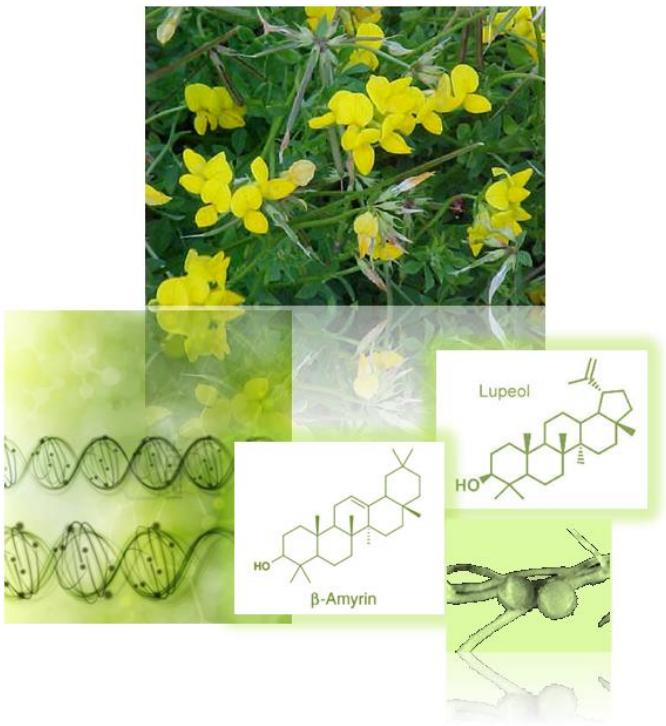
University of Thessaly

Department of Biochemistry and Biotechnology

Afrodite Krokida

GENOME ORGANIZATION, FUNCTIONAL ANALYSIS OF BIOSYNTHETIC GENES AND METABOLIC DIVERSITY OF TRITERPENES IN LEGUMES



2013

Genome Organization, Functional Analysis of Biosynthetic Genes and Metabolic Diversity of Triterpenes in Legumes

Three member committee

***Papadopoulou Kalliope**, Associate Professor, Department of Biochemistry & Biotechnology, University of Thessaly, Expertise: Plant Biotechnology*

***Osbourn E. Anne**, Professor/Head, Department of Metabolic Biology, University of East Anglia/John Innes Centre*

***Karpouzas Dimitrios**, Assistant Professor, Department of Biochemistry & Biotechnology, University of Thessaly, Expertise: Environmental Microbiology & Biotechnology*

Seven member committee

***Haralampidis Kosmas**, Assistant Professor at University of Athens, Faculty of Biology, Department of Botany, Molecular Plant Development Laboratory*

***Hatzopoulos Polydeukis**, Professor at Agricultural University of Athens, Department of Biotechnology, Director at Laboratory of Molecular Biology*

***Kanellis Angelos**, Professor of Biotechnology of Pharmaceutical Plants, Laboratory of Pharmacognocy, Department of Pharmaceutical Sciences, Aristotle University of Thessaloniki*

***Karpouzas Dimitrios**, Assistant Professor, Department of Biochemistry & Biotechnology, University of Thessaly, Expertise: Environmental Microbiology & Biotechnology*

***Osbourn E. Anne**, Professor/Head, Department of Metabolic Biology, University of East Anglia/John Innes Centre*

***Papadopoulou Kalliope**, Associate Professor, Department of Biochemistry & Biotechnology, University of Thessaly, Expertise: Plant Biotechnology*

***Psarra Anna-Maria**, Assistant Professor, Biochemistry, Department of Biochemistry & Biotechnology, University of Thessaly*

Afrodite Krokida

Genome Organization, Functional Analysis of Biosynthetic Genes and Metabolic Diversity of Triterpenes in Legumes

Afrodite Krokida

2013

"Genome Organization, Functional Analysis of Biosynthetic Genes and Metabolic Diversity of Triterpenes in Legumes"

PhD Thesis

University of Thessaly

Abstract

Triterpenes are plant natural products that are formed by cyclization of 2,3-oxidosqualene. The triterpene scaffolds formed by enzymes known as oxidosqualene cyclases (OSCs) may be subsequently modified by enzymes such as cytochrome P450s and glucosyltansferases. Genes for triterpene biosynthetic pathways exist as metabolic gene clusters in oat and *Arabidopsis thaliana* plants. Four loci of candidate metabolic gene clusters were identified in the genomes of the model legumes *Lotus japonicus* and *Medicago truncatula*, in regions flanking *OSC* genes. Amongst these was a *L. japonicus* cluster containing the *AMY2 OSC* gene along with genes for two different classes of cytochrome P450 and a reductase. This cluster represents a new pathway for triterpene biosynthesis in legumes. Expression of *AMY2* cluster genes in *Nicotiana benthamiana* identified a novel triterpene structure, dihydrolupeol, produced by AMY2, and allowed the biochemical characterization of a new plant cytochrome P450, CYP71D353, which catalyses the formation of 20-hydroxybetulinic acid in a sequential three-step oxidation of 20-hydroxylupeol. The cluster genes are highly co-expressed during root and nodule development, in hormone-treated plants and under different environmental stresses. A possible role for the gene cluster in plant development was revealed by the production of RNA silencing lines. These experiments also revealed an unexpected mechanism for the regulation of cluster gene expression by long distance transcriptional gene silencing.

The role of *βAS* gene of *M. truncatula* was identified. The expression pattern of *βAS* gene was studied under various developmental conditions. Knockout

experiments in which transgenic *M. truncatula* for *βAS* were generated through *A. rhizogenes* transformation. Furthermore, silenced plant lines shown an increased nodule number when compared with control plants, disclosing a possible involvement of *βAS* in nodulation.

The role of lupeol synthase, encoded by *OSC3*, and its product, lupeol, was identified in developing roots and nodules of the model legume *L. japonicus*. The expression patterns of *OSC3* in different developmental stages of uninfected roots and in roots infected with *Mesorhizobium loti* were determined. The tissue specificity of *OSC3* expression was analyzed by *in situ* hybridization. Functional analysis, in which transgenic *L. japonicus* roots silenced for *OSC3* were generated, was performed. The absence of lupeol in the silenced plant lines was determined by GC-MS. The expression of *ENOD40*, a marker gene for nodule primordia initiation, was increased significantly in the *OSC3*-silenced plant lines, suggesting that lupeol influences nodule formation. Silenced plants also showed a more rapid nodulation phenotype, consistent with this. Exogenous application of lupeol to *M. loti*-infected wild-type plants provided further evidence for a negative regulatory effect of lupeol on the expression of *ENOD40*. The synthesis of lupeol in *L. japonicus* roots and nodules can be solely attributed to *OSC3*. Taken together, our data suggest a role for lupeol biosynthesis in nodule formation through the regulation of *ENOD40* gene expression.

Number of Preliminary Pages: 17

Number of Pages of PhD Thesis: 213

Number of Tables: 11

Number of Figures: 70

Number of Photographs: 5

Number of Appendices: 2

Number of References: 629

Prologue

This PhD thesis was conducted at the research group of Plant & Environmental Biotechnology, at Department of Biochemistry & Biotechnology, University of Thessaly. The functional analysis of *AMY2* cluster genes in *Nicotiana benthamiana* and the GC-MS analysis of *OSC3*-silenced plant lines were carried out at John Innes Centre, Norwich, UK. George Lomonossoff provided the pM81-FSC1 plasmid vector, Thomas Ott the pUBI-GWS-GFP vector and Peter M. Waterhouse the pHannibal vector. The work has been supported by the ESF / NSRF programme Heracleitus II.

Declaration

I declare that the work contained in this thesis, submitted by me for the degree of Ph.D., is my own original work, except where due reference is made to other authors, and has not been submitted by me for a degree at this or any other university.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co-financed by Greece and the European Union

Acknowledgements

I would like to thank my supervisor Papadopoulou Kalliope for all of her help and support over the last five years, through the good times and the bad. I would also like to thank Osbourn E. Anne and Karpouzas Dimitrios for their guidance and advice all of these years. I would like to thank the seven member committee, Haralampidis Cosmas, Hatzopoulos Polydeukis, Kanellis Angelos, Karpouzas Dimitrios, Osbourn E. Anne, Papadopoulou Kalliope and Psarra Anna-Maria, for their invaluable comments and remarks that improved the quality of this work.

I would like to thank Dr Delis Costas and Dr Tsikou Daniela for their wise words and invaluable advice during the conduction of the experimental procedures. I would also like to thank Dr Luis M. Peña-Rodríguez for completing all the GC-MS analyses, Dr Katrin Geisler for her help in infiltration and metabolite extraction experiments and all the members of Osbourn E. Anne's laboratory.

I would like to thank the present and past members of the research group of Plant & Environmental Biotechnology, PhD candidate Katsarou Dimitra, PhD candidate Roussidou Konstantina, PhD candidate Perruchon Chiara, MSc student Psarrakou Ioanna-Sofia, Dr Mirtziou Ioanna, Karagounis Constantine, PhD candidate at the University of Oxford and Koufogiannis Asterios, MSc student at the Aristotle University of Thessaloniki, for their help, excellent collaboration and enjoyable time in the lab. I would especially like to thank PhD candidate Karas Panagiotis for his help, support and friendship for all of these years. I would also like to thank anyone in the Department of Biochemistry & Biotechnology that contributed in any way to this work.

I would like to thank my friends Chanika Eleni, Georgiadou Dafni, Kyriakopoulou Zacharoula and Tomatsidou Anastasia for their advice, long conversations and for standing by me upon joys and difficulties.

Last but not least, the biggest thank you to my parents, Ioannis and Harikleia, and my sister, Vasiliki, for their continuing support throughout all of my studies.

To my parents, Ioannis and Harikleia,

and my sister, Vasiliki

Table of contents

Table list

Figure list

List of photographs

A. INTRODUCTION

CHAPTER A. INTRODUCTION

A.1. *Lotus japonicus*: a model plant for legume research

A.1.1. Legumes and their agricultural importance

Fabaceae or Leguminosae is the third largest family of flowering plants, consisting of 650 genera and 18.000 species, many of which are of agronomic importance. Legume plants serve as source of food for humans, feed for livestock and raw materials for industries (Graham and Vance, 2003). Legumes are capable to fix nitrogen in a symbiotic association with the soil bacteria of genus *Rhizobium*, providing these plants and subsequent crops with free and renewable quantities of nitrogen (N). It is estimated that cultivated legumes are able to fix between 40 million to 60 million tons of N annually (Smil, 1999). Some of the globally important food legumes are bean (*Phaseolus vulgaris*), soybean (*Glycine max*), pea (*Pisum sativum*), chickpea (*Cicer arietinum*), broad bean (*Vicia faba*), pigeon pea (*Cajanus cajan*), cowpea (*Vigna unguiculata*) and lentil (*Lens esculenta*). Soybean is the most widespread legume as most of the seeds produced are used as fodder for livestock and for oil extraction. Legumes produce diverse groups of natural products, some of which exhibit health-promoting properties, such as isoflavones (Dixon, 2003; Dixon 2004).

Legume productivity is being affected by numerous biotic and abiotic barriers, as soil salinity, drought, nutrient limitations, and various diseases and pests (Graham and Vance, 2003). Developing tolerant plants to these impediments is a main effort of breeding programs. Research on the non-legume model species *Arabidopsis thaliana* and rice can investigate many fundamental phenomena of plant biology. However, non-legume plants are not able to provide insights into many aspects of legume biology, including symbiotic nitrogen fixation. Legume plants are responsible for answering these intriguing rising questions. Cultivated legumes are considered as poor model systems for investigation, due to the often large genome sizes, the tetraploidity or genome duplications of some species, the low – density culture and the difficulties in plant transformation and regeneration (Udvarti *et al*., 2005). As a result, two other species, *Lotus japonicus* (Handberg and Stougaard, 1992) and *Medicago truncatula*

(Barker *et al*., 1990) have been chosen as models for legume research. The main biological differences between the two model species are: perennial growth and plant rejuvenation through side shoots, indeterminate flowering and ample seed production, straight seed pods with easily accessible seeds and determinate nodulation in *Lotus* versus annual growth, solid spiral seed pods and indeterminate nodulation in *Medicago* (Udvarti *et al*., 2005).

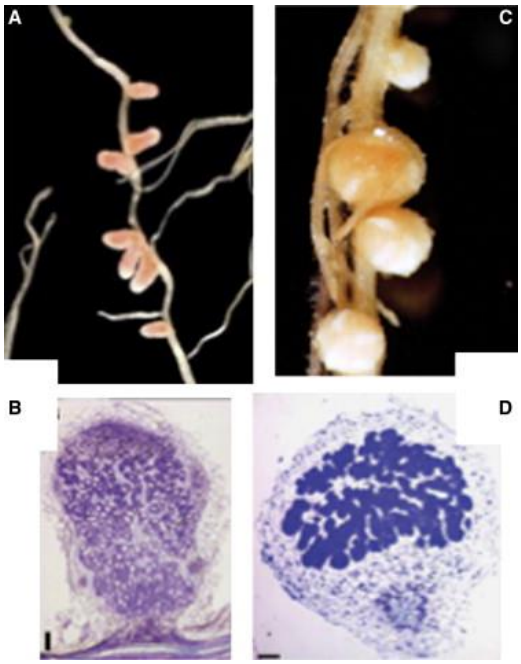


Figure 1. Legumes develop two main types of nodules upon inoculation. *M. truncatula* develop elongated nodules (A) of the indeterminate type (B), while *L. japonicus* develop spherical nodules (C) of the determinate type (D) (Desbrosses and Stougaard, 2011).

A.1.2. The model legume plant *Lotus japonicus* (*L. japonicus*)

L. japonicus was discovered centuries ago at the ancient capital of Japan Kyoto. The first researchers who established *L. japonicus* (Gifu progeny B 129) as model plant were Kurt Handberg and Jens Stougaard (1992, Aarhus University, Denmark). Since then, *L. japonicus* Gifu B-129 had a lab worldwide distribution. The Ecotype Miyakojima (MG-20) is also an extensively used ecotype especially bred to get progeny after its crossing with Gifu ecotype (Kawagushi, 2000). The derived population is useful for the construction of genetic linkage mapping, for map based cloning and QTL identification. Moreover, MG-20 was the ecotype used for the genome sequencing of *L. japonicus* in

2008 (Sato *et al.*, 2008). Finally MG-20 is the ecotype serving for the *Lotus* sequencing genome program undertaken by the Kazusa DNA Research Institute (http://www.kazusa.or.jp/e/; Udvardi *et al.,* 2005).

Figure 2. *L. japonicus* plant morphology (A), flowers (B) and nodules (C) established on the roots.

The genus *Lotus* is an Angiosperm true dicotyledonous plant belonging to the Subfamily Faboideae or Papilioneae of the Fabales (Figure 3). The morphology of *L. japonicus* plant is shown in Figure 2. As mentioned above, *L. japonicus* have been chosen as a typical model legume plant because of its useful characteristics. It is an autogamous diploid legume with small genome size (~470Mb) with simple architecture. In addition, it has small stature, large seed set, short generation time (~2-3 months) and it is self-fertile (Handberg and Stougaard, 1992). *L. japonicus* recourses for molecular genetics and genomics, such as genetic maps, genome sequences and large-scale expressed sequence tags (ESTs) are being developed (Hayashi *et al.*, 2001; Sato *et al.*, 2001; Asamizu *et al.*, 2004), and a protocol for a reliable transformation technique with *Agrobacterium* strains has been established (Aoki *et al.*, 2003). TILLING and LORE1 mutant collections have been established (McCallum *et al.*, 2000a, b; Perry *et al.*, 2003; Le Signor *et al.*, 2009; Urbanski *et al.*, 2012) and mutants in diverse biological processes specific to legumes such as nitrogen fixation (Heckmann *et al.*, 2006; Perry *et al.*, 2009), and those common to flowering plants such as flower morphogenesis or seed production have been characterized and the responsible genes have been identified (Vriet *et al.*, 2010; Chen and Dubcovsky, 2012; Credali *et al.*, 2013).

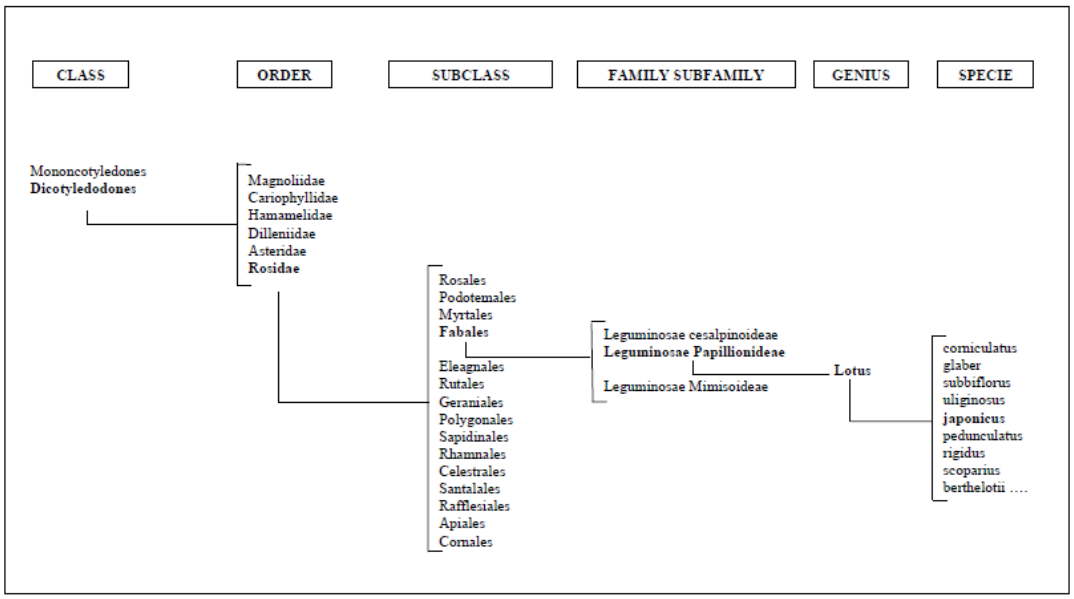


Figure 3. Taxonomic classification of *L. japonicus* (Aramrarri, 1999).

A.1.3. The soil bacteria of the genera *Mesorhizobium loti* (*M. loti*)

Rhizobia designate a collective name of the genera *Rhizobium, Sinorhizobium, Mesorhizobium* and *Bradyrhizobium*. These prokaryotes colonize the rhizosphere where they perform nitrogen fixing symbiosis with leguminous plants. Both *M. loti* and *B. loti* can inoculate *L. japonicus*. However *B. loti* induces only infected non N-fixing nodules. By contrast, *M. loti* is able to form determinant-type globular nodules and perform nitrogen fixation on several *Lotus* species (Marquez, 2005).

The genome of *M. loti* consists of a single chromosome (7,036,071 bp) and two plasmids, designated as pMLa (351,911 bp) and pMLb (208,315 bp). The chromosome consists of 6,752 potential protein coding genes, two sets of rRNA genes and 50 tRNA genes representing 47 tRNA species. The genome of plasmids pMLa and pMLb, contains 320 and 209 potential protein-coding genes, respectively. To date, the common strains used globally in laboratories are: R7A, NZP2235, JRL501 (Niwa *et al.*, 2001; Kawagushi *et al.,* 2000), MAF303099, and TONO (Kawaguchi, 2000). Most of these strains are equipped with GUS and GFP fusion reporter genes to monitor the process of infection from the beginning until the formation of the nodules.

A.1.4. Nodule formation and symbiosis

The first plant organs involved in the nodulation process are the root hairs. The root hairs of higher plants represent an important extension of the root surface, which are able to sense and uptake nutrient from a continuously fluctuating rhizosphere. During symbiosis, root hairs contained in a fragment of the root at 0.5 cm far from the root tips, start a new morphogenesis process representing the first event of the symbiotic interaction (Szczyglowski and Amyot, 2003). The first steps involve the reciprocal recognition between the symbiont and the host plant. Bacteria are attracted by root exudates (flavonoids, sugars, volatile compounds). Plants synthesize these compounds under biotic (e.g. pathogen attacks) or abiotic stimulations (e.g. loss of the oligo-mineral retention capacity of the rhizosphere). These secreted plant flavonoids and phenolic compounds are able to induce the expression of *Nod* genes in rhizobia (Downie, 1998) and subsequently the synthesis of Nod factors (NFs). The initial response of root hairs to the production of NFs involves the establishment of *de novo* polar root hair tip growth and curling, which leads to the formation of typical ''shepherd's crook'' structures (Lhuissier *et al.,* 2001). The curled root hairs entrap the bacteria and serve as a starting point for the initiation of the infection procedure. In a process that is not well understood, the rhizobium gain entry into the plant by invagination of the root hair membrane and the formation of an infection thread (IT). The IT grows down through the root hair and eventually enters and ramifies in the plant root cortex. Cortical cells are triggered to divide in advance of the infection thread, eventually forming the nodule primordium. The infection thread terminates in a cell that will become infected. The rhizobia are released via endocytosis into the infected cell. The plant cell forms a membrane that surrounds the intracellular symbiont forming the symbiosome. It is within the symbiosome that the rhizobia fix nitrogen that is provided to the plant. The nodule is a true organ with tissue differentiation. For example, there are infected cells where nitrogen fixation takes place and associated uninfected cells that play a role in nitrogen assimilation. The plant gains a steady supply of nitrogen, which is often a limiting nutrient in many environments. In return,

the bacterium gains a steady supply of carbon (Boogerd and van Rossum, 1997; Guinel and Geil, 2002).

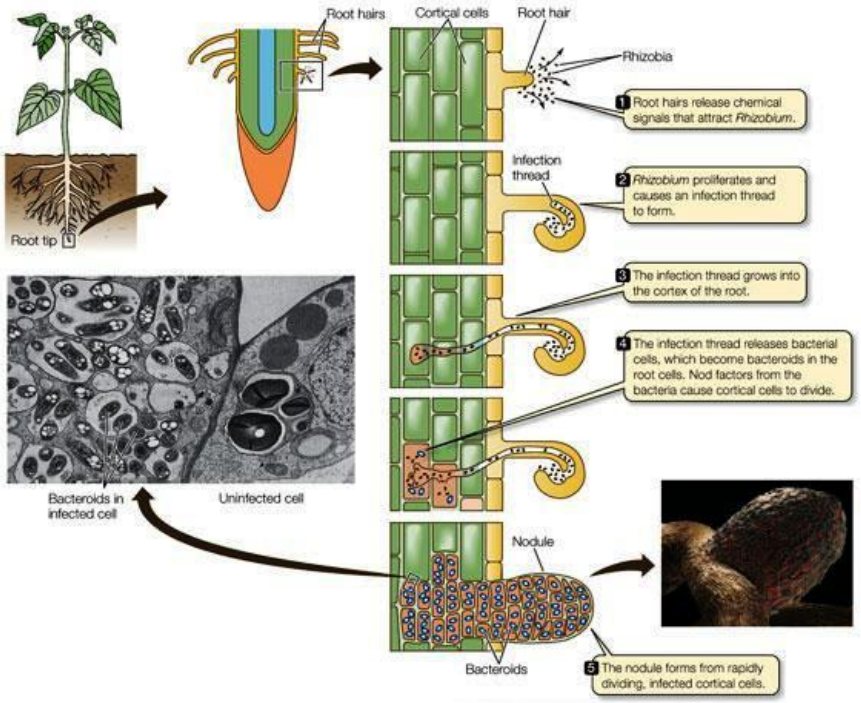


Figure 4. Steps of nodule formation.

A.2. Plant secondary metabolism

A.2.1. Secondary metabolites in general

Plant metabolic pathways are partitioned into two major classes. Intermediary or primary metabolism includes vital reactions similar in most organisms (Luckner, 1972). Primary metabolism facilitates plant to utilize water, carbon dioxide and minerals in order to form nucleic acids, amino acids and proteins, carbohydrates and carboxylic acids (Seigler, 1977). By these means, plant organisms are able to make and maintain cells. Primary compounds have ancient origins and genes required for their formation are highly conserved across all known plants.

By definition, secondary metabolites or natural products are low-molecular compounds, generally not essential for the basic processes of plant growth and development. Some decades ago it was considered that secondary metabolites represent either the end products of metabolism without function or the detoxification products of simple products or plant metabolism (Seigler, 1977). Alternatively to the

"waste product" hypothesis, Whittaker and Feeny were the first to express the opinion that secondary compounds are involved in plant defense against animals, fungi, bacteria and other plants (Whittaker and Feeny, 1971). Eventually, there is a vast body of information showing that secondary metabolites in plants exist in a state of dynamic equilibrium and are not static end products of metabolism. Many of the *Arabidopsis thaliana* secondary substrates have been demonstrated to function in the primary processes of growth and development or resistance to antibiotic stresses (Landry *et al*., 1995; Fiskus and Booker, 2002; D'Auria and Gershenzon, 2005). For example, representatives of both phenylpropanoids (sinapate esters) and flavonoids (flavonols) are involved in protection against UV-B radiation (Landry *et al*., 1995; Fiskus and Booker, 2002). The class of flavonoids is implicated in primary physiological functions, such as auxin transport, regulation of seed longevity, and dormancy (D'Auria and Gershenzon, 2005). A role in regulating seed longevity and dormancy has been attributed to the proanthocyanidins of the *A. thaliana* seed coat (Debeaujon *et al*., 2003). Abscisic acid (ABA) and gibberellin acid (GA) are the primary endogenous factors that regulate the transition from dormancy to germination, and they regulate this process antagonistically (Bewley, 1997; Gubler *et al*., 2005; Seo *et al*., 2006; Yano *et al*., 2009). ABA is essential for the induction and maintenance of seed dormancy, while GA is required for the release of dormancy and for the initiation of seed germination (Seo *et al*., 2006; Yano *et al*., 2009). Strigolactones (SLs) promote the establishment of mycorrhizal symbiosis which mainly facilitates the phosphate acquisition from the soil (Akiyama *et al*., 2005). SLs were also found to play a key role in shoot branching inhibition (Gomez-Roldan *et al*., 2008; Umehara *et al*., 2008). Their biological functions were further explored and it was discovered that they also exert their effects on different developmental processes including root development, seed germination, hypocotyl elongation, and secondary growth (Cheng *et al*., 2013).

Plants elaborate a vast array of over 500.000 natural products (Rasbery, 2007). Most secondary compounds are derived from the isoprenoid, phenylpropanoid, alkaloid or fatty acid/polyketide pathways. This rich diversity of molecules formed is an evolutionary process. The core structure of secondary molecules is further modified by

oxidative enzymes, such as cytochrome P450 monooxygenases and 2-oxoacid-dependent dioxygenases, also resulting in enrichment of structure diversity. Natural inhibitory compounds can be synthesized during normal growth and development (preformed antimicrobial compounds or phytoanticipins) (VanEtten, 1994; Osbourn, 1996; Papadopoulou, 1999). Otherwise, secondary chemicals may be absent from healthy plants, accumulating only in response to pathogen attack or stress (phytoalexins) (Muller and Borger, 1940; Paxton, 1981).

Collectively, secondary compounds are frequently associated with the distinguishing features of particular taxa and serve as disease resistance determinants, scents, colors and flavors. In general, it is considered that secondary metabolites have important ecological functions, participating in plant responses to biotic and/or abiotic barriers (Harborne, 1999; Field *et al.*, 2006), and the ability of different plant species/lineages to synthesize distinct secondary metabolites is likely to have been key to the survival and diversification of plant species. However, some of these molecules accumulate during normal plant growth and development, suggesting additional functions in such processes (Abe *et al.*, 1988; Guhling *et al.*, 2006). The ecological importance of secondary compounds is further elucidated by observations of mutants unable to form specific secondary metabolites. These mutants either fail to survive in their environment, or vanish in competition with wild type plants because of their defective genotypes (Brown *et al.*, 2005; Zuest *et al.*, 2011). Therefore, secondary metabolites are considered necessary for the successful existence of plant species in the natural environment, like primary compounds.

A.2.2. Terpenes and triterpenoids

According to Croteau *et al.* (2000), terpenes and terpenoids, alkaloids and phenolic compounds are the three broad categories of plant secondary metabolites. Terpenes or isoprenoids are the most numerous and structurally diverse plant natural products. Based on the abundance and diversity, terpenes offer much potential in an array of industrial and medicinal applications (Croteau *et al.*, 2000). The distribution of terpenes in nature has been studied thoroughly. In loblolly pine (*Pinus taeda*) core samples, the highest concentrations of terpenes were detected in heartwood, lowest in outer

sapwood, and moderate levels in the inner sapwood (Thompson *et al*., 2006). Methyl jasmonate was applied onto foliage of Norway spruce trees (*Picea abies*) resulting in a two-fold increase of terpenes within the needles (Martin *et al*., 2003). Additionally, it been shown that flowers can emit terpenoids to attract pollinating insects (Maimone and Baran 2007). Interestingly, terpenoids have also been shown to attract beneficial insects, which feed on the herbivorous ones (Kappers *et al*., 2005). Kessler and Baldwin (2001) have shown that herbivorous insects can induce the release of terpene exudates and signaling molecules, which attract predatory species. According to Cheng *et al*. (2007), terpenes exhibit roles above and below ground in attracting predatory species upon herbivore attack. Additionally, it is proposed that terpenes may act as chemical messengers that influence the expression of genes involved in plant defense mechanisms or even influence gene expression of neighboring plants (Cheng *et al*., 2007).

Except for their contribution to the plant status, terpenes offer potential in industries and medicine. Pharmaceutical and food industries have exploited them for their potentials and effectiveness as medicines and flavor enhancers. Perhaps the most widely known terpene is rubber, which has been used extensively by humans. Other important terpenes include camphor, menthol, pyrethrins (insecticides), cleaners, antiallergenic agents, and solvents (Croteau *et al*., 2000). Murata *et al*. (2008) extracted numerous compounds from stem bark of the cape ash (*Ekebergia capensis)* growing in Kenya. Ten of these were triterpenes, showing promising application towards a newer and more effective drug against *Plasmodium falciparum*, the causative agent of malaria. Artemisinin, a sesquiterpene originally sourced from *Artemisia annua*, is used in combination therapy for malaria (Ro *et al*., 2006; Ansari *et al*., 2013). Taxol (paclitaxel) is a diterpenoid used against numerous cancers (Vanisree *et al*., 2004).

Triterpenes are a major subgroup of the terpene superfamily (Xu *et al*., 2004), counting more than 20.000 members with chemical diverse structures. During plant growth and development, triterpenoids are essential precursors for cell membranes and steroid hormones (Nes and Heftmann, 1981; Benveniste, 2004; Suzuki *et al*., 2006). Additionally, they play roles in defense against pathogen attack (Phillips *et al*., 2006)

and they participate in protection against abiotic stimulations (Wang *et al.*, 2010). Their pharmacological properties, such as anti-cancer, anti-inflammatory, anti-oxidant activities (Waller and Yamasaki, 1996; Oliviera *et al.*, 2005; Shai *et al.*, 2008; Augustin *et al.*, 2011; Osbourn *et al.*, 2011), and their applications in food industry, render triterpenoids as a very important group of natural products (Figure 5). Triterpenes are synthesized from mevalonate via a 30-carbon intermediate, 2,3-oxidosqualene (2,3-OS). The biosynthetic intermediate 2,3-OS is region- and stereo-specifically cyclized by various 2,3-OS synthases or cyclases (OSCs), giving rise to either tetracyclic sterols, or pentacyclic triterpenes, like α- and β-amyrin and lupeol (Abe *et al.*, 1993; Abe, 2007; Xu *et al.*, 2004). Simple triterpenes molecules can be further modified and accumulate as triterpene glycosides (triterpenoid saponins). It has been demonstrated that triterpenoid saponins are implicated in plant defense in cereals (Papadopoulou *et al.*, 1999). Legumes produce a vast array of secondary compounds, including the triterpenes α- and β-amyrin, and lupeol (Phillips *et al.*, 2006). The biosynthetic pathway and the possible functions of triterpenes are described in detail in the next sections.

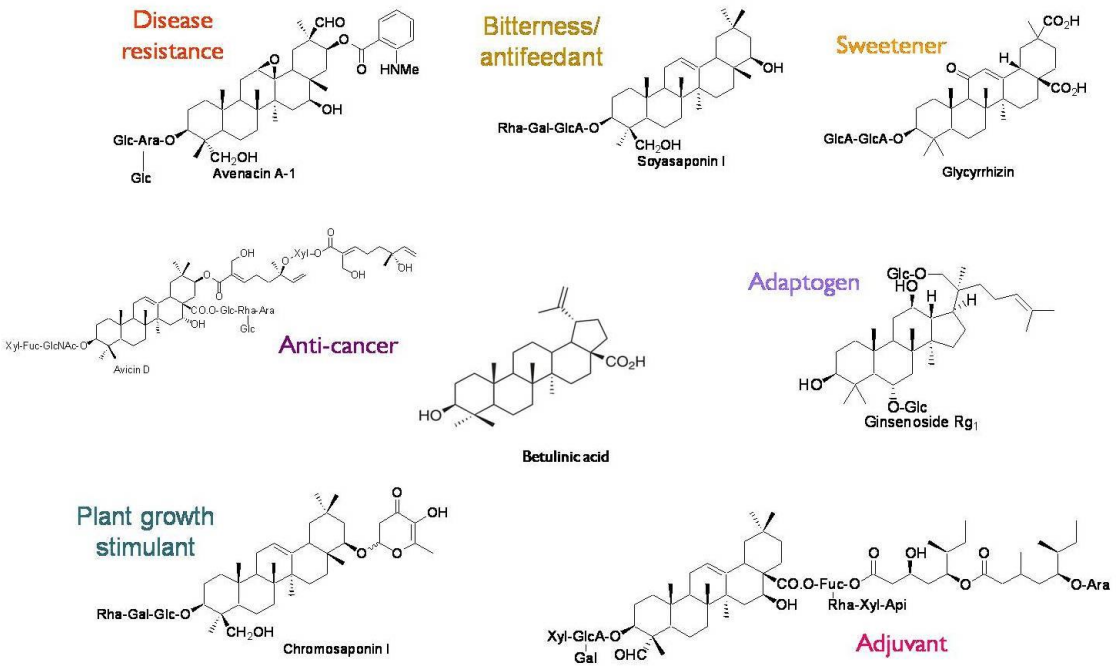


Figure 5. Paradigms of different biological functions of triterpenoids (Osbourn AE, personal communication).

A.3. Pathway to triterpene biosynthesis

A.3.1. Triterpene and sterol biosynthesis is the branch point between primary and secondary metabolism

The isoprenoid family is classified into different groups according to the number of isoprene units in their structure. One isoprene unit (C5) consists of five carbon atoms and eight hydrogen atoms. The major groups within the isoprenoid family are the monoterpenes (C10) consisting of two isoprene units, sesquiterpenes (C15) containing three isoprene units, diterpenes (C20) built by four isoprene units, triterpenes (C30) containing five isoprene units and polyterpenes built by more than eight isoprene units (McGarvey and Croteau, 1995). Regardless of their structure, all isoprenoids are synthesized from two universal precursors, the isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). Although IPP and DMAPP are the common precursors for isoprenoid production, there are two distinct pathways for isoprenoid biosynthesis in plant kingdom. The one is the well-studied mevalonate pathway (MVA) and the other is the relatively new methyl-erythritol-phosphate pathway (MEP). The MEP pathway was recently unveiled in prokaryotes capable of accumulating hopenes (Rohmer *et al*., 1993, 1996), which are the equivalent of eukaryotic sterols. The MEP pathway has been confirmed to exist in plants, is localized in plastids (Lichtenthaler, 1990) and synthesizes monoterpenes, diterpenes, tetraterpenes (carotenoids) and polyprenols (Rohmer, 1999). The MVA pathway predominates in cytosol and appears to be responsible for the biosynthesis of sesquiterpenes, sterols and triterpenes (Dubei *et al*., 2003; Liu *et al*., 2005).

In MVA pathway, mevalonate is a six-carbon intermediate, arising from the sequential condensation of three acetyl-CoA units to generate 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA), which is converted to mevalonate in an irreversible reaction catalyzed by HMG-CoA reductase (HMGR). The six-carbon mevalonate is sequentially phosphorylated and decarboxylated to generate IPP (Chappell, 1995). For triterpene and sterol biosynthesis IPP (C5) is added in a head-to-tail fashion with DMAPP (C5) to form geranyl diphosphate (GPP) (C10). Further condensation of GPP with IPP produces the larger prenyl diphosphate, farnesyl

diphosphate (FPP) (C15). FPP then dimerises and synthesizes squalene (C30) (Liu *et al.*, 2005; Kirby and Keasling, 2009). The enzyme squalene epoxidase adds an oxygen across the first carbon-carbon double bond of squalene to form the epoxide, 2,3-oxidosqualene (2,3-OS) (Laden *et al.*, 2000) (Figure 6).

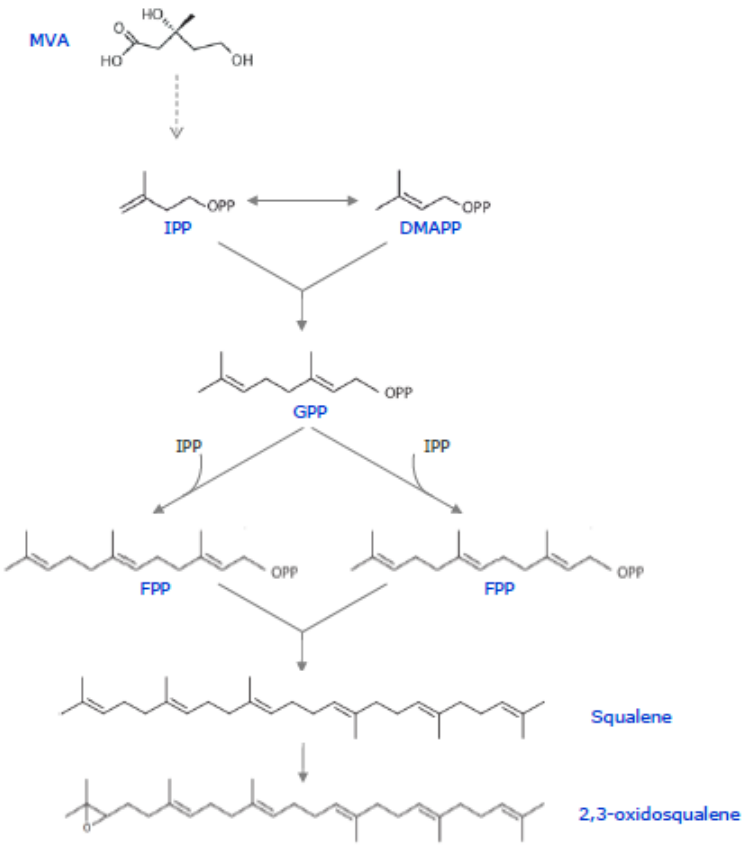


Figure 6. Biosynthesis of 2,3-oxidosqualene via the mevalonate pathway.

2,3-OS is a linear molecule that can be cyclized by specific enzymes in order to give a number of potential polycyclic products. In this stage, sterol and triterpene biosynthetic pathways diverge; the former produces membrane sterols and hormones (primary metabolites) and the latter produces a diverse range of triterpenes (secondary metabolites). These cyclization events of 2,3-OS represent the branchpoint between primary and secondary metabolism. Cyclization of 2,3-OS to distinct products is catalyzed by specialized enzymes called oxidosqualene cyclases or synthases (OSCs) (Haralampidis *et al.*, 2002), cycloartenol synthase (E.C. 5.4.99.8) and lanosterol synthase (E.C. 5.4.99.7) (for sterols) and triterpene synthases (for triterpenoids) (Abe *et al.*, 1993). Sterol and triterpene biosynthesis is illustrated in Figure 7 (Phillips *et al.*, 2006).

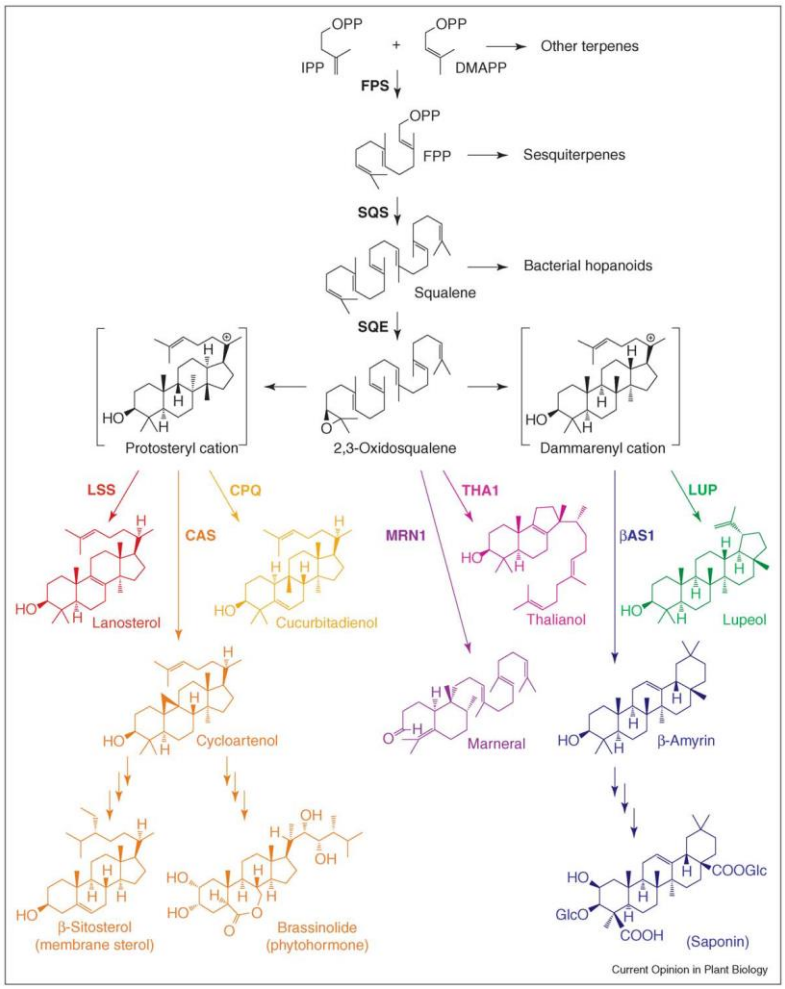


Figure 7. A simplified scheme of plant triterpenoid biosynthesis. Farnesyl diphosphate synthase (FPS) isomerizes IPP and DMAPP to form FPP, which squalene synthase (SQS) converts to squalene. Squalene epoxidase (SQE) oxidizes squalene to 2,3-OS. OSC enzymes cyclize 2,3-OS. bAS1, β-amyrin synthase; LUP, lupeol synthase; MRN1, marneral synthase; PP, diphosphate; THA1, thalianol synthase (Phillips *et al.*, 2006).

A.3.1.1. Biosynthesis of sterol precursors

Sterols have been studied extensively and although their biosynthetic pathway is still considered to be complex and incomplete, the basis outlines of their synthesis are fairly well understood (Chappell, 2002). Cyclization events of 2,3-OS by lanosterol synthase (E.C. 5.4.99.7) and cycloartenol synthase (E.C. 5.4.99.8) result in the production of the tetracyclic sterol precursors lanosterol and cycloartenol respectively. Lanosterol is the precursor of the membrane sterols cholesterol and ergosterol, in animals and fungi respectively and cycloartenol is the precursor of a number of phytosterols in plants. According to this apparently clear distinction between animal, fungal and plant

sterol OSC synthases, it was previously thought that lanosterol synthases are restricted to animals and fungi and cycloartenol synthases to plants. Recent studies identified that functional lanosterol synthases also exist in plants (Kolesnikova *et al.*, 2006) and have arisen by convergent evolution from cycloartenol synthase (Sawai *et al.*, 2006a). The role of lanosterol in plant metabolism is currently not understood. The presence of a lanosterol synthase, in addition to a cycloartenol synthase that already provides the biosynthetic pathway to membrane sterols, suggests that it may plays an alternative role in plants, probably in defense responses (Kolesnikova *et al.*, 2006). Sitosterol, stigmasterol and campesterol constitute 60 – 80% of all sterols in almost all plants and plant tissues and a mixture of 6 – 12 other chemical variants make up the rest (Chappell, 2002).

A.3.1.2. Biosynthesis of triterpene precursors

The family of triterpene OSC synthases has many more members than the sterol OSC synthase family. As a result, cyclization of 2,3-OS by the triterpene OSCs produces an extended range of cyclic products. Triterpene structures are classified into 11 main classes according to their basic carbon skeleton: dammaranes, tirucallanes, lupanes, hopanes, oleananes, taraxasteranes, ursanes, cycloartanes, lanostanes, cucurbitanes and steroids (Vincken *et al.*, 2007, Figure 8). Biosynthesis and differences between the basic skeletons are discussed in section A.3.2. β-amyrin synthases (E.C. 5.4.99.B1) and lupeol synthases (E.C. 5.4.99.B3) are two main groups of triterpene OSCs identified in monocots and dicots and are responsible for biosynthesis of the pentacyclic β-amyrin and lupeol respectively. Many triterpene OSCs produce multiple triterpene products, usually one or two major and a number of minor products. A number of new triterpene OSC synthases have been identified, mainly in *A. thaliana*. For example, thalianol synthase (E.C. 5.4.99.31) is a specific enzyme, that converts 2,3-OS to the triterpene thalianol when expressed in yeast (Fazio *et al.*, 2004). Some examples of OSC synthases that produce multiple polycyclic products are the baccharis oxide synthase (E.C. 5.4.99.51) from *Stevia rebaudiana*, and baurol synthase (Lodeiro *et al.*, 2007), marneral synthase (E.C. 5.4.99.53) (Shibuya *et al.*, 2007) and arabidiol synthase (E.C. 4.2.1.124) (Xiang *et al.*, 2006) from *A. thaliana*.

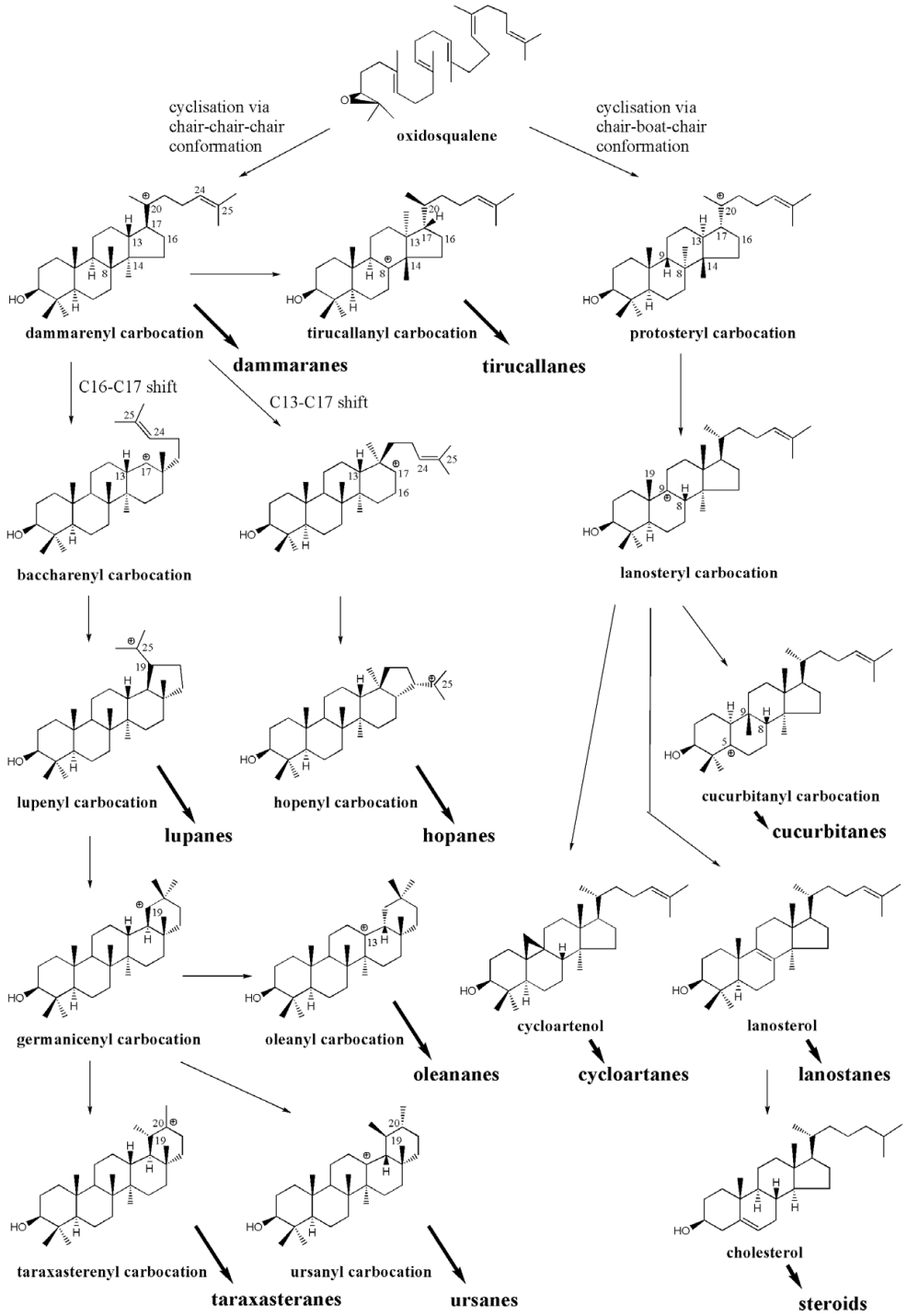


Figure 8. Cyclization events and triterpene heterogeneity (Vincken *et al.*, 2007).

A.3.1.3. Functions of sterols and triterpenes

Plants are known to accumulate a vast range of sterols and triterpenes, which raises questions about why so many variants might be needed. Until 1995 it was considered that sterols serve as structural components of membranes and as precursors of steroidal hormones in both plants and animals, and have a role in altering the fluid

dynamics of phospholipid bilayers (Chappell, 2002). In 1996 this view of sterol metabolism changed and investigations suggested the existence of sterol signal molecules that are capable of controlling developmental programs (Li *et al*., 1996; Szekers *et al*., 1996; Schrick *et al*., 2000; Jang *et al*., 2000).

By contrast, the biological functions of plant triterpenes are less well understood (Lindsey *et al*., 2003). Simple triterpenes, such as α- and β-amyrin, have been shown to be structural membrane constituents during normal growth and development in legumes, and their synthesis is up-regulated during the establishment of rhizobial and mycorrhizal symbioses (Baisted, 1971; Hernandez and Cooke, 1996; Grandmougin-Ferjani *et al*., 1999; Iturbe-Ormaetxe *et al*., 2003). Simple triterpene compounds can also be elaborated further and accumulate as structurally complex triterpene glycosides (triterpenoid saponins). Biologically, plant saponins are considered defensive compounds against pathogenic microbes and herbivores (Osbourn, 1996; Papadopoulou *et al*., 1999). Triterpenoid saponins have attracted much of scientific interest because of data showing their diverse biological activities and beneficial properties for humans, which include antifungal, antiviral, antibacterial, antitumor and antifeedant activities (Suzuki *et al*., 2002; Sparg *et al*., 2004; Huhman *et al*., 2005). Except for drugs and medicine, saponins are utilized as foaming agents, sweeteners, taste modifiers and cosmetics (Osbourn, 2003).

A.3.2. Oxidosqualene cyclases (OSCs) contribute to the triterpene heterogeneity

Oxidosqualene cyclases or synthases (OSCs) are responsible for the regio- and stereo-specific cyclization of 2,3-OS to triterpene alcohols with up to six carbocyclic rings. Squalene synthases (SCs) are the enzymes involved in the direct cyclization of squalene to pentacyclic triterpenes in prokaryotes (Ourisson *et al*., 1987). Plants encode multiple OSCs with distinct product specificity that results in diverse triterpene skeletons. According to the isoprene rule (Ruzicka, 1953) qualified OSCs fulfill three conditions for catalyzing the cyclization process. Firstly, they are characterized by a catalytic acid that initiates the cyclization process by protonating 2,3- OS. Secondly, they contain a specialized catalytic cavity that, by primarily spatial constraints of the

active site, guides the cyclizing backbone through defined intermediate stages that eventually lead to formation of specific cyclization end products. Finally, they are able to shield the reactive intermediates during the cyclization in order to prevent interfering side reactions (Ruzicka, 1953).

Mechanism-based irreversible inhibitors and mutational analysis with OSCs have indicated that the highly conserved amino acid motif DCTAE is required for substrate binding (Herrera *et al*., 1998; Abe and Prestwich, 1994) and the conserved aspartate residue within this motif (D456) has been implicated as the likely electrophilic activator in the generation of the protosteryl cation for LS (Abe and Prestwich, 1994; Corey *et al*., 1997). Similar experiments in *Alicyclobacillus acidocaldarius* indicate that the aspartate residues 376 and 377 of the DDTAV motif at the homologous position of the SC amino acid sequence are also necessary for enzyme activity (Feil *et al*., 1996). Interestingly, targeted mutations that convert the DDTAV motif of SC to DCTAE motif of OSC result in a change in substrate specificity from squalene to 2,3-OS. In addition to the DCTAE/DDTAV motifs, the highly conserved QW motif in β-strand turn occurs in all OSCs and SCs and is repeated four to eight times. These repeats are likely to be important for protein structure and stability and also for catalytic activity (Poralla *et al*., 1994; Sato *et al*., 1998). It has been proposed that the aromatic amino acids of the QW motif constitute sites of negative point charge that may interact with the intermediate cations during the cyclization process (Poralla *et al*., 1994).

OSCs can be classified as "accurate" and "multifunctional". These classifications discriminate OSCs that either exclusively catalyze cyclization of 2,3-OS into one single product or synthesize several structural distinct cyclization products in comparable amounts. However, some OSCs are able to give rise to several minor products in addition to a few major products. For example, the *A. thaliana* OSC BARS1/At4g15370 was identified to produce 22 byproducts co-occurring with the main cyclization product baruol (~90%) (Lodeiro *et al*., 2007). Enzymes such as BARS1, which give rise to several cyclization products of which one is clearly predominant, are hardly categorized into the previously groups of "accurate" and "multifunctional". As a result, the

separation of OSCs into "highly accurate", "moderately accurate" and "multifunctional" was proposed (Lodeiro *et al*., 2007; Augustin *et al*., 2011).

OSCs can be additionally divided into two groups according to the cation intermediate they utilize for 2,3-OS cyclization. Cyclization via "chair-boat-chair" produces the tetracyclic protosteryl C20 carbocation, whereas cyclization via "chair-chair-chair" results in the tetracyclic dammarenyl C20 carbocation (Xu *et al*., 2004; Phillips *et al*., 2006; Vincken *et al*., 2007). An important difference between these two skeletons is related to the stereochemistry, which is most clearly displayed by the configurations of the C8 and the C14 atoms. After cyclization of the "chair–chair–chair" conformation, the methyl group at the C8 atom is pointing upwards and the one at the C14 atom is pointing downwards, whereas the opposite happens in the case of the "chair–boat–chair" conformation (Figure 8).

A.3.2.1. OSCs that utilize the protosteryl cation

The biosynthesis of tetracyclic triterpene structures from the protosteryl cation has been studied extensively because these compounds are the initial cyclic intermediates in sterol biosynthesis. From the proton-initiated cyclization of the "chair– boat–chair" conformation of oxidosqualene, the protosteryl carbocation is obtained, which undergoes a series of hydride and methyl shifts ultimately leading to the intermediate C9 lanosteryl carbocation. This carbocation can undergo further shifts of a methyl group and a hydride to the C5 cucurbitanyl carbocation. All structures derived from this carbocation are classified as cucurbitane type compounds. The lanosteryl carbocation may also undergo deprotonation of the C19 methyl group leading to formation of a cyclopropane ring as is found in cycloartenol. All structures derived from cycloartenol are classified as cycloartane type saponins. Deprotonation of the lanosteryl carbocation gives lanosterol and all compounds derived from lanosterol are classified as lanostane type saponins. Lanosterol can also undergo demethylation and isomerisation of the double bond, leading to cholesterol. The saponins derived from this skeleton are classified as steroid type saponins (Figure 8) (Xu *et al*., 2004; Vincken *et al*., 2007 and references within).

Lanosterol synthase (LS E.C. 5.4.99.7) gives rise to lanosterol after 2,3-OS cyclization. Cycloartenol synthase (CS E.C. 5.4.99.8) that converts 2,3-OS to cycloartenol was considered to be the basal plant OSC from which others derived (Xiong *et al.*, 2005). The *Arabidopsis CAS1* cDNA was isolated and cloned, facilitating the subsequent homology based identification of others OSCs. *CAS* plant genes have now been cloned and characterized from numerous eudicots, several monocots and a gymnosperm. The *in planta* roles of cycloartenol were confirmed by the characterization of *cas1* knock-out *Arabidopsis* mutants. A *cas1* knock-out mutant is a male gametophyte-lethal mutant due to the depletion of sterols, indicating that CAS1 is essential for cell viability. Analysis of weak or conditional *cas1* mutants further reinforced the implication of CAS1 in plastidial thylakoid membrane biosynthesis (Babiychuk *et al.*, 2008). On the contrary, the *Arabidopsis* lanosterol synthase knock-out mutants did not exhibit significant morphological phenotypes (Suzuki *et al.*, 2006), revealing that plants preferentially utilize cycloartenol over lanosterol, which is a precursor for the biosynthesis of membrane sterols and steroid hormones in animals and fungi.

A.3.2.2. OSCs that utilize the dammarenyl cation

Most of the widespread triterpene alcohols in plants, such as lupeol and β-amyrin originate from the dammarenyl cation. A proton-initiated cyclization of the "chair–chair–chair" conformation results in the dammarenyl carbocation, and all structures derived from this carbocation are classified as dammarane type saponins. A series of hydride and methyl shifts in the dammarenyl carbocation leads to the tirucallenyl C8 carbocation, and all structures derived from this carbocation are classified as tirucallane type saponins. The 5-membered ring next to the C20 dammarenyl carbocation can expand either by a shift of the C16-C17 bond, or by a shift of the C13-C17 bond. A shift of the C16-C17 bond leads to the tetracyclic C17 baccharenyl carbocation and can be followed by a reaction with the C24-C25 double bond to produce the pentacyclic C25 lupenyl carbocation. All compounds derived from this carbocation are classified as lupane type triterpenes. The lupenyl carbocation can be rearranged further, first to the C18 germanicenyl carbocation, and then via a series of hydride shifts to the C13 oleanyl carbocation. All structures derived from this oleanyl carbocation are classified as

oleanane type triterpenes. Oleanane type skeleton is also called to as β-amyrin skeleton and it is the most abundant type of structures in nature. A shift of the methyl group in the germanicenyl carbocation produces the C20 taraxasterenyl carbocation which can be deprotonated to generate taraxasterane type saponins. A methyl shift in the germanicenyl carbocation, followed by several hydride shifts, ultimately yields the C13 carbocation, which can be deprotonated to ursane type structures. The ursane skeleton is also referred as the α-amyrin skeleton. The α-amyrin and β-amyrin skeletons are the cyclization products of different and distinct enzymes, α-amyrin synthase (α-AS, E.C. 5.4.99) and β-amyrin synthase (β-AS, 5.4.99.B1), respectively (Haralampidis *et al*., 2002). A shift of the C13-C17 bond in the C20 dammarenyl carbocation yields to a C17 carbocation, which can be cyclized by a reaction with the double bond in the side chain to form the C25 pentacyclic hopenyl carbocation. All saponins derived from this carbocation are classified as hopane type saponins (Figure 8) (Vincken *et al*., 2007 and references wherein).

A.3.2.3. Evolution of OSC enzymes in plants

To date, more than 40 OSC enzymes have been isolated from various plant species, including α- and β-amyrin synthases, lupeol synthases and sterol synthases (Phillips *et al*., 2006; Shibuya *et al*., 2009). *A. thaliana* harbors 13 *OSC* genes, which have been well characterized in a yeast mutant lacking lanosterol synthase activity. Eight of these genes encode enzymes producing a predominant triterpenoid, such as cycloartenol, lanosterol, marneral, camelliol C, β-amyrin, arabidiol, thalianol, or tirucalladienol (Corey *et al*., 1993; Fazio *et al*., 2004; Suzuki *et al*., 2006; Xiang *et al*., 2006; Xiong *et al*., 2006; Shibuya *et al*., 2009). Thalianol and marneral synthases were found to be organized in operon-like gene clusters in *Arabidopsis* genome (Field and Osbourn, 2008; Field *et al*., 2011). The remaining five OSCs have been characterized as multifunctional enzymes involved in the formation of several triterpenoids (Herrera *et al*., 1998; Kushiro *et al*., 2000; Segura *et al*., 2000; Husselstein-Muller *et al*., 2001; Ebizuka *et al*., 2003; Kolesnikova *et al*., 2006; Lodeiro *et al*., 2007; Shibuya *et al*., 2007; Morlacchi *et al*., 2009). The potential of triterpene biosynthesis was investigated in *Oryza sativa* (*O. sativa*, rice) and a total of 12 OSC genes were discovered, including β-amyrin synthases (Inagaki *et*

al., 2011). Among them, a new enzyme, isoarborinol synthase was discovered (Xue *et al*., 2012), a cycloartenol synthase was identified, while further two have been shown to produce the triterpenes parkeol and achilleol B in the yeast strain GIL77 that lacks lanosterol activity (Ito *et al*., 2011). The *OSC* genes as well as their products reported in the literature are reviewed by Augustin *et al*, 2011.

A phylogenetic analysis was conducted, which resulted in classifying the 96 OSCs from higher plants into 10 groups (groups I-X, Xue *et al*., 2012). The analysis included 53 functionally defined OSCs, 13 OSCs of *A. thaliana* with known functions, 11 predicted OSCs of *O. sativa*, and the 12 OSCs of *Sorghum bicolor* and 7 OSCs of *Brachypodium distachyon* which have been disclosed by manual annotation of the genes present in the genome. In addition, the predicted CS from the lower plants species *Chlamydomonas reinhardtii*, *Physcomitrella patents*, *Adiantum capillus-veneris* and *Popypodiodes niponica* were added in the analysis. In dicots, OSC enzymes are grouped into CSs (I), cucurbitadienol synthases (II), LSs (VIII) and a pentacyclic triterpene synthase-like group (X). Five more groups of OSCs are defined in monocots in addition to the CS group (III). One group is defined as being of unknown function (IV), another one contains parkeol synthases (V), including the rice parkeol synthase (Ito *et al*., 2011; Xue *et al*., 2012) and the third group (VI) includes the rice isoarborinol synthase (Xue *et al*., 2012). Most OSC members from the Poaceae species belong to a pentacyclic triterpene synthase-like group (VII) and are predicted to produce various triterpene structures. Interestingly, a group of unknown function (IX) comprises of four monocot sequences and is closely related to the dicot pentacyclic triterpene synthase-like group (X) and LS group (VIII) (Figure 9, Xue *et al*., 2012). It is feasible to investigate the evolutionary history of the *OSC* genes and predict the duplication events during *OSC* gene family evolution owing to the availability of *A. thaliana*, rice, *S. bicolor* and *B. distachyon* whole-genome sequences. It is highly supported that one duplication event (D1) must have occurred 140 million years ago, before the divergence of dicots and monocots (Moore *et al*., 2007; Jiao *et al*., 2011; Xue *et al*., 2012), giving rise to two ancient OSC genes, the ancestral cycloartenol synthase (ACS) gene and the ancestral LS-like (ALSL) gene. The ancestral ACS and ALSL then provided the foundation for the

two distinct groups, D1-1 and D1-2 (Figure 9). After the divergence of monocots from dicots, the ACS gene was duplicated many times, resulting in the expansion of *OSC* genes in monocot species. Another ancient duplication event (D2) is proposed for the ALSL gene before the divergence of monocots from dicots. The original LS gene was maintained in many dicot species, while the duplicated gene is likely to have been the origin of most of the dicot triterpene synthase genes. The function of the genes within monocot group IX, closely related to the dicot LSs (VIII), remains inexplicit. The possibility that IX group contains LSs cannot be nullified. It is also likely that the original LS gene was lost in monocots and that IX group is derived from a duplicated gene. The analysis also demonstrated that the dicot triterpene synthases, including lupeol, dammarenediol and β-amyrin synthase, may have originated from the ALSL gene via three successive gene duplication events (D2, D10 and D11, Figure 9). These findings oppose to the previous data that the dicot triterpene synthases are directly derived from ACS, as previously claimed by Sawai *et al.* (2006). Interestingly, 11 out of the 13 *A. thaliana OSC* genes and 20 out of 36 Poaceae *OSC* genes are grouped into functional groups (X) and (VII) / (VI) respectively. These data point out that a major spreading of the OSC gene family has occurred after the divergence of monocots and dicots. The phylogenetic genome-wide duplication and codon substitution analyses displayed that local tandem gene duplication has contributed greatly to the expansion of the *OSC* gene family (Xue *et al*., 2012). These data are in agreement with the observation that gene families involved in the biosynthesis of secondary metabolites tend to arise by gene duplication, forming tandem clusters within the plant genome (Ober, 2005).

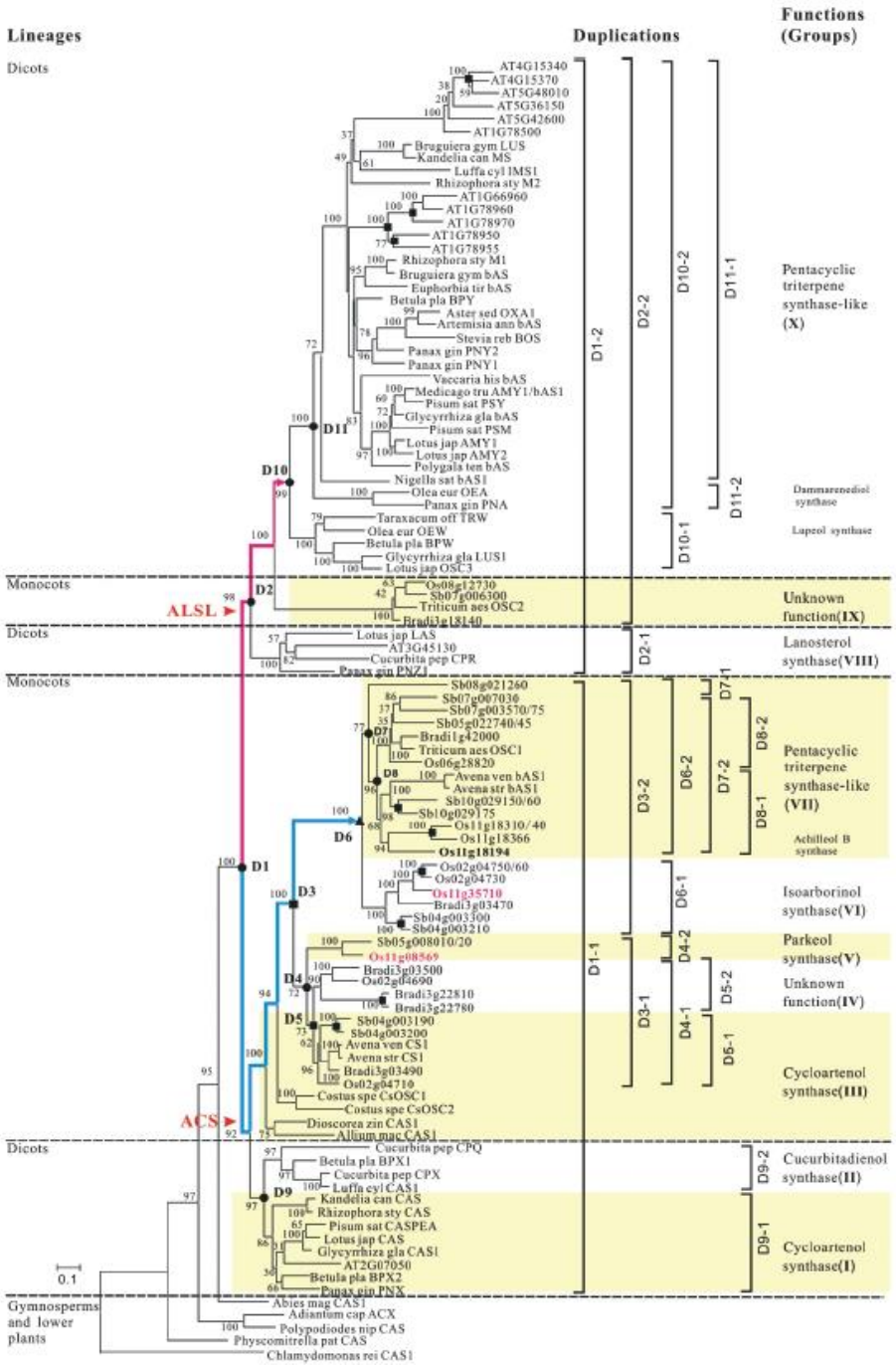


Figure 9. Phylogenetic analysis of OSCs in higher and lower plants. Broken lines separate monocots and dicots. D1–D11 display the gene duplication events. Black squares display the tandem duplications, the triangle the segmental or whole-genome duplication, and the dots unknown types of duplications. The red and blue lines show the evolutionary paths of OSCs in dicots and monocots, respectively (Xue *et al*., 2012).

A.3.3. Further elaboration of triterpene backbone structures

After 2,3-OS cyclization the triterpene backbone undergoes a series of modification steps which include oxidation, alkylation, glycosylation and acylation, resulting in accumulation of complex triterpene glycosides. Hill and Connolly (2012) have covered

all the triterpenoid structures isolated and determined in a review paper. Elaboration of the triterpene skeleton to produce the final product requires a number of steps and enzymes as cytochrome P450-dependent monooxygenases, uridine diphosphate glycosyltransferases (UGTs), and occasionally other enzymes (Haralampidis *et al*., 2002) and most of them remain uncharacterized.

A.3.3.1. Cytochrome P450 enzymes in triterpenoid biosynthesis

At present, 5100 sequences of plant cytochrome P450s (CYPs) have been annotated and named. The number of the annotated plant CYPs is significantly greater than is seen in other taxa: 1461 in vertebrates, 2137 in insects, 2960 in fungi, 1042 in bacteria, 27 in Archaea and two in viruses (Nelson, 2011). When investigating an individual plant genome, CYPs form the third largest family of plant genes. Typically the "CYPome" from angiosperms consists of around 300 genes grouping in about 50 families, a number estimated to represent up to 1% of the total gene annotations of each plant species (Nelson and Wreck-Reichhart, 2011).

P450s are hemethiolate enzymes involved in numerous biosynthetic and xenobiotic pathways found in all organisms from bacteria to humans. The reactions they catalyze are extremely diverse, but usually based on activation and heterolytic cleavage of molecular oxygen with insertion of one of its atoms into the substrate and reduction of the other to form water (Werck-Reichhart and Feyereisen, 2000; Bernhardt, 2006). They are classified as monooxygenases. Bacterial P450s are soluble proteins, but all described plant P450s are bound to membranes on the endoplasmic reticulum through a short hydrophobic segment of their N-terminus, and possibly a hydrophobic loop of the protein (Williams *et al*., 2000). In canonical P450s, heme-bound O_2 is activated by the successive transfer of two electrons from NADPH via the NADPH-cytochrome P450 reductase. In some cases, the second electron can be provided by an alternative electron transport chain involving NADH, and NADH cytochrome b_5 reductase and cytochrome b_5 (De Vetten *et al*., 1999).

The natural substrates of plant P450 enzymes include precursors of membrane sterols and of structural polymers such as lignin, cutin, suberin and sporopollenin. P450s also contribute to the homeostasis of phytohormones and signaling molecules

by controlling their biosynthesis (e.g. gibberellins, auxin, brassinosteroids, cytokinins, strigolactones and jasmonate) and catabolism (e.g. gibberellins, brassinosteroids and abscisic acid). They are involved in the biosynthesis of volatiles, antioxidants, and defense compounds, including phenolics and their conjugates, flavonoids, coumarins, lignans, glucosinolates, cyanogenic glucosides, benzoxazinones, isoprenoids, alkaloids (Schuler and Werck-Reichhart, 2003; Morant *et al*., 2003; Mizutani and Ohta, 2010). In addition to their physiological substrates, cytochromes P450s can detoxify xenobiotics such as pesticides and pollutants (Morant *et al*., 2003; Powles and Yu, 2010).

Plant P450s were initially classified in two main clades, the A-type and the non-A-type (Durst and Nelson, 1995; Paquette *et al*., 2000). The majority of P450s involved in secondary metabolic pathways seemed to be found in the A-type group. In contrast, the non-A-type included a much more divergent group of sequences consisting of several individual clades that sometimes show more local similarity to non-plant P450s than to other plant P450s and function in lipid or hormone metabolism (Paquette *et al*., 2000). Sequencing of more plant genomes changed the initial view of plant P450 phylogeny. It turned up that the "A-type" group was just the largest out of 14 P450 clans. These clans eventually were named according to their lowest-numbered family member (Nelson *et al*., 2004). The initial A-type clade became the CYP71 clan. The ten P450 clans found in vascular plants, and also represented in *A. thaliana*, form two groups. Six of them contain only a single family (CYP51, CYP74, CYP97, CYP710, CYP711 and CYP727) usually required for ancestral functions such as lipid, sterol and carotenoid metabolism and signaling (Nelson and Werck-Reichhart, 2011). The four additional clans CYP71, CYP72, CYP85 and CYP86 have evolved by intensive gene duplication and diversification with a major evolutionary burst in Angiosperms (Bak *et al*., 2011).

Several P450 enzymes, which participate in triterpenoid biosynthesis by adding the majority of functional groups at the core triterpene structure, have been isolated and characterized (Shibuya *et al*., 2006; Qi *et al*., 2006; Seki *et al*., 2008; 2011; Field and Osbourn, 2008; Field *et al*., 2011; Carelli *et al*., 2011; Fukushima *et al*., 2011; 2013; Geisler *et al*., 2013). In biosynthetic pathway of glycyrrhizin the enzyme CYP88D6 has been characterized as a β-amyrin C-11 oxidase (Seki *et al*., 2008). Seki *et al* (2011)

identified a second relevant P450 enzyme (CYP72A154) and shown to be responsible for the C-30 oxidation in this biosynthetic pathway. CYP93E1, a C-24 hydrolase has been identified as an oxidase in soyasaponin biosynthetic pathway in *Glycine max* (Shibuya *et al*., 2006). A gene encoding a member of the CYP51 family, *AsCyp51H10*, is required for the production of the avenacins in oats (Qi *et al*., 2006). AsCYP51H10 is able to catalyze both hydroxylation and epoxidation of β-amyrin to give 12,13β-epoxy-3β,16β-dihydroxy-oleanane (12,13β-epoxy-16β-hydroxy-β-amyrin) (Geisler *et al*., 2013).

In thalianol biosynthesis in *Arabidopsis* two P450 enzymes have been characterized. The first, CYP708A2, belongs to the CYP708 family (CYP85 clan) of Brassicaceae and is required for the conversion of thalianol to the downstream pathway intermediate thalian-diol. The second P450 enzyme belongs to the CYP705 family (CYP71 clan) of Brassicaceae and is responsible for the conversion of thalian-diol to desaturated thalian-diol (Field and Osbourn, 2008). Similarly, in biosynthetic pathway of marneral in *Arabidopsis* two P450 enzymes have been characterized. The marneral oxidase, CYP71A16, belongs to the CYP71 clan and is involved in the conversion of marneral to downstream pathway intermediates. The second, CYP705A12, belongs to a different family of CYP71 clan and is also implicated in marneral metabolism (Field *et al*., 2011).

A cytochrome P450 gene, *CYP716A12*, involved in the biosynthetic pathway of hemolytic saponins in *M. truncatula* was identified by combining an activation tagging method and a reverse genetic TILLING approach (Carelli *et al*., 2011). *In vitro* enzymatic activity assays indicate that CYP716A12 catalyzes the oxidation of β-amyrin and erythrodiol at the C-28 position, yielding oleanolic acid (Carelli *et al*., 2011; Fukushima *et al*., 2011). Fukuskima *et al* (2011) confirmed CYP716A12 activity *in vivo*, by expressing *CYP716A12* in transgenic yeast cells that endogenously produce β-amyrin, and identified CYP93E2 as key enzyme in non-hemolytic sapogenin biosynthetic pathway. Additionally, CYP716A12 was evaluated for its potential α-amyrin- and lupeol-oxidizing activities. It was found that CYP716A12 is able to modify α-amyrin and lupeol in order to produce ursolic acid through uvaol and, possibly, ursolic aldeyde and betulinic acid through betulin, respectively (Fukushima *et al*., 2011). CYP72A61v2 and CYP72A68v2 were also identified, because their expression is highly correlated with CYP93E2 and

CYP716A12 respectively. Yeast strains were constructed expressing bAS, CPR, CYP93E2 and CYP72A61v2, and bAS, CPR, CYP716A12 and CYP72A68v2. These transgenic yeast strains produced soyasapogenol B and gypsogenic acid, respectively. Therefore two CYP72A subfamily enzymes have been identified: CYP72A61v2, which modifies 24-OH-β-amyrin, and CYP72A68v2, which modifies oleanolic acid (Fukushima *et al.*, 2013).

Enzyme	CYP family / clan	Enzymatic activity	Product
CYP716A2	CYP716A subfamily / CYP85 clan	β-amyrin 28-oxidase (Carelli *et al.*, 2011; Fukushima *et al.*, 2011)	Oleanolic acid (Carelli *et al.*, 2011; Fukushima *et al.*, 2011)
		α-amyrin oxidase (Fukushima *et al.*, 2011)	Ursolic acid (Fukushima *et al.*, 2011)
		Lupeol oxidase (Fukushima *et al.*, 2011)	Betulinic acid (Fukushima *et al.*, 2011)
CYP93E2	CYP71 clan	β-amyrin 24-oxidase (Fukushima *et al.*, 2011)	24-OH-β-amyrin (Fukushima *et al.*, 2011)
CYP72A61v2	CYP72A subfamily / CYP72 clan	24-OH-β-amyrin oxidase (Fukushima *et al.*, 2013)	Soyasapogenol B (Fukushima *et al.*, 2013)
CYP72A68v2	CYP72A subfamily / CYP72 clan	Oleanolic acid oxidase (Fukushima *et al.*, 2013)	Gypsogenic acid (Fukushima *et al.*, 2013)

Table 1. CYPs implicated in triterpenoid biosynthesis in *M. truncatula*.

A.3.3.2. UDP-glycosyltransferases in triterpenoid biosynthesis

Glycosyltransferases (GTs) catalyze the transfer of a sugar residue of an activated sugar donor to an acceptor molecule. The GTs constitute one of the most large family of enzymes. Classification of the GTs into subfamilies has been performed according to the degree of primary sequence identity (Campbell *et al.*, 1997; Coutinho *et al.*, 2003). Members of the families of GT enzymes can be found at the CAZY database (http://cazy.org/), which includes over 90 different GT families (Osmani *et al.*, 2009). Family 1 GTs is found in all phyla (Campbell *et al.*, 1997). The plant family 1 GTs is able to glycosylate a large array of different small molecules. These include terpenoids, alkaloids, cyanogenic glucosides and glucosinolates as well as flavonoids, isoflavonoids and other phenylpropanoids. Many family 1 GTs utilize an UDP activated sugar as donor in the glycosylation reaction, and most of these belong to a group of GTs referred to as the UGTs (UDP-dependent glycosyltransferases) (Lim and Bowles, 2004; Mackenzie *et al.*, 1997). The plant UGTs are characterized by sharing a highly conserved motif denoted the PSPG motif (Plant Secondary Product Glycosyltransferase motif) that

comprises the Prosite UGT defining sequence (Hughes and Hughes, 1994; Paquette *et al*., 2003).

Glycosylation is a key modification of plant natural products during their biosynthesis, and plant secondary metabolites are often decorated with sugars. Glycosylation enhances their solubility and stability and facilitates their storage and accumulation in plant cells (Jones and Vogt, 2001; Bowels *et al*., 2005), and is also one of the major factors determining natural product bioactivity and bioavailability (Ross *et al*., 2001; Bowels *et al*., 2006). Glycosylation also plays a role in the regulation of the active levels of several hormones, including auxin, abscisic acid, cytokinins, brassinosteroids and the stress-related hormone salicylic acid (Bowels *et al*., 2006). The significance and functional implications of these modifications often remain speculative and controversial (Xu *et al*., 2002).

To date only a few UDP-GTs genes have been identified for saponin biosynthesis. The first soyasaponin UDP-GT has been purified from *G. max* and is involved in triterpenoid saponin production (Kurosawa *et al*., 2002). Soyasaponin β_g, the main soyasaponin in *G. max*, is soyasapogenol B that attaches three sugar molecules, glucuronic acid, galactose, and rhamnose, at the C-3 hydroxyl group. UGT73P2 and UGT91H4 are the responsible enzymes for attaching the second and third sugars in the sugar chain, respectively (Shibuya *et al*., 2010). Additionally Sayama *et al*., (2012) reported two novel UDP-GTs, UGT73F4 and UGT73F2, which catalyze the addition of xylose (Xyl) and glucose (Glc) respectively at the C-22 position of the group A saponins in *G. max*. The two recently characterized UDP-GT enzymes from *M. truncatula*, UGT71G1 and UGT73K1, utilize the triterpene aglycones hederagenic/medicagenic acid and soyasapogenols B and E/hederagenin respectively as acceptors for saponin biosynthesis (Achnine *et al*., 2005). An additional UDP-GT, UGT73F3, acts in the *in vivo* glucosylation at the C-28 carboxylic groups of multiple sapogenins in *M. truncatula* (Naoumkina *et al*., 2010). UGT74M1 catalyzes the glucosylation of sapogenins at the C-28 carboxylic groups in *Saponaria vaccaria* (Meesapyodsuk *et al*., 2007). Two UDP-GTs, UGT73C10 and UGT73C11, which have high catalytic activity and substrate specificity and regiospecificity for catalyzing 3-*O*-glucosulation of the sapogenins oleanolic acid

and hederagenic, have been recently identified in *Barbarea vulgaris* (Augustin *et al.*, 2012). *Sad10* gene encodes UGT74H5, which is a GT enzyme belonging to clade L of family 1 in oat. It was shown that UGT74H5 is an *N*-methylanthranilic acid *O*-glucosyltransferase implicated in avenacin A biosynthesis (Owatworakit *et al.*, 2012). The second enzyme identified in oat genome, UGT74H6, is likely required for the synthesis of other avenacins that are acylated with benzoic acid (Owatworakit *et al.*, 2012). Finally, investigations are needed in order to identify UGT candidate genes involved in glycyrrhizin biosynthesis. It can be concluded that P450 and UGT enzymes are the key factors for the continually increasing diversity of plant natural products (Figure 10).

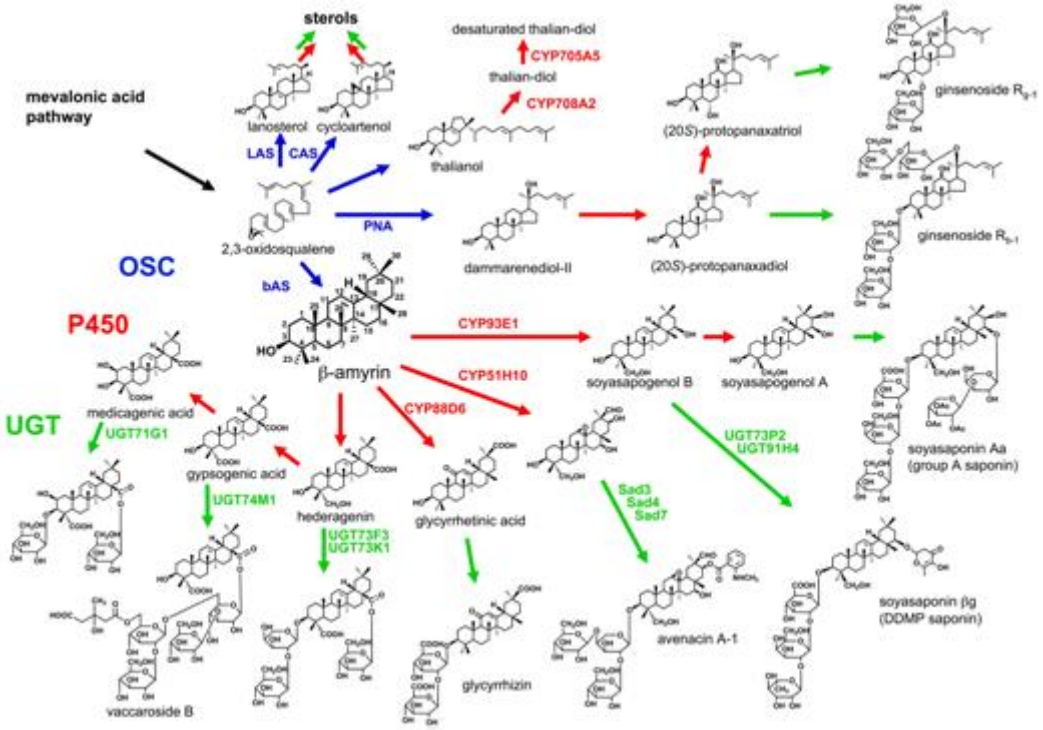


Figure 10. Triterpenoid biosynthetic pathway. After the cyclization of 2,3-OS catalyzed by OSC, a triterpenoid undergoes various modifications including P450-catalyzed oxidation and UGT-catalyzed glycosylation. Blue arrows, OSC-catalyzed steps; red arrows, P450-catalyzed steps; green arrows, additional modifications including UGT-catalyzed steps (Sawai and Saito, 2011).

A.3.3.3. Other enzymes in triterpenoid biosynthesis

In the current conception, the precursor 2,3-OS is cyclized to a limited number of core structures, which are subsequently decorated with functional groups by P450 enzymes, and finally activated with glycosyl groups added by GT enzymes (Augustin *et*

al., 2011). Occasionally, other enzymes are required in triterpene biosynthetic pathways but in most cases they remain uncharacterized (Haralampidis *et al*., 2002). In *A. thaliana* a gene encoding a BADH acyltransferase (ACT), which is implicated in secondary metabolism, has been identified. It is likely that is implicated in modification of desaturated thalian-diol in thalianol metabolism, but it has not been verified yet (Field and Osbourn, 2008).

In addition, an SCPL acyltransferase encoded by *Sad7*, SCPL1, is required for the acylation of avenacins in oat (Mugford *et al*., 2009). SCPL1 catalyzes the transfer of *N*-methyl anthranilate from an *O*-Glc ester onto triterpenoid backbones to produce avenacins A-1 and B-1. It is also required for the formation of the two benzoylated forms of avenacin, A-2 and B-2 (Mugford *et al*., 2009). Recently, *Sad9* gene encoding an anthranilate *N*-methyltransferase was characterized. This enzyme acts together with the UGT74H5 glucosyltransferase and the SCPL1 acyltransferase in the final steps of the synthesis of avenacin A-1 (Mugford *et al*., 2013).

