The *R2R3-MYB TT2b* and the *bHLH TT8* genes are the major regulators of proanthocyanidin biosynthesis in the leaves of *Lotus* species

Authors

Francisco José Escaray[1], Valentina Passeri[2], Ana Perea-García[1], Cristian Antonelli[1], Francesco Damiani[2], Oscar Adolfo Ruiz[†1] and Francesco Paolocci[*†2]

[*] *Corresponding author (Phone: +39 075 5014861 Fax: 39 075 5014869*)

[†] Equal contributors as senior authors

[1] IIB-INTECH/CONICET-UNSAM, Chascomús, Bs. As, Argentina.

[2] Institute of Biosciences and BioResources (CNR-IBBR), Perugia, Italy.

Author Contribution Statement

FJE, VP, CA, FD and FP cloned the genes and performed phylogenetic analysis; FJE,VP, APG and FP performed qRT PCR analysis; FJE and FD performed statistical analyses; FJE, APG and FP wrote the paper; FJE, OAR and FP conceived and designed research. All authors read and approved the manuscript.

Main Conclusion

By exploiting interspecific hybrids and their progeny, here we report on the identification of key regulatory and transporter genes intimately related to proanthocyanidin biosynthesis in leaves of *Lotus* spp.

Abstract

Proanthocyanidins (PAs), known as condensed tannins, are polymeric flavonoids enriching forage legumes of key nutritional value to prevent bloating in ruminant animals. Unfortunately, major forage legumes such as alfalfa and clovers lack PAs in edible tissues. Therefore, engineering the PA trait in herbage of forage legumes is paramount to improve both ecological and economical sustainability of cattle production system. Progresses on the understanding of genetic determinants controlling

PA biosynthesis and accumulation have been mainly made studying mutants of Arabidopsis, *Medicago truncatula* and *Lotus japonicus*, model species unable to synthesize PAs in the leaves. Here, we exploited interspecific hybrids between *Lotus corniculatus,* with a high levels of PAs in the leaves, and *Lotus tenuis,* with no PAs in these organs, and relative F_2 progeny, to identify among candidate PA regulators and transporters the genes mainly affecting this trait. We found that the levels of leaf PAs significantly correlate with the expression of *MATE1*, the putative transporter of glycosylated PA monomers, and, among the candidate regulatory genes, with the expression of *TT2b* (*MYB*) and *TT8* (*bHLH*). Our study unveils previously unrecognized regulators of PA biosynthesis and highlights differences in the regulation of PA biosynthesis in our *Lotus* genotypes with respect to model species. This information opens new avenues for breeding bloat safe forage legumes.

Keywords

Interspecific hybrids; MATE; MBW complex; catechins; proanthocyanidins; transcription factors.

Introduction

Proanthocyanidins (PAs), also called condensed tannins, are an important class of structurally complex secondary metabolites. They are flavonoids that largely share their biosynthetic pathways with anthocyanins and that might occur as oligomers or polymers. PA polymers are generated by the sequential addition of leucoanthocyanidin, 2,3-trans-(+)-flavan-3-ols (e.g. catechin), or 2,3-cis-(-)-flavan-3-ols (e.g. epicatechin) extension units to starter 2,3-cis-(-)-flavan-3-ols or 2,3-trans-(+)-flavan-3-ols units (Dixon et al. 2005). Flavan-3-ol units are biosynthesized by two different pathways, branching from common leucoanthocyanidin (flavan 3, 4-diol) intermediates. While 2,3-trans-(+)-flavan-3-ols are derived from the direct reduction of leucoanthocyanidin via the activity of leucoanthocyanidin reductase (LAR) (Tanner et al. 2003), 2,3-cis-(-)-flavan-3-ols formation occurs via anthocyanidin synthesis and reduction. The two biosynthetic steps leading to 2,3-cis-(-)-flavan-3-ols are catalysed by anthocyanidin synthase (ANS; syn. leucoanthocyanidin synthase LDOX) and anthocyanidin reductase (ANR), encoded by *LDOX*/*ANS* and *BAN*/*ANR* genes, respectively (Xie et al. 2003; Abrahams et al. 2003). Anthocyanidins can be either glycosylated by UDP glycosyltransferases (UGTs) to produce anthocyanins or reduced by ANR to give off 2,3-cis-(-)-flavan-3-ols.

Regulation of the flavonoid pathway appears to be primarily at the level of transcription of the structural genes that encode enzymes for each biosynthetic step (Davies and Schwinn 2003). Flavonoid biosynthesis is typically controlled by the tissue-specific expression of transcription factors belonging to the R2R3-MYB, basic helix-loop-helix (bHLH), and WD-repeat protein families (Lepiniec et al. 2006; Xu et al. 2015). These physically interact as a complex, known as MBW complex. Different MBW complexes regulate the biosynthesis of different flavonoid end products (Allan et al. 2008; Dubos et al. 2010). Indeed, the MYB partner generally provides specificity to the complex (Broun 2005; Feller et al. 2011). In *A. thaliana*, three regulators, namely AtMYB123/TT2 (R2R3MYB), TT8 (bHLH), and TTG1 (WD40 protein), play a central role for the activation of PA-specific genes in the seed coat. If TT2 is the key component to specify PA biosynthesis, TT8 and TTG1 take part to the MBW complexes that also control anthocyanins and mucilage biosynthesis (Nesi et al. 2000; Nesi et al. 2001; Zhang et al. 2003). Orthologs of TT2, TT8 and TTG1 have been recently characterized in *Lotus japonicus*, *Medicago* spp. and *Trifolium* spp. (Yoshida et al. 2010a; Verdier et al. 2012; Hancock et al. 2012). Nevertheless, studies in *A. thaliana* and other model species have shown that the accumulation of flavonoids is finely orchestrated by numerous regulatory genes, with MYB members of the MBW complex that act either as activators or repressor (Xu et al. 2015; Zhou et al. 2015).

In planta, PAs have been related to defence mechanisms against pathogens, fungi and herbivores, as well to abiotic stress tolerance (Dixon and Paiva 1995). PAs are among the compounds that strongly affect the nutritional value of forage legumes; even though high concentratio, moderate quantities of these metabolites (about 5 mg PAs/g DW) prevent ruminal bloating (Barry and Mcnabb 1999). The binding of PAs to the proteins reduces in fact the protein fermentation rate in the rumen of animals fed with fresh forage and, consequently, the production of gases of fermentation; notably methane produced by cattle farming systems accounts about 17-37% of the anthropogenic production of greenhouse gases (GHG) (Aerts et al. 1999; Steinfeld and Wassenaar 2007). In essence, with reference to ruminant farming systems PAs increase the conversion efficiency of plant protein into animal protein, reduce the need to include supplemental protein in the diet and decrease livestock environmental impact (Kingston-Smith and Thomas 2003).

Unfortunately, the major forage legumes of temperate clime belonging to *Medicago* and *Trifolium* genus, i.e., alfalfa and clovers, accumulate PAs mainly in seed pods, but not in the herbage (Pang et al. 2007). Due to the lack of intraspecific polymorphism for PAs in the herbage and the complexity of the genetic control of this trait, neither classical nor biotechnological-based approaches have succeeded to breed bloat-safe forage legumes within these genus yet (Pang et al. 2007; Hancock et al. 2012). In stark contrast to *Medicago*, in which anthocyanin biosynthesis appears to be predominant over PA biosynthesis (Li et al. 2016), the *Lotus* genus presents species polymorphic for the foliar PA trait, and in the leaves that accumulate PAs anthocyanins are absent (Sivakumaran et al. 2006; Gruber et al. 2008). The widely cultivated tetraploid *Lotus corniculatus* L accessions present moderate levels of PAs, whereas the diploid species *Lotus tenuis* Waldst. et Kit is unable to accumulate adequate levels of PAs in its leaves (Blumenthal and McGraw 1999; Escaray et al. 2012b). Despite this limit, but by virtue of its adaptability to alkaline soils and flooding condition, *L. tenuis* is a largely cultivated forage legume particularly in marginal areas such as the Flooding Pampas in South

America where is regarded as a “keystone species” for cattle production (Escaray et al. 2012a).

Until now, the genetic determinants controlling PA biosynthesis have been mainly revealed via mutants of Arabidopsis, *Medicago truncatula* and *Lotus japonicus,* model species unable to synthesize PAs into the leveas. Here, we exploited the innovative plant material previously produced by intercrossing *L. tenuis* with a diploid, wild and PA- rich accession of *L. corniculatus* to sort among candidate regulators of the MBW complex and PA transporters those that play a prominent role in controlling the biosynthesis and transport of PAs in edible tissues of *Lotus* spp. (Escaray et al. 2014). To reach this goal, we cloned the *R2R3-MYBs TT2a, TT2b*, *TT2c*, *MYB14* and *PAP2*, the *bHLH TT8,* the *WDR TTG1* and the multidrug and toxic compound extrusion (*MATE1)* genes from both parental plants and assessed whether their expression correlated with those of the structural genes and the levels of PAs in the leaves from parents down to F_2 plants with contrasting levels of these metabolites. By adding new MYB players and highlighting differences in the regulation of this trait in our *Lotus* genotypes with respect to model species, present study opens new avenues for breeding bloat safe forage legumes.

