Title: A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis

Afrodite Krokida[1], Costas Delis[1], Katrin Geisler[2], Constantine Garagounis[1], Daniela Tsikou[1], Luis M. Peña-Rodríguez[3], Dimitra Katsarou[1], Ben Field[2,‡], Anne E. Osbourn[2], Kalliope K. Papadopoulou[1*]

[1] University of Thessaly, Department of Biochemistry and Biotechnology, Ploutonos 26 & Aeolou Str., Larisa 41221, Greece

[2] Department of Metabolic Biology, John Innes Centre, Colney Lane, Norwich, NR4 7UH, UK

[3] Unidad de Biotecnología, Centro de Investigación Científica de Yucatán, Mérida, Yucatán, México

[‡]Present address: Laboratoire de Génétique et de Biophysique des Plantes (LGBP), Unité,Mixte de Recherche 6191 Centre National de la Recherche Scientifique-Commissariat àl'Energie Atomique et aux Energies Alternatives-Université de la Méditerranée, 13009, Marseille, France.

* Corresponding author

Kalliope K. Papadopoulou

University of Thessaly

Department of Biochemistry and Biotechnology

Ploutonos 26 & Aeolou, 41221, Larissa, Greece

Tel: +30 2410565244, email: kalpapad@bio.uth.gr

Total word count: 6639

Introduction: 713

Materials and Methods: 1601

Results: 2369

Discussion: 1843

Ackowledgements: 113

Number of Tables: 2

Number of Figures: 8

Supporting Information: Fig. S1-S8, Table S1

SUMMARY

- Genes for triterpene biosynthetic pathways exist as metabolic gene clusters in oat and *Arabidopsis thaliana* plants. We characterized the presence of an analogous gene cluster in the model legume *Lotus japonicus*.
- In the genomic regions flanking the oxidosqualene cyclase *AMY2* gene, genes for two different classes of cytochrome P450 and a gene predicted to encode a reductase gene were identified. Functional characterization of the cluster genes was pursued by heterologous expression in *Nicotiana benthamiana.* Gene expression pattern was studied under different developmental and environmental conditions. The physiological role of the gene cluster in nodulation and plant development was studied in knockdown experiments.
- A novel triterpene structure, dihydrolupeol, was produced by AMY2. A new plant cytochrome P450, CYP71D353, which catalyses the formation of 20-hydroxybetulinic acid in a sequential three-step oxidation of 20-hydroxylupeol was characterized. The genes within the cluster are highly co-expressed during root and nodule development, in hormone-treated plants and under various environmental stresses. A transcriptional gene silencing mechanism that appears to be involved in the regulation of the cluster genes was also revealed.
- A tightly co-regulated cluster of functionally related genes is involved in legume triterpene biosynthesis, with a possible role in plant development.

Keywords *Lotus japonicus*, triterpene, cytochrome P450, nodule, transcriptional silencing

Introduction

Triterpenes are a major subgroup of the terpene superfamily of plant secondary metabolites (Xu et al., 2004). The formation of the skeleton structure of these compounds is catalysed by oxidosqualene cyclases (OSCs) (Iturbe-Ormaetxe et al., 2003) through the cyclization of 2,3-oxidosqualene. These OSC "signature" enzymes (Osbourn, 2010) give rise to either tetracyclic sterols, catalysed by cycloartenol synthases (CS, E.C 5.4.99.8) and lanosterol synthases (LS, E.C 5.4.99.7), or to triterpenes, catalysed by enzymes such as β-amyrin synthase (β-AS, E.C 5.4.99.B1), α-amyrin synthase (α-AS, E.C 5.4.99) and lupeol synthase (LuS, E.C 5.4.99.B3) (Phillips et al., 2006; Ohyama et al., 2009). Subsequent modifications of the basic backbone of these precursor scaffold molecules are carried out by enzymes such as cytochrome P450s, acyltransferases, glycosyltransferases, and methyltransferases, collectively known as "tailoring" enzymes (Osbourn, 2010).

A number of genes and enzymes for triterpene synthesis have been characterized from plants (Osbourn, 2010; Augustin et al., 2011; Sawai and Saito, 2011). Interestingly, genes for certain triterpene biosynthetic pathways exist as metabolic gene clusters in oat (*Avena strigosa*) *and Arabidopsis thaliana* plants (Papadopoulou et al., 1999; Qi et al., 2004; Qi et al., 2006; Osbourn, 2010; Chu et al., 2011). Over the last few years, metabolic gene clusters for the synthesis of other classes of secondary metabolites have also been identified from diverse plant species. These include the phytocassane and momilactone diterpenes in rice (Wilderman et al., 2004; Shimura et al., 2007; Swaminathan et al., 2009), cyanogenic glucosides in *Lotus japonicus*, cassava (*Manihot esculenta*) and sorghum (*Sorghum bicolor*) (Takos et al., 2011), and recently the alkaloid noscapine in poppy (Winzer et al., 2012).

100 Clearly the occurrence of gene clusters in plants is a recurring phenomenon and, thus,
101 implies some functional significance for this form of genetic organization.

102 Legumes produce a huge variety of secondary metabolites (Wink and
103 Mohamed, 2003). Among these, the triterpenoid saponins have received considerable
104 interest due to their antinutrient and, simultaneously, health-giving properties that can
105 affect the quality of food and forage legumes (Rochfort and Panozzo, 2007).
106 Similarly, the triterpenoid saponin content of the model legume *Medicago truncatula*
107 has been studied in great detail (Huhman and Sumner, 2002; Pollier et al., 2011).
108 Simple triterpene skeletons such as β-amyrin have been detected in the roots of
109 legumes during the establishment of rhizobial and mycorrhizal symbioses (Baisted,
110 1971; Hernandez and Cooke, 1996; Grandmougin-Ferjani et al., 1999; Iturbe-
111 Ormaetxe et al., 2003). Furthermore, we recently demonstrated that the simple
112 triterpene lupeol is involved in negative regulation of early nodulation processes in *L.*
113 *japonicus*. Interestingly, heterologous expression of a β-amyrin synthase from aster in
114 *M. truncatula* resulted in nodulation enhancement (Confalonieri et al., 2009). As
115 regards the biosynthesis of triterpenes in legumes, co- expression analysis has led to
116 the identification of candidate genes involved in the production and modification of
117 triterpenes in *M. truncatula* (Naoumkina et al., 2010). Similarly, in *L. japonicus, in*
118 *silico* analysis has revealed several predicted OSC genes (Sawai et al., 2006). Among
119 these, *AMY1/OSC1* and *AMY2*/*OSC8* have previously been characterized by
120 heterologous expression in yeast; AMY1 was shown to be a β-amyrin synthase while
121 AMY2 was reported to be a mixed-function OSC capable of synthesising both β-
122 amyrin and lupeol (Iturbe-Ormaetxe et al., 2003).

123 In this study, the genomic regions encompassing members of the *OSC* gene
124 family in the model legume *L. japonicus* were investigated for candidate gene clusters

for triterpene biosynthesis. This led us to identify two loci in which copies of the previously characterized *AMY2* gene are flanked by genes encoding candidate tailoring enzymes. We then focused our investigations on one of these candidate clusters in *L. japonicus*, which contained the *AMY2* gene along with genes for two different classes of cytochrome P450 (belonging to the CYP88D and CYP71D families) and a gene predicted to encode a reductase. The expression pattern of the genes within the cluster was studied under different developmental and environmental conditions. Using heterologous *in planta* approaches we then showed that this gene cluster probably represents a new pathway for the biosynthesis of a novel triterpene structure, dihydrolupeol, and its subsequent conversion to 20-hydroxybetulinic acid, catalyzed by the sequential activity of AMY2 and CYP71D353. Knockdown experiments suggested a role for this pathway in plant development. Finally, we unexpectedly identified a transcriptional gene silencing mechanism that appears to be involved in the regulation of the cluster genes.

