Published as: *Nat Chem Biol*. 2008 October ; 4(10): 617–623.

# Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases

**P.E. O'Maille**[1], **A. Malone**[1], **N Dellas**[1,2], **B. A. Hess Jr.**[3], **L. Smentek**[3,4], **I. Sheehan**[1], **B.T. Greenhagen**[5], **J. Chappell**[5], **G. Manning**[6], and **J.P. Noel**[1]

[1]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, Jack H. Skirball Center for Chemical Biology & Proteomics, La Jolla, CA 92037 USA [2]University of California San Diego, Department of Chemistry, La Jolla, CA 92093 USA [3]Department of Chemistry, Vanderbilt University, Nashville, Tennessee, USA [4]Institute of Physics, Nicolaus Copernicus University, 87-100 Toruń, Poland [5]University of Kentucky, Department of Plant & Soil Sciences, Plant Biology Program, Lexington, KY 40546 USA [6]The Salk Institute for Biological Studies, Razavi Newman Center for Bioinformatics, La Jolla, CA 92037 USA

## Abstract

Throughout molecular evolution, organisms create assorted chemicals in response to varying ecological niches. Catalytic landscapes underlie metabolic evolution, wherein mutational steps alter the biosynthetic properties of enzymes. We report the first systematic quantitative characterization of a catalytic landscape underlying the evolution of sesquiterpene chemical diversity. Based on our previous discovery of a set of 9 naturally occurring amino acid substitutions that functionally inter-converted orthologous sesquiterpene synthases from *Nicotiana tabaccum* and *Hyoscyamus muticus*, we created a library of all possible residue combinations ($2^9$ = 512) in the *N. tabaccum* parent. The product spectra of 418 active enzymes to reveal a rugged landscape where several minimal combinations of the 9 mutations encode convergent solutions to the inter-conversions of parental activities. Quantitative comparisons indicate context dependence for mutational effects - epistasis - in product specificity and promiscuity. These results provide a measure of the mutational accessibility of phenotypic variability among a diverging lineage of terpene synthases.

---

The acquisition of innovative metabolic activities is a major force shaping evolutionary change, but one that is poorly understood. Metabolic adaptability is particularly crucial for a plant's capacity to synthesize specialized chemicals affording protection against microbial

---



**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P.N. (noel@salk.edu).

**Author Contributions**

P.E.O. designed the study, conducted experiments, analyzed data, and wrote the manuscript; A.M. conducted experiments, developed small-scale protein purification; N.D. conducted expertiments, analyzed data, and contributed revisions to the manuscript; B.A.H conducted QM calculations and contributed revisions to the manuscript; L.S. conducted QM calculations; I.S. conducted experiments; B.T.G. and J.C. designed the study and contributed revisions to the manuscript; G.M. analyzed data, developed the biosynthetic tree and chemical distance analysis, and contributed revisions to the manuscript; J.P.N. designed the study, analyzed the data, and wrote the manuscript.

pathogens1–3, elicitation of symbiotic relationships4, attractive5 and repellent6 activities and physiological responses to local environments (as reviewed7). Understanding the evolution of metabolic function at the molecular level requires knowledge of the distribution of enzymatic properties through accessible mutational changes, and hence defining the catalytic landscape. Currently, there is no reported measurement of the catalytic landscapes underlying secondary (specialized) metabolism. The physicochemical constraints relating sequence variation and metabolic output raise important fundamental questions concerning catalytic complexity and natural selection. For instance, how does a particular biosynthetic property like catalytic efficiency or substrate/product specificity vary amongst extant and probable ancestral sequences? Moreover, how likely is the emergence of new product specificities in a family of diverging biosynthetic enzymes?

To experimentally approach these questions in the current work, we relied upon i.) a model system composed of a pair of closely-related secondary metabolic enzymes, ii.) a simplified set of naturally occurring mutations which interconvert a defined catalytic property that is functionally divergent between the parental sequences, iii.) structure-based combinatorial protein engineering (SCOPE)8 to provide a means of creating an enzyme lineage representing putative evolutionary intermediates in one parental background, and iv.) a gas chromatography-mass spectrometry (GC-MS) method9 for measuring the catalytic properties (recording the chemical readout) of the enzyme library. Therefore, quantitative comparison of catalytic properties of enzymes across these simplified and probable lineages provides a direct measure of functional variation likely correlated with phenotypic variation in the defense response of parental species. Moreover, this comprehensive study explores the mutational accessibility of alternative biosynthetic properties over this experimentally defined region of sequence space.

Terpene synthases of secondary metabolism are a diverse enzyme family responsible for the biosynthesis of complex chemicals. Tobacco 5-epiaristolochene synthase (TEAS) and henbane premnaspirodiene synthase (HPS) are closely related (75% amino acid identity) enzymes yet cyclize ionized farnesyl diphosphate (FPP, **1**) to form 5-epi-aristolochene (5-EA, **2**) and premnaspirodiene (PSD, **3**), respectively. These structurally distinct terpene hydrocarbons are precursors of antifungal phytoalexins in solanaceous plants10,11. Mechanistically, TEAS and HPS share a common carbocation intermediate during an electrophilic cyclization reaction, but differ in directing either a methyl or methylene migration, respectively (Fig. 1a). Density functional theory calculations indicate both of these alkyl shifts to be endothermic (~3 kcal/mol), with the methylene shift's transition state of lower energy (Fig. 1b,c and Supplementary Fig. 1 online).

We previously used a structure-guided approach to identify a functionally linked subset of nine amino acid residues out of the 135 naturally occurring amino acid differences between TEAS and HPS (Fig. 2). Mutational swaps of this nine-residue subset are sufficient to interconvert the product specificities of the encoded mutant proteins in the background of each parent enzyme12. Eight of these nine amino acid substitutions are achievable by single non-synonymous nucleotide changes per codon (Fig. 2a). However, position 406 (TEAS numbering) requires a two base change to interconvert between Tyr and Leu in TEAS and HPS, respectively, suggesting the possible intermediacy of Phe at this position in a common

ancestor. Interestingly, only two of the nine amino acid differences are localized on the active site surface, while the remainder are scattered throughout the second tier (Fig. 2d). Replacing these two active site residues of TEAS with their HPS counterparts redirects the final deprotonation-neutralization step to produce 4-epi-eremophilene (4-EE, **4**), a terpene not previously identified in nature12. Thus, the resulting 4-epi-eremophilene synthase (EES) represents an intermediate enzyme with hybrid parental activities (Fig. 1a).

