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 corniculatus

Francesco Paolocci*, Mark P. Robbins, Laura Madeo, Sergio Arcioni, Stefan Martens, and Francesco Damiani

National Research Council, Plant Genetics Institute, 130–06128 Perugia, Italy (F.P., L.M., S.A., F.D.); Institute of Grassland and Environmental Research, Aberystwyth, Ceredigion SY23 3EB, United Kingdom (M.P.R.); and Philipps Universität Marburg, Institut für Pharmazeutische Biologie, 35037 Marburg/Lahn, Germany (S.M.)

Proanthocyanidins (PAs) are plant secondary metabolites and are composed primarily of catechin and epicatechin units in higher plant species. Due to the ability of PAs to bind reversibly with plant proteins to improve digestion and reduce bloat, engineering this pathway in leaves is a major goal for forage breeders. Here, we report the cloning and expression analysis of anthocyanidin reductase (ANR) and leucoanthocyanidin 4-reductase (LAR), two genes encoding enzymes committed to epicatechin and catechin biosynthesis, respectively, in Lotus corniculatus. We show the presence of two LAR gene families (LAR1 and LAR2) and that the steady-state levels of ANR and LAR1 genes correlate with the levels of PAs in leaves of wild-type and transgenic plants. Interestingly, ANR and LAR1, but not LAR2, genes produced active proteins following heterologous expression in Escherichia coli and are affected by the same basic helix-loop-helix transcription factor that promotes PA accumulation in cells of palisade and spongy mesophyll. This study provides direct evidence that the same subclass of transcription factors can mediate the expression of the structural genes of both branches of PA biosynthesis.

Flavonoids represent one of the oldest, largest, and most diverse families of plant secondary metabolites. These compounds fulfill a multitude of functions during plant development (Winkel-Shirley, 2001) and are also of significant interest as antioxidant and anticancer agents in the human diet (Dixon et al., 2005; Tohge et al., 2005). Proanthocyanidins (PAs), also known as condensed tannins, are oligomeric and polymeric end products of the flavonoid biosynthetic pathway. They are widespread in nature, occurring in numerous plant species, including many important plant-derived food materials (Tanner, 2004).

The beneficial effects of PAs on human health and the significant pharmacological activities of PAs have been recently reviewed (Marles et al., 2003; Dixon et al., 2005). In planta, PAs act as protectants against pathogens, pests, and diseases and additionally control seed permeability and dormancy (Debeaujon et al., 2000). These compounds strongly affect plant quality traits, and the palatability and nutritive value of forage legumes are highly influenced by their concentration and structure (Barry and McNabb, 1999). High concentrations of PAs can decrease the palatability and digestibility of plants. By contrast, moderate quantities of PAs (2%–4% dry matter) in forage prevent proteolysis during ensiling and rumen fermentation, thereby protecting ruminants against pasture bloat (Tanner et al., 1995). In fact, by binding to proteins in the rumen, PAs reduce the rate of fermentation and increase the levels of proteins passing through the rumen of grazing animals. PAs therefore make the conversion of plant protein into animal protein more efficient, with reduced methane production (Kingston-Smith and Thomas, 2003). However, PAs only accumulate in the seed coats of the most valuable forage species, such as alfalfa (Medicago sativa) and clovers (Trifolium spp.), and are absent from their leaves. The environmental

---

1 This work was supported by the Italian Ministry of University and Research (Fondo per gli Investimenti della Ricerca di Base project “Post genomica di leguminose foraggere” code n.RBAU01B/KAN to the Plant Genetics Institute) and by the Biotechnology and Biological Science Research Council (to the Institute of Grassland and Environmental Research for work in the area of plant natural products). Travel funding from Consiglio Nazionale delle Ricerche Short-Term Mobility Program. This is contribution number 77 from Plant Genetics Institute-Consiglio Nazionale delle Ricerche.

* Corresponding author; e-mail francesco.paolocci@igv.cnr.it; fax 39–075–5014869.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Francesco Paolocci (francesco.paolocci@igv.cnr.it).

[W] The online version of this article contains Web-only data. www.plantphysiol.org/cgi/doi/10.1104/pp.106.090886
and agronomical benefits that could be derived from triggering the accumulation of a moderate amount of PAs in forage leaves are of considerable importance in the protection and nutrition of ruminant animals (Damiani et al., 2000).

