Linking Chemistry and Genomics for the Study of Secondary Metabolism in Aromatic and Medicinal Plants

Efraim Lewinsohn
Newe Ya'ar Research Center
Agricultural Research Organization
P.O. Box 1021
Ramat Yishay, 30095
Israel

Keywords: Metabolism, gene discovery, natural products, expressed sequence tags, functional expression

Abstract

The study of the biosynthesis of secondary metabolites and the genes involved in these processes has been greatly facilitated by novel genomic approaches developed during the last years. Many of the biosynthetic pathways dedicated to secondary metabolism, and the enzymes involved in these pathways, have apparently evolved from the much better studied primary biosynthetic pathways. Therefore by exploiting similarities between functionally-related genes, it has been possible to isolate novel genes involved in the formation of unique natural products. To implement this novel approach, appropriate tissues in the proper physiological state, where the compounds of interest are produced in significant levels, is identified. Next, sequence information on large numbers (thousands) of different ESTs (expressed sequence tags) originating in these tissues is obtained. The information obtained is *en masse* examined using bioinformatic computer algorithms. Predictions on the physiological and biochemical role of individual ESTs are then made based on DNA similarities, and the patterns of expression of individual ESTs. Identity and biochemical function of the particular EST in question can then be confirmed by functional expression experiments. A few examples of such genomic projects aimed at isolating and characterizing genes involved in the formation of key metabolites are reviewed. Some of the genes responsible for the formation of the volatile phenylpropenes prominent in the essential oil of sweet basil and in the formation of the many compounds that compose the fragrance of roses have been identified utilizing this approach. The potential of utilizing genes that code for the formation of volatile compounds, for the improvement of the quality properties of aromatic plants and other agricultural produce, are discussed.

INTRODUCTION

Plants produce and accumulate an enormous variety of secondary metabolites (often referred as "natural products"). Many of these products have well defined ecological roles in plant defense and in mediating the plant's interactions with other organisms, while others have yet unknown biological roles. More than 45,000 different chemical structures of natural products have been identified. The largest group of natural products is the terpenes, with more than 25,000 structures elucidated. Additionally, more than 2,000 alkaloids and about 8,000 phenolic derivatives are known (Croteau et al., 2000). Despite the large number of different proven chemical structures, there are an amazingly low number of biochemical pathways by which these compounds are biosynthesized.

This apparent inconsistency can be easily explained by the fact that plants produce all their metabolites starting from water, CO_2, a few minerals and solar energy. It is seemingly that during the course of evolution, the same basic biosynthetic pathways have developed to allow for the production of different metabolites, by only minor changes in the basic core of the pathways (Pichersky and Gang, 2000).

Proc. XXVI IHC – Future for Medicinal and Aromatic Plants
Eds. L.E. Craker et al.
Acta Hort. 629, ISHS 2004
Publication supported by Can. Int. Dev. Agency (CIDA)

CHEMICAL BIOSYNTHESIS

Terpenes are biosynthesized by the isopentenoid pathway, that includes two main metabolic branches, the mevalonic acid pathway, by which sesqui- and triterpenes are biosynthesized, and the deoxyxylulose diphosphate pathway,by which mono-, di-, and tetra-terpenes are formed (Croteau et al., 2000). Most phenolic derivatives are biosynthesized either by the shikimic acid pathway or through the malonate-acetate pathway and most alkaloids are biosynthesized from amino acids.

Monoterpenoids such as linalool, menthol, thymol, and limonene are derived from geranyl diphosphate (GPP, Fig. 1). The different conversions from GPP are catalyzed by a group of enzymes termed monoterpene synthases. These enzymes share many properties, such as cofactor requirements, molecular size and protein sequence similarity, but still they are different enough to allow for the catalysis of the different products according to their specificity (McGarvey and Croteau, 1995). Sesqui- and triterpenes, as well as many triterpene-derived metabolites such as the saponins and sterols, are synthesized from the 15-carbon intermediate farnesyl diphosphate (FPP).

Similarly to monoterpene synthases, minor changes in the sesquiterpene synthase enzymes (and their genes) are responsible for catalyzing the conversion of FPP to the different sesquiterpenes respectively. Still, at least in angiosperms, sesquiterpene synthase genes are significantly similar to each other, and also similar (to a lesser extent) to angiosperm monoterpene synthases (Bohlmann et al., 2000a).

In an analogous way, it seems that the ubiquitous pathway to lignin has been specifically adapted in many plants for the production of unique natural products, again, by small but important modifications of existing genes and enzymes, to allow for the specific conversions (Fig. 2). Thus, phenylpropanoids are mostly derived from L-phenylalanine, not only a precursor of lignin, anthocyanins and other flavonoids, but also, a precursor of *t*-anethole in anise (*Pimpinella anisum*) (Manitto et al., 1974a) and in fennel (*Foeniculum vulgare*) (Kaneko, 1960). L-phenylalanine is also a precursor of cinnamaldehyde in cinnamon (*Cinnamomum zeylanicum,* Lauraceae Senanayake et al., 1977*)*, and of estragole in sweet basil (*Ocimum basilicum,* Lamiaceae (Manitto et al., 1974b).

