Supplemental Information

Materials

E. coli DH5α was obtained from the University of Illinois at Urbana-Champaign Biochemistry Department's Media Preparation Facility (Urbana, IL). *Hypericum perforatum* var. Topas seeds were purchased from Sand Mountain Herbs (Fyffe, AL). *Eucalyptus robusta* and *Eucalyptus camaldulensis* var. Silverton Province plants were obtained from Windmill Outback Nursery (Louisa, VA). Plasmid pET26b(+) was obtained from Novagen (Madison, WI). All restriction endonucleases, as well as T4 DNA ligase and antarctic phosphatase, were purchased from New England Biolabs (Beverly, MA). The QIAprep Spin Plasmid Mini-prep Kit, QIAquick PCR Purification Kit, QIAquick Gel Extraction Kit, and RNeasy MiniElute Cleanup Kit were purchased from Qiagen (Valencia, CA). The Total RNA Isolation Mini Kit was obtained from Agilent Technologies (Santa Clara, CA). All primers were synthesized by Integrated DNA Technologies (Coralville, IA). 5 PRIME MasterMix was purchased from 5 PRIME Inc. (Gaithersburg, MD). Unless otherwise noted, all other reagents were purchased from Sigma-Aldrich (St. Louis, MO).

Degenerate Primer Design

The mRNA sequences of fourteen enzymes in the plant Type III PKS family were used as the templates to design degenerate primers: *Hypericum androsoemum* benzophenone synthase (BPS), *Hordeum vulgare* homoeriodictyol/eriodictyol synthase (HEDS), *Medicago sativa* chalcone synthase 2 (CHS2), *Arachis hypogea* stilbene synthase (STS), *Pinus sylvestris* stilbene synthase (STS), *Gerbera hybrida* 2-pyrone synthase (2-PS), *Aloe arborescens* octaketide synthase (OKS), *Aloe arborescens* pentaketide chromone synthase (PCS), *Rheum palmatum* aloesone synthase (ALS), *Rheum palmatum* benzalacetone synthase (BAS), *Humulus lupulus* isovalerophenone synthase (VPS), *Hydrangea macrophylla* stilbenecarboxylate synthase 1 (STCS1), *Hydrangea macrophylla* coumaroyl triacetic acid lactone synthase (CTAS), and *Huperzia serrata* acridone synthase (ACS). The phylogenetic relationship of these fourteen enzymes and their main polyketide products are shown in **Figure S1**.

The mRNA sequences of the fourteen above enzymes were aligned using the ClustalW tool in the Biology Workbench suite (http://workbench.sdsc.edu).[1] Highly similar regions were identified from the alignment, and the six regions chosen to create degenerate primers are mapped onto the *M. sativa* CHS2 mRNA and protein sequences in **Figure S2**. The conserved regions were denoted F1, F2, F3, R1, R2, and R3, and used to design primers of the same name (see **Figure S3** and **Figure S4**). To keep degeneracy as low as possible, less conserved bases were replaced with deoxyinosine, with a maximum of three deoxyinosine nucleotides per primer. The primer sequences, their levels of degeneracy, and definitions of mixed bases are given in **Table S1**. Three sets of primers were designed to perform the nested PCR, which was essential because repeated use of F1-R3 cannot amplify desired fragment and using the inner primer set F3-R1 initially cannot generate enough amounts of the desired fragment either. Restriction sites were added to primers PKS-F3 (*Nhe*I) and PKS-R1 (*Xho*I) to facilitate cloning the smallest fragments into a vector for sequencing; those primers are PKS-F3-F and PKS-R1-F. Primers both with and without flanking bases were generated because it was unclear whether non-hybridizing, flanking nucleotides would inhibit the amplification reaction. Based on the *M. sativa* CHS2 mRNA (**Figure S2**), the fragment amplified by PKS-F1 and PKS-R3 is 803 nt in length; the fragment amplified by PKS-F2 and PKS-R2 is 776 nt; and the fragment amplified by PKS-F3(-F) and PKS-R1(-F) is 407 nt.

