

# Rice Metabolomics

Akira Oikawa · Fumio Matsuda · Miyako Kusano ·  
Yoza Okazaki · Kazuki Saito

Received: 25 March 2008 / Accepted: 9 July 2008 / Published online: 15 August 2008  
© Springer Science + Business Media, LLC 2008

**Abstract** Metabolomics is a recently developed technology for comprehensive analysis of metabolites in organisms. Plant metabolites that are produced for the growth, development, and chemical defense of plants against climatic alterations or natural predators are also useful to us as nutrients or medicines; hence, it is important to comprehend the amounts and varieties of plant metabolites. Besides providing an understanding of the metabolic state in plants under various circumstances, metabolomic techniques are applicable to the clarification of the functions of unknown genes by using natural variants or mutants of the target plants. Furthermore, a metabolomic approach might be useful in the breeding of crops, since valuable plant traits such as taste and yield are closely related to metabolic conditions. Here, we describe the methodology of metabolomics including sample extraction and preparation, metabolite detection, and data processing and analysis, and introduce the application of metabolomic studies to rice.

**Keywords** Metabolomics · Mass spectrometry ·  
Bioinformatics · QTL · GM rice

## Introduction

Metabolomics is one of the “omics” studies used for acquiring biological information in a comprehensive manner. In the narrow sense, metabolomics implies a qualitative and quantitative approach for all metabolites present in organisms; contrastingly, metabolite profiling or metabolic profiling is a term used to describe the quantitative and qualitative analysis of only the identified metabolites [9, 29, 35, 41, 80]. In this review, we use the term “metabolomics” for comprehensive metabolite analysis, irrespective of whether the detected metabolites have been identified or not.

The basic motivation for rice studies is crop breeding to acquire valuable traits. Forward and reverse genetics have been utilized to select better rice cultivars among natural varieties and artificial mutants [5, 27, 47, 105]. Conventional genetic methods revealed biological functions of unknown genes that are related to valuable qualities, such as rice grain production [6] and tomato yield [34], leading to the generation of new genetically modified (GM) crops. However, it is difficult to clarify the functions of all genes by only genetic techniques because estimation of functions from genome sequences has limitations. Furthermore, other omics technologies may serve to understand the functions of unknown genes and may refine the candidate genes having valuable traits. Metabolomics provides the possibility of clarifying gene functions directly connected to rice quality, since metabolome, implying the varieties and amounts of all metabolite in organisms, is closely related to the important traits of rice such as the yield, nutrient content, and defense mechanism [27, 105]. Therefore, metabolomics is thought to be a technique that may be applicable to rice breeding by a combination of other omics approaches.

---

A. Oikawa (✉) · F. Matsuda · M. Kusano · Y. Okazaki ·  
K. Saito (✉)

Metabolomics Research Group, RIKEN Plant Science Center,  
Suehiro-cho 1-7-22 Tsurumi-ku,  
Yokohama 230-0045, Japan  
e-mail: oikawa@psc.riken.jp

K. Saito  
Department of Molecular Biology and Biotechnology,  
Graduate school of Pharmaceutical Sciences, Chiba University,  
Yayoi-cho 1-33 Inage-ku,  
Chiba 263-8522, Japan

Among the omics studies in plant science, genomics was the firstly emerged and is the mostly advanced technique; it has uncovered the genome sequences of several organisms including *Arabidopsis* [3] and rice [37, 49, 82, 107]. There should be no doubt that the development of an analytical instrument, i.e., the DNA sequencer, has led to the current progress of genomic studies. Other omics studies have also proceeded as a result of technical innovations. Two-dimensional electrophoresis and mass spectrometer (MS) have become universal methods for proteomic studies. Furthermore, the technological development of DNA microarray has led to the emergence of a new omics area, namely, transcriptomics, which encyclopedically analyzes the messenger RNA (mRNA) expression in biological samples. Similarly, MS and nuclear magnetic resonance spectrometer (NMR) have been facilitating metabolomic studies. However, metabolites have various chemical characters, particularly in the plant metabolites whose estimated number is up to 200,000 [20, 29]. Besides, most of the plant metabolites are still unknown. Due to these reasons, no unified analytical instrument for metabolomics has existed thus far. On the other hand, the techniques of metabolomics have been recognized to be useful for basic researches in the area of plant physiology and biology [9]. Additionally, they have been found to be applicable to the safety assessment of GM crops. Therefore, many positive studies on plant metabolomics have been performed, and therefore, many recent review articles on plant metabolomics can be consulted for general statements [9, 28, 31, 35, 41, 80, 90], methodology [13, 18, 30, 46, 69, 81, 89, 99, 101], and applications [42, 67, 76, 79, 84, 97].

In this review, we describe the current technological developments in plant metabolomics, including sample preparations, analytical methods, and informatics techniques. Furthermore, we show several examples and possibilities of the application of the techniques of metabolomics to the study of rice.

### Technological developments in metabolomics

Metabolomics is a complex field of analytical chemistry and bioinformatics in which advanced techniques to determine the levels of a wide range of metabolites are employed in a series of procedures, including sample extraction and preparation, metabolite detection using analytical instruments, and data processing and mining by means of bioinformatics techniques. Each step includes some alternatives; for example, using 100% or 80% aqueous methanol for extraction and MS or NMR for metabolite analysis. Depending on the combination of these options in each step, subtly or completely different results can be acquired from the same sample itself. Hence, we

should carefully select a proper method in the respective studies. In this section, we mention the details of the metabolomic methods that can be used in each step.

