Max-Planck-Institut für molekulare Pflanzenphysiologie
Arbeitsgruppe Alisdair Fernie

Quantitative Trait Loci (QTL) for Metabolite Accumulation and Metabolic Regulation: Metabolite Profiling of Interspecific Crosses of Tomato

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Biochemie"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Nicolas Schauer

Potsdam, den 07.03.2006

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und unter Verwendung keiner anderen als den von mir angegebenen Quellen und Hilfsmitteln verfasst habe.
Ferner erkläre ich, dass ich bisher weder an der Universität Potsdam noch anderweitig versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

Nicolas Schauer
Potsdam, 07.03.2006

Table of Contents

1 Summary

The advent of large-scale and high-throughput technologies has recently caused a shift in focus in contemporary biology from decades of reductionism towards a more systemic view. Alongside the availability of genome sequences the exploration of organisms utilizing such approach should give rise to a more comprehensive understanding of complex systems. Domestication and intensive breeding of crop plants has led to a parallel narrowing of their genetic basis. The potential to improve crops by conventional breeding using elite cultivars is therefore rather limited and molecular technologies, such as marker assisted selection (MAS) are currently being exploited to re-introduce allelic variance from wild species. Molecular breeding strategies have mostly focused on the introduction of yield or resistance related traits to date. However given that medical research has highlighted the importance of crop compositional quality in the human diet this research field is rapidly becoming more important. Chemical composition of biological tissues can be efficiently assessed by metabolite profiling techniques, which allow the multivariate detection of metabolites of a given biological sample.

Here, a GC/MS metabolite profiling approach has been applied to investigate natural variation of tomatoes with respect to the chemical composition of their fruits. The establishment of a mass spectral and retention index (MSRI) library was a prerequisite for this work in order to establish a framework for the identification of metabolites from a complex mixture. As mass spectral and retention index information is highly important for the metabolomics community this library was made publicly available. Metabolite profiling of tomato wild species revealed large differences in the chemical composition, especially of amino and organic acids, as well as on the sugar composition and secondary metabolites. Intriguingly, the analysis of a set of *S. pennellii* introgression lines (IL) identified 889 quantitative trait loci of compositional quality and 326 yield-associated traits. These traits are characterized by increases/decreases not only of single metabolites but also of entire metabolic pathways, thus highlighting the potential of this approach in uncovering novel aspects of metabolic regulation. Finally the biosynthetic pathway of the phenylalanine-derived fruit volatiles phenylethanol and phenylacetaldehyde was elucidated via a combination of metabolic profiling of natural variation, stable isotope tracer experiments and reverse genetic experimentation.

2 Introduction

2.1 Crop Improvement through Access to Natural Variation

The improvement of crop species has been a human pursuit since cultivation began with great advances achieved both in crop yield and production traits. Coincident to these advances a narrowing of the genetic basis of most crops has occurred. Therefore, the potential to further improve crop traits by conventional breeding strategies has become somewhat limited and the developments of molecular techniques to introduce defined genes, or genomic regions, from more exotic sources are currently being evaluated. To date the majority of such studies have focussed on improving crop yield or resistance, however, given that medical research has highlighted the importance of crop compositional quality for human health this research field is of rapidly increasing significance(Demmig-Adams and Adams 2002; Rousseaux *et al*. 2005; Spencer *et al*. 2005).

2.1.1 Metabolic Engineering in Plants

In the last two decades an enormous research effort has been expanded on approaches for genetically engineering plants. These studies can largely be partitioned into three strategies, those intent on increasing (i) yield, (ii) resistance or (iii) the content of nutritionally or economically important metabolites. Whilst the vast majority of these approaches have failed to achieve their goal, there are many recent examples of successful transgenic improvements (Ye *et al*. 2000; Holmstrom *et al*. 2000; Bovy *et al*. 2002a; Geigenberger *et al*. 2005; Paine *et al*. 2005). For example the identification and transgenic modification of key genes in the fibre composition of *Arabidopsis* and tomato (for a review see Scheible and Pauly 2004), the overexpression of a citrate synthase in tobacco resulted in a 10-fold increase of citrate in root tips conferring enhanced aluminium tolerance (delaFuente *et al*. 1997), the antisense inhibition of threonine synthase yielded high-methionine potatoes (Zeh *et al*. 2001), whilst the antisense inhibition of a potato plastidial adenylate kinase resulted in high starch and amino acid contents (Regierer *et al*. 2002) and tomato plants were recently generated that exhibited fruit specific increases in carotenoid content (Davuluri *et al*. 2005). Further

examples are the overproduction of long-chain unsaturated fatty acids and omega-3-fatty acids in *Brassica* (Wu *et al*. 2005), as well as the production of an insecticidal protein derived from *Bacillus thurengiensis* in maize (Koziel *et al*. 1993). However, caution must be taken in evaluating the results of these studies since many of the transformants described have often only been tested under relatively stable environmental conditions (Nunes-Nesi *et al*. 2005; Van Camp 2005). Furthermore, at least in Europe, public concern about the use of GM technology is likely to minimize the chances of such manipulated crops reaching the market. That said it should not be forgotten that large areas of the world are growing high acreage of GM crops and the impact of transgenic agriculture is potentially considerable at the global level (Cohen 2005). An alternative source of genetic diversity that has long been exploited in agriculture is the use of mutations. An example of the use of mutants in agriculture is that of maize (Zamir 2001) which contains at least five mutations that distinguish it from its direct ancestor – teosinte (Doebley and Stec 1993). Whilst many mutations are prevalent in crop cultivars, some of the most obvious are those in genes for biosynthesis of pigmentation (Hirschberg 2001). The emergence of TILLING (Targeted induced local lesions in genomes), a highly efficient and rapid way of detecting genetic polymorphism and mutations, is bridging the gap between identifying mutants phenotypically and elucidating their genetic identity (McCallum *et al*. 2000; Emmanuel 2002; Comai *et al*. 2004; Slade and Knauf 2005; Xu *et al*. 2005). Whilst it is likely that much research effort will continue to be spent on the cloning of mutants (Ronen *et al*. 2000; Isaacson *et al*. 2002; Liu *et al*. 2004) an alternative strategy is offered by the development of marker assisted selection (Francia *et al*. 2005).

2.1.2 The Unexploited Resources of Wild Species

Domesticated varieties of plants (cultivars), have been selected by humans for the last 10 000 years: a fact that inevitably resulted in shallow gene pools in the majority of cases (Tanksley and McCouch 1997; Zamir 2001). This narrowing of the genetic basis of crop species is principally the result of two phenomena. First, unusual or even extreme phenotypes are often selected by humans and maintained in cultivars either for aesthetic reasons or due to properties that

facilitate harvesting. Although such extreme phenotypes confer the feeling that the cultivars that bear them are highly diverse the truth of the matter is that domestication usually results in a genetic bottleneck. Secondly, the fact that those cultivars are normally grown in uniform agricultural environments, and as such are less subject to the diversity of environment, tends to further narrow the genetic basis (McCouch 2004).

Landraces are the earliest form of cultivar and represent the first step in the domestication process. Landraces are thus more closely related to wild species than they are to the modern high-yielding varieties. Despite successes in the improvement of modern cultivars over the last century, it is important to stress that these improvements are generally only relevant under such carefully controlled conditions. Looking at the world population explosion and change in climate it is increasingly apparent that crop growth in more severe environments will become even more important. Early landrace varieties and wild species provide a broad representation of the natural variation that occurs in the species as a whole. The study of such natural variation being used increasingly both in crop species (Singh and Ocampo 1997; Septiningsih *et al*. 2003a; Huang *et al*. 2003; Pillen *et al*. 2003; Concibido *et al*. 2003; Rao *et al*. 2003; Borevitz and Chory 2004), and in the model species *Arabidopsis thaliana* (Koornneef *et al*. 2004; Kliebenstein *et al*. 2005), as a means of facilitating gene discovery and function.

Natural Variation in Breeding. In the early twentieth century the potential of wild species as a source for genetic variation was already recognized (Zamir 2001; McCouch 2004). First attempts met with severe problems including cross sexual incompatibility between the wild species and cultivated crop. However, today numerous examples exist in which wild introgression breeding has played a considerable contribution to the development of modern day varieties. These contributions include introgression of close to 30 independent genes for disease resistance in wheat (Fedak 1999), increased wheat yield by the introgression of the homologous arm from rye of the short arm of chromosome B1 (Villareal *et al*. 1995), commercial hybrids of tomato including different combinations of 15 independently introgressed disease resistance genes originating from diverse wild sources (Pan *et al*. 2000). There are even examples of the potential of this approach in maize, a naturally outcrossing species, since crosses of cultivated

maize to populations of its direct ancestor, teosinte, yield progeny with some traits of agricultural value (Ray *et al*. 1999). Whilst the use of exotic germplasm as a source for monogenic traits has been extensively exploited relatively little has been carried out with respect to the complex traits influenced by quantitative trait loci (QTL). Traits such as yield, compositional quality and stress resistance show complex inheritance patterns resulting from the segregation of numerous interacting QTL and possible epistatic effects (Zamir 2001). In recent years fruit from introgression lines of tomato, in particular the *Solanum lycopersicum* x *S. pennellii* substitution lines (Eshed 1995), have been evaluated for a relatively wide range of traits including morphological, transcriptional and a limited number of compositional traits (Fridman *et al*. 2000; Causse *et al*. 2004; Baxter *et al*. 2005a). Each of these 76 lines contain a single homozygous chromosomal segment from the green-fruited species *Solanum pennellii* in a *S. lycopersicum* background, thus covering the whole genome (Eshed 1995). Similar populations have been developed containing introgressions of other wild species tomato into the elite cultivar (Bernacchi *et al*. 1998; Pertuze *et al*. 2002; Frary *et al*. 2003; Canady *et al*. 2005; Fridman *et al*. 2005), as well as exotic libraries of wild species introgressions into other crop species including rice, barley, chickpea, soybean, lettuce and pepper (Singh and Ocampo 1997; Septiningsih *et al*. 2003a; Huang *et al*. 2003; Pillen *et al*. 2003; Concibido *et al*. 2003; Rao *et al*. 2003; Jeuken and Lindhout 2004). Moreover, analogous approaches have been used in mammalian systems with several different populations of congenic or chromosomal substitution lines of mice having been generated and genotyped (Singer *et al*. 2004).

2.1.3 Examples of Natural Variation in Breeding

Yield and Stress Resistance. Several key findings concerning the genetics of crop yield were made in recent years. These observations cover a wide range of species including tomato, rice, barley, chickpea, soybean and pepper (Singh and Ocampo 1997; Frary *et al*. 2000; Septiningsih *et al*. 2003b; Huang *et al*. 2003; Pillen *et al*. 2003; Concibido *et al*. 2003; Rao *et al*. 2003; Gur and Zamir 2004), but the most progress has been made by the detailed studies of tomato fruit size (Tanksley 2004). In this species wild species introgressions have played an

important role in the identification, and subsequent exploitation, of the genetics underpinning yield.

In the tomato, as in most plant species, domestication and selection correlated with a dramatic increase in fruit size. Whilst it has been known for some time that tomato fruit size is quantitatively controlled, it is only in the last decade that the genetic loci responsible for this were delineated (Tanksley 1993). One of the identified QTL, *fw2.2*, changes fruit weight by up to 30% and appears to have been responsible for a key transition during domestication. When transformed into large fruited elite species a cosmid derived from the *fw2.2* region of the small fruited species *S. pennellii* resulted in a reduction of fruit size by a predictable amount. The cause of the QTL could be denoted to a single gene, *ORFX*, which controls carpel cell number (Frary *et al*. 2000). Creation of an artificial gene dosage series, by means of transgenesis, further confirmed that this gene is responsible for tomato fruit size (Liu *et al*. 2003a). More recently a genetic map has been constructed for candidate genes and QTLs involved in tomato fruit size (and composition), giving some hints of genes that co-localise with some of the other fruit weight QTL (Causse *et al*. 2004).

In a recent study (Gur and Zamir 2004), demonstrated that by pyramiding three independent yield-promoting genomic regions introduced from the drought-tolerant *S. pennellii* were able to dramatically elevate tomato yield. The yield of the pyramided hybrid was more than 50% higher than that of a control market leader variety. This example has been confirmed under conditions of abiotic stress and clearly shows the potential of this approach for biotechnological improvement of crop yield (Morandini and Salamini 2003).

Forward genetics has had similar impact on identification of genetic factors underlying biotic resistance in plants (Gebhardt and Valkonen 2001; Bernal et al. 2005). Examples of the analysis of these traits include nineteen single dominant genes (*R* genes) for resistance to viruses, nematodes and fungi, which have been mapped on the chromosome of potato using DNA markers (Gebhardt and Valkonen 2001). The extensive characterisation of the *Pto* gene conferring resistance in tomato to strains of *Pseudomonas syringae* was also identified via

such genetic screens (Bernal *et al*. 2005). These examples are however only representative of a much larger number studies aimed at understanding of both biotic (Robert *et al*. 2001; McCouch 2004; Rose *et al*. 2005) and abiotic stresses (Singh and Ocampo 1997; Frankel *et al*. 2003; Vinocur and Altman 2005).

2.1.4 Metabolite Composition

The nutritional status of crop plants is ultimately dependent on their metabolic composition (Fernie *et al*. 2004). Whilst traits associated with yield and resistance have been the focus of much research, quality traits dependent on chemical composition are less well studied. A few exceptions are studies of carotenoid content in tomato (Liu *et al*. 2003b), protein content in maize (Moose *et al*. 2004), starch content in potato (Schäfer-Pregl *et al*. 1998; Chen 2001) and rice (Wang *et al*. 1995; Duan and Sun 2005; Ishimaru *et al*. 2005), being the subject of much attention. Moreover, seed soluble oligosaccharide, phytate, phosphate and cationic mineral content have all been measured across a wide range of *Arabidopsis* ecotypes (Bentsink *et al*. 2003; Vreugdenhil *et al*. 2004).

Protein, Starch, Cell Wall and Oil. The Illinois long term selection experiment for protein and oil content in maize experiment is the longest continuous genetic experiment in higher plants. It comprises of over 100 cycles of selection producing nine related populations exhibiting phenotypic extremes for grain composition and correlated traits (Moose *et al*. 2004). Beginning in 1896, 163 ears from the open-pollinated variety Burr's White were analysed for oil and protein concentration. The 24 highest and lowest-scoring ears for protein and oil formed the ILLINOIS HIGH PROTEIN (IHP), ILLINOIS LOW PROTEIN (ILP), ILLINOIS HIGH OIL (IHO), ILLINOIS LOW OIL (ILO) strains respectively. Another example is the case of protein introgression of a high-grain protein QTL from wild emmer wheat, *Triticum dicoccoides,* markedly improves the quality of pasta made from flour of wheat carrying the QTL (Kovacs *et al*. 1998), whereas QTL for oil content have been observed for many species including sunflower, soybean, rapeseed, pea, oat and maize (see for example Hyten *et al*. 2004).

A further study on storage carbon metabolism was provided by Hazen *et al* (Hazen *et al*. 2003). Here, the authors evaluated the cell wall composition of maize pericarp using the IBM recombinant inbred line population identifying QTL for xylose, arabinose, galactose and glucose content of the cell wall. Considering the wide physiological importance of the cell wall (Somerville *et al*. 2004), both approaches will provide an important basis for the analysis of the functional significance of cell wall composition across a range of biological processes. As the cell wall is also an important source of fibre in the diet, such investigations could allow the improvement of crop yield and compositional quality.

Soluble Carbohydrates. Soluble carbohydrates are an important determinant of crop quality from both high energy and taste perspectives. Given this fact and the relatively simple quantification methods, it is not of surprise that many QTL studies on the sugar composition across a wide range of crop species have been performed (Krapp *et al*. 2005; Salvi and Tuberosa 2005).

Fruit of the wild species tomato display considerable high variance in sugar content, this fact was exploited by the introgression of the sucrose accumulator gene (*sucr*) from *S. chmielewskii* into *S. lycopersicum* (Chetelat *et al*. 1995b) which dramatically changed the ratio of sucrose to hexose in the fruit. Similar research efforts have been placed on determining genetic factors responsible for glucose to fructose ratios in fruits; this is of applied interest since fructose is twice as sweet as its isomer.

QTLs for sugar content have been identified for a wide range of species including potato, tomato, melon and sugarcane (see (Monforte *et al*. 2004) for a recent example). Recently, a moderate tomato QTL for Brix (total soluble solids content), of which sugars and acids represent the major constituents, was mapped to a 486bp region of the cell wall invertase gene LIN5 (Fridman *et al*. 2000). Subsequent high-resolution studies, using a range of introgressions from different wild species germplasm, delineated this trait to a QTN (Quantitative trait nucleotide) that conferred altered kinetic properties to the enzyme (Fridman *et al*. 2004). Physiological studies of the *S. pennellii* introgression line harbouring this trait revealed that the increased Brix in the ripe fruit was due to an increase in sucrose and glucose. Enhanced invertase activity in the fruit columella led to a greater capacity to take up sucrose unloaded from the phloem (Baxter *et al*.

2005a). This example highlights the power of resolution available within introgression populations as well as the potential that the introgression of wild species alleles has on compositional improvement of crops. Several important questions remain open including the understanding of the mechanisms underlying other sugar or Brix QTL in tomato. First steps have been made in addressing this question since many sugar/ Brix QTLs have been defined in tomato fruit and positional cloning of sixty three genes involved in carbon metabolism (Causse *et al*. 2004), as well as transcriptional profiling of selected introgression lines along a developmental time series (Baxter *et al*. 2005b) have facilitated the identification of candidate genes underlying these traits.

Organic and Amino Acids. In contrast to the situation regarding soluble and storage carbohydrates, manipulation of organic and amino acid content via means of wide crosses has received little attention. A recent example of using wide crosses to change fruit acidity comes from melons, *Cucumis melo.* Sweet melon varieties are characterized by low organic acid content, whereas high acid melons do not accumulate high levels of sugars. Intriguingly, fruit acidity and the sugar: acid ratio is a fruit quality trait in essentially all fruits. A single recessive gene is responsible for high levels of sugar in melons, while high acidity is determined by a single dominant gene (Burger *et al*. 2002). A combination of both genes provides the melon with a unique taste due to a sugar: acid ratio not found in other sweet melons.

A handful of studies in tomato and peach have addressed the genetics behind the importance of these key components of taste (Fulton *et al*. 2002; Causse *et al*. 2004; Quilot *et al*. 2004). It should also be noted that the organic acids and possibly also glutamate form major constituents of the total soluble solid or Brix trait described above. More direct studies identified QTLs for titratable acidity, pH and citrate and malate content (Causse *et al*. 2002). Recently such studies have been integrated with taste analyses by trained panels, which identified the glutamate: sugar ratio as an important component in the flavour of tomatoes (Fulton *et al*. 2002), as well as alongside analysis of plant volatiles.

Vitamins, Pigments and Antioxidants. Vitamins, pigments and antioxidants have been mostly studied in highly coloured, genetically tractable crops such as

tomato, peach and melon (Lewinsohn *et al*. 2005a). This was initially driven by the desire to have attractive looking food; however research in the medical field suggests that pigments such as lycopene in addition to vitamins confer health-benefits to the consumer. Relatively little information has been achieved to date on vitamins and antioxidants, where most studies have concentrated on Vitamin C and E (Liu *et al*. 2004; Rousseaux *et al*. 2005), glutathione (Davey and Keulemans 2004) and total phenolic contents (Rousseaux *et al*. 2005). In all of these studies QTL were identified for enhanced metabolite content suggesting that manipulation of these compounds is feasible.

The *S. pennellii* introgression lines defined above were analysed with respect to carotenoid and lycopene content. Based on trials in different environments, 16 QTL that modified the intensity of the red colour of ripe fruit were assigned to chromosomal regions. Candidate sequences associated with the carotenoid biosynthesis pathway were mapped to 23 loci. This example reveals the presence of several candidate genes for the accumulation of pigments that are unrelated to the structural genes of the known biosynthetic pathways of these compounds. The above studies suggest that the use of wild species alleles is a potentially important strategy for elevating the levels of these compounds in food crops.

Volatiles. As discussed above the flavour of many fruits is highly dependent on sugar and acid contents and also on the sugar: acid ratio. However, the correlation between consumer judgements and fruit and vegetable flavour is relatively weak- a fact that justifies the study of volatile composition. A recent study has demonstrated that GC/MS is able to quantify almost 400 such volatile compounds (Tikunov *et al*. 2005), however, extensive research has only been carried out on a handful of these (Baldwin *et al*. 2000). Utilizing a recombinant inbred line population generated from an intraspecific cross between a cherry tomato line with a good overall aroma intensity and an inbred line with a common taste but with bigger fruits, allowed the identification of major QTL for six aroma volatiles (Lecomte *et al*. 2004), however, up to date the exact mechanisms underlying these traits remains unresolved. Intriguingly, a broader analysis recently revealed that carotenoid pigmentation in both tomato and watermelon fruits affects the volatile composition of these fruits with lycopene-containing fruit displaying greater

contents of non-cyclic norisoprenoids, such as geranial and neral, than those fruit that do not contain lycopene (Lewinsohn *et al*. 2005b).

The identification of *malodorous*, a wild species allele affecting tomato aroma, allowed the identification, within the *S. pennellii* introgression lines, of a QTL displaying a markedly undesirable flavour (Tadmor *et al*. 2002). This trait corresponded to a 60-fold increase in phenylethanol and phenylacetaldehyde as compared to the cultivated variety providing a genetic explanation for one of the aroma changes that occurred during domestication. Whilst the above examples highlight the importance of wild germplasm for agronomic application and in improving fundamental understanding of metabolic pathway structure and regulation, it is also a great resource for assessing trait evolution. Given that, highlights the importance of wild germplasm for crop improvement, further they are a valuable resource for the fundamental understanding of metabolic pathways and regulation (Fernie *et al*. 2006).

2.2 Techniques in Contemporary Biology

Biological research has undergone a dramatic change from the gathering of small-scale datasets to routine large-scale data acquisition of recent years. This process started in the mid 90's with multinational efforts to sequence entire genomes. Up to date over 200 genomes, covering species from plants, microbes, funghi and mammals, have been fully sequenced (Goffeau *et al*. 1996; Blattner *et al*. 1997; Adams *et al*. 2000; The Arabidopsis Genome Initiative 2000; McPherson *et al*. 2001; Goff *et al*. 2002; Yu *et al*. 2002).

The genome information for crop species is very poor, with rice being the only crop for which the full genome sequence is publicly available. However, in the last few years initiatives and consortia have been funded to start the sequencing of other crop plants. For example, in the beginning of 2004 the International *Solanaceae* Genomics Project (SOL) started to sequence the genome of *S. lycopersicum* (Mueller *et al*. 2005a), which will act as a model to the entire *Euasterid* clade.

Post-genomic technological advances of recent years facilitated the multiparallel determination of the expression levels of many thousands of genes by mRNA or

transcript profiling (Celis 2000). Moreover, achievements in protein research have led to more and more sophisticated high-throughput proteomics approaches reliant on highly sensitive mass spectrometry platforms (Sickmann *et al*. 2003; Aebersold and Mann 2003).

Until very recently, metabolite analysis has focussed on a handful of compounds that were of possible interest in a given experiment. Whilst this approach yielded important insights into the pathway of interest to the researcher, it was simultaneously limited to the pathway under study and often very labour intensive. The recent hyphenation of chromatographic separation methods to mass spectrometry, whilst already proposed years ago (Jellum *et al*. 1975; Jellum 1977), has only recently become routinely applicable (Roessner *et al*. 2001a; Roessner *et al*. 2001b; Soga and Imaizumi 2001; Tolstikov 2002; Sato *et al*. 2004). This achievement enabled the multivariate analysis of complex sample mixtures with hundreds of analytes of known or unknown chemical structure. Estimations of the number of naturally occurring compounds are currently of the order of 200,000 (De Luca and St Pierre 2000). Whereas only a subset of metabolites are present in one single organism, the estimated numbers for different eukaryotes currently range from 4000 to 20,000 metabolites, with plant species being the method of scale (Fernie *et al*. 2004).

The investigation of the metabolome of the cell can be defined by a range of terms – metabolite profiling, metabolite fingerprinting, metabonomics and metabolomics (Fiehn 2002; Sumner *et al*. 2003). Here the term metabolite profiling is used to describe the rapid, but accurate quantification of metabolites using a single analytic technique.

2.2.1 Metabolite Profiling Platforms

Metabolomics is the general term for the study of the global metabolite composition of a system under a given set of conditions (Sumner *et al*. 2003). As metabolite profiling platforms have only been recently developed a great deal of innovation and method improvement is still in process. Nowadays platforms range from different separation systems to various detection components. Separation systems include liquid chromatography (LC), gas chromatography (GC) and capillary electrophoresis (CE). On the detector side different types of mass

spectrometry (MS), nuclear magnetic resonance (NMR) and ultraviolet light spectroscopy (UVIS) systems exist.

The complex nature of biological compounds does not allow the measurement of all metabolites of one organism by a single comprehensive platform but rather requires the analysis by more than one technique. Hence, the selection of a suitable system depends on the biological question and in general represents a compromise between selectivity, speed and sensitivity (Sumner *et al*. 2003). While NMR systems are rapid, selective and non-destructive they have low sensitivity and are very cost intensive. In a LC/NMR analysis of plant extract 2700 putative compounds could be detected, however less than 50 were identified (Noteborn 2000). Nevertheless NMR is of unparalleled importance in structure elucidation and for *in viv*o determination of metabolite levels. CE/MS and LC/MS systems have been shown to be highly sensitive and are able to separate compounds with high molecular weight, but these set-ups have still problems with chromatographic reproducibility and resolution, as well as relatively high separation time, selectivity and costs. Further advance might overcome these problems and the problems of identification of tremendous number of unknown peaks. GC/MS, in contrast, has proven to be a highly reliable and robust platform, providing good separation and high sensitivity at comparatively low costs. For this reasons it is now widely and routinely applied to study metabolic phenotypes in plants (Roessner *et al*. 2000b; Roessner-Tunali *et al*. 2003a; Chen *et al*. 2003), microbes (Strelkov *et al*. 2004; Barsch *et al*. 2004; Stephanopoulos *et al*. 2004; Bolling and Fiehn 2005) and mammals (Jiye *et al*. 2005; Fustinoni *et al*. 2005). Nevertheless there are drawbacks in the study of the metabolome by GC/MS. In order to visualize the metabolic content of one sample the compounds have to be volatile. This requirement is accomplished by chemical derivatisation, but at the cost of additional time and chemical modification.

As the metabolome of an organism is constituted of a complex mixture of highly diverse and often stereomeric chemical compounds the identification presents a grand challenge to the metabolomics community. The majority of components covered by metabolite profiling techniques remain to be identified. Whilst 10-15% of the unknowns can be designated to a specific compound class, the rest are

compounds with completely unknown properties. Thus it is of major importance to identify unknown peaks, but at the burden of time and cost for which must be met by the wide metabolomics community. This demands the establishment of an efficient and standardized way to exchange information (Bino *et al*. 2004; Kopka *et al*. 2005), i.e. mass spectral and retention index information and the warehousing of this information in public repositories, similar to genome, transcriptome and proteome databases (Gollub *et al*. 2003; Schomburg *et al*. 2004; Mewes *et al*. 2004; Zimmermann *et al*. 2004).

2.2.2 Applications of Metabolite Profiling

Metabolite profiling is becoming increasingly more important in many research areas. Its application ranges from diagnostics through gene function identification to systems biology. In diagnostics, metabolite profiling has been used to study the mode of action of various herbicides (Sauter *et al*. 1988). Barley seedlings were treated with LC_{50} dosages of various herbicides and were then comparatively profiled with untreated plants. In conjunction with bioinformatic tools acetyl CoA carboxylase (ACC) and acetolactate synthase (ALS) inhibitors could be discriminated. Similar studies on human blood samples led to the identification of inborn errors of metabolism by metabolomic assessment (Rashed *et al*. 1997). Loss of function of a gene can facilitate the identification of its function. The Keio knock-out population of *E. coli* consisting of approximately 4000 strains is currently being profiled at the metabolome level to study the effect of gene loss (pers. communication, T. Soga, Japan, 2005). Gain of function analysis by metabolite profiling similarly provides very powerful information. A transgenomic approach where every single gene of *E. coli* and yeast were independently expressed in *Arabidopsis thaliana* highlights this fact, for example the expression of a threonine aldolase resulted in an expected increase in threonine, but also revealed an increase in the metabolites of the methionine pathway (homoserine and methionine) and a decrease in the isoleucine biosynthesis, thus revealing novel information of the network of amino acid metabolism (Fernie *et al*. 2004). A further example is the gene function analysis of the Myb-like transcription factor PAP1 by correlative transcript and metabolite profiling (Tohge *et al*. 2005). Here the authors report changes caused by over-expression of PAP1 on the transcript and

metabolite level, with further discussion on the function and role of the upregulated genes within a network context. However, causality cannot be implied and the evaluation of these correlations has to be cautiously considered, especially given aspects of post-transcriptional modification. Nevertheless, these examples demonstrate that gene function annotation can be facilitated by using metabolite profiling. Lately, the analysis of stress responses by metabolic analysis could reveal new insight into the reaction of a plant to cold stress (Kaplan *et al*. 2004; Scholz *et al*. 2005). Kaplan *et al*. showed that following 1h cold stress *Arabidopsis* plants responded with an early increase of maltose. In a more recent study the effect of nitrogen limitation in tomato plants was investigated. Nitrate starvation has a dramatic effect on leaf metabolism and resulted in decreases of amino and organic acids, whereas levels of carbohydrates increased, these findings confirm and extend studies of nitrate nutrition over many years (Stitt *et al*. 2002; Urbanczyk-Wochniak and Fernie 2005).

The term systems biology, is variously defined, however here the definition of Sweetlove *et al*. - the comprehensive multidimensional analysis of the inventory of the cell – will be used (Sweetlove *et al*. 2003). Systems biology approaches are now starting to include metabolomic analysis more frequently. Metabolite-transcript correlations from a large data set collected throughout development in wild type and transgenic potato tubers, which were engineered to have enhanced sucrose metabolism, allowed the identification of candidate genes for biotechnology approaches (Urbanczyk-Wochniak *et al*. 2003). In this study the transcript levels of approximately 280 transcripts were compared to changes in metabolite levels in paired samples. 517 metabolite-transcript correlations (P<0.01) out of over 26,616 were identified to be significant. Most of the correlations were known, but some strong correlations between genes and nutritional important metabolites were new and thus could lead to the identification of novel candidate genes for metabolic engineering for a broad number of traits.

These examples show the power of metabolite profiling in the understanding of biological systems and proves to be a tool, in conjunction with other –omics approaches, which can help to gain more insight into the control and regulation of metabolic networks in living organisms. Metabolite profiling can also be extended to study and explore the phenotypic relevance of genomic regions by utilizing introgression lines or other mapping populations, e.g. *S. pennellii* introgression

lines created by Dani Zamir (Eshed 1995). This approach may therefore lead to the identification of environmentally stable QTL and with the help of map based cloning to the discovery of regulatory genes that modulate metabolite content (Fernie *et al*. 2004).

2.3 Aim of the Thesis

To investigate the nutritional composition in the cultivated tomato and the species of tomato that can be readily crossed with elite cultivated species – the *Solanum lycopersicum* complex- a GC/MS based metabolite profiling technique was applied. As a prerequisite for this study a mass spectral library with mass spectral and retention index (RI) information for metabolite identification of known and unknown structure was generated. Further, to investigate the power of natural variation a set of tomato introgression lines, which harbour one single defined chromosome segment substitution from the green-fruited wild species *S. pennellii* in the background of *S. lycopersicum cv.* M82 was subjected to metabolite profiling. Finally, the pathway of the phenylalanine derived volatile phenylethanol was elucidated by assessment of the levels of this metabolite across this population in conjuncture with stable isotope trace analysis and reverse genetics.

3 GC/MS libraries for the rapid identification of metabolites in complex biological samples

Nicolas Schauer[1$], Dirk Steinhauser[1], Sergej Strelkov[2], Dietmar Schomburg[2], Gordon Allison[3], Thomas Moritz[4], Krister Lundgren [4], Ute Roessner-Tunali[5], Megan G. Forbes [5], Lothar Willmitzer[1], Alisdair R. Fernie[1], Joachim Kopka[*1]

[1]Max Planck Institute of Plant Molecular Physiology, Am Muehlenberg 1, D-14476 Golm, Germany
[2]University of Köln, CUBIC - Institute of Biochemistry, Zuelpicher Str. 47, D-50674 Köln, Germany
[3]Institute of Grassland and Enviromental Research, SY23 3EB, Aberystwyth, Wales, UK
[4]Umea Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE–901 83 Umea, Sweden
[5]Australian Centre for Plant Functional Genomics, School of Botany, University of Melbourne, Victoria 3010, Australia
$ Contribution by NS to *Solanum lycopersicum* MSRI library and preparation of the manuscript

Abbreviations: GC, gas chromatography; MS, mass spectrometry; MST; mass spectral metabolite tag; RI, retention time index; TOF, time of flight

3.1 Abstract

Gas chromatography – mass spectrometry based metabolite profiling of biological samples is rapidly becoming one of the cornerstones of functional genomics and systems biology. Thus the technology needs to be available to many laboratories and open exchange of information is required such as those achieved for transcript and protein data. The key-step in metabolite profiling is the unambiguous identification of metabolites in highly complex metabolite preparations with composite structure. Collections of mass spectra, which comprise frequently observed identified and non-identified metabolites, represent the most effective means to pool the identification efforts currently performed in many laboratories around the world. Here we describe a platform for mass spectral and retention time index libraries that will enable this process (MSRI; http://csbdb.mpimp-golm.mpg.de/gmd.html). This resource should ameliorate many of the problems that each laboratory will face both for the initial

establishment of metabolome analysis and for its maintenance at a constant sample throughput.

3.2 Introduction

In the last decade the maturity of genomic technologies generated a vast amount of sequence data and thus allowed full insight into the finite number of genes which constitute organisms. As a consequence biological science went through a paradigm-change and today focuses on unravelling gene function and regulation. With these tasks at hand technologies have been developed which aim at comprehensive and non-biased monitoring of gene expression, and coinciding effects on protein composition and changes in metabolism. Consequently new fields emerged in biological science which we today call, transcriptomics, proteomics and metabolomics. With increasing amount and diversity of "-omics" data the need for standardization by the research community arises and availability of tools for a user-friendly, open access to the flood of information have become essential.

