PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol x(x): xx–xx
https://doi.org/10.14719/pst.2115

HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

Evolutionary Insights into the Enzymes involved in the Biosynthesis of the Volatile Organic Compounds Isoprene and Pinene in Plants

Arunima Bhattacharya[1,3,#], Pragyasree Bhowmick[1,#], Sayak Ganguli[1,*], Arup Kumar Mitra[2]

[1]Department of Biotechnology, St. Xavier's College (Autonomous), Kolkata-700 016, India
[2]Department of Microbiology, St. Xavier's College (Autonomous), Kolkata-700 016, India
[3]Current Address: Univ. Bordeaux, Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, ARNA, UMR 5320, U1212, Institut Européen de Chimie et Biologie, Pessac 33607, France

[#]Contributed equally
*Email: sayakgan3@gmail.com

OPEN ACCESS

ARTICLE HISTORY

Received: 12 September 2022
Accepted: 24 December 2022

Available online
Version 1.0 : 14 January 2023

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS etc.
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE

Bhattacharya A, Bhowmick P, Ganguli S, Mitra A K. Evolutionary Insights into the Enzymes involved in the Biosynthesis of the Volatile Organic Compounds Isoprene and Pinene in Plants. Plant Science Today (Early Access). https://doi.org/10.14719/pst.2115

Abstract

Volatile organic compounds (often abbreviated as VOCs) are emitted as secondary metabolites by plants, and contribute to a wide range of ecological processes, owing to their pivotal role in plant interactions with biotic and abiotic variables. As a result, they differ greatly between species and explain disparities in ecological strategy. In an effort to comprehend their genesis and assess potential evolutionary trends, this work probes into the enzymatic pathways that lead to their synthesis. Correspondingly, we adopt and propose an *in-silico* approach to analyze connections between the species evolution and the gene evolution of two major plant volatile organic compounds. We lay focus on isoprene and pinene, volatile organic compounds synthesized by two common yet compartmentally isolated pathways - the methylerythritol phosphate (MEP) pathway and the mevalonic acid (MVA) pathway, respectively. Analyses of gene-specific and protein-specific phylogenetic trees of the enzymes involved in these pathways thereby indicate a mixed trend in the evolution as per the APG IV (Angiosperm Phylogeny Group IV) system. These results and the *in-silico* pipeline thus provide us with future opportunities to explore different networks of plant communication for a holistic understanding of intraspecific and interspecific interactions in different natural ecosystems .

Keywords

Volatile organic compounds; phylogenetic tree; APG IV; isoprene; pinene; methylerythritol phosphate; mevalonic acid; plant communication.

Introduction

The evolutionary history of plants provides us with the best evidence that these organisms prefer to survive in structured communities. The sustenance of these communities is heavily dependent on finding mechanisms of effective communication between the present members of a plant association (1). Volatile organic compounds (VOCs), one of the most significant plant secondary metabolites, are considered to be the key players in this cascade of communication leading to robust evolutionary events as they are released *via* their interactions with biotic and abiotic stimuli. At room temperature, these lipophilic substances have a low molecular weight and high vapour pressures. The physical attributes of these chemicals allow them to pass readily through cellular membranes and be discharged into the surrounding environment (1 - 3). More than 1700 VOCs have been found in various angiosperm and gymnosperm taxa, spanning 90 families and 38 orders (2,4). They are usually emitted by flowers, although they can also be found in fruits, leaves,

stems, and roots. The chemical landscapes of various ecosystems are shaped by VOCs, which participate in intra- and interspecific interactions, are extremely context-dependent and operate in both direct and indirect ways from the landscape to the intrafloral scale (2,3). Plant VOCs influence plant-pollinator, plant-herbivore, plant–plant, and other interactions, as well as plant fitness. The most significant and well-understood of its roles is the attraction of pollinators, which ensures the reproductive success of the plant. Plant-pollinator specialization, as well as outcrossing and reproductive isolation, is aided by floral constancy (1,2). As a result, such a sexual signal is subject to intense selection pressure, and it is essential for plant development and adaptability to the environment. Despite this, VOCs have a vital role in plant defence against herbivores and pathogen protection. Due to their inherent or augmented release during herbivory attacks, VOCs can act as repellents (3,4). Moreover, certain VOCs can function as indirect defenses by attracting parasitoids and predators, lowering the number of herbivores, notably insects and larvae. Furthermore, for diagnostic objectives, plants emit bacteria-specific VOCs, which activate defensive signaling pathways and operate as direct inhibitors of bacterial growth, making plants more impervious to pathogen invasion (4).

Biosynthesis of VOCs

VOC biosynthesis is reliant on the availability of carbon, nitrogen, and sulphur, as well as energy from primary metabolism. The accessibility to these building blocks has a considerable impact on the concentration of any secondary metabolite, including VOCs, due to the high degree of crosstalk between primary and secondary metabolism. In the production of the vast variety of VOCs, merely a few fundamental metabolic pathways are involved (4). VOCs are categorized into many families based on their biosynthetic origin, including terpenoids, phenylpropanoids/ benzenoids, fatty acid derivatives, and amino acid derivatives, as well as a few species-/genus-specific compounds not included in those major classes (2,5).