### Sample extraction and preparation

To encompass as many plant metabolites as possible, it is important to carefully perform the operations in the sample extraction and preparation steps before metabolomic analysis. First, we need to select an optimal buffer for sample extraction because the extraction efficiency of metabolites is dependent on the type of the extraction buffer used due to the various chemical characters of plant metabolites. The most frequently used buffer is 80% or 100% aqueous methanol, which enables the extraction of a relatively wide range of metabolites from plant tissues. Extraction performed twice, i.e., once each with polar and non-polar buffers, for one sample is surely considered to be more effective to encompass many metabolites, as compared to extraction at one time with one buffer, although the time and labor for the analysis of each sample are doubled. In this case, hydrophilic and hydrophobic buffers such as 50% methanol and 100% acetonitrile are used [1]. Second, the sample preparation step must be taken into account. This step is conducted to remove the unnecessary compounds and residues that interfere with the separation of metabolites and decrease the sensitivity of metabolite detection. Filtration, ultrafiltration, liquid–liquid extraction, and solid-phase extraction are the main techniques used for sample preparation in plant metabolomics. Depending on the instruments used and the samples, the methods involved in sample preparation have slight variations; for example, when capillary electrophoresis (CE)–MS is used for metabolomic analysis, ultrafiltration and liquid–liquid extraction are employed to remove high-molecular-weight compounds such as proteins and hydrophobic compounds such as chlorophyll, respectively, which adversely affect metabolite separation by CE [88]. In case of sample preparation for gas chromatography (GC)–MS analysis, chemical derivatization should be employed to improve the sensitivity of detection of hydrophilic metabolites. Nevertheless, monitoring the recovery rate during sample preparation by use of suitable internal standards is essential for precise quantification of each detected metabolite.

### Analytical technologies

For detecting as many metabolites as possible, we should consider the physicochemical properties of metabolites, i.e., their hydrophobicity, volatility, and acidity. Therefore, various separation and detection techniques, such as GC–MS [31, 65], liquid chromatography (LC)–MS [62, 70], CE–MS [83, 92], Fourier transform ion cyclotron resonance

(FT-ICR) MS [1, 73, 75] and NMR [61, 85] have been applied to metabolomics. Among them, GC–MS is the most popular method in current metabolomic studies [31]. In addition to volatile compounds, GC–MS can detect hydrophilic metabolites, such as amino acids, organic acids, and sugars, by chemical derivatization of these metabolites. Moreover, in GC–MS analysis using the electron-impact ionization method, it is possible to detect many fragment peaks derived from each compound; these fragment peaks provide information of the structure of the metabolites and thereby help to identify the detected peaks as known metabolites. LC–MS is a powerful tool for profiling of hydrophobic secondary metabolites including alkaloids, flavonoids, and phenylpropanoids due to its high chromatographic performance for separation of these compounds. CE–MS can separate and detect ionic metabolites including amino acids, organic acids, nucleotides, and sugar phosphates without any experiments for chemical derivatization. Most of these metabolites belong to a central metabolic pathway such as the glycolytic pathway and the tricarboxylic acid cycle. Consequently, CE–MS analysis leads to the acquisition of a considerable amount of information on the metabolites of a central metabolic pathway. FT-ICR MS has an ultra-high performance, particularly in terms of mass resolution and sensitivity, leading to the detection of a huge number of peaks without the need for any steps for the separation of metabolites from a mixture by chromatography or electrophoresis. Additionally, FT-ICR MS needs only a few seconds for one scan; thereby, it enables us to perform high-throughput and high-resolution analysis. NMR is also a conducive technique for metabolomics, since it is able to determine the atomic state of compounds and enables the identification of metabolites that are otherwise unidentifiable by MS analysis. Among the various pulse sequence programs,  $^1\text{H-NMR}$  has been routinely employed for high-throughput metabolomic studies due to its relatively short acquisition time per analysis. Alternatively, several conventional as well as innovative analytical instruments, such as Fourier transform infrared spectrometry [51], direct-infusion time-of-flight (TOF) MS [24, 77], ion mobility MS [25], and LC–FT-ICR MS [48, 91], have recently been applied to metabolomic studies.

Meanwhile, thus far, there have been no breakthroughs in the technology for the detection of unstable metabolites or the metabolites that are present in low concentrations. In case of unstable compounds, we have to treat the samples carefully; for example, by avoidance of exposure to strong light, high temperature, and contact with oxygen during sample extraction. On the contrary, because of the narrow dynamic range of the currently available analytical instruments, it is difficult to simultaneously detect the metabolites that are present in high and low concentrations. The existing analytical instruments are inadequate to encompass

all plant metabolites in a single analysis, considering the natural dynamic range of plant metabolites. Additionally, it is laborious and time consuming to analyze diluted or concentrated samples in several batches besides the original one, although this method might compensate for the narrow dynamic range of the analytical instruments and may enable the detection of metabolites independent of their concentrations. Thus, the development of an innovative instrument with a wide dynamic range and a high sensitivity is expected.

#### Informatics techniques

We encounter a large amount of data after instrumental analysis, particularly in case of high-performance instruments that can detect tiny peaks with high resolution and thus increase the number of signals detected. Since these large datasets cannot be handled manually, we need an automatic high-throughput precise software that is capable of picking up peaks from mass or NMR spectra, aligning the peaks among the samples, and identifying or estimating and quantifying each metabolite. Therefore, informatics is also an essential tool to deal with large metabolomic datasets [4]. Recently, a number of software that can be downloaded from websites are available [15]; for example, “DrDmassPlus” ([75], <http://kanaya.naist.jp/DrDMassPlus/>) can be used to treat metabolome data from FT-ICR MS, including the functions of peak picking, peak alignment, and some statistical analyses. Even if a metabolomic dataset contains a large amount of data derived from over thousands of detected peaks, DrDmass-Plus can process it at high throughput. In addition to such data-processing softwares, we should take advantage of the compounds database in metabolomic studies. Recently, some databases have been constructed for metabolomics. KNApSAcK ([87], <http://kanaya.naist.jp/KNApSAcK/>) is one of the comprehensive metabolite databases containing information on 21,599 metabolites (March 21, 2008) including their names, chemical formulae, molecular weights, CAS numbers, the organisms from which they are derived, and the reference research papers that have reported them. Furthermore, we can project the detected  $m/z$  values to the KNApSAcK database and whether this can easily line up the candidate metabolites for each detected peak.

