Journal of Experimental Botany, Vol. 60, No. 1, pp. 87–97, 2009
doi:10.1093/jxb/ern263 Advance Access publication 10 December, 2008
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Modules of co-regulated metabolites in turmeric (*Curcuma longa*) rhizome suggest the existence of biosynthetic modules in plant specialized metabolism

Zhengzhi Xie[1,2], Xiaoqiang Ma[1] and David R. Gang[1,*]

[1] Department of Plant Sciences and BIO5 Institute, 1657 E. Helen Street, University of Arizona, Tucson, AZ 85721, USA
[2] Department of Pharmaceutical Sciences, University of Arizona, Tucson, AZ 85721, USA

Received 16 June 2008; Revised 29 September 2008; Accepted 2 October 2008

Abstract

Turmeric is an excellent example of a plant that produces large numbers of metabolites from diverse metabolic pathways or networks. It is hypothesized that these metabolic pathways or networks contain biosynthetic modules, which lead to the formation of metabolite modules—groups of metabolites whose production is co-regulated and biosynthetically linked. To test whether such co-regulated metabolite modules do exist in this plant, metabolic profiling analysis was performed on turmeric rhizome samples that were collected from 16 different growth and development treatments, which had significant impacts on the levels of 249 volatile and non-volatile metabolites that were detected. Importantly, one of the many co-regulated metabolite modules that were indeed readily detected in this analysis contained the three major curcuminoids, whereas many other structurally related diarylheptanoids belonged to separate metabolite modules, as did groups of terpenoids. The existence of these co-regulated metabolite modules supported the hypothesis that the 3-methoxyl groups on the aromatic rings of the curcuminoids are formed before the formation of the heptanoid backbone during the biosynthesis of curcumin and also suggested the involvement of multiple polyketide synthases with different substrate selectivities in the formation of the array of diarylheptanoids detected in turmeric. Similar conclusions about terpenoid biosynthesis could also be made. Thus, discovery and analysis of metabolite modules can be a powerful predictive tool in efforts to understand metabolism in plants.

Key words: Biosynthesis, *Curcuma longa*, curcumin, metabolite module, metabolomics, rhizome, specialized metabolism.

Introduction

A very important but still largely unanswered question in plant metabolism is: how is the large number (>200 000 or more has been claimed) and diversity of metabolites observed in the plant kingdom produced, given the relatively small number of genes in plant genomes? Plant metabolism has most often been viewed as consisting of pathways or networks of specific reactions leading from common precursors to specific end-products. In this view, diversity is partially explained by enzyme promiscuity or by gene duplication followed by divergent evolution across the plant kingdom, leading to variations on common pathways or networks. In the case of plants like *Arabidopsis* and rice, where around 5000 metabolites have been hypothesized to be produced by the plant as a whole, the genome, with ~30% of the genes dedicated to metabolism, may be able to account for the number of metabolites present. In the case of plants like turmeric and ginger, two medicinal plants in the Zingiberaceae with genome sizes comparable to rice but with metabolic capacity far exceeding *Arabidopsis* or rice, the situation becomes less clear. Rhizome extracts of ginger and turmeric contain thousands of easily detectable metabolites (Jiang *et al.*, 2005, 2006*b*, *c*, 2007; Ma and Gang, 2005, 2006) whose levels and composition change through development, and are very different between tissue types. Although we have learned much about the major branches of the plant metabolic network over the last several decades, the mechanisms responsible

* To whom correspondence should be addressed: E-mail: gang@ag.arizona.edu

for the formation of this large array of compounds in these plants are still not fully defined, hence the great interest by many groups around the world to use modern tools to address unanswered questions in plant metabolism (Dixon *et al.*, 2005; Hirai *et al.*, 2005*a*, *b*; Deavours *et al.*, 2006; Sawada *et al.*, 2006; Kusano *et al.*, 2007; Tohge *et al.*, 2007; Farag *et al.*, 2008; Saito *et al.*, 2008; Yamazaki *et al.*, 2008). Important questions that still remain largely unanswered for most plant metabolites are: how are their pathways structured and organized, what controls these pathways, and are there higher order organizations to these pathways or within these pathways that can be understood and then used to predict how they function to produce specific molecules?

Based on the concept of biosynthesis/biosynthetic modules put forward by Reiko Tanaka and John Doyle (Tanaka, 2005; Tanaka *et al.*, 2005), on the suggestion of hierarchical modularity of metabolic pathways in data presented by Tikunov *et al.* (2005), and on recent work in our laboratory related to the control of production of different classes of compounds in specific cell types (Xie *et al.*, 2008), we hypothesized that many compounds produced by complex biological networks or a series of parallel metabolic pathways could be produced and may be detectable in biological systems in what we call 'metabolite modules'. Such metabolite modules would consist of groups of metabolites whose production and further metabolism would be co-regulated under a series of defined conditions in the organism. One benefit that the existence of such metabolite modules present to plant metabolism investigations would be that identification of one compound within such a module would allow for the rapid identification of other members of the module, because they would be biosynthetically and structurally linked. When it is considered that only around 4–8% of all plants have been investigated in any detail for the metabolites that they produce (422 000 plant species estimated, 35 000 species tested for anti-cancer activity by NCI, 15 254 registered in the KNApSAck database), having such a tool in hand could lead to great strides in our understanding, not only of what compounds plants produce but also of how such compounds are produced and how their production is regulated. It has been known for quite some time, for example, that the activity of enzymes such as HMG-CoA reductase (HMGR) and phenylalanine ammonia lyase (PAL) influence the rates of production of a large number and a wide variety of downstream compounds (Camm and Towers, 1973; Stermer *et al.*, 1994; Fukasawa-Akada *et al.*, 1996; Britton *et al.*, 1998; Weisshaar and Jenkins, 1998; Harker *et al.*, 2003; Winkel, 2004). It could be argued that these 'key' enzymes regulate large metabolite modules that represent entire biosynthetic pathways. However, they are not the only components in the pathways that contribute to metabolic flux control and compound production rates, and the determination of sub-groups of compounds that follow alternative production profiles can be used to predict additional organizational structures of the metabolic networks in question. This will be demonstrated below.

Due to the complex nature apparent in the metabolism of members of the Zingiberaceae, we thought that turmeric (*Curcuma longa* L.), which is of great general interest due to its important medicinal properties (Arora *et al.*, 1971; Reddy and Lokesh, 1992; Jayaprakasha *et al.*, 2005; Sharma *et al.*, 2005; Shishodia *et al.*, 2005; Xia *et al.*, 2005), would represent an ideal organism with which to test this hypothesis, to see if such metabolite modules could be easily detected and if so to see if their presence and organization could suggest anything about the biosynthesis of metabolites in plants. The most characteristic and abundant compounds in turmeric rhizomes are the non-volatile curcuminoids (curcumin **1**, demethoxycurcumin **2**, and bisdemethoxycurcumin **3**) (Srinivasan, 1952, 1953; Kosuge *et al.*, 1985; He *et al.*, 1998; Ma and Gang, 2006; Pothitirat and Gritsanapan, 2006; Tayyem *et al.*, 2006; Jagetia and Aggarwal, 2007), belonging to the larger class of compounds called diarylheptanoids. Several other diarylheptanoids have also been detected and identified from turmeric as more minor constituents (Masuda *et al.*, 1993; Nakayama *et al.*, 1993; Park and Kim, 2002; Jiang *et al.*, 2006*b*, *c*; Ma and Gang, 2006). By contrast, the volatile oils of turmeric rhizomes contain sesquiterpenoids, monoterpenoids, and fatty acids (Jayaprakasha *et al.*, 2005).