Materials and Methods

Plant material and growth conditions

L. japonicus (cultivar Gifu B-129 or MG20) plants uninoculated and inoculated with *Mesorhizobium loti* (strain R7A) were grown as described by Delis et al, 2011.

In situ hybridization

Sections (7 μm) of nodules from plants 14 and 28 days post-infection were hybridized with antisense and sense RNA probes, labeled with digoxigenin (DIG)-11-rUTP (ROCHE Mannheim, Germany) according to (Delis et al., 2006). Pairs of gene

specific primers were designed, AMY2isF/AMY2isR, LjCYP88D5isF/LjCYP88D5R, LjCYP71D353isF/LjCYP71D353isR (Supporting Information Table S1) and used for the *in vitro* transcription of RNA probes.

Real time PCR experiments

Real time PCR experiments were conducted as previously described (Delis et al., 2011). Relative transcript levels in different samples for the gene of interest were calculated as a ratio to the ubiquitin (*UBQ)* gene transcripts. Data were analyzed according to (Pfaffl, 2001) and the reactions efficiency have been estimated with LinRegPCR (Ramakers et al., 2003). For all samples a triplicate of PCR reaction was performed for each gene.

Roots, nodules and leaves at different developmental stages of plants grown as described were collected and grounded in liquid nitrogen. Gene specific primers were designed with Beacon designer v 7.01 software for the *AMY2*, *P-450-88D5* and *P-450-71D353* genes AMY2rtF, AMY2rtR, LjCYP88D5rtF, LjCYP88D5R, LjCYP71D353rtF, LjCYP71D353rtR, respectively (Supporting Information Table S1). For the different developmental stages and tissues, total RNA was isolated from organs of 50 plants for each sample using RNeasy extraction Kit (QIAGEN). The experiment was repeated once.

Exogenous hormone and abiotic stress treatments

L. japonicus (cultivar Gifu B-129) seeds were pregerminated and seedlings were grown for 7 days on petri dishes containing MS with 1% sucrose substrate at 22^{o}C in a 16h/8h dark/ light photoperiod. Then the seedlings were transferred on petri dishes supplemented with 10 μM or 25 μM methyl jasmonate (MeJA) (Duchefa, Haarlem,

175 The Netherlands), 0.6 mg L^{-1} benzylaminopurine (BA), 0.15 mg L^{-1} 2,4
176 dichlorophenoxyacetic acid (2, 4 – D) or 150nM paraquat (all from Sigma Chemical
177 Co., St Louis, MO, USA) for 7 days. Control plants were grown on petri dishes
178 containing MS supplemented with the respective amount of ethanol for the MeJA and
179 2,4-D treatments. For salt stress treatments, plants were grown in MS with 1% sucrose
180 supplemented with 25 mM, 50 mM or 75 mM NaCl for 7 days. For the heat/cold
181 stress, seedlings were grown for 11 days on petri dishes containing MS substrate at
182 22^{o}C in a 16h/8h dark/ light photoperiod. Then the seedlings were exposed to 37^{o}C
183 for heat stress and to 4^{o}C for cold stress for 24 hours. Roots from 20-50 seedlings per
184 treatment were collected and subjected to total RNA isolation using the QIAGEN
185 Rneasy Mini Kit (Qiagen, Hilden, Germany) and subjected to Real-time PCR as
186 described above. The experiments were repeated three times.
187
188 **Heterologous expression into *Nicotiana benthamiana* leaves**
189 Plasmids pBinPS1NT, pBinPS2NT and pBinPS2NT2AGFP, containing full-
190 length copies of CPMV RNA-1 (35S-RNA-1), RNA-2 (35S-RNA-2) and RNA-2-
191 GFP (35S-RNA-2-GFP), respectively, in the binary transformation vector pBINPLUS
192 (van Engelen et al., 1995), have been described previously (Liu and Lomonossoff,
193 2002). The creation of deleted versions of RNA-2, based on the vector pN81S2NT
194 containing the complete sequence of RNA-2, have been described before (Canizares
195 et al., 2006). Full-length *AMY2, P450-88D5* and *P450-71D353* genes were obtained
196 by appropriate digests and by using two pairs of specific primers LjCYP88D5Fl-
197 F/LjCYP88D5Fl-R and LjCYP71D353Fl-F/LjCYP71D353Fl-R, respectively,
198 (Supporting Information Table S1) and cloned into pM81-FSC1. The pM81-FSC1
199 derived plasmids were digested with AscI / PacI and the fragments were transferred to

vector pBINPLUS (van Engelen et al., 1995). The derivative plasmids were maintained in *A. tumefaciens* strain LBA4404 and agroinfiltration into *N. benthamiana* leaves was carried out as previously described (Canizares et al., 2006). 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. Leaf tissue was harvested and frozen in liquid nitrogen after six days.

***Agrobacterium rhizogenes* plant transformation**

A polyubiquitin promoter-based binary vector, pUBI-GWS-GFP, which allows for GFP overproduction thus facilitating the detection of transgenic roots generated via the infection of *A. rhizogenes* (Maekawa et al., 2008), was used for silencing of the *AMY2* and *LjCYP88D5* genes. 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 (Supporting Information Table S1). *AMY2* and *CYP88D5* ORF sections named AMY2-2, AMY2-3, LjCYP88D5-*1* and LjCYP88D5-*3* were cloned into the *KpnI* - *XhoI* restriction sites of pENTR4 (Invitrogen) replacing the ccdB gene of the original vector. The four new clones, pENTRY-AMY2.2/2.3 and pENTRY-LjCYP88D5-1/3 were then used in an LR Clonase reaction (Invitrogen) with destination vector pUBI-GWS-GFP in order to create the final expression vectors that were used in a plant binary transformation system. Hairy root transformation of *L. japonicus* (cultivar MG20) utilising *A. rhizogenes* strain LBA 1334 and the binary vectors was performed according to (Martirani et al., 1999). Control plant lines were obtained following the same procedure and *A. rhizogenes* transformed with empty

pUBI-GWS-GFP vector. Following transformation, wild-type roots were removed and only one transgenic root was allowed to grow further to produce a new transformed root system per plant. The transformation procedure was repeated twice. Root tissue from all transformed plants, silenced for the two genes (*AMY2* and *LjCYP88D5*) was retained for RNA extraction and the plants, together with fourteen control plants, were allowed to grow further for another 20 days as described before (Delis et al., 2011). Real time PCR experiments were conducted as described above. Nodule numbers were counted at 20 and 40 days in all transformed plants in comparison to control plants.

***Agrobacterium tumefaciens* stable transformation**

PCR amplicon was produced using cDNA from 14 days old roots as template and a pair of specific primers AMY2-2F and AMY2-2R (Supporting Information Table S1). 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 subcloned into pCambia 1300. The resultant binary vector plasmid was transferred into *A. tumefaciens* strain AGL1 by electroporation. Plants were transformed following procedures previously described by (Lombari et al., 2003; Barbulova and Chiurazzi, 2005). Control plants were transformed with the empty vector. 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. 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 (Supporting Information Table S1), respectively. Total RNA was

249 isolated from the four identified transformed plants using RNeasy extraction Kit
250 (QIAGEN). Real time PCR experiments were conducted as described above.
251
252 **DNA methylation assays-Bisulfite sequencing**
253 Genomic DNA samples from wild type and mutant roots were extracted using
254 the Nucleospin Plant II kit (Macherey-Nagel). Sodium bisulfite treatment of the DNA
255 was conducted using the EpiTect Bisulfite kit (Qiagen), following manufacturer's
256 instructions. The target regions (all in coding sequences, 280-350 bp long) were
257 amplified using the primer pairs described in Supporting Information Table S1.
258 Amplified fragments were TA cloned using the pGEM-T-easy Vector System
259 (Promega). Ten clones were sequenced for each amplicon to determine levels of
260 methylation (percentage of all methylated deoxycytidine 5mdC in relation to the total
261 deoxycytidines in all ten clones in mutant).
262
263 **Metabolite extraction and GC-MS Analysis**
264 *N. benthamiana* leaf material was harvested, ground in liquid nitrogen, and
265 lyophilized. The dry plant tissue (100-250 mg) was saponified in 10% KOH (w/v) in
266 80% EtOH (v/v) with 0.5 mg/ml butylated hydroxytoluene (Sigma-Aldrich, Poole,
267 UK) at 65°C for 2 hours and extracted with hexane as previously described (Field and
268 Osbourn, 2008). After hexane extraction, hydrochloric acid was added to the aqueous
269 solution to lower the pH ~ 2.0 and another round of hexane extractions were
270 performed to obtain an acid extraction fraction. The alkaline and acid hexane extracts
271 were concentrated and derivatised with Tri-Sil Z (Pierce, Cambridge, UK) prior to
272 GC-MS analysis. GC-MS analyses were conducted on an Agilent 5973 MSD
273 (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 was analyzed with the AMDIS software package (http://chemdata.nist.gov/mass-spc/amdis/).

