View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Wageningen University & Research Publications

Combining Genomics and Metabolomics for the Discovery of Regulatory Genes and Their Use in Metabolic Engineering to Produce 'Healthy Foods'

C. Martin[1], Jie Luo[1], B. Lebouteiller[1], H.P. Mock[2], A. Matros[2], S. Peterek[2], E.G.W.M. Schijlen[3], R. Hall[3], L. Shintu[4], I. Colquhoun[4], B. Weisshaar[5] and E. Butelli[1]

[1] John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom

[2] Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstrasse 3, D-06466 Gatersleben, Germany

[3] Plant Research International, Business Unit Bioscience, PO Box 16, 6700 AA Wageningen, The Netherlands

[4] Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, United Kingdom

[5] Department of Genome Research, University of Bielefeld, D-33594 Bielefeld, Germany

Keywords: metabolic engineering, antioxidants, secondary metabolites

Abstract

Plants often accumulate their natural products to relatively low levels, so there is a lot of interest in breeding or engineering plants that produce higher levels. It has been shown that the most effective way to increase the accumulation of secondary metabolites is to increase the activity of genes that regulate the activity of the biosynthetic pathways that make different natural products. Regulatory genes of this type encode proteins called transcription factors. The biggest bottleneck in using this strategy to develop plants that accumulate significantly higher levels of important natural products is that not many transcription factors regulating secondary metabolism have yet been identified at the molecular level. Genes encoding transcription factors can be identified from model plants with sequenced genomes. The ability of such genes to regulate metabolism can be assayed by examination of mutants (reverse genetics) and by investigating the metabolic effects of high levels of expression of the genes. The combined techniques of metabolic fingerprinting and metabolite profiling of mutant and transgenic plants are allowing us to identify new genes encoding transcription factors controlling secondary metabolism, that can be used as tools for engineering natural product accumulation.

INTRODUCTION

Plants produce a very broad array of metabolites, which are not essential for growth, but which are used to provide protection against stress and pathogens, to attract pollinators and dispersal agents and as signals for development. These are often referred to as 'secondary metabolites' but are known more generally as plant 'natural products'. Natural products have recently become recognised as important components of the diet, offering protection against cardiovascular diseases, certain cancers and age-related degenerative diseases. They are also important components of beauty products and natural remedies for diseases.

Natural products are often made at low levels and so there is interest in breeding or engineering plants that produce higher levels. The most effective way to increase the accumulation of secondary metabolites is to increase the activity of genes that regulate the activity of their biosynthetic pathways. Regulatory genes of this type encode proteins called transcription factors.

Transcription factors are proteins that modify the expression of target genes by binding to cis-acting, regulatory DNA motifs within their target loci. Eukaryotic transcription factors fall into different classes of protein on the basis of containing conserved domains involved in protein-DNA binding or protein-protein interaction. Despite the fact that, for the majority of plant transcription factors, target genes remain

Proc. 1st IS on Genetic Modifications - Challenges and Opportunities for Horticulture in the World
Ed.: A.K. Hvoslef-Eide
Acta Hort. 941, ISHS 2011

unknown, it is assumed that transcription factors generally modify the activity of more than one target gene and serve to integrate the expression of genes with related biochemical functions in any particular process or metabolic pathway.

Why is it important to identify transcription factors regulating secondary metabolism? It is now well accepted that the control of flux along metabolic pathways is usually vested in more than one biosynthetic step. In addition, major control points in a pathway may change with prevailing environmental, metabolic or developmental conditions (Kacser and Burns, 1973). This broad distribution of control makes modification of flux through engineering of specific structural genes (i.e., those genes encoding enzymes in a particular pathway) very difficult to achieve, because the increases in productivity achievable by increasing the activity of individual biosynthetic steps are usually very limited (in the order of up to 2-fold). Transcription factors modulate the activity of genes involved in entire branches of metabolism, and so engineering their activity provides a far more effective way of engineering secondary metabolism. This can be seen in several reports on the use of transcription factors to engineer flavonoid metabolism including use of Lc and C1 from maize in *Arabidopsis* and tobacco which increased flavonoid production 23-fold (Lloyd et al., 1992), their use in tomato (Bovy et al., 2002) which gave increases of 30-fold (up to 130 µg/g fwt) and their use in soybean (Yu et al., 2003). The use of related genes from *Antirrhinum* induces anthocyanin biosynthesis more than 150-fold in tomato fruit (C. Martin, unpublished results) and upregulation of *AtMYB34* (*ATR1*) in *Arabidopsis* can increase indole glucosinolate accumulation 15-fold (Celenza, 2005). Clearly the use of transcription factors to improve accumulation of natural products in plants through genetic engineering or through marker assisted breeding is enormous. The limitation on this strategy is the number of transcription factors that have been identified as regulating secondary metabolism. Almost all reports focus on the use of the first plant transcription factors identified (C1 and Lc) which regulate anthocyanin biosynthesis in maize. So the major bottleneck in exploiting this potentially powerful tool is in the identification of transcription factors whose biological function is the regulation of the different branches of plant secondary metabolism.

