Doctoral Dissertation

Comparative analysis of legume NADPH-cytochrome P450 reductase classes towards improving heterologous production of triterpenoids in transgenic yeast

Pramesti Istiandari

May 2022

Division of Advanced Science and Biotechnology

Graduate School of Engineering,

Osaka University

Thesis Committee

Professor Toshiya Muranaka, Ph. D

Laboratory of Cell Technology, Department of Biotechnology,

Graduate School of Engineering, Osaka University

Professor Kazuhito Fujiyama, Ph.D

Laboratory of Applied Microbiology, International Center for Biotechnology,

Graduate School of Engineering, Osaka University

Professor Kohsuke Honda, Ph.D

Laboratory of Molecular Microbiology, International Center for Biotechnology,

Graduate School of Engineering, Osaka University

List of abbreviations

aAS : α-Amyrin synthase

ATP : Adenosine triphosphate

ATR : Arabidopsis CPR

bAS : β-Amyrin synthase

CAS : Cycloartenol synthase

Cas9 : CRISPR-associated protein 9

Cb5 : Cytochrome b_5

CPR : NADPH-cytochrome P450 reductase

CRISPR : Clustered regularly interspaced short palindromic repeats-Cas9

CslM : Cellulose synthase-like family M

CSyGT : Cellulose-synthase superfamily-derived glycosyltransferase

CYP : Cytochrome P450 monooxygenases

DMAPP : Dimethylallyl pyrophosphate

DNA : Deoxyribonucleic acid

ER : Endoplasmic reticulum

ET : Electron transfer

FAD : Flavin adenine dinucleotide

FMN : Flavin mono-nucleotide

FPP : Farnesyl diphosphate

GC-MS : Gas chromatography-mass spectrometry

gRNA : Guide-RNA

GT : Glycosyltransferase

HMA : Heteroduplex mobility assay

IPP : Isopentenyl diphosphate

LAS : Lanosterol synthase

LORE1 : Lotus retrotransposon 1

LUS : Lupeol synthase

MeJA : Methyl jasmonate

MEP : Methylerythritol phosphate pathway

MVA : Mevalonate

NADPH : Nicotinamide adenine dinucleotide phosphate

NCBI : National Center for Biotechnology Information

NIST : National Institute of Standards and Technology

OSC : Oxidosqualene cyclases

PCC : Pearson correlation coefficient

PCR : Polymerase chain reaction

RNA : Ribonucleic acid

RT-PCR : Real-time PCR

SQE : Squalene epoxidase

SQS : Squalene synthase

UDP : Uridine diphosphate

UGT : UDP-glucosyltransferase

WT : Wild-type

Table of Contents

List of Figures

List of Tables

List of Supplementary Tables

List of Supplementary Figures

Chapter 1

General Introduction

1.1 Plant triterpenoids

Despite being a sessile organism, plants possess a wide array and complex biosynthetic pathways that manufacture more than one million compounds of plant specialized metabolisms (Afendi et al. 2012). Plant specialized metabolism is also known as plant secondary metabolism which is not essential for the primary development of plants. Among different groups of plant specialized metabolites, triterpenoids comprise the most diverse groups with more than 20,000 compounds known to exist in nature (Oldfield and Lin 2012). Generated from 30 carbon atoms that are polymerized into six isoprene units, triterpenoids can be grouped into linear, monocyclic, dicyclic, up to pentacyclic structures Triterpenoids are commonly found in plants as triterpenoids saponins, which is are glycosides consisting of a sugar moiety (glycone) and a triterpenoid compound (aglycone). Triterpenoid saponins are more rarely found in monocotyledons and more abundantly found in many dicotyledonous families such as Araliaceae, Ranunculaceae, Leguminosae, Caryophyllaceae, Cucurbitaceae, and Umbelliferae (Du et al. 2014). Some plant families can produce unique structure of triterpenoids which can further be classified into diverse groups such as cucurbitanes, cycloartanes, dammaranes, euphanes, friedelanes, holostanes, hopanes, isomalabaricanes, lanostanes, limonoids, lupanes, oleananes, protostanes, sqalenes, tirucallanes, ursanes and miscellaneous compounds (Bishayee et al. 2011).

1.1.1 The importance of plant triterpenoids

Triterpenoids saponins are not only found in our daily diet such as beans, soybeans, spinach, lentils, and oats, but also in phytomedicines such as *Radix et Rhizoma Ginseng*, *Radix et rhizoma glycyrrhizae*, and *Radix astragli* (Du et al. 2014). Compared to other terpenoids, such as monoterpenes, sesquiterpenes, and diterpenes, triterpenoids recently emerged as a promising phytochemical with various pharmacological activities with promising results in preclinical studies (Bishayee et al. 2011). These triterpenoids are traditionally used as medicine to treat number of diseases such as cancer, diabetes, cardiac and cerebral vascular diseases, inflammation, and viral and bacterial infections. Natural triterpenoids provide an attractive alternative anticancer agent due to its favorable efficacy and safety profiles, compared to conventional anticancer agents that is known to be highly toxic not only to tumor cells, but also to normal cells. Among different types of triterpenes, studies on bioactivity of pentacyclic terpenes are increasing in recent years and thus attracting the most attention. As evidence, several pentacyclic terpenes are currently being marketed as therapeutical agents and dietary supplement (Sheng and Sun 2011).

Table 1-1. The current profile of natural pentacyclic triterpenes as therapeutic agents and dietary supplements (modified from Sheng and Sun 2011).

Name of active compound	Category	Indications	Current status	Region
Oleanolic acid	Drug	Liver diseases	Registered (OTC drug)	China
Glycyrrhizin	Drug	Liver diseases	Registered	China, Japan
Carbenoxolone	Drug	Gastritis and peptic ulcers	Registered	Asia, Europe
Asiaticoside	Drug	Wound-healing	Registered (OTC drug)	China, Europe
Corosolic acid	Dietary supplement	Diabetes, obesity	Marketed	Asia, America

Gymnemic acid	Dietary supplement	Diabetes	Marketed	Asia, America, Europe

Although highly diverse structures of triterpenoid compounds are found in plants, they are produced in very small amount in natural sources due to their strict biosynthetic regulation. Plant extraction is considered not environmentally friendly and the cost is also expensive because of the low yield. Therefore, despite their superior potential as therapeutic agents, the supply of natural triterpenes is hardly meet the demand and create drawback for their widespread application. Many approaches have been conducted to improve yield of natural compounds, which can be achieved through various applications of plant biotechnology or synthetic biology. Biotechnology application *in planta* can be performed by propagation of high-producing cultivars and the production and/or elicitation of (transgenic) plant (cell) cultures. On the other hand, synthetic biology includes metabolic engineering of the biosynthetic pathway, particularly in microbes (Moses et al. 2013). Either way, deeper understanding of triterpenoids biosynthesis is very crucial to enhance productivity of natural triterpenoids.

1.1.2 Triterpenoids biosynthesis in plants

Triterpenoids and phytosterols are derived from common precursor 2,3-oxidosqualene through two distinct pathways: the plastidial 2-C-methyl-d-erythritol 4-phosphate (MEP) and cytosolic mevalonate (MVA) pathways (Vranová et al. 2012). Oxidosqualene cyclases (OSC) catalyze the first step in triterpenoid biosynthesis by performing cyclization of 2,3-oxidosqualene to various triterpenol scaffolds. This step involves complex reactions that lead to the formation of polycyclic molecules as triterpene backbones. Depends on the type of OSC, 2,3-oxidosqualene can

be either converted towards triterpenoids or sterol biosynthesis (Figure 1-1) (Fukushima et al. 2011)

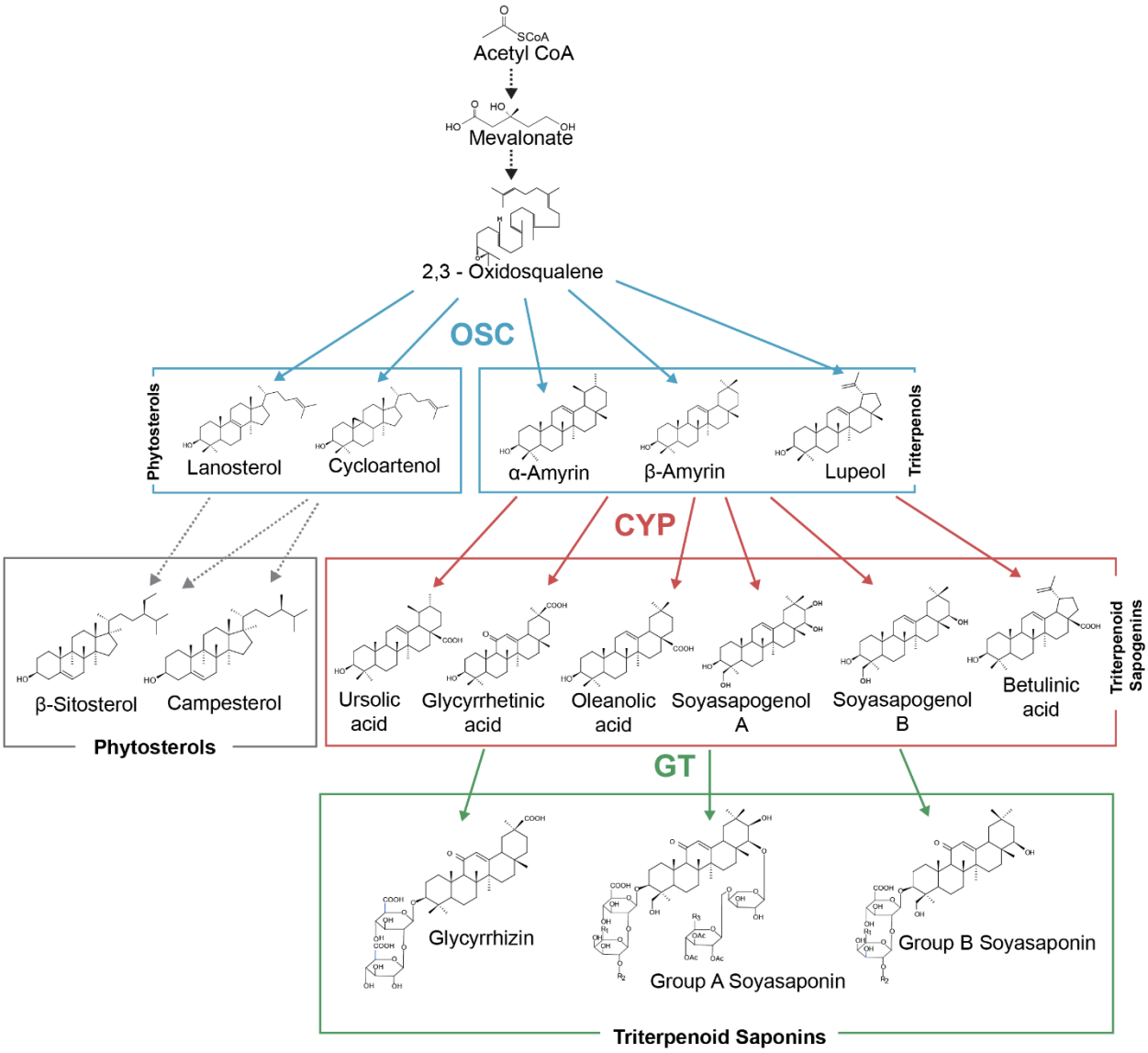

Figure 1-1. Overview of triterpenoids and phytosterol biosynthetic pathway in plants.

As the first step in triterpenoids biosynthesis, oxidosqualene cyclases (OSC) convert 2,3-oxidosqualene into various triterpenols as triterpene backbones. In the second step, cytochrome P450 monooxygenases (CYP) catalyze site-specific oxidation on triterpenols to further product highly diverse triterpenoids sapogenins. Lastly, glycosyltransferases (GT) then decorate these sapogenins with sugar moieties to produce triterpenoid saponins.

The next step in triterpenoids biosynthesis is the site-specific oxidations catalyzed by CYPs, which is the most critical step for generating structural diversity of triterpenoids (Xu et al. 2004). Cytochromes P450 (CYPs) are the largest family of enzymes containing heme as a cofactor that functions as monooxygenases and involved in various plant metabolism, representing 1% of protein coding genes of the plant genome. CYPs are grouped into different families based on their sequence homology and phylogenetic criteria. In plants, not only in primary metabolism such as sterols, CYPs also involve in almost all plant specialized metabolic and detoxification pathways (Moktali et al. 2012). The triterpenols undergo a various of scaffold-, regio-, and stereo-specific oxidations catalyzed by CYPs, leading to triterpene scaffold decoration with various functional groups such as hydroxyl, carbonyl, carboxyl, and epoxy moieties (Figure 1-1). These triterpene scaffolds are also called triterpenoid sapogenins. The reactions catalyzed by the CYPs were found to be extremely diverse in nature, including oxidation, desaturation, and C–C bond cleavage (Ghosh 2017).

The final committed step in the triterpenoids biosynthesis is the addition of various sugar moieties (glycone) to the triterpenoids sapogenins (aglycone) by glycosyltransferases (GT) to generate triterpenoid saponins. There are different types of GT that play key role in diversification of triterpenoid saponins. The majority of saponins are diversified through glycosylation at hydroxyl and/or carboxyl groups by UGTs, which catalyze glycosylation using a UDP-sugar donor such as UDP-glucose, UDP-galactose, UDP-arabinose, UDP-rhamnose, UDP-xylose or UDP-glucuronic acid. Similar to the CYPs, UGTs also constitute large family enzymes in plants. As with P450s, the nomenclature of UGTs is based on the homology of their amino acid sequences (Seki et al. 2015). However, just recently it was revealed that another type of GT also involves in triterpenoids saponin biosynthesis. Cellulose synthase superfamily-derived glycosyltransferase

(CSyGT) catalyses 3-O-glucuronosylation of triterpenoid aglycones and has been reported to be involved in saponin biosynthesis of *Glycyrrhiza uralensis* and *Lotus japonicus* to produce glycyrrhizin and soyasaponin, respectively. The cellulose-synthase superfamily is the glycosyltransferase 2 superfamily and consists of cellulose-synthase and cellulose-synthase-like family, which is involved in biosynthesis of cellulose and hemicellulose, respectively. All CSyGTs were classified into the cellulose-synthase-like M subfamily (CslM) which specifically selects UDP-glucuronic acid as the sugar donor (Chung et al. 2020).

1.1.3 Triterpenoids in legume plants

Among dicotyledonous families, Fabaceae or leguminous plants are known to accumulate diverse triterpenoids saponin such as soyasaponin. Legumes are also a part of human daily life dietary food, such as beans and soybeans. Soyasaponin in soybean has been reported to promote several health functions as antioxidant, lowering cholesterol level, reducing blood glucose levels, and anti-inflammatory (Kamo et al. 2014). *G. uralensis* or licorice is a well-known traditional medicinal legume that produced glycyrrhizin, which is found only in the Glycyrrhiza species. Glycyrrhizin shows high pharmacological activities and is used as a natural sweetener because its sweetness is 150 times higher than that of sucrose (Seki et al. 2008). Due to its economic value and high demand, studies on glycyrrhizin biosynthesis have been attracting attention. Other legume such as *Centella asiatica* produces unique sapogenins, Asiatic acid, which has been registered as wound-healing drug and is now available on market (Sheng and Sun 2011). Model legumes *Medicago truncatula* and *Lotus japonicus* are also have been intensively studied, resulting in numerous characterized CYPs involved in triterpenoid saponin biosynthesis (Suzuki et al. 2019; Vo et al. 2017). These legumes have a comprehensive genomic and transcriptomic database, as well as mutant library available (Carelli et al. 2013; Urbański et al. 2012). Therefore, *Medicago*

truncatula and *Lotus japonicus* are very good plant source for deeper study on triterpenoids biosynthesis.

The major pentacyclic triterpene backbones are β-amyrin, α-amyrin, and lupeol (Figure 1-1) which generate oleanane, ursane, and lupane tritepene group, respectively. To date, about 55 P450s are identified that involve on plant pentacyclic triterpenes (Ghosh 2017). CYP716As and CYP72As family are widespread in plants and show diverse functions in triterpenoids biosynthesis. In the other hand, other CYP families such as CYP93Es and CYP88Ds are known to be more unique to certain plant family, such as legumes (Seki et al. 2015). Sapogenins in *M. truncatula* are differentiated into two groups based on the position of the functional groups and sugars. Sapogenins with hydroxyl group at the C-24 position such as soyasapogenols are called non-hemolytic sapogenins and is catalyzed by CYP93E2. On the other hand, the carboxylation of β-amyrin at the C-28 position by CYP716A12 leads to oleanolic acid production, the precursor of hemolytic saponins (Carelli et al. 2011; Fukushima et al. 2011). Oleanolic acid can be further oxidized by CYP72A67 and CYP72A68, which catalyze hydroxylation of the oleanane backbone at the C-2 position and a three-step carboxylation at the C-23 position, respectively, to produce another hemolytic sapogenin, medicagenic acid (Fukushima et al. 2011; Biazzi et al. 2015). In the nonhemolytic branch, 24-hydroxy β-amyrin is further oxidized by CYP72A61 that catalyzes hydroxylation at the C-22 position, producing the major soyasaponin aglycone soyasapogenol B (Fukushima et al. 2011). *L. japonicus* CYP716A51 and *G. uralensis* CYP716A149 are homologue to CYP716A12 which catalyze the formation of oleanolic acid, ursolic acid and betulinic acid (Fukushima et al. 2011; Tamura et al. 2017; Suzuki et al. 2019). *L. japonicus* CYP93E1, *G. uralensis* CYP93E3, and *Glycine max* CYP93E1 are homologue and have same function with CYP93E2 (Shibuya et al. 2006; Seki et al. 2008). Other CYPs such as, CYP88D6 and CYP72A154

catalyze C-11 and C-30 oxidation of β-amyrin, respectively, to produce glycyrrhizin (Seki et al. 2008; Seki et al. 2011). As the first CYP responsible for glycyrrhizin biosynthesis, CYP88D6 is unique only to *Glycyrrhiza sp.*, and not found in other legume species. However, *M. truncatula* CYP72A63 is homologous to and show similar function with stronger oxidation activity than CYP72A154. While C-30 oxidase is not found in *L. japonicus*, it has CYP72A61 which homologue to and show same function with *M. truncatula* CYP72A61.

1.2 NADPH-cytochrome P450 reductase

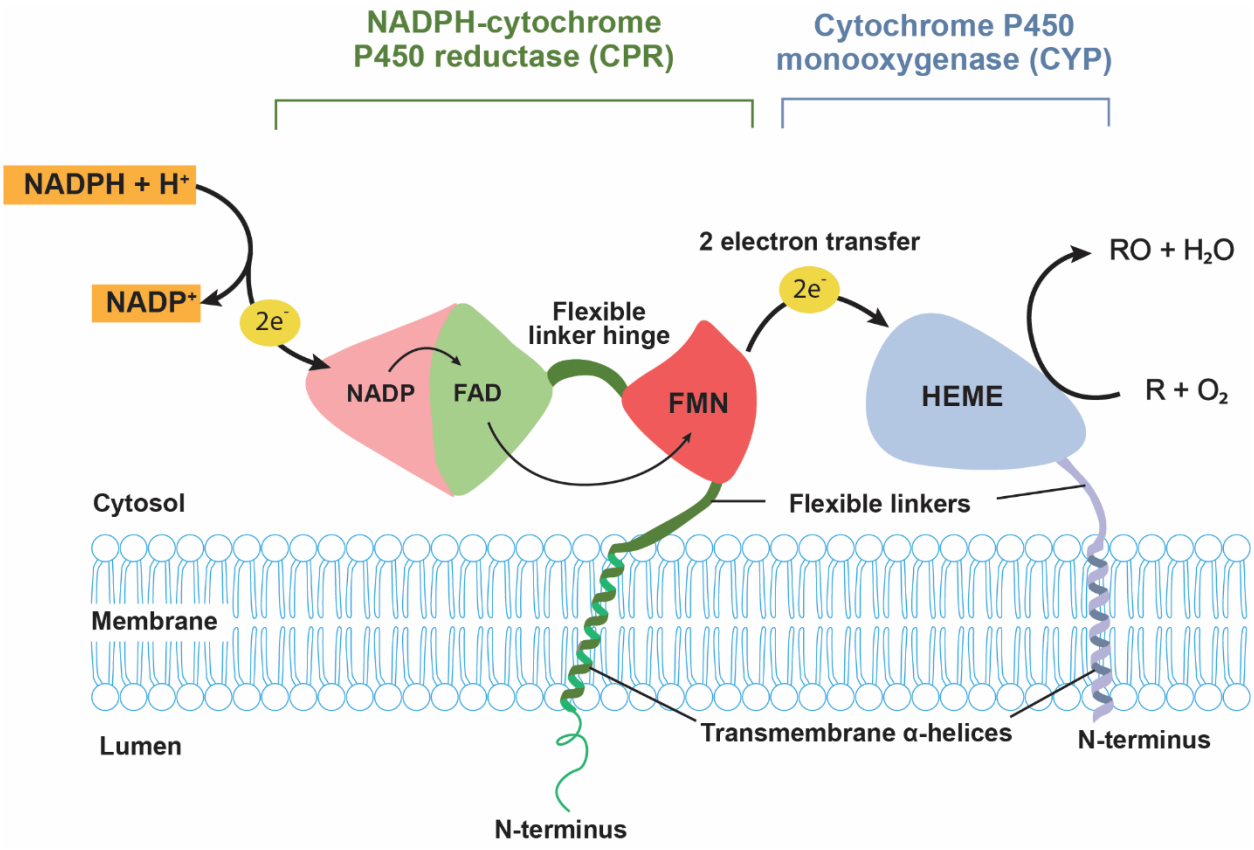

Figure 1-2. Simplified scheme of catalysis of cytochrome P450 (CYP) and NADPH-cytochrome P450 reductase (CPR) system (modified from Pandian et al. 2020; Mukherjee et al. 2020)

CYPs containing heme cofactor receive two electrons derived from nicotinamide adenine dinucleotide phosphate (NADPH) through CPR to catalyze the oxidation reaction $R + O_2 + NADPH \rightarrow RO + H_2O + NADP^+$; where R is the substrate and RO is the product of the oxidation reaction.

To perform site-specific oxidation, CYPs require two electrons transferred by its redox partner, NADPH-cytochrome P450 reductase (CPR). CPR contains three cofactor binding domains: NADP, FAD, and FMN domain (Mukherjee et al. 2021); Figure 1-2). NADP domain containing nicotinamide adenine dinucleotide phosphate (NADP) cofactor acts as binding site for NADPH. As NADPH bounds to NADP domain of CPR, FAD domain with the flavin adenine dinucleotide (FAD) cofactor reduces NADPH and accepts a pair of electrons as hydride ion, resulting in reduced $NADP^+$. Then, the electron is transferred to FMN domain with the flavin mononucleotide (FMN) cofactor via FAD-FMN flexible linker hinge (Figure 1-2), which is the rate-limiting ET (electron transfer) in plant CPR system (Haque et al. 2012). From FAD and FMN, the electron is then transferred one at a time to the heme cofactor of CYPs (Wang et al. 1997). The iron-containing heme cofactor of CYP is reduced from ferric to ferrous state by the first electron from CPR, leading to rapid binding of oxygen to form oxyferrous CYP. The oxyferrous CYP accepts the second electron from CPR and the reduced oxyferrous leads to water formation and enables CYP to oxidize various substrates (Hamdane et al. 2008).

CYPs and CPRs are both integral membrane proteins. In human, CPR is expressed in all tissues and the endoplasmic reticulum (ER) of the liver is found to have the highest expressions of CPR (Gan et al. 2009). As integral protein, CPR anchors to the membrane at its N-terminal membrane-binding domain. To connect this transmembrane α-helices to the FMN domain, there is a flexible tether region between those two domains. Not only CPR, CYP also has the transmembrane-binding domain which has been reported to be crucial for efficient ET from CPR to CYP (Mukherjee et al. 2020; Figure 1-2). The hydrophobic N-terminal membrane of CPR and CYP is responsible for its association and orientation in the ER. This transmembrane helix involves in the preliminary association with target CYPs, ensuring correct orientation of the

CPR:CYP binding motifs on the FMN domain to ensure effective electron transfers to the CYPs (Esteves et al. 2020). The truncated N-terminal membrane domain of rat CPR allows CPR to reduce soluble cytochrome *c*, but incapable of donating electrons to the membrane bound CYPs (Laursen et al. 2011).

During the electron transfer, CPR must undergo conformational changes between an open and closed form for interflavin electron transfer (Figure 1-3). NADPH binds to CPR and transfer electrons to FAD domain in its 'closed' conformation, and the electrons are transferred from FAD to FMN. The oxidized state of FAD will move away from FMN, resulting in its 'open' conformation, allowing the FMN domain to rotate away from the FAD domain for electron transfer to the CYPs (Quehl et al. 2017). CPR and CYP interactions are based on electrostatic forces. Therefore, the amino acid residues on solvent exposure of the FMN domain is a crucial for successful interaction with CYPs. The FMN-binding domain is a strong dipole rich in negatively charged residues such as Glu and Asp, which is critical for proper interaction with CYPs. These negatively charged residues interact with positively charged amino acids surrounding the heme cofactor of CYPs, which in rich in Arg and Lys (Laursen et al. 2011). Since the interflavin electron transfer is the rate limiting step in plant CPRs, the flexibity of the flexible FAD-FMN hinge is one of the most important factors in the optimization of CPR conformational states. Differences in amino acid residues in this hinge affected CPR flexibility and activity, including the reaction rate of CPRs (Ebrecht et al. 2019).

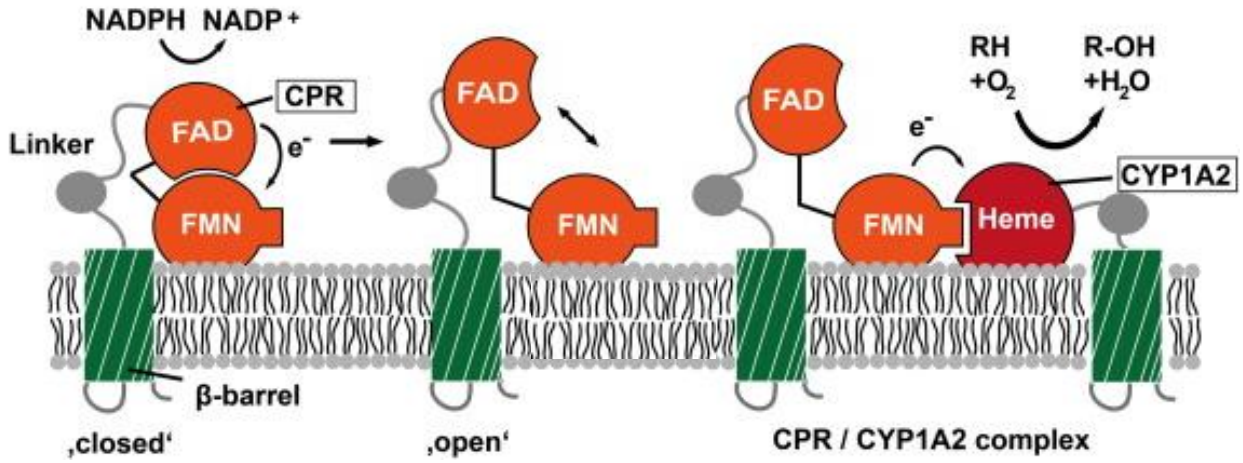

Figure 1-3. Schematic depiction of the biocatalysis of CPR and CYP1A2 on the lipid bilayer membrane. (Partially taken from (Quehl et al. 2017).

Other than CYPs, CPR is also capable to reduce microsomal heme oxygenase, cytochrome b_5 (Cytb5), and cytochrome c. Cytochrome b_5 is a microsomal hemoprotein and can also act as redox partner for some CYPs, making the relationship of CYP-CPR-Cytb5 is more complex. In some cases, Cytb5 is very essential in modulating CYP activity in human-drug metabolism, such as CYP3A4 (Gan et al. 2009). Cytb5 only capable to deliver the second electron to oxyferrous CYP but not strong enough to donate the first electron (Hamdane et al. 2008). CPR and Cytb5 has nonidentical but overlapping binding site in CYPs, leading to competition as redox partner. Therefore, CPR still acts as the obligatory redox partner of CYPs since it can transfer the first electron, but Cytb5 can either interfere CPR binding to CYP depending on Cytb5 concentration (Hamdane et al. 2008).

CPRs from different species such as human, yeast, and plants show high sequence homology, which is located in the highly conserved domain: the hydrophobic N-terminal membrane-binding domain and the hydrophilic C-terminal catalytic domain. The N-terminal membrane-binding domain also similar to bacterial flavodoxin, which is suggested to be important

for FMN binding. The C-terminal domain comprises of several functional domain, NADP and FAD-binding site. These highly conserved amino acid sequence among different species indicates the importance of CPR throughout the course of evolution (Wang et al. 1997).

1.2.1 Plant NADPH-cytochrome P450 reductase

Compared to other species, plants possess larger number of CYPs to help them organize more complex and various metabolisms, especially plant secondary metabolisms such as phenylpropanoids, alkaloids, plant hormone, including terpenoids biosynthesis (Rana et al. 2013). Various CYP families are involved in biosynthetic pathway and these eukaryotic CYPs are mostly anchored to the same endoplasmic reticulum as CPRs. Interestingly, unlike human, insects and yeast that only possess a single copy of CPR gene (Porter et al. 1990), several isoforms of CPRs were found in plants (Mizutani and Ohta 1998). One to three paralogs of CPR have been reported in angiosperms (Mizutani and Ohta 1998; Rana et al. 2013; Ro et al. 2002). Three CPR paralogs from plants were first discovered and isolated from *Arabidopsis thaliana*. While *ATR1* and *ATR2* were found to involve in plant biosynthetic pathways, *ATR3* was reported to be involved in the germination process and it possesses the most distinct sequences among all paralogs (Mizutani and Ohta 1998). Later it is known that *ATR3* encodes a diflavin reductase which is essential for embryogenesis (Niu et al. 2017)

Plant CPRs homologous to the first two paralogs of *Arabidopsis thaliana* CPRs were then categorized as CPR class I and II which showed different clade in the phylogenetic tree based on their differences in N-terminal membrane sequences. Among these CPR classes, at least the expression of one of them is inducible during mechanical wounding or stress treatment. Interestingly, all inducible CPRs so far are reported to be of CPR class II, while CPR class I expression is constitutive and uninducible (Rana et al. 2013; Parage et al. 2016; Huang et al. 2012).

CPR class II was also reported to highly correlated with genes involved in specialized metabolisms, compared to CPR class I. In contrast to *ATR1*, expression of *ATR2* in *Arabidopsis* stems is coregulated with the general phenylpropanoids and monolignol pathway, including CYP genes *C4H*, *C3H1*, and *F5H1* (Sundin et al. 2014). The expression of *GhCPR2* was increased in seed after 15-20 DPA, which has been reported to be highly active in gossypol (cotton) biosynthesis (Yang et al. 2010). While *GhCPR1* expression level did not show any significant different in gland and glandless cultivars, *GhCPR2* expressed highly in cotton gland, and 40% lower in glandless petals and ovules (Yang et al. 2010). On the other hand, *CPR1* expression in *Capsicum spp.* showed correlation with basal pungent alkaloid biosynthesis (Mazourek et al. 2009; Lee et al. 2014). However, both *GhCPR* classes actually showed reductase activity against *CYP73A25*, a cinnamate 4-hydroxylase of cotton *in vitro*, implying both CPR classes can support CYPs involved in specialized metabolisms (Rana et al. 2013).

Both CPR classes are required to support huge number of CYPs in plants, involved in both primary and specialized metabolisms. Since CPR class II, not CPR class I, seems to be closely related to be inducible by abiotic and biotic factors, it is now believed that CPR class I may play more important role in plant primary and basal specialized metabolisms, while CPR class II are mosre responsible for plant adaption and defense mechanisms such as stress or injury (Ro et al. 2002). In rat liver, CPR and CYP have been reported to present in microsomes in a ratio of 1:15. This high ratio of CYP to CPR implies CYP competitions of CPR and that they must be particularly organized in order to ensure electron supply from CPRs to abundant CYPs at once (Shephard et al. 1983). The expansion of plant CPRs may reflect the numerous and diversity of CYPs involved in different biosynthetic pathway and more CPRs are required to meet the high demand of electron supply during primary growth or secondary metabolism in response to biotic and abiotic stress

(Rana et al. 2013). Therefore, the physiological relevance of the presence of multiple CPRs in plant is important to be elucidated to understand their relationship with various CYPs.

1.3 Functional analysis *in planta*

As plant genetic engineering and next generation sequencing technology are advancing, many powerful approaches are now available to study gene functional analysis *in planta*. More sophisticated DNA recombinant technology in plants lead to the development of various reverse genetics strategies. Strategies to investigate gene function in plants by genetic modification is to examine the effect of loss-of-function or gain-of-function gene mutation to the phenotypic changes. Loss-of-function method is more commonly used as a direct method to elucidate the function of the mutated target gene by genome editing technology such as insertion or deletion mutations. In contrast, gain-of-function mutants method allows a target gene to be inserted into the plant genome and the correlation will be drawn with the resulted phenotypic change. This method benefits in investigating function of gene with functional redundancy which might come from same gene families. Gain-of-function mutants might reveal gene functions that cannot be achieved by conventional loss-of-function mutants (Kuromori et al. 2009).

1.3.1 Loss-of-function gene mutation

Compared to gain-of-function gene mutation method, gene disruption in loss-of-function gene mutation method provides more powerful tool to obtain knockout mutants that provide direct causal relationship between gene sequence and function (Radhamony et al. 2005). CRISPR-Cas9 technology has been widely applied to induce targeted mutagenesis in plant genes. The Cas9 nuclease guided with sgRNA can induce double-stranded break of the targeted genes, resulted in gene mutation such as indels or point mutations which generate loss-of-function mutants due to

frameshift mutations or early termination. The guide RNAs and Cas nucleases can be delivered into the plant cell nucleus by agrobacterium-mediated transformation, particle bombardment, or PEG-mediated transfer. The most common method is the Agrobacterium-mediated transformation and the technique varies across species, such as floral dip in Arabidopsis, tissue transformation from callus, leaf, or stem, and hairy root transformation (Anjanappa and Gruissem 2021; Soyars et al. 2018). However, the plant ability to recover both mutated and unmutated cells in *Agrobacterium tumefaciens*-mediated transformed tissue culture increases the possibility for the chimeric plants to be recovered. Rice tissue culture and Arabidopsis floral-dip mutants showed chimeric mutations after prolonged culture period (Zhang et al. 2020).

Agrobacterium rhizogenes mediated transformation to generate hairy root mutant is now becoming more popular for functional genomics. The Ri plasmid of *A. rhizogenes* contains Ri T-DNA that induces hairy roots, which can be transferred into plant cells along with Ti T-DNA that carries both CRISPR-Cas9 and gRNA expression cassettes which give advantage for the *A. rhizogenes* transformation. The fast and infinite proliferation of hairy roots also provide benefits to obtain higher mass in short time, about 5-10 days. *A. rhizogenes* transformation method is also considered easier compared to *A. tumefaciens* mediated transformation, making hairy roots genome editing system has been widely applied to different plants (Zhang et al. 2020). Some of the plants with well-established hairy root system is legumes plants, such as *M. truncatula* (Zhang et al. 2020), *L. japonicus* (Okamoto et al. 2013), and soybean (Cai et al. 2015).

Another method for generating loss-of-function mutant is by insertional mutagenesis. This method enables us to locate the mutated genes by known inserted fragments. Arabidopsis and rice mutant resources are publicly available and mostly are insertion tagged lines, which makes the investigation of gene function is fast and convenient (Kuromori et al. 2009). Mostly used tag for

insertional mutagenesis are transfer DNA (T-DNA) tag and transposon tag. In legumes, *M. truncatula* and *L. japonicus* also have their mutant library publicly available. *M. truncatula* mutant library was constructed using Targeting Induced Local Lesions in Genomes (TILLING) technology, which is a reverse-genetic method consists of chemical random mutagenesis and a high throughput screening of point mutations in the regions of interest (Carelli et al. 2013). *LORE1* insertion mutant library of *L. japonicus* was generated by inserting *LORE1* retrotransposon into the germline and these ~5000 bp on random exonic insertions provide a null-mutant resources with random loss-of-function genes (Urbański et al. 2012). With the availability of genome editing platform and mutant library in legume plants such as *L. japonicus*, deeper studies on triterpenoids biosynthesis, including the involvement of different plant CPR classes would be possible to be carried out.

1.4 Heterologous triterpenoids production in transgenic yeast

A plant natural compounds such as secondary metabolites are very important to promote various health benefits in human, ensuring a sufficient supply to meet the demand is highly essential. However, extraction of metabolites from its plant natural resources are very environmentally unfriendly and costly, because the inadequate yield of the compound in the plant requires abundant plant biomass. Complex secondary metabolism in plants also requires further purification of the desired metabolites from similar chemical compounds. Complexity of plant natural compounds also causes the chemical synthesis is highly difficult. As a result, there is a strong need for alternative production platform for plant natural products (Cravens et al. 2019; Moses et al. 2017).

Microbial host has been widely used for plant metabolite production. *Saccharomyces cerevisiae* or budding yeast, has been proven to be the most suitable heterologous host for

producing diverse plant secondary metabolites among other microorganisms. Yeast is also commonly used in industrial applications such as for food and beverage. As a key model organism for fundamental molecular biology, genome of *S. cerevisiae* has been completely sequenced. The available of complete genomic database of yeast has supported researchers to reveal the detailed models of yeast metabolic pathway and processes, which ease the way for metabolic network control and manipulation towards diverse plant secondary metabolite production (Siddiqui et al. 2012).

CYPs redox system are classified into several classes. Class I CYP system is prokaryotic CYPs consisting of cytosolic and soluble three-component CYP–Fdx–FDR electron chain. Fdx is an Fe-containing ferredoxin and FDR is a flavin adenine dinucleotide (FAD)-containing ferredoxin. On the other hand, class II CYPs system is eukaryotic CYPs consisting of CYP-CPR membrane bound proteins, which exists in plants (Finnigan et al. 2020). Most of the fungal CYPs belong to eukaryotic CYPs class II and perform extremely diverse catalytic reactions, including in *Saccharomyces cerevisiae* (Durairaj et al. 2016). Therefore, compared to prokaryotic *Escherichia coli*, eukaryotic yeast is particularly suitable for functional expression of heterologous eukaryotic enzymes, including CYPs and CPRs that require endomembrane to exhibit their activity. Post-translational modifications of expressed plant enzymes can also be carried out in yeast cell. Moreover, the availability of some pathways in yeast compared to *E. coli* provides more suitable environment for CYPs activity towards plant secondary metabolism, such as terpenoids (Siddiqui et al. 2012). Yeast contains endogenous enzymes involve in the biosynthetic pathway of the common precursor for terpenoids production. Similar to plants, yeast produces isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) as squalene precursor from acetyl-coA through mevalonate (MVA) pathway (Carsanba et al. 2021). Yeast does not contain any

terpenoids biosynthetic enzymes, providing ways for heterologous expression of terpenoids enzymes without distubing with yeast native metabolic pathway. Proven to be an ideal host for secondary metabolite production, diverse terpenoids have been successfully produced in transgenic yeast by applying creative metabolic engineering strategies (Zhang and Hong 2020; Table 1-2).

Most of these strategies are focused on engineering the oxidizing enzymes such as CYPs, HMGR, and squalene synthase. However, very few attempts have been tried to consider the importance of CPR as CYP redox partner. Since the knowledge of different plant CPR classes starts to surface, investigating the effect of heterologous expression of different plant CPR classes in transgenic yeast would be required to further improve plant secondary metabolite production, including triterpenoids.

Table 1-2. Strategies for production of various terpenoids in *S. cerevisiae* (modified from Zhang and Hong 2020)

Product	Strategy and features	Culture conditions	Titer or Improvement
8-hydroxygeraniol	Mitochondrial compartmentalization by targeting the geraniol biosynthetic pathway to the mitochondria	Fed-batch fermentation	227 mg/L
Geraniol	Protein structure analysis, site-directed mutation, overexpression of tHMGR and IDI	Fed-batch fermentation	1.68 g/L
Limonene	Regulation of ERG20 by PHXT1 promoter (glucose-sensitive)	Fed-batch fermentation	917.7 mg/L 6-fold
	N-degron-mediated destabilization of ERG20	Batch fermentation	76 mg/L
Amorpha-4,11-diene	Optimization of [NADPH]/[$NADP^+$] ratios by introducing mutations into phosphofructokinase (PFK) along with overexpression of ZWF1	Shake flasks	497 mg/L
	Mitochondria compartmentalization by targeting the whole FPP pathway together with Amorpha-4,11-diene synthase (ADS) into mitochondria	Shake flasks	427 mg/L

Zerumbone	Regulation of ERG9 by PHXT1 promoter	Fed-batch fermentation	40 mg/L
Farnesene	Increase the availability of acetyl-CoA by removing the native source of cytosolic acetyl-CoA (Δ*RHR2*) and overexpressing xPK, PTA, ADA and NADH-HMGr	Fed-batch fermentation	2.24 g/L/h >130 g/L
Oxygenated taxanes	*E. coli–S. cerevisiae* co-culture by dividing the synthetic pathway for the acetylated diol paclitaxel precursor into two modules	Co-culture in bioreactor	33 mg/L
Nerolidol	Minimizing competition for FPP by destabilizing squalene synthase, degrade ER membrane-integrating protein.	Two-phase flask	4–5.5 g/L
Casbene	Regulation of ERG20 and ERG9 by P*HXT1* and P*ERG1* promoters	Deepwell microplate	108.5 mg/L
Jolkinol C	Optimize soluble expression of Cbsp using protein tagging strategies, codon-optimization of CYPs	Milliliter plates	800 mg/L
Carotenoid	Colorimetric-based promoter strength comparison system; inducer/repressor-free sequential control strategy by combining a modified GAL regulation system and a PHXT1-controlled squalene synthetic pathway	Fed-batch fermentation	1156 mg/L
Lycopene	Lipid engineering; Improve triacylglycerol metabolism	Fed-batch fermentation	2.37 g/L
	Scaffold-free enzyme assemblies (IDI and CrtE);	Fed-batch fermentation	2.3 g/L
Medicagenic acid	Endoplasmic reticulum (ER) engineering; expand the ER by disrupting the phosphatidic acid phosphatase	Tube cultures	27.1 mg/L 6-fold
β-Carotene	Tri-functional CRISPR system combines transcriptional activation, transcriptional interference, and gene deletion	Tube cultures	2.8-fold
Squalene	ER engineering; expand the ER by overexpressing a key ER size regulatory factor, INO2.	Shake flasks	634 mg/L

1.5 Objective and strategy

The low accumulation of beneficial triterpenoids in plants requires a new strategy for improving productivity. *S. cerevisiae* as microbial host provides an ideal alternative for heterologous triterpenoids production. Some modifications through heterologous expression of plant genes can be applied to improve triterpenoids productivity in transgenic yeast. In addition to their health benefits, legumes also provide a good plant source for triterpenoids biosynthetic genes. Most of the yeast metabolic engineering attempts to produce terpenoids are focused on CYPs, as the most responsible enzymes for structural diversification of triterpenoids. However, very little attention has been paid to its redox partner, CPR, which is very essential for CYP catalytic activity. The increasing knowledge on involvement of different plant CPR classes towards plant secondary metabolism suggests that different CPR classes might also affect triterpenoids biosynthesis. Understanding the physiological relevance of different CPR classes *in planta* might propose a strategy for improving heterologous triterpenoids production in transgenic yeast.

Therefore, the objective of this study is to perform comparative analysis of legume CPR classes towards improving heterologous production of triterpenoids in transgenic yeast. There are three main strategies to achieve this objective. The first is to perform structural and gene co-expression analysis to give insight on sequence-functional relationship of different legume CPR classes. Then, elicitation experiment on *L. japonicus* hairy root was performed to reveal triterpenoids biosynthetic genes relationship, followed by functional analysis using two types *L. japonicus* cpr loss-of-function mutants, *LORE1* insertion mutant plants and CRISPR-Cas9 hairy root knockout mutants. By reverse genetics, loss of *cpr* genes can be correlated with the change in triterpenoids composition of the mutants. Finally, heterologous expression in transgenic yeast is conducted by co-expressing different legume CPR classes in pairs with different CYPs to

investigate the effect of different CYP-CPR pair on heterologous triterpenoids production. Different plant CPR classes were also co-expressed in engineered yeast strain to further explore the possibility to enhance the triterpenoids production.

1.6 Thesis outline

This thesis is divided into five chapters. Chapter 1 describes the general introduction of the importance of triterpenoids and their biosynthesis, especially in legumes, and the involvement of CPRs as the obligatory redox partner for CYPs in triterpenoids biosynthesis. Introduction of functional analysis *in planta* and heterologous expression in transgenic yeast as the main strategies for this study are also described briefly in this chapter.

In chapter 2, several plants CPRs from different plant families were mined from different plant genome database to perform phylogenetic tree and multiple sequence alignment analysis. One gene for each CPR class were mined for each dicotyledonous plant species. Numerous legume CPRs were then mined to perform motif analysis on the differently conserved amino acids between legume CPR class I and II. Among these different motifs, several amino acids that were considered to be important for CYP-CPR interaction reported in previous studies were chosen for further analysis. By homology modelling, MtCPR class I and II were analyzed in silico to investigate the effect of highlighted amino acid difference between legume CPR class I and II. These structural differences provide insights of how each CPR class might exhibit different specificity or affinity towards particular CYPs.

