Integrative Metabolomics, Genomics, and Transcriptomics Analysis Unravels Anti-Cancer Potential of Secondary Metabolites in Dillenia Suffruticosa

Norhayati Ahmad

norhayati.ahmad@ubd.edu.bn

Universiti Brunei Darussalam

Sabri Mohd. Ali

https://orcid.org/0000-0001-5283-3483

Khwanta Kaewnarin

Abner Lim

Jing Han Hong

Duke-NUS Medical School, Singapore https://orcid.org/0000-0002-2173-9192

Cedric Ng

https://orcid.org/0000-0002-7150-4674

Nur Izzah Amal Afiqah Hamdani

Nurul Hazlina Zaini

Mohammad Amiruddin Ruslan

Boon Yu Soon

Nor Sarah Irdina Ibrahim

Wei Liu

Odelia Yuan Xuan Chung

Boon Yee Lim

Jason Yongsheng Chan

National Cancer Centre Singapore https://orcid.org/0000-0002-4801-3703

Hussein Taha

Bin Tean Teh

National Cancer Centre Singapore https://orcid.org/0000-0003-1514-1124

Article

Keywords:

Posted Date: October 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3430002/v1

Additional Declarations: There is **NO** Competing Interest.

Abstract

The genus *Dillenia*, native to Southeast Asia and the Indian Ocean islands tropics, lacks genomic information despite its wide-ranging medicinal and ornamental applications. This study presents a comprehensive genomics, transcriptomics and metabolomics profiling of *Dillenia suffruticosa* which is widely used in the local populace and highly regenerative in secondary forests of Brunei Darussalam. The assembled genome spans a size of 596 Mb (N50: 20.8 Mb) with 30,490 genes. Multi-omics profiling revealed metabolites were discovered in *D. suffruticosa*, including phenolics, alkaloids, flavonoids, and terpenoids, alongside their biosynthetic pathways. Additionally, the study examined the cytotoxic effects of *D. suffruticosa* extracts on ten types of cancer cell lines. The findings indicate that extracts derived from the root organ, which contains higher levels of terpenoids trigger cancer cell death through the NF-kB pathway. In conclusion, this study enriches the chemogenomic and plant metabolites understanding of *D. suffruticosa* for pharmacological applications in a multidisciplinary approach.

Introduction

Dillenia is a genus of about 100 flowering plant species natively distributed across the tropical forests of South East Asia and the Indian Ocean Islands. Species in *Dillenia* are renowned for their vast phytochemical diversity, which underpins their ethnobotanical and pharmacological uses ranging from anti-microbial to ornamental purposes [1–6]. In Brunei's secondary or cleared forests, *Dillenia suffruticosa* (Simpur Bini or Simpoh Air) is a species of *Dillenia* that stands out for its large penni-veined leaves, bright yellow flowers, and pink star-shaped 'fruits'. Due to its prevalence and utility, this species is significant in Brunei Darussalam's local culture and the regional forest ecosystem.

Traditionally, the plant's large, robust leaves have been harnessed to ferment culinary staples such as rice, creating the dish known as 'tapai', and soybeans to create 'tempeh'. Beyond their culinary uses, these leaves have long been a natural remedy in the local populace's repertoire. They are widely used for healing minor cuts and wounds, providing relief from itching, and reducing inflammation, demonstrating the broad scope of their therapeutic value.

D. suffruticosa is an evergreen shrub that can reach a height of 10 meters. Its leaves are elliptical to obovate and have a folded petiole that opens at the leaf base. The plant is frequently seen alongside Macarangas and thrives in alluvial habitats such as tropical heath forests (Kerangas), white sandy areas, and ridge and hillside clayey soils [6–8]. In 2000, the flower of *D. suffruticosa* was chosen as the national flower of Brunei Darussalam during the Asia Pacific Economic Corporation (APEC), symbolizing the country's economic growth, prosperity and resilience. This particular species also plays a crucial role in natural forest regeneration following deforestation or forest fires, often colonizing areas that have been cleared [9,10].

D. suffruticosa has been leveraged for various pharmacological and ethnobotanical applications in traditional practices [2,6,11–13]. Existing studies have illuminated the cytotoxic effects of *D. suffruticosa*

extracts on breast and ovarian cancer cell lines [5,13–17]. To bridge these knowledge gaps, this study aims to assemble the genome of *D. suffruticosa*, identify and understand the biosynthesis of its phytochemicals based on genomic and metabolic profiling, and test the anti-cancer efficacy of extracts from different organs of *D. suffruticosa* against multiple cancer cell lines. This approach elucidates the underlying anti-cancer mechanism of action and assesses its safety for use. This study is an early prototype and a critical first step toward paving the way for novel metabolite candidates for treatments or complementary therapies from *D. suffruticosa*.

Results

Genome assembly and annotation of D. suffruticosa

Leaf tissues used for Whole Genome Sequencing of *D. suffruticosa* were authenticated from field studies and collected at Universiti Brunei Darussalam (Fig. 1A). *D. suffruticosa* has elliptic to obovate large leaves measuring 15–25 cm by 8-12cm. There are 12–20 pairs of lateral veins and pinnately veined leaves with a dentate margin on the leaves. The apex is acuminate and has sharp tips, measuring between 3–7 mm. The leaf base is either cuneate or rounded with a petiole measuring 2–6 cm long and 1.5 cm broad and wings that mainly open on the lower and upper parts near the stem and leaf base. The flowers are large, with a diameter of 8–11 cm with pale to bright yellow petals. The petals are thin and large, measuring 40–50 by 25–30 mm (Fig. 1B).

De novo assembly of *D. suffruticosa* was performed using ~ 60.4x coverage (36 Gb) of PacBio HiFi reads and 120 Gb Omni-C libraries generated from leaf tissues. Before assembly, the genome characteristics of *D. suffruticosa* were assessed using Smudgeplot [18] (Supplementary Fig. 1) and Genomescope [18] (Supplementary Table 1, Supplementary Fig. 2). the estimated size of the genome based on Genomescope is 520 Mb. The initial assembly spanned 651 Mb with 443 contigs, and a N50 of 18.1 Mb. The assembly was further refined by removing duplicate contigs and scaffolding with omni-c. This resulted in a final assembly of 42 scaffolds spanning 596 Mb, with an improved N50 of 24 Mb (Fig. 2, Supplementary Table 2). The quality of the assembled genome was also assessed by searching for $(CCCTAAA)_n$ telomeric motifs on the end of the assembled scaffolds with lengths exceeding 1Mb (Supplementary Table 3) and completeness with BUSCO [19]. 16 scaffolds had telomeric motifs on both ends, six scaffolds on one end, and the remaining two lacked telomeric ends. BUSCO [19] confirmed a high degree of completeness with 97% of conserved genes identified in the assembly from the Viridiplantae and Embryophyta lineages of the odb10 database (Supplementary Table 4).

The assembled genome was annotated by searching for Transposable elements (TEs) with RepeatMasker [20] and a *de novo* repeat library built using repeatmodeler2 [21]. Transposable elements occupied 52% of the genome which was soft-masked before structural annotation of the genes (Supplementary Table 5). The masked genome was then annotated with BRAKER2 [22] using protein evidence and RNA-sequencing data from various plant organs (roots, leaf, stem, and petal of *D.*

suffruticosa) for structural annotation. Eventually, functional annotation were successfully assigned to 97% (30,490) of the genes identified by BRAKER (Supplementary Table 6) using EggNOG [23,24].

Deciphering Orthologous Gene Families and Their Evolutionary Dynamics in D. suffruticosa

Insights into the evolutionary mechanisms in *D. suffruticosa* were identified using Orthofinder [25]. The tool organized the genes predicted by BRAKER [22] into distinct orthologous gene sets. Coding sequences from *D. suffruticosa* and nine other plant species, namely *Ananas comosus*, *Arabidopsis thaliana*, *Eucalyptus grandis*, *Glycine max*, *Musa acuminata*, *Sorghum bicolor*, *Theobroma cacao*, and *Vitis vinifera* were used to identify orthologous genesets for analysis. The pool of 428,917 genes across the 9 species was categorized into 31,020 unique orthologous gene sets. Of these gene sets, 17 contain single-copy genes which served as a foundation for constructing the phylogenetic tree **(**Fig. 3A**)**.

Subsequently, a detailed investigation involving changes within these orthologous gene families was analyzed using CAFE5 [26]. An ultrametric rooted tree supplemented with a tabulated count file of the orthologous gene sets was used in this analysis. An expansion of 4,461 and a contraction of 1,929 orthologous gene families specific to *D. suffruticosa*. To enhance our understanding of the dynamics of these evolutionary changes in biological processes, an enrichment analysis using clusterProfiler [27] was performed using the expanded and contracted gene sets. This allowed for an investigation into the biological processes that were either gained or lost throughout evolution compared to the other 10 species. In summary, functions associated with secondary metabolite biosynthesis were observed to be dynamically changed through this investigation. **(**Fig. 3B**)**.

Enrichment of secondary metabolite biosynthesis pathways in D. suffruticosa roots

RNA-seq from the plant tissues of *D. suffruticosa* including leaf, petal, stem and root was used to identify tissue-specific molecular signatures. Each plant tissue was profiled with two biological replicates to identify tissue-specific gene enrichment, and a PCA plot using the top 500 variable genes from each library was performed for quality control. Substantial variation across tissue types was observed. The identification of differentially expressed genesets was achieved using DEseq2 [28]. Statistically enriched genes were identified pairwise for downstream pathway analysis if the gene demonstrated a two-fold change and a *p*-adjusted value of < 0.05 from DEseq2 statistical testing. DESeq2 identified 1,375 genes significantly expressed in the petal, 1,727 in leaf tissues, 1,257 in the root tissues and 13 in the stem.