Generally plant transcription factors belong to large families of proteins which share domains for DNA binding and protein-protein interaction. While biological function is not conserved over entire families, the most closely related members fall into phylogenetic sub-families (for examples see Fig. 1), and sub-families share common biological functions dictated by their recognition of common DNA binding sites and their common interactions with other proteins. The R2R3MYB gene family is a good example from plants. Overall this family of proteins is involved in a wide range of processes including the regulation of meristem activity and dorsi-ventral polarity, cellular specification and morphogenesis, intracellular signalling and the control of secondary metabolism, particularly phenylpropanoid metabolism. A number of subgroups of R2R3MYB proteins are involved in controlling phenylpropanoid metabolism. Different members of particular subgroups control the expression of the same sets of target genes but within different tissues or under different environmental conditions (paralogous proteins). Examples are *PAP1* and *PAP2* (*AtMYB75* and *AtMYB90*, respectively) which control anthocyanin biosynthesis particularly in senescing leaves of *Arabidopsis* (Borevitz et al., 2000). Members of structurally related subgroups control related branches of secondary metabolism, as for example in the activity of AtMYB12 (which is closely related to PAP1 and PAP2; Fig. 1) and which controls flavonol biosynthesis (Mehrtens et al., 2005).

These simple rules defining the regulatory activity of transcription factors can be used in functional genomics to identify tentatively new regulatory activities in secondary metabolism. Thus, where one protein has been identified as playing a role in the regulation of particular steps in secondary metabolism, new activities regulating related steps are best sought amongst transcription factors that are, structurally, very similar. Additionally, where a protein has been shown to have a regulatory role in secondary metabolism in one plant species, the best place to look for equivalent or related regulatory roles in other species is in the most similar (orthologous) proteins. Consequently the basic functional genomics approach to the identification of novel regulatory functions (knock out or over-express each gene encoding a transcription factor and characterise the phenotype) can be refined to focus on more limited numbers of likely candidates. Where characterisation of phenotypes involves detailed metabolite analysis, such focus is really the key to success. While whole genome approaches may identify transcription factors with major impacts on biological processes, particularly developmental processes, considerable focus and attention to detail is required to identify genes with metabolic phenotypes, which may explain why so few have currently been identified, despite considerable public sector and commercial investment. Our approach is to couple selective analysis of transcription factor function (through targeted analysis of knockout and over-expression lines) with sensitive and robust metabolite fingerprinting to identify new activities regulating secondary metabolism in *Arabidopsis thaliana*.

WHICH TRANSCRIPTION FACTORS AND WHY?

We have selected 38 genes encoding transcription factors in *Arabidopsis*, for which there is already good preliminary evidence that they regulate different branches of secondary metabolism. The majority of these genes encode proteins of the R2R3MYB family, and of these, many probably function in regulating different branches of phenylpropanoid metabolism, although members of a selected R2R3MYB sub-family regulate glucosinolate metabolism (Celenza et al., 2005) and another regulates the activity of the shikimate pathway (Verdonk et al., 2005). We have also selected to study five members of the AP2 (EREBP) transcription factor family, because their closest relatives in the rosy periwinkle, *Catharanthus roseus*, regulate alkaloid biosynthesis (Menke et al., 1999; van der Fits and Memelink, 2000). Groups of genes encoding transcription factors assigned to discrete sub-families have been selected, especially sub-families for which there is strong evidence that one or more members are involved in regulating secondary metabolism. Different members of discrete sub-families are being considered because there is good evidence that closely related transcription factors share common biological functions.

Amongst the R2R3MYB family we have chosen AtMYB13, At MYB14 and AtMYB15 as likely regulators of general phenylpropanoid metabolism. There is biochemical evidence that the closely related R2R3MYB protein NtMYB2 from tobacco regulates expression of the gene encoding phenylalanine ammonia lyase (PAL) the first enzyme of general phenylpropanoid metabolism (Sugimoto, 2000). The three *Arabidopsis* genes are most highly expressed in rapidly dividing, undifferentiated cells (like those found in meristems) but are also closely associated with transcriptional responses to pathogen challenge (Genevestigator; Gene Atlas results and results of REGIA consortium EU FP5). We have isolated knockout mutants of all three genes and we have already produced the double mutants for metabolite analysis.