GENE CODING

At times, the same compound is present in many unrelated organisms. For example, eugenol, a phenylpropanoid compound possessing a very strong pungent aroma reminiscent of cloves, is indeed one of the major components of clove essential oil, but it is also present in lower levels in banana, cinnamon leaf, pimento and other plants (Senanayake et al., 1977, Tucker et al 1991, Jordan et al., 2001). At times, small chemical differences in the metabolites might cause dramatic differences in the biological activity or the aroma of the compound in question. For example, high levels of eugenol are detrimental to the aroma of tomato, rending it reminiscent of cloves. In contrast, its methoxylated derivative methyl eugenol, has a soft scent, reminiscent of cut grass (Fig. 3). Interestingly, the enzyme EOMT (SAM: eugenol O-methyltransferase is able to transfer a methyl group from S-adenosylmethionine (SAM) to release methyl eugenol, converting a pungent compound (eugenol) into a mild-scented one (methyl eugenol) (Fig. 3). This enzyme and its gene have been characterized in sweet basil varieties that accumulate methyl eugenol in their essential oil (Lewinsohn et al., 2000, Gang et al., 2001). The enzyme also accepts other substrates especially *para* substituted phenols. EOMT can also transfer a methyl group to the *p*-position of chavicol, releasing estragole, a major constituent of sweet basil essential oil. Thus, it seems that the same enzyme can produce either estragole and/or methyl eugenol, depending on the availability of the corresponding substrate acceptor.

The gene that encodes for the EOMT enzyme has been isolated and characterized. It belongs to a relatively large gene family that includes many *O*-methyltransferases, specific for diverse phenolic and flavonoid substrate acceptors. A very closely related gene, also isolated from sweet basil is 90 % similar to EOMT at the protein level (Gang et

al., 2002a). The enzyme encoded by this gene catalyzes the transfer of a methyl group to the *para* position of chavicol, but is inefficient in the transfer of a methyl group to eugenol, in contrast to EOMT that accepts both substrates and is even more efficient in transferring a methyl group to eugenol than to chavicol. In fact, site directed mutagenesis experiments indicated that the substitution of only one amino acid dramatically changed the substrate specificity of the enzymes with respect to either chavicol or eugenol. Substitution of a phenylalanine residue at the position 260 of CVOMT (chavicol *O*-methyltransferase) for a serine, resulted in a mutant enzyme that preferentially accepted eugenol as a substrate, effectively converting it into an eugenol-*O*-methyltransferase. Reciprocally, a substitution of a serine at position 261 of EOMT (eugenol-*O*-methyl-transferase) for a phenylalanine, resulted in a mutant enzyme that preferentially accepted chavicol as a substrate effectively converting it into a chavicol-*O*-methyltransferase. Thus it was possible to change the substrate specificity of the *O*-methyltransferases to accept the opposite substrate preference by changing only one amino acid in their sequence (Gang et al., 2002a) (Fig. 3).

Most *O*-methyltransferases studied are dimeric soluble enzymes that possess similar subunit molecular masses of approximately 40 KDa, utilize *S*-adenosylmethionine as a methyl donor, but differ in the efficiency towards acceptor substrates or the position of methylation. Moreover, genes that code for methyltransferases involved in primary metabolism, such as DNA methylation, also share common sequence motifs, conserved throughout evolution (Ibrahim and Muzac, 2000). These motifs probably reflect areas in the DNA that code for important common features of the enzymes such as the active sites or ligand-binding sites; while the non-conserved sequences often reflect the regions that determine the substrate specificity differing among the enzymes. The phenomenon of sequence conservation during gene evolution has been very useful in studies aimed at discovering new *O*-methyltransferase genes, acting on novel substrates (Ibrahim and Muzac, 2000, Lavid et al., 2002). The advent of new advances in high throughput methodologies, by which large number of genes can be examined, has allowed the discovery of many novel genes, using a similar approach (see below).

Genomics

By ‘genomics’ we usually designate novel high-throughput technologies, through which a large number of genes can be examined and characterized. The genome of a plant, animal or microbe is the totality of its genetic information including all the genes as well as the non-transcribed regions of the DNA.

This information has recently became available utilizing many novel tools that aim at representing a complete picture of the process being studied. Sequence information of complete genomes is currently available and these include not only information on 90 fully sequenced prokaryotic genomes, but also on eight fully sequenced eukaryotic genomes, including the mouse, and human genomes, as well as genomes of a plant, arabidopsis (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/org.html). The sequencing of the rice genome is close to completion. Full genomic information on other important plants such as maize, tomato, and *Medicago truncatula*, (an annual relative of alfalfa but with a genome size of only half that of alfalfa) are projected to be available in the near future.

Moreover, sequence information for more than 7,000,000 EST's (expressed sequence tags) is publicly available. EST’s are generated by massive sequencing of cDNA’s generated from the mRNA of the tissue of interest. EST’s are typically short (normally only partially represent the full-length clones) and are of relatively low sequencing quality. The majority of EST’s represent only the 3’ untranslated region of the respective genes, which often complicates their accurate functional annotation.

Nevertheless, ESTs are useful molecular landmarks. They provide a profile of the mRNA population at a particular time and offer a quick method for cloning a large number of genes known to be expressed in a particular cell population or tissue. In particular, useful information can be found by clustering EST’s and mRNAs on the basis

of sequence overlaps, to yield sequences that are longer and more accurate (contigs), and therefore, better represent the underlying genes. To date there is information on more than 300,000 EST's from barley (*Hordeum vulgare* subsp. *vulgare*), 290,000 EST's from soybean (*Glycine max*) 260,000 from wheat (*Triticum aestivum*), about 150,000 from tomato (*Lycopersicon esculentum*) and many other plants. In the case of medicinal and aromatic plants, EST information still lacks, although more than 20,000 are known from *Capsicum annuum* and 5,500 from *Stevia rebaudiana* (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

The promising potential of these technologies will undoubtedly yield increasing number of available also for other aromatic and medicinal plants. The information gained so far is readily searchable and includes the sequences of genes being expressed at different stages of development, different tissues, and different physiological states of different plants, and has enlightened our understanding of the expression patterns and the control of genes involved in simple physiological processes, biosynthetic pathways and other biological processes as well.

