Molecular Diversity (2026) 30:789–807
https://doi.org/10.1007/s11030-025-11215-5

ORIGINAL ARTICLE

sCentInDB: a database of essential oil chemical profiles of Indian medicinal plants

Shanmuga Priya Baskaran[1,2] · Geetha Ranganathan[1] · Ajaya Kumar Sahoo[1,2] · Kishan Kumar[1] · Jayalakshmi Amaresan[1] · Kundhanathan Ramesh[1] · R. P. Vivek-Ananth[1,2,3] · Areejit Samal[1,2]

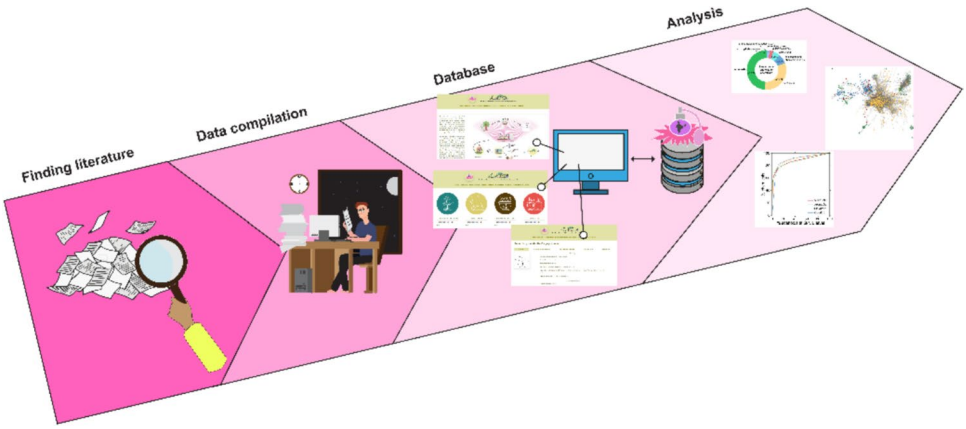
Received: 11 March 2025 / Accepted: 3 May 2025 / Published online: 9 May 2025

Abstract

Essential oils are complex mixtures of volatile compounds produced by aromatic plants and widely used in personal care, food flavoring, and pharmaceutical industry due to their odor and therapeutic properties. As a high-value and low-volume organic product, optimizing plant yield and modifying composition by leveraging knowledge on chemical profiles of essential oils can lead to enhanced bioproducts. Additionally, overharvesting of wild medicinal plants, especially in India, threatens biodiversity. Essential oil profiles of such plants can help regulate their exploitation. Here, we present sCentInDB, a manually curated FAIR-compliant DataBase of Essential oil Chemical profiles of Medicinal plants of India, compiled from published literature. sCentInDB contains data on 554 Indian medicinal plants at the plant part level, encompassing 2170 essential oil profiles, 3420 chemicals, 471 plant-part-therapeutic use associations, 120 plant-part-odor associations, and 218 plant-part-color associations. sCentInDB also compiles metadata such as sample location, isolation, and analysis methods. Subsequently, an extensive analysis of the chemical space in sCentInDB was performed. By constructing a chemical similarity network, terpenoids were found to be distributed across the network, indicating greater structural diversity. Moreover, a comparison of the scaffold diversity of chemicals in sCentInDB was performed against three other aroma libraries using cyclic system retrieval curves. Altogether, sCentInDB will serve as a valuable resource for researchers working on plant volatiles and employing genetic engineering to enhance oil yield and composition. Further, sCentInDB will aid in the establishment of quality standards for essential oils and provide vital insights for therapeutic and perfumery applications. sCentInDB is accessible at https://cb.imsc.res.in/scentindb/.

Graphical abstract

Graphical abstract summarizing the workflow for constructing the sCentInDB database and analysis of the associated chemical space.



Keywords Essential oil · Database · Chemical composition · Aromatic plants · Odor · Therapeutic use

Extended author information available on the last page of the article

Introduction

Essential oils are complex mixtures of volatile compounds produced by plants, and are characterized by strong odor and therapeutic properties [1–3]. These volatile substances are secreted and accumulated through specialized anatomical structures [2]. The term "essential" reflects the fact that it encapsulates the fragrance essence of the plant [4]. Essential oils are commonly obtained from various plant parts via distillation or a mechanical process [5]. These oils play a crucial role as internal messengers, in plant defense, plant-plant interactions, and attracting pollinators [2, 4, 6–8]. Notably, essential oils have been reported to exhibit antimicrobial, anti-inflammatory, and anticancer properties, and therapeutic effects against cardiovascular diseases and other health conditions [9–15]. Due to their distinct odor and therapeutic properties, essential oils have found applications in perfumery, personal care, pharmaceuticals, and food and cosmetics industries [16–20]. In particular, these oils have been traditionally used in aromatherapy to promote health and holistic well-being [1, 21]. India, with its rich biodiversity and numerous endemic plant species, has a long-standing tradition of utilizing medicinal plants in various traditional systems of medicine [22, 23]. While significant research has been conducted to characterize essential oils from Indian medicinal plants [24, 25], much of this knowledge remains scattered across the literature, underscoring the need for its systematic compilation and utilization.

Typically, essential oils consist of chemical components with low molecular weight and volatility at ambient temperature [1, 26, 27]. These oils contain specific classes of chemicals, including terpenoids, shikimates, polyketides and lipids, produced through major pathways of plant secondary metabolism [2, 28, 29]. Variations in these chemical components are influenced by factors such as plant developmental stage, temperature, water availability, and climatic conditions, which contribute to the unique composition, yield, and aroma of the oil [5, 21, 30–33]. The chemical constituents of the oil are generally categorized into major and minor components, with the minor components typically comprising less than 5% of the total composition [34]. Though the major chemical components of the oil primarily determine its biological properties, minor components can also play a crucial role by exerting antagonistic, additive, or synergistic effects, thereby influencing the oil's overall bioactivity [14, 35]. Notably, these minor components can also greatly contribute to the odor of the oil [36].

Chemical profiling of an essential oil is crucial for ensuring the quality standards [37], identifying the odorous chemicals responsible for the oil fragrance [38, 39], optimizing the yield and composition of the oil through biotechnology application [40], and determining the suitability of the oil for medicinal use or as food flavoring [2, 41, 42]. Essential oils are high-value, low-volume bioproducts due to low plant yields, and hence, changing the yield and composition of essential oils is important [43–45]. Additionally, modifying their composition helps reduce antinutritional factors [46] or undesirable components, thereby improving the overall quality and usability of the oils. While chemical synthesis and microbial cultures can produce essential oil constituents, they are limited by high costs [47], structural complexity of phytochemicals [48], and gaps in understanding cellular processes [49], making biotechnology-based methods a promising alternative. To this end, tissue culture techniques help to genetically improve and obtain fully grown, mature plants with desirable traits [40, 50, 51]. Additionally, modified plant cell cultures are often utilized to produce secondary metabolites of interest, such as terpenoids [51]. These approaches require detailed knowledge of the chemical composition and biological activities of the oil.

To date, a few databases have been specifically developed to catalogue the chemical profiles of essential oils, aside from broader databases that compile information on natural products and plants [52–60]. However, these essential oil databases are not region-specific or do not focus exclusively on medicinal plants, leading to limited coverage of medicinal plant species. Among the notable databases, EssOilDB [4] and AromaDb [61] provide valuable information on essential oils. EssOilDB [4] was the first publicly available resource containing information on essential oils from plants worldwide, curated from published research articles. AromaDb [61] specifically captures the aroma molecules in essential oils of medicinal and aromatic plants. Apart from these freely accessible databases, a few proprietary resources such as EOUdb (https://essentialoils.org/plans) and AromaWeb (https://www.aromaweb.com/) have been developed to provide information on essential oils and aromatherapy. However, no dedicated effort has been made to develop an essential oil database for Indian medicinal plants, in particular, one that focuses on oil data derived from plant samples collected within India.

In this contribution, we compiled from published literature the chemical components of essential oils from Indian medicinal plants, with a particular focus on plants whose parts were also sourced within India. Additionally, we gathered data on their therapeutic uses, odor, and color from published literature. To assist chemists in utilizing this data, we annotated the chemical compounds with relevant information such as physicochemical properties, drug-likeness characteristics, ADME properties, and molecular descriptors. The compiled resource represents the most comprehensive database to date in terms of plant coverage and detailed

chemical information on essential oil-bearing medicinal plants from India, providing a valuable knowledgebase for scientific research. The compiled database is accessible via the associated website: https://cb.imsc.res.in/scentindb/.

Methods

Data collection

To compile the chemical profiles of essential oils of Indian medicinal plants, we manually curated information from 778 published research articles. To search the published literature for essential oil profiles, the list of Indian medicinal plants was compiled from the IMPPAT database [57, 62], information provided by the Ministry of AYUSH, Government of India, and books on the subject. This resulted in a non-redundant list of around 7500 Indian medicinal plants. Thereafter, published articles providing chemical profiles of essential oils of the Indian medicinal plants were searched from various sources, including PubMed (https://pubmed.ncbi.nlm.nih.gov/) and Google Scholar (https://scholar.google.com/). This search for articles was conducted until March 2022. From the identified literature sources, information on the plant part from which the essential oil was extracted, the location from which the sample was collected, isolation method, analysis method, chemical constituents, retention index, retention time, and percentage of the individual chemicals in an essential oil was collected. During this compilation, we included essential oil data from Indian medicinal plants, only when the location of the plant sample collection was reported to be within India. In a few instances, the chemical composition of the essential oil in terms of individual chemicals summed up to more than 100% (as reported in the published source), and such inconsistent information was omitted in this compilation. Furthermore, information on the therapeutic uses, color, odor, and general uses of an essential oil was compiled from the corresponding published source, along with additional literature sources and books. In sum, through this extensive manual effort, we curated chemical profiles of essential oils from Indian medicinal plants, focusing exclusively on plant samples collected within India (Fig. 1). Figure 1 summarizes the workflow of this study, from literature mining to compilation, curation, and harmonization of the data contained in sCentInDB.