A.4. Triterpenes and saponins in legumes

Legumes produce a huge variety of natural products, including triterpenes (Phillips *et al*., 2006). α- and β-amyrin have been shown to be structural components during normal growth and development and their synthesis is up-regulated during the establishment of symbiotic relations with rhizobia bacteria or mycorrhizal fungi (Baisted, 1971; Hernandez and Cooke, 1996; Grandmougin-Ferjani *et al*., 1999; Iturbe-Ormaetxe *et al*., 2003).

Lupeol is related with root nodulation in several plants. In *Vicia faba* nodule outer cortex, 7% of the organic-soluble material is lupeol and 82% is betulin, a lupeol metabolite (Hartmann *et al*., 2002). In *Glyccyrrhiza glabra* (*G. glabra*), lupeol synthase is highly expressed in nodules and cultured cells, the same areas in which the lupeol metabolite betulinic acid is accumulated. It is proposed that lupane-type triterpenes play important roles as hydrophobic and antibiotic barriers in the development of symbiosis in legumes (Hayashi *et al*., 2004). In *L. japonicus*, lupeol synthase (*OCS3*) is preferentially expressed in root and nodules (Iturbe-Ormaetxe *et al*., 2003; Sawai *et al*.,

2006b). A *M. truncatula* EST clone, sharing homology to other lupeol synthase genes from olive leaves and dandelion roots, is expressed during nodule formation (Gamas *et al*., 1996; Shibuya *et al*., 1999). Conclusively, the role of lupeol in root nodule biology remains to be further elucidated.

From an analysis of *Pisum sativum* nodules, Hernandez and Cooke (1996) detected β-amyrin in peribacteroid membranes and the microsomal fraction of nodule cells, which mostly corresponded to plant cell endoplasmic reticulum. β-amyrin was not detected in free living bacteria or in the plasma membrane of the roots, thus it was suggested that β-amyrin is synthesized by the host plant following plant colonization by rhizobia (Hernandez and Cooke, 1996). In *G. glabra*, the expression pattern of β-amyrin synthase was examined in different tissues and high levels were detected in root nodules, along with triterpene saponin accumulation (Hayashi *et al*., 2004). Confalonieri *et al*. (2008) demonstrated that the expression of a β-amyrin synthase gene from *Aster sedifolius*, *AsOXA1* has an effect on the symbiotic nodulation performance when introduced in *M. truncatula* using *Agrobacterium*-mediated transformation. The ectopic expression of *AsOXA1* in transgenic plants led to enhanced root nodulation. In addition, the transgenic plants showed an increased production of triterpenoid saponins in leaves and roots (Confalonieri *et al*., 2008). The role of β-amyrin in nodules remains unclear.

In legumes, β-amyrin is considered the precursor of saponins that have been identified so far. Saponins have been characterized as chemical compounds that participate in plant defense. However, some of these molecules may be involved in different processes. For example, saponins from pea have been proposed to regulate gravitropism and cellulose synthesis in plants (Ohana *et al*., 1998; Rahman *et al*., 2001). Soyasaponin I was reported to be a phytochrome killer in pea (Yokota *et al*., 1982). Chromosaponin I (CSI) is a γ-pyronyl-triterpenoid saponin isolated from pea (Tsurumi *et al*., 1991, 1992) and other leguminous plants (Kudou *et al*., 1992; Massiot *et al*., 1992) that has been shown to influence the growth of roots in several plants (Tsurumi and Wada, 1995). CSI increases the mechanical extensibility of cell walls in roots, increases the length of the cells and reduces the diameter of the root, and stimulates the root

growth (Tsurumi *et al*., 1996; Tsurumi and Ishizawa, 1997). Rahman *et al* (2001) reported that CSI specifically interacts with AUX1 protein in regulating the gravitropic response of *Arabidopsis* roots. In contrast to the soyasaponin I and chromosaponin I, nothing is known about the possible physiological function of glycyrrhizin in *G. glabra* roots.

Glycyrrhizin is an oleanane-type saponin present in a large amount (2 – 8% of the dry weight) in the underground parts of licorice and it is one of the most important crude drugs in the world (Gibson, 1978; Seki *et al*., 2008). Since the glycyrrhizin content in the thickened roots is very high, it is proposed that this chemical compound may act as a preformed chemical defense substance against pathogens as has been shown for avenacins (Hayashi *et al*., 2004). Glycyrrhizin is used worldwide as a natural sweetener and flavoring adding because of its sweet taste. Additionally, it presents various pharmacological activities, including anti-inflammatory (Matsui, 2004), immunomodulatory (Takahara *et al*., 1994), antiulcer (He *et al*., 2001), antiallergy (Park *et al*., 2004) and antiviral against various DNA and RNA viruses including HIV (Baba *et al*., 1988). The biosynthetic pathway of glycyrrhizin from β-amyrin involves hydroxylations at C-11 and C-30, and two steps of glucuronyl transfers to the hydroxyl group at C-3 (Seki *et al*., 2008; 2011).

The model legume *M. truncatula* has an interesting profile of triterpenoid saponins, from which sapogenins are differentiated into hemolytic and non-hemolytic types, according to the position of their functional groups and hemolytic properties. Soyasapogenols have a hydroxyl group at the C-24 position and are non-hemolytic sapogenins. On the contrary, hemolytic sapogenins have a hydroxyl group at C-23 and a carboxyl group at C-28 (Carelli *et al*., 2011). Saponins isolated from different species of the genus *Medicago* have been reported to possess antimicrobial, cytotoxic and insecticidal activities (Tava and Avato, 2006). Recently, their nematicidal potential has designated them as important bio-active compounds in agro-industry (Argentieri *et al*., 2008; D'Addabbo *et al*., 2009). All these triterpenic compounds are synthesized from the isoprenoid pathway via the cyclization of 2,3-oxidosqualene to the triterpene β-amyrin. The β-amyrin skeleton is then elaborated in the aglycones skeletons by cytochrome P450 monooxygenases (Tava *et al*., 2011). Five different triterpene

aglycones constitute the core of the 37 saponins identified in roots and in aerial parts, soyasapogenol B, soyasapogenol E, medicagenic acid, hederagenin and bayogenin (Huhman and Sumner, 2002). Several enzymes participating in triterpenoid saponin biosynthesis in *M. truncatula* have been biochemically characterized and are referred in the previous section. The proposed biosynthetic pathway of *M. truncatula* triterpenoids is shown in Figure 11 (Fukushima *et al.*, 2013).

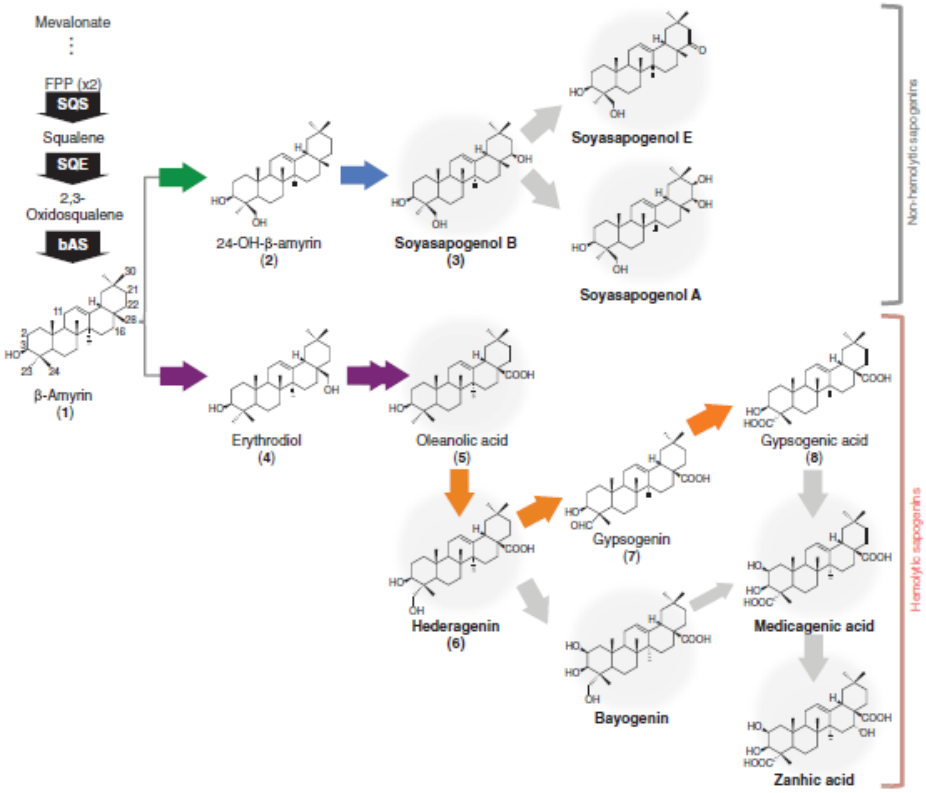


Figure 11. *M. truncatula* triterpenoids. Arrows indicate the enzymes responsible for each reaction. Colored arrows correspond to P450s described in Fukushima work (CYP716A12, purple; CYP72A68v2, orange; CYP93E2, green; CYP72A61v2, blue). Light gray arrows represent as yet unidentified enzymes. The structures in the gray bracket are non-hemolytic sapogenins and those in the pink bracket are hemolytic sapogenins (Fukushima *et al.*, 2013).

Glycine max (*G. max*) accumulates β-amyrin-derived oleanane-type triterpenoid saponins termed soyasaponins (Kitagawa *et al.*, 1982; Taniyama *et al.*, 1988; Kudou *et al.*, 1992; Kikuchi *et al.*, 1999). Soyasaponins are divided into two groups, 2,3-dihydro-2,5-dihydroxy-6-methyl-4*H*-pyran-4-one (DDMP) group and group A saponins. DDMP saponins have soyasapogenol B as their aglycone and group A saponins have

soyasapogenol A as their aglycone. Soyasapogenol B is a C-22 and C-24-hydroxylated β-amyrin and soyasapogenol A has an additional hydroxyl group at C-21. DDMP saponins and their derivatives present various beneficial activities on human health, such as prevention of dietary hypercholesterolemia (Murata *et al*., 2005), suppression of colon cancer cell proliferation (Ellington *et al*., 2005) and antiperoxidation of lipids and liver-protecting action by acceleration of secretion of thyroid hormones (Ishii and Tanizawa, 2006). On the other hand, some saponins of the group A are unfavorable because of their bitter and astringent taste (Okubo *et al*.,, 1992). Thus, it is important to manipulate saponin composition and content in order to improve soybean quality and function. Unfortunately, the responsible enzymes in soyasaponin biosynthesis are mostly unknown.

Triterpene structures have been also identified from other legume species. The triterpenoid compounds friedelin and lupeol have been isolated from *Machaerium hirtum* leaves and branches. Anti-inflammatory activities of plant extracts and purified substances were evaluated using the mouse ear edema model. In addition, the triterpenoid friedelin displayed activity against cancer cells (Ignoato *et al*., 2012). Triterpenoid structures like lupeol, friedelin, α-amyrin and betulinic acid have been isolated from species of the genus *Cassia* (Sob *et al*., 2010 and references within). Additionally, a novel oleanane type triterpenoid glycoside has been isolated from butanolic seed extracts of *Cassia angustifolia*. This extracted saponin was tested for its antifungal activity against human and plant pathogenic fungi and the maximum inhibition was found against *Colletotrichium dematium* (Khan and Srivastava, 2009). A number of triterpene derivatives have been isolated and structurally determined from species of the genus *Lupinus* (Woldemichael *et al*., 2003). The hepatoprotective effects of saponins in lupins have been reported (Kinjo *et al*., 1999). In addition, three saponin compounds in lupin seeds indicated moderate antifungal activity against *Candida albicans* and soyasaponin I was found weakly hemolytic (Woldemichael and Wink, 2002). Two new triterpenoid saponins, namely caraganoside C and caraganoside D were isolated from the seeds of *Caragana microphylla* and structurally elucidated. Both saponins exhibited moderate inhibitory activity against NO production in LPS-

stimulated RAW264.7 cells. In addition, caraganoside C showed weak cytotoxicity against MCF-7, HL-60, HCT116, and A549 cell lines (Jin *et al*., 2010).

A.5. Characterization, functional analysis and gene organization of OSCs of *L. japonicus* and *M. truncatula*.

Three triterpene synthases have been cloned and characterized from the model legume species *L. japonicus* and *M. truncatula* (Iturbe-Ormaetxe *et al*., 2003). *MtAMY1* is predicted to encode a β-amyrin synthase from *M. truncatula* and *LjAMY2* is a multifunctional OSC from *L. japonicus*. A partial cDNA predicted to encode a β-amyrin synthase was identified in *L. japonicus* (LjAMY1, Iturbe-Ormaetxe *et al*., 2003), which was further investigated by Sawai *et al*., 2006. MtAMY1 shares 91 and 95% amino acid identity with LjAMY1 and LjAMY2, respectively, and the two *L. japonicus* clones share 94% identity. MtAMY1, LjAMY1 and LjAMY2 all contain the DCTAE motif (Abe and Prestwich, 1994) and the characteristic *β*-strand turn QW motifs of the OSC family (Poralla *et al*., 1994), suggesting that these enzymes are likely to be β-amyrin synthases (Haralampidis *et al*., 2001; Iturbe-Ormaexte *et al*., 2003). *MtAMY1* and *LjAMY2* were expressed in yeast, in order to identify their metabolite potential. MtAMY1 generated a product identical to that of β-amyrin, while LjAMY2 accumulated almost equal amounts of lupeol and β-amyrin and a number of other minor products. Northern blot analysis revealed signals only in roots for LjAMY1 and LjAMY2, while MtAMY1 was expressed in all tissues analyzed, with highest levels in shoot meristem and stem tissue (Iturbe-Ormaexte *et al*., 2003).

Triterpenoid saponins derived from the β-amyrin skeleton are present in the roots of *M. truncatula* (Huhman and Sumner, 2002). The saponin content of *M. truncatula* leaves has not yet been characterized extensively, but related species such as *Medicago sativa* does contain *β*-amyrin-derived saponins in the foliage (Oleszek *et al*., 1992). Therefore, the expression of the *β*-amyrin synthase gene *MtAMY1* in the roots and leaves of *M. truncatula* attributes a role for MtAMY1 in the synthesis of triterpenoid saponins in these tissues. More detailed analysis of the saponin content *L. japonicus*

tissues and nodules is required in order to evaluate the role of LjAMY1 and LjAMY2 (Iturbe-Ormaexte *et al*., 2003).

The *OSC* genes of *L. japonicus* were identified and functionally elucidated. Eight *OSC* genes (*OSC1 – OSC8*) were analyzed (Sawai *et al*., 2006) and, *OSC1* and *OSC8* correspond to the previously reported *LjAMY1* and *LjAMY2* (Iturbe-Ormaexte *et al*., 2003). The *OSC* cDNAs were expressed in yeast. The results displayed that the yeast cells expressing *OSC1, OSC3* and *OSC5* accumulated β-amyrin, lupeol and cycloartenol respectively (Sawai *et al*., 2006). Eight *OSC* genes (*OSC1–OSC8*) form two sets of clusters in chromosomes 2 and 3 of *L. japonicus* (Figure 12), which suggests that they have been formed by repeated local gene duplication followed by accumulated nucleotide substitution, insertion and deletion. Sawai *et al*., obtained several genomic clones, LjT03N08, LjT07E09, LjT43B09, LjB16L08 and LjT11L01, using the primer sets for *OSC1*, *OSC2*, *OSC3*, *D9FR* and *LjAMY2*, respectively. Sequence analysis of these clones showed that LjB16L08, LjT07E09 and LjT03N08 constitute a contig in this order and that *OSC1*, *OSC2* and *OSC4* are located within a 160 kb region. *OSC2* was unveiled to be 1 kb shorter than other functional *OSC* genes, lacking seven exons; therefore it was thought to be a pseudogene. This LjT07E09 contig is localized on chromosome 3. Parts of the nucleotide sequences of LjB16L08 remain unclear, and the complete coding sequence of *OSC4* is not available. LjT43B09 is located on chromosome 2 and is found to contain *OSC3*, *OSC5* and a part of an *OSC*-like gene, determined as *OSC6*. LjT43B09 overlaps another clone LjT16A21, which contains the *OSC6* gene and another *OSC*-like gene, defined as *OSC7* gene. The *OSC8* gene shares a coding sequence highly similar to *LjAMY2* (99%) and is localized in LjT11L01 clone, which is mapped at 0.8 cM from LjT07E09 on chromosome 3 (Figure 12, Sawai *et al*., 2006). It is proposed that *OSC3* gene in chromosome 2 is derived from an ancestor *OSC5* gene or, *OCS6* and *OSC7* genes. Additionally, *OSC1*, *OSC2*, *OSC4* and *OSC8* share a common ancestor, which would have been derived from an ancestor of *OSC3* and generated the *OSC* group in chromosome 3 (Sawai *et al*., 2006).

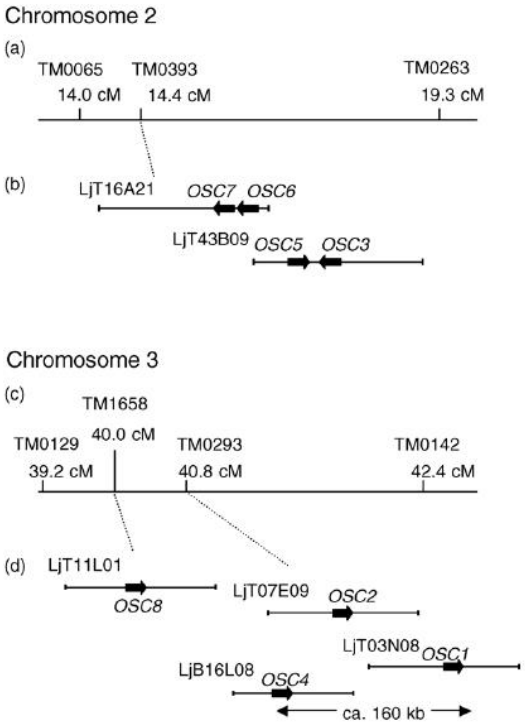


Figure 12. Organization of *OSC* genes in *L. japonicus* genome (Sawai *et al.*, 2006).

A.6. Gene clusters for secondary metabolic pathways in plant kingdom

A.6.1. Gene clustering is a widespread phenomenon among prokaryotes and eukaryotes

Clusters containing genes that do not share homology but are functionally related are common features in bacterial genomes, known as operons. The term operon was established by Jacob and Monod, who characterized the first defined classical operon, the *lac* operon in *Escherichia coli*. Bacteria produce a vast array of secondary metabolites with interesting biological properties, such as antibiotics, anticancer agents, insecticides, immunosuppressants and herbicides (Osbourn, 2010a). Clustering of the genes implicated in biosynthesis of these compounds has been recognized. For example, members of the genus *Streptomyces*, belonging in the group of Gram-positive filamentous bacteria, were found to contain more than 20 - 30 such clusters (Bentley *et al.*, 2002; Ikeda *et al.*, 2003).

Functional gene clusters have been also described in eukaryotes (Hurst *et al*., 2004; Osbourn and Field, 2009). Most of them consist of paralogs that presumably have evolved by tandem gene duplication and divergence. For example, the mammalian *Hox* locus is required for embryonic patterning and β-globin cluster is necessary for haemoglobin production (Osbourn and Field, 2009). However, in eukaryotic genomes clusters of functionally related but non-homologous genes have also been qualified. The best known example is the major histocompatibility complex (MHC) in animals, which protects against pathogens in innate and adaptive immunity (Horton *et al*., 2004). By far the most abundant clusters unveiled are involved in the biosynthesis of secondary metabolites in filamentous fungi. These include clusters for the synthesis of important pharmaceutical compounds, such as the β-lactam antibiotics penicillin and cephalosporin, and for the production of toxins, such as aflatoxin and host-selective toxins related to virulence on plants (Hoffmeister and Keller, 2007; Turgeon and Bushley, 2010). In addition, the *DAL* and *GAL* gene clusters in the yeast *Saccharomyces cerevisiae* allow the proper use of allantoin and galactose respectively (Hittinger *et al*., 2004; Wong and Wolfe, 2005). Gene clusters of this type are regarded as adaptive gene clusters because they are associated with growth and survival under certain environmental conditions. Particularly, clusters are the means for the exploitation of new environments or the management of interactions with other organisms (Field and Osbourn, 2010).

A.6.2. "Operon-like" gene clusters in plants

Until about 15 years ago, it was considered that gene clusters in plants were restricted to tandem duplicates, such as the arrays of Leucine-rich repeat genes involved in disease resistance. The existence of gene clusters necessary for the biosynthesis of secondary metabolites has emerged as a novel and increasing theme in plant biology (Osbourn 2010a, b; Chu *et al*., 2011). Clustering of genes for plant metabolic pathways is not the rule. The genes implicated in the well-characterized anthocyanin, phenylpropanoid and glucosinolate pathways are unlinked. However, several gene clusters for secondary metabolic pathways have been reported in plants, such as for the synthesis of cyclic hydroxamic acids in maize (Frey *et al*., 1997), the

triterpene avenacin production in oat (Qi *et al*., 2004; 2006; Mylona *et al*., 2008; Mugford *et al*., 2009; Chu *et al*., 2011; Mugford *et al*., 2013), the synthesis of two different diterpenes in rice (momilactones and phytocassanes) (Wilderman *et al*., 2004; Shimura *et al*., 2007; Swaminathan *et al*., 2009; Wang *et al*., 2011) and the biosynthesis of cyanogenic glucosides in *L. japonicus* (Takos *et al*., 2011). It seems that this number will increase as the volume of available genome sequences continues to grow. These "operon-like" clusters are composed of non-homologous genes that are physically linked, co-regulated and responsible for a common function (Osbourn and Field, 2009; Field and Osbourn, 2010). A critical difference between clusters and operons is that, as far as we know, the genes are all transcribed as separate mRNAs rather than as a single mRNA in bacteria.

Secondary metabolic gene clusters contain genes encoding enzymes that catalyze the synthesis of the skeleton structures of the different classes of secondary metabolites. These genes are referred as "signature" genes and are considered as the founders of secondary metabolic clusters (Osbourn, 2010a). The signature genes all share homology with genes from plant primary metabolism, thus they have presumably recruited directly or indirectly from primary metabolism by gene duplication, neofunctionalization and genome reorganization (Gierl and Frey, 2001; Qi *et al*., 2004, 2006; Field and Osbourn, 2008; Osbourn and Field, 2009; Swaminathan *et al*., 2009). Alternatively, it is likely that the primary metabolic genes and their signature counterparts evolved from a common ancestor (Chu *et al*., 2011). In addition, clusters include genes encoding the proper "tailoring" enzymes that elaborate the secondary metabolite skeleton, such as oxidoreductases, methyltransferases, acyltransferases and glycosyltransferases. These enzymes are responsible for the successive conversion of the products of the signature enzymes into the respective pathway end products (Osbourn, 2010a).

A.6.2.1. The significance of clustering

The arising question is why genes participating in some biosynthetic pathways are clustered, while others are not. Co-inheritance and co-regulation are considered the primary benefits conferred by clustering of genes (Osbourn 2010a, b). Clustering of

functionally related genes will facilitate the co-inheritance of favorable combinations of gene alleles placed in these multigene loci. The final products of most of the clusters described are involved in plant defense, thus the formation and maintenance of the most effective gene set is expected to be necessary and beneficial and confers a selective advantage in nature. When the fitness of an allele at one locus depends on the genotype at another locus, this may lead to selection for genomic rearrangements that limit the distance between the pathway genes (Nei, 1967; 2003). Such gene organization in clusters ensures that the end protective product of the biosynthetic pathway is produced in the right place and the right time. Similarly, clustering of genes of a biosynthetic pathway with toxic/bioactive reaction intermediates would reduce the likelihood of its disruption by segregation (Osbourn 2010b; Takos *et al*., 2011). Interestingly, it have been observed that accumulation of triterpene pathway intermediates in oat and *A. thaliana* can result in severe effects on growth and development (Field and Osbourn, 2008; Mylona *et al*., 2008; Osbourn, 2010b). Thus, disruption of these pathways can have detrimental effects in addition to compromised disease resistance, and this may further enhance selection for clustering (Chu *et al*., 2011).

A further advantage of gene clustering is that it may provide various other mechanisms for co-ordinated transcriptional regulation, including shared long-distance regulatory elements, or regulation through chromatin structure or nuclear organization (Hurst *et al*., 2004; Sproul *et al*., 2005; Janga *et al*., 2008; Osbourn and Field, 2009; Osbourn, 2010a,b). Organization of genes into clusters may also facilitate the co-ordinated manipulation of transcripts of the physically linked genes, from transcription through processing and export to protein synthesis. In yeast, filamentous fungi and animals it is clear that expression of cluster genes is associated with various histone modifications, and is mediated by chromatin remodeling complexes (reviewed in Osbourn and Field, 2009). Non-coding RNAs have also been implicated in the recruitment of chromatin complexes, and in animals *Hox* gene expression can be controlled post-transcriptionally and epigenetically, by non-coding RNAs and Polycomb group proteins (Yekta *et al*., 2008; Fraser *et al*., 2009). In filamentous fungi,

regulation of secondary metabolic clusters at chromatin level is well demonstrated, and the use of drugs and mutations that affect chromatin remodeling has been established as a powerful means of pathway activation and identification (Cichewicz, 2010). Thalianol and avenacin clusters, which are developmentally regulated, provide more proofs towards implication of chromatin remodeling in the regulation of plant secondary metabolic clusters. The thalianol cluster has marked histone H3 lysine 27 (H3K27) trimethylation, a fact that insinuates the coordinated regulation at the chromatin level (Field and Osbourn, 2008). In addition, high-resolution DNA fluorescence in situ hybridization (FISH) experiments in oat have revealed that expression of the avenacin cluster is associated with chromatin decondensation (Wegel *et al*., 2009).

A.6.2.2. Description of the secondary metabolic clusters characterized in plants

The Benzoxazinone gene cluster in maize: The first cluster involved in plant secondary metabolism was described for the synthesis of cyclic hydroxamic acids in maize (Frey *et al*., 1997). In the Graminae, the cyclic hydroxamic acids 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) and 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) are the predominant forms of the well-studied class of the secondary metabolites Benzoxazinones (Bxs). They confer resistance to several pathogens and insects, like northern corn leaf blight (*Helminthosporium turcicum*), maize plant louse (*Rhophalosiphum maydis*) and stalk rot (*Diplodia maydis*). In maize DIBOA is synthesized by a series of five distinct genes. The first gene in the pathway, *Bx1*, encodes the enzyme that catalyses the conversion of indole-3-glycerolphosphate to indole (Frey *et al*., 1997). Bx1 is a homologue of the Trp synthase α-subunit (TSA) in tryptophan biosynthesis. Bx1 and TSA are both characterized as indole-3-glycerolphosphate lyases (IGLs) encoded by paralogous enzymes. In the following reactions four oxygen atoms are introduced into the indole moiety by four cytochrome P450 monooxygenases, termed Bx2 - Bx5, which belong to the CYP71 clan. Bx2 catalyses the conversion of indole to indolin-2-one, which is transformed to 3-hydroxy-indolin-2-one by Bx3. Bx4 is responsible for the unusual expansion of 3-hydroxy-indolin-2-one to 2-hydroxy-2-1,4-benzoxazin-3-one (HBOA, Spiteller *et al*., 2001). Bx5

enzyme catalyses the *N*-hydroxylation of HBOA to DIBOA. The resulting DIBOA is glucosylated by the UDP-glucosyltransferases Bx8 and Bx9 (von Rad *et al*., 2001). The conversion of DIBOA to DIMBOA requires hydroxylation and methylation at C-7 of DIBOA. Bx6 accepts DIBOA-glc as substrate producing the hydroxy-derivative 2,4,7-trihydroxy-2*H*-1,4-benzoxazin-3-(4*H*)-one-glc (TRIBOA-glc) (Frey *et al*., 2003). Bx7 is characterized as an *O*-methyltransferase (OMT) and catalyses the formation of DIMBOA-glc from TRIBOA-glc. It is proposed that glucosylation may reduce phytotoxicity and so be important for storage (von Rad *et al*., 2001). DIMBOA biosynthetic genes (*Bx1 – Bx8*) are clustered within 6 cM on the short arm of chromosome 4 in maize. The *OMT Bx7* gene is more loosely associated, but also localized on the short arm of the same chromosome. The second glucosyltransferase gene, *Bx9*, is located on chromosome 1 (Frey *et al*., 2009).

A gene cluster for avenacin metabolism in oat: As mentioned in previous section, avenacins are protective saponins that accumulate in the root tips of oat (Papadopoulou *et al*., 1999). Avenacins are produced from the already described isoprenoid pathway. From genetic analysis, several loci were identified for avenacin biosynthesis (*Sad1* to *Sad10*). *Sad1*, *Sad2*, *Sad3* and *Sad4*, *Sad7* and *Sad10* encode for β-amyrin synthase (Haralampidis *et al*., 2001), a cytochrome P450 monooxygenase (Qi *et al*., 2006), glucosyltransferases (Mylona *et al*., 2008), an SCPL acyltransferase (Mugford *et al*., 2009) and a family 1 glycosyltransferase (UGT74H5, Owatworakit *et al*., 2012). Interestingly, *Sad1* and *Sad2*, along with five of the six other loci identified (*Sad3*, *Sad10* and *Sad5* to *Sad8*) are clustered in the oat genome. *Sad2*, *Sad6*, *Sad7* and *Sad8* were 0 cM from the *Sad1* locus, with the *Sad3* locus was less closely linked at 3,6 cM from the *Sad1* locus. *Sad2* is located 69,5 kb upstream of the *Sad1* gene, *Sad7* is localized 61,6 kb downstream of the *Sad1* gene and *Sad10* is adjacent to *Sad7* gene. A gene encoding an anthranilate N-methyltransferase (MT1, encoded by *Sad9*) was recently identified and it was shown that this enzyme acts together with the UGT74H5 glucosyltransferase and the SCPL1 acyltransferase in the final steps of the synthesis of avenacin A-1. These three genes are adjacent within the wider avenacin biosynthetic gene cluster and so represent an acylation module (Mugrofd *et al*., 2013).

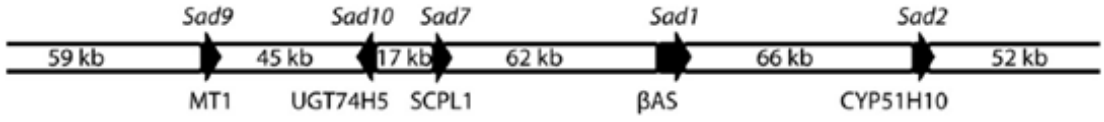


Figure 13. A BAC contig showing the avenacin gene cluster (Mugford *et al*., 2013).

Expression of the avenacin cluster genes is tightly regulated and is restricted to the epidermal cells of the root tip, the site of avenacin accumulation (Osbourn *et al*., 1994; Haralampidis *et al*., 2001; Mugford *et al*., 2009). The avenacin biosynthetic pathway has evolved from gene components of the primary metabolism, by gene duplication and neofunctionalization, and not by horizontal gene transfer (Haralampidis *et al*., 2001; Qi *et al*., 2004, 2006). Therefore, this cluster is an example of a rapidly evolving co-adapted gene complex (Mylona *et al*., 2008). It is possible that the avenacin cluster has evolved because the end product of the pathway confers an advantage (pest and pathogen resistance). *Sad1* mutants have normal root morphology, but mutations in *Sad3* and *Sad4* loci resulted in stunted root growth, membrane trafficking defects in the root epidermis and root hair deficiency. Regarding these phenotypes it can be concluded that avenacin gene cluster is likely to hamper the accumulation of deleterious intermediates (Mylona *et al*., 2008).

***The thalianol and marneral clusters in A. thaliana*:** As mentioned above, the genome of *Arabidopsis* contains 13 predicted *OSC* genes, which have been functionally characterized (Molracchi *et al*., 2009). The genome regions around these genes were investigated. Four *OSC* genes of them are flanked by genes predicted to encode other classes of enzymes, which participate in secondary metabolic pathways. Particularly, the investigations were focused on two genomic regions of *Arabidopsis*. The first region contains four contiguous genes predicted to encode an OSC, two cytochrome P450 enzymes and a BAHD acyltransferase (ACT) (Field and Osbourn, 2008). Microarray expression analysis of these genes revealed that the expression of all four genes is highly correlated and primarily detected in root tissues. Thus, the genes are functionally related (Eisen *et al*., 1998). The *OSC* gene encodes thalianol synthase (THAS), an enzyme that previously had been shown to convert 2,3-OS to thalianol when expressed

in yeast (Fazio *et al*., 2004). The first CYP450 encodes the enzyme CYP708A2, which belongs to the functionally uncharacterized CYP708 family of the Brassicaceae. CYP708A2 is required for the conversion of thalianol to the down-stream pathway intermediate thalian-diol. The second cytochrome P450 enzyme belongs to the CYP705 family, another uncharacterized CYP450 Brassicaceae-specific clade. The CYP705 and CYP708 families belong to different clans and especially to CYP71 and CYP85 respectively. Therefore, CYP705 is not a tandem duplicate of CYP708 and is functionally characterized as thalianol hydroxylase (THAH) (Field and Osbourn, 2008). The second CYP450 is responsible for the conversion of thalian-diol to desaturated thalian-diol, by introducing a double bond to carbon 15, thus is functionally characterized as thalian-diol desaturase (THAD). The fourth cluster gene encodes a BAHD acyltransferase, also belonging to an enzyme family specific to the Brassicaceae. It is likely that BAHD is involved in further modifications of desaturated thalian-diol (Field and Osbourn, 2008). *Arabidopsis* plants overexpressing thalianol or thalian-diol have pronounced dwarf phenotypes, suggesting that there is a need for high co-ordinate regulation of the pathway.

The second genomic region of *Arabidopsis* examined contains an *OSC* gene which is flanked by two cytochrome P450 genes that belong to different P450 families (Field *et al*., 2011). These three genes are specifically expressed in root tissue, in a pattern similar to the thalianol cluster. The *OSC* gene encodes marneral synthase (MRN1) which converts 2,3-OS to marneral when expressed in yeast. The first co-expressed gene is *CYP71A16*, predicted to encode a cytochrome P450 enzyme belonging to the greatly expanded Brassicaceae clan CYP71. CYP71A16 is involved in the conversion of marneral to downstream pathway intermediates. CYP71A16 is primarily characterized as marneral oxidase, possibly generating multiple isomers (MRO). The second co-expressed gene, immediately adjacent to *MRO*, is the cytochrome P450 gene *CYP705A12*. The predicted product of *CYP705A12* belongs to the same cytochrome P450 family as THAD in the thalianol cluster (Field *et al*., 2011). Similarly to THAS accumulation, overexpression of *MRN1* or *MRN1* and *MRO* together resulted in a dwarf phenotype. The role of MRN1 was further investigated in *Arabidopsis mrn1* knock-out

mutant displaying round-shaped leaves, late flowering, short anther filaments, root growth retardation and delayed embryo development (Go *et al*., 2012). It is proposed that the thalianol and marneral clusters have been founded by duplication of an ancestral *OSC/CYP705* gene pair. However, the evolutionary events to drive the establishment of the present-day clusters are independent. Both thalianol and marneral clusters are positioned in dynamic chromosomal regions, considerably abundant in transposable elements (TEs) (Field *et al*., 2011; Field and Osbourn, 2012). It is reported that TEs can launch ectopic recombination and gene trans-duplication. These findings are fully consistent with the already mentioned gene clusters in oat and maize. The avenacin and Benzoxazinone clusters are both located in subtelomeric regions, in which chromosomal recombinations and segmental duplications (SD) are observed. It is likely that dynamic chromatin domains accelerate the sampling of different gene combinations and the functional optimization of gene clusters, and facilitate coordinate regulation of gene expression (Field *et al*., 2011).

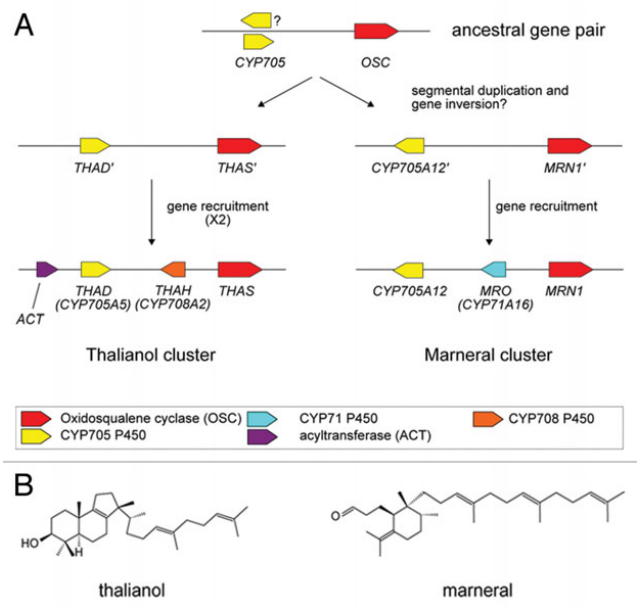


Figure 14. (A) Proposed scheme for formation of the thalianol and marneral clusters, based on the assumption that the two clusters were founded by duplication of an ancestral *OSC/CYP705* gene pair. (B) Structures of thalianol and marneral (Field *et al*., 2011; Field and Osbourn, 2012).

Biosynthetic gene clusters for labdane-related diterpenoids in rice: Rice produces a complex mixture of diterpenoid phytoalexins in response to fungal attack

(*Mangaporthe grisae*), bacterial infection (*Xanthomonas campestris*) (Peters, 2006; Toyomasu, 2008) or under exposure to UV irradiation (Tamogami *et al*., 1993; Koga *et al*., 1995). These compounds are grouped into four structurally distinct types based on the structures of their diterpene hydrocarbon precursors: phytocassanes A to E, oryzalexins A to F, momilactone A and B, and oryzalexin S, the founding members of which are the gibberellin phytormones. Biosynthesis of rice phytoalexins is initiated by sequential cyclization steps catalyzed by specific diterpene cyclases (Peters, 2006; Toyomasu, 2008). Firstly, the acyclic universal diterpene precursor GGPP is cyclized by class II diterpene cyclases, typically a labdadienyl copalyl diphosphate (CPP) synthase (CPS). Secondly, further cyclization of this bicyclic intermediate occurs, by stereospecific class I diterpene synthases, also termed kaurene synthase-like (KSL) because of their relationship to the ancestral enzyme involved in gibberellin metabolism. Bioactive natural products are produced when oxygen atoms are inserted into the resulting diterpene molecules by heme-thiolate cytochrome P450 monooxygenases, as well as subsequently acting short chain alcohol dehydrogenases/reductases. Rice contains four CPS, ten KSL, several biochemically characterized CYPs and many functionally unidentified CYPs involved in phytoalexin biosynthesis (Swaminathan *et al*., 2009). Two gene clusters are located in rice genome, containing genes encoding CPS, KSL and CYP (Sakamoto *et al*., 2004). The gene cluster located on chromosome 4 is involved in momilactone biosynthesis. It contains the relevant sequentially acting *syn*-CPP synthase OsCPS4 and *syn*-pimaradiene synthase OsKSL4, along with two CYPs (CYP99A2 and CYP99A3) and s SDR. SDR is shown to catalyze the final step in momilactone A production. CYP99A3 as a diterpene oxidase reacts with *syn*-pimaradiene to transform C19 from a methyl to carboxylic acid, forming *syn*-pimaradien-19-oic acid, which is presumably an early intermediate in momilactone biosynthesis. In addition, CYP99A3 carries out the same series of reactions to form a C19 carboxylic acid with *syn*-stemod-13(17)-ene, providing evidence regarding the currently unknown metabolic fate of this diterpene in rice (Wang *et al*., 2011).

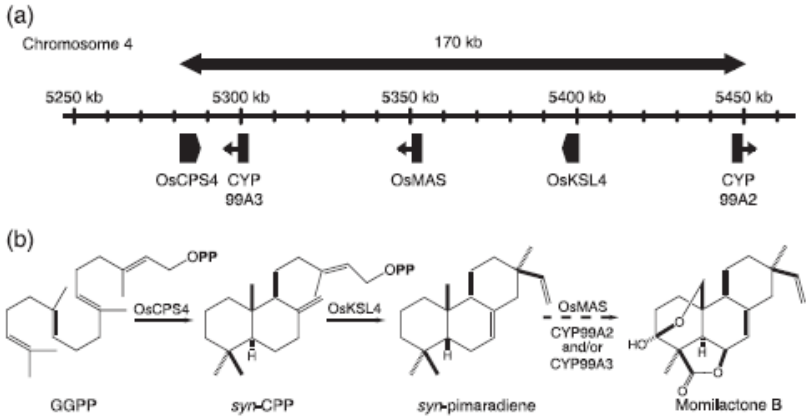


Figure 15. (a) Biosynthetic gene cluster on chromosome 4. (b) Momilactone B biosynthesis in rice (Wang *et al.*, 2011).

The second biosynthetic gene cluster is localized on chromosome 2 and contains genes that encode enzymes implicated in multiple biosynthetic pathways. Specifically, this cluster contains OsCPS2, which produces the common *ent*-CPP precursor (Prisic *et al.*, 2004; Otomo *et al.*, 2004), as well as the subsequently acting OsKSL5-7, which produce *ent*-pimaradiene, *ent*-isokaurene, and *ent*cassadiene, respectively (Cho *et al.*, 2004; Xu *et al.*, 2004). OsKSL6 and OsKSL7 then catalyze the committed step in oryzalide and phytocassane biosynthesis, respectively. In addition, this region further contains six CYP – CYP76M5-8 and CYP71Z6 and CYP71Z7 (Swaminathan *et al.*, 2009; Wu *et al.*, 2011). This cluster is quite different from other operon-like plant clusters, which contain genes dedicated to a single biosynthetic pathway as described in the previous cases.

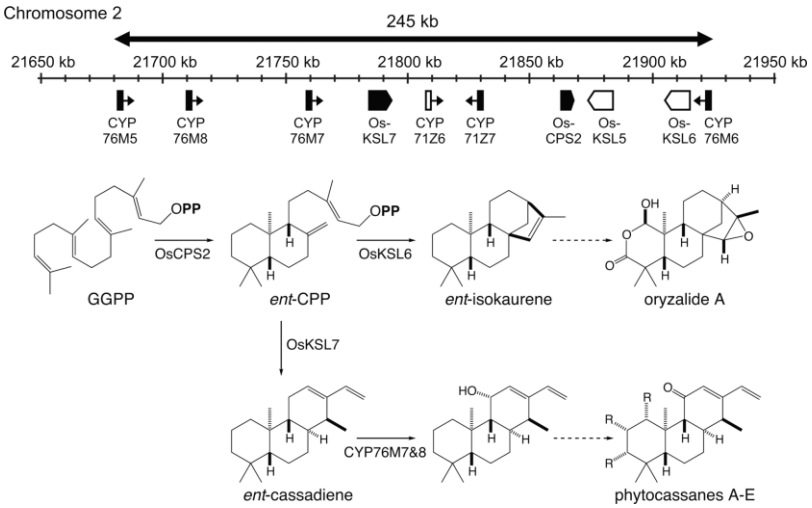


Figure 16. Rice chromosome 2 labdane-related diterpenoid gene cluster and associated biosynthetic pathways (Wu *et al.*, 2011).

The gene cluster for cyanogenic glucosides biosynthesis: Gene clusters involved in cyanogenic glucosides biosynthesis have been reported in *L. japonicus*, *Manihot esculent* (cassava) and *Sorghum bicolor* (sorghum) (Takos *et al*., 2011). Cyanogenesis is regarded as an ancient defense mechanism in plants. Upon tissue disruption, specific enzymes known as β-glucosides are responsible for the degradation of cyanogenic glucosides and the release of hydrogen cyanine, which participates in the plant defense against generalist herbivores (Morant *et al*., 2008). In all plants examined, the conversion of an amino acid into an oxime in the first committed step requires a cytochrome P450 enzyme belonging to the CYP79 family (Bak *et al*., 2006). In sorghum, the enzyme CYP79A1 catalyzes the conversion of tyrosine into an oxime intermediate in the first committed step. This oxime is converted into a hydroxynitrile by a second cytochrome P450, CYP71E1, and finally is stabilized by glucosylation via the UDP-glucosyltransferase UGT85B1, resulting in dhurrin formation. In cassava, biosynthesis of the cyanogenic glucosides linamarin and lotaustralin from the amino acids valine and isoleucin respectively, requires the activity of CYP79D1 and CYP79D2. Functional homologues of the sorghum CYP71E1 and UGT85B1 have been recently reported in cassava. In *L. japonicus*, *CYP79D3* and its paralogue *CYP79D4* encode the enzymes which catalyze the conversion of valine and isoleucine to the corresponding oximes. *CYP79D3* is separated from *CYP79D4* by a gap of 240 kb in CM0241 contig. The genomic region containing *CYP79D3* and *CYP79D4* also contains a UDP-glucosyltransferase (*UGT85*) and the *CYP736A2* gene, which completes the functional pathway for cyanogenic glucosides biosynthesis. In addition, the genetic locus *Rho* involved in natural variation in rhodiocyanoside biosynthesis is localized in the same genomic region (Takos *et al*., 2011). Clustering of genes participating in biosynthesis of cyanogenic glucosides is proposed in *L. japonicus*, cassava and sorghum. In all three species the genomic regions differ in the size, gene density, identity of additional genes and presence of transposon sequences. These findings suggest that cyanogenic glucoside biosynthetic pathway has evolved independently multiple times. Additionally, it is proposed that gene assembly in this cluster is a selective advantage because of the toxicity of the pathway intermediates (Takos *et al*., 2011).

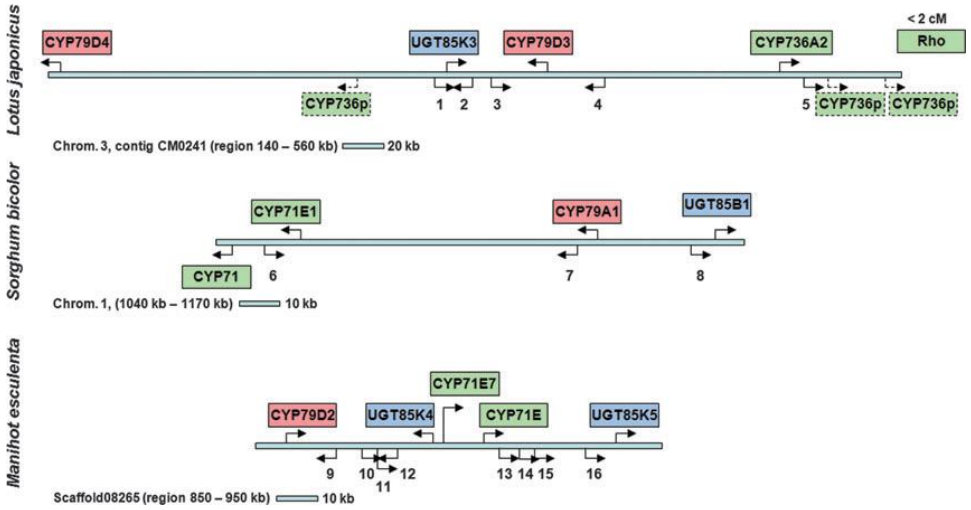


Figure 17. The cyanogenic glucoside cluster in *L. japonicus, S. bicolor* and *M. esculenta* genomes. Functional genes are presented by arrows indicating their orientation. Confirmed genes biosynthesis are labeled above each bar, with *CYP79* genes in pink, *CYP71E* and *CYP736* genes in green, and *UGT85* genes in blue. Pseudo-genes are indicated below the *L. japonicus* and *S. bicolor* bars (Takos *et al*., 2011).

A gene cluster for the synthesis of the alkaloid noscapine in Papaver somniferum: Transcriptome analysis of the high noscapine-producing poppy variety HN1 identified the exclusive expression of 10 genes encoding five distinct enzyme classes involved in noscapine biosynthesis. The corresponding genes as three O-methyltransferases (*PSMT1, PSMT2* and *PSMT3*), four cytochrome P450s (*CYP82X1*, *CYP82X2, CYP82Y1* and *CYP719A21*), an acetyltransferase (*PSAT1*), a carboxylesterase (*PSCXE1*) and a short chain dehydrogenase/reductase (*PSSDR1*) were putatively identified. These genes are not presented in EST libraries from the high morphine variety HM1 and high thebaine variety HT1. The organization of these 10 genes in a cluster was confirmed (Figure 18, Winzer *et al*., 2012). According to the homology and intron - exon structure of *CYP82* and *PSMY* genes, it is concluded that they have generated by tandem duplication after genome reorganization of the progenitor genes. A similar hypothesis can be made for *PSCX1* gene and its homologous, *PSCX2*, in the flanking area of the cluster. *CYP82Y1* gene lies 45 kb from *CYP82X2*, and *PSAT1* and *PSMT2* are located between them. *PSMT3* and *PSMT2* have a 73 kb distance, and the *CYP82Y1* lies between them. In the area between the 10 genes retrotransposons and DNA transposable elements are localized. These elements may have same function in gene rearrangement for cluster

formation, as in the cases of the thalianol and marneral clusters in *Arabidopsis* (Winzer *et al.*, 2012).

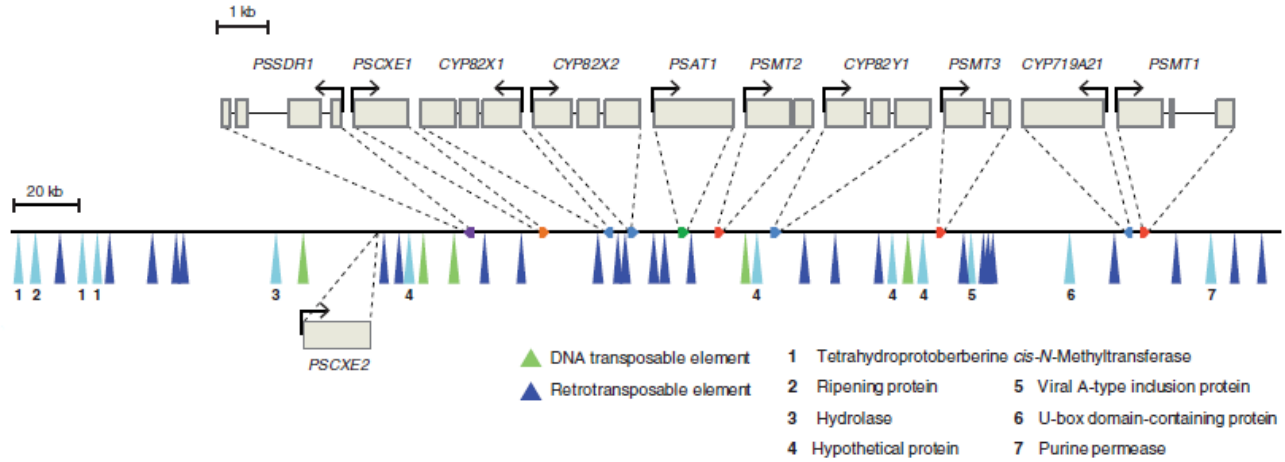


Figure 18. The HN1 gene cluster (401 kb). The 10 HN1-specific genes are shown above the central black line, exons are represented by solid gray boxes and introns by fine black lines. Arrows indicate the 5′ to 3′ orientation of each gene (Winzer *et al.*, 2012).

The novel noscapine HN1 biosynthetic pathway was proposed, with some steps remaining unaccounted. However, HN1 cluster identification extends the involvement of gene clusters to the alkaloid class of secondary metabolites in higher plants (Winzer *et al.*, 2012).

A gene cluster for the biosynthesis of the alkaloids in solanaceous crops: The most recent operon-like gene cluster reported is the cluster for the synthesis of alkaloids in solanaceous plants, like *Solanum tuberosum* (potato) and *S. lycopersicum* (tomato) (Itkin *et al.*, 2013). Steroidal glycoalkaloids (SGAs) in widely used vegetable crops, such as potato and tomato, are a class of antinutritional compounds that remain in our food chain and daily diet (Friedman, 2006). The glycoalkaloids a-solanine and a-chaconine are the principal toxic substances in potato. These SGAs cause gastrointestinal and neurological disorders and can be lethal to humans when accumulate at high concentrations. SGA biosynthesis requires genes encoding uridine 5´-diphosphate (UDP)–glycosyltransferases (UGTs), which decorate the steroidal alkaloid (SA) skeleton with various sugar moieties. The tomato GLYCOALKALOID METABOLISM 1 (GAME1) glycosyltransferase is the homolog of the potato SGT1 (McCue *et al.*, 2005), which catalyzes galactosylation of the alkamine tomatidine (Itkin *et al.*, 2011). Genes

associated with SGA biosynthesis in potato and tomato were identified according to co-expression network analysis (Itkin *et al*., 2013). Sixteen genes from each species were co-expressed with *GAME1/SGT1* gene. One of these genes, which we termed as GLYCOALKALOID METABOLISM 4 (*GAME4*), encodes a member of the CYP88D subfamily of cytochrome P450 proteins. Expression of *GAME4* and *GAME1/SGT1* is very similar in tomato and potato. The *GAME1/SGT1* and *GAME4* genes are positioned in chromosomes 7 and 12 in tomato and potato, respectively, and several co-expressed genes lie next to them (Figure 19). Silencing of *GAME4* in potato by RNA interference (RNAi, *GAME4i* plants), resulted in a reduction in the levels of a-solanine/chaconine and other SGAs in both leaves and tubers. Similarly, in *GAME4i* tomato plants, the dominant SGA in leaves and mature green fruit (a-tomatine) was severely reduced. This strategy may provide a means for deletion of unsafe, antinutricional compounds detected in staple vegetable crops (Itkin *et al*., 2013).

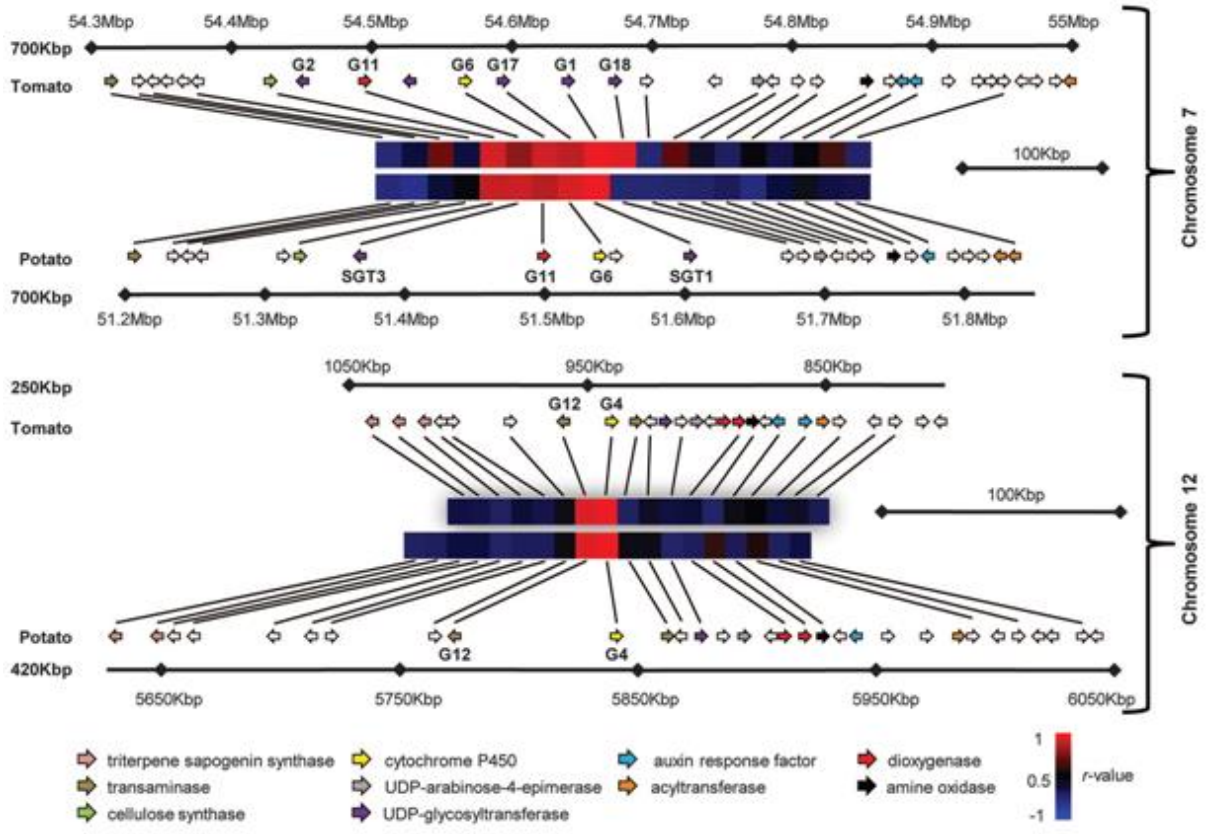


Figure 19. Schematic map of genes identified in the duplicated genomic regions in tomato and potato and their co-expression. *GAME1/SGT1* is located on chromosomes 7, while *GAME4* on chromosomes 12. Specific gene families are indicated by colored arrows, whereas members of other gene families are shown by white arrows. Note the homology in genes flanking the high co-expression regions and positioned in a matching sequence along the genome, suggesting a common origin of the regions on both chromosomes (Itkin *et al*., 2013).

A.7. Genetic manipulation in plants

A.7.1. Plant transformation of legumes

Legumes offer unique opportunities for the study of plant-microbe interactions such as symbiotic nitrogen fixation, mycorrhizal associations and legume-pathogen interactions. The development of efficient plant transformation and regeneration methods is a prerequisite for functional genomics. *L. japonicus* and *M. truncatula* are susceptible to *Agrobacterium*-mediated transformation. The molecular basis of the neoplasmic transformation of plant cells by the crown gall bacterium *A. tumefaciens* is the transfer and stable maintenance of T-DNA, a well defined segment of the bacterial Ti plasmid, in plant cells (Chilton *et al*., 1977). Ti plasmid can be utilized as a host vector system for introducing foreign DNA in plant cells (Hernalsteens *et al*., 1992). Transgenic plants are usually made by *Agrobacterium*-mediated hypocotyl transformation (Handberg and Stougaard, 1992; Stiller *et al*., 1997) and selection for antibiotic resistance using *nptII*, *hptII* and *bar* genes. More recently, a new transformation procedure has been reported, in which root explants are used for *Agrobacterium*-mediated transformation (Lombardi *et al*., 2003). The new method results in a tenfold increase of in the number of transformants within about 4 months.

Unlike *A. tumefaciens*, *A. rhizogenes* generates adventitious, genetically transformed hairy roots at the site of inoculation. Upon expression of the *root locus* (*rol*) genes carried on the root inducing (Ri) plasmid-derived T-DNA, roots are formed of which a certain number will be transformed with the binary vector carrying the gene of interest (Limpens *et al*., 2004). *A. rhizogenes*-mediated transformation results in the production of "composite plants", a term derived from the fact that transformed roots are induced on a non-transformed plant (Boisson-Dernier *et al*., 2001). Although the generation of "composite plants" is a rapid tool to test gene and promoter function in roots, the transgenic trait cannot be transmitted to the progeny. Furthermore, the roots can also be chimeric (Limpens *et al*., 2004). The *A. rhizogenes*-mediated transformation procedure has been successfully utilized several times (Shimomura *et al*., 2006; Govindarajulu *et al*., 2009; Wang *et al*., 2013).

A.7.2. Heterologous expression in *Nicotiana benthamiana*

Two principal means have been developed for the expression of heterologous proteins in plants, stable genetic transformation and utilization of autonomously replicating viral vectors. In transgenic plants the foreign gene is stably incorporated into the plant genome and inherited by the next generations. Constraints on the size or complexity of the proteins that can be expressed are not generally encountered, but high expression levels cannot easily achieved (Canizares *et al*., 2006). Viral vectors can lead to high expression levels of the foreign gene, but there are still limitations on the size of the protein that can be expressed before genetic instability becomes a problem. Furthermore, concerns have been expressed about the possibilities of viral vectors carrying foreign genes to spread in the environment (Canizares *et al*., 2006). As a result, attention has recently turned toward the development of defective versions of viral RNAs which, though able to replicate, cannot spread in the environment (Gleba *et al*., 2004; Canizares *et al*., 2006).

Several viruses have been utilized for the plant-based protein production, for example the tobacco mosaic virus (TMV) (Lindbo, 2007), *Potato virus X* (Chapman *et al*., 1992) and cowpea mosaic virus (CPMV) (Liu *et al*., 2005). A significant disadvantage of vectors based on monopartite viruses, such as TMV, is their inability to co-express multiple proteins. This limitation can be overcome by using vectors based on two different viruses that exist synergistically in nature, such as TMV and *Potato virus X* (Pruss *et al*., 1997). The separate H and L chains of a tumor-specific IgG were expressed in TMV and *potato virus X*-based vectors (Giritch *et al*., 2006).

Cowpea mosaic virus (CPMV) infects a number of legume species and the experimental host, *N. benthamiana*. The genome of CPMV consists of two separately encapsidated positive-strand RNA molecules of 5889 (RNA-1) and 3481 (RNA-2) nucleotides. RNA-1 encodes proteins required for the replication of viral RNAs and polyprotein processing. RNA-2 encodes the 48K movement protein and the two coat proteins, which are necessary for cell to cell movement and systemic spread (Sainsbury *et al*., 2009). The high level productivity makes CPMV a very attractive system for the

production of pharmaceutical proteins. The CPMV system has been used to express a number of recombinant proteins, such as the hepatitis B core antigen (HBcAg) and small immune proteins (SIPs), in cowpea plants (Alamillo *et al*., 2006; Mechtcheriakova *et al*., 2006; Monger *et al*., 2006). Many therapeutic proteins, such as antibodies, are heteromeric, consisting of complexes of more than one polypeptide. The CPMV-based systems raise the possibility of the production of multichain and heteromeric complexes, as there is no obvious limit on the number of CPMV RNA-2-based constructs that can be co-infiltrated (Sainsbury and Lomonossof, 2008). For example, it was possible to use two different RNA-2 molecules containing the heavy and light chains to express assembled IgG in plants (Sainsbury *et al*., 2008). Overall, the CPMV-based systems that have been developed offer a wide range of options for the expression of foreign proteins in plants. The levels of foreign gene expression achieved, although probably sufficient for most applications, are not as high as those reported for chloroplast transformation (Daniell *et al*., 2002). However, unlike chloroplast transformation, it permits post-translational modification of proteins. The ability to target proteins to the secretory pathway, as well as allowing complex post-translational modifications, such as glycosylation, to take place, often significantly increases their stability and hence increases their levels of accumulation (Canizares *et al*., 2006).

A.8. Aim of this present work

Plants are characterized by a vast array of complex biochemical pathways resulting in the production of a plethora of natural products or secondary metabolites. The ability of different plant lineages to synthesize distinct secondary metabolites is likely to have been a key to the survival and diversification of plant species. The largest group of natural products is the terpenes and the major subgroup of them is the triterpenes (Xu *et al*., 2004). Triterpenes are synthesized from mevalonate via 2,3-oxidosqualene. A number of genes and enzymes for triterpene biosynthesis have been characterized in plants (Osbourn, 2010; Augustin *et al*., 2011). Unlike sterols, the biological functions of plant triterpenes remain less well understood. Legumes produce a huge variety of triterpenes (Phillips *et al*., 2006). Among them, triterpenoid saponins receive

considerable interest because of their diverse biological activities and beneficial properties. The triterpenoid saponin content of *M. truncatula* has been greatly studies (Huhman and Sumner, 2002), whereas the information for triterpenoid saponins of *L. japonicus* remains scant. β-amyrin has been detected during normal growth and development, but is up-regulated during the establishment of symbiotic relations with rhizobia bacteria or mycorrhizal fungi (Baisted, 1971; Hernandez and Cooke, 1996; Iturbe-Ormaetxe *et al*., 2003). Lupeol is frequently associated with root nodulation in several plants (Hartmann *et al*., 2002; Iturbe-Ormaetxe *et al*., 2003; Hayashi *et al*., 2004; Sawai *et al*., 2006b). Interestingly, genes for certain triterpene biosynthetic pathways exist as metabolic gene clusters in oat and *A. thaliana* (Qi *et al*., 2004; 2006; Field and Osbourn, 2008; Osbourn, 2010; Field *et al*., 2011; Chu *et al*., 2011).

This study aims to investigate the potential of organization of *OSC* genes in secondary metabolic gene clusters in legumes. In addition, this study is an effort to unravel metabolic diversity in legumes and possible roles of simple triterpenes, like lupeol and β-amyrin. For this purpose, the genomic regions flanking members of *OSC* genes in *L. japonicus* and *M. truncatula* were initially investigated for candidate gene clusters involved in triterpene biosynthesis. This led to the identification of four loci where *OSC* genes are flanked by genes encoding candidate tailoring enzymes. Further investigations were focused on one of these in *L. japonicus*, which contained the *AMY2* gene along with genes for two different classes of CYP450, belonging to the CYP88D and CYP71D families, and a gene predicted to encode a reductase. The expression pattern of *AMY2*, *LjCYP88D5*, *LjCYP71D353* and *LjSDRt* was studied under different developmental and environmental conditions, revealing coordinate regulation of *AMY2* cluster genes. Functional characterization of the *AMY2* cluster genes was accomplished by heterologous expression in *N. benthamiana*, resulting in novel enzyme functions in triterpene biosynthesis. A novel triterpene structure, dihydrolupeol, and its subsequent conversion to 20-hydroxynetulinic acid catalyzed by the sequential activity of AMY2 and LjCYP71D353 were uncovered. Using functional genomics approaches a possible role for this pathway in plant development was revealed. Finally, a transcriptional gene

silencing mechanism that appears to be involved in the regulation of *AMY2* cluster genes was identified.

In addition, some initial steps have been realized for the verification of the existence of an operon-like gene cluster in *βAS* loci of *M. truncatula*. Knockout experiments in which transgenic *M. truncatula* for *βAS* were generated through *A. rhizogenes* transformation. The expression pattern of *βAS* gene was studied under various developmental conditions. Furthermore, silenced plant lines shown an increased nodule number when compared with control plants, disclosing a possible involvement of *βAS* in nodulation.

In parallel, the role of lupeol synthase and its product, lupeol, in developing roots and nodules of *L. japonicus* was investigated. The expression patterns of *OSC3* in different developmental stages of uninfected roots and roots infected with *M. loti* were determined. Knockout experiments were carried out, in which transgenic *L. japonicus* roots silenced for *OSC3* were generated through *A. rhizogenes* transformation. The effects of gene silencing on nodule formation and function were investigated. The expression of *ENOD40*, a marker gene for nodule primordia initiation, was increased significantly in the *OSC3*-silenced plant lines, suggesting that lupeol influences nodule formation. Silenced plants also exhibited a more rapid nodulation phenotype, consistent with this. All these experimental data suggest a role for lupeol biosynthesis in nodule formation through the regulation of *ENOD40* gene expression.

B. MATERIALS AND METHODS

CHAPTER B. MATERIALS AND METHODS

B.1. Plant material and growth conditions

L. japonicus (Cultivar Gifu B-129 or MG20) seeds were scarified for 5 min with sulfuric acid (H_2SO_4), followed by several washes in cold sterile water, immediately sterilized for 20 min in 5% NaOCl - 0.02% Tween 20 solution and rinsed with sterile distilled water in laminar flow conditions. The seeds were pre-germinated in Petri dishes containing 1% water-agar in the dark for 1 day at 4° C and 2 days at 22 ºC, and then were transferred to humid sterile perlite/vermiculite (1:1) into plastic trays. For inoculation with *Mesorhizobium loti* (strain R7A), pre-germinated 3 day-old seedlings were inoculated with 0.1 OD600 suspension culture in PBS (phosphate buffered saline) (4.3 mM Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, 1.4 mM KH_2PO_4, pH 7.4) solution and the plants were grown in a modified nitrogen-free Hoagland nutrient solution. Both infected and uninfected plants were grown at 22 °C under a photoperiod of 16/8 h light/dark cycle, at 200 µmol photons m^{-2} s^{-1} for one month. The uninfected plants were watered periodically with a full strength Hoagland nutrient solution.

Seeds of *M. truncatula* cv A17 genotype were surface-sterilized in H_2SO_4 for 10 min, followed by several washes in cold sterile water. Then they sterilized for 10 min in commercial bleach and washed again with sterile distilled water in laminar flow conditions. Seeds were imbibed in the dark at 4° C for two days on wet 3MM Whatman paper and then transferred in Petri dishes containing Murashige and Skoog medium (MS, Duchefa, Haarlem, The Netherlands). *M. truncatula* seedlings were used in exogenous hormone treatments and *Agrobacterium rhizogenes* plant transformation; procedures described thoroughly in next sections. Inoculation with *Sinorhizobium meliloti* (strain RCR2011) and plant growth were conducted as in the case of *L. japonicus* plants.

B.2. PCR conditions

PCR segments used for plant transformation and in situ hybridization were amplified using gene-specific primers (Table A1) and Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA), following the manufacturer's instructions. PCR cycling started with initial denaturation at 98° C for 30 s, followed by 35 cycles of 98° C for 10 s, the corresponding primer annealing temperature for 20 s and 72° C for 15 s, and ended up with the final extension step at 72° C for 5 min.

The full-length cDNAs for *LjCYP88D5* and *LjCYP71D353* were obtained by PCR amplification with gene-specific primers (Table A1) based on genomic DNA sequence information. The cDNA used as template was synthesized from root RNA 14 days old. PCR amplification was performed with the Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) and the following PCR conditions: an initial denaturation at 98° C for 30 s, followed by 35 cycles of 98° C for 10 s, the corresponding primer annealing temperature for 30 s and 72° C for 30 s, and ended up with the final extension step at 72° C for 10 min.

The target regions of *LjCYP71D353* and *LjSDRt* for determination of DNA methylation were amplified using the HotStart® DNA Polymerase (Qiagen, Hilden, Germany) and the primer pairs described in Table A1. PCR cycling started with an initial activation step at 95° C for 15 min, followed by 35 cycles of 94° C for 30 s, approximately 5° C below the corresponding primer annealing temperature for 30 s and 72° C for 1 min, and ended up with the final extension step at 72° C for 10 min. The PCR amplicons of all procedures described were inserted into the pGEM-T Easy vector (Promega, Madison, WI, USA), maintained into *E. coli* strain DH5a and the complete sequences were determined.

The PCR reaction to test the complete DNA removal from RNA samples was conducted using the KAPATaq DNA Polymerase according to the protocol instructions: an initial denaturation at 95° C for 2 min, followed by 35 cycles of 95°

C for 30 s, 56° C for 30 s and 72° C for 30 s, and ended up with the final extension step at 72° C for 2 min.

B.3. *In situ* hybridization

Nodule segments from plants 14 and 28 days post-infection were fixed in 4% (w/v) paraformaldehyde supplemented with 0.25% glutaraldehyde in 10 mM sodium phosphate buffer pH 7.4, for 1 h, in a vacuum aspirator. Fixed tissues were dehydrated in ethanol series, and then exchanged with xylene before embedding in paraffin. Pairs of *AMY2*, *LjCYP88D5*, *LjCYP71D353* and *OSC3* gene specific primers were designed, AMY2isF/AMY2isR, LjCYP88D5isF/LjCYP88D5R, LjCYP71D353isF/LjCYP71D353isR and OSC3isF/ OSC3isR (Table A1) producing 206 bp, 331 bp, 345 bp and 300 bp amplicons, respectively. Fragments were ligated into the pGEM-T Easy vector (Promega, Madison, WI, USA) and the complete sequences of the products were determined. Antisense and sense RNA probes were labeled with digoxigenin (DIG)-11-rUTP (ROCHE Mannheim, Germany) by in vitro transcription of PCR products derived from the *AMY2*, *LjCYP88D5*, *LjCYP71D353* and *OSC3* clones using the SP6 and T7 promoters of the pGEM-T Easy vector. Sections (7µm) were placed on poly-L-lysine slides, digested with proteinase K for 30 min at 37° C, treated with acetic anhydride, dried in ethanol, and then hybridized with the appropriate gene-specific DIG-labelled probes overnight, at 42° C. After washing with 4XSSC containing 5 mM DTT, the slides were treated with RNase A for 30 min at 37° C, washed with RNase A buffer 500 mM NaCl, 1 mM EDTA (ethylenediaminetetraacetic acid), 10 mM TRIS–HCl, pH 7.5, containing 5 mM DTT, and then processed for revealing the DIG antigen. This involved blocking with DIG-blocking reagent and bovine serum albumin, followed by incubation with an anti-DIG antibody conjugated to alkaline phosphatase, and washing with blocking reagent. The color revealed by incubation in 5-bromo-3- chloro-3-indolyphosphate nitroblue tetrazolium (BCIP/NBT) (ROCHE Mannheim, Germany) and the reactions were stopped with water, the slides dehydrated, air-dried, and then mounted in

DPX (BDH Chemicals, England) before viewing. The slides were observed in a Leica DM2000 microscope and photographs were taken using a Leica DFC 490 camera and the corresponding Leica Microsystems software.

B.4. Real-time PCR experiments

Roots, nodules and leaves at different developmental stages of *L. japonicus* plants, grown as described previously, were collected and ground in liquid nitrogen. Total RNA was isolated from the organs of 50 plants for each sample using the Qiagen RNeasy Extraction Kit (Qiagen, Hilden, Germany). Root tissues of successfully transformed and control *L. japonicus* and *M. truncatula* plants at 28 dpi with *M. loti* or *S. meliloti*, as well as root tissues of exogenous hormone and abiotic stress treatments were subjected to the same procedure. The total RNA and protein concentrations for all samples were quantified spectrophotometrically by measuring the absorbance at 260 nm, as well as the absorbance ratio 260:280 nm, in a Biophotometer (Eppendorf, Hamburg, Germany) and on 2% w/v agarose gel. To eliminate genomic DNA contamination, the samples were treated with DNaseI (TaKaRa, Otsu Shiga, Japan or Invitrogen, Carlsbad, CA, USA) at 37^{o} C for 30 min. Subsequently *UBIQUITIN* (*UBQ*) and *Mtc27* (constitutively expressed gene in *M. truncatula*) primers UBQF/UBQR and Mtc27F/Mtc27R were employed in PCR to test the complete DNA removal using *L. japonicus* and *M. truncatula* genomic DNA respectively as positive controls.

First-strand cDNA was reverse transcribed with AffinityScript^{TM} Multi Temperature (Stratagene, Santa Clara, CA, USA) or SuperScript^{TM} II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) from 2 μg of DNase-treated total RNA using Oligo(dT)$_{12-18}$ Primer (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The resulting first-strand cDNA was normalized for the expression of the housekeeping gene of *UBQ*. Gene specific primers for the *UBQ*, *AMY2*, *LjCYP88D5, LjCYP71D353*, *LjSDRt*, *ENOD2*, *ENOD40 L. japonicus* genes and *Mtc27*, *MtBasI M. truncatula* genes were designed with Beacon designer v7.01

software (Premier Biosoft, CA, USA): UBQrtF, UBQrtR, AMY2rtF, AMY2rtR, LjCYP88D5rtF, LjCYP88D5R, LjCYP71D353rtF, LjCYP71D353rtR, LjSDRTF, LjSDRTR, ENOD2F, ENOD2R, ENOD40F, ENOD40R, Mtc27F, Mtc27R, MtBasIF and MtBasIR, respectively.

Quantitative reverse transcription-polymerase chain reactions (RT-PCRs) were performed on an MX 3005P system (Stratagene, CA, USA) using Brilliant II QPCR Master Mix (Stratagene, CA, USA) and gene-specific primers, following the manufacturer's instructions. PCR cycling started with initial polymerase activation at 95^{o} C for 10 min, followed by 40-45 cycles of 95^{o} C for 15 s, 60^{o} C for 20 s and 72^{o} C for 10 s. The primer specificity and the formation of primer dimers were monitored by dissociation curve analysis and agarose gel electrophoresis on a 3% (w/v) gel. In all samples, a single amplicon was detected.

The expression levels of an *L. japonicus UBQ* gene and *M. truncatula Mtc27* gene were used as internal standards to normalize small differences in cDNA template amounts. For the relative quantification of gene expression, a modification of the comparative threshold cycle method was used. The relative abundance of all transcripts amplified was normalized to the constitutive expression level of *UBQ* or *Mtc27* mRNA. Relative transcript levels in different samples for the gene of interest were calculated as a ratio to the *UBQ* or *Mtc27* gene transcripts. Data were analyzed according to Pfaffl (2001) and the reaction efficiencies were estimated with LinRegPCR (Ramakers *et al*., 2003). For all samples, triplicate PCRs were performed for each gene. The sequences of all primers used for Real-time experiments are recorded in Table A1.

B.5. Exogenous hormone and abiotic stress treatments

L. japonicus (cultivar Gifu B-129) and *M. truncatula* cv A17 genotype seeds were pre-germinated as described above. Seedlings were grown for 7 days on Petri dishes containing MS (Duchefa, Haarlem, The Netherlands) with 1% sucrose substrate at 22^{o}C in a 16h/8h dark/light photoperiod. Then the seedlings were

transferred on Petri dishes containing MS with 1% sucrose and 10 μM and 25 μM methyl jasmonate (MeJA) (Duchefa, Haarlem, The Netherlands), 0.6 mg/L^{-1} benzylaminopurine (BA) (Sigma Chemical Co., St Louis, MO, USA) or 0.15 mg/L^{-1} 2,4 dichlorophenoxyacetic acid (2, 4 – D) (Sigma Chemical Co., St Louis, MO, USA) for 7 days. Control plants were grown on Petri dishes containing MS supplemented with the respective amount of ethanol for the MeJA and 2,4-D treatments.

For salt stress treatments, *L. japonicus* seedlings were grown for 7 days on Petri dishes containing MS (Duchefa, Haarlem, The Netherlands) with 1% sucrose substrate at 22^{o}C in a 16h/8h dark/light photoperiod. Then the seedlings were transferred on Petri dishes containing in MS (Duchefa, Haarlem, The Netherlands) with 1% sucrose supplemented with 25 mM, 50 mM or 75 mM NaCl for 7 days.

For the heat/cold stress, *L. japonicus* seedlings were grown for 11 days on Petri dishes containing MS (Duchefa, Haarlem, The Netherlands) substrate at 22^{o}C in a 16h/8h dark/light photoperiod. Then the seedlings were exposed to 37^{o}C for heat stress and to 4^{o}C for cold stress for 24 hours.

For oxidative stress, *L. japonicus* seedlings were grown for 7 days on Petri dishes containing MS (Duchefa, Haarlem, The Netherlands) with 1% sucrose substrate at 22oC in a 16h/8h dark/light photoperiod. Then the seedlings were transferred on petri dishes containing MS (Duchefa, Haarlem, The Netherlands) with 1% sucrose and 150nM paraquat (Sigma Chemical Co., St Louis, MO, USA) for 7 days.

Roots from 20-50 seedlings per treatment were collected and subjected to total RNA isolation using the Qiagen RNeasy Mini Kit (Qiagen, Hilden, Germany). Genomic DNA was removed from RNA samples, cDNA samples were prepared from isolated RNA and their concentration was normalized. Real-time experiments were carried out as described above.

B.6. Physiological measurements

The numbers of nodules of *OSC3*-silenced plants lines and of control plants were determined 20 days after inoculation with *M. loti*. Nodule numbers were reassessed

at 40 days and nodules were collected for nitrogen fixation efficiency determination (performed according to Balis *et al*., 1996). The number of nodules and the length of transgenic roots in *βAS*-silenced plant lines were determined 20 days after inoculation with *S. meliloti*.

B.7. Heterologous expression into *Nicotiana benthamiana* leaves

Plasmids pBinPS1NT, pBinPS2NT and pBinPS2NT2AGFP, containing full-length copies of CPMV RNA-1 (35S-RNA-1), RNA-2 (35S-RNA-2) and RNA-2- GFP (35S-RNA-2-GFP), respectively, in the binary transformation vector pBINPLUS (van Engelen *et al*., 1995), have been described previously (Liu and Lomonossoff, 2002). The creation of deleted versions of RNA-2, based on the vector pN81S2NT containing the complete sequence of RNA-2, have been described before (Canizares *et al*., 2006). Plasmids used in heterologous *N. benthamiana* expression system are described in detail in Appendix 1. Full-length gene of *AMY2* was digested with *StuI* / *BstBI* and full-length genes of *LjCYP88D5* and *LjCYP71D353*¸ obtained using two pairs of specific primers LjCYP88D5Fl-F / LjCYP88D5Fl-R and LjCYP71D353Fl-F / LjCYP71D353Fl-R, respectively, were digested with *StuI* / *Bsu36i*. pM81-FSC1 vector was also digested with the same enzymes. In this way, the whole of the RNA-2 ORF downstream of AUG 512 encoding both the movement protein and the viral coat proteins was excised by digestion and replaced by the sequences of interest with compatible ends. The pM81-FSC1 derived plasmids were digested with *AscI* / *PacI* and the fragments containing *AMY2*, *LjCYP88D5* and *LjCYP71D353* sequences were transferred to the similarly digested vector pBINPLUS (van Engelen *et al*., 1995).

The derivative plasmids were maintained in *A. tumefasiens* strain LBA44. Cultures were co-infiltrated with an *Agrobacterium* culture carrying the pBIN61-P19 plasmid which encodes for the P19 silencing suppressor protein (Voinnet *et al*., 2003). For co-infiltration experiments the used cultures were mixed to an equal density. Agroinfiltration into *N. benthamiana* leaves was carried out as previously

described (Canizares *et al*., 2006). *N. benthamiana* plants were grown in greenhouses with supplemental light to 16 h at a constant temperature of 24°C. *A. tumefaciens* cultures carrying the pBINPLUS-AMY2, pBINPLUS-LjCYP88D5 and pBINPLUS-LjCYP71D353 plasmids were grown to stationary phase. *A. tumefaciens* cells were centrifuged and resuspended in MMA [10 mM MES (2-[N-morpholino] ethanesulfonic acid; Sigma-Aldrich) pH 5.6, 10 mM $MgCl_2$, 100 µM Acetosyringone (Sigma-Aldrich)] to an OD_{600} of 1.0. Cultures were incubated for at least 2 h at room temperature. *A. tumefaciens* cultures carrying the pBINPLUS-AMY2, pBINPLUS-LjCYP88D5 and pBINPLUS-LjCYP71D353 plasmids were mixed in equivalent volumes of equal-density *A. tumefaciens* culture of the pBIN61-P19 plasmid. The underside of *N. benthamiana* leaves was nipped with a razor blade or small pipette tip and the wounds were infiltrated against a counter pressure (provided by a finger) with the Agrobacteria mixture using a 1 or 2 ml syringe. Leaf tissue was harvested, frozen in liquid nitrogen after six days and metabolite extraction was conducted.

B.8. *Agrobacterium rhizogenes* plant transformation

A polyubiquitin promoter-based binary vector, pUBI-GWS-GFP which allows for GFP overproduction (Maekawa *et al*., 2008) was used for silencing of the *AMY2*, *LjCYP88D5* and *LjCYP71D353 L. japonicus* genes, as well as the *βAS* gene from *M. truncatula*. pUBI-GWS-GFP vector contains the 35S promoter fused with the GFP gene for the expression of the reporter gene. The expression of the silencing construct is attributed to the activity of the promoter of *Ljubq1*. The high transcriptional activity of polyubiquitin promoters has been investigated in plants and *Ljubq1* possesses higher activity than the CaMV35S promoter in roots and nodules. In addition, pUBI-GWS-GFP has tandem GATEWAY cassettes, in opposite orientation to each other. The intron AtWRKY33 is located between the cassettes (Maekawa *et al*., 2008). As a result, pUBI-GWS-GFP is able to produce double-stranded RNA with the intron and, consequently, efficient gene silencing (Wesley *et*

al., 2001). Additionally, the GATEWAY-compatibility of pUBI-GWS-GFP limits the usage of restriction enzymes during the entire cloning process and the negative selection marker *ccdB* gene eliminates the original vector after bacterial transformation. The GFP reporter gene allows the easier selection of hairy root transformants. In entirely strong GFP-expressed roots and in chimeric or weak GFP-expressed roots, 84% and 65% of roots are expected to be silenced (Maekawa *et al*., 2008). More details for the GATEWAY technology are described in Appendix 1.

PCR amplicons were produced using cDNA from 14 days old roots as template and pairs of specific primers AMY2-3F, AMY2-3R, AMY2-2F, AMY2-2R, LjCYP88D5-1F, LjCYP88D5-1R, LjCYP88D5-3F, LjCYP88D5-3R, LjCYP71D353RNAiF, LjCYP71D353RNAiR, MtBasIRNAiF and MtBasIRNAiR (Table A1). Two *AMY2*, two *LjCYP88D5* and one *MtBasI* ORF sections named *AMY2-2*, *AMY2-3, LjCYP88D5-1*, *LjCYP88D5-3*, *LjCYP71D353-1* and *MtBasI-1* were cloned into the *KpnI – XhoI* restriction sites of the pENTR4 plasmid. The *LjCYP71D353* was cloned into *NcoI – SalI* restriction sites of the pENTR4 vector (Invitrogen, Carlsbad, CA, USA) replacing the *ccdB* gene of the original vector. The four new clones, pENTRY-AMY2.2/2.3, pENTRY-LjCYP88D5-1/3, pENTRY-LjCYP71D353-1 and pENTRY-MtBasI-1 were then used in an LR Clonase reaction (Invitrogen, Carlsbad, CA, USA) with destination vector pUBI-GWS-GFP in order to create the final binary expression vectors that were used in a plant binary transformation system.

pCambia 1300 vector was used for silencing of *OSC3* gene of *L. japonicus*. A PCR amplicon was produced using, as template, cDNA from 7-d-old roots and a pair of specific primers, OSC3ri1F / OSC3ri1R. The PCR product, named hereafter OSC3ri, was digested with *XbaI／HindIII* and *XhoI／KpnI* and subcloned into pHannibal in the antisense and sense orientation, respectively. The complete construct with Cauliflower mosaic virus (CaMV) 35S promoter and Octopine Synthase (OCS) terminator from pHannibal was digested with *SacI ／ PstI* and subcloned into pCambia 1300, yielding vector pCOSC3. All binary vectors described were transferred to the appropriate *Agrobacterium rhizogenes* strains by electroporation.

