

Parallel evolution of cannabinoid biosynthesis

In the format provided by the authors and unedited

Parallel evolution of cannabinoid biosynthesis

In the format provided by the authors and unedited

Table of Contents

1. Supplementary Methods.....	1–8
Chemicals.....	1
Targeted and non-targeted MS approaches for identification of metabolites.....	1–2
General instrument setups and operation for purification of metabolites.....	2
NMR spectroscopy.....	3
Trichome enrichment.....	3
Bioinformatics methods.....	3–4
3' RNA sequencing and gene co-expression network analysis.....	4
Circos and gene cluster plots.....	4
Phylogenetic analyses of functionally tested enzymes.....	4
<i>In vitro</i> enzyme assays.....	4–7
UPLC chromatographic conditions.....	7
Operational parameters of the Orbitrap IQ-X Tribrid MS.....	7–8
2. Supplementary Figures.....	9–43
Supplementary Fig. 1. Identification and labeling of major additional terpenophenols from different classes in <i>H. umbraculigerum</i>	9–10
Supplementary Fig. 2. Differentiation between chalcones and flavanones via UV absorbance and thermal stability.....	11
Supplementary Fig. 3. Identification of CBGA-type alkyl homologues in <i>H. umbraculigerum</i>	12
Supplementary Fig. 4. Identification of cannabichromenic acid (CBCA 15) and CBCA-like metabolites in <i>H. umbraculigerum</i>	13
Supplementary Fig. 5. Identification of cannabiprenylic acid (CBPA 19) and heli-cannabiprenylic acid (heliCBPA 20) in <i>H. umbraculigerum</i>	14
Supplementary Fig. 6. Identification of hydroxylated cannabinoids and amorfrutins in <i>H. umbraculigerum</i>	15–16
Supplementary Fig. 7. Identification of cyclocannabigerolic acid (cycloCBGA 47) and cyclocannabigerolic acid-like metabolites in <i>H. umbraculigerum</i>	17
Supplementary Fig. 8. Identification of geranylated-acyl homologue phloroglucinoids in <i>H. umbraculigerum</i>	18–19
Supplementary Fig. 9. Identification of acylphloroglucinoids with different prenylations in <i>H. umbraculigerum</i>	20–21
Supplementary Fig. 10. Identification of hydroxylated and dihydroxylated prenyl-acyl-phloroglucinoids in <i>H. umbraculigerum</i>	22
Supplementary Fig. 11. Identification of chalcones with different prenylations in <i>H. umbraculigerum</i>	23–24
Supplementary Fig. 12. Identification of flavanones with different prenylations in <i>H. umbraculigerum</i>	25–26
Supplementary Fig. 13. Major intermediates (92-101 before prenylation) from all the five metabolic routes identified in <i>H. umbraculigerum</i>	27
Supplementary Fig. 14. Genome size estimation of <i>H. umbraculigerum</i> by flow cytometry.....	28
Supplementary Fig. 15. Read Length vs Read Quality kde plot of the three SMRT cells sequenced.....	29
Supplementary Fig. 16. Hi-C contact frequency matrix represented as heatmap along all <i>H. umbraculigerum</i> scaffolds of the primary assembly sorted by length.....	30

Supplementary Fig. 17.	BUSCO completeness values of each of the assemblies.....	31
Supplementary Fig. 18.	Over Representation Analysis of module M4.....	32
Supplementary Fig. 19.	Coupled PKS-PKC reactions and PKC phylogeny.....	33
Supplementary Fig. 20.	Comparison of total ion chromatograms of six different tissues from <i>H. umbraculigerum</i>	34
Supplementary Fig. 21.	Identification of glucosylated intermediates, cannabinoids and amorfrutins in <i>H. umbraculigerum</i>	35–36
Supplementary Fig. 22.	MS/MS spectra of monoglucosides following <i>in vitro</i> assays with UGTs from <i>H. umbraculigerum</i> , stevia and rice.....	37
Supplementary Fig. 23.	<i>In vitro</i> production of di-glucosides with the purified UGTs.....	38
Supplementary Fig. 24.	Identification of <i>O</i> -acylated cannabinoids in <i>H. umbraculigerum</i> .	39–40
Supplementary Fig. 25.	Identification of hydroxylated and dihydroxylated <i>O</i> -acylated cannabinoids in <i>H. umbraculigerum</i>	41
Supplementary Fig. 26.	Identification of <i>O</i> -acylated amorfrutins in <i>H. umbraculigerum</i>	42–43
3.	Supplementary Table Captions.....	44
4.	Supplementary NMR Data.....	45–86
	Supplementary NMR Data- General Note.....	45
	Supplementary NMR Data 1. Data of CBGA 1	46–48
	Supplementary NMR Data 2. Data of heliCBGA 2	49–51
	Supplementary NMR Data 3. Data of geranylphlorocaprophenone 4	52–54
	Supplementary NMR Data 4. Data of geranylpinocembrin chalcone 6	55–57
	Supplementary NMR Data 5. Data of 12-OH-cycloCBGA 26	58–60
	Supplementary NMR Data 6. Data of geranylphloro-2-methylbutyrophenone 54	61–62
	Supplementary NMR Data 7. Data of geranyl-phloretin 82	63–65
	Supplementary NMR Data 8. Data of 6-prenylpinocembrin 86	66–68
	Supplementary NMR Data 9. Data of 6-prenylnaringenin 87	69–71
	Supplementary NMR Data 10. Data of 6-geranylnaringenin 88	72–74
	Supplementary NMR Data 11. Data of Glc-OA 102	75–77
	Supplementary NMR Data 12. Data of Glc-DHSA 103	78–80
	Supplementary NMR Data 13. Data of <i>O</i> -MeButCBGA 120	81–83
	Supplementary NMR Data 14. Data of <i>O</i> -MeButHeliCBGA 138	84–86
5.	Supplementary Orthology Data.....	87–90
	Supplementary Orthology Data 1. Rooted species tree inferred with Orthofinder.....	87
	Supplementary Orthology Data 2. Orthogroup OG0014461 to which HuCoAT6 belongs	88
	Supplementary Orthology Data 3. Orthogroup OG0000313 to which HuTKS4 belongs...	89
	Supplementary Orthology Data 4. Orthogroup OG0002538 to which HuCBGAS4 belongs.....	90
6.	References.....	91–93

Supplementary Methods

1. Chemicals

Unless otherwise stated, all the analytical metabolites were $\geq 95\%$ pure. CBGA **1**, CBCA **15**, CBDA, acetic acid, propionic acid, butyric acid, pentanoic acid, hexanoic acid, heptanoic acid, octanoic acid, ± 2 -methyl butyric acid, phenylalanine, hexanoic-D₁₁ acid (D>98%), GPP, IPP, FPP, phloretin **98**, naringenin **96**, malonyl-CoA ($\geq 90\%$), acetyl-CoA ($\geq 93\%$), butyryl-CoA ($\geq 90\%$), hexanoyl-CoA ($\geq 85\%$), octanoyl-CoA, iso-valeryl CoA ($\geq 90\%$), olivetol and sodium hexanoate were purchased from Sigma-Aldrich (Rehovot, Israel). Δ^9 -THCA was purchased from Silicol Scientific Equipment Ltd. (Or Yehuda, Israel). Acetic-D₃ acid (D>99%), propionic-D₅ acid (D>99%), butyric-D₅ acid (D>98%), pentanoic-D₉ acid (D>98%), heptanoic-D₅ acid (D>99%), octanoic-D₅ acid (D>99%), iso-butyric-D₇ acid (D>98%), ± 2 -methyl butyric-D₉ acid (D>99%), iso-valeric-D₉ acid (D>98%), iso-caproic-D₁₁ acid (D>98%) were purchased from C/D/N isotopes (Quebec, Canada). Phenylalanine-D₅ (D>98%) and phenylalanine-¹³C₉,¹⁵N₁ (¹³C,¹⁵N>99%) were synthesized by Cambridge Isotope Laboratories (Andover, MA). HeliCBGA **2** (NP009525, 90%) was purchased from Analyticon Discovery GmbH (Potsdam, Germany). APHA **3** was reported as an impurity (NP015136, 5%) in the heliCBGA analytical metabolite. OA **92** (>90%), VA (>90%) and iso-butyryl-CoA were purchased from Cayman Chemical (Ann Arbor, MI, USA). PCP **95**, naringenin chalcone **97** and pinocembrin chalcone **100** were purchased from Wuhan ChemFaces Biochemical Co Ltd. (Hubei, China). Cinnamoyl-CoA and Coumaroyl-CoA were purchased from TransMIT GmbH (Hesse, Germany).

2. Targeted and non-targeted MS approaches for identification of metabolites

We applied in this study both a targeted and a non-targeted approach using mass spectrometry (MS) to identify unknown peaks. In the non-targeted practice, we used the MISO package³⁵ developed by our team to compare the LC-MS/MS data with and without feeding *H. umbraculigerum* leaves with stable isotopic labeled precursors to identify new peaks that incorporated the labeled metabolites. The MS/MS spectra of the labeled metabolites also helped in the assignment of fragmentation structures. We also used the Weizmass library of plant metabolites⁴⁶ to screen the plant extract's spectra. In the targeted approach, we first established a list of potential masses according to previously identified metabolites in *H. umbraculigerum*. We added to this list masses of additional putative structures according to possible metabolic pathways. For example, we assumed that metabolites with different chain lengths may exist similar to *Cannabis* and may also have different prenylations. We also considered potential modifications of the main terpenophenols, including hydroxylation, methylation, acylation, glycosylation, and others. We then screened the chromatogram for possible masses and identified unknown metabolites by accurate mass, relative retention time (RT), MS/MS, and UV spectra. Since most of the terpenophenols identified in this study had no commercially available analytical standards, we purified selected metabolites and elucidated their structure via one- and two-dimensional NMR. These then served as standards for identifying chemically similar metabolites by relative RT and MS/MS fragmentation.

In the assignment of structures according to LC-MS/MS data, we used similar guidelines as previously suggested for putative identification of cannabinoids in *Cannabis*¹⁵. For homologues with different alkyl chain lengths, we observed analogous fragmentation patterns with mass shifts corresponding with the chain length. The same was also observed for similar metabolites with different prenylations. Each terpenophenol group had some typical fragments and fragmentation structures. We also found that the relative order of elution of metabolites from the reversed-phase

(RP) column can predict the metabolite's structure. For example, increasing alkyl chain length (C1→C7) and prenyl group (isoprenyl→monoprenyl→sesquiterpene→diterpene) result in increasing RT. Branching of the alkyl chain slightly reduces RT compared to the linear configuration for the same carbon number, while reducing a double bond (like in the phenethyl group) slightly increases RT. For the same chain length and prenylation (similar accurate mass), cannabinoids elute before phloroglucinoids. The same also occurs for amorfrutins versus chalcones and flavanones. The order of elution of flavanones versus chalcones with similar masses cannot be predicted. Since they have similar MS/MS fragmentation patterns, the best method to differentiate between the two groups is via UV spectra and observing differences in appearance/disappearance of peaks following heating. Modifications on the structure affect the time of elution depending on the change. For example, hydroxylation reduces the metabolite's lipophilicity, leading to faster elution, whereas *O*-acylation increases the RT. Oxidocyclization, like in the case of CBGA **1**→CBCA **15**, leads to longer RTs.

3. General instrument setups and operation for purification of metabolites

The Büchi Sepacore MPLC System was equipped with two C-605 pump modules, a C-620 control unit, C-660 fraction collector, C-640 UV photometer (Büchi Labortechnik AG, Switzerland), and a C18 manually packed column. The mobile phase consisted of acetonitrile:water (5:95, v/v; phase A) and acetonitrile (phase B), with the following multistep gradient method: initial conditions were 0% B for 10 min, raised to 99% B until 530 min, and slowly raised to 100% B until 660 min. The flow rate was 15 ml min⁻¹, the injection volume was 15 ml and the wavelengths were: 210, 224, 270 and 350 nm. Fractions of 100 ml were collected throughout the run and analyzed by UPLC–qTOF to select specific metabolites for purification. The selected fractions were evaporated using a rotary evaporator at 40 °C, lyophilized, reconstituted in ethanol or methanol, and filtered through a 0.22 µm syringe filter.

The Agilent 1290 Infinity II UPLC system was equipped with a quaternary pump, autosampler, diode array detector, a Bruker/Spark Prospekt II LC–SPE system (Spark) and Impact HD UHR-QqTOF MS (Bruker) connected via a Bruker NMR MS Interface (BNMI-HP). Method development was performed by acquisition of both MS and UV signals. MS spectra were acquired in negative full scan mode between *m/z* 50 and 1,700. HPLC columns were either XBridge (BEH C18, 250 × 4.6 mm i.d., 5 µm; Waters) or Luna (C18, 250 × 4.6 mm i.d., 5 µm; Phenomenex), and the conditions were adjusted and optimized for each metabolite. In this system, the eluent with the metabolites of interest were mixed with a makeup-flow of 1.8 ml min⁻¹ water and then trapped on solid phase extraction (SPE) cartridges (10 × 2 mm Hysphere resin GP cartridges). Each cartridge was loaded four times with the same metabolite, and 36–72 cartridges were used for trapping one metabolite, depending on the concentration of the sample injected. After collection, SPE cartridges were dried with a stream of N₂, and eluted with 150 µl methanol.

The Waters Acquity UPLC system was equipped with a binary pump, an autosampler, a fraction manager and a diode array detector. A UPLC BEH C18 column (100 mm × 2.1 mm i.d., 1.7 µm; Waters) was used on for purification, apart from metabolites Glc-OA **102** and Glc-DHSA **103**, which were fractionated on a Luna Phenyl-Hexyl column (150 mm × 2 mm i.d., 3 µm; Phenomenex). The flow rate was 0.3 ml min⁻¹, and the column temperature was kept at 35 °C. All other conditions were adjusted and optimized according to the sample. The eluent with the metabolite of interest was collected in 2 ml HPLC vials.

4. NMR Spectroscopy

Purified metabolites were resuspended in 300 μ l of Methanol- d_4 , dried under a stream of N_2 , reconstituted in 70 μ l Methanol- d_4 with 0.01% of 3-(trimethylsilyl)propionic-2,2,3,3- d_4 acid sodium salt (TMSP, used as an internal chemical shift reference for 1H and ^{13}C) and transferred into 1.7 mm micro NMR test tubes for structure elucidation. NMR spectra were collected on a Bruker AVANCE NEO-600 NMR spectrometer equipped with a 5 mm TCI-xyz CryoProbe. All spectra were acquired at 298 K. The structures of the different metabolites were determined by one dimensional (1D) 1H NMR spectra, as well as various two-dimensional (2D) NMR spectra: 1H - 1H Correlation Spectroscopy (COSY), 1H - 1H Total Correlation Spectroscopy (TOCSY), 1H - 1H Rotating Frame Nuclear Overhauser Spectroscopy (ROESY), 1H - ^{13}C Heteronuclear Single Quantum Coherence (HSQC), and 1H - ^{13}C Heteronuclear Multiple Bond Correlation (HMBC) spectra.

One dimensional 1H NMR spectra were collected using 16,384 data points and a recycling delay of 2.5 s. Two-dimensional COSY, TOCSY and ROESY spectra were acquired using 16,384–8,192 (t_2) by 400–512 (t_1) data points. 2D TOCSY spectra were acquired using isotropic mixing times of 100–300 ms. A T-ROESY experiment was used in this study, TOCSY-less ROESY that effectively suppresses TOCSY transfer in ROESY experiments. T-ROESY spectra were recorded using spin lock pulses of 100–400 ms. 2D HSQC and 2D HMBC spectra were collected using 4,096 (t_2) by 400–512 (t_1) data points. Multiplicity editing HSQC enables differentiating between methyl and methine groups that give rise to positive correlation, versus methylene groups that appear as negative peaks. HMBC delay for evolution of long-range couplings was set to observe long-range couplings of $J_{H,C} = 8$ Hz. All data were processed and analyzed using TopSpin 4.1.1 software (Bruker).

5. Trichome enrichment

Trichomes were enriched following Balcke et al. guidelines⁴⁷ with modifications. Briefly, young leaves were harvested and soaked in ice-cold, distilled water and then abraded using a BeadBeater machine (Biospec Products, Bartlesville, OK). The polycarbonate chamber was filled with 15 g of plant material, and filled with half the volume with glass beads (0.5 mm diameter), XAD-4 resin (1 g g^{-1} plant material), and ethanol 80% to full volume. Leaves were beaten by 2-4 pulses of operation of 1 min each. This procedure was carried out at 4 $^{\circ}C$, and after each pulse the chamber was allowed to cool on ice. Following abrasion, the contents of the chamber were first filtered through a kitchen mesh strainer and then through a 100- μ m nylon mesh to remove the plant material, glass beads, and XAD-4 resin. The residual plant material and beads were scraped from the mesh and rinsed twice with additional ethanol 80% that was also passed through the 100- μ m mesh. The presence of enriched glandular trichome secretory cells was checked by visualization in an inverted optical microscope.

6. Bioinformatics methods

Briefly, random sequencing errors were corrected using Rcorrector⁴⁸ and uncorrectable reads were removed. Adaptor and quality trimming were performed using TrimGalore!⁴⁹. Ribosomal RNA was filtered by discarding reads mapping to SILVA_132_LSURef and SILVA_138_SSURef non-redundant databases using bowtie2⁵⁰. Fastq quality checks on each of the steps were performed using MultiQC⁵¹. The remaining reads were pooled and used for genome-guided and genome-independent *de novo* transcriptome assembly using Trinity⁵².

The Iso-Seq data was obtained from four of the tissues (Supplementary Table 8) and processed with isoseq3⁵³. Fused and unspliced transcripts were removed, and only polyA-positive transcripts were kept for a unique set of high-quality isoforms. Iso-Seq and Trinity transcripts were aligned to the assembly using minimap2⁵⁴ and the BAM files were incorporated into the PASA pipeline⁵⁵ to generate RNA-based gene model structures. In addition, *de novo* gene structures were obtained using the software braker2⁵⁶ and the BAM file alignments of long and short reads as extrinsic training evidence. *Ab initio* and RNA-based gene models were combined using EvidenceModeler followed by a final round of PASA pipeline⁵⁵. Gene functional annotation was performed for the predicted mature transcripts using TransDecoder⁵⁷, which takes into account HMMER⁵⁸ hits against PFAM⁵⁹ and BLASTP⁶⁰ hits against UniProt⁶¹ databases for similarity retention criteria. Further annotation of protein-coding transcripts was performed by taking the best hit of BLASTP searches against other plant protein databases (Uniprot protein fasta files of sunflower id UP000215914_4232, Arabidopsis id UP000006548_3702, tomato id UP000004994_4081, rice id UP000059680_39947 and *Cannabis* NCBI id GCF_900626175.1_cs10). Signal peptides were predicted with SignalP⁶², transmembrane domains were predicted with TMHMM⁶³, and GO and KEGG terms were obtained with Trinotate⁶⁴. BUSCO⁶⁵ was used at multiple stages of the analysis to assess the completeness of the different versions of both the transcriptome and the genome.

7. 3' RNA sequencing and gene co-expression network analysis

UMI-based 3' RNAseq of three replicates of the seven tissues was obtained similarly as described⁶⁶. Adaptor and quality trimming was performed using TrimGalore!⁴⁹ in two steps, including PolyA trimming mode. Reads were mapped to the genome using STAR⁶⁷ UMI-deduplicated using UMI-tools⁶⁸, and counts were obtained with featureCounts⁶⁹. Normalization was performed with the varianceStabilizingTransformation algorithm of DESeq2⁷⁰, and the CEMITools⁷¹ package was used for co-expression analysis (dissimilarity threshold of 0.6, pvalue of 0.1).

8. Circos and gene cluster plots

Gene and TEs density were calculated by intersecting the corresponding gff files with 0.1 Mb non-overlapping windows using bedtools makewindows and bedtools intersect⁷². True-seq and Trans-seq coverage were calculated using bedtools genomecov in BedGraph format⁷². The circos plot was made with the R circlize⁷³ package and the gene clusters plots were made with the gggenes⁷⁴ package.

9. Phylogenetic analyses of functionally tested enzymes

The selection of the proteins for each of the families analyzed in this study was based on functionally tested enzymes according to studies referenced in each Figure. The full list of IDs can be found in Supplementary Table 15. The Maximum Likelihood trees were constructed with 100 bootstrap tests based on a MUSCLE multiple alignment using the MEGA11 software. The evolutionary distances were computed using the JTTmatrix-based method.

10. *In vitro* enzyme assays

AAE enzyme assays

Recombinant AAE assays were performed in a 20 µl reaction mix that contained 0.1 µg recombinant AAE, 50 mM HEPES pH 9.0, 8 mM ATP, 10 mM MgCl₂, 0.5 mM CoA and 4 mM of the sodium salt of the respective acid (acetic, butyric, hexanoic, octanoic, cinnamic and coumaric acids) for 10 min at 40 °C. Reactions were terminated with 2 µl of 1 M HCl and stored on ice until analysis. After centrifugation at 15,000 g for 5 min at 4 °C, the samples were diluted

1:100 in water and analyzed on the TQ-S system in MRM mode using a similar column as previously described. The system was operated with an aqueous buffer pH 7.0 (10 mM ammonium acetate, 5 mM NH₄HCO₂, phase A) and acetonitrile (phase B) according to UPLC Method 4). The instrument was operated in positive mode with a capillary voltage of 3.0 kV, and a cone voltage of 50 V. Two different transitions were used for analysis of: acetyl-CoA (810.52 > 303.30, 27.0V; 810.52 > 428.25, 24.0V); butyryl-CoA (838.58 > 331.30, 28.0 V; 838.58 > 331.30, 25.0 V); hexanoyl-CoA (866.65 > 359.40, 28.0 V; 866.65 > 428.25, 26.0 V); octanoyl-CoA (894.65 > 387.55, 30.0 V; 894.65 > 428.25, 28.0 V); coumaroyl-CoA (914.59 > 407.37, 30.0 V; 914.59 > 428.25, 28.0 V); cinnamoyl-CoA (898.59 > 391.37, 30.0 V; 898.59 > 428.25, 28.0 V). Metabolite identity was confirmed with authentic standards.

PKS and PKC enzyme assays

Individual and coupled HuPKS and PKC (HuOACs or CsOAC) assays were carried out as described by Gagne et al. (2012)⁷ with some modifications. Enzyme assays were performed in 50 µl with 20 mM HEPES at pH 7.2, 5 mM DTT, 1.8 mM malonyl CoA and 0.6 mM of hexanoyl-CoA. HuPKSs (5 µg) and PKCs (10 µg), were added either individually or in combination. Reaction mixtures were incubated at 30 °C for 3 h. Reactions were stopped by extraction with 100 µl methanol, vortexing and centrifugation at 15,000 g for 10 min. The supernatant was filtered and analyzed with both UPLC–qTOF and triple-Quad systems. The column and mobile phase were as for the metabolic profiling. Metabolites were analyzed using UPLC Method 5. UPLC–qTOF was run in both polarities with MS or MS/MS modes using similar parameters as previously described. The TQ-S system was operated in MRM mode in both positive (for olivetol) and negative modes with a capillary voltage of 3.5 or 1.5 kV, respectively, and a cone voltage of 40 or 20 V, respectively. Two different transitions were used for analysis of: OA 92 (223.1 > 179.1, 15.0 V; 223.1 > 137.1, 20.0 V); PDAL (181.2 > 137.1, 10.0 V; 181.2 > 97.1, 20.0 V); HTAL (223.1 > 179.1, 10.0 V; 223.1 > 125.1, 10.0 V); PCP 95 (223.1 > 179.1, 20.0 V; 223.1 > 81.0, 25.0 V); olivetol (181.1 > 111.0, 10.0 V; 181.1 > 71.2, 10.0 V). Olivetol, OA **92** and PCP **95** identities were confirmed with authentic standards.

PT enzyme assays

Genes encoding *HuPT1-4* were separately cloned into pESC-TRP vector. Microsomal preparations from yeast cells transformed with pESC-TRP vectors were performed following Jozwiak et al. (2020)⁷⁵. Briefly, Yeast cells transformed with pESC-TRP:HuPT1-4 were grown for 24 h at 30 °C in 50 ml of Synthetic Defined (SD) minimal media supplemented with appropriate amino acids and 2% glucose. The cell culture was centrifuged for 10 min at 700 g, pellet resuspended in 50 ml of H₂O and pelleted again. Washed yeast cells were used to inoculate 2 l culture of SD minimal media with 2% galactose. After 48 h cells were collected by centrifugation at 700 g and washed with cold PBS pH 7.4. Each pellet was resuspended with ice-cold lysis buffer (1× PBS, 1 mM DTT, 2 mM PMSF and 1× cOmplete, EDTA-free Protease Inhibitor Cocktail: Roshe) in the ratio 1:3 (w:v). An equal amount of acid-washed glass beads (425–600 µm, Sigma-Aldrich) was added to the cell suspension and vortexed at the highest speed for 1 min. The procedure was repeated 12 times with 1-min breaks on ice. The suspension with broken cells was transferred to new tubes and centrifuged at 17,000 g for 10 min at 4 °C. The supernatant was transferred to ultracentrifuge tubes and centrifuged for 1.5 h at 160,000 g at 4 °C. The microsome pellet was resuspended and homogenized in a reaction buffer (20 mM Tris-HCl pH 7, 100 mM NaCl, 10% glycerol).

For the preparation of DHSA **93** we used an almond β -glucosidase (≥ 6 U mg^{-1} , Sigma Aldrich) as follows. MPLC fractions (50 ml each) containing Glc-DHSA **103** were evaporated using a rotary evaporator at 40 °C, lyophilized and reconstituted in 15 ml McIlvaine buffer (20 mM, pH 5.0). Reactions were performed in separate 20 ml vials incubated at 45 °C for 24 h. Each reaction consisted of 6 ml of McIlvaine buffer (pH 5.0), 3 ml of 0.1 mg ml^{-1} of the almond β -glucosidase solution in McIlvaine buffer, and 1.5 ml of the fractions containing Glc-DHSA **103**. The metabolites were extracted using 3 volumes of ethyl acetate:diethyl ether 1:1, evaporated using a rotary evaporator and reconstituted in 5 ml methanol. The products from the reaction contained a mixture of both glucosylated and non-glucosylated metabolites. DHSA **93** was therefore purified using System 2, and reconstituted in 100 μl methanol for the enzymatic assay. The purified DHSA **93** was analyzed via UPLC–qTOF to verify that the purified fraction did not contain Glc-DHSA **103**.

PT enzymatic assays were carried out as described previously for CsPT4⁸ with some modifications. The microsomes from yeasts expressing the HuPTs were resuspended in 3.3 ml buffer (10 mM Tris-HCl, 10 mM MgCl₂, pH 8.0, 10% glycerol) and homogenized with a tissue grinder. The enzyme assays were performed in 50 μl with 2 μl of the respective membrane preparations dissolved in the reaction buffer (50 mM Tris-HCl, 10 mM MgCl₂, pH 8.0), with 500 μM of the aromatic acceptor (OA **92**, VA, DHSA **93**, PCP **95**, naringenin chalcone **97** or pinocembrin chalcone **100**) and 500 μM of the isoprenoid (IPP, GPP or FPP). Samples were incubated for 1 h at 30 °C. Kinetic assays were similarly performed with 1 mM of GPP and varying (0.5 μM –1.5 mM) concentrations of OA **92**, with 15 min incubation at 30 °C. Samples were extracted with 100 μl ethanol followed by vortexing and centrifugation. The supernatant was filtered and analyzed via UPLC–qTOF as for the terpenophenols (UPLC Method 1).

UGT enzyme assays

The UGT enzyme assays were performed as described by Cai et al. (2021)⁷⁶ with some modifications. UGT assays using different aromatic substrates were performed by mixing 1.5 μl of the UDP-Glc solution (80 mM, final concentration: 2.5 mM), 27.5 μl Tris buffer (100 mM, pH 8.0), 1 μl of each of the substrates (50 mM, final concentration: 1 mM) and 20 μl of the lysate enzyme solution. The reactions were incubated at 30 °C for 1 h. Reactions were stopped by extraction with 100 μl methanol, vortexing and centrifugation at 15,000 g for 10 min. The supernatant was filtered and analyzed via UPLC–qTOF using UPLC Method 2. The assay with the purified UGTs was performed by mixing 2 μl of the cannabinoid acceptors (OA **92**, DHSA **93**, CBGA **1**, heliCBGA **2**, CBDA, Δ^9 -THCA, CBCA **15**, olivetol, CBG, CBD or Δ^9 -THC, PCP **95**, naringenin chalcone **97** or pinocembrin chalcone **100**) in the presence of 1.5 μl UDP-Glc 80 mM, 46.5 μl Tris buffer (100 mM, pH 8.0) and 1 μl of each enzyme. The metabolites were extracted and analyzed as previously described. Kinetic assays were performed with the purified enzymes (1.5 $\mu\text{g } \mu\text{l}^{-1}$) dissolved in 45 μl Tris buffer (100 mM, pH 8.0) and substrates were added using varying (0.5 μM –3 mM) and constant (1 mM) concentrations of OA **92** and UDP-Glc and the total reaction volume was 50 μl . To stop the reactions, 100 μl methanol was added to each tube, and the metabolites were extracted and analyzed as previously described.

AAT enzyme assays

Recombinant AAT assays were performed by mixing 7 μl of the cannabinoid acceptors (CBGA **1** or heliCBGA **2**, 1 mg ml^{-1}) with 58 μl of a potassium phosphate buffer (100 mM, pH 7.4), 5 μl of the acyl-CoA donors (butyryl-CoA or, hexanoyl-CoA, 10 mM) and 30 μl of the enzyme solutions. The reactions were incubated at 30 °C for 3 h. Samples were extracted with 100 μl ethanol followed

by vortexing and centrifugation. The supernatant was filtered and used for UPLC–qTOF analysis using a similar column, mobile phase and MS parameters as previously described for terpenophenols and UPLC Method 6. The assay with the purified HuCBAT5 enzyme was performed by mixing 2 μ l of the cannabinoid acceptors (OA **92**, CBGA **1**, heliCBGA **2**, CBDA, Δ^9 -THCA or CBCA **15**) with 2 μ l of the acyl-CoA donors (butyryl-CoA, iso-butyryl-CoA, hexanoyl-CoA, iso-valeryl-CoA, or acetyl-CoA, 10 mM), 44 μ l of a potassium phosphate buffer (100 mM, pH 7.4), and 2 μ l of the purified HuCBAT5 enzyme solution. The reactions were incubated at 30 °C for 3 h. To stop the reactions, 50 μ l ethanol was added to each tube and the acylated metabolites were extracted and analyzed via UPLC–qTOF as for the terpenophenols (UPLC Method 1) in both MS and MS/MS modes. Extracted ion chromatograms using the major products were selected from the LC-MS/MS analyses as follows: cannabinoid acceptors without CoAs: OA **92**>179.107, CBGA **1**, CBCA **15**>191.107, heliCBGA **2**>225.092, CBDA, Δ^9 -THCA>245.154; acylated cannabinoids: OA **92**>179.107, CBGA **1**>231.102, heliCBGA **2**>265.086, CBDA>245.154, Δ^9 -THCA>245.154, CBCA **15**>191.107).

11. UPLC chromatographic conditions

UPLC Method 1: Initial conditions were 40% B for 1 min, raised to 100% B until 23 min, held at 100% B for 3.8 min, decreased to 40% B until 27 min, and held at 40% B until 29 min for re-equilibration of the system. The flow rate was 0.3 ml min⁻¹, and the column temperature was kept at 35 °C.

UPLC Method 2: Initial conditions were from 0% to 28% B over 22 min, raised to 100% B until 36 min, held at 100% B for 2 min, decreased to 0% B until 38.5 min, and held at 40% B until 40 min for re-equilibration of the system. The flow rate was 0.3 ml min⁻¹, and the column temperature was kept at 35 °C.

UPLC Method 3: Initial conditions were 57% B raised to 85% B until 4 min, raised to 100% B until 4.2 min, held at 100% B until 6 min, decreased to 67% B until 6.2 min, and held at 67% B until 7 min for re-equilibration of the system. The flow rate was 0.6 ml min⁻¹, and the column temperature was kept at 40 °C.

UPLC Method 4: Initial conditions were 1% B raised to 35% B until 10.5 min, and then raised to 100% B until 11 min, held at 100% B for 1 min, decreased to 1% B until 12.5 min, and held at 1% B until 15 min for re-equilibration of the system. The flow rate was 0.3 ml min⁻¹, and the column temperature was kept at 25 °C.

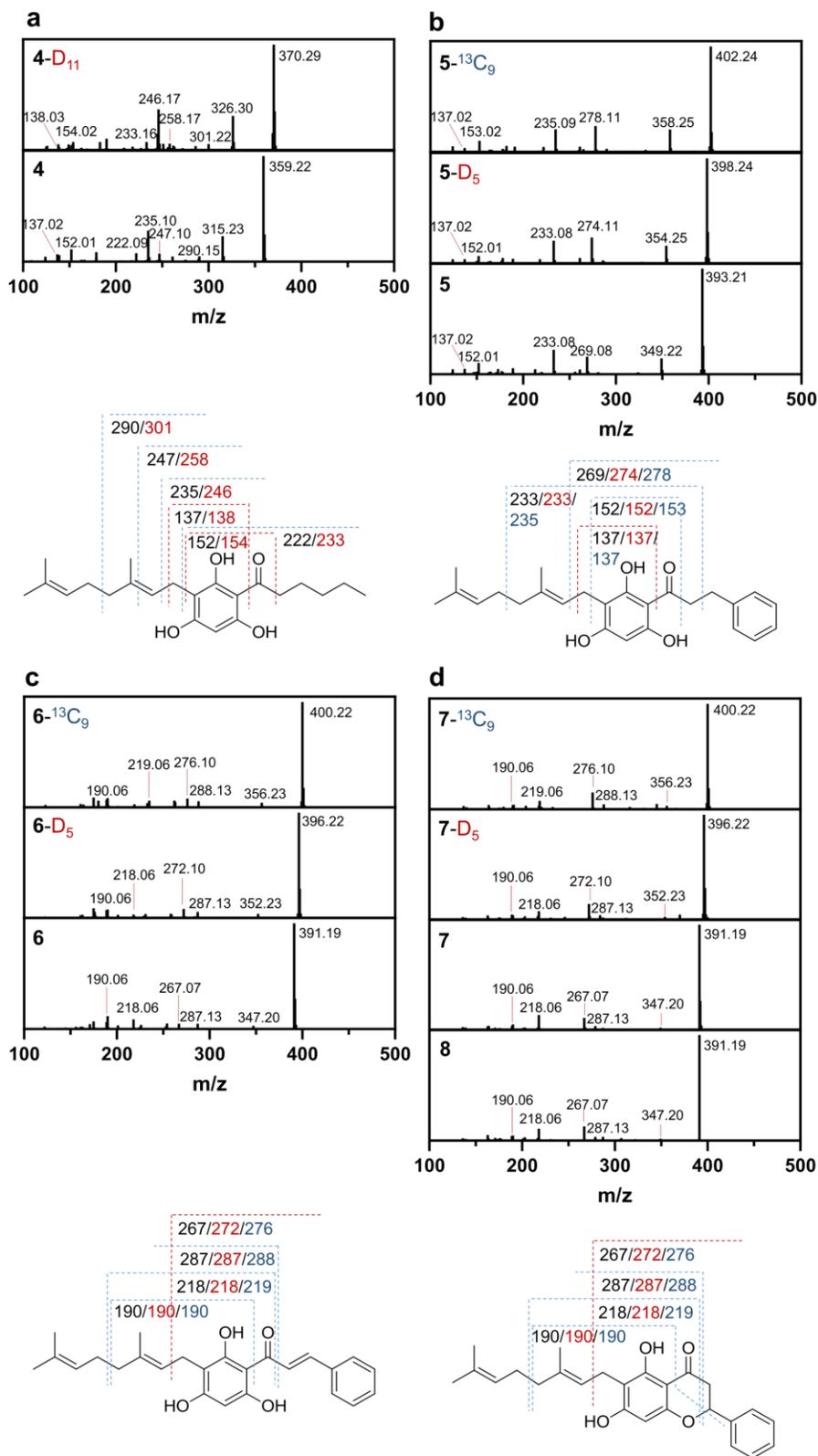
UPLC Method 5: Initial conditions were 10% B raised to 70% until 6 min, raised to 100% B until 6.2 min, held at 100% B until 8 min, decreased to 10% B until 8.5 min, and held at 10% B until 11 min for re-equilibration of the system. The flow rate was 0.3 ml min⁻¹, and the column temperature was kept at 35 °C.

UPLC Method 6: Initial conditions were 40% B for 1 min, raised to 100% B until 14 min, held at 100% B for 3.8 min, decreased to 40% B until 18 min, and held at 40% B until 20 min for re-equilibration of the system. The flow rate was 0.3 ml min⁻¹, and the column temperature was kept at 35 °C.

12. Operational parameters of the Orbitrap IQ-X Tribrid MS

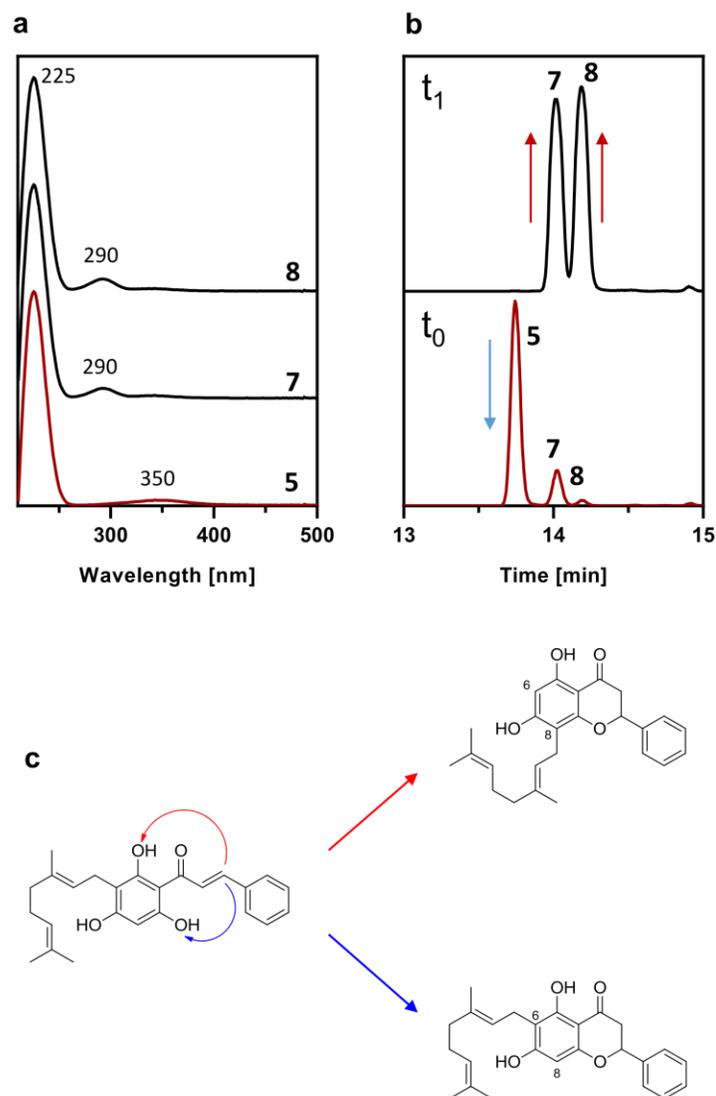
The source parameters were: sheath gas flow rate, auxiliary gas flow rate and sweep gas flow rate: 45, 10 and 1 arbitrary units, respectively; vaporizer temperature: 300 °C; ion transfer tube temperature: 275 °C; spray voltage: 2.3 kV. The instrument was operated in full MS¹ with data dependent MS/MS (MS-dd-MS²). Data acquisition in full MS¹ mode was 60,000 resolution, the

scan range 100–1000 m/z , normalized automatic gain control (AGC) target of 25% and a maximum injection time (IT) of 50 ms. Data acquisition in dd-MS² mode was with 15,000 resolution, a normalized AGC target of 20%, maximum IT of 150 ms, isolation window of 1.5 m/z and normalized collision energy of 40.

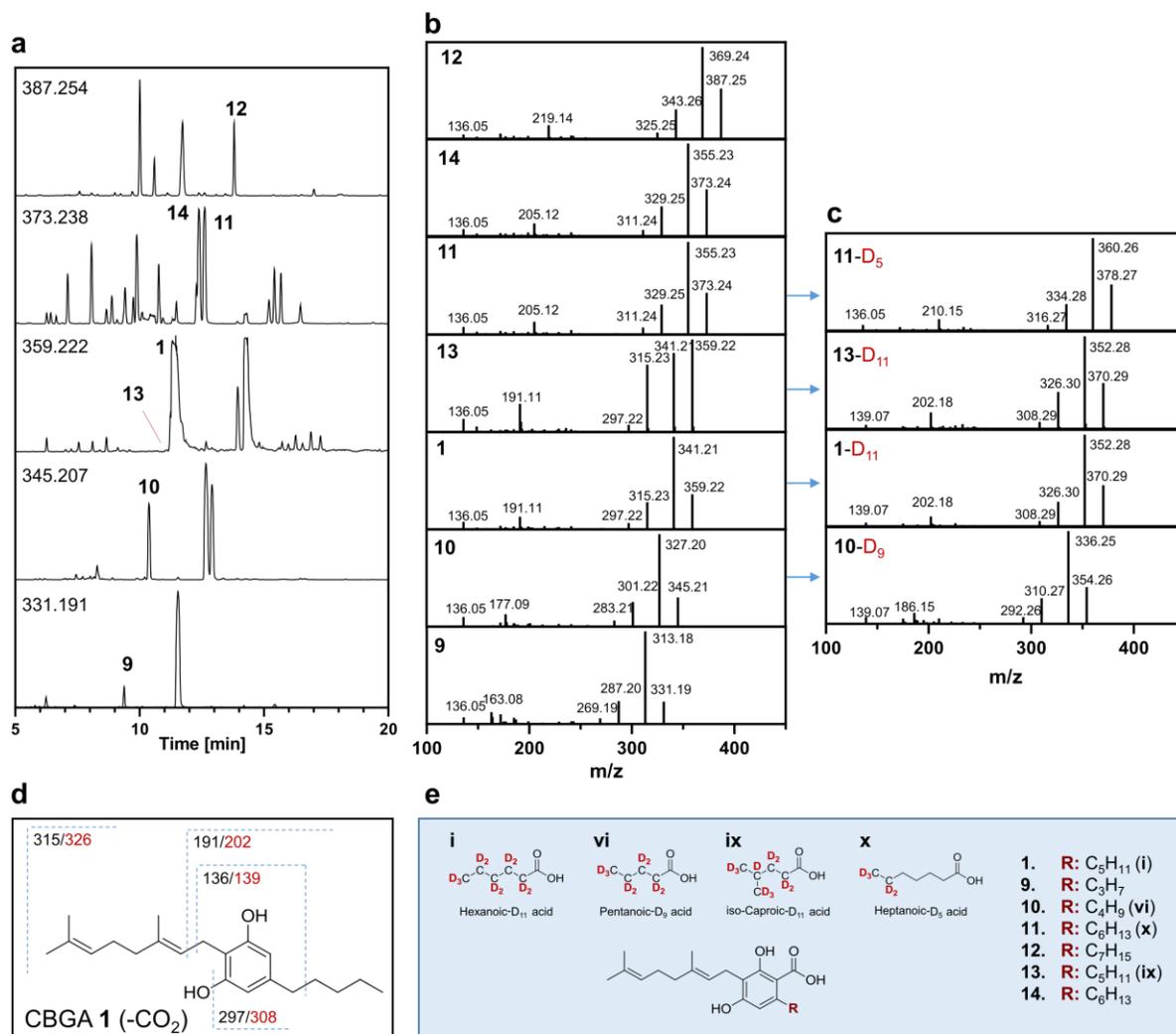


The legend of Supplementary Fig. 1 appears on the next page.

Supplementary Fig. 1. Identification and labeling of major additional terpenophenols from different classes in *H. umbraculigerum*. MS/MS spectra in negative polarities of unlabeled and isotopically labeled **a.** geranylphlorocaprophenone **4**, **b.** geranyl-2',4',6'-trihydroxydihydrochalcone **5**, **c.** geranyl-pinocembrin chalcone **6**, **d.** 6- and 8-geranylpinocembrin **7** and **8**, and their predicted fragmentation structures. The labeled **7** and **8** metabolites had similar MS/MS spectra, therefore we include here only the spectra of **7**. Geranylphlorocaprophenone **4** was labeled following feeding with hexanoic-D₁₁ acid, and the chalcones and flavanones were labeled following dual feeding with phenylalanine-D₅ or phenylalanine-¹³C₉. Fragments colored in red or blue correspond to the *m/z* of the specific fragment in the labeled metabolite. The structures of geranylphlorocaprophenone **4** and geranyl-pinocembrin chalcone **6** were further confirmed by NMR (Supplementary NMR Data 3,4).

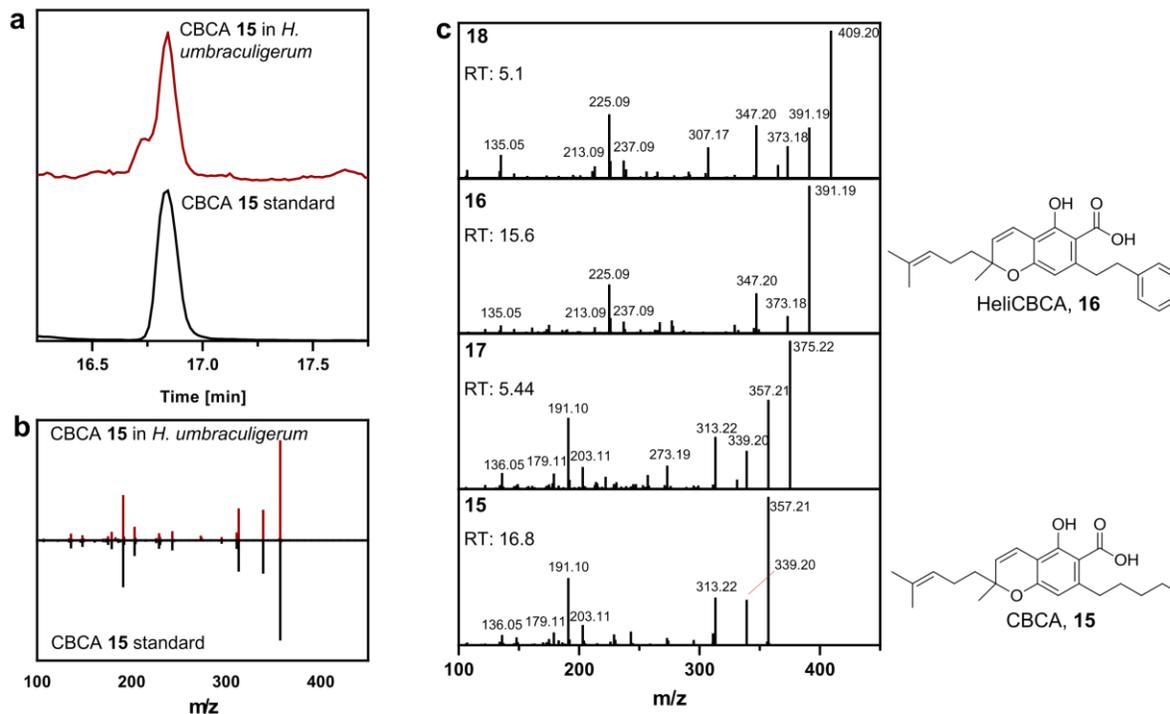


Supplementary Fig. 2. Differentiation between chalcones and flavanones via UV absorbance and thermal stability. **a.** Photo-diode array (PDA) spectra of geranyl-2',4',6'-trihydroxydihydrochalcone **5** versus 6/8-geranylpinocembrin **7** and **8**. Geranyl-2',4',6'-trihydroxydihydrochalcone **5** has a minor absorption band at 340–390 nm, whereas 6/8-geranylpinocembrin **7** and **8** absorb in the range of 270–310 nm. **b.** Extracted ion chromatogram of $m/z [M-H]^- = 391.191$ Da before (t_0) and after (t_1) heating of *H. umbraculigerum* leaves (120 °C for 1 h). As shown, geranyl-2',4',6'-trihydroxydihydrochalcone **5** disappears following heating while the relative abundances of **7** and **8** increase. **c.** Prenylchalcones can isomerize to yield flavanones with the prenyl group in one of two possible positions.