PAs vary widely in composition, depending upon the nature (stereochemistry and hydroxylation pattern) of the flavan 3-ol starter and extension units, the position and stereochemistry of the linkage to the lower unit, the degree of polymerization, and modifications of the 3-hydroxyl group (for recent and extensive reviews, see Tanner, 2004; Dixon et al., 2005). The steps involved in PA polymerization, hydroxylation, and sequestration are still to be fully elucidated. Conversely, understanding of the biosynthesis of the building blocks of PAs has progressed significantly over the last few years. The building blocks or starter units of most PAs are the flavan-3-ols (+) catechin and (−) epicatechin, whose last common biosynthetic intermediate is leucoanthocyanidin (an intermediate in anthocyanin biosynthesis also referred to as flavan 3,4-diol), which results from the reduction of flavonol by dihydroflavonol 4-reductase (DFR). Subsequently, catechin is derived from the reduction of leucoanthocyanidin through the activity of leucoanthocyanidin reductase (LAR), an enzyme that has been purified and characterized in the legume Desmodium uncinatum (Tanner et al., 2003). Epicatechin formation occurs via anthocyanidin synthesis and reduction, two steps catalyzed by anthocyanidin synthase (ANS) and anthocyanidin reductase (ANR), encoded by leucoanthocyanidin dioxygenase (LDOX)/ANS and BANYULS (BAN)/ANR genes, respectively (Abrahams et al., 2003; Xie et al., 2003). Anthocyanidins are consequently essential for PA biosynthesis in species that accumulate only epicatechin-based PAs.

Although the PA biosynthetic pathway has been best characterized in Arabidopsis (Arabidopsis thaliana), this species accumulates only epicatechin-based starter units and lacks any obvious LAR ortholog (Abrahams et al., 2003; Tanner et al., 2003). Thus, Arabidopsis does not offer a system for the complete understanding of the genetic and environmental determinants that control PA biosynthesis in important crops where parallel catechin- and epicatechin-based pathways contribute to PA biosynthesis (Fig. 1). Additionally, Arabidopsis accumulates PAs only in seed coats, whereas these polymers accumulate in vegetative tissues of some important forage crops and pathway regulation may differ between seed coats and other nonfloral tissues, such as leaves, stems, and roots.

The analysis of Arabidopsis seed coat mutants has yielded critical information regarding the transcriptional control of genes involved in PA biosynthesis (Lepiniec et al., 2006). A ternary transcriptional complex comprising TTG1 (encoding a WD-repeat protein), TT2 (encoding a R2R3 MYB-class protein), and TT8 (encoding a basic helix-loop-helix [bHLH] class protein) has been shown to regulate tissue-specific accumulation of PAs. Mutations in any of these three genes result in a transparent testa phenotype with either a reduced or complete absence of PA polymers. The MYB, bHLH, and WD complex also regulates anthocyanin biosynthesis, trichome initiation, and non-root hair cell specification in Arabidopsis. In these cases, the WD-repeat protein (TTG1) is the same in all of the different functional complexes, whereas the bHLH and MYB proteins are flexible in their participation. Therefore, mutations in TTG1 affect the functioning of all the pathways, whereas mutations in the genes encoding the bHLH proteins affect overlapping subsets of the pathways and the specificity in regulating the different pathways in different tissues is conferred by different MYB proteins (Zhang et al., 2003). In this context, mutations in TTG1 and TT8 also affect anthocyanin accumulation in seedlings, consistent with the expression of TT8 and TTG1 in both seed and seedlings. Conversely, TT2 is specifically expressed in the innermost cell layer of the inner seed tegument where it controls PA accumulation by regulating the expression of DFR, LDOX, and BAN. TT8 is similarly important for the expression of DFR and BAN in siliques and TTG1 controls BAN expression mainly by affecting TT8 function (Baudry et al., 2004).

The genus Lotus offers a wide range of options for studying PAs and includes species that accumulate different quantities of condensed tannins; some species accumulate these end products in the majority of plant tissues, whereas others (Lotus japonicus, Lotus tenuis) do not produce condensed tannin polymers in leaf tissues (Harborne, 1988). Another point of interest is that, in Lotus corniculatus and L. japonicus, PA polymers are synthesized from both epicatechin and catechin starter units (Foo et al., 1996). In L. corniculatus, natural or transgenic lines polymorphic for PA quantity in vegetative tissues have already been selected and characterized (Carron et al., 1994; Robbins et al., 2003). We have demonstrated previously that the ectopic expression of the maize (Zea mays) flavonoid regulator Sn, a bHLH gene, increases the levels of PAs in the vegetative tissues of L. corniculatus and regulates the expression of DFR and ANS (Paolocci et al., 2005).

