III.5 Metabolomics and Gene Identification in Plant Natural Product Pathways

R.A. DIXON, L. ACHNINE, B.E. DEAVOURS, and M. NAOUMKINA¹

1 Introduction

Collectively, plants produce more than 100,000 natural products (also known as secondary metabolites). The underlying genetic basis of their chemical elaboration appears at first to be dauntingly complex. However, the rich diversity of many chemical structures found in the plant kingdom arises from a number of chemical scaffolds (of many types in terpene biosynthesis, of a much more limited number in flavonoid biosynthesis) modified by a limited number of chemical substitution types (hydroxylation, glycosylation, acylation, prenylation, *O*-methylation, etc.) (Fig. 1). Much of the chemical diversity is brought about by the substrate- and/or regio-specificities of the substitution enzymes. Functional genomics of plant natural product pathways therefore centers in large part on identifying genes encoding the substitution enzymes that determine the chemical complexity of a given plant species.

This chapter highlights the problems of how to assign metabolic function to gene sequences that appear to encode enzymes of secondary metabolism. We argue that metabolomic analysis is an essential complement to “genomic” approaches for functional annotation of genes involved in plant natural product biosynthesis (see Fig. 2 for a summary of the concept and potential strategies). However, developments in this field have lagged far behind those for gene discovery per se.

2 Gene Discovery – Past and Present Strategies

The classical biochemical approach for characterization of metabolic pathway genes relied on assay-directed purification of the protein followed by protein sequence determination, synthesis of gene-specific oligonucleotides based on the protein sequence, and screening of cDNA or genomic libraries (or use of polymerase chain-reaction (PCR) approaches) to clone the corresponding gene. This approach has been very successful in the past (see Kutchan 2002 for an excellent summary in relation to alkaloid biosynthesis), but is somewhat laborious and “low throughput”. Manipulation of proteins is often more

¹ Plant Biology Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA, e-mail: radixon@noble.org

Fig. 1. Simplified outline of plant natural product biosynthesis, showing the interface between primary and secondary metabolism and the intermediacy of common scaffolds. In flavonoid biosynthesis, the great diversity of chemical structures depends largely on the modification and substitution of the scaffold, often catalyzed by the products of members of large gene families (an example is given on the right). Note that in terpenoid biosynthesis, much of the diversity arises from the formation of different scaffolds arising from differential cyclization products formed by different but closely related terpene cyclase enzymes, and further changes most often involve oxidation and reduction reactions

Source material:

Fig. 2. Strategies for the application of metabolomics to assign gene function in plant natural product biosynthesis

difficult, and less intuitive, than that of DNA, and low abundance membrane proteins with poor stability, such as some cytochrome P450s (Kochs and Grisebach 1986), pose particular problems. The advantage of the approach is that it leads directly and unequivocally to a gene encoding a protein of known catalytic activity. However, that catalytic activity has already been defined by the experimenter, and may not always reflect the *in vivo* activity of the enzyme. Furthermore, it is still necessary to confirm function of the gene by expression in *E. coli* or an alternative heterologous system such as yeast or insect cells.

The availability of large collections of cDNAs (e. g., expressed sequence tag [EST] libraries), which can be prepared in “Gateway” vectors for direct transfer, via site-specific recombination, into a variety of “destination vectors” for heterologous expression (Karimi et al. 2002), efficiently by-passes the need for initial protein purification if methods are available for selecting candidates for a particular catalytic activity. DNA sequence-based comparison and annotation of EST clones can give an overall list of genes potentially encoding enzymes of a specific class, although, as the annotation in the databases is based purely on sequence comparisons, it may sometimes be incorrect. For

example, some genes annotated as serine proteases are now known to be acyl-transferases involved in plant secondary metabolism (Li and Steffens 2000; Shirley et al. 2001), and genes annotated as encoding chalcone synthase (CHS, the polyketide synthase (Fig. 1) at the entry point of the flavonoid pathway) may encode related polyketide synthases (Tropf et al. 1994; Schröder 1997; Schröder et al. 1998).

Generally speaking, sequence-based functional predictions have to be tested by heterologous expression followed by enzyme assay. This becomes problematic when the gene is a member of a large gene family, such as a cytochrome P450 or glycosyltransferase (GT), and the experimenter still has to make the decision as to which substrates will be tested, a decision influenced by the often limited availability of potential substrates, or their instability in vitro. Furthermore, the substrate specificity of the recombinant enzyme may differ from that of the enzyme purified from the plant as a result of in planta post-translational modifications (Vogt 2004).