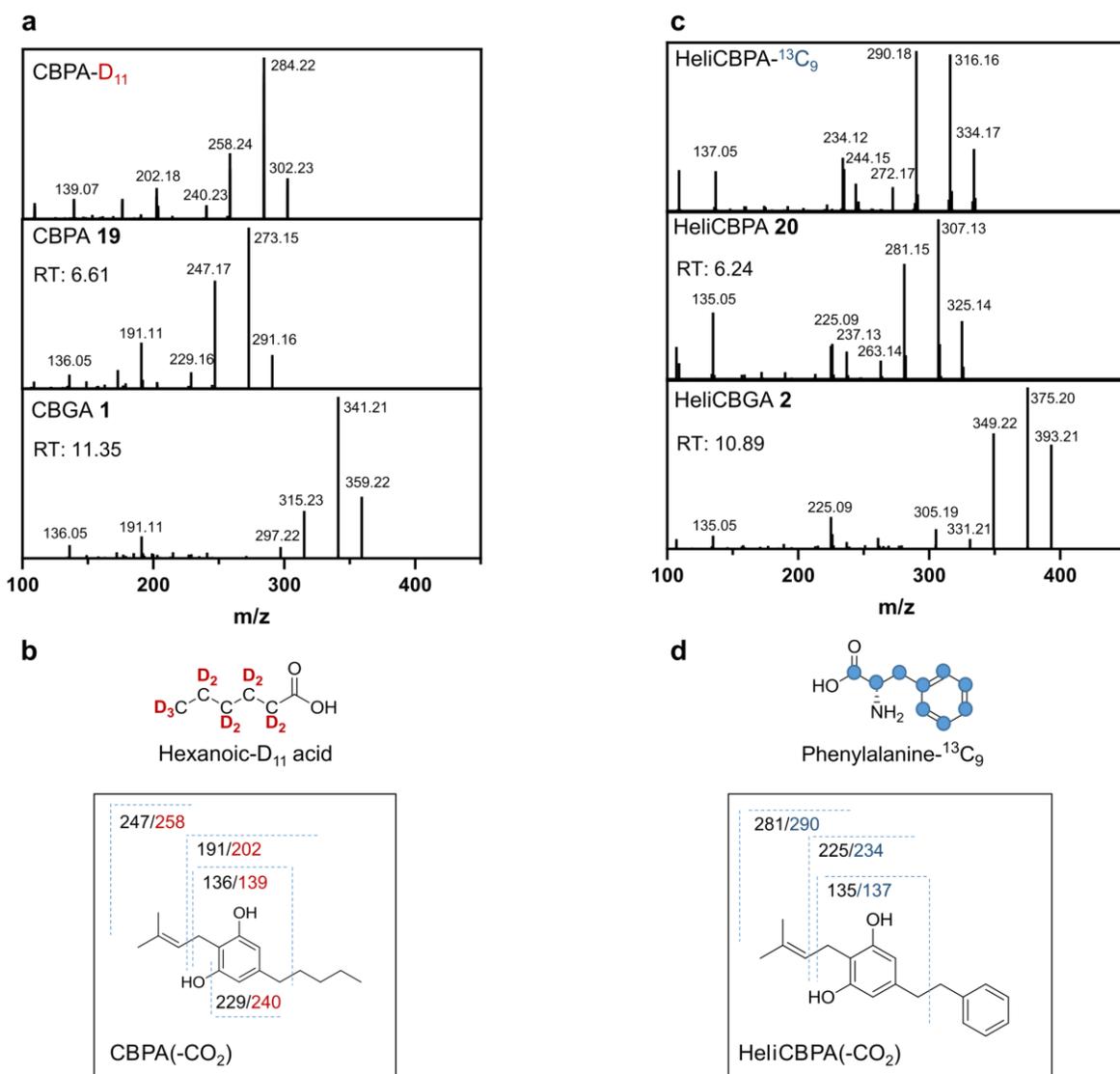


Supplementary Fig. 3. Identification of CBGA-type alkyl homologues in *H. umbraculigerum*.

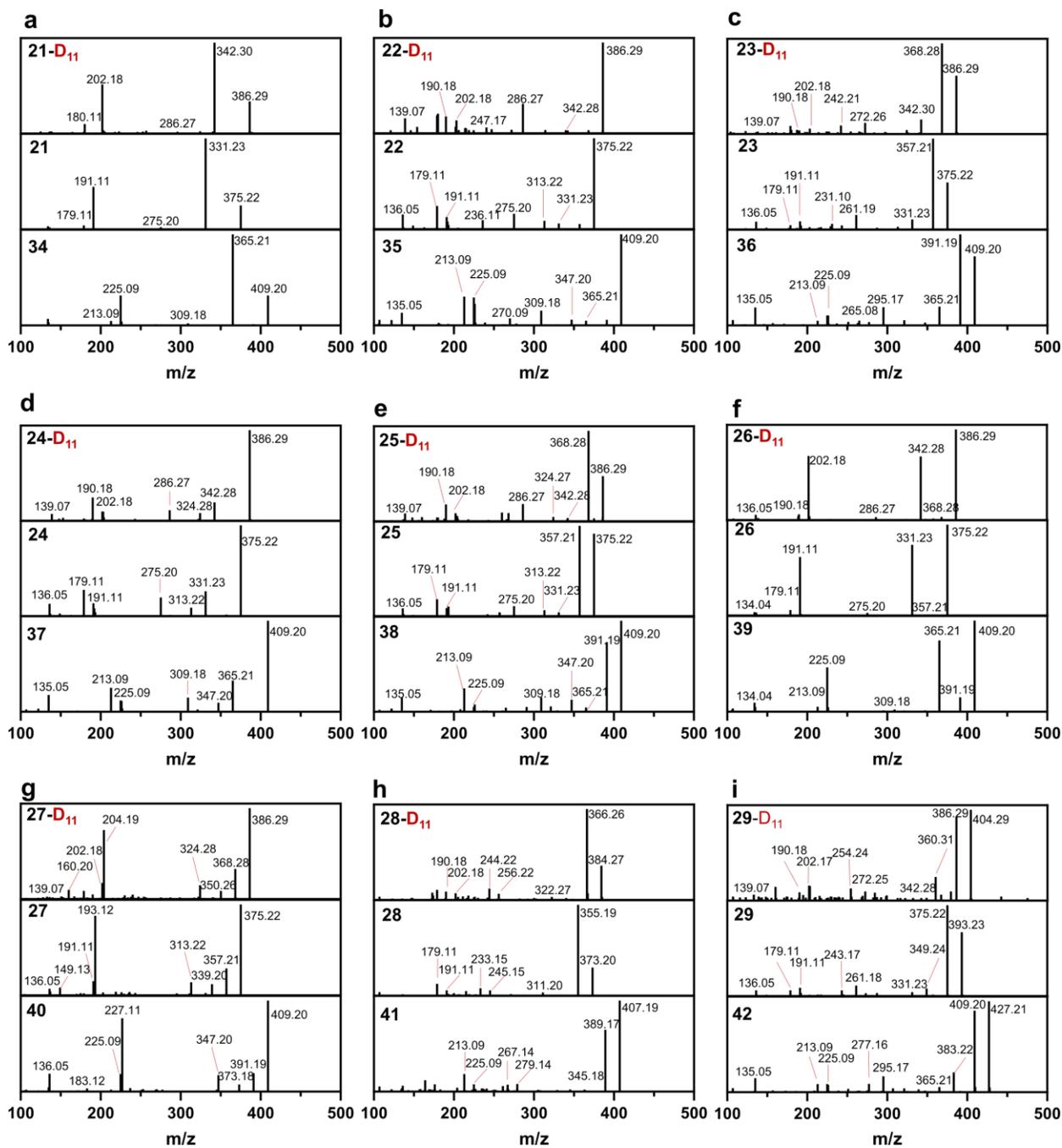
a. Extracted ion chromatograms of $[M-H]^- = 331.191, 345.207, 359.222, 373.238,$ and 387.254 Da. The marked peaks in each chromatogram correspond with the detected **1,9-14** cannabinoids. As shown, the alkyl homologues elute from the column in order of chain length as a result of increasing lipophilicities. MS/MS spectra in negative polarity of **b.** unlabeled and **c.** isotopically labeled metabolites. Not all labeled metabolites were detected probably due to low abundance in the plant. **d.** Suggested fragmentation structure of CBGA **1** according to MS/MS spectra and labeling. Fragments colored in red correspond to the m/z of the specific fragment in the cannabinoid labeled with hexanoic-D₁₁ acid. For all alkyl homologues, an appropriate m/z shift in the MS/MS spectra of all the product ions that include the alkyl chain was observed. **e.** Structures of the observed cannabinoids and isotopically labeled precursors of the identified metabolites.



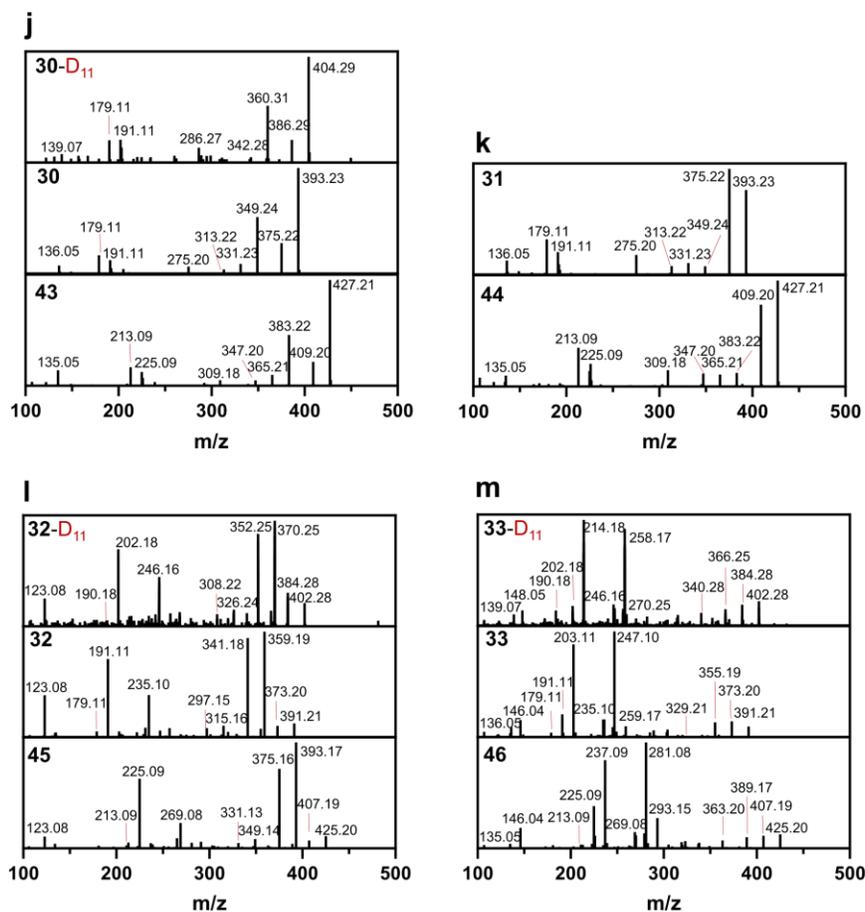
Supplementary Fig. 4. Identification of cannabichromenic acid (CBCA 15) and CBCA-like metabolites in *H. umbraculigerum*. **a.** Extracted ion chromatograms ($[M-H]^- = 357.207$ Da) and **b.** MS/MS spectral matching of CBCA 15 standard versus a *H. umbraculigerum* leaf extract. **c.** Identification of OH-CBCA 17, heliCBCA 16, and OH-heliCBCA 18 by comparison of MS/MS fragmentations. As shown, heliCBCA 16 exhibited analogous fragmentation patterns to CBCA 15 with mass shifts corresponding with the aromatic versus five-carbon tails (m/z difference of 33.984 Da), and both hydroxylated metabolites exhibited similar fragmentation patterns with the addition of one hydroxyl. In addition, hydroxylation reduced the retention time (RT) considerably, and the aromatic metabolites eluted at shorter RT compared to the aliphatic ones.



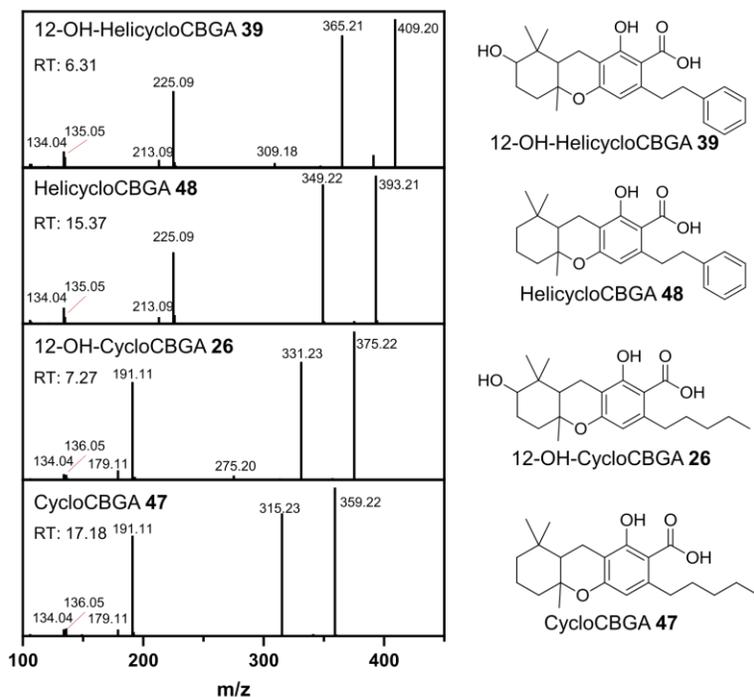
Supplementary Fig. 5. Identification of cannabiprenylic acid (CBPA 19) and heli-cannabiprenylic acid (heliCBPA 20) in *H. umbraculigerum*. **a.** MS/MS spectra in negative polarity of unlabeled and isotopically labeled CBPA 19 and **c.** heliCBPA 20 and their suggested fragmentation structures (**b** and **d**, respectively). CBPA 19 was labeled following feeding with hexanoic-D₁₁ acid (fragments colored in red), and heliCBPA 20 was labeled following feeding with phenylalanine-¹³C₉ (fragments colored in blue). Identification of CBPA 19 and heliCBPA 20 was by comparison to the MS/MS fragmentation patterns of CBGA 1 and heliCBGA 2, respectively. As shown, the isoprenylated metabolites exhibited analogous fragmentation patterns to the geranylated ones, with mass shifts corresponding with one prenyl group (m/z difference of 68.063 Da). In addition, the isoprenylated metabolites eluted from the UPLC column several minutes before the geranylated ones, as a result of increasing lipophilicity with prenylation, and heliCBPA 20 eluted prior to CBPA 19 as observed for the aromatic versus five-carbon tails. RT, retention time.



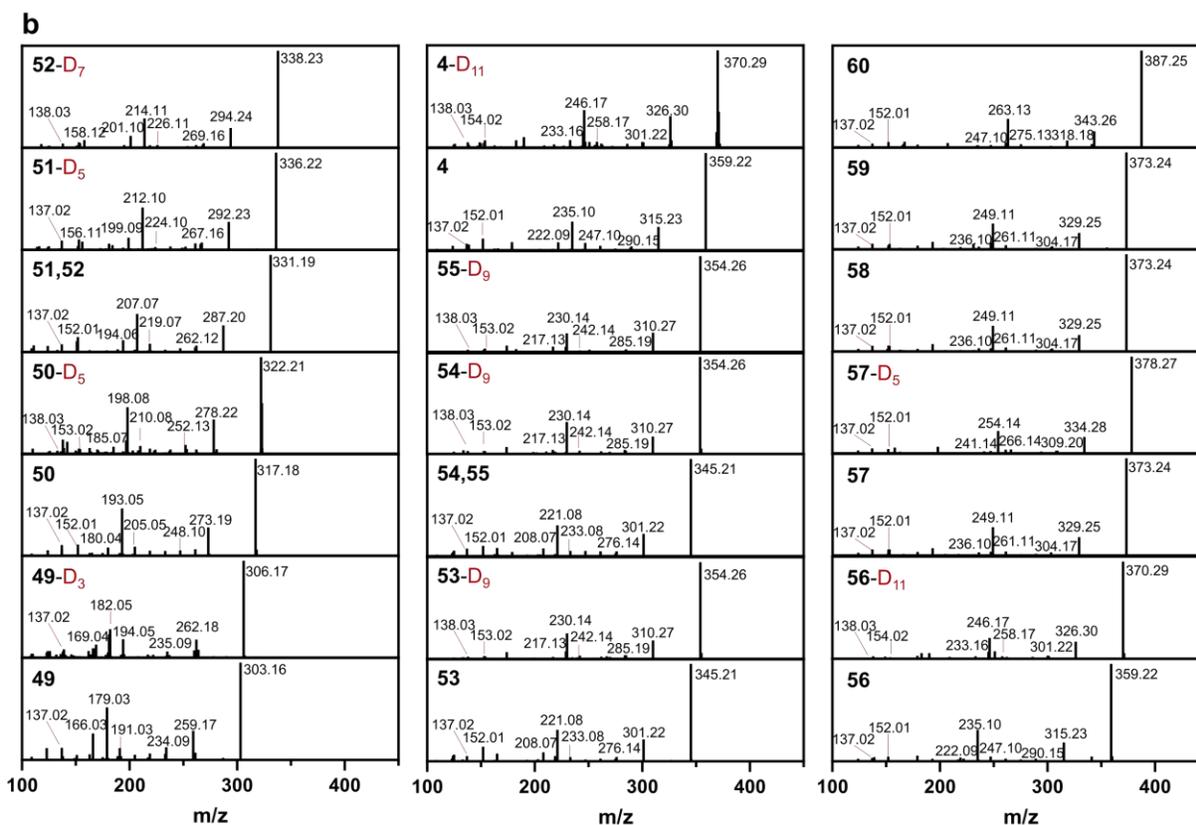
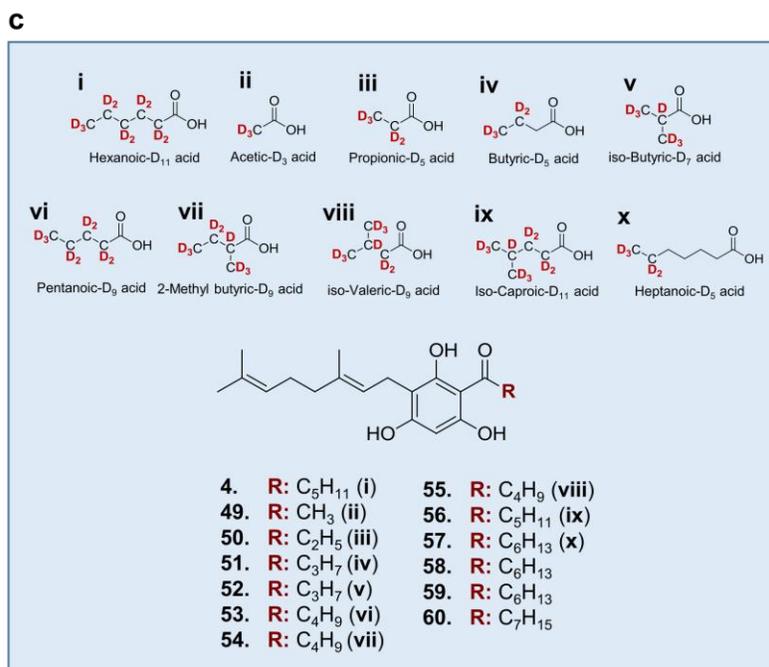
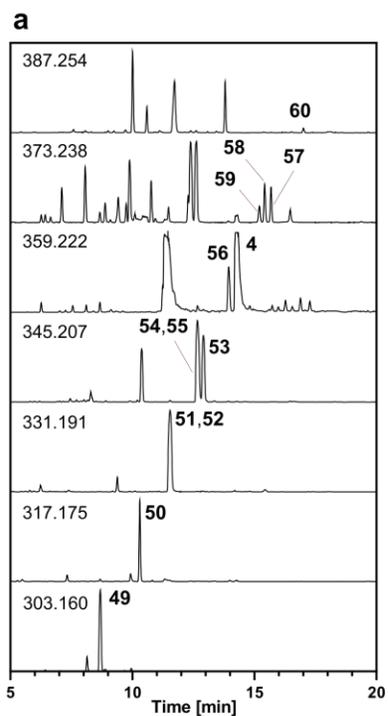
The continuation and legend of Supplementary Fig. 6 appears on the next page.



Supplementary Fig. 6. Identification of hydroxylated cannabinoids and amorfrutins in *H. umbraculigerum*. MS/MS spectra in negative polarity of unlabeled and isotopically labeled **a.-h.** hydroxylated and **i.-m.** dihydroxylated five-carbon cannabinoids (**21-33** labeled with hexanoic-D₁₁ acid, **31-D₁₁** was not observed probably due to low abundance), and their corresponding amorfrutins (**34-46**). The amorfrutins were identified by similar fragmentation patterns as the cannabinoids shifted in masses in correspondence to the aromatic versus five-carbon tails (m/z difference of 33.984 Da). The aralkyl peaks were observed at constant time intervals from the alkyl ones with similar fragmentation patterns.

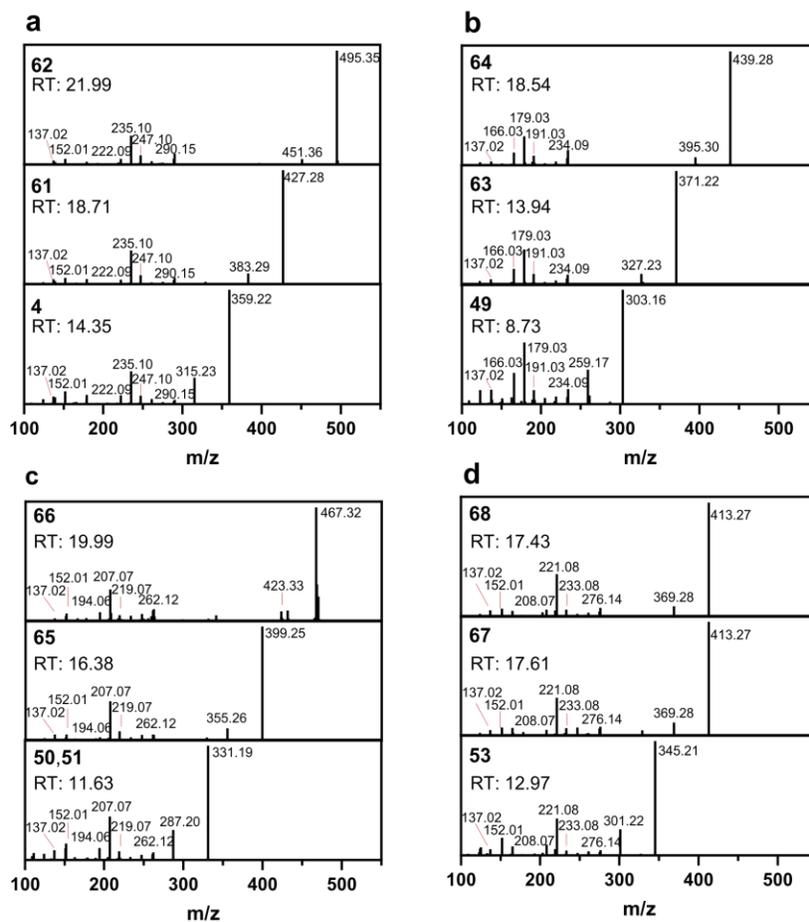


Supplementary Fig. 7. Identification of cyclocannabigerolic acid (cycloCBGA **47) and cyclocannabigerolic acid-like metabolites in *H. umbraculigerum*.** Metabolite 12-OH-cycloCBGA **26** was purified and identified by NMR as a novel tetrahydroxanthane cannabinoid (12-OH-cyclocannabigerolic acid, Supplementary NMR Data 5). According to MS/MS fragmentation patterns and relative retention times (RTs), we also putatively identified cyclocannabigerolic acid (cycloCBGA **47**) and analogous amorfrutin types [12-OH-helicyclocannabigerolic acid **39** and helicyclocannabigerolic acid **48**, respectively].

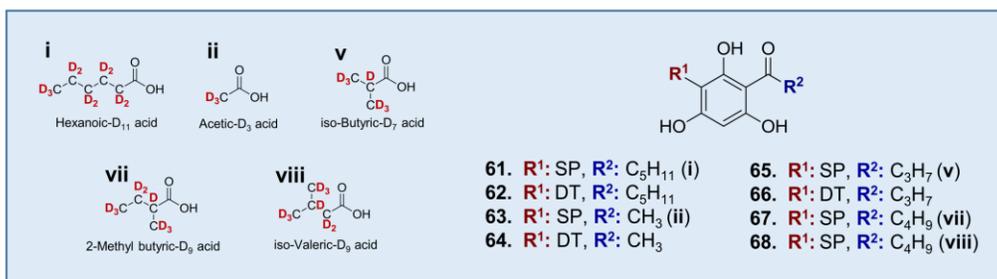


The legend of Supplementary Fig. 8 appears on the next page.

Supplementary Fig. 8. Identification of geranylated-acyl homologue phloroglucinoids in *H. umbraculigerum*. **a.** Extracted ion chromatograms of $[M-H]^- = 303.160, 317.175, 345.207, 359.222, 373.238, \text{ and } 387.254$ Da. The marked peaks in each chromatogram correspond with the detected **4, 49-60** phloroglucinoids. As shown, the alkyl homologues elute from the UPLC column in order of chain length because of increasing lipophilicities. **b.** MS/MS spectra in negative polarity of unlabeled and isotopically labeled phloroglucinoids. Not all labeled phloroglucinoids were detected probably due to low abundance in the plant. The suggested fragmentation structure of geranylphlorocaprophenone **4** according to MS/MS spectra and labeling appears in Supplementary Fig. 1a. For all alkyl homologues, an appropriate m/z shift in the MS/MS spectra of all the product ions that include the alkyl chain was observed. **c.** Structures of the observed phloroglucinoids and isotopically labeled precursors of the identified metabolites. To confirm the assignment of phloroglucinoids, metabolite **54** was further purified and its ^1H and ^{13}C NMR spectra were compared to the spectra of geranylphlorocaprophenone **4** (Supplementary NMR Data 6).

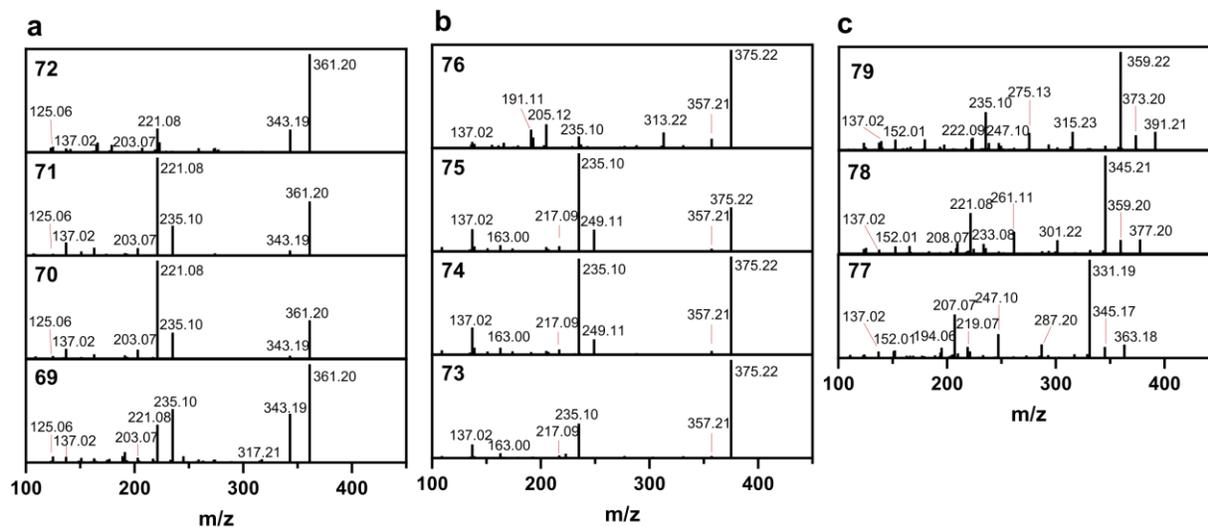


e

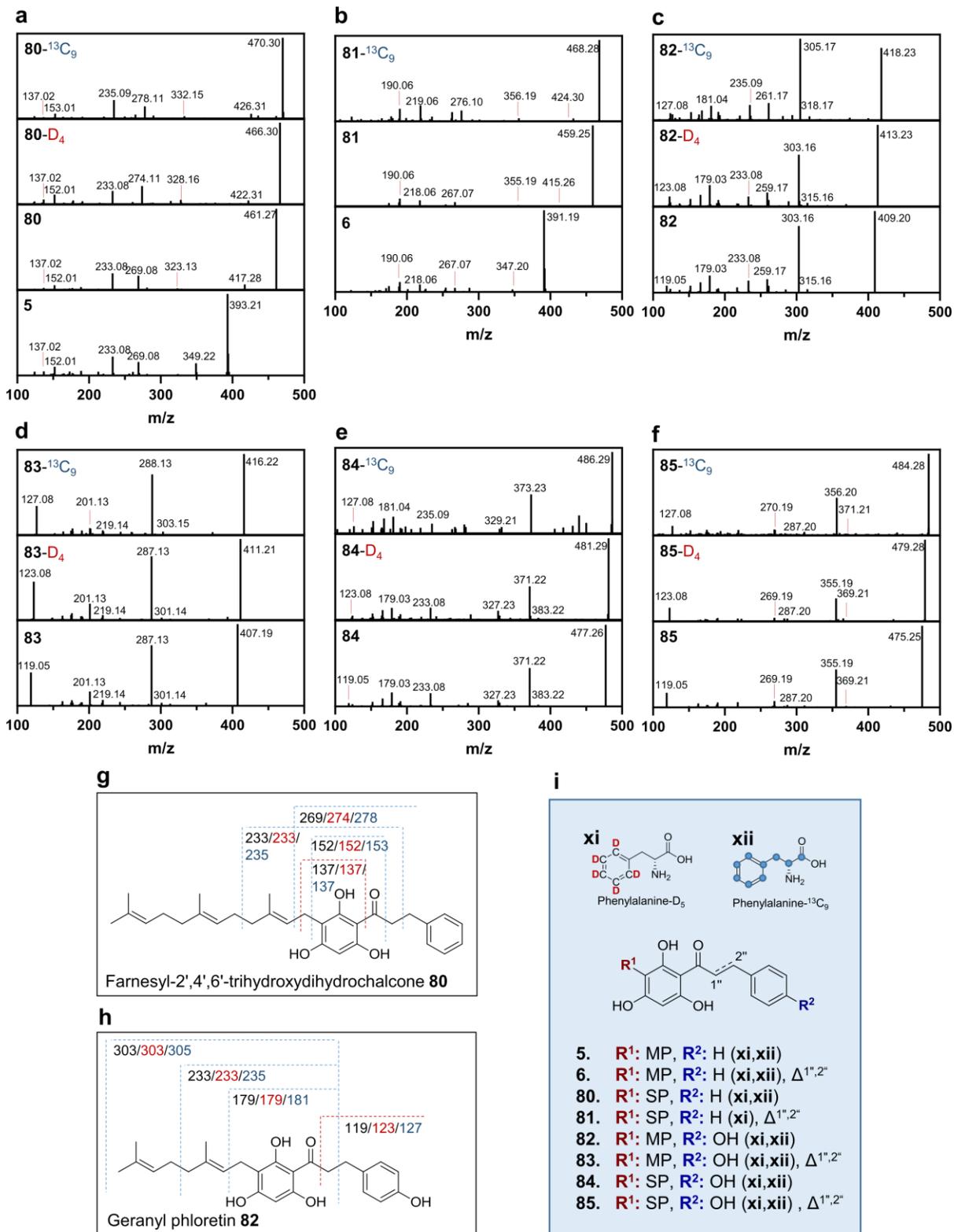


The legend of Supplementary Fig. 9 appears on the next page.

Supplementary Fig. 9. Identification of acylphloroglucinoids with different prenylations in *H. umbraculigerum*. MS/MS spectra in negative polarity of monoprenyl (as reference), sesquiprenyl, and diterpene phloroglucinoids with **a.** five- (**4**, **61** and **62**, respectively), **b.** one- (**49**, **63** and **64**, respectively), and **c.** three-carbon atom tails (**50**, **51**, **65** and **66**, respectively); and **d.** four-carbon atom tail monoprenyl (as reference) and sesquiprenyl phloroglucinoids (**53** and **67-68**, respectively). As shown, the sesquiprenyl and diterpene phloroglucinoids exhibited analogous fragmentation patterns to the geranylated ones, with mass shifts corresponding with one and two prenyl groups (m/z difference of 68.063 and 136.126 Da, respectively). In addition, the order of elution from the UPLC column for the prenylated phloroglucinoids was monoprenyl \rightarrow sesquiprenyl \rightarrow diterpene because of increasing lipophilicity with prenylation. Not all labeled metabolites were detected probably due to low abundance. **e.** Structures of the observed prenyl-acyl-phloroglucinoids and the isotopically labeled precursors of the identified metabolites. SP, sesquiprenyl; DT, diterpene.

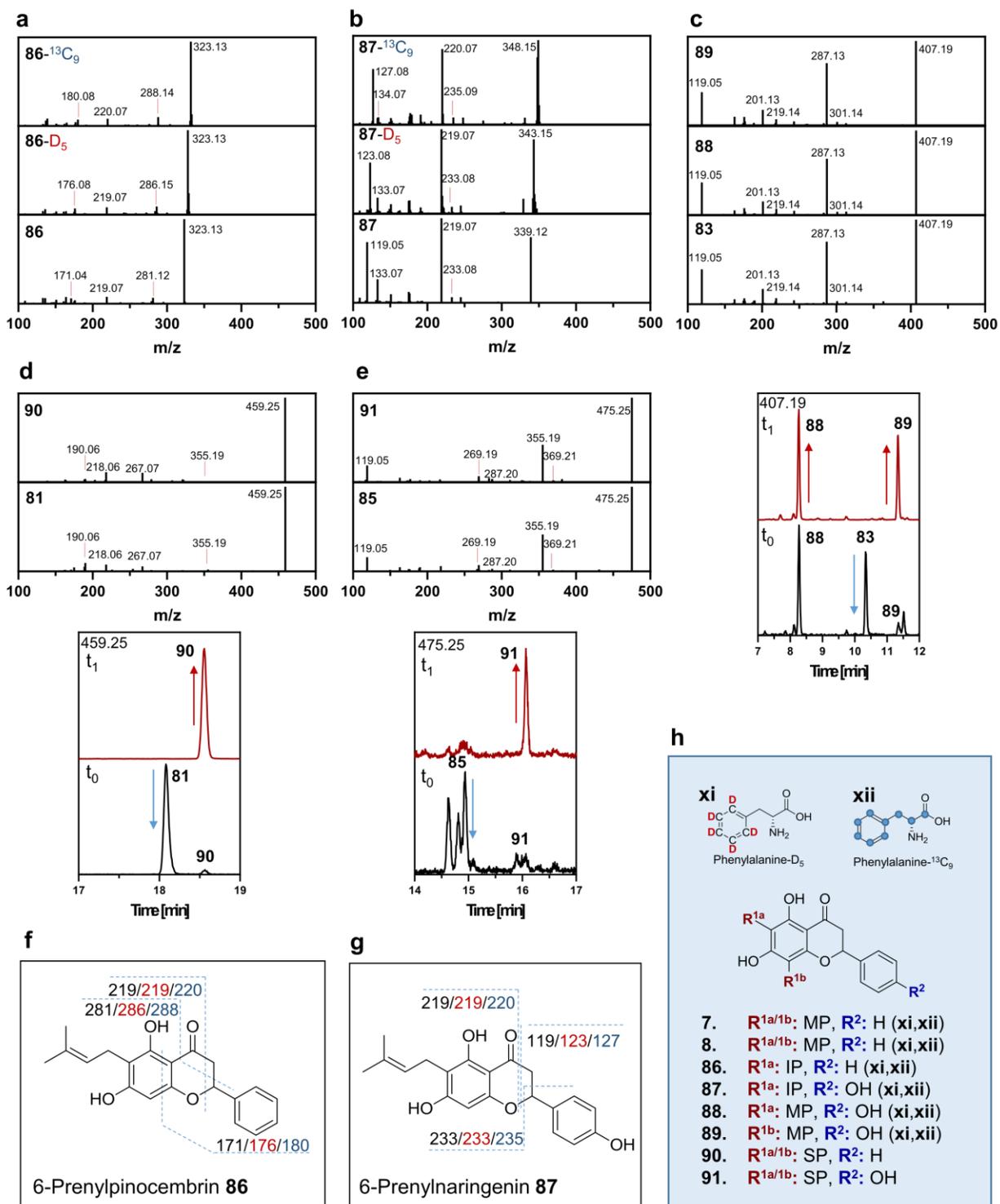


Supplementary Fig. 10. Identification of hydroxylated and dihydroxylated prenylacylphloroglucinoids in *H. umbraculigerum*. MS/MS spectra in negative polarity of hydroxylated **a.** four- (**69-72**) and **b.** five-carbon tail (**73-76**) monopenylphloroglucinoids; and dihydroxylated **c.** three- (**77**), four- (**78**), and five-carbon tail (**79**) geranylphloroglucinoids.



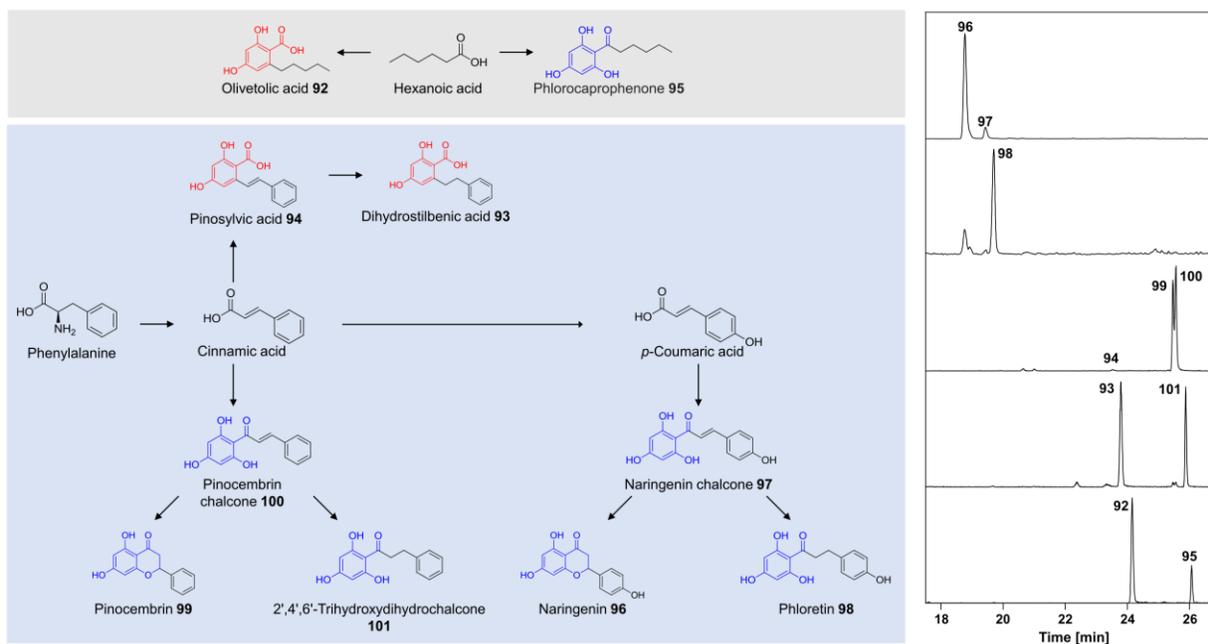
The legend of Supplementary Fig. 11 appears on the next page.

Supplementary Fig. 11. Identification of chalcones with different prenylations in *H. umbraculigerum*. **a-f** MS/MS spectra in negative polarity of unlabeled and isotopically labelled **80-85** (**5** and **6** are shown for reference). **5**, **6**, **80** and **81** originate from cinnamoyl CoA, and **82-85** from coumaroyl CoA. The chalcones were identified by specific fragmentation patterns as exemplified for **g.** farnesyl-2',4',6'-trihydroxydihydrochalcone **80** and **h.** geranyl-phloretin **82** according to MS/MS spectra and labeling [the fragmentation structure of farnesyl-2',4',6'-trihydroxydihydrochalcone **80** was determined according to that of geranyl-2',4',6'-trihydroxydihydrochalcone **5** (Supplementary Fig. 1b) and the structure of geranyl-phloretin **82** was confirmed by NMR (Supplementary NMR Data 7)]. The sesquiprenyl chalcones exhibited analogous fragmentation patterns to the geranylated ones, with mass shifts corresponding with one prenyl group (m/z difference of 68.063 Da). In addition, the order of elution from the column for the prenylated chalcones was monoprenyl→sesquiprenyl because of increasing lipophilicity with prenylation. **i.** Structures of the observed chalcones and the isotopically labeled precursors of the identified metabolites. MP, monoprenyl; SP, sesquiprenyl. Fragments colored in red or blue correspond to the m/z of the specific fragment in the chalcone labeled with phenylalanine-D₅ or phenylalanine-¹³C₉, respectively.

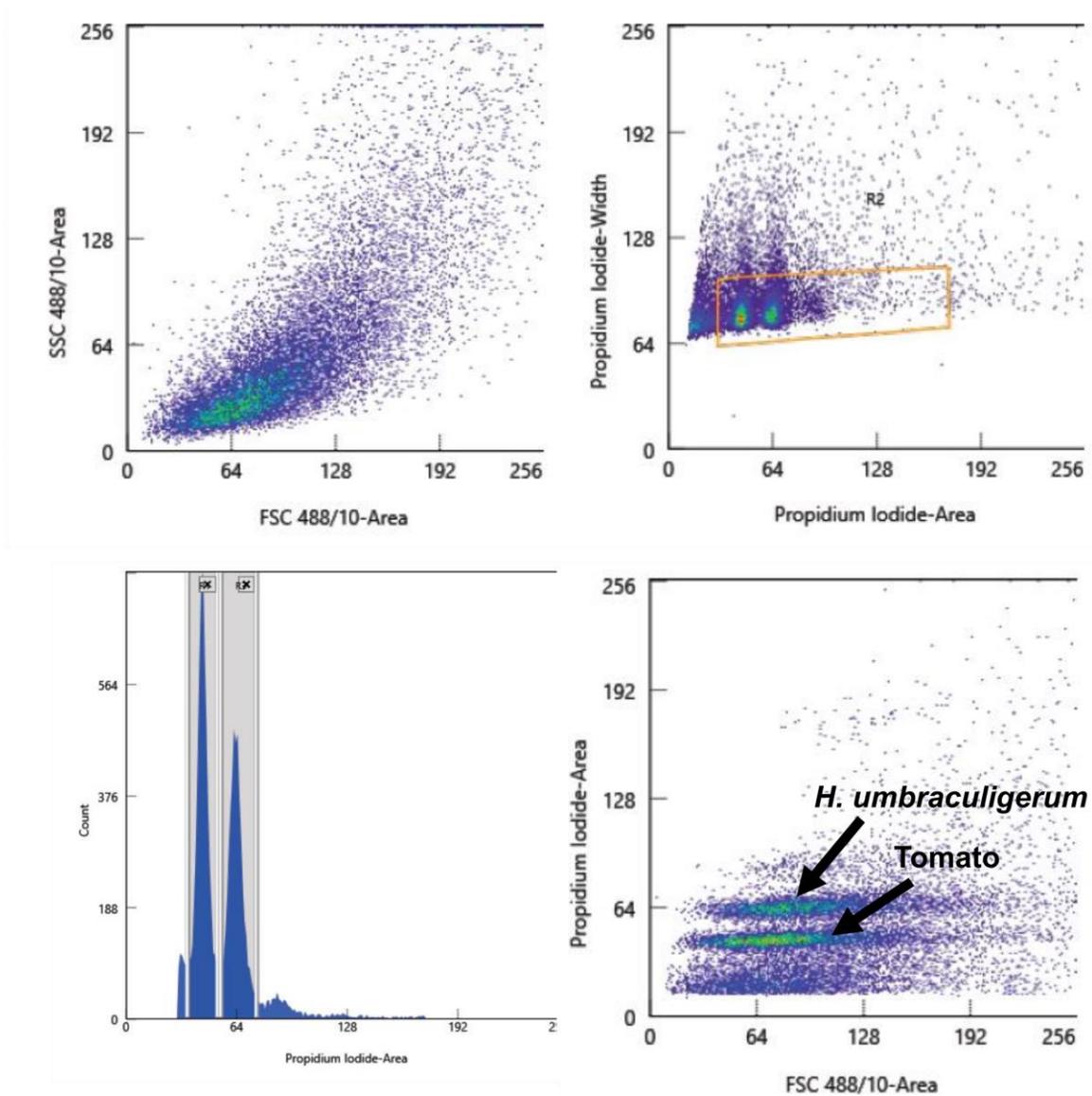


The legend of Supplementary Fig. 12 appears on the next page.

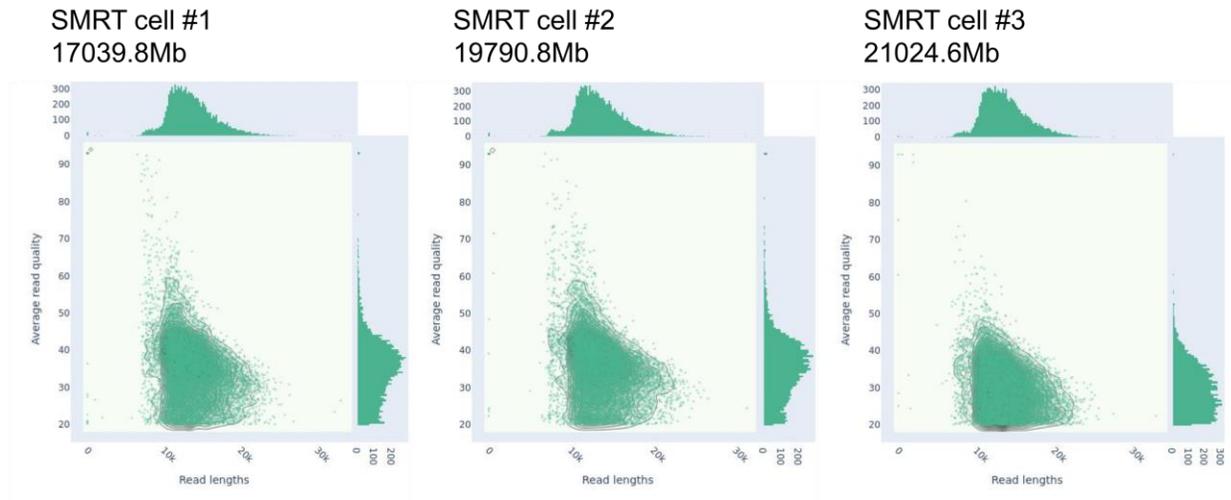
Supplementary Fig. 12. Identification of flavanones with different prenylations in *H. umbraculigerum*. **a.-e.** MS/MS spectra in negative polarity of unlabeled and isotopically labelled flavanones **86-91**. **c.-e.** show also extracted ion chromatograms of m/z $[M-H]^- = 407.19, 459.25$ and 475.25 Da before (t_0) and after (t_1) heating of *H. umbraculigerum* leaves ($120\text{ }^\circ\text{C}$ for 1 h). The flavanones were identified by UV absorbance, thermal stability and specific fragmentation patterns as exemplified for **f.** 6-prenylpinocembrin **86** and **g.** 6-prenylnaringenin **87** according to MS/MS spectra and labeling [the structures of **86-88** were confirmed by NMR (Supplementary NMR Data 8-10)]. Fragments colored in red or blue correspond to the m/z of the specific fragment in the metabolite labeled with phenylalanine- D_5 or phenylalanine- $^{13}C_9$, respectively. As shown, the chalcones **83**, **81** and **85** disappear following heating while the relative abundances of the respective flavanones increase. **h.** Structures of the observed flavanones and the isotopically labeled precursors of the identified metabolites. Prenyl chalcones can isomerize to yield flavanones with the prenyl group in one of two possible positions as shown for $R^{1a/1b}$. IP, isoprenyl; MP, monoprenyl; SP, sesquiprenyl.



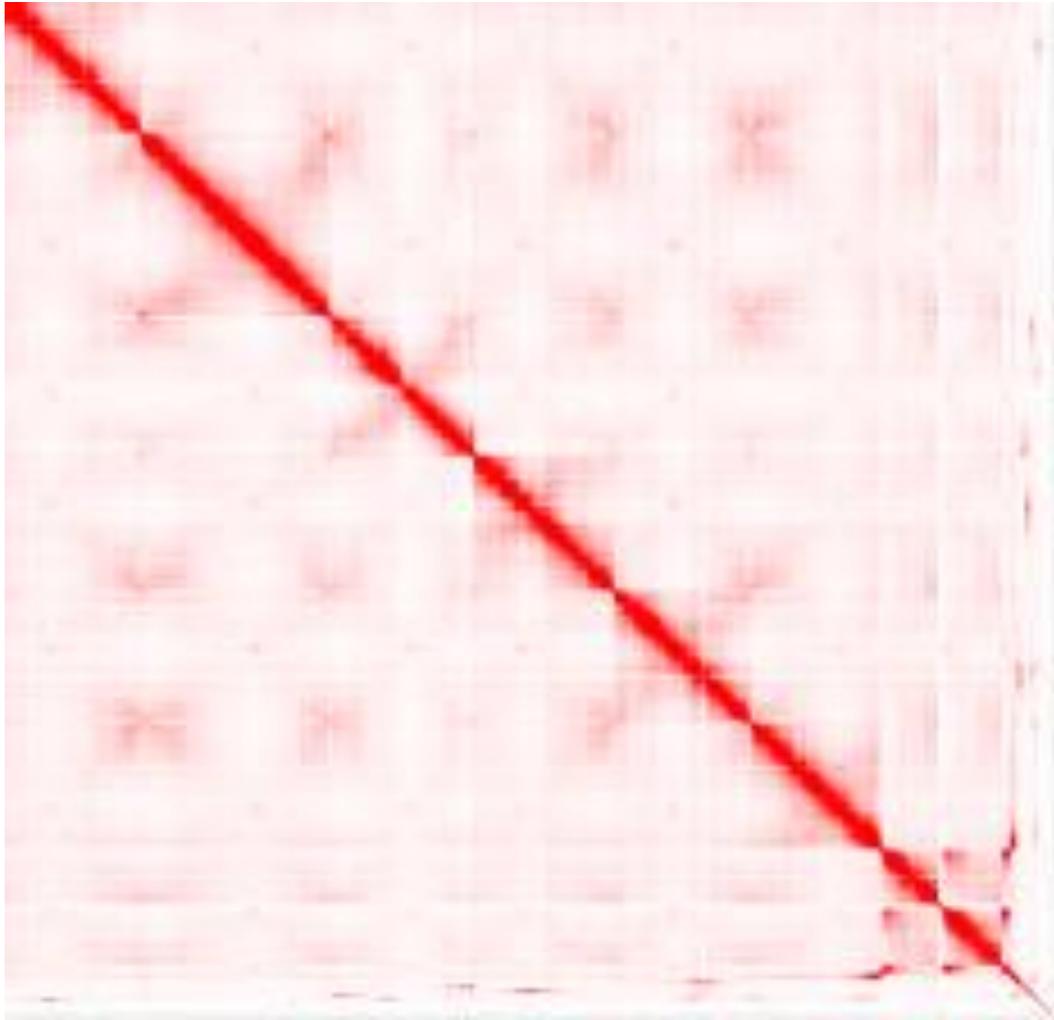
Supplementary Fig. 13. Major intermediates (92-101 before prenylation) from all the five metabolic routes identified in *H. umbraculigerum*. The intermediates were analyzed using UPLC Method 2. Identification and validation of assignment was by analytical standards and the Weizmass library of plant metabolites⁴⁶ for the available metabolites (metabolites **94** and **96-101**). Full description of the identified metabolites and their fragments appear in Supplementary Table 6. Backbone structures showing aldol or Claisen cyclizations are marked in red or blue, respectively.