Phylogenetic Analysis

The alignment and phylogenetic analysis of the cytochrome P450s and OSCs protein sequences were performed with MEGA 5.05 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 *L. japonicus* 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), and *L. japonicus* (pseudomolecules v2.5) genome databases with expectation value 0.001. Selected nucleotide sequences was analysed with Gevo algorithm for syntenic genomic regions.

Statistical analysis

All experiments were conducted at least twice and analysed by analysis of variance (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 ($P\leq0.01$).

Results

Analysis of the genomic regions encompassing *OSC* genes in legumes

We carried out *in silico* analysis using the public genome sequence databases for *L. japonicus* to investigate the genomic regions encompassing all predicted *OSC* genes present in the genomes of the model plant species (Sawai et al., 2006). 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/). Sequences with expectation values of $< 1 \times e^{-10}$ were excluded from further analysis. A region of approximately 300 Kb flanking each side of the OSC genes was screened and analysed using FGENESH gene prediction software (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) and GeneScanwebserver (Burge and Karlin, 1998). For all predicted ORFs and translated amino acid sequences (Expasy translation tool http://web.expasy.org/translate/) blast searches were run against *L. japonicus* EST databases (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=l_japonicus) 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).

Two genomic regions were identified, in which genes potentially implicated in triterpenoid biosynthesis are assembled in cluster formation, located on chromosome

3 of *L. japonicus* (Fig. 1a). The first gene cluster consists of genes encoding the previously characterized AMY2 enzyme (Iturbe-Ormaetxe et al., 2003) and two cytochrome P450 enzymes, LjCYP88D5 and LjCYP71D353 (Fig. 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*), is also located in this region, in between *AMY2* and the two cytochrome P450 genes. *LjSDRt* is predicted to encode an enzyme with extensive similarity (82%) to an *A. thaliana* short chain dehydrogenase/reductase (SDR)- like protein, HCF173, (Q9FWQ6, At1G16720) (Fig. 1b). 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* (81% similarity for the coding amino acid sequences) (Fig.1b). 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). 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.

The cytochrome P450 gene identified in the two clusters were subjected to phylogenetic analysis by comparison with cytochrome P450 gene 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) (Fig. 2a, b, Supporting Information S1). LjCYP88D4/5 belong to the Fabaceae-specific CYP88D subfamily (CYP85 clan), which has been implicated in triterpene biosynthesis, and exhibit 71% and 72% similarity respectively with GuCYP88D6, a

licorice β-amyrin 11-oxidase (Seki et al., 2011; 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.

Phylogenetic analysis was also performed for the predicted SDR-like LjSDRt (Fig. 2c, Supporting Information S1). 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.

The AMY2 cluster genes are co-expressed in roots and nodules of L. japonicus and in response to hormone and abiotic stress treatments

Previously we and others have shown that *AMY2* is highly expressed in the root and nodules of *L. japonicus* plants that have been inoculated with the symbiotic bacterium *Mesorhizobium loti* (Iturbe-Ormaetxe et al., 2003; Sawai et al., 2006). We therefore investigated whether the genes that are clustered with *AMY2* are similarly expressed in leaves and roots of *L. japonicus* plants and in *M. loti*-inoculated roots and nodules at various developmental stages (Fig. 3). These experiments revealed that the *AMY2*, *LjCYP88D5* and *LjCYP71D353* genes show similar expression patterns in all tissues and developmental stages examined (Fig. 3, Supporting Information Fig. S2), indicating that the three genes are co-expressed. Thus, gene expression is higher in the roots of 7 day-old seedlings than in the roots of 14- and 28-day old seedlings (Fig. 3; a,c,e), while transcript levels were also detectable in the leaves at all three time points but at much lower levels than in roots. Transcript levels were highest in inoculated with *M. loti* 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 (Fig. 3; b,d,f). Accordingly, co-localization of *AMY2*, *LjCYP88D5* and *LjCYP71D353* gene transcripts was detected by *in situ* hybridization in the vascular bundles, parenchymatic cells (inner parenchyma) and also in uninoculated cells of the central tissue (Supporting Information Fig. S3). Differences in transcript levels for *LjSDRt* were less marked in the different tissues and developmental stages compared and there is no common expression pattern with the three other cluster genes (Supporting Information Fig. S4).

We then investigated the effect of hormone treatment on the expression of the three cluster genes. In plants treated with three different types of hormones, i.e. 2,4 dichlorophenoxyacetic acid (2,4D), benzylaminopurine (BA) and methyl jasmonic acid (MeJA), the transcript levels for all three genes were significantly increased compared to control mock-treated plants (Fig. 4a-c). We extended our investigations to abiotic stess treatments that included heat, cold, oxidative and salt stress. All three cluster genes were co-ordinately down-regulated in response to cold, heat and oxidative stress (Fig. 4d-f) while no significant change was observed in the transcript levels of any of the genes in response to salt stress (Supporting Information Fig. S5). All together these data suggest that the *AMY2*, *LjCYP88D5* and *LjCYP71D353* genes comprise a tightly co-regulated cluster of functionally related genes.

Functional analysis in *Nicotiana benthamiana*

By analogy with previously reported plant metabolic gene clusters, we hypothesized that the two cytochrome P450 enzymes that are encoded by the *AMY2* cluster genes may act on the product of the signature enzyme *AMY2* and therefore participate in the synthesis of triterpene secondary metabolites in *L. japonicus*. The

401 full-length cDNAs of *AMY2, LjCYP88D5* and *LjCYP71D353* were cloned into the
402 pBinP-NS-ER-GFP vector for transient expression in *N. benthamiana* leaves
403 infiltrated with *Agrobacterium tumefaciens* (Canizares et al., 2006). Following
404 expression, metabolites were extracted and analysed by GC/MS. All candidate genes
405 were introduced on their own and also in combination with each other in co-
406 expression experiments.

407 Previously, AMY2 was assigned a multi-functional oxidosqualene cyclase
408 function, since it produces both β-amyrin and lupeol when expressed in yeast (Iturbe-
409 Ormaetxe et al., 2003). In *N. benthamiana* leaves, AMY2 produced β-amyrin (peak 2,
410 Figure 5) as expected and the mass spectrum of peak 2 matched that of the
411 trimethylsilyl (TMS) ether derivative of a β-amyrin standard (Fig. 5b). Surprisingly,
412 lupeol was not detected. Instead, a less polar compound was detected in leaf extracts
413 of AMY2 expressing plants (peak 1, Fig. 5). Analysis of the mass spectrum of the
414 trimethylsilyl (TMS) ether derivative of the less polar compound yielded a parent ion
415 of m/z 500 (Figure 5B), a number of fragment ion peaks characteristic for C-3
416 hydroxy lupanes (e.g. *m/z* 279, 207, 220), and two significant fragment ion peaks at
417 *m/z* 457 and m/z 191, that suggested a saturated lupane structure. The fragment ion
418 peak at m/z 457 is known to result from the loss of a propyl group, following a
419 fragmentation favoured only in saturated lupanes (Budzikiewicz et al., 1964);
420 similarly, the fragment ion peak at m/z 191 corresponds to a fragment having rings D
421 and E of a saturated lupane structure (Budzikiewicz et al., 1964; Dantanarayana et al.,
422 1981). On the basis of these results, the less polar component (peak 1, Fig. 5) detected
423 in the leaf extracts of LjAMY2-expressing plants is proposed to be dihydrolupeol.