To investigate the natural distribution of these activities in the current work, we constructed a phylogenetic tree from available TEAS-like and HPS-like sequences from related solanaceous plants (Fig. 3a and Supplementary Table 1 online). While the product spectra of terpene synthases cannot be readily predicted from traditional phylogenetic analyses13,14, a clear functional division was apparent between the tobacco and pepper synthases as compared to their orthologues in tomato, potato and henbane. This division was also apparent at the level of our nine-residue subset with the exception of the *Capsicum annuum* synthase (Fig. 3b). This TEAS-like enzyme differs from both TEAS and HPS-like groups at 3 positions, and most notably, contains a threonine at position 438 like HPS, suggesting the first mutational steps in the TEAS-HPS divergence likely occurred at these positions. Evaluating the functional divergence of TEAS and HPS within the context of these nine amino acid substitutions provides a simplified experimental system to ask the broader question: how prevalent and hence evolvable are these parental and alternative biosynthetic activities throughout the intervening lineages connecting these extant enzymes? Measuring the distribution of biosynthetic activities over this sequence space defines a functionally relevant portion of the overall catalytic landscape, and provides a window into the complex functional terrain underlying the evolution of these enzymes. While variation at other positions may have contributed to the functional divergence of TEAS and HPS in meaningful ways, this focused set of functionally important residues makes it experimentally tractable to quantitatively characterize a catalytic landscape of secondary metabolism to biochemical resolution.

## RESULTS

### Creation and characterization of the M9 lineage

Using structure-based combinatorial protein engineering (SCOPE), we constructed a gene library encoding all permutations of nine amino acid substitutions in TEAS ($2^9 = 512$ combinations) previously demonstrated to functionally interconvert TEAS and HPS15. The resulting library, termed the M9 lineage, represent the nodes along all possible mutational pathways (9! = 362,880) between wild type TEAS and the nine mutant HPS-like form (TEAS M9). Therefore, the combinatorial exploration of variation at these diverging positions with SCOPE captures a portion of the functionally relevant genetic variation leading to the current extant sequences. Mutant genes were cloned and identified by sequencing resulting in 432 unique combinations (Supplementary Table 2 online). Individual mutants were expressed and purified from *E. coli*, leading to the recovery of 418 active proteins. We developed an *in vitro* biochemical assay for increased sample throughput and analysis of terpene synthases using gas chromatography–mass spectrometry (GC-MS)9,

which provided quantified chemical fingerprints and catalytic activities of the M9 library proteins (Supplementary Tables 3,4 online).

To quantitatively asses product specificity, the catalytic property defined here as the relative proportion of products, we calculated the product percentages of each mutant from their respective total ion chromatograms. 5-EA (**2**), 4-EE (**4**), and PSD (**3**) were the dominant products observed across a wide spectrum of mutants accompanied by a collection of minor products which were grouped together and treated as a single product class for simplicity. As a caveat, this assumption may introduce error, as the ionization efficiencies of these minor components are as yet unknown, however, their inclusion contributes to the measure of product specificity and promiscuity.

A three-dimensional scatter plot illustrates how the product specificities of mutants distribute throughout chemical space (Fig. 4a). The 3 dominant products 5-EA (**2**), 4-EE (**4**), and PSD (**3**) define a two dimensional triangular plane, whereas the collective minor products contribute the third dimension of the tetrahedron. Mutants with varying degrees of catalytic promiscuity radiate uniformly from a cluster of TEAS-like activities, together forming a continuum with more HPS-like mutants. By contrast, EES-like enzymes are rare, appearing as a sparsely populated subgroup. Subdividing the scatter plot into three smaller tetrahedrons of equal volume geometrically defines product specificity as >50% 5-EA (**2**), 4-EE (**4**) or PSD (**3**), whereas the central volume represents promiscuous activities (Fig. 4b). The majority of mutants are promiscuous (51%) showing expanded product distributions and up-regulation of other TEAS minor products, predominantly germacrene A (**5**)16, a neutral intermediate along the TEAS and HPS cyclization pathways as well as the major product of a closely-related family of plant synthases. Kinetic analyses of select members of the library with diverse product specificities reveals that most mutants possess catalytic activities (Kcat) within 10-fold of wild type TEAS, indicating that most combinations of mutations alter product specificity without significantly compromising the overall catalytic rate (Supplementary Table 4 online).

## Biosynthetic tree of the M9 lineage

To quantitatively examine the distribution of biosynthetic activities across the M9 library, we performed a sum of least squares pairwise comparison of chemical product proportions. The resulting 'chemical' distance matrix was condensed to produce an un-rooted neighbor-joining "biosynthetic tree" (Fig. 5). This tree shows several distinct clusters or clades separating TEAS and HPS-like activities at either end. Annotating each clade with chromatograms from representative mutants highlights the common product profiles that define its members. For example, a promiscuous clade near the tree center is marked by elevated production of germacrene A (**5**) in mutant "(8)". Sequence analysis of the HPS-like clade reveals a clustering of mutants into distinct groups based on sequence, indicating that several convergent solutions exist along a subset of synthetic lineages (Supplementary Fig. 2 online). By comparison, members of the EES-like clade generally possess diverse sequences but exhibit a strict dependence on the T402S active site mutation for EES activity.

### Chemical distances of mutational effects

Values from the 'chemical' distance matrix are a measure of changes in product spectra between mutants and hence provide a quantitative basis to assess the influence of each mutation on product outcome. Is one of the nine positions most critical for controlling product specificity? Consider the effect of mutating a particular position in the background of all other possible combinations of mutations. We calculated the average chemical distance of each of the nine positions in this manner and found them to be comparable, each having large standard deviations of ~50% or greater than each residue's average distance (Supplementary Table 5 online). This result indicates that no single position is more important than another, implicating a context-dependence for a position's influence on the control of product specificity.

### Quantifying mutational context

Given the importance of context, how accessible are alternative product specificities in various mutant backgrounds? To quantitatively examine context effects, chemical distances were tabulated for a subset of 236 mutants for which there are complete data for all single mutational neighbors (permutations that differ at only one position). For example, all TEAS single mutants have been characterized, and represent the mutational neighbors that are a single, coding mutational step away from TEAS wild type. However some permutations of the 512 combinations were not identified by sequencing or failed to produce recombinant protein, so are absent from the final data set and hence omitted from this analysis.

