Review

Genome-wide analysis of phenylpropanoid defence pathways

MARINA A. NAOUNKINA, QIAO ZHAO, LINA GALLEGOL-GIRALDO, XINBIN DAI, PATRICK X. ZHAO AND RICHARD A. DIXON*

Plant Biology Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA

SUMMARY

Phenylpropanoids can function as preformed and inducible antimicrobial compounds, as well as signal molecules, in plant-microbe interactions. Since we last reviewed the field 8 years ago, there has been a huge increase in our understanding of the genes of phenylpropanoid biosynthesis and their regulation, brought about largely by advances in genome technology, from whole-genome sequencing to massively parallel gene expression profiling. Here, we present an overview of the biosynthesis and roles of phenylpropanoids in plant defence, together with an analysis of confirmed and predicted phenylpropanoid pathway genes in the sequenced genomes of 11 plant species. Examples are provided of phylogenetic and expression clustering analyses, and the large body of underlying genomic data is provided through a website accessible from the article.

INTRODUCTION

Among their many functions in plants, phenylpropanoid compounds play important roles in resistance to pathogen attack (Dixon et al., 2002). Plant defence compounds have been classified into three main groups, signalling molecules, constitutive or preformed phytoanticipins, and phytoalexins, which are synthesized de novo in response to microbial attack (VanEtten et al., 1994); phenylpropanoid compounds are found among all three groups.

Flavonoids (C6-C3-C6) represent one of the largest classes of phenylpropanoid-derived, plant-specialized metabolites (with an estimated 10,000 structurally different members) (Tahara, 2007). They consist of two main groups: the 2-phenylchromans (flavonoids) and the 3-phenylchromans (isoflavonoids) (Fig. 1). The structural diversity of flavonoids is derived by complex substitution of these carbon skeletons through hydroxylation, glycosylation, methylation, prenylation, etc. The enzymes that catalyse these reactions often belong to large gene families, which can be recognized in expressed sequence tag (EST) and genome datasets through family-specific conserved sequence motifs.

In our earlier review (Dixon et al., 2002), the available knowledge on the genomics of phenylpropanoid metabolism was limited primarily to EST expression data and information from the Arabidopsis and rice genome sequencing projects. During the past 8 years, there has been a major increase in such information, brought about largely by advances in genome technology, from whole-genome sequencing to massively parallel gene expression profiling. Here, we provide an update of the field by first reviewing our current understanding of how phenylpropanoid compounds are synthesized and function in plant defence, followed by a discussion of phenylpropanoid biosynthesis across whole plant genomes, with an emphasis on flavonoids. Because of the massive amount of data now available, we have generated a website (http://bioinfo.noble.org/manuscript-support/mpp/) to present the information in a readily accessible manner. It is important to note that genes identified on the basis of bioinformatics approaches alone require subsequent functional identification by biochemical or genetic approaches; genes with proven functions are annotated as such on the website.

FUNCTIONS OF PHENYLPROPAANOIDS COMPOUNDS IN PLANT DEFENCE

Lignin and lignans

Lignin and lignans share monolignols as common precursors (Figs 1 and 2). Lignification makes the cell wall more resistant to mechanical pressure applied during fungal penetration (Becherer et al., 1999). A lignified cell wall is water resistant and thus less accessible to cell wall-degrading enzymes from plant pathogens.

Natural lignin contains three major units: p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S). The structural analysis of lignins induced by elicitor from Rhizosphaera kalkhoffii in spruce revealed that the defence lignins are significantly different from
the lignins that are laid down during plant development (Lange et al., 1995). The specific activation of an S lignin biosynthesis pathway in elicitor-treated wheat leaves was suggested by the induction of a sinapyl alcohol-specific isoform of cinnamyl alcohol dehydrogenase (CAD) (Mitchell et al., 1999). Wheat leaves infected with stem rust fungus or exposed to a derived fungal elicitor subsequently accumulated S lignin (Menden et al., 2007).

The function of lignification during plant defence was originally assumed to be as a physical barrier (Moerschbacher et al., 1990), although lignins or their precursors might also function in chemical defence. Induction of monolignol biosynthesis genes by pathogen infection has been reported in a number of plant species (Buhiyan et al., 2009; Dixon and Paiva, 1995; Koutanomi et al., 2007; Lauvergeat et al., 2001; Menden et al., 2007). For example, 5-hydroxyguaiaeryl 3-O-methyltransferase [also known as caffeic acid 3-O-methyltransferase 1 (COMT1)] expression increased on infiltration of Arabidopsis leaves with flagellin (flg22) or harpin (HrpZ), or with the necrosis-inducing Phytophthora protein 1 (NPP1) (Engelhardt et al., 2009; Fellbrich et al., 2002; Zipfel et al., 2004); thus, transcription of COMT1 appears to correlate with the onset of the plant’s pathogen-associated molecular pattern (PAMP)-triggered immune (PTI) response.

Cinnamoyl-CoA reductase (CCR), the first enzyme specific for monolignol synthesis, was shown to be involved in pathogen defence signalling in rice (Oryza sativa). OsCCR1 is an effector of OsRac1, a small guanosine triphosphatase (GTPase) from the Rac/Rop family, which is known to play a critical role in defence responses and to which OsCCR1 binds through a guanosine triphosphate (GTP)-dependent physical interaction (Kawasaki et al., 2006).

In some cases, the downregulation of monolignol biosynthesis genes not only compromises resistance to host pathogens, but also allows the penetration of nonhost pathogens. For instance,
RNA interference (RNAi)-mediated gene silencing of \textit{COMT} in \textit{wheat} led to successful penetration of \textit{Blumeria graminis} f. sp. \textit{hordei}, a nonhost pathogen, in leaf cells, whereas no penetration was observed in control plants (Bhuiyan \textit{et al.}, 2009). However, the downregulation of lignin biosynthesis genes does not necessarily impair plant defence and, in some cases, reduced lignin plants have increased resistance to plant pathogens (Funnell and Pedersen, 2006). The mechanisms for this are not always clear. \textit{Arabidopsis} \textit{Comt1} mutants accumulated more soluble 2-O-5-hydroxyferulol-L-malate than did wild-type plants, and this, rather than changes in lignin, accounted for the \textit{Comt1} mutant’s enhanced resistance to \textit{Hyaloperonospora arabidopsisidis} (Quentin \textit{et al.}, 2009).