One of the first "-omics" databases, BRENDA, was developed in 1987 (Schomburg *et al*. 2004). BRENDA is a powerful database of enzyme and metabolic information initially published as a series of books, now adapted to a relational database and accessible through the worldwide-web. BRENDA hosts about 83000 different enzymes from 9800 different organisms and describes enzyme function, taxonomy, sequences and enzyme ligands. The Munich Information Centre for Protein Sequences (MIPS) provides databases related to protein sequences based on whole genome analysis and annotation. For example, MIPS hosts databases of *Saccharomyces cerevisiae* and *Neurospora crassa* which comprise maps of protein-protein interactions, protein localization, and information on transcription factors, cDNA libraries and gene homology (Mewes *et al*. 2004). Transcript profiling rapidly evolved into a worldwide accepted and generally applied laboratory tool. Subsequently database were designed and established, which efficiently deal with transcriptome data. The Stanford Microarray Database (SMD), which hosts data of over 3500 DNA-Microarrays of twelve distinct organisms, including bacteria, plants and animals, was the first implementation to fulfil this aim (Gollub *et al*. 2003).

With the full availability of the human genome sequence the need and opportunity to understand the structure and function of all proteins, beyond those with enzymatic properties, was met with respective initiatives. For example HPI, the human protein initiative, focuses on the annotation of both the human genome and proteome. As proteins are generally regarded to determine cellular function, the full exploration of the proteome will be crucial. The goal of HPI is to deliver this information in high quality to facilitate further investigations of the genomic and proteomic data (Hermjakob *et al*. 2004).

Presently, a wealth of databases houses information gathered at the genomic, transcriptomic, proteomic and metabolomic, e.g. (Schomburg *et al*. 2004; Kanehisa *et al*. 2004), levels of life. However, there is a significant lack of a metabolome database, capable of storing the flood of data arising from analysis of biological samples using established gas chromatography – mass spectrometry (GC/MS) techniques for metabolome (Roessner *et al*. 2000a; Roessner *et al*. 2001a; Strelkov *et al*. 2004; Gullberg *et al*. 2004) and fluxome analysis (Fischer and Sauer 2003; Sauer 2004; Kromer *et al*. 2004). Most promisingly, first efforts have already been made by the plant metabolomics community to agree on conventions for data formats and the description of metabolomics experiments (Bino *et al*. 2004; Jenkins *et al*. 2004).

3.3 GC/MS based metabolome analysis: application and key challenge

GC/MS based metabolome analysis has profound applications in discovering the mode of action of drugs or herbicides and helps unravel the effect of altered gene expression on metabolism and organism performance in biotechnological applications. The prerequisite and thus key challenge of metabolite profiling is the rapid, reliable and unambiguous identification of hundreds of metabolites in highly complex preparations, such as blood plasma, intracellular microbial extracts, or complex plant and animal samples. Identification is routinely performed by time-consuming standard addition experiments using commercially available or purified metabolite preparations. Thus a strong need for a publicly accessible database exists, harbouring the evidence and underlying metabolite identification in complex GC/MS profiles from diverse biological sources. In addition the non-supervised

collection of as yet unidentified mass spectra of metabolites, “so-called” mass spectral metabolite tags (MSTs), will most likely be highly effective for future identification efforts and discovery of novel metabolic markers. In this report we present a platform of mass spectral and retention time index (MSRI) libraries, generated using identical types of capillary GC columns, however, utilizing two independent GC/MS detection technologies, namely quadrupole GC/MS (Roessner *et al*. 2000a; Fiehn 2000; Roessner *et al*. 2001a; Strelkov *et al*. 2004) and GC-TOF (time of flight)-MS (Wagner *et al*. 2003; Gullberg *et al*. 2004). In the following study we will present three test cases which illustrate the general applicability of this library for the key processes of GC/MS based metabolite profiling, (i) identification or preliminary classification of all MST components, which are present in any given biological sample, (ii) query for those biological samples that contain a certain metabolite, (iii) matching of metabolite identifications made on different GC/MS systems and by different laboratories.

3.4 Mass spectral and retention time index (MSRI) libraries for GC/MS

We propose public exchange and open access of mass spectral identifications from GC/MS metabolite profiles, for example, through a web-based platform of mass spectral and retention time index libraries (MSRI; http://csbdb.mpimp-golm.mpg.de/gmd.html (Kopka *et al*. 2005)). In addition, we provide downloadable files, which can be imported into the currently leading and widely accepted NIST02 mass spectral search program or AMDIS, the automated mass spectral deconvolution and identification system (National Institute of Standards and Technology, Gaithersburg, MD, USA) (Ausloos *et al*. 1999) (Stein 1999). Both software systems are publicly available from http://chemdata.nist.gov/mass-spc/amdis/ and http://chemdata.nist.gov/mass-spc/Srch_v1.7/index.html. Our libraries are classified according to technology and degree of manual mass spectral identification that was required for the library construction. After import into NIST02 the current libraries may be fused into one or customized subsets generated. Q_MSRI and T_MSRI libraries contain MSTs, which were either generated on three identically configured quadrupole (Q_MSRI) GC/MS systems or on a single time of flight (T_MSRI) system. All systems were run with identical

settings except for the temperature program and scanning rate. Mass spectral libraries, which exclusively comprise manually evaluated, identified or classified MSTs, are assigned to ID-libraries, indicative of supervised identifications. Libraries which were generated exclusively by automated deconvolution were assigned NS indicative of the non-supervised mode of construction. The NS-libraries may contain deconvolution errors, such as multiple mass spectra for single components, accidental deconvolutions, due to random fluctuations of background noise, or partial and mixed, in other words, chimeric mass spectra of metabolic components. In addition, detailed information on processed biological samples, source of pure reference compounds, respective collaborators and previous citations is provided. For those queries on the current mass spectral collection, which cannot be performed within NIST02 we offer a tab delimited compilation of the manually evaluated mass spectra and access through web query forms. Currently we support queries within ID-libraries, such as compound search, mass spectral search using names or mass spectra and customized library generation for subsets of mass spectra.

We previously demonstrated that both mass spectrum and retention time index are required for unequivocal metabolite identification in GC/MS profiles (Wagner *et al*. 2003). This feature was not available in commercial mass spectral comparison software. Therefore, the central feature of our web search forms is optional restriction of searches to RI windows and sorting of hit lists according to RI deviation and mass spectral similarity. For a shortlist of the currently implemented matching tools and queries please refer to (Kopka *et al*. 2005).

The present version of the Q_MSRI_ID library contains 1166 identified compounds or annotated MSTs, which represent 574 non-redundant compounds. Of these compounds 306 are unambiguously identified, while the residual MSTs are annotated with the best the mass spectral match from a commercially available mass spectral collection (National Institute of Standards and Technology, Gaithersburg, MD, USA). The T_MSRI_ID collection has a similar size, namely 855 MSTs with 229 identifications within the set of 632 non-redundant components. The non-supervised collections comprise close to 30,000 MSTs from a range of plant organs, root, leaf, tuber, stolon, flower, fruits in different developmental stages, and suitable non-samples controls. Plant species covered are model plants, crops and related wild species, such as *Lotus japonicus*,

Arabidopsis thaliana, *Solanum tuberosum*, *Nicotiana tabacum*, *Solanum lycopersicum*, *Solanum pennellii*, *Solanum parviflorum*, *Solanum pimpinellifolium*, *Solanum habrochaites and Solanum neorickii*.

3.5 Test cases

3.5.1 Test case 1: Analysis of sample composition

Non-supervised MSRI data allow screening for differences in various samples. A given biological sample can be compared to the non-supervised library. All MSTs, which match in their mass spectrum and RI, within certain thresholds, such as mass spectral match > 650 and RI deviation < 3.0, will be presented as possible hits, thus allowing the evaluation of whole biological samples for differences in composition with respect to mass spectral datasets from the established MSRI library. In the following we applied the above thresholds for automated identification but still performed additional manual verification on each of the best hits.

A supervised database is a valuable tool to identify compounds with known RI and mass spectra in specific biological samples. A typical example of the metabolite composition from polar bacterial extracts demonstrates the scope of GC/MS based metabolite profiling (Figure 1). To further illustrate the power of this tool, we have chosen the plant specific flavonol (kaempferol), a phytosterol (ß-sitosterol) and vitamin E (α-tocopherol). Taking into account that sheep are herbivores, we expected to find ß-sitosterol and kaempferol also in sheep plasma samples. To test our hypothesis we have performed a MSRI library searching for these compounds in sheep blood plasma samples, resulting in mass spectral hits for ß-sitosterol and kaempferol in the plasma composition (Table 1).

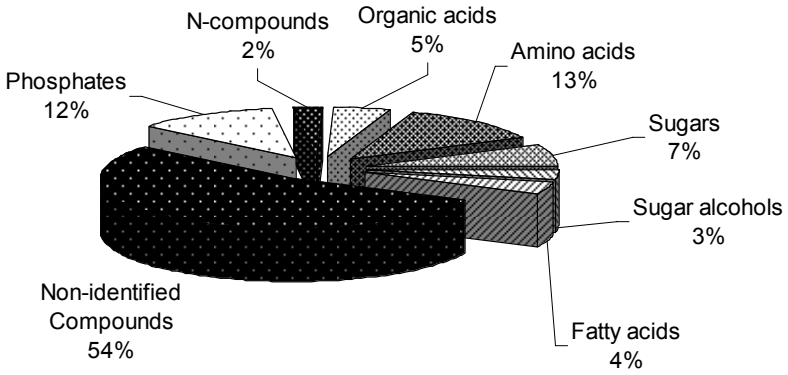

Figure 1. The distribution of major compound classes, which were identified in GC-TOF/MS profiles in cellular extracts from *Corynebacterium glutamicum*. Note that more than 50 % of the metabolites that are covered by GC/MS are currently non-identified mass spectral tags (MSTs).

It is also known that tocopherol and its derivatives play an important role in the human diet and thus are important targets for novel nutrigenomics approaches (Muller and Kersten 2003; Davis and Milner 2004). Tocopherol is additionally widely hypothesized to be helpful in preventing diseases associated with oxidative stress. Therefore the question arose, whether this substance can be easily identified in mammal tissues? Here we demonstrate the power of a MSRI library to search for α-tocopherol in different samples from animal, as well as plant tissues. Considering the importance of this compound for mammals we expected to find α-tocopherol in animal samples and querying the library indeed resulted in the identification of α-tocopherol in blood plasma sample from sheep (Table 1).

Amino Acids	**Organic acids**	**Phosphates**
2-Aminobutyric acid	2-Ketoglutaric acid	Adenosine-5-monophosphate
4-Hydroxyproline	4-Hydroxybenzoic acid	Glycerol-3-phosphate
Alanine	Benzoic acid	myo-Inositol-phosphate
β-Alanine	Citric acid	Phosphoric acid
Glycine	Erythronic acid	**Sugars***
Alanine	Fumaric acid	
Arginine	Gluconic acid	Arabinose
Asparagine	Glucuronic acid	Fructose
Cysteine	Glutaric acid	Glucose
Glutamic acid	Glyceric acid	Raffinose
Glutamine	Gulonic acid	Ribose
Homoserine	Isocitric acid	**N-compounds**
Isoleucine	Itaconic acid	
Leucine	Malic acid	Allantoin
Lysine	Threonic acid	Hypoxanthine
Methionine	trans-Sinapinic acid	Inosine
Phenylalanine		Thymine
Proline		**Alcohols**
Serine	**Lipids**	
Threonine		Erythritol
Tryptophan	9,12-(cis,cis)-Octadecadienoic acid	Glycerol
Tyrosine	Hexadecanoic acid	myo-Inositol
Valine	Octadecanoic acid	Sorbitol
N-Acetylglycine	α-Tocopherol	Threitol
Ornithine	β-Sitosterol	Xylitol
Pyroglutamic acid	Campesterol	
S-Methyl-cysteine	Cholesterol	Kaempferol
* With rare exceptions DL stereoisomers are not separated on the current choice of GC capillary column.		

Table 1. Metabolites found in preparations of blood plasma samples from domestic sheep.

3.5.2 Test case 2: Analysis of metabolite occurrence

A laboratory which maintains a GC/MS based metabolite profiling facility will continually need to identify metabolites. Frequently the identity of previously non-identified MSTs will be discovered and the question will arise, if these MSTs were found in previous experiments (Fiehn *et al*. 2000; Roessner *et al*. 2001a; Wagner *et al*. 2003) or by other laboratories (Gullberg *et al*. 2004; Strelkov *et al*. 2004). It will therefore be important to identify the type and source of sample, which showed this MST. For this purpose non-supervised mass spectral libraries, which may hold independently repeated analyses of each type of sample, will be valuable tools. We chose chlorogenic acid, a typical secondary product of *Solanaceous* species, and the ubiquitous precursor quinic acid to demonstrate the possible gain of knowledge to be retrieved from non-supervised mass spectral libraries (Table 2). Our analysis indicated the presence, above detection limit, of quinic acid in almost all profiles analysed, whereas caffeic acid, the second precursor of chlorogenic acid, was present above detection limit only in leafs and Lotus japonicus nodules. In agreement with expectations chlorogenic acid and its positional isomers were found with good mass spectral match and RI deviation in *Solanum* samples.

Species	Organ	Quinic acid		Caffeic acid		Chlorogenic acid		4-Caffeoylquinic acid		5-Caffeoylquinic acid	
		Match	Δ RI	Match	Δ RI	Match	Δ RI	Match	Δ RI	Match	Δ RI
Arabidopsis thaliana (L.) Heynh.	Leaf	649	1.9								
Arabidopsis thaliana (L.) Heynh.	Root	723	2.0								
Lotus japonicus	Root lateral										
Lotus japonicus	Root primary										
Lotus japonicus	Nodule			752	-2.4						
Lotus japonicus	Flower	795	0.4								
Lotus japonicus	Leaf developing	685	0.1								
Lotus japonicus	Leaf mature	565	0.2								
Solanum lycopersicum	Root	963	-0.2			974	-0.4				
Solanum lycopersicum	Leaf	854	0.5	838	-0.8	975	0.1	939	1.1	822	0.7
Solanum lycopersicum	Green fruit	967	1.4			970	1.6				
Solanum lycopersicum	Orange fruit	930	1.3			965	1.6				
Solanum lycopersicum	Red fruit	964	0.7			949	1.8	916	2.0		
Solanum neorickii	Fruit 45DAF	964	0.9			976	1.1	555*	2.1		
Solanum neorickii	Leaf	950	-0.9	862	-1.9	974	-1.2	971	0.9	827	1.0
Solanum habrochaites	Fruit 45DAF	959	0.6			949	-0.4	920	1.2		
Solanum habrochaites	Leaf	909	0.5	825	-1.9	936	0.6	962	0.9	747	0.9
Solanum parviflorum	Fruit 45DAF	964	0.8			888	0.8				
Solanum parviflorum	Leaf	926	-1.0	856	-1.6	975	-0.2	971	1.2	802	1.8
Solanum pennellii	Fruit 45DAF	799	0.0			974	-0.6				
Solanum pennellii	Leaf	953	-0.4	848	-0.6	973	1.3	925	1.5		
Solanum pimpinellifollium	Fruit 45DAF	778	-0.7			977	0.9				
Solanum pimpinellifollium	Leaf	912	-1.6	847	-2.4	966	0.1	961	1.0	840	0.4

* Low mass spectral match results from mixed mass spectra with a co-eluting compound (presence of compound was manually verified).

Table 2. Occurrence of metabolites in non-supervised libraries of *Arabidopsis thaliana, Lotus japonicus* and *Solanum* species. Presence of a metabolite is validated by best mass spectral match on a scale 0-1000 (Match) and smallest deviation of retention time index (Δ RI).

3.5.3 Test case 3: GC/MS system transfer of metabolite identifications

Almost all metabolites were analysed either in different laboratories or on two GC/MS technology platforms, GC-QUAD-MS and GC-TOF-MS. The resulting information on retention time indices from both technology platforms clearly demonstrated strict linearity in a comparative analysis of both systems, provided the same type of capillary column was used (Figure 2). Thus RI prediction through regression appears highly feasible for different GC/MS systems, but only when identical column types are used. Nevertheless, we detected compound specific deviations from the prediction (Figure 2). On average we observed an error of ~ 5.4 RI units, but most deviations were minor and within the expected range taking the typical reproducibility of retention time indices within one system into consideration (Wagner *et al*. 2003), namely up to 2.0 RI units (standard deviation), depending mostly on changes in metabolite amount. In addition, typical metabolite classes, such as sugars, fatty acids or amino acids (Fiehn 2000; Roessner *et al*. 2000a; Roessner *et al*. 2001a; Wagner *et al*. 2003; Strelkov *et al*. 2004; Gullberg *et al*. 2004), mostly exhibited common positive or negative trends of deviation. Therefore RI information obtained from one technology platform will allow good prediction of retention time indices, if reference compounds are already mapped on both systems. Use and implementation of RI systems for different GC column types is ongoing effort in our laboratories (data not shown) but RI prediction will require other methods than regression, because the elution-sequence of compounds is known to change.

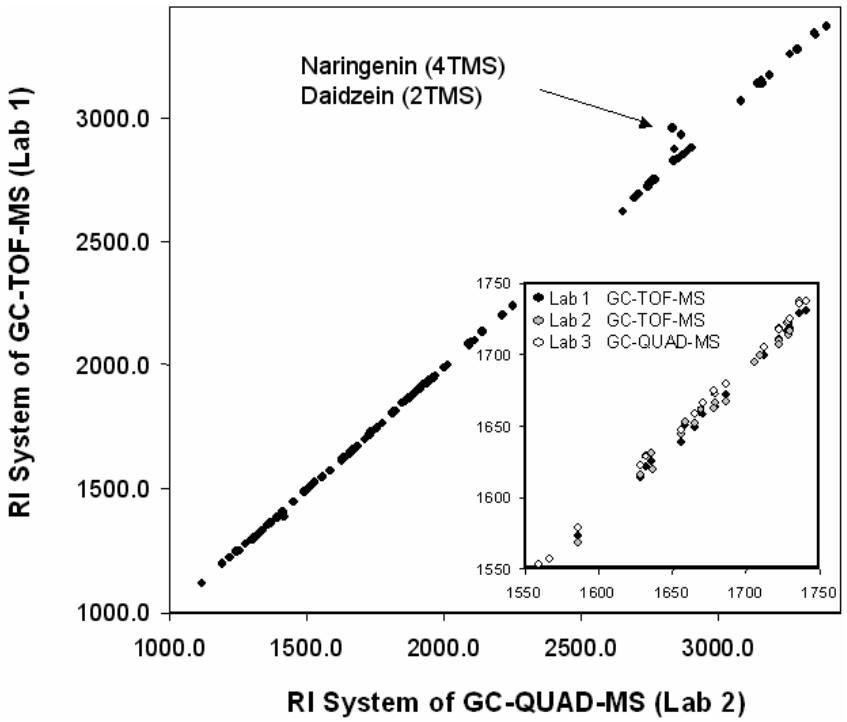

Figure 2 Comparative analysis of retention time indices (RI), which were determined in parallel on two GC/MS technology platforms in different laboratories, demonstrates that knowledge of one RI system allows good prediction of RI in a second system. Retention time indices were determined on time of flight (TOF) and quadrupole (QUAD) GC/MS systems. GC capillary columns were of identical build. Lab 2 operates TOF and QUAD GC/MS systems.

3.6 Conclusions

The hypothesis and test cases described here present, for the first time, a comprehensive MSRI library database covering MSTs of GC/MS metabolite profiles from mammals, corynebacteriae and major plant species. It includes in total more than 2000 fully evaluated mass spectral data sets obtained using two distinct technology platforms with 1089 non-redundant and 360 identified MSTs. The database is designed to be continuously extended by additional accessory information as it becomes available. We demonstrated the use of this MS/RI library to screen biological samples for known compounds and showed the appliance of the non-supervised library for screening samples for known or recently identified mass spectra.

This library is constantly being updated with every new biological sample and application run in-house. Because even slight changes in GC/MS settings, such as

carrier flow, temperature ramp, and dimension/ make of capillary columns, induce shifts in retention behaviour of substances, GC/MS systems need to be recalibrated after each change. As there is currently no solution - other than recalibration - addressing the problem of RI shifts using different GC/MS machines we would like to offer to the biological and metabolite profiling community to perform qualitative analysis of any biological sample using our currently running protocols.

In addition to offering this service to the community we believe that the data presented here demonstrate three general applications of such libraries, which will help to advance the field. (i) The composition of still non-characterized biological samples, for example blood plasma, or microbial extracts (data not shown) can be screened for identified constituents, and tentative best matching compounds. (ii) Occurrence of identified metabolites can be analysed in a large range of biological samples, such as different plant organs or species. For this purpose we provide libraries comprising samples from tomato, related wild type species, and other *Solanacea*, collections of different organs of *Lotus japonicus*, *Arabidopsis thaliana*, and preparations from microbial species. (iii) Subsequent analysis of samples on two different GC/MS systems facilitates transfer of identifications made on the first system to the second. We present data on identifications, which were made in-parallel on quadrupole GC/MS and GC-TOF-MS systems in different laboratories worldwide. This is therefore the first validation that metabolite profiling, when carried out with appropriate care, can yield comparable results between laboratories. Given the number of independent laboratories involved in this study we believe that it offers similar reassurance as provided to the microarray community by the multi-laboratory Affymetrix microbial gene expression study. We are convinced that the effort described here will be useful on several levels. Not only will it meet a recently expressed demand within the metabolomics community (Bino *et al*. 2004; Fernie *et al*. 2004), which was already apparent in earliest metabolomics applications in clinical diagnostics, but it will also aid laboratories entering the field of metabolomics (Jellum 1977).

4 Metabolic profiling of leaves and fruit of wild species tomato: a survey of the Solanum lycopersicum complex

Nicolas Schauer[1], Dani Zamir[2] and Alisdair R. Fernie[1*]

[1]Department of Lothar Willmitzer, Max-Planck-Institut für Molekular Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany

[2]The Otto Warburg Center for Biotechnology, Faculty of Agriculture, The Hebrew University of Jerusalem, PO Box 12, Rehovot 76100, Israel

Abbreviations: gas-chromatography (GC)/mass spectrometry (MS), γ- aminobutyric acid (GABA), 3-phosphoglycerate (3PGA), days after flowering (DAF), fresh weight (FW), *Solanum lycopersicum* (LYC), *S. pimpinellifolium* (PIM), *S. neorickii* (NEO), *S. chmielewskii* (CHM), *S. habrochaites* (HAB), *S. pennellii* (PEN)

4.1 Abstract

The domestication of the tomato *Solanum lycopersicum* and associated selective pressures eventually led to the large fruited varieties cultivated today. *Solanum lycopersicum* varieties are generally red-fruited but display considerable variance in fruit colour intensity shape and quality. The increase in productivity on cultivation is however somewhat offset by the narrowing of the crops genetic base which leads to increased susceptibility to biotic and abiotic stresses. Since *Solanum lycopersicum* can be easily crossed with its wild species relatives these exotic germplasm can provide a valuable source for the improvement of agriculturally important traits. Here we present a GC/MS based survey of the relative metabolic levels of leaves and fruit of *Solanum lycopersicum* and five wild species tomato that can be crossed with it (*S. pimpinellifolium*, *S. neorickii*, *S. chmielewskii*, *S. habrochaites* and *S. pennellii*). Changes in metabolite contents were identified in the wild species that are potentially important with respect to stress responses as well as in the levels of nutritionally important metabolites. The significance of these changes is discussed with respect to the use of the various wild species for metabolic engineering within wide breeding strategies.

4.2 Introduction

Fresh-market and processing varieties of tomato both originated from the cherry tomato *Solanum lycopersicum* "*cerasiforme*". The domestication of wild cherry types in Mexico spread to Europe and through the process of selection eventually led to large fruited varieties (Frary *et al*. 2000). Most *S. lycopersicum* varieties are red fruited and display considerable variance with respect to fruit colour intensity, shape, quality, growth habit and leaf morphology (Grandilio *et al*. 1996; Van Der Knaap *et al*. 2002; Frankel *et al*. 2003; Holtan and Hake 2003; Yates *et al*. 2004). However, it is important to note that the increases in productivity conferred by evolution of *S. lycopersicum* under domestication, like those of all crop species, are somewhat offset by the narrowing of the crops genetic basis (Zamir 2001). Whilst exotic germplasm resources often carry many agriculturally undesirable alleles, genetic studies are increasingly being employed in the identification of both undesirable (see for example (Tadmor *et al*. 2002) and agriculturally valuable traits (see for example (Rick 1974; Kovacs *et al*. 1998; Ray *et al*. 1999; Sebolt *et al*. 2000). Once identified, selected traits of exotic resources can be incorporated into commercial elite varieties via introgression breeding (Eshed 1995; Zamir 2001). In the case of tomato several examples of the utility of wild species alleles have been realised for some time with several traits being successfully integrated to date. Examples of such integrations include commercial tomato hybrids that contain different combinations of up to 15 wild disease resistance genes (Pan *et al*. 2000) and the introduction of important genes from *S. pennellii* which increase the fruit soluble solids content by 15-25 % (Fridman *et al*. 2000) and provitamin A (ß-carotene) level by more than 15-fold (Ronen *et al*. 2000).

S. lycopersicum can be easily crossed with a range of other *Solanum* species including *S. pimpinellifolium*, *S. neorickii*, *S. habrochaites*, *S. chmielewskii* and *S. pennellii* – sometimes collectively referred to as the *Lycopersicum* complex. These species display markedly different phenotypes from *S. lycopersicum* most notably with respect to their fruit. All wild species bear fruits that are dramatically smaller than those of the domesticated species and only a few of the species are red-fruited. Comparative physiology has been carried out on many of the wild species

with particular attention being paid to comparisons of plant performance under different environmental conditions such as aridity (Frankel *et al*. 2003), high salinity (Monforte *et al*. 1997a, 1997b; Foolad and Chen 1999) and chilling (Venema *et al*. 1999), as well as susceptibility to biotic stresses such as viral resistance (Legnani *et al*. 1996). Far less attention has, however, been paid to comparison at the biochemical level with comprehensive surveys essentially restricted to soluble solid contents (Fridman *et al*. 2000), sugars (Chetelat *et al*. 1995b) and glycoloids (Courtney and Lambeth 1977). Here we report the metabolic profiles of leaves and fruit of *Solanum lycopersicum* and five wild species tomato that can be crossed with this elite variety (*S. pimpinellifolium*, *S. neorickii*, *S. habrochaites*, *S. chmielewskii* and *S. pennellii*). To perform this study we grew all plants alongside each other under carefully controlled growth conditions and harvested leaf material six hours into the light period from the different species after 6 weeks of growth and fruits 45 days after flowering. Subsequently we characterized these tissues by utilizing a gas-chromatography (GC)/mass spectrometry (MS) protocol that we have recently established for tomato tissues (Roessner-Tunali *et al*. 2003a). This method allows the detection and robust quantification of over 90 metabolites of known chemical structure including organic acids, sugars, sugar alcohols, amino acids and a few soluble secondary metabolites (Fernie 2003; Stitt and Fernie 2003; Fernie *et al*. 2004). Differences in metabolite composition amongst these species will be discussed in the context of their utility for the selection of near isogenic introgression lines for breeding purposes.

4.3 Results

4.3.1 Experimental design

The plants were all grown simultaneously in a climate-controlled growth chamber under conditions that allowed normal fruit development in all species and which were close to optimal for *S. lycopersicum*. Material was harvested from equivalent fully expanded source leaves and from fruits 45 days after flowering at which stage the *S. lycopersicum* fruits were ripe. This time point was chosen for the fruit harvest on the basis of the fact that the metabolite content of *S. lycopersicum* is relatively constant in the days preceding and following it (F Carrari, AR Fernie;

personal communication), and we therefore reasoned that differences in developmental age between the genotypes would be minimised.

Time after flowering was preferred to other developmental parameters as the criterion for harvesting since it was difficult to choose a suitable parameter for the comparison of such morphologically diverse species. Differences in plant and fruit morphology are illustrated in Figures 3A and 3B, respectively (fruits were harvested 45 days after flowering). After 6 weeks and 45 days after flowering leaf and fruit samples were taken respectively. Tissues were dissected and rapidly snap-frozen in liquid nitrogen. The midrib was removed from leaf material prior to freezing whilst the fruit was skinned and seeds removed and only the pericarp tissue was taken for subsequent metabolite analysis.

Figure 3. **Leaf (A) and fruit (B) phenotypes of the *S. lycopersicum* complex.** (I) *S. chmielewskii*, (II) *S. habrochaites*, (III) *S. lycopersicum*, (IV) *S. pimpinellifolium*, (V) *S. neorickii*, and (VI) *S. pennellii*.

4.3.2 Starch and protein content in fruits of the wild species

As a first experiment we quantified the starch and importantly the protein content in samples taken from *S. lycopersicum* and the wild species. As would be expected given the loss of photosynthetic on the conversion of chloroplasts to chromoplasts during the ripening process (Obiadalla-Ali *et al*. 2004), the green fruited species contained significantly higher levels of starch than *S. lycopersicum* and *S. pimpinellifolium* although only marginally so in the case of *S. habrochaites*. The protein content was less variable across the species with only *S. pimpinellifolium* (lower) and *S. pennellii* (higher) exhibiting significantly different

levels of protein in the fruit relative to *S. lycopersicum* (Figure 4). Given that the fruit protein content was in approximately the same range between the species subsequent metabolite data will only be presented here per gram fresh weight (however values expressed per mg protein can be viewed on our webpage: www.mpimp-golm.mpg.de).

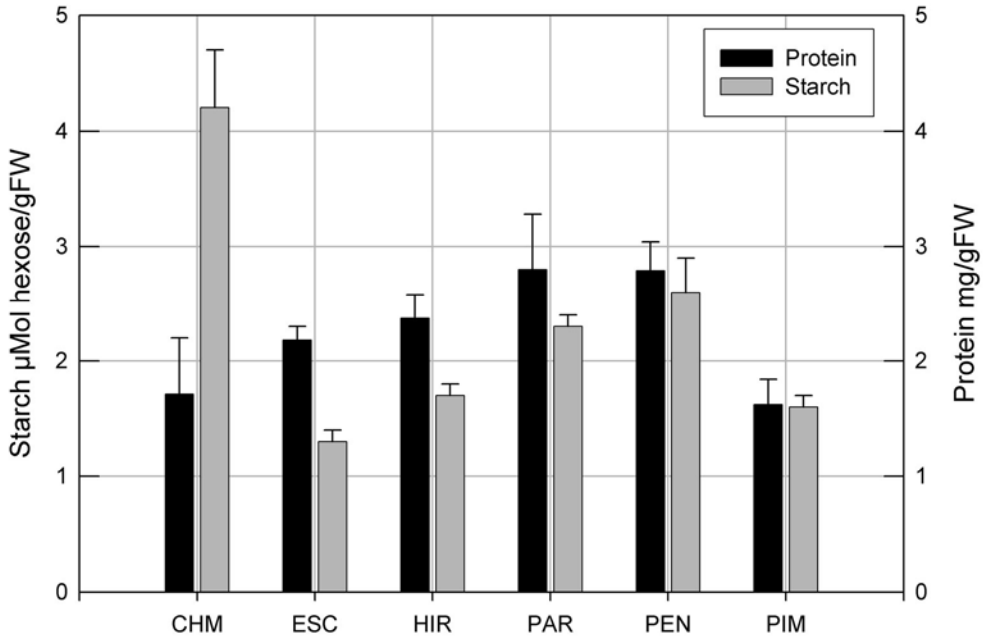

Figure 4. Protein and starch levels of fruits of the *S. lycopersicum* complex. Six independent fruit samples were measured. Fruits were harvested 45 DAF and 6h into the light. Protein values (black bars) are presented as mg protein g^{-1} FW. Starch values (grey bars) are presented as µmol hexose g^{-1} FW.

4.3.3 Leaf metabolite contents

4.3.3.1 Organic acid contents in wild species leaves

For the purposes of this study we generated calibration curves for every organic acid (and every metabolite for which we could obtain standards – a total of 64 metabolites). As would be expected the major organic acids in leaves of all species were malate, citrate, succinate, various forms of ascorbate and glycerate. Considerable lower levels were observed in the levels of aconitate (*S. habrochaites*, *S. pimpinellifolium* and *S. neorickii*), citrate (*S. pennellii*) and dehydroascorbate (*S. habrochaites*), galacturonate (*S. chmielewskii* and *S.*

pennellii) and isocitrate (*S. pennellii*) with respect to the levels determined for *S. lycopersicum*. Conversely, significantly greater amounts were observed for citramalate and glycerate (*S. chmielewskii* and *S. habrochaites*), glycolate (in all wild species with the exception of *S. pennellii*), 2-oxoglutarate (*S. habrochaites* and *S. neorickii*), maleate (*S. neorickii*), malate (*S. chmielewskii* and *S. habrochaites*), succinate (*S. chmielewskii* and *S. neorickii*) and threonate (*S. habrochaites*) with respect to the amounts determined for *S. lycopersicum*.

4.3.3.2 Sugar and sugar alcohol content in wild species leaves

The major sugars and sugar alcohols in *S. lycopersicum* and the wild species were glucose, fructose and sucrose and inositol as would be expected with the exception of the sucrose accumulator (*S. chmielewskii*; (Chetelat *et al*. 1995a)) the contents of glucose and fructose are much higher than those of sucrose most probably due to the high invertase activity present in tomato leaves. Significant decreases were observed in the levels of arabinose (all wild species with the exception of *S. pimpinellifolium*), fructose (dramatically decreased in all wild species with the exception of *S. chmielewskii*), fucose (below the level of detection in *S. pimpinellifolium* and *S. pennellii*), galactose (below the level of detection in *S. neorickii* and *S. pennellii*), glucose (*S. chmielewskii* and *S. habrochaites*), maltose (*S. pennellii*), mannose (in all wild species), mannitol, rhamnose and ribose(*S. neorickii* and *S. pennellii*), xylose (*S. chmielewskii*, *S. habrochaites* and *S. pennellii*), with respect to the levels determined for *S. lycopersicum*. Conversely, significant increases were observed in the levels of ribose (*S. habrochaites*) and sucrose (*S. chmielewskii*) with respect to the levels determined for *S. lycopersicum* whereas isomaltose was only above the detection limit in *S. neorickii* and *S. pimpinellifolium* and trehalose was only detected in *S. chmielewskii* and *S. habrochaites*. The levels of galactitol, gentiobiose (an oligosaccharin recently identified in ripening tomato fruit; (Dumville and Fry 2003)), glycerol, inositol, maltitol, melezitose and sorbitol were, however, remarkably similar across the species.