The average inter-neighbor distance (AID) was calculated for each mutant and specific examples illustrate how this index relates to chemical space (scatter plot) and sequence space (alignment with mutational neighbors) (Fig. 6a–c). For a mutant with a low AID, most mutations registered negligible to modest effects on the product output as evident from the clustering of most mutational neighbors in a small region of chemical space (Fig. 6a). By contrast, mutants with a high AID show a broad scattering of mutational neighbors throughout chemical space (Fig. 6c) with demonstrable long jumps between highly specific TEAS-like to EES- or HPS-like activities by single mutational steps. Hence, the activities of some mutants are highly sensitive to mutational perturbations. For a promiscuous mutant with a moderate AID (Fig. 6b), nearly half of the mutational steps tighten product specificity. Considering the larger trends throughout the M9 library, the AID for the subset of 236 mutants was plotted as a simple histogram (Fig. 6d). Plotting the AID as a function of the number of mutations reveals that the distribution of averages is similar across the library (Fig. 6e).

## DISCUSSION

The emerging picture from these experiments is a rugged landscape whereby alternative catalytic specificities are mutationally accessible, requiring as little as a single base change in the coding gene. Single mutations on average exert moderate effects, relaxing product specificity while up-regulating 5-EA (**2**), 4-EE (**4**), PSD (**3**) or other TEAS minor products consistent with postulates that specific activities arise from catalytically promiscuous ancestors11–19. While most mutations are additive, rapid or punctuated changes in product

output are not rare. In fact, such hot spots (average inter-neighbor chemical distance <50) account for 7% of the mutants analyzed thus far, indicating considerable biosynthetic potential for rapid evolutionary jumps throughout the M9 lineage. This rapid adaptability may be unique to terpene cyclases, stemming in part from the subtle energetic differences between competing cyclization pathways (Fig. 1b and Supplementary Fig. 1 online) that ultimately govern product specificity. By implication, TEAS-HPS predecessors had the potential for frequent shifts between PSD (**4**), 4-EE (**3**) and 5-EA (**2**) biosynthesis to generate rapidly changing chemical repertoires throughout evolution.

While both additive and punctuated specificity changes have been observed in terpene cyclases20–22, this is the first effort to quantitatively measure the frequency and distribution of these enzymatic properties over a catalytic landscape. To quantitatively describe this landscape, it was essential to employ a simple Euclidean distance metric, a 'chemical' distance which is generally applicable to mapping how any experimentally defined enzymatic property is distributed throughout sequence space. In the current work, this metric provided a means to quantitatively compare product spectra of terpene synthases, assess the effects of mutations in different backgrounds, particularly mutational neighbors, and to construct a biosynthetic tree to quantitatively organize the M9 enzyme lineage by functional relatedness.

Structural and phylogenetic information has been invaluable in guiding mutagenesis experiments leading to the successful inter-conversion of terpene cyclase substrate24 and product specificities12,23,24. In the absence of phylogenetically derived models, applying saturation mutagenesis to the active site of a notably promiscuous terpene cyclase, γ-humulene synthase, has made significant engineering advances to generate specific activities21. By contrast, the work here explores phylogenetically-relevant biosynthetic interrelationships which extend product specificity control beyond the active site. Characterizing the reciprocal M9 lineages in HPS will be of great future interest to address the contribution of alternative protein backgrounds to the product specificity landscape.

Only recently have efforts been made to characterize the underlying adaptive landscapes of molecular evolution25–28 or to trace the evolutionary origins of the four fundamental isoprenoid-based coupling reactions29. Our study provides the first experimental measure of the complex functional terrain evident in secondary metabolism from the construction and biochemical characterization of intervening lineages between a pair of extant and diverging enzymes. While it is tempting to speculate that 4-EE (**4**) was the dominant product of a TEAS-HPS common ancestor on the basis of its hybrid mechanistic origins, this activity represents less than 3% of the total library. Also, greater product specificity for PSD (**3**) is achievable with fewer than 9 mutations. For example, an M8 (mutant 226, Supplementary Tables 3,7 online) produces 81% PSD (**3**) versus 72% by M9. However, the native henbane enzyme produces 97% PSD (**3**), indicating that additional mutations beyond the 9 considered here contributes to this high degree of product specificity. Importantly, considering the facile phenotypic change from minimal sequence changes uncovered by our work suggests that it may be extremely difficult to make accurate assignments of ancestral function in this pervasive class of secondary metabolic enzymes. This result has broader implications for reconstructing ancestral proteins and ascribing ancient functions; one must

consider the extent of phenotypic variation among a population of putative intermediates encompassed by a "probabilistic guess" of the most likely ancestor.

Connecting catalytic landscapes of secondary metabolism to fitness landscapes of organisms remains an enormous future challenge. While antibiotic resistance or primary metabolic functions in microbes have a direct survival consequences easily measured in laboratory evolution experiments, the fitness contributions of secondary metabolism, which are of particular relevance to speciation in complex organisms, are intrinsically more difficult to study. This arises in part from the myriad roles of secondary metabolites in the greater chemical ecology of host organisms. For the current work, relating the *in vitro* results to *in vivo* functions involves several caveats; numerous factors like transcription, translation, solubility, etc surely contribute to enzyme evolution, while possible effects of the cellular environment on enzymatic activity must also be considered. There is precedence, however, for the correspondence of *in vitro* and *in vivo* product profiles of terpene cyclases23, so the observed plasticity of terpene cyclase enzymatic function from *in vitro* biochemical measurements likely approximates the activities of these enzymes in their native environment. More extensive sequencing efforts and biochemical annotation of terpene synthases from the larger family of solanaceous plants will both address the degree to which mutant combinations of the M9 lineage reflect the actual evolutionary lineages, and provide valuable insights towards understanding the role of secondary metabolism in shaping the evolution of the *Solanaceae*.