Labelling studies and enzyme assays have suggested that diarylheptanoids, such as curcumin, are formed from a one-carbon unit and two phenylpropanoids, with the one-carbon unit being derived from malonate (Holscher and Schneider, 1995; Kamo *et al.*, 2000; Brand *et al.*, 2006; Ramirez-Ahumada *et al.*, 2006), suggesting the action of polyketide synthases or similar enzymes in the biosynthesis of the backbone structure of these compounds. Based on this, we proposed a putative biosynthetic pathway for curcuminoids in turmeric (Ramirez-Ahumada *et al.*, 2006), which has been modified as a result of the data presented here (Fig. 1). The activities of some of the important enzymes in the proposed pathway, such as phenylalanine ammonia lyase (PAL), *p*-coumaroyl-CoA:*p*-coumaroyl-5-*O*-shikimate transferase (CST), curcuminoid synthase (a polyketide synthase), and hydroxycinnamoyl-CoA thioesterase, have been identified from turmeric (Ramirez-Ahumada *et al.*, 2006). However, it was not clear when the 3-methoxyl groups on the aromatic rings are formed, whether before or after the formation of the diarylheptanoid backbone (Ramirez-Ahumada *et al.*, 2006).

In this report, we show that metabolite modules do exist in turmeric rhizomes, supporting the hypothesis that biosynthetic modules do indeed exist in natural plant systems. Several of these metabolite modules in turmeric rhizomes contain specific groups of diarylheptanoids, including one module that contains the three major curcuminoids and a separate module that contains those diarylheptanoids that would be intermediates in the pathway to curcumin if the methoxyl groups were to be added after the action of the polyketide synthase(s). The presence of these compounds in separate metabolite modules, however, suggests that these compounds are not directly biosynthetically linked and supports the hypotheses that the methoxyl groups are indeed added prior to diarylheptanoid backbone formation and that several different polyketide synthases are involved

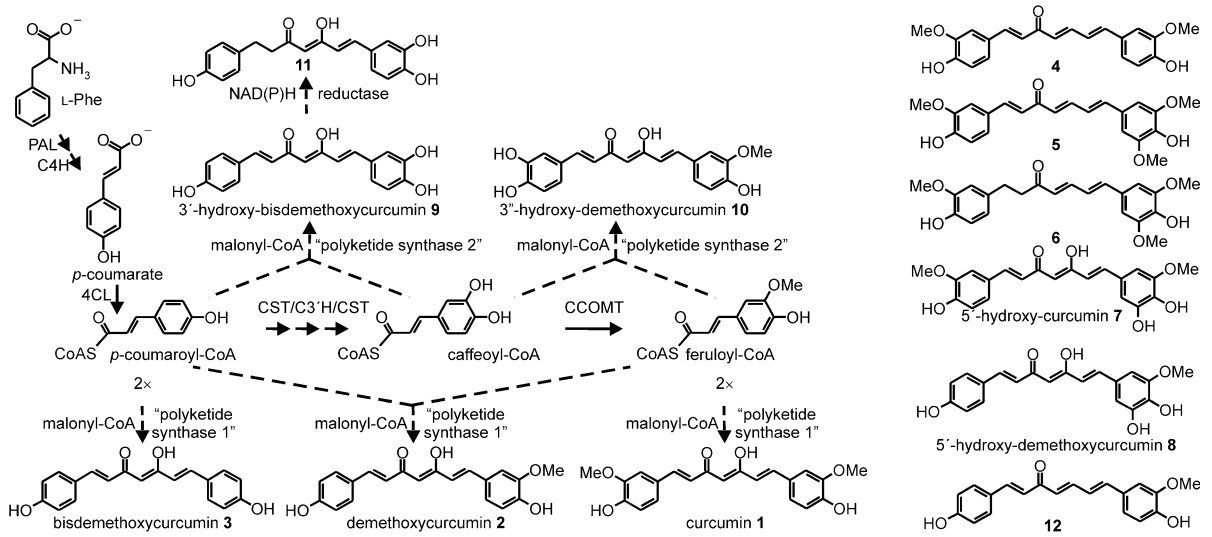

Fig. 1. Proposed biosynthetic pathway to selected diarylheptanoids in the turmeric rhizome. Solid and dashed arrows are for established and proposed conversions, respectively. Note that compounds **9** and **10** are not proposed to be intermediates in the biosynthesis of curcumin **1** because they belong to a different metabolite module. Compounds derived from this pathway, but which would require several additional steps are shown to the right. Structures of the diarylheptanoids are drawn in keto-enol tautomer form, which is how they would exist in solution (Jiang *et al.*, 2006*a*), although they are typically named after their β-diketide tautomeric forms.

in the production of the large array of diarylheptanoids that are produced in turmeric.

Materials and methods

Acetonitrile and methanol (B&J ACS/HPLC certificated solvent) were purchased from Burdick and Jackson (Muskegon, MI). Methyl *t*-butyl ether (MTBE, High Purity Solvent) was purchased from EMD Chemicals Inc (Gibbstown, NJ). Authentic standards of curcumin, demethoxycurcumin, and bisdemethoxycurcumin were purchased from ChromaDex, Inc. (Santa Ana, CA).

Plant material

Turmeric plants were grown in a single greenhouse under conditions described previously (Ma and Gang, 2005, 2006; Jiang *et al.*, 2006*c*). Four types of fertilizer treatments were applied to plants from two turmeric cultivars (TMO and HRT). Fresh rhizome samples were collected 5 months and 7 months after planting, and were immediately frozen in liquid nitrogen after harvest. The frozen samples were stored in –80 °C until analyzed.

Sample preparation

Frozen rhizome samples were ground to a fine powder in a mortar and pestle under N_2(l). Exactly 4.0 g of the rhizome powder were transferred to a 20 ml glass vial sealed with a cap lined with a Teflon septum and extracted three times sequentially with 16 ml MeOH by shaking (200 rpm, orbital shaker) at room temperature overnight. The MeOH extractions were centrifuged in the 20 ml vials at 2060 *g* for 30 min. The supernatants from the three extractions per sample were combined and dried under nitrogen gas. The dry extracts were resuspended in 20 ml of LC-MS grade MeOH. 100 µl of the suspension was diluted with 1.9 ml of LC-MS grade MeOH, filtered through 0.2 µm PTFE membranes, and stored at –20 °C until analyzed using LC-PDA. The rest of each suspension was dried under nitrogen gas and resuspended in 2 ml of MeOH. The suspensions were centrifuged at 2060 *g* for 30 min, and the supernatants were filtered through 0.2 µm PTFE membranes, and stored at –20 °C until analyzed using LC-MS and LC-MS/MS. Two grams of the rhizome powder were extracted with 4 ml MTBE overnight with shaking at room temperature. The MTBE extracts were filtered through 0.2 µm PTFE membranes, and stored at –20 °C until analyzed using GC-MS.