Enrichment analysis was performed on GO [29] terms and KEGG [30] pathways using clusterProfiler [27]. Differentially expressed genes in the stem yielded no interesting biological processes **(Supplementary Table 8)**. Enrichment analysis of the petal revealed pathways significantly enriched for biological processes. This includes regulation of photosynthetic activity, activity in light receptors, pigmentation and biosynthesis of secondary metabolites such as carotenoids, flavonoids and tetraterpenoids **(Supplementary Table 9)**.

Gene enrichment in the leaf tissue of *D. suffruticosa* revealed a complex landscape of metabolic and physiological processes with many pathways associated with photosynthesis and related functions, reflecting the fundamental role of leaves in energy capture and conversion. Other significantly enriched pathways include those connected with thylakoid membrane organization, chloroplast organization, and plastid organization, suggesting active maintenance and adaptation of these crucial structures within the leaf tissue [31,32]. Interestingly, several pathways related to the metabolism of pyridine-containing compounds and tetrapyrrole binding were enriched **(Supplementary Table 10)**.

Examination of the root revealed pathways related to cell wall regeneration and regulation and biosynthesis processes involving phenylpropanoid, lignin, terpenoid, isoprenoid, flavonoids, phenolic compounds and triterpenoids. Notably, secondary metabolic biosynthetic processes were enriched in the roots of *D. suffruticosa* to warrant further investigation into their pharmacological properties **(Supplementary Table 11)**.

Metabolite profiling across tissues of D. suffruticosa

Ethanol and SFE extractions from the organs (root, leaf, stem, and flower) of *D. suffruticosa* were investigated using UPLC-MS/MS in positive and negative ionization modes for their metabolic profiles. In ethanol extracts, 2,321-3,360 and 682–792 features were detected in positive and negative modes respectively. SFE extracts only had 558–846 positive and 104–262 negative features identified. Among the organs profiled, the leaf has the highest detectable features in ethanol (3,360) and SFE (846) extracts compared to the other organs.

MS/MS library search was performed in MS-DIAL and MS-FINDER for feature annotation (**Supplementary Table 12)**. Across the extracts profiled from the organs of *D. suffruticosa*, there were 151 annotated metabolites in the leaf, 134 metabolites in the flower, 134 metabolites in the root, and 137 metabolites in the stem across ethanolic extracts. On average, the number of metabolites identified across the organs of *D. suffruticosa* was 5–7 times lower in SFE extracts than in ethanol.

Six major classes of compounds were identified from *D. suffruticosa* extracts across various organs. This includes 1) primary metabolites (lipids, nucleosides and amino acids), 2) phenolic compounds, 3) flavonoids, 4) terpenoids, 5) alkaloids and 6) organic compounds covering masses ranging from 120–900 Da (Fig. 4**).** Flavonoids and terpenoids are the two main compounds in ethanol extracts of *D. suffruticosa*. Terpenoids were the major compound enriched in ethanol extracts of root (78%) and flower (36%) organs, while leaf and stem were mainly flavonoids (39%) and organic compounds (39%), respectively **(**Fig. 4A**)**.

The terpenoids found in ethanol extracts *D. suffruticosa* can be classified into seven subclassifications: 1) monoterpenoids, 2) sesquiterpenoids, 3) diterpenoids, 4) triterpenoids, 5) terpenoid glycosides, 6) terpenoid lactones and 7) terpenoid conjugates. Triterpenoids were highly abundant in the root, which features unique pentacyclic triterpenoids (dillenic, betulin, lupeol, maslinic acid, hederagenin, 4,4'-diapophytoene and ursolic acid). In SFE, the six main terpenoid metabolite subclasses include 1) terpene

glycosides, 2) sesquiterpenoids, 3) diterpenoids, 4) diterpenoid glycosides, 5) triterpenoids and 6) triterpenoid glycosides. Similarly, triterpenoids were the major constituent of *D. suffruticosa* SFE root extracts. Despite this, there was limited overlap between the titerpenoids of SFE and ethanol in the root extracts suggesting that the choice of extraction method is crucial in determining the metabolite extracted (Fig. 5).

Besides triterpenoids, another major constituent identified in *D. suffruticosa* extracts was flavonoids which is a class of phytochemicals with a C6-C3-C6 chemical structure. Seven groups of flavonoid derivatives were found in *D. suffruticosa* extracts: 1) free flavones, 2) free flavanones, 3) free flavanols, 4) free flavonols, 5) free isoflavones, 6) flavonoid conjugates (bioflavonoids and polyflavonoids and methylated flavonoids) and 7) C-glycosylated flavonoids. Across the extracts profiled, leaf tissues are highly enriched in flavonoids (Figs. 4 & 5), with an abundance of C-glycosylated flavonoids with 22 identified metabolites.

SFE extracts mainly contain primary metabolites and alkaloid molecules (Fig. 4B). Alkaloids in *D. suffruticosa* can be categorized into six classes based on their chemical structures: 1) indole alkaloids (Ajmaline-sarpagine, Harmala and Vallesaman), 2) quinine alkaloids (Cinchona), 3) amino acid alkaloids (Phenethylamines), 4) quinoline alkaloids (Phenylquinolines and Quinoline carboxylic acids), 5) quinolizidine alkaloids (Sparteine), 6) isoquinoline alkaloids (morpholines). The root SFE extract has the highest abundance of alkaloid metabolites compared to the other tissues. Among the primary metabolites, lipids classes, including medium chain fatty acids, furanoid fatty acids and phospholipids subclasses, were the primary constituents. Stem and flower SFE extracts exhibited higher lipid abundance than other plant organs.

Root specific expression of triterpenoid Biosynthesis in D. suffruticosa

The metabolic abundance of triterpenoids in root extracts of *D. suffruticosa* can be correlated to the enzymes in the triterpenoid biosynthesis pathway. The triterpenoid biosynthetic pathway starts with the precursor molecule Acetyl-CoA. Through a series of enzymatic reactions, Acetyl-CoA is converted to Farnesyl diphosphate, which is then transformed into 2,3-oxidosqualene. From 2,3-oxidosqualene, the pathway diverges to produce various triterpenoids, including Lupeol, α-amyrin, and β-amyrin. The gene expression data (average TPM) from RNA-Seq reveals that genes responsible for synthesizing these triterpenoids are predominantly expressed in the roots (Fig. 6). After their formation, these triterpenoids are transformed into their respective phytochemicals such as Betulinic acid, α-amyrin transforms into Ursolic acid, and β-amyrin gets converted to Oleanolic acid.

Dillenia suffruticosa modulate NF-κB activity in cancer cell lines

To investigate the anti-cancer efficacy of *D. suffruticosa*, ten cancer cell lines from different cancer indications were treated with extracts obtained from the various organs of the plant. The cancer cell lines examined are cholangiocarcinoma (CCA), hepatocellular carcinoma (HCC), clear cell renal cell carcinoma (ccRCC), gastric cancer, colon cancer, prostate cancer, breast cancer, lung cancer, natural

killer T cell lymphoma (NKTL) and diffused large B cell lymphoma (DLBCL). Upon treatment with these extracts, extracts obtained from the root displayed the most promising effects against cancer. This was evident with lower IC50 values for most cancer cell lines (Fig. 7A). Ethanol extracts showed better anti-cancer effects compared to SFE (Fig. 7A). Among all the cancer cell lines tested, NKTL was the least sensitive to SFE and ethanol extracts of *D. suffruticosa* (Fig. 7A). Normal cell lines, HK-2 and human peripheral blood mononuclear cells (PBMCs), were found to survive well, with significantly higher IC50 values **(Supplementary Table 13)**, at effective anti-cancer concentrations of the different extracts (Fig. 7A), suggesting that the *D. suffruticosa* are selectively anti-cancer and likely to be safe for therapy.

To elucidate how *D. suffruticosa* root extracts exert anti-cancer effects, the transcriptomics landscape of the treated cell lines was examined using RNA sequencing **(Supplementary Table 14)**. Principal component analysis of the control, ethanol extracts treated and SFE extracts treated cell lines showed that the different extracts induce different transcriptomics programs in the cancer cells **(Supplementary Figs. 3–12)**. DEseq2 analysis showed an average of 374 upregulated and 387 downregulated genes after treatment with ethanol root extracts and an average of 192 upregulated and 242 downregulated genes after treatment with SFE extracts across all cell lines **(Supplementary File 2)**. Gene set enrichment analysis (GSEA) was performed using hallmark genesets to discover enriched pathways after treatment with *D. suffruticosa* extracts (Fig. 7B). Only enriched pathways with *p*-values less than 0.05 were considered.

After treatment with both *D. suffruticosa* ethanol or SFE root extracts, upregulation of pathways related to cancer signaling, cellular damage, cellular death, cellular responses, cell metabolism, cell movement, cell cycle and immune were observed (Fig. 7B). Among the significant pathways enriched, Hallmark TNFα signaling via NFκB geneset was highly enriched and activated in most cell lines with a higher overall gene ratio, followed by the Hallmark p53 geneset. (Fig. 7B). Hallmark hypoxia, unfolded protein response and DNA repair genesets were found to be enriched and activated when cellular damage are present (Fig. 7B). Hallmark apoptosis, a mechanism of death, was also upregulated in most cancer cell lines after treatment (Fig. 7B). Genesets relating to cellular metabolism, cell movement, cell cycle and immune were upregulated in a few cancer cell lines but not the others.

Heatmaps **(Supplementary Fig. 13)** showed the expression of the genes associated with NFκB, apoptosis and p53 signaling after treatment, corroborating with GSEA analysis. After treatment, the higher expressions of these genes suggest that *D. suffruticosa* root extracts activate NFκB activities in cancer cells to increase viability through p53-mediated apoptosis.

