2007

In Silico Analysis of Terpene Synthase Genes in Arabidopsis Thaliana

Sam Zwenger

Chhandak Basu
University of Northern Colorado

Follow this and additional works at: http://digscholarship.unco.edu/biofacpub

Part of the Botany Commons

Case report:

IN SILICO ANALYSIS OF TERPENE SYNTHASE GENES IN *ARABIDOPSIS THALIANA*

Sam Zwenger, Chhandak Basu[*]

University of Northern Colorado, School of Biological Sciences, Ross Hall 2480, Box 92
501 20 St., Greeley, CO 80631, United States, Tel: +1-970-351-2716, FAX: +1-970-351-2335,
e-mail: chhandak.basu@unco.edu (*corresponding author)

ABSTRACT

Terpenes are defense chemicals found in wide groups of plants. Terpenoids play a large role in plant development and stress response. The terpene synthase family comprises a diverse set of genes, all which contribute to production of terpenoids. We have used tools of bioinformatics and performed an *in silico* analysis of developmental and tissue specific terpene synthase gene expression in *Arabidopsis thaliana*, as well as those expressed due to biotic and abiotic environmental stimuli. Using software tools from Genevestigator, a powerful microarray analyzer, we used multiple tool sets to better understand terpene synthase expression in *Arabidopsis*, which will hopefully open the genetic door to further wet laboratory investigations. The data can be used to predict roles of terpene synthase genes in plant cell division and growth. The data presented here can be used to model for terpene synthesis expression in other plant species and can also be used to integrate basic plant physiology, and 'omics' disciplines.

Keywords: *Arabidopsis thaliana*, terpene synthase, Genevestigator

INTRODUCTION

Terpenoids are the largest class of plant natural products (Trapp and Croteau, 2001) and more than 30,000 terpenoid compounds have been identified (Buckingham, 1998). Terpenoids like essential oils and resins have commercial and industrial values (Zinkel and Russell, 1989; Dawson 1994). The main function of terpenes and their derivatives is widely recognized in plant defense mechanism. Some terpenes are of interest to biotechnological applications. For instance, terpenes can be released by plants after being attacked by herbivorous mites and insects, which in turn act as chemical attractants for predatory arthropods (Degenhardt et al., 2003). However, it is also known that terpene synthesis is involved not only in secondary metabolites, but primary metabolites as well (Trapp and Croteau, 2001). Additionally, terpenes have been proposed to play a role in attracting pollinating insects in *Arabidopsis* flowers (Chen et al., 2003). Plant terpenes have also been used as models for phylogenetic studies (Trapp and Croteau, 2001).

Many terpene synthase genes have been deposited in GenBank from various sources (Altschul et al., 1997). Cytochrome P450 monooxygenases (P450s) participate in oxidation reactions in terpene biosynthesis (Whitbred and Schuler, 2000). A PubMed (http://www.ncbi.nlm.nih.gov/sites/entrez) search yields more than fifty terpene synthase genes in *Arabidopsis*. The diversity in

location and size of terpene synthase genes in *Arabidopsis* poses significant challenges when doing genetic analysis for genomic expression assays.

To better cope with the accumulation of large amounts of genetic data, which has been deposited in databases, new and powerful software tools have been created. This includes such software as Genevestigator, which is capable of performing large assays on thousands of microarrays giving repeatable and reliable results (Zimmermann et al., 2004).

Using the microarray data via Genevestigator, we present developmental, tissue specific, and stimuli-induced terpene synthase gene expression in *Arabidopsis thaliana*.

MATERIALS AND METHODS

The web based Genevestigator software was used for data analysis purposes following developer's protocol at http://www.genevestigator.ethz.ch/. After establishing a user account the "start analysis tool" button was selected in the lower center area of the home page. After entering the account username and password the Genvestigator V3 program appears on the screen. By clicking the green "+" button the "Array Selection" screen appears. Users may choose from organisms including *Mus musculus*, *Rattus norvegicus*, *Hordeum vulgare*, or, as in our analysis, *Arabidopsis thaliana*. For the array type, the 22k array was selected to yield more robust results. For quality control we selected high quality arrays only. To add genes pertaining only to the terpene synthase family we typed all genes which resulted from our PubMed search phrase, "Arabidopsis: terpene synthase", which yielded fifty one genes. No probe set could be found for six of the genes from the PubMed search (At1g78955, At3g31425, At3g32030, At3g14515, At3g14540, and At4g13300). Therefore, we

used the web-based software tools to analyze the expression of 45 different genes based on *Arabidopsis* microarrays.