4.3.3.3 Amino acid contents in wild species leaves

Leaves from *S. lycopersicum* are characterised by relatively high contents of aspartate, glutamine, glutamate, oxoproline, proline, threonine and serine. Significantly lower levels of the following metabolites were observed in leaves from the wild species: γ- aminobutyric acid (GABA), glutamate and lysine (*S. pennellii*). In contrast, the levels of the following metabolites were higher in leaves from the wild species: alanine and arginine (*S. chmielewskii*, *S. habrochaites* and *S. neorickii*), asparagine (*S. chmielewskii* and *S. neorickii*), aspartate (*S. habrochaites*), glutamate, glycine, lysine and ornithine (*S. habrochaites*), glutamine and isoleucine (*S. chmielewskii*), methionine (*S. habrochaites* and *S. neorickii*), proline (*S. pimpinellifolium*), ß-alanine, hydroxyproline and valine (*S. chmielewskii* and *S. habrochaites*), serine (*S. neorickii*). The differences observed in absolute values did, however, not always reflect important metabolite ratios. For example asparagine: aspartate and glycine: serine ratios are lower in all wild species. Differences were, however, observed in ratios of glutamine: glutamate, where ratios are lower in *S. lycopersicum* (1:0.3), *S. pennellii* (1:0.6), *S. pimpinellifolium* (1:0.8) and *S. habrochaites* (1:0.7), but elevated in *S. chmielewskii* (1:2.7) and *S. neorickii* (1:1.5). Despite the fact that the levels of the majority of amino acids were higher in the wild species levels of leucine, 5-oxoproline, phenylalanine, threonine, tryptophan and tyrosine were essentially the same in all species.

	ESC		SE	CHM		SE	HIR		SE	PAR		SE	PEN		SE	PIM		SE
									µmol /g FW									
alanine	1,26	±	0,00	**3,28**	±	**0,00**	**4,27**	±	**0,00**	**2,92**	±	**0,00**	1,53	±	0,00	2,31	±	0,00
arginine	0,24	±	0,00	**0,61**	±	**0,00**	**0,68**	±	**0,00**	**0,79**	±	**0,00**	0,25	±	0,00	0,39	±	0,00
asparagine	0,23	±	0,00	**0,80**	±	**0,00**	1,14	±	0,00	**0,73**	±	**0,00**	0,18	±	0,00	0,25	±	0,00
aspartate	6,36	±	0,01	8,77	±	0,00	**15,55**	±	**0,00**	11,17	±	0,00	2,63	±	0,00	8,66	±	0,00
ß-alanine	0,11	±	0,00	**0,30**	±	**0,00**	**0,38**	±	**0,00**	0,15	±	0,00	0,19	±	0,00	0,12	±	0,00
GABA	1,45	±	0,00	1,29	±	0,00	1,21	±	0,00	1,38	±	0,00	**0,75**	±	**0,00**	1,89	±	0,00
glutamate	11,51	±	0,01	13,41	±	0,00	**19,67**	±	**0,00**	17,45	±	0,00	**2,26**	±	**0,00**	16,11	±	0,00
glutamine	3,39	±	0,01	**35,99**	±	**0,00**	12,91	±	0,00	26,78	±	0,00	1,46	±	0,00	**13,08**	±	**0,00**
glycine	1,89	±	0,00	4,03	±	0,00	**5,67**	±	**0,00**	3,50	±	0,00	1,94	±	0,00	3,12	±	0,00
isoleucine	0,93	±	0,00	**1,14**	±	**0,00**	1,23	±	0,00	0,97	±	0,00	0,46	±	0,00	0,88	±	0,00
leucine	0,53	±	0,00	0,44	±	0,00	0,78	±	0,00	0,46	±	0,00	0,17	±	0,00	0,31	±	0,00
lysine	0,06	±	0,00	0,12	±	0,00	**0,09**	±	**0,00**	0,10	±	0,00	**0,00**	±	**0,00**	0,07	±	0,00
methionine	0,05	±	0,00	0,06	±	0,00	**0,10**	±	**0,00**	**0,13**	±	**0,00**	0,16	±	0,00	0,07	±	0,00
ornithine	0,31	±	0,00	0,57	±	0,00	**0,97**	±	**0,00**	0,90	±	0,00	0,49	±	0,00	**0,48**	±	**0,00**
5-oxoproline	6,33	±	0,01	6,72	±	0,00	9,17	±	0,00	8,92	±	0,00	4,95	±	0,00	7,41	±	0,00
phenylalanine	0,74	±	0,00	1,14	±	0,00	1,29	±	0,00	1,34	±	0,00	0,45	±	0,00	0,90	±	0,00
proline	18,59	±	0,04	17,37	±	0,01	**31,10**	±	**0,01**	22,35	±	0,01	9,14	±	0,01	**29,80**	±	**0,00**
serine	4,88	±	0,01	10,05	±	0,00	**12,84**	±	**0,00**	**12,42**	±	**0,00**	3,60	±	0,00	6,86	±	0,00
hydroxyproline	1,58	±	0,00	**3,26**	±	**0,00**	**4,43**	±	**0,00**	3,18	±	0,00	1,00	±	0,00	2,26	±	0,00
threonine	14,97	±	0,03	32,09	±	0,00	28,84	±	0,00	25,72	±	0,00	9,47	±	0,00	18,60	±	0,00
tryptophan	2,80	±	0,01	0,47	±	0,00	0,70	±	0,00	1,71	±	0,00		nd		1,41	±	0,00
tyrosine	0,15	±	0,00	0,10	±	0,00	0,17	±	0,00	0,11	±	0,00	0,03	±	0,00	0,13	±	0,00
valine	1,10	±	0,00	**2,59**	±	**0,00**	**3,12**	±	**0,00**	1,47	±	0,00	1,06	±	0,00	1,27	±	0,00
aconitate	0,02	±	0,00	0,05	±	0,00	**0,00**	±	**0,00**	**0,05**	±	**0,00**	0,04	±	0,00		nd	
citramalate	0,06	±	0,00	**0,19**	±	**0,00**	**0,36**	±	**0,00**	0,06	±	0,00	0,06	±	0,00	0,08	±	0,00
citrate	4,06	±	0,01	4,14	±	0,00	2,63	±	0,00	5,55	±	0,00	**1,24**	±	**0,00**	4,12	±	0,00
dehydroascorbate	2,71	±	0,00	3,62	±	0,00	**2,12**	±	**0,00**	2,96	±	0,00	3,77	±	0,00	3,10	±	0,00
fumarate	0,14	±	0,00	0,44	±	0,00	0,33	±	0,00	0,40	±	0,00	0,19	±	0,00	0,39	±	0,00
galacturonate	0,76	±	0,00	**0,27**	±	**0,00**	0,58	±	0,00	0,36	±	0,00	**0,23**	±	**0,00**	**0,44**	±	**0,00**
glycerate	1,77	±	0,00	**4,29**	±	**0,00**	**4,32**	±	**0,00**	3,46	±	0,00	1,28	±	0,00	3,99	±	0,00
glycolate	0,07	±	0,00	**0,14**	±	**0,00**	**0,25**	±	**0,00**	**0,13**	±	**0,00**	0,05	±	0,00	**0,14**	±	**0,00**
isocitrate	0,06	±	0,00	0,07	±	0,00	0,09	±	0,00	0,07	±	0,00	**0,02**	±	**0,00**	0,05	±	0,00
2-oxoglutarate	0,30	±	0,00	0,34	±	0,00	**0,87**	±	**0,00**	**0,66**	±	**0,00**	0,46	±	0,00	0,34	±	0,00
L-ascorbate	1,50	±	0,00	1,81	±	0,00	1,46	±	0,00	1,62	±	0,00	1,35	±	0,00	1,58	±	0,00
maleate	0,25	±	0,00	0,00	±	0,00	0,13	±	0,00	**0,97**	±	**0,00**	0,13	±	0,00	0,28	±	0,00
malate	28,93	±	0,04	**33,92**	±	**0,01**	**39,10**	±	**0,01**	39,37	±	0,01	17,17	±	0,00	30,02	±	0,00
succinate	2,16	±	0,00	**3,28**	±	**0,00**	3,28	±	0,00	**3,57**	±	**0,00**	2,60	±	0,00	3,89	±	0,00
threonate	1,35	±	0,00	1,73	±	0,00	**2,42**	±	**0,00**	1,27	±	0,00	0,43	±	0,00	**2,25**	±	**0,00**
arabinose	0,11	±	0,00	**0,07**	±	**0,00**	**0,05**	±	**0,00**	**0,05**	±	**0,00**	**0,05**	±	**0,00**	**0,08**	±	**0,00**
fructose	26,20	±	0,04	29,14	±	0,01	**0,15**	±	**0,00**	**0,07**	±	**0,01**	**0,05**	±	**0,00**	**0,21**	±	**0,01**
fucose	0,28	±	0,00	0,40	±	0,00	0,42	±	0,00	0,32	±	0,00		nd			nd	
galactose	0,17	±	0,00	0,18	±	0,00	**0,13**	±	**0,00**	**0,10**	±	**0,00**	**0,10**	±	**0,00**	**0,17**	±	**0,00**
gentiobiose	0,02	±	0,00	0,01	±	0,00	0,03	±	0,00	0,02	±	0,00	0,01	±	0,00	0,02	±	0,00
glucose	59,97	±	0,13	50,65	±	0,03	**10,98**	±	**0,04**	**14,70**	±	**0,03**	**14,23**	±	**0,03**	**38,91**	±	**0,02**
glycerol	0,32	±	0,00	1,04	±	0,00	1,15	±	0,00	0,27	±	0,00	0,22	±	0,00	0,24	±	0,00
inositol	15,40	±	0,03	**16,17**	±	**0,00**	**15,29**	±	**0,00**	15,15	±	0,00	11,21	±	0,00	15,76	±	0,00
isomaltose		nd			nd			nd		0,01	±	0,00		nd		0,01	±	0,00
maltitol	0,01	±	0,00	0,01	±	0,00	0,01	±	0,00	0,01	±	0,00		nd			nd	
maltose	0,07	±	0,00	0,04	±	0,00	0,05	±	0,00	0,03	±	0,00	**0,02**	±	**0,00**	0,04	±	0,00
mannitol	0,17	±	0,00	0,09	±	0,00	0,17	±	0,00	**0,09**	±	**0,00**	**0,09**	±	**0,00**	0,14	±	0,00
mannose	0,16	±	0,00	**0,05**	±	**0,00**	**0,02**	±	**0,00**	**0,03**	±	**0,00**	**0,01**	±	**0,00**	0,06	±	0,00
melezitose	0,01	±	0,00	0,01	±	0,00	0,01	±	0,00	0,01	±	0,00	nd	±	nd	0,01	±	0,00
rhamnose	0,04	±	0,00	0,04	±	0,00	0,05	±	0,00	**0,03**	±	**0,00**	**0,02**	±	**0,00**	0,03	±	0,00
ribose	0,04	±	0,00	0,03	±	0,00	**0,09**	±	**0,00**	0,02	±	0,00	**0,20**	±	**0,00**	**0,01**	±	**0,00**
sorbitol/galactitol	0,04	±	0,00	0,05	±	0,00	0,05	±	0,00	0,03	±	0,00	0,04	±	0,00	0,04	±	0,00
sucrose	16,70	±	0,04	**42,06**	±	**0,01**	13,72	±	0,00	22,23	±	0,01	14,44	±	0,01	25,94	±	0,01
trehalose		nd		0,01	±	0,00	0,01	±	0,00		nd			nd			nd	
xylose	0,28	±	0,00	**0,06**	±	**0,00**	**0,08**	±	**0,00**	0,16	±	0,00	**0,09**	±	**0,00**	0,19	±	0,00

	LYC	SE	CHM	SE	HAB	SE	NEO	SE	PEN	SE	PIM	SE
	Relative change with respect to LYC											
fructose-6-P	1,00 ±	0,18	1,04 ±	0,56	**4,15** ±	**0,12**	0,67 ±	0,19	0,55 ±	0,18	0,38 ±	0,19
glucose-6-P	1,00 ±	0,20	1,02 ±	0,55	**4,34** ±	**0,14**	0,68 ±	0,19	0,57 ±	0,18	0,37 ±	0,21
glycerol-1-P	1,00 ±	0,20	4,42 ±	0,56	**9,88** ±	**0,17**	2,18 ±	0,30	1,18 ±	0,21	1,46 ±	0,28
inositol-1-P	1,00 ±	0,18	1,01 ±	0,45	**2,94** ±	**0,13**	0,93 ±	0,16	0,41 ±	0,31	nd	
3PGA	1,00 ±	0,16	1,30 ±	0,10	1,68 ±	0,19	**1,99** ±	**0,21**	**2,45** ±	**0,10**	**2,45** ±	**0,11**
phosphate	1,00 ±	0,24	6,19 ±	0,21	5,35 ±	0,11	0,85 ±	0,19	1,45 ±	0,18	0,42 ±	0,40
FA 16:0	1,00 ±	0,05	1,05 ±	0,08	1,13 ±	0,06	1,11 ±	0,08	2,26 ±	0,56	1,04 ±	0,06
FA 18:0	1,00 ±	0,05	1,08 ±	0,09	1,22 ±	0,07	1,19 ±	0,08	1,03 ±	0,13	1,13 ±	0,07
quinate	1,00 ±	0,16	0,97 ±	0,05	1,22 ±	0,04	1,27 ±	0,12	0,98 ±	0,09	1,19 ±	0,08
shikimate	1,00 ±	0,15	**1,53** ±	**0,22**	**2,80** ±	**0,09**	**1,91** ±	**0,21**	**2,53** ±	**0,09**	**2,42** ±	**0,12**
putrescine	1,00 ±	0,09	0,87 ±	0,07	0,93 ±	0,11	**1,71** ±	**0,10**	1,38 ±	0,21	1,06 ±	0,06
uracil	1,00 ±	0,19	1,79 ±	0,10	1,86 ±	0,22	nd		1,16 ±	0,36	1,41 ±	0,13

Table 3 Metabolite composition in leaves from species of the *S. lycopersicum* complex

Single leaf samples of six plants were measured. Leaves were harvested 6 h into the light period from fully-expanded mature leaves of 6-week-old plants. Values are presented as the mean ±SE of six independent biological determinations. Those metabolites that are significantly different to *S. lycopersicum* are set in bold type. nd indicates metabolites were not detected.

4.3.3.4 Miscellaneous metabolite contents in wild species leaves

We additionally determined the relative levels of a further twelve metabolites including phosphorylated intermediates, fatty acids, putrescine and uracil. The hexose glucose 6-phosphate and fructose 6-phosphate were marginally (yet not significantly), lower in the wild species with the exception of *S. habrochaites* in which they were three- to four-fold higher. Similar trends were also observed in glycerol 1-phosphate and inositol 1-phosphate. The level of 3-phosphoglycerate (3PGA) was significantly higher in *S. neorickii*, *S. pennellii* and *S. pimpinellifolium*, whilst the level of putrescine was significantly higher in *S. neorickii* than in *S. lycopersicum*. Significantly higher levels of shikimate could be observed throughout the wild species. Total phosphate content, fatty acids 16:0 and 18:0, quinate and uracil levels were constant across the species.

4.3.4 Fruit metabolite contents

4.3.4.1 Organic acid contents in wild species fruits

The major organic acids of the fruit were somewhat different from those found in the leaf since although citrate and malate were present at high levels; the level of succinate was lower in the fruit. Conversely the levels of galacturonic acid,

gluconate and isocitrate are considerably higher in the fruit than in the leaf. Furthermore, several metabolites such as chlorogenate and nicotinate were only detectable in the fruit.

Significantly lower levels of the following metabolites were observed in fruits from the wild species when compared on a per gram fresh weight basis to *S. lycopersicum* (data are additionally expressed per mg protein on our webpage: www.mpimp-golm.mpg.de): dehydroascorbate (*S. habrochaites*), gluconate (*S. habrochaites*, *S. pennellii* and *S. pimpinellifolium*), L-ascorbate (*S. habrochaites*), maleate (*S. pennellii*), succinate (*S. chmielewskii*, *S. habrochaites* and *S. pimpinellifolium*) and threonate (*S. habrochaites*). In contrast, higher levels of the following metabolites were observed in fruits from the wild species than in *S. lycopersicum*: aconitate (*S. chmielewskii*, *S. habrochaites* and *S. pimpinellifolium*), citramalate and threonate (*S. neorickii*, *S. pennellii* and *S. pimpinellifolium*), citrate and dehydroascorbate (all wild species with the exception of *S. habrochaites*), chlorogenate (massively higher in all lines), fumarate and salicylate (*S. chmielewskii*, *S. neorickii* and *S. pennellii*.), galacturonate (*S. neorickii* and *S. pimpinellifolium*), gluconate (*S. neorickii*), glycerate (*S. chmielewskii* and *S. neorickii*), isocitrate (all species), 2-oxoglutarate (*S. pennellii*), L-ascorbate (*S. pennellii* and *S. neorickii*), malate (massively higher in all species with the exception of *S. pennellii*), nicotinate (dramatically higher in all wild species), shikimate (all species with the exception of *S. pimpinellifolium*) and succinate (*S. pennellii*).

4.3.4.2 Sugar and sugar alcohol content in wild fruits

The major sugar and sugar alcohol content of the fruits is very similar to that observed in the leaves with glucose, fructose, sucrose and inositol again being the major constituents. However, several metabolites such as erythritol and raffinose were observed in the fruit despite being present below the level of detection (if at all) in the tomato leaf. Significantly lower levels of the following metabolites were observed in fruits from the wild species: arabinose (*S. habrochaites*, *S. neorickii* and *S. pennellii*), fructose and glucose (all species with the exception of *S. pimpinellifolium*), fucose (*S. neorickii* and *S. pimpinellifolium*), galactose (*S. pennellii*), gentiobiose (not detected in *S. chmielewskii* or *S. neorickii*), mannose

(*S. habrochaites, S. neorickii* and *S. pimpinellifolium*), raffinose (all with the exception of *S. pimpinellifolium* which increases massively), ribose (all species) and trehalose (*S. habrochaites* and *S. pennellii*). Conversely, higher levels of the following metabolites were observed in fruits from the wild species: fructose and glucose (*S. pimpinellifolium*), galactose (*S. neorickii* and *S. pimpinellifolium*), glycerol and rhamnose (all wild species), inositol, maltose and xylose (all wild species with the exception of *S. pimpinellifolium*), isomaltose (*S. chmielewskii, S. pennellii* and *S. pimpinellifolium*), mannose (*S. pimpinellifolium*) sucrose (*S. chmielewskii, S. habrochaites* and *S. neorickii*). The levels of erythritol, maltitol, mannitol, sorbitol and galactinol were observed to be invariant across the species.

4.3.4.3 Amino acid contents in wild species fruits

The pattern of amino acids of the fruits was largely similar to that of the leaves with the same amino acids being major constituents of each tissue. However, the levels of GABA and the derivatives of proline were notably higher in the fruit than in the leaves. In sharp contrast to the situation observed in the leaves, the majority of amino acids were found at lower quantity in the wild species than in *S. lycopersicum* with significant differences being observed in the following metabolites: arginine, asparagine, isoleucine, leucine and tyrosine (in all wild species with the exception of *S. habrochaites*), aspartate (*S. habrochaites, S. pennellii* and *S. pimpinellifolium*), ß-alanine, GABA and methionine (in all wild species with the exception of *S. pennellii*), glutamine, glutamate, lysine, phenylalanine, serine, hydroxyproline and threonine (in all wild species, dramatically so in the cases of glutamine and glutamate), glycine (all wild species with the exception of *S. neorickii*), ornithine and 5-oxoproline (all wild species with the exception of *S. habrochaites*), proline (*S. pennellii* and *S. pimpinellifolium*), tyrosine (*S. neorickii, S. pennellii* and *S. pimpinellifolium*) and valine (*S. chmielewskii, S. neorickii* and *S. pimpinellifolium*). The exception to this trend was the fact that alanine, tryptophan and valine were all present at significantly higher levels in *S. habrochaites* than in *S. lycopersicum*. Studying important metabolite ratios of the fruits revealed differences across the species. In the case of the asparagine: aspartate *S. chmielewskii, S. neorickii* and *S. pimpinellifolium* display much lower ratios, whereas *S. habrochaites* and *S. pennellii* display much higher

ratios than those observed for *S. lycopersicum*. A similar picture emerges in the case of glutamine: glutamate with the exception that the ratio in *S. pennellii* is not significantly different from *S. lycopersicum*. Conversely, the glycine: serine ratio was much lower in all wild species with the exception of *S. habrochaites.*

4.3.4.4 Miscellaneous metabolite contents in wild species fruits

We additionally determined the relative levels of a further eighteen metabolites including the same miscellaneous metabolites detected in the leaves in addition to α-tocopherol, spermidine, tyramine, fatty acid 18:2 and dopamine. The hexose phosphates glucose 6-phosphate and fructose 6-phosphate were significantly lower in *S. pennellii*, and glycerol 1-phosphate was lower in *S. pennellii* and *S. pimpinellifolium* than in the cultivated tomato but glycerol 1-phosphate levels were significantly higher in *S. habrochaites* (than in the cultivated tomato). *S. chmielewskii*, *S. habrochaites* and *S. neorickii* all displayed higher levels of inositol 1-phosphate than *S. lycopersicum*. The contents of free phosphate and fatty acid 16:0, 18:0, and 18:2 were higher in all of the wild species with the exception of *S. pimpinellifolium*. In contrast, the levels of α-tocopherol were lower in all the wild species, but only marginally so in the case of *S. chmielewskii.* Changes were also observed in dopamine (which was higher in *S. chmielewskii* and *S. neorickii* but was not detected in *S. pennellii* and *S. pimpinellifolium*), putrescine (which was lower in *S. chmielewskii*, *S. neorickii* and *S. pennellii*), tyramine (massively increased in *S. pimpinellifolium*) and uracil (which was higher in *S. chmielewskii*, *S. neorickii* and *S. pennellii*).

	ESC	SE	CHM	SE	HIR	SE	PAR	SE	PEN	SE	PIM	SE
						µmol /g FW						
alanine	0,67	± 0,01	0,67	± 0,00	**1,01**	± **0,00**	0,70	± 0,00	0,60	± 0,00	0,75	± 0,00
arginine	1,06	± 0,00	**0,12**	± **0,00**	0,82	± 0,00	**0,23**	± **0,00**	**0,18**	± **0,00**	**0,05**	± **0,00**
asparagine	16,50	± 0,12	**0,53**	± **0,04**	8,11	± 0,05	**1,25**	± **0,04**	**1,74**	± **0,07**	**0,21**	± **0,06**
aspartate	39,16	± 0,11	31,75	± 0,03	**6,57**	± **0,02**	**7,01**	± **0,02**	**1,86**	± **0,05**	**3,99**	± **0,02**
ß-alanine	1,01	± 0,01	**0,23**	± **0,00**	**0,25**	± **0,00**	**0,18**	± **0,00**	1,08	± 0,00	**0,13**	± **0,00**
GABA	34,77	± 0,40	**2,11**	± **0,05**	**0,60**	± **0,06**	**0,69**	± **0,04**	**151,77**	± **0,09**	**0,92**	± **0,06**
glutamate	62,50	± 0,13	**4,80**	± **0,03**	**6,82**	± **0,03**	**4,72**	± **0,02**	**1,13**	± **0,02**	**7,49**	± **0,02**
glutamine	75,35	± 0,23	**0,78**	± **0,06**	**17,74**	± **0,09**	**2,13**	± **0,09**	**1,40**	± **0,12**	**0,40**	± **0,01**
glycine	1,74	± 0,02	1,28	± 0,01	**0,82**	± **0,00**	1,04	± 0,00	**0,73**	± **0,00**	**0,73**	± **0,00**
isoleucine	4,67	± 0,04	**0,42**	± **0,00**	4,11	± 0,01	**0,49**	± **0,00**	**0,89**	± **0,01**	**0,56**	± **0,01**
leucine	2,65	± 0,02	**0,72**	± **0,00**	2,19	± 0,00	**1,07**	± **0,00**	**0,42**	± **0,00**	**0,26**	± **0,00**
lysine	1,98	± 0,01	**0,11**	± **0,00**	**0,53**	± **0,00**	**0,20**	± **0,00**	**0,13**	± **0,00**	**0,10**	± **0,00**
methionine	0,63	± 0,00	**0,12**	± **0,00**	**0,23**	± **0,00**	nd		0,32	± 0,00	**0,10**	± **0,00**
ornithine	0,09	0,00	nd		0,08	± 0,00	**0,02**	± **0,00**	**0,02**	± **0,00**	**0,01**	± **0,00**
5-oxoproline	37,22	± 0,07	2,97	± 0,06	33,10	± 0,02	11,18	± 0,02	6,66	± 0,02	11,79	± 0,01
phenylalanine	10,45	± 0,09	**0,21**	± **0,01**	**0,58**	± **0,01**	**0,15**	± **0,01**	**0,19**	± **0,01**	**1,42**	± **0,02**
proline	7,48	± 0,11	6,53	± 0,04	10,86	± 0,05	4,35	± 0,04	**0,33**	± **0,03**	3,81	± 0,02
serine	7,31	± 0,06	**3,11**	± **0,00**	**3,35**	± **0,01**	**2,07**	± **0,01**	**0,78**	± **0,01**	**1,72**	± **0,01**
hydroxyproline	20,83	± 0,07	**1,62**	± **0,01**	2,95	± 0,02	**1,71**	± **0,01**	**0,91**	± **0,01**	**0,88**	± **0,01**
threonine	26,42	± 0,16	**5,05**	± **0,01**	**12,03**	± **0,02**	**2,86**	± **0,01**	**3,99**	± **0,05**	**3,12**	± **0,02**
tryptophan	9,45	± 0,05	5,68	± 0,01	**110,24**	± **0,02**	**0,48**	± **0,01**	**0,84**	± **0,02**	**1,19**	± **0,01**
tyrosine	0,76	± 0,01	**0,16**	± **0,00**	2,80	± 0,00	**0,12**	± **0,00**	**0,21**	± **0,00**	**0,06**	± **0,00**
valine	1,51	± 0,02	**0,73**	± **0,00**	**2,64**	± **0,00**	**0,54**	± **0,00**	1,58	± 0,00	**0,51**	± **0,00**
aconitate	0,09	± 0,00	**0,28**	± **0,00**	**0,24**	± **0,00**	**0,17**	± **0,00**	0,12	± 0,00	0,11	± 0,00
citramalate	0,05	± 0,00	0,08	± 0,00	0,04	± 0,00	**0,08**	± **0,00**	**0,02**	± **0,00**	**0,09**	± **0,00**
citrate	47,93	± 0,22	**83,05**	± **0,01**	**60,65**	± **0,02**	**85,83**	± **0,02**	**68,46**	± **0,01**	55,20	± 0,02
dehydroascorbate	2,36	± 0,01	**4,26**	± **0,00**	**1,39**	± **0,00**	**4,74**	± **0,00**	**6,06**	± **0,00**	**5,80**	± **0,00**
fumarate	0,09	± 0,00	**0,23**	± **0,00**	0,13	± 0,00	**0,28**	± **0,00**	**0,22**	± **0,00**	0,04	± 0,00
galacturonate	4,45	± 0,06	8,70	± 0,02	2,93	± 0,00	**9,14**	± **0,01**	2,25	± 0,01	**12,68**	± **0,00**
gluconate	6,18	± 0,03	7,80	± 0,00	**3,48**	± **0,01**	**18,26**	± **0,00**	**1,66**	± **0,00**	**2,98**	± **0,01**
glycerate	0,08	± 0,00	**0,14**	± **0,00**	0,08	± 0,00	**0,21**	± **0,00**	0,07	± 0,00	0,13	± 0,00
glycolate	0,10	± 0,00	**0,17**	± **0,00**	0,13	± 0,00	**0,18**	± **0,00**	0,15	± 0,00	0,08	± 0,00
isocitrate	2,92	± 0,03	**6,62**	± **0,00**	**6,82**	± **0,00**	**5,44**	± **0,00**	**4,23**	± **0,00**	**4,33**	± **0,00**
2-oxoglutarate	0,33	± 0,00	0,53	± 0,00	0,26	± 0,00	0,53	± 0,00	**1,07**	± **0,00**	0,32	± 0,00
L-ascorbate	1,53	± 0,01	3,16	± 0,00	**0,82**	± **0,00**	**6,68**	± **0,00**	**3,48**	± **0,00**	2,38	± 0,00
maleate	0,19	± 0,00	0,16	± 0,00	0,11	± 0,00	0,13	± 0,00	0,14	± 0,00	**0,08**	± **0,00**
malate	12,59	± 0,11	**263,43**	± **0,01**	**96,91**	± **0,02**	**412,58**	± **0,01**	**310,12**	± **0,01**	8,27	± 0,02
saccharate	4,21	± 0,05	3,26	± 0,01	**7,73**	± **0,00**	**6,83**	± **0,00**	2,83	± 0,01	2,45	± 0,00
salicylate	0,01	± 0,00	**0,02**	± **0,00**	0,02	± 0,00	**0,02**	± **0,00**	**0,02**	± **0,00**	0,02	± 0,00
succinate	0,99	± 0,01	**0,75**	± **0,00**	**0,60**	± **0,00**	0,88	± 0,00	**1,54**	± **0,00**	**0,66**	± **0,00**
threonate	0,09	± 0,00	0,16	± 0,00	**0,05**	± **0,00**	**0,21**	± **0,00**	**0,18**	± **0,00**	**0,26**	± **0,00**
arabinose	0,22	± 0,00	0,18	± 0,00	**0,11**	± **0,00**	**0,13**	± **0,00**	**0,08**	± **0,00**	0,19	± 0,00
erythritol	0,20	± 0,00	0,26	± 0,00	0,18	± 0,00	0,22	± 0,00	0,21	± 0,00	0,20	± 0,00
fructose	116,61	± 0,93	**26,06**	± **0,06**	**25,47**	± **0,09**	**27,83**	± **0,09**	**37,84**	± **0,13**	**169,82**	± **0,03**
fucose	0,12	± 0,00	0,11	± 0,00	0,07	± 0,00	**0,07**	± **0,00**	0,11	± 0,00	**0,07**	± **0,00**
galactose	5,23	± 0,08	4,94	± 0,00	5,47	± 0,01	**8,30**	± **0,01**	**2,54**	± **0,00**	**15,95**	± **0,01**
gentiobiose	0,12	0,00	nd		0,05	± 0,00	nd		0,07	± 0,00	0,05	± 0,00
glucose	531,73	± 1,19	**62,54**	± **0,09**	**46,75**	± **0,19**	**93,75**	± **0,11**	477,22	± 0,12	**703,80**	± **0,04**
glycerol	1,18	± 0,01	**2,93**	± **0,00**	**1,67**	± **0,00**	**3,02**	± **0,00**	7,82	± 0,00	**1,64**	± **0,00**
inositol	9,71	± 0,07	**4,28**	± **0,00**	**32,01**	± **0,00**	**13,27**	± **0,00**	**68,57**	± **0,00**	8,49	± 0,00
isomaltose	0,02	± 0,00	**0,05**	± **0,00**	0,01	± 0,00	0,02	± 0,00	**0,03**	± **0,00**	**0,04**	± **0,00**
maltitol	0,09	± 0,00	0,09	± 0,00	0,07	± 0,00	0,04	± 0,00	0,04	± 0,00	0,13	± 0,00
maltose	0,27	± 0,00	**0,43**	± **0,00**	**0,96**	± **0,00**	**0,49**	± **0,00**	**0,53**	± **0,00**	0,27	± 0,00
mannitol	0,43	± 0,00	0,48	± 0,00	0,42	± 0,00	0,48	± 0,00	0,40	± 0,00	0,35	± 0,00
mannose	2,32	± 0,01	2,32	± 0,00	**1,17**	± **0,00**	**1,70**	± **0,00**	1,92	± 0,00	**12,01**	± **0,00**
raffinose	0,11	± 0,00	0,09	± 0,00	0,07	± 0,00	0,04	± 0,00	**0,01**	± **0,00**	**0,48**	± **0,00**
rhamnose	0,04	± 0,00	**0,07**	± **0,00**	**0,17**	± **0,00**	**0,06**	± **0,00**	**0,06**	± **0,00**	**0,05**	± **0,00**
ribose	0,24	± 0,00	0,25	± 0,00	**0,09**	± **0,00**	**0,43**	± **0,00**	**0,10**	± **0,00**	**0,16**	± **0,00**
sorbitol/galactitol	0,10	± 0,00	0,03	± 0,00	0,03	± 0,00	0,04	± 0,00	0,04	± 0,00	0,02	± 0,00
sucrose	39,39	± 0,34	**831,25**	± **0,08**	**146,78**	± **0,04**	**988,22**	± **0,05**	35,20	± 0,02	42,20	± 0,06
trehalose	0,03	± 0,00	0,03	± 0,00	**0,01**	± **0,00**	0,03	± 0,00	**0,01**	± **0,00**	0,03	± 0,00
xylose	0,53	± 0,00	**1,47**	± **0,00**	**1,40**	± **0,00**	**0,64**	± **0,00**	**1,87**	± **0,00**	0,52	± 0,00
a-tocopherol	20,61	± 0,11	16,18	± 0,00	**2,31**	± **0,04**	**11,61**	± **0,01**	**7,13**	± **0,02**	**10,25**	± **0,01**

	ESC	SE	CHM	SE	HIR	SE	PAR	SE	PEN	SE	PIM	SE
	Relative change with respect to LYC											
fructose-6-P	1,0	± 0,15	1,13	± 0,18	1,01	± 0,13	1,37	± 0,08	0,67	± 0,28	**0,47**	± **0,09**
glucose-6-P	1,0	± 0,18	1,11	± 0,21	0,88	± 0,13	1,47	± 0,09	0,65	± 0,27	**0,48**	± **0,09**
glycerol-1-P	1,0	± 0,11	0,95	± 0,14	**1,67**	± **0,14**	1,23	± 0,10	**0,56**	± **0,14**	**0,65**	± **0,04**
inositol-1-P	1,0	± 0,14	**2,49**	± **0,12**	**1,92**	± **0,16**	**3,00**	± **0,08**	1,10	± 0,17	0,85	± 0,07
3PGA	1,0	± 0,28	1,12	± 0,30	0,91	± 0,26	2,10	± 0,24	1,24	± 0,07	**0,21**	± **0,60**
phosphate	1,0	± 0,08	2,79	± 0,26	**1,53**	± **0,12**	**1,64**	± **0,05**	**1,40**	± **0,07**	0,75	± 0,15
FA 16:0	1,0	± 0,05	**2,05**	± **0,05**	**1,34**	± **0,09**	**1,53**	± **0,05**	**1,28**	± **0,04**	1,04	± 0,08
FA 18:0	1,0	± 0,06	**2,05**	± **0,05**	**1,46**	± **0,06**	**1,66**	± **0,04**	**1,35**	± **0,06**	0,96	± 0,04
FA 18:2	1,0	± 0,08	**5,44**	± **0,25**	**2,22**	± **0,19**	**2,99**	± **0,13**	**1,86**	± **0,11**	1,39	± 0,28
chlorogenate	1,0	± 0,12	**112,65**	± **0,17**	**1730,26**	± **0,08**	**52,51**	± **0,22**	**1304,79**	± **0,25**	**16,14**	± **0,10**
nicotinate	1,0	± 0,09	**5,30**	± **0,18**	**1,85**	± **0,07**	**8,44**	± **0,05**	**1,90**	± **0,10**	**2,23**	± **0,08**
quinate	1,0	± 0,13	**10,07**	± **0,31**	1,39	± 0,11	**4,88**	± **0,22**	**0,53**	± **0,11**	0,61	± 0,26
shikimate	1,0	± 0,16	**11,63**	± **0,10**	**5,51**	± **0,08**	**23,71**	± **0,15**	**9,62**	± **0,12**	3,74	± 0,34
dopamine	1,0	± 0,28	**4,42**	± **0,06**	**8,77**	± **0,13**	**4,35**	± **0,15**	nd		nd	
putrescine	1,0	± 0,11	**0,43**	± **0,18**	1,10	± 0,13	**0,39**	± **0,11**	**0,53**	± **0,18**	0,90	± 0,15
spermidine	1,0	± 0,25	nd		nd		nd		3,12	± 0,37	nd	
tyramine	1,0	± 0,57	1,16	± 0,11	**16,00**	± **0,09**	2,16	± 0,24	1,16	± 0,20	**3,74**	± **0,23**
uracil	1,0	± 0,06	**1,40**	± **0,04**	1,20	± 0,10	**1,65**	± **0,06**	**1,66**	± **0,11**	1,08	± 0,07

Table 4. Metabolite composition in fruit pericarp from species of the *Lycopersicum* complex. Single fruit pericarp samples of six plants were measured. Fruits were harvested 45 DAF and 6h into the light period. Values are presented as the mean ± SE of six independent biological determinations. Those metabolites that are significantly different to *S. lycopersicum* are set in bold type. nd indicates metabolites were not detected.