Discussion

Traditionally, *D. suffruticosa* has been used for many indications. For example, the leaves have been used for wound healing, rheumatism relief and treatment for fever, while the fruits have been thought to treat cancer growth [3,6,13,33–35]. In addition, the leaves also reportedly possess antimicrobial, dengue antivirus, anti-inflammatory and anti-oxidant properties [3,6,11–13,36,37]. Among these pharmacological

characteristics, the anti-cancer effects of *D. suffruticosa* are perhaps the most intriguing and have been reported to be efficacious in breast cancer and ovarian cancer cell-line studies [3,5,13–17]. In addition to the cancer cell lines, we found that *D. suffruticosa* extracts are also selectively effective for cholangiocarcinoma, hepatocellular carcinoma, clear cell renal cell carcinoma, gastric cancer, colon cancer and lung cancer, with little toxic effects to normal kidney (HK2) cells and PBMCs. Despite extensive research on *Dillenia* species, little attention has been paid to genomics and chemical profiling of *D. suffruticosa*. Therefore, this study combines genomic and metabolomics to deepen our pharmacological understanding of *D. suffruticosa* and the chemical composition of both ethanol and SFE extracts.

Chemical profiling in various tissues of *D. suffruticosa* was performed using the high throughput mass spectrometry (MS) technique in this study. When performing MS analysis in positive mode, secondary metabolites tend to be ionized, forming protonated adducts with higher stability and thus improved detectability [38]. As a result, more features are typically reported in positive ionization mode resulting in a broader range of spectra in the library compared to negative mode [39]. This observation was consistent across both ethanolic and SFE extracts in this study. Among secondary metabolites detected in *D. suffruticosa* extracts, flavonoids and terpenoids are the most abundant, similar to previously reported studies.

Across MS profiling of *D. suffruticosa* plant organs, the unique pentacyclic triterpenoids (ursane, lupane and oleanane classes) were mainly found in the root samples compared to other parts. These pentacyclic triterpenes have been proposed to have a wide range of biological actions, including anti-tumor, anti-inflammatory, and antioxidant properties. Ursolic acid has been reported against proliferation in various cancers, such as breast, prostate and liver [40].

Interestingly, extracts from root organs were most effective for cancer among the extracts of the four major plant organs investigated. This was also previously reported in two breast cancer with *D. suffruticosa* extracts [5,14]. This could be owing to the high concentration of triterpenoid chemicals found in the roots [5,14]. This was also evident from RNA-seq expression and enrichment analysis of differentially expressed genes in the root revealing the enrichment of phenylpropanoid, lignin, terpenoid, isoprenoid, flavonoids, phenolic and triterpenoids within the root. This study shows ethanol extraction has a higher anti-cancer efficacy than SFE attributed to the tendency to isolate a higher abundance of flavonoids and terpenoids in ethanol extracts. This suggests that the metabolite composition across organs of *D. suffruticosa* can have various biological activities.

Naturally occurring phytochemicals are known to be effective for cancer by regulating cancer cell biological processes [41]. To elucidate the mechanism of action of ethanol extracts of the roots of *D. suffruticosa*, we integrated data from mass spectrometry to discover the phytochemicals among the different extracts with the transcriptomic landscape of the cancer cells altered by treatment with the extracts. In most cancer cells, we observed that NFκB activity was activated by treatment with ethanol extracts of the roots of *D. suffruticosa*, accompanied by enrichment in p53 and apoptosis pathways.

NFκB activation typically inhibits cell death. However, NFκB can induce cell death in certain conditions [42–45]. For example, apoptosis can also be induced by p53, which has been reported to be modulated by the activation of NFκB [14,46–48]. Flavonoids and terpenoids were found to be overrepresented in ethanol extracts of the roots of *D. suffruticosa*. Many terpenoids have been reported to increase apoptosis and reduce the viability of cancer cells. Flavonoids also play essential roles as antioxidants and have been reported to be effective for many cancers, including breast cancer, hepatocellular carcinoma and colorectal cancer [49]. Importantly certain flavonoids, such as artocarpin have been shown to increase NF-κB activity to induce apoptosis [50]. We also observed pathways related to cellular damage upregulated, which could also affect NF-κB activity and vice versa. For example, hypoxia is known to activate NF-κB pathways, but NF-κB can also activate hypoxia through direct regulation of HIF-1α expression [51,52]. Interestingly, cellular metabolism and immunity pathways were observed to be upregulated in a minority of the cancer cell lines, suggesting cancer-specific signaling of these pathways. For example, immune-related pathways were only found to be upregulated in NKTL cell line, an immune cell-related cancer, after treatment with root ethanol extracts of *D. suffruticosa.* Therefore, we hypothesize that the main mechanism of action of *D. suffruticosa* root extracts is through the enriched triterpenoids activating NF-κB directly or via p53 to induce apoptosis and reduce the viability of cancer cells. In follow-up studies, the triterpenoids can be directly isolated from *D. suffruticosa* to assay for NF-κB activation capabilities and kinetics. Given the multitude of biological pathways in cancer cells altered by the root ethanol extracts of *D. suffruticosa*, it is probable that the entire root is essential for the anti-cancer effects rather than a single phytocompound alone. The synergism and antagonism of the different phytocompounds discovered from this study warrant deeper examinations. Although the contributions of other altered pathways found through GSEA remain to be determined, this study lays the scientific groundwork to study the pharmacological effects of *D. suffruticosa* in in-vivo models to investigate the biological pathways changed at the cellular and systemic levels.

A potential application of this study is to explore the benefits and effects of consuming *D. suffruticosa* as a health supplement to cancer therapies. However, any short or long-term side effects of *D. suffruticosa* consumption should be studied carefully in appropriate *in-vitro* and *in-vivo* preclinical models. In addition, the impact of *D. suffruticosa* use on changes at the cellular and systemic levels in preclinical models should also be investigated. Clinical trials may also be required to assess the safety and performance of *D. suffruticosa* and to identify, minimize and manage risks. In conclusion, combining genomics and metabolomics can reveal the biosynthesis, pharmacological properties and any possible toxicity of plants. Many plants, particularly those in the tropics, have been reported to possess pharmacological properties, including anti-cancer effects. This study provides a framework for characterizing and curating secondary metabolites in plants that may have health benefits and safety.

Methods

Plant Materials

D. suffruticosa was authenticated and collected from 5 nearby sites within the vicinity of Universiti Brunei Darussalam (4°58'34.7"N 114°53'28.4"E, 4°58'28.0"N 114°53'23.7"E, 4°58'26.0"N 114°53'24.5"E, 4°58'23.1"N 114°53'30.2"E, 4°58'33.9"N 114°53'30.6"E). Samples used for High Molecular Weight (HMW) DNA extraction and Omni-C sequencing were taken from a single individual *D. suffruticosa* (4°58'33.9"N 114°53'30.6" E). Samples used for metabolomic and transcriptomic profiling were collected across multiple sites to account for biological replicates in this project.

Modified HMW DNA Extraction and Sequencing

Fresh *D. suffruticosa* mature leaf tissue was frozen with liquid nitrogen and ground into a fine powder using pestle and mortar. The frozen tissue powder was treated with sorbitol before nuclei isolation using a Tris-HCl buffer and subsequently incubated with a 2% CTAB buffer at 65℃ for 60 minutes. An equal volume of chloroform: isoamyl alcohol (24:1) was added and mixed with gentle inversion. Isopropanol was subsequently added into the solution DNA pellet to form. The pellet was washed twice with 70% ethanol before eluting in 50 uL TE buffer and dissolved at room temperature overnight. The extracted DNA was cleaned using 0.45x Mag-Bind® Total Pue NGS Cleanup Magnetic Bead and quantified and assessed for purity using Qubit and Nanodrop. Whole genome sequencing of HMW DNA was performed with PacBio Sequel II generating 36 Gb of hifi sequences with an average N50 of 18kb.

RNA Extraction and Library Preparation

RNA sequencing was conducted for *D. suffruticosa* mature flowers, stem, leaf, and root profiles. RNA was extracted using the Monarch Total RNA Miniprep Kit (New England Biolabs), following the manufacturer's recommended protocol. Total RNA was extracted from the flower, stem, leaf and root of *D. suffruticosa* and sent to NovogeneAIT for transcriptomic profiling. RNA sequencing libraries were obtained from processing total RNA samples using the TruSeq RNA Sample Preparation kit (Illumina) and sequencing using Illumina NovaSeq 6000.

Omni-C Library Preparation and Sequencing

The Omni-C library was prepared using the Dovetail Omni-C Proximity Ligation Assay (Dovetail Genomics, Scotts Valley, CA, USA) following the manufacturer’s protocol (manual version 1.0 for non-mammalian samples) and using nuclei pellets isolated from *D. suffruticosa* mature leaf tissue. The chromatin was fixed with formaldehyde in the nucleus, extracted and digested with sequence-independent endonucleases. After digestion, 5' overhangs were filled with biotinylated nucleotides, and the free blunt ends were ligated. Formaldehyde cross-links were reversed, and the DNA was purified and treated to remove biotin, not internal to ligated fragments. Finally, sequencing libraries were constructed using Illumina-compatible adapters (Illumina). Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced using Illumina NovaSeq 6000 to produce ~ 400 million 2 × 150 bp paired-end reads.

Genome Assembly and Annotation

PacBio hifi sequences were assembled into contigs with Hifiasm 0.19.6-1 in Hi-C mode without additional parameters. Haplotype 1 of the assembled contigs was subsequently used for all downstream analyses. The remaining duplicated sequences were removed from the assembled contigs using purge_dups v1.2.6 with the parameters -a 80 to preserve BUSCO scores. The contigs were subsequently scaffolded using Dovetail Genomics Omni-C libraries.

