Xiaoquan Qi · Xiaoya Chen
Yulan Wang *Editors*

Plant Metabolomics

Methods and Applications

Chemical Industry Press
Springer

Plant Metabolomics

Xiaoquan Qi · Xiaoya Chen
Yulan Wang
Editors

Plant Metabolomics

Methods and Applications

Chemical Industry Press
Springer

Editors
Xiaoquan Qi
Institute of Botany
Chinese Academy of Sciences
Beijing
China

Yulan Wang
Wuhan Institute of Physics and Mathematics
Chinese Academy of Sciences
Wuhan, Hubei
China

Xiaoya Chen
Shanghai Institute for Biological Sciences
Chinese Academy of Sciences
Shanghai
China

ISBN 978-94-017-9290-5 ISBN 978-94-017-9291-2 (eBook)
DOI 10.1007/978-94-017-9291-2

Library of Congress Control Number: 2014951658

Springer Dordrecht Heidelberg New York London

Jointly published with Chemical Industry Press, Beijing
ISBN: 978-7-122-21436-2, Chemical Industry Press, Beijing

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Life sciences progress quickly with each passing day. The improvement of genomics and related analytical techniques greatly promoted the rapid development of transcriptomics, proteomics, metabolomics, and phenomics. Thus, the means of system integration can be employed to reveal life phenomenon at multiple levels. The above research thoughts and methods gave birth to systems biology. Metabolomics is an important part of systems biology. Metabolites are closest to phenotype, thus the change in metabolites can more directly reveal the function of genes. And metabolic markers have important application values in the early diagnosis of diseases. There are a wide range of plant species in nature. Different groups of plant species synthesize different special compounds. It is estimated that there are 0.2–1 million kinds of metabolites synthesized by plants. The structural and physicochemical properties vary widely, making plant metabolomics research more challenging. Since the year 2002 when the first International Plant Metabolomics Conference was held in Wageningen, analytical techniques and methods of plant metabolomics have been developing rapidly and applied in several areas, such as plant scientific research, biotechnology safety assessment, crop breeding, etc., and play important roles in the study of gene function and the analysis of metabolic pathway and metabolic network regulation. Plant metabolomics research in China started around 2005, and currently has a good development trend. This book written in cooperation by researchers active in plant metabolomics in China, not only introduces the latest advances in plant metabolomics and analyzes the development trend in the next few years, but also demonstrates new studies of authors in their respective scientific projects, reflecting the current study level of China very well.

This book includes three parts introducing and demonstrating plant metabolomics. The first part includes an overview of plant metabolomics and the principles, methods, issues, considerations, and developments of metabolite analytical technologies, which mainly include mass spectrometry and nuclear magnetic resonance; the second part includes metabolomics data analysis, metabolites determination, metabolomics database and metabolic network study; the third part includes detailed application examples in plant metabolomics, which mostly are the current research

achievements in recent years. We strive to be realistic and practical in this book, and hope that this book can promote the rapid development of plant metabolomics in China. Immense thanks to the authors of each chapter for taking the time from busy research and teaching tasks. Many authors of this book received the funding of "Metabolism and Regulation of Special Crop Nutrients (2007CB108800)" and the funding of "The Formation Mechanism and Control Approaches of Harmful Substances in Animal Products (2009CB118800)" from "973" plans of Ministry of Science of China. This book received the funding of "Metabolism and Regulation of Special Crop Nutrients" project office, too. Warm thanks to Prof. Chun-Ming Liu from the Institute of Botany of Chinese Academy of Sciences for his concern and help in the publishing of this book. Readers are welcome to criticize and correct errors and shortcomings in this book.

November 2010

Xiaoquan Qi
Xiaoya Chen
Yulan Wang

Contents

Chapter 1
Overview

Xiaoya Chen, Xiaoquan Qi and Li-Xin Duan

Metabolism reflects all the (bio) chemical changes during life activities, and metabolic activity is the essential characteristics and material basis of life. The central dogma of molecular biology believes that life information flows from deoxyribonucleic acids (DNAs) to messenger ribonucleic acids (mRNAs), then to proteins, and then to metabolites catalyzed by enzymes (mostly protein enzymes). And finally, these products converge and interact to produce a wide variety of biological phenotypes. DNA, as the carrier of life information, plays a crucial role in above-mentioned process. Genomics that comprehensively analyzing the constitution and function of DNAs in various species is the earliest 'omics.' Genomics research has greatly accelerated the development of life sciences. The success of genomics has also promoted the development of many other 'omics,' such as transcriptomics, proteomics, metabolomics, and phenomics (Fig. 1.1). Therewith systems biology that integrating above-mentioned 'omics' means to multi-levelly and comprehensively reveal biological phenomena come into being.

Metabolomics or metabonomics aims to study all the small molecular metabolites and their dynamic changes in an organism or a tissue or even a single cell (Oliver et al. 1998; Fiehn 2002). As early as in 300 AD, the ancient Greeks realized that they can predict disease through observing changes of body fluids or tissues. The idea is consistent with metabolomics in disease diagnosis (Nicholson and Lindon 2008). As an interdisciplinary combination of organic chemistry, analytical chemistry, chemometrics, informatics, genomics and transcriptomics, metabolomics has penetrated into all aspects of life sciences research. Metabolomics is a very important node of systems biology and the nearest node to phenotype. Therefore,

X. Chen (✉)
Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
e-mail: xychen@sibs.ac.cn

X. Qi · L.-X. Duan
Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
e-mail: xqi@ibcas.ac.cn

L.-X. Duan
e-mail: nlizn@ibcas.ac.cn

X. Qi et al. (eds.), *Plant Metabolomics*, DOI 10.1007/978-94-017-9291-2_1

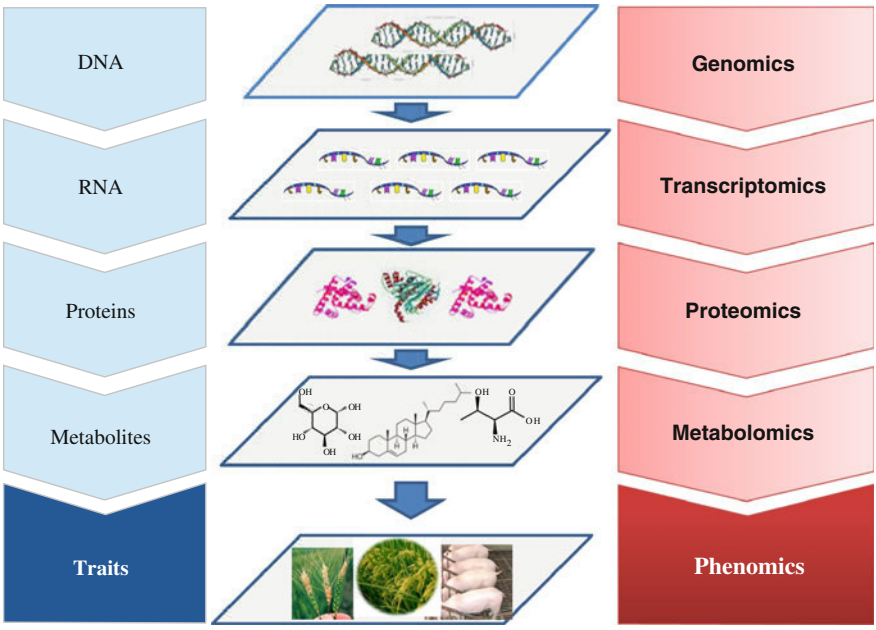

Fig. 1.1 Metabolomics is a branch of systems biology. The multilevel and systematic study is expected to reveal the molecular and metabolic basis for the formation of plant traits

metabolomics research is able to reveal gene function more holistically, thereby providing scientific basis to the application of biotechnology.

Plant metabolomics is one of important parts of metabolomics research. It is estimated that there are about 0.2–1 million metabolites produced by about 0.3 million known plant species in the world (unknown plant species are not included yet) (Dixon and Strack 2003). Of note is that the structures of secondary metabolites are widely different. In terms of the current (or after a period of time) level of instrument analysis, there is no any analytic method that is able to detect all the metabolites, which makes plant metabolomics research more challenging. Other chapters in this book will detail the analysis techniques of plant metabolites, data analysis methods, and application examples of metabolomics. This chapter summarizes the main developments and challenges that plant metabolomics face.