GC-MS analysis

450 µl of the filtered MTBE extracts of turmeric rhizomes were mixed with 50 µl of internal standard solution (*p*-chlorotoluene in MTBE, 0.1 mg ml^{-1}) and then analyzed by GC-MS as previously described (Ma and Gang, 2005, 2006; Jiang *et al.*, 2006*c*). Before data processing, all data files were exported to NetCDF format using the file converter in Xcalibur (Version 1.4, Thermo Electron). A target spectral library with retention time information was built up in AMDIS (version 2.65) based on compound identification using NIST Mass Spectral library Version 2.0 (NIST/EPA/NIH, USA) and an essential oil GC-MS mass spectra library from Dr. Robert P. Adams (Adams, 2004), as well as by referral to the literature (Jolad *et al.*, 2004; Jiang *et al.*, 2006*c*; Ma and Gang, 2006). The parameters in AMDIS were: (i) Deconv.: component width, 32; resolution, low; shape

requirement, low; (ii) Identif.: use retention time; (iii) Instr: scan direction, low to high; (iv) Other: default. A compound was considered identified only when the match score of its spectrum was larger than 800. Compounds failing to meet this criterion were considered unidentified and code names were assigned according to standard metabolite profiling nomenclature rules (Bino *et al.*, 2004).

Quantitative analysis of the GC-MS results was performed using MET-IDEA (version 1.2.0). An ion-retention time list was generated using AMDIS and then manually processed to exclude redundant peaks (R^2 >0.8 and ΔRt <0.2 min) and unreliable peaks (Rt <5 min; Rt >42 min; or peak purity <50%) after the first round of MET-IDEA analysis. The refined ion-retention time list was used for a second round of MET-IDEA analysis to collect peak area information. The parameters for MET-IDEA were: (i) chromatography: GC; average peak width, 0.1; minimum peak width, 0.3; maximum peak width, 6; peak start/stop slope, 1.5; adjusted retention time accuracy, 0.95; peak overload factor, 0.3; (ii) mass spec: quadrupole; mass accuracy, 0.1; mass range, 0.5; (iii) AMDIS: exclude ion list, 73, 147, 281, 341, 415; lower mass limit, 50; ions per component, 1. The peaks of internal standard *p*-chlorotoluene were used for retention time calibration.

LC-MS and LC-MS/MS analysis

75 μl of the concentrated MeOH extracts of turmeric rhizomes were mixed with 75 μl of internal standard solution (6-benzylaminopurine in MeOH, 0.25 mg ml^{-1}) and 5 μl of these mixtures were analyzed by LC-MS using a Thermo-Electron Surveyor MS HPLC coupled to a ThermoElectron LCQ Advantage ion trap and an in-line PDA detector (San Jose, CA, USA) as previously described (Ma and Gang, 2005, 2006; Jiang *et al.*, 2006*c*).

Representative samples were selected for analysis using LC-MS/MS for compound identification using the same extraction and solvent conditions, except that no internal standard was added. Both positive and negative modes were performed under collision gas pressure, *c.* 10^{-5} torr. Mass ranges for positive mode were: 100–307; 282–450; 312–337; 342–365; 370–450; 440–630; 620–820; 810–1000. Mass ranges for negative mode were: 100–304; 280–450; 310–335; 340–365; 370–450; 440–630; 620–820; 810–1000. Data dependent scanning was used to acquire the MS/MS spectra of the top 1–3 and 3–5 most abundant ions in a precursor ion scan at each of the multiple mass scan ranges in both positive and negative mode. Therefore, four files were generated for each mass range.

Diarylheptanoids in the rhizome sample were identified based on their MS/MS spectra and fragmentation rules reported previously (Jiang *et al.*, 2006*a*, *b*). Quantitative analysis of LC-MS was performed using an R package, xcms (version 1.6.1) with the following parameters: snthresh=6, fwhm=18, bw=10, minfrac=0.4, and span=0.5. The results of xcms were manually processed to eliminate isotopic peaks (0.5 <ΔM <1.5, ΔRt <18 s) and unreliable peaks (Rt <600 s or Rt >3300 s).

Data analysis

Hierarchical cluster analysis (HCA) and the creation of heatmaps of data from non-targeted analysis (LC-MS and GC-MS) were performed using two R packages, Heatplus, and gplots. All data were autoscaled. Pearson's correlation coefficients, which represent the similarity of the abundance patterns of compounds in the rhizome samples, were calculated for all compound pair-wise comparisons within the analysis type (LC-MS or GC-MS). Two-way HCA analysis of correlation coefficients was carried out separately for LC-MS and GC-MS data using Euclidean distance and Ward's method (Ward, 1963). The data were then sorted according to cluster membership. Using the sorted data, correlation heatmaps were generated. Correlation heatmaps were created using the 'bluered' color scheme in the 'gplots' package.

Results and discussion

To determine whether metabolite modules exist and are readily detected in plants, and to evaluate the utility of using metabolite modules to investigate plant metabolism if they do exist, the metabolite content of rhizomes obtained from turmeric plants that had been subjected to 16 different growth and development treatments was analyzed. This produced a dataset with the complexity required to test for the presence of metabolite modules. In these experiments, the composition and levels of metabolites of rhizome samples that were collected at two different developmental stages from two different turmeric varieties that were grown under four different fertilizer treatment regimes were compared. Both volatile and non-volatile compounds were analyzed using GC-MS and LC-MS^n. Correlations between product ion profiles of all compound pairings were then determined and used to derive metabolite modules.

Production of metabolic profiles and identification of metabolite modules

Combined metabolic profiles were produced for all samples in this investigation, where a total of 136 and 113 compounds were detected, respectively, in LC-MS and GC-MS analyses. A typical LC-MS result for turmeric rhizome samples is shown in Fig. 2A, where the majority of the detected peaks formed four clusters in the 3D chromatograms based on *m/z* ratio, elution time, and peak intensity. All of the diarylheptanoids identified, including the three major curcuminoids, are located in area 1 of the 3D chromatograms. However, most of the peaks in the LC-MS results represent unidentified metabolites. Because these compounds were detected in negative ionization mode in the electrospray source under acidic conditions (pH of the mobile phase ~3.3), most of these compounds probably contain carboxyl, phenolic hydroxyl or other readily ionizable groups. However, a carboxyl group typically affords a neutral loss of 44 (CO_2) in MS/MS analysis (Bandu *et al.*, 2004; Zeng *et al.*, 2006), which was not frequently observed in our MS/MS results. Therefore, many of these unknown

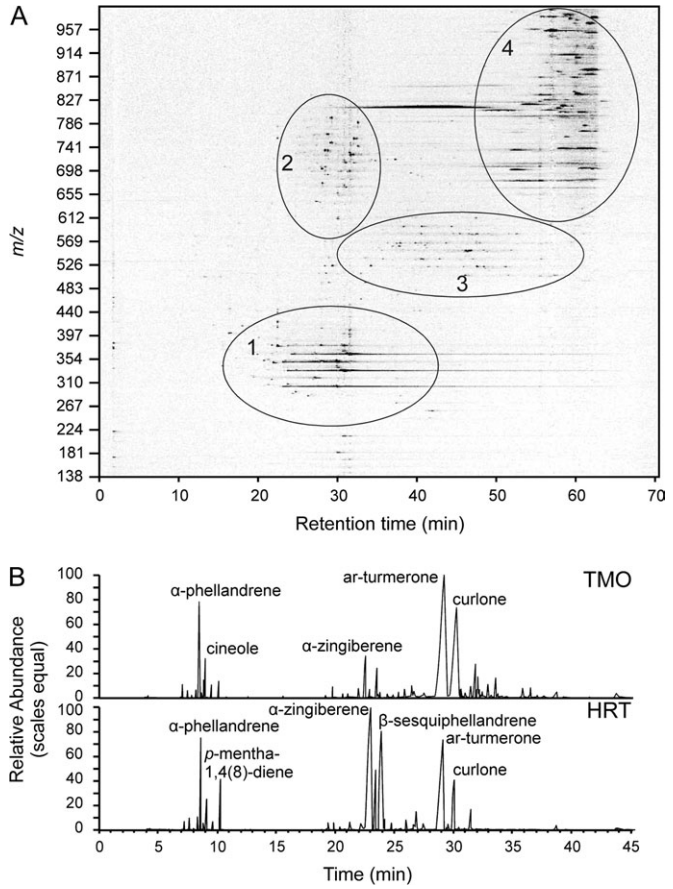

Fig. 2. Typical chromatograms from LC-MS (A) and GC-MS (B) analysis of turmeric rhizome extracts. As can be seen in (A), most of the compounds identified in the LC-MS analysis belong to four major clusters, as determined by elution time and mass range.

