Plant Molecular Biology (2006) 60:773–792
DOI 10.1007/s11103-005-5772-4
© Springer 2006

Conversion of MapMan to allow the analysis of transcript data from *Solanaceous* species: effects of genetic and environmental alterations in energy metabolism in the leaf

Ewa Urbanczyk-Wochniak[1,‡], Björn Usadel[1,‡,*], Oliver Thimm[1,5], Adriano Nunes-Nesi[1], Fernando Carrari[1,6], Marcus Davy[2], Oliver Bläsing[1], Magdalena Kowalczyk[3], Daniel Weicht[1], Anna Polinceusz[1], Svenja Meyer[4], Mark Stitt[1] and Alisdair R. Fernie[1,*]
[1]*Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, Golm, 14476, Germany (*author for correspondence; e-mails usadel@mpimp-golm.mpg.de; fernie@mpimp-golm.mpg.de);* [2]*Hort Research Mt Albert, Private Bag 92 169, Auckland, New Zealand;* [3]*Department of Plant Genetics, Breeding and Biotechnology, Faculty of Horticulture, Warsaw Agricultural University, Nowoursynowska 166, Warsaw, 02-787, Poland;* [4]*RZPD German Resource Centre for Genome Research, Heubnerweg 6, Berlin S, D-14059, Germany;* [5]*CNAP, Department of Biology (Area 7), University of York, York, YO10 5YW, UK;* [6]*Instituto de Biotecnología, CICVyA INTA, Castelar, B1712WAA, Argentina;* [‡]*These author contributed equally*

Received 23 September 2005; accepted in revised form 8 December 2005

Key words: photosynthesis, *Solanum tuberosum*, *Solanum lycopersicon*, transcript analysis, tricarboxylic acid cycle, visualisation

Abstract

The tomato microarray TOM1 offers the possibility to monitor the levels of several thousand transcripts in parallel. The microelements represented on this tomato microarray have been putatively assigned to unigenes, and organised in functional classes using the MapMan ontology (Thimm *et al.*, 2004. Plant J. 37: 914–939). This ontology was initially developed for use with the *Arabidopsis* ATH1 array, has a low level of redundancy, and can be combined with the MapMan software to provide a biologically structured overview of changes of transcripts, metabolites and enzyme activities. Use of this application is illustrated using three case studies with published or novel TOM1 array data sets for *Solanaceous* species. Comparison of previously reported data on transcript levels in potato leaves in the middle of the day and the middle of the night identified coordinated changes in the levels of transcripts of genes involved in various metabolic pathways and cellular events. Comparison with diurnal changes of gene expression in *Arabidopsis* revealed common features, illustrating how MapMan can be used to compare responses in different organisms. Comparison of transcript levels in new experiments performed on the leaves of the cultivated tomato *S. lycopersicum* and the wild relative *S. pennellii* revealed a general decrease of levels of transcripts of genes involved in terpene and, phenylpropanoid metabolism as well as chorismate biosynthesis in the crop compared to the wild relative. This matches the recently reported decrease of the levels of secondary metabolites in the latter. In the third case study, new expression array data for two genotypes deficient in TCA cycle enzymes is analysed to show that these genotypes have elevated levels of transcripts associated with photosynthesis. This in part explains the previously documented enhanced rates of photosynthesis in these genotypes. Since the *Solanaceous* MapMan is intended to be a community resource it will be regularly updated on improvements in tomato gene annotation and transcript profiling resources.

Abbreviations: cDNA, complementary DNA; GO, gene ontology; EST, expressedsequence tag

Introduction

Microarray expression analyses offer an unprecedented tool for the systematic evaluation of biological responses. The tomato, expressed sequence tag (EST) collection described in van der Hoeven *et al.* (2002) is available in printed form on the TOM1 microarray, and represents a unique resource for the tomato research community. Currently, this EST resource is curated by CGEP (Centre for Gene Expression Profiling) at the BTI, Cornell University, The Geneva Agricultural Experiment Station and The USDA Federal Plant and Nutrition Laboratory. The first data sets utilizing this platform were recently published (Alba *et al.*, 2004; Fei *et al.*, 2004; Baxter *et al.*, 2005; Carbone *et al.*, 2005; Lemaire-Chamley *et al.*, 2005; Uppalapati *et al.*, 2005; Urbanczyk-Wochniak *et al.*, 2005). Smaller sets of ESTs have also been utilized in earlier transcript profiling studies (Mysore *et al.*, 2002; Urbanczyk-Wochniak *et al.*, 2003). These studies are providing important insights into the regulation of metabolism, photosensory receptors and light signalling proteins, defence responses to *Pseudomonas syringae*, phytohormone responses and events during fruit ripening. Due to extensive sequence homology, this array can be applied not only to tomato but also to related *Solanaceae* species, including potato as has been shown earlier (Urbanczyk-Wochniak *et al.*, 2005).

The huge number of transcripts measured in a single experiment raises challenges with respect to the evaluation and visualisation of the data. Numerous tools exist to group and cluster genes or experiments on the basis of their expression (*see for example* the TIGR multi-experiment from the TM4 suite; Saeed *et al.*, 2003). Clustering is well suited for finding classes of genes and/or responses, but it does not put the data into a meaningful background. Tools are also emerging that use ontologies to interpret expression data in terms of the known biological background (e.g. GOminer; Zeeberg *et al.*, 2003). To facilitate statistical analyses, such ontologies should be as non-redundant as possible.

The MapMan visualization software for *Arabidopsis thaliana* transcript and metabolite profiling (Thimm *et al.*, 2004; Usadel *et al.*, 2005) offers the possibility to paint out microarray and/or metabolite profiling experiments onto diagrams of pathways or processes, and visualize the responses of gene expression in a biological context. It is supported by a plant specific ontology, in which redundancy has been minimized. The principle of the MapMan ontology is a hierarchical ''BIN''-based structure. Each BIN comprises items of similar biological function e.g. glycolysis, and can be further split into subBINs corresponding to submodes of the biological function. In contrast to the GO ontology (The Gene Ontology Consortium, 2000), the MapMan BIN structure minimizes the redundancy between BINs. The hierarchical structure is reflected in its ''BINCode'' by separating dots similar to EC numbers. Thus BIN 1 (photosynthesis) can, for example, be subdivided to 1.1 (Photosynthesis.light reaction) and 1.2 (Photosynthesis.Calvin cycle). This allows the relationship between subBINs and BINS to be deduced immediately from the name of the BIN as well as from its numerical BINCode. This software can be used as a web-based application, or downloaded at http://gabi.rzpd.de/projects/MapMan/ to display and interpret *Arabidopsis* microarray expression experiments.

A very important goal is to extend this software platform to allow it to be applied to crop plants. This would allow knowledge generated by analysis of the model system *Arabidopsis* to be used to support crop genomic research. The present paper reports the classification of the spots represented on the tomato TOM1 array according to the MapMan ontology, and the establishment of a visualisation system for *Solanaceous* transcript profiling data based on the currently available tomato microarrays. The first part of the paper describes the approach we have taken to establish potential similarities between sequences in tomato and *Arabidopsis* genes, in order to build a classification scaffold for the spots represented on the TOM1 array. The second part presents three applications based on TOM1 experiments carried out in our laboratory, which illustrate the utility of this platform in facilitating biological interpretation of experiments. These include (i) previously published datasets from transcript profiling of potato through a diurnal cycle (Urbanczyk-Wochniak *et al.*, 2005), (ii) an unpublished dataset resulting from the comparison of the transcriptome of the elite cultivar *Solanum lycopersicum* with that of its wild species relative *S. pennellii* and (iii) an unpublished dataset resulting from the

transcript profiling of biochemically well characterized tomato leaves exhibiting genetic perturbations in the TCA cycle.

Materials and methods

Plant material

Solanum tuberosum L. cv. Desiree was obtained from Saatzucht Lange AG (Bad Schwartau, Germany). Plants were maintained in tissue culture with a 16-h light, 8-h dark regime on MS medium (Murashige and Skoog, 1962), which contained 2% sucrose. All plants were grown in the greenhouse under the same light regime with a minimum of 250 μmol photons m^{-2} s^{-1} at 22 °C. *Solanum lycopersicum* L. cv Moneymaker was obtained from Meyer Beck (Berlin, Germany), *Solanum pennellii* (accession LA0716) and the *Aco*1 mutant were obtained from the true breeding monogenic stocks maintained by the Tomato Genetic Stock Centre (University of California, Davies). Plants were handled as described in the literature (Carrari *et al*., 2003).

Metabolite analysis

Tissue samples were rapidly frozen in liquid nitrogen. The relative levels of metabolites were subsequently determined as described in Roessner *et al*. (2001a), with the modifications for tomato tissue documented in Roessner-Tunali *et al*. (2003). Data are presented normalized to wild type as described by Roessner *et al*. (2001b).

RNA isolation

Total RNA from potato leaves was isolated using Trizol (BioRad, München). RNA was hybridized against either our custom macroarray (Urbanczyk-Wochniak *et al*., 2003) or glass slide microarrays (Urbanczyk-Wochniak *et al*., 2005).

Custom macroarray library selection

The generation of our custom *Solanaceous* macroarray based largely on the EST library described by Van der Hoeven *et al*. (2002) has been described previously (Urbanczyk-Wochniak *et al*., 2003), for the complete list of selected clones see the minimal information about microarray experiment (MIAME) protocol http://www.mpimp-golm.mpg.de/fernie). Conditions for data acquisition and analysis were described previously (Urbanczyk-Wochniak *et al*., 2005), results coming from the macroarray have been published previously in this publication.

Glass slide microarray

Glass slides containing arrayed tomato ESTs were obtained directly from The Center for Gene Expression Profiling (CGEP) at the Boyce Thompson Institute (BTI), Cornell University, The Geneva Agricultural Experiment Station, and The USDA Federal Plant and Nutrition Laboratory. The tomato array contains 13,440 spots randomly selected from cDNA libraries isolated from a range of tissues including leaf, root, fruit, and flowers and representing a broad range of metabolic and developmental processes. Technical details of the spotting are provided as MIAME http://www.mpimp-golm.mpg.de/fernie. Further annotation of this file was carried out to provide gene identities and putative functions for the ESTs described on the Solanaceae Genomics Network (http://soldb.cit.cornell.edu/) website. Fluorescent probe preparation and microarray hybridization was exactly as described previously (Urbanczyk-Wochniak *et al*., 2005). Detailed information is included into MIAME http://www.mpimp-golm.mpg.de/fernie. The data for these experiments is presented for the first time in the current paper.

