASN1* operate to scavenge the ammonium that is released by GDH, and store it as the N-rich amino acid Asn.

Inhibition of sulphate assimilation and the synthesis of S-containing amino acids

Both treatments repressed many of the genes involved in S assimilation (BIN 14; Figure 4b,c). Cysteine synthesis involves synthesis of O-acetylserine by serine acyl transferase (SAT), followed by incorporation of sulphide by O-acetyl-(thiol)lyase (OAS) (BIN 13.5.3.1). One member of the SAT family was repressed. Most of the OAS family were unaltered or decreased but, curiously, one was induced. Many genes assigned to methionine and S-adenosylmethionine synthesis were also repressed (BIN 13.3.4.1).

Reprogramming of respiratory metabolism to allow flexible use of a range of substrates

Low sugar led to decreased expression of many genes assigned to carbohydrate breakdown, glycolysis, the tricarboxylic acid (TCA) cycle and mitochondrial electron transport and ATP synthesis (Figure 4b,c). The Supplementary Material and Figures 5 and 6 provide a detailed breakdown of these pathways, resolving them at the level of the individual reactions, and showing the gene or (in almost all cases) the gene family annotated to each step. There was a general trend to decreased expression. The extent varied in a member-specific manner, and in some cases was at or below the threshold.

Unexpectedly, a small subset of genes was induced. This included some of the genes annotated as neutral invertases (see Supplementary Material; interestingly, an invertase inhibitor homolog was repressed), a member of the large family annotated to pyrophosphate-fructose-6-phosphate phosphotransferase (PFP, this enzyme catalyses a reversible P_{Pi}-dependent interconversion of Fru6P and Fru1,6bisP, Stitt, 1990), and two members of the large family annotated to pyruvate kinase (Figure 6). In the TCA cycle (Figure 7) individual genes that encode enzymes in the first part of the cycle (pyruvate dehydrogenase, citrate synthase, aconitase) were induced. Several genes that encode NADH dehydrogenases were induced in the mitochondrial electron transport chain. One of two genes assigned to uncoupling proteins (UCPs) was also induced (see Figure 4). A member of the UCP family is also induced in starvation in humans (Dulloo *et al.*, 2001), and has been suggested to be involved in transport of fatty acids during lipid catabolism (see below for further discussion).

Of the genes annotated to gluconeogenesis, pyruvate Pi dikinase, one of five ATP-citrate lyases (Figure 7) and (see above, Figure 6) cFBPase were strongly induced. Many other genes involved in gluconeogenesis (phosphoenolpyruvate carboxykinase, various malic enzymes) were unaffected. Genes involved in the glyoxylate cycle (isocitrate lyase, malate synthase, peroxisomal malate dehydrogenase) remained below detection or did not change

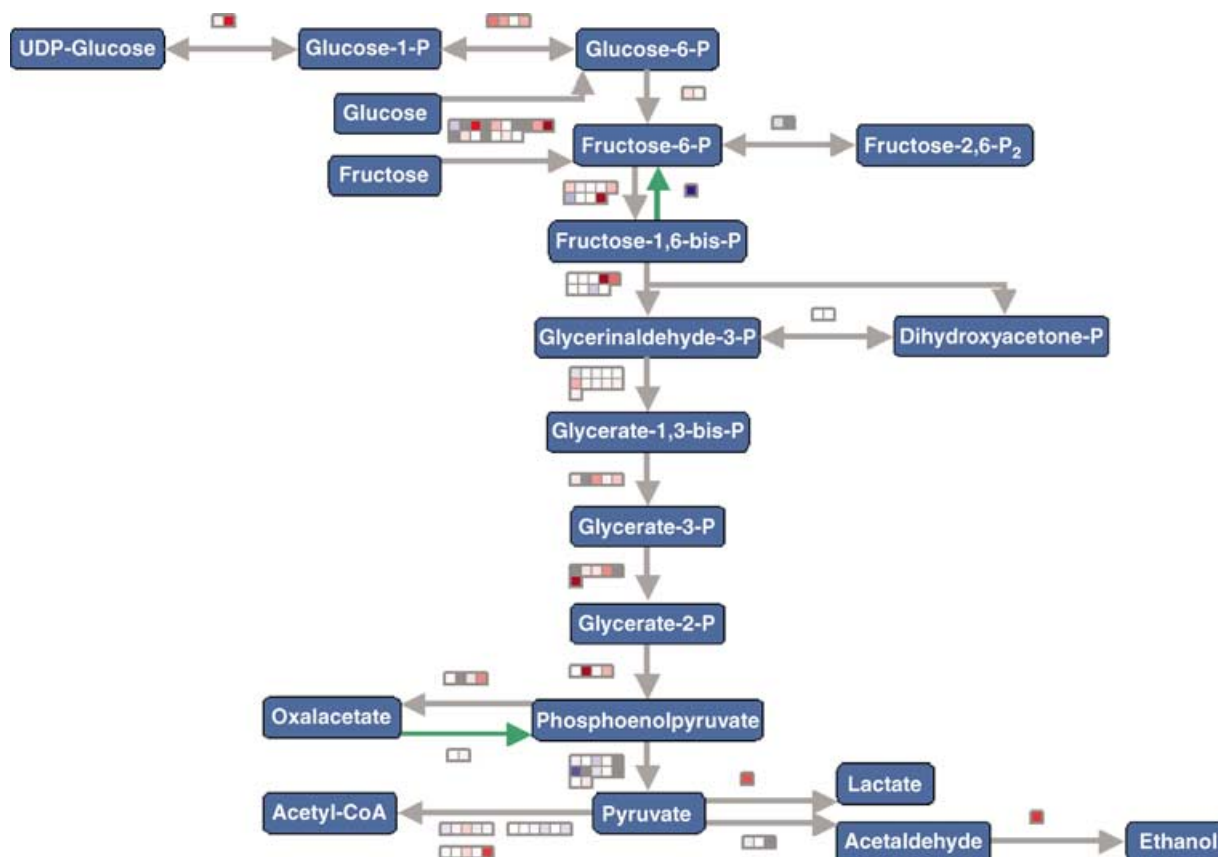


Figure 6. Pathway level display of genes involved in glycolysis.

Change of transcript levels in wild-type Col-0 after a 6-h extension of the night, compared to the end of the night. For scales, see Figure 5. Overlay plots comparing the response in an extended night and in the *pgm* mutant are available in Supplementary Material. This figure is best viewed and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

(Figure 7). This differs from lipid-storing seeds, when the glyoxylate cycle enzymes (isocitrate lyase, malate synthase) and PEP carboxykinase are coordinately induced during germination (Rylott *et al.*, 2001).

Taken together, these results reveal a trend to decreased expression of many genes involved in carbohydrate breakdown, glycolysis, the TCA cycle, mitochondrial electron transport and ATP synthesis. This is consistent with a general slowing down of respiration energy metabolism. At the same time, several genes are induced encoding enzymes that move carbon back or forward between hexose phosphates and 3-carbon intermediates in glycolysis (PFK, cytosolic FBPase), between PEP and pyruvate (pyruvate Pi dikinase, pyruvate kinase), and between citrate and oxaloacetate and acetyl-CoA (ATP-citrate lyase). This response indicates that central metabolism is being re-organised to allow flexible use of carbon skeletons from different sources.

We next extended the analysis of the data to learn whether there was a widespread switch from anabolism to catabolism. The following discussion can best be followed using the downloadable MAPMAN package to access the individual genes. Figure 8 shows excerpts from

screen shots of displays, in which the overall response of groups of genes assigned to the synthesis and degradation of nucleotides, lipids and cell wall components are summarized as frequency histograms (see Figures 2 and 5 for a full description of this display mode).

Nucleotide metabolism

Low sugar repressed many genes assigned to purine and pyrimidine synthesis (BIN 23.1), and slightly stimulated expression of many genes assigned to nucleotide breakdown (BIN 23.2; see Figures 4b,c and 8). The only exceptions were two genes (annotated as 'CTP-synthase-like and GMP-synthase-like'), which were induced in the biosynthesis subset, and one gene (a 5-bisphosphate nucleotidase), which was repressed in the degradation subset.

Switch from lipid synthesis to lipid breakdown

Low sugar repressed a subset of the genes assigned to fatty acid synthesis (Figures 4b,c and 8). Although expression of genes encoding acetyl-CoA carboxylase and enzymes in the