Supplementary Fig. 14. Genome size estimation of *H. umbraculigerum* by flow cytometry. Preparation of nuclei from *Helichrysum* and *Tomato* was performed as described by Doležel et al. (2005)⁷⁷. The histogram of relative DNA content was obtained after flow cytometric analysis of propidium iodide-stained nuclei of *H. umbraculigerum* and *Tomato*, which were isolated, stained and analysed simultaneously. *Tomato* (haploid size 950 Mbp) served as internal reference standard. From the ratio of G1 peak means, we estimated that the haploid genome size of *H. umbraculigerum* is 1330 Mbp. Number of nuclei were labeled with Propidium Iodide at saturating concentration without wash. Labeled nuclei were analyzed immediately following labeling on ZE5 cell analyzer (Bio-Rad Laboratories). Excitation of Propidium Iodide was done with 561 nm laser line while emitted light was collected through 577/15 nm bandpass filter. Analysis was performed on FlowJo v10.7.1 (BD Biosciences).

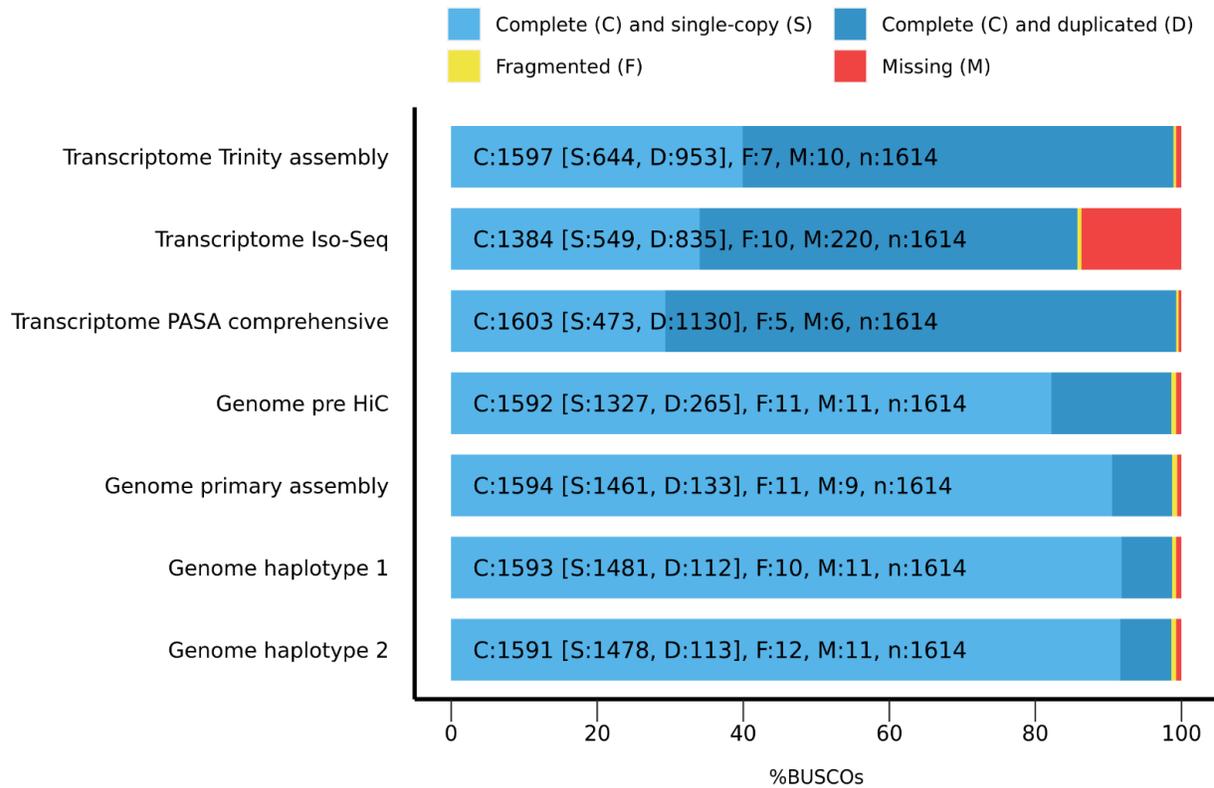


Supplementary Fig. 15. Read Length vs Read Quality kde plot of the three SMRT cells sequenced. A total of 57855.2Mb were obtained, yielding ~44x haploid genome coverage.

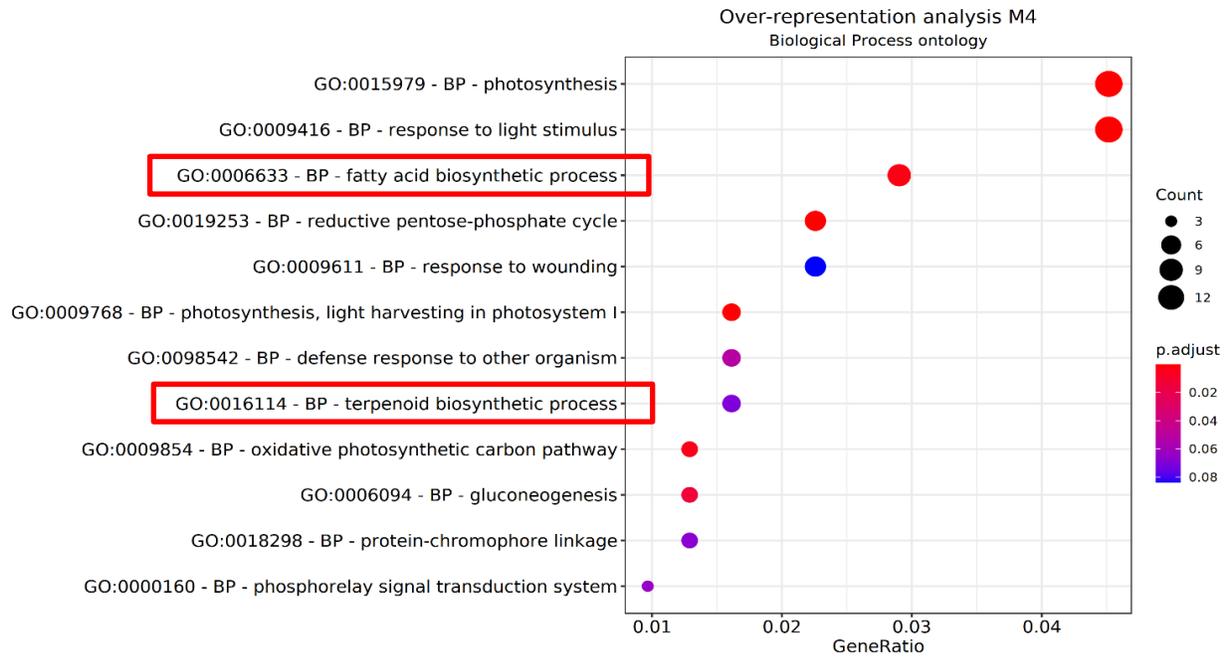


Supplementary Fig. 16. Hi-C contact frequency matrix represented as heatmap along all *H. umbraculigerum* scaffolds of the primary assembly sorted by length. Scaffolding was performed with SALSA and the .hic file was used for visualization with Juicebox.

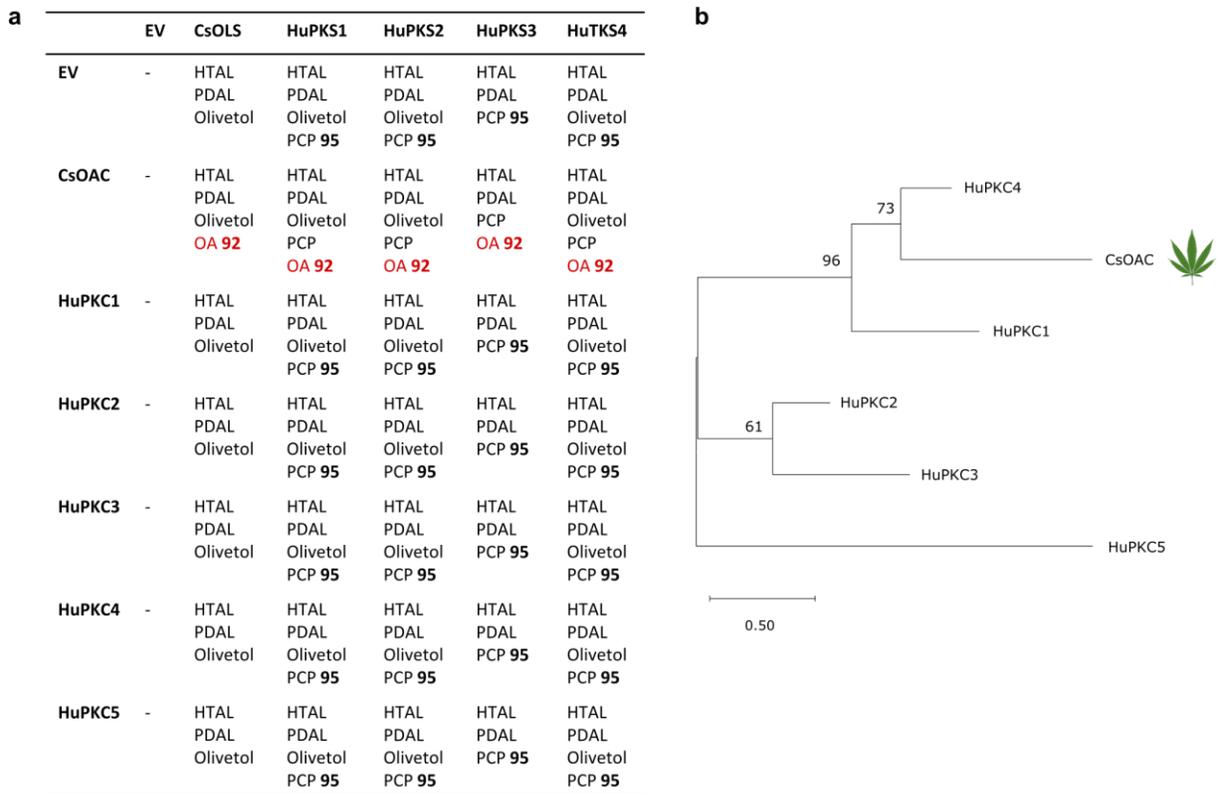
BUSCO Assessment Results



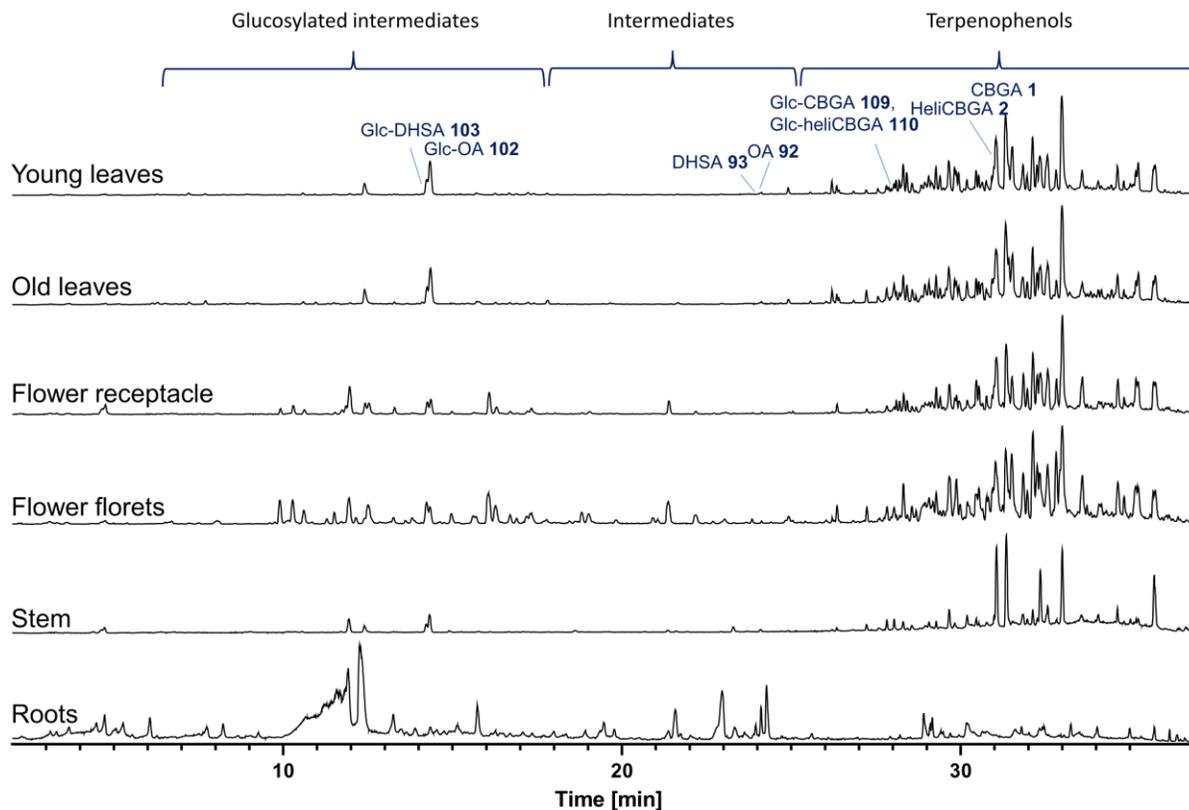
Supplementary Fig. 17. BUSCO completeness values of each of the assemblies. The y-axis shows the different datasets: three transcriptomes (Illumina TrueSeq Trinity assembled transcripts, Iso-Seq polished transcripts, and PASA generated comprehensive dataset) and four genomes (before HiC scaffolding, and after scaffolding, for the primary assembly and the two haplotypes). The x-axis (%BUSCO) indicates the percentage of present genes in each category (Complete and single-copy, Complete and duplicated, Fragmented, and Missing). BUSCO version was 5.2.2 and the database was “embrophyta_odb10”.



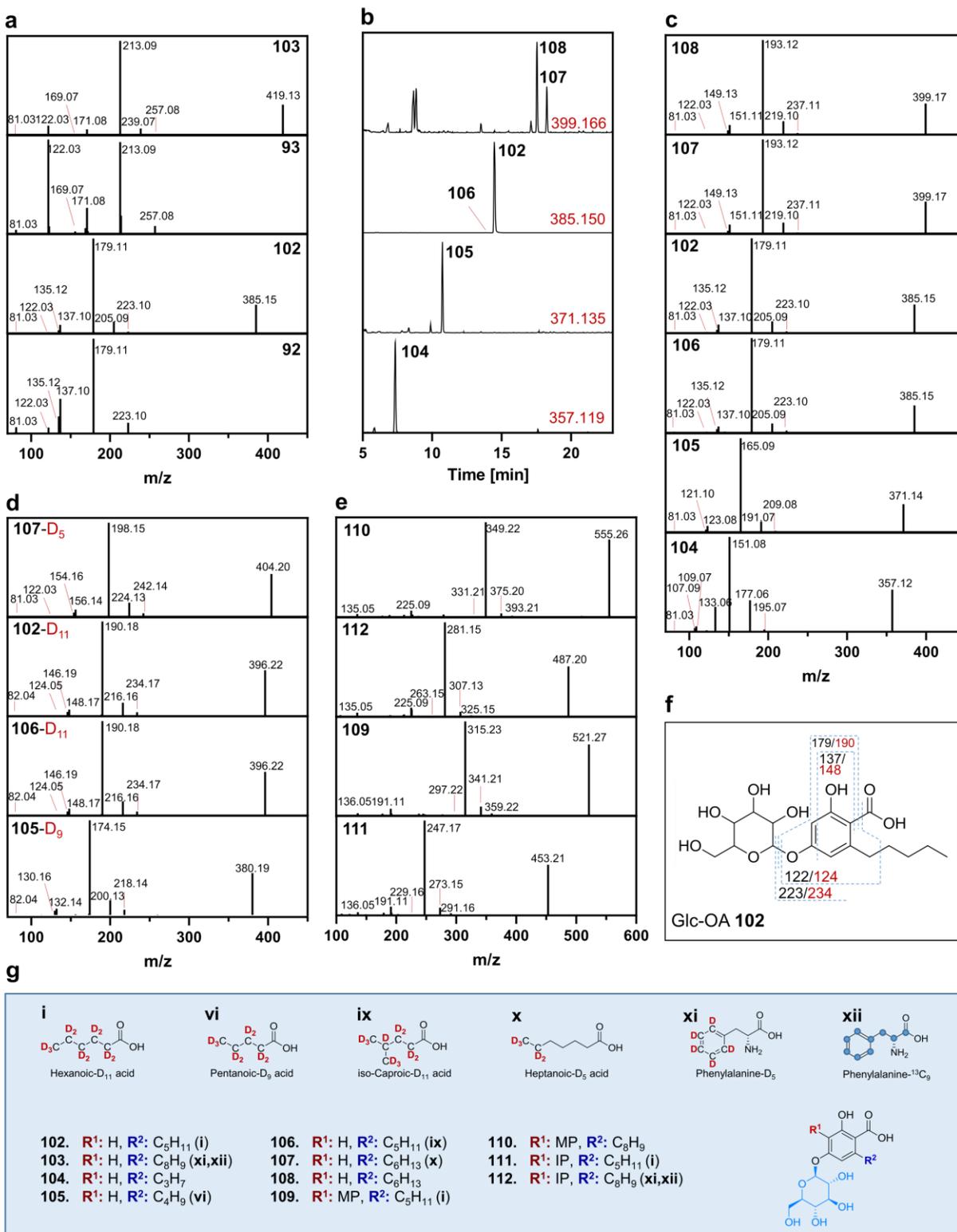
Supplementary Fig. 18. Over Representation Analysis of module M4. Dot plot shows the enriched GO terms of biological processes in the co-expressed genes in module M4. The color scale represents the pvalue while the size of the dot represents the gene count.



Supplementary Fig. 19. Coupled PKS-PKC reactions and PKC phylogeny. **a.** Combinations of PKS-PKC *in vitro* assays tested using hexanoyl-CoA and malonyl-CoA as substrates and the observed products for each combination. Increase of **OA 92** concentration was only observed in coupled assays with CsOAC. **b.** Phylogenetic analysis of candidate PKC enzymes from *H. umbraculigerum* versus CsOAC from *Cannabis*. Bootstrap values are indicated at the nodes of each branch. The *Cannabis* leaf marks the active CsOAC.

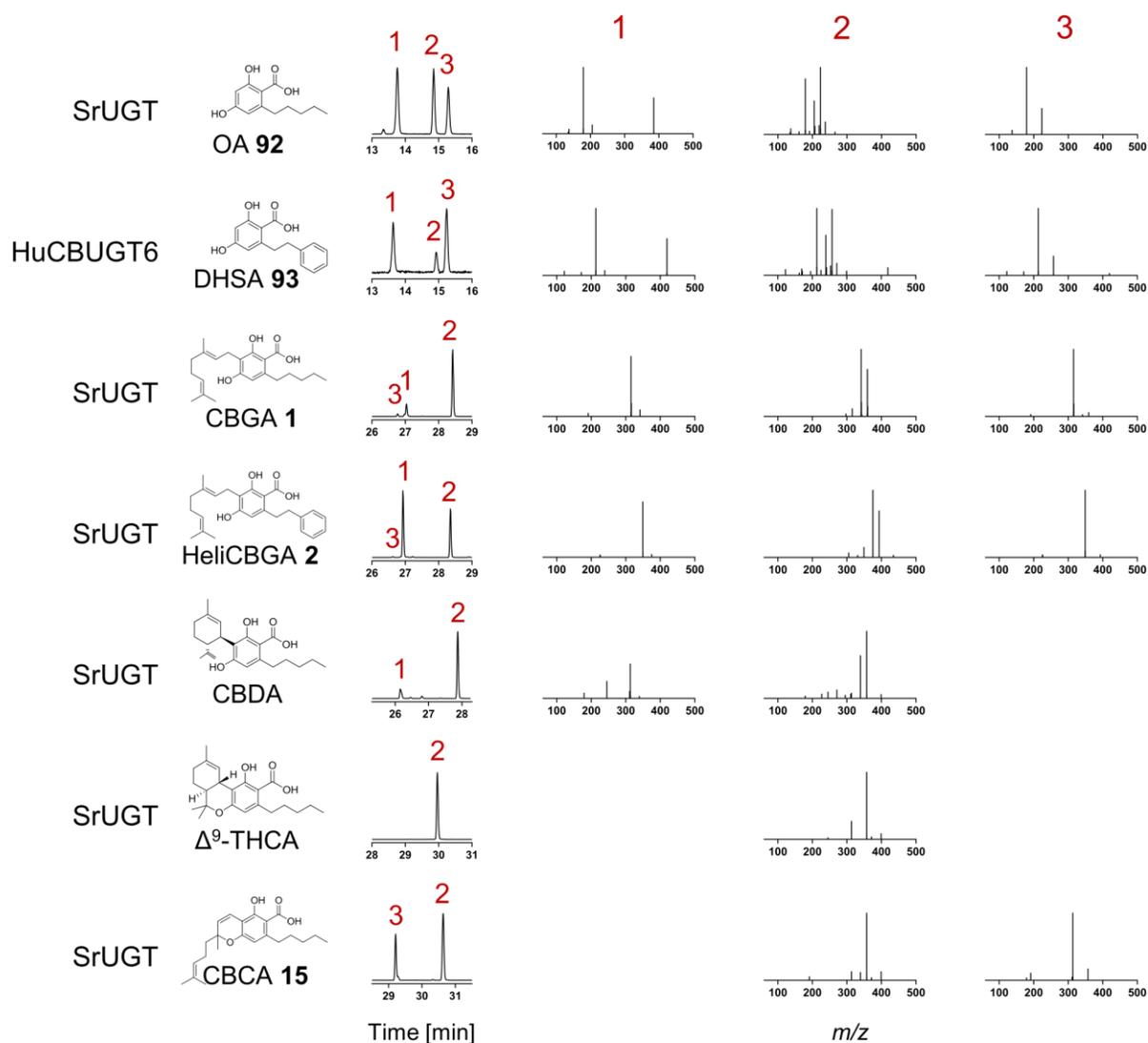


Supplementary Fig. 20. Comparison of total ion chromatograms of six different tissues from *H. umbraculigerum*. Samples were analyzed using UPLC Method 2. Retention time ranges of elution of the glucosylated intermediates, intermediates and terpenophenols and some selected metabolites are annotated. As shown, all the tissues exhibited similar metabolite profiles with different relative abundances, excluding roots that revealed no terpenophenols.

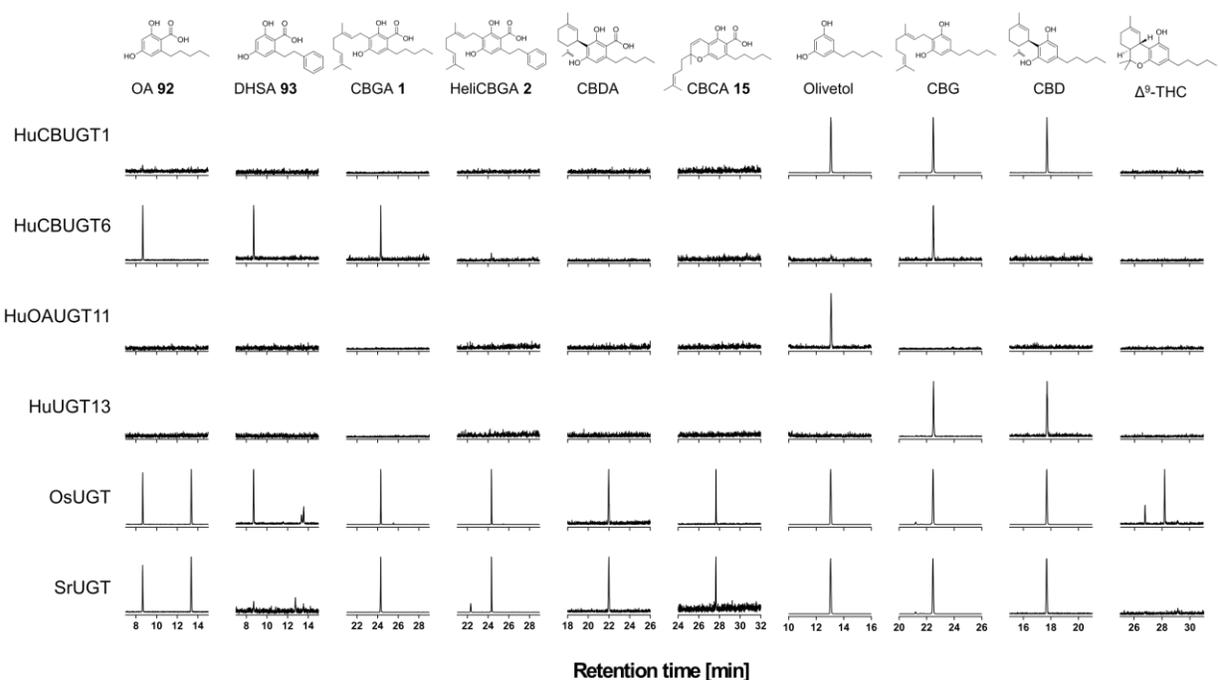


The legend of Supplementary Fig. 21 appears on the next page.

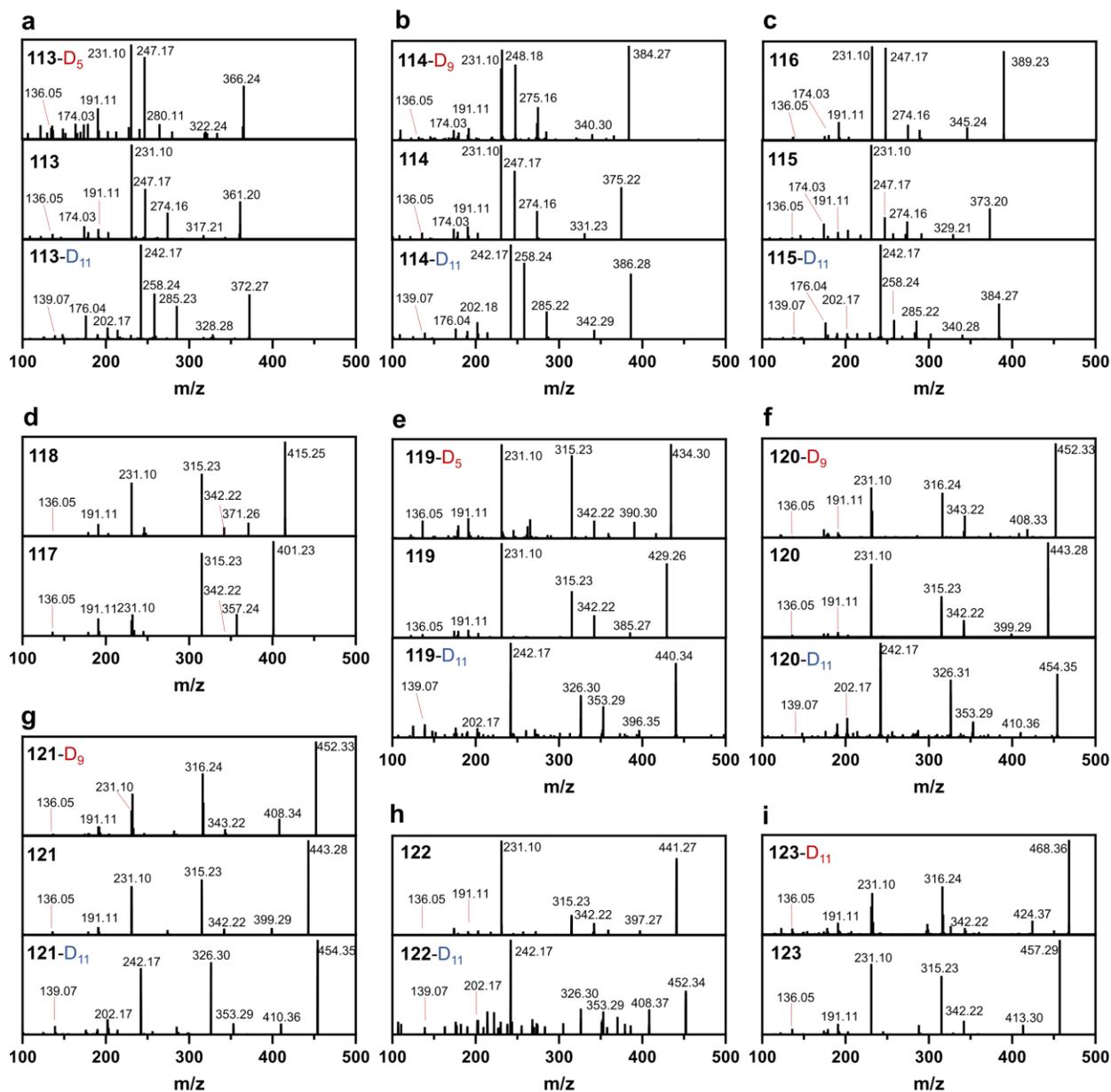
Supplementary Fig. 21. Identification of glucosylated intermediates, cannabinoids and amorfrutins in *H. umbraculigerum*. **a.** Comparison of MS/MS spectra in negative polarity of Glc-OA **102** and Glc-DHSA **103** versus OA **92** and DHSA **93**, respectively. As shown, the glucosylated metabolites exhibited neutral losses of 162.053 Da corresponding to the loss of a hexose and similar fragments as the non-glucosylated metabolites. To confirm the identification, Glc-OA **102** and Glc-DHSA **103** were purified and analyzed by NMR (Supplementary NMR Data 11,12). **b.** Extracted ion chromatograms of $[M-H]^- = 357.119, 371.136, 385.150$ and 399.166 Da, and MS/MS spectra of **c.** unlabeled and **d.** isotopically labeled glucosylated alkyl intermediates. The marked peaks in each chromatogram correspond with the detected glucosylated intermediates. As shown, the alkyl homologues elute from the column in order of chain length because of increasing lipophilicities. For all alkyl homologues, an appropriate m/z shift in the MS/MS spectra of all the product ions that include the alkyl chain was observed. Metabolites **106** and **108** had similar masses and MS/MS fragmentations as Glc-OA **102** and Glc-HA **107** and were assigned as branched short-chain FAs according to feeding experiments, and in agreement with the identified cannabinoids (Supplementary Fig. 3). **e.** MS/MS spectra in negative polarity of Glc-CBPA **111**, Glc-CBGA **109**, Glc-heliCBPA **112** and Glc-heliCBGA **110**. Identification was by similar MS/MS fragmentations as the non-glucosylated metabolites and relative retention time. **f.** Suggested fragmentation structure of Glc-OA **102** according to MS/MS spectra and labeling. Fragments colored in red correspond to the m/z of the specific fragment in the metabolite labeled with hexanoic-D₁₁ acid. The fragments are similar to those observed for the non-glucosylated metabolites. **g.** Summary of the identified glucosylated metabolites in *H. umbraculigerum* leaf extracts and respective isotopically labeled precursors. IP, isoprenyl; MP, monoprenyl.



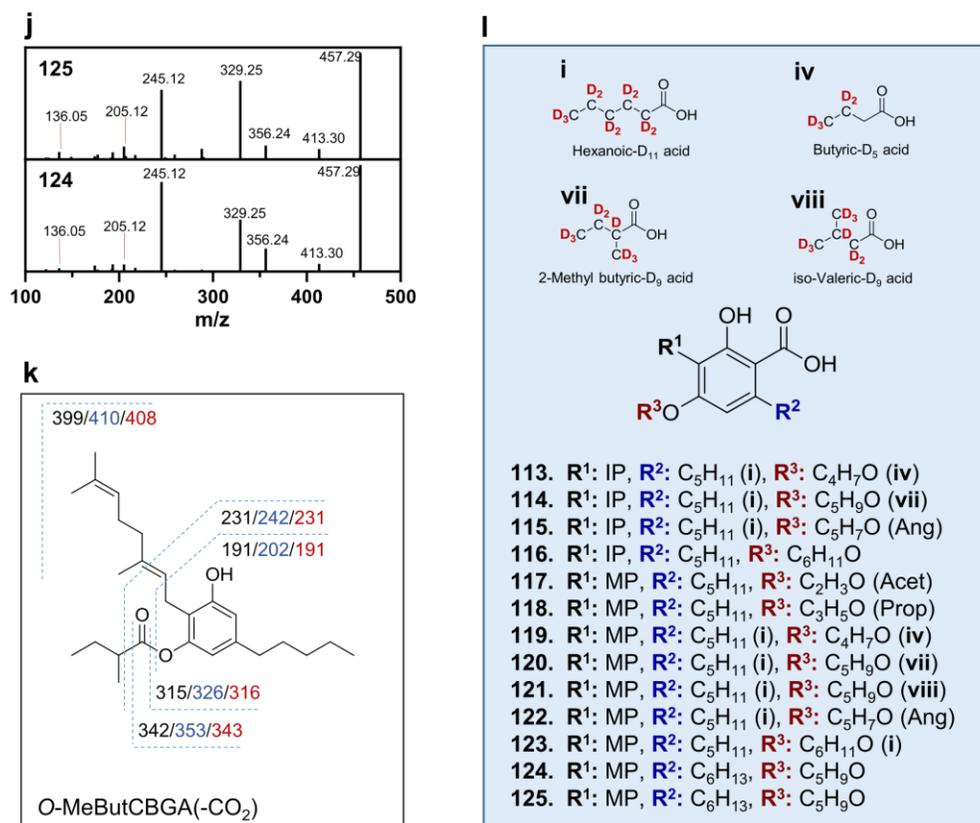
Supplementary Fig. 22. MS/MS spectra of monoglucosides following *in vitro* assays with UGTs from *H. umbraculigerum*, stevia and rice. Assignment of peaks (1-3) was according to MS/MS fragmentation patterns and the m/z difference between the parent and fragment1 (Supplementary Table 20). The relative retention times of peaks 1 and 2 were constant, whereas peak 3 eluted at different relative retention times. The extracted ion chromatograms for each substrate following a reaction with SrUGT or HuCBUGT6 are shown as reference.



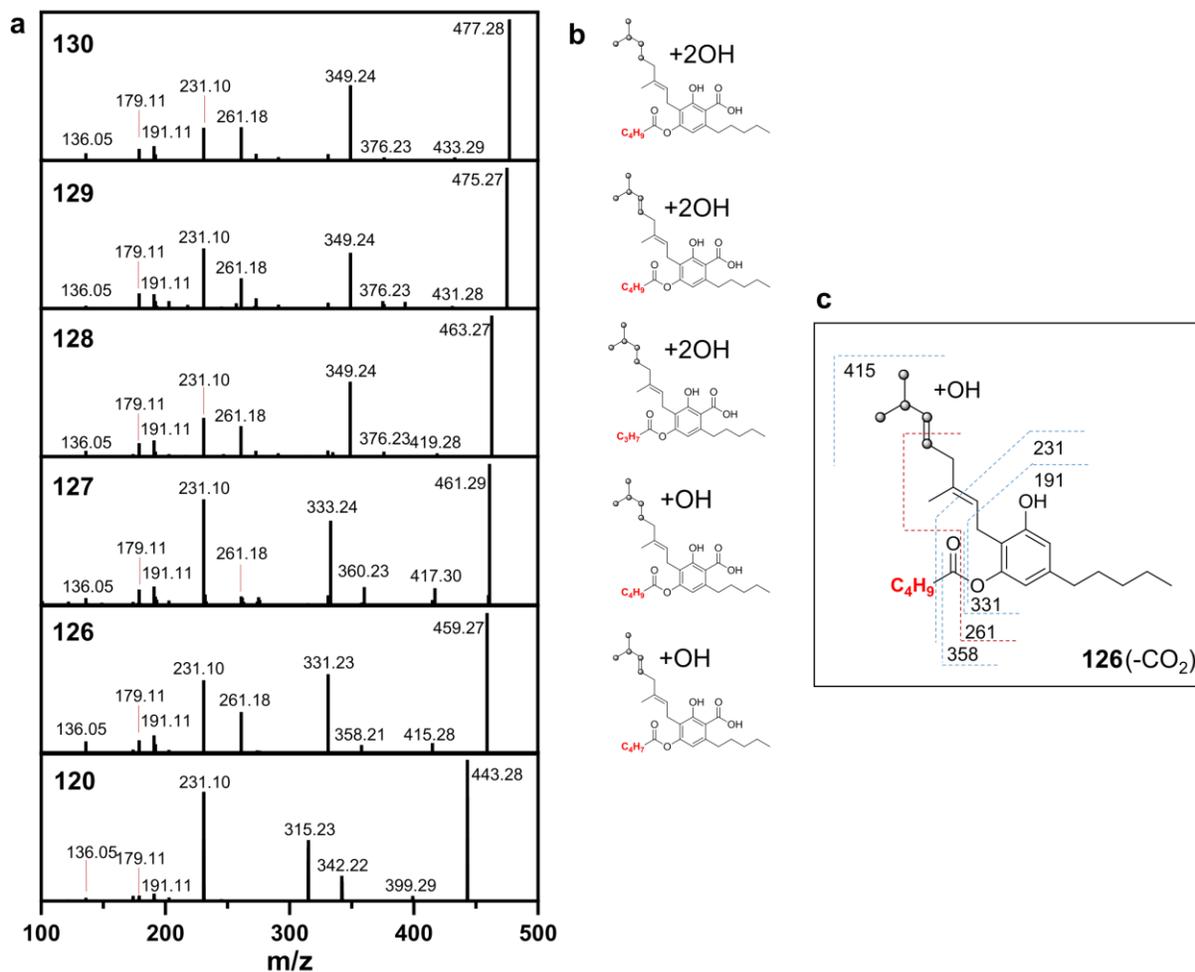
Supplementary Fig. 23. *In vitro* production of di-glucosides with the purified UGTs. Extracted ion chromatograms of the observed di-glucosides following enzymatic assays with the purified enzymes in the presence of UDP-Glc and the cannabinoid acceptors. All LC-MS chromatograms were selected for the theoretical m/z values of the respective metabolites of interest.



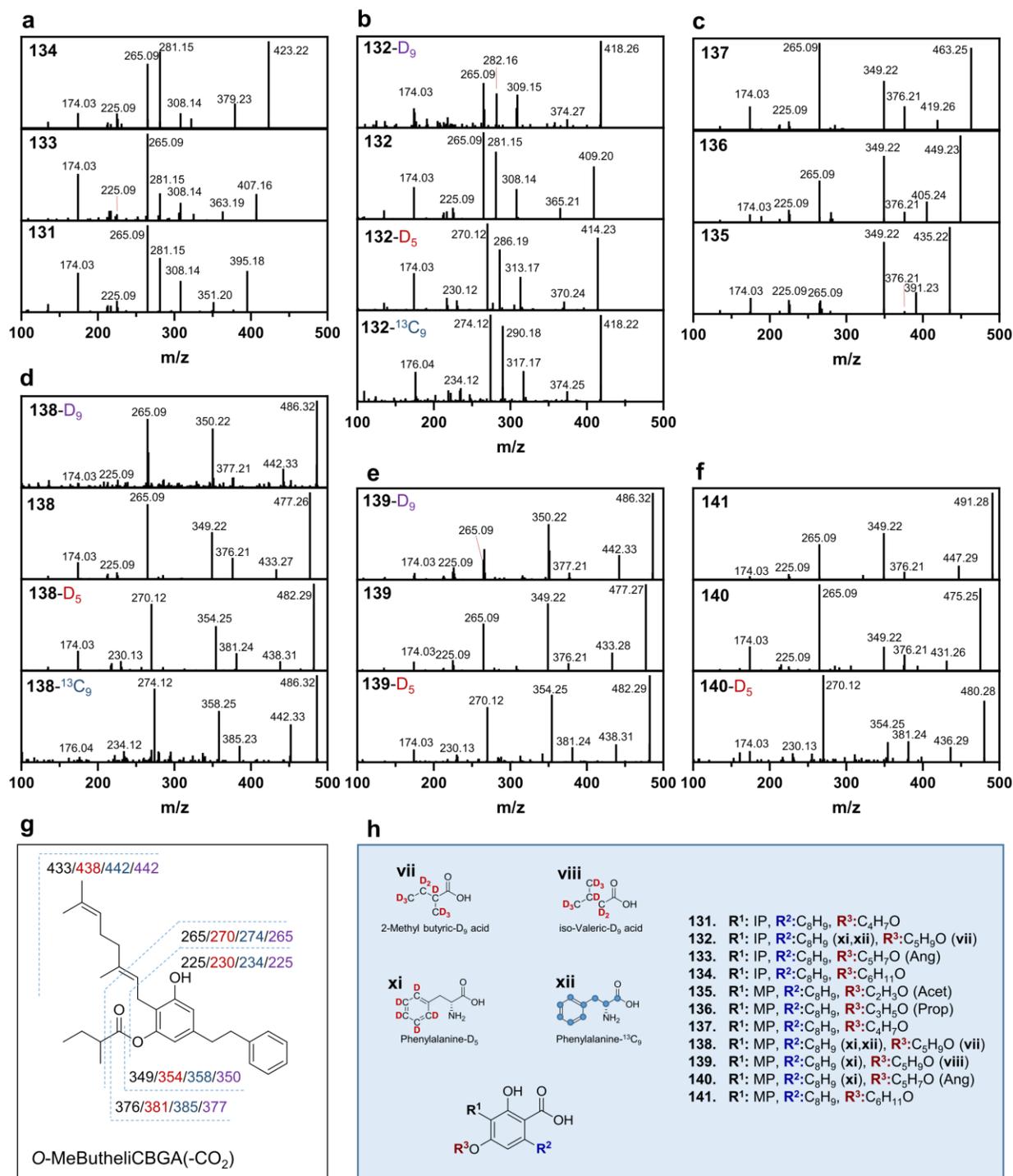
The continuation and legend of Supplementary Fig. 24 appears on the next page.



Supplementary Fig. 24. Identification of *O*-acylated cannabinoids in *H. umbraculigerum*. **a-j.** MS/MS spectra in negative polarity of unlabeled and isotopically labeled *O*-acylated cannabinoids (**113-125**). The metabolites were identified by specific fragmentation patterns as exemplified for *O*-MeButCBGA **120** in **k**, according to MS/MS spectra and labeling [the structure of *O*-MeButCBGA **120** was confirmed by NMR (Supplementary NMR Data 13)]. Fragments colored in red or blue correspond to the *m/z* of the specific fragment with labeled acyl group or alkyl tail, respectively. The isoprenylated cannabinoids **113-116** exhibited analogous fragmentation patterns to the geranylated ones (**119, 120, 122** and **123**, respectively), with mass shifts corresponding with one prenyl group (*m/z* difference of 68.063 Da). In addition, the isoprenylated cannabinoids eluted several minutes before the monoprenylated ones, as a result of increasing lipophilicity with prenylation, and the relative order of elution was in relation to fatty acid as previously described. Cannabinoids **113-123** had five-carbon tails (according to labeling with hexanoic-D₁₁ acid and/or MS/MS fragmentation) and **124-125** had six-carbon tails. Cannabinoids **117** and **118** were not labeled following feeding probably due to low abundance in the plant. However according to accurate mass and MS/MS fragmentation they contain acetyl and propyl groups, respectively. Cannabinoids **115** and **122** potentially contain acyl groups from angelic acid [as in previously identified amorfrutins by Bohlmann and Hoffmann (1979)¹²]. **l.** Summary of the identified *O*-acylated cannabinoids in *H. umbraculigerum* leaf extracts and respective isotopically labeled precursors. IP, isoprenyl; MP, monoprenyl.



Supplementary Fig. 25. Identification of hydroxylated and dihydroxylated *O*-acylated cannabinoids in *H. umbraculigerum*. **a.** MS/MS spectra in negative polarity of hydroxylated (**126**, **127**) and dihydroxylated (**128-130**) cannabinoids. The MS/MS spectrum of *O*-MeButCBGA **120** is shown as reference. The labeled cannabinoids were not observed probably due to low abundance in the extracts, however, according to the observed fragmentation patterns we putatively assigned the structures presented in **b.** with the addition of one or two hydroxyls at the marked possible positions. **c.** Suggested fragmentation structure of **126** according to its observed MS/MS spectrum in relation to the elucidated fragmentation structure of *O*-MeButCBGA **120** (Supplementary Fig. 24).



The legend of Supplementary Fig. 26 appears on the next page.

Supplementary Fig. 26. Identification of *O*-acylated amorfrutins in *H. umbraculigerum*. a-f. MS/MS spectra in negative polarity of unlabeled and isotopically labeled *O*-acylated amorfrutins (**131-141**). The metabolites were identified by specific fragmentation patterns as exemplified for **g. *O*-MeButheliCBGA 138** according to MS/MS spectra and labeling [the structure of *O*-MeButheliCBGA **138** was confirmed by NMR (Supplementary NMR Data 14)], and by following the same fragmentation patterns and relative retention times observed for cannabinoids (Supplementary Fig. 24). Fragments colored in purple, red or blue correspond to the *m/z* of the specific fragment in the metabolite labeled with either deuterated fatty acyl, phenylalanine-D₅ or phenylalanine-¹³C₉, respectively. Amorfrutins **135** and **136** were not labeled following feeding probably due to low abundance in the plant. However according to accurate mass and MS/MS fragmentation they contain acetyl and propyl groups, respectively. Metabolites **133** and **140** potentially contain acyl groups from angelic acid [as in previously identified amorfrutins by Bohlmann and Hoffmann (1979)¹²]. **h.** Summary of the identified *O*-acylated amorfrutins in *H. umbraculigerum* leaf extracts and respective isotopically labeled precursors. IP, isoprenyl; MP, monoprenyl.