The work on chapter 3 focused on elucidating the differential function of plant CPR class I and II *in planta* by using *L. japonicus*. To observe the CPRs involvement with other triterpenoids biosynthetic genes, *L. japonicus* hairy root is treated with methyl jasmonate (MeJA) as phytohormone elicitors. Gene expression and triterpenoids profile analysis of the elicited hairy

roots were then performed which revealed different set of genes which strongly correlated with either *LjCPR1* and *LjCPR2*. To confirm the involvement of LjCPR class I and II in different triterpenoids biosynthesis, metabolite analysis of *Ljcpr* loss-of-function mutants was performed. *Ljcpr* loss-of-mutants were obtained by cultivating soil- and hydroponic-cultured of *LORE1* insertion mutants and generating *Ljcpr* knockout hairy root mutants. The functional analysis of *Ljcpr* loss-of-function mutants strengthen the hypothesis of different involvement of LjCPR class I and II towards triterpenoids biosynthesis.

To apply the information obtained from chapter 3 for heterologous triterpenoids production, in chapter 4, different pairs of legume CPR class and CYP family were co-expressed in transgenic yeast and metabolite analysis was performed. Different triterpenoids conversion ratio was observed in different CYP-CPR pairs, which suggested that the CPR pair is essential in heterologous triterpenoids production. Some triterpenoids showed better conversion ratio when the CYPs were paired with CPR class I, while other triterpenoids showed better conversion ratio with CPR class II. The metabolite analysis of N-terminal membrane domain swapping between MtCPR1 and MtCPR2 suggested that this domain is responsible for CYP-CPR protein-protein interaction, rather than affecting CPR catalytic activity. Similar CYP-CPR pairs were also co-expressed in engineered yeast strain, previously constructed to enhance triterpenoids production. The engineered yeast co-expressing certain CPR showed significantly higher triterpenoids conversion ratio for pre-cursor production of high value compounds.

Finally, in chapter 5 the conclusions of this study and the future perspectives were proposed.

Chapter 2

Structural and gene co-expression analysis of legume CPRs

2.1 Introduction

Cytochrome P450 monooxygenases (CYPs) are the most important enzymes responsible for the structural diversity of triterpenoids (Xu et al. 2004). CYPs are classified based on their sequence similarities (Moktali et al. 2012). The CYP716A and CYP72A subfamily are widespread in plants and show diverse functions in triterpenoid biosynthesis, while other CYP subfamilies such as CYP93E and CYP88D are known to be unique to legumes (Seki et al. 2015) (Figure 2-1). To perform catalysis, CYPs require the electrons that are transferred by CPRs. Electrons from NADPH flow through FAD to the FMN domains of CPRs, which are finally transferred to the heme center of CYPs (Paine et al. 2005).

CPR and CYP are both membrane-bound proteins that have been reported to be present in microsomes in a ratio of 1:15 (Shephard et al. 1983). This high ratio of CYP to CPR implies that CYPs compete for CPRs and that they must be particularly organized in order to ensure electron supply from CPRs to huge numbers of CYPs at once (Shephard et al. 1983). Previous study on human CPR (hCPR) showed that specific mutations on FMN domain improved interactions with a specific CYP (Esteves et al. 2020). Several point mutations were found to be co-localized with mutations found in naturally occurring hCPR variants that caused CYP isoform-dependency (Pandey and Flück 2013; Burkhard et al. 2017) and negatively charged residues that has been previously suggested to be important in CPR:CYP interactions (Hamdane et al. 2009; Hong et al. 2010; Jang et al. 2010).

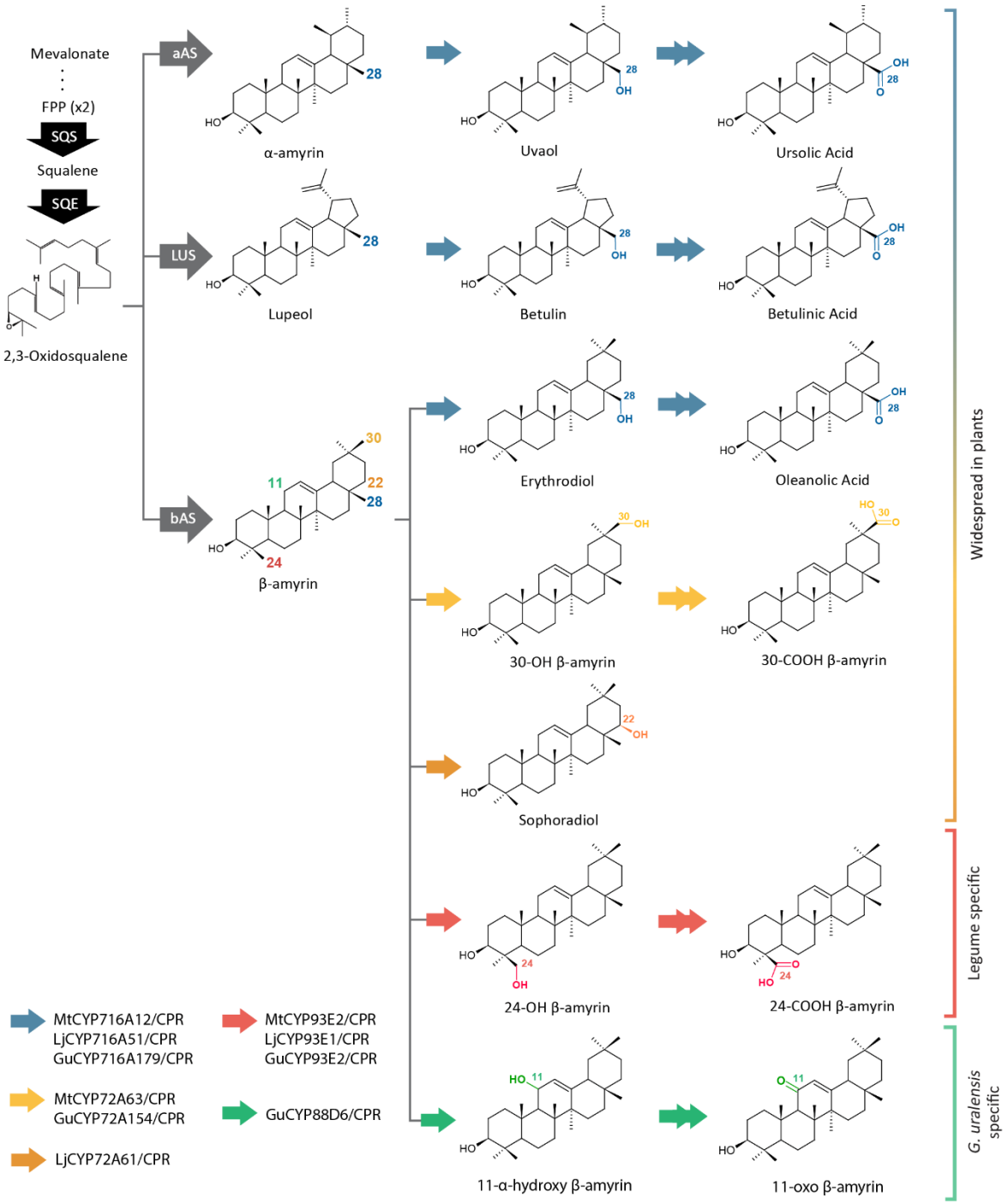

Figure 2-1. Triterpenoid biosynthetic pathways of several CYP families from *M. truncatula* (Mt), *L. japonicus* (Lj), and *G. uralensis* (Gu).

Single and double arrows indicate one and two oxidation steps, respectively. CYP, cytochrome P450; FPP, farnesyl pyrophosphate; SQS, squalene synthase; SQE, squalene epoxidase; bAS, β-amyrin synthase; aAS, α-amyrin synthase; LUS, lupeol synthase.

Unlike mammals and fungi with one CPR gene, plants have multiple CPR genes, depending on the species (Mizutani and Ohta 1998; Jensen and Møller 2010). Plant CPRs are branched into two classes, CPR class I and class II (Qu et al. 2015). CPR class I generally has a shorter N-terminal membrane sequence than CPR class II (Parage et al. 2016). CPR class I is reported to be constitutively expressed and plays a role in primary or basal constitutive metabolism, while CPR class II is inducible by environmental stimuli and involves more defense mechanisms through plant secondary metabolism (Ro et al. 2002; Qu et al. 2015; Yang et al. 2010; Rana et al. 2013; Parage et al. 2016). Different tissue expression profiles of CPR class I and II were reported in *Withania somnifera*, *Panax ginseng*, *Camelia sinensis*, and *Catharantus roseus*. The differences in the protein sequence and expression level of CPR class I and II suggest a possibility of their different roles in plants. Therefore, insight on differential gene co-expression and protein structures of legume CPRs provide a good basis to understand the difference between legume CPR class I and II.

2.2 Materials and Method

2.2.1 Data mining and phylogenetic analysis of plant CPRs

Plant CPR sequences from 20 plant species were obtained from the web-based resource for Arabidopsis P450, cytochrome b_5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (www.P450.kvl.dk), NCBI (https://www.ncbi.nlm.nih.gov/), and each plant genome database. Accession numbers of CPR genes and genome databases used in this study are listed in Table S1. Multiple sequences were aligned using ClustalW and were used for tree construction using MEGA7, with the neighbor-joining method.

2.2.2 Motif and structural analysis of legume CPRs

The position of protein helices was predicted using an in silico transmembrane helix prediction tool (http://www.cbs.dtu.dk/services/TMHMM/) (Figure S1). The conserved membrane anchor, FMN-, FAD-, CYP-, and NADPH-binding domains were predicted as described by Qu et al. (2015). Motif analysis of FMN domain of legume CPR class I and II were performed with the MEME Suite online software (http://meme-suite.org) using minimum of 37 sequences for each CPR class from 24 different legume species (Figure S2). MtCPRs structure model were constructed using 5gxu.1.A (crystal structure of Arabidopsis ATR2) as template with >71% identity in SWISS-MODEL (Arnold et al. 2006). PyMol (DeLano 2020) was used to visualize the amino acids in the active site of the enzyme.

2.2.2 Gene co-expression analysis of legume CPRs

Co-expression analysis of *M. truncatula* and *L. japonicus* was performed using an online transcriptomic database from the gene expression atlas web servers https://lipm-browsers.toulouse.inra.fr/pub/expressionAtlas/app/mtgeav3 and https://lotus.au.dk/expat/ (Gifu v1.2 and Miyakojima MG20 v 3.0), respectively. Transcriptomic data for *M. truncatula*, *L. japonicus*, and *G. uralensis* were obtained from RNA-seq data submitted to the DNA Data Bank of Japan (DDBJ) Sequence Read Archive (DRA) under the accession numbers of DRA012266 (Istiandari et al. 2021)

2.3Result and discussion

2.3.1 Phylogenetic analysis of plant CPRs

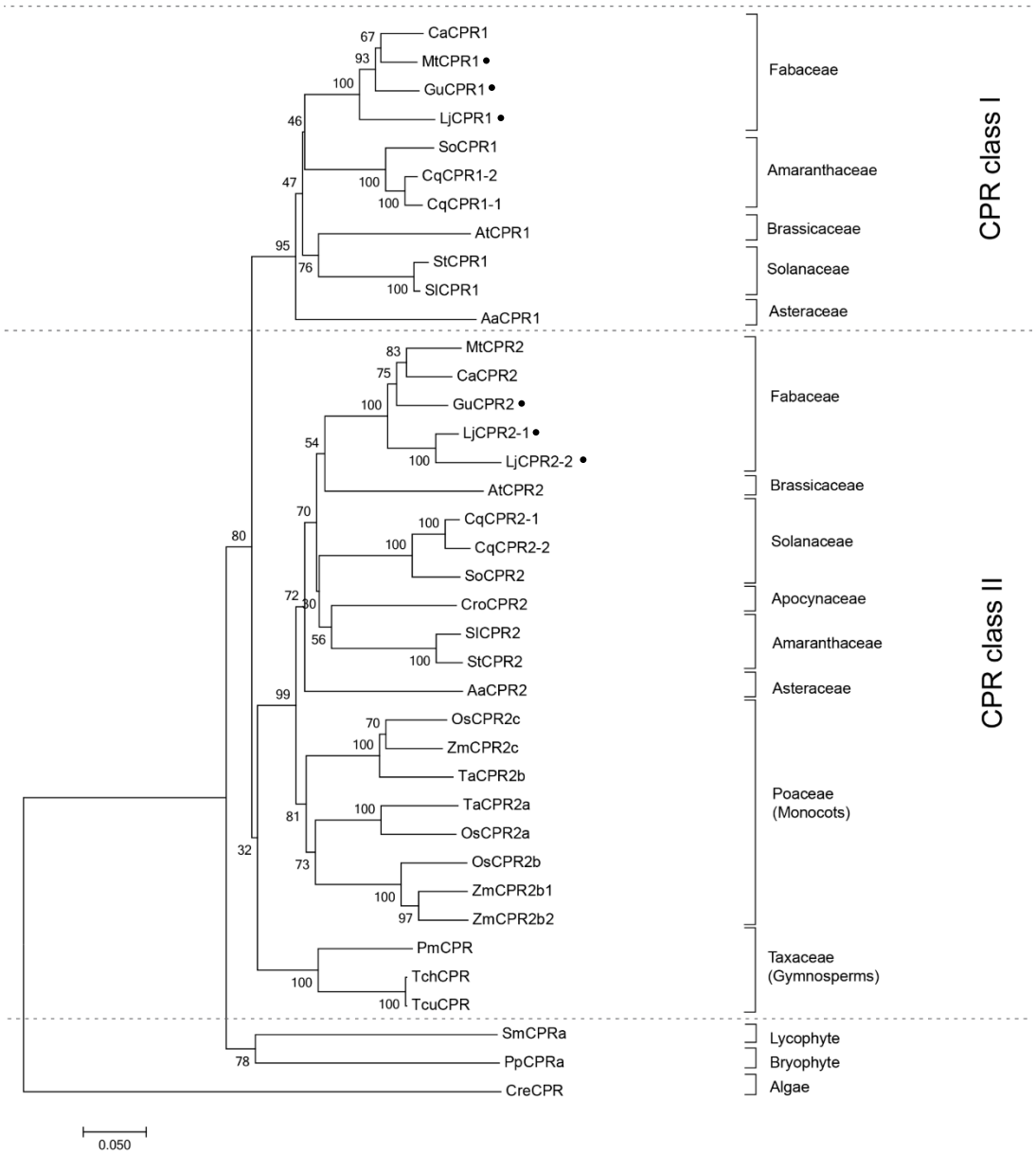

Figure 2-2. Molecular phylogenetic tree of 38 CPR amino acid sequences from 20 different plant species shows clear branching of class I and II CPRs in higher plants.

Sequences of CPR classes I and II from *M. truncatula*, *L. japonicus*, and *G. uralensis* that were used for further analysis in this study are indicated as black dots. CPR, cytochrome P450 reductase; Aa, Artemisia annua; ATR1, Arabidopsis thaliana CPR1; ATR2, Arabidopsis thaliana CPR2; Ca, Cicer arietinum; Cre, Chlamydomonas

reinhardtii; Cro, Catharantus roseus; Cq, Chenopodium quinoa; Gu, *G. uralensis*; Lj, *L. japonicus*; Mt, *M. truncatula*; Os, Oryza sativa; Pm, Pseudotsuga menziesii; Pp, Physcomitrella patens; Sl, Solanum lycopersicum; Sm, Selaginella moellendorffii; St, Solanum tuberosum; Ta, Triticum aestivum; Tch, Taxus chinensis; Tcu, Taxus cupidata; Zm, Zea mays

The phylogenetic analysis results from 20 different plant species showed clear branching of CPR class I and II sequences (Figure 2-2). Each dicot plant species was found to have a minimum of one CPR sequence in each class. Monocots, gymnosperms, and mosses possess several copies of CPR class II genes, but no CPRs belong to CPR class I was found up until now (Jensen and Møller 2010). Amino acid sequences of plant CPRs from different plant families were compared to create a table of similarity matrices (Table S2). Same CPR classes of the same plant family were found to have 75-95% similarity, while same CPR classes of different plant families were found to have 64-75% similarity. On the other hand, different CPR classes of the same or different plant families were found to have 55-65% similarity. Motif analysis was performed on different plant CPR classes (Figure S2). Multiple sequence alignments from several plant CPRs showed that CPR class I has approximately 10–20 amino acids shorter N-terminal membrane sequences than CPR class II (Figure 2-3). Genes from each CPR class for *M. truncatula*, *L. japonicus*, and *G. uralensis* were obtained and chosen for further analysis of legume CPRs (Figure 2-2). One gene of each CPR class was found in *M. truncatula* and *G. uralensis* genome database, while from the latest version of *L. japonicus* Gifu v1.2 genome database, one CPR class I and two CPR class II genes were found.

Based on the phylogenetic results of plant CPRs, it was identified that dicotyledons possess both CPR class I and II, while so far, only CPR class II genes were identified in monocotyledons, gymnosperms, and mosses. Among angiosperms, as compared to dicotyledons, monocotyledons are well known to be more tolerant to abiotic (Agrawal and Agrawal 1999; Rao et al. 2006; Wright et al. 2010). Since CPR class II is believed to play a more important role in defense and adaptation

mechanisms compared to CPR class I (Rana et al. 2013; Parage et al. 2016). The fact that multiple copies of CPR class II were found in all monocotyledons while some dicotyledons might only possess single CPR class II (Figure 2-2; Jensen and Møller, 2010), suggest that the role of CPR class II is very prominent in supporting abiotic resistance in monocots. In addition, the more sophisticated translocation and root system of dicotyledons (Scarpella and Meijer 2004), might be supported by the presence of CPR class I found in dicotyledons. Nevertheless, it remains to be shown whether monocots really only possess CPR class II.

2.3.2 Structural analysis of MtCPR class I and II

In this study, we focused on the structural analysis of FMN domain of the legume CPRs using CPR from a model legume, MtCPR. Based on motif analysis of amino acid from CPR class I and II from 24 legume species, there are some differences in amino acid residues between CPR classes I and II in the FMN domain which are conserved in each CPR classes (Figure S2). By multiple sequence alignment with hCPR, we highlighted 8 amino acid residues of MtCPR1 and MtCPR2 which are co-localized with the position of hCPR mutations and negatively charged amino acid residues on FMN domain which are reported to play role in CPR:CYP specific interactions (Esteves et al. 2020; Campelo et al. 2018) (Figure 2-3). The location of each highlighted amino acid in the protein structure can be observed through structural analysis of MtCPRs (Figure 2-4).

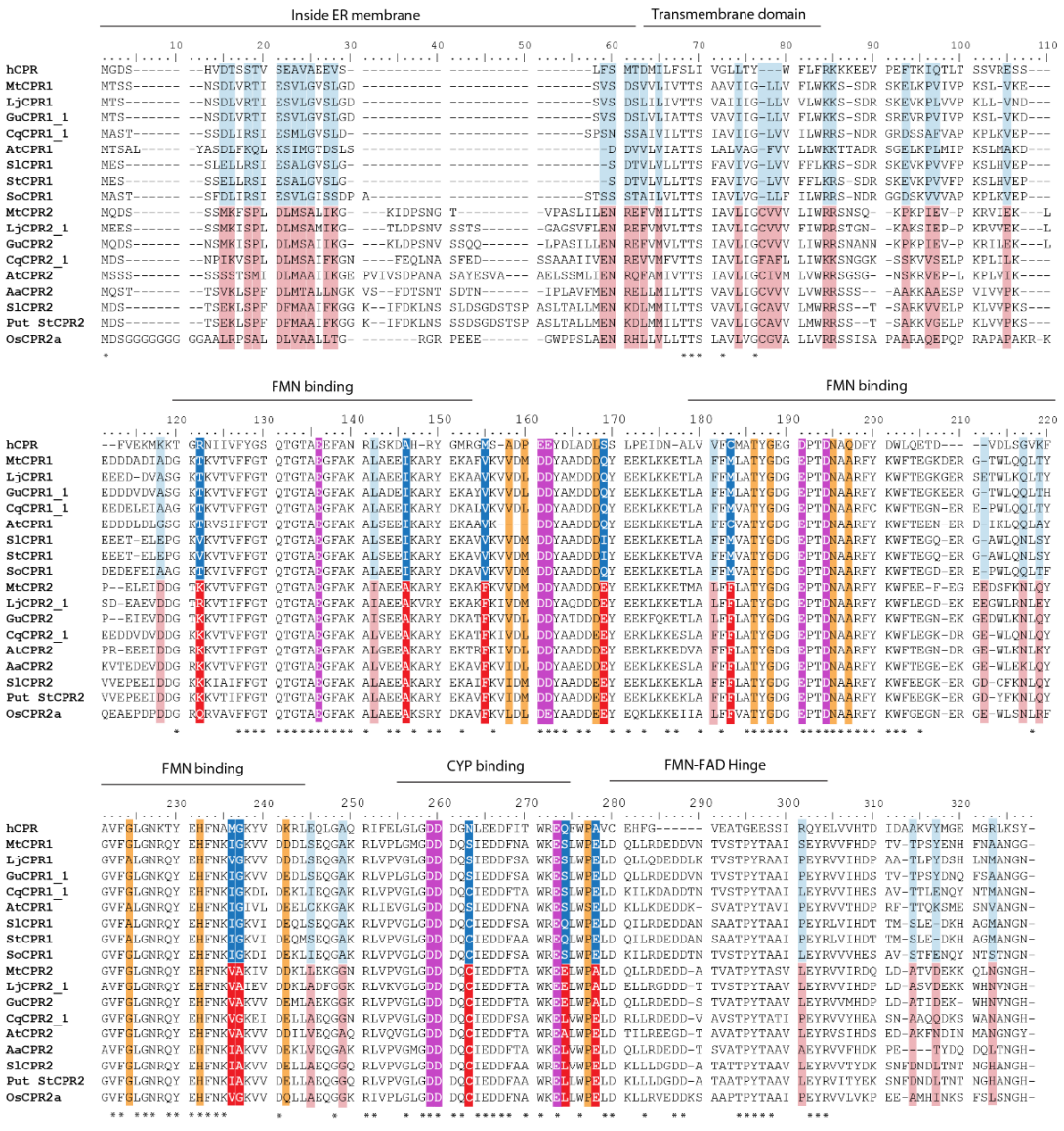

Figure 2-3. Multiple sequence alignment of first 327 amino acid sequences of CPR class I and II containing N-terminal and FMN domains using the ClustalX color scheme for amino acid alignments.

Dark blue, red, light blue, and pink colors indicate the different conserved amino acids in CPR class I and II in legume family based on motif analysis (Figure S2). Dark blue and red colors are the eight highlighted key residue in this study. Magenta color indicates acidic residue formerly reported to be important in CYP:CPR interaction in human CPR (hCPR). Orange color indicates point mutations in hCPR that are reported to improve interaction with a specific CYP. Asterisks indicate the positions which have a single, fully conserved residue. CPR, cytochrome P450 reductase; Aa, Artemisia annua; ATR1, Arabidopsis thaliana CPR1; ATR2, Arabidopsis thaliana CPR2; Ca, Cicer arietinum; Cq, Chenopodium quinoa; Gu, *G. uralensis*; Lj, *L. japonicus*; Mt, *M. truncatula*; Os, Oryza sativa; Sl, Solanum lycopersicum; St, Solanum tuberosum

The FMN domain of CPRs have several conserved patches of acidic residues located on its solvent exposed surface, suggested to be involved in the electrostatic interaction with its redox partners, including CYPs (Waskell and Kim, 2015). It was shown that in human, every CYP is interacting with CPR in a specific manner, based on the difference in the binding motifs located in the FMN domain (Esteves et al. 2020). The motif analysis from 24 different legume species showed that there are some differences in the conserved amino acid residues between legume CPR class I and II in the FMN domain. Among them, 8 residues either located on the conserved binding domain or near the formerly suggested residue to be important of CYP:CPR interactions in human were highlighted (Table 2-1). CPR from a model legume, MtCPRs, were used for the 3D structural analysis. The first highlighted amino acids, T81 in MtCPR1 and K92 in MtCPR2, are located on the first residue in the FMN binding domain (Figure 2-4). This difference in the residues is conserved in CPR I and II of legumes and other species as well. Based on the structural analysis, these residues are located near the FMN-FAD hinge (Figure 2-4). It was previously reported that the interflavin electron transfer (ET) is the rate-limiting step in plant CPR (Simtchouk et al. 2013; Whitelaw et al. 2015). This highly flexible hinge also has been suggested to play role in formations of multiple open conformations to form complexes with CYP in an isoform-dependent manner (Campelo et al. 2018). Therefore, the difference in size and charge property of T81 in MtCPR1 and K92 in MtCPR2 might greatly affect the conformational flexibility of the interflavin hinge during ET. The next highlighted residues are I105, V114, and M142 in MtCPR1 while they are A116, F125, and F153 in MtCPR2, respectively (Figure 2-4). Even though these amino acid residues are all hydrophobic and only exhibit small differences in the chain length, these different residues of MtCPR1 and MtCPR2 are conserved in CPR I and II of legumes and other plant families (Figure 2-3). These residues are located in the α-helix of the FMN-binding domain and

close to the acidic residues important for ET and these residues are facing each other (Figure 2-4). The difference in these residues might cause minor changes in the geometrical shape of α-helix of the FMN-binding domain, leading to improve interactions with specific CYP, in which small conformational perturbations of the α-carbon backbone might accommodate change in substrate binding (Bart and Scott 2018). One of the most important highlighted residues are Q128 in MtCPR1 and E139 in MtCPR2 (Figure 2-4), which also are conserved not only in legume, but also in other plant families (Figure 2-3). These residues are located in the acidic-rich residues on the solvent exposed surface which highly possible to serve as binding motifs for CPR:CYP interaction (Figure 2-4). The glutamate (E139) residue in CPR II will provide stronger negative charged to form ionic bond with the respective CYP. Changing this to uncharged glutamine residue as in CPR I (Q128) will greatly affect the CYP interaction. The last three highlighted residues are S221, S232, E236 in MtCPR1 and C231, E242, and A246 in MtCPR2, respectively (Figure 4). Interestingly, these residues seem to be conserved only in CPR I and II of legume family (Figure 2-3; Figure S2). These residues are located on the solvent exposed surface of FMN domain and near the acidic residue important for CPR:CYP interaction (Figure 2-4). C231 in MtCPR2 can form disulfide bridges with the interacting CYP, which cannot be achieved by S221 in MtCPR1. Serine to cysteine mutations is also reported to have stabilizing effect due to optimization of van der Waals interactions and packing (Santos et al. 2007). Hence, the C231 in CPRII might account for more stable interaction with specific CYP, compared to CPRI. The last two residues are acidic residue of E242 in MtCPR2, while it is S232 in MtCPR1, and acidic residue of E236 in MtCPR1, while it is A246 in MtCPR2. The difference between the negatively charged glutamate (E) residue and short aliphatic alanine (A) or uncharged serine residue (S) will greatly affect the ionic interaction between CPR and its ET partner. The position of the acidic residues in CPR I and II are different

but parallel in the α-helix (Figure 2-4), which highly possible to lead the CYP isoform-specific interaction. The fact that these positions are conserved only in legume CPRs suggest that CPR from each plant families might exhibit unique CYP binding motifs to interact with specific CYP. Since CYPs are classified into different families and subfamilies based on sequence homology, the deduction is that different CYP families might show preference towards different CPR classes due to these sequence-structure relationship.

Table 2-1. Summary of the eight selected different conserved amino acid residues in the FMN domain protein structure of MtCPR1 and MtCPR2.

<table>
<tr><th>No.</th><th>MtCPR1</th><th>Amino acid characteristics</th><th>MtCPR2</th><th>Amino acid characteristics</th><th>Position</th><th>Amino acid conservation</th><th>Effect</th></tr>
<tr><td>1</td><td>T81</td><td>Small, polar-uncharged</td><td>K92</td><td>Large, positively charged</td><td>FMN-binding domain, near FAD-FMN hinge</td><td rowspan="5">Conserved in different plant families</td><td>Electron transfer from FAD to FMN</td></tr>
<tr><td>2</td><td>I105</td><td>Small, hydrophobic</td><td>A116</td><td>Small, hydrophobic</td><td>FMN-binding domain</td><td rowspan="3">The conformation of FMN-binding domain towards CYP</td></tr>
<tr><td>3</td><td>V114</td><td>Small, hydrophobic</td><td>F125</td><td>Large, hydrophobic</td><td rowspan="6">Solvent exposed surface, close to important acidic residue for CYP binding</td></tr>
<tr><td>4</td><td>Q128</td><td>Polar-uncharged</td><td>E139</td><td>Negatively charged</td><td rowspan="6">CYP:CPR specific interaction</td></tr>
<tr><td>5</td><td>M142</td><td>Small, hydrophobic</td><td>F153</td><td>Large, hydrophobic</td></tr>
<tr><td>6</td><td>S221</td><td>Small, polar-uncharged</td><td>C231</td><td>Small, sulfur-containing</td><td rowspan="3">Conserved in legumes only</td></tr>
<tr><td>7</td><td>S232</td><td>Small, polar-uncharged</td><td>E242</td><td>Large, negatively charged</td></tr>
<tr><td>8</td><td>E236</td><td>Large, negatively charged</td><td>A246</td><td>Small, hydrophobic</td></tr>
</table>

Amino acids abbreviations: A, alanine; C, cysteine; E, glutamate; F, phenylalanine; I, isoleucine; K, lysine; M, methionine; S, serine; T, threonine; V, valine; Q, glutamine.

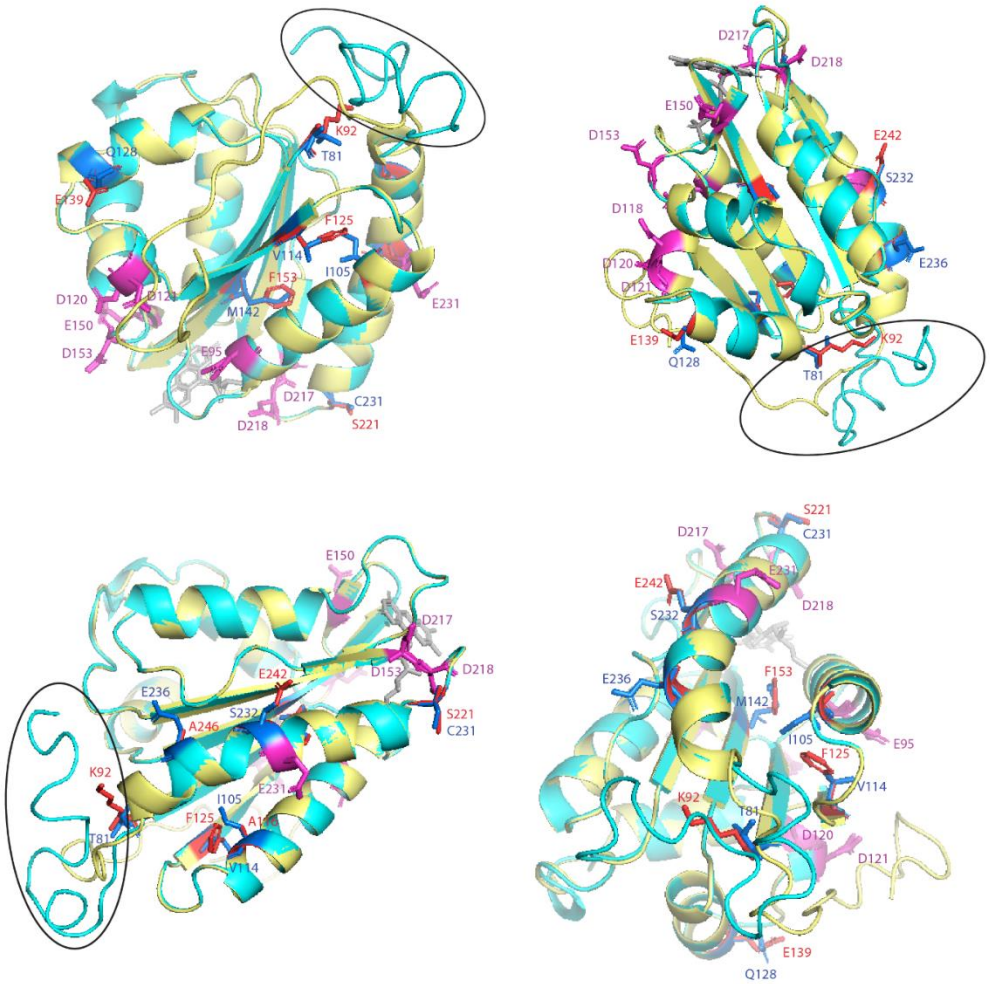

Figure 2-4. Structural analysis of key amino acid residues in the FMN-domain of MtCPRs model from four different angles.

FMN-domain of MtCPR1 is colored in yellow and MtCPR2 in cyan. Specific residues are depicted as stick with the following color code: magenta for acidic residues formerly indicated to be involved in CPR:CYP interactions, blue for key residue in MtCPR1, and red for key residue in MtCPR2. The FMN molecule is colored in grey. Black circle indicates the FMN-FAD hinge.

2.3.3 Gene co-expression analysis of legume CPRs

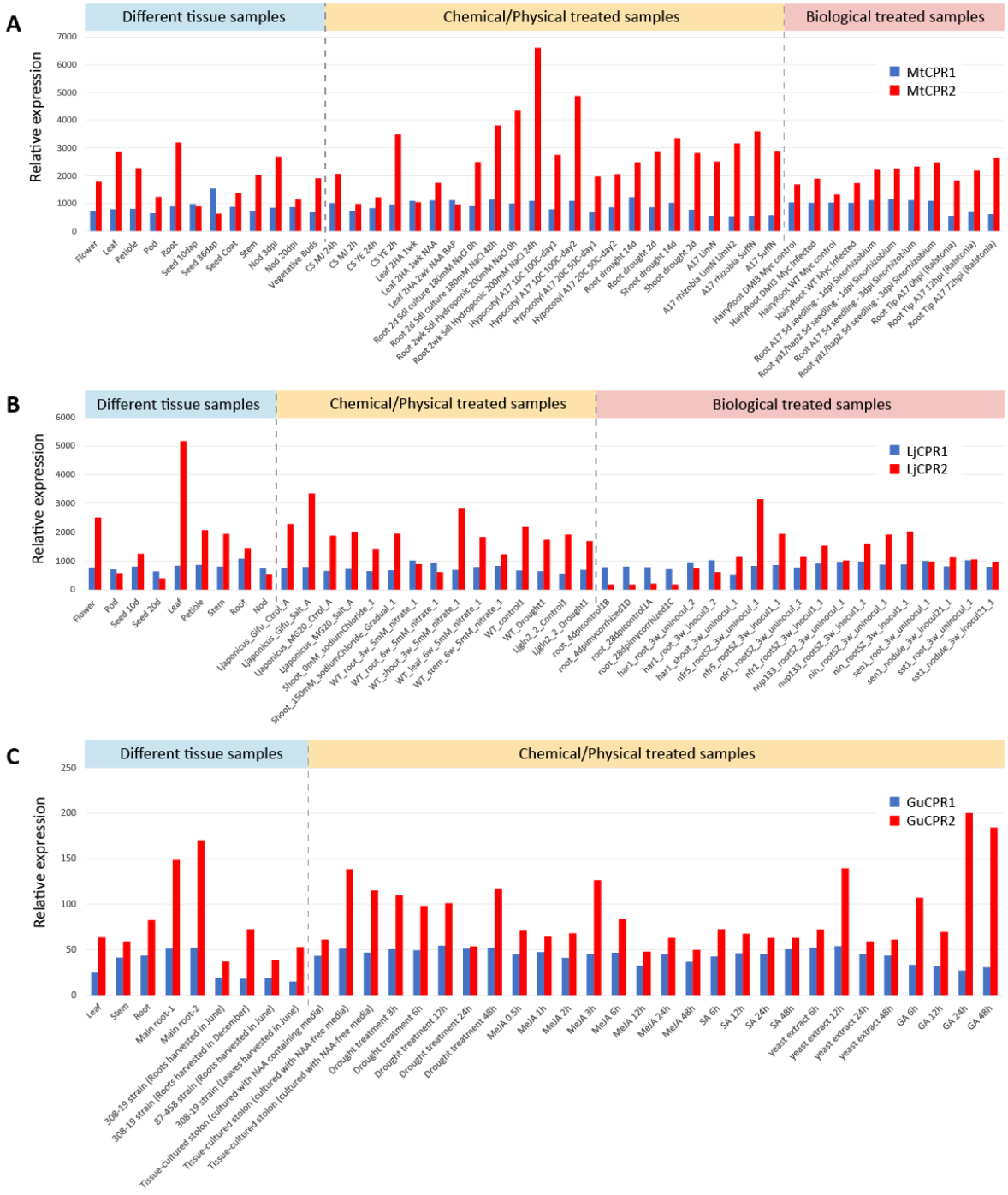

Figure 2-5. Co-expression analysis of CPR class I and II from (A) *M. truncatula*, (B) *L. japonicus*, and (C) *G. uralensis*.

This gene co-expression analysis was performed using selected transcriptomic data-sets representing samples from different tissue, chemical/physical treatments, and biological treatments. Transcriptomic data-sets of *M. truncatula*

and *L. japonicus* were obtained from https://lipm-browsers.toulouse.inrae.fr/pub/expressionAtlas/app/v2 and https://lotus.au.dk/ of Miyakojima MG20 v 3.0. The *G. uralensis* transcriptomic data-set was obtained using RNA-seq analysis obtained in Istiandari et al., 2021 submitted to the DNA Data Bank of Japan (DDBJ) Sequence Read Archive (DRA) under the accession numbers of DRA012266.

Table 2-2. Correlation strength[a] between different CPR classes and CYP families in A) *M. truncatula*, B) *L. japonicus*, and C) *G. uralensis*.

(A)

Gene	*PCC value*					
	bAS	***LUS***	***CYP716A12***	***CYP72A61***	***CYP72A63***	***CYP93E2***
MtCPR1	0.12	0.05	-0.14	-0.28	0.11	-0.30
MtCPR2	0.44	-0.11	0.02	0.18	0.13	0.09

(B)

Gene	*PCC value*				
	bAS	***LUS***	***CYP716A51***	***CYP72A61***	***CYP93E1***
LjCPR1	0.47	0.55	0.54	-0.16	-0.25
LjCPR2	0.34	-0.21	-0.36	0.30	0.33

(C)

Gene	*PCC value*					
	bAS	***LUS***	***CYP716A179***	***CYP72A154***	***CYP93E3***	***CYP88D6***
GuCPR1	-0.26	0.15	-0.23	-0.33	0.34	-0.48
GuCPR2	0.25	0.72	0.33	0.29	0.70	-0.08

[a] Correlation strength of different CPR classes with different CYP families are based on PCC. PCC values range from -1 to 1. Positive values were positively correlated, whereas negative values were negatively correlated. PCC, Pearson's correlation coefficient; Mt, *M. truncatula*; Lj, *L. japonicus*; Gu, *G. uralensis*

To analyze the gene expression level of different CPR classes in legumes, available transcriptomic data of *M. truncatula* was obtained from and *L. japonicus* were retrieved from the respective transcriptome database, while transcriptome data of *G. uralensis* were obtained in previous study (Istiandari et al. 2021). *M. truncatula* and *G. uralensis* transcriptomic database

comprised of 1559 and 37 number of samples, respectively. Using the most recent and high-quality *L. japonicus* Gifu v1.2 genome database constructed by 100x PacBio read coverage and RNA-seq analysis (Kamal et al. 2020), both *LjCPR2-1* and *LjCPR2-2* can be annotated in the transcriptomic data. However, since the Miyakojima MG20 v3.0 genome was based on a hybrid assembly of Sanger TAC/BAC, Sanger shotgun and Illumina shotgun sequencing data generated from the Miyakojima-MG-20 accession (Kamal et al. 2020), this database contained a large fraction of unanchored contigs and many pseudogenes. In Miyakojima MG20 v3.0 transcriptomic database annotated by microarrays (Kamal et al. 2020), only one *LjCPR2* gene was annotated. Due to high sequence similarity between *LjCPR2-1* and *LjCPR2-2*, the probe ID only identified them as one gene. In this chapter, transcriptomic data from Miyakojima MG20 v 3.0 genome version was used for *L. japonicus* gene co-expression analysis instead of Gifu v1.2 version, since it contains more data with 74 experiment samples, compared to only 35 samples in Gifu v1.2 version. For comparison, the result of *L. japonicus* gene co-expression analysis of using Gifu v1.2 version can be found in Table S3 and Figure S3.

Gene co-expression analysis of CPR class I and II of *M. truncatula*, *L. japonicus*, and *G. uralensis* showed distinct expression patterns (Figure 2-5). CPR class I were found to be constitutively expressed with lower and more stable expression levels. In contrast, CPR class II were generally found to have higher expression levels than CPR class I, which varied depending on tissues and treatments. Similarly observed in *L. japonicus* Gifu v1.2 transcriptomic database, *LjCPR2-1* and *LjCPR2-2* showed more inducible expression than *LjCPR1*. However, *LjCPR2-2* expression is even lower than *LjCPR1* in some of the samples, but seems to be tissue-specific, with the highest expression was observed in immature flower and pod (Figure S3). *LjCPR2-1*

expressions are generally higher in almost all samples compared to *LjCPR2-2* and *LjCPR1*, implying *LjCPR2-1* is the dominant LjCPR class II in *L. japonicus* (Figure S3).

Global analysis of the genes from each legume species was performed by calculating the PCC of each CPR class with each of the transcripts found in the assembly. CPR class I showed strong correlation (PCC >0.70) with genes involved in basal metabolism or molecular functions, such as ATP-binding/-carrier proteins. In contrast, CPR class II strongly correlated with genes such as ubiquitin, glutamate, and jasmonate receptor, which are known to play a role in plant defense responses (Guo et al. 2018; Qiu et al. 2020; Doroodian and Hua 2021) (Table S4). Interestingly, despite its critical role in providing electrons for CYPs, the global transcriptomic analysis of genes correlated with CPRs showed no correlation with CYPs, at a PCC >0.4.

To analyze the correlation strength between the different CPR classes and CYPs in each species, we calculated the PCC values of different CPR classes in combination with different OSCs, *bAS* and *LUS*, and CYP families 716A, 72A, and 93E from the three legume species, in addition to *CYP88D6* from *G. uralensis*. These CYPs were chosen since they are involved in the majority of triterpenoid compounds in legumes. All *bAS* and *LUS* from these legume species convert 2,3-oxidosqualene into triterpene backbone β-amyrin and lupeol, respectively. MtCYP716A12, LjCYP716A51, and GuCYP716A179 catalyze the same reaction of three steps of oxidation at the C-28 position of α-amyrin, β-amyrin, and lupeol to produce the final product of ursolic acid, oleanolic acid, and betulinic acid, respectively (Fukushima et al. 2011; Tamura et al. 2017; Suzuki et al. 2019) (Figure 2-1). Both MtCYP72A63 and GuCYP72A154 oxidize the C-30 position of β-amyrin to produce 30-OH β-amyrin, but only MtCYP72A63 catalyzes the further oxidation to produce 30-COOH β-amyrin (Seki et al. 2011). LjCYP72A61 and MtCYP72A61 catalyzes the one-step oxidation at the C-22 position of β-amyrin to produce sophoradiol (Ebizuka et al. 2011).

MtCYP93E2, LjCYP93E1, and GuCYP93E3 catalyze the same reaction of three steps of oxidation on the C-24 position of β-amyrin to produce the final product of 24-COOH β-amyrin (Seki et al. 2008; Fukushima et al. 2011; Suzuki et al. 2019). GuCYP88D6 catalyzes the two-step oxidation at the C-11 position of β-amyrin to produce 11-oxo β-amyrin (Seki et al. 2008).

The PCC table shows that different CPR classes had different correlation values with different CYPs (Table 2-2). The probeset ID used for PCC calculation is presented in Table S5. In *M. truncatula*, *MtbAS* showed higher correlation with *MtCPR2* compared to *MtCPR1*. Similar correlation value was observed for *MtCYP716A12* with both *MtCPR1* and *MtCPR2*, while *MtCYP72A61*, *MtCYP72A63*, and *MtCYP93E2* had a slightly higher positive value with *MtCPR2*, compared to *MtCPR1* (Table 2-2A). From the *L. japonicus* database MG20 v3.0, *LjCPR1* showed much higher correlation values with *LjCYP716A51* and *LjLUS*, and slightly higher *LjbAS* compared to *LjCPR2* (Table 2-2B). In contrast, *LjCYP93E1* and *LjCYP72A61* showed higher correlation with *LjCPR2* compared to *LjCPR1*. Based on the transcriptome data of *L. japonicus* in the Gifu version 2.0, *LjCPR1* also showed higher positive correlation value with *LjCYP716A51* and *LjLUS*, as compared to *LjCPR2* (Table S3). *GuCPR2* showed higher correlation with all genes: *GuLUS*, *GubAS*, *GuCYP716A179*, *GuGCYP72A154*, *GuCYP93E3*, and *GuCYP88D6* as compared to *GuCPR1*. *GuCPR1* showed very significant negative correlation with *GuCYP88D6* (Table 2-2C).