1.1 The Development of Metabolite Analysis Techniques

1.1.1 The Automation of Sample Preparation

Sampling, metabolite extraction, and pretreatment (derivatization) are the three critical steps of sample preparation in metabolomics and the premise for obtaining reliable data. In order to rapidly and efficiently sample and extract metabolites and keep the

metabolites with good uniformity and stability, generally add the extraction solution rapidly after the plant tissue is quickly freeze in liquid nitrogen and ground into powder. The commonly used extraction solutions are methanol–chloroform–water, methanol–isopropanol–water, and methanol–water–formic acid. The extract must be dried and derivatized prior to be analyzed by gas chromatography–mass spectrometry (GC–MS). A two-step derivatization method is commonly used: The first step is adding methoxyamine pyridine solvent to reduce the cyclization of reductive sugar and protect carbonyl group; the second step is trimethyl silanization reaction to reduce the boiling point of analyte. Commonly used silanization reagents are bis (trimethylsilyl) trifluoroacetamide (BSTFA) and N-Methyl-N- (trimethylsilyl) trifluoroacetamide (MSTFA). The silanization effect of BSTFA and MSTFA is similar, but as the boiling point of MSTFA is lower than BSTFA, the chromatographic peak times of silanization reagents and by-products is earlier, and thus, the influence to metabolite analysis is less. Prior to be analyzed by liquid chromatography–mass spectrometry (LC–MS), the extract needs to be filtrated to remove insolubles to prevent blocking the separation column. Some efficient sampling, extraction, and derivatization methods have been developed to realize the globality, reproducibility, and high throughput that are required for metabolomics analysis. For example, (Weckweth et al. 2004) based on methanol–chloroform–water (2.5:1:1, v/v/v) extraction solution simultaneously extracted metabolites, proteins, and RNAs from 30 to 100 mg Arabidopsis (*Arabidopsis Thaliana*) fresh leaves, which can be subjected to metabolomics, proteomics, and transcriptomics analysis, respectively, providing convenience for systems biology research.

The automated sampling, extraction, and pretreatment technology come into being to prevent large deviation caused from the tedious and complicated sample preparation process (Nikolau and Wurtele 2007). The online derivatization and autoinjection system armed with the multifunctional autosampler decreases error that is caused from the tedious manual derivatization and the time difference of derivatization. The mechanization and automation of sample preparation technology is the tendency, because it can minimize the experimental error, thus making data more stable and reproductive.

1.1.2 Plant Single-Cell Separation Technology and High-Resolution Imaging Mass Spectrometry Techniques

At present, the main sources of plant metabolomics samples are plant organs, tissues, and suspension cultured cell lines which contain plant cells with different types and in different developmental stages and are treated with different environmental and experimental stimuli or treatment. The types and quantities in different functional cells are different. Therefore, metabolomics samples are different among themselves (Saito and Matsuda 2010). Furthermore, metabolites can be transported through vascular bundle between plant tissues. For example, methionine glucosinolate mainly accumulates in Arabidopsis seeds and buds, but the

biosynthesis-related genes are found expressing in internode vascular bundle (Nour-Eldin and Halkier 2009). Although traditional methods of sampling and extraction in spite of the difference of developmental stages and cell types are quick and easy in sample preparation, to a greater degree, it weakens the ability of metabolomics in plant uncovering life activities.

The single-cell separation technology and high-resolution imaging MS are the trend of plant metabolomics development. With the combination of precise cell and tissue separation technology and ultra-high sensitive detection technology, researchers can study the intracellular metabolism and the intercellular transportation of metabolites. Recently, the imaging MS is used to detect the spatial distribution of metabolites, which uses a continuous laser to scan the surface of the plant tissue. The excitation of the laser makes metabolites of the surface of the plant tissue ionized and detected by mass spectrometer. And the distribution of metabolites in tissue or single cell can be observed by means of mass spectrum intensity. An example of such a technology is the use of the colloidal graphite-assisted laser desorption ionization mass spectrometry (MS) imaging technology to detect specific accumulation of flavonoids in flowers and petals (Saito and Matsuda 2010).

1.1.3 The Comprehensive Two-Dimensional Gas Chromatography High-Resolution Time-of-Flight Mass Spectrometry Coupling Technique (GC × GC-TOF/MS)

The separation and detection of metabolites are two core components of plant metabolomics analysis technology. The separation technique consists of a variety of chromatographic separation methods, such as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE). And detection technique consists of MS and nuclear magnetic resonance (NMR). The efficient combination of above two techniques is able to meet the fundamental needs of plat metabolomics analysis. Metabolomics aims at comprehensively, qualitatively, and quantitatively analyzing all the metabolites of samples. The problem is that there are about hundreds to thousands of metabolites in a single plant cell and different tissues, different cell types, different subcellular organelles and cells biosynthesize and accumulate different metabolites. In the meanwhile, the type and content of metabolites that are biosynthesized and accumulated are influenced by the changes of developmental stages and the difference of growth environment. Metabolites extracted from plant tissue materials have a huge number of types (at least thousands of types), complicated structures, various analogues, and extremely wide range of variation in content (it is estimated that the range of content variation is about 10^7) (Hegeman 2010). To comprehensively, qualitatively, and quantitatively analyze metabolites, the separation and detection instruments must have such features as good stability, good qualitative ability to compounds, high resolution and sensitivity, fast detection speed, and wide dynamic range of detection.

MS is such a technique that by detecting the size and abundance of charge-to-mass ratio of molecular ions or fragment ions after the ionization of interested components analyzes interested components qualitatively and quantitatively. MS is classified into various types depending on the type of mass analyzer. Compared with other analysis techniques, MS has higher sensitivity, faster detection speed, and wider dynamic range and can be combined with other techniques, such as GC and LC, thereby improving the analysis ability to complex matrices (Ekman et al. 2009). The time-of-flight (TOF) MS technique is superior to the quadrupole MS detector that was commonly used in earlier plant metabolomics analysis. The TOF detector analyzes the charge-to-mass ratio of different ions according to the difference of flight time of charged ions in a vacuum flight tube with extremely high sensitivity and fast scan speed (the data acquisition rate can achieve up to 500 full spectrograms per second). Thereby, it is beneficial for the fast analysis and improving the effect of spectrogram deconvolution. The dynamic range of mass detection can be over 10^5. With the high-resolution (a general capillary column has about more than 1 million theoretical plates) and good stability of GC, when GC is coupled with TOF/MS detector, the required qualification of metabolomics will be achieved. Generally, a standard voltage for electron impact ionization (−70 eV) is employed, the mass spectrum for a compound is usual stable and can be used as the structure characteristics of the compound. Generic compound libraries, such as National Institute of Standards and Technology Mass Spectral Database, have been established, which greatly ease the difficulties of compound qualitative analysis in the plant metabolomics research. Weckwerth and colleagues (Weckwerth et al. 2004) quantified about 1,000 compounds using GC time-of-flight mass spectrometer (GC-TOF/MS) and discriminated sugar biosynthesis isomerase mutant by metabolic network analysis. Wagner and colleagues qualitatively analyzed metabolites by the retention indices and mass spectrum of GC-TOF/MS (Wagner et al. 2003) and established a plant metabolome database, i.e., the Golm Metabolome Database (Kopka et al. 2005).

The maturely developed comprehensive two-dimensional GC (GC × GC) further strengthens the ability to detect complex metabolites. GC × GC is such a technique that connecting in series two chromatographic columns which with different stationary phases and are independent of each other. Each component separated from the first column is trapped and focused by modulator and then enter into the second column in pulse mode. The second column is very short, so separation is quick. Then, the separated component is subjected to MS scan in a speed of up to 500 spectrograms per second to get the two-dimensional GC data. Comprehensive two-dimensional GC combined with time-of-flight MS with high separation capacity and sensitivity is one of the most powerful separation tool widely used in the separation and analysis of complex systems such as metabolomics (Wang et al. 2010). Recently, Zoex Company introduced a instrument which combine comprehensive two-dimensional GC with high-resolution time-of-flight mass spectrometer detector, GC × GC-HiResTOF/MS. The resolution of TOF/MS is 4,000–7,000. The mass precision can be in three decimal places. The precise mass number can be used to speculate molecular formula. The highly precise mass of

fragment ion peaks makes the deconvolution of overlap peaks more accurate and easier. As a result, the qualitative ability of MS to compounds is greatly enhanced. It can be predicted that this type of equipment will be widely applied in plant metabolomics research and will be a member of the plant metabolomics technology mainstream.