Dovetail Genomics Omni-C was aligned to the assembled contigs with bwa-mem2 with the – 5SP parameter. The aligned sequences were subsequently sorted, and sources of PCR and optical duplicates were subsequently identified and marked using Hi-C Pro. YAHS v1.1 was used to scaffold the assembled contigs with default parameters. Manual curation for any remaining scaffolding errors was performed using Juicebox v1.9 using a Hi-C contact map generated with juicer_tools v1.6. Gaps from Omni-C scaffolding were subsequently filled with PacBio hifi reads using close_scaffold_gaps.sh script distributed as part of the MasuRcA genome assembler to generate the final assembly.

Repetitive elements from the assembled genome were identified using RepeatModeler2 and soft-masked using RepeatMasker. Gene structure annotation was subsequently performed using BRAKER2 using RNA-seq generated from various plant organs and protein evidence from the viridiplantae odb10 database. The identified coding genes were functionally annotated usint EggNOG emapper v2.1.6.

Sample preparation for metabolomics

Fresh *D. suffruticosa* leaf, root, stem, and flower (10–12 g) were flash-frozen with liquid nitrogen and pulverized with a mortar and pestle before being transferred to a 50 mL conical tube and frozen in a -20 freezer overnight. The frozen samples were lyophilized for five days. The dried samples were stored in a -80°C freezer until further examination.

Briefly, 3 g of powder samples were extracted with 50 mL of absolute ethanol using an ultrasonication bath at 40°C for 30 min. The crude extracts were centrifuged at 6,000 rpm for 5 min. The supernatant was transferred to a new conical tube and the residue was re-extracted with 50 mL of absolute ethanol. The two extracts were combined and dried using a vacuum concentrator (Labconco, MO, U.S.A). The dried residue was weighed and stored in -80°C freezer. 10 mg of dried residue was reconstituted in 1 mL of methanol. The supernatant was spiked with salicylic acid-d6 (20 μg/mL) served as the internal standard. Samples were filtered through a 0.22 μm hydrophilic poly (vinylidene fluoride) syringe filter before UPLC-QTOF-MS analysis.

Supercritical fluid extraction (SFE) was performed on a Shimadzu Nexera-UC system (Shimadzu, Corporation, Japan). 0.6–1.8 g of powder samples were introduced into the extraction vessel for each run. The extraction was performed at a pressure of 10 MPa with 10% ethanol as a co-solvent. The average SC-CO2 flow rate is 2mL/min with a temperature of 40°C. The extracts were collected within 8 min. The solvent was evaporated using a vacuum concentrator.

Instrumental conditions and data acquisition

Data were acquired using an Agilent 1290 Infinity II LC System coupled with G6540B QTOF. Waters CORTECS T3 1.6 µm 2.1 mm x 100 mm column was deployed and maintained at 40°C for gradient elution using 0.1% aqueous formic acid as mobile phase A and acetonitrile containing 0.1% formic acid as mobile phase B. The gradient elution program was set as follows: 1% B (0–4 min), 1–70% B (4–17 min), 70–100% B (17-20.5 min), 100% B (20.5–23.9 min), 100-1% B (23.9–24 min), 1% (24–28 min), with a flow rate of 0.4 mL/min. The sample injection volume was set at 10 uL.

The electrospray ion source was used for compound ionization. Data analysis was performed in autoMSMS using the following settings: m/z range: 50 – 1,000 Da, gas temperature: 300°C, gas flow rate: 8 L/min, nebulizer: 35 psig, sheath gas temperature: 350°C, sheath gas flow: 11 L/min, capillary voltage: 3500 V, Nozzle voltage 0 V (positive mode) 2000 V (negative mode), fragmentor 75 V. Agilent ES-Tof Reference Mass solution was used for mass correction (121.0508, 922.0097 for positive ion mode; 112.9855, 1033.9881 for negative ion mode).

Data processing and compound annotation identification

LC-MS data were first processed with MS-DIAL version 4.9. MS1 and MS2 tolerance were set to 0.01 and 0.05 Da in centroid mode for each dataset. The peak was aligned on a quality control (QC) reference file with an RT tolerance of 0f 0.1 min and mass tolerance of 0.015 DA. The minimum peak width of 5 and peak height of 1500 were set for peak detection, with mass slice width of 0.1 Da for peak spotting. Adduct detection was set to [M + H] + or [M-H]- for positive or negative modes. Peaks were exported to MS-FINDER version 3.52 for compound identification. MS-FINDER parameters were set to search through formula prediction and structural elucidation through in silico fragmentation, with a mass tolerance of 20 ppm for MS1 and MS2. Formula calculation was performed under default settings. A cutoff score of 4.0 was applied for structural elucidation, with a spectral match cutoff of 25% being imposed. The compound search was focused on the following databases: FooDB, PlantCyc, NPA, GNPS, UNPD, KNApSAcK, NANPDB, and COCONUT. Hits were manually curated before being assigned to the respective chromatographic peaks.

Cell lines

Egi-1 (DSMZ, ACC385) was cultured in DMEM media (Gibco) supplemented with 10% FBS (Gibco). PBMCs (National Cancer Centre Singapore), HK-2 (ATCC, CRL-2190), SNU475 (ATCC, CRL-2236), SN12C (National Cancer Institute), AGS (ATCC, CRL-1739), SW620 (ATCC, CCL-227), PC3 (ATCC, CRL-1435), MB-MDA-231 (ATCC, HTB-26), A549 (ATCC, CCL-185) were cultured in RPMI media (Gibco) supplemented with 10%FBS (Gibco). Natural Killer/T cell lymphoma (NKTL) cell line NKS1 cells (National Cancer Centre Singapore) were maintained in modified DMEM supplemented with 10% fetal bovine serum (GIBCO, Cat#10270106), 10% horse serum (Gibco, Cat#16050122), 1% L-glutamine (Gibco, Cat#25030081), and 1% penicillin-streptomycin (Gibco, Cat#SV30010). Diffuse Large B cell lymphoma (DLBCL) U2932 cells (National Cancer Centre Singapore) were maintained in modified RPMI supplemented with 10% fetal bovine serum, 1% L-glutamine, and 1% penicillin-streptomycin. The cell lines were maintained in a 37ºC incubator with 5% CO_2 and free from mycoplasma contamination.

IC50 values determination

The plant extracts were dissolved in DMSO. 1000 cells were seeded in 96-well plates. After 24 hours, the dissolved plant extracts were added to the seeded cells. After 72 hours of treatment, cell viability was determined using CellTiter-Glo® Luminescent Cell Viability Assay (Promega) per the manufacturer's instructions. IC50 values were calculated using GraphPad Prism 9. Three biological replicates were performed, and the averages were reported as a heatmap.

RNA extraction, sequencing and analysis

3×10^5 cells were seeded in 6 well plates. After 24 hours, the dissolved indicated plant extracts at 300µg/ml were added to the seeded cells. The cells were harvested 72 hours after treatment. RNA was extracted using TRIzol (Ambion, Cat#15596018) to separate RNA aqueous phase and Qiagen RNeasy extraction kit (Qiagen) per the manufacturer's instructions. RNA bioanalyzer (Agilent) was used to determine the RNA integrity (RIN) values and quantified using the Nanodrop 2000 Spectrophotometer. Approximately 20 µg of RNA with RIN values over 7.0 were sent to NovogeneAIT Singapore for RNA library construction and sequencing using NovaSeq PE150. Two biological replicates were performed per cell line per condition. For analysis, sequencing data was aligned to the GRCh38 transcriptome using STAR[53], and the transcript abundance was subsequently counted using Salmon [54]. Pathway enrichment was performed using ClusterProfiler [27] from DESeq2 [28] differentially expressed genes.

Declarations

Acknowledgments

We would like to thank the Brunei Darussalam's Forestry Department and the Ministry of Primary Resources and Tourism (MPRT) for entry and sampling permits REF:[JPH/IND/SPL/FORM7/2023/005] and export permits (REF:JPH/HOB/TAD/51-218) and the Brunei Darussalam's Department of Agriculture and Agrifood, MPRT for phytosanitary certificate (PC/05/2023). The study team would also like to acknowledge Shimadzu Singapore for using their Supercritical fluid extraction and LC-MS profiling services.

Competing interests

All authors declared no conflict of interest.

Author contributions

Manuscript Draft: S.M.A, K.K, A.L, J.H.; Sample Collection and Identification: C.N, I.H, N.H.Z, M.A.R, S.B.Y, N.S.I; Plant nucleic acid extraction and library preparation: O.C, W.L, I.H; Metabolomic Profiling and

Analysis: K.K; Cellular Bioassay Testing and transcriptomic profiling: L.B, J.H, J.C; Bioinformatics Analysis: A.L; Conceptualization: C.N, J.C, B.T, H.T, N.A. All authors contributed to the manuscript editing and proofreading.

Data availability

The genome assembly of *Dillenia suffruticosa*, sequencing data generated from PacBio, Dovetail Genomics and transcriptomics have been deposited NCBI sequence read archive under the following bioproject accession PRJNA1020456.

