Xiaoquan Qi · Xiaoya Chen
Yulan Wang *Editors*

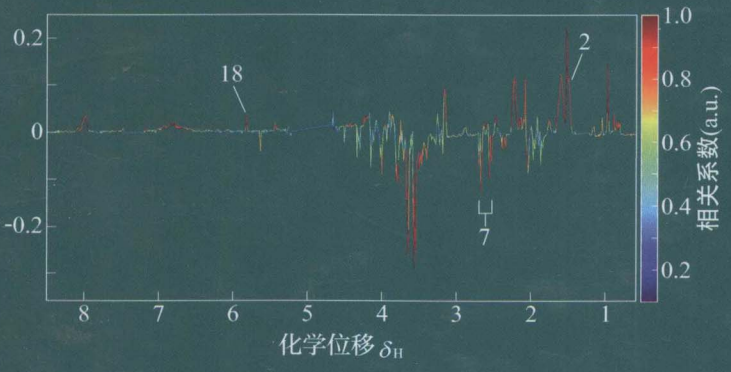
漆小泉　陈晓亚　王玉兰　主编

Plant Metabolomics

Methods and Applications

植物代谢组学

——方法与应用



化学工业出版社

Springer

[redacted]

Plant Metabolomics

Methods and Applications

植物代谢组学

——方法与应用

[redacted]

Plant Metabolomics
Methods and Applications

植物代谢组学
——方法与应用

Xiaoquan Qi · Xiaoya Chen
Yulan Wang *Editors*

漆晓泉　陈晓亚　王玉兰　主编

化学工业出版社
·北京·

Springer

代谢组学是近年来发展起来的“组学”之一，它是系统生物学中非常重要的一个环节，是基因与表型之间的桥梁，代谢组学不仅可以揭示基因的功能，也为生物技术的应用提供科学依据。本书由活跃在国内植物代谢组学领域的一线学者撰写，按照代谢物分析方法、数据分析及应用三部分深入浅出地介绍了植物代谢组学的基本概念、最新研究方法和发展趋势及其应用中的注意事项，该书还以实际的科研项目为实例，展现了我国植物代谢组学研究现状。该书内容贴切实际、语言深入浅出，使初学者能够较快掌握植物代谢组学的基本知识和技能。

本书适合高等院校有关专业的本科生、研究生，也可供科学研究工作者、专业技术人员参考学习。

图书在版编目（CIP）数据

植物代谢组学：方法与应用=Plant Metabolomics: Methods and Applications : 英文 / 漆小泉，陈晓亚，王玉兰主编. — 北京：化学工业出版社，2015.3
ISBN 978-7-122-21436-2

Ⅰ. ①植… Ⅱ. ①漆… ②陈… ③王… Ⅲ. ①植物-代谢-研究-英文 Ⅳ. ①Q946

中国版本图书馆CIP数据核字（2014）第 168143 号

本书由化学工业出版社与德国Springer出版公司合作出版。本版本仅限在中华人民共和国境内（不包括中国台湾地区和中国香港、澳门特别行政区）销售。

责任编辑：吴 刚 李 丽　　装帧设计：韩 飞

出版发行：化学工业出版社（北京市东城区青年湖南街13号 邮政编码100011）
印　　装：北京京华虎彩印刷有限公司
710mm×1000mm 1/16　印张 20½　字数 656 千字　2015年3月北京第1版第1次印刷

购书咨询：010-64518888（传真：010-64519686）　售后服务：010-64518899
网　　址：http://www.cip.com.cn
凡购买本书，如有缺损质量问题，本社销售中心负责调换。

定　　价：280.00元

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

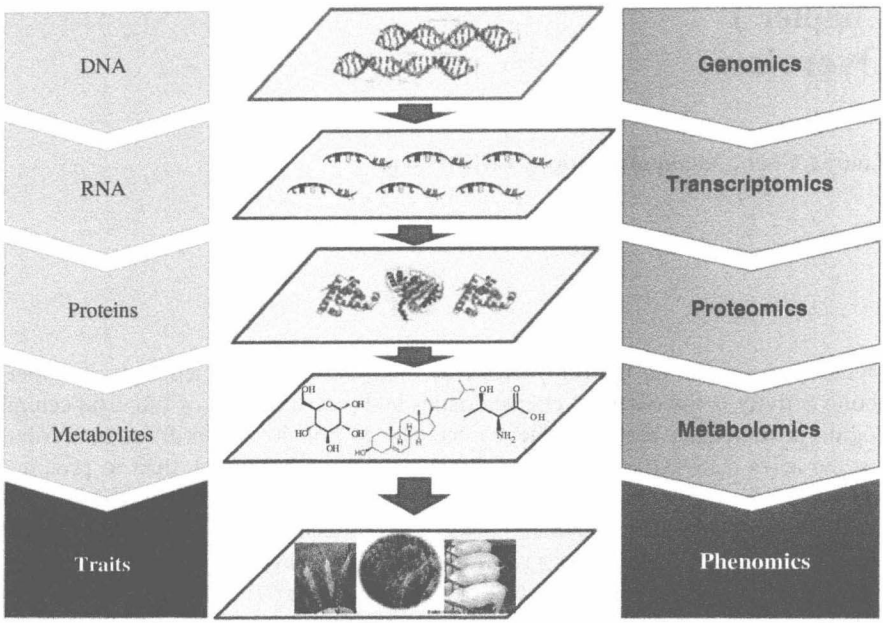


Fig. 1.1 Metabolomics is a branch of systems biology. The multilevel and systematic study is expected to reveal the molecular and metabolic basis for the formation of plant traits

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