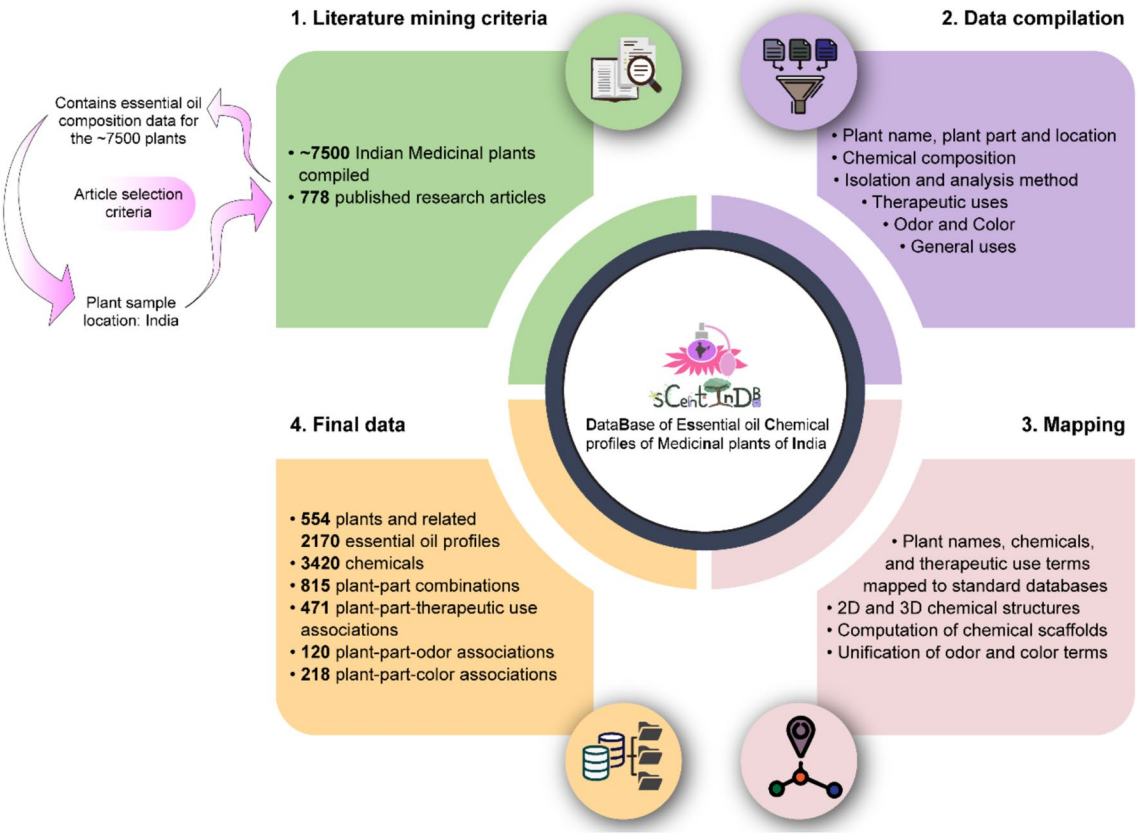


Fig. 1 Workflow summarizing the entire process of sCentInDB database construction involving literature mining, data compilation, mapping, and final dataset

Plant annotation

We mapped the reported names of medicinal plants with essential oil profiles to their corresponding accepted scientific names using three resources, namely, The World Flora Online (WFO) (https://www.worldfloraonline.org/), Plants of the World Online (https://powo.science.kew.org/), and Tropicos (https://tropicos.org/). This resulted in a non-redundant list of 554 medicinal plants with essential oil profiles in this compiled dataset. Furthermore, common names, taxonomic information, and conservation status or extinction risk of these medicinal plants were compiled from Flowers of India (https://www.flowersofindia.net/), WFO (https://www.worldfloraonline.org/), and IUCN Red List (https://www.iucnredlist.org/en), respectively. Moreover, information on the use of medicinal plants in various traditional Indian systems of medicine, such as Ayurveda and Siddha, was compiled from the IMPPAT database (https://cb.imsc.res.in/imppat/). Additionally, for these medicinal plants with reported essential oil profiles, we provided links to other resources namely, WFO (https://www.worldfloraonline.org/), Tropicos (https://tropicos.org/), Medicinal Plant Name Services (https://mpns.science.kew.org/), Plants of the World Online (https://powo.science.kew.org/), International Plant Names Index (https://www.ipni.org/) and Gardeners' World (https://www.gardenersworld.com/).

Annotation of essential oil chemicals

We obtained a non-redundant list of 3420 chemicals across essential oil profiles compiled in sCentInDB. The two-dimensional (2D) structure of the essential oil chemicals was obtained from PubChem [63] (https://pubchem.ncbi.nlm.nih.gov/), ChemIDplus (https://pubchem.ncbi.nlm.nih.gov/source/ChemIDplus), and ChemSpider (https://www.chemspider.com/) databases. For chemicals that could not be mapped to information in these chemical databases, we manually generated their 2D structures using the Chemicalize Marvin JS (https://chemicalize.com) software, based on the structural representations provided in the corresponding published research articles or books. Moreover, the 2D structure information of chemicals was stored in various formats such as SMILES, InChI, InChIKey, SDF, MOL, and MOL2, which were generated using Open Babel [64]. For the chemicals, the images of their 2D structure were generated using RDKit (https://www.rdkit.org/). Further, the three-dimensional (3D) structures of the chemicals were retrieved from PubChem. If the 3D structure of a chemical was not available in PubChem, the 3D structure was generated from its 2D structure by employing RDKit using the ETKDG method and the MMFF94 force field. Moreover, the 3D structures of chemicals were stored in various formats such as SDF, MOL, MOL2, PDB, and PDBQT using Open Babel. Of the 3420 chemicals, we were able to obtain the 3D structure for 3417 chemicals. Moreover, the chemicals were mapped to identifiers in several databases, including PubChem, ChEMBL [65, 66] (https://www.ebi.ac.uk/chembl/), ChEBI (https://www.ebi.ac.uk/chebi/), FDA SRS (https://precision.fda.gov/uniisearch), ChemSpider, Molport (https://www.molport.com/), SureChEMBL (https://www.surechembl.org/), and ZINC [67–69] (https://zinc15.docking.org/). Further, the chemicals were associated with published spectrum information obtained via different experimental techniques, by using compiled information in the SpectraBase (https://spectrabase.com/) database.

ClassyFire [70] was used to obtain a hierarchical classification for the chemicals, namely, kingdom, superclass, class, and subclass. Of the 3420 chemicals in our compilation, ClassyFire could not compute the classification for 49 chemicals. Further, physicochemical properties and drug-likeness properties of chemicals were computed using RDKit (https://www.rdkit.org/). Moreover, absorption, distribution, metabolism, and excretion (ADME) properties of chemicals were predicted using SwissADME [71]. Of the 3420 chemicals in our compilation, SwissADME failed to compute the ADME properties for 13 chemicals. Furthermore, we computed 1875 molecular descriptors (both 2D and 3D descriptors) for each chemical using PaDEL software [72]. The chemicals were subject to a cleaning protocol using MayaChemTools [73] to remove invalid structures, duplicates, and salts, resulting in a final set of 3414 chemicals. For each of these, we computed the natural product likeness score or NP-likeness [74] score (https://github.com/rdkit/rdkit/tree/master/Contrib/NP_Score) using RDKit. Moreover, we computed the molecular scaffolds for the 3414 chemicals at the graph/node/bond (G/N/B) level, graph/node (G/N) level, and the graph level, using RDKit [75, 76]. G/N/B has connectivity, element and bond information, G/N has connectivity and element information, and the graph level has only connectivity information.

Chemical similarity network

The structural similarity between any two essential oil chemicals was ascertained by computing the pairwise Tanimoto coefficient (Tc) [77] using the extended circular fingerprints (ECFP4) [78] in RDKit. Thereafter, a chemical similarity network was generated wherein the nodes are chemicals and two nodes are connected with an edge if they have $Tc \geq 0.5$.

Web interface and database management

We have encapsulated the compiled data in an online database, namely, the *DataBase of Essential oil Chemical profiles of Medicinal plants of India* (sCentInDB), with a

user-friendly web interface to widely disseminate the compiled dataset on essential oil profiles of Indian medicinal plants and associated metadata (Fig. 2). sCentInDB can be accessed at: https://cb.imsc.res.in/scentindb. In the back-end, we employed the open-source relational database management system MariaDB (https://mariadb.org/) to store the compiled data. Moreover, the front-end of the website was created using the open-source CSS framework Bootstrap 5.3.0 (https://getbootstrap.com/docs/5.3/) customized with in-house HTML, PHP (https://www.php.net/), JavaScript, and jQuery (https://jquery.com/) scripts. Furthermore, JSME Molecule Editor [79] has been incorporated into the web interface to enable drawing of chemical structures. sCentInDB is hosted on an Apache (https://httpd.apache.org/) webserver running on Debian 9.4 Linux Operating System.

Results and discussion

sCentInDB: a FAIR compliant database of essential oil chemical profiles of medicinal plants of India

We have built a user-friendly database, namely, sCentInDB, which stands for DataBase of Essential oil Chemical profiles of Medicinal plants of India. The database contains compiled information on essential oils of Indian medicinal plants and is openly accessible for academic research at: https://cb.imsc.res.in/scentindb/. Users can access the compiled information on the website through three options, namely, Browse, Basic search, and Advanced search (Fig. 2a). Supplementary Figure S1 displays the Entity-Relationship Diagram (ERD) for the different entities in sCentInDB.

Browse

The Browse tab allows users to browse essential oil information in four different ways: Indian medicinal plant, chemical, therapeutic use, and odor (Fig. 2b). By choosing an Indian medicinal plant from the drop-down menu via the 'Choose an Indian medicinal plant' option, the user will be redirected to a page with information on the plant part, essential oil identifier, chemicals found in the oil, and the published literature reference. By clicking on the plant name, the user can open a detailed page about the plant. Similarly, by clicking on the oil identifier, the user will be redirected to a detailed page about the oil, including all the associated metadata for that oil (Fig. 2c). Similarly, the chemical name is hyperlinked to a detailed page about the chemical. By choosing a chemical from the drop-down menu via the 'Choose a Chemical' option, the user will be displayed a table containing the chemical identifier, essential oil profiles containing the chemical, and a link that redirects to a detailed page about the chemical. This includes summary information on the chemical, its physicochemical properties, its drug-likeness properties, its ADME properties, and its molecular descriptors (Fig. 2d). By choosing a therapeutic use term from the drop-down menu via the 'Choose a Therapeutic use' option, the user will be displayed a page with information on essential oil at the plant part level with the corresponding therapeutic use, therapeutic use identifiers in external databases, and a link to published literature reference. By choosing an odor term from the drop-down menu via the 'Choose an Odor' option, the user will be displayed a page containing information on essential oil at the plant part level with corresponding odor, related uniformized odor terms, and a link to published literature reference.

Basic search

The basic search tab allows text-based search to retrieve the curated information on essential oil via two options, namely, by chemical and by therapeutic use (Fig. 2e). The 'Search by Chemical' option allows users to enter the complete or partial name of a plant, plant part, geographical location, chemical name, and chemical composition in percentage. This generates a table containing information on essential oils relevant to the user query. The 'Search by Therapeutic use' option enables users to search for therapeutic uses of essential oils at the plant part level with a link to published literature reference.

Advanced search

The advanced search tab enables users to filter and retrieve a subset of chemicals in sCentInDB based on various features, including physicochemical properties, drug-likeness properties, chemical similarity, and molecular scaffolds (Fig. 2f). The 'Physicochemical filter' allows users to filter chemicals based on molecular weight, Log P, topological polar surface area (TPSA), number of hydrogen bond acceptors, number of hydrogen bond donors, number of heavy atoms, number of heteroatoms, number of rings, and number of rotatable bonds. The 'Drug-like filter' helps in retrieving chemicals based on a combination of various drug-likeness scoring schemes. The 'Chemical similarity filter' allows users to draw a chemical's structure using the molecular editor or provide input in SMILES format, and thereafter, retrieve structurally similar chemicals present in sCentInDB (Fig. 2g). The 'Scaffold filter' helps in retrieving chemicals based on molecular scaffold at Graph/Node/Bond (G/N/B), Graph/Node (G/N), or Graph levels.