Supplementary Table Captions

Supplementary Table 1. Full description of the identified cannabinoids in *H. umbraculigerum*

Supplementary Table 2. Full description of the identified amorfrutins in *H. umbraculigerum*

Supplementary Table 3. Full description of the identified prenylacylphloroglucinoids in *H. umbraculigerum*

Supplementary Table 4. Full description of the identified prenylchalcones in *H. umbraculigerum*

Supplementary Table 5. Full description of the identified prenylflavanones in *H. umbraculigerum*

Supplementary Table 6. Full description of the identified precursors and glucosylated metabolites in *H. umbraculigerum*

Supplementary Table 7. Quantitative metrics of the haplotype resolved and primary assemblies generated using hifiasm assembler and salsa2 scaffold

Supplementary Table 8. Number of reads obtained in each type of RNAseq library

Supplementary Table 9. Number of mapped and assigned True-seq reads using STAR

Supplementary Table 10. Number of mapped, assigned and UMI deduplicated Tran-seq reads using STAR and UMIttools

Supplementary Table 11. Summary Statistics of the different types of Transposable Element Annotation in the primary assembly using EDTA

Supplementary Table 12. Gene annotation of cannabinoid related genes (AAEs, PKCs, PTs, PKSs), UGTs and AATs which are also co-expressed in module M4

Supplementary Table 13. Types of transposons present in the regions where cluster of PKSs were identified in scaffold_1

Supplementary Table 14. Gene sequence and annotation of cannabinoid related genes (AAEs, PKCs, PTs, PKSs) cloned in this study

Supplementary Table 15. Gene ids of the protein sequences used for phylogenetic trees

Supplementary Table 16. Orthogroup OG0014461 to which HuCoAT6 belongs

Supplementary Table 17. Orthogroup OG0000313 to which HuTKS4 belongs

Supplementary Table 18. Orthogroup OG0002538 to which HuCBGAS4 belongs

Supplementary Table 19. Gene annotation of selected UGTs and AATs

Supplementary Table 20. Full description of the metabolites synthesized *in vitro* by HuUGTs

Supplementary Table 21. Full description of the metabolites synthesized *in vitro* by HuCBAT5

Supplementary Table 22. Proteomes used in orthology and synteny analyses

Supplementary Table 23. List of primers used in this study

Supplementary NMR Data- General Note

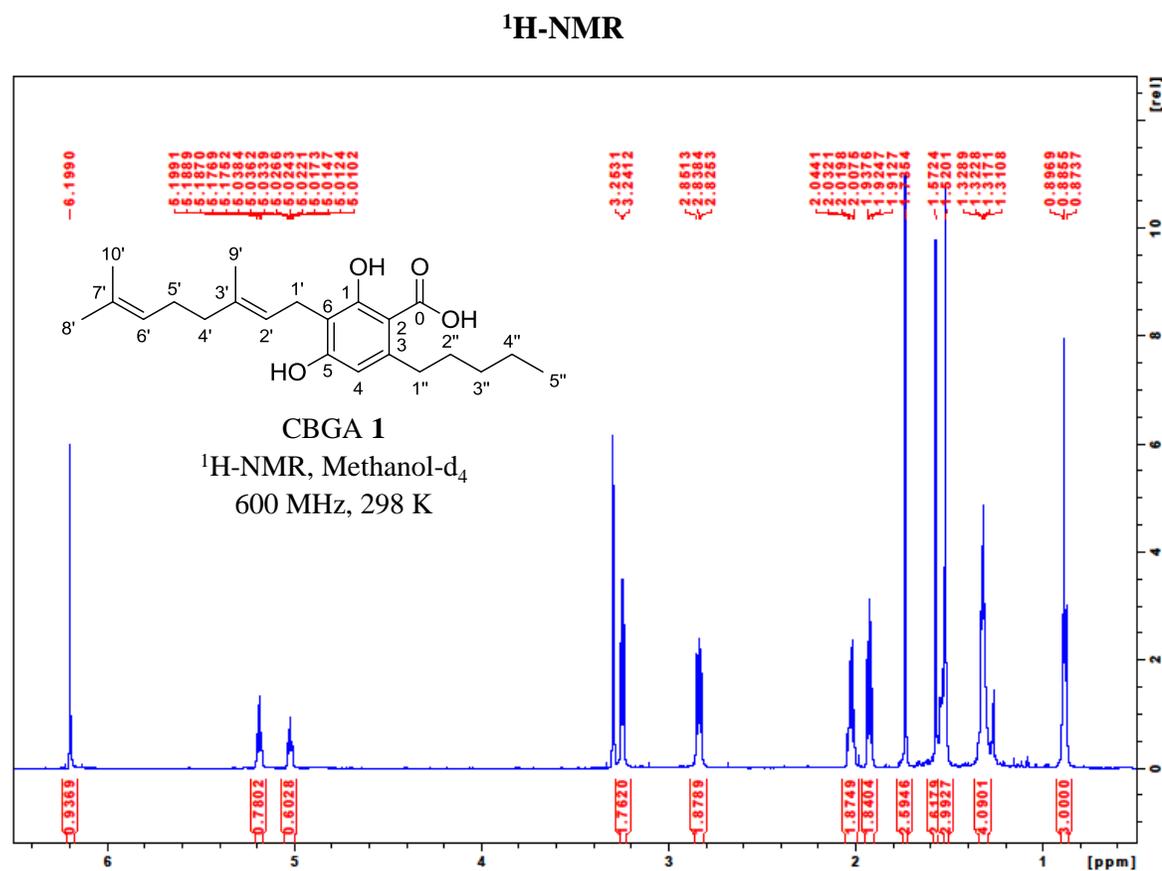
^1H and ^{13}C chemical shift assignment was based on information derived from all the NMR spectra (Supplementary NMR Data 1-14). $^1\text{H} - ^{13}\text{C}$ correlations observed in HMBC spectra are marked by arrows, and $^1\text{H} - ^1\text{H}$ correlations observed in COSY spectra are shown in bold (Supplementary NMR Data 1-14). Data for ^1H NMR spectra are reported as follows: chemical shift (δ , ppm) (multiplicity, coupling constant (Hz), integration). Multiplicity abbreviations are as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. * (asterisk) indicates reduced integral value, due to partial saturation (following solvent presaturation).

Supplementary NMR Data 1. CBGA 1

Data of CBGA 1: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 6.12 (s, 1H), 5.19 (td, $J=7.2, 1.2$ Hz, 1H), 5.03 (tt, $J=7.2, 1.4$ Hz, 1H), 3.25 (d, $J=7.2$ Hz, 2H), 2.84 (m, 2H), 2.03 (m, 2H), 1.92 (t, $J=7.7$ Hz, 2H), 1.74 (s, 3H), 1.57 (s, 3H), 1.53 (m, 2H), 1.52 (s, 3H), 1.30-1.34 (m, 4H), 0.88 (t, $J=6.9$ Hz, 3H).

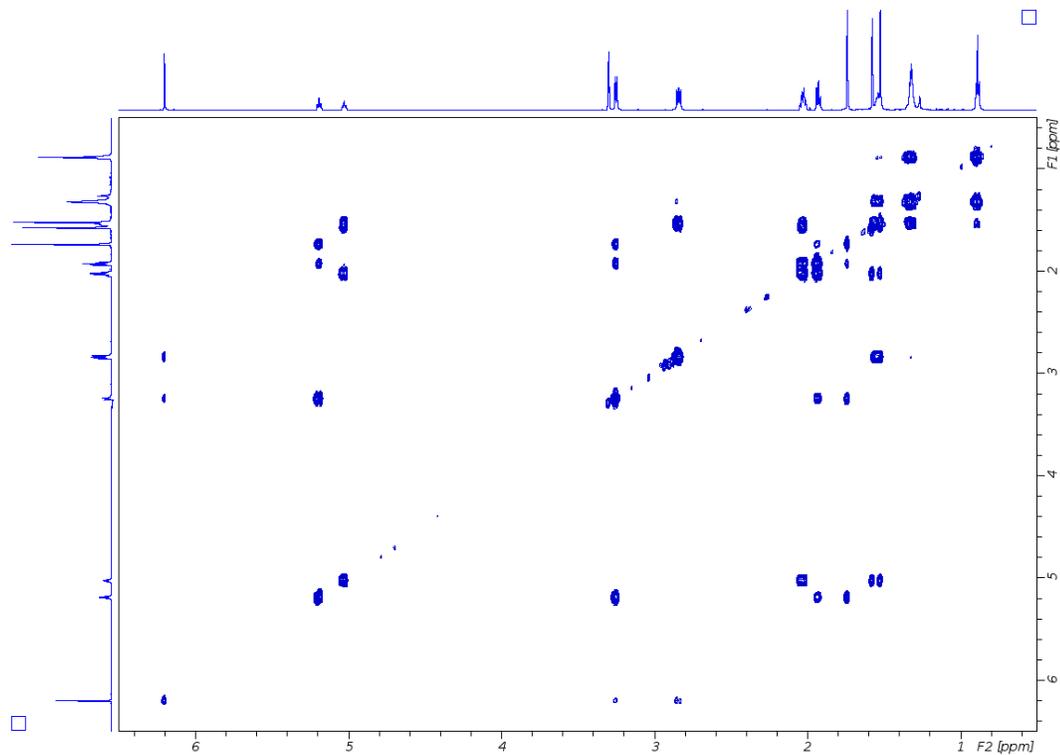
$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 177.6, 166.5, 162.9, 148.8, 137.0, 133.9, 127.7, 126.3, 116.0, 112.9, 106.8, 42.9, 39.6, 35.2, 34.9, 29.8, 27.6, 25.6, 24.9, 19.6, 18.3, 16.2.

HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{22}\text{H}_{31}\text{O}_4$: 359.2222, found 359.2226.

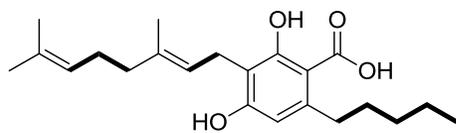
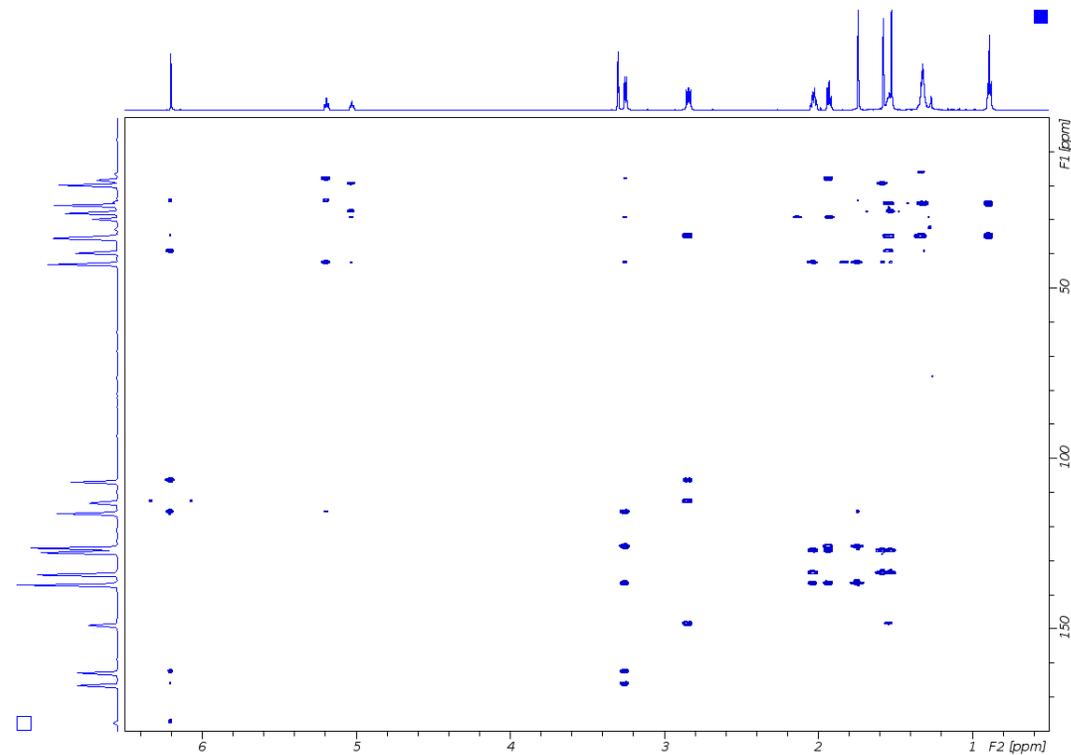


IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
0	177.6	
1	162.9	
2	106.8	
3	148.8	
4	112.9	6.12 (s)
5	166.5	
6	116.0	
1'	24.9	3.25 (d)
2'	126.3	5.19 (td)
3'	137.0	
4'	42.9	1.92 (t)
5'	29.8	2.03 (m)
6'	127.7	5.03 (tt)
7'	133.9	
8'	276	1.57 (s)
9'	18.3	1.74 (s)
10'	19.7	1.52 (s)
1''	39.7	2.84 (m)
2''	34.9	1.53 (m)
3''	35.2	1.30-1.34 (m)
4''	25.6	1.30-1.34 (m)
5''	16.2	0.88 (t)

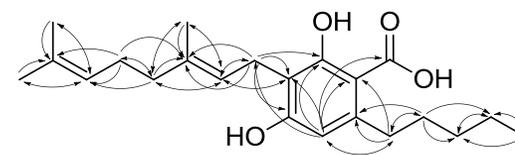
^1H - ^1H COSY



^1H - ^{13}C HMBC

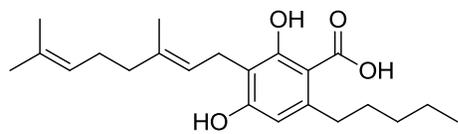
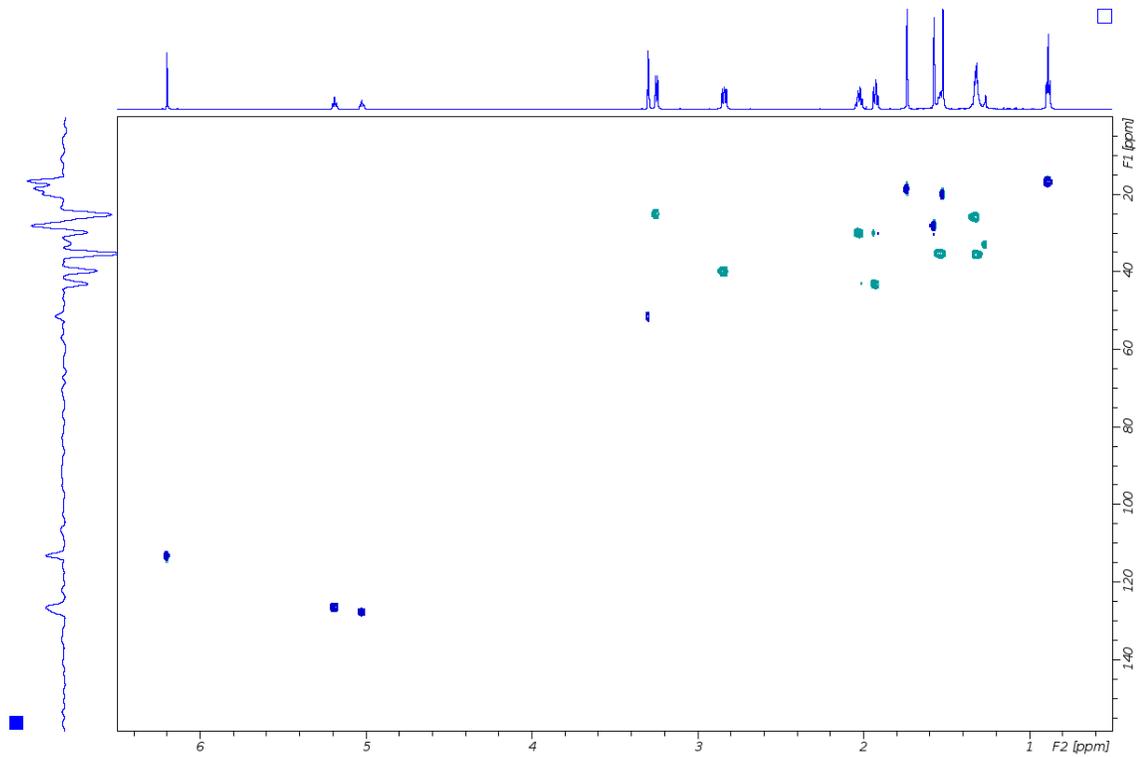


CBGA 1
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K



CBGA 1
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



CBGA 1
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

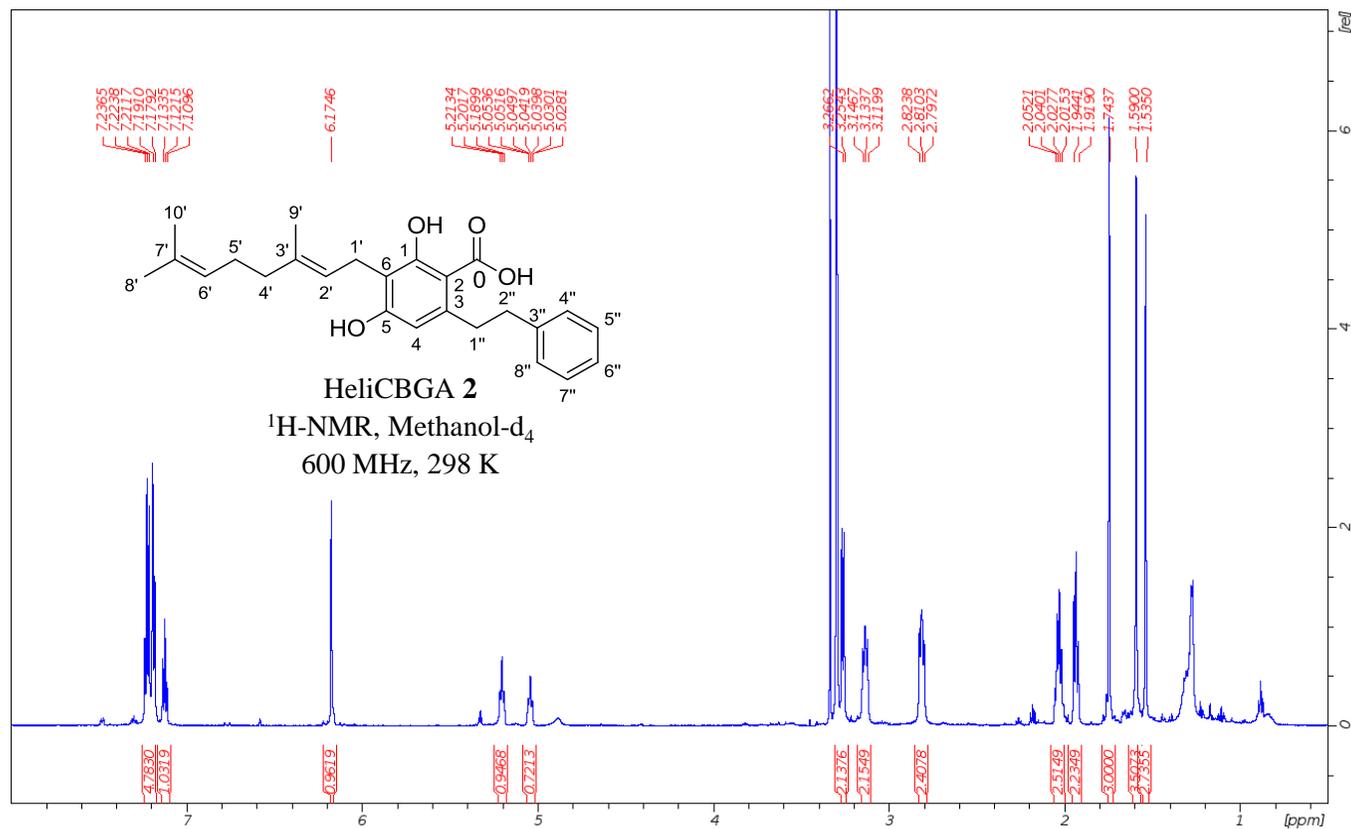
Supplementary NMR Data 2. HeliCBGA 2

Data of heliCBGA 2: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 7.19-7.22 (m, 4H), 7.12 (t, $J = 7.1$ Hz, 1H), 6.17 (s, 1H), 5.20 (t, $J = 7.1$ Hz, 1H), 5.04 (t, $J = 7.1$ Hz, 1H), 3.26 (d, $J = 7.1$ Hz, 2H), 3.11-3.16 (m, 2H), 2.79-2.83 (m, 2H), 1.99-2.06 (m, 2H), 1.93 (t, $J = 7.4$, 2H), 1.74 (s, 3H), 1.59 (s, 3H), 1.54 (s, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 167.1, 162.5, 147.5, 146.3, 136.8, 134.1, 131.5, 131.2, 128.7, 127.5, 126.2, 116.3, 113.1, 107.4, 43.1, 42.7, 41.7, 29.8, 27.9, 24.9, 19.7, 18.3.

HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{25}\text{H}_{29}\text{O}_4$: 393.2065, found 393.2071.

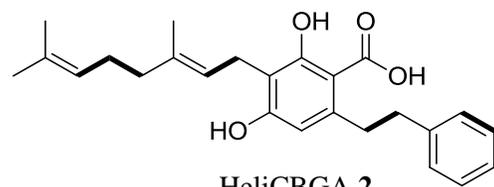
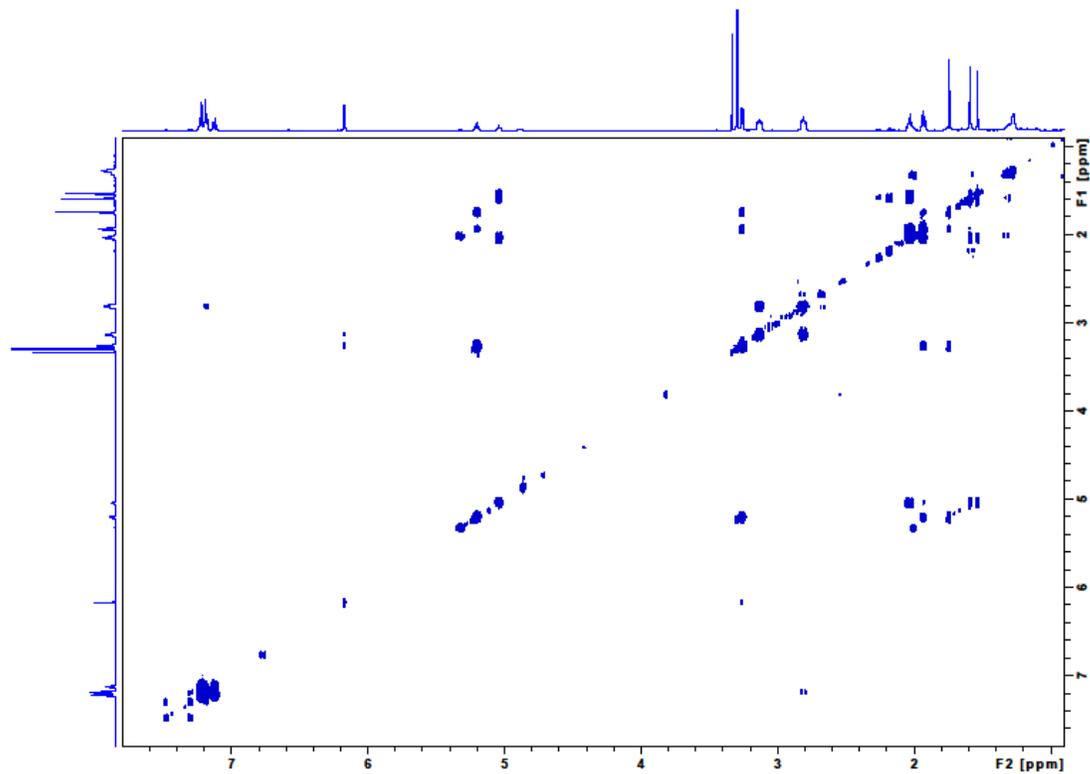
$^1\text{H-NMR}$



IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
0	nd ^a	
1	167.1	
2	107.4	
3	147.5	
4	113.1	6.17 (s)
5	162.5	
6	116.3	
1'	24.9	3.26 (d)
2'	126.3	5.20 (t)
3'	136.8	
4'	43.1	1.93 (t)
5'	29.8	1.99-2.06 (m)
6'	127.5	5.04 (t)
7'	134.1	
8'	27.9	1.59 (s)
9'	18.3	1.74 (s)
10'	19.7	1.54 (s)
1''	42.2	3.11-3.16 (m)
2''	41.7	2.79-2.83 (m)
3''	146.3	
4'', 8''	131.5	7.19-7.22 (m)
5'', 7''	131.2	7.19-7.22 (m)
6''	128.7	7.12 (t)

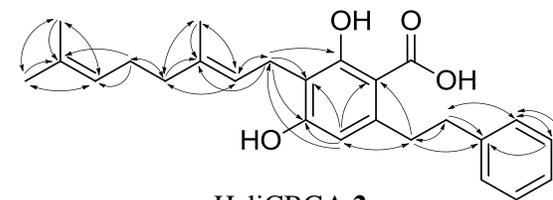
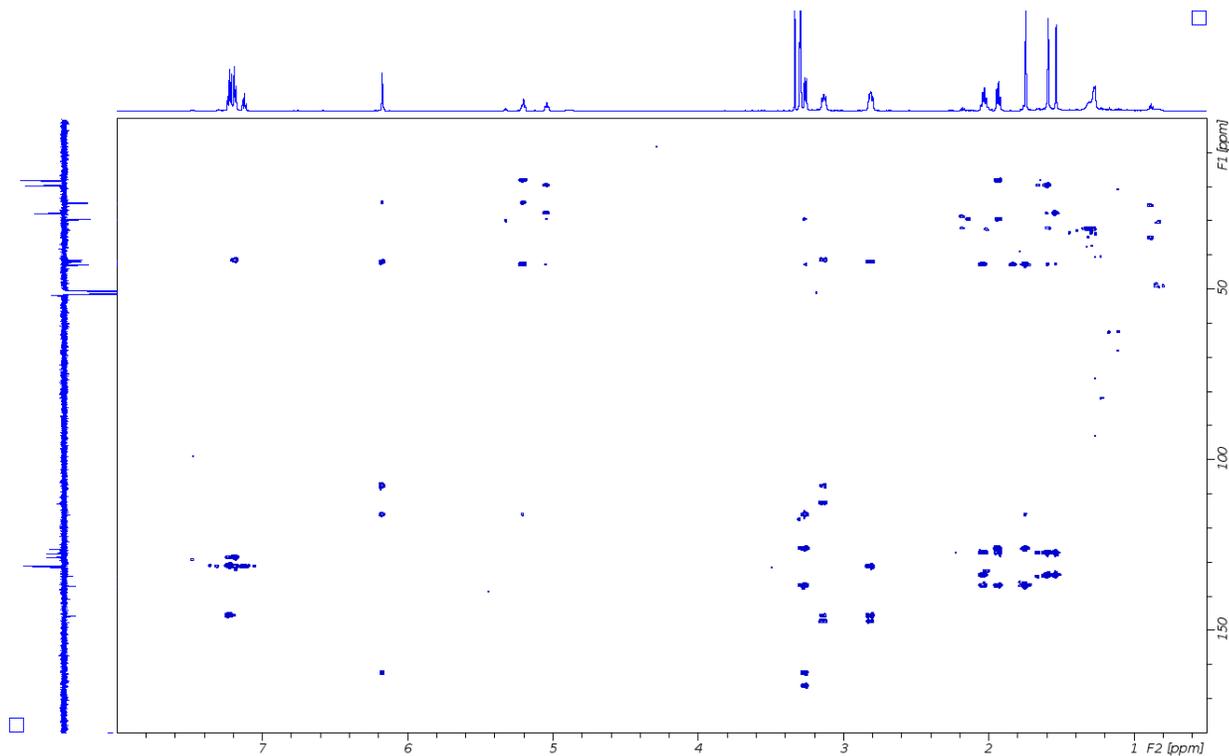
^aThe ^{13}C carbonyl signal was not observed in the NMR spectra, however LC-MS/MS spectra and chemical formula confirm the presence of this group.

^1H - ^1H COSY



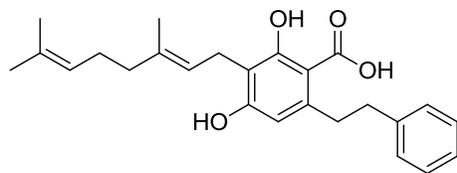
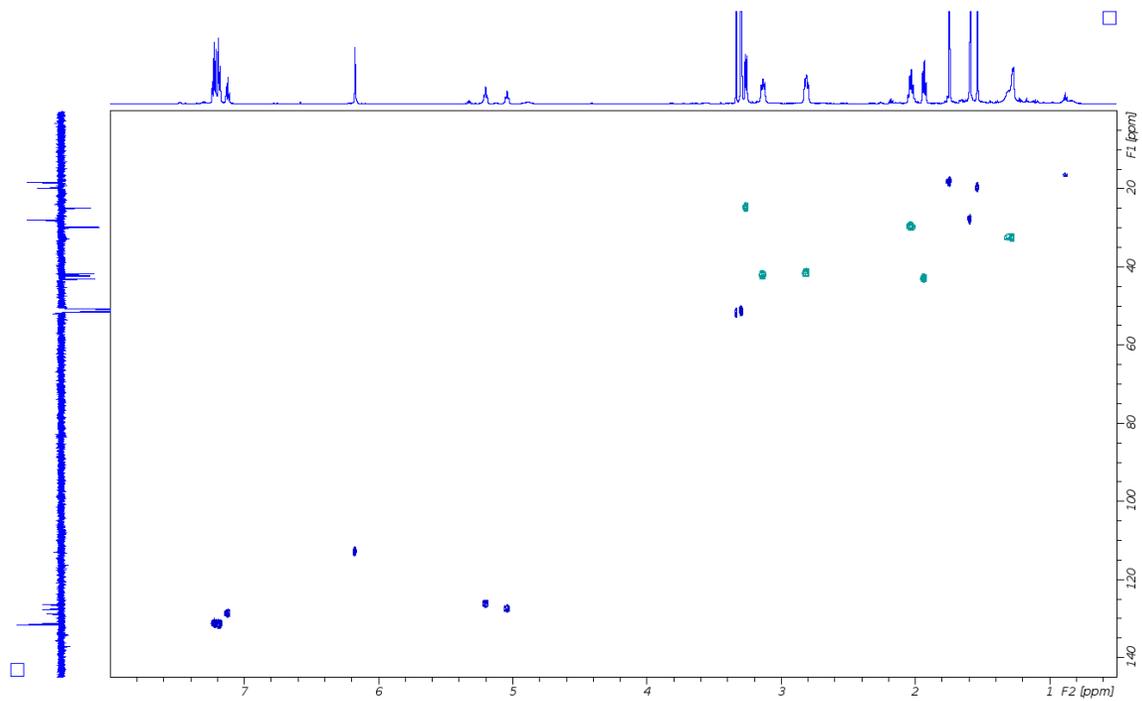
HeliCBGA 2
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



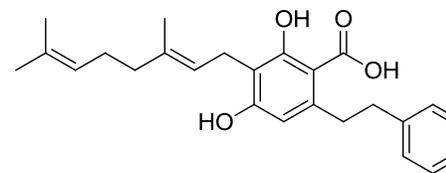
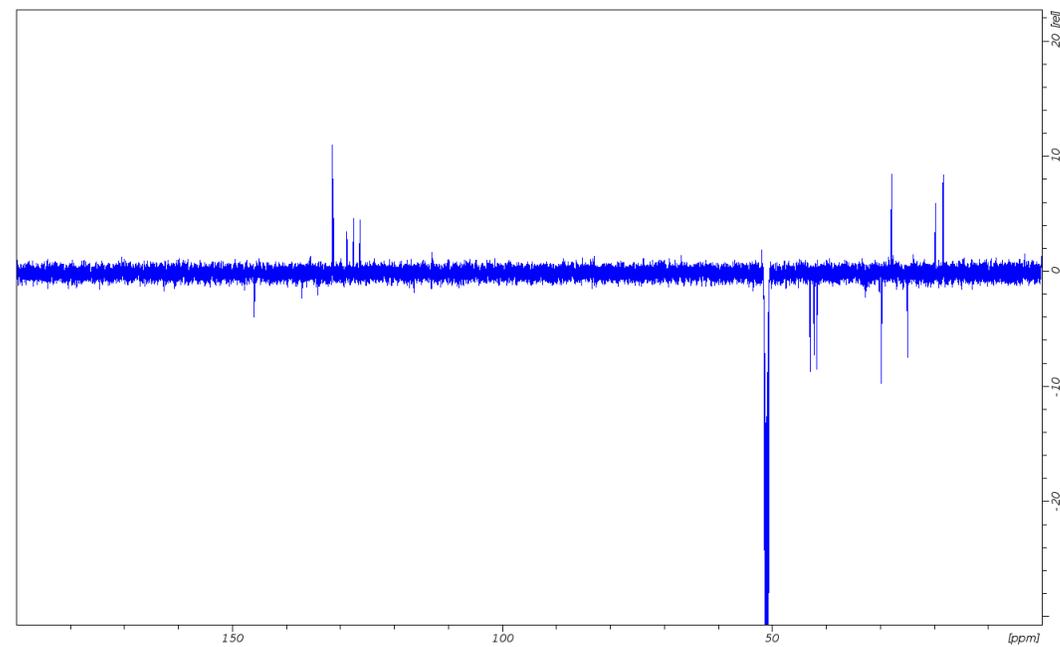
HeliCBGA 2
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



HeliCBGA 2
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

^{13}C NMR



HeliCBGA 2
 ^{13}C NMR, 150.9 MHz,
Methanol- d_4 , 298 K

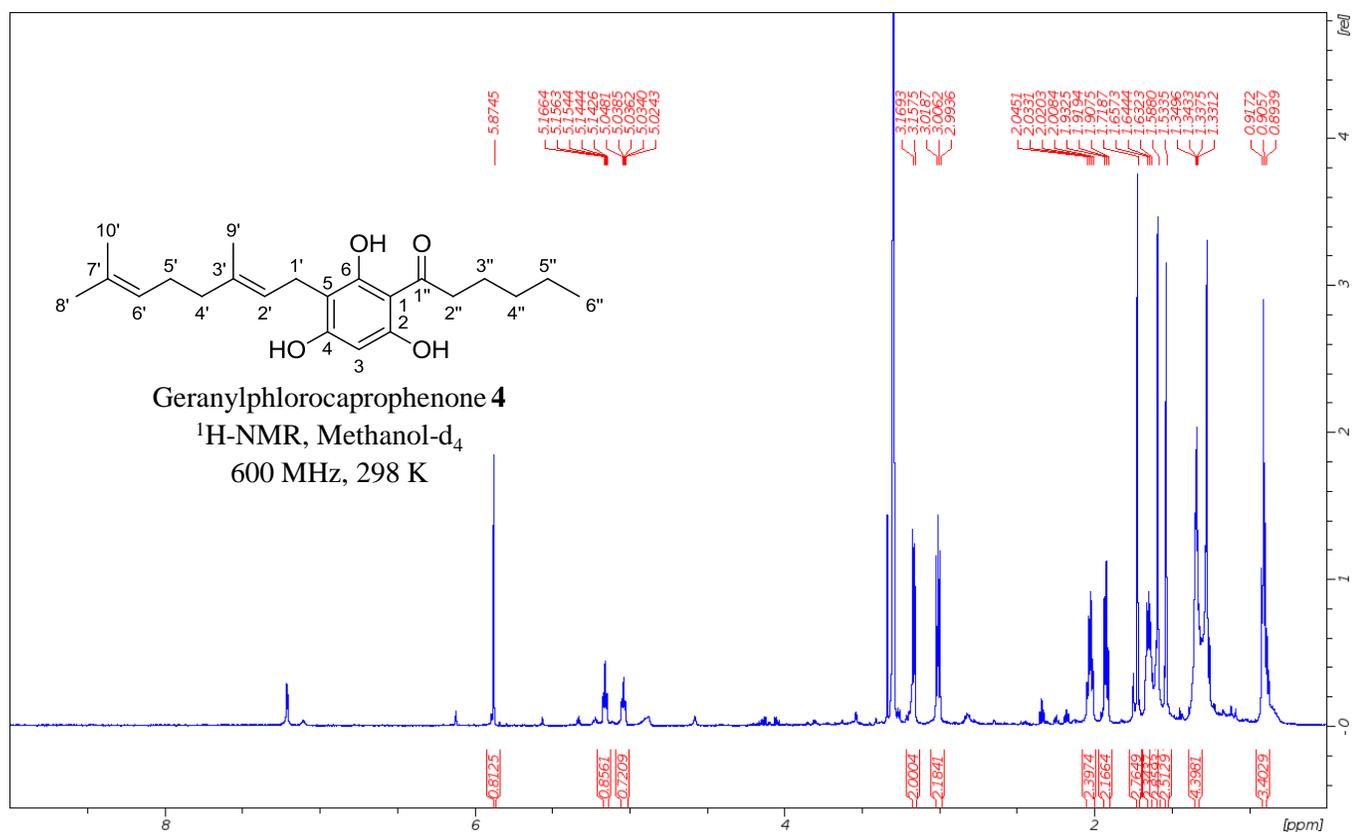
Supplementary NMR Data 3. Geranylphlorocaprophenone 4.

Data of geranylphlorocaprophenone 4: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 5.87 (s, 1H), 5.15 (td, $J=7.07$, 1.16 Hz, 1H), 5.03 (tt, $J = 7.26$, 1.29 Hz, 1H), 3.16 (d, $J = 7.07$ Hz, 2H), 3.01 (m, 2H), 2.03 (m, 2H), 1.92 (m, 2H), 1.72 (s, 3H), 1.63-1.66 (m, 2H), 1.59 (s, 3H), 1.54 (s, 3H), 1.31-1.36 (m, 4H), 0.91 (t, $J = 6.9$ Hz, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 209.5, 167.1, 165.7, 163.3, 136.6, 133.9, 127.5, 126.7, 110.0, 107.0, 96.2, 46.1, 42.9, 34.9, 29.6, 28.1, 27.7 25.6, 24.0, 19.6, 18.2, 16.3.

HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{22}\text{H}_{31}\text{O}_4$: 359.2222, found 359.2226.

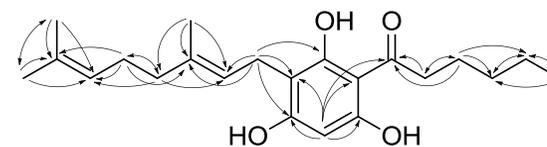
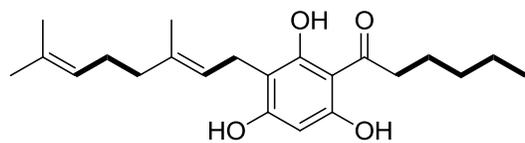
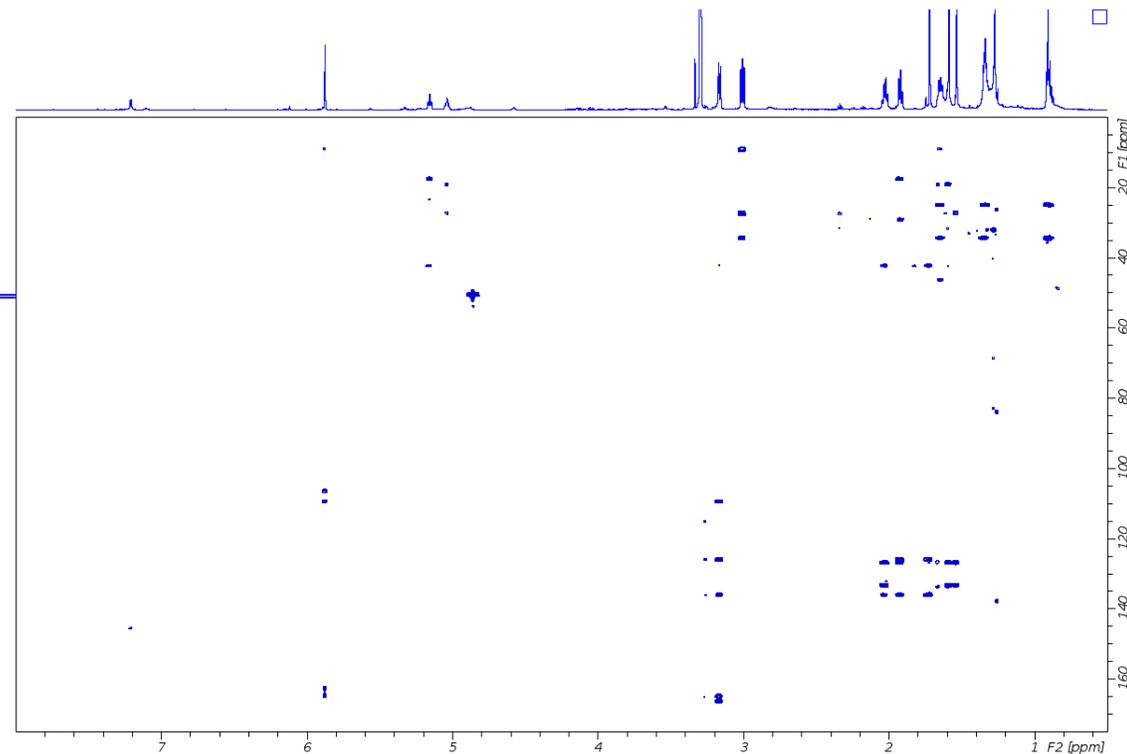
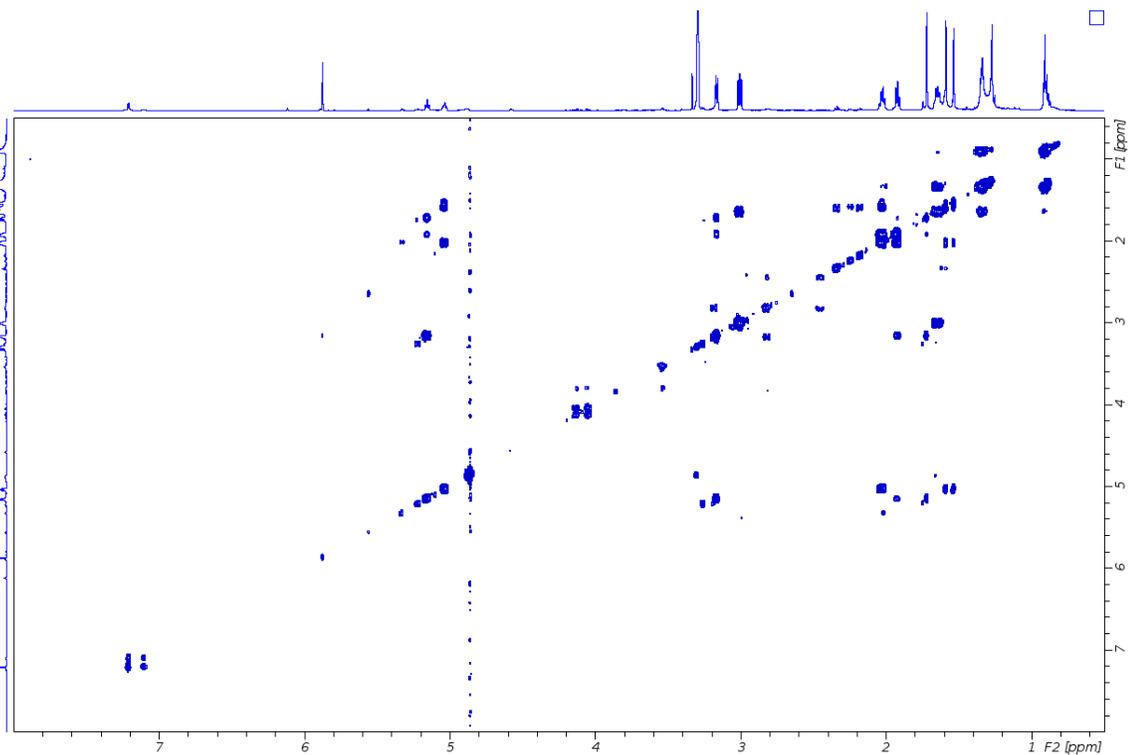
$^1\text{H-NMR}$



IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
1	107.0	
2	163.3	
3	96.8	5.87 (s)
4	165.7	
5	110.0	
6	167.1	
1'	24.0	3.16 (d)
2'	126.7	5.15 (td)
3'	136.6	
4'	42.9	1.92 (m)
5'	29.6	2.03 (m)
6'	127.5	5.03 (tt)
7'	133.9	
8'	27.6	1.59 (s)
9'	18.2	1.72 (s)
10'	19.6	1.54 (s)
1''	209.5	
2''	46.8	3.01 (m)
3''	28.1	1.63-1.66 (m)
4''	25.6	1.31-1.36 (m)
5''	34.9	1.31-1.36 (m)
6''	16.3	0.91 (t)

^1H - ^1H COSY

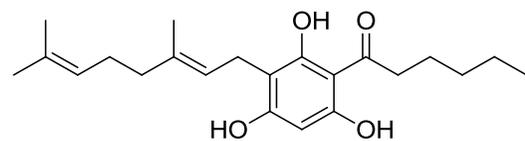
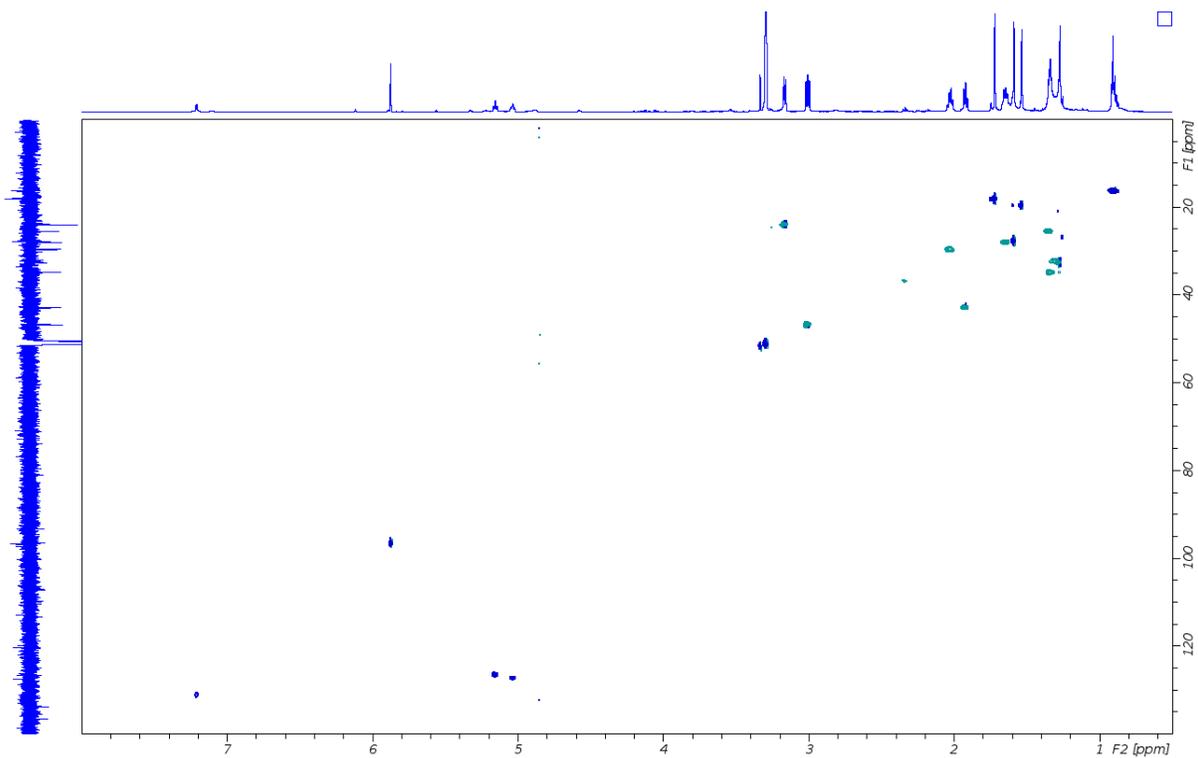
^1H - ^{13}C HMBC



Geranylphlorocaprophenone **4**
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

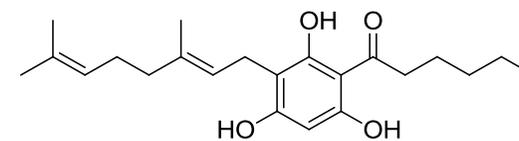
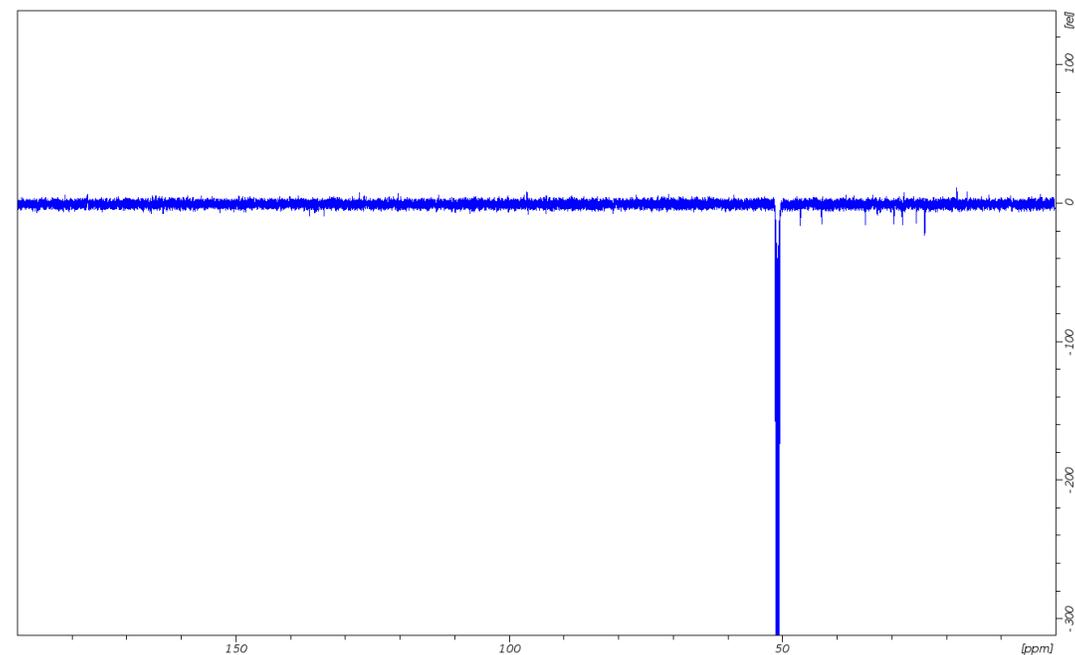
Geranylphlorocaprophenone **4**
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



Geranylphlorocaprophenone 4
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

^{13}C NMR



Geranylphlorocaprophenone 4
 ^{13}C NMR, 150.9 MHz,
Methanol- d_4 , 298 K

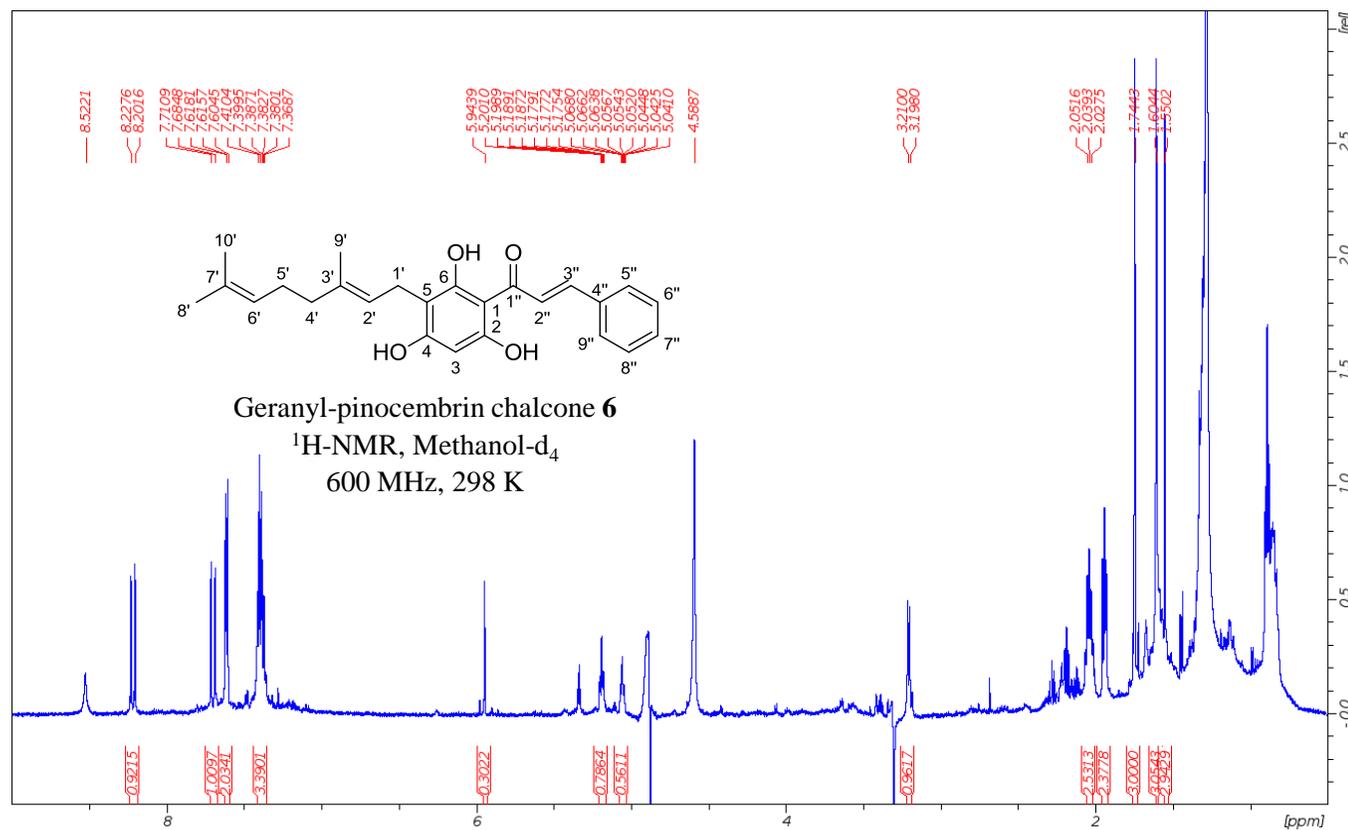
Supplementary NMR Data 4. Geranyl-pinocembrin chalcone 6.