In metabolomic studies, statistical analysis is often employed to understand the comprehensive differences in the detected metabolites among various samples [29, 35, 41]. Various statistical methods used in conventional genetic studies are applicable to metabolomic data by considering the amount of each metabolite as a trait value. Principal component analysis (PCA), one method of multivariate analyses, is commonly used in metabolomic

studies. There have been many reports on the application of PCA to metabolomic data [7, 12, 19, 54, 65, 71, 75, 93, 94, 95]. In addition, several statistical analytical methods have been used for analysis of metabolomic datasets; for example, hierarchical cluster analysis (HCA) [39, 78], partial least squares discriminant analysis (PLS-DA) [52, 65], and batch-learning self-organizing map (BL-SOM) [43, 44, 54]. Depending on the objective of each study, the most appropriate statistical analytical method should be exploited to analyze the metabolomic data.

### Application of metabolomics to rice study

Since rice is one of the most important crops in the world, there is a constant challenge to generate better rice cultivars with valuable traits, although it has been performed using traditional and molecular breeding techniques [105]. Recently, it has been found that metabolomic methods help to generate new rice cultivars in the context of identification of valuable gene functions and safety assessment of GM crops [7, 12, 58, 105]. In this section, we elucidate actual examples of applications of metabolomic techniques to rice studies from two viewpoints: the application of metabolomics to the study of rice biology, including the detection of metabolic changes according to the natural genetic variations in rice plants and the application of metabolomics to the production of GM rice, including the evaluation of substantial equivalence.

#### Metabolomics for rice biology

Analysis of the natural variations in rice using metabolomic techniques is thought to be not only useful to understand the biological traits of rice, such as the yield and defense responses, but is also helpful to improve rice quality, including its taste and nutritive value [27]. Kusano et al. [65] used a metabolomic method combining one-dimensional (1D) and two-dimensional (2D) GC-TOF MS for metabolic phenotyping of the 68 rice varieties in the world rice core collection (WRC) and two other varieties. The WRC is a representative set of the cultivated rice (*Oriza sativa* L.) and covers approximately 90% of the DNA polymorphisms detected in the original population [57]. In this study, in order to characterize the metabolic phenotypes of WRC, metabolomic analysis of the extracts of brown rice seeds was conducted as follows. First, 1D-GC-TOF MS analysis in combination with multivariate statistical analysis, i.e., PLS-DA, was performed to narrow down the representative varieties, which were selected on the basis of the criterion of good separation from the other varieties in PLS-DA. Second, the three selected varieties were subjected to a detailed 2D-GC-TOF/MS analysis for high-resolution me-

tabolite profiling. Finally, we statistically compared the ten metabolites whose concentrations in the three selected cultivars were different from those in a control rice cultivar (Nipponbare). Through these experiments, the selected varieties were successfully characterized in terms of the differences in their metabolite compositions. For instance, although the selected variety Urasan 1 is considered to be a Japonica variety on the basis of its DNA polymorphisms, this variety showed a unique metabolic profile when compared to other Japonica varieties. This result may suggest that the metabolomic approach can evaluate the quality of rice cultivars according to their metabolite compositions.

Changes in the concentrations of the metabolites according to the developmental or diurnal periods were also investigated using metabolomic techniques. The combination of GC-MS and statistical analyses was exploited to discover certain biomarker metabolites that occur in rice plants in the developmental period [94]. In this study, first, rice samples in various developmental stages were extracted using chloroform and water, and both hydrophobic and hydrophilic extracts were analyzed by GC-MS. Next, the metabolome data was projected onto PCA. The principal component representing the differences in the developmental period of rice was determined, and then, the loading metabolites on this axis were found to be candidates for the biomarker metabolites related to the developmental period of rice. Finally, the physiological roles of each biomarker candidate were discussed. This successive flow from sample extraction to statistical analysis appears to be the typical method for current metabolomic studies. Moreover, Sato et al. [83] applied CE-MS and CE-DAD (DAD, diode array detector) technology to detect water-soluble metabolites in rice leaves during the day and night. They measured the levels of 88 metabolites, including amino acids, organic acids, nucleotides, and sugars, whose detection limits ranged from 0.3 to 71  $\mu\text{mol}\cdot\text{l}^{-1}$ . In this study, metabolite fluctuations within a single day were discussed further by comparison with the results reported in the earlier investigations. Furthermore, there have been reports on the comprehensive metabolite analysis of root exudates [26] and flavor volatiles [36] of rice plants. Thus, recently, there have been various examples of applications of metabolomic techniques to rice plants, indicating that metabolomics is accepted to be a potent tool for the study of rice biology.