2.4 Conclusion

There was a significant difference observed in the protein structure of legume CPR class I and II which is located in close proximity with CYP:CPR binding sites. This information suggested that each CPR class might exhibit specific binding conformation with different CYP families. The difference of CPR class I and II protein structure might also responsible for their physiological

roles in plants, which is shown different expression pattern of CPR class I and II of *M. truncatula*, *L. japonicus*, and *G. uralensis*. CPR class I showed lower and stable expression level compared to the inducible and higher expression levels of CPR class II in almost all samples. In addition, based on global co-expression analysis, CPR class I showed higher correlation with genes responsible for basal growth and primary metabolisms, while CPR class II showed higher correlation with genes responsible for defense and stress-related mechanisms. Moreover, CPR class in each plant species also showed different correlation strength with different CYP families which was previously speculated by structural analysis. Therefore, elucidating the actual role of each CPR classes *in planta* is crucial to understand CPR:CYP relationship towards triterpenoids biosynthesis.

Chapter 3

Functional analysis of *L. japonicus* CPR classes *in planta*

3.1 Introduction

L. japonicus, as one of the model legumes, is known to accumulate diverse triterpenoid saponins (Suzuki et al. 2019). The availability of *L. japonicus* genome database, mutant library, and the establishment of its hairy root transformation makes this plant an excellent platform for studying triterpenoids biosynthesis and its regulatory mechanisms, including CPRs (Małolepszy et al. 2016; Urbański et al. 2013; Okamoto et al. 2013). Many triterpenoids biosynthetic genes in *L. japonicus* also have been characterized. At least five *L. japonicus* OSCs have been identified, among which *LjbAS*, *LjAaS*, and *LjLUS*, and three LjCYPs involved in triterpenoids biosynthesis, *LjCYP716A51*, *LjCYP72A61*, and *LjCYP93E1* with functions have been described before (Suzuki et al. 2019; Figure 2-1). Based on previous study, *GmLUS* and *GmCYP716As* genes that were responsible for lupeol and betulinic acid production, respectively, were upregulated in secondary aerenchyma of soybean plants under flooded condition (Nakazono 2014). Root was also known to accumulate the highest amount of total triterpenoids in *L. japonicus* plants, compared to other tissues (Suzuki et al. 2019). Thus, root part of *L. japonicus* provides the best choice to study triterpenoids biosynthesis.

Based on previous chapter, we know that legume CPR class I and II showed different tissue expression profile and CPR class I is generally more stable and has lower expression than CPR class II. However, it is also believed that class I and II CPR play more critical role in primary and secondary metabolism, respectively. The significantly low ratio of CPR:CYP in mouse liver

implied a competition between a vast number of different CYPs over a small number of CPR, in which a systematic regulation is required to perform their functions properly. Knocking down *CPR2* gene in *Catharantus roseus* significantly reduced the total monoterpene indole alkaloid content, while knocking down *CPR1* did not show any change (Parage et al. 2016). *ATR2* mutation decreased the electron transfer to three CYPs involved in lignin-related phenolic metabolites, *C4H*, *C3H1*, and F5H1, but did not much affect other CYPs involved in glucosinolate and flavonol glycoside biosynthesis in *Arabidopsis thaliana* (Sundin et al. 2014).

Transcript and metabolite profiling of stress/elicitor-treated plants or cell cultures is a powerful approach to determine gene function in secondary metabolism (Misra et al. 2014). CPR class II genes from *Withania somnifera*, *Panax ginseng*, *Camellia sinensis*, and *Catharantus roseus* were found to be highly induced by methyl jasmonate (MeJA) treatment, whereas their CPR class I genes showed less or no induction (Parage et al. 2016; Rana et al. 2013; Huang et al. 2021). Interestingly, this differential effect of phytohormone treatment was also observed in OSCs and CYPs. RT-PCR analysis of MeJA-treated *Ocimum basilicum* showed that *ObbAS1* and *ObCYP2* were significantly and continually induced until 12 h of treatment, while the phytohormone effect was not so apparent on *ObbAS2*, *ObCYP1*, and *ObCYP3* (Misra et al. 2014). These results implied that there could also be a differential regulation of plant CPR class in triterpenoids biosynthesis of *L. japonicus* upon MeJA elicitaton. Thus, elucidation of differential function of LjCPR class I and II gene *in planta* is very crucial in understanding their involvement in triterpenoids biosynthesis.

3.2 Materials and method

3.2.1 Plant materials and germination treatment

L. japonicus Gifu B-129 wild-type (WT) and *LORE1* insertion lines (Fukai et al. 2012; Urbański et al. 2012) were provided by Miyazaki University, Japan and Aarhus University, Denmark, through the National BioResource Project (NBRP). Seeds of *L. japonicus* were surface-sterilized using 2% (v/v) sodium hypochlorite and 0.02% (v/v) Tween 20 for 15 min in seesaw shaker, rinsed three times by ultrapure water obtained from a Milli-Q Synthesis system (Millipore), and placed onto a 0.8% agar plate. The seeds were allowed to germinate at 23°C for 4 d under dark conditions and for 2 d under 16 h light conditions.

3.2.2 Chemicals

β-Amyrin, α-amyrin, lupeol, erythrodiol, uvaol, oleanolic acid, ursolic acid and asiatic acid were purchased from Extrasynthese (France). Betulin and methyl jasmonate (MeJA) was purchased from Sigma-Aldrich (USA). Betulinic acid was purchased from Tokyo Chemical Industry (Japan). Soyasapogenol B and soyasapogenol A were purchased from Tokiwa Phytochemical (Japan). Sophoradiol, 24-hydroxy-β-amyrin, and soyasapogenol E were kindly provided by Dr. Kiyoshi Ohyama (Tokyo Institute of Technology, Japan).

3.2.3 Hairy root induction

Induction of hairy roots was performed as reported previously (Suzuki et al. 2019), with slight modifications. *A. rhizogenes* ATCC15834 strains were cultured on YEB plates for 2 d and suspended in sterilized water. The roots of 7-day-old WT seedlings were cutoff, and *A. rhizogenes* was infected into the cross-sections of hypocotyls. After co-cultivation for 4 d, the infected

seedlings were cultured on cefotaxime-containing hairy root elongation (HRE) solid medium for 2 weeks in 16 h light condition. After dissection, hairy roots were cultured under dark conditions. Root tip of a randomly chosen healthy wild type hairy root clone was subcultured in 5 ml cefotaxime-containing HRE liquid medium for 2 weeks and then transferred to 5 ml of HRE liquid medium without antibiotics 90 rpm. Isolated hairy roots were cultured for 2 months at room temperature with subculturing every 3–4 weeks. Finally, hairy roots were cultured in 100 ml of HRE liquid medium at 25°C with shaking at 90 rpm for 4 weeks.

3.2.4 Methyl jasmonate preparation and addition

MeJA elicitation was carried out to test its effect on triterpenoid biosynthesis in *L. japonicus* hairy roots. MeJA preparation was performed as reported previously (Akhgari et al. 2019). MeJA stock solution was made by dissolving it in 40% (v/v) ethanol to achieve a concentration of 20 mM and then filter sterilization (0.22 µm). The 4-week-culture of wild-type hairy root was cut into similar portions (roughly 200 mg fresh weight each) and cultured separately into same 100 ml of HRE liquid medium without antibiotic at 25°C with shaking at 90 rpm for another 4 weeks. Four weeks old hairy root cultures were supplemented with final concentrations of 100 µM MeJA. For control cultures equal volumes (500 µl) 40% ethanol was applied into the 100 ml culture medium. The hairy roots were incubated under the same conditions as mentioned above and collected after 0, 3, 6, and 12 h for gene expression analysis, and 0, 12, 24, and 24 h for metabolite analysis. The samples were flash-frozen in liquid nitrogen and stored at -80°C until used.

3.2.5 Quantitative real time PCR

Total RNA was extracted from 100 mg of 4-week-old frozen *L. japonicus* MeJA-treated and control hairy root using RNeasy Plant Mini Kit (Qiagen, USA). The RNA obtained was purified

using the After Tri-Reagent RNA Clean-Up Kit (Favorgen Biotech Corp., Taiwan) after digesting contaminated genomic DNA with recombinant DNase I (RNase-free) (Takara Bio, Japan). First-strand cDNA was synthesized from purified total RNA by PrimeScript RT Master Mix (Perfect Real Time) (Takara Bio, Japan). We performed qPCR analysis with The LightCycler® 96 (Roche, Germany) and FastStart Essential DNA Green Master (Roche, Germany). The primers used for qPCR analysis are listed in Table S6. The expression of the Ubiquitin (UBQ) gene was analyzed as a reference gene following Delis et al. (2011).

3.2.6 *Ljcpr1* and *Ljcpr2* loss-of-function mutant lines

Two independent lines of *Ljcpr1* (30003941 and 30059903) and *Ljcpr2-1* (30037476 and 30065390) homozygous insertion mutants were chosen from *LORE1* mutant library (lotus.au.dk). Each mutant line contains other exonic or intronic *LORE1* insertions other than *LjCPR* genes (Table S7). These chosen mutant lines were cultivated on soil and their progenies were cultivated in hydroponic system. For soil-cultured plants, the 7-day-old WT and mutant seedlings were moved to pot of the mixture of soil and vermiculite, and cultivated for 3 months. The produced seeds were then collected for hydroponic cultivation. The seed pods were counted from three independent lines for each *Ljcpr1* and *Ljcpr2-1* homozygous insertion mutants. Then, pod length was measured using digital vernier caliper from randomly selected 26 pods from each *Ljcpr1* and *Ljcpr2-1* homozygous insertion mutant lines. The 7-day-old WT and mutant seedlings of the soil-cultured plants progeny were firstly cultured in 5 ml tubes containing basal nutrient solution. After 2 weeks, the hydroponic plants were scaled-up into 50 ml tubes containing basal nutrient solution and cultivated until flowering stage. The hydroponic medium was renewed weekly.

Genomic DNA was extracted using FavorPrep™ Plant Genomic DNA Extraction Mini Kit (Favorgen Biotech Corp., Taiwan) from their leaves to screen for homozygous *Ljcpr1* and *Ljcpr2-*

1 mutants by PCR using each CPR-specific primers and a *LORE1*-specific primer (Table S6; Figure S4). PCR was performed using KOD FX Neo following the manufacturer's instructions (Toyobo, Japan). Triterpenoids were extracted from the soil- and hydroponic-cultured roots of homozygous and heterozygous mutants and were analyzed as described below.

3.2.6 *Ljcpr1* knockout mutant hairy root lines

The multiplex guide RNA (gRNA)-expressing CRISPR-Cas9 vector, pMgP237-2A-GFP (Nakayasu et al. 2018; Hashimoto et al. 2018), was used for genome editing of *L. japonicus* hairy root. The target sequences of the gRNAs (Figure 3-8) were selected from *LjCPR1* gene using the web-based tool, CRISPRdirect (https://crispr.dbcls.jp/) (Naito et al. 2015). Two gRNA target sequences were simultaneously transferred into the pMgP237-2A-GFP vector as described previously (Nakayasu et al. 2018), generating the T1/T2-pMgP237 vector. Three sets of different gRNA designs (No. 2, 4, and 5) targeting *LjCPR1* gene were constructed using primers listed in Table S6. *A. rhizogenes* ATCC15834 was transformed with the pMgP237 empty vector or the T1/T2-pMgP237 vectors.

Induction of hairy root was described as above. Crude genomic DNA extraction and PCR were performed using KOD FX Neo following the manufacturer's instructions (Toyobo, Japan). Mutagenesis was confirmed by PCR with specific primers for each gRNA designs (Table S6; Figure 3-9), and analyzed by Heteroduplex Mobility Assay (HMA) by using an MCE-202 MultiNA microchip electrophoresis system (Shimadzu, Japan) following the manufacturer's instructions. The target sequences amplified from putative mutants were cloned into the pJET1.2/blunt vector (CloneJET PCR Cloning Kit; Thermo Fisher Scientific, USA). Insertion and deletion mutations were confirmed by sequencing of several randomly selected clones.

3.2.7 Extraction of triterpenoids from *L. japonicus* plants and hairy roots

Triterpenoid extraction was performed as reported previously (Suzuki et al. 2019), with slight modifications. Plants at the flowering stage and hairy roots were lyophilized and powdered using a multibead shocker (Yasui Kikai, Japan). Powdered tissues (20.00 ± 0.3 mg) were extracted three times with 1 ml of methanol by a sonication-assisted method. Completely dry extracts were resuspended in 2 ml of MeOH : 4 M HCl (1:1). The extracts were incubated at 80°C for 1 h to remove the sugar moieties of triterpenoid saponins. The hydrolyzed products were extracted three times with hexane : EtOAc (1:1) and dried completely. The obtained pellet was resuspended in 500 µl of MeOH: chloroform (1:1). A portion of the solution was dried in a GC-MS vial. One hundred microliters of the solution were evaporated and trimethylsilylated using a mixture of 50 µl of N,N-dimethylformamide (Kishida Chemical Co., Ltd., Japan) and 50 µl of BSTFA:TMCS (99:1) (TCI) at 80°C for 30 min. For semi-quantitative analysis, an asiatic acid authentic standard was applied to the plant tissue powder before extraction.

3.2.8 GC-MS analysis

GC-MS analyses were performed as reported previously (Suzuki et al. 2019) on a 5977A MSD mass spectrometer (Agilent Technologies, USA) connected to a 7890B gas chromatograph (Agilent Technologies) with an HP-5MS UI (30 m × 0.25 mm, 0.25 µm film thickness; Agilent Technologies) capillary column for qualitative analysis. The injection temperature was set at 250°C. The column temperature program was as follows: 80°C for 1 min, an increase to 300°C at a rate of 20 °C/min, and hold for 28 min. The carrier gas was helium at a flow rate of 1.0 ml/min. The ion source temperature was 230°C and the quadrupole temperature was 150°C. One microliter of the derivatized sample was injected in splitless injection mode. Peaks were identified by

comparing their Rt and mass spectra with those of authentic standards. Samples were analyzed in SIM mode for relative quantification by extracting the mass chromatogram in respective EIC for each metabolite as listed in Table S8.

3.3 Result and discussion

3.3.1 Methyl jasmonate treatment on *L. japonicus* hairy roots

To elucidate the effect of phytohormone elicitation on *L. japonicus* triterpenoids biosynthetic genes, 1-month-old *L. japonicus* hairy roots were added with 100 μM MeJA and sampled at different time points. The qRT-PCR result of extracted RNA from these treated hairy roots showed different regulation on some of triterpenoids biosynthetic genes (Figure 3-1). While *LjCPR1* gene was quickly down-regulated by MeJA addition, *LjCPR2-1* and *LjCPR2-2* expressions were significantly upregulated up to 4 times compared to mock sample in only 3 h after MeJA addition. These results clearly showed different regulatory mechanisms of CPR class I and II in *L. japonicus* hairy root upon MeJA addition. Interestingly, similar to *LjCPR1* expression pattern, *CYP716A51* and *LUS* gene also showed down-regulation by MeJA addition even after 12 h of treatment. On the other hand, very high and quick upregulation was observed in *bAS*, *CYP93E1*, and *CYP72A61* expression levels. The expression of *bAS* was upregulated more than 20 times compared to control after 12 h of treatment, while *CYP72A61* and *CYP93E1* upregulation was reached the highest after 6 h of treatment with around 5 times higher compared to control. Another triterpene OSC genes, *aAS*, was not able to be detected on all samples. To observe the effect of MeJA on the phytosterol biosynthetic pathway, *CAS* and *LAS* expression were also analyzed. *CAS* and *LAS* are cycloartenol and lanosterol synthase, respectively, which represent the branch-off entry of phytosterol biosynthesis after 2,3-oxidosqualene cyclization (Figure 1-1). Both *CAS* and *LAS* expression were

instantly downregulated after 3 h of MeJA addition, but then the expression were increased after 6 h and returned to be similar to control level after 12 h of treatment.

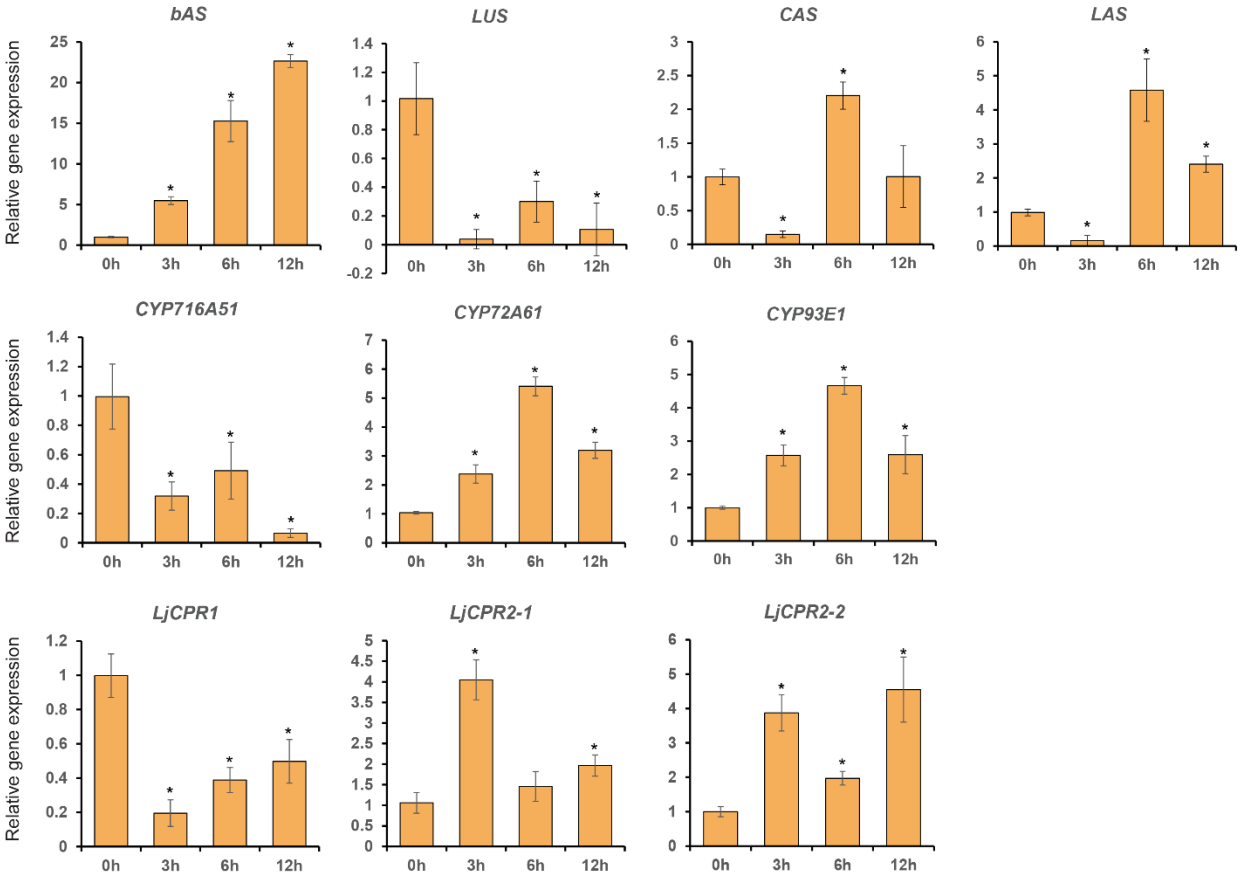

Figure 3-1. The relative expression of triterpenoids biosynthetic genes of *L. japonicus* hairy roots treated with methyl jasmonate (MeJA) on different time period.

Transcript levels of *LjCPR1*, *LjCPR2-1*, *LjCPR2-2*, *bAS*, *LUS*, *CAS*, *LAS*, *CYP716A51*, *CYP72A61*, and *CYP93E1* were analyzed by qRT-PCR in *L. japonicus* hairy roots treated with 100 µM MeJA for 0, 3, 6, and 9 h of treatment. Relative expression levels were normalized to that of Ubiquitin and are presented as fold induction relative to the control. Data represent the mean of three independent replicates ± SD. Statistical significance in comparison with the corresponding control values is indicated by Student's t-test; * $p < 0.05$. qRT-PCR, quantitative reverse transcription PCR; SD, standard deviation.

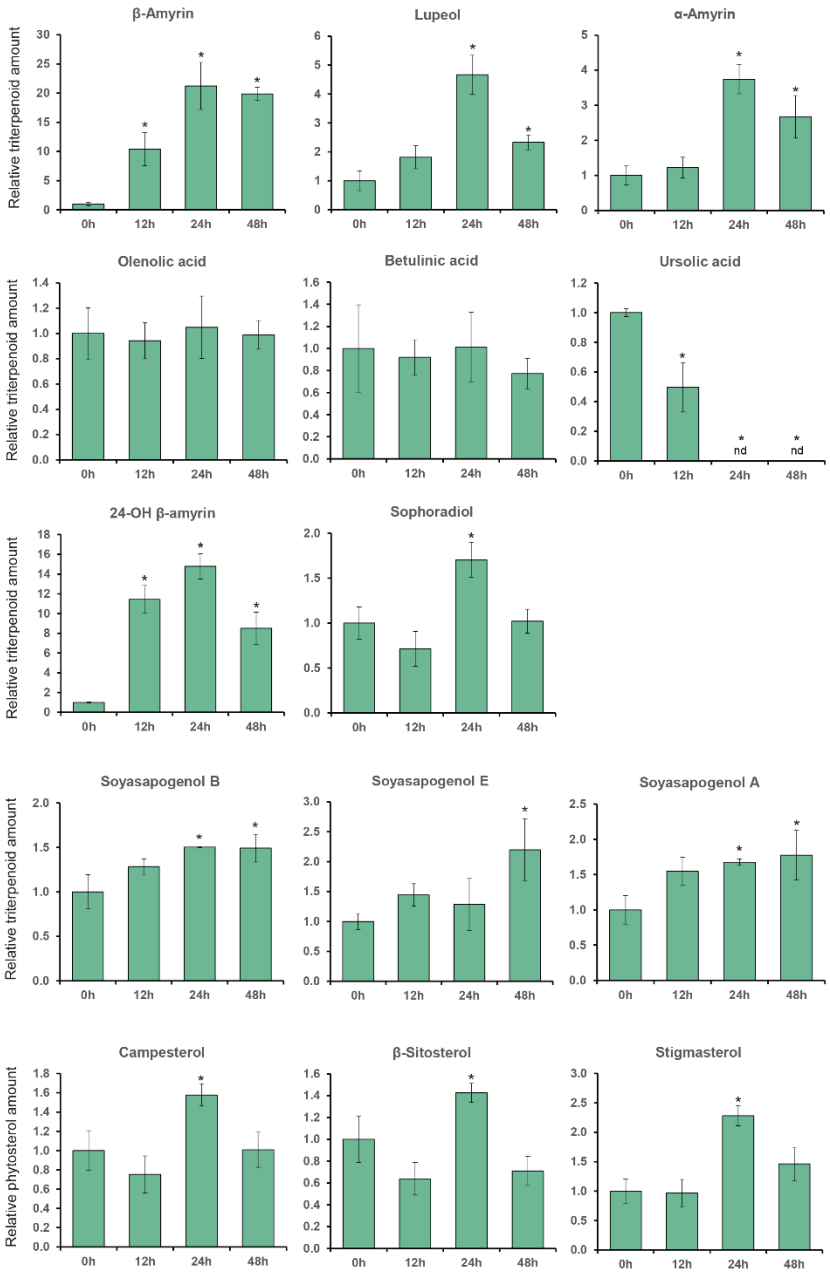

Figure 3-2. The relative amount of triterpenoids and phytosterol content of *L. japonicus* hairy roots treated with methyl jasmonate (MeJA) on different time period.

Production levels of α-amyrin, β-amyrin, lupeol, ursolic acid, oleanolic acid, betulinic acid, sophoradiol, 24-OH β-amyrin, soyasapogenols, and phytosterols were analyzed by GC-MS in *L. japonicus* hairy roots treated with 100 μM MeJA for 0, 12, 24, and 48 h of treatment. Relative triterpenoids and phytosterol amount were normalized to that of asiatic acid as internal standard and are presented as fold induction relative to the control. Data represent the mean of three biological replicates ± SD. Statistical significance in comparison with the corresponding control values is indicated by Student's t-test; * $p < 0.05$. SD, standard deviation.

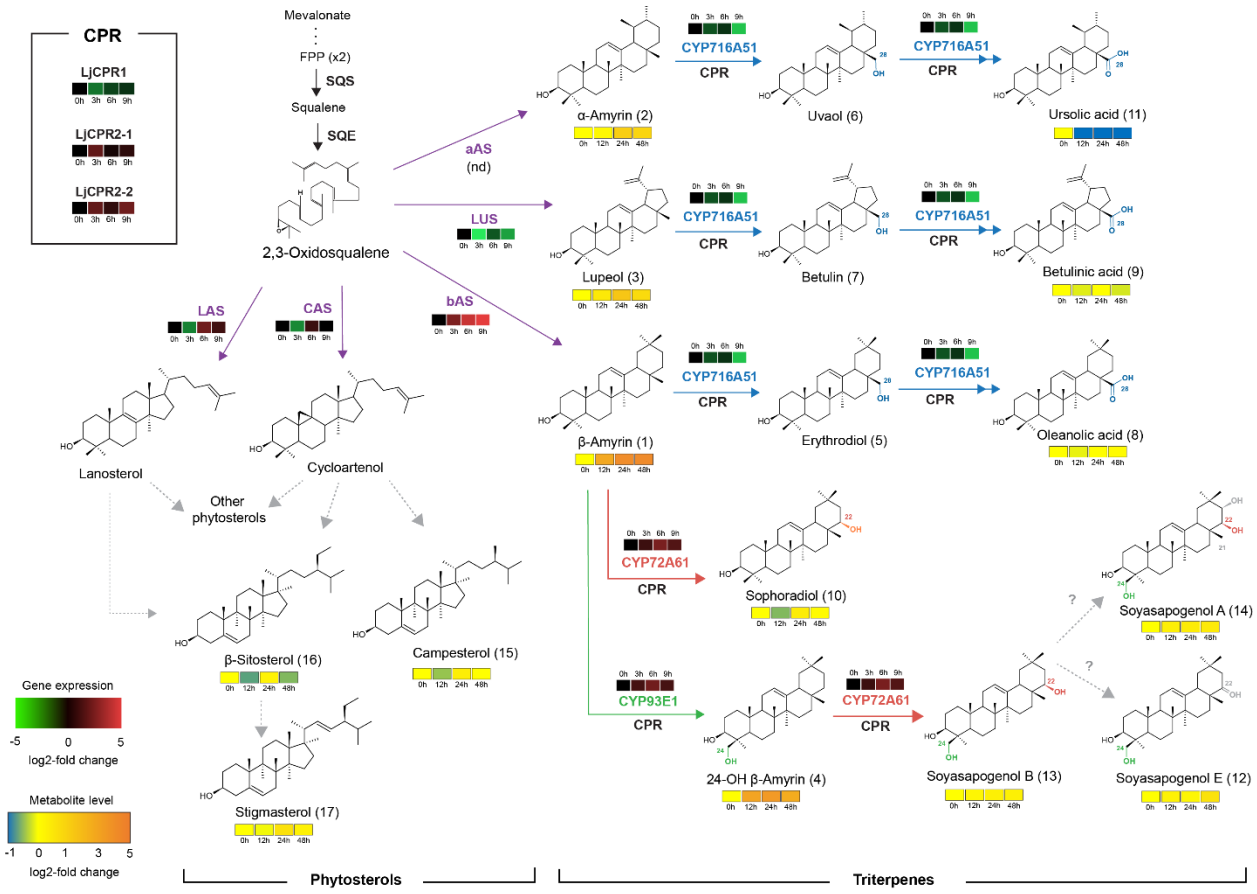

Figure 3-3. Effect of methyl jasmonate treatment on gene expression and metabolite level on triterpenoids and phytosterol biosynthesis in *L. japonicus* hairy root.

Single and double arrows indicate one and two oxidation steps, respectively. Dashed arrows indicate multiple steps. CYP, cytochrome P450; FPP, farnesyl pyrophosphate; SQS, squalene synthase; SQE, squalene epoxidase; bAS, β-amyrin synthase; aAS, α-amyrin synthase; LUS, lupeol synthase; LAS, lanosterol synthase; CAS, cycloartenol synthase.

Based on GC-MS analysis, the change in expression level of triterpenoids biosynthetic genes due to MeJA treatment affected the triterpenoids production in *L. japonicus* hairy roots (Figure 3-2). All triterpenoids' peaks were annotated and the chromatogram area was measured by comparing the retention time and mass spectrum with their authentic standards. Triterpenoids that were analyzed are the major triterpenoid constituents in *L. japonicus* (Suzuki et al. 2019), which served as the representative for triterpenoids in this study. The levels of other minor triterpenoids

were too low to be detected by GC-MS. In correlation with the increase in *bAS* and *CYP93E1* expression, a significant increase was shown in β-amyrin and 24-OH β-amyrin levels. The β-amyrin level was increased 10 times after 12 h of treatment and continued to increase until reached 20 times higher than control even after 48 h of treatment. Similar to β-amyrin, 24-OH β-amyrin level was also increased rapidly after 12 h into 10 times higher than control, and reached maximum after 24 h of treatment up to 15 times increase compared to control. The increase in *CYP72A61* expression level also increased the sophoradiol production level after 24 h of treatment. In addition, soyasapogenols showed gradually increasing production levels even after 48 h of treatment. As expected, correlated with the downregulation of *CYP716A51* expression due to MeJA addition, no change was observed in oleanolic acid and betulinic acid level, and even significantly downregulated ursolic acid levels. However, even though *LUS* showed downregulated expression, lupeol level showed a significant increase after 24 h up to 4 times compared to control, similar to α-amyrin production. The increased level was possibly due to the accumulation of unconverted lupeol or α-amyrin into betulinic acid and ursolic acid, respectively, as the *CYP716A51* expression were very low. To observe the effect of MeJA on phytosterol pathway, campesterol, β-sitosterol, and stigmasterol as the three major sterols in plant were analyzed. Due to the lack of standard compounds, the physterol peak was annotated based on similarity with mass spectrum of NIST library (Figure S7). All phytosterols showed only a small increase after 24 h, and it was returned to control level after 48 h.

The MeJA treatment on *L. japonicus* hairy roots revealed a very interesting and clear segregation between LjCPR class I and II regulatory mechanisms. Based on the similarity of regulation pattern of each gene, the result suggested that *LjCPR1*, *CYP716A51*, and *LUS* showed similar regulatory mechanisms, and they showed different regulatory mechanisms with *LjCPR2*s,

bAS, *CYP93E1*, and *CYP72A61* regarding MeJA elicitation (Figure 3-3). Interestingly, these different groups of triterpenoids biosynthetic pathway and CPR class were supported by gene co-expression analysis that showed *LjCPR1* has stronger correlation value with *CYP716A51* compared to *LjCPR2* (Table 2-2B; Table S3), while *LjCPR2* has stronger correlation value with *bAS*, *CYP93E2*, and *CYP72A61* compared to *LjCPR1* (Table 2-2B). This similar preference or correlation of different LjCPR classes towards different CYPs showed in gene co-expression analysis and qRT-PCR analysis of MeJA treated hairy root should not have been a coincidence, which suggested that different CPR classes might have specific regulatory mechanisms with different CYPs involved in triterpenoids biosynthesis.

3.3.2 *Ljcpr1* and *Ljcpr2-1* loss-of-function mutant plants

To investigate the effect of loss-of-function of either *LjCPR1* or *LjCPR2* genes, the triterpenoids profile of *LORE1* insertion mutant lines (Fukai et al. 2012; Urbański et al. 2012) were analyzed. Some of triterpenoids biosynthetic genes in soybean are regulated differently under flooded condition (Nakazono 2014). Therefore, to investigate if the effect different plant culture condition also affects LjCPRs involvement with specific CYPs, the *Ljcpr1* and *Ljcpr2-1* loss-of-function mutant lines were both cultured in soil and hydroponic system. The seeds of two independent homozygous mutant lines, 30003941 (L1-A) and 30059903 (L1-B) which contain a single *LORE1* insertion into the first and fourth exon of *LjCPR1* were obtained (Figure 3-4). Other seeds of two independent homozygous mutant lines 30037476 (L2-1A) and 30065390 (L2-1B) which contain a non-single insertion *LORE1* insertion into the first and fourth exon of *LjCPR2-1* were obtained (Figure 3-4).

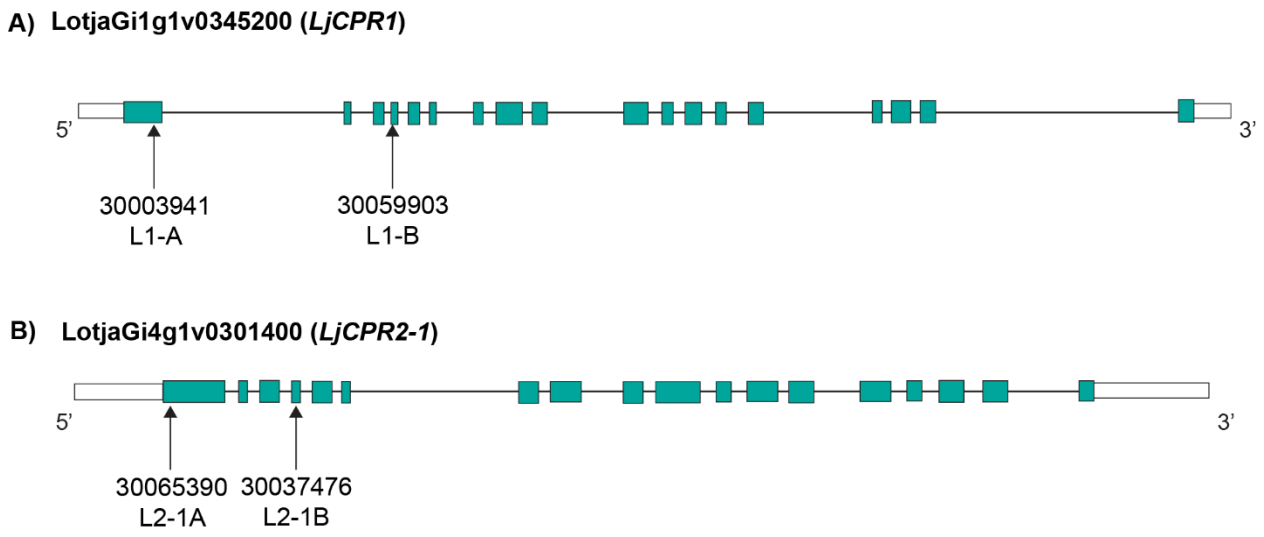

Figure 3-4. *LORE1* insertion map into *LjCPR1* and *LjCPR2-1* genes

While both *Ljcpr1* mutants are single insertion homozygous mutants (Figure S4-B), PCR genotyping result showed that there were heterozygous insertions on other genes other than *LjCPR2-1* in these mutant lines (Figure S4-C). Therefore, the effect of other gene mutations on both *Ljcpr2-1* mutant lines should not be ruled out in this study. Other seeds of two independent mutant lines which contain an *LORE1* insertion on *LjCPR2-2* exon were also screened, however no homozygous mutant was obtained. Therefore, only *Ljcpr1* and *Ljcpr2-1* loss-of-function mutants will be analyzed in this study. *Ljcpr2-1* mutant lines might serve as a representative mutant line for LjCPR class II since it has similar expression pattern with *LjCPR2-2* (Figure S3). Some WT allele-specific primer sets (F and R) and insertion allele-specific primer sets (F and P2) were designed for the *LjCPR1* and *LjCPR2-1* gene and other genes that may have *LORE1* insertions in their exons (Figure S3; Table S7) according to a previous report (Urbański et al. 2012).

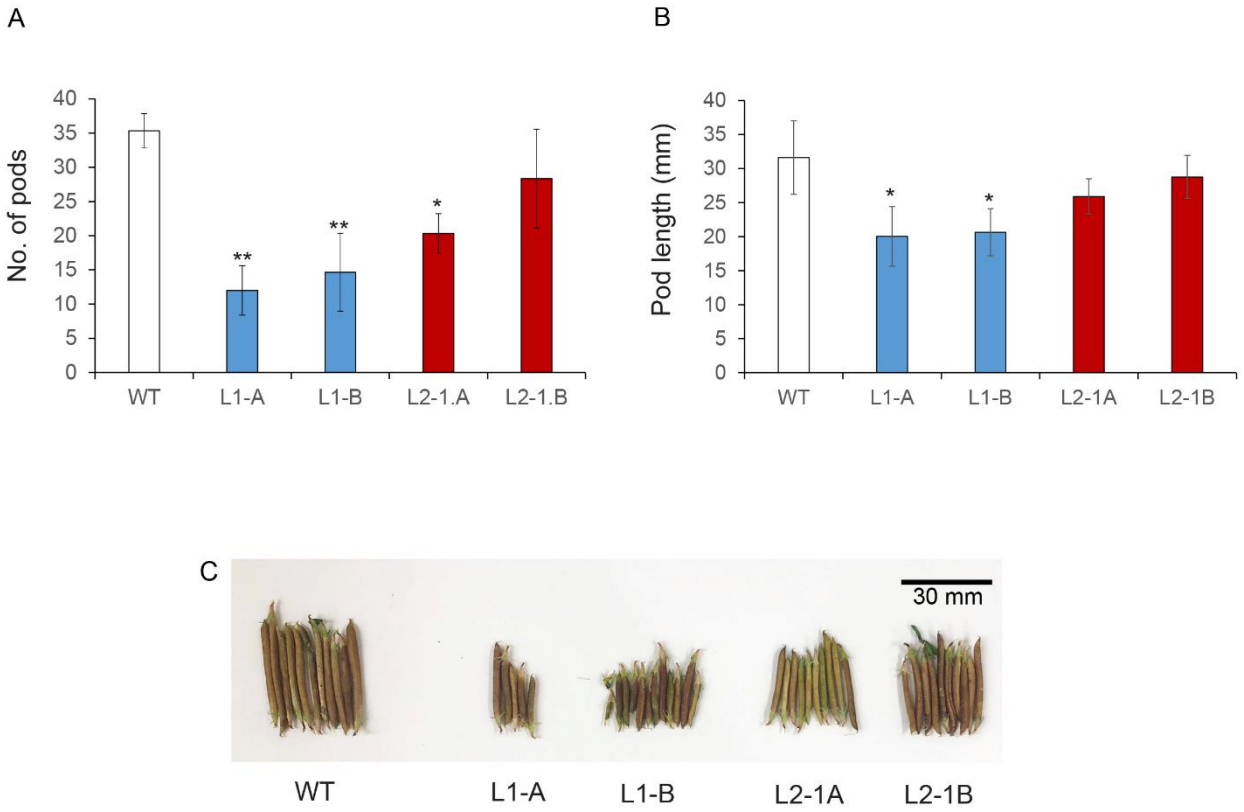

Figure 3-5. A) Pod count and B) pod length of *Ljcpr1* and *Ljcpr2-1* homozygous *LORE1* insertion mutants. C) The photo of representative mutant pods was shown in this figure.

Data represent the mean of three biological replicates in pod count (A) and *N*=26 randomly selected pods from each mutant plants in pod length measurement (B), both are ± SD. Single-factor ANOVA with Tukey's post-hoc test was used for statistical comparisons. Values were considered statistically significant at *P<0.05 and **P<0.01. SD, standard deviation.

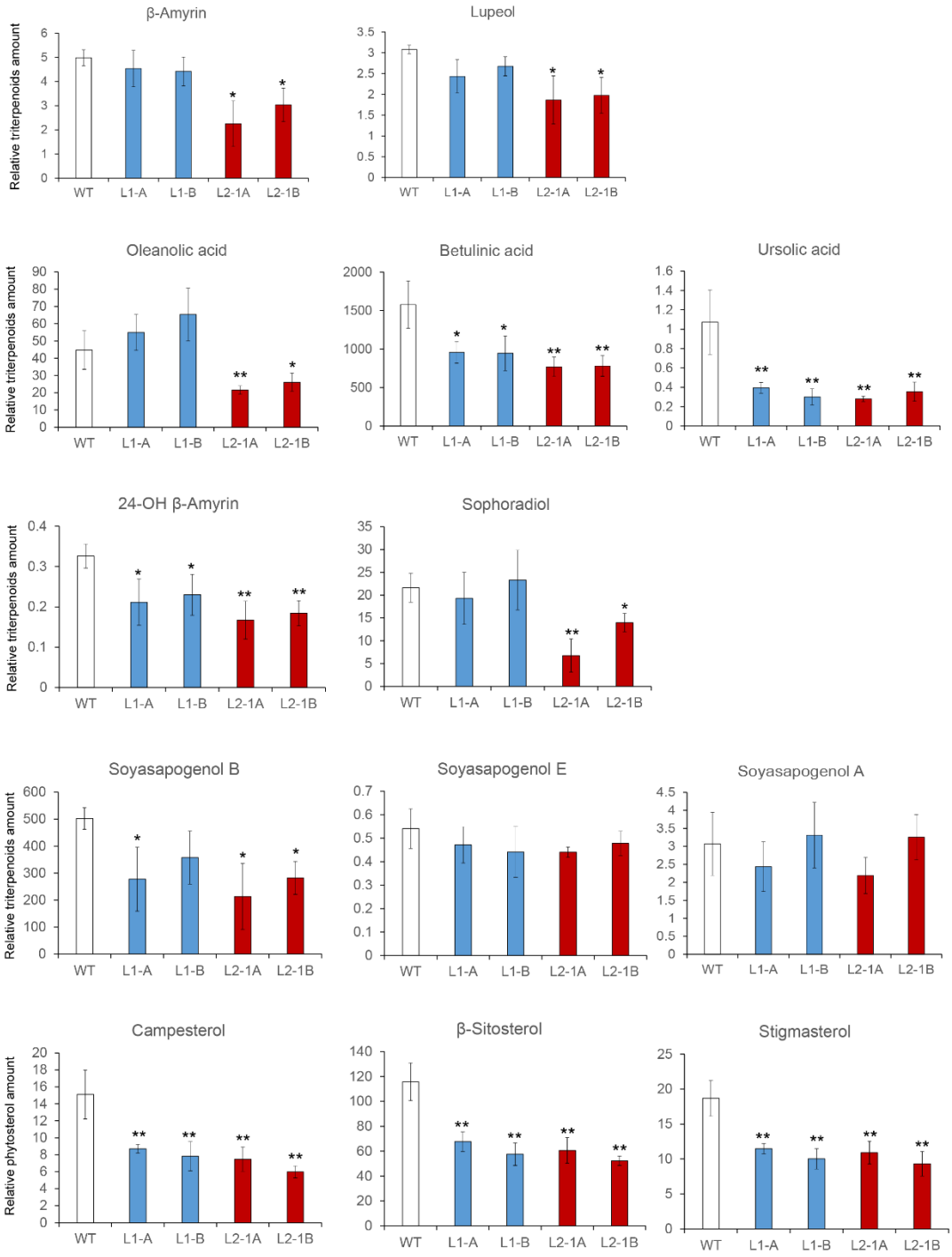

Figure 3-6. The relative amount of triterpenoids and phytosterol content of soil-cultured *L. japonicus LORE1* insertion mutant roots analyzed by GC-MS.

Relative triterpenoids and phytosterol amount were normalized to that of asiatic acid as internal standard and are presented as fold induction relative to the control. Data represent the mean of three biological replicates ± SD. Statistical significance in comparison with the corresponding control values is indicated by Student's t-test; * $p < 0.05$. SD, standard deviation.

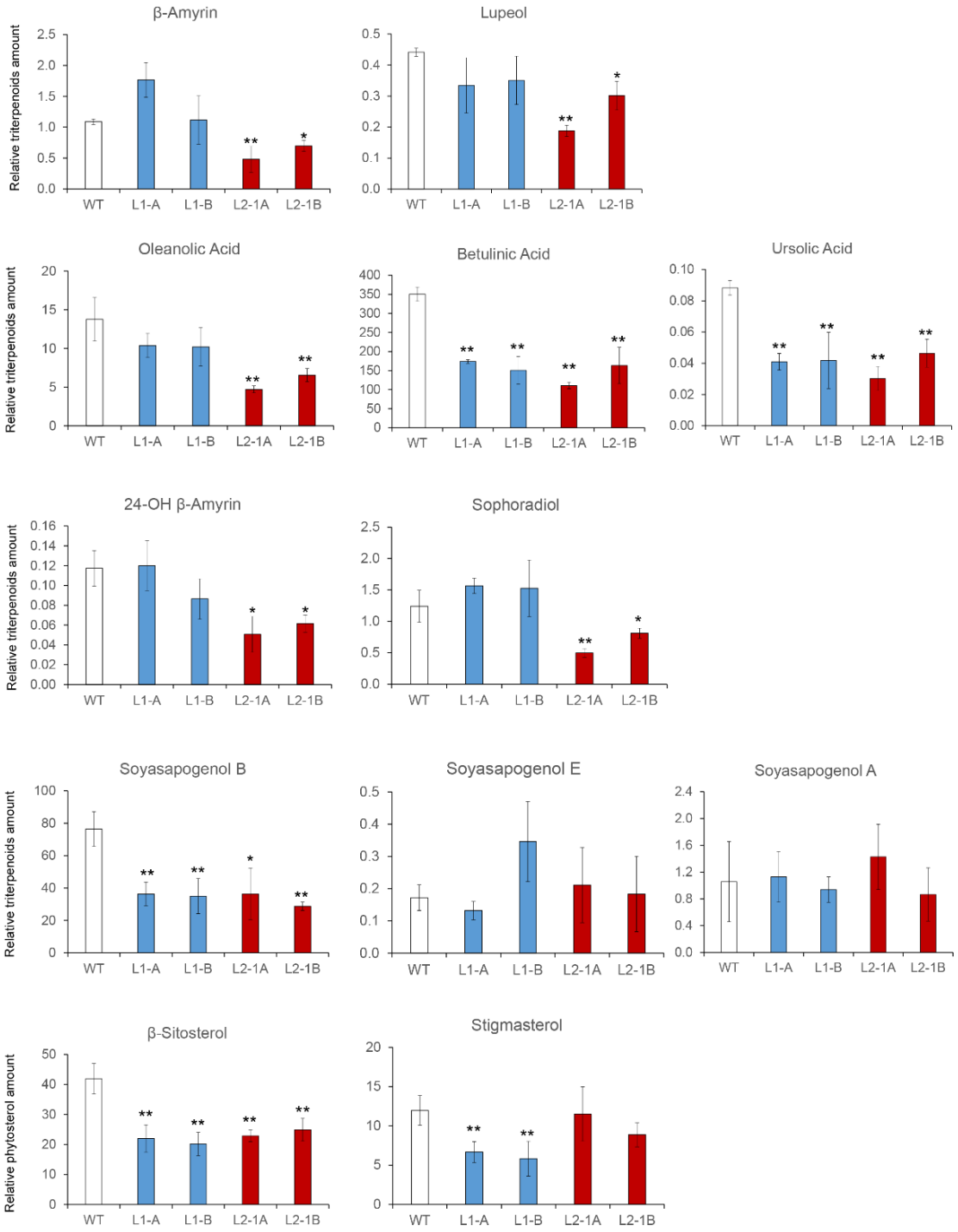

Figure 3-7. The relative amount of triterpenoids and phytosterol content of hydroponic-cultured *L. japonicus LORE1* insertion mutant roots analyzed by GC-MS.