4.4 Discussion

This study provides the first comprehensive comparative analysis of the metabolite composition of leaves and fruits from the elite tomato species *S. lycopersicum* and five wild species tomato (*S. pimpinellifolium, S. neorickii*, *S. chmielewskii, S. habrochaites, S. chmielewskii* and *S. pennellii*). We have previously used the techniques described in this paper to phenotypically characterize transgenic potato and tomato lines exhibiting altered sucrose breakdown and hexose phosphorylation, respectively (Roessner *et al*. 2001a; Roessner *et al*. 2001b; Roessner-Tunali *et al*. 2003a; Carrari *et al*. 2003). During the course of these earlier studies we optimised our GC/MS method to allow the evaluation of the relative levels of over 60 metabolites in the leaf and fruit of tomato and furthermore established that there were large differences in a broad range of metabolites during fruit development (Roessner-Tunali *et al*. 2003a). Despite the fact given that many of the wild species are green-fruited it is difficult to gauge developmental equivalence. Therefore, for the purposes of this study we harvested all fruits 45 days after flowering. Here we extended the number of metabolites evaluated in both the fruit and leaves to 82 and 71 compounds,

respectively (largely by the procurement of additional chemical standards) and furthermore evaluated the absolute concentrations of the vast majority of these compounds. In addition we also evaluated the levels of starch and protein in fruit samples. The later measurements were carried out in order to assess the influence of water content on the changes observed in metabolite levels.

There is a tremendous variance in metabolite content in both leaves and fruits the wild species. On the basis of leaf metabolite content *S. pimpinellifolium* is the closest wild species to *S. lycopersicum* (in that it shows the least significant differences) followed by *S. pennellii*, *S. chmielewskii* and *S. neorickii* whilst *S. habrochaites* is the most distinct. This may have some basis in the botanical relationship of the species with *S. pimpinellifolium* being the only red-fruited wild species and the only one to have exhibited a natural introgression with *S. lycopersicum.* In fact it is most probable that both species evolved from a common ancestor. *S. chmielewskii, S. neorickii* and *S. pennellii* are all green-fruited species native to Peru. The first two favour growth in moist conditions whilst the later evolved in hot dry environments and has been long regarded as a good source for drought resistance and insect tolerance genes. *S. habrochaites*, native to Southern Ecuador is a trichromous green fruited species with tolerance to low temperature and has been noted to be resistant to several pests as well as containing a high concentration of the naturally occurring pesticide 2-tridecanone. The general pattern of metabolite content in the fruits is however somewhat reversed with *S. habrochaites* showing the fewest significant changes compared to *S. lycopersicum* followed by *S. chmielewskii*, *S. pennellii*, *S. pimpinellifolium* and finally *S. neorickii*. It should be noted that with the exception of *S. habrochaites* all the wild species showed substantially higher levels of metabolic variation in fruits than in the leaves – a fact that probably could be expected to be due to the higher degree of morphological variance in these organs.

The fact that all the wild species studied here can interbred with *S. lycopersicum* has long been exploited by plants breeders with particular attention being paid to biotic and abiotic stresses (Monforte *et al*. 1997a, 1997b; Venema *et al*. 1999; Pan *et al*. 2000). Clear links between resistance to these phenomena and metabolite composition have been demonstrated. For example biotic resistance is often

conferred by the synthesis of molecular repellents (Mitchell-Olds 1998) whereas proline has been reported to play an important role in water stress (see for example (Brugiere *et al*. 1999) and several solutes, especially hexoses have frequently been implicated in response to cold stress (Gilmour *et al*. 2000). Whilst the data presented here may ultimately be of great use in selecting breeding material for the improvement of the abovementioned traits the biochemical basis underlying them is currently unclear. Furthermore, it is likely that such complex traits will be influenced by a wide range of genes and/or biochemical interactions. Another long-term aim of plant breeders over many years has been in fruit quality improvement (Saliba-Colombani 2001; Lecomte *et al*. 2004). Substantial progress has been made in this field and the metabolic basis of several quality traits has been established. Brix (the total soluble sugar content – mainly constituting hexoses, citrate and malate; (Fridman *et al*. 2000)), organoleptic properties and analysis of volatiles (Saliba-Colombani 2001) have all been examined in wide crosses between wild species and the cultivated *S. lycopersicum*. Quantitative trait loci (QTL) for physical and chemical traits have been established (Fridman *et al*. 2000; Saliba-Colombani 2001). In the case of Brix 9-2-5 a QTL for soluble sugar content has been delineated to a 484bp region spanning the third intron and exon of the apoplastic invertase gene LIN5 (Fridman *et al*. 2000). The wide metabolic variance of primary metabolites in fruits of the wild species suggests that similar approaches aimed at boosting the levels of nutritionally important metabolites such as lysine, methionine, ascorbate and tocopherol will stand a high chance of success. The dramatic higher levels of secondary metabolites that we were able to detect by GC/MS also suggests that these wild species represent a valuable resource both for the increase of flavour compounds such as volatiles (Saliba-Colombani 2001; Bovy *et al*. 2002b) and carotenoids (Liu *et al*. 2004) and for the natural product chemist (Lamartiniere 2000; Dixon and Sumner 2003). Rich resources of exotic libraries in which marker-defined chromosome segments from the wild species have been introgressed into the cultivated variety have already been established (Paterson *et al*. 1990; Dixon and Sumner 2003). These will undoubtedly represent an import tool for metabolic engineering particular given current problems of consumer acceptance of products modified by transgenesis.

In conclusion, in recent years there has been an increasing interest in analysing various biological properties of natural genetic diversity (Maloof 2003; Koornneef *et al*. 2004). This study provides a compendium of metabolite levels from leaves and fruits of *S. lycopersicum* and wild species tomato. In combination with prior botanical and genetic studies on adaptability to diverse climates and resistance to biotic stress these data provide correlative information that may, with further experimentation, allow the elucidation of biochemical factors underlying these phenomena. Furthermore, they provide information that may be of considerable importance for breeding-driven metabolic engineering of nutritionally important metabolites.

4.5 Materials and Methods

4.5.1 Plant Material and Growth of Plants

Tomato seeds of accession numbers LA3475 (*Solanum lycopersicum*), LA1589 (*S. pimpinellifolium*), LA2133 (*S. neorickii*), LA1028 (*S. chmielewskii*), LA1777 (*S. habrochaites*) and LA0716 (*S. pennellii*) were obtained from the true-breeding monogenic stocks maintained by the Tomato Genetics Stock Centre (University of California, Davis). The seeds were germinated on Murashige and Skoog medium (Murashige 1962) containing 2% (w/v) sucrose and were grown in a growth chamber 500µmol photons $m^{-2}s^{-1}$, 25°C under a 12h day/ 12h dark regime. Experiments were carried out on mature fully expanded source leaves from six week-old-plants and on fruits taken 45 days after flowering.

4.5.2 Chemicals

All chemicals were purchased from Sigma-Aldrich Chemie GmbH (Deisenhofen, Germany) with the exception of N-methyl-N-[trimethylsilyl]trifluoroacetamide (Macherey-Nagel GmbH & Co. KG, Düren, Germany).

4.5.3 Starch and protein measurements

Starch and protein were extracted and measured as detailed in (Fernie *et al*. 2001) and (Tauberger 2000), respectively.

4.5.4 Extraction, derivatisation and analysis of tomato leaf and fruit metabolites using gas chromatography mass spectrometry

Metabolite analysis by GC/MS was carried out by a method modified from that described by (Roessner *et al*. 2001a). Tomato leaf tissue (250mg) was homogenised using a ball mill pre-cooled with liquid nitrogen and extracted in 1400 µl of methanol, 60µl of internal standard (0.2mg ml^{-1} ribitol in water) was subsequently added as a quantification standard. The mixture was extracted for 15 min at 70°C and mixed vigorously with 1 volume of water. In order to separate polar and non-polar metabolites 750µl chloroform was then added to the mixtures. After centrifugation at 2200g the upper methanol/water phase was taken and reduced to dryness in vacuo. For tomato fruit tissue the same procedure was used with the exception that 300mg of tissue were taken and the extraction mixture was comprised entirely of methanol.

Residues following reduction were redissolved in and derivatised for 90min at 37°C (in 40µl of 20mg ml^{-1} methoxyamine hydrochloride in pyridine) followed by a 30min treatment with 60µl MSTFA (N-methyl-N-[trimethylsilyl]trifluoroacetamide) at 37°C. 8µl of a retention time standard mixture (0.029% (v/v) *n*-dodecane, *n*-pentadecane, *n*-nonadecane, *n*-docosane, *n*-octacosane, *n*–dotracontane, *n*–hexatriacontane dissolved in pyridine) was added prior to trimethylsilylation. Sample volumes of 1µl were then injected onto the GC column using a hot needle technique.

The GC/MS system used comprised an AS 2000 autosampler, a GC 8000 gas chromatograph and a Voyager quadrupole mass spectrometer (ThermoFinnigan, Manchester, UK). The mass spectrometer was tuned according to the manufacturer's recommendations using tris-(perfluorobutyl)-amine (CF43). Gas chromatography was performed on a 30m Rtx-5MS column with 0.2µm film thickness with a10m Integra precolumn (Restek, Bad Homburg, Germany). The injection temperature was set at 230°C, the interface at 250°C and the ion source adjusted to 200°C. Helium was used as the carrier gas at a flow rate of 1ml min^{-1}. The analysis was performed under the following temperature program; 5 min isothermal heating at 70°C, followed by a 5°C min^{-1} oven temperature ramp to 350°C and a final 5min heating at 330°C. The system was then temperature

equilibrated for 1 min at 70°C prior to injection of the next sample. Mass spectra were recorded at 2scan s^{-1} with a m/z50-600 scanning range. Both chromatograms and mass spectra were evaluated using the MASSLAB program (ThermoQuest, Manchester, UK) and the resulting data are prepared and presented as described in (Roessner *et al*. 2001b). The absolute concentrations of most metabolites were determined by comparison to calibration standard curves response ratios of various concentrations of standard substance solutions, including the internal standard ribitol and which were derivatised concomitantly to tissue samples.

4.5.5 Statistical Analysis

If two observations are described in the text as different this means that their difference was determined to be statistically significant (P<0.05) by the performance of Student's *t*-tests.

5 Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement

Nicolas Schauer[1,4$], Yaniv Semel[2,4], Ute Roessner[3], Amit Gur[2], Ilse Balbo[1], Fernando Carrari[1], Tzili Pleban[2] , Alicia Perez-Melis[1], Claudia Bruedigam[1], Joachim Kopka[1], Lothar Willmitzer[1], Dani Zamir[2], Alisdair R. Fernie[1]

[1] Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany.

[2] Institute of Plant Sciences and Genetics and Otto Warburg Centre for Biotechnology, Faculty of Agriculture, Hebrew University of Jerusalem, P.O. Box 12, Rehovot 76100 Israel.

[3] Present address: Australian Centre for Plant Functional Genomics, School of Botany, University of Melbourne, Victoria 3010, Australia.

[4] These authors contributed equally to this work.

$Contribution by NS to methods 5.5.2-5.5.4, 5.5.7 and preparation of the manuscript

5.1 Abstract

Tomato represents an important source of fibre and nutrients in the human diet and a central model for the study of fruit biology. To identify components of fruit metabolic composition we phenotyped tomato introgression lines (IL) containing chromosome segments of a wild species in the genetic background of a cultivated variety. Using this high diversity population we identified 889 quantitative fruit metabolic loci and 326 loci that modified yield associated traits. The mapping analysis indicated that at least 50% of the metabolic loci were associated with quantitative trait loci (QTL) that modified whole plant yield associated traits. Based on correlation analysis we generated a cartographic network that revealed whole plant phenotype-associated and independent metabolic associations including links with metabolites of nutritional and organoleptic importance. The results of this genomic survey illustrate the power of genome-wide metabolite profiling and

detailed morphological analysis for uncovering novel traits with potential for crop breeding.

5.2 Introduction

Fruit is a major portion of the human diet and tomato is a model organism for fruit bearing plants (Giovannoni 2001). In addition tomato is important from an agronomic perspective with a global production in 2004 over 120,000 tons (http://faostat.fao.org, last updated November 2005), with the vast majority of this for human consumption. The cultivated tomato, as many other crop plants, carries only a very small fraction of the genetic variation that is available in related wild species and landraces (Tanksley and McCouch 1997). A present focus in modern plant breeding is to screen wild genetic resources for valuable traits that could be introduced into modern varieties (Zamir 2001). In tomato a major target in the breeding is to improve the nutritional and taste qualities of the fruits through assays of a range of metabolites (Fulton *et al*. 2002; Causse *et al*. 2002; Schauer *et al*. 2005b; Rousseaux *et al*. 2005). Specific examples in which such breeding strategies had markedly altered metabolite contents include carotenoid content in tomato (Liu *et al*. 2003b), protein and oil content in maize (Moose *et al*. 2004), and starch content in potato and rice (Fernie 2004). Across nature, there are now several examples in which one or a few metabolic traits have been studied using this approach including the identification of QTL for energy metabolism, respiration and flight in *Drosophila*, obesity in pigs, and glucosinolate content in *Arabidopsis* (Kissebah *et al*. 2000; Montooth *et al*. 2000; Kliebenstein *et al*. 2001).

Recent technological developments have tremendously extended our ability to describe complex biological systems allowing analysis of multiple metabolites in parallel. These developments alongside measurements at the level of the other constituent entities of the cell present us with an unprecedented opportunity to improve our understanding of function (Barabasi and Oltvai 2004; Fernie *et al*. 2004). The interpretation of the large genomic datasets obtained through the implementation of these technologies is, however, still far from facile and the subsequent understanding of the cellular networks underlying this complexity remains a major scientific challenge (Guimera and Nunes Amaral 2005). Recently great advances have been made in theoretical aspects of network thinking

(Barabasi and Oltvai 2004), whilst experimental biologists have begun to collect and integrate multi-level data sets (Fernie *et al*. 2004; Gibon *et al*. 2004). To date these experiments have largely focussed on the different molecular entities that constitute the cell, concentrating mainly on transcript and protein abundance. However, although providing important information concerning the hierarchical control of biological processes (Gygi *et al*. 1999; Oksman-Caldentey and Saito 2005), and its temporal dependence (Gibon *et al*. 2004), relatively few studies to date have paid attention to morphological features of the investigated organism. Moreover, early proof-of-concept studies of this type were carried out on populations of limited sample size and mainly concerned variation resulting from exposure to different environmental conditions. Although a few such phenotyping experiments have been carried out that consider whole organismal features these exclusively utilize loss-of-function as a mechanism to generate genetic variance (Dudley and Goldstein 2005), and as such have not yet exploited the enormous potential available in natural biodiversity.

To explore the genetic basis of tomato fruit biochemistry we utilized a high-throughput gas chromatography-mass spectrometry metabolite profiling protocol (Roessner *et al*. 2001a), in parallel to whole plant phenotype characterisation. We phenotyped a tomato IL population composed of marker-defined genomic regions of the wild species *Solanum pennellii* substituting for the homologous intervals of the cultivated variety M82 (*Solanum lycopersicum*) (Eshed and Zamir 1995). The fact that the ILs constitute a nearly isogenic resource, largely devoid of the masking effects of complex whole-genome epistatic interactions, made them a unique system for the identification of genomic regions independently associated with changes in fruit metabolism. Utilizing this approach allowed the identification of 889 single metabolite QTL, in addition to the definition and identification of many metabolic QTL that affect a number of compounds in a metabolic pathway. Furthermore, correlation analysis of the combined datasets allowed the analysis of a whole plant phenotype-fruit metabolism network and suggested an important role of whole plant phenotypes on the final metabolite composition of the fruit. However, it should be noted that the results of the current study largely do not allow us to distinguish causality in these networks and it is equally possible that a large number of these associations are driven by the influence a given metabolite

has over plant development. That said irrespective of the mechanisms underlying these associations these results are of high agronomic interest given the fact that many recent biotechnological modifications of crop composition exhibit deleterious effects on yield (Stark *et al*. 1992; Brown 2002; Fernie 2004). Our analysis thus highlights the value of combined genetic, physiological and biochemical profiling to identify evolutionary diverse quality determinants of fruit chemical composition.

5.3 Results

5.3.1 Metabolite profiling

To date *metabolomic* strategies have been concerned with obtaining quantitative information on as wide as possible coverage of the metabolome on a mutant or environment basis (Raamsdonk *et al*. 2001; Fernie *et al*. 2004). Here we use a different interpretation of the phrase *metabolic genomics* since we carry out multi-parallel metabolite analysis using ILs that offer full genome coverage. This method was applied to skinned pericarp material from six independent plants per line in each of two tomato harvests (in the summers of 2001 and 2003). The ripe pericarp was selected for the assays as it is a homogeneous tissue, which constitutes the major edible portion of the fruit. A total of 74 metabolites of known chemical structure were accurately quantified in every chromatogram. These compounds include the majority of plant amino and organic acids, sugars, sugar alcohols, fatty acids and vitamins C (ascorbate) and E (α-tocopherol). The range in content of specific metabolites within the IL population was far in excess of that observed between the parental controls. The metabolite content shows a bias towards an increase in metabolite levels in the ILs on a relative basis compared to M82. Whilst from a naïve standpoint this could be expected since a recent survey of metabolite contents in the parental lines revealed that *S. pennellii* contained higher contents of several metabolites than *S. lycopersicum* (Schauer *et al*. 2005b). It is, however, interesting to note that not only the metabolites that were elevated in the wild species were increased in the ILs. For example, large increases were also apparent in TCA cycle intermediates, despite these being invariant across the tomato clade (Schauer *et al*. 2005b). The mechanism underlying these transgressive events is not apparent from the current study and requires further

investigation. The full data sets from the metabolite profiling study are presented in the overlay heat map of Figure 5. This Figure displays increases (red) or decreases (blue) of metabolite content relative to the *S. lycopersicum* parental line in a false colour scale, on a metabolite content vs. introgressed genomic region plot, with the results of the 2003 harvest superimposed on those of the 2001 harvest. The overlay of the heat maps of the separate harvests revealed that the differences in between the data sets were generally due to quantitative factors (as in only a minority of instances, 9.9%, are purple squares, indicative of contrasting results, apparent). We next performed a correlation analysis on the level of all possible metabolite pairs across the entire population. Sixty-two percent of these revealed significant correlations ($P<0.05$) between the two harvests further demonstrating the reproducibility of the results.

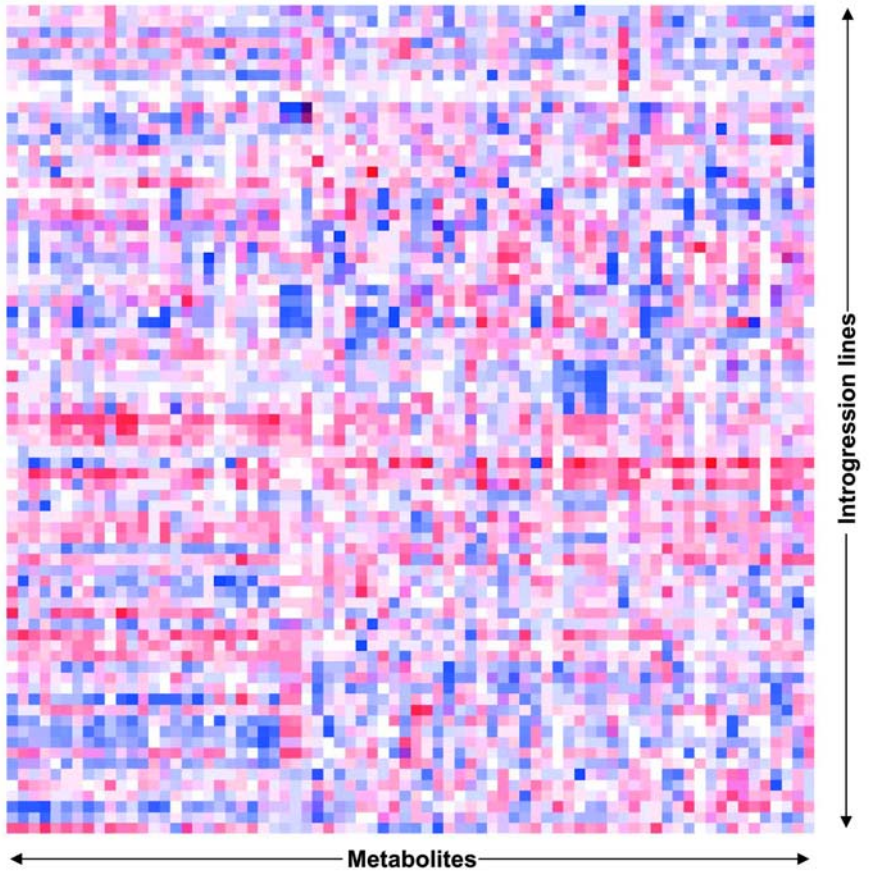

Figure 5. Overlay heat map of the metabolite profiles and other traits of the introgression lines in comparison to that of the parental control (*S. lycopersicum*). Large sections of each map are white or pale in colour reflecting the fact that many of the chromosomal segment substitutions do not have a large affect on the level of every metabolite. Regions of red or blue indicate that the metabolite content is either increased or decreased respectively following the introgression of *S. pennellii* segments. Very dark colouring indicates a large change in metabolite content that was conserved across harvests; purple coloration indicates that a metabolite was

increased relative to *S. lycopersicum* in one harvest but decreased in the other. A total of 74 metabolites were quantified by gas chromatography/mass spectrometry for each harvest, including amino acids, organic acids, fatty acids, sugars, sugar alcohols and vitamins. Fully annotated heat maps for the individual data sets are presented as Supplementary Figure 1.

It is difficult to display such a large data set in a true quantitative manner, however, the absolute difference in the level of any given metabolite ranged between 0.18 and 17.67-fold of that observed in *S. lycopersicum*. Calculation of the Gaussian distribution showed that 95% of metabolite levels were between 0.79 and 2.34-fold of those observed in the elite cultivar. QTL were determined by using ANOVA tests, at a significance level of 0.05, to statistically compare every trait of each IL to the common control (M82). This non-stringent statistical threshold was used since in the first stage of the analysis our objective was to identify general trends in the relationships and associations of a large number of QTL. Using this criterion a total of 889 single trait QTL were identified (presented as Supplementary Table 1). Whilst the majority of the QTL presented in our study are entirely novel, QTL found for fructose, glucose, citrate, glutamate and malate content in tomato have been reported previously (Fulton *et al*. 2002; Causse *et al*. 2004).

5.3.2 Integration of metabolic and whole plant phenotypic traits

In order to place the changes observed in fruit metabolism into a broader physiological context we compared the metabolic traits described above with whole plant phenotypic traits of the ILs. For this purpose we combined the IL biochemical and whole plant phenotypic databases (see Supplementary Table 2). In order to validate this combined approach we statistically verified the stability of the morphological traits of the ILs across several years harvests (Gur *et al*. 2004). It should be noted that Brix (Bx) is a measure of soluble solid content and as such is a biochemical rather than a true morphological trait. The combined whole plant phenotypic and biochemical data was then assessed by pair wise correlation analyses. This analysis weighted the correlations between 3403 pairs of traits across the entire IL population. It revealed a total of 280 positive and only 22 negative correlations at a significance threshold of 0.0001 (see Supplementary Table 3). The high bias towards positive correlation may reflect the dependence of fruit metabolic network on assimilate supply from photosynthesis. Consistent with this hypothesis is the fact that the majority of the negative correlations are

between metabolic and morphological parameters which could be expected to reflect enhanced supply of, or reduced competition for, photoassimilate. We next analysed the correlation network obtained in order to perform a less cursory discrimination of which traits were highly associated (Figure 6). For this purpose we chose to utilize the algorithm recently presented by Guimera and Nunes Amaral (Guimera and Nunes Amaral 2005), which identifies functional modules within complex networks and thereby simplifies their interpretation. The algorithm defines a module as a subset of vertices that are connected to each other more than to vertices in other modules. Starting from the initial state where each node represents a module, the algorithm performs iteration of merging, splitting and transferring nodes between modules. By maximizing the modularity of the network, it ends up with a partition that maximizes the interconnectivity of vertices within modules. Given that the algorithm is stochastic, different runs can, in principal, yield different partitioning. To verify the robustness of the algorithm we obtained 20 partitions of the network depicted in Figure 6 which revealed most pairs of traits are always classified in the same manner (Supplementary Figure 2).

The modularity obtained for the tomato network was relatively simple with three large modules and three smaller outgroups with low connectivity to the rest of the network. The three large modules can be coarsely defined by the traits that they predominantly share with one module comprising largely of whole plant phenotypic traits and phosphorylated intermediates (but also containing sucrose, γ-amino butyric acid, melezitose, proline and quinate), another of the amino acids and the third of sugars and organic acids. The fatty acids palmitate and stearate and the sub-cluster of ascorbate, maltitol, ribose and rhamnose are independent of the network, whilst benzoate and uracil are only loosely associated to it. An interesting feature of the network is the predominance of negative links emanating from the whole plant phenotypic module - in particular from harvest index.

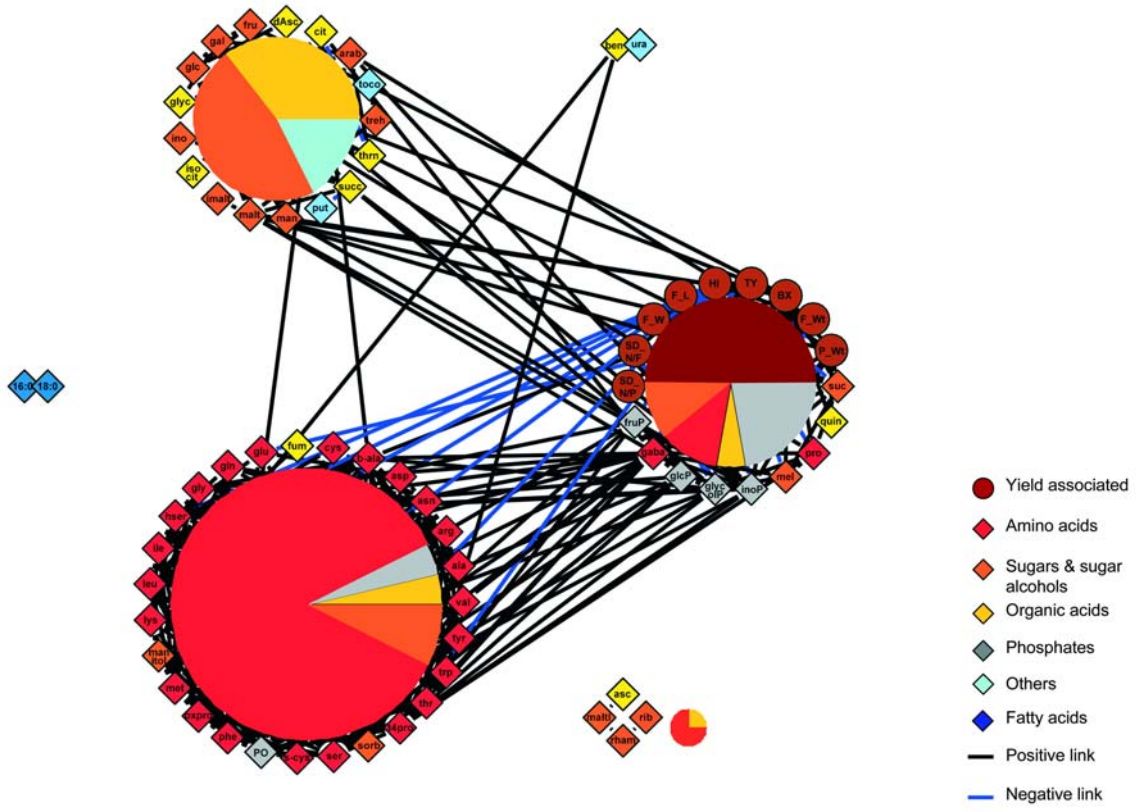

Figure 6. Cartographic representation of the combined metabolic and morphological network of the tomato. Each trait (node) is represented by a shape (morphological by a circle and metabolic a diamond) and colour as defined below. The composition of the modules is represented in the pie charts adjacent to them and pink segments representing amino acids; brown, morphological traits (and Bx), light green, sugars and sugar alcohols; yellow, organic acids; grey, phosphorylated intermediates; blue, fatty acids and light blue, miscellaneous metabolites. Interactions are indicated with lines: black representing positive correlations, blue representing negative correlations. 13 additional traits were not included in the network (they were not correlated to any of the other traits): aconitate, citramalate, erythritol, FA18_2, fucose, galacturonate, gluconate, glycerate-3-P, glycerol, malate, maleate, saccharate, shikimate.

Abbreviations: whole plant yield associated, yield associated; plant weight, P_Wt; fruitweight, F_Wt; brix, BX; total; Yield, TY; harvestindex, HI; fruitlength, F_L; fruitwidth, F_W; seednumber/fruit, SD_N/F; seednumber/plant, SD_N/P; a-tocopherol, toco; alanine, ala; arabinose, arab; arginine, arg; asparagine, asn; aspartate, asp; b-alanine, b_ala; benzoate, benz; citrate, cit; cysteine, cys; dehydroascorbate, d_Asc; FA16_0, palmitate; FA18_0, stearate; FA18_2, linoleate; fructose, fru; fructose-6-P, fruP; fumarate, fum; GABA, gaba; galactose, gal; glucose, glc; glucose-6-P, glcP; glutamate, glu; glutamine, gln; glycerate, glyc; glycerol-3-P, glycolP; glycine, gly; homoserine, hser; inositol, ino; inositol-1-P, inoP; isocitrate, isocit; isoleucine, ile; isomaltose, imalt; L-ascorbate, asc; leucine, leu; lysine, lys; maltitol, maltl; maltose, mal; mannitol, manitol; mannose, man; melezitose, mel; methionine, met; 5-oxoproline, oxpro; phenylalanine, phe; phosphate, PO; proline, pro; putrescine, put; quinate, quin; rhamnose, rham; ribose, rib; S-me-cysteine, s-cys; serine, ser; sorbitol, sorb; succinate, succ; sucrose, suc; t-4-HO-proline, t4prol; threonate, thrn; threonine, thr; trehalose, treh; tryptophan, trp; tyrosine, tyr; uracil, ura; valine, val.

5.3.3 Network evaluation

Whilst the interpretation of this network is in some cases fairly trivial, for instance the interconnectivity of the amino acids is not surprising given the exquisite multi-levelled mechanisms of regulation operating on their metabolism (Galili and Hofgen 2002), many features of the network are novel. In order to comprehensively analyse this network we evaluated the numbers of links each trait displayed to other traits both within and outside their assigned modules. Harvest index (HI), the ratio of fruit yield to total plant mass, was chosen as the focal point of this evaluation since it has both a high total number of connections (20 in total) in the cartographic model presented above and a high number of connections (7) to nodes external to the module in which it belongs. In contrast, with the exception of Bx, all other nodes, even those that are highly connected, are predominantly linked to other nodes within the same module. For this reason we contest that HI is the central pleiotropic hub of the network. Bx is linked to 15 nodes, 5 of which are to nodes external to the module in which it belongs. The reasons for the high connectivity of Bx to the modules that contain exclusively metabolites can be easily rationalised as it is an integrator trait, i.e. variance of this trait is dependent on variance in other traits, for the major metabolites of the fruit (its high connectivity to true plant morphological traits is discussed below). In contrast, none of the other morphological traits determined in this study display characteristics of a network hub, all of them exhibiting few correlations with little or no correlations to constituents of external modules. When the network is assessed from the metabolite perspective it is clear that many metabolites are highly associated within the network with arginine, asparagine, β-alanine, GABA, glycerol 3-phosphate, inositol, inositol 1-phosphate, lysine, methionine, phenylalanine, phosphate, serine, threonine and valine all containing 15 or more significant associations. As would be expected the majority of the metabolites were tightly associated to other nodes within their module, however, a handful of metabolites displayed a high number of correlations both within, and externally to, the module in which they belong. These highly connected metabolites include several that would be anticipated such as the known signal metabolites, phosphate, GABA and inositol 1-phosphate as well as others that maybe would not be expected such as glycerol 3-phosphate. The exact reason for the high-connectivity exhibited by this

metabolite and many of the others mentioned above is not clear from the current study. It is, however, known that the network of amino acid metabolism is subject to a high degree of metabolic regulation (Galili and Hofgen 2002), and that in potato tubers this includes regulation by the photoassimilate supply from the leaf (Roessner-Tunali *et al*. 2003b).