Hairy root transformation of *L. japonicus* (cultivar MG20) and *M. truncatula* cv A17 genotype, utilizing *A. rhizogenes* strains LBA 1334 and ARqua1 respectively, carrying the binary vectors described, was performed according to (Martirani *et al*., 1999). Sterilised seeds were lined on 0.8% Bactoagar plates and the plates were covered with aluminium foil. The plates were placed vertically at 22 °C under a photoperiod of 16/8 h light/dark cycle, at 200 µmol photons m^{-2} s^{-1}. Then the aluminium foil was removed and the plates were incubated in the controlled environment for 3 days more. Simultaneously, a liquid *Agrobacterium rhizogenes* culture was incubated for 2 days in 28^{o}C and then 0.5 ml of the culture were transferred to LB plate with the appropriate antibiotics. A suspension culture was made on the LB plates using 2 ml sterilized H_2O and was transferred to sterilized filter paper. Seedlings were placed on the filter paper and hypocotyls were cut above the root with a scalpel. Seedlings were transferred to B5 (Duchefa, Haarlem, The Netherlands) medium plates. Plates were covered with aluminium foil and placed horizontally to the growth chamber for 2 days. After that, the aluminium foil was removed and the plates remained in the controlled environment for 3 days more. Seedlings were transferred to B5 (Duchefa, Haarlem, The Netherlands) medium plates with cefotaxime and plates were placed vertically in the growth chamber. Seedlings were grown in B5 (Duchefa, Haarlem, The Netherlands) medium with cefotaxime for 7 days and then they were transferred to Jensen's medium (Bardulova and Chiurazzi, 2005) for 7 – 10 days until the root growth. Control plant lines were obtained following the same procedure and *A. rhizogenes* LBA 1334 and ARqua1 were transformed with empty pUBI-GWS-GFP vector. In the case of *OSC3* silencing control plant lines were obtained when inoculated with *A. rhizogenes* LBA 1334 carrying the empty pCambia 1300 vector.

Following root development, roots were examined for GFP production under UV light in a Leica DM2000 microscope, wild-type roots were removed and only transgenic roots were allowed to grow further in order to produce a new transformed root system per plant. The *L. japonicus* and *M. truncatula* plants were

transferred to sterile perlite/vermiculite (1:1) pots, inoculated with *M. loti* and *S. meliloti*, respectively, and watered periodically with a modified nitrogen-free Hoagland nutrient solution as described. The transformation procedure was repeated twice in the case of *L. japonicus* genes.

In the case for *OSC3* silencing, following transformation, the plants were transferred to sterile perlite/vermiculite (1:1) pots, inoculated with *M. loti* and watered periodically with a modified nitrogen-free Hoagland nutrient solution, as described previously. Twenty days after inoculation, plant roots were carefully washed, a small segment of plant roots was excised and the genomic DNA was isolated with Nucleospin plant (Macherey-Nagel, Germany). The plants were screened for the incorporation of hairpin construct into their genome utilizing a pairs of primers: OSC3riF / HANantis and OSC3riF / HANsense. The HANsense and HANantis primers were designed against the intron of pHannibal vector, producing PCR products of different size. Eighteen successfully transformed plants were detected.

Root tissue from all *L. japonicus* and *M. truncatula* transformed plants, was retained for RNA extraction using Qiagen RNeasy Mini Kit (Qiagen, Hilden, Germany) and the plants, together with their corresponding control plants, were allowed to grow further for another 20 days. Real time PCR experiments were conducted as described above in order to identify if the transformed plants are also silenced for the corresponding genes. Nodule numbers were counted at 20 and 40 days in all transformed plants in comparison to control plants.

B.9. *Agrobacterium tumefaciens* stable transformation

A PCR amplicon was produced using cDNA from 14 days old roots as template and a pair of specific primers AMY2-2F and AMY2-2R (Table A1). The amplification product was digested with *XbaI* / *HindIII* and *XhoI* / *KpnI* in order to obtain the antisense and sense direction respectively, and ligated into pHannibal plasmid vector. The silencing construct was digested with *PstI* / *SacI,* excised and subcloned

into pBluescript vector. The insertion of the two fragments in the right orientation was verified by sequencing using designed primers for universal T7 and SP6 primers annealing on pBluescript. pBluescript vector containing the silencing construct was digested with *PstI* / *SacI* and subcloned into pCambia 1300. pCambia 1300 carries the hygromycin resistance gene, which is more suitable for *L. japonicus* transformation, because hygromycin selection results in less escape than the canamycin selection (Handberg and Stougaard, 1992). The resultant binary vector plasmid was transferred into *A. tumefaciens* strain AGL1 by electroporation.

L. japonicus plants were transformed following procedures previously described (Lombari *et al*., 2003; Barbulova and Chiurazzi, 2005). After seed sterilization, the plants were allowed to grow for 30 to 45 days in vitro on B5/2 medium. The freshly harvested roots were transferred on CIM such that the entire root system is in contact with the medium. During this incubation, a visible increase of the root thickness was observed. For co-cultivation, the roots were dipped in the appropriate *Agrobacterium* culture (about 5 X 10^9 cells / ml) for at least 10 min and wounded by cutting into 0.5 cm pieces and by squeezing firmly with forceps. The root pieces were then blotted quickly on sterile paper and transferred on fresh CIM medium plates. After two days of co-cultivation, the bacterial mass was removed by rinsing the explants 3 times in sterile water, blotting on sterile paper and transferring on CIM medium plus 200 mg/l cefotaxime. After two days, the explants were transferred onto CIM medium containing 200 mg/l cefotaxime and 15 mg/l hygromycin. After three to four weeks, the putative transformants were scored for resistance and green sectors (0.2 - 0.4 cm) were isolated and transferred on SIM1 containing 200 mg/l cefotaxime and 15 mg/l hygromycin plus 10mM $NH^{4+}SO_4$ for 3 weeks. The calluses with regenerative structures were transferred on fresh SIM2 medium containing 200 mg/l cefotaxime and 15 mg/l hygromycin. In this step, the green parts of the calluses were cleaned carefully of the brown, necrotic tissue. Cultivation on SIM2 medium was about 15 days long, until small shoots were formed. Calli with small shoots emerging were then transferred to SEM without

selection for 10 days more, in order to be further elongated. Root induction was achieved on 2-4 cm long shoots by a ten days subculture on RIM and further subculture on B5/2 medium without hormones (REM).

Callus inducing medium (CIM)	**B5 medium**	**+ 3mg / l IAA**
		+ 0.15 mg/l 2,4-D
		+0.6 mg / l BA
		+ 0.3 mg / l IPA
		+ 3% sucrose
		+ 1% agarose
		+ 15 mg / l hygromycin (when needed)
		+ 200 mg / l cefotaxime (when needed)
Shoot inducing medium (SIM1)	MS medium	+ 0.5 mg / l TDZ
		+ 3% sucrose
		+ 1% plant agar
		+ 15 mg / l hygromycin
Shoot inducing medium (SIM2)	MS medium	+ 0.05 mg / l TDZ
		+ 3% sucrose
		+ 1% plant agar
		+ 15 mg / l hygromycin
Shoot elongation medium (SEM)	MS medium	+ 0.05 mg / l TDZ
		+ 3% sucrose
		+ 1% plant agar
Root inducing medium (RIM)	MS/2 or B5/2 medium	+ 0.1% mg / l NAA
		+ 3% sucrose
		+ 1% plant
Root elongation medium (REM)	MS/2 or B5/2 medium	+ 3% sucrose
		+ 1% plant agar

Table 2. The media used in stable transformation protocol. The pH of the listed media was adjusted to 5.8 with a KOH 10M before autoclaving. The filter sterilized selective antibiotics and hormones were added before use. The plant hormones were prepared as 10.000x stock solutions. They were solved in DMSO, aliquoted into eppendorf tubes and stored at -20^{o} C.

Control plants were transformed with the empty vector. Following transformation, the plants were transferred to sterile perlite/vermiculite (1:1) pots and watered periodically with full strength Hoagland nutrient solution. Five weeks after planting, roots were carefully washed and a small segment of plant roots was excised. Genomic DNA was isolated using the CTAB procedure (Doyle & Doyle, 1987). A PCR reaction was carried out, using pair of primers, which amplify the 35S promoter and the hygromycin gene, 35S-F, 35S-R, Hyg-F and Hyg-R (Table A1), respectively. Total RNA was isolated from the four identified transformed plants using RNeasy extraction Kit (Qiagen, Hilden, Germany). Real time PCR experiments were conducted as described above.

B.10. Protocols for bacterial transformation

B.10.1. Preparation of competent cells

The *E. coli* competent cells were prepared following the Calcium Chloride protocol. A single *E. coli* colony was inoculated into 5 ml LB in a 50 ml flask and grown overnight at 37°C. 1 ml of the *E. coli* culture was used to inoculate 100 ml of LB in a 250 ml flask. The *E. coli* culture was grown until the optical density reaches OD_{600} 0.2 – 0.3. The *E. coli* cells were collected by centrifugation, gently resuspended in 50 ml cold 0.1M $MgCl_2$ and incubated on ice for 30 min. Then, the *E. coli* cells were again centrifuged, resuspended in 25 ml cold 0.1M $CaCl_2$ and incubated on ice for 20 min. Finally, the *E. coli* cells were collected by centrifugation, gently resuspended in 10 ml cold 0.1M $CaCl_2$ / 20% glycerol, dispensed in eppendorf tubes (80 µl / tube), froze in liquid nitrogen and placed in -80°C.

The *A. tumefaciens* and *A. rhizogenes* competent cells were prepared according to the same protocol. A single colony of the *Agrobacterium* strain of choice was picked, inoculated in 5 ml of LB supplemented with the appropriate antibiotic selection and grown at 28^{o} C overnight. 1 ml of the overnight culture was inoculated into 50 ml of LB in a 250 ml flask and grown at 28^{o} C until mid-log (OD_{600} is between 0.5 and 1.0). The *Agrobacterium* cultures were chilled for 10 min

on ice, centrifuged and pellets were resuspended in 1 ml of ice cold 20 mM $CaCl_2$. 0.1 ml of the *Agrobacterium* suspension were dispensed into each of two pre-chilled 1.5 ml eppendorf tubes, froze in liquid nitrogen and placed in -80°C.

B.10.2. Transformation of competent cells

The *E. coli* strain DH5a was transformed in all cases using the heat shock procedure. For transforming with a DNA construct, 50 µl of competent cells were used. For transforming with a ligation, 100 µl of competent cells were used. An aliquot of *E. coli* competent cells was taken from –80^{o}C freezer and thaw on ice. 50 ng of circular DNA were added into *E. coli* cells, mixed gently and incubated on ice for 30 min. The eppendorf with the DNA and *E. coli* mixture was incubated into water bath at 42^{o}C for 1 min. Then the eppendorf was incubated on ice for 2 minutes to reduce damage to the *E. coli* cells. 200 µl of LB with no antibiotic was added and the eppendorf was incubated for 1 hour at 37^{o} C. The resulting culture was spread on LB plates supplemented with the appropriate antibiotic and grown overnight at 37^{o} C.

Binary vectors were transferred to the appropriate *A. rhizogenes* and *A. tumefaciens* strains by electroporation. An eppendorf containing *Agrobacterium* competent cells was thaw on ice. 60 µl of *Agrobacterium* cells and 600 ng of the binary vector were mixed gently and incubated for 3 min on ice. The plasmid DNA/cell mixture was transferred to a cuvette on ice avoiding air bubbles by gently shaking the cuvette. The cuvette was inserted into the slide of the shocking chamber and the slide was pushed into the chamber until the cuvette makes firm contact with the chamber electrodes. The "Pulse" button was pressed at the desired settings until a tone sounds indicating that the pulse has been given (*Agrobacterium* require 2.5 kV). The cuvette was immediately removed, 1 ml cold LB without antibiotics was added and the solution was mixed by gently pipetting up and down. The cells were transferred to an eppendorf and incubated at 28^{o} C for approximately 4 h. Then the cells were collected by centrifuge, resuspended in 200

ml LB medium, plated on LB plates with the appropriate antibiotics and incubated at 28^{o} C. The *Agrobacterium* colonies appeared in 2-3 days.

B.11. Metabolite extraction and GC-MS analysis

N. benthamiana leaf material was harvested, ground in liquid nitrogen, and lyophilized. The dry plant tissue (100–250 mg) was saponified at 70°C for 2 hours in 40 ml 10% KOH (w/v) in 80% EtOH (v/v) with 0.5 mg/ml butylated hydroxytoluene (Sigma-Aldrich, Poole, UK). Where appropriate an internal standard (betulin, Sigma-Aldrich) was incorporated during saponification. After hydrolysis the mixture was diluted with water (10 ml) and extracted three times with an equal volume of hexane. The hexane fraction was washed three times each with water and saturated NaCl solution. The hexane was then evaporated under reduced pressure (Field and Osbourn, 2008). After hexane extraction, hydrochloric acid was added to the aqueous solution to lower the pH~2.0 and another round of hexane extractions were performed to obtain an acid extraction fraction. The alkaline and acid hexane extracts were concentrated and derivatised with Tri- Sil Z (Pierce, Cambridge, UK) prior to GC-MS analysis. GC-MS analyses were conducted on an Agilent 5973 MSD (Agilent, Stockport, UK) coupled to an Agilent 6890 Gas Chromatograph. The GC was fitted with an Agilent DB-17 column (30 m x 0.25 mm internal diameter, 0.15 µm film). The injector port, source and transfer line temperatures were set at 230° C and an oven temperature program from 180° C (2.0 min) to 320° C (3.0 min) at 8° C / min was used. The flow rate of the helium carrier gas was set to a constant flow of 0.6 ml/min and mass spectral data were acquired for the duration of the GC program from m/z 50–800. Raw GC-MS data were analyzed with the AMDIS software package (http://chemdata.nist.gov/mass-spc/amdis/).

Metabolites from the OSC3-silenced plant lines, as well as from non-transformed plants were also extracted according to Field and Osbourn, (2008). GC-MS analyses of extracts were performed on an Agilent Technologies 6890N gas chromatograph, connected to an MSD 5975 and a GC/MSD ChemStation 61701DA.

Separations were carried out using an Ultra 1 column (J&W Scientific, Folsom, CA, USA; 30 m × 0.32 mm id; film thickness, 0.52 µm) and the following GC conditions: split injection of 1 µl of sample; flow rate of 1.5 $ml.min^{-1}$ (helium); temperature program: initial temperature (iT), 150°C (5 min); final temperature (fT), 300°C (25 min); rate, 8°C min^{-1}. The identification of lupeol was made by comparing the retention times and MS fragmentation patterns of the various components with those of an authentic sample (Sigma Chemical Co., St Louis, MO, USA) and / or by comparing the MS fragmentation pattern with those included in the database of the equipment.

B.12. DNA methylation assays - Bisulfite sequencing

Genomic DNA samples from wild type and *AMY2* stable transformed roots were extracted using the Nucleospin Plant II kit (Macherey-Nagel, Düren, Germany). Sodium bisulfite treatment of the DNA was conducted using the EpiTect Bisulfite kit (Qiagen, Hilden, Germany), following manufacturer's instructions. This procedure comprises of a few simple steps: bisulfite-mediated conversion of unmethylated cytosines, binding of the converted single-stranded DNA to the membrane of an EpiTect spin column, washing, desulfonation of membrane-bound DNA, washing of the membrane-bound DNA to remove desulfonation agent and elution of the pure, converted DNA from the spin column. Particularly, the protocol "Sodium Bisulfite Conversion of Unmethylated Cytosines in DNA from Low-Concentration Solution" in EpiTect Bisulfite Handbook was used.

The target regions of *LjCYP71D353* and *LjSDRt* (all in coding sequences, 280–350 bp long) were amplified using the HotStart® DNA Polymerase (Qiagen, Hilden, Germany) and the primer pairs described in Table A1. Amplified fragments were TA cloned using the pGEM-T-easy Vector System (Promega, Madison, WI, USA). Ten clones were sequenced for each amplicon to determine levels of methylation (percentage of all methylated deoxycytidine 5mdC in relation to the total deoxycytidine content in all 10 clones in mutant).

B.13. Phylogenetic analysis

The alignment and phylogenetic analysis of the cytochrome P450s and OSCs protein sequences were performed with the MEGA v5.05 software package (Tamura *et al*., 2011). The multiple alignment parameters were adjusted with gap cost 10 and gap extension 1. The phylogenetic trees were constructed using the neighbor-joining and maximum likelihood algorithm with bootstrap analysis of 1000 replicates.

The synteny of the specialized metabolic genes clusters of legumes and *A. thaliana* was examined with CoGe (Lyons and Freeling, 2008, http://genomevolution.org/CoGe/index.pl). For OSCs protein sequences, tblastn analysis was performed against *A. thaliana* (ncbi unmasked v1), *M. truncatula* (Medicago.org unmasked v3.5.1) and *L. japonicus* (pseudomolecules v2.5) genome databases with expectation value 0.001. Selected nucleotide sequences were analyzed with Gevo algorithm for syntenic genomic regions.

B.14. Promoter analysis

Genomic regions located about or 2 Kb upsteam of the start codon of each *L. japonicus* gene (*AMY2*, *LjCYP88D5*, *LjCYP71D353* and *LjSDRt*) in LjT11L01 clone, were examined for repeats and inverted repeats in Dotlet web page (http://myhits.isb-sib.ch/cgi-bin/dotlet). Prediction and recognition of plant regulatory motifs in putative promoter regions was conducted with NSITEM-PL(http://linux1.softberry.com/berry.phtml?topic=nsitemp&group=programs&subgroup=promoter), PLACE: a Database of Plant Cis-acting Regulatory DNA elements (http://www.dna.affrc.go.jp/PLACE/) and PlantPAN (http://plantpan.mbc.nctu.edu.tw/). Predicted motifs and transcription factors were inspected for their possible role and involvement in plant processes; they were compared with reported transcription factors in PlnTFTB 3.0 - Plant Transcription Factor Database (http://plntfdb.bio.uni-potsdam.de/v3.0/), PlantTFTB (*L. japonicus*

factors http://planttfdb.cbi.edu.cn/index.php?sp=Lja) and ProFITS (http://bioinfo.cau.edu.cn/ProFITS/).

B.15. Statistical analysis

All experiments were conducted at least twice and analyzed by ANOVA followed by Duncan multiple comparison tests ($a < 0.05$). Standard errors were calculated for all mean values and t-tests were performed for pairwise comparisons of means at different time points ($P \leq 0.01$).

B.16. Appendix 1

B.16.1. Table A1: Primers used in all experimental procedures described

Experimental procedure	Primer name	Sequence (5'>3')
***In situ* hybridization**	AMY2isF	GGACTCGAGTCTAGACAACAGGATATTACTGGAGTATACG
	AMY2isR	TTAGGTACCAAGCTTGAACTTGTTTGATTATTTTATTTGCATG
	LjCYP88D5isF	TGGTCGGAAACTGGAGGATGG
	LjCYP88D5R	AGTCTGCGAGGACACTCTTTAACC
	LjCYP71DisF	AACGCTGGCTACTTGTGATTAG
	LjCYP71DisR	CCTCAACATACCCTCTGCTACC
	OSC3isF	AGGACCTCAACACAAGAAG
	OSC3isR	AGAACTCTGGTGGAAGGG
Plant transformation	AMY2-2F	GGACTCGAGTCTAGACAACAGGATATTACTGGAGTATACG
	AMY2-2R	TTAGGTACCAAGCTTGAACTTGTTTGATTATTTTATTTGCATG
	AMY2-3F	GGACTCGAGTCTAGAGTACAGAAATAATTTTTTTACAACGATGG
	AMY2-3R	AAGGTACCAAGCTTGCAACAAACCGACACTAAATAC
	LjCYP88D5-1F	AAGACTCGAGTCTAGATGGTCGGAAACTGGAGGATGG
	LjCYP88D5-1R	TTAGGTACCAAGCTTAAGTCTGCGAGGACACTCTTACC
	LjCYP88D5-3F	CCGACTCGAGTCTAGAACCCACACATCTTGAATAAAGC
	LjCYP88D5-3R	AAAGGTACCAAGCTTTATTGGCACACCGCAACG
	LjCYP71D353RNAiF	GGACCATGGTACACCACCTCATTGGCTCC
	LjCYP71D353RNAiR	CGAGTCGACAGGACCAAGAAGCATATATC
	MtBasIRNAiF	ACGCCATGGTAAGCTATAAGCTATAAGCTATGAG
	MtBasIRNAiR	GCATAGGTAGAATATTTACTCGAGCGCT
	OSC3riF	GACTCGAGTCTAGATAGGACCTCAACACAAGAAG
	OSC3riR	AGGTACCAAGCTTAGAACTCTGGTGGAAGGG
RT-PCR	UBQF	ATGCAGATCTTTTGTGAAGAC
	UBQR	ACCACCACGGAAGACGGAG
	35S-F	TGTGATAACATGGTGGAGCA

	35S-R	GGTGATTTCAGCGTGTCCTC
	Hyg-F	GACCAATGCGGAGCATATACG
	Hyg-R	CAGCTTCGATGTAGGAGGGC
	HANantis	TTAGTCGAACATGAATAAACAAGG
	HANsense	TCAAACCAGCTAGAATTACTATTATG
Q-PCR	UBQrtF	TTCACCTTGTGCTCCGTCTTC
	UBQrtR	AACAACAGCACACACAGCCAATCC
	AMY2rtF	GCAGTTTAACTTGTAAAGATAGC
	AMY2rtR	GGCAACAAACCGACACTAAATAC
	LjCYP88D5rtF	TAGTGTTCTGGAAGTCAATGATG
	LjCYP88D5rtR	AGATGTGTGGGTGTTGTGTAAG
	LjCYP71DrtF	ACATTAAAGCCGTTCTTCAGGAC
	LjCYP71DrtR	CCTCAACATACCCTCTGCTACC
	LjSDRtF	CGCTTTGAGCCTAGAAAACAG
	LjSDRtR	GCTTGGCTTTTAAGAACCTGCT
	Mtc27F	TGAGGGAGCAACCAAATACC
	Mtc27R	GCGAAAACCAAGCTACCATC
	MtBasIF	ACGCCATGGTAAGCTATAAGCTATGA
	MtBasIR	AGCGCTCGAGTAAATATTCTACCTATGC
	OSC3rtF	TCAGAGGCTTACAAACTTCATTTAG
	OSC3rtR	GTTGGTCCATACTCTTCACTC
	Enod2F	CACAGGGCATCCGACTCC
	Enod2R	AGGTTCATAAATGGGTGGTTTCTC
	Enod40F	GTAGCAGAAGAAGCAGATGAGC
	Enod40R	ATAGTTTAGAGTGAGATTTAGCC
Cloning of full-length genes	LjCYP88D5Fl-F	ATGGAACTATACTGGGCTTGG
	LjCYP88D5Fl-R	TAATTACATGAAACCTTTATCACC
	LjCYP71DFl-F	TGCCCTTTTGCTAATGATGG
	LjCYP71DFl-R	TTATTCAACAGAAACAGGATTGTAAG
Heterologous expression in *Nicotiana benthamiana*	LjAMY2Fl-BstBi	AATATTCGAACATGTGGAAGCTGAAGGTAG
	LjAMY2Fl-StuI	TTATAGGCCTTGCACACAGCTATCTTTACAAG
	LjCYP88D5Fl-Bsu36i	TACACCTGAGGAATGGAACTATACTGGGCTTGG
	LjCYP88D5Fl-StuI	TTAGAGGCCTTTAATTACATGAAACCTTTATCACC
	LjCYP71DFl-Bsu36i	TACACCTGAGGTGCCCTTTTGCTAATGATGG
	LjCYP71DFl-StuI	TTATAGGCCTTTTTATTCAACAGAAACAGGATTGTAAG
Bisulfite sequencing	LjCYP71D-Bs-F	TAATCACTGCTCTCCCTCCC
	LjCYP71D-Bs-R	CAACACCTCTTTGGCAATTTC
	LjSDRt-Bs-F	CCCGAAAACTACTTTTTGCA
	LjSDRt-Bs-R	CGATTAGTATTCGCTTAAACC

Table A1. Pairs of primers used in experimental procedures.

B.16.2. Plasmid vectors utilized in the described experimental processes

B.16.2.1. pGEM-T easy ™: the vector for sequences determination

pGEM-T easy ™ vector is a linearized vector with small size (3.0 kb) and a single 3′-terminal thymidine at both ends. It offers bacterial ampicillin resistance and gives the opportunity of direct incorporation of DNA fragments via TA cloning. It contains an initiation of replication site and T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. Insertional inactivation of the α-peptide allows identification of recombinants by blue/white screening on indicator plates.

Figure A1. The pGEM-T easy cloning vector.

B.16.2.2. *Agrobacterium rhizogenes* plant transformation and Gateway technology

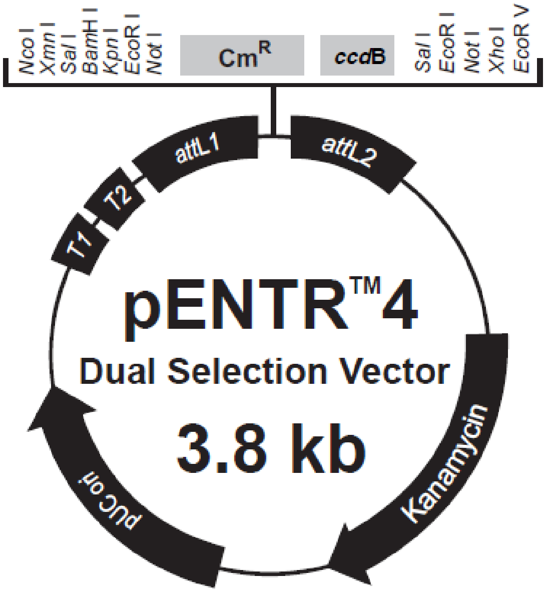
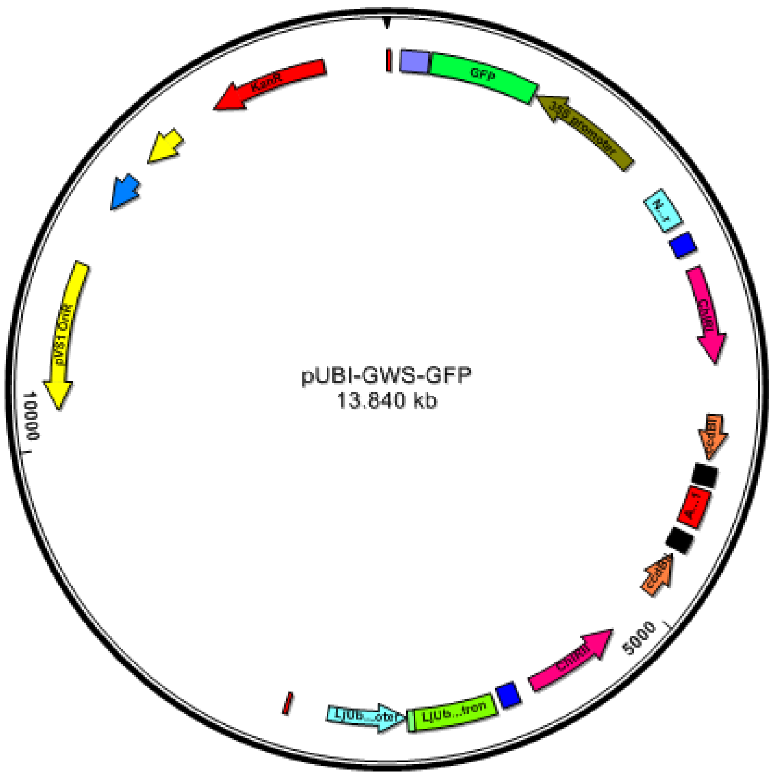
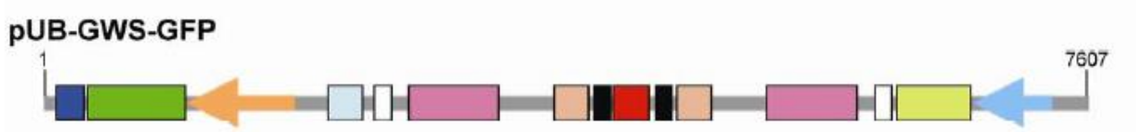


Figure A2. The entry pENTR™4 vector used for Gateway® Technology.





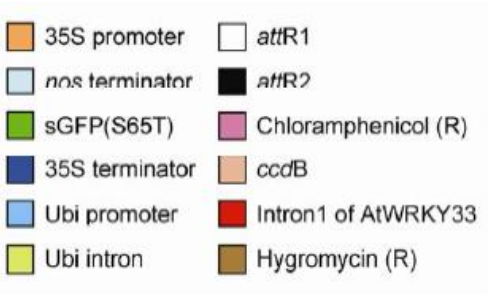


Figure A3. The destination vector pUBI-GWS-GFP used for Gateway® Technology (Maekawa *et al.*, 2008).

The Gateway® Technology is based on the bacteriophage lambda site-specific recombination system which facilitates the integration of lambda into the *E. coli* chromosome and the switch between the lytic and lysogenic pathways. In the Gateway® Technology, the components of the lambda recombination system are modified to improve the specificity and efficiency of the system. Lambda-based recombination involves two major components, the DNA recombination sequences (*att* sites) and the proteins that mediate the recombination reaction (*i.e.* Clonase™ enzyme mix). Lambda recombination is catalyzed by a mixture of enzymes that bind to these *att* sequences, bring together the target sites, cleave them, and covalently attach the DNA. Recombination occurs following two pairs of strand exchanges and ligation of the DNAs in a novel form. The recombination proteins involved in the reaction differ depending upon whether lambda utilizes the lytic or lysogenic pathway. The lysogenic pathway is catalyzed by the bacteriophage λ Integrase (Int) and *E. coli* Integration Host Factor (IHF) proteins (BP Clonase™ enzyme mix) while the lytic pathway is catalyzed by the bacteriophage λ Int and Excisionase (Xis) proteins, and the *E. coli* Integration Host Factor (IHF) protein (LR Clonase™ enzyme mix). Lambda recombination occurs between site-specific attachment (*att*) sites: *att*B on the *E. coli* chromosome and *att*P on the lambda chromosome. Upon lambda integration, recombination occurs between *att*B and *att*P sites to give rise to *att*L and *att*R sites. The actual crossover occurs between homologous 15 bp core regions on the two sites, but surrounding sequences are required as they contain the binding sites for the recombination proteins. The presence of the *ccd*B gene

allows negative selection of the donor and destination (and some entry) vectors in *E. coli* following recombination and transformation. The CcdB protein interferes with *E. coli* DNA gyrase, thereby inhibiting growth of most *E. coli* strains (*e.g.* DH5α, TOP10). When recombination occurs (*i.e.* between a destination vector and an entry clone or between a donor vector and an *att*B-PCR product), the *ccd*B gene is replaced by the gene of interest. Cells that take up unreacted vectors carrying the *ccd*B gene or by-product molecules retaining the *ccd*B gene will fail to grow. This allows high-efficiency recovery of the desired clones. The LR reaction used facilitates recombination of an *att*L substrate (entry clone, pENTR4) with an *att*R substrate (destination vector, pUBI-GWS-GFP) to create an *att*B-containing expression clone. This reaction is catalyzed by LR Clonase™ enzyme mix and is shown below.

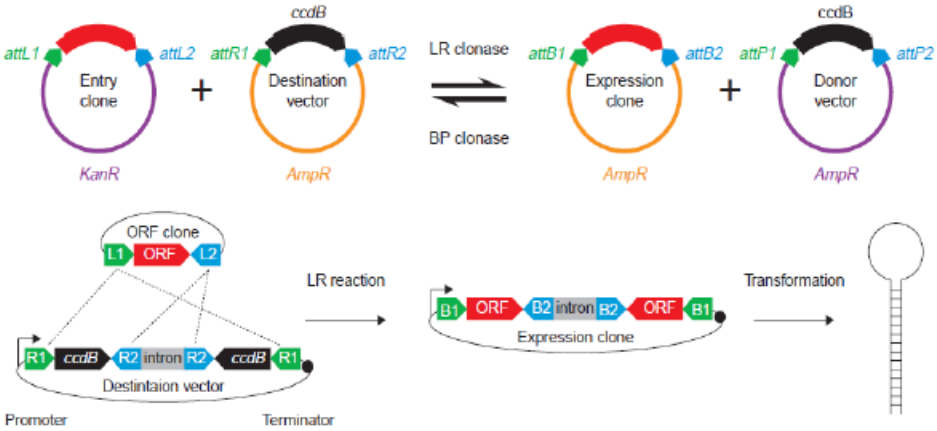


Figure A4. The LR recombination between the entry and the destination vectors.

B.16.2.3. *Agrobacterium tumefaciens* plant transformation

The pHannibal vector (~5.8 kb) is a hairpin vector for conventional cloning of PCR fragments into hairpin expression cassettes, with a bacterial ampicillin resistance marker gene. It contains a 35S promoter-driven inverted repeat cassette with restriction sites for directional cloning of fragments from the target gene on either side of an intron (880bp) of the pyruvate dehydrogenase kinase gene (*PDK*). It also contains a termination of transcription site (OCS terminator) at the end of the

second cloning site. The hairpin cassette described can be cloned into a binary vector using the appropriate restriction enzymes.

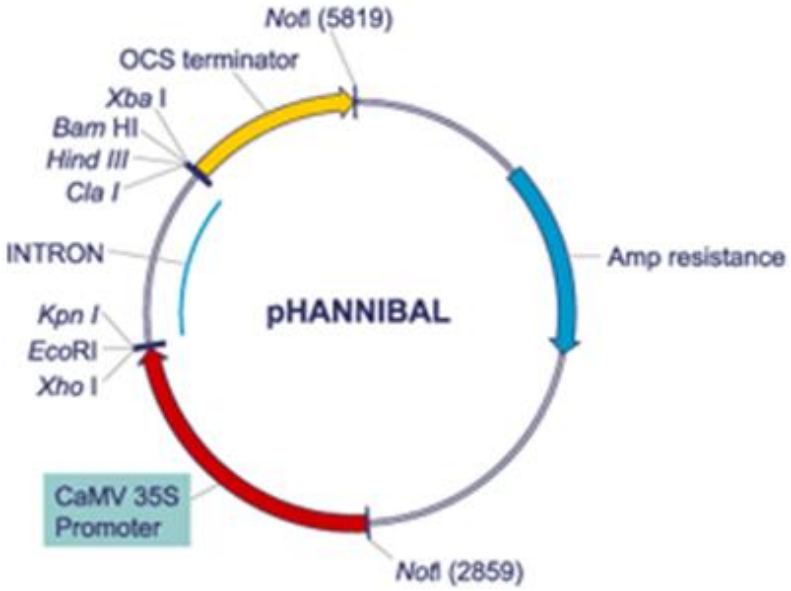


Figure A5. The pHannibal hairpin vector.

The pCambia 1300 vector a small size (~9.0 kb) T-DNA binary vector and its backbone is derived from the pPZP vectors. It carries the right and left borders that enable the transfer and incorporation of DNA regions into the plant genome. It offers high copy number in *E. coli* for high DNA yields and a bacterial selection with kanamycin. It contains a pVS1 replicon for high stability in *Agrobacterium* strains, restriction sites designed for modular plasmid modifications and small but adequate poly-linkers for introducing the gene of interest. Transformant plant lines are selected with the hygromycin B gene, which is driven by a double-enhancer version of the CaMV35S promoter and terminated by the CaMV35S polyA signal.

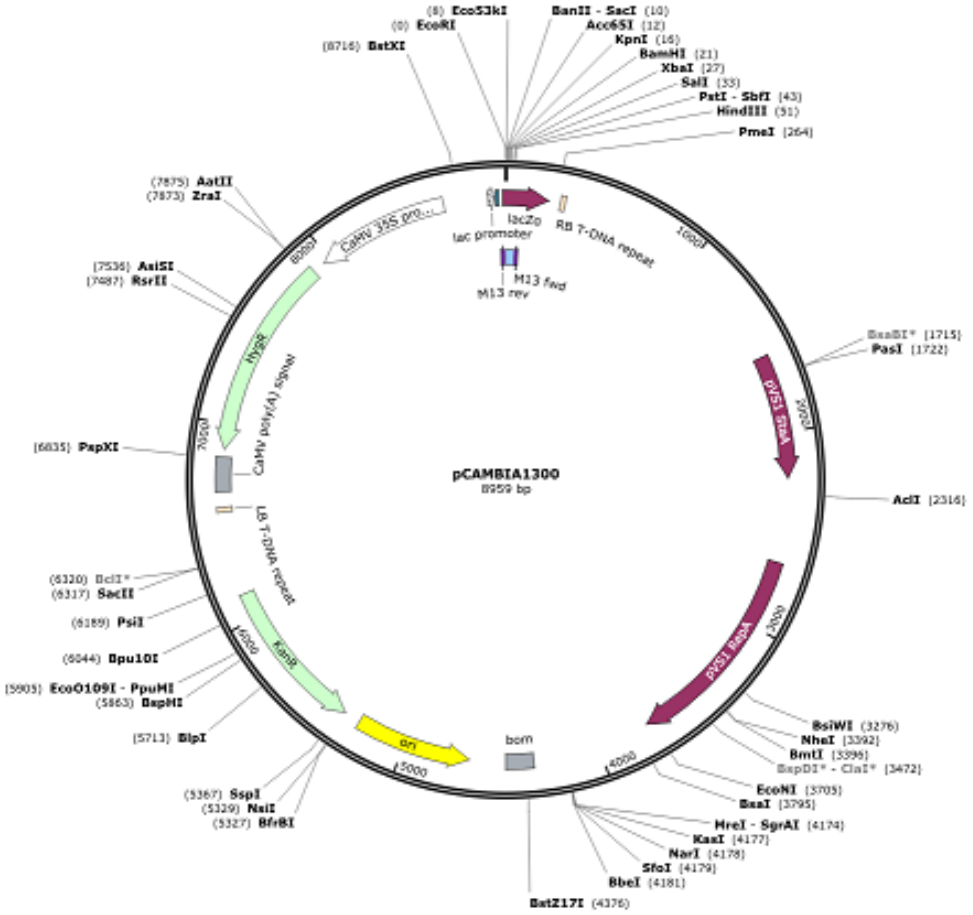


Figure A6. The binary vector pCambia 1300.

B.16.2.4. Heterologous expression in *N. Benthamiana*

The genome of cowpea mosaic virus (CPMV) consists of two separately encapsidated positive-strand RNA molecules of 5889 (RNA-1) and 3481 (RNA-2) nucleotides. The RNAs each contain a single open reading frame (ORF) and are expressed through the synthesis and subsequent processing of precursor polyproteins. RNA-1 encodes proteins involved in the replication of viral RNAs and polyprotein processing. RNA-2 encodes the 48K movement protein and the two coat proteins, which are essential for cell to cell movement and systemic spread (Sainsbury *et al.*, 2009). The development of CPMV-based systems for the expression of whole proteins has focused on modifying the sequence of RNA-2. Replication functions are provided by co-inoculating the RNA-2 constructs with unmodified RNA-1. Two types of system have been developed based on either full-length or deleted versions of RNA-2. In the first instance, the foreign gene to be

expressed is added to the normal complement of RNA-2 proteins, resulting in an increase in size of RNA-2. In the second case, the foreign protein replaces most of the RNA-2 ORF, allowing larger inserts to be incorporated but preventing the ability of the virus to spread both within and between plants. This second concept is based on the observation that the sequences necessary for replication of RNA-2 by the RNA-1-encoded replicase, lie exclusively at the 5´ and 3´ ends of the RNA. This fact permits most of the RNA-2 ORF to be deleted without affecting the ability of RNA-2 to be replicated (Sainsbury *et al*., 2009). To create a useful cloning vector for the expression of foreign proteins from a pBinP-1-GFP-based plasmid, the complete sequence of RNA-2 flanked by the Cauliflower mosaic virus (CaMV) 35S promoter and nopaline synthase (*nos*) terminator from pBinP-S2NT (Liu and Lomonossoff, 2002) was inserted into the mutagenesis plasmid pM81W (Liu and Lomonossoff, 2006) as an *AscI* / *PacI* fragment. The resulting plasmid, pM81W-S2NT, was subjected to a single round of mutagenesis which simultaneously introduced four changes to give pM81B-S2NT-1. The mutagenesis removed two *BspHI* sites from the vector backbone and introduced a *BspHI* site around AUG 512 and a *StuI* site after UAA 3299, the termination codon for the RNA-2-encoded polyprotein. Subsequently, the *BamHI* / *AscI* fragment was excised from pBinP-NS-1 and ligated into similarly digested pM81BS2NT-1, yielding pM81-FSC-1 (Liu and Lomonossoff, 2006).

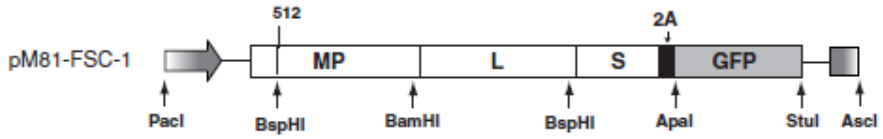


Figure A7. The pM81-FSC1 vector. MP movement protein, L large coat protein, S small coat protein (Sainsbury *et al*., 2009)

This vector allows the whole RNA-2 ORF downstream of AUG 512 to be digested with *BspHI* / *StuI* and replaced with any sequence with these compatible ends, and expresses the foreign gene in plants. The pM81-FSC-1-derived plasmid is digested with *AscI* / *PacI* and the fragment containing the foreign sequences transferred to similarly digested pBinD and the resulting binary plasmids are finally transformed

into *A. tumefaciens*. In this way, an *Agrobacterium* suspension of the deleted version of RNA-2 containing the inserted gene can be co-infiltrated into *N. benthamiana* leaves (Sainsbury *et al.*,2009) in the presence of suspensions of bacteria containing RNA-1 and the suppressor of silencing from *Potato virus Y* HC Pro (Voinnet *et al.*, 2003).

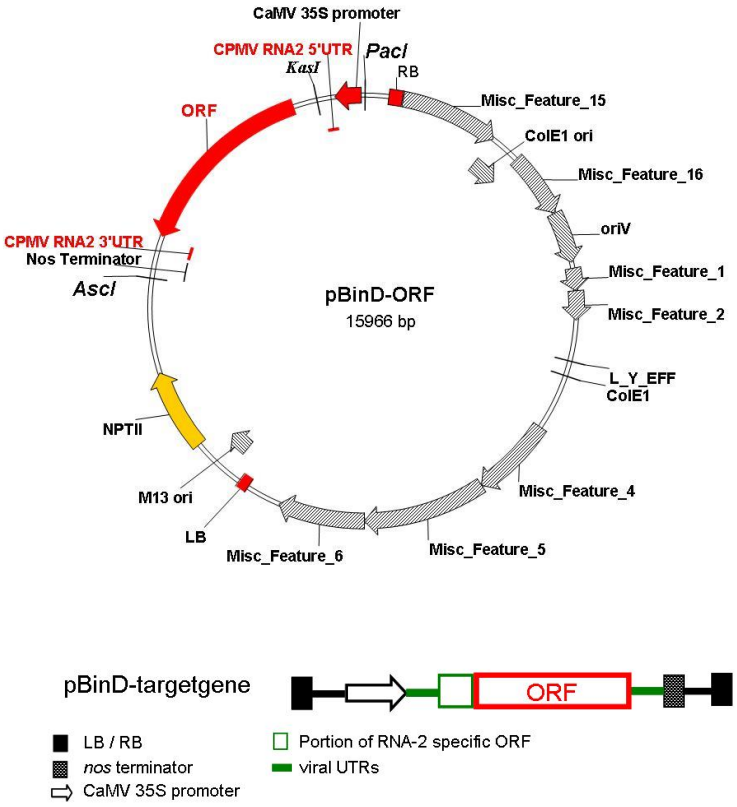
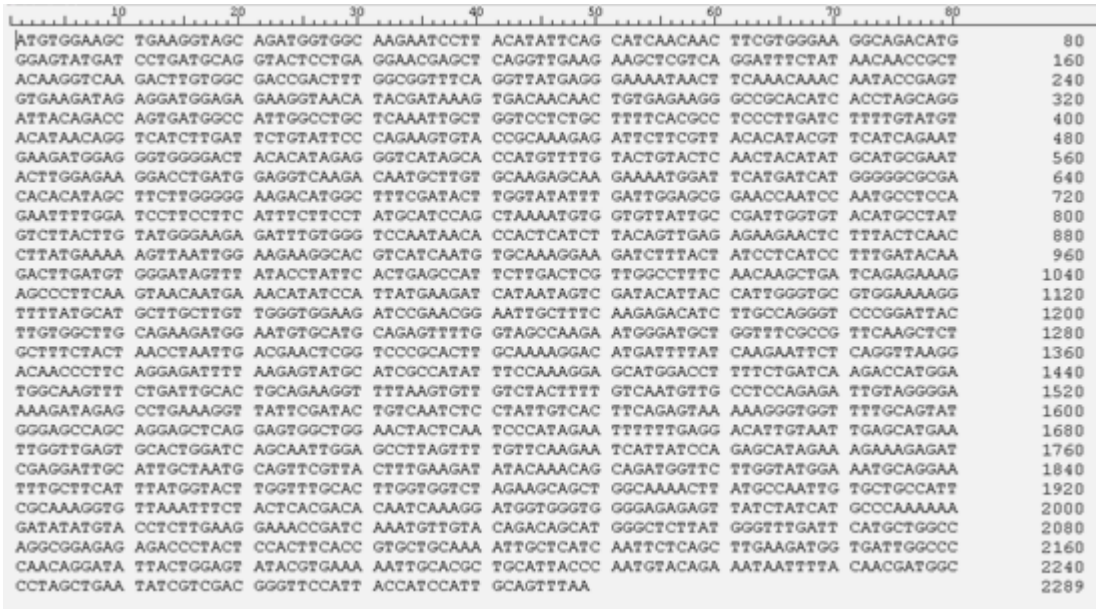


Figure A8. The binary vector pBinD containing the gene of interest.

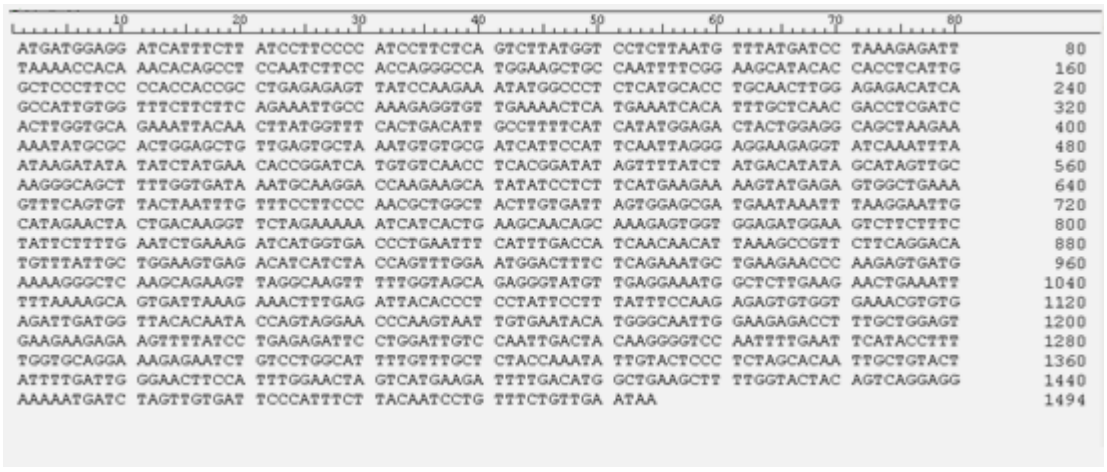
B.16.3. Sequences of the cluster genes

The "signature" *AMY2* gene in *L. japonicus* is flanked by genes encoding candidate "tailoring" enzymes. Particularly, the flanking genes *LjCYP71D353* and *LjCYP88D5* are genes for two different classes of CYP450, belonging to the CYP71D and CYP88D families, and the *LjSDRt* gene is predicted to encode a reductase. Their sequences are shown in the next figures (A9-A12). In addition, the *OSC3* gene of *L. japonicus* and *βAS* gene of *M. truncatula* gene used in experimental procedures are shown in figures A13 and A14 respectively.



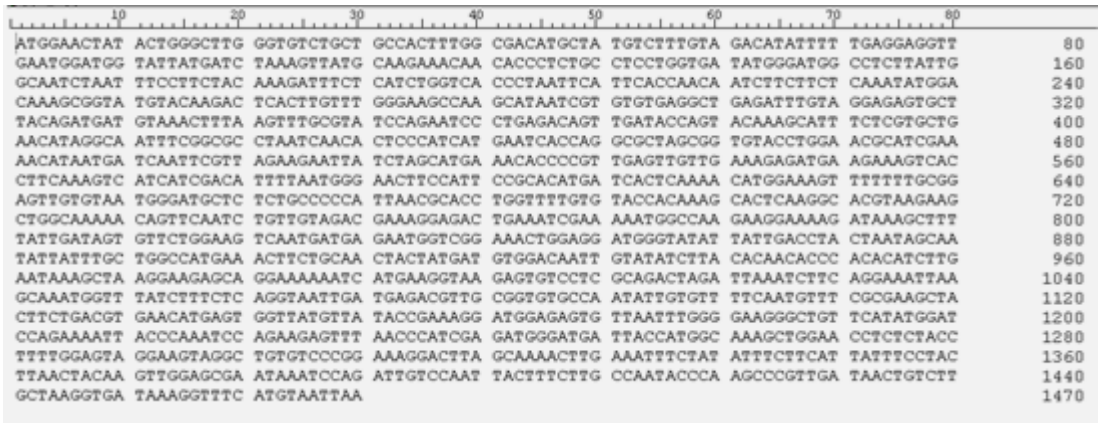
```
ATGTGGAAGC TGAAGGTAGC AGATGGTGGC AAGAATCCTT ACATATTCAG CATCAACAAC TTCGTGGGAA GGCAGACATG      80
GGAGTATGAT CCTGATGCAG GTACTCCTGA GGAACGAGCT CAGGTTGAAG AAGCTCGTCA GGATTTCTAT AACAACCGCT     160
ACAAGGTCAA GACTTGTGGC GACCGACTTT GGCGGTTTCA GGTTATGAGG GAAAATAACT TCAAACAAAC AATACCGAGT     240
GTGAAGATAG AGGATGGAGA GAAGGTAACA TACGATAAAG TGACAACAAC TGTGAGAAGG GCCGCACATC ACCTAGCAGG     320
ATTACAGACC AGTGATGGCC ATTGGCCTGC TCAAATTGCT GGTCCTCTGC TTTTCACGCC TCCCTTGATC TTTTGTATGT     400
ACATAACAGG TCATCTTGAT TCTGTATTCC CAGAAGTGTA CCGCAAAGAG ATTCTTCGTT ACACATACGT TCATCAGAAT     480
GAAGATGGAG GGTGGGGACT ACACATAGAG GGTCATAGCA CCATGTTTTG TACTGTACTC AACTACATAT GCATGCGAAT     560
ACTTGGAGAA GGACCTGATG GAGGTCAAGA CAATGCTTGT GCAAGAGCAA GAAAATGGAT TCATGATCAT GGGGGCGCGA     640
CACACATAGC TTCTTGGGGG AAGACATGGC TTTCGATACT TGGTATATTT GATTGGAGCG GAACCAATCC AATGCCTCCA     720
GAATTTTGGA TCCTTCCTTC ATTTCTTCCT ATGCATCCAG CTAAAATGTG GTGTTATTGC CGATTGGTGT ACATGCCTAT     800
GTCTTACTTG TATGGGAAGA GATTTGTGGG TCCAATAACA CCACTCATCT TACAGTTGAG AGAAGAACTC TTTACTCAAC     880
CTTATGAAAA AGTTAATTGG AAGAAGGCAC GTCATCAATG TGCAAAGGAA GATCTTTACT ATCCTCATCC TTTGATACAA     960
GACTTGATGT GGGATAGTTT ATACCTATTC ACTGAGCCAT TCTTGACTCG TTGGCCTTTC AACAAGCTGA TCAGAGAAAG    1040
AGCCCTTCAA GTAACAATGA AACATATCCA TTATGAAGAT CATAATAGTC GATACATTAC CATTGGGTGC GTGGAAAAGG    1120
TTTTATGCAT GCTTGCTTGT TGGGTGGAAG ATCCGAACGG AATTGCTTTC AAGAGACATC TTGCCAGGGT CCCGGATTAC    1200
TTGTGGCTTG CAGAAGATGG AATGTGCATG CAGAGTTTTG GTAGCCAAGA ATGGGATGCT GGTTTCGCCG TTCAAGCTCT    1280
GCTTTCTACT AACCTAATTG ACGAACTCGG TCCCGCACTT GCAAAAGGAC ATGATTTTAT CAAGAATTCT CAGGTTAAGG    1360
ACAACCCTTC AGGAGATTTT AAGAGTATGC ATCGCCATAT TTCCAAAGGA GCATGGACCT TTTCTGATCA AGACCATGGA    1440
TGGCAAGTTT CTGATTGCAC TGCAGAAGGT TTTAAGTGTT GTCTACTTTT GTCAATGTTG CCTCCAGAGA TTGTAGGGGA    1520
AAAGATAGAG CCTGAAAGGT TATTCGATAC TGTCAATCTC CTATTGTCAC TTCAGAGTAA AAAGGGTGGT TTTGCAGTAT    1600
GGGAGCCAGC AGGAGCTCAG GAGTGGCTGG AACTACTCAA TCCCATAGAA TTTTTTGAGG ACATTGTAAT TGAGCATGAA    1680
TTGGTTGAGT GCACTGGATC AGCAATTGGA GCCTTAGTTT TGTTCAAGAA TCATTATCCA GAGCATAGAA AGAAAGAGAT    1760
CGAGGATTGC ATTGCTAATG CAGTTCGTTA CTTTGAAGAT ATACAAACAG CAGATGGTTC TTGGTATGGA AATGCAGGAA    1840
TTTGCTTCAT TTATGGTACT TGGTTTGCAC TTGGTGGTCT AGAAGCAGCT GGCAAAACTT ATGCCAATTG TGCTGCCATT    1920
CGCAAAGGTG TTAAATTTCT ACTCACGACA CAATCAAAGG ATGGTGGGTG GGGAGAGAGT TATCTATCAT GCCCAAAAAA    2000
GATATATGTA CCTCTTGAAG GAAACCGATC AAATGTTGTA CAGACAGCAT GGGCTCTTAT GGGTTTGATT CATGCTGGCC    2080
AGGCGGAGAG AGACCCTACT CCACTTCACC GTGCTGCAAA ATTGCTCATC AATTCTCAGC TTGAAGATGG TGATTGGCCC    2160
CAACAGGATA TTACTGGAGT ATACGTGAAA AATTGCACGC TGCATTACCC AATGTACAGA AATAATTTTA CAACGATGGC    2240
CCTAGCTGAA TATCGTCGAC GGGTTCCATT ACCATCCATT GCAGTTTAA                                     2289
```

Figure A9. The sequence of *AMY2* gene.



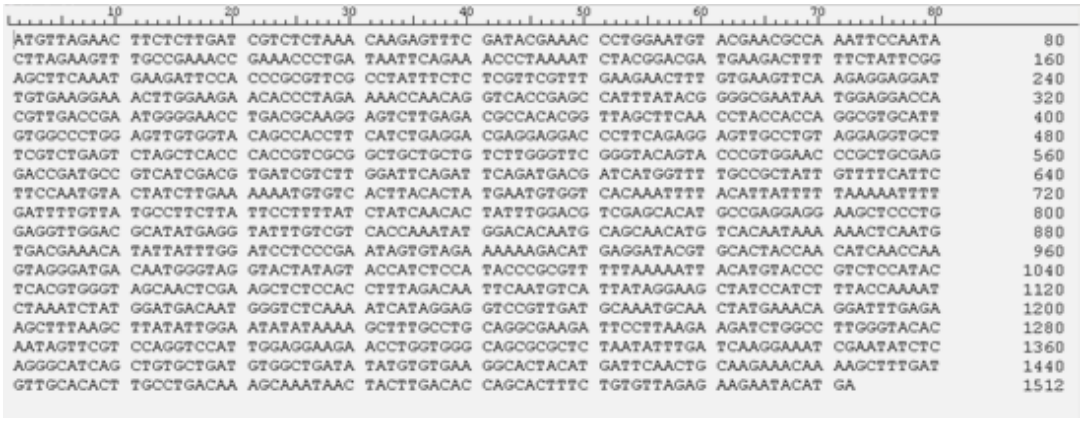
```
ATGATGGAGG ATCATTTCTT ATCCTTCCCC ATCCTTCTCA GTCTTATGGT CCTCTTAATG TTTATGATCC TAAAGAGATT      80
TAAAACCACA AACACAGCCT CCAATCTTCC ACCAGGGCCA TGGAAGCTGC CAATTTTCGG AAGCATACAC CACCTCATTG     160
GCTCCCTTCC CCACCACCGC CTGAGAGAGT TATCCAAGAA ATATGGCCCT CTCATGCACC TGCAACTTGG AGAGACATCA     240
GCCATTGTGG TTTCTTCTTC AGAAATTGCC AAAGAGGTGT TGAAAACTCA TGAAATCACA TTTGCTCAAC GACCTCGATC     320
ACTTGGTGCA GAAATTACAA CTTATGGTTT CACTGACATT GCCTTTTCAT CATATGGAGA CTACTGGAGG CAGCTAAGAA     400
AAATATGCGC ACTGGAGCTG TTGAGTGCTA AATGTGTGCG ATCATTCCAT TCAATTAGGG AGGAAGAGGT ATCAAATTTA     480
ATAAGATATA TATCTATGAA CACCGGATCA TGTGTCAACC TCACGGATAT AGTTTTATCT ATGACATATA GCATAGTTGC     560
AAGGGCAGCT TTTGGTGATA AATGCAAGGA CCAAGAAGCA TATATCCTCT TCATGAAGAA AAGTATGAGA GTGGCTGAAA     640
GTTTCAGTGT TACTAATTTG TTTCCTTCCC AACGCTGGCT ACTTGTGATT AGTGGAGCGA TGAATAAATT TAAGGAATTG     720
CATAGAACTA CTGACAAGGT TCTAGAAAAA ATCATCACTG AAGCAACAGC AAAGAGTGGT GGAGATGGAA GTCTTCTTTC     800
TATTCTTTTG AATCTGAAAG ATCATGGTGA CCCTGAATTT CATTTGACCA TCAACAACAT TAAAGCCGTT CTTCAGGACA     880
TGTTTATTGC TGGAAGTGAG ACATCATCTA CCAGTTTGGA ATGGACTTTC TCAGAAATGC TGAAGAACCC AAGAGTGATG     960
AAAAGGGCTC AAGCAGAAGT TAGGCAAGTT TTTGGTAGCA GAGGGTATGT TGAGGAAATG GCTCTTGAAG AACTGAAATT    1040
TTTAAAAGCA GTGATTAAAG AAACTTTGAG ATTACACCCT CCTATTCCTT TATTTCCAAG AGAGTGTGGT GAAACGTGTG    1120
AGATTGATGG TTACACAATA CCAGTAGGAA CCCAAGTAAT TGTGAATACA TGGGCAATTG GAAGAGACCT TTGCTGGAGT    1200
GAAGAAGAGA AGTTTTATCC TGAGAGATTC CTGGATTGTC CAATTGACTA CAAGGGGTCC AATTTTGAAT TCATACCTTT    1280
TGGTGCAGGA AAGAGAATCT GTCCTGGCAT TTTGTTTGCT CTACCAAATA TTGTACTCCC TCTAGCACAA TTGCTGTACT    1360
ATTTTGATTG GGAACTTCCA TTTGGAACTA GTCATGAAGA TTTTGACATG GCTGAAGCTT TTGGTACTAC AGTCAGGAGG    1440
AAAAATGATC TAGTTGTGAT TCCCATTTCT TACAATCCTG TTTCTGTTGA ATAA                               1494
```

Figure A10. The sequence of *LjCYP71D353* gene.



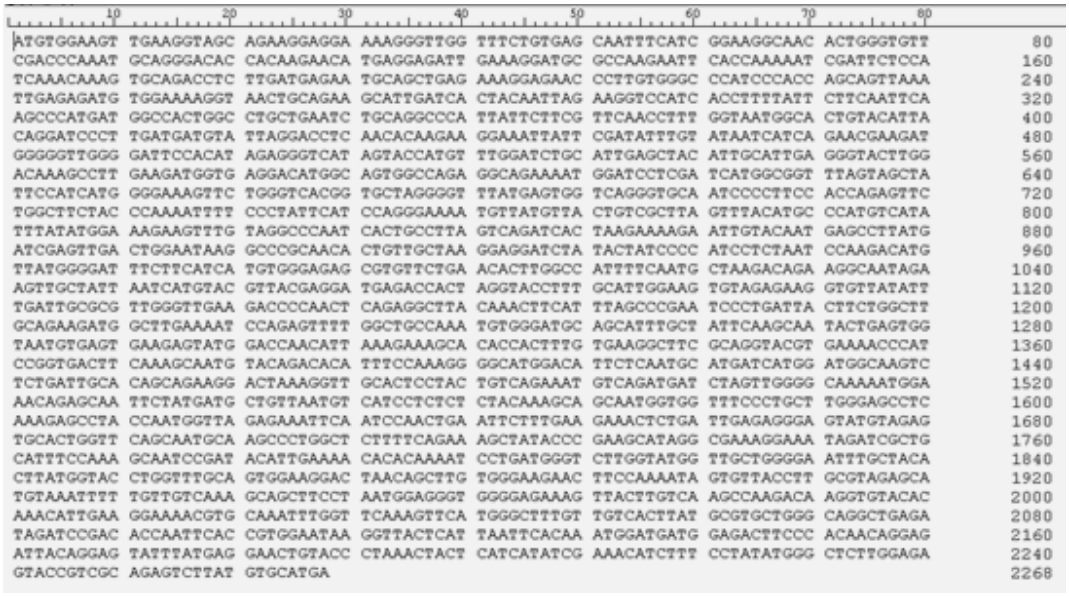
```
ATGGAACTAT ACTGGGCTTG GGTGTCTGCT GCCACTTTGG CGACATGCTA TGTCTTTGTA GACATATTTT TGAGGAGGTT      80
GAATGGATGG TATTATGATC TAAAGTTATG CAAGAAACAA CACCCTCTGC CTCCTGGTGA TATGGGATGG CCTCTTATTG     160
GCAATCTAAT TTCCTTCTAC AAAGATTTCT CATCTGGTCA CCCTAATTCA TTCACCAACA ATCTTCTTCT CAAATATGGA     240
CAAAGCGGTA TGTACAAGAC TCACTTGTTT GGGAAGCCAA GCATAATCGT GTGTGAGGCT GAGATTTGTA GGAGAGTGCT     320
TACAGATGAT GTAAACTTTA AGTTTGCGTA TCCAGAATCC CTGAGACAGT TGATACCAGT ACAAAGCATT TCTCGTGCTG     400
AACATAGGCA ATTTCGGCGC CTAATCAACA CTCCCATCAT GAATCACCAG GCGCTAGCGG TGTACCTGGA ACGCATCGAA     480
AACATAATGA TCAATTCGTT AGAAGAATTA TCTAGCATGA AACACCCCGT TGAGTTGTTG AAAGAGATGA AGAAAGTCAC     560
CTTCAAAGTC ATCATCGACA TTTTAATGGG AACTTCCATT CCGCACATGA TCACTCAAAA CATGGAAAGT TTTTTTGCGG     640
AGTTGTGTAA TGGGATGCTC TCTGCCCCCA TTAACGCACC TGGTTTTGTG TACCACAAAG CACTCAAGGC ACGTAAGAAG     720
CTGGCAAAAA CAGTTCAATC TGTTGTAGAC GAAAGGAGAC TGAAATCGAA AAATGGCCAA GAAGGAAAAG ATAAAGCTTT     800
TATTGATAGT GTTCTGGAAG TCAATGATGA GAATGGTCGG AAACTGGAGG ATGGGTATAT TATTGACCTA CTAATAGCAA     880
TATTATTTGC TGGCCATGAA ACTTCTGCAA CTACTATGAT GTGGACAATT GTATATCTTA CACAACACCC ACACATCTTG     960
AATAAAGCTA AGGAAGAGCA GGAAAAAAATC ATGAAGGTAA GAGTGTCCTC GCAGACTAGA TTAAATCTTC AGGAAATTAA    1040
GCAAATGGTT TATCTTTCTC AGGTAATTGA TGAGACGTTG CGGTGTGCCA ATATTGTGTT TTCAATGTTT CGCGAAGCTA    1120
CTTCTGACGT GAACATGAGT GGTTATGTTA TACCGAAAGG ATGGAGAGTG TTAATTTGGG GAAGGGCTGT TCATATGGAT    1200
CCAGAAAATT ACCCAAATCC AGAAGAGTTT AACCCATCGA GATGGGATGA TTACCATGGC AAAGCTGGAA CCTCTCTACC    1280
TTTTGGAGTA GGAAGTAGGC TGTGTCCCGG AAAGGACTTA GCAAAACTTG AAATTTCTAT ATTTCTTCAT TATTTCCTAC    1360
TTAACTACAA GTTGGAGCGA ATAAATCCAG ATTGTCCAAT TACTTTCTTG CCAATACCCA AGCCCGTTGA TAACTGTCTT    1440
GCTAAGGTGA TAAAGGTTTC ATGTAATTAA                                                          1470
```

Figure A11. The sequence of *LjCYP88D5* gene.



```
ATGTTAGAAC TTCTCTTGAT CGTCTCTAAA CAAGAGTTTC GATACGAAAC CCTGGAATGT ACGAACGCCA AATTCCAATA      80
CTTAGAAGTT TGCCGAAACC GAAACCCTGA TAATTCAGAA ACCCTAAAAT CTACGGACGA TGAAGACTTT TTCTATTCGG     160
AGCTTCAAAT GAAGATTCCA CCCGCGTTCG CCTATTTCTC TCGTTCGTTT GAAGAACTTT GTGAAGTTCA AGAGGAGGAT     240
TGTGAAGGAA ACTTGGAAGA ACACCCTAGA AAACCAACAG GTCACCGAGC CATTTATACG GGGCGAATAA TGGAGGACCA     320
CGTTGACCGA ATGGGGAACC TGACGCAAGG AGTCTTGAGA CGCCACACGG TTAGCTTCAA CCTACCACCA GGCGTGCATT     400
GTGGCCCTGG AGTTGTGGTA CAGCCACCTT CATCTGAGGA CGAGGAGGAC CCTTCAGAGG AGTTGCCTGT AGGAGGTGCT     480
TCGTCTGAGT CTAGCTCACC CACCGTCGCG GCTGCTGCTG TCTTGGGTTC GGGTACAGTA CCCGTGGAAC CCGCTGCGAG     560
GACCGATGCC GTCATCGACG TGATCGTCTT GGATTCAGAT TCAGATGACG ATCATGGTTT TGCCGCTATT GTTTTCATTC     640
TTCCAATGTA CTATCTTGAA AAAATGTGTC ACTTACACTA TGAATGTGGT CACAAATTTT ACATTATTTT TAAAAATTTT     720
GATTTTGTTA TGCCTTCTTA TTCCTTTTAT CTATCAACAC TATTTGGACG TCGAGCACAT GCCGAGGAGG AAGCTCCCTG     800
GAGGTTGGAC GCATATGAGG TATTTGTCGT CACCAAATAT GGACACAATG CAGCAACATG TCACAATAAA AAACTCAATG     880
TGACGAAACA TATTATTTGG ATCCTCCCGA ATAGTGTAGA AAAAAGACAT GAGGATACGT GCACTACCAA CATCAACCAA     960
GTAGGGATGA CAATGGGTAG GTACTATAGT ACCATCTCCA TACCCGCGTT TTTAAAAATT ACATGTACCC GTCTCCATAC    1040
TCACGTGGGT AGCAACTCGA AGCTCTCCAC CTTTAGACAA TTCAATGTCA TTATAGGAAG CTATCCATCT TTACCAAAAT    1120
CTAAATCTAT GGATGACAAT GGGTCTCAAA ATCATAGGAG GTCCGTTGAT GCAAATGCAA CTATGAAACA GGATTTGAGA    1200
AGCTTTAAGC TTATATTGGA ATATATAAAA GCTTTGCCTG CAGGCGAAGA TTCCTTAAGA AGATCTGGCC TTGGGTACAC    1280
AATAGTTCGT CCAGGTCCAT TGGAGGAAGA ACCTGGTGGG CAGCGCGCTC TAATATTTGA TCAAGGAAAT CGAATATCTC    1360
AGGGCATCAG CTGTGCTGAT GTGGCTGATA TATGTGTGAA GGCACTACAT GATTCAACTG CAAGAAACAA AAGCTTTGAT    1440
GTTGCACACT TGCCTGACAA AGCAAATAAC TACTTGACAC CAGCACTTTC TGTGTTAGAG AAGAATACAT GA            1512
```

Figure A12. The sequence of *LjSDRt* gene.



```
ATGTGGAAGT TGAAGGTAGC AGAAGGAGGA AAAGGGTTGG TTTCTGTGAG CAATTTCATC GGAAGGCAAC ACTGGGTGTT      80
CGACCCAAAT GCAGGGACAC CACAAGAACA TGAGGAGATT GAAAGGATGC GCCAAGAATT CACCAAAAAT CGATTCTCCA     160
TCAAACAAAG TGCAGACCTC TTGATGAGAA TGCAGCTGAG AAAGGAGAAC CCTTGTGGGC CCATCCCACC AGCAGTTAAA     240
TTGAGAGATG TGGAAAAGGT AACTGCAGAA GCATTGATCA CTACAATTAG AAGGTCCATC ACCTTTTATT CTTCAATTCA     320
AGCCCATGAT GGCCACTGGC CTGCTGAATC TGCAGGCCCA TTATTCTTCG TTCAACCTTT GGTAATGGCA CTGTACATTA     400
CAGGATCCCT TGATGATGTA TTAGGACCTC AACACAAGAA GGAAATTATT CGATATTTGT ATAATCATCA GAACGAAGAT     480
GGGGGTTGGG GATTCCACAT AGAGGGTCAT AGTACCATGT TTGGATCTGC ATTGAGCTAC ATTGCATTGA GGGTACTTGG     560
ACAAAGCCTT GAAGATGGTG AGGACATGGC AGTGGCCAGA GGCAGAAAAT GGATCCTCGA TCATGGCGGT TTAGTAGCTA     640
TTCCATCATG GGGAAAGTTC TGGGTCACGG TGCTAGGGGT TTATGAGTGG TCAGGGTGCA ATCCCCTTCC ACCAGAGTTC     720
TGGCTTCTAC CCAAAATTTT CCCTATTCAT CCAGGGAAAA TGTTATGTTA CTGTCGCTTA GTTTACATGC CCATGTCATA     800
TTTATATGGA AAGAAGTTTG TAGGCCCAAT CACTGCCTTA GTCAGATCAC TAAGAAAAGA ATTGTACAAT GAGCCTTATG     880
ATCGAGTTGA CTGGAATAAG GCCCGCAACA CTGTTGCTAA GGAGGATCTA TACTATCCCC ATCCTCTAAT CCAAGACATG     960
TTATGGGGAT TTCTTCATCA TGTGGGAGAG CGTGTTCTGA ACACTTGGCC ATTTTCAATG CTAAGACAGA AGGCAATAGA    1040
AGTTGCTATT AATCATGTAC GTTACGAGGA TGAGACCACT AGGTACCTTT GCATTGGAAG TGTAGAGAAG GTGTTATATT    1120
TGATTGCGCG TTGGGTTGAA GACCCCAACT CAGAGGCTTA CAAACTTCAT TTAGCCCGAA TCCCTGATTA CTTCTGGCTT    1200
GCAGAAGATG GCTTGAAAAT CCAGAGTTTT GGCTGCCAAA TGTGGGATGC AGCATTTGCT ATTCAAGCAA TACTGAGTGG    1280
TAATGTGAGT GAAGAGTATG GACCAACATT AAAGAAAGCA CACCACTTTG TGAAGGCTTC GCAGGTACGT GAAAACCCAT    1360
CCGGTGACTT CAAAGCAATG TACAGACACA TTTCCAAAGG GGCATGGACA TTCTCAATGC ATGATCATGG ATGGCAAGTC    1440
TCTGATTGCA CAGCAGAAGG ACTAAAGGTT GCACTCCTAC TGTCAGAAAT GTCAGATGAT CTAGTTGGGG CAAAAATGGA    1520
AACAGAGCAA TTCTATGATG CTGTTAATGT CATCCTCTCT CTACAAAGCA GCAATGGTGG TTTCCCTGCT TGGGAGCCTC    1600
AAAGAGCCTA CCAATGGTTA GAGAAATTCA ATCCAACTGA ATTCTTTGAA GAAACTCTGA TTGAGAGGGA GTATGTAGAG    1680
TGCACTGGTT CAGCAATGCA AGCCCTGGCT CTTTTCAGAA AGCTATACCC GAAGCATAGG CGAAAGGAAA TAGATCGCTG    1760
CATTTCCAAA GCAATCCGAT ACATTGAAAA CACACAAAAT CCTGATGGGT CTTGGTATGG TTGCTGGGGA ATTTGCTACA    1840
CTTATGGTAC CTGGTTTGCA GTGGAAGGAC TAACAGCTTG TGGGAAGAAC TTCCAAAATA GTGTTACCTT GCGTAGAGCA    1920
TGTAAATTTT TGTTGTCAAA GCAGCTTCCT AATGGAGGGT GGGGAGAAAG TTACTTGTCA AGCCAAGACA AGGTGTACAC    2000
AAACATTGAA GGAAAACGTG CAAATTTGGT TCAAAGTTCA TGGGCTTTGT TGTCACTTAT GCGTGCTGGG CAGGCTGAGA    2080
TAGATCCGAC ACCAATTCAC CGTGGAATAA GGTTACTCAT TAATTCACAA ATGGATGATG GAGACTTCCC ACAACAGGAG    2160
ATTACAGGAG TATTTATGAG GAACTGTACC CTAAACTACT CATCATATCG AAACATCTTT CCTATATGGG CTCTTGGAGA    2240
GTACCGTCGC AGAGTCTTAT GTGCATGA                                                            2268
```

Figure A13. The sequence of *OSC3* gene.

```
ATGTGGAAGC TGAAGATTGG AGAGGGAAAG AATGAGCCAT ACTTATTCAG CACAAATAAT TTTGTGGGAA GACAAACATG      80
GGAGTATGAT CCAGAAGCAG GTAGTGAAGA AGAAAGAGCA CAAGTTGAAG AAGCTCGTAA GAATTTCTAT GACAACAGAT     160
TCAAGGTCAA GCCCTGTGGT GACCTCCTAT GGCGTTTTCA GGTTCTAAGA GAAAATAACT TTATGCAAAC AATAGATGGA     240
GTGAAGATAG AAGATGGAGA AGAGATAACA TATGAGAAAG CAACGACAAC GTTGAGAAGG GGCACACACC ATCTAGCAGC     320
ATTGCAAACC AGTGATGGCC ATTGGCCTGC TCAAATTGCA GGTCCTCTAT TTTTCATGCC TCCCTTGGTT TTCTGTGTCT     400
ACATTACTGG ACATCTTGAT TCCGTCTTCC CACGAGAACA TCGCAAAGAG ATTCTTCGTT ACATTTACTG TCACCAAAAT     480
GAAGATGGAG GATGGGGGCT ACACATTGAG GGTCACAGCA CCATGTTTTG TACTGCACTT AACTATATAT GTATGCGAAT     560
TCTCGGAGAA GGACCTGATG GCGGTCAAGA CAATGCTTGT GCTAGAGCCA GAAACTGGAT TCGGGCACAC GGTGGTGTCA     640
CATATATACC TTCATGGGGA AAAACTTGGC TTTCGATACT TGGTCTCTTT GATTGGTTGG GAAGCAACCC AATGCCCCCT     720
GAGTTTTGGA TCCTTCCTTC ATTTCTTCCT ATGCATCCAG CTAAAATGTG GTGTTATTGT CGATTGGTAT ACATGCCTAT     800
GTCTTACTTG TACGGGAAGA GATTTGTGGG TCCGATCACA CCACTCATCT TACAGTTGAG AGAAGAACTC CATACTCAGC     880
CTTATGAAAA AATTAACTGG ACGAAATCAC GTCACCTATG TGCAAAGGAA GATATTTACT ATCCCCATCC TTTGATACAA     960
GATCTGATAT GGGATAGCTT ATACATATTT ACCGAGCCGC TTCTCACTCG CTGGCCTTTC AACAAGCTGG TCAGAAAAAG    1040
AGCCCTTGAA GTTACAATGA AGCATATCCA CTACGAGGAT GAGAACAGTC GATACCTAAC CATTGGGTGT GTGGAAAAGG    1120
TATTATGTAT GCTTGCTTGT TGGGTGGAAG ATCCAAATGG AGATGCTTAC AAGAAGCATC TTGCAAGGGT CCAAGATTAC    1200
TTGTGGATGT CAGAAGATGG AATGACCATG CAGAGTTTTG GTAGCCAAGA ATGGGATGCT GGTTTTGCCG TTCAAGCTTT    1280
GCTTGCCGCT AACCTAAATG ATGAAATCGA ACCTGCACTT GCCAAAGGAC ATGATTTCAT TAAGAAATCT CAGGTTACAG    1360
AGAACCCTTC TGGAGATTTT AAGAGTATGC ATCGTCATAT TTCTAAAGGC TCATGGACCT TCTCCGATCA AGACCATGGA    1440
TGGCAAGTTT CTGATTGCAC CGCTGAAGGT TTGAAGTGTT GTCTACTTTT ATCAATGTTG CCTCCAGAGA TTGTGGGGGA    1520
AAAGATGGAA CCAGAAAGGT TATATGATTC GGTCAATGTC TTGTTGTCGC TTCAGAGTAA AAAGGGTGGT TTGGCAGCAT    1600
GGGAGCCCGC AGGAGCTCAA GAGTGGTTAG AACTACTCAA TCCCACTGAG TTTTTTGCGG ACATTGTTGT TGAGCATGAA    1680
TATGTTGAGT GCACTGGATC AGCAATTCAA GCTTTAGTTT TGTTCAAGAA GCTATATCCA GGGCATAGGA AGAAAGAGAT    1760
AGAGAATTTC ATCTCCGAGG CAGTTCGATT CATTGAAGAT ATACAAACAG CCGATGGTTC ATGGTATGGA AACTGGGGAG    1840
TTTGCTTCAC TTATGGTTCT TGGTTTGCTC TTGGTGGTTT AGCAGCTGCT GGCAAGACTT ATACCAATTG CGCTGCTATT    1920
CGCAAAGCTG TTAAATTTCT TCTCACAACA CAGAGAGAGG ATGGTGGGTG GGGGGAGAGC TATCTTTCAA GCCCAAAAAA    2000
GATATATGTA CCTCTCGAAG GAAGCCGATC CAATGTTGTA CATACTGCAT GGGCTCTTAT GGGTTTAATT CATGCCGGCC    2080
AGGCAGAGAG AGACCCTACT CCTCTCCATC GTGCTGCAAA ATTGCTCATC AATTCCCAGT TGGAAGAAGG CGATTGGCCC    2160
CAACAGGAAA TCACAGGAGT ATTCATGAAA AATTGTATGT TGCATTACCC AATGTATAGA GATATTTACC CCTTGTGGGC    2240
TCTAGCCGAG TATCGTAGAC GGGTTCCATT GCCTTCCACT GCAGTTTAA                                     2289
```

Figure A14. The sequence of *βAS* gene.

C. RESULTS

CHAPTER C. RESULTS

C.1. The *AMY2* gene cluster

C.1.1. Genome organization of the *AMY2* cluster

An *in silico* analysis was carried out, using the public genome sequence databases for *L. japonicus* and *M. truncatula,* in order to investigate the genomic regions encompassing all predicted *OSC* genes present in the genomes of these two model plant species (Sawai *et al.*, 2006; Naoumkina *et al.*, 2010). Previously biochemically characterized and predicted *OSC* genes from *L. japonicus* (Iturbe-Ormaetxe *et al.*, 2003; Sawai *et al.*, 2006; Sato *et al.*, 2008), *M. truncatula* (Naoumkina *et al.*, 2010), *A. thaliana* and oat (Field and Osbourn, 2008; Field *et al.*, 2011; Qi *et al.*, 2004) were used as query sequences against the *L. japonicus* genome databases (all clones and contigs, *L. japonicus* genome assembly 1.0, http://www.kazusa.or.jp/lotus/) and the *M. truncatula* genome databases (pseudomolecule MT3.0, http://www.medicago.org/genome/cvit_blast.php), respectively. Sequences with expectation values of < 1 x e^{-10} were excluded from further analysis. A region of approximately 300 Kb flanking each side of the *OSC* genes was screened and analyzed using FGENESH gene prediction software (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) and GeneScan webserver (http://genes.mit.edu/GENSCAN.html) (Burge and Karlin, 1998). For all predicted Open Reading Frames (ORFs) and translated amino acid sequences (ExPASy translation tool, http://web.expasy.org/translate/) blast searches were run against *L. japonicus* and *M. truncatula* EST databases (http://compbio.dfci.harvard.edu/tgi/cgibin/tgi/gimain.pl?gudb=l_japonicus; http://compbio.dfci.harvard.edu/cgibin/tgi/gimain.pl?gudb=medicago) and Expasy connected protein databases of UniProtKB/Swiss-Prot (http://web.expasy.org/blast/). Confirmed amino acid sequences were aligned to biochemically characterized proteins, and thus genes with potential roles in secondary metabolism were identified (Table 1).

CLUSTER	*OSC*	*CYP*	*GDSL motif Lipase Hydrolase*	*Oxidoreductase*	*Epoxidase Hydrolase*
***L. japonicus AMY2* cluster**	AMY2 (Iturbe-Ormaetxe *et al.*, 2003)	LjCYP88D5 (72% similarity to GgCYP88D6 (Seki *et al.*, 2008). LjCYP71D353 (50% similarity to GmCYP71D9 (Latunde-Dada *et al.*, 2001)		LjSDRt (82% similarity to HCF173 (Schult *et al.*, 2007)	
***L. japonicus AMY2* 2nd cluster**	AMY2 interrupted (Iturbe-Ormaexte *et al.*, 2003)	LjCYP88D4 (71% similarity to GgCYP88D6 (Seki *et al.*, 2008); 80.7% similarity to LjCYP88D5, this study			
***M. truncatula bAS1* cluster**	BAS1 (Suzuki *et al.*, 2002)	MtCYP88D1 (60% similarity to GgCYP88D6 (Seki *et al.*, 2008); 58% similarity to LjCYP88D5		MtSDRt (83% similarity to HCF173 (Schult *et al.*, 2007); 80.7% similarity to LjSDRt	Epoxide Hydrolase (59% similarity to At3G05600)
***M. truncatula CAS* cluster**	CAS (81% similarity to *A. thaliana* CAS1)	MtCYP97B (78% similarity to LUT1 (Tian *et al.*, 2004)	GDSL esterase/lipase (83% similarity to At4G10950GNH hydrolase-type esterase superfamily, 54% similarity to CLIP7 (Oh *et al.*, 2005)	Oxidoreductase (67% similarity to At4G10020, HSDS; 50% similarity to MtSDRt)	

Table 1. Predicted polypeptides encoded by the genes present in the four clusters identified in *L. japonicus* and *M. truncatula* genomes and similarities with characterized proteins and enzymes in legumes and other plants.

A total of four genomic regions were identified, in which genes potentially implicated in triterpenoid / sterol biosynthesis are assembled in cluster formation. Two of these regions are located on chromosome 3 of *L. japonicus*, and the other on chromosomes 4 and 5 of *M. truncatula* (Figure 1). The first gene cluster of *L. japonicus* consists of genes encoding the previously characterized AMY2 enzyme (Iturbe-Ormaetxe *et al.*, 2003) and two cytochrome P450 enzymes, LjCYP88D5 and LjCYP71D353 (Figure 1b). The two cytochrome P450 genes are adjacent to each other, 60Kb from *AMY2*, and probably share a 2 kb common regulatory region. A fourth gene (*LjSDRt*), predicted to encode an enzyme with extensive similarity (82%) to an *A. thaliana* short chain dehydrogenase/reductase (Q9FWQ6, At1G16720), is also located in this region, in between *AMY2* and the two cytochrome P450 genes (Figure 1b, Table

1). The second putative cluster in *L. japonicus* is located approximately 200 Kb from the first and consists of just two genes: one that is highly similar to *AMY2* (99% similarity for the deduced amino acid sequences) and a cytochrome P450 gene *LjCYP88D4* that is very similar to *LjCYP88D5* (97% similarity for the coding amino acid sequences) (Figure 1b).

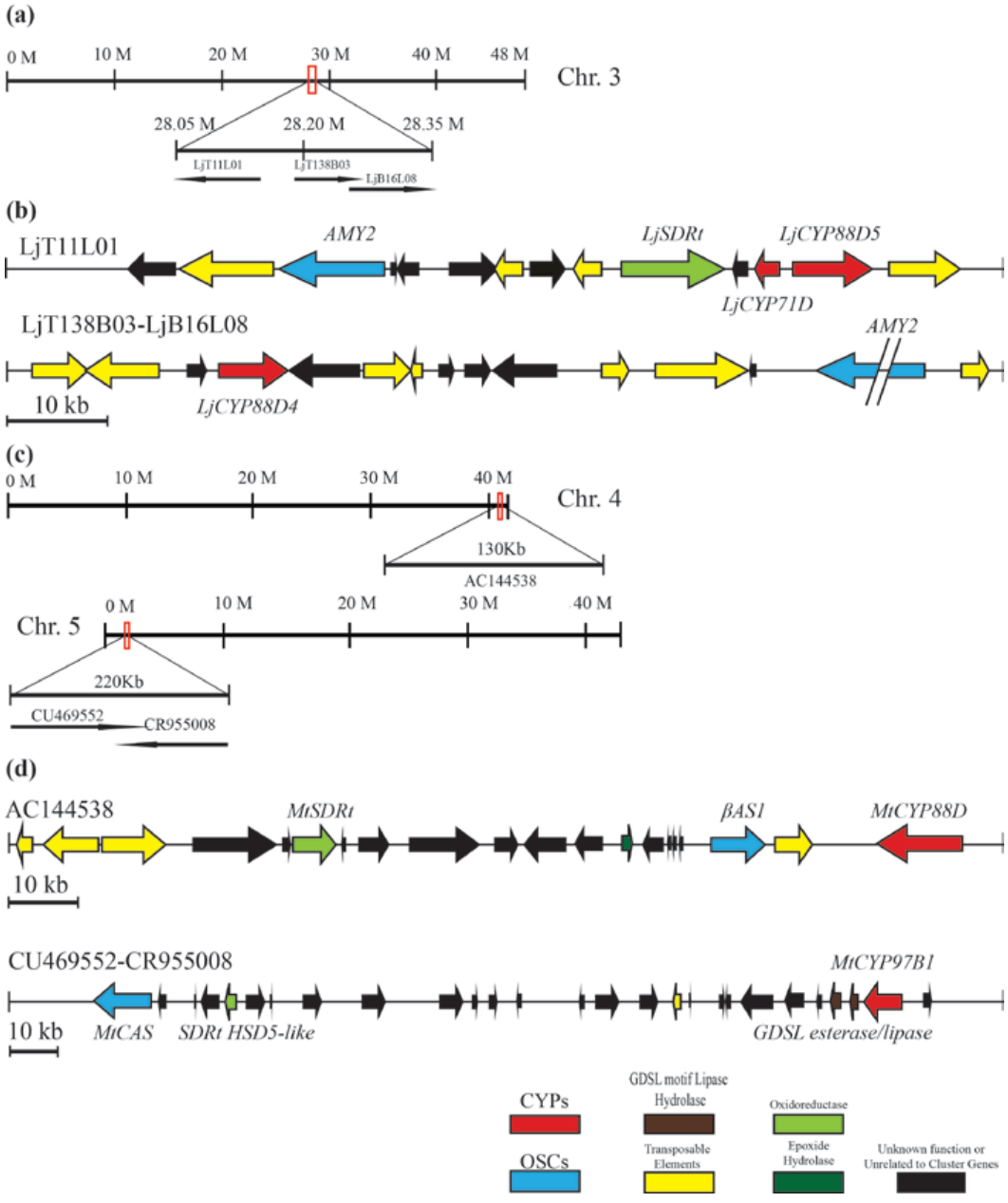


Figure 1. Structure of the genomic regions containing *OSCs* are flanked by genes putatively involved in triterpene / sterol metabolism in legumes. (a), (c) Maps of candidate gene clusters on chromosome 3 of *L. japonicus* and chromosome 4, 5 of *M. truncatula*, analyzing *L. japonicus* LjT11L01 and the continuous LjT138B03-LjB16L08 and *M. truncatula* AC144538 and the continuous CU469552-CR955008 genomic clones (b), (d) Organization of genes in *OSCs* flanking genomic regions.