#### Metabolomics for GM rice

Molecular breeding of GM rice has been attempted in order to improve its nutritional quality [17, 105]. One of the manipulation targets is the engineering of a metabolic pathway to produce rice cultivars with a high content of

valuable metabolites, such as  $\beta$ -carotene in “golden rice” [104] and tryptophan. High tryptophan rice is a prospective GM crop because tryptophan is an essential amino acid for mammals and a growth-encouraging fodder for livestock. Because tryptophan can presently be produced industrially by fermentation, high tryptophan GM rice may serve as a functional food or fodder with a health- or growth-promoting benefit. Normally, a low level of tryptophan is maintained in plants as a result of the feedback inhibition of anthranilate synthase (AS) by tryptophan [10, 66]. Hence, AS has been considered to be the prime target in order to increase the amount of tryptophan [14, 108]. Tozawa et al. [96] cloned 2 AS  $\alpha$ -subunit genes (*OASA1*, *OASA2*) from rice (*Oryza sativa* cv. Nipponbare) and produced transgenic rice expressing a feedback-insensitive *OASA1* gene (*OASAIID*) in which aspartate-323 is replaced by asparagine [96]. Detailed evaluation of *OASAIID* transgenic rice revealed that the levels of free tryptophan in the calli, seedlings, and seeds were increased up to 35- to 300-folds [96, 100], while the other agronomic traits in the field-grown transgenics were essentially identical to those of untransformed variety [100]. Besides the visible phenotypes, it might be expected that the recombination of tryptophan biosynthetic genes affected other invisible phenotypes, such as the concentration of metabolites that are related or unrelated to the tryptophan biosynthetic pathway [50]. Indeed, it has been reported that a change in the pool size of tryptophan in potato activated the biosynthesis of phenylpropanoid compounds [103]. In addition, since tryptophan is a precursor for the biosynthesis of various plant metabolites including auxin, indole-3-acetic acid, [8] and indole alkaloids [98], some of which are toxic or bitter for mammals, the overaccumulation of tryptophan is likely to be associated with the production of these by-products. However, prediction of the side effects has been difficult due to our incomplete understanding of the regulation of metabolism in plants [50]. Thus, the use of metabolomic techniques as well as other omics tools to survey the wide range of metabolites is important to evaluate the substantial equivalence of GM crops [7, 12, 19]. Microarray and targeted metabolite analyses of *OASAIID* rice seedlings showed that the effect of the transgene was very limited, probably due to the low activity of the tryptophan-utilizing pathways in rice seedlings [23]. Profiling analyses of UV-active metabolites in the rice calli and seeds expressing the *OASAIID* transgene also demonstrated that drastic accumulation of tryptophan had little effects on the profiles of other UV-active metabolites except for the accumulation of a few minor indole alkaloids in these tissues [72, 100]. From these results, it can be predicted that the unchanged metabolic profiles are the consequences of the low secondary metabolic activity in rice, suggesting that transformation with *OASAIID* might

prove effective in the breeding of crops having an increased tryptophan content.

In addition to the studies mentioned above, there have been a number of reports regarding the application of metabolomics to the production of GM rice; for example, to understand the correlation between gene overexpression and the concentrations of metabolites, the GM rice overexpressing the *YK1* gene, which is known as the homologue of the HC-toxin reductase gene and has a dihydroflavonol-4-reductase activity, was subjected to metabolomic analysis using FT-ICR MS and CE-MS [92, 93]. Slight changes were detected in the amount of amino acids and organic acids in the leaves, roots, seeds, and calli of this GM rice. At the same time, increase in the level of glutathione derivatives was found in the calli. The observation on the relationship between the overexpression of the *YK1* gene and accumulation of glutathione derivatives was hence summarized in this report. Moreover, for the elucidation of the varieties of mutants that have a small amount of phytic acid considered to be an antinutrient factor in rice grains, several types of low phytic acid-containing mutants (*lpa*) were analyzed by means of metabolite profiling using GC [33]. The metabolic changes in the mutant rice grown in different fields were also investigated. Using these metabolomic data, the effects of the fields in which the rice varieties were grown and the mutations on the concentrations of various metabolites were discussed. As is clear from these studies, metabolomic approaches may enable the understanding of the influences of genetic modifications on plants metabolite concentrations.

Molecular breeding of GM rice by introducing various transgenes is expected to be a promising approach for the production of rice cultivars with preferred agronomical traits [17, 105]. However, before GM rice enters the market, its potential risks need to be assessed for ensuring safety [63, 64]. Metabolomics combined with other allergenicity and toxicity tests [22, 38] plays an important role in risk management [19]. However, it should be noted that although the survey of the complete metabolome is required for detailed scrutiny of risks, the number of detectable metabolites by currently available metabolomic techniques is limited. Likewise, since the scientific procedures or guidelines for assessment of possible risks using metabolome data have not been established yet [21], ensuring substantial equivalence tends to require considerable work and sometimes might be impossible [38]. These drawbacks indicate that further improvement in metabolomics is essential for more comprehensive profiling and interpretation of rice metabolites.

### Future prospects

Plant metabolomics has a possibility to be applied to various purposes in addition to the studies exemplified