5.3.4 Whole plant phenotypic-associated metabolic traits

Given that so many of the metabolic traits appear to be associated with at least one whole plant phenotypic trait we decided to classify these as either morphologically-associated or –independent. It is important to note that we do not use this phrase to imply that the metabolite content is causally determined by the whole plant phenotype, merely that the behaviour of these traits is highly linked. We took two approaches for classification. First, we looked at the correlation between metabolic and whole plant phenotypic traits across the entire IL population. The correlation of changes in each metabolic and whole plant phenotypic trait was assessed at two different significance levels: 0.005 (strict) and 0.05 (permissive). Metabolites that were significantly correlated to at least one whole plant phenotypic trait (at the strict level) were considered whole plant phenotype associated, those that were not correlated to any whole plant phenotypic trait (even at the permissive level) were considered whole plant phenotype independent and those that were significantly correlated to at least one whole plant trait, at the permissive but not the strict level, were classified as intermediate. Using these criteria we identified 50% of the metabolites as whole plant associated, 23% as intermediate and 27% as independent (Figure 7). In the second approach, we used the Chi method, to statistically evaluate the significance of each metabolic trait co-mapping with whole plant phenotypic traits by analysing the number of shared QTL between two traits. If a pair of metabolic-whole plant phenotypic traits had more common QTL than would be expected due to chance alone the shared metabolite QTL were considered to be whole plant phenotype associated. Using this alternative approach, 46% of the 889 fruit metabolite QTL were considered to be whole plant-associated (see Supplementary Table 4) which is close agreement to the value obtained on the evaluation of the entire population.

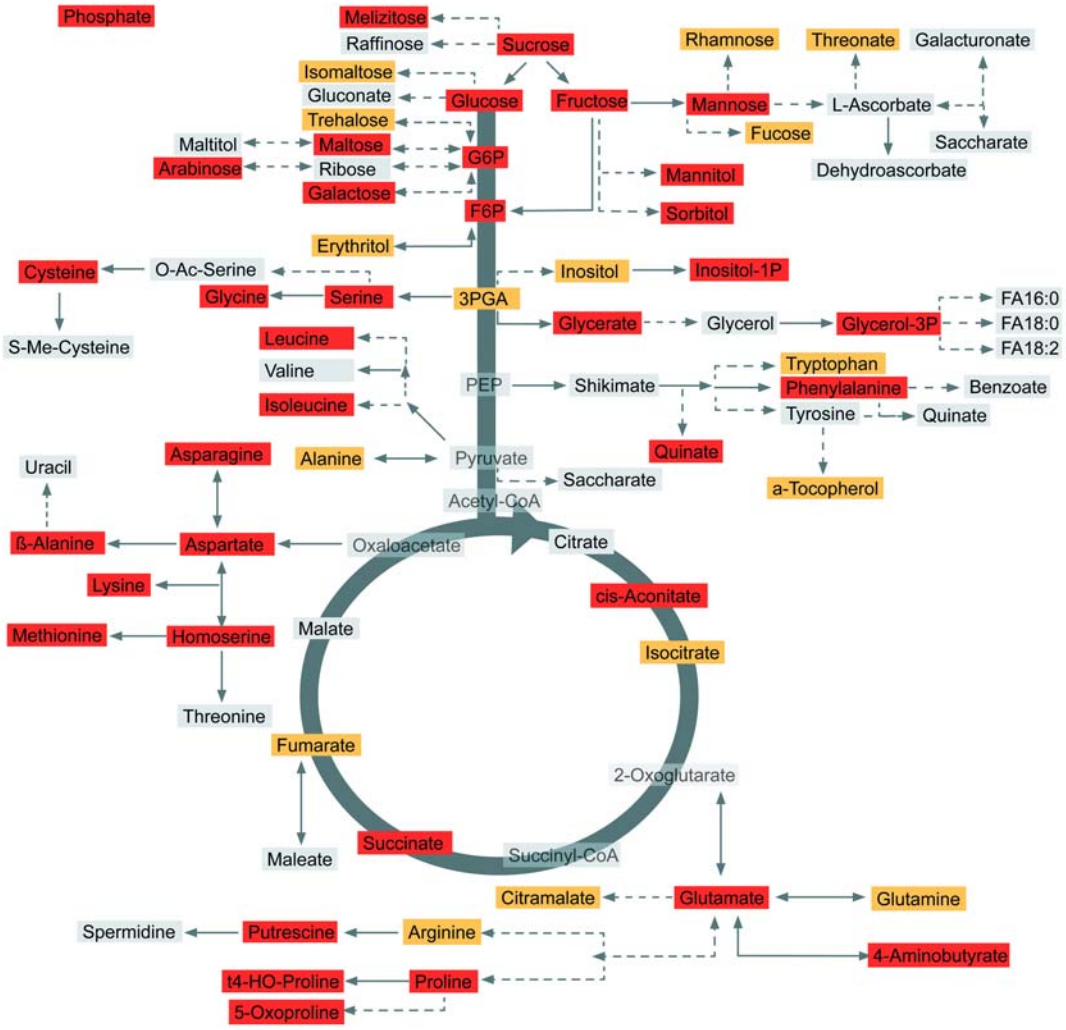

Figure 7. Morphologically-associated and independent metabolites. Two different thresholds were applied in the determination of this association. Those metabolites marked in red were determined to be morphologically associated since they correlated significantly to at least one of the true morphological traits at the strict (0.005 level), those marked in orange were correlated significantly to at least one of the true morphological traits at the permissive level (0.05), whereas those marked in grey were determined to be morphologically independent since neither correlated to any of the true morphological traits at the strict nor the permissive threshold. Traits coloured as pale grey were not analyzed in this study.

Abbreviations: 3PGA, 3-phosphoglycerate; F6P, fructose-6-phosphate; FA16:0, palmitate; FA18:0, stearate; FA18:2, linoleate; G6P, glucose-6-phosphate; PEP, phosphoenolpyruvate**.**

Also in agreement with the correlation approach was the fact that HI was the most highly connected whole plant phenotypic trait with 210 associations to metabolite QTL (followed by plant weight [with 128 associations] and seed number per fruit [117 associations]). Furthermore, the metabolite traits that co-mapped with the major whole plant phenotypic QTL, IL3-4 (which is almost completely sterile), and IL7-2 and 7-3 (which have low HI) are predominantly those which were

demonstrated to have high connectivity via the correlation analyses described above, namely GABA, glycerol 3-phosphate, phosphate and sucrose.

Regardless of the approach taken the majority of the traits that we defined as whole plant phenotype-associated belong to the central metabolic pathways whereas those that are questionable or independent are generally removed from these pathways for example the metabolites associated with vitamin metabolism and the minor sugars (Figure 7 and Supplementary Table 4). Interestingly, those metabolites that correlate strongly with whole plant yield associated traits appear to be more stable; however, further research is required to clarify the biological significance of this observation. The finding that central metabolic pathways are strongly associated to whole plant phenotypes has direct implications for crop improvement strategies, irrespectively of the causality inherent in these associations. Breeding of fruit with high vitamin content is of vast nutritional importance, while on the other hand the major economic determinant of processing tomatoes is Bx and the levels of some of the compounds that constitute this are clearly negatively correlated with the measure of efficiency in partitioning of assimilated photosynthate to harvestable product - the harvest index (HI).

5.3.5 The molecular nature of QTL that regulate fruit metabolism

Having mapped both metabolic and whole plant phenotypic QTL to the genomic regions defined by the ILs and identified metabolites that appeared to be associated with or independent of whole plant yield associated traits we next chose to further evaluate selected regions by the use of the bin-mapping approach combined with more stringent statistical thresholds. The 76 ILs partition the tomato genome into 107 discrete marker-defined mapping bins with a unique composition (http://www.sgn.cornell.edu), thus allowing a greater level of resolution than achievable via IL mapping. We carried out this mapping for two exemplary loci that modify traits important for tomato breeding IL7-4 and IL 6-3, (Figure 8). In the chromosome-7 case there are four overlapping ILs that delimit bin 7B, this bin only contains a single QTL– that for malate content, which is a whole plant-independent trait, that is therefore likely to be mediated by a metabolic or regulatory associated gene. Malate is a major contributor to fruit acidity a trait which is important for both

fresh-market and processing tomatoes. Assessment of the metabolically associated genes mapped to this bin reveals very few candidate genes, with only sucrose phosphate synthase and the vacuolar pyrophosphatase co-localising with the QTL. Of these the second would seem a more likely candidate gene given that the majority of the malate is stored in the vacuole. It should, however, be noted that only relatively few genes associated with metabolism have been positionally mapped in tomato with 93 genes associated with primary metabolism and 58 genes of secondary metabolism mapped to date. That said it can be anticipated that data from the tomato genome sequencing initiative (Mueller *et al*. 2005b) will greatly facilitate the elucidation of the genetic factors that determine metabolite accumulation in the fruit.

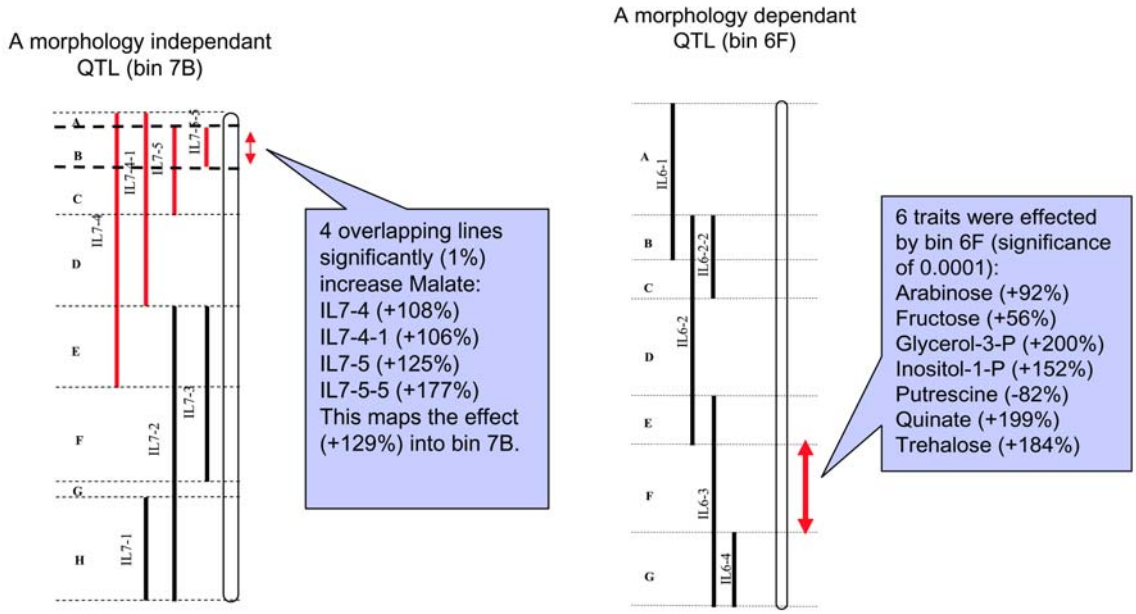

Figure 8. Fine evaluations of genomic regions that harbour morphologically associated and independent metabolite QTL.

Two exemplary regions were chosen:

A. Four overlapping lines that increase (significance of 0.01) the content of malate. The effect is mapped into the overlapped region (bin 7B). Neither of these lines was affected by HI. Malate was determined (Figure 7) to be morphology independant.

B. IL6-3 was affected dramatically by HI (reduction of more than 50%). Seven metabolites were affected by this IL (using very strict significance level - 0.0001). This effect is mapped to bin 6F (as neither of the other overlapping lines affected these metabolites). Six out of these 7 metabolites were determined to be morphology dependant (only trehalose was classified in the intermediate group). The significance level used in the first example was not that strict (0.01 compared to 0.0001 in the second example), because the 4 overlapping lines add more confidence to that mapping.

In the second example, of the major QTL on chromosome 6, far more associated effects were detected following higher-stringency mapping; Bin 6F containing a total of seven metabolite QTL in addition to a dramatic reduction in HI and an increase in Bx. Of these seven metabolites, all with the exception of trehalose (which was deemed intermediate), were defined as whole plant phenotype associated. Interestingly the molecular reason that the harvest index is compromised at this locus is known since the *SELF-PRUNING (SP)* gene of tomato has been cloned (Pnueli *et al*. 1998). *SP* is the tomato homolog of *CENTRORADIALIS (ATC)*, and belongs to a small family of genes involved in the vegetative to reproductive transition including *FLOWERING TIME* (Wigge *et al*. 2005) (Huang *et al*. 2005). Mutation of *SP* results in "determinant" plants in which sympodial segments develop progressively fewer nodes until the shoot is terminated by two consecutive inflorescences thus limiting the fruit yield of the plant but providing for uniform ripening which is the foundation of the processing tomato industry that relies on machine harvest. Furthermore, the closest paralog to *SP*, which co-localised to IL9-2-5, has previously been shown to be a candidate QTL for Bx (Fridman *et al*. 2002). In order to directly assess causality between *SP*, harvest index and Bx three isogenic sets of tomato (*sp* recessive mutants in Gardener and VFNT varities and M82), were planted in the field alongside their respective wild type controls and these parameters were investigated (Supplementary Figure 3). In each instance the harvest index of the *SP* line was much lower than that of the *sp* recessive mutation whilst the Bx value was always higher. These data show that there is a strong negative association between these traits and as such are in support of earlier studies of inheritance of these traits (Emery and Munger 1970). Although this experiment reinforces the conclusion that Bx (and therefore the major fruit metabolites) and HI are associated and it seems likely that, at least in this instance, the change in plant phenotype precedes that of the changes in fruit metabolites, there is still no clear proof of causality in this relationship. In order to address this issue we carried out a second experiment with an F2 population of M82 which segregated for a recessive male sterile mutation ("*ms35"*). This population was also planted in the field. The fertile plants were selfed naturally, whilst flowers of the male sterile plants were pollinated with M82 pollen. An average of 200 fruit per plant were produced by the fertile plants whereas only 50 were produced by the sterile plants; the fertile plants with a HI of

0.85 exhibited an average Bx of 4.2 whilst that of the sterile plants (HI of 0.35) had Bx values of 5.4. When taken together the data from these three experiments provide compelling evidence that plant morphology contributes strongly to the determination of metabolic composition of fruits at harvest time.

5.4 Discussion

This study demonstrates that metabolite profiling has utility beyond diagnostics (Brindle *et al*. 2002; Fernie *et al*. 2004) and gene function annotation (Raamsdonk *et al*. 2001; Fernie *et al*. 2004). Understanding metabolic regulatory circuitry is of interest both for the fundamental comprehension of the complexity of metabolic regulation and for applications such as improving crop quality or the discovery of drugs to counter metabolic disease. Whilst crop yield has been increased significantly in the past twenty years, by harnessing the diversity inherent in wild species for genetic improvement (Tanksley and McCouch 1997; Grandillo *et al*. 1999), attempts to improve tomato quality are generally based both on mutant characterisation and transgenesis approaches (Bovy *et al*. 2002a; Davuluri *et al*. 2004). Here we evaluate fruit pericarp metabolite levels alongside whole plant parameters in an IL population composed of marker-defined regions of the wild species *S. pennellii* substituting for the homologous regions of the cultivated variety M82 (*S. lycopersicum*). The identification power for genomic regions that are significantly associated with quantitative traits in the ILs is higher than in populations that segregates simultaneously for multiple QTL scattered throughout the genome where independent loci can mask the effects of one another due to epistatic interactions. In the IL population the lines are identical for their entire genome except for a single introgressed region and as a result, all the phenotypic variation is associated with the introduced segment and the identification power for individual QTL is increased. Furthermore, due to the permanent nature of the population and the large diversity phenotypes collected from different years are generally reproducible thus facilitating data integration from independent groups. The parallel analysis of multiple traits, once fully integrated in searchable databases, will ultimately allow the selection of lines that contain increases in compounds that are important from a nutritional or organoleptic perspective without compromising the harvest index (such as Malate, Figure 8). Furthermore,

whilst this observation is true for some of the metabolites, changes in the contents of several organic acids in the ILs are in the order of four- to five-fold that found in the elite cultivar and as such are similar to the maximal changes observed via transgenesis (Koyama *et al*. 2000; Weckwerth *et al*. 2004). The combination of marker-assisted selection and metabolite profiling, therefore, represents a viable alternative to GM strategies for metabolic engineering (Zamir 2001). Whilst not wanting to underplay the importance of this observation it is perhaps secondary to the insights into network regulation provided by utilizing metabolic profiling of a population known to exhibit continuous variation in complex traits. Perhaps primary amongst these insights was the fact that a large proportion of the fruit metabolite QTL are strongly associated with whole plant phenotypes. HI, the measure of efficiency in partitioning of assimilated photosynthate to harvestable product, was identified as the major pleiotropic hub in the combined network of metabolic and whole plant phenotypic traits. Moreover, 50% of the metabolites measured were considered to be associated with at least one whole plant yield associated trait. Many studies indicate the influence of metabolites, such as hormones, on whole plant or fruit morphology during early stages of development (Alba *et al*. 2005; Lumba and McCourt 2005). However despite the fundamental importance the analysis of such relationships is beyond the scope of the current study, which focuses on quality aspects at harvest time. That HI is a regulator of the metabolite content of the mature fruit pericarp can be rationalised, since it represents the efficiency of partitioning of assimilated photosynthate to harvestable product, and - as seen in both the male sterile and sp recessive mutants - plants with lower HI have a high Bx. Consistent with this theory, the metabolites that were designated morphology-independent tend to belong to pathways that are relatively distant from the major photoassimilates imported into the fruit and all ILs with low HI contain high levels of the morphology dependant metabolites. The suggestion that a large proportion of fruit metabolic traits are controlled by source sink partitioning implies that targeted manipulation of the content of metabolites of central pathways of primary metabolism should be directed towards this process. In contrast, as evidenced by recent transgenic experimentation the levels of less central metabolites can be altered without paying attention to whole plant physiology (Davuluri *et al*. 2004). The importance of source sink partitioning was previously demonstrated by the cloning, kinetic

characterisation and physiological studies of allelic variance in the apoplastic invertase LIN5 (Fridman *et al*. 2000; Fridman *et al*. 2004; Baxter *et al*. 2005a). Whilst the previous studies focussed in on elucidating the exact mechanism underlying the moderate Brix QTL harboured by this gene, the current survey demonstrates the general importance of plant morphology and source sink relationships in the regulation of fruit metabolite content.

A similar conceptual approach to the one we have taken here was adopted in a recent study of polygenic gene expression networks underlying the nervous system in recombinant inbred mice (Chesler *et al*. 2005). In this study the authors performed global gene expression profiling in recombinant inbred strains with extensive SNP and haplotype data and used graph theoretical approaches to study shared genetic modulation of networks involved in neural synapses. In doing so they were able to identify QTL for transcript abundance of up to 1650 transcripts in neuronal tissue and to relate some of these to previously characterised behavioural traits of the mice. The research presented here demonstrates that to gain more comprehensive understanding it is important to assess the functional interactions in the organism as a whole. The observation that plant morphology is a major factor affecting fruit metabolic profiles, suggests that this phenotype might regulate biological processes at various molecular levels. Thus morphology of permanent populations should be adopted, as the basis on which to superimpose different levels of molecular information, in attempts to understand and breed complex biological systems. In the case of crop biotechnology the end phenotype must also remain an important factor since successful manipulation of chemical composition is highly compromised if it occurs at the cost of plant performance or fertility. Metabolite profiling is essentially species-independent (Stitt and Fernie 2003), and a wide range of breeding populations exploiting natural diversity have been created for various crop species (Zamir 2001; McCouch 2004). Therefore, the approach described here could be readily utilized in the general understanding of the genetics underlying metabolite abundance and ultimately in the selection of breeding stock for improved crop quality.

5.5 Methods

5.5.1 Growth conditions

The presented metabolite data set is based on field grown introgression lines from two seasons, 2001 and 2003. The field trials were conducted in Akko (Israel) at Western Galilee Experimental Station. Plants were grown in a completely randomized design with 1 plant per 1m². Seedlings were grown in greenhouse for 35-40 days and then transferred in the field. The field was irrigated with 320m³ of water per 1000m² field area throughout the season.

Nine morphology traits were measured on the IL population. *Plant weight*. Plant vegetative weight (kg/plant) was determined by weighing only the vegetative tissue (after harvesting of the fruits), without the roots. *Total yield*. Total weight of the red and green fruits (kg/plant). *Brix*. Concentration of total soluble solids (in percent Brix) was measured using the digital refractometer (RFM-80 BS) from a random sample of 10 fruits per plant. *Fruit weight*. Mean fruit weight (g) calculated from a random sample of 10 representative fruits per plant. *Harvest index*. Calculated as *total yield*/*(total yield + plant weight)*. *Seed number per fruit.* Five fruits were collected from each plant and all the seeds were extracted and incubated in plastic cups for fermentation. After 30h, the seeds were washed and dried on a filter paper. A sample of 100 seeds was weighed using a semi-analytic mass balance (Shimadzu Company; range of 0-120g, accuracy of 0.001g). All seeds collected from the five fruits were weighed and their number was estimated based on the 100 seeds weight sample. *Seed number per plant*. Estimated by multiplying *seed number per fruit* and *fruit number* (which is the total yield divided by fruit weight). *Fruit length*, *Fruit width.* Six representative ripe fruits were collected from each replication and subjected to fruit morphology analysis. A longitudinal section of three fruits and a transverse section of three different fruits were cut and placed on a standard Epson1600 desktop scanner. WinFOLIA image analysis software (http://www.regentinstruments.com/products/folia/FOLIA.html) was used for measuring fruit maximum length and fruit maximum width. The first 5 traits (Plant weight (PW), Total yield (TY), Brix (BX), Fruit weight (FW), Harvest index (HI)) were measured in two harvests (2004 and 2000), while the other four (Seed number per fruit (SN/fruit), Seed number per plant (SN/plant), Fruit length, Fruit

width) were measured only in 2004 harvest. Experimental procedures of these experiments were similar to those described above.

5.5.2 Metabolite measurement by gas chromatography-mass spectrometry

The determination of relative metabolite content was carried out essentially as described in (Roessner *et al*. 2001a) following modifications for tomato described in (Roessner-Tunali *et al*. 2003a).

5.5.3 Statistics

Statistical analyses were performed using the JMP IN 5.1 software package (SAS Institute, 2003, R statistical software (http://www.R-project.org) or Microsoft Excel 7.0 (Microsoft Corp., 2000). Distribution of changes was calculated using Gaussian distribution with a ratio of 0.95. Correlations across the entire population were calculated using Pearson Product Moment Correlation (Pearson's ρ).

5.5.4 IL mapping

For mapping of metabolites a two-way ANOVA has been used to partition metabolic variation into genotype, environment and genotype x environment interaction effects. This method has been commonly applied in transcriptional analysis and shows excellent robustness. A metabolic effect in one specific IL has been assigned significant when the control and an IL have at least four replications in each year and the genotype factor for the combined two years analysis is significant (a=0.05), whereas the interaction factor (a=0.01) was not significant (or if the IL was significantly different (0.05) than M82 in each year separately (*t*-test) and in the same direction). Genotypes missing in one year or have less than four replications have not been subjected to ANOVA.

5.5.5 Network analysis

Correlation of all traits (metabolite + morphologic) pairs were calculated using IL means (total of 76 lines). The 83 traits yield 3403 pairs, so we choose strict

significance levels (0.0001). The 302 resulted pairs were considered as a network in which a vertex corresponds to trait, and a link between 2 vertices corresponds to significant correlations between these 2 traits. This network was then subjected to the cartography algorithm (Guimera and Nunes Amaral 2005). Basically, this algorithm divides the network into modules, which are groups of vertices that are connected between themselves more than to nodes from other modules. This yields a cartographic representation of a complex network. Whilst, implementing this algorithm, negative correlations were considered equal to positive correlations.

5.5.6 Chi value

The Chi value is calculated to define the association between two traits using their shared QTL by subtracting the number of expected QTL from the number of observed QTL and then dividing this value by the expected QTL (Chi = (observed QTL – expected QTL)/expected QTL). The expected QTL value is calculated by (the number of QTL of Trait1/Number of all ILs)* Number of QTL of trait2. Observed QTL is the number of shared QTL between the 2 traits. It can be either the number of shared QTL that are in the same direction (both increase or both decrease compared to the control) or in the opposite direction (one increase and the other decrease). To determine this, both types of shared QTL were counted and the type with the higher number was selected. This determined the association between these 2 traits – if the 'same direction' type is prevalent, there is a positive association; if the 'opposite direction' is prevalent, then it is negative association. Metabolites with less than five QTL were excluded from the analysis. To test the chance of two metabolite QTL sharing the same IL a binomial probability (BP) was calculated using following formula:

$$BP= \binom{n}{k} * p^k * (1-p)^{n-k}.$$

Where p is the probability to get a shared QTL for the 2 traits in a specific IL, calculated by p = (QTL_number_of trait_1 / n)*(QTL_number_of trait_2 / n)*0.5. (The reason to multiply it by 0.5 is that in order to be considered as a shared QTL it should comply with the association between the traits; in positively associated traits, the shared QTL must be in the same direction for both of them, while in a negatively associated, it should be in the opposite direction). k is the number of shared QTL and n is the number of all ILs (76). Association between traits were

termed significant when the BP value was below 0.05 and the Chi value was greater than 0 (to exclude cases in which there are significantly less shared QTL than expected).

5.5.7 Heatmap

Heat maps were calculated using the 'heatmap' module of the statistical software environment R (http://www.r-project.org) version 1.9. False colour imaging was performed on the log10-transformed metabolite data. Data was internally scaled on a column-basis to have mean zero and standard deviation one. A colour range of 25 steps was shown to be the most useful for presentation. The overlay heatmap was generated using Adobe Photoshop's (Adobe Systems Inc, 2002) multiplication function on the single heat maps.

6 Aromatic amino acid decarboxylases participate in synthesis of the flavour and aroma volatiles, 2-phenylethanol and 2-phenylacetaldehyde in tomato fruits

Denise Tieman, Mark Taylor, Nicolas Schauer*[$], Alisdair R. Fernie*, Andrew D. Hanson, Harry J. Klee[†]

University of Florida, Horticultural Sciences, PO Box 110690, Gainesville, FL 326110690 USA

* Max Planck Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm-Potsdam, Germany

$ Contribution by NS to methods 6.5.6 and 6.5.7 and to the discussion

Abbreviations: AADC, aromatic amino acid decarboxylase; GST, glutathione S-transferase

6.1 Abstract

2-Phenylethanol is an important phenylalanine-derived volatile compound produced by plants. It is a major contributor to flavour in many foods, including fresh fruits such as tomato, and an insect-attracting scent in roses and many other flowers. Despite the centrality of 2-phenylethanol to flavour and fragrance, plant genes involved in its synthesis have not been identified. Here we identify a biosynthetic pathway for 2-phenylethanol and other phenylalanine-derived volatiles in tomato fruits, and a small family of decarboxylases that can mediate its first step. These enzymes catalyze conversion of phenylalanine to phenethylamine as well as tyrosine to tyramine. Although tyrosine is the preferred substrate *in vitro*, phenylalanine levels in tomato fruits far exceed those of tyrosine, indicating that phenylalanine is a physiological substrate. Consistent with this view, overexpression of two of the decarboxylases in transgenic tomato plants resulted in fruits with up to ten-fold increased emissions of the products of the pathway including 2-phenylacetaldehyde, 2-phenylethanol, and 1-

nitro-2-phenylethane. Besides establishing a biosynthetic route, these results show that it is possible to change phenylalanine-based flavour and aroma volatiles in plants by manipulating expression of a single gene.

6.2 Introduction

Human perception of flavour involves integration of multiple chemical stimuli from taste and olfactory receptors. While taste receptors respond to a limited set of cues, olfactory receptors respond to thousands of chemicals and provide the diversity of unique food flavours. For example, there are about 15-20 volatile compounds that together constitute the unique flavour of fresh tomatoes (Buttery *et al*. 1971; Buttery 1993; Baldwin *et al*. 2000; Baldwin *et al*. 2004; Tikunov *et al*. 2005). These volatiles are derived from various precursors, including fatty acids, carotenoids, and amino acids. Several of the most important tomato aroma volatiles, including 2-phenylacetaldehyde and 2-phenylethanol, come from phenylalanine (Buttery 1993). 2-Phenylethanol is also a major flavour constituent of such diverse foods as cheese, bread, wine and olive oil. Both 2-phenylacetaldehyde and 2-phenylethanol have pleasant fruity/floral odours, and are major contributors to scent in many flowers, e.g., 2-phenylacetaldehyde in hyacinths and 2-phenylethanol in roses (Knudsen 1993). Because of its desirable aroma and association with flowers, 2-phenylethanol is the most used fragrance chemical in cosmetic products (Clark 1990). As a consequence, there is much interest in natural sources of 2-phenylethanol for the flavour and fragrance industry.

2-Phenylacetaldehyde and 2-phenylethanol have important biological functions in plants. 2-Phenylethanol has long been known to have antimicrobial properties (Berrah and Konetzka 1962) and its presence in plant reproductive structures suggests a protective role for flowers and fruits. Both 2-phenylacetaldehyde and 2-phenylethanol are potent insect attractants (see http://www.pherobase.com/) and each attracts different sets of pollinating and predatory insects (Zhu *et al*. 2005). The presence of 2-phenylacetaldehyde and 2-phenylethanol in ripening fruits is also probably related to their attractiveness to mammals and other seed dispersers (Goff and Klee 2006).

Such multiple roles in defence and reproduction suggest that regulation of their synthesis likely to be critical to the plant.

Despite the importance of 2-phenylacetaldehyde and 2-phenylethanol to flavour and aroma, it is not clear how plants make them. The yeast *Saccharomyces cerevisiae* produces 2-phenylethanol from phenylalanine via phenylpyruvate and 2-phenacetaldehyde (Vuralhan *et al*. 2003). Deuterium labelling studies in rose (*Rosa damascene* Mill.) indicated that there might be as many as four pathways of synthesis (Vuralhan *et al*. 2003). In addition to the yeast pathway, (Watanabe *et al*. 2002) reported synthesis via a phenethylamine/2-phenylacetaldehyde route and a *trans*-cinnamic acid/phenyllactate pathway. As plants contain many aromatic L-amino acid decarboxylases (AADCs) (Facchini *et al*. 2000), a pathway that begins with phenylalanine decarboxylation is reasonable *a priori*. Here we demonstrate that tomato (*Solanum lycopersicum*) indeed uses a pathway whose first step is decarboxylation of phenylalanine to phenethylamine. This reaction is catalyzed by a set of related AADCs. Overexpression of the corresponding genes in transgenic tomato plants led to accumulation of significantly higher levels of 2-phenylacetaldehyde and 2-phenylethanol as well as the related compounds 2-phenylacetonitrile and 1-nitro-2-phenylethane.

6.3 Results

6.3.1 The pathway for synthesis of 2-phenylethanol in tomato fruits

Biochemical considerations indicate that the first step in 2-phenylethanol synthesis most probably involves either decarboxylation or deamination of phenylalanine. In the former case, the predicted reaction product is phenethylamine and in the latter case it is phenylpyruvate. In order to determine the preferred tomato pathway, we examined fruits for the presence of possible intermediates. (Tadmor *et al*. 2002) described a line, IL8-2-1, that contains a single introgressed portion of chromosome 8 derived from the wild relative *S. pennellii*. This line contains a locus, *malodorous*, that is associated with a large increase in emissions of 2-phenylacetaldehyde and 2-phenylethanol. Since we have routinely observed more than 1000-fold increases in these two volatiles over multiple growing seasons (Tieman *et al*. 2006), we used this

line and its near isogenic parent, M82, for subsequent analyses. In addition to these two volatiles, IL8-2-1 emitted greatly increased amounts of two other phenylalanine-derived volatiles, 2-phenylacetonitrile and 1-nitro-2-phenylethane (Figure 9).

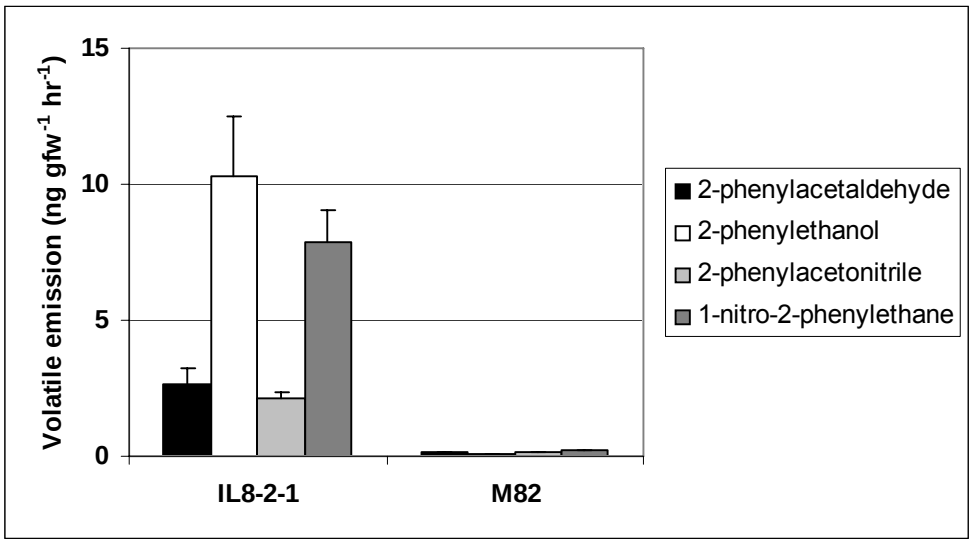

Figure 9. Levels of phenylalanine-derived volatile emissions from M82 and IL8-2-1 fruit.

These latter two volatiles are almost surely derived from phenethylamine rather than phenylpyruvate because they have a nitrogen atom in the side chain. Therefore, the most likely pathway for synthesis of 2-phenylethanol is via phenethylamine, as shown in Figure 10.