Relative triterpenoids and phytosterol amount were normalized to that of asiatic acid as internal standard and are presented as fold induction relative to the control. Data represent the mean of three biological replicates ± SD. Statistical significance in comparison with the corresponding control values is indicated by Student's t-test; * $p < 0.05$. SD, standard deviation.

There was no significant difference on the physiology of the *Ljcpr* mutants could be observed on the plant leaf, stem, or root. However, a notable change can be observed on the seed pods. The collected seed pods from soil-cultured mutants were then measured and counted to observe the effect of *Ljcpr* loss on the seed physiology. Number of pods and pod length of two homozygous *Ljcpr1* mutant lines showed more significant reduction compared to *Ljcpr2-1* mutant seed pods (Figures 3-5A and B). The physiology change of the seed pod can be observed in Figure 3-5C. This significant involvement of *LjCPR1* gene in pod development might be related to the higher expression of *LjCPR1* in pod and seeds compared to *LjCPR2* (Figure 2-5). *LjCPR1* is also found to be strongly correlated with adenylate translocator with PCC value > 0.75 (Table S4), which was reported to be responsible to translocate starch accumulation in maize endosperms (Shannon et al. 1998), as the storage to nourish the embryo in the seeds (Yan et al. 2014). Loss of *LjCPR1* gene might compromised adenylate translocator function towards seed and pod development. Furthermore, the fact that the loss of *LjCPR1*, but not *LjCPR2*, compromised pod number and length in *L. japonicus* also suggest that *LjCPR1* is crucial to support CYPs and other electron acceptors involved in seed development.

Triterpenoid sapogenins profile of each soil-cultured and hydroponic-cultured mutant plants were analyzed by GC-MS and the result showed different profile between *Ljcpr1* and *Ljcpr2-1* loss-of-function mutant lines (Figures 3-6 and 3-7). A significant difference was similarly shown both in soil-cultured and hydroponic-cultured mutant plants whereas β-amyrin, oleanolic acid, lupeol, 24-OH β-amyrin, and sophoradiol levels were decreased significantly in *Ljcpr2-1* mutant lines but showed little or no change in *Ljcpr1* mutants (Figures 3-6 and 3-7). Interestingly, both *Ljcpr1* and *Ljcpr2-1* mutant lines showed significant decrease in betulinic acid and ursolic acid, with no change in soyasapogenol E and A (Figures 3-6 and 3-7). Both mutants also showed lower

level of soyasapogenol B, however, the decrease of soyasapogenol B is more significant in hydroponic-cultured compared (Figure 3-7) to soil-cultured mutant roots (Figure 3-6). To check the effect of *Ljcpr* loss-of-function effect to the primary metabolisms, the level of campesterol, β-sitosterol, and stigmasterol as three major phytosterols were also analyzed and annotated based on mass spectra from NIST library (Figure S5). However, due to shift in retention time, campesterol could not be detected in hydroponic-cultured samples. Campesterol, β-sitosterol, and stigmasterol were shown to be significantly reduced in both *Ljcpr1* and *Ljcpr2-1* loss-of-function soil-cultured mutants (Figure 3-6). Interestingly, whereas both *Ljcpr1* and *Ljcpr2-1* mutants showed reduction in β-sitosterol content, only *Ljcpr1* mutants showed lower level of stigmasterol in hydroponic-cultured mutants (Figure 3-7). In *A. thaliana*, stigmasterol is synthesized from β-sitosterol by CYP710A, while campesterol and β-sitosterol biosynthesis do not involve CYPs (Morikawa et al 2006). Therefore, the involvement of CYP71 family in stigmasterol biosynthesis pathway might be correlated stronger with *LjCPR1* compared to *LjCPR2*. Further experiment must be conducted to elucidate the differential role of CPR classes in phytosterol biosynthesis (Morikawa et al. 2006).

Previous result on gene co-expression analysis and MeJA treatment showed that *LjCPR2*s showed strong correlation with *bAS*, *CYP93E2*, and *CYP72A61* genes, which were supported by the analysis of *Ljcpr2-1* loss-of-function mutants that showed significant reduction of β-amyrin, 24-OH β-amyrin, and sophoradiol compared to *Ljcpr1* mutants and wild-type. On the other hand, it was previously suggested that *LjCPR1* showed strong correlation with *CYP716A51*, which was shown by the significant reduction in betulinic acid and ursolic acid level of *Ljcpr1* loss-of-function mutants. However, betulinic acid and ursolic acid were also reduced in *Ljcpr2-1* mutants, suggesting that *LjCPR1* is not acting alone in supporting *CYP716A51*. There might be synergistic work of *LjCPR1* and *LjCPR2-1* in supporting *CYP716A51*. But in case of *CYP93E2* and

CYP72A61, the vital role of *LjCPR2-1* cannot be complemented by the presence of *LjCPR1*. The level of oleanolic acid which was not changed in *Ljcpr1* mutants, which might due to the high level of β-amyrin in *Ljcpr1* mutants compared to *Ljcpr2-1* mutants. However, it was known that CPR should only support CYPs, not OSCs which implied that there might be a more complex regulatory mechanism on how CPR works with CYPs. The possibility is that the *CYP716A51* can make a specific complex with different OSCs (*bAS*, *aAS*, *LUS*) and create a specific metabolon. Metabolon is a temporary multi-protein complexes of sequential enzymes that mediate substrate channeling (Zhang and Fernie 2021). Metabolons have been found in several primary metabolisms such as monolignol biosynthesis in *Populus trichocarpa* (Lin et al. 2021) or secondary metabolisms such as camalexin biosynthesis in *Arabidopsis thaliana* (Mucha et al. 2019) and dhurrin biosynthesis in *Sorghum bicolor* (Nielsen et al. 2008) which involved membrane bound CYPs, CPRs, and other enzymes. The significant reduction of betulinic acid and ursolic acid but not in oleanolic acid level suggested that *LjCPR1* specifically involves in betulinic acid and/or ursolic acid biosynthetic pathway, but not with oleanolic acid biosynthesis. Instead, it was suggested that *LjCPR2-1* might involve more closely in oleanolic acid biosynthetic pathway with *bAS* as the first enzyme step in converting 2,3-oxidosqualene into β-amyrin as substrate for *CYP716A51*. However, more studies need to be carried on to confirm this hypothesis. Protein-protein interaction analysis such as using bimolecular fluorescence complementation or protoplast two-hybrid assay should be conducted to reveal the structural proteins of the metabolon scaffold *in planta* (Nielsen et al. 2008; Mucha et al. 2019).

3.3.3 Knockout of the *Ljcpr1* gene in transgenic hairy roots

To directly confirm the involvement on different LjCPR classes on triterpenoids biosynthetic pathway, we used the CRISPR-Cas9 system. We previously described a CRISPR-Cas9 vector, pMgP237-2A-GFP, which can express multiplex gRNAs (Nakayasu et al. 2018; Hashimoto et al. 2018). Since *LjCPR2-1* and *LjCPR2-2* gene have very similar sequence identity, it was very difficult to obtain single or complete null-mutant of double knockout *Ljcpr2*s due to less probability of getting rid all the intact sequences. Thus, in this study, only *Ljcpr1* knockout mutants were successfully obtained and analyzed. Total of nine target sequences on *LjCPR1*, (Figure 3-8; Table S8), were selected using CRISPRdirect software (Naito et al. 2015). However, only five targets were successfully cut to produce frameshift mutants (Figure 3-8). Two target sequences were simultaneously integrated into the vector to generate double tgRNA-pMgP237. Transgenic hairy roots were induced by *A. rhizogenes* ATCC15834 harboring double tgRNA-pMgP237 or the empty vector as a control.

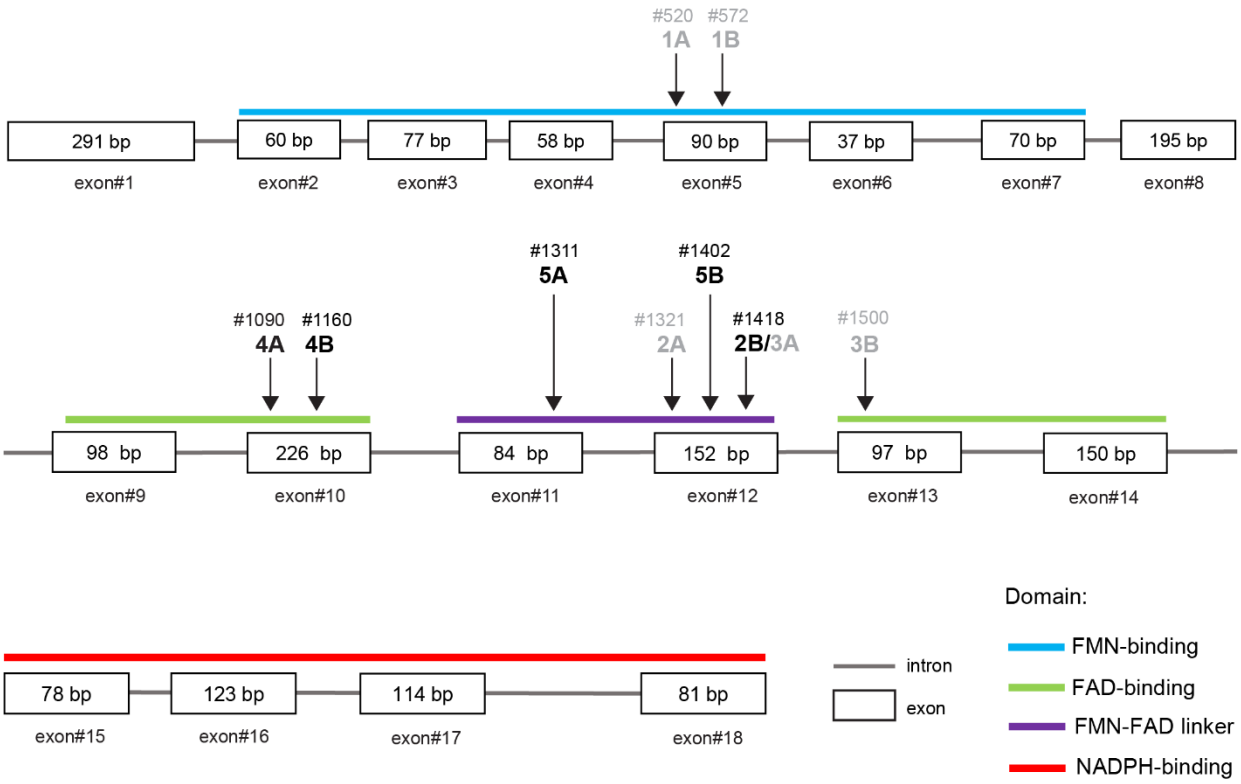

Figure 3-8. The gRNA design targeting *LjCPR1* gene.

Five sets of duplex-gRNA targeting LjCPR1 gene were designed based on https://crispr.dbcls.jp/. The number in hashtag represents the base pair position of the target gRNA sequence in the exon part.

Putative *Ljcpr1* mutant hairy root lines were selected by PCR and electrophoresis (heteroduplex mobility assay, HMA). Some extra bands were observed in the nine hairy root lines (L1-2.1, 2.2, 2.3, 2.4, 2.5, 4.1, 4.2, 5.1 and 5.2), but not in the vector control hairy root lines (EV-1 and EV-2) (Figure 3-9), suggesting that mutations occurred in the *LjCPR1* gene and produced heteroduplex PCR fragments. Genomic DNA fragments around the target sites were cloned and sequenced. Mutated alleles were not found in the control lines, EV-1 (Figure 3-9). No wild-type (WT) sequences were detected in all obtained mutants (Figure 3-9). Six hairy root lines showed frameshift mutations, with other three hairy root lines were non-frameshift mutants. L1-2.2 and L1-4.2 mutants showed no frameshift mutations in one of the alleles (-12 bp and -72 bp,

respectively) and L1-2.5 showed no frameshift mutation in both alleles (-6 bp and -12 bp) (Figure 3-9). These non-frameshift mutants were displayed to give contrast in the result in compare to the wild-type and frameshift mutants.

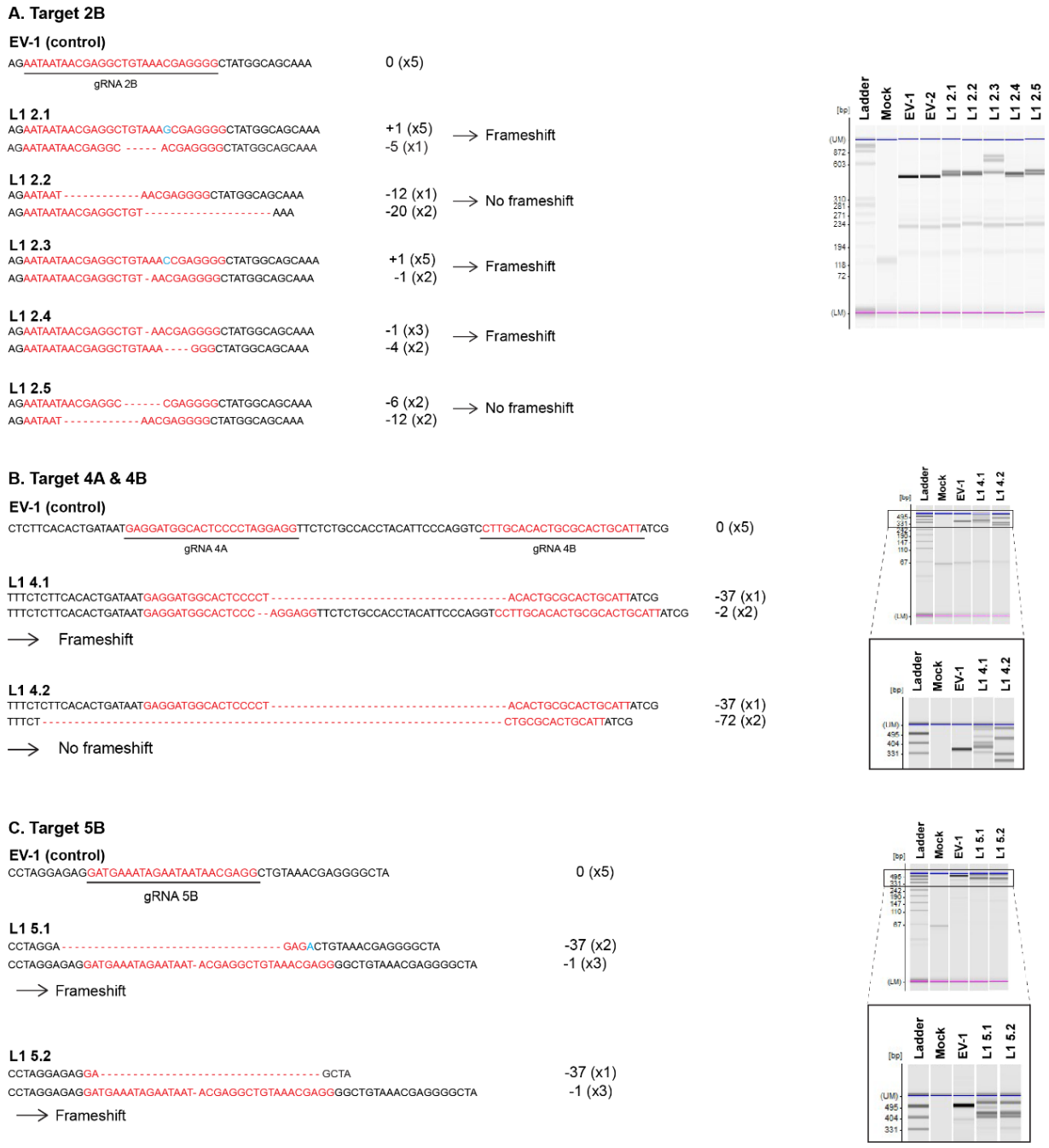

Figure 3-9. Disruption of the *LjCPR1* gene in transgenic *L. japonicus* hairy roots by CRISPR-Cas9 system using gRNA target number A) 2B, B) 4A and 4B, and C) 5B.

The deleted sequences are annotated with red colors. The numbers inside the brackets represent number of clones with the sequence patterns observed. The figure on the right-hand side is HMA result of selected mutant lines.

The triterpenoids composition of control and *Lcpr1* hairy root mutants were then analyzed by GC-MS. The hairy root culture and GC-MS analysis were conducted separately for *Ljcpr1* hairy root mutant lines target 2 (L1 2.1-2.5, Figure 3-10) and target 4 and 5 (L1 4.1, 4.2, 5.1, 5.2, Figure 3-11). All the frameshift mutants from different target region showed significantly lower betulinic acid and ursolic acid content compared to control. A significantly lower lupeol content compared to control was also observed in all L1-2 frameshift mutants (L1-2.1, 2.3, 2.4) (Figure 3-10). Both frameshift mutant lines showed similar level of lupeol, α-Amyrin and soyasapogenols compared to control (Figures 3-10 and 3-11) with the exception of L1-4.1 which showed significantly lower level of those triterpenes compared to control (Figure 3-11). All frameshift mutant lines showed similar level of β-amyrin, 24-OH β-amyrin, and sophoradiol compared to control (Figures 3-10 and 3-11). Interestingly, some of the non-frameshift mutants L1-2.2 and L1-4.2, and frameshift mutants L1-2.3 and L1-2.4 showed significantly higher oleanolic acid compared to control (Figure 3-10), while other mutants showed similar level of oleanolic acid compared to control (Figures 3-10 and 3-11). Even one non-frameshift mutant, L1-4.2, showed significantly higher betulinic acid content compared to control (Figure 3-11). In the phytosterol biosynthesis pathway, the knockout of *Ljcpr1* genes showed significantly lower β-sitosterol and stigmasterol amount compared to control (Figures 3-10 and 3-11). The TIC chromatogram showing significant decrease in betulinic acid level of *Ljcpr1* knockout mutant hairy roots can be seen in Figure S6.

Different triterpenoids profile among mutants might be caused by the difference of mutation patterns, which caused an unknown effect to the amino acid changes in the mutated *Ljcpr1*. The changes in the amino acid also highly possible to alter the CYP:CPR binding motifs (Figure S2),

which gave rise to possibility that either these non-functional *Ljcpr1* mutant proteins still bind to the CYPs, or they were replaced by another redox proteins, such as LjCPR2s or cytochrome b_5 due to change in protein-protein affinity (Esteves et al. 2020; Campelo et al. 2018). However, the similar pattern that can be observed from all of the frameshift mutants was the low capability of the knockout *Ljcpr1* mutants to produce betulinic acid and ursolic acid compared to control. Moreover, the knockout of *Ljcpr1* gene did not alter the production of β-amyrin, 24-OH β-amyrin and sophoradiol content in all mutants. However, the increase of oleanolic acid in some of the frameshift mutants L1-2.3 and L1-2.4 (Figure 3-10) still remained a question, of how is the involvement of LjCPR2s in bAS and CYP716A51 possible metabolon, which seemed to be not affected by the loss of *Ljpr1* gene, unlike betulinic acid and ursolic acid biosynthesis. Despite of some of the differences in triterpenoids profile of each mutant, these results clearly confirmed previous experiment in *Ljcpr1* loss-of-function *LORE1* insertion mutant plants, which showed strong involvement of *LjCPR1* in betulinic acid and ursolic acid production, but not vital for β-amyrin, oleanolic acid, 24-OH β-amyrin, and sophoradiol production.

Both CPR classes were reported to be able to support the *in vitro* activities of CYPs involved in specialized metabolism (Rana et al. 2013). However, based on numerous functional analyses *in planta*, CPR class I was believed to be responsible for basal or constitutive metabolisms, while CPR class II was more responsible for adaptation and defense mechanisms, involving numerous specialized metabolisms (Rana et al. 2013; Ro et al. 2002; Qu et al. 2015; Sundin et al. 2014; Parage et al. 2016; Huang et al. 2012). Functional analysis of LjCPR classes *in planta* showed that LjCPR1 is closely involved with CYP716A51, a C-28 oxidase, which is considered to be involved in triterpenoids biosynthesis, one of the specialized metabolisms in *L. japonicus*. Based on the elicitor treatment, the LjCPR class I showed no induction, which is in line with previous notion,

that only CPR class II in inducible. The result that CYP716A51 was not inducible during methyl jasmonate treatment, might suggest other function of C-28 oxidized triterpenes in *L. japonicus* other than defense mechanism. In line with phenotypic change observed in seed of *Ljcpr1* mutant plants and expression level changes in mature seed, *LjCPR1* and *CYP716A51* might also involve in physiological role of seed development of *L. japonicus*.

In *M. truncatula*, *CYP716A12* involves in hemolytic sapogenin biosynthesis which is the derivative of oleanane type saponins. During developmental stage such as reproductive phase, *CYP716A12* showed significant increase in expression level and increase in hemolytic sapogenin content. Ten-week-old *M. truncatula cyp716a12* mutant plant showed dwarf phenotype compared to wild-type. This finding suggested a possible dual role of hemolytic sapogenins in defense and plant developmental growth in *M. truncatula* (Carelli et al. 2011). Functional analysis and phenotypic observation of *cyp17a51* mutant in *L. japonicus* might provide insight on *LjCPR1* and *CYP716A51* involvement in plant primary metabolism.

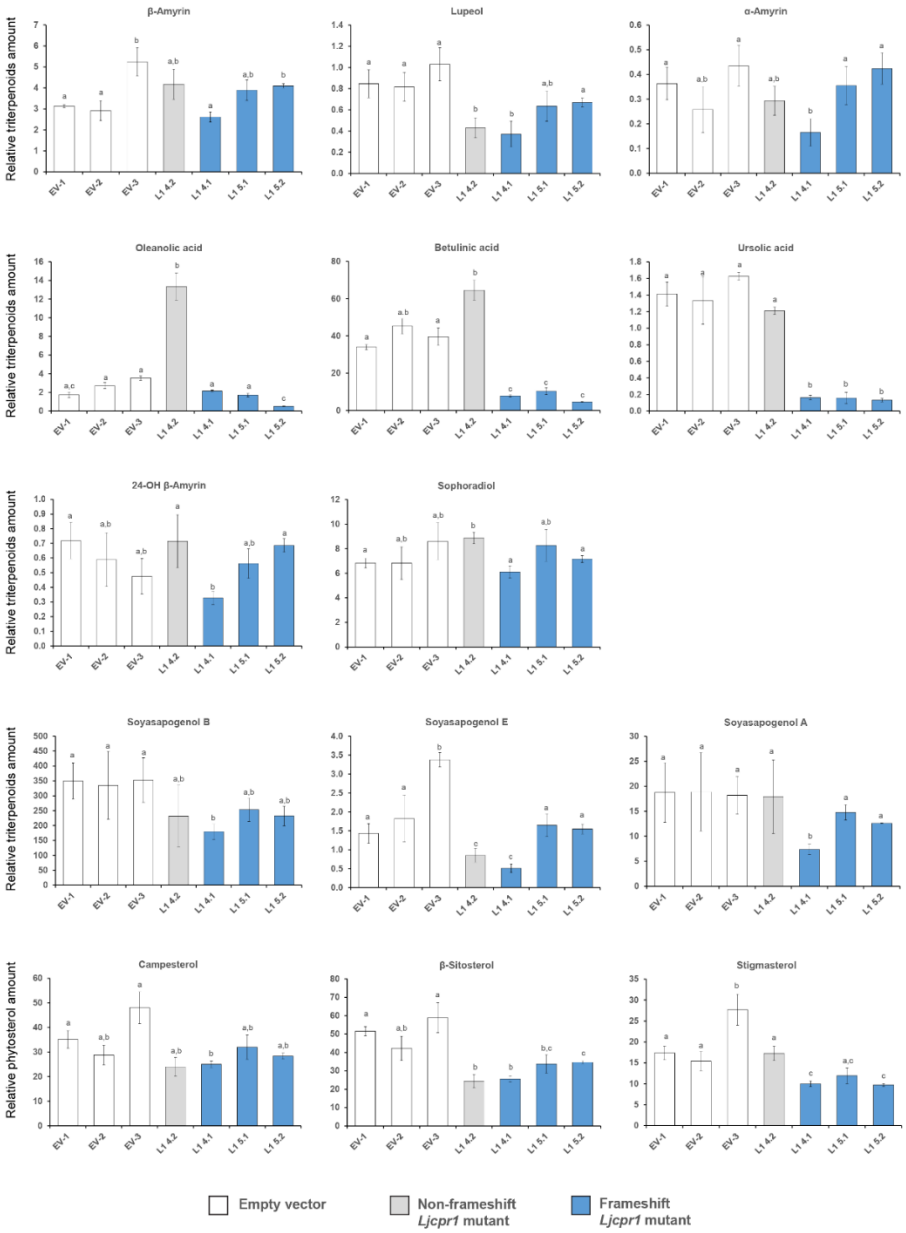

Figure 3-10. The relative amount of A) triterpenoids and B) phytosterol content hairy root *Ljcpr-1* (target 2) mutants analyzed by GC-MS.

Relative triterpenoids and phytosterol amount were normalized to that of asiatic acid as internal standard and are presented as fold induction relative to the control. Data represent the mean of three technical replicates ± SD. Statistical significance in comparison with the corresponding control values is indicated by Student's t-test; * $p < 0.05$. SD, standard deviation.

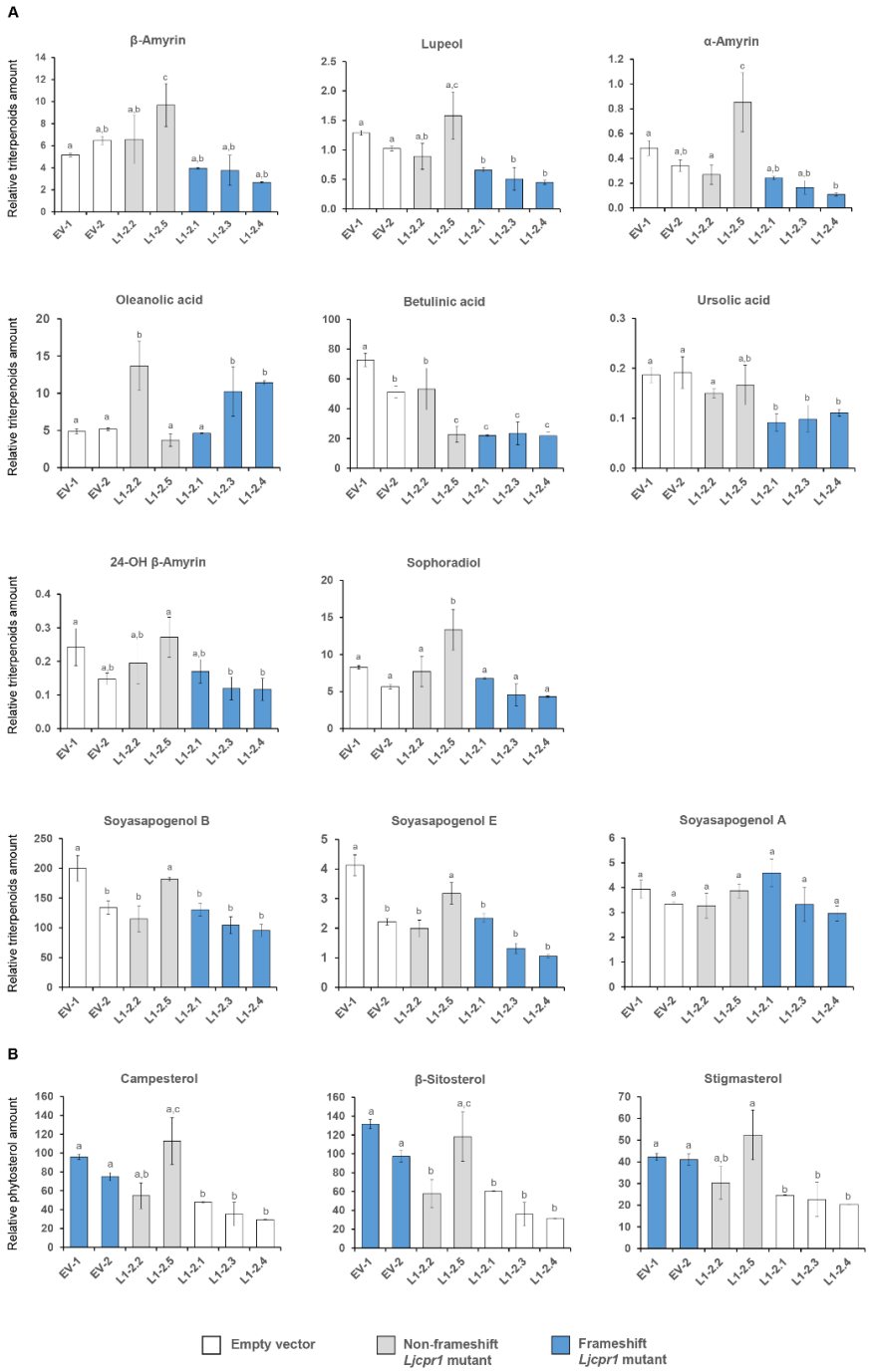

Figure 3-11. The relative amount of A) triterpenoids and B) phytosterol content hairy root *Ljcpr-1* (target 4&5) mutants analyzed by GC-MS.

Relative triterpenoids and phytosterol amount were normalized to that of asiatic acid as internal standard and are presented as fold induction relative to the control. Data represent the mean of three technical replicates ± SD. Statistical significance in comparison with the corresponding control values is indicated by Student's t-test; * $p < 0.05$. SD, standard deviation.

3.4 Conclusion

Elicitor experiment on *L. japonicus* hairy roots reveal different set of triterpenoids biosynthetic genes correlated with either LjCPR class I or II. *LjCPR1* was observed to be closely correlated with *CYP716A51* and *LUS*, where these genes are down-regulated by MeJA treatment and no change is observed in betulinic acid and lupeol content. In contrast, *LjCPR2*s were closely correlated with *bAS*, *CYP93E2*, and *CYP72A61*, since they are significantly upregulated by MeJA treatment, followed by significant increase of β-amyrin, 24-OH β-amyrin, and sophoradiol content. These different set of genes might exhibit different transcriptional regulation upon elicitation with MeJA. The downregulation of *LjCPR1*, *CYP716A51*, and *LUS* suggested that these genes are not responsible in defense response which involves jasmonate signalling. The involvement of *LjCPR1* and *LjCPR2* in the triterpenoids biosynthetic pathway was then confirmed by loss-of-function mutant experiment, revealing that *LjCPR1* and *LjCPR2* are both important for betulinic acid and lupeol production, with possibly synergistic effect, in which loss of either one of the CPR classes cannot restore betulinic acid and lupeol content. On the other hand, the loss of *LjCPR2* function cannot be complemented by the presence of *LjCPR1* towards β-amyrin, 24-OH β-amyrin, and sophoradiol production.

Furthermore, *LjCPR1* was found to be crucial for seed development, but not *LjCPR2*, supporting previous notion that CPR class I might support plant basal metabolism. Interestingly, higher *LjCPR1* involvement with *CYP716A51* compared to *LjCPR2* suggested a deviance from previous notion that stated CPR class II is more correlated with CYPs involved in specialized metabolism, including triterpenoids, compared to CPR class I. However, other possibility arises that *CYP716A51* might also involve in plant developmental stage, similar to *CYP716A12* of *M.*

truncatula. This information implies that different CPR class has distinct involvement in triterpenoids biosynthesis which might be beneficial to propose a novel strategy to improve heterologous production of our target triterpenoids.

Chapter 4

Co-expression of different plant CPRs for heterologous triterpenoids production in transgenic yeast

4.1 Introduction

Previous studies have demonstrated a different correlation between LjCPR class I and II towards different CYPs. *LjCPR1* showed stronger involvement with *CYP716A51* and *LUS*, while *LjCPR2-1* showed stronger involvement with *bAS*, *CYP93E1*, and *CYP72A61*. These tendencies might not only be observed *in planta*, but also by co-expressing them in other platform, such as other plants of microorganisms. The different correlation strength or preference of each CPR classes towards different triterpenoid sapogenins biosynthesis might affect their heterologous production. As previously mentioned, N-terminal membrane sequence of CPR class I have around 10 amino acids shorter than CPR class II. The long N-terminal membrane sequence of CPR class II show to be rich in Ser/Thr sequences that might involve in protein catalytic activity. CPR class II from *Arabidopsis thaliana* and *Gossypium heliantum* exhibits higher reductase activity towards cytochrome *c* in the presence of NADPH compared to its CPR class I (Urban et al. 1997; Yang et al. 2010). However, until now there are no clear reports about the effect of different N-terminal membrane sequence between legume CPR class I and II towards triterpenoids biosynthesis.

Many CYPs from legumes have already been used for heterologous production of triterpenoids. Recently, glycyrrhizin has been successfully produced de novo in yeast by the combinatorial expression of different enzymes (Chung et al. 2020). Glycyrrhizin shows high pharmacological activities and is used as a natural sweetener because its sweetness is 150 times

higher than that of sucrose (Kitagawa 2002). Improving the production of glycrrhizin by yeast metabolic engineering would be very beneficial for production companies.

Pairing different CPRs with different CYPs has also been reported to result in different productivity of plant specialized metabolites in heterologous yeast (Zhu et al. 2018). In previous studies, CPRs were randomly selected without consideration of the classes (Zhu et al. 2018; Jin et al. 2019). Therefore, the CPR screening was still a subject of trial and error, which might lead to an ineffective strategy for improving heterologous production. Therefore, by performing comparative analysis of different CPR classes of legumes, achieved by co-expression different pairs of legume CPRs and CYPs, might give insight into the differential functions of CPR classes towards improving the heterologous production of triterpenoids in transgenic yeast.

4.2 Material and methods

4.2.1 Chemicals

β-Amyrin, erythrodiol, oleanolic acid, α-amyrin, uvaol, ursolic acid, and lupeol were purchased from Extrasynthese (France). Betulin, methyl jasmonate, and salicylic acid were purchased from Sigma-Aldrich (USA). Betulinic acid and gibberellin were purchased from Tokyo Chemical Industry (Japan). Sophoradiol, 24-OH β-amyrin, 24-COOH β-amyrin, 11-oxo-β-amyrin, 30-OH β-amyrin, and 11-deoxoglycyrrhetinic acid (30-COOH β-amyrin) were kindly gifted by Dr. Kiyoshi Ohyama (Tokyo Institute of Technology, Japan).

4.2.2 N-terminal domain swapping of CPR-I and II

The position of protein helix inside, transmembrane, and outside of the endoplasmic reticulum (ER) were predicted using *in silico* transmembrane helix prediction

(http://www.cbs.dtu.dk/services/TMHMM/) (Figure S1). In this study, the N-terminal domain are represented by the protein located inside ER together with the transmembrane helix. To swap the N-terminal domain of MtCPR1 and MtCPR2, the first 46 amino acids of MtCPR1 were swapped with the first 60 amino acids of MtCPR2 using NEBuilder® HiFi DNA Assembly cloning to produce N-terminal sequence of MtCPR1 fused with truncated-N-terminal MtCPR2 (M1N-M2C) and N-terminal sequence of MtCPR2 fused with truncated-N-terminal MtCPR1 (M2N-M1C). Two different fragments of each designed chimeric CPRs (M2N-M1C and M1N-M2C) were amplified using pENTR-MtCPR1 and pENTR-MtCPR2 as template and primers listed in Table S11. A minimum of 20-bp overlapping nucleotides of each CPR mutant and pENTR™ as vector backbone were designed for fusion cloning.

4.2.3 Cloning and vector construction

CPR class I and II genes from *M. truncatula*, *L. japonicus*, and *G. uralensis* were amplified from the respective cDNA using PCR with PrimeSTAR® Max DNA Polymerase (Takara Bio, Japan) and named MtCPR1, MtCPR2, LjCPR1, LjCPR2, GuCPR1, and GuCPR2. *LjCPR2-1* sequence was used for this experiment as LjCPR2. The primers used in this study are listed in Table S11. Wild-type and chimeric CPR genes were cloned into the pENTR™ vector using the D-TOPO® (Thermo Fisher Scientific) or NEBuilder® HiFi DNA Assembly (USA) cloning method, to produce entry clones of CPR genes. Yeast expression clones of CPR, using pAG415-GAL-ccdB (plasmid number 14145, Addgene, USA) as the destination vector, were constructed using LR reaction with LR Clonase™ II Enzyme mix (Thermo Fisher Scientific), to produce pAG415-CPR. pAG415GAL-ccdB was a gift from Susan Lindquist. Yeast expression clones of MtCYP716A12, MtCYP72A63, LjCYP716A51, LjCYP93E1, LjCYP72A61, and GuCYP88D6 were constructed using LR reaction with LR Clonase™ II Enzyme mix (Thermo Fisher Scientific) into pYES-

DEST52 (Invitrogen, USA) and a Gateway™-compatible version of pESC-HIS generated previously in our laboratory (Yasumoto et al. 2016) as the destination vectors, to produce pYES-DEST52-CYP and pESC-HIS-CYP (Seki et al. 2008; Fukushima et al. 2011; Suzuki et al. 2019).

4.2.4 Yeast strain construction

Two different yeast strains were used in this research, INV*Sc*1 (MATa his3D1 leu2 trp1-289 ura3-62; Thermo Fisher Scientific) and PSIII strain. PSIII strain was generated by transforming pESC-TRP[PGAL10/tHMG1-T2A-upc2-1][PGAL1/LjbAS]) as described in Supplementary Methods 2 into PSI strain (BY4742/ Perg7::PMET3-ERG7/ trp1::PACT1-Gal4dbd-ER-VP16) constructed in a previous study (Srisawat et al. 2020). INV*Sc*1 strain carrying pYES3-ADH-aAS, pYES3-ADH-OSC1, or pYES3-ADH-LUS (Seki et al. 2008) and PSIII strain were then co-transformed with pAG415-CPR, pYES-DEST52-CYP, and pESC-HIS-CYP consecutively using Frozen-EX Yeast Transformation II™ (Zymo Research, USA).

4.2.5 Yeast cultivation

In case of INV*Sc*1, yeast sample with cell density at an OD600 value in the range of 2.2–2.5 (late logarithmic phase) were cultured in 5 ml appropriate synthetic defined medium (Clontech, USA) containing 2% glucose at 30°C and 220 rpm for 24 h. Yeast cells were collected by centrifugation, resuspended in 5 ml of appropriate synthetic defined medium (Clontech) containing 2% galactose and then incubated at 30°C and 220 rpm for 48 h. In case of the PSIII strain, yeast strains with cell density at an OD_{600} value in the range of 1.4–1.6 (mid-logarithmic phase) were cultured in 5 ml appropriate synthetic defined medium (Clontech) containing 2% glucose, 100 nM β-estradiol, and 1 mM L-methionine and then incubated at 30°C and 220 rpm for 48 h. For the feeding assay, INV*Sc*1 yeast was cultured as described previously and supplemented with 10 μM

of erythrodiol at the same time as galactose addition, and yeast were then cultured at 30°C and 220 rpm for 60 h.

4.2.6 Metabolite extraction

Before extraction, 20 ppm of ursolic acid or uvaol were added to 5 ml yeast culture as internal standards. Triterpenoid metabolite extraction was performed as previously reported (Fanani et al. 2019). The dried extract was dissolved in 500 μl of chloroform:methanol 1:1 (v/v). Upon Gas chromatography–mass spectrometry (GC-MS) analysis, 100 μl of the sample was evaporated to dryness and trimethylsylated with 50 μl of N-methyl-N-(trimethylsilyl) trifluoro acetamide (Sigma-Aldrich) for 30 min at 80°C.

4.2.7 GC-MS analysis

GC-MS was performed as previously reported (Fanani et al. 2019). For peak identification, authentic standards were used to confirm the peak in the sample and compare the retention time and mass fragmentation patterns. The relative amount of triterpenoids of each strains were compared by calculating the area of each peak from the extracted chromatogram based on m/z values 203 (Kim et al. 2018). Each sample of the different CYP-CPR combinations was analyzed using three biological replicates.

4.3 Results and discussion

4.3.1 Co-expression of different legume CPRs and CYPs in transgenic yeast

To analyze the effect of different CPR classes on heterologous triterpenoid production, different CYPs and CPR classes from *M. truncatula* and *L. japonicus* were co-expressed in transgenic *S. cerevisiae* INV*Sc*1, a commonly used strain for recombinant protein expression. By comparing the relative value of each metabolite by semi-quantitative method with an internal standard, we compared the accumulation of oxidized triterpenoids resulting from the heterologous expression of different CPR-CYP pairs in transgenic yeast (Figure 4-1). The amounts of the triterpenoid backbone in control samples were considered to be 100%. In yeast strain expressing MtCYP716A12 and LjCYP716A51, the intermediate compounds oleanolic aldehyde and ursolic aldehyde could be detected based on the mass spectrum profile (Figure S3) compared to previously reported mass spectrum of oleanolic aldehyde and ursolic aldehyde authentic standard (Misra et al. 2017). However, betulinic aldehyde could not be detected in the yeast extract and known to be co-eluted together with betulinic acid in yeast extract (Suzuki et al. 2018), hence the betulinic acid amount in Figure 4-2 corresponds to the sum of both compounds. C-30 and C-24 aldehydes were excluded due to the instability of the compound, lack of authentic standards, and no reports on the mass spectrum profile.

The results show that upon using INV*Sc*1 β-amyrin-producing yeast, MtCYP716A12 and LjCYP716A51 showed higher conversion ratio of β-amyrin into erythrodiol and oleanolic acid when paired with MtCPR1, as compared to when these were paired with MtCPR2, with a comparable amount of oleanolic aldehyde (Figure 4-1A-B). Pairing MtCYP716A12 with LjCPR1 showed higher conversion of β-amyrin into erythrodiol compared to pairing it with LjCPR2. However, pairing MtCYP716A12 with LjCPR2 showed higher conversion ratio of erythrodiol into

oleanolic aldehyde and oleanolic acid than with LjCPR1 (Figure 4-1A). On the other hand, pairing LjCYP716A51 with LjCPR2 showed higher conversion ratio of β-amyrin into erythrodiol than with LjCPR1, while the conversion ratio of erythrodiol into oleanolic acid is higher when pairing LjCYP716A51 with LjCPR1 than LjCPR2 (Figure 4-1B). These trends are also showed to be similar in the yeast feeding assay result of INV*Sc*1 strain supplemented with erythrodiol as a substrate, even though the difference is not as significant as *in vivo* assay (Figure S8). In contrast, pairing MtCYP72A63 with CPR class II resulted in significantly higher conversion of β-amyrin into 30-OH β-amyrin and 30-COOH β-amyrin levels (Figure 4-1C) than pairing it with CPR class I. Similar to MtCYP72A63, pairing LjCYP72A61 with CPR class IIs resulted in significantly higher conversion of β-amyrin into sophoradiol (Figure 4-1D). In case of the legume-specific CYP subfamily, CYP93E1 paired with MtCPR2 showed higher conversion of β-amyrin into 24-COOH β-amyrin, as compared to CYP93E1 paired with MtCPR1. However, CYP93E1 paired with LjCPR1 and LjCPR2 did not show any significant difference in the conversion ratio of β-amyrin into 24-OH and 24-COOH-β-amyrin (Figure 4-1E).

CYP716As can also catalyze oxidations on different triterpenoid backbone. Therefore, to confirm that the effect of the CPR on the conversion ratio can also be applied to different triterpenoid backbone, we then co-expressed MtCYP716A12 in α-amyrin and lupeol-producing strains (Figure 4-2). The result showed that MtCYP716A12 paired with MtCPR1 and LjCPR1 showed higher conversion of α-amyrin into ursolic acid and uvaol, respectively (Figure 4-2A). Similar trend is also observed in lupeol-producing strain, that MtCYP716A12 paired with MtCPR1 showed higher conversion ratio of betulin into betulinic acid compared to MtCPR2, and MtCYP716A12 paired with LjCPR1 showed higher conversion ratio of lupeol into betulin and

betulinic acid compared to LjCPR2 (Figure 4-2B). All the chromatograms of *yeast in vivo* result can be found in Figure S9-S18.

