Amino Acids (2010) 39:1023–1028
DOI 10.1007/s00726-010-0566-7

ORIGINAL ARTICLE

Principal transcriptional regulation and genome-wide system interactions of the Asp-family and aromatic amino acid networks of amino acid metabolism in plants

Hadar Less · Ruthie Angelovici · Vered Tzin · Gad Galili

Received: 7 January 2010 / Accepted: 10 March 2010 / Published online: 4 April 2010

Abstract Amino acid metabolism is among the most important and best recognized networks within biological systems. In plants, amino acids serve multiple functions associated with growth. Besides their function in protein synthesis, the amino acids are also catabolized into energy-associated metabolites as well we into numerous secondary metabolites, which are essential for plant growth and response to various stresses. Despite the central importance of amino acids in plants growth, elucidation of the regulation of amino acid metabolism within the context of the entire system, particularly transcriptional regulation, is still in its infancy. The different amino acids are synthesized by a number of distinct metabolic networks, which are expected to possess regulatory cross interactions between them for proper coordination of their interactive functions, such as incorporation into proteins. Yet, individual amino acid metabolic networks are also expected to differentially cross interact with various genome-wide gene expression programs and metabolic networks, in respect to their functions as precursors for various metabolites with distinct functions. In the present review, we discuss our recent genomics, metabolic and bioinformatics studies, which were aimed at addressing these questions, focusing mainly on the Asp-family metabolic network as the main example and also comparing it to the aromatic amino acids metabolic network as a second example (Angelovici et al. in Plant Physiol 151:2058–2072, 2009; Less and Galili in BMC Syst Biol 3:14, 2009; Tzin et al. in Plant J 60:156–167, 2009). Our focus on these two networks is because of the followings: (i) both networks are central to plant metabolism and growth and are also precursors for a wide range of primary and secondary metabolites that are indispensable to plant growth; (ii) the amino acids produced by these two networks are also essential to the nutrition and health of human and farm animals; and (iii) both networks contain branched pathways requiring extensive regulation of fluxes between the different branches. Additional views on the biochemistry, regulation and functional significance of the Asp-family and aromatic amino acid networks and some of their associated metabolites that are discussed in the present report, as well as the nutritional importance of Lys and Trp to human and farm animals, and attempts to improve Lys level in crop plants, can be obtained from the following reviews as examples (Radwanski and Last in Plant Cell 7:921–934, 1995; Halkier and Gershenzon in Annu Rev Plant Biol 57:303–333, 2006; Ufaz and Galili in Plant Physiol 147:954–961, 2008; Jander and Joshi in Mol Plant 3:54–65, 2010).

Keywords Amino acids · Aspartate family pathway · Lysine · Aromatic amino acids · Bioinformatics · Plants

H. Less · R. Angelovici · V. Tzin · G. Galili (✉)
Department of Plant Sciences, The Weizmann Institute of Science, 76100 Rehovot, Israel
e-mail: Gad.Galili@weizmann.ac.il

Gene coordination—a novel bioinformatics tool adapted to analyzing complex interactions within and between gene expressions networks

Organisms generally live in changing environments and have to adjust their gene expression networks to such changes. Such adjustments are particularly important in plants because plants are sessile organisms that continuously interact with the environment. Classical bioinformatics tools, such as Pearson correlation, are not optimal to analyze the adjustments of system interactions between

plant gene networks in response of the dynamically changing environments because such system interactions may include both negative and positive expression correlations between pairs of genes in response to different cues. This is illustrated schematically in Fig. 1a, which depicts the potential compound regulation of two genes encoding two enzymes in a branch point of a given metabolic network (gene 1 of one branch and gene 2 of the second branch). Under some biological conditions, the plant may need to operate both branches and hence, under these conditions the expression of both gene 1 and gene 2 expects to be positively correlated (Fig. 1b, top panel). Under other biological conditions, the plant may need to operate only one of the two branches and shut down the other branch and hence, under such conditions the expression of both gene 1 and gene 2 expects to be negatively correlated (Fig. 1b, middle panel). Combining these two sets of biological conditions together shows no significant correlation between genes 1 and 2 (Fig. 1b, bottom panel).