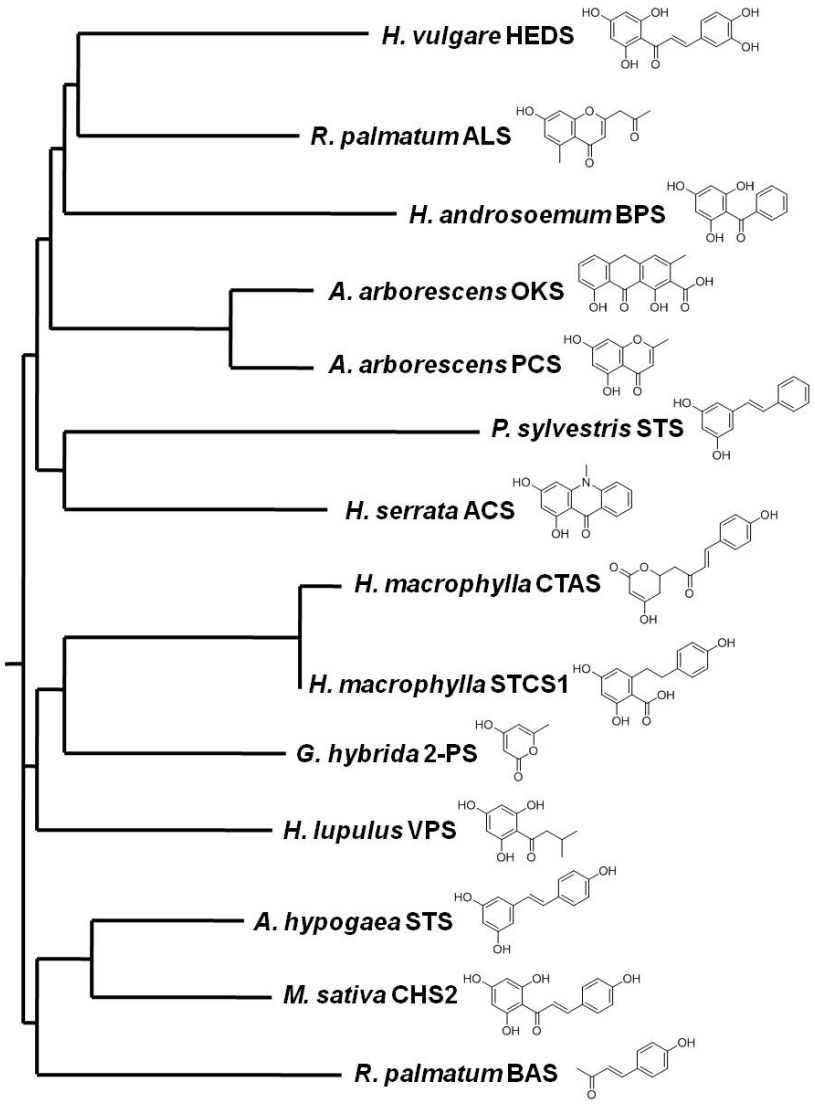

Figure S1 Phylogenetic analysis of the fourteen Type III PKS enzymes used to create degenerate primers in this study. The primary products of the enzymes are shown. The enzymes are, in order: HEDS, homoeriodictyol/eriodictyol synthase; ALS, aloesone synthase; BPS, benzophenone synthase; OKS, octaketide synthase; PCS, pentaketide chromone synthaes; STS, stilbene synthase; ACS, acridone synthase; CTAS, coumaroyl triacetic acid synthase; STCS1, stilbenecarboxylate synthase 1; 2-PS, 2-pyrone synthase; VPS, isovalerophenone synthase; CHS2, chalcone synthase 2; BAS, benzalacetone synthase.