The most rigorous, unbiased approach to gene function is the analysis of knock-out lines in the gene of interest, possible for *Arabidopsis* where comprehensive collections of such lines exist covering almost the entire genome (Alonso et al. 2003), but this is not possible for most other species. In some cases, knock-outs or knock-downs in secondary metabolic pathways have clear biochemical and visible developmental phenotypes (Chapple et al. 1992; Franke et al. 2002). In other cases there may be a strong developmental phenotype but the biochemical basis for this may be unclear (Woo et al. 1999). Even in *Arabidopsis* with its small genome, genetic redundancy is a problem for functional identification. For example, knock-outs in several *Arabidopsis* GTs with ascribed in vitro activities fail to yield a clear phenotype (D.J. Bowles, personal communication). Lack of a discovered phenotype in knock-out lines may also reflect subtle biochemical changes in the mutant that escape targeted metabolite profiling.

3 Enzyme Promiscuity in Natural Product Pathways

In vitro biochemical analyses provide information on substrate preference and catalytic properties determined for the substrates selected by the experimenter. An enzyme with a relatively high K_m or low K_{cat}/K_m value may have more favorable kinetic constants for other substrates unknown at the time. Furthermore, the exact in vitro conditions chosen may, in some cases, have profound effects on relative substrate specificity (Lukacin et al. 2004). In cases where enzymes are promiscuous, a range of alternative in vivo substrates may exist. For example, developing strawberry fruits contain lignified achenes and vascular bundles, and an *O*-methyltransferase was cloned from the fruit with substrate specificity for ortho-diphenols including caffeic acid and caffeoyl CoA (Wein et al. 2002). It was thought likely that this enzyme is involved in

lignification in the fruit. However, the enzyme was also active with the vanillin precursor protocatechuic aldehyde, and could also methylate 2,5-dimethyl-4-hydroxy-3(2*H*)-furanone (DMHF). The latter two compounds are involved in flavor production in strawberry. In spite of the relatively low K_{cat}/K_m value for DMHF, it was concluded, from temporal and spatial examination of expression patterns, that the promiscuous OMT may play an important role in flavor production (Wein et al. 2002). Clearly, for any cell type within the fruit, knowing the presence, absence or relative levels of the potential OMT substrates would be instructive for assigning a biochemical function.

A recent example from mammalian cells nicely illustrates the problems of enzyme promiscuity, and how these may be addressed by a metabolomics approach. Untargeted LC/MS was used to analyze lipophilic compounds extracted from the brains and spinal cords of wild-type and transgenic mice in which the enzyme fatty acid amide hydrolase (FAAH) had been knocked-out (Saghatelian et al. 2004). Peaks seen in the knock-out but not in the wild type samples were in fact FAAH substrates. Surprisingly, the relative hydrolytic activity of FAAH shown for lipid metabolites *in vitro* was not necessarily indicative of the specific contribution of this enzyme *in vivo* (Saghatelian et al. 2004).

4 Examples of the Use of Metabolomics in the Elucidation of Gene Function

It will be clear from the above sections that ascribing gene function in secondary metabolism is not straightforward because activity *in vitro* may not always reflect activity *in vivo*. Particularly for promiscuous enzymes, knowledge of the cellular levels of all potential substrates may be critical for assigning *in vivo* function. Unfortunately, the depth of transcriptomic analyses (which can be close to genome wide for model species such as *Arabidopsis*, rice, *Medicago* and poplar) far exceeds that of metabolomic analyses at the present time. This is particularly true in natural product biosynthesis, where compounds tend to be identified on an “at need basis”, rather than globally.

4.1 The Isoflavonoid Pathway in *Medicago*

Isoflavonoids comprise a class of plant natural products with important biological activities including health promotion in humans and antimicrobial activity against plant pathogens (Dixon 1999, 2004; Dixon and Ferreira 2002). They are found primarily in leguminous plants, where they function as pre-formed or inducible antimicrobial or anti-insect compounds, as inducers of the nodulation genes of symbiotic *Rhizobium* bacteria, or as allelopathic agents (Dixon 1999). Isoflavonoids originate from a flavanone intermediate (either liquiritigenin or naringenin, Fig. 3a) that is ubiquitously present in plants. For entry

into the isoflavonoid pathway, flavanone undergoes migration of the B-ring to the 3-position followed by hydroxylation at the 2-position, catalyzed by the cytochrome P450 enzyme CYP93C1 (2-hydroxyisoflavanone synthase, commonly termed isoflavone synthase [IFS]). The resulting 2-hydroxyisoflavanone is dehydrated to the corresponding isoflavone (Kochs and Grisebach 1986) (Fig. 3a), which is then modified by substitution, reduction, ring cyclization and glycosylation to yield the range of isoflavone, isoflavanone and pterocarpan compounds illustrated in Fig. 3a.

