Chapter 1
Metabolomics Reveals Hemiterpenoids as New Players in the Carbon–Nitrogen Economy

Michael H. Beale and Jane L. Ward

Abstract In plant metabolic engineering, a holistic view of plant metabolism can be provided by modern metabolomics. We discuss a recent application of combined nuclear magnetic resonance–mass spectrometry (NMR–MS) metabolomics to hydroponically grown *Arabidopsis thaliana* as a means to study metabolic reprogramming in plant shoots in response to root stress. This led to the discovery of a novel, specifically regulated overflow from the chloroplastic methylerythritol phosphate (MEP) pathway that resulted in the biosynthesis of hemiterpenoid glycosides (HTGs) from hydroxymethylbutenyl diphosphate. The induction of these compounds in leaves is quite specific to a few conditions that lead to low foliar nitrate, implicating the involvement of this anion in the molecular switch. The reprogramming of isoprenoid metabolism in shoots was also correlated with the production of the phenylpropanoids scopolin and coniferin in roots. These results are discussed against a survey of the literature on hemiterpenoids, including isoprene, in order to develop a unified model of metabolic switching of the MEP pathway that allows carbon overflow from a number of points, depending on the species.

1.1 Introduction

In living systems, the metabolites, whether in the free form, or as components of structural and functional macromolecules, are the end products of a complex self-replicating machinery (genes and proteins), which is itself also built from metabolic products. The complement of low-molecular-weight compounds in a living system is known as the metabolome [1]. Metabolomics is the study of the metabolome, usually in context with genetic, physiological, and environmental influences.

In plants, biochemistry is highly geared to assimilate sunlight, carbon dioxide, nitrate, and other nutrients into metabolites and ultimately into biomass. Most metabolic flux is thus dedicated to these primary processes. However, plants are also the source of a rich secondary metabolism, which is highly species dependent and

J. L. Ward (✉) · M. H. Beale
The National Centre for Plant and Microbial Metabolomics, Rothamsted Research, West Common, Harpenden, Herts Al5 2JQ, UK
e-mail: Jane.ward@rothamsted.ac.uk

D. R. Gang (ed.), *50 Years of Phytochemistry Research*,
Recent Advances in Phytochemistry 43, DOI 10.1007/978-3-319-00581-2_1,

has evolved to give plants an ecological advantage in the constant pressure from microbial pathogens and herbivores. One of the key areas of metabolic control is concerned with sensing and adapting to variable carbon (C) and nitrogen (N) supply. Both are crucial for the normal operation of fundamental biochemical and cellular activities, and internal monitoring of C/N balance is critical. It is recognized that C and N metabolism is coordinated. The C/N balance theory [2] was developed to explain the allocation of resources to secondary metabolism for defense. Over the years, there has been much discussion of the veracity of this hypothesis, for example, regarding plant herbivore interactions. Of particular concern has been the inability to predict quantities of individual secondary compounds [3]. However, outside of this discussion on the cost–benefits of plant defense, it is still widely recognized that the primary assimilation of C and N is monitored and balanced [4, 5] and this is supported by transcriptomic studies that demonstrate genome-wide reprogramming of metabolism in response to nitrate levels [6–9]. One of the key applications of plant metabolomics is in the assessment of this type of metabolic reprogramming that occurs when plants are subjected to environmental pressures and/or genetic alterations. An approach to this problem, at the whole plant level (roots and shoots), is described in this perspective.

1.2 Analyzing the Plant Metabolome

The technology used in metabolomics combines data collected using analytical spectroscopies, such as nuclear magnetic resonance (NMR)and mass spectrometry (MS)with that from genomic and/or phenotype studies. Application of multivariate and correlative statistics to these data allows discovery of metabolites, metabolic processes, and genes associated with developmental programming or environmental effects at the tissue or whole organism level. Increasing levels of sophistication in analytical technologies and our ability to generate and utilize data from large sample numbers, as well as in network modeling, have brought metabolomics to the forefront of plant systems biology research [10].

Obviously, metabolism is highly dynamic and the observed metabolome arises from the modulation of the network of biochemical pathways by genetic and environmental influences. Individual metabolite profiles are thus snapshots of the state of the biochemical network. The primary goal of metabolomics analysis is to assay as many metabolites as possible in order to maximize the view of the network. In reality, even when a panel of extraction, separation, and detection strategies is employed, not all metabolites can be detected and a truly holistic measurement of the metabolome cannot be obtained. However, technologies are improving rapidly and well-found laboratories, with access to high-field NMR and a range of MS techniques (especially LC–MS–MS and GC–MS; LC, liquid chromatography; GC, gas chromatography) and ever-growing spectral libraries of authentic metabolites, are able to measure an impressive number of metabolites, especially when the “lipidome”—the large group of homologous compounds for which specific

techniques have been developed—is included [11]. It is fair to say that metabolomics data are more meaningful, in terms of tracking the dynamics of the biochemical network in a systems biology approach, when they are collected from time-separated samples of tissue, as the network adjusts to genetic programming or environmental change, including stress, disease, and other predatory pressures.

In plants, there are significant challenges associated with simultaneous analysis of the primary metabolome, which is, qualitatively, fairly consistent across species, and the secondary metabolome which is generally species specific. However, closely related species contain related secondary metabolites (e.g., glucosinolates are associated with the Brassicaceae and isoflavonoids with legumes), and this forms the basis for much of the historical chemotaxonomic literature. The primary metabolome, in green parts of the plant, is intimately linked to photosynthesis and varies diurnally over the day–night cycle. There is also dependence on the growth or developmental stage of the plant. Secondary metabolism can be tissue specific (e.g., trichomes are rich sources of terpenoids) and is often induced by biotic and abiotic stresses. There is also dependence on developmental stage, e.g., older leaves can have different secondary metabolite compositions compared to younger leaves. Thus, plant metabolomics experiments have to be carefully designed to suit the biological question that is to be addressed, and the data interpreted against these background developmental and environmental effects.

In terms of the technologies utilized to measure the plant metabolome, the spectroscopic choices are no different to those faced in human and microbial metabolomics, or indeed in classical natural product discovery and identification. Whether collecting data directly on unpurified solvent extracts, or after separation using chromatography-linked spectroscopic techniques, there are pros and cons for every method in terms of metabolome coverage, dynamic range, accuracy of quantitation, and throughput. A summary of these considerations has been given for *Arabidopsis* [12], but is applicable to most plant systems. There are two different, but complementary, approaches to plant metabolomics. The first involves analysis of tissue extracts against a list of known compounds and, depending on the technique, can give quantitative data on several hundred metabolites, of types ranging from central metabolism to secondary products and lipids. The second approach is to compare metabolite signatures, containing thousands of signals from both known and unknown metabolites, by multivariate statistics to select discriminatory metabolite signals explaining the biology/biochemistry under test. Both approaches are hampered by the lack of standards of known compounds that can be used to annotate metabolite signals (i.e., the MS, NMR, or chromatography peaks). Thus, identification of unknowns and the building of spectral libraries have become major challenges in the quest for maximum coverage of the metabolome. This problem is of course much greater in plants and microbes, where there are large differences in secondary metabolomes.

Here, we describe the use of an *Arabidopsis* experimental setup that allows the rapid assessment of metabolic reprogramming in both roots and shoots when stresses are applied to the root or shoot. The data obtained have brought a root-to-shoot signaling perspective to plant metabolomics, and, thus, a more complete ‘systems’ view of the reallocation of metabolic resources.

