Supporting Information for:

Probing the promiscuity of *ent*-kaurene oxidases via combinatorial biosynthesis

Sibongile Mafu, Meirong Jia, Jiachen Zi, Dana Morrone, Yisheng Wu, Meimei Xu, Matthew L. Hillwig and Reuben J. Peters

Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, Iowa, 50011, USA.

Content

Detailed Materials and Methods

General. Unless otherwise noted, chemicals were purchased from Fisher Scientific and molecular biology reagents from Invitrogen. Gas chromatography (GC) was performed with a Varian 3900 GC with Saturn 2100 ion trap mass spectrometer (MS) in electron ionization (70 eV) mode. Samples (1 μL) were injected in splitless mode with the injection port at 250 °C and oven at 50 °C, after holding for 3 min at 50 °C, the oven temperature was raised at a rate of 15 °C/min to 300 °C, where it was held for an additional 3 min. MS data from 90 to 600 mass-to-charge ratio (*m/z*) were collected starting 14 min after injection until the end of the run.

Screening. Initial screening was carried out using pCDF-Duet1/DEST::KO/CPR constructs (with either AtCPR1 (1) and AtKO or OsCPR1 (2) and OsKO2). These were co-transformed with a pGG*x*C vector that leads to production of one of the three common stereoisomers of CPP (pGGnC to normal CPP, pGG*e*C to *ent*-CPP, and pGG*s*C to *syn*-CPP), along with a subsequently acting class I diterpene synthase, carried in a pET-based pDEST vector, for production of the potential LRD substrate, much as previously reported (3)(see Supplemental Table S1 for the relevant combinations). Note that, while the previously identified production of *syn*-pimara-9(11),15-diene from *syn*-CPP by TaKSL4 from wheat exhibits poor yield (4), fortuitously, it was found here that a previously identified class I diterpene synthase from the bacterium *Salinaspora arenicola* (SaDTS)(5) efficiently reacts with *syn*-CPP to produce a 4:6 mixture of *syn*-pimara-9(11),15-diene and *syn*-stemodene (Figure S11), and this was used here. In addition, while the native rice *syn*-stemarene/*endo*-stemodene synthase OsKSL8 also exhibits poor yield (6), leading to no observable turnover with AtKO, employment of a codon-optimized version was found here to provide sufficient yield for observation of CYP activity, although the resulting mixture of products was complex enough to preclude straightforward isolation of sufficient amounts of any

individual compound for NMR structural analysis. The final results reported here in Figure 3 were obtained using pET-Duet1/CPR/KO constructs (same pairings as above, but both inserted via a "cut-and-paste" sub-cloning strategy using NcoI/NotI and NdeI/XhoI restriction sites, respectively), the utility of which for recombinant expression in *E. coli* has been recently reported (7). For this purpose, the relevant class I diterpene synthases were sub-cloned by directional recombination from pENTR vectors into pCDF-Duet1/DEST/AtCPR1 for co-expression. Although no qualitative changes were observed using the pET based KO expression constructs, there were significant increases in flux to KO product, examples of which are depicted in Figure S6 & S7.

For screening, recombinant bacteria were grown in 20 mL of TB medium (with the appropriate antibiotics) in 125 mL Erlenmeyer flasks to mid-log phase (OD_{600} ~0.8) at 37 °C, then transferred to 16 °C for 1 h prior to induction with 0.5 mM isopropylthiogalactoside (IPTG) and the addition of aminolevulinic acid (75 mg/L) and riboflavin (2 mg/L). Thereafter, the cultures were fermented for an additional 72 hours at 16 °C, and then extracted with an equal volume of hexanes. The organic extract was separated and dried under a stream of N_2, derivatized with diazomethane to examine potential acid formation, and re-dissolved in hexane (100 μL) for analysis by GC-MS.

Diterpenoid production. Novel enzymatic products were obtained in sufficient amounts for structural determination by NMR by both increasing flux into isoprenoid metabolism and scaling up cultures, which were grown in parallel batches, each consisting of 500 mL TB media (with the appropriate antibiotics) in 2.8 L Fernbach flasks, largely as described above. However, flux toward isoprenoid biosynthesis was increased by incorporation of the previously described

pMBI, which encodes for the bottom half of the mevalonate-dependent isoprenoid precursor pathway from yeast *Saccharomyces cerevisiae* (8). This enables production of the isoprenoid precursors isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate, such that feeding the cultures 10 mM mevalonolactone (upon induction) significantly increases diterpenoid production, as previously described (9). The resulting diterpenoids were extracted from 3-4 L of culture (media and cells) with an equal volume of hexanes, and the organic extract obtained using a separatory funnel, and then dried by rotary evaporation. The resulting residue was resuspended in 5 mL hexanes and fractionated over a 4 g-silica column by flash chromatography, using a Reveleris system (Grace, Deerfield, IL) with UV detection, and a hexane:acetone step gradient as the mobile phase. The hydroxylated diterpenoids eluted at either 10% or 20% acetone, as determined by GC-MS analysis. Fractions of interest were dried under N_2. The resulting residue was dissolved in 1 mL acetonitrile and the hydroxylated diterpenoids purified by HPLC, using an Agilent 1200 series instrument equipped with autosampler, fraction collector, and diode array UV detection, over a ZORBAX Eclipse XDB-C8 column (4.6 × 150 mm, 5 mm) at a 0.5 mL/min flow rate. The column was pre-equilibrated with 50% acetonitrile/distilled water, sample loaded, then the column washed with 50% acetonitrile/distilled water (0–2 min), and eluted with a gradient of 50% to 100% acetonitrile (2–7 min), followed by a 100% acetonitrile wash (7–30 min). The fractions containing the diterpenoid of interest were identified by GC-MS analysis. Each purified compound was dried under a gentle stream of N_2, and then dissolved in 0.5 mL deuterated chloroform ($CDCl_3$; Sigma-Aldrich), with this evaporation-resuspension process repeated twice more to completely remove the protonated acetonitrile solvent.

NMR analysis. Spectra for the diterpenoids were recorded at 25 °C on a Bruker Avance 700 spectrometer equipped with a 5-mm HCN cryogenic probe for ^{1}H and ^{13}C. All samples were placed in NMR microtubes (Shigemi) and purged with nitrogen gas for analyses, and chemical shifts were referenced using known chloroform-d [^{13}C 77.23, ^{1}H 7.24 p.p.m] signals offset from tetramethylsilane. Structural analysis was performed using 1D ^{1}H, and 2D DQF-COSY (double-quantum-filtered correlation spectroscopy), HSQC (heteronyclear single-quantum coherence), HMBC (heteronuclear multiple bond correlation) and NOESY experiment spectra acquired at 700 MHz, and 1D ^{13}C spectra (174 MHz) using standard experiments from the Bruker TopSpin v1.4 software. Correlations from the HMBC spectra were used to propose a partial structure, while DQF-COSY correlations between protonated carbons and HSQC spectra were used to complete the partial structure and assign proton chemical shifts. Correlations from NOESY (nuclear overhauser effect) dipole-dipole signals observed were used to assign stereochemistry. The acquired 1D ^{1}H and ^{13}C spectra can be found in the Supplemental Information (Figures S24 – S35).

Feeding studies. To verify that the observed carboxylic acid products from reaction of AtKO with *ent*-cassadiene and *ent*-sandaracopimaradiene correspond to *ent*-cassadien-19-oic and *ent*-sandaracopimaradien-19-oic acids, feeding studies with the structurally characterized *ent*-cassadien-19-ol and *ent*-sandaracopimaradien-19-ol precursors were carried out. For this purpose, *ent*-cassadien-19-ol or *ent*-sandaracopimaradien-19-ol was fed to cultures co-expressing AtKO and AtCPR1, using the pET-Duet1/AtCPR1/AtKO construct described above, in the C43 Overexpress strain of *E. coli*. Recombinant cultures (10 mL) were grown as described above, and hydroxylated diterpenoid fed upon induction (to a final concentration of 25 μM). Cultures were

extracted with an equal volume of hexanes after 3 days of incubation, and the production of *ent*-cassadien-19-oic and *ent*-sandaracopimaradien-19-oic acids verified by GC-MS analysis as described above.

Relative Kinetics. Substrates were obtained by extraction from *E. coli* metabolically engineered for their production [i.e., using the relevant combination of pGG*x*C and pDEST listed in Table S1, along with pIRS, which overexpresses a number of genes from the endogenous isoprenoid precursor pathway (9)], with purification as described above (in most cases flash chromatography was sufficient to obtain > 90% pure compound for this purpose), and the resulting compounds quantified by GC analysis with flame ionization detection. AtKO was co-expressed with AtCPR1, using the pCDF-Duet1/DEST::AtKO/AtCPR1 construct described above, in the C43 Overexpress strain of *E. coli*. Recombinant cultures were grown as described above. The resulting cells were harvested via centrifugation at 5000 g for 15 min, and were then resuspended in one-tenth of the culture volume of buffer (100 mM, Tris-HCl, pH 7.8, 20 % glycerol and 0.5 mM EDTA). Lysis was carried out using a French press homogenizer (Emulsiflex-C5; Avestin), with the cell suspension passed through twice at 15000 psi (1 psi = 6.9 kPa). The resulting lysates were clarified via centrifugation (20 min at 15000 *g*). The supernatant (cell-free lysate) was used in the subsequent studies. Assays also included a previously described NADPH regeneration system (10 mM glucose-6-phosphate, 1 mM $NADP^+$, and yeast glucose-6-phosphate dehydrogenase)(1). The diterpene substrate was added from a 100x stock dissolved in 50:50 (v/v) methanol and DMSO. Substrates (*ent*-cassadiene, *ent*-sandaracopimaradiene, *syn*-aphidicolene and pimara-7,15-diene) were reacted over a range of 0 to 160 μM, with assays times ranging from 15-30 min (as determined by initial optimization). A control reaction with

ent-kaurene was carried out concurrently, utilizing 40 μM of substrate with a 3 min reaction time, in parallel with each set of kinetic assays, enabling normalization of V_{max} (i.e., to the rate observed with *ent*-kaurene, as presented in Table 1).

Molecular phylogenetic analyses. The nucleotide sequences coding for the characterized CYP701 family members (see Supplemental Table S14) were aligned by codons using the MUSCLE algorithm, as implemented in the MEGA 6.06 software package for Mac (10). Specifically, with the default settings – i.e., a gap opening penalty of -2.9, gap extension penalty of 0, hydrophobicity multiplier of 1.2, and UPGMB clustering with a minimum diagonal length (lambda) of 24. This alignment was used to calculate a phylogenetic tree via the Maximum Likelihood method, with the reliability tested by bootstrapping with 100 replicates, and using the calculated optimal parameters – the Kimura 2-parameter model (K2) with discrete Gamma distribution to model evolutionary rate differences between sites (+G), allowing some sites to be evolutionarily invariable (+I). All possible codon positions were included, although any position with less than 95% coverage was excluded, leaving a total 1443 sites in the final dataset. Evolutionary pressure was probed by codon-based Z-tests of selection using the Nei-Gojobori method with variance calculated by bootstrapping with 100 replicates, revealing strong purifying selection, with an overall $d_S - d_N = 29.38$ and $p < 10^{-10}$. Pairwise comparisons similarly supported only purifying selection. Lineage specific adaptive evolution was investigated using the PAML 4.8 software package for Mac (11), particularly the CODEML program, employed largely as outlined (12). The fit of various models was assessed by comparison of the calculated log likelihood scores (l), with statistical significance estimated by likelihood ratio tests ($2\Delta l$) under the assumption of χ^2 distribution, where appropriate (i.e., between models with different

degrees of freedom). All model fits were run using the sequence alignment and phylogenetic tree described above, with empirical estimation of the codon frequency (i.e., CodonFreq = 3) and transition/transversion rate ratio (i.e., fix_kappa = 0), as well as d_N/d_S or ω (i.e., fix_omega = 0). Comparison of models assuming a constant ω value across the CYP701 family (i.e., model = 0), versus a separate ω value for CYP701A3/AtKO (i.e., model = 2, with the relevant branch designated in the given treefile), did not indicate a significant improvement ($2\Delta l = 0.16$) despite the additional degree of freedom (i.e., free parameter). On the other hand, using distinct branch models (i.e., fitting a separate ω value) for rice CYP701A sub-family members did provide significant improvements, with $2\Delta l > 8$ ($p < 0.01$). The rice CYP701A paralogs seems to have arisen from an initial gene duplication event, with neo-functionalization of one of this original pair to a role in more specialized or secondary metabolism, including loss of KO activity. Subsequent duplication of both then led to the observed two pairs of functional genes, with CYP701A6/OsKO2 and CYP701A19 retaining KO activity (13), while CYP701A8 and CYP701A9 are functionally distinct (7, 14). Comparison of l for a separate ω value for just CYP701A6/OsKO2 versus either the lineage leading to both this and CYP701A19 ($\Delta l = 3.9$), or this latter versus just the lineage leading to CYP701A8 & 9 ($\Delta l = 39.1$), suggests that the selective pressure was most changed for these functionally distinct paralogs. Indeed, allowing separate ω values for each pair of paralogs further improves the fit ($\Delta l = 11.4$), with ω = 0.51 for the CYP701A8 & 9 lineage versus 0.33 for the CYP701A6 & 19 lineage, and 0.12 for the rest of the family, consistent with a greater change in the former (functionally distinct) lineage. However, l is not improved for this three-branch model relative to one in which only a single separate ω value is fit to the lineage leading to all the rice paralogs ($\Delta l = -0.5$). Accordingly, this last two-branch model assuming a change in selective pressure for all the rice CYP701A sub-

family members was used in further analysis. In particular, even without such a separate branch model (i.e., model = 0), it was evident that there is variable selective pressure over the sequences, as this site model (NSsites = 3 and ncatG = 3) exhibits a very significant improvement in fit ($2\Delta l = 662$), regardless of the four additional degrees of freedom. However, there is no evidence for positive selection, as l was unchanged between models with variable selective pressure but no positive selection (NSsites = 1), relative to that with positive selection (NSsites = 2), despite the two additional degrees of freedom. Similarly, the combined branch-site model (model = 2) exhibited clear evidence for variable selective pressure over the sequences ($2\Delta l = 602$), but also unchanged l for comparison of such viable selective pressure with positive selection to the null model where one ω value is set to 1 (fix_omega = 1 and omega = 1). Indeed, even when all ω values were allowed to vary (fix_omega = 0), the one of the values was estimated to be 1.00, regardless of the initial starting value (omega = 0.2 or 2). Nevertheless, this branch-site model is favored over the site model only ($2\Delta l = 104$), regardless of the additional degree of freedom. Intriguingly, this branch-site model also still suggests that a few residues may be under positive selection (six with $P > 0.95$), although the only one with a highly significant probability ($P > 0.99$; corresponding to the isoleucine at position 315 of AtKO) is unchanged in OsKO2 (and its closest paralog, CYP701A19), and is only changed to leucine or alanine in the functionally distinct CYP701A8 & 9, respectively.

Table S1: Diterpene synthases used in producing LRDs via modular metabolic engineering.

pGG*x*C (product)	Class I diterpene synthase (reference)	Product(s)	*Species* (kingdom; p = plant, b = bacteria)
pGG*e*C (*ent*-CPP)	AtKS (15)	*ent*-kaur-16-ene	*Arabidopsis thaliana* (p)
	OsKSL6 (6)	*ent*-(iso)kaur-15-ene	*Oryza sativa* (p)
	PpCPSKS (16)	*ent*-kauran-16b-ol (70%) *ent*-kaur-16-ene (30%)	*Physcomitrella patens* (p)
	RcKSL4 (17)	*ent*-beyer-15-ene	*Ricinus communis* (p)
	RcKSL2 (17)	*ent*-trachylobane (70%) *ent*-kaurene (30%)	*Ricinus communis* (p)
	OsKSL7 (6)	*ent*-cassa-12,15-diene	*Oryza sativa* (p)
	OsKSL5j (6)	*ent*-pimara-8(14),15-diene	*Oryza sativa* (p)
	OsKSL10 (6)	*ent*-sandaracopimaradiene	*Oryza sativa* (p)
pGG*s*C (*syn*-CPP)	OsKSL10 (18)	*syn*-labda-8(17),12*E*,14-triene	*Oryza sativa* (p)
	OsKSL11 (19)	*syn*-stemod-13(17)-ene	*Oryza sativa* (p)
	OsKSL8 (6)	*syn*-stemar-13-ene (60%) *syn*-stemod-12-ene (35%)	*Oryza sativa* (p)
	OsKSL4 (20)	*syn*-pimara-7,15-diene	*Oryza sativa* (p)
	OsKSL4:T696I (21)	*syn*-aphidicol-15-ene (80%) *syn*-pimara-7,15-diene (20%)	*Oryza sativa* (p)
	SaDTS (5)	*syn*-pimara-9(11),15-diene (40%) *syn*-stemod-13(17)-ene (60%)	*Salinispora arenicola* (b)
pGGnC (CPP)	SmKSL (22)	miltiradiene (abieta-12,14-diene)	*Salvia miltiorrhiza* (p)
	AgAS (23)	abieta-8(14)-en-13-ol	*Abies grandis* (p)
	OsKSL4 (18)	pimara-8(14),15-diene	*Oryza sativa* (p)
	SaDTS (5)	isopimara-8,15-diene	*Salinispora arenicola* (b)
	TaKSL1 (4)	isopimara-7,15-diene	*Triticum aven* (p)
	OsKSL11 (18)	sandaracopimaradiene (isopimara-8(14),15-diene)	*Oryza sativa* (p)
	SmCPSKSL1 (24)	labda-7,13*E*-dien-15-ol	*Selaginella mollendorffii* (p)
	KgTS (25)	sclarene (labda-8(17),13(16),14-triene)	*Kitasatospora griseola* (b)

Figure S1: Comparative turnover of AtKO versus OsKO2 with *ent*-kaurene. GC-MS chromatogram (extracted ion, *m/z* = 257) of extract from *E. coli* engineered for production of *ent*-kaurene co-expressing N-terminally modified A) AtKO and AtCPR1; B) OsKO2 and OsCPR1. Peaks correspond to successive oxidations: Peak 1, *ent*-kaurene; Peak 2, methyl ester of *ent*-kauren-19-oic acid. C) Authentic methyl ester of *ent*-kauren-19-oic acid.

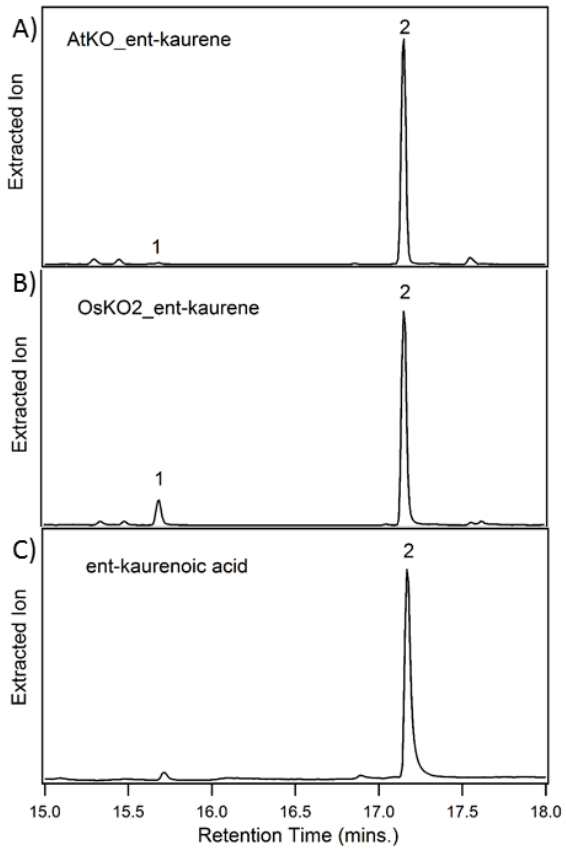


Figure S2: Comparative turnover of AtKO versus OsKO2 with *ent*-isokaurene. GC-MS chromatogram (extracted ions, *m/z* = 257 & 316) of extract from *E. coli* engineered for production of *ent*-isokaurene co-expressing N-terminally modified A) AtKO and AtCPR1; B) OsKO2 and OsCPR1. Peaks correspond to successive oxidations: Peak 1, *ent*-isokaurene; Peak 2, *ent*-isokauren-19-al; Peak 3, methyl ester of *ent*-isokauren-19-oic acid; Peak 4, *ent*-isokauren-19-ol. C) Authentic methyl ester of *ent*-isokauren-19-oic acid.

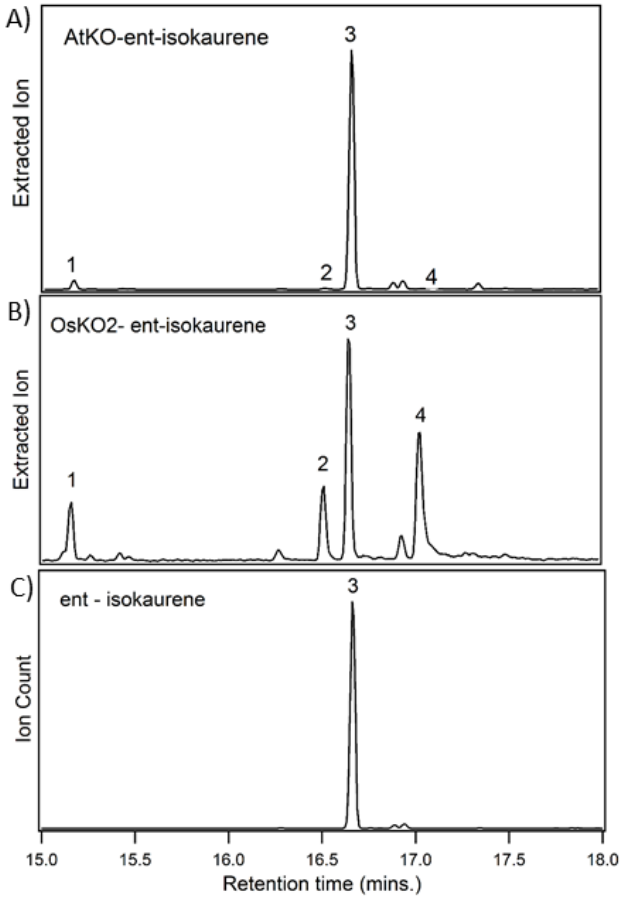


Figure S3: Comparative turnover of AtKO versus OsKO2 with *ent*-trachylobane. GC chromatogram (extracted ions, *m/z* = 272, 288 & 316) of extract from *E. coli* engineered for production of *ent*-trachylobane co-expressing N-terminally modified A) AtKO and AtCPR1; B) OsKO2 and OsCPR1. Peaks correspond to successive oxidations: Peak 1, methyl ester of *ent*-trachyloban-19-oic acid; as compared to authentic standards: C) Peak 2, *ent*-trachylobane; D) Peak 3, *ent*-trachyloban-19-al; E) Peak 4, *ent*-trachyloban-19-ol.