Not only information on the DNA, genome organization and genes expression patterns are being made available. Systematic modifications of existing technologies have been designed to allow for the processing of thousands of samples simultaneously and these new emerging technologies have effectively shortened the time and cut of analysis of thousands of samples.

The new developments include the use of microarrays, or cDNA chips, that allow monitoring the expression levels of tens of thousands of genes simultaneously; generating what is called 'transcriptome' profiles. Developments in automatization of two dimensional polyacrylamide gel electrophoresis, coupled to mass spectrometry adapted to proteins and other high-molecular weight molecules has resulted in a technique often described as 'proteomics', by which it is possible to monitor abundance patterns and often identify thousands of proteins simultaneously.

Additionally, developments in mass spectrometry has resulted in an approach often called 'metabolomics' by which the levels of thousands of low molecular weight metabolites including sugars, acids, and other volatile and non-volatile plant-constituents are monitored simultaneously (Fiehn et al., 2000, Summer et al., 2002, Aharoni et al., 2002).

Advances in Bio-informatics have provided many useful computer programs and algorithms to process the unprecedentedly high biological information accumulated. Interdisciplinary research that combines the use of all the above methodologies has greatly advanced our understanding of gene regulation and the processes that control physiological phenomena. In a way, these developments have allowed to examine biological processes utilizing a broader perspective, much difficult to achieve utilizing the conventional techniques. The simultaneous and concerted analyses of a large number of samples has corroborated much of previously known information, but also has provided much new information, almost impossible to obtain earlier. Assessment of this information has allowed us to conclude that many biosynthetic pathways to secondary metabolites are not only similar to each other but are probably evolutionary derived from primary metabolism. Moreover, taking advantage of the sequence similarities between known genes, it has been possible to identify new genes involved in primary and secondary metabolism with previously unknown functions.

In order to take optimal advantage of the high throughput technologies available to discover genes involved in secondary metabolism, a tissue where the genes of interest are expressed must be identified to serve as the source of genes. Sequence information on thousands of different EST's is generated. *En masse* sequencing methodologies have reduced the costs of this process. For example, mRNA obtained from isolated glandular trichomes from sweet basil, has been used to generate sequence information on more than 1,200 EST's expressed in these tissues, and this has been the key in identifying many genes involved in the biosynthesis of essential oil components of this plant (Gang et al 2001, 2002ab). Similar projects aimed at elucidating genes involved in secondary

metabolism from many plants are currently been pursued. They include studies in peppermint (Lange et al., 2000, Lange and Ketchum, 2002), rose petals (Lavid et al., 2002, Scalliet et al., 2002, Channeliere et al., 2002, Guterman et al., 2002), strawberry (*Fragaria x annanassa)* (Aharoni et al., 2000*), Bixa orellana* (Jako et al., 2002), *Stevia rebaudiana* (Brandle et al., 2002), tea tree (*Melaleuca alternifolia)* (Shelton et al., 2002), and opium poppy (*Papaver somniferum*) (Kutchan, 2002) among others. These studies have corroborated previous evidence on the involvement of specific genes and biosynthetic pathways in the formation of mono- di- and tetra terpenes, volatile phenolics alkaloids, and other secondary metabolites, but have also uncovered genes coding for novel methyltransferases, hydroxylases, acetyltransferases, sesquiterpene synthases and many other genes involved in the formation of natural products.

Bioinformatic methodologies are crucial in compiling and processing the information needed for the initial identification of the EST's obtained (see Koltai and Volpin, 2002). The sequence information generated is massive and to be efficient has to be examined with the aid of powerful computer programs. Utilizing specific programs and algorithms, combined with the growing and efficient database mining methodologies, many putative biochemical functions are assigned to thousands of individual EST's, based on similarity to known genes and their expression patterns.

Nevertheless, the final confirmation of EST identity and function is normally made by functional expression utilizing homologous or heterologous systems. Although for most cases, similarity with known genes is sufficient to predict the role of a new available gene with a great degree of accuracy, this process is not always simple or trivial, especially when annotating genes involved in secondary metabolism. For example two putative caffeic acid *O*-methyltransferases, (predicted based on sequence similarity), isolated from arabidopsis and barley were overexpressed in *E. coli,* the gene product possessed *O*-methyltransferase activity, but caffeic acid was a poor substrate in both cases. The Arabidopsis enzyme primarily accepted apigenin and other flavonoids, methylating at the 7 position, while the barley enzyme primarily accepted quercetin and other flavonol substrates, generating 3'-methylated products (Ibrahim and Muzac, 2002).

In a parallel way, an Arabidopsis gene putatively identified as "limonene synthase" based on similarity with the spearmint and other limonene synthase genes was functionally identified as a monoterpene synthase, but mainly producing myrcene and β-ocimene (as well as low levels of limonene and other cyclic monoterpenes) in vitro, (Bohlmann et al., 2000b). Similarly, a putative sesquiterpene synthase was isolated from rose petals; based on similarity with the known cotton δ-cadinene synthase gene and other sesquiterpene synthase genes.

However, the gene-product of this gene, produced the sesquiterpene germacrene D when heterologously expressed in *E. coli* and fed with the appropriate substrate farnesyl diphosphate (FPP) (Guterman et al., 2002)

The above examples point to the necessity to functionally express the putative genes in order to verify/assess their biochemical function. Moreover, these functions represent those obtained in vitro under specific conditions and not necessary reflect the real biochemical role of the specific EST in vivo, since other factors such as pH, substrate availability, and other cofactors might have an influence on the product obtained. If post-translational modifications are needed to gain the respective biochemical function, ectopic expression in eukaryotic cells, such as yeast or plants can also be used. Additionally the biological function of the EST in question can also be determined in experiments aiming at silencing the gene in the original plant utilizing anti-sense, co-suppression or RI constructs.

