

## **Quantification of The Plant Terpenome: Predicted Vs. Actual Emission Potentials**

### **Authors**

Piyush Priya, Sangita Kumari & Gitanjali Yadav\*

### **Running Head:**

The Plant Terpenome

### **Affiliation**

Computational Biology Laboratory, National Institute of Plant Genome Research (NIPGR), New Delhi, India- 110067

### **Full Address for Correspondence**

Gitanjali Yadav, Computational Biology Laboratory, National Institute of Plant Genome Research  
New Delhi, India- 110067; Tel: +91-11-26735103; Fax: +91-11-26741658

Email: [gy@nipgr.ac.in](mailto:gy@nipgr.ac.in)

\*- To Whom Correspondence should be addressed

### **Abstract**

Plant essential oils are complex mixtures of volatile organic compounds, which play indispensable roles in communication, defense, and adaptive evolution. The complete chemical library produced by a plant is referred to as its terpenome. The potential biological information stored in essential oil composition data can provide an insight into the silent language of plants, as well as roles of terpene emissions in direct and indirect defense, and for playing a crucial role in adaptive evolution. In this work, we have attempted to measure the plant terpenome from a global perspective. One way of measuring the terpenome is to observe and record actual emissions in natural conditions, and this has been in practice for over a century through variously evolving methods of comprehensive GC-MS and HPLC. An alternative method is a knowledge-based prediction of the terpenome, and this method has gained popularity in recent years, with the advent of large-scale genome sequencing technologies. Over the past decade, our laboratory has been involved in compilation and investigation of the plant terpenome using both these methods and this has offered us the opportunity to compare and contrast data from actual and potential emissions, in order to better understand the terpenome and its roles in primary, secondary and adaptive metabolism. We have used emission data in conjunction with genomic data in order to understand how a plant creates the so-called final terpenome, specific to itself, and whether or not plants tap the complete potential for terpene biosynthesis at their disposal according to their genomes. For measuring actual emissions, we have used EssOilDB (the ESSential OIL DataBase), the largest contextual web resource for phytochemicals and for measuring the total plant potential for emissions, we have used TERZYME, an

automated algorithm for identification and analysis of genes and proteins involved in isoprenoid biosynthesis.

## **INTRODUCTION**

The isoprenoid family represents one of the most ancient and widespread classes of structurally and functionally rich biomolecules known to man. Although these natural products are synthesized in all organisms, the plant kingdom exhibits tremendous variation in their chemistry and roles, ranging from primary metabolism to secondary metabolism and specialized ecological interactions with the environment, and a majority of terpenoids, specially sesquiterpenes and monoterpene derivatives are released by plants in the form of volatile emissions. The isoprenoid biosynthetic pathways in plants are capable of generating a diverse spectrum of volatile organic compounds (VOCs or essential oils, also known as ethereal oils) for maintaining communication and fruitful adaptive interaction with the environment. These compounds can be grouped together with other chemicals of plant origin, into the category of 'phytochemicals'. Extensive research on phytochemicals released through plant volatile emission has shown these to be important protective and signaling molecules under oxidative, thermal and pathogen stress, and can act both as allelochemical and neighbor detection signals, thus constituting a platform for plant-plant interaction (Leitner et al., 2008; Vickers et al., 2009). Herbivore-challenged plants are known to emit volatile phytochemicals not only to invite the natural enemies of herbivores (parasitoids and predators), but also to aid neighboring plants by inducing defense response in them (Pare and Tumlinson 1999; Kessler and Baldwin, 2001; Baldwin et al., 2006). Alongside of such roles in defense, volatile phytochemicals have enormous applications in pharmaceutical, industrial and agribiotechnological areas. Since antiquity, plant-originated natural products have been used for their medicinal and organoleptic properties. They are extensively used in food and cosmetics industry for generating flavor and fragrances due to their pleasant scent (Caputi and Aprea, 2011). The significant role of these phytochemicals in the prevention and treatment of human diseases is attributed to their specific activities against microbes, cancer, cardiovascular problems and diabetes (Reddy et al., 1998; Bardon et al., 1998; Bardon et al., 2002; Abdollahi et al., 2003). Besides these properties, they have also shown to be associated with hepatoprotective, spasmolytic and carminative activities (Ozbek et al., 2003; Lahlou et al., 2003). Reports on the analysis of their free radical scavenging capacity have suggested that they can act as promising natural antioxidants as well (Ruberto and Baratta, 2000; Botsoglou et al., 2003; Mimica-Dukic et al., 2003). Isoprenoids are derived from the universal C5 precursor isoprene, and terpene synthase (TPS) enzymes are key players in the generation of isoprenoid diversity, catalyzing one of the most complex reactions known to chemistry and biology. Infact, TPSs can generate such a tremendous diversity of compounds using the regiochemical and stereochemical intricacy of their cyclization reactions that the complete terpenoid chemical library of a given plant has been

