Review

Application of Metabolomics to Improve Tomato Fruit Productivity and Quality

Yoko Iijima and Koh Aoki*

Kazusa DNA Research Institute, Kazusa-Kamatari, Kisarazu 292-0818, Japan

Productivity and quality of tomato (Solanum lycopersicum) fruit largely depend on the quantity and composition of its metabolites. Recent development of mass spectrometry technology has facilitated the simultaneous detection of a large number of metabolites. This opens the possibility of investigating the mechanisms of metabolic regulation that affect fruit productivity and quality, in conjunction with genetic and omics approaches. In this article, we review recent progress in the analysis of tomato fruit metabolites based on comprehensive metabolic profiling and the identification of genes and quantitative trait loci that alter tomato fruit metabolite profiles.

Key Words: metabolomics, primary metabolite, secondary metabolite, tomato, volatile.

Introduction

A major goal of tomato metabolite analysis is to elucidate genes and biosynthetic pathways that can be exploited for crop productivity and quality improvement. Tomato productivity involves the net accumulation of photosynthetically synthesized compounds, such as sugars, and the polymeric compounds derived from them, cellulose and starch. Fruit quality is determined by a complex blend of traits including flavor, color, and texture. These traits are mostly attributed to metabolite composition. For example, lipid, sugar, amino acid, alcohol, aldehyde, and isoprenoid contents affect fruit flavor. Levels of carotenoids and flavonoids influence fruit color. Thus, productivity and quality can be defined in terms of metabolic profiles.

Recent development of chromatography-coupled mass spectrometry has revealed complex profiles of tomato metabolites. This facilitates a comparison of metabolite composition across cultivar or variety at the whole metabolite level (Saito et al., 2008c). It also allows for an investigation of the relationship between metabolic phenotype and genotype. Additionally, in conjunction with parallel analysis of transcripts, a comprehensive metabolite analysis allows for the identification of genes involved in metabolic regulation (Carrari et al., 2006; Saito et al., 2008b; Schauer et al., 2006). Thus, a metabolomics approach is a valuable part of tomato functional genomics that aims to modify metabolic pathways affecting fruit productivity and quality.

In this review, we present an overview of the status of tomato metabolomics. First, a brief explanation of the technical aspects of metabolite extraction and detection is provided. We then describe recent advances in the metabolomic analyses of primary and secondary (including non-volatile and volatile) metabolites based on comprehensive profiling approaches. Finally, application of metabolomics to phenomics-assisted breeding is discussed.

Tools developed for metabolomics

Table 1 lists the mass spectrometry abbreviations that appear in this section.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>CE</td>
<td>capillary electrophoresis</td>
</tr>
<tr>
<td>EI</td>
<td>electron ionization</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>APCI</td>
<td>atmospheric pressure chemical ionization</td>
</tr>
<tr>
<td>APP1</td>
<td>atmospheric pressure photo ionization</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>Q</td>
<td>quadrupole</td>
</tr>
<tr>
<td>IT</td>
<td>ion trap</td>
</tr>
<tr>
<td>TOF</td>
<td>time-of-flight</td>
</tr>
</tbody>
</table>

Received; September 4, 2008. Accepted; November 20, 2008.
* Corresponding author (E-mail: kaoki@kazusa.or.jp).
1) Metabolite extraction

The tomato contains a diverse range of metabolites that have different molecular sizes and polarities. Optimum extraction conditions differ for different metabolite types. To extract carbohydrates, amino acids and fatty acids for measurement with gas chromatography-mass spectrometry (GC-MS), an extraction protocol originally developed for the potato tuber (Roessner et al., 2000) is widely used. Tomato tissues are extracted with methanol, which gives final methanol concentrations ranging from 85% to 93% (Urbanczyk-Wochniak et al., 2003). The extracts are then fractionated into polar and non-polar phases using chloroform. A detailed review of extraction conditions is available for Arabidopsis samples for GC-MS (Gullberg et al., 2004). To extract secondary metabolites for non-targeted analysis by liquid chromatography-mass spectrometry (LC-MS), an extraction using 75% methanol (final concentration) is widely used (Bino et al., 2005; Iijima et al., 2008a; Moco et al., 2006). However, more lipophilic extraction is required to extract non-polar secondary metabolites such as carotenoids (Fraser et al., 2000; Long et al., 2006). To extract volatiles, solid phase micro extraction (Tikunov et al., 2005), and solvent extraction (e.g., propanol, pentane) (Schmelz et al., 2003) were established.