424 No new metabolites were observed when LjCYP88D5 or LjCYP71D353 were
425 expressed alone in *N. benthamiana*. However, simultaneous expression of *AMY2* and

426 *LjCYP71D353* did result in the production of two novel metabolites (peak 3-5, Fig.
427 5a). We used basic and acid procedures to ensure extraction of a wide range of
428 possible metabolites. Peaks 3 and 4 correspond to the same product, the TMS-
429 derivative of 3,20-lupandiol (commonly referred to as 20-hydroxy-lupeol); both peaks
430 showed the same fragmentation pattern (Fig. 5b), with a prominent parent ion peak at
431 *m/z* 588 and characteristic fragment ion peaks at m/z 573 (M^+-CH_3), 498 (M^+-TMS-
432 H_2O), and 408 (M^+-2TMS-$2H_2O$) (Cole et al., 1991; Ulubelen et al., 1994). Peak 5,
433 which was only detected in the acidic fraction, was identified as 3,20-dihydroxy-28-
434 lupanoic acid (commonly referred to as 20-hydroxy-betulinic acid) after a detailed
435 analysis of the mass spectrum of its TMS derivative, which showed fragment ion
436 peaks at *m/z* 619, 513, and 408; these fragment ion peaks can be explained by the loss
437 of a TMS protecting group following a McLafferty-type rearrangement from a
438 protonated parent ion peak, the combined loss of a second TMS protecting group, a
439 molecule of water and a methyl group, and by a fragment having the fully-substituted
440 rings D and E, respectively (Supporting Information Fig. S6) (Budzikiewicz et al.,
441 1964; Tsichritzis & Jakupovic, 1990). 20-hydroxylupeol and 20-hydroxybetulinic
442 acid were only detected when *LjCYP71D353* was expressed together with *AMY2*. This
443 indicates that LjCYP71D353 catalyses the formation of 20-hydroxylupeol from
444 dihydrolupeol in a single oxidation reaction. In addition, LjCYP71D353 catalyses the
445 formation of 20-hydroxybetulinic acid in a three-step sequential oxidation at the C-28
446 position of 20-hydroxylupeol (Fig. 6). To further verify that LjCYP71D353 acts on
447 dihydrolupeol and not on the β-amyrin produced by AMY2, we co-expressed
448 LjCYP71D353 together with the oat β-amyrin synthase gene AsbAS1, which
449 produces only β-amyrin. No new products were detected. Similarly, β-amyrin was not
450 recognized as a substrate of LjCYP88D5 (Supporting Information Fig. S7). No further

products were detected when AMY2, LjCYP71D353 and LjCYP88D5 were co-expressed together.

Physiological significance of the AMY2 secondary metabolic gene cluster

Previously we have shown by gene silencing that lupeol has a role in suppression of nodule formation in *L. japonicus* (Delis et al., 2011). Others have 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. We employed a hairpin RNA gene silencing strategy to investigate a putative role of the cluster genes in nodulation. Two silencing constructs were made for the *AMY2* gene and transgenic roots were generated using an *A. rhizogenes* transformation protocol. 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 Fig. 7a). 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 severe retardation of the rate of hairy-root growth was recorded as compared to 20dpi. We next produced, by using two different hairpin constructs, transgenic roots with reduced levels of the *LjCYP88D5* expression (Fig. 7e). We found that silencing of *LjCYP88D5* caused no obvious effects on plant development or nodulation.

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. Four transformed plant lines with reduced levels of *AMY2* transcript were obtained (Fig. 8a). These plants did not flower (and so we were unable to obtain seed) and consistently exhibited a short, stunted root phenotype, indicating that

476 silencing of *AMY2* in stably transformed plants has clear effects on growth and
477 development.
478
479 **_AMY2_ cluster genes are epigenetically regulated**
480 Based on our results indicating strong co-ordinated regulation of the
481 expression profile of the cluster genes, we investigated their expression profile in the
482 silenced plant lines. Strikingly, we observed that the co-ordinate regulation of the
483 cluster genes was manifested in a very unusual fashion. Thus, a decrease in *AMY2*
484 transcription levels due to *A. rhizogenes*-mediated gene silencing, in all plant lines
485 examined, was accompanied by significant reductions in the transcript levels of other
486 cluster genes, namely *LjCYP88D5, LjCYP71D353* and *LjSDRt* (Fig. 7b, c, and d
487 respectively). Similarly, plant lines silenced for *LjCYP88D5* had reduced transcript
488 levels for *AMY2* (Fig. 7f), *LjCYP71D353* (Fig. 7g) and *LjSDRt* (Fig. 7h). These
489 experiments were repeated for both genes over a period of more than two years and
490 we always observed such a "spreading" of transcription silencing in the cluster.
491 Furthermore, we also observed the *cis* spreading of transcript level repression in
492 *AMY2* silenced stably transformed plants (Fig. 8b-d). The transcript levels of *LjSDRt*
493 and *LjCYP71D353* (but not of *LjCYP88D5* in this case) were both significantly
494 reduced in lines amy2-3 and amy2-4, while those of *LjSDRt* were also reduced in
495 lines amy2-1 and amy2-2. To further investigate this phenomenon we used bisulfite
496 sequencing to determine the DNA methylation levels of *LjSDRt* and *LjCYP71D353*
497 genes in the wild type and in two *AMY2* silenced plants, amy2-3 and amy2-4. We
498 detected a significant increase in the degree of DNA methylation in both of the genes
499 with reduced expression levels in the silenced plant lines compared to the methylation
500 level of the genes in wild type plants (Table 2). 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) (Wessenegger et al, 1994; Dalakouras and Wessenegger, 2013), thus promoting transcriptional gene silencing (TGS). The spreading of RdDM into the adjacent regions by TGS transitivity would silence the adjacent genes.

DISCUSSION

Triterpene gene clusters in legumes

In this paper we have mined the genome sequences of the model legume *L. japonicus* for triterpene biosynthesis gene clusters. We functionally validated the *AMY2* gene cluster by the characterization and analysis of three of the genes comprising the cluster. A second *L. japonicus* cluster consists of a corrupted *AMY2*-like gene, flanked by a single cytochrome P450 gene (*LjCYP88D4*). Due to the corrupted AMY2 gene and the absence of a *LjCYP71D353* homologue, we propose that this second cluster is either incomplete or in decay. Furthermore, the low degree of conservation in genome structure between the two regions suggests that the ancestral *AMY2* 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. *LjCYP71D353* is phylogenetically related to *CYP71A16* that forms part of the marneral gene cluster in *A. thaliana* (Field et al., 2011). However, the low degree of synteny (Supporting