termed as the ‘terpenome’ (Cane and Ikeda, 2011; Christianson, 2008). The ongoing genomic revolution has provided an unprecedented opportunity for genome mining efforts toward discovery and characterization of the terpenome.

In this work, an attempt has been made to quantitatively measure the plant ‘terpenome’ or phytochemical dynamics through a comprehensive analysis of available data on essential oils (where the term ‘essential’ denotes that the oil has the fragrant essence of the source plant), treating them as multicomponent chemical combinations having terpenes as their major constituents along with other non-terpene compounds. We have used emission data in conjunction with genomic data in order to understand how a plant creates the so-called final terpenome, specific to itself, and whether or not plants tap the complete potential for terpene biosynthesis at their disposal according to their genomes. Thus, We have performed comparison of actual and potential phytochemical emissions; where the term ‘actual’ refers to the observed and recorded plant emissions in natural conditions, a strategy that has been in practice for over a century through variously evolving methods of comprehensive GC-MS and HPLC. Essential oils are generally obtained from plants using an appropriate extraction method (Vigan, 2010). Traditional extraction methods like hydrodistillation, steam distillation, solvent extraction and expression under pressure, are more common although advanced techniques such as supercritical fluid and subcritical water extraction have been used in recent times (Edris, 2007; Coelho et al., 2012). For this, we have used an in-house repository developed over the past decade through a systematic compilation of documented emission profiles along with the additional information about the respective source plants, geographical locations, extracted compounds and their compositions, existing biotic and abiotic factors at the time of extraction, and several others, in order to gain insights into product diversity. The massive data thus collected data has been integrated into an online repository named ‘EssoilDB’ (Kumari et al, 2014). Although more than 5000 individual compounds are recorded in the database, the major constituents of essential oils are terpenoid compounds and their derivatives. The alternative method of terpenome quantification involves a knowledge-based prediction of the genomic complement of terpenes, and this method has gained popularity in recent years, with the advent of large-scale genome sequencing technologies. Since the terpene biosynthetic machinery regulates both quantity and diversity of terpenes present in a plant, we define the ‘terpenome’ to signify the full complement of TPSs genes present in a genome. Plants genomes, on account of repeated events of duplication, tend to have very large sizes, and this in turn, leads to the existence of gene families in the plant kingdom. The TPSs family is one such example, with members that have similar sequences and structures, and yet show clear divergence in different lineages (Bohlmann *et al.*, 1998a). Current knowledge of the TPSs in plants, with the exception of the moss genome *Physcomitrella patens* (which has a single functional TPS gene), indicates that the TPS gene family is a mid-size family, with gene

numbers ranging from approximately 20 to 150 (Chen *et al.*, 2011). Rapidly increasing numbers of plant nuclear genomes, in various stages of sequencing provide an unprecedented opportunity for genome mining efforts toward discovery of novel terpene synthases, and their annotations or classification as per the various established subfamilies. Towards this, we have performed comprehensive identification and functional prediction of TPSs in 42 plant genomes, and the method has been developed as a freely available online web resource, namely, 'Terzyme' (Priya, 2016, Under Review).

In the following sections, we present a comparative assessment of predictions from Terzyme and actual emission data from EssOilDB, for 19 plant species for which such an assessment was feasible, based on availability of the full genome as well as availability of emission data. We compare and contrast emission data with genomic data in order to confirm that plants do indeed create a so-called final terpene, specific to each species, and also that, plants have the unique ability to selectively tap the potential for terpene biosynthesis at their disposal according to their requirement. We discuss the advantages and limitations of both methods, followed by specific examples to illustrate the biological implications of this work plant communication and defense, and also to highlight its applicability and in the current era of comparative genomics.