Transcript Scavenger

All work was carried out on a state-of-the-art laptop (Fujitsu-Siemens S series lifebook) or a comparable PC with 1 GB RAM, running Microsoft windows or Linux. Combinations of gene entries, assignments as well as translations were made with the standard Microsoft Office package (Redmond, WA) and perl (www.perl.org). Sequence similarity searches were conducted with the standalone NCBI-Blast version 2.2.10. Tentative ortholog information was gathered using Inparanoid (Remm *et al*., 2001) with standard settings. As a basis the TAIR *Arabidopsis* proteome dataset (release TIGR5) and the predicted

unigene peptide set from SGN (release #3) was used.

Origins of BINS and subBIN

The 34 MapMan BINs currently used for the *Arabidopsis* MapMan classification (Thimm *et al.*, 2004; Usadel *et al.*, 2005), were also used for tomato. As in *Arabidopsis* each BIN has a numerical code ranging from 1 to 35. These BINs can be extended in a hierarchical manner into subBINs. Therefore like in the first MapMan release for *Arabidopsis* BINs representing metabolism are represented by the codes 1–14, 16–19 as well as 22–26. However, some of these BINs have been broken down to individual enzymes, to facilitate the view of individual pathways.

Transfer of tomato unigenes to the MapMan BIN system

In order to transfer the MapMan BIN system to the tomato unigenes represented on the TOM1 array, the UniGene Set release 3 was downloaded from the SGN ftp site as a flat file in FASTA format (downloaded 7. Feb, 2005). From the FASTA file both annotations and sequences were extracted using a simple perl script. The annotations which were available consisted of the description of an *Arabidopsis* BLAST hit as well as an nr BLAST hit with their corresponding e-values. The descriptions were truncated by removing the hit to the *Arabidopsis* database as well as most accession numbers and the e-values, to increase legibility of corresponding entries. Subsequently, all unigenes represented on the TOM1 chip were determined by using the SGN bulk download tool (http://soldb.cit.cornell.edu/cgi-bin/bulk/input.pl?mode=microarray) with all available spot identifiers of the TOM1 chip (downloaded 9. Feb., 2005). The corresponding unigene sequences were used in an exhaustive BLASTx search (Altschul *et al.*, 1990) against the *Arabidopsis* proteome, release TIGR 5, downloaded from the TAIR website (ftp.tair.org, ATH1_pep_cm_20040228). An e-value of 10^{-10} was chosen as an initial cut-off. Using this BLAST search results, every tomato unigene could be either assigned to one best matching *Arabidopsis* gene, or to no gene at all.

If a corresponding *Arabidopsis* gene was thus found, and if the best hit from the nr database found by SGN, was the same *Arabidopsis* gene, this was indicative of the *Arabidopsis* gene being the closest known relative to the tomato gene. Therefore, in such a case the BIN class from *Arabidopsis* was imported and not checked against its automatic SGN annotation. Furthermore, predicted putative unigene peptide sequences from the SGN resource were used to identify *Arabidopsis* orthologs and "in-paralogs" using the Inparanoid software (Remm *et al.*, 2001). Afterwards, the assignments to MapMan BINS, were checked manually, using the SGN description (September 2005), the Interpro domain information from SGN as well as the ortholog information output from Inparanoid . (see Table 1 of the supplemental material, giving BIN Code, BIN Name as well as putative orthologs and Interpro domains).

This resulted in binning of 7513 unigenes into MapMan BINS. For an overview of this process see supplementary Figure 1, the BINing including additional information is reproduced in supplementary table 2. Moreover, a MapMan mapping file based on unigenes is available as well as a perl tool (upon request) to average spots from the TOM1 array belonging to the same unigene.

Construction of a MapMan hierarchy for the TOM1 chip

To construct a MapMan hierarchy for the TOM1 chip, the classified unigenes were first connected to the Spot-Identifiers present on the TOM1 chip. The Spot Identifiers on the TOM1 chip are four numbers separated by dots, which indicate grid-row, grid column, spot-row and spot-column (e.g. 3.4.5.6), as is the case for many cDNA microarrays. For the construction of the final MapMan classification the spots on the chip were used as a primary identifier. To obtain their corresponding UniGene release #3 identifiers, the Bulk download tool at SGN was used. In brief, all spots represented on the TOM1 chip were pasted into this tool and the result was downloaded. In cases where a sequence attached to a particular array spot was clearly not based on an INRA sequence read, this was labelled with "jjj".

Finally the resulting assignment was compared to that of the downloadable annotations for the TOM1 chip (ted.bti.cornell.edu; file: Genlist.xls,

representing a tomato release #2 UniGene set, downloaded 7. February 2005), to obtain information for spots where no UniGene release #3 was available using the array spot identifier. In cases, where a clone mentioned in #2 UniGene set was not identified with the SGN array spot bulk download tool, the clones mentioned in the table were pasted into the SGN clone bulk download tool to obtain a corresponding unigene from the latest set. Subsequently, these were labelled by "OOO" for old annotation. In the few cases where the clones mentioned in #2 UniGene set could not be obtained using the bulk download tool, the clones were used to query the SGN site individually. These cases usually represented DNA of very low quality. A unigene was then inferred by using the BLAST facility of the SGN site. The best matching UniGene release #3 was taken, provided that there was a significant similarity. These spots were then labelled by "HHH" to mark them as spots of lowest quality. All spots which had a different unigene assigned to their five and three prime reads were flagged with "flag_XH" to mark potential chimera and/or cross-hybridization problems and both the five and the three prime unigene descriptions were incorporated into the MapMan identifier description (for an overview see supplementary figure. 2).

Comparison of day-night experiments

For a comparison of different organisms' response to the diurnal cycle, the $\log_2$ transformed expression values for each transcript belonging to a series were normalized to the average expression of that series. Subsequently, for each experiment the average expression within each major BIN was calculated. For a comparison the results were loaded into the R statistical environment (R development core team, 2005) and a principal component analysis was performed on the average BIN values of all BINs except BIN 35, which represents not classified and/or unknown transcripts.

Tomato Metabolites

All metabolites which had been classified in the *Arabidopsis* release have been imported for tomato, since essentially they should have a similar role in tomato. For a conversion of arbitrary metabolite names to the canonical metabolite names used in MapMan, a tool for name conversion is available at http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/tools/gmd_tools.html (Kopka *et al.*, 2005).

Statistical analysis

Microarray experiment slides were normalized with print tip loess and moving minimum background subtraction using the Bioconductor limma package framework (Gentleman *et al.*, 2004). Microarray slides were subsequently scale normalized, adjusting the log-ratios to have the same median absolute deviation across arrays (Yang *et al.*, 2002), (Smyth and Speed, 2003). Moderated *t*-statistics were used to detect any genes likely to be differentially expressing (Smyth, 2004). MapMan files were constructed from resulting analysis $\log_2$ fold change values, where any poor quality spots created during the experimental process were down-weighted. In the case of the comparison of the two tomato species *Solanum lycopersicon* and *S. penellii*, background correction was conducted using simple background subtraction, after down-weighing all spots, not having a detectable signal in penellii in all three microarrays. Moreover, only spots were averaged for display in MapMan, that had detectable signals in both channels over all arrays.

Results and discussion

MapMan Categorization

The MapMan ontology and software were developed for *Arabidopsis* ATH1 arrays (Thimm *et al.* 2004). Whereas the full genome sequence was available in *Arabidopsis*, gene models in tomato are mainly derived from unigene builds based on clustering ESTs, plus a few full length clones. Further, the *Arabidopsis* ATH1 array is based on oligomers of known composition, whereas the TOM1 is a cDNA based chip. The underlying cDNAs of the TOM1 chip have been sequenced twice, initially by the TIGR institute (www.tigr.org/tigr-scripts/tgi; Bartley and Ishida, 2002) and more recently by INRA-Toulouse (and deposited on SGN; www.sgn.cornell.edu; Mueller *et al.*, 2005a). The latter provided greater precision of the individual sequence reads.

To assign the 13,440 spots present on the TOM1 chip, which represent approximately 7500 unigenes, to MapMan BINS and subBINS, a combination of a blast search against *Arabidopsis*, protein domain and putative ortholog information were used (for details see Materials and Methods).

However, for 3958 spots, no sequence information could be derived from the SGN resource. In some cases, these spots may not have been resequenced, and in others the spots might represent sequences of such low quality that they have been removed from the database. The spots for which no information was available from the SGN bulk download resource are provided as a list in the supplemental material. This list is in close agreement with the results of the resequencing by the INRA institute, which resulted in high-quality sequence information for only 9172 clones (http://bti.cornell.edu/CGEP/array%20info.htm), and as such revealed that information was missing for over 3000 clones.

The remaining 9482 spots could be assigned to unigenes. However, in 2095 cases, individual spots the five and three prime reads of the corresponding clone showed similarity to different unigenes, resulting in assignment of the spot to two unigenes. This is in agreement with the earlier observation that, from 6,695 reads from both ends, >2000 could not be assigned to the same unigene (http://bti.cornell.edu/CGEP/array%20info.htm). These spots are flagged in the MapMan description line with the term flag_XH to make users aware of potential problems. These multiple assignments are a direct result, of the lack of complete sequence information of the tomato genome, which makes it necessary to rely on unigene builds that are inherently prone to error and building mistakes (Sorek and Safer, 2003). They also reflect a general problem of cDNA based chips because some cDNAs may represent chimeras or products of differential splicing (Kothapalli *et al.*, 2002).