Concluding, the utilization of high throughput analyses, including genomics, transcriptomics, proteomics, and metabolomics, coupled to bioinformatic tools and database mining allow us to make predictions on the biochemical roles of thousands of individual EST's. Confirmation of the identity and function of the gene involved is performed by functional expression in heterologous or homologous systems. Utilizing an EST data bases generated from basil glandular trichomes, two novel OMT genes were

isolated, acting on chavicol and eugenol respectively (Gang et al., 2001, 2002a).

In a similar fashion novel genes coding for related *O*-methyltransferases involved in the formation of volatile phenolic ethers emitted by rose flowers have also been characterized (Scalliet et al., 2002, Lavid et al., 2002). The potential ofutilizing this technique for the discovery of novel genes involved in secondary metabolism is very promising.

Genetic Engineering to Modifying Aromas of Agricultural Produce

A few of the genes involved in the formation of aroma chemicals and other secondary metabolites have recently been identified. These genes have been key in assessing the utilization of genetic engineering to modify the aromas of agricultural produce, including fruits, flowers, and aromatic plants. Usually genetic engineering is limited to the transfer of one or a few genes, and therefore, it is of importance that sufficient levels of precursors will be available for the process to succeed. Thus, genetic engineering presents an elegant solution to the restoration and augmentation of original aromas to agricultural produce and to improve other quality traits such as color, vitamin content and other with minimal interference with other traits (Galili et al., 2002).

Monoterpenes are key determinants of the aromas of many aromatic plants, vegetables and fruits. Linalool is an acyclic monoterpene alcohol that imparts an aroma with a sweet floral alcoholic note. Linalool is a major component of the scent of many flowers (Dobson, 1993; Knudsen et al., 1993) and is also present in many edible fruits, such as guava, peach, plum, pineapple, kubo and passion fruit (Bernreuther and Schreier, 1991, Ninio et al., 2002). Linalool is a chiral compound, naturally appearing in two forms (*S*- and *R*-linalool) that slightly differ in their aroma.

The enzyme that catalyzes the formation of *S*-linalool from the ubiquitous precursor geranyl diphosphate (Fig. 1) has been purified (Pichersky et al., 1995), and its gene (*LIS*) isolated from the flowers of a small Californian annual plant *Clarkia breweri* (Dudareva et al., 1996). This gene is a promising candidate for future attempts to manipulate monoterpene metabolism in transgenic plants and has been used in attempts to modify the scents of lowers and the aroma of tomatoes (Lavy et al., 2002, Lucker et al., 2001, Lewinsohn et al., 2001).

Many modern tomato varieties have impaired aromas, due to lack of many common volatiles, such as linalool, present in the older tomato varieties. The *Clarkia LIS* gene, under the control of the tomato late-ripening specific promoter *E8,* has been transformed into tomatoes and this has resulted in fruits that produce *S*-linalool (Fig. 4). Unexpectedly, the expression of *LIS* also caused the accumulation of 8-hydroxylinalool, a linalool derivative possibly produced by allylic hydroxylation of linalool via an unknown endogenous enzyme (Fig. 4). The levels of other terpenoids present in high levels in tomato fruits, such as lycopene, other carotenoids and tocopherols remained essentially unaffected by the transgenesis, since only a small fraction of the metabolic flow through the terpenoid pathway was diverted into the production of linalool and 8-hydroxylinalool (Lewinsohn et al., 2001). These studies also illustrate a case in which the introduction of one gene can cause the production of unexpected novel metabolites in the transgenic plants. This was yet sufficient to modify the aroma, since the threshold levels for the perception of linalool are very low (6 ppb) (Buttery et al., 1971). In fact, many other volatiles also have very low thresholds of detection (Buttery et al., 1971) and therefore, aroma enhancement may be relatively easily achieved by diverting only a small fraction of the existing metabolic flow to production of volatiles, with negligible perturbation of the general metabolism of the plant.

The potential of genetic engineering for the improvement of aroma and taste properties of agricultural products is just beginning to be investigated. With the advent of more genes encoding for key enzymes involved in the production of volatile aroma chemicals, the potential to utilize genetic engineering for the manipulation of crops is very promising.

CONCLUSION

Aromatic Plants have for generations been an important source of aroma and flavor compounds, and that is one of the main reason they are collected and cultivated worldwide. The potential use of aromatic and medicinal plants as a source of nutraceuticals and medicinal compounds has also been well recognized. By combining the genomic methodologies, and utilizing the tools of modern biotechnology, a new key role of aromatic and medicinal plants has become evident. Aromatic plants can also be an unlimited source of genes that affect the flavor, and quality of our foods and other agricultural produce. The vast biotechnological potential of such genes has started to be noted and utilized in the last years to obtain agricultural products with improved quality traits without compromising other important agronomical traits (Galili et al 2002). The biotechnological utilization of these genes and the information and the tools obtained are not only used for genetic engineering, they are also key in aiding conventional breeding programs, aimed at improving the aroma of flowers, vegetables and fruits. In order to manipulate the levels of desired metabolites, it is important to understand the biosynthetic pathways involved in scent metabolites formation. Most natural products require the involvement of many biosynthetic steps for their biosynthesis, but some natural products are biosynthesized by only one or a few enzymes from ubiquitous precursors. Due to the fact that small chemical modifications of the fragrance molecules (position of ring closure, degree of saturation, oxidation, reduction, methylation, etc.) often have profound effects in the aroma characteristics of volatile compounds (Croteau and Karp, 1991), it is possible to limit gene transfer to one or two genes. Therefore, the implementation of genetic engineering to modify the level and composition of aroma metabolites produced in plants is very promising. Compounds of commercial interest, available only from exotic species or in limited quantities, can be potentially genetically engineered into domesticated, agronomically developed crops (Galili et al., 2002). Moreover, it is conceivable that using recombinant DNA technology, it will be possible to produce expensive fragrances in genetically engineered microorganisms.