1.3 The *Arabidopsis* Root–Shoot Metabolome Under Stress

A recent publication [13] describing our work on understanding the effects of different stresses on the *Arabidopsis* root and shoot metabolomes highlighted the importance of not confining studies to a single tissue type or time point. A series of experiments utilizing the commercially available Araponics system (http://www.araponics.com) allowed us to develop a reliable protocol for generating *Arabidopsis* root and shoot tissues in enough quantities to allow for a comprehensive metabolomics analysis to be carried out. The system also allowed for easy access of the root system allowing an investigation on the effect of root perturbation on the shoot metabolome and *vice versa*. Early experiments identified common discriminatory metabolites and also revealed key unknown metabolites, the identities of which were determined *via* a classical natural product chemistry approach. The biological provenance of these metabolites was established and results have highlighted a new link with C–N balance and the terpenoid pathway. The work is described in more detail in the subsequent sections of this perspective.

1.3.1 Effect of Total Nutrient Withdrawal

As a relatively simple initial experiment, the effect of total nutrient withdrawal was explored by switching healthy *Arabidopsis* plants, previously grown under full nutrient supply, to water. NMR–MS analysis of extracts of shoot tissue, harvested at two time points (3 days and 7 days), revealed striking and repeatable differences in metabolite fingerprints between nutrient-deficient and nutrient-supplied plants. Increases in common carbohydrates, flavonoids (kaempferol glycosides), and in particular the amino acid phenylalanine and decreases in sinapoyl malate and the amino acids alanine, threonine, aspartate, and glutamate were clearly evident in the ^{1}H NMR data. However, the most striking feature of the NMR data set was the presence of new signals for olefinic hydrogens and aliphatic methyl groups, which were structurally linked to each other and to other signals in the carbohydrate region, in the nutrient-deprived plants. In fact, further inspection of the data indicated that a small family of related compounds had been induced by the nutrient withdrawal, and that there appeared to be associated reprogramming of metabolism involving phenylalanine, flavonoids, and other primary metabolites. Direct infusion electrospray ionization–MS (DI-ESI–MS) data also displayed biomarkers for nutrient deprivation. A decrease in the ion *m/z* 640 (negative ion mode), a novel flavonoid species that was shown to be related to kaempferol 3,7-dirhamnoside (KRR), by MS–MS fragmentation was observed. However, the most prominent feature was in the positive ion fingerprints where two significant ions at *m/z* 287 and 303 were observed only in nutrient-deprived plants.

1.3.2 Identification of the "Unknowns"

MS data collected at higher resolution showed that the ions at *m/z* 287 and 303 actually arose from sodium and potassium adducts of a compound with the empirical formula $C_{11}H_{20}O_7$. Two-dimensional (2D)-NMR experiments on the nutrient-deprived plant extracts indicated that the discriminatory signals identified in the 1D-NMR fingerprints could be assigned to at least two closely related novel glycosides of a C_5-unsaturated diol, present in about a 4:1 ratio. Chemical shift, connectivity, and coupling data indicated that the major compound could be the hemiterpenoid (2*E*)-4-hydroxy-2-methyl-2-buten-1-yl-*O*-D-glucopyranoside (1) and further confirmation of the formula and the presence of two isomers were obtained from GC–MS analysis of the plant material.

The molecular ion was absent but fragments were characteristic of trimethylsilyl glycosides. The key fragment, *m/z* 157, present in both isomers, due to the aglycone had the empirical formula $C_8H_{17}OSi$ and corresponded to the structural fragment $[(CH_3)_3SiOCH_2CH=C(Me)CH_2]^-$ that was consistent with the C_5-enediol glycoside structure. 1D nuclear Overhauser enhancement spectroscopy (NOESY) NMR data of the major isomer indicated the *trans* (*E)* arrangement of the double bond and heteronuclear multiple bond correlation NMR spectroscopy (HMBC) placed the glycosidic linkage at the 1-position. The structures of this compound and the minor isomer were confirmed by synthesis. This was accomplished from (2*E*)-hydroxy-2-methylbut-2-enyl, 4-acetate, an intermediate previously utilized by us in the synthesis of (*E*)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMBPP) [14]. The synthesis involved Koenigs–Knorr coupling of acetylbromoglucose to the 4-monoacetate, followed by deacetylation and, serendipitously, two isomers were produced that were identical to the two hemiterpenoids observed in *Arabidopsis*. The availability of synthetic material enabled comprehensive 2D-NMR investigations of both isomers to be completed in isolation from contaminating plant carbohydrate. This led to the definitive conclusion that the naturally occurring compounds were the two regional isomers of the hemiterpenoid glycoside (HTG)-*(2E)* -2-methylbut-2-en-1,4-diol glycoside, the major isomer being the 1-glycoside (1) and the minor isomer being the 4-glycoside structure (2) as shown in Fig. 1.1. 2-Methylbut-2-en-1,4-diol (3) was a by-product of the synthesis and was also available from separate deacetylation of the starting monoacetate. Small amounts of this compound were also detected in the NMR spectra of nutrient-deprived *Arabidopsis* extracts.

1.3.3 Nitrate, Nitrite, or N? Establishing the Link with HTG Production in Arabidopsis.

To explore the relationship between nutrient deprivation and the formation of HTGs, experiments were carried out where individual nutrients were omitted from the full growth medium. Using metabolomic screening to monitor the production of the identified biomarkers, it was demonstrated that the prime inducer of the HTGs

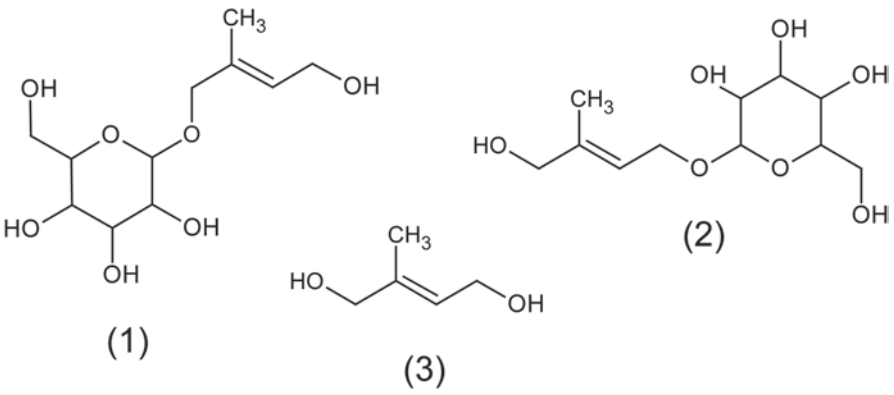


Fig. 1.1 Structures of hemiterpene glycosides identified in *Arabidopsis*

in shoots was root NO_3^- deprivation. The HTGs were still produced when $NH_4^+$was substituted for NO_3^- as a nitrogen source, indicating that nitrate ion sensing was the underlying mechanism behind the synthesis of these compounds. Alternating periods of nitrate deprivation and resupply led to a stepwise accumulation of HTGs under periods of starvation, rising to approximately 1 % dry weight. The HTGs did not appear to be substantially re-assimilated on nitrate resupply. Other primary metabolites also changed during nitrate resupply. Alanine, threonine, aspartate, asparagine, glutamine, and glutamate which were all depleted on nitrate deprivation recovered during the resupply period. Conversely, levels of metabolites that increased during nitrate deprivation (malate, phenylalanine, sucrose, glucose, stachyose, and maltose) were seen to drop during the resupply period, to a level similar to that observed in control plants. Thus, a clear synchronization of certain primary metabolites with nitrate availability and HTG production was observed.