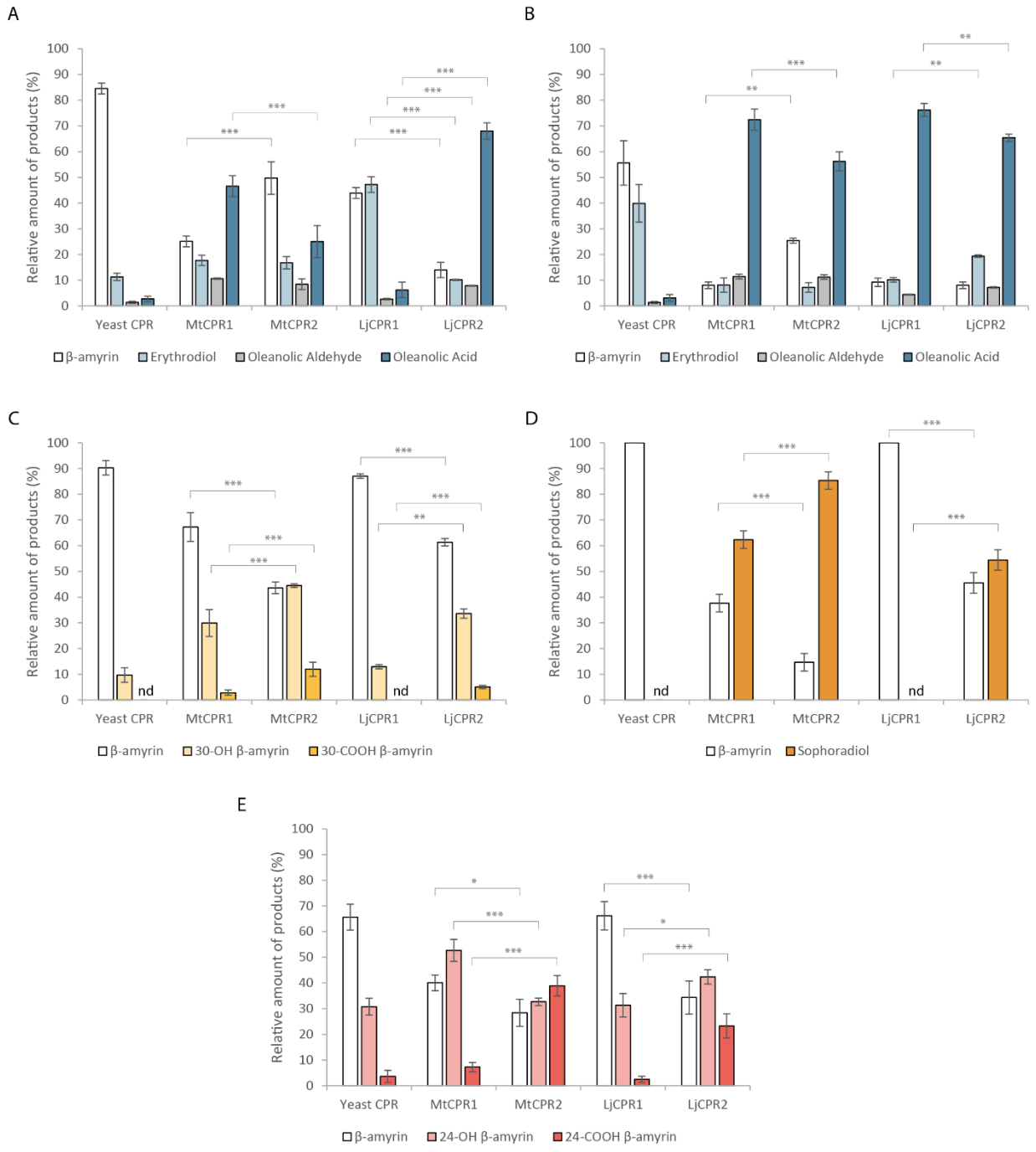

Figure 4-1. The relative amount of triterpenoid produced using combinations of (A) MtCYP716A12, (B) LjCYP716A51, (C) MtCYP72A63, (D) LjCYP72A61, and (E) LjCYP93E1

co-expressed with different CPRs from *M. truncatula* and *L. japonicus* in β-amyrin-producing yeast (INV*Sc*1 strain).

Triterpenoids content were measured relative to uvaol as internal standard. Data have been presented as mean ± SE (n=3). nd, signal below detection limit. Single-factor ANOVA was used for statistical comparisons. Values were considered statistically significant at *P<0.05, **P<0.01, and ***P<0.001.

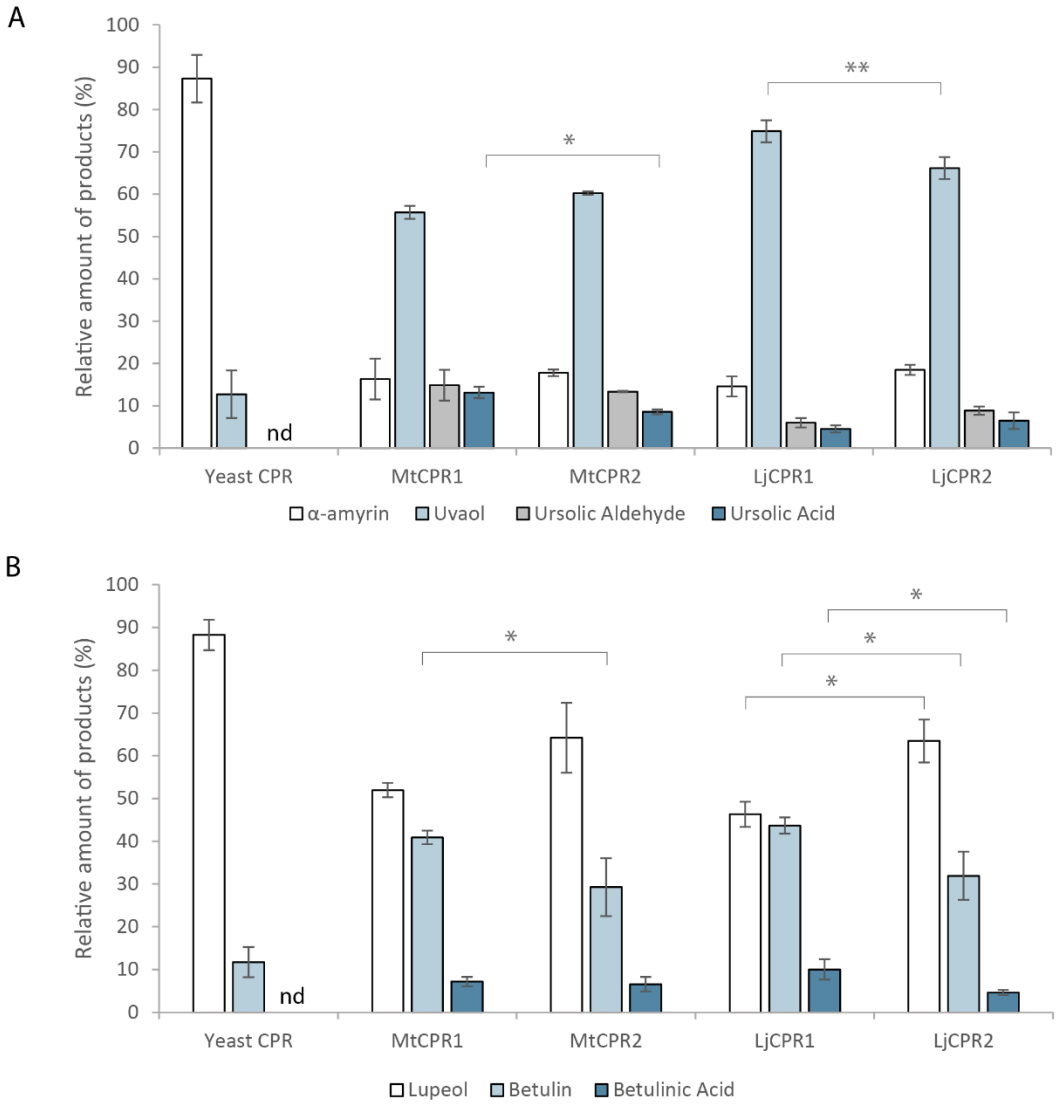

Figure 4-2. The relative amount of triterpenoid produced by co-expressing MtCYP716A12 with different CPRs from *M. truncatula*, *L. japonicus*, and *G. uralensis* in (A) α-amyrin-producing and (B) lupeol-producing yeast (INV*Sc*1 strain).

Triterpenoids content were measured relative to erythrodiol as internal standard. Data have been presented as mean ± SE (n=3). nd, signal below detection limit. Single-factor ANOVA was used for statistical comparisons. Values were considered statistically significant at *P<0.05 and **P<0.01.

This study emphasizes that different CYP shows preference towards certain CPR class. Based on the yeast *in vivo* assay in the INV*Sc*1 β-amyrin-producing strain, CYP716As paired with CPR class I generally showed higher conversion ratio of triterpenoids backbone when they are paired with CPR class II, with the exception of MtCYP716A12 paired with LjCPR1 showed lower conversion ratio of triterpenoids compared to LjCPR2 (Figure 4-1). This might be due to the incompatibility of CYP716A12 from *M. truncatula* with LjCPR1 due to species difference, which suggested that CYP:CPR binding motifs in CYP716A families are species-specific. However, the preference of CPR class I over CPR class II of MtCYP716A12 is also observed in INV*Sc*1 α-amyrin and lupeol-producing strain (Figure 4-1). On the other hand, CYP72A63, CYP72A61, and CYP93E1 showed higher conversion ratio when co-expressed with CPR class II (Figure 4-1).

Interestingly, LjCYP72A61 paired with MtCPR2 showed the highest conversion ratio of β-amyrin into sophoradiol compared to its cognate reductase, LjCPR2. This result suggested that CPR:CYP interaction is not only species specific, but mainly due to CYP:CPR specific binding motifs. *M. truncatula* also has MtCYP72A61, homologue to the LjCYP72A61, which might explain why MtCPR2 can work well with LjCYP72A61 as well. MtCPR2 might contain binding motifs which is suitable for LjCYP72A61, which suggested that CYP:CPR binding motifs of CYP72A families are similar across species. MtCPR2 also might have higher reduction strength or electron transfer activity than LjCPR2, resulting in higher conversion ratio of β-amyrin into sophoradiol, which has to be confirmed by *in vitro* kinetic analysis. Since *L. japonicus* has two LjCPR class II genes, while *M. truncatula* has only one copy of MtCPR class II, it might be reasonable to suggest that *MtCPR2* has higher involvement with *CYP72A61*, compared to LjCPR

class II in which the involvement might be shared between the two genes, *LjCPR2-1* and *LjCPR2-2*.

This yeast *in vivo* assay might give us insight of CPR-CYP preferences which better be justified by their protein expression level. Quantifying CPR protein itself would not be enough since the triterpenoids are the products of CYP. However, the protein quantification of CYP is still challenging. We propose this study to analyze the effect CPR class on heterologous production in yeast, not on the CPR or CYP catalytic activity. We admit that this is the limitation of yeast *in vivo* study. Therefore, further deep *in vitro* analysis of the effect of CPR class to the CYP activity such as using peroxidase assay (Manoj et al. 2010; Mishin et al. 2014) or by incorporating CPR and CYP into ER liposomes and bilayer (Barnaba et al. 2017) might be needed in future experiments to give more justifiable result.

4.3.2 N-terminal domain swapping of MtCPR1 and MtCPR2

Based on the multiple sequence alignment result, the main difference of CPR class I and II is located in the N-terminal domain. CPR class I has shorter amino acid sequence than CPR class II (Figure 2-3). To analyze whether this N-terminal domain is responsible for the differences in the conversion ratio of β-amyrin when CPR was co-expressed with different CYPs in transgenic yeast (Figure 4-1), we investigated the effect of this domain by swapping the N-terminal membrane sequence of MtCPR1 and MtCPR2. MtCPRs were used as the representative since *M. truncatula* is a model legume. We speculated if the N-terminal domain of CPR is indeed responsible for the CPR activity, swapping the N-terminal domain of CPR class I and II will affect the triterpenoids profile.

The location of transmembrane helix of MtCPR1 and MtCPR2 was predicted in silico (Figure S1). Amino acids of MtCPR1 residue 1-26 are located inside the ER, residue 27-46 are the

transmembrane helix, and residue 47-692 are located outside the ER (Figure S1-A). Amino acids of MtCPR2 residue 1-40 are located inside the ER, residue 41-60 are the transmembrane helix, and residue 61-701 are located outside the ER (Figure S1-B). In this study, the first 46 amino acids of MtCPR1 located inside the ER together with the transmembrane helix were swapped with the first 60 amino acids of MtCPR2. The resulting chimeric MtCPRs were N-terminal membrane sequence of MtCPR1 fused with truncated-N-terminal MtCPR2 (M1N-M2C) and N-terminal membrane sequence of MtCPR2 fused with truncated-N-terminal MtCPR1 (M2N-M1C) (Figure 8A).

Yeast *in vivo* assay was then performed by co-expressing the chimeric MtCPRs with CYP716As and CYP72As to observe its effect on triterpenoids conversion ratio (Figure 4-3B-E). The results showed that by swapping the N-terminal domain of CPR class I with that of class II, the product from CYP716A12 and CYP716A51 were significantly reduced (Figure 4-3B-C). The production of erythrodiol by CYP716A51 co-expressed with mutant CPRs was even lower than the background level observed in native yeast CPR (Figure 4-3C). A similar effect was also seen in CYP72A61, where the chimeric CPR caused a significant decrease in conversion of β-amyrin into sophoradiol (Figure 4-3D). Interestingly, only CYP72A63 was not negatively affected by the chimeric MtCPRs. Moreover, the results showed that CYP72A63 paired with N-terminal membrane sequence of MtCPR2 fused with truncated-N-terminal MtCPR1 (M2N-M1C) showed higher conversion of β-amyrin into of 30-OH β-amyrin and 30-COOH β-amyrin (Figure 4-3E).

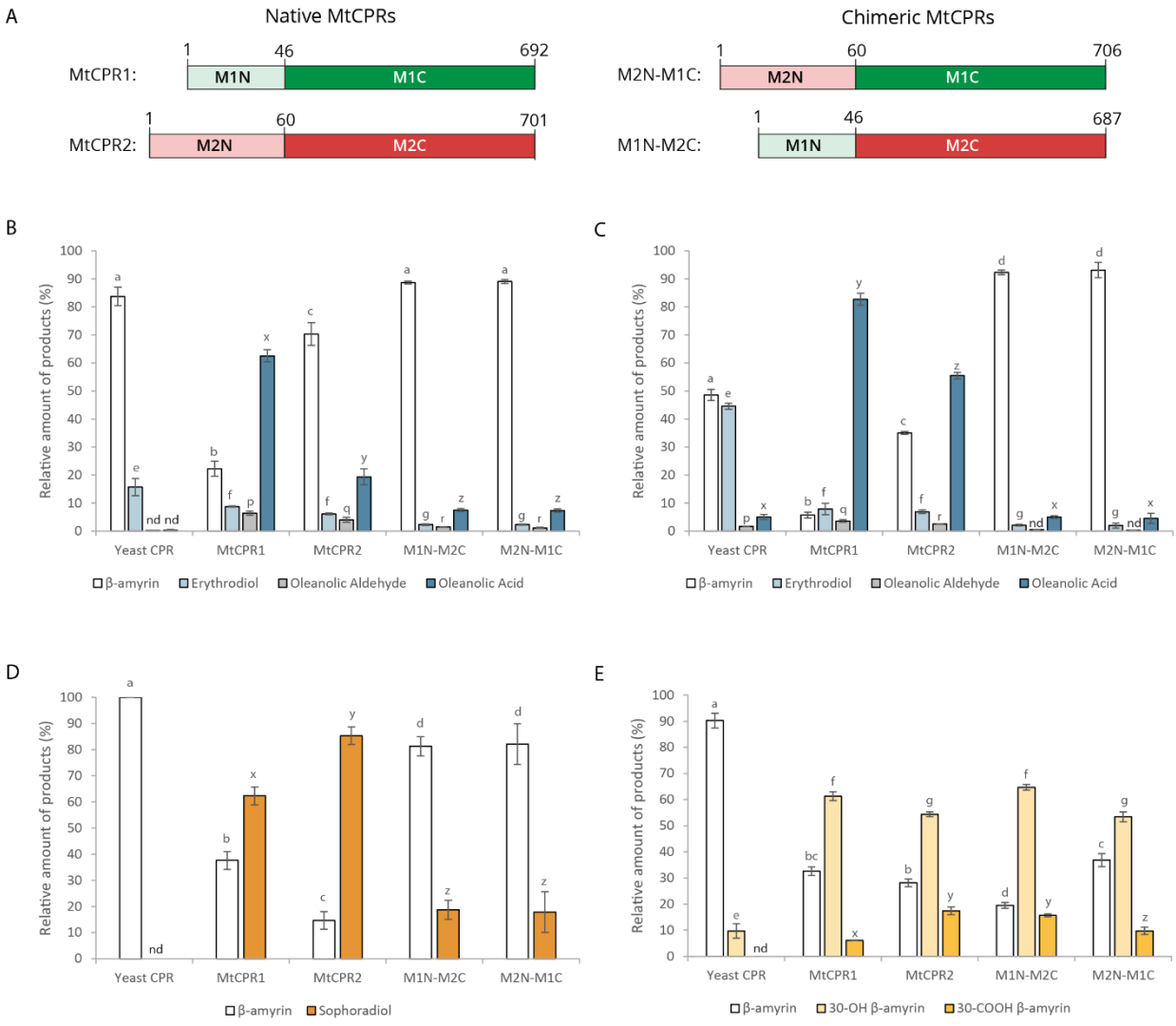

Figure 4-3. The relative amount of triterpenoid produced by co-expressing (A) N-terminal swapped chimeric MtCPRs with (B) MtCYP716A12, (C) LjCYP716A51, (D) LjCYP72A61, and (E) MtCYP72A63 in β-amyrin-producing yeast (INV*Sc*1 strain).

(A) N-terminal domain of MtCPR classes I and II were swapped to generate two chimeric CPRs, M2N-M1C and M1N-M2C. The numbers indicate the position of amino acid residue. Triterpenoids content were measured relative to uvaol as internal standard. Data have been presented as mean ± SE (n=3). Nd, signal below detection limit. Single-factor ANOVA was used for statistical comparisons. The different letters indicate significant differences ($P<0.05$, one-way ANOVA followed by Tukey's test). Letter a-d are used to indicate significancy for β-amyrin, letter e-g are for erythrodiol and 30-OH β-amyrin, letter p-r are for oleanolic aldehyde, and letter x-z are for oleanolic acid, sophoradiol, and 30-COOH β-amyrin.

It has previously been reported that the N-terminal of ATR2 (*Arabidopsis thaliana* CPR2), which contains high number of Ser/Thr residues, significantly enhances the ATR2 activity (Urban

et al. 1997), which might lead to higher conversion ratio. However, N-terminal domain swapping of MtCPR1 and MtCPR2 showed that the longer N-terminal membrane sequence of MtCPR2 did not result in increase of conversion ratio of β-amyrin. In contrast, these chimeric CPRs caused a reduction in the β-amyrin conversion ratio of the CYP716As and CYP72A61, while only showed little or no effect when co-expressed with CYP72A63. Interestingly, chimeric CPR M2N-M1C showed higher conversion ratio of 30-COOH β-amyrin than native MtCPR1, implying the N-terminal domain of MtCPR2 might corresponds to the increase in conversion ratio. This result suggests that there might be a specific CYP-CPR protein-protein interaction that was affected by the sequence-structural relationship in the N-terminal domain of CPRs. A previous study on human CPR-FMN-domain mutants has also shown that different CYP isoforms interact with CPR in a specific manner due to docking elements in the binding motifs of the FMN-domain, which affect the CYP-CPR affinity (Ritacco et al. 2019). The N-terminal domain swapping between MtCPR1 and MtCPR2 might cause a structural change of the CYP:CPR specific binding motif as shown in Figure 2-3. Further studies, such as mutagenesis experiments on CPR, are needed to understand the mechanisms underlying this phenomenon.

4.3.3 Improving triterpenoid production using an engineered yeast

To further analyze the effect of different CPR classes on improving heterologous production, different CPR classes were co-expressed with different CYP families in the PSIII strain. PSIII is an engineered yeast strain that has been optimized for triterpenoid biosynthesis by means of some modifications in the upstream mevalonate pathway. This strain is modified from the PSII strain constructed in previous research, which can produce a 100× times higher triterpene backbone as a substrate than its parental strain (Srisawat et al. 2020). In contrast to the INV*Sc*1 results, in the PSIII strain, almost all the triterpenoid sapogenin production profiles had a lower ratio to β-amyrin

and the trends of CYP-CPR preferences observed in INV*Sc*1 strain seems to be less obvious in this strain (Figure 4-4). Nevertheless, some notable changes were observed in the PSIII strain. MtCYP716A12 paired with LjCPR1 that previously showed different conversion ratio of triterpenoids with LjCPR2 in INV*Sc*1 strain, showed similar conversion ratio profile in PSIII strain (Figure 4-4A). The ratio of oleanolic acid to erythrodiol was lower in PSIII strain co-expressing MtCYP716A12 (Figure 4-4A). The PSIII strain co-expressing MtCYP72A63 resulted in much lower 30-COOH β-amyrin to 30-OH β-amyrin ratio, as compared to the INV*Sc*1 strain (Figure 4-4C). Co-expressing all CPRs with LjCYP72A61 and MtCYP716A51 resulted in similar conversion ratio in PSIII strain (Figure 4-4D). However, similar to INV*Sc*1 strain, co-expressing MtCYP716A12 with MtCPR1 in PSIII strain showed higher oleanolic aldehyde and oleanolic acid conversion ratio compared to MtCPR2. Also, co-expressing MtCYP72A63 with LjCPR2 in PSIII strain showed higher 30-OH β-amyrin and 30-COOH β-amyrin conversion ratio compared to LjCPR1.

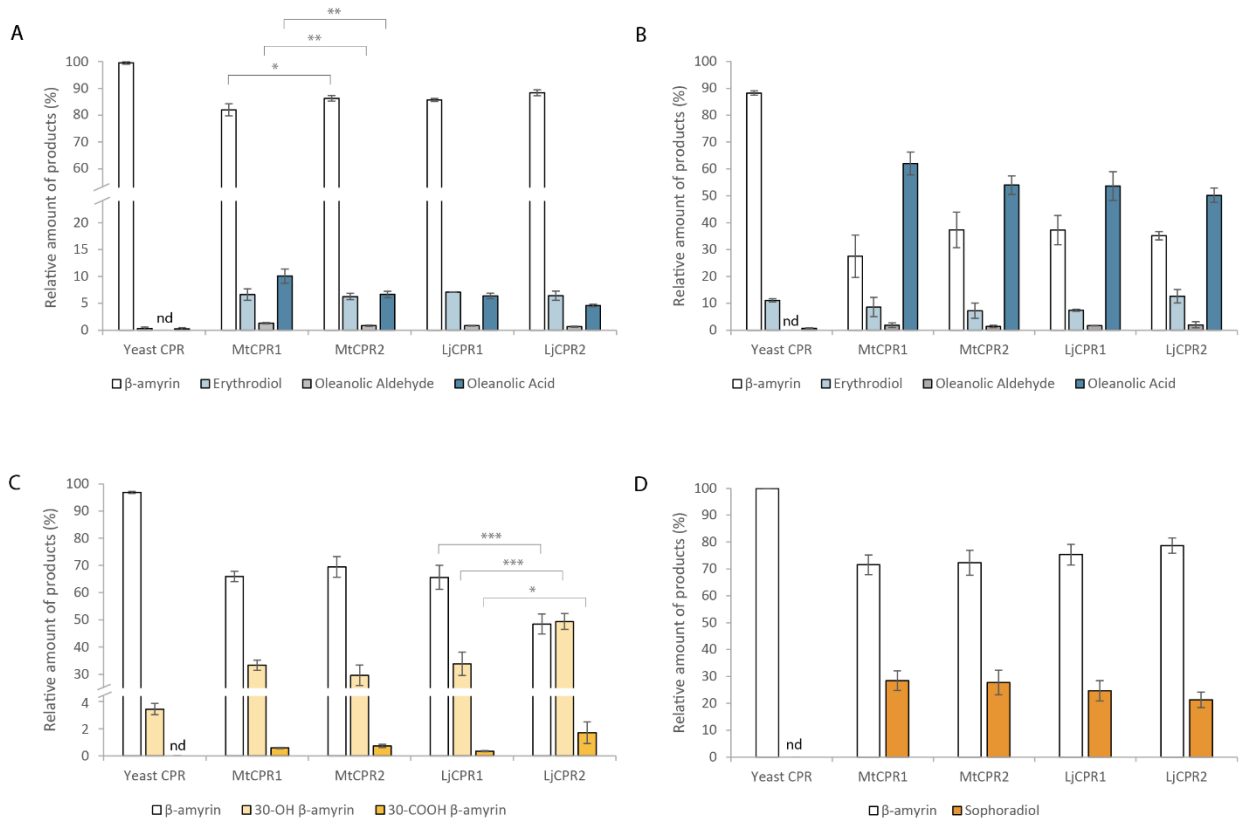

Figure 4-4. The relative amount of triterpenoids from A) MtCYP716A12, B) LjCYP716A51, C) MtCYP72A63, and D) LjCYP72A61 co-expressed with different CPRs from *M. truncatula* and *L. japonicus* in the engineered PSIII yeast strain.

Triterpenoids content were measured relative to uvaol as internal standard. Data have been presented as mean ± SE (n=3). nd, signal below detection limit. Single-factor ANOVA was used for statistical comparisons. Values were considered statistically significant at *P<0.05, **P<0.01, and ***P<0.001.

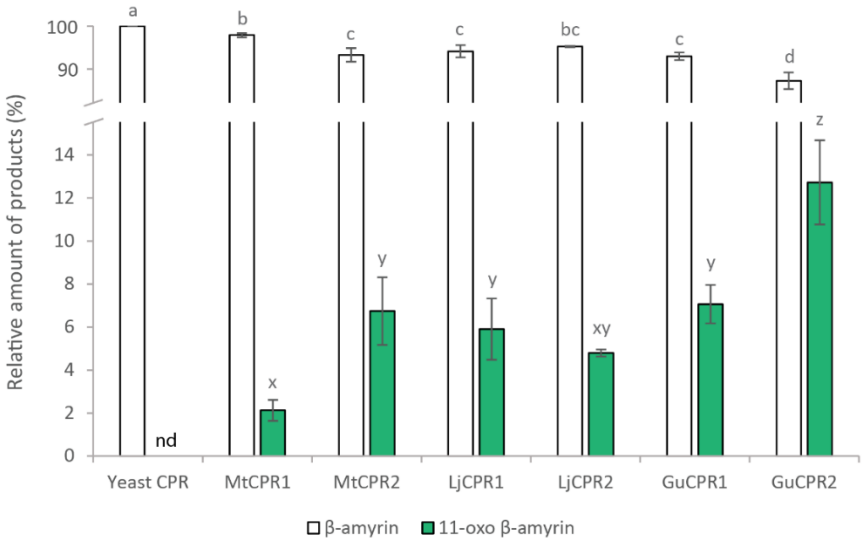

Figure 4-5. Improving 11-oxo-β-amyrin production using the engineered PSIII yeast strain. The relative amount of triterpenoid in PSIII strain co-expressing different combinations of GuCYP88D6 with different CPRs.

11-oxo-β-amyrin content was measured relative to an internal standard (uvaol). Data have been presented as mean ± SE (n=3). nd, signal below detection limit. The different letters indicate significant differences (P<0.05, one-way ANOVA followed by Tukey's test) of the triterpenoid content produced by GuCYP88D6 paired with different CPRs. Letter a-d and x-z are used to indicate significancy for β-amyrin and 11-oxo-β-amyrin content, respectively.

Co-expressing different CYP and CPR classes from model legumes, *M. truncatula* and *L. japonicus* have suggested that CYP shows preference towards certain CPR class. Therefore, to further study the effect of CYP-CPR pairs, we added CPRs from a non-model legume species, *G. uralensis*, which produces glycyrrhizin, a highly valuable metabolite found only in the *Glycyrrhiza sp*.. Furthermore, unlike CYP716As and CYP72As that were commonly found in plant species, GuCYP88D6 uniquely presents in *Glycyrrhiza sp*. and very specific in function, which is interesting to facilitate contrasting results. GuCYP88D6 catalyzes the two-step oxidation at the C-11 position of β-amyrin to produce 11-oxo β-amyrin, the precursor for glycyrrhizin (Seki et al. 2008). The PCC value of GuCYP88D6 with GuCPRs also indicates that GuCYP88D6 might have different preferences for GuCPR classes I and II, which shows negative correlation with GuCPR1

(Table 2-2C). Corresponding to this hypothesis, the results showed that co-expressing GuCYP88D6 with class II GuCPR in the PSIII strain showed the highest conversion of β-amyrin into 11-oxo β-amyrin, as compared to the other CPRs tested in this study (Figure 4-5).

There are several aspects that might affect heterologous gene expression between INV*Sc*1 and PSIII strain. The INV*Sc*1 strain is a diploid, whereas BY4742, the parental strain of PSIII, is a haploid strain. Diploid yeast co-expressing similar plasmids resulted in more stable and approximately two times higher copy number than the haploid strain (Karim et al. 2013). The engineered pathway of PSIII strain also significantly affects triterpenoid production. Benefiting from the GEV chimeric transcriptional activator system, this strain can utilize glucose as a carbon source, while the triterpenoid biosynthetic genes are being activated by β-estradiol (McIsaac et al. 2011). Meanwhile, INV*Sc*1 utilizes galactose as its growth medium to active GAL promoters, which presents the drawback of catabolic repression in the presence of glucose (Escalante-Chong et al. 2015). Upon growing in glucose, the PSIII strain exhibited a much higher growth rate than INV*Sc*1, which might affect its ability to produce triterpenoids. One of the factors affecting enzyme activity is substrate concentration (Zhu et al. 2018). In the PSIII strain, the abundant β-amyrin concentration might drive the CYP-CPR pairs, which otherwise show lower conversion ratio of β-amyrin in the INV*Sc*1 strain. The fact that CYP716As and CYP72As are both widespread in plants and possess diverse functions in triterpenoid site-specific oxidations might be the reason they work well with all CPR classes in the engineered yeast strain.

Despite many variables affecting gene expression and metabolic regulation in yeast, this study achieved the highest conversion of β-amyrin into 11-oxo-β-amyrin by pairing GuCYP88D6 with GuCPR2. Previous research has attempted to screen different CPRs for glycyrrhetinic acid production and reported that GuCPR1 is the best fit for CYP88D6 (Zhu et al. 2018). However, this

study revealed that pairing CYP88D6 with GuCPR2 showed 1.8 times higher conversion ratio of β-amyrin into 11-oxo-β-amyrin than pairing with GuCPR1. This result also reveals that the conversion of β-amyrin in PSIII is still inefficient, as compared to the INV*Sc*1 strain, as shown in the lower ratio of the oxidized products of β-amyrin. This implies that there is still an opportunity to improve triterpenoid production in yeast by increasing the conversion efficiency of β-amyrin.

4.4 Conclusion

This study revealed that pairing different plant CYP families with different plant CPR classes results in different triterpenoid conversion ratio in heterologous yeast. This is the first study to compare different pairs of CPR classes and CYPs from legumes in transgenic yeast. The comparison data of this study did not cover enough CPR-CYP pairs from different species to generalize which CPR class is better for which CYP. Nevertheless, this study highlighted that CYP does have preference towards certain CPR class. A strategy that could be proposed through this study is that performing screening for both CPR classes from the same species as their target CYPs would increase their chance to find the best CPR pair for their heterologous production system. This study cannot generalize for other species yet, but at least for legume species, CYP72A family have higher chances to work better with CPR class II, while CYP716A also has higher probability to work better with CPR class I of the same species. Starting with that statement might help other researchers to find their best CPR pairs.

Chapter 5

Conclusion

In the previous notion, CPR class I is believed to play a more important role in basal and constitutive specialized metabolism, while CPR class II is involved more in defense response by supporting inducible specialized metabolism. This study revealed that both *L. japonicus* CPR class I and II involve in triterpenoids biosynthesis, one of plant specialized metabolites. However, *LjCPR1* and *LjCPR2* showed different correlations with CYPs involved in the triterpenoids pathway, which might be affected by different CYP:CPR binding motifs of CPR class I and II. A strong correlation between *LjCPR1*, *CYP716A51*, and *LUS* genes was observed in which their expressions were similarly down-regulated by MeJA treatment, suggesting *CYP716A51* low involvement in defense response. In contrast, MeJA treatment significantly upregulated *LjCPR2*, *CYP93E2*, and *CYP72A61* expressions, inferring a strong correlation between these stress-inducible genes. Both *LjCPR1* and *LjCPR2* are necessary for betulinic acid and lupeol production. However, only *LjCPR2* plays a vital role in β-amyrin, 24-OH β-amyrin, and sophoradiol production, in which the function cannot be complemented by *LjCPR1*. *LjCPR1* showed a more substantial physiological role in the pod and seed development of *L. japonicus* than *LjCPR2*, which is in line with the previous notion that CPR class I plays a more important role in basal metabolism than CPR class II.

In line with the *in planta* result, co-expressing different legume CPR and CYP in transgenic yeast showed that CYP716A has a better conversion ratio when paired with CPR class I from the same species, while CYP72A has a better conversion ratio when co-expressed with CPR class II. As the most distinct part of CPR class I and II, the N-terminal membrane domain might play an

essential role in CYP:CPR protein interaction since swapping this domain in MtCPR1 and MtCPR2 significantly compromised CYP716A catalytic activity, but not CYP72A. By screening different CPR classes in engineered yeast, the conversion ratio of CYP88D6 was highest when paired with GuCPR2 to produce 11-oxo β-amyrin as the precursor of glycyrrhizin as a high-value compound. Therefore, screening at least the genes of each CPR class from the same species with the CYP of interest is essential in improving the CYP conversion ratio to improve the heterologous triterpenoid production.

However, the underlying mechanisms of how a CPR prefers a specific CYP have not yet been fully elucidated. This study is limited to the legume species and triterpenoids biosynthesis and might not be generalized to other plant families and plant secondary metabolism. Therefore, more CPRs and CYPs from different plant families need to be studied in order to gain a deeper understanding of CPR class preferences towards CYPs. Comprehensive studies regarding the physiological role of different CPR classes in different plant species and other metabolic pathways are required to understand the involvement of CPR as the obligatory redox partner of numerous CYPs. Even though this study revealed a different set of genes that might be correlated differently with each CPR class in triterpenoids biosynthesis, deeper analysis to confirm their protein-protein interactions are required. Protein-protein interaction assay can be performed by heterologous system such as yeast two-hybrid assay or *in planta* system such as protoplast two-hybrid and bimolecular fluorescence complementation assay. The possibility of metabolon structure can also be validated through these methods.

Furthermore, since different conserved amino acids between CPR class I and II were observed, a mutagenesis experiment can be conducted to confirm the importance of these amino acids in CYP:CPR interaction. Protein engineering on these amino acids can also increase

specificity or affinity between a particular CPR and CYP of interest. The effect of CPR and CYP pairs should also be reinvestigated in different heterologous hosts. While quantifying CYP proteins *in vitro* is still challenging work, the molecular ratio between CPR and CYP should be considered to ensure optimum electron supply in the heterologous host. Nevertheless, the information of this study is hoped to be beneficial to drive more profound research on the utilization of plant CPRs in improving the production of various plant natural products.

Acknowledgement

All Praise to ALLAH S.W.T the Almighty, The Most Gracious and The Most Merciful, for giving me the blessing, the strength, the chance, and endurance to complete this study.

I would like to express my gratitude to Prof. Muranaka Toshiya, Assoc. Prof. Ery Odette Fukushima, Assoc. Prof. Seki Hikaru, and Assist. Prof. Yasumoto Shuhei, for their continuous support, guidance, patience, and valuable advices during my study until I can finish my doctoral dissertation. I would also like to thank MEXT for the financial support throughout my study.

I would like to thank my seniors and alumni of Muranaka Laboratory, Much Zaenal Fanani, Pisanee Srisawat, Soo Yeon Chung, Suzuki Hayato, and Jekson Robertlee for their moral support and confidence in me for me to strongly endure the hardship during my study, and for their continuous advices to improve the quality of my research and publication.

To Game, as my partner in the same batch, who has fought with me in this journey together since the beginning and always had my back and supported in me technically and mentally. I am grateful for his kindness, work attitude, optimism, and perseverance that always motivates me to be better. To my beloved husband, Malik, for his continuous support, motivation, patience, and love for me. Words are not enough to describe how grateful I am to have him as my support system during my doctoral study in Japan. To Callan, Jun, Yen, and other Muranaka Laboratory members that always brighten up my days during my study. Finally, I would like to thank my family: papap, mamah, nana, ade, for their continuous support and care, that always become my motivation to finish my study and being a better person every day.

References

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L.K., Saito, K., Kanaya, S.: KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology*, **53**, e1 (2012).

Agrawal, S.Bhushan., Agrawal, Madhoolika.: *Environmental pollution and plant responses*, Boca Raton, Fla.: Lewis Publishers (1999).

Akhgari, A., Laakso, I., Maaheimo, H., Choi, Y.H., Seppänen-Laakso, T., Oksman-Caldentey, K.M., Rischer, H.: Methyljasmonate elicitation increases terpenoid indole alkaloid accumulation in Rhazya stricta hairy root cultures. *Plants*, **8**, 534 (2019).

Anjanappa, R.B., Gruissem, W.: Current progress and challenges in crop genetic transformation. *Journal of Plant Physiology*, **261**, 153411 (2021).

Arnold, K., Bordoli, L., Kopp, J., Schwede, T.: The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201 (2006).

Barnaba, C., Martinez, M.J., Taylor, E., Barden, A.O., Brozik, J.A.: Single-Protein Tracking Reveals That NADPH Mediates the Insertion of Cytochrome P450 Reductase into a Biomimetic of the Endoplasmic Reticulum. *Journal of the American Chemical Society*, **139**, 5420–5430 (2017).

Bart, A.G., Scott, E.E.: Structures of human cytochrome P450 1A1 with bergamottin and erlotinib reveal active-site modifications for binding of diverse ligands. *Journal of Biological Chemistry*, **293**, 19201–19210 (2018).

Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., Scotti, C., Calderini, O.: CYP72A67 catalyzes a key oxidative step in Medicago truncatula hemolytic saponin biosynthesis. *Molecular Plant*, **8**, 1493–1506 (2015).

Bishayee, A., Ahmed, S., Brankov, N., Perloff, M.: Triterpenoids as potential agents for the chemoprevention and therapy of breast cancer. *Frontiers in Bioscience*, **Jan 1**, 980–996 (2011).

Burkhard, F.Z., Parween, S., Udhane, S.S., Flück, C.E., Pandey, A. v.: P450 Oxidoreductase deficiency: Analysis of mutations and polymorphisms. *Journal of Steroid Biochemistry and Molecular Biology*, **165**, 38–50 (2017).

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., Han, T., Hou, W.: CRISPR/Cas9-mediated genome editing in soybean hairy roots. *PLoS ONE*, **10**, e0136064 (2015).

Campelo, D., Esteves, F., Palma, B.B., Gomes, B.C., Rueff, J., Lautier, T., Urban, P., Truan, G., Kranendonk, M.: Probing the role of the hinge segment of cytochrome P450 oxidoreductase in the interaction with cytochrome P450. *International Journal of Molecular Sciences*, **19**, 3914 (2018).

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., Graham, N., Odoardi, M., Piano, E., Arcioni, S., May, S., Scotti, C., Calderini, O.: Medicago truncatula CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell*, **23**, 3070–3081 (2011).

Carelli, M., Calderini, O., Panara, F., Porceddu, A., Losini, I., Piffanelli, P., Arcioni, S., Scotti, C.: Reverse Genetics in Medicago truncatula Using a TILLING Mutant Collection. In R. J. Rose, ed. *Legume Genomics: Methods and Protocols*. (*pp. 101–118*). Totowa, NJ: Humana Press.(2013).

Carsanba, E., Pintado, M., Oliveira, C.: Fermentation strategies for production of pharmaceutical terpenoids in engineered yeast. *Pharmaceuticals*, **14**, 295 (2021).

Chung, S.Y., Seki, H., Fujisawa, Y., Shimoda, Y., Hiraga, S., Nomura, Y., Saito, K., Ishimoto, M., Muranaka, T.: A cellulose synthase-derived enzyme catalyses 3-O-glucuronosylation in saponin biosynthesis. *Nature Communications*, **11**, 1–8 (2020).

Cravens, A., Payne, J., Smolke, C.D.: Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nature Communications*, **10**, 2142 (2019).

DeLano, W.L.: The PyMOL Molecular Graphics System, Version 2.3. *Schrödinger LLC* (2020).

Delis, C., Krokida, A., Georgiou, S., Peña-Rodríguez, L.M., Kavroulakis, N., Ioannou, E., Roussis, V., Osbourn, A.E., Papadopoulou, K.K.: Role of lupeol synthase in *Lotus japonicus* nodule formation. *New Phytologist*, **189**, 335–346 (2011).

Doroodian, P., Hua, Z.: The ubiquitin switch in plant stress response. *Plants*, **10**, 246 (2021).

Du, J.R., Long, F.Y., Chen, C.: Research Progress on Natural Triterpenoid Saponins in the Chemoprevention and Chemotherapy of Cancer. *Enzymes*, **36**, 95–130 (2014).

Durairaj, P., Hur, J.S., Yun, H.: Versatile biocatalysis of fungal cytochrome P450 monooxygenases. *Microbial Cell Factories*, **15**, 125 (2016).

Ebizuka, Y., Shibuya, M., Wakita, E.: C-22 hydroxylase.(2011).

Ebrecht, A.C., van der Bergh, N., Harrison, S.T.L., Smit, M.S., Sewell, B.T., Opperman, D.J.: Biochemical and structural insights into the cytochrome P450 reductase from Candida tropicalis. *Scientific Reports*, **9**, 20088 (2019).

Escalante-Chong, R., Savir, Y., Carroll, S.M., Ingraham, J.B., Wang, J., Marx, C.J., Springer, M.: Galactose metabolic genes in yeast respond to a ratio of galactose and glucose.

Proceedings of the National Academy of Sciences of the United States of America, **112**, 1636–1641 (2015).

Esteves, F., Campelo, D., Gomes, B.C., Urban, P., Bozonnet, S., Lautier, T., Rueff, J., Truan, G., Kranendonk, M.: The Role of the FMN-Domain of Human Cytochrome P450 Oxidoreductase in Its Promiscuous Interactions With Structurally Diverse Redox Partners. *Frontiers in Pharmacology*, **11**, 299 (2020).

Fanani, M.Z., Fukushima, E.O., Sawai, S., Tang, J., Ishimori, M., Sudo, H., Ohyama, K., Seki, H., Saito, K., Muranaka, T.: Molecular Basis of C-30 Product Regioselectivity of Legume Oxidases Involved in High-Value Triterpenoid Biosynthesis. *Frontiers in Plant Science*, **10**, 1–16 (2019).

Finnigan, J.D., Young, C., Cook, D.J., Charnock, S.J., Black, G.W.: Cytochromes P450 (P450s): A review of the class system with a focus on prokaryotic P450s. *Advances in Protein Chemistry and Structural Biology*, **122**, 289–320 (2020).

Fukai, E., Soyano, T., Umehara, Y., Nakayama, S., Hirakawa, H., Tabata, S., Sato, S., Hayashi, M.: Establishment of a *Lotus japonicus* gene tagging population using the exon-targeting endogenous retrotransposon LORE1. *Plant Journal*, **69**, 720–30 (2012).

Fukushima, E.O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., Saito, K., Muranaka, T.: CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant and Cell Physiology*, **52**, 2050–2061 (2011).

Gan, L., von Moltke, L.L., Trepanier, L.A., Harmatz, J.S., Greenblatt, D.J., Court, M.H.: Role of NADPH-cytochrome P450 reductase and cytochrome-b 5/NADH-b5 reductase in variability of CYP3A activity in human liver microsomes. *Drug Metabolism and Disposition*, **37**, 90–6 (2009).

Ghosh, S.: Triterpene structural diversification by plant cytochrome P450 enzymes. *Frontiers in Plant Science*, **8**, 1886 (2017).

Guo, Q., Yoshida, Y., Major, I.T., Wang, K., Sugimoto, K., Kapali, G., Havko, N.E., Benning, C., Howe, G.A.: JAZ repressors of metabolic defense promote growth and reproductive fitness in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E10768–E10777 (2018).

Hamdane, D., Xia, C., Im, S.C., Zhang, H., Kim, J.J.P., Waskell, L.: Structure and function of an NADPH-cytochrome P450 oxidoreductase in an open conformation capable of reducing cytochrome P450. *Journal of Biological Chemistry*, **284**, 11374–11384 (2009).

Hamdane, D., Zhang, H., Hollenberg, P.: Oxygen activation by cytochrome P450 monooxygenase. *Photosynthesis Research*, **98**, 657–666 (2008).

Haque, M.M., Fadlalla, M.A., Aulak, K.S., Ghosh, A., Durra, D., Stuehr, D.J.: Control of electron transfer and catalysis in neuronal nitric-oxide synthase (nNOS) by a hinge connecting its FMN and FAD-NADPH domains. *Journal of Biological Chemistry*, **287**, 30105–30116 (2012).