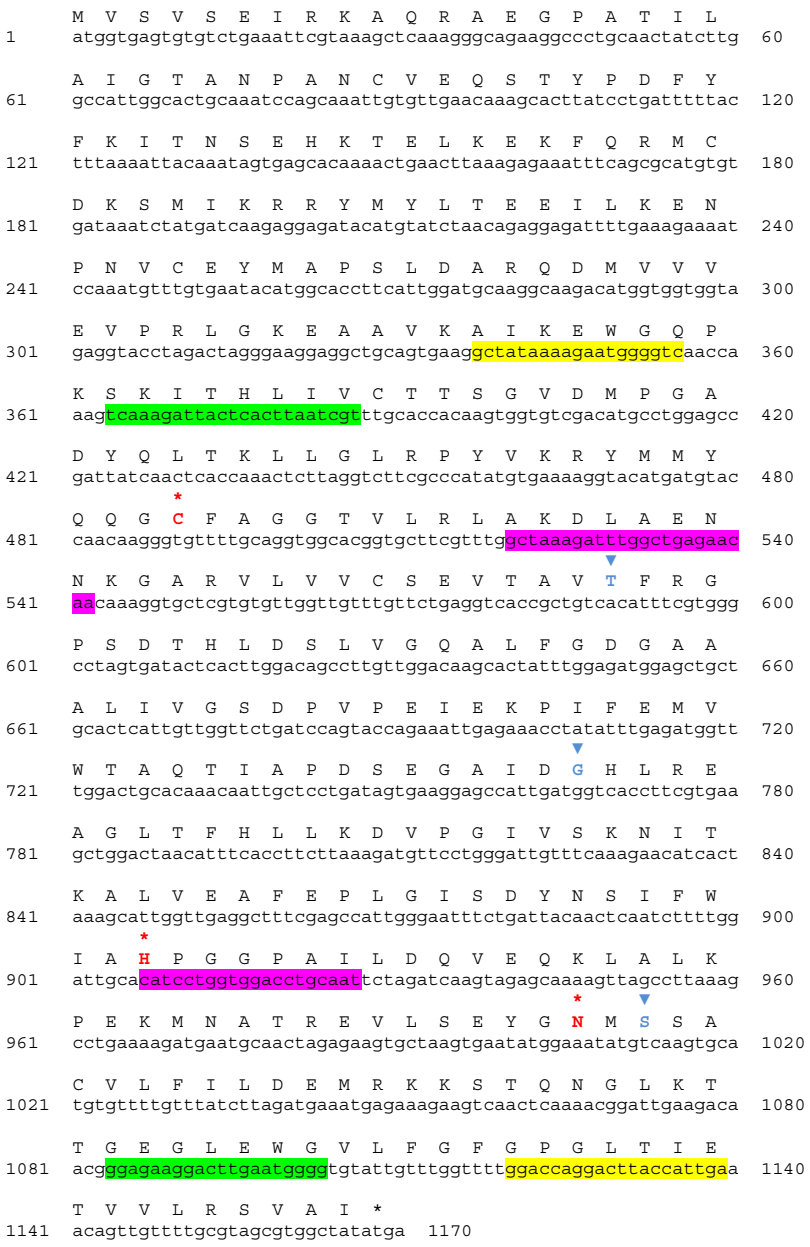

Figure S2 Locations of the three sets of conserved regions used to create degenerate primers, mapped onto the *M. sativa* CHS2 gene and protein, are shown in yellow, green, and pink, corresponding to F1, F2, F3, R1, R2, and R3, respectively. Active site residues Cys-152, His-305, and Asn-338 are shown in red and marked with asterisks. Residues which are known to affect substrate preference and elongation number, Thr-197, Gly-256, and Ser-338, are shown in blue and marked with a triangle.

Figure S3 Genetic diversity, consensus sequence, and primers designed at conserved regions F1, F2, and F3. In PKS-F3, italicized bases are extra flanking bases and the underlined bases are the *Nhe*I restriction site.

PKS R3 5′ TCIAYNGTNADICCIGGNCC 3′

Figure S4 Genetic diversity, consensus sequence, and primers designed at conserved regions R1, R2, and R3. In PKS-R1-F, italicized bases are extra flanking bases and the underlined bases are the *Xho*I restriction site.

Table S1 DNA sequence and degeneracy of primers used in this study.