2) GC-MS

GC-MS is applicable to volatiles, such as alcohols, monoterpenes, and esters, as well as non-volatile polar metabolites (primary metabolites), such as amino acids, sugars, lipids, and organic acids through derivatization by methylation or trimethylsilylation. The electron ionization (EI) technique that is usually coupled with GC provides reproducible fragmentation patterns. In addition, the retention index procedure (Kovats, 1958) compensates for the fluctuation of retention times by normalizing them with respect to the internal standards. Metabolites can be identified by comparing fragment patterns and retention indices with those of standard compounds in compound databases.

3) LC-MS

LC-MS is a versatile technology used to analyze non-volatile secondary metabolites. Most LC-MS instruments are equipped with ion sources for electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and/or atmospheric pressure photo ionization (APPI). Using these ionization methods, LC-MS detects quasi-molecular ions and is suitable for analyzing relatively large metabolites (m/z 500–2000). Additionally, many LC-MS instruments are equipped with MS/MS, which enables the fragmentation of precursor ions that provide useful information for predicting partial metabolite structures. High-resolution MS systems, such as time-of-flight (TOF), Fourier transform ion cyclotron resonance (FTICR) and orbitrap Fourier transform (orbitrap FT) coupled with the high-precision LC are proving useful in analyzing secondary plant metabolites containing large numbers of isomers (Bino et al., 2005; Iijima et al., 2008a; Iijima et al., 2008b; Suzuki et al., 2008). However, in contrast to GC-MS, the lack of a searchable LC-MS database is one of the obstacles in identifying metabolites by LC-MS.

4) CE-MS

Capillary electrophoresis-mass spectrometry (CE-MS) has been employed to analyze water-soluble metabolites, particularly organic acids, nucleotides, amino acids, sugars, and sugar phosphates. Although the application of CE-MS to plant metabolite profiling is relatively limited (Sato et al., 2004; Takahashi et al., 2006; Williams et al., 2007), CE-MS will be used in a broader field of application due to its applicability to the polar metabolites and ease of sample preparation.

5) NMR

Nuclear magnetic resonance (NMR) is frequently used for metabolic profiling of tomatoes (Mattoo et al., 2006). Unfortunately, the application of NMR to plant metabolomics is currently limited to the profiling of major metabolites because of its low sensitivity compared to MS-based methods. Ongoing improvements in instrumentation include the use of multi-dimensional NMR to increase resolution (Fan et al., 2001; Kikuchi and Hirayama, 2007), and the coupling of LC separation with NMR (Bailey et al., 2000). Although NMR remains underused in plant metabolomics, NMR’s advantages have been demonstrated for the analysis of metabolic flux by combining a stable isotope labeling method (Sekiyama and Kikuchi, 2007).

6) Analytical strategies

The data acquired by MS or NMR are subjected to data mining procedures. There are two strategies for data mining: metabolic profiling, which essentially quantifies the metabolites present in a given biological sample (Hall, 2006), and metabolite annotation, which aims to label peaks with metadata and reliable predictions with respect to the identity and structure of the metabolites (Fiehn et al., 2005; Iijima et al., 2008a).

In metabolic profiling, all detected peaks, including unknown ones, can be quantified and utilized as unannotated variables for further statistical analysis. Approaches such as principal component analysis and hierarchical clustering are the most widely used to clarify the difference or similarity between the metabolite profiles of multiple samples.

