Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GC-MS chemoprofiling, chemotaxonomic differentiation and PASS online prediction of biological activities of essential oils from five Kenyan Hypericum species

Jared Misonge Onyancha

jonyancha@mku.ac.ke

Mount Kenya University

Geoffrey Sengera Ogeto

Kisii University

Peter Maina Mbugua

Mount Kenya University

May Chao Mungulle

Mount Kenya University

Norman Wairagu Wachira

Forest Products Research Centre

Research Article

Keywords: Hypericaceae, medicinal plants, phytochemistry, quality control, volatile oils

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11063421/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

GC-MS chemoprofiling, chemotaxonomic differentiation and PASS online prediction of biological activities of essential oils from five Kenyan Hypericum species

Jared Misonge Onyancha[1], Geoffrey Sengera Ogeto[2], Peter Maina Mbugua[3], May Chao Mungulle[1] and Norman Wairagu Wachira[4]

[1]Department of Pharmacognosy, Mount Kenya University, P.O Box 342-01000, Thika, Kenya
[2]Department of Pharmacy, Kisii University, P. O. Box 408-40200, Kisii, Kenya
[3]Centre for Malaria Elimination, Mount Kenya University, P.O Box 342-01000, Thika, Kenya.
[4]Forest Products Research Centre, Kenya Forest Research Institute, Nairobi, Kenya
Corresponding author: Jared Misonge Onyancha
Email of the corresponding author: jonyancha@mku.ac.ke

Abstract

Hypericum species are potential sources of essential oils. Whereas there exists reported work on volatile constituents of Hypericum essential oils (Eos) from other parts of the world, the species of Hypericums in Kenya remain largely unexplored. The current study assessed the chemical variability of Eos from five Hypericum species collected in Kenya using multivariate chemometric approaches. Eos were extracted by hydrodistillation and analyzed using gas chromatography–mass spectrometry (GC–MS). Further, prediction of activity spectra for substances was done. After data cleaning, including merging duplicate compounds, removing absent compounds, and autoscaling, a dataset of 113 compounds was subjected to principal component analysis (PCA), Pearson correlation analysis, hierarchical cluster analysis, phytochemical class profiling, diversity indices, Venn analysis, and discriminating-compound assessment. The volatile profiles showed clear interspecific variation. Sesquiterpenes predominated in *Hypericum kiboense*, *Hypericum roeperianum*, *Hypericum peplidifolium*, and *Hypericum scioanum*, whereas *Hypericum revolutum* contained a stronger monoterpene component dominated by α-pinene (43.04%) and caryophyllene

(19.40%). Characteristic major constituents included valencene in *H. kiboense* (28.77%), 6-hydroxy-1-phenylhexane-1,3-dione in *H. roeperianum* (11.73%), caryophyllene in *H. peplidifolium* (15.06%), and caryophyllene oxide in *H. scioanum* (10.68%). The first two principal components explained 68.2% of total variation and clearly separated the species. Correlation analysis revealed generally weak similarities among species, except for the strong relationship between *H. peplidifolium* and *H. scioanum* (r = 0.81). Cluster analysis identified three major chemotypic groups. Diversity metrics further highlighted substantial differences in richness, evenness, and overall diversity among species. The five Hypericum species possess distinct but partially overlapping volatile oil fingerprints and are not chemically interchangeable. The results provide a valuable foundation for chemotaxonomy, quality control, and future bioactivity-guided investigations. The predicted biological activity profile showed that antieczematic activity had the highest number of occurrences across the investigated *Hypericum* essential oils, followed by antineoplastic and apoptosis agonist activities. The ranking table further showed that some less frequent activities, such as spasmolytic urinary activity and mucomembranous protection, had high weighted mean Pa scores, suggesting strong but more narrowly distributed predicted bioactivity.

Keywords: *Hypericaceae,* medicinal plants, phytochemistry, quality control, volatile oils

Introduction

Since antiquity, aromatic plants accumulate essential oils (Eos) that are applied globally as ingredients in the pharmaceutical, nutraceutical and cosmeceutical formulations. The Hypericaceae family contains 700 species of which about 590 are in the genus Hypericum, therefore making it to contain the largest number of plants in the family represented by herbs,

shrubs and rarely trees (1,2). The plants of the genus Hypericum have a worldwide distribution as either native or introduced plant species with exception of Antarctica. The genus is known for it is contribution to both traditional complementary and alternative medicine and conventional medicine. They have diverse secondary metabolites responsible for the numerous pharmacological activities, the main characteristic compounds being such hypericin and pseudohypericin (naphthodianthrones) and others classes of secondary metabolized like phloroglucinols (hyperforin and adhyperforin), flavonoids and their glycosidic derivatives (quercetin, rutin, quercitrin, hyperoside), and biflavonoids (amentoflavone)(3). In addition, Hypericum plants are classified as aromatic plants. though being low essential oil yielding plants, they are sources of valuable essential oils, some of which as from *H. perforatum* being available commercially application on wound and burns (4).

The study examined five Hypericum species indigenous to Kenya for Eos. With the exception of *Hypricum revolutum* the chemical composition of the essential oils of *Hypricum kiboense, Hypricum roeperianum Hypricum peplidifolium, Hypricum scioanum* have not been reported to date (5–7). The aim of this study was to demonstrated the volatile fingerprints the Eos of five *Hypericum* species endemic to Kenya. Therefore, adding to the database of phytochemical knowledge for the genus. In addition to this, a comparison between the volatile oil components of the Hypericum species were performed as a potential quality parameter related to chemotaxonomic classification. Prediction of activity spectra for substances (PASS) was also done for the most abundant volatile compounds.

Materials and Methods

Area of study

The plants in this study were collected mainly from two counties in Kenya. Kericho county, at Kedowa area and Nyandarua county at Mutubio gate in Aberdare National Park. Kedowa is located along the busy Kericho-Nakuru Highway. Kedowa area is characterized with high-elevation of upto 2,306 meters and cool and temperatures typically ranging between 12°C to 24°C (54°F to 75°F) and frequent afternoon showers and thunderstorms during wet seasons.



Figure 1: Map showing the geographical area where the *Hypericum* species were collected

Plant Materials

The leaves of *H. kiboense*, *H. scioanum*, *H. peplidifolium*, and *H. revolutum* were collected from mature and flowering plants in the Month of October, 2025. *H. roperianum* was also, collected in the same Month. However, from a far much different location at Kipchorian River near the road between Kedowa town and Ringa secondary school in Kipkelion East subcounty, Kericho county (-0.228511, 35.548528) at an elevation of 2125.066 m. The collection was done during day time from different sites at Aberdare ranges in Kenya. The plants were Identified at the Herbarium Department in the East Africa Herbarium in Nairobi where the voucher specimens (JM0-1-2025,

JMO-2-2025, JMO-3-2025, 002J2024 and JMO-5-2025 respectively. The duplicate voucher specimens were deposited at Mount Kenya University Herbarium in the School of Pharmacy.

Isolation and GCMS analyses of essential oils

Approximately 250 g of shade dried and coarsely ground leaves of *H. kiboense*, *H. roperianum*, *H. peplidifolium, H. scioanum*, and *H. revolutum.* The essential oils were isolated by hydrodistillation using a Clevenger-like apparatus for four hours. The extracted oil was dried using anhydrous sodium sulfate in a filter. The essential oils then put drained into air tight vials and stored in 4 ^{0}C until further analysis. The volatile chemical component in the essential oils from the four different species collected from Kenya were analysed using Japan's Shimadzu QP2010SE GC-MS coupled to an autosampler as described in our early studies (5).

Statistical analyses

The raw GC-MS data matrix of 128 volatile compounds identified in 5 species of Hypericum was systematically preprocessed before the multivariate analysis. Initial recordings of compound abundances were as the percentage relative peak areas. On the five species studied, the abundance values for *H. kiboense*, *H. roeperianum*, *H. peplidifolium*, *H. scioanum* and *H. revolutum* were summarized in a species/compound matrix. In cases where duplicate data of the same compound were found in a species, the abundance values were added to produce a single representative value. The items that did not occur in all five species, as shown by zero abundance in all samples, were

not included in the matrix to avoid distorting the multivariate space. The resulting working matrix was 113 compounds present with non-zero abundance in at least one species. To deal with the varying scales of the abundances of the compounds, the data were subjected to autoscaling, with each variable being mean-centered and divided by its standard deviation. This conversion was done such that any compounds, irrespective of the absolute abundance magnitude, were balanced to subsequent analysis.

Principal Component Analysis

The scaled species x compound matrix was subjected to Principal Component Analysis using the PCA function in the FactoMineR package of R statistical software. The analysis reduced the initial 113-dimensional space of variables to a reduced set of uncorrelated (or, in other words, orthogonal) principal components that explained as much variance as possible in the data. The Kaiser criterion was used to decide on the number of components to be retained for interpretation, and the components whose eigenvalues exceeded one were retained, which was further justified by inspecting the scree plot, which shows the decreasing eigenvalues of successive components. The two major components, which together accounted most of the overall variance, were chosen to be visualized and interpreted to find the separation patterns in the species.

Loadings of the components were pulled out to determine the volatile compounds that were most strongly correlated with each main axis. Variables that have an absolute loading that is greater than 0.7 were deemed to be strong contributors to a particular component. The relative contribution of each compound was further determined with the squared cosine which is the quality of representation of each variable on the subspace of the principal components. Compounds whose cos^2 was high (usually greater than 0.5) were said to be well-represented in the reduced dimensional space.

Species Correlation Analysis

To test the pairwise correlations between species using their volatile profiles, Pearson product-moment correlation matrix was calculated. The abundance values of all 113 compounds were used in the analysis as each species is a variable in multidimensional space. The resulting coefficient r was between -1 and +1 where positive coefficients showed no difference in chemical profiles between species pairs and negative coefficients showed inverse relationships between abundance. Correlation coefficients were postulated as follows: a value of higher than 0.7 meant strong positive relationship, a value of between 0.5 and 0.7 meant moderate relationships, a value of between 0.3 and 0.5 meant weak relationships, and a value of lower than 0.3 meant none or insignificant linear relationships. The importance of each correlation was measured at the 95 percent confidence level, but the main focus was made on the strength of association since this was an exploratory analysis.