Primer	Sequence, 5' → 3'	Degeneracy
PKS-F1	GCIATIVMDSARTGGGGNC	288
PKS-F2	TCIVRRATIACICAYBTNRT	576
PKS-F3	GYIAARGAYITNGGNGARAACAA	256
PKS-F3-F*	*GCGTTC*GCTAGCGYIAARGAYITNGGNGARAACAA	256
PKS-R1	ATNSCISKICCICCRGGRTG	128
PKS-R1-F**	*GCGTTC*CTCGAGATNSCISKICCICCRGGRTG	128
PKS-R2	CCSMWITCIWDNCCITCNCC	768
PKS-R3	TCIAYNGTNADICCIGGNCC	384

[±]Mixed bases B: C, G, T; D: A, G, T; H: A, C, T; K: G, T; M: A, C; N: A, C, G, T; R: A, G; S: C, G; V: A, C, G; W: A, T ; Y: C, T.

*Additional flanking bases are in italics; the *Nhe*I restriction site is underlined.

**Additional flanking bases are in italics; the *Xho*I restriction site is underlined.

Overview of the PCR-based Method for Discovering New Type III PKS Genes

An overview of the PCR-based method for discovering new Type III PKS genes is shown in **Figure S5**. Plant tissue was collected, total RNA extracted, and cDNA prepared as described below. The cDNA library was then used as the template for a nested PCR amplification (primers PKS-F1 and PKS-R3; PKS-F2 and PKS-R2; and PKS-F3-F and PKS-R1-F). In parallel, the cDNA library was used as a template for amplification of a ~1000 nt fragment of the gene encoding RuBisCo (Fwd primer 5' – CTTGGCAGCATTCCGAGTAACTCC – 3'; Rev primer 5' – CGGTCAGA GCAGGCATATGC – 3') as a control for the presence of mRNA. The ~400 nt fragments obtained from the final nested PCR were ligated into pET26b(+). The fragments were sequenced using primers specific to the T7 terminator in pET26b(+). Finally, translation of the resulting ~400 nt sequence gave a ~130 amino acid sequence of the Type III PKS enzyme.