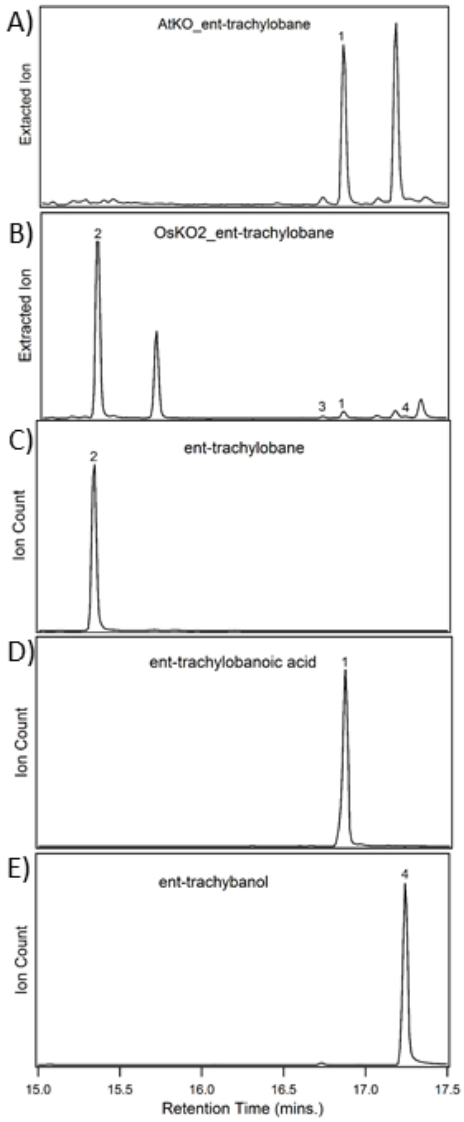


Figure S4: Comparative turnover of AtKO versus OsKO2 with *ent*-beyerene. GC-MS chromatogram (total ion count) of extract from *E. coli* engineered for production of *ent*-beyerene co-expressing N-terminally modified A) AtKO and AtCPR1; B) OsKO2 and OsCPR1. Peaks correspond to successive oxidation: Peak 1, *ent*-beyerene; Peak 2, methyl ester of *ent*-beyeren-19-oic acid; Peak 3, *ent*-beyeren-19-ol.

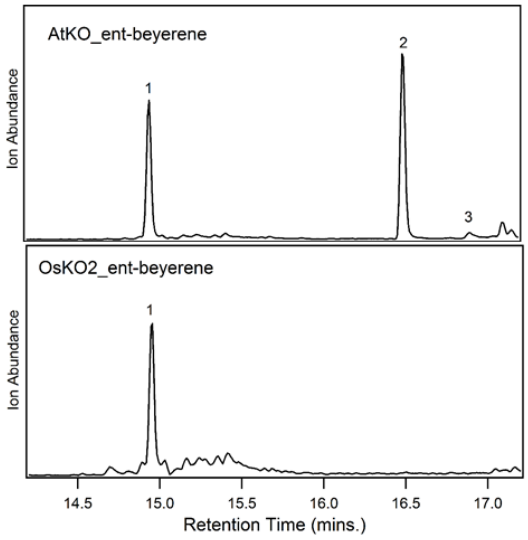


Figure S5: AtKO products with *ent*-cassadiene. GC-MS chromatogram (extracted ion, *m/z* = 257) of extract from *E. coli* engineered for production of *ent*-cassadiene co-expressing N-terminally modified A) pET-Duet/AtKO/AtCPR1; B) pCDF-Duet/AtKO/AtCPR1. Mass spectra for numbered peaks in the chromatogram. C) Peak 1, *ent*-cassadiene; D) Peak 2, *ent*-cassadien-19-al; E) Peak 3, *ent*-cassadien-19-oic acid; F) Peak 4, *ent*-cassadien-19-ol; G) Peak 5, unidentified hydroxylated product.

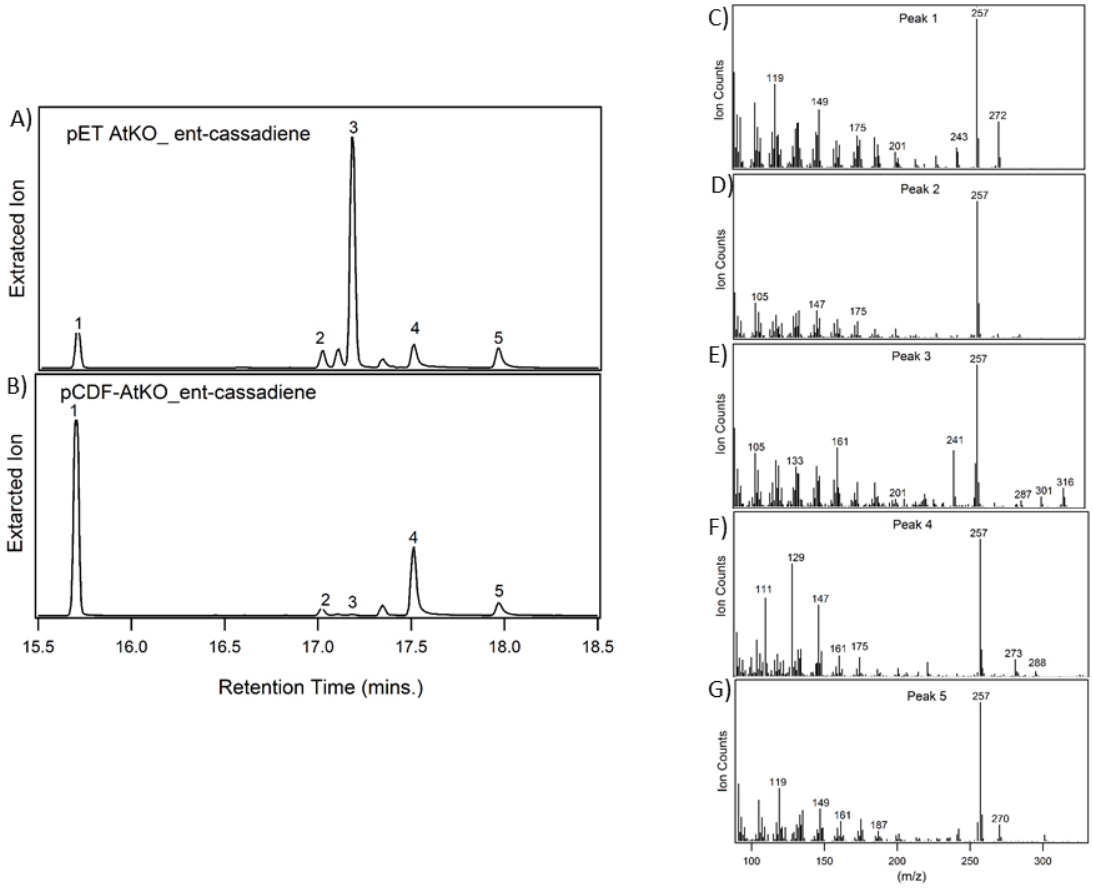


Figure S6: AtKO products with *ent*-sandaracopimaradiene. GC-MS chromatograms (total ion count) of extract from *E. coli* engineered for production of *ent*-sandaracopimaradiene co-expressing N-terminally modified A) pET-Duet/AtKO/AtCPR1; B) pCDF-Duet/AtKO/AtCPR1. Mass spectra for numbered peaks in the chromatogram. C) Peak 1, *ent*-sandaracopimaradiene; D) Peak 2, *ent*-sandaracopimaradien-19-al; E) Peak 3, *ent*-sandaracopimaradien-19-oic acid; F) Peak 4, unidentified carboxylate of *ent*-sandaracopimaradiene; G) Peak 5, *ent*-sandaracopimaradien-19-ol.

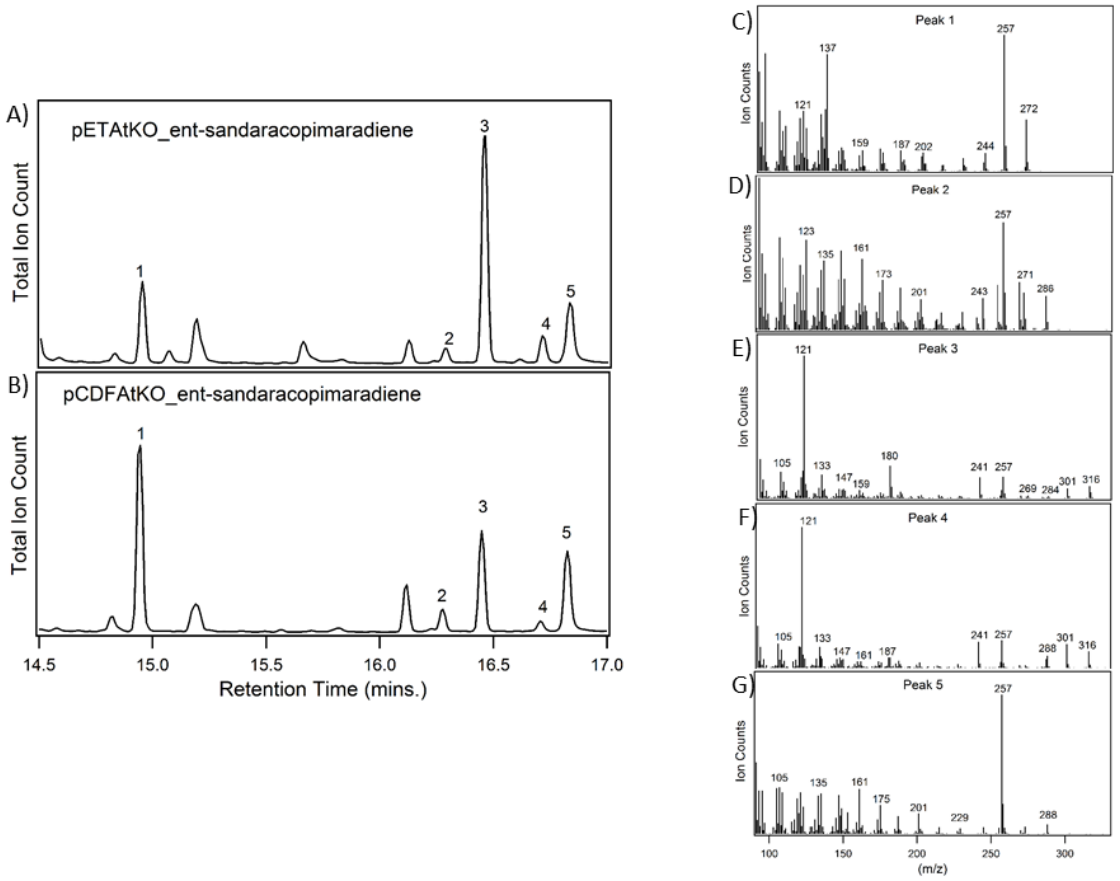


Figure S7: AtKO products with *syn*-pimaradiene and *syn*-aphidicolene. GC-MS chromatograms of extract from *E. coli* engineered for production of A) *syn*-pimara-7,15-diene (extracted ions, *m/z* = 272, 286, 288 & 316); B) *syn*-aphidicol-15-ene (total ion count), co-expressing N-terminally modified AtKO and AtCPR1. Mass spectra for numbered peaks in the chromatogram. C) Peak 1, *syn*-pimara-7,15-diene; D) Peak 2, unknown methyl ester acid; E) Peak 3, unknown 286 compound; F) Peak 4, 2α-hydroxy-*syn*-pimara-7,15-diene; G) Peak 5, 3β-hydroxy-*syn*-pimara-7,15-diene; H) Peak 6, *syn*-aphidicolene; I) Peak 7, unknown 286 compound; J) Peak 8, 3β-hydroxy-*syn*-aphidicolene.

Note: syn-aphidicolene is produced by OsKSL4:T696I which is a mutant of OsKSL4 and still produces ~ 20% syn-pimaradiene. Accordingly, peaks 1, 3, 4 & 5 in the chromatogram shown in panel B are products derived from syn-pimaradiene.

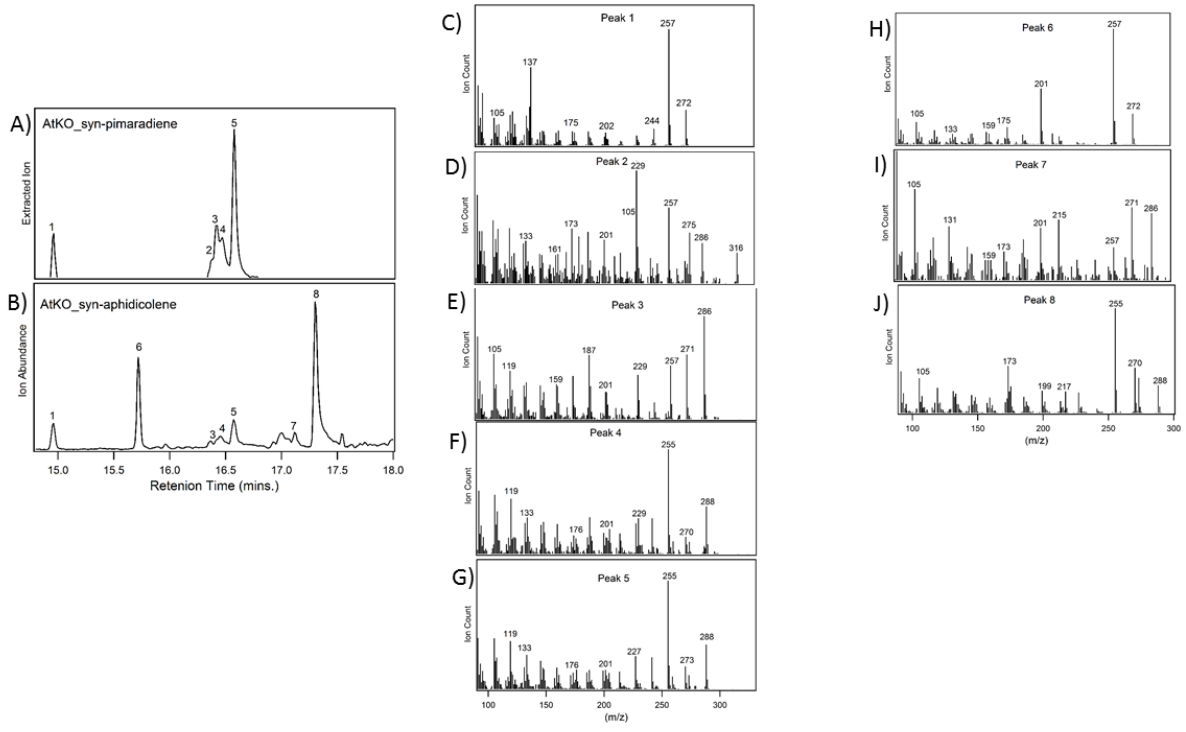


Figure S8: AtKO products with *syn*-stemod-13(17)-ene. GC-MS chromatograms (total ion count) of extract from *E. coli* engineered for production of *syn*-stemodene co-expressing N-terminally modified A) AtKO and AtCPR1; B) CYP99A3 and AtCPR1. Mass spectra for numbered peaks in the chromatogram, those corresponding to successive oxidation of the C4β methyl (C19) are compared to the known products from CYP99A3. C) Peak 4, 3β-hydroxy-*syn*-stemodene; D & E) Peaks 1 & 6, *syn*-stemodene; F & G) Peak 2 compared to CYP99A3 Peak 7, *syn*-stemoden-19-al; H & I) Peak 3 compared to CYP99A3 Peak 8, *syn*-stemoden-19-oic acid; J & K) Peak 5 compared to CYP99A3 Peak 9, *syn*-stemoden-19-ol.

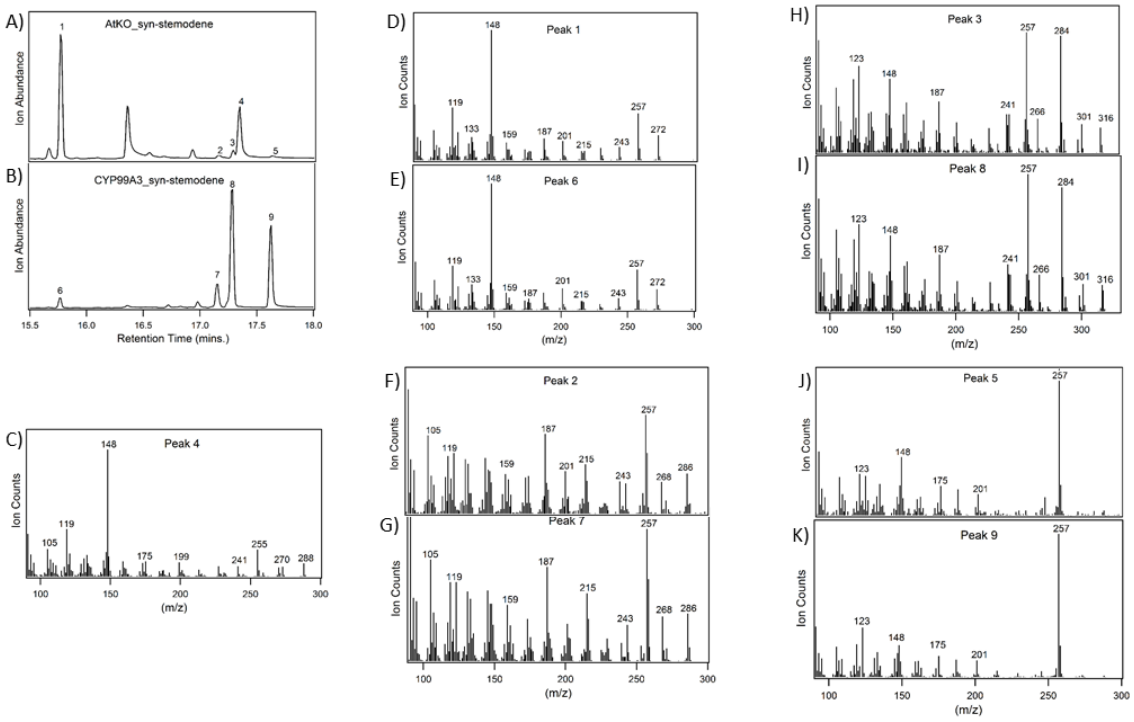


Figure S9: AtKO products with *syn*-stemar-13-ene and *syn*-stemod-12-ene. A) GC-MS chromatogram (extracted ion count, *m/z* = 288) of extract from *E. coli* engineered for production of *syn*-stemar-13-ene and *syn*-stemo-12-ene co-expressing N-terminally modified AtKO and AtCPR1. Mass spectra for numbered peaks in the chromatogram. B) Peak 1, unidentified hydroxylated product; C) Peak 2, unidentified hydroxylated product; D) Peak 3, unidentified hydroxylated product.

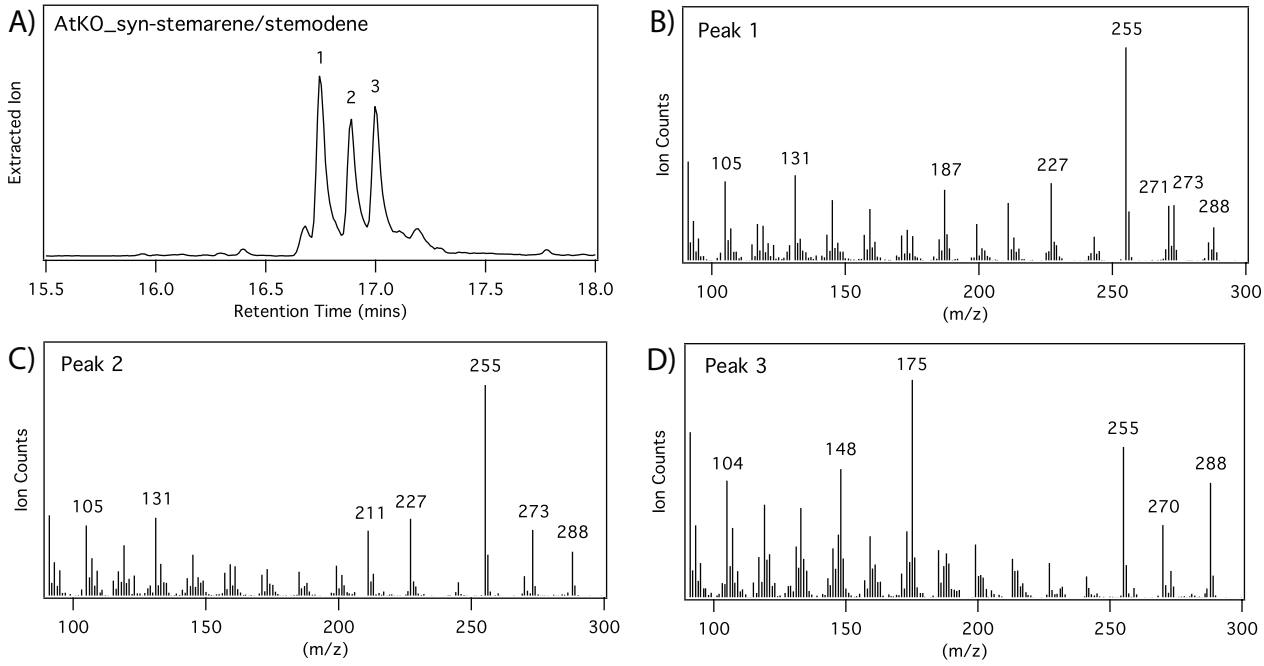


Figure S10: AtKO products with *syn*-labda-8(17),12*E*,14-triene. A) GC-MS chromatogram (extracted ions *m/z* = 255, 257 & 288) of extract from *E. coli* engineered for production of *syn*-labdatriene co-expressing N-terminally modified AtKO and AtCPR1. Mass spectra for numbered peaks in the chromatogram. B) Peak 1, *syn*-labdatriene; C) Peak 2, 3β-hydroxy-*syn*-labda-8(17),12*E*,14-triene.