## METHODOLOGY

**Emission Data Collection:** Essential oil profiles were collated from EssOilDB, the largest online contextual library of phytochemicals, encompassing literature records spanning a century of published reports of essential oil profiles, starting from early 1900s to date (Kumari et al., 2014). EssOilDB contains data from over 30 scientific journals and 1667 plant species, subspecies or varieties representing about 80 distinct taxonomic families encompassing the entire range from ancient Magnolids, lower plants like chlorophytes and mosses, to the gymnosperms and angiosperms.

**TPS Data Compilation:** Terzyme (Priya et al., Under review) is hidden markov model-based tool developed for identification and analysis of genes and proteins involved in isoprenoid emissions and was run on 42 available nuclear genomes leading to identification of more than 3000 unique terpene synthase enzymes in the plant kingdom.

**Data Analysis:** Comparative analysis of EssOilDB and Terzyme data was performed for identification of common plant species for which both actual and potential emission data would be available for further analysis. Emissions from these species were further categorized based on the terpene product specificity, and the genes identified in these plants were interactively mapped using IGMAP (Priya et al., 2015). Product level clustering of emissions and TPS predictions were performed using Terzyme, Pubchem classification as well as in-house shell scripts. Monoterpene specific emission pattern visualization was done in R. Quantification of the volatile terpene in this manner led to a dataset of

terpene emission products and corresponding genomic signatures for each compound in each plant, and this data was manually compared in order to identify patterns, if any, between actual and predicted emission potentials.

## RESULTS & DISCUSSION

*Association of essential oil data: actual and potential emissions:* In all, 19 plant species were identified for which, both emission datasets, as well as genomic datasets were available for comparison. These 19 species are listed in Table 1. As described in the introduction, it is important to understand the actual volatile terpene emissions of a plant, as much as the TPS complement within the respective genome that is responsible for production of the emitted volatiles. Towards this, we have compared the two datasets, one representing the complete genomic complement of TPS genes in a given plant (as predicted by TERZYME described in methods), and the second representing actual emission profiles obtained from GC/GC-MS based essential oil compositional data from published literature records for the corresponding plant, as compiled from EssoilDB (Kumari et al., 2014). These two datasets are termed as the ‘potential’ and the ‘actual’ terpenome, respectively and the comparative analysis can only be performed for plants for which detailed essential oil compositional data is available, in addition to availability of full nuclear genome sequence. In all, the comparison could be performed for 19 plants where both datasets were available. Table 1 depicts the non-redundant number of actual compounds observed to be emitted from each given plant (observed terpenome) compared with the total number of identified terpene biosynthetic enzyme coding sequences in the genome (potential terpenome).

It is clear from Table 1 that the quantified terpenome across all 19 plants shows a high degree of variability between actual number of compounds recorded as released, compared with the potential number of compounds that can, in principle, be biosynthesized by the species. For example, in case of the model dicot Arabidopsis, 36 distinct TPSs were identified in the nuclear genome, whereas literature records contain information about 29 unique, non-redundant terpenoid compounds known to have been emitted by the plant. It is well known that plants do not utilize their full phytochemical potential at a given time and that essential oil compositions are often influenced by bio-geological conditions at the time of extraction, such as environmental and physiological circumstances, plant parts used for oil extraction, stress/disease and geographical distribution etc. Our method of quantification has been designed to take into account any or all of these factors and combine them before performing a proportional estimate in terms of the genomic potential. This allows us to get an accurate estimate and gain insights into how these factors individually or in concert, influence phytochemical dynamics of the species in question. Thus, as per data provided in Table 1, Arabidopsis has been observed, under various