Components that are suitable for GC analysis are those that can be easily vaporized, which have low polarity and low boiling point. The examples of such components are various volatile compounds, or those have low boiling point after derivatization, such as amino acids, organic acids, sugars, and alcohols. GC can detect some primary metabolites of plant extract, but GC–MS alone cannot fully reveal the changes of all the plant metabolites.

1.1.4 Ultra-High-Performance Liquid Chromatography with Tandem Quadrupole Time-of-Flight Mass Spectrometry

Compared with GC, LC is not influenced by the volatility and thermal stability of the sample, pretreatment of sample is very simple, and injection can be directly done after filtering. Therefore, LC combined with MS can effectively analyze the abundant plant secondary metabolites, including various terpenoids, alkaloids, flavonoids, and glucosinolates. LC–MS-based analytical equipments have been progressed to be the indispensable analytical equipment for metabolomics.

There are a variety of MS types combined with LC, such as quadrupole mass spectrometry (Q/MS), tandem triple quadrupole mass spectrometry (QQQ/MS), ion trap mass spectrometry (IT/MS), time-of-flight mass spectrometry (TOF/MS), tandem quadrupole time-of-flight mass spectrometry (Q-TOF/MS), tandem ion trap time-of-flight mass spectrometry (IT-TOF/MS), fourier transform ion cyclotron resonance mass spectrometry (FT-ICR/MS), and linear trap quadrupole orbitrap mass spectrometer (LTQ-Orbitrap/MS). There are also a variety of ion sources that can be used for LC-MS, such as electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), matrix-assisted laser desorption ionization (MALDI), and atmospheric pressure photo ionization (APPI). And there are a lot of scan modes, such as selected ion monitoring (SIM), selected reaction monitoring (SRM), multiple reaction monitoring (MRM), full scan and tandem mass spectrometry MS/MS, or multistage MS MS^n scanning. Among many MS types, the high-resolution quadrupole time-of-flight mass spectrometer (Q-TOF/MS) is able to meet the plant metabolomics research requirement to the greatest extent. The present Q-TOF has a scanning speed of about 20 spectrograms per second. The latest triple TOF has a scanning speed of 100 spectrograms per second and a resolution of over 40 thousands and a wide dynamic range of over 10^5. LC-Q-TOF/MS has become the widely used analytical equipment for plant metabolomics research and has been successfully used in tomato metabolomics research (Moco et al. 2006). Diode array

(PDA) detector was connected behind LC, and ESI source was used for systematic analysis of metabolites with moderate polarity in tomato. With the combination of retention time, accurate mass, UV spectra, and double MS, a tomato metabolite database MoTo DB was built. Similarly, a comprehensive analysis of metabolites in Arabidopsis root and leaf was done with the use of Waters CapLC combined with Q-TOF/MS (von Roepenack-Lahaye et al. 2004). LC-Q-TOF/MS has been successfully used for the analysis of metabolite changes in vegetative growth of 14 ecotypic accessions and 160 recombinant inbred lines of *A. thaliana* and found that 75 % of the mass peaks are stably heritable and are mapped to Arabidopsis genome by means of metabolite quantitative trait locus (QTL) (Keurentjes et al. 2006; Fu et al. 2007).

Chromatography linear trap quadrupole mass spectrometer (LC-LTQ-MS/MS^2) can finish 3 full spectrum scans per second or 3 double MS scans per second and provide quantitative and qualitative information in the mean time (Evans et al. 2009).

LTQ-Orbitrap-MS and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) have ultra-high resolution of up to 60,000–2,500,000 and the MS^n capacity of up to 10. From the above two techniques, the precise molecular weight of the compound can be obtained for the prediction of molecular formula, the qualitative analysis of compounds, and the establishment of compound MS database. However, in order to obtain ultra-high resolution, a longer scanning time is required.

The development of ultra-high performance liquid chromatography (UPLC or UHPLC) technology is the icing on the cake for the metabolomics analytic technology. The UPLC makes use of a chromatographic column with the particle size of packing <2.0 μm and has overcome the traditional HPLC pressure limit. Column pressure can be increased to 15,000 psi. Therefore, the column efficiency is enhanced, peak widths are narrower, chromatographic resolution is increased, and the analysis time is shorter. It is very ideal for UPLC to combine with Q-TOF/MS which has a high scanning speed for the high-throughput analysis in plant metabolomics research. Recently, there is a breakthrough for the chromatographic packings technology. The hydrophilic interaction liquid chromatography (HILIC) employs a kind of polar stationary phase (such as silica gel and amino-bonded silica gel) and water and polar organic solvent as the mobile phase. It is particularly suitable for the separation of strong polar and strong hydrophilic small molecules. It is a supplement for the reverse chromatography (Tolsticov and Fiehn 2002; Cubbon et al. 2010).

In short, LC–MS has simple sample extraction requirements, easy to implement high-throughput and automation, can detect most of the plant metabolites, and is bound to play a greater role in plant metabolomics research. There have been various combinations of LC with MS, and still, there will be newcomers in the future development. The ultra-high-performance LC coupled with high-resolution tandem quadrupole time-of-flight MS technology will be the mainstream platform of plant metabolomics analysis.

1.1.5 Analysis of Other Special-Purpose Technology

(1) **Capillary Electrophoresis Mass Spectrometry (CE-MS) Technique**. CE technology is a new separation technology that is developed in the early 1980s, which based on the difference of mobility and distribution behavior between components to be separated. It has such features as high speed, high efficiency, high resolution, good reproducibility, and easy to automation, etc. The main advantage of CE-MS is the ability to detect ionic compounds, such as phosphorylated sugars, nucleotides, organic acids, and amino acids. Researchers have detected 200 metabolites from Arabidopsis and have identified 70–100 compounds from the 200 metabolites using CE-MS (Ohkama-Ohtsu et al. 2008).

(2) **Nuclear Magnetic Resonance (NMR) Technique**. NMR technology is a non-biased, universal analytical technique, with simple pretreatment requirement and a variety of detecting method. NMR includes liquid high-resolution NMR, high-resolution magic angle spinning NMR (HR-MAS), and in vivo magnetic resonance spectroscopy (MRS) technology. NMR methods also have their limitations, for example, its detecting sensitivity is low and the dynamic detecting range is limited, which makes it difficult to detect components with great difference in content in the same sample at the same time (Zhu et al. 2006). Recently, combined with LC separation, solid phase extraction (SPE) enrichment, full deuterated solvent elution, and online LC-UV-SPE-NMR-MS have been used in the structure identification of plant metabolites (Exarchou et al. 2003; Lin et al. 2008).

(3) **Fourier Transform-InfraRed (FT-IR) Technique**. FT-IR is based on the mechanism: The infrared ray gives rise to the vibrations of chemical bonds in the molecule or rotational energy level transitions, which lead to the production of absorption spectrum. The infrared spectrum of a plant sample is the superposition of the infrared spectra of all compounds therein, therefore having the fingerprint characteristics. FT-IR is capable of screening metabolic mutants from a large population, as it can conduct fast and high-throughput scanning of more than 1,000 samples per day without destroying them (Allwood et al. 2008). The disadvantage is that it is difficult to identify and discriminate metabolites with similar structure types.

There is a big chemical diversity of plant metabolites. The content of some components is minimal and the dynamic range is wide. The biosynthesis and accumulation of metabolites are vulnerable to the external environment. Metabolomics cannot predict the structure of metabolites from genomic information just as proteomics and transcriptomics do. Currently, the panorama qualitative and quantitative analysis cannot be done by one single mean, but using a variety of analytical tools that can complement each other to monitor and track changes of plant metabolites as many as possible.