References

1. Apu, A. S. *et al.* Antimicrobial Activity and Brine Shrimp Lethality Bioassay of the Leaves Extract of Dillenia indica Linn. *J. Young Pharm.* **2**, 50–53 (2010).
2. Sabandar, C. W., Jalil, J., Ahmat, N. & Aladdin, N.-A. Medicinal uses, chemistry and pharmacology of Dillenia species (Dilleniaceae). *Phytochemistry* **134**, 6–25 (2017).
3. Armania, N. *et al.* Dillenia suffruticosa exhibited antioxidant and cytotoxic activity through induction of apoptosis and G2/M cell cycle arrest. *J. Ethnopharmacol.* **146**, 525–535 (2013).
4. De, U. C., Debnath, S., Ghosh, R. & Debnath, S. Preliminary screening for in vitro anti-enteritic properties of a traditional herb Dillenia pentagyna Roxb. fruit extracts. *Asian Pac. J. Trop. Med.* **7**, S332–S341 (2014).
5. Foo, J. B. *et al.* Induction of cell cycle arrest and apoptosis by betulinic acid-rich fraction from Dillenia suffruticosa root in MCF-7 cells involved p53/p21 and mitochondrial signalling pathway. *J. Ethnopharmacol.* **166**, 270–278 (2015).
6. Saiful Yazan, L. & Armania, N. Dillenia species: A review of the traditional uses, active constituents and pharmacological properties from pre-clinical studies. *Pharm. Biol.* **52**, 890–897 (2014).
7. Wickramathilake, B. A. K., Weerasinghe, T. K. & Ranwala, S. M. W. Impacts of Woody Invader Dillenia suffruticosa (Griff.) Martelli on Physio-chemical Properties of Soil and, Below and Above Ground Flora. *J. Trop. For. Environ.* **3**, (2014).
8. Goh, M. P. Y., Basri, A. M., Yasin, H., Taha, H. & Ahmad, N. Ethnobotanical review and pharmacological properties of selected medicinal plants in Brunei Darussalam: Litsea elliptica, Dillenia suffruticosa, Dillenia excelsa, Aidia racemosa, Vitex pinnata and Senna alata. *Asian Pac. J. Trop. Biomed.* **7**, 173–180 (2017).
9. Shono, K., Davies, S. J. & Kheng, C. Y. Regeneration of native plant species in restored forests on degraded lands in Singapore. *For. Ecol. Manag.* **237**, 574–582 (2006).
10. Jambul, R., Limin, A., Ali, A. N. & Slik, F. Invasive Acacia mangium dominance as an indicator for heath forest disturbance. *Environ. Sustain. Indic.* **8**, 100059 (2020).

11. Muliawan, S. Y. Effect of Dillenia suffruticosa extract on dengue virus type 2 replication. *Universa Med.* **27**, 1–5 (2008).
12. Abubakar, S., Al-Mansoub, M. A., Murugaiyah, V. & Chan, K.-L. The phytochemical and anti-inflammatory studies of Dillenia suffruticosa leaves. *Phytother. Res.* **33**, 660–675 (2019).
13. Armania, N. *et al.* Dillenia suffruticosa extract inhibits proliferation of human breast cancer cell lines (MCF-7 and MDA-MB-231) via induction of G2/M arrest and apoptosis. *Molecules* **18**, 13320–13339 (2013).
14. Foo, J. B. *et al.* Dillenia suffruticosa dichloromethane root extract induced apoptosis towards MDA-MB-231 triple-negative breast cancer cells. *J. Ethnopharmacol.* **187**, 195–204 (2016).
15. Tor, Y. S. *et al.* Induction of Apoptosis in MCF-7 Cells via Oxidative Stress Generation, Mitochondria-Dependent and Caspase-Independent Pathway by Ethyl Acetate Extract of Dillenia suffruticosa and Its Chemical Profile. *PloS One* **10**, e0127441 (2015).
16. Tor, Y. S. *et al.* Induction of apoptosis through oxidative stress-related pathways in MCF-7, human breast cancer cells, by ethyl acetate extract of Dillenia suffruticosa. *BMC Complement. Altern. Med.* **14**, 55 (2014).
17. Wan Nor Hafiza, W. A. G. *et al.* Endoplasmic reticulum stress-induced apoptotic pathway and mitochondrial dysregulation in HeLa cells treated with dichloromethane extract of Dillenia suffruticosa. *Pharmacogn. Mag.* **12**, S86-95 (2016).
18. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1–10 (2020).
19. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
20. Smit, A., Hubley, R. & Green, P. RepeatMasker at http://repeatmasker. org. *Nucleic Acids Res* (2016).
21. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457 (2020).
22. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma.* **3**, lqaa108 (2021).
23. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
24. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
25. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).

26. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
27. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
28. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
29. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
30. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
31. Austin, J. R., Frost, E., Vidi, P.-A., Kessler, F. & Staehelin, L. A. Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes. *Plant Cell* **18**, 1693–1703 (2006).
32. Kobayashi, K. Role of membrane glycerolipids in photosynthesis, thylakoid biogenesis and chloroplast development. *J. Plant Res.* **129**, 565–580 (2016).
33. Ahmad, F. B. & Holdsworth, D. K. Traditional medicinal plants of Sabah, Malaysia part III. The Rungus people of Kudat. *Int. J. Pharmacogn.* **33**, 262–264 (1995).
34. BINTIHUSAIN, N. ANTI-COLORECTAL CANCER PROPERTIES OF DILLENIA SUFFRUTICOSA (GRIFFITH EX HOOK. F. & THOMSON) MARTELLI WATER EXTRACT IN IN VITRO AND IN VIVO MODELS. (2011).
35. Mat-Salleh, K. & Latiff, A. Tumbuhan Ubatan Malaysia. (2002).
36. Johnny, L., Yusuf, U. K. & Nulit, R. The effect of herbal plant extracts on the growth and sporulation of Colletotrichum gloeosporioides. *J. Appl. Biosci.* **34**, 2218–2224 (2010).
37. Wiart, C. *et al.* Antimicrobial screening of plants used for traditional medicine in the state of Perak, Peninsular Malaysia. *Fitoterapia* **75**, 68–73 (2004).
38. Mannochio-Russo, H. *et al.* Untargeted metabolomics sheds light on the diversity of major classes of secondary metabolites in the Malpighiaceae botanical family. *Front. Plant Sci.* **13**, (2022).
39. Tripathi, A. *et al.* Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nat. Chem. Biol.* **17**, 146–151 (2021).
40. Shanmugam, M. K. *et al.* Ursolic acid in cancer prevention and treatment: molecular targets, pharmacokinetics and clinical studies. *Biochem. Pharmacol.* **85**, 1579–1587 (2013).
41. Majrashi, T. A. *et al.* Insight into the Biological Roles and Mechanisms of Phytochemicals in Different Types of Cancer: Targeting Cancer Therapeutics. *Nutrients* **15**, 1704 (2023).
42. Bessho, R. *et al.* Pyrrolidine dithiocarbamate, a potent inhibitor of nuclear factor κB (NF-κB) activation, prevents apoptosis in human promyelocytic leukemia HL-60 cells and thymocytes. *Biochem. Pharmacol.* **48**, 1883–1889 (1994).
43. Bian, X. *et al.* NF-κB activation mediates doxorubicin-induced cell death in N-type neuroblastoma cells. *J. Biol. Chem.* **276**, 48921–48929 (2001).

44. Hettmann, T., DiDonato, J., Karin, M. & Leiden, J. M. An essential role for nuclear factor κB in promoting double positive thymocyte apoptosis. *J. Exp. Med.* **189**, 145–158 (1999).
45. Ivanov, V. N. & Ronai, Z. p38 protects human melanoma cells from UV-induced apoptosis through down-regulation of NF-κB activity and Fas expression. *Oncogene* **19**, 3003–3012 (2000).
46. Fujioka, S. *et al.* Stabilization of p53 is a novel mechanism for proapoptotic function of NF-κB. *J. Biol. Chem.* **279**, 27549–27559 (2004).
47. Ryan, K. M., Ernst, M. K., Rice, N. R. & Vousden, K. H. Role of NF-kappaB in p53-mediated programmed cell death. *Nature* **404**, 892–897 (2000).
48. Barkett, M. & Gilmore, T. D. Control of apoptosis by Rel/NF-κB transcription factors. *Oncogene* **18**, 6910–6924 (1999).
49. Roszkowski, S. Application of Polyphenols and Flavonoids in Oncological Therapy. *Mol. Basel Switz.* **28**, 4080 (2023).
50. Tsai, M.-H. *et al.* Artocarpin, an isoprenyl flavonoid, induces p53-dependent or independent apoptosis via ROS-mediated MAPKs and Akt activation in non-small cell lung cancer cells. *Oncotarget* **8**, 28342–28358 (2017).
51. Cummins, E. P. *et al.* Prolyl hydroxylase-1 negatively regulates IkappaB kinase-beta, giving insight into hypoxia-induced NFkappaB activity. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18154–18159 (2006).
52. Rius, J. *et al.* NF-kappaB links innate immunity to the hypoxic response through transcriptional regulation of HIF-1alpha. *Nature* **453**, 807–811 (2008).
53. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
54. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