To analyze terpene synthesis in development we created scatter plots for each gene using the meta-analysis tool. This revealed genetic expression early in life (germinated seed and seedling), during life (rosette and flowering), and into late life (mature siliques). Developmental periods were divided into nine categories. A bicluster analysis showing developmental expression of genes was also created.

The cluster analysis tool was used to determine similar expression profiles of genes expression in tissues and organs. These included reproductive and transport tissues and organs such as the root and stem.

A bicluster data matrix was also generated from analysis of gene expression due to biotic and abiotic stimuli. The wide variety of factors included ozone, elevated CO_2 varying light intensities, nutrient depravation, wounding stress, etc.

RESULTS AND DISCUSSION

Developmental gene expression

As of August, 2007, 2509 microarrays were available, all of which were included in our analysis. All genes queried were involved with terpene synthesis and no two genes shared the exact expression levels through each stage of development. Throughout the seedling, young flower, developed flower, and mature siliques, most genes are expressed, albeit at various levels (Fig. 1).

At2g07050 was continually expressed at high levels (> 10.8) throughout all developmental stages. At2g07050 has been shown to play a role in sterol synthesis in *Arabidopsis* (Suzuki et al., 2006).
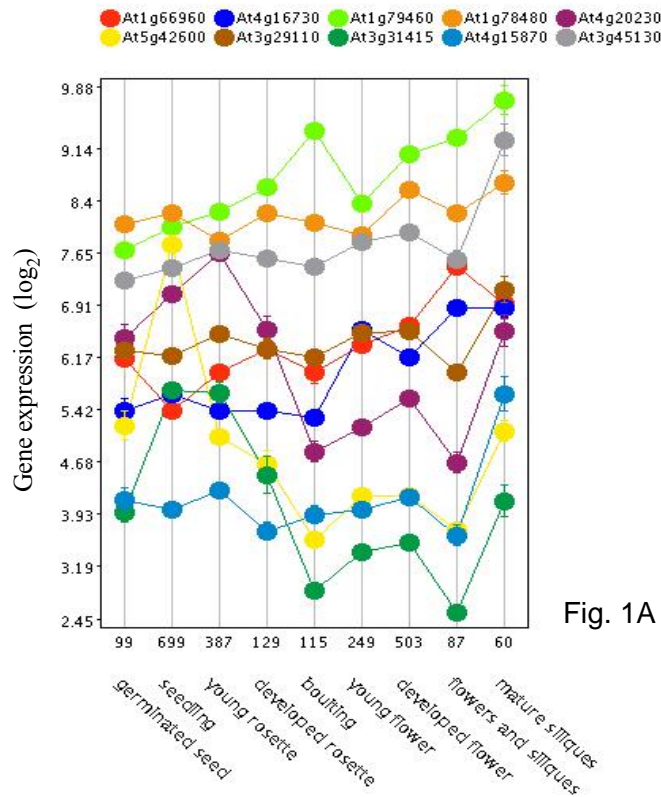
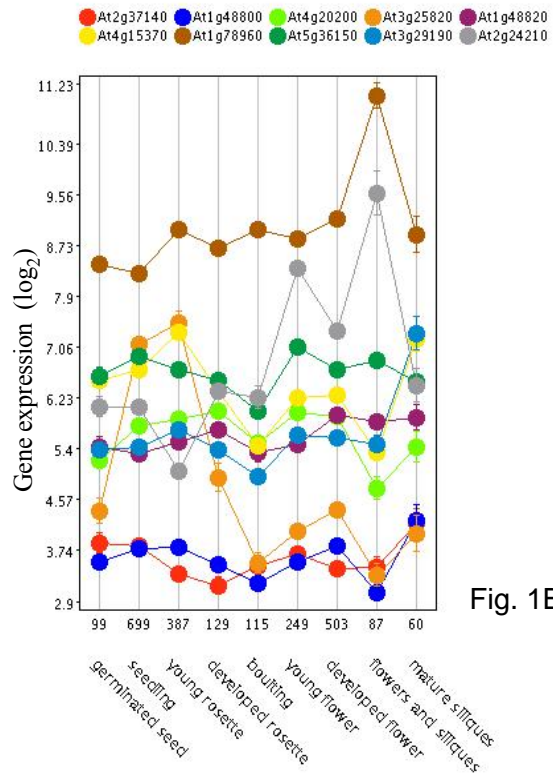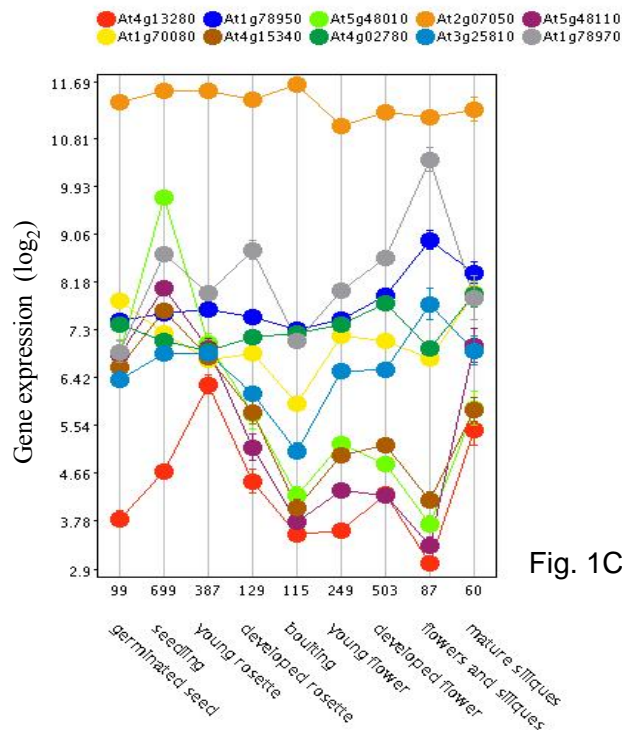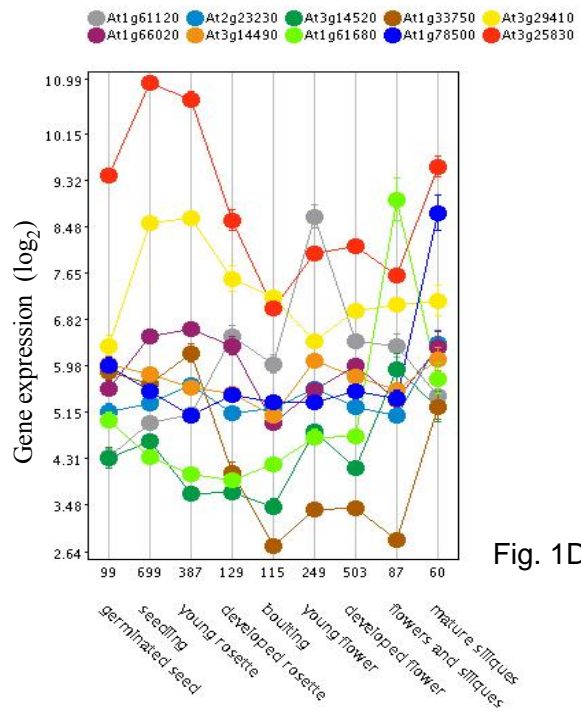**Figure 1:** Scatterplots for the 45 terpene synthase genes in *Arabidopsis thaliana*. The left axis represents the change of gene expression using base 2 logarithm. The numbers on the bottom axis represent the number of arrays per developmental stage.
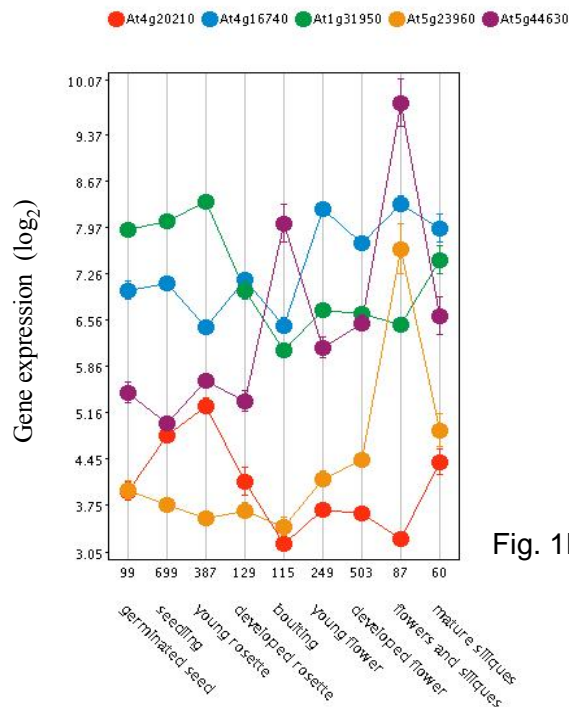
Fig. 1E

