Chapter 13
Genomics of Secondary Metabolism in Soybean

Terry Graham, Madge Graham, and Oliver Yu

An Overall Perspective on Plant Secondary Metabolites

Secondary product pathways in plants are extraordinarily diverse. Although classical chemical and biochemical analyses have gone a long way in delineating the major pathways and metabolites in many plant species, no plant species has been completely characterized for all of the secondary products that it produces *or is capable of producing*. While examinations of plants at the metabolic level (e.g., through metabolic profiling/metabolomics) is powerful in that it provides a picture of the ultimate net accumulation of endproducts in various cells, tissues and organs, unexpected or new metabolites are often found in a given species when the plant is examined under previously unexamined conditions. A very simple example is the production of large quantities of formononetin by soybean following treatment by the disease resistance-inducing herbicide lactofen (Landini et al. 2002). While this metabolite is found at high levels in some other legumes, such as chickpea, it is normally undetectable or at trace levels in soybean. Thus, it is very difficult by analyses of metabolites per se to know all of the potential metabolites that can be produced by a plant. This inherent limitation is an aspect where genomic analyses will have particular impact in that the presence of a gene for a particular enzyme may be suggested even though the product of its action is not.

Quantification of the secondary metabolites in classical biochemical or metabolomics approaches is also a challenge. While quantitative data is readily obtained from various chromatographic and spectroscopic techniques, the actual tissue concentration of individual metabolites requires the use of standards for each metabolite, something that is extremely difficult to achieve across all metabolites and tissues. Moreover, many important biological processes involve changes in metabolites in individual cells or groups of cells, which is beyond the limits of detection of most

Terry Graham
Department of Plant Pathology, Ohio State University, Columbus, OH 43210, USA
e-mail: graham.1@osu.edu

G. Stacey (ed.), *Genetics and Genomics of Soybean*,

biochemical analyses. New techniques such as laser capture microdissection (LCM, see e.g., Klink et al. 2005) promise to provide genome wide mRNA expression information at a cellular level. Such analyses may provide additional critical and complementary insights.

Although secondary product metabolism across the plant kingdom is diverse, extensive characterization of the natural products made by plants has suggested that the major metabolites normally produced by a given plant family, genus or species are somewhat limited. While fundamental natural products such as some of the simple flavonoids or anthocyanins are widespread in plants, beyond these common metabolites, plant families tend to specialize in a few major pathways. As just one example, the phenylpropanoid derived isoflavones are predominant metabolites in the Leguminosae. Likewise the major defense related phytoalexins in the legumes are coumestans and pterocarpans further derived from the isoflavonoids. In contrast, in the Solanaceae, phenylpropanoid-derived metabolites are relatively simple and alkaloids and sesquiterpenoid-derived phytoalexins predominate. Further specialization is found even within a family. Within the family Leguminosae, to which soybeans belong, the specific major isoflavones and pterocarpans that predominate are different at a genus level. Thus, the specific pterocarpan phytoalexins that predominate in soybean (the glyceollins) are different from those that predominate in alfalfa (the medicarpins), etc. At the ultimate level of specialization, even different cultivars of a specific species, for instance soybeans (*Glycine max*), can produce very different mixtures of specific endproducts (e.g., different flavonols and flavonol glycosides, Buttery and Buzzell 1973, 1975).

In this short review, we will focus on (1) the information on secondary product enzymes, pathways and their expression that can be mined from the extensive soybean EST and derived databases and (2) the information that is just starting to emerge from a few interesting genome wide microarray studies on secondary products produced in *Glycine max*. Since the phenylpropanoid pathway in soybean has been particularly well characterized at a biological and biochemical level, we focus primarily on this pathway and its many branch pathways (e.g., for the flavonoids, isoflavonoids, anthocyanins, etc). It is beyond the scope of this current review to cover other major secondary metabolites, such as the saponins, waxes, fatty acids, and alkaloids.

Secondary Product Metabolism in Soybean

Figures 13.1A and 13.1B outline the major secondary metabolic pathways in soybean that will be the focus of this review. Much of the work on these pathways in soybean was the result of their importance to various phenotypes, such as flower color or seed coat pigmentation (anthocyanins), protection against UV light (flavonols), and interaction with various pests or symbionts (isoflavones). In Fig. 13.1A, we show an abbreviated schematic for lignin, flavonol and anthocyanin

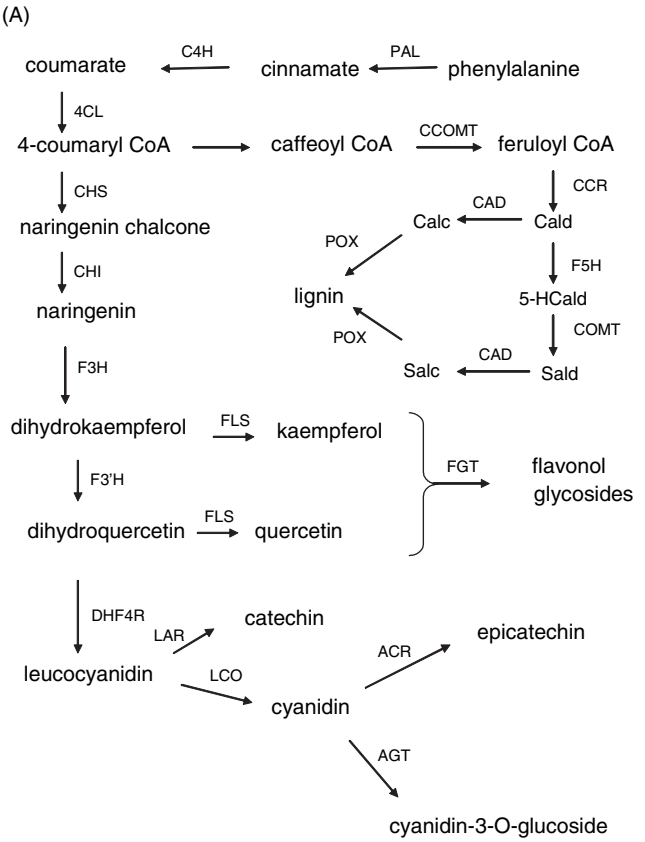

Fig. 13.1A Lignin, flavonol and anthocyanin biosynthesis Abbreviations: 4CL, 4-coumaryl-CoA ligase; ACR, anthocyanidin reductase; AGT, anthocyanidin-3-O-glucosyltransferase; C4H, cinnamic acid 4-hydroxylase; CAD, cinnamyl alcohol dehydrogenase; CCOMT, caffeoyl-CoA O-methyltransferase; CCR, cinnamoyl-CoA reductase; CHI, chalcone isomerase; CHS, chalcone synthase; COMT, caffeic acid O-methyltransferase; DHF4R, dihydroflavonol 4-reductase; F3′H, flavonoid 3′-hydroylase; F3H, flavanone 3-hydroxylase; F5H, ferulate 5-hydroxylase; FLS, flavonol synthase; LAR, leucoanthocyanidin reductase; LCO, leucocyanidin oxygenase; POX, lignin peroxidase; PAL, phenylalanine ammonia lyase

biosynthesis and in Fig. 13.1B we show a simplified pathway for the formation of the isoflavones and the pterocarpans, including the phytoalexin glyceollin. Although we tried to illustrate most of the key enzymes, each of these pathways is considerably more complex than shown. In the sections below, we first describe each branch of these pathways in more detail, including some important aspects of their genetics, biology and biochemistry. We then address them from a genomics perspective, including information that can be mined from the various EST libraries and recent microarray work.

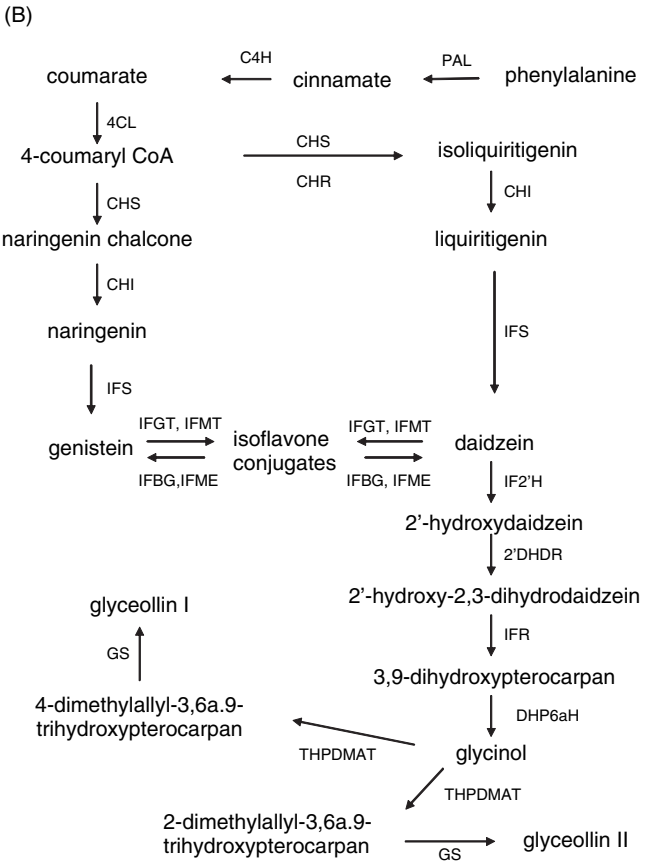

Fig. 13.1B Isofavone and glyceollin biosynthesis Abbreviations: 2′DHDR, 2′-hydroxydihydrodaidzein reductase; CHR, chalcone reductase; DFR, dihydroflavonol-4-reductase, DHP6aH, 3,9-dihydroxypterocarpan 6a-hydroxylase; GS, glyceollin synthase; IF2′H, isoflavone 2′-hydroxylase; IFBG, isoflavone 7-O-beta glucosidase; IFGT, isoflavone 7-0-glucosyl transferase; IFME, isoflavone malonyl esterase; IFMT, isoflavone malonyl transferase; IFR, isoflavone reductase; IFS, isoflavone synthase; THPDMAT, trihydroxypterocarpan dimethylallyltransferase

Early Phenylpropanoids

The early phenylpropanoids include those metabolites from phenylalanine to the simple phenylpropanoic acids. Early metabolic profiling of various soybean organs (Graham 1991a, b) suggested that these metabolites do not normally accumulate to high levels, suggesting that they are efficiently used as substrates for the more complex metabolites shown in Fig. 13.1. Phenylalanine ammonia lyase (PAL) is considered the entry point for phenylpropanoid metabolism and a major point of regulation for loading the entire metabolic grid deriving from the simple phenylpropanoids.