Members of a second sub-family that includes AtMYB36, AtMYB87, AtMYB84, AtMYB68, AtMYB37 and AtMYB38, are all highly expressed in roots, and some members are expressed exclusively in roots (Genevestigator; Gene Atlas results and results of REGIA consortium EU FP5). A mutation in one, AtMYB68, results in increased root biomass and a reported increase in lignin production (Feng et al., 2004). Somewhat paradoxically a close relative in tomato is encoded by the *Blind* locus, which controls lateral branching of the shoot (Schmitz, 2002). One possible explanation is that transcription factors belonging to this R2R3MYB sub-family (XIV) are involved in regulating the levels of plant hormones (which are plant secondary metabolites). It is possible that through activating the synthesis of cytokinins some members may influence shoot branching (through antagonism of the negative regulation by auxin) while others influence root growth (the main site of cytokinin synthesis). Alternatively this sub-family may be involved in negative regulation of auxin accumulation either through repressing its synthesis or through promoting its turnover. A MYB transcription factor in

Arabidopsis (ATR1/AtMYB34) regulates the indole glucosinolate branch of tryptophan secondary metabolism[5], suggesting that the involvement of other members of the R2R3MYB transcription factor family in other branches of glucosinolate metabolism is likely. This suggestion for the function of R2R3MYB sub-family XIV has recently been confirmed by metabolic and transcriptomic profiling of mutants and overexpressing lines, which showed that AtMYB28 and AtMYB29 regulate aliphatic glucosinolate production (Gigolashvili et al., 2007; Hirai et al., 2007).

AtMYB5 does not belong to a clear R2R3MYB subgroup (Fig. 1), but is relatively unique in the *Arabidopsis* genome. Its predicted protein structure suggests that it interacts with a bHLH protein in modulating activity of its target genes. It has been shown to interact with the TT8 and AtbHLH0012 bHLH transcription factors (both of which are believed to regulate branches of phenylpropanoid metabolism) and it is one of the proteins most closely related to the PAP transcription factors that regulate anthocyanin biosynthesis (Zimmermann et al., 2004). It is more or less expressed exclusively in seed tissue (Genevestigator; Gene Atlas results and results of REGIA consortium EU FP5). Preliminary targeted metabolite analysis of an atmyb5 knockout line revealed a metabolic phenotype restricted to seeds, with large increases in the content of sinapoyl glucose, suggesting that this transcription factor regulates hydroxycinnamate metabolism in conjunction with bHLH transcription factors (H.P. Mock and C. Martin, unpublished results).

A MYB-related transcription factor, ODORANT1, from petunia has recently been shown to regulate the expression of genes of the shikimate pathway (Mehrtens et al., 2005). Members of the same sub-family in *Arabidopsis* are AtMYB99, AtMYB85, AtMYB42, AtMYB21 and AtMYB20. It is very likely that these *Arabidopsis* proteins will play similar roles to ODORANT 1 in the regulation of metabolism in *Arabidopsis*. Identification of regulators of this metabolic pathway and their target genes may prove to be particularly useful for plant biotechnology, due to the vital nature of shikimate to plants and the importance of herbicides that target this pathway.

Members of R2R3MYB sub-family IV include AtMYB3, AtMYB4, AtMYB6, AtMYB7 and AtMYB32. AtMYB4 is a negative regulator of general phenylpropanoid metabolism, targeting in particular the expression of the gene encoding cinnamate 4-hydroxylase (C4H) (Jin et al., 2000). AtMYB4 serves a specific role in regulating phenylpropanoid metabolism in response to UV-B light, derepressing the synthesis of sinapoyl malate sunscreens in *Arabidopsis*. The structure of the other members of the subgroup is closely related to AtMYB4 and they all contain a conserved C-terminal repression motif, suggesting that they all negatively regulate the expression of their target genes. Different members of this subgroup are expressed in response to different stresses, including UV-B (AtMYB4), osmotic stress (AtMYB3, AtMYB7), bacterial infection (AtMYB4, AtMYB6) and cold stress (AtMYB32) (Genevestigator; Gene Atlas results and results of REGIA consortium EU FP5). The most likely roles for this subgroup of R2R3MYB proteins is that they all negatively regulate hydroxycinnamate or flavonoid biosynthesis, but they may have different target genes. Our overexpression analysis of AtMYB3 and AtMYB32 in *Arabidopsis* shows that such lines have white necrotic lesions on their leaves and are very stunted in growth (Fig. 2). These are both symptoms shown by AtMYB4 overexpression in *Arabidopsis* and tobacco, suggesting that AtMYB3 and AtMYB32 have very similar, negative regulatory roles on hydroxycinnamic acid biosynthesis. We are currently analysing the metabolic fingerprints of the overexpression lines to see if they are the same. We are also testing which target genes AtMYB3 and AtMYB32 regulate to determine whether the different members of this subfamily are truly ‘functionally redundant’ or whether there is specificity in the target genes that each member represses.