Material and methods

cDNA samples

The pools of cDNA samples from leaves of the parental plants *L. tenuis* and *L. corniculatus,* their interspecific hybrids (LH1; LH2; LH3 and LH4) and F_2 progeny (A27, A120, A186, B3, B5 and B186) were previously employed to clone and study the expression of both early and late genes of the flavonoid pathways as well as the relationship between their expression and the overall accumulation of soluble and insoluble PAs (Escaray et al. 2014). Here, we used these tools to clone and assess the expression levels of master genes related to PA biosynthesis and transport. In essence, plant material, plant growing conditions and cDNA synthesis were as previously described (Escaray et al. 2014). Briefly, the *L. tenuis* parent plants are from a commercial variety (cv. Pampa INTA), bred to be tolerant to the saline-alkaline soils of Argentinean Pampa, the *L. corniculatus* parental plants from a wild, diploid population from Devesa de El Saler (Latitude 39°20′41″ N; Longitude 00°19′12″ W) Valencia, Spain. The hybrid plants investigated are four randomly selected plants, named LH1, LH2, LH3 and LH4 from an F_1 population consisting of 50 individuals, whereas the six selected F_2 plants are from a population of 200 individuals obtained by randomly intercrossing F_1 hybrids. Three F_2 plants are with high levels (A27, A120 and A186), named as A class plants, and three with low levels of foliar PAs (B3, B5 and B186), named as B class plants.

The parental plants (*L. corniculatus* and *L. tenuis*), the selected F_1 hybrid plants (LH1, LH2, LH3 and LH4) and F2 progeny were propagated by cuttings and grown either under growth chamber condition, for the parental plants and F_1 hybrids, or in pots under field condition for the parental plants and F_2 hybrids, as previously reported (Escaray et al. 2014).

Cloning of PA - related genes

Sequences of regulatory genes *TT2, TT8, PAP2, MYB14* and *TTG1* and of *TT12* (*MATE1*) gene from Arabidopsis, *L. japonicus* or other reference species were downloaded from GenBank and aligned using the BioEdit software (Hall 1999). *L. japonicus* sequences homologs to the clover *MYB14,* the Arabidopsis *PAP2* and *M. truncatula MATE1* genes were searched in the *L. japonicus* genome by BLASTn search (http://www.kazusa.or.jp). For each target gene, a specific primer pair was designed on the most conserved nucleotide residues among species (Supplemental Table S1) and used to amplify partial cDNA sequences from leaves of both parental plants. The PCR conditions and sequencing of the resulting amplicons were as reported in Paolocci and colleagues (2007). To clone the 5’ and/or the 3’ ends of the cDNA from the parental plant *L. corniculatus* the partial cDNAs of *TT2*, *TT8*, *MYB14* and *MATE* genes were extended by Rapid Amplification of cDNA Ends (RACE) using the SMARTer® RACE cDNA amplification kit (Clontech, Auckland, New Zealand, http://www.clontech.com/) following the manufacturer’s instructions and using the gene-specific primer pairs reported in Supplemental Table S1. Then, the open reading frames of *TT2*, *PAP2, MYB14*, *TT8* and *TTG1* genes from the diploid *L. corniculatus* were PCR amplified using specific primers designed to embrace the first and the last codons of each gene, cloned into the pENTR/D TOPO vector (Thermo Fisher Scientific) and sequenced.

Sequence analysis

Double-strand sequence analysis was carried out, either directly on PCR products or on amplicons previously cloned in the pGEM-T Easy Vector System I (Promega), or pENTR/D TOPO using the Big Dye Terminator Cycle Sequencing Kit and an ABI Prism 310 Sequence Analyzer (Applied Biosystems) according to the supplier's instructions. Gene-specific primers, Sp6 and T7 primers for amplicons cloned into pGEM-T Easy and M13Forward and M13Reverse for amplicons cloned into pENTR D/TOPO were employed in the sequencing reactions. Partial and full-length cDNA sequences were deposited into Gene Bank under the following accession numbers: [GenBank: KY196482] (*LtTT2a*), [GenBank: KY196474] (*LcTT2a*), [GenBank: KY196483] (*LtTT2b*), [GenBank: KY196475] (*LcTT2b*), [GenBank: KY196484] (*LtTT2c*), [GenBank: KY196476] (*LcTT2c*), [GenBank: KY196480] (*LtMYB14*), [GenBank: KY196472] (*LcMYB14*), [GenBank: KY196481] (*LtPAP2*), [GenBank: KY196473] (*LcPAP2*). [GenBank: KY196485] (*LtTT8*), [GenBank: KY196477] (*LcTT8*), [GenBank: KY196486] (*LtTTG1*), [GenBank: KY196478] (*LcTTG1*). [GenBank: KY196479] (*LtMATE1*), [GenBank: KY196471] (*LcMATE1*).

Phylogenetic Analysis

The phylogenetic tree was built using the amino acidic sequences of *TT2s*, *MYB14*, *PAPs*, *TT8*, *TTG1* and *MATE1* from legume and reference species retrieved from public databases. Additionally, the Unigene19239 and CL2952 relative to two MYB transcriptional factors from tetraploid *L. corniculatus* (Wang et al. 2013) were used. The Clustal IV multiple sequence alignment was obtained using the default parameters of the MEGA package (Tamura et al. 2007). The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al. 1992). The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the taxa analysed (Felsenstein 2010). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016).

Quantitative RT-PCR analysis

Gene-specific primer pairs for q-RT PCR assay, designed on conserved sites between the two parental plants with the help of OligoExpress Software (Applera Biosystems) are given in Supplementary Table S2 where are also reported the three primer pairs designed to be specific to the three different *TT2* gene members present in *Lotus* spp. (Yoshida et al. 2008, this work). All these primer pairs were initially checked for their amplification efficiency and specificity using leaf cDNA from both parents. PCR efficiency was assessed by performing a standard curve for each gene using six dilution points, each one replicated four times. To test their specificity, a dissociation analysis was performed after each amplification run, followed by amplicon sequencing. Then, an aliquot of 5 μL of 1:10 diluted cDNAs obtained as reported in Escaray and colleagues (2014) was used in the PCR reaction, which was made up using the Power Sybr-Green PCR core mix (Applied Biosystems) according to the supplier's instructions, in 20 μL of final volume in the presence of 2.5 pmol of each primer. Three biological replicates were analysed per each *Lotus* genotype. Cycling parameters were two initial steps of 50°C for 2 min and 95°C for 2 min, a two-step cycle of 95°C for 15 s and 60°C (or 62.5ºC for *TT2a*, *TT2b* and *TT2c*) for 1 min repeated 50 times and a final step for 10 min at 60°C. Afterward, the dissociation protocol was performed. Amplifications were performed on ABI PRISM 5700 SDS apparatus (Applied Biosystems). For each transcript, the average threshold cycle (Ct) was determined. The gene quantification method based on the relative expression of the target gene versus the reference gene *EFACT*, was adopted according to Escaray and colleagues (2014).

Statistical analysis

Analysis of relative gene expression was performed according to Pfaffl et al. (2002). The expression levels of *MATE1* and regulatory genes here determined were correlated with the levels of transcripts for flavonoid structural genes retrieved from Escaray et al. (2014), being the two expression analyses performed over the same pools of cDNAs. Likewise, the expression of *MATE1* and the regulatory genes were correlated with the levels of PAs reported in Escaray et al. (2014). Correlation analysis was performed using the Pearson test. All statistical analysis was performed using the Infostat program (Di Rienzo et al. 2010).