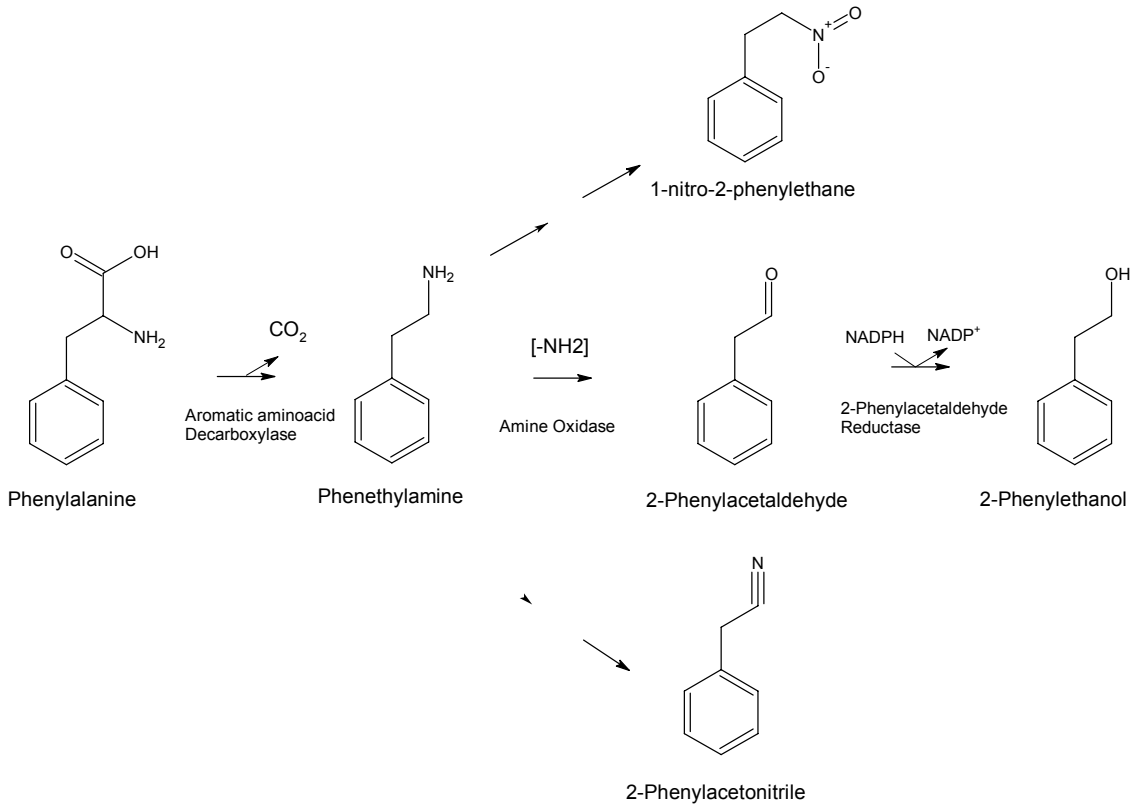

Figure 10. Proposed pathway for production of the volatiles compounds phenylacetaldehyde and 2-phenylethanol in plants. Phenylalanine is decarboxylated to form phenethylamine by an AADC. Phenethylamine is converted to 2-phenylacetaldehyde by removal of an amine group, followed by conversion to 2-phenylethanol by phenylacetaldehyde reductase. The volatile compounds 1-nitro-2-phenethane and 2-phenylacetonitrile are co-ordinately synthesized and are likely derived from further metabolism of phenethylamine.

To validate the hypothesized pathway, fruit pericarp discs from M82 and IL8-2-1 were fed [^{13}C]phenylalanine. After four hours, tissues were examined for the presence of and flux through [^{13}C]phenethylamine (Table 5). Labelled phenethylamine was detected in both lines with substantially more product in IL8-2-1. The samples were also examined for the presence of labelled phenylpyruvate and phenyllactate. Neither product was detected. By defining the limits of detection for each compound (Supplementary Table 5), it was determined that levels of phenylpyruvate must be at least ten-fold lower and phenyllactate 100-fold lower than phenethylamine, if present at all.

Flux Phe(PEA)	M82	IL8-2-1
nmol g^{-1} h^{-1}	0,15±0,12	0,87±0,33

Table 5. Flux through phenethylamine following [^{13}C]phenylalanine feeding of M82 and IL8-2-1 pericarp discs. The data are expressed as nmol of phenylalanine metabolized to phenethylamine (PEA) per gram fresh weight per hour ± SE.

M82 and IL8-2-1 were then examined for the capacity to decarboxylate phenylalanine *in vivo*. Pericarp discs were fed [^{14}C]phenylalanine for eight hours and the amount of $^{14}CO_2$ generated over the period determined. The amount of [^{14}C]phenethylamine in the tissues was also measured (Table 6). Both samples decarboxylated phenylalanine and both contained [^{14}C]phenethylamine with significantly higher quantities of each generated byIL8-2-1, consistent with higher activity in IL8-2-1. The non-stoichiometric amounts of products can be attributed to further metabolism to 2-phenylacetaldehyde and 2-phenylethanol. Taken together, the above data are consistent with the major, if note exclusive, route to synthesis of 2-phenylethanol in tomato fruits being via a phenethylamine pathway that is more active in IL8-2-1.

Line	nCi CO_2	nCi phenethylamine
M82	0.96±0.05	0.36±0.02
8-2-1	5.16±2.36	2.08±0.39

Table 6. Phenylalanine decarboxylase activity of M82 and IL8-2-1 tomato pericarp disks fed [14C]phenylalanine. Pericarp disks were fed 1µCi (2.17 nmol) universally labelled [^{14}C]phenylalanine for 8 h, and amounts of $^{14}CO_2$ and [^{14}C]phenethylamine produced were determined. Data are means of four replicates ± SE. Note that because the phenylalanine contains 9 carbon atoms and was universally labelled, a 1:1 molar ratio of CO_2:phenethylamine corresponds to a 1:8 ratio of ^{14}C.

6.3.2 Identification of AADC candidate genes by screening in *E. coli*

The tomato EST database (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi-?species=tomato) contains many clones annotated as amino acid decarboxylases, based on homology to known enzymes. Several full length cDNA clones representing different subgroups of enzymes were isolated and expressed as recombinant proteins in *E. coli*. Bacterial cultures were grown in media supplemented with phenylalanine. Following growth, cells and culture media were extracted with hexanes and assayed for the presence of phenethylamine. Cultures expressing proteins annotated as

histidine decarboxylases (clones cLEC73K23 andcLEC75E21) readily converted phenylalanine to phenethylamine; those expressing proteins annotated as tyrosine/dopa decarboxylases did not (Figure 11).

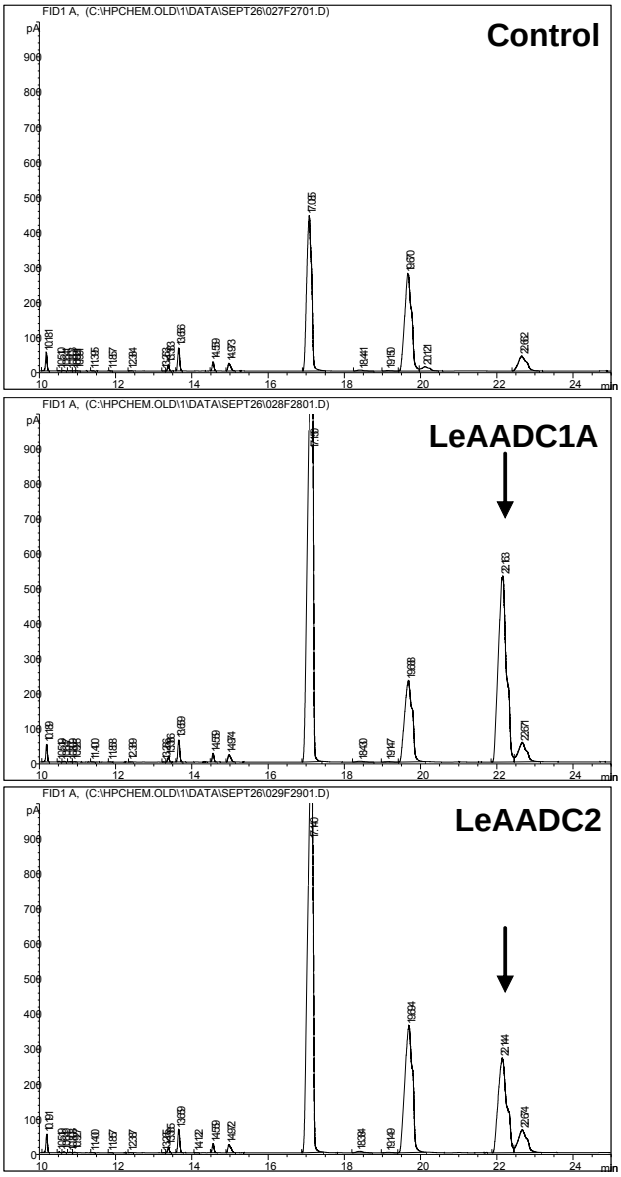

Figure 11. Gas chromatography profiles of volatile compounds extracted from *E. coli* cultures expressing the tomato LeAADC1A and LeAADC2 cDNAs. Only cultures expressing these genes produce phenethylamine (arrows) when grown in media supplemented with phenylalanine.

Further screening of the available EST clones indicated that there are two highly related genes, designated *LeAADC1A* (GenBank accession) and *LeAADC1B* (GenBank accession). The proteins encoded by these two genes are 95% identical. Both genes map to the same position on Chromosome 8 (http://www.sgn.cornell.edu),

indicating that they likely arose by a duplication event. A third cDNA encoding an enzyme with phenylalanine decarboxylase activity (GenBank accession), with 81% identity to LeAADC1A, was named *LeAADC2* (Supplementary Table 5).

6.3.3 Activities of AADC enzymes

A histidine-tagged version of LeAADC1A was produced in *E. coli*, purified, and kinetically characterized using tyrosine and phenylalanine as substrates. Tyrosine was preferred, giving far greater activity than phenylalanine at low substrate concentrations (Table 7). The K_m value for tyrosine was around 1mM, but was not measured precisely because of substrate inhibition (Figure 12). The K_m for phenylalanine was clearly much higher since saturation was not reached at 40mM, the highest concentration tested (Figure 12). Partial characterization of LeAADC2 likewise indicated a strong preference for tyrosine (Table 7). Neither enzyme attacked histidine (not shown). In connection with substrate preference, it should be noted that tomato fruits contain much more phenylalanine than tyrosine (10.45 vs. 0.76µmol g^{-1} fresh weight, respectively) (Schauer *et al*. 2005b). We will return to this point in the Discussion. To assess the stoichiometry of the reaction, uniformly labelled [^{14}C]phenylalanine was used as substrate for LeAADC1A and the molar ratio of CO_2 to phenethylamine was determined. This ratio was found to be 1.04 ± 0.04 (mean of five replicates ± SE).

Enzyme	Tyrosine activity (nmol CO_2 min^{-1} mg^{-1} protein)		Phenylalanine activity (nmol CO_2 min^{-1} mg^{-1} protein)	
	1 mM	10 mM	1 mM	10 mM
LeAADC1A	120	100	0.32	3.5
LeAADC2	0.90	0.54	0.0054	0.035

Table 7. Comparison of the activities of LeAADC1A and LeAADC2 with tyrosine or phenylalanine as substrates. Activities were measured at pH 7.9 using the $^{14}CO_2$ release assay. The LeAADC1A protein was partially purified by Ni^{2+} affinity chromatography. The LeAADC2 protein came from extracts of *E. coli* cells.

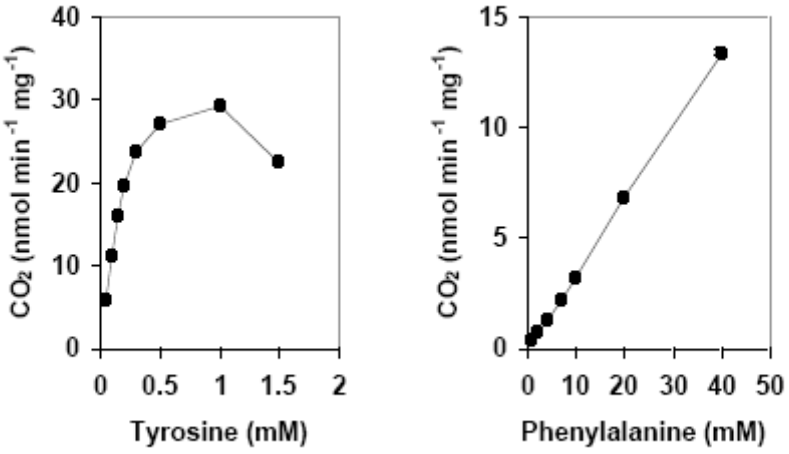

Figure 12. Comparison of the activity of LeAADC1A with tyrosine or phenylalanine as substrate. Activities were measured at pH 7.9 using the $^{14}CO_2$ release assay. The LeAADC1A protein was isolated by Ni^{2+} affinity chromatography.

6.3.4 *In vivo* functions of *AADC* genes

The above *in vitro* data indicated that the LeAADC enzymes can convert phenylalanine to phenethylamine. To validate these observations *in vivo*, expression vectors containing full length *LeAADC1A* and *LeAADC2* cDNA clones under control of a constitutive 35S promoter were constructed and transgenic tomato plants were produced. Multiple independent lines overexpressing each cDNA were identified (Supplementary Figure 5) and the volatile profiles of ripe fruits from each line were determined. The results showed that overexpression of either enzyme significantly enhanced production of multiple phenylalanine-derived volatiles (Table 8). Each of the lines overexpressing either gene product had significantly higher emissions of 2-phenylethanol and 1-*nitro*-2-phenylethane. The 2-phenylacetaldehyde levels were usually not elevated. This latter result is consistent with observations of transgenic plants altered in phenylacetaldehyde reductase expression indicating that this enzyme is not normally limiting for 2-phenylethanol synthesis (Tieman *et al*. submitted). Although 2-phenylacetonitrile emissions were significantly higher in IL8-2-1, the transgenic lines exhibited variable emissions that were not significantly increased relative to the controls. The levels of phenethylamine in selected transgenic lines were also determined. Ratios of phenethylamine in red ripe fruits were

1.68±0.25, 1.1±0.18, 1.67±0.27, 1.3±0.15 and 1.35±0.11-fold higher than control M82 fruit.

Line	2-Phenylacet-aldehyde	2-Phenylethanol	2-Phenylacetonitrile	1-Nitro-2-phenylethane
M82	0.12±0.04	0.07±0.03	0.13±0.05	0.19±0.06
1A-6306	0.45±0.16*	4.02±1.16**	0.25±0.05	1.44±0.31**
1A-6404	0.16±0.12	0.67±0.14**	0.21±0.03	1.04±0.17**
1A-6410	0.19±0.08	0.75±0.34**	0.23±0.16	1.58±1.44*
2-6610	0.30±0.14	0.47±0.14**	0.19±0.05	0.97±0.21**
2-6612	0.16±0.05	0.50±0.28*	0.08±0.02	0.85±0.20**
2-6721	0.14±0.09	0.62±0.32**	0.24±0.11	2.01±0.77**

Table 8. Volatile emissions from fruits of control (M82) and transgenic lines overexpressing either LeAADC1A (prefix 1A-) or LeAADC2 (prefix 2-). Values are presented as ng g^{-1} FW h^{-1} ± SE. Values significantly different from those for M82 are indicated (* P<0.05; ** P<0.01)

Real time RT-PCR analysis of expression patterns for each of the genes indicated that all three are expressed in ripening fruit tissues (Figure 13). Since multiple *AADC* genes are expressed in ripening fruit tissues, the pathway for synthesis has redundancy. Consistent with this redundancy, single gene loss-of-function transgenic lines were not significantly different from control fruits (data not shown).

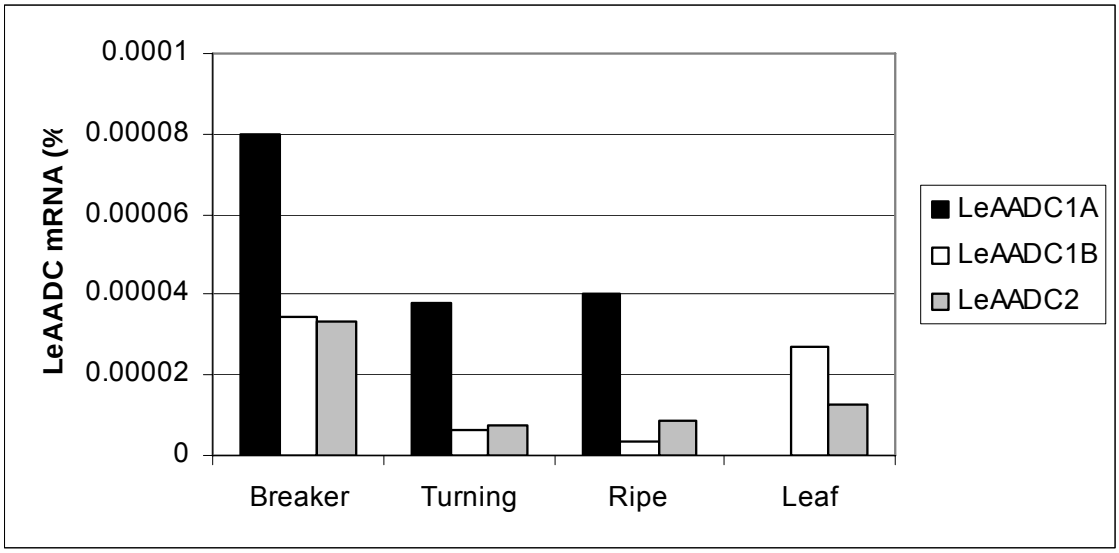

Figure 13. LeAADC expression in ripening fruits and leaves of M82 tomato plants. Levels of mRNA were determined using TaqMan real time quantitative RT-PCR.

6.4 Discussion

The results presented here establish the major pathway for synthesis of 2-phenylethanol from phenylalanine in tomato fruits. This pathway involves enzymatic decarboxylation of phenylalanine by aromatic amino acid decarboxylases to produce phenethylamine. Although we have not identified genes encoding the second step, phenethylamine is presumably converted by an amine oxidase, dehydrogenase or transaminase to 2-phenylacetaldehyde. The final step in the pathway to 2-phenylethanol is catalyzed by 2-phenylacetaldehyde reductase. Analysis of introgression and transgenic lines overproducing AADC enzymes further indicates common intermediates for synthesis of the related volatiles 1-nitro-2-phenylethane and 2-phenylacetonitrile.These latter two volatiles are most likely derived from phenethylamine (Figure 10). Both have been previously identified in tomato fruits and 1-nitro-2-phenylethane is considered to be a major contributor to tomato flavour, as are 2-phenylacetaldehyde and 2-phenylethanol (Buttery 1993; Baldwin *et al*. 2000). Although we cannot rule out alternative pathways to synthesis of 2-phenylethanol, we were unable to detect either phenylpyruvate or phenyllactate in fruit tissue. Based on the limits of detection of these compounds, we can say that these latter pathways for 2-phenylethanol synthesis are relatively minor, if they exist at all in tomato fruits.

The first step in 2-phenylethanol synthesis is catalyzed by a family of aromatic amino acid decarboxylases. Although currently annotated in databases as histidine decarboxylases, they are active against phenylalanine and tyrosine and do not decarboxylate histidine. This initial misannotation is not surprising since both histidine and aromatic amino acid decarboxylases belong to the same class of pyridoxal phosphate-dependent amino acid decarboxylases (Sandmeier *et al*. 1994). In tomato there are two highly homologous enzymes that are likely derived from a gene duplication event (LeAADC1A and LeAADC1B) as well as a third distinct enzyme (LeAADC2). Both LeAADC1A and LeAADC2 catalyze conversion of phenylalanine to phenethylamine *in vitro*. Although we have not tested LeAADC1B for activity, its high identity with LeAADC1A (95%) suggests that it should have similar enzyme activity. All three genes are expressed at comparable levels in fruits while *LeAADC1B* and *LeAADC2* are also expressed in leaves. We cannot rule out the existence of additional *AADC* genes in the tomato genome.

Most aromatic amino acid decarboxylases convert their substrates to the corresponding amine and nothing more. However, Kaminaga *et al*. (N. Dudareva, personal communication) have recently reported a petunia AADC that converts phenylalanine directly to phenylacetaldehyde, having amine oxidase as well as decarboxylase activity. Although petunia and tomato are quite closely related (both belong to the family *Solanaceae*), all evidence indicates that the tomato enzymes described here simply convert phenylalanine to phenethylamine. Specifically: (i) the tomato enzyme LeAADC1A produces CO_2 and phenethylamine in a 1:1 molar ratio; (ii) representative plants overproducing LeAADC enzymes have expanded pools of phenethylamine; (iii) two volatile compounds, 1-nitro-2-phenethane and 2-phenylacetonitrile, which are produced at very high levels in IL8-2-1 and to a lesser extent in the transgenic lines, are almost certainly derived from phenethylamine. In our analyses of petunia volatiles, we have not detected emissions of 1-nitro-2-phenylethane and 2-phenylacetonitrile. It therefore appears that petunia and tomato have evolved distinct pathways to 2-phenylethanol that start with different enzymes, the petunia enzyme being peculiarly specialized.

AADC activity appears to exert major control over the flux from phenylalanine to multiple volatile compounds. IL8-2-1 has significantly higher AADC activity as well as emissions of 2-phenylacetaldehyde, 2-phenylethanol, 1-nitro-2-phenylethane and 2-phenylacetonitrile. This increased flux occurs in the absence of any change in the pool of free phenylalanine in IL8-2-1 relative to M82 (Chapter 5). Transgenic plants overexpressing either *LeAADC1A* or *LeAADC2* also exhibited increases in these volatiles. However, none of the transgenic plants synthesized these volatiles to levels even remotely approaching those of IL8-2-1. Although *LeAADC1A* and *LeAADC1B* map to the segment of chromosome 8 corresponding to the *S. pennellii* introgression, we have no evidence that the *AADC* orthologues are responsible for the *malodorous* phenotype.

The high activity against tyrosine exhibited by the tomato enzymes indicates that they have the potential to synthesize tyramine as well as phenethylamine *in vivo*. Tyramine is a precursor for many plant alkaloids (Facchini *et al*. 2000) and may have a role in synthesis of metabolites associated with defences against pathogenic organisms (von Roepenack-Lahaye *et al*. 2003). The IL8-2-1 fruits would be expected to have elevated levels of tyrosine-derived alkaloids. Expression of the *AADC* genes in vegetative tissues may be tied to other defence related secondary metabolites.

The availability of transgenic plants synthesizing a range of phenylalanine-derived volatiles should facilitate evaluation of their roles in many plant-related processes. Both 2-phenylethanol and 2-phenylacetaldehyde attract pollinating insects and repel feeding insects (Zhu *et al*. 2005). It will also be possible to critically evaluate the roles of these volatiles in human taste preferences. Increased production of these important volatiles can be accomplished by expression of a single gene, opening the possibility for engineering enhanced scent production in flowers such as rose where scent has been lost in many varieties in the course of breeding. Finally, these genes will likely be useful as markers for flavour and scent in breeding programs aimed at quality improvement of food and ornamental crops.

6.5 Materials and Methods

6.5.1 Volatile collection

Tomato (*S. lycopersicum* cv. M82), and the *S. pennellii*-derived introgression line IL8-2-1 (Eshed 1994) were grown in the greenhouse or field under standard conditions. Tomato fruit volatiles were collected from approximately 100g chopped ripe tomato fruit with nonyl acetate as an internal standard as described (Schmelz *et al*. 2001). Fruits were enclosed in glass tubes. Air filtered through a hydrocarbon trap (Agilent, Palo Alto, CA) flowed through the tubes for 1h with the aid of a vacuum pump. Volatiles were collected on a Super Q column and subsequently eluted with methylene chloride. Volatiles were separated on an Agilent (Palo Alto, CA) DB-5 column and analyzed on an Agilent 6890Ngas chromatograph (GC); retention times were compared to known standards. Identities of volatile peaks were confirmed by Gas Chromatography/Mass Spectrometry (GC/MS) as described (Schmelz *et al*. 2003). Standards were purchased from Sigma-Aldrich (St. Louis, MO).

6.5.2 *In vivo* phenylalanine decarboxylase assays

Tomato (M82 or IL8-2-1) pericarp discs were incubated with 1 µCi (460 mCi/mmol) [U-^{14}C]phenylalanine (Amersham Bio-sciences, Piscataway, NJ) for 8h in sealed flasks, each with a 1-cm diameter filter paper disk impregnated with 20µl of 2 N KOH suspended in the head-space. $^{14}CO_2$ trapped on the filter paper was quantified by scintillation counting. [^{14}C]phenylalanine and [^{14}C]phenethylamine were extracted from the tissue and separated using an AG-1 (OH^-) column in series with a BioRex-70 (H^+) column as described (Rontein *et al*. 2001). The identity of [^{14}C]phenethylamine was confirmed by co-migration with an authentic standard in thinlayer chromatography on silica gel 60 F_{254} plates in methylenechloride:methanol:triethylamine (80:10:1 v/v/v).

6.5.3 *LeAADC* expression in *E. coli*

Full-length *LeAADC* cDNAs were identified by sequencing putative clones from the TIGR database. Following sequence analysis, the full-length coding sequences were PCR-amplified, and cloned into vector pENTR/D-TOPO. The coding regions were then cloned into vector pDEST15 containing a GST tag (Invitrogen, Carlsbad, CA) by recombination, and transformed into *E. coli* BL21-AI (Invitrogen) for inducible protein expression. Control *E. coli* BL21-AI strains contained pDEST15 with a β-glucuronidase gene inserted. Production of recombinant protein was confirmed by protein blotting with anti-GST antibodies. Activity of each enzyme was determined by growing *E. col*i expressing each LeAADC in media containing 20mM phenylalanine. Volatile compounds were extracted from the cultures using an equal volume of hexanes. Extracts were concentrated and analyzed on an Agilent 6890N gas chromatograph. Identification of phenethylamine was confirmed by GC/MS as described (Schmelz *et al*. 2001).

6.5.4 Protein purification and enzyme assays

The coding region of *LeAADC1A* or *LeAADC2* was cloned into vector pENTR/D-TOPO (Invitrogen). The coding region was then recombined into vector pDEST17 containing a His tag (Invitrogen), and transformed into *E. coli* strain BL21-AI (Invitrogen) for inducible protein expression. To purify His-tagged protein, bacterial cultures were centrifuged at 5000*g* for 5min, followed by sonication for 1min in 1xPBS buffer and centrifugation at 10000 g for 15 min. The His-tagged protein was bound to Ni-NTA resin (Invitrogen) according to the manufacturer's instructions with the inclusion of 200μM pyridoxal 5' phosphate in all solutions. After elution from the Ni-NTA column, the protein was desalted using PD-10 columns (Amersham Biosciences, Piscataway, NJ) equilibrated with 50mM Tris-HCl pH 8.0 containing 200μM pyridoxal 5' phosphate. Protein purity was analyzed by SDS-PAGE followed by staining with Coomassie brilliant blue and protein blotting with anti-His tag antibodies. LeAADC2 was assayed as a crude extract from *E. coli*. The *E. coli* containing the control β-glucuronidase gene had no detectable activity. Decarboxylase activity was determined by the method of Facchini *et al*. (Facchini *et al*. 1999) using [U-^{14}C]phenylalanine or [U-^{14}C]tyrosine as a substrate and measuring the release of $^{14}CO_2$. It was first determined that the

reaction was linear for 4h at 30°C; subsequent assays were run for 3h. To determine the stoichiometry of the reaction [^{14}C]phenethylamine was separated from the reaction mix using the two-column system described above. [^{14}C]phenethylamine data were corrected for recovery from the columns; the recovery was determined to be 60.3% (mean of triplicate observations) using authentic [^{14}C]phenethylamine.

6.5.5 RNA expression analysis

Total RNA was extracted using a Qiagen (Valencia, CA) total RNA extraction kit followed by DNase treatment to remove any contaminating DNA. *LeAADC* mRNA levels were measured by real-time quantitative RT-PCR using Taqman One-Step RT-PCR reagents and a GeneAmp 5700 Sequence Detection System(Applied Biosystems, Foster City, CA). Primers and probe for real-time PCR were as follows: LeAADC1A Taqman probe, 5'-FAM-CCGAACGTGGACA-ACAAGAAACAGAAAATG-3'-BHQ; LeAADC1A forward primer, 5'-AGCGCGACGACGATTGTT-3'; LeAADC1A reverse primer, 5'-GGTCCTGC-ACCTGGTTGTG-3'. LeAADC1B Taqman probe, 5'-FAM-TTTAGCACGACGA-AGATTGTTTCCAAATGTG-3BHQ, LeAADC1B forward primer, 5'-GATTTTGAGC-CATCACCTATGACA-3', LeAADC1B reverse primer, 5'-TGTTCCACCTT-CTGTTTTTTGTTG-3', LeAADC2 Taqman probe, 5'-FAM-TTGGATTGTACATTGATGAATTATATTGATACACTCACCC-3BHQ, LeAADC2 forward primer, 5'-CAGTGACGGAGCCAGGAAA-3', LeAADC2 reverse primer, 5'-TGGATAACCGATATGATAGTTGATACG-3'. For absolute quantification of RNA a standard curve was constructed from *LeAADC* RNAs. *LeAADC* RNAs were synthesized by *in vitro* transcription of the coding sequence as described previously (Tieman *et al*. 2001).

6.5.6 GC/MS analyses of non-volatile plant metabolites

For ^{13}C-labeling studies, pericarp disks were incubated with 10µmol ring-labelled [$^{13}C_6$]phenylalanine (Cambridge Isotope Laboratories, Andover, MA) for 4h. Metabolites were extracted from pericarp tissue and quantified as described previously (Roessner-Tunali *et al*. 2003a), with the exception that for low

abundance metabolites a substantially higher extract concentration (up to 1000-fold) was injected onto the GC/MS. The absolute concentration of metabolites was determined by comparison to standard concentration curves as defined in (Schauer *et al*. 2005b). For metabolite analysis, mass spectral peaks were compared to mass spectral tag (MST) libraries housed in the Golm Metabolome Database (Schauer *et al*. 2005a) (Kopka *et al*. 2005). In addition, the metabolites (phenethylamine, phenylpyruvate, phenyllactate, phenylacetaldehyde) for which no MST information was available were identified by analysis of identically derivatized authentic standards (purchased from Sigma-Aldrich, Munich, Germany). These metabolites were subsequently quantified using the 178, 117, 193 and 193 m/z ions of their derivatives, respectively. For metabolites that could not be detected in fruit extracts, the limit of detection of the method was determined.

6.5.7 Analysis of [U-$^{13}C_6$]phenylalanine-labelled samples

Tomato pericarp discs were extracted as described above. After centrifugation the supernatant was dried under vacuum, and the resulting residue was derivatized for 120min at 37°C (in 50 µl of 20 mg ml^{-1} methoxyamine hydrochloride in pyridine) followed by a 30min treatment at 37°C with 50µl of *N*-Trimethylsilyl-*N*-methyl trifluoroacetamide. GC/MS analysis of the derivatized samples was carried out as described previously (Roessner-Tunali *et al*. 2003a). Uncorrected molar percentage enrichments of metabolites were evaluated as described previously (Giege *et al*. 2003) by comparison of the ^{12}C spectral fragments and the isotopic spectral fractions of non-labelled control incubation with the fragmentation patterns of the ^{13}C-fed tomato pericarp discs as detailed previously (Roessner-Tunali *et al*. 2004). The reaction rates from metabolic precursors through intermediates to end-products was estimated by dividing the amount of label accumulating in the product by the calculated average proportional labelling of the precursor pool.

6.5.8 Transgenic plants

The full-length LeAADC1A or LeAADC2 coding region was cloned in a vector under the control of the Figwort Mosaic Virus 35S promoter (Richins *et al*. 1987)

and followed by the *Agrobacterium* nopaline synthase (*nos*) 3' terminator. The transgene was introduced into tomato cultivar M82 by the method of McCormick *et al*. (Mccormick *et al*. 1986) with kanamycin resistance as a selectable marker. Transgenic plants were grown to maturity in a greenhouse under standard horticultural practices and fruit collected for further analysis.

7 Discussion and Outlook

Whilst each chapter has a discussion focussing on the achievement and the impact for the research contained within it, this final chapter will discuss the thesis integrity and detail its findings in a broader context. In addition it will give an outlook of ongoing and future research planned using the work described herein as its foundation.

7.1 Relation of the work in a broader context

Functional analysis of genes in plants has been facilitated by mutant populations, transgenesis and complete genome sequences. Over the last decades individual genes and their respective function were primarily studied on the basis of their phenotype of the gene, protein or metabolite level. The study of single genes has increased our knowledge, but provided little insight into the relationship and interaction with other genes and corresponding complex phenotypes. Therefore, the exploitation of natural variation in plant biology and, especially, in crop research is becoming highly important. Recent maturation of transcriptomic and proteomic technologies allow the comprehensive phenotypic analysis of genotypes on an integrative level. Further, recent advancements have led to metabolomics platforms which are able to analyse the complex chemical composition of biological systems on a broader scale than was previously obtainable. Up to now this approach has been largely used as a diagnostic tool to study transgenic or mutant lines and stress responses in plants. In diagnostics, metabolite profiling is mostly used in combination with bioinformatic tools to distinguish biological groups. In plant research, a proof of concept study revealed that four different *Arabidopsis thaliana* genotypes could be discriminated by this approach (Fiehn 2000). Similarly an early approach documented that transgenic potato tubers could be distinguished by their differences in the chemical composition using data-mining tools, like hierarchical cluster analysis and principal components analysis (Roessner *et al*. 2001a).

Abiotic and biotic stress responses in plants are relatively poorly understood with most studies highly directed to metabolites that are believed to be of importance. Studies investigating the transcript and metabolite levels on the response of

tomato plants to nitrate starvation (Urbanczyk-Wochniak and Fernie 2005) or the influence of stress factors such as yeast elicitor or methyl jasmonate on *Medicago truncatula* cell cultures (Suzuki *et al*. 2005) or cold stress in *Arabidopsis thaliana* (Kaplan *et al*. 2004) highlight the range of application of metabolite profiling platforms in the characterisation of various stress responses.

Furthermore this approach can be used to study gene functions. For example, the role of an electron-transfer flavoprotein:ubiquinone oxidoreductase (ETFQO) in *Arabidopsis thaliana* could be elucidated by studying insertional mutant lines of *Arabidopsis thaliana* ETFQ during dark induced senescence by an comprehensive approach including transcript and metabolite analysis. This study revealed that ETFQ is needed in plants under dark-induced sucrose starvation where it is involved in both leucine and chlorophyll catabolism (Ishizaki *et al*. 2005). Another such example is the analysis of the Myb-like transcription factor PAP1 by correlative transcript and metabolite profiling, which was discussed earlier (Tohge *et al*. 2005). Metabolite profiling can further be extended to study complex phenotypes of genomic regions, interactions of genes and gene products, in conjunction with gene function annotation (Koornneef *et al*. 2004).