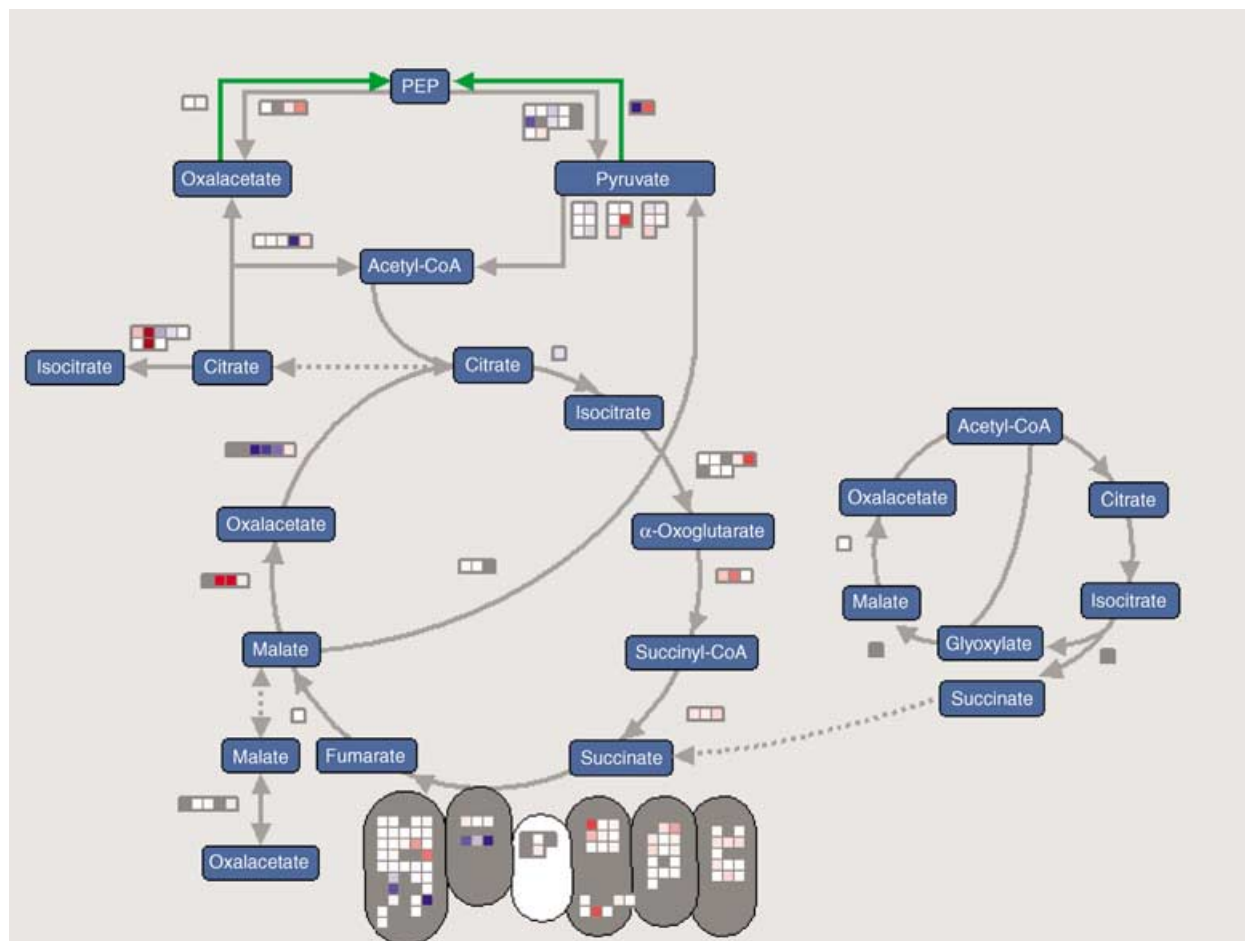


Figure 7. Pathway level display of genes involved in the TCA cycle, glyoxylate cycle, gluconeogenesis and other organic acid transformations. Change of transcript levels in wild-type Col-0 after a 6-h extension of the night, compared to the end of the night. For scales, see Figure 5. Overlay plots comparing the response in an extended night and in the *pgm* mutant are available in Supplementary Material. This figure is best viewed, and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

plastid pathway of fatty acid synthesis was not strongly affected, many genes assigned to fatty acid transfer and acyl-CoA elongation in the cytosol were repressed (BIN 11.1). Many genes assigned to fatty acid desaturation (BIN 11.2) and phospholipid and galactolipid synthesis (BIN 11.3) were also repressed. Of the three genes in these subBINs that were induced are annotated as choline kinases and might equally well be involved in recycling choline released during phospholipid breakdown (see next paragraph).

Genes annotated to lipid breakdown were subdivided into four subgroups (Figure 4, see also the histogram frequency plots in Figure 8). One contained genes annotated as containing a GDSL-motif in the protein, with Gly-, Asp-, Ser-(Leu), as active site residues (BIN 11.9.1). These comprise a distinct class of lipases/esterases found in prokaryotes, fungi and plants (Beisson *et al.*, 1997). Many genes in this group were repressed, and none were induced. This might indicate that they are mainly involved in biosynthesis. A second group comprised all other genes annotated as 'lipase' (BIN 11.9.2). Some were repressed and others induced.

The third group contained genes annotated as phospholipases and lysophospholipases (BIN 11.9.3). Many were induced in low sugars. The fourth set contained genes whose annotation indicates a role in fatty acid β -oxidation (BIN 11.9.4). Many of these were induced, including several acyl-CoA oxidases, which catalyse the first step in the peroxisomal pathway. There was a marked increase of acyl-CoA dehydrogenase, which catalyses the flavoprotein-dependent reduction of NADH as the first step in the mitochondrial pathway. One or more members of the small families for enoyl-CoA hydratase, multifunctional protein (MFP)2 and 3-ketoacyl-CoA thiolase, which catalyse the subsequent steps in the β -oxidation pathway, were induced. These results point to activation of lipid catabolism.

Cell wall breakdown

Several members of the UDP glucose dehydrogenase family, which catalyse the first step in synthesis of precursors for pectin synthesis family, were repressed

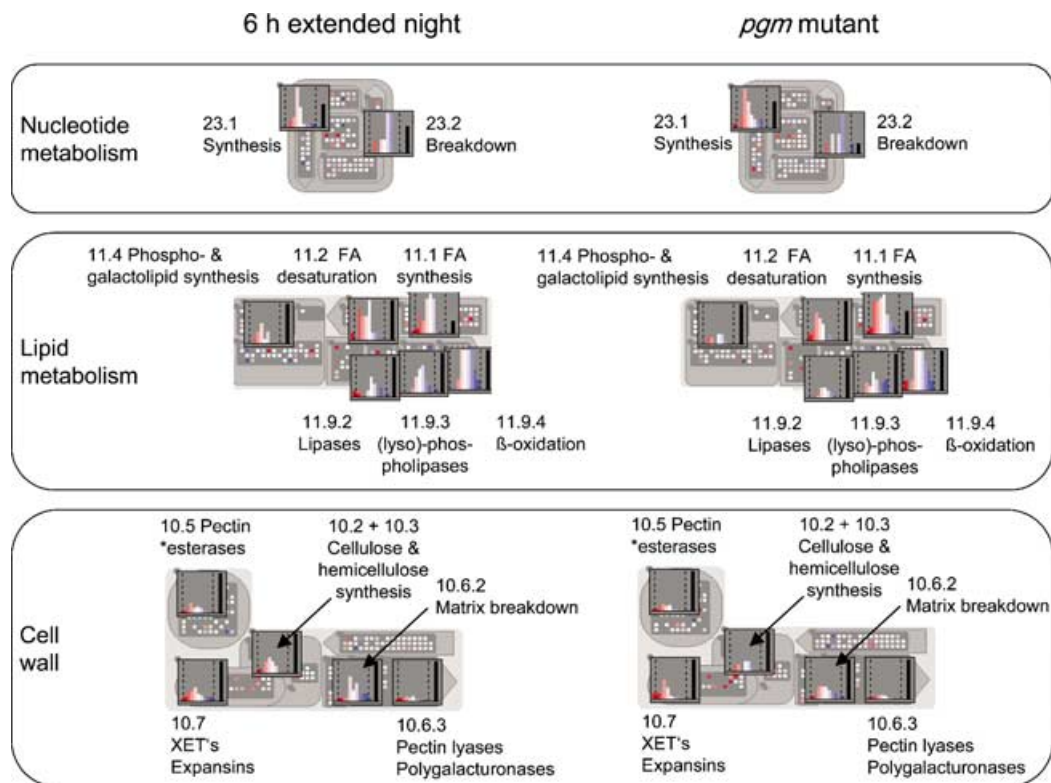


Figure 8. Frequency distribution display of genes involved in nucleotide, lipid and cell wall metabolism, and cell wall modification.

Genes that change by less than a filter value (<0.27 and >-0.27 on a \log_2 scale) are grouped in the central white bar, genes that increase a series of blue bars on the right-hand side (corresponding on this scale to changes between 0.27–0.83, 0.83–1.38, 1.38–1.94, 1.94–2.5 and >2.5 , respectively), genes that decrease by a similar set of red bars on the left-hand side, and genes called 'not present' by the Affymetrix software are shown as a black bar on the right-hand side. This figure is a combination of excerpts. This figure is best viewed and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures and legend to Figure 5, or the general instructions on the website). After opening the data sets and displaying the data for a particular subBIN as a frequency histogram, a mouse-over action will give the total number of genes in the category and, for each group displayed as a bar in the histogram, gives the limits and the number of genes in the group.

(Figure 4b,c, the first signals reading left to right in subBIN 10.1). Expression of other genes involved in UDP and GDP sugar metabolism did not show a clear trend, with some decreasing and others rising. Histogram plots (Figure 8) of genes in subBINs 10.2, 10.3 and 10.6.2 reveal a trend to preferential induction of genes involved in the synthesis and repression of genes assigned to the breakdown of the cell wall. Nearly all of the genes annotated as cellulose synthase and cellulose synthase-like decreased (subBINs 10.2 and 10.3). Several genes annotated as xylo(endo)glucanases, arabinosidases and xylosidases were induced (subBIN 10.6.2), as were genes annotated as ribulokinases, galactokinases and xylokinases (see subBIN 3.7), indicating they might be involved in phosphorylating sugars released during cell wall breakdown. There was no clear trend in expression of genes annotated as cellulases or endo alpha 1,4 glucanases (subBIN 10.6.1).