Results

Isolation of PA regulatory genes from *L. tenuis* and *L. corniculatus* parental plants

For these studies, members of MBW regulatory complex cloned from *A. thaliana* and different legumes were retrieved from public databases and used to clone their homologs from the *Lotus* parental genotypes. Five R2R3-MYB genes, one bHLH and one WDR genes were considered. In *L. japonicus TT2*, the master regulator of PAs in Arabidopsis, is present with three differentially expressed paralogs, named as *TT2a*, *TT2b* and *TT2c* (Yoshida et al. 2008; Yoshida et al. 2010b). The use in qRT-PCR of primers designed on conserved sequences among the *TT2* genes from several species followed by the melting analysis of the resulting amplicons provided us with the first evidence in support of the thesis that also in the diploid *L. corniculatus* different *TT2* paralogs were present (Escaray et al. 2014). In the present study, the partial *TT2* cDNA fragments from the diploid and tetraploid *L. corniculatus* and a *TT2* genomic region from *L. tenuis* previously identified (Paolocci et al. 2011; Escaray et al. 2014) were aligned with those from *L. japonicus* to design primers and clone *TT2* cDNA fragments from both *Lotus* parental species. As a result of this approach, amplicons of about 330 bp, encompassing the R2R3-MYB domain region containing the polymorphic sites among the three *Lj TT2* paralogs, from both parental species were obtained and cloned in *E. coli*. About twenty colonies *per* species were screened and sequenced. From this analysis, it emerged that the *TT2* fragments from both parental plants could be sorted into three classes, highly similar to the three *Lj TT2* paralogs (Supplemental Table S3). Additionally, *MYB* homologs of the PA and anthocyanin regulators *MYB14* from *Trifolium* spp. (Hancock et al. 2012) and *PAP2* from Arabidopsis (Quattrocchio et al. 2006; Gonzalez et al. 2008) respectively, were first searched in the model species *L. japonicus* through a BLASTn search in the *L. japonicus* databases (http://www.kazusa.or.jp). These queries provided *Lj0g3v0094629.1* and JCVI-FLLj-7C5 as the closest *L. japonicus* homologues sequences with 56.0% and 41.5% of sequence similarity to *TaMYB14* and *AtPAP2*, respectively. In turn, these *L. japonicus* sequences were used as queries to interrogate the library of *L. corniculatus* unigenes produced by Wang and colleagues (2013) to give off the Unigene19239 and Unigene2952, highly similar (93.6% and 99.8%) to the *L. japonicus MYB14* and *PAP2* genes, respectively. The same library was used to retrieve the homologs of the *L. japonicus TT8* and *TTG1* genes (Yoshida et al. 2010a) and the Unigene23867 and Unigene322 with 99.6% and 99.1% of sequence similarity to the first and second query sequences, respectively, were identified. Finally, primers designed on the most conserved nucleotide residues between all the TFs mentioned above from *L. japonicus* and *L. corniculatus* were employed to amplify partial cDNAs from the diploid *L. corniculatus* and *L. tenuis* genotypes. As reported in Supplementary Table S3, the partial cDNA fragments for all these genes were highly homologues between the two parental genotypes, with similarity ranging from 96.7% to 99.0%.

Phylogenetic analysis of PA regulatory genes

To perform phylogenetic analysis, the open reading frames of the regulatory genes *TT2*, *PAP2*, *MYB14* and *TTG1* were amplified from the PA rich diploid *Lotus* parent. To clone the full length cDNA sequences for the three *TT2 MYB*s present in the diploid *L. corniculatus*, 5’ and 3’ RACE analyses were performed and then a forward primer, starting from the first codon common to all the members, and three reverse primers, each specific to one of the gene members, were employed in an end-to-end RT-PCR. Sequencing analysis proved that the open reading frame of the three resulting amplicons consisted of 907, 778 and 745 nucleotides, and they had a high similarity to the *L. japonicus TT2a*, *TT2b* and *TT2c*, respectively (similarity of 97.5% for *TT2a*; 98.7% for *TT2b*; and 97.4% for *TT2c*).

By taking advantage of the high sequence similarities between the full length cDNA sequences of *PAP2* and *TTG1* from *L. japonicus* and tetraploid *L. corniculatus,* primers embracing the first and the last codons on both genes were designed to clone the homologous sequences from the diploid *L. corniculatus* parental genotype. The open reading frame of *PAP2* from the diploid *L. corniculatus* was 808 nucleotides long and showed 99.1% of identity with *JCVI-FLLj-18L10* clone from *L. japonicus*. The open reading frame of *TTG1* was 1051 nucleotides long and showed high identity (98.5%) with its homologous from *L. japonicus*. Conversely, due to the lack of the 5’ end in Unigene23867 (*TT8*), and of the 3’ end in Unigene19239 (*MYB14*), a RACE analysis was performed to identify these extremes in the corresponding genes from the diploid *L.*

corniculatus. Upon alignment of these termini between *L. japonicus* and tetraoploid *L. corniculatus* sequences, primers were then designed to clone the open reading frames of *TT8* and *MYB14* from the diploid *L. corniculatus*. *MYB14* full CDS from this parent was 937 nucleotides long with 98.6% of identity to the *L. japonicus* homolog (*Lj0g3v0094629.1*), whereas that of *TT8* was 2038 nucleotides long with a 97.5% identity with its *L. japonicus* homolog.

The deduced amino acid sequences obtained for the MYB regulators of *L. corniculatus* were aligned to other R2R3-MYBs and used to construct a phylogenetic tree (Fig. 1A). Among the PA regulators, the TT2s from all *Lotus* spp. formed a well-supported cluster apart from that grouping MYB14s. Yet, within this last cluster MYB14s from *Trifolium* and *Medicago* spp. were clearly separated from MYB14 from *Phaseolus vulgaris*, *Glycine soja* and *Lotus* spp., with the proteins from the latter genus forming a specific subclade. Therefore, the identity in the amino acid sequences between MYB14 of *L. corniculatus* and *L. japonicus* was high (98.1%) but very low (41.2%) when these proteins were compared to TaMYB14 (AFJ53053.1). Finally, *L. corniculatus* PAP2 aminoacidic sequence was 98.4% identical to *JCVI-FLLj-18L10* clone from *L. japonicus* and 98.3% to CL2952 gene (Wang et al. 2013) from *L. corniculatus*. All these proteins grouped with AtMYB75 (PAP2) and away from the clade containing LAPs, the MYBs controlling anthocyanin accumulation in *Medicago* spp. (Peel et al. 2009) (Fig. 1A).

The deduced amino acid sequence of all the MYBs considered above displayed the typical motifs indicative of PA- or anthocyanin-regulating R2R3-MYB factors, in particular, the $[D/E]L_{x2}[R/K]_{x3}L_{x6}L_{x3}R$ motif required for the interaction with bHLH transcription factors (Zimmermann et al. 2004) (Fig. 1B).

The deduced amino acid sequences of *L. corniculatus TT8* and *TTG1* were aligned with bHLHs and WDRs from different model and crop species (Supplemental Fig. S1 and S2). Interestingly, the highly identical *L. corniculatus* and *L. japonicus* (BAH28881.1) TT8s (97.8%) clustered apart from AtTT8 as well as bHLHs TAN1, thought to be related to PA biosynthesis in *L. japonicus* and *L. angustissimus* (Yoshida et al. 2010a). TTG1 from the diploid *L. corniculatus* was almost identical (98.5% of identity) to its *L. japonicus* homologous (BAH28880.1) and, with a few exceptions, the topology of the tree built for this protein turned out similar to that of TT8.

Isolation and characterization of *MATE1*

In *A. thaliana*, *TT12* gene encodes a tonoplastic MATE enzyme that can act as a proton antiporter to mediate the vacuolar accumulation of PA precursor epicatechin 3'-O-glucoside (E3'G) (Marinova et al. 2007). In *M. truncatula*, two *MATE* genes have been characterized, *MATE1* that complemented *tt12* mutant in Arabidopsis and *MATE2*, which showed higher transport capacity for anthocyanins and lower efficiency for PA precursors with respect to *MATE1* (Zhao and Dixon 2009a; Zhao et al. 2011). Although the ortholog of *TT12* in *L. japonicus* has not been characterized yet, the BLASTn search using *AtTT12* and *MtMATE1* sequences as queries retrieved the predicted gene *Lj2g3v3070850.3* with 74.2% and 79.6% of sequence similarity to *AtTT12* and *MtMATE1*, respectively. In turn, the *Lj2g3v3070850.3* gene was used to identify a partial cDNA (CL2359) from the *L. corniculatus* transcriptomic database produced by Wang and colleagues (2013). These *Lotus* genes were aligned with those from Arabidopsis and *M. truncutula* and primers to amplify a fragment of 1069 bp of this gene from both *Lotus* parents were designed along with primers to clone the open reading frame from *L. corniculatus*. The 1069 bp long amplicons from the two parents were highly similar (identity higher than 96%) each other (Supplementary Table S3). The open reading frame of *LcTT12* was 1509 bp long with the deduced amino acid sequence highly similarity to its homologous from *L. japonicus* (85.3%) and to *AtTT12* and *MtTT12* (68.9 and 79.5%, respectively). The phylogenetic tree resulting from the alignment of *TT12* amino acidic sequences from different species confirmed the closer relatedness of obtained sequences to the *MtMATE1*, the transporter of E3'G, rather than to the *MtMATE2*, the transporter of anthocyanins (Supplemental Fig. S3).

Expression of the *R2R3-MYB* members of MBW complex

The relative expression levels of the transporter and regulatory genes reported above were investigated by qRT-PCR analysis. In a previous study, by means of a primer pair unable to sort among the three different *TT2* paralogs, we showed that the expression levels of this gene cluster correlated with the overall PA accumulation in the *Lotus* herbage, flowers and roots (Escaray et al. 2014). In the present work, the expression of the different *TT2* genes was analysed separately by designing three member-specific primer pairs. The specificity of each primer pair to its target sequence was confirmed by the melting curve analysis coupled to the direct sequencing of the resulting amplicons (Supplemental Fig. S4). When the expression of the different *TT2* genes was investigated on parental and F_1 plants grown under growth chamber condition it emerged that: a) all of these genes were significantly more expressed in the PA-rich parental genotype (Lc) than in PA-depleted one (Lt); b) the difference in the expression levels between parents were the highest for *TT2b* and the lowest for *TT2c;* c) *TT2b* was the only *TT2* gene to be significantly more and less expressed in F_1 plants than in Lt and Lc parents, respectively (Fig. 2A, B and C).