Terpenoids: Terpenoids are the most abundant and diversified family of secondary metabolites, with many volatile compounds produced from two common five-carbon precursors, isopentenyl diphosphate (IPP) and its allylic isomer, dimethylallyl diphosphate (DMAPP). Plants produce these C5-isoprene building blocks using two independent, compartmentally isolated pathways: methylerythritol phosphate (MEP) and mevalonic acid (MVA) (5, 6). The MVA route yields volatile sesquiterpenes (C15), whereas the MEP pathway generates volatile hemiterpenes (C5), monoterpenes (C10), and diterpenes (C20). The MEP route is referred to as "plastidic" since experimental findings and predicted subcellular localization implies that a complete set of the corresponding enzymes occurs solely in plastids. The subcellular location of the MVA route, on the other hand, is less known. This pathway was formerly thought to be cytosolic; however, fresh data reveals that it is dispersed throughout the peroxisomes, endoplasmic reticulum, and cytosol (4,6,7). The MVA pathway is comprised primarily of eight enzymatic activities that start with the progressive condensation of three molecules of acetyl-CoA to 3-hydroxy-3-methylglutaryl-CoA, which is subsequently reduced to MVA by two sequential phosphorylations and a decarboxylation/elimination step, yielding IPP. The *Arabidopsis* genome has two genes that produce acetoacetyl-CoA thiolase (AACT), one of which (AACT2) catalyses the first step in the MVA pathway and is found in peroxisomes, according to proteome analysis. The *Arabidopsis* genome encodes two genes that synthesize acetoacetyl-CoA thiolase (AACT), one of which (AACT2) catalyses the first step in the MVA pathway and is present in peroxisomes, according to proteome analysis. The MEP pathway is composed primarily of nine enzymatic steps that begin with the condensation of D-glyceraldehyde-3-phosphate (GAP) and pyruvate (Pyr) to form 1-deoxy-D-xylulose-5-phosphate, which would then be isomerized/reduced to form the process's unique intermediate, MEP (8). Converting MEP to IPP and DMAPP is a five-step process. Pyr and GAP are supplied to the MEP system by primary metabolism, with the latter acquired via both glycolysis and the pentose phosphate pathway (PPP). Short-chain prenyltransferases can then create geranyl diphosphate (GPP), which are prenyl diphosphate precursors, alongside farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP), that can subsequently be acted on by a vast family of terpene synthases/cyclases (TPSs). For example, although IPP is transformed to pinene through a geranyl diphosphate intermediate in the MVA route, the MEP pathway's direct precursor of isoprene is DMAPP (5, 6). While the MEP pathway yields both IPP and DMAPP in a 6: 1 ratio, the MVA route however solely generates IPP. IPP, DMAPP, and short prenyl diphosphates (GPP and FPP) act as bridging metabolites, allowing metabolic interaction between the compartmentally separated MVA and MEP pathways. Due to the isoprenoid biosynthetic pathways being interconnected, the MEP pathway, which typically has a higher carbon flux than the MVA route, can facilitate terpenoid synthesis in the cytosol (5,9).

Phenylpropanoids/ Benzenoids: The second largest class of plant VOCs is composed of phenylpropanoid and benzenoid compounds. This is derived from the aromatic amino acid phenylalanine (Phe). Phe is linked to central carbon metabolism *via* seven enzyme processes of the shikimate pathway and three of the arogenate pathway. Phosphoenolpyruvate (PEP) and D-erythrose-4-phosphate (E4P), the shikimate pathway's immediate precursors, are produced from glycolysis and the PPP, respectively (4). Given that the same mechanisms provide precursors for the MEP system, the latter has to contest for carbon allocation with the shikimate/phenylpropanoid route, especially given that Phe gets 30% of photosynthetically fixed carbon, mostly to form lignin. The first gene in the shikimate pathway, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (DAHP synthase), is critical in adjusting carbon intake in the process (5,10). However, the molecular processes behind this control in plants are mostly unexplored.

Derivatives of Volatile Fatty Acids: Plant VOCs even include fatty acid derivatives generated from C-18 unsaturated fatty acids, linoleic or linolenic acids, such as 1-hexanal, cis-3-hexenol, nonanal, and methyl jasmonate. In the pro-

duction of these fatty acids, the plastidic pool of acetyl-CoA produced from Pyr, the end product of glycolysis, is utilized. After entering the lipoxygenase (LOX) pathway, unsaturated fatty acids undergo stereospecific oxygenation, releasing 9-hydroperoxy and 13-hydroperoxy intermediates, which are subsequently metabolized by the two branches of the LOX pathway to form volatile compounds (11). The allene oxide synthase branch uses the 13-hydroperoxy intermediate solely to generate jasmonic acid (JA), which is then converted to methyl jasmonate by JA carboxyl methyl transferase. The hydroperoxide lyase branch, on the other hand, converts both types of hydroperoxide fatty acid derivatives into C6 and C9 aldehydes, which are usually reduced to alcohols before being converted to their esters by alcohol dehydrogenases. The saturated and unsaturated C6/C9 aldehydes and alcohols are also known as green leaf volatiles and are typically generated in the green organs of plants in response to injury, but they also offer the distinctive 'fresh green' aroma of fruits and vegetables (5,11).

Volatile branched-chain amino acids derivatives: Numerous volatile chemicals, particularly those seen in flower smells and fruit odours, are produced from amino acids or intermediates in their formation and contain nitrogen and sulphur. The manufacture of these amino acid-derived volatiles in plants is considered to be analogous to that reported in bacteria or yeast, where these pathways have garnered more attention (12). As in microbes, aminotransferases catalyse the first deamination or transamination of the amino acids, resulting in the synthesis of the matching α-ketoacid (13). Such α-ketoacids can be further decarboxylated, then reduced, oxidised, and/or esterified to produce aldehydes, acids, alcohols, and esters (3).

Significance of VOCs

Insect-Plant Interactions: Floral VOCs have long been recognized as important in attracting pollinators. The importance of tri-trophic interactions among plants, insect herbivores, and natural enemies of herbivores was not recognized until the 1980s when Price along with other researchers highlighted the importance of tri-trophic interactions among plants, insect herbivores, and natural enemies of herbivores (2). Plant VOCs appear to have a prominent role in plant-plant dissemination of defensive responses, for example. When it comes to VOC detection in biological systems, insects have been shown to detect scents through their peripheral nervous system (14). Several subsequent investigations established the significance of herbivore-induced plant VOCs (HIVOCs) as host location cues for parasitoids and herbivore predators. This indirect chemical defense is likely to be as important as or more important than direct chemical and physical defenses in terms of reducing herbivore damage. When a herbivore consumes a plant, many VOC production pathways are activated. While terpenes and sesquiterpenes play a pivotal role, several additional types of volatiles are created depending on the plant species attacked (15).

Plant to Plant Communication: While animal reactions to VOCs have been widely studied, particularly in insects, plant-to-plant communication is significantly less understood. In the last 10-15 years, it has been conclusively demonstrated that plants detect and respond to VOCs and that this capability is an important defense mechanism. Plants exposed to VOCs from neighbouring plants that have been harmed by insect herbivores might build their defenses and respond more quickly and powerfully if attacked later by the herbivores. As a result, HIVOCs not only attract natural herbivore enemies, but also act as warning signals for adjacent plants. In rare cases, VOCs can act as intra-plant signals between injured leaves and those far from the source of harm (15).

However, it is still unclear how plants detect and respond to various VOCs. Nonetheless, studies have demonstrated altered plant defense in response to mVOCs or nearby plant VOC emissions (16 – 21). Brosset and Blande (22) evaluate current information on plant-plant signaling where the authors describe 40 distinct plant-derived VOCs that have been shown to modify receiver plant resistance or defense-related internal signals. As a result, it is probable that detection of diverse VOCs and consequent changes in plant internal signaling rely on more than one or more unique receptors in plants. Brosset and Blande (18) review recent evidence that suggests a distinct method of VOC detection in receiver plants, including absorption and subsequent conversion of VOCs to soluble metabolites. Metabolizing VOCs might be a useful functional method of plant-plant communication since it allows compound-specific responses in a dose-dependent way. It is still difficult to ascertain how the various volatile cues are incorporated into a plant response (23).