Figures

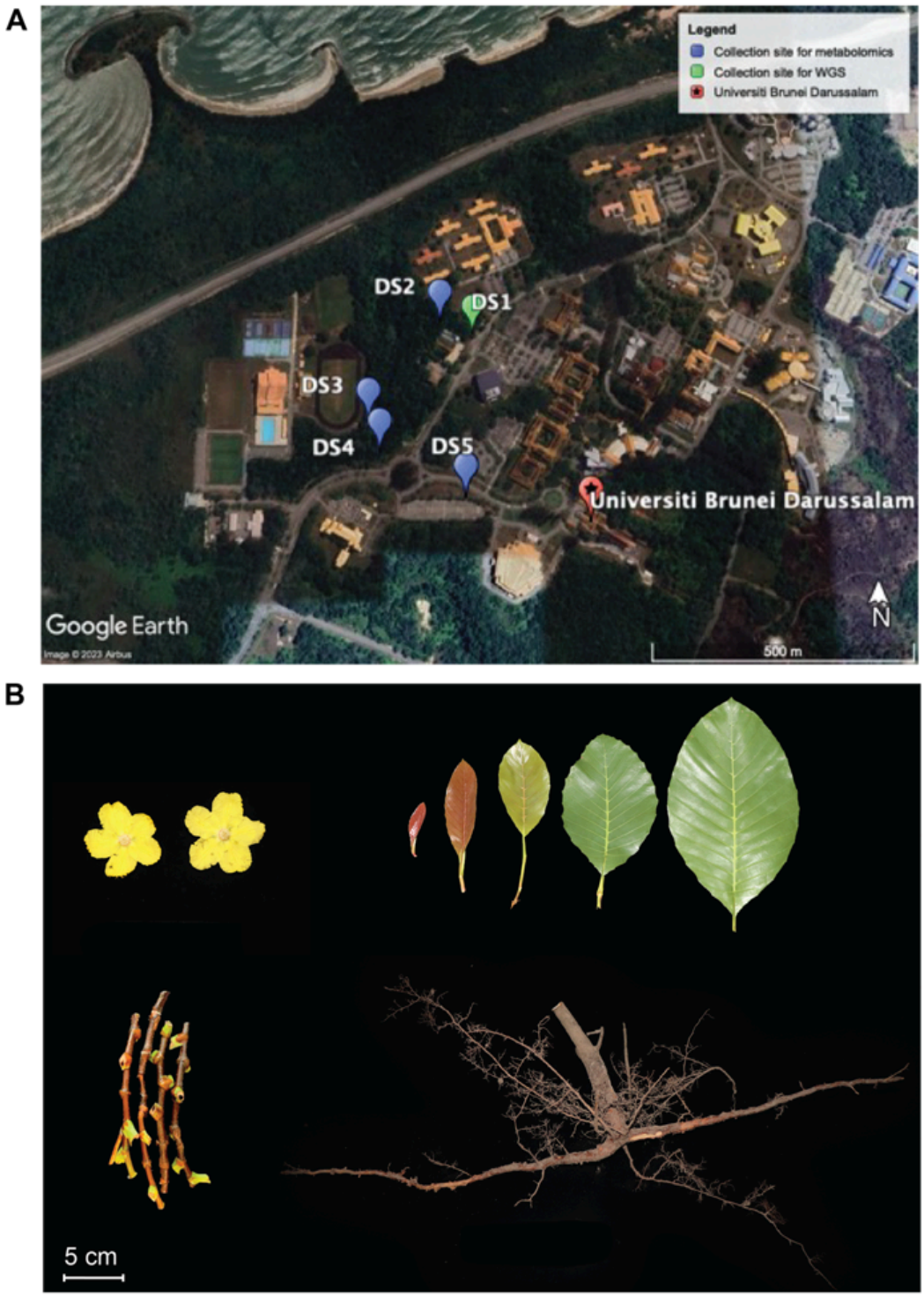

Figure 1

1A: Overview of the Universiti Brunei Darussalam campus, marking the five distinct collection sites (DS1 to DS5) for metabolomics and whole genome sequencing (WGS) studies. The designated sampling areas are proximate to the main academic buildings and coastal region.

1B: Detailed morphological representations of *D. suffruticosa*, showcasing yellow flowers, various developmental stages of leaves, stems and intricate root system. A scale of 5 cm is provided for size reference.

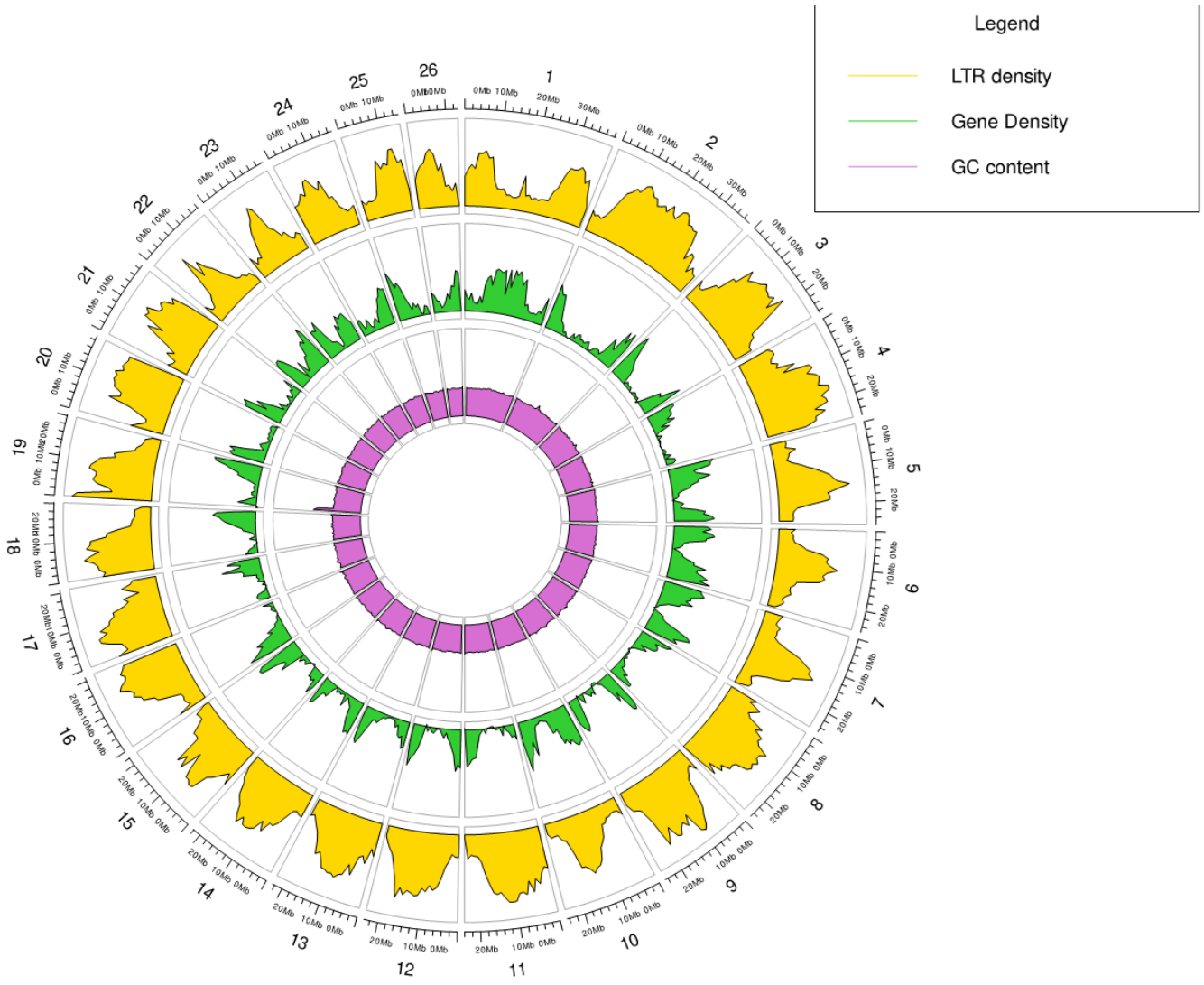

Figure 2

Circos plot of *D. suffruticosa*'s 26 assembled scaffolds. From outer track to inner track 1) Long Terminal Repeat density, 2) Gene density and 3) GC content across the 26 scaffolds.

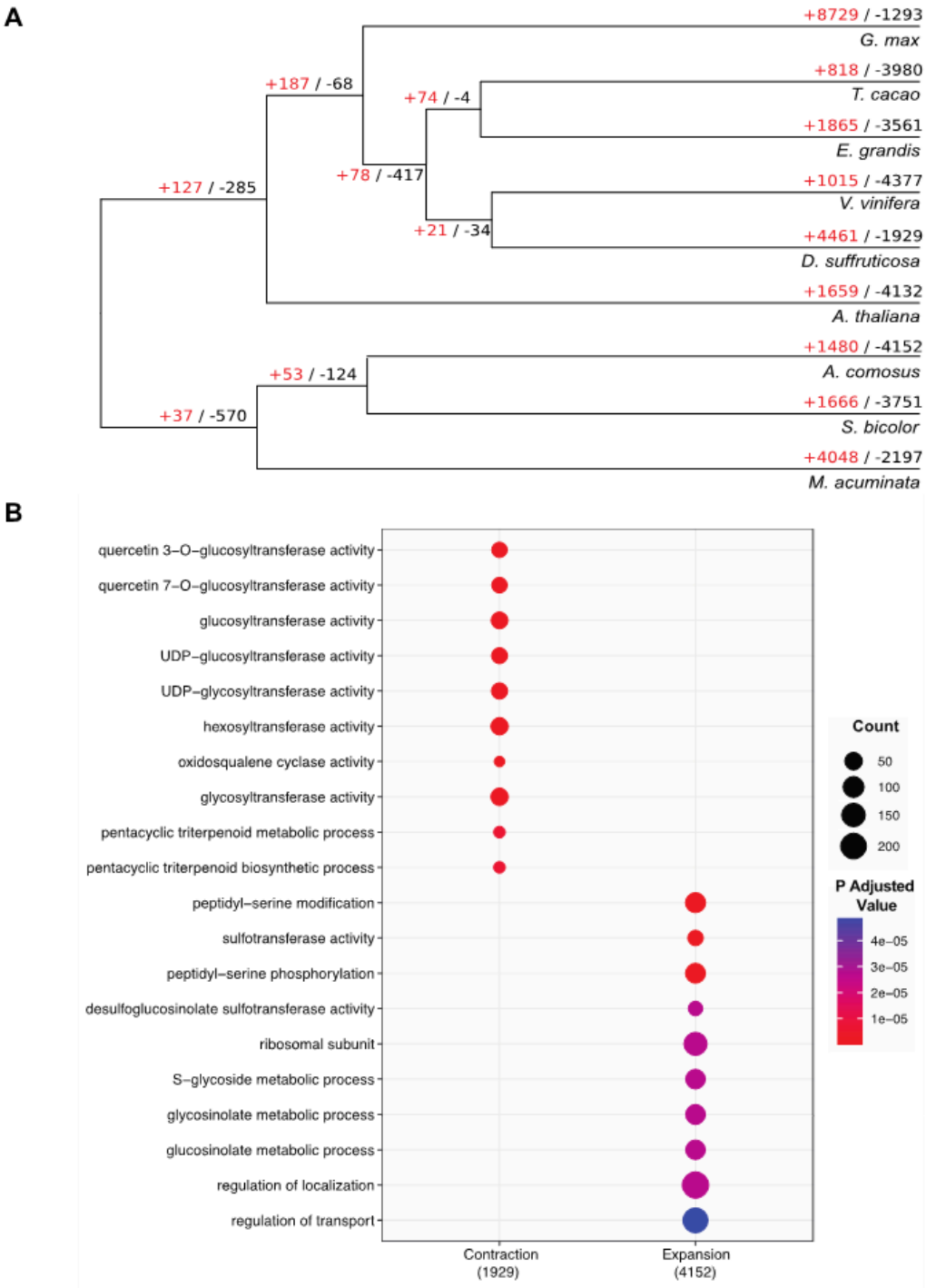

Figure 3

3A: Phylogenetic tree illustrating the evolutionary relationship of *D. suffruticosa* with eight other species from the Viridiplantae lineage. Numbers in red and black depict gene expansion and contraction events, respectively, highlighting the dynamic evolution of gene families among these species.