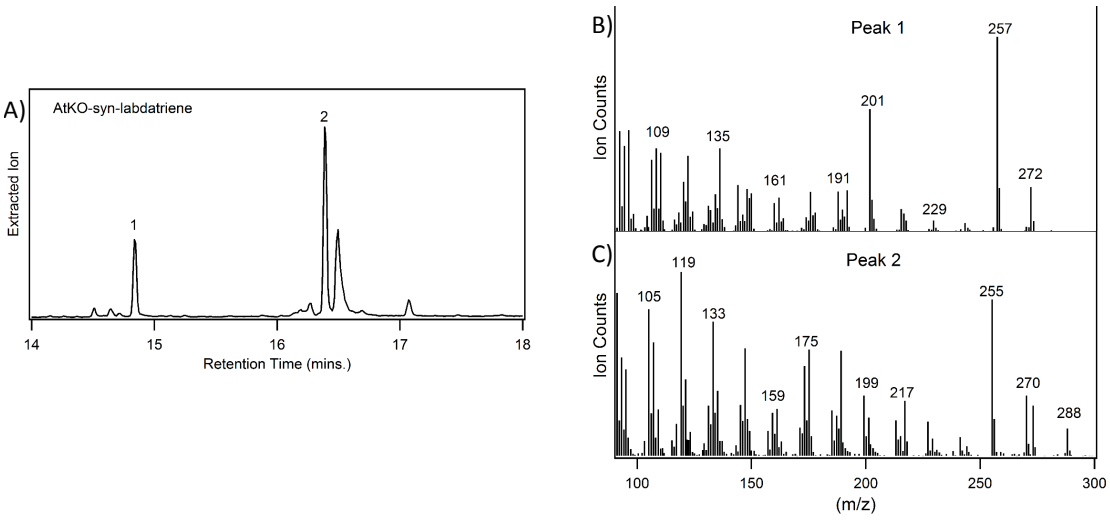


Figure S11: AtKO products with *syn*-pimara-9(11),15-diene. A) GC-MS chromatogram (extracted ion, *m/z* = 255 & 288) of extract from *E. coli* engineered for production of *syn*-pimara-9(11),15-diene co-expressing N-terminally modified AtKO and AtCPR1. Mass spectra for numbered peaks in the chromatogram. B, C & D) unidentified hydroxylated *syn*-pimara-9(11),15-diene products.

*Note that production of syn-pimara-9(11),15-diene relied on use of a previously reported bacterial class I diterpene synthase, SaDTS, which also produces significant amounts of syn-stemodene - * indicates peaks corresponding to syn-stemodene and derived products.*

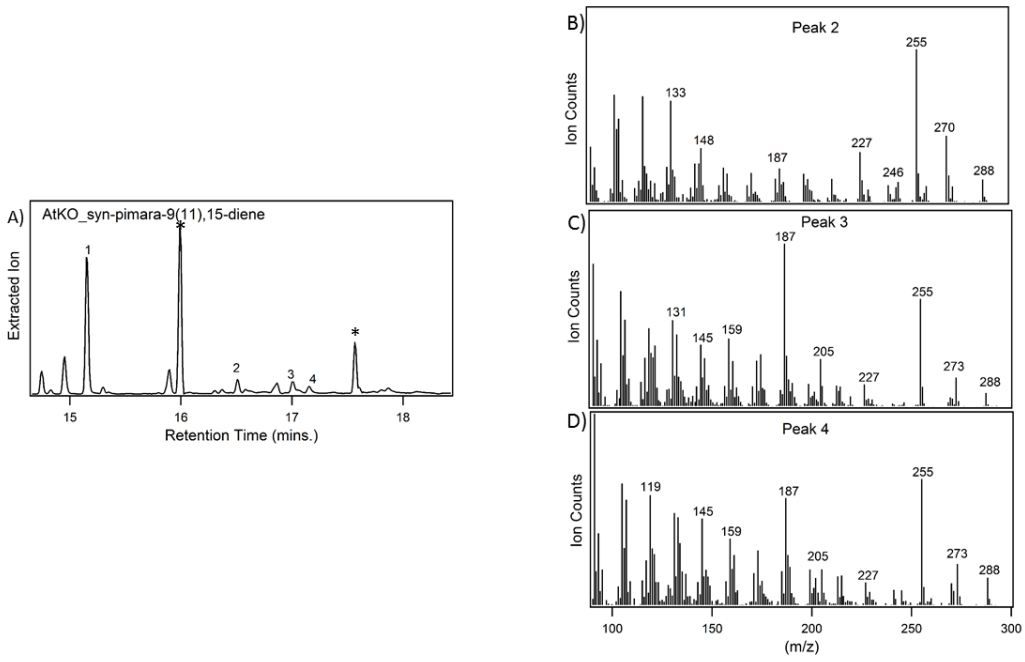


Figure S12: AtKO products with isopimara-8,15-diene. A) GC-MS chromatogram (total ion count) of extract from *E. coli* engineered for production of pimara-8,15-diene co-expressing N-terminally modified AtKO and AtCPR1. Mass spectra for numbered peaks in the chromatogram. B) Peak 1, isopimara-8,15-diene C) Peak 2, 3β-hydroxy-isopimara-8,15-diene.

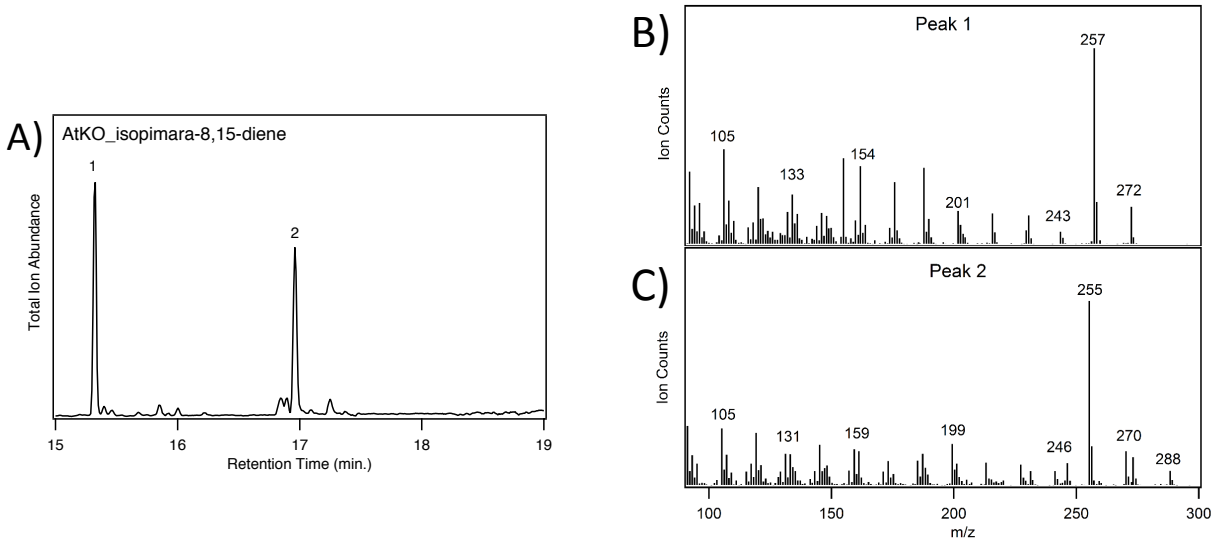


Figure S13: AtKO products with isopimara-7,15-diene. A) GC-MS chromatogram (total ion count) of extract from *E. coli* engineered for production of pimara-7,15-diene co-expressing N-terminally modified AtKO and AtCPR1. Mass spectra for numbered peaks in the chromatogram. B) Peak 1, isopimara-7,15-diene; C) Peak 2, isopimara-7,15-dien-2-one; D) Peak 3, 2α-hydroxy-isopimara-7,15-diene.

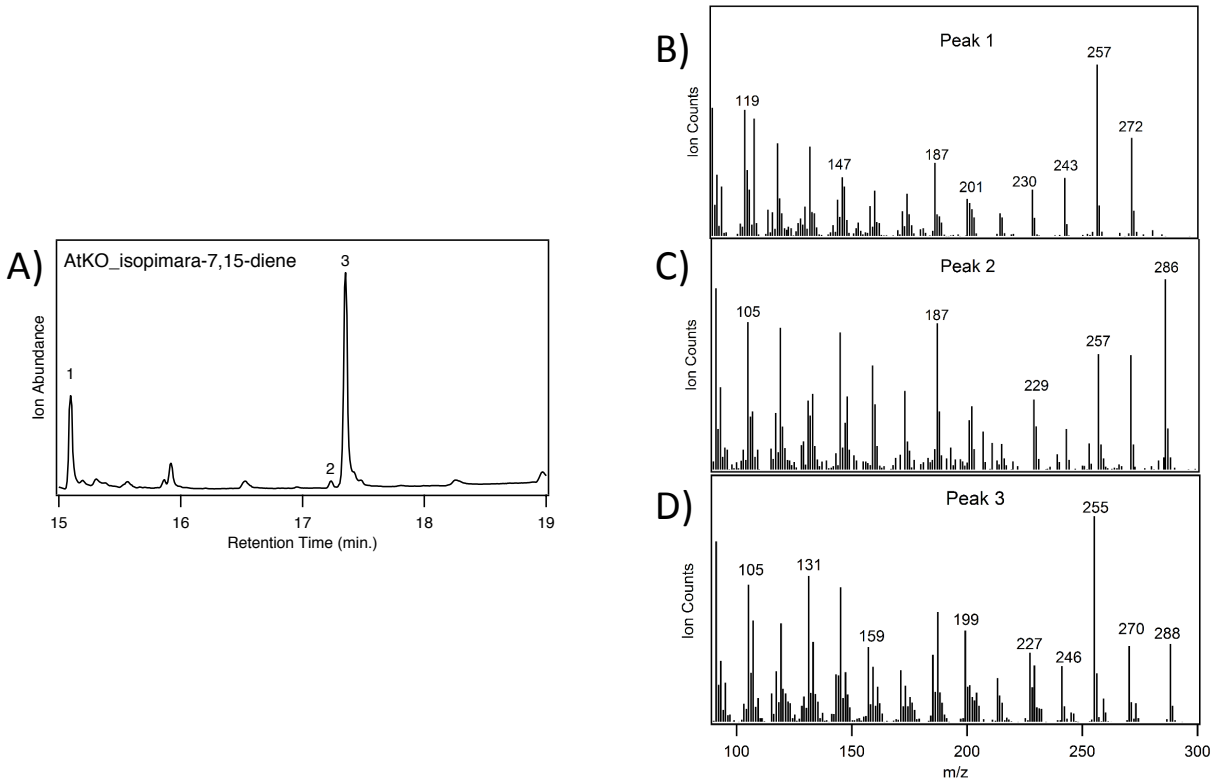


Figure S14: AtKO products with sandaracopimaradene. A) GC-MS chromatogram (extracted ion, *m/z* = 288) of extract from *E. coli* engineered for production of sandaracopimaradene co-expressing N-terminally modified AtKO and AtCPR1. Mass spectra for numbered peaks in the chromatogram. B) Peak 1, 2α-hydroxy-sandaracopimaradiene; C) Peak 2, 3β-hydroxy-sandaracopimaradiene; D) Peak 3, unidentified hydroxylated product; E) Peak 4, methyl ester acid of unidentified carboxylic acid product.

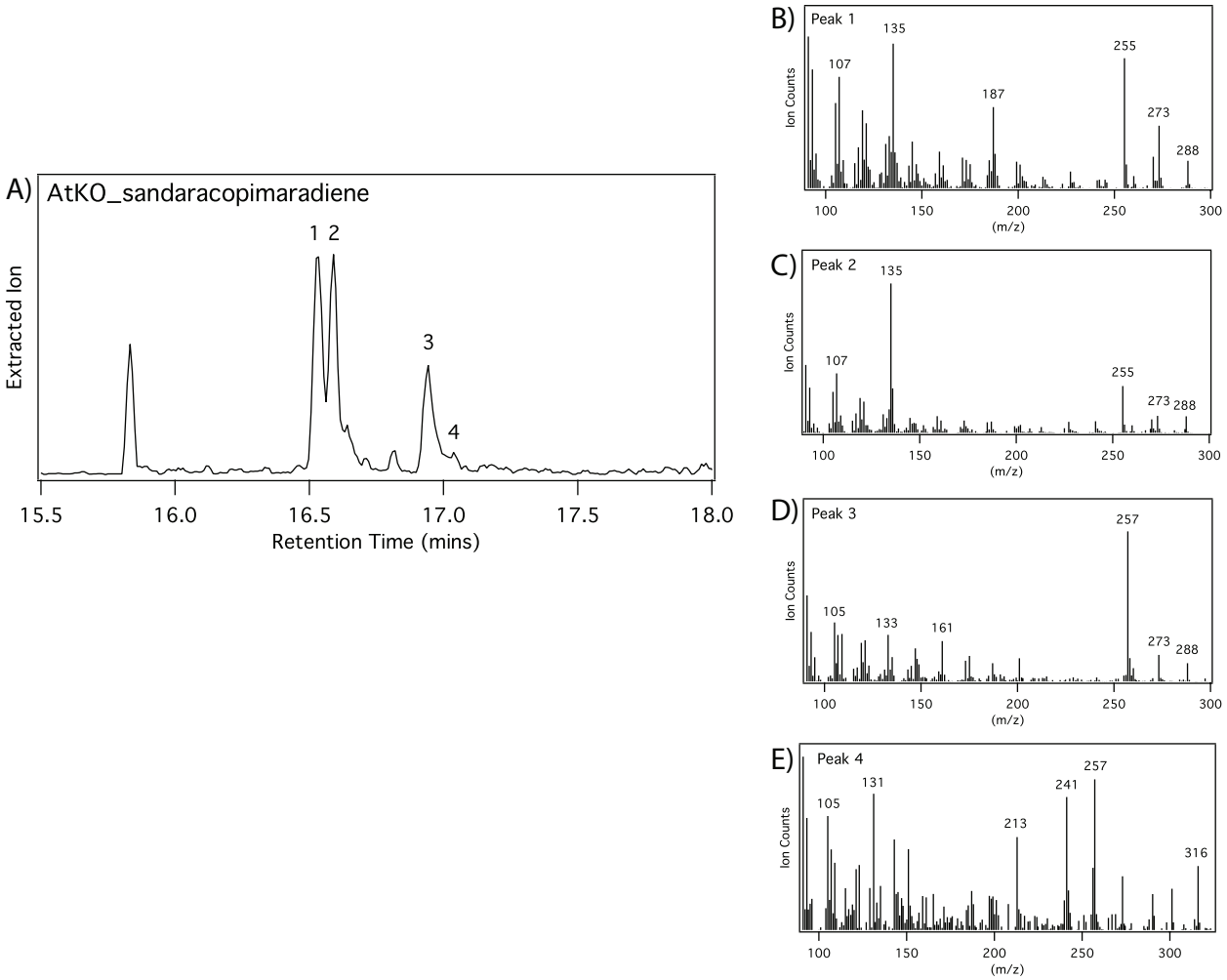
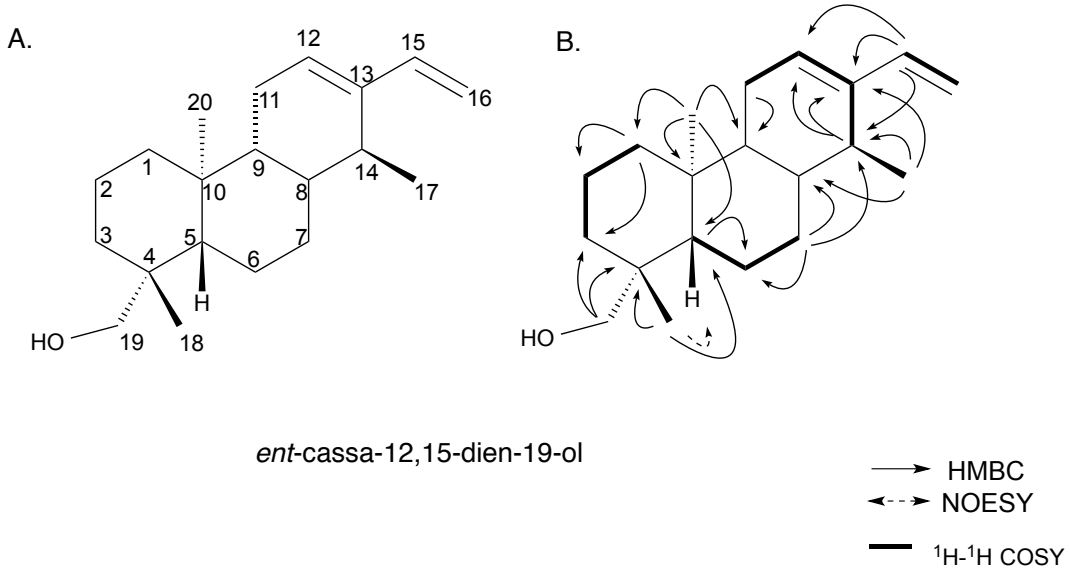
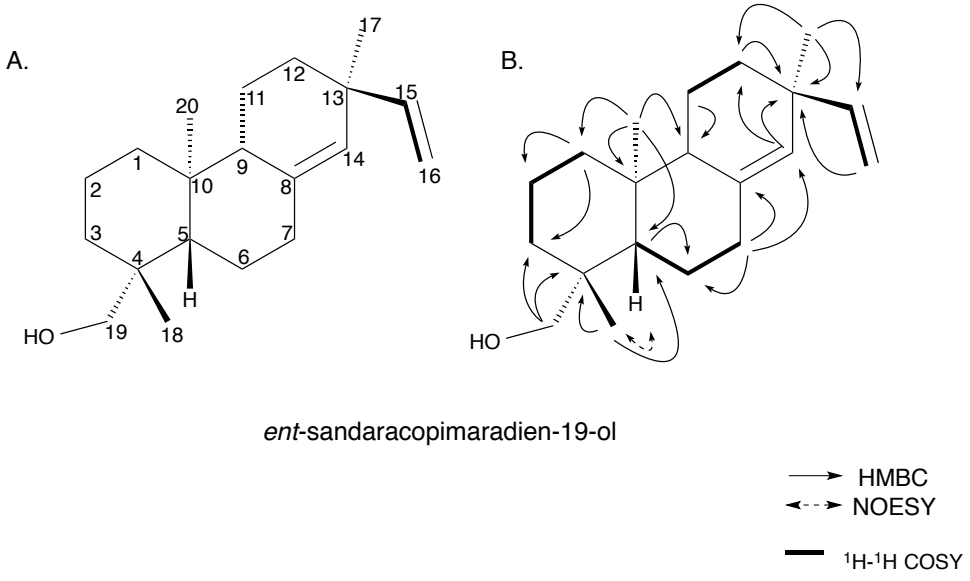


Figure S15: AtKO product with *ent*-cassadiene. A) numbering;) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.



ent-cassa-12,15-dien-19-ol

Figure S16: AtKO product with *ent*-sandaracopimaradiene. A) Numbering; B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.



ent-sandaracopimaradien-19-ol

Figure S17: AtKO products for *syn*-pimaradiene. A & B) Numbering; C & D) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.

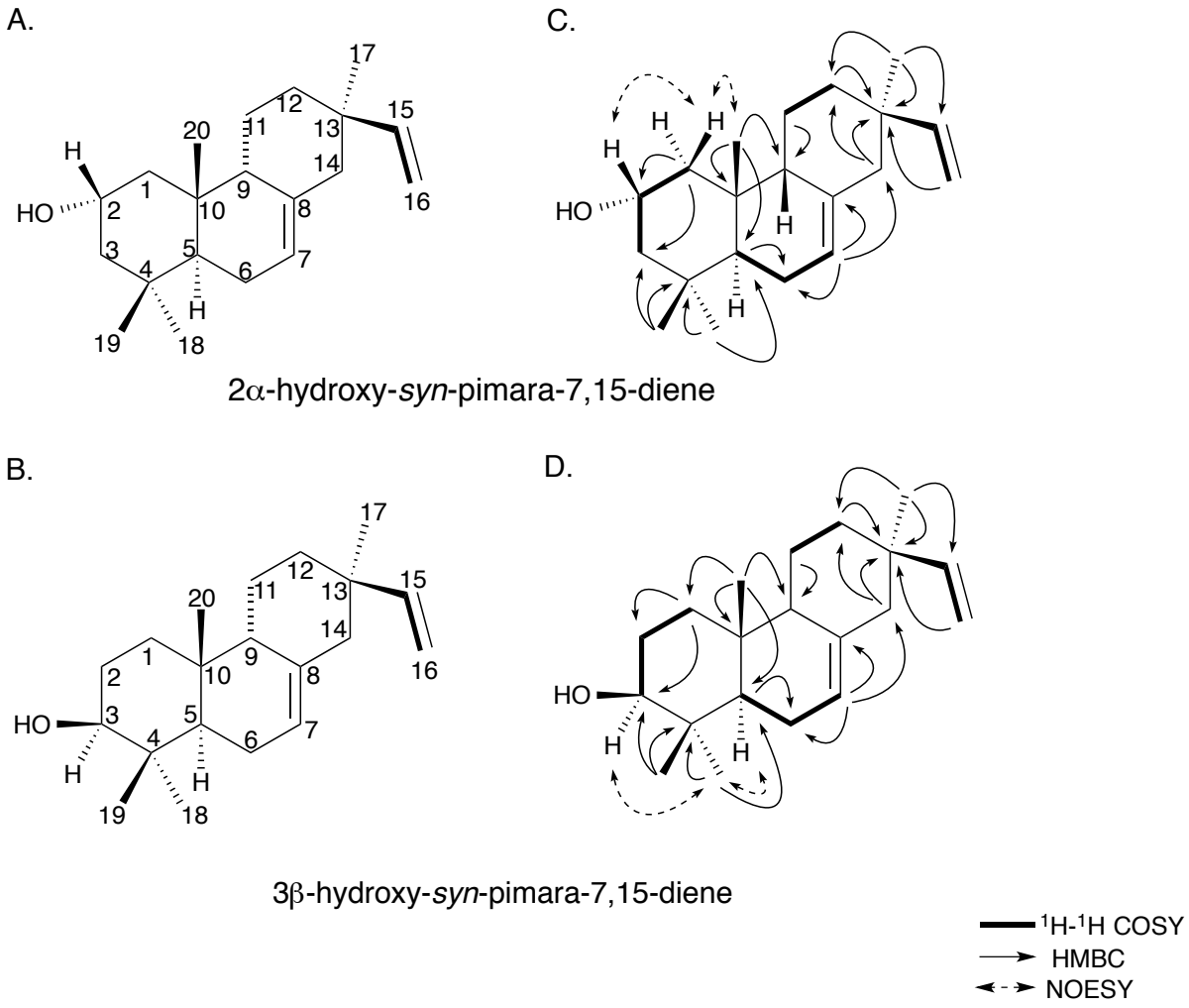


Figure S18: AtKO product with *syn*-aphidicol-15-ene. A) Numbering; B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.