Interestingly, this second *AMY2*-like gene copy appears to be interrupted at amino acid residue 228 and has an insertion of approximately 5 Kb of a gene coding for a predicted U-box protein (FGENESH publicly available software: http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind). The distance between the *AMY2*-like gene and *LjCYP88D4* is approximately 70 Kb. No other genes with predicted functions in secondary metabolism were identified in this region.

In *M. truncatula*, one of the gene clusters is located on chromosome 4 (Figure 1c) and contains the previously reported β-amyrin synthase gene, *βAS1* gene (Suzuki *et al.*, 2002). A cytochrome P450 gene *MtCYP88D1* is adjacent to *βAS1,* together with a short chain dehydrogenase / reductase gene (*MtSDRt)* that shares 80.7% amino acid sequence similarity with the respective *LjSDRt* gene present in the *AMY2* gene cluster. Genes encoding other putative tailoring enzymes, a methyltransferase and an epoxide hydrolase are also present in this region (Figure 1d, Table 1). A second candidate region in *M. truncatula* (Figure 1d) is located on chromosome 5 and consists of a putative cycloartenol synthase gene *MtCAS*. MtCAS shares 81% amino acid sequence similarity compared to *A. thaliana* cycloartenol synthase CAS1, 90% similarity *to L. japonicus* OSC5 and 82% similarity to *Betula platyphylla* BPX2 (Corey *et al.*, 1993), i.e. a potential enzyme involved in sterol biosynthesis. This gene is also flanked by an oxidoreductase gene, which shares 50% amino acid similarity with the predicted MtSDRt present in the *βAS1* gene cluster. The only cytochrome P450 gene in this region is *MtCYP97B*, which is located ~160Kb away from *MtCAS1* (Figure 1d). A common feature of all four candidate clusters is the presence of several transposable elements in the proximity of the OSC and cytochrome P450 genes (Figure 1b, d).

The cytochrome P450 gene products identified in the four clusters were subjected to phylogenetic analysis by comparison with cytochrome P450 gene products that: (a) have been shown to be involved in triterpene biosynthesis, and / or (b) have been reported to be co-expressed with *OSC* genes in transcriptome analysis of plants with known genome sequence (Qi *et al.*, 2006; Ehlting *et al.*, 2008; Field and Osbourn, 2008; Field *et al.*, 2011) (Figure 2b). LjCYP88D4/5 and MtCYP88D1 belong to the Fabaceae-

specific CYP88D subfamily (CYP85 clan), which has been implicated in triterpene biosynthesis, and exhibit 72% and 60% similarity respectively with GgCYP88D6, a licorice β-amyrin 11-oxidase (Seki *et al*., 2008; Nelson and Werck-Reichhart, 2011). LjCYP71D353 enzyme belongs to the CYP71 clan of P450s and is phylogenetically close to a CYP71A16 from *A. thaliana,* which is a member of the gene cluster for the marneral pathway (Field *et al*., 2011). *A. thaliana* CYP71A16 is a marneral oxidase.

The *M. truncatula* putative cycloartenol cyclase MtCAS and associated cytochrome P450 MtCYP97B1 were also included in this phylogenetic analysis. MtCYP97B1 has 73% similarity with the AtCYP97B3 gene product, which has been implicated in carotenoid biosynthesis in *A. thaliana* (Kim *et al*., 2010). There are evidence that CYP97B3 plays a role in the hydroxylation of the b-ring of a-carotene and b-carotene in plants (Kim *et al*., 2010). Interestingly, *AtCYP97B3* also lies within a candidate triterpene metabolic gene cluster in *A. thaliana,* containing the *BARS1 / PEN2 OSC* genes for synthesis of baruol and arabidiol (Husselstein-Muller *et al*., 2001; Ehlting *et al*., 2008; Field *et al*., 2011). However, it remains unidentified if *CYP97B1* encodes an enzyme involved in the modification of cycloartenol or other triterpenes.

Phylogenetic analysis was also performed for the predicted SDR-like LjSDRt (Figure 2c). BLAST analyses (Altschul *et al*., 1994) indicated several LjSDRt/HCF173 homologs in many plants, including members of the green algae, although none of the homolog proteins in other plants has been characterized as yet.

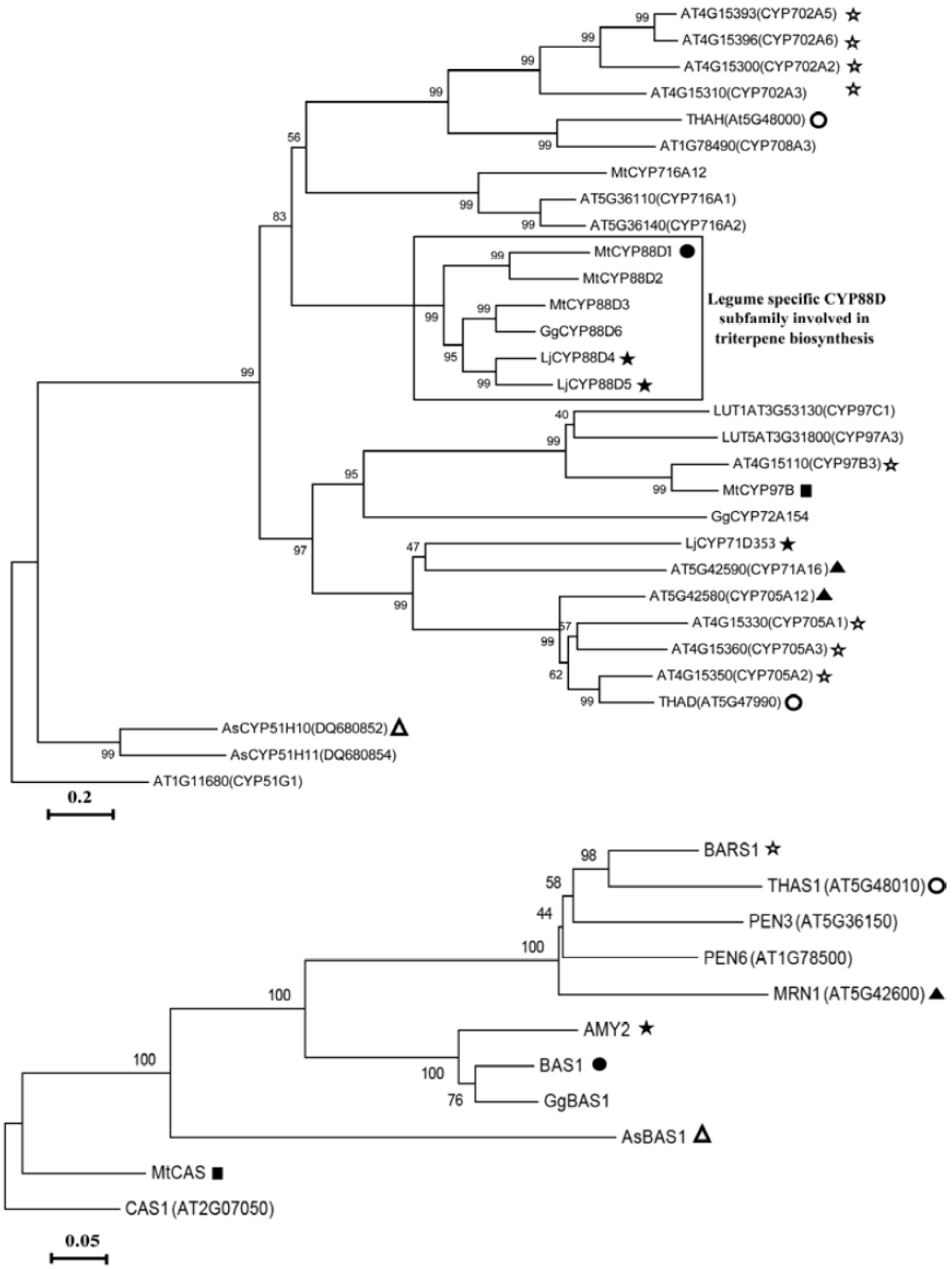


Figure 2b. Phylogenetic trees of cytochrome P450s and OSCs proteins constructed by the neighbour-joining method with a 1000 bootstrap replicates. The scale bar indicates 0.2 substitutions per site. Cytochrome P450s from *A. thaliana, M. sativa*, *L. japonicus, G. glabra* and *A. sativa* adjacent to/highly co-expressed with OSCs or previously found to participate in triterpene biosynthesis were used for the phylogenetic analysis. The open and black stars, open and black circles, open and black triangles and the black square indicate the cytochrome P450s clustered together with the BARS1, AMY2, THAS1, BAS1, AsBAS1, MRN1 and MtCAS1, respectively. The black box indicates the legume specific cytochrome P450 subfamily.

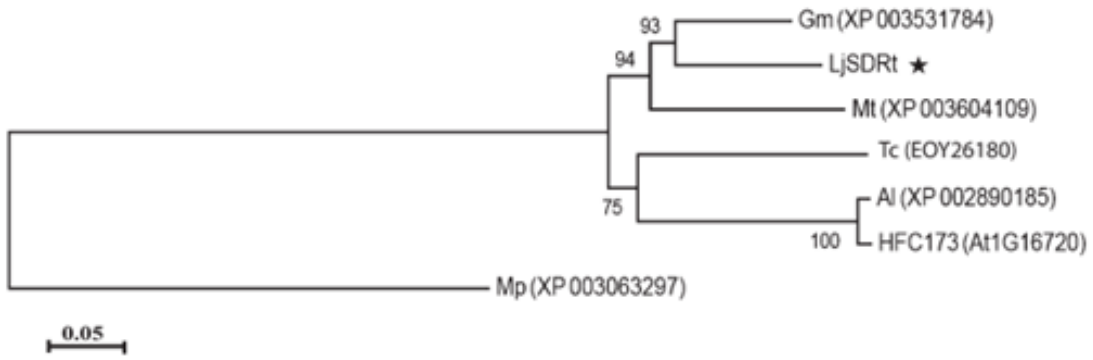


Figure 2c. Phylogenetic tree of SDR-like proteins constructed by the neighbor-joining method with 1000 bootstrap replicates. SDR-like proteins from legumes *L. japonicus*, *M. sativa*, *Glycine max*, from *A. thaliana*, *A. lyrata*, *Theobroma cacao* as well as the green algae *Micromonas pusilla* are depicted.

A syntenic analysis of *L. japonicus AMY2* cluster and *M. truncatula βAS1* cluster was conducted with the Gevo algorithm (Figure 3A). *AMY2* and *bAS1*, *LjCYP88D5* and *MtCYP88D1* are conserved genes in the *AMY2* and *bAS1* gene clusters respectively, which indicates that they have been maintained by evolution despite speciation. Despite this conservation, a conserved synteny was not observed. The high level of sequence homology between the orthologs *AMY2* and *bAS1* and between *LjCYP88D5* and *MtCYP88D1* suggests that these two gene clusters share a common ancestor that evolved independently in these two species.

LjCYP71D353 is phylogenetically related to *CYP71A16* that forms part of the marneral gene cluster in *A. thaliana* (Field *et al.*, 2011). A syntenic analysis of *L. japonicus AMY2* cluster and *A. thaliana* gene clusters of marneral and thalianol synthase was carried out (Figure 3B). The phylogenetic distance between the *OSC* genes in these clusters and the low degree of synteny proposes that these three gene clusters do not share a common origin.

The *AMY2* cluster located on LjT11L01 genomic clone was selected for further experimental investigations.

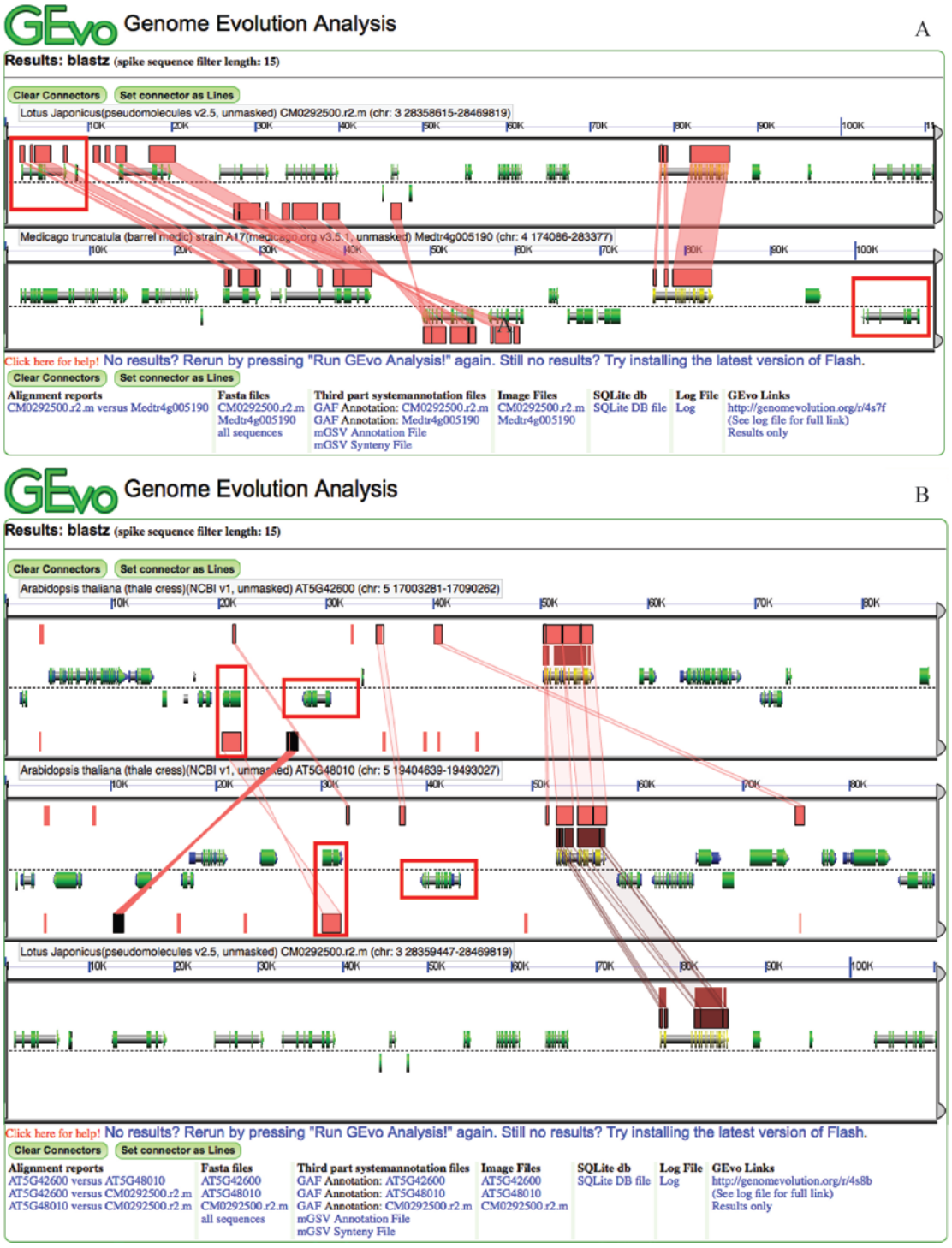


Figure 3. Syntenic analysis of multiple genomic regions encompassing *OSC* genes in *L. japonicus* and *M. truncatula* (described in this study) (A) and in *L. japonicus* and *A. thaliana* (thalianol and marneral gene clusters), (B) using the Gevo algorithm. Cytochrome P450 genes are indicated with red squares and OSC genes with yellow exons.

C.1.2. Verification of the *AMY2* gene cluster

C.1.2.1. Gene expression of the *AMY2* gene cluster partners

C.1.2.1.1. The *AMY2* cluster genes are co-expressed in roots and nodules of *L. japonicus*

The expression patterns of the *AMY2*, *LjCYP88D5*, *LjCYP71D353* and *LjSDRt* genes were examined in leaves and roots, as well as in nodules of *L. japonicus* plants at different developmental stages. Roots, nodules and leaves at various time points of plants grown were collected. Total RNA was isolated and reverse transcribed, the concentration was normalized between the samples and Real-time PCR experiments were conducted. Relative gene expression was measured with respect to *UBQ* transcripts.

In accordance with previous data (Iturbe-Ormaetxe *et al*., 2003; Sawai *et al*., 2006) the expression of *AMY2* is detected in roots by quantitative RT-PCR. Transcript accumulation is higher in the roots of 7 day-old seedlings than in the roots of 14 and 28-day old seedlings (Figure 4). *AMY2* transcript levels are also detectable in the leaves at all three time points but at much lower levels than in roots. The expression patterns of *LjCYP71D353* and *LjCYP88D5* are similar (Figure 4), indicating that the three genes are co-expressed. Transcript levels for *LjSDRt* were not different developmental at the various stages and gene expression is detected in both leaves and roots (Figure 6a).

Previously it had been shown that *AMY2* gene is also highly expressed in the nodules of *L. japonicus* plants that have been inoculated with the symbiotic bacterium *M. loti* (Iturbe-Ormaetxe *et al*., 2003; Sawai *et al*., 2006). It was, therefore, investigated whether the genes that are clustered with *AMY2* are similarly expressed in *M. loti*-infected roots and nodules at various developmental stages. These experiments demonstrated again that the *AMY2*, *LjCYP88D5* and *LjCYP71D353* genes show similar expression patterns in *L. japonicus* roots following infection with *M. loti* strain R7A (Figure 5). Transcript levels were highest in infected roots 7 days post infection (dpi) (which include young nodules) and in nodules at 14 dpi and then decreased in mature nodules at 28 dpi (Figure 5). Again, *LjSDRt* expression did not follow a similar pattern (Figure 6b).

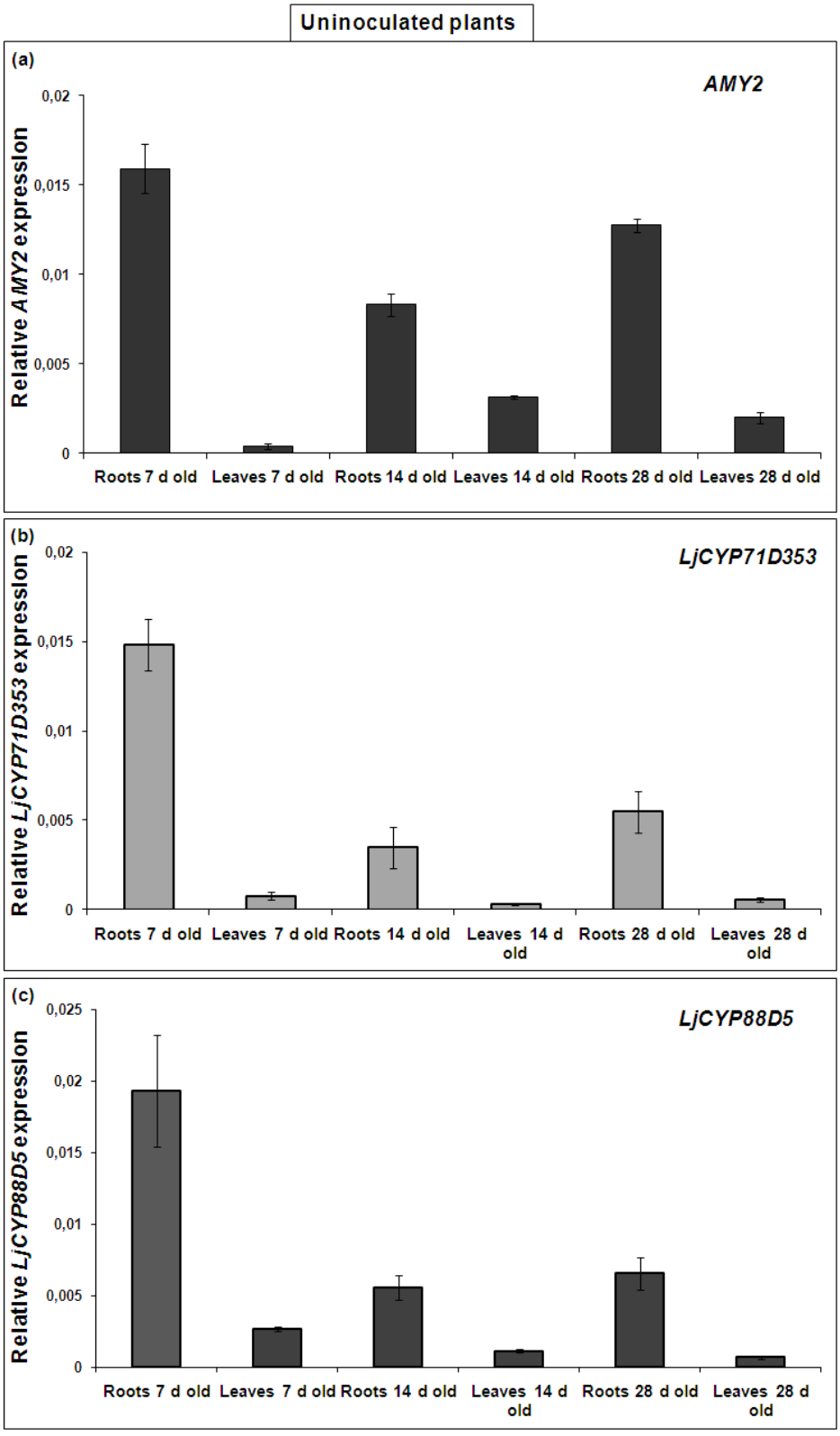


Figure 4. *AMY2* cluster gene expression in different developmental stages. Gene transcript levels of *AMY2, LjCYP71D353* and *LjCYP88D5*, constituting the *AMY2* gene cluster, are detected in uninfected (a, b, c) *L. japonicus* roots, leaves and nodules. Mean values ± SD are shown (n=3).

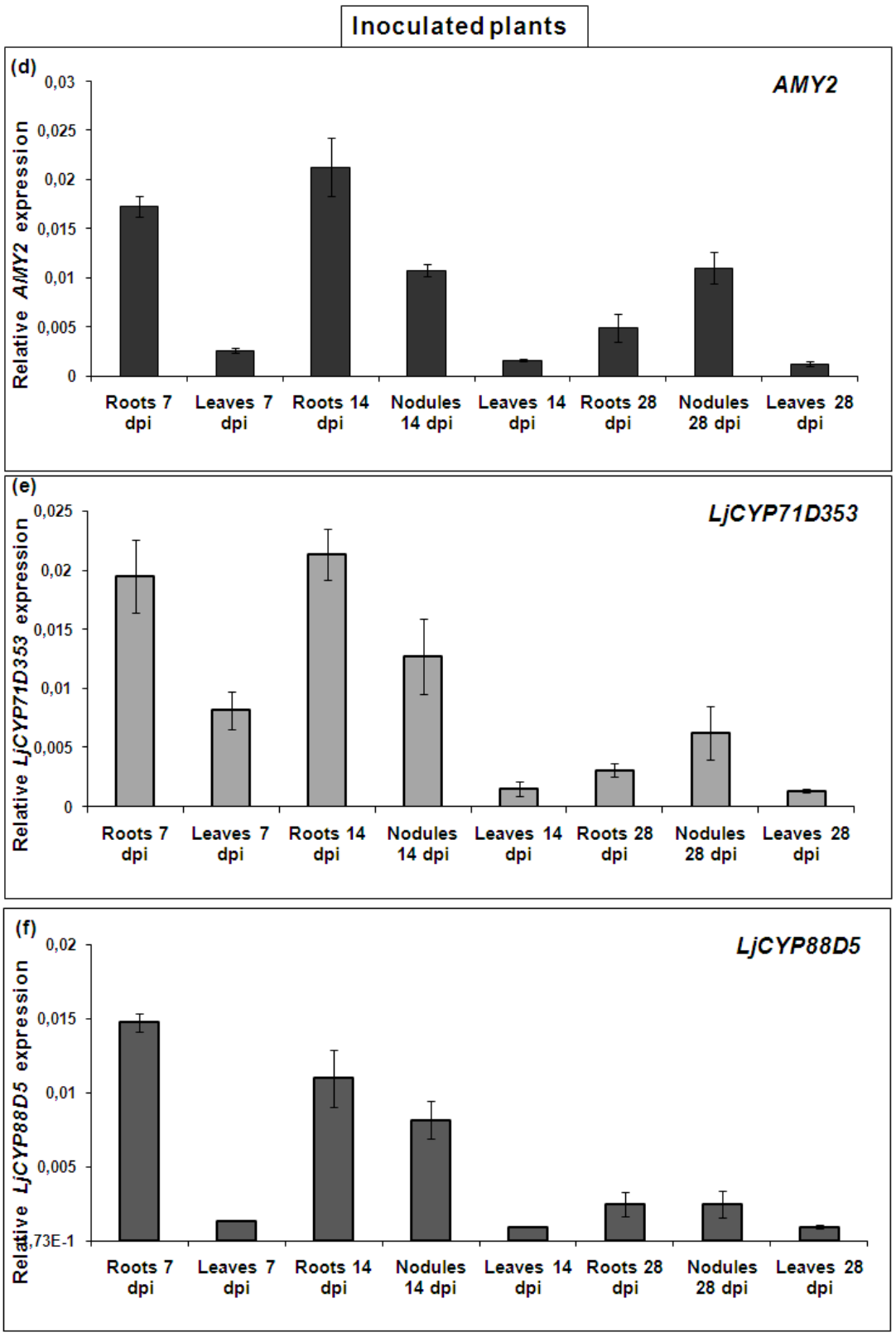


Figure 5. *AMY2* cluster gene expression in different developmental stages. Gene transcript levels of *AMY2, LjCYP71D353* and *LjCYP88D5*, constituting the *AMY2* gene cluster, are detected in infected with *M. loti* (d, e, f) *L. japonicus* roots, leaves and nodules. Uninfected and infected plants are of the same age at the stages of 7 days old (days old) - 7dpi (days post infection), 14 days old - 14 dpi and 28 days old - 28 dpi, respectively. Mean values ± SD are shown (n=3).

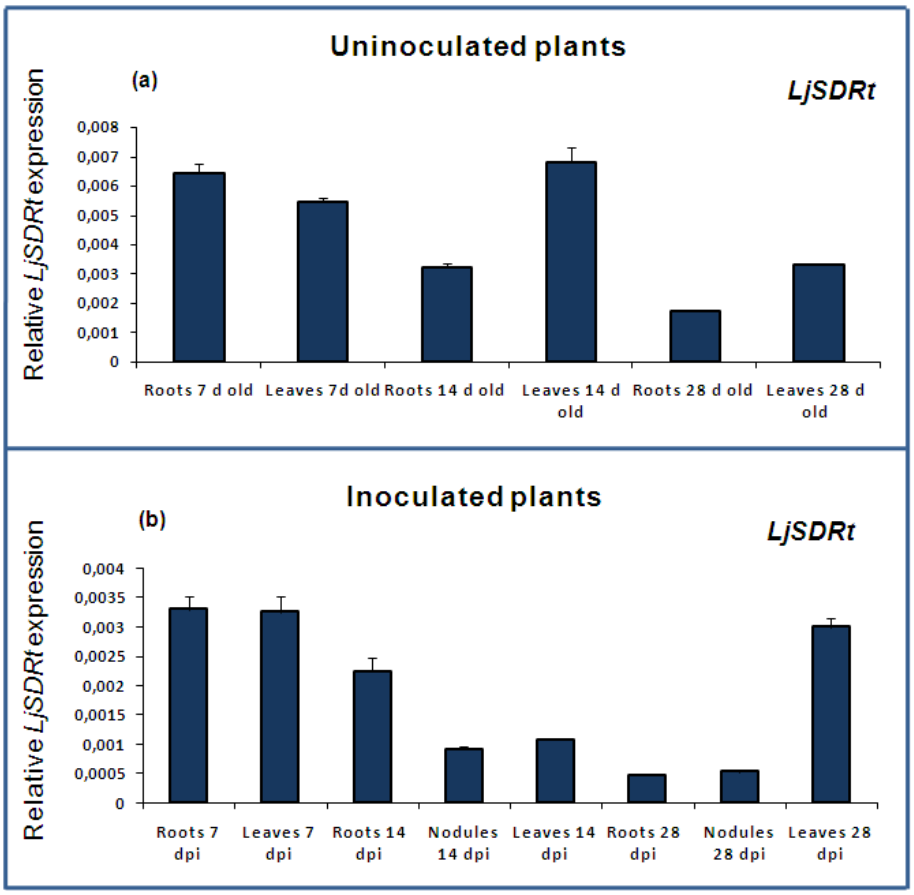


Figure 6. Gene transcript levels of *LjSDRt*, present in the AMY2 gene cluster, are detected in both uninfected and infected with *M. loti* but they do not follow the same profile. Uninfected and infected plants are of the same age at the stages of 7 days old (days old) - 7dpi (days post infection), 14 days old - 14 dpi and 28 days old - 28 dpi, respectively. Mean values ± SD are shown (n=3).

C.1.2.1.2. The *AMY2* cluster genes are co-expressed in *L. japonicus* roots in response to hormone treatments

The effect of three different type of hormones, 2,4 dichlorophenoxyacetic acid (2,4-D), benzylaminopurine (BA) and methyl jasmonic acid (MeJA), on the expression of *AMY2* cluster genes was investigated. *L. japonicus* seedlings were grown on Petri dishes containing MS substrate and then were transferred to Petri dishes containing MS supplemented with 2,4-D, BA or MeJA. Total RNA was isolated from the roots, cDNA were prepared and normalized, and real-time experiments were carried out. In plants treated with these three types of hormones, the transcript levels of *AMY2*, *LjCYP71D353* and *LjCYP88D5* are significantly increased, when compared to control mock-treated plants (Figure 7).

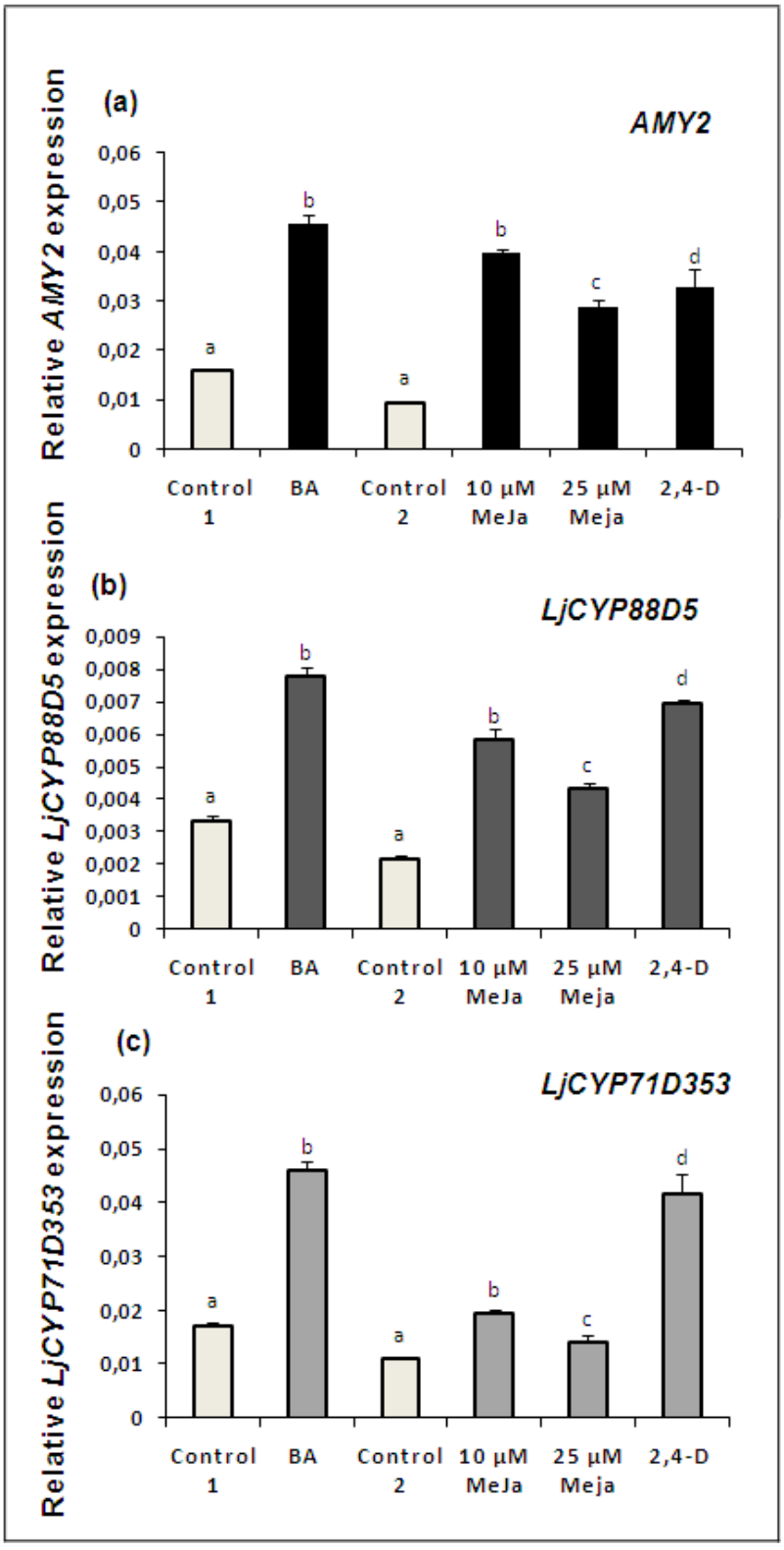


Figure 7. *AMY2* cluster gene expression in *L. japonicus* root tissues subjected to various exogenously applied plant hormones. Gene transcript levels of *AMY2*, *LjCYP88D5* and *LjCYP71D353* in root tissues (14 days old) treated with various concentrations of 2,4-D, BA and MeJA (a–c). Data from a single representative experiment are presented; three experimental repeats yielded similar results. Statistical comparisons were performed by Duncan tests ($a < 0.05$). Indicator letters in common denote lack of significant difference. Bars indicate + SEM ($n = 3$).

C.1.2.1.3. The *AMY2* cluster genes are co-expressed in *L. japonicus* roots in response to abiotic stress treatments

The affection of abiotic stress stimuli, such as heat, cold, oxidative and salt stress on the transcript levels of *AMY2* cluster genes was also investigated. *L. japonicus* seedlings were grown on MS substrate and then, either transferred to MS substrate supplemented with NaCl or paraquat, or exposed to high or low temperature. In accordance with the hormone treatments, total RNA from roots was reverse transcribed, the concentration of the samples was normalized and real-time procedures were conducted.

AMY2, *LjCYP71D353* and *LjCYP88D5* are coordinately down regulated in response to heat, cold and oxidative stress in comparison with control plants (Figure 8). No significant alteration was noticed in the expression profile of *AMY2* cluster genes in response to salt stress (Figure 9 a, b, c).

Collectively, it can be concluded that the *AMY2*, *LjCYP71D353* and *LjCYP88D5* genes comprise a tightly co-regulated cluster of functionally related genes.

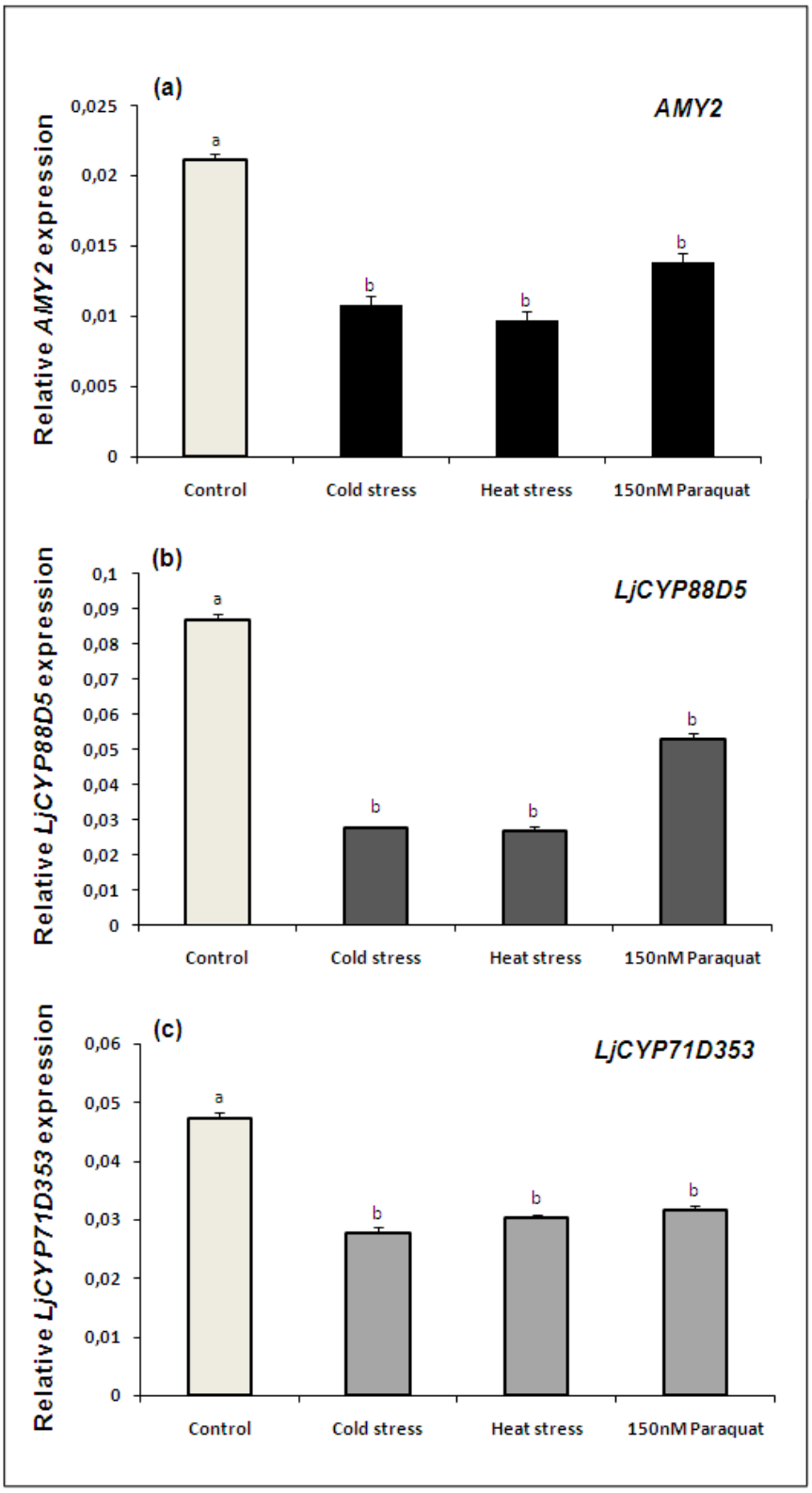


Figure 8. *AMY2* cluster gene expression in *L. japonicus* root tissues subjected to various environmental cues. Gene transcript levels of *AMY2*, *LjCYP88D5* and *LjCYP71D353* in root tissues treated with heat and cold stress (12 days old, treated 24 h) and paraquat (14 days old) (a-c). Data from a single representative experiment are presented; three experimental repeats yielded similar results. Statistical comparisons were performed by Duncan tests ($a < 0.05$). Indicator letters in common denote lack of significant difference. Bars indicate + SEM ($n = 3$).

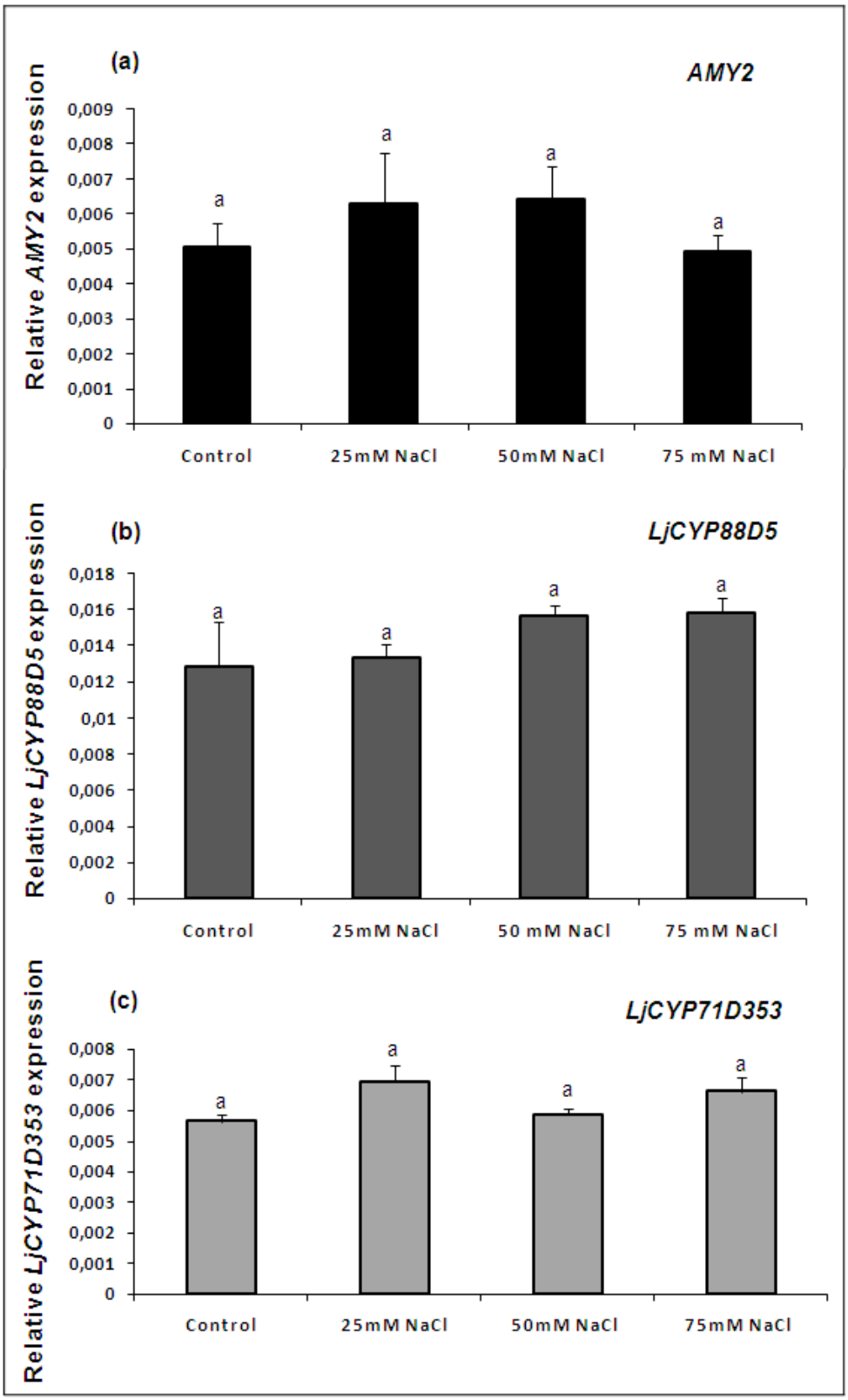
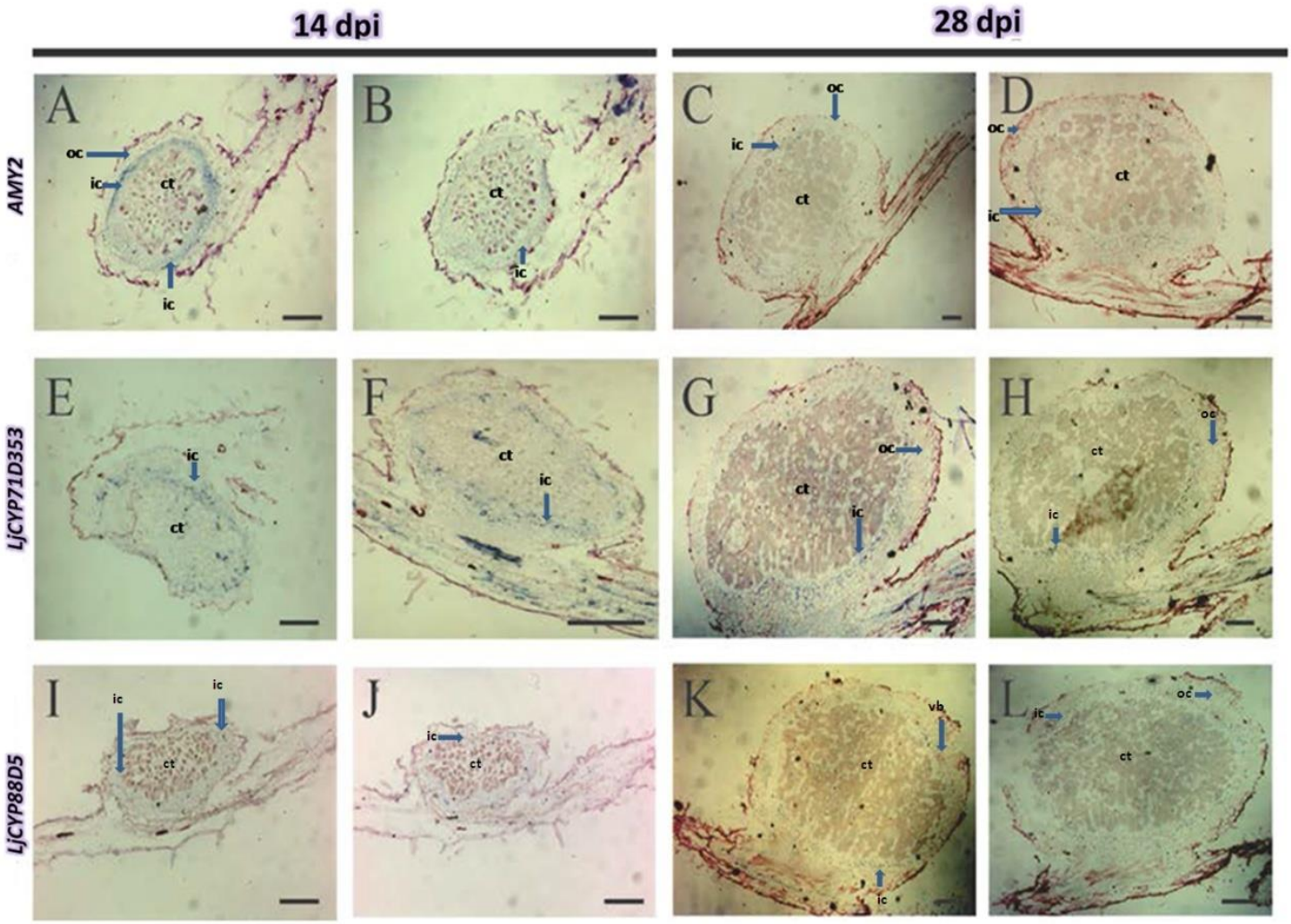


Figure 9. *AMY2* cluster gene expression in *L. japonicus* root tissues subjected to salt stress. Gene transcript levels of *AMY2*, *LjCYP71D353* and *LjCYP88D5* in root tissues (14 days old, treated 7 days). Statistical comparisons were performed by Duncan tests ($a < 0.05$). Bars indicate + SEM ($n = 3$).

C.1.2.1.4. The *AMY2* cluster genes are co-localized in nodules of *L. japonicus*

The localization of *AMY2*, *LjCYP71D353* and *LjCYP88D5* was explored in *L. japonicus* developing (14 dpi) and mature nodules (28 dpi), using mRNA *in situ* hybridization. Sections (7 µm) of nodules were hybridized with antisense and sense RNA probes, labeled with digoxigenin (DIG)-11-rUTP. Hybridization signal is visible as blue-purple precipitate. Pairs of gene-specific primers for the cluster genes were designed and used for the in vitro transcription of RNA probes. For both developmental stages, *AMY2*, *LjCYP88D5* and *LjCYP71D353* gene transcripts were localized in the vascular bundles, parenchymatic cells (inner parenchyma) and also in uninfected cells of the central tissue (Figure 12 A, B, C, D; 12 E, F, G, H and 12 I, J, K, L respectively). The spatial patterns of expression of the three genes were remarkably similar.



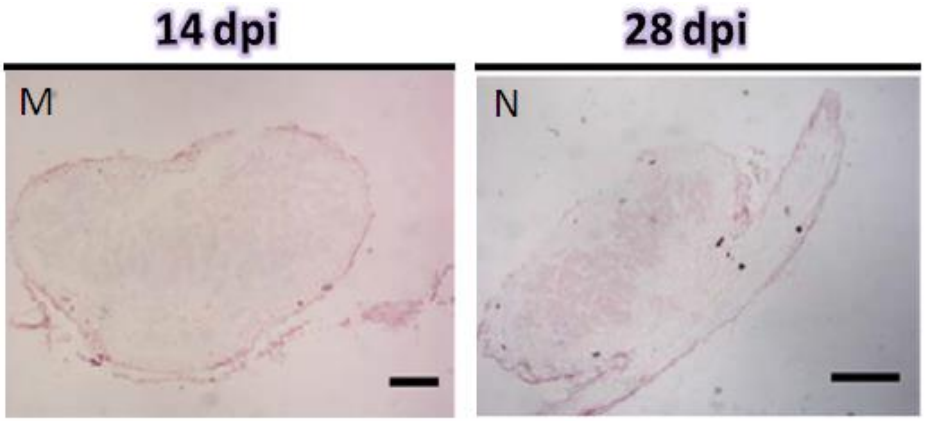


Figure 10. *In situ* hybridization of *AMY2* (A, B, C, D), *LjCYP71D353* (E, F, G, H) and *LjCYP88D5* (I, J, K, L) gene transcripts in developing 14dpi and mature 28dpi *L. japonicus* nodules. Eight µm thin sections were hybridized with DIG-11rUTR labeled anti-sense RNA *in vitro* transcribed from of PCR products of *AMY2, LjCYP71D353* and *LjCYP88D5*. Hybridization signal was visualized using an alkaline phosphate reaction product (blue-purple colour). For all genes in mature and developing nodules the hybridization signal was detected in the inner cortex (ic), in vascular bundles (vb) and in uninfected cells of central tissue (ct). M, N: As a negative control, thin sections of 14dpi (M) and 28 dpi (N) *L. japonicus* nodules were hybridized with DIG-11rUTR sense RNA *in vitro* transcribed from PCR products of *AMY2, LjCYP71D353* and *LjCYP88D5* clones. In this case, no significant hybridization signal is visible. Scale bar 100µm.

C.1.2.2. *In silico* prediction of the putative promoter regions and regulatory motifs of *AMY2* cluster genes

Genomic regions located about 2 kb upsteam of the start codon of each *L. japonicus* gene (*AMY2*, *LjCYP88D5*, *LjCYP71D353* and *LjSDRt*) in LjT11L01 clone, were examined for repeats and inverted repeats (http://myhits.isb-sib.ch/cgi-bin/dotlet). Inverted repeats are usually found in DNA regions, to which transcription factors bind. For example, the G-box is a cis-acting element found within the promoters of many plant genes where it mediates expression in response to a variety of different stimuli. This palindromic DNA motif (CCACGTGG) is composed of two identical half sites, the base pairs of which we have numbered -4 to +4 (numbering from 5' to 3'). Both half sites are involved in the binding of the bZIP protein GBF1, a member of the GBF family of *A. thaliana* (Schindler *et al.*, 1992). Additionally, the DNA of many transposons is flanked by inverted repeats. Finally, inverted repeats are commonly found in duplicated genes. Interestingly, many repeats and inverted repeats are found in the putative promoter domains of *AMY2* cluster genes, revealing their dynamic nature (Figure 11). LjT11L01 clone contains many transposons (Figure 1).

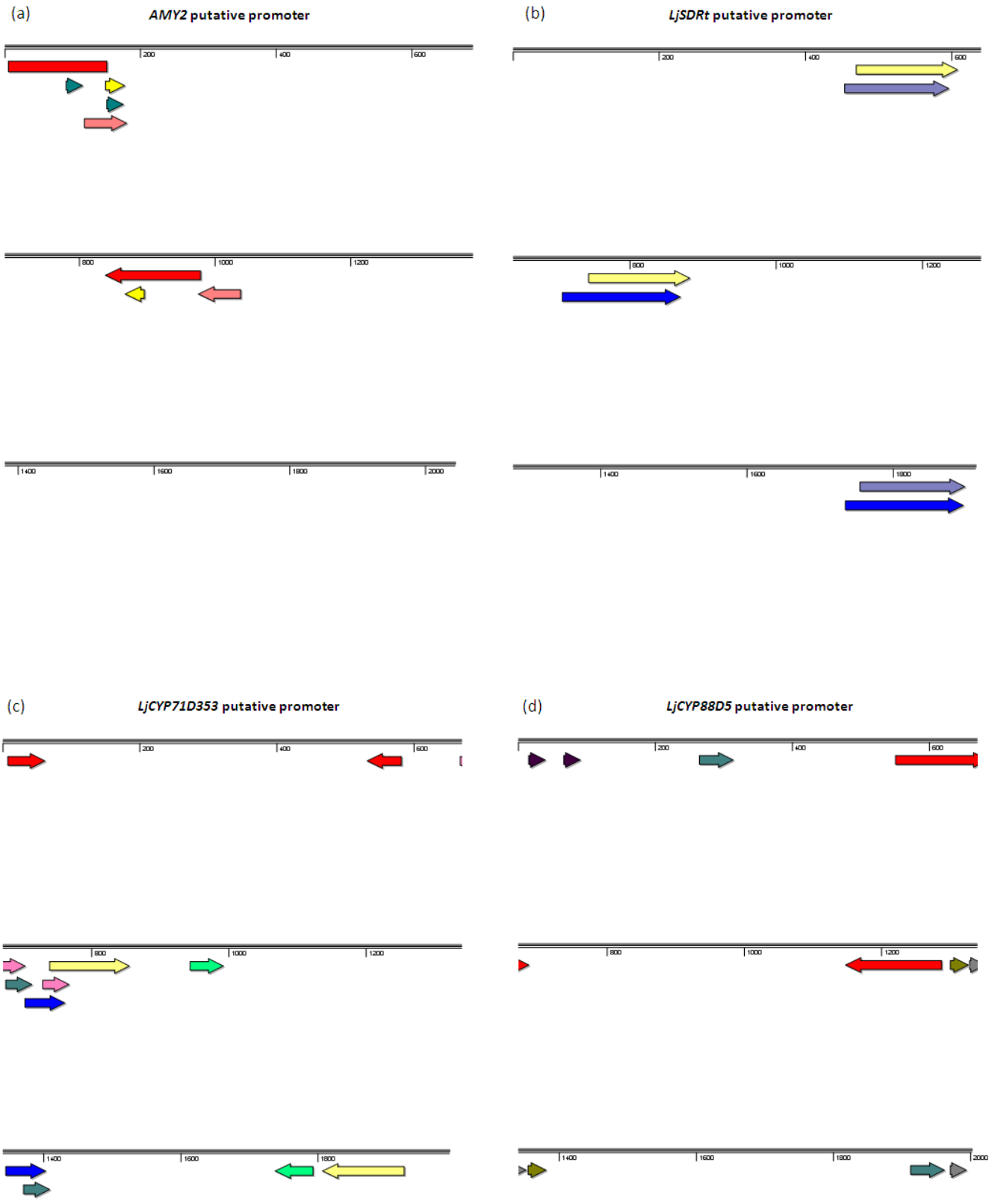


Figure 11. Repeats in *AMY2* (a), *LjSDRt* (b), *LjCYP71D353* (c) and *LjCYP88D5* (d) putative promoter regions. The same colors of underlined areas indicate a repeat in each promoter. Additionally, the direction of the arrows indicates the 5′ – 3′ or 3′ – 5′ reading of the partners in a pair of repeats.

The location of the TATA box in *AMY2*, *LjCYP71D353*, *LjCYP88D5* and *LjSDRt* putative promoter regions was predicted using the TSSP software of SoftBerry database (http://linux1.softberry.com/berry.phtml?topic=tssp&group=programs&subgroup=promoter) (Table 2). For most genes transcribed by RNA polymerase II, the most important *cis*-acting element is called the TATA box on the basis of its consensus sequence. TATA box is a promoter sequence that specifies where transcription begins. In many eukaryotic genes the TATA box is located 25-35 bp before the site of the transcription start. In addition, in the case of *AMY2* and *LjCYP88D5*, the position of the enhancers was also predicted. Enhancer *cis*-elements although not promoters themselves, can enormously increase the effectiveness of promoters. Interestingly, the positions of enhancers relative to promoters are not fixed; they can vary substantially. Enhancers play key roles in regulating gene expression in a specific tissue or developmental stage (Stryer, Biochemistry, 5th edition, 2002).

Promoter	Position of TATA box	Position of enhancer
AMY2	243	1539
		365
LjCYP71D353	1586	Not predicted
LjCYP88D5	593	718
LjSDRt	54	Not predicted

Table 2. Putative TATA box location in *AMY2* cluster promoters.

Putative promoters of *AMY2* cluster genes were screened in PlantPAN transcription factor library selecting *A. thaliana* and *M. truncatula* as species for comparison (http://plantpan.mbc.nctu.edu.tw/), and in NSITEM-PL (http://linux1.softberry.com/berry.phtml?topic=nsitemp&group=programs&subgroup=promoter) and PLACE (http://www.dna.affrc.go.jp/PLACE/). Transcription factors (TFs) that bind to *cis*-regulatory DNA sequences are responsible for either positively or negatively influencing the transcription of specific genes, essentially determining whether a particular gene will be turned "on" or "off" in an organism. After TFs bind to promoter or enhancer regions of the DNA, they interact with other bound TFs and

recruit RNA polymerase II. A typical TF has multiple functional domains, not only for recognizing and binding to the appropriate DNA strand, but also for interactions with other TFs, with proteins called co-activators, with RNA polymerase II, with chromatin remodeling complexes, and with small non-coding RNAs. Some TFs are believed to act as tethering elements between distant enhancers and promoters by forming connections with other proteins (Chen and Rajewsky, 2007). The distribution of predicted binding sites for TFs in *AMY2* cluster promoter sequences is shown in Figure 12.

Figure 12. Location of predicted binding sites for TFs across the putative promoter domains of *AMY2* cluster genes.

Many TFs were predicted to bind on *AMY2*, *LjCYP71D353*, *LjCYP88D5* and *LjSDRt* promoters, as it is demonstrated in Figure 12. The possible roles of predicted TFs were identified when screening the PlnTFTB 3.0, PlantTFTB and PlantPAN databases, as many of them have been already reported. In addition, a comparison between the TFs identified to bind in all promoters was done. Interestingly, the four genes of *AMY2* cluster share many common TFs, involved in various plant processes. The common TFs are all presented in Table 3. Among them, the ARFAT is an auxin response factor that

regulates the expression of auxin response genes (Goda *et al*., 2004). The ARR1AT is a response regulator that is involved in an early step of cytokinin transduction (Sakai *et al*., 2001). The T/GBOXATPIN2 is a T/G box *cis*-element that is involved in jasmonate induction of *LAP* genes (Boter *et al*., 2004). The WRKY71OS belongs to WRKY transcription factors, which play important roles in plant responses to biotic and abiotic stresses (Zhang *et al*., 2004; 2005). Promoter motifs expressed specifically in roots and root hairs are also recorded. The OSE1ROOTNODULE motif is a characteristic *cis*-element of the promoters activated in infected cells of root nodules (Vieweg *et al*., 2004; Fehlberg *et al*., 2005). The RHERPATEXPA7 is a root hair-specific *cis*-element (Kim *et al*., 2006) and ROOTMOTIFTAPOX1 is a *cis*-element found in the promoter of rolD gene in roots (Elmayar *et al*., 1995). These findings are only preliminary results and further investigations are required.

Transcription factor	Consensus sequence	Description
ABRELATERD1	ACGTG	ABRE-like sequence required for etiolation-induced expression of erd1 (early responsive to dehydration) in *A. thaliana*. ABRE; etiolation; erd; Arabidopsis thaliana ACGTG Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant J. 33: 259-270 (2003) PubMed: 12535340; Nakashima K, Fujita Y, Katsura K, Maruyama K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of Arabidopsis. Plant Mol Biol. 60:51-68 (2006). PubMed: 16463099.
ACGTATERD1	ACGT	ACGT sequence (from -155 to -152) required for etiolation-induced expression of erd1 (early responsive to dehydration) in *Arabidopsis*; ACGT; etiolation; erd; *Arabidopsis thaliana* ACGT Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Two different novel cis-acting elements of erd1, a clpA homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant J. 33: 259-270 (2003) PubMed: 12535340.
ANAERO3CONSENSUS	AAACAAA	One of 16 motifs found in silico in promoters of 13 anaerobic genes involved in the fermentative pathway (anaerobic set 1) (Mohanty *et al*., 2005); Arbitrary named ANAERO3CONSENSUS by the PLACEdb curator; See also S000477, S000478, S000480, S000481; anaerobic; Zea mays (maize); *Arabidopsis thaliana*; *Pisum sativum* (pea); *Hordeum vulgare* (barley); *Oryza sativa* (rice); *Petunia hybrida* (petunia); Lycopersicon esculentum (tomato); TCATCAC Mohanty B, Krishnan SP, Swarup S, Bajic VB. Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. Ann. Bot. (Lond).96: 669-681. (2005) in silico PubMed: 16027132.
ARFAT	TGTCTC	Auxin response factors (ARF) are transcription factors that regulate the expression of auxin response genes. ARFs bind with specificity

		to TGTCTC auxin response elements (AuxRE) in promoters of these genes and function in combination with Aux/IAA (auxin/indole acetic acid) repressors, which dimerize with ARF activators in an auxin-regulated manner. Goda H, Sawa S, Asami T, Fujioka S, Shimada Y, Yoshida S. Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in Arabidopsis. Plant Physiol. 2004 Apr;134(4):1555-73.
ARR1AT	NGATT	ARR1 is a principal transcription factor-type response regulator that is involved in an early step of cytokinin signal transduction, possibly as a partner of the sensor histidine kinase CRE1 in *L. japonicus.* Hiroe Sakai, Takashi Honma, Takashi Aoyama, Shusei Sato, Tomohiko Kato, Satoshi Tabata, Atsuhiro ARR1, a Transcription Factor for Genes Immediately Responsive Cytokinin Science, Vol 294, Issue 5546, 1519-1521, 16 November 2001.
CAATBOX1	CAAT	CAAT promoter consensus sequence found in *legA* gene of pea. NF-Y is a bifunctional transcription factor capable of activating or repressing transcription. NF-Y specifically recognizes CCAAT box motifs present in many eukaryotic promoters. The mechanisms involved in regulating its activity are poorly understood. Shirsat A, Wilford N, Croy R, Boulter D Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Mol Gen Genet. 1989 Jan;215(2):326-31.
DOFCOREZM	AAAG	Dof proteins are DNA binding proteins, with presumably only one zinc finger, and are unique to plants. Dof proteins have been reported to participate in the regulation of gene expression in processes such as seed storage protein synthesis in developing endosperm, light regulation of genes involved in carbohydrate metabolism, plant defense mechanisms, seed germination, gibberellin response in post-germinating aleurone, auxin response and stomata guard cell specific gene regulation. Yanagisawa S. and Schmidt RJ Diversity and similarity among recognition sequences of Dof transcription factors. Plant J. 17, 209-214 (1999).
DPBFCOREDCDC3	ACACNNG	A novel class of bZIP transcription factors, DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2) binding core sequence; Found in the carrot (D.c.) Dc3 gene promoter; Dc3 expression is normally embryo-specific, and also can be induced by ABA in *A. thaliana*. Kim SY, Chung HJ, Thomas TL Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system Plant J 11: 1237-1251 (1997) PubMed: 9225465. The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor Plant Cell 12: 599-609 (2000) PubMed: 10760247; Lopez-Molina L, Chua NH A null mutation in a bZIP factor confers ABA-insensitivity in Arabidopsis thaliana Plant Cell Physiol 41: 541-547 (2000) PubMed: 10929936.
EBOXBNNAPA	CANNTG	E-box of napA storage-protein gene of *Brassica napus.* Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. Planta 199:515-519 (1996) PubMed: 8818291; Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B. Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol Biol. 57: 155-171 (2005). PubMed: 15821875.
GT1GMSCAM4	GAAAAA	GT-1 motif found in the promoter of soybean (*Glycine max*) CaM isoform, SCaM-4; Plays a role in pathogen- and salt-induced SCaM-4 gene expression. Park HC, Kim ML, Kang YH, Jeon JM, Yoo JH, Kim MC, Park CY, Jeong JC, Moon BC, Lee JH, Yoon HW, Lee SH, Chung WS, Lim CO, Lee SY, Hong JC, Cho MJ. Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1

		box that interacts with a GT-1-like transcription factor. Plant Physiol. 135: 2150-2161 (2004). PubMed: 15310827.
IBOXCORE	GATAAG	I-box is a conserved sequence upstream of light-regulated genes. Sequence found in the promoter region of rbcS of tomato and Arabidopsis; I box. Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore ARProc Natl Acad Sci U S A. 1988 Oct;85(19):7089-93. An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene.
MYB1AT	WAACCA	MYB recognition site found in the promoters of the dehydration-responsive gene *rd22* and many other genes in *Arabidopsis*; Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15: 63-78 (2003) PubMed: 12509522.
MYCCONSENSUSAT	CANNTG	MYC recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in *Arabidopsis*. Binding site of ICE1 (inducer of CBF expression 1) that regulates the transcription of CBF/DREB1 genes in the cold in *Arabidopsis* (Chinnusamy et al., 2004). Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15: 63-78 (2003) PubMed: 12509522; Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev. 17: 1043-1054 (2003) PubMed: 12672693; Chinnusamy V, Schumaker K, Zhu JK. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. J Exp Bot. 55: 225-236 (2004).
NODCON2GM	CTCTT	One of two putative nodulin consensus sequences. Sandal NN, Bojsen K, Marcker KA. A small family of nodule specific genes from soybean. Nucleic Acids Res. 15:1507-1519 (1987). consesus; in silico; PubMed: 3822835 Stougaard J, Jorgensen JE, Christensen T, Kuhle A, Marcker KA. Interdependence and nodule specificity of cis-acting regulatory elements in the soybean leghemoglobin lbc3 and N23 gene promoters. Mol Gen Genet. 220: 353-360 (1990). PubMed: 2338938.
OSE1ROOTNODULE	AAAGAT	One of the consensus sequence motifs of organ-specific elements (OSE) characteristic of the promoters activated in infected cells of root nodules. Vieweg MF, Fruhling M, Quandt HJ, Heim U, Baumlein H, Puhler A, Kuster H, Andreas MP. The promoter of the Vicia faba L. leghemoglobin gene VfLb29 is specifically activated in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots from different legume and nonlegume plants. Mol Plant Microbe Interact. 17: 62-69 (2004). PubMed: 14714869 Fehlberg V, Vieweg MF, Dohmann EM, Hohnjec N, Puhler A, Perlick AM, Kuster H. The promoter of the leghaemoglobin gene VfLb29: functional analysis and identification of modules necessary for its activation in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots. J Exp Bot. 56:799-806 (2005) PubMed: 15668224.
POLASIG1	AATAAA	Plant polyA signal; Consensus sequence for plant polyadenylation signal. Loke JC, Stahlberg EA, Strenski DG, Haas BJ, Wood PC, Li QQ. Compilation of mRNA polyadenylation signals in *Arabidopsis* revealed a new signal element and potential secondary structures. Plant Physiol. 138: 1457-1468. (2005) in silico PubMed: 15965016.
PYRIMIDINEBOXOSRAMY1A	CCTTTT	Pyrimidine box found in rice (O.s.) alpha-amylase (RAmy1A) gene; Gibberellin-respons cis-element of GARE and pyrimidine box are partially involved in sugar repression; Found in the promoter of barley alpha-amylase (*Amy2/32b*) gene which is induced in the aleurone layers in response to GA; BPBF protein binds specifically to this site.

RHERPATEXPA7	KCACGW	Right part of RHEs (Root Hair-specific cis-Elements) conserved among the *Arabidopsis thaliana* A7 (AtEXPA7) orthologous (and paralogous) genes from diverse angiosperm species with different hair distribution patterns. Kim DW, Lee SH, Choi SB, Won SK, Heo YK, Cho M, Park YI, Cho HT. Functional Conservation of a Root Hair Cell-Specific cis-Element in Angiosperms with Different Root Hair Distribution Patterns. Plant Cell. 18:2958-2970 (2006) PubMed: 17098810.
ROOTMOTIFTAPOX1	ATATT	Motif found both in promoters of *rolD* gene in root and *rolD* gene in *Agrobacterium rhizogenes.* Elmayan T, Tepfer M Evaluation in tobacco of the organ specificity and strength of the rol D promoter, domain A of the 35S promoter and the 35S^2 promoter Transgenic Res 4:388-396 (1995) Sequence similarity; PubMed: 7581519.
SEF3MOTIFGM	AACCCA	SEF3 binding site; Soybean consensus sequence is found in the 5' upstream region of beta-conglycinin (7S globulin) gene. Allen RD, Bernier F, Lessard PA, Beachy RN Nuclear factors interact with a soybean beta-conglycinin enhancer. Plant Cell 1:623-631 (1989) PubMed: 2535514; Lessard PA, Allen RD, Bernier F, Crispino JD, Fujiwara T, Beachy RN Multiple nuclear factors interact with upstream sequences of differentially regulated beta-conglycinin genes. Plant Mol Biol 16:397-413 (1991) PubMed: 1893110.
SURECOREATSULTR11	GAGAC	Core of sulfur-responsive element (SURE) found in the promoter of SULTR1.1 high-affinity sulfate transporter gene in Arabidopsis; SURE contains an auxin response factor (ARF) binding sequence (GAGACA). Maruyama-Nakashita A, Nakamura Y, Watanabe-Takahashi A, Inoue E, Yamaya T, Takahashi H. Identification of a novel cis-acting element conferring sulfur deficiency response in Arabidopsis roots.Plant J. 42: 305-314 (2005).
T/GBOXATPIN2	AACGTG	T/G-box found in tomato proteinase inhibitor II (*pin2*) and leucine aminopeptidase (*LAP*) genes; Involved in jasmonate (JA) induction of these genes; bHLH-Leu zipper JAMYC2 and JAMYC10 proteins specifically recognize this motif. Boter M, Ruiz-Rivero O, Abdeen A, Prat S. Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes Dev 18: 1577-1591 (2004). PubMed: 15231736.
TAAAGSTKST1	TAAAG	TAAAG motif found in promoter of *Solanum tuberosum KST1* gene; Target site for trans-acting StDof1 protein controlling guard cell-specific gene expression; *KST1* gene encodes a K^+ influx channel of guard cells. Plesch G, Ehrhardt T, Mueller-Roeber B Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression Plant J 28: 455-464 (2001) PubMed: 11737782.
WBOXATNPR1	TTGAC	W-box found in promoter of *Arabidopsis thaliana NPR1* gene; Located between +70 and +79 in tandem; They are recognized specifically by salicylic acid (SA)-induced WRKY DNA binding proteins. Yu D, Chen C, Chen Z Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression Plant Cell 13: 1527-1540 (2001) PubMed: 11449049; Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budworth PR, Tao Y, Xie Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses Plant Cell 14: 559-574 (2002) PubMed: 11910004; Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. Trends Plant Sci. 5:199-206. (2000) Review. PubMed: 10785665 Chen C, Chen Z. Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced Arabidopsis transcription factor. Plant Physiol. 129:706-716 (2002) PubMed: 12068113 Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA. The transcriptome of Arabidopsis thaliana during systemic acquired resistance. Nat Genet. 26:403-410 (2000) PubMed:

		11101835 Xu X, Chen C, Fan B, Chen Z Physical and Functional Interactions between Pathogen-Induced Arabidopsis WRKY18, WRKY40, and WRKY60 Transcription Factors. Plant Cell 18:1310-1326 (2006) See also S000142.
WBOXNTERF3	TGACY	W box found in the promoter region of a transcriptional repressor ERF3 gene in tobacco; May be involved in activation of ERF3 gene by wounding. Nishiuchi T, Shinshi H, Suzuki K. Rapid and transient activation of transcription of the ERF3 gene by wounding in tobacco leaves: Possible involvement of NtWRKYs and autorepression. J Biol Chem. 279: 55355-55361 (2004) PubMed: 15509567.
WRKY71OS	TGAC	WRKY transcription factors, originally isolated from plants contain one or two conserved WRKY domains, about 60 amino acid residues with the WRKYGQK sequence followed by a C_2H_2 or C_2HC zinc finger motif. Evidence is accumulating to suggest that the WRKY proteins play significant roles in responses to biotic and abiotic stresses, and in development. Zhang ZL, Xie Z, Zou X, Casaretto J, Ho TH, Shen QJ. A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. Plant Physiol. 134:1500-1513(2004) PubMed: 15047897 Xie Z, Zhang ZL, Zou X, Huang J, Ruas P, Thompson D, Shen QJ. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol. 137:176-189 (2005) PubMed: 15618416 Eulgem T, Rushton PJ, Schmelzer E, Hahlbrock K, Somssich IE. Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors. EMBO J. 18:4689-4699 (1999) PubMed: 10469648 Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. Trends Plant Sci. 5:199-206. (2000) Review. PubMed: 10785665.

Table 3. The common transcription factors identified to bind in the four putative promoter regions. The sequence that they recognize and a brief description are recorded.

C.1.3. Functional analysis in *Nicotiana benthamiana*

C.1.3.1. Cloning of *AMY2* cluster genes in pBinP-NS-ER-GFP

By analogy with previously reported plant metabolic gene clusters, the first thought was that the two cytochrome P450 enzymes that are encoded by the *AMY2* cluster genes may act on the product of the signature enzyme AMY2. Therefore, it was assumed that LjCYP88D5 and LjCYP71D353 participate in the synthesis of triterpene secondary metabolites in *L. japonicus*. The full-length cDNAs of *AMY2*, *LjCYP88D5* and *LjCYP71D353* (Materials and Methods, Appendix B.16.3) were amplified by PCR reaction using specific pairs of primers containing the suitable restriction enzymes and then, they were first cloned into the pM81-FSC1 vector. The generated plasmids pM81-AMY2, pM81-LjCYP71D353 and pM81-LjCYP88D5 (Figure 13) were excised with the appropriate restriction enzymes and then, they were cloned into pBinP-NS-ER-GFP

vector. The resultant binary plasmids pBinP-AMY2, pBinP-LjCYP71D353 and pBinP-LjCYP88D5 (Figure 14) maintained in *A. tumefaciens* and infiltrated in *N. benthamiana* leaves for transient expression (Canizares *et al.*, 2006).

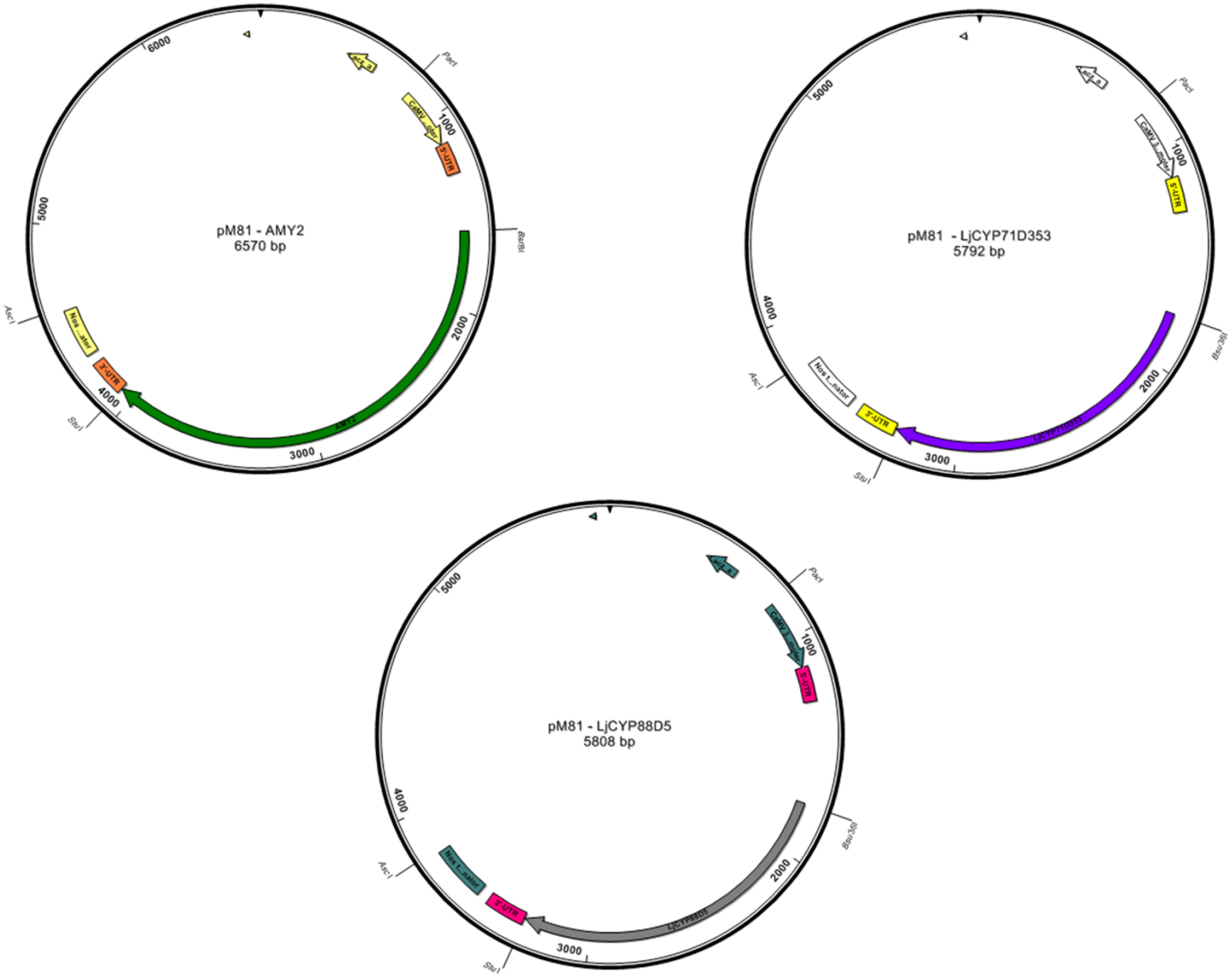


Figure 13. pM81-FSC1 vector containing the *AMY2* cluster genes between the viral 3'- and 5'-UTR regions. *AMY2* was cloned into *StuI / BstBI* restriction sites and *LjCYP88D5* and *LjCYP71D353* were cloned into *StuI / Bsu36i* restriction sites as shown. pM81-AMY2, pM81-LjCYP88D5 and pM81-LjCYP71D353 were digested with the *AscI / PacI* restriction enzymes and ligated into the birary vector pBinP-NS-ER-GFP. The digested regions contain the CaMV promoter, the 5'-UTR, the full-length gene, the 3'-UTR and the NOS terminator.

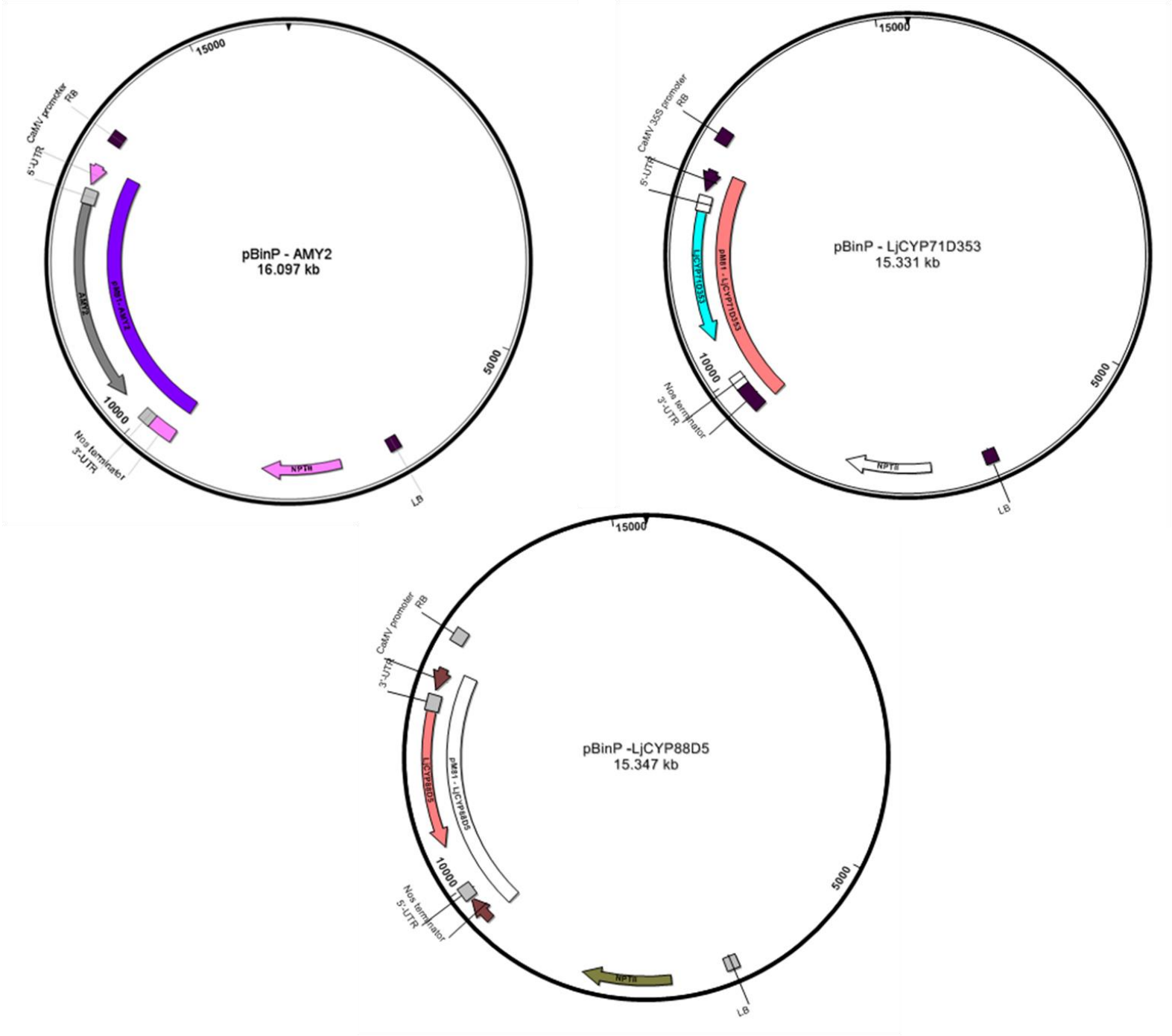


Figure 14. The binary plasmids pBinP-AMY2, pBinP-LjCYP88D5 and pBinP-LjCYP71D353 resulting from the described cloning steps. Plasmids were maintained in *A. tumefaciens* and were used for heterologous expression of *AMY2* cluster genes in *N. benthamiana* leaves.

C.1.3.2. GC/MS analysis of the metabolites of *AMY2* cluster genes

Following transformation, *N. benthamiana* leaves were collected, subjected to metabolite extraction and analyzed by GC-MS. The experimental procedures were conducted mainly by Dr Luis M. Peña-Rodríguez and Dr Katrin Geisler at the John Innes

Centre. In *N. benthamiana* leaves AMY2 produced β-amyrin and dihydrolupeol, instead of lupeol. No new metabolites were observed when *LjCYP88D5* or *LjCYP71D353* were expressed alone. Simultaneous expression of *AMY2* and *LjCYP71D353* resulted in the production of 20-hydroxybetulinic acid in three sequential oxidations steps of 20-hydroxylupeol. More details about the GC-MS results are discussed in Appendix 2.

C.1.4. The functional role of *AMY2* gene cluster

C.1.4.1. *Agrobacterium rhizogenes* mediated plant transformation

It has been previously shown by gene silencing that lupeol has a role in suppression of nodule formation in *L. japonicus* (Delis *et al*., 2011). Additionally, it has been reported that heterologous expression of a β-amyrin synthase from aster in *M. truncatula* resulted in enhanced nodulation (Confalonieri *et al*., 2009). Thus, simple triterpenes have different and opposing effects on nodule development.

A hairpin RNA gene-silencing strategy was employed to investigate a putative role of the *AMY2* cluster genes in nodulation. Two silencing constructs were made for the *AMY2* gene, two constructs were made for *LjCYP88D5* gene and one silencing construct for *LjCYP71D353* gene. Specific primers carrying restriction sites were used and the PCR – amplified fragments were cloned into the *KpnI / XhoI* restriction sites of pENRT4 vector, replacing its *ccdB* gene. The location of the PCR amplicons in the corresponding full-length genes is shown in Figure 15. The resultant plasmids pENTRY-AMY2-2, pENTRY-AMY2-3, pENTRY-LjCYP71D353, pENTRY-LjCYP88D5-1 and pENTRY-LjCYP88D5-3 were used in an LB clonase reaction with destination vector pUBI-GWS-GFP and the binary plasmids pUBI-AMY2-2, pUBI-AMY2-3, pUBI-LjCYP71D353, pUBI-LjCYP88D5-1 and pUBI-LjCYP88D5-3 were maintained in *A. rhizogenes* LBA1334 (Figure 16). Control plant lines were inoculated with *A. rhizogenes* carrying the pUBI-GWS-GFP empty vector. Transgenic roots were generated using an *A. rhizogenes* transformation protocol (Martirani *et al*., 1999).