The observation that HPS-like activity is achievable by many combinations of mutations lying outside of the active site (Supplementary Tables 6,7 online) highlights the critical importance of epistasis. This phenomenon has been documented and described in other enzyme systems25–27, and is manifested here as a highly interdependent network of interactions in the protein that ultimately control product specificity. These functionally critical yet distal mutations are not surprising given precedence for the effects of distant mutations on protein and enzyme function30,31. Modulation of isoprenoid cyclization by discrete ensembles of peripherally distributed residues is suggestive of energetic networks spread throughout the protein structure32, which may allow greater adaptive capabilities. As recently noted, the interface of enzyme adaptability with intrinsic and induced substrate reactivity underlies the emergence of cyclic diversity in secondary metabolism33. Our quantitative exploration of the catalytic landscape of the M9 lineage provides a first glimpse into the functional effects of quantum evolutionary change on specialized biosynthesis.

## METHODS

### Library construction

SCOPE mutant gene synthesis followed published methods15. A library encoding 512-mutants comprising all permutations of the original TEAS M9 mutants12 was produced with 432 unique combinations being identified by DNA sequencing (Supplementary Table 2 online). The library was created as gene sets consisting of a series of discrete mixtures to reduce over-sampling. Mutant TEAS genes were sub-cloned into pH9GW, an in-house expression vector encoding nine N-terminal histidines, via the Gateway system (Invitrogen) according to manufacturer's instructions.

### Biosynthetic tree construction

Terpene products from GCMS analyses were quantified by integration of product peak areas and transformed into percentages, where 5-EA (**2**), 4-EE (**4**), PSD (**3**) and all remaining products represent four groups adding up to 100% (Supplementary Table 3 online). A distance between every pair of mutants was calculated by the sum of least squares:

$$d=\mathrm{sqrt}\left[(w_1-w_2)^2+(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2\right]$$

where product profile 1 is point $w_1$, $x_1$, $y_1$, $z_1$ and product profile 2 has coordinates $w_2$, $x_2$, $y_2$, $z_2$. The variables w, x, y, and z correspond to 5-EA (**2**), 4-EE (**4**), PSD (**3**) and the sum of remaining products, respectively. A large NxN matrix was dimensionally-reduced into a standard phylogenetic tree, to show which mutants cluster together in space. An un-rooted N-J tree was produced using MEGA 3.1 software.

### Sequencing

Plasmid DNA was prepared by the Microarray Core Facility at The Salk Institute and DNA sequencing performed by Eton Biosciences.

### Vial assay characterization

In vitro assays of purified recombinant enzyme were conducted in duplicate according to a previously published method9. Products were quantified by integration of peak areas from the GC trace using Agilent Chemstation software and expressed as percentage of the total products. Of note, germacrene A (**5**) was detected as the thermally rearranged product β-elemene. Authentic standards of 5-EA (**2**), 4-EE (**4**), PSD (**3**) were used for instrument calibration and absolute product quantitation for kinetic measurements (Supplementary Methods online).

### Other methods

Protein expression, Ni-NTA affinity chromatography purification of library proteins, and kinetic measurements are provided in Supplementary Methods online.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank M. Austin and J. Melnick for critical review of the manuscript, Y. Zhai for computational support, J. Gullberg and A. Nordstom for insightful discussions. This work was supported by a National Institutes of Health Grants GM54029 to J.C. and J.P.N. J.P.N. is an investigator of the Howard Hughes Medical Institute.

## References

1. Grayer RJ, Kokubun T. Phytochemistry. 2001; 56(3):253. [PubMed: 11243452]
2. Pedras MSC, Okanga FI, Zaharia IL, et al. Phytochemistry. 2000; 53(2):161. [PubMed: 10680168]
3. Harborne JB. Biochemical Systematics and Ecology. 1999; 27(4):335.

4. Akiyama K, Matsuzaki K, Hayashi H. Nature. 2005; 435(7043):824. [PubMed: 15944706]
5. Mumm R, Hilker M. Chemical Senses. 2005; 30(4):337. [PubMed: 15788712]
6. Feeny, P. Herbivores: Their Interactions with Secondary Plant Metabolites. 2nd ed. San Diego: Academic Press; 1992.
7. Gershenzon J, Dudareva N. Nat. Chem. Biol. 2007; 3(7):408. [PubMed: 17576428]
8. O'Maille PE, Bakhtina M, Tsai MD. J. Mol. Biol. 2002; 321:677. [PubMed: 12206782]
9. O'Maille PE, Chappell J, Noel JP. Anal. Biochem. 2004; 335(2):210. [PubMed: 15556559]
10. Back KW, He SL, Kim KU, et al. Plant and Cell Physiology. 1998; 39(9):899. [PubMed: 9816674]
11. Facchini PJ, Chappell J. Proc. Natl. Acad. Sci. USA. 1992; 89(22):11088. [PubMed: 1438319]
12. Greenhagen BT, O'Maille PE, Noel JP, et al. Proc. Natl. Acad. Sci. USA. 2006; 103(26):9826. [PubMed: 16785438]
13. Dudareva N, Martin D, Kish CM, et al. Plant Cell. 2003; 15(5):1227. [PubMed: 12724546]
14. Bohlmann J, Meyer-Gauen G, Croteau R. Proc. Natl. Acad. Sci. USA. 1998; 95(8):4126. [PubMed: 9539701]
15. O'Maille PE, Tsai MD, Greenhagen BT, et al. Methods Enzymol. 2004; 388:75. [PubMed: 15289063]
16. O'Maille PE, Chappell J, Noel JP. Arch. Biochem. Biophys. 2006; 448(1–2):73. [PubMed: 16375847]
17. Copley SD. Current Opinion in Chemical Biology. 2003; 7(2):265. [PubMed: 12714060]
18. Jensen RA. Annual Review of Microbiology. 1976; 30:409.
19. O'Brien PJ, Herschlag D. Chemistry & Biology. 1999; 6(4):R91. [PubMed: 10099128]
20. Wilderman PR, Peters RJ. J. Am. Chem. Soc. 2007; 129(51):15736. [PubMed: 18052062]
21. Yoshikuni Y, Ferrin TE, Keasling JD. Nature. 2006; 440(7087):1078. [PubMed: 16495946]
22. Hyatt DC, Croteau R. Archives of biochemistry and biophysics. 2005; 439(2):222. [PubMed: 15978541]
23. Kollner TG, Schnee C, Gershenzon J, et al. Plant Cell. 2004; 16(5):1115. [PubMed: 15075399]
24. Kampranis SC, Ioannidis D, Purvis A, et al. Plant Cell. 2007; 19(6):1994. [PubMed: 17557809]
25. Weinreich DM, Delaney NF, DePristo MA, et al. Science. 2006; 312(5770):111. [PubMed: 16601193]
26. Ortlund EA, Bridgham JT, Redinbo MR, et al. Science. 2007; 317(5844):1544. [PubMed: 17702911]
27. Bershtein S, Segal M, Bekerman R, et al. Nature. 2006; 444(7121):929. [PubMed: 17122770]
28. Miller SP, Lunzer M, Dean AM. Science. 2006; 314(5798):458. [PubMed: 17053145]
29. Thulasiram HV, Erickson HK, Poulter CD. Science. 2007; 316(5821):73. [PubMed: 17412950]
30. Agarwal PK, Billeter SR, Rajagopalan PTR, et al. Proc. Natl. Acad. Sci. USA. 2002; 99(5):2794. [PubMed: 11867722]
31. Rajagopalan PTR, Lutz S, Benkovic SJ. Biochemistry. 2002; 41(42):12618. [PubMed: 12379104]
32. Lockless SW, Ranganathan R. Science. 1999; 286(5438):295. [PubMed: 10514373]
33. Austin MB, O'Maille PE, Noel JP. Nat. Chem. Biol. 2008; 4(4):217. [PubMed: 18347585]