Hierarchical Cluster Analysis

Hierarchical cluster analysis, which was done to supplement the PCA results, was used to show the natural cluster structure of the five species without the need to define a set number of clusters. The Euclidean distance was used to compute the distance matrix in which the dissimilarity between species pairs was computed. This is the measure of choice since it maintains the geometric aspects of the scaled data space and is suitable with continuous abundance data. Agglomerative hierarchical clustering of the distance matrix was then done with Ward minimum variance as the linkage criterion. This algorithm aims at creating clusters by achieving a minimum within-cluster sum of squares at each fusion step, resulting in well separated compact spherical clusters in the multivariate space.

Chemical Class Profiling

The 113 volatile compounds were categorized into one of four phytochemical classes (monoterpenes C10 hydrocarbons and oxygenated derivatives), (sesquiterpenes C15 hydrocarbons and oxygenated derivatives), (diterpenes C20 compounds and derivatives, and miscellaneous volatiles not belonging Relative abundance of both chemical classes was determined in each species by adding together the percentage abundances of all compounds of each type. Class totals were then put into 100% by dividing each class by the total sum of the compounds in that species and multiplying by 100.

Diversity Indices

Three complementary indices that measure various facets of community structure were used to assess the chemical diversity of each species. The richness of species was defined as the number of different volatile compounds in each species, which is the number of chemical structures found in each species, not the proportions of these structures. Shannon diversity index was determined using the formula $H' = -\sum_{i=1}^{S} p_i \ln(p_i)$

whereby, p_i is the proportional abundance of the i-th compound in the total abundance of all compounds in the particular species and S is the total number of compounds. The index takes into consideration the richness and evenness with higher scores reflecting a greater diversity of chemical in cases where there is a more even distribution of compounds. The Simpson diversity index was calculated as $D = 1 - \sum_{i=1}^{S} p_i^2$ Where S is the total number of compounds/species and p_i is the proportional abundance of the i-th compound/species which is the chance that two randomly chosen compounds of a species are of different chemical entities. This index is more

sensitive to the prevalence of dominant compounds as compared to the one by Shannon. Evenness, which is the fairness of the number of compounds in the species, was estimated as $J = \frac{H'}{\ln S}$

where H^1 is the Shannon diversity index and S is richness. Evenness values close to one show that compounds are present in equal relative abundance and values close to zero show dominance by a few compounds.

Compound Abundance Heatmap

To represent the distribution of the most abundant volatile compounds among the five species, a heatmap was made using the top ten compounds as determined by the total sum of abundance of the compounds across all species. The z-score transformation of the data matrix of these ten compounds row-standardized the data whereby the abundance of each compound in all the species was transformed to the mean of zero and standard deviation of one. This standardization enabled straightforward comparison of relative abundance patterns among compounds that initially varied significantly in absolute concentration. A blue-white-red color gradient was used in the heatmap, where blue meant abundances less than the mean of the compound, white meant abundances equal to the mean, and red meant abundances greater than the mean. Hierarchical clustering of both rows (compounds) and columns (species) was done with Euclidean distance and complete linkage to show natural groupings based on similarity of abundance patterns. The values of numerical abundance were plotted on each cell of the heatmap to retain quantitative data, but still to visualize the relative differences.

Statistical Software and Implementation.

R version 4.5 was used to carry out all statistical analyses. Specific analytical components were used as follows: FactoMineR and factoextra to perform PCA and the related visualizations,

corrplot to visualize correlation matrices, pheatmap to create heatmaps, ggplot2 and ggrepel to create custom graphical outputs, and vegan to perform additional diversity calculations where necessary.

Biological Activity Prediction Using PASS Online

The selected major phytoconstituents identified by GC–MS analysis were further screened for their potential biological activities using the Prediction of Activity Spectra for Substances (PASS) online server (https://www.way2drug.com/passonline/). The SMILES structures retrieved from the PubChem database were submitted to the PASS platform to evaluate their probable pharmacological activities. The predicted activities were expressed as the probability of activity (Pa) and probability of inactivity (Pi). Activities with Pa values greater than Pi were considered biologically significant and were selected for further interpretation. PASS prediction was employed to estimate the potential antibacterial activity of the selected phytoconstituents as well as prediction of other relevant biological activities (5,8).

Results

The PCA score plot displays a clear pattern of multivariate differentiation of the five Hypericum species investigated, with the first two principal components (PCs) accounting for 68.2% of the total variance, thus defining the major pattern of variability in the dataset. There is a spatial segregation where strong horizontal segregation (PC1) between *H. kiboense* and *H. roeperianum* reflects large differences in dominant volatile compounds, as seen by the positioning of these two species on the opposite ends of the axis. *H. kiboense*, positioned in the upper-right quadrant, shows strong positive scores on both PC1 and PC2, and thus a highly specific compositional profile that is defined by variables contributing positively to both axes. Conversely, *H. roeperianum* located

in the upper-left quadrant, with high negative PC1 and positive PC2 scores, exhibits a different dominant composition, but some similarity along the vertical axis.

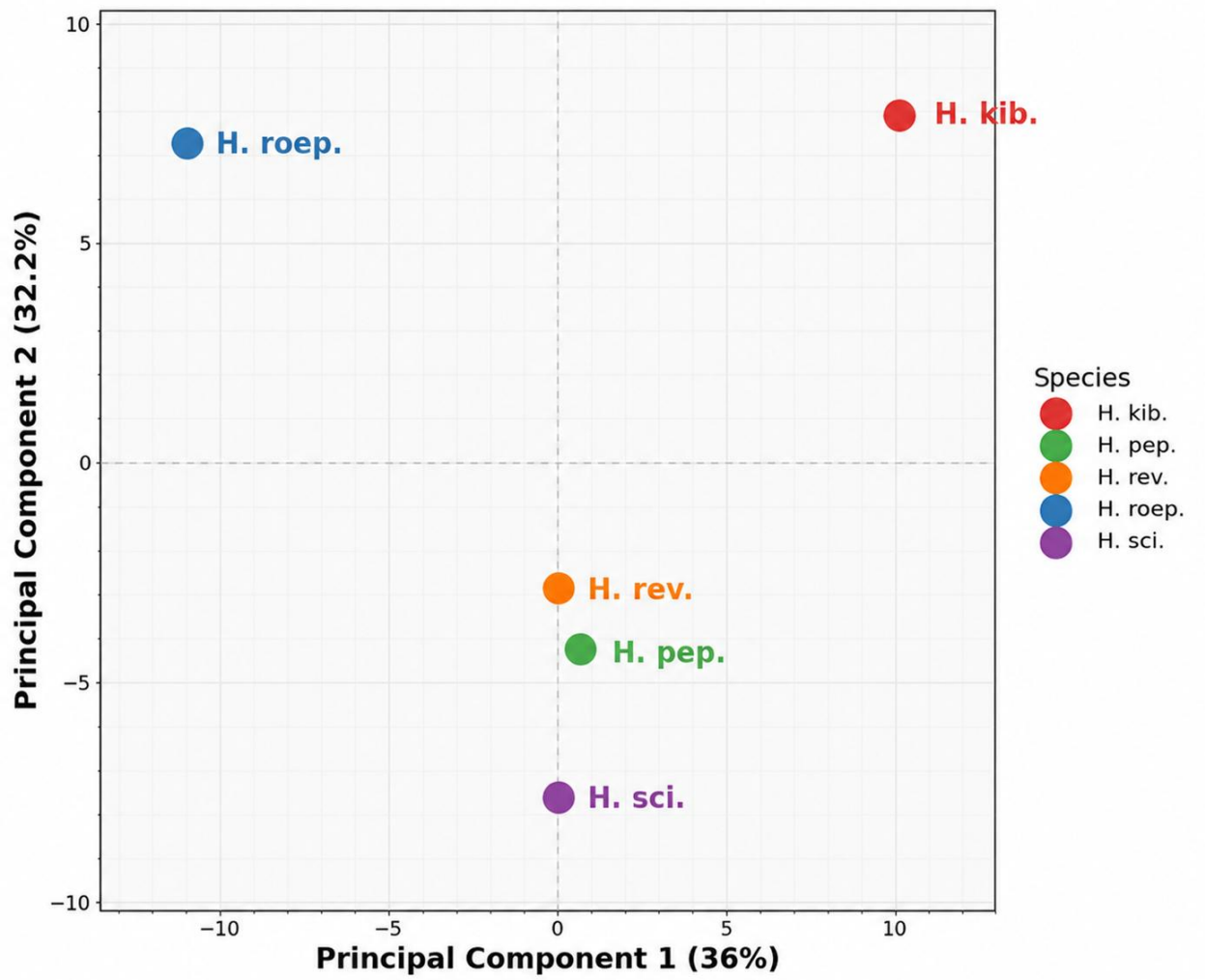


Figure 2. PCA score plot

Principal component analysis showing separation of the five Hypericum species based on their GC-MS volatile profiles. Points represent individual species based on GC-MS profiles. The plot highlights overall chemical differentiation among species, with some species forming distinct chemotypes and others showing closer compositional similarity.

On the vertical axis, vertical separation (PC2) sets apart *H. scioanum*, *H. peplidifolium* and *H. revolutum,* which are driven by variation in secondary or minor metabolites, with *H. scioanum*

showing the strongest negative score on PC2, indicating the greatest compositional difference along the secondary axis. *H. peplidifolium* and *H. revolutum,* both in the lower part of the plot, are separated mainly along PC2, suggesting that the variation between these two species is more related to secondary composition than the main gradient. Moreover, central versus peripheral shows that *H. revolutum* is chemically more intermediate and *H. kiboense* and *H. roeperianum* are more highly specialised, by the central location of *H. revolutum*, which indicates an intermediate or integrative chemical profile that shares traits with several other species. By contrast, the peripheral position of *H. kiboense* and *H. roeperianum* shows their large contributions to variability and their unique chemotypes. Overall, these spatial positions show significant variability and well-resolved compositional separations between the species.