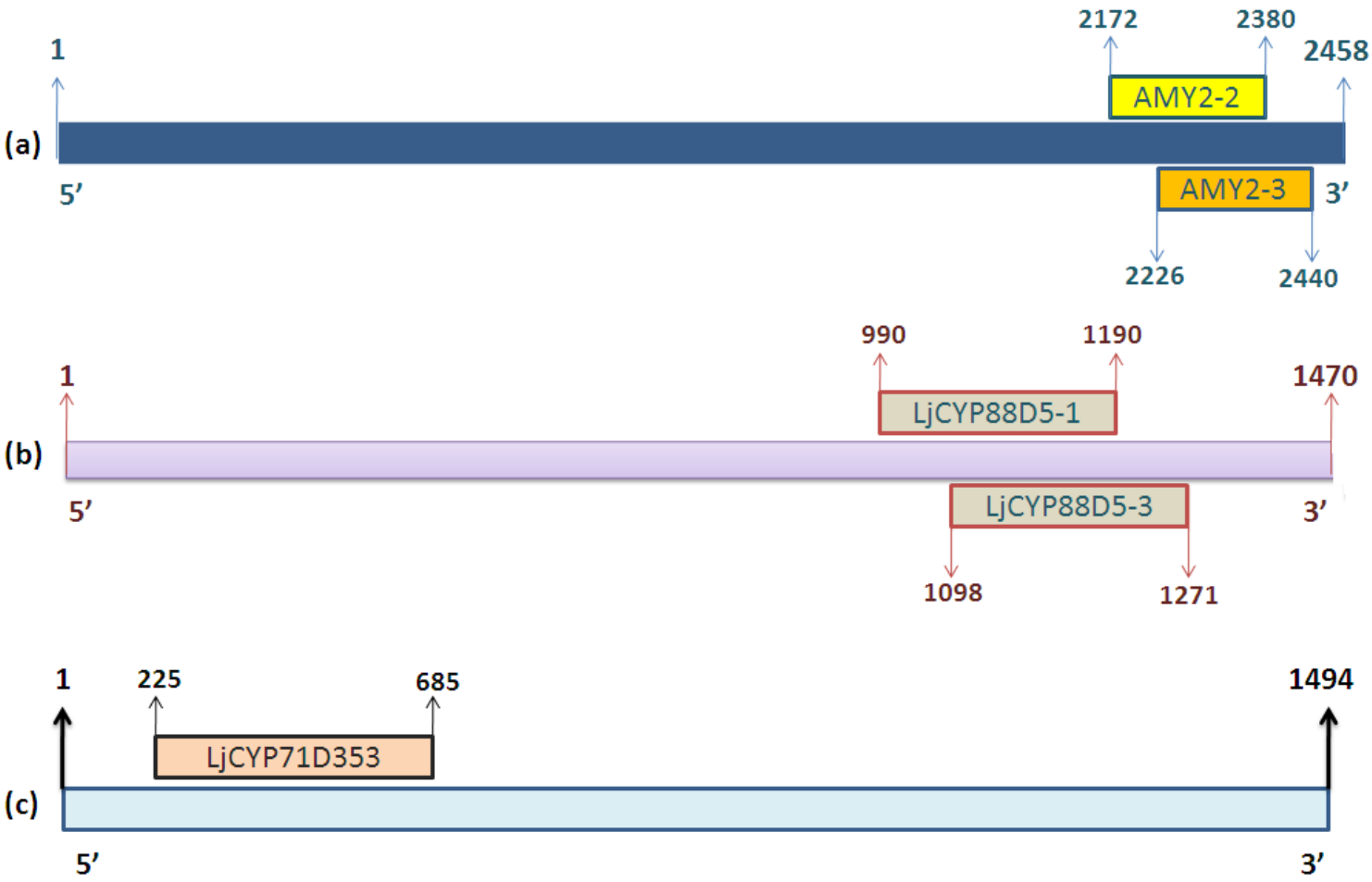


Figure 15. The PCR amplicons located at the corresponding full-length clones of *AMY2* cluster, (a) two PCR amplicons, one at the nucleotide position 2172 and one at the nucleotide position 2226 of *AMY2* gene, (b) the PCR segment at the nucleptide position 238 of *LjCYP71D353* gene and (c) two PCR fragments, one at the nucleotide position 990 and the other at the nucleotide position 1098 of *LjCYP88D5* gene.

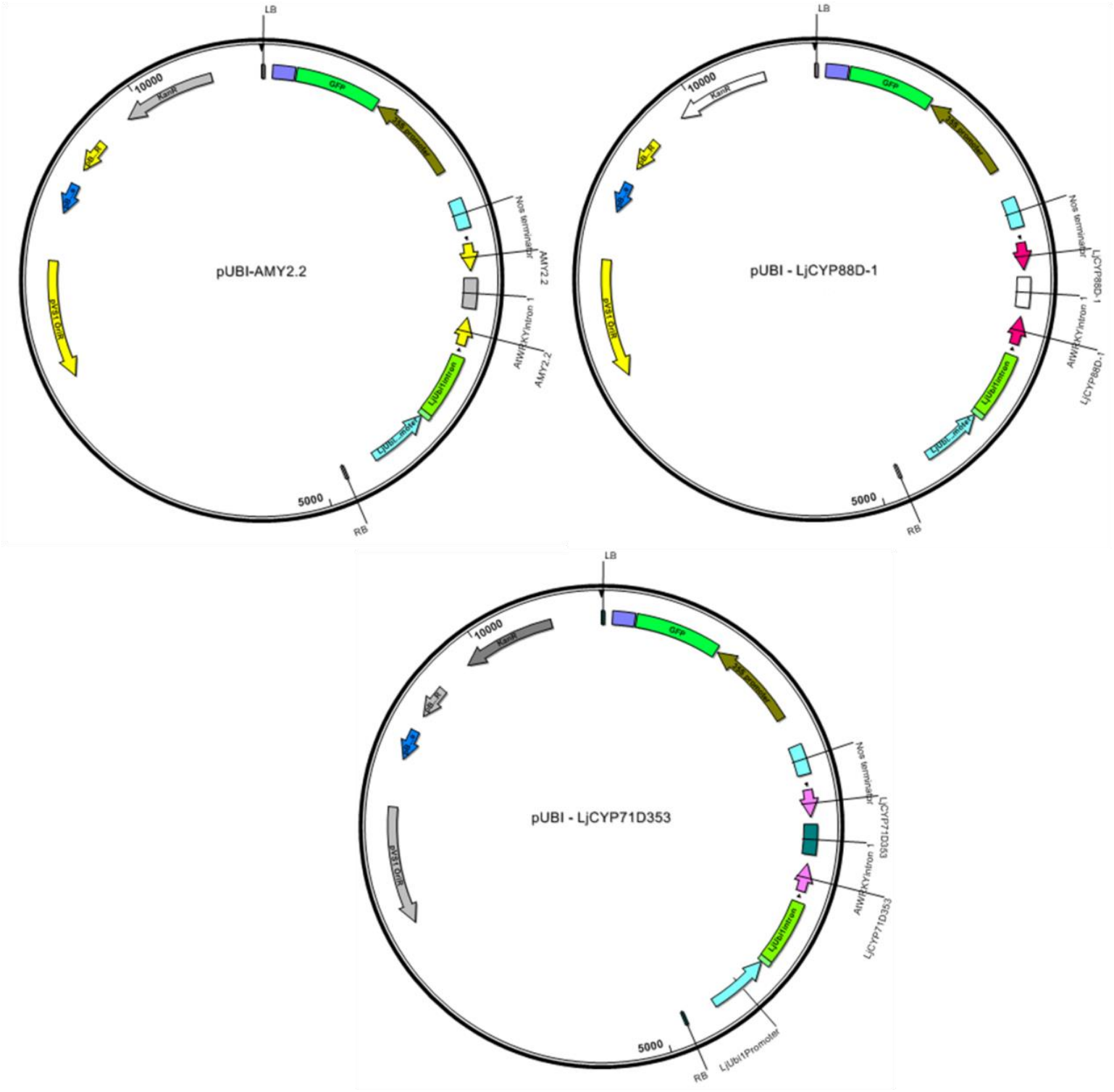


Figure 16. The binary plasmids pUBI-AMY2, pUBI-LjCYP71D353 and pUBI-LjCYP88D5 used for generation of transgenic *L. japonicus* roots. Plasmids contain between the right (RB) and left borders (LB): the LjUbi1 promoter, the LjUbi1 exon, the LjUbi1 intron, the attb1 recognition site of LB clonase, the PCR amplicons of the corresponding gene in 5′ – 3′ and in 3′ – 5′ direction on both sides of the AtWRKY intron1 the 2nd attb1 site of LB clonase recognition and the Nos terminator for the expression of the silencing cassette, and the 35S promoter fused with the GFP gene for the expression of the reporter gene.

Following transformation the root system was observed under the microscope for GFP production (Figure 17) and wild type root were removed.

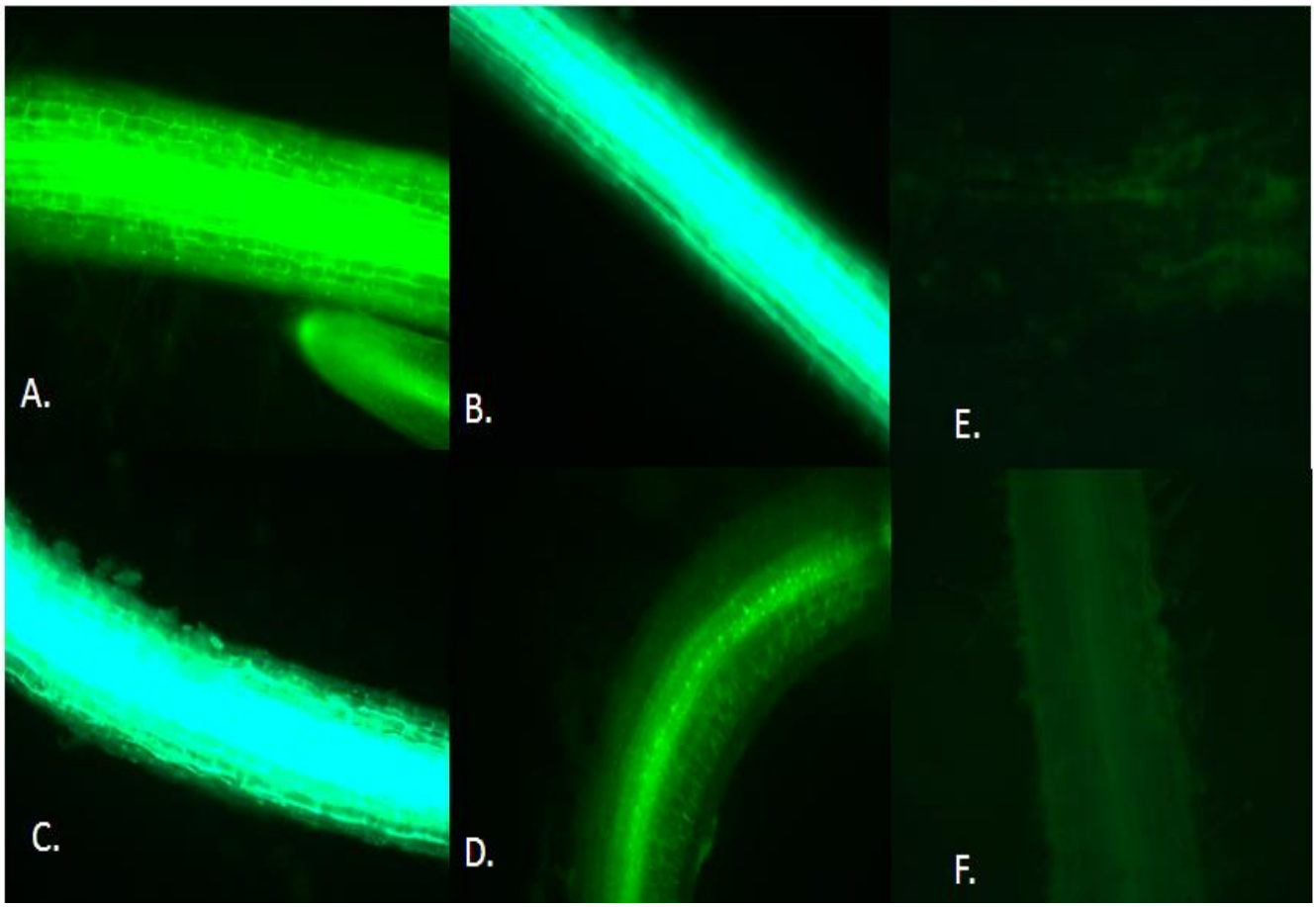


Figure 17. Observation of GFP production in *L. japonicus* roots. (A, B, C, D): transformed roots. (E, F): non – transformed roots. The difference of GFP accumulation between transformed and non – transformed roots is conspicuous.

L. japonicus plants were grown for 20 days in order to produce their transformed root system. *A. rhizogenes* T-DNA genes are capable of inducing the hairy root phenotype (Figure 18).

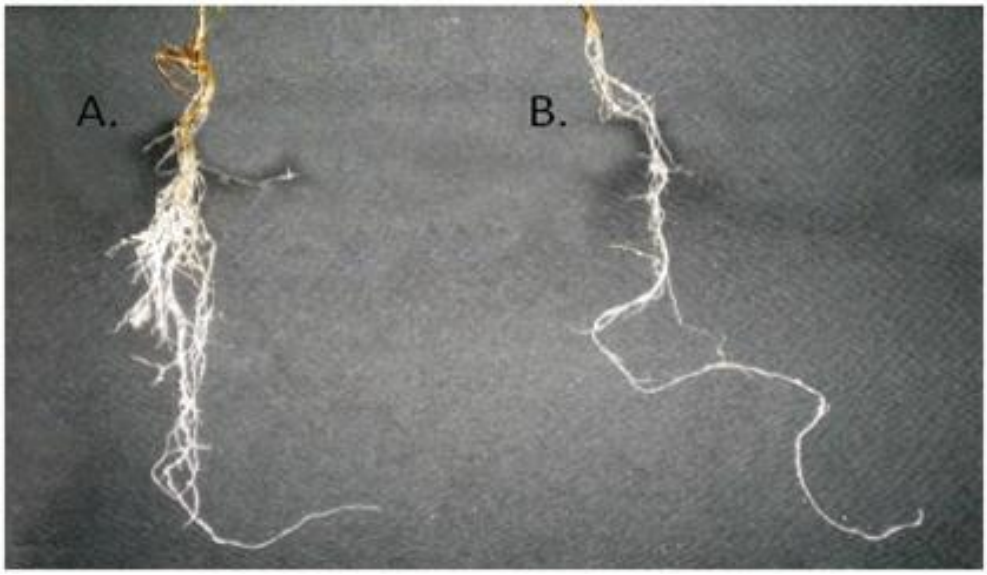


Figure 18. (A) *L. japonicus* transformed root 28 days old, with the hairy root phenotype, (B) *L. japonicus* non – transformed root 28 days old.

Transgenic roots were collected, total RNA was extracted and reverse transcribed, the concentration was normalized between the samples, and Real-time experiments were carried out. Significantly decreased levels of *AMY2* transcript were detected in almost all of these transgenic plants (three plants exhibiting different degree of silencing are shown in Figure 19a). No obvious effects were observed with regard to the nodulation process (i.e. nodule number at 20 and 40 dpi). Interestingly, at 40 dpi a more severe retardation of the rate of hairy-root growth was recorded than at 20 dpi. Transgenic roots with reduced levels of *LjCYP71D353* or *LjCYP88D5* transcripts (Figure 23b, c respectively) were also obtained. It was found that silencing of *LjCYP88D5* or *LjCYP71D353* caused no obvious effects on plant development or nodulation. No different effects among the two different silencing constructs of *AMY2* and *LjCYP88D5* were noticed.

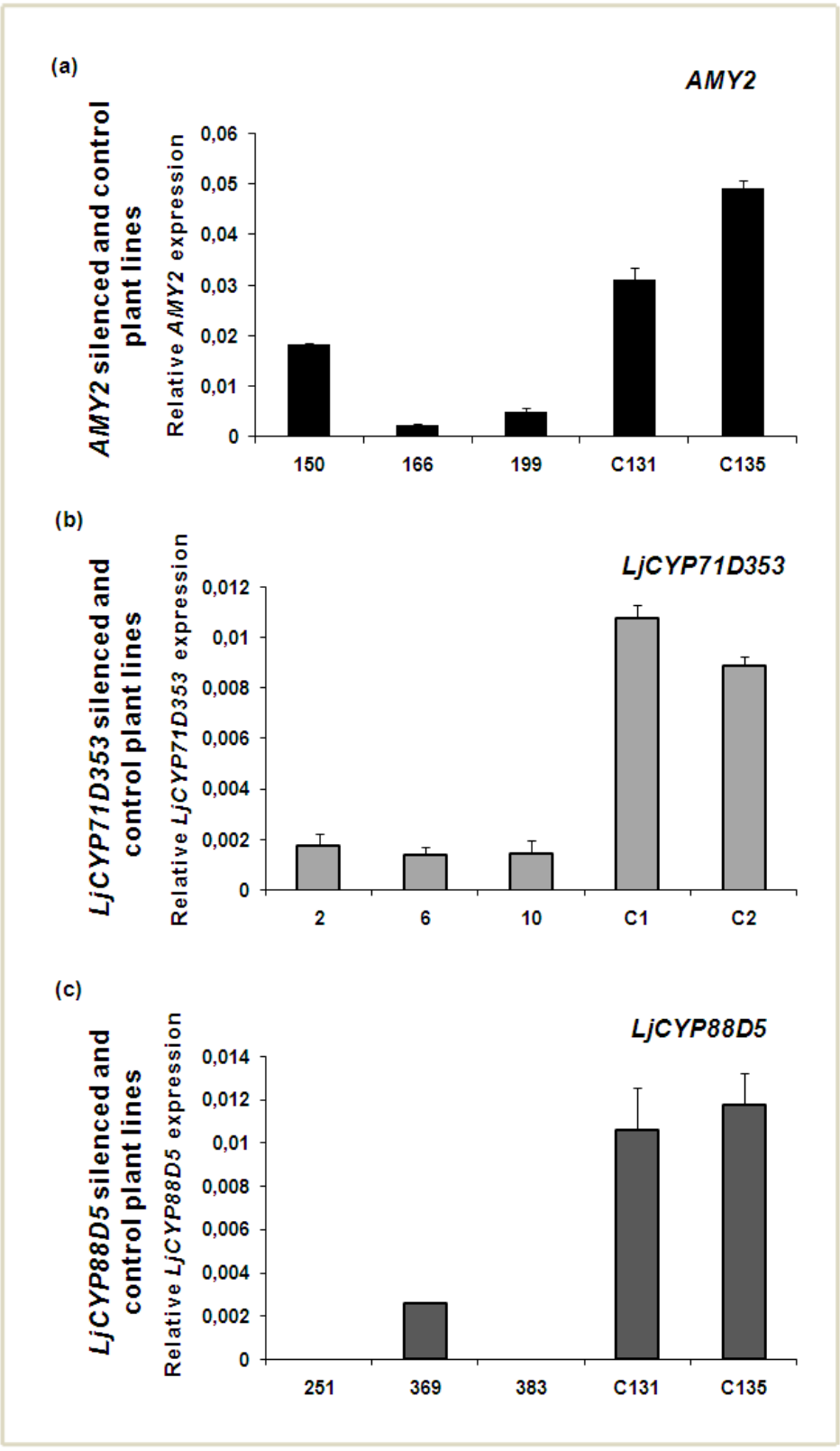


Figure 19. Plants silenced either for *AMY2* (a), *LjCYP71D353* (b) or *LjCYP88D5* (c) were generated by *A. rhizogenes* mediated transformation. Three representative plants for each hairpin construct (150, 166, 199), (2, 6, 10) and (251, 369, 383) with varying levels of silencing for *AMY2*, *LjCYP71D353* and *LjCYP88D5* respectively, all presenting statistically significant gene expression when compared to control plant lines (C131, C135, C1, C2) (*t*-test, $P < 0.01$), are shown. Error bars represent ± SE of means of technical repeats (n = 3).

C.1.4.2. *Agrobacterium tumefaciens* mediated plant transformation

In order to further examine the physiological role of the *AMY2* gene cluster, stably transformed lines silenced for *AMY2* were also obtained via *A. tumefaciens* transformation. The PCR amplicon *AMY2-2* that was used for hairy root transformation was also used for *A. tumefaciens* mediated transformation. *AMY2-2* amplification product was cloned into pHannibal vector in the antisense and sense direction. The silencing cassette was then excised, ligated into pBluescript plasmid and sequenced to verify the proper incorporation of *AMY2-2* fragments. Finally, the silencing construct was excised with the appropriate enzymes and cloned into pCambia 1300 (Figure 20). The *A. tumefaciens* strain AGL1 was transformed with the resultant binary plasmid pCambia 1300 – AMY2-2 and used for regeneration from *L. japonicus* root explants. L. japonicus explants were transformed following procedures previously described (Lombari *et al*., 2003; Barbulova and Chiurazzi, 2005).

A. tumefaciens strains that contain the Tumor inducing plasmid (Ti) are the causal agents of crown gall disease. Hygromycin was used for growth and differentiation of the transgenic calli. Several plant growth hormones were used for regeneration of the plant organs from the transgenic calli.

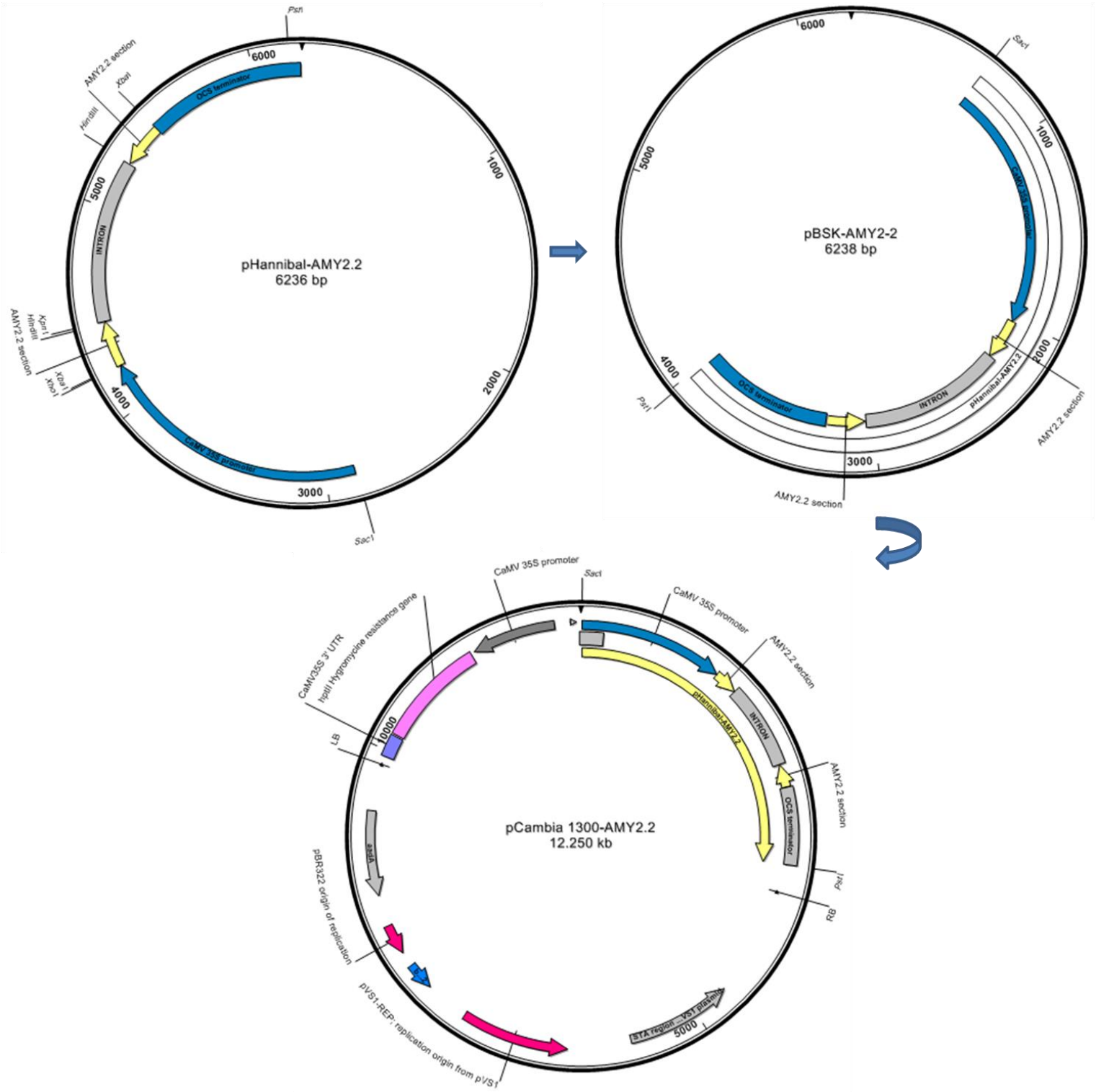


Figure 20. Depiction of the mediated cloning steps for the creation of the appropriate silencing constructs. The binary plasmid pCambia 1300-AMY2-2 carries the silencing cassette under the CaMV 35S promoter control and the hygromycin resistance gene under the 35S promoter control for the selection of transformed explants.

Five weeks after planting small segments of plant roots were excised and genomic DNA was isolated. PCR reactions were conducted in order to prove whether the plants are transformed. Subsequently, total RNA was isolated from the identified transformed plants; reverse transcribed and Real-time experiments were carried out. Four transformed plant lines with reduced levels of *AMY2* transcript were obtained (Figure 21). These plants did not flower (and so it was unable to obtain seed) and consistently exhibited a short, stunted root phenotype, indicating that silencing of *AMY2* in stably transformed plants has clear effects on growth and development.

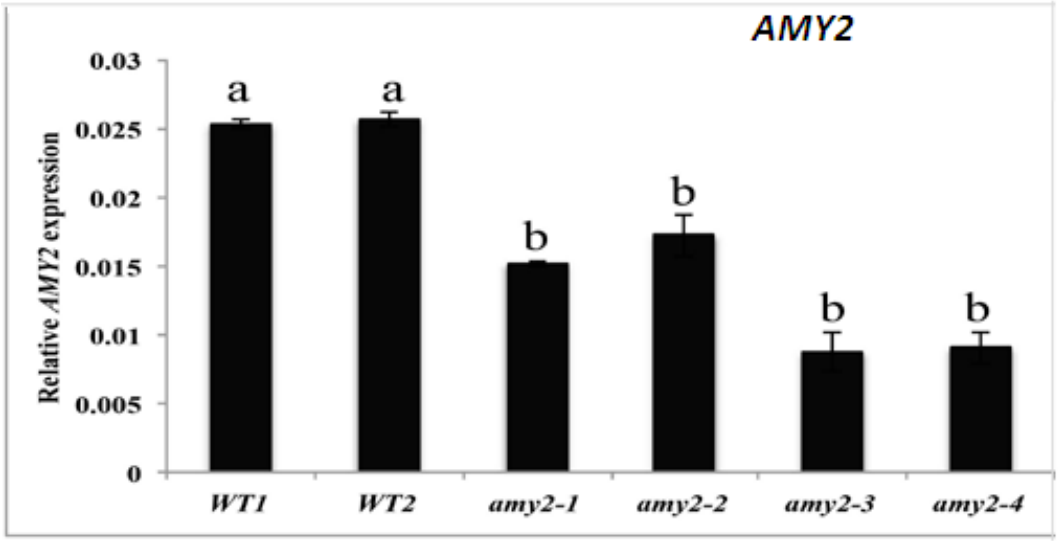


Figure 21. The reduced *AMY2* transcript levels of the four transformed plant lines generated by *A. tumefaciens* – mediated transformation where compared with wild type plants (WT1, WT2). Statistical comparisons within plant lines were performed by Duncan tests ($a<0.05$). Indicator letters in common denote lack of significant difference. Bars indicate standard errors of means (n=3).

C.1.5. *AMY2* cluster genes are epigenetically regulated

Based on our results indicating strong co-ordinated regulation of the expression profile of the *AMY2* cluster genes, their expression profile was investigated in the silenced plant lines. Real-time experiments were carried out using specific primers for *LjCYP71D353*, *LjCYP88D5* and *LjSDRt* and, strikingly, it was observed that the co-ordinate regulation of the cluster genes was manifested in a very unusual fashion. Thus, a decrease in *AMY2* transcription levels due to *A. rhizogenes* – mediated gene silencing, in all plant lines examined, was accompanied by significant reductions in the transcript

levels of other cluster genes, namely *LjCYP88D5, LjCYP71D353* and *LjSDRt* (Figure 22 a, b, c respectively).

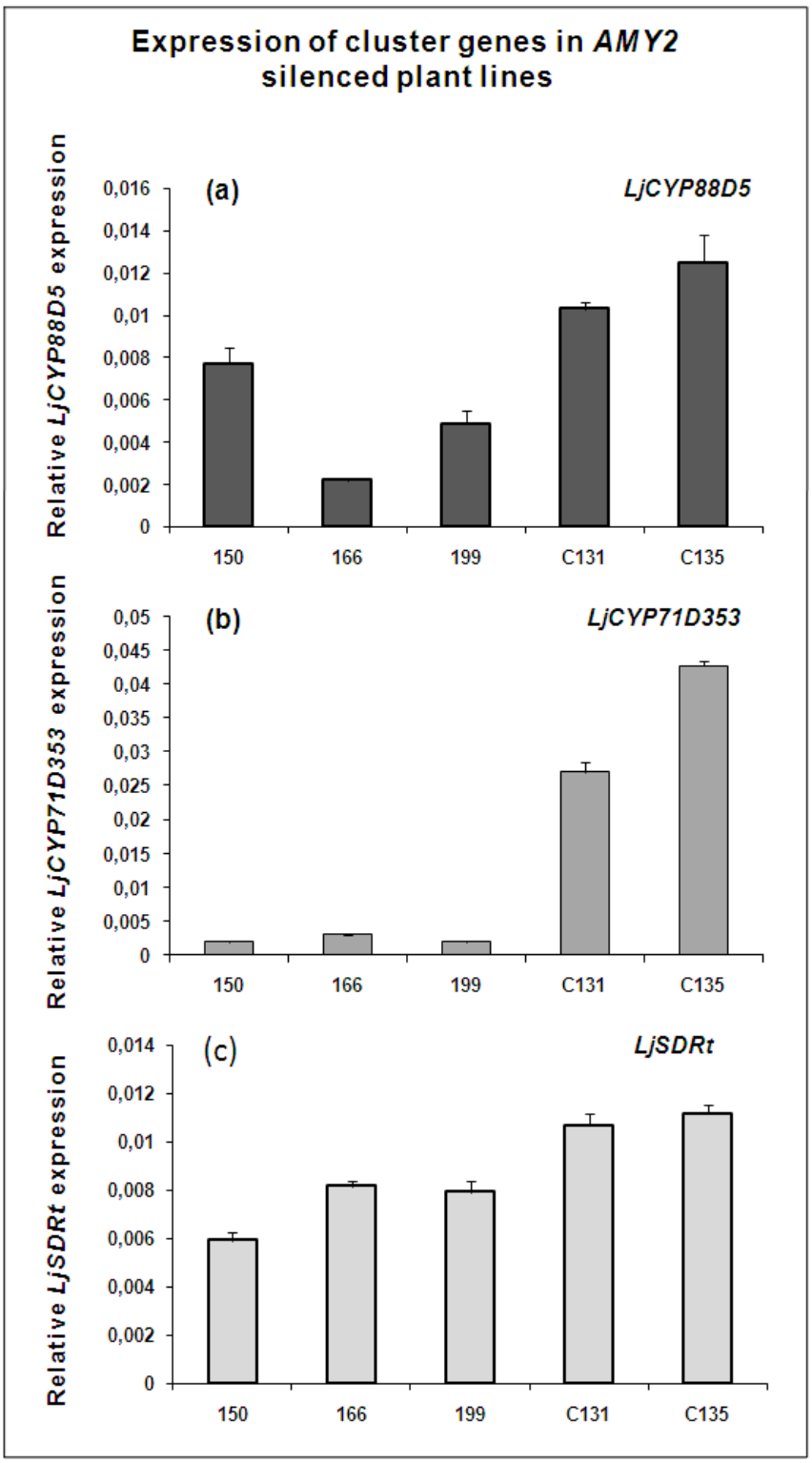


Figure 22. *AMY2* cluster gene expression in *L. japonicus* hairy-root tissues. Plants silenced for *AMY2* were generated by *A. rhizogenes* mediated transformation. Expression levels of *LjCYP88D5* (a), *LjCYP71D353* (b) and *LjSDRt* (c) are down-regulated in *AMY2* silenced roots (150, 166 and 199) when compared to their expression in control plant lines (C131, C135). Error bars represent standard error of means of technical repeats (n=3).

Similarly, plant lines silenced for *LjCYP88D5* had reduced transcript levels for *AMY2* (Figure 23a), *LjCYP71D353* (Figure 23b) and *LjSDRt* (Figure 23c) and plant silenced lines for *LjCYP71D353* had reduced transcript levels for *AMY2* (Figure 24d), *LjCYP88D5* (Figure 24e) and *LjSDRt* (Figure 24f).

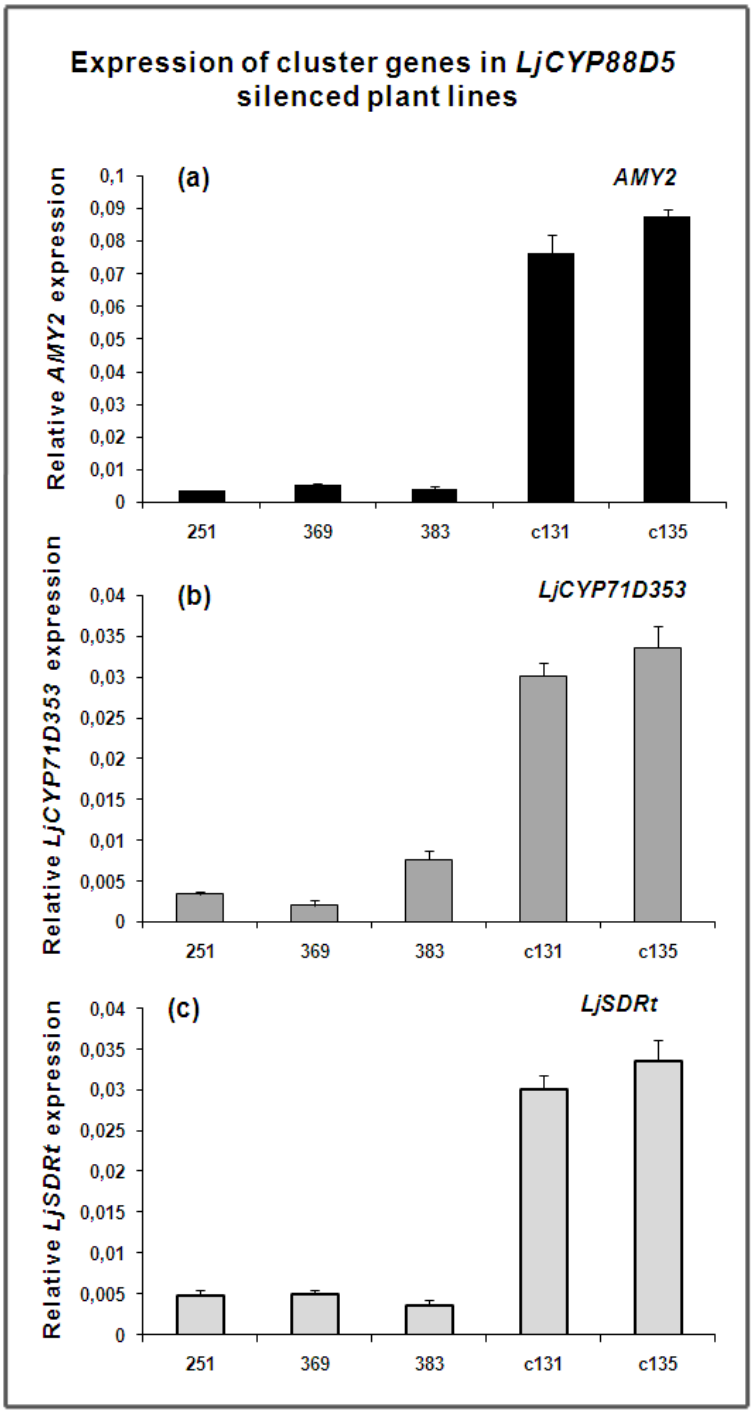


Figure 23. *AMY2* cluster gene expression in *L. japonicus* hairy-root tissues. Plants silenced for *LjCYP88D5* were generated by *A. rhizogenes* mediated transformation. Expression levels of *AMY2* (a), *LjCYP71D353* (b) and *LjSDRt* (c) are down-regulated in *LjCYP88D5* silenced roots (251, 369 and 383) when compared to their expression in control plant lines. Error bars represent standard error of means of technical repeats (n=3).

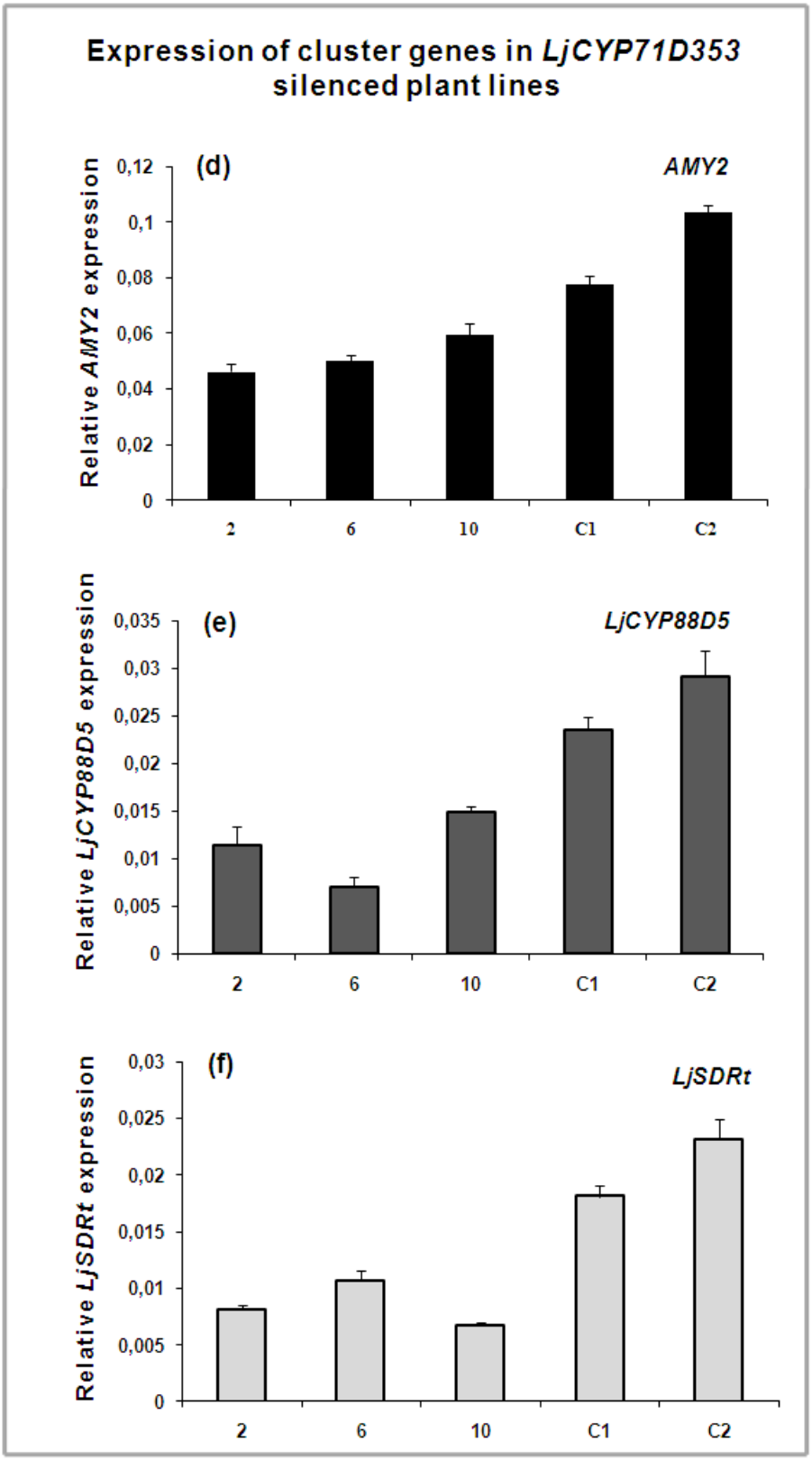


Figure 24. *AMY2* cluster gene expression in *L. japonicus* roots silenced for *LjCYP71D353* generated by *A. rhizogenes* mediated transformation. In accordance with the expression pattern of cluster genes in *AMY2* and *LjCYP88D5* silenced lines, similar expression profile of cluster genes is observed in *LjCYP71D353* silenced roots (2,6 and 10) (d, e, f). Error bars represent standard error of means of technical repeats (n=3).

These experiments were repeated for these three genes over a period of more than two years and it was always observed such a "spreading" of transcription silencing in the cluster. Furthermore, this *cis* spreading of transcript level repression was also observed in *AMY2* silenced stably transformed plants (Figure 25). The transcript levels of *LjSDRt* and *LjCYP71D353* (but not of *LjCYP88D5* in this case) were both significantly reduced in lines amy2-3 and amy2-4, while those of *LjSDRt* were also reduced in lines amy2-1 and amy2-2.

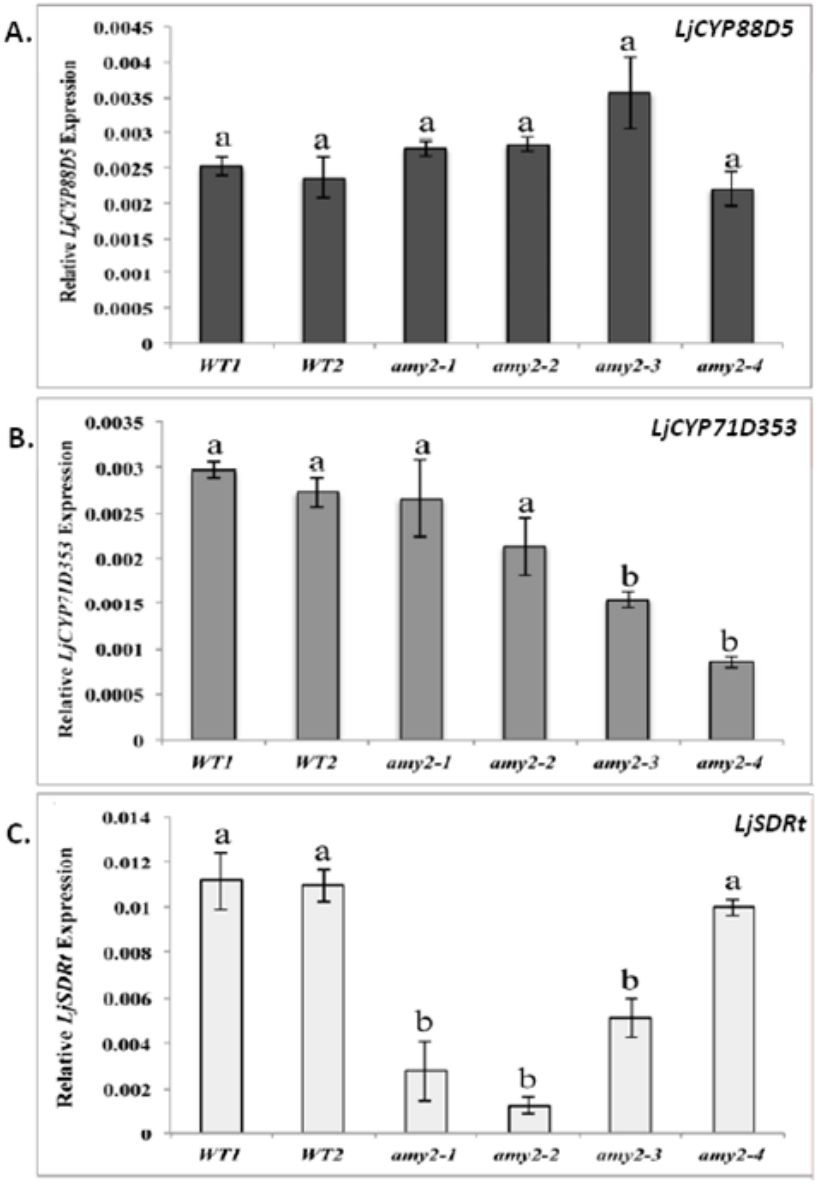


Figure 25. Expression levels of *LjCYP88D5* (A), *LjCYP71D353* (B) and *LjSDRt* (C) are variably down-regulated in *AMY2* silenced roots. Statistical comparisons within plant lines were performed by Duncan tests ($a<0.05$). Indicator letters in common denote lack of significant difference. Bars indicate standard errors of means (n=3).

To further investigate this phenomenon bisulfite sequencing was used to determine the DNA methylation levels of *LjSDRt* and *LjCYP71D353* genes in the wild type and in two *AMY2* silenced plants, amy2-3 and amy2-4. Genomic DNA was isolated from roots of wild type plants and, amy2-3 and amy2-4 silenced plant lines and treated with sodium bisulfite. PCR reactions were carried out using specific primers for *LjSDRt* and *LjCYP71D353* genes (Table A1). The PCR amplicons were cloned into pGEM® T-easy vector and sequenced to determine levels of methylation (percentage of all methylated deoxycytidine 5mdC in relation to the total deoxycytidine content in all 10 clones in mutant). A significant increase in the degree of DNA methylation was detected in both of the genes with reduced expression levels in the silenced plant lines compared to the methylation level of the genes in wild type plants (Table 4).

%DNA methylation		
	Genes	
Plant genotype	*LjSDRt*	*LjCYP71D353*
Wild type 1	48	78
Wild type 2	56.6	80.7
amy2-3	79*	100*
amy2-4	40	100*

Table 4. Degree of DNA methylation of *LjSDRt* and *LjCYP71D353* in wild type and stable silenced plant lines. (*) indicates statistical significant increase in DNA methylation degree between the silenced and wild type genotypes (*t* - test, $P < 0.01$).

Moreover, in the plant line amy2-4 in which the transcript levels of *LjSDRt* are not significantly reduced, the degree of DNA methylation was the same as in wild type plants. Our cautious interpretation of these results is that the hairpin-derived siRNAs introduced by two different experimental approaches (i.e. hairy-roots and stable transformation) induce RNA-directed DNA methylation (RdDM; Wassenegger *et al.*, 1994; Dalakouras and Wassenegger, 2013) thus promoting transcriptional gene silencing (TGS). The spreading of RdDM into the adjacent regions by TGS transitivity would silence the adjacent genes.

C.2. Role of the β-amyrin synthase gene *βAS* in *M. truncatula*

C.2.1. Gene expression in response to hormone treatment

MeJA is a well-known elicitor for the production of many plant secondary metabolites and regarding the triterpene biosynthesis, it induces the expression of the *M. truncatula βAS* gene (Suzuki *et al*., 2002). The effect of the other two different types of plant hormones, BA and 2,4-D, on *M. truncatula βAS* gene expression was investigated. *M. truncatula* seedlings were grown on Petri dishes containing MS substrate and then were transferred to Petri dishes containing MS supplemented with 2,4 D, BA or MeJA. Total RNA was isolated from the roots, cDNA were prepared and normalized, and real-time experiments were carried out. Except for MeJA, the BA and 2,4-D hormone did not significantly altered the expression of *βAS* gene (Figure 26).

As it is mentioned above, in *βAS1* gene cluster the cytochrome P450 gene *MtCYP88D1* is adjacent to *βAS1*. The *in silico* analysis of the AC144538 genomic clone of *M. truncatula* revealed that there is a gap in the genomic locus of *MtCYP88D1*. The study of *MtCYP88D1* expression, although valuable, is not feasible because the sequence of *MtCYP88D1* locus is not as yet available.

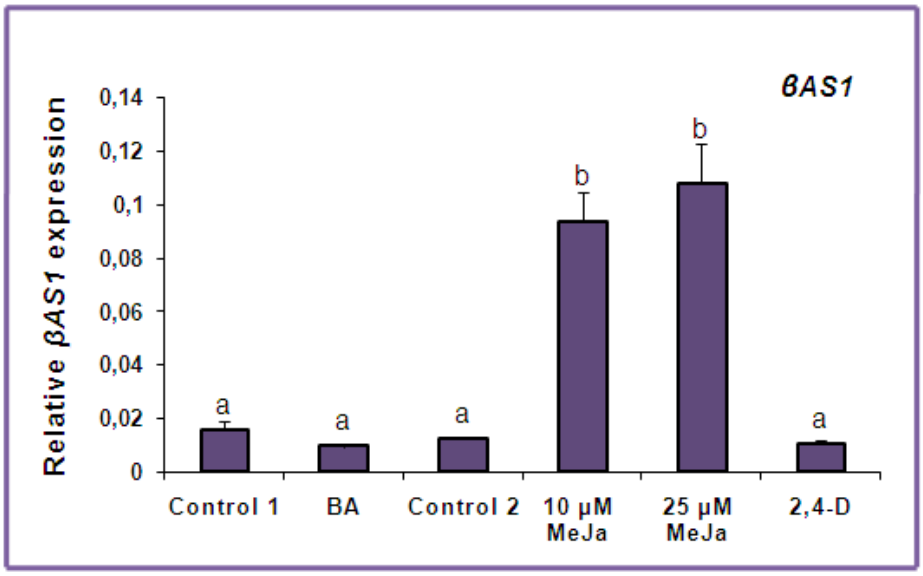


Figure 26. *βAS* expression in *M. truncatula* root tissues (14 days old) subjected to various to various exogenously applied plant hormones. Data from a single representative experiment are presented; two experimental repeats yielded similar results. Statistical comparisons were performed by Duncan tests ($a < 0.05$). Indicator letters in common denote lack of significant difference. Bars indicate + SEM ($n = 3$).

C.2.2. RNAi silencing of *βAS* in *M. truncatula* roots

Transgenic *M. truncatula* roots with reduced *βAS* transcript levels were successfully acquired with the *A. rhizogenes* mediated transformation protocol. The PCR segment of *βAS* gene used in the LB clonase reaction, as well as the resultant binary pUBI-βAS1 plasmid are shown in Figure 27.

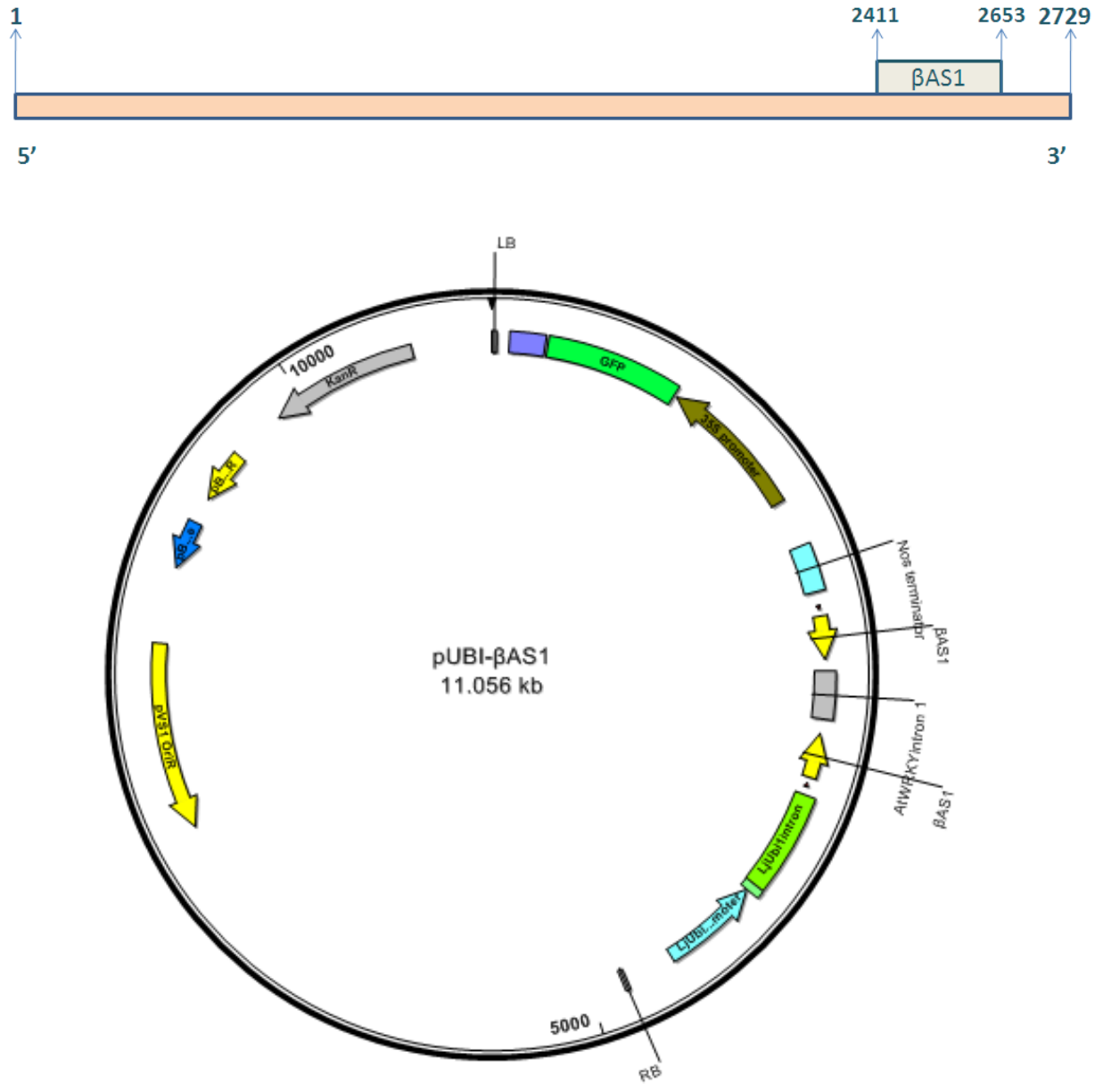


Figure 27. (a) The PCR *βAS1* amplicons is located at the nuclotide position 2411 of *βAS* gene. (b) The binary plasmid pUBI-βAS1 contain between the right (RB) and left borders (LB): the LjUbi1 promoter, the LjUbi1 exon, the LjUbi1 intron, the attb1 recognition site of LB clonase, the PCR amplicons of the *βAS1* gene in 5' – 3' and in 3' – 5' direction on both sides of the AtWRKY intron1 the 2nd attb1 site of LB clonase recognition and the Nos terminator for the expression of the silencing cassette, and the 35S promoter fused with the GFP gene for the expression of the reporter gene.

By analogy with *L. japonicus*, total RNA was isolated, cDNA were prepared and Real-time experiments were conducted. Significantly decreased levels of *βAS* transcripts were detected in five transgenic plants exhibiting different degree of silencing when compared to control plant lines (inoculated with *A. rhizogenes* carrying the pUBI-GWS-GFP empty vector) (Figure 28).

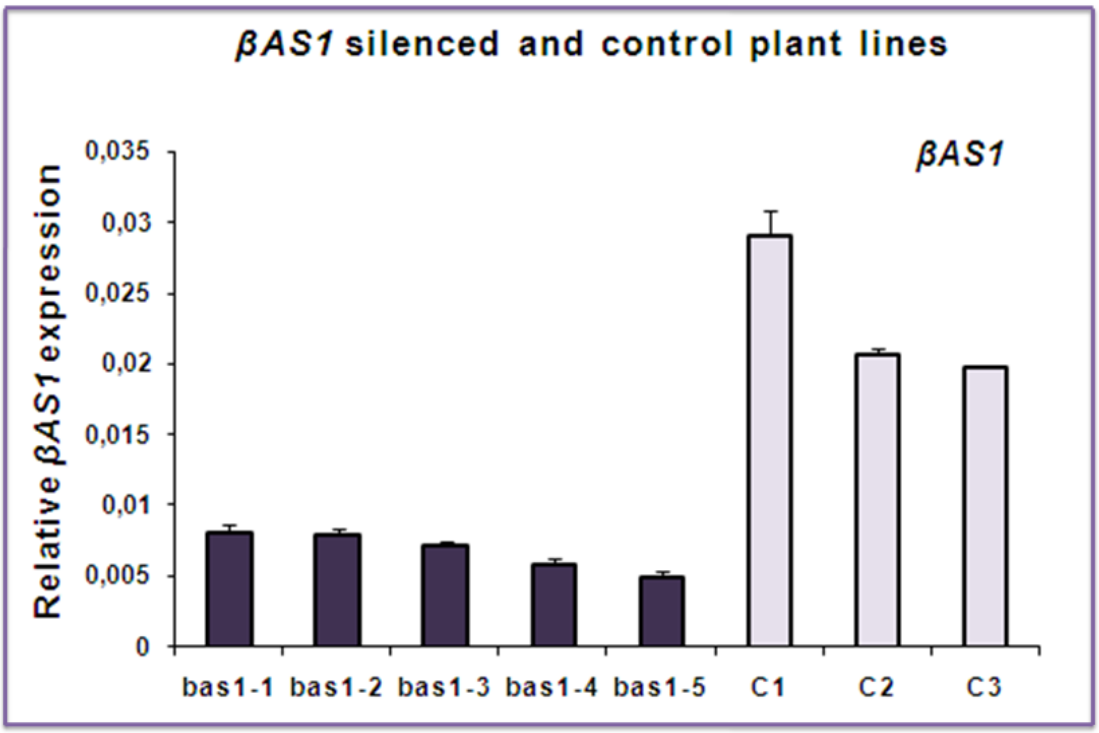
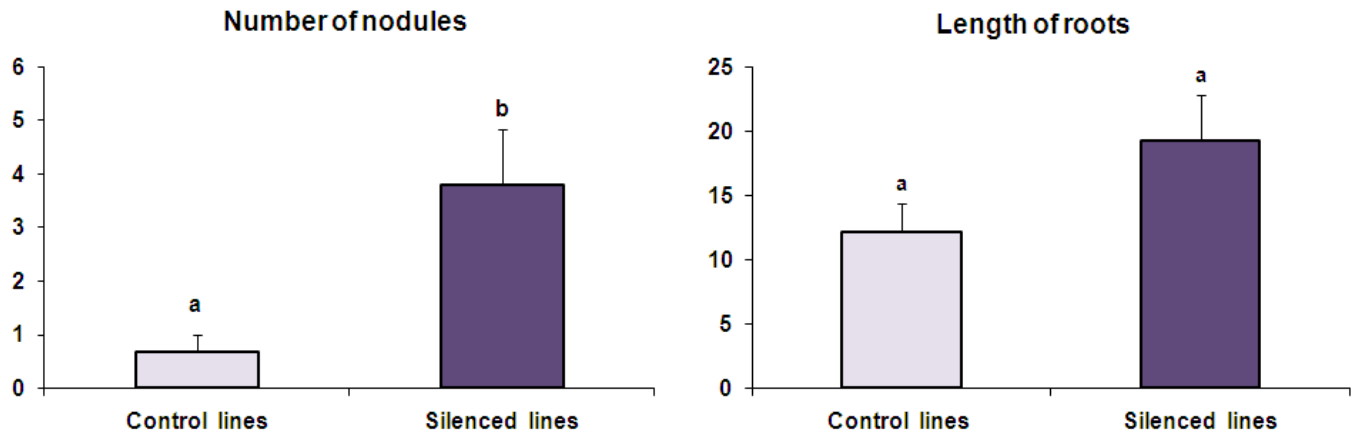


Figure 28. *βAS* expression in *M. truncatula* hairy root cultures. Plants with reduced *βAS* transcript levels were generated by *A. rhizogenes* mediated transformation. Five representative plants for *βAS1* hairpin construct, with varying levels of silencing, all presenting statistically significant gene expression when compared to control plant lines (*t*-test, $P < 0.01$), are shown. Error bars represent ± SE of means of technical repeats ($n = 3$).

The number of nodules and the length of transgenic roots in *βAS1* silenced plant lines were determined 20 days after inoculation with *S. meliloti*. The difference in root length of the silenced and control lines is not statistically significant. Interestingly, the *βAS1* silenced lines showed an increased nodule number when compared to control lines (Figure 29). This enhanced nodulation in the *βAS1* silenced plant lines needs further investigations.



	Nodules / plant 20 dpi	Root length / plant
Plant lines		
Silenced for βAS	2.28 ± 1.02	19.28 ± 3.5
Control (transformed with empty vector	0.67 ± 0.33	12.125 ± 2.24
	s*	ns*

Figure 29. Effect of *βAS* on the number of nodules and length root at 20 days post – inoculation of *M. truncatula* with *S. meliloti*; indicator letters denote the significant difference, s, significant, ns, not significant, *t*-test, α = 0.05.

C.3. Investigating the role of *OSC3* in developing roots and nodules

C.3.1. *OSC3* is exclusively expressed in roots and nodules of *L. japonicus*

Expression of *OSC3* was investigated in different plant organs by Q-PCR and *in situ* hybridization. The accumulation of *OSC3* transcripts was observed in all of the developmental phases of the root and nodule tissues examined (Figure 30). By contrast, *OSC3* transcripts were not detected in leaves. In uninfected plants, the highest levels of *OSC3* expression were detected in 28-d-old roots, whereas, in plants infected with *M. loti* (strain R7A), the highest level of gene expression was detected in the roots and nodules at 14 dpi. The expression of *OSC3* was highly induced in plants infected with *M. loti* compared with uninfected plants of the same developmental stage. The increase was greatest in younger plants, most notably in roots at 14 dpi, where an increase of more than 10-fold was observed (Figure 30). Furthermore, developing nodules at 14 dpi accumulated higher levels of *OSC3* transcripts when compared with mature nodules at 28 dpi (Figure 30b).

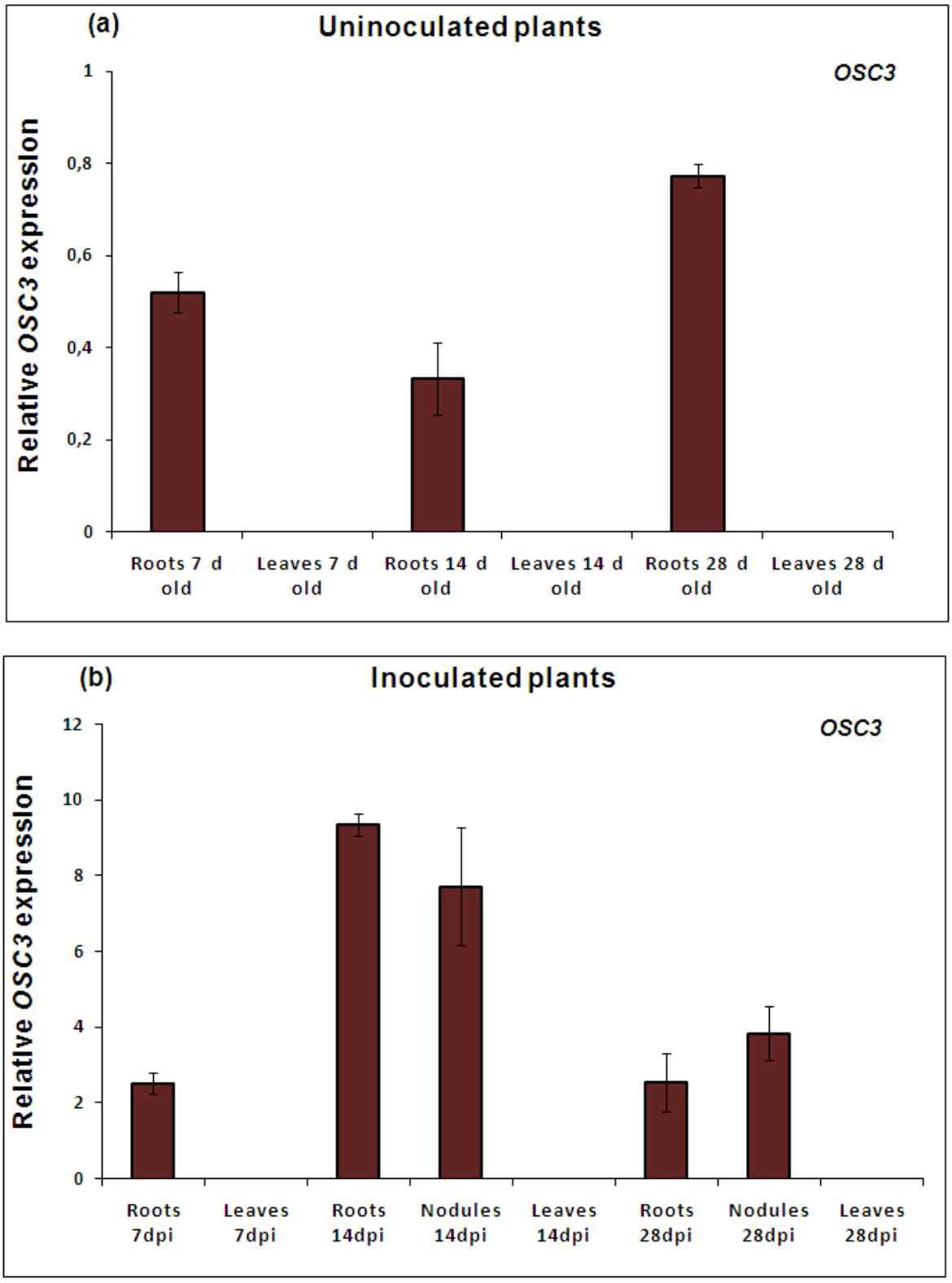


Figure 30. Accumulation of *OSC3* gene transcripts in various organs and developmental stages of uninfected plants (a) and plants infected (b) with *M. loti* R7A on the day of sowing. Uninfected and infected plants are of the same age at the 7-d-old / 7 d post infection (dpi), 14-d-old / 14 dpi and 28-d-old / 28 dpi stages, respectively. Total RNA was reverse transcribed, normalized to the constitutively expressed gene of *UBQ* transcripts and subjected to real-time PCR. Relative OSC3 expression was measured with respect to *UBQ* transcripts. Mean values ± SD are shown (n = 3).

The substantial increase in *OSC3* expression in the roots of *M. loti*-infected plants and the high levels of *OSC3* expression detected in nodules led us to examine the spatial localization of *OSC3* gene transcripts in order to investigate the possible physiological role of lupeol biosynthesis in symbiosis. *In situ* hybridization was conducted to determine the nodular tissue in which the *OSC3* gene is expressed. Thin sections of 14- and 28-dpi-old *L. japonicus* nodules were hybridized with either antisense or sense DIG-11-rUTP-labelled RNA. In both developing (14 dpi) and mature (28 dpi) nodules, *OSC3* transcripts were localized in the vascular bundles, nodule inner cortex cells and uninfected cells of central tissue (Figure 31). Hybridization signals were not detected in the outer cortex parenchyma or in infected cells in any of the nodules examined (Figure 31b). In the vascular bundles, the signal was detected in phloem cells, but not in the xylem (Figure 31e).

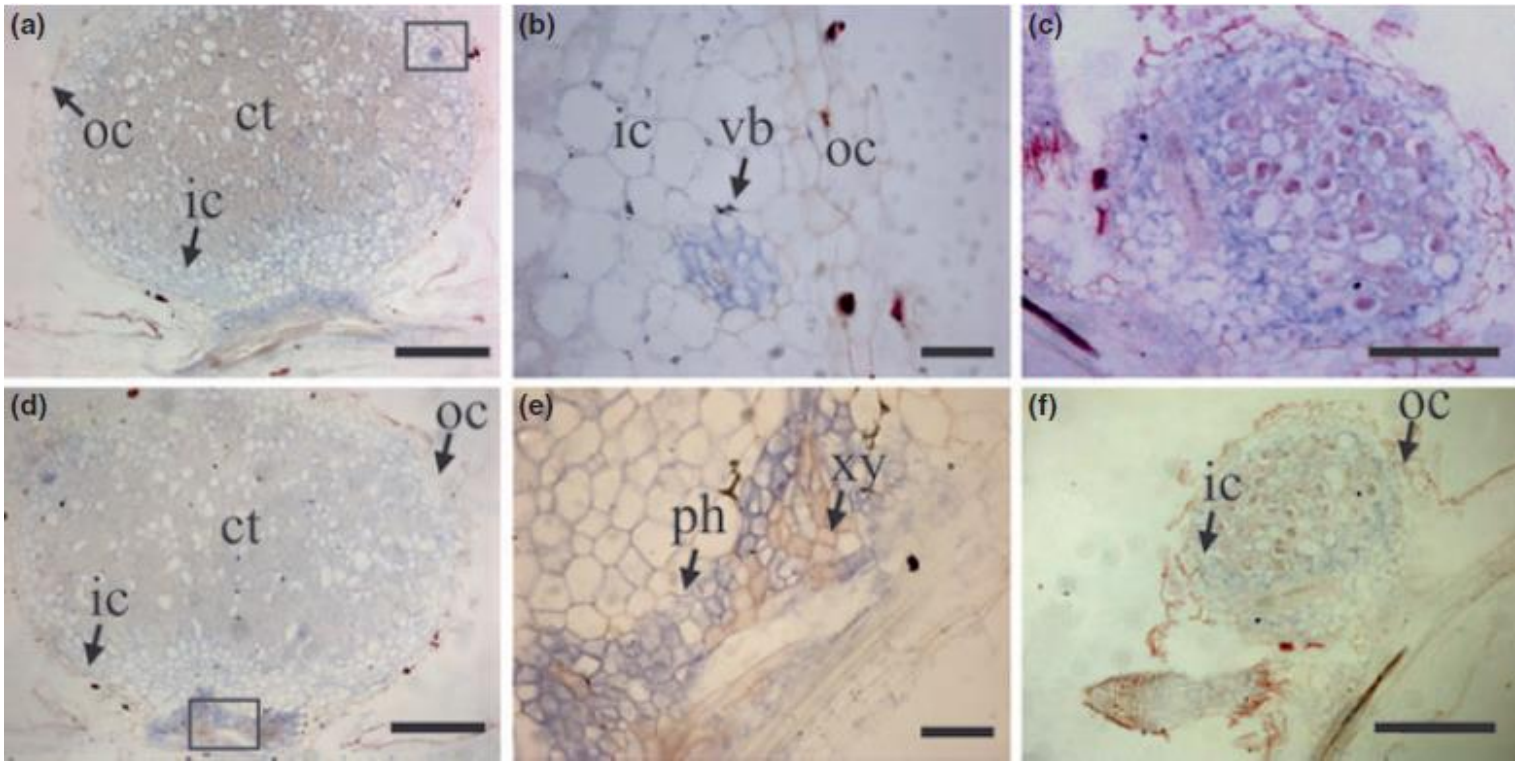


Figure 31. *In situ* hybridization of *OSC3* gene transcripts in mature 28 d post-inoculation (dpi) (a, b, d, e) and developing 14 dpi (c, f) *L. japonicus* nodules. Thin sections (8 µm) were hybridized with DIG-11-rUTP-labelled antisense RNA, in vitro transcribed from a PCR product for *OSC3*. The hybridization signal was visualized using an alkaline phosphate reaction product (blue–purple color). In all cases, the hybridization signal was detected in the inner cortex (ic), phloem (ph) cells of vascular bundles (vb) and uninfected cells of the central tissue (ct). The hybridization signal was not detected in the infected cells of the central tissue, outer cortex parenchyma (oc) and xylem (xy). Boxes in (a) and (d) indicate enlarged areas as depicted in (b) and (e), respectively. Scale bar: 100 µm for (a, c, d, f); 10 µm for (b, e).

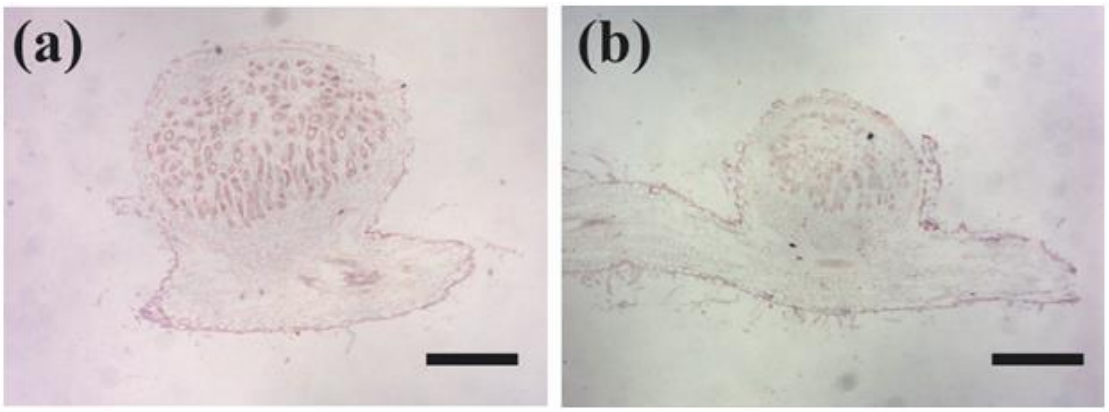


Figure 32. Negative control for the *in situ* hybridization of *OSC3* gene transcripts in mature 28 dpi (a) and developing 14 dpi (b) *L. japonicus* nodules. Eight µm thin sections were hybridized with DIG-11rUTR labeled sense RNA *in vitro* transcribed from of a PCR product for *OSC3*. No signal was seen with this control.

C.3.2. RNAi silencing of *OSC3* in *L. japonicus* roots

An *A. rhizogenes*-mediated gene silencing strategy was employed to investigate the role of *OSC3* in roots and nodule physiology. This system is highly amenable for functional analysis as *OSC3* shows organ-specific expression in roots and nodules. 18 plants were obtained that had developed a transgenic root system and were successfully transformed with the hairpin construct. Both DNA extraction (to verify hairpin integration) and RNA extraction (to perform gene expression analyses) were conducted using the same tissue of a single root. *OSC3* expression was determined quantitatively in these plants (lines 10, 23, 38, 65, 70 and 99 are shown in Figure 33) and in plants that had been transformed with the empty vector (lines C1 and C2) or with the *GUS* gene at 20 dpi with *M. loti* (Figure 33). The latter plant line had no obvious phenotype other than the expression of GUS and was included as a further control line. Plant line 89, also depicted in Figure 33, represents a plant line which was transformed with the hairpin construct but was not silenced. In the different independent plant lines that had been transformed with the gene silencing construct, *OSC3* expression could not be detected, with the exception of plant line 99, which had substantially reduced expression levels. *OSC3* expression was readily detectable in control plants (Figure 33). The analysis of independent plant lines that had each been transformed with the hairpin construct provides independent biological replicates, so enabling to ensure the reproducibility of the outcome of *OSC3* silencing. *OSC3*

expression was dramatically lower in the silenced plant lines than in control plants (*t*-test, a = 0.001).

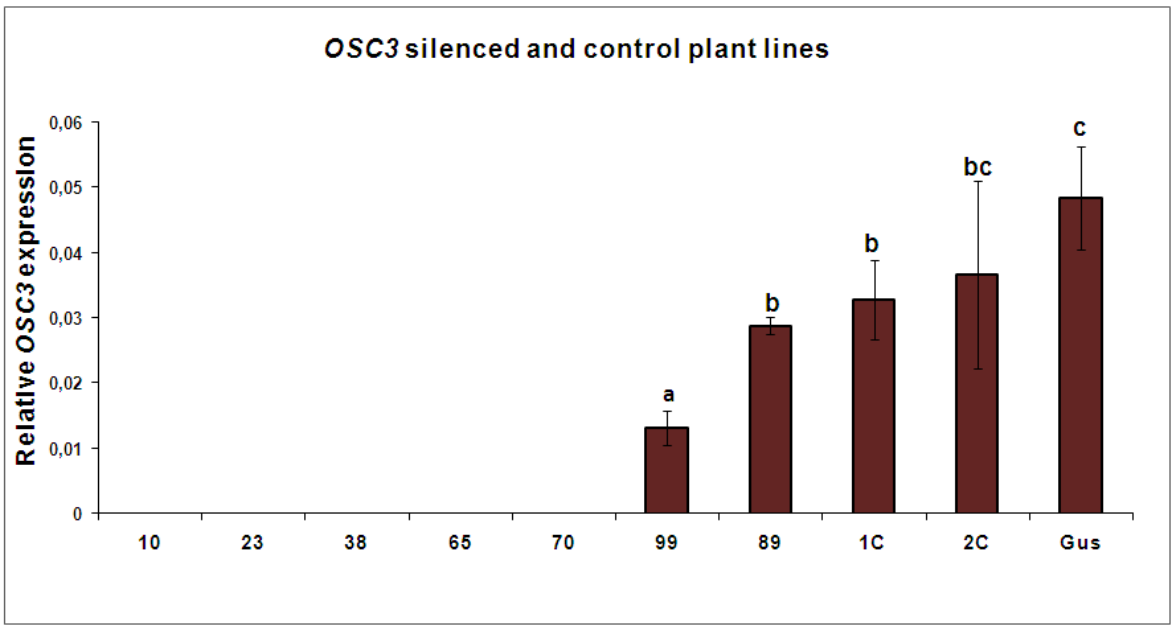


Figure 33. *OSC3* expression in plant lines found to be transformed for the hairpin structure (10, 23, 38, 65, 70, 99), in the non-silenced plant line (89) and in the control line transformed with the empty pCambia 1300 vector (1C, 2C) and pCambia 1303 vector (expressing GUS). Plant line 99 is only partially silenced. Total RNA isolated from plant roots, 20 d post-inoculation with *M. loti*, was reverse transcribed, normalized to the constitutively expressed gene of *UBQ* transcripts and subjected to real-time PCR. Relative *OSC3* expression was measured with respect to *UBQ* transcripts. Bars indicate standard errors of the means for three technical replicates (n = 3). Statistical comparisons within biological replicates (i.e. the individual silenced / control lines) were performed by Duncan tests (a < 0.05). Indicator letters in common denote a lack of a significant difference.

C.3.3. A role of lupeol in symbiosis

The effect of silencing of *OSC3* on the nodule number was then assessed. The numbers of nodules formed on roots of *OSC3*-silenced and control plants were counted 20 and 40 dpi with *M. loti*. At 20 dpi, silenced plants exhibited significantly more visible nodules than control plants (a = 0.05). However, the silenced and control plants showed similar numbers of nodules at 40 dpi (Table 5). At this later stage, the nodules had a healthy appearance in both silenced and control lines. This was further verified by estimating the efficiency of nitrogen fixation using an acetylene reduction assay, which indicated that there were no significant differences between the different plant lines.

	Nodules / plant	
Plant lines	20dpi	40dpi
Silenced for *OSC3*	2.05±0.54 (n=18)	3.94±0.54
Wild-type (transformed with empty vector)	1.21±0.28 (n=14)	3.85±0.37
	s*	ns*

Table 5. Effect of *OSC3* on the number of nodules at 20 and 40 d post-inoculation of *L. japonicus* with *M. loti*. *ns, non-significant, s, significant, *t*-test, α=0.05.

To further examine the role of *OSC3* in nodule formation and development, the expression levels of two early nodulin genes, *ENOD2* and *ENOD40*, were determined quantitatively in silenced and control plants. The *ENOD2* gene encodes a hydroxyproline-rich protein, and its expression is induced in specific tissue layers in nodules, that is in nodule parenchyma surrounding the infected zone. It is considered to be a "late" early nodulin gene and is used routinely as a developmental marker for progression through late symbiotic differentiation (van de Wiel *et al*., 1990; Niwa *et al*., 2001; Mitra and Long, 2004; Tirichine *et al*., 2006). *ENOD40* plays an important and, as yet, unidentified role in nodule formation and subsequent nodule organogenesis, and *ENOD40*-silenced plants exhibit poor nodulation, but it has no role in early infection events (Charon *et al*., 1999; Kumagai *et al*., 2006; Wan *et al*., 2007). *ENOD40* has also been used as a marker gene for nodule primordia initiation (Radutoiu *et al*., 2003; Murray *et al*., 2007). Interestingly, the expression of *ENOD40* was increased significantly in plant lines that were silenced for *OSC3* when compared with control plant lines (Figure 34). In plant line 99, which displayed partial silencing of *OSC3* expression, the expression levels of *ENOD40* were similar to those of control plant lines. Expression of *ENOD2* was slightly, but significantly, reduced in the transformed plant lines in all silenced plant lines examined. On the contrary, the expression levels of *ENOD2* were not affected in the partially silenced plant line 99 when compared with control lines.

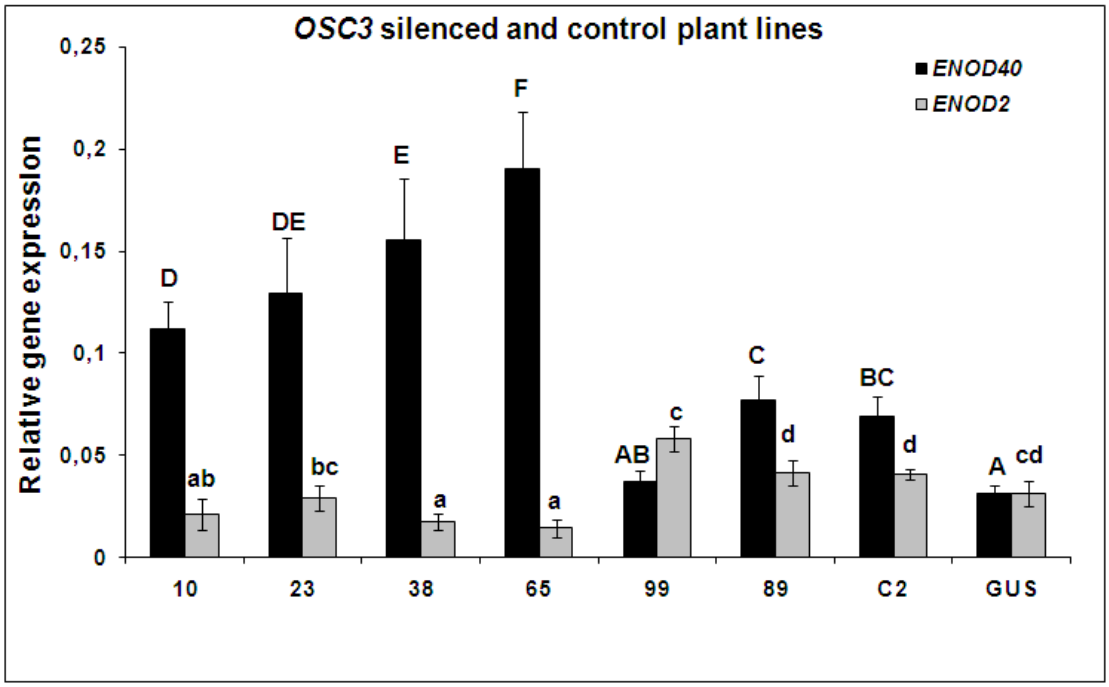


Figure 34. *ENOD40* and *ENOD2* expression in plant lines silenced for *OCS3* (10, 23, 38, 65, 99) and in lines transformed with the empty pCambia 1300 vector (2C) and pCambia 1303 vector (GUS). Plant line 89 was not silenced. Total RNA was isolated from plant roots, 20 d post-inoculation (dpi) with *M. loti*, and was reverse transcribed and normalized to the constitutively expressed gene of *UBQ* transcripts.

Collectively, these data suggest that lupeol may play a role in nodule formation. The effect of exogenously applied lupeol was next examined by assessing *ENOD40* and *ENOD2* expression in the roots of *M. loti*-inoculated wild type plants. As a control, a different triterpene, β-amyrin, was also tested. The triterpenes were added to the roots at different time points (10 min, 6 h or 12 h) after plant inoculation with *M. loti*. *ENOD40* and *ENOD2* gene expression levels were estimated by Q-PCR 48 h post-inoculation. A clear decrease in *ENOD40* expression was observed in plants treated with lupeol when compared with non-treated plants, but *ENOD2* expression was unaffected (Figure 35). This effect was evident very rapidly after the addition of lupeol, at 10 min post-inoculation, and was even more pronounced at 6 and 12 h post-inoculation. On the other hand, plants treated with β-amyrin exhibited an elevated level of *ENOD40* at 10 min post-inoculation when compared with untreated plants. This effect was only transient and was not observed when β-amyrin was applied to the plants at 6 or 12 h following plant inoculation with *M. loti* (Figure 35).

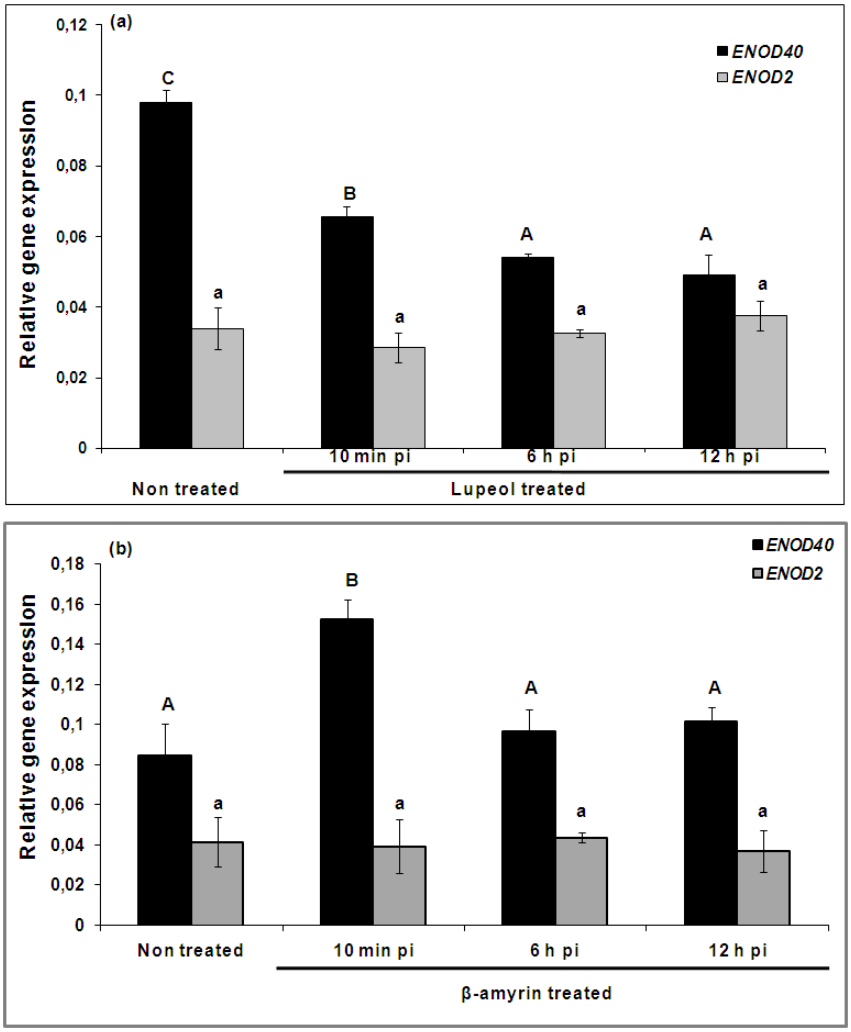


Figure 35. *ENOD40* and *ENOD2* expression in wild-type plants inoculated with *M. loti* and treated with lupeol (a) and β-amyrin (b), 10 min, 6 h and 12 h post-inoculation with the symbiont and in plants without the addition of lupeol / β-amyrin. The expression of *ENOD40* and *ENOD2* was measured 48 h after infection with rhizobia. Relative *ENOD40* and *ENOD2* expression was measured with respect to *UBQ* transcripts. In all samples, total RNA, derived from 40 plants per treatment, was reverse transcribed, normalized to the constitutively expressed gene of UBQ and subjected to real-time PCR. Bars indicate standard errors of the means (n = 3). Statistical comparisons within plant lines or treatments for *ENOD40* or *ENOD2* expression levels were performed by Duncan tests ($a < 0.05$). Indicator letters (capitals for *ENOD40*, lower case for *ENOD2*) in common denote a lack of a significant difference. The experiment was repeated with similar results.