526 Information Fig. S8) and the phylogenetic distance between the *OSC* genes in these
527 two clusters (*MRN1*, in the marneral cluster, and *AMY2* in the *L. japonicus* cluster)
528 suggests that *LjCYP71D353* and *AMY2* are not derived from the tandem duplication
529 of a common ancestral pair of *OSC* and cytochrome P450 genes and do not share a
530 common origin with the marneral cluster. Thus, the gene clusters identified in *L.*
531 *japonicus* further support the theory that specialized metabolic gene clusters in plants
532 are likely to have arisen *de novo* within recent evolutionary history (Chu et al., 2011).
533
534 **Co-ordinate regulation of cluster genes**
535 Consistent with previous findings for the co-ordinated regulation of cluster
536 activity (Osbourn and Field, 2009), the *AMY2* cluster genes are co-ordinately
537 expressed in response to developmental and environmental cues. Of the four genes
538 comprising the *AMY2* gene cluster, three were co-ordinately expressed and induced by
539 abiotic stresses and rhizobium infection while one (*LjSDRt*) was not. The genes for
540 several other plant metabolic gene clusters have also been reported to be induced by
541 biotic or abiotic treatments. For example, some (but not all) genes of the momilactone
542 diterpene gene cluster in rice are up-regulated in response to treatment with a chitin
543 oligosaccharide elicitor (Swaminathan et al., 2009). The *AMY2* cluster genes are also
544 developmentally co-regulated and their expression is co-ordinately induced by plant
545 growth hormones. MeJA is a well- known elicitor for the production of many plant
546 secondary metabolites, typically manifested when plants are under environmental
547 stresses (Zhao et al., 2005). Regarding the induction of triterpene biosynthesis, MeJA
548 induces the expression of the *M. truncatula bAS* gene (Suzuki et al., 2002). It is
549 expected that, through the hormonal "crosstalk" networks that operate in plants, other
550 hormones are likely to participate in the amplitude and specificity of such elicitations

551 (Pauwels et al., 2009). Auxin and cytokinins have been reported to alter alkaloid
552 content when applied exogenously to *Cantharanthus roseus* cell cultures (Whitmer et
553 al, 1998; Yahia et al, 1998). To our knowledge, auxin and cytokinin have not been
554 previously reported to have functions in inducing triterpene production.
555
556 **Mechanisms of gene cluster regulation**
557 Very little is known about the mechanisms that control regulation of plant
558 metabolic gene clusters. Common upstream *cis* elements that are recognized by a
559 transcription factor required for regulation of diterpene biosynthesis in rice have been
560 identified (Okada et al., 2009), but this transcription factor appears to be a global
561 regulator of diterpenes and is not specific for the two characterized rice diterpene
562 clusters. Transcriptional regulators for other plant gene clusters have not yet been
563 defined. Additionally, the cluster for noscapine biosynthesis in poppy is co-ordinately
564 regulated with regard to the homozygous/heterozygous state of the plant genotype,
565 indicating an additional level of regulation beyond the transcriptional level (Winzer et
566 al., 2012). There is also evidence that co-ordinate expression of plant secondary
567 metabolic gene clusters is likely to be regulated at the level of chromatin
568 modification/remodelling in oat and *A. thaliana* (Field and Osbourn, 2008; Wegel et
569 al., 2009; Field et al., 2011). Nevertheless, systematic analysis of the type of
570 chromatin modifications that are important for the expression of plant secondary
571 metabolic gene clusters has not yet been reported. We report here such a modification,
572 i.e. DNA methylation that appears to play a role in the *AMY2* gene cluster regulation. .
573 We observed a "global-silencing" phenomenon, in which hairpin-mediated gene
574 silencing of any of the genes in the cluster induced DNA methylation and, thus,
575 repression of gene expression in the adjacent genes within the cluster. It should be

noted, though, that our conditions involved a transgenic RNAi –mediating silencing procedure which may not directly represent a natural regulatory event.

During post-transcriptional gene silencing, spreading of RdDM has been reported and is associated with the production of ~21-22 nt secondary siRNAs that require RDR6 (Vaistij et al., 2002; Eamens et al., 2008; Voinnet, 2008). 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 and Jacobsen, 2008; Daxinger et al., 2009). It has been suggested that locus-specific effects, may account for the mechanism that underlies RdDM, which is initiated using hairpin constructs (Daxinger et al., 2009). Perhaps the chromatin condensation state could be responsible for the 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). 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.

Function of cluster genes

Expression of the *L. japonicus AMY2* gene in *N. benthamiana* revealed that in addition to β-amyrin, AMY2 also produced dihydrolupeol. It is possible that non-specific endogenous enzymatic activity in *N. benthamiana* results in the formation of the dihydrolupeol as compared to the formation of lupeol in yeast cells by AMY2. 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. Further characterization of the enzyme kinetics and action, as well as the production of a dihydrolupeol standard, which proved challenging as yet, will allow the confirmation of the molecule identity. Our attempts to detect dihydrolupeol in MeJA induced *L. japonicus* roots were not successful; this could be due to conversion to downstream products.

Not many members of the CYP71D subfamily have been functionally characterized but the involvement of certain members in the biosynthesis of terpenoids has been reported before (Lupien et al., 1999; Takahashi et al., 2005). LjCYP71D353 represents a novel cytochrome P450 enzyme acting on an unusual triterpene skeleton, i.e. 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* needs investigation. Successive hydroxylation/oxidation reactions catalyzed by plant cytochrome P450 enzymes towards other substrates have been documented (Bak et al., 1997; 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., 20011). 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 catalysing 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 the avenacin gene cluster of oat, CYP51H10 catalyses

626 epoxidation of β-amyrin at C-12/C-13 and also hydroxylation at C-16 (Geisler &
627 Osbourn, unpublished results). This enzyme is an addition to the arsenal of
628 biosynthetic enzymes needed for the production of novel triterpenes by synthetic
629 biology approaches.

630 The cytochrome P450 genes *LjCYP88D4/5* within the *AMY2* cluster belong to
631 the legume-specific CYP88D subfamily, members of which have previously been
632 characterized and assigned roles in triterpene biosynthesis. Specifically, GuCYP88D6
633 (Seki et al., 2008) exhibiting 72% similarity with LjCYP88D5 is a β-amyrin 11-
634 oxidase. Unexpectedly, we did not detect any activity towards β-amyrin when both
635 enzymes are expressed in *N. benthamiana* with *AMY2*. This could be attributed to the
636 different heterologous system employed. Alternatively LjCYP88D5 may have a
637 different function than GuCYP88D6 in *Glycyrrhiza uralensis*.

638 The fourth gene identified in the *AMY2* gene cluster has extensive similarity to a
639 SDR-like protein that has been identified and characterized in *A.thaliana,* namely
640 HCF173 (Schult et al., 2007). HCF173 has been predicted to have lost its
641 dehydrogenease activity and is reported as an RNA binding protein, with a regulatory
642 role in the translational activity of the mRNA (*pbsA)* that directly interacts with as
643 part of a high molecular weight complex. The high similarity between LjSDRt and
644 HCF173 suggests that LjSDRt may also represent a regulatory protein, having lost a
645 metabolic function. However, no other homologs to HCF173 in other plants have
646 been characterized as yet to allow for a justified functional prediction for LjSDRt.

647 In our functional analysis using gene silencing, we were able to detect a root
648 growth defect in the plants silenced for *AMY2.* Our investigation for a phenotype was
649 not exhaustive and we have only looked macroscopically for defects in plant growth
650 and nodulation. The recent generation of *LORE*1 reverse genetic resource for *L.*

651 *japonicus* (Ubranski et al, 2012), in which mutant lines for the cluster genes have
652 already been identified, offer a new opportunity to further characterize the role of the
653 cluster genes in the plant development and physiology as well as validate the
654 functional relation amongst them.