The work presented in this thesis represents the first large scale, multiparallel metabolite analysis combining natural variation, by using introgression lines of *S. pennellii* (ILs), and the metabolite profiling approach to unravel such phenotypic relationships. A prerequisite for this study was the establishment of a MSRI library for the targeted analysis of metabolites in complex biological systems, especially within the *Solanaceae* family (Chapter 3). Eventually, this library was stored in the GMD@CSB.DB (Kopka *et al*. 2005), which is publicly available and serves the metabolomics community as a reference. Application of these libraries within investigation of tomato wild species, that can be easily crossed with the tomato elite cultivar - the *Solanum lycopersicum* complex, allowed the documentation of differences within the natural genetic variance (Chapter 4). These results endorse the strategy to investigate interspecific *S. pennellii* introgression lines on the metabolite level (Chapter 5). The results from this survey are discussed in detail in the following paragraphs as is an example wherein this approach facilitated the identification of genes playing a key role in the biosynthesis of volatile compounds of tomato fruits (Chapter 6).

7.2 QTL analysis

In this study the phenotypic analysis of the ILs at the metabolite level resulted in the characterization of a huge number of metabolic traits. In total over 10,000 metabolic traits and around 1368 morphological traits were recorded over two years harvests. Statistical analysis revealed 889 significant chemical compositional traits and 326 significant morphological traits (ANOVA $P<0.05$). Overall we found 302 correlations between trait pairs ($P<0.0001$) shedding some light on the interconnectivity within the metabolic networks. Currently, the identification of the genetic basis of QTL in tomato is somewhat hampered by the paucity of metabolic genes localized on the genetic map. Although this number is continuously being expanded by the *Solanaceae* research community with numerous genes for fruit size and compositional traits, such as ascorbate, malate and citrate content being localised in the recent years (Causse *et al*. 2004; Zou *et al*. 2006). Despite this paucity of information, elucidation of the genes underlying QTL is achievable in this population and traits can be pinned down to genes. For example Brix9-2-5, a moderate QTL for total soluble solid content has been characterized. This QTL was mapped to a 484bp of the cell wall invertase gene – LIN5 (Fridman *et al*. 2000) and further to a single quantitative trait nucleotide (QTN), that is responsible for different kinetical properties of the enzyme (Fridman *et al*. 2004). A complementary biochemical study revealed that the more highly efficient invertase encoded by the *S. pennellii* allele leads to an increased accumulation of sucrose in the fruit columella and a greater capacity of phloem unloading (Baxter *et al*. 2005a).

Future steps in elucidating the QTL identified in this broad-range metabolite analysis will follow a similar approach as described for the above identification of the Brix9-2-5 (Fridman *et al*. 2000; Fridman *et al*. 2004). Further steps include the generation of tomato lines with smaller introgressed segments of *S. pennellii* to narrow down the responsible region for the QTL in question. Simultaneously, the exploration of the genetics in this set of introgression lines and others, alongside a detailed analysis on the protein level utilizing a high-throughput robotics platform for enzymatic assays (Gibon *et al*. 2004) will hopefully reveal the underlying genetic basis of the QTL of interest.

Despite the recent success in identifying a QTN for Brix9-2-5 it is difficult to predict whether such kinetic effects are a prevalent mechanism that underlies metabolite

QTL. The interaction of hundreds of genes in the ILs could potentially lead to epistatic effects, which could in principle contribute to a QTL. However, the IL population structure used here minimizes this possibility. On the other hand the regulation of pathways appears to be strongly co-ordinately controlled by key regulatory genes and the elucidation of these will give greater understanding of complex biological systems and will proof highly useful for metabolic engineering.

7.3 Compositional quality traits

In tomato, as in many other fruits and vegetables, the chemical composition is a major quality trait on the nutritional and organoleptic level for both the salad market and also for the processing industry. Organic acids and sugar contribute to the total soluble solids or Brix index, which is relevant for both markets. The recent mapping of 15 ascorbate genes to the ILs and co-localisation with identified ascorbate QTL in this and other studies should contribute to future breeding strategies of Vitamin C composition (Rousseaux *et al*. 2005; Zou *et al*. 2006). In the current study several QTL of particular consequences for the chemical composition of the fruit were found and a few examples are highlighted in the following section.

(i) An "*organoleptic*" QTL, with significant increases in citrate and isocitrate, as well as hexoses could be identified in IL5-3 (Figure 14A). This observation is in close agreement with an increased Brix (Eshed and Zamir 1995). Given GC/MS metabolite profiling cannot be readily applied to the study of some glycolic and tricarboxylic acid cycle (TCA) intermediates a second approach was taken to study those compounds. For this purpose IL5-3 was subjected to CE/MS analysis (during a stay at Keio University, Tsuruoka, Japan, 2005). Preliminary CE/MS results revealed no change in glycolic intermediates, but a decrease of TCA intermediates after the conversion to isocitrate. The NAD and adenylate pool sizes were, however, equivalent to that of the elite cultivar, although the ATP: ADP ratio appeared to be slightly lower (unpublished data). These results suggest that the TCA cycle is perturbed downstream of the conversion of aconitate to isocitrate in this line.

(ii) IL1-4 shows an increase of malate content by 173% compared to the elite cultivar Figure 14B). A putative mitochondrial malate dehydrogenase (mMDH)

mapped to the introgressed region on chromosome 1 (Causse *et al*. 2004) and as such represents a plausible candidate gene. The putative mMDH gene from *S. pennellii* and the elite cultivar *S. lycopersicum cv.* M82 were amplified from both species. A subsequent protein sequence comparison of the open reading frames of the *S. pennellii* and *S. lycopersicum cv.* M82 mMDH shows 99% homology with only one amino acid substitution at position P226H in the elite cultivar. Preliminary qRT-PCR experiments suggest that the IL exhibits decreased levels of mMDH transcripts. However, despite these results it is too early to propose the mechanism behind the malate increase. Further, analysis of recombinants of the species introgressions and heterologous expression in yeast KO mutants will be carried out to investigate this gene. It is hoped that this combined approach will reveal (i) whether mMDH is responsible for the phenotype and if so (ii) the precise mechanism by which this is achieved.

7.4 Regulatory traits

Concerted metabolite upregulation is a phenotype observed more than once through in the set of ILs. These pathway regulations seem logical considering the need for full control and regulation of metabolism within a biological system - particularly a sessile one such as a plant. Such coordinated regulation was observed for the branched chain amino acids, valine, leucine and isoleucine (BCAA). Although striking, these observations can be best explained by the fact that the biosynthetic pathways of all three amino acids share four common enzymes. Tight regulation in the flow of carbon is therefore required for the homeostatic biosynthesis of the BCAAs (Singh and Shaner 1995). Since a long time it has been known that *Arabidopsis* exhibits tightly controlled catabolism of BCAAs, thought to be regulated by the first and second enzymes of the degradation pathway, branched-chain amino acids aminotransferase (BCAT-1) and α-ketoacid dehydrogenase (BCKDH) (Graham and Eastmond 2002). Recent studies have revealed that the ETF/ETFQO system participates in the degradation of these amino acids (Ishizaki *et al*. 2005), in analogy to the operation of the human system (Goodman *et al*. 1994), most likely by a mechanism involving electron transfer.

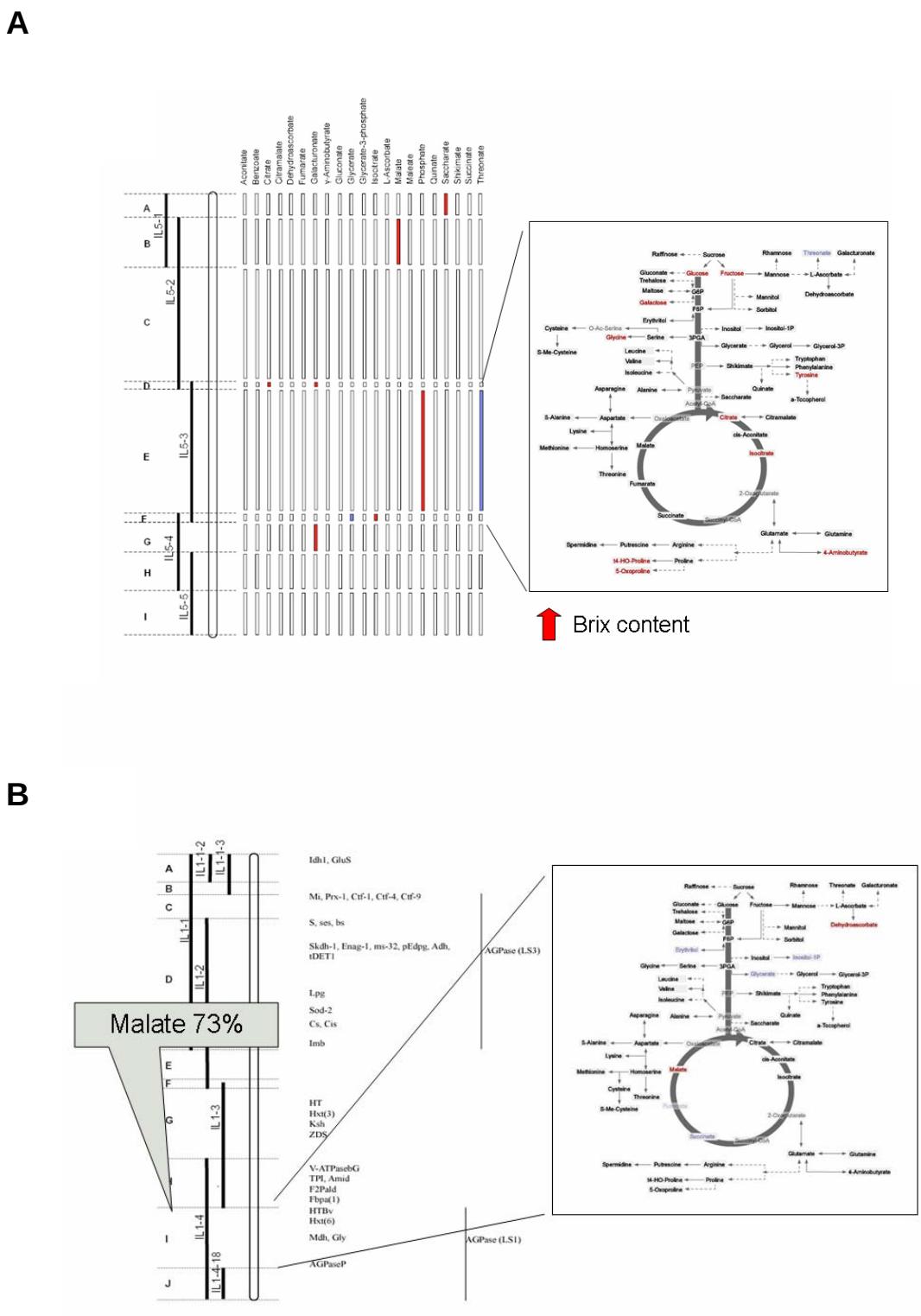

Figure 14. Pathway and chromosomal representation of changes in metabolite content of A) IL5-3 "organoleptic" QTL and B) IL1-4 "Malate" QTL. Red bars or letters represent an increase and blue bars or letters a decrease in metabolite content. The chromosome is visualized alongside introgressd regions (thick black bars) on the left hand side of each figure.

Further steps in elucidating the regulation of BCAA in plants should therefore include localizing map positions of these enzymes to check for the co-localisation with metabolite traits.

Another IL revealed a concerted increase in the levels of several metabolites of the same pathway that from citrate to GABA and proline. Although not easily understood the genetic basis underlying this trait will hopefully lead to the identification of a key regulatory gene. One possible physiological explanation for this phenomenon could be a permanently activated stress response resulting in the stress signals GABA and proline, however, considerable work will be required in order to better understand this phenomenon.

7.5 *Malodorous* QTL

The "*malodorous*" phenotype of IL8-2-1 is characterized by increased levels of the volatiles phenylacetaldehyde and phenylethanol (Tadmor *et al*. 2002). Although of high interest the biosynthesis of volatiles in plants is little understood. A fact is that volatiles in tomato are almost all derived from nutritional important compounds, like amino acids, fatty acids and carotenoids (Goff and Klee 2006). In this thesis some steps of the biosynthesis of phenylethanol could be elucidated using the "*malodorous*" QTL. Headspace GC/MS and stable isotope labelling analysis of tomato fruits from the elite cultivar and IL8-2-1 could show that phenylethanol is largely derived from phenethylamine via phenylacetaldehyde (Chapter 6). Furthermore the high sensitivity of the GC/MS experiment described here suggests that other potential pathways, such as those described for roses (Watanabe *et al*. 2002), are essentially absent in the fruit. In conjunction with mapping studies and a reverse genetic approach two AADCs in tomato could be mapped to the introgressed region on chromosome 8 and via heterologous expression demonstrated to show activity for the decarboxylation of phenylalanine. This study illustrates the power of GC/MS alongside stable isotope labelling experiments for pathway elucidation studies that are fundamental for a full biochemical understanding of the cell.

7.6 Metabolite Assisted Breeding (MAB)

Over the past decades conventional breeding strategies have focused on the identification of phenotypes and selection of genotypes. The process of phenotyping is time-consuming and the selection of appropriate genotypes is difficult, because of the quantitative nature of agronomic traits. Breeding, nowadays, employs not only the fore mentioned methods, but also natural variation, QTL and haplotype studies, alongside modern molecular techniques, like genome sequences, Marker-Assisted Selection (MAS), forward and reverse genetic approaches, as well as transcript profiling and thus is called molecular breeding. MAS is a powerful method to identify genes of interest, associated with a specific phenotype, in breeding populations using molecular markers and defining its location on the chromosome (for a review see (Edmeades *et al*. 2004; Varshney *et al*. 2005)). In conjunction with transcript profiling this strategy appears to be fruitful for the rapid identification of agronomically desirable lines. Nevertheless, due to the complex inventory of a biological system and massive regulatory control the phenotype might not show the desired trait, this can be especially the case for nutritional important metabolic traits. The combination of MAS and metabolite profiling might prove much more effective in the long run for the identification of metabolic traits in crops. However to fully understand the power of this approach inheritance studies are required for the identified QTL. A first preliminary study shows that most of the detected QTL are inherited in a heterozygous F1 IL population in a truly dominant manner. Although, inheritance studies have been carried out at the metabolite level these have generally focussed on single secondary metabolites, such as carotenoids or flavonoids (LeRosen *et al*. 1941; Jones *et al*. 2003) or metabolite classes (Taylor *et al*. 2002). The continuance of the work defined in this thesis will therefore represent the first analysis of inheritance at such a broad level and could be anticipated to shed light on the mode of inheritance of genes in primary metabolism. This pilot study demonstrates that high-throughput and multivariate analysis of metabolites by GC/MS, or other metabolomics platforms, in combination with molecular breeding tools will facilitate the time- and cost intensive process of selection and transferring of agronomically desirable traits into cultivated crops species.

7.7 Conclusion

The application of sophisticated analytical tools has begun to extend the utility of exotic genetic material to the understanding, and ultimately the improvement, of crop compositional quality. The use of natural variation in molecular breeding is clearly a feasible alternative to GM technology particularly given the public concern over the use of GM crops. One further advantage of permanent populations such as ILs is that phenotyping can be carried out at many different levels and laboratories, allowing a comprehensive systems biology approach to analyse the regulation and function of molecular networks. In conjunction future molecular breeding strategies, like MAS, and metabolite profiling will be highly powerful to achieve a more comprehensive view of the mechanisms underlying crop compositional quality.

The analysis of the collected dataset presented here is, however, in its infancy and will likely prove highly informative towards understanding the underlying network of metabolism in tomato fruits. Furthermore these studies indicate that this population can be used to investigate relationships of the cellular inventory of the tomato pericarp cell and its interaction with other organs. Moreover it may increase our knowledge both of key regulatory hubs in scale-free cellular metabolic networks and the mode of inheritance of metabolic traits. When taken together these advances may greatly facilitate crop improvement strategies in the coming years.

8 References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF *et al*. (2000) The genome sequence of *Drosophila melanogaster*. Science 287: 2185-95.

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422: 198-207.

Alba R, Payton P, Fei ZJ, McQuinn R, Debbie P, Martin GB, Tanksley SD, Giovannoni JJ (2005) Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. Plant Cell 17: 2954-2965.

Ausloos P, Clifton CL, Lias SG, Mikaya AI, Stein SE, Tchekhovskoi DV, Sparkman OD, Zaikin V, Zhu D (1999) The critical evaluation of a comprehensive mass spectral library. J Am Soc Mass Spectrom 10: 287-99.

Baldwin EA, Goodner K, Plotto A, Pritchett K, Einstein M (2004) Effect of volatiles and their concentration on perception of tomato descriptors. Journal of Food Science 69: S310-S318.

Baldwin EA, Scott JW, Shewmaker CK, Schuch W (2000) Flavor trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. Hortscience 35: 1013-1022.

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101-13.

Barsch A, Patschkowski T, Niehaus K (2004) Comprehensive metabolite profiling of *Sinorhizobium meliloti* using gas chromatography-mass spectrometry. Funct Integr Genomics 4: 219-30.

Baxter CJ, Carrari F, Bauke A, Overy S, Hill SA, Quick PW, Fernie AR, Sweetlove LJ (2005a) Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant Cell Physiol 46: 425-37.

Baxter CJ, Sabar M, Quick WP, Sweetlove LJ (2005b) 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. Journal of Experimental Botany 56: 1591-1604.

Bentsink L, Yuan K, Koornneef M, Vreugdenhil D (2003) The genetics of phytate and phosphate accumulation in seeds and leaves of *Arabidopsis thaliana*, using natural variation. Theoretical and Applied Genetics 106: 1234-1243.

Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Sayama H, Uhlig J, Zamir D *et al*. (1998) Advanced back-cross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L-pimpinellifolium* (vol 97, pg 170, 1998). Theoretical and Applied Genetics 97: 1191-1196.

Bernal AJ, Pan QL, Pollack J, Rose L, Kozik A, Willits N, Luo Y, Guittet M, Kochetkova E, Michelmore RW (2005) Functional analysis of the plant disease resistance gene Pto using DNA shuffling. Journal of Biological Chemistry 280: 23073-23083.

Berrah G, Konetzka WA (1962) Selective and Reversible Inhibition of Synthesis of Bacterial Deoxyribonucleic Acid by Phenethyl Alcohol. Journal of Bacteriology 83: 738-&.

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH *et al*. (2004) Potential of metabolomics as a functional genomics tool. Trends Plant Sci 9: 418-25.

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF *et al*. (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453-74.

Bolling C, Fiehn O (2005) Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. Plant Physiology 139: 1995-2005.

Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Current Opinion in Plant Biology 7: 132-136.

Bovy A, de Vos R, Kemper M, Schijlen E, Almenar Pertejo M, Muir S, Collins G, Robinson S, Verhoeyen M, Hughes S *et al*. (2002a) High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. Plant Cell 14: 2509-26.

Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, Clarke S, Schofield PM, McKilligin E, Mosedale DE *et al*. (2002) Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med 8: 1439-44.

Brown JKM (2002) Yield penalties of disease resistance in crops. Current Opinion in Plant Biology 5: 339-344.

Brugiere N, Dubois F, Limami AM, Lelandais M, Roux Y, Sangwan RS, Hirel B (1999) Glutamine synthetase in the phloem plays a major role in controlling proline production. Plant Cell 11: 1995-2012.

Burger Y, Saar U, Katzir N, Paris HS, Yeselson Y, Levin I, Schaffer AA (2002) A single recessive gene for sucrose accumulation in Cucumis melo fruit. J Am Soc Hortic Sci 127: 938-943.

Buttery RG (1993) Quantitative and Sensory Aspects of Flavor of Tomato and Other Vegetables and Fruits. In: Acree TE, Teranishi, R., editor. Flavor science: Sensible principles and techniques: Amer. Chem. Soc., Washington, D.C. pp. 259-286.

Buttery RG, Seifert RM, Guadagni DG, Ling LC (1971) Characterization of Volatile Pyrazine and Pyridine Components of Potato Chips. Journal of Agricultural and Food Chemistry 19: 969-&.

Canady MA, Meglic V, Chetelat RT (2005) A library of Solanum lycopersicoides introgression lines in cultivated tomato. Genome 48: 685-697.

Carrari F, Nunes-Nesi A, Gibon Y, Lytovchenko A, Loureiro ME, Fernie AR (2003) Reduced expression of aconitase results in an enhanced rate of photosynthesis and marked shifts in carbon partitioning in illuminated leaves of wild species tomato. Plant Physiol 133: 1322-35.

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55: 1671-85.

Causse M, Saliba-Colombani V, Lecomte L, Duffe P, Rousselle P, Buret M (2002) QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J Exp Bot 53: 2089-98.

Celis JE, Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaars, M., Thykjaer, T., Gromov, P., Yu, J., Pàlsdòttir, H., Magnusson, N., Orntoft, T.F. (2000) gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. FEBS letters 480: 2-16.

Chen F, Duran AL, Blount JW, Sumner LW, Dixon RA (2003) Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis. Phytochemistry 64: 1013-21.

Chen X, Salamini, F. & Gebhardt C. (2001) A potato molecular-function map for carbohydrate metabolism and transport. Theor Appl Genet 102: 284-295.

Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, Hsu HC, Mountz JD, Baldwin NE, Langston MA *et al*. (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet 37: 233-42.

Chetelat RT, Deverna JW, Bennett AB (1995a) Effects of the *Lycopersicon-Chmielewskii* Sucrose Accumulator Gene (Sucr) on Fruit Yield and Quality Parameters Following Introgression into Tomato. Theoretical and Applied Genetics 91: 334-339.

Chetelat RT, Deverna JW, Bennett AB (1995b) Introgression into Tomato (*Lycopersicon-Esculentum*) of the *L-Chmielewskii* Sucrose Accumulator Gene (Sucr) Controlling Fruit Sugar Composition. Theoretical and Applied Genetics 91: 327-333.

Clark GS (1990) Vanillin. Perfumer and Flavorist Journal 15: 45-54.

Cohen JI (2005) Poorer nations turn to publicly developed GM crops. Nature Biotechnology 23: 27-33.

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR *et al*. (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J 37: 778-86.

Concibido VC, La Vallee B, Mclaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theoretical and Applied Genetics 106: 575-582.

Courtney WH, III, Lambeth VN (1977) Glyco alkaloid content of mature green fruit of lycopersicon-spp. Hortscience 12: 550-551.

Davey MW, Keulemans J (2004) Determining the potential to breed for enhanced antioxidant status in Malus: Mean inter- and intravarietal fruit vitamin C and glutathione contents at harvest and their evolution during storage. Journal of Agricultural and Food Chemistry 52: 8031-8038.

Davis CD, Milner J (2004) Frontiers in nutrigenomics, proteomics, metabolomics and cancer prevention. Mutat Res 551: 51-64.

Davuluri GR, van Tuinen A, Fraser PD, Manfredonia A, Newman R, Burgess D, Brummell DA, King SR, Palys J, Uhlig J *et al*. (2005) Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. Nature Biotechnology 23: 890-895.

Davuluri GR, van Tuinen A, Mustilli AC, Manfredonia A, Newman R, Burgess D, Brummell DA, King SR, Palys J, Uhlig J *et al*. (2004) Manipulation of DET1 expression in tomato results in photomorphogenic phenotypes caused by post-transcriptional gene silencing. Plant J 40: 344-54.

De Luca V, St Pierre B (2000) The cell and developmental biology of alkaloid biosynthesis. Trends Plant Sci 5: 168-73.

delaFuente JM, RamirezRodriguez V, CabreraPonce JL, HerreraEstrella L (1997) Aluminum tolerance in transgenic plants by alteration of citrate synthesis. Science 276: 1566-1568.

Demmig-Adams B, Adams WW (2002) Antioxidants in photosynthesis and human nutrition. Science 298: 2149-2153.

Dixon RA, Sumner LW (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol 131: 878-85.

Doebley J, Stec A (1993) Inheritance of the Morphological Differences between Maize and Teosinte - Comparison of Results for 2 F2 Populations. Genetics 134: 559-570.

Duan M, Sun SS (2005) Profiling the expression of genes controlling rice grain quality. Plant Mol Biol 59: 165-78.

Dudley NR, Goldstein B (2005) RNA interference in *Caenorhabditis elegans*. Methods Mol Biol 309: 29-38.

Dumville JC, Fry SC (2003) Gentiobiose: a novel oligosaccharin in ripening tomato fruit. Planta 216: 484-95.

Edmeades GO, McMaster GS, White JW, Campos H (2004) Genomics and the physiologist: bridging the gap between genes and crop response. Field Crop Res 90: 5-18.

Emery GC, Munger HM (1970) Effects of Inherited Differences in Growth Habit on Fruit Size and Soluble Solids in Tomato. J Am Soc Hortic Sci 95: 410-&.

Emmanuel E, Levy, A.A. (2002) Tomato mutants as tolls for functional genomics. Current opinion in Plant Biology 5: 112-117.

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141: 1147-62.

Eshed Y, Zamir D. (1994) Introgressions from *Lycopersicon pennellii* can improve the soluble-solids yield of tomato hybrids. Theor Appl Genet 88: 891-897.

Facchini PJ, Huber-Allanach KL, Tari LW (2000) Plant aromatic L-amino acid decarboxylases: evolution, biochemistry, regulation, and metabolic engineering applications. Phytochemistry 54: 121-38.

Facchini PJ, Yu M, Penzes-Yost C (1999) Decreased cell wall digestibility in canola transformed with chimeric tyrosine decarboxylase genes from opium poppy. Plant Physiology 120: 653-663.

Fedak G (1999) Molecular aids for integration of alien chromatin through wide crosses. Genome 42: 584-591.

Fernie AR (2003) Metabolome characterisation in plant system analysis. Functional Plant Biology 30: 111-120.

Fernie AR, Roessner U, Geigenberger P (2001) The sucrose analog palatinose leads to a stimulation of sucrose degradation and starch synthesis when supplied to discs of growing potato tubers. Plant Physiol 125: 1967-77.

Fernie AR, Tadmor Y, Zamir D (2006) Natural genetic variation for improving crop quality. Curr Opin Plant Biol.

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Innovation - Metabolite profiling: from diagnostics to systems biology. Nature Reviews Molecular Cell Biology 5: 763-769.

Fernie AR, Willmitzer, L. (2004) Carbohydrate metabolism. In: Christou P, Klee, H.K., editor. The Handbook of Plant Biotechnology. Chichester, U.K.: Wiley.

Fiehn O (2002) Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol 48: 155-71.

Fiehn O, Kopka J, Trethewey RN, Willmitzer L (2000) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal Chem 72: 3573-80.

Fiehn O, Kopka, J., Dörmann, P., Altmann, T., Trathewey, R.N., Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. Nature Biotech 18: 1157-1161.

Fischer E, Sauer U (2003) Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. Eur J Biochem 270: 880-91.

Foolad MR, Chen FQ (1999) RFLP mapping of QTLs conferring salt tolerance during the vegetative stage in tomato. Theoretical and Applied Genetics 99: 235-243.

Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Vale G (2005) Marker assisted selection in crop plants. Plant Cell Tissue and Organ Culture 82: 317-342.

Frankel N, Hasson E, Iusem ND, Rossi MS (2003) Adaptive evolution of the water stress-induced gene Asr2 in *Lycopersicon* species dwelling in arid habitats. Mol Biol Evol 20: 1955-1962.

Frary A, Doganlar S, Frampton A, Fulton T, Uhlig J, Yates H, Tanksley S (2003) Fine mapping of quantitative trait loci for improved fruit characteristics from Lycopersicon chmielewskii chromosome 1. Genome 46: 235-43.

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85-8.

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305: 1786-9.

Fridman E, Liu YS, Carmel-Goren L, Gur A, Shoresh M, Pleban T, Eshed Y, Zamir D (2002) Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol Genet Genomics 266: 821-6.

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proceedings of the National Academy of Sciences, USA 4718-4723.

Fridman E, Wang JH, Iijima Y, Froehlich JE, Gang DR, Ohlrogge J, Pichersky E (2005) Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. Plant Cell 17: 1252-1267.

Fulton TM, Bucheli P, Voirol E, Lopez J, Petiard V, Tanksley SD (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica 127: 163-177.

Fustinoni S, Campo L, Colosio C, Birindelli S, Foa V (2005) Application of gas chromatography-mass spectrometry for the determination of urinary ethylenethiourea in humans. J Chromatogr B Analyt Technol Biomed Life Sci 814: 251-8.

Galili G, Hofgen R (2002) Metabolic engineering of amino acids and storage proteins in plants. Metab Eng 4: 3-11.

Gebhardt C, Valkonen JPT (2001) Organization of genes controlling disease resistance in the potato genome. Annu Rev Phytopathol 39: 79-102.

Geigenberger P, Regierer B, Nunes-Nesi A, Leisse A, Urbanczyk-Wochniak E, Springer F, van Dongen JT, Kossmann J, Fernie AR (2005) Inhibition of de novo pyrimidine synthesis in growing potato tubers leads to a compensatory stimulation of the pyrimidine salvage pathway and a subsequent increase in biosynthetic performance. Plant Cell 17: 2077-88.

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JH, Palacios N, Cross J, Selbig J, Stitt M (2004) A Robot-based platform to measure multiple enzyme activities in *Arabidopsis* using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. Plant Cell 16: 3304-25.

Giege P, Heazlewood JL, Roessner-Tunali U, Millar AH, Fernie AR, Leaver CJ, Sweetlove LJ (2003) Enzymes of glycolysis are functionally associated with the mitochondrion in *Arabidopsis* cells. Plant Cell 15: 2140-51.

Gilmour SJ, Sebolt AM, Salazar MP, Everard JD, Thomashow MF (2000) Overexpression of the *Arabidopsis* CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiol 124: 1854-65.

Giovannoni J (2001) Molecular biology of fruit maturation and ripening. Annu Rev Plant Phys 52: 725-749.

Goff SA, Klee HJ (2006) Plant volatile compounds: sensory cues for health and nutritional value? Science 311: 815-9.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. *ssp. japonica*). Science 296: 92-100.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M *et al*. (1996) Life with 6000 genes. Science 274: 546, 563-7.

Gollub J, Ball CA, Binkley G, Demeter J, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaloper M, Matese JC *et al*. (2003) The Stanford Microarray Database: data access and quality assessment tools. Nucleic Acids Res 31: 94-6.

Goodman SI, Axtell KM, Bindoff LA, Beard SE, Gill RE, Frerman FE (1994) Molecular-Cloning and Expression of a Cdna-Encoding Human Electron-Transfer Flavoprotein-Ubiquinone Oxidoreductase. European Journal of Biochemistry 219: 277-286.

Graham IA, Eastmond PJ (2002) Pathways of straight and branched chain fatty acid catabolism in higher plants. Prog Lipid Res 41: 156-81.

Grandilio S, Ku HM, Tanksley SD (1996) Characterization of fs8.1, a major QTL influencing fruit shape in tomato. Molecular Breeding 2: 251-260.

Grandillo S, Zamir D, Tanksley SD (1999) Genetic improvement of processing tomatoes: a 20 years perspective. Euphytica 110: 85-97.

Guimera R, Nunes Amaral LA (2005) Functional cartography of complex metabolic networks. Nature 433: 895-900.

Gullberg J, Jonsson P, Nordstrom A, Sjostrom M, Moritz T (2004) Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. Anal Biochem 331: 283-95.

Gur A, Semel Y, Cahaner A, Zamir D (2004) Real time QTL of complex phenotypes in tomato interspecific introgression lines. Trends in Plant Science 9: 107-109.

Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. Plos Biol 2: 1610-1615.

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19: 1720-1730.

Hazen SP, Hawley RM, Davis GL, Henrissat B, Walton JD (2003) Quantitative trait loci and comparative genomics of cereal cell wall composition. Plant Physiology 132: 263-271.

Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, Moore S, Orchard S, Sarkans U, von Mering C *et al*. (2004) The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data. Nat Biotechnol 22: 177-83.

Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Current Opinion in Plant Biology 4: 210-8.

Holmstrom KO, Somersalo S, Mandal A, Palva TE, Welin B (2000) Improved tolerance to salinity and low temperature in transgenic tobacco producing glycine betaine. Journal of Experimental Botany 51: 177-185.

Holtan HEE, Hake S (2003) Quantitative trait locus analysis of leaf dissection in tomato using Lycopersicon pennellii segmental introgression lines. Genetics 165: 1541-1550.

Huang T, Bohlenius H, Eriksson S, Parcy F, Nilsson O (2005) The mRNA of the *Arabidopsis* gene FT moves from leaf to shoot apex and induces flowering. Science 309: 1694-1696.

Huang XQ, Wang LX, Xu MX, Roder MS (2003) Microsatellite mapping of the powdery mildew resistance gene Pm5e in common wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics 106: 858-865.

Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME (2004) Seed quality QTL in a prominent soybean population. Theoretical and Applied Genetics 109: 552-561.

Isaacson T, Ronen G, Zamir D, Hirschberg J (2002) Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of beta-carotene and xanthophylls in plants. Plant Cell 14: 333-342.

Ishimaru K, Kashiwagi T, Hirotsu N, Madoka Y (2005) Identification and physiological analyses of a locus for rice yield potential across the genetic background. J Exp Bot 56: 2745-53.

Ishizaki K, Larson TR, Schauer N, Fernie AR, Graham IA, Leaver CJ (2005) The critical role of Arabidopsis electron-transfer flavoprotein: Ubiquinone oxidoreductase during dark-induced starvation. Plant Cell 17: 2587-2600.

Jellum E (1977) Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry, with special reference to organic acids. J Chromatogr 143: 427-62.

Jellum E, Helland P, Eldjarn L, Markwardt U, Marhofer J (1975) The development of a computer-assisted search for anomalous compounds (CASAC). J Chromatogr 112: 573-80.

Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino RJ, Hall R *et al*. (2004) A proposed framework for the description of plant metabolomics experiments and their results. Nat Biotechnol 22: 1601-6.

Jeuken MJW, Lindhout P (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theoretical and Applied Genetics 109: 394-401.

Jiye A, Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, Marklund SL, Moritz T (2005) Extraction and GC/MS analysis of the human blood plasma metabolome. Analytical Chemistry 77: 8086-8094.

Jones CM, Mes P, Myers JR (2003) Characterization and inheritance of the Anthocyanin fruit (Aft) tomato. J Hered 94: 449-456.

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32: D277-80.

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol 136: 4159-68.

Kissebah AH, Sonnenberg G, Funahashi T, Jacob HJ, Martin LJ, Blangero J, Matsuzawa Y, Comuzzie AG (2000) Novel genetic pathway for the abdominal obesity metabolic syndrome (AMOS): Preliminary analysis from the TOPS/MRC-OB genes project. Obes Res 8: 4S-4S.

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. Plant Physiology 126: 811-825.

Kliebenstein DJ, Kroymann J, Mitchell-Olds T (2005) The glucosinolate-myrosinase system in an ecological and evolutionary context. Current Opinion in Plant Biology 8: 264-271.