conditions, to emit 29 of the 36 potential terpenes. It may be noted that the remaining seven (36-29) compounds have not “yet” been observed by any experimental study, and may be highly specialized compounds that are only emitted under very specific conditions that have not been recorded in literature till date. Future studies may be able to shed light on the specific conditions under which these compounds are released. A detailed perusal of these seven compounds revealed very interesting patterns that have been presented in the next section. In contrast, the case of the green bean (*P.vulgaris*) is quite striking since it has been known to emit only three of the 49 terpene based compounds that its genome has the potential to biosynthesize (Table 1). This is a case where a miniscule portion of the prospective terpenome has been recorded by experimental studies. It is quite possible that much more of the potential terpenome is being emitted regularly by the green bean, but has not been recorded in literature due to lack of sufficient number of studies on essential oil composition. However, it is also possible that much of the potential terpenome of the bean is latent and awaiting specific conditions in which to be expressed. It is quite possible that emission data from these plants may have been collected under normal conditions lacking any stress factors, such as herbivory, thereby resulting in a discordance between the observed and predicted terpenome.

***Emission patterns of Monoterpenes:*** It may be noted that the present comparative analysis relies extensively on observed/published emissions of phytochemicals, and it is known that monoterpenes constitute the major fraction of terpene volatile emission, since these are lowest MW compounds. Accordingly, we divided the dataset in Table 1 based on the “kind” of terpene emitted, and a large fraction of the data could be deciphered for better comprehension of individual patterns and emission signatures. This grouping has been presented in Table 2, where the data is same as in Table 1, except that monoterpene, sesquiterpene and diterpene emissions have been separately compared. Further analysis of the patterns was done between corresponding values of identified Monoterpene synthase proteins present in these 19 plant species. The analysis of Arabidopsis emissions, discussed earlier, presents a very interesting case based on data in Table 2. Of the 36 TPS sequences that were identified/predicted in the genome of Arabidopsis, seven are monoTPSs, 6 are DiTPSs, and 23 are SesquiTPSs. Of these predicted TPSs, none of the diterpenes and only half of the sesquiterpenes have been observed in emission profile studies (see Table 2). However, a total of 14 distinct monoterpene compound emissions have been observed in the plant, as compared to only seven monoTPS sequences in the genome – clearly suggesting that observed emissions are higher than the predicted number of biosynthetic proteins in the genome. In general, based on the above considerations, the monoterpene emission data in Table 2 can be categorized into three groups, as shown in Figure 1. The group of plants in Figure 1A represent cases where the actual and observed emissions are comparable in quantity. Plants in this category have attained the full potential

of terpene volatile emission, as per available experimental records. The second group, shown in Figure 1B, is represented by majority of the 19 plants investigated, and this group represents cases where the observed volatile terpenome is much smaller than potential terpenome. It can be inferred that the genome has the potential to produce far more terpenoids than those actually recorded for these plants. As mentioned earlier, it is well known that terpene emission profiles are condition specific and spatio-temporal, often altered based on plant developmental stage, source parts and various geomorphological situations. Also each plant species is known to occupy its own unique ecological niche, and thus it aptly evolves the ability to synthesize sets of compounds that assist it in interaction and adaptations with its biotic and abiotic environment (Chen et al., 2011). Finally, in plants like *Arabidopsis thaliana*, *Oryza sativa*, *Fragaria vesca*, *Citrus clementina* and *Cicer arietinum* the observed terpenome was found to be much higher than the potential terpenome. This presents an ironic situation, where the genome was found to contain fewer TPS genes than the minimum number required to produce the spectrum of actual emissions observed in these plants. Figure 1C depicts this case in various plants investigated. In such cases, it may be conjectured that although the total number of TPSs available for each plant, viz. seven in case of *Arabidopsis thaliana* and one in case of *Oryza sativa*, is relatively low when compared to their respective emission profiles, these specific TPS enzymes may possess the ability to form multiple terpene products, thus bestowing the plant to adapt in its specific local niche. These examples also offer opportunities for further research leading to new insights and a fuller understanding of TPS regulation and product complexity.