The relative expression of the *MYB14* transcription factor that in *Trifolium* and *Medicago* spp. controls PA biosynthesis (Hancock et al. 2012) was also evaluated on our genotypes. Although higher (around 4 fold more) in *L. corniculatus* than in *L. tenuis,* its expression in LH1, LH2, LH3 and LH4 hybrids was lower or not dissimilar to that in PA depleted parent (Fig. 2D).

Then, the expression of the different *TT2s* and *MYB14* was analyzed in the parents and in the six plants from the segregating population consisting of three F_2 plants with high PA levels (A27, A120 and A186) and three plants with low PA levels (B3, B5 and B186) grown under field condition. As shown previously, the levels of foliar PAs were: a) increased in *L. corniculatus* (around a 20%) with respect to growth chamber condition, whereas no significant change was observed in *L. tenuis*; b) higher in the three class A plants than in *L. tenuis* (ranging from 20 and 40 fold); c) not dissimilar to that of *L. tenuis* in the three class B plants (Escaray et al. 2014). Under this treatment, the relative expression of each *TT2* gene was always higher in *L. corniculatus* than in *L. tenuis,* with *TT2a* and *TT2b* showing the highest differences between the parents (Figs. 3A, B and C). Notably, under field condition the differences in the expression levels of both *TT2a* and *TT2b* between the parents significantly increased with respect to growth chamber condition (Fig. 2A and B, Fig. 3A and 3B). A different expression pattern was observed for the three *TT2* genes in the F_2 plants: the clearest relationship with PA levels was observed for *TT2b* gene (Fig. 3B). Class A genotypes showed the highest *TT2b* mRNA levels, close to that in *L. corniculatus*, whereas B3, B5 and B147 displayed the lowest values, close or even slightly lower than that in *L. tenuis*. *TT2a* expression pattern differed from that of *TT2b* because in F_2 plants with low PAs its expression was markedly higher than in *L. tenuis* (Fig. 3A). The relative expression of *TT2c* and *MYB14* transcription factors showed a similar trend: their expressions were higher in *L. corniculatus* and F_2 plants than in *L. tenuis*, but did not mirror the levels of PAs displayed by the six F_2 plants (Fig. 3C and D).

Finally, the expression of *PAP2*, the R2R3-MYB involved in the anthocyanin regulation, was analysed. Under growth chamber condition, its relative expression was higher in *L. corniculatus* than *L. tenuis*, and intermediate in F_1 hybrid plants (Fig. 4A). Under field condition, the difference in *PAP2* expression between the parents was confirmed, whereas among the F_2 hybrids the steady state levels of *PAP2* mRNAs varied regardless of the levels of PAs. In A186, it was lower than in B147 despite the fact that the former plants accumulated higher levels of PAs than the latter (26.53 vs 0.96 mg PA/g DM) and was similar between A120 and B3 with 19.94 and 0.52 mg PA/g DM, respectively (Fig. 4B).

Expression of *bHLH* and *WDR* members of MBW complex

Beside the R2R3-MYBs, the ternary complex (MBW) is made up of basic helix–loop–helix (bHLH) and WD-repeat (WDR) proteins. Therefore, the relative expression of *TT8* (bHLH class) and *TTG1* (WDR class) genes was measured by qRT-PCR (Fig. 5).

Under both growing conditions, *TT8* expression was higher (almost 8 fold more) in *L. corniculatus* than *L. tenuis* (Fig. 5A and B). Considering the F_1 hybrid plants, in those with the lowest levels of PAs, namely LH2 and LH3, *TT8* mRNA steady state levels were intermediate between the parents, whereas were not dissimilar to the *L. corniculatus* parent in those with the highest PA levels (LH1 and LH3) (Fig. 5A). Considering the F_2 plants, the relative expression of *TT8* was consistent with PA levels: its expression was similar between A class plants (A27, A120 and A186) and *L. corniculatus*, and between B class plants (B3, B5 and B147) and *L. tenuis* (Fig. 5B).

As far as the WDR component is concerned, *TTG1* gene expression was different between parental genotypes under growth chamber, but not under field condition (Figs. 5C and D). Neither in F_1 nor in F_2 plants the expression of this gene was consistent with the levels of PAs (Fig. 5C and D).

Relative expression of *MATE1* transporter

The relative expression of the putative *MATE1* gene that encodes an enzyme related to the transport of glycosylated PA monomers to vacuole was measured as described for the other genes (Fig. 6). Analysis performed on F_1 hybrid and parental plants showed that the relative expression of this gene was the highest in *L. corniculatus*, intermediate in *L. tenuis* x *L. corniculatus* plants and the lowest in *L. tenuis*. In turn, between F_1 hybrid plants, the relative expression of *MATE1* gene was higher in LH1 and LH4 than LH2 and LH3, according to the foliar PA levels (Figs. 1E and 5A). A similar pattern was observed in F_2 plants: *L. corniculatus* and A class hybrid plants showed a higher expression (around 8 fold more) than *L. tenuis* and the B class plants.

Global analysis of gene expression related to PA content in leaves

The MBW transcriptional complex controls the transcription of genes that code for key enzymes of the flavonoid pathway (reviewed by Xu et al. 2015). According to their position in the pathway, these genes are classified as early and late genes. The former include, among others, *PAL* and *CHS* that code for phenylalanine ammonia-lyase (PAL) and chalcone synthase (CHS), respectively. These early genes are involved into the biosynthesis of many secondary metabolites. Late genes are directly involved into anthocyanin and PA biosynthesis and include *DFR*, *ANS*, *ANR*, *LAR1* and *LAR2* genes. Here, we investigated whether, and to what extent, the expression of the three *TT2 MYBs*, *MYB14*, *PAP2*, *TT8* and *TTG1*, among the regulatory proteins, and of *MATE1* correlated with the expression of both early and late structural genes as well as with the levels of soluble and insoluble PAs in F_2 plants.

Considering the different *TT2 R2R3-MYB* genes, none of them correlated with early genes (Table 1). As far it concerns late genes, *TT2b* but not *TT2a* showed a positive correlation with *DFR* (r=0.68; p=0.0019), *LAR2* (r=0.85; p<0.0001) and *MATE1* (r=0.83; p<0.0001), whereas both correlated positively with *ANS*, *ANR* and *LAR1* (Table 1). Differently from the other two *TT2* gene members, the expression of *TT2c* did not show any relationship with the mRNA levels of the late genes (Table 1). The transcription factor *MYB14* positively correlated with *PAL* (r=0.54; p=0.02) among early genes, *ANS* (r=0.85; p<0.0001), *ANR* (r=0.87; p<0.0001), *LAR1* (r=0.56; p<0.02) and *MATE1* (r=0.54; p<0.02) among the late genes. The correlation with PA levels was significant for *TT2a*, *TT2b* and *MYB14* genes; in particular, the expression of *TT2b* gene showed the strongest correlation coefficient (r) with PAs levels (around 0.9 for the three fractions, $p<0.0001$), whereas the r-values between the expression of *TT2a* and *MYB14* and PAs levels were low (r=0.49; p=0.04 for *TT2a* and r=0.52; p=0.03 for *MYB14*), and not significant for *TT2c* (Table 1).

The levels of anthocyanins in the leaves of the *Lotus* material under investigation are very low and generally undetectable utilizing the spectrophotometric method, but their accumulation is visible at naked eye and quantifiable by spectrophotometric analysis in the stems (Supplemental Fig. S5). Thus, the expression levels of *PAP2* in the leaves were compared with the total levels of anthocyanins in the stems. A significant correlation between these two variables did not emerged (p=0.059 for parental and F_1 plants under growth chamber condition; p=0.920 for parental and F_2 plants under field condition). Also, when the expression of this gene was compared with the expression of structural genes or with soluble, insoluble and total PAs in F_2 plants no significant values emerged (Table 1).

When the relationship between the mRNA levels of the *MYB* genes was examined, it turned out: a) that *TT2a* but not *TT2b* correlated with *TT2c*; b) that *MYB14* correlated with all *TT2*s, with the highest coefficient for *TT2a* (r=0.84; $p\leq0.01$), the lowest for *TT2c* (r=0.48; p=0.04); and c) that *PAP2* correlated only with *TT2c* (r=0.67; $p\leq0.01$).