The lignans (Figs 1 and 2) represent a structurally diverse class of plant-specialized metabolites which, unlike lignins, are not found ubiquitously in all land plants. The primary physiological role of lignans in \textit{planta} is probably in defence, and several lignans have been reported to have antibacterial and antifungal activities (Akiyama \textit{et al.}, 2007; Carpinella \textit{et al.}, 2005; Ralph \textit{et al.}, 2006). Lignans can act as both phytoanticipins and phytoalexins. The heartwood of many tree species, such as loblolly pine and western red cedar, contains significant levels of lignans, and constitutive deposition of lignans reinforces durability, longevity and resistance of the heartwood of many tree species against wood-rotting fungi. On the other hand, \textit{de novo} formation of lignans in response to fungal attack has been reported in both woody and nonwoody species (Gang \textit{et al.}, 1999).

**Coumarins**

Scopoletin and its glucosylated form scopolin (Fig. 1), which are induced under various stress conditions (Gachon \textit{et al.}, 2004; Shimizu \textit{et al.}, 2005), have antimicrobial and antioxidative activities and appear to play an important role in disease resistance (Chong \textit{et al.}, 2002). In addition to effects on fungi and bacteria, scopoletin accumulation may also be part of a mechanism for virus restriction in \textit{planta}. Exogenously applied scopoletin inhibits \textit{Tobacco mosaic virus} (TMV) replication in tobacco BY2 protoplasts (Chong \textit{et al.}, 2002), and transgenic tobacco plants overaccumulating scopoletin show enhanced resistance.

---

\textit{Phenylpropanoid defence pathways}

Fig. 2 The monolignol biosynthetic pathway. The enzymes are as follows: CAD, cinnamyl alcohol dehydrogenase; CCaOMT, caffeoyl CoA 3-O-methyl transferase; CCR, cinnamyl CoA reductase; C3H, 4-coumaroyl/3-shikimate 3-hydroxylase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; COMT, caffeic acid O-methyltransferase; DP, dirigent protein; FSH, ferulic acid 5-hydroxylase; HCT, hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase; LAC, laccase; PAL, L-phenylalanine ammonia-lyase; PER, peroxidase; PLR, pinoresinol/lariciresinol reductase; SIRD, secoisolariciresinol dehydrogenase.
against infection with *Potato virus Y* (PVY) (Matros and Mock, 2004).

Scopoletin may be involved in the regulation of reactive oxygen intermediate (ROI) accumulation, one of the earliest changes underlying hypersensitive cell death (Dangl et al., 1996). Tobacco plants expressing an antisense construct to a coumarin-specific glycosyltransferase exhibited reduced levels of scopolin and scopoletin, and showed more intense ROI accumulation relative to control plants after virus infection (Chong et al., 2002).

Flavonoids

Crude extracts from plants rich in flavonoids have been used for centuries to treat microbial diseases. Different compounds may target different components and functions of the bacterial cell, although individual flavonoids may target multiple sites (Cushnie and Lamb, 2005; Tsuchiya and Linuma, 2000). The flavonol quercetin (Fig. 1) binds to the Gyrb subunit of *Escherichia coli* DNA gyrase and inhibits the enzyme’s adenosine triphosphatase (ATPase) activity (Plaper et al., 2003). Quercetin from leaves of *Lotus* has been described as a major antimicrobial compound against five microorganisms, namely *Actinobacillus actinomycetemcomitans*, *Actinomyces viscosus*, *Porphyromonas gingivalis*, *Fusobacterium nucleatum* and *Actinomyces naeslundii* (Li and Xu, 2008).

Some flavonoids inhibit bacterial cytoplasmic membrane function. The antibacterial activity of the prenylated flavanone sophoraflavanone G (SFG, Fig. 1) involves the reduction of bacterial membrane fluidity (Tsuchiya and Linuma, 2000). Catechins (flavan-3-ols, Fig. 1) from tea can damage bacterial membranes by directly penetrating them and disrupting their barrier function. Alternatively, catechins may promote membrane fusion, a process that results in leakage of intramembranous materials and aggregation (Tamba et al., 2007).

Different modifications of flavonoid skeletons have a significant impact on their antimicrobial activity. For example, prenylation generally increases inhibitory activity against bacteria and fungi (Sehn et al., 2004); glycosylation of flavonols from *carnation* increases their activity against *Fusarium oxysporum* (Galeotti et al., 2008); methoxylation drastically decreases the antibacterial activity of flavonoids (Alcaraz et al., 2000).

Pterocarpsans, such as medicarpin from alfalfa (*Medicago sativa*) (Fig. 3), pisatin from pea (*Pisum sativum* L.) (Fig. 1) and maackiain from chickpea (*Cicer arietinum*), constitute a major class of antimicrobial isoflavonoid phytoalexins in legumes. Genetic evidence now supports the hypothesis that pisatin production plays an important role in disease resistance of pea. For instance, pea hairy roots expressing an antisense construct to 6a-hydroxymaackiain O-methyltransferase (HMM, the last enzyme in pisatin biosynthesis) exhibit a large reduction in (+)-pisatin content, and are more susceptible than controls to damage caused by the fungus *Nectria haematococca* (Wu and VanEtten, 2004). Furthermore, detoxification of pisatin by the cytochrome P450 enzyme pisatin dimethylase (PDA) is a virulence factor for *N. haematococca* in pea (Funnell et al., 2002). PDA transcripts are induced in infected pea tissues or by *in vitro* application of pisatin to the fungus, and other genes necessary for pathogenesis on pea display similar temporal patterns to PDA gene expression both *in planta* and *in vitro*, consistent with a coordinated regulation of these genes by pisatin during pathogenesis (Liu et al., 2003b). A Zn(2)-cysteine(6) zinc binuclear DNA-binding protein is involved in pisatin-responsive activation of PDA gene expression (Khan et al., 2003).