Fig. 2 Snapshots of the web interface of sCentInDB. **a** Home page of sCentInDB providing a brief description of the database. **b** Browse page of sCentInDB showing four different ways for browsing the compiled information. **c** Snapshot of the detailed page on an essential oil. The displayed page corresponds to the leaf oil of the plant *Curcuma longa*, and contains information on its identifier, plant name, part name, analysis method, chemical composition, general uses, odor, and color. **d** Snapshot of the detailed page corresponding to a chemical with five different tabs for summary, physicochemical properties, drug-likeness properties, ADME properties, and descriptors. **e** Basic search page of sCentInDB displaying two different options to search the database. **f** Advanced search page of sCentInDB which enables filtering and retrieval of chemicals via four different options. **g**) Chemical similarity filter option in the advanced search page, allowing users to retrieve structurally similar chemicals in the database based on a specific user query

Compliance with FAIR principles

While creating the database, we ensured that sCentInDB adheres to FAIR [80] principles, which stands for Findable (F), Accessible (A), Interoperable (I), and Reusable (R). sCentInDB is compliant with the FAIR principles due to the following reasons:

- sCentInDB has a stable and permanent web address (F, A).
- All entries in sCentInDB can be accessed by stable URLs (F).
- Plants, chemicals, and therapeutic use terms are linked to unique external identifiers in standard databases (F, I).
- Data within sCentInDB can be accessed through the Browse, Basic Search, and Advanced Search tabs by providing a user query (A).
- Users do not need to create an account to access sCentInDB (A).
- sCentInDB has a Help page explaining how to use the data (A).
- Data within sCentInDB is available in human readable format (html) and can be downloaded in machine readable format (.csv), and therefore, can be reused for further research (A, R).
- License of the sCentInDB database also allows users to freely use the associated data for academic research (R).
- The data curation workflow is explained in detail in the associated publication (R).

Exploration of Indian medicinal plants with documented essential oil profiles in sCentInDB

We compiled detailed information on essential oils from 554 Indian medicinal plants, including their chemical composition and reported therapeutic uses from published literature (Methods). Additionally, the Indian medicinal plants were annotated with information on their taxonomic classification, their use in traditional systems of medicine, and categorization in the IUCN Red List of Threatened Species. We observed that the 554 plants with essential oil profiles in sCentInDB belong to 86 taxonomic families. Figure 3a shows the taxonomic families with at least 5 plants in the database. Lamiaceae, commonly known as the mint family, has the largest representation (95 plants) in sCentInDB. As one of the largest flowering plant families in the world, Lamiaceae holds significant economic importance in cosmetics, food, and medicine [81, 82]. Further, the Asteraceae family has the second largest representation (76 plants) in sCentInDB (Fig. 3a). Moreover, Angiosperms or flowering plants constitute 96.2%, Gymnosperms account for 3.6%, and Pteridophytes comprise only 0.2% of the plants in sCentInDB (Fig. 3b). Many plants in sCentInDB are documented to be used in traditional systems of medicine, including 212 plants in Ayurveda, 187 plants in Siddha, 145 plants in Unani, 78 plants in Sowa Rigpa, and 60 plants in Homeopathy (Fig. 3c). According to the IUCN Red List of Threatened Species, 6 plants in the sCentInDB are classified as critically endangered, and 13 are listed as endangered (Fig. 3d).

Analysis of the essential oil information compiled in sCentInDB

sCentInDB compiles detailed information on essential oils of Indian medicinal plants, including their chemical composition, therapeutic uses, odor, and color, from published literature. Each essential oil was assigned a unique identifier, corresponding to the specific plant and the plant part from which it was derived. In addition, we assigned sample identifiers to each essential oil profile, taking into account variations in location, sampling conditions, and different published references, to ensure that different samples of the same essential oil across published literature were clearly distinguished. For example, in sCentInDB, the essential oil derived from the rhizome of the plant *Zingiber officinale* is assigned the identifier 'ESO0822'. The profile for this oil is based on 88 samples, which are denoted by sample identifiers 'ESO0822_01' to 'ESO0822_88', and capture the variations in the geographical locations from where the plant samples were sourced, sampling conditions, and literature source. Altogether, sCentInDB catalogs 815 essential oils at the plant part level, which expands to 2170 when considering the multiplicity of samples reported in published literature. Among the different plant parts, leaf is the most used part for essential oil extraction, being utilized in a maximum number (266) of plants of the database (Fig. 4a). Further, among the different Indian medicinal plants with essential oil profiles, the database compiles essential oils for a maximum of six different plant parts for *Mentha arvensis* (Fig. 4b). Moreover, the chemical caryophyllene is documented to be present in maximum number of essential oils (627) compiled in the database (Fig. 4c). Among the 12 extraction methods recorded in sCentInDB, hydrodistillation is the most commonly used, accounting for 672 out of 831 oil-isolation method combinations, followed by steam distillation with 133 instances. Distillation involves exposing plant material to water or steam to release essential oils through evaporation. In hydrodistillation, the plant material is in direct contact with boiling water. In contrast, steam distillation uses steam produced in a separate boiler, which then contacts the plant material in another container. Steam distillation gives more yield and is less susceptible to hydrolysis when compared with hydrodistillation. However, hydrodistillation is much faster than steam distillation [83].

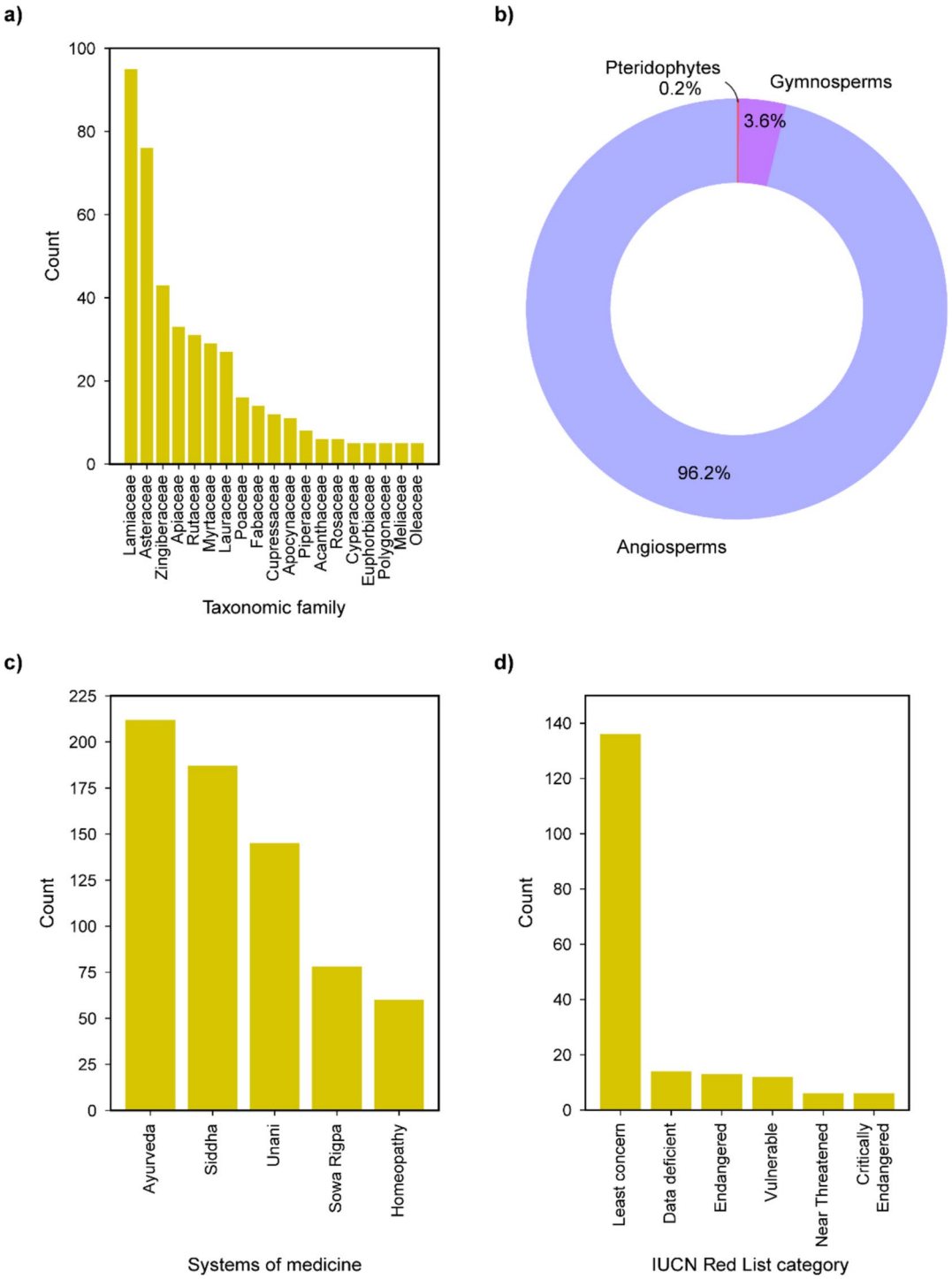


Fig. 3 Basic statistics of medicinal plants in sCentInDB. **a** Bar plot depicting the number of plants across taxonomic families with at least five plant species in the database. **b** Pie chart showing the distribution of plants across groups in the database. **c** Bar plot depicting the distribution of plants across various traditional systems of Indian medicine. **d** Bar plot showing the distribution of plants across different IUCN Red List categories

