농학박사학위논문

쪽(*Polygonum tinctorium* L.)과 후추(*Piper nigrum* L.)에서 알칼로이드와 세스퀴테펜 생합성효소의 유전자 동정과 특성

Gene identification and characterization of enzymes involved in alkaloid and sesquiterpenoids biosyntheses in *Polygonum tinctorium* L. and *Piper nigrum* L.

2018년 2월

서울대학교 대학원

농생명공학부 응용생명화학전공

Jin Zhehao

A Dissertation for the Degree of Doctor of Philosophy

Gene identification and characterization of enzymes involved in alkaloid and sesquiterpenoids biosyntheses in *Polygonum tinctorium* L. and *Piper nigrum* L.

February 2018

Zhehao Jin

Applied Life Chemistry Major

Department of Agricultural Biotechnology

Seoul National University

쪽(*Polygonum tinctorium* L.)과 후추(*Piper nigrum* L.)에서 알칼로이드와 세스퀴테펜 생합성효소의 유전자 동정과 특성

Gene identification and characterization of enzymes involved in alkaloid and sesquiterpenoids biosyntheses in *Polygonum tinctorium* L. and *Piper nigrum* L.

지도교수 김 수 언

이 논문을 농학박사학위논문으로 제출함
2018년 2월

서울대학교 대학원
농생명공학부 응용생명화학전공

Jin Zhehao

Jin Zhehao의 박사학위논문을 인준함
2018년 2월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Gene identification and characterization of enzymes involved in alkaloid and sesquiterpenoids biosyntheses in *Polygonum tinctorium* L. and *Piper nigrum* L.

Advisor: Soo-Un Kim

A Dissertation Submitted in Partial Fulfillment
of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

to the Faculty of
Applied Life Chemistry Major,
Department of Agricultural Biotechnology
at
SEOUL NATIONAL UNIVERSITY
by

Zhehao Jin

Date Approved

_____ _____

_____

_____

_____

_____

Abstract

This thesis presents enzymes involved in biosynthesis of plant alkaloids, indigo and piperine from *Polygonum tinctorium* L. and *Piper nigrum* L., respectively, and those in sesquiterpene synthesis in *P. nigrum*. The first chapter presents indole synthase from *P. tinctorium*. Indigo is an old natural blue dye produced by plants such as *P. tinctorium*. The first key step in plant indigoid biosynthesis is the production of indole by indole-3-glycerol phosphate lyase (IGL). Two tryptophan synthase α-subunit homologs, *PtIGL*-short and -long forms on genome of the plant contained two genes each coding for IGL. The short and the long forms respectively encoded 273 and 316 amino acid residue-long proteins. The short form complemented *E. coli* Δ*tnaA* Δ*trpA* mutant on tryptophan-depleted agar plate, signifying the production of free indole, and thus was named indole synthase gene (*PtINS*). The long form, either intact or without the transit peptide sequence, did not complement the mutant. It was tentatively named *PtTSA*. PtTSA is transported to the chloroplast as predicted by 42 amino acid residues of targeting sequence, whereas PtINS is localized in cytosol. Genomic structure analysis suggested that a *TSA* duplicate acquired splicing sites during the course of evolution toward *PtINS* so that the targeting sequence-containing pre-mRNA segment was deleted as an intron. *PtINS* had about two to five-fold higher transcript level than that of *PtTSA*, and treatment of 2,1,3-benzothiadiazole caused the relative transcript level of *PtINS* over *PtTSA* significantly enhanced in the plant.

The second and the third chapter respectively focuses on sesquiterpene synthesis imparting characteristic peppery bouquet and piperine alkaloid biogenesis responsible for pungent taste of black pepper. The unripe peppercorn was submitted to transcriptome analysis utilizing the Illumina next-generation sequencing (NGS). Compared with gene cloning based on rapid amplification of cDNA ends (RACE)-PCR, NGS technology offers more cost-effective and time-saving alternative to

identify specific gene by collecting massive sequencing data. Using Local tBLASTn routine against query genes with similar biochemical functions, I have found three full-length of sesquiterpene synthase (sesqui-TPS) clones (*PnTPS1* through *3*) and four kinds of enzymes putatively involved in piperine biosynthesis (PnMCHL; PnPKS1 and 2; Pn4CL3; PnNAT6 and 7). The genes were expressed in *E. coli* to obtain proteins for *in-vitro* assay. In addition, they were expressed in the engineered yeast and *E. coli* to directly produce the expected products or to effect bioconversion *in-vivo*. In particular, Part II fully describes sesqui-TPSs mentioned above. PnTPS1 produced caryophyllene as a major product and minor humulene, and thus was named caryophyllene synthase (PnCPS). Likewise, PnTPS2 and PnTPS3 were named cadinol/cadinene synthase (PnCO/CDS) and germacrene D synthase (PnGDS). PnGDS expression in yeast system yielded β-cadinene and α-copaene also found in pepper extract. They were verified as rearrangement products of germacrene D not found in pepper.

Part III describes transcriptome-based gene mining for elucidation of piperine biosynthesis. At first, *P. nigrum* 3,4-methylenedioxycinnamic acid (MDCA) hydratase-lyase (PnMCHL) responsible for conversion of MDCA to piperonal, was identified and described. Piperonylic acid, possibly an oxidation product from piperonal *in-planta*, could undergo $2\times C_2$ extension in the side chain to arrive at C_6C_5 carbon skeleton of piperic acid, as opposed to the common belief that piperic acid skeleton would be the results of one C_2 extension from MDCA. Also described is 3,4-methylenedioxyphenyl-specific 4-coumaroyl-coenzyme A ligase (Pn4CL3) which converted piperic acid into piperoyl-CoA. Finally, two clones coding enzymes for transfer of piperoyl-CoA to piperidine are described.

Keywords: Biosynthesis, Indole synthase, *Piper nigrum* L., *Polygonum tinctorium* L., Piperine, Sesquiterpene.

***Student Number*: 2013-31316**

Contents

List of Figures

List of Tables

List of Abbreviations

4CL	4-coumarate:coenzyme A ligase
bp	Base pair
BX1	Bezoxazinless 1
BTH	2,1,3-Benzothiadiazole
cDNA	Complementary DNA
CO/CDS	Cadinol/cadinene synthase
CHS	Chalcone synthase
CPS	Caryophyllene synthase
CTAB	Hexadecyl trimethyl ammonium bromide
CYP	Cytochrome p450
DCS	Diketide CoA synthase
DTT	Dithiothreitol
E. coli	*Escherichia coli*
FDC1	Ferulic acid decarboxylase 1
FPP	Farnesyl diphosphate
GAP	Glyceroldehyde-3-phosphate
GC-MS	Gas chromatography-mass spectrometry
GDS	Germacrene D synthase
GFP	Green fluorescence protein
gDNA	Genomic DNA
HML	Histidine, methionine, leucine
IGL	Indole-3-glycerol phosphate lyase
IGP	Indole-3-glycerol phosphate
INS	Indole synthase
kDa	Kilodalton
LB	Luria-Bertani
LC-MS	Liquid chromatography-mass spectrometry
MBP	Maltose binding protein
MCHL	MDCA hydratase-lyase
MDCA	3,4-methylenedioxycinnamic acid

MW	Molecular weight
NAT	*N*-acyltransferase
NGS	Next generation sequencing
OD	Optical density
ORF	Open reading frame
PAD1	Phenylacrylic acid decarboxylase 1
PAL	Phenylalanine ammonia lyase
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PKS	Polyketide synthase
Q20	Transcriptome sequencing accuracy of 99%
qRT-PCR	Quantitative real-time PCR
RACE	Rapid amplification of cDNA ends
RIN	RNA integrity number
RT-PCR	Reverse transcription polymerase chain reaction
SAR	Systemic acquired resistance
SD (DO)	Synthetic dropout (dropout)
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
tnaA	Tryptophanase gene
t_R	Retention time
trpA	Tryptophan synthase α-unit gene
trpB	Tryptophan synthase β-unit gene
TIC	Total ion chromatogram
TPS	Terpene synthase
TSA	Tryptophan synthase α-unit
TSS	Translation start site
UTR	Untranslated region

Part I: Isolation and functional studies of indole synthase from *Polygonum tinctorium* L.

Introduction

***Polygonum tinctorium* L.**

Indigo is the only natural blue dye known to human beings (Ensley et al. 1983). *P. tinctorium* had been cultivated in Europe and East Asia for the production of indigo (natural blue dye) up to the 17th century. After this time, indigo from the *Indigofera tinctoria* was used because of the superior quality of the dyestuff (Schmidt 1997). However, the use of natural indigo was discontinued with the advent of cheaper synthetic indigo, which almost completely replaced natural indigo at the end of the 19th century (Schrott 2001). However, in Far East countries, *P. tinctorium* has been the most important indigo dye from plant (Kim et al. 2007).

Nevertheless, the esthetic hue of the natural indigo dye enabled the cultivation of indigo plant to continue until now, especially in Korea, Japan, and some European countries. The current consumption of the dye is enormous due to the popularity of blue jeans, which are mostly dyed with synthetic indigo. Natural fabric dyeing recently attracted attention from public due to the concern over the potential environmental damage from chemical indigo dyeing process. More recently, because of the importance of natural indigo, considerable research has been performed to replace chemical synthesis of indigo by application of biotechnological methods (Chae et al. 2000; Thwe et al. 2012; Sakamoto et al. 2017). In Korea, city of Naju, sees the natural indigo industry as a highly profitable business and encourages local farmers to cultivate *P. tinctorium* as a cash crop.

Indole biosynthetic pathway

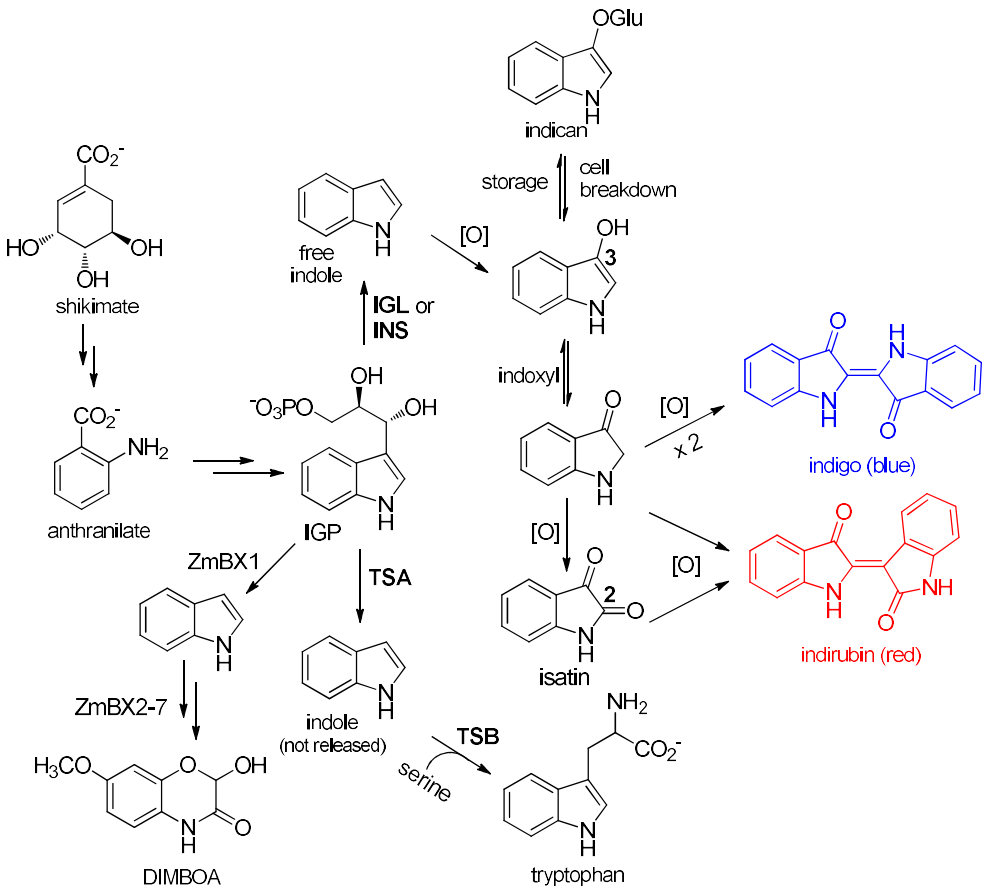
The biosynthetic precursor of indigo, free indole, is directly derived from indole-3-glycerol phosphate (IGP) by indole-3-glycerol phosphate lyase (IGL) (Xia and Zenk 1992). The indole-producing pathway shares central metabolic pathways that produce the precursors of aromatic amino acid, shikimate pathway. Also indicated in

the Fig. 1.1 is the cleavage of tryptophan produced *de novo* or supplied exogenously to indole via tryptophanase (Berry et al. 2002). Free indole must be oxidized to indoxyl, which is very unstable molecule, prior to conversion into indigo. The oxidation reactions from indoxyl to indigo presumably must occur extracellularly (Bolwell et al. 1994). Two molecules of indoxyl combines to form indigo (blue color), whereas indoxyl and isatin (indoxyl oxidation products) lead to the synthesis of indirubin (red color). In indigo producing plants, indoxyl exists as indoxyl-β-D-glucoside, indican, for storage in the vacuole (Maugard et al. 2002). *P. tinctorium* contains indican which serves as a starting material for indigo production. Damage of leaves exposes indoxyl-β-D-glucosides to β-glucosidase, and the synthesis indigo ensues (Maier et al. 1990). In the xiamycin producing *Streptomyces* sp., a flavin-dependent bacterial indole terpenoid cyclase XiaF catalyzes formation of indoxyl, which was spontaneously condensed to indigo (Kugel et al. 2017). In animal, the indole ring was produced by gut bacteria, such as *Bifidobacterium* species, expressing tryptophanase. Indole thus produced was oxidized by hepatic cytochrome P450s, for example, porcine CYP2A6 (Gillam et al. 2000) or in human CYP2E1 (Chen et al. 2008; Wikoff et al. 2009), to produce indoxyl. A cytochrome P450 (*ItB24*) from another indigo-producing plant, *Isatis tinctoria,* causes oxidation at C-3 position (Fig. 1.1) of indole to form indoxyl which ultimately forms indigo when co-expressed with a TSA from the same plant in *E. coli* (Multani 2015).

Indole and indole compounds have become the subject of increasing interest since the discovery of their association with plant growth hormones, specifically auxins (Kende and Zeevaart 1997). Indirubin as potent inhibitors of cyclin-dependent kinase of DNA synthesis in several cell lines warranted it as anticancer drug (Hoessel et al. 1999). In bacteria, indole reduces biofilm formation at the low growth temperature (30°C) by activating stress responses (Lee et al. 2008).

Figure 1.1. Biosynthesis of indole and its derivatives in plants.

BX1, benzoxazin less 1; IGL, IGP lyase; IGP, indole-3-glycerol phosphate; INS, indole synthase; Glu, glucose; [O], oxidation by air or cytochrome P450; TSA, tryptophan synthase α-subunit; TSB, tryptophan synthase β-subunit.


OGlu
indican
storage
cell breakdown
free indole
[O]
OH
3
IGL or INS
indoxyl
shikimate
anthranilate
IGP
[O]
x 2
indigo (blue)
ZmBX1
TSA
[O]
2
isatin
[O]
indirubin (red)
ZmBX2-7
indole (not released)
TSB
serine
DIMBOA
tryptophan

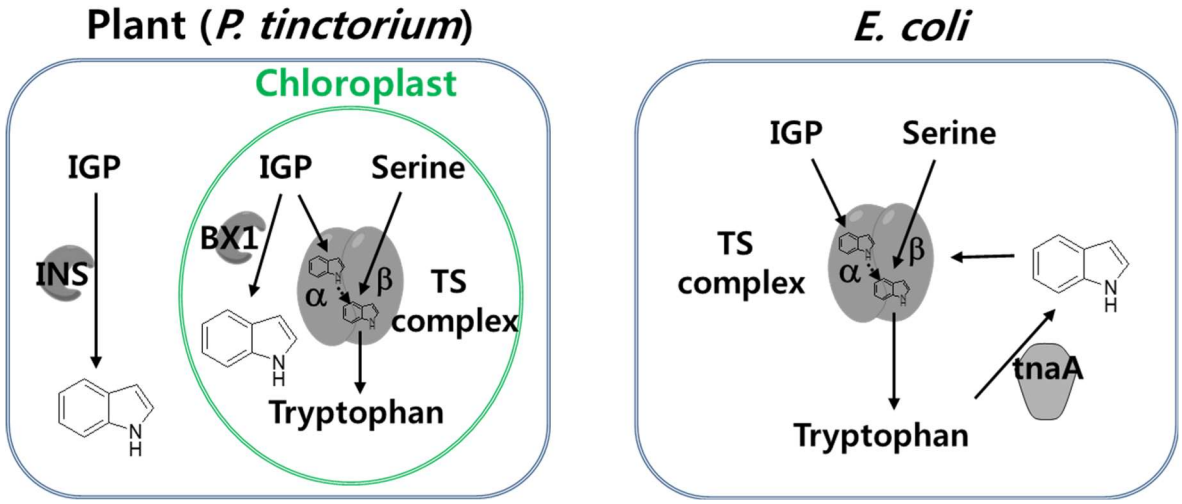
Indole synthase

Indole-3-glycerol phosphate lyase (IGL) is a generic term to indicate free-standing enzymes that cleave indole-3-glycerol phosphate (IGP) to indole and glyceraldehyde-3-phosphate (GAP) (Frey et al. 2000). Whereas α-subunit of tryptophan synthase (TSA) is localized in chloroplast in *Arabidopsis* mesophyll cells, IGL is observed in cytosol (Zhang et al. 2008). IGL and BX1 have the same function as TSA. While BX1 and IGL cleaves IGP to form free indole and GAP, TSA catalyzes the conversion of IGP into indole that travels through a tunnel connecting the active sites of the α- and β-subunits to combine with serine to form tryptophan (Tuominen et al. 2011) (Fig. 1.1 and 1.2). Therefore, indole is not released from the TS complex (Gierl and Frey 2001). BX1 has high homology with TSA, and catalyzes the formation of free indole, defining the branch point from primary to secondary metabolites (Frey et al. 2000; Nomura et al. 2003; Kulik et al. 2005). Indole moiety of the most of indole-containing secondary metabolites, such as indole-3-acetic acid, are derived from essential amino acid tryptophan (Li et al. 1995; Frey et al. 1997).

In bacteria, *tnaA* gene plays key role in indigo production. *tnaA* cleaves tryptophan into indole and serine. The indole oxidation is affected by naphthalene dioxygenase to indoxyl. The oxidation reactions from indoxyl to indigo presumably occur extracellularly. However, TS complex composed of subunits trpA and trpB catalyzes the conversion of IGP and serine to tryptophan (Fig. 1.2), (Berry et al. 2002).

Figure 1.2. Indole-related enzymes in *P. tinctorium* and *E. coli*.

IGP, indole-3-glycerol phosphate; BX1, benzoxazinless 1; TS complex, tryptophan synthase complex; INS (or IGL), indole synthase (or IGP lyase); tnaA, tryptophanase


Plant (*P. tinctorium*)
Chloroplast
IGP
INS
IGP
Serine
BX1
α
β
TS
complex
Tryptophan
E. coli
IGP
Serine
TS
complex
α
β
tnaA
Tryptophan

The purposes of research

The purpose of this study was to isolate indole synthase (INS) from *P. tinctorium* that was putatively involved in biosynthesis of free indole to understand indole metabolism in the plant. Two candidate genes, phylogenetically very close to each other but with distinctive size, function, subcellular localization, and genomic sequence, were isolated and their functions were confirmed.

Materials and Methods

Plant material and growth conditions

P. tinctorium seeds, gift from Professor Kwan-Soo Kim at Mokpo National University, Korea, were germinated at 26°C on wet filter paper in glass Petri dish. Germinated seeds were transferred to 20 cm plastic pots filled with Biosoil (Hungnong Seeds®, Pyeongtaek, Korea). Seedlings were grown in growth chamber at 24°C with 16/8 h light/dark regimen, until the 5th leaf fully opened.

For elicitation experiment, aqueous 0.3 mM BTH containing 0.05% (v/v) of Tween-20 was sprayed onto the plants until the leaves became wet. The leaves were collected on the third day after the treatment for RNA isolation (Jin et al. 2012).

Nicotiana benthamiana for transient expression was grown until 5-6 leaf-stage in a plant growth chamber with 16 h light /8 h dark cycle at 24°C.

Bacterial strains and culture media

Bacterial strains used in this experiment were *Escherichia coli* DH10B [F^{-} *mcrA Δ*(*mrr-hsdRMS-mcrBC*) *Φ80d lacZΔM15 ΔlacX74 recA1 deoR Δ(ara,leu)7697 araD139 galU galK rpsL λ*$^{-}$ *endA1 nupG*] for the transformation and plasmid propagation, and *E. coli ΔtnaA ΔtrpA* mutant [*supE44 hsdR thi-1thr-1 leuB6 lacY1 tonA21 recD1009* Δ*tnaA* Δ*trpA*] (Jin 2013) for the complementation assay. *E. coli* was grown in Luria-Bertani (LB) medium (MBcell, Korea) and solid LB medium plate was made with 1.5% (w/v) Micor-agar (Duchefa, Netherlands). The transformants were cultured in LB medium containing antibiotics. The concentration and the kinds of antibiotics will be described in the proper section below.

Plasmids

pUC19 (Fermentas, Canada) and *pMD20* T-vector (TaKaRa, Japan) were used for the sub-cloning. *pMW118* expression vector (Nippon gene, Japan) was used for

complementation assay. Subcellular localization experiment employed *pEAQ*-GFP (Sainsbury et al. 2009).

Enzyme and chemicals

Restriction enzymes, T4 DNA ligase, dNTPs, Taq polymerase and Prime STAR HS DNA polymerase were purchased from TaKaRa (Japan). BTH (2,1,3-Benzothiadiazole), Kovac's reagent (or indole test reagent), and others chemicals were purchase from Sigma-Aldrich.

Oligonucleotides

Oligonucleotides used for DNA sequencing and polymerase chain reaction (PCR) were purchased from Macrogen Korea Co. (Seoul, Korea). The sequence of the oligonucleotides was listed in Table 1.1.

Table 1.1. List of oligonucleotides used in *PtIGL* research.

Primer	(5'→3')
long 5UTR F	caaacatcaccccattttccccaatt
long 5UTR R	cgagcgaggccatgatcttggt
short 5UTR F	agcccagctccgtcgtcgcct
short 5UTR R	gcgagagaggccatgatcttggt
PtIGL L-F	atggctgtctccctcagatca
PtIGL S-F	atggcctctctcgcgacttcga
PtIGL-R	tcaaacaagggcagatttcagggatt
pMW118:Pt L-F	ggtacccggggatctatggctgtctccctcagat
pMW118:Pt S-F	ggtacccggggatcgatggcctctctcgcgactt
Pt L-Δ42-F	ggtacccggggatcgatggcctcgctcgctacttc
pMW118:Pt-R	gactctagaggatcctcaaacaagggcagatttc
pMW118:AtIGL-F	ggtacccggggatcgatggatcttctcaagactc
pMW118:AtIGL-R	gactctagaggatcctcaagagacaagagcaga
pMW118:AtTSA-F	ggtacccggggatcgatggcgattgctttcaaat
pMW118:AtTSA-R	cgactctagaggatcctcaagaagagcagatta
Pt LR-F	ctttgttttccctaattcggcaa
Pt LR-R	cggcgctggagaaagcaagt
Pt SR-F	atggcctctctcgcgacttc
Pt SR-R	ccacaagtgtcaagcaacttcaa
pEAQ:GFP L-F	gcccaattcgcgaccatggctgtctccctcagat
pEAQ:GFP S-F	gcccaattcgcgaccatggcctctctcgcgact
pEAQ:GFP-R	ctcctttagtcataccaacaagggcagatttcag

Genomic DNA, total RNA isolation and cDNA synthesis

Total RNA was isolated from the leaves of one-month old *P. tinctorium* through CTAB method (Chang et al. 1993). The total RNA was treated with RNase-free DNase I (PhileKorea, Korea) and purified by phenol-chloroform method (Tozawa et al. 2001). Two micrograms of purified total RNA was reverse-transcribed using an oligo dT_{17} primer (SuperScript Reverse Transcriptase Kit, Invitrogen, USA) with the following PCR conditions to complete the synthesis of the first strand cDNA; 7 min at 65℃, 2 min on an ice bath, 60 min at 37℃, 5 min at 94℃, and finally hold at 16℃. Genomic DNA was extracted with Wizard® Genomic DNA Purification Kit (Promega, USA) from one gram of plant leaves ground with liquid nitrogen.

Isolation of *IGL* and *UTR* sequences

Two full ORF sequences were amplified with combination of PtIGL-R and PtIGL L-F or PtIGL S-F. Truncated PtIGL-long devoid of 42-residue-long transit peptide (PtIGL-longΔ42) was cloned by using primer pair of Pt L-Δ42-F and pMW118:Pt-R.

Primer pairs based on uppermost 5' UTR sequence of each IGL form (long and short 5UTR F) and about 13 bp downstream from deduced second translation starting points (long and short 5UTR R) were employed to clone genomic IGL sequence near 5' ends. The PCR products were purified and cloned into *pMD20* T-vector (TaKaRa, Japan).

AtIGL (AT4G02610) and *AtTSA* (AT3G54640) ORF sequences were cloned as above by using the corresponding primer pairs (AtIGL-F or AtTSA-F and R) and cloned into *pMW118*, a low copy expression vector for complementation assay in *E. coli* Δ*tnaA* Δ*trpA*.

Complementation assay in *E. coli ΔtrpA ΔtnaA*

The *IGL*- or *TSA*-harboring *pMW118* was transformed into *E. coli* Δ*trpA* Δ*tnaA* double mutant by using heat-shock method. The selection was done on LB medium supplemented with 32 μg/ml of chloramphenicol, 50 μg/ml of kanamycin, and 100 μg/ml of ampicillin. Finally, individual colony was selected and streaked on tryptophan-depleted M9CG agar medium (Kramer and Koziel 1995), containing the above-mentioned concentrations of antibiotics.

QRT-PCR

Transcript levels of *PtIGL*s were measured by using qRT-PCR. To assess *PtIGL*-long and -short transcript levels by qRT-PCR, primer pairs Pt LR-F and -R and Pt SR-F and -R were used. The PCR reactions were carried out in quadruplicates for 40 cycles on Rotor-Gene 2000 Real Time Cycler (Corbett Research, Australia) by using QuantiMix SYBR Kit PCR system (PhileKorea, Korea).

Determination of subcellular localization

*PtIGL*s were amplified by PCR and fused to upstream of the *GFP* reporter gene of *pEAQ*-HT vector (Sainsbury et al. 2009) as follows. pEAQ:GFP L-F and pEAQ:GFP S-F and pEAQ:GFP-R primer pairs were employed to assemble long and short IGL-GFP fusion proteins into *pEAQ-GFP* by using Gibson Assembly Kit (NEB, USA). Assembly reaction condition was incubation at 50℃ for 20 min followed by on ice bath for 2 min. *Agrobacterium tumefaciens* LBA4404 harboring the recombinant plasmid was cultured at 28℃ in 5 ml of LB medium containing kanamycin (50 μg/ml) and rifampicin (50 μg/ml). The cultures were centrifuged at 3,500 rpm (F1202 Rotor, Beckman) at room temperature for 5 min, and then resuspended in infiltration buffer (10 mM MES, 10 mM $MgCl_2$ and 100 μM acetosyringone, pH 5.6) to adjust OD_{600} to 0.2. The bacteria were infiltrated into the leaves of *N. benthamiana* by using 1 ml syringe (Sainsbury et al. 2009). After 5 days of incubation in the plant growth

chamber, the signals for GFP and chloroplast autofluorescence were examined under a confocal microscope (SP8 X STED, Leica, Germany) equipped with 10× ocular and 100× oil immersion objective operating at 488 nm for excitation and 509 nm and 670 nm for GFP and chlorophyll emissions, respectively. The image was processed by using Leica LAS AF Lite image view software.

Bioinformatics analyses

The alignment of nucleotide and amino acid sequences was performed using the Vector NTI Advance® software (Thermo Fisher Scientific, USA). ChloroP (Emanuelsson et al. 1999; Emanuelsson et al. 2007) was used to predict the subcellular localization of protein. The phylogenetic tree was constructed with MEGA6 (www.megasoftware.net) and sequence analyses (translation, reverse complement and restriction summary) were done at www.bioinformatics.org/sms2/ web-site.

Results and Discussion

Cloning of *IGL*s

PCRs of cDNA library using combination of 5'- and 3'-end sequences yielded two clones each with distinctive 5' UTR sequences but highly homologous ORF regions (Fig. 1.3). One cDNA, tentatively named *PtIGL*-long (GenBank accession No. KU049663), was 1102 bp-long and contained an open-reading frame (ORF) for 316 amino acid residues. The other cDNA sequence (*PtIGL*-short; KU049662) contained an ORF coding for a protein consisted of 273 amino acid residues. The estimated MW of the long and the short form proteins were 33.6 and 28.8 kD, respectively. Alignment of amino acid sequences of both forms indicated that translation start site (TSS) of the short form corresponded to the second methionine of the long form (Fig. 1.3).

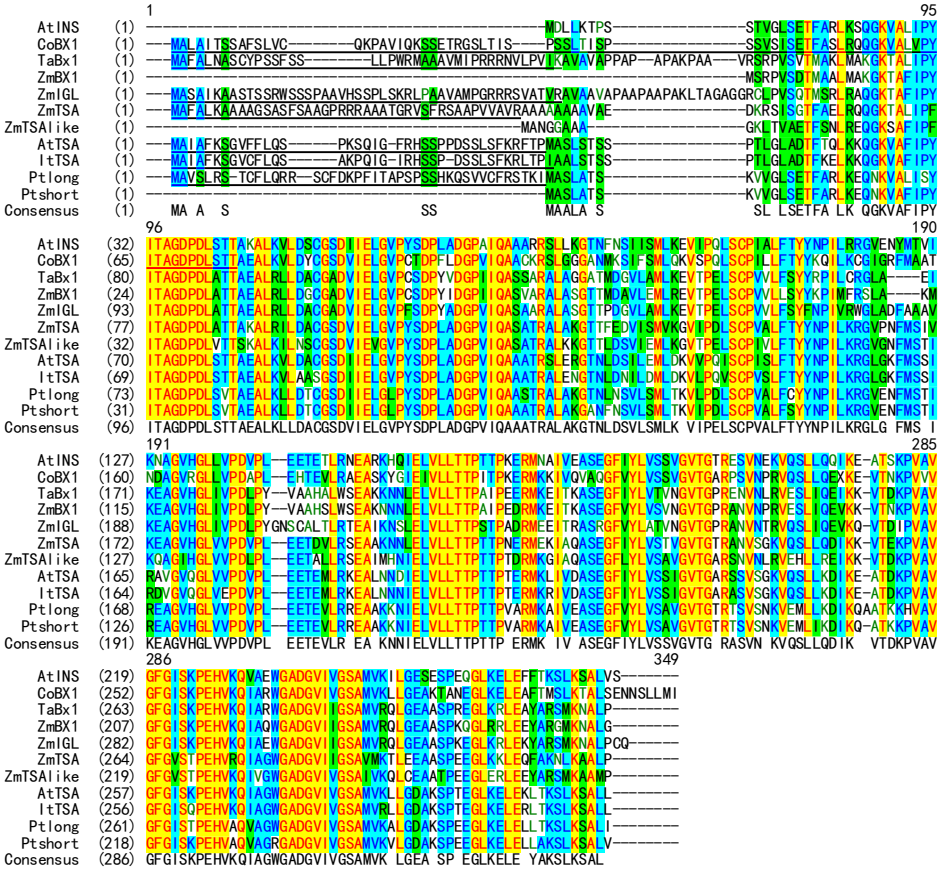
The biosynthetic precursor of indigo in higher plants was identified 25 years ago as indole that did not originate from tryptophan (Xia and Zenk 1992). Indole acts not only as a precursor of several plant secondary metabolites such as indigoid pigments (Maier et al. 1990), benzoxazinoids (Frey et al. 2009), and auxin (Mashiguchi et al. 2011) but also is a defense molecule emitted from plants in a tritrophic interaction system (Frey et al. 2000). Indole also appears transiently in the reaction of tryptophan synthase complex: TSA cleaves IGP to generate the indole that is directly delivered into TSB without diffusing into reaction medium (Swift and Stewart 1991). Generation of free indole in plants is achieved by TSA homologs generally termed as IGL or more specifically INS. Therefore, it is highly possible that a free indole-generating IGL exists to specifically feed indole into the indigoid pathway (Fig. 1.1). In an effort to clone an *IGL* gene producing free indole for indigo biosynthesis in *Isatis tinctoria*, Salvini et al. (2008) found a TSA-related gene, but its function did not specifically correlate with indigo accumulation. In the present study, we cloned

two *IGL*s from *P. tinctorium*, one of which was positively identified as INS possibly supplying indole for indigoid biosynthesis.

PtINS retains Asp64-Pro65 residues known to play a crucial role in intersubunit communication (Rowlett et al. 1998). The residues are highly conserved among plant IGLs (Fig. 1.3). Kulik et al. (2005) compared *Salmonella typhimurium* TSA and ZmBX1 to understand the structural basis of the catalytic difference. They found that maintenance of active conformation of catalytic Glu135 and closed αL6 loop were responsible for 1,400 times higher indole producing activity compared to *S. typhimurium* TSA. However, when we compare primary structure of plant ZmBX1 with those of plant TSA and INS, no clear pattern in amino acid residues that was suggested to contribute to the active conformation of ZmBX1 was found. Therefore, comparison of three dimensional structures of PtINS and PtTSA, with very small change in the primary structure, could provide a clue to identify mutation(s) responsible for IGL-cleaving activity without the presence of TSB.

Figure 1.3. Alignment of the putative *P. tinctorium* IGLs with IGL-like proteins.

AtINS, AAP04082; AtTSA, AAC49117; CoBX1, ACJ02769; ItTSA, CAH56478; Ptlong, KU049663; Ptshort, KU049662; TaBx1, BAC81205; ZmBX1, AFW60209; ZmIGL, AAG42687; ZmTSA, ACA25187; ZmTSAlike, DAA51946. Underlined indicates chloroplast transit peptide.



Analysis of *UTR* sequences

PtIGL-short appeared in two variant forms that differ only in 5' UTR length (data not shown). One variant was shorter by 77 bp from -14 through -90 compared to the other variant, which was 1017 bp-long containing 195 bp-long 5' UTR (Fig. 1.4).