At2g37140 and At1g48800 were expressed at low levels (< 4.6) throughout all developmental stages. At2g37140 has been shown to play a role in isoprenoid synthesis (Lange and Ghessemian, 2003). Although Chen et al. (2003) found no expression of At1g48800 in leaves, flowers, siliques, stems, or roots, our results show a low, continual expression.

Other genes, such as At1g78960 and At5g23960, showed a steady, continuous expression but increased in flowers and siliques. At1g78960 is known to play a role in the biosynthesis of terpenes, specifically that of acting as a cyclase in triterpene productions (Phillips et al., 2006). At5g23960 is responsible for multiple terpene products, including α-humulene and (−)-(E)-β-caryophyllene, which comprise nearly half of the total volatiles emitted from *Arabidopsis* flowers (Aharoni et al., 2003).

Chen et al. (2003) discussed At2g23230 and showed that no signal was present in leaves, flowers, siliques, stems, or roots. This contrasts with our findings depicted in the scatter plot in Figure 1., which shows a steady (5.15) expression through all developmental stages and a slight rise in mature siliques. At1g78500 showed a similar trend with a sharper rise in mature siliques.

At3g25830 showed a high (~ 9.3) expression signal in the germinated seed, increased expression (~ 11.0) in the seedling stage, and then gradually decreased in the young rosette through boulting stages. At325830 encodes a monoterpene synthase, which has shown to play a role in *Arabidopsis* roots (Chen et al., 2004).

At5g23960 and At5g44630 showed a strong increase in expression levels in the flowers and siliques. Chen et al. (2003) found these two genes are expressed predominantly in flowers.

Bicluster analysis for developmental gene expression shows four genes (At4g16740, At3g31415, At1g70080, and At4g15340) which are down-regulated during various stages of development, and others (At3g31415, At1g70080, and At4g15340) which are down-regulated in flowers or siliques. Additionally, At4g16740 and At1g70080 in young rosette stage and At4g16740, At1g70080, and At4g15340 in boulting stage are all down-regulated.

Unfortunately, previous research has not discussed the mechanisms of At4g15340 or At3g31415, although both are known to be in the terpene synthase family. To increase the understanding of these two genes experiments focusing on developmental expression are highly encouraged, specifically the boulting to flowering stage, in which gene expression levels decreased.

Tissue specific gene expression

In total, 40 tissues and organs were analyzed for each gene. Thirty genes were up-regulated in 27 tissues and organs (Fig. 2). Nine genes were up-regulated in all tissue and organ categories; At1g31950, At4g16730, At1g79460, At1g78480, At5g42600, At3g29110, At3g25820, At1g78960, and At1g61680.
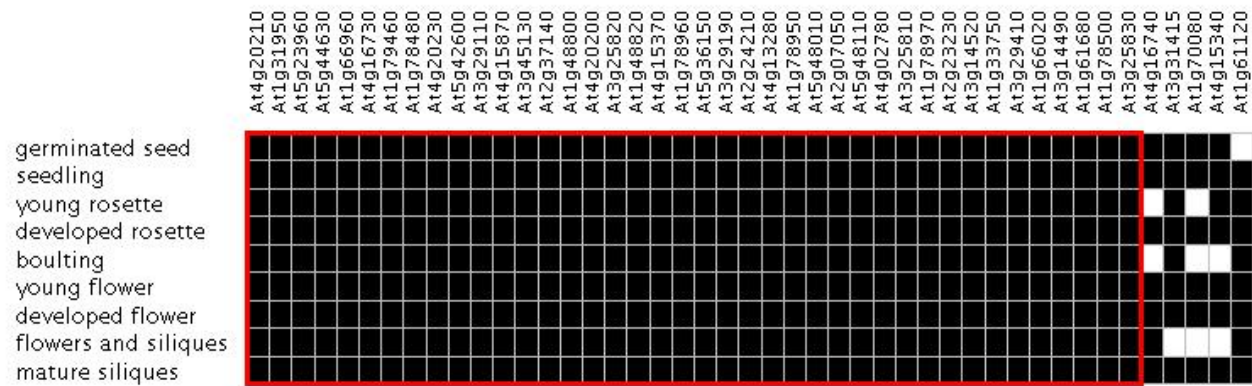
Figure 2: Bicluster analysis data matrix for developmental gene expression. Black squares represent above threshold values and white below threshold values.