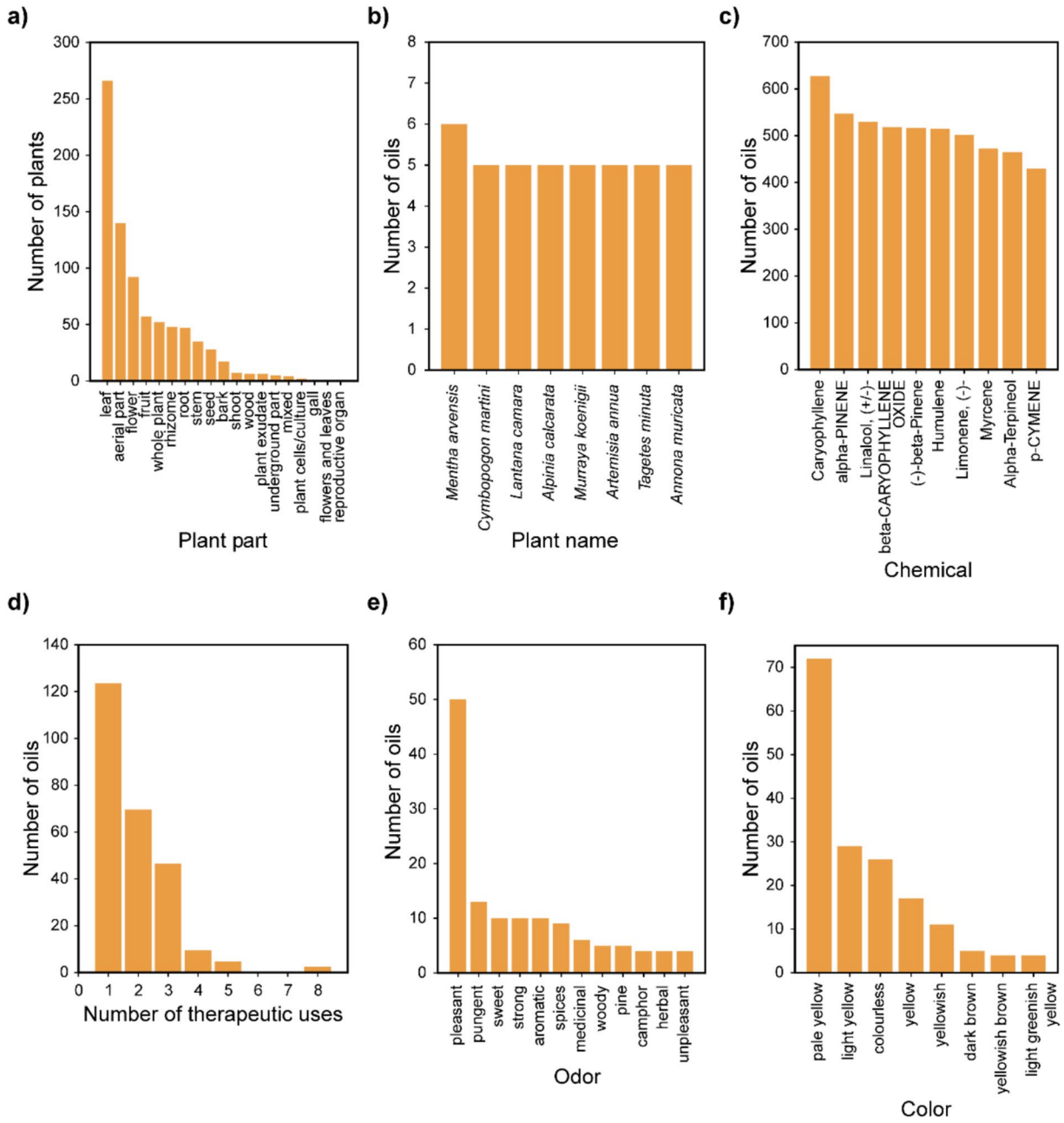


Fig. 4 Basic statistics of essential oil data compiled in sCentInDB. **a** Bar plot showing the distribution of the plants across different plant parts in the database, from which the essential oils are derived. **b** Bar plot depicting the top Indian medicinal plants with documented information based on the number of essential oils at the level of plant parts. **c** Bar plot showing the top most represented chemicals in the essential oil data. **d** Bar plot depicting the distribution of the essential oils in terms of their number of therapeutic uses. **e** Bar plot showing the distribution of the essential oils across the most represented odor terms associated with the essential oils. **f** Bar plot showing the distribution of the essential oils across the most represented color terms associated with the essential oils

Therapeutic uses of essential oils in sCentInDB

The therapeutic use information for the essential oils at the level of plant parts was compiled from published literature. Among the 554 essential oil-bearing plants in sCentInDB, we were able to compile the therapeutic use information for oils corresponding to 209 plants. In terms of essential oils, of the 815 oils in the database, we were able to collect the therapeutic use information for 253 oils. In order to standardize these therapeutic use terms, we mapped them to corresponding terms in Medical Subject Headings (MeSH) (https://www.ncbi.nlm.nih.gov/mesh/), Unified Medical

Language System (UMLS) (https://www.nlm.nih.gov/research/umls/index.html), and International Classification of Diseases 11 th Revision (ICD-11) (https://icd.who.int/browse/2024-01/mms/en). This resulted in a non-redundant list of 41 standardized therapeutic use terms and 471 plant-part-therapeutic use associations in the sCentInDB database. We observed that the majority of the oils in sCentInDB have only one therapeutic use term (Fig. 4d), and anti-bacterial and antifungal activity are the most represented therapeutic use terms across the oils in sCentInDB.

Odor and color of essential oils in sCentInDB

The odor data of the essential oils at the level of plant parts was compiled from published research articles. Among the 815 oils in sCentInDB, we were able to compile odor data for 115 oils. To standardize these odor data, we primarily relied on OlfactionBase [84], refining them to 54 uniformized odor terms. Notably, the odor of essential oils is often described using a combination of odor terms, including the odor families, intensity, etc. It is important to note that given an essential oil, different research papers may have reported its odor differently. This is because odor perception is a subjective psychophysical process [85]. Additionally, variations in odor descriptions may arise due to differences in sampling conditions across the published literature [86]. We observed that the odor term 'pleasant' is associated with maximum number of oils (50) in sCentInDB (Fig. 4e).

Similarly, the color data of the essential oils at the level of plant parts was compiled from published research articles. Among the 815 oils in sCentInDB, we were able to compile the color data for 210 oils. Thereafter, we manually uniformized the color data into 48 distinct color terms. Similar to the case of odor, the color of essential oil can have variations across published literature due to the differences in sampling conditions [87, 88]. We also observed that the color term 'pale yellow' is associated with maximum number of oils (72) in sCentInDB (Fig. 4f).

Altogether, sCentInDB has compiled chemical composition information for 2170 essential oils (including multiple samples for a particular oil) at the plant part level. Additionally, it captures information on therapeutic use, odor, and color information for essential oils from published literature.

Analysis of the chemical space of essential oils documented in sCentInDB

sCentInDB compiles an in silico, non-redundant and stereo-aware library of 3420 unique chemicals found in the essential oils of Indian medicinal plants at the plant part level. To curate this chemical library, we obtained the 2D structures of these chemicals from multiple sources (Methods). For completeness, we also accounted for all chemicals contributing to the essential oil profile, even if their structures could not be retrieved. Furthermore, these chemicals were checked for structural similarity, and thereafter, unique structures were resolved based on their stereochemistry using InChI representation. This led to a non-redundant list of 3420 chemicals with their structural information, across 2170 essential oils (including oil samples) at the level of plant parts, which accounts for 39,438 plant-part-chemical associations. Supplementary Figure S2a shows the distribution of the number of chemicals across Indian medicinal plants documented in sCentInDB. From the figure, it can be observed that there is one plant with 352 chemicals, *Jasminum sambac*, and 3 other plants produce 332 chemicals each. Interestingly, the top 20 plants in our database, in terms of the number of documented phytochemicals found in their essential oils, cover nearly one-third of the total phytochemicals in the database.

The phytochemicals in sCentInDB were given unique chemical identifiers, with each identifier corresponding to a chemical name, respective structural features, and hyperlinks to external standard chemical databases. The 2D and 3D structures of the chemicals are provided in five different file formats (Methods). Further, we computed the molecular scaffold(s), the core structure of a molecule, which has extensive use in the medicinal chemistry field. We have adapted the scaffold definition given by Lipkus et al. [75, 76] and computed the scaffold of the chemicals at three different levels (Methods).

Supplementary Figure S2b–g show the distributions of six physicochemical properties for the 3420 chemicals in sCentInDB. The chemicals in sCentInDB were classified into 18 superclasses, 103 classes, and 158 subclasses using ClassyFire (Fig. 5a). Among these 18 superclasses, the top three superclasses are lipids and lipid-like molecules, organic oxygen compounds, and benzenoids, comprising 1708, 455, and 371 chemicals, respectively, in sCentInDB (Fig. 5a). Furthermore, to obtain a natural product specific classification, we used NPClassifier [89] to classify the chemicals in sCentInDB based on their biosynthetic pathways. The top three biosynthetic pathways were terpenoids, fatty acids, and shikimates and phenylpropanoids corresponding to 1662, 1099, and 264 chemicals, respectively, in sCentInDB (Fig. 5b).

NP-likeness [74] score is a measure to quantify the natural product likeness of a chemical. This score ranges from -5 to $+5$, wherein the higher the score for a chemical, the more likely it is to be a natural product [74]. In general, NP-likeness scores of natural product libraries are predominantly positive [90, 91]. As expected, $>90\%$ of the chemicals in sCentInDB have NP-likeness scores to be positive, as the database is specific for natural products. Figure 5c shows the distribution of NP-likeness scores for chemicals in sCentInDB. It can be seen that this distribution is skewed toward the positive side, and similar to the distribution

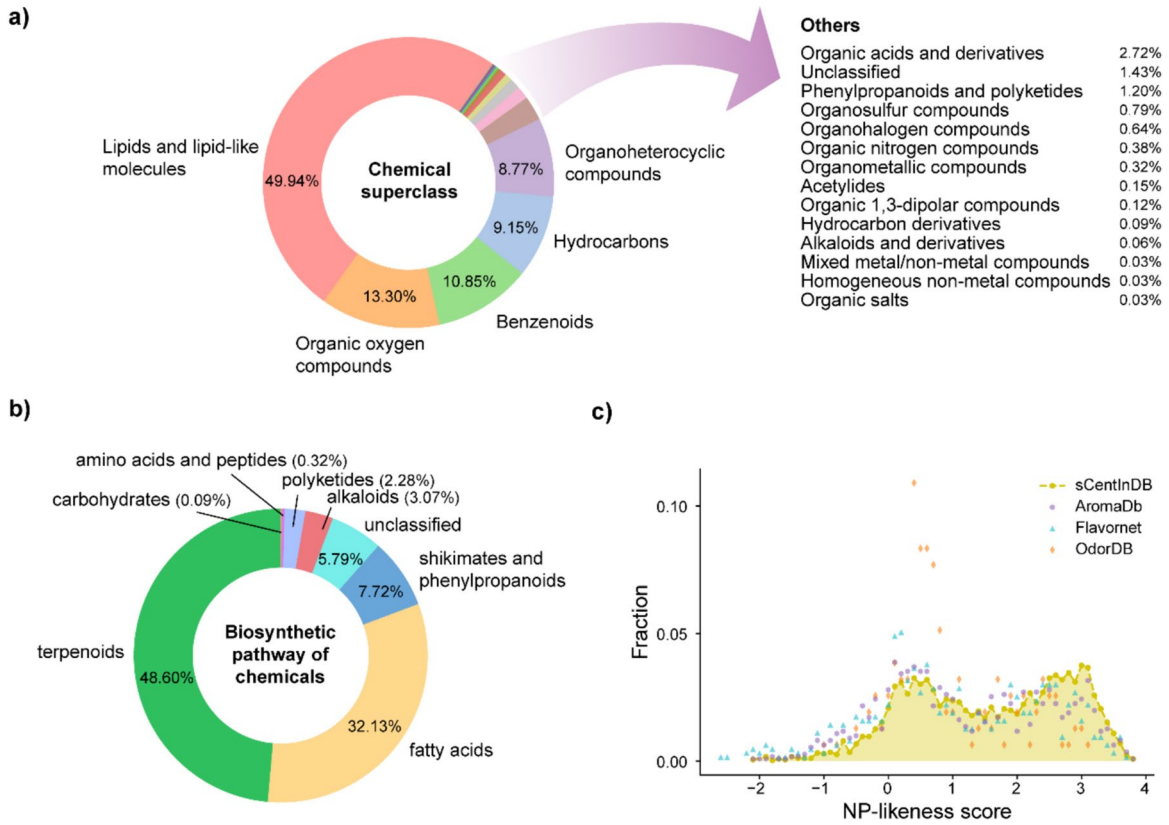


Fig. 5 Chemical superclass, biosynthetic pathways, and natural product-likeness scores for the chemicals in sCentInDB. **a** Distribution of the chemical superclass for chemicals as predicted by ClassyFire. **b** Distribution of the biosynthetic pathways of chemicals as predicted by NPClassifier. **c** Distribution of NP-likeness scores of chemicals in sCentInDB in comparison to other odor or aroma molecule libraries

for other chemical libraries containing aroma molecules (Fig. 5c).