Repression of many cell wall-modifying enzymes

There were also clear trends for five gene families, whose members modify the properties of the cell wall

matrix (Figure 4b–d, see also the frequency histograms in Figure 8). Many pectinesterases (subBIN 10.8) were repressed in both treatments. Many pectin lyases and polygalacturonases (subBIN 10.6.3) were repressed by low sugars, while two genes annotated as polygalacturonase inhibiting proteins were induced (these are the two 'induced' genes in this subBIN in the overlay plot in Figure 4c). These results indicate that pectinesterases, polygalacturonases and pectin lyases are not involved in breaking down cell walls during the early starvation response but may instead be involved in growth processes, which are inhibited when sugars are exhausted. Xyloglucan endotransglycosylases (XET) and expansins (subBIN 10.7) facilitate rearrangement of the xyloglucan matrix and loosening of the bonds between the matrix and the cellulose fibrils during cell elongation. Many were repressed, indicating that cell wall extensibility is decreased when sugars fall. Interestingly, three specific members of the XET family were induced in both treatments, and two expansins were strongly induced in the extended night treatment (although not in *pgm*). These genes might be involved in specific responses that are triggered by low sugar.

Trehalose metabolism

Striking changes were found for genes assigned to trehalose metabolism (BIN 3.2). Both treatments strongly induced several of the 11 members of the trehalose phosphate synthase (TPS) family, including the ones whose expression is highest in leaves, slightly repressed the most strongly expressed of the trehalose-6-phosphate (Tre6P) phosphatase family and induced trehalase. In yeast, Tre6P regulates the entry of carbon into metabolism by inhibiting hexokinase (Bonini *et al.*, 2000). Recent evidence reveals an important role in plants too, although the mode of action and role are still unclear. Knockout mutants in *TPS1* are embryo lethal (Eastmond and Graham, 2003), whereas ectopic over-expression of heterologous TPS and TPP revealed that increased Tre6P promotes growth on exogenous sugar (Schluepmann *et al.*, 2003). Our results prompt the hypothesis that falling sugars lead to an increase of Tre6P, which might play a role in starvation responses.

Secondary metabolism

Both treatments slightly repressed the major member of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*DXR*) family and induced acetoacetyl-CoA-thiolase. The extended night treatment also weakly induced 3-hydroxymethylglutaryl-CoA reductase (Figure 4b,c, these genes are located at the upper left-hand corner of BIN 16.1). This indicates there may be a shift from the plastid to the non-plastidic pathway for terpene biosynthesis. There was a trend in both treatments to repression of genes further downstream in terpene metabolism. This was particularly clear for carotene metabolism. Interestingly, one of the two genes encoding neoxanthin cleavage enzyme (involved in abscisic acid (ABA) synthesis, see below) was strongly induced in the extended night treatment.

Both treatments repressed two to three of the four genes that encode phenylalanine ammonia-lyase (BIN 16.2). Genes annotated to steps further downstream in phenylpropanoid and flavonoid (BIN 16.3) metabolism showed highly individual responses, with many rising and others falling. The response was strongly conserved between the extended night and *pgm* treatments. These results point to extensive re-programming of phenylpropanoid and flavonoid metabolism.

Plants possess large gene families for cytochrome P450s, UDP glucosyl transferases, alcohol dehydrogenases, glucosidases, *O*-methyl transferases, nitrilases, cyanohydrin lyases, berberine bridge enzymes/reticuline oxidases, troponine reductase-like proteins, acetyltransferases, beta 1,3-glucan hydrolases and peroxidases. Although the precise role of the individual members is rarely known, they presumably catalyse reactions in various biosynthetic and secondary pathways. Figure 9 shows the overlay plot of the response

of individual members of these families to both treatments (see Supplementary Material for the original plots). A relatively high percentage was below the detection limit. Most of those that were expressed showed reproducible changes of transcript levels in both an extended night and *pgm*, with some rising and others falling. This confirms the conclusion that low sugar rapidly leads to extensive transcriptional re-programming of biosynthetic and secondary metabolism.

Integrated changes in the expression of genes involved in transport and metabolism

Genes annotated as transporters showed changes in expression that are coordinated with the changes in metabolism (see Supplementary Material). Low sugar repressed several nitrate, sulphate and phosphate transporters (see Supplementary Material). This parallels the repression of enzymes involved on nitrate and sulphate assimilation (see above, Figure 4). Sucrose, hexose, amino acid and peptide transporters showed a differentiated response, with some falling and others rising. The latter might transport metabolites formed during protein, amino acid and cell wall catabolism. Several plastid envelope membrane transporters were repressed, including the glucose-6-phosphate: phosphate transporter, two PEP:phosphate transporters, the ATP/ADP transporter, an oxoglutarate/malate exchanger and a putative glycerol-3-phosphate permease (see Supplementary Material), which is consistent with down-regulation of biosynthetic pathways in the plastid. Many transporters assigned to mitochondria were also repressed, including several genes encoding oxoglutarate/malate and other putative dicarboxylate carriers, which is consistent with downregulation of the TCA cycle.

Integration of metabolism with major cellular functions

Figure 10 provides an overview of the expression of about 18 000 genes, sorted into 25 BINs or subBINs that reflect major cellular or functional processes. The results are summarised as frequency histogram. A fuller description of this display mode is given in the legend to Figures 2 and 5, and the accompanying text (see above). A very high proportion of genes were called present for protein synthesis (89%), amino acid activation (86%), vesicle transport (85%), central metabolism (84%, not shown) and protein targeting (82%), a somewhat lower proportion in the groups 'regulation of transcription', 'regulation' (62%), DNA synthesis (46%), development (55%), abiotic (55%) and biotic (55%) stresses, and even fewer of the genes in BIN 35.1 (not assigned: no ontology, 31%). Interestingly, 64% of the genes in BIN 35.2 (not assigned: unknown or hypothetical protein) were detected on the arrays. Marked changes of the expression pattern were found in all of these categories.

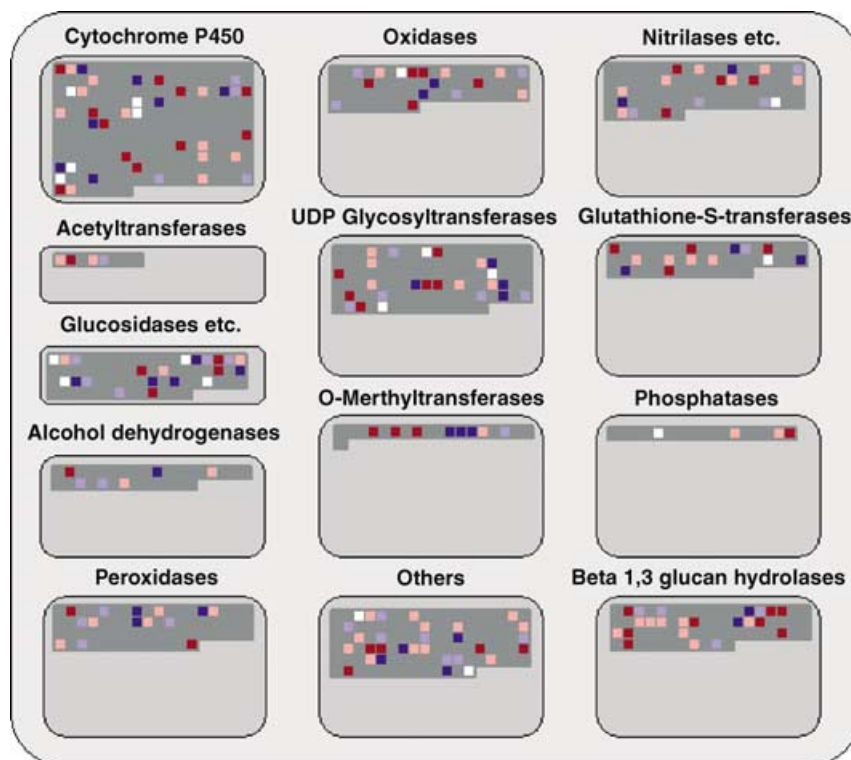


Figure 9. Display of genes assigned to families for large enzyme classes.

The figure shows an overlay of the response in an extended night and in the *pgm* mutant (for procedure and scale, see legend to Figure 4c). Plots of the original data for the extended night treatment are available in Supplementary Material. This figure is best viewed, and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

RNA and protein synthesis

Both treatments led to a preferential repression of genes involved in amino acid activation and protein synthesis, and a slight preferential induction of genes assigned to protein degradation (Figure 10). A similar spectrum of genes was repressed and induced by both treatments (not shown). There was a slight preferential repression of genes assigned to RNA synthesis and (in the case of *pgm*) RNA processing. There was also a slight preferential repression of genes assigned to cell division, regulation of cell cycle (only in the case of *pgm*), DNA synthesis and DNA repair. The results reveal that a general transcriptional inhibition of cellular activity, especially protein synthesis, is initiated within a few hours of sugars being exhausted.