An ion at *m/z* 640, routinely seen in untreated *Arabidopsis* leaf DI-ESI–MS negative ion spectra, was absent in the nitrate-deprived plants. From MS–MS experiments, it was determined that this was formed in the MS source and corresponded to a nitrate adduct of the major flavonoid glycoside KRR and that its intensity was dependent on the nitrate concentration in the leaf tissue. The titer of the *m/z* 640 ion thus serves as a second biomarker for NO_3^- starvation in *Arabidopsis* extracts, and independent measurement of nitrate in the samples by conventional means supported this hypothesis. A weak negative correlation of intensity of ions relating to the HTGs (*m/z* 287 and 303) and *m/z* 640 provided a semi-empirical means of relating high HTG levels directly with low nitrate ion concentration in the same tissue.

It was already apparent that ammonium ion was not involved in the metabolic switch to HTGs. However, there was a possibility that another product of nitrate assimilation—nitrite ion—was the signal. In order to investigate the possibility that levels of nitrite rather than nitrate may drive HTG production, HTG levels in extracts from tissue of nitrate reductase (NR) mutants of *Arabidopsis* were examined. Leaf tissue from these mutants, which cannot assimilate NO_3^-, typically

contain higher levels of free nitrate in the foliar tissue. Under starvation conditions the double (*nia1/nia2*) mutant, retaining only 0.5 % of wild-type NR activity [15], did not produce HTGs at all and the levels of nitrate in the foliar tissue remained high presumably as enough nitrate had built up prior to the starvation period (when the plants were growing on nitrate-containing media) which could not then be assimilated. Together, the results pointed to the fact that it was nitrate levels *per se* and not a product of assimilation, such as nitrite or ammonium ion, which were linked with the HTG formation.

1.3.4 Occurrence of the HTGs Under Alternative Stress Conditions and Evidence of Pathway Reprogramming

The Araponics system, and the ability to readily detect the HTG biomarkers, allowed the application of a number of different stresses within the same experiment and measurement of the concentration of HTGs produced. The results of the experiment are shown as a hierarchical cluster analysis in Fig. 1.2. Levels of the HTGs in leaves exceeded 0.5 % dry weight under nitrate deprivation. Wounding of the roots, but not of leaves, induced formation of the HTGs in leaves but other root stresses such as osmotic stress and salt treatment did not induce HTG biosynthesis despite showing characteristic metabolite changes illustrating that these plants were highly stressed. Likewise, cold treatment of the plants did not induce the new metabolite. Potassium deprivation induced HTGs, but the magnitude was less than that of nitrate starvation. However, oxidative stress induced by hydrogen peroxide treatment of roots induced the HTGs in leaves, to levels reaching approximately the same as those observed with nitrate starvation. Most importantly, we determined that high HTG production was always associated with low foliar nitrate levels. In contrast, no HTGs were detected in any of the root samples from this experiment but other major changes in root tissues were apparent. Increases in scopolin and coniferin were not only present in the same samples as those producing HTGs but were also very well correlated with HTG production. These metabolites are products of the phenylpropanoid pathway in *Arabidopsis* and were already known compounds in *Arabidopsis* roots [16].

Although previously associated with light treatment [17], in our study these metabolites were also evidence of a reprogramming of root metabolism that occurs on nitrate deprivation which is in agreement with observations of a scopolin increase in tobacco roots, in response to nitrogen starvation, as early as 1970 [18]. Analogous results on the redirection of carbon flow into phenylpropanoids under nitrate deficiency have also been reported in tobacco leaves [19]. In *Arabidopsis*, data from transcriptomic studies on nitrate limitation are available [6, 9, 20, 21] and these indicate a similar reprogramming, demonstrating that, as chlorophyll biosynthesis was repressed, phenylpropanoid biosynthesis genes were upregulated. Clearly, there is much more work needed to completely understand this relationship. Unpublished work from our laboratory has indicated that HTGs can be induced *via* a

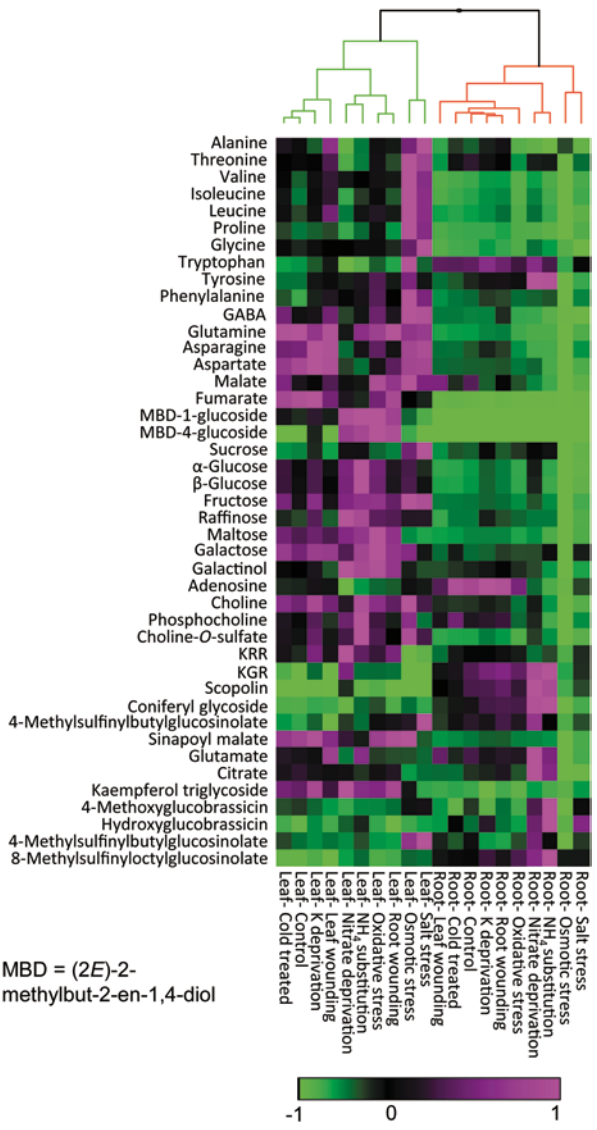


Fig. 1.2 Hierarchical cluster analysis of major metabolites in *Arabidopsis* leaf and root tissue as a result of perturbations from a range of abiotic stresses. Data from ^{1}H NMR (characteristic chemical shift regions) and ESI-MS (*m/z* ions) have been scaled to unit variance and mean centered

high light stress to leaves. Furthermore, despite being supplied with normal nitrate concentrations *via* the roots, foliar nitrate was again very low in the light-stressed plants. Regarding the order of molecular events leading to HTG induction, it is possible that high light induces oxidative stress and as a consequence low foliar nitrate status within the leaves. Alternatively, low nitrate levels caused by an acceleration of photoassimilation in high light induce a pseudo-oxidative stress which in turn triggers the production of HTGs.