Here, we report on the cloning and characterization of genes encoding ANR and LAR from L. corniculatus. Expression analysis of these genes in wild-type and transgenic lines showing different levels of leaf PAs and the analysis of cell types in leaf tissues, which accumulate detectable levels of PAs, have developed our understanding of the genetic control of PA biosynthesis in this forage legume. Our findings may also be of importance in understanding the genetic mechanisms controlling PA accumulation in other tanniferous plants, where both branches of the PA pathway are operating.

RESULTS
Cloning of ANR cDNAs
By aligning the ANR cDNA sequences from Medicago truncatula (AY184243) and Phaseolus coccineus...
degenerate primers were designed and used to amplify an orthologous leaf cDNA fragment from *L. corniculatus* S41, which was then extended to the 5' and 3' ends by RACE. The ANR amplicons resulting from end-to-end PCR on leaf cDNA were cloned in a plasmid vector and six clones were selected. Four clones harbored a fragment of about 1.2 kb, whereas the other two harbored fragments of about 1,100 bp and 500 bp, respectively. Sequence analysis showed that length differences among clones resided basically in the coding sequences. The 5' and 3' untranslated regions (UTR) ranged from 73 to 79 bp and from 104 to 109 bp; the longest clones, named *LcANR1-1* to *LcANR1-4*, encoded a 338-amino acid protein and the other two, named *LcANR2* and *LcANR3*, encoded proteins of 302 and 107 amino acids, respectively (Supplemental Fig. S1). The four *LcANR1* clones differed slightly from each other, with a total of six conservative amino acid substitutions, and all showed a high (82%) sequence identity at the amino acid level to the ANR cDNAs from *M. truncatula* and *P. coccineus*. The identity of *LcANR2* and *LcANR3* to the above-mentioned cDNAs was still high. However, *LcANR2* differed from *LcANR1* and ANR cDNAs from other species, such as *M. truncatula*, *P. coccineus*, and grapevine (*Vitis vinifera*), because it lacked most of the coding sequence that showed homology to the NAD-dependent epimerase/dehydratase domain as per the National Center for Biotechnology Information (NCBI) conserved domain search. The same search did not show any putative conserved domain in *LcANR3*.

**Cloning of cDNAs for LAR**

A BLASTn similarity search was carried out to identify putative homologous sequences to *LAR* cDNA from *D. uncinatum* (AJ550154), the only *LAR* gene present in public databases when this research was initiated. This search revealed that at least five regions in the *M. truncatula* bacterial artificial chromosome clone (BN000164), degenerate primers were designed and used to amplify an orthologous leaf cDNA fragment from *L. corniculatus* S41, which was then extended to the 5' and 3' ends by RACE. The ANR amplicons resulting from end-to-end PCR on leaf cDNA were cloned in a plasmid vector and six clones were selected. Four clones harbored a fragment of about 1.2 kb, whereas the other two harbored fragments of about 1,100 bp and 500 bp, respectively. Sequence analysis showed that length differences among clones resided basically in the coding sequences. The 5' and 3' untranslated regions (UTR) ranged from 73 to 79 bp and from 104 to 109 bp; the longest clones, named *LcANR1-1* to *LcANR1-4*, encoded a 338-amino acid protein and the other two, named *LcANR2* and *LcANR3*, encoded proteins of 302 and 107 amino acids, respectively (Supplemental Fig. S1). The four *LcANR1* clones differed slightly from each other, with a total of six conservative amino acid substitutions, and all showed a high (82%) sequence identity at the amino acid level to the ANR cDNAs from *M. truncatula* and *P. coccineus*. The identity of *LcANR2* and *LcANR3* to the above-mentioned cDNAs was still high. However, *LcANR2* differed from *LcANR1* and ANR cDNAs from other species, such as *M. truncatula*, *P. coccineus*, and grapevine (*Vitis vinifera*), because it lacked most of the coding sequence that showed homology to the NAD-dependent epimerase/dehydratase domain as per the National Center for Biotechnology Information (NCBI) conserved domain search. The same search did not show any putative conserved domain in *LcANR3*.