AtMYB28, AtMYB29, AtMYB34, AtMYB51, AtMYB76, and AtMYB122 form another closely related sub-family (XII) of MYB proteins in *Arabidopsis*. AtMYB34 (ATR1) is involved in the regulation of tryptophan secondary metabolism and specifically indole glucosinolate biosynthesis (Celenza et al., 2005). The genes of this sub-family are expressed more or less ubiquitiously in *Arabidopsis*, suggesting that different members may be responsible for regulating metabolism of other glucosinolates, or perhaps the different routes for auxin biosynthesis. Very recent publications by independent groups have confirmed AtMYB28, AtMYB29 and AtMYB76 to be regulating aliphatic glucosinolate biosynthesis production (Gigolashvili et al., 2007; Hirai et al., 2007) and it seems likely that AtMYB34, AtMYB51 and AtMYB122 regulate indole glucosinolate biosynthesis.

The final sub-family of R2R3MYB proteins that we are studying includes AtMYB50, AtMYB55, AtMYB61, and AtMYB86. Most is known about AtMYB61 which is required for the extrusion of the mucus surrounding the seed coat (Penfield et al., 2001). The other members of this sub-family (XIII) are all most highly expressed in roots, where they may also regulate the production of mucus, as for example in the root cap slime. However, over expression of *AtMYB61* results in ectopic lignification, suggesting that AtMYB61 may also regulate monolignol synthesis or accumulation (Newmann et al., 2004).

Members of one AP2 protein sub-family are also being studied, because these proteins have been shown to regulate the transcription of genes of alkaloid biosynthesis in *Catharanthus roseus* (Menke et al., 1999; van der Fits and Memelink, 2000). Since *Arabidopsis* does not synthesise these alkaloids it should be very interesting to determine what metabolic pathways (if any) they regulate in *Arabidopsis*. The proteins include AtAP2L51, AtAP2L52, AtAP2L53, AtAP2L54 and AtAP2L55. The genes encoding all these proteins are expressed most strongly in embryonic tissues, but most are induced by different stresses, UV, cold, salt, jasmonate (*AtAP2L51*), salt and senescence (*AtAP2L54*) and salt and cold (*AtAP2L55*), while one is repressed in expression in response to salt and cold (*AtAP2L53*).

The first step in our strategy is to identify which transcription factor members of each sub-family have redundant functions and which have distinct functions in regulating secondary metabolism. This is being determined by over-expression of each member of each sub-family followed by systematic metabolic profiling to determine their effects. One of the products of the EU-FP5 REGIA project is 1200 cDNAs encoding transcription factors from *Arabidopsis* in GATEWAY entry vectors. Each gene has been recombined into a GATEWAY over-expression destination cassette we have constructed in the pBin19 binary vector. A single recombination has produced each transcription factor gene driven (for constitutive, high level expression) by the double 35S promoter in the binary vector, ready for transfer to *Agrobacterium tumifaciens*. These tools have allowed the efficient and rapid generation of 38 transformed *Arabidopsis* populations over-expressing each selected transcription factor gene. Transformed lines are now being screened for equivalent levels of over expression of the different transcription factors, and two independent high-level expression lines are being selected for biological replication of metabolite analysis.

Biological function is being assessed by metabolite analysis and comparison between the different transgenic lines. Genes are judged to have the same biochemical functions if their over-expression gives rise to equivalent changes in metabolites. Metabolite analysis is initially being focussed to those areas indicated by preliminary evidence to be likely regulated by the particular transcription factors under consideration.

METABOLOMIC ANALYSIS OF OVER-EXPRESSION LINES OF *ARABIDOPSIS*

All samples are being examined by a combination of the four profiling methods available to us: 600 MHz ^{1}H NMR, LC/MS, LC/UV and GC/MS. This provides more comprehensive coverage of the metabolome than could be given by any single method. There is some redundancy regarding compounds detected by more than one technique but for this stage the main methods are NMR and LC/MS. The NMR is being used for metabolite fingerprinting because it is rapid, simple and non-targeted and provides an ideal initial pass screening method. This is particularly useful in assessing the effects of over-expression of the different transcription factors from each sub-family. It is being used to determine whether the effects of over-expression of closely related genes on metabolite levels is the same or distinguishable. These decisions are made using principal component analysis (PCA) to compare the profiles from different over-expression lines to controls, and to each other. Where necessary supervised data analysis methods are employed (PLS-DA or genetic algorithm). Where signals responsible for differences are identified (e.g., from PC loadings) the significance of the result is examined for that signal by ANOVA. This informs us of whether the activities of the genes are the same (functionally redundant) or distinct. Because of the often structurally-informative pattern of signals in NMR, it is also possible to identify the origin of any significant unknowns in the NMR spectra. Even where it is not immediately possible to identify the origin of the signal, NMR can direct the search for the structure to LC/MS or GC/MS.