Very little information exists for At1g31950, At1g78480, and At5g42600. Chen et al. (2003) have found expression of At4g16730, At1g61680, and At1g78480 in *Arabidopsis* flowers; however, the exact function of At1g79460 has yet to be determined. Previously, At1g79460 was expressed in floral and other organs (Chen et al., 2003).

Although in this analysis At3g29110 was up-regulated in all organs, Chen et al. (2003) found no expression in various organs. Interestingly, this gene has also been noted to play a role in snapdragon floral scent (Dudareva et al., 2003). As can be seen in Fig. 3, some genes are not significantly up-regulated in multiple tissues and organs. At1g70080 and At3g29410 are two examples. At170080 and At3g29410 have been shown by Ro et al. (2005) to be expressed in flowers and root tissues, respectively. Although our data matrix shows At170080 down-regulated in the flower and up-regulated in the inflorescence.

Expression patterns due to biotic and abiotic stress

Our data matrix included 107 different stress variables (Fig. 4). Two genes,

At1g48800 and At1g61680, were expressed when exposed to the same biotic stresses, nearly all of which were chemical. The biclustering of these two genes can be seen in Fig. 4.

Sharkey et al. (2005) have previously shown At1g48800 to play a role in isoprene synthesis and give evidence that phenylalanine is involved in the active site of isoprene synthase. At1g61680 was previously shown to be involved with S-linalool synthesis (Phillips et al., 2006).

In our analysis At2g37140 was expressed in response to many stressors. Stress was responsible for multiple terpene synthase expression, including drought, nutrient, ozone and mechanical stress. Although At2g37140 was up-regulated in response to multiple stressors, including cold, drought, and osmotic stress, little research has been done on this gene.

Although this analysis pertained to *Arabidopsis*, there has been a variety of other plants which have been investigated in regard to terpenoids. It has been reported that neither water stress, nor nutrient stress was responsible for observed variation in terpene content in *Eucalyptus polybractea* (King et al., 2004). However, mechanical stress has been shown to cause a decrease of

Figure 3: Bicluster analysis of tissue specific gene expression in *Arabidopsis thaliana*. The red box includes genes expressed in the same organs or tissues.

terpene production in *Daucas carota* (Seljasen et al., 2001). Only transient changes in terpene concentrations were detected in spruce seedlings (*Picea abies*) when exposed to gases normally emitted by automobiles (Viskari et al., 2000). Other environmental stress like ozone stress on Norway spruce (*Picea abies* L. Karst) was found to have no effect on terpene emission (Lindskog and Potter, 1995).