Cinnamate-4-hydroxylase (C4H) adds a hydroxyl group to the 4 position, giving rise to *p*-coumaric acid. Following this, additional hydroxylations and/or methylations of the aromatic ring occur, which at least partially determine if these simple phenylpropanoids are destined to be diverted into lignin/suberin, flavonoids/anthocyanins or simple esters (Hahlbrock 1981). Formation of the CoA derivatives through the action of 4-coumarate:CoA ligase (4CL) is the entry point for the simple phenylpropanoids into the various alternative pathways leading to more complex metabolites.

Phenolic Polymers: Lignin and Suberin

Lignin is synthesized mostly as a component of secondary cell walls, which provide additional structural rigidity to support the plants. Although the biosynthesis of lignin involves relatively few enzymes, there are some unresolved complexities in the actual enzymatic pathways (Humphreys and Chapple 2002; Boerjan et al. 2003). Basically, the process involves a series of reductions of CoA derivatives of phenylpropanoic acids first to an aldehyde, through the action of cinnamoyl-CoA reductase (CCR), and then to an alcohol by cinnamyl alcohol dehydrogenase (CAD). This is followed by reoxidation and polymerization by ligninperoxidases. Much of the complexity of the pathway involves how and where in the grid of enzymes the hydroxylation and methylation reactions occur that are needed for the formation of the corresponding caffeic, ferulic, 5-hydroxyferulic and sinapic acids derivatives. Although ferulate 5-hydroxylase (F5H), caffeic acid O-methyl transferase (COMT) and caffeoyl-CoA O-methyltransferase (CCOMT) are thought to be involved, other enzymes are possible (Humphreys and Chapple 2002; Boerjan et al. 2003). Laccases (diphenol oxidases) are also thought to play a potential role in lignin formation (Gavnholt and Larsen 2002).

Unlike lignin which has a polyaromatic structure, suberin consists of both polyaromatic and polyaliphatic domains. The polyaliphatic domain makes suberin a very hydrophobic polymer, which participates primarily in exclusion of water, for instance in cell walls of the casparian strip in roots. The composition of the monomeric constituents of suberin varies in different species. Common aliphatic monomers include α-hydroxyacids and α, ω-diacids. Like lignin, the monomeric constituents of the polyaromatic domain are hydroxycinnamic acids and derivatives (for a recent review, see Ma and Peterson 2003).

As a part of general defense mechanisms, lignin content increases markedly during biotic (e.g., infection) and abiotic (e.g., wounding) stress. Formation of lignin and suberin-like polymers in plants is developmentally important for the formation of the cell walls of many cells, but particularly in secondary cell walls (Humphreys and Chapple 2002) and the endodermis (Ma and Peterson 2003), respectively. In soybean, the formation of these phenolic polymers is also a relatively early event in infected or pathogen elicitor treated tissues (Graham and Graham 1991a; Mohr

and Cahill 2001; Lozovaya et al. 2004). Some genomic aspects of the induction of enzymes for lignin formation in wounded tissues and pathogen resistance are discussed in detail below.

Flavonols

The precise roles of the flavonols are not fully understood, although they are very widespread in the plant kingdom. In soybean, the flavonols are present in all mature aerial tissues, but not in roots or seedling organs (Graham 1991b). A possible function of flavonols in protection against UV-B irradiation was suggested (see e.g., Landry et al. 1995; Logemann et al. 2000). The flavonols are formed through the action of chalcone synthase (CHS) and chalcone isomerase (CHI) which lead to the flavanone naringenin. Flavanone 3-hyroxylase (F3H) then leads to dihydrokaempferol, which can be further hydroxylated by flavonoid 3′-hydroxylase (F3′H) to form dihydroquercetin. Kaempferol and quercetin are then formed through the action of flavonol synthase (FLS). Not all soybean cultivars have both kaempferol and quercetin. The presence of quercetin requires the presence of the *T* gene (Buttery and Buzzell 1973; Buzzell et al. 1987), which encodes F3′H (Zabala and Vodkin 2003). Soybean TC216289, which encodes a F3′H, is a close homolog to the *TT7* gene in *Arabidopsis* which helps impart testa (seed coat) color. The flavonol, isorhamnetin, was also identified in certain soybean lines (Le- Van, N. and Graham, T. L., unpublished).

Like many end products of secondary product metabolism, the flavonols are usually present as conjugates (e.g., as glycosides). The patterns of glycosylation can be very complex, perhaps suggesting roles that we still do not fully understand. Several of the genes involved in the accumulation of flavonol 3-O-glycosides in soybean leaves were identified through a combination of biochemistry and classical genetics. Buttery and Buzzell (1975) identified nine different 3-O-glycosides for each of the flavonols, kaempferol or quercetin. Genes for glycosylation of kaempferol or quercetin are the same. The kaempferol monoglucoside (designated as K5) is found in all soybean cultivars, and both K5 and Q5 are found in all cultivars carrying the *T* gene (for F3′H). At least four additional enzymes give rise to the various diglycosides and triglycosides by simple additions to these monoglucosides. These include glucosyl transferases that transfer glucose in β linkages to the 6 position *(FG1)* or the 2 position *(FG3)* of the monoglucoside and rhamnosyl transferases that transfer rhamnose in α linkages to either the 6 position *(FG2)* or the 2 position *(FG4)* of the monoglucoside. Taken together, the *T* and various *FG* genes can give rise to over 32 flavonoid combinations for kaempferol and quercetin alone.

Flavonols can also serve as co-pigments in flowers. A *wm* locus that is associated with magenta flower color dramatically reduces flavonol glycoside levels in leaves and flowers, but does not affect their presence in pod pubescence (Buttery and Buzzell 1987). The magenta flower color is in fact thought to be due to the lack of flavonol glycosides in the otherwise purple flower background (caused by

anthocyanins, Buzzell et al. 1977). Thus the *WM* locus seems to regulate the organ specific formation or glycosylation of the flavonols. In summary, the genetics of flavonol glycoside formation in soybean generally support the notion that addition of O-glycosidic bonds to flavonoids occurs after synthesis of the aglycone, and that specific sugar linkages are added in single steps, each conferred by a different gene.

While the biochemistry and genetics of flavonol glycoside metabolism is fairly well established, the accompanying genomic aspects are very limited due to the fact that the cultivars used in the various current EST libraries are quite limited. Although we will not treat the leaf flavonols further in this chapter, this is an important area for further molecular genetic analyses.

Another type of common co-pigments of flowers are the flavones. However, unlike other legumes such as *Medicago truncatula* where various flavones are dominant flavonoid compounds in leaves and roots (Farag et al. 2007), flavones have not been reported in soybean. Consistent with this, the key enzyme for flavone synthesis, flavone synthase (FNS, CYP93B), was not found in the soybean EST genomic database. It is possible that soybean does not have a flavone biosynthesis pathway.

Anthocyanins and Proanthocyanidins

The anthocyanins and proanthocyanidins (condensed tannins) are downstream from the flavonols (Fig. 13.1) and contribute to flower and seed coat color in soybean. In addition, proanthocyanidins are important to protect ruminant animals from pasture bloat (Waghorn and McNabb 2003). There are interesting parallels between the genes controlling the synthesis of anthocyanins and proanthocyanidins in soybeans and those responsible for the pigmentation of the testa in Arabidopsis. The genes for testa pigmentation in Arabidopsis were identified through analysis of a series of transparent testa (*tt*) mutants. As with Arabidopsis, several major genes control the accumulation of flavonols in the seed coat of soybean. These include the *T* gene (Buttery and Buzzel 1973), which as noted already is responsible for the F3′H-mediated conversion of dihydrokaempferol to dihydroquercetin (Zabala and Vodkin 2003). Both of these flavonoids are also precursors for the anthocyanins, although we only show the pathway for cyanidin formation from dihydroquercetin in Fig. 13.1. The *TT7* gene, which has an analogous function in Arabidopsis, was cloned and shown to be a cytochrome P450 that indeed functions as a flavonoid 3′-hydroylase (Schoenbohm et al. 2000). It and a related, though unpublished, Arabidopsis sequence (AF155171) have by far the highest homology (tblastx, 1.4e-166) to a single TC (TC216289) in the soybean EST database and somewhat lower homology to a soybean singleton derived from AY117551. Thus, it is likely, by homology to the Arabidopsis gene that TC216289 or AY117551 may function in the conversion of dihydrokaempferol to dihydroquercetin in the seed coat.

For biosynthesis of proanthocyanidin, the key enzymes are the *Banyuls* homologs that function as anthocyanidin reductase (ACR or BAN, TC220896 and TC226859), and the leucoanthocyanidin reductase (LAR, TC232038). The expression of the

homologs of these two genes was recently investigated in the legume *Lotus corniculatus* (Paolocci et al. 2007), but not in soybean.

The function of another gene that controls anthocyanin accumulation in soybean, the soybean *R* gene, remains unknown although it is thought to function in the anthocyanin pathway before accumulation of the anthocyanins per se (Todd and Vodkin 1993). Finally, in soybean, an extra copy of the chalcone synthase *CHS1* gene, *ICHS1*, at the I (inhibitor) locus (Senda et al. 2002) actually causes suppression of pigment accumulation through homology-dependent gene silencing of CHS (Tuteja et al. 2004; Clough et al. 2004). The presence of this duplicated *CHS1* gene in most commercial soybean lines is responsible for the yellow seed coat of these lines. In homozygous recessive (*i*) lines, the entire seed coat is pigmented, whereas additional *i* alleles restrict pigmentation to specific areas of the seed. CHS is also required for anthocyanin/proanthocyanidin based pigmentation of the testa in Arabidopsis and TT4 encodes a CHS.