Leaves of *Arabidopsis* over-expressing lines (pooled plants) grown under standardised conditions for three weeks are freeze-dried and then extracted in 70% d_{4-} methanol/buffer in D_2O to provide a supernatant that will be examined directly by NMR profiling, following centrifugation. Internal (chemical) references and electronic references (ERETIC method) are used for quantitative, instrument-related intensity corrections. 2D NMR experiments (COSY, TOCSY, J-resolved, $^1H/^{13}C$ correlations) are carried out at 600 MHz to extend assignments of *Arabidopsis* metabolites already made at 400 MHz (Le Gall et al., 2004) and a library of 600 MHz reference spectra is being compiled from standards and compounds isolated by LC/SPE.

Samples for LC/MS and LC/UV are extracted in 70% methanol. We previously analysed *Arabidopsis* leaf material by LC/UV/DAD with a method that can measure four of the major groups of *Arabidopsis* metabolites in a single run (Le Gall et al., 2004). The LC gradient is a multipurpose chromatographic method that detects glucosinolates at 227 nm, sinapate esters at 325 nm, flavonols at 370 nm and anthocyanins at 520 nm on a C18 5 μm reversed-phase silica column. Compound identification was by LC/MS with the MS run in the negative ionisation mode for the glucosinolates and in positive ionisation mode for the phenylpropanoids. By LC/MS we detected 14 glucosinolates, 5 sinapate esters, 9 flavonol glycosides and 11 anthocyanins in wild type (Ws-0) *Arabidopsis*. Currently we are using LC/MS with the accurate mass MS (Bruker microTOF) for quantitative profiling of secondary metabolites. The TOF MS can, of course, detect and quantify metabolites that are present at much lower levels than those seen by NMR. It also provides unambiguous empirical formulae for unknown metabolites based on a matching of measured and predicted isotope profiles as well as on accurate mass measurement. A separate LC/MS system (Micromass Quattro II triple quad) is available if LC/MS/MS is required.

Whereas direct data input of NMR spectra to multivariate analysis programmes is straightforward (the entire spectral trace may be input) there are two quite lengthy additional steps required for LC/MS and GC/MS data. Deconvolution of chromatogram files can be carried out, e.g., with AMDIS software, to identify characteristic ions for quantification (it allows quantification of compounds that overlap in the chromatogram). Then quantitative information (integrated intensity of characteristic ions) has to be extracted and tabulated from the raw data files for every compound in every sample. This has the advantage over NMR that an explicit intensity is available for every compound (even unknown compounds may be 'indexed' and recognised each time they are present), making ultimate interpretation simpler. At present we do this data extraction in a semi-automated fashion using the Excalibur (Thermo) software but detailed manual checking of many integrations is still necessary. We are currently evaluating MetAlign software (http://www.metalign.nl) as a possible fully automated solution to this data analysis bottleneck. After various pre-processing operations MetAlign uses univariate statistics to provide the location of the significant differences in metabolite profiles which in our case are between over-expressing lines and controls or between the over-expressing lines for different transcription factors from the same sub-family of proteins. Alternatively it can provide a data table suitable for multivariate analysis, after which it is possible to proceed as already described for NMR. The advantage of having accurate mass data for LC/MS profiling, especially for low abundance compounds, has been demonstrated (Wang et al., 2003) but is not exploited in MetAlign.

METABOLITE PROFILING OF KNOCK-OUT MUTANTS

Once a subset of genes encoding transcription factors with unique activities in regulating secondary metabolism has been identified, the biological functions of the encoded proteins can be established by examining knock out mutants. It is often much easier to interpret the regulatory role of a transcription factor from a knock out mutant than from over-expression studies, because the latter may be complicated by metabolic spill-over into related pathways, and transcription factors may lose some target specificity when expressed at high levels (Andersson et al., 1999; Hirai et al., 2007).