Several research approaches that combine recent developments in biology and genetics are currently available for the mass isolation and characterization of genes that code for the formation of fragrances. The isolation of genes coding for respective key enzymes has opened new options in attempting the genetic manipulation of the biosynthetic pathways. In essence, using genetic engineering, it is possible to change the aroma properties of fruits, and flowers by introducing genes form other organisms, changing the expression pattern of existing genes, and repressing the expression of endogenous genes, thus diverting the flow of metabolite to desired directions (Lewinsohn et al., 1996, Galili et al., 2002). Still, there are many obstacles that need to be addressed before the full potential of these novel technologies can be implemented. Most of the obstacles might be overcome as a result of more research aimed at solving technical and fundamental problems, while others involve patents, regulation and other legal issues, as well as public acceptance. The gathering of knowledge related to the molecular genetics, the regulation and biochemistry of the pathways to natural products in aromatic, medicinal and other plants has been greatly facilitated by genomics. In the next decades we will most likely see many examples of the implementation of this knowledge for human benefit.

Literature Cited

Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., Van Houwelingen, A.M.M.L., De Vos, R.C.H., Van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., Van Tunen, A.J. and O'Connel, A.P. 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell, 12:647-661.

Aharoni, A.R., de Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R. and Goodenowe, D.B. 2002. Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. Omics, A Journal of Integrative Biology 6:217-234

Bernreuther, A. and Schreier, P. 1991. Multidimensional gas chromatography/mass spectrometry: a powerful tool for the direct chiral evaluation of aroma compounds in plant tissues. II. Linalool in essential oils and fruits. Phytochemical Analysis 2:167-170.

Bohlmann, J., Gershenzon, J. and Aubourg, S. 2000a. Biochemical, molecular, genetic and evolutionary aspects of defense-related terpenoid metabolism in conifers. In: J.T. Romeo, R. Ibrahim, L. Varin, eds, Evolution of Metabolic Pathways. Recent Advances in Phytochemistry vol 34, Elsevier Science, Oxford, pp. 109-150.

Bohlmann, J., Martin, D., Oldham, N.J. and Gershenzon, J. 2000b. Terpenoid secondary metabolism in Arabidopsis thaliana: cDNA cloning, characterization, and functional expression of a myrcene/(*E*)-beta-ocimene synthase Archives of Biochemistry and Biophysics 375:261-269.

Brandle, J.E., Richman, A., Swanson A.K. and Chapman B.P. 2002. Leaf ESTs from *Stevia rebaudiana*: a resource for gene discovery in diterpene synthesis. Plant Molecular Biology 50:613–622.

Buttery, R.G., Siefert, R.M., Guadagni, D.G. and Ling, D.G. 1971. Characterization of additional volatile components of tomato. J. Agric. Food. Chem. 19:524-529.

Channeliere, S., Riviere, S., Scalliet, G., Jullien, F., Szecsi, J., Dolle, C., Vergne, P., Dumas, C., Bendahmane, M., Hugueney, P. and Cock, J.M. 2002. Analysis of gene expression in rose petals using expressed sequence tags. FEBS Letters 515:35-38.

Croteau, R. and Karp, F. 1991. Origin of natural odorants. In PM Muller, D Lamparsky eds, Perfumes. Art, Science and Technology. Elsevier Applied Science. London, pp. 101-126.

Croteau, R., Kutchan, T.M. and Lewis, N.G. 2000. Natural products (secondary metabolites). In: B. Buchanan, W. Gruissem and R. Jones (eds), Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists. Rockville, MD, pp. 1250-1318.

Dobson, H.E.M. 1993. Floral volatiles in insect biology. In: E. Bernays, ed, Insect-Plant Interactions, Vol. 5, CRC Press, Boca Raton FL. pp. 47-81.

Dudareva, N., Cseke, L., Blanc, V.D. and Pichersky, E. 1996. Evolution of floral scent in *Clarkia*: Novel patterns of *S*-linalool synthase gene expression in the *C. breweri* flower. Plant Cell 8:1137-1148.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. 2000. Metabolite profiling for plant functional genomics. Nat. Biotechnol. 18:1157-1161.

Galili, G., Galili, S., Lewinsohn, E. and Tadmor, Y. 2002. Genetic, molecular, and genomic approaches to improve the value of plant foods and feeds. Critical Reviews in Plant Sciences 21:167–204.

Gang, D.R., Wang, J., Dudareva, N., Nam, K.H., Simon, J., Lewinsohn, E. and Pichersky, E. 2001. An investigation of the storage and biosynthesis of phenylpropenes in sweet basil (*Ocimum basilicum* L.). Plant Physiology 125:539-555.

Gang, D., Lavid, N., Zubieta, C., Cheng, F., Beuerle, T., Lewinsohn, E., Noel, J.P. and Pichersky, E. 2002a. Characterization of phenylpropene *O*-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant *O*-methyltransferase family. Plant Cell 14:505-519.

Gang, D.R., Beuerle, T., Ullmann, P., Werck-Reichhart, D. and Pichersky E. 2002b. Differential production of *meta* hydroxylated phenylpropanoids in sweet basil peltate glandular trichomes and leaves is controlled by the activities of specific acyltransferases and hydroxylases. Plant Physiol. 130:1536-1544.

Guterman, I., Shalit, M., Menda, M., Piestun, D., Dafny-Yelin, M., Shalev, G., Bar, E., Davydov, O., Ovadis, M., Emanuel, M., Wang, J., Adam, Z., Pichersky, E., Lewinsohn, E., Zamir, D., Vainstein, A. and Weiss, D. 2002. Rose Scent: Genomics approach to discovering novel floral fragrance-related genes. Plant Cell 14:2325-2338.