Most, but not all, enzymes in the pathway to medicarpin are known (Dixon 1999). Several of these characterized enzymes are encoded by large multigene families; these include P450s such as IFS, OMTs and isoflavone reductases (Dixon et al. 2002). These genes were first discovered using classical biochemical approaches, such that only one member of the family was initially isolated and functionally characterized. Questions exist as to whether the other family members provide redundancy, tissue-specificity, or even encode enzymes with different catalytic properties. One way to address these questions is to link gene-specific transcript analysis with metabolic profiling that targets products and intermediates of the pathway in different tissues and/or in tissues in which the pathway is induced in response to biotic or abiotic stimuli. This approach has recently proved instructive for addressing gene function in sulfur metabolism (including glucosinolate biosynthesis) in *Arabidopsis* (Hirai et al. 2004) and in pyridine alkaloid biosynthesis in tobacco (Goossens et al. 2003).

► **Fig. 3.** Integration of transcriptomics and metabolomics for gene identification in the isoflavonoid pathway: **a** scheme for isoflavonoid biosynthesis in *Medicago sativa* (alfalfa) and *M. truncatula*. Enzymes are: PAL, L-phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CHR, “chalcone reductase”; IFS, “isoflavone synthase”; HI4’OMT, 2-hydroxyisoflavanone 4’-O-methyltransferase; DH, dehydratase; I2’H, isoflavone 2’-hydroxylase; IFR, isoflavone reductase; VR, vestitone reductase; GT, glycosyltransferase; MT, malonyltransferase; P450, cytochrome P450; OMT, O-methyltransferase. *Boxed structures* show compounds for which enzymes remain uncharacterized in *Medicago*. MG, medicarpin glucoside; FGM, formononetin glucoside malonate. Compounds marked with a * have been observed in metabolomic analysis; **b** *color-coded panels* show DNA microarray analysis of transcripts encoding functionally assigned genes of the isoflavonoid pathway (only one gene family member is shown for each enzyme apart from CHS) and candidates for genes encoding predicted O-methyltransferase, cytochrome P450 and GT genes involved in isoflavone modification, based on co-induction with identified metabolites. *Color coding* reflects relative expression level at the times shown (hours) after exposure to elicitor or water (control). Normalization was performed using GeneTraffic software. Signal intensities between two fluorescent images (Cy3 reference, Cy5 experimental) were normalized using LOWESS sub-grid normalization. The *color scale* indicates normalized signal intensities (log₂ ratio of fold change between experimental and reference samples); **c,d** HPLC-UV profiling of (iso)flavonoids in alfalfa cell suspension cultures 48 h after exposure to water (control, C) or yeast elicitor (D); **e** a hydrolyzed extract from cells that had been fed with labeled liquiritigenin (L), with label incorporation in formononetin (F), 2’hydroxyformononetin (2’HF) and medicarpin (M) indicating de novo synthesis in response to elicitation. IL, isoliquiritigenin

Targeted metabolite profiling in alfalfa (*Medicago sativa*) (Fig. 3c–e) and *Medicago truncatula* cell suspension cultures reveals induction of the same major isoflavonoid metabolites following exposure of the cultures to an elicitor from yeast cell walls (Liu and Dixon 2001; Suzuki et al. 2005; M. Farag and L.W. Sumner, unpublished results). The profiling methods developed in the past for flavonoids/isoflavonoids, and shown in Fig. 3b–d, rely primarily on HPLC with UV/visible detection (Graham 1991). This is a simple and reliable method for cases where metabolites have already been characterized and their properties (HPLC retention times and UV spectra) are known and authentic standards are available, e. g., for fingerprinting specific metabolites. However, mass spectrometry offers many advantages as a detection system for a more unbiased metabolomic approach, including greatly improved sensitivity and better structural resolution, particularly when employing tandem MS (Fiehn et al. 2000a,b; Sumner et al. 2003). Thus, LC/MS/MS analysis of elicited *M. truncatula* cell suspension cultures resulted in identification not only of the group of isoflavones, pterocarpan and their glycosides revealed by UV/visible analysis (Fig. 3a,c–e), but also of a number of additional isoflavones with unexpected A-ring methylation and methelenedioxy substitution, as well as isoflavene and aurone metabolites (M. Farag and L.W. Sumner, unpublished results).

DNA microarray analysis of *M. truncatula* cell cultures harvested at a range of times post-elicitation revealed induction of several members of the multi-gene families encoding early phenylpropanoid pathway, flavonoid branch and isoflavonoid-specific branch pathway enzymes (Fig. 3b). Integration of metabolite and transcript data from such experiments in an interrogable database allows in silico comparison of transcript induction kinetics with appearance of specific metabolites. For example, yeast elicitor induced accumulation of a range of methylated isoflavones, suggesting the need for multiple isoflavone OMTs, and potential candidate genes can be identified from the microarray dataset (Fig. 3a). Likewise, it is possible to identify a number of candidate *cytochrome P450* and *GT* genes potentially involved in the formation of the various glycosylated, oxidized isoflavone derivatives accumulating in response to elicitation (Fig. 3b).