Based on the metabolite profiles established for the over-expression alleles, extracts of specific tissues can be compared between mutants and wild type controls, using targeted metabolite profiling with calibration standards for absolute quantification of selected compounds. Therefore as well as metabolomic analysis for relative quantification of a very broad range of compounds, methods such as the GC/MS and LC/UV procedures quoted above are involved. In addition the high mass resolution of the TOF MS allows LC/MS analyses to be performed with high sensitivity and the specificity of MS/MS multiple reaction monitoring, but the experiment is simpler to set up. We plan to focus on analysis of metabolites in those tissues (and under those environmental conditions) in which each gene is (normally) most highly expressed, and compare these profiles to metabolite profiles for equivalent wild type tissues. Specialised analysis will be focused on those compounds most likely to be altered, which we will identify from the over-expression analysis. Data will be analysed using ANOVA, PCA and allied methods to facilitate identification of metabolites that are consistently altered between wild type and mutant plants. The regulatory roles of the different transcription factors will be established principally from the metabolic profiling conducted on the knock-out lines. However, the metabolomic data obtained from the over-expression lines will be used to support or refute these interpretations of function.

IDENTIFICATION OF TARGET GENES OF TRANSCRIPTIONAL REGULATORS OF SECONDARY METABOLISM

Where the metabolic analysis from knock-out and over-expression alleles indicates clearly the branches of secondary metabolism in which the specific transcription factor is involved, the target genes of the metabolic pathways can be analysed by comparing transcript profiles on Affymetrix arrays of mutant, over-expressing lines and controls. Transcript responses can be integrated with the data from metabolite profiling to confirm the regulatory function of the transcription factors and to identify the most important parameters dictating the rate and direction of flux in particular secondary metabolic pathways.

CONCLUSION

The aim of this integrated genomic and metabolomic approach to identify transcription factors controlling secondary metabolism is to identify tools that can be used for effective engineering of natural product accumulation in plants. We have demonstrated the effectiveness of metabolic engineering using genes encoding transcription factors in crops through the production of high-flavonoid tomatoes, by a variety of strategies, which have 3-4 fold higher antioxidant capacities. These are predicted to offer protection against a range of diseases and are currently being assessed on animal models.

Literature Cited

Andersson, K.B., Berge, T., Matre, V. and Gabrielsen, O.S. 1999. Sequence selectivity of c-Myb in vivo. Resolution of a DNA target specificity paradox. J. Biol. Chem. 274:21986-21994.

Blakeslee, J.J., Bandyopadhyay, A., Peer, W.A., Makam, S.N. and Murphy, A.S. 2004. Relocalization of the PIN1 auxin efflux facilitator plays a role in phototropic responses. Plant Physiology 134:28-31.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A. and Lamb, C. 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. The Plant Cell 12:2383-2393.

Bovy, A., de Vosa, R., Kempera, M., Schijlena, E., Almenar Pertejoa, M., Muirb, S., Collins, G., Robinson, S., Verhoeyen, M., Hughes, S., Santos-Buelga, C. and van Tunen, A. 2002. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. The Plant Cell 14:2509-2526.

Celenza, J.L., Quiel, J.A., Smolen, G.A., Merrikh, H., Silvestro, A.R., Normanly, J. and Bender, J. 2005. The *Arabidopsis* ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. Plant Physiology 137:253-262.

Feng, C., Andreasson, E., Maslak, A., Mock, H.P., Mattsson, O. and Mundy, J. 2004. *Arabidopsis* MYB68 in development and responses to environmental cues. Plant Science 167:1099-1107.

Gigolashvili, T., Yatusevich, R., Berger, B., Müller, C. and Flügge, U.-I. 2007. The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. The Plant Journal 51:247-261.

Hirai, M.Y., Sugiyama, K., Sawada,Y., Tohge, T., Obayashi, T., Suzuki, A., Araki, R., Sakurai, N., Suzuki, H., Aoki, K., Goda, H., Nishizawa, O.I., Shibata, D. and Saito, K. 2007. Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. PNAS 104:6478-6483.

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., Tonelli, C., Weisshaar, B. and Martin, C. 2000. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. The EMBO Journal 19:6150-6161.

Kacser, H. and Burns, J.A. 1973. The control of flux. Symp. Soc. Exp. Biol. 27:65-104.

Le Gall, G., Colquhoun, I.J. and Defernez, M. 2004. Metabolite profiling using 1H NMR spectroscopy for quality assessment of green tea, *Camellia sinensis* (L.). J. Agric. Food Chem. 52:692-700.

Lloyd, A.M., Walbot. V. and Davis, R.W. 1992. *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators R and C1. Science 258:1773-1775.

Mehrtens, F., Kranz, H., Bednarek, P. and Weisshaar, B. 2005. The *Arabidopsis* transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. Plant Physiology 138:1083-1096.

Menke, F.L.H., Champion, A., Kijne, J.W. and Memelink, J. 1999. A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. The EMBO Journal 18:4455-4463.