655 Most of the plant metabolic gene clusters discovered so far are for synthesis of
656 compounds that are implicated in innate and induced disease resistance, insect
657 resistance, abiotic stress tolerance and/or allelopathy (Chu et al., 2011; Kliebenstein
658 and Osbourn, 2012). A number of apparently opposed activities have been reported
659 for the role of triterpenes in plant developmental processes. For example, in the
660 avenacin cluster the accumulation of late triterpene pathway intermediates can result
661 in deleterious effects on plant growth (Mylona et al., 2008); elevated accumulation of
662 the triterpene thalianol in *A. thaliana* results in enhanced root length (Field and
663 Osbourn, 2008; Mylona et al., 2008; Field et al., 2011); simple triterpenes, like lupeol,
664 act as negative regulators of nodule formation, and thus cell proliferation (Delis et al.,
665 2011). These observations open up further questions about whether growth
666 inhibition/promotion mediated by triterpenes occurs via different pathways or through
667 antagonistic effects on a common pathway (Osbourn et al., 2011). The inducibility of
668 the *AMY2* gene cluster provides the opportunity to further pursue and define such a
669 regulatory role, as yet unidentified, in plant growth and development processes.
670
671
672 **Acknowledgements:** We would like to thank George Lomonossoff for the pM81-
673 FSC1 vector, Thomas Ott for the pUBI-GWS-GFP vector and P. Waterhouse for the
674 pHANNIBAL vector. This work has been supported by the ESF/NSRF programme
675 Heracleitus II (to AK and KKP), the Postgraduate Programmes "Biotechnology:

Quality of Nutrition and the Environment" and "Applications of Molecular Biology-Genetics, Diagnostic Biomarkers" (DBB) (to KKP), a joint Engineering and Physical Sciences Research Council/National Science Foundation award to AEO as part of the Syntegron consortium (EP/H019154/1) NSF/EPRSC, the BBSRC Institute Strategic Programme Grant "Understanding and Exploiting Plant and Microbial Secondary Metabolism" (BB/J004596/1 to AEO) and the John Innes Foundation (AEO). KG was supported by a Danish Research Agency International Studentship.

701 **References**

702 **Altschul SF, Boguski MS, Gish W, Wootton JC. 1994.** Issues in searching molecular
703 sequence databases. *Nature Genetics* **6:**119-129.

704 **Anzai Y, Li SY, Chaulagain MR, Kinoshita K, Kato F, Montgomery J, Sherman DH.**
705 **2008.** Functional analysis of MycCI and MycG, cytochrome P450 enzymes involved in
706 biosynthesis of mycinamicin macrolide antibiotics. *Chemistry and Biology* **15**:950-959.

707 **Augustin JM, Kuzina V, Andersen SB, Bak S. 2011.**Molecular activities, biosynthesis and
708 evolution of triterpenoid saponins. *Phytochemistry* **72**: 435-457.

709 **Baisted DJ. 1971.** Sterol and triterpene synthesis in developing and germinating pea seed.
710 *Biochemical Journal* **124**: 375-383.

711 **Bak S, Kahn RA, Olsen CE, Halkier BA. 1997.** Cloning and expression in Escherichia coli
712 of the obtusifoliol 14 alpha-demethylase of *Sorghum bicolor* (L.) Moench, a cytochrome
713 P450 orthologous to the sterol 14 alpha-demethylases (CYP51) from fungi and mammals.
714 *Plant Journal* **11**: 191-201.

715 **Barbulova A, Chiurazzi M 2005.** A procedure for in vitro nodulation studies. *Lotus*
716 *japonicus Handbook*: 83-86.

717 **Budzikiewicz H, Djerassi C, Williams DH. 1964.** *Structure elucidation of natural products*
718 *by mass spectrometry, v. II, Steroids, terpenoids, sugars, and miscellaneous classes*. San
719 Francisco: HoldVL - Day.

720 **Burge CB, Karlin S. 1998.** Finding the genes in genomic DNA. *Current Opinion in*
721 *Structural Biology* **8**: 346-354.

722 **Canizares MC, Liu L, Perrin Y, Tsakiris E, Lomonossoff GP. 2006.** A bipartite system
723 for the constitutive and inducible expression of high levels of foreign proteins in plants. *Plant*
724 *Biotechnology Journal* **4**: 183-193.

725 **Carlson JC, Fortman JL, Anzai Y, Li SY, Burr DA, Sherman DH. 2010.** Identification of

726 the tirandamycin biosynthetic gene cluster from Streptomyces sp 307-9. ChemBioChem **11:**
727 564-572.**Chu HY, Wegel E, Osbourn A. 2011.** From hormones to secondary metabolism:
728 the emergence of metabolic gene clusters in plants. *Plant Journal* **66**: 66-79.
729 **Cole BJW, Bentley MD, Hua Y, Bu L. 1991.** Triterpenoid constituents in the outer bark of
730 *Betula alleghaniensis* (Yellow Birch). *Journal of Wood Chemistry and Technology* **11**: 209-
731 223.
732 **Confalonieri M, Cammareri M, Biazzi E, Pecchia P, Fevereiro MP, Balestrazzi A, Tava**
733 **A, Conicella C. 2009.** Enhanced triterpene saponin biosynthesis and root nodulation in
734 transgenic barrel medic (*Medicago truncatula* Gaertn.) expressing a novel beta-amyrin
735 synthase (AsOXA1) gene. *Plant Biotechnology Journal* **7**: 172-182.
736 **Dalakouras A, Wassenegger M. 2013.** Revisiting RNA-directed DNA methylation. *RNA*
737 *biology* **10:** 453-455.
738 **Dantanarayana AP, Kumar NS, Sultanbawa MUS, Balasubramaniam S. 1981.**
739 Structures of four new oxygenated lupanes from *Pleurostyliaopposita*(Celastraceae). *Journal*
740 *of the Chemical Society, Perkin Transactions* **10**: 2717-2723.
741 **Daxinger L, Kanno T, Bucher E, van der Winden J, Naumann U, Matzke AJ, Matzke**
742 **M. 2009.** A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of
743 DNA methylation. *EMBO Journal* **28**: 48-57.
744 **Delis C, Dimou M, Flemetakis E, Aivalakis G, Katinakis P. 2006.** A root- and hypocotyl-
745 specific gene coding for copper-containing amine oxidase is related to cell expansion in
746 soybean seedlings. *Journal of Experimental Botany* **57**: 101-111.
747 **Delis C, Krokida A, Georgiou S, Pena-Rodriguez LM, Kavroulakis N, Ioannou E,**
748 **Roussis V, Osbourn AE, Papadopoulou KK. 2011.** Role of lupeol synthase in *Lotus*
749 *japonicus* nodule formation. *New Phytologist* **189**: 335-346.
750 **Eamens A, Wang MB, Smith NA, Waterhouse PM. 2008.** RNA silencing in plants:

751 yesterday, today, and tomorrow. *Plant Physiology* **147**: 456-468.

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

755 **Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE. 2011.**
756 Formation of plant metabolic gene clusters within dynamic chromosomal regions.
757 *Proceedings of the National Academy of Sciences of the United States of America* **108**:
758 16116-16121.

759 **Field B, Osbourn AE. 2008.** Metabolic diversification--independent assembly of operon-like
760 gene clusters in different plants. *Science* **320**: 543-547.

761 **Grandmougin-Ferjani A, Dalpe Y, Hartmann MA, Laruelle F, Sancholle M. 1999.**Sterol
762 distribution in arbuscular mycorrhizal fungi.*Phytochemistry* **50**: 1027-1031.

763 **Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ. 2001.** The CYP88A
764 cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin
765 biosynthesis pathway. *Proceedings of the National Academy Sciences of the USA* **98**: 2065-
766 2070.

767 **Henderson IR, Jacobsen SE. 2008.** Tandem repeats upstream of the *Arabidopsis* endogene
768 SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes and*
769 *Development* **22**: 1597-1606.

770 **Hernandez LE, Cooke DT. 1996.** Lipid composition of symbiosomes from pea root
771 nodules. *Phytochemistry* **42**: 341-346.

772 **Huhman DV, Sumner LW. 2002.** Metabolic profiling of saponins in *Medicago sativa* and
773 *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer.
774 *Phytochemistry* **59**: 347-360.