Knudsen JT, Tollsten, L. & Bergstron, L. G. (1993) Floral scents—A checklist of volatile compounds isolated by head—space techniques. Phytochemistry 33: 253-280.

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. Annual Review of Plant Biology 55: 141-172.

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, Dormann P, Weckwerth W, Gibon Y, Stitt M *et al*. (2005) GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 21: 1635-8.

Kovacs MIP, Howes NK, Clarke JM, Leisle D (1998) Quality characteristics of durum wheat lines deriving high protein from a *Triticum dicoccoides* (6b) substitution. Journal of Cereal Science 27: 47-51.

Koyama H, Kawamura A, Kihara T, Hara T, Takita E, Shibata D (2000) Overexpression of mitochondrial citrate synthase in Arabidopsis thaliana improved growth on a phosphorus-limited soil. Plant Cell Physiol 41: 1030-7.

Koziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, Crossland L, Dawson J, Desai N, Hill M, Kadwell S *et al*. (1993) Field Performance of Elite Transgenic Maize Plants Expressing an Insecticidal Protein Derived from *Bacillus Thuringiensis*. Bio-Technology 11: 194-200.

Krapp A, Saliba-Colombani V, Daniel-Vedele F (2005) Analysis of C and N metabolisms and of C/N interactions using quantitative genetics. Photosynthesis Research 83: 251-263.

Kromer JO, Sorgenfrei O, Klopprogge K, Heinzle E, Wittmann C (2004) In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. J Bacteriol 186: 1769-84.

Lamartiniere CA (2000) Protection against breast cancer with genistein: a component of soy. Am J Clin Nutr 71: 1705S-7S; discussion 1708S-9S.

Lecomte L, Saliba-Colombani V, Gautier A, Gomez-Jimenez MC, Duffe P, Buret M, Causse M (2004) Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato. Molecular Breeding 13: 1-14.

Legnani R, Gognalons P, Selassie KG, Marchoux G, Moretti A, Laterrot H (1996) Identification and characterization of resistance to tobacco etch virus in *Lycopersicon* species. Plant Dis 80: 306-309.

LeRosen AL, Went FW, Zechmeister L (1941) Relation between genes and carotenoids of the tomato. P Natl Acad Sci USA 27: 236-242.

Lewinsohn E, Sitrit Y, Bar E, Azulay Y, Ibdah M, Meir A, Yosef E, Zamir D, Tadmor Y (2005a) Not just colors - carotenoid degradation as a link between pigmentation and aroma in tomato and watermelon fruit. Trends in Food Science & Technology 16: 407-415.

Lewinsohn E, Sitrit Y, Bar E, Azulay Y, Meir A, Zamir D, Tadmor Y (2005b) Carotenoid pigmentation affects the volatile composition of tomato and watermelon fruits, as revealed by comparative genetic analyses. Journal of Agricultural and Food Chemistry 53: 3142-3148.

Liu J, Cong B, Tanksley SD (2003a) Generation and analysis of an artificial gene dosage series in tomato to study the mechanisms by which the cloned quantitative trait locus fw2.2 controls fruit size. Plant Physiol 132: 292-9.

Liu YS, Gur A, Ronen G, Causse M, Damidaux R, Buret M, Hirschberg J, Zamir D (2003b) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnol J 1: 195-207.

Liu YS, Roof S, Ye ZB, Barry C, van Tuinen A, Vrebalov J, Bowler C, Giovannoni J (2004) Manipulation of light signal transduction as a means of modifying fruit nutritional quality in tomato. P Natl Acad Sci USA 101: 9897-9902.

Lumba S, McCourt P (2005) Preventing leaf identity theft with hormones. Current Opinion in Plant Biology 8: 501-505.

Maloof JN (2003) Genomic approaches to analyzing natural variation in *Arabidopsis thaliana*. Curr Opin Genet Dev 13: 576-82.

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol 123: 439-42.

Mccormick S, Niedermeyer J, Fry J, Barnason A, Horsch R, Fraley R (1986) Leaf Disk Transformation of Cultivated Tomato (*Lycopersicon-Esculentum*) Using *Agrobacterium-Tumefaciens*. Plant Cell Reports 5: 81-84.

McCouch S (2004) Diversifying selection in plant breeding. Plos Biol 2: 1507-1512.

McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK *et al*. (2001) A physical map of the human genome. Nature 409: 934-41.

Mewes HW, Amid C, Arnold R, Frishman D, Guldener U, Mannhaupt G, Munsterkotter M, Pagel P, Strack N, Stumpflen V *et al*. (2004) MIPS: analysis and annotation of proteins from whole genomes. Nucleic Acids Res 32: D41-4.

Mitchell-Olds T, Gershenzon, J., Baldwin, I., Boland, W. (1998) Chemical ecology in the molecular era Trends in Plant Science 3: 362-365.

Monforte AJ, Asins MJ, Carbonell EA (1997a) Salt tolerance in *Lycopersicon* species .5. Does genetic variability at quantitative trait loci affect their analysis? Theoretical and Applied Genetics 95: 284-293.

Monforte AJ, Asins MJ, Carbonell EA (1997b) Salt tolerance in *Lycopersicon* species .6. Genotype-by-salinity interaction in quantitative trait loci detection: constitutive and response QTLs. Theoretical and Applied Genetics 95: 706-713.

Monforte AJ, Oliver M, Gonzalo MJ, Alvarez JM, Dolcet-Sanjuan R, Arus P (2004) Identification of quantitative trait loci involved in fruit quality traits in melon (*Cucumis melo* L.). Theor Appl Genet 108: 750-8.

Montooth KL, Clark AG, Marden JH (2000) Physiological genetics of flight performance in *Drosophila melanogaster*. Am Zool 40: 1135-1136.

Moose SP, Dudley JW, Rocheford TR (2004) Maize selection passes the century mark: a unique resource for 21st century genomics. Trends Plant Sci 9: 358-64.

Morandini P, Salamini F (2003) Plant biotechnology and breeding: allied for years to come. Trends in Plant Science 8: 70-75.

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y *et al*. (2005a) The SOL Genomics Network. A Comparative Resource for *Solanaceae* Biology and Beyond. Plant Physiol 138: 1310-7.

Mueller LA, Tanksley SD, Giovannoni JJ, van Eck J, Stack S, Choi D, Kim BD, Chen MS, Cheng ZK, Li CY *et al*. (2005b) The Tomato Sequencing Project, the first cornerstone of the International *Solanaceae* Project (SOL). Comp Funct Genom 6: 153-158.

Muller M, Kersten S (2003) Nutrigenomics: goals and strategies. Nat Rev Genet 4: 315-22.

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

Noteborn HPJM, Lommen, A., van der Jagt, R.C., Weseman, J.M. (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. J Biotechnol 77: 103-114.

Nunes-Nesi A, Carrari F, Lytovchenko A, Fernie AR (2005) Enhancing crop yield in *Solanaceous* species through the genetic manipulation of energy metabolism. Biochemical Society Transactions 33: 1430-1434.

Obiadalla-Ali H, Fernie AR, Kossmann J, Lloyd JR (2004) Developmental analysis of carbohydrate metabolism in tomato (Lycopersicon esculentum cv. Micro-Tom) fruits. Physiol Plant 120: 196-204.

Oksman-Caldentey KM, Saito K (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. Curr Opin Biotechnol 16: 174-9.

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL *et al*. (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nature Biotechnology 23: 482-487.

Pan QL, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R (2000) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: Tomato and Arabidopsis. Genetics 155: 309-322.

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124: 735-42.

Pertuze RA, Ji YF, Chetelat RT (2002) Comparative linkage map of the *Solanum lycopersicoides* and S-sitiens genomes and their differentiation from tomato. Genome 45: 1003-1012.

Pillen K, Zacharias A, Leon J (2003) Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 107: 340-352.

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal M, Zamir D, Lifschitz E (1998) The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. Development 125: 1979-89.

Quilot B, Wu BH, Kervella J, Genard M, Foulongne M, Moreau K (2004) QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species P-davidiana. Theoretical and Applied Genetics 109: 884-897.

Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ *et al*. (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19: 45-50.

Rao NK, Reddy LJ, Bramel PJ (2003) Potential of wild species for genetic enhancement of some semi-arid food crops. Genetic Resources and Crop Evolution 50: 707-721.

Rashed MS, Bucknall MP, Little D, Awad A, Jacob M, Alamoudi M, Alwattar M, Ozand PT (1997) Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles. Clin Chem 43: 1129-41.

Ray JD, Kindiger B, Sinclair TR (1999) Introgressing root aerenchyma into maize. Maydica 44: 113-117.

Regierer B, Fernie AR, Springer F, Perez-Melis A, Leisse A, Koehl K, Willmitzer L, Geigenberger P, Kossmann J (2002) Starch content and yield increase as a result of altering adenylate pools in transgenic plants. Nat Biotechnol 20: 1256-60.

Richins RD, Scholthof HB, Shepherd RJ (1987) Sequence of Figwort Mosaic-Virus DNA (Caulimovirus Group). Nucleic Acids Research 15: 8451-8466.

Rick CM (1974) High soluble solids content in large fruited tomato lines derived from a wild green fruited species. Hilgardia 42: 493-510.

Robert VJM, West MAL, Inai S, Caines A, Arntzen L, Smith JK, St Clair DA (2001) Marker-assisted introgression of blackmold resistance QTL alleles from wild *Lycopersicon cheesmanii* to cultivated tomato (*L-esculentum*) and evaluation of QTL phenotypic effects. Molecular Breeding 8: 217-233.

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

Roessner-Tunali U, Liu J, Leisse A, Balbo I, Perez-Melis A, Willmitzer L, Fernie AR (2004) Kinetics of labelling of organic and amino acids in potato tubers by gas chromatography-mass spectrometry following incubation in (13)C labelled isotopes. Plant J 39: 668-79.

Roessner-Tunali U, Urbanczyk-Wochniak E, Czechowski T, Kolbe A, Willmitzer L, Fernie AR (2003b) De novo amino acid biosynthesis in potato tubers is regulated by sucrose levels. Plant Physiol 133: 683-92.

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie A (2001a) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13: 11-29.

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2000a) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant Journal 23: 131-142.

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2000b) Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J 23: 131-42.

Roessner U, Willmitzer L, Fernie AR (2001b) High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies. Plant Physiol 127: 749-64.

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to *beta*-carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold color mutations in tomato. P Natl Acad Sci USA 97: 11102-11107.

Rontein D, Nishida I, Tashiro G, Yoshioka K, Wu WI, Voelker DR, Basset G, Hanson AD (2001) Plants synthesize ethanolamine by direct decarboxylation of serine using a pyridoxal phosphate enzyme. Journal of Biological Chemistry 276: 35523-35529.

Rose LE, Langley CH, Bernal AJ, Michelmore RW (2005) Natural variation m the Pto pathogen resistance gene within species of wild tomato (*Lycopersicon*). I. Functional analysis of Pto alleles. Genetics 171: 345-357.

Rousseaux MC, Jones CM, Adams D, Chetelat R, Bennett A, Powell A (2005) QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theor Appl Genet: 1-13.

Saliba-Colombani V, Causse, M., Langlois, D., Philouze, J., Buret, M. (2001) Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. Theor Appl Genet 102: 259-272.

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10: 297-304.

Sandmeier E, Hale TI, Christen P (1994) Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylases. Eur J Biochem 221: 997-1002.

Sato S, Soga T, Nishioka T, Tomita M (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. Plant J 40: 151-63.

Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. Curr Opin Biotechnol 15: 58-63.

Sauter H, Lauer M, Fritsch H (1988) Metabolic Profiling of Plants - a New Diagnostic-Technique. Abstr Pap Am Chem S 195: 129-AGRO.

Schäfer-Pregl R, Ritter E, Concilio L, Hesselbach J, Lovatti L, Walkemeier B, Thelen H, Salamini F, Gebhardt C (1998) Analysis of quantitative trait loci (QTLs) and quantitative trait alleles (QTAs) for potato tuber yield and starch content. Theor Appl Genet 97: 834-846.

Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K, Roessner-Tunali U, Forbes MG, Willmitzer L *et al*. (2005a) GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett 579: 1332-7.

Schauer N, Zamir D, Fernie AR (2005b) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. J Exp Bot 56: 297-307.

Scheible WR, Pauly M (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. Current Opinion in Plant Biology 7: 285-295.

Schmelz EA, Alborn HT, Banchio E, Tumlinson JH (2003) Quantitative relationships between induced jasmonic acid levels and volatile emission in Zea mays during *Spodoptera exigua* herbivory. Planta 216: 665-673.

Schmelz EA, Alborn HT, Tumlinson JH (2001) The influence of intact-plant and excised-leaf bioassay designs on volicitin- and jasmonic acid-induced sesquiterpene volatile release in *Zea mays*. Planta 214: 171-179.

Scholz M, Kaplan F, Guy CL, Kopka J, Selbig J (2005) Non-linear PCA: a missing data approach. Bioinformatics 21: 3887-3895.

Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 32: D431-3.

Sebolt AM, Shoemaker RC, Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Science 40: 1438-1444.

Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR (2003a) Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theoretical and Applied Genetics 107: 1419-1432.

Septiningsih EM, Trijatmiko KR, Moeljopawiro S, McCouch SR (2003b) Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theoretical and Applied Genetics 107: 1433-1441.

Sickmann A, Mreyen M, Meyer HE (2003) Mass spectrometry--a key technology in proteome research. Adv Biochem Eng Biotechnol 83: 141-76.

Singer JB, Hill AE, Burrage L, Olszens KR, Song JH, Justice M, O'Brien WE, Conti DV, Witte JS, Lander ES *et al*. (2004) Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304: 445-448.

Singh BK, Shaner DL (1995) Biosynthesis of Branched Chain Amino Acids: From Test Tube to Field. Plant Cell 7: 935-944.

Singh KB, Ocampo B (1997) Exploitation of wild *Cicer* species for yield improvement in chickpea. Theoretical and Applied Genetics 95: 418-423.

Slade AJ, Knauf VC (2005) TILLING moves beyond functional genomics into crop improvement. Transgenic Research 14: 109-115.

Soga T, Imaizumi M (2001) Capillary electrophoresis method for the analysis of inorganic anions, organic acids, amino acids, nucleotides, carbohydrates and other anionic compounds. Electrophoresis 22: 3418-25.

Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredez A, Persson S, Raab T *et al*. (2004) Toward a systems approach to understanding plant-cell walls. Science 306: 2206-2211.

Spencer JPE, Kuhnle GGC, Hajirezaei M, Mock HP, Sonnewald U, Rice-Evans C (2005) The genotypic variation of the antioxidant potential of different tomato varieties. Free Radical Research 39: 1005-1016.

Stark DM, Timmerman KP, Barry GF, Preiss J, Kishore GM (1992) Regulation of the Amount of Starch in Plant-Tissues by ADP Glucose Pyrophosphorylase. Science 258: 287-292.

Stein SE (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J Am Soc Mass Spectr 10: 770-781.

Stephanopoulos G, Alper H, Moxley J (2004) Exploiting biological complexity for strain improvement through systems biology. Nat Biotechnol 22: 1261-7.

Stitt M, Fernie AR (2003) From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions. Curr Opin Biotechnol 14: 136-44.

Stitt M, Muller C, Matt P, Gibon Y, Carillo P, Morcuende R, Scheible WR, Krapp A (2002) Steps towards an integrated view of nitrogen metabolism. Journal of Experimental Botany 53: 959-970.

Strelkov S, von Elstermann M, Schomburg D (2004) Comprehensive analysis of metabolites in *Corynebacterium glutamicum* by gas chromatography/mass spectrometry. Biol Chem 385: 853-61.

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62: 817-36.

Suzuki H, Reddy MS, Naoumkina M, Aziz N, May GD, Huhman DV, Sumner LW, Blount JW, Mendes P, Dixon RA (2005) Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. Planta 220: 696-707.

Sweetlove LJ, Last RL, Fernie AR (2003) Predictive metabolic engineering: a goal for systems biology. Plant Physiol 132: 420-5.

Tadmor Y, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E (2002) Identification of *malodorous*, a wild species allele affecting tomato aroma that was selected against during domestication. Journal of Agricultural and Food Chemistry 50: 2005-2009.

Tanksley SD (1993) Mapping Polygenes. Annu Rev Genet 27: 205-233.

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16 Suppl: S181-9.

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277: 1063-6.

Tauberger E, Fernie, A.R., Emmermann, M., Kossmann, J., Willmitzer, L., Trethewey, R.N. (2000) Antisense inhibition of plastidial phosphoglucomutase provides compelling evidence that potato tuber amyloplasts import carbon from the cytosol in form of glucose-6-phosphate. Plant Journal 23: 43-53.

Taylor J, King RD, Altmann T, Fiehn O (2002) Application of metabolomics to plant genotype discrimination using statistics and machine learning. Bioinformatics 18: S241-S248.

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

Tieman DM, Ciardi JA, Taylor MG, Klee HJ (2001) Members of the tomato LeEIL (EIN3-like) gene family are functionally redundant and regulate ethylene responses throughout plant development. Plant J 26: 47-58.

Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P, Kirst M, Klee HJ (2006) Identification of loci affecting flavour volatile emissions in tomato fruits. J Exp Bot: doi:10.1093/jxb/erj074.

Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant Physiology 139: 1125-1137.

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M *et al*. (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J 42: 218-35.

Tolstikov VV, Fiehn, O. (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. Anal Biochem 301: 298-307.

Urbanczyk-Wochniak E, Fernie AR (2005) Metabolic profiling reveals altered nitrogen nutrient regimes have diverse effects on the metabolism of hydroponically-grown tomato (*Solanum lycopersicum*) plants. J Exp Bot 56: 309-21.

Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Rep 4: 989-93.

Van Camp W (2005) Yield enhancement genes: seeds for growth. Current Opinion in Biotechnology 16: 147-53.

Van Der Knaap E, Lippman ZB, Tanksley SD (2002) Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. Theor Appl Genet 104: 241-247.

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10: 621-30.

Venema JH, Posthumus F, de Vries M, van Hasselt PR (1999) Differential response of domestic and wild Lycopersicon species to chilling under low light: Growth, carbohydrate content, photosynthesis and the xanthophyll cycle. Physiologia Plantarum 105: 81-88.

Villareal RL, DelToro E, MujeebKazi A, Rajaram S (1995) The 1BL/1RS chromosome translocation effect on yield characteristics in a *Triticum aestivum* L cross. Plant Breeding 114: 497-500.

Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. Current Opinion in Biotechnology 16: 123-132.

von Roepenack-Lahaye E, Newman MA, Schornack S, Hammond-Kosack KE, Lahaye T, Jones JDG, Daniels MJ, Dow JM (2003) p-coumaroylnoradrenaline, a novel plant metabolite implicated in tomato defense against pathogens. Journal of Biological Chemistry 278: 43373-43383.

Vreugdenhil CJ, Scheper FY, Hoogstraatte SR, Smolders M, Gikunda S, Cobelens FG, Kager PA (2004) Comparison of the parasitologic efficacy of amodiaquine and sulfadoxine-pyrimethamine in the treatment of *Plasmodium falciparum* malaria in the Bungoma district of western Kenya. American Journal of Tropical Medicine and Hygiene 71: 537-541.

Vuralhan Z, Morais MA, Tai SL, Piper MD, Pronk JT (2003) Identification and characterization of phenylpyruvate decarboxylase genes in *Saccharomyces cerevisiae*. Appl Environ Microbiol 69: 4534-41.

Wagner C, Sefkow M, Kopka J (2003) Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. Phytochemistry 62: 887-900.

Wang ZY, Zheng FQ, Shen GZ, Gao JP, Snustad DP, Li MG, Zhang JL, Hong MM (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. Plant J 7: 613-22.

Watanabe S, Hayashi K, Yagi K, Asai T, MacTavish H, Picone J, Turnbull C, Watanabe N (2002) Biogenesis of 2-phenylethanol in rose flowers: incorporation of [2H8]L-phenylalanine into 2-phenylethanol and its beta-D-glucopyranoside during the flower opening of Rosa 'Hoh-Jun' and Rosa damascena Mill. Biosci Biotechnol Biochem 66: 943-7.

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci U S A 101: 7809-14.

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. Science 309: 1056-1059.

Wu GH, Truksa M, Datla N, Vrinten P, Bauer J, Zank T, Cirpus P, Heinz E, Qiu X (2005) Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. Nature Biotechnology 23: 1013-1017.

Xu YB, McCouch SR, Zhang QF (2005) How can we use genomics to improve cereals with rice as a reference genome? Plant Molecular Biology 59: 7-26.

Yates HE, Frary A, Doganlar S, Frampton A, Eannetta NT, Uhlig J, Tanksley SD (2004) Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. Euphytica 135: 283-296.

Ye XD, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287: 303-305.

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. *ssp. indica*). Science 296: 79-92.

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nature 2: 983-989.

Zeh M, Casazza AP, Kreft O, Roessner U, Bieberich K, Willmitzer L, Hoefgen R, Hesse H (2001) Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. Plant Physiology 127: 792-802.

Zhu J, Obrycki JJ, Ochieng SA, Baker TC, Pickett JA, Smiley D (2005) Attraction of two lacewing species to volatiles produced by host plants and aphid prey. Naturwissenschaften 92: 277-281.

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

Zou LP, Li HX, Ouyang B, Zhang JH, Ye ZB (2006) Cloning and mapping of genes involved in tomato ascorbic acid biosynthesis and metabolism. Plant Science 170: 120-127.

9 Supplementary information

Supplementary information is provided online. The data can be retrieved from www.mpimp-golm.mpg.de/fernie.

Supplementary Figure 1. Heat maps of the metabolite profiles of the introgression lines in comparison to that of the parental control (*S. lycopersicum*) from the individual data sets of A) 2001 and B) 2003. Large sections of each map are white or pale in colour reflecting the fact that many of the chromosomal segment substitutions do not have a large affect on the level of every metabolite. Regions of red or blue indicate that the metabolite content is either increased or decreased respectively following the introgression of S. pennellii segments. A total of 74 metabolites were quantified by gas chromatography-mass spectrometry for each harvest, including amino acids, organic acids, fatty acids, sugars, sugar alcohols and vitamins.

Supplementary Figure 2. The cartography algorithm is stochastic, so different runs yield, in principle, different partitions. To test the robustness of the algorithm, we obtain 20 partitions of the network depicted in Figure 2 and plot, for each pair of nodes in the network, the fraction of times that they are classified in the same module. As shown in the figure, most pairs of nodes are either always classified in the same module (red) or never classified in the same module (dark blue), which indicates that the solution is robust.

Supplementary Figure 3. HI and BX levels in three different genotypes of the recessive self-pruning (SP) allele of tomato plants in comparison to their respective wild type control plants. Tomato 3 genotypes were IL6-3 and the elite cultivars Gardener and VFNT. Isogenic seeds of the three genotypes were planted on a field in Israel. Values are presented as an average of replicate plants per line; * indicates values that were determined by *t*-test to be significantly different ($P < 0.05$) from the reference genotype.

Supplementary Figure 4. Amino acid alignments of the three aromatic amino acid decarboxylases.

Supplementary Figure 5. Levels of mRNAs in ripe fruit from LeAADC1A (6306, 6404 and 6410) and LeAADC2 (6610, 6612 and 6721) transgenic overexpressing lines.

Supplementary Table 1. Metabolite QTL table - Levels of significant ILs ($p<0.05$) in 74 metabolic traits. Values are presented as percent of control means of two independent harvests (each harvest was normalized to the mean response calculated for *S. lycopersicum*) with each six independent replications. The table is sorted by IL and than by trait name.

Supplementary Table 2. Yield associated QTL table - Levels of significant ILs ($p<0.01$) of 9 morphology traits. 5 of them (Plant weight (PW), Total yield (TY), Brix (BX), Fruit weight (FW), Harvest index (HI)) were measured in 2 harvests (2004 and 2000). The remaining 4 traits were measured only in the 2004 harvest (Seed #/fruit (SN/fruit), Seed #/plant (SN/plant), Fruit length, Fruit width). Values are presented as percent of control means of two (or one) independent harvests (each harvest was normalized to the mean response calculated for *S. lycopersicum*). Six replications in 2000 and 9 in 2004 were considered. The table is sorted by IL and than by trait name.

Supplementary Table 3. Association between pairs of traits as is determined by correlation (using IL means) and by shared QTLs (chi method – see methods). The 'type' column determine the pairs' type (met – for 2 metabolites; mor – for 2 morphology; mix – for one morphology and one metabolite). 'Correlation' is the correlation coefficient value, 'count' is the correlation sample size and 'prob' is its probability. 'QTL (left)' is the number of QTLs of the first trait (on the left), and 'QTL (right)' is the number of QTL of the second traits (right). The next 3 columns were calculated using the Chi method (see methods for details):'shared QTLs' is the number of shared QTLs, 'association' determine whether the association between traits is positive (1) or negative (-1), and 'binomial probability' is the probability to get the observed shared QTLs by chance alone.

Supplementary Table 4. Dependence of metabolite QTLs in morphology traits. This table lists all 889 metabolite QTLs, and indicates for each of them whether it is associated by any of the morphological traits (1 indicates association of the corresponding morphology trait, and 0 indicates it is not associated). The table is sorted decreasingly according to the number of morphology traits associated to each QTL. Determination of association was done as follows: Each pair of metabolite and morphology traits were analyzed according to the Chi method (see methods and Supplementary Table 3), if the association was significant (5%), all the shared QTLs of these pair of traits were considered as associated to morphology.

Supplementary Table 5. Limits of detection of phenylalanine and phenylalanine-derived metabolites.

Deutsche Zusammenfassung

Die Einführung von Hochdurchsatzmethoden zur Analyse von biologischen Systemen, sowie die umfangreiche Sequenzierung von Genomen haben zu einer Verlagerung der Forschung „im Detail“ zu einer ganzheitlicheren Betrachtungsweise auf Systemebene geführt. Aus einer jahrhundertlangen, intensiven Züchtung und Selektion von Nutzpflanzen resultierte gleichzeitig eine Abnahme der genetischen Varianz. Daraus resultierend sind Nutzpflanzen anfälliger gegenüber Stressfaktoren, wie Pathogenen, hohen Salzkonzentrationen oder Trockenheit, als ihre Wildarten. Das Potential konventioneller Züchtung scheint somit heute an seine Grenzen gekommen zu sein. Daher versucht man mittels moderner Molekulartechnik, wie zum Beispiel Marker-gestützte Selektion, Gene oder ganze Genombereiche von Wildarten mit hoher genetischer Variation in Nutzpflanzen einzukreuzen, vornehmlich mit dem Ziel einer Ertrags- bzw. Resistenzsteigerung. Neueste medizinische Studien belegen, dass die Ernährung eine wesentliche Rolle für die menschliche Gesundheit spielt. Besonders wichtig sind hierbei die gesundheitsfördernden Substanzen in pflanzlichen Nahrungsmitteln. Aus diesem Grund kommt der Erforschung der biochemischen Zusammensetzung von biologischen Proben eine immer größere Bedeutung zu. Diese Untersuchung kann elegant durch Metabolitenprofile, welche die multivariate Analyse komplexer biologischer Proben erlauben, durchgeführt werden.

In dieser Arbeit wurde zur Untersuchung der biochemischen Zusammensetzung von Tomatenwildarten und interspezifischen *S. pennellii* Tomatenintrogressions-linien (IL) eine GC/MS basierte Metabolitenanalyseplattform verwendet. Hierzu war es zunächst notwendig eine Massenspektrenbibliothek, zur Annotierung von Massenspektren und Retentionsindices von, in pflanzlichen Proben vorkommenden, Metaboliten anzulegen. Die Analyse der Tomatenwildarten ergab große Unterschiede gegenüber der Kulturtomate im Hinblick auf den Gehalt an Amino- und organischen Säuren, sowie der Zuckerzusammensetzung und den Gehalt an Sekundärmetaboliten. Die darauf folgende Analyse der ILs, von den jede ein genau definiertes genomisches Segment von *S. pennellii* beinhaltet, bestätigte diese enorme Variation mit 889 metabolischen und 326 ertragsassozierten-Veränderungen in den ILs. Die metabolischen Veränderungen zeichneten sich durch abnehmende bzw. steigende Gehalte von einzelnen Metaboliten, aber auch durch eine koordinierte Änderung aus. In dieser Arbeit wurde weiterhin der Biosyntheseweg der Volatilenstoffe Phenylethanol und Phenylacetaldehyd mit Hilfe einer IL untersucht. Hierbei konnten durch stabile Isotopenmarkierung und eines „reverse genetics“-Ansatzes Gene bzw. Enzyme identifiziert werden, die für die Dekarboxylierung des Eduktes Phenylalanin verantwortlich sind. Diese Arbeit beschreibt erstmals die umfassende Analyse von biochemischen Komponenten auf Genombasis in Tomatenintrogressionslinien und zeigt damit ein Werkzeug auf zur Identifizierung von qualitativen biochemischen Merkmalen in der modernen molekularen Züchtung.

Curriculum vitae

First name:	Nicolas
Family name:	Schauer
Nationality:	German
Date of birth:	April 03, 1976
Place of birth:	Düsseldorf, Germany
Address:	Chausseestr. 19, 10115 Berlin, Germany
Email:	schauer@mpimp-golm.mpg.de

Current position:

04/2003-04/2006	PhD thesis "Quantitative Trait Loci (QTL) for Metabolite Accumulation and Metabolic Regulation: Metabolite Profiling of Interspecific Crosses of Tomato" at Max-Planck-Institute of Molecular Plant Physiology in Potsdam-Golm, Germany under supervision of Prof Dr Lothar Willmitzer & Dr Alisdair R Fernie.

University education:

10/1998-03/2003	Biology (Diploma) at University of Tübingen.
03/2003	Diploma thesis "The Regulation of P-type H^+ATPase by 14-3-3 Proteins" in Biology at University of Tübingen, Germany, Faculty of Biology, ZMBP under supervision of Prof Dr Wolf B Frommer and Prof Dr Claudia Oecking.

School education:

1982-1994	Rudolf-Steiner Schule, Düsseldorf
08/1994-08/1997	Kollegschule Kikweg, Düsseldorf Graduation, Abitur (HSC)

Fellowships:

08/2004	IPP-IPS – International PhD program - Integrative Plant Science, University of Potsdam and MPI-MP.
06/2005	Short-term fellowship of the Japanese Society of Promoting Science (JSPS).
Other activities:	ECBA, vdbiol

Publication list

Peer reviewed articles

Geigenberger P, Regierer B, Lytovchenko A, Leisse A, **Schauer N**, Springer F, Kossmann J, Fernie AR (2004). Heterologous expression of a *keto*hexokinase in potato plants leads to inhibited rates of photosynthesis, severe growth retardation and abnormal leaf development. Planta 218: 569-78.

Bender-Machado L, Bauerlein M, Carrari F, **Schauer N**, Lytovchenko A *et al*. (2004). Expression of a yeast acetyl CoA hydrolase in the mitochondrion of tobacco plants inhibits growth and restricts photosynthesis. Plant Molecular Biology 55: 645-662.

Ishizaki K, Larson TR, **Schauer N**, Fernie AR, Graham IA, Leaver CJ (2005). The critical role of *Arabidopsis* electron-transfer flavoprotein: Ubiquinone oxidoreductase during dark-induced starvation. Plant Cell 17: 2587-2600.

Lytovchenko A, **Schauer N**, Willmitzer L, Fernie AR (2005). Tuber-specific cytosolic expression of a bacterial phosphoglucomutase in potato (*Solanum tuberosum* L.) dramatically alters carbon partitioning. Plant Cell Physiol 46: 588-97.

Kopka J, **Schauer N**, Krueger S, Birkemeyer C, Usadel B *et al*. (2005). GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 21: 1635-8.

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

Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G *et al*. (2005b). GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett 579: 1332-7.

Carrari F, Coll-Garcia D, **Schauer N**, Lytovchenko A, Palacios-Rojas N, Balbo I, Rosso M, Fernie AR (2005). Deficiency of a plastidial adenylate kinase in *Arabidopsis* results in elevated photosynthetic amino acid biosynthesis and enhanced growth. Plant Physiol. 137: 70-82.

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (in press). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol.

Ottmann C, Marco, S, Jaspert N, **Schauer N**, Vandermeeren C, Boutry M, Wittinghofer A, Rigaud JL, Oecking C (submitted). Structure of a 14-3-3 coordinated hexamer of the plant plasma membrane H+-ATPase by combining x-ray crystallography and electron cryo-microscopy.

Junker BH, Wuttke R, Nunes-Nesi A, Steinhauser D, **Schauer N**, Bussis D, Willmitzer L, Fernie AR (2006). Enhancing Vacuolar Sucrose Cleavage within the Developing Potato Tuber has Only Minor Effects on Metabolism. Plant Cell Physiol.

Hackel A, **Schauer N**, Carrari F, Fernie AR, Grimm B, Kuhn C (2006). Sucrose transporter LeSUT1 and LeSUT2 inhibition affects tomato fruit development in different ways. Plant Journal 45: 180-192.

Schauer N, Soga T, Baran R, Tomita M, Fernie AR (in preparation). Combinatorial elucidation of the tomato metabolome by GC/MS and CE/MS.

Tieman D, Taylor M, **Schauer N**, Fernie AR, Hanson AD, Klee HJ (accepted). Aromatic amino acid decarboxylases participate in synthesis of the flavor and aroma volatiles, 2-phenylethanol and 2-phenylacetaldehyde in tomato fruits. PNAS.

Ishizaki K, **Schauer N**, Larson TR, Graham IA, Fernie AR, Leaver CJ (submitted). The functional association between *Arabidopsis* electron transfer flavoprotein (ETF) and electron transfer flavoprotein ubiquinone oxidereductase (ETFQO) during dark induced starvation. Plant Journal.

Book chapters and invited reviews

Schauer N, Fernie, AR (2005) Intraspezifische Introgressionen – wie die Genetik biochemische Phänotypen beeinflusst. GenomXPress 05: 14-16.

Carrari F, **Schauer, N**, Willmitzer, L, Fernie, AR (2006) Systems Biology: A Renaissance of the Top-down Approach for Plant Analysis. In: Saito K, Dixon, R.A., Willmitzer, L., editor. Plant Metabolomics. Berlin Heidelberg: Springer Verlag.

Schauer N, Willmitzer, L, Fernie, AR (2006b) Essay: Metabolic Profiling of Plant Cells. In: Taiz L, Zeiger, E., editor. Plant Physiology. 4th Edition ed: Sinauer Associates Inc.

Schauer N, Fernie, AR (in preparation) Biological function and mechanism: application of metabolomics in plant biochemistry and genetics. Trends in Plant Science.