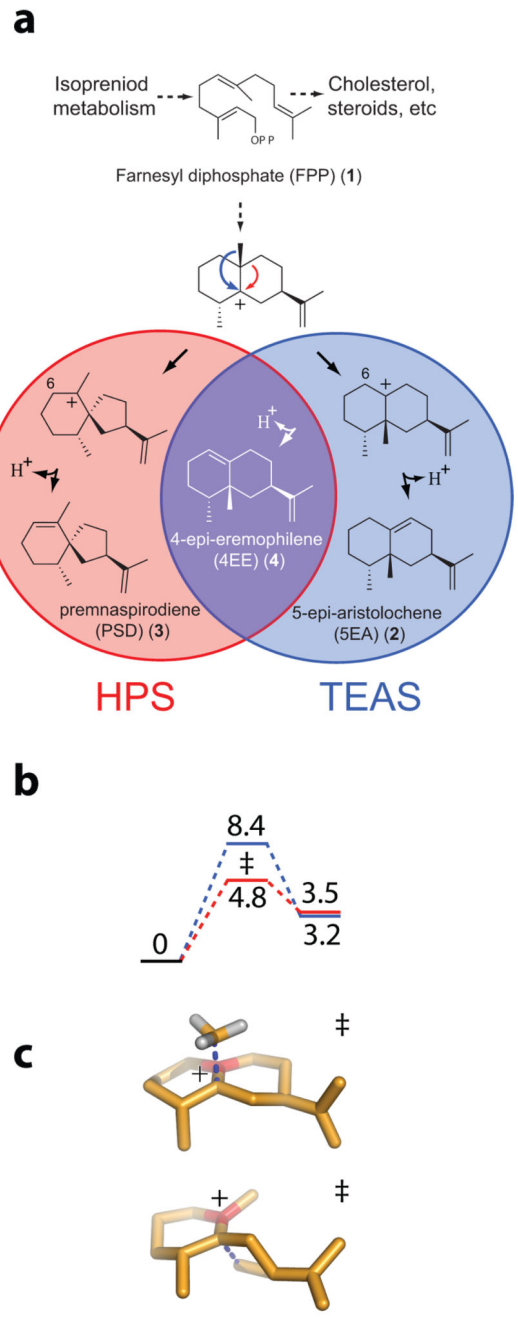


**Figure 1. Terminal cyclization steps of TEAS and HPS terpene synthases**

(**a**) TEAS and HPS exert differential conformational control on a common carbocation intermediate to produce 5-epi-aristolochene (5-EA, **2**) and premnaspirodiene (PSD, **3**), in the blue and red circles, respectively. The discovery of 4-epi-eremophilene (4-EE, **4**) biosynthetic activity supports hybridization of the final two biosynthetic steps (the purple intersection) in TEAS and HPS involving a methyl migration shared with TEAS and a final deprotonation at carbon 6 shared with HPS. (**b**) Proposed reaction coordinate for the methyl (blue trace) and alkyl (red trace) migrations extending from a common carbocation

intermediate (defined as zero energy) through a transition state (‡) leading to the penultimate carbocations of their respective reaction pathways. Calculated energies are expressed in units of Kcal $mol^{-1}$. (**c**) Conformations of the methyl (top) or alkyl (bottom) migration transition states as calculated from density functional theory (DFT) calculations (Supplementary Fig. 1 online). Carbon atoms are shown in gold with the carbocation center colored red (marked by a plus sign); dashed blue lines indicate newly forming bonds, and hydrogen atoms are omitted for clarity.