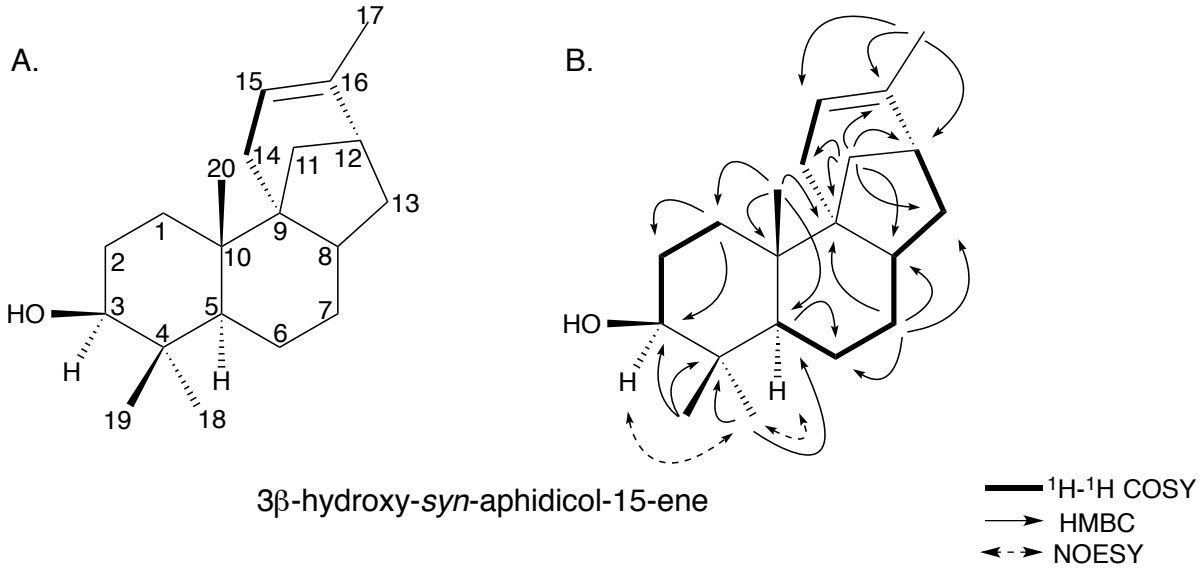


Figure S19: AtKO product with *syn*-labda-8(17),12*E*,14-triene. A) Numbering; B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.

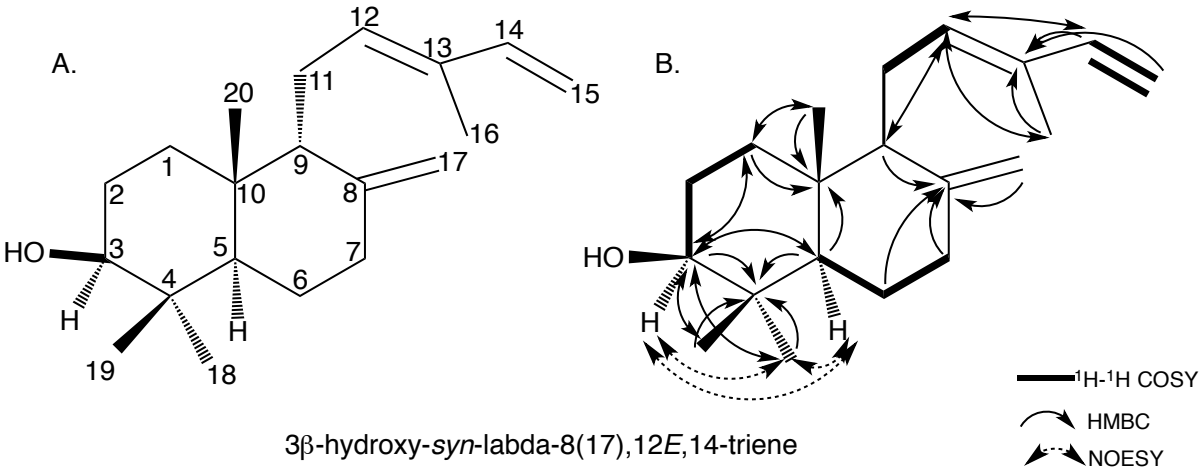


Figure S20: AtKO product with *syn*-stemod-13(17)-ene. A) Numbering; B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.

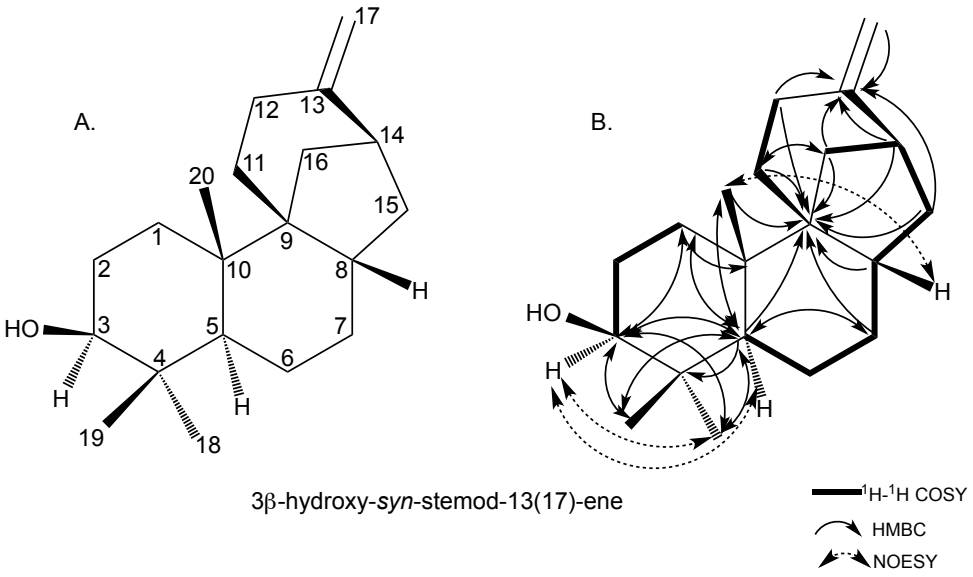


Figure S21: AtKO product with isopimara-8,15-diene. A) Numbering; B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.

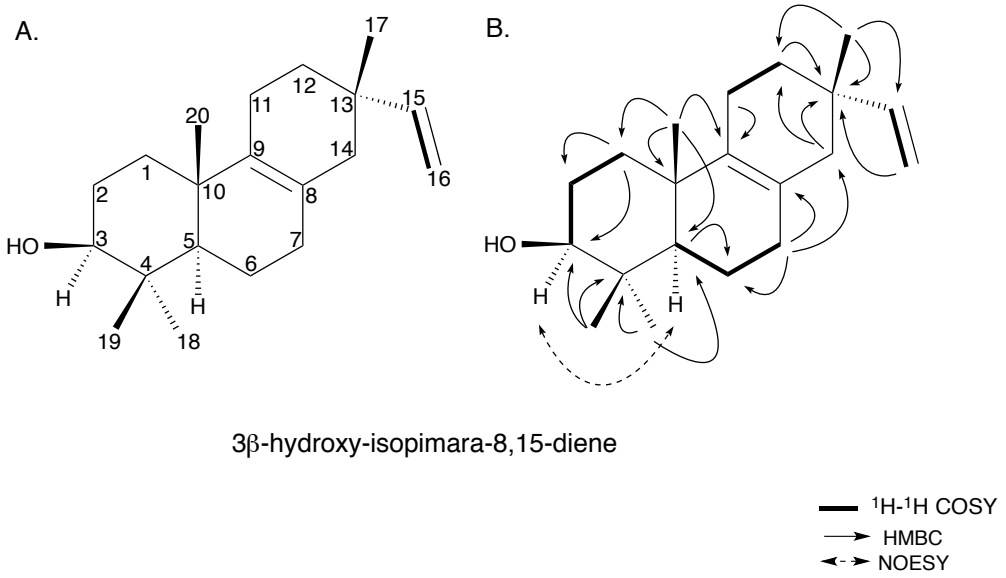


Figure S22: AtKO products with isopimara-7,15-diene. A & C) Numbering; B & D) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.

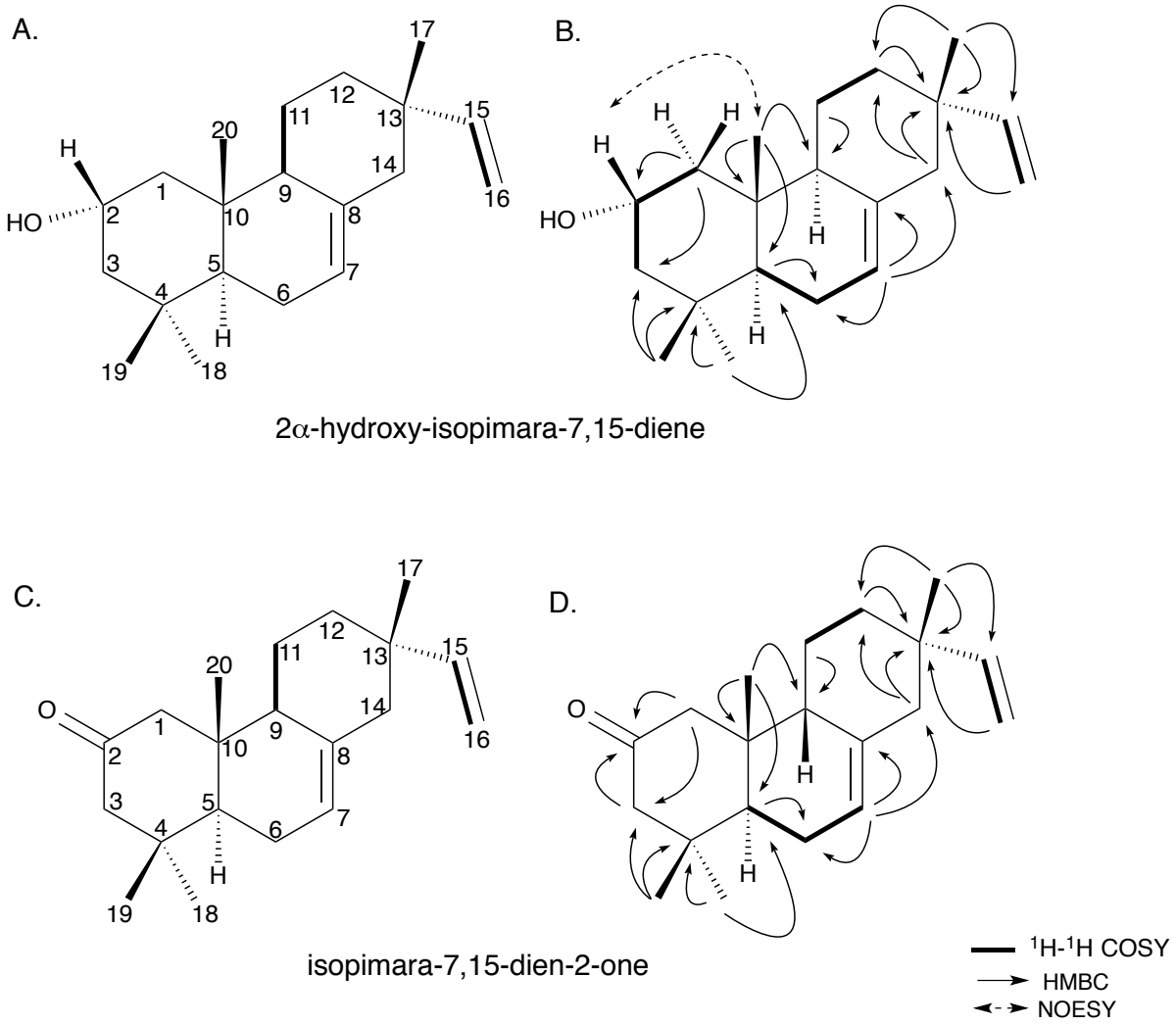


Figure S23: AtKO products with sandaracopimara-8(14),15-diene A & B) Numbering, C & D) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations.

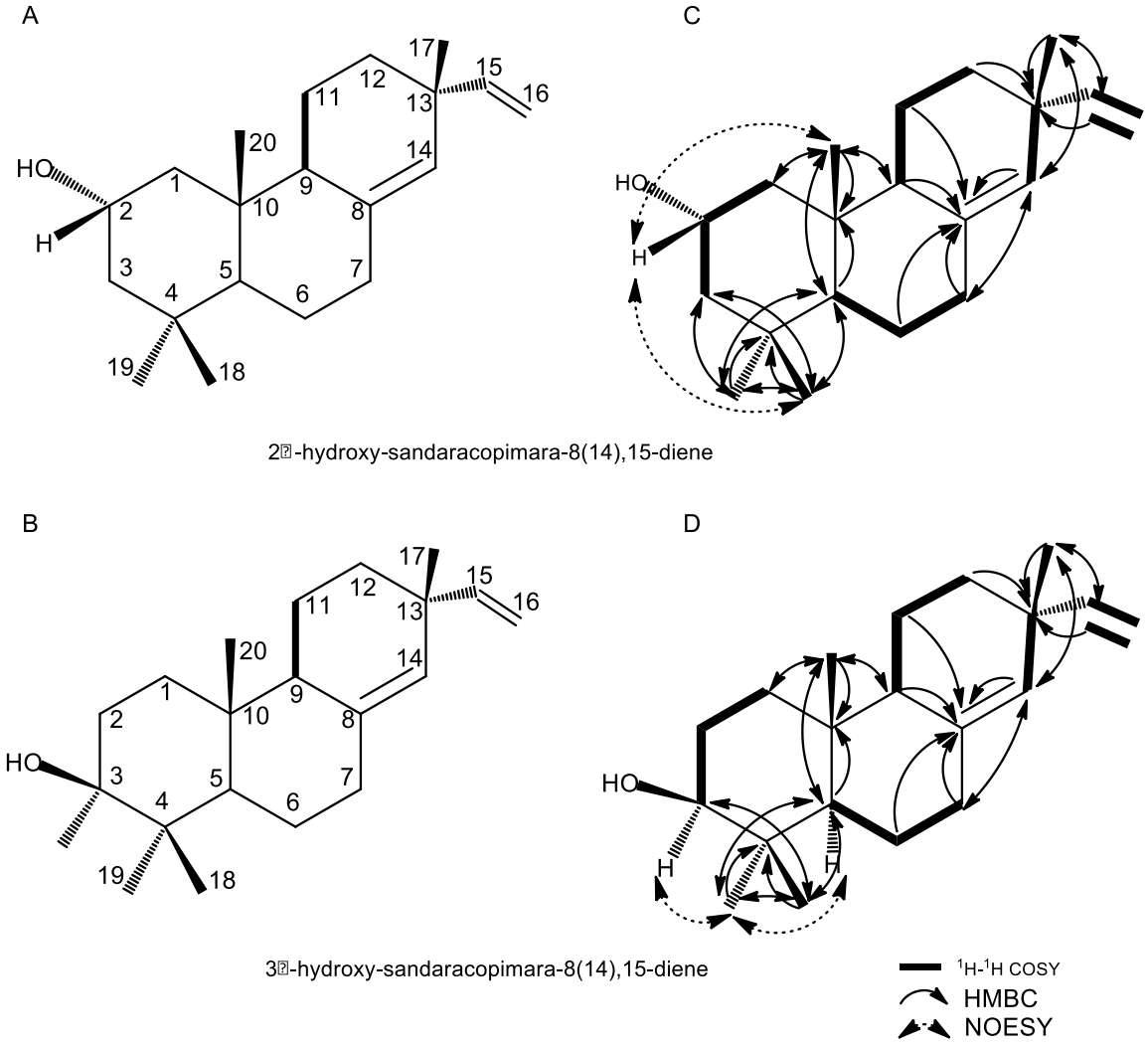


Table S2: ^{1}H and ^{13}C NMR assignments for *ent*-cassa-12,15-dien-19-ol.

	***ent*-cassa-12,15-dien-19-ol**	
Position	δ_H	δ_C
1 a	1.63 (1H, m)	39.4
b	0.96 (1H, d)	
2 a	1.47 (1H, m)	18.6
b	1.41 (1H, m)	
3 a	1.80 (1H, m)	35.7
b	0.95 (1H, m)	
4		38.7
5	1.03 (1H, m)	56.5
6 a	1.71 (1H, m)	21.8
b	1.33 (1H, m)	
7 a	1.63 (1H, m)	31.6
b	1.32 (1H, m)	
8	1.51 (1H, m)	35.3
9	1.30 (1H, m)	44.6
10		37.2
11 a	2.09 (1H, m)	25.6
11 b	1.90 (1H, ddd, J = 19.9, 9.3, 3.4 Hz)	
12	5.58 (1H, t, J = 3.8 Hz)	128.7
13		142.1
14	2.37 (1H, m)	32.2
15	6.20 (1H, dd, J = 17.6, 10.8 Hz)	138.9
16 a	5.05 (1H, d, J = 17.6 Hz)	109.7
b	4.88 (1H, d, J = 10.8 Hz)	
17	0.90 (3H, s)	14.7
18	0.96 (3H, s)	27.1
19	3.78 (1H, d, J = 10.9 Hz)	65.8
	3.44 (1H, d, J = 10.9 Hz)	
20	0.77 (3H, s)	15.1

Table S3: ^{1}H and ^{13}C NMR assignments for *ent*-sandaracopimaradien-19-ol.

	***ent*-sandaracopimaradien-19-ol**	
Position	δ_H	δ_C
1 a	1.66 (1H, m)	39.9
b	0.98 (1H, m)	
2 a	1.42 (1H, m)	19.3
b	1.39 (1H, m)	
3 a	1.77 (1H, br.d, J = 13.7 Hz)	36.0
b	0.89 (1H, m)	
4		39.2
5	1.13 (1H, dd, J = 12.9, 1.9 Hz)	56.4
6 a	1.60 (1H, m)	23.1
b	1.21 (1H, m)	
7 a	2.17 (1H, ddd, J = 14.1, 4.3, 1.9 Hz)	36.9
b	1.93 (1H, m)	
8		137.4
9	1.63 (1H, m)	51.3
10		38.8
11 a	1.52 (1H, m)	19.5
b	1.39 (1H, m)	
12	1.37 (1H, m)	35.1
	1.27 (1H, dd, J = 12.5, 3.7 Hz)	
13		38.0
14	5.13 (1H, m)	129.4
15	5.69 (1H, dd, J = 17.5, 10.6 Hz)	149.6
16 a	4.82 (1H, d, J = 17.5 Hz)	110.6
b	4.80 (1H, d, J = 10.6 Hz)	
17	0.95 (3H, s)	26.6
18	0.90 (3H, s)	27.6
19	3.75 (1H, d, J = 10.8 Hz)	65.8
	3.35 (1H, d, J = 10.8 Hz)	
20	0.68 (3H, s)	16.7

Table S4: ^{1}H and ^{13}C NMR assignments for 2α-hydroxy-*syn*-pimara-7,15-diene.

Position	**2α-hydroxy-*syn*-pimara-7,15-diene**	
	δ_H	δ_C
1 a	1.46 (1H, m)	46.8
b	1.27 (1H, t, *J* = 11.3 Hz)	
2	3.84 (1H, tt, *J* = 11.3, 3.3 Hz)	65.9
3 a	1.68 (1H, dt, *J* = 12.2, 3.3)	52.5
b	1.04 (1H, dd, *J* = 12.2, 11.3 Hz)	
4		35.2
5	1.12 (1H, dd, *J* =12.0, 4.5 Hz)	43.5
6 a	1.94 (1H, m)	24.21
b	1.76 (1H, m)	
7	5.24 (1H, br.d, *J* = 5.3 Hz)	120.3
8		137.1
9	1.31 (1H, m)	53.7
10		37.6
11 a	1.63 (1H, m)	25.7
11 b	1.20 (1H, m)	
12	1.44 (1H, m)	38.2
	1.37 (1H, td, *J* = 13.1, 4.2 Hz)	
13		39.4
14 a	1.92 (1H, m)	48.4
b	1.75 (1H, *J* = 12.3, 2.2 Hz)	
15	5.74 (1H, dd, *J* = 17.5, 10.7 Hz)	150.9
16 a	4.85 (1H, d, *J* = 17.5 Hz)	109.8
b	4.78 (1H, d, *J* = 10.7 Hz)	
17	0.81 (3H, s)	22.4
18	0.870 (3H, s)	34.1
19	0.873 (3H, s)	24.18
20	0.90 (3H, s)	23.5

Table S5: ^{1}H and ^{13}C NMR assignments for 3β-hydroxy-*syn*-pimara-7,15-diene.