**RECENT ADVANCES IN OUR UNDERSTANDING OF PHENYLPROPAANOID BIOSYNTHESIS**

Lignin and lignans

Lignins are synthesized by the oxidative coupling of hydroxycinnamyl alcohol monomers. Although some researchers have proposed that a variety of units can be incorporated into the lignin polymer, and that lignin assembly is essentially combinatorial, this is still a matter of some controversy (Davin et al., 2008; Ralph et al., 2008).

Monolignols share the general phenylpropanoid biosynthetic pathway with other phenolic compounds, such as anthocyanins, condensed tannins and isoflavones (Fig. 2). This pathway includes the entry-point enzyme L-phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H) and 4-coumarate:CoA ligase (4CL). The specific portion of the monolignol synthesis pathway includes the enzymes hydroxycinnamoyl CoA:shikimate hydroxycinnamoyltransferase (HCT), 4-coumaroylshikimate 3-hydroxylase (C3H),CCR, CAD, COMT, caffeoyl CoA 3-O-methyltransferase (CCoAOMT) and ferulic acid 5-hydroxylase (FSH) (Fig. 2). Several reviews have described the monolignol pathway in much detail (Boerjan et al., 2003; Raes et al., 2003). Recent evidence has suggested that simple phenolic acids are derived by reduction of aldehyde intermediates of monolignol biosynthesis; thus, ferulic acid is now believed to originate from the reduction of coniferaldehyde rather than sequential ring modification of cinnamic acid, and sinapic acid may be derived similarly (Nair et al., 2004). Esterification with malate leads to the formation of the sinapate esters found in the seeds and leaves of the Brassicaceae, and recently implicated in the modulation of pathogen reproduction *in planta* (Quentin et al., 2009). After synthesis, monolignols are transported to the cell wall by mechanisms yet unknown. They then undergo oxidative polymerization, catalysed by peroxidases and/or laccases (Gabaládon et al., 2006; Sato et al., 2006). There are 73 peroxidase genes in the Arabidopsis genome (Tognoli et al., 2002), and laccases are
Fig. 3 Biosynthesis of flavonoids and isoflavonoids. The enzymes are as follows: C4H, cinnamate 4-hydroxylase; CHI, chalcone isomerase; CHR, chalcone reductase; CHS, chalcone synthase; 4CL, 4-coumarate:CoA ligase; DMID, 7,2'-dihydroxy-4'-methoxy-isoflavonol dehydratase; F3H, flavone synthase II; 2HID, 2-hydroxyisoflavonone dehydratase; HI4'OMT, 2-hydroxyisoflavonone 4'-O-methyltransferase; IFR, isoflavone reductase; IFS, isoflavone synthase; I2'H, isoflavone 2'-hydroxylase; PAL, L-phenylalanine ammonia-lyase; VR, vestitol reductase.

also encoded by multigene families in plants, making it difficult to identify isoforms that are specifically involved in monolignol oxidation. Some peroxidases can cause ectopic lignification when overexpressed in planta (Elfsstrand et al., 2002). However, solid genetic evidence is still missing to identify unambiguously the specific isoforms of peroxidases and/or laccases involved in lignification.

Because monolignol biosynthesis is a metabolically costly process and its carbon investment is not reversible, the activation of this pathway must be tightly regulated. Many of the genes encoding lignin biosynthetic enzymes, such as PAL, C4H, COMT, CCR and CAD, share a conserved motif known as the AC element, in their promoters that binds lignin pathway regulatory transcription factors (Chen et al., 2000; Hatton et al., 1995; Lauvergeat et al., 2002). A network of transcription factors is responsible for the regulation of secondary wall biosynthesis in Arabidopsis. In this network, two NAC transcription factors, NST1 and NST3, act as a master switch to turn on the entire process of secondary wall biosynthesis, including monolignol biosynthesis (Mitsuda et al., 2007; Zhong et al., 2007, 2008). Two MYB transcription factor genes, MYB58 and MYB63, are directly regulated by NST1 and NST3, and can specifically activate the monolignol biosynthesis pathway (Zhou et al., 2009). Some lignin gene family members are involved in constitutive lignification, whereas other gene family members are only expressed in response to pathogen infection (Lauvergeat et al., 2001). Therefore, it is likely that lignin biosynthesis in vascular tissue development is regulated differentially from that occurring during plant defence responses.

Dimeric lignans result from the stereoselective dirigent protein-directed oxidative coupling by laccases of two coniferyl alcohol units (Davin et al., 1997) (Fig. 2), a process that has been well characterized in flax (Linum usitatissimum) (Ford et al., 2001). In recent years, a large number of dirigent protein genes have been identified in various plant species (Kim et al., 2002; Ralph et al., 2006). Many complex substituted lignan structures exist. In one commonly found pathway, dimerization of two coniferyl alcohol units produces pinoresinol, which is then reduced to lariresinol and secoisolariciresinol by pinoresinol/ lariresinol reductase (PLR), and finally into matairesinol by secoisolariciresinol dehydrogenase (SID) (Hano et al., 2006; von Heimendahl et al., 2005) (Fig. 2). Secoisolariciresinol is
mainly found in its diglucoside form (SDG), and is stored as a hydroxymethyl glutaryl ester-linked complex (Ford et al., 2001).

Although the dirigent protein determines the stereoselectivity of the oxidative free radical coupling reaction, the enantiomeric composition of pinosinol differs in various plants (Umezawa, 2003). However, the fact that some lignans derived from pinosinol, such as matairesinol, are optically pure indicates that PLR and SIRD, which catalyze subsequent metabolic steps, also contribute to the enantiomeric composition of lignans (von Heimendahl et al., 2005; Nakatsubo et al., 2008). PLR normally reduces both pinosinol and lariciresinol efficiently (Hemmati et al., 2007; von Heimendahl et al., 2005). However, the two Arabidopsis PLRs show strict substrate preference towards pinosinol and no activity towards lariciresinol, in contrast with conventional PLRs of other plants (Nakatsubo et al., 2008).