Newman, L.J., Perazza, D.E., Juda, L. and Campbell, M.M. 2004. Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the det3 mutant phenotype. Plant Journal 37:239-250.

Penfield, S., Meissner, R., Shoue, D.A., Carpita, N.C. and Bevan, M.W. 2001. MYB61 is required for mucilage deposition and extrusion in the *Arabidopsis* seed coat. The Plant Cell 13:2777-2791.

Schmitz, G., Tillmann, E., Carriero, F., Fiore, C., Cellini, F. and Theres, K. 2002. The tomato Blind gene encodes a MYB transcription factor that controls the formation of lateral meristems. PNAS 99:1064-1069.

Sugimoto, K., Takeda, S. and Hirochika, H. 2000. MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. The Plant Cell 12:2511-2527.

van der Fits, L. and Memelink, J. 2000. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science 289:295-297.

Verdonk, J.C., Haring, M.A., van Tunen, A.J. and Schuurink, R.C. 2005. ODORANT1 regulates fragrance biosynthesis in petunia flowers. The Plant Cell 17:1612-1624.

Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T.A., Hill, L.R., Norton, S., Kumar, P., Anderle, M. and Becker, C.H. 2003. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal. Chem. 75:4818-4826.

Zimmermann, I.M., Heim, M., Weisshaar, B. and Uhrig, J.F. 2004. Comprehensive identification of *Arabidopsis* MYB transcription factors interacting with R/B-like BHLH proteins. Plant Journal 40:22-34.

Tables

Table 1. Transcription factors of *Arabidopsis* being analysed for potential roles in the regulation of secondary metabolism.

Transcription factor subfamily	Gene member	Targetted area of secondary metabolism
R2R3MYB SF XIII	*AtMYB50* *AtMYB61* *AtMYB55* *AtMYB86*	Monolignol/lignin biosynthesis
R2R3MYB SF XIV	*AtMYB36* *AtMYB87* *AtMYB84* *AtMYB68* *AtMYB37* *AtMYB38*	Cytokinins, auxins, indole metabolism
R2R3MYB SF V	*AtMYB5*	Hydroxycinnamate/sinapate metabolism
R2R3MYB SF IV	*AtMYB6* *AtMYB8* *AtMYB3* *AtMYB4* *AtMYB7* *AtMYB32*	Hydroxycinnamate/flavonoid metabolism
R2R3MYB SF II	*AtMYB15* *AtMYB14* *AtMYB13*	General phenylpropanoid metabolism
No SF name assigned	*AtMYB99* *AtMYB85* *AtMYB42* *AtMYB21* *AtMYB20*	Shikimate biosynthesis
R2R3MYB SFXII	*AtMYB34* *(ATR1)* *AtMYB122* *AtMYB51* *AtMYB28* *AtMYB76* *AtMYB29*	Glucosinolate/indole metabolism
AP2/EREBP	*AtAP2L51* *AtAP2L52* *AtAP2L53* *AtAP2L54* *AtAP2L55*	Camalexin/glucosinolate metabolism
bHLH SF IIIf	*AtbHLH042* *(TT8)* *AtbHLH012* *(AtMYC1)*	Flavonoid/condensed tannin/ hydroxycinnamate metabolism