775 **Iturbe-Ormaetxe I, Haralampidis K, Papadopoulou K, Osbourn AE. 2003.** Molecular

776 cloning and characterization of triterpene synthases from *Medicago truncatula* and *Lotus*
777 *japonicus*. *Plant Molecular Biology* **51**: 731-743.
778 **Kliebenstein DJ, Osbourn A. 2012.** Making new molecules - evolution of pathways for
779 novel metabolites in plants. *Current Opinionin Plant Biology* **5**:415-423
780 **Kudo F, Motegi A, Mizoue K, Eguchi T. 2010**. Cloning and characterization of the
781 biosynthetic gene cluster of 16-membered macrolide antibiotic FD-891: Involvement of a
782 dual functional cytochrome P450 monooxygenase and hydroxylation. *ChemBioChem* **11,**
783 1574-1582
784 **Latunde-Dada AO, Cabello-Hurtado F, Czittrich N, Didierjean L, Schopfer C,**
785 **Hertkorn N, Werck-Reichhart D, Ebel J. 2001.** Flavonoid 6-hydroxylase from soybean
786 (Glycine max L.), a novel plant P-450 monooxygenase. *Journal of Biological Chemistry* **276**:
787 1688-1695.
788 **Liu L, Lomonossoff G. 2002.**Agroinfection as a rapid method for propagating Cowpea
789 mosaic virus-based constructs. *Journal of Virological Methods* **105**: 343-348.
790 **Lombari P, Ercolano E, El Alaoui H, Chiurazzi M. 2003.** A new transformation-
791 regeneration procedure in the model legume *Lotus japonicus*: root explants as a source of
792 large numbers of cells susceptible to Agrobacterium-mediated transformation. *Plant Cell*
793 *Reports* **21**: 771-777.
794 **Lupien S, Karp F, Wildung M, Croteau R. 1999.**Regiospecific cytochrome P450 limonene
795 hydroxylases from mint (Mentha) species: cDNA isolation, characterization, and functional
796 expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. *Archives of*
797 *Biochemistry and Biophysics* **368**: 181-192.
798 **Maekawa T, Kusakabe M, Shimoda Y, Sato S, Tabata S, Murooka Y, Hayashi M.**
799 **2008.**Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in
800 *Lotus japonicus*. *Molecular Plant Microbe Interactions* **21**: 375-382.

Martirani L, Stiller J, Mirabella R, Alfano F, Lamberti A, Radutoiu S, Iaccarino M, Gresshoff PM, 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. *Molecular Plant Microbe Interactions* **12:**275-284.

Mylona P, Owatworakit A, Papadopoulou K, Jenner H, Qin B, Findlay K, Hill L, Qi X, Bakht S, Melton R, Osbourn A. 2008. Sad3 and Sad4 are required for saponin biosynthesis and root development in oat. *Plant Cell* **20**: 210-212

Naoumkina MA, Modolo LV, Huhman DV, Urbanczyk-Wochniak E, Tang Y, Sumner LW, Dixon RA. 2010. Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* **22**: 850-866.

Nelson D, Werck-Reichhart D. 2011.A P450-centric view of plant evolution.*Plant Journal* **66**: 194-211.

Ohyama K, Suzuki M, Kikuchi J, Saito K, Muranaka T. 2009. Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 725-730.

Okada A, Okada K, Miyamoto K, Koga J, Shibuya N, Nojiri H, Yamane H. 2009. OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoidphytoalexins in rice. *Journal of Biological Chemistry* **284**: 26510-26518.

Osbourn A. 2010. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends in Genetics* **26**: 449-457.

Osbourn A, Goss RJ, Field RA. 2011. The saponins: polar isoprenoids with important and diverse biological activities. *Natural Product Reports* **28**: 1261-1268.

Osbourn AE, Field B. 2009. Operons. *Cellular and MolecularLife Science* **66**: 3755-3775.

Papadopoulou K, Melton RE, Leggett M, Daniels MJ, Osbourn AE. 1999. Compromised

826 disease resistance in saponin-deficient plants. *Proceedings of the National Academy of*
827 *Sciences of the United States of America* **96**: 12923-12928.
828 **Pauwels L, Inze D, Goossens A. 2009.** Jasmonate-inducible gene: What does it mean?
829 *Trends in Plant Science* **1**: 87-91.
830 **Pfaffl MW. 2001.** A new mathematical model for relative quantification in real-time RT-
831 PCR. *Nucleic Acids Research* **29**: e45.
832 **Phillips DR, Rasbery JM, Bartel B, Matsuda SPT. 2006.**Biosynthetic diversity in plant
833 triterpene cyclization. *Current Opinion in Plant Biology* **9**: 305-314.
834 **Pollier J, Morreel K, Geelen D, Goossens A. 2011.** Metabolite profiling of
835 triterpene saponins in Medicago truncatula hairy roots by liquid chromatography
836 Fourier transform ion cyclotron resonance mass spectrometry. *Journal of Natural*
837 *Products* **74:** 1462-1476.
838 **Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A. 2004.** A gene cluster for
839 secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants.
840 *Proceedings of the National Academy of Sciences of the United States of America* **101**: 8233-
841 8238.
842 **Qi X, Bakht S, Qin B, Leggett M, Hemmings A, Mellon F, Eagles J, Werck-Reichhart D,**
843 **Schaller H, Lesot A, Melton R, Osbourn A. 2006.** A different function for a member of an
844 ancient and highly conserved cytochrome P450 family: From essential sterols to plant
845 defense. *Proceedings of the National Academy of Sciences of the United States of America*
846 **103**: 18848-18853.
847 **Ramakers C, Ruijter JM, Deprez RH, Moorman AF. 2003.** Assumption-free analysis of
848 quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**: 62-
849 66.

Rochfort, S., and Panozzo, J. 2007. Phytochemicals for health, the role of pulses. *Jorunal of Agricultural and Food Chemistry* **55:** 7981-7994.

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, Tabata S. 2008. Genome structure of the legume, *Lotus japonicus*. *DNA Research* **15**: 227-239.

Sawai S, Saito K. 2011. Triterpenoid biosynthesis and engineering in plants. *Frontiers in Plant Science***2**: 25.

Sawai S, Shindo T, Sato S, Kaneko T, Tabata S, Ayabe S, Aoki T. 2006. Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*.*Plant Science* **170**: 247-257.

Schult K, Meierhoff K, Paradies S, Toller T, Wolff P, 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.

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

Shimura K, Okada A, Okada K, Jikumaru Y, Ko KW, Toyomasu T, Sassa T, Hasegawa M, Kodama O, Shibuya N, Koga J, Nojiri H, Yamane H. 2007. Identification of a biosynthetic gene cluster in rice for momilactones. *Journal of Biological Chemistry* **282**: 34013-34018.

Suzuki H, Achnine L, Xu R, Matsuda SP, Dixon RA. 2002. A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Journal* **32**:

1033-1048.

Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ. 2009. CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* **21**: 3315-3325.

Takahashi Y, Nishida Y, Ishii Y, Ishikawa K, Asai S. 2005. Monitoring expression of cytochrome P450 genes during postischemic rat liver reperfusion using DNA microarrays. *Journal of Pharmacological Science* **97**: 153-156.

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jorgensen K, Moller BL, Rook F. 2011. Genomic clustering of cyanogenicglucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *Plant Journal* **68**:273-286.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731-2739.

Tokai T, Koshino H, Takahashi-Ando N, Sato M, Fujimura M, 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.

Tsichritzis F, Jakupovic J. 1990. Diterpenes and other constituents fromRelhania species. *Phytochemistry* **29**: 3173-3187.

Ulubelen A, Topcu G, Lotter H, Wagner H, Eris C. 1994. Triterpenoids from the aerial parts of *Salvia montbretii*. *Phytochemistry* **36**: 413-415.

Urbanski DF, Malolepszy A, Stougaard J, Andersen SU. 2012. Genome-wide *LORE1*retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *Plant Journal* **69**: 731-741.

Vaistij FE, Jones L, Baulcombe DC. 2002. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* **14**: 857-867.

vanEngelen FA, Molthoff JW, Conner AJ, Nap JP, Pereira A, Stiekema WJ. 1995.pBINPLUS: an improved plant transformation vector based on pBIN19. *Transgenic Research* **4**: 288-290.

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.

Voinnet O, Rivas S, Mestre P, 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.

Wassenegger M, Heimes S, Riedel L, Sanger HL. 1994. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76:**567-576.**Wegel E, Koumproglou R, Shaw P, Osbourn A. 2009.** Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. *Plant Cel l***21**: 3926-3936.

.