## **CONCLUSION & FUTURE DIRECTIONS**

Plants produce a diverse array of phytochemicals for both general and specialized functions. Environmental factors such as light, drought, temperature and several biotic conditions can greatly influence the yield, composition and emission of phytochemicals. The emission of volatile terpenes is estimated to account for more than half of the total emission of essential oils. Terpenoid emission mainly comprises monoterpenes and sesquiterpenes whereas their derivatives such as aldehydes, alcohols and ketones are present in various amounts. Biosynthesis of terpenes results in the generation of an enormous chemical and structural diversity of natural compounds, mainly on account of the highly complex machinery that regulates formation of isoprenoid products. This machinery comprises of the unique class of enzymes known as Terpene synthases (TPSs). With the availability of sequenced genomes and high throughput data analysis tools, the automated large scale search and characterisation for genome wide TPSs is now becoming feasible. The complete chemical library of terpenes, or the Terpenome, is highly complex consisting of about 60,000 complex chemical structures but any given plant species can

synthesize only a small fraction of it. Thus, for a single species TPSs have evolved unique capacity to synthesize a specific array of terpenes.

Comparison of actual terpenome with the potential terpenome, as performed in this study, revealed how plants can modulate their TPSs expression based on condition or environment-specific needs. As can be seen from Table 1 and 2, in most plant species, the available emission profiles appear to only partially capture the full terpene potential. For example, in case of Eucalyptus, the emission of only 21 monoterpenes have been recorded in literature, whereas the genome of this plant contains genes that code for more than double this number of monoterpene synthases. This is the case with most other plants analysed in this effort, and it may be explained by limitations of record compilations or the unavailability of data under stress for the plant, since these are secondary metabolites induced for specific functions and a large number of biosynthetic genes may lie dormant under natural conditions. Interestingly, a few plants like the legume crop chickpea (*Cicer arietinum*) show a contrasting trend, wherein more terpene emissions have been recorded in literature than the predicted genomic potential, means that the number of TPS gene models in the plant genome is less than the total number of terpenes released by that plant, reflecting in turn, the unique ability of TPS enzymes to synthesize multiple products from a single substrate. A single predicted TPS gene product can thus generate many terpene compounds rather than just one, and this ability to create complex odor profiles from a minimal gene set can lead to an expansion of the terpene potential of the respective plant. Further characterization and detailed analysis of such genes is currently underway in our laboratory and may provide insights into the unique molecular mechanism of TPSs that remains poorly understood.

In summary, our data provides evidence that plant can intelligently engineer their actual terpenome by driving the expression of particular TPSs based on the specific factors like stress and other spatio-temporal environmental conditions. Attainment of full potential of terpenome is condition specific for most plants and out of many potential TPSs genes, it may be that, only few specific genes express that potentially help plants to counter the specific condition to which plant has been exposed. We also identified plants in which recorded terpene emissions are fairly higher than the potential product-formation capacity of the corresponding genome, suggesting very interesting scenarios, which might directly relate to TPS promiscuity, and subtle modulations within TPS active sites, resulting in a plasticity that eventually enables the plant to generate multiple products from minimal substrate pool. It may be recalled that TPSs employ various strategies for generation of terpenoid diversity, as reviewed in the introduction and hence provide an elegant example of combinatorial chemical synthesis. To maximize the number of final products, they have the ability to create multiple products from minimal/single substrate pool and minimal genetic alterations (Degenhardt et al., 2009; Sacchettini and Poulter, 1997). Thus, this



unique aspect associated with terpene synthases makes them promiscuous in function (Yoshikuni and Keasling, 2007). Finally, although we are still far from a systems-level modeling of the terpenoid pathway, this work has led to the generation of the most comprehensive and high quality atlas of the plant terpenome till date and will pave the way for further studies into characterization of multi-product TPS enzymes, as predicted by our data, through metabolic engineering.

### **Acknowledgements**

Authors' thanks are due to Director, NIPGR for encouragement, the SERB Women's Excellence Award grant of DBT, Govt of India, to GY for financial support, the Biotechnology Information System Network (BTISNET) program of Dept of Biotechnology (DBT), Govt of India, for computational resources, and the Council of Scientific & Industrial Research (CSIR) for Senior Research Fellowship (SRF) to SK and PP.