Position	**3β-hydroxy-*syn*-pimara-7,15-diene**	
	δ_H	δ_C
1 a	1.50 (1H, m)	34.8
b	1.16 (1H, m)	
2	1.56 (2H, m)	28.0
3	3.12 (1H, t, *J* = 7.6 Hz)	80.2
4		39.4
5	1.13 (1H, m)	43.8
6 a	1.93 (1H, m)	24.0
b	1.88 (1H, m)	
7	5.24 (1H, br.d, *J* = 4.8 Hz)	120.1
8		137.3
9	1.28 (1H, m)	53.6
10		35.6
11 a	1.61 (1H, m)	25.7
11 b	1.14 (1H, m)	
12	1.42 (1H, m)	38.3
	1.36 (1H, *J* = 13.1, 4.0 Hz)	
13		39.2
14 a	1.92 (1H, m)	48.4
b	1.75 (1H, *J* = 12.2, 2.2 Hz)	
15	5.74 (1H, dd, *J* = 17.5, 10.8 Hz)	150.9
16 a	4.84 (1H, d, *J* = 17.5 Hz)	109.8
b	4.78 (1H, d, *J* = 10.8 Hz)	
17	0.802 (3H, s)	22.4
18	0.93 (3H, s)	28.9
19	0.799 (3H, s)	16.2
20	0.85 (3H, s)	22.8

Table S6: ^{1}H and ^{13}C NMR assignments for 3β-hydroxy-*syn*-aphidicol-15-ene.

Position	**3β-hydroxy-*syn*-aphidicol-15-ene**	
	δ_H	δ_C
1 a	1.68 (1H, td, *J* = 12.4, 5.5 Hz)	32.4
b	1.14 (1H, m)	
2	1.56 (2H, m)	28.6
3	3.08 (1H, dd, *J* = 10.1, 6.0 Hz)	79.9
4		39.5
5	1.21 (1H, m)	45.1
6 a	1.54 (1H, m)	23.3
b	1.28 (1H, m)	
7 a	1.58 (1H, m)	26.1
b	1.36 (1H, qd, *J* = 12.6, 4.4 Hz)	
8	2.00 (1H, m)	42.4
9		48.8
10		40.5
11 a	1.87 (1H, td, *J* = 11.3, 7.5 Hz)	39.0
b	0.93 (1H, d, *J* = Hz)	
12	2.00 (1H, m)	39.6
13	1.44 (1H, dd, *J* = 10.3, 5.4 Hz)	36.0
	1.17 (1H, m)	
14 a	2.19 (1H, m)	28.5
b	2.08 (1H, m)	
15	4.90 (1H, br.s)	118.4
16		148.0
17	1.56 (3H, s)	22.3
18	0.90 (3H, s)	29.2
19	0.72 (3H, s)	15.8
20	0.92 (3H, s)	16.3

Table S7: ^{1}H and ^{13}C NMR assignments for 3β-hydroxy-*syn*-labda-8(17),12*E*,14-triene.

Position	**3β-hydroxy-*syn*-labda-8(17),12*E*,14-triene**	
	δ_H	δ_C
1 a	1.67 (1H,m)	34.7
b	1.15 (1H, d, *J* = 11.1 Hz)	
2	1.66 (2H, m)	28.0
3	3.23 (1H, dd, *J* = 10.6, 4.3 Hz)	79.5
4		39.1
5	1.29 (1H, m)	45.3
6 a	1.63 (1H, m)	23.4
b	1.37 (1H, m)	
7 a	2.17 (1H, d, *J* = 14.8 Hz)	31.3
b	2.04 (1H, td, *J* = 13.6, 5.7 Hz)	
8		148.5
9	1.67 (1H, m)	58.0
10		38.1
11 a	2.37 (1H, m)	25.9
b	2.13 (1H, dd, *J* = 16.8, 8.6 Hz)	
12	5.34 (1H, t, *J* = 7.11 Hz)	132.9
13		133.8
14	6.32 (1H, dd, *J* = 17.3, 10.8 Hz)	141.9
15 a	5.03 (1H, d, *J* = 17.3 Hz)	110.1
b	4.87 (1H, d, *J* = 10.8 Hz)	
16	1.7 (3H, s)	12.1
17 a	4.66 (1H, br.s)	110.3
b	4.50 (1H, br.s)	
18	1.0 (3H, s)	28.6
19	0.77 (3H, s)	15.8
20	0.91 (3H, s)	22.5

Table S8: ^{1}H and ^{13}C NMR assignments for 3β-hydroxy-*syn*-stemod-13(17)-ene.

Position	**3β-hydroxy-*syn*-stemod-13(17)-ene**	
	δ_H	δ_C
1 a	1.63 (1H, dt, *J* = 13.2, 3.4 Hz)	34.2
b	1.28 (1H, m)	
2 a	1.55 (1H, m)	27.6
b	1.46 (1H, m)	
3	3.17 (1H, dd, *J* = 12.0, 4.3 Hz)	79.4
4		38.7
5	1.21 (1H, dd, *J* = 12.0, 1.6 Hz)	47.3
6 a	1.40 (1H, m)	22.4
b	1.27 (1H, m)	
7 a	1.91 (1H, m)	37.0
b	1.10 (1H, m)	
8	1.84 (1H, m)	39.1
9		51.5
10		33.5
11 a	1.55 (1H, m)	32.3
b	1.45 (1H, m)	
12 a	2.26 (1H, m)	28.3
b	2.07 (1H, m)	
13		155.6
14	2.67 (1H, t, *J* = 6.6 Hz)	43.6
15 a	1.76 (1H, dd, *J* = 13.6, 8.2 Hz)	39.7
b	1.31 (1H, m)	
16 a	2.04 (1H, m)	37.8
b	1.17 (1H, d, *J* = 11.4 Hz)	
17 a	4.43 (1H, t, *J* = 2.2 Hz)	102.0
b	4.35 (1H, t, *J* = 2.2 Hz)	
18	0.99 (3H, s)	29.6
19	0.83 (3H, s)	16.6
20	0.95 (3H, s)	18.9

Table S9: ^{1}H and ^{13}C NMR assignments for 3β-hydroxy-isopimara-8,15-diene.

Position	**3β-hydroxy-isopimara-8,15-diene**	
	δ_H	δ_C
1 a	1.66 (1H, m)	35.3
b	1.06 (1H, td, *J* = 12.9, 3.5 Hz)	
2 a	1.59 (1H, m)	28.3
b	1.51 (1H, m)	
3	3.15 (1H, dd, *J* = 11.3, 3.1 Hz)	79.6
4		39.4
5	1.02 (1H, d, *J* = 12.2 Hz)	51.6
6 a	1.60 (1H, m)	19.2
b	1.40 (1H, m)	
7	1.86 (2H, m)	33.2
8		125.2
9		137.0
10		37.9
11	1.80 (2H, m)	21.8
12 a	1.41 (1H, m)	35.5
b	1.23 (1H, m)	
13		35.7
14 a	1.73 (1H, d, *J* = 17.3 Hz)	42.4
b	1.64 (1H, d, *J* = 17.3 Hz)	
15	5.65 (1H, dd, *J* = 17.3, 10.7 Hz)	146.8
16 a	4.81 (1H, d, *J* = 10.7 Hz)	111.3
b	4.76 (1H, d, *J* = 17.3 Hz)	
17	0.884 (3H, s)	28.6
18	0.92 (3H, s)	28.5
19	0.73 (3H, s)	16.1
20	0.877 (3H, s)	20.0

Table S10: ^{1}H and ^{13}C NMR assignments for 2α-hydroxy-isopimara-7,15-diene.

Position	**2α-hydroxy-isopimara-7,15-diene**	
	δ_H	δ_C
1 a	2.05 (1H, dt, *J* = 11.9, 2.8 Hz)	49.6
b	0.88 (1H, m)	
2	3.77 (1H, m)	65.8
3 a	1.68 (1H, dt, *J* = 12.0, 2.9 Hz)	51.8
b	1.06 (1H, t, *J* = 12.0 Hz)	
4		35.3
5	1.03 (1H, dd, *J* = 12.0, 4.0 Hz)	50.3
6 a	1.87 (1H, m)	23.9
b	1.78 (1H, m)	
7	5.30 (1H, br.d, *J* = 3.1 Hz)	122.2
8		135.9
9	1.63 (1H, br.d, *J* = 9.8 Hz)	52.6
10		38.0
11 a	1.48 (1H, m)	20.8
11 b	1.28 (1H, m)	
12	1.41 (1H, m)	36.6
	1.29 (1H, m)	
13		37.4
14 a	1.89 (1H, m)	46.6
b	1.84 (1H, dd, *J* = 13.9, 2.5 Hz)	
15	5.72 (1H, dd, *J* =17.5, 10.8 Hz)	150.9
16 a	4.85 (1H, d, *J* = 17.5 Hz)	109.9
b	4.79 (1H, d, *J* = 10.8 Hz)	
17	0.78 (3H, s)	22.1
18	0.85 (3H, s)	34.2
19	0.88 (3H, s)	23.8
20	0.82 (3H, s)	16.4

Table S11: ^{1}H and ^{13}C NMR assignments for isopimara-7,15-dien-2-one.

	isopimara-7(8),15-dien-2-one	
Position	δ_H	δ_C
1 a	2.35 (1H, dd, *J* = 12.6, 2.3 Hz)	54.9
b	2.13 (1H, d, *J* = 12.6 Hz)	
2		212.3
3 a	2.31 (1H, d, *J* = 12.5 Hz)	57.0
b	2.03 (1H, dd, *J* = 12.5, 2.4 Hz)	
4		40.0
5	1.64 (1H, *J* = 11.7, 4.3 Hz)	50.7
6 a	1.98 (1H, m)	24.3
b	1.90 (1H, m)	
7	5.33 (1H, br.d, *J* = 4.1 Hz)	122.0
8		136.1
9	1.85 (1H, m)	52.7
10		42.7
11 a	1.38 (1H, m)	20.9
11 b	1.30 (1H, m)	
12	1.42 (1H, m)	36.5
	1.29 (1H, m)	
13		37.3
14 a	1.91 (1H, m)	46.4
b	1.89 (1H, m)	
15	5.72 (1H, dd, *J* = 17.5, 10.7 Hz)	150.6
16 a	4.86 (1H, d, *J* = 17.5 Hz)	110.1
b	4.80 (1H, d, *J* = 10.7 Hz)	
17	0.78 (3H, s)	22.1
18	0.98 (3H, s)	33.7
19	0.88 (3H, s)	23.8
20	0.82 (3H, s)	16.1

Table S12: ^{1}H and ^{13}C NMR assignments for 2α-hydroxy-sandaracopimara-8(14),15-diene.

	2α-hydroxy-sandaracopimara-8(14),15-diene	
Position	δ_H	δ_C
1 a	2.00 (1H, m)	48.8
b	0.96 (1H, t, *J* = 11.8 Hz)	
2	3.81 (1H, tt, *J* = 11.6, 4.1 Hz)	65.6
3 a	1.73 (1H, m)	51.3
b	1.12 (1H, t, *J* = 11.6 Hz)	
4		35.2
5	0.99 (1H, dd, *J* = 12.5, 2.7 Hz)	54.3
6 a	1.56 (1H, m)	22.5
b	1.24 (1H, m)	
7 a	2.25 (1H, ddd, *J* = 14.0, 4.6, 1.8 Hz)	36.0
b	2.02 (1H, m)	
8		136.7
9	1.75 (1H, m)	50.8
10		40.0
11 a	1.60 (1H, m)	19.1
b	1.50 (1H, m)	
12 a	1.44 (1H, m)	34.7
b	1.34 (1H, td, *J* = 12.5, 3.6 Hz)	
13		37.6
14	5.22 (1H, brs)	129.5
15	5.75 (1H, dd, *J* = 17.6, 10.6 Hz)	149.2
16 a	4.88 (1H, dd, *J* = 17.6, 1.5 Hz)	110.3
b	4.86 (1H, dd, *J* = 10.6, 1.5 Hz)	
17	1.01 (3H, s)	26.2
18	0.87 (3H, s)	23.3
19	0.92 (3H, s)	34.0
20	0.81 (3H, s)	16.1

Table S13: ^{1}H and ^{13}C NMR assignments for 3α-hydroxy-sandaracopimara-8(14),15-diene.

	3β-hydroxy- sandaracopimara-8(14),15-diene	
Position	δ_H	δ_C
1	1.24 (2H, m)	31.8
2 a	1.88 (1H, m)	25.7
b	1.60 (1H, m)	
3	3.42 (1H, m)	76.3
4		37.2
5	1.45 (1H, m)	48.4
6 a	1.47 (1H, m)	22.4
b	1.33 (1H, m)	
7 a	2.25 (1H, m)	36.1
b	2.04 (1H, m)	
8		137.3
9	1.77 (1H, m)	50.6
10		38.1
11 a	1.60 (1H, m)	18.9
b	1.48 (1H, m)	
12 a	1.43 (1H, m)	34.8
b	1.34 (1H, m)	
13		37.4
14	5.20 (1H, brs)	128.9
15	5.76 (1H, dd, J = 17.6, 10.6 Hz)	149.2
16 a	4.88 (1H, dd, J = 17.6, 1.5 Hz)	110.2
b	4.86 (1H, dd, J = 10.6, 1.5 Hz)	
17	1.01 (3H, s)	26.2
18	0.86 (3H, s)	22.9
19	0.94 (3H, s)	28.8
20	0.79 (3H, s)	15.1

Figure S24A: [1]H Spectrum of *ent*-cassa-12,15-dien-19-ol.

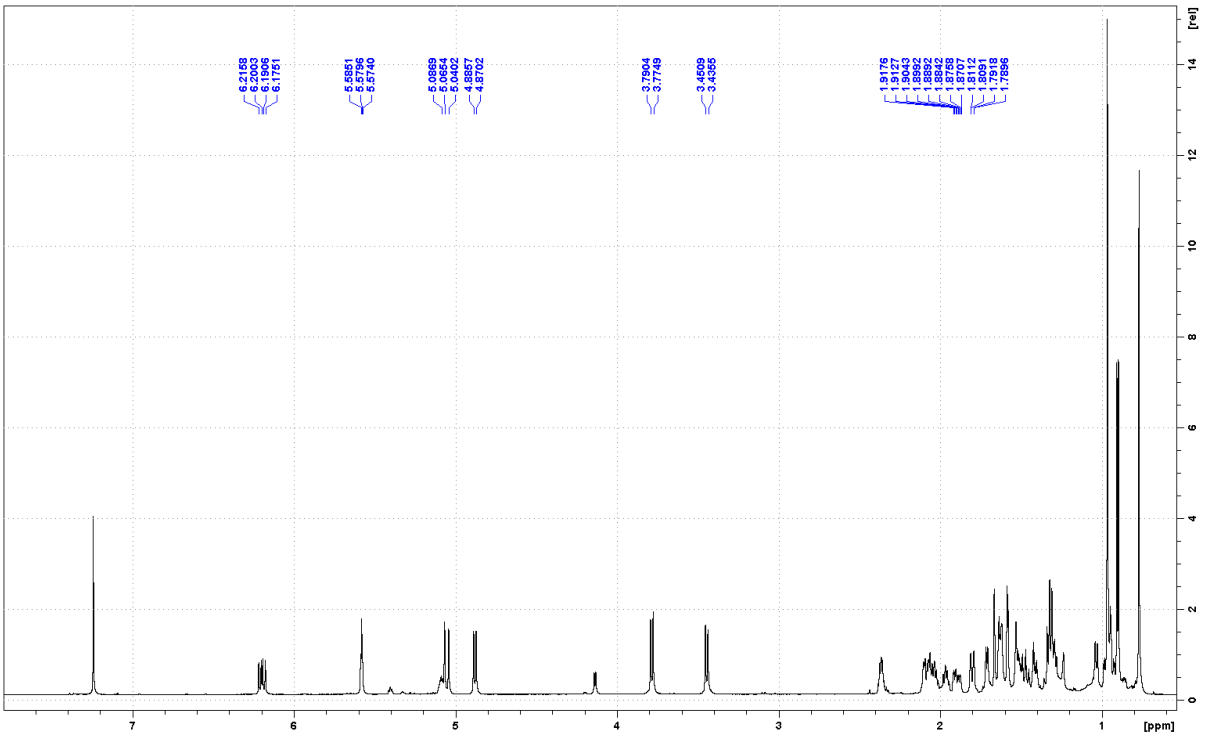


Figure S24B: [1]C Spectrum of *ent*-cassa-12,15-dien-19-ol.

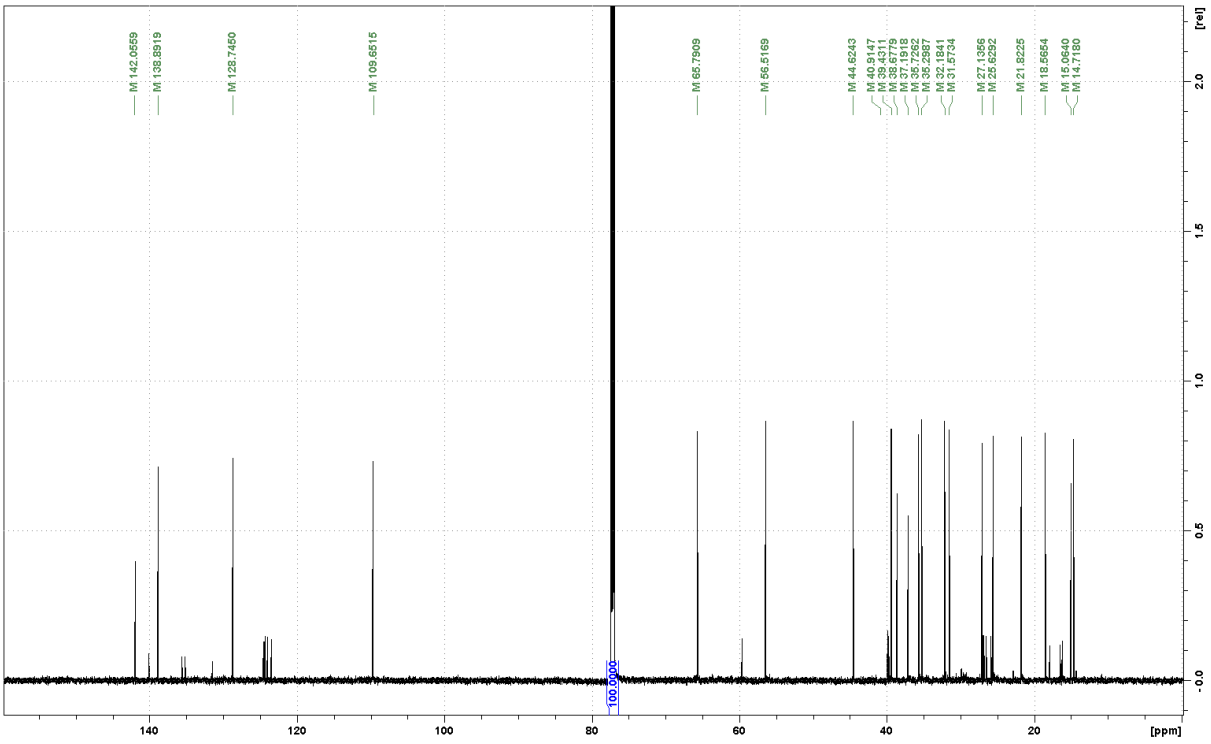


Figure S25A: ^{1}H Spectrum of *ent*-sandaracopimaradien-19-ol.

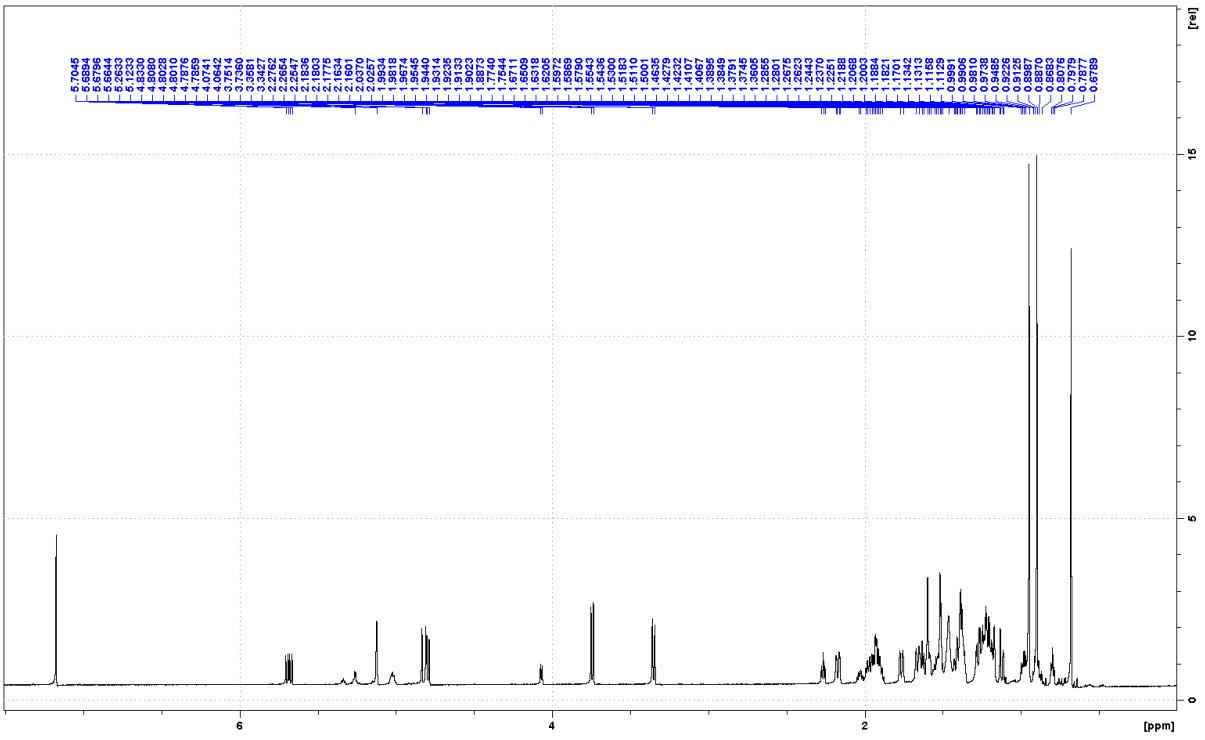


Figure S25B: ^{1}C Spectrum of *ent*-sandaracopimaradien-19-ol.