Data of geranyl-pinocembrin chalcone **6**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 8.22 (d, $J = 15.5$ Hz, 1H), 7.70 (d, $J = 15.5$ Hz, 1H), 7.61 (dd, $J = 6.8, 1.75$ Hz, 2H), 7.37-7.41 (m, 3H), 5.94 (s, 1H), 5.19 (td, $J = 6.99, 1.26$ Hz, 1H), 5.05 (tt, $J = 7.18, 1.26$ Hz, 1H), 3.21 (d, $J = 6.99$ Hz, 1H*), 2.04 (m, 2H), 1.94 (m, 2H), 1.74 (s, 3H), 1.61 (s, 3H), 1.55 (s, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 196.2, 167.4, 166.0, 144.4, 138.9, 136.7, 133.8, 133.1, 131.9, 131.3, 131.3, 127.5, 126.6, 110.3, 107.9, 97.2, 43.0, 29.7, 27.8, 24.1, 19.7, 18.1.

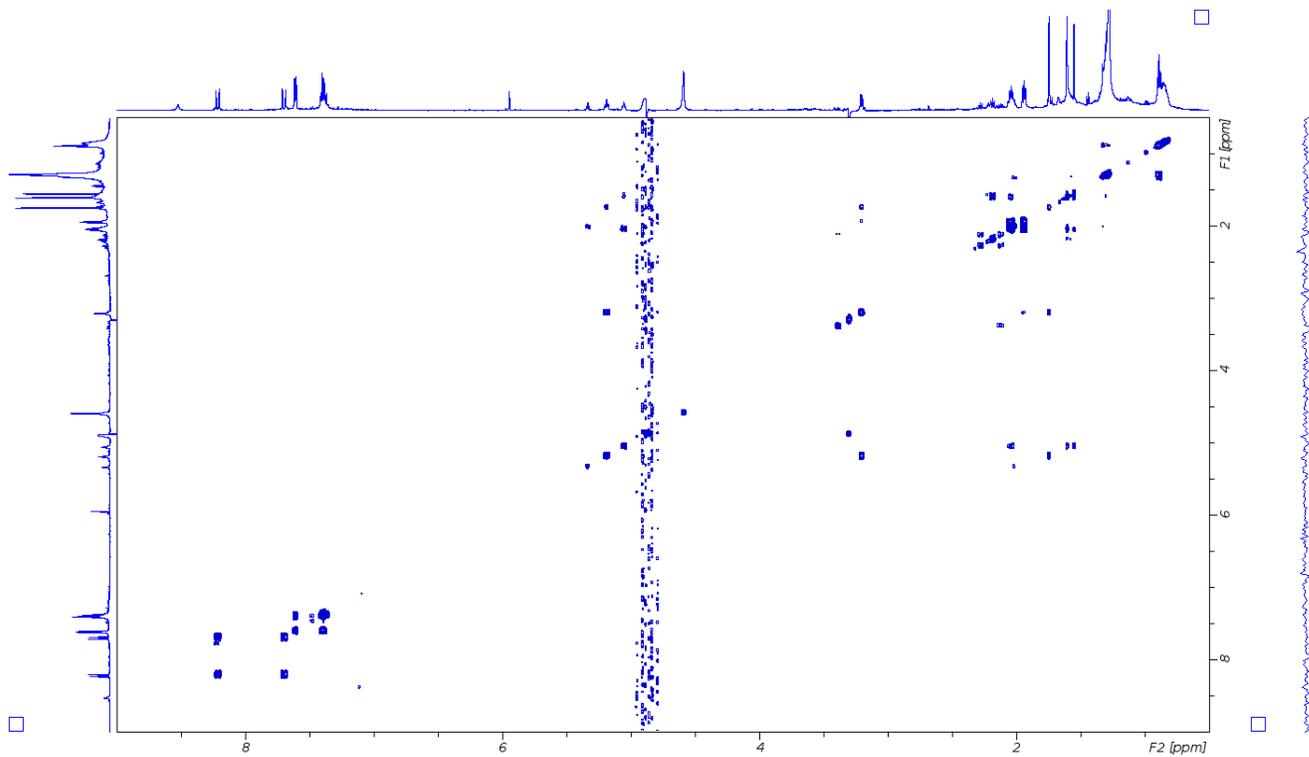
HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{25}\text{H}_{27}\text{O}_4$: 391.1909, found 391.1904.

$^1\text{H-NMR}$



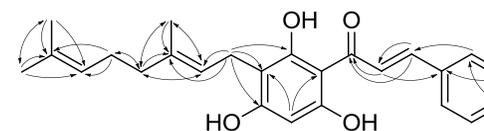
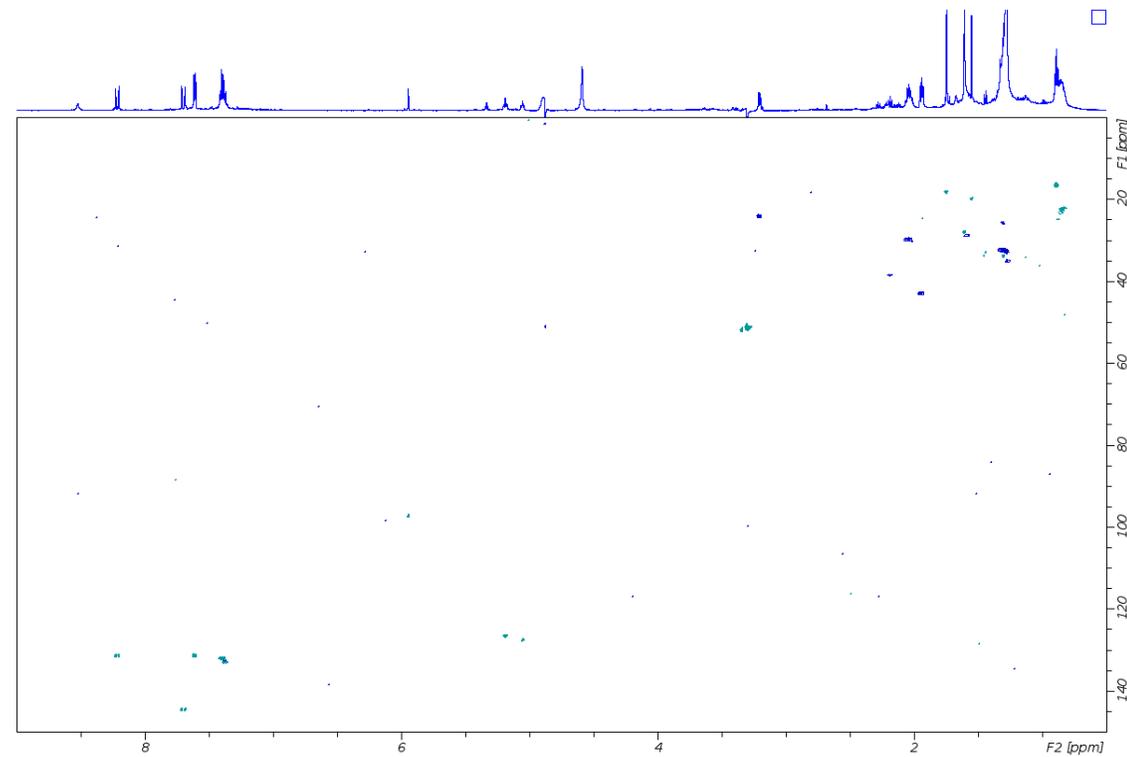
IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
1	107.9	
2	nd	
3	97.2	5.94 (s)
4	166.1	
5	110.3	
6	167.4	
1'	24.1	3.21 (d)
2'	126.6	5.19 (td)
3'	136.7	
4'	43.0	1.94 (m)
5'	29.7	2.04 (m)
6'	127.5	5.05 (tt)
7'	133.8	
8'	27.8	1.61 (s)
9'	18.1	1.74 (s)
10'	19.7	1.55 (s)
1''	196.2	
2''	131.3	8.22 (d)
3''	144.4	7.70 (d)
4''	138.9	
5'', 9''	131.3	7.61 (dd)
6'', 8''	131.9	7.37-7.41 (m)
7''	133.1	7.37-7.41 (m)

^1H - ^1H COSY



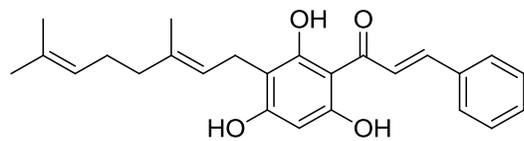
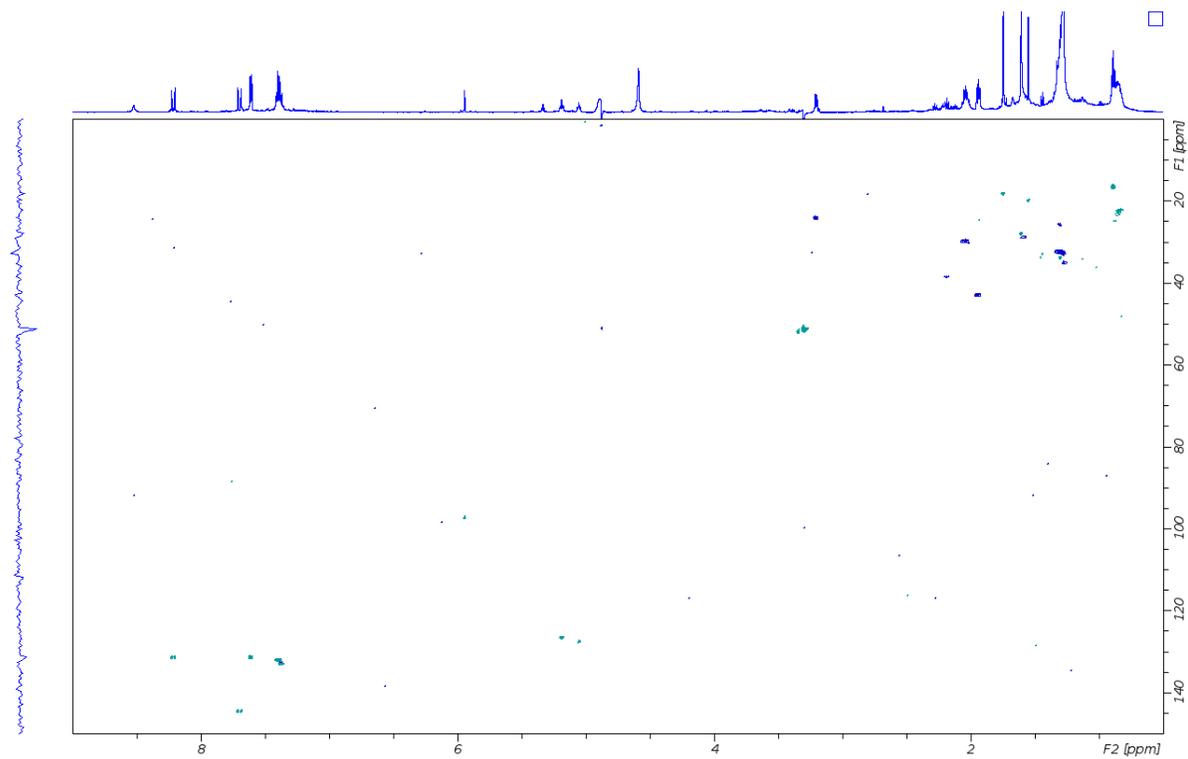
Geranyl-pinocembrin chalcone **6**
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



Geranyl-pinocembrin chalcone **6**
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



Geranyl-pinocembrin chalcone **6**

^1H - ^{13}C HSQC,

Methanol- d_4 , 298 K

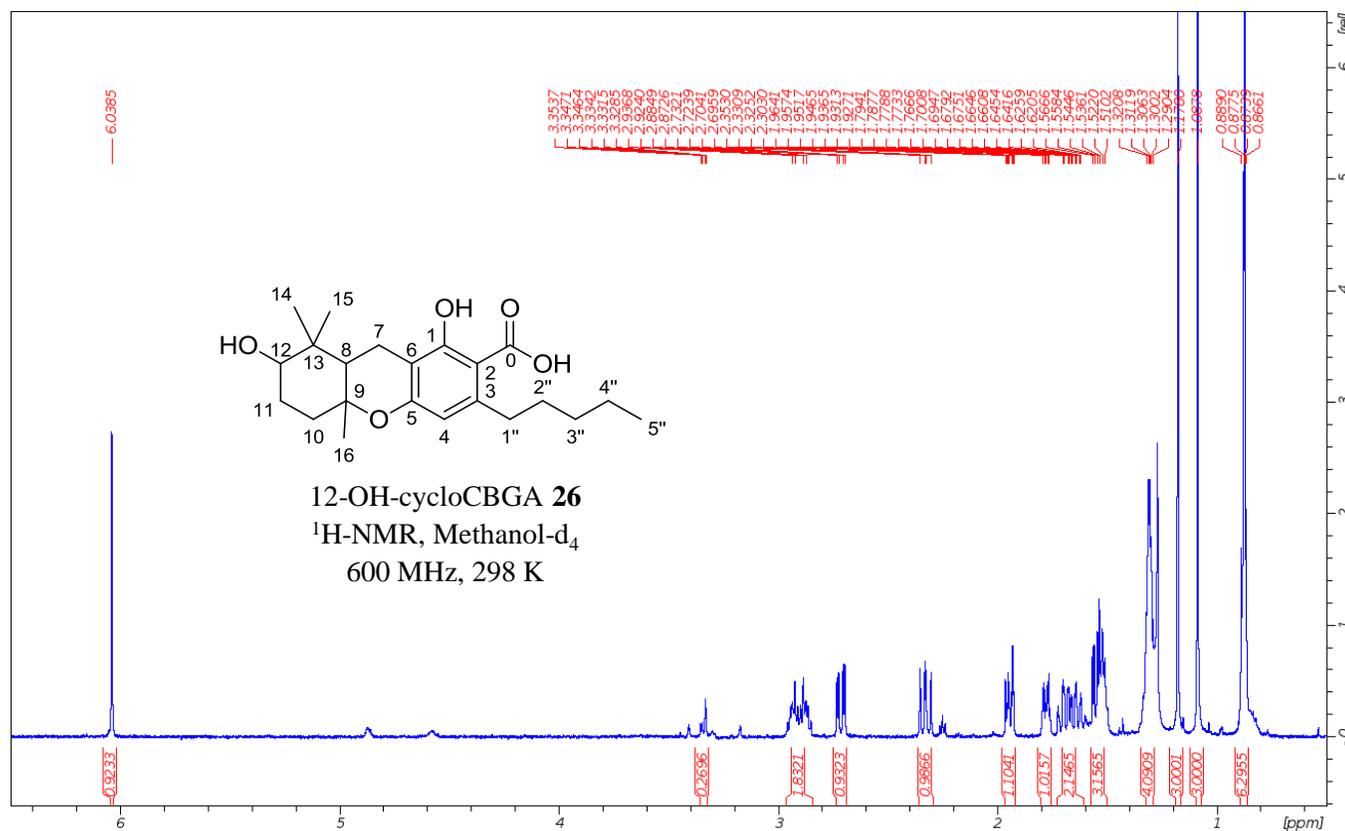
Supplementary NMR Data 5. 12-OH-CycloCBGA 26.

Data of 12-OH-cycloCBGA **26**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 6.04 (s, 1H), 3.34 (m, 1H*), 2.90 (m, 2H), 2.72 (dd, $J = 16.6, 5.0$ Hz, 1H), 2.33 (dd, $J = 16.6, 13.3$ Hz, 1H), 1.94 (m, 1H), 1.78 (m, 1H), 1.73 (td, $J = 13.7, 3.3$ Hz, 1H), 1.64 (m, 1H), 1.56 (dd, $J = 13.3, 5.0$ Hz, 1H), 1.52 (m, 2H), 1.29-1.33 (m, 4H), 1.18 (s, 3H), 1.09 (s, 3H), 0.88 (t, $J = 6.9$ Hz, 3H), 0.87 (s, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 165.6, 159.5, 148.7, 113.4, 110.4, 109.5, 80.9, 79.9, 49.9, 41.5, 40.9, 39.0, 35.3, 34.9, 30.9, 30.9, 29.9, 25.6, 22.1, 20.7, 16.8, 16.5.

HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{22}\text{H}_{31}\text{O}_5$: 375.2171, found 375.2167.

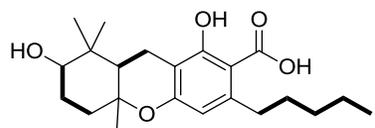
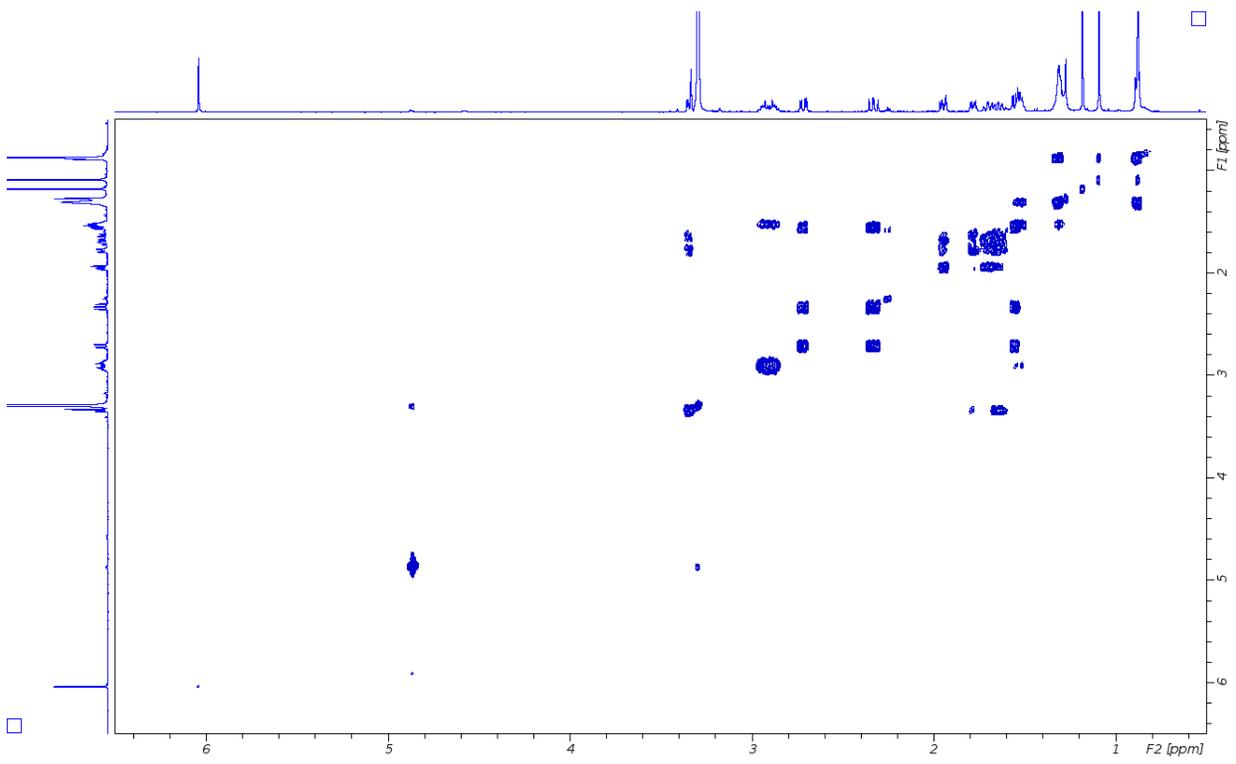
$^1\text{H-NMR}$



IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
0	nd ^a	
1	165.6	
2	109.5	
3	148.7	
4	113.4	6.04 (s)
5	159.5	
6	110.4	
7	20.7	2.33 (dd)
7	20.7	2.72 (dd)
8	49.9	1.56 (dd)
9	79.9	
10	40.9	1.73 (td)
10	40.9	1.94 (m)
11	30.9	1.64 (m)
11	30.9	1.78 (m)
12	80.9	3.34 (m)
13	41.5	
14	29.9	1.09 (s)
15	16.8	0.87 (s)
16	22.1	1.18 (s)
1''	39.0	2.90 (m)
2''	34.9	1.52 (m)
3''	35.3	1.29-1.33 (m)
4''	25.6	1.29-1.33 (m)
5''	16.5	0.88 (t)

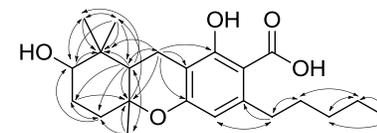
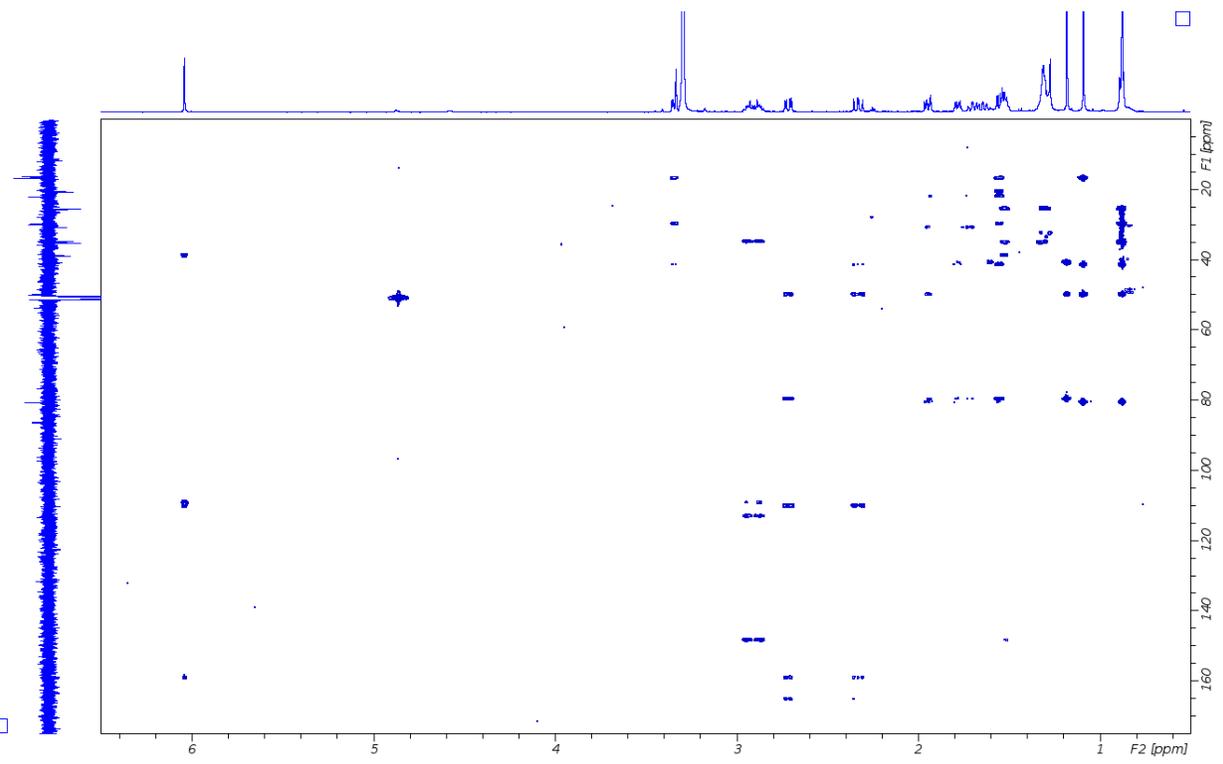
^aThe ^{13}C carbonyl signal was not observed in the NMR spectra, however LC-MS/MS spectra and chemical formula confirm the presence of this group.

^1H - ^1H COSY



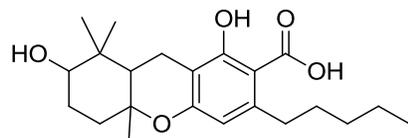
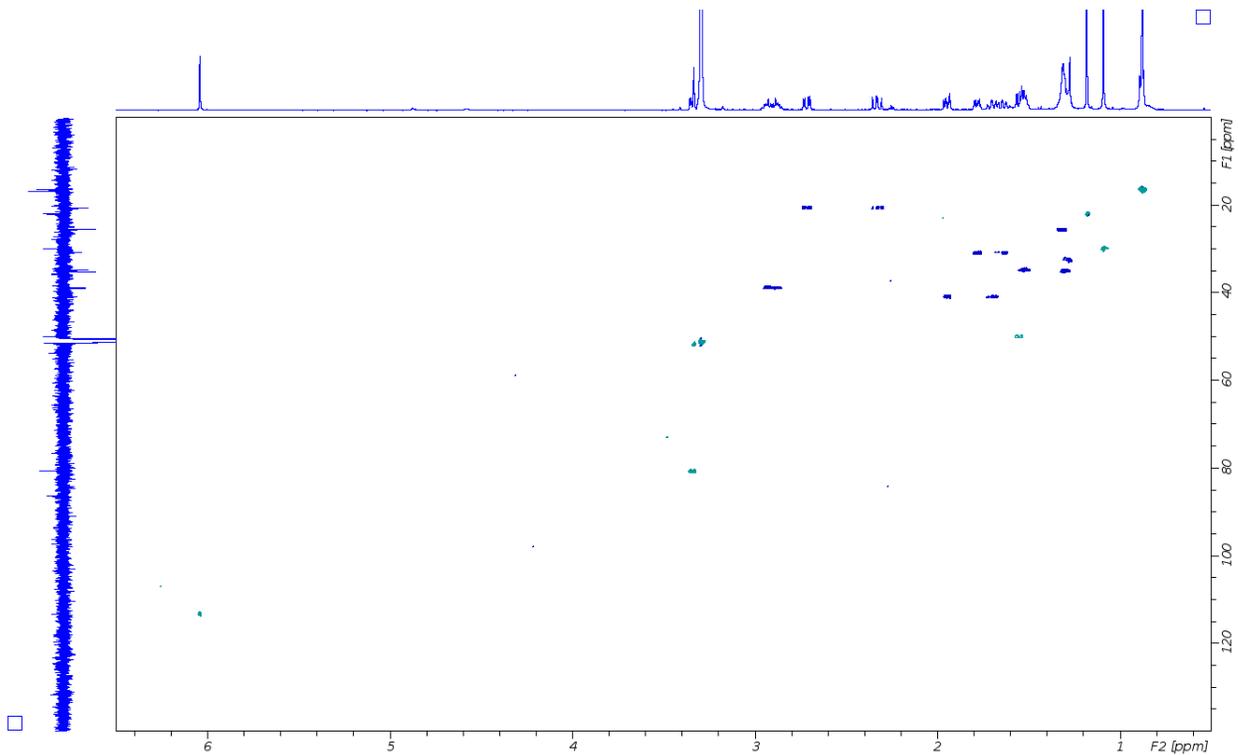
12-OH-cycloCBGA **26**
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



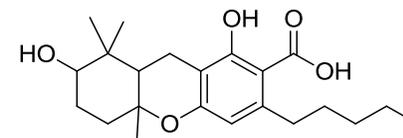
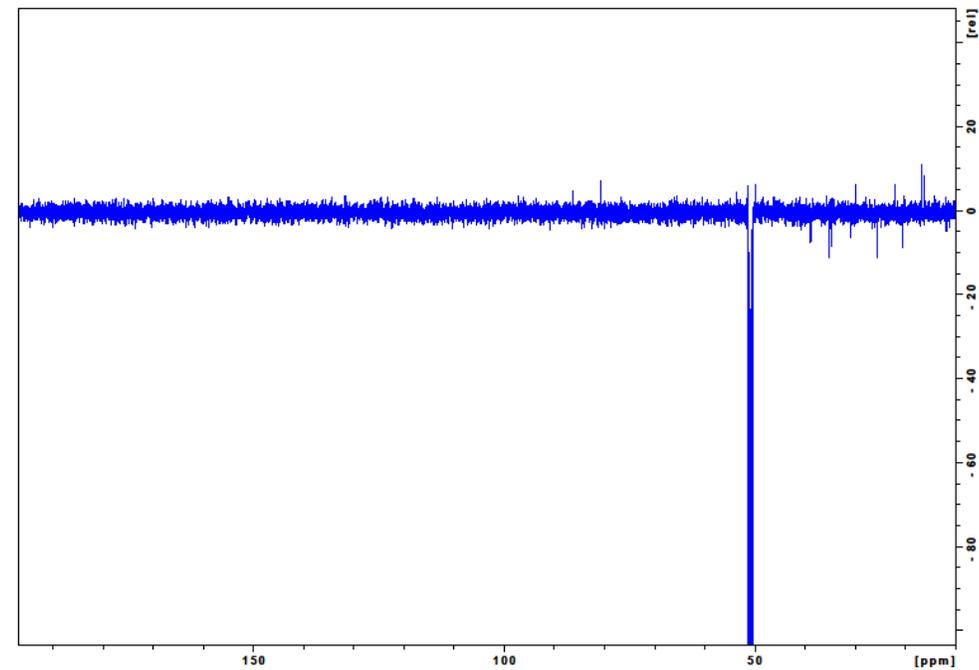
12-OH-cycloCBGA **26**
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



12-OH-cycloCBGA **26**
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

^{13}C NMR



12-OH-cycloCBGA **26**
 ^{13}C NMR, 150.9 MHz,
Methanol- d_4 , 298 K

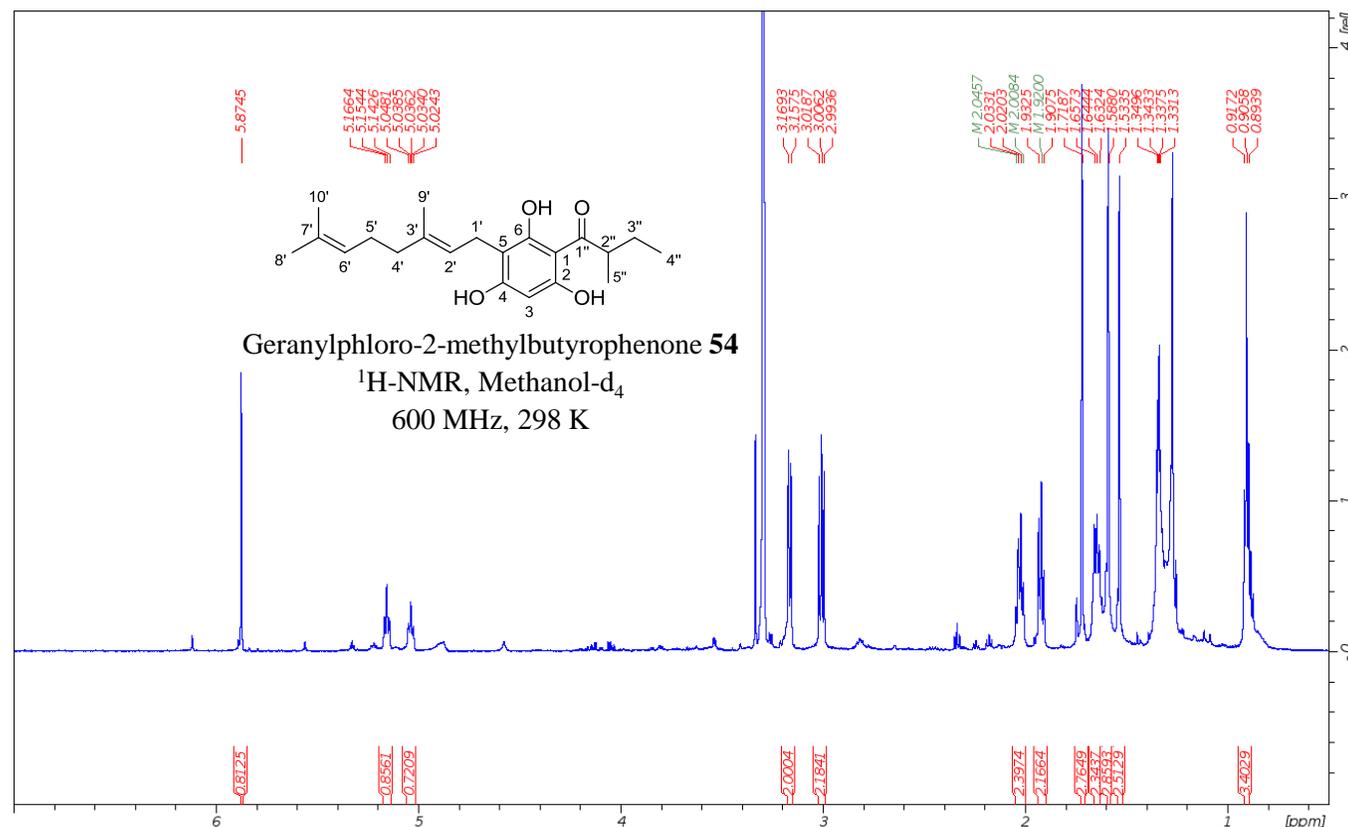
Supplementary NMR Data 6. Geranylphloro-2-methylbutyrophenone 54.

Data of geranylphloro-2-methylbutyrophenone **54**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 8.54 (br, 1H), 5.88 (s, 1H), 5.16 (tm, $J = 7.2, 1.1$ Hz, 1H), 5.02 (tt, $J = 7.2, 1.3$ Hz, 1H), 4.59 (br, 1H), 3.85 (q, $J = 6.8$ Hz, 1H), 3.20 (d, $J = 7.2$ Hz, 2H*), 2.01 (m, 2H), 1.93 (m, 2H), 1.79 (m, 1H), 1.72 (s, 3H), 1.58 (d, $J = 0.82$ Hz, 3H), 1.55 (s, 3H), 1.30 (m, 1H), 1.12 (d, $J = 6.8$ Hz, 3H), 0.91 (t, $J = 7.4$ Hz, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 128.2, 126.3, 97.0, 48.1, 42.7, 30.3, 29.7, 27.5, 23.7, 19.6, 18.1, 14.3.

HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{21}\text{H}_{29}\text{O}_4$: 345.2066, found 345.2073.

$^1\text{H-NMR}$

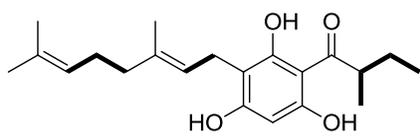
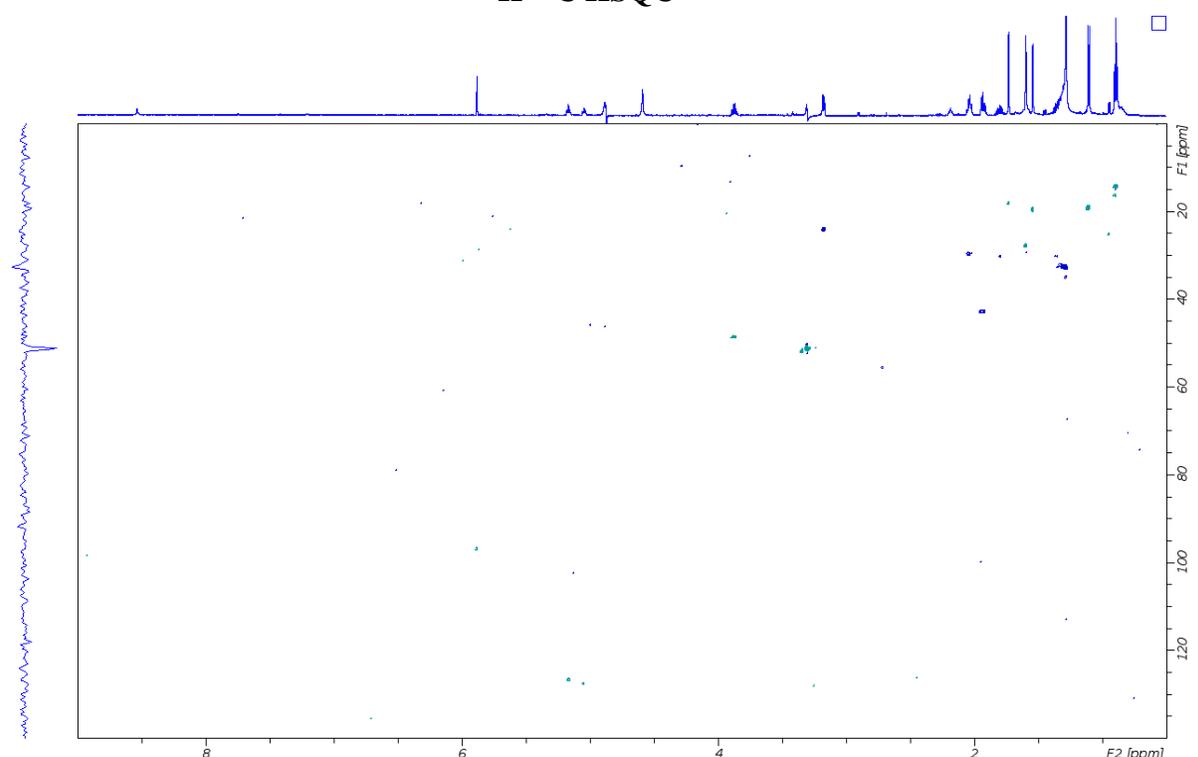
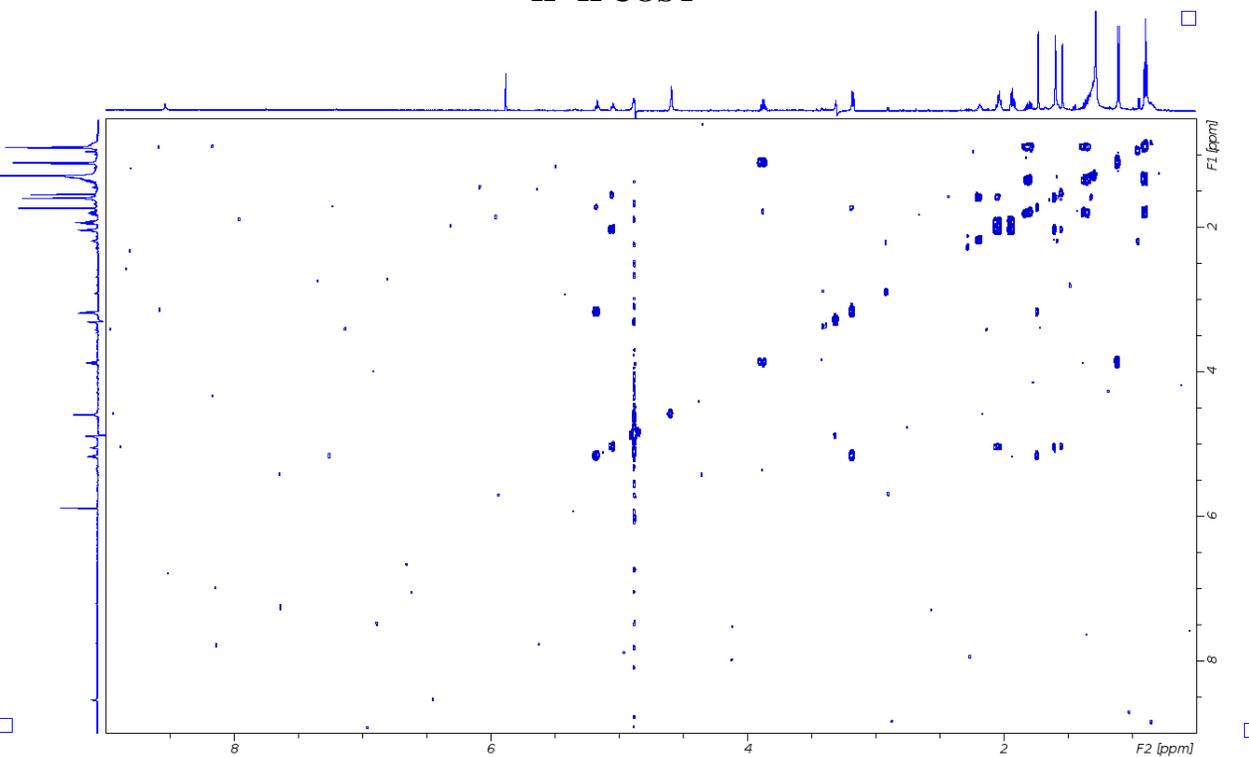


IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
1	na	
2	na	
3	97.0	5.88 (s)
4	na	
5	na	
6	na	
1'	23.7	3.20 (d)
2'	126.3	5.16 (tm)
3'	na	
4'	42.7	1.93 (m)
5'	29.7	2.01 (m)
6'	128.2	5.02 (tt)
7'	na	
8'	27.5	1.58 (d)
9'	18.1	1.72 (s)
10'	19.6	1.55 (s)
1''	na	
2''	48.1	3.85 (q)
3''	30.3	1.30 (m)
3''	30.3	1.79 (m)
4''	14.3	0.91 (t)
5''	19.6	1.12 (d)

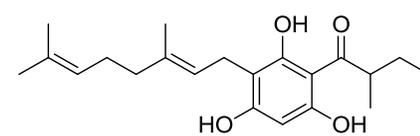
Due to low sample concentration, 2D HMBC spectrum could not be recorded, therefore quaternary carbon shifts could not be assigned. However, the acyl group was identified and the acquired LC-MS/MS spectra and chemical formula confirm the elucidated structure. The ^1H and ^{13}C chemical shifts compare also with those of geranylphlorocaprophenone **4** (Supplementary NMR Data 3).

^1H - ^1H COSY

^1H - ^{13}C HSQC



Geranylphlorocaprophenone 4
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K



Geranylphlorocaprophenone 4
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

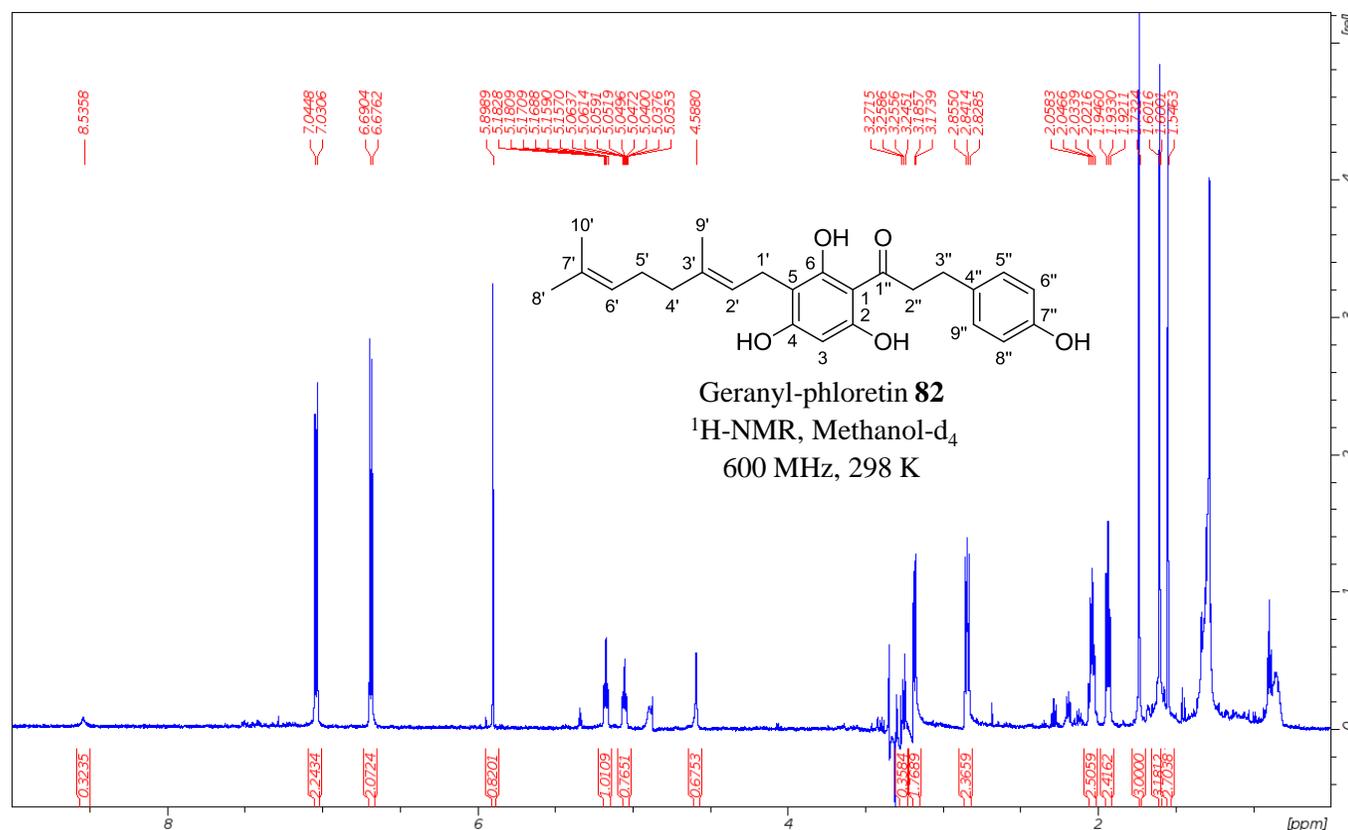
Supplementary NMR Data 7. Geranyl-phloretin **82**.

Data of geranyl-phloretin **82**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 8.54 (br), 7.04 (dt, $J = 8.5, 2.1$ Hz, 2H), 6.68 (dt, $J = 8.5, 2.1$ Hz, 2H), 5.90 (s, 1H), 5.17 (tq, $J = 7.16, 1.15$ Hz, 1H), 5.05 (tt, $J = 7.18, 1.4$ Hz, 1H), 4.59 (br), 3.24-3.28 (m, 2H*), 3.18 (d, $J = 7.16$ Hz, 2H), 2.84 (dd, $J = 7.9$ Hz, 2H), 2.01-2.07 (m, 2H), 1.91-1.95 (m, 2H), 1.73 (s, 3H), 1.60 (d, $J = 0.99$ Hz, 3H), 1.54 (s, 3H);

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 208.3, 167.2, 166.0, 163.3, 158.4, 136.9, 133.7, 132.2, 127.56, 126.7, 118.1, 110.0, 107.0, 96.8, 49.5, 42.9, 33.7, 29.7, 27.8, 24.1, 19.8, 18.2;

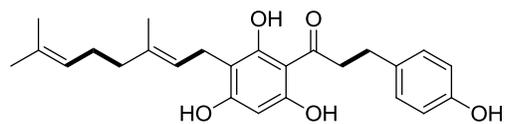
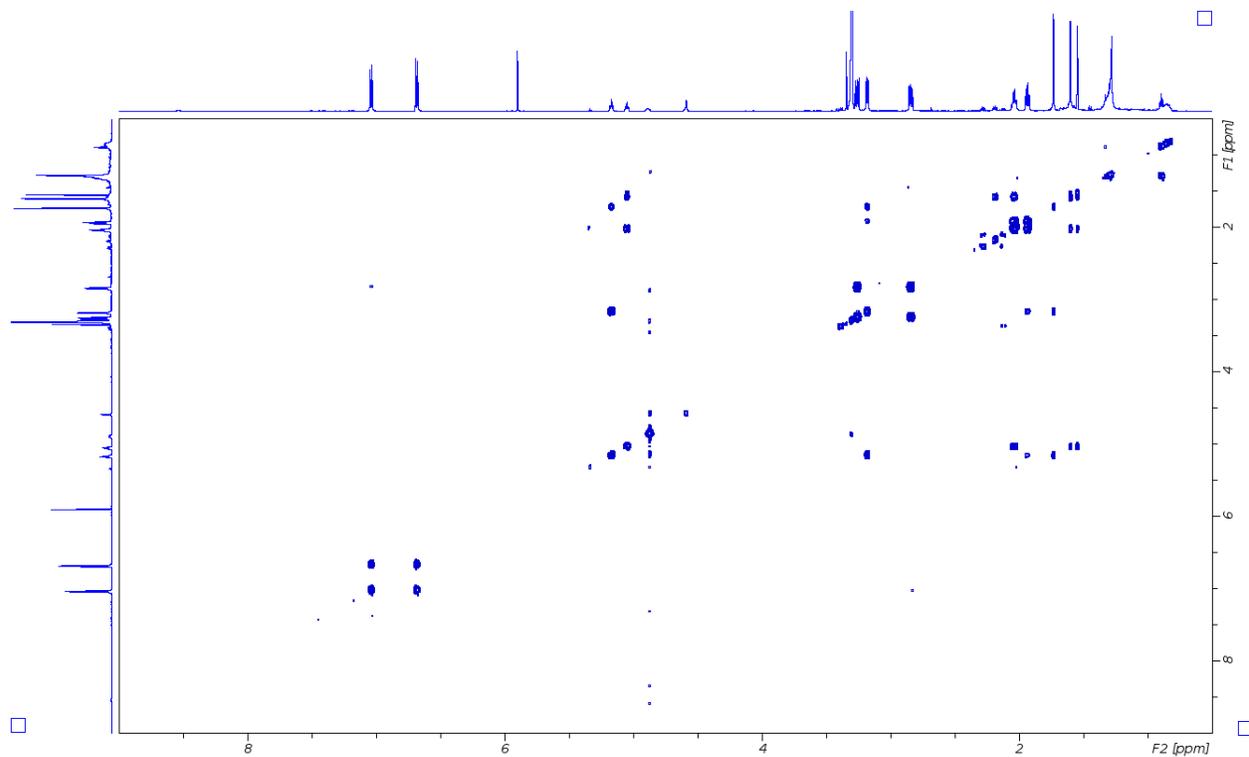
HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{25}\text{H}_{29}\text{O}_5$: 409.2015, found 409.2017.