Metabolite annotation aims at the collection and integration of a given metabolite’s information to provide chemical and structural specification. Unlike transcripts and proteins, metabolites cannot be identified on the basis of the genome sequence. Thus, integrated interpretation of mass spectral data is needed for
specifying metabolites detected with MS (Iijima et al., 2008a). Metabolite annotation facilitates the systematic analysis of unknown metabolites and interprets their relationship to known metabolites.

We need to mention that no single extraction/detection technology is sufficient for the complete coverage of metabolites. Additionally, some high molecular weight-metabolites that affect fruit quality, such as polysaccharides, pectins and lignins, are not targets of above mentioned MS technologies. Thus, combination of multiparallel technologies and development of new analytical technology are required to obtain a broader picture of tomato metabolome.

**Application of metabolomics approaches to tomato**

1) **Tissue-dependent accumulation of metabolites**

Detailed metabolic profiling of tomato fruit tissues at different developmental stages has demonstrated spatial and temporal specificity in the accumulation of endogenous metabolites (Moco et al., 2007). The authors reported that metabolic variation between tissues was more pronounced than developmental variation. For example, carotenoids, flavonoids, glycoalkaloids and ascorbic acid are abundant specifically in the epidermis. Abscisic acid-hexose and zeatin-hexose were detected at relatively high levels in the jelly parenchyma. The level of total chlorophyll was highest in the vascular attaching region at post-breaker stages. This study provides an experimental basis to investigate tissue-specific regulation of metabolic pathways.

2) **Primary metabolites**

The expectation has been that cultivated varieties and wild species of tomatoes display a large variation in metabolite content. A comparison of the profiles of primary metabolites obtained by GC-MS analysis in *S. lycopersicum* and five wild species (*S. pimpinellifolium*, *S. neorickii*, *S. habrochaites*, *S. chmielewskii*, and *S. pennellii*) has been reported (Schauer et al., 2005). Fruit metabolite profiles showed a higher level of variance than leaves. Accumulation levels of sucrose, isocitrate, chlorogenate, and shikimate were higher in the wild species than in *S. lycopersicum*. On the other hand, accumulation levels of glucose, fructose, α-tocopherol, and putrescine were lower in the wild species than in *S. lycopersicum*.

Studies using tomato introgression line (IL) populations have identified quantitative trait loci (QTL) for the accumulation levels of primary metabolites. An IL population crossing *S. pennellii* and *S. lycopersicum* (‘M82’), consisting of 76 near-isogenic lines, is one of the most intensely studied IL for agronomic traits (Eshed and Zamir, 1995). Some loci that are related to tomato metabolism, such as the Brix 9-2-5 (Fridman et al., 2000), fvo2.2 (Alpert and Tanksley, 1996), and B locus (Ronen et al., 2000), were mapped using this *S. pennellii* IL population. Based on the quantification of 74 metabolites from 76 introgression lines, the presence of 889 QTL, covering 74 metabolites, was demonstrated (Schauer et al., 2006). The authors reported that the metabolites whose IL contents were higher than *S. lycopersicum* were not limited to the metabolites that were more abundant in *S. pennellii* than *S. lycopersicum*. The authors combined the analytical results from harvests in 2001, 2003, and 2004, and estimated the hereditability and the mode of inheritance of the metabolic QTL (Schauer et al., 2008). Estimation of heritability (Semel et al., 2006) demonstrated that sugars, inositol, glycerol, and fatty acids exhibit strong heritability, while stress-related metabolites such as proline, shikimate, aconitate, fumarate, and malate exhibit low heritability. This finding demonstrates the different influences of environment on metabolic variation. Estimation of the mode of inheritance (recessive, additive, dominant, and overdominant) demonstrated that the majority of metabolic QTL exhibit an additive or dominant mode of inheritance. Limited numbers of metabolites (e.g., glycine, leucine, valine, γ-aminobutyrate, glycerol, glucose, and adenosine-5-phosphate) exhibited recessive QTL. Overdominant QTL were rarer still (e.g., cysteine, homoserine, benzoate, citrate, dehydroascorbate, glucose, mannose, trehalose, phosphate, and squalene). These results provide many practical implications for crop improvement strategies.