4.2 Deciphering the Triterpene Pathway in *Medicago*

Triterpene saponins are a class of plant natural products with a wide range of bioactivities, from allelopathic and anti-microbial to anticancer and anticholesterolemic (Waller et al. 1993; Behboudi et al. 1999; Haridas et al. 2001; Osbourn 2003), and are important components of a number of herbal medicines (Xu 2001; Chan et al. 2002). Most of the genes involved in the biosynthesis of these complex molecules remain uncharacterized.

The saponins of *M. truncatula* and alfalfa exist as glycosides of at least five different triterpene aglycones; medicagenic acid, hederagenin, soyasa-

pogenol B, soyasapogenol E and bayogenin (Huhman and Sumner 2002). These compounds are derived from β -amyrin, the cyclization product of 2,3-oxidosqualene (Kushiro et al. 1998; Suzuki et al. 2002). The downstream reactions in the biosynthesis of *M. truncatula* saponins include a number of cytochrome P450 dependent hydroxylations/oxidations and several glycosyl transfer reactions catalyzed by uridine diphosphate-dependent GTs. Based on current EST and partial genome sequence information, *M. truncatula* contains P450 and GT supergene families each with approximately 300 members. It is more than likely that most of the genes involved in triterpene biosynthesis in *Medicago* are already physically present in EST and genomic library collections, and approaching the identification from these resources, rather than taking a protein purification approach, is attractive in view of the relative instability of P450s and GTs, and the insoluble nature of the former. However, the numbers of potential candidate genes is problematic. An integrated approach involving comparison of transcript and metabolite behavior in response to a metabolic perturbation represents one way to address this problem.

Exposure of *M. truncatula* cell suspension cultures to methyl jasmonate (MeJA) induces triterpene saponin accumulation preceded by induction of the triterpene cyclase β -amyrin synthase (β -AS) (Suzuki et al. 2002, 2005). By coupling DNA array approaches to profile transcripts corresponding to all 100-plus GTs in MeJA-induced and control cell cultures with metabolite (saponin) profiling and in silico expressed sequence tag (EST) data mining, two GTs, designated UGT71G1 and UGT73K1, were selected and subsequently found to be active with *Medicago* triterpene aglycones (Achnine et al. 2005). The basis of the selection was to compare tissue-specificity (in nearly 40 cDNA libraries used for EST sequencing) and induction kinetics (assuming co-induction) of candidate GTs compared to the triterpene cyclase β -AS. A genomics approach has also recently been used to identify diterpene GTs from *Stevia rebaudiana* (Richman et al. 2005).

Of more than 40 potential triterpene and phenolic acceptor molecules tested in in vitro assays, UGT73K1 only showed activity against three triterpenes (soyasapogenol E, soyasapogenol B, and hederagenin), consistent with the appearance of glucosides of these compounds in the MeJA-treated cell cultures (Achnine et al. 2005). However, on the basis of in vitro kinetic measurements, UGT71G1 appeared to have a clear preference for flavonoid compounds (quercetin and 5-hydroxyisoflavones) as compared to triterpenes (Fig. 4a). The K_{cat}/K_m ratio of UGT71G1 was more than 30-fold higher for quercetin than for hederagenin. This raised the question of whether the identification of UGT71G1 as a triterpene GT was erroneous, and simply a reflection of in vitro enzyme promiscuity. However, comparisons of the levels of UGT71G1 transcripts, triterpenes, and isoflavones in *M. truncatula* cell cultures responding to MeJA or yeast elicitor (a strong inducer of isoflavones) suggested a lack of association of UGT71G1 with isoflavone glycoside formation, and quercetin glycosides were not detected in the cultures (Achnine et al. 2005) (Fig. 4b). Thus, metabolomic analysis resolved the issue of in vivo substrate specificity

a **b**

	Yeast elicitor	Methyl Jasmonate
UGT7G1	—	++
Quercetin glucosides	—	—
Genistein-7-O-glucoside	++	—
Biochanin A-7-O-glucoside	—	—
3-Glc-Hederagenin	—	++
28-Glc-Hederagenin	—	++
3-Glc-28-Glc-Medicagenic acid	—	++

Fig. 4. Metabolomics for gene identification in the triterpene pathway: **a** substrates and products of UGT71G1 in vitro; **b** summary of transcript and metabolite analyses in vivo. Yeast elicitor induces the formation of genistein glucoside, but not UGT71G1 transcripts, whereas methyl jasmonate induces hederagenin glucosides and UGT71G1. Other GTs are likely involved in the formation of isoflavone glucosides in yeast-elicited cells

in this cell culture. Even so, a full understanding of the specificity of UGT71G1 in vivo must await metabolomic analysis of plants in which its gene is selectively down-regulated.