To overcome this problem, we developed a novel bioinformatics tool, termed "Gene Coordination", which is adapted to analyzing the response of gene expression to different cues (for more detail see Less and Galili 2009). We then used this tool to analyze publicly available microarray expression data from several sources, containing hundreds of expression profiles representing over 200 different biological perturbations, most of which are associated with various stresses (Less and Galili 2009). The principle of this tool is illustrated in Fig. 1c, by comparing the expression response of two genes (*LKR/SDH* gene of Lys catabolism and another gene encoding a DNA-binding

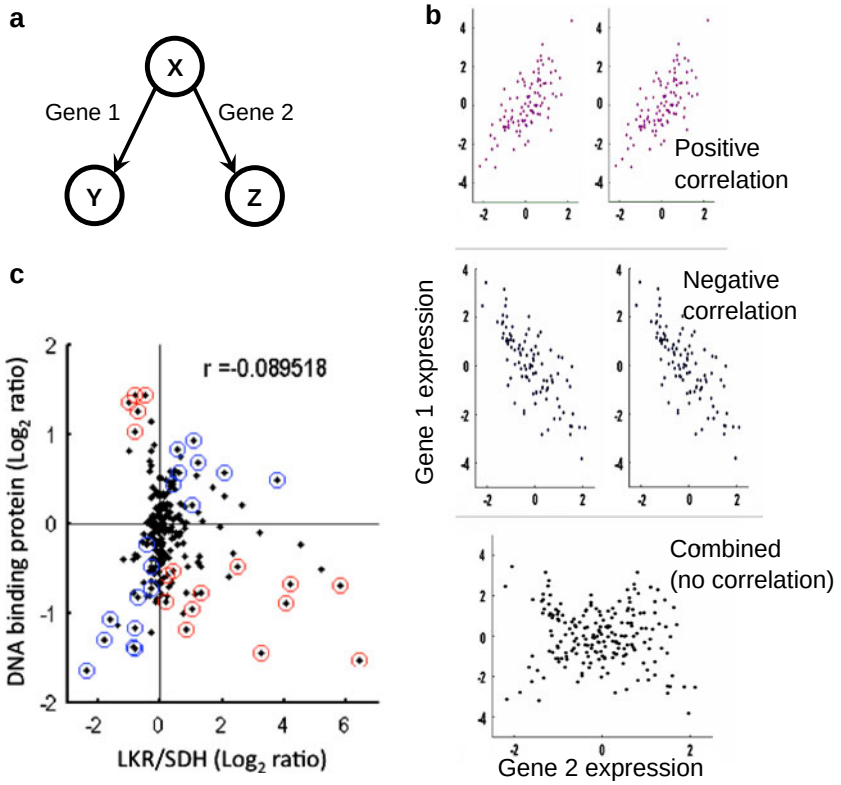

Fig. 1 Illustration of the gene coordination principal, using the LKR/SDH gene as an example. **a** Schematic representation of a metabolic branch point converting metabolite X either to metabolite Y or metabolite Z via the enzymes encoded by Gene 1 and Gene 2, respectively. **b** Illustration of the limitation of Pearson correlation for analyzing gene expression regulation associated with metabolic branch point, exemplified for the expression of a hypothetic Gene 1 versus Gene 2. Under some growth conditions, when both branches are active, the expression of Gene 1 and Gene 2 will be positively correlated (*top panel*). In contrast, under other growth conditions when only one branch is active and the second is suppressed, the expression of Gene 1 and Gene 2 will be negatively correlated (*middle panel*). Combination of the two growth conditions used in the top and middle panels will result in non-significant Pearson correlation (*bottom panel*). **c** Relationships between the expression differences of the Arabidopsis *LKR/SDH* gene and a second Arabidopsis gene encoding a DNA-binding protein, used as an example, across a set of 211 biological perturbations (for details of the biological perturbations see Less and Galili 2009). Each *black dot* indicates the expression difference (treatment vs. control) in response to a single specific biological perturbation. *Black dots inside red circles* indicate perturbations that contribute to a negative coordination, while *black dots inside blue circles* indicate perturbations that contribute to a positive coordination between each of the two compared genes. The Pearson correlation value is indicated on the *top* of the graph. (*Panel* **c** of this figure is derived with permission from Less and Galili 2009) (color figure in online)