Genomic DNA PCR of each form of PtIGL yielded unique genomic sequence (Fig. 1.4). The sequence of *PtIGL*-long, containing 151 bp-long 5' UTR upstream of TSS (1st ATG), was identical to that of cDNA. On the other hand, the *PtIGL*-short genomic sequence had a 969 bp-long segment upstream of TSS (2nd ATG). Interestingly, at the site corresponding to 1st ATG of *PtIGL*-long, an ATG was also present in the genomic *PtIGL*-short sequence. It turned out that *PtIGL*-short sequence was interrupted by two overlapping introns as a result of alternative splicing. One intron covered 774 bp-long segment, from -91 [ag(I)] to -864, and the other contained-additional 77 bp from -14 [ag(II)] to -90 (Fig. 1.4). Both variants for *PtIGL*-short were found in cDNA library as mentioned above. Identity of PtIGLs to IGL proteins from other species ranged 49-87%. Within both PtIGL-long and short forms, the amino acid identity was as high as 93.8% excluding 42 amino acid residue-long signal peptide. The phylogenetic tree indicated that both *PtIGL*s are closely clustered together within dicot IGL family including TSAs (Fig. 1.5).

The presence of two IGL transcripts with high homology and different 5′ UTR structures prompted the examination of partial genomic structure upstream of TSS. It turned out that the structure of 5′ UTR region in genomic *PtTSA* was identical to that of *PtTSA* cDNAs, and among the two putative in-frame ATGs, the upstream first ATG was most likely a true TSS as it was consistently aligned with TSS of other plastid-targeted IGLs (Fig. 1.3). However, the genomic *PtINS* 5′ UTR was considerably longer than the 5′ UTR sequence of cDNA suggesting the presence of an intron. It was found that there are splicing consensus sequences that easily explain 5′ UTR structure of the short form cDNA: two overlapping introns flanked by (C)AG|GT(T)AG(C) at 5′ splicing site and two alternative CAG|(C) and CAG|(A) at

3′ site (Reddy 2007) (Fig. 1.4). Two possible 3′ splicing sites, denoted as AG(I) and AG(II), would provide an opportunity for alternative splicing and explain two cDNA variants of *PtINS* (Fig. 1.4). It is noteworthy that splicing out of either intron from *PtINS* pre-mRNA resulted in deletion of the first ATG. The nucleotide sequence between the first and the second ATG in genomic *PtINS* could be translated into 42 amino acid residue-long segment which was 86% identical to the corresponding sequence of *PtTSA*. Therefore, it is clear that evolution of *PtINS* from *PtTSA* duplicate was achieved by deletion of the first ATG through splicing out of intron from pre-mRNA, not by direct mutation on genomic TSS sequence.

It is interesting that genomic *PtINS* sequence suggests the alternative splicing and the corresponding cDNA variants were found among the cDNA library (Fig. 1.4). Alternative splicing of pre-mRNA in plants is known to be affected by stresses and developmental cues (Reddy 2007). For example, alternative splicing of the N gene in tobacco that confers resistance to tobacco mosaic virus is activated by infection for complete resistance (Dinesh-Kumar and Baker 2000). The alternative splicing is also known to give variation in promoter structure which can be under differential translational regulation (de Klerk and ‘t Hoen 2015).

Figure 1.4. The genomic 5' UTR sequences of *PtIGL*-long and -short.

Regions in red corresponds to cDNA. The 774 bp-long intron segment of *PtIGL*-short appeared in black, from -91 [ag(I)] to -864, was spliced out to result in the long variant mRNA. Alternative splicing from -14 [ag(II)] to -864 was responsible for the short variant.

PtIGL-Long

−151 caaacatcaccccatttttccccaattcattgcattctttgttttccctaattcggcaaaaattaaaattaaaaaa

−76 gcaattggctctctgatctgagcgatcccactgcaaaataaaccccctttcgaagaagctatcaacatcctctgt

−1 a1st ATGgctgtctccctcagatctacttgctttctccagcgccgcagctgtttcgacaagccatttatcaccgctcc

75 ttcgccttcttcccataaacaatcggtggtttgcttcagatctaccaagatc2nd ATG

PtIGL-Short

−969 agcccagctccgtcgtcgcctctgtttgttgtcgcctgcagtctcgcgtcgctgccgagctccgtcgccgtcgcc

−894 tgtcccgctttctcacaggttgaacctcaggttagcattcgattctccgcccgttatatctctcacaagctactc

−819 acaacttcaacattgcttgaatttgagattttcttcttcttcaaaacataggttaaattaaaccttctgcttcct

−744 gtcttcaatttccatgtgttttcttttccttcctggcttcaatttctctttcctggttaaattaaagcttatcct

−669 gcttgcttcaatttccctgggttgctgtcacaatcagaggcggttcctggcttcaatttctctttcctgatatgg

−594 ctataaaaagattaaacgaattaggatttatgaatattggttattctttgaatttaaatgaatgtgtgatagatt

−519 ttgcaatttccaaatacaacttcgtgggtttttagtatgtgactttatccggactgtgggtttgaatgtgtatgt

−444 cagtttgcaatttcctttcatgtgctattgaacttaaatgaatgtgtgatagattttggtggttctttgaattta

−369 acattcgggatatgaaaacccgtaaaatggatttacctcctttatcgtcgtcgagttgaactatttttgtataaa

−294 actctaattaggaatgaataactctttgtgctagtgaaatctttactttttatcgctattttgtggtattgaac

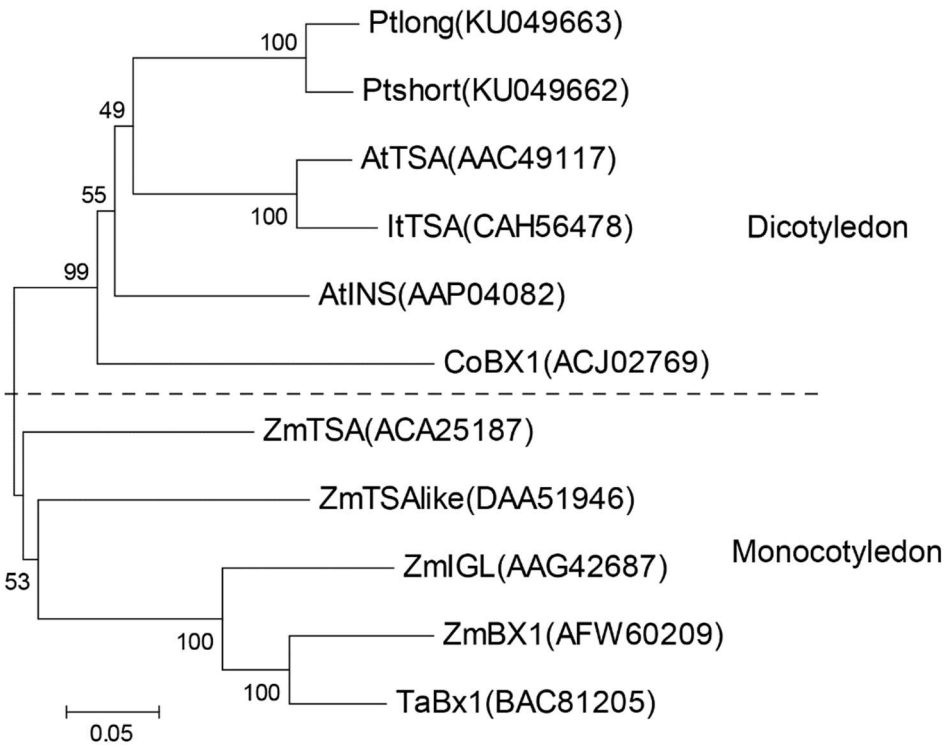
−219 acttttgtattatatgagtttgtgctattcactttctttgatgtcggtgtgcccccattctcaaagagctctg

−144 gctccgtcactgtctgtaA1st TGgcggtttccctcaaatcaacttgctttctccaagcgccgcgcctgtttcgacaag

−69 caatttctcaccactccttcgccttcttctcacaaacaatcggcggtttgcttcagatccaccaagattA2nd TG

Figure 1.5. Phylogenetic tree of the IGL-like proteins.

Only the proteins functionally characterized are listed.
At, *Arabidopsis thaliana*; Co, *Consolida orientalis*; It, *Isatis tinctoria*; Pt*, P. tinctorium*; *Ta, Triticum aestivum*; Zm, *Zea mays.*


Ptlong(KU049663)
100
Ptshort(KU049662)
49
AtTSA(AAC49117)
55
100
ItTSA(CAH56478)
Dicotyledon
99
AtINS(AAP04082)
CoBX1(ACJ02769)
ZmTSA(ACA25187)
ZmTSAlike(DAA51946)
Monocotyledon
ZmIGL(AAG42687)
53
100
ZmBX1(AFW60209)
100
TaBx1(BAC81205)
0.05

Complementation of *E. coli ΔtnaA ΔtrpA* by PtIGL-short

E. coli Δ*tnaA* Δ*trpA* was constructed from FS1576 strain through the homologous recombinantory insertion of antibiotic resistance gene (Jin 2013) inside *tnaA* and *trpA.* We introduced *tnaA* mutation in addition to *trpA* mutation to eliminate chance of generating free indole from tryptophan. *TrpB* could salvage indole for tryptophan synthesis (Swift and Stewart 1991) to complicate the interpretation of complementation experiment (Fig. 1.2). The knock-out of *tnaA* also minimized ambiguity when indole production was assayed by color reaction with Kovac's reagent. The resulting *E. coli* Δ*tnaA* Δ*trpA* grew on LB medium containing kanamycin and chloramphenicol, confirming the insertion of the resistance genes.

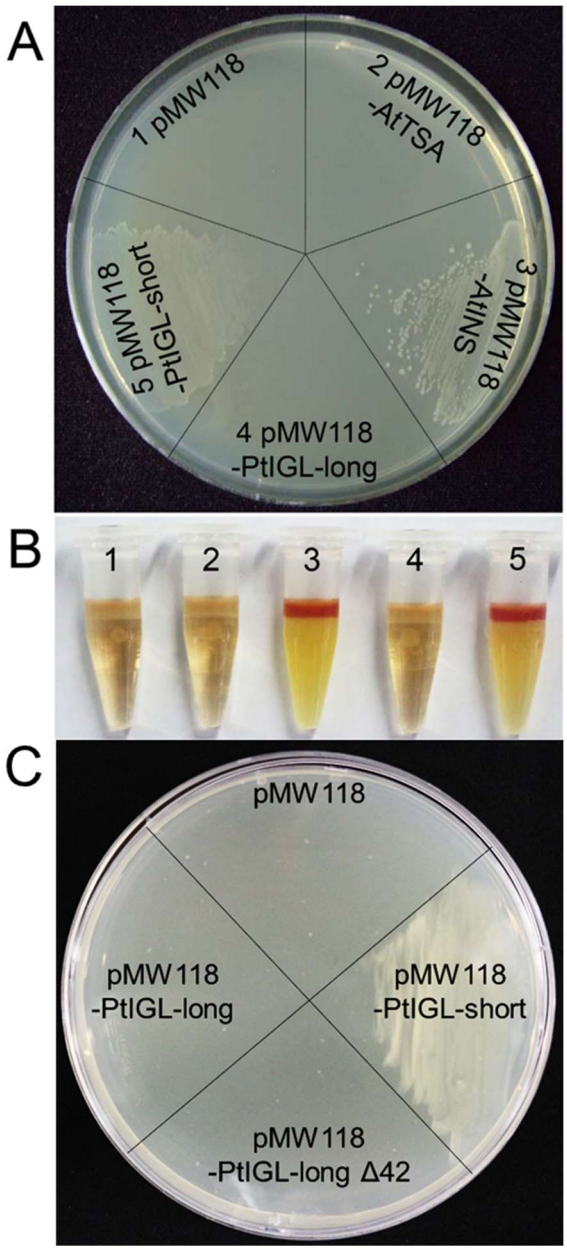
E. coli Δ*tnaA* Δ*trpA* mutant was transfected with the *PtIGL*-harboring *pMW118* plasmids. The transformants harboring *pMW118* vector, *pMW118-AtTSA*, *pMW118-PtIGL*-long or *pMW118-PtIGL*-longΔ42 did not grow on tryptophan-free M9CG medium (Vega et al. 2012), whereas the plasmid with *AtINS* and *PtIGL*-short could rescue the double mutant (Fig. 1.6A, C). The reaction with Kovac's reagent also indicated that PtIGL-short as well as AtINS produced indole (Fig. 1.6B).

To assess the catalytic activity of the cloned *PtIGL* genes, complementation assay using *E. coli* mutant was done. The complementation experiment took advantage of *E. coli* trpB dimer (β_2) catalyzing tryptophan synthesis from free indole and serine (Swift and Stewart 1991) although TSA, in general, is active only in the presence of TSB (Leopoldseder et al. 2006). Because plant TSA cannot form catalytically active complex with *E. coli* trpB (Elaine R. Radwanski et al. 1995), *PtIGL*-long and its truncated form did not complement *E. coli* Δ*tnaA* Δ*trpA*. *PtIGL*-long was thereby tentatively labeled as *PtTSA.* In addition to the complementation experiment, the high homology of the long form with plant TSA and its localization in chloroplast (*vide infra*) further supported TSA-nature of the long form. On the other hand, *PtIGL*-short successfully rescued *E. coli* Δ*tnaA* Δ*trpA* and produced indole in culture medium (Fig. 1.6). Therefore, the short form protein generating free indole as

standalone was clearly identified as indole synthase (PtINS). High amino acid homology between PtINS and PtTSA Δ42 at 94% (Fig. 1.3) suggests that PtTSA Δ42 by itself could also cleave IGP. However, PtTSA Δ42 failed to rescue the *E. coli* double mutant, indicating dependency of PtTSA Δ42 on TSB for catalytic activity.

Figure 1.6. Complementation assay of *PtIGL*s on tryptophan-depleted M9CG plate.

A, *PtIGL*-short and *AtIGL* could rescue the double mutant, *E. coli* Δ*tnaA* Δ*trpA,* whereas *AtTSA* and *PtIGL*-long could not.
B, The double mutant and the mutant transformed with *PtIGL*-long or *AtTSA* could not produce indole as visualized by reaction with Kovac's reagent, whereas transformation with *AtINS* or *PtIGL*-short could.
C, *E. coli* Δ*trpA* Δ*tnaA* harboring pMW118-*PtIGL*-long Δ42 could not grow on M9CG plate without indole.


A
1 pMW118
2 pMW118 -AtTSA
3 pMW118 -AtNS
4 pMW118 -PtIGL-long
5 pMW118 -PtIGL-short
B
1
2
3
4
5
C
pMW118
pMW118 -PtIGL-long
pMW118 -PtIGL-short
pMW118 -PtIGL-long Δ42

PtIGL transcript levels among plant organs

To assess transcript levels of two forms of *PtIGL*, two PCR primer pairs were designed. One was based on the sequence common to both forms, and the other was designed specifically for *PtIGL*-long. The latter consisted of a forward primer corresponding to long form-specific sequence upstream of TSS (the first ATG) and a reverse primer based on sequence between the first and the second ATG. In the roots and the stem, the short form *IGL* predominated in number by a factor of 2 compared to the long form (Fig. 1.7), and in the leaves further by a factor of 5. The absolute copy numbers of both forms were highest in the leaves compared to roots and shoot (Fig. 1.7).

Change in *IGL* transcription upon BTH treatment

The BTH (2,1,3-Benzothiadiazole) is a plant elicitor functionally analogous to salicylic acid to enhance the accumulation of phenolic compounds (Iriti et al. 2004; Hukkanen et al. 2007). In wheat, for example, BTH treatment induces disease resistance through activation of PAL (phenylalanine ammonia lyase) leading to accumulation of toxic phenylpropanoids which inhibits infection by *Blumeria graminis* (Stadnik and Buchenauer 2000). BTH treatment to *P. tinctorium* caused transcript level of *PtIGL*-long in the leaves decreased by 70% at 3 days after treatment, when the indigoid content is known to rebound after initial decrease (Jin et al. 2012), whereas little change was observed in the roots and stem. However, the transcript level of *PtIGL*-short followed opposite trend compared to the long form. While the short form transcript levels in the root and stem increased dramatically, the level in the leaves did not change appreciably upon BTH treatment (Fig. 1.7). However, if the ratios between the numbers of the short form transcript over the long form before and after BHT treatment were compared, striking trend emerged. The ratio between the short over the long form after the treatment became 10, 15, and 15 for roots, stem, and leaves, respectively, from 2, 2, and 5 before the treatment.

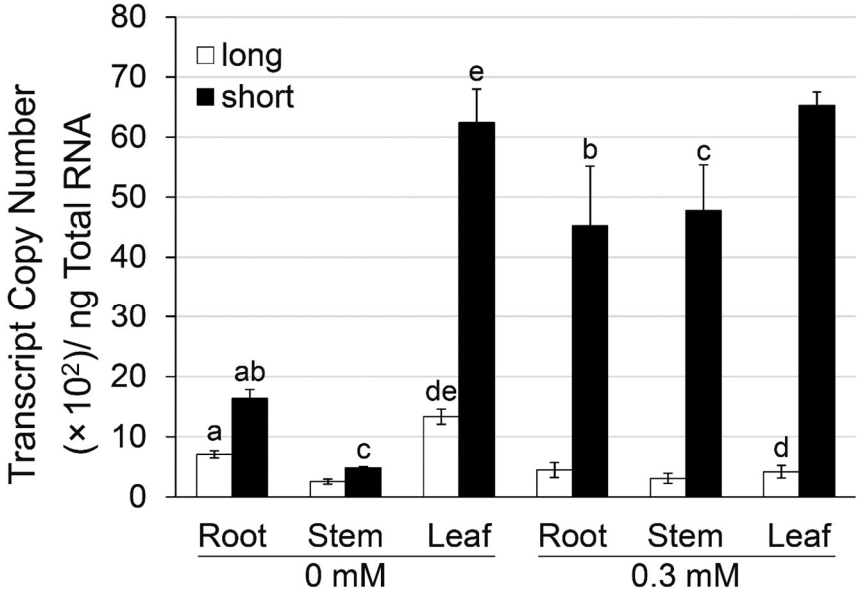
Elicitation experiment further supports role of PtINS in indigoid metabolism. Treatment of BTH to *P. tinctorium* is known to enhance indigoid accumulation by a factor of two (Jin et al. 2012). At resting state, the transcript level of *PtINS* was highest in the leaves of *P. tinctorium* (Fig. 1.7), where indigo precursor indican accumulation is also highest among *P. tinctorium* organs (Minami et al. 2000). Upregulation of *PtINS* transcription in stem and root may indicate increased accumulation of indican also in these plant organs. BTH treatment to the plant significantly increased relative transcript level of *PtINS* over *PtTSA* by a factor of three in the leaves. BTH treatment to plants is known to cause retardation of growth with concomitant acquisition of systemic acquired resistance (SAR), the process known as allocation of fitness costs (Heil et al. 2000). Therefore, the decrease in *PtTSA* transcription upon BTH treatment could represent depressed tryptophan synthesis, a process in primary metabolism. At the same time, relative upregulation of *PtINS* transcription, correlating with enhanced indigoid biosynthesis, reflected increased secondary metabolism as a part of SAR.

A number of evidence is arguing against involvement of PtINS in auxin biosynthesis. Most of all, comparatively lower contribution of auxin through tryptophan-independent auxin synthesis to total auxin reservoir, as suggested by (Wang et al. 2015), is not compatible with high *PtINS* transcript level. In the case of Arabidopsis, *INS* transcript level was lower than *TSA1* level in rosette by a factor of seven (Zhang et al. 2008), whereas in *P. tinctorium INS* transcript level was higher than *TSA* by a factor of five (Fig. 1.7). The second argument comes from evolutionary consideration. If the tryptophan-independent auxin production is common to most plant species, appearance of INS orthologous for the auxin production must have occurred before speciation of dicot plants. On the contrary, phylogenetic tree of functionally characterized TSA-like proteins indicated that separation of dicot IGL clade from monocot clade preceded separation of AtINS from TSA clade as pointed out by (Nonhebel 2015) (Fig. 1.5). Furthermore, PtINS

and PtTSA belong to a same subclade signifying that the duplication of PtTSA to gain new INS function must had occurred rather recently after speciation of *P. tinctorium*. Recent duplication of TSA in *P. tinctorium* is consistent with the fact that *P. tinctorium* is the only known indigo-producing species in *Polygonum* (Fig. 1.5). Then, what would be the possible role of PtINS, if not a part of tryptophan-independent auxin biosynthesis? The preceding argument suggests that PtINS is supplying free indole that is specifically required by *P. tinctorium*, and we know such a need: indigoid biosynthesis.

Figure 1.7. Changes in transcript numbers in various organs of *P. tinctorium* three days after BTH treatment.

0 mM, control; 0.3 mM, BTH treated. The same letter denotes significant difference between two bars at $p<0.05$ (a and b) and $p<0.01$ (c, d, and e). Biological sample was four replicates.


Transcript Copy Number
(× 10²)/ ng Total RNA
80
70
60
50
40
30
20
10
0
long
short
a
ab
c
de
e
b
c
d
Root
Stem
Leaf
Root
Stem
Leaf
0 mM
0.3 mM

Intracellular localization of PtIGLs

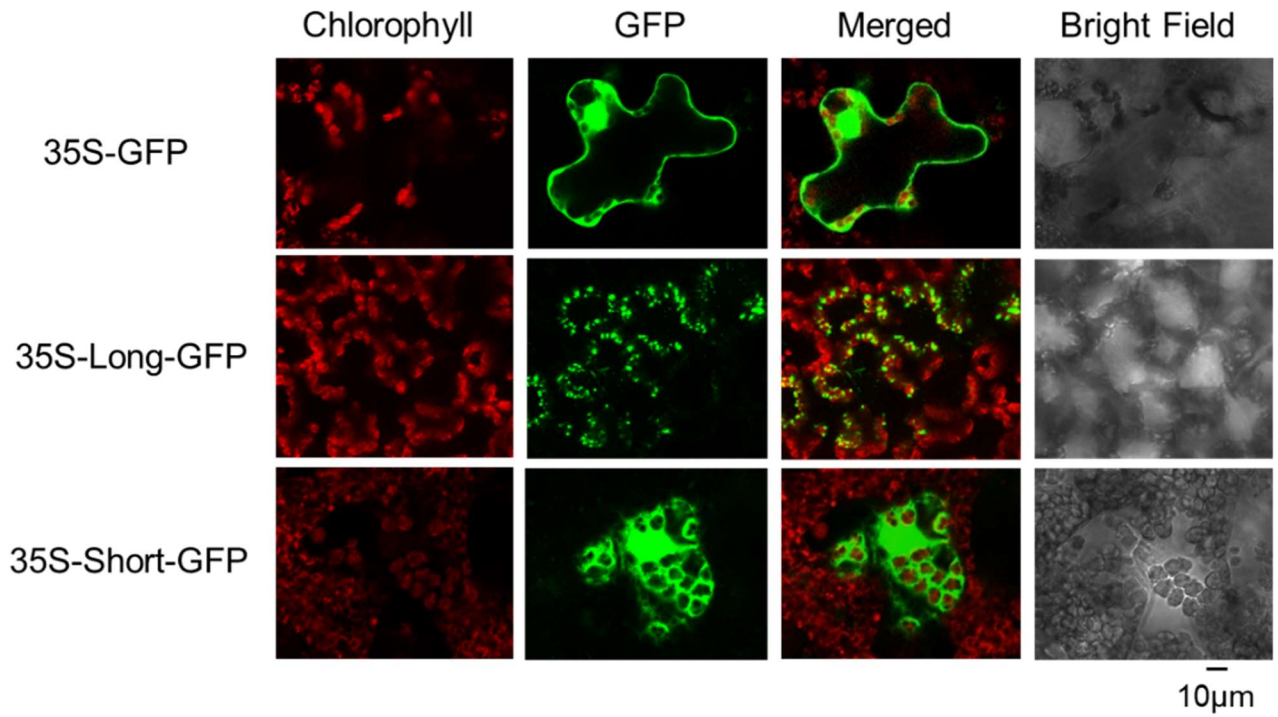
ChloroP predicted the presence of 42 amino acid residue-long putative transit peptides for PtIGL-long (Fig. 1.3). To experimentally verify the computational prediction, GFP fused to the C-terminal of each form was transiently expressed in tobacco leaves. It was found that PtIGL-long was delivered into chloroplasts, whereas the short form was targeted to cytosol (Fig. 1.8).

Distinctive role of each form was again suggested by their subcellular localization. Tryptophan synthesis in plants takes place in plastid (Elaine R. Radwanski et al. 1995), therefore TSA is localized in plastid. TSA homologs with chloroplast targeting sequence have also been recorded and some of them were characterized; ZmBX1 (Frey et al. 2009), ZmIGL (Frey et al. 2009), and OsIGL (Zhuang et al. 2012). The chloroplast-targeting IGLs other than TSA so far known are suggested to produce volatile indole (ZmIGL and OsIGL) or provide benzoxamide precursor (ZmBX1). PtTSA was also delivered into chloroplast in the present study (Fig. 1.8), thus suggesting a role in tryptophan biosynthesis. However, there are several IGLs that are cytosolic. For example, AtINS (Zhang et al. 2008) is shown to be localized in cytosol. ZmTSAlike is also localized in cytosol but it does not exhibit *in-vitro* activity (Kriechbaumer et al. 2008). AtINS and its sequence homologs are found exclusively in Brassicaceae (Nonhebel 2015). PtINS in the present study is thus the only cytosolic INS found out of Brassicaceae up to date. Because indican is likely to be synthesized in cytosol and stored in vacuole (Minami et al. 2000), it is logical to posit that indican precursor, indole, is produced in cytosol. Therefore, PtINS with cytosolic targeting and indole generating activity is qualified to play a role in indigoid biosynthesis. Recently, AtINS is suggested to generate indole for tryptophan-independent auxin biosynthesis (Wang et al. 2015). (Nonhebel 2015) instead proposed that free indole generated by AtINS might be removed by a type 2 TSB (Yin et al. 2010) to synthesize tryptophan for an unknown physiological function or used in the biosynthesis of Brassicaeae-specific indole glucosinolates.

The splicing of *PtINS* pre-mRNA resulted in alternative TSS and change of subcellular distribution of PtINS from chloroplast to cytosol.

Figure 1.8. Subcellular localization of PtIGL-long and -short visualized by transient expression as GFP-fused protein.

Confocal microscopic images of *N. benthamiana* leaf infected with *Agrobacterium* harboring the GFP fusion constructs. GFP emission, auto-fluorescence of chloroplast, and their merged images are shown. All images were taken at the same magnification. Scale bar, 10 μm.


Chlorophyll
GFP
Merged
Bright Field
35S-GFP
35S-Long-GFP
35S-Short-GFP
10μm

Part II: Cloning and functional analysis of three sesquiterpene synthases identified by transcriptome sequencing of peppercorn

Introduction

Piper nigrum L.

Black pepper, dried fruit of *Piper nigrum* L. belonging to Piperaceae family, has been one of the most valued spices in the world from ancient times. It was found with the mummy of Ramesses II of the ancient Egypt, thus confirming its historic uses at least as early as 13th century BCE (Fitzgerald 2008). According to Pliny the Elder in his *Natural History,* black pepper, costing as high as 4 denari (1 denari = 3.4 g of silver) per pound, was already wide-spread seasoning in Roman Empire (Pliny the Elder 1979). *P. nigrum* is native to India, but its leading producer nowadays is Vietnam which produces 40% of the world-wide production of around 523,000 metric ton (Nedspice 2017). It has been used in both cooking and traditional medicine. Black pepper is claimed to cure constipation, toothaches, oral abscesses and sunburn, among others (Turner 2008).

Transcriptome sequencing

For non-model plants, such as black pepper, with little or no molecular information available, next-generation sequencing (NGS) technologies offer a great opportunity for the rapid access to DNA sequence information. Transcriptomics data mining is recently facilitating the discovery of genes encoding enzymes involved in various metabolic pathways as high-throughput next-generation sequencing technologies become popularized. NGS technology provides RNA sequences on a massive scale with considerable sequence depth. Despite these advantages, the sequence reads obtained from NGS are often short (36-400bp) compared with Sanger sequencing (>700 bp) (Goodwin et al. 2016).

Black pepper sesquiterpenoids

The pepper fruit, *aka* peppercorn, contains various aroma-rendering terpenes, mostly

mono- and sesquitepenes, including pinene, limonene, nopinene, caryophyllene, copaene, and cadinene. Recently, rotundone, a sesquiterpene of guaiane skeleton, was identified to contribute peppery aroma to pepper and Shiraz (or Syrah) wine (Siebert et al. 2008).

Terpenes or terpenoids of plant origin are the largest and structurally most diverse group of natural products. Their carbon skeleton was known to originate from condensation of two simple isomeric five-carbon precursors, dimethylallyl diphosphate and isopentenyl diphosphate. For example, sesquiterpene with fifteen-carbon skeleton arises from farnesyl diphosphate precursor, composed of three C_5 units, by sesquiterpene synthases (sesqui-TPS). To understand the make-up of volatile terpene components in black pepper, it is necessary to characterize the TPSs that biosynthesize the terpenes in peppercorn.

The purposes of this study

In the present study, the non-model plant transcriptome sequencing was utilized to identify new terpene metabolizing genes. The next generation sequencing (NGS) analysis of the black pepper transcriptome can greatly enhance our knowledge in the Piperaceae family, which is known as basal angiosperms with little available genomic data. With the aid of black pepper transcriptome data, genes responsible for the secondary metabolism in black pepper were cloned and characterized. In particular, from transcriptome data of unripe green pepper fruit, three full-length cDNA clones putatively coding sesqui-TPSs and their enzymatic functions were identified.

Materials and Methods

Plant material and growth conditions

Fresh immature fruits of *P. nigrum* and other organs (root, stem and leaves), obtained from garden of Prince of Songkla University, Thailand, were kindly provided by Professor Juraithip Wungsintaweekul.

Nicotiana benthamiana for transient expression was grown on Biosoil (Hungnong Seeds®, Pyeong taek, Korea) until 5 leaf-stage in a plant growth chamber with 16 h light /8 h dark cycle at 24°C.

Bacterial, yeast strains and culture media

Bacterial strains used in this experiment was *Escherichia coli* DH10B [F^- *mcrA Δ*(*mrr-hsdRMS-mcrBC*) *Φ80d lacZΔM15 ΔlacX74 recA1 deoR Δ(ara,leu)7697 araD139 galU galK rpsL* λ^- *endA1 nupG*] for subcloning and *E. coli* Rosetta2(DE3) [F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)] for protein expression. Engineered yeast strain of EPY300 (S288C, MATα his3Δ1 leu2Δ0 PGAL1-tHMG1::δ1 PGAL1-upc2-1:: δ2 erg9::PMET3-ERG9::HIS3 PGAL1-ERG20::δ3 PGAL1-tHMG1::δ4) (Ro et al. 2006) was employed for *in-vivo* sesquiterpenoids production.

E. coli were grown in Luria-Bertani (LB) medium (MBcell, Korea) and solid LB medium plate was made with 1.5% (w/v) Micor-agar (Duchefa, Netherlands). Plasmid transformed strain were cultured in LB broth with ampicillin (100 μg/ml) or kanamycin (50 μg/ml). Yeast was cultured at Yeast Dropout Medium Supplements and Minimal SD Base from Clonetech (Madison, WI), Bacto™ Agar from Becton Dickinson.

Enzyme and chemicals

Restriction enzymes, T4 DNA ligase, Ex Taq polymerase, Phusion® High-Fidelity DNA Polymerase were purchased from TaKaRa (Japan) or NEB (USA). Farnesyl diphosphate (FPP) was from Echelon Biosciences (USA) and *n*-hexane from Duksan (Korea). Acrylamide/Bis solution (40%) to cast SDS-PAGE gel was purchased from Bio-Rad (USA). Adenine hemisulfate, $MgCl_2$, MES monohydrate, and malachite green phosphate assay kit were from Sigma-Aldrich.

Oligonucleotides

Primer pairs used for polymerase chain reaction (PCR) were synthesized by Cosmogenetech (Seoul, Korea). The sequences of the oligonucleotides are listed in Table 2.1.

Table 2.1. List of oligonucleotides used in *PnTPS* research.

Primer	(5'→3')
PnTPS1 ORF F	atgaatatggcttgtgtctctgatct
PnTPS1 ORF R	tcatactggtattgggtccttcaacag
PnTPS2 ORF F	atggatgccgtttcatgtgctat
PnTPS2 ORF R	tcaaacggaatgggatccttgtaca
PnTPS3 ORF F	atgggcttttctttgtaacaaatgct
PnTPS3 ORF R	ttagaggggataggttggacaag
pET21:PnTPS1 F	agaaggagatatacatatgaatatggcttgtgtctctgatct
pET21:PnTPS1 R	tggtggtggtggtgctcgagtactggtattgggtccttcaacag
pET21:PnTPS2 F	agaaggagatatacatatggatgccgtttcatgtgctat
pET21:PnTPS2 R	tggtggtggtggtgctcgagaacggaatgggatccttgtaca
pET21:PnTPS3 F	agaaggagatatacatatgggcttttctttgtaacaaatgct
pET21:PnTPS3 R	tggtggtggtggtgctcgaggaggggataggttggacaag
pESC-Leu2d:PnTPS1 F	aggagaaaaaaccccggatccgatgaatatggcttgtgtctctgatct
pESC-Leu2d:PnTPS1 R	ttagagcggatcttagctagctactggtattgggtccttcaacag
pESC-Leu2d:PnTPS2 F	aggagaaaaaaccccggatccgatggatgccgtttcatgtgctat
pESC-Leu2d:PnTPS2 R	ttagagcggatcttagctagcaaccggaatgggatccttgt
pESC-Leu2d:PnTPS3 F	aggagaaaaaaccccggatccgatgggcttttctttgtaacaaatgct
pESC-Leu2d:PnTPS3 R	ttagagcggatcttagctagcgaggggataggttggacaag
pEAQ-HT:PnTPS1 F	ctgcccaattcgcgaccggtatgaatatggcttgtgtctctgatct
pEAQ-HT:PnTPS1 R	tggtgatggtgatgcccgggtactggtattgggtccttcaacag
pEAQ-HT:PnTPS2 F	ctgcccaattcgcgaccggtatggatgccgtttcatgtgctatc
pEAQ-HT:PnTPS2 R	tggtgatggtgatgcccgggaaccggaatgggatccttgtaca
pEAQ-HT:PnTPS3 R	ctgcccaattcgcgaccggtatgggcttttctttgtaacaaatgct
pEAQ-HT:PnTPS3 R	tggtgatggtgatgcccggggaggggataggttggacaag
PnTPS1 QRT-F	aatatggcttgtgtctctgatcttg
PnTPS1 QRT-R	ggatgaaaagctgcagatcgac
PnTPS2 QRT-F	tggatgccgtttcatgtgcta
PnTPS2 QRT-R	tggaaaagtaaccgacttccgg
PnTPS3 QRT-F	cgaccttgaccactgattcct
PnTPS3 QRT-R	gatgcatggtacccttgctg

Total RNA isolation and cDNA preparation

Total RNA was isolated from immature black peppercorn through the through CTAB method (Chang et al. 1993). The total RNA was treated with RNase-free DNase I (PhileKorea, Korea) and purified by phenol-chloroform methods (Tozawa et al. 2001). Purified total RNA (~30 μg) was evaluated for integrity on Bioanalyzer 2100 (Agilent, USA). Samples with RNA Integrity Number (RIN) higher than 7.0 was used in further experiment (Schroeder et al. 2006). Illumina cDNA library was prepared using total RNA by Beijing Genomics Institute Genomic Center in Hong Kong, and resulting cDNA library was sequenced on Illumina HiSeq 2000 platform.