Regarding the bHLH component, the expression of *TT8* gene did not correlate with mRNA levels of early genes or with the expression of *DFR*; however, it positively correlated with those of the remaining late genes ($p\leq0.01$) (Table 1). Consequently, the correlation of *TT8* and total PA content was high and highly significant (r=0.86; $p<0.0001$). The expression of this transcriptional factor also significantly correlated with the *MYBs* *TT2a* (r=0.76, $p<0.0001$), *TT2b* (r=0.95, $p<0.0001$) and *MYB14* (r=0.80, $p<0.0001$), but not with *TT2c* and *PAP2*. Finally, the expression of *TTG1* was not consistent with the expression of early, late genes and other regulatory genes), or with the PA content.

Statistical analyses showed that *MATE1* expression correlated with soluble (r=0.85; $p<0.0001$), insoluble (r=0.91; $p<0.0001$) and total PAs (r=0.90; $p<0.0001$), with the regulatory genes *TT2b* and *TT8*, and, although less markedly, but still significantly with *MYB14*.

Altogether, these results showed that, among all the regulatory genes tested, *TT2b* and *TT8* presented the strongest positive correlations with the levels of transcripts relative to the key structural genes of both catechin and epicatechin branches of the PA pathway, the putative *MATE1* transporter and, by translation, with the levels of soluble and insoluble PAs in the leaves of *Lotus* spp. (Table 1).

Discussion

The increasing global meat demand, coupled to the incoming global warming, calls for a more sustainable and productive ruminant farming systems (Opio et al. 2013). Both aspects are related to rumen efficiency. Engineering the biosynthesis of PAs in the most valuable forage legumes is a natural alternative to chemical feed additives to improve rumen fermentation efficiency (Aerts et al. 1999; Patra and Saxena 2010). Thus, it is crucial to deep our knowledge on genetic and environmental determinants controlling the accumulation of these pigments in the herbage. Here, by taking advantage of innovative plant material, consisting of interspecific hybrids between PA rich and PA depleted *L. corniculatus* and *L. tenuis* species, respectively, and their segregating population (Escaray et al. 2014), we identified key PA regulators on *Lotus* genotypes of agronomical relevance. Present results suggest a different regulation of PAs between species that accumulate almost exclusively epicatechin based PA from those in which both catechin and epicatechin based PAs are present as well as those accumulating PAs in the seed coat *vs.* those in which these pigments are also synthesized into the leaves.

Lotus interspecific hybrid and their progeny: a model system to disclose key players of PA biosynthesis in legumes

The regulation of PA biosynthesis is controlled by the MBW complex made up of MYB, bHLH and WDR proteins in all species investigated. Nevertheless, the cloning from legumes of more and more transcription factors that affect the accumulation of these metabolites has prompted several authors to argue that there is some degree of promiscuity for the components of the MBW complex, at least when their function is investigated using heterologous systems (Liu et al. 2014). In light of these considerations, we decided to take advantage of *L. corniculatus* x *L. tenuis* interspecific hybrids and relative segregating progeny to sort, among regulators known to control PA biosynthesis in legumes, those that play a prominent role on this trait in *Lotus* genotypes of agronomic relevance. Differently from previous studies, based on the effects of the ectopic expression of *MYB* or *bHLH* PA regulators in tetraploid *L. corniculatus* accessions (Robbins et al. 2003; Paolocci et al. 2007; Paolocci et al. 2011) or on the diploid model species *L. japonicus* that is no committed to accumulate PAs in leaves (Yoshida et al. 2008; Yoshida et al. 2010b), present study employs diploid accessions only, and it does not rely on manipulation of single genes at a time. Rather, it relies on the genomic re-assortment originated by interspecific hybridization that enabled us to address the regulation of PA biosynthesis more globally and in a context that might mimic what occurred during the differentiation of *Lotus* spp. Likewise, differently from *Medicago* species used as model in other works, in which PAs are accumulated only in the seed coat and are composed of epicatechin (Liu et al. 2016), present study employs a parental species, *L. corniculatus,* committed to accumulate PAs in leaves based on both catechin and epicatechin units, and F_2 plants polymorphic for the levels of soluble and insoluble PAs (Foo et al. 1996; Paolocci et al. 2007; Escaray et al. 2014). Thus, being the regulation of the different branches of the flavonoid pathway different across genus and species, and leaves the primary target organs into which PAs have to be engineered to produce bloat safe legumes, the plant material employed in the present study offers an innovative model system to tackle the regulation of both PA branches in forage legumes.

MYBs controlling PAs in *Lotus*

With their seminal works, Yoshida and colleagues (2008; 2010b) have shown that the homologous of the Arabidopsis *TT2* gene is organized into three paralogs in *L. japonicus* genome. Out of these three members, the *TT2a* paralog plays a major role as positive regulator of PA biosynthesis in this model legume. Throughout transient expression assays using Arabidopsis leaf protoplasts, these authors showed in fact that *LjTT2a* and, to a lesser extent, *LjTT2c* caused stronger activation than *LjTT2b* of the promoters of *DFR*, *ANS* and *ANR* genes, regardless of the combination of bHLH and WDR factors used. Moreover, TT2a or TT2c, but not TT2b or AtTT2, alone activates the *AtANR* promoter without the co-expression of TT8 and TTG1. The activation of *ANR* promoter by *TT2b* was recovered when three specific amino acid residues in its MYB domain were replaced with those from *TT2a.* Finally, if ectopically expressed in Arabidopsis, *LjTT2a* induces the expression of both early and late genes of the flavonoid pathway, hence the biosynthesis of anthocyanins in seedlings and of PAs in leaves and stems of Arabidopsis plantlets (Yoshida et al. 2010b). In preliminary work done to characterize genes differentially expressed between the *Lotus* parental plants, their hybrids and progenies, we showed that also in other *Lotus* species *TT2* could be arranged as a small gene family, and that the expression of *TT2* paralogs *in toto* correlated with the expression of *ANS*, *ANR*, *LAR1* and *LAR2* genes (Escaray et al. 2014). In the present study, we cloned from *L. tenuis* and the diploid *L. corniculatus* cDNA sequences that are highly homologues to the three *TT2*s present in *L. japonicus* and employed primer pairs specific to each of them in qRT-PCR analyses to disentangle the contribution of each *TT2* paralog to PA biosynthesis in leaves from parental plants down to F_2 progenies. What emanated from this approach is that *TT2b* rather than the other two paralogs plays a major role for the PA trait in leaves. *TT2b* expression pattern fits better than those of the other two paralogs with PA levels in parental plants grown under controlled and field conditions, their hybrids and F_2 progeny. More specifically, the correlation analysis performed on F_2 plants with contrasting PA levels in the leaves shows that *TT2b* has a higher correlation coefficient than *TT2a* both with soluble and insoluble PAs (Table 1). Notably, *TT2b* correlates with all late genes, serving both the catechin (*LAR1* and *LAR2*), and epicatechin (*ANS* and *ANR*) branches of the PA pathway and with the putative *MATE1* transporter. Conversely, *TT2a* correlates primarily with those deputed to the epicatechin pathway and less significantly with *LAR1* and not with *MATE1* (Table 1). Even more contrasting with the results of Yoshida and colleagues (2010b) is the role of *TT2c,* the most expressed *TT2* member in the leaves of any genotypes here investigated (data not shown). *TT2c* expression in fact neither correlates with PAs nor with the expression of the early and late biosynthetic genes tested. In stark contrast, its homolog *LjTT2c* directly regulates the late enzymatic genes of PA biosynthesis (Yoshida et al. 2008; Yoshida et al. 2010b). An additional aspect that differentiates the regulation of PA between the model *L. japonicus* and our model system concerns the activation of *LAR*. In transient assays, none of *LjTT2s* can activate *LjLAR* promoter, regardless of the combination of MBW partners used, and *in planta* the expression of *LjLAR* gene correlates neither with those of the three *TT2*s nor with the levels of PAs (Yoshida et al. 2008). Thus, according to these authors, in *L. japonicus* ANR acts as key enzyme to regulate PA biosynthesis, whereas LAR likely has only a supplementary function. Conversely, the expression of *LAR1* and, albeit less so, that of *LAR2* gene families correlate with the levels of PAs in diploid as well as tetraploid, wild type or transgenic, for the overexpression of either the *bHLH* activator *Sn* from maize or the *MYB* repressor *FaMYB1* from strawberry, *L. corniculatus* plants

(Paolocci et al. 2007; Paolocci et al. 2011; Escaray et al. 2014). From one hand, present data reinforce our contention of a tightly co-regulation of the catechin and epicatechin branches of PAs in *Lotus* leaves. On the other, they lend support to the notion that *TT2b* is the MYB primarily deputed to the control of the catechin branch (Fig.7).