1.2 Current Development and Challenges of Metabolomics Data Analysis

1.2.1 The Experimental Design and Standardization in Metabolomics Research

Just like transcriptomics, plant metabolomics research faces the needs and challenges in experimental design and standardization. On the one hand, different metabolites are biosynthesized and accumulated in different developmental stages, different tissues, different organs, and different cell types, and the content of them is extremely vulnerable to the growth environment. On the other hand, with a variety of instruments and analysis conditions being applied in metabolite analysis, a large number of non-comparable data can be easily produced.

The strict experimental design is the first step to achieve the success of metabolomics experiments. The experimental design requires to: ① set substantially identical plant growth and environmental conditions. And if the completely identical conditions cannot be achieved in each experiment, do ensure the identical growth and environmental conditions of different treatment or materials within the same experiment; ② set the experimental replications, typically 4–6 replications, which will further eliminate the environmental and experimental operation error and obtain statistically significant data. In order to control and monitor errors coming from sample extraction, pre-treatment and instrument analysis generally require to: ① set blank control: blank control is only free of the sample to be analyzed. It can detect the purity of the organic solvent, the miscellaneous peak of derivatization reagent, the plasticizers and other foreign contaminants from plastic tubes and pipette tips; ② set quality control samples: the quality control sample is a mixture of different types of standard materials. It can also be the mixture of a small part of each test sample, which contains all the types of compounds to analyze test samples. The systemic drift and deviation, the reduced response to metabolites and other unknown changes that are easily caused by the large number of test samples in one metabolomics experiment and the too long instruments running time, column pollution, reduced column efficiency, and the pollution of inlet as well as the aging of the mass detector can affect the detection of metabolism profile. Therefore, the quality control sample plays a very important role in the detection and correction of metabolomics data; ③ set the internal standard: by adding a known amount of an internal standard substance in the plant during the extraction process, errors produced in the process of extraction and analysis can be detected; ④ add the standard substance for the retention time index: The said standard substance is generally n-alkane or homologues of saturated fatty acid methyl ester, which can be used to calculate the retention index (RI), qualitatively analyze metabolites, and correct the drift of retention time.

Metabolomics develops rapidly, and the number of papers in various areas of metabolomics increases rapidly. Metabolomics mostly studies nontargeted unknown components. The data analytic methods, the metabolites qualification

standard, and experimental report formats from different researchers may not be the same and hence hinder the exchange of data, the peer evaluation, and the reproduction of experimental results. Therefore, it is necessary to standardize metabolomics experiments and reports. John C. Lindon of the British Empire University of London took the lead and established the Metabolomics Standards Initiative (MSI) and proposed the standard metabolic reporting structure (SMRS) for metabolomics or metabonomics standardization (Lindon et al. 2005). Jenkins et al. proposed the architecture for metabolomics (ArMet) for plant metabolomics report (Jenkins et al. 2004). ArMet focus on the organization of metabolomics data, while SMRS puts forward more detailed description of data, including which parameter or data are necessary (Fiehn et al. 2007). Three papers that are published in the journal Metabolomics proposed, respectively, the MSI organization (Fiehn et al. 2007), the minimum reporting standards for chemical analysis (Sumner et al. 2007), and the minimum reporting standards for data analysis (Goodacre et al. 2007) in metabolomics. The minimum reporting standards for chemical analysis detailed the methods and technical parameters used in chemical analysis and proposed a lot of new standards and guiding principles (such as the structural determination level of metabolites and naming guidance of metabolites). The minimum reporting standards for data analysis describe in detail methods of univariate statistics, multivariate statistics, and informatics and defined many terms, such as deconvolution and pretreatment.

1.2.2 Metabolic Pathways and Metabolite Database

The establishment of metabolites database is conductive to the connection between metabolomics and other systems biology branches. At present, there are over 100 online databases (Tohge and Fernie 2009), in which the one with good comprehensiveness and containing metabolic pathways database is the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/). KEGG allows query of metabolic pathways, which include carbohydrates pathway, nucleotides pathway, amino acids pathway, and secondary metabolites pathway. KEGG provides not only all possible metabolic pathways, but also comprehensive annotations of enzymes involved in each step of the catalytic reaction, amino acid sequences, and the link to PDB database. KEGG is a comprehensive bioinformatics database for genome annotation.

MetaCyc (http://metacyc.org/) is a sub-database of BioCyc, which is a database of metabolic pathways and enzymes. It expounds metabolic pathways of more than 1,600 organism species and contains metabolic pathways, reactions, enzymes, and substrates that obtained from a large number of documents and online resources, including more than 1,200 pathways, 5,500 enzymes, 5,100 genes, and 7,700 metabolites. PlantCyc (http://www.plantcyc.org/) database and query system for plant were also established, including 12 species of Arabidopsis, poplar, rice, sorghum, and other species.

In addition to the large commercial databases on metabolites identification, such as NIST and WILEY, databases specifically for metabolomics that are established by experimental groups or research centers also come into being.

The Golm Metabolome Database (GMD) (http://gmd.mpimp-golm.mpg.de/) contains GC-MS and GC-TOF-MS mass spectrogram database of metabolites after derivatization. The current GMD database contains more than 2,000 evaluated mass spectrograms provided by both quadrupole and TOF mass spectrometry technical platform, including 1,089 non-redundant mass spectral tags (MSTs) and 360 identified MSTs. Additionally, GMD database also includes mass spectra, retention time and retention index, greatly improving the identification of structurally similar compounds. GMD can be queried free.

The METLIN metabolite database (http://137.131.20.83/metabo_search_alt2.php) (SCRIPPS) that is established by the Biological Mass Spectrometry Center of SCRIPPS Institute includes more than 23,000 endogenous and exogenous metabolites, small molecule drugs and drug metabolites, small peptides (about 8,000 dipeptides and tripeptides) and the like. METLIN contains corresponding LC-MS, MS/MS, and FTMS mass spectrometry data of each compound and can be freely retrieved by the mass, chemical formula, and structure of a compound online.

Madison-Qingdao Metabolomics Consortium Database (MMCD), http://mmcd.nmrfam.wisc.edu/) database is developed and maintained by the National Magnetic Resonance Facility at Madison of Wisconsin-Madison University and Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences. It contains data of over 20,000 small molecule metabolites that are gathered from electronic databases and the scientific literatures and provides search engines of texts, chemical structures, NMR data, MS data, etc.

KNApSAcK (http://kanaya.naist.jp/KNApSAcK/) covers the majority of connections between plant species and metabolites, which including information on more than 40,000 metabolites and 8,000 plant species and can be easily queried by users to obtain reported information of a certain plant species.

MassBank (http://www.massbank.jp/) mass spectrogram database is jointly established by many universities and research institutions in Japan. It mainly collects high-resolution mass spectrogram produced by various mass spectrometer, such as the ESI-Q-TOF-MS/MS, ESI-QqQ-MS/MS, ESI-IT-$(MS)^n$ (Ion Trap, IT), GC-EI-TOF-MS, and LC-ESI-TOF-MS. The reference spectrum comprises the information of multistage MS. So far, 24,993 mass spectrograms of more than 12,000 primary metabolites and secondary metabolites have been obtained in positive and negative ion modes. Users can input text format of mass spectrogram to search and compare to the three-dimensional mass spectrogram.

But, so far, the number of metabolites in the plant metabolites database is not enough, and most of the metabolic pathways of most of plant species are not complete, which is the main challenges and opportunities faced by the plant metabolomics research.

1.2.3 The Structural Identification of Metabolite Is the Bottleneck of Metabolomics Development

The determination of metabolites is the focus and difficult point of metabolomics research. The minimum reporting standards for chemical analysis bases on the degree of determination divided metabolites into four categories, i.e., identified compounds, putatively annotated compounds, putatively characterized compound classes, and unknown compounds (Sumner et al. 2007).

The main approach of metabolite identification using GC–MS is by software deconvolution to obtain pure mass peak and search in the existing database. The more higher the similarity of an interested component to the compound in the database, the more possible it can be regard as the same compound. In recent years, the RI is used to assist determination, which to some extent can differentiate isomers with similar mass spectragram. Aligning with a standard is the most accurate determination method. For LC-MS, there is no standard spectrogram database, and thus, the determination is mainly done by accurate mass and MS^n spectral tree. Therefore, the application of high-resolution mass spectrometry instruments to determine the molecular weight of the compound is essential. In addition, the online HPLC-DAD-SPE-NMR-MS can be combined with high-resolution mass spectrometry and NMR technology to accelerate the determination and analysis of metabolite structure.