Strongly preferential induction of genes involved in ethylene and ABA signalling

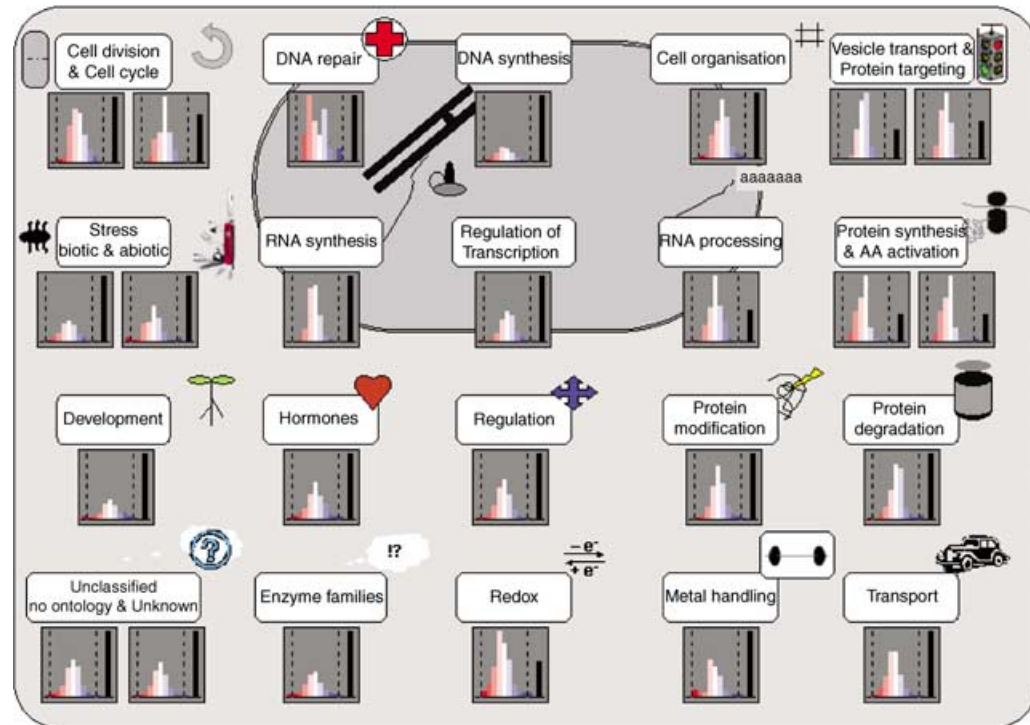
A remarkably large proportion of the genes assigned to hormone synthesis/sensing genes showed changes in expression during an extended night. Many of these were also seen in the *pgm* mutant (Figure 11, see also Figure 12 for histogram frequency plots). In both treatments, there was a trend to decreased expression of tRNA isopentenyl transferases, which are involved in cytokinin synthesis (BIN 17.4). There was a trend to increased expression of genes involved in ABA synthesis and sensing. The key regulatory step in ABA synthesis is catalysed by neoxanthin cleavage enzyme (Iuchi *et al.*, 2001; Schwartz *et al.*, 2001). One of the two genes that encode this enzyme was strongly induced

Figure 10. Frequency distribution display of general cellular responses.

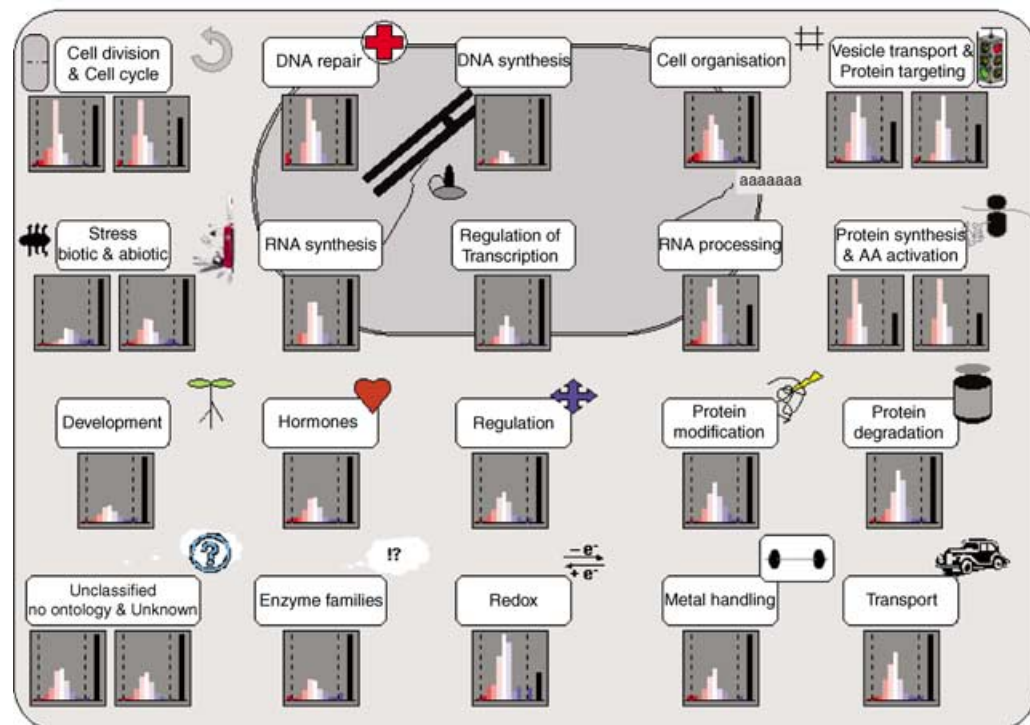
(a) Change of transcript levels in wild-type Col-0 after a 6-h extension of the night compared to the end of the night.

(b) Change of transcript levels in the starchless *pgm* mutant, compared to wild-type Col-0. The groups of genes are identified in the figure, and the respective BIN or subBIN and the numbers of genes currently assigned are given in Table 1. For each functional group, the response is shown as a frequency histogram. Genes that change by less than 20.5% of the signal scale are grouped in the central white bar, and genes that increase and decrease more are shown as a series of blue and red bars (the sequential bars moving out from the central white bar along the x-axis correspond to genes whose transcript level changes between 0.27–0.83, 0.83–1.38, 1.38–1.94, 1.94–2.5 and >2.5 on a log₂ scale) on the left- and right-hand sides, respectively. Plots of the original data are available in Supplementary Material. Exploration using the MAPMAN tool (see legend to Figure 2 for details) allows the numbers of genes in each BIN or subBIN, and the limits and number of genes in each bar of the frequency plot to be called up by mouse-over action. This figure is best viewed, and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

(a)



(b)



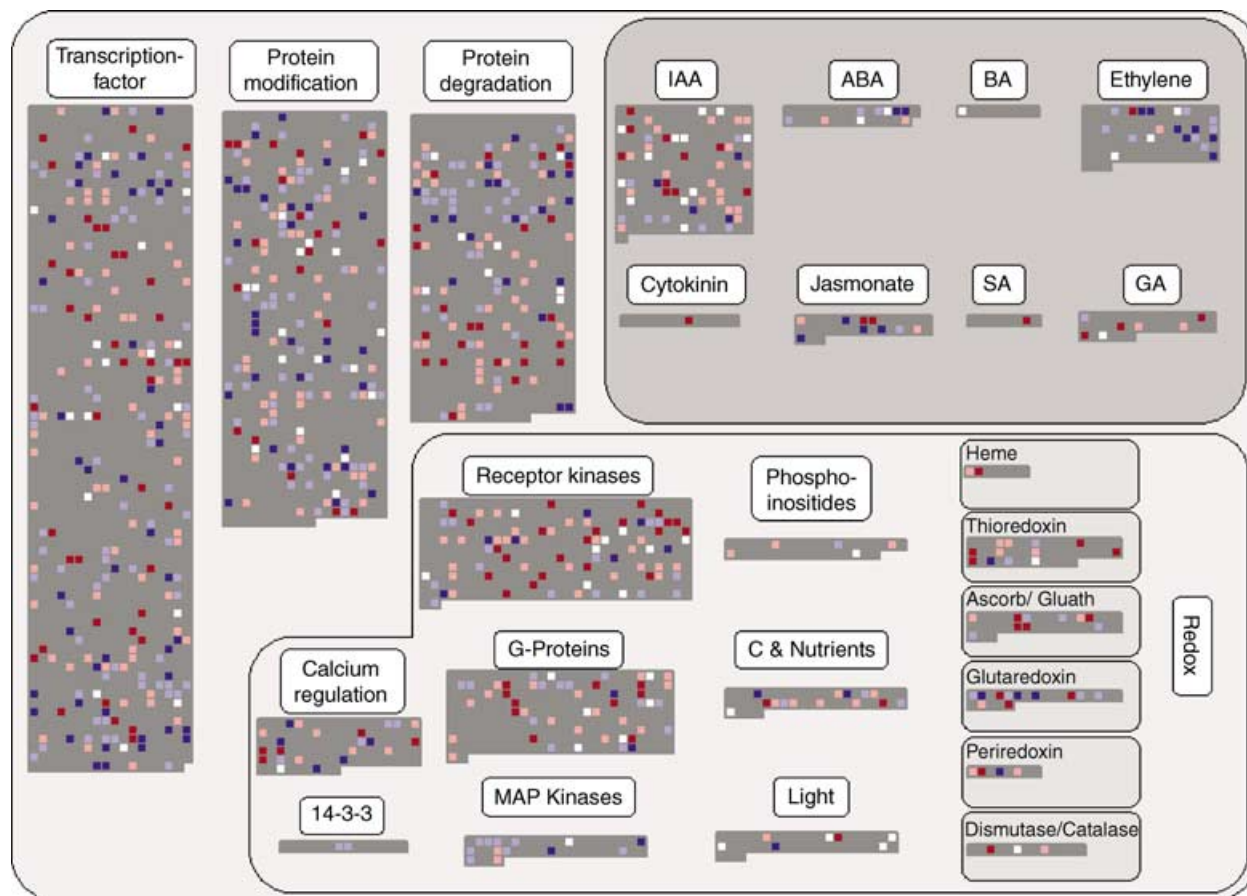


Figure 11. Display of genes involved in general regulatory processes.