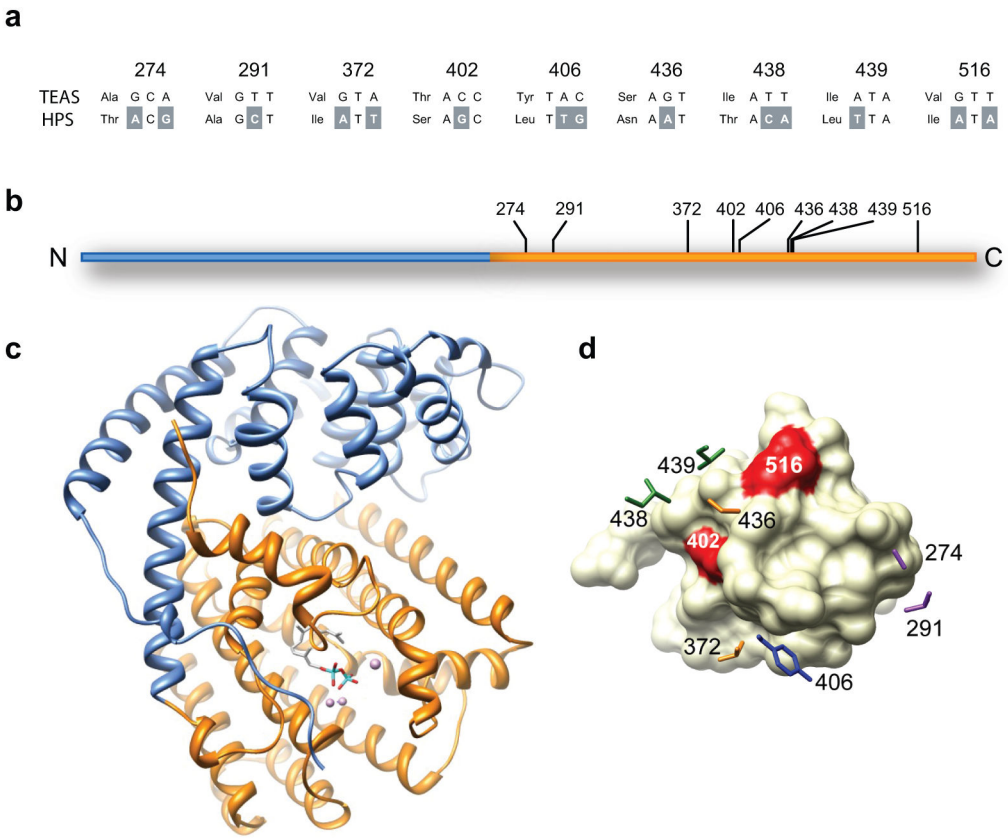


**Figure 2. Overall structure of TEAS, location and identity of M9 residues**

(**a**) Nucleotide and amino acid identity of substitutions between TEAS and HPS. Shading is used to indicate nucleotide substitutions in HPS relative to the TEAS reference sequence. (**b**) The primary structure is composed of N-terminal (blue) and C-terminal (gold) terpenoid synthase domains. Amino acid positions of the M9 library are indicated using TEAS numbering. (**c**) Tertiary structure of TEAS (pdb id 5eat) shown as ribbons with domains colored (as in **a**) and Mg2+ and FPP (**1**) modeled into the active site. (**d**) The spatial distribution of M9 library residues is depicted, where the active site is rendered as a continuous van der Waal's surface (positions 402 and 516 highlighted in red) and second tier residues (colored side chains) are arrayed behind the active site proper.

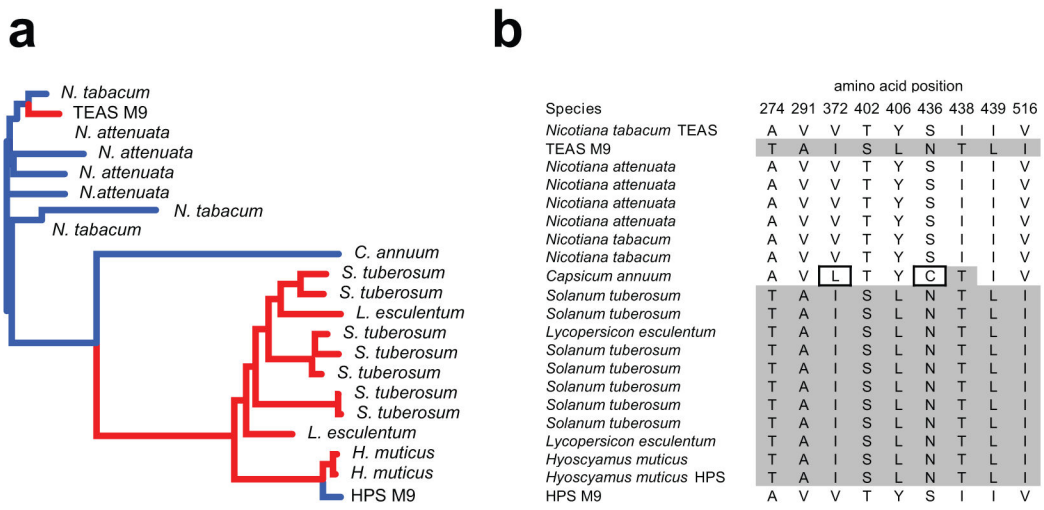


**Figure 3. Phylogenetic distribution of *Solanaceous* TEAS and HPS-like terpene synthases**

(**a**) An un-rooted phylogenetic tree of 5-EA and PSD terpene synthases was created form available sequences (Supplementary Table 1 online) where branches are colored according to the established or putative functions as TEAS-like (blue) or HPS-like (red). (**b**) Sequence alignment of the M9 residue positions of the sequences (panel **a**) with HPS like residues shaded in grey.