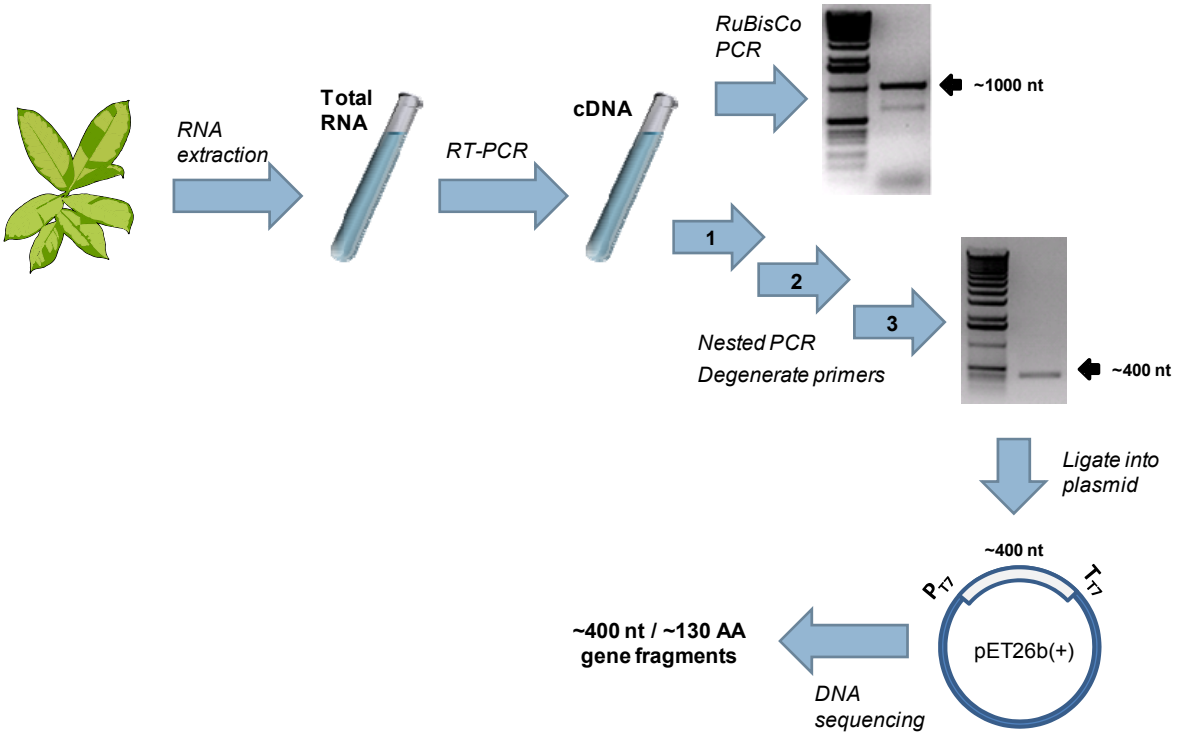

Figure S5 Overview of the PCR-based method for discovering new Type III PKS genes.

Collection of Plant Tissue

Plant tissue (leaf) was cut from live plants and immediately frozen on liquid nitrogen to prevent degradation of RNA. Frozen tissue was crushed to fine powder using a mortar and pestle cooled to freezing using liquid nitrogen. Crushed plant tissue was stored at -80 °C until RNA extraction.

Total RNA Isolation and Concentration and cDNA Synthesis

Total RNA was purified from crushed plant tissue using the Agilent Total RNA Isolation Mini Kit according to the manufacturer's instructions. Approximately 100 mg of plant tissue was used in each RNA purification. To purify and concentrate RNA, the Qiagen RNeasy MinElute Cleanup Kit was used according to the manufacturer's instructions. Total RNA was visualized by agarose gel electrophoresis and concentration and purity were determined using a NanoDrop (Thermo Fisher Scientific, Waltham, MA). Purified, concentrated total RNA was used as the template for cDNA synthesis using Transcriptor II reverse transcriptase enzyme (Invitrogen, Carlsbad, CA) and anchored Oligo-dT_{20} primer (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

Screen Validation Using *Hypericum perforatum* var. Topas

Before germinating, *H. perforatum* var. Topas seeds were sterilized by a four-step process: (1) seeds were washed in a solution of 70% ethanol and 0.1% Tween-20 with vigorous mixing for 1 minute; (2) seeds were washed in a solution of 5% NaOCl (bleach) with vigorous mixing for 10 minutes; (3) seeds were then rinsed with three washes of distilled, deionized water which was sterilized through a 0.22 µm syringe filter; (4) seeds were soaked in a solution of Agribrom antifungal dissolved in sterile distilled, deoinized water and incubated in the dark at 4 °C for 48 hours. Subsequently the sterilized seeds were rinsed in sterile distilled, deionized water, plated on Murashige-Skoog (no sucrose) plates (Murashige-Skoog basal salts, MES, 0.7% agar, pH 5.7) sealed with gas-permeable tape (3M, St. Paul, MN) and incubated in a lighted room to induce germination. After approximately four weeks, seedlings were removed from agar plates and plant tissue was collected and cDNA prepared as described above.

To verify their expression, full-length genes for *H. perforatum* CHS, BPS, PKS1, and PKS2 were amplified from *H. perforatum* cDNA using the specific primers shown in Table S2 and 5 PRIME MasterMix according to the manufacturer's instructions (PCR program: 95 °C, 3 min; [94 °C, 1 min; 50 °C 1 min; 68 °C 1 min 30 sec] x 50; 68 °C, 7 min). Approximately 2 µg cDNA was used in a 25 µL PCR reaction. PCR products were analyzed by agarose gel electrophoresis and then bands of the appropriate size were purified by gel purification and extraction. DNA sequencing reactions were performed using the same primers initially used to amplify the genes (Table S2) and the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. DNA was sequenced by the Biotechnology Core DNA Sequencing Laboratory at the University of Illinois at Urbana-Champaign (Urbana, IL). Comparing the sequencing results versus published sequences confirmed that all four PKSs were expressed in the *H. perforatum* seedlings.