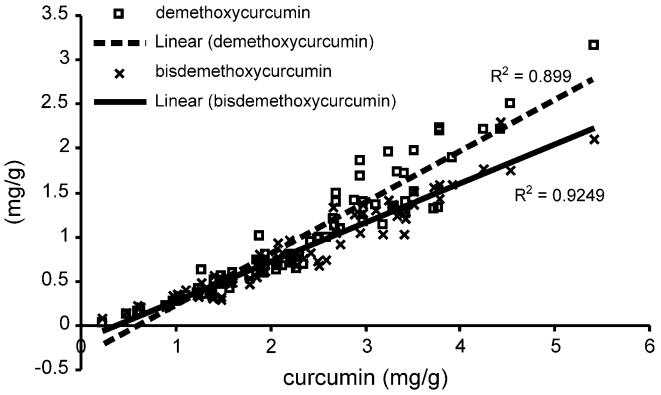

Fig. 3. The three major curcuminoids show strong correlation in production profiles across the 16 different growth and development treatments.

compounds are likely to contain phenolic functional groups, and therefore may be structurally and perhaps biosynthetically related. Typical GC-MS results for the two turmeric lines are shown in Fig. 2B, where most identified compounds were mono- and sesquiterpenoids, although many other compounds, such as eugenol, were also detected. Most of the unidentified compounds also appeared to be terpenoids, based on mass spectra features, but they could not be conclusively identified because the resulting spectra were not found in the GC-MS spectral databases.

The content of the three major curcuminoids (curcumin **1**, demethoxycurcumin **2**, and *bis*demethoxycurcumin **3**) was determined by LC-PDA-MS analysis. Interestingly, the levels of the three curcuminoids displayed similar production profiles across the different treatment combinations, where a strong linear correlation (R >0.94) was observed (Fig. 3), suggesting that the accumulation and biosynthesis of the three major curcuminoids are closely associated with each other, i.e. that they formed a metabolite module. This is an expected result, based on previous research (Ramirez-Ahumada *et al.*, 2006).

Based on these results, it was reasonable to propose that similar correlations may exist in the production profiles of other groups of compounds, including other diarylheptanoids. To test this hypothesis, HCA analysis was performed with Pearson correlation coefficients calculated for all pairs of metabolites identified in our analysis, using normalized compound levels in determining the correlation coefficients. The HCA results and 'correlation heatmaps' clearly show the existence of modules of co-regulated metabolites in both the LC-MS and GC-MS data sets (Fig. 4A, B). Metabolites within the same module had abundance patterns across treatments that were highly correlated with each other, and they had similar relationships to other compounds. For example, almost all of the compounds in module 3 in the LC-MS correlation heatmap had a negative correlation with compounds in modules 2 and 7, but positive correlations with compounds in module 4 (Fig. 4A). These metabolite modules also appeared to be hierarchical, i.e. large modules contained smaller sub-modules in which compounds were more closely associated with each other (e.g. modules 7, 8, and 10 in the LC-MS dataset, modules 11 and 11-1 in the GC-MS dataset).

Similar co-regulated metabolite modules can be detected in other plants, such as in a GC-MS data set from tomato fruit (Tikunov *et al.*, 2005), and our results showed that metabolite modules are readily determined in both LC-MS and GC-MS data sets from turmeric rhizomes. Therefore, modules of co-regulated metabolites may be a universal feature in plant metabolism.

Use of metabolite modules in biosynthetic pathway prediction

Compounds in a metabolite module can be expected to be structurally and biosynthetically related to each other (Tikunov *et al.*, 2005). For many of the apparent metabolite modules detected in our LC-MS data sets, this was observed to be the case. Series of possible compound analogues were identified, which differed by mass shifts that represent common biosynthetic modifications such as reduction (+2), dehydrogenation (–2), oxidation (+16), hydration (+18), methylation (+14), and methoxylation (+30), among others (Fig. 4A). Therefore, identification of co-regulated metabolite modules can provide valuable information for the

