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

R.A. Dixon, L. Achnine, B.E. Deavours, and M. Naoumkina

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:

- Knock-out or knock-down lines
- Plants or cell lines overexpressing biosynthetic enzymes or transcription factors
- Abiotic/biotic perturbation
- Metabolically-specialized cells or tissues
- Developmental stages

RNA

Gene transcripts
- EST counts
- Microarray
- cDNA-AFLP
- SAGE

Metabolites

Metabolic profiling
- LC/UV
- LC/MS
- GC/MS
- FTIC MS

Identification of candidate genes

- Homology/structure based
- Differential expression
- Induction kinetics

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 acyltransferases 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(2H)-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-hydroxyisoflavonone synthase, com-
monly termed isoflavone synthase [IFS]). The resulting 2-hydroxyisoflavonone
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 biochem-
ical approaches, such that only one member of the family was initially isolated
and functionally characterized. Questions exist as to whether the other fam-
ily 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 prod-
ucts 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 sul-
fur 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; CAH, cinnamate 4-hydroxylation;
4CL, 4-coumarate: CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CHR, "chalcone
reductase"; IFS, "isoflavone synthase"; H14'OMT, 2-hydroxyisoflavonone 4'-O-methyltransferase;
DH, dehydratase; I2'H, isoflavone 2'-hydroxylase; IFR, isoflavone reductase; VR, vestitene
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 metabolic analysis; b color-coded panels show DNA
microarray analysis of transcripts encoding functionally assigned genes of the isoflavonoid path-
way (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 isoflavonone modification, based on co-induction with identified metabolites. Color coding re-
fects relative expression level at the times shown (hours) after exposure to elicitor or water
(control). Normalization was preformed 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 (log 2 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 liquisiritigenin (L), with label in-
corporation in formononetin (F), 2'-hydroxyformononetin (2'HP) 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/iso flavonoids, 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 s 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 methylenedioxy 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 multigene 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 anti-cholesterolemic (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 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
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)

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

<table>
<thead>
<tr>
<th>Tissue</th>
<th>EST library</th>
<th>EST #</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Mixed tissues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-challenged</td>
<td>Cotyledon and leaf</td>
<td>2143</td>
</tr>
<tr>
<td>Abiotic challenged</td>
<td>Drought stressed seedlings</td>
<td>9520</td>
</tr>
<tr>
<td></td>
<td>UV Irradiated seedlings</td>
<td>6748</td>
</tr>
</tbody>
</table>

* 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 monoterpene s 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


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