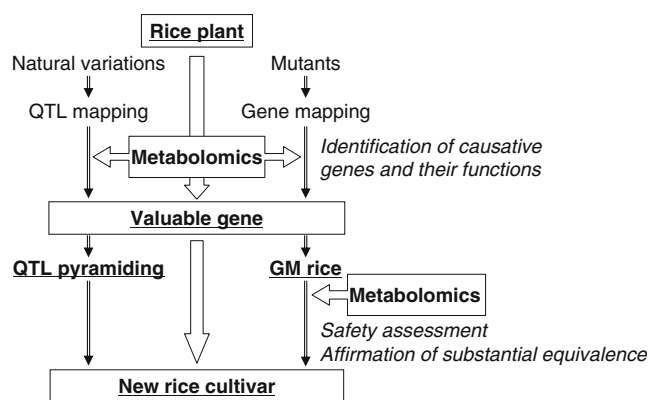
above. Recently, several *Arabidopsis* researches have demonstrated that integration analysis of transcriptomic and metabolomic datasets is a promising strategy for the functional identification of metabolism-related genes; for example, Hirai et al. [43, 44] processed the transcriptomic and metabolomic data of the *Arabidopsis* plants under the sulfur- and nitrogen-deficient conditions. In these studies, BL-SOM analysis clearly revealed the coordinated changes in the mRNA and metabolite levels in the stressed plants, thereby enabling the identification of the functions of several unknown genes [43]. Considerable transcriptomic data on *Arabidopsis* can be easily acquired on the web site of AtGenExpress (<http://www.arabidopsis.org/info/expression/ATGenExpress.jsp>), and the expression correlation data is now available in other databases such as ATTED-II [74]. The data in these databases makes it possible to identify the candidate genes for unknown biological processes by evaluating the correlation of the expression of these genes with the functionally identified genes [41, 76, 81]; these databases have facilitated the hunting of secondary metabolism-related genes of *Arabidopsis* [106]. In case of rice, Rice Expression Database (<http://red.dna.affrc.go.jp/RED/>) has been already released for the processing and mining of rice DNA microarray data. The integration of transcriptomic and metabolomic data should be a robust tool in rice studies in the foreseeable future.

In a modern *Arabidopsis* research, the loss-of-function or gain-of-function mutant lines play a very important role in the functional genomics study. A combination of the metabolomic approach and these bioresources has been demonstrated to be a smart strategy to uncover the role of a functionally unknown gene. Using GC–MS and CE–MS, Watanabe et al. [102] obtained a metabolome dataset of *Arabidopsis* seedlings of a wild-type plant, and several T-DNA insertion mutants having immature *bsas* ( $\beta$ -substituted alanine synthase) genes. Statistical analyses revealed that one unknown metabolite was not accumulated in one of the *bsas* mutants, namely, *bsas3;1*, whereas this unknown metabolite was detectable in sufficient quantities in the wild-type and other mutants. Next, they identified the unknown metabolite as a unique dipeptide based on a comparison with an enzymatically synthesized dipeptide. The *bsas3;1* gene was eventually estimated to be included in the biosynthesis of this dipeptide. Correspondingly, by the exploitation of large bioresources of rice mutants [2, 45, 47], the metabolomic approach might be able to reveal the functions of various rice genes.

Furthermore, in addition to the clarification of the function of single genes, metabolomics has the potential to determine the location of quantitative trait loci (QTL). QTL have traditionally been statistically analyzed by comparison between the genotypes of marker genes and the occurrence of the trait among naturally occurring

genetic variants [5, 11, 27, 59]. Recently, transcriptomic information has been used for QTL, which is referred to as expression QTL (eQTL), thereby clarifying the genes directly related to metabolism, such as those coding for enzymes and transporters, and the regulator genes such as the transcription factors [16, 40, 53, 55, 56, 86]. In addition to transcriptomic data, metabolomic data may be capable of application to QTL analysis. Very recently, Liseč et al. [68] reported the results of metabolic QTL analysis of *Arabidopsis*. In this study, using recombinant inbred lines and introgression lines of *Arabidopsis*, biomass, and metabolic QTL were compared. Eventually, it was suggested that the amounts of several metabolites were strongly linked to *Arabidopsis* biomass. Because the variety and amounts of metabolites are directly connected to the important traits of plants such as biomass, nutrition, and taste, QTL analysis combined with metabolomics should be a promising method.

Figure 1 summarizes the possible role of metabolomics in a sequence of investigations for rice breeding. Conventional forward and reverse genetics, such as QTL mapping derived from the phenotypes of natural rice variants or target gene mapping of spontaneously or artificially occurring mutants, have revealed the functions of valuable genes [11, 47]. Since metabolomics is a potent tool for the identification of valuable gene functions, the role of beneficial genes that can be utilized to generate a useful rice cultivar might be also elucidated by metabolomic approaches. Meanwhile, crossbreeding based on QTL pyramiding or genetic modification of target genes are the current techniques that are used to develop improved cultivars. Affirmation of substantial equivalence is essential for safety assessment of the generated rice, particularly GM rice. Metabolomic techniques might be able to play a role in ascertaining whether unexpected and undesirable compounds are accumulated in GM rice.



**Fig. 1** The sequence of steps in rice breeding. Metabolomic techniques are applicable to develop useful rice cultivars having superior qualities and particularly for the clarification of the functions of valuable genes and safety assessment of GM rice.

The methodology in metabolomics tends to converge several analytical instruments such as directly infused high-performance MS, hyphenated MS, and NMR. However, the development of a more high-throughput and high-sensitive method is expected now. Simultaneously, with the innovations in analytical technologies and methods, a more useful and high-throughput data processing software is also expected. Furthermore, besides the existing metabolomic database that can be used for an encyclopedia of metabolites, such as the KNApSACk, Golm Metabolite Database ([60], <http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html>), and MoToDB ([70], <http://appliedbioinformatics.wur.nl/moto/>), more metabolomic databases of other plants including rice should hopefully become available. On the other hand, the necessity of quality control in plant metabolomics has recently been proposed [32]. Standardization of metabolomic experiments from the biological study design to data analysis will not only facilitate metabolomic studies but will also exploit the existing metabolomic data. These technical improvements in metabolomic studies will contribute to the progress of the basic and applied studies on rice.