5 Single Cell or Isolated Tissue Metabolomics

The above examples make a strong case for the argument that knowledge of which compounds are, and which are not, present in a particular plant tissue may be essential for ascribing function to a cloned gene. This statement also implies that it is necessary to be able to correlate presence of an enzyme or its transcripts with those metabolites made in the same cells or cell types.

Many of the plant EST projects currently accessible through publicly available web sites, such as the TIGR Plant Gene Indices (<http://www.tigr.org/tdb/tgi/plant.shtml>), provide expression data as EST counts in a wide range of cDNA libraries prepared from different tissue, developmental stages, and responses to various biotic and abiotic stresses and stimuli. A list of currently available *M. truncatula* EST libraries is given in Table 1. It is informative to contrast this list with the paucity of information, and almost non-existent tissue resolution, on the *M. truncatula* metabolome. A survey of the literature, including a comprehensive phytochemical dictionary (ILDIS 1994), can provide a list of secondary metabolites found in different legume species, but the conditions of the tissues from which the metabolites were identified is generally poorly defined, and, without exception, the studies reported were targeted and non-comprehensive. A similar situation holds for those other species, such as *Arabidopsis*, rice, corn, soybean and poplar, for which extensive genomic and EST sequence information is available.

Furthermore, the degree of tissue and treatment resolution currently found in EST databases such as indicated in Table 1 is of itself insufficient to provide the kind of integration necessary to allow for meaningful correlations between transcriptome and metabolome. For example, proanthocyanidins are found in alfalfa in specific regions of the seed coat (Debeaujon et al. 2003) and in the heads of glandular trichomes (Aziz et al. 2005), but have not been shown to be present in other parts of the plant. As described above, several studies have utilized cell suspension cultures to obtain a more homogeneous cellular background for integrated transcript and metabolome profiling (Goossens et al. 2003; Achnine et al. 2005; Suzuki et al. 2005). However, although the kinetics of changes in many metabolites can be shown to cluster with changes in transcript levels, most of the metabolites remain unknown. Spatially resolved metabolomic and transcriptomic analysis is technically challenging but is an essential approach for understanding the relationship between gene expression and metabolism in whole plants. Laser capture microdissection (LCM) techniques can now provide tissue samples for such analyses (Kerek et al. 2003), and techniques are available for the construction of cDNA libraries from minute tissue samples (Belyavsky et al. 1989).

Trichomes provide an excellent system with which to develop technologies for integrated transcriptomics/metabolomics on a tissue or even single cell level. Trichomes are epidermal appendages found on the aerial organs of many plants. Glandular trichomes have a high capacity to synthesize, store and secrete secondary metabolites that help protect the plant against insect

Table 1. *M. truncatula* EST libraries as of May 1st, 2003 (TIGR Release 7.0), with a total of 189,714 EST sequences (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=medicago)

Tissue	EST library	EST #
1. Leaf		
Non-challenged	Developing leaf ^a	9415
Biotic challenged	Insect herbivore attacked leaf	10,309
	<i>Colletotrichum trifolii</i> infected leaf	6003
	<i>Phoma medicaginis</i> infected leaf	3281
Abiotic challenged	Phosphate starved leaf	10,188
2. Root		
Non-challenged	Developing root, no symbiosis ^a	3054
	KV0, non-nodulated root	2752
	MtRHE, root hair-enriched	899
Biotic challenged		
	KV1, root – four day post-nodulation	2840
	KV2, root – two days post-nodulation	3330
	KV3, root – three days post-nodulation	4315
	MtBB, root – four days post-nodulation	7807
	GVN, one-month-old root nodules	6468
	GVSN, senescent nodules	2788
	R108, young root nodules	438
	Nodulated root, mixed	3299
	MHAM, <i>Glomus versiforme</i> infected root	7368
	MtBC, mycorrhizal root	8601
	DSIR, fungus-elicited root	2463
	BNIR, nematode-infected root	3154
Abiotic challenged	MtBA, nitrogen starved root	7939
	Rootphos(-), phosphate starved root	1967
	MHRP-, phosphate starved root	2658
	MGHG, β -glucan-elicited root	2687
	HOGA, oligogalacturonide-elicited root	2861
3. Stem		
Non-challenged	Developing stem ^a	10,783
4. Seed		
Non-challenged	GESD, developing seed ^a	2672
	GLSD, developing seed	2944
	Germinating seed	6619
5. Flower		
Non-challenged	Developing flower ^a	6724
6. Cell culture		
Biotic challenged	Yeast elicited cell culture ^a	9859
Abiotic challenged	Methyl jasmonate-induced cell culture ^a	6900
7. Pods		
Non-challenged	Developing pod ^a	1915