Amplifications with the degenerate primers were performed using the following PCR program: 95 °C, 3 min; [94 °C, 1 min; 50 °C 1 min; 68 °C 1 min 30 sec] x 50; 68 °C, 7 min. DNA fragments were amplified from cDNA using the primers shown in Table S1 and 5 PRIME MasterMix according to the manufacturer's instructions. The RuBisCo gene is highly conserved among plant species. The mRNA sequences of *Eucalyptus globulus* RuBisCo and *Hypericum perforatum* RuBisCo are available. These two genes share approximately 90% sequence identity and were used to design the RuBisCo primers used in this work. Specifically, the two mRNAs were aligned using Clustal W and two small, highly conserved regions (99-100% identity) were selected for primer annealing. Amplification of the RuBisCo fragment was performed using specific primers (FWD: 5'-CTT GGC AGC ATT CCG AGT AAC TCC-3', REV: 5'-CGG TCA GAG CAG GCA TAT GC-3') for the RuBisCo gene in conjunction with the first

amplification using PKS-F1 and PKS-R3 primers, to verify the amplification was successful and cDNA was present. The primers were also used to amplify the *Eucalyptus* RuBisCo mRNA and were used as a control in *Eucalyptus* amplifications. For amplifications using cDNA as the template, 1 µl of cDNA was used; alternatively, 5 µl of PCR product was used as the template for subsequent nested PCR amplifications using primers PKS-F2 and PKS-R2 or PKS-F3-F and PKS-R1-F.

Table S2 Gene-specific primers for *H. perforatum* Type III PKSs.

Primer	Sequence, 5' → 3'
CHS-F	ATGGTGACCGTGGAAGAAGTCAG
CHS-R	TTAATATGCGACACTGTGAAGGACCAC
BPS-F	ATGGCCCCGGCGATG
BPS-R	CTGGAGAATTGGGACACTCTGGAG
PKS1-F	ATGTCTAACTTGGAGACCAATGGCTC
PKS1-R	TCATAGGCATAGGCTTCGAAGAAGG
PKS2-F	ATGGGTTCCCTTGACAATGGTTC
PKS2-R	TTAGAGAGGCACACTTCGGAGC

PCR-amplified ~400 nt DNA fragments were cloned into the pET26b(+) vector to facilitate DNA sequencing. The ~400 nt fragment was digested with *Nhe*I and *Xho*I (10 units enzyme per µg DNA), and the vector was digested with *Xba*I and *Xho*I (10 units enzyme per µg DNA). Note that *Nhe*I and *Xba*I have compatible cohesive ends. After ligation using T4 DNA ligase the plasmid was transformed into *E. coli* DH5α and single colonies were picked. Presence of an insert of the correct size was verified by colony PCR. Briefly, cells were collected from 200 µl overnight culture and resuspended in 200 µl sterile distilled, deionized water. The samples were boiled for ten minutes, cooled to room temperature, and then spun at max speed in a benchtop microcentrifuge for three minutes to pellet cell debris. A PCR amplification using *Taq* DNA polymerase (Invitrogen, Carlsbad, CA) was set up according to the manufacturer's instructions with 5 µl of the soluble cell lysate as template and using primers specific to the T7 promoter and the T7 terminator. The PCR product was analyzed by agarose gel electrophoresis. A plasmid with resulting PCR product of size ~500 nt was purified and submitted for DNA sequencing. Sequencing reactions were performed using primers specific to the T7 terminator and the BigDye Terminator v3.1 Cycle Sequencing Kit. DNA was sequenced by the Biotechnology Core DNA Sequencing Laboratory at the University of Illinois at Urbana-Champaign (Urbana, IL).