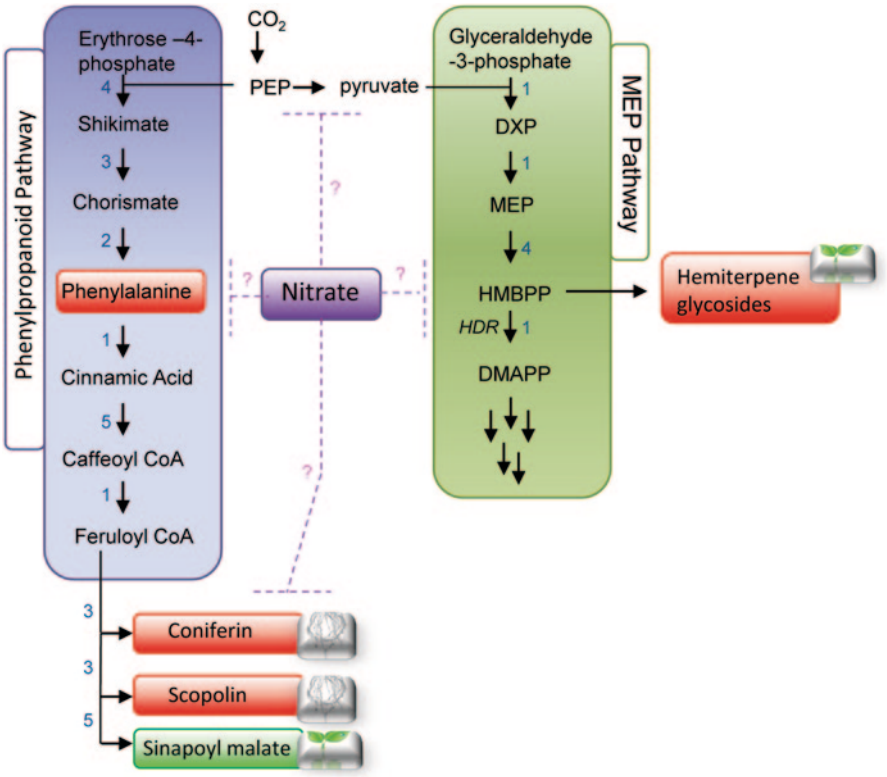
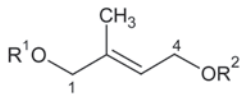


Fig. 1.3 Model of putative pathway interactions in *Arabidopsis*, under foliar nitrate control. Metabolites boxed in red increased under nitrate starvation while those in green decreased. Arrow numbers represent the number of reaction steps between metabolites

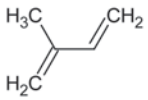
The HTGs are products of the chloroplast terpenoid pathway, known as the methylerythritol phosphate (MEP) pathway (see later). The relationship between nitrate and the co-regulation of MEP and phenylpropanoid pathways is shown in Fig. 1.3. The common link between the pathways involves photoassimilates such as phosphoenol pyruvate (PEP) and partitioning of this type of primary precursor may also be part of the nitrate sensing process.

1.4 Hemiterpenoid Diversity

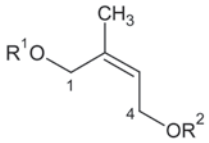
Terpenoids are one of the largest known groups of secondary metabolites. They are incredibly diverse in their structures and biological activities. Despite their structural complexity, however, they all share a common building block, that of the C_5 unit isoprene (4) (Fig. 1.4). While monoterpenes are comprised of two such building



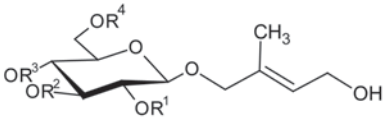
(1) R^1= Glc, R^2 = H
(2) R^1= H, R^2 = Glc
(3) R^1= R^2 = H



(4) Isoprene

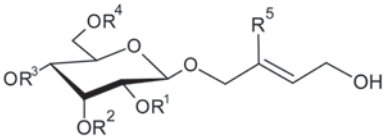


(5) R^1= Glc, R^2 = H
(6) R^1= H, R^2 = Glc
(7) R^1= R^2 = H

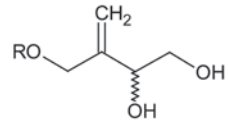


(8) R^1 = R^2 = 4-HPA, R^3 = R^4 = H
(9) R^1= R^2 = 3,4-DHPA, R^3 = R^4 = H
(10) R^1= 3,4-DHPA, R^2 = 4-HPA, R^3 = R^4 =H
(11) R^1= 4-HPA, R^2 = 3,4-DHPA, R^3 = R^4 = H
(12) R^1= R^4 = 3,4-DHPA, R^2 = R^3 = H
(13) R^1= R^4 = 3,4-DHPA, R^2 = 4-HPA, R^3 = H
(14) R^1= 3,4-DHPA, R^2 = R^3 = R^4 =H
(15) R^2= 3,4-DHPA, R^1 = R^3 = R^4 =H
(16) R^4= 3,4-DHPA, R^1 = R^2 = R^3 =H
(17) R^1 = 4-HPA, R^2 = R^3 = R^4 = H
(18) R^4 = 4-HPA, R^1 = R^2 = R^3 = H

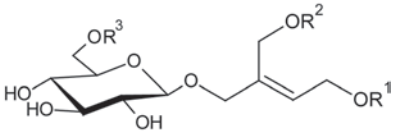
HPA = hydroxyphenylacetyl
DHPA = dihydroxyphenylacetyl



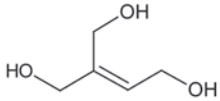
(19) R^1 = R^2 = R^4 = H, R^3 = caffeoyl, R^5 = CH_3
(20) R^1 = R^2 = R^3 = H, R^4 = 3,4-DHPA, R^5 = CH_2OH
(21) R^1 = R^3 = R^4 = H, R^2 = caffeoyl, R^5 = CH_2OH
(22) R^1 = R^2 = R^4 = H, R^3 = caffeoyl, R^5 = CH_2OH
(23) R^1 = R^2 = R^3 = H, R^4 = caffeoyl, R^5 = CH_2OH
(24) R^1 = R^2 = R^3 = R^4 = H, R^5 = CH_2OH
(25) R^1 = R^2 = R^4 = 3,4-DHPA, R^3 = H, R^5 = CH_2OH
(26) R^1 = H, R^2 = R^3 = R^4 = 3,4-DHPA, R^5 = CH_2OH



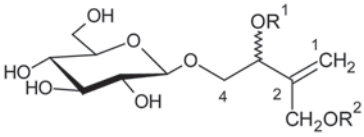
(27) R = glucopyranose
(28) R = allopyranose



(29) R^1 = caffeoyl, R^2 = R^3 = H
(30) R^1 = R^3 = H, R^2 = caffeoyl
(31) R^1 = R^3 = caffeoyl, R^2 = H



(32)



(33) R^1 = H, R^2 = caffeoyl
(34) R^1 = caffeoyl, R^2 = H

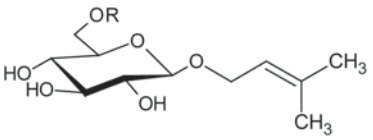
Fig. 1.4 Structures of plant hemiterpenoids

blocks and sesquiterpenes three isoprene units, hemiterpenes are based on a single C_5 structure. These basic units may of course be saturated or unsaturated, isomerized, or bear functional groups which may themselves allow for conjugation to an array of other structural moieties. The family of hemiterpenoids is quite large but here we discuss key HTG structures which have been previously reported.

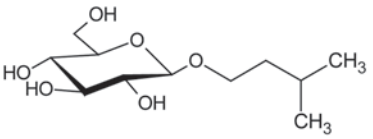
The simplest of all hemiterpenes is the compound isoprene (4) which is emitted from the leaves of many plants. How this emission arises and whether it is as a result of a specific function has, over recent years, been the subject of much discussion and speculation [22, 23]. *Arabidopsis*, however, is a non-emitter of isoprene but our results have demonstrated that, under certain conditions, reasonably high levels of the related conjugated hemiterpenoids can be produced in the aerial tissue of the plant. The compounds described earlier had not previously been detected in *Arabidopsis* but have been reported, mostly at low levels, earlier in a number of less-tractable plant species as demonstrated in Fig. 1.4. The 1-*O*-glycoside (1) was initially reported in 1992 [24] and was isolated from the aerial tissue of *Ornithogalum montanum*, a perennial herbaceous bulb found in Italy and the Balkans. The identical compound was reported in the *Torillis japonica* fruit some years later, along with the simple diol (3) in [25]. The 4-*O*-glycoside (2) has also been previously reported as a deterrent of the bean aphid in stems and leaves of *Vicia hirsuta* [26].