Chemical similarity network

We generated the chemical similarity network (CSN) of the 3420 chemicals in sCentInDB to analyze their structural diversity (Methods; Fig. 6). We observe that the CSN has 183 connected components (with two or more chemicals each) and 618 isolated nodes. Notably, the largest connected component of the CSN consists of 1452 chemicals, with the majority of these chemicals (947 of 1452) produced through the fatty acid biosynthetic pathway. Interestingly, among the 1099 chemicals belonging to the fatty acid biosynthetic pathway in the database, 947 are clustered into a single component in the CSN, while chemicals belonging to the terpenoid biosynthetic pathway are more dispersed across the CSN, indicating greater structural diversity of terpenoids [92, 93]. Additionally, within the largest connected component of the CSN, *Jasminum sambac* is documented to produce the highest number of chemicals (285), the majority of which belong to the fatty acid biosynthetic pathway.

Molecular scaffold based diversity analysis

Molecular scaffold based diversity analysis of a chemical space can help identify small molecules with novel scaffolds of interest [95, 96]. We followed Lipkus et al. [75, 76] to compute Bemis and Murcko scaffolds [97] at three different levels, namely, graph/node/bond (G/N/B) level, graph/node (G/N) level, and graph level (Methods). We find that the chemical space of sCentInDB comprises of 693 scaffolds at the G/N/B level, 479 scaffolds at the G/N level, and 362 scaffolds at the graph level. We compared the scaffold diversity of sCentInDB with three other chemical libraries containing aroma molecules, namely, AromaDb, Flavornet, and OdorDB. At the G/N/B level, sCentInDB is second only to AromaDb in terms of the fraction of scaffolds per molecule (N/M) and the fraction of singleton scaffolds per molecule (N_{sing}/M) (Table 1). sCentInDB and AromaDb have very similar area under the curve (AUC) values, suggesting similar scaffold diversity for both the essential oil libraries. The higher AUC values for sCentInDB and AromaDb in comparison to other natural product libraries [57, 95] can be attributed to the fact that essential oil chemicals are

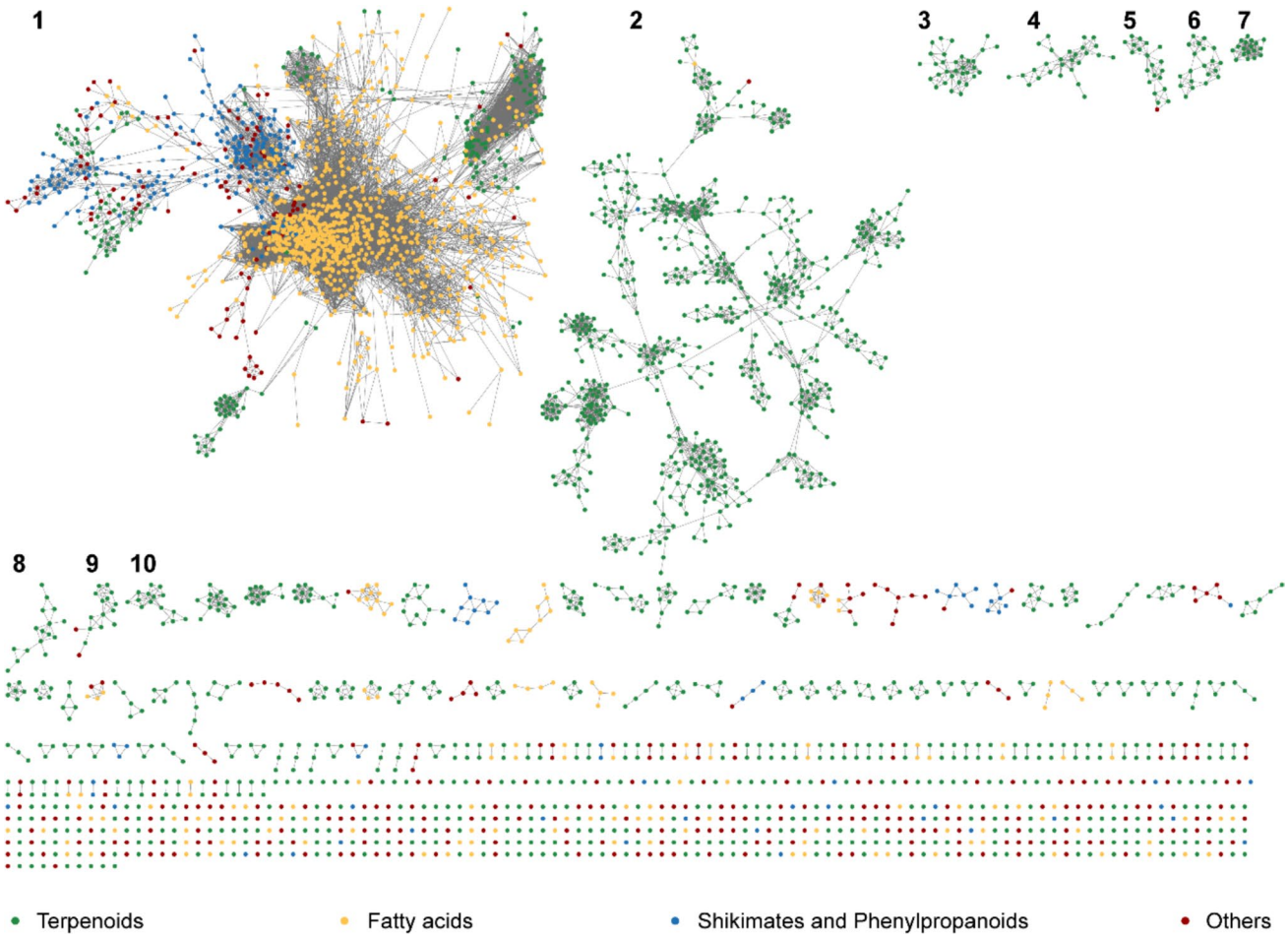


Fig. 6 Chemical similarity network (CSN) comprising of the 3420 chemicals in sCentInDB. This network visualization was generated using Cytoscape [94]. The nodes in the network correspond to the chemicals, and two nodes are connected if the chemical similarity computed using the Tanimoto coefficient (Tc) is ≥ 0.5. The nodes in the CSN are colored based on their biosynthetic pathway predicted by NP classifier: green for terpenoids, yellow for fatty acids, blue for shikimates and phenylpropanoids, and dark red for others. In this CSN visualization, the top 10 connected components are numbered

Table 1 Comparison of the scaffold diversity of sCentInDB with other chemical libraries containing aroma molecules

Chemical library	M	N	N_{sing}	N/M	N_{sing}/M	N_{sing}/N	AUC	P_{50}
sCentInDB	3414	695	471	0.204	0.138	0.678	0.876	0.576
AromaDb	1109	239	167	0.216	0.151	0.699	0.874	0.84
Flavornet	633	87	51	0.137	0.081	0.586	0.899	2.326
OdorDB	156	20	12	0.128	0.077	0.6	0.891	5.263

Note that the scaffolds for each chemical library are computed at the G/N/B level. Here, M is the Number of molecules, N is the Number of scaffolds, N_{sing} is the Number of singleton scaffolds, AUC is the Area under the curve, and P_{50} is the Percentage of scaffolds that account for 50% of the chemical library

produced by a specific set of pathways and from a limited number of precursors [98].

To further analyze the structural diversity of the four chemical libraries, we plotted the cyclic system retrieval (CSR) curves for the scaffolds computed at the three levels. We observed that the CSR curves follow a similar trend across the four libraries considered here (Fig. 7). However, the CSR curve for OdorDB shifts closer to the diagonal line compared to other libraries as one moves from G/N/B to the graph level (Fig. 7). In an ideal scenario, the CSR curve of a chemical library, wherein each chemical has a unique scaffold, will correspond to the

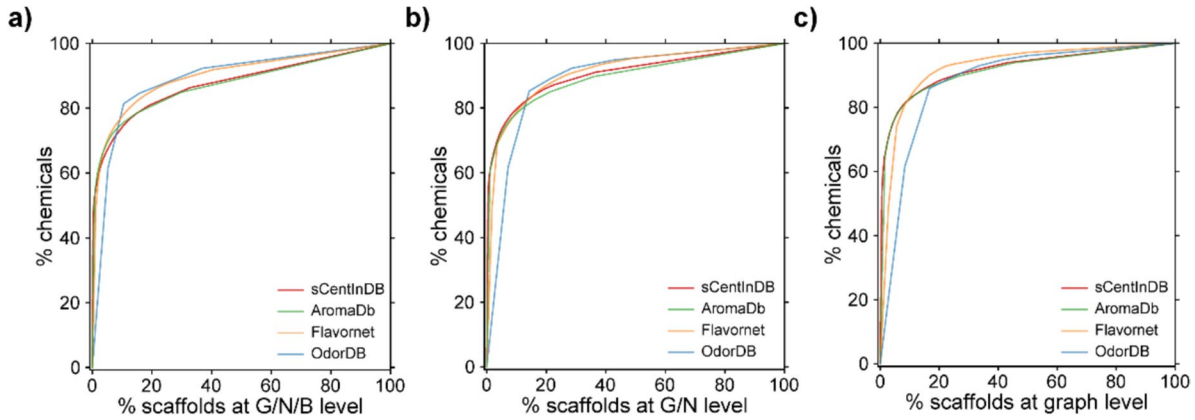


Fig. 7 Cyclic system retrieval (CSR) curves comparing the scaffold diversity across four chemical libraries containing aroma molecules namely, sCentInDB, AromaDb, Flavornet, and OdorDB. The x axis of the CSR curve is the percent of scaffold and the y axis is the percent of compounds that contain those scaffolds. **a** CSR curves for scaffolds at G/N/B level. **b** CSR curves for scaffolds at G/N level. **c** CSR curves for scaffolds at graph level

diagonal line. Similarly, the CSR curve of a chemical library, wherein scaffolds are evenly distributed among the chemicals, will also correspond to the diagonal line. This suggests the need for multiple metrics to fully assess the scaffold diversity [99]. Figure 8 is a molecular cloud visualization of the scaffolds that occur in five or more chemicals in sCentInDB. We observed that the benzene scaffold dominates sCentInDB and is present in 387 chemicals.