Mimicry and Eavesdropping: Although a molecule or mixture of molecules may operate as a very specific pheromone for members of the species that generates it, predators and parasites routinely intercept these signals that dependably identify the presence of their prey or hosts. Some orchids produce sex pheromones that attract male bees and hence aid in pollination. Other blossoming plants smell like decaying meat or animal waste, which attracts flies and other insects that consume and/or oviposit on these items (16).

As a result, VOC signals transmit a wide variety of communications between and among organisms from the planet's most diversified spectrum of life forms. Depending on the receiver, the same chemical or chemical combination may have several meanings. They are widespread in nature and critical to ecosystem functioning (16).

However, even though scientists have been studying the composition and function of VOC signals for over 50 years, we are yet to resolve the evolutionary origin of these compounds. Through this study, laying our focus primarily on the MVA and MEP pathways, we investigate and discern a possible correlation between the species and gene evolution to aid our understanding of the origin of volatile organic compounds and evaluate its future trend patterns.

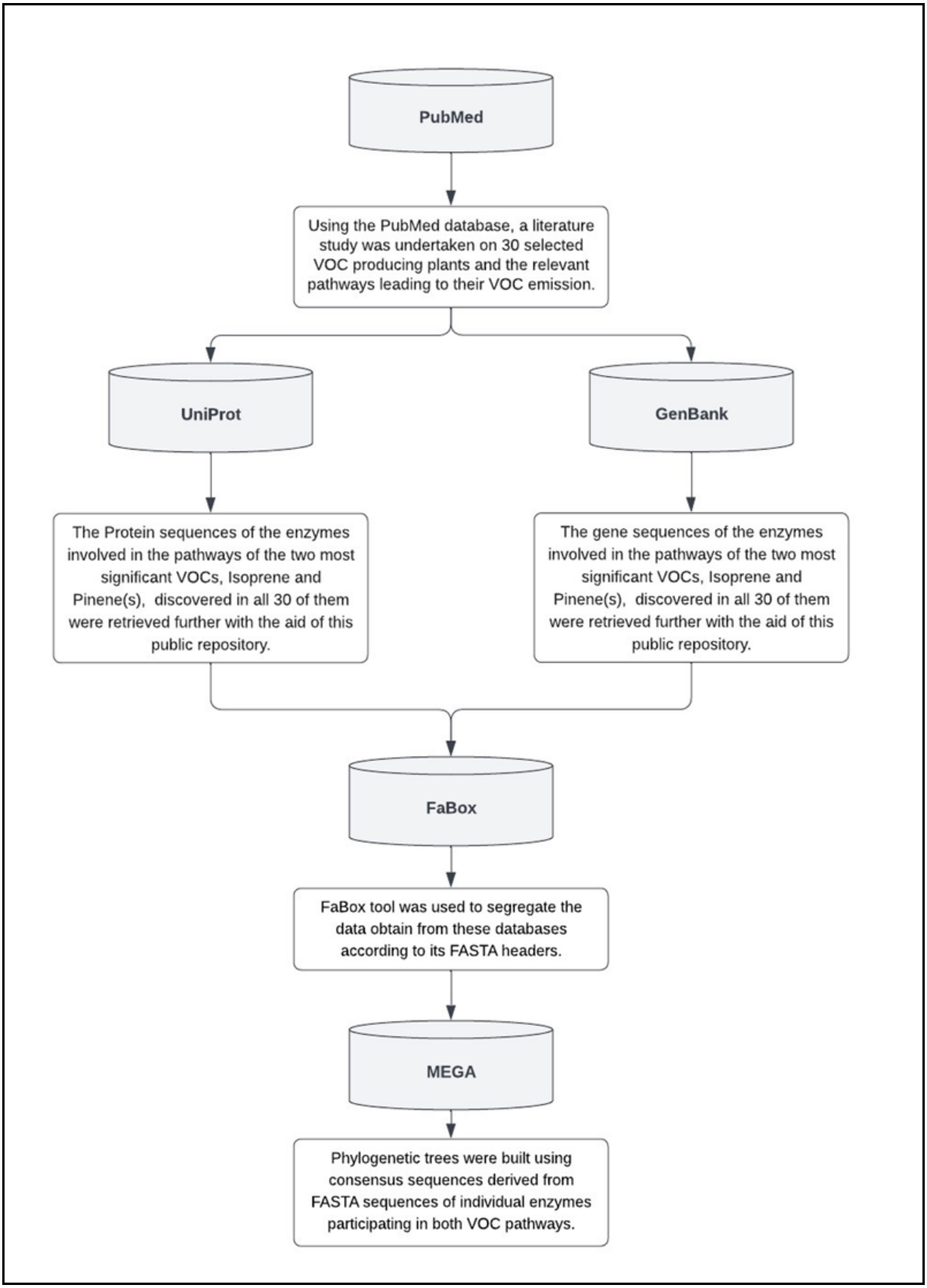


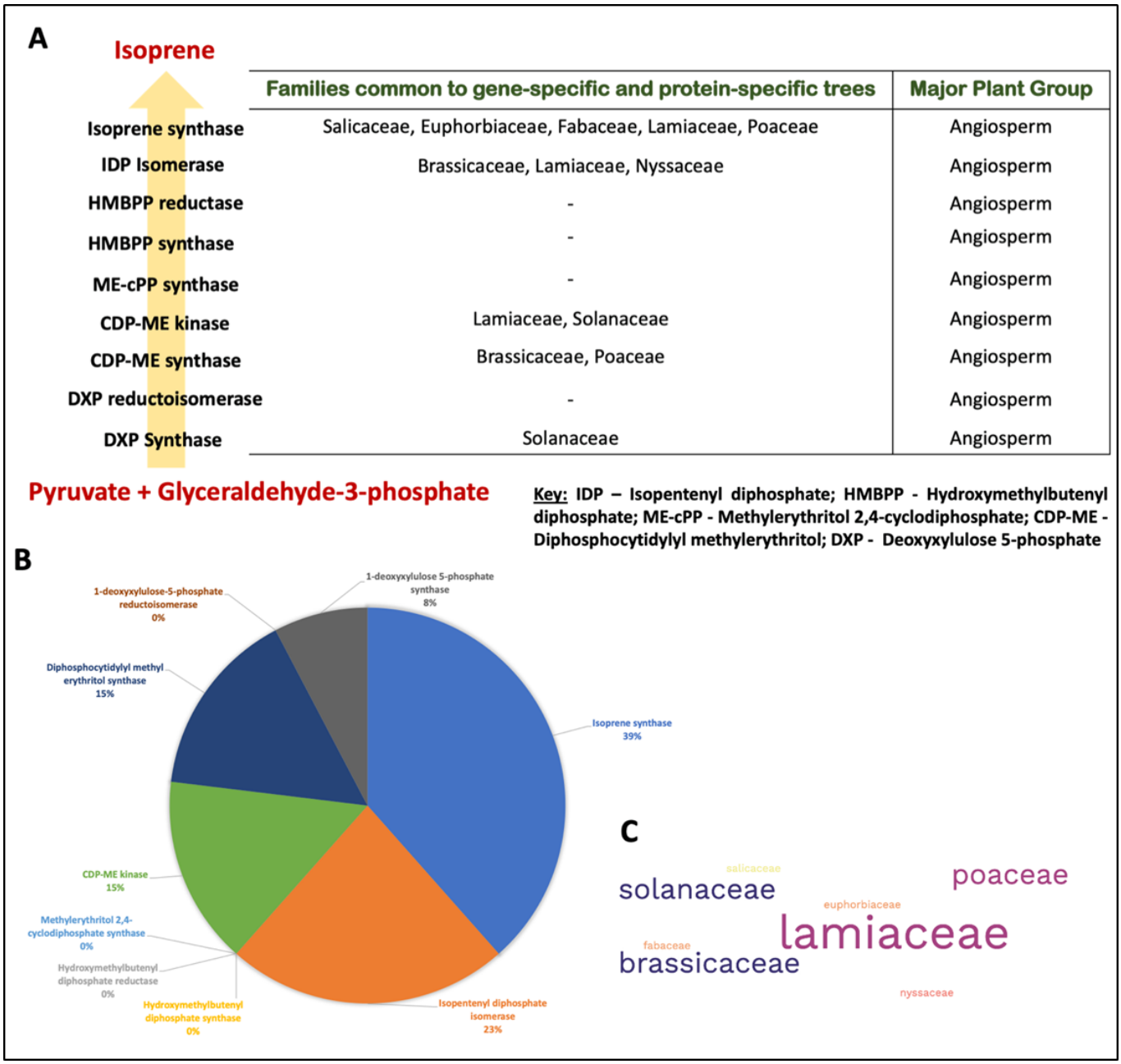
Fig. 1: Flowchart depicting and explaining the databases that were used in the analysis pipeline

Materials and Methods

Initially, a text mining-based approach was adopted to identify the top 28 plants reported in literature emitting VOCs. This list was then further curated based on the type of VOC emitted (Supplementary Table 1). This step allowed us to identify that isoprene and pinene were the two most common VOCs that were emitted ubiquitously by the 28 plants identified in step one. Following this, the pathways of synthesis of these two compounds were studied in detail to note down the enzymes that play key roles. Next, NCBI GenBank and UniProt databases were mined to obtain the sequences of the enzymes of the pathways. These sequences were then subjected to multiple sequence alignment followed by phylogenetic reconstruction using

Table 1: Description of gene-specific and protein-specific phylogenetic trees for enzymes in the methylerythritol phosphate (MEP) and mevalonic acid (MVA) pathways

MEP Pathway		
Enzyme	**Gene-specific tree**	**Protein-specific tree**
Deoxyxylulose 5-phosphate (DXP) Synthase	Mostly composed of dicots, with an occasional presence of monocots, algae and gymnosperms	Mixed clustering of plant families with sister groups from both monocots and dicots.
Deoxyxylulose 5-phosphate (DXP) Reductoisomerase	Small tree – not suitable for analysis	No tree generated – less than 4 sequences retrievable.
Diphosphocytidylyl methyl erythritol (CDP-ME) Synthase	Distinct clustering of plant families – one of them composed majorly of gymnosperms and algae, the others composed of sister groups from both monocots, dicots and algae.	Mixed clustering of plant families with sister groups from both monocots and dicots