Isoflavones and Isoflavone-Derived Metabolites

Due to their pharmacological activities, there has been much interest in recent years in the biosynthesis of isoflavones in soybean seed. However, very early work on the isoflavone metabolic pathways in soybean was greatly stimulated by an interest in accumulation of defense-related metabolites, including the phytoalexins. The soybean-*Phytophthora sojae* association was one of the classical interactions in which early host-pathogen biochemistry focused. This was in part due to the identification of the cell wall glucan elicitor (Ps-WGE) from this oomycetic pathogen (Ayers et al. 1976), an event that not only provided one of the earliest characterized pathogen-derived elicitors, but greatly facilitated biochemical characterization of the pathways stimulated by this elicitor. Pioneering work in several laboratories (see especially work from the Albersheim, Grisebach, Keen, Yoshikawa and Ward labs, reviewed in Ebel 1986) was instrumental in making this system one of the best understood in terms of the metabolic pathways involved and their regulation. More recent work led to more detailed insight into the importance of these pathways in *P. sojae* resistance (for instance, recent work from the labs of Ebel, Hahn, Graham and Yoshikawa, partially reviewed in Hammerschmidt 1999) and their roles in soybean's interactions with the symbiont *Bradyrhizobium japonicum* (reviewed in Stacey et al. 2006).

The major pathways involved in isoflavone and isoflavone-derived metabolism are outlined in Fig. 13.1B. While the presence of isoflavones in soybean seed was known for some time, the finding of isoflavone conjugates as major metabolites in all seedling organs (Graham 1991b) highlighted the importance of the flux of these metabolites into and out of the multiple conjugated forms. In particular, daidzein is a precursor for a variety of other metabolites, including the coumestans (coumestrol, not shown in Fig. 13.1B) and the pterocarpans, including the soybean phytoalexins, the glyceollins (Ebel 1986). For the sake of simplicity, enzymes for the formation

of glycitein, an additional isoflavone, are not shown in Fig. 13.1B. Glycitein contributes 5–15 % of total isoflavone contents in soybean seeds. The enzymes responsible for isoflavone-6-hydroxylation and related methyltransferase have not been identified.

The enzymes for the glyceollin branch pathway starting from the isoflavone daidzein are also shown in Fig. 13.1B. In simple terms, these reactions include a cyclization to form the pterocarpan structure, the addition of an isopentenyl side chain, followed by an additional cyclization to form glyceollin I or II. Of the various enzymes involved, TCs for IF2′H, 2′DHDR, IFR, and DHP6aH can be identified based on homologies to cloned sequences. The genes for THPDMAT and GS have not yet been identified so it is not possible to assign a TC to them.

The cytochrome P450 proteins are important to many of the enzymatic steps outlined in Fig. 13.1B. This is a very complex family of proteins (Ralston and Yu 2006) with many subfamilies based on sequence homology. The soybean P450's have been classified (Kim et al. 2004) and their expression examined in response to various stimuli (Kim et al. 2004) including Ps-WGE (Schopfer and Ebel 1998a). The cytochrome P450's are particularly difficult to work with, due to the very high degree of homology among family members even at the DNA level. Most P450 enzymes are associated with ER-membranes, which also makes them difficult to study. Thus expression and enzymatic analyses are usually needed to rigorously establish the function and identity of family members. As an example, although 9 cytochrome P450's were identified by differential display following treatment with Ps-WGE (Schopfer and Ebel 1998a), only the enzymatic function of two of them, cinnamate 4-hydroxylase (C4H, Schopfer and Ebel 1998a) and dihydroxypterocarpan 6a-hydroxylase (DHP6aH, Schopfer et al. 1998b), were verified by functional expression. So, while it is possible to broadly classify these genes and their products, this is an area where a combination of genomics with in depth biochemistry and techniques such as gene silencing (see, e.g., Subramanian et al. 2005) will be necessary for further functional analysis.

Transcription Factors Regulating the Phenylpropanoid Pathways

Several classes of transcription factors were shown to coordinately regulate biosynthesis in the flavonoid pathways. Among them, Myb-like transcription factors are the most extensively studied (Jin and Martin 1999). The *PAP1* gene of Arabidopsis and *C1* gene of maize are both Myb-like transcription factors that regulate anthocyanin biosynthesis by recognizing consensus *cis*-elements that exist in the promoters of almost all phenylpropanoid structural genes (Borevitz et al. 2000; Martin and Paz-Ares 1997). Ectopic expression of these transcription factors led to specific activation of the pathway and accumulation of anthocyanins in the tissues where the transcription factors are activated. However, Myb-like transcription factors are one of the biggest transcription factor families in plants. It will be difficult to predict the function of these regulators simply based on their sequence

homology (Stracke et al. 2001). There are at least 92 Myb-like transcription factors in soybean EST database. The functions of these Myb genes have not been reported. Furthermore, soybean isoflavonoids and flavonoids clearly have different functions during growth and development. The defense related isoflavonoid compounds may be regulated by a different set of transcription factors than the anthocyanin specific Myb genes. None of these regulators have been identified.

Genomics of Phenylpropanoid Pathways

In addition to GenBank, there are several useful web-based soybean genomic browsers, including one maintained at Southern Illinois University (http://soybean-genome.siu.edu/) and SoyBase at Iowa State University (http://soybase.agron.iastate.edu/). Finally, the public soybean EST and related databases discussed immediately below can be accessed at http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=soybean.

Mining the Soybean EST Database

The public soybean EST database is comprised of over 80 cDNA libraries with over 300,000 expressed sequence tags (ESTs) derived from mRNA expressed in a variety of tissues from various cultivars under a diverse conditions. Private EST databases (DuPont and Monsanto) are also currently available for analysis under a signed agreement. The ESTs within the public database were assembled into tentative contigs (TCs) many of which were assigned possible coding sequences for individual genes (Shoemaker et al. 2003). Information on the ESTs comprising these TCs and many tools for their analysis can be found at the DFCI Soybean Gene Index, formerly the TIGR Soybean Gene Index (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=soybean). Information available includes descriptions of the EST libraries, putative functional annotations for the TCs, etc. A useful tool is provided for mining the EST database to estimate the relative level of "expression" of any particular TC in a given library. This is accomplished electronically by determining the number of ESTs for that TC in the library of interest. For instance, for TC203618 (Chalcone synthase 7), there are 4 ESTs in the Phytophthora infection library, E5T. Since this library has a total of 3524 ESTs, this TC accounts for 0.11 % of the ESTs sequenced in that library. Importantly, there are several considerations that limit the usefulness of such analyses to more global or qualitative interpretation. However, they can be useful to gain an initial overall perspective on the relative importance of different pathways in different libraries and can in some cases lead to the identification of specific family members of a gene for further verification by functional analysis.

Of the genomic resources currently available for soybean, the EST and related databases have been very useful. We used it in this review to "mine" global information on the secondary product enzymes present and then some qualitative

information on their expression under certain conditions. We began by examining the entire public soybean EST database (all libraries) for TCs that putatively encode enzymes in the pathways shown in Fig. 13.1. This was done by keyword searching of the TC annotations as well as independently BLASTing the database (tblastx) with known homologs of key enzymes from GenBank. Generally, we excluded from our final list TCs with less than a 40–50 % stretch of similarity to known genes and/or with a tblastx e-value greater than e −30 to a known gene. The results are grouped by branch pathway and then by specific enzyme in Table 13.1. A preliminary grouping of secondary product TCs from soybean and a number of other species was previously compiled for comparison (Dixon et al. 2002), but the TC numbers for soybean have changed completely from this previous tabulation due to more detailed and accurate sequence alignments.

Nearly all of the enzymes in Fig. 13.1 are represented in Table 13.1. Several specific branch enzymes not included in the simplified pathways in Fig. 13.1 are also listed. While the biochemical function of a few cytochrome P450's were verified (e.g., C4H, Schopfer and Ebel 1998a, and dihydroxypterocarpan 6a-hydroxylase, Schopfer et al. 1998b), due to their complexity and relatively high degree of homology to one another, we simply listed most of the unidentified cytochrome P450's as a separate group. Likewise, the functions of the very diverse peroxidases were not verified, so we again simply listed them a separate class. We also included a class of enzymes in a "Miscellaneous" category. These are enzymes for which the precise fit in the pathways (Fig. 13.1) is unknown. In a few cases (e.g., with the isoflavone synthase family members) we listed specific family members corresponding to a particular TC. This kind of analysis can provide a preliminary road map toward understanding the biosynthetic pathways.

Global Perspectives on Phenylpropanoid Pathways Associated with Various Selected Phenotypes

We next undertook a comparative analysis of the expression of the various genes in Fig. 13.1 (and Table 13.1) by examination of selected key libraries for a few of the phenotypes or events most often associated with various branches of these pathways. Specifically, we focused on libraries corresponding to seed coats, flowers and disease. Due to its potential interest to disease resistance, we also included a salicylic acid response library. Because of the lack of definitive libraries for differentiation of enzymes for the leaf flavonols, we did not examine these. Details of the libraries that we examined in depth are shown in Table 13.2. Although other potentially relevant libraries exist, since library size can limit the usefulness of expression evaluation by EST counts, we rejected any library with fewer than 1,000 ESTs. After deriving complete lists of the TCs for secondary product metabolic enzymes from each of the libraries of interest, we examined EST expression levels for the same TCs in target and "control" (or reference) libraries whenever relevant (e.g., untreated hypocotyls for the *P. sojae* hypocotyl infection libraries, see Table 13.2). After subtraction of

Table 13.1 TCs for soybean phenylpropanoid/miscellaneous secondary product enzymes