Table 1. (continued)

Tissue	EST library	EST #
8. Mixed tissues		
Non-challenged	Cotyledon and leaf	2143
Abiotic challenged	Drought stressed seedlings	9520
	UV Irradiated seedling	6748

^a These tissue types are the only ones to date for which metabolomic analysis has been initiated. Preliminary results have led to identification of approximately 200 primary metabolites, and a significantly smaller number of secondary metabolites from each tissue source (ILDIS 1994; Huhman and Sumner 2002; Achnine et al. 2005; Broeckling et al. 2005; Suzuki et al. 2005; L.W. Sumner, C. Broeckling, D.V. Huhman and M. Farag, unpublished results)

For EST libraries descriptions, refer to CEDA (Comparative EST Data Analysis in *M. truncatula*) at <http://bionfo.noble.org/CEDA.htm>

predation and other biotic challenges (Wagner 1991; Ranger and Hower 2001; Wagner et al. 2004). For example, the peltate glandular trichomes of mint produce a suite of defensive monoterpenes that are the major components of, and give the characteristic smell and flavor to, mint oil (McCaskill et al. 1992; Voirin and Bayet 1996); trichomes from tomato species collectively produce a number of insecticidal sesquiterpenes, acyl sugars and methyl ketones (Li et al. 1999; Antonious 2001; Maluf et al. 2001); and tobacco trichomes produce diterpenes and acyl sugars (Kandra et al. 1990; Guo and Wagner 1995).

cDNA libraries have been constructed from trichomes of mint (Lange et al. 2000), sweet basil (Gang et al. 2001), and wild and cultivated tomatoes (http://www.tigr.org/tigr-scripts/tgi/T{_}index.cgi?species=tomato). The mint and tomato trichomes show a strong preponderance of transcripts (represented by ESTs) encoding enzymes of terpene metabolism. Because of the highly specialized biosynthetic functions of the trichomes from these two species, considerable biosynthetic information was obtained by sequencing only a relatively small number of ESTs (1000–2500). This is likely to be the case for other species that produce biochemically-specialized trichomes or other secretory cell types, such as hops (Hirosawa et al. 1995) and vanilla orchid (Joel et al. 2003). Combining in-depth EST sequencing and metabolite profiling will provide a powerful approach for gene discovery that, because of the specialized nature of the trichome, will directly address bioactive secondary metabolites.

6 Concluding Remarks

We argue for the wider use of metabolomics in the context of natural product pathway gene discovery. In depth profiling of plant tissues for natural products, coupled to parallel analysis of gene transcript levels using EST count, microarray, cDNA-AFLP or SAGE analysis, is a powerful tool when applied to a biological system in which “differential display” of metabolites and gene transcripts can be visualized and compared. Such systems include knock-out mutants, plants or cell lines over-expressing biosynthetic enzymes or transcription factors, and plants, tissues or cell lines exposed to biotic or abiotic perturbations. Comparisons of tissues at different developmental stages should also be addressed. In this context, it would be very valuable to obtain in depth metabolite profiles for all tissue and treatment types that have been sampled in EST sequencing projects, thereby providing a resource for initial in silico analysis pairing enzymes with potential substrates and products as a tool to assist functional annotation. The major problem is that many metabolites are unknown and, where they can be predicted (e.g., as intermediates in a complex pathway, Fig. 4a), standard compounds are generally unavailable.

Acknowledgements. We thank David Huhman and Drs Lloyd Sumner and Mohamed Farag for assistance with mass spectrometry, sharing results prior to publication, and useful discussions. Work in the authors’ laboratory was supported by grants to RAD from the National Science Foundation (#DBI 0109732), Forage Genetics International, the Oklahoma Center for the Advancement of Science and Technology Health Sciences Program, and the Samuel Roberts Noble Foundation.

References

- Achnine L, Huhman DV, Sumner LW, Blount JW, Dixon RA (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J* 41:875–887
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar HL, Schmid M, Weigel D, Carter DE, Marchand T, Risseuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301:653–657
- Antonious GF (2001) Production and quantification of methyl ketones in wild tomato accessions. *J Environ Sci Health B36*:835–848
- Aziz N, Paiva NL, May GD, Dixon RA (2005) Profiling the transcriptome of alfalfa glandular trichomes. *Planta* 221:28–38
- Behboudi S, Morein B, VillacresEriksson MC (1999) Quillaja saponin formulations that stimulate proinflammatory cytokines elicit a potent acquired cell-mediated immunity. *Scand J Immunol* 50:371–377
- Belyavsky A, Vinogradova T, Rajewsky K (1989) PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res* 17:2919–2933