Whitmer S, Verpoorte R, Canel C. 1998. Influence of auxins on alkaloid accumulation by a transgenic cell line of *Catharanthusroseus*. *Plant Cell, Tissue and Organ Culture* **53**: 135-141.

Wilderman PR, Xu M, Jin Y, Coates RM, Peters RJ. 2004. Identification of syn-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiology* **135**: 2098-2105.

Wink M, Mohamed GIA. 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.**Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE, Walker C, Bowser TA, Graham IA. 2012.** A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* **336**: 1704-1708.

Xu R, Fazio GC, Matsuda SP. 2004. On the origins of triterpenoid skeletal diversity.*Phytochemistry* **65**: 261-291.

Yahia A, Kevers C, Gaspar T, Chenieux J-C, Rideau M, Creche Jl. 1998. Cytokinins and ethylene stimulate indole alkaloid accumulation in cell suspension cultures of *Catharanthus roseus*by two distinct mechanisms. *Plant Science* **133**: 9-15.

Zhao J, Davis LC, Verpoorte R. 2005. Elicitor signal transduction leading to production of plant secondary metabolites.*Biotechnology Advances* **23**: 283-333.

Supporting Information

Figures

Fig. S1: Phylogenetic trees of OSCs, cytochrome P450s, and SDR-like proteins constructed by the maximum likelihood method with a 1000 bootstrap replicates.

Fig. S2: Gene transcript levels of *AMY2, LjCYP88D5* and *LjCYP71D353*, constituting the *AMY2* gene cluster, expressed relative to the level of each gene expression in seven-days-old root tissues.

Fig. S3: *In situ* hybridization of *AMY2, LjCYP71D353* and *LjCYP88D5* gene transcripts in mature 28dpi and developing 14dpi *L. japonicus* nodules.

Fig. S4: Gene transcript levels of *LjSDRt*, present in the *AMY2* gene cluster, detected in both uninoculated and inoculated with *M. loti.*

Fig. S5: *AMY2* cluster gene expression in *L. japonicus* root tissues subjected to salt stress.

Fig. S6: Main fragments in the mass spectrometry fragmentation patterns of the TMS-derivatives of 20-hydroxy-lupeol and 20-hydroxy-betulinic acid.

Fig. S7: GC-MS analysis of *N.benthamiana* leaf extracts infiltrated with different *AsbAS1* construct combinations.

Fig. S8: Syntenic analysis of multiple genomic regions encompassing OSC genes in *L.japonicus* and *A.thaliana*

Tables

Table S1: Primers used in experimental procedures

977 **Figure legends**

978

979 **Figure 1**

980 Structure of genomic regions encompassing *OSCs* flanked by genes putatively
981 involved in triterpene metabolism in legumes. (a) Map of candidate gene clusters on
982 chromosome 3 of *L. japonicus*, analysing LjT11L01 and the continuous LjT138B03-
983 LjB16L08 genomic clones and (b) organization of genes in *AMY2* flanking genomic
984 regions.

985

986 **Figure 2**

987 Phylogenetic trees of OSCs (a), cytochrome P450s (b) and SDR-like proteins (c),
988 constructed by the neighbour-joining method with a 1000 bootstrap replicates. The
989 scale bar indicates the number of amino acid substitutions per site. Cytochrome P450s
990 from *A. thaliana, M. sativa*, *L. japonicus, G. uralensis* and *A. sativa* adjacent to /
991 highly coexpressed with OSCs or previously found to participate in triterpene
992 biosynthesis were used for the phylogenetic analysis. SDR-like proteins from legumes
993 *L. japonicus, M. sativa, Glycine max,* from *A. thaliana, A. lyrata, Theobroma cacao*
994 as well as the green algae *Micromonas pusilla* are shown in (c). The open and black
995 stars, open circles, open and black triangles indicate the cytochrome P450s clustered
996 together with the BARS1, AMY2, THAS1, AsBAS1 and MRN1, respectively. The
997 black box indicates the legume specific cytochrome P450 subfamily.

998

999 **Figure 3**

1000 AMY2 cluster gene expression in different developmental stages. Gene transcript
1001 levels of *AMY2, LjCYP88D5* and *LjCYP71D353*, constituting the *AMY2* gene cluster,

1002 were detected in both uninoculated (a,c,e) and inoculated with *M. loti* (b,d,f) *L.*
1003 *japonicus* roots, leaves and nodules. Uninoculated and inoculated plants are of the
1004 same age at the stages of 7d old (days old)-7dpi (days post infection), 14d old-14 dpi
1005 and 28d old-28 dpi, respectively, but are analysed in different real-time PCR
1006 reactions. Total RNA was reverse transcribed, the concentration was normalized
1007 between samples and then real-time PCR was performed. Relative gene expression
1008 was measured with respect to *UBQ* transcripts. Mean values ± SD are shown (n=3).
1009
1010 **Figure 4**
1011 *AMY2* cluster gene expression in *L. japonicus* root tissues subjected to various
1012 exogenously applied plant hormones and environmental cues. Gene transcript levels
1013 of *AMY2, LjCYP88D5* and *LjCYP71D353* in root tissues (14 days old) treated with
1014 various concentrations of 2,4D, BA and MeJA (a-c) and in root tissues treated with
1015 heat and cold stress (12 days old, treated 24 hours) and paraquat (14 days old) (d-f).
1016 Total RNA from roots (20-50 seedlings per treatment) was reverse transcribed, the
1017 concentration was normalized between samples and then real-time PCR was
1018 performed. Relative gene expression was measured with respect to *UBQ* transcripts.
1019 Data from a single representative experiment are presented; three experimental
1020 repeats yielded similar results. Statistical comparisons were performed by Duncan
1021 tests ($a<0.05$). Indicator letters in common denote lack of significant difference. Bars
1022 indicate standard errors of means (n=3).
1023
1024 **Figure 5**
1025 GC-MS analysis of saponified *Nicotiana benthamiana* leaf extracts after transient
1026 expression of AMY2, LjCYP71D353 and/or LjCYP88D5. (a) Total ion
1027 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-hydroxy-lupeol (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).

Figure 6

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. CYP71 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.

Figure 7

AMY2 cluster gene expression in *L. japonicus* hairy-root tissues. Plants, silenced either for *AMY2* (a) or *LjCYP88D5* (e) were generated by *A. rhizogenes* mediated transformation. Expression levels of *LjCYP88D5* (b), *LjCYP71D353* (c) and *LjSDRt* (d) are down-regulated in *AMY2* silenced roots. Similarly, expression levels of *AMY2* (f), *LjCYP71D353* (g) and *LjSDRt* (h) are down-regulated in *LjCYP88D5* silenced roots. Three representative plants for each 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 standard error of means of technical repeats (n=3).

Figure 8

AMY2 cluster gene expression in *L. japonicus AMY2* silenced stable transformants. Four plant lines (amy2-1 to 2-4) silenced for *AMY2* (a) were generated by *A. tumefaciens* mediated transformation. Expression levels of *LjCYP88D5* (b), *LjCYP71D353* (c) and *LjSDRt* (d) 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).

Table 1: Predicted polypeptides encoded by the genes present in the two cluster identified in *L. japonicus* genome and similarities with characterized proteins and enzymes in legumes and other plants

CLUSTER	***OSC***	***CYP***	***Oxidoreductase***
L. japonicus ***AMY2* cluster**	AMY2 (Iturbe-Ormaetxe et al, (2003)	LjCYP88D5 (72% similarity to GuCYP88D6; (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* –like cluster**	AMY2 interrupted (Iturbe-Ormaetxe et al, (2003)	LjCYP88D4 (71% similarity to GuCYP88D6; (Seki et al, (2008) (81% similarity to LjCYP88D5, this study)	

Table 2: Degree of DNA methylation of Lj*SDRt* and Lj*CYP71D353* in wild type and silenced plant lines

	% 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*

(*) indicates statistical significant increase in DNA methylation degree between the silenced and wild type genotypes (*t*-test, $p < 0.01$)