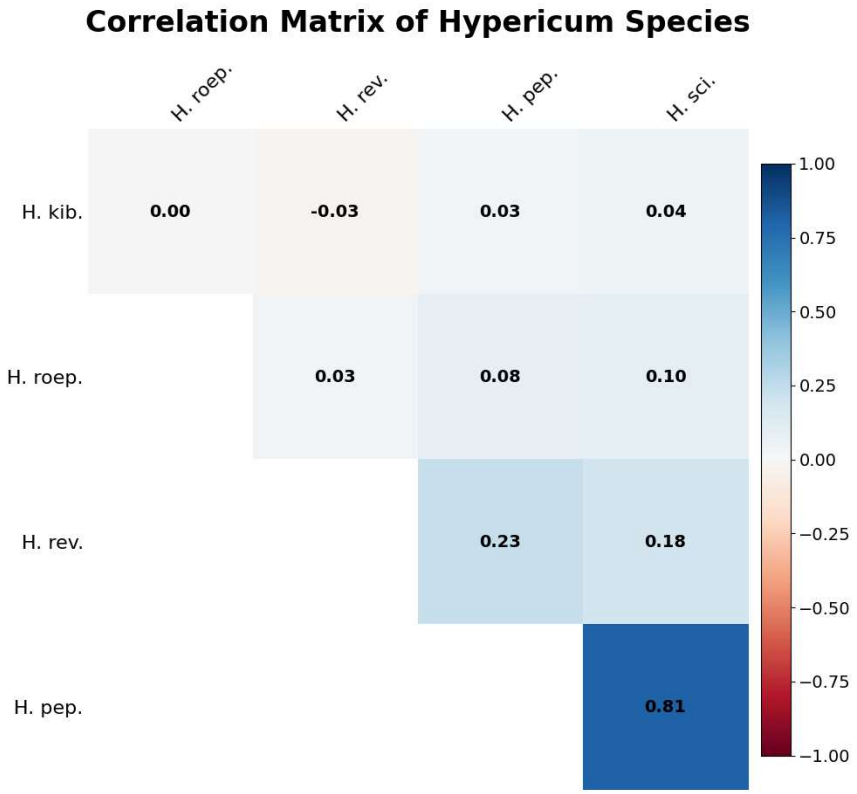


Figure 3: Pearson correlation matrix plot

Pearson correlation matrix showing pairwise similarity in volatile compound profiles among the five Hypericum species based on GC-MS relative abundance data. Correlation coefficients range from −1 to +1, with stronger positive values indicating greater chemical similarity. Most species pairs showed weak or near-zero correlations, reflecting strong interspecific chemical differentiation.

The correlation matrix shows that most of the Hypericum species are weakly correlated with each other, suggesting that their composition profiles are uncorrelated and chemically dissimilar. The majority of pairwise correlation coefficients are close to zero, such as between *H. kiboense* and *H. roeperianum* (0.00), *H. kiboense* and *H. revolutum* (-0.03), and *H. kiboense* with both *H. peplidifolium* (0.03) and *H. scioanum* (0.04). These low values indicate low shared variance among these species, reinforcing the view of *H. kiboense* being chemically distinct and not fitting well with the rest. The same can be said for *H. roeperianum* which shows weak positive correlations with the rest (0.03-0.10). This suggests some shared compositional traits but not to the extent of significant similarity. The slightly higher correlations with *H. revolutum* (e.g. with *H. peplidifolium* (0.23) and *H. scioanum* (0.18) indicate a moderate degree of similarity, suggesting that *H. revolutum* might be intermediate in composition, as shown in the PCA. The most notable element of the matrix is the high positive coefficient between *H. peplidifolium* and *H. scioanum* (0.81). This indicates a high degree of compositional similarity between these two species, suggesting that a large number of variables contribute in a similar way to both species. This association is distinct from the otherwise low correlations observed and suggests the possibility that *H. peplidifolium* and *H. scioanum* may be part of a chemotypic cluster. The matrix reveals a pattern where most of the species are poorly correlated, and thus chemically different, while the single strong correlation of *H. peplidifolium* and *H. scioanum* indicates that these species are

similar in a local sense, within the overall variability. This supports the existence of strong inter-specific differentiation, as well as selective grouping, and diversity and partial overlap in compositional characteristics.

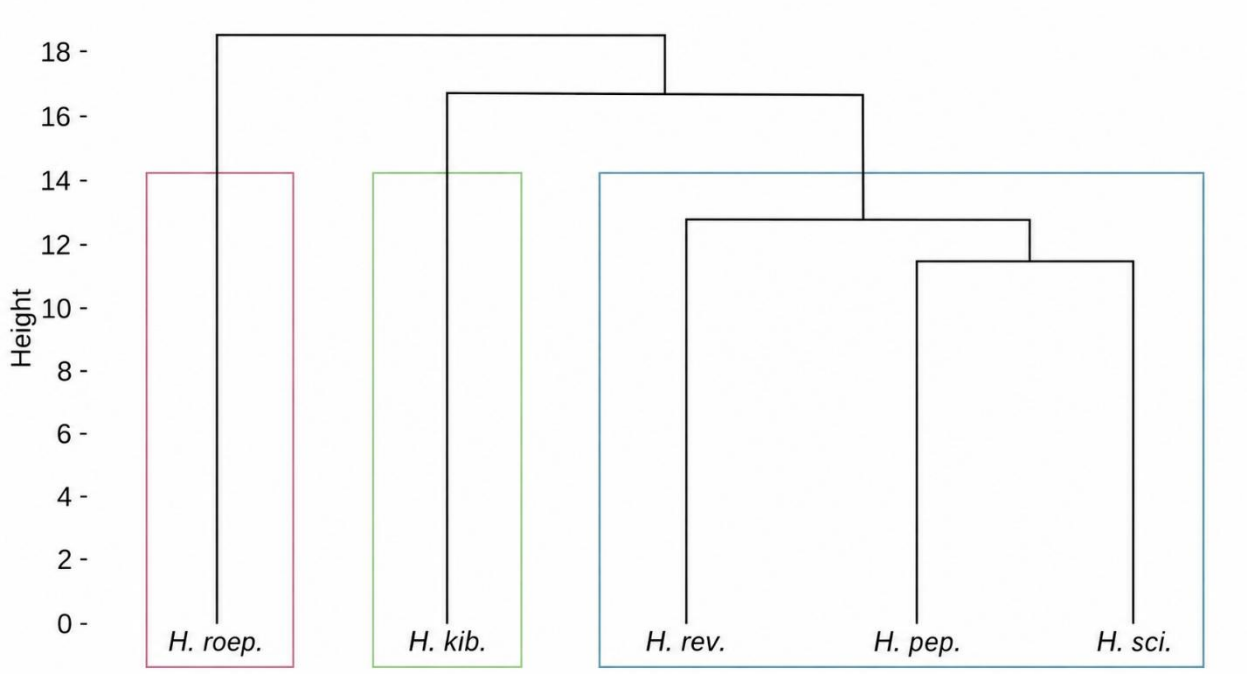


Figure 4. *Hierarchical clustering dendrogram showing three main groupings among the five Hypericum species. H. roperianum and H. kiboense appear as separate clusters, while H. revolutum, H. peplidifolium, and H. scioanum form a third cluster, with H. peplidifolium and H. scioanum showing the closest relationship.*

The hierarchical clustering of GC-MS volatile fingerprints shows a meaningful and evolutionarily significant pattern of chemical variation in the five Hypericum species. At the threshold dissimilarity level, three major chemotypic clusters are evident, reflecting significant compositional differences in their essential oils. *H. roeperianum* constitutes a single, early-branching cluster, with the highest linkage distance (i.e. the greatest botanical dissimilarity) to the

rest of the species, and suggests that this plant has a distinctive chemical profile, with either unique components or considerably different ratios of shared metabolites. *H. kiboense* also forms a distinct cluster at a lower level of dissimilarity, suggesting a more moderate but distinct chemotype. By contrast, *H. revolutum, H. peplidifolium,* and *H. scioanum* form a cluster, indicative of a relatively conserved volatile profile among these species. In particular, *H. peplidifolium* and *H. scioanum* show the closest relationship and merge first (at the lowest distance) implying strong compositional likeness and possibly shared biosynthetic pathways or environmental adaptations. *H. revolutum* merges in this cluster at a slightly higher distance, implying some degree of chemical overlap with some differences. Overall, the cluster analysis reveals that the chemical diversity in these species is not gradual, but discrete and species-specific, suggesting the presence of different chemotypes in Kenyan Hypericum species. These results are consistent with the study goal of resolving phytochemical relationships through multivariate analysis to provide a foundation for correlation studies and chemotaxonomy.

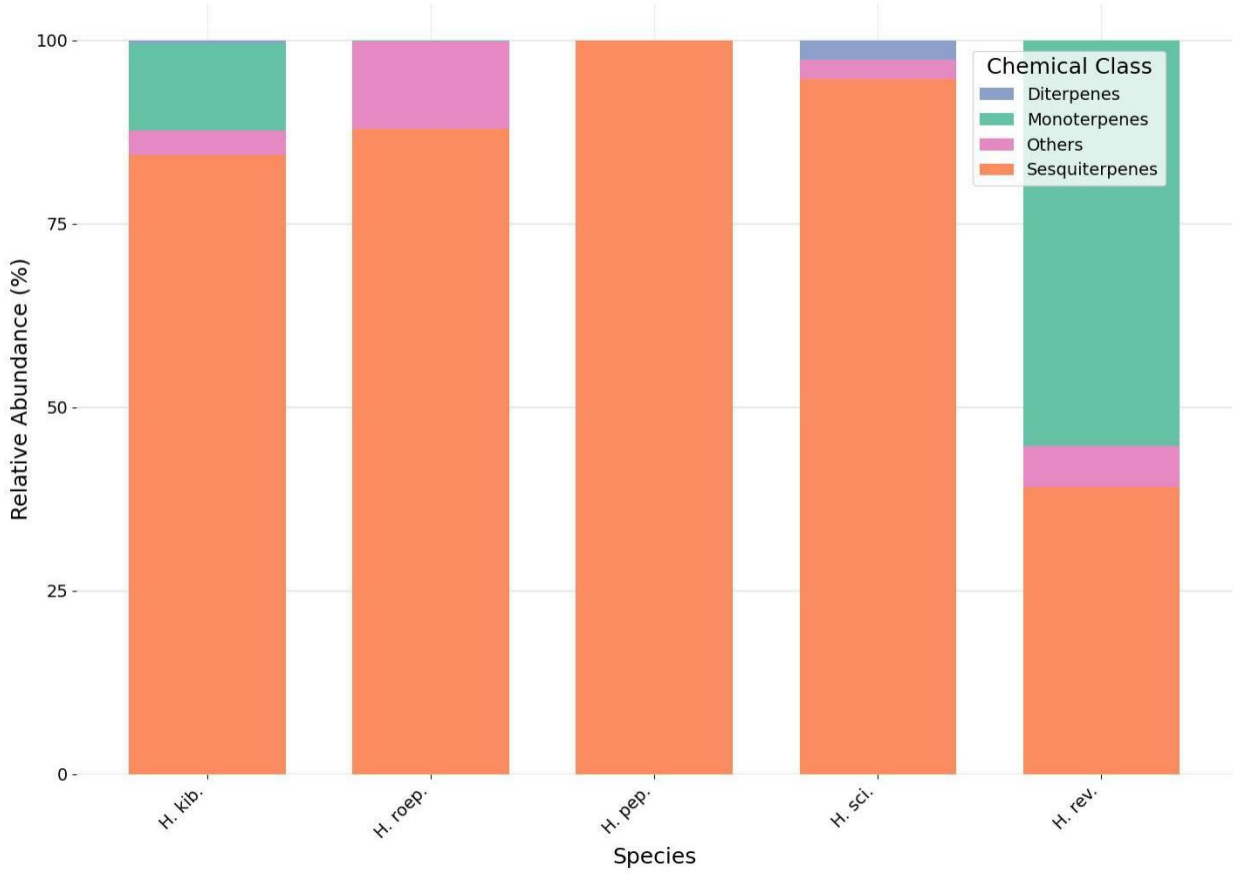


Figure 5. *Distribution of phytochemical classes across the five Hypericum species. Sesquiterpenes dominate H. kiboense, H. roeperianum, H. peplidifolium, and H. scioanum, while H. revolutum shows a higher proportion of monoterpenes compared with the other species.*