The first challenge of metabolomics bioinformatics is the comprehensive annotation of metabolites and the dissection of metabolic pathways. At present, the rebuilding of metabolic pathways of a plant species can be done by ways of hunting-related functional genes in genomic information and EST data as the KEGG and PlantCyc do. On the other hand, the previously reported metabolic pathways-related literature data can be used not only as the reference for the identification of metabolites, but also to infer metabolic pathways; The second challenge is the collection of structure identification data of natural product, including multistage mass spectrometry data; the third challenge is to develop an algorithm for the similarity comparison and search of multistage mass spectrometry; the fourth challenge is the de novo resolution of mass spectrogram, which start the automated structure identification from mass spectrogram and predict the mass spectrogram of compounds with known structure (Saito and Matsuda 2010).

1.2.4 The Metabolomics Data Annotation Facing the Challenge

The main research content of the metabolomics data annotation and the main goal of metabolomics research are to obtain a large number of qualitative and quantitative data of metabolites, classify these disorder data according to plant metabolic

pathways or metabolic network and the functions involved, and then to get data with biological significance.

Plants synthesize a large number of compounds through metabolic pathways. Metabolites in the same pathway interrelated to each other to complete one or more biological functions. At present, only a small part of metabolic pathways have been dissected, which have been collected by KEGG and PlantCyc. On the one hand, the metabolic basis of certain biological functions can be systematically revealed by the annotation of differential metabolites that are collected from metabolomic analysis into metabolic pathways; on the other hand, some metabolic pathways can be predicted according to the association between metabolomics data and biological function, so as to accelerate the dissection of plant metabolic pathways. For example, after the type and content changes of tomato fruit metabolites were analyzed using GC–MS and annotated into metabolic pathways, the relationship of metabolic pathways and main agronomic traits was established (Schauer et al. 2006). Plants initiate certain metabolic pathways and produce certain disease-resistant metabolites immediately after the pathogen invasion. The concern to above-said metabolic pathways is beneficial for the dissection of disease resistance mechanism in terms of metabolites. For example, metabolomics studies have shown that the expression of salicylic acid and the regulation of metabolic pathways are closely associated with plant disease resistance (Kachroo and Kachroo 2009; Thomma et al. 1998). Fatma Kaplan et al. used GC–MS to study the changes of metabolic profile caused by low- and high-temperature stimuli in Arabidopsis and found some temperature stress-related metabolites (Kaplan et al. 2004). And then, the above-said data were co-analyzed with transcriptomics data to reveal the mechanism of temperature stress (Kaplan et al. 2007). The metabolomics data annotation will help resolving the function of new genes. For example, Jander et al. 2004 identified a threonine aldolase gene using LC–MS to high-throughput screen Arabidopsis mutation population. In addition, the annotation of metabolic pathways can be used for the determination of the activity of the enzyme involving in the metabolic pathway for the metabolic flux analysis.

The biological metabolism is a complete system, in which various metabolic pathways are interrelated and constitute a complex metabolic network. Combined with genomics, transcriptomics and metabolomics information, the annotation or reconstruction of metabolic network can be done. Patrick May et al. reconstructed the metabolic network of *Chlamydomonas reinhardtii* by the analysis of 1,069 proteins using LC–MS and 159 metabolites using GC × GC-TOF/MS and the integrating of genomic information (May et al. 2008). The correlation between metabolites can be established through correlation analysis of metabolic pathways. For example, some associated RNAs, metabolites, and closely correlated metabolic pathways have been identified in the ripening process of tomato by the study and statistical analysis of the content changes of primary metabolites and secondary metabolites and the combination with transcriptomics data (Ciarrari et al. 2006).

1.3 Plant Metabolomics Applications

Since the term metabolomics was proposed (Oliver et al. 1998), metabolomics has developed rapidly in aspects of analytical technique, data analysis, and application, got widely attention in the field of human disease study and disease diagnosis, and has broad application prospects (Nicholson and Lindon 2008). Plant metabolomics has been gradually applied to the basic biology study, such as the study of gene function, the dissection of metabolic pathway, and the regulation mechanism of metabolic network (Fiehn et al. 2000; Naoumkina et al. 2010; Keurentjes et al. 2006; Fernie et al. 2004); plant metabolomics also started to be applied to the field of plant breeding, such as crop yield and nutritional components (Schauer et al. 2006; Tarpley et al. 2005).

1.3.1 Plant Metabolomics Applications in Basic Biology

As an essential part of systems biology, plant metabolomics plays an increasingly important role in uncovering the basic vital movement and mechanism of life. Information on gene expression patterns and metabolites accumulation can be obtained by the analysis of expression profiles and metabolome. The comprehensive analysis of gene expression patterns, metabolites accumulation, and genomic data provides an effective way for the analysis of gene function. It is generally believed that certain genes, proteins, or metabolites involving in a biological process coexist in the same control system, with the relationship of coordinated regulation and co-expression. Therefore, if an unknown gene co-express with known genes, it can be assumed that the unknown gene may involve in the same biological process with the co-expressed known genes. The co-occur principle can be further applied in the co-accumulation relationship. If a metabolic pathway was modified by gene mutations or environmental changes, the modification process can be shown through changes of the metabolic profile. The analysis of gene expression profiles and metabolic profile can predict which gene(s) may be involved in this modification process. Candidate genes can be obtained by correlation analysis between metabolites and genes, and the function of the candidate genes can be studied by reverse genetics or reverse biochemical methods (Saito and Matsuda 2010). In the study of sulfur hunger, the change of metabolites and genes related to certain amino acids, lipids, and secondary metabolites such as glucosinolates and flavonoids was found, and 10,000 transcriptomics data (DNA chip) and 1,000 metabolites data (HPLC and CE-FT-MS) were analyzed by batch learning and self-organizing maps, which predicted genes involved in the biosynthesis of glucosinolates, such as genes encoding the sulfotransferase (Hirai et al. 2005), two MYB transcription regulating factor (Hirai et al. 2007), side chain extension-related enzymes (Sawada et al. 2009a), and a putative glucosinolates transporter (Sawada et al. 2009b).

Metabolomics combined with molecular genetics can resolve Arabidopsis metabolic pathways and metabolic network (Keurentjes et al. 2006). Keurentjes et al. using LC-Q-TOF-MS analyzed the full metabolite profile of 14 Arabidopsis accessions derived from different locations around the world and identified 2,475 different mass spectrum peaks, 706 of which are unique for certain accession, only 331 of which are common for the 14 accessions. So the variation of metabolite composition among the 14 accessions is very large. The broad-sense heritability analysis showed that most of the mass peak variations derived from genetic factors. In the 160 recombinant inbred lines (F_{10}) of Landsberg *erecta* (L*er)* and Cape Verde Islands (Cvi), a total of 2,129 mass peaks were detected, 853 of which are recombinant inbred population-specific, indicating that they are generated from genetic recombination. Genetic mapping of quantitative trait loci analysis found that the 1,592 mass peaks (74.8 %) have at least one QTL control ($P < 0.0001$, $q < 0.0002$). The mapped genetic factors at least partially explained the quantity and quality variation of metabolites. The analysis of the distribution of these detected QTL in genome found that they tend to locate in some specific regions of the genome, i.e., the hot spot region.

If a large number of highly relevant mass peaks were located to the same QTL of Arabidopsis, it can be inferred that these mass peaks are regulated by the same key genes. Co-regulated metabolites may be subject to a special regulatory factor control, or a particular step in metabolic pathways may be affected. Kroymann carefully analyzed these intermediates involved in the glucosinolates metabolic pathway of *A. thaliana* and found that all the aliphatic glucosides locate in the *MAM* locus of chromosome 5 and *AOP* locus of chromosome 4. It is known that the *MAM* controls chain extension (Kroymann et al. 2001), and *AOP* controls the modification of the side chain (Kliebenstein et al. 2001). The glycoside metabolic network structure rebuilt by metabolomics analysis is consistent with reported results. Therefore, it is confirmed that the colocation of Arabidopsis glucosinolates is due to a particular step of the metabolic pathway. In addition, the metabolomics analysis can infer the relationship between *AOP* and *MAM*. As shown in Fig. 1.2, metabolites at the *AOP* locus are also mapped at *MAM* locus, the reverse, however, does not apply, and thus, it can be estimated that *AOP* locates at the downstream of *MAM*.