Enzyme class	Specific enzyme	Abbreviation	EC number	TCs
Early phenylpropanoid	Phenylalanine ammonia lyase	PAL	4.3.1.5	TC216852, TC225162, TC225165, TC225168
	Cinnamate-4-hydroxylase	C4H	1.14.13.11	TC230826, TC204341
	4-coumarate-CoA ligase	4CL	6.2.1.12	TC205641, TC216083, TC216190, TC219164, TC226636, TC226637, TC227531
Lignin	Caffeoyl-CoA O-methyltransferase	CCOMT	2.1.1.104	TC204613, TC204614, TC221351, TC225339, TC225342, TC225343, TC225345, TC228561, TC228993
	Cinnamoyl-CoA reductase	CCR	1.2.1.44	TC204999, TC208377, TC210404, TC219427, TC225292, TC225293, TC226045, TC226046, TC226919
	Caffeate O-methyltransferase	COMT	2.1.1.68	TC226262, TC226264, TC226266
	Cinnamyl-alcohol dehydrogenase	CAD	1.1.1.195	TC204437, TC204440, TC206457, TC208568, TC215695, TC225589, TC225590, TC225591, TC227396, TC229375, TC229704
	Laccase (Diphenol oxidase)	LAC	1.10.3.2	TC206546, TC209344, TC210109, TC217098, TC217105, TC219245, TC222550, TC229326
Flavonol	Chalcone synthase	CHS	2.3.1.74	TC203259, TC203618 (CHS7), TC210490, TC214207, TC214443 (CHS1), TC216931, TC215169, TC225256

Table 13.1 (continued)

Enzyme class	Specific enzyme	Abbreviation	EC number	TCs
	Chalcone isomerase	CHI	5.5.1.6	TC205496 (CHI4), TC205536 (CHI1a), TC219615, TC216784, TC216785, TC216488, TC216489
	Flavanone 3-hydroxylase	F3H	1.14.11.9	TC205989, TC208325, TC209692, TC226020
	Flavonoid 3′-hydroxylase	F3′H	1.14.13.21	TC211750, TC216289
	Flavonol synthase	FS	1.14.11.23	TC208467
	Flavonol 3-O-glucosyltransferase	F3GT	2.4.1.91	TC208101, TC203723, TC208339, TC229034, TC231164, TC234035
	Chalcone O-methyltransferase	ChOMT	2.1.1.154	TC208485, TC208486
Anthocyanin	Dihydroflavonol-4-reductase	DHF4R	1.1.1.219	TC217923, TC228195
	Leucoanthocyanidin reductase	LAR	1.17.1.3	TC232038
	Leucoanthocyanidin dioxygenase	LCO	1.14.11.19	TC208819, TC217747, TC217748, TC218608, TC220111, TC220906, TC221541, TC232276
	Anthocyanidin reductase	ACR (BAN)		TC220896, TC226559
	Anthocyanin glucosyl transferase	AGT	2.4.1.91	TC234097
Isoflavone	Chalcone reductase	CHR	1.1.1._	TC203399, TC227002, TC229251
	Isoflavone synthase	IFS	1.14.14._	TC215321 (IFS1), TC204612 (IFS2)
	Isoflavone-7-O-methytransferase		2.1.1.150	TC206764, TC206765, TC210019, TC215839
	Trihydroxyisoflavone O-methyltransferase	THIOMT	2.1.1.46	TC216576
	Isoflavone 7-O-glucosyl transferase	IFGT	2.4.1._	TC211462, TC220510, TC231296
Glyceollin	Isoflavone 2′-hydroxylase	I2′H	1.14.13.89	TC204662, TC204663, TC204664, TC204665
	2′-hydroxydihydrodaidzein reductase	2′DHDR	1.1.1._	TC224892, TC226859, TC213000
	Isoflavone reductase	IFR	1.3.1.45	BE611038, TC204692, TC206041, TC207832, TC210161, TC215001, TC215002, TC215003, TC215004, TC215005, TC215632, TC217830, TC226270
	Dihydroxypterocarpan 6a-hydroxylase	DHP6aH		TC227063

Table 13.1 (continued)

Enzyme class	Specific enzyme	Abbreviation	EC number	TCs
Cytochrome P450s	Cytochrome P450	P450	1.14.-.-	TC204664, TC205134, TC205337, TC205410, TC205411, TC206973, TC207603, TC207959, TC209682, TC211282, TC212296, TC216063, TC216663, TC217041, TC219004, TC219972, TC220823, TC221496, TC227063, TC228930, TC229792, TC230633, TC231530, TC231675
Peroxidases	Peroxidase	POX	1.11.1.7	TC203271, TC203397, TC203489, TC204153, TC204257, TC204493, TC204494, TC204496, TC204669, TC204671, TC205158, TC205794, TC205795, TC205894, TC206016, TC206181, TC206280, TC206281, TC206384, TC206605, TC206606, TC207185, TC207747, TC209015, TC209413, TC212050, TC213104, TC214084, TC214085, TC214302, TC214651, TC214657, TC214752, TC214928, TC215014, TC215015, TC215016, TC215017, TC216323, TC216324, TC216834, TC216945, TC217066, TC217976, TC218583, TC218639, TC218640, TC219609, TC221519, TC221869, TC225212, TC225213, TC226619, TC226717, TC226843, TC226844, TC227439, TC227916, TC228199, TC228613, TC229562, TC229710, TC229834, TC230675, TC230907, TC231488, TC232832, TC233091, TC234043, TC221178, TC230884
Miscellaneous	Beta-glucosidase		3.2.1.21	TC205953, TC206214, TC208892
	Hydroxycinnamoyl transferase		2.3.1.110	TC215706, TC215707, TC215709
	Carboxy-lyase		4.1.1.–	TC208106
	cis-trans-Isomerase		5.2.1.–	TC209912
	Oxidoreductase acting on the CH-CH group of donors		1.3.–.–	TC208996

Table 13.2 EST libraries used for comparisons of phenylpropanoid pathways

Library	Total ESTs	Cultivar	Tissue	Age	Treatment	Lab
	Berberine bridge enzyme-like protein			TC222911		
81E (Gm-c1074)	6953	Will 82	Leaves	9–11 day	9–11 day old seedlings that were induced for HR by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glycinea* carrying the *avrB* gene (Genetics 141:1597–1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally.	S. Clough/L. Vodkin
8J2 (Gm-c1076)	4354	Will 82	Cotyledon	11 day	Wounded cotyledons were treated with 2 ugs/ml of a crude glucan elicitor preparation isolated from the mycelial walls of *Phytophthora sojae.* Treated surface tissue was harvested after 24 hr in the dark.	M. Hahn
9IS (Gm-c1084)	4484	Will 82	Hypocotyl	NA	Tissue was inoculated with *Phytophthora sojae* (race 1, incompatible) and tissues were harvested 2 and 4 hrs following infection. The library is the pool of these two time points.	M. Bhattacharyya
E5T	3524	Harosoy	Hypocotyl	NA	Hypocotyls were inoculated with *Phytophthora sojae* culture P6497 (race 2, compatible) and harvested 48 hr following infection.	B. Tyler

Table 13.2 (continued)

Library	Total ESTs	Cultivar	Tissue	Age	Treatment	Lab
2567 (Gm-c1023)	3556	T157	Seed Coats	NA	mRNA was isolated from seed coats (100–200 mgs) of greenhouse grown plants	L. Vodkin, A. Khanna
5649 (GmaxSC)	2051	Harosoy 63	Seed Coats	NA	Library was constructed from polyA+ enriched mRNA from green seed coats in mid to late developmental stage, average fresh weight 250 mg per seed. Traces of pod and embryo tissue also present.	M. Gijzen
GM18 (Gm-c1019)	5385	Williams	Seed Coats	NA	cDNA library was constructed from mRNA isolated from immature seed coats (200–300 mgs) of greenhouse grown plants	L. Vodkin, A. Khanna
6112 (Gm-c1061)	4045	Raiden	Flowers	NA	mRNA isolated from mature flowers of field grown plants	R. Shoemaker
GM16 (Gm-c1016)	8532	Williams 82	Flowers	NA	mRNA isolated from immature flowers of field grown plants	R. Shoemaker, J. Erpelding
GM20 (Gm-c1015)	5334	Williams 82	Flowers	NA	mRNA isolated from mature flowers of field grown plants	R. Shoemaker, J. Erpelding
DE8 (Soybean induced by Salicylic Acid)	28520	Kefeng 1		2 weeks	mRNA isolated from two-week seedlings (cultivar Kefeng 1) treated by spraying 2.0 mM salicylic acid for 24, 36, 48 and 72 hr	S.-Y. Chen
Some reference libraries used						
Gm-c1013	2922	Williams	Whole seedlings	2–3 weeks	mRNA isolated from whole seedlings of 2–3 week old greenhouse grown plants	R. Shoemaker, J. Erpelding

Table 13.2 (continued)

Library	Total ESTs	Cultivar	Tissue	Age	Treatment	Lab
5567 (Gm-c1045)	5384	Will 82	Hypocotyl	9–10 day	mRNA isolated from etiolated hypocotyl tissue of 9–10 day old seedlings of the cultivar Williams 82	R. Shoemaker
GM26 (Gm-c1027)	7571	Williams	Cotyledon	3 and 7 day	mRNA isolated from cotyledons of 3- and 7-day-old Williams seedlings which were propagated on paper towels with distilled water. The cotyledons were flash-frozen in liquid nitrogen, then lyophilized for 72 hrs. Unequal amounts of mRNA was used for cDNA synthesis	P. Keim, V. Coryell
6109 (Gm-c1050)	3393	Clark	Leaf	Various develop stages	mRNA isolated from leaf tissue at various developmental stages of 3 week old greenhouse grown plants	R. Shoemaker, J. Specht, P. Keim

the reference EST expression levels, we finally rejected from these TC lists any TC for which the number of ESTs matching that TC did not comprise at least 0.05 % of the library. By using this percent, we normalized "expression" to library size. With a very few exceptions, we found that enzymes of the phenylpropanoid pathways are highly expressed in the libraries examined, with comparatively very low constitutive expression in control or reference libraries. In fact, expression of a given TC in the libraries chosen for our examinations often accounted for 50–100 % of the total expression for that TC in all soybean EST libraries.

Employing these various limitations, Fig. 13.2A shows a Venn diagram for the presence of secondary metabolic TCs in various global categories of libraries. For this global analysis, we combined the TC lists for each category (e.g. all flower libraries were combined, etc). From this global perspective, only 2 TCs were shared by all libraries, which represents only 4 % of those in *P. sojae* libraries, 7 % of Flower libraries and 17 % of the Seed Coat libraries. While the numbers of TCs involved in these correlations may seem relatively small compared to other genome-wide comparisons, this is a result of limiting the comparisons to enzymes of phenyl-propanoid metabolism and to setting relatively stringent limitations on inclusion of a specific TC.