Ibrahim, R.K. and Muzac, I. 2000. The methyltransferase gene superfamily: a tree with multiple branches. In: J.T. Romeo, R. Ibrahim, L. Varin and V.D. Luca (Eds.), Evolution of Metabolic Pathways, Elsevier Science Ltd, , pp. 349-364.

Jako, C., Coutu, C., Roewer, I., Reed, D.W., Pelcher, L.E. and Covello, P.S. 2002.

Probing carotenoid biosynthesis in developing seed coats of *Bixa orellana* (Bixaceae) through expressed sequence tag analysis Plant Science 163:141-145.
Jordan, M.J., Tandon, K., Shaw, P.E. and Goodner, K.L. 2001. Aromatic profile of aqueous banana essence and banana fruit by gas chromatography-mass spectrometry (GC-MS) and gas chromatography-olfactometry (GC-O). J. Agric. Food Chem. 49:4813-4817.
Kaneko, K. 1960. Biogenetic studies of natural products. V. Biosynthesis of anethole by *Foeniculum vulgare.* Chem Pharm Bull (Japan) 8:875-879.
Knudsen, J.T., Tollsten, L. and Bergstrom, G. 1993. Floral scents - a check list of volatile compounds isolated by headspace techniques. Phytochemistry 33**:**253-280.
Koltai, H. and Volpin, H. Agricultural genomics: an approach to plant protection. 2002. Eur. J. Plant Pathol. (In Press).
Kutchan, T.M. 2002. Sequence based approaches to alkaloid biosynthesis gene identification. In: Phytochemistry in the Genomics and Post Genomics Eras. Romeo, J.T. and Dixon, R.A. (eds.) Recent Advances in Phytochemistry vol 36. Pergamon, Amsterdam, pp. 163-178.
Lange, B.M. and Ketchum, E.B. 2002. Functional genomics approaches to unravel essential oil biosynthesis. In: Phytochemistry in the Genomics and Post Genomics Eras. Romeo, J.T. and Dixon, R.A. (eds.), Recent Advances in Phytochemistry vol 36. Pergamon, Amsterdam, pp. 145-162.
Lange, B.M., Wildung, M.R., Stauber, E.J., Sanchez, C., Pouchnik, D. and Croteau, R. 2000. Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. Proc. Natl. Acad. Sci. USA. 97:2934-2939.
Lavid, N., Wang, J., Shalit, M., Guterman, I., Bar, E., Beuerle, T., Menda, N., Shafir, S., Zamir, D., Adam, Z., Vainstein, A., Weiss, D., Pichersky, E. and Lewinsohn, E. 2002. *O*-Methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals. Plant Physiol 129:1899-1907.
Lavy, M., Zuker, A., Lewinsohn, E., Larkov, O., Ravid, U., Vainstein, A. and Weiss, D. 2002. Linalool and linalool oxide production in transgenic carnation flowers expressing the *Clarkia breweri* linalool synthase gene. Molecular Breeding 9:103-111.
Lewinsohn, E. 1996. Molecular biology for the improvement of medicinal and aromatic plants. In: L.E. Craker, L. Nolan, K. Shetty, eds, Proceedings of the Intl. Symposium on Medicinal and Aromatic Plants. Amherst, MA, USA, Acta Hort. 426:443-466.
Lewinsohn, E., Ziv-Raz, I., Dudai, N., Tadmor, Y., Lastochkin, E., Larkov, O., Chaimovitsh, D., Ravid, U., Putievsky, E., Pichersky, E. and Shoham, Y. 2000. Biosynthesis of estragole and methyl-eugenol in sweet basil (*Ocimum basilicum* L). Developmental and chemotypic association of allylphenol *O*-methyltransferase activities. Plant Science 160:27-35.
Lewinsohn, E., Schalechet, F., Wilkinson, J., Matsui, K., Tadmor, K., Nam, K.H., Amar, O., Lastochkin, E., Larkov, O., Ravid, U., Hiatt, W., Gepstein, S., and Pichersky, E. 2001. Enhanced levels of the aroma and flavor compound *S*-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. Plant Physiology 127:1256-1265.
Lücker, J., Bowmeester, J., Schwab, W., Blaas, J., van der Plas, L.H.W. and Verhoeven, H.A. 2001. Expression of *Clarkia S*-linalool synthase in transgenic petunia plants results in the accumulation of S-linalyl-β-D-glucopyranoside. Plant J 27:315-324.
Manitto, P., Monti, D. and Gramatica, P. 1974a. Biosynthesis of anethole in *Pimpinella anisum* L. Tetrahedron Lett 17:1467-1568.
Manitto, P., Monti, D. and Gramatica, P. 1974b. Biosynthesis of phenylpropanoid compounds. Part I. Biosynthesis of eugenol in *Ocimum basilicum*. JCS. Perkin I. 1727-1731.
McGarvey, D.J. and Croteau, R. 1995. Terpenoid metabolism. Plant Cell 7:1015-1026.
Ninio, R., Lewinsohn, E., Mizrahi, Y. and Sitrit, Y. 2002. Changes in sugars, acids and volatiles during ripening of koubo (*Cereus peruvianus* [L.] Miller) fruits. J. Agric. Food Chem. (In Press).