The figure shows an overlay of the response in an extended night and in the *pgm* mutant (for procedure and scale see legend to Figure 4c). Plots of the original data for the extended night treatment are available in Supplementary Material. This figure is best viewed, and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

in an extended night (see above, BIN 16.1). Several genes implicated in ABA sensing (BIN 17.1) were induced, including *ABI1*, two *ABI3*-binding proteins and an ABA-responsive element binding protein. One of the *ABI3*-binding genes was induced 350-fold, making it one of the most strongly upregulated genes in the entire response. There was also a trend to increased expression of genes involved in ethylene synthesis and sensing. Extension of the night induced several genes in the families that encode 1-aminocyclopropane-1-carboxylate synthase, and 1-aminocyclopropane-1-carboxylate oxidase (BIN 17.5), as well as several genes implicated in ethylene sensing, including the putative ethylene sensor ethylene response sensor (ERS)2, the transduction protein ethylene insensitive (EIN)3 and an EIN3-like protein 1, ethylene response factor 1 and an ethylene responsive factor (ERF)1 homologue, ethylene responsive element binding protein (EREBP)-1, AtERF2, an EREBP-3-like protein, AtERF4 and two EREBP-4 homologues, AtERF5 and a further EREBP5 homologue, AtERF6,

and four ER6-like proteins. Simultaneously, an EREBP5-like protein is repressed. A large subset of these genes was also induced in the *pgm* mutant.

This trend to repression of genes involved in cytokinin synthesis and induction of genes involved in ABA and ethylene synthesis and sensing is of interest for two reasons. First, these two sets of hormones have reciprocal effects on growth and senescence. Our results indicate that one of the early responses to low sugars is a transcriptional re-programming of hormone synthesis and sensing, which may inhibit growth and predispose towards senescence. Second, there is increasing evidence from genetic studies for cross-talk between sugar sensing and the ABA- and ethylene-sensing pathways (Brocard *et al.*, 2002; Laby *et al.*, 2000; Leon and Sheen, 2003; Rook *et al.*, 2001). Our results provide physiological evidence that sugars indeed modify the synthesis and response to these hormones. The changes in expression in low sugar are so widespread, however, that they raise the question whether the interac-

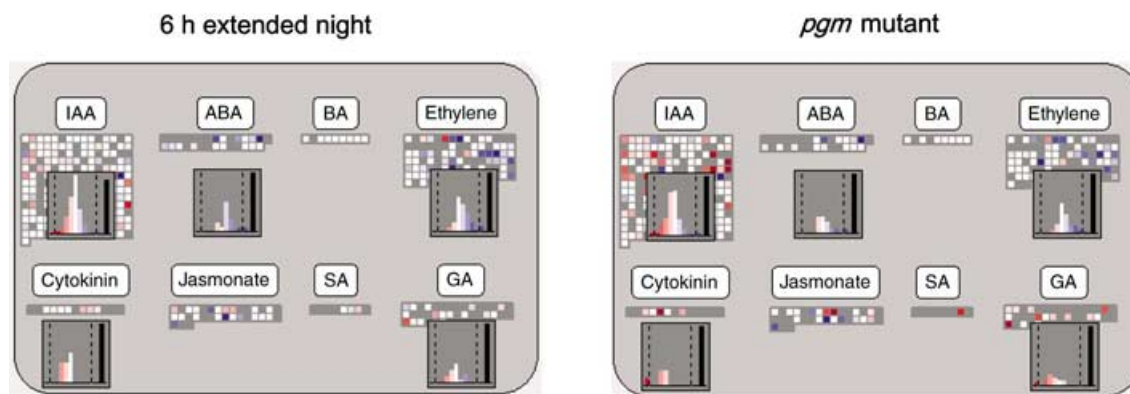


Figure 12. Frequency histogram showing the reciprocal responses of genes assigned to auxin, gibberellic acid and cytokinin synthesis and sensing, and ABA and ethylene synthesis and sensing.

The white bar corresponds to genes that change by <0.33 on a \log_2 scale, and the coloured bars to genes that change between 0.33–0.99, 0.99–1.66, 1.66–2.33, 2.33–3.0 and >3 . This figure is best viewed, and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

tions found in mutants are because of cross-talk between specific signalling pathways or reflect a more general interdependence of sugar and hormone status.

Induction and repression of large numbers of further genes encoding transcription factors, protein kinases and components of the protein degradation machinery

Both treatments led to widespread marked changes in expression of a wide range of transcription factors, protein kinases and components of the protein degradation machinery, receptor kinases and components of signalling pathways (Figure 11). These results indicate that low sugar triggers, within hours, a far-reaching rewiring of many regulatory networks. The subpanel 'C & Nutrients' (Figure 11) summarises the responses of some genes, which are already implicated in the regulation of carbon or nutrient responses. They both slightly repressed P_{II} (signal transduction protein), which by analogy to fungi (Ninfa and Atkinson, 2000) may be involved in the regulation of carbon–nitrogen interactions (Hsieh *et al.*, 1998; Smith *et al.*, 2003). AMP-regulated protein kinases and sucrose nonfermenting (SNF)1-like kinases play an important role in the regulation of central metabolism in mammals and yeast (Halford *et al.*, 2003). They slightly repressed a putative AMP-activated protein kinase homologue, repressed *AtSRPK1* (an SNF1-related protein kinase), and induced a related protein kinase (*AKIN10*). They also repressed an *NIRF3* (N,95/KBP90-like RING finger protein) homologue and repressed slightly repressor of gal-3, DELLA family protein (*RGA1*), which may be involved in the regulation of nitrogen metabolism (Truong *et al.*, 1997).

Display of metabolite data

Figure 13 shows the changes of metabolites in Col-0 wild-type rosette 8 h into an extended night, compared to the

end of the night. The raw data are provided in Supplementary Material. As expected, sucrose, hexose phosphates and most organic acids fall to low levels. Maltose decreased over 800-fold between the end of the night, falling below the detection limit after 8-h extended darkness. This is consistent with it being a product of starch breakdown. Pyruvate remained high for the first 8 h and then decreased (data not shown).

Many amino acids rose markedly in the first 8 h of the extended night, and this became even more accentuated after 24 h (not shown). This is because of protein catabolism (darkening led to a gradual decrease of total leaf protein, data not shown). Several intermediates in the amino acid biosynthesis pathways have been identified in the GC/liquid chromatography (LC) MS profiles. Three of these decreased, providing indirect evidence for an inhibition of amino acid synthesis. The fivefold decrease of *O*-acetylserine (the carbon acceptor for the sulphide) underlines the importance of the repression of sulphate transport and assimilation. Several intermediates in amino degradation pathways have been identified. Of these, four (urea, allantoin, indole-3-acetonitrile and β -alanine) decreased, which is consistent with the proposed stimulation of amino acid degradation. β -Alanine is also a participant in nucleotide breakdown.

Strikingly, most minor sugars remained unaltered and some, including xylose, ribose and rhamnose, increased. This is consistent with increased cell wall degradation, as proposed already on the basis of the expression array data. Glycerol, glycerol-3-phosphate and several fatty acids decrease, indicating that these metabolites are being remobilised for respiration. There was a marked decrease of ascorbate, indicating that this redox protectant is being re-mobilised as a carbon source. This was accompanied by a trend to decreased expression of enzymes involved in ascorbate synthesis and turnover, and of ascorbate-dependent peroxidases (see Figure 11, BIN 21.1). Another striking

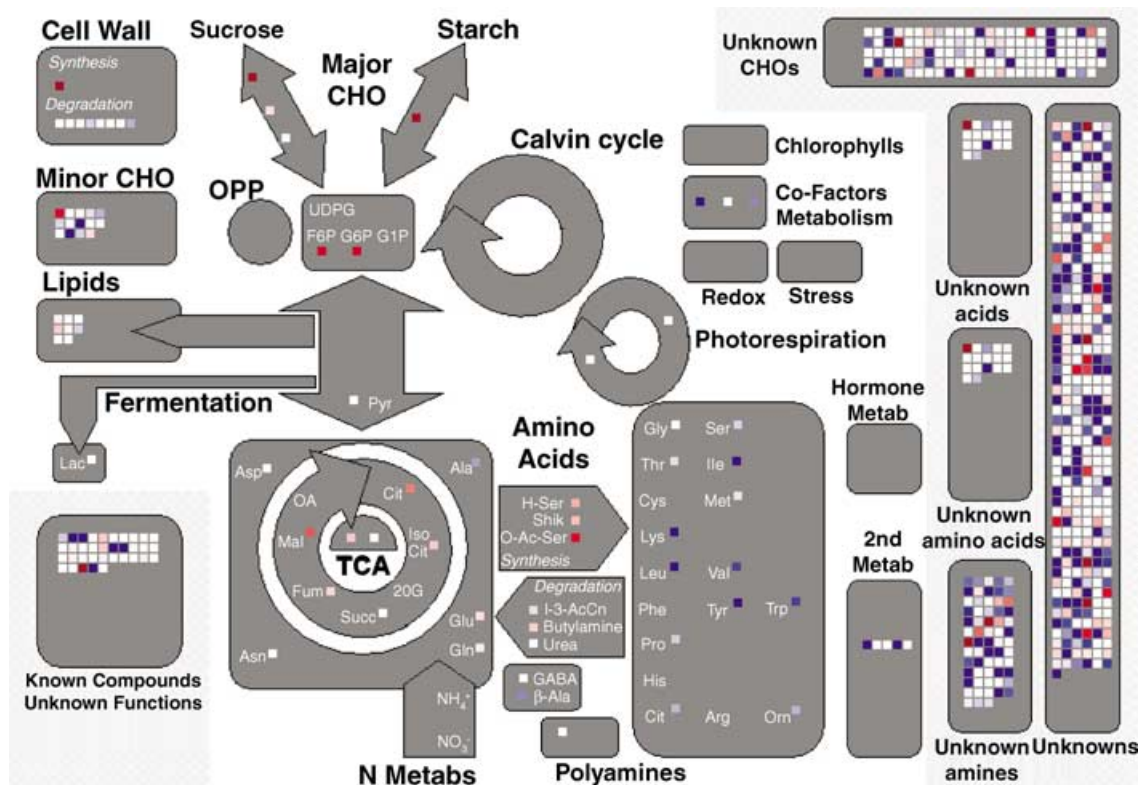


Figure 13. Display of changes in metabolites revealed by GC/MS metabolite profiling.