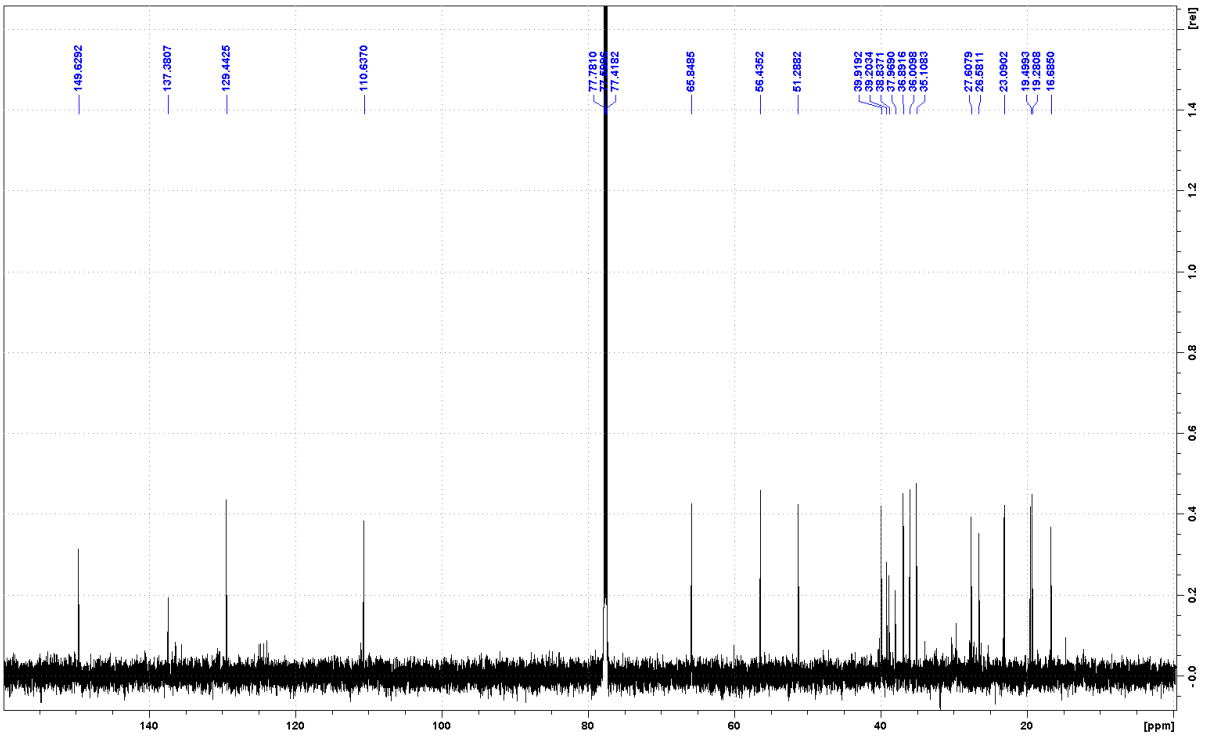


Figure S26A: ^{1}H Spectrum of 2α-hydroxy-*syn*-pimara-7,15-diene.

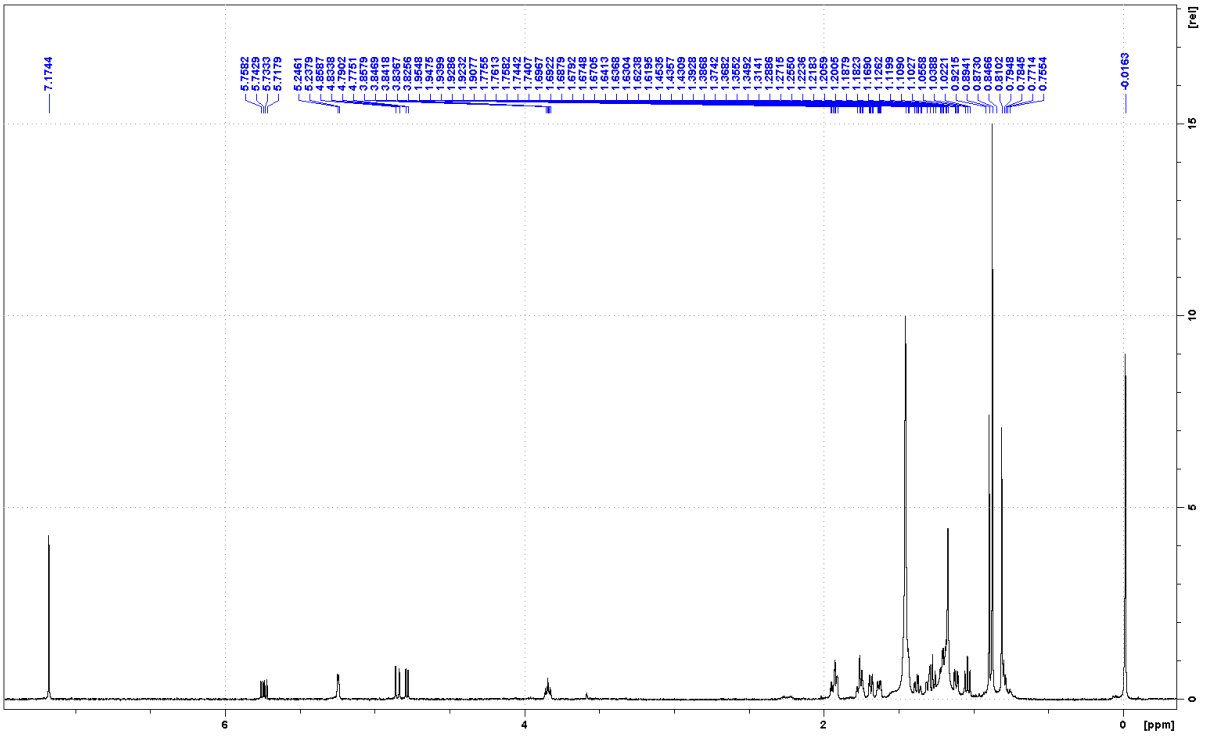


Figure S26B: ^{1}C Spectrum of 2α-hydroxy-*syn*-pimara-7,15-diene.

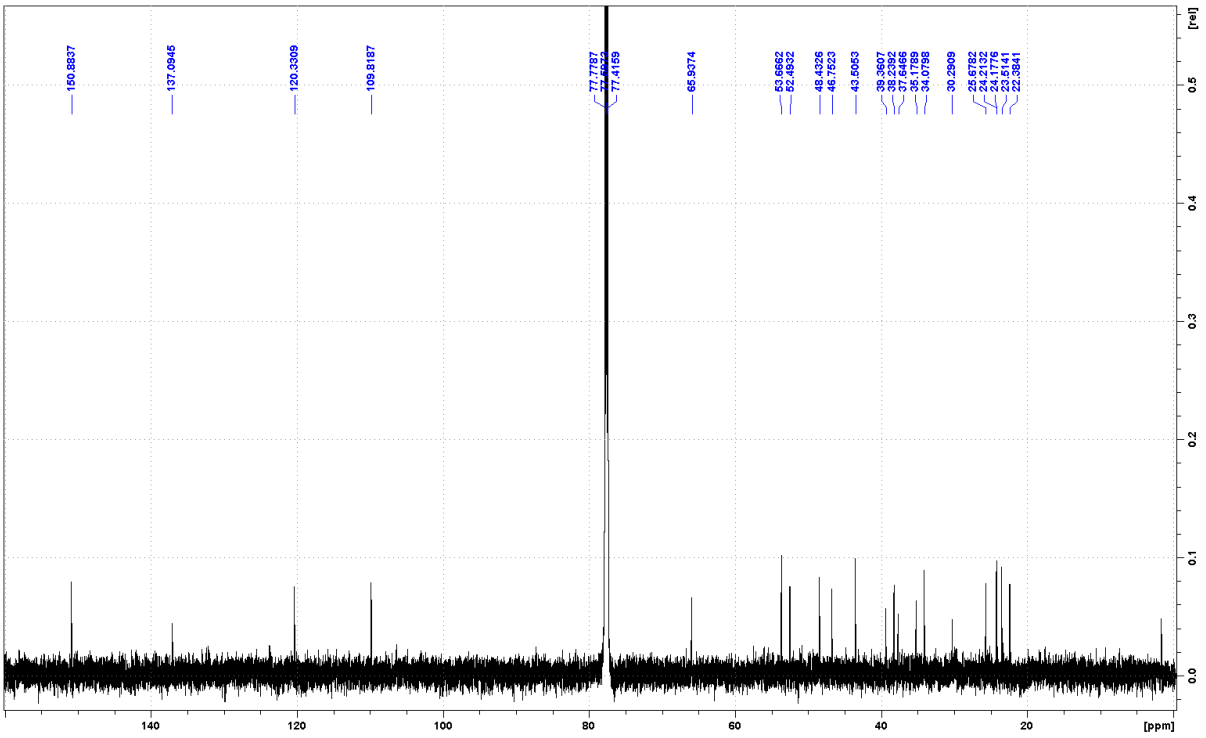


Figure S27A: [1]H Spectrum of 3β-hydroxy-*syn*-pimara-7,15-diene.

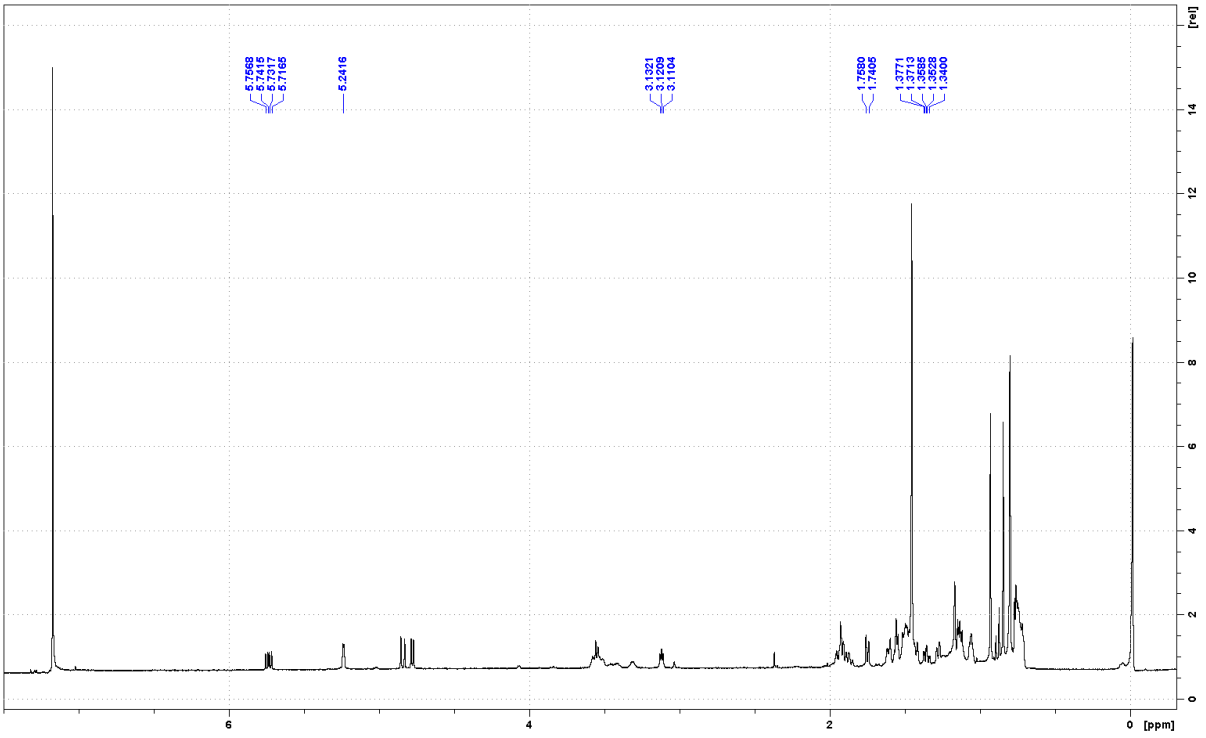


Figure S27B: [1]C Spectrum of 3β-hydroxy-*syn*-pimara-7,15-diene.

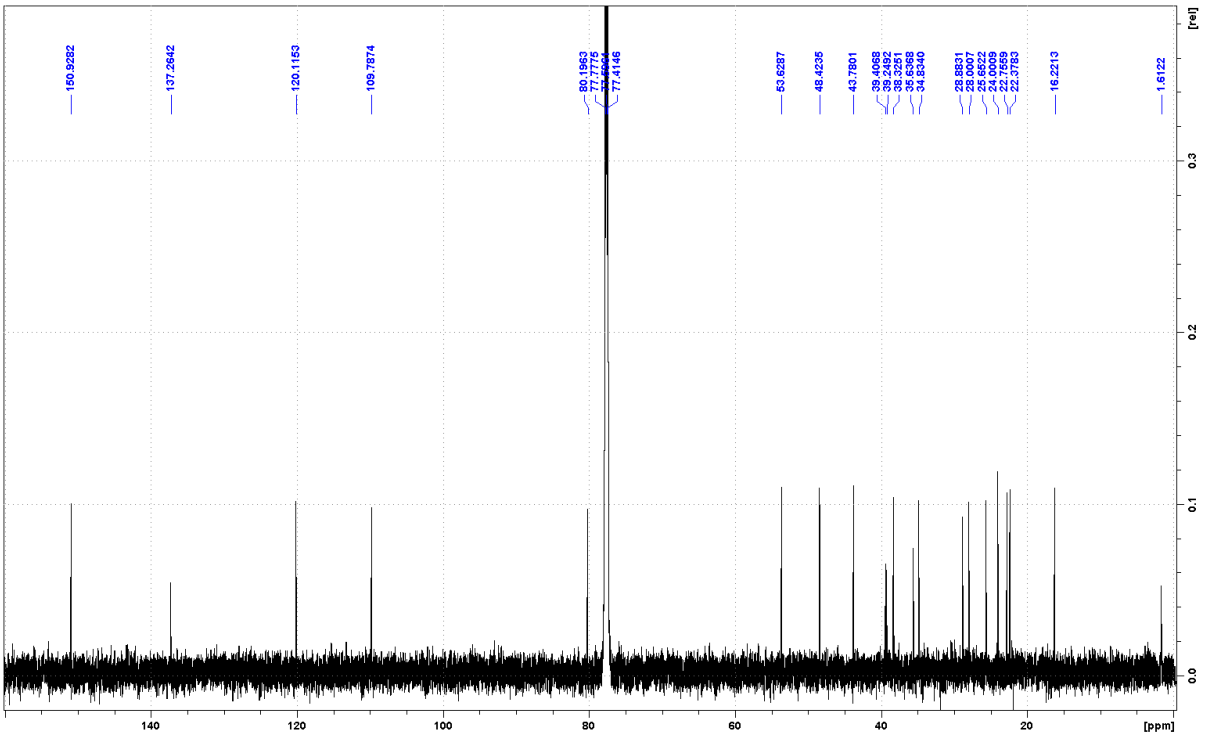


Figure S28A: ^{1}H Spectrum of 3β-hydroxy-*syn*-aphidicol-15-ene.

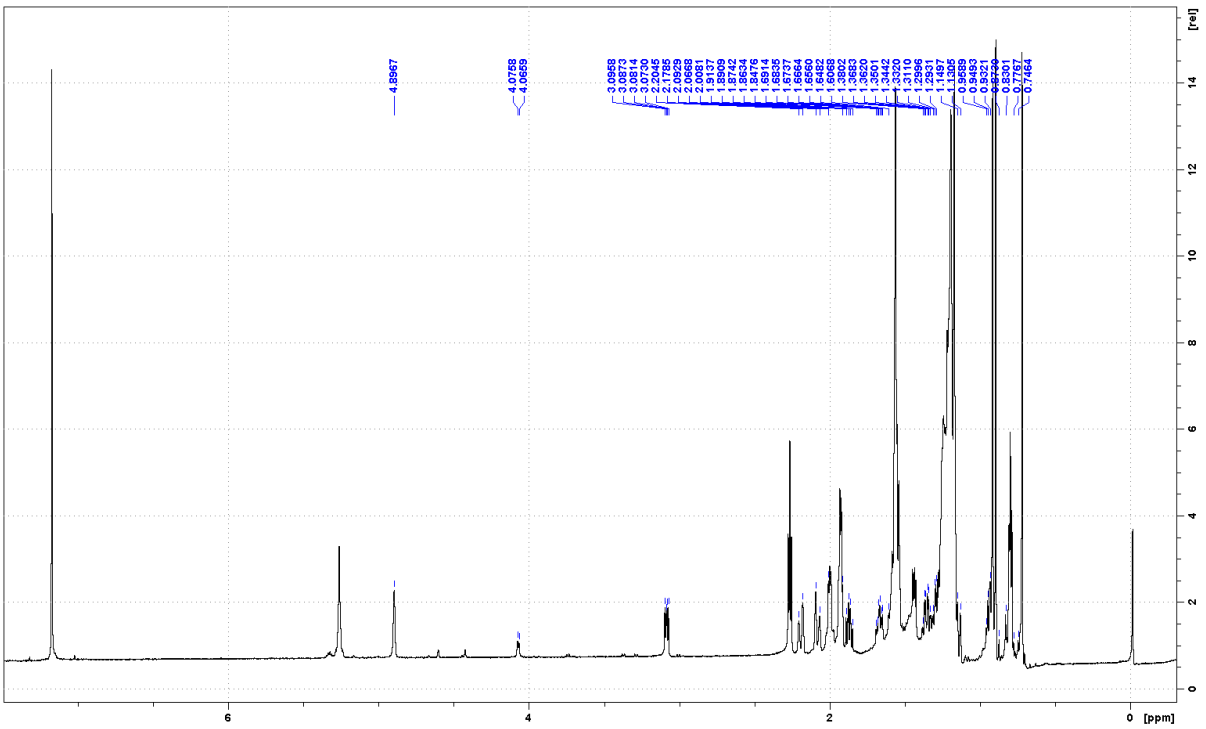


Figure S28B: ^{1}C Spectrum of 3β-hydroxy-*syn*-aphidicol-15-ene.

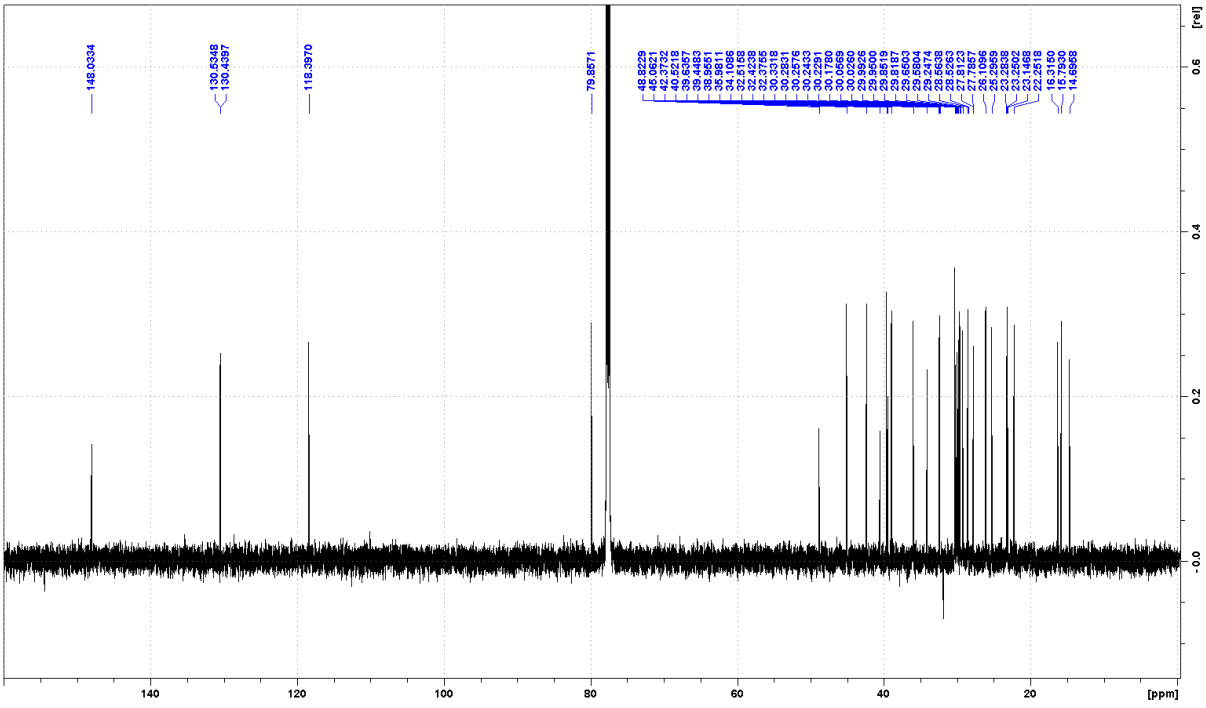


Figure S29A: ^{1}H Spectrum of 3β-hydroxy-*syn*-labda-8(17),12*E*,14-triene.