protein) to the different biological perturbations. For every single biological perturbation, we analyzed the average change in expression level of each of these two genes (repression or stimulation of treatment vs. control; $\log_2$ ratio) over the entire set of biological perturbations by a statistical t test. Each black dot in Fig. 1c represents the response of the two genes to a single biological perturbation. The dots exhibiting statistically significant co-regulated expression (induction or repression) of the two genes in response to a single biological perturbation are encircled by blue circles, while the dots exhibiting oppositely regulated expression of the two genes in response to a single biological perturbation (one is induced and the second repressed) are encircled by red circles (Fig. 1c). As shown further in Fig. 1c, the expression response of the two genes to the entire set of biological perturbation shows no significant Pearson correlation coefficient ($r = -0.089518$). Yet, the two genes exhibit either significant "positive coordination" (co-regulated expression; blue dots) or significant "negative coordination" (opposite expression; red dots) in response to different subsets of the cues, which is larger than that calculated for a random background (Less and Galili 2009).

Transcriptional regulation of the Asp-family network exhibits two distinct expression signatures regulated by two oppositely expressed subsets of its genes

The Asp-family network (Fig. 2) consists of three major branches. The first branch leads to Lys biosynthesis and its further catabolism into Glu and acetyl CoA. The second branch leads to the biosynthesis of Met and its further conversion into (i) glucosinolates; (ii) Ile on route to energy production; and (iii) S-adenosyl Met (SAM) on route to the biosynthesis of polyamines, the hormone ethylene and also to a donation of methyl group to multiple other metabolites. The third branch leads to the biosynthesis of Thr and its further conversion into Gly as well as into Ile on route to energy production. Interestingly, the gene coordination analysis (Less and Galili 2009) identified two small groups of genes, encoding enzymes of the Asp-family network, which principally exhibit co-regulated expression (positive coordination) within each of the groups and oppositely regulated expression (negative coordination) between the two groups. Based on their regulated expression patterns, these genes were named as "highly coordinated genes" (HCGs) (Less and Galili 2009). The names and location of the enzymes encoded by these two groups of HCGs on the Asp-family metabolic network are indicated in gray and black boxes in Fig. 2. The first group of HCGs (Fig. 2, gray boxes) includes seven genes. Three of these genes encode biosynthetic enzymes, namely, aspartate kinase (AK; the

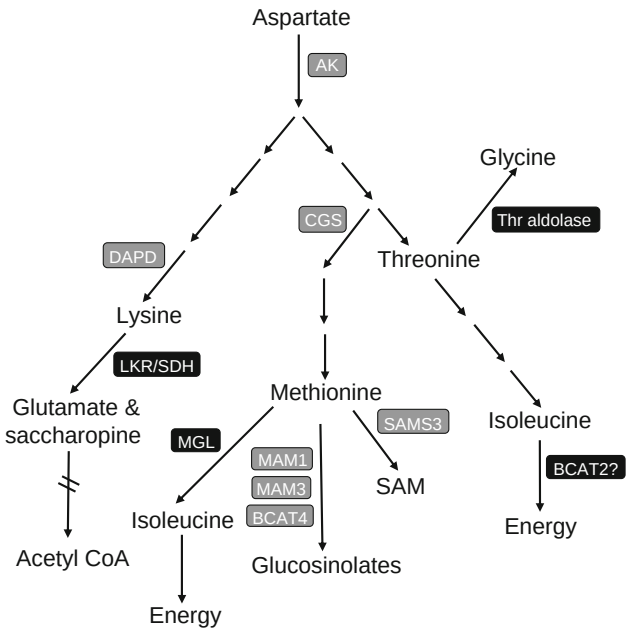

Fig. 2 Schematic representation of the Asp-family network and the two groups of highly coordinated genes within this network. The two groups of the highly coordinated genes are marked in *gray* and *black boxes*, respectively. Not all enzymes are indicated. *Arrows* represent single enzymatic steps, while *crossed arrows* represent several enzymatic steps. *AK* Asp kinase, *BCAT2* branched-chain amino acid transaminase isozyme 2, *BCAT4* enzyme originally defined as branched-chain amino acid transaminase isozyme 2, but has been proven to be a methylthioalkylmalate synthase isozyme, *CGS* cystathionine γ-synthase, *DAPD* diaminopimelate decarboxylase, *LKR/SDH* the bifunctional enzyme Lys ketoglutarate reductase/saccharopine dehydrogenase, *MAM1 & MAM3* methylthioalkylmalate synthase isozymes 1 & 3, *MGL* Met γ-lyase, *SAMS3* the isozyme S-adenosylmethionine synthase-3