Coumarins

Coumarins share hydroxycinnamoyl CoA intermediates with the monolignol pathway. Recently, the production of scopoletin and scopolin in Arabidopsis roots has been shown to require the activity of C3H; in C3H T-DNA insertion mutants, the content of scopoletin and scopolin decreased to around 3% of that in the wild-type (Kai et al., 2006).

The role of glucosyltransferase that catalyses the glucosylation of scopoletin has been demonstrated by both loss- and gain-of-function approaches in transgenic tobacco. The down-regulation of tobacco scopoletin glucosyltransferase (TGGT) activity was accompanied by decreased virus resistance (Chong et al., 2002). Conversely, the overexpression of TOGT increased virus resistance (Matros and Mock, 2004).

Flavonoids

Chalcone synthase (CHS) catalyses the first committed enzyme of flavonoid biosynthesis (Fig. 3). CHSs are type II polyketide synthases (PKSs) with two independent active sites that catalyse a series of decarbonylation, condensation and cyclization reactions (Jefferson et al., 2002). Structural biology has recently become an important tool for understanding the mechanisms and evolution of enzymes involved in the biosynthesis/modification of phenylpropanoid skeletons (Noel et al., 2005; Ritter and Schulz, 2004; Shao et al., 2005). Guided by the three-dimensional structure of CHS, substitution of the active site phenylalanine-215 (Phe215) by serine (Ser) yielded a mutant CHS that preferentially accepts the CoA-thioester substrate to generate a novel alkaloid, namely N-methylanthraniloyltriacetic acid lactone (Jefferson et al., 2002). Thus, a single point mutation in CHS dramatically shifts the molecular selectivity of this enzyme.

5-Deoxy(iso)flavonoids, which are predominantly found in legumes, have strong antimicrobial properties and also act as signals in plant-microbe interactions. Chalcone reductase (CHR) is the critical enzyme for the 5-deoxyflavanoid pathway, reducing an unstable intermediate of the CHS reaction to ultimately form 6′-deoxychalcone (isoliquiriigenin, Fig. 3).

Chalcone isomerases (CHIs) catalyse the stereospecific isomerization of chalcones into their corresponding flavanones (Shimada et al., 2003). Type I CHIs isomerize only 6′-hydroxychalcone to 5-hydroxyflavanone, and type II CHIs, generally found in legumes, act on both 6′-deoxychalcone and 6′-hydroxychalcone, yielding 5-deoxyflavanone and 5′-hydroxyflavanone, respectively (Fig. 3).

The conversion of flavanones to isoflavones is catalysed by the isoflavone synthase (IFS), a cytochrome P450 enzyme of the CYP93C subfamily. In the first step, IFS catalyses the 2-hydroxylation of flavanone, followed by 1,2-aryl migration to yield a 2-hydroxyisoflavonone. All the P450s of the CYP93 family so far functionally characterized are involved in flavonoid/isoflavonoid biosynthesis. For example, the CYP93B3 flavanone 2-hydroxylase (Akashi et al., 1998b) and flavone synthase II (Martiens and Forcken, 1999) catalyse the hydroxylation at C-2 or introduction of a double bond between C-2 and C-3 of the flavanone substrate, respectively. CYP93A is a soybean pterocarpan 6a-hydroxylase (Schopfer et al., 1998).

To date, there are no crystal structures available for plant membrane-associated cytochrome P450s. However, molecular modelling studies have suggested that two active site residues of licorice IFS, Ser310 and Iysine375 (Lys375), are critical for the aryl migration of the flavanone substrate (Sawada and Ayabe, 2005; Sawada et al., 2002). The Lys375Thr mutant (Thr, threonine) produced only 3-hydroxyflavanone, a byproduct of the IFS reaction. Thus, introduction of Lys375 was a critical event in the development of the unique catalytic function of IFS which was then fixed in the Leguminosae because of the selective advantage of the presence of isoflavonoids.

Pea, alfalfa, licorice and Lotus produce their isoflavonoid phytoalexins from a 4′-methoxyisoflavanone, formononetin, whereas soybean synthesizes 4′-hydroxy-type isoflavonoid phytoalexins (glyceollins) from daidzein (Fig. 3). The dehydration (1,2-elimination) step from 2-hydroxyisoflavonone to isoflavone is now known to be enzyme-catalysed rather than a spontaneous process (Akashi et al., 2005; Hakamatsuka et al., 1998). Licorice 2-hydroxyisoflavonone dehydratase (HID) is specific for 2,7-dihydroxy-4′-methoxyisoflavonone, whereas soybean HID has broader specificity for both 4′-hydroxylated and 4′-methoxylated 2-hydroxyisoflavonones, reflecting the structures of the isoflavones in each species. Site-directed mutagenesis studies and phylogenetic analysis revealed that HID proteins evolved from the plant carboxylesterase family (Akashi et al., 2005). HID appears to be critical for isoflavone biosynthesis in hairy root cultures of Lotus japonicus (Shimada et al., 2007). In contrast,
Isoflavones are formed in Arabidopsis expressing IFS in the absence of HID (Liu et al., 2002).

Hydroxylation of flvanoid skeletons by NADPH-dependent cytochrome P450 monoxygenases is important in the biosynthesis of complex flavonoids. Hydroxylation at the 2′ or 3′ position of the B-ring is essential for the formation of pterocarpanes or coumestrols (Guo et al., 1994; Hinderer et al., 1987). 6a-Hydroxylation of pterocarpan substrates occurs in the biosynthesis of the glyceollins in soybean and of pisatin in pea. The first isoflavone 2′-hydroxylase (I2':1; CYP81E1) was cloned and characterized from licorice (Akashi et al., 1998a). Subsequently, three CYP81E subfamily members were identified in Medicago truncatula by mining of EST databases (Liu et al., 2003a). Two of the corresponding enzymes, when expressed in yeast, utilized the same isoflavone substrates, but produced different products hydroxylated at the 2′ and/or 3′ positions of the B-ring. Unique tissue-specific expression patterns suggested the involvement of these enzymes in different defence processes in Medicago (Liu et al., 2003a).