Hashimoto, R., Ueta, R., Abe, C., Osakabe, Y., Osakabe, K.: Efficient multiplex genome editing induces precise, and self-ligated type mutations in tomato plants. *Frontiers in Plant Science*, **9**, 1–10 (2018).

Hedison, T.M., Scrutton, N.S.: Tripping the light fantastic in membrane redox biology: linking dynamic structures to function in ER electron transfer chains. *FEBS Journal*, **286**, 2004–2017 (2019).

Hong, Y., Li, H., Yuan, Y.C., Chen, S.: Sequence-function correlation of aromatase and its interaction with reductase. In *Journal of Steroid Biochemistry and Molecular Biology*. (*pp. 203–206*).(2010).

Huang, F.C., Sung, P.H., Do, Y.Y., Huang, P.L.: Differential expression and functional characterization of the NADPH cytochrome P450 reductase genes from Nothapodytes foetida. *Plant Science*, **190**, 16–23 (2012).

Huang, R., Liu, L., He, X., Wang, W., Hou, Y., Chen, J., Li, Y., Zhou, H., Tian, T., Wang, W., Xu, Q., Yu, Y., Zhou, T.: Isolation and Functional Characterization of Multiple NADPH-Cytochrome P450 Reductase Genes from Camellia sinensis in View of Catechin Biosynthesis. *Journal of Agricultural and Food Chemistry*, **69**, 14926–14937 (2021).

Istiandari, P., Yasumoto, S., Srisawat, P., Tamura, K., Chikugo, A., Suzuki, H., Seki, H., Fukushima, E.O., Muranaka, T.: Comparative Analysis of NADPH-Cytochrome P450 Reductases From Legumes for Heterologous Production of Triterpenoids in Transgenic Saccharomyces cerevisiae. *Frontiers in Plant Science*, **12**, 762546 (2021).

Jang, H.H., Jamakhandi, A.P., Sullivan, S.Z., Yun, C.H., Hollenberg, P.F., Miller, G.P.: Beta sheet 2-alpha helix C loop of cytochrome P450 reductase serves as a docking site for redox partners. *Biochimica et Biophysica Acta - Proteins and Proteomics*, **1804**, 1285–93 (2010).

Jensen, K., Møller, B.L.: Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry*, **71**, 132–141 (2010).

Jin, C.C., Zhang, J.L., Song, H., Cao, Y.X.: Boosting the biosynthesis of betulinic acid and related triterpenoids in Yarrowia lipolytica via multimodular metabolic engineering. *Microbial Cell Factories*, **18**, 77 (2019).

Kamal, N., Mun, T., Reid, D., Lin, J.S., Akyol, T.Y., Sandal, N., Asp, T., Hirakawa, H., Stougaard, J., Mayer, K.F.X., Sato, S., Andersen, S.U.: Insights into the evolution of

symbiosis gene copy number and distribution from a chromosome-scale *Lotus japonicus* Gifu genome sequence. *DNA Research*, **27**, dsaa015 (2020).

Kamo, S., Suzuki, S., Sato, T.: The content of soyasaponin and soyasapogenol in soy foods and their estimated intake in the Japanese. *Food Science and Nutrition*, **2**, 289–297 (2014).

Karim, A.S., Curran, K.A., Alper, H.S.: Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. *FEMS Yeast Research*, **13**, 107–116 (2013).

Kim, O.T., Um, Y., Jin, M.L., Kim, J.U., Hegebarth, D., Busta, L., Racovita, R.C., Jetter, R.: A novel multifunctional C-23 Oxidase, CYP714E19, is involved in asiaticoside biosynthesis. *Plant and Cell Physiology*, **59**, 1200–1213 (2018).

Kitagawa, I.: Licorice root. A natural sweetener and an important ingredient in Chinese medicine. *Pure and Applied Chemistry*, **74**, 1189–1198 (2002).

Kuromori, T., Takahashi, S., Kondou, Y., Shinozaki, K., Matsui, M.: Phenome analysis in plant species using loss-of-function and gain-of-function mutants. *Plant and Cell Physiology*, **50**, 1215–31 (2009).

Laursen, T., Jensen, K., Møller, B.L.: Conformational changes of the NADPH-dependent cytochrome P450 reductase in the course of electron transfer to cytochromes P450. *Biochimica et Biophysica Acta - Proteins and Proteomics*, **1814**, 132–8 (2011).

Lee, G.Y., Kim, H.M., Ma, S.H., Park, S.H., Joung, Y.H., Yun, C.H.: Heterologous expression and functional characterization of the NADPH-cytochrome P450 reductase from Capsicum annuum. *Plant Physiology and Biochemistry*, **82**, 116–22 (2014).

Lin, C.Y., Sun, Y., Song, J., Chen, H.C., Shi, R., Yang, C., Liu, J., Tunlaya-Anukit, S., Liu, B., Loziuk, P.L., Williams, C.M., Muddiman, D.C., Lin, Y.C.J., Sederoff, R.R., Wang, J.P., Chiang, V.L.: Enzyme Complexes of Ptr4CL and PtrHCT Modulate Co-enzyme A Ligation of Hydroxycinnamic Acids for Monolignol Biosynthesis in Populus trichocarpa. *Frontiers in Plant Science*, **12**, 727932 (2021).

Małolepszy, A., Mun, T., Sandal, N., Gupta, V., Dubin, M., Urbański, D., Shah, N., Bachmann, A., Fukai, E., Hirakawa, H., Tabata, S., Nadzieja, M., Markmann, K., Su, J., Umehara, Y., Soyano, T., Miyahara, A., Sato, S., Hayashi, M., et al.: The LORE1 insertion mutant resource. *Plant Journal*, **88**, 306–317 (2016).

Manoj, K.M., Gade, S.K., Mathew, L.: Cytochrome P450 reductase: A harbinger of diffusible reduced oxygen species. *PLoS ONE*, **5**, e13272 (2010).

Mazourek, M., Pujar, A., Borovsky, Y., Paran, I., Mueller, L., Jahn, M.M.: A dynamic interface for capsaicinoid systems biology. *Plant Physiology*, **150**, 1806–21 (2009).

McIsaac, R.S., Silverman, S.J., McClean, M.N., Gibney, P.A., Macinskas, J., Hickman, M.J., Petti, A.A., Botstein, D.: Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae. *Molecular Biology of the Cell*, **22**, 4447–4459 (2011).

Mishin, V., Heck, D.E., Laskin, D.L., Laskin, J.D.: Human recombinant cytochrome P450 enzymes display distinct hydrogen peroxide generating activities during substrate independent NADPH oxidase reactions. *Toxicological Sciences*, **141**, 344–52 (2014).

Misra, R.C., Maiti, P., Chanotiya, C.S., Shanker, K., Ghosh, S.: Methyl jasmonate-elicited transcriptional responses and pentacyclic triterpene biosynthesis in sweet basil. *Plant Physiology*, **164**, 1028–1044 (2014).

Misra, R.C., Sharma, S., Sandeep, S., Garg, A., Chanotiya, C.S., Ghosh, S.: Two CYP716A subfamily cytochrome P450 monooxygenases of sweet basil play similar but nonredundant roles in ursane- and oleanane-type pentacyclic triterpene biosynthesis. *New Phytologist*, **214**, 706–720 (2017).

Mizutani, M., Ohta, D.: Two isoforms of NAPDH:cytochrome p450 reductase in Arabidopsis thaliana gene structure, heterologous expression in insect cells, and differential regulation. *Plant Physiology*, **116**, 357–367 (1998).

Moktali, V., Park, J., Fedorova-Abrams, N.D., Park, B., Choi, J., Lee, Y.H., Kang, S.: Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC Genomics*, **13**, 525 (2012).

Morikawa, T., Mizutani, M., Aoki, N., Watanabe, B., Saga, H., Saito, S., Oikawa, A., Suzuki, H., Sakurai, N., Shibata, D., Wadano, A., Sakata, K., Ohta, D.: Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato. *Plant Cell*, **18**, 1008–1022 (2006).

Moses, T., Mehrshahi, P., Smith, A.G., Goossens, A.: Synthetic biology approaches for the production of plant metabolites in unicellular organisms. *Journal of Experimental Botany*, **68**, 4057–4074 (2017).

Moses, T., Pollier, J., Thevelein, J.M., Goossens, A.: Bioengineering of plant (tri)terpenoids: From metabolic engineering of plants to synthetic biology in vivo and in vitro. *New Phytologist*, **200**, 27–43 (2013).

Mucha, S., Heinzlmeir, S., Kriechbaumer, V., Strickland, B., Kirchhelle, C., Choudhary, M., Kowalski, N., Eichmann, R., Hückelhoven, R., Grill, E., Kuster, B., Glawischnig, E.: The Formation of a Camalexin Biosynthetic Metabolon. *The Plant cell*, **31**, 2697–2710 (2019).

Mukherjee, G., Nandekar, P.P., Wade, R.C.: Electron transfer from cytochrome P450 reductase to cytochrome P450: Towards a structural and dynamic understanding. *Communications Biology*, **4**, 55 (2021).

Naito, Y., Hino, K., Bono, H., Ui-Tei, K.: CRISPRdirect: Software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*, **31**, 1120–1123 (2015).

Nakayasu, M., Akiyama, R., Lee, H.J., Osakabe, K., Osakabe, Y., Watanabe, B., Sugimoto, Y., Umemoto, N., Saito, K., Muranaka, T., Mizutani, M.: Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiology and Biochemistry*, **131**, 70–77 (2018).

Nakazono, M.: Investigation of Enzymes for Biosynthesis of Triterpenoids (Lupeol and Betulinic acid) in Soybean. *Soy Protein Research*, **17**, 2050–2061 (2014).

Nielsen, K.A., Tattersall, D.B., Jones, P.R., Møller, B.L.: Metabolon formation in dhurrin biosynthesis. *Phytochemistry*, **69**, 88–98 (2008).

Niu, G., Zhao, S., Wang, L., Dong, W., Liu, L., He, Y.: Structure of the Arabidopsis thaliana NADPH-cytochrome P450 reductase 2 (ATR2) provides insight into its function. *FEBS Journal*, **284**, 754–765 (2017).

Okamoto, S., Yoro, E., Suzaki, T., Kawaguchi, M.: Hairy Root Transformation in *Lotus japonicus*. *BIO-PROTOCOL*, **3** (2013).

Oldfield, E., Lin, F.Y.: Terpene biosynthesis: Modularity rules. *Angewandte Chemie - International Edition*, **51**, 1124–1137 (2012).

Paine, M.J.I., Scrutton, N.S., Munro, A.W., Gutierrez, A., Roberts, G.C.K., Roland Wolf, C.: Electron transfer partners of cytochrome P450. In P. R. de Montellano, ed. *Cytochrome P450: Structure, Mechanism, and Biochemistry: Third edition*. (*pp. 115–148*). Boston, MA: Springer US.(2005).

Pandey, A. v., Flück, C.E.: NADPH P450 oxidoreductase: Structure, function, and pathology of diseases. *Pharmacology and Therapeutics*, **138**, 229–54 (2013).

Pandian, B.A., Sathishraj, R., Djanaguiraman, M., Prasad, P.V.V., Jugulam, M.: Role of cytochrome P450 enzymes in plant stress response. *Antioxidants*, **9**, 454 (2020).

Parage, C., Foureau, E., Kellner, F., Burlat, V., Mahroug, S., Lanoue, A., de Bernonville, T.D., Londono, M.A., Carqueijeiro, I., Oudin, A., Besseau, S., Papon, N., Glévarec, G., Atehortùa, L., Giglioli-Guivarc'h, N., St-Pierre, B., Clastre, M., O'Connor, S.E., Courdavault, V.: Class II cytochrome P450 reductase governs the biosynthesis of alkaloids. *Plant Physiology*, **172**, 1563–1577 (2016).

Porter, T.D., Beck, T.W., Kasper, C.B.: NADPH-Cytochrome P-450 Oxidoreductase Gene Organization Correlates with Structural Domains of the Protein. *Biochemistry*, **29**, 9814–9818 (1990).

Qiu, X.M., Sun, Y.Y., Ye, X.Y., Li, Z.G.: Signaling Role of Glutamate in Plants. *Frontiers in Plant Science*, **10**, 1743 (2020).

Qu, X., Pu, X., Chen, F., Yang, Y., Yang, L., Zhang, G., Luo, Y.: Molecular cloning, heterologous expression, and functional characterization of an NADPH-cytochrome P450 reductase gene from Camptotheca acuminata, a camptothecin-producing plant. *PLoS ONE*, **10**, 1–19 (2015).

Quehl, P., Schüürmann, J., Hollender, J., Jose, J.: Improving the activity of surface displayed cytochrome P450 enzymes by optimizing the outer membrane linker. *Biochimica et Biophysica Acta - Biomembranes*, **1859**, 104–116 (2017).

Radhamony, R.N., Prasad, A.M., Srinivasan, R.: T-DNA insertional mutagenesis in Arabidopsis: A tool for functional genomics. *Electronic Journal of Biotechnology*, **8** (2005).

Rana, S., Lattoo, S.K., Dhar, N., Razdan, S., Bhat, W.W., Dhar, R.S., Vishwakarma, R.: NADPH-Cytochrome P450 Reductase: Molecular Cloning and Functional Characterization of Two Paralogs from Withania somnifera (L.) Dunal. *PLoS ONE*, **8**, e57068 (2013).

Rao, K.M., Raghavendra, A., Reddy, K.J.: *Physiology and molecular biology of stress tolerance in plants*, Dordrecht: Springer (2006).

Ritacco, I., Spinello, A., Ippoliti, E., Magistrato, A.: Post-Translational Regulation of CYP450s Metabolism As Revealed by All-Atoms Simulations of the Aromatase Enzyme. *Journal of Chemical Information and Modeling*, **59**, 2930–2940 (2019).

Ro, D.K., Ehlting, J., Douglas, C.J.: Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar. *Plant Physiology*, **130**, 1837–1851 (2002).

Santos, J., Risso, V.A., Sica, M.P., Ermácora, M.R.: Effects of serine-to-cysteine mutations on β-lactamase folding. *Biophysical Journal*, **93**, 1707–1718 (2007).

Scarpella, E., Meijer, A.H.: Pattern formation in the vascular system of monocot and dicot plant species. *New Phytologist*, **164**, 209-242 (2004).

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., Akashi, T., Aoki, T., Saito, K., Muranaka, T.: Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 14204–14209 (2008).

Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T., Sudo, H., Fukushima, E.O., Akashi, T., Aoki, T., Saito, K., Muranaka, T.: Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell*, **23**, 4112–4123 (2011).

Seki, H., Tamura, K., Muranaka, T.: P450s and UGTs: Key players in the structural diversity of triterpenoid saponins. *Plant and Cell Physiology*, **56**, 1463–71 (2015).

Shannon, J.C., Pien, F.M., Cao, H., Liu, K.C.: Brittle-1, an adenylate translocator, facilitates transfer of extraplastidial synthesized ADP-glucose into amyloplasts of maize endosperms. *Plant Physiology*, **117**, 1235–1252 (1998).

Sheng, H., Sun, H.: Synthesis, biology and clinical significance of pentacyclic triterpenes: A multi-target approach to prevention and treatment of metabolic and vascular diseases. *Natural Product Reports*, **28**, 543–93 (2011).

Shephard, E.A., Phillips, I.R., Bayney, R.M., Pike, S.F., Rabin, B.R.: Quantification of NADPH:cytochrome P-450 reductase in liver microsomes by a specific radioimmunoassay technique. *Biochemical Journal*, **211**, 333–340 (1983).

Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T., Ebizuka, Y.: Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. *FEBS Journal*, **273**, 948–59 (2006).

Siddiqui, M.S., Thodey, K., Trenchard, I., Smolke, C.D.: Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Research*, **12**, 144–170 (2012).

Simtchouk, S., Eng, J.L., Meints, C.E., Makins, C., Wolthers, K.R.: Kinetic analysis of cytochrome P450 reductase from Artemisia annua reveals accelerated rates of NADH-dependent flavin reduction. *FEBS Journal*, **280**, 6627–42 (2013).

Soyars, C.L., Peterson, B.A., Burr, C.A., Nimchuk, Z.L.: Cutting edge genetics: Crispr/cas9 editing of plant genomes. *Plant and Cell Physiology*, **59**, 1608–1620 (2018).

Srisawat, P., Yasumoto, S., Fukushima, E.O., Robertlee, J., Seki, H., Muranaka, T.: Production of the bioactive plant-derived triterpenoid morolic acid in engineered Saccharomyces cerevisiae. *Biotechnology and Bioengineering*, **117**, 2198–2208 (2020).

Sundin, L., Vanholme, R., Geerinck, J., Goeminne, G., Höfer, R., Kim, H., Ralph, J., Boerjan, W.: Mutation of the inducible ARABIDOPSIS THALIANA CYTOCHROME P450 REDUCTASE2 alters lignin composition and improves saccharification. *Plant Physiology*, **166**, 1956–1971 (2014).

Suzuki, H., Fukushima, E.O., Shimizu, Y., Seki, H., Fujisawa, Y., Ishimoto, M., Osakabe, K., Osakabe, Y., Muranaka, T.: *Lotus japonicus* Triterpenoid Profile and Characterization of

the CYP716A51 and LjCYP93E1 Genes Involved in Their Biosynthesis in Planta. *Plant and Cell Physiology*, **60**, 2496–2509 (2019).

Suzuki, H., Fukushima, E.O., Umemoto, N., Ohyama, K., Seki, H., Muranaka, T.: Comparative analysis of CYP716A subfamily enzymes for the heterologous production of C-28 oxidized triterpenoids in transgenic yeast. *Plant Biotechnology*, **35**, 131–139 (2018).

Tamura, K., Seki, H., Suzuki, H., Kojoma, M., Saito, K., Muranaka, T.: CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of Glycyrrhiza uralensis. *Plant Cell Reports*, **36**, 437–445 (2017).

Urban, P., Mignotte, C., Kazmaier, M., Delorme, F., Pompon, D.: Cloning, yeast expression, and characterization of the coupling of two distantly related Arabidopsis thaliana NADPH-cytochrome P450 reductases with P450 CYP73A5. *Journal of Biological Chemistry*, **272**, 19176–19186 (1997).

Urbański, D.F., Małolepszy, A., Stougaard, J., Andersen, S.U.: Genome-wide LORE1 retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *Plant Journal*, **69**, 731–41 (2012).

Urbański, D.F., Małolepszy, A., Stougaard, J., Andersen, S.U.: High-throughput and targeted genotyping of *Lotus japonicus* LORE1 insertion mutants. *Methods in Molecular Biology*, **1069**, 119–46 (2013).

Vo, N.N.Q., Fukushima, E.O., Muranaka, T.: Structure and hemolytic activity relationships of triterpenoid saponins and sapogenins. *Journal of Natural Medicines*, **71**, 50–58 (2017).

Vranová, E., Coman, D., Gruissem, W.: Structure and dynamics of the isoprenoid pathway network. *Molecular Plant*, **5**, 318-333 (2012).

Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S.S., Kim, J.J.P.: Three-dimensional structure of NADPH-cytochrome P450 reductase: Prototype for FMN- and FAD-containing enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 8411–8416 (1997).

Waskell, L., Jung-Ja P. Kim: Electron transfer partners of cytochrome P450. In P. R. O. De Montellano, ed. *Cytochrome P450*. (*pp. 33–68*). Basel: Springer.(2015).

Whitelaw, D.A., Tonkin, R., Meints, C.E., Wolthers, K.R.: Kinetic analysis of electron flux in cytochrome P450 reductases reveals differences in rate-determining steps in plant and mammalian enzymes. *Archives of Biochemistry and Biophysics*, **584**, 107–15 (2015).

Wright, T.R., Shan, G., Walsha, T.A., Lira, J.M., Cui, C., Song, P., Zhuang, M., ArnoldN, N.L., Lin, G., Yau, K., Russell, S.M., Cicchillo, R.M., Peterson, M.A., Simpson, D.M., Zhou, N., Ponsamuel, J., Zhang, Z.: Robust crop resistance to broadleaf and grass

herbicides provided by aryloxyalkanoate dioxygenase transgenes. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 20240–20245 (2010).

Xu, R., Fazio, G.C., Matsuda, S.P.T.: On the origins of triterpenoid skeletal diversity. *Phytochemistry*, **65**, 261–291 (2004).

Yan, D., Duermeyer, L., Leoveanu, C., Nambara, E.: The functions of the endosperm during seed germination. *Plant and Cell Physiology*, **55**, 1521–33 (2014).

Yang, C.Q., Lu, S., Mao, Y.B., Wang, L.J., Chen, X.Y.: Characterization of two NADPH: Cytochrome P450 reductases from cotton (Gossypium hirsutum). *Phytochemistry*, **71**, 27–35 (2010).

Yasumoto, S., Fukushima, E.O., Seki, H., Muranaka, T.: Novel triterpene oxidizing activity of Arabidopsis thaliana CYP716A subfamily enzymes. *FEBS Letters*, **590**, 533–540 (2016).

Zhang, C., Hong, K.: Production of Terpenoids by Synthetic Biology Approaches. *Frontiers in Bioengineering and Biotechnology*, **8**, 347 (2020).

Zhang, H., Cao, Y., Zhang, H., Xu, Y., Zhou, C., Liu, W., Zhu, R., Shang, C., Li, J., Shen, Z., Guo, S., Hu, Z., Fu, C., Sun, D.: Efficient Generation of CRISPR/Cas9-Mediated Homozygous/Biallelic Medicago truncatula Mutants Using a Hairy Root System. *Frontiers in Plant Science*, **11**, 294 (2020).

Zhang, Y., Fernie, A.R.: Metabolons, enzyme–enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism. *Plant Communications*, **2**, 100081 (2021).

Zhu, M., Wang, C., Sun, W., Zhou, A., Wang, Y., Zhang, G., Zhou, X., Huo, Y., Li, C.: Boosting 11-oxo-β-amyrin and glycyrrhetinic acid synthesis in Saccharomyces cerevisiae via pairing novel oxidation and reduction system from legume plants. *Metabolic Engineering*, **45**, 43–50 (2018).

List of publications

Original paper

Istiandari, P., Yasumoto, S., Srisawat, P., Tamura, K., Chikugo, A., Suzuki, H., Seki, H., Fukushima, E. O., & Muranaka, T. (2021). Comparative Analysis of NADPH-Cytochrome P450 Reductases From Legumes for Heterologous Production of Triterpenoids in Transgenic Saccharomyces cerevisiae. Frontiers in Plant Science, 12. https://doi.org/10.3389/fpls.2021.762546.

Presentations:

1. **Istiandari, P.,** Fukushima, E. O., Yasumoto, S., H., Seki, H., & Muranaka, T. Comparative analysis of plant NADPH-cytochrome P450 reductase classes of legumes towards triterpenoids biosynthesis. (Oral) *62nd Annual Meeting of JSPP in Matsue (online). 2021.*

2. **Istiandari, P.,** Fukushima, E. O., Yasumoto, S., H., Seki, H., & Muranaka, T. Comparative analysis of plant NADPH-cytochrome P450 reductases towards heterologous triterpenoid production in transgenic yeast. (Oral presentation) *The 22nd Academic Exchange Seminar Between Shanghai Jiao Tong University and Osaka University (online). 2020.*

3. **Istiandari, P.,** Fukushima, E. O., Yasumoto, S., H., Seki, H., & Muranaka, T. Comparative analysis of plant NADPH-cytochrome P450 reductases towards heterologous triterpenoid production in transgenic yeast. (Poster presentation) *女子大学院生と企業等との交流会. 2019. Osaka, Japan.*

4. **Istiandari, P.,** Fukushima, E. O., Yasumoto, S., H., Seki, H., & Muranaka, T. Comparative analysis of plant NADPH-cytochrome P450 reductases towards heterologous triterpenoid production in transgenic yeast. (Oral presentation) *37th Annual Meeting of the Japanese Society for Plant Cell and Molecular Biology. 2019. Kyoto, Japan.*

Supplementary

Table S1. List of accession numbers of CPR genes and amino acid sequences used for phylogenetic analysis and multiple sequence alignment in this study. Yellow color shows genes not available in NCBI.

Gene	Gene accession number / Gene ID	Protein accession number	Source
Arabidopsis thaliana CPR1 (ATR1)	NM_118585.4	NP_194183.1	NCBI
Arabodipsis thaliana CPR2 (ATR2)	NM_119167.4	NP_194750.1	NCBI
M. truncatula CPR1	XM_003602850.3	XP_003602898.1	NCBI
M. truncatula CPR2	XM_003610061.4	XP_003610109.1	NCBI
L. japonicus CPR1	Lj1g3v1548790.1	-	lotus.au.dk (Miyakojima MG20 v 3.0)
	LotjaGi1g1v0345200	-	lotus.au.dk (Gifu v1.2)
L. japonicus CPR2-1	AB433810.1	BAG68945.1	NCBI
	Lj4g3v2107200 / Lj0g3v0139899	-	lotus.au.dk (Miyakojima MG20 v 3.0)
	LotjaGi4g1v0301400	-	lotus.au.dk (Gifu v1.2)
L. japonicus CPR2-2	LotjaGi4g1v0301300	-	lotus.au.dk (Gifu v1.2)
G. uralensis CPR1	KY798117.1	AUG98241.1	NCBI
G. uralensis CPR2	MH401048.1	QCZ35624.1	NCBI
Cicer arietinum CPR1	XM_004501597.3	XP_004501654.1	NCBI
Cicer arietinum CPR2	XM_004507801.3	XP_004507858.1	NCBI
Chenopodium quinoa CPR1	XM_021904070.1	XP_021759762.1	NCBI
Chenopodium quinoa CPR2	XM_021867713.1	XP_021723405.1	NCBI
Spinacia oleracea CPR1	XM_021999727.1	XP_021855419.1	NCBI

Spinacia oleracea CPR2	XM_022003966.1	XP_021859658.1	NCBI
Solanum lycopersicum CPR1	XM_004237953.3	XP_004238001.1	NCBI
Solanum lycopersicum CPR2	XM_004242883.4	XP_004242931.1	NCBI
Solanum tuberosum CPR1	XM_006337990.2	XP_006338052.1	NCBI
Solanum tuberosum CPR2	PGSC0003DMT400035801	-	www.plantgdb.org
Artemisia annua CPR1	PKPP01006895.1 (Whole genome sequence)	PWA55016.1	NCBI
Artemisia annua CPR2	EF104642.1	EF104642.1	NCBI
Catharantus roseus CPR2	X69791.1	CAA49446.1	
Oryza sativa CPR2a	CM000134.1	EAZ10065.1	NCBI
Oryza sativa CPR2b	AP008214.2	BAF23260.1	NCBI
Oryza sativa CPR2c	AL606690.3	CAE03554.2	NCBI
Triticum aestivum CPR2a	AJ303373.1	CAC83301.1	NCBI
Triticum aestivum CPR2c	AF123610.1	AAG17471.1	NCBI
Zea Mays CPR2b 1	EU955593.1	ACG27711.1	NCBI
Zea Mays CPR2b 2	EU956822.1	ACG28940.1	NCBI
Zea Mays CPR2c	BT061122.1	ACN25819.1	NCBI
Pseudotsuga menziesii CPR	CAA89837.3	Z49767.3	NCBI
Taxus chinensis CPR	AAX59902.1	AY959320.1	NCBI
Taxus cuspidata CPR	AAT76449.1	AY571340.1	NCBI
Physcomitrella patens CPR	EDQ49310.1	DS545408.1	NCBI
Selaginella moellendorffii CPR	XP_002978784.2	XM_002978738.2	NCBI
Chlamydomonas reinhardtii CPR	XP_042928682.1	XM_043058768.1	NCBI
Human CPR	NM_001395413.1	NP_001382342.1	NCBI

Table S2. Similarity matrix of amino acid sequences of different CPR classes from different plant families.

CPR comparison		Average Similarity (%)	Max (%)	Min (%)
Fabales CPR-I	Fabales CPR-I	84	93	73
Fabales CPR-II	Fabales CPR-II	84	91	79
Fabales CPR-I	Fabales CPR-II	63	66	56
Amaranthaceae CPR-I	Amaranthaceae CPR-I	92	93	91
Amaranthaceae CPR-II	Amaranthaceae CPR-II	92	93	91
Amaranthaceae CPR-I	Amaranthaceae CPR-II	65	66	64
Fabales CPR-I	Amaranthaceae CPR-I	76	79	66
Fabales CPR-II	Amaranthaceae CPR-II	74	76	73
Fabales CPR-I/2	Amaranthaceae CPR-II/1	65	67	57
Fabales CPR-I	Other species CPR-I	75	79	64
Fabales CPR-II	Other species CPR-II	71	75	38
Fabales CPR-I/II	Other species CPR-II/I	62	66	51

Table S3. Correlation strength between *LjCPR1* and *LjCPR2s* with different triterpenoids biosynthetic enzymes using *L. japonicus* database Gifu v1.2 genome version (lotus.au.dk).

Gene ID	Gene	*LjCPR1*	*LjCPR2-2*	*LjCPR2-1*	*CYP716A51*	*CYP72A61*	*CYP93E3*	*bAS*	*LUS*	*αAS*
LotjaGi1g1v0345200	*LjCPR1*	1.00								
LotjaGi4g1v0301300	*LjCPR2.2*	0.07	1.00							
LotjaGi4g1v0301400	*LjCPR2.1*	0.01	0.59	1.00						
LotjaGi4g1v0438900	*CYP716A51*	0.44	-0.31	-0.39	1.00					
LotjaGi3g1v0557600	*CYP72A61*	-0.25	-0.18	-0.33	0.16	1.00				
LotjaGi1g1v0588600	*CYP93E3*	0.22	-0.37	-0.42	0.52	0.77	1.00			
LotjaGi3g1v0340000	*bAS*	-0.14	-0.18	-0.02	0.05	0.74	0.69	1.00		
LotjaGi2g1v0235400	*LUS*	0.37	-0.23	-0.34	0.78	-0.04	0.16	-0.21	1.00	
LotjaGi3g1v0537500	*αAS*	0.05	0.91	0.58	-0.27	-0.21	-0.38	-0.14	-0.19	1.00

Table S4. Co-expression analysis of closely correlated genes with CPR class I and II in different tissues of a) *M. truncatula*, b) *L. japonicus*, and c) *G. uralensis.*

a) *M. truncatula*

No.	***MtCPR1*** **(**Mtr.10548.1.S1_at)			***MtCPR2*** **(**Mtr.16806.1.S1_at)		
	Probeset	**PCC value**	**GO annotation**	**Probeset**	**PCC value**	**GO annotation**
1	Mtr.1520.1.S1_at	0.92	GO:0008150 (biological_process), GO:0003674 (molecular_function)	Mtr.34798.1.S1_at	0.87	GO:0016567 (protein ubiquitination), GO:0005488 (binding), GO:0004842 (ubiquitin-protein ligase activity)
2	Mtr.30028.1.S1_at	0.89	GO:0005515 (protein binding), GO:0006499 (N-terminal protein myristoylation), GO:0008026 (ATP-dependent helicase activity)	Mtr.8676.1.S1_s_at	0.86	GO:0005575 (cellular_component), GO:0008150 (biological_process), GO:0006952 (defense response), GO:0005524 (ATP binding), GO:0005515 (protein binding)
3	Mtr.8673.1.S1_at	0.89	GO:0008152 (metabolic process), GO:0008194 (UDP-glycosyltransferase activity), GO:0016757 (transferase activity, transferring glycosyl groups), GO:0035251 (UDP-glucosyltransferase activity), GO:0010294 (abscisic acid glucosyltransferase activity)	Mtr.14547.1.S1_at	0.84	GO:0008150 (biological_process), GO:0003824 (catalytic activity), GO:0016208 (AMP binding), GO:0008152 (metabolic process), GO:0003824 (catalytic activity)

4	Mtr.14582.1.S1_at	0.88	GO:0019787 (small conjugating protein ligase activity), GO:0005515 (protein binding), GO:0005634 (nucleus), GO:0005737 (cytoplasm), GO:0008150 (biological_process), GO:0005515 (protein binding), GO:0008270 (zinc ion binding)	Mtr.40581.1.S1_at	0.84	GO:0009816 (defense response to bacterium, incompatible interaction), GO:0009817 (defense response to fungus, incompatible interaction), GO:0005515 (protein binding)
5	Mtr.44464.1.S1_at	0.87	GO:0006414 (translational elongation), GO:0005739 (mitochondrion), GO:0003746 (translation elongation factor activity), GO:0006414 (translational elongation), GO:0003746 (translation elongation factor activity), GO:0008135 (translation factor activity, nucleic acid binding)	Mtr.8656.1.S1_s_at	0.84	GO:0003837 (beta-ureidopropionase activity), GO:0006807 (nitrogen compound metabolic process), GO:0003837 (beta-ureidopropionase activity)
6	Mtr.3434.1.S1_at	0.87	GO:0008284 (positive regulation of cell proliferation), GO:0045941 (positive regulation of transcription)	Mtr.40166.1.S1_s_at	0.84	GO:0009699 (phenylpropanoid biosynthetic process), GO:0045548 (phenylalanine ammonia-lyase activity)
7	Mtr.10292.1.S1_at	0.87	GO:0008150 (biological_process), GO:0005488 (binding), GO:0005515 (protein binding)	Mtr.33344.1.S1_at	0.83	GO:0006952 (defense response), GO:0005524 (ATP binding), GO:0005515 (protein binding),

						GO:0004888 (transmembrane receptor activity)
8	Mtr.51424.1.S1_at	0.86	GO:0019684 (photosynthesis, light reaction), GO:0010207 (photosystem II assembly), GO:0016168 (chlorophyll binding)	Mtr.37895.1.S1_at	0.83	GO:0009617 (response to bacterium), GO:0009620 (response to fungus), GO:0009816 (defense response to bacterium, incompatible interaction), GO:0009817 (defense response to fungus, incompatible interaction), GO:0005515 (protein binding)
9	Mtr.31523.1.S1_at	0.86	GO:0006350 (transcription), GO:0003899 (DNA-directed RNA polymerase activity)	Mtr.11218.1.S1_s_at	0.83	GO:0005524 (ATP binding), GO:0005576 (extracellular region), GO:0005886 (plasma membrane), GO:0006468 (protein amino acid phosphorylation), GO:0019199 (transmembrane receptor protein kinase activity), GO:0006499 (N-terminal protein myristoylation), GO:0006468 (protein amino acid phosphorylation), GO:0016301 (kinase activity)
10	Mtr.10754.1.S1_at	0.86	GO:0030612 (arsenate reductase (thioredoxin) activity), GO:0009793 (embryonic development ending in seed dormancy),	Mtr.27884.1.S1_at	0.83	GO:0004674 (protein serine/threonine kinase activity), GO:0005887 (integral to plasma membrane), GO:0009737

			GO:0003674 (molecular_function)			(response to abscisic acid stimulus), GO:0016301 (kinase activity)

b) *L. japonicus*

No.	***LjCPR1*** (Ljwgs_006504.2_at)			***LjCPR2*** (Ljwgs_068084.1_at)		
	Probeset	**PCC value**	**Initial annotation during chip design**	**Probeset**	**PCC value**	**Initial annotation during chip design**
1	TM1224.12_at	0.75	*L. japonicus* similar to At3g08580: adenylate translocator	Ljwgs_109412.1_at	0.68	*L. japonicus* similar to At3g51480: glutamate receptor like protein → defense against pathogens, reproduction, control of stomata aperture and light signal transduction
2	TM1224.12.1_at	0.75	*L. japonicus* similar to At3g08580: adenylate translocator	Ljwgs_016866.2_at	0.67	*L. japonicus* similar to At5g58870: cell division protein - like
3	chr1.TM0430.17.1_at	0.75	*L. japonicus* similar to At5g48900: pectate lyase	TC10072_at	0.66	homologue to UP|Q863B4 (Q863B4) Trefoil factor 3, partial (15%)
4	chr2.CM0056.38_at	0.72	*L. japonicus* similar to At1g60070: hypothetical protein	chr1.CM0591.55_at	0.66	*L. japonicus* similar to Q40983: (Q40983) METALLOENDOPEPTIDASE: amyloid precursor protein catabolic process

5	chr1.CM0105.95_at	0.72	*L. japonicus* similar to At3g54770: RNA binding protein - like	Ljwgs_051871.1_at	0.66	*L. japonicus* similar to At1g50360: myosin, putative
6	TM0759.8_at	0.71	*L. japonicus* similar to At2g26640: putative beta-ketoacyl-CoA synthase	Ljwgs_023382.1_at	0.65	*L. japonicus* similar to O04434: (O04434) PUTATIVE NADPH-CYTOCHROME P450 REDUCTASE
7	Ljwgs_081701.1_at	0.71	*L. japonicus* similar to At4g12420: pollen-specific protein - like predicted GPI-anchored protein	Ljwgs_043693.1_at	0.65	*L. japonicus* similar to At2g32400: ionotropic glutamate receptor (GLR5)
8	Ljwgs_022220.1_at	0.71	*L. japonicus* similar to At4g00710: unknown protein	Ljwgs_074438.1_at	0.65	*L. japonicus* similar to At1g30360: ERD4 protein (ERD4: Early-responsive to dehydration stress protein)
9	chr5.CM0328.80_at	0.71	*L. japonicus* similar to At5g08680: H+-transporting ATP synthase beta chain (mitochondrial) -like protein	chr1.CM0591.54_at	0.64	*L. japonicus* similar to At5g42390: pitrilysin
10	Ljwgs_089550.1_at	0.71	*L. japonicus* similar to At4g28650: receptor protein kinase-like protein	Ljwgs_058241.1_at	0.63	*L. japonicus* similar to At2g39190: ABC transporter like protein

c) *G. uralensis*

No.	***GuCPR1*** **(**TRINITY_DN18227_c7_g1)	***GuCPR2*** (TRINITY_DN21433_c2_g4)

	Unigene ID	PCC value	Gene description[1]	Unigene ID	PCC value	Gene description[1]
1	TRINITY_DN22119 _c0_g2	0.88	Tudor2, AtTudor2, TSN2 | Arabidopsis thaliana TUDOR-SN protein 2, TUDOR-SN protein 2	TRINITY_DN25698 _c0_g1	0.86	PAP26, ATPAP26 | purple acid phosphatase 26, PURPLE ACID PHOSPHATASE 26
2	TRINITY_DN19174 _c0_g1	0.86	AVA-P3, VHA-C3, ATVHA-C3 | vacuolar-type H(+)-ATPase	TRINITY_DN17844 _c6_g2	0.85	4CL3 | 4-coumarate:CoA ligase 3
3	TRINITY_DN20160 _c1_g3	0.85	SHY3, ATKT2, KT2, KUP2, ATKUP2, TRK2 | potassium transporter 2	TRINITY_DN17669 _c0_g6	0.85	No symbol available | no full name available
4	TRINITY_DN13072 _c0_g1	0.85	No symbol available | no full name available | chr3:6325858-6327666	TRINITY_DN11745 _c0_g1	0.84	SG1 | SLOW GREEN 1
5	TRINITY_DN19072 _c0_g3	0.85	ATPAH2, PAH2 | PHOSPHATIDIC ACID PHOSPHOHYDROLASE 2, phosphatidic acid phosphohydrolase 2	TRINITY_DN22602 _c1_g3	0.84	EFE, ACO4, EAT1 | ethylene forming enzyme, ethylene-forming enzyme
6	TRINITY_DN19747 _c2_g2	0.85	SAPX | stromal ascorbate peroxidase	TRINITY_DN20044 _c5_g1	0.83	JAZ11, TIFY3A | jasmonate-zim-domain protein 11
7	TRINITY_DN23561 _c2_g1	0.84	Symbols: FUT12, FUCT2, ATFUT12, FUCTB | fucosyltransferase 12	TRINITY_DN19740 _c1_g1	0.83	METK1, SAM-1, AtSAM1, SAM1, MAT1 | S-adenosylmethionine synthetase 1, S-

						ADENOSYLMETHIONINE SYNTHETASE-1
8	TRINITY_DN21594_c5_g2	0.83	Symbols: no symbol available | no full name available | chr1:25028538-25029857 REVERSE LENGTH=1320	TRINITY_DN24462_c0_g1	0.81	No symbol available | no full name available |
9	TRINITY_DN15591_c0_g3	0.83	Symbols: ARO4 | armadillo repeat only 4 | chr3:9769666-9772112 FORWARD LENGTH=2447	TRINITY_DN18249_c1_g4	0.81	HSFA1B, HSF3, ATHSF3, ATHSFA1B | ARABIDOPSIS HEAT SHOCK FACTOR 3, CLASS A HEAT SHOCK FACTOR 1B, heat shock factor 3, ARABIDOPSIS THALIANA CLASS A HEAT SHOCK FACTOR 1
10	TRINITY_DN17892_c5_g1	0.82	Symbols: CSE, LysoPL2, AtMAGL3 | Caffeoyl Shikimate Esterase, lysophospholipase 2	TRINITY_DN22602_c1_g2	0.81	EFE, ACO4, EAT1 | ethylene forming enzyme, ethylene-forming

[1]Gene description was obtained from **by blastn query on Araport11 transcripts (DNA) sequences**. Query was performed by the The Arabidopsis Information Resource (TAIR) using DNA contig sequences from the highly correlated *G. uralensis* unigenes. For full BLAST options and parameters, refer to the NCBI BLAST Documentation. BLAST top hit with > 70% identity was chosen.