DNA sequencing of 26 fragments in pET26b(+) identified the ~400 bp fragments as belonging to the genes for *H. perforatum* CHS, *H. perforatum* BPS, and an additional, previously unknown gene, based on sequence homology (**Table S3**). The other two enzymes, PKS1 and PKS2, were not identified, probably due to lower relative expression (and thus abundance of mRNA) in the young plants. *H. perforatum* PKS1 and PKS2 were known to be expressed in mature plants, with expression particularly high in flower buds.[2]

Table S3 Sequence identity (%) of ~400 nt fragments to known *H. perforatum* PKSs.

#	BPS	CHS	PKS1	PKS2	#	BPS	CHS	PKS1	PKS2
1	94	50	41	38	**14**	95	48	37	38
2	93	50	43	34	**15**	54	87	45	43
3	95	52	42	39	**16**	54	88	46	44
4	55	88	47	43	**17**	55	87	45	43
5*	55	54	35	30	**18**	58	80	46	46
6	53	84	44	41	**19**	70	48	34	43
7	55	81	46	45	**20**	53	84	45	42
8*	58	49	36	43	**21**	54	88	46	44
9	54	88	45	44	**22**	93	52	42	40
10	90	49	35	40	**23**	94	48	36	38
11	94	51	40	42	**24**	93	48	32	33
12**	-	-	-	-	**25**	93	49	38	40
13	90	47	32	37	**26****	69	45	37	41

*Previously uncharacterized Type III PKS
**Low quality sequencing data.

The sequence identities of the four *H. perforatum* PKSs are: BPS/CHS = 61%, BPS/PKS1 = 35%, BPS/PKS2 = 47%, CHS/PKS2 = 52%, CHS/PKS1 = 40%, PKS1/PKS2 = 36%.

Screening *Eucalyptus* Species

Because the polyketide compounds of interest were generally isolated from *Eucalyptus* leaves,[3] leaf tissues from *E. robusta* and *E. camaldulensis* var. Silverton Province trees (~1.5 m height) were used for RNA isolation. Collection of plant tissue, RNA isolation, and cDNA synthesis were completed as described in the sections above. Degenerate primer nested PCR, cloning of ~400 nt fragments, and DNA sequencing were performed as described above for *H. perforatum*. A total of 27 individual clones were submitted for sequencing from *E. robusta* and 94 from *E. camaldulensis*. Clustal W and BLAST analysis using the Biology Workbench Suite[1] were used to analyze the resulting DNA and translated protein sequences. Depending on the threshold (i.e. the number of different amino acids in the cloned DNA fragments), the number of different Type III PKS genes could range from 5 to 97 (**Table S4**). If the threshold in the validation experiment was used as a reference (i.e. genes with 80% sequence identity or 27 different amino acids out of ~136 are considered the same), the identified DNA fragments could be classified into 11 groups, representing 11 unique putative Type III PKSs genes. Phylogenetic analysis of all the Type III PKS fragments is given in **Figure S6**.

Table S4 The number of expected different Type III PKS genes.

Number of different amino acids (out of ~136)	Number of expected different genes
1	97
5	31
10	16
15	12
27	11
30	9
50	5