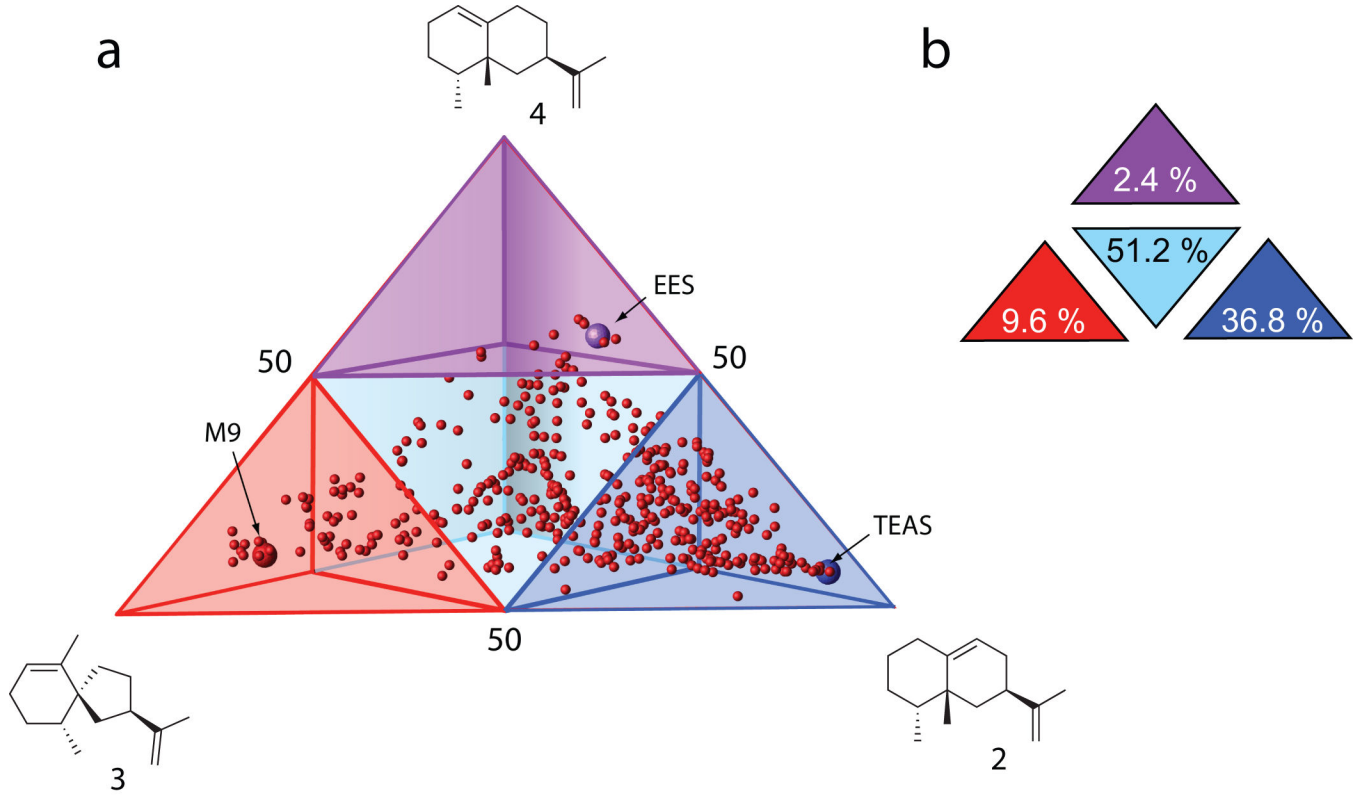


**Figure 4. Activities of the M9 lineage**

(**a**) A 3-D scatter plot of the product output (chemical space) was constructed where the x, y, and z-axes correspond to percentages of the major products 5-EA (**2**), 4-EE (**4**), and PSD (**3**), respectively (Supplementary Table 3 online). Each sphere represents one of the 418 active mutant proteins from the M9 library with wild type TEAS, M9 and EES highlighted as enlarged spheres. The tetrahedron encompassing the scatter plot was partitioned to represent each of the major reaction products by choosing the midpoint of each axis for subdividing into geometrically equivalent tetrahedrons. Each shaded volume, blue (5-EA, **2**), purple (4-EE, **4**), or red (PSD, **3**) indicates product specificity of 50% or greater. Mutants in the remaining central volume (cyan) are defined as promiscuous. (**b**) Schematic of the scatter plot (panel **a**) summarizing the distribution of activities where the number of mutants in each quadrant is expressed as a percentage of the total number characterized.

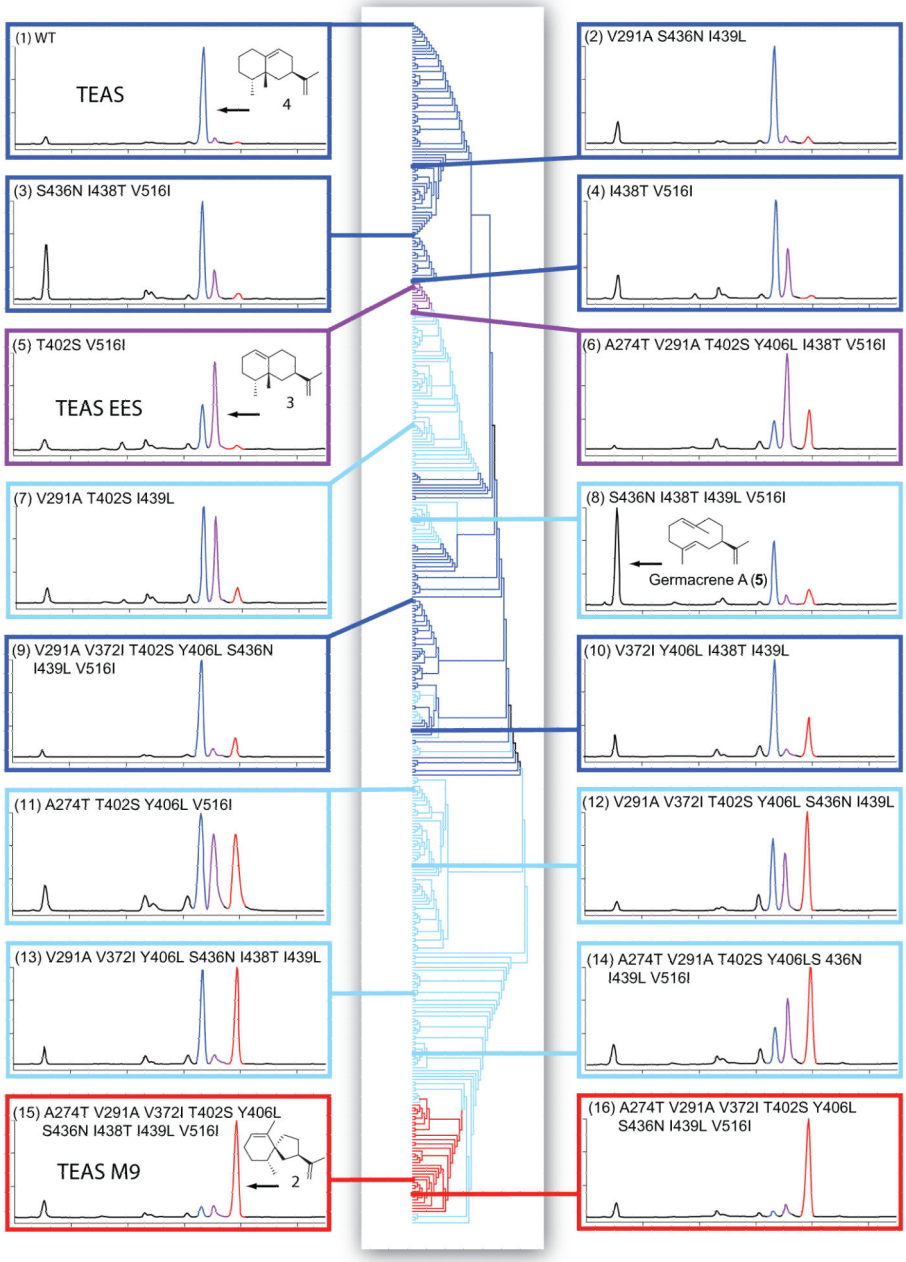


**Figure 5. Biosynthetic tree of the M9 library**

A similarity-based cluster diagram was constructed to quantitatively organize the M9 library according to terpene product spectra from the pair-wise alignment of product proportions for each of the 418 active mutants (described in Methods). Clades are colored according to the major reaction product (defined in Fig. 4a), with representative chromatograms identified and numbered branching off each major clade. Product peaks in the chromatograms are colored blue (5-EA, **2**), purple (4-EE, **4**) or red (PSD, **3**).