**Figure 1.** Pathways to PA starter units (catechin type and epicatechin type) from flavan 3,4-diols in two model species. Superscript letters give references for genes encoding these enzymatic activities in given species and/or reference for structure of PA polymers. a, Abrahams et al. (2003); b, Devic et al. (1999); c, Tanner (2004); d, this paper; e, Paolocci et al. (2005); f, Foo et al. (1996); and g, Meagher et al. (2004).
mth2-103j7 (AC146570.4) displayed high nucleotide similarity with the query sequence (data not shown). From these highly homologous regions, degenerate primer pairs were designed to amplify leaf cDNA from the S41 genotype. The cDNA amplicons resulting from amplification with the primer pairs 2LarF/3larR (about 800 bp long) and 5larF/6larR (about 500 bp long) were then used as targets for 5′- and 3′-RACE PCR amplification. Primers were designed for each set of both 5′ and 3′ ends at the extremity of the longest amplicons and used to amplify the respective full-length cDNAs. Using the different forward/reverse primer combinations, amplicons of two lengths were recovered, the longest, generated by 5Lar1cl/3larbis primers, was name LcLAR1, the other, named LcLAR2, was generated by the 14fwUTR/x3Lar primer combination.

Both LAR1 and LAR2 fragments were cloned and two clones for each gene, named LcLAR1-1, LcLAR1-2 and LcLAR2-1, LcLAR2-2, were sequenced. The 1,418-bp-long LcLAR1-1 cDNA contained 99 nucleotides upstream of the first in-frame ATG, with an open reading frame (ORF) of 1,047 bp (349 amino acids) and a 3′ UTR of 272 bp, the poly(A) tail not included. Both LAR1-1 and LAR1-2 showed the microsatellite-like motif (TG)4(TGAG)5 in the 5′ UTR. LAR1-2 differed from LAR1-1 by two amino acid substitutions (T-S and N-K at positions 192 and 333, respectively). As regards the LcLAR2 amplification products, LAR2-1 cDNA was 1,120 bp long and contained an ORF of 339 amino acids with a 14-bp 5′ UTR and 79 nucleotides after the stop codon and before the poly(A) tail. LAR2-2 cDNA was 1,120 bp long and contained an ORF of 339 amino acids with a 19-bp 5′ UTR and 81 nucleotides after the stop codon and before the poly(A) tail. At a protein level, three amino acid substitutions were detected between the LAR2-1 and LAR2-2 cDNA clones (Q-L, I-V, and P-S at positions 37, 79, and 229, respectively). The LAR gene cloned from Desmodium was most closely related to isoflavone reductase-like proteins (Tanner et al., 2003) and all the Lotus LAR cDNA clones in this study displayed the RFLP, ICCN, and THD motifs conserved in putative LAR sequences, but absent from the related isoflavone reductase proteins (Bogs et al., 2005). The cloned LAR1 and LAR2 sequences showed 71% sequence identity at the amino acid level when compared with each other. A BLASTP search showed that both LAR1 genes had the highest sequence identity at the amino acid level to a LAR gene cloned from Lotus uliginosus (96%) and, to a lesser extent, to the LAR genes from M. truncatula and P. coccineus (66%). LAR1-1 cDNA, by contrast, displayed a higher sequence identity to LAR genes cloned from M. truncatula and P. coccineus (78% and 76%, respectively). A phylogenetic tree showing the relation of LcLAR1 and LcLAR2 with LAR sequences from other legume species is shown in Figure 2.

Organization of ANR and LAR Genes

The primer pairs designed to amplify the full-length cDNA of ANR and of the two LAR genes were also used to amplify and clone the corresponding genomic regions from L. corniculatus to gain insight into their organization and structure within the genome. Amplification from the S41 genome using end-to-end PCR with ANR-specific primers produced two products. Alignment of the 2,860-bp-long genomic sequence, named LcANR1, with the longest cDNAs, showed the ANR coding region is organized into six exons and that the ORF of LcANR2 encodes a 302-amino-acid polypeptide that lacks the N-terminal part of the second exon with respect to the longest ANR cDNAs (LcANR1). The short ANR genomic fragment, LcANR2, perfectly matched the short ANR cDNA fragment (LcANR3) and was about 540 bp long. In addition to extensive deletion on the 3′ end of the first exon, LcANR2 lacked the exons 2, 3, and 4, and the 5′ end of the fifth exon, while retaining the sixth exon in its entirety with respect to LcANR1 (Fig. 3A).

The LAR1-specific primer pair gave rise to amplicons highly divergent in length. Their sequencing and alignment with the respective LcLAR1 cDNA showed that two genomic fragments coded for LAR1, although with differences in terms of both intron structure and length. The shortest gene (about 2,300 bp), named LcLAR1-1, exhibited a single intron. By contrast, in the second LAR1 gene, the coding sequence was split into five exons (Fig. 3B). The microsatellite-like motif in the 5′ UTR of both genomic fragments was (TG)4(TGAG)5.