The profile of phytochemical classes among the five Hypericum species exhibits a highly unbalanced composition dominated by sesquiterpenes, but with considerable variation among the species. *H. peplidifolium* is the only species with a highly skewed composition, solely comprised of sesquiterpenes, suggesting a highly focused metabolic output and tight regulation of a single pathway. A similarly high but not exclusive dominance is evident in *H. scioanum*, with almost complete representation by sesquiterpenes and only minor representation by other classes, showing low, though present, metabolic diversification. *H. kiboense* and *H. roeperianum* also exhibit sesquiterpene dominance; however, both species also contain minor amounts of monoterpenes and

other compounds, suggesting a more extensive, albeit sesquiterpene-dominated, metabolic profile. Conversely, *H. revolutum* exhibits a different pattern, with a prominent contribution of monoterpenes along with a low share of sesquiterpenes, revealing the most distinct profile in terms of class-level composition. The minor presence of diterpenes and other (minor) classes in some species also adds to the interspecific diversity, without dominating the overall pattern. In summary, the above patterns show that while sesquiterpenes form the backbone of chemical composition across the genus, the presence of other, less dominant classes introduces variation and accordingly provides evidence for discrete chemotypes within the genus, as well as reinforcing the multivariate nature of the species' chemical profiles that have been revealed by the cluster analysis.

Species	Richness	Shannon	Simpson	Evenness
H. kib.	50	2.964	0.891	0.758
H. roep.	49	3.354	0.952	0.862
H. pep.	21	2.680	0.917	0.880
H. sci.	33	3.121	0.944	0.893
H. rev.	29	2.047	0.749	0.608

Figure 6. *Chemical diversity indices of volatile constituents in the five Hypericum species. H. kiboense shows the highest richness, H. roeperianum has the highest Shannon and Simpson diversity values, and H. scioanum shows the highest evenness, while H. revolutum records the lowest diversity and evenness values.*

Diversity indices indicate significant inter-specific differences in the diversity and balance of volatile constituents in the Hypericum species. The highest species richness is found in *H. kiboense* (50) and *H. roeperianum* (49), which corresponds to a wide range of detected constituents, while low richness (21) is observed in *H. peplidifolium*, which is consistent with a chemically less diverse

profile; *H. scioanum* (33) and *H. revolutum* (29) fall between these extremes. Taking into account both the number of constituents and their proportions, *H. roeperianum* shows the highest Shannon diversity (3.354) and Simpson index (0.952), indicating a highly diverse, balanced profile. *H. scioanum* is also highly diverse, with a Shannon index of 3.121 and Simpson index of 0.944, despite having moderate richness, implying an even distribution of its constituents. By comparison, *H. revolutum* has the lowest Shannon index (2.047) and Simpson value (0.749), suggesting low diversity and uneven distribution of constituents. *H. kiboense*, while the richest, surprisingly exhibits a lower Shannon index (2.964) and Simpson value (0.891), which suggests that the high number of compounds are not uniformly distributed. Interestingly, despite having the lowest species richness (21), *H. peplidifolium* has a Shannon index of 2.680 and a Simpson value of 0.917, suggesting that its few constituents have a relatively even distribution. This is in line with the evenness values, with the most balanced distributions observed in *H. scioanum* (0.893) and *H. peplidifolium* (0.880), followed by *H. roeperianum* (0.862) and *H. kiboense* (0.758), while *H. revolutum* displays the lowest evenness (0.608), confirming a heavily skewed composition. Overall, such quantitative patterns reveal that the chemotypic diversity of the species is controlled by the richness and distribution of their constituents, confirming the ordered variability among species in the cluster analysis and phytochemical class distribution analysis.

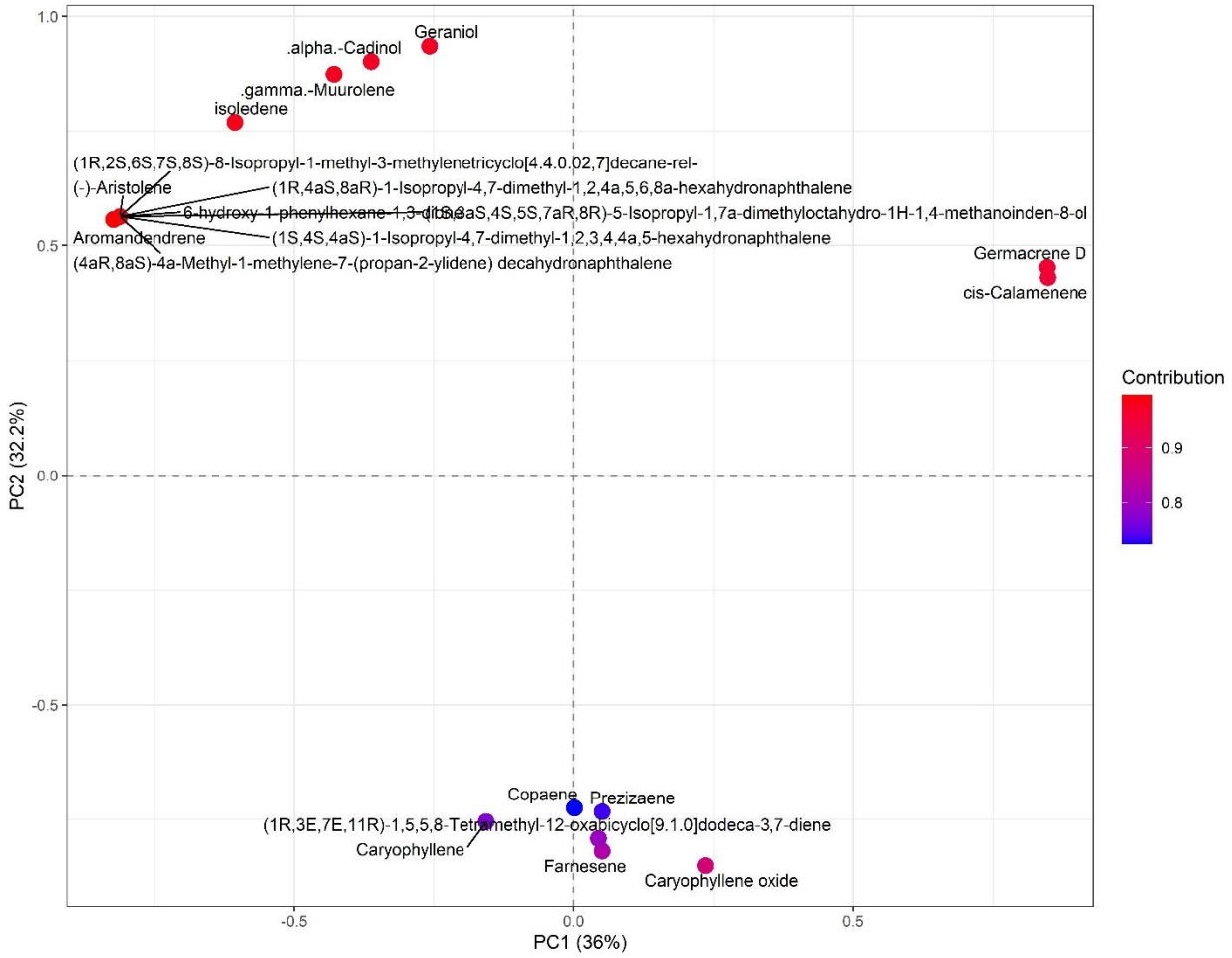


Figure 7. *PCA loadings plot showing the volatile compounds contributing most to species separation. Germacrene D and cis-calamenene load strongly on positive PC1, while geraniol, α-cadinol, γ-muurolene, isolongifolene, aromadendrene, caryophyllene, farnesene, and caryophyllene oxide contribute mainly to separation along PC2.*

The PCA loading plot of the Hypericum species illustrates the compounds that are responsible for the discrimination between species and show distinct patterns along both the principal components. The variability observed in the four quadrants suggests that species separation is not uniformly due to all compounds, but rather to specific groups of compounds. On the positive axis of PC1, compounds such as Germacrene D and cis-Calamenene are clearly highlighted, revealing that these compounds play a significant role in this direction of the separation. Their comparatively high loadings indicate that species positioned along the positive end of PC1 are closely linked to these components, which serve as compositional signatures. On the other hand, species along the

negative side of PC1 are associated with compounds like aromadendrene, aristolene, and several related sesquiterpenoid compounds, suggesting a distinct composition for species located on this side of the plot. This distinct separation along PC1 highlights the change in primary drivers of variability from one group of compounds to another.

PC2 provides a secondary source of variability, with compounds including geraniol, α-cadinol, γ-muurolene, and isolongifolene being positioned in the upper part of the plot. These compounds have high positive loadings on PC2, implying that species with positive PC2 scores are rich in these compounds. By contrast, the lower part of the plot is populated by compounds such as caryophyllene oxide, farnesene, copaene and caryophyllene, which contribute negatively to PC2. This vertical split suggests that secondary differentiation of species is due to a different set of compounds independent of those driving differentiation on PC1. The grouping of multiple compounds in the same regions of the plot implies that several compounds vary together, probably due to shared biosynthetic pathways or co-occurrence. For instance, the clustering of oxygenated sesquiterpenes in the upper quadrant suggests that these compounds work together to define a compositional space that describes species in that direction, while the grouping of hydrocarbons and their derivatives in the lower quadrant represents a different compositional pattern. When combined with the score plot, the loadings plot suggests that species are separated based on chemical signatures. For example, species in the upper-right quadrant are likely to be affected by compounds such as Germacrene D, in conjunction with the compounds loading positively on PC2, while those in the upper-left quadrant are more likely to be affected by compounds that load negatively on PC1, but still affected by compounds loading positively on PC2. Likewise, species in the lower quadrants are related to compounds with negative influence on PC2, such as caryophyllene oxide and farnesene. The loadings plot shows that the multidimensional

discrimination of species is structured by a well-defined group of compounds, with PC1 representing primary compositional differences and PC2 representing secondary compositional differences. This reveals the existence of specific compositional signatures in species separation and demonstrates the efficacy of the multivariate method in detecting these signatures.