Until our work in *Arabidopsis*, the two positional isomers of the glycosylated hemiterpene had not been reported in the same species. This is unlike the corresponding *cis* isomers found in *Vitis vinifera* leaves where both the *cis*-1-*O*-glycoside (5) and the *cis*-4-*O*-glycoside (6) were detected and found to exist in a 3:1 ratio [27]. The *cis*-1-*O*-glycoside (5) was also reported in leaves of the South American cangorosa plant *Maytenus ilicifolia* where it was described as one of three ilicifolinosides isolated [28]. Unlike the *trans* diol (3), the *cis* diol (7) does not appear to have been isolated from plants, but has been synthesized as part of structure determinations of compounds above [26].

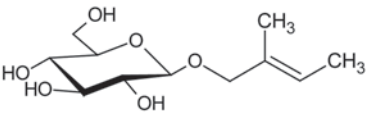
Clearly, the common factor with the compounds discussed so far is the aglycone, 2-methyl-but-2-ene-1,4-diol moiety, whether in the *cis* or *trans* configuration. Further, more complex analogs based on the same agylcone skeleton have also been reported from methanolic extracts of the Japanese Fern, *Hymenophyllum barbatum* [29]. These analogs (8–18) include a *trans* 1-*O*-substituted glucose moiety. The family of compounds isolated, named hymenosides, included a mixture of mono- and dihydroxyphenylacetylated glucose substituents with the position of the substitution varying around the carbohydrate ring. A much wider selection of these so-called hymenosides, bearing a hydroxylated aglycone skeleton and a substituted allose moiety, were also isolated from the same plant species [29–31]. These compounds bore a mixture of caffeoyl- or dihydroxyphenylacetyl-substituted allose arrangements and were conjugated to either 2-methylbut-2-ene-1,4-diol (19), 2-hydroxymethyl-but-2-ene-1,4-diol (20–26) or 2-methylene-butane-1,3,4-triol (27–28). Caffeoyl-substituted hemiterpene glycosides have also been isolated from the bark and roots of two different *Ilex* (holly) species within the Aquifoliaceae family. Three new caffeoyl-substituted HTGs were isolated from methanolic extracts of the dried bark of *Ilex macropoda* [32]. These metabolites (29–31), based on the triol structure (32), bear caffeoyl substituents at one (C-4 or C-5) or two (C-4 and C-5) positions alongside the glucose substitution at C-1. Two further HTGs, named pubescenosides, were isolated from the roots of *Ilex pubescens* [33]. These compounds (33, 34) possessed potent antiplatelet aggregation activities and were based on the aglycone structure 2-hydroxymethyl-3,4-dihydroxy-but-1-ene.



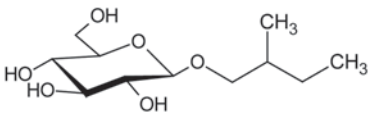
(35) R = H
(39) R = apiofuranose



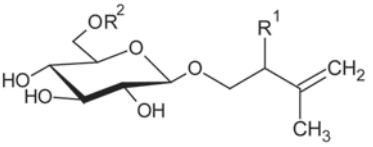
(36)



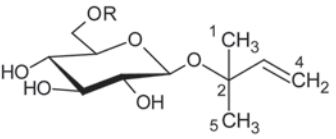
(37)



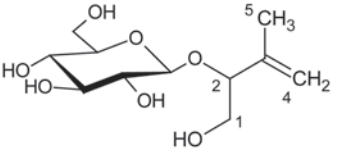
(38)



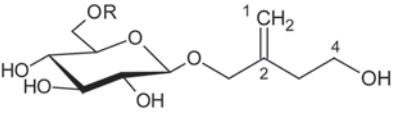
(40) R^1 = H, R^2 = apiofuranose
(42) R^1 = OH, R^2 = H



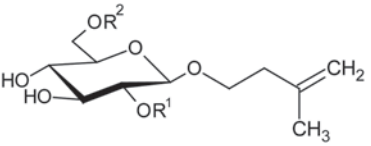
(41) R = apiofuranose



(43)



(44) R = H
(45) R = CO-C(CH_2)CH_2CH_2OH



(46) R^1 = R^2 = H
(47) R^1 = glucopyranose, R^2 = H
(48) R^1 = H, R^2 = xylopyranose
(48) R^1 = H, R^2 = xylofuranose

Fig. 1.5 Structures of plant hemiterpenoids, continued

Further HTGs (Fig. 1.5) have been reported in a number of diverse species. Perhaps the most relevant of these is a prenyl glycoside based on a 3-methylbut-2-en-1-ol (dimethylallyl alcohol) (35). Originally reported as a fragrance precursor

in flower buds of *Citrus unshiu* [34] and subsequently in 1977 [35] *via* Soxhlet extraction of the aerial parts of *Ferula loscosii,* a prevalent species in Spain, this metabolite was again reported as being part of a collection of metabolites present in the water-soluble fraction of the methanol extract of fennel fruit (*Foeniculum vulgare,* Umbelliferae) alongside the fully saturated analog (36) and the positional isomer, 2-methyl-but-2-en-1-ol glucoside (37), and the analog (38) [36]. Early reports of these saturated HTGs generally associated these types of compounds with a variety of medicinal properties such as antimicrobial and antifungal activities. Such an example [37] was the isolation from leaves of the Canarian shrub *Bystropogon plumosus* of 2-methylbutan-1-yl-β-D-glucoside (38) to levels of around 0.01 %.

An apiofuranose analog (39) of the dimethylallylalcohol glycoside (35) was isolated from *Vitis vinifera* wine [38] alongside the similar analog from the 3-methylbut-3-en-1-ol glycoside (40). An apiofuranose analog (41) of 2-methylbut-3-en-2-ol glycoside has also been isolated from the water-soluble extract of the root and rhizome tissue of *Glehnia littoralis* [39]. The position of the double bond within the aglycone structure varies and there have been a number of reports of such metabolites with a terminal double bond structure. The position of the double bond and subsequent substitution are presumably as a result of the biosynthetic route utilized in particular tissues. Aerial tissue of *Lamium album* afforded the 2-*O*-glucoside of 3-methylbut-3-en-1,2-diol (43) named by the authors as hemialboside [40] while the extraction of fresh rhizomes of *Coptis japonica* var. *dissecta* allowed the isolation of a group of close analogs (42, 44–45) [41]. Finally, a series of hemiterpene glycosides (46–49) based on the 3-methylbut-3-en-1-ol parent structure were isolated recently from fruits of noni (Indian mulberry) [42]. This set of metabolites included the 1-*O*-glucoside (46) and conjugates further extended with an additional glucopyranose (47), xylopyranose (48), or xylofuranose (49).

In summary, we have highlighted examples of compounds that appear to be structurally related to the main C_5 building blocks of isoprenoid biosynthesis. Other related structures have been summarized by Dembitsky [43]. Of particular interest in metabolomics and carbon flow is the provenance of the variety of skeleta observed, and this is discussed in the next section.