Subsequently, cDNA was prepared using cDNA synthesized (PhileKorea, Korea) for gene cloning. Two micrograms of purified total RNA were reverse-transcribed using an oligo (dT) primer with the following PCR conditions 45 min for 42℃, 10 min 70℃ to denature RTase activity, finally hold at 16℃ conditions to complete the synthesis of the first strand cDNA.

Isolation of sesqui-*TPS*s

With amino acid sequence of amorpha-4,11-dinene synthase (Genbank accession number ABM88787) (Chang et al. 2000) as query, homologous gene contigs were searched in transcriptome database using tBLASTn (BioEdit, version 7.2.0). Three contigs, suggested to code full-length open reading frame with high homology against the query gene, were selected. Based on the sequence information of the contigs, primer pairs were designed (Table 2.1.) to clone respective ORFs from cDNA library under the following conditions using Ex Taq Polymerase (Takara, Japan): 98℃ for 5 min for initial, 98℃ for 15 s, 58℃ for 30 s, 72℃ for 2 min for 35 cycles and 72℃ for 5 min and 16℃ hold for final. PCR products were purified by using Inclone™ Gel & PCR purification kit (Inclone Biotech, Korea) and finally TA cloned into *pMD20* T-vector (Takara, Japan).

Yeast transformation and fermentation

Yeast expression plasmid was constructed by inserting sesqui-*TPS* ORF without stop codon into the yeast expression vector *pESC*-LEU2d at *Bam*HI and *Nhe*I double digestion site. Engineered yeast EPY300 (Ro et al. 2006) was then transformed by using LiAc/SS carrier DNA/PEG method (Gietz and Schiestl 2007). All transformants were grown on SC medium (Nguyen et al. 2012) lacking histidine (H), methionine (M), and leucine (L) incubate at 30℃ select positive transformants. Inoculation of single colony of transformants yeast in a test tube with 3 ml of lacking HML with 2% glucose for 16h. Transfer 600 µl of the inoculation to a 150 ml flask containing 30 ml of lacking HML with 2% galactose, 0.2% glucose and 1mM methionine, 30℃, 200 rpm, three days of shaking incubator. The culture medium was extracted with 3ml of *n*-hexane.

Transient expression in *N. benthamiana*

pEAQ-HT binary vector was employed for transient expression of *PnTPS*s in *N. benthamiana*. The ORF of terpene synthase was cloned between *Age*I and *Xma*I double digested enzyme sites on the vector. The transient expression of sesqui-TPS in *N. benthamiana* were performed as described by using *Agrobacterium tumefaciens* LBA4404 (Sainsbury et al. 2009). At fifth day of cultivation in the plant growth chamber after agro-infiltration, two gram of leaves were grinded liquid nitrogen used mortar and pestle, extracted with 2 ml *n*-hexane mixed at shaker of 100 rpm 10 min. Centrifugation of 5,000 rpm (Beckman, F1202 Rotor) for 6 min at 4℃ collected organic phase.

Heterologous expression and *in-vitro* assay

ORF of each *PnTPS* was cloned into *pET21*a(+) vector at *Nde*I and *Xho*I double digestion site with C-terminal 6×His tag by using Gibson Assembly® Cloning Kit (NEB, USA). The ORF-containing vector was then transformed into *E. coli*

Rosetta2(DE3). Single colony was taken to inoculate 5 ml LB medium supplemented with 100 μg/ml ampicillin and incubated at 37℃ overnight. The overnight culture was inoculated 500 μl to 50 ml fresh LB medium with the same concentration of antibiotics. Incubation was continued at 37℃ until OD_{600} reached 0.5, when the culture was added with IPTG to 0.1 mM and the incubating temperature was lowered to 20℃. The culture was further incubated for 16 h, and the cells were collected by centrifugation at 5,000 rpm (Beckman F1202 Rotor) for 5 min at 4℃. The cell pellet was resuspension in 5 ml of lysis buffer (20 mM Tris-HCl, pH 7.0, supplemented with 300 mM NaCl and 10 mM imidazole). The suspension was then sonicated (Sonic Dismembrator 550, Fisher Scientific) for total of 10 min with repeated 3 s on and 1 s off interval. Soluble protein fraction was collected by centrifugation for 30 min at 7,000 rpm and 4℃. The supernatant was applied to 60 Ni Superflow Resin (Takara, Japan) and the TPS fraction was eluted by increasing concentration of imidazole. Finally, the eluted TPS was desalted by buffer exchange on a centrifugal filter (Amicon® Ultracell-30k, Merck Millipore, Ireland). The proteins were subjected to SDS-PAGE (10% gel) and the purity of the protein was estimated by analyzing the image of the gel using ImageJ2x program version 2.1.4.7 (NIH ImageJ, USA).

The enzymatic reaction was done in glass vial using 1 μg of the purified protein in a final volume of 1 ml of assay buffer (100 mM Tris-HCl, pH 7.0) supplemented with 1 mM DTT (dithiothreitol), 10 mM $MgCl_2$, and 100 μM FPP (Echelon, USA). The reaction mixture was gently overlaid with 0.5 ml *n*-hexane and allowed to stand for 2 h at 30℃. The reaction mixture was then vortexed and the organic phase was separated by centrifugation. The extract was analyzed on GC-MS (*vide infra*).

Steady-state kinetics

Steady-state kinetics of sesqui-TPS was analyzed by determining released free phosphate with malachite green phosphate assay kit (Sigma-Aldrich, Cat. No.

MAK307) (Vardakou et al. 2014). One microgram of the purified sesqui-TPS and FPP (50 μM to 1.56 μM) in 500 μl of assay buffer (100 mM Tris-HCl, pH 7.0) were reacted at 30℃ for 10 min and the reaction was terminated by addition of 10 μl of 5 M HCl. Two hundred microliters of the enzyme reaction mixture were mixed with 50 μl of working reagent (1 ml of Reagent A and 10 μl of Reagent B from the kit) and incubated for 30 min at room temperature before determining A_{620} on a spectrophotometer (X-ma 3000PC, Human Corp., Korea). Standard curve was constructed with the inorganic phosphate in the kit (0-40 μM). The data was fit to the Michaelis-Menten equation by using SigmaPlot 13 (Systat Software, USA) to obtained V_{max} and K_m.

GC-MS analysis

GC System (Agilent model 6890, USA) equipped with a mass spectrometer was used to analyze the terpene products under the following conditions: He at 1 ml/min, 1μl injection, a Zebron ZB-5MSi capillary column (60 m × 0.25 mm × 0.25 μm), injector temperature at 220℃, and a temperature program from 50℃ to 110℃ at 5℃/min (5 min hold) and then at 10℃/min to 240℃ (10 min hold). Fruit total extract was analyzed by using temperature program of 50℃ to 260℃ at 4℃/min.

The coupled mass spectrometer was a Mass Selective Detector (Hewlett-Packard model 5973, USA) with transfer line temperature set at 320℃, source temperature at 250℃, quadrupole temperature at 150℃, and ionization potential at 70 eV. Scan range was 50 to 300 atomic mass units. Products were identified by using MS Search Program v.2.0 (NIST Standard Reference Database, USA).

QRT-PCR

Transcript levels of *PnTPS*s were determined by QRT-PCR using primer pairs described on Table 2.1. The PCR in a total reaction volume of 20 μl containing 10 μl of 2×QuantiMix SYBR Kit PCR system (PhileKorea, Korea), 1 μl diluted cDNA,

0.25 µM of each primer, and double distillated H_2O. The PCR reaction was run on Rotor-Gene 2000 Real Time Cycler (Corbett Research, Australia) with a temperature program of 5 min at 95°C, 40 cycles of 15 s at 95°C, 20 s at 58°C, and 20 s at 72°C. The standard curve was made by running the PCR of each gene ranging from 10^2 to 10^8 copy/µl, and the copy numbers were calculated as described (Yin et al. 2001). Each data point was obtained from five biological samples, each composed of four technical replicates.

Bioinformatics analyses

The alignment of nucleotide and amino acid sequences was performed using the Vector NTI Advance® software (Thermo Fisher Scientific, USA). ChloroP (Emanuelsson et al. 1999; Emanuelsson et al. 2007) was used to predict the subcellular localization of protein. The phylogenetic tree was constructed with MEGA6 (www.megasoftware.net) and sequence analyses (translation, reverse translation, and restriction summary) were done at www.bioinformatics.org/sms2/ web-site.

Results and Discussion

Transcriptome analysis

Sequencing of the immature peppercorn transcriptome yielded total of 49,413,668 clean reads (Table 2.2.). All clean reads were *de novo* assembled into contigs by the Trinity method (Grabherr et al. 2011). The clean read assembly afforded 119,354 contigs with all the isoforms included. These contigs represented a total of 66,787 unigenes, the length ranging from 300 to 3000 nt with a mean length of 647 nt, for downstream analysis. The number of unigenes longer than 1 kb was 15,101 (22.6% of the total unigenes). The quality control of transcriptome sequencing error ratio (Q20 value), the results of 95.67% was produced in this study (Table 2.2.). This result was comparable to the previously published transcriptome data from black pepper roots (SOLiD; unigene number 10,338, mean length=168) (Gordo et al. 2012), leaves (Illumina; unigene number 128,157, mean length=449) (Joy et al. 2013), and fruits (Illumina; unigene number 44,061, mean length=1,345) (Hu et al. 2015). Hu et al. (2015) discussed lysine/ornithine metabolic genes with respect to piperine biosynthesis.

Table 2.2. Summary of NGS reports generated by immature black pepper fruit transcriptome sequencing.

	Number	Percentage
Raw reads	55,077,258	
Clean reads	49,413,668	
Total clean length	4,447,230,120	
Contig	119,354	
Total unigenes	66,787	
Total length	43,197,636	
300-500 nt	35,002	52.42
600-900 nt	16,684	24.98
1000-2000 nt	13,627	20.40
2100-3000 nt	1,313	1.96
>3000 nt	161	0.24
Mean length	647	
N50 length	875	
GC content		48.26
Q20 percentage		96.57

Screening of sesqui-*TPS*

Search for the sesqui-TPSs from the database yielded nineteen candidate cDNAs, including three full length *TPS*s (Fig. 2.1A). One of the full length contigs, named *PnTPS1* (GenBank accession number KU953957), was composed of 560 amino acids (predicted mass of 64.9 kDa), another contig, PnTPS2 (KU953958) 563 amino acids (64.6 kDa), and the third PnTPS3 (MF104556) 563 amino acids (64.7 kDa) (Fig. 2.1B). Amino acid identities between PnTPS1 and 2 was 55%, but PnTPS3 had 39% identity with the other two TPSs. PnTPSs had well-conserved regions characteristic for sesqui-TPS (Fig. 2.2).

Figure 2.1. A, Agarose gel electrophoresis of three *PnTPS*s isolated by PCR. B, SDS-PAGE of proteins from *E. coli* expressing PnTPSs.

The proteins were tagged with 6×His. PnTPS1, lane 1, 4, 7; PnTPS2, lane 2, 5, 8; PnTPS3, lane 3, 6, 9. The purity of the proteins, 82.7, 80.7, and 95.0% for PnTPS1 through 3 respectively, were estimated by analyzing gel image with ImageJ2x program (version 2.1.4.7).

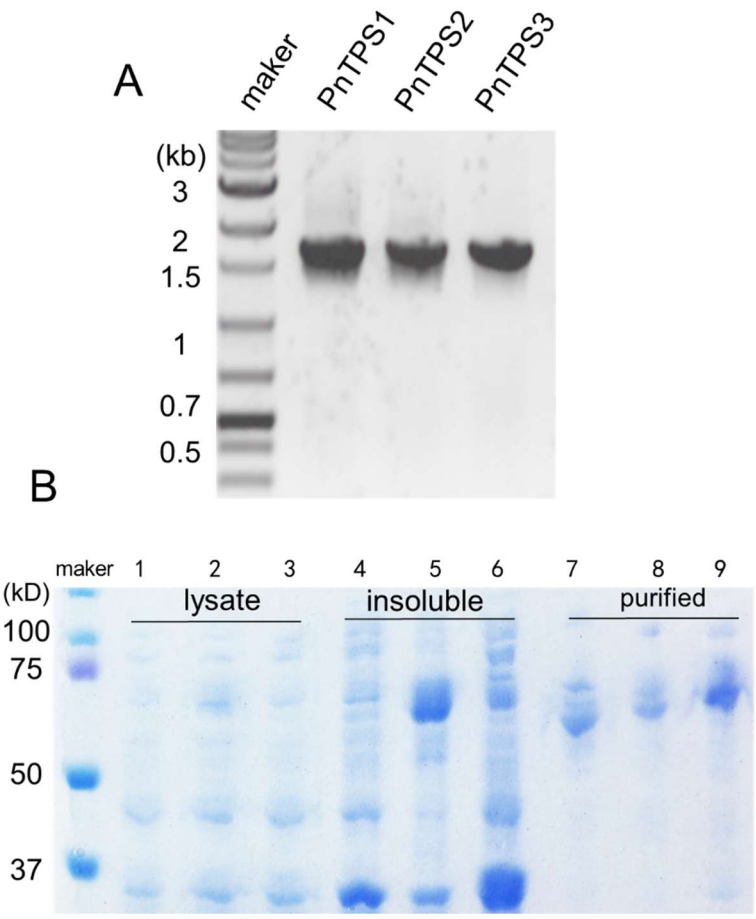
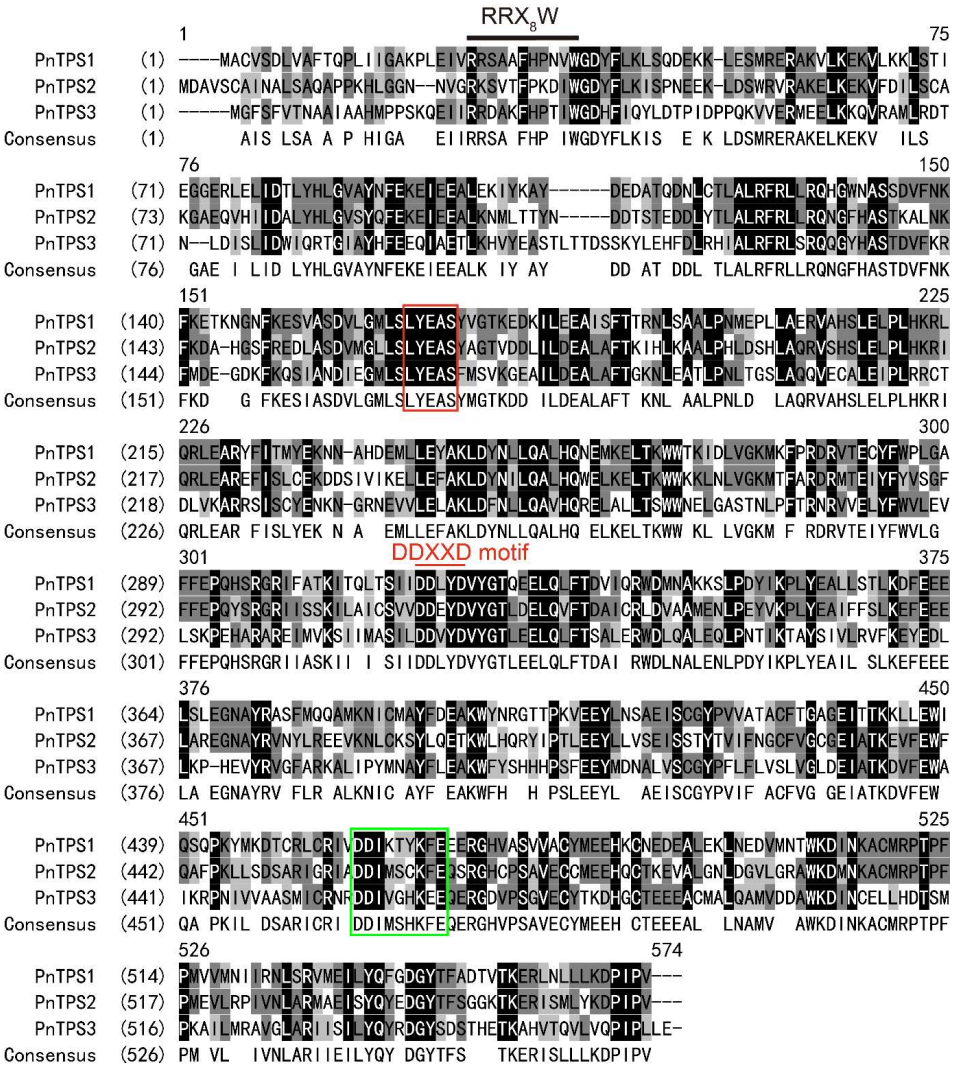

A
maker
PnTPS1
PnTPS2
PnTPS3
(kb)
3
2
1.5
1
0.7
0.5
B
maker
1
2
3
4
5
6
7
8
9
(kD)
lysate
insoluble
purified
100
75
50
37

Figure 2.2. Alignment of PnTPS sequences.

Identities between the TPSs were: PnTPS1 and 2, 55.1%; PnTPS1 and 3, 39.4%; PnTPS2 and 3, 38.2%. Conserved regions known for sesqui-TPS, RRX_8W (black line), DDXXD motif (red line), LYEAS (red box), and DDIX(S,T,G)XXXE (green box) were shown.


RRX8W
DDXXD motif

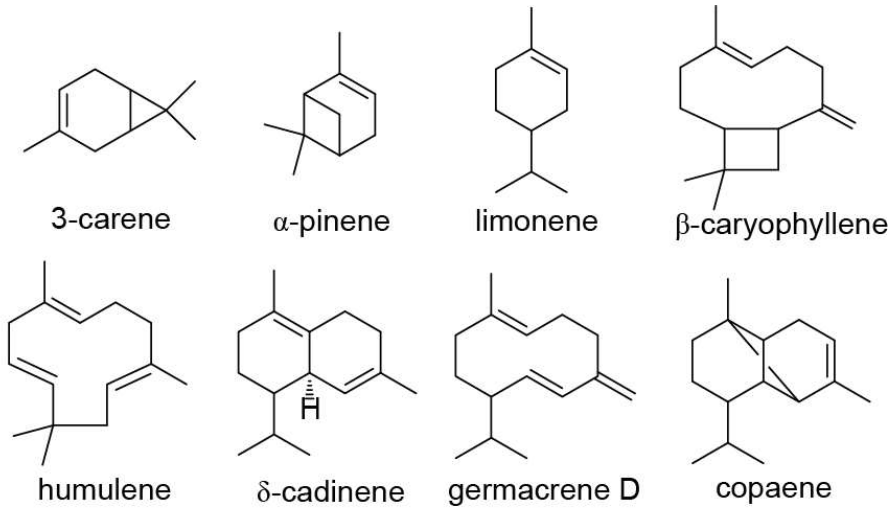
Analysis of sesquiterpenes in pepper fruit

To correlate the terpene composition and the enzyme products, terpene metabolites in the unripe peppercorn were profiled. Identification was based on mass spectrum data and the literature (Jeleń and Gracka 2015). In the fresh unripe fruit, thirteen sesquiterpenes including β-caryophyllene, δ-elemene, α-copaene, cubebene, α-humulene, δ-cadinol, and γ- and δ-cadinenes were found (Fig. 2.3A and B, Table 2.3.). Caryophyllene was the most abundant compound among sesquitepenes (65.8% of the total), followed by copaene (8.3%) (Table 2.3.). Even though the reported composition of peppercorn varies from literature to literature, β-caryophyllene was always the most abundant sesquiterpene. Rotundone or its α-guaiene skeletal homologs were not detected in the present study nor described in the previous analysis reports (Kollmannsberger et al. 1992; Liu et al. 2007; Gupta et al. 2013; Jeleń and Gracka 2015).

Figure 2.3. A, Volatile mono- and sesquiterpenes found in dried green peppercorn. B, GC-MS profiling of terpenes from immature black pepper fruit extract in this study.

Only major representative terpenes are shown (Liu et al. 2007; Jeleń and Gracka 2015).

A


3-carene
α-pinene
limonene
β-caryophyllene
humulene
δ-cadinene
germacrene D
copaene

B

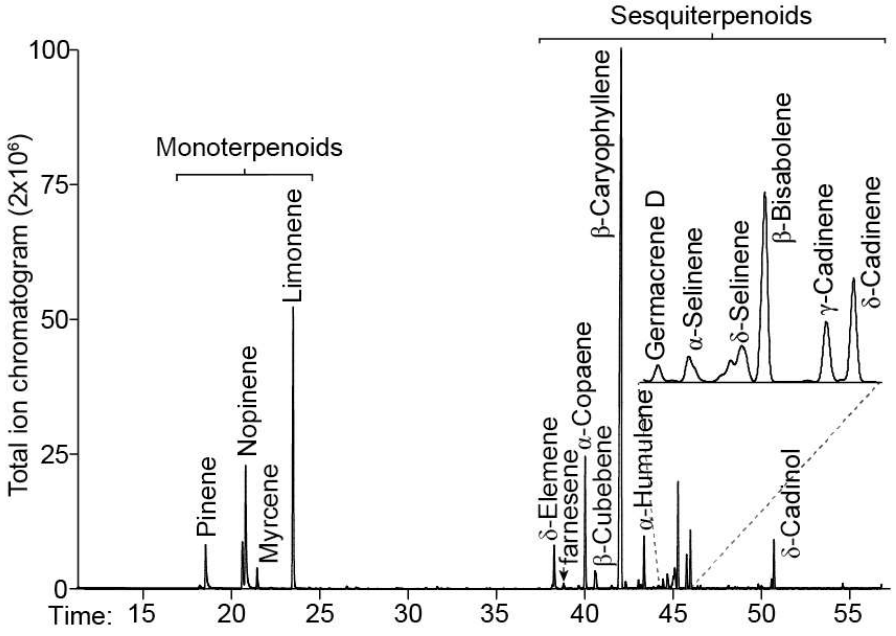

Sesquiterpenoids
Monoterpenoids
Total ion chromatogram (2x10⁶)
100
75
50
25
0
Time: 15 20 25 30 35 40 45 50 55
Pinene
Nopinene
Myrcene
Limonene
δ-Elemene
farnesene
α-Copaene
β-Cubebene
β-Caryophyllene
α-Humulene
Germacrene D
α-Selinene
δ-Selinene
β-Bisabolene
γ-Cadinene
δ-Cadinene
δ-Cadinol

Table 2.3. **Sesquiterpene contents from pepper fruits.**

No.	compound	relative content (%) among total sesquiterpene		
		Liu et al. 2007[a]	Jeleń and Gracka 2015[b]	Present study[c]
1	copaene	0.5	ND	8.3
2	cubebene	2.6	4.4	1.2
3	β-elemene	5.6	3.4	ND
4	δ-elemene	ND	4	2.4
5	β-farnesene	2.0	ND	0.7
6	β-caryophyllene	82.3	54.2	65.8
7	α-humulene	0.1	ND	2.6
8	germacrene D	5.1	ND	0.7
9	α-guaiene	ND	1.7	ND
10	α-caryophyllene	ND	6.7	ND
11	β-eudesmene	ND	5.7	ND
12	α-selinene	ND	4.6	0.9
13	δ-selinene	ND	ND	1.4
13	β-bisabolene	ND	2.9	6.9
14	γ-cadinene	ND	ND	2.2
15	δ-cadinene	1.2	3.6	2.9
16	δ-cadinol	ND	ND	4.0
17	caryophyllene oxide	0.6	4.8	ND
18	ledene oxide (?)	ND	4.0	ND

The sesquiterpenes are listed according to the elution order shown in Figure 2.3B.
a,b; dried peppercorn. c; unripe fresh peppercorn

Functional analyses of sesqui-TPSs

The engineered yeast strain EPY300 that has been used as sesquiterpene production platform (Nguyen et al. 2012) was employed to characterize catalytic function of the putative PnTPSs. In the heterologous yeast system, high level of farnesol (R_t = 28 min) appeared in the background. *PnTPS1*-transformant produced caryophyllene (97%) and trace of humulene (3%), and *PnTPS2* transformant accumulated δ-cadinol as a major product (91%) and smaller amounts of α-cadinene (2%) and δ-cadinene (7%) (Fig. 2.4 left panel and Supplementary Data). The products from *PnTPS3*-transformant were copaene (59%), β-cadinene (38%), and trace of cubebene (3%).

Because acidic condition prevailed (pH 4.6) at the end of yeast fermentation, we conducted *in-vitro* enzyme assay using partially purified enzymes with higher than 80% purity (Fig. 2.1B) to avoid the possible acid-catalyzed rearrangement of initial enzymatic product(s). For PnTPS1 and PnTPS2, the products were same as those from yeast fermentation. This finding allowed us to rename PnTPS1 to PnCPS and PnTPS2 to PnCO/CDS. CPSs have been commonly found from plant (Cai et al. 2002) to microorganism (Nakano et al. 2011). PnTPS2 showed higher α-cadinene (12%) and δ-cadinene (10%) compared to *in-vivo* product composition, and thus was promiscuous in reaction product. δ-Cadinol synthase has been previously described from fungus *Boreostereum vibrans* (Zhou et al. 2016). In the case of PnTPS3, however, the *in-vitro* reaction yielded germacrene D as the major product (91%) and small amounts of copaene (6%) and cubebene (3%). Therefore, PnTPS3 was named PnGDS. GDS have been cloned from various sources such as *Streptomyces* (He and Cane 2004), goldenrod (Prosser et al. 2004), and grape (Lücker et al. 2004).

Next we used *N. benthamiana* transient expression system for the same purpose. *Agrobacterium* LBA4404 strains each harboring sesqui-TPS full length coding sequence under the of 35S promoter of *pEAQ-HT* binary vector were used to infiltrated *N. benthamiana* leaves (Sparkes et al. 2006; Sainsbury et al. 2009). The infiltrated leaves were harvested at seven days after infiltration and analyzed by GC-

MS after extraction (Fig. 2.4 middle panel). PnTPS1 produced caryophyllene against high noise background which did not allow identification of humulene. In the PnTPS2 transformant two cadinene isomers were detected without cadinol being found. PnTPS3, however, produced copaene and cadinene as major product as seen in yeast system. It was therefore evident that the plant system metabolized some of the sesquiterpenes heterologously produced in *N. benthamiana*. Glucosylation of xenobiotic alcohols into corresponding glucoside is one of well-known example of such reactions and could explain lack of cadinol in the leaf extract (van Herpen et al. 2010). To further characterize the initial products of PnTPS reactions decisively, we finally turned to *in-vitro* assay.

Germacrene D is a labile compound to easily undergo rearrangement (Bülow and König 2000; Adio 2009). In particular, in the presence of acid, germacrene D rearranges into copaene and various cadinene-type products (Adio 2009), which explains the appearance of prominent copaene and β-cadinene peaks from the PnTPS3 yeast fermentation. Therefore, the copaene and cadinene found in PnTPS3, either *in-vivo* or *in-vitro*, are most likely the rearrangement products of germacrene D. To confirm the possible rearrangement of germacrene D under the acidic condition, we conducted *in-vivo* enzyme reaction under at pH 6.0 (Fig. 2.5). Reaction at the lower pH clearly favored formation of copaene and β-cadinene. Thus large copaene and β-cadinene peaks in the yeast experiment (Fig. 2.4 lower panel) could be justified by acidification of broth during fermentation. The moderate copaene content and the lack of β-cadinene in unripe peppercorn may reflect relatively neutral cellular environment for PnGDS *in-planta*. This observation on PnTPS3 exemplarily demonstrates that heterologous *in-vivo* product in yeast does not necessarily advocate the identities of initial enzymatic reaction product(s) (Nguyen et al. 2010; Pickel et al. 2012).

We could not identify a gene for guaiene skeleton, which could be a direct precursor to rotundone. The low concentration of rotundone at 1.2 ppm (Siebert et al. 2008)

and δ-guaiene (<0.1%) (Kollmannsberger et al. 1992) in black pepper suggests extremely low abundance of the transcript. Nevertheless, the importance of rotundone in flavor industry warrants further search for the gene.

Figure 2.4. Total ion chromatogram of fermentation broth extract (left column), tobacco transient expression (middle column), and *in-vitro* enzyme assays (right column).

PnTPS1: peak **1**, caryophyllene; peak **2**, humulene. PnTPS2: peak **3**, α-cadinene (cadina-4,9-diene); peak **4**, δ-cadinene (cadina-1(10),4-diene); peak **5**, δ-cadinol. PnTPS3: peak **6**, copaene; peak **7**, cubebene; peak **8**, β-cadinene (cadina-3,9-diene); peak **9**, germacrene D. The chemical structures are illustrated in Figure 2.6.

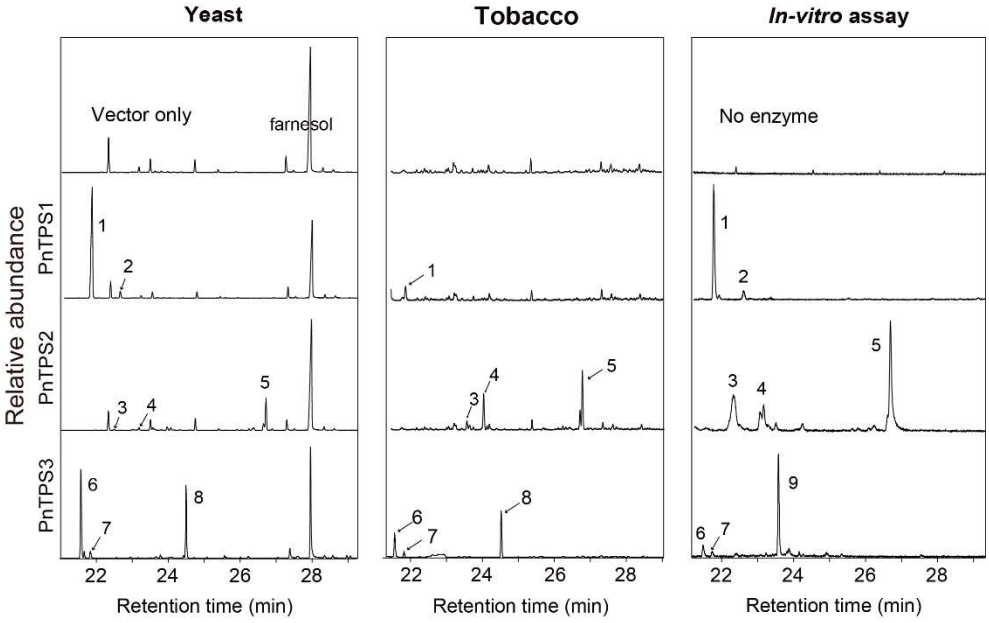
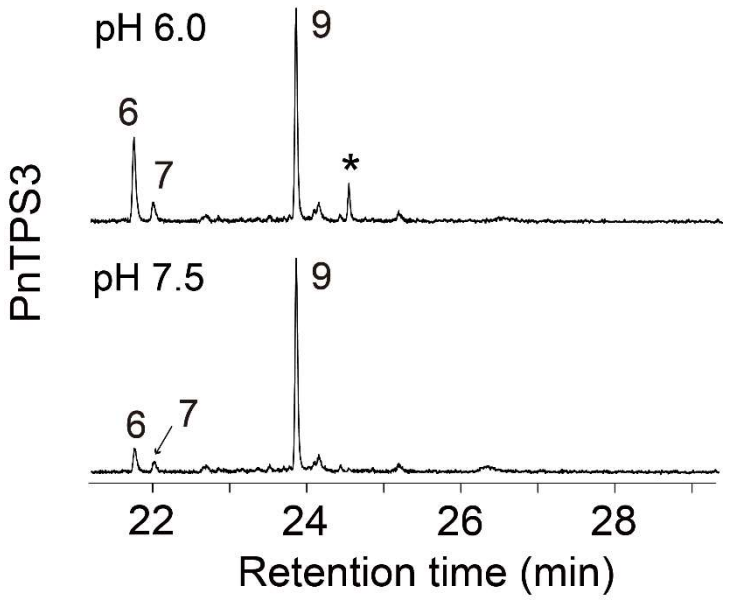

Yeast
Tobacco
In-vitro assay
Vector only
farnesol
No enzyme
Relative abundance
PnTPS1
PnTPS2
PnTPS3
1
2
3
4
5
6
7
8
9
22
24
26
28
Retention time (min)

Figure 2.5. Total ion chromatogram (TIC) of PnTPS3 reaction product under different pH conditions.

Peak 6, copaene; peak 7, cubebene; peak 9, germacrene D. Asterisk indicates β-cadinene (cadina-3,9-diene). Chromatogram for reaction at pH 7.0 is taken from Fig. 2.2.


pH 6.0
9
6
7
*
PnTPS3
pH 7.5
9
6
7
22
24
26
28
Retention time (min)

Kinetic parameters

Steady-state kinetic parameters of sesqui-TPSs were determined by measuring the release of diphosphate from FPP on cyclization of the substrate into terpene products (Vardakou et al. 2014). The enzymes had K_m values ranging from 9 to 21 μM and k_{cat} values from 0.17 to 0.56 s^{-1} (Table 2.4). The affinity to substrate and the turn-over numbers were within the range of other sesqui-TPSs (Vardakou et al. 2014; Manczak and Simonsen 2016).

Catalytic mechanism

Mechanistic pathways for three PnTPSs to explain the formation of various products are depicted in Fig. 2.6. Caryophyllene synthases that co-produce humulene have been previously described (Wang et al. 2008). The cyclization of FPP to form caryophyllene by PnGDS apparently calls for so-called 1,10-ring closure. Here we depicted cyclization process catalyzed by PnTPSs follows 1,10-closure mechanism assuming that the present PnTPSs have evolved from a common ancestor to share common mechanistic feature. Nevertheless, 1,6-ring closure mechanism is also possible for PnCO/CDS (Faraldos et al. 2012).