entry enzyme of the Asp-family network); LL-diaminopimelate decarboxylase (DAPD) of Lys biosynthesis; and cystathionine γ-synthase (CGS) of Met biosynthesis. The rest four genes of this group encode three methylthioalkylmalate synthase isozymes that utilize Met for the synthesis of glucosinolates (MAM1, MAM3 and BCAT4) (Kroymann et al. 2001; Schuster et al. 2006; Textor et al. 2007) as well as one S-adenosylmethionine synthase 3 isozyme (SAMS3; Arabidopsis possesses four genes encoding SAMS isozymes) (Goto et al. 2002; Shen et al. 2002) that utilizes Met for the synthesis of polyamines and the hormone ethylene as well as for the donation of methyl groups for DNA replication and for the synthesis of a wide range of secondary metabolites, such as the cell wall component lignin (Fig. 2; enzymes indicated within gray boxes) (Kroymann et al. 2001; Schuster et al. 2006; Textor et al. 2007). The second group of HCGs (Fig. 2, black boxes) includes four genes encoding only catabolic enzymes, namely (i) the bifunctional lysine-ketoglutarate reductase/saccharopine dehydrogenase (LKR/SDH) enzyme that

catabolizes Lys into glutamate and acetyl CoA (Arruda et al. 2000) as well as into multiple additional, yet unknown, metabolites; (ii) Met γ-lyase that catabolizes Met into multiple metabolites, including methanethiol, alpha-keto-butyrate, S-methylmethionine, S-methylcysteine as well as into Ile, which is a major donor of energy (Rebeille et al. 2006; Joshi and Jander 2009); (iii) Thr aldolase that converts Thr into Gly (Joshi et al. 2006); and (iv) BCAT2 whose activity has not yet established experimentally, but based on the Plant Metabolic Network database (http://www.plantcyc.org:1555/ARA/NEW-IMAGE?type=NIL&object=AT1G10070) it is suggested to function in Ile catabolism toward energy production (Rebeille et al. 2006; Joshi and Jander 2009) (Fig. 2; enzymes located within black boxes). The Arabidopsis genes encoding LKR/SDH and BCAT2 belonging to the second group of HCGs (see above) are induced together by dehydration stress and by the hormone ABA (Urano et al. 2009).

The Asp-family network possesses extensive transcriptional interactions with genome-wide gene expression networks

Amino acid metabolism is among the most important and best recognized networks within biological systems. However, elucidation of the regulation of amino acid metabolism within the context of the entire system, particularly transcriptional regulation, is still in its infancy. To address this issue, we also used the gene coordination approach to discover system interactions between the expression of the HCGs of the Asp-family network (Fig. 2) and other functionally annotated genome-wide genes. This was achieved by selecting genome-wide genes that are either significantly positively or negatively coordinated with all of the genes within each one of the two groups of HCGs of the Asp-family network (Fig. 2). Using the publicly available gene ontology (GO) annotation (http://www.arabidopsis.org), we found that (i) the first group of HCGs of the Asp-family network (containing the biosynthetic genes and some of the catabolic genes) was generally positively coordinated with genes controlling growth-promoting processes, while negatively coordinated mostly with genes controlling stress-associated processes; and (ii) the second group of HCGs (containing only catabolic genes) was generally positively coordinated with stress-associated genes, while negatively coordinated with growth-associated genes (Less and Galili 2009). Hence, our bioinformatics study proposes that the Asp-family network possesses two distinct expression signatures; one includes increased expression of specific biosynthesis and catabolic genes and is associated with genome-wide transcriptional networks that control anabolic processes associated with active growth, while the second includes induction of some catabolic genes (others than the catabolic genes associated with the active growth), and is associated with the response of the plant metabolome to stress conditions. Since the LKR/SDH and BCAT2 enzymes are suggested to, respectively, convert Lys and Ile into the energy-associated TCA-cycle metabolites (Arruda et al. 2000; Rebeille et al. 2006; Joshi and Jander 2009), the stimulation of expression of their genes under stress conditions apparently contributes to the shortage of energy resulting from such stresses.