1.5 Hemiterpenes—Products of Internal Regulation of the Chloroplast Isoprenoid Pathway?

The structural resemblance of the hemiterpene glycosides to HMBPP, a key intermediate in the chloroplastic MEP pathway, indicates a biosynthetic pathway involving hydrolysis of the diphosphate group followed by glycosylation at either end of the 1,4-diol. (Fig. 1.6). Support for this route was provided [13] by feeding intermediates—HMBPP, desoxyxylulose phosphate (DXP), and 1,4-diol—to leaves excised from nitrate-depleted *Arabidopsis* plants. Although, in feeding studies such as these, phosphorylated compounds are unlikely to be taken up into plant chloroplasts, the results indicated that the distal precursor DXP was taken up and

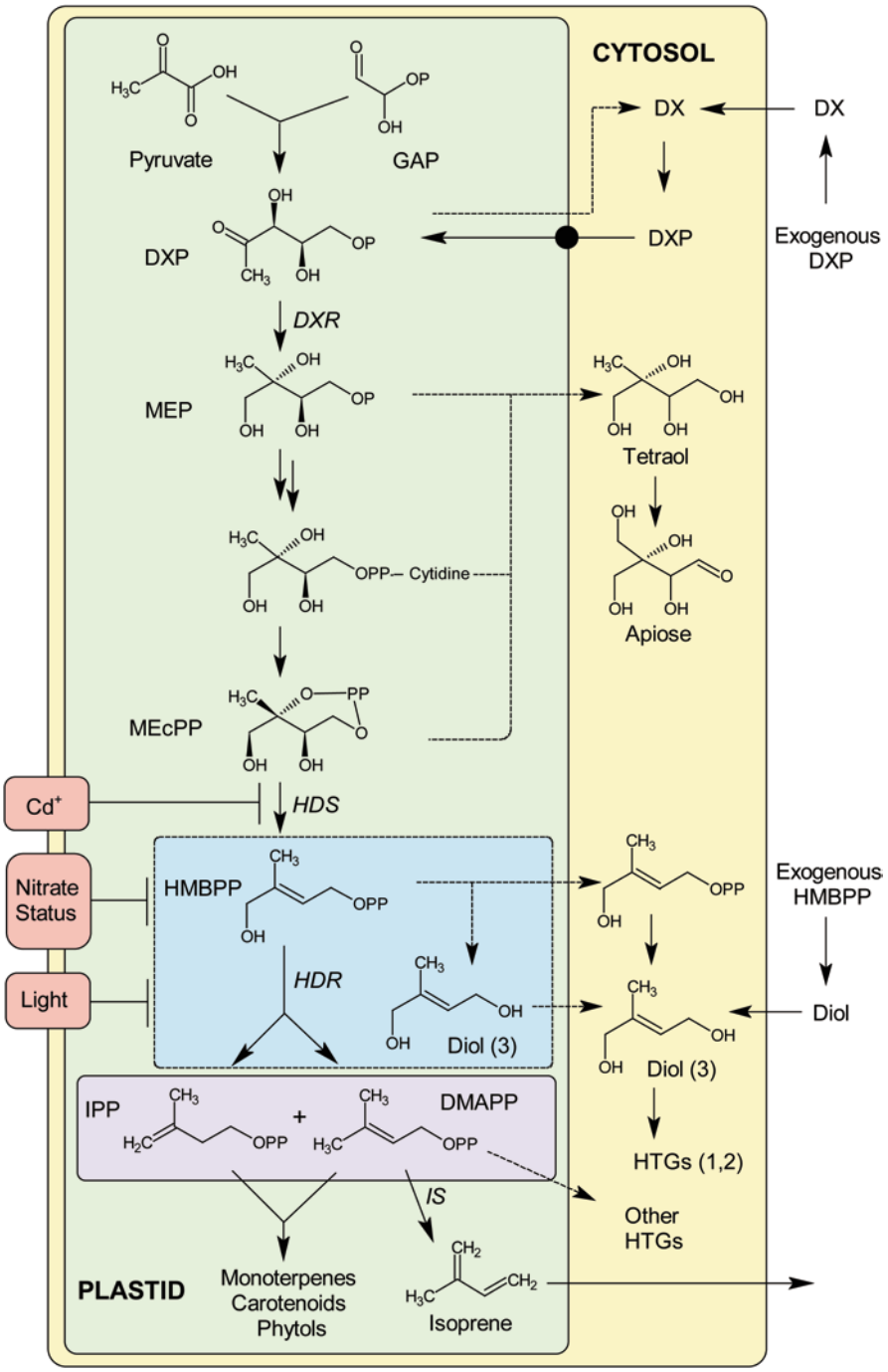


Fig. 1.6 The chloroplast MEP pathway and proposed provenance of hemiterpene alcohols and derivatives *via* stress-induced pathway shunts

resulted in the formation of some 3–4 times more hemiterpene glycosides in nitrate-depleted tissue when compared with controls. There are precedents for DXP uptake and turnover in plant tissue and the mechanism is thought to involve extracellular dephosphorylation to the alcohol and then uptake into the cell, followed by re-phosphorylation to DXP which is then actively transported into the chloroplast by a carrier protein [44, 45]. When either HMBPP or the 1,4-diol was fed to cut leaves of nitrate-depleted *Arabidopsis*, Ward et al. [13] observed high turnover of both compounds to the hemiterpene glycosides, providing firm evidence for the proposed pathway. However, there was no difference in turnover between nitrate-depleted and control plants.

Given that HMBPP has been shown not to be transported across chloroplast membranes [46] and it can be assumed that it also was dephosphorylated outside of the cell, this result indicated that the glycosylation of the diol occurred outside of the chloroplast. Together, the feeding results from nitrate-depleted plants support the hypothesis that nitrate signaling decreased flux through the enzyme hydroxymethylbutenyl diphosphate reductase (HDR; Fig. 1.6) causing HMBPP to build up and overflow to the HTGs. HMBPP generated *de novo* within the chloroplast (from the endogenous pathway or from exogenous DXP) was subject to the nitrate-regulated shunt mechanism. HMBPP added exogenously does not appear to reach the chloroplast but is dephosphorylated and glycosylated irrespective of nitrate levels. The experiments could not distinguish between low-nitrate induction of an HMBPP phosphatase in the chloroplast and the alternate mechanism of low-nitrate-induced translocation of excess HMBPP from chloroplast to cytosol (Fig. 1.6). The results observed in *Arabidopsis* parallel those observed in virus-induced gene silencing of HDR reductase activity in *Nicotiana benthamiana* [47]. In this system, silencing of HDR led to the conversion of HMBPP to the diol, via the monophosphate HMBP, indicating similar regulatory mechanisms to those observed in *Arabidopsis*.

Many of the known hemiterpene structures depicted in Figs. 1.4 and 1.5 can be accommodated by similar biosynthetic conversions from HMBPP, dimethylallyl diphosphate (DMAPP), or isopentenyl diphosphate (IPP). Some structures (e.g., 41, Fig. 1.5) can arise from allylic rearrangement of the double bond in the dephosphorylation step or at the alcohol stage. Most difficult to rationalize is the biosynthesis of the dihydro analogs (e.g., 36, Fig. 1.5) and *cis*-diol-based structures (e.g., 5–7, Fig 1.4). Both appear to arise out of the downstream modification of dimethylallyl alcohol by either hydrogenation or hydroxylation.

Other naturally occurring C_5 polyols such as apiose and the related tetraol (Fig. 1.6) are also potentially formed from MEP pathway intermediates as shown [25]. This opens up the possibility of other overflow mechanisms from different places in the MEP pathway. It has also been reported that 2-C-methyl-D-erythritol 2,4 cyclodiphosphate, a potential precursor of such tetraol structures, builds up in high light, high temperatures, and also on cadmium stress in spinach tissue [48]. The production of isoprene itself is well established from chloroplastic DMAPP *via* the action of a specific enzyme isoprene synthase and has been suggested to be a shunt product when DMAPP pool size becomes too large [22]. There has been much discussion concerning the raison d'être for isoprene emission from plants. This

occurs in many tree species and contributes to greenhouse gases as well as the summer haze over forests. There is evidence that isoprene emission has a function in the protection of plant membranes under heat and light stress [23], but recent results of field trials with poplar trees where isoprene synthase genes have been suppressed by RNA interference (RNAi) techniques indicate that there is no growth and development advantage from isoprene emission, and perhaps the only benefit of isoprene emission may come from interactions with herbivores [49]. Despite the different routes of formation of isoprene (induction of a specific chloroplast enzyme) and the HTGs (potentially via nonspecific extra-chloroplastic phosphorylases and glycosyl transferases), there are parallels that can be drawn. Both processes appear to be governed by the relevant isoprenoid diphosphate pool size [13, 50] and furthermore appear to be a consequence of plant nitrate status [13, 51] as well as induced by high light [52, Ward et al. unpubl.].