Isoflavone reductase (IFR) is a key enzyme in the latter part of the pterocarpan phytoalexin biosynthetic pathway (Paiva et al., 1991). In species producing medicarpin, it catalyses the NADPH-dependent reduction of 2′-hydroxyformononetin to the isoflavone (3R)-vestitone. An engineered IFR lacking residues 39–47, which constitute a flexible region, was crystallized, and the structure of Δ39–47 IFR was shown to comprise two domains (N- and C-terminal) which form a cleft in which both NADPH and 2′-hydroxyformononetin are presumed to bind (Wang et al., 2006). The study provides a structural basis for understanding the enzymatic mechanism and substrate specificity of IFRs, as well as the functions of IFR-like proteins (Wang et al., 2006).

Enzymatic O-methylation of (iso)flavonoids is catalysed by O-methyltransferases (OMTs), which transfer a methyl group from S-adenosyl-L-methionine (SAM) to a hydroxyl moiety of the acceptor molecule. Two different groups of small-molecule OMTs can be distinguished from plants: proteins of group I do not require a metal ion for activity, whereas OMTs of group II are Mg2+ dependent (Noel et al., 2003). 4′-O-Methylation of the isoflavonoid B-ring by a SAM-dependent 2′-hydroxyisoflavonone 4′-O-methyltransferase (H14′OMT) is a critical step at the entry point into the formation of isoflavonoid phytoalexins in pea, alfalfa, M. truncatula and chickpea (Akashi et al., 2003; Dixon, 1999). However, among these four species, only pea catalyses A-ring 3-O-methylation that converts the pterocarpan 6a-hydoxymaackiaein to pisatin (Fig. 1). The H14′OMT from M. truncatula shares 90% amino acid sequence identity with (+)-6a-hydoxymaackiaein 3-O-methyltransferase from pea, which is responsible for the final step in the biosynthesis of (+)-pisatin. Although (+)-6a-hydoxymaackiaein is not synthesized in M. truncatula, cell suspension cultures accumulate pisatin when fed with 6a-hydoxymaackiaein (Liu et al., 2006). Recombinant H14′OMT of M. truncatula shows comparable activity towards both 2,7,4′-trihydroxyisoflavone and the unnatural substrate (+)-6a-hydoxymaackiaein. Protein X-ray crystal structures of H14′OMT substrate complexes revealed similar bound conformations for the seemingly distinct chemical substrates, explaining how leguminous plants can use homologous enzymes for two different biosynthetic reactions (Liu et al., 2006).

Vestitone reductase (VR) is the central enzyme for the conversion of nonenantiomeric isoflavones to enantiomeric pterocarpan. It is a member of the short-chain dehydrogenase/reductase (SDR) superfamily and catalyses the NADPH-dependent reduction of (3R)-vestitone in the biosynthesis of medicarpin. It acts with high stereospecificity, and recognizes only the 3R conformation of vestitone, not (3S)-vestitone (Shao et al., 2007). A VR homologue, sorhol reductase (SOR), was identified for the biosynthesis of pisatin in pea. SOR prefers (+)-sorhol (Fig. 1) over (+)-sorhol as substrate, converting it to (−)-7,2′,3′,4′-dihydroxy-4′,5′-methylenedioxyisoflavonol for the production of (+)-pisatin (DiCenzo and VanEtten, 2006). Homologues of VR and SOR have recently been described in additional species, including soybean (Glycine max) (Liu, 2009). The three final steps in pisatin biosynthesis are catalysed by SOR, pterocarpan 6a-hydroxylase and (+)-6a-hydroxymaackiaein-3-O-methyltransferase (HMM). The catalytic specificity of two genetically linked HMMs in pea suggested gene duplication for (+)-pisatin biosynthesis (Akashi et al., 2006).

The prenylation of flavonoids enhances their antibacterial, antifungal and other biological activities by increasing their lipophilicity and membrane permeability (Sohn et al., 2004). Recently, Sasaki and colleagues isolated the first cDNA of a flavonoid-specific plant prenyltransferase, SFIN8T1, involved in SFG (Fig. 1) biosynthesis in Sophora flavescens (Sasaki et al., 2008). Phylogenetic analysis revealed that the naringenin dimethylallyltransferase SFIN8T1 has the same evolutionary origin as prenyltransferases for vitamin E and plastoquinone biosynthesis (Sasaki et al., 2008). This discovery facilitated the subsequent identification of the pterocarpan 4-dimethylallyltransferase (G4DT) catalysing the critical prenylation step in glyceollin biosynthesis in soybean (Akashi et al., 2009).

Family I uridine diphosphate (UDP) glycosyltransferases (UGTs) transfer sugars to a wide range of acceptors. Plants contain very large UGT gene families: over 100 putative UGT genes have been identified in Arabidopsis and more than 300 in M. truncatula. Significant progress has been made during recent years in the isolation and structural characterization of (iso)flavonoid UGTs from legume species (He et al., 2006; Modolo et al., 2007; Noguchi et al., 2007; Shao et al., 2005). Most of the identified UGTs to date prefer UDP-glucose as sugar donor. Relationships between the structure and function in M. truncatula.
(iso)flavonoid UGTs have been explored by X-ray crystallography and site-directed mutagenesis (He et al., 2006; Li et al., 2007; Shao et al., 2005).