In addition to TT2 from *L. japonicus*, MYB regulators of PA biosynthesis from other forage legumes have been recently characterized. Among them, MYB14 from *T. arvense* and *M. truncatula*, which are functionally orthologs of AtTT2 (Liu et al. 2014). In *T. arvense*, *MYB14* expression coincides with *ANR* and *LAR* expression as well as PA accumulation. Moreover, its constitutive expression in *Nicotiana tabacum*, *M. sativa*, and *T. repens* results in ectopic accumulation of PAs in leaf tissues of these plants without the need for co-expression of additional factors (Hancock et al. 2012). In light of all these considerations, via *MYB14* from *Trifolium* spp. as query, here we first identified its homologs in *L. japonicus* genome (Lj0g3v0094629.1) and in the transcriptomic database of the tetraploid *L. corniculatus* (Unigene19239) to finally clone it from the diploid *L. corniculatus.* The phylogenetic tree resulting from the alignment of MYB14 from a number of different legumes confirms that all these proteins belong to a different clade with respect to that of TT2 or the clades grouping the PAP and LAP MYBs that regulate anthocyanin biosynthesis. Moreover, from the same tree it, clearly emerges that MYB14 from *Lotus* spp., soybean and *Phaseolus vulgaris* form a distinct subclade with respect to the functionally characterized MYB14 from *Trifolium* and *Medicago* spp. Indeed, according to Wang and colleagues (2013) the Unigene19239 is a transcription factor related to flavonoid biosynthesis, but unrelated with foliar PA accumulation, being strongly expressed in flowers and very faintly in leaves. Here, we show that, if in the two parental plants *MYB14* expression coincides with the levels of PAs, regardless of their growing conditions, the same does not hold true in the hybrids and F_2 progeny. Notwithstanding, the analysis run on F_2 plants suggests that *MYB14* expression correlates, although to a far less extent than *TT2b*, with the levels of PAs, as well as the transcript levels of *ANS*, *ANR* and, less markedly yet still significantly, *LAR1*, *MATE1* and the early gene *PAL* (Table 1; Fig. 7). Thus, while waiting for the functional characterization of the different *TT2* genes and *MYB14* from *L. corniculatus*, present data let us to argue that in *Lotus* spp. the contribution of the latter gene to the accumulation of PAs is supplementary or accessory to that of TT2s. Conversely, MYB14 contribution appears to be more relevant on those legume species where efficient PA production requires more, if not only, anthocyanidin precursors. It is noteworthy indeed that in *M. sativa myb14* mutants, wherein content of soluble PAs in the seed coat decreases by 50%, *ANR* is downregulated but not *LAR* (Li et al. 2016). Interestingly, in *M. truncatula* LAR functions as a regulator of PA extension by controlling the production of epicatechin starter units rather than in the biosynthesis of catechins (Liu et al. 2016).

In *M. truncatula*, for a high-level transactivation of PA biosynthesis a quaternary complex made up of the two MYBs, MYB5 and MYB14, TT8 and WD40-1 is required (Liu et al. 2014). To the best of our knowledge, if in *Lotus* spp. the co-expression of different MYBs is necessary to upregulate this biosynthetic pathway is an issue not addressed yet. Our correlation data point to a possible involvement of more than a single MYB or, embracing a less conservative hypothesis, of a possible self-reinforcing mechanism among MYBs, to activate or upregulate the PA pathway, in particular its epicatechin branch. We note in fact that *TT2a* shows a higher correlation than *TT2b* with *MYB14*, and that *TT2a*, but not *TT2b* correlates with *TT2c*. The comparative analysis of PAs and gene expression levels on a higher number of segregating plants would help us to assess whether TT2s and MYB14 act synergistically to activate the transcription of MYBs and PA structural genes. Concomitantly, this issue would be tackled by cloning the promoter regions of MYBs and of the key structural genes *ANR* and *LAR*s in front of reporter genes to study by *in vitro* and *in vivo* assays their activation as a consequence of the presence of different combinations of MYB, bHLH and WDR partners.

Beyond those intimately related to PA biosynthesis, here we also determined if the expression of *MYB*s known to regulate the expression of genes upstream *ANR* and *LAR*, correlates with the PA content in *Lotus* leaves. Although the accumulation of anthocyanins, if any, is negligible in the leaves of *Lotus* spp., in order to address this point we focused on *PAP2*, a member of the R2R3-MYB clade that regulates the anthocyanin biosynthesis in Arabidopsis (Gonzalez et al. 2008). Indeed, the relative expression of *PAP2* in our genotypes neither correlates with anthocyanin levels in the stems (p=0.91), nor with those of PAs in leaves. Therefore, the levels of *PAP2* mRNAs do not appear to be the rate limiting factor for PA biosynthesis to occur. This is congruent with our hypothesis that differently from *Medicago* spp., for an efficient accumulation of PAs in *Lotus* spp. more anthocyanidin precursors might not be necessary.

TT8 but not the WDR component is limiting PA biosynthesis in *Lotus* spp.

To investigate the contribution of WDR and bHLH components of the MBW complex to PA biosynthesis in our *Lotus* genotypes, we cloned the putative ortologs of Arabidopsis and *Medicago TTG1* and *TT8* genes from both parents and assayed their expression levels in our PA polymorphic plant material. The relative expression of *TTG1* is high as shown by the low threshold cycle observed in the qRT-PCR experiments (data not shown). Furthermore, it does not correlate with PAs or the expression levels of all genes tested, to suggest that this component is constitutively expressed in *Lotus* spp. Constitutive expression of *MtWD40-1* has been demonstrated in *M. truncatula* (Pang et al. 2009). Concerning the bHLH partners, we have previously shown that the expression of the putative PA regulator *TAN1* gene does not correlate the PA levels in our plant material (Escaray et al. 2014). These result are in keeping with those provided by Yoshida and colleagues (2010b). By using the three LjTT2s and LjTTG1 in transient assays as effectors, these Authors showed that out of the three *L. japonicus* bHLHs with putative activity on PAs tested, namely LjTAN1, LjEGL3 and LjTT8, the first one could slightly activated the *ANR* promoter only in the presence of LjTT2a and LjTTG1. The strongest activation of this promoter was conversely obtained when LjTT8 protein was employed, regardless of the LjTT2 used as effector. This outcome was to interpreted to mean that “LjTT8 could be an innate PA regulator” (Yoshida et al. 2010a). These data prompted us to clone *TT8* from our *Lotus* parental plants and test its expression throughout their hybrids and F_2 progenies. Indeed, *TT8* expression dovetails nicely with the pattern of PA accumulation and, similarly to *TT2a* and *TT2b,* it correlates with both soluble and insoluble PAs. We also observe that *TT8* shares with *TT2b* the same set of genes with which they correlate, namely *DFR* and, more markedly and significantly, with all the other late genes tested (Fig. 7). More interestingly, the levels of *TT8* mRNAs significantly correlate with those of *TT2b* and, although less markedly, with those of *TT2a* and *MYB14*, but not with *TT2c* and *PAP2*. The MBW complex operates within a gene regulation network that involves reinforcement and repression (Xu et al. 2015). Current models in *T. repens* suggest that the MBW complexes that control anthocyanins and proantochyanidins activate the expression of the bHLH clade to which TT8 belongs (Albert et al. 2014; Albert 2015). Our transcriptomic data can therefore be interpreted to signify that the *Lotus* MBW complex formed primarily, but not exclusively, by TT2b as the R2R3-MYB partner, activates the transcription of *TT8* in a self-reinforcement mechanism. The cloning of *TT8* promoter from *Lotus* spp. is in progress to test by transient assay which, and to that extent, are the endogenous MBW components that trigger its activation.

MATE1 mRNAs correlate with PA key regulators

Whether PA polymerization occurs enzymatically or non- enzymatically and the sub-cellular localization of the oligomerization process remain matters of debate. According to several authors, the polymerization of PA precursors occurs in vacuole (Pourcel et al. 2005). Conversely, the non-enzymatic oligomerization process between epicatechin starter and extension units recently described in *Medicago* would likely occur in the cytoplasm (Liu et al. 2016). The nature of PA intermediates that are transported to the vacuole is consequently not fully understood yet. Glucosyltransferases, glutathione S-transferases, ATP binding cassette (ABC) transporter, tonoplast proton pump P-ATPase and MATE transporter proteins have been shown genetically to be involved in PA precursor transport in Arabidopsis seed coat (Debeaujon et al. 2001; Abrahams et al. 2003; Goodman et al. 2004; Marinova et al. 2007; Pang et al. 2008; Zhao and Dixon 2009 and 2009b; Appelhagen et al. 2015). The Arabidopsis *TT12* codes for MATE1 that functions to transport epicatechin 3'-O glucoside into the vacuole and the *Medicago* MATE1 complements the seed proanthocyanidin phenotype of the Arabidopsis tt12 mutant both qualitatively and quantitatively (Zhao and Dixon 2009a). Here, we cloned and analysed the expression of the putative ortholog of *TT12* and *MtMATE1* from *Lotus* spp. as one of the master genes for PA transport. The phylogenetic analysis supports our contention that the *Lj2g3v3070850.3* gene we have identified in *L. japonicus* genome and its homolog we have cloned from the diploid *L. corniculatus* are indeed the putative orthologs of MATE1 from *M. truncatula*. Not only the expression patter of this gene nicely coincides with the PA quantity in the parents, F_1 and F_2 plants, but also statistical analysis depicts a strong correlation between the steady state levels of its mRNAs with both soluble and insoluble PAs in F_2 plants. In these plants, *MATE1* correlates with the expression of the transcription factors *TT2b* and *TT8* and, although less markedly but still significantly, with *MYB14*. These data are in keeping with data from other legumes showing a MYB or bHLH -mediated regulation not only of biosynthetic genes but also of genes related to modification and transport of flavonoids (Lepiniec et al. 2006; Liu et al. 2014; Li et al. 2016). Thus, *MATE1* is part of the *Lotus* MBW regulon that controls the transcription of PA biosynthesis genes (Fig. 7). Consistently with the composition of PAs, based only on epicatechin units both in *Medicago* and Arabidopsis (Abrahams et al. 2003; Li et al. 2016), *MATE1* from *M. truncatula* and *TT12* from Aradidopsis function to transport epicatechin 3'-O-glucoside as precursors for PA biosynthesis (Zhao and Dixon 2009a). Present investigation calls for specific studies aimed at verifying whether *MATE1* from species that accumulates catechin such as *L. corniculatus* could transport additional glycosylated flavan 3-ols to epicatechin 3'-O-glucoside.