Diphosphocytidylyl methyl erythritol (CDP-ME) Kinase	Mixed clustering of plant families with sister groups from both monocots, dicots and algae	Small tree – not suitable for analysis
Methylerythritol 2,4-cyclodiphosphate (ME-cPP) Synthase	Small tree – not suitable for analysis	Mixed clustering of plant families with sister groups from both monocots and dicots
Hydroxymethylbutenyl diphosphate (HMBPP) Synthase	Mixed clustering of plant families with sister groups from both monocots and dicots	No tree generated – less than 4 sequences retrievable.
Hydroxymethylbutenyl diphosphate (HMBPP) Reductase	Two major clusters of plant families – one composed of sister groups majorly of asterids, the other composed majorly of sister groups from commelinids	No tree generated – less than 4 sequences retrievable.
Isopentenyl diphosphate Isomerase	Mostly composed of dicots, with an occasional presence of algae	Mostly composed of dicots, with an occasional presence of monocots
Isoprene Synthase	Mixed clustering of plant families with sister groups from both monocots and dicots	Mixed clustering of plant families with sister groups from both monocots and dicots
MVA Pathway		
Enzyme	**Gene-specific tree**	**Protein-specific tree**
Acetoacetyl-CoA thiolase	Mixed clustering of plant families with sister groups from both monocots and dicots	Small tree – not suitable for analysis
3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) Synthase	Mostly composed of dicots, with an occasional presence in monocots, algae and gymnosperms	Small tree – not suitable for analysis
3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) Reductase	Two major clusters of plant families – one composed of eudicots, and the other of monocots	Two major clusters of plant families – one composed of eudicots, and the other of monocots
Mevalonate-5-kinase	Mixed clustering of plant families with sister groups from both monocots and dicots	Small tree – not suitable for analysis
Phosphomevalonate kinase	Mixed clustering of plant families with sister groups from both monocots and dicots	Small tree – not suitable for analysis
Mevalonate pyrophosphate decarboxylase	Mixed clustering of plant families with sister groups from both monocots and dicots	Small tree – not suitable for analysis
Geranyl diphosphate synthase	Distinct clustering of plant families – one composed majorly of gymnosperms, the other is composed of sister groups from both monocots and dicots	Distinct clustering of plant families – one composed majorly of gymnosperms, the other is a mixed cluster, composed of sister groups from both monocots and dicots
Pinene synthase	Two major clusters of plant families – one composed of sister groups majorly from gymnosperms, the other composed of sister groups from core eudicots	Two major clusters of plant families – one composed of sister groups majorly from gymnosperms, the other composed of sister groups from core eudicots