Using the same limitations for the data, Fig. 13.2B shows a Venn diagram for comparison of three separate *P. sojae* related libraries (combined for the analysis in Fig. 13.2A). As described in Table 13.2, these include two hypocotyl infection libraries (an incompatible infection at 2–4 hr and a compatible infection at 48 hr) and a cotyledon *P. sojae* cell wall glucan elicitor (Ps-WGE) response library. Based on expression analysis of responses to infection or Ps-WGE by many years of HPLC metabolic profiling or mRNA analyses, it was concluded that Ps-WGE is of primary importance in soybean – *P. sojae* interactions (see, e.g., Graham et al. 1990; Graham

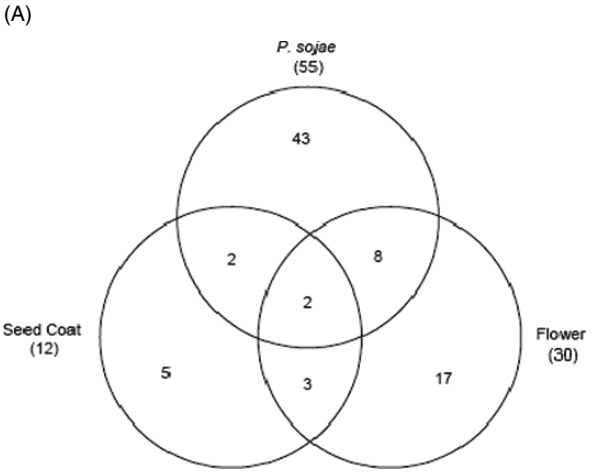

Fig. 13.2A Relationships of secondary product metabolic genes between various broad categories of EST libraries

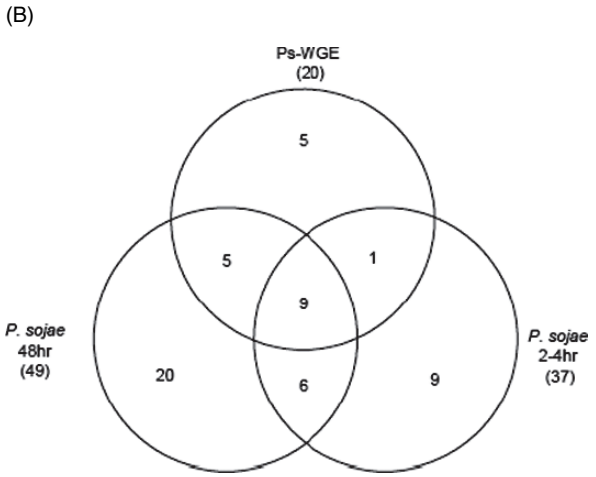

Fig. 13.2B Relationships of secondary product metabolic genes between defense related EST libraries

1991a,Graham and Graham1991b; Graham et al. 2003). Indeed, the majority (15/20) of TCs expressed in the Ps-WGE library are common to either one or the other of the infection libraries. There are 9 TCs common to all libraries, which represents 45 % of the TCs expressed in the Ps-WGE library, 24 % of those in the 2–4 hr *P. sojae* infection library and 18 % of those in the 48 hr *P. sojae* infection library. Thus, there is considerable overlap between the *P. sojae* related libraries, particularly that of the elicitor with the two infections. On the other hand, the two infection libraries share only 15 TCs, amounting to ca. 30–40 % of these two libraries. This is not surprising given the different nature and timing of the two libraries (compatible at 48 hr or incompatible at 2–4 hr). In this same context, however, it is interesting that the compatible library shows as much expression of phenylpropanoid metabolism genes as it does. This may reflect the common observation, in many studies of the phenylpropanoid responses in soybean – *P. sojae* interactions, that the differences in incompatible and compatible infections is more a matter of timing, with the latter expressing key phenylpropanoid responses later, after the infection front has passed through the tissue. It is also consistent with the fact that with more pathogen biomass in the compatible infection, it is likely that there will also be much more Ps-WGE present. These issues cannot be easily addressed from the study of whole organs and is an excellent example of where use of laser capture microdissection techniques will be useful. Recently, we completed microarrays on Ps-WGE treated cotyledon tissues (4 hr post treatment). The resulting data are discussed below in comparison to the TC expression shown in Fig. 13.2B.

Finally, another analysis of global interest is the relationship of infection by *P. sojae* to that by *Pseudomonas syringae* pv. *glycinea* (Psg) and the relationships of both to salicylic acid treatment. Figure 13.2C shows a Venn diagram comparing the TCs expressed in library Gm-C1084 (*P. sojae*, incompatible, 2–4 hr), library

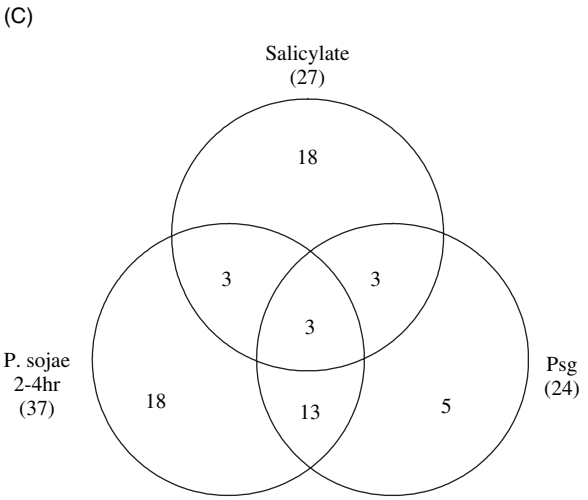

Fig. 13.2C Relationships of secondary product metabolic genes between pathogen and salicylic acid EST libraries

8IE (Psg, incompatible, mixed time points, 2–53 hr) and library DE8 (SA induced). There is little apparent overlap in EST expression in the SA library compared to the two incompatible infection libraries. There are only 3 TCs in common to all three libraries, representing 8 %, 11 %, and 12 % of selected TCs for the *P. sojae*, SA and Psg libraries, respectively. On the other hand the Psg and *P. sojae* infection libraries have 16 TCs in common, accounting for 55 % and 42 % of the Psg and *P. sojae* libraries respectively. Thus, at least in terms of secondary product response genes, there seem to be some similarities between resistance responses induced by infection with either oomycetic or bacterial pathogens. The lack of a clear correlation of SA responses to either infection library is consistent with our long term negative results with SA in terms of activation of phenylpropanoid responses or local or systemic resistance in soybean (T.L. Graham, unpublished). In contrast, ACC oxidase (involved in ethylene generation) and 12-oxophytodienoic acid 10,11-reductase (OPR, involved in jasmonic acid accumulation) are both strongly activated in the various disease resistance libraries, consistent with published conclusions (Creelman et al. 1992; Graham et al. 2003) that these wound signals participate in up-regulation of phenylpropanoid pathway enzymes. These results are further discussed in the comparisons of EST library expression and microarray results described below.

While it is tempting to examine the specific enzymes that contribute to the differences and similarities suggested by the Venn diagrams, due to the limited nature of the data, this is really not feasible. However, we found that examination of differences in the branch pathways for which TCs were present gave some additional and interesting global insights. In Fig. 13.3A, we first compare the classes of metabolic

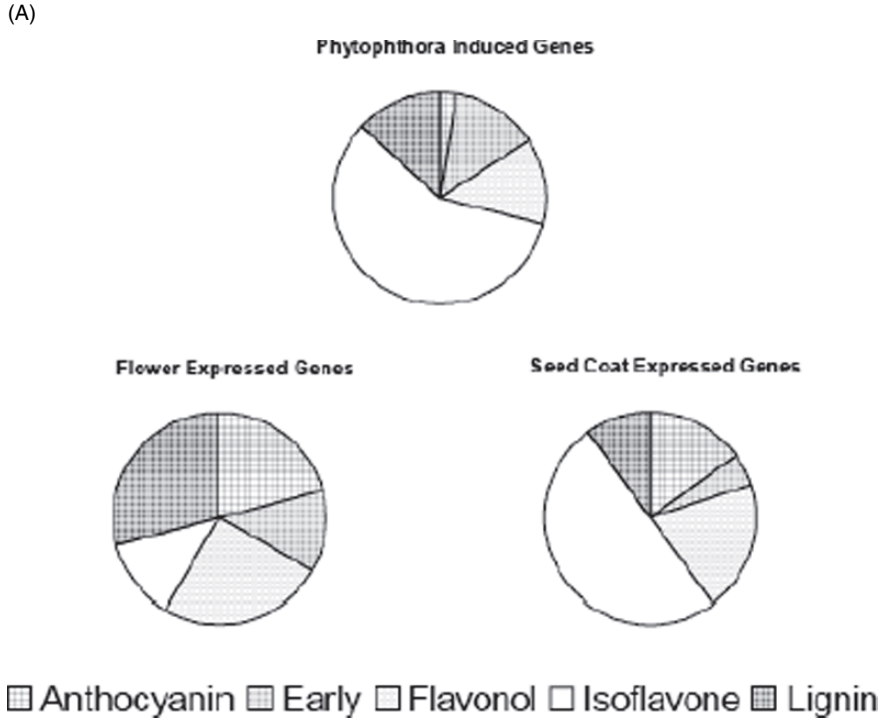

Fig. 13.3A Comparison of metabolic genes expressed in widely different libraries

genes expressed in the various *P. sojae* infection libraries to those in the seed coat and flower libraries. The enzymes for isoflavone metabolism (including the pterocarpans) predominate in the *P. sojae* infection libraries. In the flower libraries the expression of various genes is much more balanced, with a clear and much stronger contribution from the anthocyanin pathway. Although anthocyanin genes also contribute to the seed coat libraries, an interesting finding was that isoflavone enzymes also contribute strongly to these libraries. It is difficult to know if this is actually true or if it is due to slight contamination of this library with tissues from the underlying endosperm, which is known to be producing high levels of isoflavones (Dhaubhadel et al. 2003). Indeed, the description of one of the seed coat libraries (Table 13.2) points out this possibility.