Despite these limitations, the distribution of the spots on the tomato EST array between different MapMan BINS resembled the distribution for the full *Arabidopsis* genome (Figure 1). Some errors

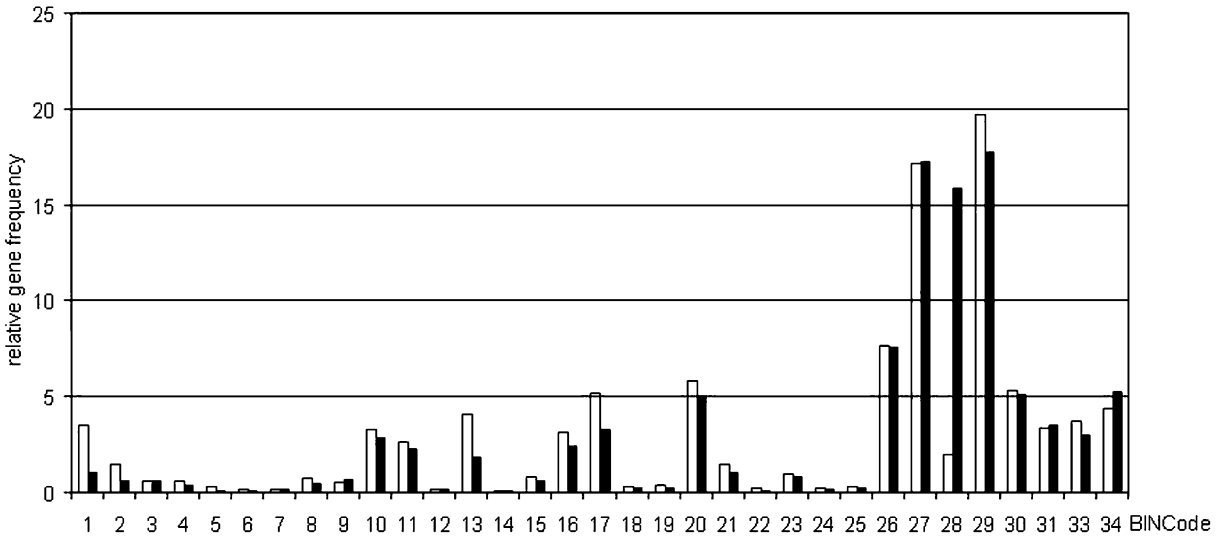

Figure 1. Comparison of the distribution of *A.thaliana* and *S. esculentum* genes in MapMan BINS. All genes of *A. thaliana* and all Unigenes represented on the tomato chip, were classified into MapMan BINS and for each BIN, the number of items was counted. Only major BINs are represented here. *A. thaliana* items are represented as solid bars, whereas *S. esculentum* values are represented as open bars. All values are scaled to 100% to facilitate better comparison, between the organisms. BIN 1, Photosynthesis; BIN 2, Major carbohydrates; BIN 3, Minor carbohydrates; BIN 4, Glycolysis; BIN 5, Fermentation; BIN 6, Gluconeogenesis/ glyoxylate cycle; BIN 7, Oxidative Pentose Phosphate pathway; BIN 8, TCA cycle/ organic acid transformations; BIN 9, Mitochondrial electron transport/ ATP synthesis; BIN 10, Cell wall; BIN 11, Lipid metabolism; BIN 12, Nitrogen assimilation; BIN 13, Amino acid metabolism; BIN 14, S-assimilation; BIN 15, Metal handling; BIN 16, Secondary metabolism; BIN17, Hormones; BIN 18, Cofactor and vitamin synthesis;, BIN 19, Tetrapyrrole synthesis; BIN 20, Stress; BIN 21, Redox; BIN 22, Polyamine synthesis; BIN 23, Nucleotide metabolism; BIN 24, Biodegradation of xenobiotics; BIN 25, C1-metabolism; BIN 26, Miscellaneous enzyme families; BIN 27, RNA; BIN 28, DNA;BIN 29, Protein; BIN 30, Signalling; BIN 31, Cell; BIN 33, Development; BIN 34, Transport; BIN 35, Not assigned.

may be masked, for example, if the clones representing two unigenes fall into the same MapMan subBIN, they will only be represented once. The relatively course structure of the MapMan BINS is in this sense advantageous, because it decreases the number of multiple assignment. This will also decrease the number of false assignment for genes with a similar sequence. In total there were 1248 multiple assignments (this corresponds to approximately 10% of the clones), due in part to potential chimeric cDNA clones and/or unigenes, but also to multiple roles of proteins in different biological processes. For example many genes that act as transcription factors (BIN 27.3) also have a role in development (BIN 33). Such multiple assignments are also present in the original *Arabidopsis* build.

Comparison of the distribution of the genes between different BINS in the *Arabidopsis* genome and the clones on the tomato TOM1 array reveals a similar overall distribution of genes (Figure 1). As shown in Figure 1, of the spots that had been categorized into a meaningful category (i.e. all categories except BIN 35, which represents unknown or unclassified proteins) most items either fell into the major BINS 27 (RNA) or 29 (protein). This is not surprising, given the high number of transcription factors found in plants (*see for example* Czechowski *et al.*, 2004) which are classified into BIN 27 and proteins playing a role in post-translational modification or protein degradation (*see for example* Nuhse *et al.*, 2004), represented in BIN 29. These BINS also contain a very large number of genes in *Arabidopsis*. One notable difference between *Arabidopsis* and tomato is the higher proportion of genes being classified as belonging to the DNA category in the former. This discrepancy might be explained by the high number of transposable elements in the *Arabidopsis* genome (Arabidopsis Genome Initiative, 2000), which are probably not highly represented on the tomato chip. Overall, the similar distribution of elements between different functional categories in the *Arabidopsis* ATH1 array and TOM1 array indicates that the latter is a reliable representation of the tomato genome. However, the final proof requires the availability of reliable predictions of all tomato genes and thus the completion of the sequencing of the tomato genome (Mueller *et al.*, 2005b).

The gene assignments must be viewed as tentative, because they were made based on homology to sequences from other organisms and/or domain information, instead of genuine information about tomato proteins. In the latest UniProtKB/Swiss-Prot relase 48, only 270 tomato entries exist (http://www.expasy.org/sprot/ppap/ppap_stat.html; Schneider *et al.*, 2005). The *Solanaceous* MapMan will need to be constantly updated, even prior to the availability of full genome data. This will facilitate the analysis of other platforms based on the EST collection described by van der Hoeven *et al.* (2002), as they become available. Also, as for the first build of the original *Arabidopsis* version (Thimm *et al.*, 2004), it is probable that several deficits exist in the first build of the *Solanaceous* MapMan. These are due to the current quality of gene annotation, inaccuracies in BIN assignments and crude substructuring of many areas of function. We hope that, as for the *Arabidopsis* MapMan, this can be addressed by mobilizing expert input. This could be especially valuable for biological processes for which tomato and its close relatives in the *Solanaceae* family serve as important model systems, for example hormonal control of fruit development (Fray and Grierson, 1993; Willkinson *et al.*, 1995; Zegzouti *et al.*, 1999; Tieman *et al.*, 2000), or processes underlying tuberisation, dormancy and sprouting in the potato (Fernie and Willmitzer, 2001; Sonnewald, 2001). Moreover, in the case of polyamine synthesis, where *Arabidopsis* seems to be one of very few organisms, which has lost ornithine decarboxylase (Hanfrey *et al.*, 2001), the consensus pathway could be completed and was incorporated into MapMan as a additional pathway.

The TOM1 MapMan ontology allows the visualisation of results obtained using this array using the MapMan software and diagrams ('Maps'), which can be downloaded at http://gabi.rzpd.de/projects/MapMan/. It should be pointed out that one of the features of MapMan is that the user can create new diagrams of metabolic or cellular processes as a template to display their data. It is hoped that such additional 'maps' will be made available to the community via this web site. To illustrate the use of MapMan in visualising and interpreting data, three case studies with the TOM1 array will now be presented.

Evaluation of changes in transcript level between the middle of the day and middle of the night in potato

Urbanczyk-Wochniak *et al.* (2005) showed that the TOM1 microarray can be used for transcript profiling in potato, and used it to analyse changes in transcript levels between leaf samples from wild type potato harvested in the middle of the day and the middle of the night. The quality resulting from hybridising potato on tomato ESTs is considerable and such experiments, including quantitative PCR validation, have been published before (see Urbanczyk-Wochniak *et al.*, 2003, 2005) whilst hybridisations in more divergent *Solanaceous* species such as pepper and eggplant have also recently been carried out (see Moore *et al.* 2005). Point-by-point analysis of our own previous experiments allowed several major trends to be identified, such as the downregulation of genes involved in the light reactions of photosynthesis and the increase in the transcript levels of many genes associated with respiration at night (Urbanczyk-Wochniak *et al.*, 2005). This original interpretation missed a surprisingly large number of responses, which are readily observable in the metabolism overview map in MapMan (Figure 2A). For this, and all other examples described here, changes are initially described at the BIN level since this allows description of coordinated changes before zooming in on the most variable individual gene by gene changes. A critical assessment of the functional importance of the various transcriptional changes was not used as a criteria for their description. However, in the limited number of instances where such assessments have been recorded in the literature we have documented them within the description of our results. The changes that were not fully described in our initial interpretation, of the potato diurnal experiment, include a decrease in the night of transcripts for genes involved in chlorophyll synthesis, and an increase of transcripts for many of the genes involved in the Calvin cycle, carbonic anhydrases, nucleotide metabolism, flavanoid and terpene synthesis.as well as amino acid biosynthesis (most notably phosphoserine amino transferase and histidine phosphate transaminase were upregulated whereas cysteine synthase was strongly downregulated).

A recently added feature of MapMan (Usadel *et al.*, 2005) allows the responses of the genes in a particular BIN, to be compared with the overall response of all the genes on the microarray. This statistical analysis reveals whether the population of genes in the BIN behave in a significantly different manner to the response of all of the genes on the array. The BINS can be ranked, according to the significance of the response of the genes assigned to them. This tool was applied to the comparison of night and day potato leaf samples. The photosynthesis BINs showed the most significant change, followed by starch degradation, amino acid synthesis and finally nucleotide metabolism (see Supplementary data for details). This provides a more rigorous analysis than the comparison of single spots, or the inspection of lists.

Visualisation in MapMan also facilitated the detection of changes in the transcript levels of genes that are not associated with metabolism, for example a large number of regulatory genes were found to exhibit altered transcript levels (Figure 2B). The changes seen in the transcript levels are largely what would be predicted with differences between the day and night emerging for many transcription factors, and some genes involved in protein modification and degradation, hormone metabolism or sensing, and light and redox associated genes. While we cannot exclude the possibility that our failure to identify some of these changes in the original paper was due to our own poor interpretation of the EST identity, the re-interpretation of this experiment in Figure 2 exemplifies the utility of MapMan as an aid to interpreting data from the TOM1 array.