Families common to gene-specific and protein-specific trees	Major Plant Group
Salicaceae, Euphorbiaceae, Fabaceae, Lamiaceae, Poaceae	Angiosperm
Brassicaceae, Lamiaceae, Nyssaceae	Angiosperm
-	Angiosperm
-	Angiosperm
-	Angiosperm
Lamiaceae, Solanaceae	Angiosperm
Brassicaceae, Poaceae	Angiosperm
-	Angiosperm
Solanaceae	Angiosperm

Fig. 2: The MEP pathway. (A) The major enzymes involved in the conversion of pyruvate and glyceraldehyde-3-phosphate to isoprene via the MEP pathway have been shown. The table illustrates the families common to both gene-specific and protein-specific phylogenetic trees for each of the enzymes in the MEP pathway, alongside the major plant group/ evolutionary lineage of occurrence, identified independently for each enzyme. (B) The enrichment profile of the enzymes common to both gene-specific and protein-specific trees has been represented as a pie chart. Isoprene synthase is the most enriched enzyme in the MEP pathway, with an enrichment value of 39%. (C) The prevalence of common families has been represented as a tag cloud, which shows that Lamiaceae is the most represented family.

the neighbour-joining method with 100 bootstrap iterations using MEGA and the consensus sequences of each were obtained. Following this, the trees were analyzed and an enrichment table (Supplementary Table 2) was prepared in a family-specific manner which embodied the occurrence of the enzymes corresponding to each family. This enabled us to compare the families with the APG IV (Angiosperm Phylogeny Group) system (18, 21) to check the phylogenetic and evolutionary fidelity. A summary of the workflow is provided in Fig. 1.

Results

Gene-specific and protein-specific phylogenetic trees were constructed for 9 major enzymes in the MEP pathway and 8 major enzymes in the MVA pathway (Supplementary Figures). Out of the 34 expected trees (including 18 from the MEP pathway and 16 from the MVA pathway), 3 protein-specific trees in the MEP pathway corresponding to DXP reductoisomerase, HMBPP synthase and HMBPP reductase could not be generated due to the non-availability of a minimum number of sequences. Thus, a total of 31 phylogenetic trees were taken into consideration and analysed. Thereafter, the clustering of members from different families of plants was identified for comparison with the APG IV classification for the orders and families of flowering plants (20, 23). A brief description of the trees has been included in Table 1.

Next, we identified families that were common to both types of trees for each enzyme in the pathways,

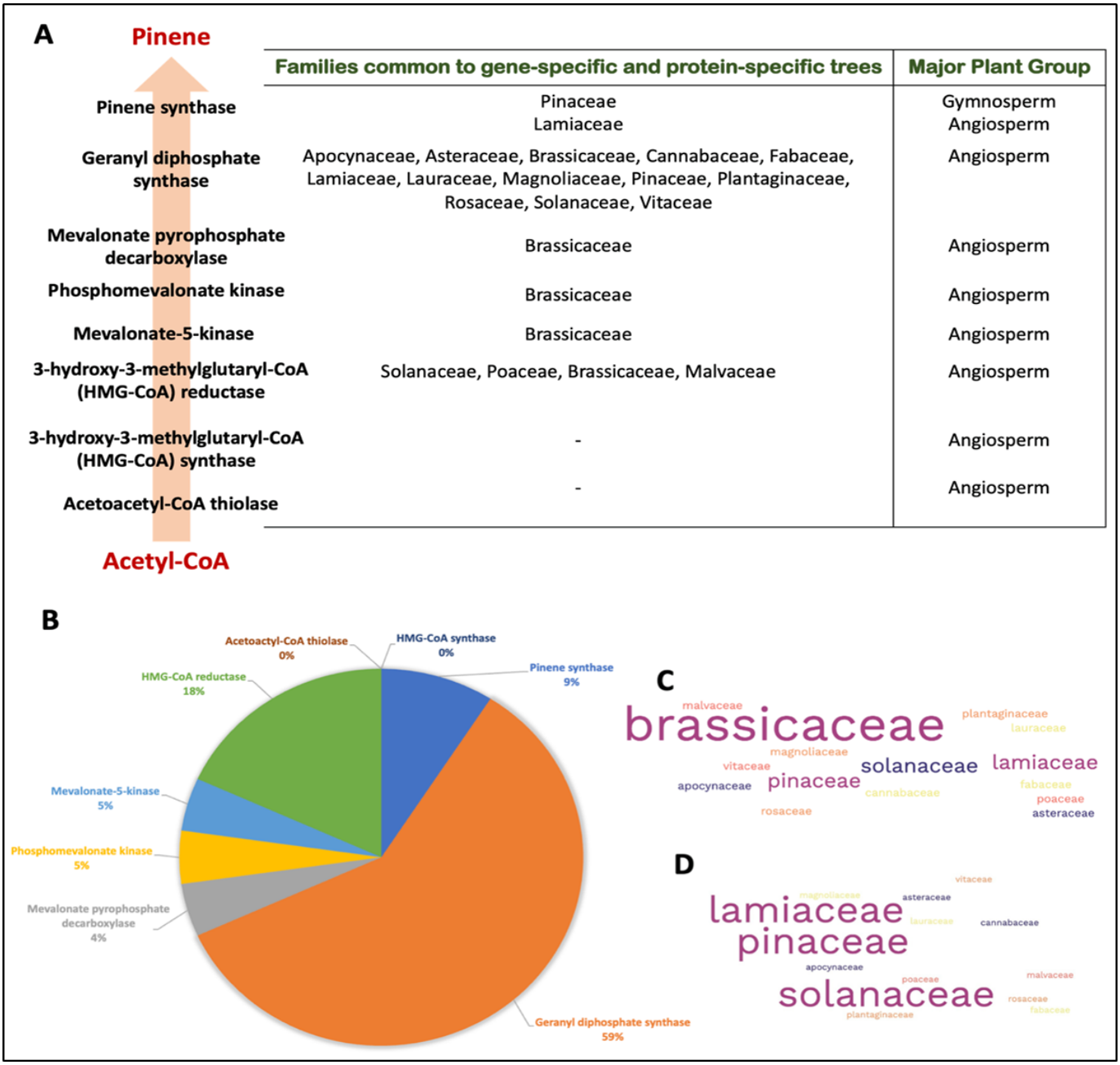


Fig. 3: The MVA pathway. (A) The major enzymes involved in the conversion of acetyl-CoA to pinene via the MVA pathway have been shown. The table illustrates the families common to both gene-specific and protein-specific phylogenetic trees for each of the enzymes in the MVA pathway, alongside the major plant group/ evolutionary lineage of occurrence, identified independently for each enzyme. **(B)** The enrichment profile of the enzymes common to both gene-specific and protein-specific trees has been represented as a pie chart. Geranyl diphosphate synthase is the most enriched enzyme in the MVA pathway, with an enrichment value of 59%. **(C)** The prevalence of common families has been represented as a tag cloud, which shows that Brassicaceae is the most represented family. **(D)** If Brassicaceae is selectively removed from consideration due to probable experimental bias (see text), the tag cloud indicates a high prevalence of the families Lamiaceae, Pinaceae and Solanaceae.

and an enrichment profile of the enzymes comparing the number of common families was developed. While isoprene synthase was the most enriched enzyme in the MEP pathway, with an enrichment value of 39%, and having 5 common families in the gene-specific and protein-specific trees, geranyl diphosphate synthase was the most enriched in the MVA pathway with a value of 59% and having 13 common families.