In Fig. 13.3B, the libraries for various defense responses are compared. Isoflavone enzymes predominate for both *P. sojae* and Psg, consistent with the overlap in the Venn diagrams. This is also true for the Ps-WGE library. On the other hand, the SA library has a more balanced distribution, with much lower expression of isoflavone related genes and significant contributions from the lignin and anthocyanin pathways. This mirrors the discussion above based on Venn diagrams and the conclusion that the SA and disease resistance libraries are quite different.

(B)

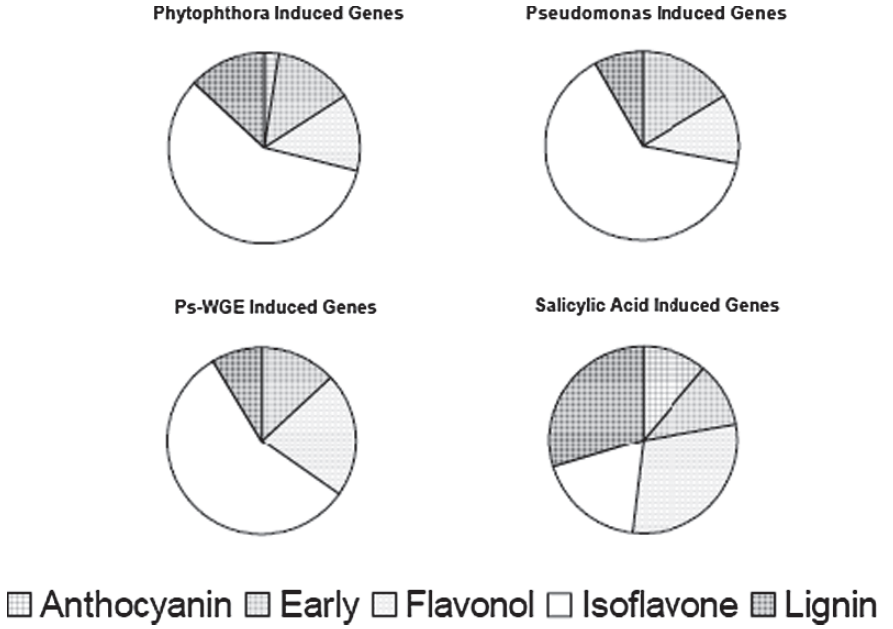

Fig. 13.3B Comparison of metabolic genes induced in defense libraries

Phenylpropanoid Pathways in EST Libraries and Microarray Experiments Associated with Pathogen Resistance

To compare expression in the various *P. sojae* and Psg EST libraries to that in three microarray experiments, we first tabulated the TCs that had 0.1 % EST expression or greater in at least one of these defense EST libraries. We then examined expression for each of these TCs in three microarray experiments: (1) a *P. sojae* hypocotyl infection experiment (Moy et al. 2004) involving a custom array of a smaller number of soybean genes previously seen to be induced in an earlier cDNA library (E5S); (2) an Affymetrix microarray experiment (MY Graham, unpublished) involving a 4 hr treatment of cotyledons by either wounding alone or wounding and WGE elicitor treatment; (3) a microarray experiment involving the unigene re-racked cDNA microarrays and infection by Psg (Zou et al. 2005; Zabala et al. 2006). The TCs chosen for these analyses and information on them is presented in Table 13.3. The results of the comparisons of their expression across the selected libraries and arrays are shown in Table 13.4.

For comparative purposes, Array 1 is somewhat related to both *P. sojae* infection EST libraries. All are hypocotyl infections. The soybean – *P. sojae* EST library E5T is a compatible infection at 48 hr, whereas library 91S is an incompatible infection at 2–4 hr. Array 1 is an incompatible infection over a time frame of 3–48 hr. The data shown in Table 13.3 was derived from the 48 hr data for better comparison to E5T.

Table 13.3 TCs examined in depth in various defense libraries

TC number	Group	Enzyme	Specific information	GB # of Closest homolog	Tblastx e-value	Genus
TC204341	Early	C4H	elicitor induced cinnamate 4-hydroxylase	X92437	0	Glycine
TC216852		PAL	Phenylalanine ammonia-lyase 1	X52953	0	Glycine
TC225162		PAL	Phenylalanine ammonia-lyase (class II)	AB236800	0	Trifolium
TC225165		PAL	Phenylalanine ammonia-lyase (class II)	AB236800	0	Trifolium
TC215695	Lignin	CAD	Cinnamyl-alcohol dehydrogenase.	AF053084	0	Malus
TC225589		CAD	Cinnamyl-alcohol dehydrogenase.	AF053084	8e-169	Malus
TC204999		CCR	Cinnamoyl-CoA reductase	NM_127952	3e-50	Arabidopsis
TC204153		POX	Peroxidase.	AJ011939	5e-174	Trifolium
TC205496	Flavonoid	CHI	Chalcone isomerase 4	AY595417	0	Glycine
TC205536		CHI	Chalcone isomerase 1A	AY595413	0	Glycine
TC203618		CHS	Chalcone synthase 7	M98871	5e-163	Glycine
TC214443		CHS	Chalcone synthase 1	DQ239918	0	Glycine
TC215169		CHS	6′-deoxychalcone synthase	X55730	0	Glycine
TC225256		CHS	6′-deoxychalcone synthase	X55730	2e-145	Glycine
TC224552		SAMS	S-adenosylmethionine synthetase	Z71271	0	Catharanthus
TC215321	Isoflavone	IFS	Isoflavone synthase 1	AF195798	0	Glycine
TC204612		IFS	Isoflavone synthase 2	AF135484	0	Glycine
TC226859	Glyceollin	2′DHDR	2′-hydroxydihydrodaidzein reductase	AJ003246	0	Glycine
TC204663		I2′H	Isoflavone 2′-hydroxylase	AY278227	0	Medicago
TC204664		I2′H	Isoflavone 2′-hydroxylase	DQ340238	0	Glycine
TC204692		IFR	Isoflavone reductase-like protein	AB265592	3e-176	Lotus
TC215001		IFR	Isoflavone reductase homolog 2	AF202183	0	Glycine
TC215002		IFR	Isoflavone reductase	AJ003245	1e-178	Glycine
TC215632		IFR	Isoflavone reductase homolog 1	AF202184	0	Glycine
TC227063		DHP6aH	Dihydropterocarpan 6a-hydroxylase	D83968/Y10492	0	Glycine
TC206973	P450	P450	Elicitor induced P450	Y10493	0	Glycine
TC216063		P450	Elicitor induced P450	Y10491	0	Glycine
TC225031	Signaling	ACCO	1-aminocyclopropane-1-carboxylate oxidase	AB062359	0	Phaseolus
TC225757		OPR	12-oxophytodienoate reductase (OPR2)	NM_106319	0	Arabidopsis

Table 13.4 Expression of selected genes in various defense libraries and arrays*

			Expression in *P. sojae* – related EST libraries			Expression in *P. sojae* microarray (Array 1)	Expression in WGE microarray (Array 2)	Expression in wound microarray (Array 2)	Expression in Psg EST library	Expression in Psg microarray (Array 3)	
TC number	Group	Enzyme	Library E5T	Library 91S	Library 8J2	48 hr time point, 2X or greater	4 hr time point, 3X or greater	4 hr time point, 3X or greater	Library 8IE	I, 24 hr	C, 24 hr
TC204341	Early	C4H	0.23	0.18	0.05		X	X	0.03		
TC216852		PAL	0.06	0.09		X		X			
TC225162		PAL	0.60	0.02	0.07				0.03	2.3	0.0
TC225165		PAL	0.14		0.05				0.01		
TC215695	Lignin	CAD	0.11								
TC225589		CAD	0.11						0.01	2.0	0.0
TC204999		CCR	0.06	0.18				X			
TC204153		POX	0.23	0.22	0.34			X	0.04		
TC205496	Flavonoid	CHI 4	0.03	0.09				X	0.03	7.0	0.6
TC205536		CHI 1a	0.11	0.09	0.05			X	0.01	13.0	6.0
TC203618		CHS 7	0.11	0.02	0.05			X	0.03		
TC214443		CHS 1	0.34	0.42	0.02			X	0.03	17.0	6.0
TC215169		CHS	0.11	0.18	0.02	X			0.04	9.0	4.0
TC225256		CHS	0.20	0.09	0.11				0.09	31.0	3.0
TC224552		SAMS	0.06	0.13	0.09			X	0.09		

Table 13.4 (continued)

TC number	Group	Enzyme	Expression in *P. sojae* – related EST libraries			Expression in *P. sojae* microarray (Array 1)	Expression in WGE microarray (Array 2)	Expression in wound microarray (Array 2)	Expression in Psg EST library	Expression in Psg microarray (Array 3)	
			Library E5T	Library 91S	Library 8J2	48 hr time point, 2X or greater	4 hr time point, 3X or greater	4 hr time point, 3X or greater	Library 8IE	I, 24 hr	C, 24 hr
TC215321	Isoflavone	IFS 1	0.26	0.13	0.07	X		X	0.01	6.0	3.0
TC204612		IFS 2	0.31	0.18	0.14			X	0.09	17	7
TC226859	Glyceollin	2′HDDR	0.11	0.04	0.11	X	X		0.04		
TC204663		I2′H	0.51		0.09		X		0.01		
TC204664		I2′H	0.11		0.02	X	X		0.01		
TC204692		IFR	0.06	0.13							
TC215001		IFR	0.11		0.05		X		0.10		
TC215002		IFR	0.09	0.36	0.05			X	0.06	16.0	0.0
TC215632		IFR	0.14		0.02				0.04		
TC227063		DHP6aH	0.20	0.02	0.02	X	X		0.04		
TC206973	P450	P450	0.14	0.02	0.02		X		0.01		
TC216063		P450	0.17	0.04	0.07	X			0.01		
TC225031	Signaling	ACCO	0.06	0.09	0.18			X	0.24		
TC225757		OPR	0.17	0.2	0.07			X	0.10		

*Values for EST expression libraries represent the percent of the total ESTs for that library which correspond to the specific TC. Values for the Psg microarrays are the fold difference for the treatment versus a mock inoculation (Zabala et al. 2006).