Genome-wide analysis ofphenylpropanoid metabolism

The last decade has seen remarkable advances in the structural analysis of model plant genomes. Post-sequencing, functional genomics approaches are increasingly being applied to the discovery of new gene functions, and comprehensive web-based software tools have become available for biologists to explore gene expression across a wide variety of biological contexts. Reference expression databases for the meta-analysis of transcriptomes, such as Expression Profiler (http://www.ebi.ac.uk/ expressionprofiler) or the Medicago truncatula Gene Expression Atlas (http://bioinfo.noble.org/gene-atlas/v2), have provided new tools for the analysis of the expression of genes in many different contexts, for the clustering of genes into expression modules, and for the modelling of expression responses in the context of metabolic and regulatory networks.

Because of the exponential increase in plant genome information since the publication of our previous review (Dixon et al., 2002), it is no longer possible to summarize this field in hard copy text and tables. We have therefore created a database of the genes involved in the biosynthesis of phenylpropanoid natural products from 11 plant species with fully or near-fully sequenced genomes; the moss Physcomitrella patens, the spike moss Selaginella moellendorffii, three monocots (Oryza sativa, Sorghum bicolor, Zea mays) and six dicots (Arabidopsis thaliana, Populus trichocarpa, Vitus vinifera, Glycine max, Medicago truncatula and Lotus japonicus). For gene annotation, proteins predicted from genome sequences were searched by BLASTX against the uniprot trembl database (http://www.ebi.ac.uk/uniprot) (e-value < 1e-04), and the top five hits with meaningful annotations were retrieved as annotation. Genes were sorted as representing proteins involved in the various branches of the phenylpropanoid pathway, including flavonoid, isoflavonoid, lignin, etc., based on predicted function. The database is available online (http://bioinfo.noble.org/manuscript-support/mpp/); the information includes the source of genome sequences, enzyme name, gene identification, redundancy, Affymetrix probeset ID, location on chromosome (if possible), length of sequence, five top hit annotations and the sequence itself. It is critical to note that this analysis uses informatics approaches to predict function; annotations should be viewed as tentative unless true biochemical functions have been ascribed to a particular gene. For example, many genes are now being given ‘like’ annotations, such as CCR-like or IFR-like, highlighting the fact that they have sequence identity to functionally characterized genes, but that their protein products have not been shown to catalyse the predicted reaction. Gene products with proven function are indicated in the database and, where appropriate, in the following figures in the text.

Genomic evidence for the evolution of flavonoids as protective compounds

To illustrate how genomic data can be mined to reveal details of phylogenetics, chromosomal organization and expression patterns of relevance to plant defence, we have chosen the flavonoid pathway as an example. It is believed that flavonoids initially evolved as protective compounds against UV irradiation (Koes et al., 1994; Markham, 1988; Shirley, 1996), or as hormone regulators (Stafford, 1991). During evolution, plants adopted flavonoids for many different functions, including defence against pathogens.

Mosses are one of the oldest plant groups that produce flavonoids, including chalcones, flavonols and flavones (Fig. 1) (Markham, 1988). Flowering plants extended the array of flavonoids, also producing aurones, anthocyanidins, proanthocyanidins and isoflavonoids (the latter mainly in legume species). CHS, encoding the initial enzyme of the pathway, exhibits strong sequence similarity to bacterial genes encoding fatty acid PKSs. PKSs are classified into three types according to their architectural configurations (Fischbach and Walsh, 2006); type III (which includes CHS-related enzymes) is present in bacteria, plants and fungi (Austin and Noel, 2003; Funa et al., 2007; Seshime et al., 2005).

Gene duplication followed by sequence divergence has been used as a model to explain how novel biochemical functions have arisen from CHSs. Multiple copies of CHS-related genes are found in many species, including the 11 analysed in the present work (Fig. 4). In general, the numbers of CHS-related genes in the genomes correlate with total genome size (Fig. 4A,B). However, soybean, with double the genome size of M. truncatula, rice and sorghum, contains fewer CHS-related genes (Fig. 4A,B). CHS-related genes are scattered throughout the genomes; however, they form clusters of up to five to seven genes in rice, M. truncatula and soybean that may reflect ongoing duplication events (Fig. 4C). The Zea mays genome (the largest genome analysed in this work) is not fully sequenced, and so we do not yet know how many CHS-related genes exist.

Four residues, cysteine-164 (Cys164), Phe215, histidine-303 (His303) and asparagine-336 (Asn336), are highly conserved in CHS proteins (Fig. S1, see Supporting Information; conserved residues are highlighted by blue rectangles) and define the active site (Ferrer et al., 1999). Stilbene synthases (STSs) are among the closest related enzymes to CHSSs. They likewise condense one coumaroyl- and three malonyl-CoA molecules, but the reaction involves a decarboxylation leading to the formation of resveratrol. STS shows high amino acid sequence identity to
Fig. 4 Distribution of chalcone synthase (CHS)-related genes in multiple sequenced genomes. (A) The number of CHS-related and true CHS genes in the genomes of 11 species. (B) Size of the genomes of the 11 species in (A). (C) Map positions of CHS-related genes on chromosomes of *Medicago truncatula*, *Oryza sativa*, *Arabidopsis thaliana* and *Glycine max*.
CHS; however, it differs in several critical amino acids, such as aspartic acid-255 (Asp255), His266 and leucine-268 (Leu268) in the β1d–β2d region of the CoA-binding tunnel (Ferrer et al., 1999) (see alignment in Fig. S1; critical amino acids highlighted by red rectangles and STSs by yellow rectangles). It has been suggested that STSs evolved from CHSs several times independently, as they group phylogenetically with CHSs from the same or related plants (Tropf et al., 1994). Our analyses indicate that STS distributes into the CHS cluster, but as a separate subcluster (Fig. 5A).

Cloning of CHS from the moss *P. patens* revealed an important transition from the CHSs present in microorganisms to those present in higher plants (Jiang et al., 2006). Of the 12 CHS-related genes found in the *P. patens* genome, 10 are distributed...
into the true CHS cluster, but form a separate subcluster with the
CHS-related genes from Selaginella moellendorffii (Fig. 5A, moss
CHSs coloured green). Moss CHSs showed more than 60% amino acid identity to CHSs from higher plants, although one of the
critical amino acids in the CoA-binding tunnel, Leu268, is
substituted with methionine, which probably does not affect the enzyme
function (Fig. S1). Mosses typically exhibit a simple flavonoid profile, and the physiological significance of the presence
of multiple CHSs in moss requires further investigation.