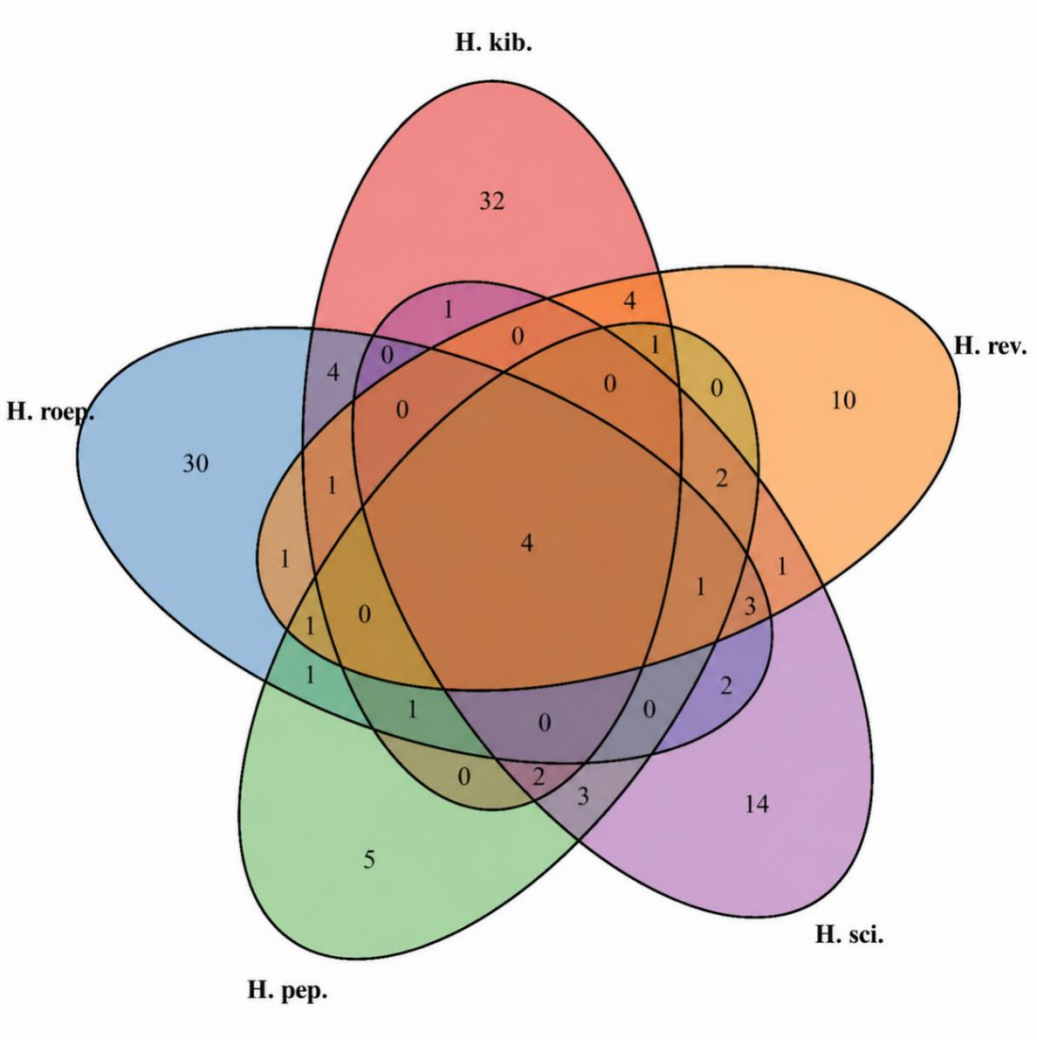


Figure 8. *Venn diagram showing shared and unique volatile compounds among the five Hypericum species. H. kiboense and H. roeperianum contain the highest numbers of unique compounds, while four compounds are shared across all five species.*

The Venn diagram shows a complex distribution of compounds among the five Hypericum species, with strong uniqueness and selective sharing. The most prominent aspect of the diagram

is the high number of unique compounds for each species, especially in *H. kiboense* (32) and *H. roeperianum* (30), implying high chemical individuality and suggesting that these species have highly specialized compositions. *H. scioanum* is also remarkable for its number of unique compounds (14), whereas *H. revolutum* (10) and *H. peplidifolium* (5) have fewer unique compounds, suggesting a higher level of compositional overlap or lower level of specialization. Despite high individuality, the diagram's centre illustrates that a relatively small number of compounds (4) are present in all five species, indicating a shared core of chemicals that may reflect shared biosynthetic pathways within the genus. But the small number of compounds common to all species relative to the high number of unique ones highlights the degree of differentiation.

Pairwise and partial overlaps add detail to this story. Modest overlap between some species pairs, like *H. kiboense* with *H. revolutum* and *H. scioanum*, and between *H. revolutum*, *H. peplidifolium*, and H. *scioanum* suggests localised overlap in composition. Conversely, sharing between *H. roeperianum* and other species appears to be sparse and weak, further emphasizing its uniqueness, which is supported by its high number of unique compounds. The low values of overlaps (one to four compounds) indicate that chemical overlaps are generally minimal and not shared among large groups of species. The uneven distribution of overlaps also reflects varying relationships between the species. H. revolutum seems to play a central role, connecting with several other species, while *H. kiboense* and *H. roeperianum* are more distinct. In contrast, the increased overlap between *H. peplidifolium* and *H. scioanum* is in agreement with their closer resemblance in other multivariate biplots, and suggests a degree of compositional overlap between these two species. In conclusion, the diagram reveals a richly diverse system in which species retain distinct identities while sharing a minor core. This unique and shared compositional nature is symptomatic of both

shared metabolic profiles and species-specific adaptations, and helps confirm the presence of distinct (albeit partly overlapping) compositional profiles in the group.

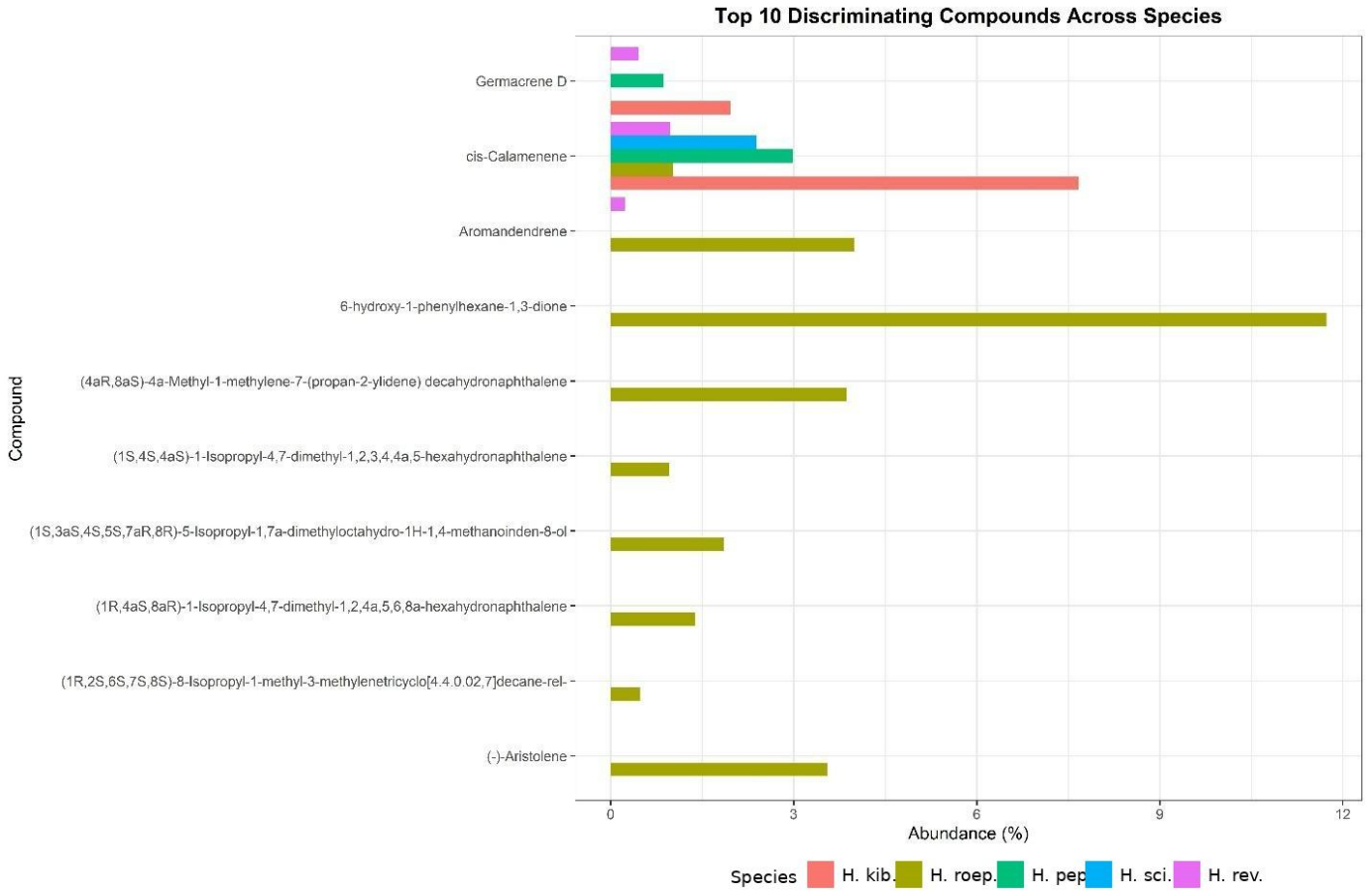


Figure 9. *Bar plot showing the top ten discriminating volatile compounds across the five Hypericum species. H. kiboense shows the highest cis-calamenene abundance, while H. roeperianum is distinguished by high levels of 6-hydroxy-1-phenylhexane-1,3-dione and related sesquiterpenoid compounds.*

The bar plot of the top ten discriminating compounds illustrates the distribution of key compounds among the five Hypericum species, and offers a foundation for interpreting the factors responsible for species differentiation. One notable aspect of the plot is the strong dominance of some compounds in certain species, confirming the presence of chemical specialization. A key feature of *H. kiboense* is the remarkably elevated level of cis-calamenene, one of the largest peaks in the

data. This high abundance, along with moderate abundances of Germacrene D, suggests that a limited number of compounds are responsible for its compositional makeup. This is suggestive of a specialization strategy where a relatively small number of major constituents are involved in discriminating among species.