Pichersky, E. and Gang, D.R. 2000. Genetics and biochemistry of specialized metabolites in plants: an evolutionary perspective. Trends in Plant Sci. 5:439-445.
Pichersky, E., Lewinsohn, E. and Croteau, R. 1995. Purification and characterization of *S*-linalool synthase, an enzyme involved in the production of floral scent in *Clarkia breweri*. Arch. Biochem. Biophys. 316**:**803-807.
Scalliet, G., Journot, N., Jullien, F., Baudino, S., Magnard, J.L., Channeliere, S., Vergne, P., Dumas, C., Bendahmane, M., Cock, J.M. and Hugueney, P. 2002. Biosynthesis of the major scent components 3,5-dimethoxytoluene and 1,3,5-trimethoxybenzene by novel rose O-methyltransferases. FEBS Lett. 523:113-118.
Senanayake, U.M., Wills, R.B.H. and Lee, T.H. 1977. Biosynthesis of eugenol and cinnamic aldehyde in *Cinnamomum zeylanicum*. Phytochemistry 16:2032-2033.
Shelton, D., Leach, D., Baverstock, P. and Henry R. 2002. Isolation of genes involved in secondary metabolism from *Melaleuca alternifolia* (Cheel) using expressed sequence tags (ESTs). Plant Science 162:9–15.
Summer, L.W., Duran, A.L., Huhman, D.V. and Smith, J.T. 2002. Metabolomics: a developing and integral component in functional genomic studies of *Mendicago truncatula.* In: Phytochemistry in the Genomics and Post Genomics Eras. Romeo J.T. and Dixon R.A. (eds.) Recent Advances in Phytochemistry Vol. 36. Pergamon, Amsterdam, pp. 31-61.
Tucker, A.O., Maciarello, M.J. and Landrun L.R. 1991. Volatile Leaf Oils of Caribbean Myrtaceae. II *Pimenta dioica* (L.) Merr. of Jamaica J. Essent. Oil Res. 3:195-196.

Figures

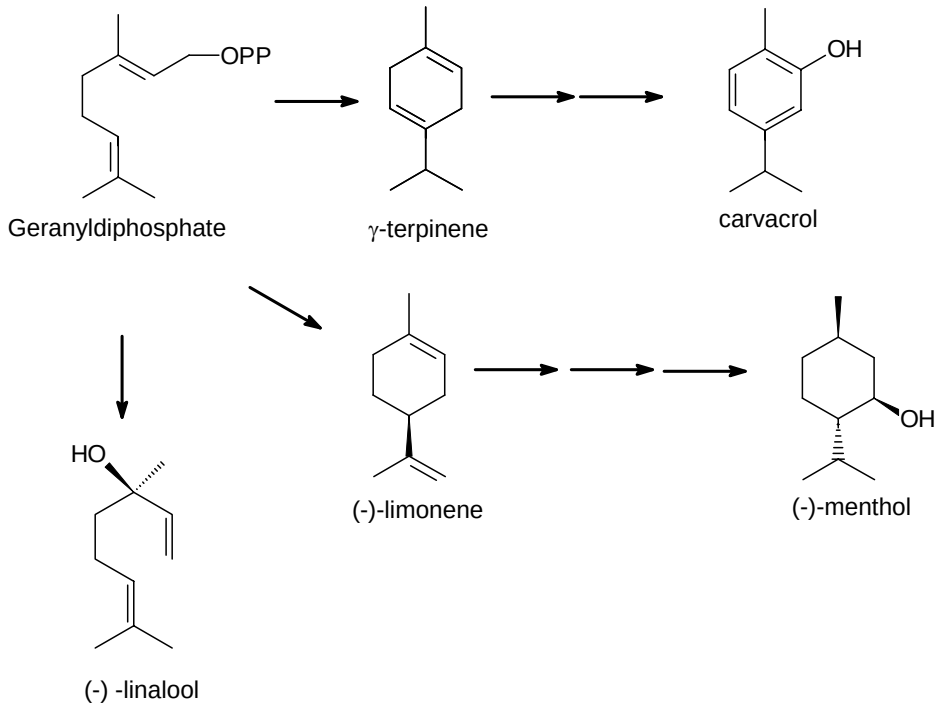

Fig. 1. All monoterpenes such as carvacrol, menthol and linalool are biosynthesized in plants from the precursor geranyldiphosphate (= geranylpyrophosphate, GPP).

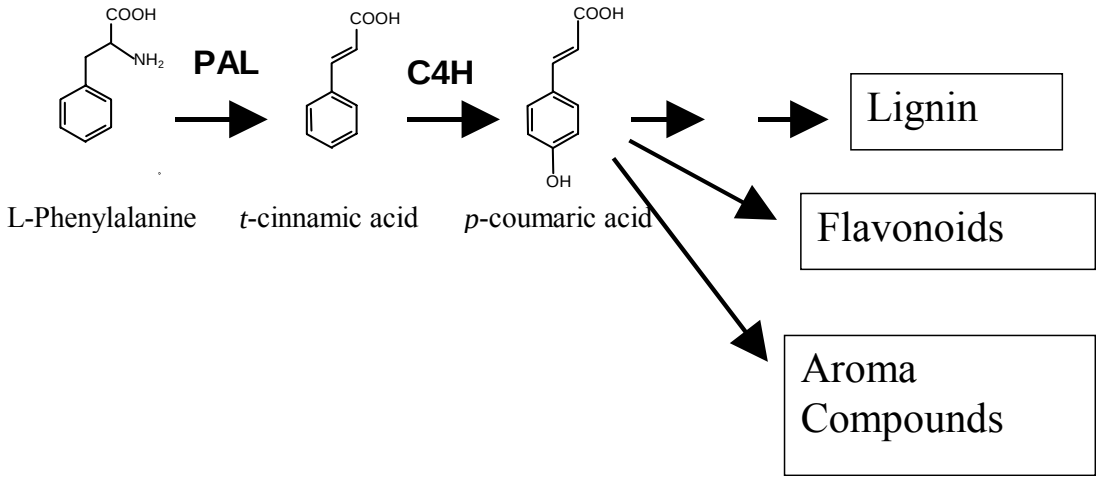

Fig. 2. The phenylpropanoid pathway. L-Phenylalanine is not only a precursor to lignin, anthocyanins and other flavonoids in plants. Many aroma compounds such as estragole, *t*-anethole, cinnamaldehyde and eugenol are also formed from L-phenylananine. The pathway and enzymes involved are often similar. PAL: phenylananine ammonia lyase, C4H: cinnamate 4-hydroxylase.