Obviously more work needs to be done on flux through the MEP pathway, but this work and increasing evidence in the literature indicate that hemiterpenoid production is the result of a multilevel intermediate pool size management system that is responsive to a number of external and internal signals.

1.6 Future Prospects

This work has demonstrated the power of unbiased metabolomics not only to discover new natural products but also to provide a "systems" overview of metabolism as it adjusts to pressure from environmental and nutritional stress. Until this work, flux through the terpenoid pathway in *Arabidopsis* was thought to be low, relative to many plants that accumulate large amounts of secondary terpenoids. Despite the presence of greater than 30 mono- and sesquiterpene synthase genes [53], oxidized terpenes have not been reported in the model plant, although volatile mono- and sesquiterpene hydrocarbons are emitted at low levels by flowers and roots [54]. Aside from the extensive research on isoprene, hemiterpenoids have been a somewhat neglected research area. However, the discovery of HTGs in the model plant and the specific nature of the induction of their biosynthesis will stimulate renewed interest in these molecules. The possibility that the HTG formation in *Arabidopsis* is a paradigm for isoprene production in trees will bring a greater range of genetic tools and resources to bear on carbon overflow mechanisms. This will be of direct interest to terpene engineering whether the strategic outlet is the accumulation of valuable compounds or for manipulation of interactions with predators. The production of isoprene and other linear terpenoids as biofuels is also a growing research topic. Genetic engineering is beginning to shed light on the 'raison d'être' for isoprene production by tree species [49], and the debate on carbon overflow versus thermoprotectant continues. Delineation of the regulatory mechanism of HTG production in *Arabidopsis* will surely inform on this discussion.

The use of the metabolomics screen for HTGs will allow a detailed genetic, temporal, and environmental study of the induction of the low foliar nitrate state

that leads to HTG synthesis. Nitrate sensing and signaling in leaf biochemistry has long been mooted as a key factor in metabolic control in plants [7]. The work described here presents a new opportunity to explore this, using the extensive genetic resources available for *Arabidopsis*. The links between high light, oxidative stress, and low foliar nitrate revealed in this study can also now be explored by using HTG production as readout. The high correlation between HTG production in the shoots with diversion of phenylpropanoid flow to coniferin and scopolin in the roots perhaps indicates that aspects of the plant defense response cross over with the nitrate response, and that nutrient limitation may activate the flow of carbon into secondary defense metabolites.

In the discussion, we have suggested that a palette of regulatory mechanisms modulate the chloroplastic MEP pathway. We suggest that different plants use different routes to off-load excess flux in stress conditions. The determinant appears to be the pool size of the relevant isoprenoid diphosphate. *Arabidopsis* and tobacco [47] produce HMBPP-derived products. Isoprene is emitted by many tree species, and DMAPP pool size appears to be a factor [50] in isoprene emission. Similarly, hemiterpenoids apparently derived from DMAPP (and IPP) have been widely reported (Figs. 1.4 and 1.5). Also of interest are the tetraols that could be derived from a number of the higher precursors (Fig. 1.6). Future developments in metabolomic analysis of the MEP pathway metabolites will be necessary as an aid to understanding the extent of this form of regulation. Also of interest, in the biosynthesis of HTGs, are the dephosphorylation and glycosylation steps. Evidence for the presence or absence of specific and inducible enzymes for these steps needs to be sought and combined transcriptomic–metabolomic studies of stressed plants will aid in this endeavor.

Acknowledgments Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the UK.

References

1. Oliver SG, Winson MK, Kell DB, Banganz F (1998) Systematic functional analysis of the yeast genome. Trends in Biotech 16:373
2. Bryant JP, Chapin FS, Klein DR (1983) Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. Oikos 40:357
3. Hamilton JG, Zangerl AR, DeLuca EH, Berenbaum MR (2001) The carbon-nutrient balance hypothesis: its rise and fall. Ecol Lett 4:86
4. Stitt M (1999) Nitrate regulation of metabolism and growth. Curr Opin Plant Biol 2:178
5. Zheng ZL (2009) Carbon and nitrogen nutrient balance signalling in plants. Plant Signal Behav 4:584–91
6. Scheible WR, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, Schindelsach D, Thimm O, Udvardi MK, Stitt (2004) Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. Plant Physiol 136: 2483

7. Vidal EA, Gutierrez RA (2008) A systems view of nitrogen nutrient and metabolite responses in Arabidopsis. Curr Opin Plant Biol 11:521
8. Coruzzi GM, Zhou L (2001) Carbon and nitrogen sensing and signaling in plants: emerging 'matrix effects' Curr Opin Plant Biol 4:247
9. Wang R, Okamoto M, Xing X, Crawford NM (2003) Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron and sulfate metabolism. Plant physiol. 132:556
10. Redestig H, Szymanski J, Hirai MY, Selberg J, Willmitzer L, Nikoloski Z, Saito K (2011) Data integration, metabolic networks and systems biology. In: Hall RD (ed) Annual Plant Reviews, vol43, Biology of Plant Metabolomics. Wiley-Blackwell, Oxford, pp 261–316
11. Welti R, Shah J, Li W, Li M, Chen J, Burke JJ, Fauconnier M-L, Chapman K, Chye M-L, Wang X (2007) Plant lipidomics: Discerning biological function by profiling plant complex lipids using mass spectrometry. Frontiers in Bioscience 12:2494
12. Beale MH, Sussman MR (2011) Metabolomics of *Arabidopsis Thaliana*. In: Hall RD (ed) Annual Plant Reviews, vol43, Biology of Plant Metabolomics. Wiley-Blackwell, Oxford, pp 157–180
13. Ward JL, Baker JM, Llewellyn AM, Hawkins ND, Beale MH (2011) Metabolomic analysis of *Arabidopsis* reveals hemiterpenoid glycosides as products of a nitrate ion-regulated, carbon flux overflow. Proc Nat Acad Sci 108:10762
14. Ward JL, Beale MH (2002) Synthesis of (*2E*)-4-hydroxy-3-methylbut-2-enyl diphosphate, a key intermediate in the biosynthesis of isoprenoids. J Chem Soc Perkin Trans I 6:710
15. Wilkinson JQ, Crawford HM (1993) Identification and characterisation of a chlorate-resistant mutant of *Arabidopsis thaliana* with mutations in both nitrate reductase structural genes *NIA1* and *NIA2*. Mol Gen Genet 239:289
16. Bednarek P, Schneider B, Svatoš A, Oldham NJ, Hahlbrock K (2005) Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoidsecondary metabolism in Arabidopsis roots. Plant Physiol138:1058
17. Hemm MR, Rider SD, Ogas J, Murry DJ, Chapple C (2004) Light induces phenylpropanoid metabolism in Arabidopsis roots. Plant J 38:765
18. Armstrong GM, Rohrbaug LM, Rice EL, Wender SH (1970) Effect of nitrogen deficiency on concentration of caffeoylquinic acids and scopolin in tobacco. Phytochemistry 9:945
19. Fritz C, Palacios-Rojas N, Feil R, Stitt M (2006) Regulation of secondary metabolism by the carbon-nitrogen status in tobacco: NO3- inhibits large sectors of phenylpropanoid metabolism. The Plant Journal 46:533
20. Bi YM, Wang RL, Zhu T, Rothstein SJ (2007) Global transcription profiling reveals differential responses to chronic nitrogen stress and putative nitrogen regulatory components in Arabidopsis. BMC Genomics 8:Article Number: 281
21. Peng MS, Bi YM, Zhu T, Rothstein SJ (2007) Genome-wide analysis of Arabidopsis responsive transcriptome to nitrogen limitation and its regulation by the ubiquitin ligase gene NLA. Plant Mol Biol 65:775
22. Sanadze GA (2004) Biogenic isoprene (a review). Russ J Plant Physiol 51:729
23. Sharkey TD, Wiberley AE, Donohue AR (2008) Isoprene emission from plants: Why and how. Ann Bot (Lond) 101:5
24. Nicoletti M, Tomassini L, Foddai S (1992) A new hemiterpene glucoside from *Ornithogalum montanum*. Planta Med. 58:472
25. Kitajima J, Suzuki N, Ishikawa T, Tanaka Y (1998a) New hemiterpenoid pentol and monoterpenoid glycoside of *Torillis japonica* fruit and consideration of the origin of apiose. Chem Pharm Bull 46:1583
26. Ohta N, Mori N, Kuwahara Y, Nishida R (2006) A hemiterpene glucoside as a probing deterrent of the bean aphid, *Megoura crassicauda*, from a non-host vetch, *Vicia hirsute*. Phytochemistry 67:584
27. Messerer M, Winterhalter P (1995) (*2Z*)-4-Hydroxy-2-methyl-2buten-1-yl-β-D-glucoside and (*2Z*)-1-Hydroxy-2-methyl-2buten-4-yl-β-D-glucoside: Two new hemiterpene glucosides from *Vitis vinifera* leaves. Nat Prod Lett 5:241