In our bicluster data matrix two genes, At1g78480 and At1g70080, were not

significantly up-regulated with any application of stressor. However, this could be due to lack of experimental data within the microarray databases, which contributed to our data matrix.
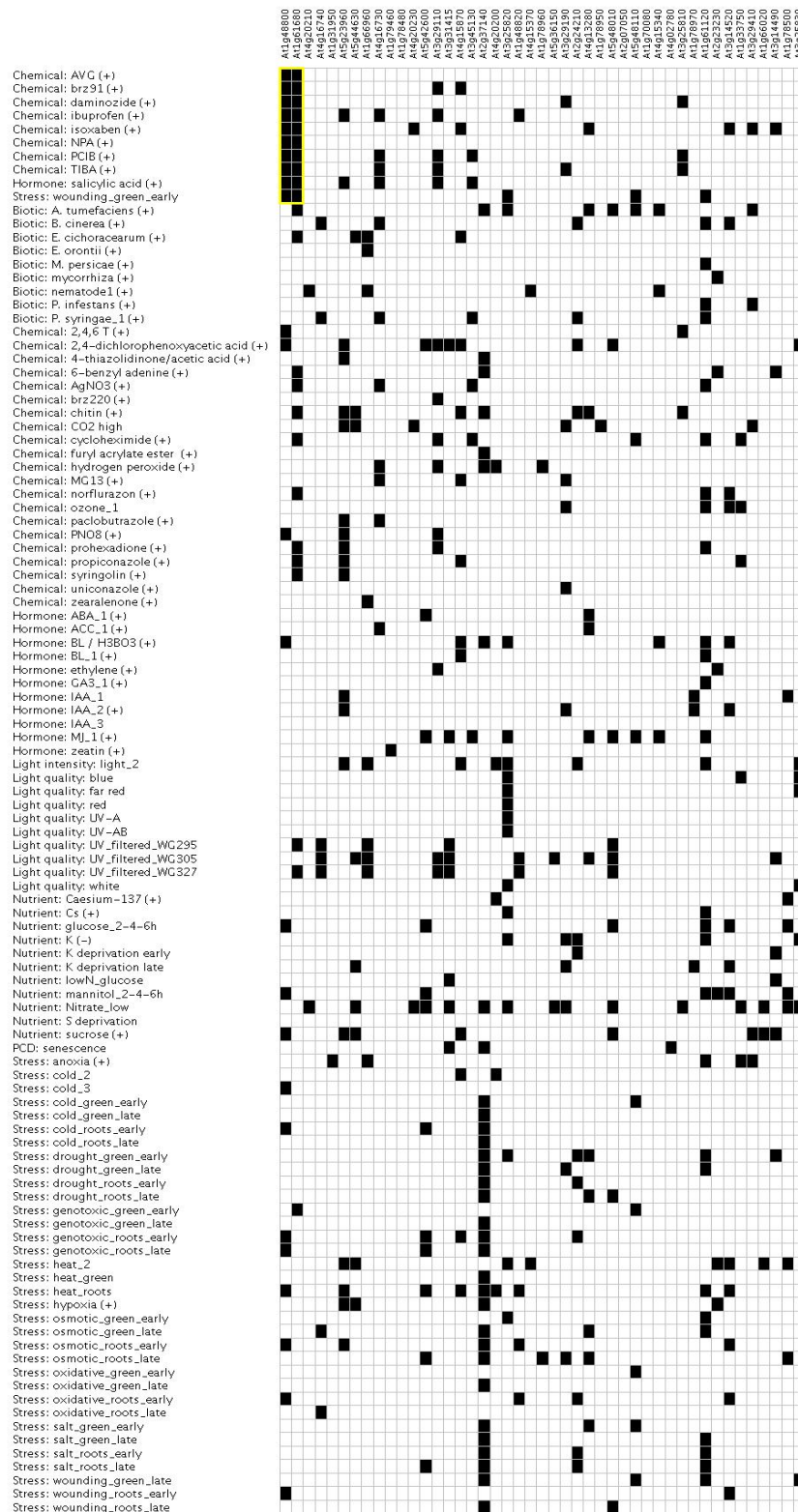

Figure 4: Bicluster data matrix showing genes up-regulated in response to stimuli. Clustering of At1g48800 and At1g61680 were due to similar up-regulation in response to chemicals; AVG amino-ethoxvinylglycine), brz91 (brassinozole), daminozide, ibuprofen, isoxaben, NPA (nonyl phenoxy acetic acid), PCIB (p-chlorophenoxyiso-butyric acid), and TIBA (tri-isobutyl aluminum).

CONCLUSION

As shown by Aubourg et al. (2004), many genes are involved in the complex production of terpenoids in *Arabidopsis* and we have presented 45 of them here. To more rapidly analyze the plethora of microarray data, which is being accumulated, the application of powerful software is critical for finding expression trends within a gene set. Based on the *in silico* analysis of expression of terpene synthesis genes in *Arabidopsis* using Genevestigator, many factors contribute to expression in development, various anatomical locations, and stimulus response. Further research of genes that currently lack experimental studies (At1g31950, At1g78480, At2g37140, At3g31415, and At4g15340) may prove helpful to understand terpene synthesis in development, tissues, and stress response.