Array 2 is similar in treatment to the Ps-WGE EST library 8J2, except that the time point is much earlier in Array 2 (4 hr compared to 24 hr). In addition, in library 8J2 wounded cotyledon effects were not separated out as they were in Array 2. Array 3 is similar to the Psg infection EST library 81E, except that in the array work both incompatible and compatible infections were examined (Zabala et al. 2006). Array 3 involved sampling over the period 2–53 hr. For the data shown in Table 13.3 for Array 3, we averaged expression in these arrays for each TC at 24 hr.

Although expression of individual genes varies across the various libraries and arrays, an interesting result from the analyses is that nearly all (26/30) of the major genes up-regulated in the various *P. sojae* EST libraries were also up-regulated in at least one of the microarray experiments. Moreover, all of the major pathways (early phenylpropanoid, lignin, flavonoid, isoflavonoid, glyceollin) showed enhanced expression, consistent with these pathways representing the major commitments of phenylpropanoid metabolism in infected tissues (see Venn diagrams and pie charts above). Other enzymes that are up-regulated in at least one library or microarray (data not shown) include homologs of other key enzymes in the early phenylpropanoid (4CL) or lignin pathways (CCR and COMT). That these enzymes were more library/array specific might reflect differences in timing, pathogen and/or the organ examined. The up-regulation of the putative I2′H TCs in library E5T and the Ps-WGE microarray was among the strongest of all genes. Up-regulation of TC 204663 was 0.51 % of E5T and over 500X for the Ps-WGE array. Up-regulation of TC 204664 was somewhat lower, being 0.11 % and over 300X, respectively. This is quite interesting since this is the first committed enzyme for the mobilization of daidzein into the glyceollin branch pathway. It suggests that this critical step is under very strong induction control. Again, it is interesting that there is such strong induction of this enzyme in the compatible infection library, E5T. As discussed previously, this may reflect the possibility that quite a bit of Ps-WGE may be released in the compatible infection (in which more pathogen biomass is present) and the late time point for this library. These observations once again emphasize the need for analysis of more specific tissues than whole organs, for instance those at the infection front.

Confirming the analyses discussed above, salicylic acid was a weak inducer of all of the enzymes shown in Table 13.3. On the other hand, TCs related to two wound induced defense signal molecules, ethylene and jasmonic acid were dominantly induced in many of the libraries or microarrays (Table 13.3, under “Signaling” genes). These included ACCO and OPR, involved in the synthesis of ethylene and jasmonic acid respectively. Interestingly, while there are 22 TCs for ACCO in the soybean EST database, the specific TC for this enzyme induced in the various libraries shown in Table 13.3 appears to be strongly correlated to defense.

Other Relevant Macroarray or Microarray Studies

There are several other studies involving macroarrays or microarrays that describe some general aspects of phenylpropanoid related responses. For the most part these studies do not go into specifics of the phenylpropanoid enzymes up-regulated, but discuss them as part of the general categories of genes that are expressed.

A study of root responses to *Fusarium solani* f. sp. *glycines*, which causes sudden death syndrome (SDS) examined the transcript abundance of 191 ESTs on macroarrays at five time points post inoculation (Iqbal et al. 2005). Root responses of both susceptible (Essex) and partially resistant line RIL23 were compared. The ESTs examined were chosen based on earlier studies from the same lab (Iqbal et al. 2002). Expression of some of the enzymes of early phenylpropanoid metabolism (PAL, C4H, 4CL) and isoflavone metabolism (CHS, IOMT) was noted as being stronger in the partially resistant line, although no quantitative data is presented. However, some interesting specific quantitative analyses of the induction of phenylalanine ammonia lyase (PAL) are described. Much as we described above for the *P. sojae* and Psg libraries, we also undertook an electronic examination of expression of ESTs in the EST libraries for sudden death syndrome [catalog 81C (Gm-c1072) and 81D (Gm-c1073)] corresponding to cultivar PI567374 (partially resistant) and Williams 82 (susceptible) to the disease, respectively. These libraries were constructed from pooled RNA isolated at 4 time points after inoculation. Consistent with the study by Iqbal et al. (2002), expression of phenylpropanoid enzymes (including those of early, lignin, isoflavone and glyceollin pathways) was much higher in the partially resistant line. Interestingly, the pie chart for expression of genes by category (see Fig. 13.3) was quite similar to that for *P. sojae* and Psg (not shown), again suggesting that phenylpropanoid defense responses to a range of pathogens (bacterial, fungal and oomycetic) may involve a similar basic grid of phenylpropanoid enzymes.

Another interesting set of experiments involve microarray analyses of responses of the susceptible cultivar Kent to soybean cyst nematode (Khan et al. 2004; Alkharouf et al. 2005). Although the discussions in these papers are again largely limited to the broad categories of genes, an online resource is available for examination of the microarray data for all ca. 6,000 genes examined. Moreover, microarray data for both resistant and susceptible cultivars are available at the website: http://psi081.ba.ars.usda.gov/SGMD/Publications/OLAP/. There are also soybean EST libraries, though relatively small, for soybean cyst nematode interactions that can be explored (Catalog #'s 9NO, C1G, C1H, C1I).

There are also a series of EST libraries for soybean interactions with *Bradyrhizobium japonicum* that could be mined for interesting data on phenylpropanoid metabolic enzymes. These include libraries for both the cultivar Bragg and its supernodulating mutant (Catalog #s 9DO, 9DP, 9F9, 9G3, GM28).

Recently, a transcriptome analysis following soybean embryo development was carried out using a set of soybean cDNA microarrays (Dhaubhadel et al. 2007). By selecting two cultivars different in isoflavone content, it was discovered that the expression patterns of key isoflavonoid biosynthesis enzymes, such as PAL and IFS were higher at 70 day after pollination in both the cultivars. Therefore, expression of these genes coincides with the onset of isoflavone accumulation. The most interesting discovery was that only CHS7 and CHS8 are highly expressed in the high isoflavone containing lines, suggesting that these two genes may represent the rate limiting enzyme of the pathway.

Finally, an interesting observation from a microarray analysis of soybean leaves grown under elevated carbon dioxide was that two enzymes required for lignin

formation, caffeic acid methyl transferase and cinnamoyl-CoA reductase are both moderately up-regulated under high carbon dioxide concentrations (Ainsworth et al. 2006).

Quantitative Trait Loci Associated with Isoflavones

The total isoflavone levels in soybean are clearly controlled by multiple factors. Several groups have identified soybean quantitative trait loci (QTL) for isoflavone levels. Kassem et al. (2004) established more than 100 recombinant inbred lines using the high isoflavone cultivar Essex with the low isoflavone cultivar Forrest. In two separate reports with a total of 390 markers, the group identified 8 QTL markers that relate to the levels of isoflavones (Meksem et al. 2001). Unfortunately, earlier soybean genetic maps and physical maps were notoriously difficult to match. Later, QTL mapping of isoflavone content using a different soybean population (AC756 X RCAT Angora) was reported (Primomo et al. 2005). These QTL markers will be useful for molecular breeding, but there are still major challenges to derive the underlying genes that control isoflavone levels.

Conclusions and Future Perspectives

Our understanding of the genomics of secondary product metabolism in soybean is still in its infancy. Here, we provide a few interesting analyses and perspectives on the data that can currently be mined from the EST and related databases and a few microarray experiments. Much of the promise of the genomic analysis of secondary product pathways in soybean lies in future work and will in many cases (e.g., particularly with the cytochrome P450's and peroxidases) require the careful integration of complementary information from microarrays, laser capture microdissection, biochemical verification of enzymatic function, and the complementary use of metabolomics with emerging tools such as gene silencing (see, e.g., Subramanian et al. 2005).

References

Ainsworth, E. A., Rogers A., Vodkin, L.O., Walter, A., and Schurr, U. (2006) The effects of elevated CO2 concentration on soybean gene expression. An analysis of growing and mature leaves. Plant Physiology 142, 135–147.

Alkharouf, N.W., Jamison, D. C. and Matthews B.F. (2005) Online analytical processing (OLAP): a fast and effective data mining tool for gene expression databases. J Biomed. Biotechnol. 2005(2): 181–188.

Ayers A.R., Ebel J., Valent B.S., Albersheim P. (1976) Host-pathogen interactions. X. Fractionation and biological activity of an elicitor isolated from the mycelial walls of *Phytophthora megasperma* var. *sojae*. Plant Physiol. 57, 760–765.

Boerjan, W., Ralph, J., Baucher, M. (2003) Lignin biosynthesis. Annual Review of Plant Biology 54, 519–546.

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell. 2000 Dec; 12, 2383–2394.

Buttery, B.R. and Buzzell, R. I. (1973). Varietal differences in leaf flavonoids of soybeans. Crop Sci. 13, 103–106.

Buttery, B.R. and Buzzell, R. I. (1975) Soybean flavonol glylcosides: identification and biochemical genetics. Can. J. Bot. 53, 219–224.

Buttery, B. R. and Buzzell, R. I. (1987) Leaf traits associated with flavonol glycoside genes in soybean. Plant Physiol. 85, 20–21.

Buzzell, R. I., Buttery, B. R. and Bernard, R. L. (1977) Inheritance and linkage of a magenta flower gene in soybean. Can. J. Genet. Cytol. 19, 749–751.

Buzzell R. I., Buttery B. R., Mactavish D.C. (1987) biochemical genetics of black pigmentation of soybean seed. Journal of Heredity 78, 53–54.

Clough, S. J., Tuteja, J. H., Li, M., Marek, L. F., Shoemaker, R. C., Vodkin, L. O. (2004) Features of a 103-kb gene-rich region in soybean include an inverted perfect repeat cluster of CHS genes comprising the I locus. Genome 47, 819–831

Creelman R.A., Tierney M.L., Mullet J.E. (1992) Jasmonic Acid/Methyl Jasmonate Accumulate in Wounded Soybean Hypocotyls and Modulate Wound Gene Expression. Proc. Natl. Acad. Sci. 89, 4938–4941.

Dhaubhadel, S., McGarvey, B. D., Williams, R., Gijzen, M. (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Molecular Biology 53,733–743.

Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M. (2007) Transcriptome Analysis Reveals a Critical Role of CHS7 and CHS8 Genes for Isoflavonoid Synthesis in Soybean Seeds. Plant Physiology 143, 326–338.