3) **Secondary metabolites**

The secondary metabolites can be classified into two groups: non-volatile and volatile. Since the analysis of non-volatile and volatile secondary metabolites requires LC-MS and GC-MS techniques, respectively, a separate description of studies on non-volatile and volatile secondary metabolites is provided in this section.

Metabolites detected using LC-ESI-FTICR-MS in a miniature tomato, ‘Micro-Tom’, were subjected to a comprehensive metabolite annotation procedure (Iijima et al., 2008a). Metabolite annotations including predicted molecular formula, predicted structure and database-hits were attached to 869 individual metabolites. A database search demonstrated that 494 out of the 869 metabolites do not match compounds registered in public metabolite databases such as KEGG, KNApSaCk, and PubChem (September, 2007), suggesting that they are novel metabolites. Based on the m/z value and molecular formula predictions, modification of metabolites by amino group, caffeate, malonate, and hexose appears to occur frequently. Additionally, new metabolic pathways were hypothesized for flavonoid- and glycoalkaloid-groups. Particularly, the presence of a biosynthetic pathway from tomatine to esculeoside A was supported by the finding of intermediate metabolites on the pathway, which were also reported in tomato metabolic profiling studies using LC-Q-TOF-MS (Mintz-Onon et al., 2008; Moco et al., 2007). According to the
glycoalkaloid profiling using ripening mutants (Iijima et al., 2008a), it was suggested that the conversion of tomatine to esculeoside A depended on the ethylene, of which biosynthesis is controlled by internal feedback mechanisms (Inaba, 2007).

To identify QTL for the content of secondary metabolites, particularly antioxidants such as ascorbate, phenolic compounds, lycopene, and β-carotene, the S. pennellii IL populations (Eshed and Zamir, 1995) were analyzed (Rousseaux et al., 2005). The authors identified six QTLs for ascorbate, nine for total phenolics, four for lycopene, and two for β-carotene. However, none of the QTL for lycopene increased the content relative to a parental S. lycopersicum. The authors also pointed out that the identified QTL were under strong environmental control, since most of the QTL were not consistently observed throughout the three harvests, or were only observed in field trials and not greenhouse trials. This result was expected, since many genes involved in antioxidant biosynthesis are controlled by environmental factors including oxidative stress, temperature, and light. Thus, with stricter control of growth environment, it is anticipated that additional QTL will be identified for a broader range of secondary metabolites.

Genetic engineering of secondary metabolism pathways (Fig. 1) has been a major focus in tomato fruit research. Genetic manipulation of carotenoid pathway genes is of special interest since the pathway produces the predominant tomato fruit pigments, including lycopene and β-carotene (Fraser et al., 2002; Ralley et al., 2004; Tian et al., 2004). The Phytoene synthase 1 gene (Psy-1) is responsible for carotenoid formation during tomato fruit ripening. To characterize the effect of Psy-1’s constitutive expression, quantitative, comprehensive metabolite data were obtained in parallel with gene expression data and enzymatic activity data (Fraser et al., 2007). The results demonstrate that constitutive expression of Psy-1 enhances the accumulation of carotenoids (particularly, from phytoene to β-carotene) in green fruit. Early depletion of chlorophyll with increased carotenoid levels in Psy-1 overexpressing fruit suggested that geranylgeranyl diphosphate is preferentially used in carotenoid rather than chlorophyll formation.