Thereafter, the prevalence of the common families was determined for each pathway, and it was observed that while Lamiaceae was the most represented family in the MEP pathway, Brassicaceae is highly prevalent in the MVA pathway (Figures 2 and 3). However, it should be noted that the genus *Arabidopsis*, a model experimental plant and a member of Brassicaceae, may have contributed to an over-representation of the family, leading to an experimental bias. In that case, the other prevalent families in the MVA pathway were found to be Lamiaceae, Solanaceae and Pinaceae. Finally, the major evolutionary lineages of plants (angiosperm/gymnosperm) from the two types of phylogenetic trees were independently identified (Figs. 2 and 3). Overall, the highest number of common genes and proteins were identified in angiosperms, except the enzyme pinene synthase in the mevalonate pathway, which had a considerable representation in gymnosperms alongside angiosperms. This bias may also be attributed to the data available in the data repositories where a large number of angiosperms have been studied, for which more data were available for that group.

Discussion

It is largely believed that the origin of scent compounds in plants can be attributed to a couple of evolutionary phenomena - the first being gene duplication and the other being divergence - convergence (24, 26). Genes have also been reported to have undergone structural changes because of pollinator interactions, which may have led to specific speciation events (26). Some workers believe that alteration in the expression levels of the genes under specific conditions may also have a long-term impact on the evolution of certain compounds (28, 29). The changes in the expression level may have epigenetic causes as well which if found to be beneficial for the plant will be transmitted to the next generation and thus inherited and evolutionarily accepted (30). We do find such evidence in *Brassica rapa* (31). Another key contributor is whole genome duplication which, in synergy with epigenetic regulators, influence changes in plant floral traits.

Analysis of the phylogenetic trees generated in this analysis allowed their segregation into three major types – a) trees that entirely showed similar evolutionary patterns as per the APG IV system (24) (b) trees that showed a mixed pattern with only small stretches showing similar evolutionary patterns as per the APG IV system, and c) small trees with only a few recurring members which could not be analyzed for gaining evolutionary insights.

Trees with evolutionary patterns similar to APG IV

In the MEP pathway, none of the trees obtained was found to be strictly consistent with the APG IV system although, in the MVA pathway, pinene synthase showed accordance with the APG IV system entirely, both in the gene-specific and the protein-specific trees. In each tree, two distinct clusters were observed – the larger cluster was dominated by members from Pinaceae, which is a gymnosperm family, while the other cluster was composed of sister groups from the closely-related core eudicots, belonging to asterids and rosids, with no contribution of monocots, magnoliids, or more ancient orders, which may, however, be attributed to data deficiency due to lesser experiments with such plants (24). Another example is the gene-specific tree of HMG CoA reductase, which has two major clusters – one with eudicots from Brassicaceae, Plantaginaceae, Lamiaceae, Cucurbitaceae, Solanaceae, Fabaceae, Malvaceae, Rubiaceae, and another smaller cluster from monocots, especially commelinids like Melanthiaceae, Poaceae, Orchidaceae.

Such evolutionary patterns underline the concept of molecular evolution, indicating that the complexity of the plant system increased in tandem with the increase in diversity of the sequence of the enzyme encoded as earlier reported in acyl-activating enzymes (32) and diacylglycerol acyltransferases (33).

Trees with mixed evolutionary patterns

The MEP pathway was dominated by this pattern with evolutionarily advanced and primitive families placed close together, as was seen in the case of the gene-specific trees of CDP-ME kinase and synthase where a mixed evolutionary pattern was observed, composed of algae (Chlorophyceae), gymnosperms (Taxaceae, Ginkgoaceae, and Pinaceae), a wide variety of dicots and monocots such as Euphorbiaceae, Lamiaceae, Solanaceae, Brassicaceae, Asteraceae, Apocynaceae, Rubiaceae, Brassicaceae, Poaceae, Euphorbiaceae, Solanaceae, Ranunculaceae, Aizoaceae, Rutaceae and Zingiberaceae. The gene-specific tree of the enzyme isoprene synthase, which likewise displayed a mixed clustering pattern, shared 5 common families with the gene-specific tree (Figure 2), achieving the greatest enrichment value in the MEP pathway. In addition to this, the gene-specific trees of HMBPP reductase (dominated by asterids and commelinids), HMBPP synthase, DXP synthase, and isopentenyl diphosphate Isomerase - comprise sister groups from algae, dicots, monocots (Poales), and gymnosperms (Pinaceae). Several workers attribute this apparent polyphyletic pattern to the coevolution of these plant groups based on habitat conditions exemplified in plant receptor-like kinases (30 - 33) and the players of the plant immune system (36).

In the MVA pathway as well, most of the enzymes exhibited this pattern. For example, both the gene-specific and protein-specific trees of geranyl diphosphate synthase showed a mixed evolutionary pattern. In the gene-specific tree, one cluster was composed of members of Pinaceae, which are gymnosperms, while the other larger cluster had members from a wide variety of monocot and dicot families, such as Lamiaceae, Plantaginaceae, Rubiaceae, Brassicaceae, Lauraceae, Vitaceae, Solanaceae, Salicaceae, Rosaceae, Apocynaceae, Canabaceae, Asteraceae, Fabaceae, and Orchidaceae. The protein-specific tree also had a similar pattern, with 13 common families with the gene-specific tree (Figure 2), and the highest enrichment value in the MVA pathway. Similarly, gene-specific trees of most other enzymes in the MVA pathway showed a mixed pattern – e.g., mevalonate pyrophosphate decarboxylase (dominated by rosids and asterids, but also has monocots like Alismatales and Poales, and Ranunculales, which is a transition between monocots and eudicots), phosphomevalonate kinase, mevalonate-5-kinase, HMG-CoA synthase, and acetoacetyl CoA thiolase– these have sister groups from core eudicots like asterids and rosids, as also monocots (Poales, Alismatales), and occasionally from Pinaceae (gymnosperms). These enzymes may have the property of enzyme promiscuity, where at very low levels, they catalyze non-canonical reactions thus adding to the metabolic noise on which natural selection acts, selecting for certain enzymes (36). The environmental conditions have also been reported to influence the organism's “fitness landscape” and the more flexible the metabolic enzymes are, the better adapted the organism is to manoeuvre across and achieve adaptive evolution (37). These promiscuous modules contribute significantly

toward the "evolvability" of the metabolic pathways and diversity of specialised metabolism (17).