A similar approach was employed to isolate and characterize the S41 genomic fragments corresponding to LAR2. The two LAR2 genomic fragments cloned had a quite similar length (about 2,400 bp) and were organized into five exons (Fig. 3C).

Southern Analysis

To determine the gene copy numbers of LAR1, LAR2, and ANR in this tetraploid species, gene-specific probes were used to hybridize against genomic DNA restricted with XbaI and SacI restriction sites, which our sequence analysis showed were absent from the genes. To discriminate between LAR1 and LAR2 genes, the respective probes were PCR amplified from regions of intron sequences that did not show appreciable sequence homology between these two genes. Southern analysis showed the presence of multiple hybridizing bands for all probes (Supplemental Fig. S2). However, due to the high molecular weight of the hybridizing fragments, it was difficult to discriminate single from comigrant bands and determine the exact number of copies.

Expression Profiles of LAR and ANR Genes in Wild-Type S41 Plants

Northern analyses were first performed to assess steady-state expression levels for LAR and ANR in different S41 plant organs harvested at late-flowering stages. Given the sequence similarity between the LAR1 and LAR2 genes, probes for differentially labeling the
respective mRNAs in northern analysis were sought at the 5’ end of both cDNAs, where nucleotide similarities between the two coding sequences were only about 55%. The hybridizing signals from northern blots were faint, especially when the filters were hybridized with the LAR2 probe. A clearer picture emerged when they were probed with ANR, where a transcript of approximately 1,200 bp was detected in all tissues analyzed. Importantly, this analysis demonstrated that the shortest ANR mRNAs, identified through the initial end-to-end reverse transcription (RT)-PCR analyses, were only poorly expressed (data not shown).

Due to the very low expression levels of both LAR1 and LAR2 genes and, to a lesser extent, ANR, we took advantage of sensitivity of the real-time RT-PCR analysis to monitor their expression profiles in the different tissues of S41 plants obtained after botanical fractionation. The LAR1- and LAR2-specific primer pairs were therefore designed against the respective 5’ gene ends and the primer specificity was tested against the cloned LAR1 and LAR2 cDNAs. To test for ANR expression by real-time RT-PCR, specific primers were designed against the 5’ gene end common to all ANR cDNAs cloned from L. corniculatus. Figure 4 shows the expression profiles of LAR1, LAR2, and ANR within each plant organ using elongation factor-1α (EF-1α) as a reference. The figure shows that, regardless of the tissue, ANR was always more highly expressed than the two LAR genes and the expression of LAR2 was always and significantly lower and with a narrower expression window. The presence of ANR and LAR
transcripts in all plant organs analyzed is consistent with the ability of *L. corniculatus* to synthesize PA in these tissues (Morris and Robbins, 1997). The highest steady-state levels of *ANR* and *LAR1* mRNAs were shown by seedless pods, whereas stems showed the lowest levels for both *LAR1* and *LAR2* genes (Fig. 4). When a nuclear ITS sequence was used as a constitutively expressed control gene, the relative expression patterns of the three target genes was not altered across the samples analyzed (data not shown), indicating that *EF-1α* is a reliable housekeeping gene for normalization of the expression of flavonoid genes in different plant organs.