Although the rice genome contains the largest number of
CHS-related genes, only two encode true CHSs (Fig. 5A). With
a few exceptions, CHSs from mosses, monocots (rice, sorghum and
maize) and legumes form separate subclusters (Fig. 5A; Fig. S2,
see Supporting Information). Importantly, there has been
a notable expansion of CHS genes in the legumes: of 17 CHS-
related genes, 13 are true CHSs in soybean, and only true CHSs
are present in the genomes of M. truncatula and L. japonicus
(Fig. 4A). The presence of multiple copies of CHS in separate
clusters reflects massive duplication events after speciation,
indicative of selective evolutionary pressure in favour of flavonoids in legume species.

CHR originated from the aldo/keto reductases (Bomati et al.,
2005). Sixteen CHR genes are found in L. japonicus, seven in M.
truncatula and 11 in soybean (see database online). CHR-like
genes are also found in the genomes of nonlegumes, but their functions in these species are obscure. Phylogenetic analysis
of putative CHRs with functionally characterized enzymes clearly
places only legume enzymes in the true CHR cluster (Fig. S8).

CHI is also encoded by a multigene family which forms gene
clusters in some legume species. For example, a cluster of genes
coding for type I and II CHIs is found in L. japonicus (Shimada et al., 2003); a cluster of seven genes encoding both types of CHI
is observed at the same locus on chromosome 1 of M. truncata-
ula. In general, more CHL genes are present in the genomes of
legumes than nonlegumes (see database online), probably as a
result of the occurrence of type II CHIs in legumes (Fig. 5C). It
has been suggested that type II CHIs evolved from an ancestral
CHI by gene duplication, and the resultant production of
5-deoxyflavonoids paralleled the establishment of the Fabaceae
(Shimada et al., 2003).

Three later enzymes of flavonoid biosynthesis, flavanone
3’-hydroxylase (F3’H), flavonol synthase (FLS) and anthocyanin
synthase (ANS), are all members of the family of
2-oxoglutarate-iron-dependent oxygenases. They are closely
related by sequence and all catalyse the oxidation of the flavonoid
‘C-ring’; these enzymes also exhibit overlapping substrate and
product specificity (Tambull et al., 2004), making annotation
of their genes confusing. In our database, we placed these
genes in the flavonoid 2-oxoglutarate-dependent dioxygenase
family, with putative annotation of some of them based on
phylogenetic analysis with functionally characterized enzymes
[see database and Fig. S3 (Supporting Information)]

Genes encoding the entry enzyme into isoflavonoid biosyn-
thesis, IFS, are present in low copy number in legume genomes
(see database online); furthermore, in Medicago, only two of the
five are true IFS genes (Fig. 6). Isoflavonoid OMTs are more
abundant: 17, 11 and three IOMT genes were identified in the
genomes of soybean, Medicago and Lotus, respectively (see
database online). A previous phylogenetic analysis (Deavours
et al., 2006) divided the M. truncatula IOMTs into two
distinct clades, with homology to either IOTM enzymes,
methylating the A-ring 7-hydroxyl of the isoflavone daidzein, or
4’OMTs enzymes, methylating the B-ring 4’-hydroxyl of 2,7,4’-
trihydroxyisoflavone.
Expression profiling of flavonoid defence responses

To evaluate the involvement of flavonoid pathway genes in defence responses, we downloaded Affymetrix microarray data specifically related to stress or pathogen studies. At present, Affymetrix microarray chips have been developed for 12 plant species: Arabidopsis, barley, citrus, cotton, grape, maize, Medicago, poplar, rice, soybean, tomato and wheat (http://www.affymetrix.com). Numerous microarray experiments are available for Arabidopsis, rice, Medicago and soybean, and we therefore selected these species for further analysis. Microarray data for rice and soybean were obtained from Expression Profiler (EP, http://www.ebi.ac.uk/expressionprofiler), a web-based platform for microarray gene expression and other functional genomics-related data analysis (Kapushesky et al., 2004). Arabidopsis and Medicago expression data were obtained from AtGenExpress (http://www.weigelworld.org/resources/microarray/AtGenExpress) and Gene Expression Atlas database version 2 (MtGEAtv2, http://bioinfo.noble.org/gene-atlas/v2/), respectively. Affymetrix probesets were mapped to the corresponding genes at individual Affymetrix probe level, using a position-weighted scoring index in which mismatches near the middle of a probe were most heavily penalized as follows: [1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,2,2,2,2,2,1,1,1,1,1] for each 25-mer Affymetrix probe. Therefore, a perfect match for a 25-mer probe yields a score of 45. Matches were declared when at least eight of 11 total probe sets had scores of 43 or higher for each gene (see database online). For better visualization of differential gene expression in response to pathogen or stress, treatment/control fold-change values were calculated. The fold-change values were converted into a log2 scale (see database online) and data were subjected to clustering analysis.

As summarized in Fig. 7 and Figs S4 and S5 (see Supporting Information), different numbers of flavonoid pathway genes responded to pathogen or elicitor treatment in each species. In Arabidopsis, putative ANS (At2g38240) was highly induced by Pseudomonas and Phytophthora treatments, whereas expression of functionally characterized CHS (At5g13930; Shirley et al., 1995) was reduced in most experiments (Fig. 7A). The weak response of flavonoid pathway genes to infection in Arabidopsis is consistent with the deployment of indole (i.e. nonflavonoid) phytoalexins in this species (Rogers et al., 1996). In rice, F3′H (Os04g49194) was highly expressed in response to infection with Xanthomonas oryzae pv. oryzae, and a cluster of genes, including IFR-Likes, DFRs and some CHS-Likes, was expressed in response to infection by the blast fungus Magnaporthe oryzae; CHS (Os11g32650) was only slightly induced under these conditions (two-fold change) (Fig. S4).