### **References:**

- Abdollahi, M., Salehnia,A., Mortazavi,S. et al. (2003) Antioxidant, antidiabetic, antihyperlipidemic, reproduction stimulatory properties and safety of essential oil of *Satureja Khuzestanica* in rat in vivo: a toxicopharmacological study. *Med Sci Monit.*, 9, 331–335.
- Baldwin, I.T., Halitschke,R., Paschold,A. et al. (2006) Volatile Signaling in Plant-Plant Interactions: "Talking Trees" in the Genomics Era. *Science.*, 311, 812-815.
- Bardon,S., Foussard,V., Fournel,S. et al. (2002) Monoterpenes inhibit proliferation of human colon cancer cells by modulating cell cycle-related protein expression. *Cancer Lett.*, 181, 187-194.
- Bardon,S., Picard,K. and Martel,P. (1998) Monoterpenes inhibit cell growth, cell cycle progression, and cyclin D1 gene expression in human breast cancer cell lines. *Nutr Cancer.*, 32, 1-7.
- Bohlmann, J., Meyer-Gauen, G. and Croteau, R. (1998) Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U. S. A.*, 95, 4126–4133.
- Botsoglou,N.A., Govaris,A., Botsoglou,E.N. et al. (2003) Antioxidant activity of dietary oreganoessential oil and R-tocopheryl acetate supplementation in long-term frozen stored turkey meat. *J. Agric. Food Chem.*, 51, 2930-2936.
- Caputi, L. and Aprea,E. (2011) Use of terpenoids as natural flavouring compounds in food industry. *Recent Pat Food Nutr Agric.*, 3, 9-16.

- Chen, F., Tholl, D., Bohlmann, J. and Pichersky, E. (2011) The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.*, 66, 212–229.
- Coelho, J.P., Cristino, A.F., Matos, P.G. et al. (2012) Extraction of Volatile Oil from Aromatic Plants with Supercritical Carbon Dioxide: *Experiments and Modeling. Molecules.*, 17, 10550-10573.
- Degenhardt, J., Köllner, T.G. and Gershenzon, J. (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, 70, 1621–1637.
- Edris, A.E. (2007) Pharmaceutical and therapeutic potentials of essential oils and their individual volatile constituents: a review. *Phytother Res.*, 21, 308-323.
- Keasling, J.D. (2008) Synthetic biology for synthetic chemistry. *ACS Chem. Biol.*, 3, 64–76.
- Kessler, A. and Baldwin, I.T. (2001) Defensive Function of Herbivore-Induced Plant Volatile Emissions in Nature. *Science.*, 291, 2141-2144.
- Kumari, S., Priya, P., Misra, G., & Yadav, G. (2013). Structural and biochemical perspectives in plant isoprenoid biosynthesis. *Phytochemistry Reviews* (Vol. 12). doi:10.1007/s11101-013-9284-6
- Kumari, S., Pundhir, S., Priya, P., Jeena, G., Punetha, A., Chawla, K., Firdos Jafaree, Z., Mondal, S. and Yadav, G. (2014) EssOilDB: a database of essential oils reflecting terpene composition and variability in the plant kingdom. *Database (Oxford).*, 2014, 1–12.
- Lahlou, S., Interaminense, L.F., Leal-Cardoso, J. et al. (2003) Antihypertensive effects of the essential oil of *Alpinia zerumbet* and its main constituent, terpinen-4-ol, in DOCA-salt hypertensive conscious rats. *Fundam Clin Pharmacol.*, 17, 323-330.
- Leitner, M., Kaiser, R., Rasmussen, M.O. et al. (2008) Microbial oligosaccharides differentially induce volatiles and signalling components in *Medicago truncatula*. *Phytochemistry.*, 69, 2029-2040.
- Mimica-Dukic, N., Bozin, B., Sokovic, M. et al. (2003) Antimicrobial and antioxidant activities of three *Mentha* species essential oils. *Planta Med.*, 69, 413-419.
- Ozbek, H., Ugras, S., Dulger, H. et al. (2003) Hepatoprotective effect of *Foeniculum vulgare* essential oil. *Fitoterapia.*, 74, 317-319.
- Pare, P.W. and Tumlinson, J.H. (1999) Plant Volatiles as a Defense against Insect Herbivores. *Plant Physiol.*, 121, 325-331.
- Reddy, B.M.V., Angers, P., Gosselin, A. et al. (1998) Characterization and use of essential oil from *Thymus Vulgaris* against *Botrytis cinerea* and *Rhizopus stolonifer* in strawberry fruits. *Phytochemistry.*, 47, 1515-1520.
- Ruberto, G. and Baratta, M.T. (2000) Antioxidant activity of selected essential oil components in two lipid model systems. *Food Chem.*, 68, 167-174.
- Sacchettini, J.C. and Poulter, C.D. (1997) Creating isoprenoid diversity. *Science*, 276, 1788–1789.