In contrast, *H. roeperianum* has a more even distribution of discriminating compounds, with moderate to high levels of several constituents. In particular, 6-hydroxy-1-phenylhexane-1,3-dione is present at a significantly higher abundance than any other compound represented in the chart, and thus is a species-specific marker. Furthermore, the presence of compounds like aromadendrene, aristolene and several structurally complex sesquiterpenoid derivatives in combination suggests a more diverse and complex, but still unique, chemical composition. This diversity of contributors is consistent with a compositional structure in which differentiation is achieved by a combination of metabolites. *H. peplidifolium* exhibits moderate levels of compounds like cis-calamenene and Germacrene D, but no single compound is predominant. This implies a more integrated compositional pattern, in which multiple compounds play a small role. This is also the case in *H. scioanum* (although with a few more cis-calamenene), suggesting some shared compositional characteristics in the two species. In contrast, *H. revolutum* has relatively lower abundances of most of the top discriminating variables, with only minor contributions from Germacrene D and cis-calamenene. This is consistent with its classification as chemically intermediate, with no strong expression of discriminating variables and a more even distribution of compounds. In conclusion, the chart shows that variability between species is due to the presence of highly abundant, species-specific compounds as well as to associations of moderately abundant compounds. Some species (*H. kiboense* and *H. roeperianum*) are strongly characterised by their markers, while others (*H. peplidifolium* and *H. scioanum*) display some degree of overlap,

with *H. revolutum* displaying a more diffuse pattern. This supports the overall pattern of distinct but partially overlapping compositional signatures of the different species.

Predicted biological activities

Figure 10 show the predicted biological spectra of activities ranked in ascending order by the number of occurrences across all quantified constituent-activity pairs. For activities with the same number of occurrences, the weighted mean Pa is used to arrange activities from higher to lower predicted probability. The predicted biological activity profile of the investigated Hypericum essential oils shows a broad and chemically meaningful spectrum of activities across the major quantified constituents. By presenting each activity as reported in the source table, the analysis preserves important distinctions between closely related terms while still allowing comparison across the five oils. This is particularly important for activities such as antieczematic and dermatologic, which are related to skin biology but should not be treated as identical outcomes.

Antieczematic activity was the most recurrent predicted activity, with 26 occurrences across all five species and the highest total linked abundance. This broad representation suggests that antieczematic potential is not restricted to one oil or to one isolated constituent, but appears repeatedly across the chemical profiles. The repeated occurrence of this activity may be relevant when considering the traditional and pharmacological interest of Hypericum species in skin-related applications. However, this interpretation should be viewed as a prediction-based indication rather than direct experimental proof.

Antineoplastic activity was also strongly represented, with 20 occurrences across all five species and a high weighted mean Pa. This indicates that several major constituents may contribute to predicted anticancer-related activity. The presence of apoptosis agonist activity in 17 occurrences

further supports a possible relationship between the chemical composition of these oils and predicted mechanisms linked to cell-death regulation. Taken together, the antineoplastic and apoptosis-related signals provide a useful basis for prioritizing constituents for more focused biological testing.

Several enzyme-related activities had fewer occurrences but relatively high weighted mean Pa values. For example, spasmolytic, urinary activity had the highest weighted mean Pa, but it was observed in only two constituent-activity pairs and represented in one species. Similarly, mucomembranous protector, prenyl-diphosphatase inhibitor, fatty-acyl-CoA synthase inhibitor and catechol 2,3-dioxygenase inhibitor activities showed high predicted probabilities but had limited distribution across species. These findings suggest that some activities may be highly predicted but chemically narrow, whereas others are more widely distributed across the investigated oils.

The species represented column adds an important layer of interpretation. Activities such as antieczematic, antineoplastic, apoptosis agonist, carminative, CYP2C12 substrate and testosterone 17beta-dehydrogenase (NADP+) inhibitor were represented across all five species, indicating a broad cross-species distribution. In contrast, activities represented in only one species may still be important, but their interpretation should be more cautious because they are linked to fewer constituents and may be more species-specific.

The total linked abundance values should also be interpreted carefully. Activities such as antieczematic, antineoplastic and apoptosis agonist had high total linked abundance values because they were associated with several major constituents. Since each constituent can contribute to more than one predicted activity, total linked abundance is not intended to represent the percentage

composition of the oil. Instead, it indicates how strongly an activity is connected to the quantified chemical profile through the abundances of the constituents carrying that activity.

Overall, the PASS-based approach provides a useful first-stage screening framework for organizing complex essential oil data into biologically interpretable patterns. Its strength lies in prioritizing activities and constituents for further investigation. Nevertheless, the results should be interpreted as computational predictions. Experimental validation using suitable in *vitro* and *in vivo* models remains necessary to confirm the biological relevance, potency, safety and mechanism of action of the identified constituents.

Ranking of biological activities across *Hypericum* essential oils

Ascending order by number of occurrences

Rank	Biological activity	n occurrences	Weighted mean Pa	Total linked abundance (%)	Species represented
1	Fatty-acyl-CoA synthase inhibitor	1	0.931	3.42	1
2	Catechol 2,3-dioxygenase inhibitor	1	0.918	11.73	1
3	Alkylglycerone-phosphate synthase inhibitor	1	0.916	11.73	1
4	Monodehydroascorbate reductase (NADH) inhibitor	1	0.912	11.73	1
5	Myc inhibitor	1	0.904	3.02	1
6	Transcription factor NF kappa B stimulant	1	0.872	3.96	1
7	Ubiquinol-cytochrome-c reductase inhibitor	1	0.855	7.67	1
8	Bone diseases treatment	1	0.846	3.87	1
9	Alpha-pinene-oxide decyclase inhibitor	1	0.810	4.45	1
10	Estrogen agonist	1	0.805	3.55	1
11	CYP2H substrate	1	0.757	3.61	1
12	Spasmolytic, urinary	2	0.977	6.93	1
13	Mucomembranous protector	2	0.952	19.03	2
14	Prenyl-diphosphatase inhibitor	2	0.937	19.03	2
15	Undecaprenyl-phosphate mannosyltransferase inhibitor	2	0.913	19.03	2
16	Aspulvinone dimethylallyltransferase inhibitor	2	0.906	7.38	1
17	Alkenylglycerophosphocholine hydrolase inhibitor	2	0.899	6.69	2
18	Acylcarnitine hydrolase inhibitor	2	0.895	9.38	2
19	Alkylacetylglycerophosphatase inhibitor	2	0.886	6.69	2
20	Antiinflammatory	2	0.823	14.38	2
21	Retinol dehydrogenase inhibitor	3	0.916	13.79	3
22	Glutamate-5-semialdehyde dehydrogenase inhibitor	3	0.908	21.73	3
23	CYP2J substrate	3	0.844	50.36	2
24	Dermatologic	3	0.775	19.77	2
25	HIF1A expression inhibitor	4	0.810	28.84	4
26	Testosterone 17beta-dehydrogenase (NADP+) inhibitor	6	0.839	68.39	5
27	CYP2C12 substrate	7	0.870	80.61	5
28	Carminative	11	0.867	61.58	5
29	Apoptosis agonist	17	0.854	131.55	5
30	Antineoplastic	20	0.910	149.10	5
31	Antieczematic	26	0.882	172.12	5

Figure 10 *ranks the predicted biological activities in ascending order according to the number of constituent–activity occurrences. The weighted mean Pa was calculated as Σ(abundance × Pa) /*

Σ(abundance), while the species represented column indicates the number of Hypericum *species in which each activity was observed among the reported major constituents.*

DISCUSSION

The current study gives a detailed chemometric explanation of the volatile chemical variability in the essential oils of five Kenyan Hypericum species: *H. kiboense*, *H. peplidifolium*, *H. scioanum*, *H. roperianum*, and *H. revolutum*. The results suggest that the volatile profiles of such species are not randomly distributed but they are organized into specific, biologically meaningful chemotypic patterns. This matters since essential oils are intricate mixtures whose taxonomic and biological significance cannot be completely comprehended in terms of discrete lists of compounds. Rather, the overall actions of dominant constituents, minor metabolites, chemical classes, and inter-compound associations can be a better reliable basis of assessing chemical variability. This interpretation is in line with the more general interpretation of Hypericum essential oils, where the marked interspecies variation, chemical polymorphism, and the formation of chemotypes has been consistently reported across different species and geographic locations (30,31). One of the strengths of the current analysis is that it considers the GC-MS dataset as a multivariate chemical fingerprint instead of being a basic inventory of detected compounds. This difference is crucial to the essential oil research since the presence of major constituents can overshadow the chromatographic profile, and the minor and moderate-abundance compounds can still play a significant role in species discrimination. The analysis of autoscaled information, principal component analysis, hierarchical clustering, Pearson correlation analysis, diversity indices and chemical class profiling thus offers a more robust analytical framework than that of relying on percentage composition alone. The recent Clausena study of (32) with GC-MS/GC-FID data and PCA, HCA, and correlation analysis, is also supportive of this approach since the chemical

variability between five species has been resolved and the marker compounds that cause the separation of the species have been identified. In that research PCA and HCA were not only used as statistical tools but also as chemotaxonomic tools to classify essential oils based on similarities and differences in volatile composition (32). The distinct separation of the five Kenyan Hypericum species into the multivariate space is what corroborates the finding that each of the species has a recognizable volatile signature. The separation must be seen as a manifestation of chemical phenotype differentiation and not as a purely numerical result of ordination. The reason behind this is that the patterns observed are supported in many independent analytical products such as, PCA, correlation analysis, hierarchical clustering, chemical class distribution, diversity indices, and discriminating pattern of compounds. Concordance between these methods minimizes the chance that the observed grouping is due to dimensionality reduction or to an isolated outlier Compounds. Similar built-in chemometric logic to demonstrate that groupings of Clausena essential oils derived by PCA were confirmed by hierarchical clustering and correlation analysis (32), thus increasing the strength of the inferred chemotaxonomic relationships. Such methodological consistency justifies the adoption of multivariate correlation analysis as a suitable measure of interpreting chemical variability in essential oils.