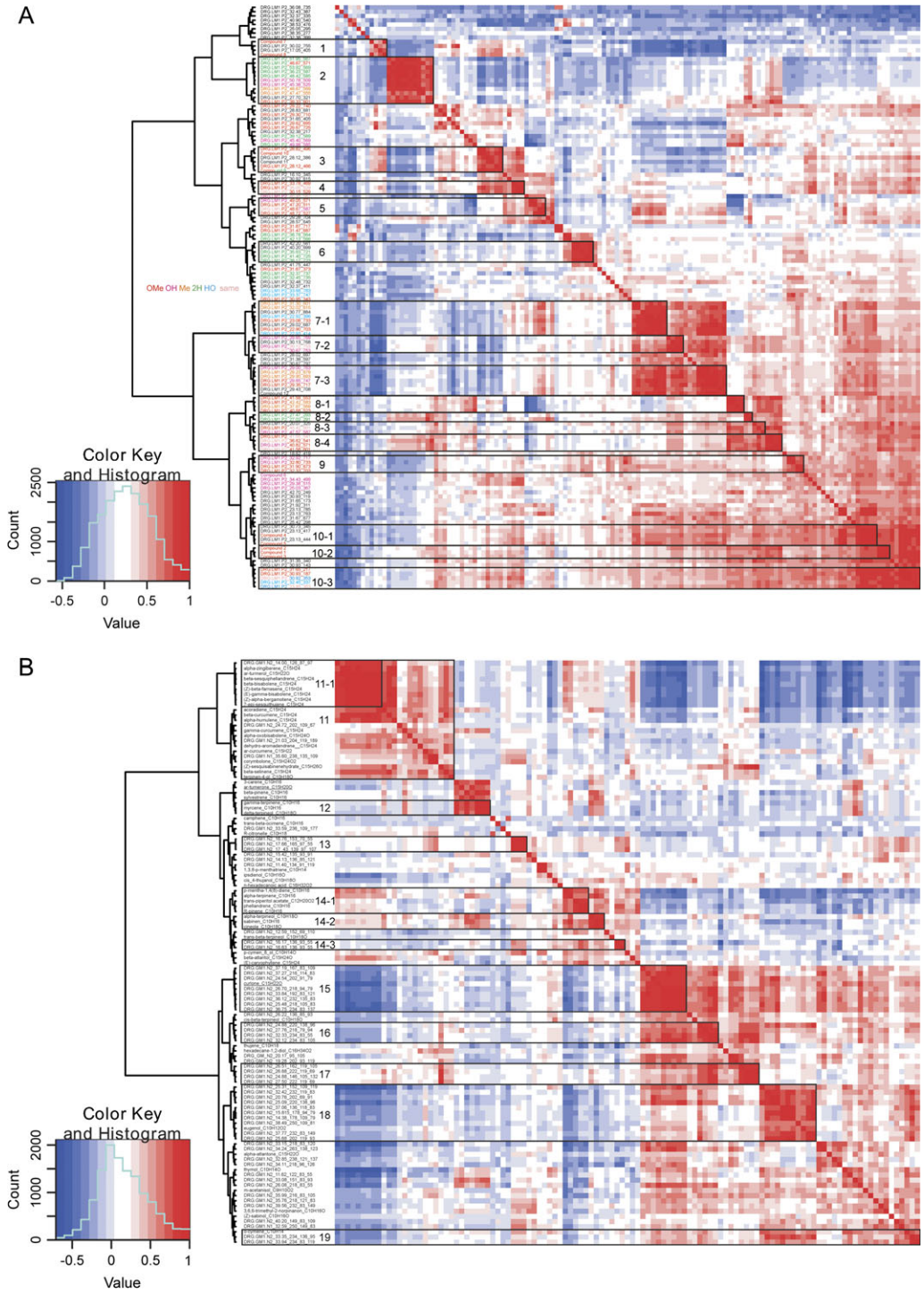

Fig. 4. Hierarchical cluster analysis of LC-MS (A) and GC-MS (B) metabolite correlation results reveal the presence of metabolite modules in turmeric rhizomes. Obvious modules are outlined and numbered. Coloring in the names of compounds in the LC-MS results indicate mass differences related to the common biosynthetic conversions between compounds within the same module or HCA cluster, as indicated near the left of the figure. The color keys and histograms show the magnitude and distribution of the correlation coefficients between pair-wise compound comparisons.

elucidation of metabolic pathways. An example of how these relationships can predict biosynthetic relationships is illustrated in Fig. 5, which shows how all but one compound in module 2 appear to be biosynthetically related. Unfortunately, the identity of any of these compounds is not yet known. However, identification of one of these compounds should allow us to identify the rest of the compounds in this module.

An excellent example of using metabolite modules to predict biosynthetic relationships can be found in the diarylheptanoid class of compounds from turmeric rhizomes. Twelve diarylheptanoids were readily detected, identified

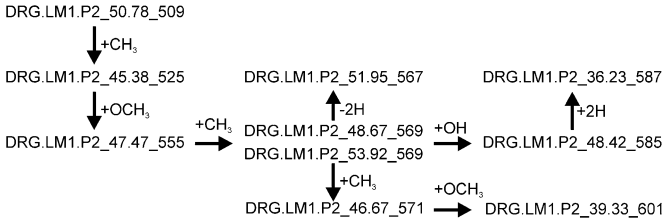

Fig. 5. Predicted biosynthetic relationships between compounds belonging to LC-MS module 2. Unidentified compounds are named according to metabolomics convention, with molecular ion mass given as the last three digits of each name, allowing for quick comparison of mass differences, which suggest biosynthetic conversions indicated over arrows.

(see Supplementary data at *JXB* online), and quantified from all samples used in this analysis. Their production profiles across 16 different growth treatments are shown in Fig. 6. In addition to the three major curcuminoids (compounds **1**–**3**), which group together in module 10-2 of Fig. 4A, the other diarylheptanoids clustered in distinct metabolite modules in the LC-MS analysis, as is clear in the following examples. First, the production of compounds **4** (1,7-*bis*(4-hydroxy-3-methoxyphenyl)-1,4,6-heptatrien-3-one) and **5** (1-(4-hydroxy-3-methoxyphenyl)-7-(4-hydroxy-3,5-dimethoxyphenyl)-1,4,6-heptatrien-3-one**)** was found to be highly co-ordinated (Fig. 6B), and was also correlated with, yet was distinct from, the production of the three major curcuminoids. Compound **4** differs from **5** by 30 Da (extra methoxyl group), and from **1** by loss of 16 Da (loss of hydroxyl group on heptanoid chain). Compound **6** (1-(4-hydroxy-3-methoxyphenyl)-7-(4-hydroxy-3,5-dimethoxyphenyl)-4,6-heptadien-3-one; Fig. 6C) does not belong to a distinct module and differs from compound **5** by 2 Da, suggesting that the production of this compound is controlled by the enzyme responsible for the reduction of the double bond between carbons 6 and 7 of the heptanoid chain. It is yet to be determined whether this putative reductase acts prior to or after the action of the polyketide synthase the forms the general backbone structure.

The second example includes compounds **7** (5′-hydroxy-curcumin) and **8** (5′-hydroxy-demethoxycurcumin), which again differ by a methoxyl group and belong to module 1. These two compounds differ from two of the major curcuminoids (compounds **1** and **2**) by the addition of a hydroxyl group in the aromatic ring that possesses a methoxyl group. In other words, it appears that these compounds may be derived from 5-hydroxy-feruloyl-CoA and either feruloyl-CoA or *p*-coumaroyl-CoA, although it is possible that the hydroxyl group is added after the formation of the diarylheptanoid backbone. As is clear from Fig. 6D and 6A, the production of compounds **7** and **8** is completely unrelated to the production of compounds **1** and **2**. It appears that either the enzyme that adds the hydroxyl group (exactly when this happens is yet to be determined) or a PKS (not curcuminoid synthase) that can utilize 5-hydroxy-feruloyl-CoA may be the control point for production of these compounds. Thus, these metabolite module data provide us with clear hypotheses regarding the potential biosynthetic steps to test for.

A third example from the LC-MS analysis involves compounds **9** (3′-hydroxy-*bis*demethoxycurcumin)**, 10** (3′-hydroxydemethoxycurcumin**)**, and **11** (3′-hydroxy-6,7-dihydro-*bis*demethoxycurcumin**)**, which belong to yet another metabolite module (no. 3) separate from module (no. 10) that contains the curcuminoids (Fig. 4A). Based on their structures, **9** and **10** could be potential intermediates in the pathway to curcumin (Fig. 1) and differ by a methoxyl group. However, they do not belong to the metabolite module (or even a closely affiliated one) that contains the three major curcuminoids (Fig. 4A), and instead compound **11**, a hydrogenated derivative of **9** that lacks one of the double bonds of the heptanoid chain, clusters with these two compounds in the same metabolite module. This suggests that compounds **9** and **10** are not intermediates in the pathway to curcumin, and instead reside on a separate branch of the diarylheptanoid biosynthetic network in turmeric (Fig. 1) that contains a molecule with a caffeoyl moiety as an important intermediate. The similar production profiles for these three compounds (Fig. 6E, A), as opposed to the very different production profiles for compounds **1**, **2**, and **3**, supports these conclusions. Furthermore, these observations suggest that the 3-hydroxyl and 3-methoxyl groups of the diarylheptanoids are added to the aromatic rings *prior to* the formation of the heptanoid backbone chain (Fig. 1), which can then be further modified to form additional classes of diarylheptanoids, such as compounds **9**, **10**, and **11**, or the three major curcuminoids, **1**, **2**, and **3**. The common precursor of **9**, **10**, and **11** is likely to be caffeoyl-CoA.