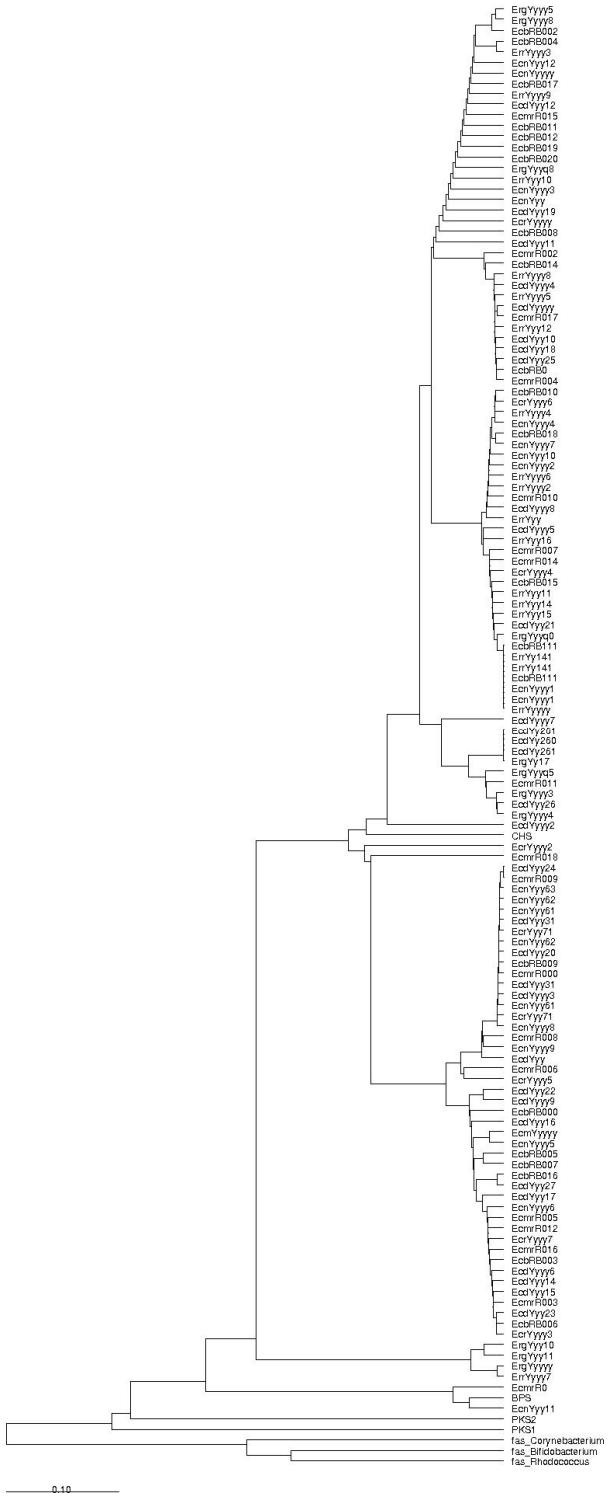

Figure S6 Phylogenetic analysis of all the cloned Type III PKS fragments. CHS, BPS, PKS1, and PKS2 were used as internal control. Fatty acid synthases (FAS) from three different species (*Cornebacterium aurimucosum* ATCC 700975, *Bifidobacterium dentium* Bd1, and *Rhodococcus opacus* B4) were used as out-group sequences.

RACE-PCR

The core fragment of EC2 was subjected to RACE using gene specific primers and the protocol of 3'/5' RACE system for Rapid Amplification of cDNA Ends Kits Version 2.0 (Invitrogen, Carlsbad, CA). The PCR program consists of 94 °C (2 min), 35 cycles at 94 °C (1 min), 55 °C (30s) and 72 °C (2 min), and a final 10 min extension at 72 °C. A plasmid with resulting PCR product of size ~1000 nt was purified and submitted for DNA sequencing. Sequencing reactions were performed using primers specific to the T7 terminator and the BigDye® Terminator v3.1 Cycle Sequencing Kit.

Table S5 Gene-specific primers for RACE PCR.

Primer	Sequence, 5' → 3'
EC2-F1	CGACACCGAGACCTACCTGGACAACC
EC2-F2	CGAGTGGGCCAGGCTCTGTTCG
EC2-F3	CGA*TTCGCTAGC*GATGGCGCCTCCTCCACC
EC2-R1	*TTCGCTAGC*TGTCCAGGTAGGTCTCG
EC2-R2	ACAGAGCCTGGCCCAC
EC2-R3	GGTGGAGGAGGCGCCATC

1. Subramaniam, S. The Biology Workbench--a seamless database and analysis environment for the biologist. *Proteins* **32**, 1-2 (1998).
2. Karppinen, K. & Hohtola, A. Molecular cloning and tissue-specific expression of two cDNAs encoding polyketide synthases from Hypericum perforatum. *J. Plant Physiol* **165**, 1079-1086 (2008).
3. Ghisalberti, E.L. Bioactive acylphloroglucinol derivatives from Eucalyptus species. *Phytochemistry* **41**, 7-22 (1996).