Genetic analysis of metabolites can help resolving unknown metabolic pathways. A series of highly relevant unknown metabolite mass peak (QTL) have been located at 88.6 cM of *A. thaliana* chromosome 1. All of these co-localized metabolites are identified as flavonol glycosides with the analysis of photodiode array (PDA) absorption signal and MS/MS mass spectrum fragments. The additive allelic effect results indicate that the genotype variation of this locus displays an opposite flavonol accumulation mode, i.e., the L*er* allelic genotype accumulates flavonols, while Cvi not. These results indicate that a previously unknown glycosyltransferase exists in L*er*, while not in Cvi. Based on the homologous sequence analysis, two putative glycosyltransferase genes, UGT79B10 and UGT79B11, were found in the QTL locus, providing candidate genes for the functional verification of genes.

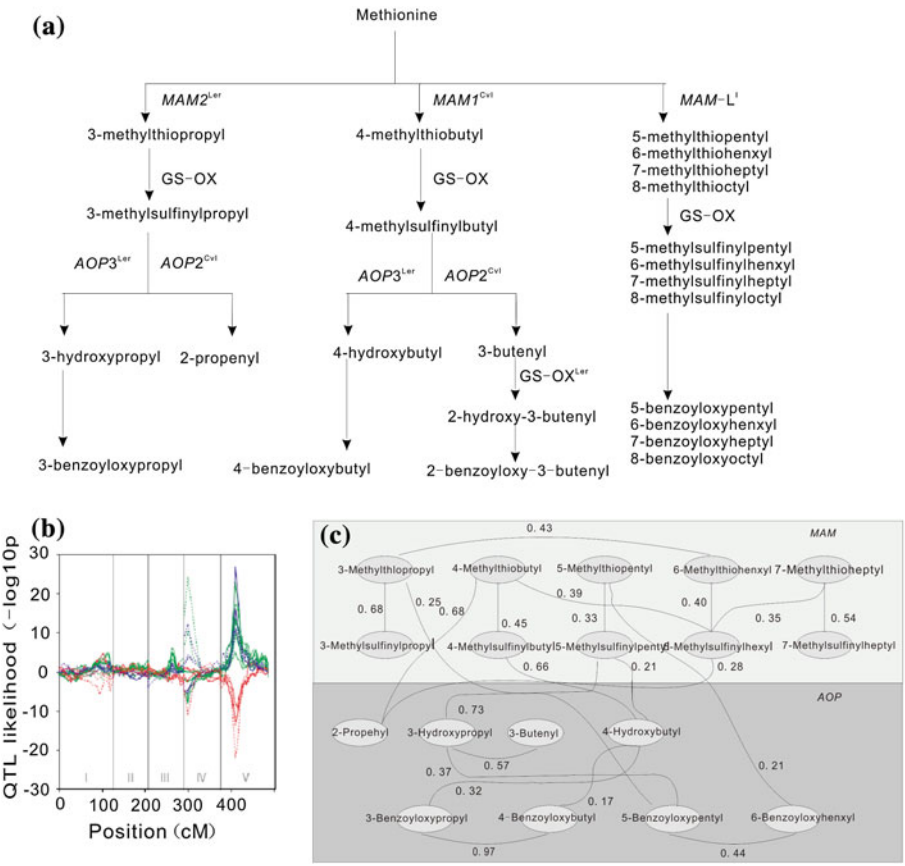

Fig. 1.2 Genetic regulation of aliphatic Glucosinolate accumulation in *A. thaliana*. **a** Scheme of aliphatic Glucosinolate formation. Corresponding loci of enzymatic steps are shown in *bold next to the arrows*. **b** QTL likelihood profiles of aliphatic Glucosinolates detected in the RIL population. The first QTL, at 303.3 cM, is at the *AOP* locus, and the second, at 409.4 cM, is at the *MAM* locus. The sign of the value is related to the additive effect at each marker (+, Cvi; –, Ler). *Solid lines* represent Glucosinolates before side chain modification and *dotted lines* Glucosinolates after side chain modification. Chromosomal borders are indicated by *vertical gray lines*. **c** Second-order genetic correlations between aliphatic Glucosinolates detected in the RIL population. *Upper panel* contains Glucosinolates before side chain modification; *lower panel* contains Glucosinolates after side chain modification. All edges depicted are significant at $\alpha = 0.05$, as determined by permutation. Corresponding correlation values are placed next to edges. In (**b**) and (**c**), colors represent different chain lengths (*red*, 3 C; *blue*, 4 C; *green*, 44 C) (Reprinted by permission from Macmillan Publishers Ltd: [Nature Genetics] (Keurentjes et al. 2006))

1.3.2 Applications of Metabolomics in Crop Breeding and Biotechnology

The safety of transgenic breeding has been the key issue hindering its wide use. Combined with allergic reactions and toxicological reaction experiments, metabolomics can be used for the comprehensive evaluation of the safety of genetically modified (GM) plants (Rischer and Oksman-Caldentey 2006). The metabolic profiles of GM, non-GM, and two cultivated varieties of potato were analyzed by principal component analysis (PCA) and hierarchical clustering analysis (HCA) and show that the most obvious differences in metabolic profiles are between the two potato varieties but not between the wild type and the GM plant (Shepherd et al. 2006).

Metabolomics can be used to study the effect of temperature, moisture, salt, sulfur, phosphorus, heavy metals, and other stress on plant metabolism. Drought resisting has been one of the most concerned issues in plant breeding. When drought occurs, the signal substances secreted by root will be transported to the aboveground part of plant through the xylem which is the channel for transporting water, minerals, and other components in plant. Alvarez et al. using metabolomics studied the changes of metabolites and proteins in maize xylem sap under drought stress (Alvarez et al. 2008) and found content changes in 31 metabolites. They found changes in most of the 31 metabolites such as the phytohormones abscisic acid (ABA) and cytokinin and the presence of high concentrations of the aromatic cytokinin 6-benzylaminopurine (BAP) at the 12 days of drought stress. Several phenylpropanoid compounds (coumaric, caffeic, and ferulic acids) were found in xylem sap. The concentrations of some of these phenylpropanoid compounds changed under drought. In parallel, an analysis of the xylem sap proteome was conducted. They also found a higher abundance of cationic peroxidases, which with the increase in phenylpropanoids may lead to a reduction in lignin biosynthesis in the xylem vessels and could affect the length of leaves and stems.

Plant initiates autoimmune response after the invasion of pathogen, in which metabolite plays a very important role. Once plant recognizes the pathogen, the cell will initiate a series of mobilization activities to activate disease resistance response to resist the invasion of pathogens. This disease resistance response requires a lot of energy, reducing power and carbon skeleton from the primary metabolic pathways to achieve the energy supplement and scheduling (Bolton 2009), and some low molecular weight antibacterial secondary metabolites from the secondary metabolic pathway such as phytoalexins (phenols, isoflavones, terpenes), and some substances that can block pathogen invasion and spread, such as lignin and callose (O'Connell and Panstruga 2006). On the other hand, after invading to the plant, pathogen will usually interfere with the normal metabolism of plant in order to meet their nutritional needs (Solomon et al. 2003; Swarbrick et al. 2006; Divon and Fluhr 2007). By the qualitative or quantitative analysis of metabolites in organism under certain conditions, metabolomics can detect changes in metabolites associated with a particular physiological and pathological reaction. In recent years, such method starts to be applied in the study of plant–microbe interaction. By the nontarget

metabolite profile analysis of different resistant varieties of wheat and barley, some disease resistance-related mark metabolites have been identified, which play a role as assisting breeding in production practice (Hamzehzarghani et al. 2005; Swarbrick et al. 2006).