The flavonoid pathway is another well-studied target of tomato genetic manipulation. Tomato fruit that overexpressed the Petunia chalcone isomerase gene (chi-a) accumulated higher levels of flavonols (quercetin- and kaempferol-glycosides) in the peel (Muir et al., 2001). Overexpression of maize transcription factors, LC and Cl, resulted in the increase of kaempferol and naringenin glycosides in the tomato’s flesh (Bovy et al., 2002; Le Gall et al., 2003a, 2003b). To manipulate the composition of flavonoids in tomato fruits, genetically modified tomatoes were produced that overexpressed the foreign flavonoid-related genes: stilbene synthase, chalcone reductase (together with chalcone synthase), and flavone synthase (together with chalcone isomerase) (Schijlen et al., 2006). In these transgenic tomatoes, flavonoids that are not normally present in tomato fruit, including stilbenes (resveratrol and piceid), deoxychalcones (butein and isoliquiritigenin), and flavones (luteolin and luteolin-7-glucoside) were successfully produced. This result shows that redirection of the

Fig. 1. Pathways of tomato secondary metabolism. Compounds in gray box indicate predominant metabolites in tomato fruit tissues. Abbreviations: -PP, diphosphate; Psy-1, Phytoene synthase 1 gene; chi-a, chalcone isomerase gene.
flavonoid pathway is possible by genetic engineering.

To manipulate carotenoid and flavonoid pathways simultaneously, Davuluri and coworkers engineered a regulatory gene, *TDET1* (Davuluri et al., 2005). The *TDET1* gene, a tomato ortholog of *Arabidopsis DETIOLATED1*, which is proposed to be a negative regulator of light signal transduction, is responsible for the phenotype of the tomato high pigment-2 (*hp-2*) mutant, in which a high accumulation of flavonoids, anthocyanins and carotenoids was observed (Bino et al., 2005; Mustilli et al., 1999). Fruit-specific RNAi-mediated silencing of *TDET1* resulted in elevated levels of lycopene, β-carotene, naringenin chalcone, quercetin derivatives, and chlorogenate up to 3.5-fold compared with wild type fruit (Davuluri et al., 2005). The whole pathway of carotenoid and flavonoid was affected by *TDET1*-silencing, suggesting that the precursor supply is not a rate-limiting step in this transgenic tomato.

4) Volatiles

The flavor of fresh tomato fruit is a mixture of various volatile metabolites (Tieman et al., 2006b). Although the biosynthetic pathways of several major volatiles have not been elucidated, flavor volatiles of tomato are believed to be synthesized from precursors, including: phenylalanine (*Phe*), leucine/isoleucine (*Leu/Ile*), lipids, isoprenoids, and carotenoids (Fig. 2). A comprehensive analysis of tomato fruit volatiles was performed using GC-MS with solid phase microextraction (SPME), and resulted in the detection of as many as 322 different volatiles from 94 different varieties (Tikunov et al., 2005). The first implication of this comprehensive volatile analysis was that the accumulation of Phe-derived volatiles is associated with the fruit-type difference between the cherry tomato and the round/ fleshy tomato. Secondly, a clustering analysis of volatile accumulation patterns across tomato varieties revealed that volatiles derived from the same precursor showed a similar accumulation pattern. This result is in accordance with the notion that the metabolic system is organized in a modular manner.

An extensive analysis of tomato QTL for volatile levels using *S. pennelli* IL populations (Eshed and Zamir, 1995) identified 25 loci for 23 volatiles (Tieman et al., 2006b). Most of the identified loci were altered in multiple volatiles that associated with the same metabolic pathway. In the case of carotenoid-derived volatiles, there was good agreement between carotenoid accumulation levels and carotenoid-derived volatile levels. These results are useful for breeding programs aimed at the improvement of fruit flavor.

Studies combining metabolomic and genomic approaches have identified the genes responsible for the production of major volatiles of tomato fruit. For Phe-derived volatiles, genes encoding the key-enzyme aromatic amino acid decarboxylase (AADC), were identified and designated as *LeAADC1* and *LeAADC2* (Tieman et al., 2006a). Genes encoding the key enzymes...
for biosynthesis of carotenoid-derived volatiles, carotenoid cleavage dioxygenase (CCD), have been identified based on the sequence similarity to Arabidopsis AtCCD1 (Schwartz et al., 2001), and are designated as LeCCD1A and LeCCD1B (Simkin et al., 2004). LeCCD1 proteins utilize both linear and cyclic carotenoids as substrates. For terpenoid-derived volatiles, genes encoding monoterpene synthase were identified and designated as LeMTS1 and LeMTS2 (van Schie et al., 2007). LeMTS1 protein generated monoterpene linalool by using the precursor geranyl diphosphate, and sesquiterpene nerolidol by using the precursor farnesyl diphosphate. On the other hand, LeMTS2 protein generated several monoterpene products including β-phellandrene, β-myrcene, and sabinene by using the precursor geranyl diphosphate, whereas did not show any sesquiterpene synthase activity.