The chemotaxonomic importance of the differentiation between *H. kiboense* and *H. roperianum* in particular is of significance. These species seem to exemplify chemically-specific profiles, indicating that the volatile composition of these species is determined by the unique biosynthetic regulation. Specialization of chemicals may be observed when a single compound or a small number of compounds are very high relative abundance, or when multiple related compounds may be collectively defined as a species-specific profile. Both scenarios are informative in essential oil chemotaxonomy. A single dominant marker can give quick species recognition; a coordinated set

of moderately abundant compounds can give a more stable and robust fingerprint. The *H. kiboense* and *H. roeperianum* as different volatile chemotypes is in line with the recent work on Hypericum, where other species have been demonstrated to express vastly different volatile profiles, including variation in sesquiterpene hydrocarbons, oxygenated sesquiterpenes, monoterpenes, alkanes, and other specialized metabolites (30,31). The close relationship between *H. peplidifolium* and *H. scioanum* is to be taken into consideration. Their chemical closeness suggests that the two species share or converge to a common volatile organization, but does not imply that the two species are chemically identical. In essential oil research, the species can clump together due to the overlap in the predominant chemical group or the similarity in its biosynthetic focus, but still differing in individual marker compounds, ratio of abundances, or minor components. Hence, the interpretation which is most defensible is that the *H. peplidifolium* and *H. scioanum* are related volatile subgroup within Kenyan Hypericum dataset. Such a relationship can be due to common genetic lineage, similar ecological selection pressures, similar regulation of enzymes in the biosynthesis of sesquiterpene, or a combination of these. The same logic were observed in the Clausena study where *C. austroindica* and *C. heptaphylla* were close together due to the same profiles of trans-anethole and, the two species were separated by other species based on the different marker compounds (32). In such a way, the interrelation between *H. peplidifolium* and *H. scioanum* can be defined as chemotypic affinity as opposed to complete chemical equivalence.

The abundance of sesquiterpenes in most of the Kenyan Hypericum oils can offer a viable biochemical explanation to the observed groupings. Sesquiterpenes hydrocarbons and oxygenated sesquiterpenes are commonly reported in Hypericum essential oils and often play a role in species-level differentiation. Individual compounds, including Germacrene D, 2-caryophyllene, caryophyllene oxide, spathulenol, 2-muurolene, and 2-alpha-cadinol have been repeatedly named

as important constituents in the genus (31). Thus, the prevalence of the chemistry of sesquiterpene in the current data is in line with the known volatile chemistry of Hypericum. Simultaneously, the clear-cut behavior of *H. revolutum* especially where its profile at the class level is at variance with the more sesquiterpene dominated species makes it clear that the chemical differentiation is not explained by the sesquiterpenes in isolation. The separation of species might also be a result of differences in monoterpenes, oxygenated derivatives and other minor volatile classes. This is in line with recent studies of Hypericum that show that the composition of essential oils can vary substantially between species, with some oils predominantly containing sesquiterpenes and others having a stronger contribution of monoterpenes, alkanes, or oxygenated compounds (30,33) .

The diversity indices also assist in further interpretation of the chemical richness and chemical evenness as different facets of essential oil variability. A species with significant numbers of compounds detected is by no means the most chemically balanced, as its oil can still be dominated by a few major constituents. On the other hand, evenness may be high in a species with fewer detected compounds, provided the compounds are not present in very different proportions. This difference is significant in that richness is used to describe the breadth of biosynthetic output, whereas evenness indicates the spread of that output evenly. Chemotaxonomic viewpoint High richness can be proposed to suggest a more balanced and possibly stable volatile profile, whereas high evenness can be proposed to indicate a wider range of metabolic capability. Biologically speaking evenness can also affect ecological activity, since the action of an essential oil may not lie in the action of a single constituent. This understanding is consistent with the essential oil literature that highlights that biological activity is often correlated with the overall chemical composition, the structure of key constituents, functional groups, and potential synergistic interactions among components (31,32).

The analysis revealed the discriminating compounds, which are chemically supporting the multivariate separation. Chemically plausible markers include constituents such as Germacrene D, cis-calamenene, caryophyllene oxide, farnesene, copaene, caryophyllene, α-cadinol, gamma-muurolene and related sesquiterpenoid compounds since some of these compounds or structural classes have been reported in Hypericum oils. The fact that they have contributed to the process of separation of species implies that the statistical groupings are not abstract mathematical clusters but are based on recognizable phytochemical differences. In this regard, the current results are similar to the Clausena study, where the specific compounds that are identified as the major contributors to the segregation of species in both the PCA and the clustering analysis include myristicin, sabinene, terpinolene, biofloratriene, and trans-anethole (32). The implication to the current Hypericum study is that compounds with high loadings or strong discriminatory behavior may be used as candidate chemical markers in future authentication, quality control or comparative chemotaxonomic studies. It has additional implications in the coexistence of unique and common compounds. The few common constituents imply that the five species share a common biosynthetic base in agreement with their grouping in a common genus. The multiplicity of species-specific compounds, however, suggests that there is a great deal of diversification beyond that common chemical core. This trend is common to plant secondary metabolism, in which the conserved metabolic pathways produce common chemical frameworks, whereas the species-specific expression of enzymes, branching of metabolic pathways, environmental selection and developmental regulation generate the chemical diversity. Thus, the current findings can be considered as the sign of the conservation and divergence. The common compounds indicate genus-level metabolic relatedness, and the unique compounds and abundance differences indicate species-level chemical identity. This balance is the center of chemotaxonomy since it explains how

related taxa can still be chemically recognizable as belonging to the same genus, and yet have distinct essential oil fingerprints. The findings also have consequences in terms of quality control and subsequent use of Kenyan Hypericum essential oils. Chemical consistency is important in medicinal and aromatic plants since a variation in volatile composition may influence aroma, biological activity, safety, and reproducibility. The discovery of individual chemotypes implies that these species are not chemically interchangeable. To illustrate, a species that is dominated by sesquiterpenes may be markedly different in potential biological action than a species which is dominated by monoterpenes or a mixed volatile profile. In the same way, species with a limited number of marker compounds may provide more straightforward targets to standardization, whereas chemically diverse species may require a broader fingerprint-based quality control. The same argument was made by (32) in the case of Clausena, where chemotype-based chemotaxonomy is found useful in quality assessment, resource development, and medicinal research. Using the same logic here, the Kenyan Hypericum species must be considered as individuals, not as a genus. Whereas the volatile profiles are indicative of possible relevance to the biological activity, the discussion should be approached in a suitably cautious manner. The literature has linked hypericum essential oils to various biological activities, such as antimicrobial, antioxidant, anti-inflammatory, and other pharmacological actions (31,33). But even the current chemical analysis on its own cannot prove biological activity. The biological effects are dependent on the concentration of the compound, stereochemistry, the method used in the extraction process, assay system, synergistic and antagonistic interactions and on the organism or biological endpoint on which the test is being conducted. As such the discriminating compounds reported herein should be read as a possible future bioactivity-guided target and not as a direct indication of

pharmacological efficacy. This conservative reading plays a critical role against overstatement and in ensuring that a distinction is made between chemical profiling and biological validation.

The identified variability can also be due to ecological and geographical factors besides genetic variation at the species level. The composition of essential oils can be influenced by the altitude, the type of soil, rainfall, temperature, season, phenological stage, part of the plant, conditions of extraction, and post-harvest treatment. Thus, although the current analysis gives solid evidence of chemical differentiation between the five species, future research should determine how stable these chemotypes are across multiple populations, collection seasons, and ecological zones. This is particularly applicable to the Kenyan Hypericum species since the African species are still underrepresented in the essential oil literature in comparison to the European and Asian species. A larger sample would also help in identifying whether the profiles that are observed are stable species-level markers or whether they contain population-specific and environment-dependent variation. However, the fact that the current multivariate analyses are consistent gives a good basis to consider the identified profiles as meaningful chemical signatures. This study demonstrates that the essential oils of *H. kiboense*, *H. peplidifolium*, *H. scioanum*, *H. roperianum* and *H. revolutum* exhibit organized chemical variability with well-defined chemotaxonomic value. The data affirm the presence of discrete yet partially overlapping volatile chemotypes, with species differentiation being driven by the dominant marker compounds, as well as coordinated variation between broader chemical classes. The findings are also in line with the recent literature on the multivariate correlation framework, which has been successfully applied in the recent Clausena essential oil study (32). This study contributes new regional knowledge on Kenyan *Hypericum* essential oils and provides a defensible basis for future work in chemotaxonomy, quality control, ecological chemistry, and bioactivity-guided screening.

Conclusion and Recommendation

Conclusion: The GC-MS profiles and multivariate analyses show that the five Kenyan Hypericum species have organized, species-specific volatile fingerprints. Although the oils share a small common chemical core, the large number of unique constituents and the uneven distribution of dominant compounds reveal clear chemotypic differentiation. *H. kiboense* and *H. roeperianum* appear as the most chemically distinctive species, while *H. peplidifolium* and *H. scioanum* show a close chemotypic affinity. *H. revolutum* occupies a more intermediate position in the multivariate analyses but remains distinctive because of its stronger monoterpene contribution. The agreement between PCA, Pearson correlation analysis, hierarchical clustering, chemical class profiling, diversity indices, Venn analysis and discriminating-compound patterns strengthens the interpretation that these are meaningful chemical signatures rather than isolated statistical patterns. Overall, the study adds important Kenyan data to the phytochemistry of Hypericum and supports the use of volatile-oil fingerprints in chemotaxonomy, authentication and quality control.

The investigated Hypericum essential oils showed a diverse predicted biological activity profile. Antieczematic activity was the most recurrent prediction and was represented across all five species, suggesting a broad skin-related bioactivity signal among the quantified constituents. Antineoplastic and apoptosis agonist activities were also widely represented, indicating that several constituents may contribute to predicted anticancer-related and cell-death-associated biological effects.

Separating antieczematic from dermatologic activity improved the accuracy of the interpretation because the two terms do not represent the same prediction. Antieczematic activity was frequent and widely distributed, whereas dermatologic activity appeared less often and was represented in fewer species. This distinction strengthens the reliability of the activity profile and avoids overgeneralizing the skin-related predictions.