Thus, these data suggest that multiple PKS-like enzymes, with different substrate selectivities, appear to be responsible for the formation of these different groups of diarylheptanoids. One PKS presumably uses caffeoyl-CoA and catalyses the formation of **9** and **10**, the latter of which could be then converted into **11** by a dehydrogenase. Alternatively, **11** could be produced from 7,8-dihydrocaffeoyl-CoA by the same polyketide synthase, similar to what may occur in the production of compound **6**, see above. A second PKS does not use caffeoyl-CoA as a substrate, but utilizes 5-hydroxy-feruloyl-CoA instead, forming compounds **7** and **8**. And a third PKS (curcuminoid synthase) cannot use either of these CoA esters as substrate and catalyses the formation of the three major curcuminoids (Ramirez-Ahumada *et al.*, 2006). These three groups of compounds formed distinctive co-regulated metabolite modules probably due to the differential regulation of the enzymatic activities of these different polyketide synthases.

Similar results are seen from the GC-MS analysis, where metabolite modules containing different groups of terpenoids can be identified. As can readily be seen in Fig. 4B, module 11 contains a large set of mostly sesquiterpenoids. Compounds in sub-module 11-1 are highly correlated with each other (Fig. 6F) and are also mostly very similar to each other structurally (Fig. 7), except for *Z*-α-bergamotene **18** and *E*-β-farnesene **19** which may be TPS derailment products, suggesting a common biosynthetic origin beyond the

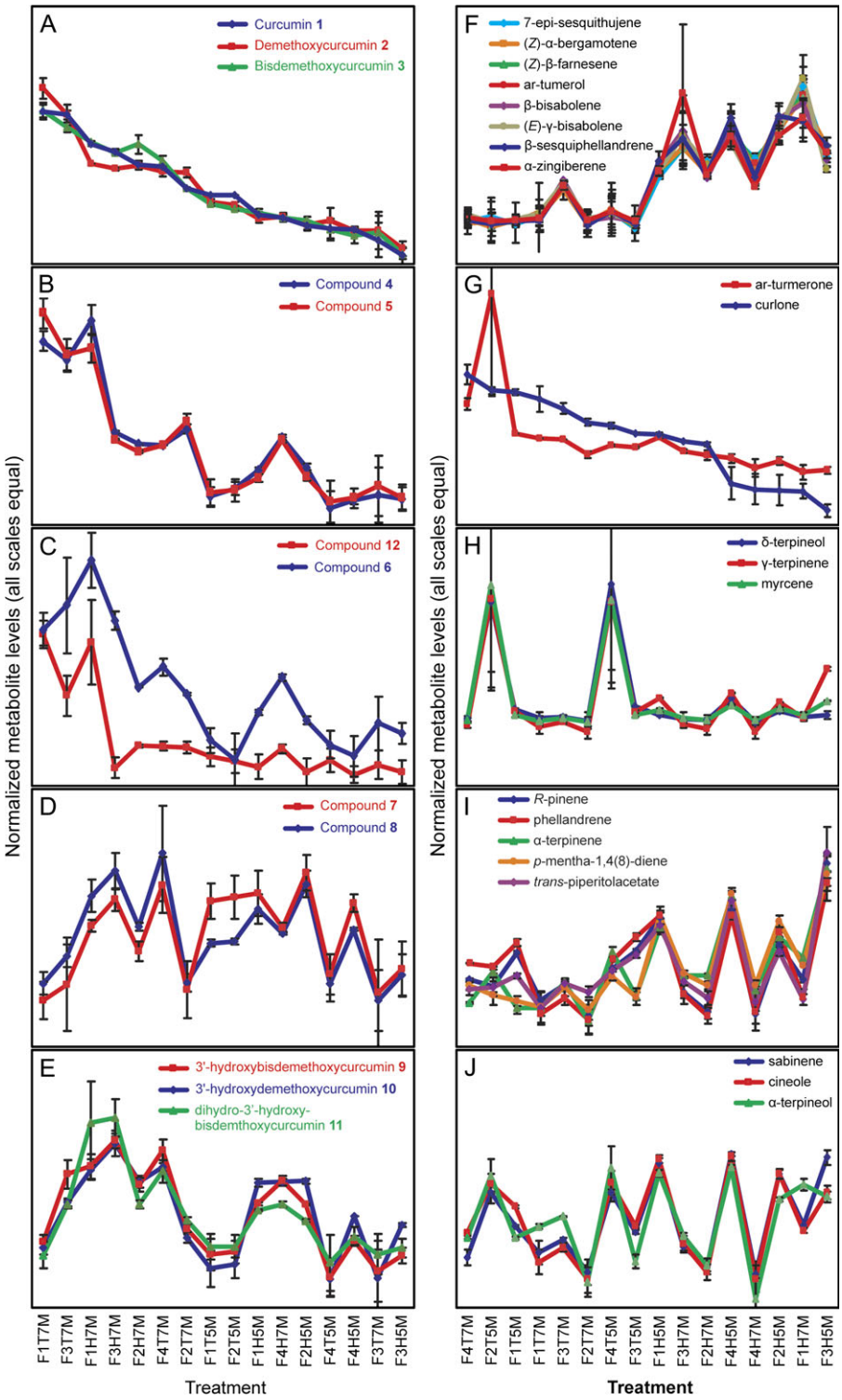

Fig. 6. Production profiles of groups of closely related diarylheptanoids (A–E) and terpenoids (F–J) support the concept of metabolite modules in specialized metabolism. Compounds were quantified by LC-MS (A–E) and GC-MS (F–J) analyses, with treatments ordered arbitrarily according to decreasing average levels of curcumin **1** and curlone **22**, respectively. Error bars are standard error from the mean (n=5–6).

availability of the common precursor farnesyl-diphosphate. These results suggest that a common terpene synthase or group of co-ordinately regulated terpene synthases is responsible for the formation of these compounds. Interestingly, the compounds ar-turmerone **21** and curlone **22** (Fig. 7) are also similar to these compounds in structure, yet they are, in fact, strongly negatively correlated with them (Fig. 4B, module 15 and just above module 12; Fig. 6G). This

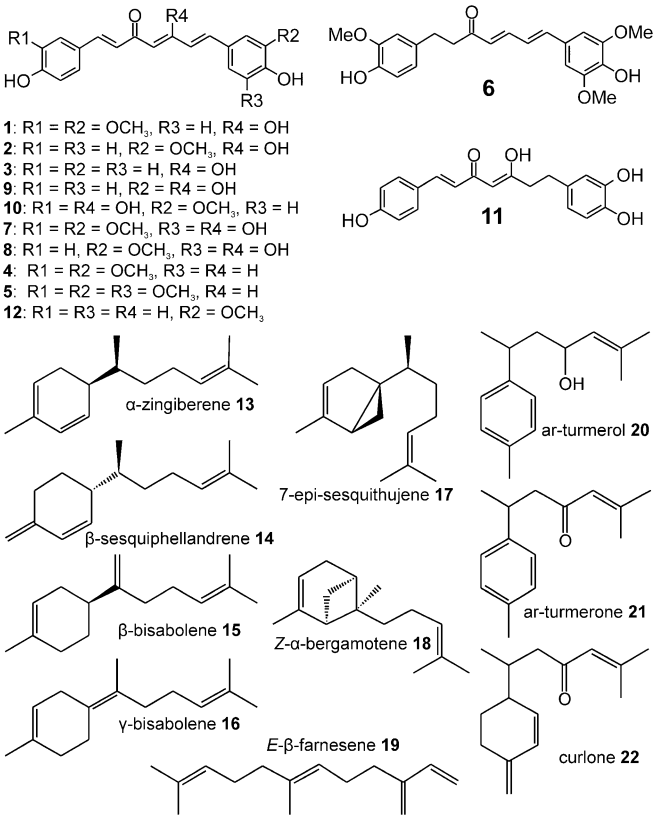

Fig. 7. Structures of curcuminoids (**1**–**3**) and other diarylheptanoids (**4**–**12**) belonging to metabolite modules shown in Fig. 4A, of sesquiterpenoids (**13**–**20**) belonging to module 11-1 of Fig. 4B, and of ar-turmerone **21** and curlone **22**.