Conclusions

PA accumulation in legume species is a complex, multigenic trait, with several, and likely species-specific, MYB factors that act within the MBW complex either as activators or repressors (Paolocci et al. 2011; Albert 2015; Zhou et al. 2015). Yet, the MBW complex is a gene network regulated by developmental signals and environmental determinants (Xie et al. 2016). The relative contribution of the different regulators and, more specifically, of MYB partners, to PAs might also change according to the tissues and organs within a given forage legume. Present study let us to conclude that in *Lotus* spp., *TT2b*, among the *R2R3-MYB*s and *TT8* among the *bHLH*s play a prominent role in the synthesis of these compounds in the herbage. Moreover, the first gene likely takes part to the MBW that activates the latter one to reinforce the mechanism. Ongoing transient assays and ectopic expression of *LcTT2b* in model and crop forage legumes will provide us with definitive evidence on its interaction with MBW partners and activation properties of the PA pathway. The involvement of additional MYBs such as MYB14 and the other two TT2s might be accessory for the biosynthesis of these compounds in leaves, but crucial for their biosynthesis in the seed coat. High throughput transcriptomic analyses are in progress on our *Lotus* genotypes in order to identify new members of the MBW networks and provide us with an atlas of the transcript levels of putative activators and repressors of PA biosynthesis in different organs of *Lotus* spp.

Acknowledgements

This work was supported by FONCYT-PICT 2011-1612, PICT 2014-3718 and 2014-3648 grants from the Agencia Nacional de Promoción Científica y Tecnológica (ANPCYT-Argentina); PIP 0980 grant from Consejo Nacional de Ciencia y Tecnologìa (CONICET-Argentina), CNR-CONICET bilateral project 2014-2016, and by Fondazione Cassa di Risparmio di Perugia grant n° 2015.0303 021

References

Abrahams S, Lee E, Walker AR, et al. (2003) The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. Plant J 35:624–636

Aerts RJ, Barry TN, Mcnabb WC (1999) Polyphenols and agriculture: beneficial effects of proanthocyanidins in forages. Agric Ecosyst Environ 75:1–12

Albert NW (2015) Subspecialization of R2R3-MYB Repressors for Anthocyanin and Proanthocyanidin Regulation in Forage Legumes. Front Plant Sci 6:1–13

Albert NW, Griffiths AG, Cousins GR (2014) Anthocyanin leaf markings are regulated by a family of R2R3-MYB genes in the genus Trifolium. New Phytol 205:882–893

Allan AC, Hellens RP, Laing W a (2008) MYB transcription factors that colour our fruit. Trends Plant Sci 13:99–102

Appelhagen I, Nordholt N, Seidel T, et al. (2015) TRANSPARENT TESTA 13 is a tonoplast P3A-ATPase required for vacuolar deposition of proanthocyanidins in Arabidopsis thaliana seeds. Plant J 82:840–849

Barry TN, Mcnabb WC (1999) The implications of condensed tannins on the nutritive value of temperate forages fed to ruminants. Br J Nutr 81:263–272

Blumenthal MJ, McGraw RL (1999) Lotus adaptation, use, and management. In: Beuselinck PR (ed) Trefoil Sci. Technol. Lotus. CSSA Special Publication Number 28, Madison, Wisconsin, pp 97–119

Broun P (2005) Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. Curr Opin Plant Biol 8:272–279

Davies KM, Schwinn KE (2003) Transcriptional regulation of secondary metabolism. Funct Plant Biol 30:913–925

Debeaujon I, Peeters AJM, Léon-kloosterziel KM, Koornneef M (2001) The TRANSPARENT TESTA12 Gene of Arabidopsis Encodes a Multidrug Secondary

Transporter-like Protein Required for Flavonoid Sequestration in Vacuoles of the Seed Coat Endothelium. Plant Cell 13:853–871

Di Rienzo JA, Casanoves F, Balzarini MG, et al. (2010) InfoStat versión 2010.

Dixon RA, Paiva NL (1995) Stress-Induced Phenylpropanoid Metabolism. Plant Cell 7:1085–1097

Dixon RA, Xie D-Y, Sharma SB (2005) Proanthocyanidins – a final frontier in flavonoid research? New Phytol 165:9–28

Dubos C, Stracke R, Grotewold E, et al. (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci 15:573–581

Escaray FJ, Menéndez AB, Gárriz A, et al. (2012a) Ecological and agronomic importance of the plant genus Lotus. Its application in grassland sustainability and the amelioration of constrained and contaminated soils. Plant Sci 182:121–133

Escaray FJ, Passeri V, Babuin MF, et al. (2014) Lotus tenuis x L. corniculatus interspecific hybridization as a means to breed bloat-safe pastures and gain insight into the genetic control of proanthocyanidin biosynthesis in legumes. BMC Plant Biol 14:14–40

Escaray FJ, Rosato M, Pieckenstain FL, et al. (2012b) The proanthocyanidin content as a tool to differentiate between Lotus tenuis and L. corniculatus individuals. Phytochem Lett 5:37–40

Feller A, MacHemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J 66:94–116

Felsenstein J (2010) Confidence limits on phylogenies: an approach using the bootstrap. Evolution (N Y) 39:783–791

Foo LY, Newman R, Waghorn G, et al. (1996) Proanthocyanidins from Lotus corniculatus. Phytochemistry 41:617–624

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM (2008) Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. Plant J 53:814–827

Goodman CD, Casati P, Walbot V (2004) A multidrug resistance-associated protein involved in anthocyanin transport in Zea mays. Plant Cell 16:1812–1826

Gruber MY, Skadhauge B, Yu M, et al. (2008) Variation in morphology, plant habit, proanthocyanidins, and flavonoids within a Lotus germplasm collection. Can J plant Sci 88:121

Hall TA (1999) BioEdit: a user-friendly biological sequence alignament editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hancock KR, Collette V, Fraser K, et al. (2012) Expression of the R2R3-MYB transcription factor TaMYB14 from Trifolium arvense activates proanthocyanidin biosynthesis in the legumes Trifolium repens and Medicago sativa. Plant Physiol 159:1204–20

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Kingston-Smith AH, Thomas HM (2003) Strategies of plant breeding for improved rumen function. Ann Appl Biol 142:13–24

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 1–11

Lepiniec L, Debeaujon I, Routaboul J-M, et al. (2006) Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol 57:405–430

Li P, Dong Q, Ge S, et al. (2016) Metabolic engineering of proanthocyanidin production by repressing the isoflavone pathways and redirecting anthocyanidin precursor flux in legume. Plant Biotechnol J 1–15

Liu C, Jun JH, Dixon RA (2014) MYB5 and MYB14 Play Pivotal Roles in Seed Coat Polymer Biosynthesis in Medicago truncatula. Plant Physiol 165:1424–1439

Liu C, Wang X, Shulaev V, Dixon RA (2016) A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. Nat Plants 2:16182

Marinova K, Pourcel L, Weder B, et al. (2007) The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. Plant Cell 19:2023–2038

Nesi N, Debeaujon I, Jond C, et al. (2000) The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. Plant Cell Online 12:1863-1878

Nesi N, Jond C, Debeaujon I, et al. (2001) The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 13:2099–2114

Opio C, Gerber P, Mottet A, et al. (2013) Greenhouse gas emissions from ruminant supply chains - A global life cycle assessment. Food and Agriculture Organization of the United Nations (FAO), Rome

Pang Y, Peel GJ, Sharma SB, et al. (2008) A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of Medicago truncatula. Proc Natl Acad Sci U S A 105:14210–14225

Pang Y, Peel GJ, Wright E, et al. (2007) Early steps in proanthocyanidin biosynthesis in the model legume Medicago truncatula. Plant Physiol 145:601–615