Comparison of diurnal changes in potato and Arabidopsis *experiments*

The standardized ontology and display mode provided by MapMan facilitates cross species comparisons. This is illustrated in Figure 1, which compares the changes in global gene expression in potato leaves between the light and dark with the diurnal changes of gene expression in *Arabidopsis*. We reasoned that even though potato and *Arabidopsis* are very different plants and the transcripts are being measured on different platforms, common themes may emerge.

Ideally, such a comparison should be done on a gene-by gene basis, but this requires that true orthologs are known between the two species. While initiatives exist that try to identify orthologs

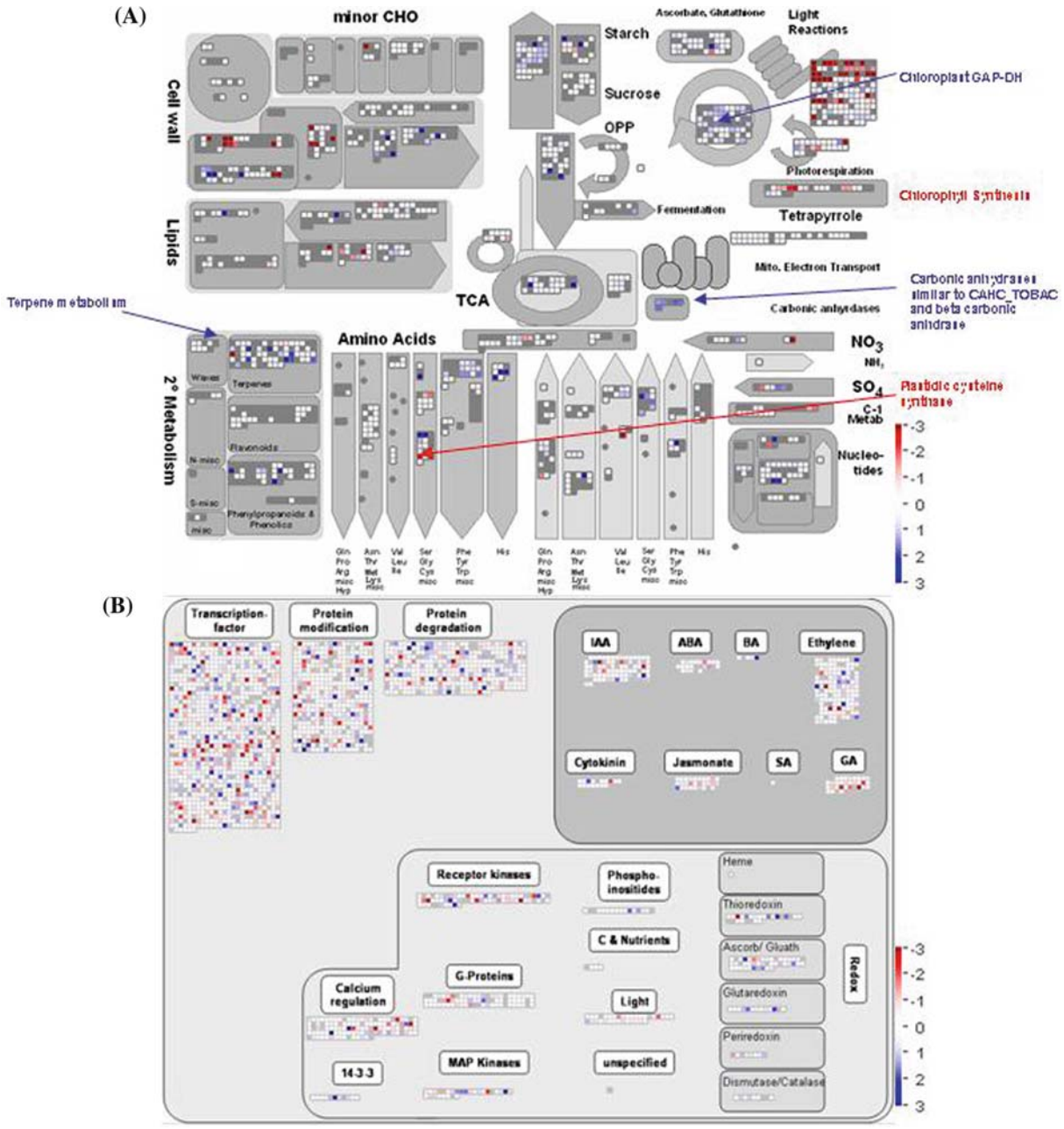

Figure 2. Changes in transcript levels in *S. tuberosum* leaves between the middle of the day and the middle of the night for genes associated with (A) metabolism, and (B) regulation. Experimental data from Urbanczyk-Wochniak *et al.* (2005). Red and blue represent a decrease and an increase of expression respectively, relative to the middle of the day. The color scale which was used is reproduced in the figure. This figure is best viewed, and all data point annotations provided at http://gabi.rzpd.de/projects/MapMan/ (see Materials and Methods). (C). Principal component analysis of *A. thaliana* and *S. tuberosum* diurnal cycle data. A principal component analysis was performed on the average, mean centred response of each of the major annotated BINS for the diurnal set of *A. thaliana* (Bläsing *et al.*, 2005, unpublished data) and *S. tuberosum* data for the middle of the day and middle of the night (Urbanczyk-Wochniak *et al.*, 2005). Day Points are displayed in grey and night time points are displayed in black. The percentage of variance explained by each component is shown in parenthesis.

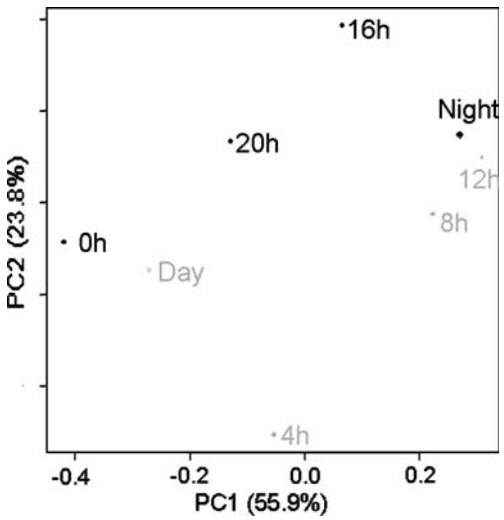

Figure 2. continued.

(O'Brien *et al.*, 2005) and the underlying algorithms have been considered in our manual curation of the BIN assignments, these ultimately rely on full genome knowledge which is currently lacking for *Solanaceous* species. Moreover, comparison of a non-full genome arrays like the ATH1 *Arabidopsis* chip (providing approximately 3/4 gene coverage) and the TOM1 (providing approximately 1/3 gene coverage) tomato chip is further hindered because the real orthologs might not be present on one or the other of the arrays. Indeed, even when full genome sequences are available, there is frequently insufficient experimental evidence to allow clear identification of orthologs, especially for members of the many large gene families encountered e.g. in *Arabidopsis* (Cannon *et al.*, 2004).

An alternative interim approach is to look for conserved or differing responses in sets of genes representing different functional classes. Transcript levels have been analysed at six times during the diurnal cycle in *Arabidopsis*, using the ATH1 array (Bläsing *et al.*, 2005). The response of all of the individual probe sets assigned to a given MapMan BIN were averaged, to obtain an "average BIN response". The signals for the entire TOM1 array clones assigned to each BIN (Urbanczyk-Wochniak *et al.*, 2005) were also averaged. Principal component (PCA) analysis of the average BIN signals (for this analysis BIN 35 (unknown) was excluded) reveals that a sample harvested during daytime from potato leaves lies closer on a PCA plot to *Arabidopsis* samples harvested during daytime (Figure 2C), and that a sample harvested during night-time from potato lies closer to *Arabidopsis* samples harvested during night. This example illustrates the potential of extending the MapMan ontology into crop species for comparative genomics, even though the low level of replication in the case of the potato experiment and the limited reliability of a PCA alone suggest that a far broader experimental data set would be required to statistically validate this cross-species comparison for individual genes. Moreover, the pending genome sequence of tomato will facilitate a higher resolution comparison of these changes, and in the interim it is important that non-comprehensive transcript profiling in the *Solanaceae* is supported by physiological or biochemical phenotyping (Urbanczyk-Wochniak *et al.*, 2005).

Evaluation of differential transcript level accumulation in S. lycopersicum *and its close wild species relative* S. pennellii

Recently we performed GC-MS based metabolite profiling of leaves and fruits of species of the *Solanum lycopersicum* complex (the wild species' of tomato that can be readily crossed with the elite species *S. lycopersicum*; Schauer *et al.*, 2005). The rationale for choosing *S. pennellii* was two-fold: (i) it is the background of the *Aco*1 mutant (Tanksley *et al.*, 1992, see also below), and also (ii) the donor species of what are, at least to date, the best characterized tomato introgression lines (Eshed and Zamir, 1995; Fridman *et al.*, 2004). This analysis revealed that the crop plant contained lower levels of many secondary metabolites, with particularly large decreases in chlorogenate and nicotinate.