3B: Dot plot representation of specific biological processes and enzymatic activities influenced in D. suffruticosa. The x-axis distinguishes between gene contraction and expansion, with each dot's size indicating the count of genes associated with a given process or activity. The color gradient reflects the adjusted p-value, demonstrating the significance of enrichment for each category.

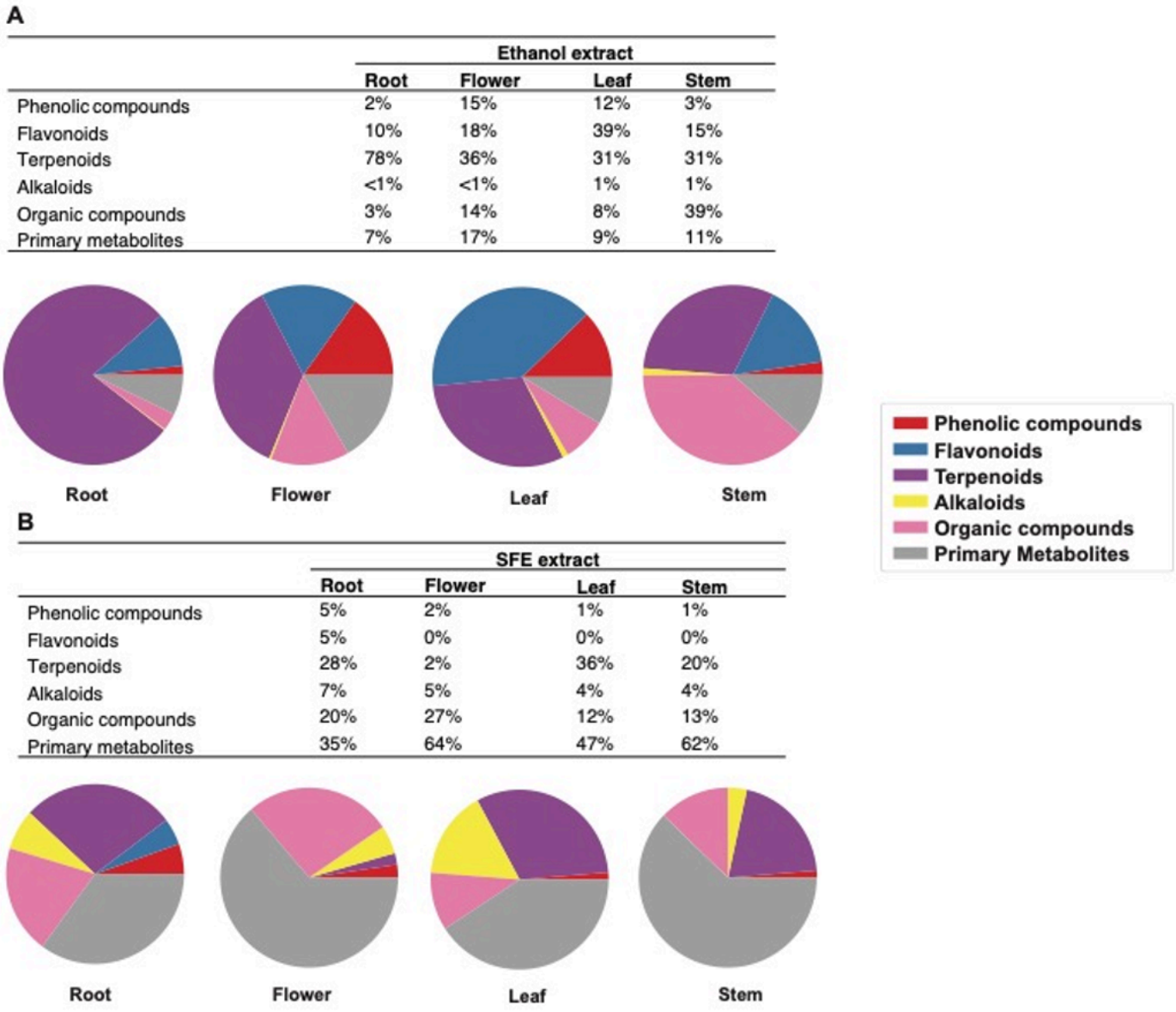

Figure 4

Distribution of metabolites in various tissues of *D. suffruticosa*. The pie charts represent the relative abundances of six key metabolite classes: phenolic compounds, flavonoids, terpenoids, alkaloids, organic compounds, and primary metabolites (comprising amino acids and lipids). These metabolite profiles were derived from two distinct extraction methods: (A) ethanol and (B) supercritical fluid extraction (SFE). Each chart corresponds to a specific plant tissue: root, flower, leaf, and stem.

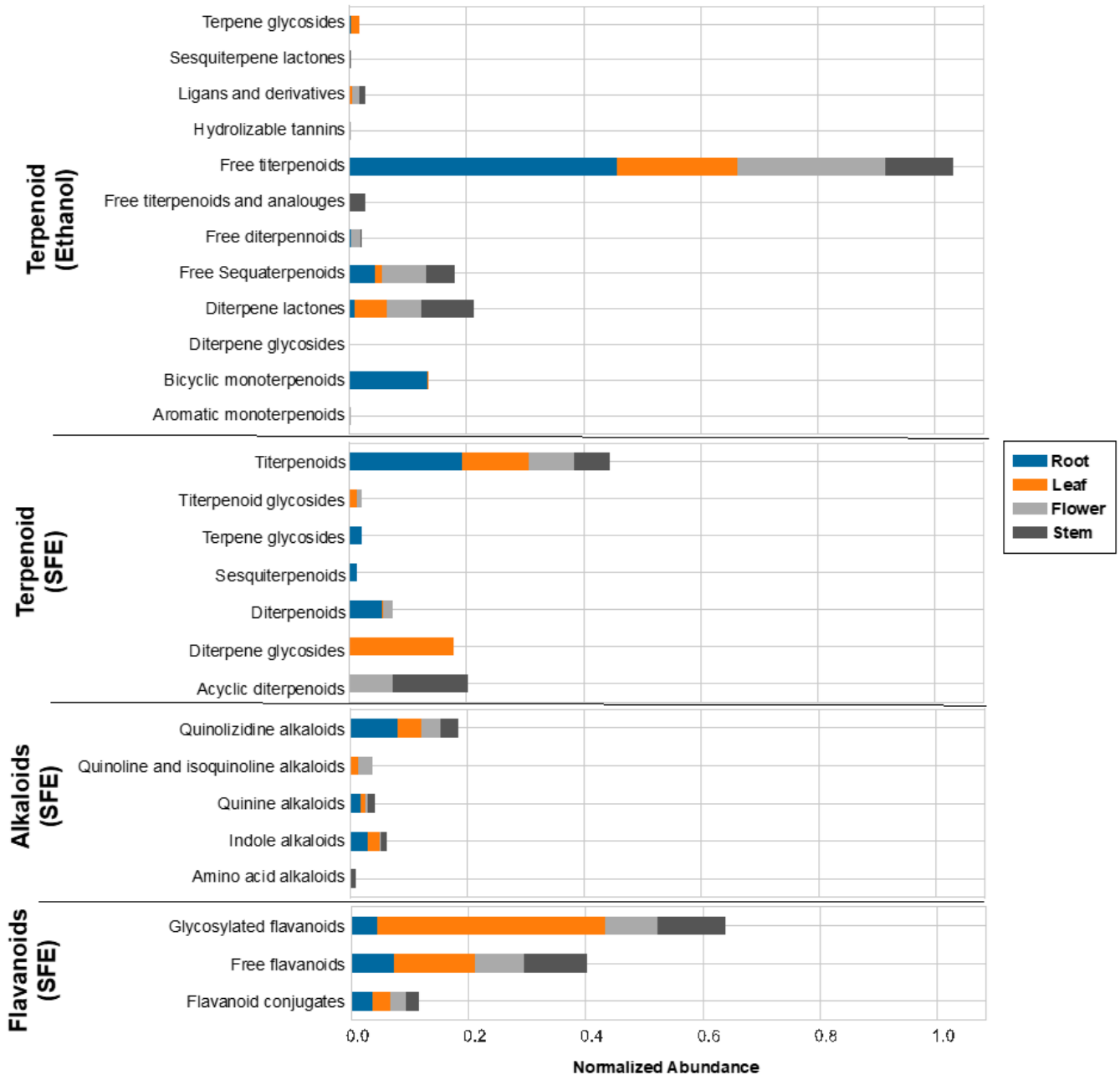

Figure 5

Stacked bar plot depicting the various classes of terpenoid, alkaloid, and flavonoid compounds identified from the root, leaf, flower, and stem tissues of *D. suffruticosa*. The abundance of each compound was normalized against the total intensities of all the identified compounds on the x-axis. Terpenoids were the primary category of secondary metabolites detected using UPLC-QTOF-MS across both ethanol and SFE extracts. Alkaloids were the second most abundant secondary metabolite in SFE extracts, while flavonoids were only detected in ethanol extract and were the second highest abundance. A total of 79 terpenoid compounds were identified from ethanol extracts, and 18 terpenoid compounds, 53 flavonoid compounds, and 13 alkaloid compounds were identified from SFE extracts.