Table S5. Probeset ID used for co-expression analysis and PCC calculation

Gene	Probeset ID	Transcriptomic Database
MtCPR1	MtrunA17_Chr3g0132841	lipm-browsers.toulouse.inra.fr/pub/expressionAtlas/app/v2
MtCPR2	MtrunA17_Chr4g0072351	
MtbAS	MtrunA17_Chr4g0000191	
MtLUS	MtrunA17_Chr6g0488871	
MtCYP716A12	MtrunA17_Chr8g0388921	
MtCYP72A61	MtrunA17_Chr4g0014051	
MtCYP72A63	MtrunA17_Chr8g0354571	
MtCYP93E2	MtrunA17_Chr7g0235551	
LjCPR1	Ljwgs_006504.2_at	lotus.au.dk (Miyakojima MG20 v 3.0)
LjCPR2	Ljwgs_068084.1_at	
LjbAS	chr3.CM0292.16_s_at	
LjLUS	chr2.CM0373.79_s_at	
LjCYP716A51	Ljwgs_038251.2_at	
LjCYP72A61	chr3.TM0797.5.1_at	
LjCYP93E1	Ljwgs_008809.1_at	
GuCPR1	TRINITY_DN18227_c7_g1	DNA Data Bank of Japan Sequence Read Archive under the accession numbers of DRA012266
GuCPR2	TRINITY_DN21433_c2_g4	
GuLUS	TRINITY_DN23588_c0_g2	
GubAS	TRINITY_DN17145_c0_g2	
GuCYP716A179	TRINITY_DN19674_c0_g1	
GuCYP72A154	TRINITY_DN18189_c2_g1	
GuCYP93E3	TRINITY_DN19088_c1_g2	
GuCYP88D6	TRINITY_DN21774_c5_g1	

Table S6. List of primers used in study of Chapter 3

Gene	Primer Sequence	Additional description
Primers for qPCR analysis of *L. japonicus* hairy root		
LjCPR1_qPCR_Fw	ATGACTTCGAATTCCGATTTGG	
LjCPR1_qPCR_Rv	GTCGTCACGATCAGAATCAGC	
LjCPR2-1_qPCR_Fw	CGAGAAGCTTAGCGACGAGG	
LjCPR2-1_qPCR_Rv	CTCTTCCGCAATCGCCTTG	
LjCPR2-2_qPCR_Fw	CGTCGACGAGGCTGAGGTTGAC	
LjCPR2-2_qPCR_Rv	GCCACAAGCGCCTTGGCGAATC	
LjUBQ_qPCR_Fw	TTCACCTTGTGCTCCGTCTTC	
LjUBQ_qPCR_Rv	AACAACAGCACACACAGACAATCC	
LjCYP716A51_qPCR_Fw	GTCTTCCCCTCATCACTCCA	
LjCYP716A51_qPCR_Rv	TGTCTCTGCGCTATGTCGTC	
LjCYP93E1_qPCR_Fw	AGCACTTTGTCAGCGTTCG	
LjCYP93E1_qPCR_Rv	ACGGCTTCACCTGTTTTTGA	
LjCYP72A61_qPCR_Fw	GTGTGATTGCTACGGTGGTG	
LjCYP72A61_qPCR_Rv	GGAGACCCTGCTGCTTCATA	
LjbAS_qPCR_Fw	TCACTTACGGTTCTTGGTTCG	
LjbAS_qPCR_Rv	CGCCATCACCTCTTTGTGTAG	
LjLUS_qPCR_Fw	TATGAGTGGTCAGGGTGCAA	
LjLUS_qPCR_Rv	GGGCATGTAAACTAAGCGACA	
LjaAS_qPCR_Fw	TGGGCTTTGATGGCTCTAATTC	
LjaAS_qPCR_Rv	TTTGCGGCATGATGAAGTGG	
LjLAS_qPCR_Fw	GGAACTGAACAAGAACGAGCTCAAG	
LjLAS_qPCR_Rv	CCATTTTCCCTCTCAAACTGGAGTC	
LjCAS_qPCR_Fw	CTGAAGAGGCTGTGGTAACAACG	
LjCAS_qPCR_Rv	CATTGGACCTCCATAATCCCCTG	
Primers for LORE1 mutant PCR genotyping		
3941_LjCPR1_Fw	TCCACCATGTATCCAAACACCCCACA	L1-A F
3941_LjCPR1_Rv	GAGGCGAAGACAGCAAATCGACGC	L1-A R
59903_LjCPR1_Fw	TTTTTGGCAATCCCTCGTTCCGGT	L1-B F
59903_LjCPR1_Rv	GCCTGTTACCCAGGGCAAAAACTCCA	L1-B R
37476_LjCPR2-1_Fw	AGGCGATTGCGGAAGAGGCAAAAG	L2-1A F
37476_LjCPR2-1_Rv	TCCACTTCAATGGCGACCTGTGTCA	L2-1A R
65390_LjCPR2-1_Fw	GACGTTGGAAGGGTCGAGTGTGCC	L2-1B F
65390_LjCPR2-1_Rv	CAGCTGCGCTCGTTTTCGATTGGT	L2-1B R
P2 *LORE1*	CCATGGCGGTTCCGTGAATCTTAGG	*LORE1* insertion specific reverse primer
chr0_141891533_F	GCTGCGAAAATGCATGCCAGTCAA	Other insertion in L1-A line
chr0_141891533_R	GGGCACTTCAAAACCTGTAGCTGCCCT	Other insertion in L1-A line
chr0_144522704_F	GGAAGATTTTCCAGCGAGGGACGA	Other insertion in L1-B line
chr0_144522704_R	GGTGGACCAGGTGTTCTTCCACGA	Other insertion in L1-B line
chr5_23067809_F	CAATAGACAGCCACACGGTGACCCC	Other insertion in L1-B line

chr5_23067809_R	TCCACCTAATCAAGACTGGTACTGAGGCA	Other insertion in L1-B line
Primers for gRNA construct targeting *LjCPR1* gene		
F2_tgRNA_2A_LjCPR1	ttgggtctcgTGCAGCTGGCTTCCAATCACCCATTG TTTTAGAGCTAGAAATAGCA	
R2_tgRNA_2B_LjCPR1	ttgggtctccAAACTCGTTTACAGCCTCGTTATTCT GCACCAGCCGGGAATCGAA	
F2_tgRNA_LjCPR1-4A	ttgggtctcgTGCAGAGGATGGCACTCCCCTAGGG TTTTAGAGCTAGAAATAGCA	
R2_tgRNA_LjCPR1-4B	ttgggtctccAAACTGCACACTGCGCACTGCATTCT GCACCAGCCGGGAATCGAA	
F2_tgRNA_LjCPR1-5A	ttgggtctcgTGCAGTCACTTCAAGTAGACTTCTCG TTTTAGAGCTAGAAATAGCA	
R2_tgRNA_LjCPR1-5B	ttgggtctccAAACCGTTATTATTCTATTTCATCTG CACCAGCCGGGAATCGAA	
Primers for checking CRISPR-Cas9 mutation in *LjCPR1*		
LjCPR1_F_982	GGAGACCATGTGGGTGTTTATGCTG	To check target gRNA LjCPR1-4
LjCPR1_R_1222	GAGCAGCTAATGCAACTAGAGCAGC	To check target gRNA LjCPR1-4
LjCPR1_F_1208	GCTGCTCTAGTTGCATTAGCTGC	To check target gRNA LjCPR1-2
LjCPR1_R_1485	CGTTGGACCACAAACCAAGGCACAAG	To check target gRNA LjCPR1-2 or LjCPR1-5
LjCPR1_F_1198	GCTGCTCTAGTTGCATTAGCTGCTC	To check target gRNA LjCPR1-5

Table S7. All *LORE1* insertions in the genome of the selected *Ljcpr1* and *Ljcpr2-1* mutant lines (lotus.au.dk)

Mutant name	Plant line ID	Genomic position	Gene ID	Insertion type	Gene annotation
L1-A	30003941	chr1_13956169_R	Lj1g3v1113880	Intronic	PREDICTED: probable inactive receptor kinase At5g58300-like isoform X1
		chr1_18540324_F	**Lj1g3v1548790**	**Exonic**	**PREDICTED: NADPH--cytochrome P450 reductase-like isoform X2 gi|502133111|ref|XP_004501654.1|**
		chr0_141891533_F	Lj0g3v0273739	Exonic	PREDICTED: ATP-dependent zinc metalloprotease FTSH 12, chloroplastic-like
L1-B	30059903	**chr1_18538152_R**	**Lj1g3v1548790**	**Exonic**	**PREDICTED: NADPH--cytochrome P450 reductase-like isoform X2 gi|502133111|ref|XP_004501654.1|**
		chr5_23067809_R	Lj5g3v1598550	Exonic	PREDICTED: PH-interacting protein-like gi|502117700|ref|XP_004495905.1|
		chr0_144522704_F	Lj0g3v0278319	Exonic	Lactoylglutathione lyase / glyoxalase I family protein gi|508699005|gb|EOX90901.1|
L2-1A	30037476	**chr4_28883580_R**	**Lj4g3v2107220**	**Exonic**	**cytochrome P450 reductase [*L. japonicus*] gi|197209812|dbj|BAG68945.1|**
		chr2_35390899_F	Lj2g3v2574420	Exonic	Receptor-type tyrosine-protein phosphatase U [Theobroma cacao] gi|508728148|gb|EOY20045.1|
		chr0_125503891_F	-	Intergenic	-
L2-1B	30065390	chr2_11997511_R	Lj2g3v0766630	Intronic	Serine incorporator [*M. truncatula*] gi|357500415|ref|XP_003620496.1|

		chr0_20652508_R	Lj0g3v0061569	Intronic	PREDICTED: ras-related protein RABH1b-like [Cicer arietinum] gi|502152574|ref|XP_004508989.1|
		chr4_28884617_F	**Lj4g3v2107220**	**Exonic**	**cytochrome P450 reductase [*L. japonicus*] gi|197209812|dbj|BAG68945.1|**
		chr0_166260633_F	Lj0g3v0318339	Exonic	-

Table S8. List of m/z values for the target ion and qualifier ion used in GC-MS analysis of Chapter 3

Compounds	Target ion	Qualifier ion
α-amyrin **(2)**	218	203
Ursolic acid **(11)**	203	320
Lupeol **(3)**	189	203
Betulinic acid **(9)**	189	203
β-amyrin **(1)**	218	203
Oleanolic acid **(8)**	203	320
24-OH β-amyrin **(4)**	218	203
Sophoradiol **(10)**	306	291
Soyasapogenol B **(13)**	306	291
Soyasapogenol A **(15)**	394	278
Soyasapogenol E **(12)**	232	278
Asiatic Acid (IS) **(14)**	320	203
Campesterol **(16)**	382	400
β-sitosterol **(17)**	397	400
Stigmasterol **(18)**	395	400

Table S9. List of primers used in study of Chapter 4

Gene	Primer Sequence	Target sequence
*MtCPR1*_for	CACCATGACTTCTTCCAATTCCGATTTAGTCCG	Amplification of *MtCPR1* for TOPO cloning
*MtCPR1*_rev	TCACCAGACATCCCTAAGGTAGCGTCCATCC	Amplification of *MtCPR1* for TOPO cloning
*MtCPR2*_for	CACCCCATGCAAGATTCAAGCTCAATG	Amplification of *MtCPR2* for TOPO cloning
*MtCPR2*_rev	GCCCGGTTCATCATTACCATACATCACG	Amplification of *MtCPR2* for TOPO cloning
*LjCPR1*_FOR	CACCATGACTTCGAATTCCGATTTGGTTCG	Amplification of *LjCPR1* for TOPO cloning
*LjCPR1*_REV	TCACCAGACATCCCTGAGGTAACGTCC	Amplification of *LjCPR1* for TOPO cloning
*LjCPR2*_for	CACCATGGAAGAATCAAGCTCCATGAAG	Amplification of *LjCPR2* for TOPO cloning
*LjCPR2*_rev	TCACCATACATCACGCAAATACCTACC	Amplification of *LjCPR2* for TOPO cloning
*GuCPR1*_FOR	CACCATGACTTCGAATTCCGATTTGGTTCG	Amplification of *GuCPR1* for TOPO cloning
*GuCPR1*_REV	TCACCAGACATCCCTGAGGTAACGTCC	Amplification of *GuCPR1* for TOPO cloning
*GuCPR2*_FOR	CACCATGCAGGATTCAAACTCCATGAAG	Amplification of *GuCPR2* for TOPO cloning
*GuCPR2*_REV	TCACCATACATCACGCAAATACCTGCCA	Amplification of *GuCPR2* for TOPO cloning
*GuCPR1*_Inf_Fw	GCCGCCCCCTTCACCATGACTTCGAATTCCGATT	Amplification of *GuCPR1* for HiFi-DNA Infusion TOPO cloning
*GuCPR1*_Inf_Rv	GGCGCGCCCACCCTTTCACCAGACATCCCTG	Amplification of *GuCPR1* for HiFi-DNA Infusion TOPO cloning
*GuCPR2*_Inf_Fw	GCCGCCCCCTTCACCATGCAGGATTCAAACTC	Amplification of *GuCPR2* for HiFi-DNA Infusion TOPO cloning
*GuCPR2*_Inf_Rv	GGCGCGCCCACCCTTTCACCATACATCACGCA	Amplification of *GuCPR2* for HiFi-DNA Infusion TOPO cloning
*MtCPR1*_for1	ACTGACAATGCCGCAAGATT	Sequencing primer for *MtCPR1*
*MtCPR1*_for2	AGGGACCGGCGTAACATAC	Sequencing primer for *MtCPR1*
*MtCPR1*_for3	CACGTAACTTGTGCCCTGGT	Sequencing primer for *MtCPR1*
*MtCPR1*_rev1	ACCTAGGCCAAAAACACCAT	Sequencing primer for *MtCPR1*
*MtCPR1*_rev2	CCAACAACTTCCCAGCTTCT	Sequencing primer for *MtCPR1*
*MtCPR1*_rev3	GGGAATAGCATTCTTCATCCA	Sequencing primer for *MtCPR1*
*MtCPR2*_for1	TCTTAGCTACATATGGTGATGGTGA	Sequencing primer for *MtCPR2*
*MtCPR2*_for2	TCAGATCGTTCTTGCACTCA	Sequencing primer for *MtCPR2*
*MtCPR2*_for3	CATCATCTCCAAGAGTGGCA	Sequencing primer for *MtCPR2*
*MtCPR2*_rev1	CGAATCTTCTTCCCCTTCAA	Sequencing primer for *MtCPR2*
*MtCPR2*_rev2	CGGATAAATTCTCACAGTAAACACC	Sequencing primer for *MtCPR2*

*MtCPR2*_rev3	CACACTCCTTGATGAATCCG	Sequencing primer for *MtCPR2*
*MtCPR2*_rev4	TTGGAATTGTCCAAAGAGCC	Sequencing primer for *MtCPR2*
*LjCPR1*_FOR1	CCAACTGACAATGCTGCAAG	Sequencing primer for *LjCPR1*
*LjCPR1*_FOR2	TGATATATCGGGGACTGGCA	Sequencing primer for *LjCPR1*
*LjCPR1*_REV1	CCAACAACTTCCCAGCTTCT	Sequencing primer for *LjCPR1*
*LjCPR1*_REV3	GGGAATAGCATTCTTCATCCA	Sequencing primer for *LjCPR1*
LjCPR_for1	CACTGGCACTTTTCTTCTTAGC	Sequencing primer for *LjCPR2*
LjCPR_for2	TTCATACTCCTGTGTCAGATCGTT	Sequencing primer for *LjCPR2*
LjCPR_for3	GATTTTATTCGATCTCATCATCTCC	Sequencing primer for *LjCPR2*
LjCPR_rev1	CTCCCTCCAGAAACCATTTG	Sequencing primer for *LjCPR2*
LjCPR_rev2	TAAACACCAACATGGTCCCC	Sequencing primer for *LjCPR2*
LjCPR_rev3	ATGAATCCTACCAGTGGGCA	Sequencing primer for *LjCPR2*
LjCPR_rev4	CTGCTCTTGCAAAATTGTGTG	Sequencing primer for *LjCPR2*
*GuCPR1*_FOR1	CCAACTGACAATGCTGCAAG	Sequencing primer for *GuCPR1*
*GuCPR1*_FOR2	TGATATATCGGGGACTGGCA	Sequencing primer for *GuCPR1*
*GuCPR1*_REV2	CCAACAACTTCCCAGCTTCT	Sequencing primer for *GuCPR1*
*GuCPR1*_REV3	GGGAATAGCATTCTTCATCCA	Sequencing primer for *GuCPR1*
*GuCPR2*_FOR1	GGAGACACTCGCACTTTTCTTT	Sequencing primer for *GuCPR2*
*GuCPR2*_FOR2	TCTGTGTCGGATCGTTCTTG	Sequencing primer for *GuCPR2*
*GuCPR2*_REV2	GAAAACACCAACATGGTCCC	Sequencing primer for *GuCPR2*
*GuCPR2*_REV3	ATGAATCCTACCAGTGGGCA	Sequencing primer for *GuCPR2*
M1N_M2C_FOR	GGACTTCTCGTTTTTCTATGGCGTAGATCCAATTCTCAAAAACC	Amplification of fragments for N-terminal switching of MtCPRs
M1N_M2C_ backbone_REV	ATTGGATCTACGCCATAGAAAAACGAGAAGTCCAATTATGACGG	Amplification of fragments for N-terminal switching of MtCPRs
M2N_M1C_FOR	CGTCGTCGTTTTAATTTGGAAGAAATCTTCGGATCGGAGC	Amplification of fragments for N-terminal switching of MtCPRs
M2N_M1C_ backbone_REV	CCGAAGATTTCTTCCAAATTAAAACGACGACGCAACCG	Amplification of fragments for N-terminal switching of MtCPRs

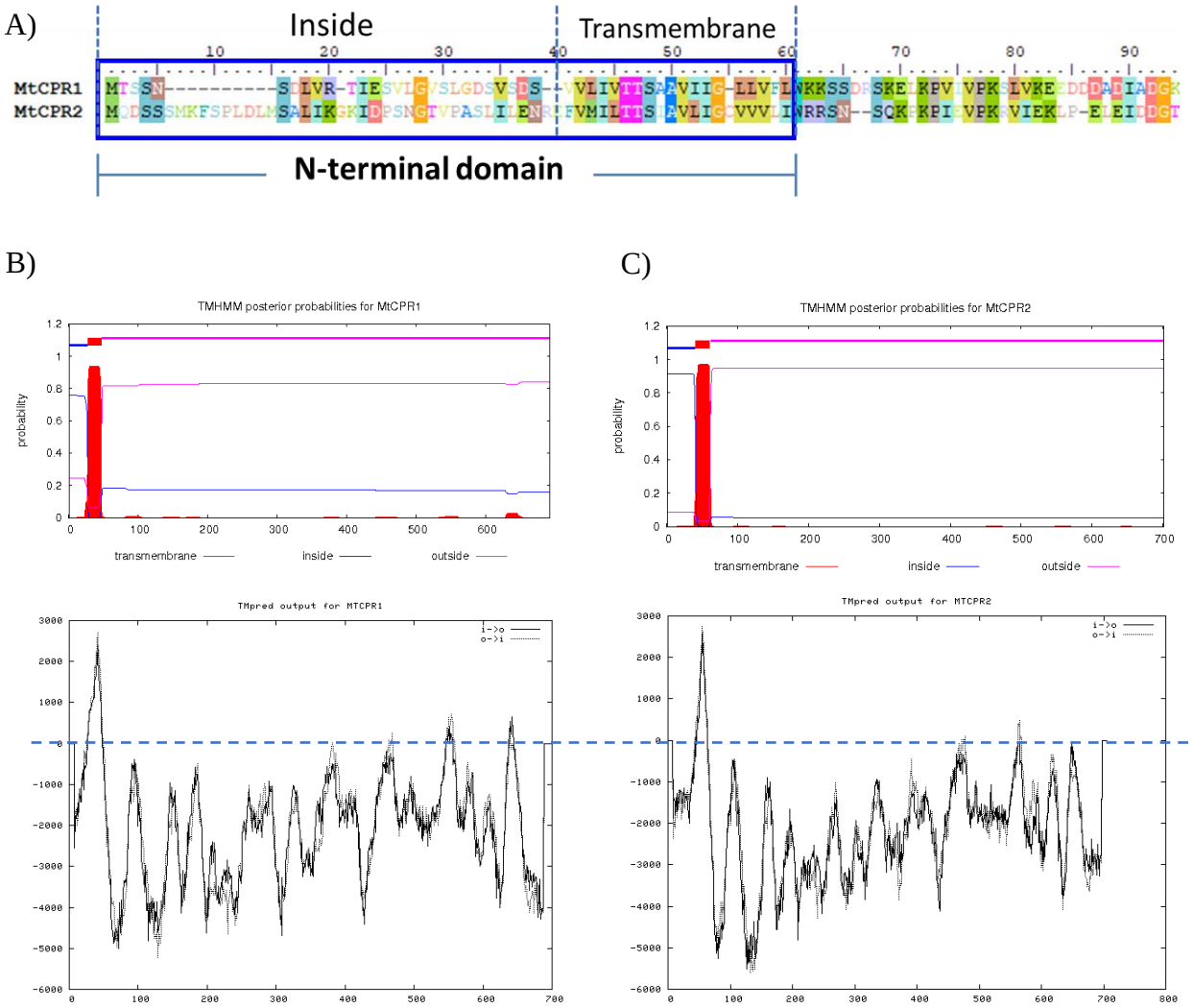

Figure S1. In silico transmembrane helix prediction of MtCPR1 and MtCPR2 (http://www.cbs.dtu.dk/services/TMHMM/). (A) The sequence of MtCPR1 and MtCPR2 position of protein helix inside, transmembrane, and outside of the endoplasmic reticulum (ER) respectively. (B) Amino acid of MtCPR1 number 1-26 is located inside the ER, 27-46 is the transmembrane helix, and 47-692 is located outside the ER. (C) Amino acid of MtCPR2 number 1-40 is located inside the ER, 41-60 is the transmembrane helix, and 61-701 is located outside the ER.

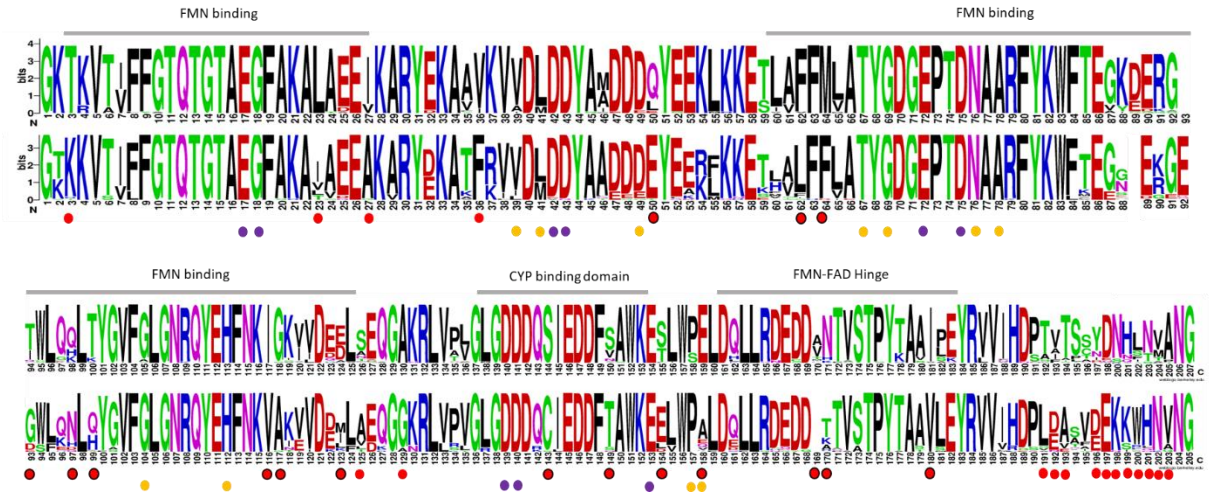

Figure S2. Motif analysis of 37 sequences of each CPR class I and II from 24 legume species. Red circles indicate the different residue that are conserved in each CPR class I and II. Purple circles indicate acidic residue formerly reported to be important in CYP:CPR interaction in human CPR (hCPR). Yellow circles indicate point mutations in hCPR that are reported to improve interaction with a specific CYP. Motif logo was created by WebLogo online software (https://weblogo.berkeley.edu/logo.cgi).

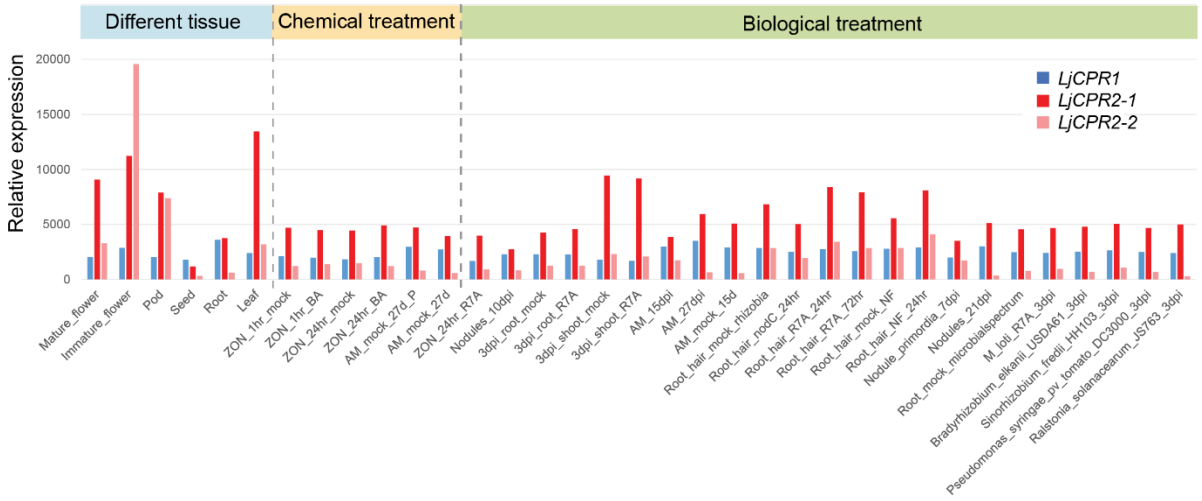

Figure S3. Gene co-expression analysis of *L. japonicus* using transcriptomic data of 35 different samples from Gifu v1.2 genome version.

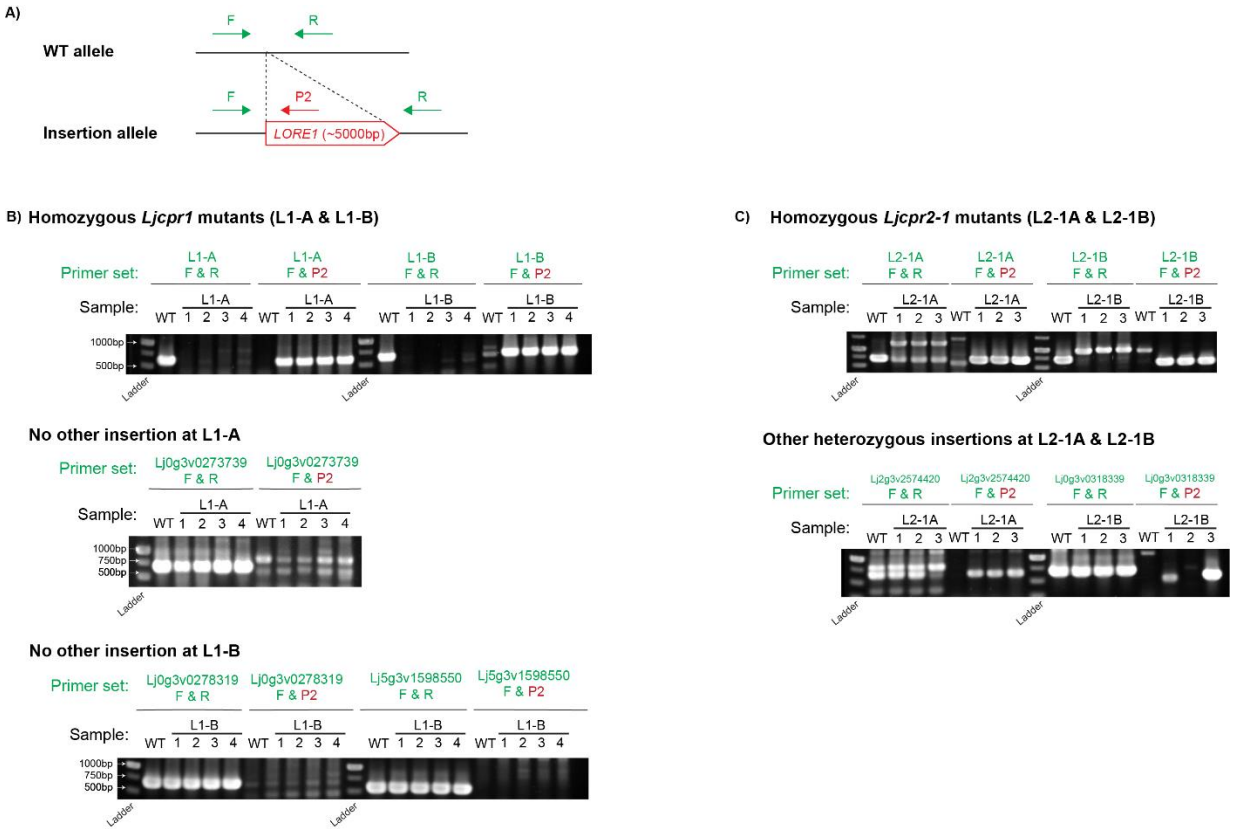

Figure S4. PCR genotyping of *LORE1* insertion *Ljcpr* mutant. A) Primer pairs for *LORE1* PCR genotyping. B) Gel electrophoresis of soil-cultured *Ljcpr1 LORE1* insertion mutants confirmed single insertion homozygous L1-A and L1-B *LORE1* insertion mutation and no other *LORE1* exonic insertion present in other expected genes. C) Gel electrophoresis of soil-cultured *Ljcpr2-1 LORE1* insertion mutants confirmed non-single insertion homozygous L2-1A and L2-1B *LORE1* insertion mutation with heterozygous *LORE1* exonic insertion presents in other expected genes. Hydroponic-cultured mutant seeds were collected from these homozygous mutant plants and also confirmed by PCR as described above. Minimum three homozygous mutant plants per lines were obtained.

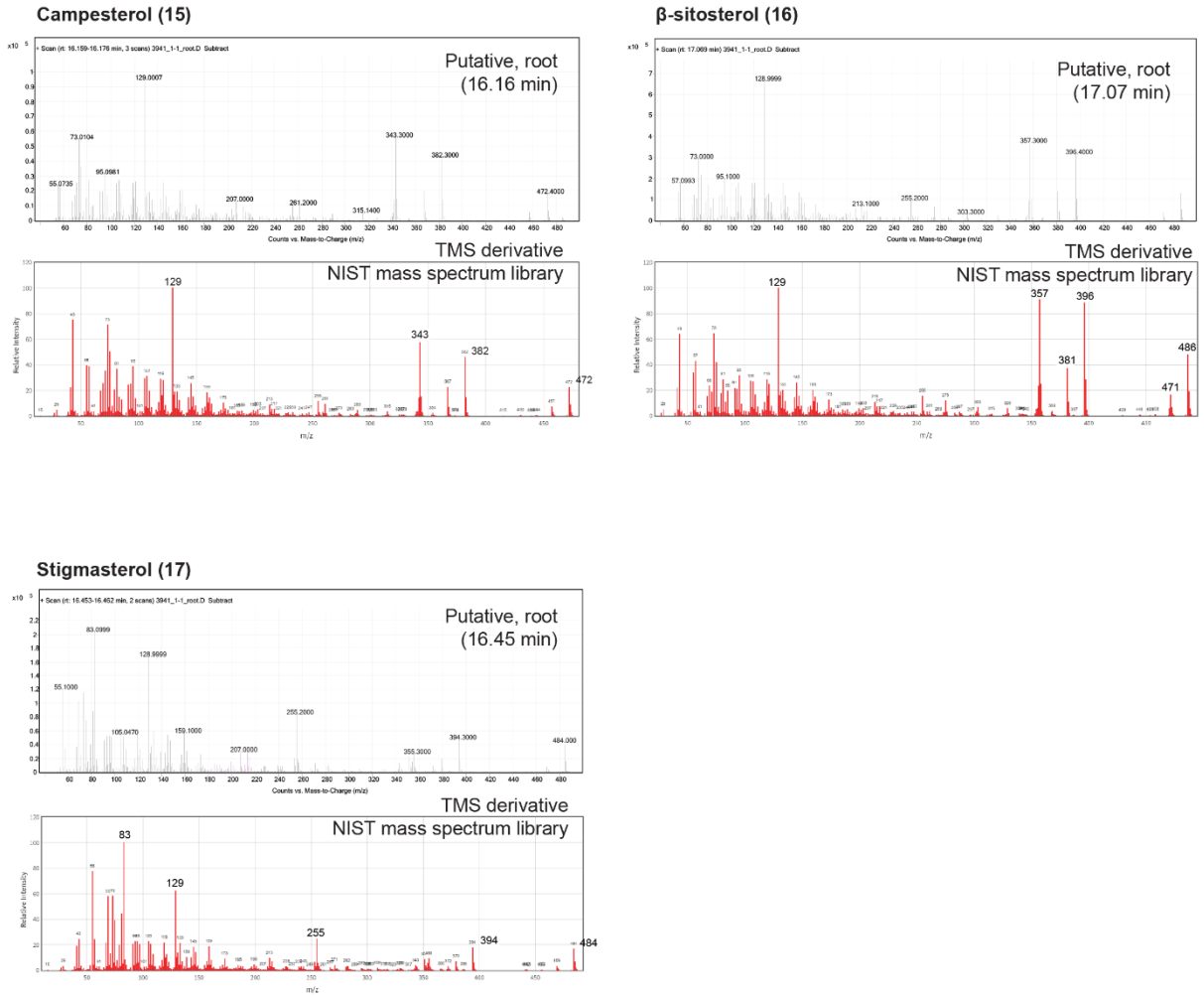

Figure S5. Mass spectra of phytosterols compared to NIST library and 24-OH β-amyrin compared to standard compound mass spectrum obtained from previous experiment.

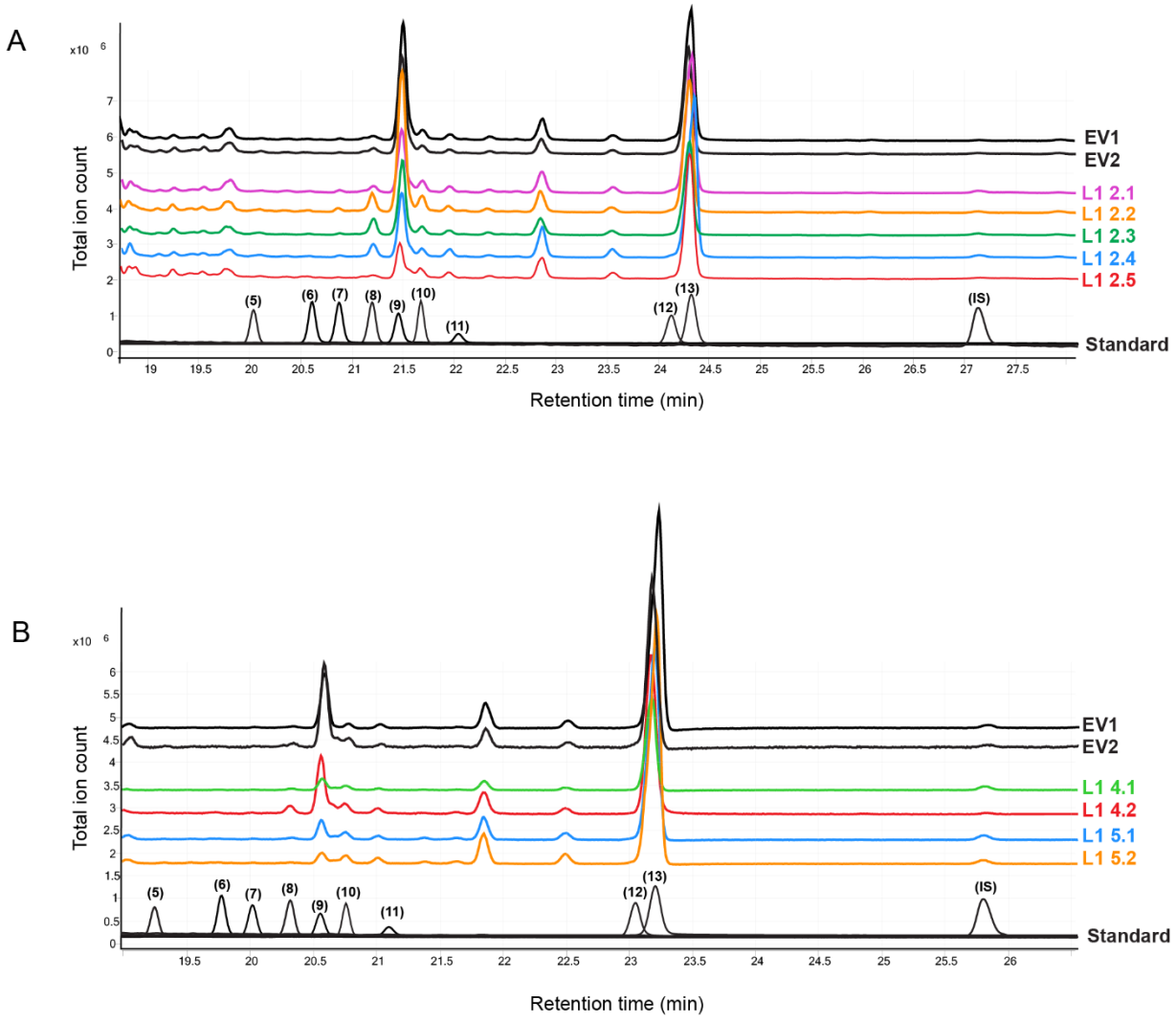

Figure S6. TIC scan of GC-MS chromatogram of *Ljcpr1* knockout hairy root mutants A) target gRNA No. 2, and B) target gRNA No. 4 and 5. Compound numbers are consistent with those in Figure 3-3. The peaks of betulinic acid decreased significantly in Ljcpr1 frameshift mutants (L1-2.1, 2.3, 2.4, 4.1, 5.1, 5.2).

A

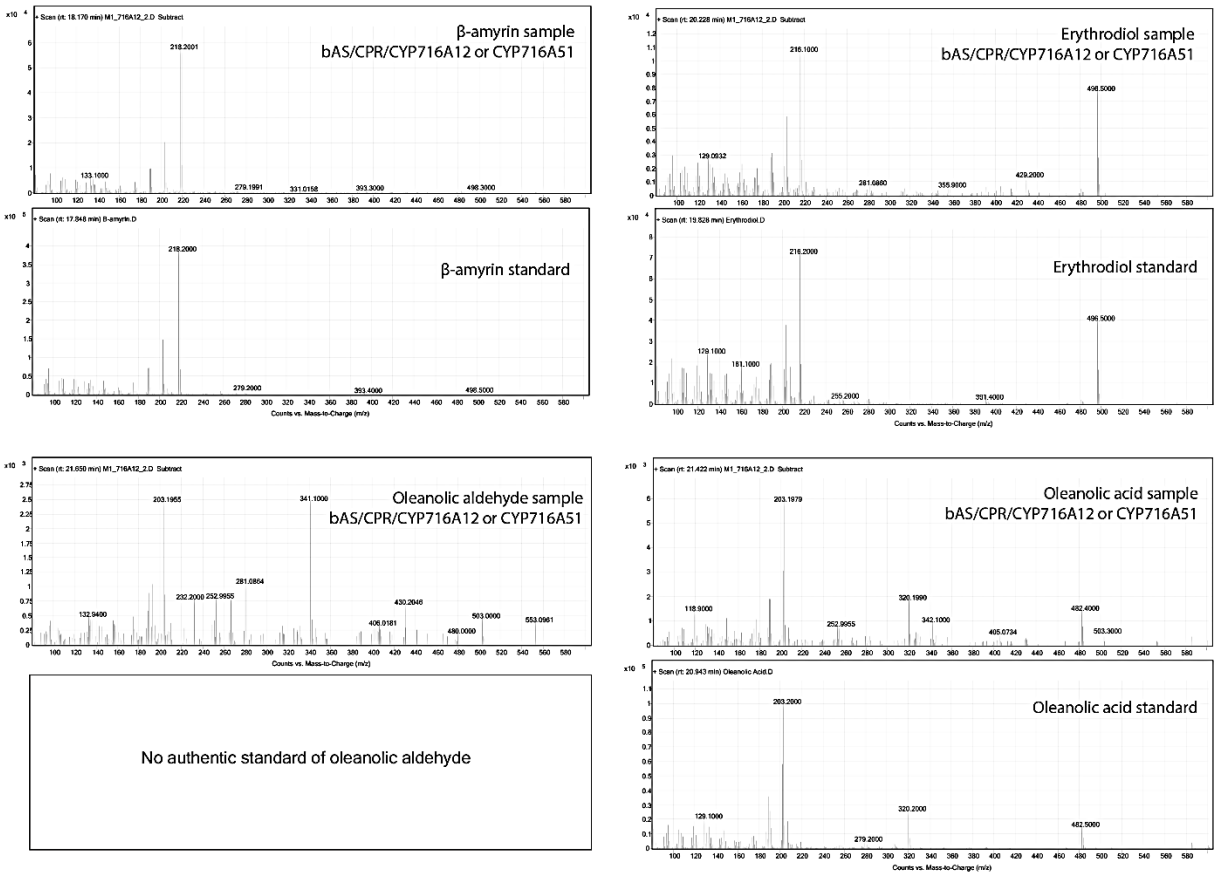

Figure S7. Mass spectra of target compounds and authentic standards of β-amyrin derivatives at (A) C-28, (B) C-30, (C) C-24, (D) C-22, and (E) α-amyrin and (F) lupeol derivatives at C-28.

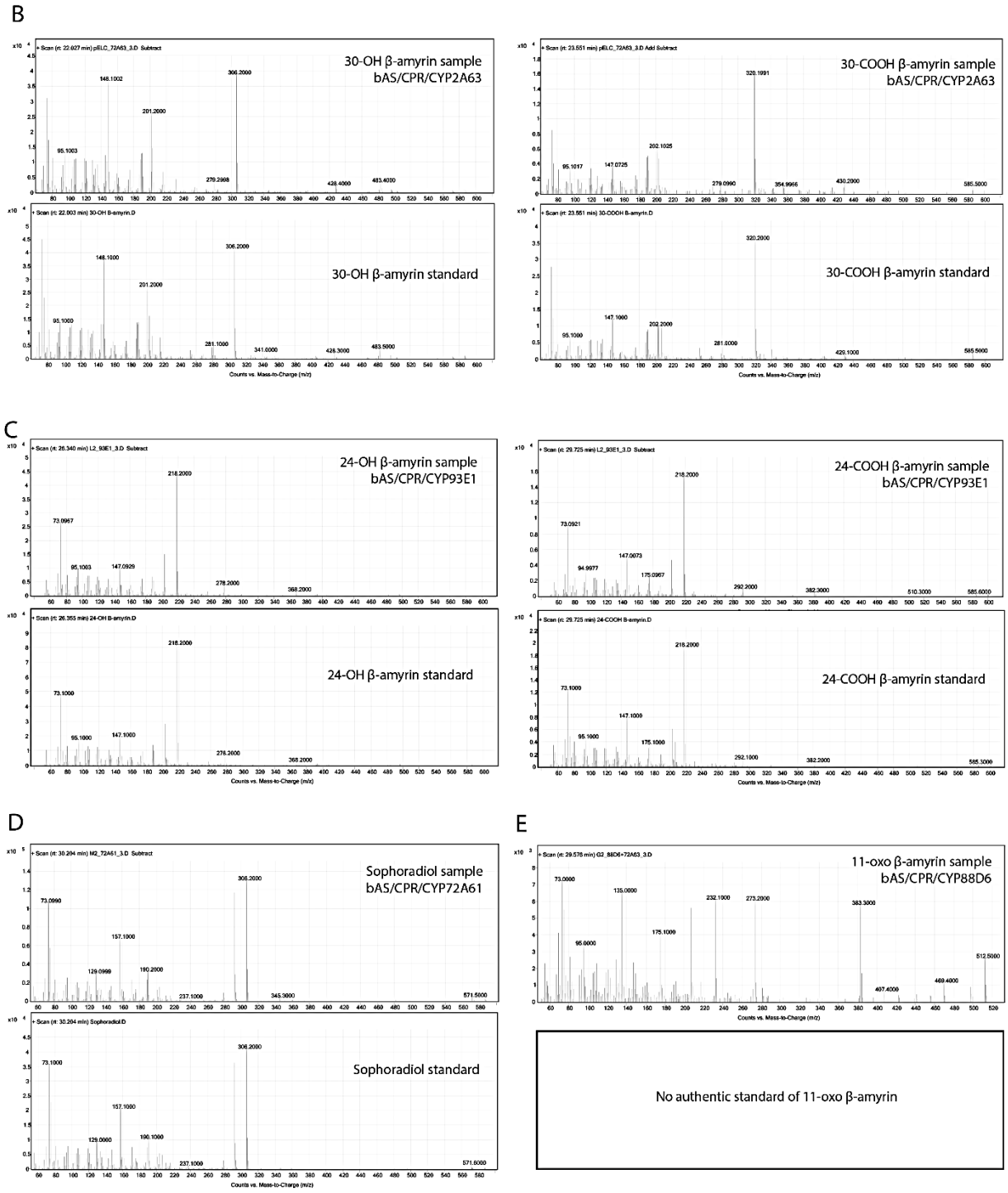

Figure S7. Cont.

F

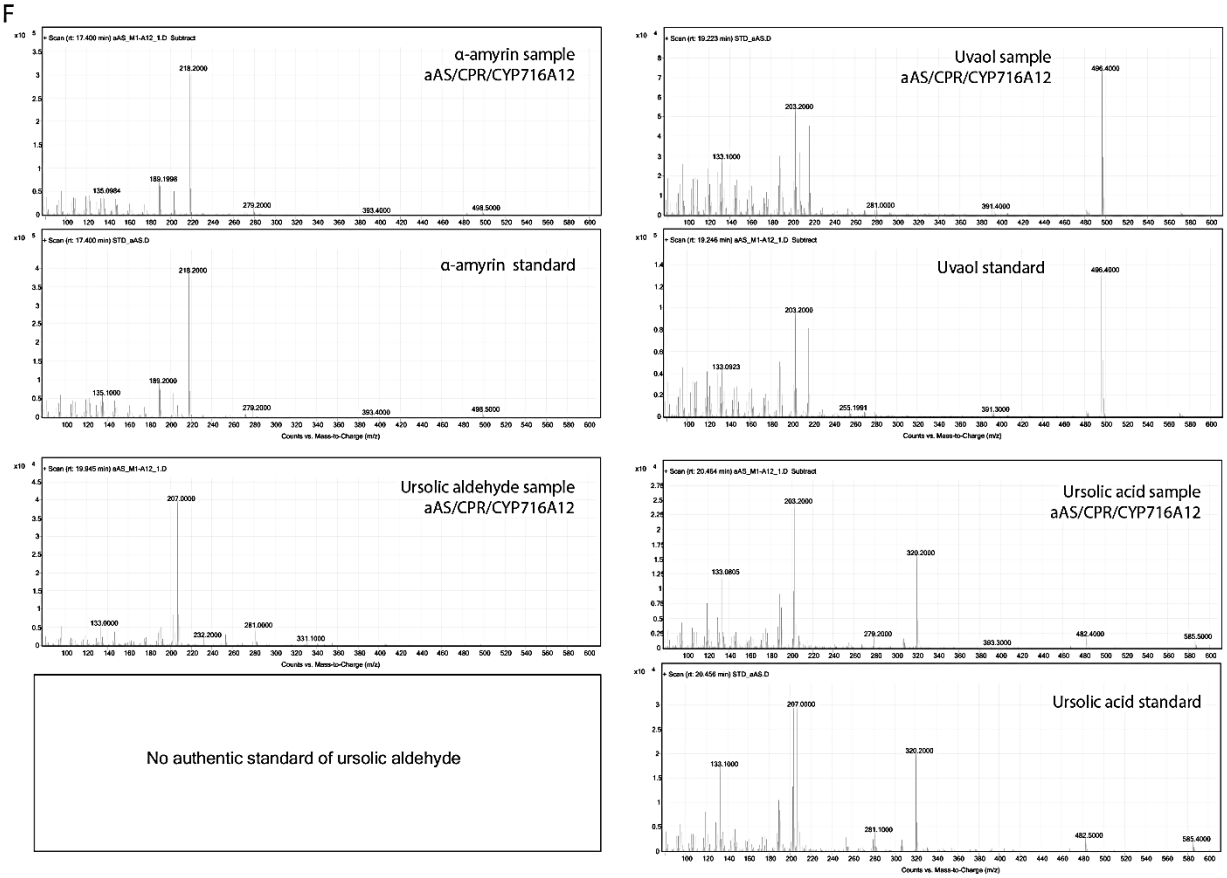

Figure S7. Cont.