Figure 3 visualizes TOM1 array data comparing transcript levels in the leaves, from six week old plants, 2 h into the dark period, of *S. pennellii* and *S. lycopersicum*. There were relatively few changes and, many of those that did occur can be explained by what we know about the selective pressures put on tomato during domestication, and match the previously reported changes of metabolites. *S. lycopersicum* has lower levels of transcript for genes involved in the metabolism of secondary metabolites, in particular terpenes,

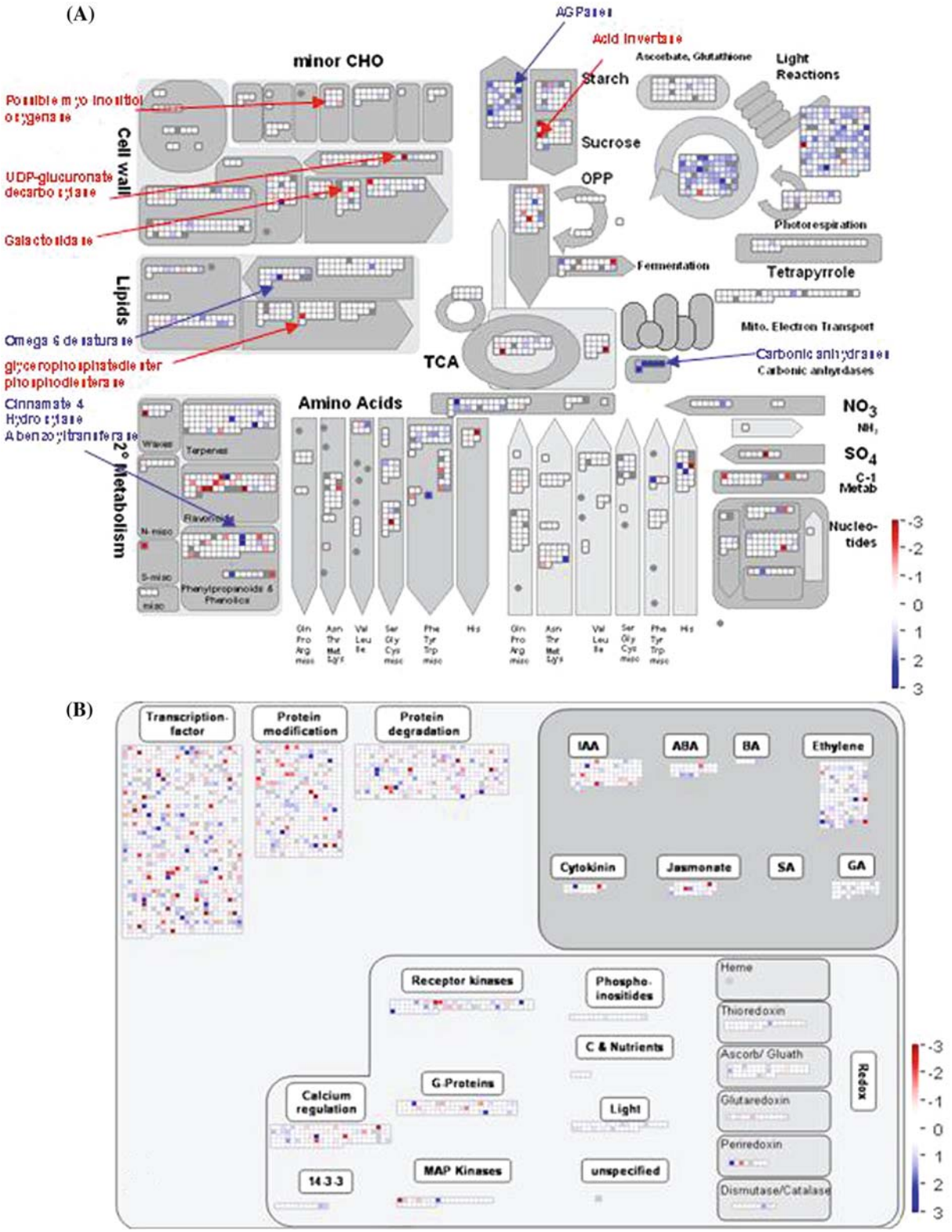
(A)
minor CHO
AGPase1
Acid Invertase
Ascorbate, Glutathione
Light Reactions
Starch
Sucrose
OPP
Photorespiration
Fermentation
Tetrapyrrole
Mito. Electron Transport
Carbonic anhydrase1
Carbonic anhydrases
TCA
Amino Acids
NO3
NH4
SO4
C-1 Metab
Nucleo-tides
Cell wall
Lipids
2° Metabolism
Terpenes
Flavonoids
Phenylpropanoids & Phenolics
Possible myo inositol oxygenase
UDP-glucuronate decarboxylase
Galactosidase
Omega 6 desaturase
glycerophosphodiester phosphodiesterase
Cinnamate 4 Hydroxylase
Abenzoyltransferase
-3
-2
-1
0
1
2
3
(B)
Transcription-factor
Protein modification
Protein degradation
IAA
ABA
BA
Ethylene
Cytokinin
Jasmonate
SA
GA
Receptor kinases
Phospho-inositides
C & Nutrients
G-Proteins
Light
MAP Kinases
unspecified
Calcium regulation
14.3.3
Heme
Thioredoxin
Ascorb/ Gluath
Glutaredoxin
Peniredoxin
Dismutase/Catalase
Redox
-3
-2
-1
0
1
2
3

Figure 3. Differences in transcript levels between leaves of *S. lycopersicum* and *S. pennellii* for genes associated with metabolism (a) and regulation (b). Both sets of material were harvested from plants grown under comprise growth conditions (*see* Schauer *et al.*, 2005 for details), in the middle of the day. Red and blue represent a decrease and an increase of expression respectively, in the *S. lycopersicum* relative to the *S. pennellii*. The colour scale which was used is reproduced in the figure. This figure is best viewed, and all data point annotations provided at http://gabi.rzpd.de/projects/MapMan/ (see Materials and methods).

phenylpropanoids and the chorismate pathway (Figure 3A). It is known that many secondary metabolites were selected against upon domestication (Tadmor *et al.*, 2002), a fact that is supported by lower levels of secondary metabolites in our earlier survey of the metabolite content of the wild species (see above). However, given the paucity of data available on the similarity of the respective genes, let alone the *in vivo* function of the proteins that they encode the full biological significance of these observations remains unclear from the current study.

There also seem to be some changes in the expression of some processes of carbohydrate metabolism. However, given that the carbohydrate levels were largely unaltered in the wild species (Schauer *et al.*, 2005), this would have to be confirmed by an independent approach. This comparison also reveals that many genes involved in regulatory processes have different transcript levels in the wild and elite crop species, including transcription factors, protein modification and degradation factors, phytohormones, and some receptor kinases and calcium regulation factors. Collection and curation of a broader number of microarray experiments in varying conditions, both within and between species, may yield more information about the functions of some of these genes. Moreover, an important caveat that should be noted is that we did not directly assess the strength of the *S. lycopersicon* vs *S. pennellii* hybridisations by an independent technique. However, we believe that the closeness of these species (in contrast to *S. lycopersicum* vs *S. tuberosum*) almost certainly implies that a large bias is not introduced due to hybridisation factors. Furthermore, as stated above the major conclusions we were able to draw here required analysis of the transcript data alongside that obtained from metabolite profiling (which in contrast to transcript profiling is essentially species-independent; Stitt and Fernie, 2003).

Evaluation of differential transcript level patterns in tomato deficient in the expression of enzymes of the mitochondrial TCA cycle

The third application, investigated the changes of transcription in two tomato genotypes that are deficient in two different enzymes of the mitochondrial TCA cycle. We have previously demonstrated photosynthesis and fruit yield is increased in a mutant line in the wild relative (*S. pennellii*) that has a lesion in the gene encoding *Aco*1, and in a transformant of the cultivated tomato (*S. lycopersicum*) that expresses the mitochondrial malate dehydrogenase gene in antisense orientation (Carrari *et al.*, 2003; Nunes-Nesi *et al.*, 2005). Figures 4 and 5 present the results of TOM1 microarray analyses of the changes in transcript levels in these two genotypes, compared to their corresponding parents.

When analysed at the BIN level, the *Aco*1 mutant was characterized by a large number of changes in transcripts for genes associated with the light reactions of photosynthesis, followed by changes for genes involved in cell wall synthesis, amino acid degradation, the TCA cycle and the Calvin cycle. A similar picture emerged for the malate dehydrogenase antisense line AL9, with the Calvin cycle and photosynthetic light reactions being the most strongly affected. AL9 displayed less marked changes of genes assigned to amino acid synthesis (aspartate family), isoprenoid metabolism and starch synthesis and degradation than the *Aco*1 mutant.

At the level of individual genes, in both genotypes the vast majority of genes involved in the photosynthetic light reactions had increased transcripts, as well as most genes for enzymes of the Calvin cycle, with the exception of chlorophyll A–B binding proteins. Both genotypes show a general trend towards decreased levels of transcripts for individual genes associated with photorespiration (most notably for the key photorespiratory enzymes serine hydroxyl methyl transferase and hydroxypyruvate reductase) and biosynthetic processes (most notably in cell wall metabolism).