In soybean, eight of the 10 available CHS genes were induced either in response to cyst nematode infection (experiment accession: E-MEXP-808) or to Asian soybean rust (Fig. S5; accession: E-TABM-230) (van de Mortel et al., 2007). Many genes involved in isoflavonoid biosynthesis in soybean, including three CHRs, CHS, two II′HS, seven IFRs, two IFSs and eight IDMTs, were induced (Fig. S5). Notably, the expression patterns of at least seven CHSs tightly correlated with those of isoflavone biosynthetic genes (P value less than 0.05, bottom subcluster in the dendrogram highlighted by the red rectangle in Fig. S5), suggesting their nonfunctionalization via cis-regulatory evolution into isoflavonoid biosynthesis. It should be noted, however, that a similar set of flavonoid genes is activated in the susceptible soybean genotype EM (Embrapa-48) and in the Rpp2-resistant genotype PI (P230970) in response to fungal infection, suggesting the involvement of flavonoids in basal soybean defences (Fig. S5).

As in soybean, many flavonoid genes are induced in Medicago by fungal infection (Phytophthora cinnamomami) or elicitors [methyl jasmonate, which mimics wound signals, and yeast elicitor (YE), which mimics pathogen signals] (Fig. 7B). In contrast, only two of this set of genes, IFS-Like (Medtr7g111290) and isoflavone 3′-hydroxylase (I′3′H, GB 33521519, CYP81E9; Liu
Fig. 7 Hierarchical clustering analysis of expression patterns of flavonoid biosynthesis genes in microarray experiments involving exposure of plants to pathogens or pathogen-derived elicitors. (A) Arabidopsis thaliana; (B) Medicago truncatula. Microarray data were obtained for A. thaliana at AtGenExpress (http://www.weigelworld.org/resources/microarray/AtGenExpress/) and for M. truncatula from MtGEAv2 (http://bioinfo.noble.org/gene-atlas/v2/). Treatment/control fold-change values were calculated, converted into a log2 scale (see website online) and subjected to hierarchical cluster analysis (pvclust; Suzuki and Shimodaira, 2006). Clusters with approximately unbiased (AU) P value probability of more than 95% are indicated by the red rectangles. Functionally characterized genes are underlined in blue; their National Center for Biotechnology Information (NCBI) gene accession numbers and primary literature citations are available at http://bioinfo.noble.org/manuscript-support/mpp/.
et al., 2003a), were induced in response to colonization by a symbiotic mycorrhizal fungus. Twelve of the 17 available Medicago CHS genes were induced by either Phymatotrichopsis omnivora or YE (Fig. 7B). Expansion of the legume CHS gene family is directly related to the functional diversity of flavonoids, in this case the evolution of isoflavonoids as protective phytoalexins. Individual CHS genes in legumes often exhibit differential expression patterns that can be correlated with the types of flavonoid compounds being synthesized (Dixon and Paiva, 1995).

Future perspectives

Functional genomics approaches are powerful tools to facilitate the understanding of secondary metabolism in plants. Enzymes involved in the same function will be temporally and spatially coexpressed, and clustering analysis of transcript profiles, as described above, can be used as a discovery tool to define gene networks associated with plant defence responses. Metabolomics is fast developing as a powerful tool for the analysis of gene function and natural genetic variation (Dixon et al., 2006; Fridman and Pichersky, 2005; Morrell et al., 2006). Combining transcript and metabolite profiles (either global or targeted) is even more informative, and is becoming an important new approach for the identification of candidate genes in complex plant secondary metabolic pathways (Saito et al., 2008; Shulaev et al., 2008; Yonekura-Sakakibara et al., 2007, 2008).

Genes encoding enzymes functioning in secondary metabolism are generally more divergent than those involved in primary metabolism. Although genes for most metabolic pathways in plants are not organized into gene clusters, a small but increasing number of ‘operon-like’ gene clusters have been identified for the synthesis of plant defence compounds (Field and Osbourn, 2008; Frey et al., 1997; Qi et al., 2004; Shimura et al., 2007), although to date these are mainly in the area of terpenoid metabolism in monocots (oats and rice), and have yet to be seen for consecutive functional enzymes of flavonoid biosynthesis. Investigation of the genomic organization of metabolic pathways may, nevertheless, be a useful additional approach to ascribe potential function to candidate pathway genes, as well as to shed further light on the evolution of chemical diversification in plants.

As shown by the above examples, genomic approaches have huge potential to aid our understanding of how plants have elaborated various chemical defence mechanisms against pathogen attack during evolution. They do not, however, prove causality, which will still require careful biochemical, genetic and physiological analysis of the plant–pathogen interaction. It is a sobering thought that the almost exponential increase in genomic information since our earlier review is likely to accelerate over the next 5-10 years, and data handling and visualisation may then become the major challenges to the biologist. The effort will be worth it, however, as being able to visualize plant–microbe interactions at the level of the whole genomes (and metabolomes) of both organisms will open up unprecedented opportunities to understand metabolic and regulatory cross-talk and the evolution of defence mechanisms, and to develop novel antimicrobial strategies.

REFERENCES


© 2010 THE AUTHORS


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Database: http://bioinfo.noble.org/manuscript-support/mpp/ Fig. S1 Alignment of the active site defined by Ferrer et al. (1999) of chalcone synthase (CHS)-related enzymes from 11 species. Fig. S2 Detailed phylogenetic tree of chalcone synthase (CHS)-related genes. Fig. S3 Phylogenetic analysis of flavonoid 2-oxoglutarate-iron-dependent oxygenases. Fig. S4 Hierarchical clustering analysis of expression patterns of flavonoid biosynthesis genes in microarray experiments involving exposure of rice to pathogens. Fig. S5 Hierarchical clustering analysis of expression patterns of flavonoid biosynthesis genes in microarray experiments involving exposure of soybean to pathogens.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.