## References

- Aharoni A, Ric de Vos CH, Verhoeven HA, Maliepaard CA, Kruppa G, Bino R, et al. Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. *OMICS* 2002;6:217–34.
- An G, Lee S, Kim SH, Kim SR. Molecular genetics using T-DNA in rice. *Plant Cell Physiol* 2005;46:14–22.
- Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000;408:796–815.
- Arita M. Additional paper: computational resources for metabolomics. *Brief Funct Genomic Proteomic* 2004;3:84–93.
- Ashikari M, Matsuoka M. Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends Plant Sci* 2006;11:344–50.
- Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, et al. Cytokinin oxidase regulates rice grain production. *Science* 2005;309:741–5.
- Baker JM, Hawkins ND, Ward JL, Lovegrove A, Napier JA, Shewry PR, et al. A metabolomic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotechnol J* 2006;4:381–92.
- Bartel B. Auxin biosynthesis. *Ann Rev Plant Physiol Plant Mol Biol* 1997;48:51–66.
- Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, et al. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 2004;9:418–25.
- Bohmann J, Lins T, Martin W, Eilert U. Anthranilate synthase from *Ruta graveolens*. Duplicated AS alpha genes encode tryptophan-sensitive and tryptophan-insensitive isoenzymes specific to amino acid and alkaloid biosynthesis. *Plant Physiol* 1996;111:507–14.
- Borevitz JO, Chory J. Genomics tools for QTL analysis and gene discovery. *Curr Opin Plant Biol* 2004;7:132–6.
- Catchpole GS, Beckmann M, Enot DP, Mondhe M, Zywicki B, Taylor J, et al. Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc Natl Acad Sci U S A* 2005;102:14458–62.
- Chen C, Gonzalez FJ, Idle JR. LC-MS-based metabolomics in drug metabolism. *Drug Metab Rev* 2007;39:581–97.
- Cho HJ, Brotherton JE, Song HS, Widholm JM. Increasing tryptophan synthesis in a forage legume *Astragalus sinicus* by expressing the tobacco feedback-insensitive anthranilate synthase (ASA2) gene. *Plant Physiol* 2000;123:1069–76.
- Codrea MC, Jimenez CR, Heringa J, Marchiori E. Tools for computational processing of LC-MS datasets: a user's perspective. *Comput Methods Programs Biomed* 2007;86:281–90.
- DeCook R, Lall S, Nettleton D, Howell SH. Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics* 2006;172:1155–64.
- DellaPenna D. Plant metabolic engineering. *Plant Physiol* 2001;125:160–3.
- Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 2007;26:51–78.
- Dixon RA, Gang DR, Charlton AJ, Fiehn O, Kuiper HA, Reynolds TL, et al. Applications of metabolomics in agriculture. *J Agric Food Chem* 2006;54:8984–94.
- Dixon RA, Strack D. Phytochemistry meets genome analysis, and beyond. *Phytochemistry* 2003;62:815–6.
- Domingo JL. Health risks of GM foods: many opinions but few data. *Science* 2000;288:1748–9.
- Domingo JL. Toxicity studies of genetically modified plants: a review of the published literature. *Crit Rev Food Sci Nutr* 2007;47:721–33.
- Dubouzet JG, Ishihara A, Matsuda F, Miyagawa H, Iwata H, Wakasa K. Integrated metabolomic and transcriptomic analyses of high-tryptophan rice expressing a mutant anthranilate synthase alpha subunit. *J Exp Bot* 2007;58:3309–21.
- Dunn WB, Overy S, Quick WP. Evaluation of automated electrospray-TOF mass spectrometry for metabolic fingerprinting of the plant metabolome. *Metabolomics* 2005;1:137–48.
- Dwivedi P, Wu P, Klopsch S, Puzon G, Xun L, Hill H. Metabolic profiling by ion mobility mass spectrometry (IMMS). *Metabolomics* 2008;4:63–80.
- Fan TW, Lane AN, Shenker M, Bartley JP, Crowley D, Higashi RM. Comprehensive chemical profiling of gramineous plant root exudates using high-resolution NMR and MS. *Phytochemistry* 2001;57:209–21.
- Fernie AR, Tadmor Y, Zamir D. Natural genetic variation for improving crop quality. *Curr Opin Plant Biol* 2006;9:196–202.
- Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L. 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. Metabolite profiling in *Arabidopsis*. *Methods Mol Biol* 2006;323:439–47.
- Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. Metabolite profiling for plant functional genomics. *Nat Biotechnol* 2000;18:1157–61.
- Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee do Y, Lu Y, et al. Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J* 2008;53:691–704.
- Frank T, Meuleye BS, Miller A, Shu QY, Engel KH. Metabolite profiling of two low phytic acid (lpa) rice mutants. *J Agric Food Chem* 2007;55:11011–9.
- Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 2004;305:1786–9.