Small trees

The protein-specific trees for CDP-ME kinase and gene-specific tree for ME-cPP synthase of the MEP pathway had very few representative families, and recurrence of the same members of the families was observed, as was the case for protein-specific trees for acetoacetyl CoA thiolase, HMG CoA synthase, mevalonate-5-kinase, phosphomevalonate kinase and mevalonate pyrophosphate decarboxylase of the MVA pathway. For example, the protein-specific tree for CDP-ME kinase had only 3 representative families Solanaceae, Lamiaceae and Brassicaceae while in the protein-specific tree for phosphomevalonate kinase, all 4 members of the tree were from different strains of *Arabidopsis thaliana*, belonging to Brassicaceae. Thus, such trees could not be analyzed for the determination of evolutionary significance due to the non-availability of sufficient comparative data.

Apart from these three types, a few interesting evolutionary anomalies were observed within the trees. For example, in the gene-specific tree for DXP synthase, one sister group consisted of members from Rutaceae (a core eudicot family, rosid) and Pinaceae (gymnosperm) alongside the protein-specific tree for geranyl diphosphate synthase, where one sister group had members from Rosaceae (a core eudicot family) and Selaginellaceae (a pteridophyte family), while another sister group had *Pseudotsuga menziesii* (Pinaceae, gymnosperm) and *Cistus creticus* subsp. *critecus* (Cistaceae, angiosperm), marking the gymnosperm – angiosperm transition in the tree. Again, in the gene-specific tree of HMG CoA synthase, an anomalous sister group was found with members from Poaceae (a monocot family) and Mesostigmataceae (an algal family) while interestingly, in the gene-specific tree of CDP-ME synthase, 4 anomalous sister groups were found with members from Taxaceae (a gymnosperm family) and Chlorophyceae (an algal family), Euphorbiaceae (a core eudicot family, rosid) and Chlorophyceae, Brassicaceae (a core eudicot family, rosid) and Chlorophyceae and finally, Poaceae (a monocot family) and Chlorophyceae.

Conclusion

The phylogenetic analyses showed a mixed trend in the evolution of the enzymes involved in the VOC synthesis cascades. Some adhere well to the standard APG IV arrangement of the families while others veer out of the trend. The authors believe that the anomalies observed and reported are largely due to the lack of experimental data and may exhibit a different pattern once more data accumulates in public repositories. We are still in the infancy of our understanding of how microbial VOCs in the rhizosphere and phyllosphere engage in crosstalk with the plant VOCs, which may also be regulated by the prevalent abiotic conditions. With the threat of climate change looming large, the build-up and sustenance of our future communities may well be reliant on the elucidation and characterization of such intraspecific and interspecific interactions in natural ecosystems.

Acknowledgements

The authors would like to acknowledge Rev. Fr. Dr. Dominic Savio SJ, Principal, St. Xavier's College (Autonomous), Kolkata for the constant encouragement. They also remain grateful to Dr. Jhimli Dasgupta, Head of the Department of Biotechnology, St. Xavier's College (Autonomous), Kolkata, for her support.

Authors contributions

AB and PB performed the analysis, while SG and AKM supervised and planned the work. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None.

References

1. Picazo-Aragonés J, Terrab A, Balao F. Plant volatile organic compounds evolution: Transcriptional regulation, epigenetics and polyploidy. Int J Mol Sci. 2020;21(23):8956. http://dx.doi.org/10.3390/ijms21238956
2. Effah Evans, Holopainen Jarmo K. Clavijo McCormick, Andrea. Potential roles of volatile organic compounds in plant competition. Perspectives in plant ecology evolution and systematics, 2019:38,58-63. http://dx.doi.org/10.1016/j.ppees.2019.04.003
3. Kigathi, R. N., Weisser, W. W., Reichelt, M., Gershenzon, J., &Unsicker, S. B. Plant volatile emission depends on the species composition of the neighboring plant community. BMC plant biology, 2019;19(1):1-17. https://doi.org/10.1186/s12870-018-1541-9
4. Dudareva N, Klempien A, Muhlemann JK, Kaplan I. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytol. 2013;198(1):16–32. http://dx.doi.org/10.1111/nph.12145
5. Vivaldo G, Masi E, Taiti C, Caldarelli G, Mancuso S. The network of plants volatile organic compounds. Sci Rep. 2017;7(1):11050. http://dx.doi.org/10.1038/s41598-017-10975-x
6. Zhao L, Chang W-C, Xiao Y, Liu H-W, Liu P. Methylerythritol phosphate pathway of isoprenoid biosynthesis. Annu Rev Biochem. 2013;82(1):497–530. http://dx.doi.org/10.1146/annurev-biochem-052010-100934
7. Jozwiak A, Lipko A, Kania M, Danikiewicz W, Surmacz L, Witek A, et al. Modeling of dolichol mass spectra isotopic envelopes as a tool to monitor isoprenoid biosynthesis. Plant Physiol. 2017;174(2):857–74. http://dx.doi.org/10.1104/pp.17.00036
8. Nagegowda DA. Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. FEBS Lett. 2010;584(14):2965–73. http://dx.doi.org/10.1016/j.febslet.2010.05.045
9. de Souza VF, NiinemetsÜ, Rasulov B, Vickers CE, Duvoisin Júnior S, Araújo WL, et al. Alternative carbon sources for isoprene emission. Trends Plant Sci. 2018;23(12):1081–101. http://dx.doi.org/10.1016/j.tplants.2018.09.012
10. Sato N, Kishida M, Nakano M, Hirata Y, Tanaka T. Metabolic engineering of shikimic acid-producing Corynebacterium