**Expression Levels of LAR and ANR in Leaves of Wild-Type and Sn Transgenic Lines**

Real-time RT-PCR analyses were performed to test the steady-state levels of *ANR*, *LAR1*, and *LAR2* genes in leaves of either naturally or ectopically induced PA polymorphic genotypes. Wild-type S50 and S41 plants displayed a different pattern of PA accumulation in leaves; the S50 genotype produces much lower leaf PA levels than S41, the clonal genotype of the *L. corniculatus* line that constitutively accumulates the highest amount of these compounds in leaves (Robbins et al., 2003). We have previously shown that, as a result of the transformation of the S50 genotype with the maize *Sn* gene, transgenic lines displaying contrasting leaf PA phenotypes with respect to the control were produced and that the same regulator had only subtle effects on anthocyanin accumulation, which were restricted to the leaf midrib, leaf base, and petiole, whereas other secondary metabolites were unaffected (Robbins et al., 2003). Interestingly, *Sn* mRNA was never detected even by real-time RT-PCR analysis in leaves of PA-suppressed lines S50/6 and S50/9, whereas *Sn* was expressed in the PA-enhanced lines S50/10 and S50/11 (Paolocci et al., 2005). Wild-type, β-glucuronidase (*GUS*), and *Sn* transgenic lines were grown contemporaneously in a glasshouse and RNA was isolated from young leaves. Because lines S50/9 and S50/6 are suppressed for PA accumulation and lines S50/10 and S50/11 are enhanced in PA accumulation when
Figure 6. ANR and LAR enzyme assays with recombinant Lotus LcANR1 and LcLAR1-1, 1-2, 2-1, and 2-2. Thin-layer chromatography separation of assay extractions was done in n-butanol:HOAc:water. A, Activity assay for LcANR1. Product identification by cochromatography with epicatechin and derivatization by subsequent spraying with DMACA. Lanes 1 and 2, ANR1, 50 and 100 μL protein, respectively; lane 3, denatured ANR1, 100 μL protein; Ref, authentic epicatechin probe after DMACA spraying. B, Coupled assay of GH_DFR and Lotus LAR1-2, 50 μL of each target gene was normalized against the levels of SnGUS transgenic lines, as well as from Sn nontransformed and S41 lines. Similarly, data from S50 nontransformed and S50 GUS transgenic lines, were pooled. Expression of each target gene was normalized against the levels of the host nontransformed or SnGUS recipient genotypes were compared. Here, we report the isolation and characterization of two Sn expressing genotypes, the number of tannin-containing cells detected in the palisade mesophyll was also significantly increased. However, PA-containing cells overlying vascular tissues (vascular mesophyll) were not substantially affected in PA-enhanced or PA-suppressed genotypes or when S50 and S41 clonal recipient genotypes were compared.

DISCUSSION

According to the most recent models, ANR and LAR participate in two separate pathways to PA biosynthesis in most species (Dixon et al., 2005). However, the specific contribution of each branch of this pathway to PA biosynthesis in different plant tissues and species is not clear. The isolation and characterization of ANS, ANR, and LAR cDNAs from the same species has been reported in grapevine (Bogs et al., 2005) and apple (Takos et al., 2006); however, whether the two branches of the PA pathway are controlled by the same or independent regulators has not yet been addressed. Here, we report the isolation and characterization of ANR and LAR cDNAs and corresponding genomic fragments from L. corniculatus, a legume species that accumulates epicatechin- and catechin-based PA in both reproductive and vegetative tissues. In addition, by means of real-time RT-PCR analyses on...
The body of knowledge on the genetic regulation of late flavonoid biosynthetic genes has been gained mainly from studies on anthocyanin-defective mutants in several species, with the control of flavonoid biosynthesis being regulated principally at the transcriptional level. The model described for MYB-bHLH interaction in the regulation of anthocyanin biosynthesis in maize (Goff et al., 1992; Grotewold et al., 2000) is mirrored by the cooperation between TT2 and TT8 in regulating PA accumulation in the seed coats of Arabidopsis (Baudry et al., 2004). In Arabidopsis, the concerted action of TT2, TT8, and TTG1 regulators positively controls the transcription of DFR, LDOX (ANS), and ANR genes (Nesi et al., 2000, 2001; Debeaujon et al., 2003; Baudry et al., 2004). However, unlike the maize R (bHLH) gene, overexpression of TT8 is not sufficient to functionally complement the Arabidopsis ttg1 mutation and the developmental competence of tissues to accumulate BAN transcripts is rate limited both spatially and temporally by TT2 expression (Nesi et al., 2001). Conversely, in maize, tissue-specific pigmentation is caused in many cases by the tissue-specific expression of alleles of the bHLH genes rather than by the MYB gene (Ludwig and Wessler, 1990). In Arabidopsis, the redundancy of bHLH gene function in TTG1-dependent pathways has been reported and reviewed by Zhang et al. (2003) and the coexistence of anthocyanin and PA biosynthesis and the structural similarity between the regulators of these two pathways has raised the question of the molecular mechanisms leading to their specific activation (Baudry et al., 2004; Gachon et al., 2005). However, the mechanisms controlling the expression of the epicatechin and catechin branches of PA biosynthesis have not been investigated in species where both pathways are operating simultaneously.

We have previously shown that the ectopic expression of Sn, a maize bHLH gene, increases the steady-state levels of expression of DFR and ANS genes, two gene families committed also to the accumulation of anthocyanins (Paolocci et al., 2005). In Lotus plants overexpressing Sn, the levels of transcripts encoding Phe ammonia lyase, acetyl-CoA carboxylase, cinnamyl alcohol dehydrogenase, cinnamoyl-CoA reductase, chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3β-hydroxylase (FHT), and chalcone reductase were fundamentally unchanged (Hughes, 2002). This situation is similar to that in Petunia, where the introduction and expression of the maize Le bHLH regulates late anthocyanin pathway genes, but has more limited effects upon early biosynthetic genes such as CHS, CHI, and FHT (Bradley et al., 1998). In Arabidopsis, the ectopic expression of Lc weakly induced the expression of the AHA10 gene, which codes for a proton pump ATPase required for accumulation of PA (Baxter et al., 2005), but did not induce BAN or accumulation of PA (Sharma and Dixon, 2005). Conversely, with reference to forage legumes, there is a report that alfalfa plants expressing the maize Lc accumulate anthocyanins and PAs in leaf tissues under conditions of environmental stress, such as cold or high light intensity (Ray et al., 2003).