Figures

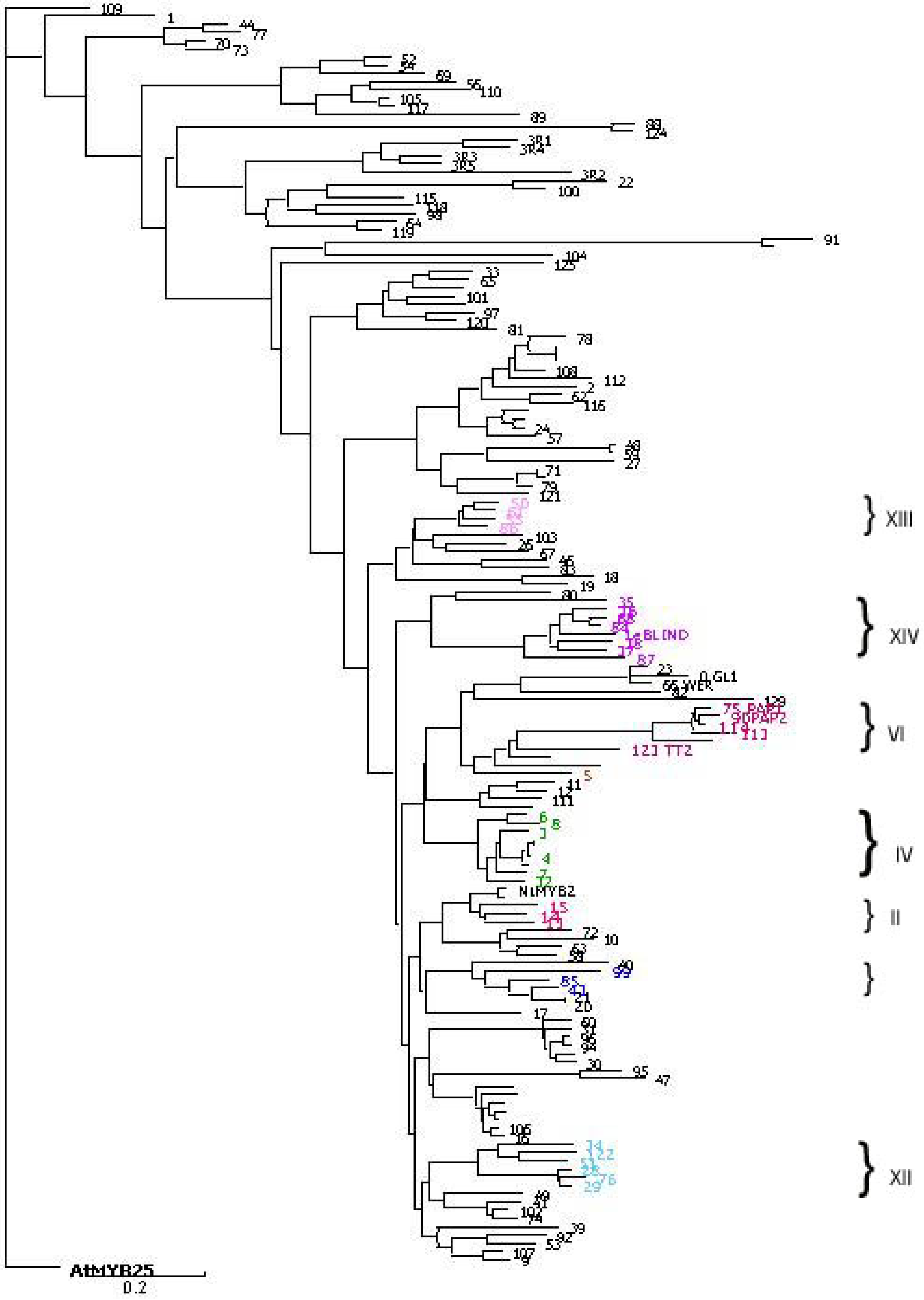

Fig. 1. Phylogenetic tree of R2R3MYB proteins from *Arabidopsis*. Proteins selected for analysis in this project are indicated in colour. Proteins of the same colour belong to a common sub-family of R2R3MYB proteins. Sub-family numbers are indicated after the brackets. Branches without numbers represent related R2R3MYB proteins from species other than *Arabidopsis*. Their names have been omitted for clarity.

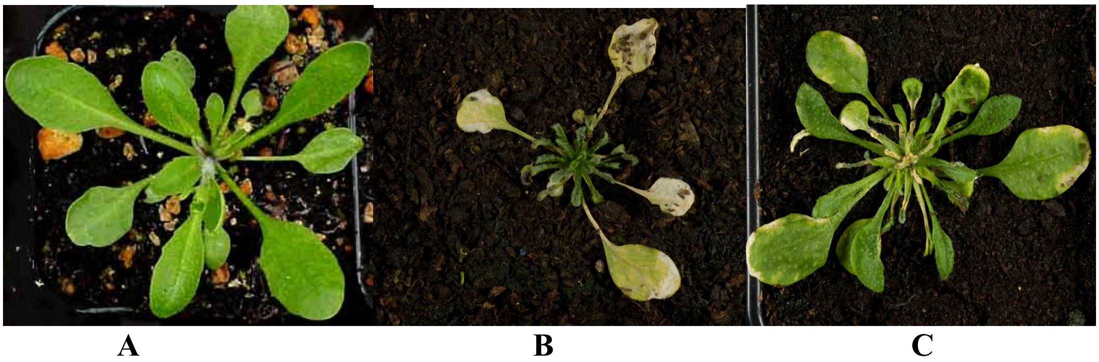

Fig. 2. Similarity in function as determined by over-expression of transcription factors. High-level expression of AtMYB32 (B) or AtMYB3 (C) gives similar white lesions and cell death on leaves compared to wild type Arabidopsis (A). Metabolite analysis will be used to identify the changes in metabolites that underpin these phenotypes, which are likely due to changes in hydroxycimmate/sinapoyl ester levels by analogy to the effects of the related transcription factor, AtMYB4 (Jin et al., 2000).