The application of genetic manipulation has extended to the introduction of genes from other plant species to alter the tomatoes’ volatile composition. Expression of the geraniol synthase gene of lemon basil (Ocimum basilicum) under the tomato polygalacturonase gene promoter enhances the production of geranyl diphosphate-derived volatiles (geranyl acetate, citronellol, citronellal, and rose oxide) in ripe fruit (Davidovich-Rikanati et al., 2007). Similarly, expression of the sesquiterpene synthase gene from lemon basil enhances the production of farnesyl diphosphate-derived volatiles, including α-zingiberene and α-bergamotene (Davidovich-Rikanati et al., 2008). These examples clearly demonstrate that improvement of tomato flavor is feasible by genetic engineering.

**Metabolite analysis in phenomics-assisted breeding**

1) Integrated omics

To link metabolomic data to functional genomic study, subsequent steps include the assignment of metabolites to biological pathways. The possibility of linking metabolite profiles with transcript profiles has been recently explored (Alba et al., 2005; Carrari et al., 2006). Carrari and co-workers performed a comprehensive parallel analysis of transcripts and metabolites during tomato fruit development (Carrari et al., 2006). The authors estimated metabolite-to-metabolite, transcript-to-transcript, and metabolite-to-transcript correlations. A comparison between the metabolite-to-metabolite and the transcript-to-transcript correlation profiles demonstrated that, in metabolite-to-metabolite correlation, the metabolites belonging to the same class are regulated in a highly coordinated manner, whereas the transcripts belonging to the same functional group displayed less coordinated behavior. Moreover, metabolite-to-transcript correlation demonstrated that the metabolite levels of certain metabolic pathways displayed low correlation with the transcript levels that are responsible for that pathway. These results suggest that metabolism is largely regulated at the posttranscriptional level. The authors, however, identified several strong correlations between metabolites and transcripts in ripening-associated processes.

2) Phenotypic databases

With the progress of tomato metabolomics, the development of a database is becoming more and more important, not only as a repository of large-scale data, but also as a tool to interpret metabolite data at a whole metabolite level. Standards for presenting and exchanging data were proposed for metabolomics, Minimum Information About a METabolomics experiment (MIAMET), so that significant results can be accessible to the community (Bino et al., 2004). In Table 2, we list tomato metabolite databases currently open to public domain. The next step in database development will be

<table>
<thead>
<tr>
<th>Database</th>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tomato Metabolomics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KOMICS (<a href="http://webs2.kazusa.or.jp/komics/">http://webs2.kazusa.or.jp/komics/</a>](<a href="http://webs2.kazusa.or.jp/komics/">http://webs2.kazusa.or.jp/komics/</a>)</td>
<td>(Iijima et al., 2008a)</td>
<td>A database of mass spectra and metabolite annotation experimentally obtained by using LC-FTICR-MS.</td>
</tr>
</tbody>
</table>

*URLs are valid on November 17, 2008.*
fully automated and interoperable analytical services. Such development will allow a synergetic integration of tomato omics data.

From a breeding-related point of view, a metabolite database will have a potential use as a phenotypic catalog for screening candidate cultivars/varieties for breeding. The analysis of metabolic phenotype will be combined with appropriate molecular markers for loci controlling metabolite content, which will in turn facilitate phenomics-assisted breeding. The analysis of metabolic phenotype will also provide beneficial returns to the cultivation techniques (Saito et al., 2008a). Additionally, a comparison of the comprehensive profiling data will possibly demonstrate the substantial equivalence of newly bred cultivars/varieties with that of parental varieties.

Literature Cited