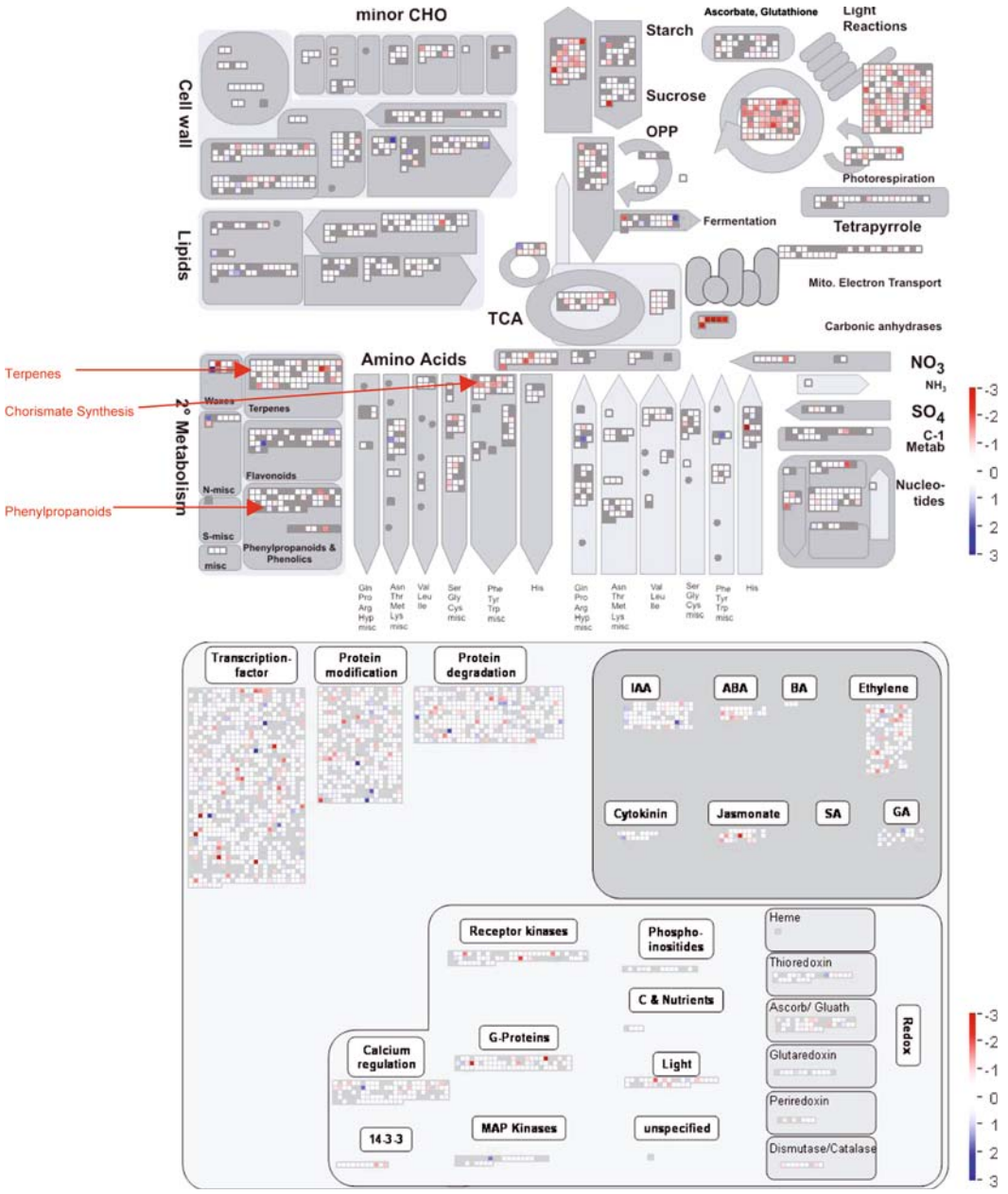

Figure 4. Differences in transcript levels between leaves of the *Aco*1 mutant of *S. pennellii* and its corresponding wild type accession for genes associated with metabolism (A) and regulation (B). Both sets of material were harvested from plants grown in the middle of the day. Red and blue represent a decrease and an increase of expression respectively, in the *Aco*1 mutant relative to its corresponding wild type accession. The colour scale which was used is reproduced in the figure. This figure is best viewed, and all data point annotations provided at http://gabi.rzpd.de/projects/MapMan/ (see Materials and methods).

Some marked differences emerged between the genotypes when changes in transcript levels were analysed on a gene by gene basis. Whereas the *Aco*1 mutant displayed relatively minor and mixed changes in transcripts associated with glycolysis and the TCA cycle, the malate dehydrogenase antisense line displayed a concerted decrease of transcripts for key regulatory enzymes of carbohydrate oxidation, starch synthesis and breakdown and fermentative metabolism (Fernie *et al.*, 2004;

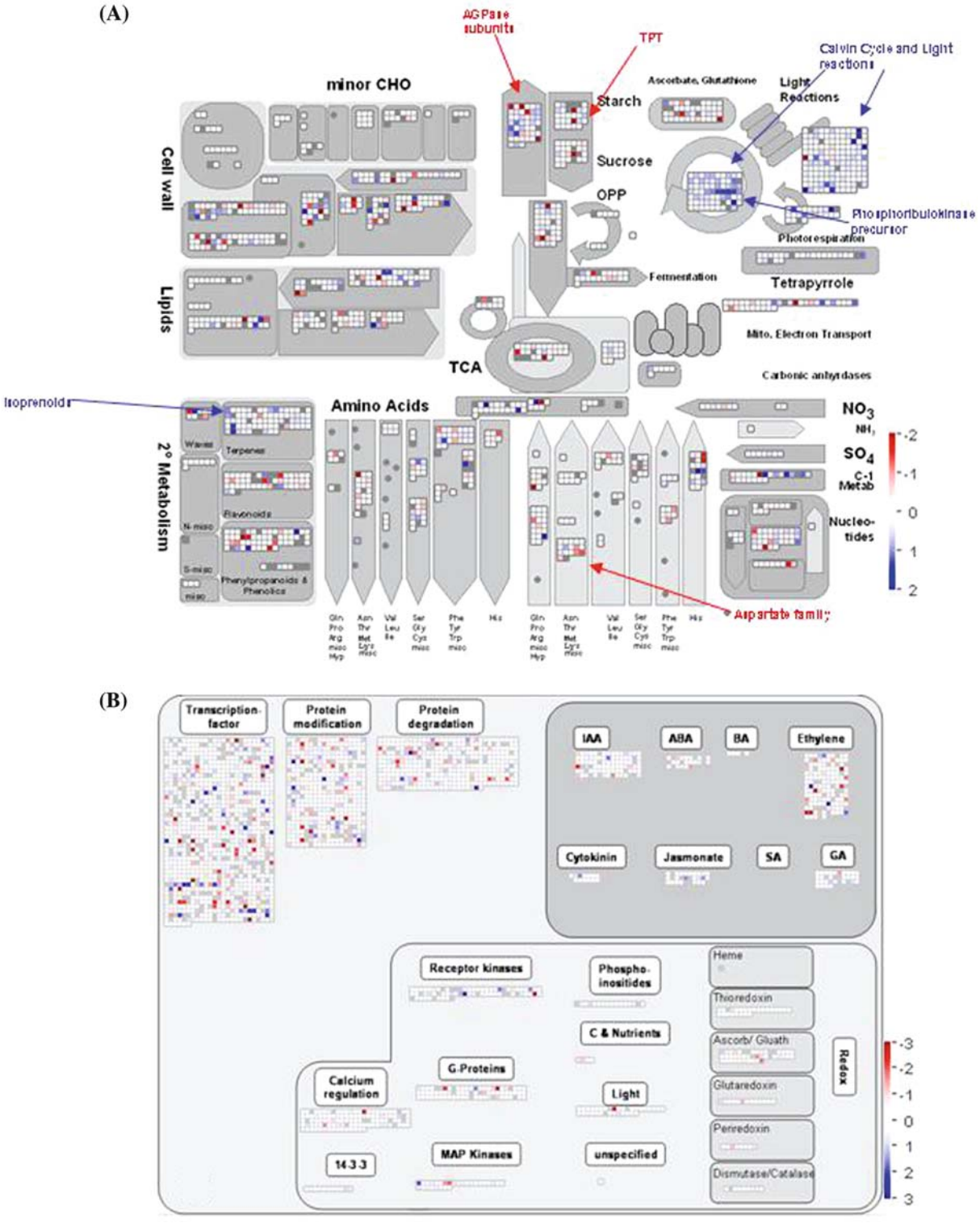
(A)
AGPase subunits
TPT
Calvin Cycle and Light reactions
minor CHO
Ascorbate, Glutathione
Light Reactions
Starch
Sucrose
OPP
Phosphoribulokinase precursor
Photorespiration
Cell wall
Lipids
Fermentation
Tetrapyrrole
Mito. Electron Transport
TCA
Carbonic anhydrases
Isoprenoids
Amino Acids
NO3
NH4
SO4
C-1 Metab
Nucleo-tides
2° Metabolism
Waxes
Terpenes
Flavonoids
N-misc
S-misc
misc
Phenylpropanoids & Phenolics
Aspartate family
(B)
Transcription factor
Protein modification
Protein degradation
IAA
ABA
BA
Ethylene
Cytokinin
Jasmonate
SA
GA
Receptor kinases
Phospho-inositides
C & Nutrients
Calcium regulation
G-Proteins
Light
14-3-3
MAP Kinases
unspecified
Heme
Thioredoxin
Ascorb/ Gluath
Glutaredoxin
Perredoxin
Dismutase/Catalase
Redox

Figure 5. Differences in transcript levels between leaves of the well characterized mictochondrial malate dehydrogenase antisense line AL–9 (Nunes-Nesi *et al*., 2005) and its respective wild type control (*S. lycopersicum* cv. Moneymaker) for genes associated with metabolism (A) and regulation (B). Both sets of material were harvested from plants grown in the middle of the day. Red and blue represent a decrease and an increase of expression respectively, in the malate dehydrogenase antisense line relative to its corresponding wild type accession. The color scale which was used is reproduced in the figure. This figure is best viewed, and all data point annotations provided at http://gabi.rzpd.de/projects/MapMan/ (see Materials and methods).

Geigenberger *et al*., 2004; Gass *et al*., 2005): namely AGPase, glucose 6-phosphate dehydrogenase, β-amylase, the triose-phosphate transporter, pyrophosphate dependent phosphofructokinase, invertase, pyruvate kinase, alcohol dehydrogenase, the pyruvate dehydrogenase complex and succinate dehydrogenase. *Aco*1, but not in the malate dehydrogenase antisense line, showed a general trend of decreased transcript levels of genes associated with terpenes and phenylpropanoids although it should be noted that the transcript levels of some genes of these BINs increased and that further analysis of the type initiated by the Bouwmeester and Campbell research groups (Aharoni *et al*., 2003; Patzlaff *et al*., 2003), is required in order to assign functionality to those transcriptional changes. There was a shift in the pattern of genes associated with ascorbate and glutathione in *Aco*1, with some being induced and other repressed. There is, furthermore, a clear tendency to decreased levels of genes associated with biosynthesis of the aspartate family amino acids in the *Aco*1 mutant, but not the malate dehydrogenase antisense line, which is consistent with the minor reduction in the steady state levels of these metabolites reported previously for the former (Nunes-Nesi *et al*., 2005).

The changes of transcripts for genes associated with regulatory processes are shown in Figures 4 and 5B. Perhaps surprisingly, different genes are induced in the *Aco*1 mutant and in the malate dehydrogenase antisense line AL9. Whilst on one hand this could be interpreted to suggest that these genes do not play an important role in establishing the phenotype observed following inhibition of the mitochondrial TCA cycle, it should be noted that these phenotypes are by no means identical to one another and that changes in the levels of these regulatory transcripts for example those associated with redox may account for at least some of the differences between the genotypes.

To summarize, comparison of the transcription changes observed following modulation in the expression of two TCA cycle enzymes revealed some commonality – as would be anticipated from the fact that at a gross phenotypic level the plants were similar (Nunes-Nesi *et al*., 2005). This is most apparent when the transcript changes were analysed at the BIN level since the most significantly altered BINs were largely conserved. One exception to this was the fact that genes associated with the Calvin cycle were more dramatically altered in the mitochondrial malate dehydrogenase plants than in the *Aco*1 mutant.