The plant–pathogen interaction systems, such as Brachypodium (*Brachypodium distachyon*)–rice blast fungus (*Magnaporthe grisea*) (Allwood et al. 2006; Parker et al. 2009), Arabidopsis (*A. thaliana*)–Arabidopsis nematode (*Heterodera schachtii*) (Hofmann et al. 2010), potato (*Solanum tuberosum* L)–potato late blight fungus (*Phytophthora infestans*) (Abu-Nada et al. 2007), have been studied by metabolomics. The study found that once the pathogen successfully invades into plant, it will seriously interfere with the normal metabolism of plants to meet their own needs of nutrition absorption and usage. Changes of metabolic profiles between resistant and susceptible cultivar after the invasion of pathogen are different. Combined with metabolomics and transcriptomics, Doehlemann et al. expounded that changes in signal transduction and metabolites in corn disease resistance response to tumor smut (*Ustilago maydis* SG200) are controlled by a major gene (Doehlemann et al. 2008). These studies suggest that metabolomics is a very effective method in the study of plant–microbe interaction.

The main traits of the crop, especially nutrition, quality, and other traits, have become the main target of metabolomics studies. Metabolomics is able to distinguish which chemical component determines the taste, such as sweet and sour. Accordingly, it has a good application prospect for breeding in improving nutrition, quality, and food quality. Using ^{1}H-NMR spectrum, combined with multi-dimensional statistical analysis can distinguish wines of different varieties and different regions (Son et al. 2009); by the application of GC–MS to analyze the metabolic profile of green tea, the quality of green tea is able to be evaluated (Pongsuwan et al. 2008); and the use of FT-IR to distinguish between different regions of olive oil ingredients (Rischer and Oksman-Caldentey 2006; Galtier et al. 2007).

The improvement of the nutritional quality and taste of the tomato can be achieved by changing the composition and content of the metabolites. The following section will be a more detailed introduction to literatures of metabolomics research in the molecular mechanisms of plant phenotype and metabolite content of tomato fruit. The study used 76 introgression lines containing chromosome segments of a wild tomato species (*Solanum pennellii*) in the genetic background of a cultivar M82 (*Solanum lycopersicum*) as the material and used high-throughput GC-MS to obtain metabolomics data of tomato fruit, combined with tomato whole plant phenotype data, and to explore the molecular genetic basis of metabolite composition formation of tomato fruit (Schauer et al. 2006). With the use of variance analysis, a total of 889 quantitative fruit metabolic loci and 326 loci that modify yield-associated traits were identified.

To further analyze the correlation among fruit metabolites, change, plant growth and development, and the important agronomic traits, correlation analysis of metabolites and 83 traits such as total fruit yield, harvest index and Brix was done, and a total of 280 positive correlations and 22 negative correlations were obtained with the significant level $P = 0.0001$ (Fig. 1.3). Module analysis and correlation

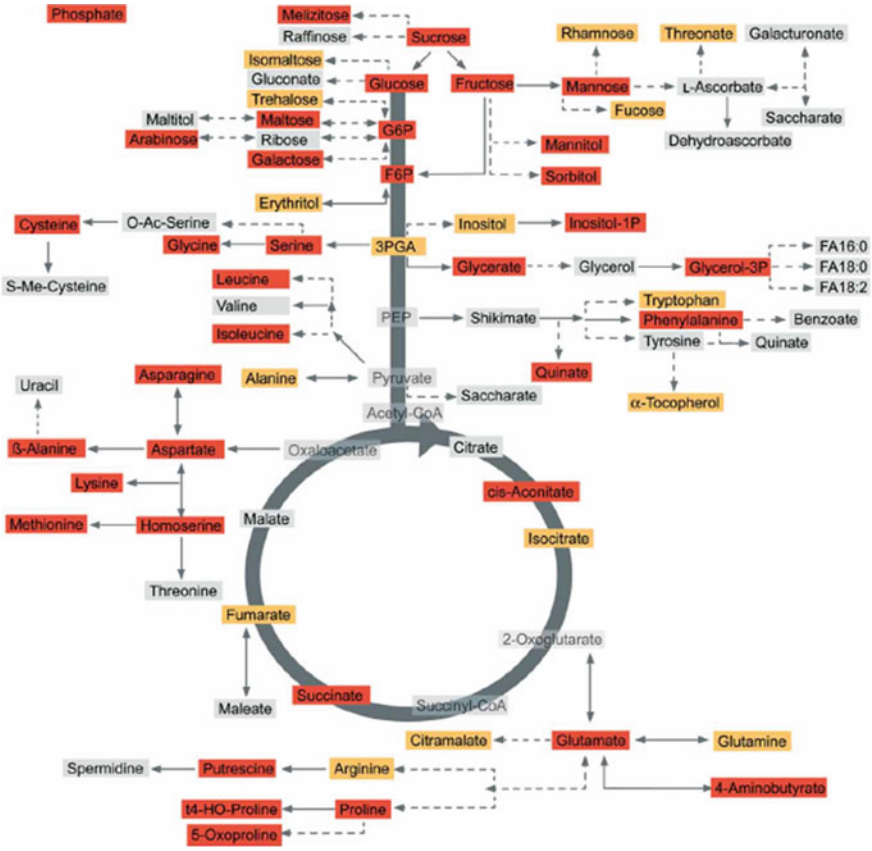

Fig. 1.3 The diagram of phenotype-associated and phenotype-independent metabolites in tomato fruit. Metabolites marked in *red* and *orange* were determined to be highly phenotype associated ($p \geq 0.005$) and phenotype associated ($p \geq 0.05$), respectively; those marked in *gray* were determined to be phenotype-independent because they did not correlate to any of the true phenotypic traits at either the strict or the permissive level. Traits colored *pale gray* were not analyzed in this study (Reprinted by permission from Macmillan Publishers Ltd: [Nature Biotechnology] (Schauer et al. 2006))

analysis show that many of the metabolic loci are associated with at least one QTL that modify whole plant phenotype traits. According to the degree of correlation between metabolic and whole plant phenotypic traits, the association was classified into three categories: whole plant phenotype associated, whole plant phenotype independent, and the intermediate. The proportion of the three types is 50, 27, and 23 %, respectively. In addition, correlation analysis of metabolic QTL and whole plant phenotypic QTL showed that 46 % of the 889 metabolic QTL are shared with whole plant phenotypic QTL.

Data from many introgression lines containing bin 6F (chromosome segment number) indicate that bin 6F contains 7 metabolic QTL associated with whole plant

phenotype traits and 1 QTL associated with decreased harvest index and increased Brix. Previous studies have shown that bin 6F contains the *SELF-PRUNING* (*SP*) gene which regulates vegetative to reproductive switching of plant (Pnueli et al. 1998). Therefore, it is not difficult to understand why the harvest index-associated QTL locates in bin 6F. In order to further assess the causality among *SP* gene, harvest index and Brix, the phenotype and metabolite content of three near-isogenic lines (*sp* recessive mutants in Gardener and VFNT varieties and M82) and wild-type control were directly compared. The data show that the harvest index of the *SP* line was much lower than that of the *sp* recessive mutant, whereas the Brix value was always higher, indicating that there is a strong negative association between these traits. For Brix determination, the content of soluble solids, i.e., the soluble metabolites, was detected. From above experiments, it can be considered that the changes of plant phenotype can also affect the changes of fruit metabolites. Results from the F_2 population of M82 that segregated for a recessive male sterile mutation also proved that plant morphology contributes strongly to determining the metabolic composition of fruits at harvest time. The above findings provide target traits for efficient selection and laid the foundation for efficient breeding.

In short, plant metabolomics is developing rapidly and is playing an increasingly important role in revealing the molecular mechanisms of plant growth and development and adversity adaptation. At present, there are a lot of technical bottleneck and challenges in terms of metabolite analysis technique, compound structure identification, and data analysis, providing research objectives for those engaged in plant metabolomics research. This chapter and other chapters only enumerate several application examples of plant metabolomics; however, with the improvement of plant metabolomics techniques and methods, it is bound to be applied to every field of botany by the majority of plant science researchers.