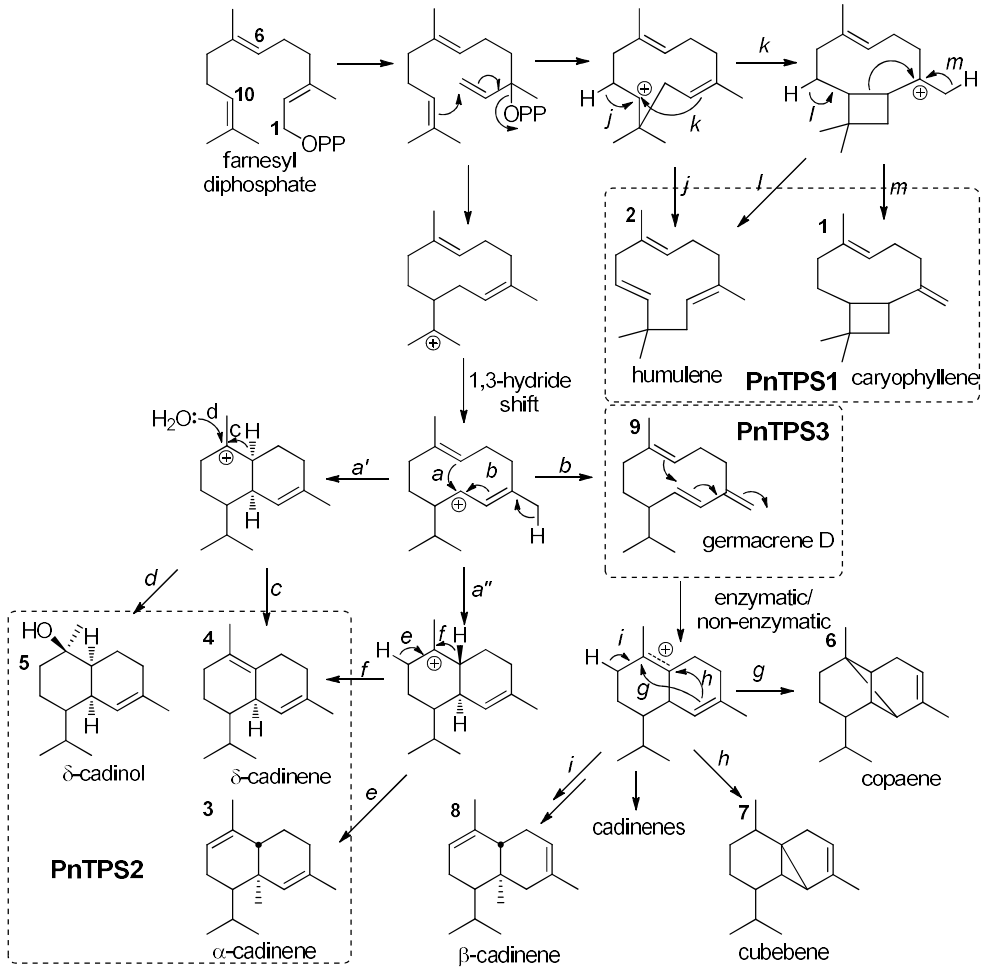
Table 2.4. **Kinetic properties of three PnTPSs.**

Data were from three replicate assays.

	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} μM^{-1})
PnTPS1	21.06 ± 4.71	0.455 ± 0.003	0.0216
PnTPS2	18.40 ± 4.94	0.694 ± 0.066	0.0377
PnTPS3	9.15 ± 1.81	0.185 ± 0.012	0.0202

Figure 2.6. Possible cyclization mechanism of sesquiterpene synthases from *P. nigrum*.

The reaction pathway was based on 1,10-ring closure pathway.

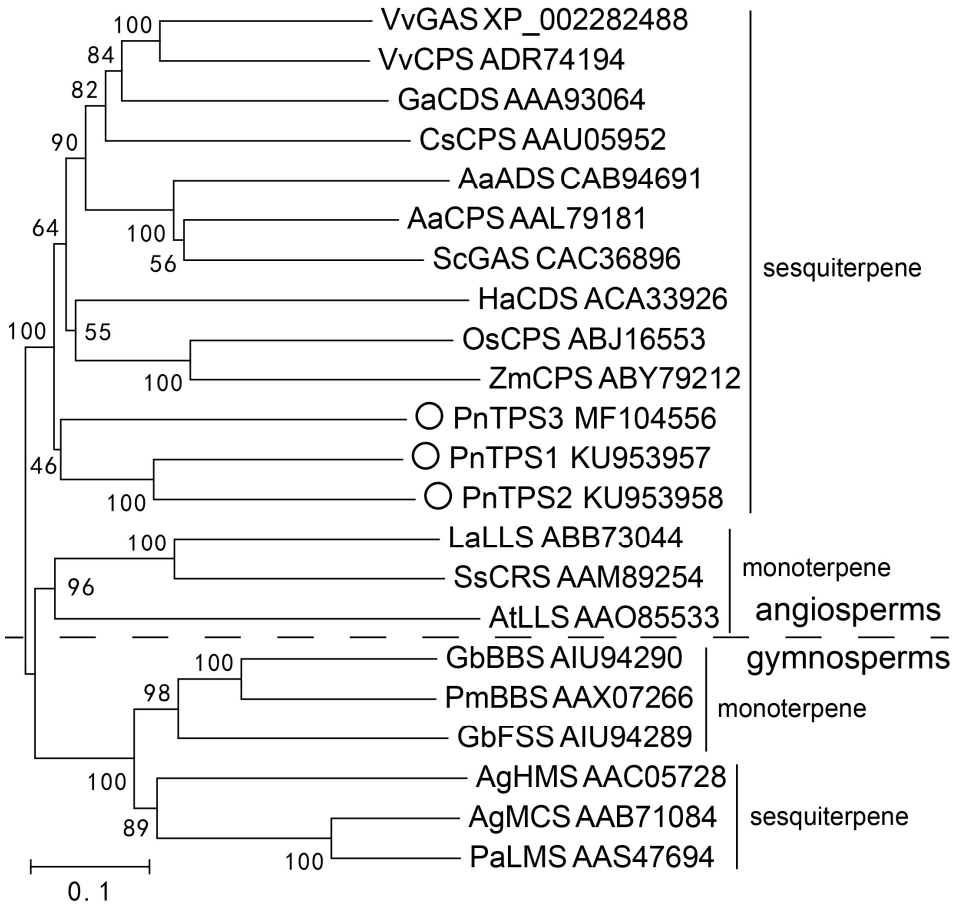

farnesyl diphosphate
OPP
1,3-hydride shift
humulene
PnTPS1
caryophyllene
PnTPS3
germacrene D
enzymatic/ non-enzymatic
δ-cadinol
δ-cadinene
α-cadinene
PnTPS2
β-cadinene
cadinenes
cubebene
copaene

Phylogenetic analysis

To construct phylogenetic tree, sesquiterpene synthases with catalytic function similar to PnTPSs were selected. The tree, constructed using neighbor-joining method (MEGA 6), placed all three PnTPSs studied in this work in a small sub-clade within angiosperm sesqui-TPS family (Fig. 2.7), suggesting that the present PnTPSs are paralogous evolved from common ancestor after speciation.

Figure 2.7. Phylogenetic tree constructed with previously reported terpene synthases.

Aa, *Artemisia annua*; Ag, *Abies grandis*; At, *Arabidopsis thaliana*; Cs, *Cucumis sativus*; Ga, *Gossypium arboreum*; Gb, *Ginkgo biloba*; Ha, *Helianthus annuus*; La, *Lavandula angustifolia*; Os *Oryza sativa*; Pa, *Picea abies*; Pm, *Pseudotsuga menziesii*; Sc, *Solidago canadensis*; Ss, *Salvia stenophylla*; Vv, *Vitis vinifera*; Zm, *Zea mays*. ADS, amorpha-4,11-diene synthase; BBS, bisabolene synthase; CDS, cadinene sythase; CPS, caryophyllene synthase; CRS, 3-carene synthase; FSS, farnesene synthase; GAS, germacrene synthase; HMS, humulene synthase; LLS, linalool synthase; LMS, (-)-limonene synthase; MCS, myrcene synthase. Hollow circle marks three TPSs from *P. nigrum*.

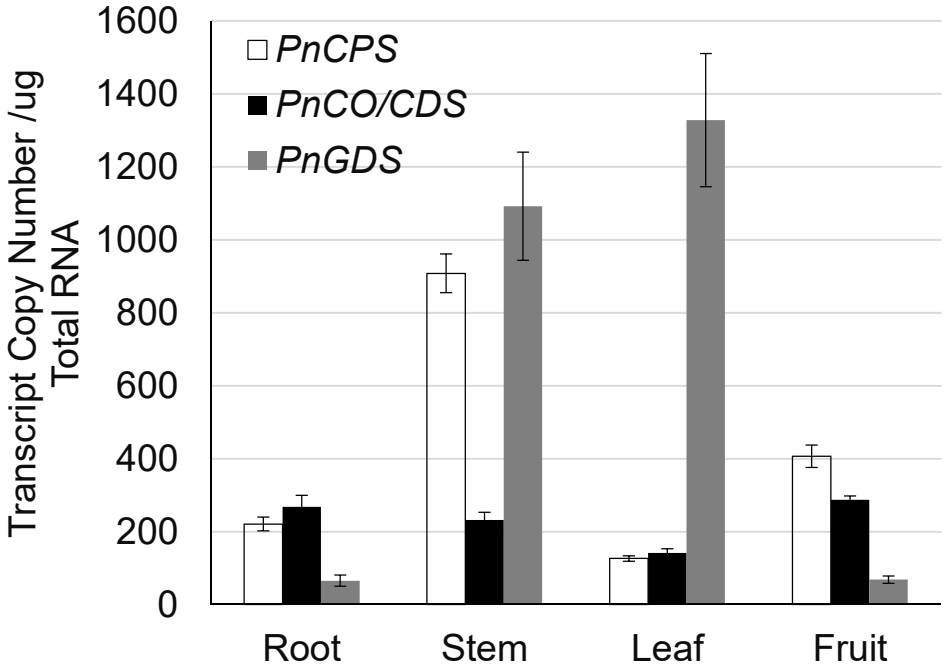

VvGAS XP_002282488
VvCPS ADR74194
GaCDS AAA93064
CsCPS AAU05952
AaADS CAB94691
AaCPS AAL79181
ScGAS CAC36896
HaCDS ACA33926
OsCPS ABJ16553
ZmCPS ABY79212
PnTPS3 MF104556
PnTPS1 KU953957
PnTPS2 KU953958
LaLLS ABB73044
SsCRS AAM89254
AtLLS AAO85533
GbBBS AIU94290
PmBBS AAX07266
GbFSS AIU94289
AgHMS AAC05728
AgMCS AAB71084
PaLMS AAS47694
sesquiterpene
monoterpene
angiosperms
gymnosperms
monoterpene
sesquiterpene
100
84
82
90
64
100
56
100
55
100
46
100
100
96
100
98
100
89
100
0. 1

***PnTPS* transcript levels among pepper organs**

PnTPS transcript levels were relatively low in root and fruit compared to stem and leaf (Fig. 2.8). Stem exhibited high comparable *PnCPS* and *PnGDS* transcript levels with low *PnCO/CDS* level, whereas leaves were characterized by nine-fold abundant *PnGDS* transcripts compared to *PnCPS* and *PnCO/CDS* transcripts. The abundant *PnGDS* transcripts in leaves explains α-cubebene content as high as 20.2% among terpenes in the leaf extract (Lim 2012). Among *PnTPSs* in the peppercorn, *PnCPS* level was the highest correlating with the high combined caryophyllene and humulene content (68.4%) in the peppercorn (Table 2.3). However, about four-fold higher *PnCO/CDS* transcript level compared to *PnGDS* level (Fig. 2.8) could not explain the comparable enzymatic product contents from PnCO/CDS (9.1% for γ- and δ-cadiene and δ-cadinol) and PnGDS (9.5% for copaene and cubebene) (Table 2.3).

Figure 2.8. Absolute transcript copy number of *PnTPS*s estimated by QRT-PCR in four different organs of *P. nigrum*.

Each data point was determined from five biological samples each consisted of four technical replicates.


1600
1400
1200
1000
800
600
400
200
0
Transcript Copy Number /ug Total RNA
PnCPS
PnCO/CDS
PnGDS
Root
Stem
Leaf
Fruit

Part III: Isolation of genes putatively involved in piperine biosynthesis in *Piper nigrum* L.

Introduction

The piperine

The spicy pungent compound of black pepper is an alkamide piperine and its homologs (Friedman et al. 2008). In nature, alkamides frequently occur as *N*-alkylamide, alkenamide or alkenylamide. These alkamides belong to a promising group of natural alkaloids derived from the primary metabolite, for example, amino acids such as phenylalanine, tryptophan, tyrosine, and lysine (Facchini et al. 2004). Modern pharmacological studies support beneficial effects of black pepper consuming to human health and wellness (Srinivasan 2007). The major alkamide in black pepper, piperine, was demonstrated to help digestion of food (Platel and Srinivasan 2000) and to have antioxidative activity (Rauscher et al. 2000). Furthermore, piperine inhibits the enzymes involved in drug biotransformation in liver, thus enhancing bioavailability of drug (Lambert et al. 2004).

Phenylalanine is a starting material for biosynthesis of a large number of alkaloid metabolites. The aromatic ring-containing moiety of piperine is provided by phenylalanine through phenylpropanoid pathway. Piperic acid has been implicated in the piperine biosynthesis. It is believed that the C_5 bridge, linking phenyl ring and piperidine, is derived from propenoyl side chain of cinnamic acid and two-carbon from a malonyl-CoA extender unit. The nitrogen-containing six-membered piperidine ring is from lysine (Dewick 2002). So far, only one report on biosynthesis of piperine has been available (Geisler and Gross 1990) describing the activity of piperoyl-CoA:piperidine *N*-piperoyltransferase in the cell-free extract of *Piper nigrum*.

The purposes of this study

In this study, isolation of piperine biosynthesis genes was attempted by using transcriptome data generated from next generation sequencing (NGS) of unripe

peppercorn to elucidate detailed biosynthetic sequence of piperine. Functional analysis of enzymes involved in four biosynthetic steps, side chain shortening, side chain elongation, phenylpropenoyl-CoA ligation, and *N*-acyltransfer, suggested a novel alternative pathway leading to piperine structure.

Materials and Methods

Bacterial, yeast strains and culture media

Bacterial strains used in this experiment were *E. coli* DH10B [F^- *mcrA Δ*(*mrr-hsdRMS-mcrBC*) *Φ80d lacZΔM15 ΔlacX74 recA1 deoR Δ(ara,leu)7697 araD139 galU galK rpsL* λ^- *endA1 nupG*] for subcloning and *E. coli* Rosetta2(DE3) [F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm* (DE3) pRARE2 (Cam^R)] for protein expression. Yeast disruption mutant YPH499*Δpad1Δfdc1* strain (*MATa ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1 Δpad1-fdc1::TRP1*) (tryptophan selective) was used for functional analysis of *PnCuE*.

E. coli was grown in Luria-Bertani (LB) medium (MBcell, Korea) and solid LB medium plate was made by adding with 1.5% (w/v) Micor-agar (Duchefa, Netherlands). Plasmid-transformed strains were cultured in LB broth supplemented with ampicillin (100 μg/ml) or kanamycin (50 μg/ml).

Wild-type yeast was cultured in YPDA medium (Clontech, USA). The mutant yeast was cultured in 0.67% of Tryptophan DO (dropout) Supplement and 2.67% of Minimal SD Base (Clonetech, USA), and transgenic yeast was cultured in the above medium lacking leucine, and solid plate medium was supplied with 2% of Bacto™ Agar (Becton, USA).

Enzyme and chemicals

Restriction enzymes, T4 DNA ligase, Ex Taq polymerase, and Phusion® High-Fidelity DNA Polymerase were purchased from TaKaRa (Japan) or NEB (USA). Acetonitrile, ethyl acetate, *n*-hexane and methylene chloride were purchased from Duksan, Korea. Acrylamide/bis solution (40%) for SDS-PAGE and Coomassie Blue protein assay reagent were purchased from Bio-Rad (USA). Adenine hemisulfate, ATP, NAD^+, MES monohydrate, $MgCl_2$, and various phenolic acids were purchased from Sigma-Aldrich. Amylose resin (NEB, USA) for maltose-binding protein (MBP)

tagged protein purification and His trap kit for 6×His tagged protein were purchased from GE Healthcare (Sweden).

Oligonucleotides

Primer pairs used for polymerase chain reaction (PCR) were synthesized by Cosmogenetech (Seoul, Korea). The sequence of the oligonucleotides is listed in Table 3.1.

Table 3.1. List of oligonucleotides used in piperine-related study.

Underlined are restriction sites. CuE, cutting enzyme (also called MCHL); PKS, polyketide synthase; CHS, chalcone synthase; 4CL, 4-coumarate:CoA ligase; NAT, *N*-acyltransferase

Primer	(5'→3')
PnCuE	
PnCuE F	atggcgtctcgcctcactctc
PnCuE noER F	atggaggagaatccgatccggctt
PnCuE R	ttacagagagagaataggataagatg
pESC-Leu2 BamHI noERCuE F	cgggatccgatggaggagaatccgatccg
pESC-Leu2 SalI CuE R	acgcgtcgaccagagagagaataggataagatg
pMBP noER CuE NdeI F	cgccatatggaggagaatccgatccggctt
pMBP noER CuE XhoI R	cccgctcgagcagagagagaataggata
PnCuE QRT F	cagcttacactcaagcaacggg
PnCuE QRT R	ggaagacctccattgcagcca
YPH499 ΔPTF F	atgctcctatttccaagaagaa
YPH499 ΔPTF R	ttatttatatccgtacctttcca
TRP SphI F	acatgcatgcaccataaacgacattactatatata
TRP SpeI R	ggactagtaatttcctgatgcggtattttc
pET28 MBP NcoI F	catgccatggatgaaaatcgaagaaggtaaact
pET28 MBP NdeI R	cgccatatgagtctgcgcgtctttcagg
PnPKS and CHS	
PnCHS&PKS F	atgtcgaagacggtagaggagattc
PnCHS&PKS R	ttagttggcctcggcgatgg
Pn DCS F	atggagaaggaagagagtagta
PnDCS R	ttacaaggttccagtttgggata
PnPKS QRT F	actcgataacatggtaggcca
PnPKS QRT R	ccccgataatgatggcagccg
PnCHS QRT F	cctcgatagcatggtcggtca
PnCHS QRT R	accgacgatgatagcagcgg
Pn DCS QRT F	atggagaaggaagagagtagta
PnDCS QRT R	gtagcttgctccacaaaagcagct
Pn4CL	
Pn4CL1 F	atgaagatggtagtagacaat

Pn4CL1 R	ctatttgggcagatcagctgcca
Pn4CL2 F	atggagaaatcaggttatgggaagga
Pn4CL2 R	ttacatcttggaccggactttctcg
Pn4CL3 F	atgatttctgtatctgctgttca
Pn4CL3 R	ttaagtatttgtcaacttggctttca
pET21 NdeI Pn4CL1 F	tgcccgcgcggcagccatatgaagatggtagtagacaat
pET21 XhoI Pn4CL1 R	tggtggtggtggtgctcgagtttgggcagatcagctgcca
pET21 NdeI Pn4CL2 F	tgcccgcgcggcagccatatggagaaatcaggttatgggaag
pET21 XhoI Pn4CL2 R	tggtggtggtggtgctcgagcatcttggaccggactttctcg
pET21 NdeI Pn4CL3 F	tgcccgcgcggcagccatatgatttctgtatctgctgttca
pET21 XhoI Pn4CL3 R	tggtggtggtggtgctcgagagtatttgtcaacttggctttca
pET21 NdeI Nt4CL2 F	tgcccgcgcggcagccatatggagaaagatacaaaacaggttg
pET21 XhoI Nt4CL2 R	tggtggtggtggtgctcgagatttggaagcccagcagcca
Pn4CL1 QRT F	gtgtacacgtacggcgaggt
Pn4CL1 QRT R	gaggaggatcatgatcacctcc
Pn4CL2 QRT F	tcgtcttgctcttgctgatgctg
Pn4CL2 QRT R	gcccaatttggacaatccatggga
Pn4CL3 QRT F	atgatttctgtatctgctgttca
Pn4CL3 QRT R	agcgttccttttcagtggata
PnNAT	
pET-PnNAT1 F	aagaaggagatatacatatgataataacagtgaaggagga
pET-PnNAT1 R	tggtggtggtggtgctcgaggaaatcataaaagagcttcttga
pET-PnNAT2 F	aagaaggagatatacatatggaggtggagatcgtta
pET-PnNAT2 R	tggtggtggtggtgctcgagcaatttgcttgttggcttcatct
pET-PnNAT3 F	aagaaggagatatacatatgagttctcttcttcttctca
pET-PnNAT3 R	tggtggtggtggtgctcgaggagcatgcttctctcttcaat
pET-PnNAT4 F	aagaaggagatatacatatggcagttgagattacca
pET-PnNAT4 R	tggtggtggtggtgctcgaggctcaatgtatgagaaacatgt
pET-PnNAT5 F	aagaaggagatatacatatggaaaaggcttttga
pET-PnNAT5 R	tggtggtggtggtgctcgagaatttggatttcctcttca
pET-PnNAT6 F	aagaaggagatatacatatggagatcatccgcaatagaatcct
pET-PnNAT6 R	tggtggtggtggtgctcgagtgccaaagagtgtgaaaccagct
pET-PnNAT7 F	aagaaggagatatacatatggaggtgaagataatcagtgcag
pET-PnNAT7 R	tggtggtggtggtgctcgaggtcatatggcggataaacgtagga

Isolation of genes putatively involved in piperine biosynthesis

With amino acid sequences of vanillin synthase [*VpVAN* (Genbank accession No. AKG47593)], chalcone synthase [*MsCHS*2 (P30074)], 4-coumarate:CoA ligase [*Nt4CL*2 (NP_001312554)], and agmatine-coumaroyl transferase [*HvACT* (AAO73071)] as queries, homologous gene contigs were searched in transcriptome database using Local tBLASTn method (BioEdit, version 7.2.0). Based on the sequence information of the contigs, primer pairs were designed (Table 3.1.) to clone respective ORFs from cDNA library under the following conditions using Takara Ex Taq Polymerase (Takara, Japan): 98℃ for 5 min for initial, 98℃ for 15 s, 58℃ for 30 s, 72℃ for 1 min/kbp for 35 cycles and 72℃ for 5 min and 16℃ hold for final. PCR products were purified by using Inclone™ Gel & PCR purification kit (Inclone Biotech, Korea) and finally TA cloned into *pMD20* T-vector (Takara, Japan).

Yeast *PAD*1, *FDC*1 disruption mutant

Yeast phenylacrylic acid decarboxylase (PAD1, YDR538W) and ferulic acid decarboxylase (FDC1, YDR539W) are closely linked to each other up and downstream of chromosome IV. Using PAD F and FDC R primer set, a fragment containing *PAD* and *FDC* genes (2.7kbp) were isolated from chromosomal DNA. The PCR product was ligated to *pMD19* T-vector. The new construct was confirmed by sequencing. In the next step, *Sph*I and *Spe*I were used to double digest *PAD1* and *FDC1* at 368 bp and at 996 bp downstream from ATG of each gene, respectively. The tryptophan synthase gene (1.06kbp) obtained from *pESC-TRP* vector was then ligated to *Sph*I and *Spe*I double digestion site. Gene fragments for knock-out experiment was thus prepared (Fig. S3). By using PAD F and FDC R primer pair, the fragment was amplified and 5 μg of PCR products was transformed into yeast YPH499 strain by using LiAc/SS carrier DNA/PEG method (Gietz and Schiestl

2007). Finally, transformants were spread on tryptophan dropout medium to select *Δpad1-fdc1* mutant named YPH499 *ΔPTF*.

Heterologous expression and protein purification

ORF of each gene was cloned into *pET21a* at *Nde*I and *Xho*I double digestion site with C-terminal 6×His tag or *Nde*I and *Xho*I double digestion site of *pET28*a(+)-MBP (maltose binding protein) vector (MBP from *pMAL-c2x* vector at N-terminal of gene at *Nco*I and *Nde*I site) by using Gibson Assembly® Cloning Kit (NEB, USA). The ORF-containing vector was then transformed into *E. coli* Rosetta2(DE3), and selected on agar plate supplemented with 100 μg/ml ampicillin or 50 μg/ml kanamycin. The selected single colony was taken and incubated at 37℃ overnight. A loop of the overnight incubated colony was inoculated into 5 ml LB medium supplemented with 100 μg/ml ampicillin for *pET21* constructs or 50 μg/ml kanamycin for *pET28* constructs. Five hundred microliters of the overnight culture was inoculated into 50 ml fresh LB medium with the same concentration of antibiotics. Incubation was continued at 37℃ until OD_{600} reached 0.5, when the culture was added with IPTG to 0.1 mM and the incubating temperature was lowered to 20℃. The culture was further incubated for 16 h, and the cells were collected by centrifugation at 5,000 rpm (Beckman F1202 rotor) for 5 min at 4℃. The cell pellet was resuspended in 5 ml of lysis buffer (20 mM Tris-HCl, pH 7.0, supplemented with 300 mM NaCl). The suspension was then sonicated (Sonic Dismembrator 550, Fisher Scientific) for total of 10 min with repeated 3 s on and 1 s off interval. Soluble protein fraction was collected by centrifugation for 30 min at 6,000 rpm (Beckman F1202 rotor) and 4℃. The supernatant was applied to 60 Ni Superflow Resin (Takara, Japan) and the target protein fraction was eluted by increasing concentration of imidazole from 100 to 500 mM. The MBP-tagged protein was purified on Amylose Resin (NEB, USA) eluted by lysis buffer containing 10 mM maltose. Finally, the

eluted protein was desalted by buffer exchange on centrifugal filter (Amicon® Ultracell-30k, Merck Millipore, Ireland).

***In-vitro* assay**

Standard assays consisted of 0.2 mM substrate, 2.5 mM DTT, 2.5 mM ATP, 5 mM $MgCl_2$, 0.2 mM coenzyme A, and 1 µg of purified proteins (PnPKSs, PnDCS, PnMCHL or Pn4CLs) in total volume of 1 ml Tris-HCl buffer (100 mM, pH 7.5). The reaction mixture was gently overlaid with 0.5 ml of *n*-hexane to trap volatile compounds. Control reaction contained boiled enzyme except other conditions were same. After 30 min incubation at 30°C, assays were terminated by boiling assay mixture. The large scale reaction (5 ml) was continued for 3 h.

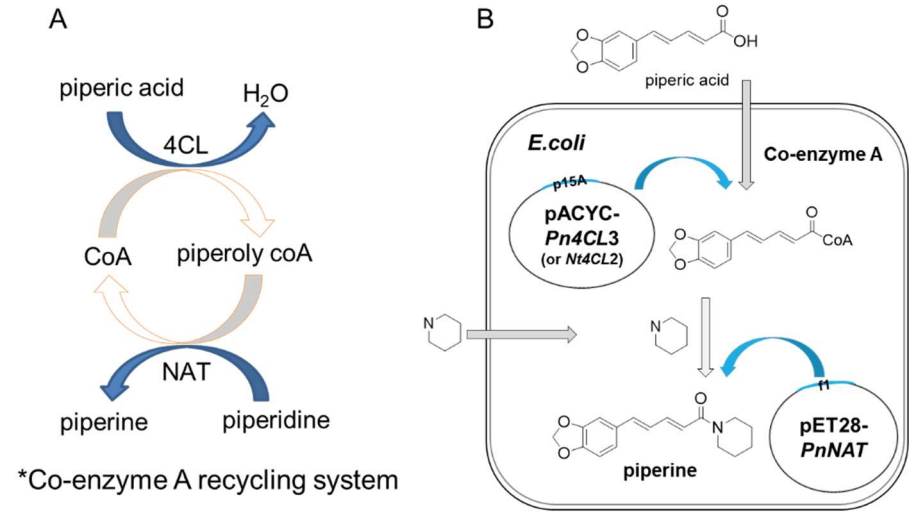
***In-vivo* bioconversion assay by co-transformants**

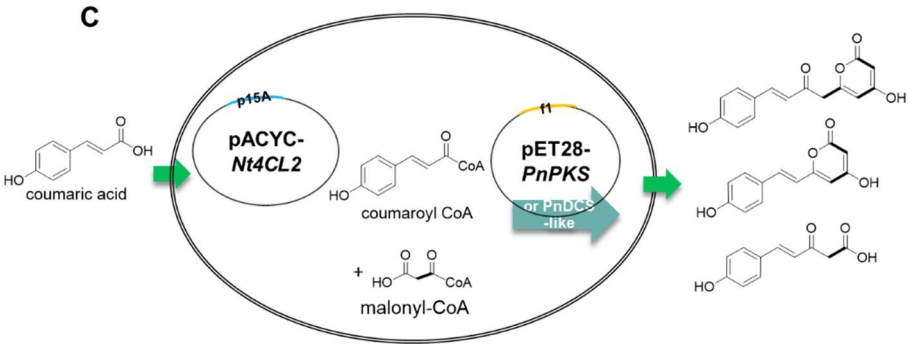
To screen PnPKS, DCS, and *N*-acyltransferase activity, *E. coli* was double-transformed with *pACYC*-4CL (Nt4CL2 or Pn4CL3) and *pET*-PKS (PnPKS or DCS) or *pET*-NAT. Specifically, Pn4CL3 was cloned into *pACYC*-chloramphenicol selective plasmid and PnNATs into *pET28*-kanamycin selection plasmid. Both plasmids were transformed into *E. coli* C43 (DE3). Double-transformants were selected on LB agar plate containing 50 µg/ml of kanamycin and 35 µg/ml of chloramphenicol (Fig. 3.1A, B). Selected single colony was taken to inoculate 5 ml LB medium supplemented with 50 µg/ml kanamycin and 35 µg/ml chloramphenicol and incubated at 37°C overnight. Five hundred microliters of the overnight culture was inoculated to 50 ml fresh LB medium with the same concentration of antibiotics. Incubation was continued at 37°C until OD_{600} reached 0.5, when the culture was supplemented with IPTG to 0.2 mM and the incubation temperature was lowered to 18°C. For NAT assay, the culture was further incubated for 16 h and was added with piperic acid and piperidine to the final concentrations of 0.2 mM and 0.3 mM, respectively. Using the same strategy, PnPKSs and DCS-like were assayed by

supplying eight phenylpropenic acid analogs (Fig. 3.1C). The culture was incubated for additional 3 days at 30°C before the culture medium was analyzed.

Figure 3.1. Schematic illustration for *in-vivo* enzyme assay using two plasmid expressing *E. coli*.

A, A model of Co-enzyme A recycling. **B**, Biotransformation strategy of piperine synthesis. **C**, Bioconversion system employing PnPKSs (or PnDCS-like) and Nt4CL2.


A
piperic acid
H_2O
4CL
CoA
piperoly coA
NAT
piperine
piperidine
*Co-enzyme A recycling system
B
piperic acid
E.coli
Co-enzyme A
p15A
pACYC-
Pn4CL3
(or Nt4CL2)
CoA
N
f1
pET28-
PnNAT
piperine


C
p15A
pACYC-
Nt4CL2
coumaric acid
coumaroyl CoA
malonyl-CoA
f1
pET28-
PnPKS
or PnDCS
-like

Analysis of metabolites

The fresh peppercorn extract for metabolite analysis was prepared as follows. Frozen peppercorn (5 g) was ground with liquid nitrogen and extracted with CH_2Cl_2 (20 ml) by gentle stirring overnight at room temperature. The CH_2Cl_2 layer was separated by centrifugation at 4°C before concentrating to 4 ml under stream of nitrogen gas.

The culture medium (50 ml) after desired biotransformation was extracted with 5 ml of ethyl acetate or CH_2Cl_2 by vortexing and the organic phase was separated by centrifugation (4,000 *g*, 5 min). The extraction was repeated once with the same volume of solvent. The combined extract was concentrated to 2 ml under mild flow of nitrogen before direct analysis. The *in-vitro* enzyme assay mixture was extracted with 500 µl *n*-hexane by vigorous vortexing for 1 min. The hexane layer was separated by centrifugation (4,000 *g*, 5 min) and the extraction was repeated. The combined extract was concentrated to 0.3 ml under mild flow of nitrogen gas before subjecting to analysis. The polyketide products were analyzed on LC-MS, and piperonal and piperine were analyzed on GC-MS. The conditions for analyses are shown below.

The phenylpropenate:CoA ligation reaction was stopped by boiling for 5 min. The supernatant was collected after centrifugation (6,000 *g*, 10 min), diluted five-fold with water, and analyzed on spectrophotometer (X-ma 3000PC, Human Corp., Korea) in scan mode to obtained absorption spectrum (Fig. 3.8).

GC-MS analysis

GC System (Agilent model 6890, USA) equipped with a mass spectrometer was used to analyze piperonal and piperine. The column was ZB-5MSi (60 m × 0.25 mm × 0.25 µm, Zebron) and helium was carrier gas at a flow rate of 1 ml/min. For enzymatic reaction product analysis, 1 µl of *n*-hexane extract was directly injected into GC with the following temperature program: injection at 220°C, and initial temperature of 50°C (5 min hold) raised to 110°C at a rate of 5°C/min and then at

10°C/min to 240°C (10 min hold). The coupled mass spectrometer was Agilent HP 5973 (USA) with transfer line temperature set at 320°C, source temperature at 250°C, quadrupole temperature at 150°C, and ionization potential at 70 eV. Scan range was 50 to 300 atomic mass units. Products were identified by using MS Search Program v.2.0 (NIST Standard Reference Database, USA).

LC-MS analysis

Twenty microliters of ethyl acetate extract was applied to an Inerssil ODS-3 column (2.1 × 150 mm, 3 μm; GL Sciences Inc., Japan) and eluted by CH_3CN–0.05% formic acid in H_2O with linear CH_3CN gradient at a flow rate of 500 μL/min by using a high performance liquid chromatography (HPLC) system coupled to an ion trap mass spectrometer with an electrospray ionization (ESI) source Thermo Electron-LTQ-Orbitrap XL Hybrid MS operating in the positive mode with mass range of *m/z* 50-400.

QRT-PCR

Transcript levels of piperine biosynthesis-related genes were determined by qRT-PCR using primer pairs described in Table 3.1. The PCR was performed in a total reaction volume of 20 μl containing 10 μl of 2×QuantiMix SYBR Kit PCR system (PhileKorea, Korea), 1 μl diluted cDNA, 0.25 μM of each primer, and double-distillated water. The PCR reaction was run on Rotor-Gene 2000 Real Time Cycler (Corbett Research, Australia) with a temperature program of 5 min at 95°C, 40 cycle of 15 s at 95°C, 20 s at 58°C, and 20 s at 72°C. The standard curve was made by running the PCR of each gene ranging from 10^2 to 10^7 copy/μl, and the copy numbers were calculated as described by (Yin et al. 2001). Individual data point was obtained from quintuple biological samples each composed of quadruple technical replicates.

Bioinformatics analyses

The local tBLASTn search for each gene was done by using BioEdit (version 7.2.0). Threshold values in tBLASTn search were set at ≥ 25% for protein identity and ≤ 5×10^{-20} for e-value. The alignment of nucleotide and amino acid sequences was performed using the Vector NTI Advance® software (Thermo Fisher Scientific, USA). The online prediction of subcellular localization was performed at www.cbs.dtu.dk/services/. The phylogenetic tree was constructed with MEGA6 (USA). The gene sequence translation, reverse translation, and restriction summary were done at www.bioinformatics.org/sms2/.