suggests that enzymes distinct from the TPS that forms the compounds in module 11 are responsible for regulating the formation of these compounds, be it other TPSs or an oxidase that forms the ketone functional group of these molecules. Similarly, three distinct modules were easily identified that contained groups of monoterpenoids (Fig. 4B, modules 12, 14-1, and 14-2) whose production profiles were very similar within the module (Fig. 6H, I, J, respectively). Interestingly, the production of all monoterpenoids in module 14-1 (see also Fig. 6I) is very tightly co-ordinated in turmeric cultivar HRT (the eight treatments on the right half of the panel are from this line), but is less co-ordinated in cultivar TMO (the eight treatments on the left half of the panel). This suggests that the production of these monoterpenoids is regulated differently in these two lines. This could be due to one TPS enzyme being responsible for the formation of all of these compounds in line HRT, whereas two or more enzymes would be involved in the production of these compounds in line TMO. Alternatively, multiple TPS enzymes could be involved in the production of these compounds in both cultivars, but these enzymes would be co-ordinately expressed in HRT but not in TMO. Thus, metabolite modules in GC-MS data lead to similar interesting conclusions as observed for metabolite modules in LC-MS data regarding the biosynthesis of specific compounds.

Conclusions

Metabolite modules may be a universal feature of plant specialized metabolism. Detection of these modules is useful for both compound identification and biosynthetic investigation. The existence of metabolite modules may provide evidence for the presence not only of co-ordinated gene expression being involved in the production of groups of compounds in plant cells but also the presence of metabolons, where suites of proteins form large macromolecular complexes whose composition yields specific metabolite production outcomes, although this is yet to be tested.

Supplementary data

Supplementary data is available online for the identification of diarylheptanoids from turmeric rhizome extracts used in this analysis.

Fig. S1. Fragmentation rules for diarylheptanoids with a 1,6-heptadiene-3,5-dione bridge (Jiange *et al.*, 2006*a*).

Fig. S2. Special comparison of **15** and **11**. (A) Positive mode, (B) negative mode. (*: peaks with the same mass; +: peaks with the mass sifted by 16).

Fig. S3. (A) Spectral comparison of **17** and **20** (*: peaks with the same mass; +: peaks with the mass shifted by 28). (B) Spectral comparison of **4** and **5** (*: peaks with the same mass; +: peaks with the mass shifted by 30; the preferred structure candidate was marked with a dashed-line square).

Fig. S4. Structure and MS/MS spectra of **7** (a new compound), **8** (a new compound; the chromatographic peak of this compound was very close to the peak of **10**, which shares the same precursor ion mass, so the MS/MS spectra of **8** were contaminated by product ions from **10**.

Acknowledgements

The authors acknowledge financial assistance from the NSF Plant Genome Program (Grant DBI-0227618 to DRG). We also thank Mr Hyun Jo Koo and Ms Jessica Schmidt for assistance with extract production. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Science Foundation.

References

Adams RP. 2004. *Identification of essential oil components by gas chromatography/mass spectrometry*, 4th edn. Carol Stream, Illinois: Allured Publishing Corporation.

Arora RB, Kapoor V, Basu N, Jain AP. 1971. Anti-inflammatory studies on *Curcuma longa* (turmeric). *Indian Journal of Medical Research* **59,** 1289–1295.

Bandu ML, Watkins KR, Bretthauer ML, Moore CA, Desaire H. 2004. Prediction of MS/MS data. 1. A focus on pharmaceuticals containing carboxylic acids. *Analytical Chemistry* **76,** 1746–1753.

Bino RJ, Hall RD, Fiehn O, *et al*. 2004. Potential of metabolomics as a functional genomics tool. *Trends in Plant Science* **9,** 418–425.

Brand S, Holscher D, Schierhorn A, Svatos A, Schroder J, Schneider B. 2006. A type III polyketide synthase from *Wachendorfia thyrsiflora* and its role in diarylheptanoid and phenylphenalenone biosynthesis. *Planta* **224,** 413–428.

Britton G, Liaaen-Jenson S, Pfander H. 1998. *Carotenoids: biosynthesis and metabolism*. Basel: Birkhauser Verlag.

Camm EL, Towers GHN. 1973. Phenylalanine ammonia lyase. *Phytochemistry* **12,** 961–973.

Deavours BE, Liu CJ, Naoumkina MA, Tang YH, Farag MA, Sumner LW, Noel JP, Dixon RA. 2006. Functional analysis of members of the isoflavone and isoflavanone *O*-methyltransferase enzyme families from the model legume *Medicago truncatula*. *Plant Molecular Biology* **62,** 715–733.

Dixon RA, Achnine L, Deavours B, Farag M, Naoumkina M, Sumner LW. 2005. Integrated transcript and metabolite profiling for gene discovery in plant natural product pathways. *Journal of Biotechnology* **118,** S143–S143.

Farag MA, Huhman DV, Dixon RA, Sumner LW. 2008. Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures. *Plant Physiology* **146,** 387–402.

Fukasawa-Akada T, Kung SD, Watson JC. 1996. Phenylalanine ammonia-lyase gene structure, expression, and evolution in *Nicotiana*. *Plant Molecular Biology* **30,** 711–722.

Harker M, Hellyer A, Clayton JC, Duvoix A, Lanot A, Safford R. 2003. Co-ordinate regulation of sterol biosynthesis enzyme activity during accumulation of sterols in developing rape and tobacco seed. *Planta* **216,** 707–715.

He XG, Lin LZ, Lian LZ, Lindenmaier M. 1998. Liquid chromatography electrospray mass spectrometric analysis of curcuminoids and sesquiterpenoids in turmeric (*Curcuma longa*). *Journal of Chromatography A* **818,** 127–132.

Hirai M, Klein M, Fujikawa Y, *et al*. 2005*a*. Functional identification of unknown genes by integration of metabolomics and transcriptomics. *Plant and Cell Physiology* **46,** S61–S61.

Hirai MY, Klein M, Fujikawa Y, *et al*. 2005*b*. Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. *Journal of Biological Chemistry* **280,** 25590–25595.

Holscher D, Schneider B. 1995. A diarylheptanoid intermediate in the biosynthesis of phenylphenalenones in *Anigozanthos preissii*. *Journal of the Chemical Society-Chemical Communications* 525–526.

Jagetia GC, Aggarwal BB. 2007. 'Spicing up' of the immune system by curcumin. *Journal of Clinical Immunology* **27,** 19–35.

Jayaprakasha GK, Jagan L, Rao M, Sakariah KK. 2005. Chemistry and biological activities of *C. longa*. *Trends in Food Science and Technology* **16,** 533–548.