Changes in the expression levels of *OSC3* were not observed in lupeol-treated plants (Figure 36).

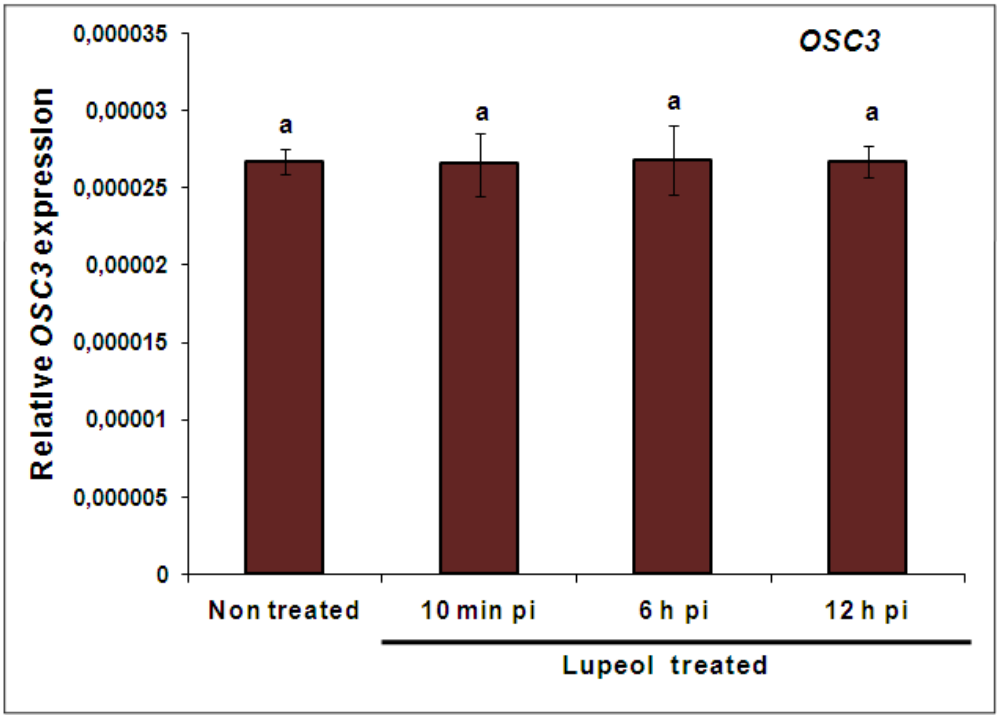


Figure 36. *OSC3* expression in wild type plants inoculated with *M. loti* and treated with lupeol 10 min, 6h and 12h post inoculation with the symbiont and in plants without the addition of lupeol/β-amyrin. The expression of *OSC3* was measured 48 h after the infection with rhizobia. Relative *OSC3* expression was measured with respect to *UBQ* transcripts. In all samples total RNA, derived from 40 plants per treatment, was reverse transcribed, normalized to constitutively expressed gene of *UBQ* and subjected to real-time PCR reaction. Bars indicate standard errors of means (n=3).

Taken together, our data suggest a role for lupeol biosynthesis in nodule formation through the regulation of *ENOD40* gene expression. Thus, silencing the LuS gene *OSC3* resulted in the absence of detectable levels of the metabolite, lupeol, and associated overexpression of *ENOD40* in the independent plant lines. An opposite effect could be mimicked by applying lupeol exogenously to *M. loti* inoculated plants, which subsequently exhibited reduced levels of *ENOD40* gene expression.

C.4. Appendix 2

C.4.1. GC/MS analysis of the metabolites of *AMY2* cluster genes

Following expression, *N. benthamiana* leaves were collected, subjected to metabolite extraction and analyzed by GC/MS. All candidate genes were introduced on their own and also in combination with each other in co-expression experiments. Previously, AMY2 was determined as a multi-functional oxidosqualene cyclase, because it produces both β-amyrin and lupeol when expressed in yeast (Iturbe-Ormaetxe *et al*., 2003). In *N. benthamiana* leaves, AMY2 produced β-amyrin (peak 2, Figure A1) as expected and the mass spectrum of peak 2 matched that of the trimethylsilyl (TMS) ether derivative of a β-amyrin standard (Figure A1). Surprisingly, lupeol was not detected. Instead, a less polar compound was detected in leaf extracts of *AMY2*-expressing plants (peak 1, Figure A1). Analysis of the mass spectrum of the trimethylsilyl (TMS) ether derivative of the less polar compound yielded a parent ion of *m/z* 500 (Figure A1.b), a number of fragment ion peaks characteristic for C-3 hydroxy lupanes (e.g. *m/z* 279, 207, 220), and two significant fragment ion peaks at *m/z* 457 and *m/z* 191, that suggested a saturated lupane structure. The fragment ion peak at *m/z* 457 is known to result from the loss of a propyl group, following a fragmentation favored only in saturated lupanes (Budzikiewicz *et al*., 1964); similarly, the fragment ion peak at *m/z* 191 corresponds to a fragment having rings D and E of a saturated lupane structure (Budzikiewicz *et al*., 1964; Dantanarayana *et al*., 1981). On the basis of these results, the less polar component (peak 1, Figure A1) detected in the leaf extracts of *AMY2*-expressing plants is proposed to be dihydrolupeol.

No new metabolites were observed when *LjCYP88D5* or *LjCYP71D353* were expressed alone in *N. benthamiana*. However, simultaneous expression of *AMY2* and *LjCYP71D353* did result in the production of two novel metabolites (peaks 3–5, Figure A1.a).

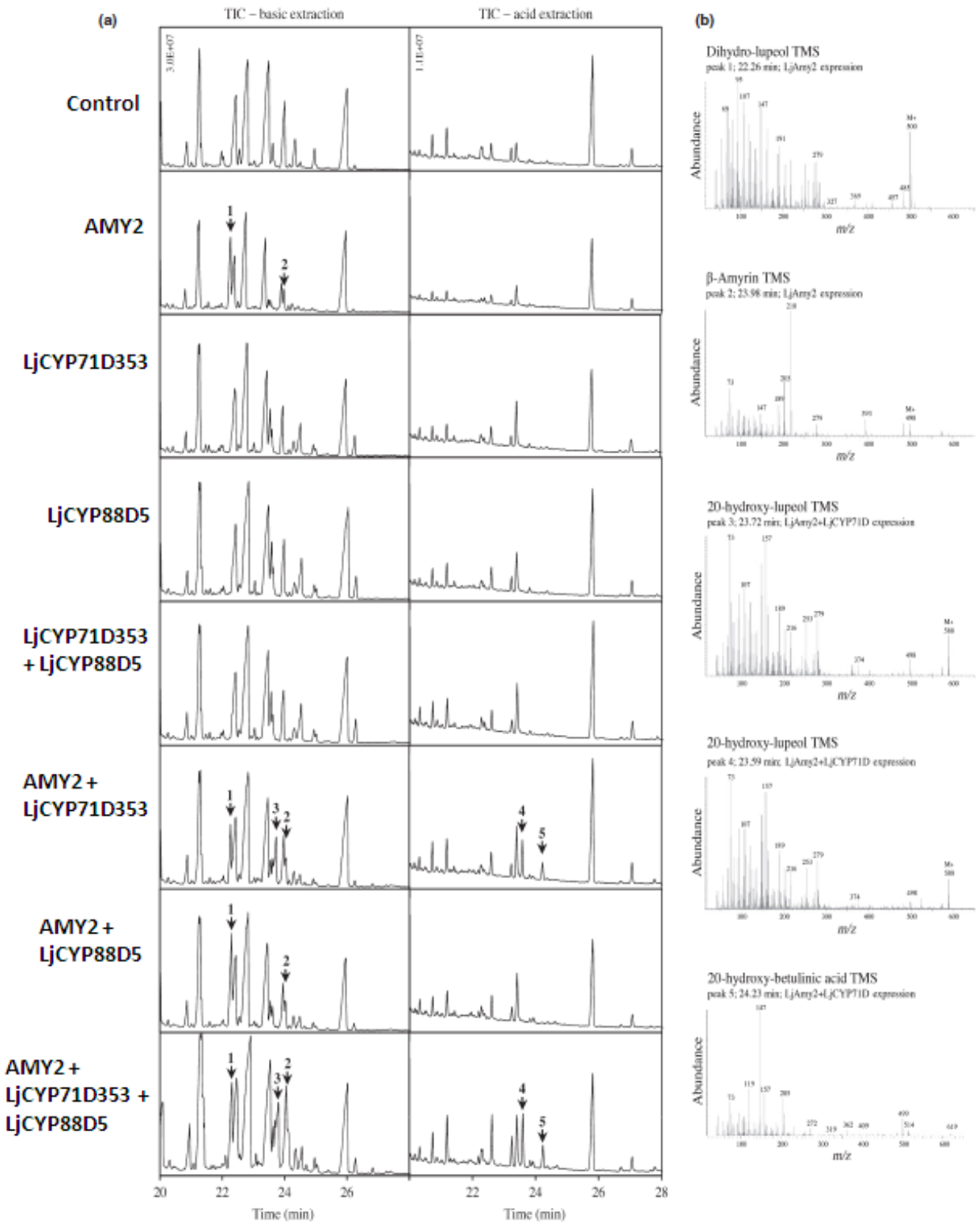


Figure A1. GC-MS analysis of saponified *N. benthamiana* leaf extracts after transient expression of *AMY2*, *LjCYP71D353* and/or *LjCYP88D5*. (a) Total ion chromatograms (TIC) of derivatised samples from basic and acidic extracts after saponification of plant material are shown. AMY2 protein expression results in accumulation of dihydrolupeol (peak 1) and β-amyrin (peak 2). Co-expression of *AMY2* and *LjCYP71D353* leads to accumulation of 20-hydroxylupeol (peak 3; RT = 23.7 min in the basic extractions and peak 4, RT = 23.59 in the acid extractions) and 20-betulinic acid (peak 5, RT = 24.22 min, acid extraction). No activity for LjCYP88D5 was detectable. Other major peaks are plant sterols. Each column of chromatograms has the same scale (ion count, indicated in the top left corner). (b) Mass spectra of peaks 1–5 from the GC profiles shown in (a).

Basic and acid procedures were used to ensure extraction of a wide range of possible metabolites. Peaks 3 and 4 correspond to the same product, the TMS-derivative of 3,20-lupandiol (commonly referred to as 20-hydroxy-lupeol); both peaks showed the same fragmentation pattern (Figure A2.b), with a prominent parent ion peak at *m/z* 588 and characteristic fragment ion peaks at *m/z* 573 (M^+–CH_3), 498 (M^+–TMS–H_2O), and 408 (M^+–2TMS–2H_2O; Cole *et al*., 1991; Ulubelen *et al*., 1994). Peak 5, which was only detected in the acidic fraction, was identified as 3,20-dihydroxy-28-lupanoic acid (commonly referred to as 20-hydroxy-betulinic acid) after a detailed analysis of the mass spectrum of its TMS derivative, which showed fragment ion peaks at *m/z* 619, 513, and 408; these fragment ion peaks can be explained by the loss of a TMS protecting group following a McLafferty-type rearrangement from a protonated parent ion peak, the combined loss of a second TMS protecting group, a molecule of water and a methyl group, and by a fragment having the fully substituted rings D and E, respectively (Figure 16; Budzikiewicz *et al*., 1964; Tsichritzis & Jakupovic, 1990). 20-hydroxylupeol and 20-hydroxybetulinic acid were detected only when *LjCYP71D353* was expressed together with *AMY2*.

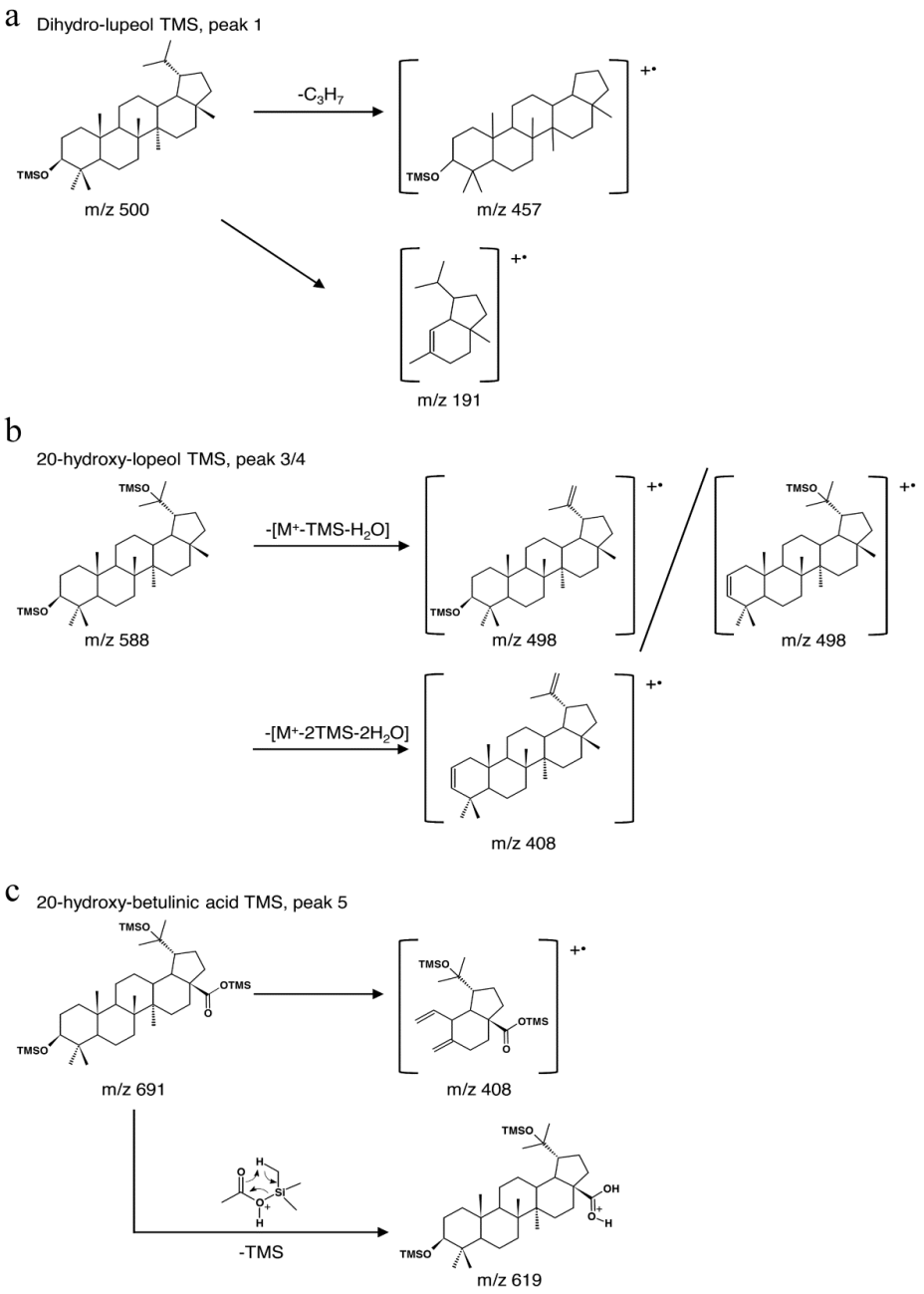


Figure A2. Main fragments in the mass spectrometry fragmentation patterns of the TMS-derivatives of dihydrolupeol (a), 20-hydroxy-lupeol (b) and 20-hydroxy-betulinic acid (c).

This indicates that LjCYP71D353 catalyses the formation of 20-hydroxylupeol from dihydrolupeol in a single oxidation reaction. In addition, LjCYP71D353 catalyses the formation of 20-hydroxybetulinic acid in a three-step sequential oxidation at the C-28 position of 20-hydroxylupeol (Figure A3).

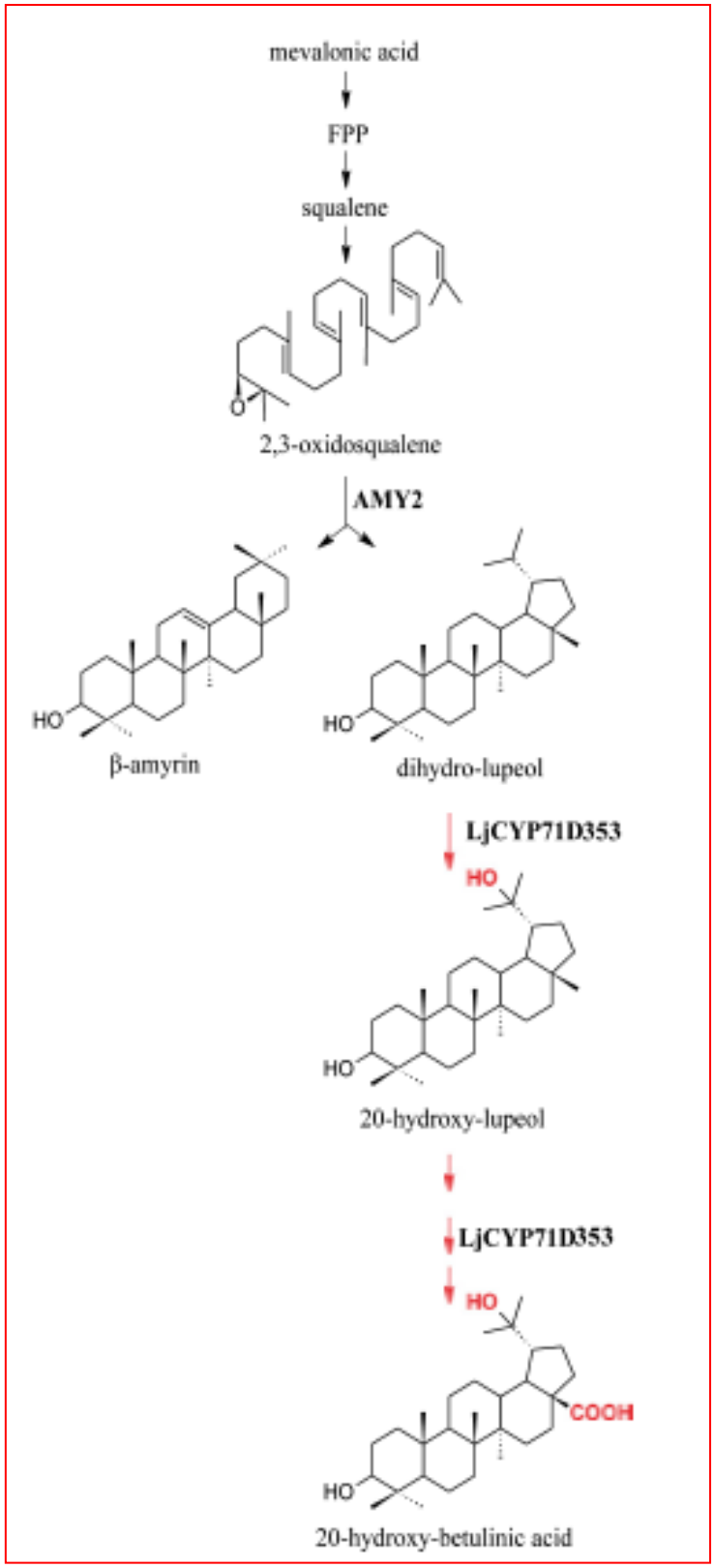


Figure A3. Proposed enzymatic reactions catalysed by AMY2 and LjCYP71D353. The structures of compounds and possible biosynthetic intermediates produced by the *L. japonicus AMY2* cluster are shown. AMY2 catalyzes the cyclization of 2,3 oxidosquealene to β-amyrin and dihydro-lupeol. LjCYP71D353 catalyzes the reaction to 20-hydroxy-lupeol and the formation of 20-hydroxybetulinic acid in a sequential three-step oxidation at C-28 of 20-hydroxylupeol.

To further verify that LjCYP71D353 acts on dihydrolupeol and not on the β-amyrin produced by AMY2, *LjCYP71D353* was co-expressed together with the oat β-amyrin synthase gene *AsbAS1*, which produces only β-amyrin. No new products were detected. Similarly, β-amyrin was not recognized as a substrate of LjCYP88D5 (Figure A4). No further products were detected when *AMY2*, *LjCYP71D353* and *LjCYP88D5* were co-expressed together.

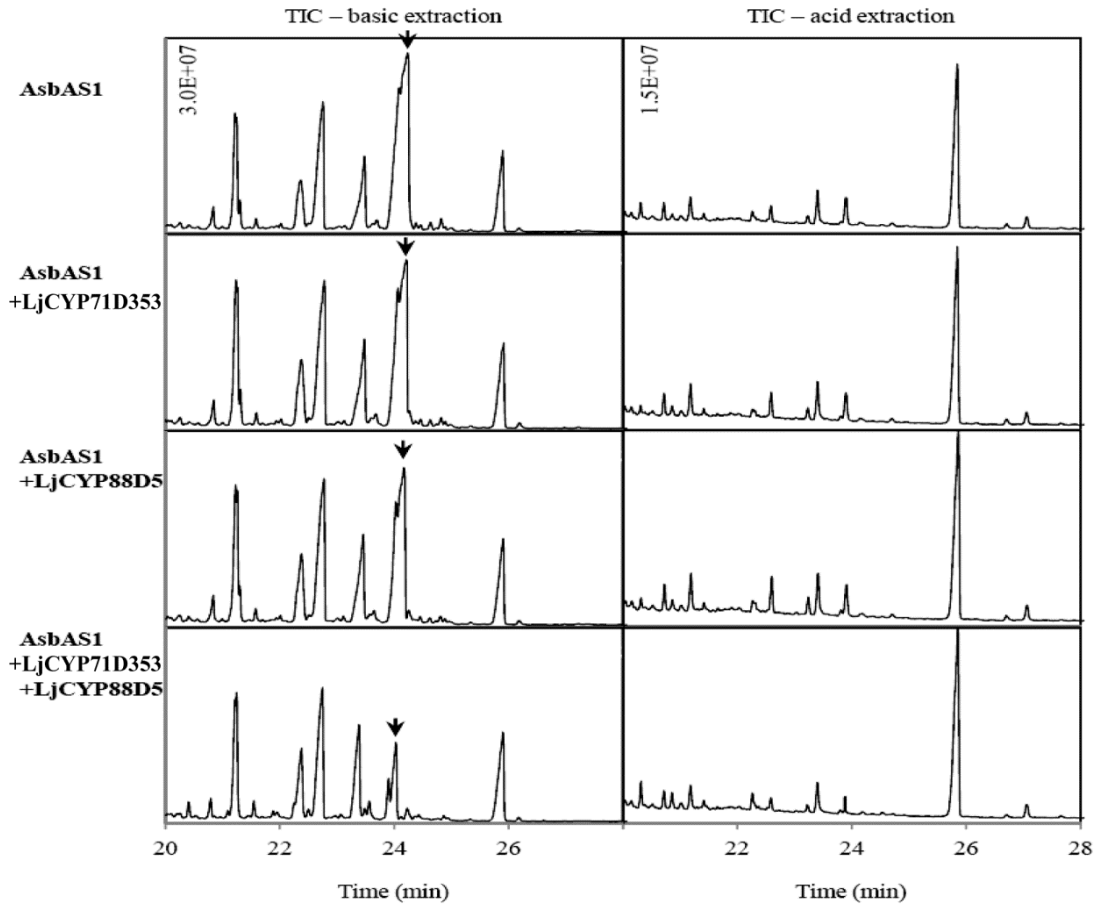


Figure A4. GC-MS analysis of saponified *N. benthamiana* leaf extracts after transient expression of *AsbAS1*, *LjCYP71D353* and/or *LjCYP88D5*. Arrows indicate the β-amyrin peak.

C.4.2. GC-MS analysis of *OSC3*-silenced plant lines

The presence or absence of lupeol in extracts from transgenic root tissues derived from control plants, as well as non-silenced plant lines and plant lines with no detectable level of *OSC3* expression, was established by GC-MS analysis (Figure A5). Although lupeol could be detected at low levels in the chromatographic profiles of the extracts from both control and non-silenced lines, none of the extracts from silenced lines showed the presence of lupeol by GC-MS. The presence or absence of lupeol in the various extracts was confirmed by searching for the parent ion peak (M^+) at *m/z* 426 in the MS fragmentation pattern of the components at the expected t_R (19.49 min) for lupeol (Figure A5). Furthermore, the possible presence of amyrin in the extracts was ruled out because none of the fragmentation patterns corresponding to the various components showed the characteristic fragment ion at *m/z* 218, resulting from the retro-Diels–Alder breakage of ring "C" of the amyrin skeleton. Thus, we are confident that, for those extracts that did not contain lupeol, the triterpene was undetectable by the method used, and it was confirmed that, in independent plant lines, transformation with the hairpin construct resulted in a lack of expression of *OSC3* and the absence of detectable levels of lupeol.

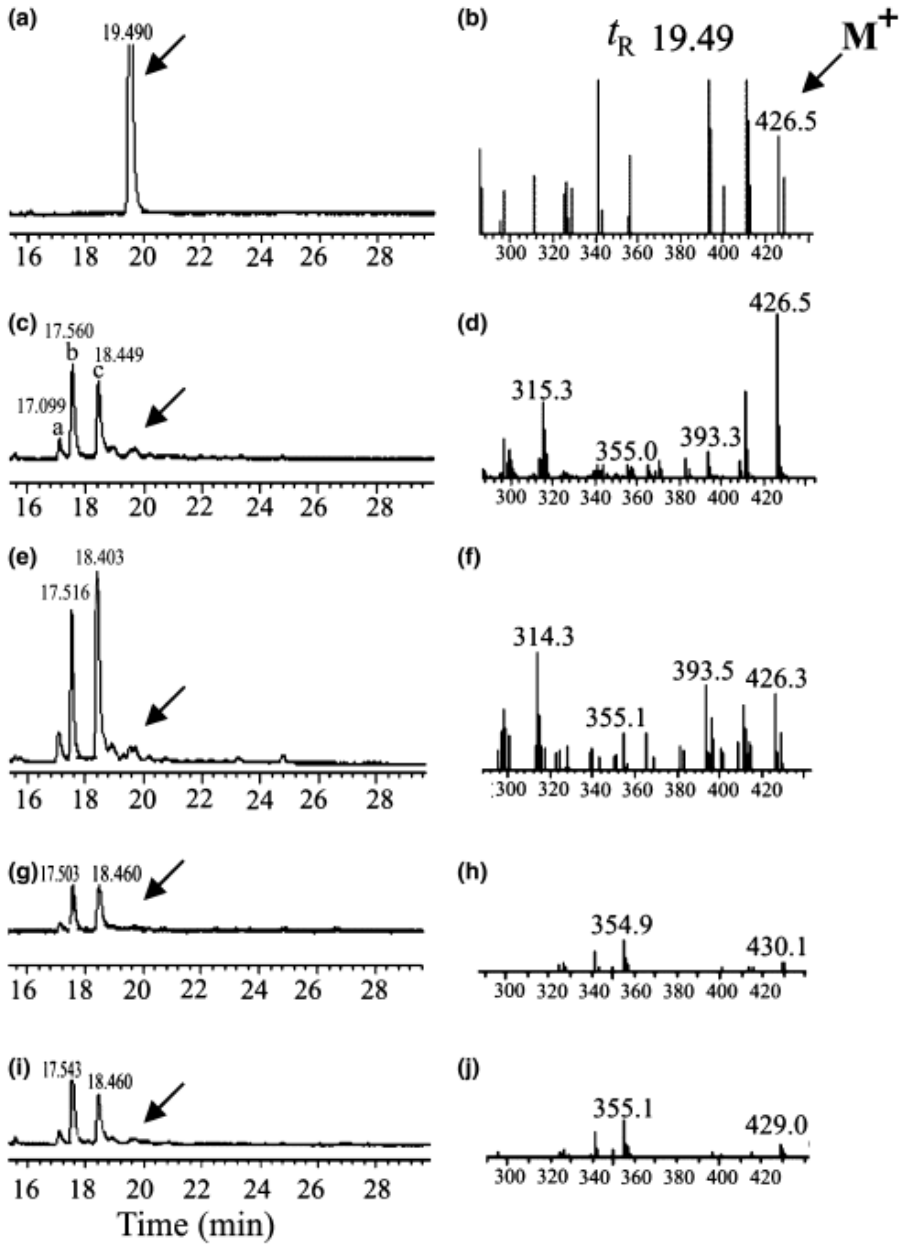


Figure A5. Gas chromatographic profiles (peak height) of lupeol (a), and extracts from control line CT2 (c), non-silenced line 89 (e) and lines silenced for lupeol synthase, 23 (g) and 65 (i), and MS fragmentation patterns of lupeol (b) and the component at t_R = 19.49 min in the chromatographic profile of the extracts of control line CT2 (d), non-silenced line 89 (f) and lines silenced for lupeol synthase, 23 (h) and 65 (j). The presence of the parent ion peak (M^+) at m/z 426 in the MS fragmentation patterns of the extracts of control line CT2 (d) and non-silenced line 89 (f) confirms the presence of lupeol. Peaks a, b and c correspond to campesterol, stigmasterol and sitosterol, respectively.

D. DISCUSSION

CHAPTER D. DISCUSSION

D.1. Triterpene gene clusters in legumes

D.1.1. Identification of gene clusters in *L. japonicus* and *M. truncatula*

Genes for triterpene biosynthetic pathways exist as metabolic gene clusters in oat and *A. thaliana* plants. This present work describes the first identification of the presence of analogous gene clusters in the legumes species *L. japonicus* and *M. truncatula*. The genome sequences of the model legumes *L. japonicus* and *M. truncatula* have been mined for triterpene gene clusters. Four candidate clusters have been identified. Three of them are potentially implicated in triterpenoid biosynthesis and the fourth is likely involved in sterol biosynthesis. The first gene cluster in *L. japonicus* consists of the previously described *AMY2* gene (Iturbe-Ormaexte *et al.*, 2003), two cytochrome P450 genes, *LjCYP71D353* and *LjCYP88D5*, and a putative oxidoreductase gene, *LjSDRt*. The second gene cluster in *L. japonicus* comprises of an *AMY2*-like gene interrupted at amino acid residue 228 and a cytochrome P450 gene, *LjCYP88D4*. In *M. truncatula*, one of the gene clusters contains the previously reported β-amyrin synthase gene, *βAS1* gene (Suzuki *et al.*, 2002) and a cytochrome P450 gene *MtCYP88D1* adjacent to *βAS1*, together with a short chain dehydrogenase / reductase gene (*MtSDRt*). This finding is in disagreement with Fukushima *et al.*, (2013), which have reported that in *M. truncatula* triterpenoid biosynthetic genes are not assembled into operon-like gene clusters. The second candidate region in *M. truncatula* consists of a putative cycloartenol synthase gene, *MtCAS*. This gene is also flanked by an oxidoreductase gene, a cytochrome P450 gene, *MtCYP97B*, and a GDSL esterase / lipase. A common trait of these four gene clusters is the presence of several scattered genes of unknown function, which could be

related to the corresponding gene cluster and participate in the metabolic pathways. The *AMY2* gene cluster in *L. japonicus* was selected and functionally validated by the characterization and analysis of the genes comprising the cluster.

D.1.2. The evolutionary origin of gene clusters

Metabolic gene clusters in other plants have likely arisen *de novo* within recent evolutionary history from plant genes by gene duplication, acquisition of new function and genome re-organization (Gierl and Frey., 2001; Field and Osbourn, 2008; Frey *et al*., 2009; Takos *et al*., 2011; Winzer *et al*., 2012) and not from microbes by horizontal gene transfer (Field and Osbourn, 2010; Osbourn, 2010 a, b; Chu *et al*., 2011). Gene duplications are a central mechanism for evolution of genes and genomes and are regarded as the major driving force in generation of evolutionary novelty (Ober, 2005; 2010; Roth *et al*., 2007). Progress in identification and characterization of genes involved in secondary metabolic pathways has provided several examples of genes that originated by gene duplication, followed by diversification of two copies. Many of these genes encode proteins that belong to families such as terpene synthases (Tholl, 2006), cytochrome P450-dependent monooxygenases (Bak *et al*., 2006), chalcone synthases (Durbin *et al*., 2000) or glucosidases (Kliebenstein *et al*., 2005). It is noteworthy that for several secondary metabolic gene clusters identified in plants the enzymes for the first committed step appear to have been recruited directly or indirectly from primary metabolism. The first committed step in the synthesis of the cyclic hydroxamic acids DIBOA and DIMBOA is the conversion of indole-3-glycerol phosphate to indole, which is catalysed by the tryptophan synthase a (TSA) homologue BX1. Bx1 is likely to have originated by duplication of the maize gene encoding TSA (Frey *et al*., 1997). The genes for the signature enzymes of the two diterpene clusters in rice (for phytocassane and momilactone synthesis) are likely to have

originated by recruitment from the gibberellin pathway (Sakamoto *et al*., 2004; Wilderman *et al*., 2004; Shimura *et al*., 2007; Swaminathan *et al*., 2009). Several genetic loci comprise the gene cluster implicated in avenacin production in oat (Papadopoulou *et al*., 1999; Haralampidis *et al*., 2001; Qi *et al*., 2004; 2006; Mugford *et al*., 2009; 2013). *Sad1* is likely to have been recruited from sterol metabolism by duplication and divergence of a plant cycloartenol synthase and this is a relatively recent evolutionary event (Haralampidis *et al*., 2001; Qi *et al*., 2004). *Sad2* encodes a cytochrome P450 enzyme belonging to the CYP51 sterol demethylase family, which is regarded as the most ancient cytochrome P450 family (Qi *et al*., 2006). Donor sequences in *Papaver somniferum* gene cluster could be recruited from genes encoding related plant enzymes, such as cytochrome P450s and *O*-methyltransferases, in a process involving gene duplication and neo-functionalization (Winzer *et al*., 2012). The arrangement of the *HN1* cluster suggests that genome reorganization is an ongoing process, occurring in some cases before duplication, as evidenced by the small gene families (*PSCXE, PSCYP82*, and *PSMT*), or after duplication, as evidenced by the single-copy genes (*PSSDR1, PSAT1*, and *CYP719A21*) (Winzer *et al*., 2012). Two functional gene clusters have been identified in the genome of *A. thaliana*, the thalianol and the marneral cluster, specialized in tricyclic triterpenes biosynthesis that so far have been only identified in *Arabidopsis* (Field *et al*., 2008; Field *et al*., 2011). These *Arabidopsis* clusters are not simply the result of the whole-scale duplication of an ancestral cluster. Instead, the two clusters were most probably founded either independently, or else by the duplication of an ancestral *OSC / CYP705* gene pair followed by independent rearrangements and the recruitment of different genes (Figure 14, Chapter A) (Field *et al*., 2011; Field and Osbourn, 2012). Recent studies revealed that the dicot triterpene synthases, including the β-amyrin synthases *AMY2* and *βAS* in *L. japonicus* and *M. truncatula* respectively, may have originated from the ancient

lanosterol synthase gene (*ALSL*), instead of cycloartenol synthases, via three successive gene duplication events, followed by positive selection and diversifying evolution (Xue *et al*., 2012).

In this present study, the gene clusters identified in legumes further support the *de novo* theory. The homology of genes in legume clusters and the synteny of genomic regions were studied. Due to the corrupted *AMY2* gene and the absence of an *LjCYP71D353* homologue in the *AMY2*-like gene cluster of *L. japonicus*, it is proposed that this second cluster is either incomplete or in decay. Furthermore, the low degree of conservation in genome structure and synteny between the *AMY2* cluster and *AMY2*-like cluster regions suggests that the ancestral *AMY2* gene was initially adjacent to a cytochrome P450 gene belonging to the CYP88D subfamily and that this region then underwent a tandem duplication followed by genome reorganization. This hypothesis is further supported by the presence of several transposable elements in these genomic regions. The *AMY2* and *bAS1*, as well as the *LjCYP88D5* and *MtCYP88D1* genes are conserved genes in the *L. japonicus* and *M. truncatula* respectively, which indicates that they have been maintained by evolution despite speciation. Despite this conservation, the genomic regions of *AMY2* and *βAS* cluster display a low degree of synteny. However, the high level of sequence homology between the orthologues of β-amyrin synthase genes in *L. japonicus* and *M. truncatula*, *AMY2* and *bAS1*, and between *LjCYP88D5* and *MtCYP88D1*, in the *AMY2* and *βAS1* gene clusters respectively, suggests that these gene clusters originate from a single cluster in a common ancestor of *M. truncatula* and *L. japonicus* that subsequently evolved independently in these two species.

The *de novo* cluster assembly hypothesis is supported by the presence of the *LjCYP71D353* gene in the *AMY2* gene cluster. There are no *LjCYP71D353* orthologues in *M. truncatula* and the relatively large phylogenetic distance between *LjCYP88D5* and *LjCYP71D353* rules out the possibility that the two

genes originate from recent gene duplication. *CYP88D* genes are present in *L. japonicus* clusters and in *M. truncatula βAS1* cluster and belong to a legume-specific cytochrome P450 subfamily. Thus, *OSC* and *CYP88* may have been in close proximity in the ancestral gene cluster and the other genes were recruited independently in the different lineages. The genes encoding the tailoring enzymes have varied distributions among different gene clusters in *M. truncatula* and *L. japonicus*, suggesting different genome reorganization events within these genetic regions. All proposed duplication and gene recruitment events occurred in these two legume model species are depicted in Figure 37.

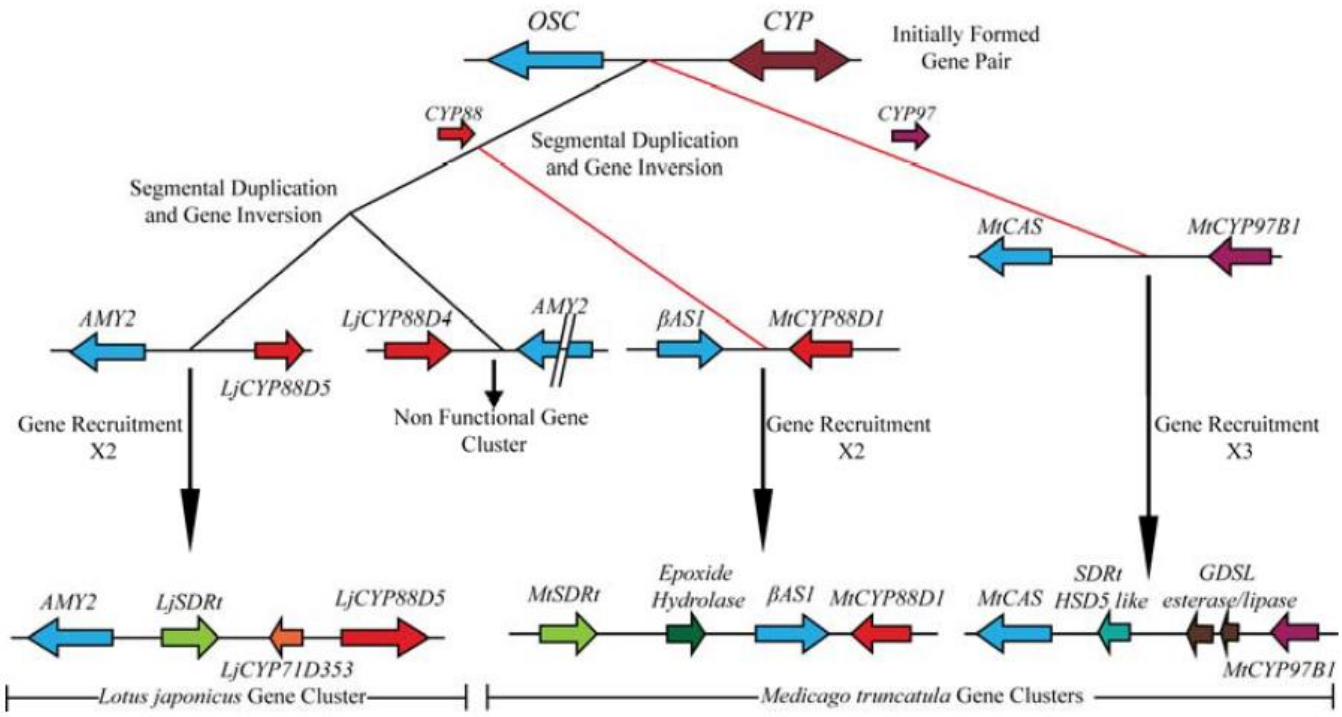


Figure 37. Formation of triterpene gene clusters in *L. japonicus* and *M. truncatula*. Proposed scheme based on gene sequence similarities and genome organization, based on the assumption that the clusters were founded by duplications and rearrangements of an ancestral *OSC/ CYP* gene pair.

LjCYP71D353 is phylogenetically related to *CYP71A16* that forms part of the marneral gene cluster in *A. thaliana* (Field *et al.*, 2011). The syntenic

analysis conducted revealed low degree of synteny between the *L. japonicus AMY2* cluster and the *A. thaliana* marneral cluster. The low syntenic conservation and the phylogenetic distance between the *OSC* genes in these two clusters (*MRN1*, in the marneral cluster, and *AMY2* in the *L. japonicus* cluster) suggests that *LjCYP71D353* and *AMY2* are not derived from the tandem duplication of a common ancestral pair of *OSC* and cytochrome P450 genes and do not share a common origin with the marneral cluster. Hence, the *LjCYP71D353* gene was most probably recruited into the *AMY2* - *LjCYP88D5* gene cluster independently.

D.1.3. Legume clusters are assembled within dynamic chromosomal regions

The second common feature of all four candidate gene clusters is the presence of several transposable elements (TEs) in the proximity of *OSC* and cytochrome P450 genes (Figure 1, Chapter C), which may have some function in gene rearrangement for cluster formation.

Genome sequencing efforts have revealed how extensively TEs have contributed to the shaping of present day genomes. The class I of retrotransposons constitutes the bulk of transposonome in most plants and they tend to concentrate in centromeres and intergenic regions (SanMiguel and Bennetzen, 1998) and, thus, they contribute principally to expansions and contractions in genome size (Bennetzen *et al*., 2005) and to intergenic sequence diversity between and within species (Morgante *et al*., 2007). On the other hand, the less abundant DNA transposons (class II) tend to associate with the euchromatic or genic components of the genome and, thus, they have closer interactions with the genes of the plant host. These interactions have resulted in TEs acquiring host sequences, forming chimeric genes through exon shuffling, replacing regulatory sequences, mobilizing genes

around the genome, and contributing genes to the host (Dooner and Weil, 2007).

The class I of retrotransposons is recognized in all four legume clusters. In *AMY2* and *AMY2*-like genomic regions few transposons of the class II are also identified. In addition, repetitive DNA sequences are found in genomic segments between the four genes in *L. japonicus AMY2* cluster, which probably correspond to the promoter regions. TEs and repetitive sequences are often associated with secondary metabolic clusters in filamentous fungi (Perrin *et al*., 2007; Shaaban *et al*., 2010), and may have functions in cluster formation and regulation in plants (Chu *et al*., 2011). The presence of transposon sequences have been observed in the clusters for cyanogenic glucosides in *L. japonicus*, sorghum and cassava (Takos *et al*., 2011). Interspersed among the ten genes in *P. somniferum* noscapine cluster are both retrotransposon and DNA TEs, which may have some function in gene rearrangement for cluster formation (Winzer *et al*., 2012). The thalianol and marneral clusters are located in an island between segments of the *Arabidopsis* genome that were duplicated in the most recent whole genome duplication event, which occurred about 47 – 69 mya. This location is unusual because only 11% of annotated genes exist outside of α-duplicated segments. Non-α regions are generally poor in genes and rich in pseudogenes and TEs (Bowers *et al*., 2003; Freeling *et al*., 2008). TE density in cluster regions was found higher than in flanking chromosomal regions or in other comparable regions between α-duplicated chromosome segments (Field and Osbourn, 2012). Thus, it can be proposed that these TEs directly contribute to the assembly of gene clusters: TEs promote ectopic recombination, and certain classes such as helitrons, which are present in both clusters, can trans-duplicate genes (Dooner and Weil, 2007; Fiston-Lavier *et al*., 2007). The TE density at *Arabidopsis* clusters suggests the dynamic nature of these regions because they accept and retain segmental duplications (SDs) more frequently

than neighboring regions (Field and Osbourn, 2012). SD acceptors have been identified in other eukaryotic genomes. In humans, they are thought to play key roles in the generation of novel gene functions (Marquez-Bonet *et al*., 2009)

In addition, the *A. thaliana* marneral and thalianol clusters are located very near to the chromosome breakpoints (Field and Osbourn, 2012). Telomeres are the nucleoprotein complexes found at chromosome ends, which are essential for chromosome stability. They prevent the ends from fusing together or being recognized and treated as damaged DNA. Most eukaryotic telomeres consist of short, tandemly repeated nucleotide sequences, with one strand (5'-3' in the direction of the chromosome end) being guanine rich (G-rich strand) (Zakian, 1995). Subtelomeric sequences are proximal to the telomeres and - in contrast to the simplicity and short lengths of telomeric repeats - their organization is complex and variable among eukaryotes (Pryde *et al*., 1997). Repetitive sequences (Alkhimova *et al*., 2004; Ohmido *et al*., 2001; Sykorova *et al*., 2003) and/or duplicated segments derived from non-homologous chromosomes (Linardopoulou *et al*., 2005; Louis, 1995) are often found within subtelomeres. Additionally, point mutations, insertions, deletions, inversions, microsatellite expansions, and/or contractions can be present (Kuo *et al*., 2006; Wu *et al*., 2009). It is generally approved that the DNA recombinations, breaks, and repair events that occur within subtelomeres can result in sequence variation at chromosome end (Carlson *et al*., 1985; Kuo *et al*., 2006). The actinomycete and fungal secondary metabolic gene clusters are often found close to the end of chromosomes (Hoffmeister and Keller, 2007). The gene cluster for cyclic hydroxamic acid production and the sorghum cluster for cyanogenic glucosides biosynthesis are localized in subtelomeres. DNA FISH suggests that the avenacin cluster may also be subtelomeric (Qi *et al*., 2004; Wegel *et al*., 2009; Field and Osbourn, 2012). The subtelomeric regions of eukaryotic chromosomes are highly dynamic, facilitating

chromosomal recombination, SDs, DNA inversion, partial deletions, translocations and other rearrangements. The enhanced turnover of SDs within such regions is likely to provide a special environment that will enable the sampling of different gene recombinations, and thus, the assembly and functional optimization of operon-like gene clusters in response to selection pressure (Field and Osbourn, 2012). On the other hand, the cluster for the biosynthesis of cyanogenic glucosides in *L. japonicus* and the two diterpene clusters in rice for phytocassane and momilactone production are not located in subtelomeric regions. The location of the cyanogenic glucosides cluster in cassava, the noscapine cluster in *P. somniferum* and the steroidal alkaloid clusters in Solanaceous crops is not currently known. The *AMY2* and *AMY2*-like gene clusters are localized in non-telomeric chromosomal DNA whereas the *βAS* and *MtCAS* gene clusters are located in chromosomal regions just proximal to the end of chromosome. Thus, although further investigation is needed as regards a putative mechanism of cluster assembly due to the location on the chromosome, the current data indicate towards a general mechanism of gene cluster assembly in all eukaryotes and the identified gene clusters in legumes could strengthen the proposed model of gene cluster assembly within dynamic chromosomal regions.

D.2. Verification of the AMY2 gene cluster

D.2.1. Coordinate regulation of cluster genes

Consistent with previous findings for the co-ordinated regulation of cluster activity (Osbourn and Field, 2009), the *AMY2* cluster genes are coordinately expressed in response to developmental and environmental stimuli. The effects of different plant hormones, various abiotic stresses and rhizobium infection on *L. japonicus AMY2* cluster genes expression were investigated.

D.2.1.1. Developmental regulation

In *P. somniferum* cluster for noscapine biosynthesis ten specific genes are co-expressed in stems and capsule tissues (Winzer *et al*., 2012). Gene co-expression analysis using transcriptome data revealed six genes co-expressed with the glycosyltransferase gene glycoalcaloid metabolism 1 (*GAME1*) from tomato fruit and five genes co-expressed with the glycosyltransferase gene homolog *SGT1* from potato tubers (Itkin *et al*., 2013). Highest transcript levels of *Bx* (*Bx1 – Bx7*) genes of benzoxizanone cluster in maize are present in the roots and shoots of young seedlings (Von Rad *et al*., 2001; Frey *et al*., 2003) and in the scutellar node (Jonczyk *et al*., 2008). The expression of all four genes comprising the thalianol cluster in *A. thaliana* is highly correlated and occurs primarily in the root epidermis, which suggests that the genes are functionally related (Field and Osbourn, 2008). A root-specific expression pattern, similar to that of the thalianol cluster, describes the marneral cluster in *A. thaliana* (Field *et al*., 2011). The avenacin biosynthetic genes are tightly regulated and expressed only in the root epidermis, the site of accumulation of the pathway end products (Haralampidis *et al*., 2001; Qi *et al*., 2004; 2006).

The expression patterns of the *AMY2*, *LjCYP88D5*, *LjCYP71D353* and *LjSDRt* genes were examined in leaves and roots, as well as in nodules of *L. japonicus* plants at different developmental stages. *AMY2*, *LjCYP88D5* and *LjCYP71D353* genes show similar expression patterns in *L. japonicus* roots and leaves at various time points, as well as in roots, leaves and nodules following infection with *M. loti* strain R7A. Gene expression of the *AMY2*, *LjCYP88D5* and *LjCYP71D353* genes is higher in the roots of 7-day-old uninfected *L. japonicus* seedlings and in inoculated with *M. loti* roots 7 dpi and 14 dpi, as well as in nodules at 14 dpi. *LjSDRt* is not coordinately expressed during plant growth and infection with *M. loti* strain R7A. The finding that *LjSDRt* follows a totally different expression pattern does not imply that this gene is not a member of *AMY2* cluster. It has been proposed that the genes in known clusters are not

always strictly co-expressed and may even be differentially regulated (VonRad *et al*., 2001; Swaminathan *et al*., 2009; Takos *et al*., 2011). Two gene members of cyanogenic glucoside biosynthetic cluster in *L. japonicus* show a similar expression profile related to leaf age, with highest expression observed in the apical leaves and transcript levels decreasing in older leaves. The expression of the other two genes remains unaltered with leaf age. Although the genes are not strictly co-regulated, the overlap in the expression patterns supports the view that these four genes encode the enzymes of the pathway for cyanogenic glucoside biosynthesis in *L. japonicus* (Takos *et al*., 2011).

AMY2, *LjCYP88D5* and *LjCYP71D353* are also developmentally co-regulated and their expression is coordinately induced by three different types of hormones, 2,4-dichlorophenoxyacetic acid (2,4-D), benzylaminopurine (BA) and methyl jasmonic acid (MeJA). Jasmonic acid and its volatile derivative MeJA, collectively called jasmonates, are plant stress hormones that act as regulators of defense responses. MeJA is a well-known elicitor for the production of plant natural products and its activity induces a wide structural variety, encompassing multiple branches of the most important secondary metabolite classes (i.e. phenylpropanoids, alkaloids and terpenoids) (Zhao *et al*., 2005). It is expected that, through the hormonal "crosstalk" networks that operate in plants, other hormones are likely to participate in the amplitude and specificity of such elicitations (Pauwels *et al*., 2009). Crosstalk between JAs and other phytormones seems to be complex and has been established with salicylic acid, ethylene, abscisic acid and auxin (Kazan and Manners, 2008) and, very recently, gibberellin (Nemhauser *et al*., 2006; Navarro *et al*., 2008). Accumulation of benzoxazinoids is induced in response to *cis*-jasmonate treatment in wheat (*Triticum aestivum*) (Moraes *et al*., 2008). Jasmonates also induce the production of diterpene (taxol) of *Taxus* spp., sesquiterpenes of *Lactuca sativa*, *Phaseolus lunatus* and *Zea mays*, and triterpenes of *Scutellaria baicalensis* (Memelink *et al*., 2001). MeJA has also been used to enhance the

ginsenoside content in cultured cells (Lu *et al*., 2001) and roots of *Panax ginseng* (Yu *et al*., 2002). Specifically regarding the induction of triterpene biosynthesis in legumes, MeJA induces the expression of the *bAS* gene and up-regulates the soyasaponin production in cultured cells of *M. truncatula* (Suzuki *et al*., 2002). It has also been reported that MeJA treatment induces soyasaponin biosynthesis in cultured cells of *Glycyrrhiza glabra* (Hayashi *et al*., 2003). It is suggested that jasmonates are a signal molecule to induce soyasaponin biosynthesis in legumes, while gibberellins may be negative regulators of the oleanane-type triterpene saponin biosynthesis (Hayashi *et al*., 2004).

Auxins have repeatedly been shown to inhibit alkaloid accumulation (Zenk *et al*., 1977; Morris, 1986; Merillon *et al*., 1986; Arvy *et al*., 1994). Auxins negatively influence the terpenoid indole alkaloid pathway primarily at transcriptional level of gene expression. Addition of naphthaleneacetic acid (NAA) to cell cultures of *Cantharanthus roseus* (*C. roseus*) led to lower levels of alkaloid accumulation. The presence of 2,4-D in the growth medium of *C. roseus* reduced culture aggregation and repressed accumulation of terpenoid indole alkaloids (Whitmer *et al*., 1998). On the contrary, the exogenous application of cytokinins in *C. roseus* cells in a 2,4-D-free medium greatly enhanced the accumulation of the indole alkaloid ajmalicine (Yahia *et al*., 1998). The influence of auxins on isoprene derivatives was investigated in callus cultures of *Vaccinium corymbosum* var bluecorp (Migas *et al*., 2006). Every auxin utilized, such as 2,4-D, NAA, indole-3-acetic acid (IAA) and naphthoxyacetic acid (NOA), except for indole-3-butyric acid (IBA), stimulated biosynthesis of steroid rather than triterpenoid compounds. A modified callus growing medium, supplemented with growth regulators of the auxin group, resulted in the production of a triterpene-rich callus culture and β-sitosterol accumulation. β-sitosterol accumulated in amounts which were approximately 10-times higher than for triterpene acids (oleanolic and ursolic acid) and

approximately 6-times higher than for amyrines (α- and β-amyrin) (Migas *et al*., 2006). Auxin and cytokinin have not been elsewhere reported to have functions in inducing triterpene production.

D.2.1.2. Coordinate response to environmental cues

In plant biology the term "elicitors" refer to chemicals from various sources that are able to trigger physiological and morphological responses. They include abiotic elicitors such as metal ions and inorganic compounds, and biotic elicitors from fungi, bacteria, viruses or herbivores, plant cell wall components, as well as chemicals that are released at the attack site by plants upon pathogen or herbivore attack. It is well-established that treatment of plants with elicitors provokes an array of defense reactions, including the accumulation of plant defensive secondary metabolites such as phytoalexins in intact plants or cell cultures.

The expression patterns of *AMY2*, *LjCYP71D353* and *LjCYP88D5* uncovered coordinate down regulation in response to heat, cold and oxidative stress in comparison with control plants, while no significant variations were noticed in the expression profile of *AMY2* cluster genes in response to salt stress. The genes comprising other plant metabolic gene clusters have also been reported to be induced by biotic or abiotic stimulations.

The transcription of some (but not all) genes of the momilactone diterpene gene cluster in rice is up-regulated in suspension-cultured cells treated with a chitin oligosaccharide elicitor or in response to challenge with pathogens (Okada *et al*., 2007; Shimura *et al*., 2007; Swaminathan *et al*., 2009). The inducible nature of phytocassane biosynthesis in response to pathogen attack has been additionally studied (Koga *et al*., 2005). The benzoxazinoids occur constitutively as glucosides. The expression of the genes responsible for their hydrolysis is induced in response to infection or physical damage to

produce DIMBOA and DIBOA, which are antimicrobial and also have pesticidal and allelopathic activity (Niemeyer, 1988; Wu *et al*., 2001).

In numerous reports, the effect of abiotic stresses on the primary metabolic processes has been well documented (i.e., photosynthesis, growth and metabolism of antioxidants) (Khan *et al*., 2009; Chen *et al*., 2012). However, environmental factors also influence production of secondary metabolites. Generally, when plants are stressed, secondary metabolism may increase, because growth is often limited more than photosynthesis and carbon fixation is predominantly invested to secondary metabolites production (Endara and Coley, 2011). A wide range of experiments have shown that plants exposed to drought stress did indeed accumulate higher concentrations of secondary metabolites. Such enhancement is reported to occur in nearly all classes of natural products, such as simple or complex phenols, numerous terpenes, as well as in nitrogen-containing substances, such as alkaloids, cyanogenic glucosides or glucosinolates (Selmar and Kleinwachter, 2013). Especially regarding the class of terpenoids, drought conditions result in strong increase of betulinic acid production in *Hypericum brasilience* (de Arbeu and Mazzafera, 2005), saikosaponin production in *Bupleurum chinense* (Zhu *et al*., 2009) and triterpene accumulation in *Prunella vulgaris* (Chen *et al*., 2011). Salinity can be considered as the major abiotic stress affecting plant physiology and, thus, plant development (Zhu, 2001; Munns and Tester, 2008). One of the effects of salinity is the alteration of the secondary metabolism involving signal molecules, oxidative stress and intermediary reactions. Glucosinolates have been shown to accumulate in plants when salinity stress is present above the tolerance levels (Yuan *et al*., 2010; Pang *et al*., 2012). The effect of salinity on biosynthesis and metabolism on individual glucosinolate merits further investigations (Del Carmen *et al*., 2013).

Environmental stresses, such as UV-irradiation, high light, wounding, nutrient deficiencies, temperature and herbicide treatment often increase the accumulation of phenylpropanoids (Dixon and Paiva, 1995). The expression of some gene partners of the momilactone cluster in rice is coordinately induced in leaves irradiated with UV light (Otomo *et al*., 2004; Sakamoto *et al*., 2004; Shimura *et al*., 2007). Nutrient stress also has a marked effect on phenolic levels in plant tissues (Chalker-Scott and Fnchigami, 1989). Deficiencies in nitrogen and phosphate directly influence the accumulation of phenylpropanoids (Dixon and Paiva, 1995). Potassium, sulfur and magnesium deficiency are also reported to increase phenolic concentrations. Low iron level can cause increased release of phenolic acids from roots (Chalker-Scott and Fnchigami, 1989). Formation of phenyl amides and dramatic accumulation of polyamines in bean and tobacco under the influence of abiotic stresses were reported, suggesting antioxidant role of these secondary metabolites (Edreva *et al*., 2000). Anthocyanin accumulation is stimulated by various environmental stresses, such as UV, blue light, high intensity light, wounding, drought, sugar and nutrient deficiency (Winkel-Shirley, 2001).

Low temperature is one of the most harmful abiotic stresses affecting plants species. During over wintering, temperate plant metabolism is redirected toward synthesis of cryoprotectant molecules such as sugar alcohols (sorbitol, ribitol, inositol) soluble sugars (saccharose, raffinose, stachyose, trehalose), and low-molecular weight nitrogenous compounds (proline, glycine betaine) (Janska *et al*., 2010). Cold stress increases phenolic production and their subsequent incorporation into the cell wall either as suberin or lignin (Griffith and Yaish, 2004). The accumulation of anthocyanins (Griffith and Yaish, 2004) and polyamime (Kovacz *et al*., 2011) during cold stress has been reported. Carotenoids in Brassicaceae, including β-carotene, were found to be slightly decreased after thermal treatments (Morison and Lawlor, 1999). Elevated temperatures increase leaf senescence, reduces

photosynthesis and biomass production of *Panax quinquefolius*, but enhances ginsenoside content (Gera *et al*., 2007). Generally, elevated or reduced temperatures have been shown to affect the synthesis and accumulation of saponins in other plants. Lower soil temperatures caused an increase in levels of steroidal furostanol and spirostanol saponins in spears of white asparagus *Asparagus officinalis*, a herbaceous perennial plant highly valued as a seasonal vegetable (Schwarzbach *et al*., 2006). Because biosynthesis of saponins is a plant response to environmental factors and part of an adaptative strategy leading to tolerance of abiotic stresses, fluctuations in the content of saponins resulting from temporal changes in environmental circumstances might be expected (Szakiel *et al*., 2011).

D.2.1.3. Coordinate expression in response to nodulation

A number of expression and biochemical studies have suggested a possible involvement of β-amyrin and derived saponins in the regulation of root nodulation. The ectopic expression of the *Aster sedifolius* β-amyrin synthase gene, *AsOXA1,* in *M. truncatula* led to enhanced root nodulation (Confalonieri *et al*., 2009). The high nodulation phenotype in *M. truncatula* transgenic lines was associated with the significance increase in the triterpenic content in roots and nodules (Confalonieri *et al*., 2009). In *G. glabra*, the expression pattern of β-amyrin synthase was examined in different tissues and high levels were detected in root nodules (Hayashi *et al*., 2004). The expression of the *Pisum sativum* β-amyrin synthase gene *PSY* was investigated during nodule development and senescence (Iturbe-Ormaetxe *et al*., 2003). PSY expression was highest in immature two-week nodules, moderate high in three-week old nodules, that were actively fixing nitrogen, and significantly reduced in senesced (five- and seven- week old nodules) nodules. Furthermore, the chemical induction of root nodule senescence was accompanied with reduced *PSY* transcript levels (Iturbe-Ormaetxe *et al*., 2003).

In this present work, *AMY2*, *LjCYP71D353* and *LjCYP88D5* were coordinately expressed and induced in the symbiotic relationship of *L. japonicus* seedlings with *M. loti*. Transcript levels were highest in infected roots 7 dpi (which include young nodules) and in nodules 14 dpi and then decreased in mature nodules at 28 dpi. The apparent increase in *AMY2* cluster genes expression in the roots of *M. loti*-infected plants and the high levels of cluster expression in nodules led to examine the spatial localization of cluster gene transcripts, in order to investigate the possible physiological role of cluster biosynthesis in symbiosis. An *in situ* mRNA hybridization experiment was carried out and revealed localization of *AMY2* cluster gene transcripts in the same tissues of *L. japonicus* developing (14 dpi) and fully mature nodules (28 dpi). For both developmental stages *AMY2*, *LjCYP71D353* and *LjCYP88D5* transcripts were localized in the vascular bundles, nodule inner cortex cells and uninfected cells of central tissue. The spatial patterns of expression of the three genes were remarkably similar. Furthermore and in accordance with their expression pattern, the hybridization signals of *AMY2* cluster genes were markedly reduced in nodules 28 dpi. The similarity of the spatial hybridization signals of *AMY2* cluster genes implies a possible involvement of the biosynthetic pathway delineated by the cluster genes in the nodulation processes.

D.2.2. Functional characterization of *AMY2* cluster genes

D.2.2.1. The signature enzyme β-amyrin synthase discloses novel triterpene structures

In this present work the *L. japonicus* full-length *AMY2*, *LjCYP71D353* and *LjCYP88D5* genes were expressed in *N. benthamiana* leaves. Expression of *AMY2* gene in *N. benthamiana* resulted in the production of β-amyrin and dihydrolupeol by AMY2 enzyme. Previously the *L. japonicus* AMY2, when

expressed in yeast, produced β-amyrin and lupeol (Iturbe-Ormaetxe *et al*., 2003). It is possible that non-specific endogenous enzymatic activity in *N. benthamiana* system results in the formation of the dihydrolupeol. Alternatively, one possibility for the formation of this triterpene backbone is that AMY2 delivers a reducing hydride to the lupenyl cation to produce dihydrolupeol. The presence of the lupenyl cation as a biosynthetic intermediate could explain the observed formation of both dihydrolupeol and β-amyrin by AMY2. On the other hand, this finding is in agreement with the *OSC3* gene silencing experiments where it was found that lupeol synthesis can be solely attributed to *OSC3* gene, since all *OSC3*–silencing plants examined by GC-MS analysis lacked the presence of lupeol traces. Further characterization of the enzyme kinetics and action, as well as the production of a dihydrolupeol standard, which has not been achieved as yet, will allow the confirmation of the molecule identity. Efforts were also made in order to detect dihydrolupeol in MeJA-induced *L. japonicus* roots but were not successful; this could be due to conversion to downstream products.

Recently a number of OSCs that catalyze the synthesis of triterpenes have been cloned and their functions confirmed by expression in yeast (Haralampidis *et al*., 2001a). Cloned triterpene synthases from *L. japonicus* offer the opportunity to address the physiological role of triterpenes. Taking into account the novel function of β-amyrin synthase, the *OSC* genes from *L. japonicus* are summarized in table 6.

Gene	Genomic clone	cDNA	Functional characterization in yeast GIL77	Functional characterization in *N. benthamiana*
OSC1	LjT03M08	*LjAMY1* (Iturbe-Ormaetxe *et al.*, 2003), *cOSC1* (Sawai *et al.*, 2006b)	β-amyrin (Sawai *et al.*, 2006b)	-
OSC2	LjT07E09	-	-	-
OSC3	LjT43B09	*cOSC3* (Sawai *et al.*, 2006b)	Lupeol (Sawai *et al.*, 2006b)	-
OSC4	LjB16L08	D9FR	Partially cloned	
OSC5	LjT43B09	*cOSC5* (Sawai *et al.*, 2006b)	Cycloartenol (Sawai *et al.*, 2006b)	-
OSC6	LjT43B09, LjT16A21	Not cloned	-	-
OSC7	LjT16A21	*cOSC7* (Sawai *et al.*, 2006a)	Lanosterol (Sawai *et al.*, 2006a)	-
OSC8	LjT11L01	*LjAMY2* (Iturbe-Ormaetxe *et al.*, 2003)	β-amyrin, lupeol (Iturbe-Ormaetxe *et al.*, 2003)	β-amyrin, dihydrolupeol (This study)

Table 6. Characterized OSCs from *L. japonicus*.

D.2.2.2. The tailoring cytochrome P450 enzymes LjCYP71D353 and LjCYP88D5

D.2.2.2.1. LjCYP71D353 belongs to the CYP71D subfamily of CYP71 clan

The CYP71 clan now represents more than half of all CYPs in higher plants and, consequently, a huge variety of functions. The CYP71 clan after an early trigger to diversification has kept growing, producing bursts of gene duplications at an accelerated rate (Nelson and Werck-Reichhart, 2011). They participate in the metabolism of aromatic and aliphatic amino acid derivatives (i.e. phenylpropanoids, glucosinolates), small isoprenoids and some triterpenoid derivatives, alkaloids and fatty acids (Nelson and Werck-Reichhart, 2011).

Several enzymes of the CYP71 clan have been functionally characterized for their involvement in the biosynthesis of terpenoids. Shibuya *et al.*, (2006) identified CYP93E1, a cytochrome P450 from soybean that participates in soyasaponin biosynthesis as a β-amyrin and sophoradiol 24-hydroxylase. Moreover, the CYP93E1 homolog in *G. uralensis*, CYP93E3, possesses β-amyrin

24-hydroxylase activity (Seki *et al*., 2008) and CYP93E2 is a key oxidase involved in sapogenin biosynthesis in *M. truncatula* (Fukushima *et al*., 2011). The second CYP450 in the *Arabidopsis* genomic region of thalianol cluster, CYP705A5, belongs to the CYP705 family and is characterized as thalian-diol desaturase (Field and Osbourn, 2008). *A. thaliana* CYP71A16 is a marneral oxidase and also belongs to the CYP71 clan (Field *et al*., 2011).

Especially concerning the CYP71D subfamily of CYP71 clan, some members have been reported for their involvement in terpenoid biosynthesis. The limonene-3-hydroxylase and limonene-6-hydroxylase are implicated in monoterpene biosynthesis in peppermint and spearmint, respectively, and are 50% identical to the nearest subfamily member (CYP71D7) from *Solanum chacoense* (Lupien *et al*., 1999). Furthermore, CYP71AV1 is involved in biosynthesis of artemisinin, a sesquiterpene lactone of *Artemisia annua* (*A. annua*), and shares a close lineage with other P450s that catalyze the hydroxylation of monoterpenoids, sesquiterpenoids and diterpenoids (Ro *et al*., 2006). LjCYP71D353 represents a novel cytochrome P450 enzyme of CYP71D subfamily, acting on an unusual triterpene skeleton – dihydrolupeol – giving rise via successive reactions to 20-hydroxy-lupeol and 20-hydroxybetulinic acid. Whether these compounds represent the final or intermediate products of the AMY2 biosynthetic pathway in *L. japonicus* warrants investigation.

Successive hydroxylation/oxidation reactions catalyzed by plant cytochrome P450 enzymes towards other substrates have been documented in *Sorghum bicolor* (Bak *et al*., 1997) and in *Arabidopsis*, maize and barley, in which CYP88A catalyses three steps in the biosynthetic pathway of gibberellins (Helliwell *et al*., 2001). Recently, CYP72A154 from liquorice was shown to catalyze three sequential oxidation steps at C-30 of 11-oxo-β-amyrin to produce glycyrrhetinic acid in yeast (Seki *et al*., 2011). LjCYP71D353 catalyzes oxidation reactions at two different positions of the triterpene skeleton

(hydroxylation at C-20 and acid formation at C-28). A range of multifunctional cytochrome P450 enzymes catalyzing different oxidation reaction at different positions have also recently been reported from bacteria (Anzai *et al*., 2008; Carlson *et al*., 2010; Kudo *et al*., 2010) and fungi (Tokai *et al*., 2007). In *Artemisia annua* the three oxidation steps converting amorphadiene to artemisinic acid by CYP71AV1 have also been described (Ro *et al*., 2006). In the avenacin gene cluster in oat, the cytochrome CYP51H10 catalyses the epoxidation of β-amyrin at C-12/C-13 and also the hydroxylation of β-amyrin at C-16 (Geisler *et al*., 2013). LjCYP71D353 enzyme is an addition to the arsenal of biosynthetic enzymes needed for the production of novel triterpenes by synthetic biology approaches.

D.2.2.2.2. CYP88D5 belongs to the CYP88D subfamily of CYP85 clan

CYP85 may have evolved from a sterol metabolizing CYP51 ancestor and is essentially devoted to the metabolism of medium to large isoprenoids, including brassinosteroids. The CYP85 clan of cytochrome P450 enzymes has expanded considerably, indicating that the diversification of CYPs parallels land plant evolution (Nelson and Werck-Reichhart, 2011).

Several members of the CYP85 clan have been functionally characterized for their involvement in the biosynthesis of triterpenoids. CYP708A2 in the *Arabidopsis* thalianol cluster is biochemically characterized as thalianol hydroxylase (Field and Osbourn, 2008). Three multifunctional enzymes belonging to the CYP716A subfamily that catalyze three sequential oxidation reactions at C-28 of β-amyrin, producing oleanolic acid in *M. truncatula* were reported (Fukushima *et al*., 2011). Additionally, two enzymes of them were able to modify α-amyrin and lupeol, generating ursolic and betulinic acid respectively, when expressed in yeast (Fukushima *et al*., 2011). Homologous enzymes to CYP716A have also been reported in other plants. CYP716A15 and CYP716A17 catalyze the formation of oleanolic acid, betulinic acid and ursolic

acid in *Vitis vinifera*, with similar catalytic activities to CYP716A12 of *M. truncatula* (Fukushima *et al*., 2011). In addition, CYP716A47 and CYP716A53v2 catalyze the formation of dammarene-type aglycones in *Panax ginseng* (Han *et al*., 2011; 2012).

The cytochrome P450 genes within the *AMY2* and *βAS* cluster, *LjCYP88D4/5* and *MtCYP88D1*, belong to the legume-specific CYP88D subfamily, members of which have previously been characterized and assigned roles in triterpene biosynthesis. Specifically, GgCYP88D6 (Seki *et al*., 2008) exhibiting 72% similarity with LjCYP88D5 is a β-amyrin 11-oxidase. Unexpectedly, no activity was detected towards β-amyrin, when *LjCYP88D5* is expressed with *AMY2* in *N. benthamiana* leaves. No further products were detected when AMY2, LjCYP71D353 and LjCYP88D5 were co-expressed together. This could be attributed to the heterologous system employed. Alternatively, LjCYP88D5 may have a different function than GgCYP88D6 in next steps of the proposed biosynthetic pathway, recognizing another molecule as a substrate. It is also likely that the LjCYP88D5 product is not detected because of the low-quantity of the resultant metabolite. Alternatively, the resultant metabolite of LjCYP88D5 reaction is quickly modified and, as a result, it cannot be detected by GC-MS analysis. The heterologous expression in yeast cells will presumably disclose a novel function for LjCYP88D5.

Saccharomyces cerevisiae through manipulation of its native ergosterol biosynthesis has already been employed for the expression of triterpenoid biosynthetic genes (Mosses *et al*., 2013). The β-amyrin-producing *S. cerevisiae* strains could be utilized as a tool for the in vivo expression and characterization of novel P450s enabling the functional characterization of LjCYP88D5.

D.2.2.2.3. Several metabolic pathways involve divergent enzymes

The CYPs within a single family or branch usually metabolize similar/related compounds. However, consecutive steps in the same metabolic pathway can also involve completely divergent enzymes (Nelson and Werck-Reichhart, 2011). The first example of the convergent evolution of P450s comes from the gibberellin biosynthesis in higher plants. In *Arabidopsis*, CYP701A3 belonging to CYP71 clan and CYP88As (CYP88A3 and CYP88A4) belonging to CYP85 clan, have been shown to catalyze very similar enzymatic reactions and two successive steps in the pathway of gibberellin biosynthesis (Helliwell *et al*., 2001). In addition, the P450 enzymes involved in triterpene saponin biosynthesis in *G. uralensis* are recruited in at least two very distant CYP families, CYP93 and CYP88, which belong to the CYP71 and CYP85 clans respectively. It is likely that the P450 enzymes, participating in triterpene saponin biosynthesis in other legumes, are also recruited in CYP71 and CYP85 clans (Seki *et al*., 2008). Interestingly and in agreement with triterpene saponin metabolic pathways in *G. uralensis* and *A. thaliana* LjCYP71D353 enzyme belongs to the CYP71 clan of P450s, whereas LjCYP88D5 enzyme, as well as LjCYP88D4 of the *AMY2*-like gene cluster and MtCYP88D1 of the *M. truncatula* β-amyrin gene cluster, to the CYP85 clan.

It is not always instructive to discuss the functions CYP families based on their sequence relatedness. As plant evolved from algae to angiosperms, the emergence of new biochemical functions has been a crucial issue. Therefore, an evolutionary approach can serve as a guide to CYP functions (Nelson and Werck-Reichhart, 2011).

D.2.2.3. Other / uncharacterized genes in *AMY2* gene cluster

The LjSDRt enzyme identified in the *AMY2* gene cluster has extensive similarity to a SDR-like protein that has been identified and characterized in *A. thaliana*, namely HCF173 (Schult *et al*., 2007). SDR proteins are enzymes which

use NAD(P)(H) as coenzymes and function as dehydrogenases, dehydratases, epimerases and isomerases of a variety of substrates (Kallberg *et al*., 2002). HCF173 does not exhibit its original dehydrogenase activity, but has possibly gained other functions during evolution (Schult *et al*., 2007).

Several SDR-related proteins have lost their dehydrogenase activity but are adapted to different functions in cells. One such example is the SDR-like enzyme CSP41, which exhibits endonucleolytic activity, is able to bind to specific RNAs and initiates degradation of plastidic mRNAs (Yang and Stern, 1997). The dinucleotide binding domain of the protein realizes the RNA binding, indicating that the domain has gained a function on chloroplast RNA metabolism (Bollenbach and Stern, 2003). RNA binding via this domain has been shown for several NAD^+-dependent dehydrogenases and (di)nucleotide binding metabolic enzymes that play an additional role in RNA metabolism (Hentze, 1994; Nagy *et al*., 2000). It is proposed that HCF173 is also likely to be an RNA binding protein, which possibly interacts with the *psbA* mRNA, affecting its translational activity and stability. The *psbA* gene encodes the D1 protein, which participates in the core complex of the photosystem II reaction. HCF173 directly interacts with *pbsA* mRNA as part of a high molecular weight complex in chloroplast membranes (Schult *et al*., 2007). The high similarity between LjSDRt and HCF173 suggests that LjSDRt may also represent a regulatory protein, having lost a metabolic function. However, no other homologues to HCF173 in other plants have been characterized as yet to allow for a justified functional prediction for LjSDRt.

Characterization of the genes required in the latter steps in terpenoid pathway, primarily cytochrome P450- and glycosyltransferase-mediated modifications, and their regulatory components will represent the next important developments in unraveling triterpenoid biosynthesis (Jenner *et al*., 2005). In LjT11L01 genomic clone of *L. japonicus* several scattered genes with unknown function are found, which can be unrelated or related to *AMY2*

cluster genes. The identification of their gene expression or function is at least tempting, as it is likely that interesting data will be unveiled. Gene co-expression analysis has emerged as a powerful tool for predicting gene function, because genes whose expression patterns are strongly correlated with each other are expected to be involved in the same biological processes. Gene co-expression analysis along with combinational biosynthesis of diverse triterpenoid structures enable the identification of the enzymes involved in their biosynthetic pathway (Fukushima *et al*., 2013).

D.2.2.4. Three modules in secondary metabolic pathways

Three modules can be distinguished in a secondary biosynthetic pathway: the branchpoint reaction, the chemical modification leading to a biological active compound and, finally, the detoxification of the active compound (Frey *et al*., 2009). In maize benzoxazinoid biosynthetic gene cluster, the pathway-specific branchpoint is generated by neo-functionalisation of a gene duplicate originating from primary metabolism, *Bx1* (Frey *et al*., 1997). Modification of the intermediates by consecutive hydroxylation is catalyzed by members of cytochrome P450 enzyme subfamily (Bx2-Bx5) (Frey *et al*., 1995). Glucosylation by UDP-glucosyltransferases (UGTs, Bx8 and Bx9) is essential for the detoxification of DIBOA and DIMBOA. Detoxification is pathway-specific and an integral part of the biosynthesis. It is tempting to speculate that evolution of a secondary metabolic pathway can happen in the presence of a detoxifying reaction (Frey *et al*., 2009).

In oat avenacin gene cluster, the branchpoint reaction between primary and secondary metabolism is the cyclization of 2,3-oxidosqualene by β-amyrin synthase (Sad1) (Haralampidis *et al*., 2011). *Sad2* encodes a cytochrome P450 enzyme belonging to the CYP51 sterol demethylase family (Qi *et al*., 2006), which recently has been biochemically characterized as a multifunctional enzyme participating in three successive oxidation reactions of β-amyrin

(Geisler *et al*., 2013). Sad3 and Sad4 are the necessary enzymes for the glycosylation step (Mylona *et al*., 2008), whereas Sad7 for acylation (Mugford *et al*., 2009). Recently, a gene encoding an anthranilate *N*-methyltransferase (MT1, encoded by *Sad9*) was identified and shown that acts together with the UGT74H5 glucosyltransferase (*Sad10*, Owatworakit *et al*., 2012) and the SCPL1 acyltransferase (*Sad 7*, Mugford *et al*., 2009) in the final steps of the synthesis of avenacin in oat (Mugford *et al*., 2013).

In thalianol and marneral biosynthesis the branchpoint reaction is the conversion of 2,3-oxidosqualene to thalianol or marneral, respectively. The chemical modifications of the triterpene skeletons require cytochrome P450 enzymes. Whether the resultant compounds represent the final or intermediate products of the thalianol and marneral biosynthetic pathways in *A. thaliana* requires further investigations (Field and Osbourn, 2008; Field *et al*., 2011).

In this present study, the branchpoint reaction in triterpene biosynthesis is the cyclization of the common sterol/triterpene precursor 2,3-oxidosqualene by β-amyrin synthase (AMY2). A new cytochrome P450 enzyme, LjCYP71D353, which catalyses the formation of 20-hydroxybetulinic acid in a sequential three-step oxidation of 20-hydroxylupeol was characterized. In other words, LjCYP71D353 participates in the chemical modification step, the second module of the secondary metabolic pathway. Taking into account the complexity and the abundance of the reactions needed in the other biosynthetic gene clusters described, it is likely that LjCYP71D353 resultant metabolites are intermediate products of the AMY2 biosynthetic pathway in *L. japonicus*. The "three module" theory supports possible functions for the third cluster member, LjCYP88D5, or the fourth partner, LjSDRt, which remain unidentified.

D.2.3. Unraveling the possible roles of *AMY2* gene cluster

The final metabolites produced by the biosynthetic pathway mediated by the *AMY2* cluster genes in the tissues and organs of *L. japonicus* have not been determined as yet. The signature enzyme of the cluster AMY2 apparently produces two different basic triterpene skeletons, β-amyrin and dihydrolupeol. These skeletons could be further elaborated by a number of tailoring enzymes, such as CYP71D353, the activity of which results in 20-hydroxy-lupeol and 20-hydroxybetulinic acid. The role of such metabolites in plant physiology has been scarcely studied.

D.2.3.1. The involvement of *AMY2* gene cluster in legume symbiotic interactions

It has been proposed that simple triterpenes act as structural components of membranes during normal growth and development (Baisted, 1971; Nes and Heftmann, 1981), as well as during the establishment of rhizobial and mycorrhizal symbiotic relations (Hernandez and Cooke, 1996; Grandmougin-Ferjani *et al*., 1999). From an analysis of *P. sativum* nodules, β-amyrin was detected in peribacteroid membranes and the microsomal fraction of nodule cells, which mostly corresponded to plant cell endoplasmic reticulum cells, and was not detected in free living bacteria, suggesting a possible role of β-amyrin in nodules (Hernandez and Cooke, 1996). Transcript levels of β-amyrin and lupeol synthase genes are induced during symbiotic relations in several plants (Iturbe-Ormaetxe *et al*., 2003; Hayashi *et al*., 2004; Sawai *et al*., 2006a; this study), also suggesting the involvement of simple triterpenes in legume/rhizobium interactions. Confalonieri *et al*. (2009) demonstrated that the expression of a β-amyrin synthase gene from *Aster sedifolius*, *AsOXA1*, has an effect on the symbiotic nodulation performance. The ectopic expression of *AsOXA1* in transgenic plants led to enhanced root nodulation in *M. truncatula*, proposing a positive involvement of β-amyrin in the regulation of nodulation

(Confalonieri *et al*., 2009). In this study, silencing of *βAS* gene resulted in enhanced nodulation in *M. truncatula*, proposing a negative involvement of β-amyrin in nodule formation. The role of β-amyrin in root nodule biology remains to be elucidated. Furthermore, lupeol were shown to act as negative regulators of nodule formation, and thus cell proliferation.

The high and coordinate expression of *AMY2* cluster genes in inoculated with *M. loti* roots and in nodules, resulted in the investigation of the tissue specificity of *AMY2* cluster expression by *in situ* hybridization. Interestingly, the *AMY2* and *OSC3* hybridization signals are localized in the same tissues; that are inner cortex, vascular bundles and uninfected cells of the central tissue. The localization of *AMY2* cluster activity in uninfected cells and in nodule vascular bundles suggests that the metabolic product(s) of the cluster are associated with structural needs in developing nodule symbiosis (Baisted, 1971; Hernandez and Cooke, 1996). As the spatial expression of *AMY2* cluster genes is reduced in mature nodules, it is possible that the structural needs are being reduced when nodules actively fix nitrogen. Alternatively, the presence of *AMY2* cluster gene transcripts in nodule vascular bundles could indicate a role in the transportation of compounds in and out of the nodule.

The *OSC3* activity and, thus, the production of lupeol in nodular tissues may be responsible for the synthesis of lupeol derivatives such as betulin, which serves as a constitutive pathogen defense barrier and possibly reducing water loss (Hartmann *et al*., 2002). It is possible that *AMY2* cluster genes are involved in the production of a hydrophobic and antibiotic triterpene derivative, which is transported to outer cortex, making the nodule surface water-repellent and resistant to microbial infection (Hayek *et al*., 1989).

D.2.3.2. The involvement of *AMY2* cluster in plant growth and development

In most cases the physiological role of triterpenoids in not clear but should be addressable by producing transgenic plants with altered triterpene synthase activities and examining the resulting phenotype. The remarkably similarity of spatial patterns of the *AMY2* cluster transcripts in nodules demands a more thorough investigation of the role of these genes in nodulation. For this reason a hairpin RNA gene silencing strategy was carried out using an *A. rhizogenes* transformation protocol. This system is highly amenable for functional analysis as the *AMY2* cluster genes show a preference of organ-specific expression in roots and nodules. Furthermore, the hairpin RNA-mediated gene silencing gives rise to a number of transformation events that are considered as separate transgenic lines. Thus, if a phenotype is replicated among the population of generated plants, it is likely that the phenotype is due to silencing of the target gene.

Two intron-spliced hairpin RNA silencing constructs were made for the *AMY2* gene, two constructs targeted to the *LjCYP88D5* gene and one construct for *the LjCYP71D353* gene. Transgenic roots exhibiting significantly decreased levels of *AMY2*, *LjCYP88D5* or *LjCYP71D353* transcripts were generated. Unexpectedly, in *AMY2–*, *LjCYP71D353–* and *LjCYP88D5–* RNAi plant lines the number of nodules did not differ when compared with control plant lines and no defect on nodule morphology was macroscopically observed. Interestingly, at 40 dpi a severe retardation of the rate of hairy-root growth was recorded in *AMY2* silenced lines, when compared to 20dpi.

In order to further examine the physiological role of the AMY2 cluster, stably transformed lines silenced for *AMY2* were also obtained via *A. tumefaciens* transformation. The four transformed plants did not flower, and so it was unable to obtain seeds, and consistently exhibited a short, stunted root phenotype, indicating that silencing of *AMY2* in stably transformed plants

has clear effects on growth and development. The investigations for a phenotype were not exhaustive and were restricted macroscopically for defects in plant growth and nodulation. Additional conditions should be studied in the future.

The recent generation of TILLING and LORE1 mutant plant lines is proved as a valuable tool for functional analysis of *L. japonicus* genes (McCallum *et al*., 2000a, b; Perry *et al*., 2003; Le Signor *et al*., 2009; Urbanski *et al*., 2012). TILLING (targeting induced local lesions in genomes) is a reverse genetic strategy for the identification of mutations throughout a genome and a screening method facilitating the localization of these mutations. Mutations are induced by the chemical mutagen ethyl methane sulphonate (EMS), and particular regions can be screened for the presence of mutations by high-throughput polymerase chain reaction (PCR) (Le Signor *et al*., 2009). LORE1 (*L. japonicus* retrotransposon 1) belongs to the Ty3-gypsy group of elements, and is a long terminal repeat (LTR) retrotransposon. Analysis of native integration sites and new insertions shows preferable insertion into genes and low-abundance sequences, rather than highly repetitive DNA in, for example, centromeres or telomeres (Madsen *et al*., 2005). LORE1 is transcribed in all organs analyzed and it does not appear to be transcriptionally up-regulated during in vitro tissue culture. Activity of LORE1 in callus and whole plants suggests that a simple insertion mutagenesis based on endogenous LORE1 elements can be established for *L. japonicus* (Madsen et al., 2005). Urbanski *et al*., (2012) have developed the *FSTpoolit* protocol and software package that efficiently detected 8935 LORE1 insertions in 3744 *L. japonicus* plants. As LORE1 transposition is identified in the germline, a complete mutant population is generated by harvesting seeds from a single founder line and cultivating progeny (Urbanski *et al*., 2012). The comprehensive reverse genetic resource eliminates the time required for generating transgenic plants via *Agrobacterium*-mediated protocols. Such TILLING and LORE1 mutant lines for

the cluster genes have already been identified, offering a new opportunity to further characterize the role of cluster genes in plant development and physiology, as well as to validate the functional relation amongst them.