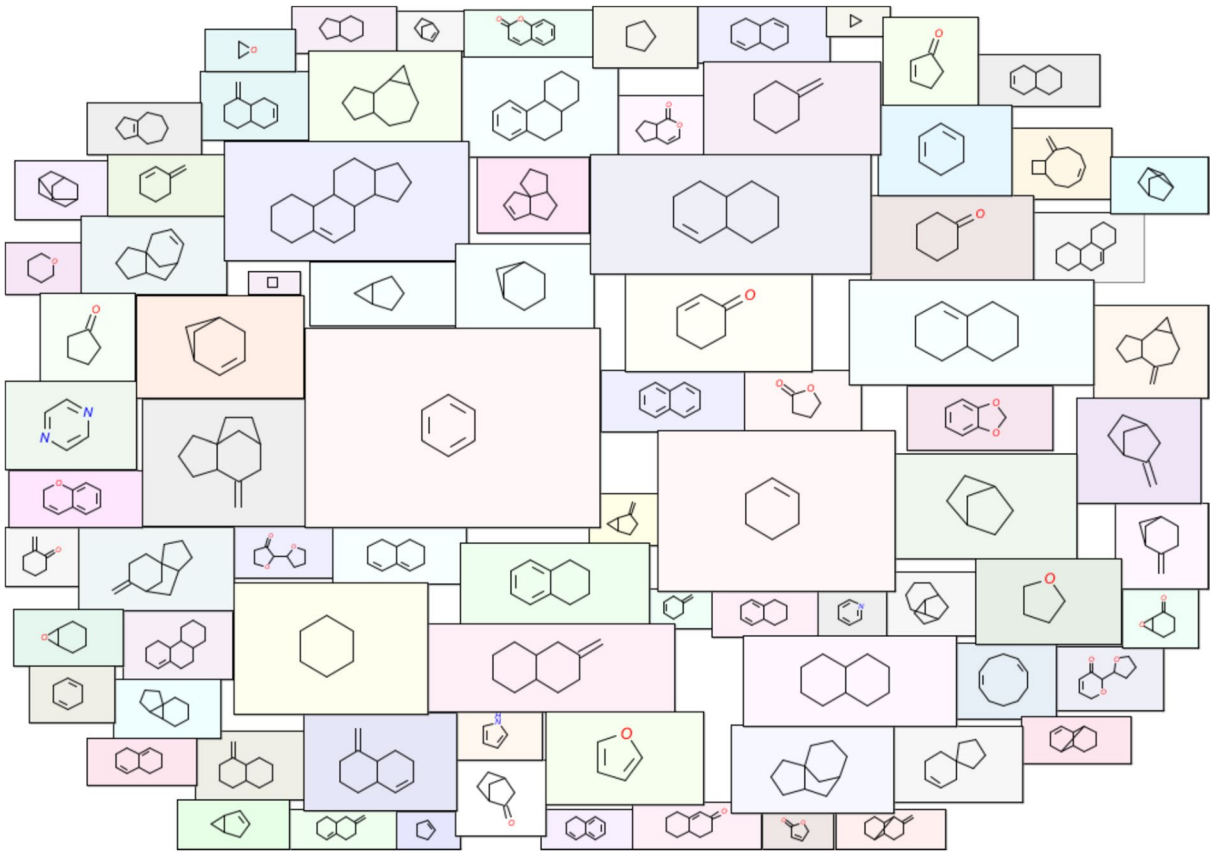


Fig. 8 Molecular cloud visualization of top scaffolds at the G/N/B level which are present in at least five chemicals in sCentInDB. In this visualization, the size of each scaffold image is proportional to its frequency of occurrence in the 3414 chemicals in the database

In addition, to identify therapeutically important scaffolds present in sCentInDB chemicals, we used the Therapeutic Target Database (TTD) [100]. We computed scaffolds for 22,561 drugs in TTD, accounting for 11,170 scaffolds at the G/N/B level. This had an overlap of 136 scaffolds with the 693 scaffolds in sCentInDB at the G/N/B level. Of these 136 shared scaffolds, 10 unique ones were associated with the terms 'Bacterial infection', 'Bacteremia', and 'Mycobacterium infection', as annotated in TTD.

Comparison of sCentInDB with related databases

We performed a comparative analysis of sCentInDB against previously built related databases. Based on this analysis, we find sCentInDB to be the most comprehensive resource for essential oils from Indian medicinal plants. Briefly, a few resources have been built previously to provide information on essential oils and their chemical constituents. Among them, EssOilDB [4] and AromaDb [61] (https://aromadb.cimapbioinfo.in/) were developed to provide detailed information about essential oils, chemical composition, and associated properties. However, the website of EssOilDB is no longer functional, while AromaDb has limited coverage in terms of essential oils and medicinal plants in comparison to sCentInDB (Table 2). Moreover, other related databases, including SuperScent [101], OdorDB (https://ordb.biotech.ttu.edu/OdorDB/), and Flavornet (https://www.flavornet.org/), focus on flavors and odor molecules rather than on

Table 2 Comparison of the features of sCentInDB with related databases, namely, AromaDb and AromaWeb

Description	sCentInDB	AromaDb	AromaWeb
	Curated information on essential oils of Indian medicinal plants	Curated information on essential oils of medicinal and aromatic plants	Information on essential oils and their use in aromatherapy
Number of essential oil profiles	2170	221	235
Number of plants	554	183	149
Chemical composition	Yes (with %)	Yes (with %)	Yes (not with %)
Odor information	Yes	Yes	Yes
Color information	Yes	No	Yes
Plant origin	Indian	Not specified	Global
Plant annotation	Yes	Yes	Yes
Plant part information	Yes	Yes	Yes
Plant sample location	Yes	No	No
Mapping of plant information to standard databases	Yes	No	No
Plant conservation status	Yes	No	Yes
Number of aroma molecules (or chemicals)	3420	1361	Not specified
Information on oil isolation method	Yes	No	Yes
Information on oil analysis method	Yes	No	No
Spectra information for chemicals	Yes	No	No
Chemical information including structure, physicochemical and ADME properties	Yes	Yes	No
Toxicological information for chemicals	No	Yes	No
Ecological information for chemicals	No	Yes	No
Hazard/safety information for chemicals	No	Yes	Yes
Mapping to chemical identifiers in standard databases	Yes	Yes	No
Receptor information for chemicals	No	Yes	No
Therapeutic uses of essential oils	Yes	Yes	Yes
Mapping of therapeutic uses to standard databases	Yes	No	No
General uses of essential oils	Yes	Yes (in oil description)	Yes (in oil description)
Published reference for chemical profiles of essential oils	Yes	Yes	Yes
Trade data for essential oils	No	Yes	No
Flat download of compiled dataset	Yes	No	No

essential oils. Furthermore, proprietary resources such as EOUdb (https://essentialoils.org/plans) and AromaWeb (https://www.aromaweb.com/) capture information on essential oil profiles and aromatherapy, but are not freely accessible for academic research, with EOUdb requiring login credentials. In Table 2, we present a detailed comparison of three resources, sCentInDB, AromaDb and AromaWeb. From this comparative analysis, sCentInDB emerges as the most comprehensive resource to date for essential oil profiles of Indian medicinal plants and associated properties.

Conclusions

sCentInDB is an elaborate resource encapsulating the essential oil chemical profiles of Indian medicinal plants. It contains manually curated information on 2170 oil profiles at the plant part level, along with extensive metadata. Additionally, we have compiled data on the therapeutic uses, odor, and color of oils at the plant part level. To enhance interoperability, we have mapped and hyperlinked plant names, chemical names, and therapeutic use terms to other standard databases, thus complying with FAIR principles. Furthermore, we analyzed the structural diversity of the chemicals in sCentInDB. From the chemical similarity network (CSN), we observed that terpenoids exhibit greater diversity than chemicals from other biosynthetic pathways. Finally, we analyzed the scaffold diversity of the chemical space in sCentInDB using cyclic system retrieval (CSR) curves, which revealed that aroma molecule libraries derived from plants typically have less diversity due to the limited number of biosynthetic pathways from which they can be synthesized.

The compilation of essential oil compositions of aromatic plants is important for both research and industrial applications. In particular, this knowledge is required when genetically modifying plants to alter the yield and composition of the oil [40]. Further, many plants used for essential oil extraction are sourced from the wild and therefore, stand the risk of extinction [102, 103]. Genetic engineering offers a viable alternative by enhancing oil yield and modifying composition to achieve desired properties. To achieve these goals, a comprehensive understanding of the chemical composition of essential oils is necessary. Prior to our work, there were a few databases containing such information. However, previous databases are specialized libraries compiled from different perspectives, and thus, have limitations. Moreover, no existing databases on essential oils specifically focus on Indian medicinal plants. Thus, sCentInDB aims to fill this gap, serving as a valuable resource for researchers working with Indian medicinal plants in therapeutic, perfumery, and cosmetic industries.

Environmental factors [30, 31] and developmental stage [32, 33] of a plant influences the composition of essential oils. However, sCentInDB does not include this information. Such data could have been valuable in determining the optimal conditions and growth stages for cultivating plant parts for essential oil extraction. Additionally, the database does not contain information on recipes or combinations of oils used in aromatherapy. While this was not the aim of the work, such data would be beneficial from a wellness perspective. Furthermore, trade-related information, such as the import and export data, was not collected, though it would have provided useful insights from a commercial standpoint.

Our database will be a highly valuable resource for the perfumery, pharmaceutical, fragrance, cosmetic, and food flavoring industries. It also enables users to find plants capable of producing their molecule of interest in high quantities. Additionally, by querying the database, chemical profiles similar to a desired plant can be identified, thereby revealing potential alternative plant sources. Although the structural diversity of essential oil chemicals is less in general [98], unique chemical entities distinct from chemical libraries of other sources can be expected due to their unique properties. Scaffold analogues or chemicals containing desired scaffolds can be retrieved from the database and further explored for drug discovery. Essential oils have been researched for their application as complementary therapy [104, 105] and for reducing the side effects of modern medicine [106]. Additionally, the chemical composition of the plants can give insights into their genetic makeup [2], making sCentInDB a valuable resource for systems biology research [107]. The data can also be used to predict the odor of new mixtures of chemicals and to identify plant combinations for desired odors using machine learning models. Lastly, the region-specific data compiled in sCentInDB can help explore the relationship between regional climate conditions and their impact on oil composition.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11030-025-11215-5.

Acknowledgements The authors thank Gokul Balaji Dhanakoti for computational assistance and Gaurav Kumar for assisting in data curation. Areejit Samal would like to acknowledge funding from the Department of Atomic Energy (DAE), Government of India through the Apex Project to The Institute of Mathematical Sciences (IMSc), Chennai, India. The funders have no role in the study design, data collection, data analysis, manuscript preparation, or decision to publish.

Author contributions Shanmuga Priya Baskaran: Conceptualization, Data Curation, Formal Analysis, Methodology, Software, Visualization, Writing—original draft, Writing—review & editing; Geetha Ranganathan: Conceptualization, Data Curation, Writing—original draft, Writing—review & editing; Ajaya Kumar Sahoo: Conceptualization, Data Curation, Formal Analysis, Methodology, Visualization, Writing—original draft, Writing—review & editing; Kishan Kumar: Conceptualization, Data Curation; Jayalakshmi Amaresan: Data Curation, Writing—original draft; Kundhanathan Ramesh: Data Curation; R.P.

Vivek-Ananth: Conceptualization, Data Curation; Writing—review & editing; Areejit Samal: Conceptualization, Formal Analysis, Methodology, Supervision, Writing—original draft, Writing—review & editing.

Funding Open access funding provided by Department of Atomic Energy. Funding was provided by Department of Atomic Energy, Government of India.