- Broeckling CD, Huhman DV, Farag M, Smith JT, May GD, Mendes P, Dixon RA, Sumner LW (2005) Metabolic profiling of *Medicago truncatula* cell cultures reveals effects of biotic and abiotic elicitors on primary metabolism. *J Exp Bot* 56:323–336
- Chan RYK, Chen WF, Dong A, Guo D, Wong MS (2002) Estrogen-like activity of ginsenoside Rg1 derived from *Panax notoginseng*. *J Clin Endocrinol Metab* 87:3691–3695
- Chapple CCS, Vogt T, Ellis BE, Somerville CR (1992) An Arabidopsis mutant defective in the general phenylpropanoid pathway. *Plant Cell* 4:1413–1424
- Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L (2003) Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. *Plant Cell* 15:2514–2531
- Dixon RA (1999) Isoflavonoids: biochemistry, molecular biology and biological functions. In: Sankawa U (ed) *Comprehensive natural products chemistry*, vol 1. Elsevier, Amsterdam, pp 773–823
- Dixon RA (2004) Phytoestrogens. *Annu Rev Plant Biol* 55:225–261
- Dixon RA, Ferreira D (2002) Molecules of Interest: Genistein. *Phytochemistry* 60:205–211
- Dixon RA, Achnine L, Kota P, Liu C-J, Reddy MS, Wang L (2002) The phenylpropanoid pathway and plant defense – a genomics perspective. *Mol Plant Pathol* 3:371–390
- Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2000a) Metabolite profiling for plant functional genomics. *Nature Biotechnol* 18:1157–1161
- Fiehn O, Kopka J, Trethewey RN, Willmitzer L (2000b) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal Chem* 72:3573–3580
- Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Chapple C (2002) The Arabidopsis *REF8* gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *Plant J* 30:33–45
- Gang DR, Wang J, Dudareva N, Nam KH, Simon JE, Lewinsohn E, Pichersky E (2001) An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. *Plant Physiol* 125:539–555
- Goossens A, Häkkinen S, Laakso I, Seppänen-Laakso T, Biondi S, de Sutter V, Lammertyn E, Nuutila AM, Söderlund H, Zabeau M, Inzé D, Oksman-Caldentey KM (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proc Natl Acad Sci USA* 100:8595–8600
- Graham TL (1991) A rapid, high resolution high performance liquid chromatography profiling procedure for plant and microbial aromatic secondary metabolites. *Plant Physiol* 95:584–593
- Guo Z, Wagner GJ (1995) Biosynthesis of cembratrienols in cell-free extracts from trichomes of *Nicotiana tabacum*. *Plant Sci* 110:1–10
- Haridas V, Higuchi M, Jayatilake GS, Bailey D, Mujoo K, Blake ME, Arntzen CJ, Gutterman JU (2001) Avicins: triterpenoid saponins from *Acacia victoriae* (Benth) induce apoptosis by mitochondrial perturbation. *Proc Natl Acad Sci USA* 98:5821–5826
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 101:10205–10210
- Hirosawa T, Saito T, Tanaka T, Matsushima H (1995) SEM observation and HPLC analysis of the accumulation of alpha- and beta-acids in the fresh developing hop (*Humulus lupulus* L.) peltate glandular trichomes. *J Electron Microsc* 44:145–147
- Huhman DV, Sumner LW (2002) Metabolic profiling of saponin glycosides in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry* 59:347–360
- ILDIS (1994) *Phytochemical dictionary of the leguminosae*. Plants and their constituents, vol 1. Chapman and Hall, London
- Joel DM, French JC, Graft N, Kourteva G, Dixon RA, Havkin-Frenkel D (2003) A hairy tissue produces vanillin. *Israel J Plant Sci* 51:157–159
- Kandra G, Severson R, Wagner GJ (1990) Modified branched-chain amino acid pathways give rise to acyl acids of sucrose esters exuded from tobacco leaf trichomes. *Eur J Biochem* 188:385–391