Interestingly, even though glucosinolates are classically associated with plant defense against herbivores (Halkier and Gershenzon 2006; Schuster et al. 2006), the genes of glucosinolates biosynthesis form Met belong to the first group of HCGs (Fig. 2), which exhibits the expression signature of anabolic processes associated with active plant growth (Less and Galili 2009). This may imply that the Met-derived glucosinolates are synthesized under favorable (non-stress) growth conditions either to be available in advance for defense against herbivores and pathogens when they attack the plant, or that the glucosinolates are also required for active plant growth. Elucidation of such potential new functions, if exist, of the glucosinolates is an interesting issue for future research.

To address further the system interaction of the Asp-family network with the global Arabidopsis transcriptome, we used maturing seeds as a developmental-inducible system to stimulate Lys synthesis and reduce its catabolism (a growth-associated pattern of Lys metabolism). This was achieved using our KD genotype, in which Lys biosynthesis is enhanced in a seed-specific manner upon expression of a bacterial feedback-insensitive dihydrodipicolinate synthase (DHDPS) under the control of a seed storage protein promoter, while Lys catabolism is inhibited by a T-DNA knockout mutation of the *LKR/SDH* gene of Lys catabolism (Angelovici et al. 2009). Notably, this manipulation of Lys metabolism was associated with increased expression of multiple genes associated with anabolic processes that stimulate plant performance and vigor (such as ribosomal proteins and translation initiation factors), while suppressing a small number of genes associated with plant stress interactions (Angelovici et al. 2009).

The aromatic amino acid network possesses relatively limited transcriptional interactions with genome-wide gene expression networks

Beside their incorporation into proteins, the three aromatic amino acids Phe, Tyr and Trp also serve in plants as precursors for a wide range of secondary metabolites having multiple functions (Herrmann 1995; Malitsky et al. 2008; Mène-Saffrané and Dellapenna 2009; Vogt 2009; Yatusevich

et al. 2009). We therefore applied our gene coordination approach to identify HCGs within the aromatic amino acid network and further elucidate their expression cross interaction with other functionally annotated genome-wide genes, as described above for the Asp-family network (Less and Galili 2009). Interestingly, the aromatic amino acids network possessed only a single group of HCGs and these also possessed expression cross interactions with a significantly smaller set of the genome-side genes, compared to the Asp-family network (Less and Galili 2009). This indicates that the Asp-family network has significantly stronger influence of the Arabidopsis transcriptome, and hence likely also a significantly stronger influence on the phenotypic response of the plants to various growth conditions, compared to the aromatic amino acids biosynthesis network.

We also tested the effect of enhancing aromatic amino acid accumulation on the Arabidopsis transcriptome. This was achieved using transgenic plants expressing an *Escherichia Coli PheA** gene encoding a bifunctional Chorismate mutase/Prephenate dehydratase (CM/PDT) enzyme of the bacterial Phe biosynthesis pathway, which converts Chorismate via Prephenate into phenylpyruvate (Tzin et al. 2009). The transgenic plants contained significantly higher levels of Phe and Tyr as well various secondary metabolites derived from these amino acids (Tzin et al. 2009). Microarray analysis showed that the increased accumulation of the aromatic amino acids and their secondary metabolites had a very minor influence, if at all, on the Arabidopsis transcriptome (Dubouzet et al. 2007; Tzin et al. 2009).

Conclusions

Our research indicates that the regulation of given metabolic networks of amino acids is executed by a group of genes encoding specific enzymes within each network, whose expression is highly coordinated within the individual metabolic networks. The expression of these genes is also highly coordinated with the expression of other genome-wide genes, enabling the synchronization of amino acid metabolism with the entire biological system of the plant. The highly coordinated genes of the Asp-family network possess more extensive intra-network as well as inter-network system interactions than the highly coordinated genes of the aromatic amino acids network, indicating a major involvement of the Asp-family network in plant growth and response to external cues.