Data availability The data comprising the database is available as flat files in the "Download" tab at https://cb.imsc.res.in/scentindb/download. Supplementary figures S1 and S2 are available in the Supplementary Information file.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Samarth RM, Samarth M, Matsumoto Y (2017) Medicinally important aromatic plants with radioprotective activity. Future Sci. https://doi.org/10.4155/fsoa-2017-0061
2. Baser KHC, Buchbauer G (2009) Handbook of essential oils: science, technology, and applications, 1st edn. CRC Press
3. Bakkali F, Averbeck S, Averbeck D, Idaomar M (2008) Biological effects of essential oils—a review. Food Chem Toxicol 46:446–475. https://doi.org/10.1016/j.fct.2007.09.106
4. Kumari S, Pundhir S, Priya P et al (2014) EssOilDB: a database of essential oils reflecting terpene composition and variability in the plant kingdom. Database 2014:bau120. https://doi.org/10.1093/database/bau120
5. Zuzarte M, Salgueiro L (2015) Essential oils chemistry. In: de Sousa DP (ed) Bioactive essential oils and cancer. Springer International Publishing, Cham, pp 19–61
6. Harrewijn P, Van Oosten AM, Piron P (2001) Natural terpenoids as messengers: a multidisciplinary study of their production, biological functions, and practical applications, 1st edn. Springer Science & Business Media, Berlin
7. Karban R, Baldwin IT, Baxter KJ et al (2000) Communication between plants: induced resistance in wild tobacco plants following clipping of neighboring sagebrush. Oecologia 125:66–71. https://doi.org/10.1007/PL00008892
8. Kessler A, Baldwin IT (2001) Defensive function of herbivore-induced plant volatile emissions in nature. Science 291:2141–2144. https://doi.org/10.1126/science.291.5511.2141
9. Mimica-Dukić N, Božin B, Soković M, Mihajlović B, Matavulj M (2003) Antimicrobial and antioxidant activities of three mentha species essential oils. Planta Med 69:413–419. https://doi.org/10.1055/s-2003-39704
10. Bardon S, Foussard V, Fournel S, Loubat A (2002) Monoterpenes inhibit proliferation of human colon cancer cells by modulating cell cycle-related protein expression. Cancer Lett 181:187–194. https://doi.org/10.1016/S0304-3835(02)00047-2
11. Bardon S, Picard K, 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. https://doi.org/10.1080/01635589809514708
12. Liao J-C, Deng J-S, Chiu C-S et al (2012) Anti-inflammatory activities of *Cinnamomum cassia* constituents in vitro and in vivo. Evid Based Complement Altern Med 2012:429320. https://doi.org/10.1155/2012/429320
13. Abdollahi M, Salehnia A, Mortazavi SR et al (2003) Antioxidant, antidiabetic, antihyperlipidemic, reproduction stimulatory properties and safety of essential oil of *Satureja khuzestanica* in rat in vivo: a oxicopharmacological study. Med Sci Monit 9(9):BR331–BR335
14. de Sousa DP, Damasceno ROS, Amorati R et al (2023) Essential oils: chemistry and pharmacological activities. Biomolecules 13:1144. https://doi.org/10.3390/biom13071144
15. De Andrade TU, Brasil GA, Endringer DC et al (2017) Cardiovascular activity of the chemical constituents of essential oils. Molecules 22:1539. https://doi.org/10.3390/molecules22091539
16. Yingngam B, Brantner AH (2015) Factorial design of essential oil extraction from agraea fragrans Roxb. flowers and evaluation of its biological activities for perfumery and cosmetic applications. Int J Cosmet Sci 37:272–281. https://doi.org/10.1111/ics.12192
17. Manou I, Bouillard L, Devleeschouwer MJ, Barel AO (1998) Evaluation of the preservative properties of *Thymus vulgaris* essential oil in topically applied formulations under a challenge test. J Appl Microbiol 84:368–376. https://doi.org/10.1046/j.1365-2672.1998.00353.x
18. Sharmeen JB, Mahomoodally FM, Zengin G, Maggi F (2021) Essential oils as natural sources of fragrance compounds for cosmetics and cosmeceuticals. Molecules 26:666. https://doi.org/10.3390/molecules26030666
19. Tongnuanchan P, Benjakul S (2014) Essential oils: extraction, bioactivities, and their uses for food preservation. J Food Sci 79:R1231–R1249. https://doi.org/10.1111/1750-3841.12492
20. Baptista-Silva S, Borges S, Ramos OL et al (2020) The progress of essential oils as potential therapeutic agents: a review. J Essent Oil Res 32:279–295. https://doi.org/10.1080/10412905.2020.1746698
21. Ali B, Al-Wabel NA, Shams S et al (2015) Essential oils used in aromatherapy: a systemic review. Asian Pac J Trop Biomed 5:601–611. https://doi.org/10.1016/j.apjtb.2015.05.007
22. Bawa KS, Sengupta A, Chavan V et al (2021) Securing biodiversity, securing our future: a national mission on biodiversity and human well-being for India. Biol Cons 253:108867. https://doi.org/10.1016/j.biocon.2020.108867
23. Chitale VS, Behera MD, Roy PS (2014) Future of endemic flora of biodiversity hotspots in India. PLoS ONE 9:e115264. https://doi.org/10.1371/journal.pone.0115264
24. Raina AP, Kumar A, Dutta M (2013) Chemical characterization of aroma compounds in essential oil isolated from "Holy Basil" *(Ocimum tenuiflorum* L.) grown in India. Genet Resour Crop Evol 60:1727–1735. https://doi.org/10.1007/s10722-013-9981-4
25. Wany A, Kumar A, Nallapeta S et al (2014) Extraction and characterization of essential oil components based on geraniol and citronellol from Java citronella (*Cymbopogon winterianus* Jowitt). Plant Growth Regul 73:133–145. https://doi.org/10.1007/s10725-013-9875-7

26. Raut JS, Karuppayil SM (2014) A status review on the medicinal properties of essential oils. Ind Crops Prod 62:250–264. https://doi.org/10.1016/j.indcrop.2014.05.055
27. Abd Manaf M, Jai J, Raslan R et al (2015) Microencapsulation methods of volatile essential oils—a review. Adv Mater Res 1113:679–683. https://doi.org/10.4028/www.scientific.net/AMR.1113.679
28. Sadgrove NJ, Padilla-González GF, Phumthum M (2022) Fundamental chemistry of essential oils and volatile organic compounds. Methods Anal Authent Plants 11:789. https://doi.org/10.3390/plants11060789
29. Dajic Stevanovic Z, Sieniawska E, Glowniak K et al (2020) Natural macromolecules as carriers for essential oils: from extraction to biomedical application. Front Bioeng Biotechnol 8:563. https://doi.org/10.3389/fbioe.2020.00563
30. Figueiredo AC, Barroso JG, Pedro LG, Scheffer JJC (2008) Factors affecting secondary metabolite production in plants: volatile components and essential oils. Flavour Fragr J 23:213–226. https://doi.org/10.1002/ffj.1875
31. Fernández-Sestelo M, Carrillo JM (2020) Environmental effects on yield and composition of essential oil in wild populations of spike lavender (*Lavandula latifolia* Medik). Agriculture 10:626. https://doi.org/10.3390/agriculture10120626
32. Mallavarapu GR, Kulkarni RN, Baskaran K et al (1999) Influence of plant growth stage on the essential oil content and composition in Davana (*Artemisia pallens* Wall.). J Agric Food Chem 47:254–258. https://doi.org/10.1021/jf980624c
33. Gershenzon J, McConkey ME, Croteau RB (2000) Regulation of monoterpene accumulation in leaves of peppermint1. Plant Physiol 122:205–214. https://doi.org/10.1104/pp.122.1.205
34. Rathore S, Mukhia S, Kumar R, Kumar R (2023) Essential oil composition and antimicrobial potential of aromatic plants grown in the mid-hill conditions of the Western Himalayas. Sci Rep 13:4878. https://doi.org/10.1038/s41598-023-31875-3
35. Dhifi W, Bellili S, Jazi S et al (2016) Essential oils' chemical characterization and investigation of some biological activities: a critical review. Medicines 3:25. https://doi.org/10.3390/medicines3040025
36. Li H, Luo L, Ma M, Zeng L (2018) Characterization of volatile compounds and sensory analysis of jasmine scented black tea produced by different scenting processes. J Food Sci 83:2718–2732. https://doi.org/10.1111/1750-3841.14340
37. Nedeltcheva-Antonova D, Gechovska K, Bozhanov S, Antonov L (2022) Exploring the chemical composition of Bulgarian Lavender absolute (*Lavandula angustifolia* Mill.) by GC/MS and GC-FID. Plants 11:3150. https://doi.org/10.3390/plants11223150
38. Raičević V, Mladenović M, Aćimović M, Radulović N (2024) New esters from the essential oil of dry flowers of elder (*Sambucus nigra* L.). J Sci Food Agric 104:1308–1321. https://doi.org/10.1002/jsfa.13012
39. Antonova DV, Medarska YN, Stoyanova AS et al (2021) Chemical profile and sensory evaluation of Bulgarian rose (*Rosa damascena* Mill.) aroma products, isolated by different techniques. J Essent Oil Res 33:171–181. https://doi.org/10.1080/10412905.2020.1839583
40. Mahmoud S, Croteau R (2001) Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. Proc Natl Acad Sci U S A 98:8915–8920. https://doi.org/10.1073/pnas.141237298
41. Cohen SM, Eisenbrand G, Fukushima S et al (2021) FEMA GRAS assessment of natural flavor complexes: origanum oil, thyme oil and related phenol derivative-containing flavoring ingredients. Food Chem Toxicol 155:112378. https://doi.org/10.1016/j.fct.2021.112378
42. Sartori Tamburlin I, Roux E, Feuillée M et al (2021) Toxicological safety assessment of essential oils used as food supplements to establish safe oral recommended doses. Food Chem Toxicol 157:112603. https://doi.org/10.1016/j.fct.2021.112603
43. Radulović NS, Mladenović MZ, Dekić MS, Boylan F (2023) Synthesis of small libraries of natural products: part II: Identification of a new natural product from the essential oil of *Pleurospermum austriacum* (L.) Hoffm. (Apiaceae). Molecules 28:4574. https://doi.org/10.3390/molecules28124574
44. Balandrin MF, Klocke JA, Wurtele ES, WmH B (1985) Natural plant chemicals: sources of industrial and medicinal materials. Science 228:1154–1160. https://doi.org/10.1126/science.3890182
45. Gounaris Y (2010) Biotechnology for the production of essential oils, flavours and volatile isolates. A review. Flavour Fragr J 25:367–386. https://doi.org/10.1002/ffj.1996
46. Gómez-Galera S, Pelacho AM, Gené A et al (2007) The genetic manipulation of medicinal and aromatic plants. Plant Cell Rep 26:1689–1715. https://doi.org/10.1007/s00299-007-0384-x
47. Yao X, Wuzhang K, Peng B et al (2022) Engineering the expression of plant secondary metabolites-genistein and scutellarin through an efficient transient production platform in *Nicotiana benthamiana* L. Front Plant Sci 13:994792. https://doi.org/10.3389/fpls.2022.994792
48. Zhu X, Liu X, Liu T et al (2021) Synthetic biology of plant natural products: from pathway elucidation to engineered biosynthesis in plant cells. Plant Communications 2:100229. https://doi.org/10.1016/j.xplc.2021.100229
49. Reed J, Stephenson MJ, Miettinen K et al (2017) A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. Metab Eng 42:185–193. https://doi.org/10.1016/j.ymben.2017.06.012
50. Kadotani N, Ikegami M (2002) Production of patchouli mild mosaic virus resistant patchouli plants by genetic engineering of coat protein precursor gene. Pest Manag Sci 58:1137–1142. https://doi.org/10.1002/ps.581
51. Lawrence BM (2006) Mint: the genus Mentha, 1st edn. CRC Press
52. Zhang R, Yu S, Bai H, Ning K (2017) TCM-Mesh: the database and analytical system for network pharmacology analysis for TCM preparations. Sci Rep 7:2821. https://doi.org/10.1038/s41598-017-03039-7
53. Pilon AC, Valli M, Dametto AC et al (2017) NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. Sci Rep 7:7215. https://doi.org/10.1038/s41598-017-07451-x
54. Zeng X, Zhang P, Wang Y et al (2019) CMAUP: a database of collective molecular activities of useful plants. Nucleic Acids Res 47:D1118–D1127. https://doi.org/10.1093/nar/gky965
55. Pilón-Jiménez BA, Saldívar-González FI, Díaz-Eufracio BI, Medina-Franco JL (2019) BIOFACQUIM: A Mexican compound database of natural products. Biomolecules. https://doi.org/10.3390/biom9010031
56. Sorokina M, Merseburger P, Rajan K et al (2021) COCONUT online: collection of open natural products database. J Cheminform 13:2. https://doi.org/10.1186/s13321-020-00478-9
57. Vivek-Ananth RP, Mohanraj K, Sahoo AK, Samal A (2023) IMPPAT 2.0: an enhanced and expanded phytochemical atlas of Indian medicinal plants. ACS Omega 8:8827–8845. https://doi.org/10.1021/acsomega.3c00156
58. Gallo K, Kemmler E, Goede A et al (2023) SuperNatural 3.0—a database of natural products and natural product-based derivatives. Nucleic Acids Res 51:D654–D659. https://doi.org/10.1093/nar/gkac1008
59. Hou D, Lin H, Feng Y et al (2024) CMAUP database update 2024: extended functional and association information of useful