Dixon, R. A., Achnine L., Kota, P., Liu, C-J., Reddy, M. S. S., Wang, L. (2002) The phenylpropanoid pathway and plant defence: A genomics perspective. Molecular Plant Pathology 3, 371–390.

Ebel J. (1986) Phytoalexin synthesis: the biochemical analysis of the induction process. Ann. Rev. Phytopathology 24, 235–264.

Farag, M. A., Huhman, D., V., Lei, Z., Sumner, L. W. (2007) Metabolic profiling and systematic identification of flavonoids and isoflavonoids in roots and cell suspension cultures of Medicago truncatula using HPLC-UV-ESI-MS and GC-MS. Phytochemistry 68, 342–354.

Gavnholt, B., Larsen, K. (2002) Molecular biology of plant laccases in relation to lignin formation. Physiologia Plantarum 116, 273–280.

Graham M.Y., Graham T.L. (1991a) Rapid accumulation of anionic peroxidases and phenolic polymers in soybean cotyledon tissues following treatment with *Phytophthora megasperma* f. sp. *glycinea* wall glucan. Plant Physiol. 97, 1445–1455.

Graham M.Y., Weidner J., Wheeler K., Pelow M.L., Graham T.L. (2003) Induced expression of pathogenesis-related protein genes in soybean by wounding and the *Phytophthora sojae* cell wall glucan elicitor Physiol. Molec. Plant Pathol. 63,141–149.

Graham T.L. (1991a) A rapid, high resolution HPLC profiling procedure for plant and microbial aromatic secondary metabolites. Plant Physiol. 95, 584–593.

Graham T.L. (1991b) Flavonoid and isoflavonoid distribution in developing soybean seedling tissues and in seed and root exudates. Plant Physiology 95, 594–604.

Graham T.L., Graham M.Y. (1991b) Glyceollin elicitors induce major but distinctly different shifts in isoflavonoid metabolism in proximal and distal soybean cell populations. Mol. Plant Microbe Interact. 4, 60–68.

Graham T.L., Kim J.E., Graham M.Y. (1990) Role of constitutive isoflavone conjugates in the accumulation of glyceollin in soybean infected with *Phytophthora megasperma*. Molec. Plant Microbe Interact. 3, 157–166.

Hahlbrock, K. (1981) Flavonoids In: P. K. Stumpf and E. E. Conn (eds.) *The Biochemistry of Plants,* Vol 7, *Secondary Plant Products*, Academic Press, NY, pp. 425–456.

Hammerschmidt, R. (1999) Phytoalexins: What Have We Learned After 60 Years? Annual Review of Phytopathology. 37, 285–306.

Humphreys, J.M., Chapple, C. (2002) Rewriting the lignin roadmap. Current Opinion in Plant Biology 5, 224–9.

Iqbal, M.J., Yaegashi, S., Njiti, V. N., Ahsan, R., Cryder, K. L. and Lightfoot, D. A. (2002). Resistance locus pyramids alter transcript abundance in soybean roots inoculated with *Fusarium solani* f.sp. *glycines*. Molecular Genetics and Genomics 268, 407–417.

Iqbal, M. J., Yaegashi, S., Ahsan, R., Shopinski, K. L., Lightfoot, D. A. (2005). Root response to Fusarium solani f. sp. glycines: temporal accumulation of transcripts in partially resistant and susceptible soybean. Theoretical and Applied Genetics 110, 1429–1438.

Jin H, Martin C. Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol. 1999; 41, 577–585.

Kassem MA, Meksem K, Iqbal MJ, Njiti VN, Banz WJ, Winters TA, Wood A, Lightfoot DA. (2004) Definition of Soybean Genomic Regions That Control Seed Phytoestrogen Amounts. Journal of Biomedicine & Biotechnology 2004(1), 52–60.

Khan, R., Alkharouf N., Beard H., MacDonald, M., Chouikha, I., Meyer, S., Grefenstette, J., Knap, H. and Matthews, B. (2004) Microarray analysis of gene expression in soybean roots susceptible to the soybean cyst nematode two days post invasion. Journal of Nematology 36, 241–248.

Kim, B.-G., Ko J.H., Hur, H-G. Ahn, J.-H. (2004) Classification and expression analysis of cytochrome P450 genes from soybean. Agricultural Chemistry and Biotechnology 47, 173–177.

Klink, V.P., Alkharouf, N., MacDonald, M., Matthews, B. (2005) Laser capture microdissection (LCM) and expression analyses of *Glycine max* (soybean) syncytium containing root regions formed by the plant pathogen Heterodera glycines (soybean cyst nematode). Plant Molecular Biology 59, 965–79.

Landini S., Graham M.Y., Graham T.L. (2002) Lactofen induces isoflavone accumulation and glyceollin elicitation competency in soybean. Phytochemistry 62, 865–874.

Landry, L.G., Chapple, C.C., Last, R.L. (1995) Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. Plant Physiology, 109, 1159–1166.

Logemann, E., Tavernaro, A., Schulz, W., Somssich, I. E., and Hahlbrock, K. (2000) UV light selectively coinduces supply pathways from primary metabolism and flavonoid secondary product formation in parsley. Proc. Natl. Acad. Sci. (USA) 97, 1903–1907.

Lozovaya, V. V., Lygin, A. V., Li, S., Hartman, G. L., Widholm, J. M. (2004) Biochemical response of soybean roots to *Fusarium solani* f. sp. *glycines* infection. Crop Science 44, 819–826.

Ma, F., Peterson, C. A. (2003) Current insights into the development, structure, and chemistry of the endodermis and exodermis of roots. Canadian Journal of Botany 81, 405–421.

Martin C, Paz-Ares J. MYB transcription factors in plants. Trends Genet. 1997 Feb; 13, 67–73.

Meksem K, Njiti VN, Banz WJ, Iqbal MJ, Kassem MM, Hyten DL, Yuang J, Winters TA, Lightfoot DA. (2001) Genomic Regions That Underlie Soybean Seed Isoflavone Content. Journal of Biomedicine & Biotechnology 2001 (1), 38–44.

Mohr, P. G. and Cahill, D. M. (2001) Relative roles of glyceollin, lignin and the hypersensitive response and the influence of ABA in compatible and incompatible interactions of soybeans with Phytophthora sojae. Physiological and Molecular Plant Pathology 58, 31–41.

Moy, P., Qutob D., Chapman, B. P., Atkinson, I. and Gijzen, M. (2004) Patterns of gene expression upon infection of soybean plants by Phytophthora sojae. Molecular Plant-Microbe Interactions 17, 1051–1062.

Paolocci F, Robbins MP, Madeo L, Arcioni S, Martens S, Damiani F. (2007) Ectopic Expression of a Basic Helix-Loop-Helix Gene Transactivates Parallel Pathways of Proanthocyanidin Biosynthesis. Structure, Expression Analysis, and Genetic Control of Leucoanthocyanidin 4-Reductase and Anthocyanidin Reductase Genes in Lotus corniculatus. Plant Physiology 143, 504–516.

Primomo, V.S., Poysa, V., Ablett, G.R., Jackson, C.J., Gijzen, M. and Rajcan, I. (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop science, 45, 2454–2464.

Ralston L., Yu O. (2006) Metabolons involving plant cytochrome P450s. Phytochemistry Review 5, 459–472.

Schoenbohm, C., Martens, S., Eder, C., Forkmann, G. and Weisshaar,B. (2000) Identification of the Arabidopsis thaliana flavonoid 3′-hydroxylase gene and functional expression of the encoded P450 enzyme. Biol. Chem. 381, 749–753.

Schopfer, C. R. and Ebel J. (1998a) Identification of elicitor-induced cytochrome P450s of soybean (Glycine max L.) using differential display of mRNA. Molecular and General Genetics 258, 315–322.

Schopfer, C.R., Kochs, G., Lottspeich, F., Ebel, J. (1998b) Molecular characterization and functional expression of dihydroxypterocarpan 6a-hydroxylase, an enzyme specific for pterocarpanoid phytoalexin biosynthesis in soybean (Glycine max L.) FEBS Lett. 432, 182–186.

Senda, M., Jumonji, A., Yumoto, S., Ishikawa, R., Harada, T., Niizeki, M., Akada, S. (2002) Analysis of the duplicated CHS1 gene related to the suppression of the seed coat pigmentation in yellow soybeans. Theoretical and Applied Genetics 104, 1086–1091.

Shoemaker, R. C., Schlueter J. A., Cregan, P., Vodkin, L. (2003) The status of soybean genomics and its role in the development of soybean biotechnologies. AgBioForum 6, 4–7.

Stacey, G., Libault M., Brechenmacher L., Wan, J. and May, G.D. (2006) Genetics and functional genomics of legume nodulation. Current Opinion in Plant Biology 9, 110–121.

Stracke R, Werber M, Weisshaar B. (2001) The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol. 2001 Oct; 4(5), 447–56.

Subramanian S., Graham M.Y., Yu O., Graham T.L. (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae.* Plant Physiol. 137, 1345–1353.

Todd, J. J. and Vodkin, L.O. (1993) Pigmented soybean (Glycine max) seed coats accumulate proanthocyanidins during development. Plant Physiol. 102, 663–670.

Tuteja, J. H., Clough S. J., Chan, W.-C., and Vodkin, L. O. (2004) Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. Plant Cell 16, 819–835.

Waghorn G.C. and McNabb, W.C. (2003) Consequences of plant phenolic compounds for productivity and health of ruminants. Proceedings of the Nutrition Society 62, 383–392.

Zabala, G. and Vodkin L. (2003) Cloning of the pleiotropic T locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3' hydroxylase. Genetics 163, 295–309.

Zabala G., Zhou, J. Jigyasa, T., Gonzalez, D.O., Clough, S.J. and Vodkin, L.O. (2006) Transcriptome changes in the phenylpropanoid pathway of Glycine max in response to Pseudomonas syringae infection. BMC Plant Biology 6:26.

Zou, J., Rodriguez-Zas S., Aldea, M., Li, M. Zhu, J., Gonzalez, D.O., Vodkin, L.O., DeLucia, E. and Clough, S.J. (2005) Expression profiling soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-specific downregulation of photosynthesis. Molecular Plant-Microbe Interactions 18, 1161–1174.