$^1\text{H-NMR}$



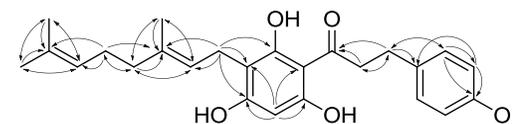
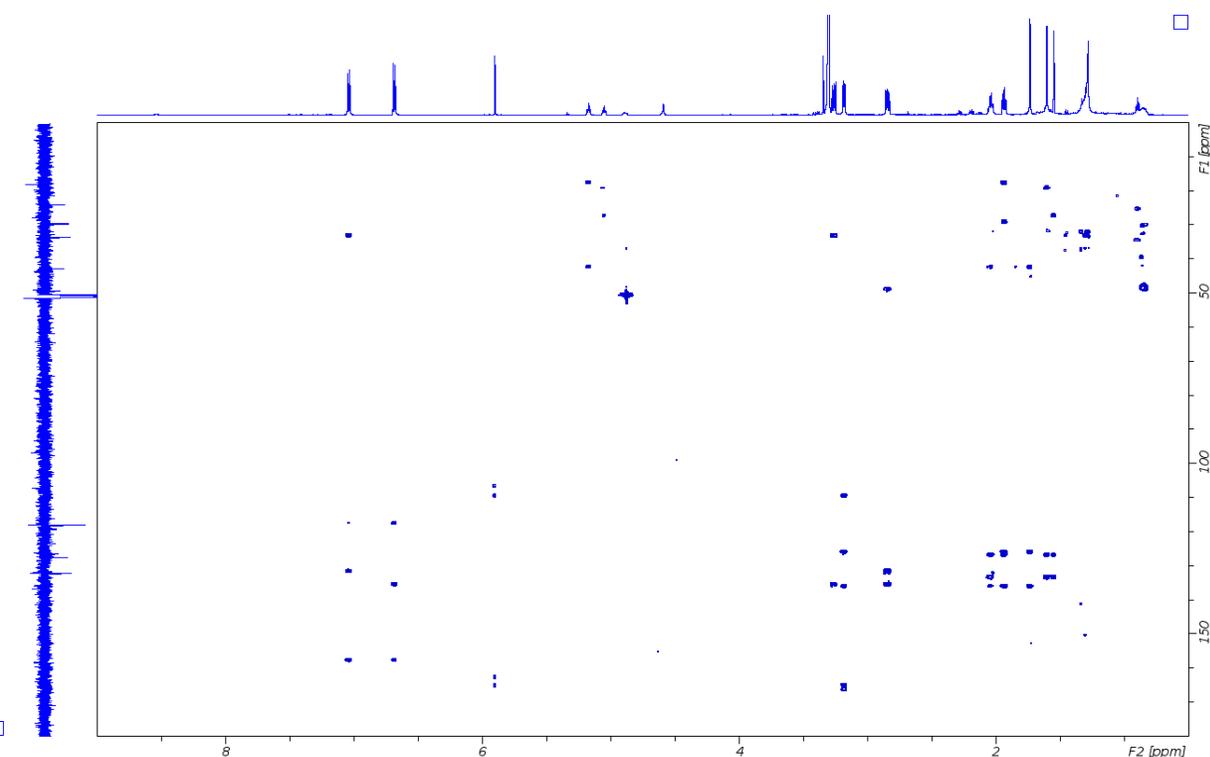
IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
1	107.00	
2	163.32	
3	96.78	5.90 (s)
4	166.01	
5	110.00	
6	167.23	
1'	24.11	3.18 (d)
2'	126.65	5.17 (tq)
3'	136.91	
4'	42.93	1.91-1.95 (m)
5'	29.71	2.01-2.07 (m)
6'	127.56	5.05 (tt)
7'	133.71	
8'	27.84	1.60 (d)
9'	18.23	1.73 (s)
10'	19.80	1.54 (s)
1''	208.32	
2''	49.45	3.24-3.28 (m)
3''	33.67	2.84 (dd)
4''	136.02	
5'', 9''	132.20	7.04 (dt)
6'', 8''	118.08	6.68 (dt)
7''	158.41	

^1H - ^1H COSY



Geranyl-phloretin **82**
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



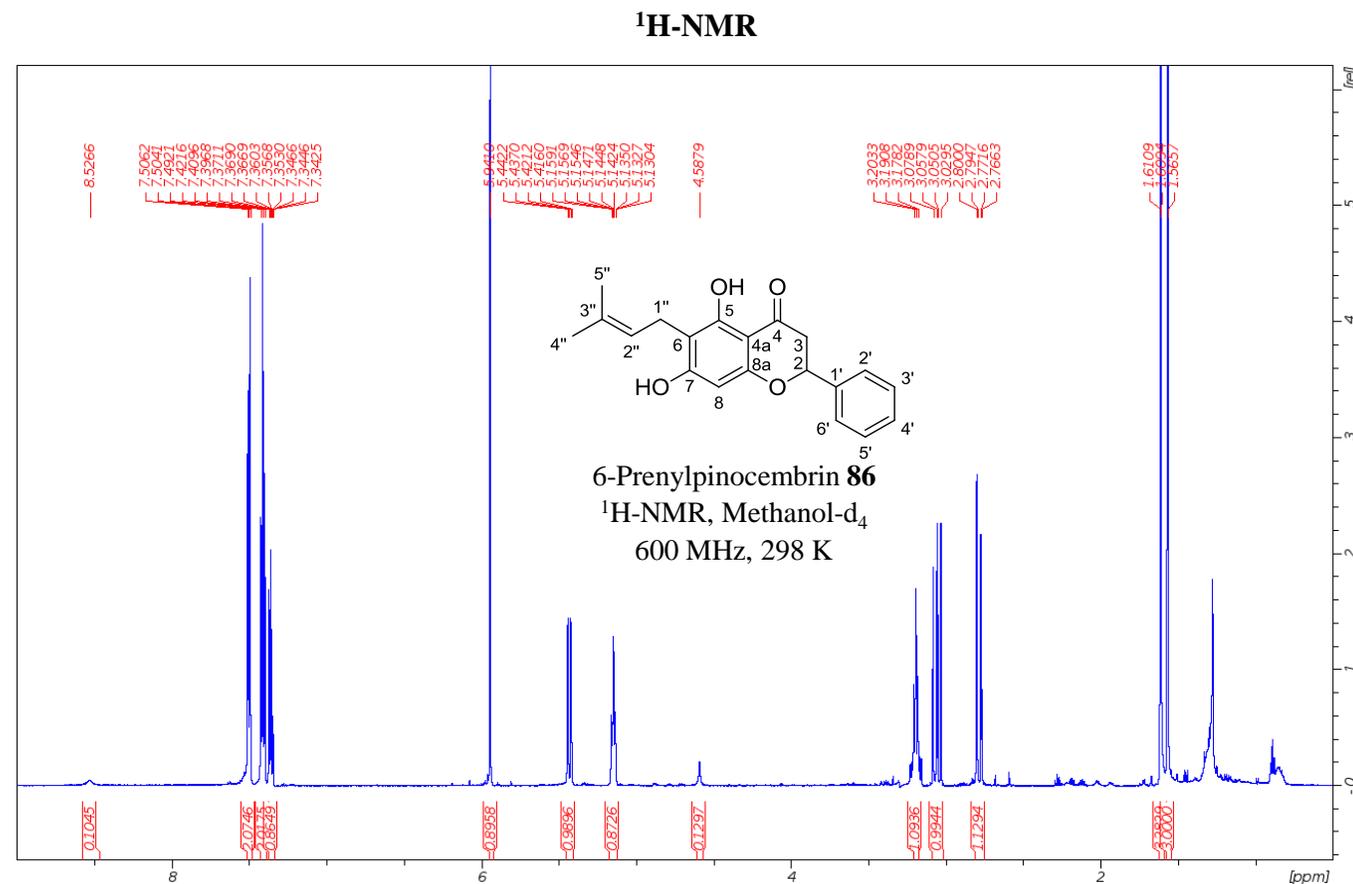
Geranyl-phloretin **82**
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

Supplementary NMR Data 8. 6-Prenylpinocembrin 86.

Data of 6-prenylpinocembrin **86**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 8.53 (br), 7.48-7.51 (m, 2H), 7.39-7.43 (m, 2H), 7.36 (tt, $J = 7.3, 1.3$ Hz, 1H), 5.94 (s, 1H), 5.43 (dd, $J = 12.7, 3.1$ Hz, 1H), 5.14 (tt, $J = 7.3, 1.3$ Hz, 1H), 4.59 (br), 3.19 (m, 2H*), 3.05 (dd, $J = 17.1, 12.7$ Hz, 1H), 2.79 (dd, $J = 17.1, 3.1$ Hz, 1H), 1.61 (d, $J = 0.9$ Hz, 3H), 1.57 (s, 3H).

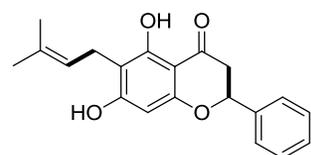
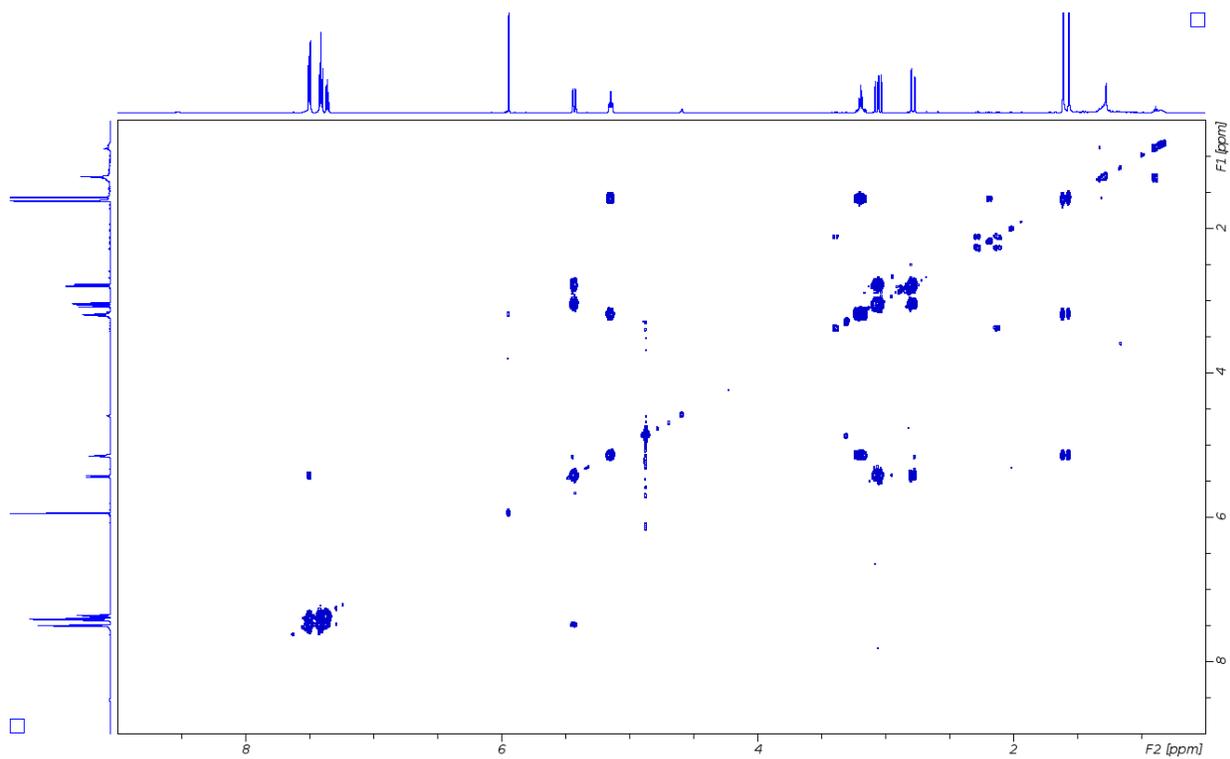
$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 199.8, 168.0, 165.2, 163.2, 142.8, 133.5, 131.5, 131.5, 129.4, 125.9, 111.0, 105.1, 98.4, 82.2, 46.2, 28.4, 24.4, 20.0.

HRMS (ESI) $[\text{M-H}]^- \text{C}_{20}\text{H}_{19}\text{O}_4$: 323.1283, found 323.1285.



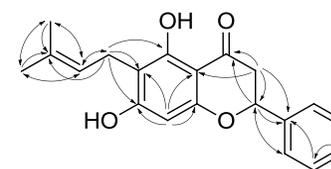
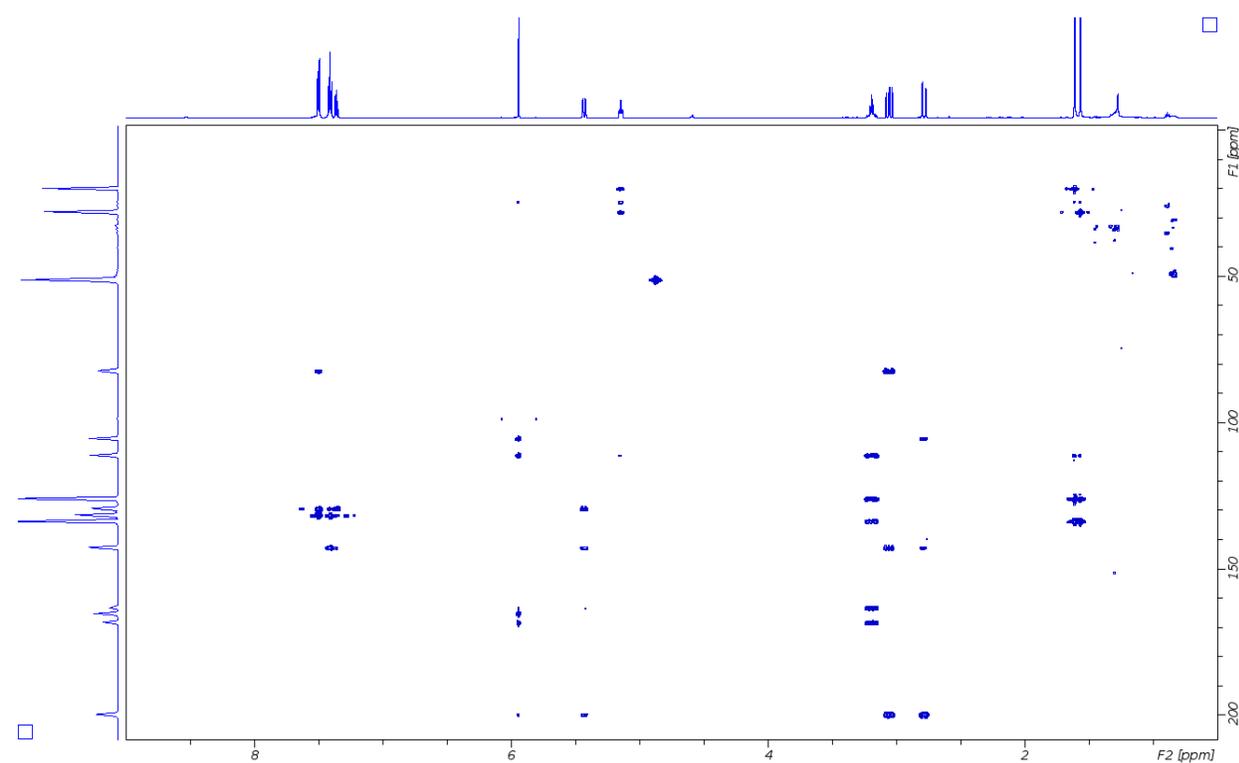
IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
2	82.2	5.43 (dd)
3	46.2	2.79 (dd)
3	46.2	3.05 (dd)
4	199.8	
4a	105.1	
5	163.2	
6	111.0	
7	168.0	
8	98.4	5.94 (s)
8a	165.2	
1'	142.8	
2',6'	129.4	7.48-7.51 (m)
3',5'	131.5	7.39-7.43 (m)
4'	131.5	7.36 (tt)
1''	24.4	3.19 (m)
2''	125.9	5.14 (tt)
3''	133.5	
4''	28.4	1.61 (d)
5''	20.0	1.57 (s)

^1H - ^1H COSY



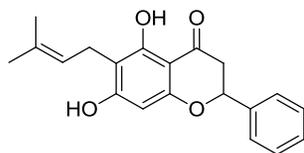
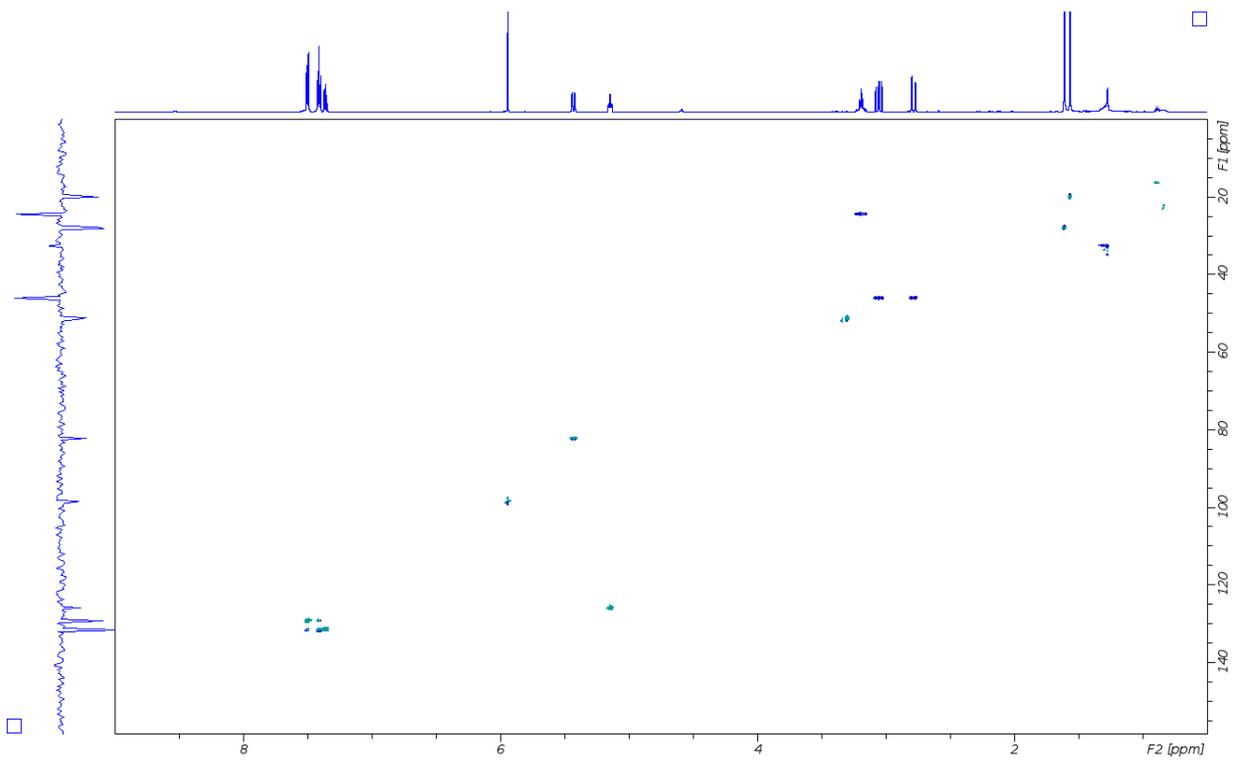
6-Prenylpinocembrin 86
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



6-Prenylpinocembrin 86
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



6-Prenylpinocembrin 86
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

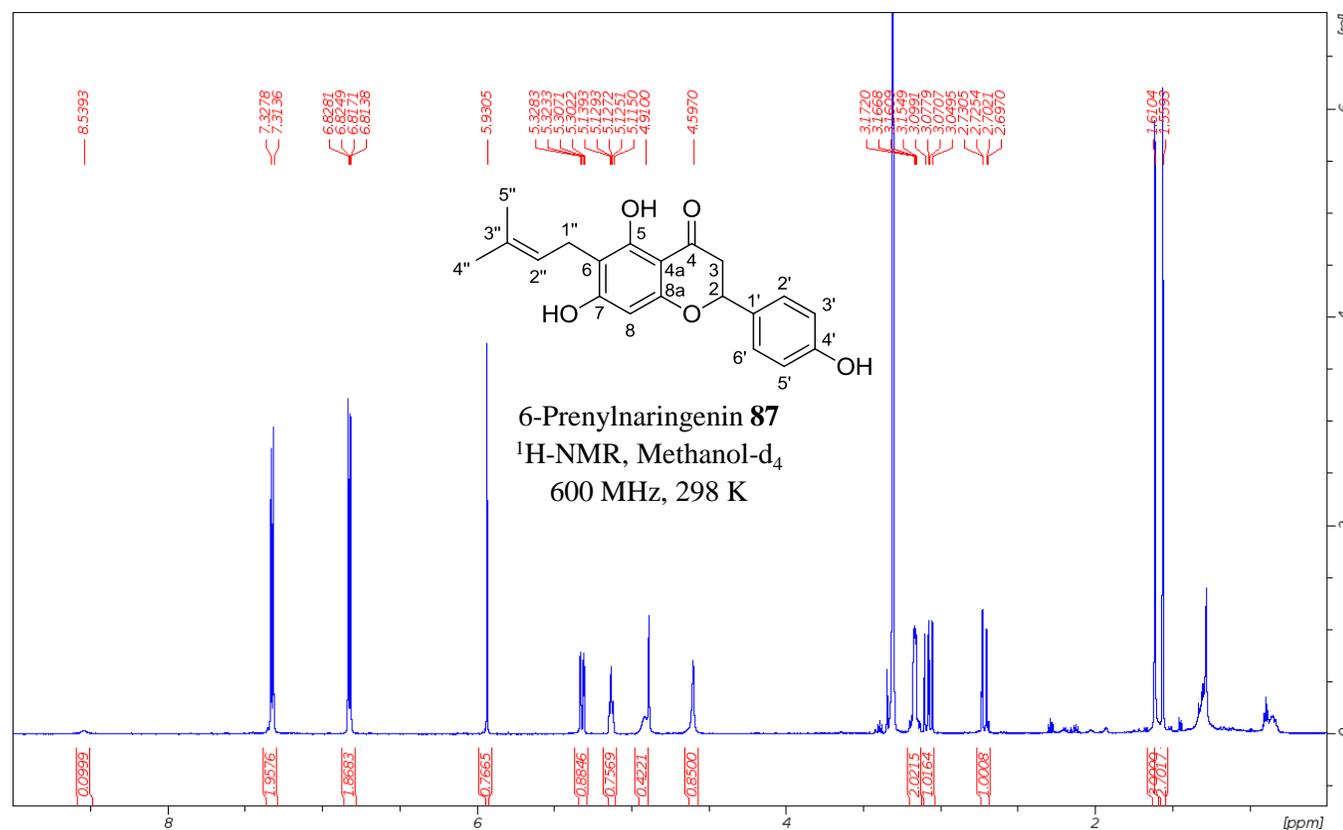
Supplementary NMR Data 9. 6-Prenylnaringenin **87**.

Data of 6-prenylnaringenin **87**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 7.32 (dt, $J = 8.6, 1.9$ Hz, 2H), 6.83 (dt, $J = 8.6, 1.9$ Hz, 2H), 5.93 (s, 1H), 5.32 (dd, $J = 12.7, 3.1$ Hz, 1H), 5.13 (tt, $J = 7.3, 1.3$ Hz, 1H*), 4.60 (br), 3.13-3.20 (m, 2H), 3.08 (dd, $J = 17.1, 12.7$ Hz, 1H), 2.72 (dd, $J = 17.1, 3.1$ Hz, 1H), 1.61 (s, 3H), 1.56 (s, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 200.1, 168.0, 164.8, 163.6, 160.9, 133.6, 133.1, 130.8, 125.7, 118.1, 111.0, 104.9, 98.3, 82.1, 45.9, 27.8, 24.3, 19.7.

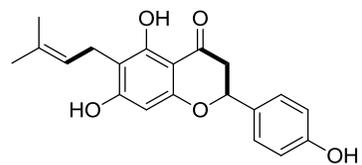
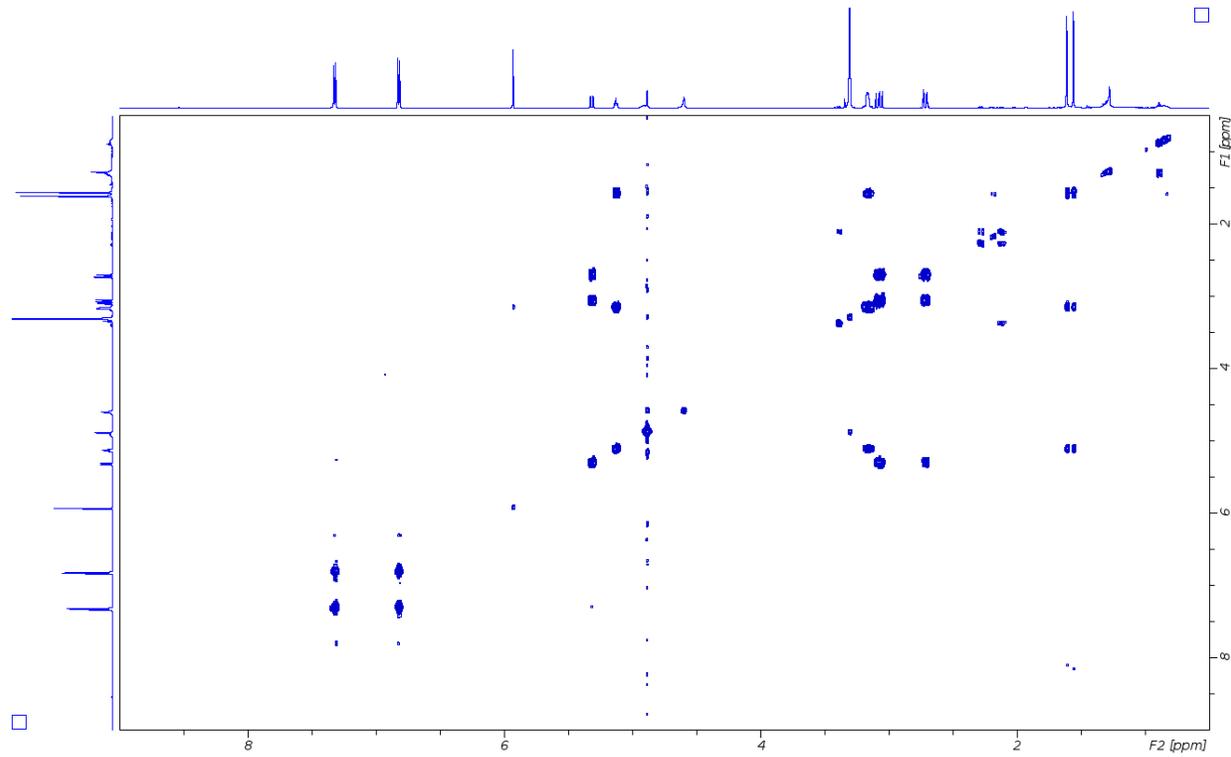
HRMS (ESI) $[\text{M-H}]^- \text{C}_{20}\text{H}_{19}\text{O}_5$: 339.1232, found 339.1242.

$^1\text{H-NMR}$



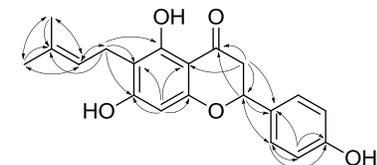
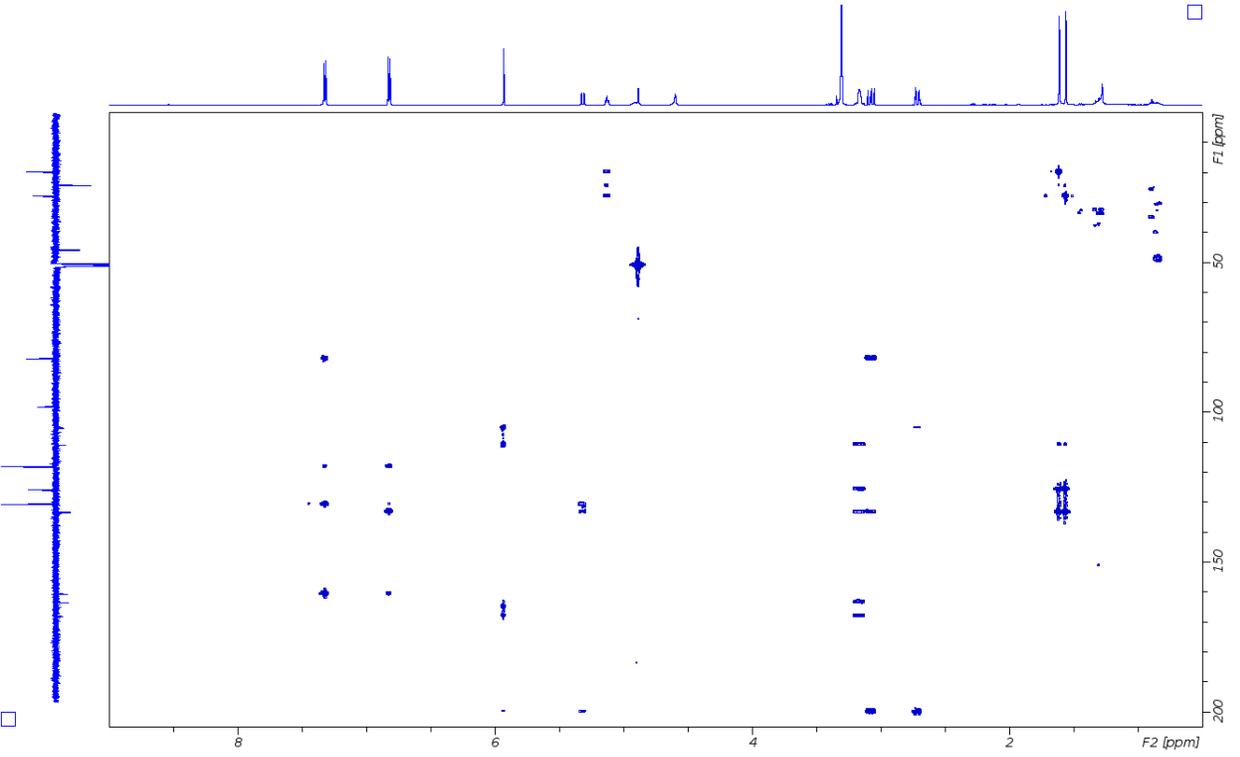
IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
2	82.1	5.32(dd)
3	45.9	2.72 (dd)
3	45.9	3.08 (dd)
4	200.1	
4a	104.9	
5	163.6	
6	111.0	
7	168.0	
8	98.3	5.93 (s)
8a	164.8	
1'	133.1	
2', 6'	130.8	7.32 (dt)
3', 5'	118.1	6.83 (dt)
4'	160.9	
1''	24.3	3.13-3.20 (m)
2''	125.7	5.13 (tt)
3''	133.6	
4''	27.8	1.61 (s)
5''	19.7	1.56 (s)

^1H - ^1H COSY



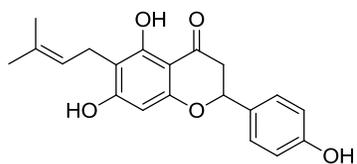
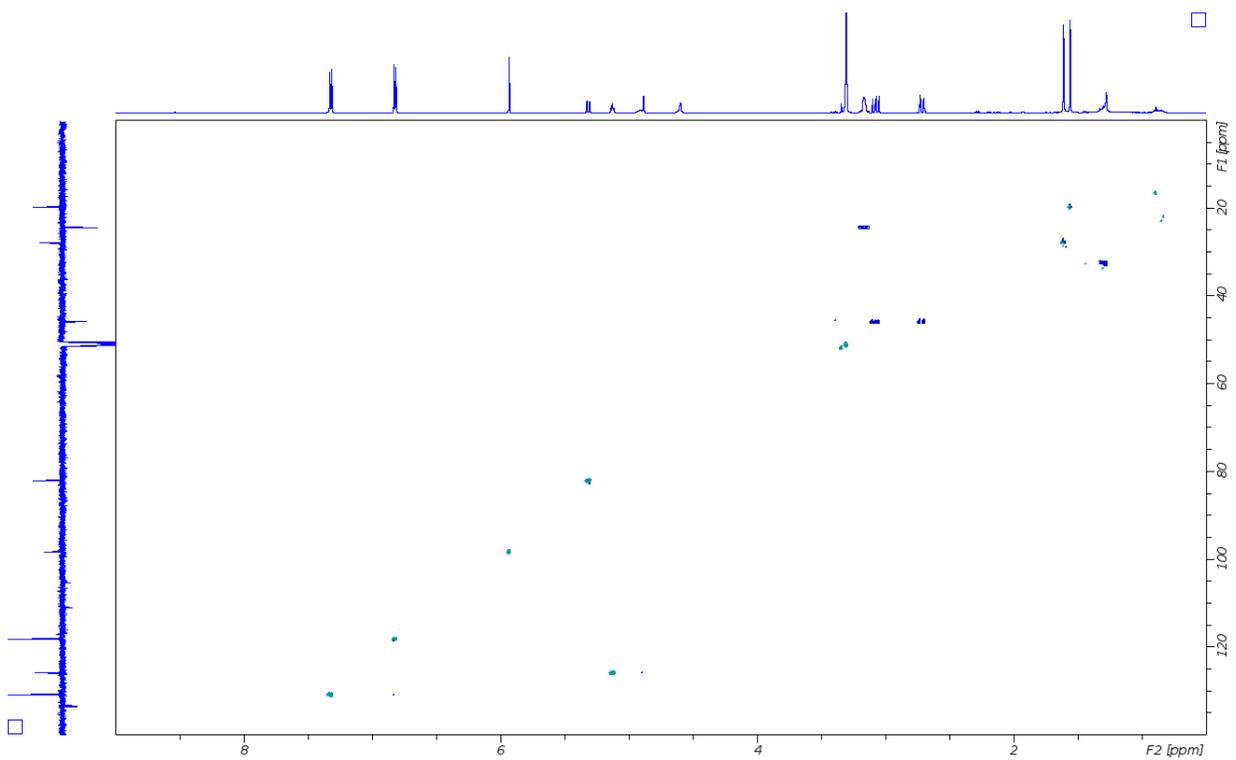
6-Prenylnaringenin **87**
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



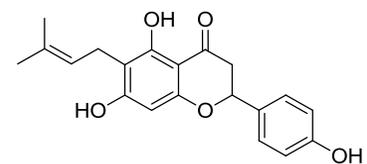
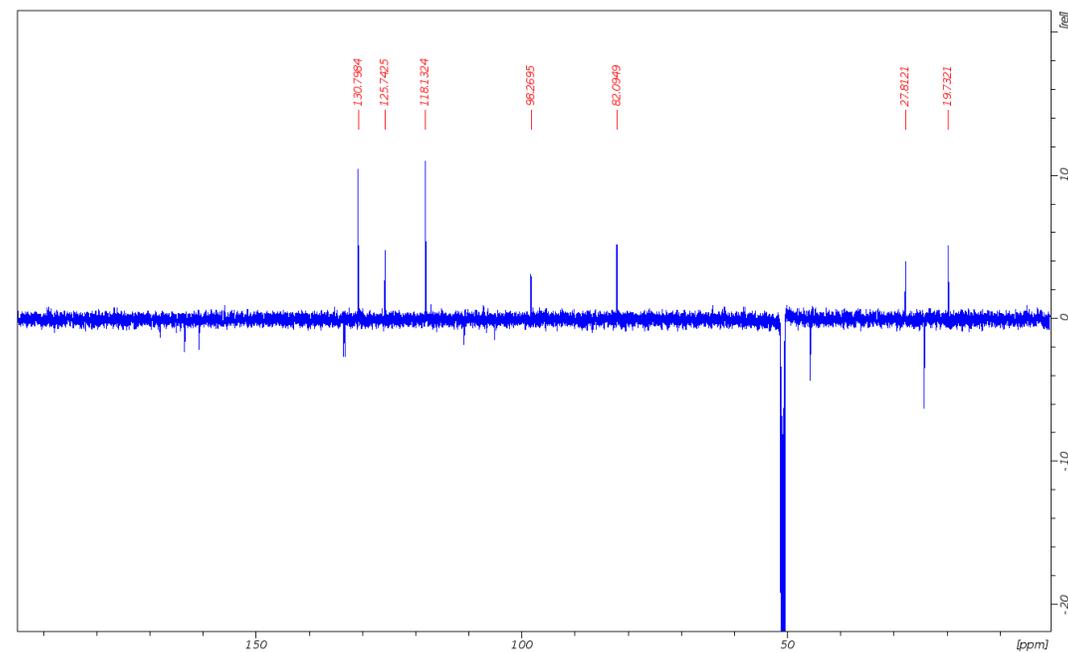
6-Prenylnaringenin **87**
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



6-Prenylnaringenin **87**
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

^{13}C NMR



6-Prenylnaringenin **87**
 ^{13}C NMR, 150.9 MHz,
Methanol- d_4 , 298 K

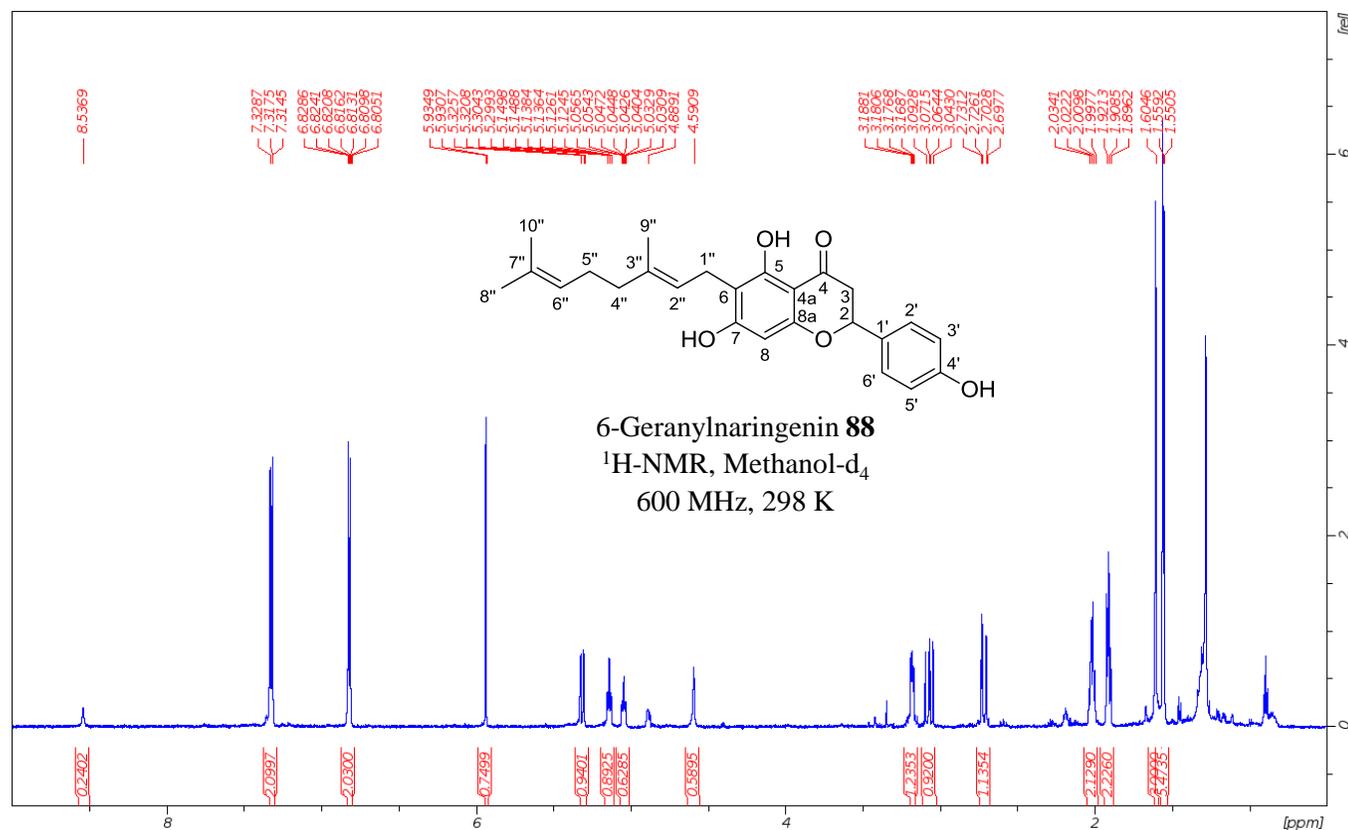
Supplementary NMR Data 10. 6-Geranylningenin 88.

Data of 6-geranylningenin **88**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 8.54 (br), 7.32 (dt, $J = 8.6, 1.8$ Hz, 2H), 6.82 (dt, $J = 8.6, 1.8$ Hz, 2H), 5.94 (s, 1H), 5.31 (dd, $J = 12.8, 3.0$ Hz, 1H), 5.14 (td, $J = 7.2, 1.2$ Hz, 1H), 5.04 (tt, $J = 7.1, 1.3$ 4Hz, 1H), 4.59 (br), 3.18 (m, 2H*), 3.07 (dd, $J = 17.0, 12.8$ Hz, 1H), 2.72 (dd, $J = 17.0, 3.0$ Hz, 1H), 2.02 (q, $J = 7.6$ Hz, 2H), 1.91 (t, $J = 7.6$ Hz, 2H), 1.61 (d, $J = 0.9$ Hz, 3H), 1.56 (s, 3H), 1.53 (s, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 200.1, 168.0, 165.1, 163.5, 160.9, 137.1, 134.0, 132.3, 130.9, 127.5, 125.9, 118.4, 111.1, 105.3, 98.4, 82.1, 45.9, 42.8, 29.6, 27.8, 24.3, 19.6, 18.1.

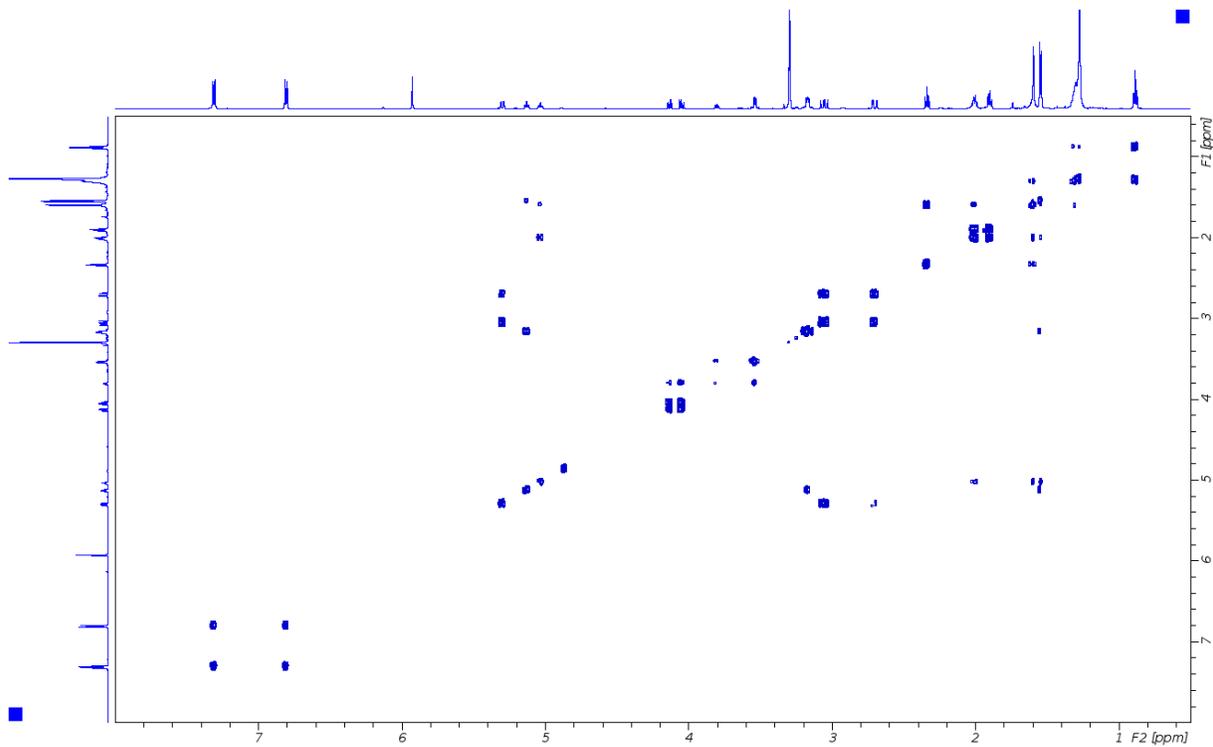
HRMS (ESI) $[\text{M-H}]^- \text{C}_{25}\text{H}_{27}\text{O}_5$: 407.1858, found 407.1861.