G

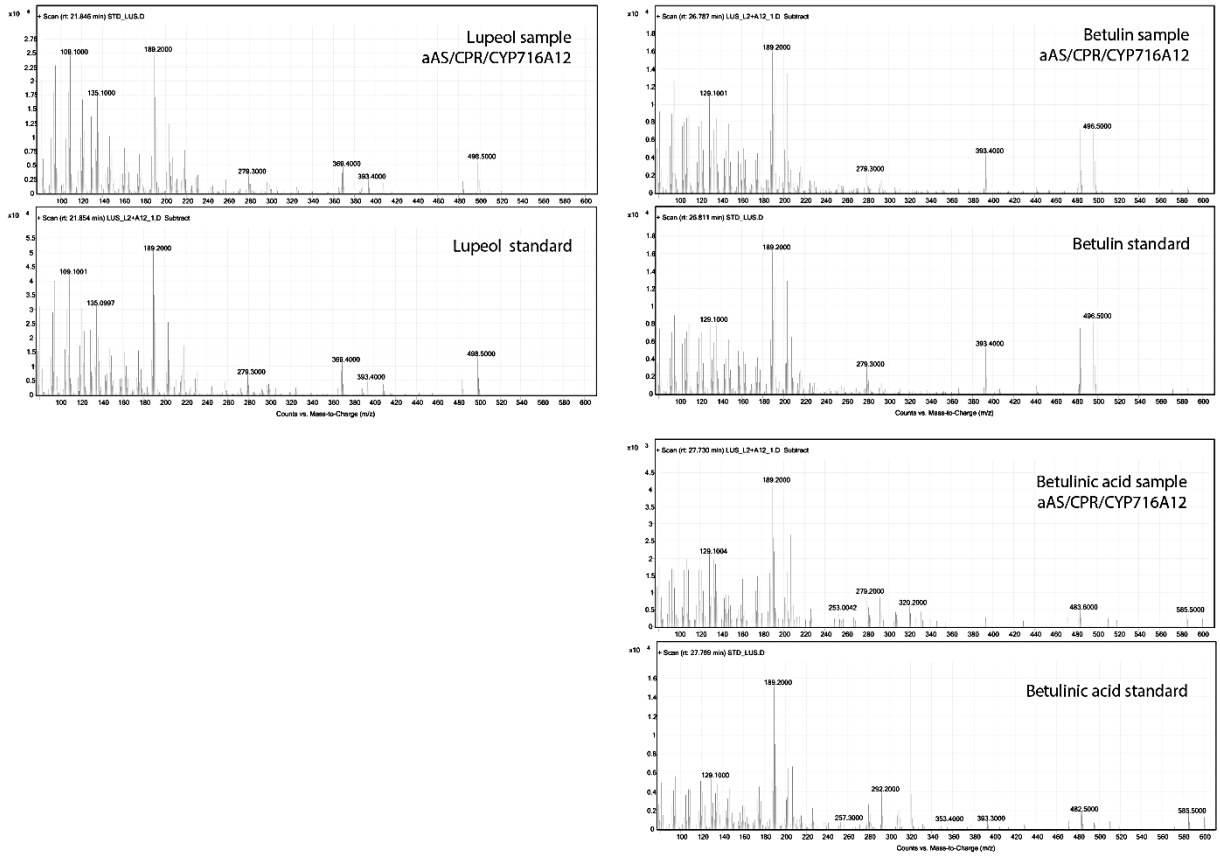

Figure S7. Cont.

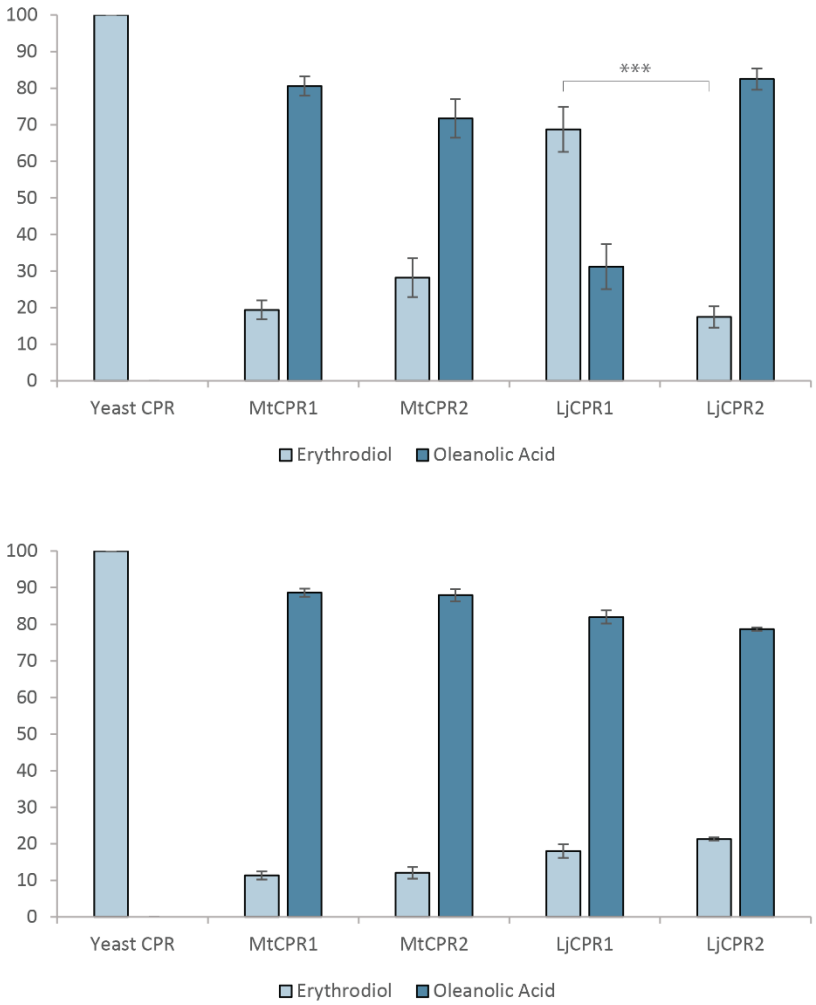

Figure S8. The relative amount of triterpenoid produced by co-expression of A) MtCYP716A12 and B) LjCYP716A51 and different CPRs from *M. truncatula* and *L. japonicus* in yeast feeding assay, by supplementing yeast INV*Sc1* strain with 10 μM of erythrodiol as substrate. Triterpenoids content were measured relative to uvaol as internal standard. Data have been presented as mean ± SE (n=3). nd, signal below detection limit. Single-factor ANOVA was used for statistical comparisons. Values were considered statistically significant at *P<0.05, **P<0.01, and ***P<0.001.

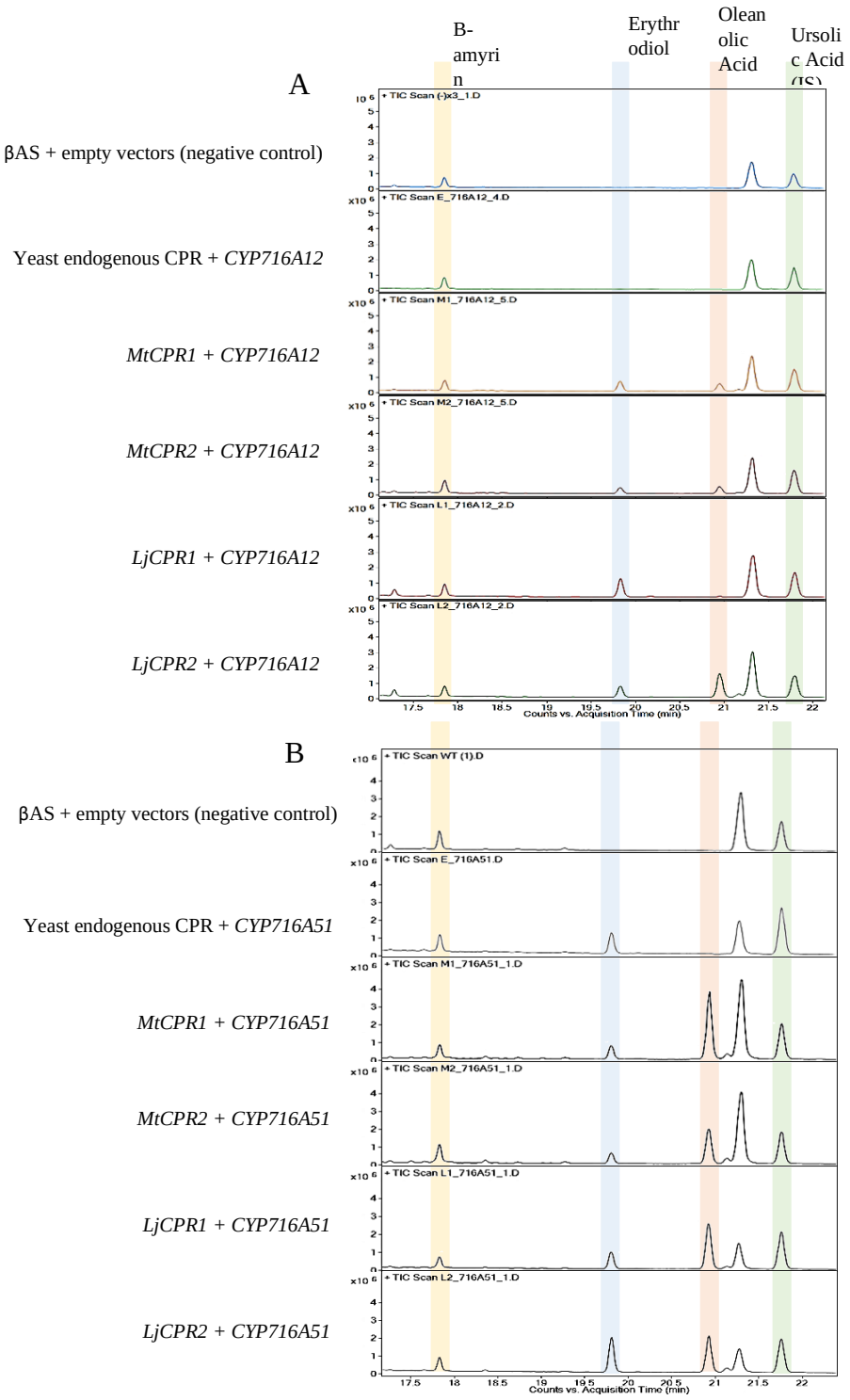

Figure S9. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing INV*Sc*1 yeast harboring MtCPRs and LjCPRs paired with A) CYP716A12 and B) CYP716A51. All samples were cultured in different time, and ran in GCMS at the same time using HP 5-MS column with common method with ursolic acid as internal standard.

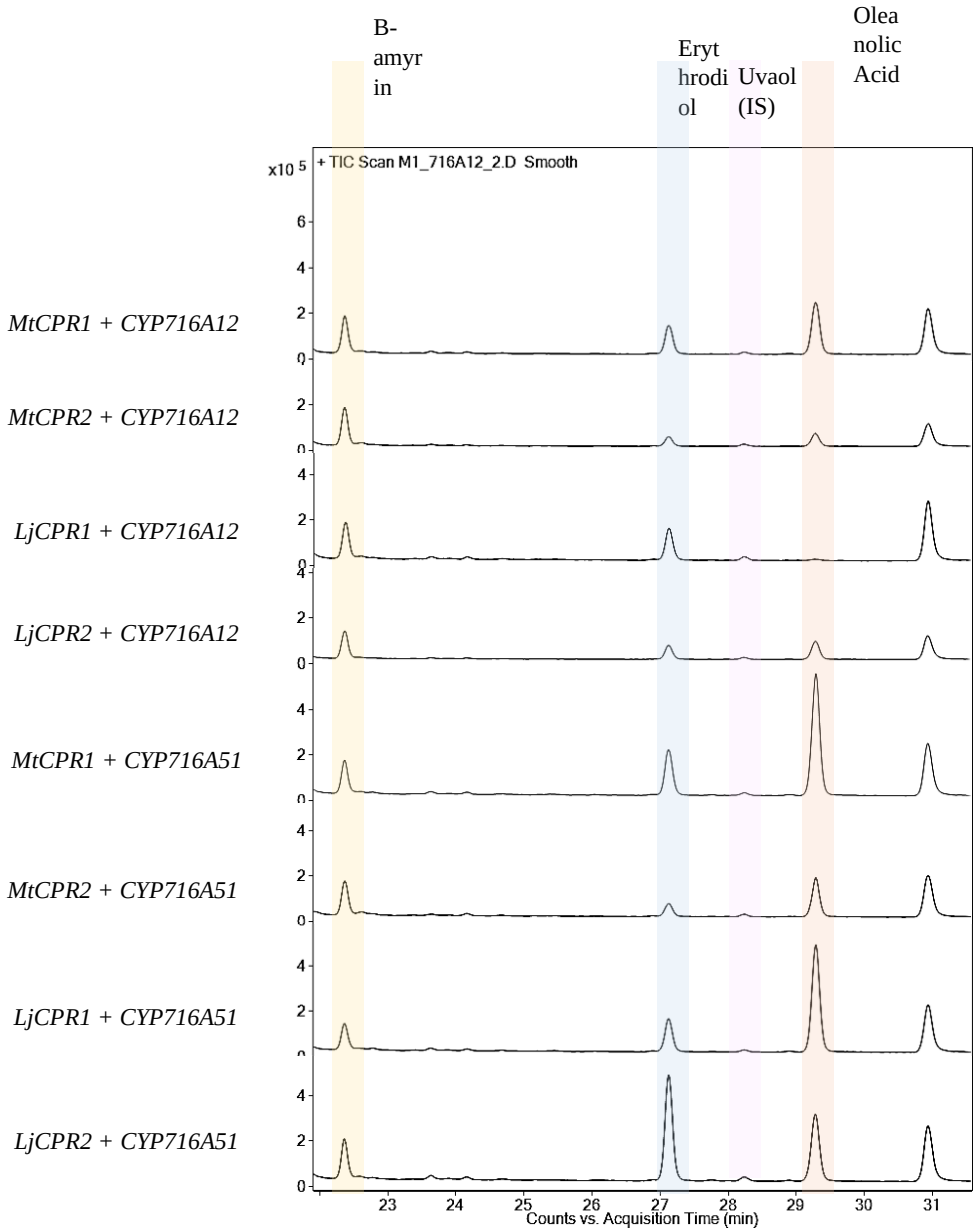

Figure S10. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing INV*Sc*1 yeast harboring MtCPRs and LjCPRs paired with CYP716A12 and CYP716A51. All samples were cultured in and ran in GCMS at the same time using HP-5 MS columin with optimized method with uvaol as internal standard.

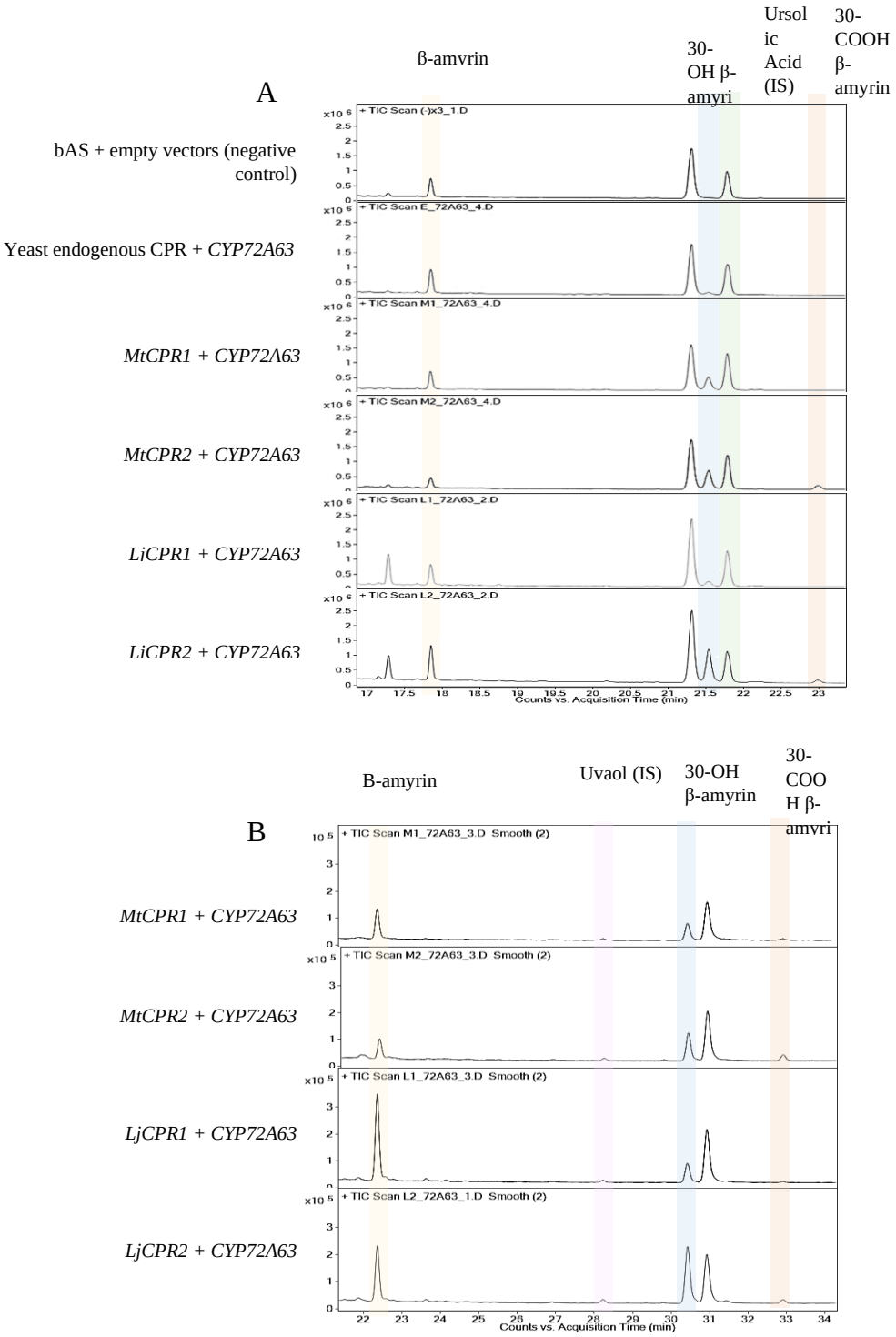

Figure S11. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing INV*Sc*1 yeast harboring MtCPRs and LjCPRs paired with CYP72A63 using HP 5-MS column with (A) common method with ursolic acid as internal standard and (B) optimized method with uvaol as internal standard.

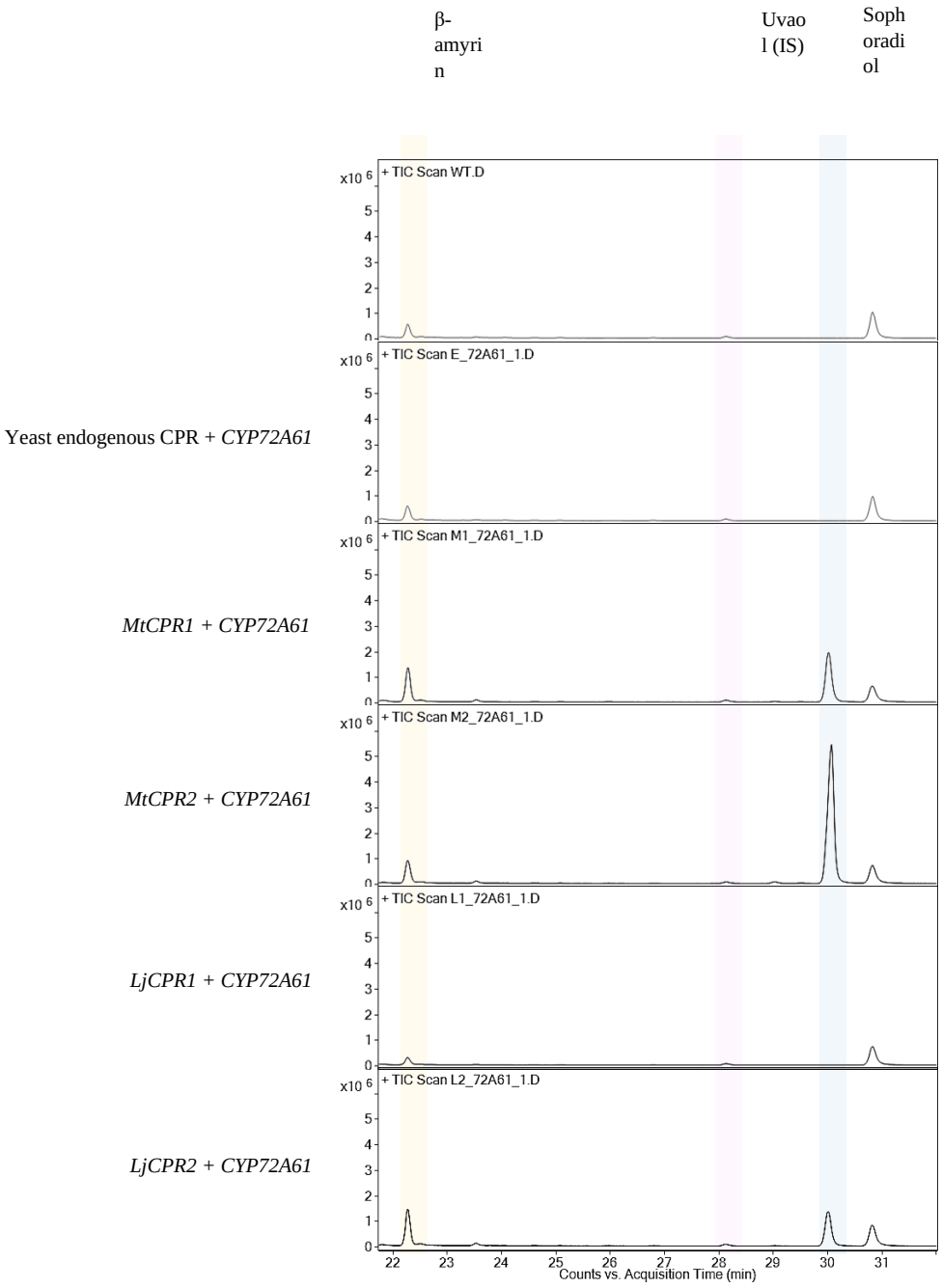

Figure S12. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing INV*Sc*1 yeast harboring MtCPRs and LjCPRs paired with CYP72A61 using HP 5-MS column with optimized method with uvaol as internal standard.

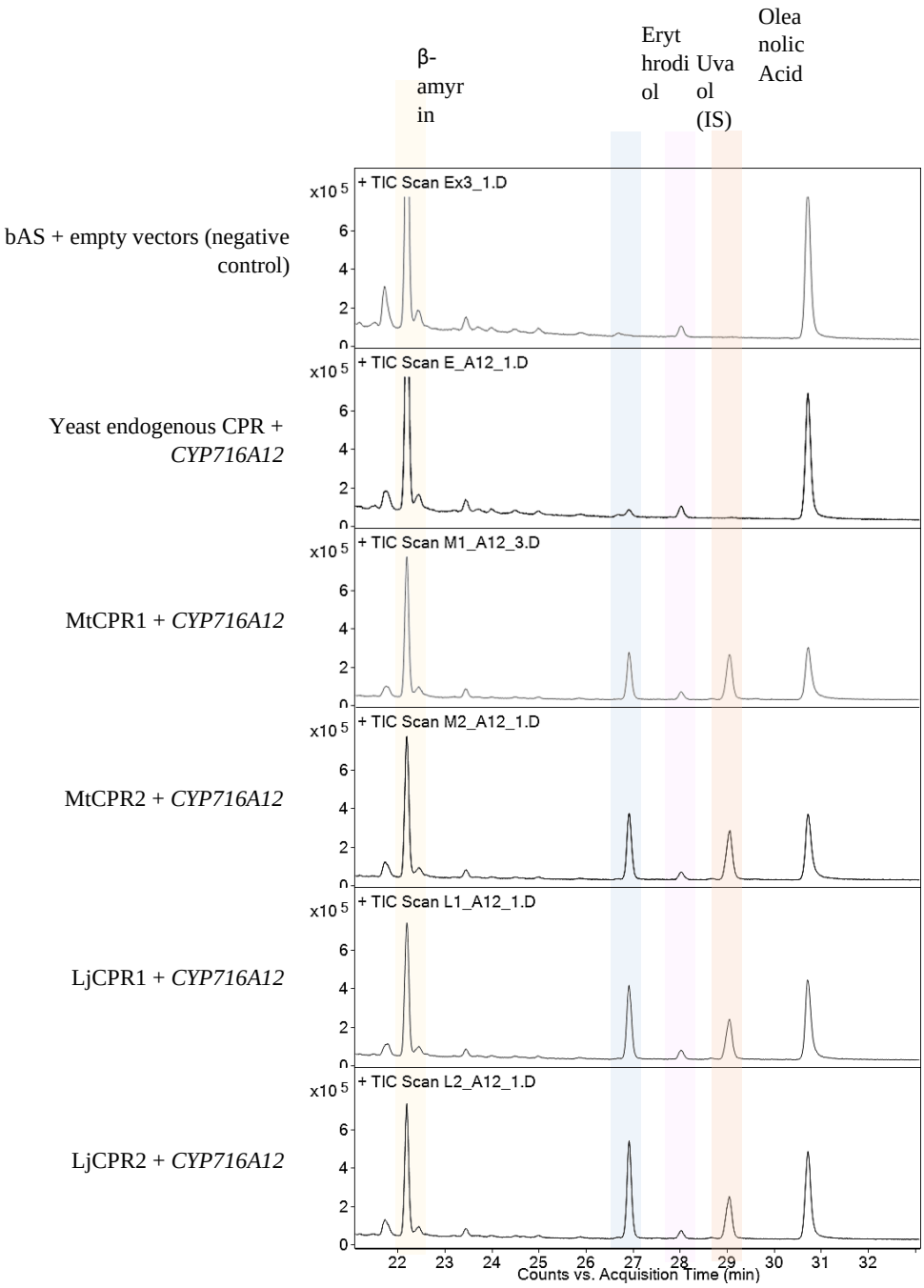

Figure S13. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing PSIII yeast harboring MtCPRs and LjCPRs paired with CYP716A12 using HP 5-MS column with optimized method with uvaol as internal standard.

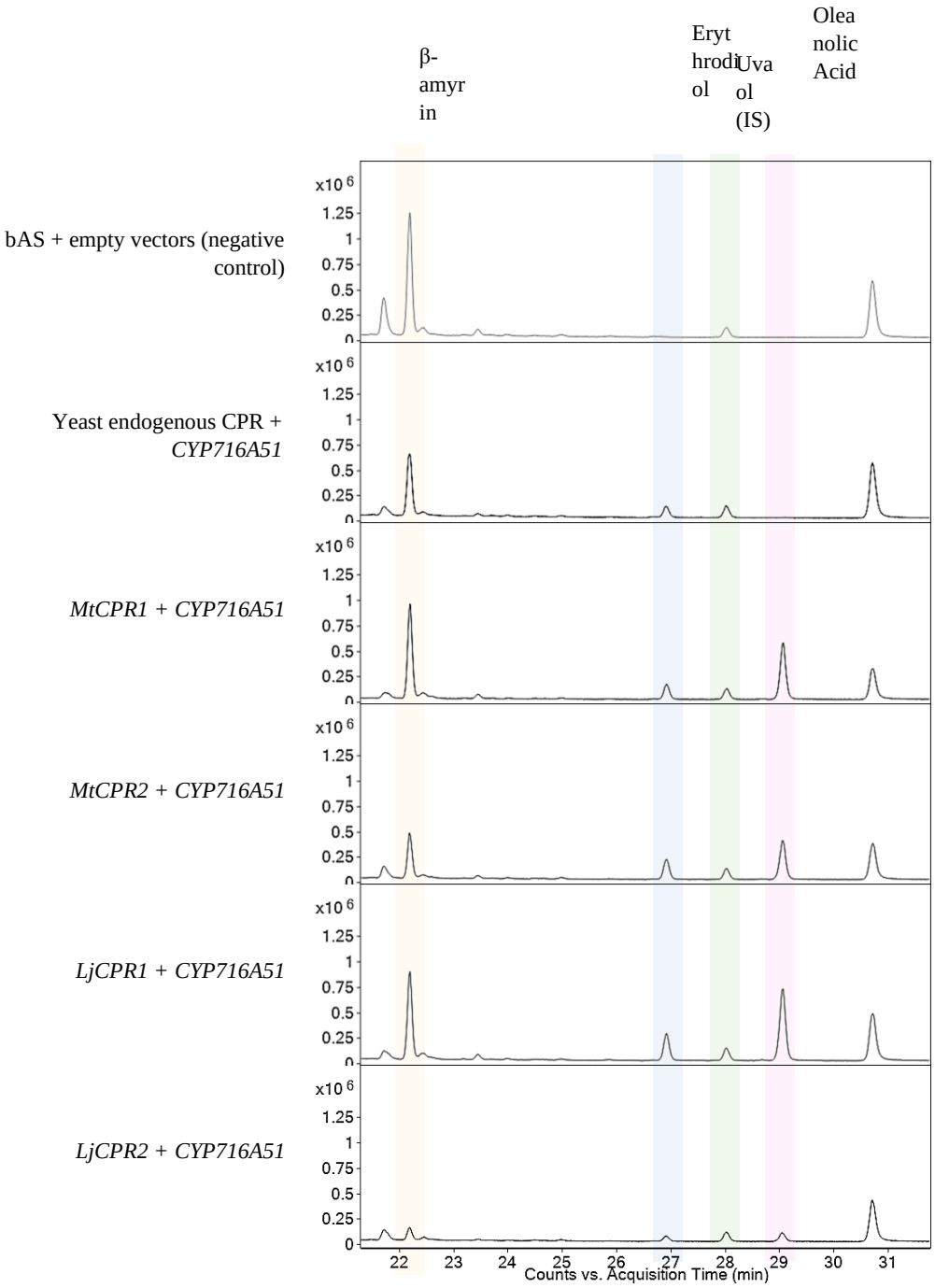

Figure S14. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing PSIII yeast harboring MtCPRs and LjCPRs paired with CYP716A51 using HP 5-MS column with optimized method with uvaol as internal standard.

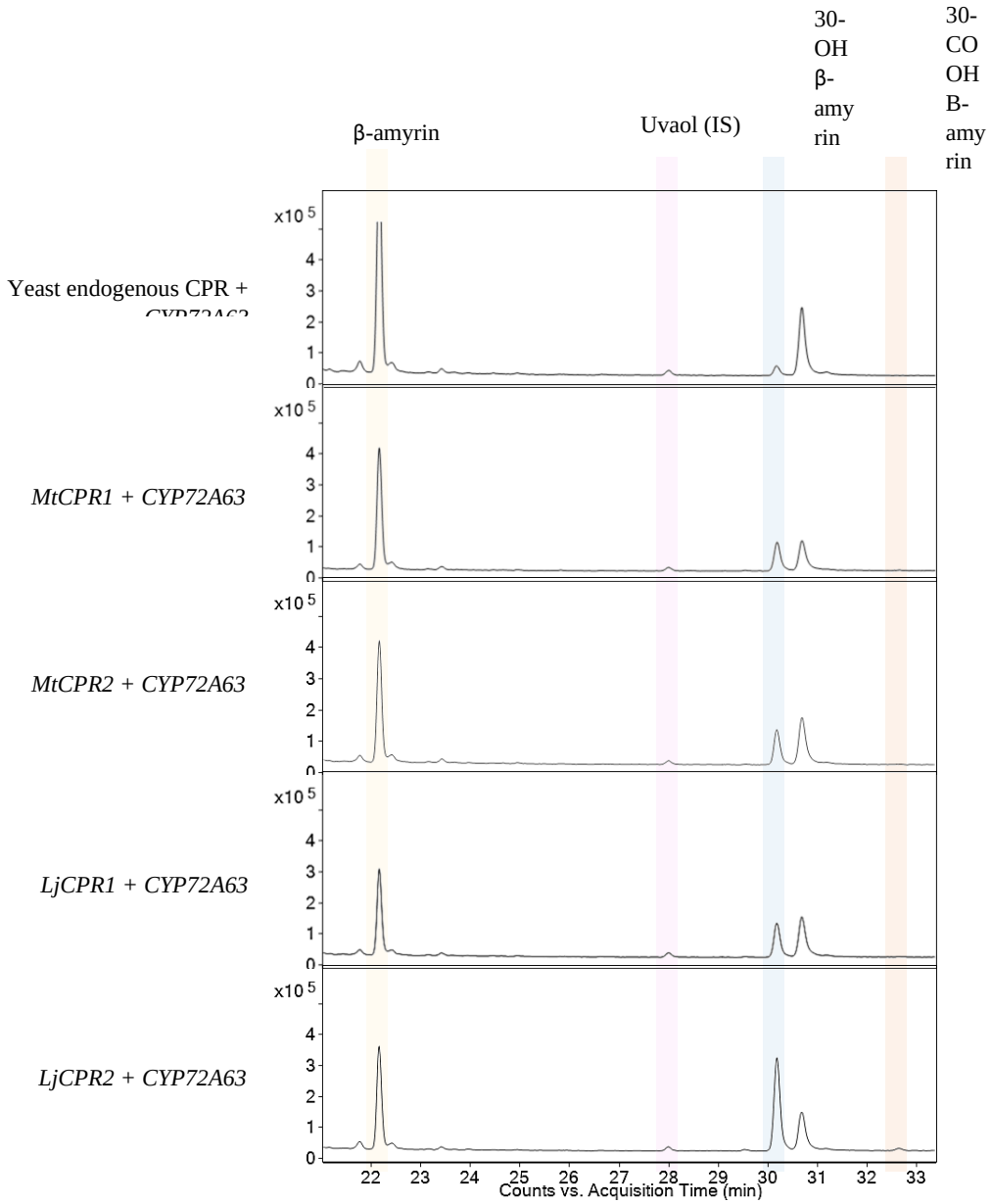

Figure S15. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing PSIII yeast harboring MtCPRs and LjCPRs paired with CYP72A63 using HP 5-MS column with optimized method with uvaol as internal standard.

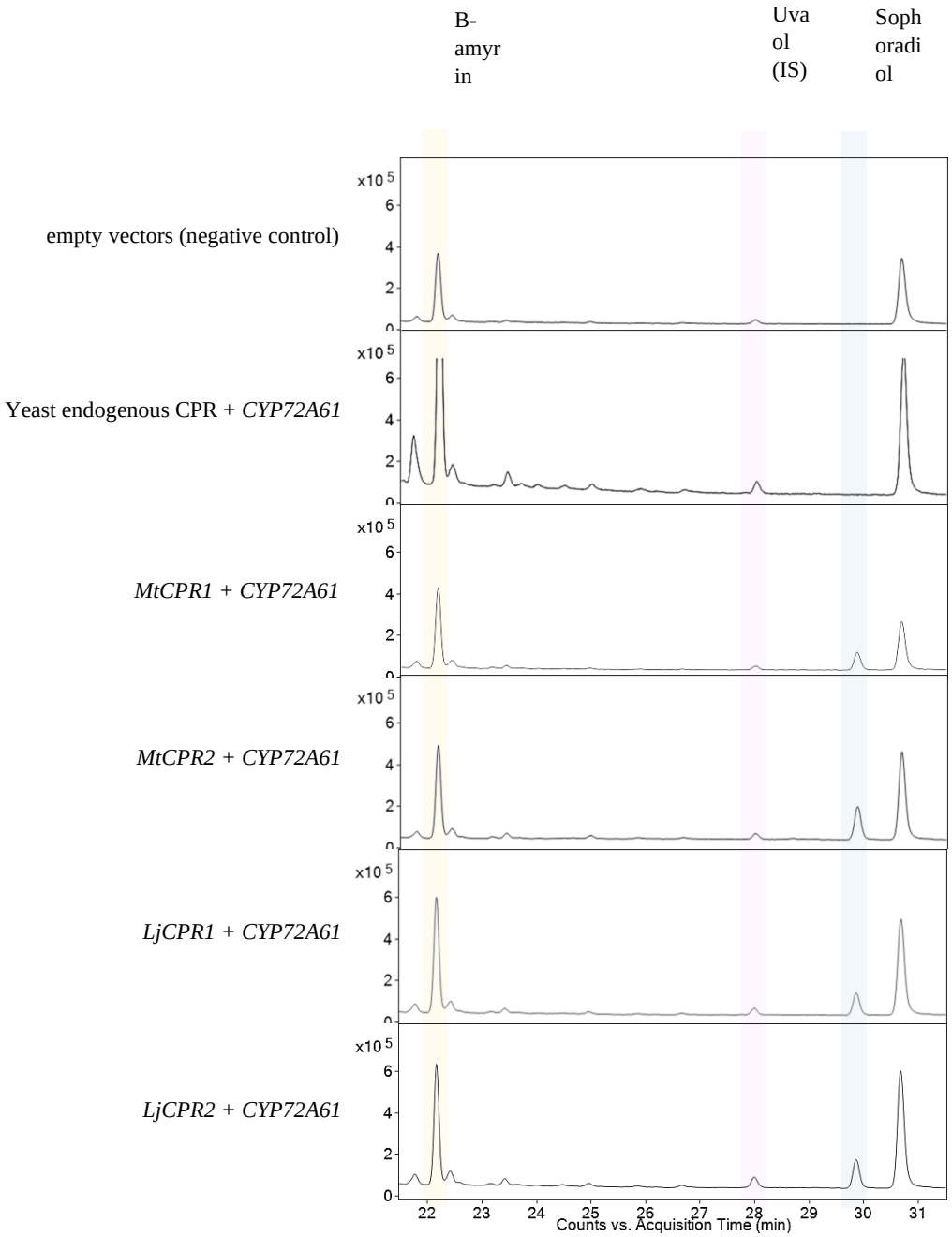

Figure S16. GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing PSIII yeast harboring MtCPRs and LjCPRs paired with CYP72A61 using HP 5-MS column with optimized method with uvaol as internal standard.

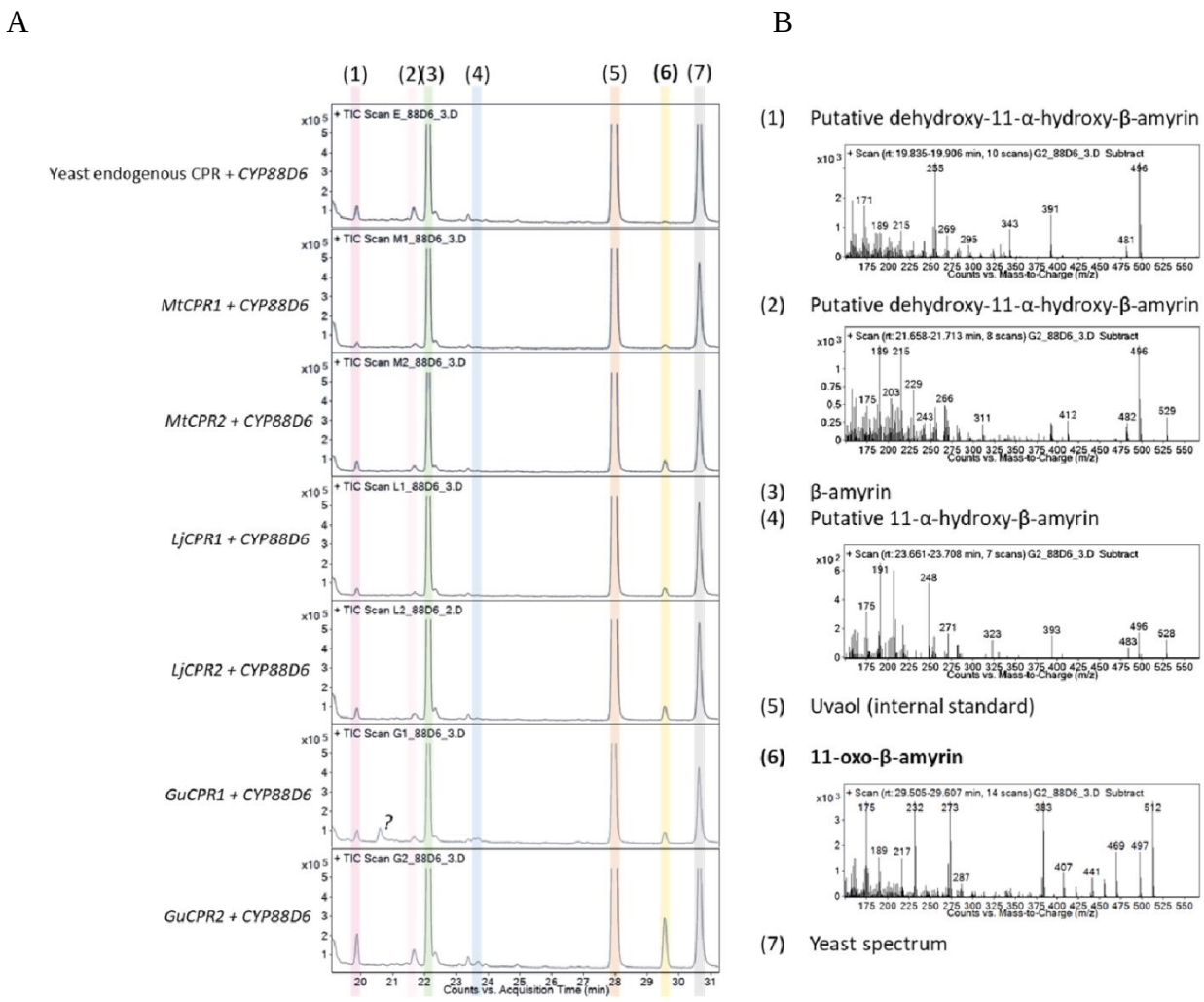

Figure S17. (A) GC-MS chromatogram of triterpenoids extracted from β-amyrin-producing PSIII yeast harboring MtCPRs, LjCPRs, and GuCPRs paired with CYP88D6 using HP 5-MS column with optimized method with uvaol as internal standard. (B) Mass spectrum of each annotated peaks.

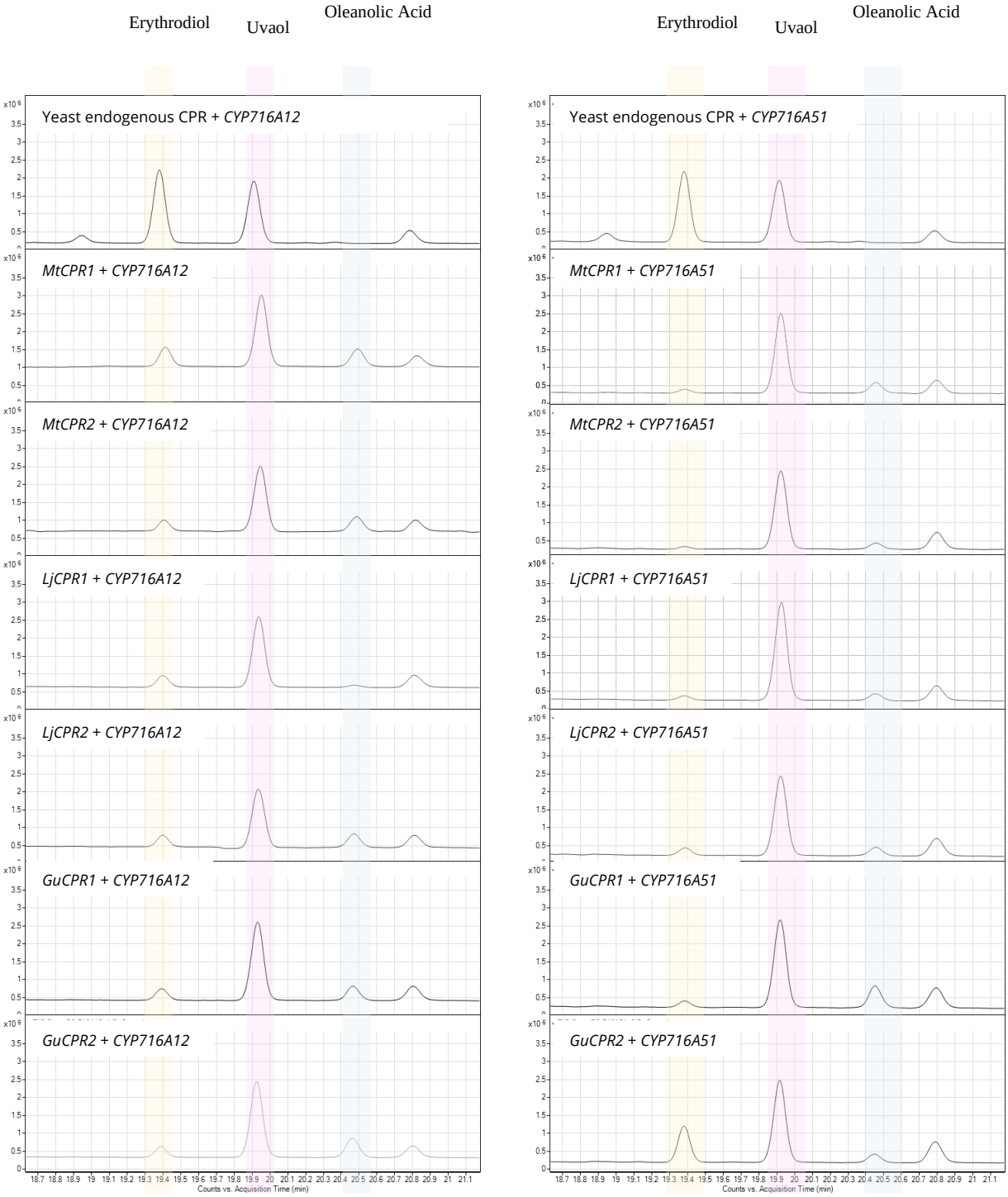

Figure S18. GC-MS chromatogram of triterpenoids extracted from INV*Sc*1 yeast harboring MtCPRs, LjCPRs, and GuCPRs paired with CYP71A612 and CYP716A51 supplemented with erythrodiol as substrate using HP 5-MS column with uvaol as internal standard.