Jiang HL, Solyom AM, Gang DR. 2005. Characterization of gingerol-related compounds in ginger rhizome (*Zingiber officinale* Rosc.) by high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **19,** 2957–2964.

Jiang HL, Somogyi A, Jacobsen NE, Timmermann BN, Gang DR. 2006*a*. Analysis of curcuminoids by positive and negative electrospray ionization and tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **20,** 1001–1012.

Jiang HL, Timmermann BN, Gang DR. 2006*b*. Use of liquid chromatography-electrospray ionization tandem mass spectrometry to identify diarylheptanoids in turmeric (*Curcuma longa* L.) rhizome. *Journal of Chromatography A* **1111,** 21–31.

Jiang HL, Timmermann BN, Gang DR. 2007. Characterization and identification of diarylheptanoids in ginger (*Zingiber officinale* Rosc.) using high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **21,** 509–518.

Jiang HL, Xie ZZ, Koo HJ, McLaughlin SP, Timmermann BN, Gang DR. 2006*c*. Metabolic profiling and phylogenetic analysis of medicinal *Zingiber* species: tools for authentication of ginger (*Zingiber officinale* Rosc). *Phytochemistry* **67,** 1673–1685.

Jolad SD, Lantz RC, Solyom AM, Chen GJ, Bates RB, Timmermann BN. 2004. Fresh organically grown ginger (*Zingiber officinale*): composition and effects on LPS-induced PGE(2) production. *Phytochemistry* **65,** 1937–1954.

Kamo T, Hirai N, Tsuda M, Fujioka D, Ohigashi H. 2000. Changes in the content and biosynthesis of phytoalexins in banana fruit. *Bioscience Biotechnology and Biochemistry* **64,** 2089–2098.

Kosuge T, Ishida H, Yamazaki H. 1985. Studies on active substances in the herbs used for Oketsu ('stagnant blood') in Chinese medicine. III. On the anticoagulative principles in curcumae rhizoma. *Chemical and Pharmaceutical Bulletin* **33,** 1499–1502.

Kusano M, Fukushima A, Arita M, Jonsson P, Moritz T, Kobayashi M, Hayashi N, Tohge T, Saito K. 2007. Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in *Arabidopsis thaliana*. *BMC Systems Biology* **1,** 53.

Ma X-Q, Gang DR. 2005. Metabolic profiling of *in vitro* micropropagated and conventionally greenhouse-grown ginger (*Zingiber officinale*). *Phytochemistry* **67,** 2239–2255.

Ma XQ, Gang DR. 2006. Metabolic profiling of turmeric (*Curcuma longa* L.) plants derived from *in vitro* micropropagation and conventional greenhouse cultivation. *Journal of Agricultural and Food Chemistry* **54,** 9573–9583.

Masuda T, Jitoe A, Isobe J, Nakatani N, Yonemori S. 1993. Antioxidative and anti-inflammatory curcumin-related phenolics from rhizomes of *Curcuma domestica*. *Phytochemistry* **32,** 1557–1560.

Nakayama R, Tamura Y, Yamanaka H, Kikuzaki H, Nakatani N. 1993. Two curcuminoid pigments from *Curcuma domestica*. *Phytochemistry* **33,** 501–502.

Park SY, Kim D. 2002. Discovery of natural products from *Curcuma longa* that protect cells from beta-amyloid insult: a drug discovery effort against Alzheimer's disease. *Journal of Natural Products* **65,** 1227–1231.

Pothitirat W, Gritsanapan W. 2006. Variation of bioactive components in *Curcuma longa* in Thailand. *Current Science* **91,** 1397–1400.

Ramirez-Ahumada MD, Timmermann BN, Gang DR. 2006. Biosynthesis of curcuminoids and gingerols in turmeric (*Curcuma longa*) and ginger (*Zingiber officinale*): identification of curcuminoid synthase and hydroxycinnamoyl-CoA thioesterases. *Phytochemistry* **67,** 2017–2029.

Reddy ACP, Lokesh BR. 1992. Studies on spice principles as antioxidants in the inhibition of lipid-peroxidation of rat-liver microsomes. *Molecular and Cellular Biochemistry* **111,** 117–124.

Saito K, Hirai MY, Yonekura-Sakakibara K. 2008. Decoding genes with coexpression networks and metabolomics: 'majority report by precogs'. *Trends in Plant Science* **13,** 36–43.

Sawada Y, Saito K, Yokota-Hirai M. 2006. Integration of transcriptomics and metabolomics for identification of leucine and glucosinolate biosynthetic genes. *Plant and Cell Physiology* **47,** S50–S50.

Sharma RA, Gescher AJ, Steward WP. 2005. Curcumin: the story so far. *European Journal of Cancer* **41,** 1955–1968.

Shishodia S, Sethi G, Aggarwal BB. 2005. Curcumin: getting back to the roots. *Natural Products and Molecular Therapy* **1056,** 206–217.

Srinivasan KR. 1952. The coloring matter in turmeric. *Current Science* **21,** 311–312.

Srinivasan KR. 1953. Chromatographic study of the curcuminoids in *Curcuma longa*. *Journal of Pharmacy and Pharmacology* **5,** 448–457.

Stermer BA, Bianchini GM, Korth KL. 1994. Regulation of HMG-CoA reductase activity in plants. *Journal of Lipid Research* **35,** 1133–1140.

Tanaka R. 2005. Scale-rich metabolic networks. *Physical Review Letters* **94,** 168101–168104.

Tanaka R, Csete M, Doyle J. 2005. Highly optimised global organisation of metabolic networks. *IEE Proceedings: Systems Biology* **152,** 179–184.

Tayyem RF, Heath DD, Al-Delaimy WK, Rock CL. 2006. Curcumin content of turmeric and curry powders. *Nutrition and Cancer* **55,** 126–131.

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

Tohge T, Yonekura-Sakakibara K, Niida R, Watanabe-Takahashi A, Saito K. 2007. Phytochemical genomics in *Arabidopsis thaliana*: a case study for functional identification of flavonoid biosynthesis genes. *Pure and Applied Chemistry* **79,** 811–823.

Ward JH. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58,** 236–244.

Weisshaar B, Jenkins GI. 1998. Phenylpropanoid biosynthesis and its regulation. *Current Opinion in Plant Biology* **1,** 251–257.

Winkel BS. 2004. Metabolic channeling in plants. *Annual Review of Plant Biology* **55,** 85–107.

Xia Q, Zhao KJ, Huang ZG, Zhang P, Dong TTX, Li SP, Tsim KWK. 2005. Molecular genetic and chemical assessment of rhizoma curcumae in China. *Journal of Agricultural and Food Chemistry* **53,** 6019–6026.

Xie Z, Kapteyn J, Gang DR. 2008. A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. *The Plant Journal* **54,** 349–361.

Yamazaki M, Shibata M, Nishiyama Y, Springob K, Kitayama M, Shimada N, Aoki T, Ayabe SI, Saito K. 2008. Differential gene expression profiles of red and green forms of *Perilla frutescens* leading to comprehensive identification of anthocyanin biosynthetic genes. *FEBS Journal* **275,** 3494–3502.

Zeng GF, Xiao HB, Liu JX, Liang XM. 2006. Identification of phenolic constituents in Radix *Salvia miltiorrhizae* by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **20,** 499–506.