glutamicum from glucose and cellobiose retaining its phosphotransferase system function and pyruvate kinase activities. Front BioengBiotechnol. 2020;8:569406. http://dx.doi.org/10.3389/fbioe.2020.569406
11. Matsui K, Koeduka T. Green leaf volatiles in plant signaling and response. SubcellBiochem. 2016;86:427–43. http://dx.doi.org/10.1007/978-3-319-25979-6_17
12. Parthasarathy A, Borrego EJ, Savka MA, Dobson RCJ, Hudson AO. Amino acid-derived defense metabolites from plants: A potential source to facilitate novel antimicrobial development. J Biol Chem. 2021;296(100438):100438. http://dx.doi.org/10.1016/j.jbc.2021.100438
13. Gonda I, Bar E, Portnoy V, Lev S, Burger J, Schaffer AA, et al. Branched-chain and aromatic amino acid catabolism into aroma volatiles in Cucumis melo L. fruit. J Exp Bot. 2010;61(4):1111–23. http://dx.doi.org/10.1093/jxb/erp390
14. Maurya A K. Application of plant volatile organic compounds (VOCs) in agriculture. *In* New Frontiers in Stress Management for Durable Agriculture (pp. 369-388). 2020. Springer, Singapore. ISBN-10 9811513244
15. Zhou S, Jander G. Molecular ecology of plant volatiles in interactions with insect herbivores. J Exp Bot. 2022; 73(2):449–62. http://dx.doi.org/10.1093/jxb/erab413
16. Tumlinson JH. The importance of volatile organic compounds in ecosystem functioning. J Chem Ecol. 2014; 40(3):212–3. http://dx.doi.org/10.1007/s10886-014-0399-z
17. Weng J-K. The evolutionary paths towards complexity: a metabolic perspective. New Phytol. 2014;201(4):1141–9. http://dx.doi.org/10.1111/nph.12416
18. Zamioudis C, Korteland J, Van Pelt JA, van Hamersveld M, Dombrowski N, Bai Y, et al. Rhizobacterial volatiles and photosynthesis-related signals coordinate MYB72 expression in Arabidopsis roots during onset of induced systemic resistance and iron-deficiency responses. Plant J. 2015; 84(2):309–22. http://dx.doi.org/10.1111/tpj.12995
19. Martínez-Medina A, Van Wees SCM, Pieterse CMJ. Airborne signals from Trichoderma fungi stimulate iron uptake responses in roots resulting in priming of jasmonic acid-dependent defences in shoots of *Arabidopsis thaliana* and *Solanum lycopersicum*. Plant Cell Environ. 2017;40(11):2691–705. http://dx.doi.org/10.1111/pce.13016
20. Riedlmeier M, Ghirardo A, Wenig M, Knappe C, Koch K, Georgii E, et al. Monoterpenes support systemic acquired resistance within and between plants. Plant Cell. 2017;29(6):1440–59. http://dx.doi.org/10.1105/tpc.16.00898
21. Frank L, Wenig M, Ghirardo A, van der Krol A, Vlot AC, Schnitzler J-P, et al. Isoprene and β-caryophyllene confer plant resistance via different plant internal signalling pathways. Plant Cell Environ. 2021;44(4):1151–64. http://dx.doi.org/10.1111/pce.14010
22. Brosset A, Blande JD. Volatile-mediated plant-plant interactions: volatile organic compounds as modulators of receiver plant defence, growth, and reproduction. J Exp Bot. 2022;73(2):511–28. http://dx.doi.org/10.1093/jxb/erab487
23. Vlot AC, Rosenkranz M. Volatile compounds-the language of all kingdoms? J Exp Bot. 2022;73(2):445–8. http://dx.doi.org/10.1093/jxb/erab528
24. Angiosperm Phylogeny Group. APG IV: Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Gbif.org*. 2017. https://doi.org/10.15468/fzuaam
25. Pichersky E, Gang DR. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci. 2000;5(10):439–45. http://dx.doi.org/10.1016/s1360-1385(00)01741-6
26. Pichersky E, Lewinsohn E. Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol. 2011; 62(1):549–66. http://dx.doi.org/10.1146/annurev-arplant-042110-103814
27. Amrad A, Moser M, Mandel T, de Vries M, Schuurink RC, Freitas L, et al. Gain and loss of floral scent production through changes in structural genes during pollinator-mediated speciation. Curr Biol. 2016;26(24):3303–12. http://dx.doi.org/10.1016/j.cub.2016.10.023
28. Pulido P, Perello C, Rodriguez-Concepcion M. New insights into plant isoprenoid metabolism. Mol Plant. 2012; 5(5):964–7. http://dx.doi.org/10.1093/mp/sss088
29. Jantzen F, Lynch JH, Kappel C, Höfflin J, Skaliter O, Wozniak N, et al. Retracing the molecular basis and evolutionary history of the loss of benzaldehyde emission in the genus Capsella. New Phytol. 2019;224(3):1349–60. http://dx.doi.org/10.1111/nph.16103
30. Quadrana L, Colot V. Plant transgenerational epigenetics. Annu Rev Genet. 2016;50(1):467–91. http://dx.doi.org/10.1146/annurev-genet-120215-035254
31. Kellenberger RT, Desurmont GA, Schlüter PM, Schiestl FP. Trans-generational inheritance of herbivory-induced phenotypic changes in Brassica rapa. Sci Rep. 2018;8(1). http://dx.doi.org/10.1038/s41598-018-21880-2
32. Shockey J, Browse J. Genome-level and biochemical diversity of the acyl-activating enzyme superfamily in plants: Biochemistry and evolution of plant AAE proteins. Plant J. 2011;66(1):143–60. http://dx.doi.org/10.1111/j.1365-313X.2011.04512.x
33. Turchetto-Zolet AC, Christoff AP, Kulcheski FR, Loss-Morais G, Margis R, Margis-Pinheiro M. Diversity and evolution of plant diacylglycerol acyltransferase (DGATs) unveiled by phylogenetic, gene structure and expression analyses. Genet Mol Biol. 2016;39(4):524–38. http://dx.doi.org/10.1590/1678-4685-GMB-2016-0024
34. Dievart A, Gottin C, Périn C, Ranwez V, Chantret N. Origin and diversity of plant receptor-like kinases. Annu Rev Plant Biol. 2020;71(1):131–56. http://dx.doi.org/10.1146/annurev-arplant-073019-025927
35. Han G-Z. Origin and evolution of the plant immune system. New Phytol. 2019;222(1):70–83. http://dx.doi.org/10.1111/nph.15596
36. Maeda HA, Fernie AR. Evolutionary history of plant metabolism. Annu Rev Plant Biol. 2021;72(1):185–216. http://dx.doi.org/10.1146/annurev-arplant-080620-031054
36. Kashtan N, Noor E, Alon U. Varying environments can speed up evolution. Proc Natl Acad Sci U S A. 2007;104(34):13711–6. http://dx.doi.org/10.1073/pnas.0611630104

§§§