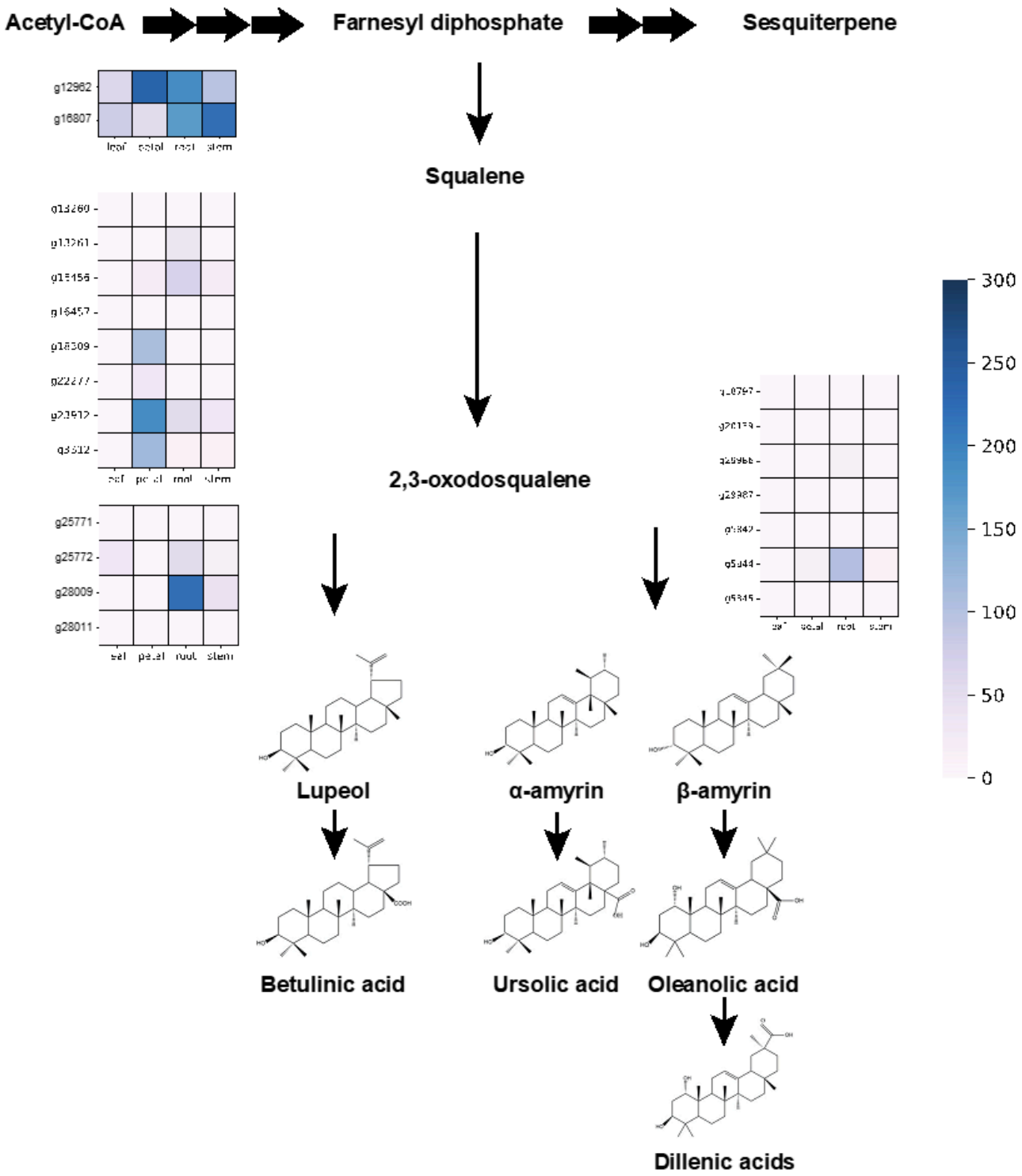

Figure 6

Biosynthetic pathway illustrating the conversion of Acetyl-CoA to various sesquiterpenoid and triterpenoid compounds in D. suffruticosa. The progression from Acetyl-CoA to sesquiterpenes, then to triterpenoids like Lupeol, α-amyrin, and β-amyrin, followed by their respective downstream compounds is depicted. Accompanying heatmaps indicate the expression levels (in TPM) of genes linked to these biosynthetic steps across different organs: leaf, petal, root, and stem.

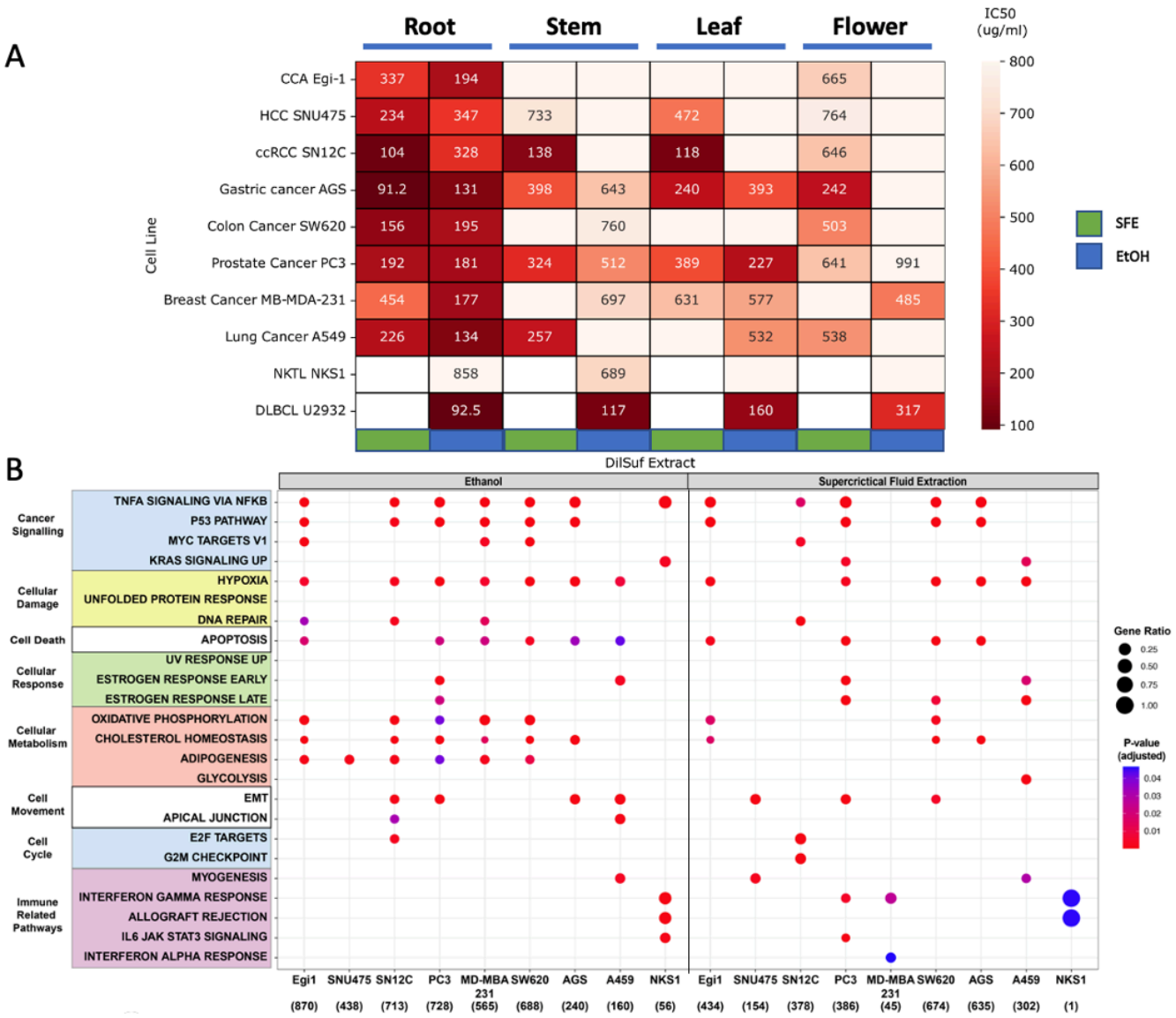

Figure 7

7A. Heatmap showing the average IC50 values across 10 primary cell lines when subjected to extracts from different parts of the D. suffruticosa plant: root, stem, leaf, and flower. Two extraction methods, Supercritical Fluid Extraction (SFE) and Ethanol (EtOH), were compared. A cell without a value denotes an IC50 value surpassing 1000µg/ml, suggesting a potential lack of anticancer activity by that extract. The abbreviations CCA, HCC, NKTL, and DLBCL correspond to cholangiocarcinoma, hepatocellular carcinoma, natural killer T cell lymphoma, and diffuse large B cell lymphoma, respectively.

7B. Dot plots representing enriched cancer hallmark pathways derived from genes displaying differential expression pre and post-treatment with *D. suffruticosa* root extracts obtained via SFE and ethanol extraction. The size symbolizes the gene ratio, while its color intensity indicates the adjusted p-value. No significant pathways were enriched from the treatment of U2932 with the extracts.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.pdf
- Supplementarytable.xlsx
- SupplementaryFile2.xlsx