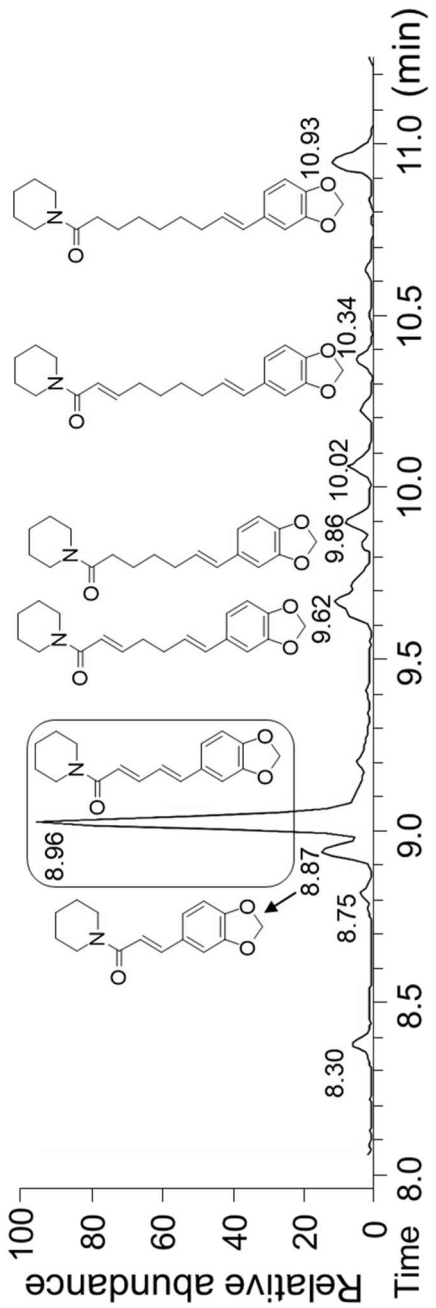
Results and Discussion

Metabolites profiling

Metabolite profiling was carried out to ensure that loci of target gene expression and biochemically related metabolites are same, *i.e.*, the edible fruit of the *Piper nigrum*. Thus, the extract of fresh peppercorns was analyzed by LC-MS. Based on total ion chromatogram (TIC) and predicted molecular mass of metabolites, it was possible to detect and confirm the presence of piperine and its homologs in the peppercorn. The piperine-series compounds were ilepcimide, piperine, pipersintenamide, piperoleine A, pipernonaline, and piperoleine B (Fig. 3.2). Structure identification of individual compound in extracts was based on retention order of the peak and EI-mass spectrum (Epstein et al. 1993; Kozukue et al. 2007; Friedman et al. 2008). In the previous report on piperine concentration in different vegetative parts of *P. nigrum*, the highest piperine accumulation was also found in fully differentiated shoots, rivaling the piperine concentration in peppercorn (Semler and Gross 1988).

Figure 3.2. LC-MS profile of immature black pepper fruit extract.

Boxed is piperine, evidently the main product of black pepper extract. Retention time (t_R) of ilepcimide 8.87 min (*m/z* 259), piperine 8.96 min (*m/z* 285), pipersintenamide 9.62 min (*m/z* 313), piperoleine A 9.86 min (*m/z* 315), pipernonaline 10.34 min (*m/z* 341), piperoleine B 10.93 min (*m/z* 343).


Relative abundance
100
80
60
40
20
0
Time
8.0
8.5
9.0
9.5
10.0
10.5
11.0 (min)
8.30
8.75
8.87
8.96
9.62
9.86
10.02
10.34
10.93

Search for genes involved in piperine biosynthesis

The piperine biosynthesis pathway includes four putative phases of molecular transformation in addition to piperidine formation: side chain extension of phenylpropenate (C_6C_3) to form piperate skeleton (C_6C_5), formation of methylenedioxy bridge, activation of piperate by CoA ligation, and transfer of piperoyl moiety to piperidine. The candidate enzymes possibly catalyzing each step were predicted against queries of MsCHS2 (for type III plant polyketide synthase), CjCYP719A1 and SiCYP81Q1 (for methylenedioxy bridge formation), Nt4CL2 (for 4-coumarate:CoA ligase), and agmatine-coumaroyltransferase HvACT (for *N*-acyltransferase). Table 3.2 lists the putative genes found as such, and additionally includes a candidate gene for piperonal or vanillin syntheses from 3,4-methylenedioxycinnamic or ferulic acid, respectively. The complete ORF sequences are presented in Supplementary Data (List 1). The successful search for homologs for these steps suggested that the transcriptome reads generated from unripe peppercorn might satisfactorily cover most gene candidates for piperine biosynthesis. After the quality of the candidate genes was evaluated and validated *in-silico*, functional validation of putative genes was then undertaken. Of the candidate genes for five steps of piperine pathway, all the enzymes except the bridging enzyme were assayed, and the results will be presented in the following sections.

Table 3.2. List of putative genes in piperine biosynthesis.

Asterisk indicates tBLASTn search query and the rest are putative genes. Enzymatic function of PnCYPs was not examined.
Vanilla planifolia VpVAN (AKG47593), *Medicago sativa* MsCHS2 (P30074), *Curcuma longa* ClDCS (AB495006), *Sesamum indicum* SiCYP81Q1 (NP_001306620), *Coptis japonica* CjCYP719A1 (BAB68769), *Hordeum vulgare* HvACT (AAO73071).

Gene name	ORF (bp)	Function (characterized or predicted)	Protein (kDa)
*VpVAN**	1,071	Vanillin synthase	39.0
PnMCHL	1,071	Side chain shortening	38.8
*MsCHS2**	1,167	Naringenin chalcone synthase 2	42.7
PnPKS1	1,185	Polyketide synthase for side chain	43.1
PnPKS2	1,185	elongation	43.1
*ClDCS**	1,167	Diketide-CoA synthesis	42.1
PnDCS	1,203	Diketide-CoA synthesis	43.0
*Nt4CL2**	1,629	Co-enzyme A ligation	59.5
Pn4CL1	1,638		59.6
Pn4CL2	1,635	Phenylpropenate-CoA ligation	59.4
Pn4CL3	1,692		60.8
*SiCYP81Q1**	1,518	Sesamin snythase	57.3
*CjCYP719A1**	1,473	Canadine synthase	55.4
PnCYP1	1,518		56.3
PnCYP2	1,587	Methylenedioxy bridge formation	60.0
PnCYP3	1,506		55.5
*HvACT**	1,317	*p*-Hydroxycinnamoylagmatine synthase	47.5
PnNAT1	1,299		47.9
PnNAT2	1,308		48.0
PnNAT3	1,365		49.8
PnNAT4	1,329	*N*-Acyltransferase	48.8
PnNAT5	1,290		47.4
PnNAT6	1,302		47.7
PnNAT7	1,314		49.0

Side chain extension of phenylpropenate

The side chain elongation of 3,4-methylenedioxycinnamate had been proposed (Dewick 2002), although chain elongation before methylenedioxy bridge formation is still possible. The enzymes performing such chain elongation is known as type-III plant polyketide synthase (PKS III), representative example being chalcone synthase (CHS). CHSs are well-known for their promiscuous nature in substrate and product specificities. Such infidelity, however, is suggested to give advantage in generating diverse structures with relatively small number of enzymes (Austin and Noel 2003; Stewart et al. 2013). One C_2 unit extension of *p*-coumaroyl-CoA in synthesis of benzalacetone by *RpBAS* (Abe et al. 2001) and of *p*-coumaroyl-CoA to *p*-coumaroyl diketide-CoA by *ClDCS* in curcumin biosynthesis (Katsuyama et al. 2009) have been documented. The three molecular malonyl-CoA extension on *p*-coumaroyl-CoA yields 3,4,5-trihydroxystilbene by stilbene synthase *PsSTS* (Schanz et al. 1992). Other examples of three C_2 extension include; coumaroyl triacetic acid synthase (HmCTAS) (Akiyama et al. 1999) and, notably, chalcone synthase (MsCHS) (Jez et al. 2000) among others. To search a PKS III catalyzing one C_2 extension on ferulate or 3,4-methylenedioxycinnamate, I looked into black pepper transcriptome data with query of MsCHS2 and ClDCS to identify full ORF sequences coding either PKS- or CHS-like enzymes, and was able to identify two putative PnPKS full clones, labeled as PnPKS1 and PKS2, and one diketide CoA synthase homolog PnDCS-like. They were expected to perform one C_2 extension. Two above-mentioned PnPKSs, with 95.4% identity between them, had 63.2% identity to MsCHS and PnDCS-like had 57.6% identity to ClDCS.

These candidate genes were cloned into *E. coli*. Because PKS works on acyl-CoA as substrate, *pACYC*-Nt4CL2 was co-transformed into the *E. coli*. Nt4CL2 is known for its broad substrate specificity accepting various phenylpropenate with or without 3,4-methylenedioxy bridge (data not shown). This two-plasmid *E. coli* system for bioconversion of phenylpropenate to oligoketide is schematically depicted in Fig.

3.1C. The products generated by PnPKS1 from various phenylpropenic acid were analyzed by LC-MS (Table. 3.3). When necessary, *in-vitro* enzyme assay using partially purified recombinant protein was performed and the chemical structure was further confirmed by NMR when there was ambiguity in MS-based identification.

PnPKS1 catalyzed extension of side chain of various phenylpropenic acid (C_6C_3) to produce bis-noryangonin-type derailed triketide products as well as corresponding flavonones in addition to minor tetraketides (Table. 3.3). Mass spectrometric and NMR data confirmed formation of benzalacetone, which calls for coumaroyl diketide (one C_2 extension product) as precursor. One C_2-extension of caffeic acids was also suggested based on MS data only. However, this extension was unlikely because reduction was additionally necessary to arrive at coumaroyl diketide structure. PnPKS1 and -2 catalyzed formation of naringenin as the major product without generating bis-noryangonin-type derailed product. PnPKS1 also could take several 3,4-methylenedioxy compounds. Especially, piperonylic acid (C_6C_1) was a substrate for two C_2-extensions resulting in C_6C_5 framework of piperic acid. This enzyme also could add two C_2-units to 3,4-methylenedioxycinnamic acid (MDCA) generating C_6C_7 skeleton of piperettic acid. Chain elongation terminated at this stage with hydrolysis of the polyketide intermediate followed by cyclization to form lactone. In short, PnPKS1 allowed formation of correct carbon framework of piperic acid from piperonylic acid and of piperettic acid from MDCA while the enzyme was functioning as a normal CHS (Fig. S2).

The formation of the derailed products involves termination of chain elongation by premature release of the polyketide intermediates to form lactones if chain length allows. However, product with C_6C_5 backbone was not found in PnPKS reaction of any phenylpropenic acids, apparently signifying both PnPKSs could not terminate chain elongation after one C_2 extension stage. Because addition of one C_2 unit had been reported for benzylacetone synthase (*RpBAS*) (Abe et al. 2001) and diketide-CoA synthase (DCS) (Katsuyama et al. 2008), the transcriptome database was further

screened with DCS as query. The search yielded one full-length ORF, *PnDCS-like*. However, PnDCS-like did not show any activity towards the substrates that had been tested with PnPKSs.

Besides phenylpropenic acids, vanillic acid (C_6C_1) and piperic acid (C_6C_5) also did not serve as the substrate of PnPKSs to yield chain elongation product. Nevertheless, PnPKS accepted piperonylic acid (C_6C_1) as substrate and catalyzed two C_2 unit elongation to form C_6C_5 compound (Table. 3.3). PnDCS did not show any *in-vivo* activity to produce diketide-CoA compounds as mentioned above (data not shown).

*PnPKS*s and *DCS*-like transcript levels were estimated by QRT-PCR. *PnPKS1* was highly expressed in stem and leaves. In fruits, PnPKS2 was the most highly expressed one while stem had high *PnPKS1* transcript level (Fig. 3.3).

Table 3.3. Summary of *in-vivo* and *in-vitro* products of PnPKS1 and PKS2 from peppercorn.

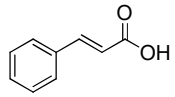
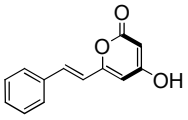
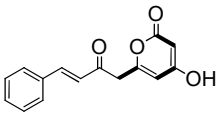
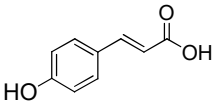
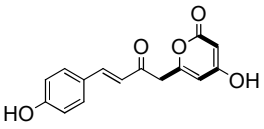
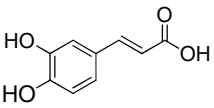
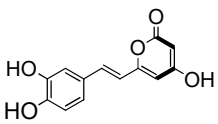
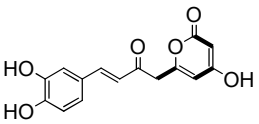
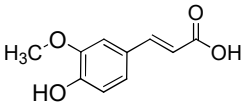
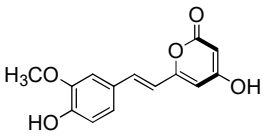
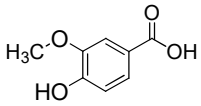
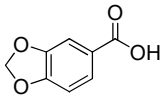
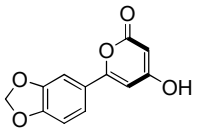
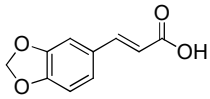
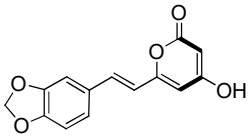
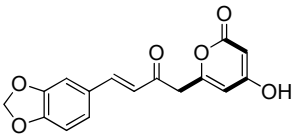
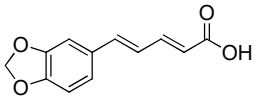
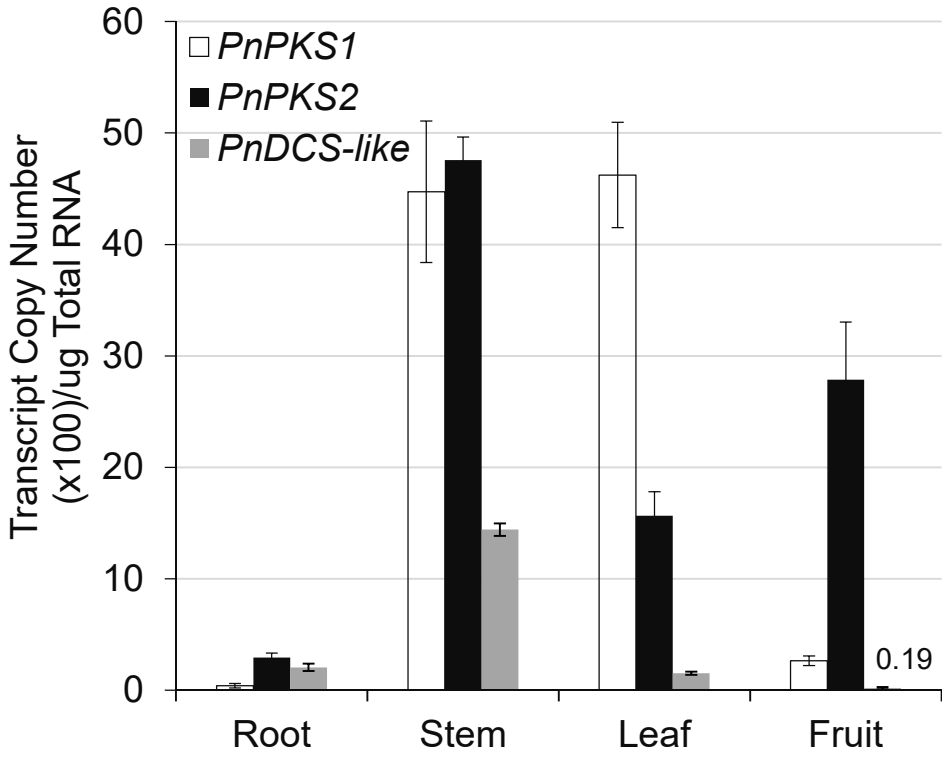
Start unit	Products
 Cinnamic acid	 
 Coumaric acid	
 Caffeic acid	 
 Ferulic acid	
 vanillic acid	not detected
 piperonylic acid	
 3,4-methylenedioxy cinnamic acid	 
 piperic acid	not detected

Figure 3.3. *PnPKS*1, *PnPKS*2, and *PnDCS*-like transcripts in four different organs of *P. nigrum*.

Each data point was determined from four biological and four technical replicates.


60
50
40
30
20
10
0
Transcript Copy Number
(x100)/ug Total RNA
PnPKS1
PnPKS2
PnDCS-like
0.19
Root
Stem
Leaf
Fruit

PnMCHL, an enzyme for 3,4-MDCA-specific side chain cleavage

Piperonal, *aka* 3,4-methylenedioxybenzaldehyde or heliotropin, is a common constituent in peppercorn presenting about 1% of pepper oil and occurs also in various plants such as vanilla and dill. Because of its pleasant floral odor, piperonal is widely added in food, cosmetics, and perfume. It is also a synthetic precursor of various pharmaceuticals including Cialis and Clarium.

The biosynthesis of piperonal remains unknown. Because the structure of piperonal is similar to vanillin, except 3,4-methylenedioxy ring in piperonal instead of 3-methoxy-4-hydroxy moiety in vanillin, it is reasonable to posit that piperonal was directly derived from 3,4-methylenedioxycinnamic acid (MDCA) in analogy of vanillin pathway in *Vanilla* pod (Gallage et al. 2014) (Fig. 3.6).

Screening of PnMCHL *in-vivo*

I have noticed that vanillin is synthesized from ferulic acid in *Vanilla planifolia* by removal of terminal two-carbon unit as acetic acid through the action of vanillin synthase (VpVAN) with hydratase-lyase (HL) activities (Gallage et al. 2014). The analogous reaction on MDCA would yield piperonal. Encouraged by this possibility, peppercorn transcriptome database was searched against VpVAN as a query. One full ORF sequence, with amino acid sequence identity of 70% to VpVAN, was found and cloned. The putative enzyme, composed of 357 amino acid residues and predicted MW of 38.7 kDa, had high similarity at 72% to cysteine proteinase superfamily to which VpVAN also belongs. For initial screening of hydratase-lyase activity, a yeast mutant named YPH499 *ΔPTF* (Fig. S3) with knocked-out phenylacrylic acid (*PAD1*) and ferulic acid (*FDC1*) carboxylase genes were constructed (Mukai et al. 2010; Lin et al. 2015). Six pheylpropenic acid homologs (cinnamic, coumaric, caffeic, ferulic, and MDC, and piperic acid) were fed to the culture of the mutant yeast harboring the HL enzyme candidate. It was found that only MDCA gave the expected product, piperonal (Figs. 3.4 and 3.6). Ferulic acid

also yielded a product but only identified as 2-methoxy-4-vinylphenol due to the residual decarboxylase activity (Fig. 3.4). This enzyme from black pepper was named PnMCHL for 3,4-methylenedioxycinnamic acid hydratase-lyase.

***In-vitro* assay of PnMCHL**

The bacterial 4-hydroxycinnamoyl-CoA hydratase-lyase (HCHL) from *Psudomonas fluorescens* strain AN103 could transform feruloyl-CoA into vanillin and acetyl CoA, requiring NAD^+ and ATP as cofactor (Gasson et al. 1998; Narbad and Gasson 1998). This bacterial enzyme also could take *p*-coumaroyl and caffeoyl-CoA *in-vitro* as substrate to produce corresponding hydroxyl benzaldehydes (Mitra et al. 1999). Therefore, PnMCHL had higher substrate specificity compared to the bacterial enzyme. To confirm if cleavage of MDCA to piperonal was artefact arising from *in-vivo* environment, MBP-PnMCHL fusion protein was heterologously expressed in *E. coli* and purified (Fig. S4) for *in-vitro* assay. Again, the putative PnMCHL cleaved MDCA to piperonal in the absence of NAD^+ and ATP, which supported position of PnMCHL in cysteine proteinase superfamily. However, ferulic acid could not serve as substrate again in *in-vivo* assay, firmly establishing substrate specificity of the putative PnMCHL (Figs. 3.5 and 3.6).

Implication of the presence of PnMCHL

Does piperonal has a role in piperine biosynthesis? The possibility might look remote, because piperonal is a degradation product of MDCA, which has been considered as precursor of piperic acid. Therefore, it would be more reasonable to conjure extension of MDCA by C_2 unit to arrive at correct piperic acid carbon framework. However, piperonylic acid, oxidized product of piperonal, was shown to be a substrate of PnPKS1 for extension of two and three C_2 units in the present experiment. In other words, it is possible that MDCA is cleaved to piperonal by PnMCHL *in-planta* followed by oxidation into piperonylic acid and, finally, two C_2 extension on

piperonylic acid by PnPKS1 or -2 would give correct piperic acid carbon framework (Fig. 3.7). In the meantime, piperettine structure (C_6C_7) also could be explained by either extension of piperonylic acid by three C_2 units or extension of MDCA by two C_2 units.

Figure 3.4. Substrate specificity of PnMCHL in yeast YPH499 *ΔPTF* mutant strain.

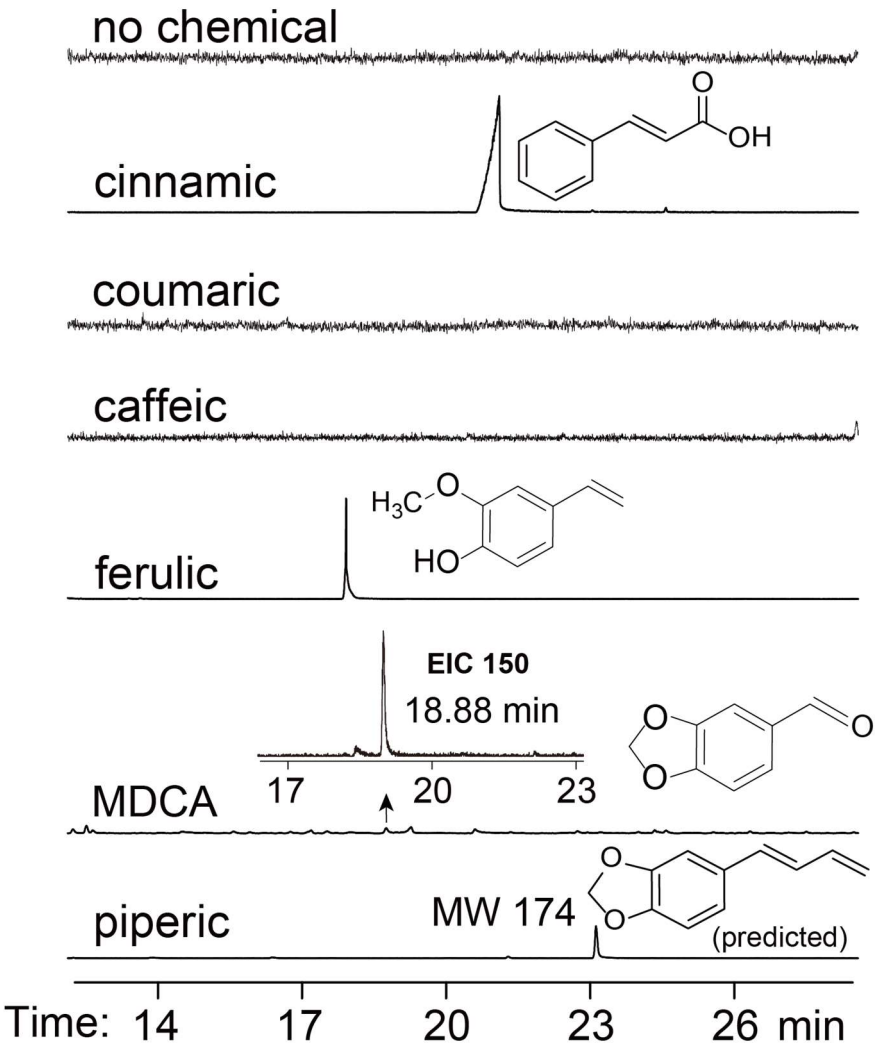

no chemical
cinnamic
O
OH
coumaric
caffeic
ferulic
H3C
O
HO
EIC 150
18.88 min
17
20
23
MDCA
O
O
O
piperic
MW 174
O
O
(predicted)
Time: 14
17
20
23
26 min

Figure 3.5. GC-MS analysis of *in-vitro* product from 3,4-methylenedioxycinnamic acid (MDCA) and ferulic acid.

Upper panel, MDCA as substrate. Lower panel, ferulic acid as substrate.

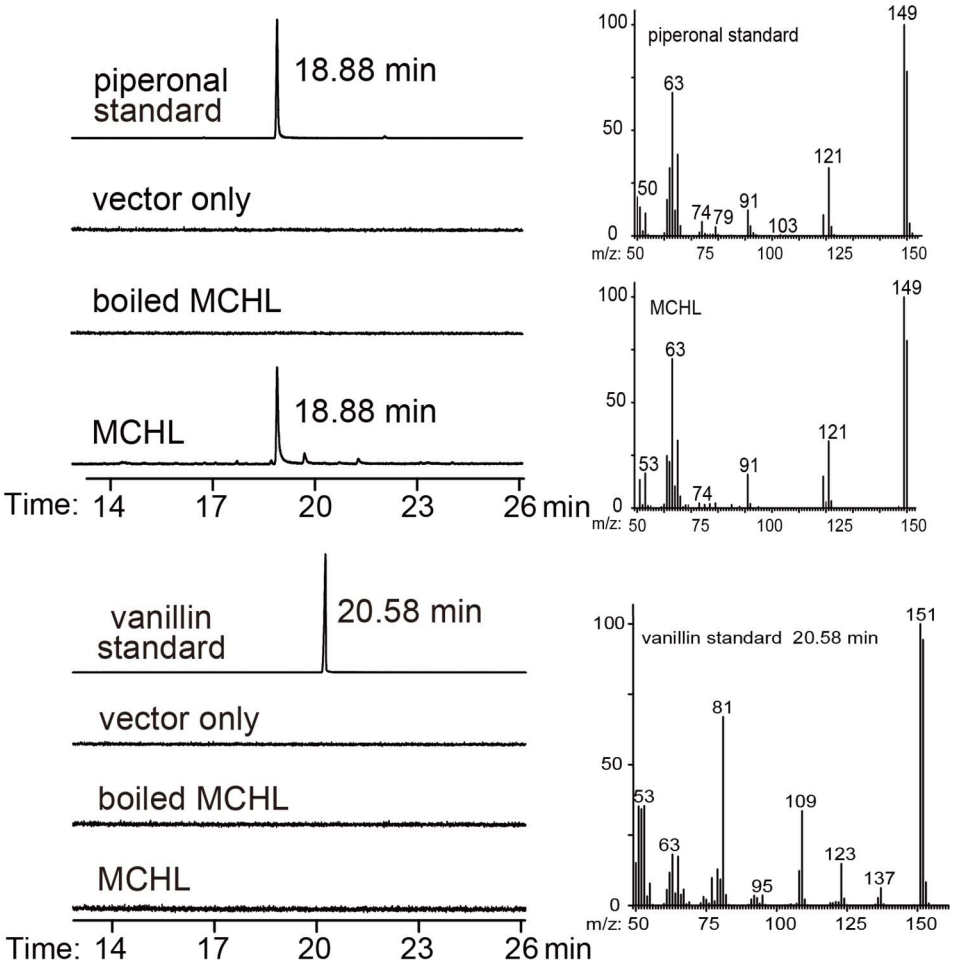

piperonal standard
18.88 min
vector only
boiled MCHL
MCHL
18.88 min
Time: 14 17 20 23 26 min
piperonal standard
MCHL
vanillin standard
20.58 min
vector only
boiled MCHL
MCHL
Time: 14 17 20 23 26 min
vanillin standard 20.58 min
m/z:

Figure 3.6. Comparatively of vanillin biosynthesis with piperonal synthesis by PnMCHL.

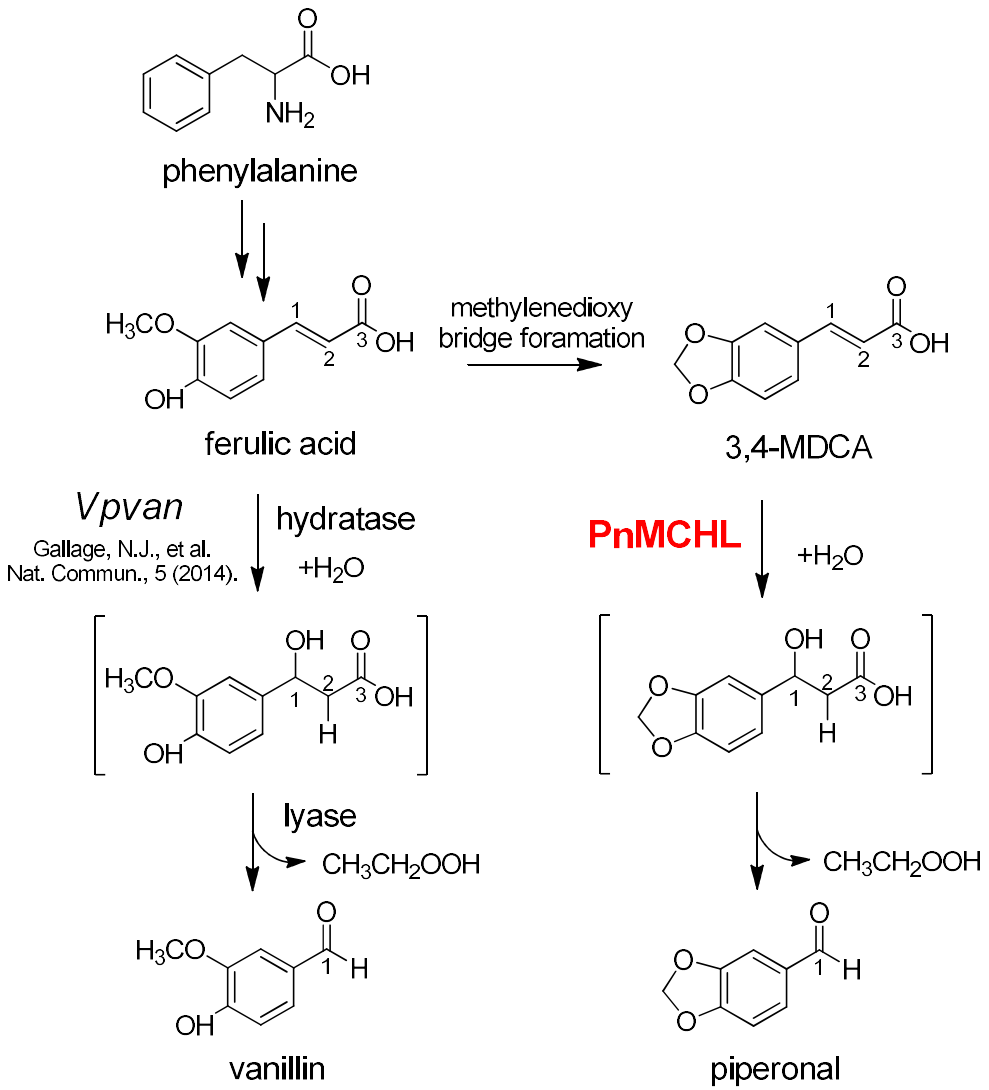
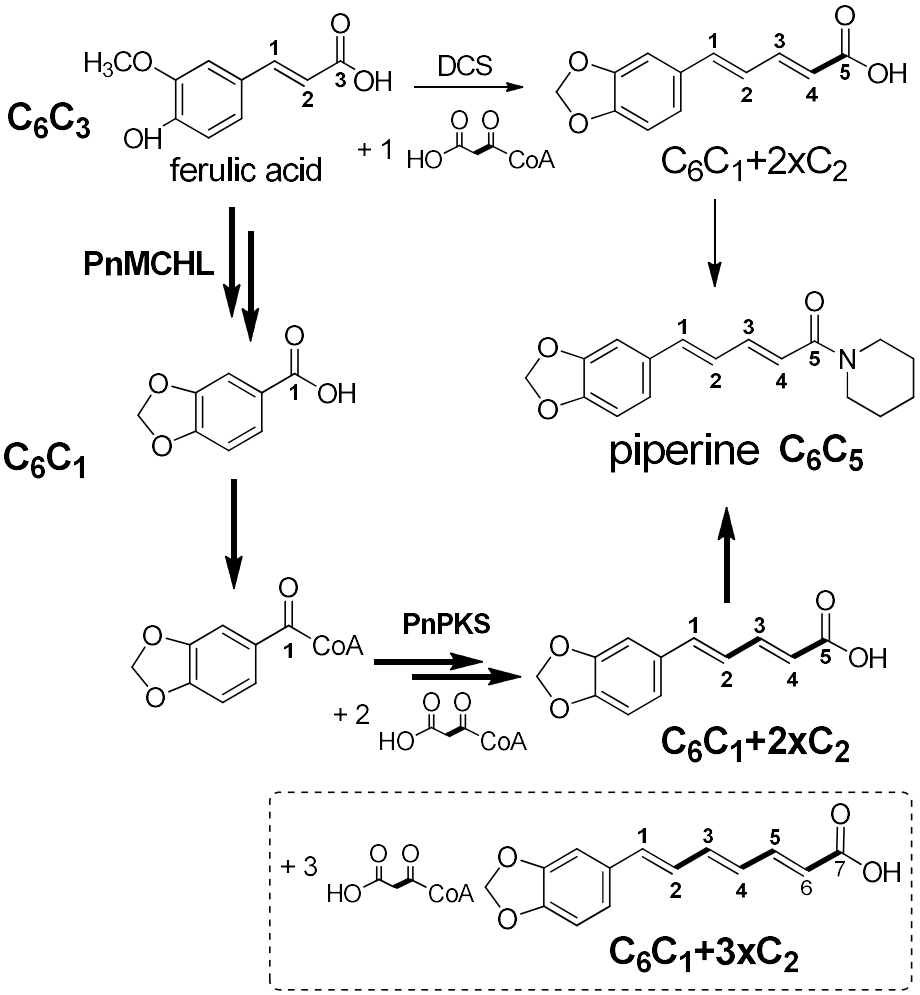

phenylalanine
ferulic acid
methylenedioxy
bridge foramation
3,4-MDCA
Vpvan
Gallage, N.J., et al.
Nat. Commun., 5 (2014).
hydratase
+H2O
PnMCHL
+H2O
lyase
CH3CH2OOH
CH3CH2OOH
vanillin
piperonal

Figure 3.7. Proposed alternative pathway (arrow in bold) to piperine involving PnMCHL, the sidechain shortening enzyme, and PnPKS.