References

Abu-Nada Y, Kushalappa EA, Marshall WD, Al-Mughrabi K, Murphy A. Murphy temporal dynamics of pathogenesis-related metabolites and their plausible pathways of induction in potato leaves following inoculation with *Phytophthora infestans*. Eur J Plant Pathol. 2007;118:375–91.

Allwood JW, Ellis DI, Heald JK, Goodacre R, Mur LAJ. Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by *Brachypodium distachyon* to challenge by *Magnaporthe grisea*. Plant J. 2006;46:351–68.

Allwood JW, Ellis DI, Goodacre R. Metabolomic technologies and their application to the study of plants and plant–host interactions. Physiol Plant. 2008;132:117–35.

Alvarez S, Marsh EL, Schroeder SG, Schachtman DP. Metabolomic and proteomic changes in the xylem sap of maize under drought. Plant Cell Environ. 2008;31:325–40.

Bolton MD. Primary metabolism and plant defense—fuel for the fire. Mol Plant Microbe Interact. 2009;22:487–97.

Carrari F, Baxter C, Usadel B, Urbanczyk-Wochniak E, Zanor M, Nunes-Nesi A, Nikiforova V, Centero D, Ratzka A, Pauly M, Sweetlove L, Fernie A. Integrated analysis of metabolite and

transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Plant Physiol. 2006;142:1380–96.
Cubbon S, Antonio C, Wilson J, Thomas-Oates J. Metabolomic applications of HILIC–LC–MS. Mass Spectrom Rev. 2010;29:671–84.
Divon HH, Fluhr R. Nutrition acquisition strategies during fungal infection of plants. FEMS Microbiol Lett. 2007;266:65–74.
Dixon RA, Strack D. Phytochemistry meets genome analysis, and beyond. Phytochemistry. 2003;62:815–6.
Doehlemann G, Wahl R, Horst RJ, Voll L, Usadel B, Poree F, Stitt M, Pons-Kuehnemann J, Sonnewald U, Kahmann R, Kämper J. Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph Ustilago maydis. Plant J. 2008;56:181–95.
Ekman R, Silberring J, Westman-Brinkmalm A, Kraj A. Mass spectrometry: instrumentation interpretation and applications. New Jersey: Wiley; 2009.
Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem. 2009;81:6656–67.
Exarchou V, Godejohann M, van Beek TA, Gerothanassis IP, Vervoort J. LC-UV-solid-phase extraction-NMR-MS combined with a cryogenic flow probe and its application to the identification of compounds present in greek oregano. Anal Chem. 2003;75:6288–94.
Fernie A, Trethewey R, Krotzky A, Willmitzer L. Innovation—Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol. 2004;5:763–9.
Fiehn O. Metabolomics—the link between genotypes and phenotypes. Plant Mol Biol. 2002;48:155–71.
Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. Metabolite profiling for plant functional genomics. Nat Biotech. 2000;18:1157–61.
Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, Sumner L, Goodacre R, Hardy N, Taylor C, Fostel J, Kristal B, Kaddurah-Daouk R, Mendes P, van Ommen B, Lindon J, Sansone S-A. The metabolomics standards initiative (MSI). Metabolomics. 2007;3:175–8.
Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational protocol for the genetic study of metabolic networks. Nat Protoc. 2007;2:685–94.
Galtier O, Dupuya N, Dr´eau YL, Ollivier D, Pinatel C, Kister J, Artaud J. Geographic origins and compositions of virgin olive oils determinated by chemometric analysis of NIR spectra. Anal. Chim. Acta. 2007;595:136–44.
Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, Bessant C, Connor S, Calmani G, Craig A, Ebbels T, Kell DB, Manetti C, Newton J, Paternostro G, Somorjai R, Sjostrom M, Trygg J, Wulfert F. Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics. 2007;3:231–41.
Hamzehzarghani H, Kushalappa AC, Dion Y, Rioux S, Comeau A, Yaylayan V, Marshall WD, Mather DE. Metabolic profiling and factor analysis to discriminate quantitative resistance in wheat cultivars against fusarium head blight. Physiol Mol Plant Pathol. 2005;66:119–1113.
Hegeman AD. Plant metabolomics—meeting the analytical challenges of comprehensive metabolite analysis. Brief Funct Genomics. 2010;9:139–48.
Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K. Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem. 2005;280:25590–5.
Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci USA. 2007;104:6478–83.
Hofmann JL, El Ashry AEN, Anwar S, Erban A, Kopka J, Grundler F. Metabolic profiling reveals local and systemic responses of host plants to nematode parasitism. Plant J. 2010;62:1058–71.

Jander G, Norris SR, Joshi V, Fraga M, Rugg A, Yu S, Li L, Last RL. Application of a high-throughput HPLC-MS/MS assay to Arabidopsis mutant screening; evidence that threonine aldolase plays a role in seed nutritional quality. Plant J. 2004;39:465–75.

Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino RJ, Hall R, Kopka J, Lane GA, Lange BM, Liu JR, Mendes P, Nikolau BJ, Oliver SG, Paton NW, Rhee S, Roessner-Tunali U, Saito K, Smedsgaard J, Sumner LW, Wang T, Walsh S, Wurtele ES, Kell DB. A proposed framework for the description of plant metabolomics experiments and their results. Nat Biotech. 2004;22:1601–6.

Kachroo A, Kachroo P. Fatty acid-derived signals in plant defense. Ann Rev Phyto. 2009;47:153–76.

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

Kaplan F, Kopka J, Sung DY, Zhao W, Popp M, Porat R, Guy CL. Transcript and metabolite profiling during cold acclimation of Arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. Plant J. 2007;50:967–81.

Keurentjes JJB, Fu J, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M. The genetics of plant metabolism. Nat Genet. 2006;38:842–9.

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell Olds T. Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate–dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell. 2001;13:681–93.

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, Dormann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D. GMD@CSB.DB: the golm metabolome database. Bioinformatics. 2005;21:1635–8.

Kroymann J. A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol. 2001;127:1077–88.

Lin Y, Schiavo S, Orjala J, Vouros P, Kautz R. Microscale LC-MS-NMR platform applied to the identification of active cyanobacterial metabolites. Anal Chem. 2008;80:8045–54.

Lindon JC, Nicholson JK, Holmes E, Keun HC, Craig A, Pearce JTM, Bruce SJ, Hardy N, Sansone S-A, Antti H, Jonsson P, Daykin C, Navarange M, Beger RD, Verheij ER, Amberg A, Baunsgaard D, Cantor GH, Lehman-McKeeman L, Earll M, Wold v, Johansson E, Haselden JN, Kramer K, Thomas C, Lindberg J, Schuppe-Koistinen I, Wilson ID, Reily MD, Robertson DG, Senn H, Krotzky A, Kochhar S, Powell J, Ouderaa FVD, Plumb R, Schaefer H, Spraul M. Summary recommendations for standardization and reporting of metabolic analyses. Nat Biotech. (2005);23:833–8.

May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupprecht J, Weiss J, Recuenco-Munoz L, Ebenhoh O, Weckwerth W, Walther D. Metabolomics- and Proteomics-assisted genome annotation and analysis of the draft metabolic network of chlamydomonas reinhardtii. Genetics. 2008;179:157–66.

Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, de Vos CHR. A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. 2006;141:1205–18.

Naoumkina MA, Modolo LV, Huhman DV, Urbanczyk-Wochniak E, Tang Y, Sumner LW, Dixon RA. Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *M. truncatula*. Plant Cell. 2010;22:850–66.

Nicholson JK, Lindon JC. Metabonomics Nature. 2008;455:1053–6.

Nikolau B, Wurtele E. Concepts in plant metabolomics. Dordrecht, The Netherlands: Springer; 2007. p. 11–5.

Nour-Eldin HH, Halkier BA. Piecing together the transport pathway of aliphatic glucosinolates. Phytochem Rev. 2009;8:53–7.

O'Connell RJ, Panstruga R. Tête à tête inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes. New Phytol. 2006;171:699–718.

Ohkama-Ohtsu N, Oikawa A, Zhao P, Xiang C, Saito K, Oliver DJ. A {gamma}-glutamyl transpeptidase-independent pathway of glutathione catabolism to glutamate via 5-oxoproline in Arabidopsis. Plant Physiol. 2008;148:1603–13.