A detailed GC-MS dataset showing changes of metabolites in the mMDH antisense plants was published in Nunes-Nesi *et al*. (2005). We carried out a more detailed evaluation of the chromatograms for *Aco*1 samples than performed previously (Carrari *et al*., 2003; Figure 6), to obtain two comparable datasets. Interestingly, similar changes were found in both genotypes, notably the increases in dehydroascorbate and changes in organic acid content, although the relative changes were much smaller in the *Aco*1 mutant. This finding is consistent with previously proposed models for L-galactono-lactone as an alternative respiratory substrate for the cytochrome pathway and hence a tight coupling of the rates of respiration and ascorbate production (Millar *et al*., 2003; Nunes-Nesi *et al*., 2005). Whilst ascorbate has previously been indicated to have large effects on the expression of genes associated with photosynthesis (Smirnoff, 2000; Kiddle *et al*., 2003), the results presented here provide a strong suggestion that the effect may be stronger on genes associated with the Calvin cycle than on genes associated with the photosynthetic light reactions. Intriguingly, the increase in the photosynthetic rate was greater in the *Aco*1 mutant than that found in the malate dehydrogenase antisense plants suggesting that factors other than those mediated by ascorbate also play a role in promoting photosynthesis. Previous experiments, from our own and other groups have however shown that ascorbate is capable of mediating effects at various levels of control (Smirnoff and Wheeler, 2000; Danna *et al*., 2003; Chen and Gallie, 2004; Nunes-Nesi *et al*., 2005).

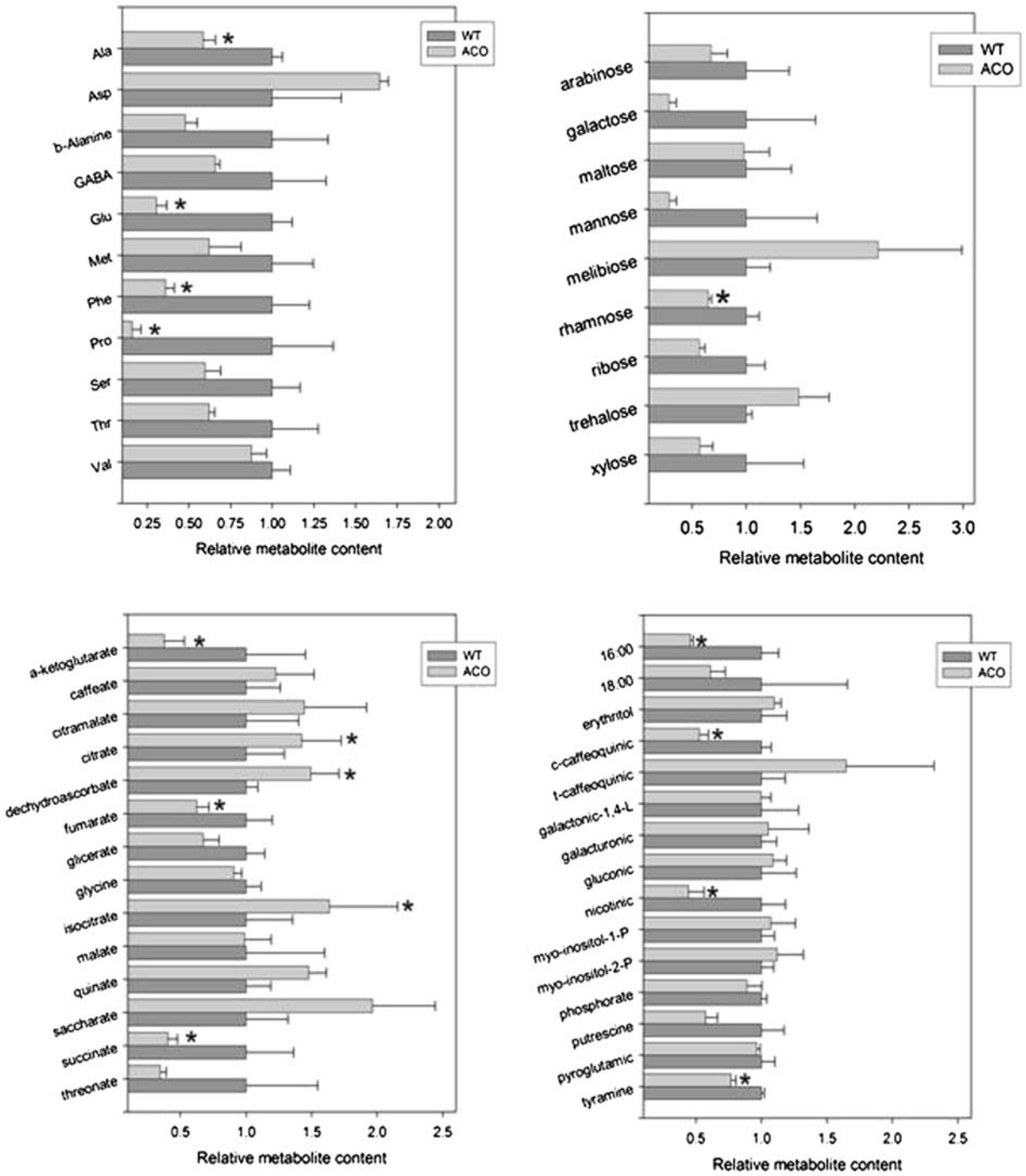

Figure 6. Relative metabolite content of the *Aco*1 mutant. Metabolites were determined as described in Materials and methods. Data are normalized with respect to the mean response calculated for the wild type. Values are presented as mean ± SE of determinations of six individual plants per line; asterisk indicates values that were determined by the *t*-test to be significantly different ($P < 0.05$) from the wild type.

Conclusion

The MapMan ontology and software provide a visual overview of the responses of genes involved in large parts of plant metabolism and/or other processes such as development and stress responses, and allow facile comparison of experiments within the same species, and between species. Using the MapMan ontology and the accompanying MapMan software tool, it was possible both to uncover interesting trends in transcript profiles of previously documented TOM1 array experiments that assessed transcript and metabolite changes during a diurnal cycle in potato leaves (Urbanczyk-Wochniak *et al.*, 2005) and two previously unpublished studies that compare the transcriptional program of leaves from two different tomato species, and the changes of transcript levels in two tomato genotypes that are deficient for two enzymes of the mitochondrial TCA cycle (Fernie *et al.*, 2004). Among the biological conclusions we were able to draw by using this software to aid interpretation of the datasets obtained from the experiments described above were (i) the clear diurnal shifts in transcripts associated with phenylpropanoid and nucleotide metabolism nucleotide metabolism and cell wall metabolism in the potato leaf, (ii) the observation that the levels of transcripts associated with secondary metabolism were generally considerably lower in the elite cultivated tomato species than in its wild species relative *S. pennellii* and (iii) the observation that changes in transcript levels in tomato plants displaying reduced expression of TCA cycle enzymes are consistent with the increased photosynthetic rate in these genotypes.

We anticipate that future developments of the MapMan software will aid further in interpretation and analysis of forthcoming *Solanaceous* microarray experiments. In order to improve the classification of the TOM1 chip, the MapMan hierarchy is currently being imported into a database to allow facile editing and sequence improvements for the structure (unpublished data). Within this database a flexible java user interface will be implemented that supports structuring, editing and correction as well as complementing of deposited data. It is anticipated that, as for *Arabidopsis* (Usadel *et al.*, 2005), the exhaustive annotation and classification of tomato proteins will require the effort of multiple laboratories and experts. Furthermore, all results represented here will also be transferable to future tomato chips since the classification presented here was performed on the latest unigene set represented on the chip.

Acknowledgements

We are grateful for the support of Lukas Mueller (Solanaceae Genomics Network, Cornell University, Ithaca, NY) during the process of tomato gene annotation. We are thankful to Helga Kulka for excellent care of the plants. Furthermore we thank Axel Nagel for excellent programming work. This work was supported by the German-Israeli Cooperativity project, DIP (to A.R.F and M.S.) by the Max-Planck-Gesellschaft (to E.U.-W., A.N.-N. and F.C.) by the BMBF GABI program Project GABI-MAPMAN (313112 and 313110) (B.U., S.M., M.S.), NZ Foundation for Research Science and Technology #CO6X0207 and ISAT #04-67 (to M.D.) and the Plant Science Group CROPNET (to M.K.) and by the PROFOOD project from the European Commission (M.S.).

References

Aharoni, A., Giri, A.P., Deuerlein, S., Griepink, F., de Kogel, W.J., Verstappen, F.W.A., Verhoeven, H.A., Jongsma, M.A., Schwab, W. and Bouwmeester, H.J. 2003. Terpenoid metabolism in wild type and transgenic Arabidopsis plants. Plant Cell 15: 2866–2884.

Alba, R., Fei, Z.J., Payton, P., Liu, Y., Moore, S.L., Debbie, P., Cohn, J., D′Ascenzo, M., Gordon, J.S., Rose, J.K.C., Martin, G., Tanksley, S.D., Bouzayen, M., Jahn, M.M. and Giovannoni, J. 2004. ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. Plant J. 39: 697–714.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–10.

Arabidopsis Genome Initiative 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

Bartley, G.E. and Ishida, B.K. 2002. Digital fruit ripening: data mining in the TIGR tomato gene index. Plant. Mol. Biol. Rep. 20: 115–130.

Baxter, C.J., Sabar, M., Quick, W.P. and Sweetlove, L.J. 2005. Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome wide

transcriptional changes and reveals links to mapped QTLs and described traits. J. Exp. Bot. 56: 1591–1604.

Bläsing, O.E., Gibon, Y., Günther, M., Höhne, M., Morcuende, R., Thimm, O., Usadel, B.,Scheible, W.-R. and Stitt, M. 2005. Sugars and Circadian Regulation make Major Contributions to the Global Regulation of Diurnal Gene Expression in Arabidopsis. Plant Cell 17: 3257–3281.

Carbone, F., Pizzichini, D., Giuliano, G., Rosati, C. and Perrotta, G. 2005. Comparative profiling of tomato fruits and leaves evidences a complex modulation of global transcript profiles. Plant Sci. 169: 165–175.

Carrari, F., Nunes-Nesi, A., Gibon, Y., Lytovchenko, A., Ehlers-Lourario, M. and Fernie, A.R. 2003. Reduced expression of aconitase results in an enhanced rate of photosynthesis and marked shifts in carbon partitioning in illuminated leaves of *Lycopersicon pennellii*. Plant Physiol. 133: 1322–1335.