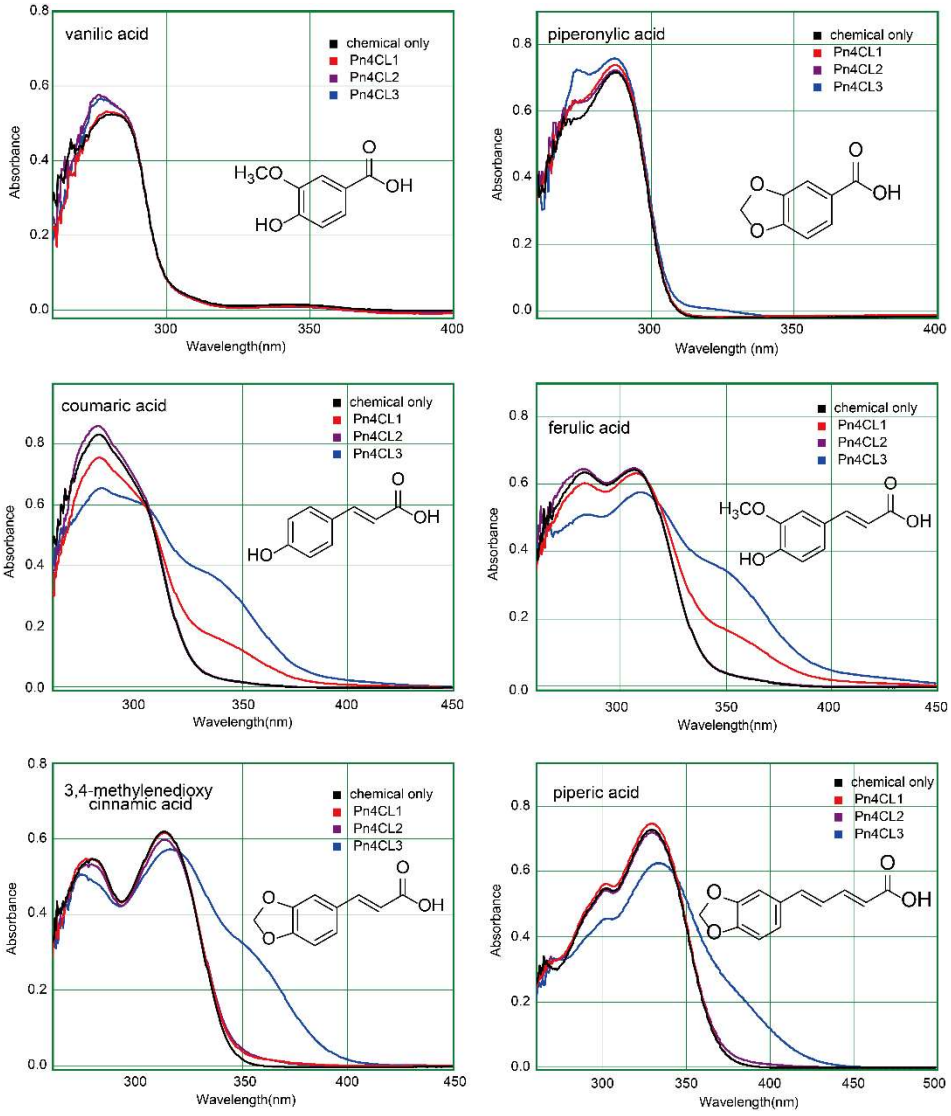

C6C3
ferulic acid
DCS
+ 1
C6C1+2xC2
PnMCHL
C6C1
piperine C6C5
PnPKS
+ 2
C6C1+2xC2
+ 3
C6C1+3xC2

Piperate:Coenzyme A ligation by Pn4CL3

The open reading frames of Pn4CL1 through -3 were cloned in *pET28a*(+) vector between *Nde*I and *Xho*I double digestion site with C-terminal 6×His tag. These constructs were expressed in *E. coli* Rosetta2 (DE3) cells. The recombinant proteins were purified through Ni^{2+} affinity column chromatography. Enzyme assays for the Pn4CLs were then carried out with phenylpropenic acid homologs as substrate: vanillic, piperonylic, coumaric, ferulic, 3,4-methylenedioxycinnamic, and piperic acids. Bathochromic shift of UV band position upon ligation of CoA to phenylpropenic acid homologs allowed spectrophotometric determination of enzymatic reaction (Fig. 3.8).

Pn4CL1 behaved as general 4CL, which ligates phenylpropanoid pathway precursors, such as coumaric and ferulic acids to CoA (Costa et al. 2005; Chen et al. 2013; Rastogi et al. 2013). However, this enzyme could not ligate 3,4-methylenedioxy compounds, such as piperic acid and MDCA. Pn4CL2 had no activity towards these substrates, although the protein had high homology (66.4% identity) to cinnamic acid-specific SbCLL7 (Zhao et al. 2016). Pn4CL3, in particular, could take piperic acid (C_6C_5) and MDCA (C_6C_3) as substrate. Piperonylic acid (C_6C_1) also exhibited minor activity, if any. These results indicate that Pn4CL3 was specialized to ligate compounds with 3,4-methylenedioxyphenyl moiety. Detailed studies to firmly identify the structure of ligated products and substrate specificity are thus necessary.

Figure 3.8. Pn4CL reaction followed by UV-Vis spectrophotometry in scan mode.


vanilic acid
chemical only
Pn4CL1
Pn4CL2
Pn4CL3
Absorbance
Wavelength(nm)
piperonylic acid
Wavelength (nm)
coumaric acid
ferulic acid
3,4-methylenedioxy
cinnamic acid
piperic acid

Piperidine-piperoyltransferase

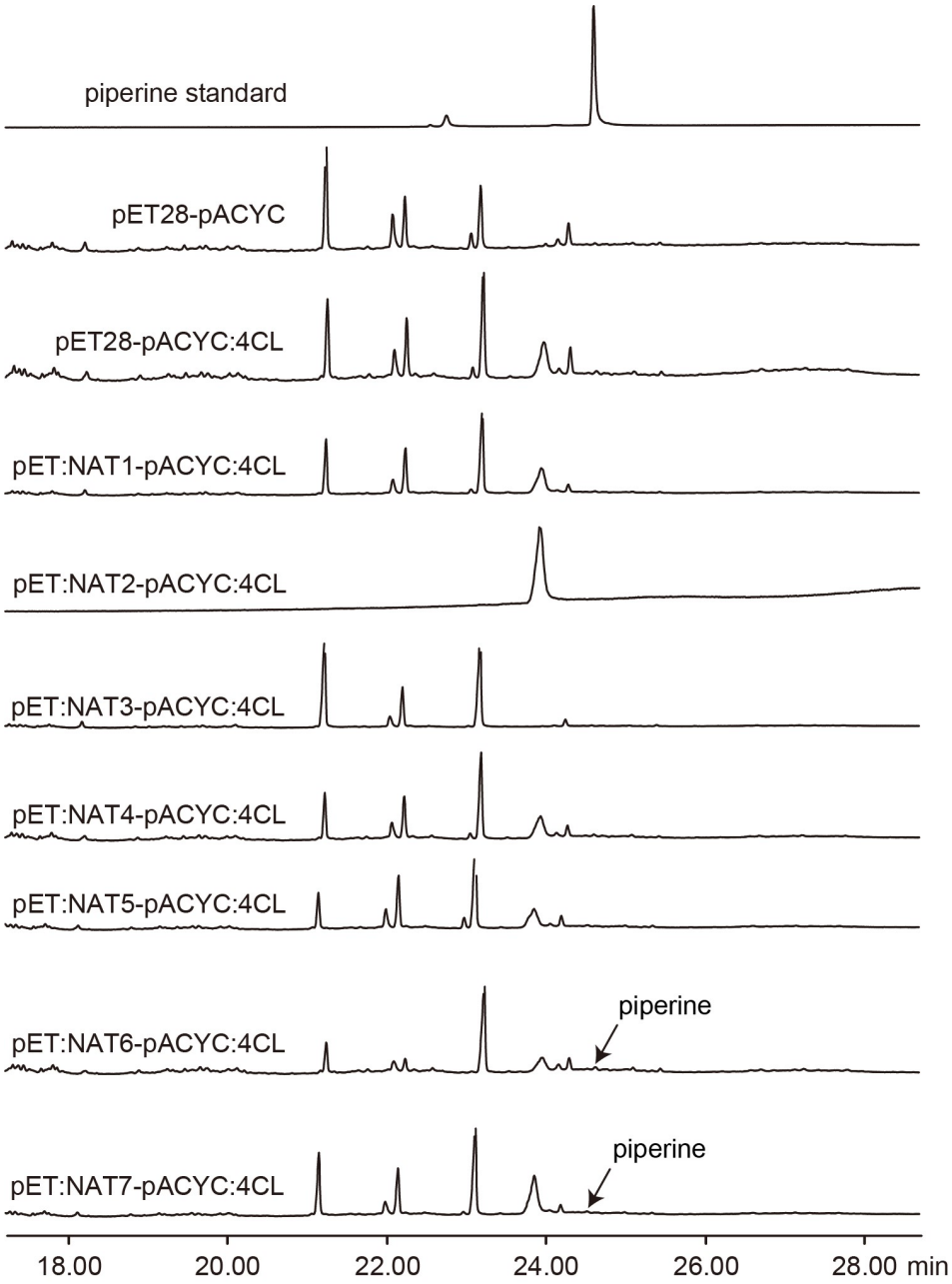
By using bioconversion system with double expression of 4CL and NAT, seven PnNAT candidates were tested. They were selected by the homology to agmatine-coumaroyltransferase, a member of BAHD-type *N*-acyltransferase superfamily (D'Auria 2006; Tuominen et al. 2011; Peng et al. 2016). Therefore, it is likely that PnNAT (piperidine-piperoryltransferase or piperine synthase) belongs to this superfamily. A large number of secondary metabolites are biosynthesized by BAHD-type *N*-acyltransferases in plant.

The *E. coli* system to identify function of PnNATs was consisted of Pn4CL3 to activate externally fed piperic acid to piperoyl-CoA and variable PnNAT to transfer piperoyl moiety to piperidine (Fig. 3.1). PnNAT6 and -7 were shown by GC-MS analysis to produce small amount of piperine under the given experimental condition; a small signal corresponding to the same retention time as authentic piperine standard at 24.5 min was detected in selective ion mode whereas control experiment did not show the signal (Fig. 3.9). However, because the conversion efficiency of the externally fed substrate into piperic acid was low, studies to optimize the conversion efficiency and to unambiguously identify the reaction product are necessary.

The total protein extract from shoots of black pepper was demonstrated to have piperine-synthesizing activity by transfering piperoyl group to piperidine (Geisler and Gross 1990). However, no ensuing study on the transferase has been reported since then. In the present study, we searched transcriptome sequencing data to putatively identify BAHD-type *N*-acyltransferases, and two of them were experimentally demonstrated to possess piperine synthesizing activity.

Figure 3.9. GC-MS analysis of piperine from *E. coli*, harboring Pn4CL3 and PnNATs, fed with piperidine and piperic acid.

The chromatograms were obtained in selected ion mode.


piperine standard
pET28-pACYC
pET28-pACYC:4CL
pET:NAT1-pACYC:4CL
pET:NAT2-pACYC:4CL
pET:NAT3-pACYC:4CL
pET:NAT4-pACYC:4CL
pET:NAT5-pACYC:4CL
piperine
pET:NAT6-pACYC:4CL
piperine
pET:NAT7-pACYC:4CL
18.00
20.00
22.00
24.00
26.00
28.00 min

Conclusion

In this research, an indole synthase from *P. tinctorium*, three sesquiterpene synthases from *P. nigrum* were isolated and characterized. Additionally, piperine-related genes were predicted for experimental characterization, leading to a proposal for a novel piperine pathway.

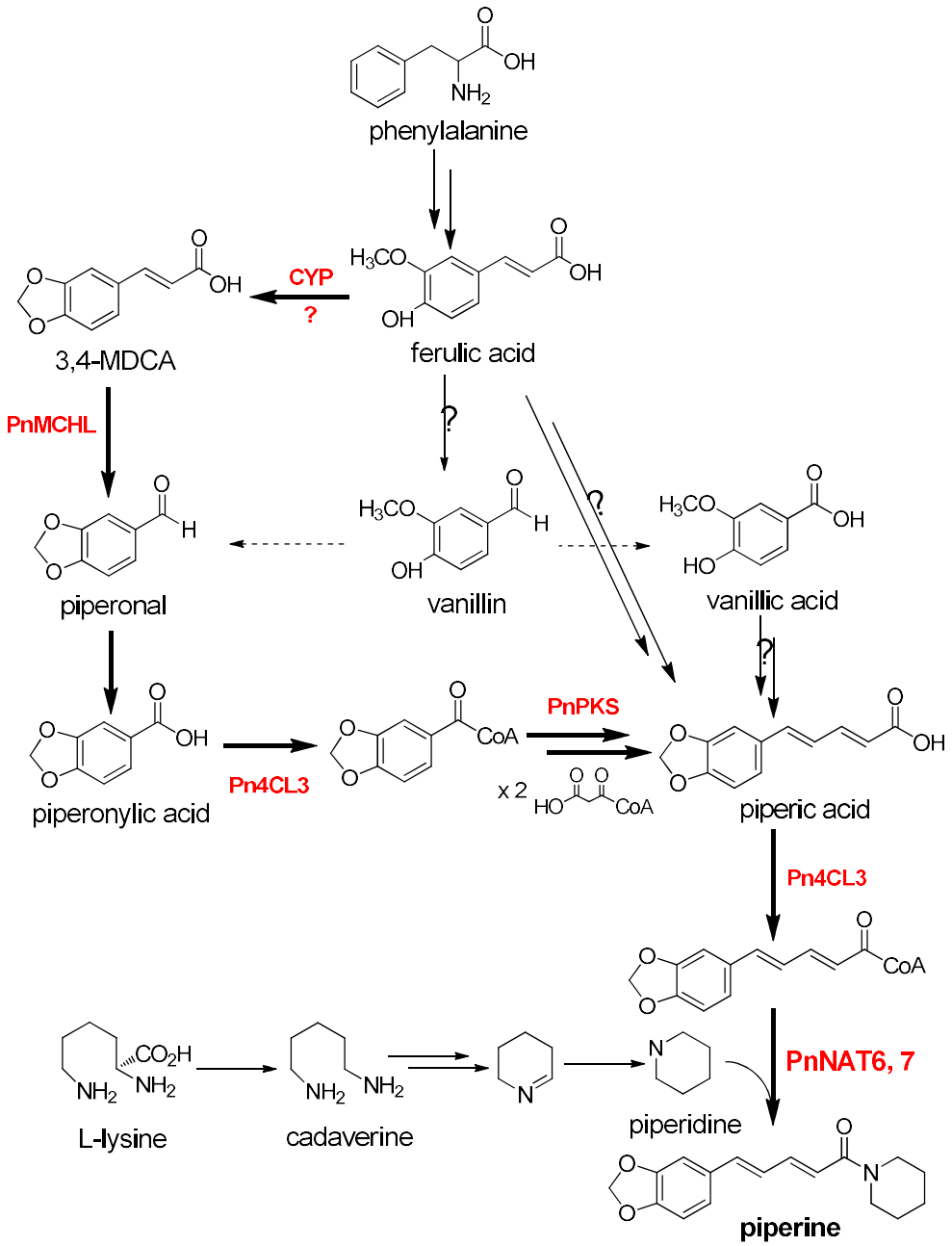
In the first part, two *IGL* cDNAs with different genomic structures from an indigo plant, *P. tinctorium*, were cloned. One of the genes was tentatively labeled as *TSA* based on localization in chloroplast and high homology with other *TSA*s. The other gene was identified through complementation assay, subcellular localization in cytosol, elicitation by BTH, and evolutionary consideration, as *INS* participating in indigoid biosynthesis. Unequivocal proof that *PtINS* is participating in indigoid pathway could be arrived at through forward and reverse genetic approaches. *In-vitro* shoot regeneration system of *P. tinctorium*, already established by Thwe et al. (2012) could be used not only for such genetic study but also for engineering of *P. tinctorium* to over-express *PtINS* in the plant.

In the second and third parts in this thesis, the utility of massive parallel sequencing was demonstrated for the study on secondary metabolism in non-model plant, *P. nigrum*. The transcriptome sequencing technology offered high quality sequence data. Isolation and functional study of three sesquiterpene synthases from unripe peppercorn were based on transcriptome data. These enzymes contributed fragrance to pepper and accounted for about half of sesquiterpenes found in black pepper in number.

In the third part, the transcriptome data were used to identify enzymes involved in biosynthesis of piperine. To be specific, candidate clones for 4CL, type III PKS, and NAT were isolated and an additional enzyme, MCHL, was proposed to participate in piperine biosynthetic pathway (Fig. 3.10).

Figure 3.10. Proposed alternative piperine biosynthetic pathway.

Bold arrow indicates new pathway proposed in the present thesis.


phenylalanine
ferulic acid
CYP
?
3,4-MDCA
PnMCHL
piperonal
vanillin
vanillic acid
piperonylic acid
Pn4CL3
PnPKS
x 2
piperic acid
Pn4CL3
PnNAT6, 7
L-lysine
cadaverine
piperidine
piperine

Supplementary Data

List 1. List of piperine biosynthesis involved genes sequences.

>PnCuE
atggcgtctcgcctcactctcatcccactcttcgtcgtcatcctcgccgccgccgccgctgggtccctatcagatgaggag
aatccgatccggcttgtcaccgacaaggcgcgggaagccgagtcggccattcaccggaccctcggcgccgcccaccatgtc
atggcctttgcccggttcgcccgaaggttcggcaagcaatacagctctgtggacgaaatcaggaagaggttcgacatcttt
gtggagaacttggagctcattcactccaccaacaagcgaggcctctcttataagcttggcatcaacaaatttgcggatttg
agctgggaggagttcaaggctcatcacttgggcgccgcacaaaattgttcagctacgaggggcacccacaagctcacacaa
gctattcttcctgagacgaaagactggagggaagaaggtatagtaagtcctgtcaaaaaccaaggccattgtggatcttgc
tggacctttagtaccactggcgcgttggaagcagcttacactcaagcaacgggaaagagcatctccctttcggagcagcag
cttgtggactgcgccagtggatttaacaattttggctgcaatggaggtcttccttcccaggcattcgagtacatcaaatac
aacggtggtctcgacacagaggagtcctatccatacgccggtgtcaatggcatctgcggatacaagatagaaaacattggt
gtcaaggtcgctgagtccgtgaatatcacagagggcgccgaagatgaattgaaacatgcagtcgctttggtccgtcccgtc
agtattgcgttccaggttgtgcacgacttccgctcatacaaaggaggggtttacacaagtcaagagtgtggtagcgctccc
atggatgtaaaccatgctgttctggctgtcggttatggtgtggagaatggcgtgccatattggctggtcaagaattcatgg
ggaaatgattgggggttgatggttattttaagatcgagctcggaaagaacatgtgtggtgttgctacttgtgcatcttat
cctattctctctctgtaa

>PnCYP1
atggaagacgccgtcctctacctcctgctactcctccctctcctcttcttcttcgcggcccgctgccttcttctacggccg
gcgaagcctccgacgggcccatcggccctcccgatcatcggccacctccacctcctcaagaagcccctccaccgggccctc
gccggcctctcccgccagtggggccccgtcctccacctccgcttcggtgcccgccccgtcctcgtcgtctcctccccggcc
gccgccgaggactgcctgtccacccacgacctcgctttctcgaaccgcccccgtggcctcctcgccggcaagcacttcggc
tccggctacaccaccatcggctgggcccccctacggcccccctctggcgcgagctccgccgcatcgccgccgtcgaggtcttc
tcccacgcccgcgccctccacttcgcccccatccgcgccgaagaggtcgccttcctcgccaaagacctcttctcccactct
cgcggcggcggctccttcgccaaggcgaagctcaggccaaggctgctggagctcactttcaacctgatgatgcggatgatc
gccgataagaggtactacggcgagagggcggaggcgtcggacgagacgcagcgattccgagacaccatcgaagagaccttt
ctcttgatgggcacgtcgaacttggccgactacctccctctcctgagatggatcgacgtcaccggcatcgagaagaagatg
attcggctcgaatcggaaaaggatgccctttacagactttgatcgaagaaaggcgcaaaaatgaaggggacggcaagaaa
gttctgattgatgttctactccaacagcaaaagacagatcccgactactacaccgatcaactcattagaggattgataaag
gcgctgctcaacgccggcacagacacatcggctctaacaatggaatgggcgatggcgctgctgctcaaccacccagacgcc
ctgcgcgcagcgcaggcggagatcgacgacaaggtcggctattcccggctggtggaggaggccgacctgccggagctcccc
tacctcaacagcgtcatccacgagaccctccggctgtgccccacggcgccgttgctgctcccccacgagtcctcctcggag
accgcagtcggcggccgggccgtgccggcggggacgatgctgctgatcaacgtctgggccgtccaccgtgatccggaggcg
tggccagcgccagaggagttccggccgacgaggttccacggcggcggggtggaggggagaaggagcggttccggtttctg
ccgttcgggtcgggtaggcgggtgtgcccaggatccgggttggcgatgcgggtggtagggttagcccttgcggttctcata
cagggggtttgggtgggcccgtgttggggaggatttggtggacttgagggagggcccgggcctgaccatgcctatggcccac
cctttggaagcccggtataagccgagggaggagatggtcccacttctgtctcagttgtga

>PnCYP2
atggccgccttctccatctcctcctcctttgcattccttacccccttcttcgccctcctcctcttctacacccaccaacaa
tggaacaagaagaagctgaccagctccaaaacgaggtcccgaactcttcccccaggcccaccttccctccccatcatcggc
aaccttcacctcctcatcctcaacagcaagaagcctctccaccggaccctcgccggagtttcccgccggtggggcgacatc
gtgttcctccgcatcggagcccgacccgtccttctaatatcttctccggcgactgcccaagaatgctttacgacccacgac
gtcacctttgccagccgcccccctcctcctcgccggcaagatcctcggctacgacttcaccaccctcggatgggcactctat
ggccctctttggcgcgagctccgccgcgtcgtcaacattgagctgttctcccaatctcgcatcctccacttctccctccat
cgctccgacgagatcctctcccttgcaaagctccttttccgacagtcaaacccgggcaatagcagtaccttcaagatgcta
gatctcagacctaggttgatggagctgaccttcaaccacataatgagggttttagctagtaagaagagctatgagcagagc
agggaagaggtggagaaggcacgggagttccaggagctgatcgagacggtgctagcactcgttggggcgtcgaacttggtt
gactacttaccggcgttgggatggttggacgtgcagggattgaggaagaagatggtggagctcgagaagaagaaggacgcg
ttcttgcaagagatggtggaggagaggaggagacgctgccatggggatgagattgggaaagaggcgcttgttgattcactt
cttgagaagcaaaagactgatccaaacttcttcaccgatcaaatcatcaagggcatcatggcggctctattagcagccgga
acagatacctccgccctaacaatggaatgggcaatggcgctcctcctcaaccacccaaaagcccttgaagcagccacggcc
gagctcgacaacgtcatcggccgaaaccgccttgtccaggactccgacctcgacaacctcccctacctccgctgcatcatc
caggagaccctccgcctcttcccctccggcccgaccatccttccccacgagtcctcatcagaatgtaccgtcgccggctac
actgtccccgcaggcacaatgctgctgaccaacatttgggccattcaccgggacccagaggcgtggccggagccagagaag
ttccgaccggaaaggttcgagaagtgggcggagatggacaaggcggtgcagttcaagtggctgccgtttggggcagggagg
agggcctgccccggggatgtgttggccatgaaagtgatggggcttacacttgcgacgctcattcaatgctttgactgggag

agggttggtgaggagaaggtggacatggaagaagggcttggccttaccatgcccatggcgcacccctggtggctatgtac
cggccccgagaggaaatgatgcaccttcttcctcaagtactccagtga

>PnCYP3
atggagcaagctcgatgggtcgacccagctctgttctctgccattgtgggcatcatcttcttcatcctcggcgtgttcttg
agaggcagatcttggagaagcaccagttctaccgagtggccagtcggccctcccaagctgcccatcatcggcaacctacac
cagctcagcgagggcggggaactgacccaccacaggctggccaagctagcccagacctacggccgcaccatgaccatttgg
gtcggcagctggcgccccatgatcgtcgttagcgacgccgaccttgcatgggaggtcctggtctccaaggcggcggacttc
gccggccgggatatgtccaaagtctcgcacttgattaatgcggagtgcaacaccgtctccacctacgacgcggggccgaac
tggcagtcgctccggagaggtctgcagcacggcccgctcggccctgcccatctttctgcgcaagcgcatttccacgaagaa
gacatgaagctcatggcgagcgagatggtgggagcggcgaggaaacgtggtgacggagtggttgaacctatggctttcatc
cggcgagccaccatccgctttatatcacgtctctgcttcgggaagaacttcaacgacgaggtattcatagagggcttggac
aacgcattagatcagacaatcagagtcaccggccatgcacgcatcgtcgacgccttctacttcacccgctacctccccatc
atccgcggccccttcattgacaccgtcaccgtcaagacagagatcgagaacttgatccgaccgctgctctccggaccggcg
ccgacagggtcttacctccacttcctcctctccaccgacgtgcctgagaacatcatcatcttccgaatcttcgaactctac
ctgctcggcgtggacagctccgcatccaccatcacctgggcccttgctttcctcgtctccaaccaaccagcgcaggagaaa
ctccacggtgagctcgctcagtattgtgctactcataacattgagactgtgaaagcggacgacgtgggcaagctgccctac
ttgctcggggtggtgaaggagacgatgcggatgaagccaatagccccgctggccgtcccccacaagacgctcaatcagacc
aagctcgacggcaagagggtggcgccggggacagccgtcgtagtcaacctctacgccatccattacaatccgaaactgtgg
ccggagccggagcagttccagccagagaggttcctcgaaagcggcggtggcgttggtggagggtcgtcggagtacatgctg
cagtcgtacctgccgtttggaggggattgagggcgtgcgcagggatggaggtggggaagatgcagatttccatggtcatc
gccaaccttgtcaaggccttcaagtggttgccggaggtggaggggaagatgccggacatggccgaggacatgaccttcgtg
cttatgatgaagaagccattggtagccaagatcgttcctcgtgcatga

>PnPKS1
atgtcgaagacggtagaggagattcgggcggcacagcgggcgaggggaccagccgcggtgctggccatcggcacggctacc
ccggccaatgtggttttccaggccgattatccggactactactttaggatcaccaagagcgagcacatgaccgagctcaag
gagaagttccaacgaatgtgtgacaagtcaatgataaggaagcggtacatgcacttgtcagaggagctgctgaaaaacaac
cctaacatctgtgcctacatggccccttccctcgacgctcgccaagatatggtggtggtggaggtacccaagctcggcaag
gaggcggccgccaaggccatcaaggagtggggtcgcccaaagtcgggcatcacccacctcatcttctgcactacctccggc
gtcgacatgcccggcgccgactaccagctcaccaagctcctcggcctccgcgcctccgtccgccgcaccatgatctatcag
cagggctgcttcgccggtggcactgtcctccgccttgccaaggacctcgcagagaacaatgcgggcgcgagggtcctcgtc
gtctgctccgagatcaccgccgtcaccttccgcggcccctcggagactcaactcgataacatggtaggccaggcgctgttc
ggcgatggcgcggctgccatcattatcggggccgaccctgaccctgccatagaaaggccactctttcaaatggtatctgca
gctcagaccattcttcctgactcggagggagccatagacggccatctccgagaagtgggtctaaccttccacctcctcaag
gacgtacctgggctcatctcaaagaacatcgagaagagcctcaaggaggagtttgcaccgctgggcatcgacgactggaac
tcgatattttggatagctcatccaggcgggcctgccattctagaccaggtggaggcgaagctgggtctgaaagaggacaag
ctgaagacaacgagaacagttctgagagagtatgggaatatgtcgagcgcttgcgtgttgttcatactggacgagatgagg
aggaggagcatggaggaagggaagacgacgaccggtgaagggttggattggggagttttgtttggttttgggccgggtttg
accgtggagacggtcgtcttgcatagtttgcccatcgccgaggccaactaa

>PnPKS2
atgtcgaagacggtagaggagattcgggcagcgcagcgagcgaggggaccagccacggtgctggccatcggcacggctacg
ccggccaatgtggtgtatcaggctgattatccggactactactttaggatcaccaagagcgaagacatgaccgagctcaag
gagaagttccgacgaatgtgtgacaagtcgatgataaggaagcggtacatgcacttgtcggaggagctgttgaaaaataac
cctaacatctgtgcctacatggccccttccctcgacgcccgccaagatatggtggtggtggaggtacccaagctcggcaag
aaggcggccgccaaggccatcaaggaatggggccgcccaaagtcggccatcacccacctcatcttctgcaccacctccggc
gtcgacatgcccggcgccgactaccagctcaccaagctactcggcctccgccactccgtccgccgcaccatgctctaccag
cagggctgcttcgccggcggtacggttctccgccttgccaaggacctcgccgagaacaatgcgggcgcgagggtcctcgtc
gtctgctccgagatcaccgccgtcaccttccgcggcccctcggagactcacctcgatagcatggtcggtcaggccctgttc
ggcgatggcgccgctgctatcatcgtcggtgccgaccccgaccctgccgtagaaaggcaactctttcaaatggtatctgcg
gctcagaccatccttcccgactcggaggggggcaatagacggccatctccgagaagtgggtctaaccttccacctcctcaag
gacgtacctgggctcatctcaaagaacatcgagaagagcctaaaggaggcgtttgcaccgctgggcatcgacgactggaac
tcgatattttggattgctcatccaggcgggccggccattctagaccaagtggaggcgaagctgggtctgaaagaggagaag
ctgaagacaacaagaacagttctgagtgagtacgggaatatgtcgagcgcttgcgtgttgttcatacttgacgagatgagg
aggagcagcatggaggaagggaagacgacgaccggtgaagggttggattggggagttttatttggttttggcccgggcttg
acagtggagacggtcgtcttgcatagtttgcccatcgccgaggccaactaa

>PnDCS
atggagaaggaagagagtagtaccgtaaggaagctatttaagccccaaagggcggaggggccggccaccgttctggccatc
ggcacggcaacgccagctgcttttgtggagcaagctacctatcccgatttctacttcggcgttaccaacagcaaccacatg

actgatctcaaagacaagttcatacgactgtgtaaggggtcgacgattgagaaacggcacatccacctaaccgaaggaata
ttgatggaaaaccgtaatctctgctcctacgtgacccctctttggatgtgcgccaagagattttggcggcggaggtgccc
aaactggccaagaaagctgccatgaaggccatcacggaatgggggcagcccatatccaagattacccacctcgtcttctcc
accacttccggcgtcgacatgcctggcgccgactacaagctcctcaacctactcggcctacaaccttcagtgaagcgcttc
atgctctaccatgttggctgcttcgccggcgccaccattctccgccttgcgaaggacctcgccgagaacaatgcaggagcc
cgggtgcttgccgtctgctctgagatctccaccgtcaccttccacggccccagctacaaccacatcgacagcctcgtcgcg
caggccatactcggggatggtgcggctgcagtcgttgtcggagccgacccactccgtgaaattgaatggccgatctttgag
attgtgtcagcgtcgcaaacaatcatcccagggagtgaagggaggatcaaggggcggcttcaggaggccggattggtatac
cacttgctcaaggacgtaccggggctggtgtcgaagaacatacataactacctaaaggaagcgtttgggcctctgggggtg
gaagattggaactctatcttctgggttgctcacccgggagggcgggcaatcctagacaaggtggaggaggagctgaagatg
agtccggagaggctggcagcttccaggaaggtgatgagggagtacgggctcatgatgagtgcgacggtactgttcgtgatg
gacgagatgaggaggaggtcggcggaggaaggggcagccacgactggggaggggcttgagtggggtgtgctcctcggcttc
ggccctggcctcacagtggagacgatggtgcttcgcagtcttcccttatcccaaactggaaccttgtaa

>Pn4CL1
atggtagtagacaatattgcaaccgatcaatgtgtataccggtctaagctgccggacattgagatcaagaacgatgtgtcg
ttgcacagttattgttttcagaacattcacgcttaccgggacaagctttgtctcatcaatggaagcaccggcgaggtgtac
acgtacggcgaggtggagacgtcggccaggcgggtggccgccgggctgcaccggatgggggtgcagcagcgggaggtgatc
atgatcctcctccccaactcgccggagttcgtcttcgccttcctcggcgcctcctaccgcggggccatgtccaccaccgcc
aacccttttctacacgccgcaggagatcgccaagcaggtcaaggcctccggcgccaagctcatcgtcaccatggccgcctac
gtcaacaaagtcagggacctggccgaggagcgcggcgtcaaagtggtgtgcgtcgacgcgccgccccgggggtgctcccac
ttctccgagctgtccggcgccgacgagtcggagctgcccgaggtggagattgaccccgacgacgtggtggcactgccgtac
tcgtccggcaccaccggcctccctaagggagtgatgctcacacaccgtggccaggtgacgagcgttgcccagcaagtcgac
ggcgagaacccgaatctctactaccggcaagacgacgtcctgctctgcgttcttcccctcttccacatctactccctcaat
tcggtgctcctctgcggcctgcgggtcggggcggcgatcctcatcatgcagaagttcgagatcacggcgctgatggagctg
gtgcagaagtacaaggtgaccattgcgcccatcgttccgcccatcgttctggccattgccaagagcccgctcgtcgacaag
tacgacttgtcgtcgattcggacgctgaagtccggcggcgctccgatggggaaggagctcgaggacgccgtccgggccaag
cttcccaacgccaagctcggccagggctatgggatgacagaggcagggccagtgctgtccatgtgtttggccttcgccaag
gagcccttcgagatcaagtctggttcttgcggcaccgtggtcaggaacgcccagctcaagatcgtcgacccagaaaccggt
gcctacctgcccagaaaccaacccggcgaaatttgcatccgaggatcccaaatcatgaaagggtatcttaatgacccggcg
gcgacgcagaggacgatcgacaaggaagggtggctgcacaccggcgacatcggctatgtcgacaacgacgacgagctcttc
atcgtcgataggttgaaggagctcattaagtacaagggcttccaagttgcccccgccgagctcgaagccatgctcattact
catcctaacattgcagacgccgccgttgtcccgatgaaagatgaggcagcaggggaagtgccagtggcatatgtggtgacg
tccaatggatcagtcatcggtgaggatgagatcaagcagttcattagcaagcaggtggtgttctacaagcgaatcaatcga
gtctttttcgttgattcaattccgaaagcaccctctgggaagatattgaggaaggatttgagggcaagattggcagctgat
ctgcccaaatag

>Pn4CL2
atggagaaatcaggttatgggaaggatgggatttacagatcacttaggccacccattgccttccccaaggatccaaacctg
tctttggtctcttttctgtttcagaattccgatgcttacccggatcgtcttgctcttgctgatgctgtttctggggagtcc
ttggacttttcccagttcaaatccttggttcagaaggtctcccatggattgtccaaattgggcgtaaagaaaggggatgtg
gtgctaatattttctccgaattccatttacttccccgtttgtttcctggccactgtagcccttgggggagttgccaccact
gtgaatcctctatatacaactgcggaaatcaaaaagcaggttgatgactctaaagccaagcttctgatcacagtccctcag
ttgtgggataaggtcaaaaatcttggcctccctgcagtgattcttgcaagaaaaatcggtggagatggaagtattgcttcc
aagactagtgtcacctactttcatgatttggttaaactgggcgggcatgtatcagaatttccgcctgtttcgattaaacaa
tcggatatggctgttcttctatactcctctggtaccactggcacgagcaaaggagttattctgacccatagaaactttatt
tcggctgcttgcatggtgattgccgatcaagactttgacggtgatggaccgaacgtgtttgtttgttttctgccaatgttc
catgtatttggacttgtgatagtttgttattctcaattaaagaggggaaatagtgtgatatcagttgagaaatttgacctt
gatgtggtgttgaaatctgtggagaaatatcggattactcacatgtttgctgtgccgcctgttatgattgctcttgctaaa
gaaaattgggggaagaagtatgatttgtcatccttgaagaggattggctctggtgctgctcctcttggaaaggaggtggta
gaagagtgtgcaaggaaatacgctcacaccgcaattgtccagggttacggtttgacagagacttgtggaattgtctcctta
gagaatccaaatggtggaggtcagaaatatggttcatcggggattctttgttcaggaattgaagctaagatagtcagtgta
gaaaagttgatacctctgcctccgaaccaactgggcgaaatatggttaagaggcccaaatattatgcaaggttatttcaac
aatccccaggcaacgaaactggcgattgataaacaaggatggtttcatacaggggatctcggttacttcgacgacggagga
aggctatatgttgttgatcggattaaggaacttatcaaatataaaggatatcaggtggctccagcagaattggaagctttg
cttctttctcatcgagagatattagatgcagtggttatcccatatcctgatgccgaagctggtgaagtccccattgcttat
gttgtgcgcgctccaagtagctcattggatgaagaagctgttaagaaatttattggtgagcaggttgcacccttaaacga
ctgcggagagtaacatttgttaacagcgtgcccaaatcagcttcggggaagatacttagaagagagcttatcgagaaagtc
cggtccaagatgtaa