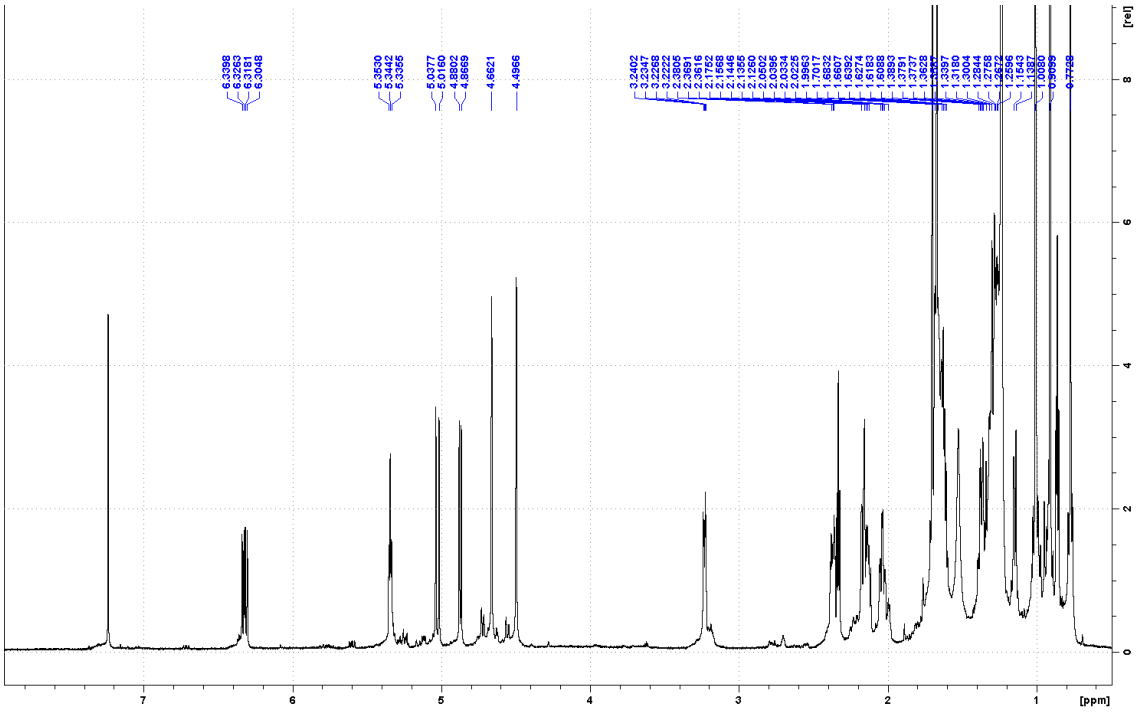


Figure S29B: ^{13}C Spectrum of 3β-hydroxy-*syn*-labda-8(17),12*E*,14-triene.

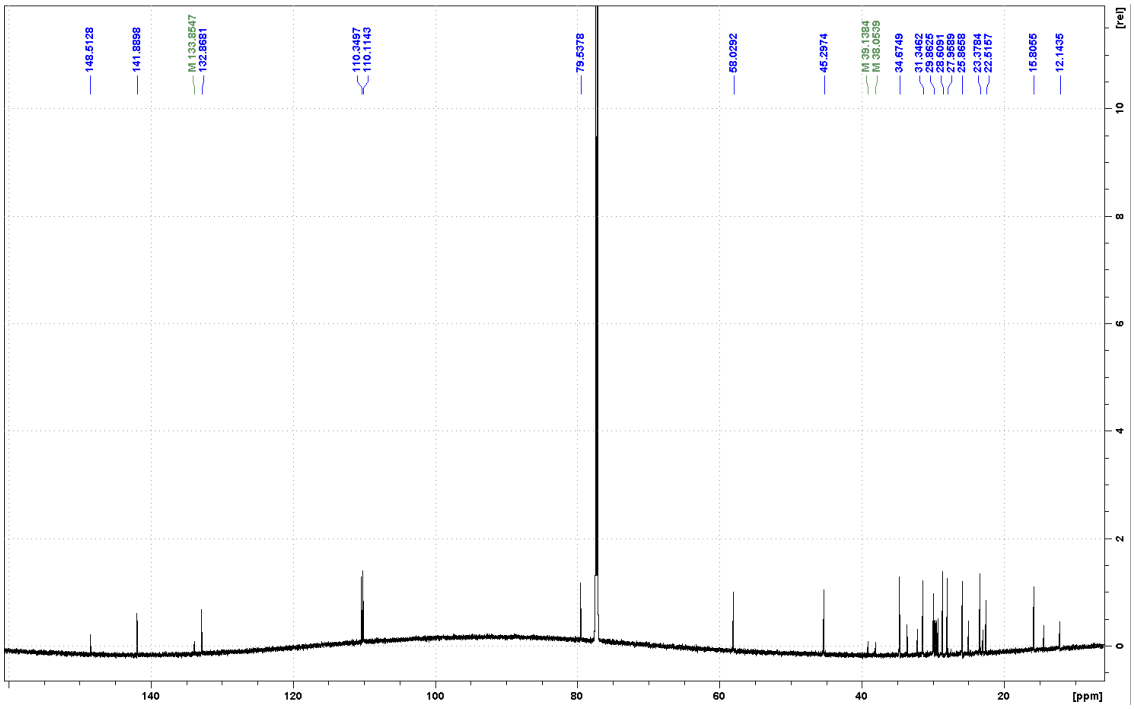


Figure S30A: ^{1}H Spectrum of 3β-hydroxy-*syn*-stemod-13(17)-ene.

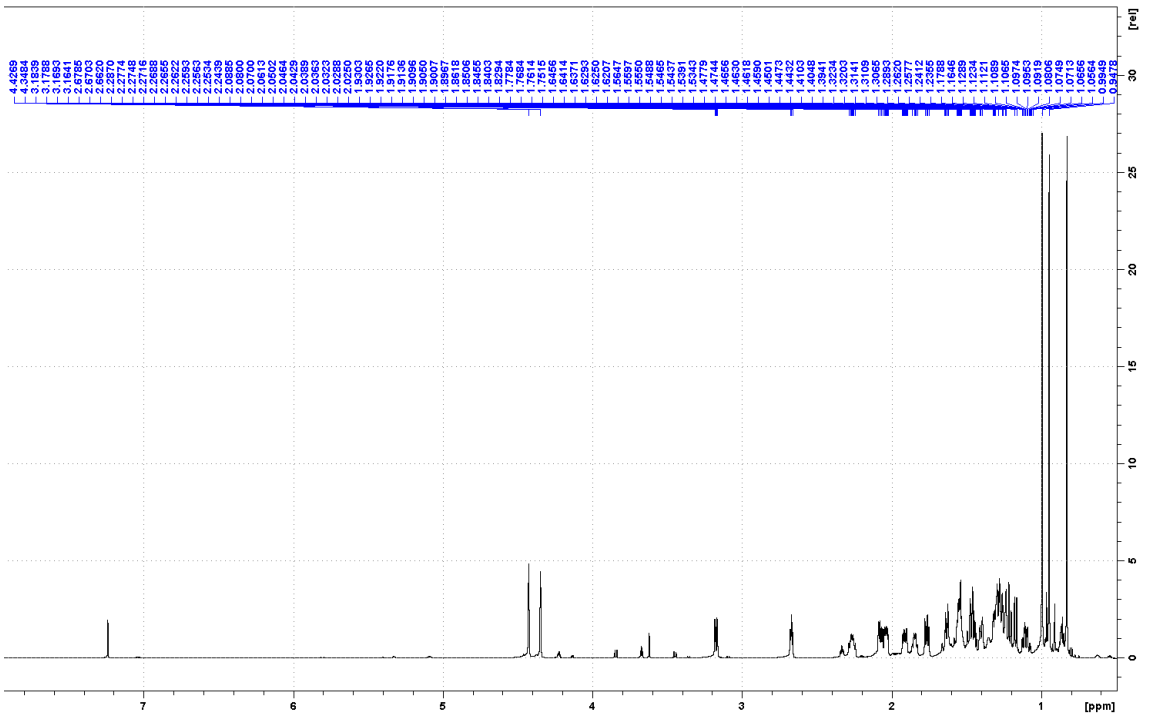


Figure S30B: ^{13}C Spectrum of 3β-hydroxy-*syn*-stemod-13(17)-ene.

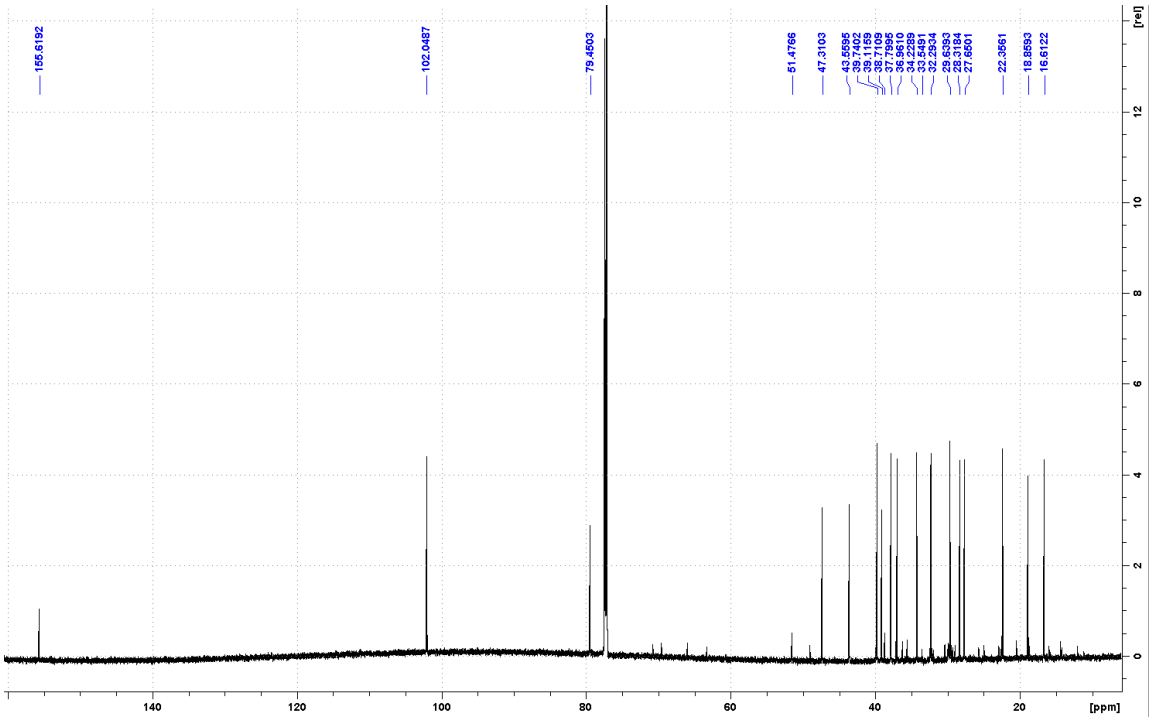


Figure S31A: [1]H Spectrum of normal 3β-hydroxy-isopimara-8,15-diene.

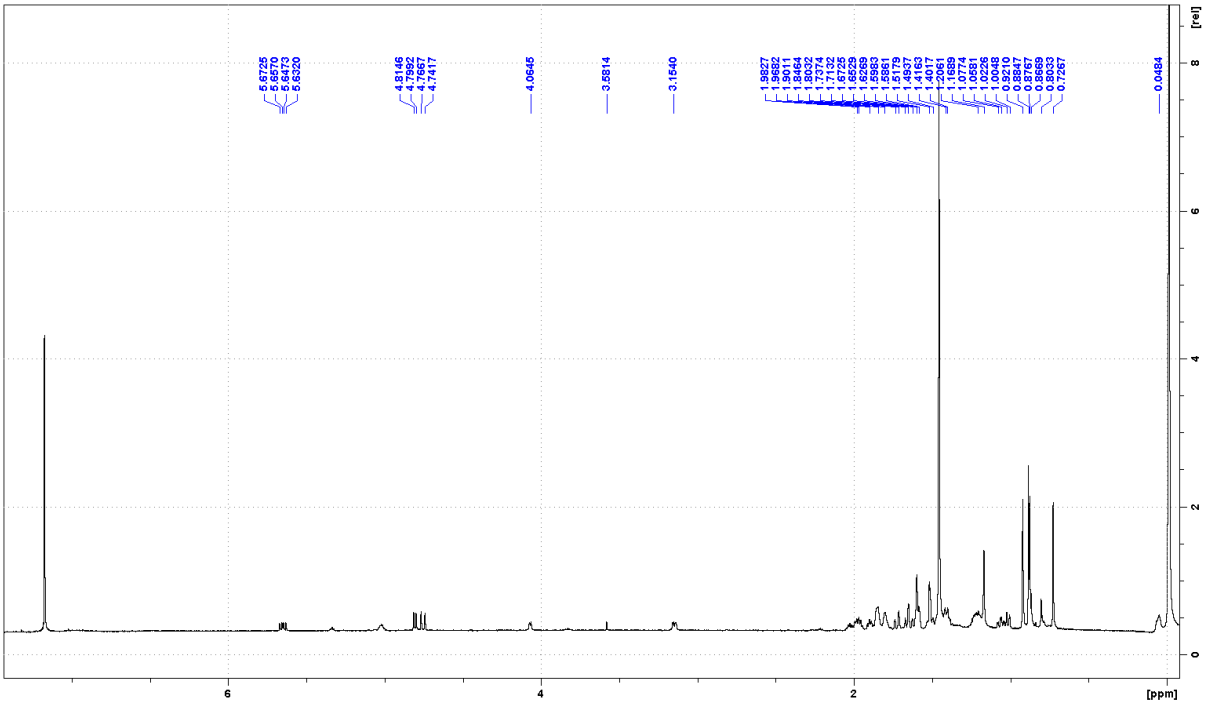


Figure S31B: [1]C Spectrum of 3β-hydroxy-isopimara-8,15-diene.

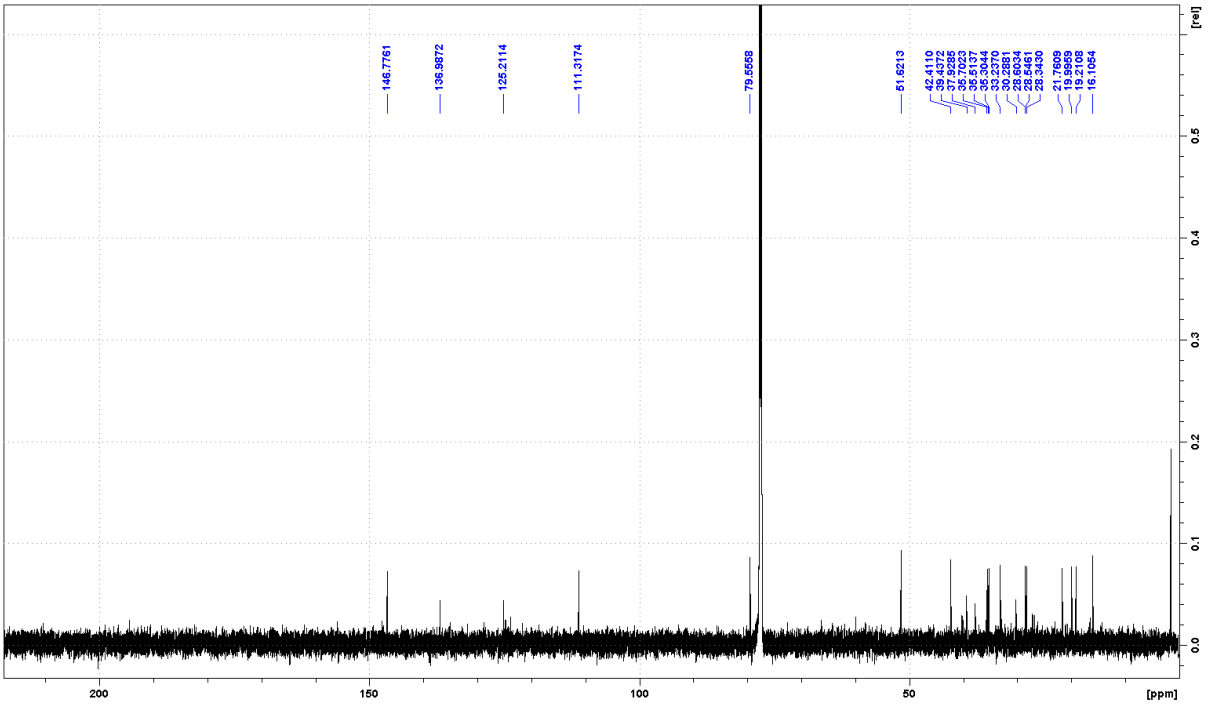


Figure S32A: [1]H Spectrum of 2α-hydroxy-isopimara-7,15-diene.

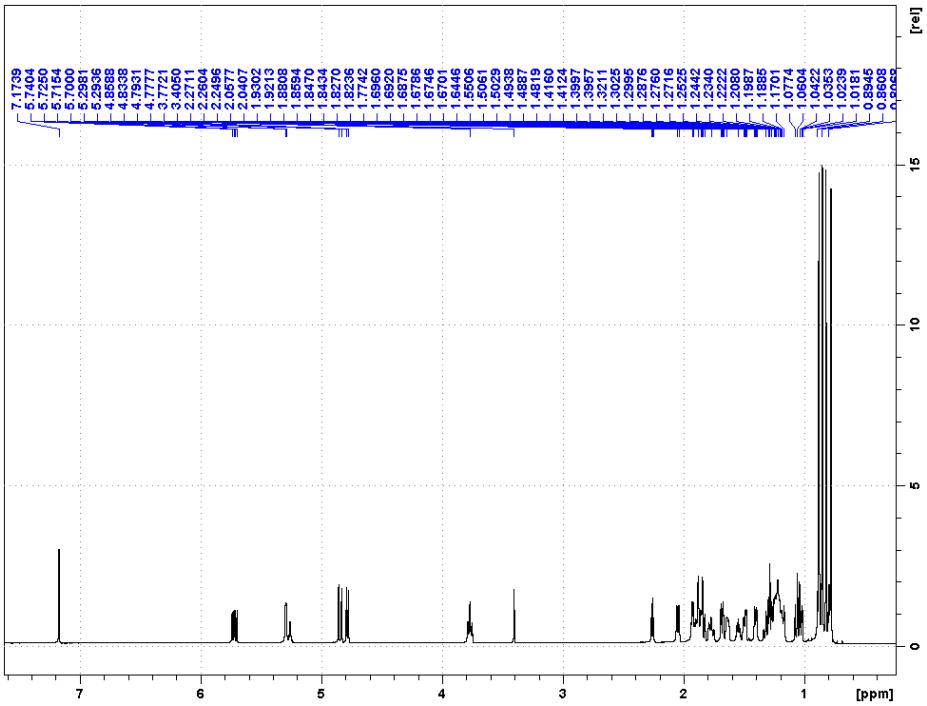


Figure S32B: [1]C Spectrum of 2α-hydroxy-isopimara-7,15-diene.

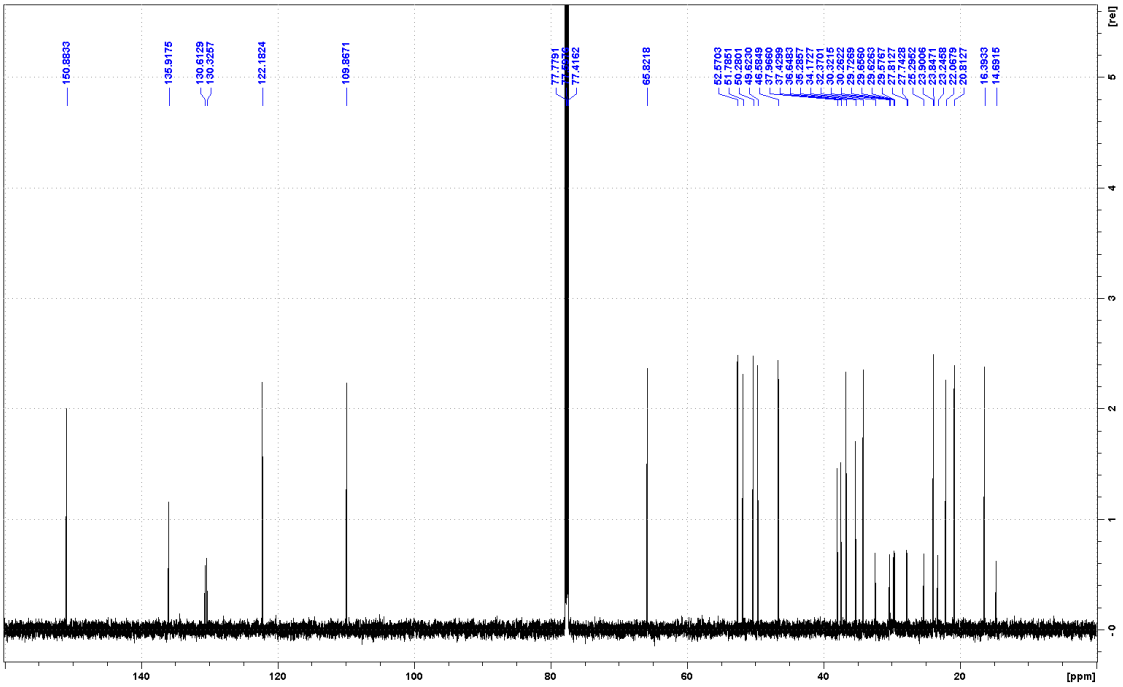


Figure S33A: ^{1}H Spectrum of isopimara-7,15-dien-2-one.

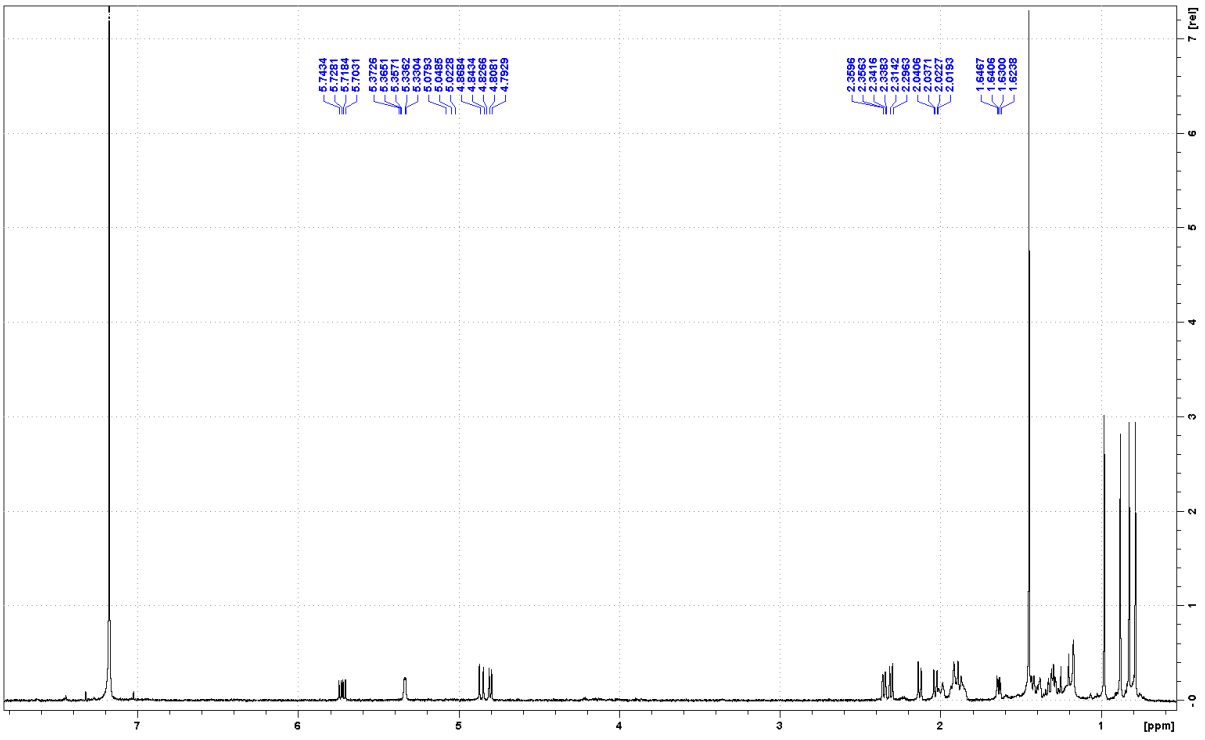


Figure S33B: ^{13}C Spectrum of isopimara-7,15-dien-2-one.