The samples were harvested at the end of the normal night and after an 8-h extension of darkness ($n = 4$ separately extracted replicates, each consisting of a pool of three plants). A colour scale was used, which in a slight coloration represents 60% change, and the response saturates at an eightfold change. This figure is best viewed, and all data point annotations are provided at <http://gabi.rzpd.de/projects/MapMan/> (see Experimental procedures).

result is that a large proportion of the unknown carbohydrates and a very large proportion of the unknown amines increased. These results indicate that many of these compounds may be involved in catabolic pathways.

MAPMAN reveals a coordinated and multilevel response to low sugar

Depletion of carbohydrates during a 6-h extension of the night is accompanied by wide-reaching and multilayered changes in gene expression. MAPMAN aided analysis of this complex response in several ways. The superimposition of different data sets in overlay plots provides a sensitive tool to identify shared features of different responses at a function-to-function or a gene-to-gene level. This revealed that very similar changes are found at the end of the night in the starchless *pgm* mutant, where leaf sugars are prematurely depleted, providing evidence that the changes are triggered by low sugar. Indeed, most of the changes can be reversed by addition of sugar (W.-R. Scheible, data not shown). By grouping genes that are probably involved in common area of function, it revealed trends towards repression or induction, which are less obvious at the single-gene level. Grouping also revealed when a large

proportion of the genes in a functional area were changing in opposite directions. This information, which cannot be obtained without grouping, provides evidence for re-organisation of many areas of function. MAPMAN allowed data to be viewed at different levels of resolution, at a global level, or at the level of discrete processes or, in cases where enough background information was available, at a very high resolution on a precise diagram of a biological process. This not only facilitated close analysis of particular processes, but also allowed changes in one area to be viewed and interpreted in the context of other processes.

One set of changes involves adjustments of genes involved in metabolism, which will stabilise energy metabolism. Within a few hours, many genes involved in photosynthesis, starch and sucrose synthesis, transport and assimilation of nitrate and sulphate, amino acid and nucleotide synthesis, lipid synthesis and synthesis of some cell wall components are repressed, as are many genes required for nutrient uptake, and metabolite import into the plastids. Many genes involved in the catabolism of amino acids, nucleotides, phospholipids and cell wall components are induced. The switch from anabolism to catabolism is accompanied by changes in the expression of genes involved in central carbon metabolism and transport, to facilitate flexible use of a wide range of

substrates. A second set of changes involves widespread changes in the expression of a large proportion of the enzymes involved in secondary metabolism. A third set involved changes in expression of genes involved in cellular functions related to growth. Examples include preferential repression of many genes involved in RNA and protein synthesis, and cell wall modification. This, together with the inhibition of amino acid, nucleotide and lipid synthesis and plastid envelope transporters, points to a broad downregulation of cellular growth. A fourth set of changes relate to processes that maintain or regulate cell or organ function. Many genes involved in ascorbate synthesis and turnover were repressed (not shown). Downregulation of genes involved in cytokinin synthesis, upregulation of genes involved in ABA and ethylene synthesis and signalling and preferential repression of genes involved in DNA repair indicate that the rosette is already preparing for senescence. Finally, there were widespread changes in genes involved in regulation. Marked changes in trehalose metabolism indicate a possible role for Tre6P as a low carbon signal. There were also changes in expression of a strikingly large proportion of genes that encode receptor kinases, transcription factors, protein kinases and phosphatases, components of protein degradation, indicating that extensive rewiring of regulatory circuits occurs as a rapid response to sugar depletion.

The finding that sugar depletion leads, within hours, to widespread changes in fundamental processes that contribute to growth and maintenance underlines the importance of an appropriate allocation of carbon between export, growth and storage during the diurnal period. Growth of the starchless *pgm* mutant in a alternating light/dark regime (Caspar *et al.*, 1985; Schulze *et al.*, 1994) is probably impaired not only because rapid respiration in the first part of the night leads to loss of carbon but also because the following period of sugar-deficiency disrupts the regulation of many important cellular processes.

Further development and availability of MAPMAN

MAPMAN requires further development to remedy deficits and extend the applications. Deficits include the current quality of gene annotation, inaccuracies in assignments and the crude substructuring of many areas of function. This will be addressed by incorporating sources of expert annotation in the public domain (e.g. <http://www.arabidopsis.org/info/genefamily/genefamily.html>, <http://aramemnon.botanik.uni-koeln.de/>, <http://www.plantbiology.msu.edu/lipids/genesurvey>) and (see below) by mobilising expert input. It will also be important to develop the SCAVENGER modules to facilitate automatic import of TIGR annotation updates and GOC releases, while screening out errors and unnecessary redundancies that necessitated extensive manual work in the first build. An unavoidable weakness of the present build is that there is not enough reliable information

to assign more than a fraction of the enzymes to a specific subcellular compartment, cell type or organ. With respect to applications, we plan to refine MAPMAN to display absolute expression levels or metabolite concentrations, and to incorporate modules that allow statistical analysis. An important application is to superimpose data from different experiments, to facilitate inspection of large data sets for global or local similarities. This application was illustrated in the present article, where it was achieved by re-organising the original EXCEL data files. It is planned to add a further software module to allow automated display of new data sets against a portfolio of reference data sets. It is also planned to develop modules that display mathematically generated clusters onto diagrams, in order to view the results of an unbiased data analysis in different biological contexts.

MAPMAN is available at a website (<http://gabi.rzpd.de/projects/MapMan/>), from which the IMAGEANNOTATOR module (including current mapping files and image maps) and instructions on how to introduce experimental data sets can be downloaded. This will allow users to analyse and view their data using the categories and assignments from the first build of the SCAVENGER modules. They can also use the IMAGEANNOTATOR module in combination with SCAVENGER modules and diagrams that they develop themselves. Improvement and correction of existing mapping files provided in the downloaded version will require access to the SCAVENGER modules. The latter will not, initially, be downloadable but will be provided on request (Experimental procedures) to any user without charge or material transfer agreement (MTA) immediately, but with a request for expert input in a timely manner into a mutually agreed section of the SCAVENGER module. This input will be incorporated into MAPMAN, the mapping files updated in the downloadable IMAGEANNOTATOR version, and the expert input acknowledged on the MAPMAN website.

Experimental procedures

TRANSCRIPTSCAVENGER

All work was carried out on a state-of-the-art laptop (Fujitsu-Siemens E series lifebook) and comparable PC with a minimum of 512 Mb RAM and 2.4 GHz Pentium IV processor. Combination of gene entries, assignments as well as categories translations were made with standard Microsoft Office package (EXCEL and ACCESS 2000), ACTIVEPERL-5.6.1.635 and JAVA 1.4.1.

Origins of the BIN and subBIN structure

Initially 23 BINs were manually defined, corresponding to different areas of metabolism (see Table 1). Each BIN was given a corresponding numerical code (e.g. 'photosynthesis' = 1), which can be extended in a hierarchical manner (e.g. the subBINs 'light reactions', 'photorespiration' and 'Calvin cycle' = 1.1, 1.2 and 1.3, respectively). Most of the BINs were initially imported from the

AraCyc (metabolism-associated genes) and the GOC (DNA, RNA, protein synthesis and modification-associated genes, signalling, transporter, redox, C1 metabolism, stress, cell, development-associated genes). 'Metabolism-associated genes' was split based into 23 areas (corresponding to BINs 1–14, 16–19 and 22–26, see Supplementary Material). Some BINs were created during subsequent activities ('metal handling', 'unknown'). Pathway annotations given by TAIR and GOC were checked manually for redundancy and compatibility to self-developed BIN structure (see Supplementary Material; based on TIGR release version 3.0).

Integration of categorised gene annotation

Gene chip annotation was downloaded from the Affymetrix homepage (<ftp://ftp.arabidopsis.org/home/tair/Microarrays/Affymetrix/>). Lists of 1476 metabolism-related genes and 2812 genes annotated to major cellular functions were downloaded as tab-delimited text files from the TAIR ftp server (<ftp://ftp.arabidopsis.org/home/tair/Genes/AraCyc/>, ftp://ftp.arabidopsis.org/home/tair/Genes/Gene_Ontology/, respectively). Genes imported from TAIR and GOC were divided between 23 manually BINs (see Table 1). The categories transport, DNA synthesis, RNA synthesis, RNA processing, regulation of RNA transcription, amino acid activation, protein synthesis, protein targeting, protein modification, protein degradation, signalling, C1 metabolism, cell, cell division, cell cycle, development and stress responses were taken over as BINs or subBINs. All genes with appropriate categories were connected via their AgiCode to the unique Affymetrix identifiers, and automatically assigned to BINs via ACCESS queries. The numbers of entries generated by the initial download from public databases are given in column 2 of Table 1.

Assignments made by text search algorithms using KEGG annotation

Another 3080 entries were recruited via a text search in the functionally categorised KEGG database. Text search with enzyme and alternative enzyme names was performed using JAVA scripts. The gene description of genes on the Affymetrix chip (TIGR release version 3.0) was searched (case-insensitive) for enzyme names (and their synonyms) provided by KEGG (<ftp://ftp.genome.ad.jp/pub/kegg/ligand/>). If a search string was found, the KEGG prediction that 'enzyme/gene y has a function in pathway/process x' and 'EC number' attributes were attached to existing list of non-assigned genes. To avoid trivial matches a list with exclusive keywords (e.g. ATPase, isomerase, PGM) was generated manually, which the script ignores. The KEGG 'function in pathway' attribute was then translated into BINs using an automatised procedure.