>Pn4CL3

atgatttctgtatctgctgttcaagctccccaatcccagcttctggatgatccttctactgcttcaattattgcagagcta
tccactgaaaaggaacgctttgtttttcgatcaaagctgccagacattatcattcccaaccatctccctcttcatgcttac
tgctttgagaaggtagaggagctctcagaccggccatgtttgatcgatggcgccggcggtagagtgtactcgtttccagag
acccatcaaatgtgcaggaaagtggcggccgggttggggaaggtgggagtcgagaagggtgatgtgatcatggtcatgcta
ccgaacagccccgagtttgtgttctgtttcatgggggcttcgatggccggggccgtcaccaccaccgtcaatcccttctac
acgccggcggagattttcaggcagctggccggctccggcgccaagctcgtcgtcactcagtcgcagttcgccgacaagctc
aaggaatccgccgagcagctcaacttcaaggttgtgacggtcgacgaacccccacccggctgcctccatttctccgacctc
ttgaactccggcgagatcaatttaacgtccgtgtcgacattcgattcggagggcgccgtggcaatgccattctcctccggc
accaccgggctccccaaggggtggtgctcactcacaagagtcttatatcaaccacagcacaacaggtggatggggagaac
ccgaattttcacatcaacgccggagacgtcgtgctctgcatgttgccgctcttccacattttctctctctcgatcctgctc
gcatccgttcgggccggggctacggttttgctaatgcataagtatgagatggggaggctattggagctggtacagcagtac
caggtgaccgtgacaccggtggtgccgccgctagtgctggccctcaccaagagtgcagcggtggagaagtatgatctgagc
tcgataaggatggtaatctccggagccgctcctcttgggaaggagcttgaagaggcactcctcaatagactccctcaagct
gtttttggccagggttatggcatgactgaggcagggccagttctgaccatgagcctgtccttggccaagaaccccttccct
accaagcccggctcatgtggcacggtcgttcgtaatgccgagctcaaggtcatcgaaccggccaccggaatctccctcgca
cggaaccagccgggcgagatatgcataagaggacctcaaatcatgaaaggatatttaaacgacgaagaggcaaccgcgttg
accattgatgtggagggctggctccacaccggcgatatcggctacgtggacgacgacgacgaggtgtttattgttgacaga
ctgaaggaggtcatcaaattcaagggctttcaggttcccccagcagagcttgagagcctccttgttagtcatccatctata
gttgatgccgccgttgtaccacaaaaagatgaagcagccggagaagtccctgtggctttcattgttcgatcaaatgggttt
gatctaactgaagaagctattaaagacttcattgccaagcaggtggtgttttacaagaagctacacaaggtctttttcgtc
cacgcaattcccaagtctccatctgggaaaatattgaggaaagaactgaaagccaagttgacaaatacttaa

>PnNAT1
atgataataacagtgaaggaggagacgatggtgacgccggcagagccgacgccggtaaggagcctctggaacgccaacgtc
gacctggtggtgcccaggttccacactcccagcgtctacttctaccgacccaacggctcccccaacttctttgagccgcag
attctcaaggacgcattggccaaggccctagtccccttctaccccatggccggccgccttggccgcgacgttgacggccgc
atcgagatcaactgcgccggccagggcgtccttttcgtcctcgccgagaccggctccgttattgatgacttcggcgacttc
gcgcccaccatggagctcaagaagctcatccccaccgtggactactctcaagatatctcctcctatccccttctcgtcctc
cagattactcaattcaaatgcggcggggtgtccctaggagtggggatgcaacaccatgcggccgatggtgcgagcgggcta
cacttcgtcaaccattggacggacgttgcgagagggctggacttgaccatcccacctttcattgaccggacgctgctccga
gcacgggacccgcccaagccggcctcccaccacgtcgagtaccagccggcgccgccattgaagcaggcaccgcccgaggag
gaccagcagcagcaccagccccagccccgccccgcagccgtctccatgttcaaaatcactcgcgaccagcttgccctgctc
aagtcccgcgctccctctttcagctcctacgagctgctggccggccacgtttggaggcgcgtgtgcgaggcccggggcctg
atcgacgaccaggacaccaagatgtacatcgccacagacggccgctcccgcctccgaccgccgcttccccagggctacttc
ggcaacgtcatcttcacggcaacccccatcgccgccgccggcgatctcaccggacgtcccccaaagttcgcagccgggaag
atccacgacgcgctggtgaggatggacgacgagtacctccggtcggcgttggattacttggagcagcagccggacctgtcg
gcgctggtgcggggcgcccacactttcaagtgccctaacttgggcatcaccagctgggcgaggctgcccatccatgacgcc
gacttcggatggggccgaccaatcttcatgggccccggcggcatcccctacgaaggtctatcgttcgtgctgcccagtgcc
accgacgatggcagcctcaccctcgccatctccctccaagaacaccatatggaccgcttcaagaagctcttttatgatttc
tga

>PnNAT2
atggaggtggagatcgttagcagcgagctcttgaagccaaccattgcccaaccaccgaacgccgccatccctctcaatatc
ttcgataaggccaccttcaacatccaagtccccatgatctacgtcggccgcacacccatgcctccaatagccaccctcaga
gaaggcctggccaaggccctcaacgacttccccgtccttgccggccgcatcggcgtcgaccagcagtctcgcccagccatc
ctcctgaacgacgccggcatccgcctcaacgtcgcgagcgtcaatgccgacatggaggagatgttcccggtctacccggcg
gagaggatgacgccactccacccttgcattgaagacatagacgagctgttcttggtgcagctcaaccagttcaagtgcggc
ggggtggtcatcggcctctccatccaccaccgggccggcgacggccagtccacgagcaacttcctcgttgcatggtcgcag
atagtccgggggctcgccgtcgaccccatgcctaaccacgaccgggccgccatcgccgccccgagaaggcatctcaacgtc
gagttcaaccataaggacatcgagttcagaaacgacccatctccaccacccaccgtcgactcctcggtcgacccaatcgaa
aacgtcatctcccactattctgtacaagacatcaaatcgttgaaatcgaccgtagaacgatcaggtcagggaagaaaggcc
actacattcgagtgtgtgctagcagacatttggaagaagatcacacttgcaaggaagatcaagggagacgagctcacccat
gtaagaatagcagttaccggccggtctagaatccaacaagtacccatggagtacttcggaaacatcgtactatgggcacac
ccaacactcaaagccaaggacctggtctctaccagcctcgcctctgtcacaaaagccatccgagactctgtttctcagatc
gataacagatacttccaatcgttcaaggactttggagcagtcgtacagccagacgaggacctggtggaggctgcgcctgac
attgccgactctttgagcccccatgtggagatcgatagctggttgaggcttcaattccatgagatcgatttcgggagcgga
gcgcctcatgctttcattccggcatgcctgcccattgagggattgatgatcttcttggcgtcttgcaccgagaagggtgct
gccgacgtggtggtggctcttcacaagagcactgtgccttatttcaggcagattgagagcgcgatgaagatgaagccaaca
agcaaattgtaa

>PnNAT3

atgagttctcttcttcttctcagcaataccgttctgaaaccgtcaagaccagcctcgaatgtcttggtaactctaaccacc
ttcgacttggccgcgatggactcttacttcaaagtcatccgcgccttccaacccccaatgccatctaattctgtcatgaaa
gatggcctcgccaatgccctcaaccacttcccccaccttgccggccgtttttgttcaaacacccaactcgaaccctgcatc
ctcttgaacggtgctggtgtccgactagtcgagactcgcctaatgacccatctctccgactggttgaagcactcttttaac
ttgtcatcgtcgccggaggcctgctgcctagcaccggagacggctggcactgaggagctgctccacattcagctcaaccgt
tttgcctgcggtgggatggtcatcggagttgtggctcaccaccgagttgccgatggcccggccatgagcttgttcttggta
gcatgggcaaggctggtccagggcttggacgtgggtcgtattccatcacatgataggtccgccattgctacacatagagac
cctccattgtgtgaattcaaccaccaggacatagaattcagaagaggaagattgcagactatgtgtgctatgaatgcctca
acctcctttgtaacgttaaaggtgcacttttcttccagcttcatctccatcctcaagctcaagatggctccattaatggca
gcagacagcaaccacgactgtcatcaccgcttcaccacctttgagtgcttgttgacccacgtctggaagaagatcacccaa
gttagaggactggacaacaagcaaaccacccaggcccgaattgccataaacggcaggcgacacggccggacaaaggcgccg
gcggagtatttcggcaacttggttttatgggcacgccccgagtttacagtcgaagacttgatgaacgaatgcaactctgcc
agtgttgctcaaacaattcatgaagccttggcctcggtcgatgagaggtacttccaatcatttgtggacttcggggagatg
gaaaggaagaagggatcatatggtgcattgtcgtcgtcgacggcggaggccggaggggactgtgagcccagatatagaggtg
gatagctggctggggatggcatcgaccagccgagggcttgattttggcggcggcaagctctgtgcagcaggtcctccagac
ctgccatatgaagggttggtgataatggtgccttcatgtgtagaagaaggtagtgttgaaatgttcttgtctctccatgat
aagtctgtggcaaggttcacggggattgtgcattctttgcaagggattgaagagagaagcatgctctaa

>PnNAT4
atggcagttgagattaccaccaacgccattctcaaacccttccccacatcccctccctccccccacaacaatgccgtggtg
cccctcacagtcttcgacaaggcggccttcgacctccacgtcgccgtcctttatgcctacagaccgccgatgcccgcgaac
gaggccatgaaagaggggctggccaaggccctgaacattttcccccacctagcaggccggctcagcacggacgaccaaggc
cgccagtgcatccaactcaatggcgctggcatccgtctcattgaggctgatgtacccgtgaccctcgccgagcggcttccc
ctcgacccgtccccggagctgatggatttccggccaccaatcgacggtgtgaaagagttgttccaaattcaactcaaccga
tttgcatgtggtggtctggtaataggccagaccgcacatcaccgggtatgtgatggccaatctatgagtgcatttttgta
acatgggccaggttggtccgtggcctagacattggacccccttccattccatgatagagcatctattgcaatcccaagaaac
ccacccatttgtgagttcaaccatagagaaattgaattcaagggtcctgataacttaaacccatcatctgacccaagtgca
aaagttgtttctttatcatcaattgacaacttggtggtaacattcccctccaacttcatagcccagctcaaatcaaatatt
gcaaaagactgcccgggccagcggttcagcacatttgagtgcctcctcacgcacctatggaagaaggtgacctacgcccga
gggctcgccggcgacgacaccacccaagtccgagtcgctgtcaacggcaggacccggatgcgtccggcggtgccgatggag
tactttgggaacatagttttgtgggcataccccaagctcaaagtgaaagaactcattgacatgaaatatggagatgtggcc
aaggccatacatgatgcagtagccaacatagatgagaggtatttccaatcatttgtggactttggggagatgatgaaagat
gggcatgaagaggataaaagtttggaggctacagcagcagaggttgggcagtctttgagccccaatatagaggtggatagc
tggttgagattccaatttcatgagcttgattttggtggtggagggccttgtgcttttctccctcccaatatgcctgtggag
gggctgctaatcatggtgccctctagcaaagaaatcggaggtgtggatgtttacatgactctcttcccagagtgtgtcccc
aagttcaaacatgtttctcatacattgagctga

>PnNAT5
atggaaaaggcttttgagcttgttgtagagcaaggagagccagttctcatcccgccatatgaagaaacagagcgagggctg
tatttcctgtccaaccttgatcaaaacatagctgttgtggtacaaactgtttactgtttcatggccaatgacaaaggcaat
gataatgctgcccaagttataaaggatgccttggccaaagtgttggttcaatactacccagttgctgggaggctcacaata
agtgcagaggggaagttgatcgtcaattgcactggggaaggagctatttttgtagaagcagatgcaaattgtgcgctggaa
gacatcggagacgtaagcacaaaaccagatcctgcaacactcgggaagcttgtttatagtgttcctggtgcaaagaatatc
ttagaaatacctcctctggtagtacaggtgacgaggttcaagtgtggaggctttgttcttgggctggccatgaaccactgc
atgtttgatggcataggcgcaatggagtttgtgaactcatggggagagattgcccggggcctcccaattgcaatcccaccc
ttcaaaagccgggcactcctcaaagcaagagaccctccaataattgagttcccccactatgaatttgctgagattgaagat
gtttccgacacctccaaactctaccaagaagagatgcagtacaggtccttctgcttcacgccagaaaagcttgaaaagctg
aaaaggttagccatggaagatgcgtacttgcccaagtgcaccaccttgaagctctctctgcttttgtatggaaggccagg
tccaaggccctcaaaatgcagcctgaccaacagaccaagctgctctttgctgttgatgggcggtccagatttgagccacca
ttgcccaaaggctactttggcaatggcattgtgttaaccaactccttgtgcagtgcaggggagatgctgggaaagcctctg
tctttcactgttgggttggtccaaaaagctgtcacaatggttacagatagctacataagatcggccattgattactttgag
gtaaccagggaaaggccctctcttgctgcaactcttctgatcactacttggtcaaggttgtccttccatactaccgatttt
ggggtggggagagcctgctcaatctggaccggtggctttgcctgagaaggaagtgatcttgttccttccccatgggaaggag
aggaagagcatcaatgttcttcttggcttgcctgcctcggccatgcctgtttttgaagaggaaatccaaatttga

>PnNAT6
atggagatcatccgcaatagaatcctgaagccgttgcatgccaaggccaatccatgcaaagcaatcccactcaccatcttc
gacaaggccgccttcaatctccatgtcaccgttctctgcgccttcctaccacccatgcccaccaacgacgccctcatacaa
ggcctcaccgaggccctagcccacttcccccacctcgccggccggctctctgacaccggcgtcgtcggccaccctaccatt
ctcctcaatggggccggtatccgtctagtggaagcgaatgtgagttctaccctggccgaggcgctgccgttgaagccgtca
gaggagatgagtttgctgcacccgcccattgaaggtgtggaggagctgctccagattcagctcaaccgatacacctgcggc
ggggtggtgatcggccagacggcccaccaccgtgtctgcgacggccaatccatgagctccttcttccttgcatgggcgagg

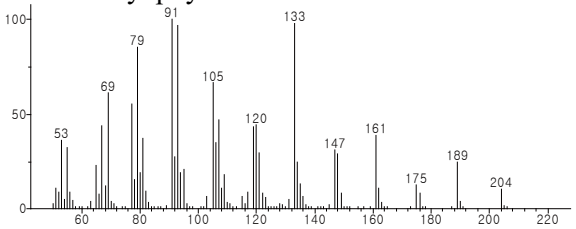
ttggtaaaggggatggaggtcgaccctattccctaccatgaccgagctgctgtctcggtcccaagaaacccaccaagaacc
gagcacaaccatggccaaatcgagttcaagaacaataaaagcgactcgactgcacttccttcttcttcctcgtctagcatt
gtcaacttggtgctccactactcggcagacttcatatccgagctcaagaagctcgtcaacgatagcgtaggcgacgacatg
ccgagtctgcgatgcaccaccttccaatgcttcctcactcacatatggaagaagatgactctagcgagagggctcaaaggc
gaagaaatgagccaagtgagggtggccgtcaatggccgctttcgaatcaagccctccgtcccagcagagtacttcggtaac
ttcgtgctctgggcgtacccaagattgaaagtggaggacttgctaaaccagagttacaagttcgtcgccaaggccatcaac
gaagcggtggccaaggtcgatagcgggtatttccagtcgttcatcgacttcggggagttgaagaaggggggaggagctgcag
cagtcggcagcggccgtggggaacatgttgtcgccaaacttggaggttgatagctggttgagcttcaggtttcatgagatt
gatttcggccatggcggcccttgtgctttccttccatgttacctgccaattgaaggtttgctgctgcttgtgccgtccccc
aaggagaaaggtgctgtggacgtcttcattgcggtagcggaagatgatgtggagcggttcaagctggtttcacactctttg
gcatag

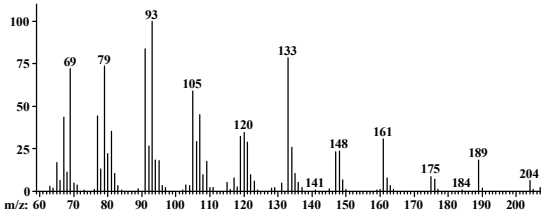
>PnNAT7
atggaggtgaagataatcagtgcagacatgatcaggcctgcttctccaacccctcctcaccttcgtcacttcaatctcagc
tatctagaccagctctccccttacatgttcatccacctcatcttgtactaccctcgccatcaaatgggcaaggctaccctt
actactgccgacatatctcgccgtctcaagacctctctctctgaaagcttggtgcaattttacccactggccggaaagttt
gtggccaaagaaaacgagatcgaatgcaacgaccaaggagtggagttttatgaaacaagagtactccattgccaattagag
gagctccttaaggaacctaacaagatcagcttgtccgagctcgttccttgttactttacatgcaccaaacactctggaaga
gaagatgctctcttggccgtgcaggccaacttcttcgactgcggcggattggcactgggagtgagcctatctcatctcatc
gctgacgctgcctctttcgccactttcctcaacggctgggcctccactgcaggcgagggggcggtcccggccaaaagtcg
ccgcttgtcgtccatcaccacctcttcccgcatataccagggctacccaggcccgtagatggtagatttgctagcaataat
atcaagatgaagaggtttgtgatcgacgagtgcgacctggcagctctccggcggaagcgaggcggcgacgccactcgcgtg
gaggccgtcacggctctggtgagcaggtgcatggcgagggcgacgaggaaaaccagcggggtcgatgttagggatgagaga
ggtttcatggtcgtccatatagtgaatttccgacggaggatggagccgccattgcccgatgatgcctttgggaacttctgg
gtctttgttgctgtggcgccggagggctcccccgctgacgaccagaagggtagtctagaggtatcactcagagacactttc
aaaaacgttgatgcagctttcgtgaaggagctgcaagggccagatgggtataagaaggcaatggagtttccgatggcacta
ataaaaaagcgcaaagagaaggaggcgtcggcctttgtcttctcgagttggtgcaagtttggatattataacgtggatttt
gggtggggaaagcctcaattggtgagccctgctcagataccttgaagaatcttttcttcttccttgatcagaaagaggga
gatggagttgaagtgtgggctaacatggaagccggagacattatcaaattcgaacaggaccaagaattgctttcctacgtt
tatccgccatatgactaa

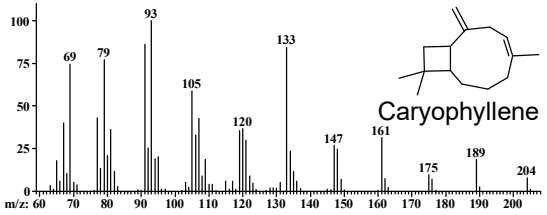
Yeast	*In-vitro* assay	NIST MS library

PnTPS1

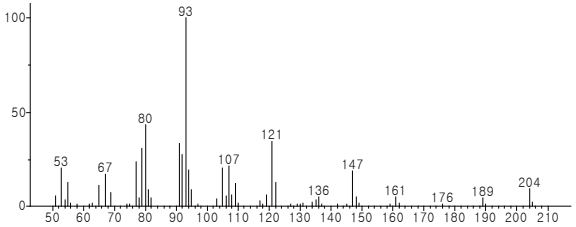
Peak 1 caryophyllene

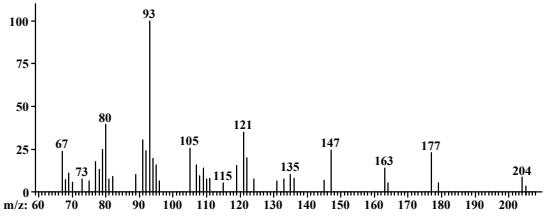


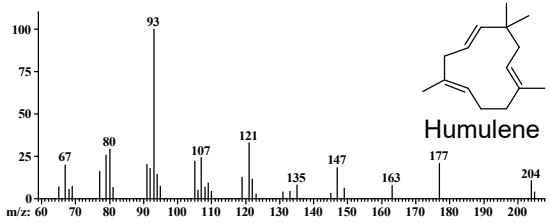




Peak 2 humulene



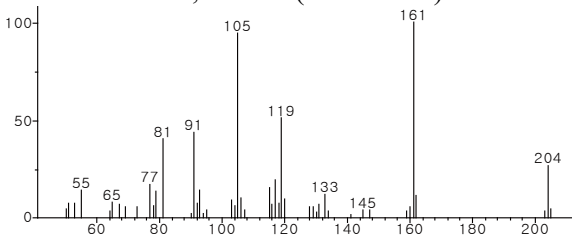


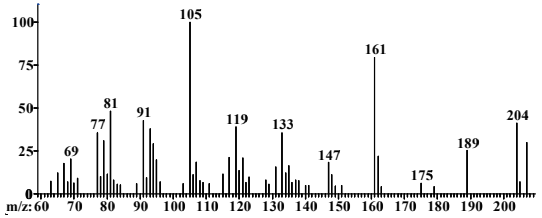


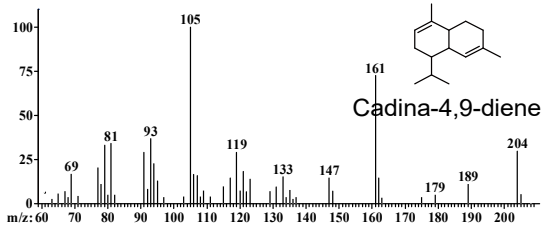
Yeast	*In-vitro* assay	NIST MS library

PnTPS2

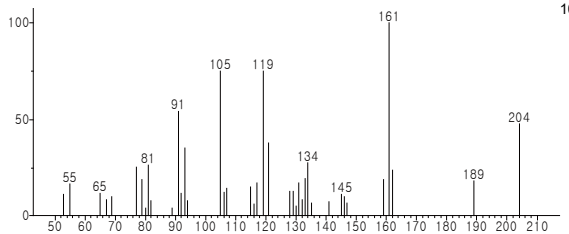
Peak 3 cadina-4,9-diene (α-cadinene)

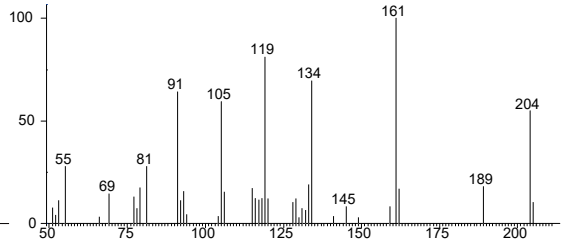


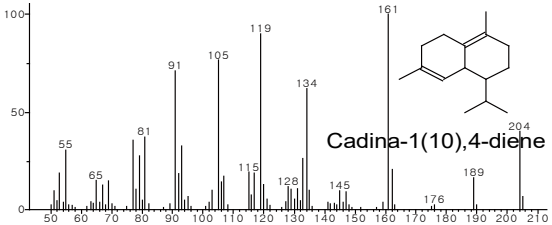




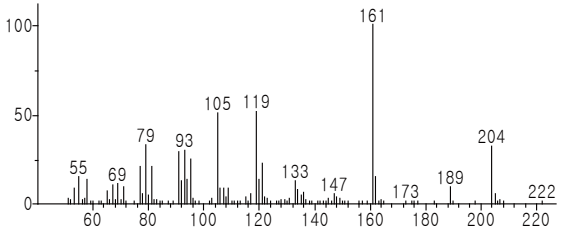
Peak 4 cadina-1(10),4-diene (δ-cadinene)

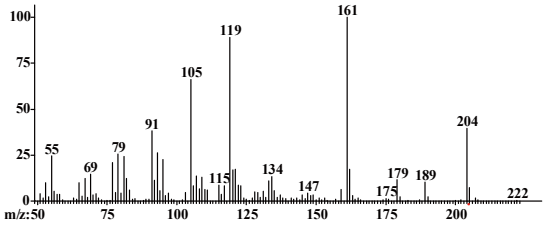


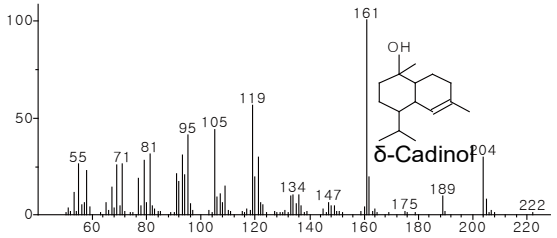




Peak 5 δ-cadinol



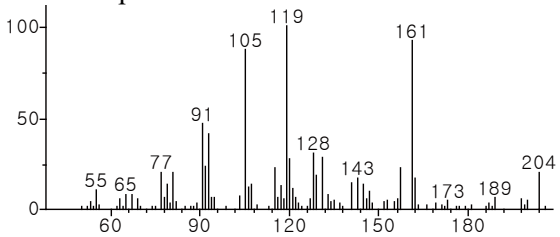


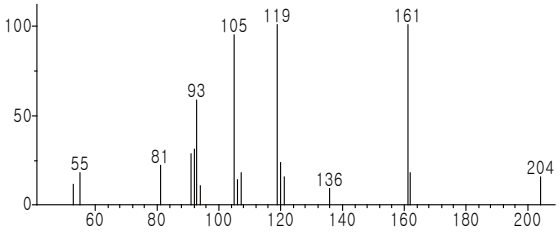


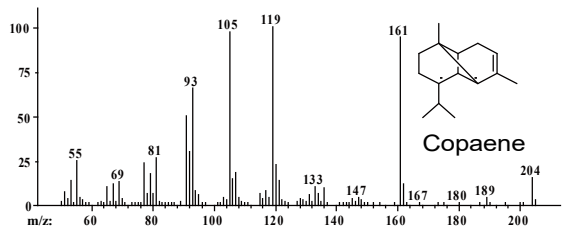
Yeast	*In-vitro* assay	NIST MS library

PnTPS3

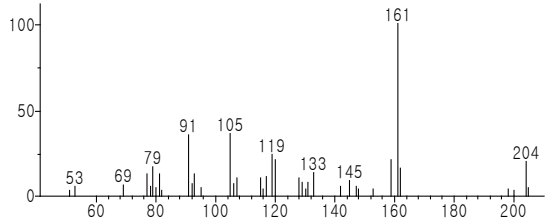
Peak 6 copaene

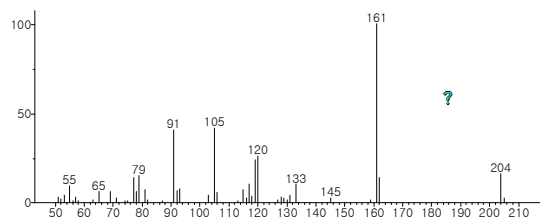


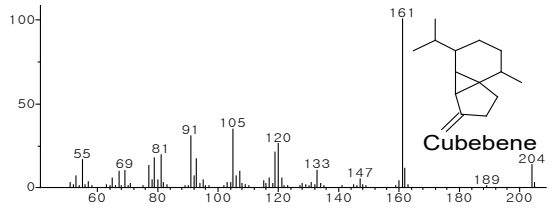




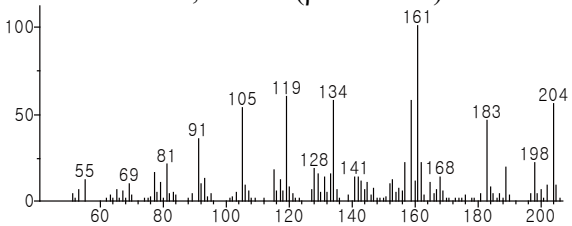
Peak 7 cubebene

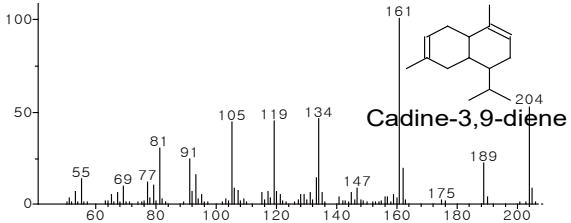






Peak 8 cadine-3,9-diene (β-cadinene)





Peak 9 germacrene D

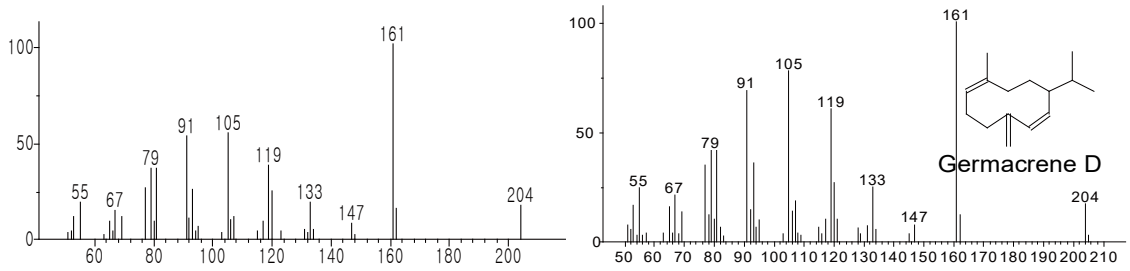


Figure S1. Mass spectra of PnTPS products, denoted as peak 1 through 9 in Fig 3.

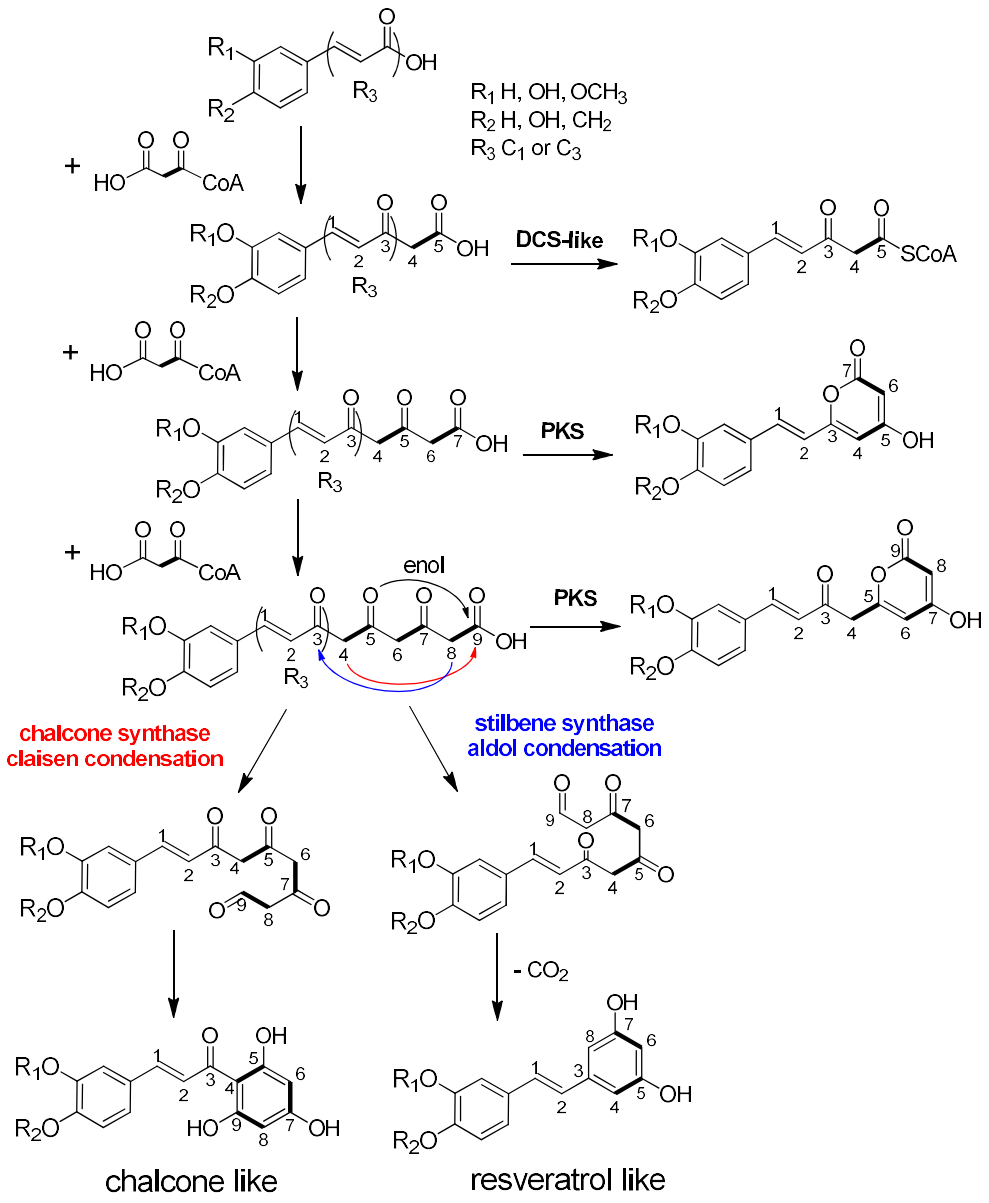


Figure S2. Chain extension of phenylpropenic acids and benzoic acids by type III polyketide synthase in plant.

DCS, diketide CoA synthase; PKS, polyketide synthase

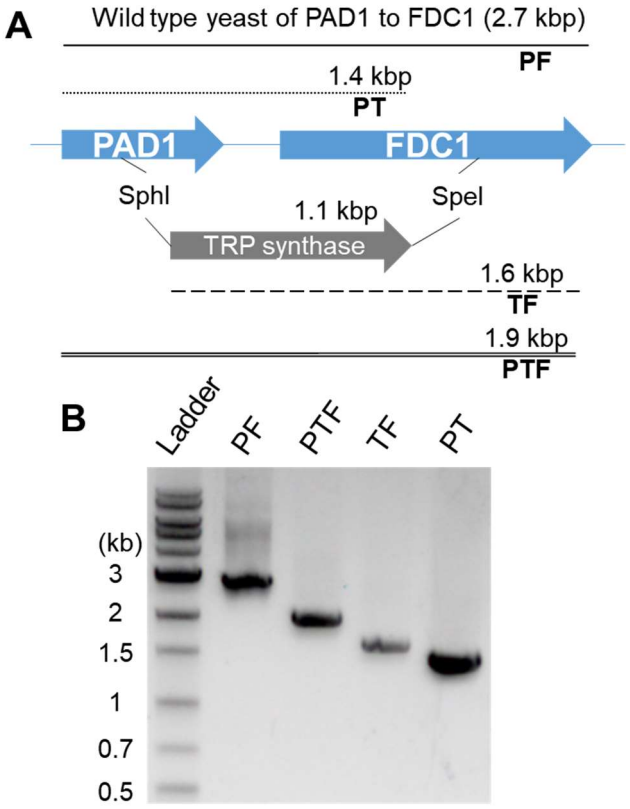


Figure S3. Yeast gene disruption strategy and knockout-mutant conformation in YPH499 strain. The mutant yeast was named yeast YPH499 *ΔPTF*.