$^1\text{H-NMR}$



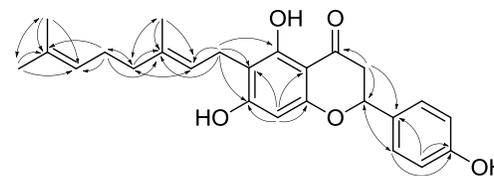
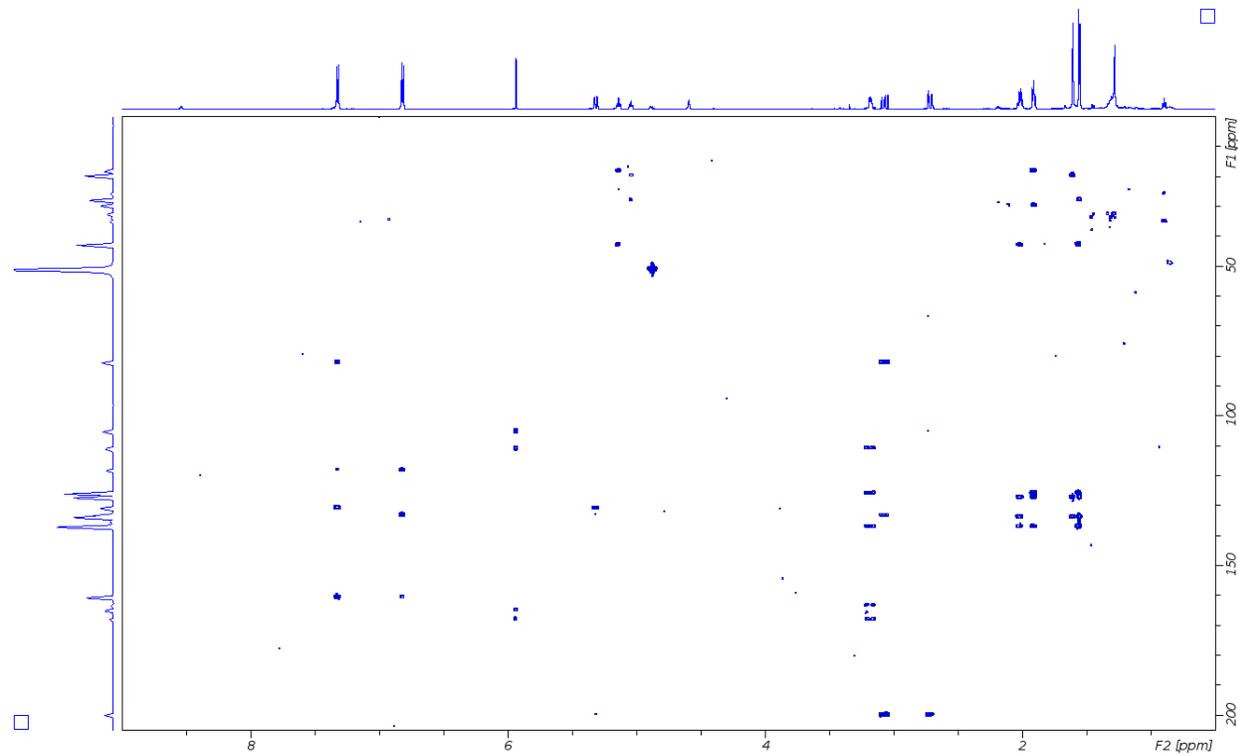
IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
2	82.1	5.31 (dd)
3	45.9	2.72 (dd)
3	45.9	3.07(dd)
4	200.1	
4a	105.3	
5	163.5	
6	111.1	
7	168.0	
8	98.4	5.94 (s)
8a	165.1	
1'	132.3	
2',6'	130.9	7.32 (dt)
3',5'	118.4	6.82 (dt)
4'	160.9	
1''	24.3	3.18 (m)
2''	125.9	5.14 (td)
3''	137.1	
4''	42.8	1.91 (t)
5''	29.6	2.02 (q)
6''	127.5	5.04 (tt)
7''	134.0	
8''	27.8	1.61(d)
9''	18.1	1.56 (s)
10''	19.6	1.53 (s)

^1H - ^1H COSY



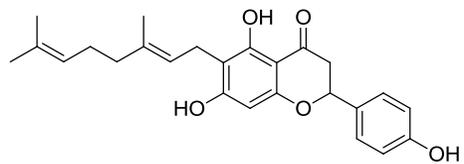
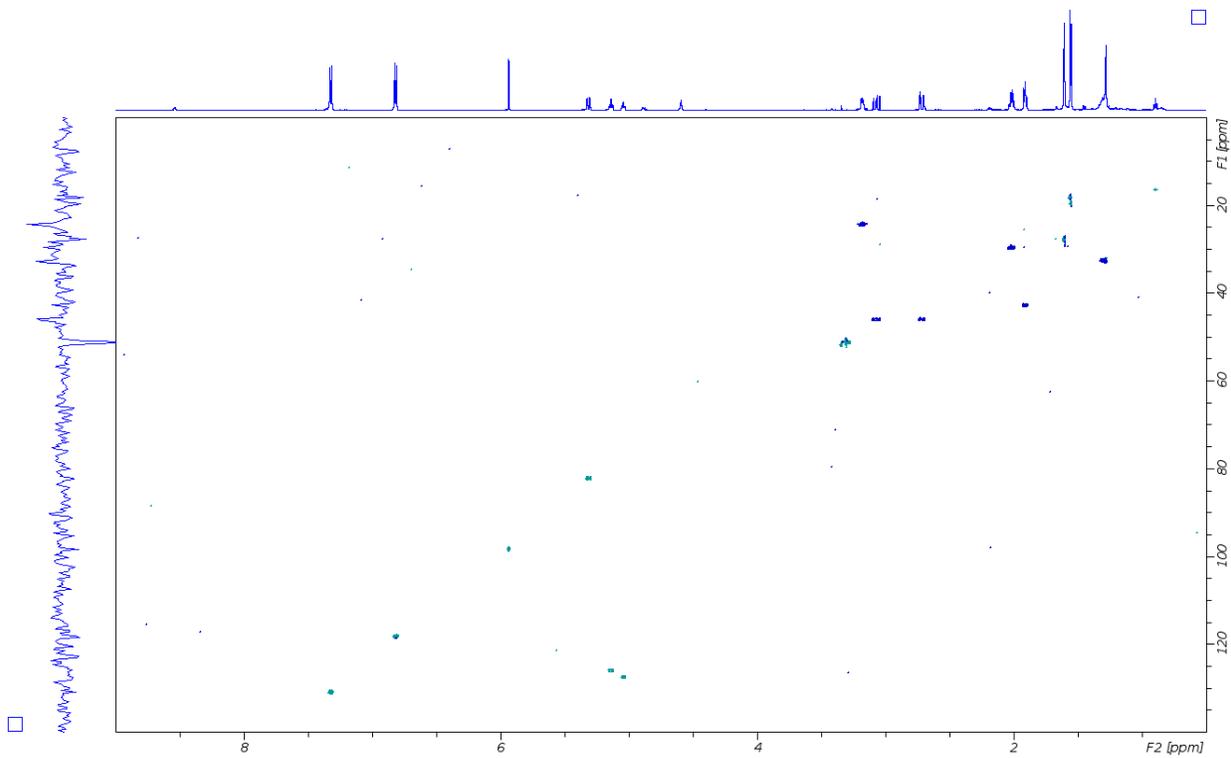
6-Geranylnaringenin 88
 ^1H - ^1H COSY
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



6-Geranylnaringenin 88
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



6-Geranylnaringenin 88

^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

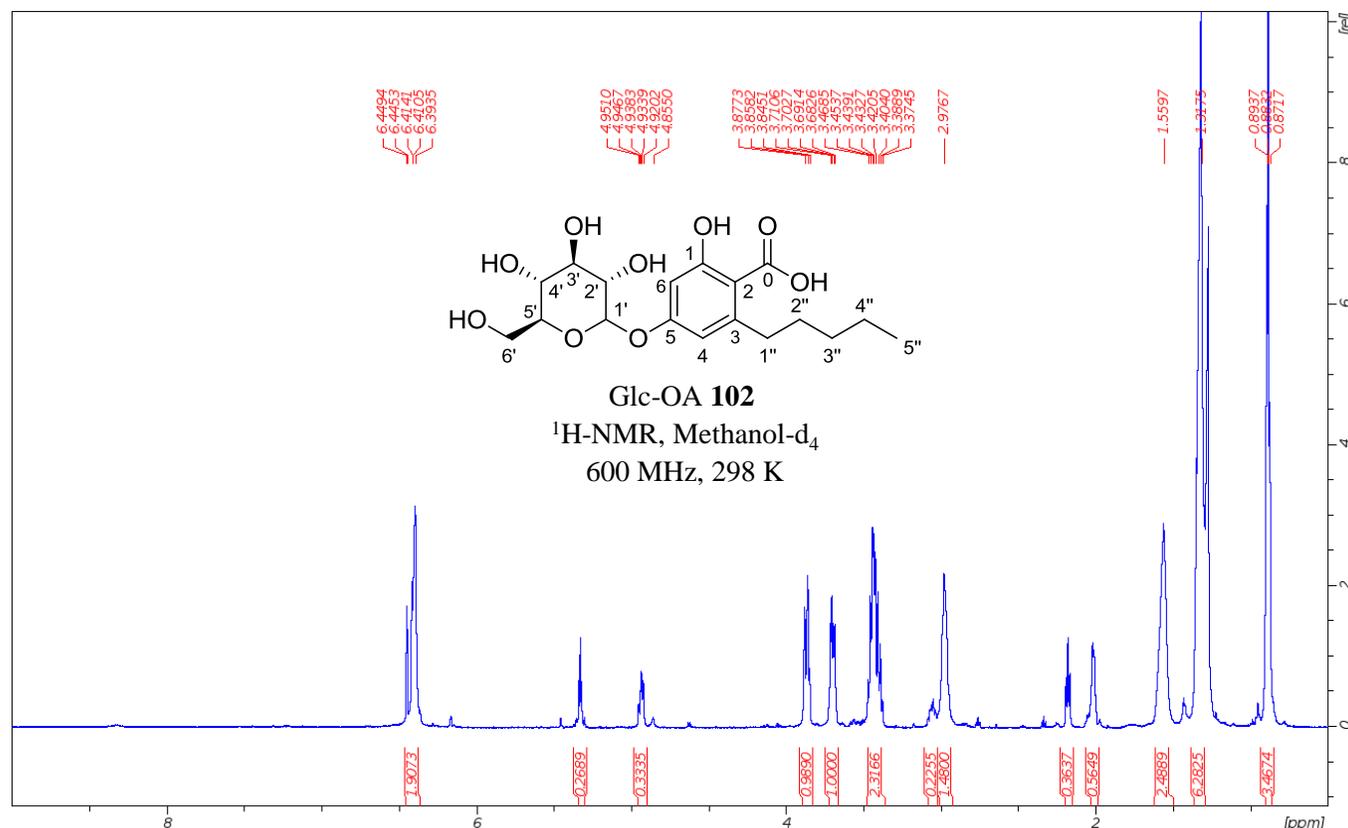
Supplementary NMR Data 11. Glc-OA 102.

Data of Glc-OA **102**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 6.44 (br, 1H), 6.41 (br, 1H), 4.91-4.96 (m, 1H*), 3.84-3.89 (m, 1H), 3.71 (dd, $J = 11.5, 4.98$ Hz, 1H), 3.45 (m, 1H*), 3.43 (m, 1H*), 3.42 (t, $J = 8.6$ Hz, 1H*), 3.39 (t, $J = 8.6$ Hz, 1H*), 2.94-3.0 (m, 2H), 1.52-1.60 (m, 2H), 1.32 (m, 4H), 0.90 (m, 3H);

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 165.3, 162.9, 152.0, 113.5, 110.5, 104.3, 103.3, 80.0, 79.7, 76.6, 73.2, 64.4, 38.8, 35.2, 34.6, 25.6, 16.7;

HRMS (ESI) $[\text{M-H}]^- \text{C}_{18}\text{H}_{25}\text{O}_9$: 385.1499, found 385.1511.

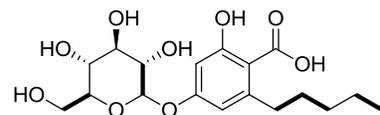
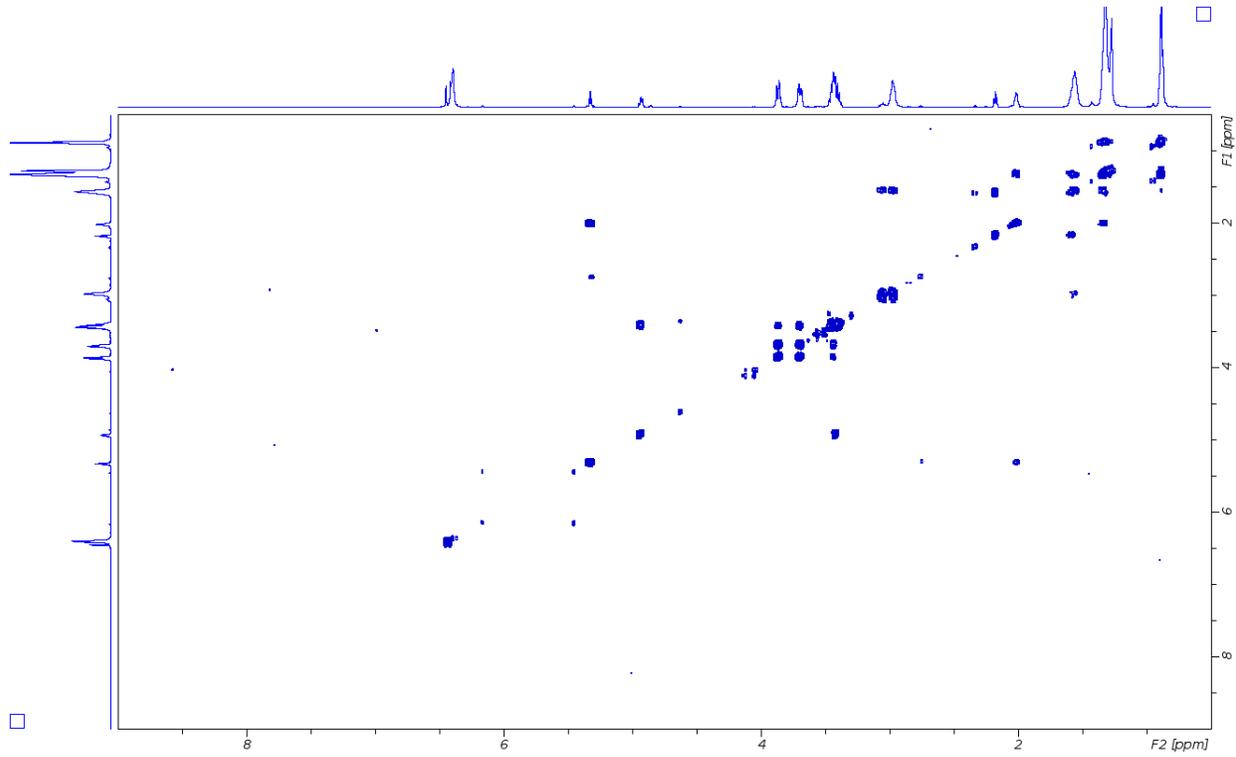
$^1\text{H-NMR}$



IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
0	nd ^a	
1	165.3	
2	110.5	
3	152.0	
4	113.5	6.41 (br)
5	162.9	
6	104.3	6.42 (br)
1'	103.3	4.91-4.96 (m)
2'	79.8	3.42 (t)
3'	76.8	3.45 (m)
4'	79.8	3.39(t)
5'	73.2	3.43 (m)
6'	64.4	3.71 (dd)
6''	64.4	3.84-3.89 (m)
1''	38.8	2.94-3.0 (m)
2''	34.6	1.52-1.60 (m)
3''	35.2	1.32 (m)
4''	25.6	1.32 (m)
5''	16.7	0.90 (m)

^aThe ^{13}C carbonyl signal was not observed in the NMR spectra, however LC-MS/MS spectra and chemical formula confirm the presence of this group.

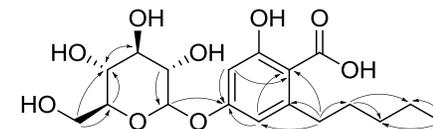
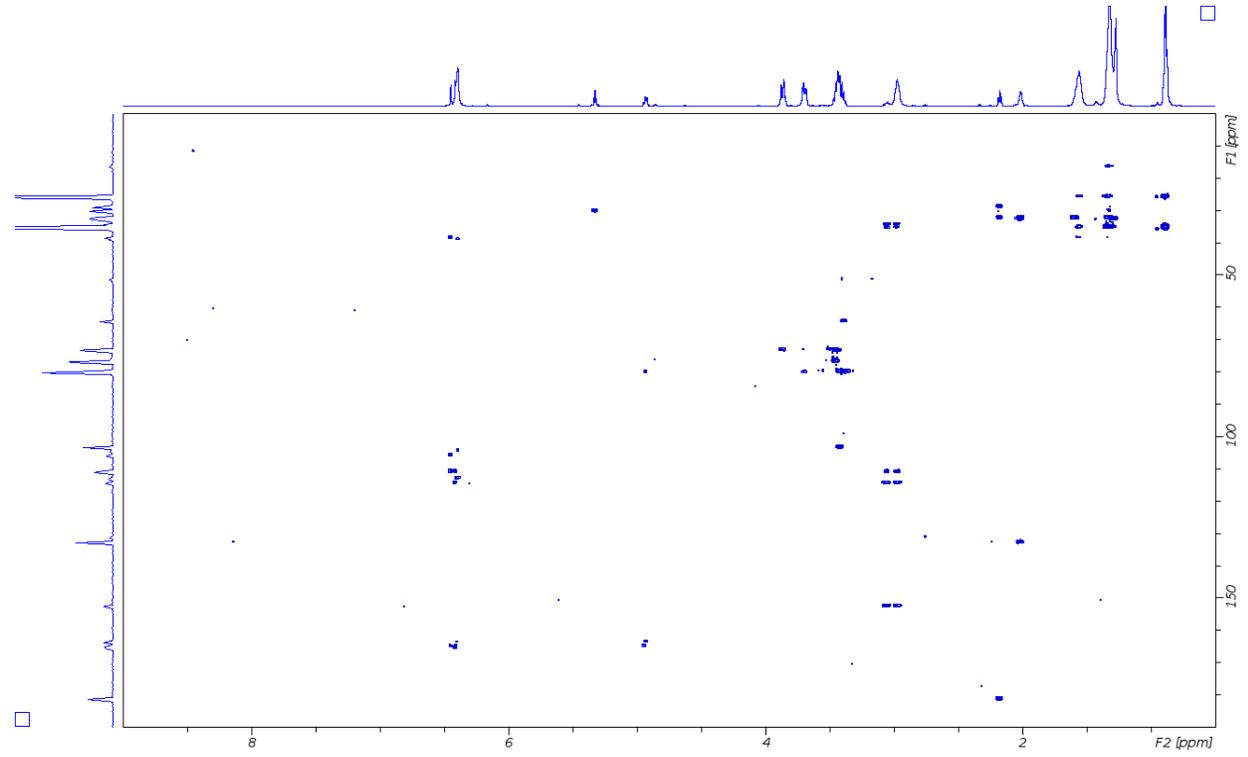
^1H - ^1H COSY



Glc-OA **102**

^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

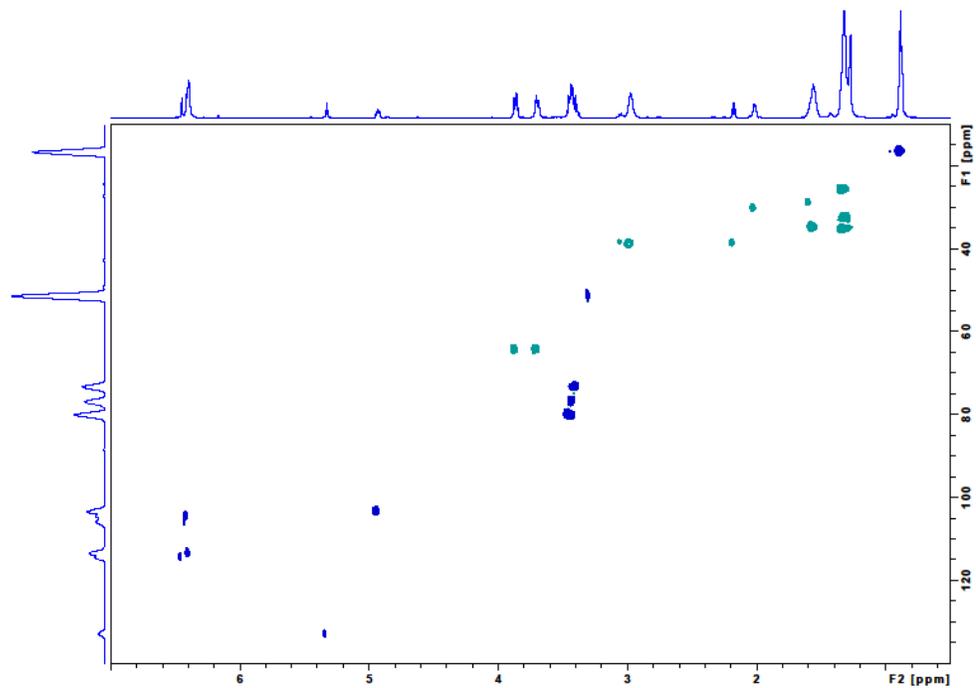
^1H - ^{13}C HMBC



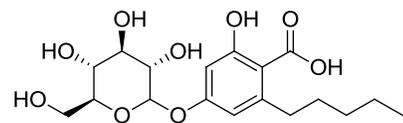
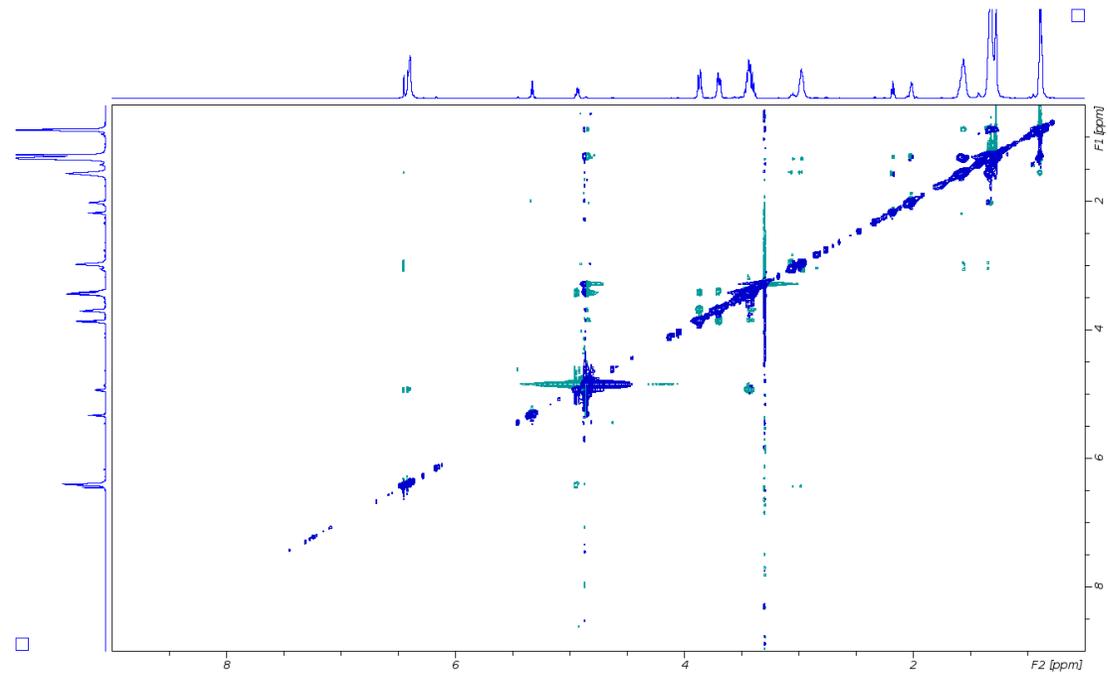
Glc-OA **102**

^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



ROESY



Glc-OA 102
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

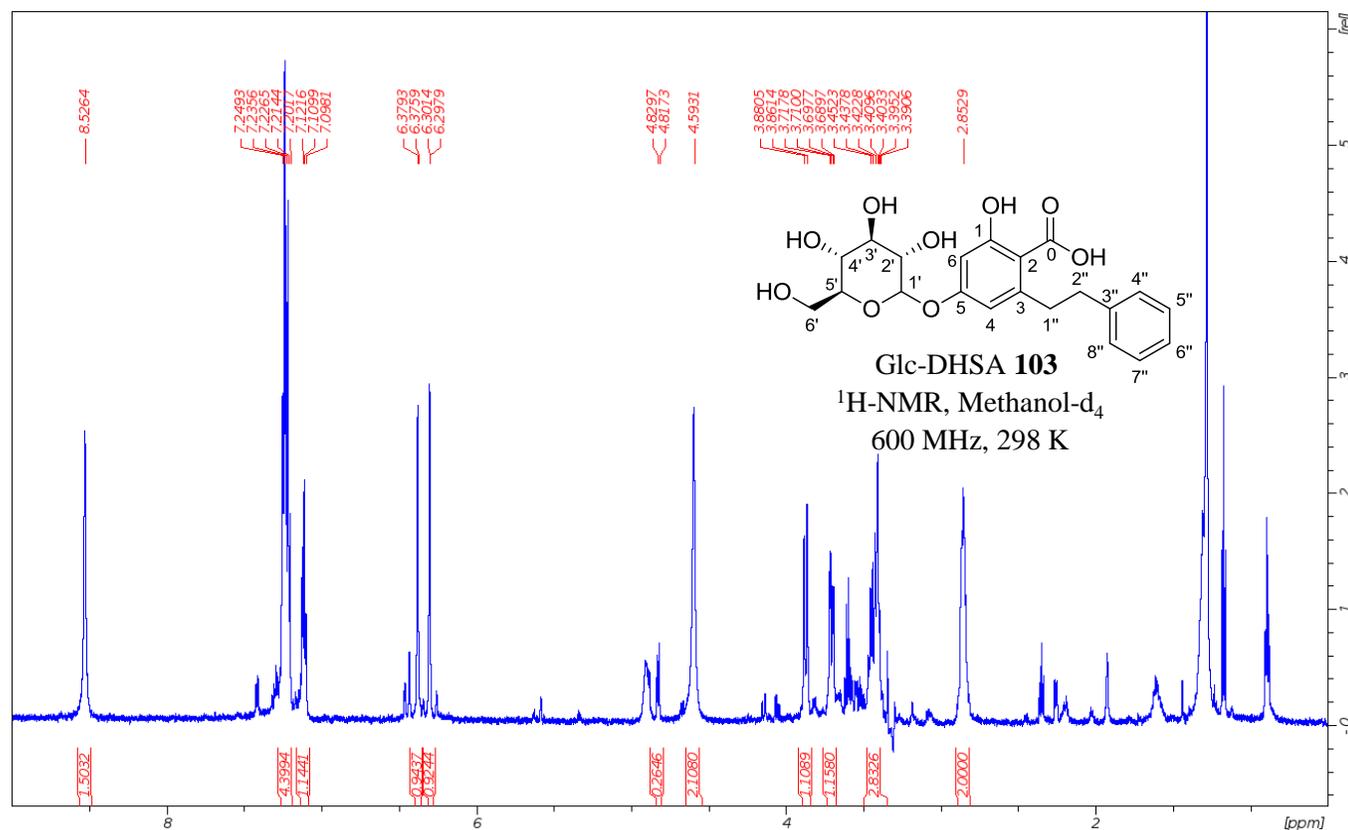
Supplementary NMR Data S12. Glc-DHSA 103.

Data of Glc-DHSA **103**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 8.53 (br), 7.19-7.26 (m, 4H), 7.11 (t, $J = 7.1$ Hz, 1H), 6.38 (d, $J = 2.1$ Hz, 1H), 6.30 (d, $J = 2.1$ Hz, 1H), 4.82 (d, $J = 7.5$ Hz, 1H*), 4.59 (br), 3.87 (d, $J = 12.0$ Hz, 1H), 3.70 (dd, $J = 12.0, 4.6$ Hz, 1H), 3.44 (m, 1H), 3.42 (m, 1H), 3.39 (t, $J = 8.9$ Hz, 1H*), 3.38 (t, $J = 4.6$ Hz, 1H*), 3.27 (d, $J = 9.6$ Hz, 1H*), 2.85 (m, 2H);

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 166.8, 162.5, 146.0, 131.6, 131.1, 128.5, 115.6, 112.9, 104.7, 103.5, 80.0, 79.8, 76.6, 73.2, 64.3, 41.3, 41.2;

HRMS (ESI) $[\text{M-H}]^- \text{C}_{21}\text{H}_{23}\text{O}_9$: 19.1342, found 419.1355.

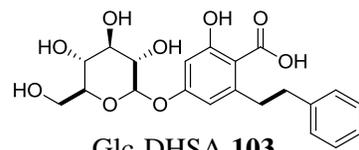
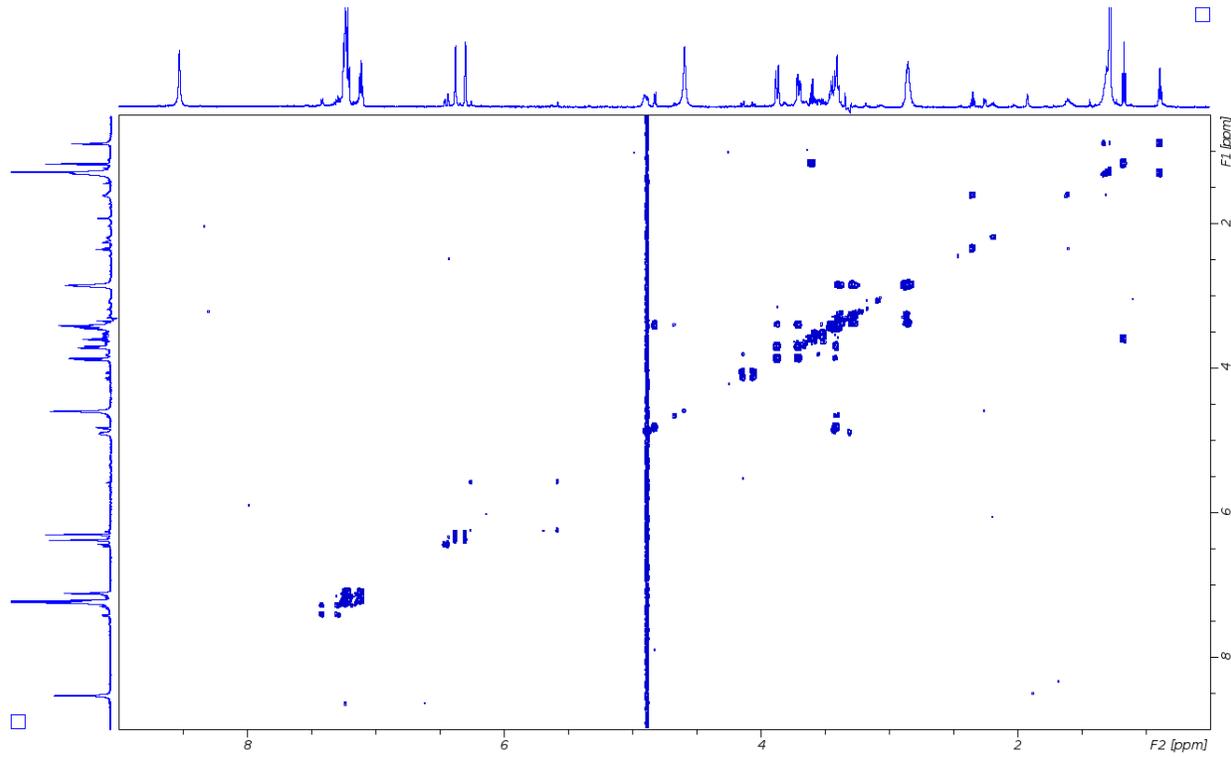
$^1\text{H-NMR}$



IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
0	nd ^a	
1	166.8	
2	115.6	
3	131.7	
4	112.9	6.30 (d)
5	162.5	
6	104.7	6.38 (d)
1'	103.5	4.82 (d)
2'	76.6	3.42 (t)
3'	79.8	3.44 (m)
4'	73.2	3.39 (t)
5'	79.8	3.42 (m)
6'	64.3	3.70 (dd)
6''	64.3	3.87 (d)
1''	41.2	2.85 (m)
2''	41.3	3.27 (d)
2'''	41.3	3.38 (t)
3''	146.0	
4'',8''	131.6	7.19-7.26 (m)
5'',7''	131.1	7.19-7.26 (m)
6''	128.5	7.11 (t)

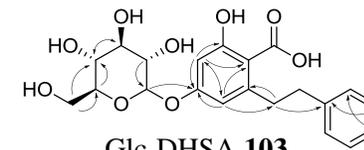
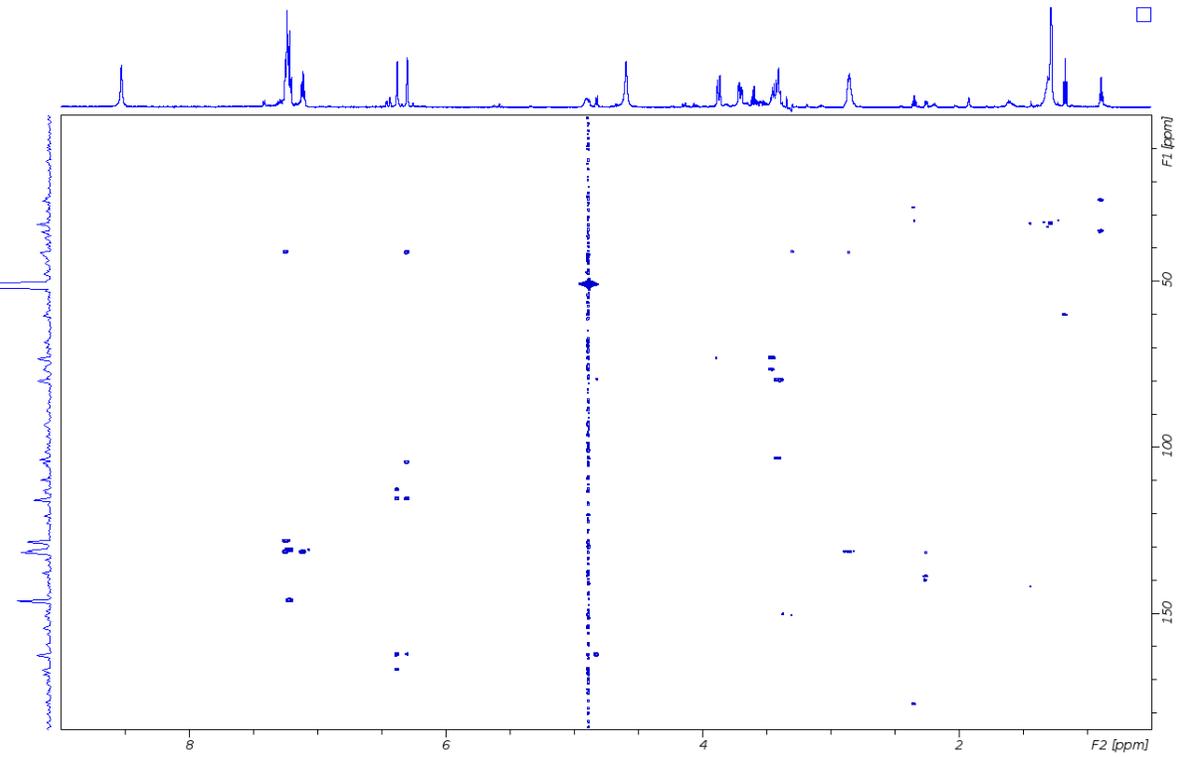
^aThe ^{13}C carbonyl signal was not observed in the NMR spectra, however LC-MS/MS spectra and chemical formula confirm the presence of this group.

^1H - ^1H COSY



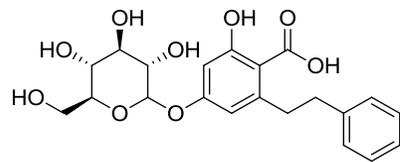
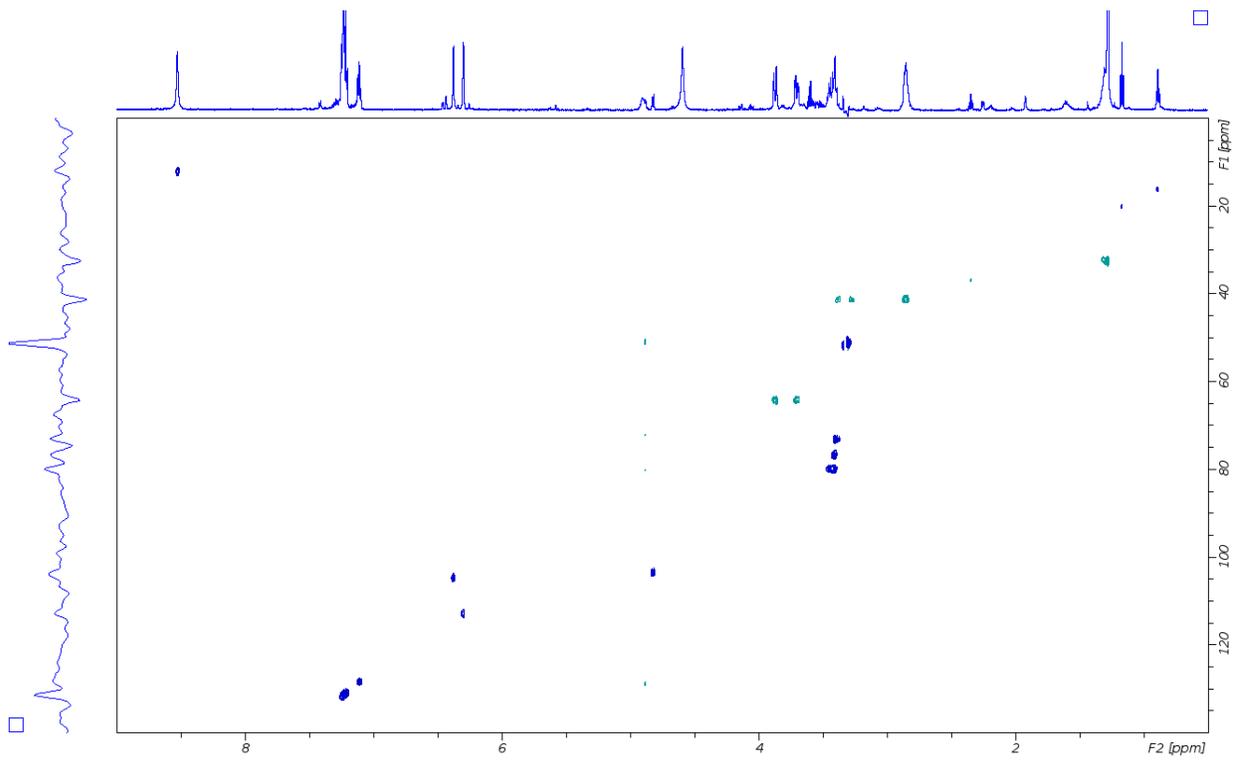
^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



Glc-DHSA 103
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

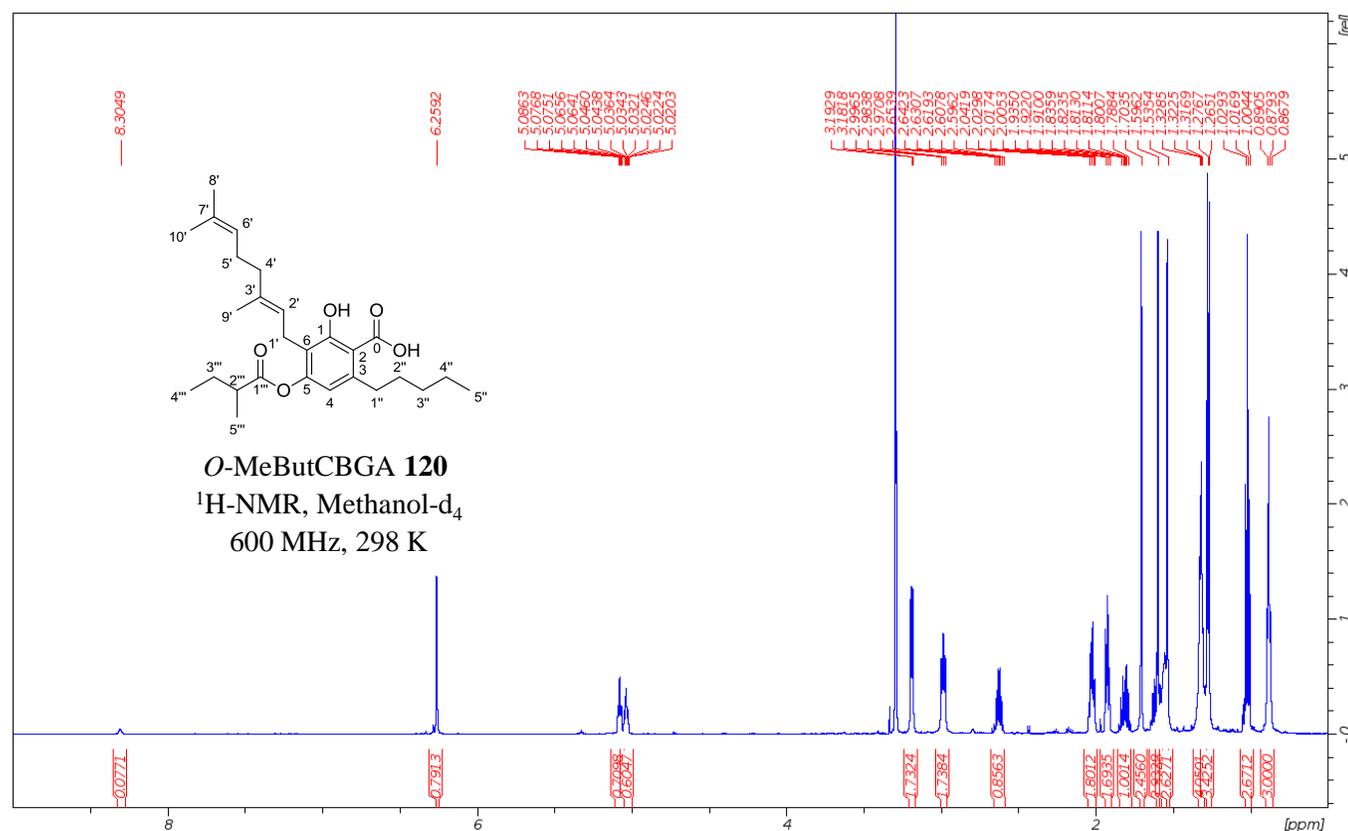
Supplementary NMR Data S13. *O*-MeButCBGA 120.

Data of *O*-MeButCBGA 120: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 6.26 (s, 1H), 5.08 (td, $J = 6.7, 0.94$ Hz, 1H), 5.04 (tt, $J = 7.06, 1.23$ Hz, 1H), 3.19 (d, $J = 6.6$ Hz, 2H), 2.98 (t, $J = 8.1$ Hz, 2H), 2.62 (sextet, $J = 7.0$ Hz, 1H), 2.02 (m, 2H), 1.92 (t, $J = 7.5$ Hz, 2H), 1.81 (m, 1H), 1.70 (s, 3H), 1.61 (m, 1H), 1.60 (s, 3H), 1.55 (m, 2H), 1.53 (s, 3H), 1.32 (m, 4H), 1.27 (d, $J = 7.0$ Hz, 3H), 1.01 (t, $J = 7.4$ Hz, 3H), 0.88 (t, $J = 6.8$ Hz, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 178.4, 164.9, 154.4, 148.0, 137.6, 133.9, 127.4, 125.7, 122.0, 117.7, 117.1, 44.5, 42.7, 38.6, 35.1, 34.6, 29.7, 29.6, 27.4, 25.7, 25.5, 19.5, 19.0, 18.5, 16.3, 14.1.

HRMS (ESI) $[\text{M-H}]^-$ $\text{C}_{27}\text{H}_{39}\text{O}_5$: 443.2797, found 443.2809.

$^1\text{H-NMR}$

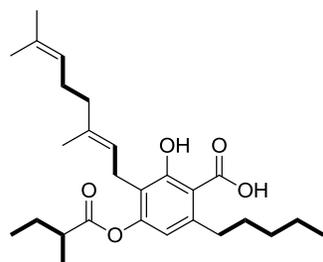
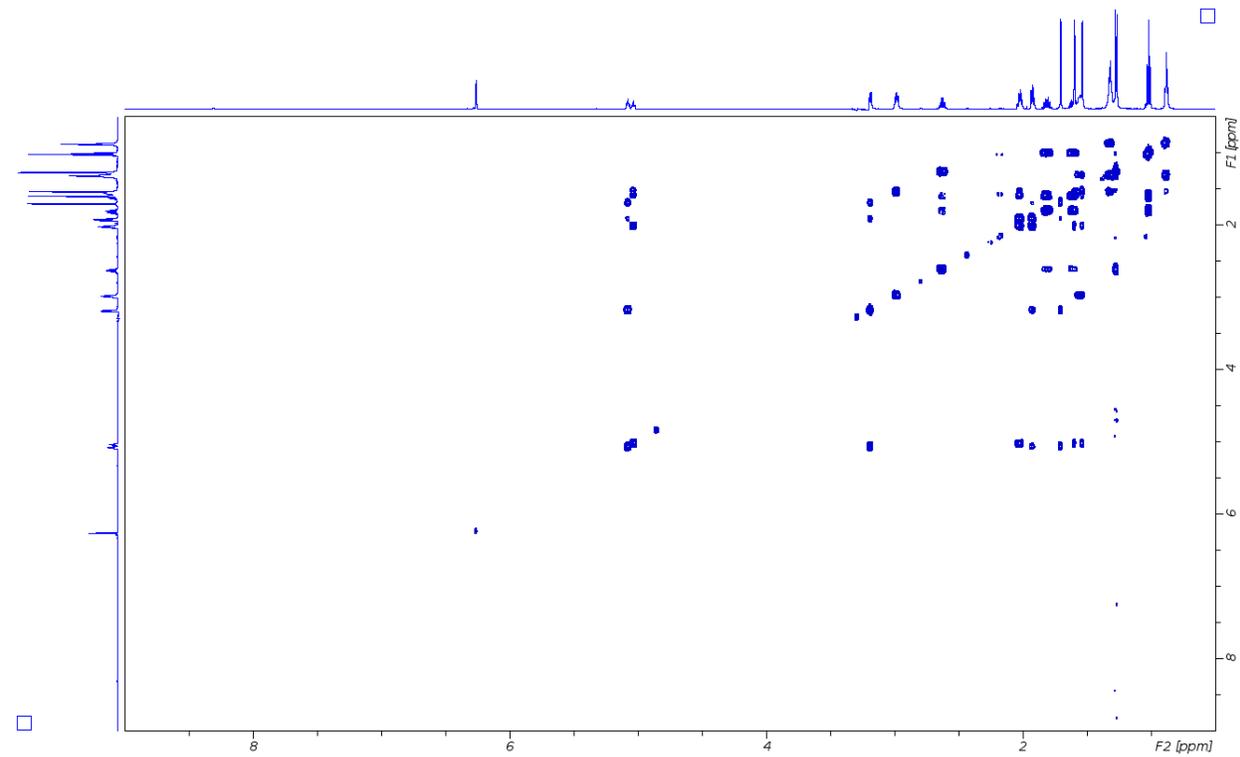


IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
0	nd ^a	
1	164.9	
2	117.1	
3	148.0	
4	117.8	6.26 (s)
5	154.4 ^b	
6	122.0	
1'	25.7	3.19 (d)
2'	125.7	5.08 (td)
3'	137.6	
4'	42.7	1.92 (t)
5'	29.6	2.02 (m)
6'	127.4	5.04 (tt)
7'	133.9	
8'	19.5	1.53 (s)
9'	18.5	1.70 (s)
10'	27.4	1.60 (s)
1''	38.6	2.98 (t)
2''	34.6	1.55 (m)
3''	35.1	1.32 (m)
4''	25.5	1.32 (m)
5''	16.3	0.88 (t)
1'''	178.4	
2'''	44.5	2.62 (m)
3'''	29.7	1.61 (m)
3'''	29.7	1.81 (m)
4'''	14.1	1.01 (t)
5'''	19.0	1.27 (d)

^aThe ^{13}C carbonyl signal was not observed in the NMR spectra, however LC-MS/MS spectra and chemical formula confirm the presence of this group.

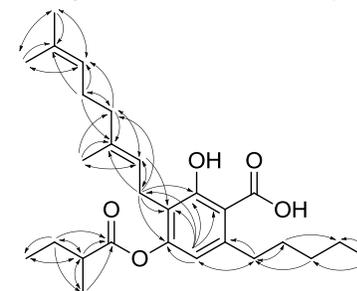
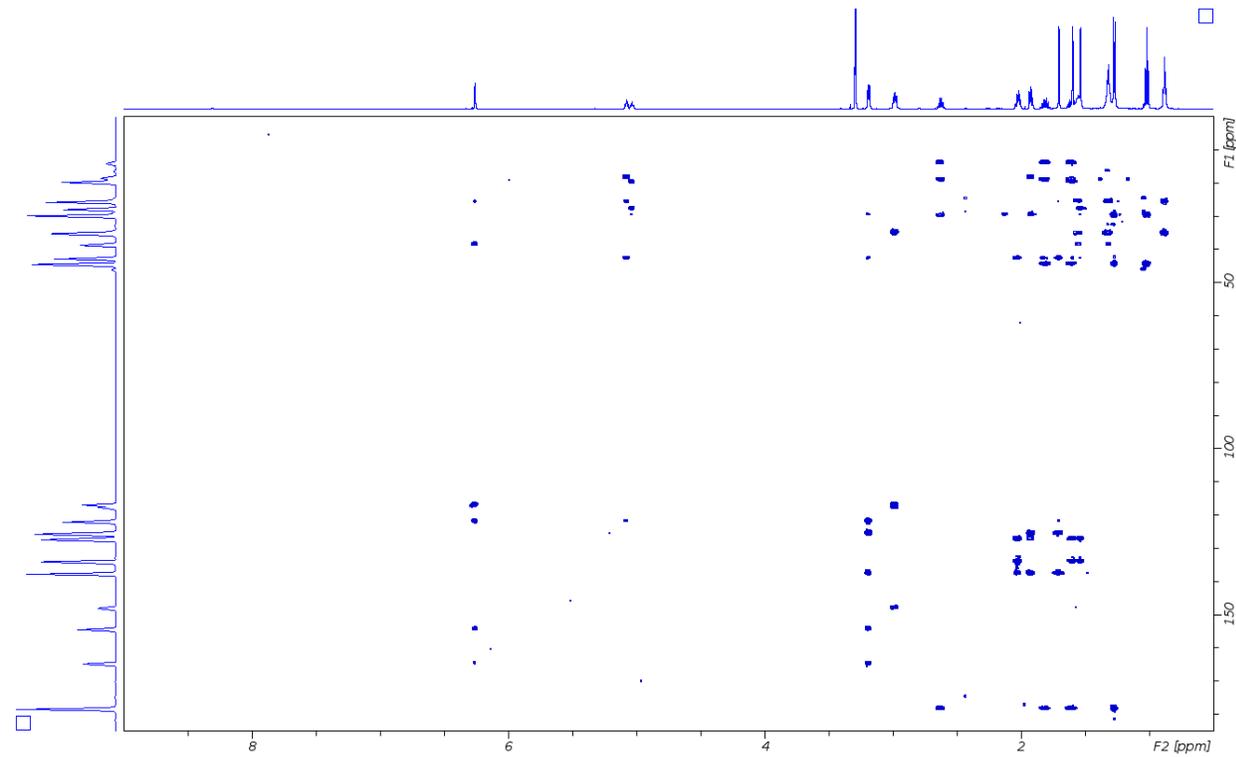
^bThe esterified phenolic carbon experienced the expected ^{13}C NMR upfield shift [$\delta^{13}\text{C}$ C5 154.4 ppm versus 166.5 ppm in CBGA 1 (Supplementary NMR Data 1)].