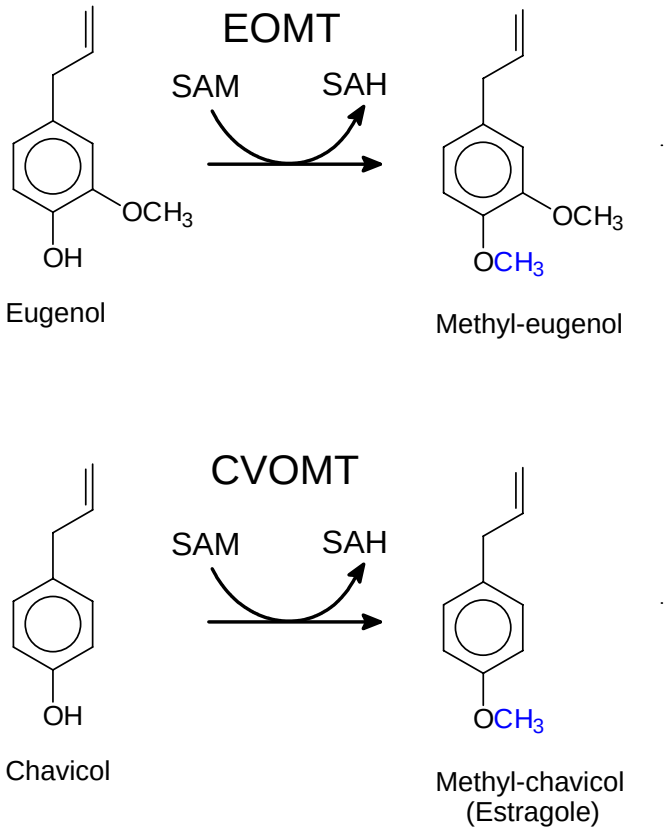

Fig. 3. The *p*-methylation of volatile phenylpropenes by specific O-methyltransferases. EOMT readily methylates eugenol (the pungent principle of cloves) to methyl-eugenol, a compound with a mild grassy smell. It is less efficient in methylating chavicol to methyl-chavicol (estragole). CVOMT efficiently methylates chavicol to generate methylchavicol, but is inefficient in methylating eugenol. Genes that code for these closely related enzymes have been isolated from sweet basil utilizing mRNA generated from glandular trichomes and a genomics EST-approach coupled to functional expression to identify the genes and their products. (Gang et al., 2001). Site directed mutagenesis experiments have indicated that it was possible to change the substrate specificity of the *O*-methyl transferases to accept the opposite substrate preference by changing only one amino acid in their sequence (Gang et al., 2002a).

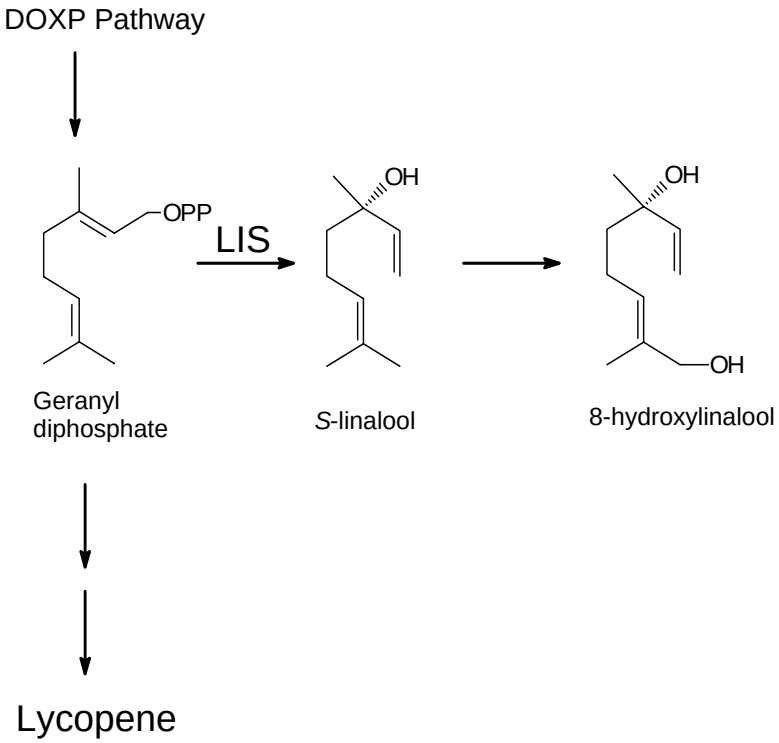

Fig. 4. Modification of the aroma in tomato fruits by genetic engineering utilizing a floral gene. Tomato fruits produce high levels of lycopene via the deoxyxylulose phosphate pathway. GPP is an intermediate in this pathway, but also a precursor to the scent compound linalool. Ripe tomato fruits were engineered to over express the *Clarkia breweri S*-linalool synthase gene (*LIS*) under the control of the *E8* ripe-fruit specific promoter. Fruits over expressing the *LIS* gene accumulated *S*-linalool and 8-hydroxylinalool, with no noticeable effect on other terpenoids present in the fruits The accumulation of s-linalool is probably due to the effect of an endogenous, yet unidentified enzyme (Lewinsohn et al., 2001).