Whereas Sn is not an ortholog of TT8, overexpression of this EGL3 (group IIIf) ortholog nevertheless appears to enhance the level of leaf PAs as well as anthocyanins in a limited number of tissues (Robbins et al., 2003). It is worth noting that TT8 and Sn present major functional differences in their ability to interact with C1 and in their requirement for TTG1 to induce high levels of BAN expression in Arabidopsis protoplasts (Baudry et al., 2004). Nevertheless, it has also been pointed out that, in these experiments, Sn can functionally replace TT8 in TTG1-dependent pathways (for review, see Quattrocchio et al., 2006).

The LcANR1 gene, selected as a representative of the ANR gene class most highly expressed in Lotus leaves, encodes a protein that will convert Cy, its expected chemical activity as an ANR (Fig. 6). We also show that in L. corniculatus there are two LAR gene families that encode proteins that share 71% identity at the amino acid level and that, within LAR1 and ANR gene families, members are differently arranged in terms of the number and position of their introns, most likely as result of gene duplication events (Fig. 3). The phylogenetic analyses of LAR proteins from different legume species show that LcLAR1 is most similar to the LAR proteins from species accumulating PAs in their leaves, such as L. uliginosus and D. uncinatum. In contrast, the LAR2 sequences cluster with proteins cloned from species that lack appreciable PA biosynthesis in stem, leaf, and root tissues (i.e. M. truncatula and P. coccineus; Fig. 2). Although LcLAR1 and LcLAR2

**Figure 7.** Number and distribution of PA-containing cells in leaves of PA polymorphic genotypes. Total leaf PA (mg/g dry weight) for the lines S41 (23.3 ± 0.3), S50 (1.92 ± 0.04), S50/10 (6.64 ± 0.11), and S50/9 (0.80 ± 0.07) determined using the butanol-HCl hydrolysis method as described by Robbins et al. (2003). PM, Palisade mesophyll; VM, vascular mesophyll; SM, spongy mesophyll.
retain the diagnostic amino acid stretches specific to LAR proteins (Bogs et al., 2005), only LcLAR1-1 and LcLAR1-2 were able to convert LCy to catechin (Fig. 6). The lack of activity shown by the products of the LAR2 genes in our in vitro tests suggests that the LAR2 genes encode enzymes with a different specificity to LAR1. Enzymes with different specificity might contribute to the production of basal levels of PAs in a range of plant tissues or to the production of other, as yet uncharacterized, metabolites. Because LAR2 genes are conserved in other legume species, it would appear to be unlikely that they encode inactive enzymes. Intriguingly, Bogs et al. (2005) reported the presence of two LAR genes actively transcribed in grapevine berries, but provided a functional evaluation of only one (VvLAR1). Sequences very similar to LcLAR1 and LcLAR2 have also been found in the L. japonicus genome and in expressed sequence tag databases (S. Sato, personal communication). The finding that in legumes there are two LAR gene families suggests a possible conserved duplication of the catechin pathway in distantly related plant species and that this branch of the PA pathway is more complex and diverse than hitherto expected. The detailed evaluation of the catalytic specificity of the LAR2 class of genes will form the basis of future investigations into functional characterization of the LAR gene family and the genetic components of the PA pathway.

Whether the LAR and ANR pathways are coordinately regulated by the same class of transcription factors is of importance in understanding the synthesis of PAs. To address these points, we first analyzed the expression profiles of LAR1, LAR2, and ANR genes in different organs of L. corniculatus and then we monitored their expression levels in leaves producing different levels of PAs. Real-time RT-PCR analyses allowed us to detect the transcripts of our target genes in all organs analyzed. High ANR expression levels combined with the observation that L. corniculatus mainly accumulates epicatechin extension units (Foo et al., 1996; Meagher et al., 2004) suggests that ANR may play a more significant role than LAR in PA synthesis. Analogously, in apple fruit skin, the pattern of abundance of MdANR and MdLAR transcripts was consistent with their roles in the synthesis of PAs, and it has been suggested that the initiation of PA synthesis is mainly due to the flux through the ANR rather than the LAR branch of the pathway (Takos et al., 2006). The slight decrease of the transcript levels of the three target genes from young unexpanded trifoliate leaves to fully expanded leaves also correlates with the observation that L. corniculatus starts to accumulate PAs at early stages of leaf development.