Assignments made by text search algorithms using user-defined keywords

A further 6131 genes were incorporated via a text search of TIGR release version 3.0 with manually pre-defined keywords. Genes that were still not in BINs were listed alphabetically to create blocks of genes with similar annotations, which were scanned manually to identify common-specific keywords. The keywords were used to assign further genes to existing BINs. Where necessary, new BINs (metal handling, unknowns) were generated manually. Using a modified JAVA script, the Affymetrix list was re-screened with these or with similar keywords. If a string match was found, the corresponding BIN was attached to the gene. This allowed another 5600 genes to be assigned to BINs.

Manual correction of annotated binned genes

The automatic downloads and searches in total assigned about 7201 genes to BINs. There were 1879 multiple assignments. Another 4367 genes, which had a specified annotation in TIGR release version 3.0 but had not yet been assigned, were provisionally placed in a subBIN 'unassigned: no ontology', and 10 082 genes that were annotated as 'hypothetical or unknown genes' were placed in the subBIN 'unassigned: unknown'. A copy of the gene list was divided in three parts: (i) list of at least twofold annotated genes; (ii) unique assigned genes; and (iii) not categorised genes. After a manual check, the corrected gene assignments were updated in the original gene list using PERL scripts.

Manual generated single gene and metabolite annotation

To display data on a single gene level, genes taking part in glycolysis, citrate cycle, glyoxylate cycle, sucrose and starch metabolism and nitrate assimilation were annotated manually. Manual annotation was also used to assign metabolites to pathways.

IMAGEANNOTATOR

The IMAGEANNOTATOR module is written in JAVA 1.4.1 and runs on different operating systems (Windows, Mac OS X, Linux, Unix, etc.). It is built in a modular fashion to allow seamless integration of modules for a web-based solution. New visualisation types can be added and instantly deployed as web services.

The IMAGEANNOTATOR is available as a JAVA application version for local application, and can be downloaded from the website <http://gabi.rzpd.de/projects/MapMan/>. It is also available as a JAVA servlet version, which is also accessible via <http://gabi.rzpd.de/projects/MapMan/>. Data are accessed across an object-oriented interface from a relational database (Oracle, version 8.1.7.; Oracle Corporation, Redwood Shores, CA, USA). The downloadable installers includes: (i) the four Affymetrix and two GC/MS experimental data sets presented in this paper; (ii) a selection of schematic maps of metabolism and cellular processes, four maps of metabolic pathways resolved to the individual step level, and a map of metabolism for display of metabolites; and (iii) three mapping files, one of which structures the *Arabidopsis* genes represented on the 22K Affymetrix™ array into BINs and subBINs for display on the schematic maps of metabolism and cellular processes, the second structures a subset of these for display on the highly resolved maps of metabolic pathways, and the third structures metabolites. The experimental data files, map files and mapping files can be moved to a user-defined directory during the program installation process. Users can choose images and mapping files from a menu, view the sample data sets provided with the database and upload their own data.

The mapping files are encrypted and only readable by the IMAGEANNOTATOR. Updated versions of mapping files will be made available at intervals for download. Changes in the mapping files require access to the SCAVENGER modules. These will not be available for automatic downloading, but will be made available on request (please enquire at mstitt@mpimp-golm.mpg.de, using the words 'request for TranscriptScavenger Mapping File' in the title) under the conditions summarised in the paper: briefly, it will be requested that the recipient undertakes to provide expert input into a mutually agreed part of the file and return the suggested improvements to the deliverer in a timely manner, and that the user does not provide the file to further users without consent. The suggested improvements will be incorporated into the next

update of the mapping file, and the contribution acknowledged on the central website. We hope to generate a system in which a large number of users provide expert input into improving a centrally available resource.

Plant growth and harvest

Arabidopsis thaliana Col-0 and the *pgm* mutant were germinated and grown on a 2 : 1 (v/v) of GS90-soil (N: 50–300 mg l⁻¹, P₂O₅: 80–300 mg l⁻¹, K₂O: 80–400 mg l⁻¹, pH 5.5–6.5; Einheitserdewerk, Uetersen, Germany) and vermiculite. Plants were germinated for 7 days (16 h light at 20°C, night at 6°C, 200 µE fluorescent light, 60–70% relative humidity), transferred to short-day condition (8 h light at 20°C, dark at 16°C, 180 µE fluorescent light, 60–70% relative humidity), picked after day 14 into pots and incubated under short day further 7 days. For harvesting Col-0 and *pgm* at the end of night, single plants grew from day 22 on for further 14 days in a 12-h photoperiod (20°C, 150 µE, 60–70% relative humidity). In the extended night experiment, the Col-0 ecotype grew from day 22 on for 11 days in a 14-h photoperiod (20°C, 150 µE fluorescent light, 60–70% relative humidity). Samples were taken at the end of night and after 6-h extended darkness. At each time, 15 rosettes were divided into 5 replicates of 3 rootless non-flowering rosettes, immediately frozen in ambient in the dark in liquid nitrogen, separately powdered under liquid nitrogen and stored at –80°C.

Preparation of RNA, cDNA and labelled cRNA

Equal amounts of five biological replicates, representing subaliquots from in total 15 rosettes, were pooled. Total RNA was prepared with TRIzol according to the manufacturer's instructions (Invitrogen Life Technologies, Karlsruhe, Germany). RNA quality and quantity was checked visually using denaturing gel electrophoresis, by analysis with the Bioanalyzer 2100 (Agilent Technologies, Böblingen, Germany) and by photometric analysis at 200–400 nm (OD_{260/280}). Ten micrograms of total RNA was used for double-stranded cDNA synthesis with the cDNA Synthesis System (Roche, Mannheim, Germany). Biotin-labelled cRNAs were synthesized by *in vitro* transcription (T7-Megascript-Kit, Ambion, Cambridgeshire, UK). Quality and quantity of each cRNA sample were determined by analysis on the Bioanalyzer 2100 (mRNA Smear Nano Assay, Agilent Technologies) according to the manufacturer's instructions, by photometric analysis at 200–400 nm (OD_{260/280}) and by hybridisation on the test3-array (Affymetrix UK Ltd.: <http://www.affymetrix.com>) to assess sample quality by examination of 3'–5' intensity ratios of housekeeping genes.

Array hybridisation and data evaluation

Fifteen micrograms of biotinylated cRNA was hybridised on the GeneChip *Arabidopsis* ATH1 Genome Array (part no. 900385, Affymetrix UK Ltd.) for 16–18 h at 45°C and 60 r.p.m. (Fluidics Station 400, Affymetrix UK Ltd.). Spike controls for *bioB*, *bioC*, *bioD* and *cre1-1* at concentrations of 1.5, 5, 25 and 100 pM were included according to the manufacturer's instructions (Affymetrix UK Ltd.). Array washing and staining were controlled by the Affymetrix Microarray Suite (MAS) 5.0 using protocols *micro1_v1* and *EukGe_WS2v4* for the test3-array and the ATH1-array, respectively. Scanning of chips was performed with the G2500A Gene Array Scanner (Agilent Technologies) controlled by MAS 5.0. Raw signal intensities were scaled to identical

trimmed mean intensity of all scaled signals, according to standard procedures (Affymetrix Microarray Suite User's Guide, version 5.0 (MAS 5.0)). For all probe sets of each chip, the median signal was scaled to a target intensity = 100 in the MAS 5.0 software. Default values for the present/absent filter were $\alpha_1 = 0.05$, $\alpha_2 = 0.065$ for 11 probe pairs/probe set. Genes with a *P*-value < 0.04 were detected as 'present' call, *P*-value > 0.06 obtained 'absent' calls and *P*-values ranging from 0.04 to 0.06 were 'marginally present'.

Calculations and criteria for the overlay displays

The EXCEL experimental files for the extended night and *pgm* treatment were organised according to the unique identifier and combined. Genes were sorted into five groups: genes whose value (on log₂ scale) was ≥ 1.0 in both treatments (markedly induced in both experiments), genes whose value was between 0.5 and 1.0 in one experiment but ≥ 0.5 in the other (both are at least slightly induced: this class includes some genes that rise markedly in one treatment and less strongly in the other), genes whose value lay between 0.5 and –0.5 in both experiments ('no change'), genes whose value was between –0.5 and –1.0 in one experiments but ≤ -0.5 in the other (both are at least slightly repressed: this class includes some genes that are repressed markedly in one treatment and less strongly in the other), genes whose value was ≤ -1.0 in both experiments (strongly repressed in both treatments), and genes whose value was ≥ 0.5 in one experiment and ≤ -0.5 in the other (opposite response). A new column was created in the EXCEL file into which a notional experimental value was entered that allowed each group to be displayed with a different false colour: 4, 1.5, 'X', –1.5, –4 and 0, which is translated via the false colour module of the IMAGEANNOTATOR into a dark blue, a light blue, a grey (X is the symbol for genes called 'not present'), light red, dark red and white, respectively. The lower cut-off of 0.5 was empirically selected by the following process. Genes in the combined EXCEL file were sorted into the above groups, but with a lower cut-off of 0.8, 0.7, 0.6, 0.5, 0.4 or 0.3. The number of genes in the groups 'opposite response', 'both at least slightly induced' and 'both at least slightly repressed' were plotted against the cut-off value (the Figure is provided in Supplementary Material). Recruitment of false positives to the shared response increases as the cut-off filter is reduced, and can be estimated because it will be similar to the number of recruits to the opposed response group. With a lower cut-off of 0.8, 0.7, 0.6, 0.5, 0.4 and 0.3, the number of genes in the shared response was 124, 173, 261, 407, 642 and 986, and the total number of genes in the 'both at least weakly induced' and 'both at least weakly repressed' were 779, 1304, 1950, 2733, 3645 and 4721, respectively. A cut-off of 0.5 was taken because lowering the cut-off from 0.6 to 0.5 led to substantial recruitment of genes to the 'shared response' groups (783), but only slightly increased (146) the number of genes in the 'opposite response' group.