35. Fukusaki E, Kobayashi A. Plant metabolomics: potential for practical operation. *J Biosci Bioeng* 2005;100:347–54.
36. Ghiasvand AR, Setkova L, Pawliszyn J. Determination of flavour profile in Iranian fragrant rice samples using cold-fibre SPME-GC-TOF-MS. *Flav Frag J* 2007;22:377–91.
37. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 2002;296:92–100.
38. Goodman RE, Vieths S, Sampson HA, Hill D, Ebisawa M, Taylor SL, et al. Allergenicity assessment of genetically modified crops—what makes sense? *Nat Biotechnol* 2008;26:73–81.
39. Grata E, Boccard J, Glauser G, Carrupt PA, Farmer EE, Wolfender JL, et al. Development of a two-step screening ESI-TOF-MS method for rapid determination of significant stress-induced metabolome modifications in plant leaf extracts: the wound response in *Arabidopsis thaliana* as a case study. *J Sep Sci* 2007;30:2268–78.
40. Guillaumie S, Charmet G, Linossier L, Torney V, Robert N, Ravel C. Colocation between a gene encoding the bZip factor SPA and an eQTL for a high-molecular-weight glutenin subunit in wheat (*Triticum aestivum*). *Genome* 2004;47:705–13.
41. Hall RD. Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol* 2006;169:453–68.
42. Hall RD, Brouwer ID, Fitzgerald MA. Plant metabolomics and its potential application for human nutrition. *Physiol Plant* 2008;132:162–75.
43. Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, et al. Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J Biol Chem* 2005;280:25590–95.
44. Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuha M, et al. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 2004;101:10205–10.
45. Hirochika H. Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 2001;4:118–22.
46. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. *Proteomics* 2006;6:4716–23.
47. Hsing YI, Chern CG, Fan MJ, Lu PC, Chen KT, Lo SF. A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol Biol* 2007;63:351–64.
48. Iijima Y, Nakamura Y, Ogata Y, Tanaka K, Sakurai N, Suda K. Metabolite annotations based on the integration of mass spectral information. *Plant J* 2008;54:949–62.
49. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* 2005;436:793–800.
50. Ishihara A, Matsuda F, Miyagawa H, Wakasa K. Metabolomics for metabolically manipulated plants: effects of tryptophan overproduction. *Metabolomics* 2007;3:319–34.
51. Johnson HE, Broadhurst D, Kell DB, Theodorou MK, Merry RJ, Griffith GW. High-throughput metabolic fingerprinting of legume silage fermentations via Fourier transform infrared spectroscopy and chemometrics. *Appl Environ Microbiol* 2004;70:1583–92.
52. Jonsson P, Gullberg J, Nordstrom A, Kusano M, Kowalczyk M, Sjostrom M, et al. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal Chem* 2004;76:1738–45.
53. Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, Van Den Ackerveken G, Snoek LB, et al. Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci U S A* 2007;104:1708–13.
54. Kim JK, Bamba T, Harada K, Fukusaki E, Kobayashi A. Time-course metabolic profiling in *Arabidopsis thaliana* cell cultures after salt stress treatment. *J Exp Bot* 2007;58:415–24.
55. Kirst M, Basten CJ, Myburg AA, Zeng ZB, Sederoff RR. Genetic architecture of transcript-level variation in differentiating xylem of a eucalyptus hybrid. *Genetics* 2005;169:2295–303.
56. Kliebenstein DJ, West MA, Van Leeuwen H, Loudet O, Doerge RW, St Clair DA. Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics* 2006;7:308.
57. Kojima Y, Ebana K, Fukuoka S, Nagamine T, Kawase M. Development of an RFLP-based Rice Diversity Research Set of Germplasm. *Breeding Science* 2005;55:431–40.
58. Kok EJ, Kuiper HA. Comparative safety assessment for biotech crops. *Trends Biotechnol* 2003;21:439–44.
59. Koornneef M, Alonso-Blanco C, Vreugdenhil D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu Rev Plant Biol* 2004;55:141–72.
60. Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, et al. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 2005;21:1635–8.
61. Krishnan P, Kruger NJ, Ratcliffe RG. Metabolite fingerprinting and profiling in plants using NMR. *J Exp Bot* 2005;56:255–65.
62. Kristensen C, Morant M, Olsen CE, Ekstrom CT, Galbraith DW, Moller BL, et al. Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. *Proc Natl Acad Sci U S A* 2005;102:1779–84.
63. Kuiper HA, Kleter GA, Noteborn HP, Kok EJ. Assessment of the food safety issues related to genetically modified foods. *Plant J* 2001;27:503–28.
64. Kuiper HA, Kok EJ, Engel KH. Exploitation of molecular profiling techniques for GM food safety assessment. *Curr Opin Biotechnol* 2003;14:238–43.
65. Kusano M, Fukushima A, Kobayashi M, Hayashi N, Jonsson P, Moritz T, et al. Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. *J Chromatogr B Analyt Technol Biomed Life Sci* 2007;855:71–9.
66. Li J, Last RL. The *Arabidopsis thaliana* trp5 mutant has a feedback-resistant anthranilate synthase and elevated soluble tryptophan. *Plant Physiol* 1996;110:51–9.
67. Lindon JC, Holmes E, Nicholson JK. Metabonomics in pharmaceutical R&D. *FEBS J* 2007;274:1140–51.
68. Liseic J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, et al. Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J* 2008;53:960–72.
69. Liseic J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 2006;1:387–96.
70. Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, et al. A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol* 2006;141:1205–18.
71. Moco S, Capanoglu E, Tikunov Y, Bino RJ, Boyacioglu D, Hall RD, et al. Tissue specialization at the metabolite level is perceived during the development of tomato fruit. *J Exp Bot* 2007;58:4131–46.
72. Morino K, Matsuda F, Miyazawa H, Sukegawa A, Miyagawa H, Wakasa K. Metabolic profiling of tryptophan-overproducing rice calli that express a feedback-insensitive alpha subunit of anthranilate synthase. *Plant Cell Physiol* 2005;46:514–21.
73. Nakamura Y, Kimura A, Saga H, Oikawa A, Shinbo Y, Kai K, et al. Differential metabolomics unraveling light/dark regulation of metabolic activities in *Arabidopsis* cell culture. *Planta* 2007;227:57–66.
74. Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, et al. ATTED-II: a database of co-expressed genes and



- cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Res* 2007;35:D863–9.
75. Oikawa A, Nakamura Y, Ogura T, Kimura A, Suzuki H, Sakurai N, et al. Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. *Plant Physiol* 2006;142:398–413.
  76. Oksman-Caldentey KM, Saito K. Integrating genomics and metabolomics for engineering plant metabolic pathways. *Curr Opin Biotechnol* 2005;16:174–9.
  77. Overy SA, Walker HJ, Malone S, Howard TP, Baxter CJ, Sweetlove LJ, et al. Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. *J Exp Bot* 2005;56:287–96.
  78. Parveen I, Moorby JM, Fraser MD, Allison GG, Kopka J. Application of gas chromatography-mass spectrometry metabolite profiling techniques to the analysis of heathland plant diets of sheep. *J Agric Food Chem* 2007;55:1129–38.
  79. Rischer H, Oksman-Caldentey KM. Unintended effects in genetically modified crops: revealed by metabolomics? *Trends Biotechnol* 2006;24:102–4.
  80. Saito K, Dixon R, Willmitzer L. *Plant metabolomics*. Heidelberg: Springer; 2006.
  81. Saito K, Hirai MY, Yonekura-Sakakibara K. Decoding genes with coexpression networks and metabolomics—'majority report by precogs'. *Trends Plant Sci* 2008;13:36–43.
  82. Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, et al. The genome sequence and structure of rice chromosome 1. *Nature* 2002;420:312–6.
  83. Sato S, Soga T, Nishioka T, Tomita M. Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *Plant J* 2004;40:151–63.
  84. Schauer N, Fernie AR. Plant metabolomics: towards biological function and mechanism. *Trends Plant Sci* 2006;11:508–16.
  85. Sekiyama Y, Kikuchi J. Towards dynamic metabolic network measurements by multi-dimensional NMR-based fluxomics. *Phytochemistry* 2007;68:2320–9.
  86. Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Lubberstedt T. Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population. *BMC Genomics* 2007;8:22.
  87. Shinbo Y, Nakamura Y, Altaf-Ul-Amin M, Asahi H, Kurokawa K, Arita M, et al. KNApSACK: a comprehensive species-metabolite relationship database. *Biotechnol Agric For* 2006; 57:166–81.
  88. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* 2003;2:488–94.
  89. Steinfath M, Groth D, Lisek J, Selbig J. Metabolite profile analysis: from raw data to regression and classification. *Physiol Plant* 2008;132:150–61.
  90. Sumner LW, Mendes P, Dixon RA. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 2003;62:817–36.
  91. Suzuki H, Sasaki R, Ogata Y, Nakamura Y, Sakurai N, Kitajima M, Takayama H, Kanaya S, Aoki K, Shibata D, Saito K. Metabolic profiling of flavonoids in *Lotus japonicus* using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Phytochemistry* 2008;69:99–111.
  92. Takahashi H, Hayashi M, Goto F, Sato S, Soga T, Nishioka T, et al. Evaluation of metabolic alteration in transgenic rice over-expressing dihydroflavonol-4-reductase. *Ann Bot (Lond)* 2006;98:819–25.
  93. Takahashi H, Hotta Y, Hayashi M, Kawai-Yamada M, Komatsu S, Uchimiya H. High throughput metabolome and proteome analysis of transgenic rice plants (*Oryza sativa* L.). *Plant Biotechnol* 2005;22:47–50.
  94. Tarpley L, Duran AL, Kebrom TH, Sumner LW. Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. *BMC Plant Biol* 2005;5:8.
  95. Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, et al. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J* 2005;42:218–35.
  96. Tozawa Y, Hasegawa H, Terakawa T, Wakasa K. Characterization of rice anthranilate synthase alpha-subunit genes OASA1 and OASA2. Tryptophan accumulation in transgenic rice expressing a feedback-insensitive mutant of OASA1. *Plant Physiol* 2001;126:1493–506.
  97. Trethewey RN. Metabolite profiling as an aid to metabolic engineering in plants. *Curr Opin Plant Biol* 2004;7:196–201.
  98. Van Der Heijden R, Jacobs DI, Snoeijer W, Hallard D, Verpoorte R. The Catharanthus alkaloids: pharmacognosy and biotechnology. *Curr Med Chem* 2004;11:607–28.
  99. Villas-Boas SG, Mas S, Akesson M, Smedsgaard J, Nielsen J. Mass spectrometry in metabolome analysis. *Mass Spectrom Rev* 2005;24:613–46.
  100. Wakasa K, Hasegawa H, Nemoto H, Matsuda F, Miyazawa H, Tozawa Y, et al. High-level tryptophan accumulation in seeds of transgenic rice and its limited effects on agronomic traits and seed metabolite profile. *J Exp Bot* 2006;57:3069–78.
  101. Ward JL, Baker JM, Beale MH. Recent applications of NMR spectroscopy in plant metabolomics. *FEBS J* 2007;274:1126–31.
  102. Watanabe M, Kusano M, Oikawa A, Fukushima A, Noji M, Saito K. Physiological roles of the beta-substituted alanine synthase gene family in Arabidopsis. *Plant Physiol* 2008;146:310–20.
  103. Yao K, De Luca V, Brisson N. Creation of a metabolic sink for tryptophan alters the phenylpropanoid pathway and the susceptibility of potato to *Phytophthora infestans*. *Plant Cell* 1995; 7:1787–99.
  104. Ye X, Al-Babili S, Klott A, Zhang J, Lucca P, Beyer P, Potrykus I. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 2000; 287:303–5.
  105. Yonekura-Sakakibara K, Saito K. Review: genetically modified plants for the promotion of human health. *Biotechnol Lett* 2006;28:1983–91.
  106. Yonekura-Sakakibara K, Tohge T, Niida R, Saito K. Identification of a Flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *J Biol Chem* 2007;282:14932–41.
  107. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 2002;296:79–92.
  108. Zhang XH, Brotherton JE, Widholm JM, Portis AR Jr. Targeting a nuclear anthranilate synthase alpha-subunit gene to the tobacco plastid genome results in enhanced tryptophan biosynthesis. Return of a gene to its pre-endosymbiotic origin. *Plant Physiol* 2001;127:131–41.