- Karimi M, Inze D, Depicker A (2002) GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* 7:193–195
- Kerek NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM (2003) Laser capture microdissection of cells from plant tissues. *Plant Physiol* 132:27–35
- Kochs G, Grisebach H (1986) Enzymic synthesis of isoflavones. *Eur J Biochem* 155:311–318
- Kushiro T, Shibuya M, Ebizuka Y (1998) β -Amyrin synthase. Cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *Eur J Biochem* 256:238–244
- Kutchan TM (2002) Sequence-based approaches to alkaloid biosynthesis gene identification. In: Romeo JT, Dixon RA (eds) *Phytochemistry in the genomics and post-genomics eras*. Elsevier Science, Oxford, pp 163–178
- Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D, Croteau R (2000) Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc Natl Acad Sci USA* 97:2934–2939
- Li AX, Steffens JC (2000) An acyltransferase catalyzing the formation of diacylglycerol is a serine carboxypeptidase-like protein. *Proc Natl Acad Sci USA* 97:6902–6907
- Li AX, Eannetta N, Ghangas GS, Steffens JC (1999) Glucose polyester biosynthesis. Purification and characterization of a glucose acyltransferase. *Plant Physiol* 121:453–460
- Liu C-J, Dixon RA (2001) Elicitor-induced association of isoflavone O-methyltransferase with endomembranes prevents formation and 7-O-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. *Plant Cell* 13:2643–2658
- Lukacin R, Matern U, Specker S, Vogt T (2004) Cations modulate the substrate specificity of bifunctional class 1 O-methyltransferase from *Ammi majus*. *FEBS Lett* 577:367–370
- Maluf WR, Campos GA, das Gracas CM (2001) Relationships between trichome types and spider mite (*Tetranychus evansi*) repellence in tomatoes with respect to foliar zingiberene contents. *Euphytica* 121:73–80
- McCaskill D, Gershenzon J, Croteau R (1992) Morphology and monoterpene biosynthetic capabilities of secretory cell clusters isolated from glandular trichomes of peppermint (*Mentha piperita* L.). *Planta* 187:445–454
- Osborn AE (2003) Molecules of interest. Saponins in cereals. *Phytochemistry* 62:1–4
- Ranger CM, Hower AA (2001) Glandular morphology from a perennial alfalfa clone resistant to the potato leafhopper. *Crop Sci* 41:1427–1434
- Richman A, Swanson A, Humphrey T, Chapman R, McGarvey BD, Pocs R, Brandle J (2005) Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. *Plant J* 41:56–67
- Saghatelian A, Trauger S, Want E, Hawkins E, Siuzdak G, Cravatt B (2004) Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry* 43:14332–14339
- Schröder J (1997) A family of plant-specific polyketide synthases: facts and predictions. *Trends Plant Sci* 2:373–378
- Schröder J, Raiber S, Berger T, Schmidt A, Schmidt J, Soares-Sello AM, Bardshiri E, Strack D, Simpson TJ, Veit M, Schröder G (1998) Plant polyketide synthases: a chalcone synthase-type enzyme which performs a condensation reaction with methylmalonyl-CoA in the biosynthesis of C-methylated chalcones. *Biochemistry* 37:8417–8425
- Shirley AM, McMichael CM, Chapple C (2001) The sng2 mutant of *Arabidopsis* is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose: choline sinapoyltransferase. *Plant J* 28:83–94
- Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62:817–836
- Suzuki H, Achnine L, Xu R, Matsuda SPT, Dixon RA (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant J* 32:1033–1048
- Suzuki H, Reddy MSS, Naoumkina M, Aziz N, May GD, Huhman DV, Sumner LW, Blount JW, Mendes P, Dixon RA (2005) Methyl jasmonate and yeast elicitor induce differential genetic and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. *Planta* 220:698–707

- Tropf S, Lanz T, Rensing SA, Schröder J, Schröder G (1994) Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. *J Mol Evol* 38:610–618
- Vogt T (2004) Regiospecificity and kinetic properties of a plant natural product O-methyltransferase are determined by its N-terminal domain. *FEBS Lett* 561:159–162
- Voirin B, Bayet C (1996) Developmental changes in the monoterpene composition of *Mentha X piperita* leaves from individual peltate trichomes. *Phytochemistry* 43:573–580
- Wagner GJ (1991) Secreting glandular trichomes: more than just hairs. *Plant Physiol* 96:675–679
- Wagner GJ, Wang E, Shepherd RW (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann Bot* 93:3–11
- Waller GR, Jurzysta M, Thorne RLZ (1993) Alleopathic activity of root saponins from alfalfa (*Medicago sativa* L.) on weeds and wheat. *Bot Bull Acad Sin* 34:1–11
- Wein M, Lavid N, Lunkenbein S, Lewinsohn E, Schwab W, Kaldenhoff R (2002) Isolation, cloning and expression of a multifunctional O-methyltransferase capable of forming 2,5-dimethyl-4-methoxy-3(2H)-furanone, one of the key aroma compounds in strawberry fruits. *Plant J* 31:755–765
- Woo HH, Orbach MJ, Hirsch AM, Hawes MC (1999) Meristem-localized inducible expression of a UDP-glycosyltransferase gene is essential for growth and development in pea and alfalfa. *Plant Cell* 11:2303–2315
- Xu Y (2001) Perspectives on the 21st century development of functional foods: bridging Chinese medicated diet and functional foods. *Int J Food Sci Technol* 36:229–242