plants for biomedical research. Nucleic Acids Res 52:D1508–D1518. https://doi.org/10.1093/nar/gkad921
60. Chandrasekhar V, Rajan K, Kanakam SRS et al (2025) COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database. Nucleic Acids Res 53:D634–D643. https://doi.org/10.1093/nar/gkae1063
61. Kumar Y, Prakash O, Tripathi H et al (2018) AromaDb: a database of medicinal and aromatic plant's aroma molecules with phytochemistry and therapeutic potentials. Front Plant Sci 9:1081. https://doi.org/10.3389/fpls.2018.01081
62. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP et al (2018) IMPPAT: a curated database of Indian medicinal plants, phytochemistry and therapeutics. Sci Rep 8:4329. https://doi.org/10.1038/s41598-018-22631-z
63. Kim S, Chen J, Cheng T et al (2023) PubChem 2023 update. Nucleic Acids Res 51:D1373–D1380. https://doi.org/10.1093/nar/gkac956
64. O'Boyle NM, Banck M, James CA et al (2011) Open babel: an open chemical toolbox. J Cheminform 3:33. https://doi.org/10.1186/1758-2946-3-33
65. Davies M, Nowotka M, Papadatos G et al (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res 43:W612–W620. https://doi.org/10.1093/nar/gkv352
66. Zdrazil B, Felix E, Hunter F et al (2023) The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res 52:D1180–D1192. https://doi.org/10.1093/nar/gkad1004
67. Irwin JJ, Shoichet BK (2005) Zinc—a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182. https://doi.org/10.1021/ci049714+
68. Irwin JJ, Sterling T, Mysinger MM et al (2012) Zinc: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757–1768. https://doi.org/10.1021/ci3001277
69. Sterling T, Irwin JJ (2015) Zinc 15—ligand discovery for everyone. J Chem Inf Model 55:2324–2337. https://doi.org/10.1021/acs.jcim.5b00559
70. Djoumbou Feunang Y, Eisner R, Knox C et al (2016) ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminform 8:61. https://doi.org/10.1186/s13321-016-0174-y
71. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. https://doi.org/10.1038/srep42717
72. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32:1466–1474. https://doi.org/10.1002/jcc.21707
73. Sud M (2016) MayaChemTools: an open source package for computational drug discovery. J Chem Inf Model 56:2292–2297. https://doi.org/10.1021/acs.jcim.6b00505
74. Ertl P, Roggo S, Schuffenhauer A (2008) Natural product-likeness score and its application for prioritization of compound libraries. J Chem Inf Model 48:68–74. https://doi.org/10.1021/ci700286x
75. Lipkus AH, Yuan Q, Lucas KA et al (2008) Structural diversity of organic chemistry. a scaffold analysis of the CAS registry. J Org Chem 73:4443–4451. https://doi.org/10.1021/jo8001276
76. Lipkus AH, Watkins SP, Gengras K et al (2019) Recent changes in the scaffold diversity of organic chemistry as seen in the CAS registry. J Org Chem 84:13948–13956. https://doi.org/10.1021/acs.joc.9b02111
77. Tanimoto TT (1957) IBM internal report. Nov 17: 1957
78. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50:742–754. https://doi.org/10.1021/ci100050t
79. Bienfait B, Ertl P (2013) JSME: a free molecule editor in JavaScript. J Cheminform 5:24. https://doi.org/10.1186/1758-2946-5-24
80. Wilkinson MD, Dumontier M, IjJ A et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18
81. Mamadalieva NZ, Akramov DKh, Ovidi E et al (2017) Aromatic medicinal plants of the Lamiaceae family from Uzbekistan: ethnopharmacology, essential oils composition, and biological activities. Medicines 4:8. https://doi.org/10.3390/medicines4010008
82. Marchioni I, Najar B, Ruffoni B et al (2020) Bioactive compounds and aroma profile of some Lamiaceae edible flowers. Plants 9:691. https://doi.org/10.3390/plants9060691
83. Reyes-Jurado F, Franco-Vega A, Ramírez-Corona N et al (2015) Essential oils: antimicrobial activities, extraction methods, and their modeling. Food Eng Rev 7:275–297. https://doi.org/10.1007/s12393-014-9099-2
84. Sharma A, Saha BK, Kumar R, Varadwaj PK (2021) OlfactionBase: a repository to explore odors, odorants, olfactory receptors and odorant–receptor interactions. Nucleic Acids Res 50:D678–D686. https://doi.org/10.1093/nar/gkab763
85. Ohloff G, Pickenhagen W, Kraft P, Grau F (2022) Scent and chemistry: the molecular world of odors. John Wiley & Sons
86. Moghaddam M, Mehdizadeh L (2017) Chapter 13—chemistry of essential oils and factors influencing their constituents. In: Grumezescu AM, Holban AM (eds) Soft chemistry and food fermentation. Academic Press, pp 379–419
87. Bernhard RA, Marr AG (1960) The oxidation of terpenes. I. Mechanism and reaction products of D-limonene autoxidation. J Food Sci 25:517–530. https://doi.org/10.1111/j.1365-2621.1960.tb00363.x
88. Turek C, Stintzing FC (2013) Stability of essential oils: a review. Compr Rev Food Sci Food Saf 12:40–53. https://doi.org/10.1111/1541-4337.12006
89. Kim HW, Wang M, Leber CA et al (2021) NPClassifier: a deep neural network-based structural classification tool for natural products. J Nat Prod 84:2795–2807. https://doi.org/10.1021/acs.jnatprod.1c00399
90. Sorokina M, Steinbeck C (2019) NaPLeS: a natural products likeness scorer—web application and database. J Cheminform 11:55. https://doi.org/10.1186/s13321-019-0378-z
91. Vanii Jayaseelan K, Moreno P, Truszkowski A et al (2012) Natural product-likeness score revisited: an open-source, open-data implementation. BMC Bioinform 13:106. https://doi.org/10.1186/1471-2105-13-106
92. Pichersky E, Raguso RA (2018) Why do plants produce so many terpenoid compounds? New Phytol 220:692–702. https://doi.org/10.1111/nph.14178
93. Masyita A, Mustika Sari R, Dwi Astuti A et al (2022) Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. Food Chem 13:100217. https://doi.org/10.1016/j.fochx.2022.100217
94. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303
95. Vivek-Ananth RP, Sahoo AK, Baskaran SP, Samal A (2023) Scaffold and structural diversity of the secondary metabolite space of medicinal fungi. ACS Omega 8:3102–3113. https://doi.org/10.1021/acsomega.2c06428
96. González-Medina M, Owen JR, El-Elimat T et al (2017) Scaffold diversity of fungal metabolites. Front Pharmacol 8:180. https://doi.org/10.3389/fphar.2017.00180

97. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39:2887–2893. https://doi.org/10.1021/jm9602928
98. Smith RL, Cohen SM, Doull J et al (2005) A procedure for the safety evaluation of natural flavor complexes used as ingredients in food: essential oils. Food Chem Toxicol 43:345–363. https://doi.org/10.1016/j.fct.2004.11.007
99. Yongye AB, Waddell J, Medina-Franco JL (2012) Molecular scaffold analysis of natural products databases in the public domain. Chem Biol Drug Des 80:717–724. https://doi.org/10.1111/cbdd.12011
100. Zhou Y, Zhang Y, Zhao D et al (2024) TTD: therapeutic target database describing target druggability information. Nucleic Acids Res 52:D1465–D1477. https://doi.org/10.1093/nar/gkad751
101. Dunkel M, Schmidt U, Struck S et al (2008) SuperScent—a database of flavors and scents. Nucleic Acids Res 37:D291–D294. https://doi.org/10.1093/nar/gkn695
102. Gowthami R, Sharma N, Pandey R, Agrawal A (2021) Status and consolidated list of threatened medicinal plants of India. Genet Resour Crop Evol 68:2235–2263. https://doi.org/10.1007/s10722-021-01199-0
103. Hamilton AC (2004) Medicinal plants, conservation and livelihoods. Biodivers Conserv 13:1477–1517. https://doi.org/10.1023/B:BIOC.0000021333.23413.42
104. Cheng H, Lin L, Wang S et al (2022) Aromatherapy with single essential oils can significantly improve the sleep quality of cancer patients: a meta-analysis. BMC Complement Med Ther 22:187. https://doi.org/10.1186/s12906-022-03668-0
105. Ou M-C, Hsu T-F, Lai AC et al (2012) Pain relief assessment by aromatic essential oil massage on outpatients with primary dysmenorrhea: a randomized, double-blind clinical trial. J Obstet Gynaecol Res 38:817–822. https://doi.org/10.1111/j.1447-0756.2011.01802.x
106. Stea S, Beraudi A, De Pasquale D (2014) Essential oils for complementary treatment of surgical patients: state of the art. Evid Based Complement Altern Med 2014:726341. https://doi.org/10.1155/2014/726341
107. Yuan JS, Galbraith DW, Dai SY et al (2008) Plant systems biology comes of age. Trends Plant Sci 13:165–171. https://doi.org/10.1016/j.tplants.2008.02.003

Authors and Affiliations

Shanmuga Priya Baskaran[1,2] · Geetha Ranganathan[1] · Ajaya Kumar Sahoo[1,2] · Kishan Kumar[1] · Jayalakshmi Amaresan[1] · Kundhanathan Ramesh[1] · R. P. Vivek-Ananth[1,2,3] · Areejit Samal[1,2]

✉ Areejit Samal
asamal@imsc.res.in

[1] The Institute of Mathematical Sciences (IMSc), Chennai 600113, India

[2] Homi Bhabha National Institute (HBNI), Mumbai 400094, India

[3] Present Address: Buck Institute for Research on Aging, Novato, CA, USA