A, mutant preparation method. B, agarose gel loading of the mutant of PCR products. PAD. (P) phenylacrylic acid decarboxylase, FDC (F) ferulic acid decarboxylase, TRP (T) tryptophan synthase.

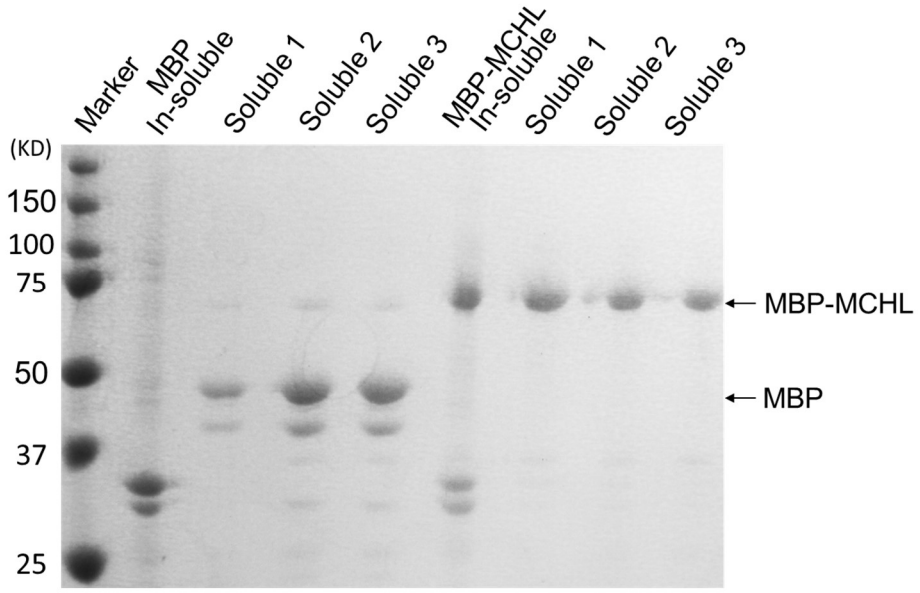


Figure S4. SDS-PAGE gel of purified PnMCHL.

Maltose-binding protein fused to N-terminal of MCHL showed molecular weight of 75kD. Soluble 1, 2, and 3 were eluted by 10% maltose containing extraction buffer.

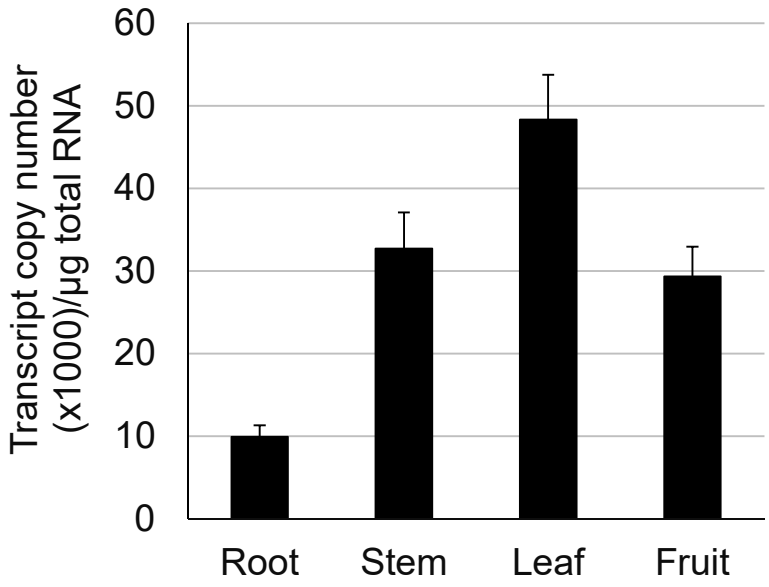


Figure S5. Transcript copy numbers of *PnMCHL* in four different organs.

Each data point was determined from four biological samples and four technical replications.

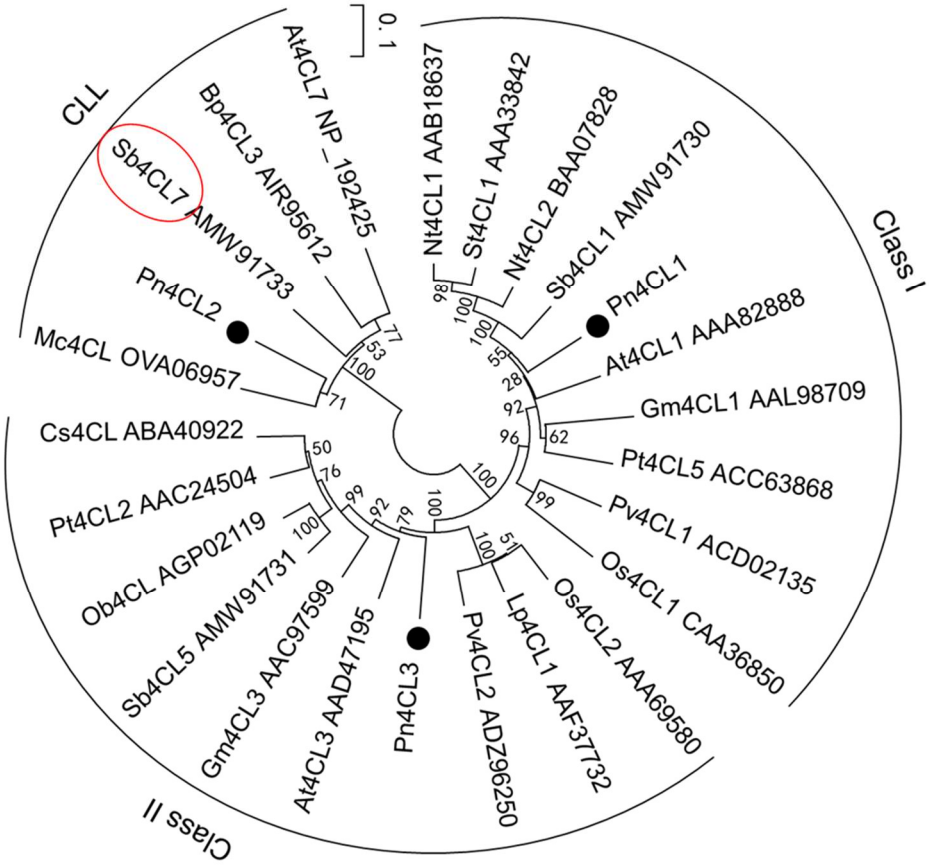


Figure S6. Phylogenetic analysis of three Pn4CLs with functional identified 4CL and CLL.

At *Arabidopsis thaliana*, Bp *Betula pendulaCLL*, Cs *Camellia sinensis*, Lp *Lolium perenne*, Nt *Nicotiana benthamiana*, Mc *Macleaya cordata*, Ob *Ocimum basilicum*, Os *Oryza sativa*, Pt *Populus tremuloides*, Pv *Panicum virgatum*, Gm *Glycine max*, Sb *Scutellaria baicalensis*, St *Solanum tuberosum*, CLL was indicated Co-enzymeA ligase-like. Black circle indicated three black pepper 4CL candidates.

References

Abe I, Takahashi Y, Morita H, Noguchi H (2001) Benzalacetone synthase A novel polyketide synthase that plays a crucial role in the biosynthesis of phenylbutanones in *Rheum palmatum*. FEBS J 268:3354-3359.

Adio AM (2009) Germacrenes A–E and related compounds: thermal, photochemical and acid induced transannular cyclizations. Tetrahedron 65:1533-1552.

Akiyama T, Shibuya M, Liu HM, Ebizuka Y (1999) *p*-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. *thunbergii*. FEBS J 263:834-839.

Austin MB, Noel JP (2003) The chalcone synthase superfamily of type III polyketide synthases. Nat Prod Rep 20:79-110.

Bülow N, König WA (2000) The role of germacrene D as a precursor in sesquiterpene biosynthesis: investigations of acid catalyzed, photochemically and thermally induced rearrangements. Phytochemistry 55:141-168.

Berry A, Dodge T, Pepsin M, Weyler W (2002) Application of metabolic engineering to improve both the production and use of biotech indigo. J Ind Microbiol Biot 28:127-133.

Bolwell GP, Bozak K, Zimmerlin A (1994) Plant cytochrome P450. Phytochemistry 37:1491-1506.

Cai Y, Jia J-W, Crock J, Lin Z-X, Chen X-Y, Croteau R (2002) A cDNA clone for β-caryophyllene synthase from *Artemisia annua*. Phytochemistry 61:523-529.

Chae Y-A, Yu H-S, Song J-S, Chun H-K, Park S-U (2000) Indigo production in hairy root cultures of *Polygonum tinctorium* Lour. Biotechnol Lett 22:1527-1530.

Chang S-j, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113-116.

Chang Y-J, Song S-H, Park S-H, Kim S-U (2000) Amorpha-4, 11-diene synthase of *Artemisia annua*: cDNA isolation and bacterial expression of a terpene synthase involved in artemisinin biosynthesis. Arch Biochem Biophys 383:178-184.

Chen G, Cue R-A, Lundstrom K, Wood JD, Doran O (2008) Regulation of CYP2A6 protein expression by skatole, indole, and testicular steroids in primary cultured pig hepatocytes. Drug Metab Dispos 36:56-60.

Chen H-C et al. (2013) Monolignol pathway 4-coumaric acid: coenzyme A ligases in *populus*. *trichocarpa*: Novel specificity, metabolic regulation, and simulation of coenzyme A ligation fluxes. Plant Physiol 161:1501-1516.

Costa MA et al. (2005) Characterization *in vitro* and *in vivo* of the putative multigene 4-coumarate: CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. Phytochemistry 66:2072-2091.

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol 9:331-340.

de Klerk E, 't Hoen PAC (2015) Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends Genet 31:128-139.

Dewick PM (2002) Medicinal natural products: a biosynthetic approach, (2en Edition.), pp.307-309, John Wiley & Sons, USA.

Dinesh-Kumar S, Baker BJ (2000) Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. P Natl Acad Sci USA 97:1908-1913.

Elaine R. Radwanski, Jianmin Zhao, Last RL (1995) *Arabidopsis thaliana* tryptophan synthase alpha gene cloning expression and subunit interaction. Mol Gen Genet 248:657-667.

Ensley B, Ratzkin B, Osslund T, Simon M, Wackett L, Gibson D (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. Science 222:167-169.

Epstein WW, Netz DF, Seidel JL (1993) Isolation of piperine from black pepper. J Chem Educ 70:598.
Facchini PJ, Bird DA, St-Pierre B (2004) Can *Arabidopsis* make complex alkaloids? Trends Plant Sci 9:116-122.
Faraldos JA, Miller DJ, González Vn, Yoosuf-Aly Z, Cascón O, Li A, Allemann RK (2012) A 1,6-ring closure mechanism for (+)-δ-cadinene synthase? J Am Chem Soc 134:5900-5908.
Fitzgerald S (2008) Ramses II, Egyptian Pharaoh, Warrior, and Builder, pp. 88-90. Capstone Point Books, ,
Frey M et al. (1997) Analysis of a chemical plant defense mechanism in grasses. Science 277:696-699.
Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A (2009) Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. Phytochemistry 70:1645-1651.
Frey M, Stettner C, Pare PW, Schmelz EA, Tumlinson JH, Gierl A (2000) An herbivore elicitor activates the gene for indole emission in maize. P Natl Acad Sci USA 97:14801-14806.
Friedman M, Levin CE, Lee S-U, Lee J-S, Ohnisi-Kameyama M, Kozukue N (2008) Analysis by HPLC and LC/MS of pungent piperamides in commercial black, white, green, and red whole and ground peppercorns. J Agr Food Chem 56:3028-3036.
Gallage NJ et al. (2014) Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. Nat Commun 5.
Gasson MJ et al. (1998) Metabolism of ferulic acid to vanillin a bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of a hydroxycinnamic acid SCoA thioester. J Biol Chem 273:4163-4170.
Geisler JG, Gross GG (1990) The biosynthesis of piperine in *Piper nigrum*. Phytochemistry 29:489-492.
Gierl A, Frey M (2001) Evolution of benzoxazinone biosynthesis and indole production in maize. Planta 213:493-498.
Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2:31-34.
Gillam EM, Notley LM, Cai H, De Voss JJ, Guengerich FP (2000) Oxidation of indole by cytochrome P450 enzymes. Biochemistry 39:13817-13824.
Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333-351.
Gordo SM et al. (2012) High-throughput sequencing of black pepper root transcriptome. BMC Plant Biol 12:168.
Grabherr MG et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644-652.
Gupta A, Gupta M, Gupta S (2013) Isolation of piperine and few sesquiterpenes from the cold petroleum ether extract of *Piper nigrum* (black pepper) and its antibacterial activity. Inter J Pharmac Phytochem 5:101-105.
He X, Cane DE (2004) Mechanism and stereochemistry of the germacradienol/germacrene D synthase of *Streptomyces coelicolor* A3(2). J Am Chem Soc 126:2678-2679.
Heil M, Hilpert A, Kaiser W, Linsenmair KE (2000) Reduced growth and seed set following chemical induction of pathogen defence: does systemic acquired resistance (SAR) incur allocation costs? J Ecol 88:645-654.
Hoessel R et al. (1999) Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases. Nat Cell Biol 1:60-67.
Hu L, Hao C, Fan R, Wu B, Tan L, Wu H (2015) *De novo* assembly and characterization of fruit transcriptome in black pepper (*Piper nigrum*). PloS One 10:e0129822.
Hukkanen AT, Kokko HI, Buchala AJ, McDougall GJ, Stewart D, Kärenlampi SO, Karjalainen RO (2007) Benzothiadiazole induces the accumulation of phenolics and

improves resistance to powdery mildew in strawberries. J Agr Food Chem 55:1862-1870.
Iriti M, Rossoni M, Borgo M, Faoro F (2004) Benzothiadiazole enhances resveratrol and anthocyanin biosynthesis in grapevine, meanwhile improving resistance to *Botrytis cinerea*. J Agr Food Chem 52:4406-4413.
Jeleń HH, Gracka A (2015) Analysis of black pepper volatiles by solid phase microextraction–gas chromatography: A comparison of terpenes profiles with hydrodistillation. J Chromatogr A 1418:200-209.
Jez JM, Ferrer J-L, Bowman ME, Dixon RA, Noel JP (2000) Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. Biochemistry 39:890-902.
Jin Z (2013) Indigoid biosynthesis in *Polygonum tinctorium*. MS Thesis, Seoul National University, Korea.
Jin Z, Kim J-H, Kim K-S, Park SU, Kim S-U (2012) Increased indigoid accumulation by plant defense activators in *Polygonum tinctorium* Lour. J Korean Soc Appl Biol Chem 55:359-362.
Joy N, Asha S, Mallika V, Soniya EV (2013) De novo transcriptome sequencing reveals a considerable bias in the incidence of simple sequence repeats towards the downstream of 'pre-miRNAs' of black pepper. PLoS One 8:e56694.
Katsuyama Y, Kita T, Funa N, Horinouchi S (2009) Curcuminoid biosynthesis by two type III polyketide synthases in the herb *Curcuma longa*. J Biol Chem 284:11160-11170.
Kende H, Zeevaart J (1997) The five" classical" plant hormones. Plant Cell 9:1197-1201.
Kim S-U, Song K-S, Jung D-S, Chae Y-A, Lee HJ (2007) Production of indoxyl derivatives in indole-supplemented tissue cultures of *Polygonum tinctorium*. Planta Med 62:54-56.
Kollmannsberger H, Nitz S, Drawert F (1992) über die Aromastoffzusammensetzung von Hochdruckextrakten. Zeitschrift für Lebensmittel-Untersuchung Und-Forschung 194:545-551.
Kozukue N, Park M-S, Choi S-H, Lee S-U, Ohnishi-Kameyama M, Levin CE, Friedman M (2007) Kinetics of light-induced cis- trans isomerization of four piperines and their levels in ground black peppers as determined by HPLC and LC/MS. J Agr Food Chem 55:7131-7139.
Kramer VC, Koziel MG (1995) Structure of a maize tryptophan synthase alpha subunit gene with pith enhanced expression. Plant Mol Biol 27:1183-1188.
Kriechbaumer V, Weigang L, Fießelmann A, Letzel T, Frey M, Gierl A, Glawischnig E (2008) Characterisation of the tryptophan synthase alpha subunit in maize. BMC Plant Biol 8:1.
Kugel S et al. (2017) Cryptic indole hydroxylation by a non-canonical terpenoid cyclase parallels bacterial xenobiotic detoxification. Nat Commun 8:ncomms15804.
Kulik V et al. (2005) On the structural basis of the catalytic mechanism and the regulation of the alpha subunit of tryptophan synthase from *Salmonella typhimurium* and BX1 from maize, two evolutionarily related enzymes. J Mol Biol 352:608-620.
Lücker J, Bowen P, Bohlmann J (2004) *Vitis vinifera* terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (−)-germacrene D synthase and expression of mono-and sesquiterpene synthases in grapevine flowers and berries. Phytochemistry 65:2649-2659.
Lambert JD, Hong J, Kim DH, Mishin VM, Yang CS (2004) Piperine enhances the bioavailability of the tea polyphenol (−)-epigallocatechin-3-gallate in mice. J Nutr 134:1948-1952.
Lee J, Zhang X-S, Hegde M, Bentley WE, Jayaraman A, Wood TK (2008) Indole cell signaling occurs primarily at low temperatures in *Escherichia coli*. ISME J 2:1007-1023.
Leopoldseder S, Hettwer S, Sterner R (2006) Evolution of multi-enzyme complexes: the case of tryptophan synthase. Biochemistry 45:14111-14119.

Li J-y, Chen S, Zhu L, Last RL (1995) Isolation of cDNAs encoding the tryptophan pathway enzyme indole-3-glycerol phosphate synthase from *Arabidopsis thaliana*. Plant Physiol 108:877-878.
Lim TK (2012) Edible medicinal andnon-medicinal plants. Vol. 4, Springer, New York, NY, USA.pp. 322-350.
Lin F, Ferguson KL, Boyer DR, Lin XN, Marsh ENG (2015) Isofunctional enzymes PAD1 and UbiX catalyze formation of a novel cofactor required by ferulic acid decarboxylase and 4-hydroxy-3-polyprenylbenzoic acid decarboxylase. ACS Chem Biol 10:1137-1144.
Liu L, Song G, Hu Y (2007) GC–MS analysis of the essential oils of *Piper nigrum* L. and *Piper longum* L. Chromatographia 66:785-790.
Maier W, Schumann B, Groger D (1990) Biosynthesis of indoxyl derivatives in *Isatis tinctoria* and *Polygonum tinctorium*. Phytochemistry 29:817-819.
Manczak T, Simonsen HT (2016) Insight into biochemical characterization of plant sesquiterpene synthases. Anal Chem Insights 11:1-6.
Maugard T, Enaud E, Sayette AdL, Choisy P, Legoy MD (2002) β-Glucosidase-catalyzed hydrolysis of indican from leaves of *Polygonum tinctorium*. Biotechnol Progr 18:1104-1108.
Minami Y, Nishimura O, Hara-Nishimura I, Nishimura M, Matsubara H (2000) Tissue and intracellular localization of indican and the purification and characterization of indican synthase from indigo plants. Plant Cell Physiol 41:218-225.
Mitra A et al. (1999) 4-Hydroxycinnamoyl-CoA hydratase/lyase (HCHL)—an enzyme of phenylpropanoid chain cleavage from *Pseudomonas*. Arch Biochem Biophys 365:10-16.
Mukai N, Masaki K, Fujii T, Kawamukai M, Iefuji H (2010) PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. J Biosci Bioeng 109:564-569.
Multani PS (2015) Production of indigoids by *Escherichia coli* expressing tryptophan synthase-like and p450 genes from *Isatis tinctoria* L. MS thesis, Southern Illinois University Carbondale, USA.
Nakano C, Horinouchi S, Ohnishi Y (2011) Characterization of a novel sesquiterpene cyclase involved in (+)-caryolan-1-ol biosynthesis in *Streptomyces griseus*. J Biol Chem 286:27980-27987.
Narbad A, Gasson MJ (1998) Metabolism of ferulic acid via vanillin using a novel CoA-dependent pathway in a newly-isolated strain of *Pseudomonas fluorescens*. Microbiology 144:1397-1405.
Nedspice (2017) Pepper crop report 2017. http://www.nedspice.com, (2017), accessed Aug. 16, 2017.
Nguyen DT et al. (2010) Biochemical conservation and evolution of germacrene A oxidase in Asteraceae. J Biol Chem 285:16588-16598.
Nguyen T-D, MacNevin G, Ro D-K (2012) *De novo* synthesis of high-value plant sesquiterpenoids in yeast. Method Enzymol 517:261-278.
Nomura T, Ishihara A, Imaishi H, Ohkawa H, Endo TR, Iwamura H (2003) Rearrangement of the genes for the biosynthesis of benzoxazinones in the evolution of Triticeae species. Planta 217:776-782.
Nonhebel HM (2015) Tryptophan-independent indole-3-acetic acid synthesis: critical evaluation of the evidence. Plant Physiol 169:1001-1005.
Peng M et al. (2016) Evolutionarily distinct BAHD N-acyltransferases are responsible for natural variation of aromatic amine conjugates in rice. Plant Cell Online:tpc. 00265.
Pickel B, Drew DP, Manczak T, Weitzel C, Simonsen HT, Ro D-K (2012) Identification and characterization of a kunzeaol synthase from *Thapsia garganica*: implications for the biosynthesis of the pharmaceutical thapsigargin. Biochem J 448:261-271.
Platel K, Srinivasan K (2000) Influence of dietary spices and their active principles on pancreatic digestive enzymes in albino rats. Food/Nahrung 44:42-46.

Pliny the Elder (1979) Natural History XII. transl. Eichholz D, Jones WHS, and Rackham H (1949-1954), pp.310-313. Harvard Iniversity Press, USA.
Prosser I, Altug IG, Phillips AL, König WA, Bouwmeester HJ, Beale MH (2004) Enantiospecific (+)-and (−)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis aspartate-rich motif. Arch Biochem Biophys 432:136-144.
Rastogi S, Kumar R, Chanotiya CS, Shanker K, Gupta MM, Nagegowda DA, Shasany AK (2013) 4-Coumarate: CoA ligase partitions metabolites for eugenol biosynthesis. Plant Cell Physiol 54:1238-1252.
Rauscher FM, Sanders RA, Watkins JB (2000) Effects of piperine on antioxidant pathways in tissues from normal and streptozotocin-induced diabetic rats. J Biochem Mol Toxic 14:329-334.
Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol 58:267-294.
Ro D-K et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440:940-943.
Rowlett R, Yang L-H, Ahmed SA, McPhie P, Jhee K-H, Miles EW (1998) Mutations in the contact region between the α and β subunits of tryptophan synthase alter subunit interaction and intersubunit communication. Biochemistry 37:2961-2968.
Sainsbury F, Thuenemann EC, Lomonossoff GP (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol J 7:682-693.
Sakamoto M, Kim D-A, Imoto K, Kitai Y, Suzuki T (2017) Comparison of two harvesting methods for the continuous production of indigo plant (*Polygonum Tinctorium*) leaves in a closed-cultivation system. Sust Agr Re 6:7.
Schanz S, Schröder G, Schröder J (1992) Stilbene synthase from Scots pine (*Pinus sylvestris*). FEBS Lett 313:71-74.
Schmidt H (1997) Indigo-100 Jahre industrielle synthese. Chem Unserer Zeit 31:121-128.
Schroeder A et al. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 7:1-14.
Schrott W (2001) Denim wieder im Blickpunkt der Textilindustrie. Melliand Textilber 82, 331.
Semler U, Gross GG (1988) Distribution of piperine in vegetative parts of *Piper nigrum*. Phytochemistry 27:1566-1567.
Siebert TE, Wood C, Elsey GM, Pollnitz AP (2008) Determination of rotundone, the pepper aroma impact compound, in grapes and wine. J Agr Food Chem 56:3745-3748.
Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1:2019-2025.
Srinivasan K (2007) Black pepper and its pungent principle-piperine: a review of diverse physiological effects. Crit Rev Food Sci 47:735-748.
Stadnik M, Buchenauer H (2000) Inhibition of phenylalanine ammonia-lyase suppresses the resistance induced by benzothiadiazole in wheat to *Blumeria graminis* f. sp. *tritici*. Physiol Mol Plant P 57:25-34.
Stewart C, Vickery CR, Burkart MD, Noel JP (2013) Confluence of structural and chemical biology: plant polyketide synthases as biocatalysts for a bio-based future. Curr Opin Plant Biol 16:365-372.
Swift S, Stewart GS (1991) The molecular biology of tryptophan synthase: A model for protein-protein interaction. Biotechnol Genet Eng Rev 9:229-294.
Thwe AA, Kim YB, Kim SU, Park SU (2012) *In vitro* shoot regeneration from stem internodes of *Polygonum tinctorium*. Life Sci J 9.
Tozawa Y, Hasegawa H, Terakawa T, Wakasa K (2001) Characterization of rice anthranilate synthase α-subunit genes *OASA1* and *OASA2*. Tryptophan accumulation in

transgenic rice expressing a feedback-insensitive mutant of OASA1. Plant Physiol 126:1493-1506.
Tuominen LK, Johnson VE, Tsai C-J (2011) Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. BMC Genomics 12.1:236.
Turner J (2008) Spice: The history of a temptation, pp. 160-165. Knopf Doubleday Publishing Group.
van Herpen TW, Cankar K, Nogueira M, Bosch D, Bouwmeester HJ, Beekwilder J (2010) *Nicotiana benthamiana* as a production platform for artemisinin precursors. PLoS One 5:e14222.
Vardakou M, Salmon M, Faraldos JA, O'Maille PE (2014) Comparative analysis and validation of the malachite green assay for the high throughput biochemical characterization of terpene synthases. MethodsX 1:187-196.
Vega NM, Allison KR, Khalil AS, Collins JJ (2012) Signaling-mediated bacterial persister formation. Nat Chem Biol 8:431-433.
Wang B et al. (2015) Tryptophan-independent auxin biosynthesis contributes to early embryogenesis in *Arabidopsis*. P Natl Acad Sci USA 112:4821-4826.
Wang G et al. (2008) Terpene biosynthesis in glandular trichomes of hop. Plant Physiol 148:1254-1266.
Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G (2009) Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. P Natl Acad Sci USA 106:3698-3703.
Xia Z-Q, Zenk MH (1992) Biosynthesis of indigo precursors in higher plants. Phytochemistry 31:2695-2697.
Yin JL et al. (2001) Real-time reverse transcriptase-polymerase chain reaction (RT–PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green I. Immunol Cell Biol 79:213-221.
Yin R, Frey M, Gierl A, Glawischnig E (2010) Plants contain two distinct classes of functional tryptophan synthase beta proteins. Phytochemistry 71:1667-1672.
Zhang R, Wang B, Ouyang J, Li J, Wang Y (2008) Arabidopsis indole synthase, a homolog of tryptophan synthase alpha, is an enzyme involved in the Trp-independent indole-containing metabolite biosynthesis. J Integr Plant Biol 50:1070-1077.
Zhao Q et al. (2016) A specialized flavone biosynthetic pathway has evolved in the medicinal plant, *Scutellaria baicalensis*. Sci Adv 2:e1501780.
Zhou H, Yang Y-L, Zeng J, Zhang L, Ding Z-H, Zeng Y (2016) Identification and characterization of a δ-cadinol synthase potentially involved in the formation of boreovibrins in *Boreostereum vibrans* of basidiomycota. Nat Prod Bioprospect 6:167-171.
Zhuang X, Fiesselmann A, Zhao N, Chen H, Frey M, Chen F (2012) Biosynthesis and emission of insect herbivory-induced volatile indole in rice. Phytochemistry 73:15-22.

Abstract in Korean

서울대학교 대학원

농생명공학부 응용생명화학전공

Jin Zhehao

본 박사학위논문에서는 쪽에서 인돌 그리고 후추에서 파이페린 및 세스퀴테펜 생합성에 관여하는 효소 발굴 및 그 효소의 기능 연구에 대한 내용이다.

첫번째 장에서는 쪽에서 인돌 생합성효소를 발굴, 규명하였다. 인디고는 아주 오래된 천연 청색 연색원료로 쪽식물로부터 생성된다. 식물에서 인디고 생합성 주요 과정은 인돌-3-글라이세롤 인산 분해효소(IGL)에 의하여 인돌이 생성된다. 트립토판 생합성효소-알파 소단위체와 유사성이 높은 PtIGL-short 과 long 이 각각 다른 IGL 단백질을 만든다. 단백질 서열분석결과 short 과 long 은 각각 273 과 316 개의 아미노산으로 구성되었다. IGL-short 는 트립토판이 없는 배지에 도말한 대장균 *ΔtnaA ΔtrpA* 돌연변이체에서 자유인돌 (free indole)을 생합성하여 인돌 생합성 유전자 (*PtINS*)라고 명명하였다. 그러나 IGL-long 은 수송펩타이를 포함 또는 제거하여도 돌연변이체 대장균이 생장 할수 없는 것으로부터 *PnTSA* 라고하였다. PtTSA 는 엽록체로 수송되는 42 개의 아미노산을 가지고 PtINS 는 세포질에 존재한다. 두 유전자의 5' UTR 분석으로 *TSA* 가 진화과정에 스프라이싱 위치가 생기고 pre-mRNA 에서 단백질수송에 관여하는 서열이 제거되면서 *PnINS* 로 진화되었다는 것을 제안하였다. 식물 강화제 (BTH) 처리와 무처리에서 *PnINS* 는 *PnTSA* 보다 2 에서 5 배 높은 유전자 전사 수준을 보였고 유의성 있는 차이를 보였다.

두번째와 세번째 장에서는 후추의 향 그리고 매운 맛을 내는 세스키터펜과 파이페린 생합성에 관한내용이다. 미성숙 후추 과실을 Illumina 의 차세대 시퀀싱 (NGS)을 이용하여 전사체 분석을 하였다. RACE-PCR 유전자 클로닝기술과 비교했을 때 NGS 기술은 시간과

비용을 절약하고 특정 유전자의 대량의 데이터를 얻을 수 있다. 전사체 데이터로부터 이미 효소의 기능이 밝혀진 단백질을 query 로 로컬 tBLASTn 방법을 사용하여 3 개의 Sesquiterpene synthase (PnTPS1 ~ 3) 및 파이페린 생합성 과정에 관연될것으로 예상되는 4 가지 효소(PnMCHL; PnPKS1,2; Pn4CL3 및 PnNAT6,7)의 유전자를 얻었다. 이 4 종의 유전자를 대장균에서 해당되는 단백질 발현하고 *in-vitro* 반응, 그리고 효모 또는 대장균 *in-vivo* 에서 물질을 생산 후 추출하여 분석하였다.

두번째 파트에서 sesqui-TPS 에서 PnTPS1 은 caryophyllene 을 주 생성물로하고 Humulene 을 부산물로 생성하여 caryophyllene synthase (PnCPS)명명하였다. 위와 동일하게 PnTPS2 와 PnTPS3 는 cadinene/cadinol synthase (PnCO/CDS)와 germacrene D synthase (PnGDS)로 명명하였다. 효모 시스템에서 PnGDS 발현은 germacrene D 의 재조합 산물 인 β-cadinene 과 α-copaene 을 생성하였다.

세번째 파트에서 transcriptome 데이터베이스를 기반으로 파이페린 생합성에 관여한다고 예상되는 유전자를 찾았다. 기질 특이성을 가진 hydratase-lyase type (PnMCHL) 효소는 3,4-methylenedioxy cinnamic acid (MDCA)를 piperonal 로 전환하는 과정이 실험으로 증명되었다. 기존의 파이페릭산 (piperic acid, C_6C_5) 생합성 가설은 화학구조로부터 예측한 MDCA (C_6C_3)에 $1\times C_2$ 연장되어 C_6C_5 를 형성한다는 것이었다. 우리는 이번 연구로부터 기존과 다른 파이페로닐산 (piperonylic acid, C_6C_1)이 $2\times C_2$ 이 연장되어 파이페릭산 탄소골격을 형성한다는 새로운 가설을 제기하게 되었다. 또한 파이페릭산을 piperoyl-CoA 로 전환하는 3,4-methylenedioxyphenyl 기질특이성 4-coumaroyl-Coenzyme A ligase (Pn4CL3)도 기술하였다. 마지막으로, piperoyl-CoA 를 파이페리딘으로 전달하기위한 효소를 암호화하는 2 개의 NAT cDNA 도 실험으로 확인되었다.

주요어 : 생합성, 세스퀴테펜, 인돌 생합성효소, 쪽, 파이페린, 후추

학 번 : 2013-31316

Curriculum Vitae

Name	: Zhehao Jin (金哲浩 김철호)
Sex	: Male
Data of Birth	: 29 Aug, 1987
Place of Birth	: Ji Lin, China
Present Address	: Graduate School of Agriculture Biotechnology Seoul National University E-mail; jinzh21@snu.ac.kr, jinzh-21@hotmail.com
Education	: Seoul National University Seoul, Korea Ph. D. in Agricultural Biotechnology (Sep, 2013 -) M. S. in Agricultural Biotechnology (Sep, 2011 - Aug, 2013) : Ji Lin Agricultural & Technology University Ji Lin, China B. S. in Bioengineering (Aug, 2007 - Jun, 2011)

Publications and Patents

Research article

3. Jin Z, Kwon MH, Lee AR, Ro DK, Wungsintaweeukl J, Kim SU (2018). Molecular cloning and functional characterization of three terpene synthases from unripe fruit of black pepper (*Piper nigrum*). Arch. Biochem. Bioph., 638(15), 35-40.

2. Jin Z, Kim JH, Park SU, & Kim SU (2016). Cloning and characterization of indole synthase (INS) and a putative tryptophan synthase α-subunit (TSA) genes from *Polygonum tinctorium*. Plant Cell Rep., 35(12), 2449-2459.

1. Jin Z, Kim JH, Kim KS, Park SU, & Kim SU (2012). Increased indigoid accumulation by plant defense activators in *Polygonum tinctorium* Lour. Appl. Biol. Chem., 55(3), 359-362.

Patent

1. 김수언, 진 제하오, 김진희, 홍성호, 권문혁, 김관수; 식물 내 인디고이드 함량 증대용 조성물 및 이를 이용한 인디고이드 함량 증가방법; 출원번호, 10-2013-0036361; 등록번호, 10-1495898-0000; 등록일자, 2015.02.16