The findings support the value of the studied Hypericum essential oils as sources of candidate constituents for further pharmacological screening. Activities with high recurrence and broad species representation may be useful for identifying shared bioactivity patterns, while activities with high weighted Pa but low occurrence may point to more specific constituents of interest. Further experimental validation is required before these predicted activities can be translated into biological or therapeutic claims.

Recommendation: Future work should test the stability of these chemotypes using larger collections from different populations, seasons, altitudes and ecological zones. The main marker compounds should also be confirmed using authentic reference standards and quantitative validation so that they can support reproducible quality-control protocols. Biological activity, toxicity and mechanism-oriented assays should be conducted cautiously to determine whether the discriminating compounds or whole-oil fingerprints have practical pharmacological relevance. Integrating GC-MS profiles with molecular taxonomy, morphology and ecological data would further strengthen species authentication and help build a reliable reference library for Kenyan Hypericum essential oil.

Author Contribution Statement

Onyancha Misonge Jared: Conceptualization, designing and supervision of the entire work. Wrote the first manuscript and was involved in the subsequent revisions till submission

Ogeto Sengera Geoffrey: Collection, identification, isolation and characterization of the volatile components in the essential oils

Mungulle Chao May: carried out in silico studies of the isolated and identified to be abundant in the essential oils. The manuscript drafts were also revised by her till the time of submission

Wachira Wairagu Norman carried out hydrodistillation and purification of the essential oils from various parts of Hypericum species. Also, he contributed to revising the manuscript to the time it was submitted.

Mbugua Maina Peter contributed to the statistical and analytical component of the study. He performed the data organization, preprocessing, statistical analysis, and interpretation of the GC-MS volatile compound data. In addition, he contributed to writing parts of the Discussion and

Conclusion sections by linking the statistical findings to chemotaxonomic relevance, chemical diversity, and future research implications.

Data Availability

The datasets used and analysed during the current study is available from the corresponding author on reasonable request.

Funding Declaration: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare no conflict of interest.

Declarations

Ethics approval and consent to participate

Ethical considerations and consent to participate was not required in this study.

Consent for publication

No consent for publication was required.

Competing interests

The authors declare no competing interests.

References

1. Jagtap T, Koche D. Powder Microscopy , Fluorescence , Qualitative Phytochemistry , and GC- MS analysis of a Neglected Ethnomedicinal Weed - Ruellia brittoniana Leonard. Haya Saudi J Life Sci. 2023;6221:71–7.

2. Celep E, Barak TH, Telci D, Yesilada E. Hypericum olympicum L. recovers DNA damage and prevents MMP–9 activation induced by UVB in human dermal fibroblasts. J Ethnopharmacol. 2019;

3. Baljak J, Bogavac M, Karaman M, Sr ¯denovi´c ˇConi´c B, Vuˇckovi´c B, Anaˇckov G, et al. Chemical Composition and Biological Activity of Hypericum Species— H. hirsutum , H. barbatum , H. rochelii. Plants. 2024;13.

4. Bertoli A, Çõrak C, Teixeira JA. Hypericum Species as Sources of Valuable Essential Oils. Med Aromat Plant Sci Biotechnol. 2011;

5. Onyancha JM, Sengera GO, Menge DM, Mandela EN, Meroka JO, Wachira NW. Hypericum revolutum essential oil as a natural anticandidal agent : Toxicological , phytochemical and antimicrobial investigation. Indian J Microbiol Res. 2026;13(1):96–107.

6. Sengera GO, Kenanda EO, Onyancha JM. Antibacterial, Antioxidant Potency, and Chemical Composition of Essential Oils from Dried Powdered Leaves and Flowers of Hypericum revolutum subsp. keniense (Schweinf.). Evidence-based Complement Altern Med. 2023;2023.

7. Moges GW, Manahelohe GM, Assege MA. Phytochemical investigation , in vitro

antioxidant , antibacterial activities of the leaf and fruit extracts of Hypericum revolutum Vahl (Amija), and essential oil composition of the leaf extract. 2025;

8. Alexey L, Alla S, Dmitrii F, Poroikov V. PASS: prediction ofactivity spectra for biologically active substances. Bioinforma Appl NOTE. 2000;16(8):747–8.

9. Filiz Ayanoğlu, Musa Türkmen DAK, Hatay. Essential oil components of Hypericum hircinum subsp. majus (Aiton) N. Robson growing in Hatay (Turkey) flora. Int Conf Adv Mater Syst. 2018;ICAMS 2018:393–6.

10. El-elimat T, El-qaderi HS, Hudaib M, Al AH. Hypericum triquetrifolium Essential Oil : Chemical Hierarchical Cluster Analysis , and Antibacterial Activity Composition ,. bioRxiv. 2025;1–21.

11. Karim H, Karim H, Kamel M, Mouna BT, Thouraya C, Brahim M. Essential oil composition of Hypericum triquetrifolium Turra . aerial parts. Ital J Biochem. 2007;56(1):1–7.

12. Gedik O. Fixed and essential oil composition of Hypericum venustum , Hypericum scabrum and Hypericum spectabile species. Prog Nutr. 2022;24.

13. Bertoli A, Çirak C, Seyis F. Hypericum spp . volatile pro fi ling and the potential signi fi cance in the quality control of new valuable raw material. Microchem J xxx. 2017;1–7.

14. Roumaissa L, Dahlab S, Saida CH, Dahlab S, Derbak L, Bordj R, et al. Examination of essential oil composition and fatty acids of Hypericum perforatum growing wild in Algeria Exame da composição de óleos essenciais e ácidos graxos de Hypericum perforatum crescendo selvagem na Argélia Examen de la composición en aceites esen.

Brazilian J Heal Rev. 2025;1–15.

15. Ferretti G, Maggi F, Tirillini B. Essential oil composition of Hypericum richeri Vill . from Italy †. Flavour Fragr J 2. 2005;20(December 2004):295–8.

16. Giuliani C, Pellegrino RM, Tirillini B, Bini LM. Natural Product Communications Volatile Organic Compounds of Hypericum androsaemum. Nat Prod Commun. 2010;5(Figure 5):107–10.

17. Schwob I, Bessière J marie, Viano J. Composition of the essential oils of Hypericum perforatum L . from southeastern France. Plant Biol Pathol. 2002;325:781–5.

18. Abreu IN, Reis MG, Marsaioli AJ, Mazzafera P. Essential oil composition of Hypericum brasiliense choise. Flavour Fragr J. 2004;19:80–2.

19. Duan J, Zhang Y, Wu W, Yao H, Li Y, Zhang C. Chemical Composition and Antioxidant Activity of the Essential Oil of Hypericum patulum (Family : Clusiaceae). Int J Agric Biol. 2018;20(July 2013):203–8.

20. Grafakou ME, Diamanti A, Antaloudaki E, Kypriotakis Z, Ćirić A, Soković M, et al. Chemical composition and antimicrobial activity of the essential oils of three closely related hypericum species growing wild on the island of Crete, Greece. Appl Sci. 2020;10(8).

21. Demirci F, Bas KHC. Volatiles of Hypericum bupleuroides Griseb . J Essent Oil Res. 2006;18(July 2005):650–1.

22. Crockett SL, Demirçi B, Baser KHC, Khan IA. Volatile Constituents of Hypericum L . Section Myriandra (Clusiaceae): Species of the H . Volatile Constituents of Hypericum L

. Section Myriandra (Clusiaceae): Species of the H . fasciculatum Lam . Alliance. J Essent Oil Res. 2008;20(3):244–9.

23. Semerdjieva I, Zheljazkov VD, Dincheva I, Piperkova N, Maneva V, Cantrell CL, et al. Essential Oil Composition of Seven Bulgarian Hypericum Species and Its Potential as a Biopesticide. Plants. 2023;12(4):1–21.

24. Chauhan RS, Vashistha RK, Nautiyal MC, Tava A, Cecotti R. Essential oil composition of hypericum perforatum L. From Cultivated Source. J Essent Oil Res. 2011;23(3):20–5.

25. Zeliou K, Koui E maria, Papaioannou C, Stavros N, Iatrou G, Tsarbopoulos A, et al. Metabolomic fingerprinting and genetic discrimination of four Hypericum taxa from Greece. Phytochemistry [Internet]. 2020;(June 2019):112290. Available from: https://doi.org/10.1016/j.phytochem.2020.112290

26. Ferraz ABF, Limberger RP, Bordignon SAL, Poser GL Von, Henriques AT. Essential oil composition of six Hypericum species from southern Brazil. Flavour Fragr J. 2005;20(February):335–9.

27. Ghiasvand A, Shadabi S, Hajipour S, Nasirian A, Borzouei M, Hassani-moghadam E, et al. Comparison of Ultrasound-Assisted Headspace Solid-Phase Microextraction and Hydrodistillation for the Identi fi cation of Major Constituents in Two Species of Hypericum. J ofChromatographic Sci. 2016;54(2):264–70.

28. Couladis M, Chinou IB, Tzakou O, Petrakis P V. Composition and Antimicrobial Activity of the Essential Oil of Hypericum rumeliacum subsp . apollinis (Boiss . & Heldr .). Phyther Res. 2003;17:152–4.

29. Kıyan HT, Demirci B, Hüsnü K, Başer C, Demirci F. The in vivo evaluation of anti-angiogenic effects of Hypericum essential oils using the chorioallantoic membrane assay. Pharm Biol. 2014;52(1):44–50.

30. Kakouri E, Daferera D, Nalbanti A, Trigas P. Volatile Constituents of Four Hypericum Species Native to Greece. Compounds. 2025;1–12.

31. Grafakou ME, Barda C, Karikas GA, Skaltsa H. Hypericum Essential Oils—Composition and Bioactivities: An Update (2012–2022). Molecules. 2022;27(16):5246.

32. Anoop SP, Joseph SM. GC Analyses of Volatile Constituents of Plants from Clausena Genus and Computing the Chemical Variability by Multivariate Correlation Analysis GC Analyses of Volatile Constituents of Plants from Clausena Genus and Computing the Chemical Variability by Multivariate Correlation Analysis. J Herbs Spices Med Plants [Internet]. 2025;31(1):79–95. Available from: https://doi.org/10.1080/10496475.2024.2433264

33. Tumbarski Y, Ivanov I, Todorova M, Gerasimova A, Dincheva I, Makedonski L, et al. Chemical Composition and Biological Activities of St John ' s Wort (Hypericum perforatum L .) Essential Oil from Bulgaria. Appl Sci. 2024;