Metabolite analysis

Samples were extracted and analysed by GC/MS time of flight (TOF) as described by Fiehn *et al.* (2000).

Acknowledgements

We thank Oliver Fiehn for aid in GC/MS profiling, Melanie Höhne and Manuela Günter for excellent support in plant growth, harvesting, extraction and analysis. We also thank the RZPO team, namely

Stefanos Petrakis for skilled programming and many useful suggestions, Florian Wagner for hybridisation procedures and data normalisation and Iris Bertram for her support in graphical design. The IMAGEANNOTATOR was developed within the BMBF-funded projects GABI-Primary Database (0312272). The TRANSCRIPTSCAVENGER and METABOLITSCAVENGER as well as the experimental studies reported in this paper were supported by the BMBF-funded project GABI Verbund Arabidopsis III 'Gauntlets, 'Carbon and Nutrient Signalling: Test Systems, and Metabolite and Transcript Profiles' (0312277A).

Supplementary Material

The following material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/TPJ/TPJ2016/TPJ2016sm.htm>

Four EXCEL tables, with raw expression data sets for the two experimental treatments:

- list of BINs, subBINs and number of assigned genes;
- raw data pgm and Col-0 – end of night.xls;
- raw data Col-0 extended night – 0 and 6 h.xls;
- raw data metabolites ext night 0vs8.xls.

Table S1 Greatly varying genes

Figure S1. Supports Figure 4.

Figure S2. (a,b) Detailed display of sucrose and starch metabolism.

Figure S3. Supports Figure 6.

Figure S4. Supports Figure 9.

Figure S5. (a,b) Supports Figure 10.

Figure S6. Supports Figure 11.

Figure S7. Supports Experimental procedures.

References

- Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
- Beisson, F., Gardies, A.-M., Teissere, M., Ferte, N. and Noat, G.** (1997) An esterase neosynthesized in post-germinated sunflower seed is related to a new family of lipolytic enzymes. *Plant. Physiol. Biochem.* **35**, 761–765.
- Bonini, B.M., Van Vaeck, C., Larsson, C., Gustafsson, L., Ma, P., Winderickx, J., Van Dijck, P. and Thevelein, J.M.** (2000) Expression of *Escherichia coli* otsA in a *Saccharomyces cerevisiae* *tps1* mutant restores trehalose 6-phosphate levels and partly restores growth and fermentation with glucose and control of glucose influx into glycolysis. *Biochem. J.* **350**, 261–268.
- Brocard, I.M., Lynch, T.J. and Finkelstein, R.R.** (2002) Regulation and role of the *Arabidopsis* abscisic acid-insensitive 5 gene in abscisic acid, sugar, and stress response. *Plant Physiol.* **129**, 1533–1543.
- Caspar, T., Huber, S.C. and Somerville, C.** (1985) Alterations in growth, photosynthesis, and respiration in a starchless mutant of *Arabidopsis thaliana* (L.) deficient in chloroplast phosphoglucomutase activity. *Plant Physiol.* **79**, 11–17.
- Caspar, T., Lin, T.P., Monroe, J., Bernhard, W., Spilatro, S., Preiss, J. and Somerville, C.** (1989) Altered regulation of beta-amylase activity in mutants of *Arabidopsis* with lesions in starch metabolism. *Proc. Natl. Acad. Sci. USA*, **86**, 5830–5833.
- Celis, J.E., Kruhoffer, M., Gromov, I. et al.** (2000) Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **480**, 2–16.
- De Risi, J.L., Iyer, V.R. and Brown, P.O.** (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.
- Dulloo, A.G., Samec, S. and Seydoux, J.** (2001) Uncoupling protein 3 and fatty acid metabolism. *Biochem. Soc. Trans.* **29**, 785–791.
- Eastmond, P.J. and Graham, I.A.** (2003) Trehalose metabolism: a regulatory role for trehalose-6-phosphate? *Curr. Opin. Plant. Biol.* **6**, 231–235.
- Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L.** (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161.
- Grosu, P., Townsend, J.P., Hartl, D.L. and Cavalieri, D.** (2002) Pathway Processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res.* **12** (7), 1121–1126.
- Halford, N.G., Hey, S., Jhurrea, D., Laurie, S., McKibbin, R.S., Paul, M. and Zhang, Y.** (2003) Metabolic signalling and carbon partitioning: role of Snf1-related (SnRK1) protein kinase. *J. Exp. Bot.* **54**, 467–475.
- Hsieh, M.H., Lam, H.M., Van De Loo, F.J. and Coruzzi, G.** (1998) A PII-like protein in *Arabidopsis*: putative role in nitrogen sensing. *Proc. Natl. Acad. Sci. USA*, **95**, 13965–13970.
- Iuchi, S., Kobayashi, M., Taji, T., Naramoto, M., Seki, M., Kato, T., Tabata, S., Kakubari, Y., Yamaguchi-Shinozaki, K. and Shinozaki, K.** (2001) Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. *Plant J.* **27**, 325–333.
- Klein, D., Krapp, A. and Stitt, M.** (2000) The regulation of NIA expression by nitrate and glutamine is overridden when sugars decrease to low levels in tobacco leaves. *Plant Cell Environ.* **23**, 863–871.
- Laby, R.J., Kincaid, M.S., Kim, D. and Gibson, S.I.** (2000) The *Arabidopsis* sugar-insensitive mutants *sis4* and *sis5* are defective in abscisic acid synthesis and response. *Plant J.* **23**, 587–596.
- Lam, H.M., Peng, S.S. and Coruzzi, G.M.** (1994) Metabolic regulation of the gene encoding glutamine-dependent asparagine synthetase in *Arabidopsis thaliana*. *Plant Physiol.* **106**, 1347–1357.
- Leon, P. and Sheen, J.** (2003) Sugar and hormone connections. *Trends Plant Sci.* **8**, 110–116.
- Melo-Oliveira, R., Oliveira, I.C. and Coruzzi, G.M.** (1996) *Arabidopsis* mutant analysis and gene regulation define a nonredundant role for glutamate dehydrogenase in nitrogen assimilation. *Proc. Natl. Acad. Sci. USA*, **93**, 4718–4723.
- Michaut, L., Flister, S., Neeb, M., White, K.P., Certa, U. and Gehring, W.J.** (2003) Analysis of the eye developmental pathway in *Drosophila* using DNA microarrays. *Proc. Natl. Acad. Sci. USA*, **100**, 4024–4029.
- Mifflin, B.J. and Habash, D.Z.** (2002) The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. *J. Exp. Bot.* **53**, 979–987.
- Müller, L.A., Zhang, P. and Rhee, S.Y.** (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* **132**, 453–460.
- Ninfa, A.J. and Atkinson, M.R.** (2000) PII signal transduction proteins. *Trends Microbiol.* **8**, 172–179.
- Rook, F., Corke, F., Card, R., Munz, G., Smith, C. and Bevan, M.W.** (2001) Impaired sucrose-induction mutants reveal the modulation of sugar-induced starch biosynthetic gene expression by abscisic acid signalling. *Plant J.* **26**, 421–433.
- Rylott, E.L., Hooks, M.A. and Graham, I.A.** (2001) Co-ordinate regulation of genes involved in storage lipid mobilization in *Arabidopsis thaliana*. *Biochem. Soc. Trans.* **29**, 283–287.

- Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M. and Wisman, E.** (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell*, **13**, 113–123.
- Schluepmann, H., Pellny, T., Van Dijken, A., Smeekens, S. and Paul, M.** (2003) Trehalose 6-phosphate is indispensable for carbohydrate utilization and growth in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*, **100**, 6849–6854.
- Schulze, W., Schulze, E.D., Stadler, J., Heilmeyer, H., Stitt, M. and Mooney, H.A.** (1994) Growth and reproduction of *Arabidopsis thaliana* in relation to storage of starch and nitrate in the wild-type and in starch-deficient and nitrate-uptake-deficient mutants. *Plant Cell Environ.* **17**, 795–809./reference>
- Schwartz, S.H., Qin, X. and Zeevaert, J.A.** (2001) Characterization of a novel carotenoid cleavage dioxygenase from plants. *J. Biol. Chem.* **276**, 25208–25211.
- Smith, C.S., Weljie, A.M. and Moorhead, G.B.** (2003) Molecular properties of the putative nitrogen sensor PII from *Arabidopsis thaliana*. *Plant J.* **33**, 353–360.
- Stitt, M.** (1990) Fructose-2,6-bisphosphate as a regulatory metabolite in plants. *Annu. Rev. Plant Physiol. Mol. Biol.* **41**, 153–185.
- Stitt, M. and Fernie, A.R.** (2003) From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon–nitrogen interactions. *Curr. Opin. Biotechnol.* **14**, 136–144.
- Truong, H.N., Caboche, M. and Daniel-Vedele, F.** (1997) Sequence and characterization of two *Arabidopsis thaliana* cDNAs isolated by functional complementation of a yeast *gln3 gdh1* mutant. *FEBS Lett.* **410**, 213–218.
- Wang, R., Okamoto, M., Xing, X. and Crawford, N.M.** (2003) Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* **132**, 556–567.