References

Angelovici R, Fait A, Zhu X, Szymanski J, Feldmesser E, Fernie AR, Galili G (2009) Deciphering transcriptional and metabolic networks associated with lysine metabolism during Arabidopsis seed development. Plant Physiol 151:2058–2072

Arruda P, Kemper EL, Papes F, Leite A (2000) Regulation of lysine catabolism in higher plants. Trends Plant Sci 5:324–330

Dubouzet JG, Ishihara A, Matsuda F, Miyagawa H, Iwata H, Wakasa K (2007) Integrated metabolomic and transcriptomic analyses of high-tryptophan rice expressing a mutant anthranilate synthase alpha subunit. J Exp Bot 58:3309–3321

Goto DB, Ogi M, Kijima F, Kumagai T, van Werven F, Onouchi H, Naito S (2002) A single-nucleotide mutation in a gene encoding S-adenosylmethionine synthetase is associated with methionine over-accumulation phenotype in Arabidopsis thaliana. Genes Genet Syst 77:89–95

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57:303–333

Herrmann KM (1995) The Shikimate pathway: early steps in the biosynthesis of aromatic compounds. Plant Cell 7:907–919

Jander G, Joshi V (2010) Recent progress in deciphering the biosynthesis of aspartate-derived amino acids in plants. Mol Plant 3:54–65

Joshi V, Jander G (2009) Arabidopsis methionine gamma-lyase is regulated according to isoleucine biosynthesis needs but plays a subordinate role to threonine deaminase. Plant Physiol 151:367–378

Joshi V, Laubengayer KM, Schauer N, Fernie AR, Jander G (2006) Two Arabidopsis threonine aldolases are nonredundant and compete with threonine deaminase for a common substrate pool. Plant Cell 18:3564–3575

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T (2001) A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol 127:1077–1088

Less H, Galili G (2009) Coordinations between gene modules control the operation of plant amino acid metabolic networks. BMC Syst Biol 3:14

Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A (2008) The transcript and metabolite networks affected by the two clades of Arabidopsis glucosinolate biosynthesis regulators. Plant Physiol 148:2021–2049

Mène-Saffrané L, Dellapenna D (2009) Biosynthesis, regulation and functions of tocochromanols in plants. Plant Physiol Biochem. doi:10.1016/j.plaphy.2009.11.004

Radwanski ER, Last RL (1995) Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. Plant Cell 7:921–934

Rebeille F, Jabrin S, Bligny R, Loizeau K, Gambonnet B, Van Wilder V, Douce R, Ravanel S (2006) Methionine catabolism in Arabidopsis cells is initiated by a gamma-cleavage process and leads to S-methylcysteine and isoleucine syntheses. Proc Natl Acad Sci USA 103:15687–15692

Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006) Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in Arabidopsis. Plant Cell 18:2664–2679

Shen B, Li CJ, Tarczynski MC (2002) High free-methionine and decreased lignin content result from a mutation in the Arabidopsis S-adenosyl-L-methionine synthetase 3 gene. Plant J 29:371–380

Textor S, de Kraker JW, Hause B, Gershenzon J, Tokuhisa JG (2007) MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in arabidopsis. Plant Physiol 144:60–71

Tzin V, Malitsky S, Aharoni A, Galili G (2009) Expression of a bacterial bi-functional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in Arabidopsis. Plant J 60:156–167

Ufaz S, Galili G (2008) Improving the content of essential amino acids in crop plants: goals and opportunities. Plant Physiol 147:954–961

Urano K, Maruyama K, Ogata Y, Morishita Y, Takeda M, Sakurai N, Suzuki H, Saito K, Shibata D, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2009) Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. Plant J 57:1065–1078

Vogt T (2009) Phenylpropanoid biosynthesis. Mol Plant 3:2–20

Yatusevich R, Mugford SG, Matthewman C, Gigolashvili T, Frerigmann H, Delaney S, Koprivova A, Flugge UI, Kopriva S (2009) Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in Arabidopsis thaliana. Plant J (in press)