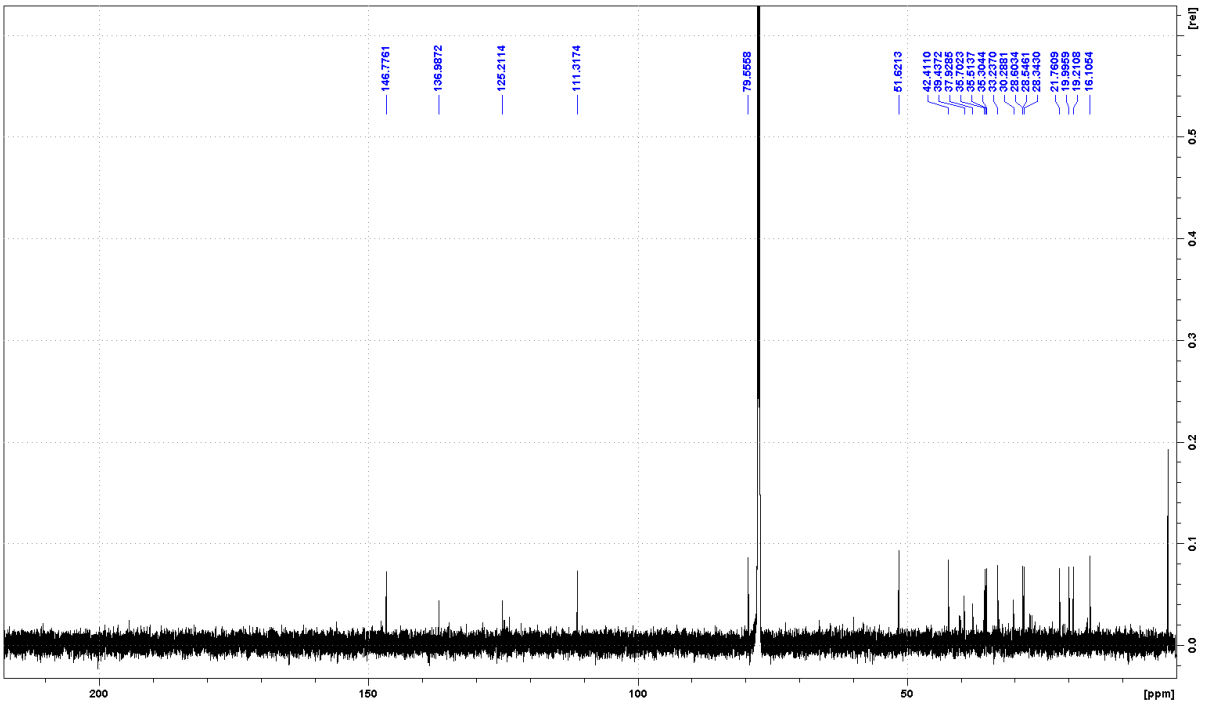


Figure S34A: ^{1}H Spectrum of 2α-hydroxy-sandaracopimara-8(14),15-diene.

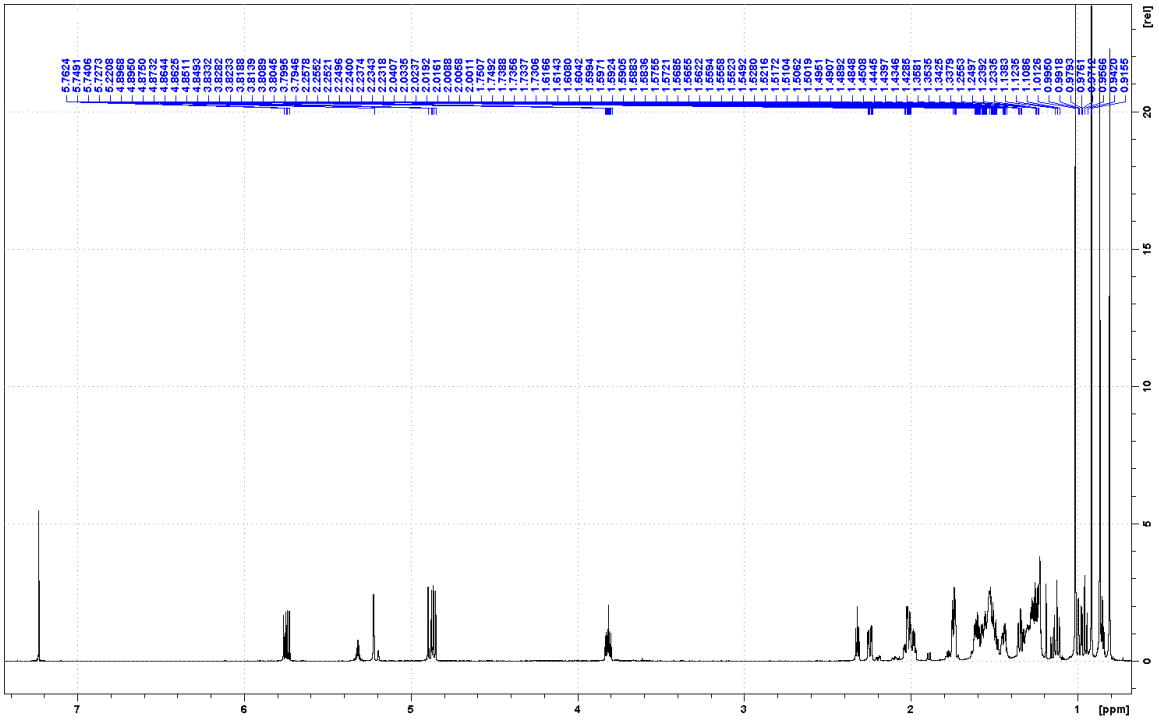


Figure S34B: ^{13}C Spectrum of 2α-hydroxy-sandaracopimara-8(14),15-diene.

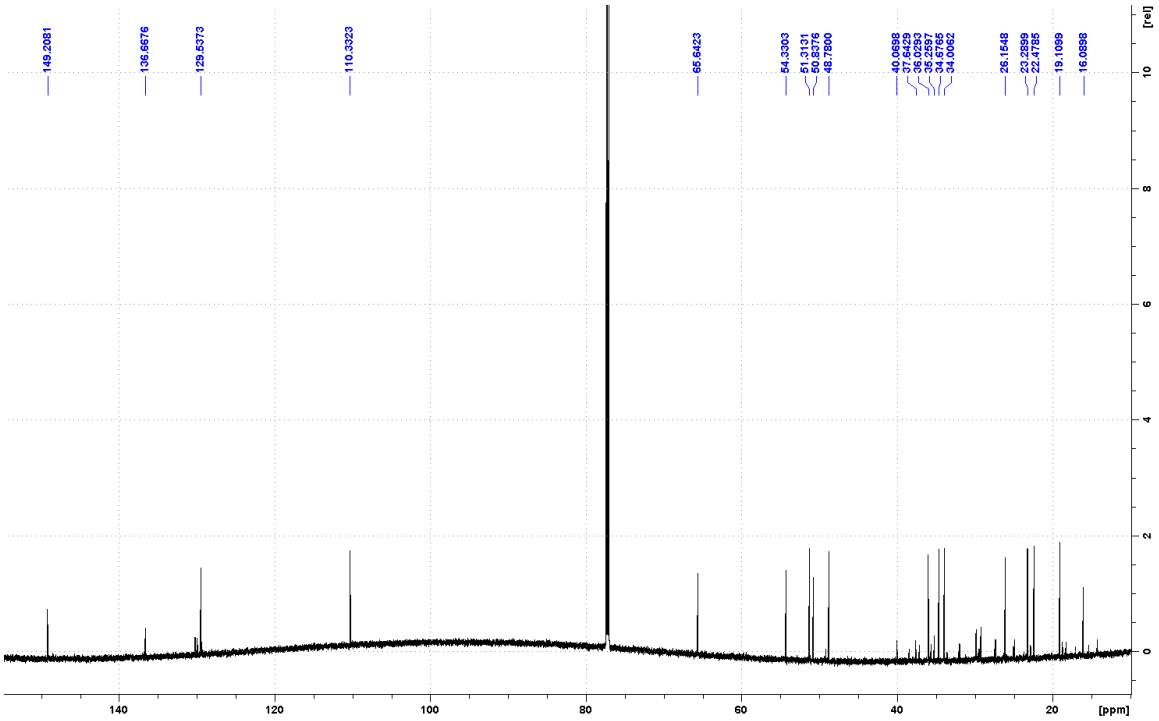


Figure S35A: ^{1}H Spectrum of 3β-hydroxy-sandaracopimara-8(14),15-diene.

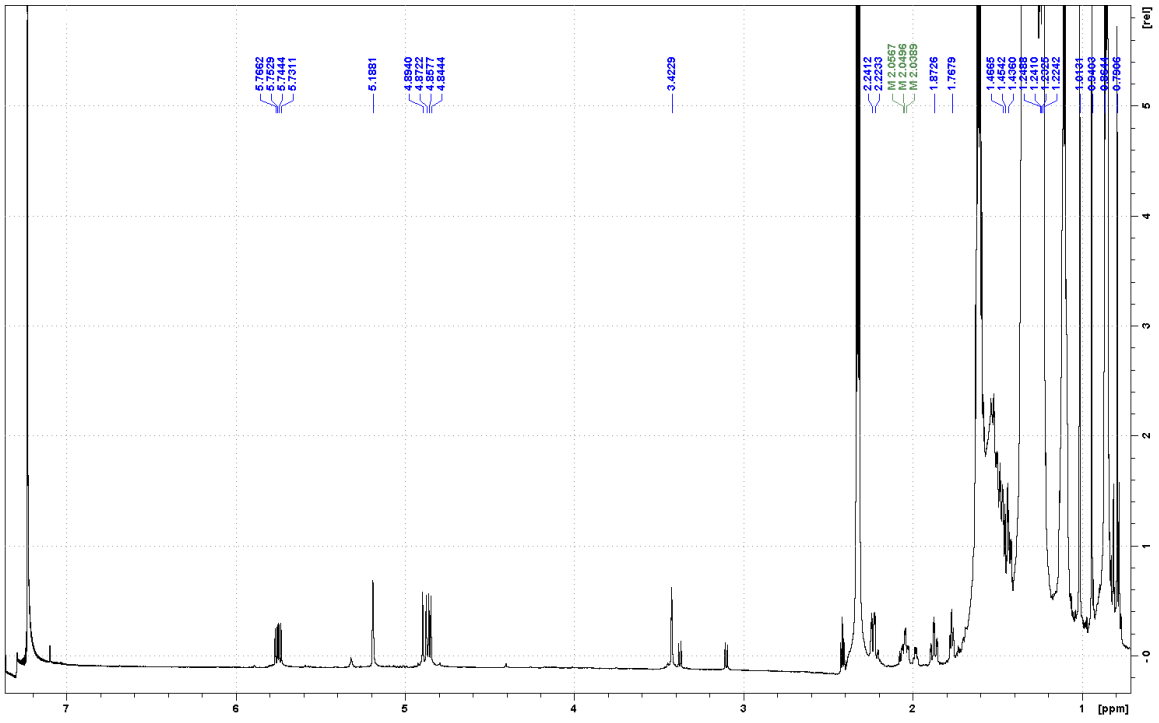


Figure S35B: ^{13}C Spectrum of 3β-hydroxy-sandaracopimara-8(14),15-diene.

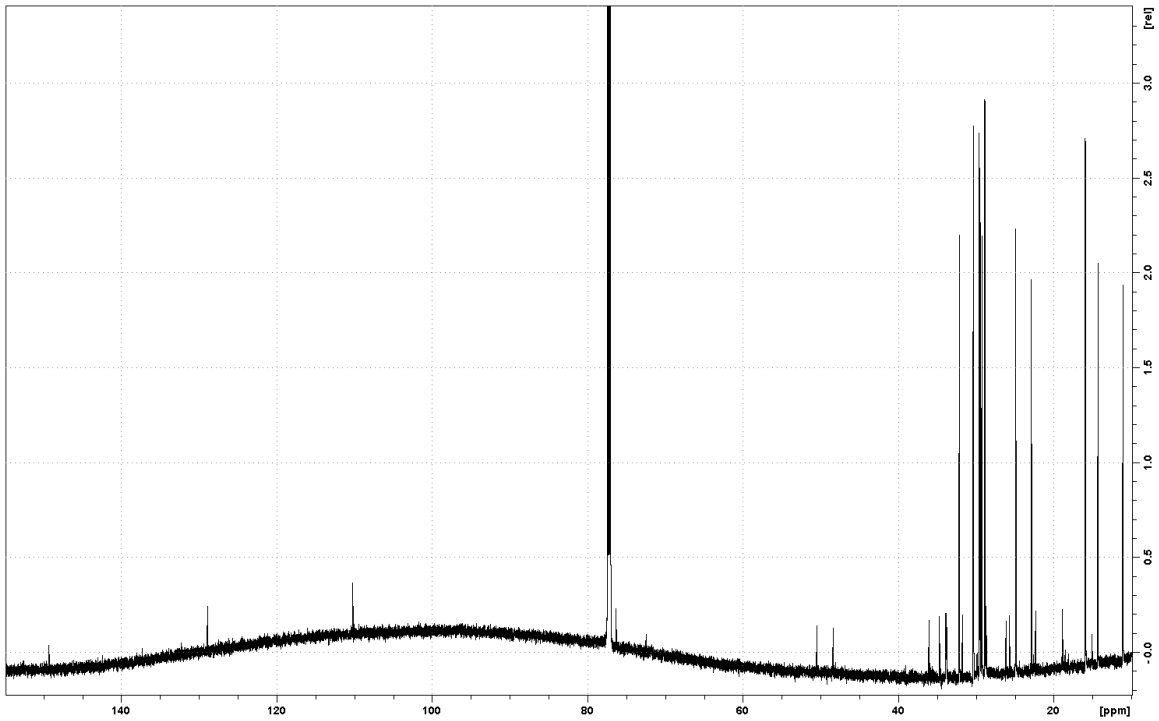


Table S14: Characterized CYP701 family members used in molecular phylogenetic analysis

CYP designation	*Species*	GenBank	Reference
CYP701A1	*Cucurbita maxima*	NM_122491	(26)
CYP701A3	*Arabidopsis thaliana*	AF212990	(27)
CYP701A5	*Stevia rebaudiana*	AY364317	(28)
CYP701A6	*Oryza sativa*	NM_001064444	(29)
CYP701A8	*Oryza sativa*	AY579214	(14)
CYP701A9	*Oryza sativa*	NM_001064438	(13)
CYP701A10	*Pisum sativum*	AY245442	(30)
CYP701A11	*Lactuca sativa*	AB370235	(31)
CYP701A12	*Lactuca sativa*	AB370236	(31)
CYP701A13	*Hordeum vulgare*	AK374293	(32)
CYP701A19	*Oryza sativa*	NM_001064442	(7)
CYP701B1	*Physcomitrella patens*	AB618673	(33)

References

1. Morrone D, Chen X, Coates RM, & Peters RJ (2010) Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis. *Biochem. J.* 431(3):337-344.
2. Swaminathan S, Morrone D, Wang Q, Fulton DB, & Peters RJ (2009) CYP76M7 is an *ent*-cassadiene C11α-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21:3315-3325.
3. Cyr A, Wilderman PR, Determan M, & Peters RJ (2007) A Modular Approach for Facile Biosynthesis of Labdane-Related Diterpenes. *J. Am. Chem. Soc.* 129:6684-6685.
4. Zhou K*, et al.* (2012) Functional characterization of wheat ent-kaurene(-like) synthases indicates continuing evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry* 84:47-55.
5. Xu M*, et al.* (2014) Characterization of an orphan diterpenoid biosynthetic operon from Salinispora arenicola. *Journal of natural products* 77(9):2144-2147.
6. Xu M*, et al.* (2007) Functional characterization of the rice kaurene synthase-like gene family. *Phytochemistry* 68(3):312-326.
7. Kitaoka N, Wu Y, Xu M, & Peters RJ (2015) Optimization of recombinant expression enables discovery of novel cytochrome P450 activity in rice diterpenoid biosynthesis. *Appl. Microbiol. Biotechnol.* 99(18):7549-7558.
8. Martin VJJ, Pitera DJ, Withers ST, Newman JD, & Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature biotechnology* 21(7):796-802.
9. Morrone D*, et al.* (2010) Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl. Microbiol. Biotechnol.* 85:1893-1906.
10. Tamura K, Stecher G, Peterson D, Filipski A, & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30(12):2725-2729.
11. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586-1591.
12. Bielawski JP (2013) Detecting the signatures of adaptive evolution in protein-coding genes. *Curr Protoc Mol Biol* Chapter 19:Unit 19 11.
13. Itoh H*, et al.* (2004) A rice semi-dwarf gene, *Tan-Ginbozu* (*D35*), encodes the gibberellin biosynthesis enzyme, *ent*-kaurene oxidase. *Plant Mol Biol* 54:533-547.
14. Wang Q, Hillwig ML, Wu Y, & Peters RJ (2012) CYP701A8: A rice *ent*-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. *Plant Physiol.* 158(3):1418-1425.
15. Morrone D*, et al.* (2009) Gibberellin biosynthesis in bacteria: Separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. *FEBS Lett.* 583(2):475-480.
16. Hayashi K*, et al.* (2006) Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*. *FEBS Lett.* 580(26):6175-6181.
17. Jackson AJ, Hershey DM, Chesnut T, Xu M, & Peters RJ (2014) Biochemical characterization of the castor bean *ent*-kaurene synthase(-like) family supports quantum chemical view of diterpene cyclization. *Phytochemistry* 103:13-21.

18. Morrone D*, et al.* (2011) Evident and latent plasticity across the rice diterpene synthase family with potential implications for the evolution of diterpenoid metabolism in the cereals. *Biochem. J.* 435:589-595.
19. Morrone D*, et al.* (2006) An unexpected diterpene cyclase from rice: Functional identification of a stemodene synthase. *Arch. Biochem. Biophys.* 448(1-2):133-140.
20. Wilderman PR, Xu M, Jin Y, Coates RM, & Peters RJ (2004) Identification of *syn*-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiol.* 135(4):2098-2105.
21. Morrone D, Xu M, Fulton DB, Determan MK, & Peters RJ (2008) Increasing complexity of a diterpene synthase reaction with a single residue switch. *J. Am. Chem. Soc.* 130:5400-5401.
22. Gao W*, et al.* (2009) A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *Org. Lett.* 11(22):5170-5173.
23. Keeling CI, Madilao LL, Zerbe P, Dullat HK, & Bohlmann J (2011) The primary diterpene synthase products of Picea abies levopimaradiene/abietadiene synthase (PaLAS) are epimers of a thermally unstable diterpenol. *J. Biol. Chem.* 286(24):21145-21153.
24. Mafu S, Hillwig ML, & Peters RJ (2011) A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from *Selaginella moellendorffii*. *ChemBioChem* 12(13):1984-1987.
25. Nakano C*, et al.* (2010) Substrate specificity of the CYC2 enzyme from *Kitasatospora griseola*: production of sclarene, biformene, and novel bicyclic diterpenes by the enzymatic reactions of labdane-and halimane-type diterpene diphosphates. *Tetrahedron Lett.* 51:125-128.
26. Helliwell CA*, et al.* (2000) Isolation of an *ent*-kaurene oxidase cDNA from *Curcurbita maxima*. *Aust. J. Plant Physiol.* 27:1141-1149.
27. Helliwell CA*, et al.* (1998) Cloning of the *Arabidopsis ent*-kaurene oxidase gene GA3. *Proc. Natl. Acad. Sci., U.S.A.* 95(15):9019-9024.
28. Humphrey TV, Richman AS, Menassa R, & Brandle JE (2006) Spatial organisation of four enzymes from Stevia rebaudiana that are involved in steviol glycoside synthesis. *Plant Mol Biol* 61(1-2):47-62.
29. Cao R*, et al.* (2010) Diterpene cyclases and the nature of the isoprene fold. *Proteins* 78(11):2417-2432.
30. Davidson SE, Elliot RC, Helliwell CA, Poole AT, & Reid JB (2003) The pea gene NA encodes ent-kaurenoic acid oxidase. *Plant Physiol.* 131(1):335-344.
31. Sawada Y*, et al.* (2008) Germination of photoblastic lettuce seeds is regulated via the control of endogenous physiologically active gibberellin content, rather than of gibberellin responsiveness. *J. Exp. Bot.* 59(12):3383-3393.
32. Spielmeyer W*, et al.* (2004) Isolation of gibberellin metabolic pathway genes from barley and comparative mapping in barley, wheat and rice. *Theor. Appl. Genet.* 109:847-855.
33. Miyazaki S, Katsumata T, Natsume M, & Kawaide H (2011) The CYP701B1 of *Physcomitrella patens* is an *ent*-kaurene oxidase that resists inhibition by uniconazole-P. *FEBS Lett.* 585(12):1879-1883.