^1H - ^1H COSY



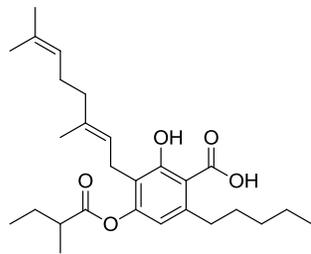
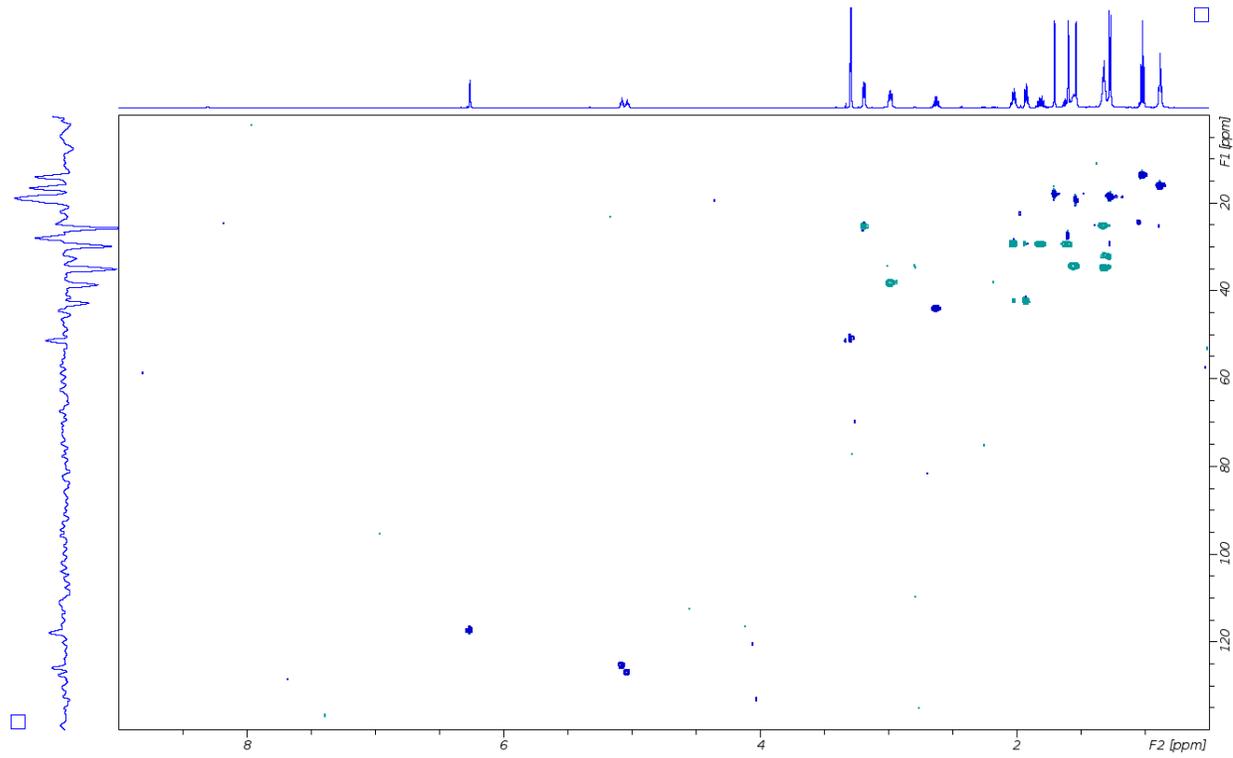
***O*-MeButCBGA 120**
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC



***O*-MeButCBGA 120**
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC



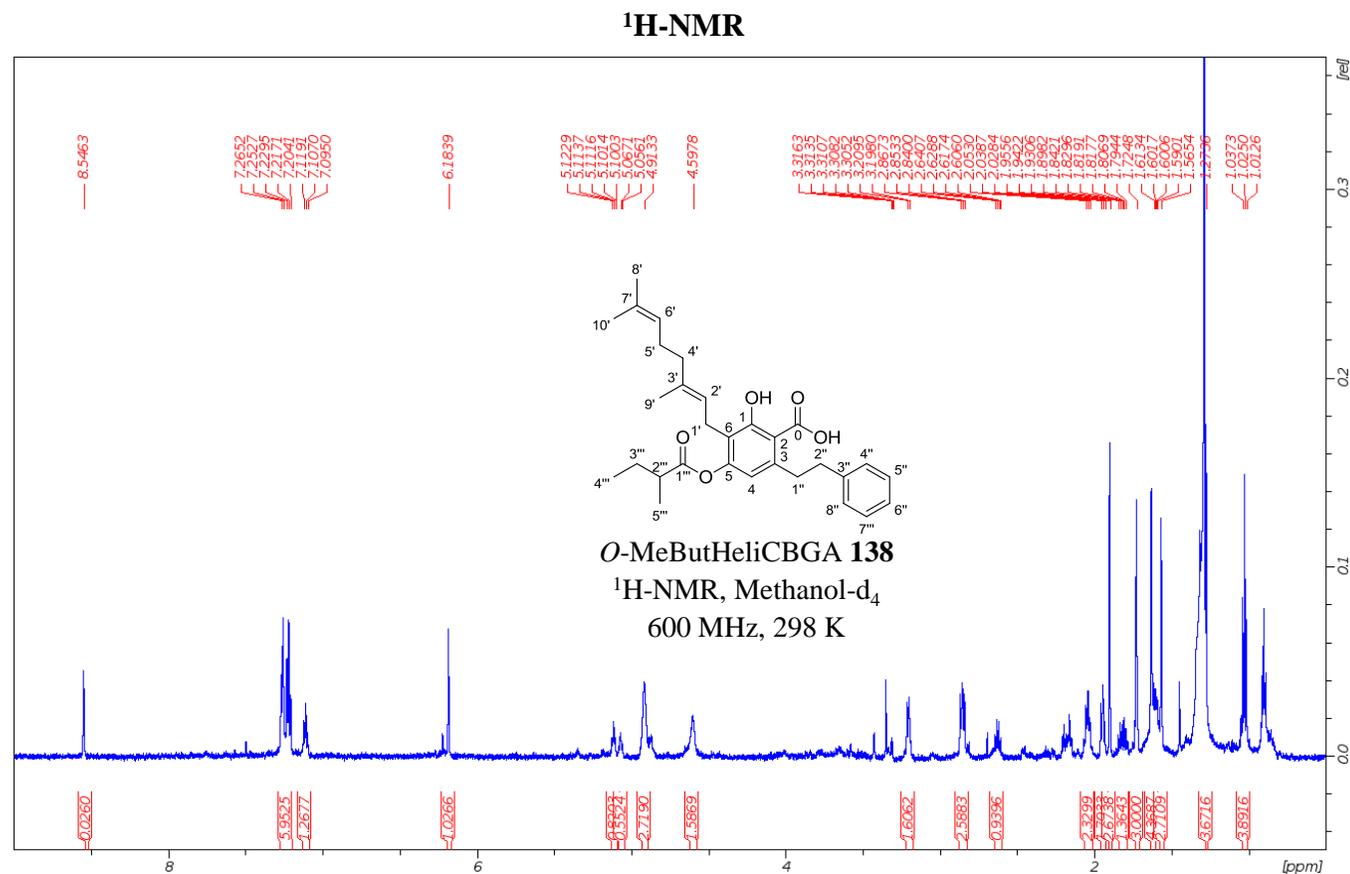
O-MeButCBGA 120
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

Supplementary NMR Data 14. *O*-MeButHeliCBGA 138.

Data of *O*-MeButHeliCBGA **138**: $^1\text{H-NMR}$ (600 MHz, Methanol- d_4) δ 8.56 (br), 7.26 (m, 2H), 7.22 (m, 2H), 7.11 (t, $J = 7.3$ Hz, 1H), 6.18 (s, 1H), 5.12 (t, $J = 6.45$ Hz, 1H), 5.07 (t, $J = 6.68$ Hz, 1H*), 4.91 (br), 4.59 (br), 3.30 (m, 2H*), 3.20 (d, $J = 7.1$ Hz, 2H*), 2.85 (m, 2H), 2.62 (q, $J = 7.0$ Hz, 1H), 2.05 (m, 2H), 1.95 (t, $J = 7.8$ Hz, 2H), 1.83 (m, 1H), 1.73 (s, 3H), 1.61 (s, 3H), 1.60 (m, 1H), 1.57 (s, 3H), 1.27 (d, $J = 7.1$ Hz, 3H), 1.03 (t, $J = 7.4$ Hz, 3H).

$^{13}\text{C-NMR}$ (150.9 MHz, Methanol- d_4) δ 178.9, 164.7, 153.7, 146.3, 146.2, 137.5, 134.0, 131.6, 131.0, 128.4, 127.2, 125.8, 121.9, 119.0, 117.1, 44.4, 42.7, 41.3, 41.1, 29.7, 29.7, 29.5, 27.8, 25.7, 19.5, 18.9, 18.3, 13.9.

HRMS (ESI) $[\text{M-H}]^- \text{C}_{30}\text{H}_{37}\text{O}_5$: 447.2641, found 447.2653.

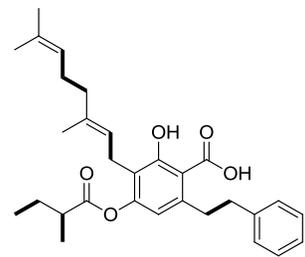
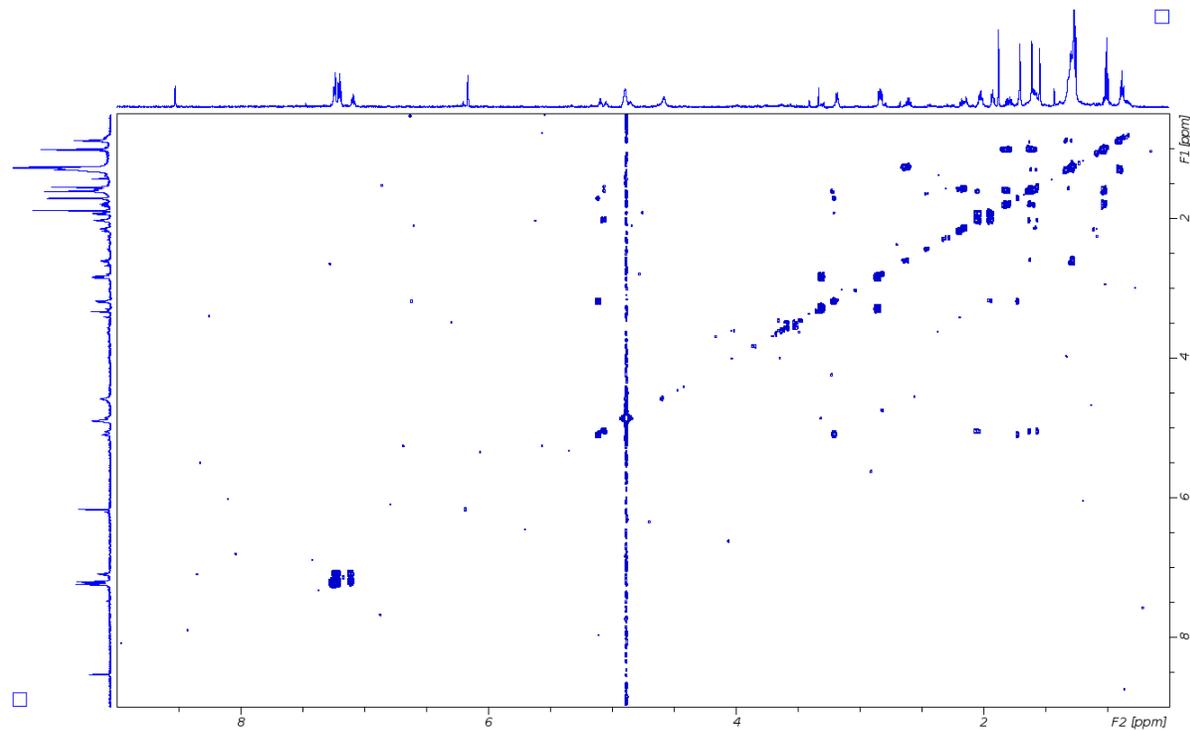


IUPAC	$\nu(^{13}\text{C})$ [ppm]	$\nu(^1\text{H})$ [ppm]
0	nd ^a	
1	164.7	
2	119.0	
3	146.3	
4	117.1	6.18 (s)
5	153.7 ^b	
6	121.9	
1'	25.7	3.20 (d)
2'	125.8	5.12 (t)
3'	137.5	
4'	42.7	1.95 (t)
5'	29.7	2.05 (m)
6'	127.2	5.07 (t)
7'	134.0	
8'	19.5	1.57 (s)
9'	18.3	1.73 (s)
10'	27.8	1.61 (s)
1''	41.1	3.30 (m)
2''	41.3	2.85 (m)
3''	146.2	
4'',8''	131.0	7.22 (m)
5'',7''	131.6	7.26 (m)
6''	128.4	7.11 (t)
1'''	178.9	
2'''	44.4	2.62 (q)
3'''	29.7	1.83 (m)
3'''	29.7	1.60 (m)
4'''	13.9	1.03 (t)
5'''	18.9	1.27 (d)

^aThe ^{13}C carbonyl signal was not observed in the NMR spectra, however LC/HRMSMS spectra and chemical formula confirm the presence of this group.

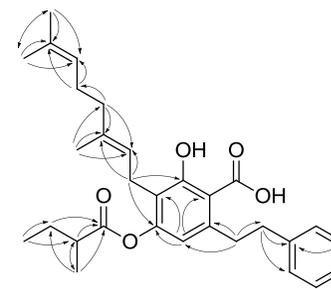
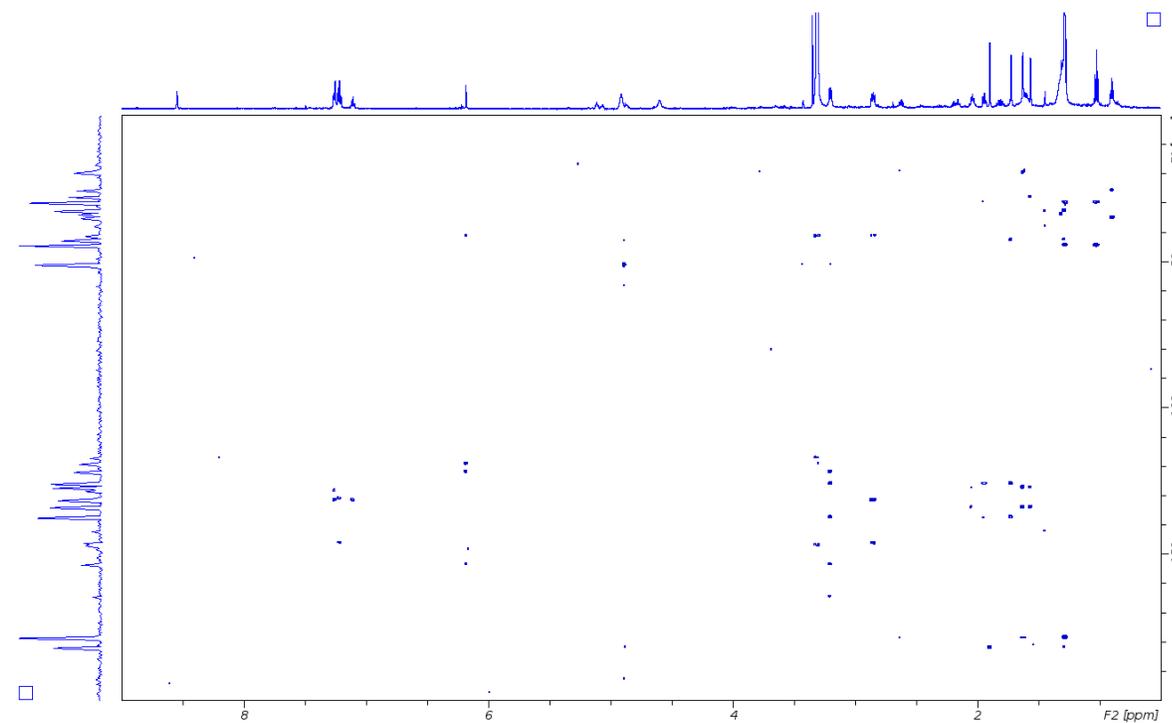
^bThe esterified phenolic carbon experienced the expected ^{13}C NMR upfield shift [$\delta^{13}\text{C}$ C5 153.65 ppm versus 162.49 ppm in heliCBGA **2** (Supplementary NMR Data 2)].

^1H - ^1H COSY



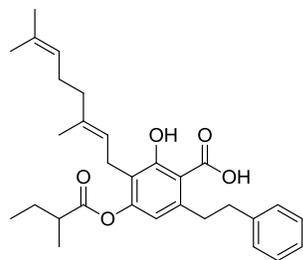
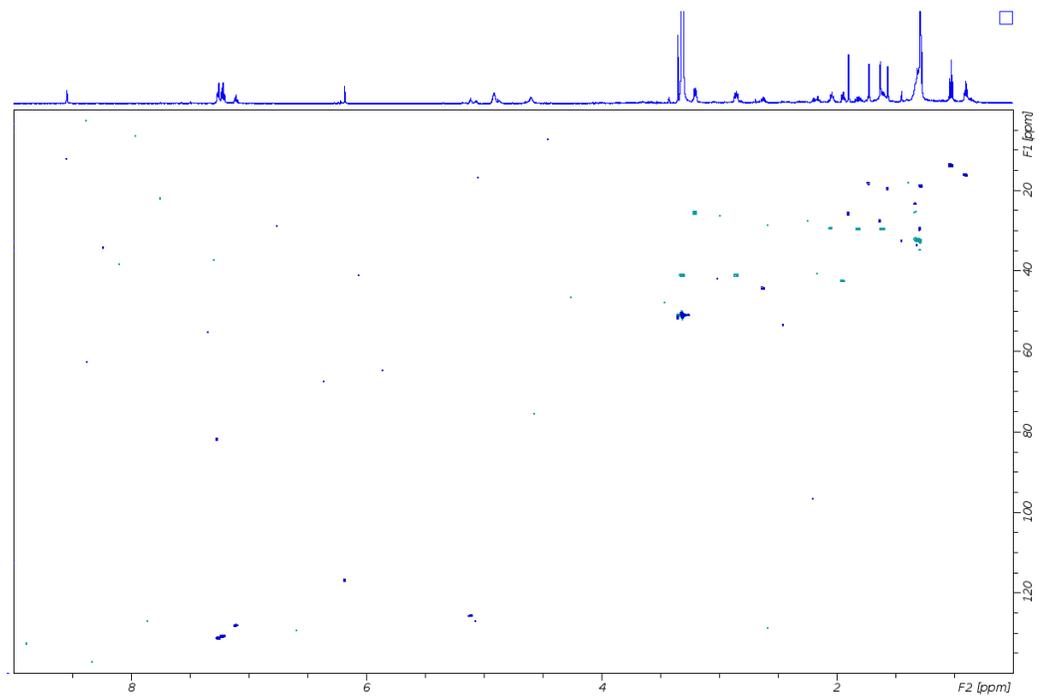
O*-MeButHeliCBGA **138*
 ^1H - ^1H COSY (bold line),
Methanol- d_4 , 298 K

^1H - ^{13}C HMBC

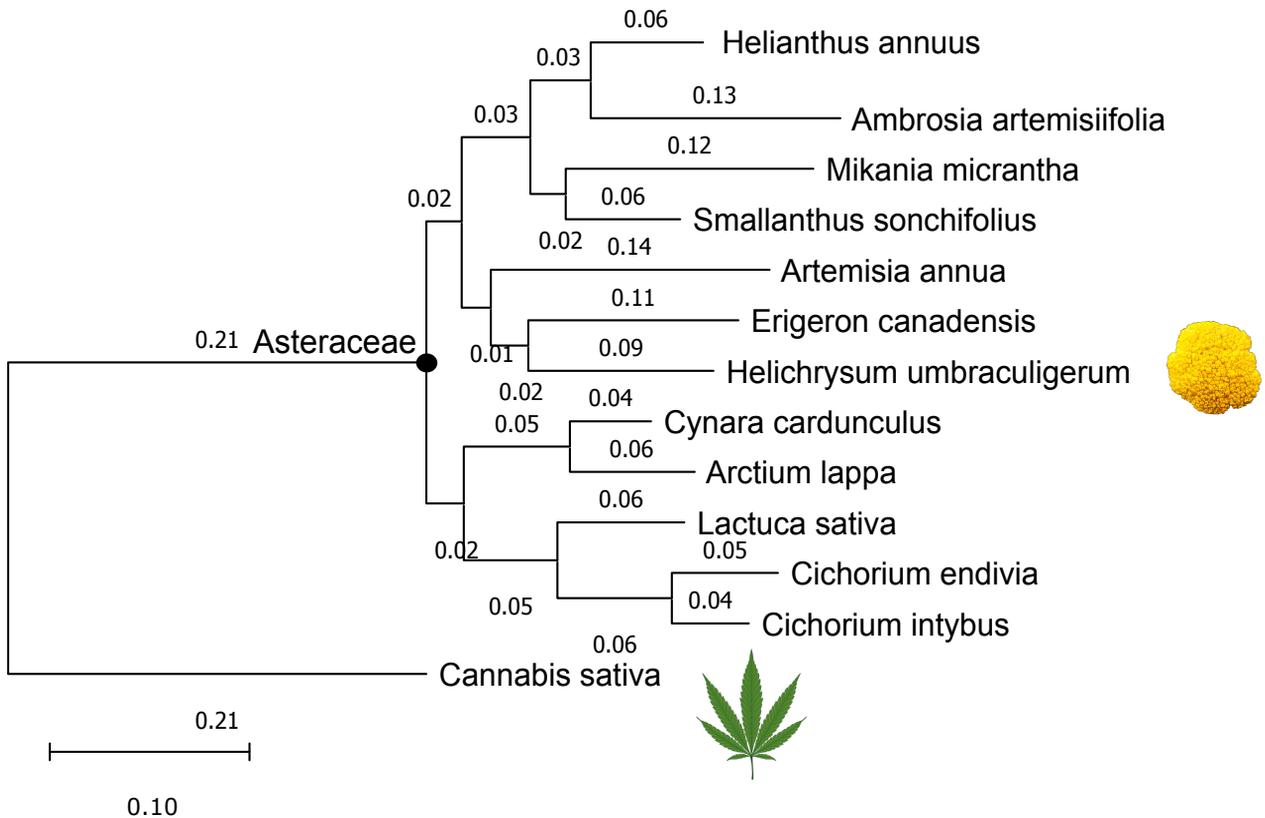


O*-MeButHeliCBGA **138*
 ^1H - ^{13}C HMBC (H to C),
Methanol- d_4 , 298 K

^1H - ^{13}C HSQC

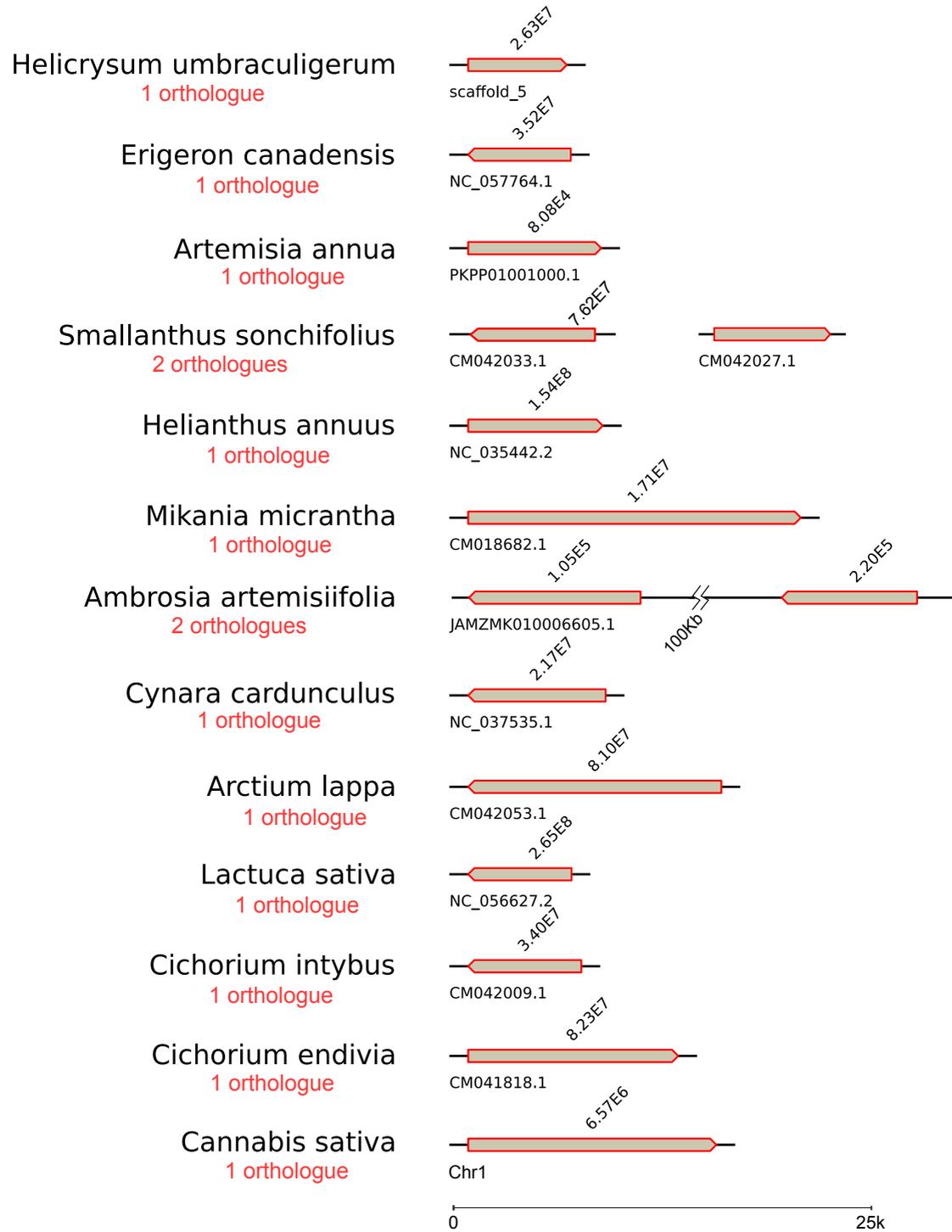


O-MeButheliCBGA 138
 ^1H - ^{13}C HSQC,
Methanol- d_4 , 298 K

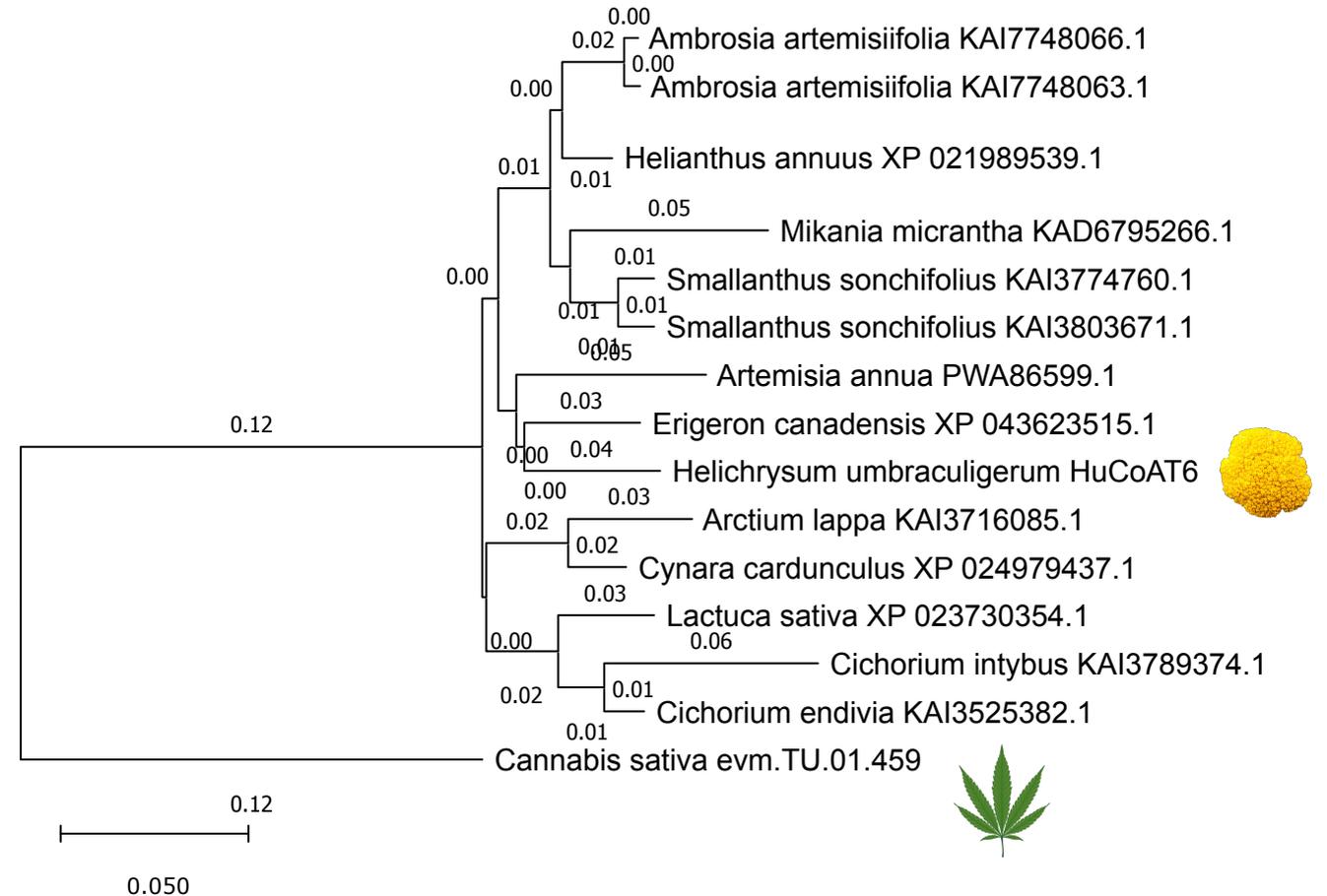


Supplementary Orthology Data 1. Rooted species tree inferred with Orthofinder. The proteomes corresponding to the genomes used for the analysis are GCA_003112345.1 (*Artemisia annua*), GCA_009363875.1 (*Mikania micrantha*), GCA_023376185.1 (*Cichorium endivia*), GCA_023525715.1 (*Cichorium intybus*), GCA_023525745.1 (*Arctium lappa*), GCA_023525975.1 (*Smallanthus sonchifolius*), GCA_024762085.1 (*Ambrosia artemisiifolia*), GCF_001531365.2 (*Cynara cardunculus* var. *scolymus*), GCF_002127325.2 (*Helianthus annuus*), GCF_002870075.4 (*Lactuca sativa*), GCF_010389155.1 (*Erigeron canadensis*) and *Cannabis sativa* GCA_900626175.1. The distance is indicated in each branch.

a

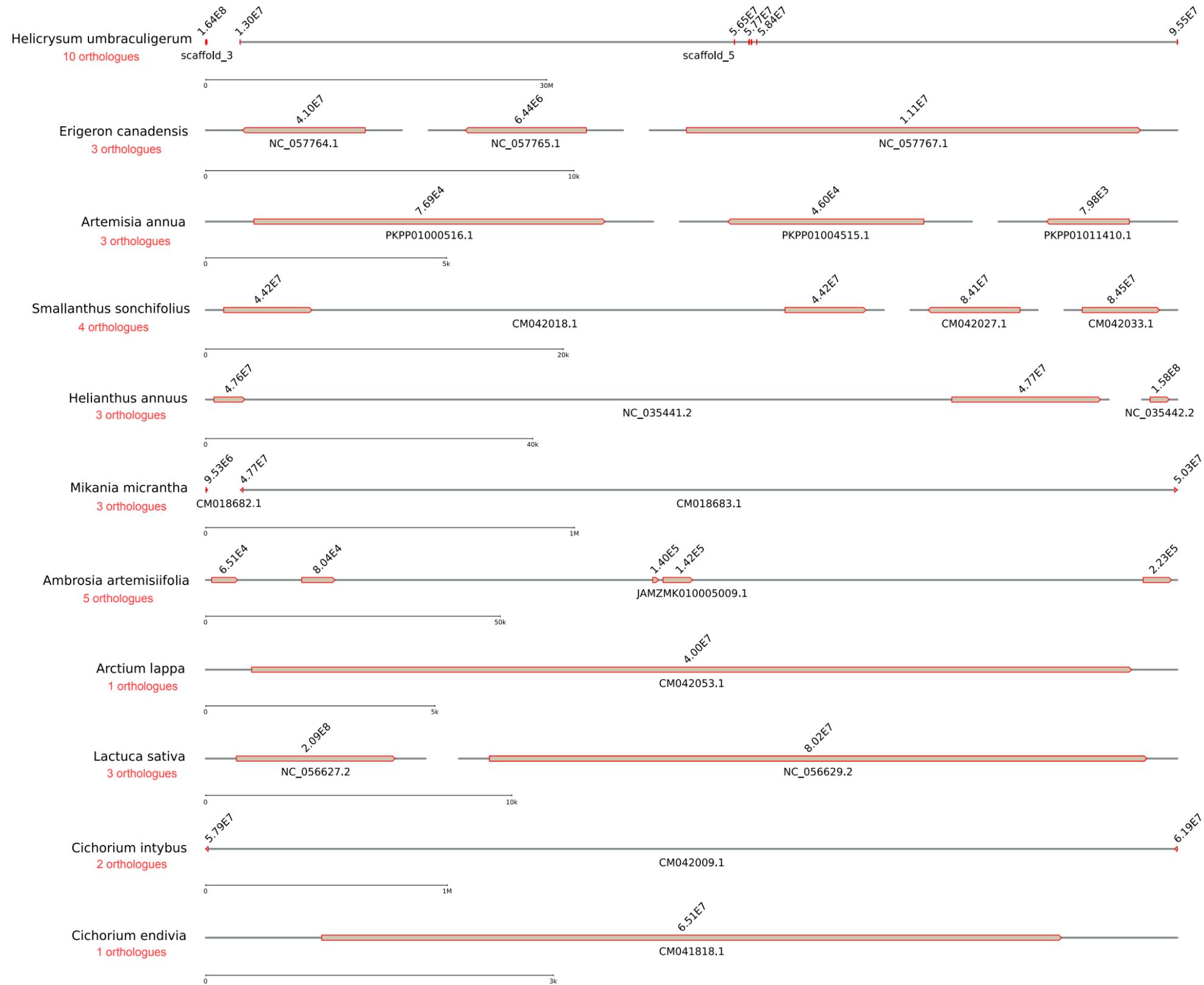


b

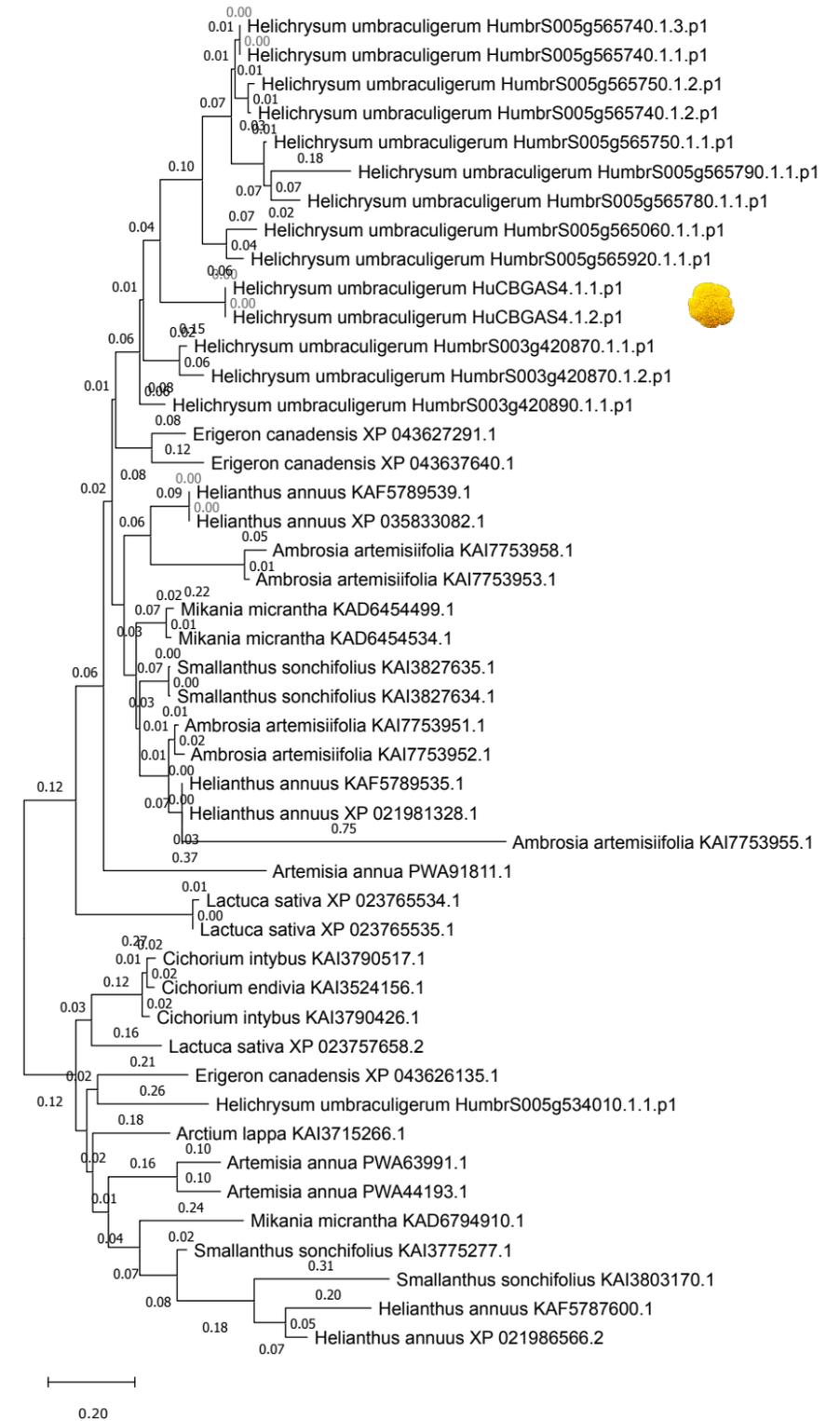


Supplementary Orthology Data 2. Orthogroup OG0014461 to which HuCoAT6 belongs. **a.** Physical position of each gene within each chromosome/scaffold in the analyzed species. The chromosomal position of each gene was taken from the corresponding GFF file. The numeric position of each gene within each scaffold/chromosome is indicated on top. **b.** Rooted gene tree of the orthogroup. The genetic distance of each branch is shown. A complete list of the genes, including the species, chromosomal positions, protein IDs and the annotated products, is described in Supplementary Table 16.

a



b



Supplementary Orthology Data 4. Orthogroup OG0002538 to which HuCBGAS4 belongs. **a.** Physical position of each gene within each chromosome/scaffold in the analyzed species. The chromosomal position of each gene was taken from the corresponding GFF file. The numeric position of each gene within each scaffold/chromosome is indicated on top. Orthologous for *Cannabis sativa* and *Cynara cardunculus* were not identified with this method. **b.** Rooted gene tree of the orthogroup. The genetic distance of each branch is shown. A complete list of the genes, including the species, chromosomal positions, protein IDs and the annotated products, is described in Supplementary Table 18.

References

1. di Marzo, V. & Piscitelli, F. The endocannabinoid system and its modulation by phytocannabinoids. *Neurotherapeutics* **12**, 692–698 (2015).
2. Ligresti, A., de Petrocellis, L. & di Marzo, V. From Phytocannabinoids to Cannabinoid Receptors and Endocannabinoids: Pleiotropic Physiological and Pathological Roles Through Complex Pharmacology. *Physiol Rev* **96**, 1593–1659 (2016).
3. Turner, S. E., Williams, C. M., Iversen, L. & Whalley, B. J. Molecular Pharmacology of Phytocannabinoids. *Prog Chem Org Nat Prod* **103**, 61–101 (2017).
4. Morales, P., Hurst, D. P. & Reggio, P. H. Molecular Targets of the Phytocannabinoids: A Complex Picture. *Prog Chem Org Nat Prod* **103**, 103–131 (2017).
5. Stout, J. M., Boubakir, Z., Ambrose, S. J., Purves, R. W. & Page, J. E. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. *Plant Journal* **71**, 353–365 (2012).
6. Taura, F. *et al.* Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Lett* **583**, 2061–2066 (2009).
7. Gagne, S. J. *et al.* Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. *Proc Natl Acad Sci U S A* **109**, 12811–12816 (2012).
8. Luo, X. *et al.* Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123–126 (2019).
9. Hanuš, L. O., Meyer, S. M., Muñoz, E., Tagliatalata-Scafati, O. & Appendino, G. Phytocannabinoids: a unified critical inventory. *Nat Prod Rep* **33**, 1357–1392 (2016).
10. Gülck, T. & Møller, B. L. Phytocannabinoids: Origins and Biosynthesis. *Trends Plant Sci* **25**, 985–1004 (2020).
11. Appendino, G., Tagliatalata-Scafati, O. & Muñoz, E. Cannabidiol (CBD) From Non-Cannabis Plants: Myth or Reality? *Natural Product Communications* vol. 17 Preprint at <https://doi.org/10.1177/1934578X221098843> (2022).
12. Bohlmann, F. & Hoffmann, E. Cannabigerol-ähnliche Verbindungen aus *Helichrysum umbraculigerum*. *Phytochemistry* **18**, 1371–1374 (1979).
13. Pollastro, F. *et al.* Amorfrutin-type phytocannabinoids from *Helichrysum umbraculigerum*. *Fitoterapia* **123**, 13–17 (2017).
14. Procaccia, S. *et al.* Cannabis for Medical Use: Versatile Plant Rather Than a Single Drug. *Frontiers in Pharmacology* vol. 13 Preprint at <https://doi.org/10.3389/fphar.2022.894960> (2022).
15. Berman, P. *et al.* A new ESI-LC/MS approach for comprehensive metabolic profiling of phytocannabinoids in *Cannabis*. *Sci Rep* **8**, 1–15 (2018).
16. de Meijer, E. P. M. & Hammond, K. M. The inheritance of chemical phenotype in *Cannabis sativa* L. (V): regulation of the propyl-/pentyl cannabinoid ratio, completion of a genetic model. *Euphytica* **210**, 291–307 (2016).
17. Kearsley, L. J. *et al.* Structure of the *Cannabis sativa* olivetol- producing enzyme reveals cyclization plasticity in type III polyketide synthases. *FEBS J* **287**, 1511–1524 (2020).
18. Yamaguchi, T. *et al.* Cross-reaction of chalcone synthase and stilbene synthase overexpressed in *Escherichia coli*. *FEBS Lett* **460**, 457–461 (1999).
19. Parvez, A., Giri, S., Bisht, R. & Saxena, P. New Insights on Cyclization Specificity of Fungal Type III Polyketide Synthase, PKSIII_{Nc} in *Neurospora crassa*. *Indian J Microbiol* **58**, 268–277 (2018).
20. Taura, F. *et al.* A Novel Class of Plant Type III Polyketide Synthase Involved in Orsellinic Acid Biosynthesis from *Rhododendron dauricum*. *Front Plant Sci* **7**, 1452 (2016).
21. Okada, Y. & Ito, K. Cloning and analysis of valerophenone synthase gene expressed specifically in lupulin gland of hop (*Humulus lupulus* L.). *Biosci, Biotechnol, Biochem* **65**, 150–155 (2001).
22. de Bruijn, W. J. C., Levisson, M., Beekwilder, J., van Berkel, W. J. H. & Vincken, J. P. Plant Aromatic Prenyltransferases: Tools for Microbial Cell Factories. *Trends Biotechnol* **38**, 917–934 (2020).
23. Lim, E. K. *et al.* The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J Biol Chem* **277**, 586–592 (2002).
24. Geissler, M., Volk, J., Stehle, F., Kayser, O. & Warzecha, H. Subcellular localization defines modification and production of Δ^9 -tetrahydrocannabinolic acid synthase in transiently transformed *Nicotiana benthamiana*. *Biotechnol Lett* **40**, 981–987 (2018).
25. Gülck, T. *et al.* Synthetic Biology of Cannabinoids and Cannabinoid Glucosides in *Nicotiana benthamiana* and *Saccharomyces cerevisiae*. *J Nat Prod* **83**, 2877–2893 (2020).
26. Tuominen, L. K., Johnson, V. E. & Tsai, C. J. Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. *BMC Genomics* **12**, 1–17 (2011).

27. Ren, G. *et al.* Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Sci Adv* **7**, (2021).
28. Nachnani, R., Raup-Konsavage, W. M. & Vrana, K. E. The Pharmacological Case for Cannabigerol. *J Pharmacol Exp Ther* **376**, 204–212 (2021).
29. Anokwuru, C. P. *et al.* Cannabigerol: a bibliometric overview and review of research on an important phytocannabinoid. *Phytochemistry Reviews* 1–25 (2022) doi:10.1007/s11101-021-09794-w.
30. Russo, E. B. *et al.* Survey of Patients Employing Cannabigerol-Predominant Cannabis Preparations: Perceived Medical Effects, Adverse Events, and Withdrawal Symptoms. *Cannabis Cannabinoid Res* (2021) doi:10.1089/CAN.2021.0058.
31. Hardman, J. M., Brooke, R. T. & Zipp, B. J. Cannabinoid glycosides: In vitro production of a new class of cannabinoids with improved physicochemical properties. *bioRxiv* 104349 (2017) doi:10.1101/104349.
32. Lourens, A. C. U., Viljoen, A. M. & van Heerden, F. R. South African *Helichrysum* species: A review of the traditional uses, biological activity and phytochemistry. *J Ethnopharmacol* **119**, 630–652 (2008).
33. Jakupovic, J. *et al.* Twenty-one acylphloroglucinol derivatives and further constituents from south african *Helichrysum* species. *Phytochemistry* **28**, 1119–1131 (1989).
34. Jakupovic, J., Kuhnke, J., Schuster, A., Metwally, M. A. & Bohlmann, F. Phloroglucinol derivatives and other constituents from South African *Helichrysum* species. *Phytochemistry* **25**, 1133–1142 (1986).
35. Dong, Y., Feldberg, L. & Aharoni, A. Miso: an R package for multiple isotope labeling assisted metabolomics data analysis. *Bioinformatics* **35**, 3524–3526 (2019).
36. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
37. Ghurye, J. *et al.* Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol* **15**, e1007273 (2019).
38. Arima Genomics. Mapping pipeline for data generated using HiC. https://github.com/ArimaGenomics/mapping_pipeline (2019).
39. Bradnam, K. R. *et al.* Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2**, 10 (2013).
40. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**, 1–18 (2019).
41. Freedman, A. & Weeks, N. Best Practices for De Novo Transcriptome Assembly with Trinity - Harvard FAS Informatics. <https://informatics.fas.harvard.edu/best-practices-for-de-novo-transcriptome-assembly-with-trinity.html>.
42. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, (2019).
43. Hackl, T. & Ankenbrand, M. gggenomes: A Grammar of Graphics for Comparative Genomics. <https://github.com/thackl/gggenomes> (2022).
44. Shockey, J. M., Fulda, M. S. & Browse, J. Arabidopsis Contains a Large Superfamily of Acyl-Activating Enzymes. Phylogenetic and Biochemical Analysis Reveals a New Class of Acyl-Coenzyme A Synthetases. *Plant Physiol* **132**, 1065 (2003).
45. Naake, T., Maeda, H. A., Proost, S., Tohge, T. & Fernie, A. R. Kingdom-wide analysis of the evolution of the plant type III polyketide synthase superfamily. *Plant Physiol* **185**, 857–875 (2021).
46. Shahaf, N. *et al.* The WEIZMASS spectral library for high-confidence metabolite identification. *Nat Commun* **7**, (2016).
47. Balcke, G. U. *et al.* Multi-Omics of Tomato Glandular Trichomes Reveals Distinct Features of Central Carbon Metabolism Supporting High Productivity of Specialized Metabolites. *Plant Cell* **29**, 960–983 (2017).
48. Song, L. & Florea, L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* **4**, 48 (2015).
49. Krueger, F. TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. <https://github.com/FelixKrueger/TrimGalore>.
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
51. Ewels, P., Magnusson, M., Lundin, S. & Källér, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
52. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* **29**, 644 (2011).
53. PacificBiosciences. IsoSeq: IsoSeq3 - Scalable De Novo Isoform Discovery from Single-Molecule PacBio Reads. <https://github.com/PacificBiosciences/IsoSeq>.
54. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

55. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7 (2008).
56. 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 Genom Bioinform* **3**, 1–11 (2021).
57. Haas, B. TransDecoder: Find Coding Regions Within Transcripts. <https://github.com/TransDecoder/TransDecoder/>.
58. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29–W37 (2011).
59. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res* **42**, D222–D230 (2014).
60. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
61. Consortium, U. The universal protein resource (UniProt). *Nucleic Acids Res* **36**, D190–D195 (2008).
62. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **2019 37:4** **37**, 420–423 (2019).
63. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567–580 (2001).
64. Hass B. Trinotate: Transcriptome Functional Annotation and Analysis. <https://github.com/Trinotate/Trinotate.github.io/blob/master/index.asciidoc>.
65. 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).
66. Tzfadia, O. *et al.* The ‘TranSeq’ 3'-end sequencing method for high-throughput transcriptomics and gene space refinement in plant genomes. *Plant Journal* **96**, 223–232 (2018).
67. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
69. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
70. Love, M. I., Anders, S. & Huber, W. *Differential analysis of count data - the DESeq2 package*. *Genome Biology* vol. 15 (2014).
71. Russo, P. S. T. *et al.* CEMiTool: A Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics* **19**, 1–13 (2018).
72. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
73. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
74. Wilcox. gggenes: Draw gene arrow maps in ggplot2. <https://github.com/wilcox/gggenes>.
75. Jozwiak, A. *et al.* Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery. *Nat Chem Biol* **16**, 740–748 (2020).
76. Cai, J. *et al.* Glycosylation of N-hydroxy-pipecolic acid equilibrates between systemic acquired resistance response and plant growth. *Mol Plant* **14**, 440–455 (2021).
77. Doležel, J. & Bartoš, J. Plant DNA flow cytometry and estimation of nuclear genome size. *Ann Bot* **95**, 99–110 (2005).
78. Montone, C. M. *et al.* Improved identification of phytocannabinoids using a dedicated structure-based workflow. *Talanta* **219**, 121310 (2020).