D.2.3.2.1. The significance of *AMY2* clustering

Similar phenotypes to those recorded in *AMY2* transgenic plant lines have been recorded in oat avenacin cluster and in *Arabidopsis* thalianol and marneral clusters. *Sad1* mutants have normal root morphology, but mutations in *Sad3* and *Sad4* loci resulted in stunted root growth, membrane trafficking defects in the root epidermis and root hair deficiency in oat. It is proposed that avenacin gene cluster is likely to hamper the accumulation of deleterious intermediates (Mylona *et al*., 2008). *Arabidopsis* plants overexpressing thalianol, thalian-diol or marneral have pronounced dwarf phenotypes, suggesting that there is a need for high co-ordinate regulation of the pathway. Additionally, the elevated accumulation of the triterpene thalianol in *A. thaliana* resulted in enhanced root length (Field and Osbourn, 2008; Field *et al*., 2011). In *Arabidopsis mrn1* knock-out mutant round-shaped leaves, late flowering, short anther filaments, root growth retardation and delayed embryo development have been displayed (Go *et al*., 2012). The severe retardation in root growth was observed in silenced plant lines for *AMY2* acquired via *A. tumefaciens* and *A. rhizogenes* transformation, pointing out that the phenotype is not an accidental. Coinheritance of favorable combinations of genes and prevention of deleterious intermediates accumulation are the selective advantages to drive the evolution of gene clusters (Field and Osbourn, 2008; Mylona *et al*., 2008; Field *et al*., 2011; Takos *et al*., 2011).

Most plant metabolic gene clusters discovered so far are for synthesis of compounds that are implicated in innate and induced disease resistance, insect resistance, abiotic stress tolerance and/or allelopathy (Chu *et al*., 2011; Kliebenstein and Osbourn, 2012). A number of apparently opposed activities

have been reported for the role of triterpenes in plant developmental processes. The stunted root growth and root hair deficiency in oat avenacin mutants (Mylona *et al*., 2008) oppose to the enhanced root length in *A. thaliana* plant lines displaying elevated accumulation of the triterpene thalianol (Field & Osbourn, 2008; Field *et al*., 2011). These observations open up further questions about whether growth inhibition/promotion mediated by triterpenes occurs via different pathways or through antagonistic effects on a common pathway (Osbourn *et al*., 2011). The inducibility of the *AMY2* gene cluster provides the opportunity to further pursue and define such a role, as yet unidentified, in plant growth and development processes for the triterpene metabolites produced by this new biosynthetic pathway.

D.2.3.3. The possible involvement of *AMY2* cluster genes in saponin biosynthesis

It is generally accepted that elevated levels of β-amyrin may be an indication of active saponin biosynthesis. In *G. glabra*, the transcript of β-amyrin synthase was highly expressed in the cultured cells, root nodules and germinating seeds, where soyasaponin accumulates, and in the thickened roots where glycyrrhizin accumulates (Hayashi *et al*., 2004). The ectopic expression of the *Aster sedifolius* β-amyrin synthase gene, *AsOXA1,* in *M. truncatula* led to an increased production of triterpenoid saponins in leaves and roots (Confalonieri *et al*., 2009). In *A. sativa* the expression of *AsSad1,* and presumably of the β-amyrin produced by Sad1 in the root epidermis, coincides with the presence of the triterpenoid avenacins in this tissues (Haralampidis *et al*., 2001).

In legumes, β-amyrin is considered the precursor of saponins that have been identified so far. Saponins from pea have been proposed to regulate gravitropism and cellulose synthesis in plants (Ohana *et al*., 1998; Rahman *et al*., 2001). Soyasaponin I was reported to be a phytochrome killer in pea

(Yokota *et al*., 1982). Chromosaponin I (CSI) is a γ-pyronyl-triterpenoid saponin isolated from pea (Tsurumi *et al*., 1991, 1992) and other leguminous plants (Kudou *et al*., 1992; Massiot *et al*., 1992) that has been shown to influence the growth of roots in several plants (Tsurumi and Wada, 1995). CSI increases the mechanical extensibility of cell walls in roots, increases the length of the cells and reduces the diameter of the root, and stimulates the root growth (Tsurumi *et al*., 1996; Tsurumi and Ishizawa, 1997). Glycyrrhizin presents various pharmacological activities, including anti-inflammatory (Matsui, 2004), immunomodulatory (Takahara *et al*., 1994), antiulcer (He *et al*., 2001), antiallergy (Park *et al*., 2004) and antiviral against various DNA and RNA viruses including HIV (Baba *et al*., 1988).

The expression of the β-amyrin synthase gene *MtAMY1* in the roots and leaves of *M. truncatula* is consistent with a role for MtAMY1 in the synthesis of triterpenoid saponins in these tissues (Iturbe-Ormaetxe *et al*., 2003). *M. truncatula lha* mutant lines, unable to synthesize hemolytic saponins, showed a retarded growth phenotype, suggesting that saponin biosynthesis plays a role in plant growth processes (Carelli *et al*., 2011). Saponins isolated from different species of the genus *Medicago* have been reported to possess antimicrobial, cytotoxic and insecticidal activities (Tava and Avato, 2006).

It is likely that the *AMY2* cluster genes may cooperate for the production of a defense-related saponin end product in roots. Alternatively, the short, stunted root phenotype recorded in *AMY2* transgenic plant lines is in agreement with *lha M. truncatula* mutants, proposing that the *AMY2* cluster genes produce a saponin molecule that plays a role in plant growth processes. Currently, the saponin content of *L. japonicus* is not known. β-amyrin-derived triterpenoid saponins are found in seeds of *Lotus* species (Jurzysta *et al*., 1973), whereas information on the saponin content of the roots and leaves of these species is scant.

D.2.4. Mechanisms of gene cluster regulation

According to several recent studies, genomes have a specific functional architecture that acts to regulate gene transcription; locally through *cis*-regulatory sequences and globally through the modification of large segments of the genome (Grewal and Moazed, 2003). *Cis*-regulatory sequences interact with transcription factors to promote or repress the transcription of target genes. Transcriptional co-regulation of genes can be influenced at different levels (Sproul *et al*., 2005). Tandem duplicates may have similar expression patterns because they have similar promoters. In some cases, co-expression of adjacent genes that are encoded on opposite DNA strands may be co-regulated by a common bidirectional promoter, although fusion of linked genes with related functions to make a single protein product represents a rather unwonted mechanism of managing co-expression (Zhang and Smith, 1998; Gross *et al*., 2006). In *AMY2* cluster, the two cytochrome P450 genes are adjacent to each other and probably share a common regulatory promoter region. None of the above mechanisms is sufficient to thoroughly describe the genomic organization and regulation of gene clusters. However, gene clustering may provide various other mechanisms for co-ordinated transcriptional regulation, including shared long-distance regulatory elements, or regulation through chromatin structure or nuclear organization (Sproul *et al*., 2005; Osbourn and Field, 2009; Osbourn 2010a, b).

D.2.4.1. Transcription factors regulate gene cluster expression

Genes with correlated expression across microarray experiments are frequently bound by common transcription factors or may be within the same regulatory pathway (Allocco *et al*., 2004). In fungi, many secondary metabolic gene clusters contain cluster specific transcription factors (TFs), often C6 zinc binuclear cluster proteins that function to activate biosynthetic genes in their respective cluster. Representative examples are the AflR TF for aflatoxin

/sterigmatocystin biosynthesis in *Aspergillus* spp. (Fernandes *et al*., 1998) or Tri6 for trichothecene biosynthesis in *Fusarium* spp. (Proctor *et al*., 1995). Several secondary metabolic gene clusters in fungi are also activated, and sometimes shut down, in response to a variety of environmental conditions, like pH, carbon source, nitrogen source, light, ROS and temperature (Calvo *et al*., 2002). Environmental stimulations are translated to the nucleus through signal transduction cascades, such as the mitogen activating protein kinase (MAPK) cascade and the cAMP mediated PkaA cascade (Tag *et al*., 2000; Shimizu *et al*., 2003; Atoui *et al*., 2008) and have been linked to activation of specific broad domain regulator factors including CreA (carbon metabolism), AreA (nitrogen metabolism) and PacC (pH sensor) (Yu and Keller, 2005).

Sharing regulatory elements and histone modifications are the described mechanisms that regulate the coordinate gene expression in mammals. Efficient β-globin transcription requires a Locus Control Region (LCR), located ~10kb upstream of the first gene in the cluster. The embryonic ε-globin requires DNaseI hypersensitive sites (DHSs) that are located in the LCR region and more to the 5′ and 3′ end of the locus, are physically close to each other in the nucleus (Carter *et al*., 2002) and can be cross-linked to one another (Tolhuis *et al*., 2002). This is thought to form the base of a looped domain to which globin genes are then sequentially recruited when they are expressed. The LCR is in proximity with the actively transcribing globin gene to form what has been termed an active chromatin hub (ACH) (Carter *et al*., 2002; Tolhuis *et al*., 2002). This structure is detectable at an early stage of erythroid differentiation and does not require transcription (Palstra *et al*., 2003). Furthermore, developmental switching of globin gene expression is reflected in chromatin fiber alterations in which active are in closer proximity to the LCR than the inactive ones (Palstra *et al*., 2003). The transcription factors GATA-1 and FOG-1 anchor a loop between distant regulatory elements to activate transcription (Vakoc *et al*., 2005). This chromatin architecture is specific and

depends on erythroid transcription factors (Drissen *et al*., 2004; Vakoc *et al*., 2005). Within the Hox clusters there are also localized control elements and enhancers that function over a few neighboring genes. However, only in the case of the homeobox D clusters (*Hoxd*) there is evidence for global regulatory elements that are located outside of the clusters. The global control region GCR that lies to the 5′ end of the *Hoxd* cluster regulates the expression of 5′ *Hoxd* genes in the distal limb bud, but it also regulates genes that are unrelated to the *Hoxd* genes (Spitz *et al*., 2003). In humans, there is evidence that comparable mechanisms can explain the inactivation of blocks of tissue-specific genes. The zinc-finger gene-specific repressor element RE-1 silencing transcription factor (REST) can mediate restriction of gene activity in non-neuronal tissues by imposing active repression through histone deacetylase recruitment. Through the recruitment of an associated co-repressor (CoREST) it can also challenge long-term gene silencing that spreads down the chromosome, affecting transcriptional units, which do not themselves contain REST response elements (Lunyak *et al*., 2002).

D.2.4.1.1. Transcription factors in plant secondary metabolic pathways

Common up-stream *cis* elements that are recognized by a TF for regulation of diterpene biosynthesis in rice have been identified (Okada *et al*., 2009), but this transcription factor appears to be a global regulator of diterpenes and it is not specific for the two characterized rice diterpene clusters. The OsTGAP1 is the chitin oligosaccharide elicitor-inducible basic leucine zipper transcription factor, which is essential for momilactone biosynthesis and regulates the expression of the five genes in the cluster. The knock-out mutant for *OsTGAP1* had almost no expression of the five clustered genes or production of momilactones upon elicitor treatment. The expression *of OsKSL7*, involved in phytocassane production, was largely affected in the knock-out mutant for *OsTGAP1*, although the phytocassane synthesis still

occurred. On the other hand, OsTGAP1-overexpressing plant lines exhibited elevated expression of the clustered genes, strong accumulation of momilactones in response to the elicitor and enhanced accumulation of phytocassanes (Okada *et al*., 2009). The *in silico* analysis of *P. somniferum* noscapine cluster for *cis*-acting DNA elements revealed the presence of WRKY elements (Winzer *et al*., 2012).

Other transcriptional regulators have been defined for several secondary metabolic pathways, but the genes participating in do not assemble in plant secondary metabolic gene clusters. For example, many genes in the phenylpropanoid pathway possess the same TF-binding *cis*-elements and are regulated by the same TF (Memelink *et al*., 2001). In *Arabidopsis*, the coordinate expression of two genes involved in the phytoalexin camalexin biosynthesis, *PAD3* and *CYP71A13*, is regulated by the AtWRKY33 TF (Qiu *et al*., 2008), but it is still unknown if the same TF is the key regulator of multiple camalexin biosynthetic genes.

It is obvious that TFs have an important role in the tight coordination of metabolic gene expression (Pauwels *et al*., 2009). The recruitment and assembly of genes to JA-induced regulons is considered as a series of evolutionary events that occurred independently in different plant species. On the contrary, the transcriptional machinery that drives the JA response seems to be conserved in different plant species, as demonstrated in *Arabidopsis*, tobacco and periwinkle (Pauwels *et al*., 2009). In terpenoid indole alkaloid metabolism and primary precursor pathways jasmonate induces gene expression and metabolism via ORCAs, which are members of the AP2/ERF-domain family of TFs. It is likely that other jasmonate-regulated secondary metabolic pathways might also be controlled by ORCA-like AP2/ERF domain TFs (Memelink *et al*., 2001). According to the *in silico* analysis in this work, no ORCA-like AP2/ERF domain TFs were recognized. Instead the T/GBOXATPIN2, a T/G box *cis*-element that is involved in jasmonate induction of *LAP* genes,

was identified (Boter *et al*., 2004). Many more other *cis*-elements, common for the *AMY2*, *LjCYP71D353*, *LjCYP88D5* and *LjSDRt* genes were also identified. For example, the ARFAT is an auxin response factor that regulates the expression of auxin response genes (Goda *et al*., 2004). The SURECOREATSULTR11 is a sulfur-responsive element found in the promoter of the SULTR11 *Arabidopsis* transporter gene, which contains an auxin-response factor binding sequence (Maruyama-Nakashita *et al*., 2005). The Dof protein DOFCOREZM is a DNA binding protein reported to participate in the regulation of gene expression in processes like plant defense, gibberellin response and auxin response (Yanagisawa and Schmidt, 1999). The ARR1AT is a response regulator that is involved in an early step of cytokinin transduction, possibly as a partner of the sensor histidine kinase CRE1 in *L. japonicus* (Sakai *et al*., 2001). The GT1GMSCAM4 is a GT-1 motif found in the promoter of SCaM-4 gene and plays a role in pathogen- and salt-induced SCaM-4 gene expression (Park *et al*., 2004). The MYCCONSESUSAT is a recognition site found in many Arabidopsis promoters and regulates the transcription of *CBF/DREB1* genes in cold conditions (Chinnusamy *et al*., 2003; 2004). The WRKY71OS belongs to WRKY transcription factors, which play important roles in plant responses to biotic and abiotic stresses (Zhang *et al*., 2004; 2005). Promoter regulatory elements expressed specifically in roots and root hairs are also recorded. The OSE1ROOTNODULE motif is a characteristic *cis*-element of the promoters activated in infected cells of root nodules (Vieweg *et al*., 2004; Fehlberg *et al*., 2005). The RHERPATEXPA7 is a root hair-specific *cis*-element (Kim *et al*., 2006) and ROOTMOTIFTAPOX1 is a *cis*-element found in the promoter of rolD gene in roots (Elmayar *et al*., 1995). The NODCON2GM is one of the two putative nodulin consensus sequences controlling the nodule-specific expression of leghemoglobin *lbc3* gene in soybean (Stougaard *et al*., 1990). The *in silico* recognition of such transcription factors could explain the coordinate expression of *AMY2* cluster genes in response to various

phytormones, abiotic stresses and nodulation. However, the performance of promoter deletion analysis will elucidate these early results.

D.2.4.1.2. Metabolons in secondary metabolism

The formation of metabolons (multienzyme complexes) and metabolic channeling in plant secondary metabolic pathways enable plants to effectively synthesize specific natural products. Several reasons explain the necessity of metabolon formation. Metabolons improve catalytic efficiency by channeling an intermediate that is formed at one active site of an enzyme to the active site of the next enzyme. In other words, they bring co-operating active sites into close proximity, resulting in the decreasing of the transit time for intermediates. In addition, they secure the fast conversion of labile and/or toxic intermediates into more stable and less toxic constituents by preventing their diffusion into the surrounding cell matrix, where chemical decomposition would take place. Furthermore, metabolons control and coordinate metabolic cross-talk that is mediated either by enzymes that function in different pathways or by intermediates that are shared between different metabolic pathways. Finally, they can guide the formation of new metabolons as might be demanded by environmental challenges (Jorgensen *et al*., 2005).

Metabolon formation typically involves specific interactions between several soluble enzymes that might be anchored to a membrane either by membrane-bound structural proteins that serve as nucleation sites for metabolon formation or by membrane-bound proteins, such as cytochrome P450s, that directly catalyze one or more of the sequential channeled reactions carried out by the metabolon (Jorgensen *et al*., 2005). Plants recruit a large number of enzymes that generate the huge diversity of plant secondary products. Many of these downstream enzymes that decorate the backbone structures are regioselective or regiospecific rather than highly substrate specific (Chau *et al*., 2004; Frick *et al*., 2001). The positioning of enzymes that

have broad substrate specificity downstream of the conserved early pivotal enzymes of plant secondary metabolism opens the possibility of producing new secondary compounds without major re-structuring of the enzyme complement. Metabolic channeling and metabolon formation provide the key to resolving and avoiding potential negative interference in plant natural product formation either by narrowing substrate specificity as a result of conformational changes upon binding or because binding into the metabolon prevents access of unwanted substrates (Jorgensen *et al*., 2005).

Isoforms of the membrane-bound enzyme 3-hydroxy-3-methylglutaryl coenzyme A reductase are localized in various types of vesicles that carry different enzymatic activities, and thereby provide a physical basis for the channeling of isoprenoid synthesis (Leivar *et al*., 2005). Most of the enzymes in phenylpropanoid metabolism appear to be organized in different metabolons (Achnine *et al*., 2004). It has also been demonstrated that flavonoid metabolism is catalyzed by an enzyme complex localized in the endoplasmic reticulum, which is composed by a chalcone synthase, a chalcone isomerase, a dihydroflavonol 4-reductase and a flavonol 3-hydroxylase (Burbulis and Winkel-Shirley, 1999). In vitro experiments demonstrated that the cyanogenic glucoside dhurrin in sorghum pathway is highly channeled (Moller and Conn, 1980). There is strong evidence that the initial steps in polyamine synthesis are organized in a metabolon (Panicot *et al*., 2002). Such protein complexes, like gene clusters, facilitate the spatial and temporal regulation of metabolic fluxes, which in turn indicates the essential role of TFs in signal transduction leading to the production of plant secondary metabolites (Zhao *et al*., 2005). It is likely that the common TFs that bind on *AMY2* cluster genes coordinate a protein complex formation, which facilitates the effective synthesis of the end pathway product.

D.2.4.2. Chromatin modifications influence the transcription of genes

The role of chromatin structure in modulating gene expression has been extensively studied at the level of histones and their post-translational modification. Chromatin can exist in two states; euchromatin is transcriptionally active and characterized by low nucleosome density, while heterochromatin is transcriptionally silent and contains densely packed nucleosomes. Heterochromatin is referred to as facultative heterochromatin because it can become activated under particular circumstances, as demonstrated by developmentally timed gene expression in the genus *Drosophila* (Troger and Reinberg, 2007). How histone modifications influence transcription is still debated. They can act as binding sites for specific regulatory proteins (Fischle *et al*., 2003) and may directly affect higher-order chromatin structure and condensation (Wolffe and Hayes, 1999; Carruthers and Hansen, 2000; Wang *et al*., 2001).

Generally, lysine acetylation correlates with transcriptional activation (Kouzarides 2007). Lysine methylation leads to either transcriptional activation or repression depending on which residues are modified and the type of modification present. For example, methylation of lysine 4 and lysine 36 on histone H3 are associated with gene activation (Zhang *et al*., 2009), whereas methylation of lysine 9 and lysine 27 are associated with gene silencing (Noma *et al*., 2001; Kouzarides 2007; Liu *et al*. 2010; Roudier *et al*., 2011; Luo *et al*., 2013). Furthermore, methylation of cytosine together with dimethylation of lysine 9 on histone H3 (H3K9me2) was found to silence transposable elements (Bernatavichute *et al*., 2008; Cokus *et al*., 2008). These generalities are flexible, as H3K4 methylation is also associated with silencing in yeast subtelomeric and rDNA regions (Mueller *et al*., 2006).

In fungi, secondary metabolic gene clusters are often located subtelomerically and, recently, it has been demonstrated that proteins

involved in chromatin remodeling, such as LaeA, CirD, CcIA and HepA, mediate gene cluster regulation. The critical role for chromatin level regulation by histone methylation of fungal secondary metabolic gene cluster genes raises the possibility that epigenetic mechanisms based on histone post-transcriptional modifications are a general mechanism to silence fungal secondary metabolic genes clusters (Bok *et al*., 2009). Finally, several data suggest that cryptic or silent secondary metabolite gene clusters in fungi are located in regions of facultative heterochromatin and can be turned on when chromatin structure is altered (Palmer and Keller, 2010). Hyperacetylation of histones at the β-globin LCR have been observed, but the levels of histone acetylation at the globin genes themselves depend on the stage of development (Forsberg *et al*., 2000). In a remarkably similar situation, induction of the temporal program of *Hoxb* expression during the differentiation of embryonic stem cells is accompanied by histone acetylation at early (3') and late (5') genes simultaneously (Chambeyron and Bickmore, 2004). Therefore, for both β-globin and *Hoxb* loci, histone acetylation could be equated with a locus-wide potentiation for transcription. Methylation of histone H3 at lysine 4 (H3K4) is also associated with the activation of gene expression at the *Hoxb* locus (Chambeyron and Bickmore, 2004).

D.2.4.2.1. Chromatin modeling regulates coordinate expression in plant gene clusters

There are recent indications that co-ordinate expression of plant secondary metabolic gene clusters is likely to be regulated at the level of chromatin modification/remodeling. Changes in chromatin structure may alter the expression status of gene clusters and coordinated regulation of the secondary pathway might be facilitated by physical proximity of the cluster gene members (Frey *et al*., 2009).

The cluster for noscapine biosynthesis in poppy is coordinately regulated with regard to the homozygous/heterozygous state of the plant genotype, indicating an additional level of regulation beyond the transcriptional level (Winzer *et al*., 2012). An F_2 population of 271 individuals was generated, using two poppy varieties as parents, the high noscapine 1 variety and the high morphine 1. Analysis of the metabolite levels of noscapine in field-grown F_2 capsules revealed very interesting data. The 150 heterozygous lines exhibited much lower levels of noscapine than expected, indicating some type of a "global" repression of the noscapine cluster in the heterozygous genotype (Winzer *et al*., 2012).

In oat gene cluster, experimental findings provide cytological evidence to link cell type-specific chromatin decondensation with expression of the avenacin cluster in nuclei of root epidermal cells, providing new insights into regulation of gene clusters in plants (Wegel *et al*., 2009). It has been suggested that opening up domains of secondary chromatin structure may place genes in a transcriptionally permissive environment, where their expression can then be triggered by transcription factors (Sproul *et al*., 2005). The gene partners of thalianol and marneral clusters of *A. thaliana* are strongly associated with repressive histone H3 lysine trimethylation (H3K27me3), but the immediate flanking genes are not (Zhang *et al*., 2007; Field and Osbourn, 2008; Field *et al*., 2011). Two chromatin-remodeling proteins, pickle (PKL) and pickle-related (PKR2), act as transcriptional activators of H3K27me3-marked genes in roots (Aichinger *et al*., 2009). PKL binds directly to at least one gene in the thalianol gene cluster (Zhang *et al*., 2007). Furthermore, the expression of the three genes composing the thalianol cluster seems to be PKL/PKR2 dependent, while the expression of the marneral cluster genes appears to be partially PKL/PKR2 dependent (Zhang *et al*., 2007).

D.2.4.3. A transcriptional gene silencing mechanism in *AMY2* gene cluster regulation

Systematic analysis of the types of chromatin modifications that are important for the expression of plant metabolic gene clusters has not yet been reported. In the present work such a modification is reported, namely DNA methylation, which appears to have a role in the *AMY2* gene cluster regulation. A "global-silencing" phenomenon was observed, in which hairpin-mediated gene silencing of any of the genes in the cluster induced DNA methylation and, thus, repression of gene expression in the adjacent genes within the *AMY2* cluster.

D.2.4.3.1. The RNA silencing in plants

RNA silencing is a nucleotide sequence-specific process of RNA degradation in animals (RNA interference-RNAi), in fungi (quelling), in higher plants (post-transcriptional gene silencing), as well as in unicellular eukaryotic algae. In higher plants, a natural role of RNA silencing is to protect against virus infections (Covey *et al*., 1997; Hamilton and Baulcombe, 1999; Vaistij *et al*., 2002). The silencing process invariably involves three core reactions, though there is a considerable variation and redundancy in the mechanisms and factors involved in each of these reactions (Brodersen and Voinnet, 2006). The first reaction involves production of various forms of double-stranded (ds) RNA, regarded as a ubiquitous trigger of RNA silencing. The dsRNA is then processed into sRNAs, a reaction catalyzed by paralogs of the RNase-III enzyme Dicer. In the third step, sRNAs are bound to an ARGONAUTE (AGO) protein in RNA-induced silencing complexes (RISCs), and act as sequence-specificity determinants that guide RISCs to partly or fully complementary RNA or DNA, resulting in inhibition of gene expression (Baulcombe, 2004).

D.2.4.3.2. The RNA-directed DNA methylation (RdDM)

RNA-directed DNA methylation (RdDM) is a RNA-mediated epigenetic modification frequently associated with transcriptional repression and gene silencing (Chan *et al*., 2005; Goll and Bestor, 2005), which is highly developed in flowering plants (Bei *et al*., 2007). The hallmarks of RdDM include methylation of cytosines in all sequence contexts (CG, CNG, CNN, where N is A, T or C) and restriction of methylation to the region of RNA-DNA sequence homology. The establishment and maintenance of RdDM postulate conventional DNA cytosine methyltransferases, histone-modifying enzymes and nuclear-localized RNAi proteins (Chan, 2008). Furthermore, various plant-specific proteins are required, especially subunits of two novel RNA polymerases termed Pol IV and Pol V (Pikaard *et al*., 2008; Wierzbichi *et al*., 2008). Spreading of silent chromatin across a chromosome is a feature of many epigenetic processes in eukaryotes but the mechanisms, including the potential poles of non-coding RNAs, are still under investigation (Talbert and Henikoff, 2006; Kwon and Workman, 2008). In fission yeast, an RNAi-mediated pathway challenges the spreading of histone H3 lysine 9 methylation in heterochromatic DNA repeats (Iida *et al*., 2008; Zhang *et al*., 2008).

Formation of approximately 21-22-nt secondary siRNAs and spreading of RdDM within transcribed regions have been observed during post-transcriptional gene silencing (PTGS) in plants and shown to require RDR6 and transcription of the target gene (Vaistij *et al*., 2002; Eamens *et al*., 2008a). Both bidirectional and unidirectional spreading has been described (Vaistij *et al*., 2002; Daxinger *et al*., 2009). Spreading of methylation in transgenic plants does not always occur and there are still discrepancies in the genetic factors that may govern the methylation spreading (Henderson & Jacobsen, 2008; Daxinger *et al*., 2009). In *A. thaliana drm1 drm2 cmt3* triple mutant (DRMs are the Domains Rearranged Methyltransferase proteins and CMT3 is the chromomethylase3 protein) an F-box gene was identified termed

SUPPRESSOR of *drm1 drm2 cmt3* (*SDC*). The *SDC* promoter contains seven direct repeats that generate siRNA and recruit DNA methylation. Bidirectional spreading of DNA methylation and siRNAs from tandem repeats occurred at the *SDC* locus, but the two phenomena occurred independently and required the CG methyltransferase MET1 and Pol IV-RDR2-DCL3, respectively (Henderson & Jacobsen, 2008). According to Daxinger *et al*., 2009, hairpin-derived primary siRNAs induce primary RdDM at the target enhancer region. This step initiates the Pol IV-RDR2-dependent turnover of a nascent RNA to produce secondary siRNAs, which trigger secondary RdDM in the downstream region. In the stepwise model of RdDM, the formation of primary and secondary siRNA and the accomplishment of the corresponding RdDM steps occur sequentially and can be uncoupled genetically. It has been suggested that locus-specific effects, still unidentified, may account for differing results with regard to the mechanism that underlies RdDM, which is initiated using hairpin constructs (Saze and Kakutani, 2007; Wierzbicki *et al*., 2008; Eamens *et al*., 2008; Daxinger *et al*., 2009). Thus, the proximity of the genes in the cluster and perhaps the chromatin condensation state could be responsible for the effectiveness of transcriptional co-silencing observed in the *AMY2* gene cluster. The latter is in accordance with the association of the expression of the avenacin cluster in oats with chromatin decondensation (Wegel *et al*., 2009). Moreover, *LjSDRt* is closer to the RNAi targeted *AMY2* gene than *LjCYP71D353* and the degree of methylation in these two genes is inversely related to the distance between the genes. Alternatively, there may be other *cis* elements present in the cluster genes that render the cluster more susceptible to methylation. Further investigation of the mechanisms governing co-ordinate gene regulation in plant genomes is likely to shed more light on this phenomenon. The reason why this type of co-ordinated down-regulation of the *AMY2* cluster genes occurs in genetically modified plant lines is not clear. It appears that it is important for the plant to maintain an intact cluster

expression pattern rather than to accumulate pathway intermediates, which can result in deleterious effects on plant growth as has been shown for the avenacin late pathway intermediates (Mylona *et al*., 2008) and for the DIMBOA and cyanogenic glycosides (von Rad *et al*., 2001; Krinstensen *et al*., 2005). It should be noted, though, that methylation conditions involved transgenic RNAi-mediating silencing procedures, which may not directly represent a natural regulatory event.

D.3. Role of the β-amyrin synthase gene βAS in M. truncatula

D.3.1. *βAS* expression in response to hormone treatment

The effect of three different types of hormones, 2,4-dichlorophenoxyacetic acid (2,4-D), benzylaminopurine (BA) and methyl jasmonic acid (MeJA) on *bAS* expression was investigated. Transcript levels of *bAS* gene were induced only by MeJA as expected (Suzuki *et al*., 2002), but not by 2,4-D or BA. MeJA is a well-known elicitor for the production of several plant natural products typically manifested when plants are under environmental stresses (Zhao *et al*., 2005). Specifically regarding the induction of triterpene biosynthesis in legumes, MeJA treatment up-regulates the soyasaponin production in cultured cells of *M. truncatula* (Suzuki *et al*., 2002) and *Glycyrrhiza glabra* (Hayashi *et al*., 2003). It is suggested that jasmonates are a signal molecule to induce soyasaponin biosynthesis in legumes (Hayashi *et al*., 2004). Furthermore, the expression *L. japonicus AMY2* gene is induced by MeJA treatment.

Auxin and cytokinin have been reported to alter alkaloid content when applied exogenously to *Cantharanthus roseus* cell cultures (Whitmer *et al*., 1998; Yahia *et al*., 1998). In callus cultures *Vaccinium corymbosum* var. *bluecrop* every auxin utilized such as 2,4-D, NAA, indole-3-acetic acid (IAA) and naphthoxyacetic acid (NOA), resulted in the production of a triterpene-rich

callus culture and β-sitosterol accumulation (Migas *et al*., 2006). Generally, auxin and cytokinin have not been reported to have functions in inducing triterpene production.

D.3.2. A role for β-amyrin in symbiosis

The role of β-amyrin in nodulation remains unclear, even if there are indications to support it. β-amyrin was detected in peribacteroid membranes and the microsomal fraction of nodule cells in *P. sativum*, which mostly corresponded to plant cell endoplasmic reticulum cells, and was not detected in free living bacteria, suggesting a possible role of β-amyrin in nodules (Hernandez and Cooke, 1996). Gene expression of β-amyrin synthase gene is induced during symbiotic relations in several plants (Iturbe-Ormaetxe *et al*., 2003; Hayashi *et al*., 2004; this study). On the other hand, lupeol has a role in suppression of nodule formation in *L. japonicus*. Thus, these simple triterpenes have different and opposing effects on nodule development.

A hairpin RNA gene silencing strategy was carried out to investigate the role of β-amyrin in nodulation. A silencing construct was made for the *βAS* gene and transgenic roots were generated using an *A. rhizogenes* transformation protocol. Thus, plants expressing these hairpin constructs in their roots were obtained. Significantly decreased levels of *βAS* transcript were detected in almost all of these transgenic plants. Interestingly, the *βAS1* silenced lines showed an increased nodule number when compared to control lines. No other obvious phenotypic differences were observed at morphological level.

Heterologous expression of a β-amyrin synthase from aster in *M. truncatula* had an effect on the symbiotic nodulation performance (Confalonieri *et al*., 2009). Two transgenic lines showed a significant enhancement of nodulation, which was accompanied by significant increase of the saponin content in roots and nodules. The high nodulation phenotype was

associated with the increase of soyasaponin B in nodules, proposing a positive involvement of β-amyrin and derived saponins in the regulation of root nodulation (Confalonieri *et al*., 2009). In this study, silencing of *βAS* gene also resulted in enhanced nodulation performance in *M. truncatula*, proposing a negative involvement of β-amyrin in the regulation of nodule development. Conclusively, β-amyrin may have different, even opposing effects on nodule formation. The enhanced nodulation in the *βAS*-silenced plant lines requires further investigations.

D.4. The role of lupeol synthase (LuS) in L. japonicus nodule formation

The LuS gene *OSC3* is expressed specifically in the roots of *L. japonicus*. The expression of this gene increases markedly in response to inoculation with the compatible symbiotic partner *M. loti*, and developing nodules exhibited the highest *OSC3* expression levels. The tissue specificity of *OSC3* expression was analyzed by *in situ* hybridization. Interestingly, *OSC3* and *AMY2* transcripts are localized in the same tissues of developing and mature nodules. The localization of *OSC3* activity in uninfected cells and in the nodule vascular bundle could indicate that the synthesis of lupeol is associated with structural needs in developing nodule symbioses (Baisted, 1971; Hernandez and Cooke, 1996; Grandmougin-Ferjani *et al*., 1999). Alternatively, the presence of *OSC3* transcripts specifically and at high levels in the phloem cells of nodule vascular bundles could indicate a role in the transportation of compounds in and out of the nodule. In *Vicia faba*, the chemical composition the nodule outer cortex (oC) cell walls was investigated. Around 90% of this material consisted of betulin, which is a lupeol derivative (Hartmann *et al*., 2002). Betulin and triterpenoids are often found at the interface of aerial (cuticles) and subsoil plant parts (roots), where they increase the transpiration barrier and

additionally act as antipathogenic agents (Hayek *et al*., 1989). This raises the idea that betulin behaves as a hydrophobic and antibiotic layer covering the cells of the oC, thus making the nodule surface water-repellent and resistant to microbial infection. This is very important since the root nodule, containing high amounts of sugars and amino acids can attract several microbes (Hartmann *et al*., 2002). The localization of *OSC3* activity and thus the production of lupeol in nodular tissues may be responsible for the synthesis of lupeol derivatives, such as betulin. Betulin is then transported to outer cortex, serving as a constitutive pathogen defense barrier and possibly reducing water loss.

Based on the *OSC3* gene expression analysis, it was postulated that *OSC3* may play a role in nodule development and function. In order to verify this assumption, RNAi silencing of the *OSC3* gene using *A. rhizogenes* transformation was carried out to investigate the effect of gene silencing in nodule formation and function. A total of 18 independent silenced plants that had undetectable levels of *OSC3* expression were obtained. Metabolite analysis verified that the silencing of *OSC3* led to a loss of ability to synthesize lupeol, as shown by the detectable presence of lupeol in control lines, but its absence in the roots of *OSC3*-silenced lines. These experiments confirm that the synthesis of lupeol in *L. japonicus* roots and nodules can be solely attributed to OSC3.

The initial hypothesis was that plants lacking lupeol might exhibit defects in nodule formation and / or function. In order to address this, the nodule number and levels of expression of two early nodulin genes, *ENOD2* and *ENOD40*, were measured quantitatively in *OSC3*-silenced lines and control lines. Surprisingly, the roots of *OSC3*-silenced plants showed more nodules than those of control plants 20 dpi with *M. loti*. No other obvious phenotypic differences were observed at the anatomical or morphological levels. Furthermore, a dramatic increase in the expression of *ENOD40* was detected

in plants silenced for *OSC3*. It has been reported that the overexpression of *ENOD40* leads to accelerated nodulation in *M. truncatula* as a consequence of extensive cortical cell division and increased initiation of primordia. In these *ENOD40*-overexpressing plants, the final total numbers of nodules were controlled by the auto-regulation or feedback regulation of the nodulation mechanism (Hayashi *et al*., 2000; Wopereis *et al*., 2000; Nishimura *et al*., 2002), and, as a consequence, the final total numbers were similar to those observed in control lines. In later stages, mature and fully functional nodules were formed (Charon *et al*., 1999). According to the experimental results, the effect of silencing of the LuS gene results in phenotypes that are similar to those observed for plants overexpressing *ENOD40*. Thus, in the absence of lupeol in *OSC3*-silenced plant lines, the elevated expression of *ENOD40* led to a more rapid nodulation phenotype. Moreover, nodule number, nodule appearance and nitrogen-fixing capacity were similar in both silenced and control plants at 40 dpi, in agreement with the function of the mechanism of the auto-regulation of nodulation, as observed in *ENOD40*-overexpressing *M. truncatula* lines (Charon *et al*., 1999; Hayashi *et al*., 2000). From these results, it is apparent that lupeol down-regulates the expression of *ENOD40*, which has been detected in proliferating symbiotic and non-symbiotic tissues in different legumes (Papadopoulou *et al*., 1996; Corich *et al*., 1998; Flemetakis *et al*., 2000). The latter include lateral root and stipule primordia and also embryonic tissues. The role of ENOD40 in these processes and in non legumes remains unclear and, to date, only subtle phenotypes have been reported in non legume plants overexpressing this protein (Wan *et al*., 2007).

The reduction in the expression of *ENOD2* in the silenced lines could be the result of the perturbation of the normal nodulation process caused by the aberrant expression of *ENOD40*. As the nodule phenotype and function were unaltered in mature nodules of silenced lines, the effect on *ENOD2* expression was apparently only transient. This is further supported by the exogenous

application of lupeol on wild-type plants, which had no effect on *ENOD2* expression levels.

The effect observed for the role of lupeol in nodule initiation appears to be specific for lupeol and not for simple triterpenes in general, as exogenous application of lupeol, but not of β-amyrin, to wild-type plants down-regulates *ENOD40* expression. More elaborated triterpenes such as betulin and betulinic acid, derivatives of lupeol, have not been included in this study but they could represent good candidates for future studies. The effect of lupeol differs from that ascribed to soyasapogenol B, a derivative of β-amyrin, as a consequence of the heterologous expression of an aster *β-AS* in *M. truncatula* (Confalonieri *et al*., 2009). It is not known whether lupeol is elaborated further in *L. japonicus* by modifying enzymes, such as sugar transferases, as the triterpenoid and saponin contents of this species have not yet been studied extensively. In addition, *OSC3* transcripts were also detected in mature nodules, but no detectable phenotype was observed in silenced plant lines at this stage of development.

D.5. Perspectives and applications

The biosynthesis of terpenoids is tightly controlled in plants, as they are involved in various functions, like plant growth, development and response to biotic and abiotic environmental factors (Nagegowda, 2010; Vranova *et al*., 2012). Because of their strict regulation, most terpenoids are produced in very small amounts in their natural sources, resulting in a wide gap between demand and supply of terpenoids, which hampers their widespread application. As terpenoid biosynthesis is strictly regulated and often controlled by specific transcription factors, one way to increase productivity is to modulate the expression of such or other regulatory factors (Broun, 2004; De Geyter *et al*., 2012). Nevertheless, the overexpression of a single transcription factor does not necessarily lead to a higher production of the compounds. To

date, *in planta* triterpenoid engineering has been hampered by the lack of knowledge about the regulatory mechanisms controlling gene expression (Sawai and Saito, 2011). Hence, a challenge for future triterpenoid research will be to identify the transcription or other regulatory factors that direct their biosynthesis. Additionally, while multiple OSCs catalyzing the cyclization of 2,3-oxidosqualene to different triterpenoid precursor backbones have been isolated already, only a few of genes corresponding to the decorating / tailoring enzymes have been identified, whereas many more must exist when considering the structural diversity of triterpenoids in the plant kingdom. Similarly, although the biosynthetic enzymes are mostly microsomal in nature, triterpenoids typically localize to the epidermal wax layer or the vacuoles, suggesting the existence of yet undiscovered transporter systems. Hence, unraveling the molecular mechanisms involved in triterpenoid biosynthesis *in planta* will facilitate their exogenous engineering (Mosses *et al*., 2013).

In this present work, more "plant" knowledge on triterpene biosynthesis, regulation and function was revealed. Genes were characterized for their involvement in triterpenoid biosynthesis. β-amyrin synthase produces a novel triterpene structure, dihydrolupeol, and LjCYP71D353 produces 20-hydroxybetulinic acid in a sequential three-step oxidation of 20-hydroxylupeol. The enzymes identified in this work could serve the goals of combinatorial biosynthesis for the *de novo* pathway construction in homologous or heterologous systems and the production of novel chemical compounds that possess improved biological activities (Pollier *et al*., 2011). The activities of the novel triterpene structures identified have to be elucidated and explored for their pharmaceutical potential since pentacyclic triterpenoids are the most potent natural product compounds showing wide ranging anti-inflammatory and anticancer activities (Shanmugam *et al*., 2012).

Naturally occurring pentacyclic triterpenoids include ursolic acid, oleanolic acid, betulinic acid, bosewellic acid, asiatic acid, α-amyrin, celastrol,

glycyrrhizin, 18-b-glycyrrhetinic acid, lupeol, escin, madecassic acid, momordin I, platycodon D, pristimerin, saikosaponins, soyasapogenol B, and avicin (Connolly and Hill, 2012). Various plant-based triterpenoids are very promising as chemopreventive and therapeutic agents, because they are able to inhibit key signaling molecules, inflammatory mediators, tumor cell proliferation, invasion, metastasis, and angiogenesis in various in vitro and in vivo models of cancer. *In vitro* antitumor activity of triterpenoids is summarized in table 2.

Pentacyclic triterpenoids	*In vivo* cancer models	Signal transduction pathways	References
Acetyl-11-keto-beta-boswellic acid (AKBA)	Suppresses the growth of subcutaneously implanted human glioma; colon tumor; leukemia; hepatocellular carcinoma; prostate tumor in nude mice. Suppresses the growth of orthotopically implanted colorectal cancer cells in nude mice.	NF-κB; STAT3	[67,68,76]
Avicin	Suppresses the growth of UVB induced skin cancer in UVB skin cancer model	NF-κB; STAT3; VEGF; NRF2-Keap1	[33,39]
Betulinic acid	Suppresses the growth of subcutaneously implanted human melanoma xenografts in nude mice; UVB induced photocarcinogenesis in nude mice.	NF-κB; ROS	[43,61,62]
Celastrol CA19	Suppresses the growth of subcutaneously implanted human melanoma tumors and metastasis in syngenic mice model and the growth of human PC3 prostate tumor; MDA-MB-231 breast tumor; and glioma in nude mice.	NF-κB; STAT3; ATF2 AKT/mTOR/p70S6K; DR4/5	[87,90,91,96,102–104]
Lupeol	Suppresses the growth of skin cancer in CD1 mice model; Suppresses the growth of orthotopically implanted head and neck cancer; Suppresses the growth of human melanoma tumor; pancreatic; hepatocellular carcinoma in nude mice.	NF-κB; PI3K/Akt; PTEN/Akt/ABCG2	[134,135,140,144,146]
Oleanolic acid Amooranin (AMR)	Suppresses the growth of subcutaneously implanted colon cancer in F344 rats and skin carcinogenesis	NF-κB; JAK/STAT3	[147,151,153]
Synthetic triterpenoids CDDO-me AMR-me	Suppresses the growth of subcutaneously implanted transgenic pancreatic tumor; and orthotopically implanted pancreatic tumor; breast cancer in BRCA1-mutated mice; HT-29 and MC38 human colon tumor; EL-4 thymoma in nude mice model. Lung cancer induced by vinyl carbamate in A/j mice model; prostate cancer in TRAMP mice model and lewis lung carcinoma in SCID-beige mice model	NF-κB; JAK/STAT3; AKT; NRF2-Keap1; DR4/5	[11,153,209–212,214,218]
Ursolic acid	Suppresses the growth of subcutaneously implanted human prostate cancer tumor in nude mice; and spontaneously developing prostate tumor in TRAMP mice model	NF-κB; JAK/STAT3; AKT	[185,186]

Table 2. In vivo effects on natural and semi-synthetic pentacyclic triterpenoids in various animal models of cancer (Shanmugam *et al.*, 2012 and references therein).

Lupeol has been identified in screens aimed at isolating phytochemicals as novel therapeutic and chemopreventive agents, and has been shown to activate the apoptotic machinery (Saleem *et al.*, 2004, 2005a, b, 2008; Lee *et al.*, 2007). Recently, it has been shown that, in human cancer cells, lupeol has a multitarget effect on different components of the β-catenin signaling pathway. It decreases the β-catenin protein level and also induces β-catenin degradation (Behrens, 2000; Conacci-Sorrell *et al.*, 2002; Saleem *et al.*, 2009). β-Catenin is required for cell proliferation, cell–cell adhesion and for the regulation of gene expression during development. The β-catenin signaling

pathway is well characterized in mammals, *Drosophila* and *Dictystelium discoideum* (Verras and Sun, 2006). Genes with homology to β-catenin have been identified in plants and, as is the case for mammalian β-catenin, their localization to the cytosol or the nucleus may depend on the availability of interacting proteins (Coates, 2003). The mode of action of these proteins in plants has not been characterized fully as yet, but proteins associated with β-catenin signaling have been described in the gibberellin and brassinosteroid signaling pathways (Amador *et al*., 2001; He *et al*., 2002; Lee *et al*., 2002), and, in *Arabidopsis*, β-catenin-related proteins (namely ARABIDILLO-1 and ARABIDILLO-2) have been shown to define a pathway that promotes lateral root development (Coates *et al*., 2006). In this study, lupeol was implicated to play a negative role in nodulation via a mechanism that involves ENOD40. Putative interaction between β-catenin-related proteins and ENOD40 is difficult to envisage. A function for b-catenin-related proteins in nodulation has also not yet been reported. Nevertheless, it is intriguing to consider that the biosynthesis of lupeol and lupeol per se may influence nodule formation and possibly also lateral root development by interfering with β-catenin-mediated signaling. Future work should include the identification of β-catenin-related proteins in *L. japonicus* and the investigation of their putative interaction with ENOD40 and lupeol in order to identify if lupeol has a signaling role in plant developmental processes. Several investigations suggest the existence of sterol signal molecules that are capable of controlling developmental programs (Li *et al*., 1996; Szekers *et al*., 1996; Schrick *et al*., 2000; Jang *et al*., 2000). Similarly, the terpenoid strigolactones were shown to have a role in various developmental processes, including nodulation (Foo and Davies, 2011; Foo *et al*., 2013; Liu *et al*., 2013). It is possible that the structurally similar triterpenes act as signaling molecules controlling developmental processes.

In conclusion, this study show that the unusual feature of metabolic gene clustering in plant genomes exists in legumes as well, further strengthening the hypothesis that specific evolution processes take place for the production of physiologically important plant special metabolites such as the triterpenes. Furhtermore, this work adds evidence to as yet largerly unrecognized functions of triterpenes in plant growth and development. Finally, it contributes with genes and enzymes for the production of novel compounds with improved or altered bioactivity via combinatorial biosynthesis or via synthetic biology approaches.

E. REFERENCES

CHAPTER E. REFERENCES

Abe I., Ebizuka Y. and Sankawa U. (1988) Purification of 2,3-oxidosqualene cycloartenol cyclase from pea-seedlings. Chemical & Pharmaceutical Bulletin 36, 5031-5034.

Abe I., Rohmer M. and Prestwich G.D. (1993) Enzymatic Cyclization of Squalene and Oxidosqualene to Sterols and Triterpenes. Chem. Rev., 93, 2189-2206.

Abe I. and Prestwich G.D. (1994) Active site mapping of affinity-labeled oxidosqualene cyclases. J. Biol. Chem. 269, 802-804.

Abe I. (2007) Enzymatic synthesis of cyclic triterpenes. Nat. Prod. Rep. 24, 1311-1331.

Achnine L., Huhman D.V., Farag M.A., Sumner L.W., Blount J.W. and Dixon R.A. (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. The Plant Journal 41, 875–887.

Achnine L, Blancaflor E.B., Rasmussen S, Dixon R.A. (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydrozylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell 16, 3098-3109.

Ahmad S., Veyrat N., Gordon-Weeks R., Zhang Y., Martin J., Smart L., Glauser G., Erb M., Flors V., Frey M. and Ton J. (2011) Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. Plant Physiol. 157, 317-27.

Aichinger E., Villar C.B., Farrona S., Reyes J.C., Hennig L. and Köhler C. (2009) CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. PLoS Genet. 5:e1000605.

Akiyama K., Matsuzaki K. and Hayashi H. (2005) Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435, 824–827.

Alamillo, J.M., **Monger, W.**, **Sola, I.**, **García, B.**, **Perrin, Y.**, **Bestagno, M.**, **Burrone, O.R.**, **Plana-Duran, J.**, **Enjuanes, L.**, **Lomonossoff, G.P.** *and* **García, J.A.** *(*2006*)* Use of virus vectors for the expression in plants of active full-length and single chain anti-coronavirus antibodies. Biotechnol. J. 1, 1103–1111.

Alkhimova O.G., Mazurok N.A., Potapova T.A., Zakian S.M., Heslop-Harrison J.S. and Vershinin A.V. (2004) Diverse patterns of the tandem repeats organization in rye chromosomes. Chromosoma 113, 42–52.

Allocco D.J., Kohane I.S. and Butte A.J. (2004) Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 5, 18.

Altschul S.F., Boguski M.S., Gish W. and Wootton J.C. (1994) Issues in searching molecular sequence databases. Nature Genetics 6, 119–129.

Amador V, Monte E, Garcia-Martinez JL and Prat S. (2001) Gibberellins signal nuclear import of PHOR1, a photoperiod-responsive protein with homology to *Drosophila* armadillo. Cell 106: 343–354.

Ansari MT, Saify ZS, Sultana N, Ahmad I, Saeed-Ul-Hassan S, Tariq I, Khanum M. (2013) Malaria and artemisinin derivatives: an updated review. Mini Rev. Med. Chem. Article In press.

Anzai Y., Li S.Y., Chaulagain M.R., Kinoshita K., Kato F., Montgomery J. and Sherman D.H. (2008) Functional analysis of MycCI and MycG, cytochrome P450 enzymes involved in biosynthesis of mycinamicin macrolide antibiotics. Chemistry and Biology 15, 950–959.

Aoki T., Kamizawa A. and Ayabe S. (2002) Efficient Agrobacterium-mediated transformation of *Lotus japonicus* with reliable antibiotic selection. Plant Cell Rep. 21, 238-243.

Arambarri A. (1999) Illustrated catalogue of *Lotus* L. Seeds (Fabaceae) In: *Trefoil*: The science and technology of *Lotus*. American Society of Agronomy pp 21-41.

Argentieri M.P., D'Addabbo T., Tava A., Agostinelli A., Jurzysta M. and Avato P. (2008) Evaluation of nematicidal properties of saponins from *Medicago* spp. Eur. J. Plant Pathol. 120, 189-197.

Ariel B., Vanesa M., Hopp E.H., Mariana del Vas and Asurmendi S. (2007) A practical approach to the understanding and teaching of RNA silencing in plants. Electronic Journal of Biotechnology. Vol.10 No.2.

Arvy M.P., Imbault N., Naudascher F., Thiersault M. and Doireau P. (1994) 2,4-D and alkaloid accumulation in periwinkle cell suspensions. Biochimie. 76, 410-416.

Asamizu E., Nakamura Y., Sato S. and Tabata S. (2004) Characteristics of the *Lotus japonicus* gene repertoire deduced from large-scale expressed sequence tags (EST) analysis. Plant Mol. Biol. 54, 405-414.

Atoui A., Bao D., Kaur N., Grayburn W.S. and Calvo A.M. (2008) *Aspergillus nidulans* natural product biosynthesis is regulated by *mpk*B, a putative pheromone response mitogen-activated protein kinase. Appl. Environ. Microb. 74, 3596-3600.

Augustin J.M., Kuzina V., Andersen S.B. and Bak S. (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry 72, 435–457.

Augustin J.M., Drok S., Shinoda T., Sanmiya K., Nielsen J.K., Khakimov B., Olsen C.E., Hansen E.H., Kuzina V., Ekstrøm C.T., Hauser T. and Bak S. (2012) UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-*O*-glucosylation in saponin-mediated insect resistance. Plant Physiol. 160, 1881-1895.

Baba M., De Cleroq S., Nakashima H. and Yamamoto N. (1988) Mechanism of inhibitory effect of glycyrrhizin on replication of human immunodeficiency virus (HIV). Antiviral Res. 10, 289-298.

Babiychuk E., Bouvier-Nave P., Compagnon V., Suzuki M., Muranaka T., Van Montagu M., Kushnir S. and Schaller H. (2008) Allelic mutant series reveal distinct functions for *Arabidopsis* cycloartenol synthase 1 in cell viability and plastid biogenesis. PNAS 105, 3163–3168.

Baisted D. J. (1971) Sterol and Triterpene Synthesis in the Developing and Germinating Pea Seed. *Biochem. J.* 124, 375-383.

Bak S., Kahn R.A., Olsen C.E. and Halkier B.A. (1997) Cloning and expression in *Escherichia coli* of the obtusifoliol 14 alpha-demethylase of *Sorghum bicolor* (L.) Moench, a cytochrome P450 orthologous to the sterol 14 alpha-demethylases (CYP51) from fungi and mammals. Plant J. 11, 191-201.

Bak S., Paquette S.M., Morant M., Morant A., Saito S., Bjarnholt N., Zagrobelny M., Jorgensen K., Osmani S., Simonsen H., Perez R., Heeswijck T., Jorgensen B. and Moller B. (2006) Cyanogenic glucosides: a case study for evolution and application of cytochromes P450. Phytochem. Rev. 5, 309-329.

Bak S., Beisson F., Bishop G., Hamberger B., Hofer R., Paquette S. and Werck-Reichhart D. (2011) Cytochromes P450. The Arabidopsis book e0144 doi: 10.1099/tab.0144.

Balis C., Chatzipavlidis J. and Flouri F. (1996) Olive mill waste as a substrate for nitrogen fixation. International Biodeterioration & Biodegradation 38, 169-178.

Barbulova A. and Chiurazzi M. (2005) A procedure for in vitro nodulation studies. In: Marquez A.J. The *Lotus japonicus* handbook. Dordrecht, the Netherlands: Springer, 83–86.

Barker D.G., Bianchi S., Blondon F., Dattée Y., Duc G., Essad S., Flament P., Gallusci P., Génier G., Guy P., Muel X., Tourneur J., Dénarié J. and Huguet T. (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the rhizobium-legume symbiosis. Plant Mol. Biol. Rep. 8, 40-49.

Batz O., Logemann E., Reinold S. and Hahlbrock K. (1998) Extensive reprogramming of primary and secondary metabolism by fungal elicitor or infection in parsley cells. Biol. Chem. 379, 1127-1135.

Baulcombe D. (2004) RNA silencing in plants. Nature 431, 356-363.

Behrens J. (2000) Control of b-catenin signaling in tumor development. Annals of the New York Academy of Sciences 910: 21–33.

Bei Y., Pressman S. and Carthew R. (2007) SnapShot: small RNA-mediated epigenetic modifications. Cell 130, 756.

Bennetzen J.L., Ma J. and Devos K.M. (2005) Mechanisms of recent genome size variation in flowering plants. Ann. Bot. (Lond) 95, 127-132.

Bentley S.D. *et al.*, (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417, 141-147.

Benveniste P. (2004) Biosynthesis and accumulation of sterols. Annu. Rev. Plant Biol. 55, 429-457.

Bernatavichute, Y., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PLoS ONE. 3, e3156.

Bernhardt, R. (2006) Cytochromes P450 as versatile biocatalysts. J. Biotechnol. 124: 128-45.

Bewley J.D. (1997) Seed germination and dormancy. Plant Cell 9, 1055–1066.

Bhuvaneswari T.V., Bhagwat A.A. and Bauer W.D. (1981) Transient susceptibility of root cells in four common legumes to nodulation by rhizobia. Plant Physiology 68, 1144–1149.

Boisson-Dernier A., Chabaud M., Garcia F., Bécard G., Rosenberg C. and Barker D.G. (2001) *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. Mol. Plant Microbe Interact. 14, 695-700.

Bok J.W, Chiang Y.M., Szewczyk E., Reyes-Domingez Y., Davidson A.D., Sanchez J.F., Lo H.C., Watanabe K., Strauss J., Oakley B.R., Wang C.C. and Keller N.P. (2009) Chroatin-level regulation of biosynthetic gene clusters. Nat. Chem. Biol. 5, 462-464.

Boller T. (1995) Chemoperception of microbial signals in plant cells. Annu.Rev. Plant Physiol. Plant Mol. Biol. 46, 189-214.

Bollenbach, T.J., and Stern, D.B. (2003) Divalent metal-dependent catalysis and cleavage specificity of CSP41, a chloroplast endoribonuclease belonging to the short chain dehydrogenase/reductase superfamily. Nucleic Acids Res. 31, 4317–4325.

Boogerd F.C. and van Rossum D. (1997) Nodulation of groundnut by *Bradyrhizobium*: a simple infection process by crack entry. FEMS Microbiol. Rev. 21, 5–27.

Bouarab K, Melton R., Peart J., Baulcombe D. and Osbourn A. (2002) A saponin-detoxifying enzyme mediates suppression of plant defences. Nature, 418.

Bowers J.E., Chapman B.A., Rong J. and Paterson A.H. (2003) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422, 433-438.

Bowles D., Isayenkova J., Lim E. and Poppenberger B. (2005) Glycosyltransferases: managers of small molecules. Curr. Opin. Plant Biol. 8, 254–263.

Bowles D., Lim E.K., Poppenberger B. and Vaistij F.E. (2006) Glycosyltransferases of lipophilic small molecules. Annu. Rev. Plant Biol. 57, 567–597.

Bowyer P., Clarke B.R., Lunness P., Daniels M.J. and Osbourn A.E. (1995) Host-Range of a Plant-Pathogenic Fungus Determined by a Saponin Detoxifying Enzyme, Science 267, 371-374.

Brodersen P. and Voinnet O. (2006) The diversity of RNA silencing pathways in plants. TRENDS in Genetics Vol.22.

Broeckling C.D., Huhman D.V., Farag M.A., Smith J.T., May G.D., Mendes P., Dixon R.A. and Sumner L.W. (2005) Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. Journal of Experimental Botany Vol. 56, No. 410, 323–336.

Broun P. (2004). Transcription factors as tools for metabolic engineering in plants. Current Opinion in Biotechnology 7, 202–209.

Brown B.A., Cloix C., Jiang G.H., Kaiserli E., Herzyk P., Kliebenstein D.L. and Jenkins G.I. (2005) A UV-B specific signaling component orchestrates plant UV protection. PNAs 102, 18225-18230.

Budzikiewicz H., Djerassi C. and Williams D.H. (1964) Structure elucidation of natural products by mass spectrometry, v. II, Steroids, terpenoids, sugars, and miscellaneous classes. San Francisco, CA, USA: Holden-Day.

Burbulis I.E. and Winkel-Shirley B. (1999) Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. PNAs 96, 12929-12934.

Burge C.B. and Karlin S. (1998) Finding the genes in genomic DNA. Current Opinion in Structural Biology 8, 346–354.

Buschhaus C. and Jetter R. (2012) Composition and Physiological Function of the Wax Layers Coating Arabidopsis Leaves: β-amyrin Negatively Affects the Intracuticular Water Barrier. Plant Physiology Vol. 160, 1120–1129.

Calvo A.M., Wilson R.A., Bok J.W. and Keller N.P. (2002) Relationship between secondary metabolism and fungal development. Microbiol. Mol. Biol. Rev. 66, 447-459.

Campbell J.A., Davies G.J., Bulone V. and Henrissat B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. 326, 929–939.

Canizares M.C., Liu L., Perrin Y., Tsakiris E. and Lomonossoff G.P. (2006) A bipartite system for the constitutive and inducible expression of high levels of foreign proteins in plants. Plant Biotechnology Journal 4, 183–193.

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

Carlson M., Celenza J.L. and Eng F.J. (1985) Evolution of the dispersed SUC gene family of *Saccharomyces* by rearrangements of chromosome telomeres. Mol. Cell Biol. 5, 2894–2902.

Carlson J.C., Fortman J.L., Anzai Y., Li S.Y., Burr D.A. and Sherman D.H. (2010) Identification of the tirandamycin biosynthetic gene cluster from *Streptomyces* sp 307-9. ChemBioChem 11, 564–572.

Carruthers L.M. and Hansen J.C. (2000) The core histone N termini function independently of linker histones during chromatin condensation. J. Biol. Chem. 275, 37285-37290.

Carter D., Chakalova L., Osborne C.S., Dai Y.F. and Fraser P. (2002) Long-range chromatin regulatory interactions *in vivo*. Nature Genet. 32, 1-4.

Chalker-Scott L. and Fnchigami LH. (1989) The role of phenolic compounds in plant stress responses. In: Paul HL, Ed. Low temperature stress physiology in crops. Boca Raton, Florida: CRC Press Inc. 1989:40.

Chambeyron S. and Bickmore W.A. (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev. 18, 1119-1130.

Chan S.W., Henderson L.R. and Jacobsen S.E. (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. Nat. Rev. Genet. 6, 351-360.

Chan S.W.L. (2008) Inputs and outputs for chromatin-targeted RNAi. Trends Plant Sci. 13, 383-389.

Chang W.C., Lee T.Y., Huang H.D., Huang H.Y. and Pan R.L. (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC Genomics* **9**, 561-574.

Chapman S, Kavanagh T, Baulcombe D. (1992) Potato virus X as a vector for gene expression in plants. Plant J. 2(4):549-57.

Chappell J. (1995) The biochemistry and molecular biology of isoprenoid metabolism. Plant Physiology 107, 1-6.

Chappell J. (1995) Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. Annu.Rev.Plant Physiol. Plant Mol.Biol. 46, 521-547.

Chappell J. (2002) The genetics and molecular genetics of terpene and sterol origami. Curr. Opin. Plant Biol. 5, 151-157.

Chapple C.C.S., Shirley B.W., Zook M., Hammerschmidt R. and Somerville S.C. (1994) Secondary metabolism in *Arabidopsis*. Cold Spring Harbor Laboratory Press 989-1030.

Charon C., Sousa C., Crespi M. and Kondorosi A. (1999) Alteration of *enod40* expression modifies *Medicago truncatula* root nodule development induced by *Sinorhizobium meliloti.* Plant Cell 11: 1953–1966.

Chau MD, Walker K, Long R, Croteau R (2004) Regioselectivity of taxoid-*O*-acetyltransferases: heterologous expression and characterization of a new taxadien-5a-ol-*O*-acetyltransferase. Arch. Biochem. Biophys. 430, 237-246.

Chen, Y., Guo, Q., Liu, L., Liao, L. and Zaibiao, Z. (2011) Influence of fertilization and drought stress on the growth and production of secondary metabolites in *Prunella vulgaris* L. J. Med. Plants Res. 5, 1749–1755.

Chen K., and Rajewsky N. (2007) The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews Genetics 8, 93–103.

Chen A. and Dubcovsky J. (2012) Wheat TILLING mutants show that the vernalization gene VRN1 down-regulates the flowering repressor VRN2 in leaves but is not essential for flowering. PLoS Genet. 8(12):e1003134.

Cheng A., Lou Y., Mao Y., Lu S., Wang L., and Chen X. (2007) Plant terpenoids: Biosynthesis and ecological functions. J. Integrative Plant Biol. 49, 179-186.

Chen, Y.Z., Pang, Q.Y., Hea, Y., Zhua, N., Branstroma, I., Yanb, X.F. and Chen, S. (2012) Proteomics and metabolomics of Arabidopsis responses to perturbation of glucosinolate biosynthesis. Mol. Plant. 5, 1138–1150.

Cheng X., Ruyter-Spira C. and Bouwmeester H. (2013) The interaction between strigolactones and other plant hormones in the regulation of plant development. Frontier in Plant Sci. 4, 1-16.

Cheong J.J. and Choi Y.D. (2003) Methyl jasmonate as a vital substance in plants. TRENDS in Genetics Vol.19 No.7.

Chilton M.D., Drummond M.H., Merio D.J., Sciaky D., Montoya A.L., Gordon M.P. and Nester E.W. (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell 11, 263-271.

Cho E.M., Okada A., Kenmoku H., Otomo K., Toyomasu T., Mitsuhashi W., Sassa T., Yajima A., Yabuta G., Mori K., Oikawa H., Toshima H., Shibuya N., Nojiri H., Omori T., Nishiyama M. and Yamane H. (2004) Molecular cloning and characterization of a cDNA encoding ent-cassa-12,15-diene synthase, a putative diterpenoid phytoalexin biosynthetic enzyme, from suspension-cultured rice cells treated with a chitin elicitor. Plant J. 37, 1-8.

Chu H.Y., Wegel E. and Osbourn A. (2011) From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. The Plant Journal 66, 66–79.

Cichewicz R.H. (2010) Epigenome manipulation as a pathway to new natural product scaffolds and their congeners. Nat. Prod. Rep. 27, 11-22.

Coates JC. (2003) Armadillo repeat proteins: beyond the animal kingdom. Trends in Cell Biology 13: 463–471.

Coates JC, Laplaze L and Haseloff J. (2006) Armadillo-related proteins promote lateral root development in *Arabidopsis*. PNAs, USA 103: 1621–1626.

Cokus, S., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C., Pradhan, S., Nelson, S., Pellegrini, M., and Jacobsen, S. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature. 452, 215–219.

Cole B.J.W., Bentley M.D., Hua Y. and Bu L. (1991) Triterpenoid constituents in the outer bark of *Betula alleghaniensis* (Yellow Birch). Journal of Wood Chemistry and Technology 11, 209–223.

Conacci-Sorrell M, Zhurinsky J, Ben-Zeev A. (2002) The cadherin–catenin adhesion system in signaling and cancer. Journal of Clinical Investigations 109: 987–991.

Conant G.C. and Wolfe K.H. (2008) Turning a hobby into a job: how dublicated genes find new functions. Nat. Rev. Genet 9, 938-950.

Confalonieri M., Cammareri M., Biazzi E., Pecchia P., Fevereiro M.P.S., Balestrazzi A., Tava A. and Conicella C. (2009) Enhanced triterpene saponin biosynthesis and root nodulation in transgenic barrel medic (*Medicago truncatula* Gaertn.) expressing a novel β-amyrin synthase (*AsOXA1*) gene. Plant Biotechnology Journal 7, 172–182.

Corey E.J., Matsuda S.P. and Bartel B. (1993) Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. PNAs 90, 11628-11632.

Corich V, Goormachtig S, Lievens S, Van Montagu M, Holsters M. (1998) Patterns of *Enod40* gene expression in stem-borne nodules of *Sesbania rostrata*. Plant Molecular Biology 37: 67–76.

Coutinho P.M., Deleury E., Davies G.J. and Henrissat B. (2003) An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. 328, 307–317.

Covey S.N., Al-Kaff N.S., Langara A. and Turner D.S. (1997) Plants combat infection by gene silencing. Nature 385, 781-782.

Crane C., Wright E., Dixon R.A. and Zeng-Yu Wang. (2006) Transgenic *Medicago truncatula* plants obtained from *Agrobacterium tumefaciens*-transformed roots and *Agrobacterium rhizogenes*-transformed hairy roots. Planta 223: 1344–1354.

Credali A., García-Calderón M., Dam S., Perry J., Díaz-Quintana A., Parniske M., Wang T.L., Stougaard J., Vega J.M. and Márquez A.J. (2013) The K^+-dependent asparaginase, NSE1, is crucial for plant growth and seed production in *Lotus japonicus*. Plant Cell Physiol. 54, 107-118.

Crombie W.M.L. and Crombie L. (1986) Distribution of Avenacins A-1, A-2, B-1 and B-2 in Oat Roots - Their Fungicidal Activity Towards Take-All Fungus, Phytochemistry 25, 2069-2073.

Croteau R., Kutchan T.M. and Lewis N.G. (2000) Natural products (secondary metabolites). In Buchanan B., Gruissem W., Jones R. (Eds.). Biochemistry and molecular biology of the plants. Rockville M.D. : American Society of Plant Physiologists, pp. 1250-1318.

D'Addabbo T., Avato P. and Tava A. (2009) Nematicidal potential of materials from *Medicago* spp. Eur. J. Plant Pathol. 125, 39-49.

Dalakouras A., Tzanopoulou M., Tsagris M., Wassenegger M. and Kalantidis K. (2011) Hairpin transcription does not necessarily lead to efficient triggering of the RNAi pathway. Transgenic Res 20, 293–304.

Dalakouras A. and Wassenegger M. (2013) Revisiting RNA-directed DNA methylation. RNA Biology 10, 453–455.

Dang T. and Prestwich G.D. (2000) Site-directed mutagenesis of squalene-hopene cyclase: altered substrate specificity and product distribution. Chem. Biol. 7, 643-649.

Daniell, H., **Khan, M.S.** *and* **Allison, L.** *(*2002*)* Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. Trends Plant Sci. 7, 84–91.

Dantanarayana A.P., Kumar N.S., Sultanbawa M.U.S. and Balasubramaniam S. (1981) Structures of four new oxygenated lupanes from Pleurostyliaopposita (Celastraceae). Journal of the Chemical Society, Perkin Transactions 10, 2717– 2723.

D'Auria J.C. and Gershenzon J. (2005) The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. Current Opinion in Plant Biology 8, 308–316.

Daxinger L., Kanno T., Bucher E., van der Winden J., Naumann U., Matzke A.J. and Matzke M. (2009) A stepwise pathway for biogenesis of 24-nt secondary si-RNAs and spreading of DNA methylation. EMBO J. 28, 48-57.

de Abreu I.N. and Mazzafera P. (2005) Effect of water and temperature stress on the content of active constituents of *Hypericum brasiliense* Choisy. Plant Physiol. Biochem. 43, 241–248.

De Geyter N, Gholami A, Goormachtig S, Goossens A. (2012) Transcriptional machineries in jasmonate-elicited plant secondary metabolism. Trends in Plant Science 17: 349–359.

De Vetten N., Ter Horst J., Van Schaik H.P., De Boer A., Mol J. and Koes, R. (1999) A cytochrome *b*5 is required for full activity of flavonoid 3',5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. PNAs 96: 778-783.

Debeaujon I., Nesi N., Perez P., Devic M., Grandjean O., Caboche M. and Lepiniec L. (2003) Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development. Plant Cell 15, 2514-2531.

Del Carmen Martínez-Ballesta M, Moreno DA and Carvajal M. (2013) The physiological importance of glucosinolates on plant response to abiotic stress in brassica. Int. J. Mol. Sci. 14(6):11607-25.

Delis C., Dimou M., Flemetakis E., Aivalakis G. and Katinakis P. (2006) A root- and hypocotyl-specific gene coding for copper-containing amine oxidase is related to cell expansion in soybean seedlings. J. Exp. Bot. 57, 101-111.

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

DellaPenna D. and O'Connor S.E. (2012) Plant gene clusters and opiates. Science 336, 1648.

Desbrosses G.J. and Stougaard J. (2011) Root nodulation: A paradigm of how plant-microbe symbiosis influences host developmental pathways. Cell Host and Microbe 10 348-358.

Ding S.W., Li H., Lu R., Li F. and Li W.X. (2004) RNA silencing: a conserved antiviral immunity of plants and animals. Virus Research 102, 109-115.

Ding S.W. (2000) RNA silencing. Current Opinion in Biotechnology 11, 152–156.

Dixon R.A. (2001) Natural products and plant disease resistance. Nature 411, 843-847.

Dixon R.A. and Sumner L.W. (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol. 131, 878-885.

Dixon R.A. (2004) Phytoestrogens. Annu Rev. Plant Biol. 55, 225-261.

Dixon RA and Paiva N. (1995) Stressed induced phenyl propanoid metabolism. Plant Cell 7, 1085-1097.

Dixon R.A. (2005). Engineering of plant natural product pathways. Curr. Opin. Plant Biol. 8, 329-336.

Dooner H.K. and Weil C.F. (2007) Give-and-take: interactions between DNA transposons and their host plant genomes. Curr. Opin. In Gen. Dev. 17, 486-492.

Downie J.A. (1998) Functions of rhizobial nodulation genes. The Rhizobiaceae, 387-402.

Doyle J.J. and Doyle J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull 19, 11-15.

Drissen R., Palstra R.J., Gillemans N., Splinter E., Grosveld F., Philipsen S. and de Laat W. (2004) The active spatial organization of the β-globin locus requires the transcription factor EKLF. Genes Dev. 18, 2485-2490.

Dubey V.S., Bhalla R. and Luthra R. (2003) An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants, Journal of biosciences 28, 637-646.

Durbin M.L., McCaig B. and Clegg M.T. (2000) Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol. Biol. 42, 79-92.

Durst F. and Nelson D.R. (1995) Diversity and evolution of plant P450 and P450-reductases. Drug Metab. Drug Interact. 12, 189-206.

Eamens A., Vaistij F.E. and Jones L. (2008a) NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in *Arabidopsis*. Plant J. 55, 596-606.

Eamens A., Wang M.B., Smith N.A. and Waterhouse P.M. (2008b) RNA silencing in plants: yesterday, today and tomorrow. Plant Physiology 147, 456-468.

Ebizuka Y., Katsube Y., Tsutsumi T., Kushiro T. and Shibuya M. (2003) Functional genomics approach to the study of triterpene biosynthesis. Pure Appl. Chem, Vol. 75, 369–374.

Edreva AM, Velikova V and Tsonev T. (2000) Phenylamides in plants. Russ. J. Plant Physiol. 54, 287-301.

Ehlting J., Sauveplane V., Olry A., Ginglinger J.F., Provart N.J. and Werck-Reichhart D. (2008) An extensive (co-)expression analysis tool for the cytochrome P450 superfamily in *Arabidopsis thaliana*. BMC Plant Biology, 8:47.

Eisen M.B., Spellman P.T., Brown P.O. and Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. PNAs 95, 14863-14868.

Ellington A.A., Berhow M. and Singletary K.W. (2005) Induction of macroautophagy in human colon cells by soybean B-group triterpenoid saponins. Carcinogenesis 26, 159-167.

Endara, M.J. and Coley, P.D. (2011) The resource availability hypothesis revisited: A meta analysis. Funct. Ecol. 25, 389–398.

Farag M.A., Huhman D.V., Richard A. Dixon, and Sumner L.W. (2008) Metabolomics Reveals Novel Pathways and Differential Mechanistic and Elicitor-Specific Responses in Phenylpropanoid and Isoflavonoid Biosynthesis in *Medicago truncatula* Cell Cultures. Plant Physiology, 146, 387–402.

Fazio G.C., Xu R. and Matsuda, S.P. (2004) Genome mining to identify new plant triterpenoids, J. Am. Chem. Soc. 126, 5678-5679.

Federico K. (2007) Gateway recombinational cloning: a biological operating system. Expert opin. Drug Discov. 2, 571-589.

Feil C., Sussmuth R., Jung G. and Poralla K. (1996) Site-directed mutagenesis of putative active residues in squalene-hopene cyclase. Eur. J. Biochem. 242, 51-55.

Fenselau de Felippes F., Wang J.W. and Weigel D. (2012) MIGS: miRNA-induced gene silencing. The Plant Journal 70, 541–547.

Fernandes M., Keller N.P. and Adams T.H. (1998) Sequence-specific binding by *Aspergillus nidulans* AflR, a C6 zinc cluster protein regulating mycotoxin biosynthesis. Mol. Microbiol. 28, 1355-1365.

Field B., Jordán F. and Osbourn A.E. (2006) First encounters – deployment of defence-related natural products by plants. New Phytologist 172, 193–207.

Field B. and Osbourn A.E. (2008) Metabolic Diversification—Independent Assembly of Operon-Like Gene Clusters in Different Plants. Science 320, 543-547.

Field B. and Osbourn A.E. (2010) Operons. Nat. Chem. Biol. 19, April 2010, DOI: 10. 1038/NCHEMBIO.359.

Field B. and Osbourn A.E. (2012) Order in the playground. Formation of plant gene clusters in dynamic chromosomal regions. Mobile Genetic Elements 2, 46-50.

Field B., Fiston-Lavier A.S., Kemen A., Geisler K., Quesneville H., and Osbourn A.E. (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. PNAS 108, 16116-16121.

Fischle W., Wang Y. and Allis C.D. (2003) Histone and chromatin cross-talk. Curr. Opin. Cell Biol. 15, 172-183.

Fiscus E.L. and Booker F.L. (2002) Growth of *Arabidopsis* flavonoid mutant is challenged by radiation longer than the UV-B band. Environ. Exp. Bot. 48, 213-224.

Fiston-Laveir A.S., Anxolabehere D. and Quesneville H. (2007) A model of segmental duplication formation in *Drosophila melanogaster*. Genome Res. 17, 1458-1470.

Flemetakis E, Kavroulakis N, Quaedvlieg NE, Spaink HP, Dimou M, Roussis A, Katinakis P. (2000) *Lotus japonicus* contains two distinct *Enod40* genes that are expressed in symbiotic, nonsymbiotic, and embryonic tissues. Molecular Plant–Microbe Interactions 13: 987–994.

Foo E. and Davies NW. (2011) Strigolactones promote nodulation in pea. Planta. 234(5):1073-81.

Foo E, Yoneyama K, Hugill C, Quittenden LJ, Reid JB. (2013) Strigolactones: Internal and external signals in plant symbioses? Plant Signal Behav. 8;8(3).

Forsberg E.C., Downs K.M., Christensen H.M., Im H., Nuzzi P.A. and Bresnick E.H. (2000) Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. PNAs 97, 14494-14499.

Fraser J., Rousseau M., Shenker S., Ferraiuolo M.A., Hayashizaki Y., Blanchette M. and Dostie J. (2009) Chromatin conformation signatures of cellular differentiation. Genome Biol. 10, R37.

Freeling M., Lyons E., Pedersen B., Alam M., Ming R. and Lisch D. (2008) Many of most genes in *Arabidopsis* transposed after the origin of the order Brassicales. Genome Res. 18, 1924-1937.

Frey M., Chomet P., Glawischnig E., Stettner C., Grün S., Winklmair A., Eisenreich W., Bacher A., Meeley R.B., Briggs S.P., Simcox K. and Gierl A. (1997) Analysis of a chemical plant defense mechanism in grasses. Science 277(5326):696-699.

Frey M., Stettner C., Pare P.W., Schmeltz E.A., Tumlinson J.H. and Gierl A. (2000) An herbivore elicitor activates the gene for indole emission in maize. PNAs 26, 14801-14806.

Frey M., Huber K., Park W.J., Sicker D., Lindberg P., Meeley R.B., Simmons C.R., Yalpani N. and Gierl A. (2003) A 2-oxoglutarate-dependent dioxygenase is integrated in DIMBOA-biosynthesis. Phytochemistry 62, 371-376.

Frey M., Schullenher K., Dick R., Fiesselmann A. and Gierl A. (2009) Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. Phytochemistry 70, 1645-1651.

Frick S, Ounaroon A, Kutchan TM (2001) Combinatorial biochemistry in plants: the case of O-methyltransferases. Phytochemistry 56, 1-4.

Friedman M. (2006) Potato glycoalkaloids and metabolites: roles in the plant and in the diet. J Agric. Food Chem. 54, 8655-8681.

Fukushima E.O., Seki H., Ohyama K., Ono E., Umenoto N., Mizutani M., Saito K. and Muranaka T. (2011) CYP716A Subfamily Members are Multifunctional Oxidases in Triterpenoid Biosynthesis. Plant Cell Physiol. 52(12): 2050-2061.

Fukushima E.O., Seki H., Sawai S., Suzuki M., Ohyama K., Saito K. and Muranaka T. (2013) Combinational biosynthesis of legume natural and rare triterpenoids in engineering yeast. Plant Cell Physiol. 54, 740-749.

Fusaro A.F., Matthew L., Smith N.A., Curtin S.J., Dedic-Hagan J., Ellacott G.A., Watson J.M., Ming-Bo Wang, Brosnan C., Carroll B.J. and Waterhouse P.M. (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defense pathway. EMBO reports 7, 1168-1175.

Gamas P., Niebel F.D.C., Lescure N. and Cullimore J.V. (1999) Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. Mol. Plant-Microbe Interact. 9, 233-242.

Geisler K., Hughes R.K., Sainsbury F., Lomonossoff G.P., Rejzek M., Fairhurst S., Olsen C.E., Motawia M.S., Melton R.E., Hemmings A.M., Bak S. and Osbourn A. (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. PNAs 110, 3360-3367.

Gera M. J., kenneth W.M. and Richard B.T. (2007) Elevated temperatures increase leaf senescence and Root secondary metabolite concentrations in the understory herb *Panax quinquefolius* (araliaceae). Am. J. Bot. 94, 819-826.

Ghildiyal M. and Zamore P.D. (2009) Small silencing RNAs: an expanding universe. Nat. Rev. Genet 10, 94-108.

Gibbons G.F., Goads L.J., Goodwin T. W. and Nes W.R. (1971) Concerning the Role of Lanosterol and Cycloartenol in Steroid Biosynthesis. The journal Of Biological Chemistry 246, 3967-3976.

Gibson M.R. (1978) Glycyrrhiza in old and new perspectives. Lloydia 41, 348-354.

Gierl A. and Frey M. (2001) Evolution of benzoxazinone biosynthesis and indole production in maize. Planta 213, 493-498.

Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y (2006) Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. PNAs 103, 14701–14706.

Gleba Y., Marillonnet S. and Klimyuk V. (2004) Engineering viral expression vectors for plants: the "full" virus and the "deconstructed" virus strategies. Curr. Opin. Plant Biol. 7, 182-188.

Go Y.S., Lee S.B., Kim H.J., Kim J., Park H.Y., Kim J.K., Shibata K., Yokota T., Ohyama K., Muranaka T., Arseniyadis S. and Suh M.C. (2012) Identification of marneral synthase, which is critical for growth and development in *Arabidopsis*. The Plant Journal 72, 791-804.

Goll M.G. and Bestor T.H. (2005) Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481-514.

Gomez-Roldan V., Fermas S., Brewer P.B., Puech-Pages V., Dun E.A., Pillot J.P., et al. (2008). Strigolactone inhibition of shoot branching. *Nature* 455, 189–194.

Govindarajulu M, Kim SY, Libault M, Berg RH, Tanaka K, Stacey G, Taylor CG. (2009) GS52 ecto-apyrase plays a critical role during soybean nodulation. Plant Physiol. 149(2):994-1004.

Graham P.H. and Vance C.P. (2003) Legumes: importance and constraints to greater use. Plant Physiology 131, 872-877.

Grandmougin-Ferjani A., Dalpe Y., Hartmann M.A., Laruelle F. and Sancholle M. (1999) Sterol distribution in arbuscular mycorrhizal fungi. Phytochemistry 50, 1027-1031.

Grewal S.I.S. and Moazed D. (2003) Heterochromatin and epigenetic control of gene expression. Science 301, 798-802.

Griffith M and Yaish MWF. (2004) Antifreeze proteins in overwintering plants: a tale of two activities. Trends Plant Sci. 9, 399-405.

Gross J., Cho W.K., Lezhneva L., Falk J., Krupinska K., Shinozaki K., Seki M., Herrmann R.G. and Meurer J. (2006) A plant locus essential for phylloquinone (vitamin K1) biosynthesis originated from a fusion of four eubacterial genes. J. Biol. Chem. 281, 17189-17196.

Gubler F., Millar A.A. and Jacobsen J.V. (2005) Dormancy release, ABA and preharvest sprouting. Curr. Opin. Plant Biol. 8, 183–187.

Guenin S., Mauriat M., Pelloux J., Olivier Van Wuytswinkel, Bellini C. and Gutierrez L. (2009) Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. Journal of Experimental Botany 60, 487–493.

Guhling O., Hobl B., Yeats T. and Jetter R. (2006) Cloning and characterization of a lupeol synthase involved in the synthesis of epicuticular wax crystals on stem and hypocotyl surfaces of *Ricinus communis*. Archives of Biochemistry and Biophysics 448, 60-72.

Guinel F.C. and Geil R.D. (2002) A model for the development of the rhizobial and arbuscular mycorrhizal symbioses in legumes and its use to understand the roles of ethylene in the establishment of these two symbioses. Can. J. Bot. 80, 695–720.

Hamilton A.J. and Baulcombe D.C. (1999) A novel species of small antisense RNA in post-transcriptional gene silencing. Science 286, 950-952.

Han J.Y., Kwon Y.S., Yang D.C., Jung Y.R. and Choi Y.E. (2006) Expression and RNA-interference-induced silencing of the dammarenediol synthase gene in *Panax ginseng*. Plant Cell Physiol. 47, 1653-1662.

Han J.Y., Kim H.J., Kwon Y.S. and Choi Y.E. (2011) The Cyt P450 enzyme CYP716A47 catalyses the formation of protopanaxatriol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Physiol. 52, 2062-2073.

Han J.Y., Hwang H.S., Choi S.W., Kim H.J. and Choi Y.E. (2012) Cytochrome P450 CYP716A53v2 catalyses the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Physiol. 53, 1535-1545.

Handberg K. and Stougaard J. (1992) *Lotus japonicus*, an autougamous, diploid legume species for classical and molecular genetics. Plant J. 2, 487-496.

Haralampidis K., Bryan G., Qi X., Papadopoulou K., Bakht S., Melton R., and Osbourn A. (2001) A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. PNAS 98, 13431-13436.

Haralampidis K., Trojanowska M. and Osbourn A.E. (2002) Biosynthesis of triterpenoid saponins in plants. Advances in biochemical engineering / biotechnology 75, 31-49.

Harborne J.B. (1999) Recent advances in clemical ecology. Natural Products Reports 16, 509-523.

Harrison D. M. (1985) The biosynthesis of Triterpenoids and Steroids. Nat Prod Rep. 2, 525-60.

Hartmann K., Peiter E., Koch K., Schobert S. and Schreiber L. (2002) Chemical composition and ultrastructure of broad bean (*Vicia faba* L.) nodule endodermis in comparison to the root endodermis. Planta 215, 14-25.

Hayashi M, Imaizumi-Anraku H, Akao S, Kawaguchi M. (2000) Nodule organogenesis in *Lotus japonicus*. Journal of Plant Research 113: 489–495.