REFERENCES

Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel WJ, Verstappen FW, Verhoeven HA, Jongsma MA, Schwab W, Bouwmeester HJ. Terpenoid metabolism in wild-type and transgenic *Arabidopsis* plants. Plant Cell 2003;15:2866-84.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-3402.

Aubourg S, Lecharny A, Bohlman J. Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. Mol Genet Genomics 2004;267:730-45.

Buckingham J. Dictionary of natural products on CD-ROM, version 6.1. London: Chapman & Hall, 1998.

Chen F, Tholl D, D'Auria JC, Farooq A, Pichersky E, Gershenzon J. Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. Plant Cell 2003;15:481-94.

Chen F, Ro DK, Petri J, Gershenzon J, Bohlmann J, Pichersky E, Tholl D. Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineol. Plant Physiol 2004;135:1956-66.

Dawson FA. The amazing terpenes. Naval Stores Rev 1994;6-12.

Degenhardt J, Gershenzon J, Baldwin IT, Kessler A. Attracting friends to feast on foes; engineering terpene emission to make crop plants more attractive to herbivore enemies. Curr Opin Biotechnol 2003;14:169-76.

Dudareva N, Martin D, Kish CM, Kolosova N, Gorenstein N, Faldt J, Miller B, Bohlmann J. (E)-β-Ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. Plant Cell 2003;15: 1227–41.

King DJ. Gleadow RM, Woodrow IE. Terpene deployment in *Eucalyptus-polybractea*; relationships with leaf structure, environmental stresses, and growth. Funct Plant Biol 2004;31:451-60.

Lange BM, Ghessemian M. Genome organization in *Arabidopsis thaliana*; a survey of genes involved in isoprenoid and chlorophyll metabolism. Plant Mol Biol 2003;51:925-48.

Lindskog A, Potter A. Terpene emission and ozone stress. Chemosphere 1995;30:1171-81.

Phillips D, Rasbery J, Bartel B, Matsuda S. Biosynthetic diversity in plant triterpene synthesis. Curr Opin Plant Biol 2006;9:305-14.

Ro DK, Ehlting J, Keeling CI, Lin R, Mattheus N, Bohlmann J. Microarray expression profiling and functional characterization of AtTPS genes: duplicated *Arabidopsis thaliana* sesquiterpene synthase genes At4g13280 and At4g13300 encode root-specific and wound-inducible (Z)-bisabolene synthases. Arch Biochem Biophys 2005;448:104-16.

Seljasen R, Bengtsson G, Hoftun H, Vogt G. Sensory and chemical changes in five varieties of carrot (*Daucuscarota* L) in response to mechanical stress at harvest and post-harvest. J Sci Food Agric 2001;81:436-47.

Sharkey TD, Yeh S, Wiberley AE, Falbel TG, Gong D, Fernandez DE. Evolution of the isoprene biosynthetic pathway in kudzu. Plant Physiol 2005;137:700-12.

Suzuki M, Xiang T, Ohyama K, Seki H, Saito K, Muranaka T, Hayashi H, Katsube Y, Kushiro T, Shibuya M, Ebizuka Y. Lanosterol synthase in dicotyledonous plants. Plant Cell Physiol 2006;47:565-71.

Trapp SC, Croteau RB. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 2001;158,811-32.

Viskari E-L, Surakka J, Pasanen P, Mirme A, Kossi S, Ruuskanen J, Holopainen J. Responses of spruce seedlings (*Picea abies*) to exhaust gas under laboratory conditions: I plant-insect interactions. Environ Pol 2000;107:89-98.

Whitbred JM, Schuler MA. Molecular characterization of CYP73A9 and CYP82A1 P450 genes involved in plant defense in pea. Plant Physiol 2000;124:47-58.

Zimmermann PM, Hirsch-Hoffmann M, Hennig L, Gruissem W. GENEVESTI-GATOR: *Arabidopsis* microarray database and analysis toolbox. Plant Physiol 2004;136: 2621-32.

Zinkel DF, Russell J. Naval stores: production, chemistry, utilization. New York: Pulp Chemicals Association, 1989.