Chen, Z. and Gallie, D.R. 2004. The ascorbic acid redox state controls guard cell signalling and stomatal movement. Plant Cell 16: 1143–1162.

Czechowski, T., Bari, R.P., Stitt, M., Scheible, W.R. and Udvardi, M.K. 2004. Real-time RT-PCR profiling of over 1,400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot- specific genes. Plant J. 38: 366–379.

Danna, C.H., Bartoli, C.G., Sacco, F., Ingala, L.R., Santa-Maria, G.E., Guiamet, J.J. and Ugalde, R.A. 2003. Thylakoid bound ascorbate peroxidase mutant exhibits impaired electron transport and photosynthetic activity. Plant Physiol. 133: 2116–2125.

Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D. and May, G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol. 4: Article no. 10.

Eshed, Y. and Zamir, D. 1995. An introgression line population of *Lycopersicum pennellii* in the cultivated tomato enables the identification and fine mapping of yield associated QTL. Genetics 141: 1147–1162.

Fei, Z., Tang, X., Alba, R.M., White, J.A., Ronning, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.J. 2004. Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. Plant J. 40: 47–59.

Fernie, A.R. and Willmitzer, L. 2001. Molecular and biochemical triggers of potato tuber development. Plant Physiol. 127: 1459–1465.

Fernie, A.R., Carrari, F. and Sweetlove, L.J. 2004. Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. Curr. Opin. Plant Biol. 7: 254–261.

Fray, R.G. and Grierson, D. 1993. Molecular genetics of tomato fruit ripening. Trends in Genetics 9: 438–443.

Fridman, E., Carrari, F., Liu, Y.S., Fernie, A.R. and Zamir, D. 2004. Zooming in on a quantitative trait for tomato yield using introspecific introgressions. Science 305: 1786–1789.

Gass, N., Glagotskaia, T., Mellema, S., Stuurman, J., Barone, M., Mandel, T., Roessner-Tunali, U. and Kuhlemeier, C. 2005. Pyruvate decarboxylase provides growing pollen tubes with a competetive advantage in petunia. Plant Cell 17: 2355–2368.

Geigenberger, P., Stitt, M. and Fernie, A.R. 2004. Metabolic control analysis and regulation of the conversion of sucrose to starch in growing potato tubers. Plant Cell Environ. 27: 655–673.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y. and Zhang, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5: Article no. 80.

Hanfrey, C., Sommer, S., Mayer, M.J., Burtin, D. and Michael, A.J. 2001. Arabidopsis polyamine biosynthesis: absence of ornithine decarboxylase and the mechanism of arginine decarboxylase activity. Plant J. 27: 551–560.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmuller, E., Dormann, P., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R. and Steinhauser, D. 2005. GMD@CSB.DB: The Golm Metabolome Database. Bioinformatics 21: 1635–1638.

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

Kothapalli, R., Yoder, S.J., Mane, S. and Loughran, T.P. 2002. Microarray results: how accurate are they? BMC Bioinformatics 3: Article no. 22.

Lemaire-Chamley, M., Petit, J., Garcia, V., Just, D., Baldet, P., Germain V., Fagard, M., Mouassite, M., Cheniclet, C. and Rothan, C. 2005. Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. Plant Physiol. 139: 750–769.

Millar, A.H., Mittova, V., Kiddle, G., Heazelwood, J.L., Bartoli, C.G., Thedoulou, F.L. and Foyer, C.H. 2003. Control of ascorbate synthesis by respiration and its implications for stress responses. Plant Physiol. 133: 443–447.

Moore, S., Payton, P., Wright, M., Tanksley, S. and Giovannoni, J 2005. Utilization of tomato microarrays for comparative gene expression analysis in the Solanaceae. J. Exp. Bot. 56: 2885–2895.

Mueller, L.A., Solow, T.H., Taylor, N., Skwarecki, B., Buels, R., Binns, J., Lin, C., Wright, M.H., Ahrens, R., Wang, Y., Herbst, E.V., Keyder, E.R., Menda, N., Zamir, D. and Tanksley, S.D. 2005a. The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol. 138: 1310–1317.

Mueller, L.A., Tanksley, S.D., Giovannini, J.J., van Eck, J., Stack, S., Choi, D., Kim, B.D., Chen, M.S., Cheng, Z.K., Li, C.Y., Ling, H.Q., Xue, Y.B., Seymour, G., Bishop, G., Bryan, G., Sharma, R., Khurana, J., Tyagi, A., Chattopadhyay, D., Singh, N.K., Stiekema, W., Lindhout, P., Jesse, T., Lankhorst, R.K., Bouzayen, M., Shibata, D., Tabata, S., Granell, A., Botella, M.A., Giullano, G., Frusciante, L., Causse, M. and Zamir, D 2005b. The tomato sequencing project, the first cornerstone of the International Solanaceae Project (SOL) Comp Funct. Genomics 6: 153–158.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A.N., Orchard, S., Pagni, M., Ponting, C.P., Quevillon, E., Selengut, J., Sigrist, C.J., Silventoinen, V., Studholme, D.J., Vaughan, R.

and Wu, C.H 2005. InterPro, progress and status in 2005. Nucleic Acids Res. 33: D201–5.

Murashige, T. and Skoog, F. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant 15: 473–497.

Mysore, K.S., Crasta, O.R., Tuori, R.P., Folkerts, O., Swirsky, P.B. and Martin, G.B. 2002. Comprehensive transcript profiling of Pto- and Prf-mediated host defense responses to infection by *Pseudomonas syringae* pv. Tomato. Plant J. 32: 299–315.

Nuhse, T.S., Stensballe, A., Jensen, O.N. and Peck, S.C. 2004. Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. Plant Cell 16: 2394–2405.

Nunes-Nesi, A., Carrari, F., Lytovchenko, A., Smith, A.M., Loureiro, M.E., Ratcliffe, R.G., Sweetlove, L.J. and Fernie, A.R. 2005. Enhanced photosynthetic performance and growth as a consequence of decreasing mitochondrial malate dehydrogenase activity in transgenic tomato plants. Plant Physiol. 137: 611–622.

O'Brien, K.P., Remm, M. and Sonnhammer, E.L. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33: 476–480.

Patzlaff, A., McInnis, S., Courtenay, A., Surman, C., Newman, L.J., Smith, C., Bevan, M.W., Mansfield, S., Whetten, R.W., Sederoff, R.R. and Campbell, M.M. 2003. Characterization of a pine MYB that regulates lignification. Plant J. 36: 743–754.

R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, http://www.R-project.org. Accessed 1.5.2005.

Remm, M., Storm, C.E. and Sonnhammer, E.L. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol 314: 1041–1052.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. and Fernie, A.R. 2001a. Metabolic profiling and phenotyping of genetically and environmentally modified systems. Plant Cell 13: 131–142.

Roessner, U., Willmitzer, L. and Fernie, A.R. 2001b. High-resolution metabolic profiling: identification of phenocopies. Plant Physiol. 127: 749–764.

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

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and Quackenbush, J. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374–378.

Schauer, N., Zamir, D. and Fernie, A.R. 2005. Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. J. Exp. Bot. 56: 297–307.

Schneider, M., Bairoch, A., Wu, C.H. and Apweiler, R. 2005. Plant protein annotation in the UniProt Knowledgebase. Plant Physiol. 138: 59–66.

Smirnoff, N. 2000. Ascorbate biosynthesis and function in photoprotection. Phil. Trans. R. Soc. Lond. B Biol. Sci. 355: 1455–1464.

Smirnoff, N. and Wheeler, G.L. 2000. Ascorbic acid in plants: biosynthesis and function. CRC Crit. Rev. Plant Sci. 19: 267–290.

Smyth, G.K. and Speed, T.P. 2003. Normalization of cDNA microarray data. In: METHODS: Selecting Candidate Genes from DNA Array Screens: Application to Neuroscience , D. Carter (Ed.). Methods Volume 31: 265–273.

Smyth, G.K. 2004. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Statistical Applications in Genetics and Molecular Biology 3, 3.

Sonnewald, U. 2001. Control of potato tuber sprouting. Trends in Plant Sci. 6: 333–335.

Sorek, R. and Safer, H.M. 2003. A novel algorithm for computational identification of contaminated EST libraries. Nucl. Acids Res. 31: 1067–1074.

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. Biotech. 14: 136–144.

Tadmor, Y., Fridman, E., Gur, A., Larkov, O., Lastochkin, E., Ravid, U., Zamir, D. and Lewinsohn, E. 2002. Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. J. Agri. Food Chem. 50: 2005–2009.

Tanksley, S.D., Ganal, M., Prince, J.P., de Vincente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, *et al.* 1992. High-density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141–1160.

The Gene Ontology Consortium 2000. Gene Ontology: tool for the unification of biology. Nature Genet 25: 25–29.

Thimm, O., Blässing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y. and Stitt, M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37: 914–939.

Tieman, D.V., Taylor, M.G., Ciardi, J.A. and Klee, H.J. 2000. The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. Proc. Natl. Acad. Sci. USA 97: 5663–5668.

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

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

Urbanczyk-Wochniak, E., Baxter, C., Kolbe, A., Kopka, J., Sweetlove, L.J. and Fernie, A.R. 2005. Profiling of diurnal patterns of metabolite and transcript abundance in potato (Solanum tuberosum) leaves. Planta 221: 891–903.

Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O.E., Palacios-Rojas, N., Selbig, J., Hannemann, J., Conceicao, P., Steinhauser, D., Scheible, W.R., Gibon, Y., Morcuede, R., Weicht, D., Meyer, S. and Stitt, M. 2005. Extension of the visualisation tool MapMan to allow statistical analysis of arrays, display of co-responding genes and comparison with known responses. Plant Physiol. 138: 1185–1194.

Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. and Tanksley, S.D. 2002. Deductions about the number,

organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell 14: 1441–1456.

Wilkinson, J.Q., Lanahan, M.B., Yen, H.C., Giovannoni, J.J. and Klee, H.J. 1995. An ethylene inducible component of signal transduction encoded by never ripe. Science 270: 1807–1809.

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

Zeeberg, B.R., Feng, W.M., Wang, G., Wang, M.D., Fojo, A.T., Sunshine, M., Narasimhan, S., Kane, D.W., Reinhold, W.C., Lababidi, S., Bussey, K.J., Riss, J., Barrett, J.C., Weinstein, J.N. 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol. 4: R28.

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