Pang Y, Wenger JP, Saathoff K, et al. (2009) A WD40 repeat protein from Medicago truncatula is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. Plant Physiol 151:1114–1129

Paolocci F, Robbins MP, Madeo L, et al. (2007) Ectopic Expression of a Basic Helix-Loop-Helix Gene Transactivates Parallel Pathways of Proanthocyanidin Biosynthesis . Structure , Expression Analysis , and Genetic Control of Leucoanthocyanidin 4-Reductase and Anthocyanidin Reductase Genes in Lotus corn. Plant Physiol 143:504–516

Paolocci F, Robbins MP, Passeri V, et al. (2011) The strawberry transcription factor FaMYB1 inhibits the biosynthesis of proanthocyanidins in Lotus corniculatus leaves. J Exp Bot 62:1189-1200

Patra AK, Saxena J (2010) A new perspective on the use of plant secondary metabolites to inhibit methanogenesis in the rumen. Phytochemistry 71:1198–1222

Peel GJ, Pang Y, Modolo L V, Dixon RA (2009) The LAP1 MYB transcription factor orchestrates anthocyanidin biosynthesis and glycosylation in Medicago. Plant J 59:136–149

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30:1–10

Pourcel L, Routaboul JM, Kerhoas L, et al. (2005) Transparent Testa10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. Plant Cell 17:2966–2980

Quattrocchio F, Baudry A, Lepiniec L, Grotewold E (2006) The regulation of flavonoid biosynthesis. In: Grotewold E (ed) Sci. Flavonoids. Springer, Columbus, Ohio, USA, pp 97–122

Robbins MP, Paolocci F, Hughes J-W, et al. (2003) Sn, a maize bHLH gene, modulates anthocyanin and condensed tannin pathways in Lotus corniculatus. J Exp Bot 54:239–248

Sivakumaran S, Rumball W, Lane GA, et al. (2006) Variation of Proanthocyanidins in Lotus Species. J Chem Ecol 32:1797–1816

Steinfeld H, Wassenaar T (2007) The role of livestock production in carbon and nitrogen cycles. Annu Rev Environ Resour 32:271–296

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tanner GJ, Francki KT, Abrahams S, et al. (2003) Proanthocyanidin Biosynthesis in Plants. J Biol Chem 278:31647–31656

Verdier J, Zhao J, Torres-Jerez I, et al. (2012) MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in Medicago truncatula. Proc Natl Acad Sci U S A 109:1766–1771

Wang Y, Hua W, Wang J, et al. (2013) Deep sequencing of Lotus corniculatus L. reveals key enzymes and potential transcription factors related to the flavonoid biosynthesis pathway. Mol Genet genomics MGG 288:131–139

Xie D-Y, Sharma SB, Paiva NL, et al. (2003) Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. Science (80-) 299:396–399

Xie Y, Tan H, Ma Z, Huang J (2016) DELLA Proteins Promote Anthocyanin Biosynthesis via Sequestering MYBL2 and JAZ Suppressors of the MYB/bHLH/WD40 Complex in Arabidopsis thaliana. Mol Plant 9:711–721. doi: 10.1016/j.molp.2016.01.014

Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci 20:176–185

Yoshida K, Iwasaka R, Kaneko T, et al. (2008) Functional Differentiation of Lotus japonicus TT2s, R2R3-MYB Transcription Factors Comprising a Multigene Family. Plant Cell Physiol 49:157–169

Yoshida K, Iwasaka R, Shimada N, et al. (2010a) Transcriptional control of the dihydroflavonol 4-reductase multigene family in Lotus japonicus. J Plant Res 123:801–805

Yoshida K, Kume N, Nakaya Y, et al. (2010b) Comparative analysis of the triplicate proathocyanidin regulators in Lotus japonicus. Plant Cell Physiol 51:912–922

Zhang F, Gonzalez A, Zhao M, et al. (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development 130:4859–69

Zhao J, Dixon RA (2009a) MATE transporters facilitate vacuolar uptake of epicatechin 3’-O-glucoside for proanthocyanidin biosynthesis in Medicago truncatula and Arabidopsis. Plant Cell 21:2323–2340

Zhao J, Dixon RA (2009b) The “ ins ” and “ outs ” of flavonoid transport. Trends Plant Sci 15:72–80

Zhao J, Huhman D, Shadle G, et al. (2011) MATE2 Mediates Vacuolar Sequestration of Flavonoid Glycosides and Glycoside Malonates in Medicago truncatula. Plant Cell 23:1536–1555

Zhou M, Wei L, Sun Z, et al. (2015) Production and transcriptional regulation of proanthocyanidin biosynthesis in forage legumes. Appl Microbiol Biotechnol 3797–3806

Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF (2004) Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. Plant J 40:22–34

Figures

Fig. 1. Phylogenetic tree of *R2R3-MYB* genes and conserved amino acidic signature among the *MYB* genes. A. Tree built using the Maximum Likelihood method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. **B.** Conserved amino acid signature ([D/E]L_{x2}[R/K]$_{x3}$$L_{x6}$$L_{x3}$R) for the interaction of MYBs with bHLH proteins. Partial amino acidic deduced sequences of TT2a, TT2b, TT2c, MYB14 and PAP2 proteins. The residues required for the interaction with bHLH transcription factors are marked with a triangle (Zimmermann et al. 2004).

Fig. 2. Relative expression of *TT2s* and *MYB14* genes in leaves of parents (Lc: *L. corniculatus*, Lt: *L. tenuis*) and F_1 hybrid plants (LH1, LH2, LH3 and LH4) under growth chamber condition. **A.** Relative expression of *TT2a*; **B.** *TT2b*; **C.** *TT2c*; **D.** *MYB14*. The *EFACT* gene was used as reference and the parental plant Lt as calibrator. mRNA levels are expressed as Log_2 (ΔCt Target / ΔCt *EFACT*). The bars represent the mean ± SD (n=3). Means with the same letter are not significantly different (P<0.05). **E.** Leaves of different *Lotus* genotypes stained with DMACA-HCl. Blue dots denote PA accumulating cells. The amount of PAs (mg PA/ g DM) are given in brackets.

Fig. 3. Relative expression of *TT2s* and *MYB14* genes in parents (Lc: *L. corniculatus*, Lt: *L. tenuis*) and F_2 plants (A27, A120, A186, B3, B5 and B147) grown under field condition. A. *TT2a*; **B.** *TT2b*; **C.** *TT2c* and **D.** *MYB14*. Gene

expression analysis and statistics are as reported in the legend of Figure 2. **E:** Leaves of different *Lotus* genotypes stained with DMACA-HCl. The levels of PAs (mg PA / g DM) are given in brackets.

Fig. 4. Relative expression of *PAP2* gene. A. *Lotus corniculatus* (Lc), *L. tenuis* (Lt) and four F_1 hybrid plants (LH1, LH2, LH3 and LH4) under growth chamber condition; **B.** Lc, Lt, three F_2 class A plants with high levels of PAs (A27, A120 and A186) and three F_2 class B plants with low levels of PAs (B3, B5 and B147). Gene expression analysis and statistics are as reported in the legend of Figure 2.

Fig. 5. Relative expression of *TT8* and *TTG1* genes in parental, F_1 and F_2 plants. A. and **B.** expression of *TT8*; **C.** and **D.** expression of *TTG1*. Lc: *Lotus corniculatus*; Lt: *L. tenuis*; LH1, LH1, LH3 and LH4: F_1 hybrid plants; A27, A120, A186, B3, B5 and B147: F_2 hybrid plants. Gene expression analysis and statistics are as reported in the legend of Figure 2.

Fig. 6. Relative expression of *MATE1* gene in leaves of parental, F_1 and F_2 plants. Gene expression analysis and statistics are as reported in the legend of Figure 2.

Fig. 7. PA structural and transporter genes regulated by TT2s, MYB14 and TT8. Scheme of the flavonoid pathway leading to PA production. Key enzymes of the pathway are given in italics and are as follows: *PAL*: L-Phenylalanine ammonialyase; *C4H*: cinnamate 4-hydroxylase; *4CL*: 4-coumarate:CoA ligase; *CHS*: chalcone synthase; *CHI*: chalcone isomerase; *F3H*: flavanone 3-hydroxylase; *F3'H*: flavonoid 3'hydroxylase; *F3'5'H*: flavonoid 3'5' hydroxylase; *FLS*: flavonol synthase; *DFR*: dihydroflavonol reductase; *LAR*: leucoanthocyanidin reductase; *ANS*: anthocyanidin synthase; *ANR*: anthocyanidin reductase; *MATE1*: multidrug and toxic compound extrusion transporter; *GT*: glucosyltransferase; H^+*ATPase*: plasma membrane H^+ pump; *GST*: glutathione-S-transferase. In bold are given the biosynthetic genes considered. Symbols indicates the transcriptional factors that correlate with the expression of structural and transporter genes (paler colour indicates lower correlation coefficient). Question mark indicates the putative role of *Lotus* LAR2 in the production of catechin units (Paolocci et al. 2007).