28. Zhu N, Sharapin N, Zhang J (1998) Three glucosides from *Maytenus Ilicifolia*. Phytochemistry 47:265
29. Toyota M, Oiso Y, Asakawa Y (2002) New glycosides from the Japanese fern *Hymenophyllum barbatum*. Chem Pharm Bull 50:508
30. Toyota M, Oiso Y, Asakawa Y (2001) New bitter-tasting hemiterpene glycosides from the Japanese Fern *Hymenophyllum barbatum*. Chem Pharm Bull 49:1567
31. Oiso Y, Toyota M, Asakawa Y (2001) Hymenosides A-F, six new hemiterpene glucosides from the Japanese fern *Hymenophyllum barbatum*. Chem Pharm Bull 49:126
32. Fuchino H, Tachibana H, Tanaka N (1997) Three new hemiterpene glycosides from *Ilex macropoda* Chem Pharm Bull 45:1533
33. Jiang Z-H, Wang J-R, Li M, Liu Z-Q, Chau K-Y, Zhao C, Liu L (2005) Hemiterpene glucosides with anti-platelet aggregation activities from *Ilex pubescens* Chem Pharm Bull 68:397
34. Yoshikawa K, Kobayashi M, Arihara S (1996) Flower fragrance precursors from flower buds of *Citrus unshiu* Marcov Natural Med 50:176
35. Pinar M, Martin-Lomas M (1977) 2-Methyl-3-buten-2-yl-β-D-glucopyranoside from *Ferula loscosii*. Phytochemistry 16:281
36. Kitajima J, Ishikawa T, Tanaka Y (1998b) Water-soluble constituents of Fennel I. Alkyl glycosides. Chem Pharm Bull 46:1643
37. Nahrstedt A, Economou D, Wray V (1990) 2-Methylbutan-1-yl-β-D-glucoside. A hemiterpene glucoside from *Bystropogon plumosus*. J Nat Prod 53:1387
38. Baltenweck-Guyot R, Trendel J-M, Albrecht P (1997) New hemiterpene glycosides in *Vitis vinifera* wine. J Nat Prod 60:1326
39. Kitajima J, Okamura C, Ishikawa T, Tanaka Y (1998c) New glycosides and furocoumarin from the *Glehnia littoralis* root and rhizome. Chem Pharm Bull 46:1939
40. Damtoft S, Jensen SR (1995) Hemialboside, a hemiterpene glucoside from *Lamium album*. Phytochemistry 39:923
41. Yoshikawa K, Kinoshita H, Arihara S (1997) Non-basic components of *Coptis rhizome* II. Four new hemiterpenoid glucosides. Two new phenylpropanoid glucosides and a new flavonoid glycoside from *Coptis japonica* var. *dissecta*. Nat Med (Tokyo) 51:244
42. Akihisa T, Seino K-I, Kaneko E, Watanabe K, Tochizawa S, Fukatsu M, Banno N, Metori K, Kimura Y (2010) Melanogenesis inhibitory activities of iridoid-, hemiterpene-, and fatty acid glycosides from the fruits of *Morinda citrifolia* (Noni). J Oleo Sci 59:49
43. Dembitsky VM (2006) Astonishing diversity of natural surfactants: 7. Biologically active hemi- and monoterpenoid glycosides. Lipids 41:1
44. Schwender J, Zeidler J, Gröner R, Müller C, Focke M, Braun S, Lichtenthaler FW, Lichtenthaler HK. (1997) Incorporation of 1-deoxy-d-xylulose into isoprene and phytol by higher plants and algae. FEBS Lett 414:129
45. Hemmerlin A, Tritsch D, Hartmann M, Pacaud K, Hoeffler J-F, van Dorsselaer A, Rohmer M, Bach TJ. (2006) A cytosolic Arabidopsis D-xylulose kinase catalyzes the phosphorylation of 1-Deoxy-D-Xylulose into a precursor of the plastidial isoprenoid pathway. Plant Physiol 142:441
46. Flügge UI, Gao W (2005) Transport of isoprenoid intermediates across chloroplast envelope membranes. Plant Biol 7:91
47. Page JE, Hause G, Raschke M, Gao W, Schmidt J, Zenk MH, Kutchan TM (2004) Functional analysis of the final steps of the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway to isoprenoids in plants using virus-induced gene silencing. Plant Physiol 134:1401
48. Rivasseau C, Seeman M, Boisson AM, Streb P, Gout E, Douce, R, Rohmer, M, Bligny R (2009) Accumulation of 2-C-methyl-d-erythritol 2,4-cyclodiphosphate in illuminated plant leaves at supra optimal temperatures reveals a bottleneck of the prokaryotic methylerythritol 4-phosphate pathway of isoprenoid biosynthesis. Plant Cell Environ 32:82
49. Behnke K, Grote R, Brüggemann N, Zimmer I, Zhou G, Elobeid M, Janz D, Polle A, Schnitzler J-P. (2012) Isoprene emission-free poplars—a chance to reduce the impact from poplar plantations on the atmosphere. New Phytologist 194:70

50. Vickers CE, Possell M, Hewitt CN, Mullineaux PM (2010) Genetic structure and regulation of isoprene synthase in Poplar (Populussp). Plant Mol Biol 73: 547
51. Rosensteil TN, Ebbets AL, Khatri WC, Fall R, Monson RK (2004) Induction of poplar leaf NO_3 reductase: a test of extra chloroplastic control of isoprene emission. Plant Biol 6:12
52. Botella-Pavia P, Besumbes O, Phillips MA, Carretero-Paulet L, Boronat A, Rodriguez-Concepcion M (2004) Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. Plant J 40:188
53. Aubourg S, Lecharny A, Bohlmann J (2002) Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. Mol Gen Genomics 267:730
54. D'Auria JC, Gershenzon J (2005) The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. Curr Opin Plant Biol. 8:308–16