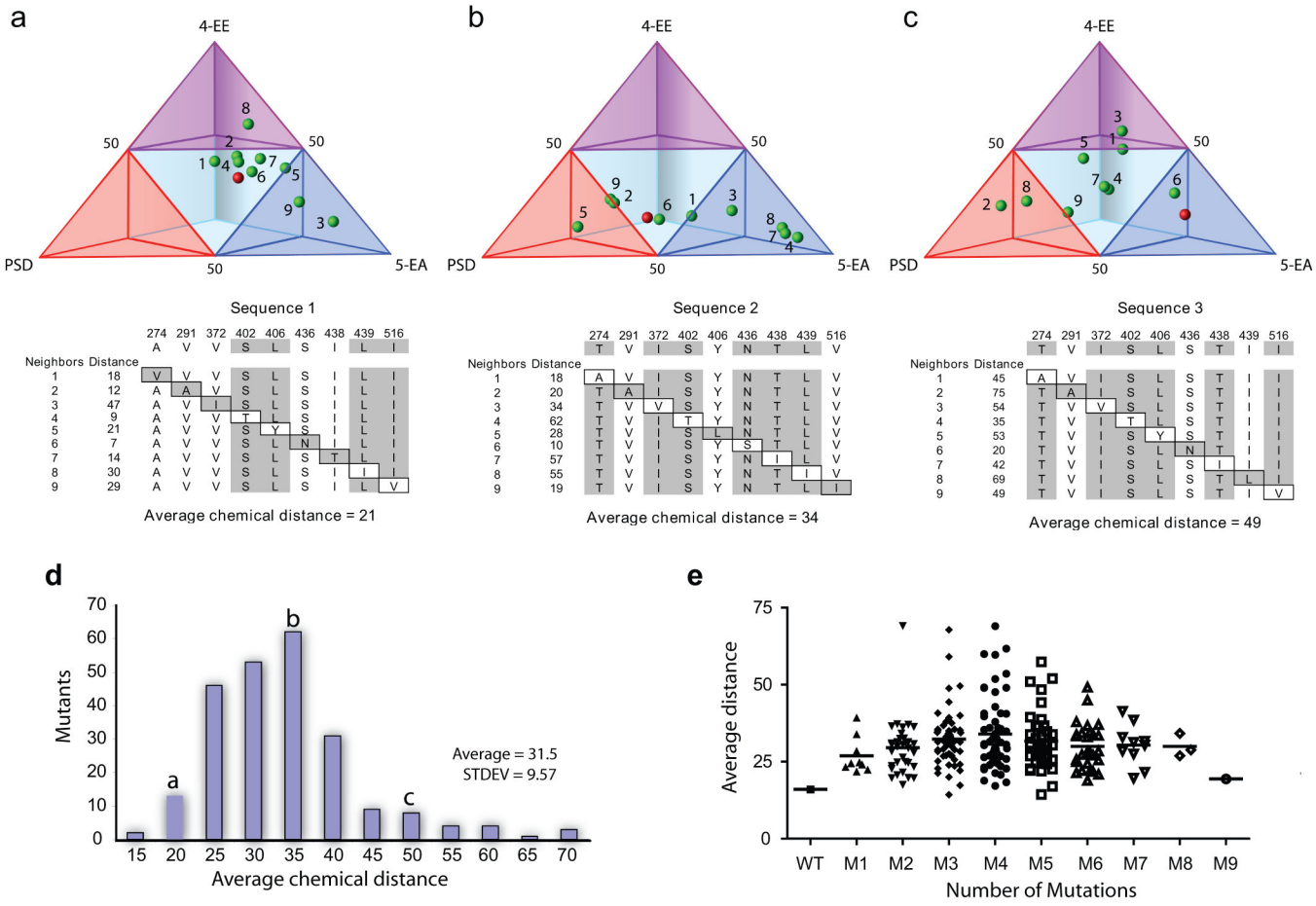


**Figure 6. Average inter-neighbor distances (AID) in chemical and sequence space**

A representative mutant (unlabeled red sphere) is shown in chemical space along with all nine possible single mutant neighbors (numbered green spheres) to illustrate **a**, short, **b**, medium, and **c**, high average inter-neighbor mutational distances (AID). Sequences of each representative mutant are referenced across the top of the three alignment tables with mutational neighbors and distances listed below. Each mutated position is boxed and residues of HPS origin are indicated with shading. (**d**) The average inter-neighbor distance (AID) for a subset of 236 mutants was plotted as a simple histogram, where the shoulders and apex of the distribution are labeled a, b and c to correspond to representative mutants above. (**e**) The distribution of AID as a function of the number of accumulated HPS substitutions was plotted, where M1 refers to all single mutants, M2 to all double mutants and so on.

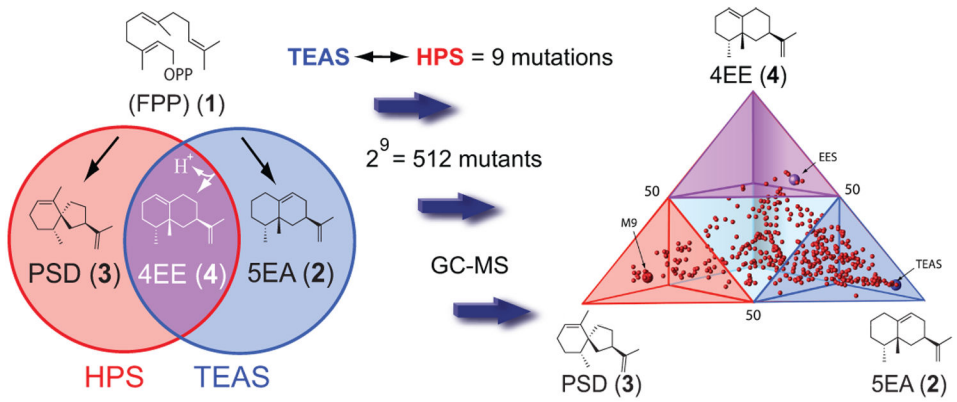


**Figure 7.**