- Vickers, C.E., Gershenzon, J., Lerdau, M.T. et al. (2009) A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nat. Chem. Biol.*, 5, 283-291.
- Vigan, M. (2010) Essential oils: renewal of interest and toxicity. *Eur J Dermatol.*, 20, 685-692.
- Yoshikuni, Y., Ferrin, T.E. and Keasling, J.D. (2006) Designed divergent evolution of enzyme function. *Nature*, 440, 1078–1082.

## FIGURE LEGENDS

**Figure 1** Comparison of observed terpenome with the potential terpenome in 19 plants. (A) The cases in which Observed Terpenome was comparable with potential Terpenome. (B) Plants in which Observed Terpenome was fairly lower than potential Terpenome, (C) Cases were Observed Terpenome was comparatively higher than potential Terpenome.

## TABLES

**Table 1:** List of plants used in this study, for comparative quantification of the Terpenome.

Plant Species	Actual Terpenome (No. of Compounds Observed)	Potential Terpenome (No. of TPSs in Genome)
<i>Arabidopsis thaliana</i>	29	36
<i>Citrus sinensis</i>	53	88
<i>Eucalyptus grandis</i>	34	111
<i>Ricinus communis</i>	5	59
<i>Populus trichocarpa</i>	22	67
<i>Phaseolus vulgaris</i>	3	49
<i>Vitis vinifera</i>	33	118
<i>Glycine max</i>	4	45
<i>Fragaria vesca</i>	41	60
<i>Malus domestica</i>	20	118
<i>Oryza sativa</i>	8	43
<i>Citrus clementina</i>	33	21
<i>Medicago truncatula</i>	12	52
<i>Capsella rubella</i>	18	42
<i>Prunus persica</i>	8	27
<i>Cucumis sativus</i>	4	32
<i>Carica papaya</i>	18	41
<i>Gossypium raimondii</i>	16	76
<i>Cicer arietinum</i>	16	19

**Table 2:** Comparative terpenome quantification with predictions and actual emissions categorized by functional groupings. O/P: Observed (Actual) Terpene Emission Records / Predicted No.of TPS genes.

<b>Plant Species</b>	<b>Monoterpenes O/P</b>	<b>Diterpenes O/P</b>	<b>Sesquiterpenes O/P</b>	<b>Total O/P</b>
<i>Arabidopsis thaliana</i>	14/07	0/6	15/23	29/36
<i>Citrus sinensis</i>	33/36	0/8	20/44	53/88
<i>Eucalyptus grandis</i>	20/40	0/19	14/52	34/111
<i>Ricinus communis</i>	5/25	0/13	0/21	5/59
<i>Populus trichocarpa</i>	12/31	0/10	10/26	22/67
<i>Phaseolus vulgaris</i>	3/21	0/10	0/18	3/49
<i>Vitis vinifera</i>	7/26	2/23	24/69	33/118
<i>Glycine max</i>	4/18	0/18	0/09	4/45
<i>Fragaria vesca</i>	33/14	2/12	6/34	41/60
<i>Malus domestica</i>	14/36	1/25	5/57	20/118
<i>Oryza sativa</i>	3/1	1/13	4/29	8/43
<i>Citrus clementina</i>	23/8	0/12	10/1	33/21
<i>Medicago truncatula</i>	3/16	0/18	9/18	12/52
<i>Capsella rubella</i>	11/08	1/9	6/25	18/42
<i>Prunus persica</i>	8/12	0/8	0/7	8/27
<i>Cucumis sativus</i>	3/11	0/4	1/17	4/32
<i>Carica papaya</i>	14/12	1/11	3/18	18/41
<i>Gossypium raimondii</i>	12/27	0/09	4/40	16/76
<i>Cicer arietinum</i>	16/2	0/6	0/11	16/19

# FIGURES

Figure 1