Hayashi H., Huang P., Kirakosyan A., Inoue K., Hiraoka N., Ikeshiro Y., Kushiro T., Shibuya M. and Ebizuka Y. (2001) Cloning and Characterization of a cDNA Encoding β-amyrin Synthase Involved in Glycyrrhizin and Soyasaponin Biosyntheses in Licorice. Biol. Pharm. Bull. 24, 912—916.

Hayashi M., Miyahara S., Sato S., Kato T., Yoshikawa M., Taketa M., Pedrosa A., Onda R., Imaizumi-Anraku H., Bachmair A., Sandal N., Stougaard J., Murooka Y., Tabata S., Kawasaki S., Kawaguchi M. and Harada K. (2001) Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F_2 population. DNA Res. 8, 311-318.

Hayashi H., Huang P. and Inoue K. (2003) Up-regulation of soyasaponin biosynthesis of methyl jasmonate in cultured cells of *Glycyrrhiza glabra*. Plant Cell Physiol. 44, 404-411.

Hayashi H., Huang P., Satoko Takada S., Obinata M., Inoue K., Shibuya M. and Ebizuka Y. (2004) Differential Expression of Three Oxidosqualene Cyclase mRNAs in *Glycyrrhiza glabra*. Biol. Pharm. Bull. 27, 1086—1092.

Hayek EWH, Jordis U, Moche W and Sauter F (1989) A bicentennial of betulin. Phytochemistry 28, 2229–2242.

He K., Guo A.Y., Gao G., Zhu Q.H., Liu X.C., Zhang H., Chen X., Gu X. and Luo J. (2010) Computational identification of plant transcription factors and the construction of the PlantTFDB database. Methods Mol Biol. 674, 351-368.

He J.X., Akao T., Nishino T. and Tani T. (2001) The influence of commonly prescribed synthetic drugs for peptic ulcer on the pharmacokinetic fate of glycyrrhizin from Shaoyao-Gancao-tang. Biol. Pharm. Bull. 24, 1395-1399.

He JX, Gendron JM, Yang Y, Li J, Wang ZY. (2002) The GSK-3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis*. PNAs, USA 99: 10185–10190.

Heckmann A.B., Lombardo F., Miwa H., Perry J.A., Bunnewell S., Parniske M., Wang T.L. and Downie J.A. (2006) *Lotus japonicus* nodulation requires two GRAS domain regulators, one of which is functionally conserved in a non-legume. Plant Physiol. 142, 1739-1750.

Helliwell C.A., Chandler P.M., Poole A., Dennis E.S. and Peacock W.J. (2001) The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. PNAs 98(4):2065-70.

Henderson I.R. and Jacobsen S.E. (2008) Tandem repeats upstream of the *Arabidopsis* endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev. 22, 1597-1606.

Hentze, M.W. (1994). Enzymes as RNA-binding proteins: A role for (di)nucleotide-binding domains? Trends Biochem. Sci. 19, 101–103.

Hernalsteens J.P., Van Vliet F., De Beuckeleer M., Depicker A., Engler G., Lemmers M., Holsters M., Van Montagu M. and Schell J. (1992) The *Agrobacterium tumefaciens* Ti plasmid as a host vector system for introducing foreign DNA in plant cells. Biotechnology. 24, 374-376.

Hernandez L.E. and Cooke D.T. (1996) Lipid composition of symbiosomes from pea root nodules. Phytochemistry 42, 341-346.

Herr A.J. and Baulcombe D.C. (2004) RNA Silencing Pathways in Plants. Cold Spring Harbor Symposia on Quantitative Biology, Volume LXIX.

Herrera J.B.R., Bartel B., Wilson W.K. and Seiichi P. T. Matsuda. (1998) Cloning and characterization of the *Arabidopsis thaliana* lupeol synthase gene. Phytochemistry 49, 1905-1911.

Higo K., Ugawa Y., Iwamoto M. and Korenaga T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Research 27, 297-300.

Hill R.A and Connolly J.D. (2012) Triterpenoids. Nat. Prod. Rep 29, 780-818.

Hittinger C.T., Rokas A. and Carroll S.B. (2004) Parallel inactivation of multiple *GAL* pathway genes and ecological diversification in yeasts. PNAs 101, 14144-14149.

Hoffmeister D. and Keller N.P. (2007) Natural products of filamentous fungi: enzymes, genes, and their regulation. Nat. Prod. Rep. 24, 393-416.

Horton R., Wilming L., Rand V., Lovering R.C., Bruford E.A., Khodiyar V.K., Lush M.J., Povey S., Talbot C.C. Jr, Wright M.W., Wain H.M., Trowsdale J., Ziegler A. and Beck S. (2004) Gene map of the extended human MHC. Nat. Rev. Genet. 5, 889-899.

Hostettmann K.A. and Marston A. (1991) Saponins (Cambridge Univ. Press, Cambridge, UK).

Hughes J. and Hughes M.A. (1994) Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta* Crantz) cotyledons. DNA Seq. 5, 41–49.

Huhman D.V. and Sumner L.W. (2002) Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry 59, 347–360.

Huhman D.V., Berhow M.A. and Sumner L.W. (2005) Quantification of Saponins in Aerial and Subterranean Tissues of *Medicago truncatula*. J. Agric. Food Chem. 53, 1914-1920.

Hurst L.D, Pál C. and Lercher M.J. (2004) The evolutionary dynamics of eukaryotic gene order. Nat. Rev.Genet. 5, 299-310.

Husselstein-Muller T., Schaller H. and Benveniste P. (2001) Molecular cloning and expression in yeast of 2,3-oxidosqualene-triterpenoid cyclases from *Arabidopsis thaliana*. Plant Mol. Biol. 45, 75-92.

Ignoato M.C., Fabrão R.M., Schuquel I.T.A., Botelho M.F.P., Bannwart G., Pomini A.M., Arruda L.L.M., Ciomar A., Amado B. and Santin S.M.O. (2012) Chemical constituents of *Machaerium hirtum Vell*. (Fabaceae) leaves and branches and its anti-inflammatory activity evaluation. Nat Prod Res 1-6.

Iida T., Nakayama J. and Moazed D. (2008) siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription. Mol. Cell 31, 178-189.

Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T., Sakaki Y., Hattori M. and Omura S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat. Biotechnol. 21, 526-531.

Imaizumi-Anraku H., Takeda N., Charpentier M., Perry J., Miwa H., Umehara Y., Kouchi H., Murakami Y., Mulder L., Vickers K., Pike J., Downie J.A., Wang T., Sato S., Asamizu E., Tabata S., Yoshikawa M., Murooka Y., Wu G.J., Kawaguchi M., Kawasaki S., Parniske M. and Hayashi M. (2005) Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. Nature 433, 527-531.

Inagaki Y.S., Etherington G., Geisler K., Field B., Dokarry M., Ikeda K., Mutsukado Y. and Osbourn A. (2011) Investigation of the potential for triterpene synthesis in rice through genome mining and metabolic engineering. New Phytologist 191, 432-448.

Invitrogen Co: Gateway® Technology Manual. [http://www.invitrogen.com].

Ishii Y. and Tanizawa H. (2006) Effects of soyasaponins on lipid peroxidation through the secretion of thyroid hormones. Biol. Pharm. Bull. 29, 1759-1763.

Islam A.K., Ali M.A., Sayeed A., Salam S.M., Islam A., Rahman M., Khan G.R. and Khatun S. (2003) An antimicrobial terpenoid from *Caesalpinia pulcherrima* Swartz.: Its characterization, antimicrobial and cytotoxic activities. Asian J. Plant Sci. 2, 17-24.

Islam S., Miyazaki T., Tanno F. and Itoh K. (2005) Dissection of gene function by RNA silencing. Plant Biotechnology 22, 443–446.

Itkin M., Heinig U., Tzfadia O., Bhide A.J., Shinde B., Cardenas P.D., Bocobza S.E., Unger T., Malitsky S., Finkers R., Tikunov Y., Bovy A., Chikate Y., Singh P., Rogachev I., Beekwilder J., Giri A.P. and Aharoni A. (2013) Biosynthesis of antinutricional alkaloids in solanaceous crops is mediated by clustered genes. Science 341, 175-179.

Ito R., Mori K., Hashimoto I., Nakano C., Sato T. and Hoshino T. (2011) Triterpene cyclases from *Oryza sativa* L.: cycloartenol, parkeol and achilleol B synthases. Org. Lett. 13, 2678-2681.

Iturbe-Ormaetxe I., Haralampidis K., Papadopoulou K. and Osbourn A.E. (2003) Molecular cloning and characterization of triterpene synthases from *Medicago truncatula* and *Lotus japonicus*. Plant Molecular Biology 51, 731–743.

Jacob F. and Monod J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. L. Mol. Biol. 3, 318-356.

Jang JC, Fujioka S, Tasaka M, Seto H, Takatsuto S, Ishii A, Aida M, Yoshida S, Sheen J (2000) A critical role of sterols in embryonic patterning and meristem programming revealed by the fackel mutants of *Arabidopsis thaliana*. Genes Dev. 14:1485-1497.

Janga S.C., Collado-Vides J. and Babu M. (2008) Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. PNAs 105, 15761-15766.

Janska A, Marsik P, Zelenkova S and Ovesna J. (2010) Cold stress and acclimation—what is important for metabolic adjustment? Plant Biol. 12, 395-405.

Jenner H., Townsend B. and Osbourn A. (2005) Unravelling triterpene glycoside synthesis in plants: phytochemistry and functional genomics join forces. Planta 22, 503-506.

Jiao Y.N., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho L.P., Hu Y., Liang H.Y., Soltis P.S. et al. (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473, 97–100.

Jin G.L., Zheng C.J., Xin W.B., Mao Z.J., Sun P.X., Zeng Z.X. and Qin P. (2010) Triterpenoid saponins from the seeds of *Caragana microphylla*. Arch. Pharm. Res. 34, 869-873.

Jonczyk R., Schmidt H., Osterrieder A., Fiesselmann A., Schullehner K., Haslbeck M., Sicker D., Hofmann D., Yalpani N., Simmons C., Frey M. and Gierl A. (2008) Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of Bx6 and Bx7. Plant Physiol. 146, 1053-1063.

Jones P. and Vogt T. (2001) Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta 213, 164–174.

Jørgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, Zagrobelny M, Bak S, Møller BL. (2005) Metabolon formation and metabolic channeling in natural product biosynthesis. Curr. Opin. Plant Biol. 8, 280-291.

Junier T. and Pagni M. (2000) Dotlet: diagonal plots in a web browser. Bioinformatics. 16, 178-179.

Jurzysta M. (1973) Isolation and chemical characterization of saponins from Lucerne seeds. Acta Societatis Boranicorum Poloniae 42, 201-207.

Kahn R.A. and Durst F. (2000) Function and evolution of plant cytochrome P450. Recent Adv. Phytochem. 34 151-189.

Kalinowska M., Zimowski J., Paczkowski C. and Wojciechowski Z.A. (2005) The formation of sugar chains in triterpenoid saponins and glycoalkaloids. Phytochemistry Reviews 4, 237–257.

Kallberg Y., Oppermann U., Jornvall H. and Persson B. (2002) Short-chain dehydrogenases/reductases (SDRs). Eur. J. Biochem. 269, 4409-4417.

Kaneko T., Nakamura Y, Sato S., Aasmizu E., Kato T., Sasamoto S., Watanabe A., Idesawa K., Ishikawa A., Kawashima., Kimura T., Kishida, Y., Kiyokawa C., Kohara M., Matsumoto, M., Matsuno A., Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M., Takeuchi C., Yamada M., and Tabata S. (2000) Complete genome structure of the nitrogen fixing symbiotic bacterium *Mesorhizobium loti*. DNA Research 7, 331-338.

Kappers I.F., Aharoni A., Van Herpen T., Luckerhoff L., Dicke M. and Bouwmeester H.J. (2005) Genetic engineering of terpenoid metabolism attracts bodyguards to *Arabidopsis*. Science 309, 2070-2072.

Kawagushi M. (2000) *Lotus japonicus* `Miyakojima' MG-20: An Early-Flowering Accession Suitable for Indoor Handling. Journal of Plant Research 113, 507-509.

Kazan K. and Manners J.M. (2008) Jasmonate signaling: toward an integrated view. Plant Physiol. 146, 1459-1468.

Kessler A. and Baldwin T. (2001) Defensive function of herbivore-induced plant volatile emission in nature. Science 291, 2141-2144.

Khan, N.A., Nazar, R. and Anjum, N.A. (2009) Growth, photosynthesis and antioxidant metabolism in mustard (*Brassica juncea* L.) cultivars differing in ATP-sulfurylase activity under salinity stress. Sci. Hort. 122, 455–460.

Khan N.A. and Srivastava A. (2009) Antifungal activity of bioactive triterpenoid saponin from the seeds of *Cassia angustifolia*. Natural Product Research 23, 1128-1133.

Kikuchi A., Tsukamoto C., Tabuchi K., Adachi T. and Okubo K. (1999) Inheritance and characterization of a null allele for group A acetyl saponins found in a mutant soybean (*Glycine max* (L.) Merrill). Breed Sci. 49, 167-171.

Kim J.E., Cheng K.M., Craft N.E., Hamberger B. and Douglas C.J. (2010) Over-expression of *Arabidopsis thaliana* carotenoid hydroxylases individually and in combination with a beta-carotene ketolase provides insight into in vivo functions. Phytochemistry 71, 168-78.

Kinjo J., Udayama M., Okawa M. and Nohara T. (1999) Study and structure-hepatoprotective relationships of oleanene-type triterpenoidal glucuronides obtained from several fabaceous plants on rat primary hepatocyte cultures. Biol. and Pharmaceut. Bulletin 22, 203-206.

Kirby J. and Keasling J.D. (2009) Biosynthesis of Plant Isoprenoids: Perspectives for Microbial Engineering. Annu. Rev. Plant Biol. 60, 335-355.

Kirti P. B. (2008) Handbook of new technologies for genetic improvement of legumes. CRC press.

Kitagawa I., Taniyama T., Nagahama Y., Okubo K., Yamauchi F. and Yoshikawa M. (1982) Saponin and sapogenol. XLII. Structures and acetyl-soyasaponins A1, A2 and A3, astringent partially acetylated bisdesmosides of soyasapogenol A, from American soybean, the seeds of *Glycine max* Merrill. Chem. Pharm. Bull 36, 2819-2828.

Kliebenstein D.J., Kroymann J. and Michael-Olds T. (2005) The glucosinolate-myrosinase system in an ecological and evolutionary context. Curr. Op. in Plant Biol. 8, 264-271.

Kliebenstein D.J. and Osbourn A. (2012) Making new molecules – evolution of pathways for novel metabolites in plants. Current Opinion in Plant Biology 15, 415–423.

Koga J., Shimura M., OshimaK., Ogawa N., Yamauchi T. and Ogasawara N. (1995) Phytocassanes A, B, C and D, novel diterpene phytoalexins from rice, *Oryza sativa*. Tetrahedron 51, 7907-7918.

Kolesnikova M.D., Xiong Q., Lodeiro S., Hua L. and Matsuda S.P.T. (2006) Lanosterol biosynthesis in plants. Archives of Biochemistry and Biophysics 447, 87–95.

Koprowski H. and Yusibov V. (2001) The green revolution: plants as heterologous expression vectors. Vaccine 19, 2735–2741.

Kovacs Z, Simon-Sarkadi L, Szucs A and Kocsy G. (2011) Differential effects of cold, osmotic stress and abscisic acid on polyamine accumulation in wheat. Amino Acids 38, 623-31.

Kristensen C., Morant M., Olsen C.E., Ekstrøm C.T., Galbraith D.W., Møller B.L. and Bak S. (2005) Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. PNAs 102, 1779-1784.

Krokida A., Delis C., Geisler K., Garagounis C., Tsikou D., Pena-Rodriguez L., Katsarou D., Osbourn A.E. and Papadopoulou K.K. (2013) A metabolic gene cluster in Lotus japonicus discloses novel enzyme functions and products in triterpene diosynthesis. New Phytologist, article in press.

Kudo F., Motegi A., Mizoue K. and Eguchi T. (2010) Cloning and characterization of the biosynthetic gene cluster of 16-membered macrolide antibiotic FD-891: involvement of a dual functional cytochrome P450 monooxygenase and hydroxylation. ChemBioChem 11, 1574–1582.

Kudou S., Tonomura M., Tsukamoto C., Shimoyamada M., Uchida T. and Okubo K. (1992) Isolation and structural elucidation of the major genuine soybean saponin. Biosci. Biotech. Biochem. 56, 142-143.

Kouzarides T. (2007) Chromatin modifications and their function. Cell 128, 693–705.

Kumagai H, Kinoshita E, Ridge RW, Kouchi H. (2006) RNAi knockdown of ENOD40s leads to significant suppression of nodule formation in *Lotus japonicus*. Plant Cell Physiology 47: 1102–1111.

Kuo H.F., Olsen K.M. and Richards E.J. (2006) Natural variation in a subtelomeric region of *Arabidopsis*: implications for the genomic dynamics of a chromosome end. Genetics 173, 401–417.

Kurosawa Y., Takahara K. and Shiraiwa M. (2002) UDP-glucuronic acid: soyasapogenol glucuronosyl-transferase involved in saponin biosynthesis in germinating soybean seeds. Planta 215, 620-629.

Kusaba M. (2004) RNA interference in crop plants. Current Opinion in Biotechnology 15, 139–143.

Kushiro T., Shibuya M. and Ebizuka Y. (1998) *β*-Amyrin synthase: Cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. Eur. J. Biochem. 256, 238-244.

Kushiro T., Shibuya M., Masuda K. and Ebizuka Y. (2000) A novel multifunctional triterpene synthase from *Arabidopsis thaliana*. Tetrahedron Letters 41, 7705–7710.

Kwon S.H. and Workman J.L. (2008) The heterochromatin protein 1 (HP1) family: put away a bias toward HP1. Mol. Cells 26, 217-227.

Laden B.P., Tang Y. and Porter T.D. (2000) Cloning, heterologous expression, and enzymological characterization of human squalene monooxygenase, Archives of biochemistry and biophysics 374, 381-388.

Lambert E., Ahmad Faizal A. and Geelen D. (2011) Modulation of Triterpene Saponin Production: In Vitro Cultures, Elicitation, and Metabolic Engineering. Appl Biochem Biotechnol 164, 220–237.

Landry L.G., Chapple C.C.S. and Last R.L. (1995) Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. Plant Physiology 109, 1159-1166.

Latunde-Dada A.O., Cabello-Hurtado F., Czittrich N., Didierjean L., Schopfer C., Hertkorn N., Werck-Reichhart D. and Ebel J. (2001) Flavonoid 6-hydroxylase from soybean (*Glycine max* L.), a novel plant P450 monooxygenase. Journal of Biological Chemistry 276, 1688–1695.

Lee L.Y. and Gelvin S.B. (2008) T-DNA Binary Vectors and Systems. Plant Physiology 146, 325–332.

Lee S, Cheng H, King KE, Wang W, He Y, Hussain A, Lo J, Harberd NP, Peng J. (2002) Gibberellin regulates *Arabidopsis* seed germination via rgl2, a gai / rga-like gene whose expression is up-regulated following imbibition. Genes Development 16: 646–658.

Lee TK, Poon RT, Wo JY, Ma S, Guan XY, Myers JN, Altevogt P, Yuen AP. (2007) Lupeol suppresses cisplatin-induced nuclear factor-kappaB activation in head and neck squamous cell carcinoma and inhibits local invasion and nodal metastasis in an orthotopic nude mouse model. Cancer Research 67: 8800–8809.

Leivar P, Gonzales VM, Castel S, Trelease RN, Lopez-Iglesias C, Arro M, Boronat A, Campos N, Ferrer A, Fernandez-Busquets X. (2005) Subcellular localization of *Arabidopsis* 3-hydroxy-3- methylglutaryl-coenzyme A reductase. Plant Physiol 137, 57-69.

Lhuissier F.G.P., De Ruijter N.C.A., Sieberer B.J., Esseling J.J. and Emons A.M. (2001) Time course of cell biological events evoked in legume root hairs by *Rhizobium* Nod factors: state of the art. Annals of Botany, 87, 289-302.

Lescot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., Rouzé P. and Rombauts S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 30, 325-327.

Le Signor C., Savois V., Aubert G., Verdier J., Nicolas M., Pagny G., Moussy F., Sanchez M., Baker D., Clarke J. and Thompson R. (2009) Optimizing TILLING populations for reverse genetics in *Medicago truncatula*. Plant Biotechnology Journal 7, 430-441.

Li J.M., Nagpal P., Virtal V., McMorris T.C. and Chory J. (1996) A role for brassinosteroids in light-dependent development of *Arabidopsis*. Science 272, 398-401.

Li Y., Luo H.M., Sun C., Song J.Y., Sun Y.Z., Wu Q., Wang N., Yao H., Steinmetz A. and Chen S.L. (2010) EST analysis reveals putative genes involved in glycyrrhizin biosynthesis. BMC Genomics 11, 268.

Li J., Huang X., Du X., Sun W. and Zhang Y. (2009) Study of chemical composition and antimicrobial activity of leaves and roots of *Scrophularia ningpoensis*, Nat Prod Res 23, 775-780.

Lichtenthaler H.K. (1990) The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annu.Rev. Plant Physiol. Plant Mol. Biol. 50, 47-65.

Lim E.K. and Bowles D.J. (2004) A class of plant glycosyltransferases involved in cellular homeostasis. EMBO J. 23, 2915–2922.

Limpens E., Ramos J., Franken C., Raz V., Compaan B., Franssen H., Bisseling T. and Geurts R. (2004) RNA interference in *Agrobacterium rhizogenes*-transformed roots of *Arabidopsis* and *Medicago truncatula*. J. Exp. Bot. 55, 983-992.

Linardopoulou E.V., Williams E.M., Fan Y., Friedman C., Young J.M. and Trask B.J. (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature 437, 94–100.

Lindbo JA (2007) TRBO: a high-efficiency tobacco mosaic virus RNA-Based overexpression vector. Plant Physiol 145, 1232–1240.

Lindsey K., Pullen M.L. and Topping J.F. (2003) Importance of plant sterols in pattern formation and hormone signaling. Trends Plant Science 8, 521-525.

Liu L. and Lomonossoff G.P. (2002) Agroinfection as a rapid method for propagating Cowpea mosaic virus-based constructs. Journal of Virological Methods 105, 343-348.

Liu Y., Wang H., Ye H.C. and Li G.F. (2005) Advances in the plant isoprenoid biosynthesis pathway and its metabolic engineering, J Integr Plant Biol 47, 769-782.

Liu C., Lu F., Cui X. and Cao X. (2010) Histone methylation in higher plants. Annu. Rev. Plant Biol. 61, 395–420.

Liu J, Novero M, Charnikhova T, Ferrandino A, Schubert A, Ruyter-Spira C, Bonfante P, Lovisolo C, Bouwmeester HJ, Cardinale F. (2013) Carotenoid cleavage dioxygenase 7 modulates plant growth, reproduction, senescence, and determinate nodulation in the model legume *Lotus japonicus*. J Exp Bot. 64(7):1967-1981.

Lodeiro S., Xiong Q., Wilson W.K., Kolesnikova M.D., Onak C.S. and Matsuda S.P. (2007) An oxidosqualene cyclase makes numerous products by diverse mechanisms: a challenge to prevailing concepts of triterpene biosynthesis, J Am Chem Soc 129, 11213-11222.

Lombari P., Ercolano E., E.l. Alaoui and Chiurazzi M. (2003) A new transformation-regeneration procedure in the model legume *Lotus japonicus*: root explants as a source of large numbers of cells susceptible to Agrobacterium-mediated transformation. Plant Cell Rep 21, 771–777.

Louis E.J. (1995) The chromosome ends of *Saccharomyces cerevisiae*. Yeast 11, 1553–1573.

Lu M.B., Wong H.L. and Teng W.L. (2001) Effects of elicitation on the production of saponin in cell culture of *Panax ginseng*. Plant Cell Rep. 20, 674-677.

Luckner M. (1972) Secondary metabolism in plants and animals. Academic Press, New York.

Lunyak V.V., Burgess R., Prefontaine G.G., Nelson C., Sze S.H., Chenoweth J., Schwartz P., Pevzner P.A., Glass C., Mandel G. and Rosenfeld M.G. (2002) Co-repressor dependent silencing of chromosomal regions encoding reuronal genes. Science 298, 1747-1752.

Luo, C., Sidote, D., Zhang, Y., Kerstetter, R., Michael, T., and Lam, E. (2013). Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. Plant J. 73, 77–90.

Lupien S., Karp F., Wildung M. and Croteau R. (1999) Regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. Arch Biochem Biophys 368, 181-192.

Lyons E. and Freeling M. (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant Journal 53, 661–673.

Mackenzie, P.I., Owens, I.S., Burchell, B., Bock, K.W., Bairoch, A., Belanger, A., Fournel-Gigleux, S., Green, M., Hum, D.W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J.R., Ritter, J.K., Schachter, H., Tephly, T.R., Tipton, K.F. and Nebert, D.W. (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics 7, 255–269.

Madsen LH, Fukai E, Radutoiu S, Yost CK, Sandal N, Schauser L and Stougaard J. (2005) LORE1, an active low-copy-number TY3-gypsy retrotransposon family in the model legume *Lotus japonicus*. Plant J. 44, 372-381.

Maekawa T., Kusakabe M., Shimoda Y., Sato S., Tabata S., Murooka Y. and Hayashi M. (2008) Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in *Lotus japonicus*. Molecular Plant-Microbe Interactions 21, 375–382.

Mahato S.M., Nandy A.K. and Roy G. (1992) Triterpenoids. Phytochemistry 31, 2199-2249.

Mahato S.B. and Sen S. (1997) Advances in Triterpenoid Research, 1990-1994. Phytochemistry 44, 1185-1236.

Maimone T.J. and Baran P.S. (2007) Modern synthetic effords toward biologically active terpenes. Nat. Chem. Biol. 3, 396-407.

Mangas S., Bonfill M., Osuna L., Moyano ., Tortoriello J., Cusido R.M., Pinol M.T.P., and Palazon J. (2006) The effect of methyl jasmonate on triterpene and sterol metabolisms of *Centella asiatica*, *Ruscus aculeatus* and *Galphimia glauca* cultured plants. Phytochemistry 67, 2041–2049.

Martin D., Tholl D., Gershenzon J. and Bohlmann J. (2003) Induction of volatile terpene biosynthesis and diurnal emission by methyl jasmonate in foliage of Norway spruce. Plant Physiology 132, 1586-1599.

Martirani L., Stiller J., Mirabella R., Alfano F., Lamberti A., Radutoiu S.E., Iaccarino M., Gresshoff P.M. and Chiurazzi M. (1999) T-DNA Tagging of Nodulation- and Root-Related Genes in *Lotus japonicus:* Expression Patterns and Potential for Promoter Trapping and Insertional Mutagenesis. Mol Plant Microbe In. 12, 275-284.

Marques-Bonet T., Girirajan S. and Eichler E.E. (2009) The origins and impact of primate segmental duplication. Trends Genet. 25, 443-454.

Marquez A.J. (Editorial Director) (2005) *Lotus japonicus* handbook, pp 53-82, 83-86.

Massiot G., Lavaud C., Benkhalem M. and Le Men-Olivier L. (1992) Soyasaponin VI, a new maltol conjugate from alfalfa and soybean. J. Nat. Prod. 55, 1339-1342.

Matsui S., Matsumoto H., Sonoda Y., Ando K., Aizu-Yokota E., Toshitsugu S. and Kasahara T. (2004) Glycyrrhizin and related compounds down-regulate production of inflammatory chemokines IL-8 and eotaxin Iin a human lung fibroblast cell line. Int. Immunopharmacol. 4, 1633-1644.

McCallum C.M., Comai L., Greene E.A. and Henikoff S. (2000a) Targeted screening for induced mutations. Nature Biotechnology 18, 455-457.

McCallum C.M., Comai L., Greene E.A. and Henikoff S. (2000b) Targeting Induced Local Lesions In Genomes (TILLING) for Plant Functional Genomics. Plant Physiol. 123, 439-442.

McCue K.F., Shepherd L.V.T., Allen P.V., Maccree M.M., Rockhold D.R., Corsini D.L., Davies H.V. and Belknap W.R. (2005) Metabolic compensation of steroidal glycoalkaloid biosynthesis in transgenic potato tubers: using reverse genetics to confirm the in vivo enzyme function of a steroidal alkaloid galactosyltransferase. Plant Science 168, 267-273.

McGarvey D.J. and Croteau R. (1995) Terpenoid metabolism, The Plant cell 7, 1015-1026.

Mechtcheriakova, I.A., **Eldarov, M.A.**, **Nicholson, L.**, **Shanks, M.**, **Skryabin, K.G. and Lomonossoff, G.P.** *(*2006*)* The use of viral vectors to produce hepatitis B virus core particles in plants. J. Virol. Methods, 131, 10–15.

Meesapyodsuk D., Balsevich J., Reed D.W. and Covello P.S. (2007) Saponin biosynthesis in Saponaria vaccaria: cDNAs encoding β-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. Plant Physiol. 143, 959-969.

Memelink J., Verpoorte R. and Kijne J.W. (2001) ORCAnization of jasmonate responsive gene expression in alkaloid metabolism. Trends in Plant Sci. 6, 212-219.

Memelink J. (2005) The use of genetics to dissect plant secondary pathways. Current Opinion in Plant Biology 8, 230–235.

Mérillon JM, Doireau P, Guillot A, Chénieux JC and Rideau M (1986) Indole alkaloid accumulation and tryptophan decarboxylase activity in *Catharanthus roseus* cells cultured in three different media. Plant Cell Rep 5, 23-26.

Metlen K.L., Aschehoug E.T. and Callaway R.M. (2009) Plant behavioural ecology: dynamic plasticity in secondary metabolites. Plant, Cell and Environment 32, 641-663.

Migas P., Łuczkiewicz M. and Cisowski W. (2006) The Influence of Auxins on the Biosynthesis of Isoprene Derivatives in Callus Cultures of *Vaccinium corymbosum* var. *Bluecrop*. Z Naturforsch C. 61, 565-570.

Mitra RM and Long SR. (2004) Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula / Sinorhizobium meliloti* symbiosis. Plant Physiology 134: 595–604.

Mizutani M. and Ohta D. (2010) Diversification of P450 genes during land plant evolution. Annu. Rev. Plant Biol. 61, 291-315.

Mizutani M. (2012) Impacts of Diversification of Cytochrome P450 on Plant Metabolism. Biol. Pharm. Bull. 35, 824-832.

Møller BL and Conn EE (1980) The biosynthesis of cyanogenic glucosides in higher plants: channeling of intermediates in dhurrin biosynthesis by a microsomal system from *Sorghum bicolor* (Linn) Moench. J Biol. Chem. 255, 3049-3056.

Monger W., Alamillo J.M., Sola I., Perrin Y., Bestagno M., Burrone O.R., Sabella P., Plana-Duran J., Enjuanes L., Garcia J.A. and Lomonossoff G.P. (2006) An antibody derivative expressed from viral vectors passively immunizes pigs against transmissible gastroenteritis virus infection when supplied orally in crude plant extracts. Plant Biotechnology Journal 4, 623–631.

Moore M.J., Bell C.D., Soltis P.S. and Soltis D.E. (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. PNAs, 104: 19363–19368.

Moraes M.C., Birkett M.A., Gordon-Weeks R., Smart L.E., Martin J.L., Pye B.J., Bromilow R. and Pickett J.A. (2008) cis-Jasmone induces accumulation of defence compounds in wheat, *Triticum aestivum*. Phytochemistry 69, 9-17.

Morant M., Bak S., Møller B.L. and Werck-Reichhart D. (2003) Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. Curr. Opin .Biotechnol. 14: 151-62.

Morant A.V., Jorgensen K., Jorgensen C., Paquette S.M., Sanchez-Perez R., Moller B.L. and Bak S. (2008) β-Glucosidases as detonators of plant chemical defense. Phytochemistry 69, 1795-1813.

Morgante M., De Paoli E. and Radovic S. (2007) Transposable elements and the plant pan-genomes. Curr. Op. in Plant Biol. 10, 149-155.

Morison JIL and Lawlor DW. (1999) Interactions between increasing CO_2 concentration and temperature on plant growth. Plant Cell Environ 22, 659-682.

Morris P. (1986) Regulation of product synthesis in cell cultures of *Catharanthus roseus*. II. Comparison of production media. Planta Med 52. 121-126.

Morita M., Shibuya M., Kushiro T., Masuda K. and Ebizuka Y. (2000) Molecular cloning and functional expression of triterpene synthases from pea (*Pisum sativum*). New α-amyrin-producing enzyme is a multifunctional triterpene synthase. Eur. J. Biochem. 267, 3453-3460.

Morlacchi P., Wilson W.K., Xiong Q., Bhaduri A., Sttiverd D., Kolesnikova M.D. and Matsuda S.P.T. (2009) Product profile of PEN3: the last unexamined oxidosqualene cyclases in *Arabidopsis thaliana*. Org. Lett. 11, 2627-2630.

Moses T, Pollier J, Thevelein JM, Goossens A. **(**2013) Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. New Phytol. 200(1):27-43.

Mroczek A., Kapusta I., Janda B. and Janiszowska W. (2012) Triterpene Saponin Content in the Roots of Red Beet (*Beta vulgaris L*.) Cultivars. J.Agric. Food Chem. 60, 12397−12402.

Mueller J.E., Canze M. and Bryk M. (2006) the requirements for COMPASS and Paf1 in transcriptional silencing and methylation of histone H3 in *Saccharomyces cerevisiae*. Genetics 173, 557-567.

Mugford S.T., Qi X., Bakht S., Hill L., Wegel E., Hughes R.K., Papadopoulou K., Melton R., Philo M., Sainsbury F., Lomonossoff G.P., Roy A.D., Goss R.J. and Osbourn A. (2009) A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. Plant Cell. 21, 2473-2484.

Mugford S.T., Louveau T., Melton R., Qi X., Bakht S., Hill L., Tsurushima T., Honkanen S., Rosser S.J., Lomonossoff G.P. and Osbourn A. (2013) Modularity of plant metabolic gene clusters: A trio of linked genes that are collectively required for acylation of triterpenes in oat. The Plant Cell 25, 1078-1092.

Muller K.O. and Borger H. (1940) Arb. Biol. Reichsasnstalt. Landw. Forstw. Berlin 23, 189-223.

Munns, R. and Tester, M. (2008) Mechanisms of salinity tolerance. Annu. Rev. Plant Physiol. 59, 651–681.

Murata M., Houdai T., Morooka A., Matsumori N. and Oishi T. (2005) Membrane interaction of soyasaponins in association with their antioxidation effect. Soy Prot. Res. 8, 81-85.

Murata T., Miyase T., Muregi F.W., Naoshima-Ishibashi Y., Umehara K., Warashina T., Kanou S., Mkoji G.M., Terada M. and Ishin A. (2008). Antiplasmodial terpenoid from *Ekebergia capensis*. J. Plant Nat. Prod. 71, 167-174.

Murray JD, Karas BJ, Sato S, Tabata S, Amyot L, Szczyglowski K. (2007) A cytokinin perception mutant colonized by rhizobium in the absence of nodule organogenesis. Science 315: 101–104.

Mylona P., Owatworakit A., Papadopoulou K.K., Jenner H., Qin B., Findlay K., Hill L., Xiaoquan Qi, Bakht S., Melton R., and Osbourn A.E. (2008) *Sad3* and *Sad4* Are Required for Saponin Biosynthesis and Root Development in Oat. The Plant Cell 20, 201–212.

Nagegowda D.A. (2010) Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. FEBS letters 584, 2965-2973.

Nagy, E., Henics, T., Eckert, M., Miseta, A., Lightowlers, R.N., and Kellermayer, M. (2000). Identification of the NAD(+)-binding fold of glyceraldehyde-3-phosphate dehydrogenase as a novel RNA-binding domain. Biochem. Biophys. Res. Commun. 275, 253–260.

Nakagawa T. and Kawaguchi M. (2006) Shoot-applied MeJA Suppresses Root Nodulation in *Lotus japonicus*. Plant Cell Physiol. 47, 176–180.

Naoumkina M.A., Modolo L.V., Huhman D.V., Urbanczyk-Wochniak E., Tang Y., Sumner L.W. and Dixon R.A. (2010) Genomic and Coexpression Analyses Predict Multiple Genes Involved in Triterpene Saponin Biosynthesis in *Medicago truncatula*. The Plant Cell 22, 850–866.

Navarro L., Bari R., Achard P., Lisón P., Nemri A., Harberd N.P. and Jones J.D. (2008) DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. Curr. Biol. 18, 650-655.

Nei M.M. (1967) Modification of linkage intensity by natural selection. Genetics 57, 625-641.

Nei M. (2003) Let's stick together. Heredity 90, 411-412.

Nelson D.R., Schuler M.A., Paquette S.M., Werck-Reichhart D. and Bak S. (2004) Comparative genomics of Rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiol. 135: 756-772.

Nelson D. (2011) Progress in tracing the evolutionary paths of cytochrome P450. Biochim. Biophys. Acta 1814, 14-18.

Nelson D. and Werck-Reichhart D. (2011) A P450-centric view of plant evolution. Plant Journal 66, 194–211.

Nemhauser J.L., Hong F. and Chory J. (2006) Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. Cell 126, 467-475.

Nes W.D. and Heftmann E. (1981) A comparison of triterpenoids with steroids as membrane components. J. Nat. Prod. 44, 377-400.

Nicot N., Hausman J.F., Hoffmann L. and Danie le Evers. (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of Experimental Botany 56, 2907–2914.

Niemeyer H.M. (1988) Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones) defense chemicals in the Gramineae. Phytochem. Anal. 27, 3349-3358.

Nishimura R, Hayashi M, Wu GJ, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M et al. (2002) *Har1* mediates systemic regulation of symbiotic organ development. Nature 420: 426–429.

Niwa S, Kawagushi M, Imaizumi-Anraku H., Chechetka S.A., Ishizaka M., Ikuta A. and Kuchi H., (2001) Responses of a model legume *Lotus japonicus* to lipochitin oligosaccharide nodulation factors purified from *Mesorhizobium loti* JRL501. MPMI 14, 848-856.

Noma K., Allis C.D. and Grewal S.I. (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 293, 1150-1155.

Ober D. (2005) Seeing double: gene duplication and diversification in plant secondary metabolism. Trends in Plant Science 10, 444–449.

Ober D. (2010) Gene duplications and the time thereafter – examples from plant secondary metabolism. Plant Biol. 12, 570-577.

Ohana R., Delmer D.P., Carlson R.W., Glushka J., Azadi P., Bacic T. and Benziman M. (1998) Identification of a Novel Triterpenoid Saponin from *Pisum sativum* as a Specific Inhibitor of the Diguanylate Cyclase of *Acetobacter xylinum*. Plant Cell Physiol. 39, 144-152.

Ohmoto T. and Ikuze M. (1970) Triterpenoids from Gramineae. Phytochemistry 9, 2137-2148.

Ohmido N., Kijima K., Ashikawa I., de Jong J.H. and Fukui K. (2001) Visualization of the terminal structure of rice chromosomes 6 and 12 with multicolor FISH to chromosomes and extended DNA fibers. Plant Mol. Biol. 47, 413–421.

Ohyama K., Suzuki M., Kikuchi J., Saito K. and Muranaka T. (2009) Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in *Arabidopsis*. PNAS 106, 725–730.

Okada A., Shimizu T., Okada K., Kuzuyama T., Koga J., Shibuya N., Nojiri H. and Yamane H. (2007) Elicitor induced activation of the methylerythritol phosphate pathway toward phytoalexins biosynthesis in rice. Plant Mol. Biol. 65, 177-187.

Okada A., Okada K., Miyamoto K., Koga J., Shibuya N., Nojiri H. and Yamane H. (2009) OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. J Biol Chem. 284, 26510-26518.

Okubo K., Iijima M., Kobayashi Y., Yoshikoshi M., Uchida T. and Kudou S. (1992) Components responsible for the undesirable taste of soybean seeds. Biosci. Biotechnol. Biochem. 56, 99-103.

Oleszek W., Jurzysta M., Ploszynski M., Colquhoun I.J., Price K.R. and Fenwick G.R. (1992) Zahnic acid tridesmoside and other dominant saponins from alfalfa (*Medicago sativa* L.) aerial parts. J. Ag. Food Chemistry 40, 191-196.

Oliveira F.A., Chaves M.H., Almeida F.R.C., Lima R.C.P., Silva R.M., Maia J.L., Brito G.A.A.C., Santos F.A. and Rao V.S. (2005) Protective effect of α- and β-amyrin, a triterpene mixture from *Protium heptaphyllum* (Aubl.) March trunk wood resin, against acetaminophen-induced liver injury in mice. Ethnopharmacol. 98, 103-108.

Osbourn A. E., Clarke B. R., Lunness P., Scorrt P. R. and Daniels M. J. (1994) An oat species lacking avenacin is susceptible to infection by *Gaeumannomyces graminis var. tritici*. Physiological and Molecular Plant Pathology 45, 457-467.

Osbourn A.E. (1996) Preformed antimicrobial compounds and plant defense against fungal attack. Plant Cell 8, 1821-1831.

Osbourn A.E. (2003) Saponins in cereals. Phytochemistry 62, 1–4.

Osbourn A.E., Qi X., Townsend B. and Qin B. (2003) Dissecting plant secondary metabolism – constitutive chemical defences in Cereals. New Phytologist 159, 101–108.

Osbourn A.E. and Field B. (2009) Operons. Cell. Mol. Life Sci. 66, 3755–3775.

Osbourn A. (2010a) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends Genet 26, 449-457.

Osbourn A. (2010b) Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. Plant Physiol 154, 531-535.

Osbourn A., Goss R.J.M. and Field R.A. (2011) The saponins – polar isoprenoids with important and diverse biological activities. Nat. Prod. Rep. 28, 1261-1268.

Osmani S.A., Bak S., Imberty A., Olsen C.A. and Lindberg Moller B. (2008) Catalytic key amino acids and UDP sugar donor specificity of a plant glucuronosyl transferase, UGT94B1. Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. Plant Physiol. 148, 1295–1308.

Otomo K., Kanno Y., Motegi A., Kenmoku H., Yamane H., Mitsuhashi W., Oikawa H., Toshima H., Itoh H., Matsuoka M., Sassa T. and Toyomasu T. (2004) Diterpene cyclases responsible for the biosynthesis of phytoalexins, momilactones A, B, and Oryzalexins A –F in rice. Biosci. Biotechnol. Biochem. 68, 2001-2006.

Otomo K., Kenmoku H., Oikawa H., Konig W.A., Toshima H., Mitsuhashi W., Yamane H., Sassa T. and Toyomasu T. (2004) Biological functions of ent- and syn-copalyl diphosphate synthases in rice: key enzymes for the branch point of gibberellin and phytoalexin biosynthesis. Plant J. 39, 886-893.

Ourisson G., Rohmer M. and Poralla K. (1987) Microbial lipids betrayed by the fossils. Microbiol. Sci. 4, 52-57.

Owatworakit A, Townsend B., Louveau T., Jenner H., Rejzek M., Hughes R.K., Saalbach G., Qi X., Bakht S., Roy A.D., Mugford S.T., Goss R.J., Field R.A. and Osbourn A. (2012) Glycosyltransferases from oat (Avena) implicated in the acylation of avenacins. J. Biol. Chem. 288, 3696-3704.

Palmer J.M. and Keller N.P. (2010) Secondary metabolism in fungi: does chromosomal location matters? Curr. Op. in Microbiol. 13, 431-436.

Pang, Q., Guo, J., Chen, S., Chen, Y., Zhang, L., Fei, M., Jin, S., Li, M., Wang, Y. and Yan, X. (2012) Effect of salt treatment on the glucosinolate-myrosinase system in *Thellungiella salsuginea*. Plant Soil 355, 363–374.

Panicot M, Minguet EG, Ferrando A, Alcazar R, Blazquez MA, Carbonell J, Altabella T, Koncz C, Tiburcio AF (2002) A polyamine metabolon involving aminopropyl transferase complexes in *Arabidopsis.* Plant Cell 14, 2539-2551.

Papadopoulou K, Roussis A, Katinakis P. (1996) Phaseolus *ENOD40* is involved in symbiotic and non-symbiotic organogenetic processes: expression during nodule and lateral root development. Plant Molecular Biology 30: 403–417.

Papadopoulou K.K., Melton R.E., Leggett M., Daniels M.J., and Osbourn A.E. (1999) Compromised disease resistance in saponin-deficient plants. PNAS 96, 12923–12928.

Paquette S.M. Bak S. and Feyereisen R. (2000) Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes *of Arabidopsis thaliana*. DNA Cell Biol. 19: 307-317.

Paquette S., Moller B.L. and Bak S. (2003) On the origin of family 1 plant glycosyltransferases. Phytochemistry 62, 399–413.

Park H.Y., Park S.H., Yoon H.K., Han M.J. and Kim D.H. (2004) Anti-allergic activity of 18beta-glycyrrhetinic acid-3-O-beta-D-glucuronide. Arch. Pharm. Res. 27, 57-60.

Pasi S., Aligiannis N., Pratsinis H., Skaltsounis A.L. and Chinou I.B. (2009) Biologically active triterpenoids from *Cephalaria ambrosioides,* Planta Med 75, 163-167.

Pathak A., Singh S.K., Biabani M.A. et al (2002) Synthesis of combinatorial libraries based on terpenoid scaffolds. Comb. Chem. High Throughput Screen 5, 241-248.

Pauwels L., Inze D. and Goossens A. (2009) Jasmonate-inducible gene: what does it mean? Trends in Plant Science 1, 87–91.

Paxton J.D. (1981) Phytoalexins – a working definition. Phytopathol. Z. 101, 106-109.

Perlova O., Gerth K., Kaiser O., Hans A. and Muller R. (2006) Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. Journal of Biotechnology 121, 174–191.

Perrin R.M., Fedorova N.D., Bok J.W., Cramer R.A., Wortman J.R., Kim H.S., Nierman W.C. and Keller N.P. (2007) Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. PloS Pathog. 3, e50.

Perry J.A., Wang T.L., Welham T.J., Gardner S., Pike J.M., Yoshida S. and Parniske M.A. (2003) TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus.* Plant Physiol. 131, 866-871.

Perry J., Brachmann A., Welham T., Binder A., Charpentier M., Groth M., Haage K., Markmann K., Wang T.L. and Parniske M. (2009) TILLING in *Lotus japonicus* identified large allelic series for symbiosis genes and

revealed a bias in functionally defective ethyl methanesulfonate alleles toward glycine replacements. Plant Physiol. 151, 1281-1291.

Peters R.J. (2006) Uncovering the complex metabolic network underlying diterpenoid phytoalexin biosynthesis in rice and other cereal crop plants. Phytochemistry 67, 2307-2317.

Pfaffl M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29(9):e45.

Phillips D.R., Rasbery L.M., Bartel B. and Matsuda S.P.T. (2006) Biosynthetic diversity in plant triterpene cyclization. Current Opinion in Plant Biology 9, 305–314.

Phillipson JD (2001) Phytochemistry and medicinal plants. Phytochemistry 56:237-243.

Pikkard C.S., Haag J.R., Ream T. and Wierzbicki A. (2008) Roles of RNA polymerase IV in gene silencing. Trends Plant Sci. 13, 390-397.

Pollier J., Mosses T. and Goossens A. (2011) Combinational biosynthesis in plants: a (p)review on its potential and future exploitation. Nat. Prod. Rep. 28, 1897-1916.

Pollier J., Morreel K., Geelen D. and Goossens A. (2011) Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. Journal of Natural Products 74, 1462–1476.

Poralla K., Hewelt A., Prestwich G.D., Abe I., Reipen I. and Sprenger G. (1994) A specific amino acid repeat in squalene and oxidosqualene cyclases. TIBS 19, 157-158.

Powles S.B. and Yu Q. (2010) Evolution in Action: Plants Resistant to Herbicides. Annu Rev Plant Biol. 61, 317-347.

Price K.R., Johnson I.T. and Fenwick G.R. (1987) The chemistry and biological significance of saponins in food and feedingstuffs. Crit. Rev. Food Sci. Nutr. 26, 27-135.

Prisic S., Xu M., Wilderman P.R. and Peters R.J. (2004) Rice contains two disparate *ent*-copalyl diphosphate synthases with distinct metabolic functions. Plant Physiol. 136, 4228-4236.

Proctor R.H., Hohn T.M., McCormick S.P. and Desjardins A.E. (1995) Tri6 encodes an unusual zinc finger protein involved in regulation of trichothecene biosynthesis in *Fusarium sporotrichiolides*. Appl. Environ. Microbiol. 61, 1923-1930.

Pryde F.E., Gorham H.C. and Louis E.J. (1997) Chromosome ends: all the same under their caps. Curr. Opin. Genet. Dev. 7, 822–828.

Qi X., Bakht S., Leggett M., Maxwell C., Melton R., and Osbourn A.E. (2004) A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. PNAS 101, 8233–8238.

Qi X., Bakht S., Qin B., Leggett M., Hemmings A., Mellon F., Eagles J., Werck-Reichhart D., Schaller H., Lesot A., Melton R. and Osbourn A.E. (2006) A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. PNAS 103, 18848-18853.

Qin B., Eagles J., Mellon F.A., Mylona P., Peña-Rodriguez L. and Osbourn A.E. (2010) High throughput screening of mutants of oat that are defective in triterpene synthesis. Phytochemistry 71, 1245-1252.

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Gronlund M, Sato S, Nakamura Y, Tabata S, Sandal N et al. (2003) Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature 425: 585–592.

Rahman A., Ahamed A., Amakawa T., Goto N. and Tsurumi S. (2001) Chromosaponin I Specifically Interacts with AUX1 Protein in Regulating the Gravitropic Response of *Arabidopsis* Roots. Plant Physiology 125, 990–1000.

Ramakers C., Ruijter J.M., Deprez R.H. and Moorman A.F. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett. 339, 62-66.

Rasbery J.M., Shan H., LeClair R.J., Norman M., Matsuda S.P.T., and Bartel B. (2007) *Arabidopsis thaliana* Squalene Epoxidase 1 Is Essential for Root and Seed Development. Journal of biological chemistry 282, 17002-17013.

Riaño-Pachón D.M., Ruzicic S., Dreyer I. and Mueller-Roeber B. (2007) PlnTFDB: an integrative plant transcription factor database. *BMC Bioinformatics* **8**, 42-51.

Ro D.K., Paradise E.M., Ouellet M., Fisher K.J., Newman K.L., Ndungu J.M., Ho K.A., Eachus R.A., Ham T.S., Kirby J., Chang M.C., Withers S.T., Shiba Y., Sarpong R. and Keasling J.D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940-943.

Rochfort S. and Panozzo J. (2007) Phytochemicals for health, the role of pulses. Journal of Agricultural and Food Chemistry 55, 7981–7994.

Rohmer M., Knani M., Simonin P., Sutter B. and Sahm H. (1993) Isoprenoid biosynthesis in bacteria — a novel pathway for the early steps leading to isopentenyl diphosphate. Biochem J. 295:, 517-524.

Rohmer M., Seemann M., Horbach S., BringerMeyer S. and Sahm H. (1996) Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway forterpenoid biosynthesis. J. Am. Chem. Soc. 118, 2564-2566.

Rohmer M. (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, Nat, Prod. Rep. 16, 565-574.

Ross J., Li Y., Lim E. and Bowles D.J. (2001) Higher plant glycosyltransferases. Genome Biol. 2, 3004.1–3004.6.

Roth C., Rastogi S., Arvestad L., Dittmar K., Light S., Ekman D. and Liberles D.A. (2007) Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. Journal of Experimental Zoology 308, 58-73.

Roudier, F., Ahmed, I., Bırard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., et al. (2011). Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. EMBO J. 30, 1928–1938.

Ruzicka L. (1953) The isoprene rule and the biogenesis of terpenic compounds. Experientia 9, 357-396.

Sainsbury F., Lavoie P.O., D'Aoust M.A., Vızina L.P. and Lomonossoff G.P. (2008) Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. Plant Biotechnology Journal 6, 82–92.

Sainsbury F, and Lomonossoff GP. (2008) Extremely high-level and rapid transient protein production in plants without the use of viral replication. Plant Physiol. 148(3):1212-1218.

Sakamoto T., Miura K., Itoh H., Tatsumi T., Ueguchi-Tanaka M., Ishiyama K., Kobayashi M., Agrawal G.K., Takeda S., Abe K., Miyao A., Hirochika H., Kitano H., Ashikari M. and Matsuoka M. (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiology 134, 1642-1653.

Saleem M, Afaq F, Adhami VM, Mukhtar H. (2004) Lupeol modulates NF-kappaB and PI3K/Akt pathways and inhibits skin cancer in CD-1 mice. Oncogene 23: 5203–5214.

Saleem M, Kaur S, Kweon MH, Adhami VM, Afaq F, Mukhtar H. (2005a) Lupeol, a fruit and vegetable based triterpene, induces apoptotic death of human pancreatic adenocarcinoma cells via inhibition of Ras signaling pathway. Carcinogenesis 26: 1956–1964.

Saleem M, Kweon MH, Yun JM, Adhami VM, Khan N, Syed DN, Mukhtar H. (2005b) A novel dietary triterpene lupeol induces fasmediated apoptotic death of androgen-sensitive prostate cancer cells and inhibits tumor growth in a xenograft model. Cancer Research 65: 11203–11213.

Saleem M, Maddodi N, Abu Zaid M, Khan N, bin Hafeez B, Asim M, Suh Y, Yun JM, Setaluri V, Mukhtar H. (2008) Lupeol inhibits growth of highly aggressive human metastatic melanoma cells in vitro and in vivo by inducing apoptosis. Clinical Cancer Research 14: 2119–2127.

Saleem M, Murtaza I, Tarapore RS, Suh Y, Adhami VM, Johnson JJ, Siddiqui IA, Khan N, Asim M, Hafeez BB et al. (2009) Lupeol inhibits proliferation of human prostate cancer cells by targeting b-catenin signaling. Carcinogenesis 30: 808–817.

SanMiguel P. and Bennetzen J.L. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. 82, 37-44.

Sato T., Kanai Y. and Hoshino T. (1998) Overexpression of squalene-hopene cyclase by the pET vector in *Escherichia coli* and first identification of tryptophan and aspartic acid residues inside the QW motif as active sites. Biosci. Biotech. Biochem. 62, 407-411.

Sato S., Kaneko T., Nakamura Y., Asamizu E., Kato T. and Tabata S. (2001) Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 Mb regions of the genome. DNA Res. 8, 311-318.

Sato S., Nakamura Y., Kaneko T., Asamizu E., Kato T., Nakao M., Sasamoto S., Watanabe A., Ono A., Kawashima K., Fujishiro T., Katoh M., Kohara M., Kishida Y., Minami C., Nakayama S., Nakazaki N., Shimizu Y., Shinpo S., Takahashi C., Wada T., Yamada M., Ohmido N., Hayashi M., Fukui K., Baba T., Nakamichi T., Mori H., and Tabata S. (2008) Genome Structure of the Legume, *Lotus japonicus*. DNA Research 15, 227–239.

Sato S. and Tabata S. (2006) *Lotus japonicus* as as platform for legume research. Curr. Opin. Plant Biol. 9, 128-132.

Sawai S., Akashi T., Sakurai N., Suzuki H., Shibata D., Shin-ichi Ayabe and Aoki T. (2006a) Plant Lanosterol Synthase: Divergence of the Sterol and Triterpene Biosynthetic Pathways in Eukaryotes. Plant Cell Physiol. 47, 673-677.

Sawai S., Shindo T., Sato S., Kaneko T., Tabata S., Shin-ichi Ayabe and Aoki T. (2006b) Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*. Plant Science 170, 247–257.

Sawai S. and Saito K. (2011) Triterpenoid biosynthesis and engineering in plants. Frontiers in Plant Sci. 2, 1-8.

Sayama T., Ono E., Takagi K., Takada Y., Horikawa M., Nakamoto Y., Hirose A., Sasama H., Ohashi M., Hasegawa H., Terakawa T., Kikuchi A., Kato S., Tatsuzaki N., Tsukamoto C. and Ishimoto M. (2012) The *Sg-1* Glycosyltransferase Locus Regulates Structural Diversity of Triterpenoid Saponins of Soybean. The Plant Cell 24, 2123–2138.

Saze H. and Kakutani T. (2007) Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin remodeling factor DDM1. EMBO J. 26, 3641-3652.

Schindler U., Terzaghi W., Beckmann H., Kadesch T. and Cashmore A.R. (1992) DNA binding site preferences and transcriptional activation properties of the *Arabidopsis* transcription factor GBF1. EMBO J. 11, 1275–1289.

Scholz M., Lipinski M., Leupold M., Luftmann H., Harig L., Ofir R., Fischer R., Prófer D. and Móller K.J. (2009) Methyl jasmonate induced accumulation of kalopanaxsaponin I in *Nigella sativa*. Phytochemistry 70, 517-522.

Schonebeck F. and Schlosser E. (1976) Preformed substances as potential protectants. In Physiological Plant Pathology. R. Heitefuss and P.H. Williams. Eds Springer-Verlag, Berlin, pp 653-678.

Schrick K, Mayer U, Horrichs A, Kuhnt C, Bellini C, Dangl J, Schmidt J,Jurgens G (2000) FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. Genes Dev. 14:1471-1484.

Schuler M.A. and Werck-Reichhart D. (2003) Functional Genomics of P450s. Annu. Rev. Plant Biol. 54: 629-637.

Schult K., Meierhoff K., Paradies S., Toller T., Wolff P. and Westhoff P. (2007) The nuclear-encoded factor HCF173 is involved in the initiation of translation of the psbA mRNA in *Arabidopsis thaliana*. Plant Cell 19, 1329–1346.

Segura M.J.R., Meyer M.M. and Matsuda S.P.T. (2000) *Arabidopsis thaliana* LUP1 Converts Oxidosqualene to Multiple Triterpene Alcohols and a Triterpene Diol. Org. Lett. 2, 2257-2259.

Seigler D. and Price P.W. (1976) Secondary compounds in plants: primary functions. Amer. Natur. 110, 101-105.

Seigler D.S. (1977) Primary roles for secondary compounds. Biochemical Systematics and Ecology 5, 195-199.

Seki H., Ohyama K., Sawai S., Mizutani M., Ohnishi T., Sudo H., Akashi T., Aoki T., Saito K., and Muranaka T. (2008) Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. PNAS 105, 14204–14209.

Seki H., Sawai S., Ohyama K., Mizutani M., Ohnishi T., Sudo H., Fukushima E.O., Akashi T., Aoki T., Saito K. and Muranaka T. (2011) Triterpene Functional Genomics in Licorice for Identification of CYP72A154 Involved in the Biosynthesis of Glycyrrhizin. The Plant Cell 23, 4112–4123.

Selmar D. and Kleinwächter M. (2013) Stress enhances the synthesis of secondary plant products: the impact of stress-related over-reduction on the accumulation of natural products. Plant Cell Physiol. 54(6):817-26.

Seo M., Hanada A., Kuwahara A., Endo A., Okamoto M., et al. (2006) Regulation of hormone metabolism in *Arabidopsis* seeds: phytochrome regulation of abscisic acid metabolism and abscisic acid regulation of gibberellin metabolism. Plant J. 48, 354–366.

Shaaban M., Palmer J.M., El-Naggar W.A., El-Sokkary M.A., Habib el. S.E. and Keller N.P. (2010) Involvement of transposon-like elements in penicillin gene cluster regulation. Fungal Genet. Biol. 47, 423-432.

Shai L., McGaw L., Aderogba M., Mdee L. and Eloff J. (2008) Four pentacyclic triterpenoids with antifungal and antibacterial activity from *Curtisia dentata* (Burm.f) CA Sm leaves. J. Ethnopharmacol. 119, 238-244.

Shanmugam M.K., Nguyen A.H., Kumar A.P., Tan B.K. and Sethi, G. (2012). Targeted inhibition of tumor proliferation, survival, and metastasis by pentacyclic triterpenoids: potential role in prevention and therapy of cancer. Cancer Lett. 320, 158-170.

Shi Z., Buntel C.J., and Griffin J.H. (1994) Isolation and characterization of the gene encoding 2, 3 – oxidosqualene lanosterol cyclase from *Saccharomyces cerevisae*. PNAs 91, 7370-7374.

Shibata S. (2001) A drug over the millennia: pharmacognosy, chemistry and pharmacology of licorice. Yakagaku Zasshi 120, 848-862.

Shibuya M., Zhang H., Endo A., Kaori Shishikura K., Kushiro T. and Ebizuka Y. (1999) Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclases. Eur. J. Biochem. 266, 302-307.

Shibuya M., Adachi S. and Ebizuka Y. (2004) Cucurbitadienol synthase, the first committed enzyme for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol biosynthesis. Tetrahedron 60, 6995–7003.

Shibuya M., Hoshino M., Katsube Y., Hayashi H., Kushiro T. and Ebizuka Y. (2006) Identification of β-amyrin and sophoradiol 24-hydrolase by expressed sequence tag mining and functional expression assay. FEBS J. 273, 948-959.

Shibuya M., Xiang T., Katsube Y., Otsuka M., Zhang H. and Ebizuka Y. (2007) Origin of structural diversity in natural triterpenes: direct synthesis of seco-triterpene skeletons by oxidosqualene cyclase, J. Am. Chem. Soc. 129, 1450-1455.

Shibuya M., Katsube Y., Otsuka M., Zhang H., Tansakul P., Xiang T. and Ebizuka Y. (2009) Identification of a product specific β-amyrin synthase from *Arabidopsis thaliana*. Plant Physiology and Biochemistry 47, 26–30.

Shibuya M., Nishimura K., Yasuyama N. and Ebizuka Y. (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. FEBS Lett. 584, 2258-2264.

Shimizu K., Hicks J.K., Huang T.P. and Keller N.P. (2003) Pka, Ras and RGS protein interactions regulate activity of AflR, a Zn(II)2Cys6 transcription factor in *Aspergillus nidulans*. Genetics 165, 1095-1104.

Shimomura K, Nomura M, Tajima S, Kouchi H. (2006) LjnsRING, a novel RING finger protein, is required for symbiotic interactions between *Mesorhizobium loti* and *Lotus japonicus*. Plant Cell Physiol. 47(11):1572-1581.

Shimura K., Okada A., Okada K., Jikumaru Y., Ko K.W., Toyomasu T., Sassa T., Hasegawa M., Kodama O., Shibuya N., Koga J., Nojiri H. and Yamane H. (2007) Identification of a biosynthetic gene cluster in rice for momilactones. J Biol Chem. 282, 34013-34018.

Smil V. (1999) Nitrogen in crop production: an account of global flows. Global Biogeochem. Cycles 13, 647-642.

Sob S.V.T., Wabo H.K., Tchinda A.T., Tane P., Ngadjui B.T. and Ye Y. (2010) Anthraquinones, sterols, triterpenoids and xanthones from *Cassia obtusifolia*. Biochemical Systematics and Ecology 38, 342–345.

Sparg S.G., Light M.E. and van Staden J. (2004) Biological activities and distribution of plant saponins. Journal of Ethnopharmacology 94, 219–243.

Spiteller P., Glawischnig E., Gierl A. and Steglich W. (2001) Studies on the biosynthesis of 2-hydroxy-1,4-benzoxazin-3-one (HBOA) from 3-hydroxyindolin-2-one in *Zea mays*. Phytochemistry 57, 373-376.

Spitz F., Gonzalez F. and Duboule D.A. (2003) A global control region defines a chromosomal regulatory landscape containing the *HoxD* cluster. Cell 113, 405-417.

Sproul D., Gilbert N. and Bickmore W.A. (2005) The role of chromatin structure in regulating the expression of cluster genes. Nat. Rev. Genet. 6, 775-781.

Stiller J., Martirani L., Tuppale S., Chian R.J., Chiurazzi M. and Gresshoff P.M. (1997) High frequency transformation and regeneration of transgenic plants in the model legume *Lotus japonicus*. J. Exp. Bot. 48, 1357-1365.

Sue M., Nakamura C. and Nomura T. (2011) Dispersed Benzoxazinone Gene cluster: Molecular Characterization and Chromosomal Localization of Glucosyltransferase and Glucosidase Genes in Wheat and Rye. Plant Phiosiology 157, 985-997.

Suzuki H., Achnine L., Xu R., Seiichi P. T. Matsuda and Dixon R.A. (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. The Plant Journal 32, 1033–1048.

Suzuki M., Xiang T., Ohyama K., Seki H., Saito K., Muranaka T., Hayashi H., Katsube Y., Kushiro T., Shibuya M. and Ebizuka Y. (2006) Lanosterol Synthase in Dicotyledonous plants. Plant Cell Physiol. 47, 576-571.

Swaminathan S., Morrone D., Wang Q., D. Fulton D.B. and Peters R.J. (2009) CYP76M7 is an ent-Cassadiene C11-α-Hydrolase Defining a Second Multifunctional Diterpenoid Biosynthetic Gene Cluster in Rice. The Plant Cell 21, 3315-3325.

Sykorova E.,Cartagena J., Horakova M., Fukui K. and Fajkus J. (2003) Characterization of telomere-subtelomere junctions in *Silene latifolia*. Mol. Gen. Genomics 269, 13–20.

Szakiel A., Paczkowski C. and Henry M. (2011) Influence of environmental abiotic factors on the content of saponins in plants. Phytochem. Rev. 10, 471–491.

Szczyglowski K. and Amyot L. (2003) Symbiosis. Inventiveness by recruitment? Plant Physiology 131, 935–940.

Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A., Altmann T., Redei G.P., Nagy F., Schehell J. and Koncz C. (1996) Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in *Arabidopsis*. Cell 85, 171-182.

Tag A., Hicks J.K., Garifullina G., Ake C., Phillips T.D., Beremant M. and Keller N.P. (2000) G-protein signaling mediates differential production of toxic secondary metabolites. Mol. Microbiol. 38, 658-665.

Takahara T., Watanabe A. and Shiraki K. (1994) Effects of glycyrrhizin on hepatitis B surface antigen: A biochemical and morphological study. J. Hepatol. 21, 601-609.

Takahashi Y., Nishida Y., Ishii Y., Ishikawa K. and Asai S. (2005) Monitoring expression of cytochrome P450 genes during postischemic rat liver reperfusion using DNA microarrays. Journal of Pharmacological Sciences 97, 153–156.

Takos A.M., Knudsen C., Lai D., Kannangara R., Mikkelsen L., Motawia M.S., Olsen C.E., Sato S., Tabata S., Jørgensen K., Møller B.L. and Rook F. (2011) Genomic clustering of cyanogenic glucoside biosynthetic genes

aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. Plant Journal 68, 273-286.

Talbert P.B. and Henikoff S. (2006) Spreading of silent chromatin: inaction at a distance. Nat. Rev. Genet. 7, 793-803.

Tamogami S., Mitani M., Kodama O. and Akatsuka T. (1993) Oryzalexin S structure: A new stemarane-type rice plant phytoalexin and its biogenesis. Tetrahedron 49, 2025-2032.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28, 2731–2739.

Tang G., Reinhart B.J., Bartel D.P. and Zamore P.D. (2003) A biochemical framework for RNA silencing in plants. Genes and Development 17, 49–63.

Taniyama T., Nagahama Y., Yoshikawa M. and Kitagawa I. (1988) Saponin and sapogenol XLIII. Acetyl-soyasaponins A4, A5, and A6, new astringent bisdesmosides of soyasapogenol A, from Japanese soybean, the seeds of *Glycine max* Merrill. Chem. Pharm. Bull 36, 2829-2839.

Tansakul P., Shibuya M., Kushiro T. and Ebizuka Y. (2006) Dammarenediol-II synthase, the first dedicated enzyme for ginsenoside biosynthesis, in *Panax ginseng*. FEBS lett. 580, 5143-5149.

Tava A., Oleszek W., Jurzysta M., Berardo N. and Odoardi M. (1993) Alfalfa Saponins and Sapogenins: Isolation and Quantification in Two Different Cultivars. Phytochemical Analysis 4, 269-274.

Tava A. and Avato P. (1996) Chemical and biological activity of triterpene saponins from *Medicago* species. Nat. Prod. Commun. 1, 1159-1180.

Tava A., Scotti C. and Avato P. (2011) Biosynthesis of saponins in the genus *Medicago*. Phytochem. Rev 10, 459-469.

Tholl D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Curr. Opin. In Plant Biol. 9, 297-304.

Tokai T., Koshino H., Takahashi-Ando N., Sato M., Fujimura M. and Kimura M. (2007) Fusarium Tri4 encodes a key multifunctional cytochrome P450 monooxygenase for four consecutive oxygenation steps in trichothecene biosynthesis. Biochemical and Biophysical Research Communications 353, 412–417.

Tolhuis B., Palstra R.J., Splinter E., Grosveld F. and de Laat W. (2002) Looping and interaction between hypersensitive sites in the active β-globin locus. Mol. Cell 10, 1453-1465.

Tompson A., Cooper J. and Ingram L.L. (2006) Distribution of terpenes ih heartwood and sapwood of loblolly pine. Forest Prod. J. 56, 7-8.

Toyomasu T. (2008) Recent advances regarding diterpene cyclase genes in higher plants and fungi. Biosci. Biotechnol. Biochem. 72, 1168-1175.

Trojanowska M.R., Osbourn A.E., Daniels M.J. and Threlfall D.R. (2000) Biosynthesis of avenacins and phytosterols in roots of *Avena sativa* cv. Image. Phytochemistry 54, 153-164.

Trojanowska M.R., Osbourn A.E., Daniels M.J. and Threlfall D.R. (2001) Investigation of avenacin-defcient mutants of *Avena strigosa*. Phytochemistry 56, 121-129.

Trojer P. and Reinberg D. (2007) Facultative heterochromatin: is there a distinctive molecular signature? Mol. Cell 28, 1-13.

Tsichritzis F. and Jakupovic J. (1990) Diterpenes and other constituents from *Relhania* species. Phytochemistry 29, 3173–3187.

Tsurumi S., Takagi T. and Hashimoto T. (1991) A new UV-B-absorbing substance in pea seedlings. Abstracts of Fourth Congress of the European Society for Photobiology, Amsterdam, p. 113.

Tsurumi S., Takagi T. and Hashimoto T. (1992) A γ-pyronyl-triterpenoid saponin from *Pisum sativum*. Phytochemistry 31, 2435-2438.

Tsumuri S. and Wada S. (1995) Chromosaponin I stimulates the elongation of cortical cells in lettuce roots. Plant Cell Physiol. 36, 925-929.

Tsurumi S., Watanabe R. and Kamisaka S. (1996) Chromosaponin I increases mechanical extensibility of lettuce root cell walls. Physiol. Plant 97, 740-746.

Tsurumi S. and Ishizawa K. (1997) Involvement of ethylene in chromosaponin-induced stimulation of growth in lettuce roots. Plant Cell Physiol. 38, 668-675.

Turgeon B.G. and Bushley K.E. (2010) Secondary metabolism. In *Cellular and Molecular Biology of Filamentous Fungi* (Borcovich K. and Eddole D., eds), pp 376-395, American Society of Microbiology.

Udvarti M.K., Tabata S., Panriske M. and Stougaard J. (2005) Lotus japonicus: legume research in the fast lane. Trends in Plant Science 10, 222-228.

Ulubelen A., Topcu G., Lotter H., Wagner H. and Eris C. (1994) Triterpenoids from the aerial parts of *Salvia montbretii*. Phytochemistry 36, 413–415.

Umehara M., Hanada A., Yoshida S., Akiyama K., Arite T., Takeda-Kamiya N., et al. (2008) Inhibition of shoot branching by new terpenoid plant hormones. *Nature* 455, 195–200.

Urbanski D.F., Malolepszy A., Stougaard J. and Andersen S.U. (2012) Genome-wide *LORE1* retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. Plant Journal 69, 731–741.

Vaistij F.E., Louise Jones L., and Baulcombe D.C. (2002) Spreading of RNA Targeting and DNA Methylation in RNA Silencing Requires Transcription of the Target Gene and a Putative RNA-Dependent RNA Polymerase. The Plant Cell 14, 857–867.

Vakoc C.R., Letting D.L., Gheldof N., Sawado T., Bender M.A., Groudine M., Weiss M.J., Dekker J. and Blobel G.A. (2005) Proximity among distant regulatory elements at the β-globin locus requires GATA-1 and FOG-1. Mol. Cell 17, 453-462.

van de Wiel C, Scheres B, Franssen H, van Lierop MJ, van Lammeren A, van Kammen A, Bisseling T. (1990) The early nodulin transcript ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. EMBO Journal 9: 1–7.

Van Engelen F.A., Molthoff J.W., Conner A.J., Nap J.P., Pereira A. and Stiekema W.J. (1995) pBINPLUS: an improved plant transformation vector based on pBIN19. Transgenic Res. 4, 288-290.

VanEtten H.D., Mansfield J.W., Bailey J.A. and Farmer E.E. (1994) Two classes of plant antibiotics: phytoalexins versus "anticipins". Plant Cell 6, 1191-1192.

Vanisree M., Lee C.Y., Lo S.F., Manohar S., Lin C.H. and Tsay S.H. (2004) Studies on the production of some important secondary metabolites from medicinal plants by plant tissue cultures. Bot. Bull. Acad. Sin. 45, 1-22.

Verras M. and Sun Z. (2006) Roles and regulation of wnt signaling and b-catenin in prostate cancer. Cancer Letters 237: 22–32.

Vincken J.P., Heng L., Groot A.D. and Gruppen H. (2007) Saponins, classification and occurrence in the plant kingdom. Phytochemistry 68, 275–297.

Voinnet O., Rivas S., Mestre P. and Baulcombe D. (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant Journal 33, 949–956.

Voinnet O. (2008) Post-transcriptional RNA silencing in plant–microbe interactions: a touch of robustness and versatility. Current Opinion in Plant Biology 11, 464–470.

Von Rad U., Hüttl R., Lottspeich F., Gierl A. and Frey M. (2001) Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. Plant Journal 28, 633-642.

Vranova E, Coman D, Gruissem W. 2012. Structure and dynamics of the isoprenoid pathway network. Molecular Plant 5, 318–333.

Vriet C., Welham T., Brachmann A., Pike M., Pike J., Perry J., Parniske M., Sato S., Tabata S., Smith A.M. and Wang T.L. (2010) A suite of *Lotus japonicus* starch mutants reveals both conserved and novel features of starch metabolism. Plant Physiol. 154, 643-655.

Wada M., Takahashi H., Md. Altaf-Ul-Amin, Nakamura K., Hirai M.Y., Ohta D. and Kanaya S. (2012) Prediction of operon-like gene clusters in the *Arabidopsis thaliana* genome based on co-expression analysis of neighboring genes. Gene 503, 56–64.

Waller G.R. and Yamasaki K. (1996) Saponins used in traditional and modern medicine. Advance in Experimental Medicine and Biology 404, New York, NY: Plenum Press.

Wan X, Hontelez J, Lillo A, Guarnerio C, van de Peut D, Fedorova E, Bisseling T, Franssen H. (2007) *Medicago truncatula* ENOD40-1 and ENOD40-2 are both involved in nodule initiation and bacteroid development. Journal of Experimental Botany 58: 2033–2041.

Wang T.L. Domoney C., Hedley C.L., Casey R. and Grusak M.A. (2003) Can we improve the nutritional quality of legume seeds? Plant Physiol. 131, 886-891.

Wang Y.F., Ni Z.Y., Dong M., Cong B., Shi Q.W., Gu Y.C. and Kiyota H. (2010) Secondary metabolites of plants from the genus saussurea: chemistry and biological activity. Chem. Biodivers. 7, 2623-2659.

Wang Q., Hillwig M.L. and Peters R.J. (2011) CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. The Plant Journal 65, 87-95.

Wang Q., Hillwing M.L., Wu Y. and Peters R.J. (2012) CYP701A8: A Rice *ent*-Kaurene Oxidase Paralog Diverted to More Specialized Diterpenoid Metabolism. Plant Physiology 158, 1418-1425.

Wang X., He C., Moore S.C. and Ausio J. (2001) Effects of histone acetylation on the solubility and folding of the chromatin fiber. J. Biol. Chem. 276, 12764-12768.

Wang C, Zhu H, Jin L, Chen T, Wang L, Kang H, Hong Z, Zhang Z. (2013) Splice variants of the SIP1 transcripts play a role in nodule organogenesis in *Lotus japonicus*. Plant Mol Biol. 82(1-2):97-111.

Wassenegger M., Heimes S., Riedel L. and Sanger H.L. (1994) RNA-directed de novo methylation of genomic sequences in plants. Cell 76, 567–576.

Wegel E., Koumproglou R., Shaw P. and Osbourn A. (2009) Cell Type–Specific Chromatin Decondensation of a Metabolic Gene Cluster in Oats. The Plant Cell 21, 3926–3936.

Werck-Reichhart, D., and Feyereisen, R. (2000) Cytochromes P450: a success story. Genome Biol. 1: reviews3003.

Wesley S.V., Helliwell C.A., Smith N.A., Wang M.B., Rouse D.T., Liu Q., Gooding P.S., Singh S.P., Abbott D., Stoutjesdijk P.A., Robinson S.P., Gleave A.P., Green A.G. and Waterhouse P.M. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J. 27, 581-590.

Whitmer S., Verpoorte R. and Canel C. (1998) Influence of auxins on alkaloid accumulation by a transgenic cell line of *Catharanthus roseus*. Plant Cell, Tissue and Organ Culture 53, 135–141.

Whittaker R.H. and Feeny P.P. (1971) Allelochemics: Chemical interactions between species. Science 171, 757-770.

Wierzbicki A.T., Haag J.R. and Pikaard C.S. (2008) Noncoding transcription by RNA polymerase IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135, 635-648.

Wilderman P.R., Xu M., Jin Y., Coates R.M. and Peters R.J. (2004) Identification of syn-pimara-7, 15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. Plant Physiology 135, 2098-2105.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E., and Mc Ree, D. E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Mol. Cell 5: 121-131.

Wink M. and Mohamed G.I.A. (2003) Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the rbcL gene. Biochemical Systematics and Ecology 31, 897–917.

Winkel-Shirley B. (2001) Flavonoid biosynthesis, A colorful model for genetics, biochemistry, cell biology and biotechnology. Plant Physiol. 26, 485-93.

Winzer T., Gazda V., Zhesi He, Kaminski F., Kern M., Larson T.R., Li Y., Meade F., Teodor R., Vaistij F.E., Walker C., Bowser T.A. and Graham I.A. (2012) A *Papaver somniferum* 10-Gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine. Science 336, 1704-1708.

Woldemichael G.M. and Wink M. (2002) Triterpene glycosides of *Lupinus angustifolius*. Phytochemistry 60, 323-327.

Woldemichael G.M., Montenegro G. and Timmermann B.N. (2003) Triterpenoidal lupin saponins from the Chilean legume *Lupinus oreophilus* Phil. Phytochemistry 63, 853–857.

Wolffe A.P. and Hayes J.J. (1999) Chromatin disruption and modification. Nucl. Acids Res. 27, 711-720.

Wong S. and Wolfe K.H. (2005) Birth of a metabolic gene cluster in yeast by adaptive gene relocation. Nat. Genet. 37, 777-782.

Wopereis J, Pajuelo E, Dazzo FB, Jiang Q, Gresshoff PM, De Bruijn FJ, Stougaard J, Szczyglowski K. (2000) Short root mutant of *Lotus japonicus* with a dramatically altered symbiotic phenotype. Plant Journal 23: 97–114.

Wu S., Haig T., Pratley J., Deidre L. and An M. (2001) Allelochemicals in wheat (*Triticum aestivum* L.): production and exudation of 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one. J. Chem. Ecol. 27, 1691-1700.

Wu Y., Hillwig M.L., Wang Q. and Peters R.J. (2011) Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. FEBS Letters 585, 3446–3451.

Wu C., Kim Y.S., Smith K.M., Li W., Hood H.M., Staben C., Selker E.U., Sachs M.S. and Farman M.L. (2009) Characterization of chromosome ends in the filamentous fungus *Neurospora crassa*. Genetics 181, 1129–1145.

Xiang T., Shibuya M., Katsube Y., Tsutsumi T., Otsuka M., Zhang H., Masuda K. and Ebizuka Y. (2006) A new triterpene synthase from *Arabidopsis thaliana* produces a tricyclic triterpene with two hydroxyl groups, Organic letters 8, 2835-2838.

Xie Z., Ma X. and Gang D.R. (2009) Modules of co-regulated metabolites in turmeric (*Curcuma longa*) rhizome suggest the existence of biosynthetic modules in plant specialized metabolism. Journal of Experimental Botany 60, 87–97.

Xiong Q., Rocco F., Wilson W.K., Xu R., Ceruti M. and Matsuda S.P.T. (2005) Structure and reactivity of the dammarenyl cation: Configurational transmission to triterpene synthesis. J. Org. Chem. 70, 5362-5375.

Xiong Q., Wilson W.K and Matsuda S. (2006) An *Arabidopsis* oxidosqualene cyclases catalyzes idiral skeleton formation by Grob fragmentation. Angew. Chem. Int. Ed. 45, 1285-1288.

Xu Z.J., Nakajima M., Suzuki Y. and Yamaguchi I. (2002) Cloning and characterization of the abscisic acid-specific glucosyltransferase gene from adzuki bean seedlings. Plant Physiol. 129, 1285-1295.

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

Xu W., Bak S., Decker A., Paquette S.M., Feyereisen R., and Galbraith D.W. (2001) Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*. Gene 272, 61-74.

Xu M., Hillwig M.L., Prisic S., Coates R.M. and Peters R.J. (2004) Functional identification of rice syn-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. Plant J. 39, 309-318.

Xue Z., Duan L., Liu D., Guo J., Ge S., Dicks J., O'Maille P., Osbourn A. and Qi X. (2012) Divergent evolution of oxidosqualene cyclases in plants. New Phytologist 193, 1022-1038.

Yahia A., Kevers C., Gaspar T., Chenieux J.C., Rideau M. and Creche J.I. (1998) Cytokinins and ethylene stimulate indole alkaloid accumulation in cell suspension cultures of *Catharanthus roseus* by two distinct mechanisms. Plant Science 133, 9–15.

Yang J. and Stern D.B. (1997) The spinach chloroplast endoribonuclease CSP41 cleaves the 3'-untranslated region of petD mRNA primarily within its terminal stem-loop structure. J. Biol. Chem. 272, 12874-12880.

Yang J.L., Gao L.L. and Zhu P. (2013) Advances in the biosynthesis research of ginsenosides. Yao Xue Xue Bao. 48, 170-178.

Yano R., Kanno Y., Jikumaru Y., Nakabayashi K., Kamiya Y., et al. (2009) CHOTTO1, a Putative Double APETALA2 Repeat Transcription Factor, Is Involved in Abscisic Acid-Mediated Repression of Gibberellin Biosynthesis during Seed Germination in *Arabidopsis*. Plant Physiol. 151, 641–654.

Yekta S., Tabin C.J. and Bartel D.P. (2008) MicroRNAs in the Hox network: an apparent link to posterior prevalence. Nat. Rev. Genet. 9, 789-796.

Yendo A.C.A., de Costa F., Gosmann G. and Fett-Neto A.G. (2010) Production of Plant Bioactive Triterpenoid Saponins: Elicitation Strategies and Target Genes to Improve Yields. Mol Biotechnol. 46, 94-104.

Yokota T., Baba J., Konomi K., Shimazaki Y., Takahashi N. and Furuya M. (1982) Identification of a Triterpenoid Saponin in Etiolated Pea Shoots as Phytochrome Killer. Plant Cell Physiol. 23, 265-271.

Young N.D., Mudge J. and Ellis T.H. (2003) Legume genomes: more than peas in a pod. Curr. Opin. Plant Biol. 6, 199-204.

Yu J.H. and Keller N.P. (2005) Regulation of secondary metabolism in filamentous fungi. Annu. Rev. Phytopathol. 43, 437-458.

Yu K.W., Gao W., Hahn E.J. and Paek K.Y. (2002) Jasmonic acid improves ginsenoside accumulation in adventitious root culture of *Panax ginseng* C.A. Meyer. Biochemical Engineering J. 3596, 1-5.

Yuan, G.F.;, Wang, X.P., Guo, R.F. and Wang, Q.M. (2010) Effect of salt stress on phenolic compounds, glucosinolates, myrosinase and antioxidant activity in radish sprouts. Food Chem. 121, 1014–1019.

Zakian V.A. (1995) Telomeres: beginning to understand the end. Science 270, 1601–1607.

Zenk M.H., EI-Shagi H., Arens H., Stockigt J., Weiler E.W. and Deus B. (1977) Formation of the indole alkaloids serpentine and ajmalicine in cell suspension culture of *Catharanthus roseus*, in: Plant tissue culture and its biotechnological application (Barz W. Reinhard E. Zenk MH, eds) Elsevier. Berlin, 250 p.

Zhang H., Shibuya M., Yokota S. and Ebizuka Y. (2003) Oxidosqualene Cyclases from Cell Suspension Cultures of *Betula platyphylla* var. *japonica*: Molecular Evolution of Oxidosqualene Cyclases in Higher Plants. Biol. Pharm. Bull. 26, 642—650.

Zhang J.D., Xu Z., Cao Y.B., Chen H.S., Yan L., An M.M., Gao P.H., Wang Y., Jia X.M. and Jiang Y.Y. (2006) Antifungal activities and action mechanisms of compounds from *Tribulus terrestris* L, J Ethnopharmacol 103, 76-84.

Zhang W. and Popovich D.G. (2009) Chemical and Biological Characterization of Oleanane Triterpenoids from Soy. Molecules 14, 2959-2975.

Zhang K., Mosch K., Fischle W. and Grewal S.I. (2008) Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat. Struct. Mol. Biol. 15, 381-388.

Zhang X., Clarenz O., Cokus S., Bernatavichute Y.V., Pellegrini M., Goodrich J. and Jacobsen S.E. (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol. 5:e129.

Zhang X. and Smith T.F. (1998) Yeast 'operons'. Microb. Comp. Genomics 3, 133-140.

Zhang, X., Bernatavichute, Y., Cokus, S., Pellegrini, M. and Jacobsen, S. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. Genome Biol. 10, R62.

Zhao J., Davis L.C. and Verpoorte R. (2005) Elicitor signal transduction leading to production of plant secondary metabolites. Biotechnology Advances 23, 283–333.

Zhu, J.K. (2001) Plant salt tolerance. Trends Plant Sci. 6, 66–71.

Zhu, Z., Liang, Z., Han, R. and Wang, X. (2009) Impact of fertilization and drought response in the medicinal herb *Bupleurum chinense* DC.: growth and saikosaponin production. Ind. Crop Prod. 29, 629–663.

Zuest T., Joseph B., Shimizu K.K., Kliebenstein D.L. and Turnbull L.A. (2011) Using knockout mutants to reveal the growth costs of defective traits. Proc. R. Soc. B-Biol. Sci. 278, 2598-2603.

Zwenger S. and Basu C. (2008) Plant terpenoids: applications and future potentials. Biotechnology and Molecular Biology Reviews 3, 001-007.