Given that all three gene families (i.e. ANR, LAR1, and LAR2) are transcribed in Lotus leaves, we sought to gain information on their genetic control by comparing their relative expression levels in leaves of natural and ectopically induced PA polymorphic L. corniculatus lines. The overexpression and silencing of the maize bHLH gene Sn in the S50 background results in a significant up- and down-regulation, respectively, of leaf PAs (Paolocci et al., 1999; Robbins et al., 2005).

**Figure 8.** Comparison of the effects of TT8 in Arabidopsis and Sn expression in Lotus on genes of PA biosynthesis. Gray line indicates the putative activity of the protein coded by LcLAR2.
et al., 2003). S41 and S50 control lines also accumulate different levels of leaf PAs and the basis for this natural, rather than engineered, difference in PA levels is currently unclear. The combined analysis of ANR, LAR1, and LAR2 levels in S41 and S50 control lines, as well as in S50-Sn transgenic lines, clearly showed that the expression of ANR and LAR1 correlates with PA accumulation and with the expression/silencing of Sn in transgenic lines. These data indicate that genes of the bHLH class control the level of PAs in the leaves of *Lotus corniculatus*. In summary, a comparison of the target metabolic genes controlled by *TT8* in Arabidopsis and by the ectopic expression of *Sn* in Lotus leaves is displayed in Figure 8.

To understand whether the ranges of PA-specific mRNA transcripts and total amounts of leaf PAs correlated with the presence of detectable tannin-accumulating cells, we recorded the number and distribution of PA-containing cells in S41, S50, S50/10 (PA-enhanced), and S50/9 (PA-suppressed) lines. The increase of steady-state levels of ANR and LAR1 mRNAs in S41 and S50/10 was paralleled by an increase in the number of PA cells in palisade and spongy mesophyll tissues when compared with the S50 control line (Fig. 6). Similarly, there were less PA cells in palisade and, most significantly, in spongy mesophyll of the S50/9 line than in S50. In stark contrast, the number of PA cells adjacent to veins remained unaltered in all genotypes. Thus, it is tempting to speculate that *Sn* activates the expression of both ANR and LAR1 in palisade and spongy mesophyll and that the high levels of PAs in the S41 genotype are due to constitutive expression of a gene encoding a protein functionally equivalent to *Sn* throughout the leaf blade. The significant reduction of levels of DFR, AN5, ANR, and LAR1 transcripts and PAs in leaves of *Sn*-suppressed lines (Paolocci et al., 2005; this study) suggests that, in these lines, the transgene has caused the silencing of an endogenous bHLH regulator. However, unlike *Sn*, the bHLH protein endogenous to Lotus may also participate in controlling LAR2 expression in leaves because this target gene was significantly down-regulated in S50 *Sn*-suppressed lines with respect to the control S50 (Fig. 5). Therefore, these studies on *Lotus corniculatus* indicate that it is likely that the accumulation of PAs in leaves is limited by the expression of a bHLH gene.

This model is in contrast to the proposal that *TTG1* is the rate-limiting regulatory factor for the expression of phenylpropanoid pathways in Arabidopsis (Zimmermann et al., 2004). The observation that the ectopic expression of *AtTTG1* in S50 has no obvious chemical or developmental phenotype, whereas a range of bHLH genes modify the expression of anthocyanin and PA pathways (G. Allison and M.P. Robbins, unpublished data) lends further support to our thesis.

The successful engineering of PAs in the leaves of forage plants requires a deep understanding of the network of structural and regulatory genes involved. Here we show the presence of two LAR gene families in *Lotus corniculatus* and provide evidence that LAR1 and ANR provide two parallel and coregulated routes for PA biosynthesis in leaves and that this regulation is at least partly under the control of a bHLH transcription factor. Analysis of gene promoters and DNA-binding proteins combined with mutant analysis in the related genetic model *L. japonicus* will assist future studies relating to the mechanisms that control the expression of PAs in the leaves of forage legumes.

### MATERIALS AND METHODS

#### Plant Material

*Lotus corniculatus* genotypes S50 and S41, isolated from the cultivar Leo and characterized for their different levels of PA accumulation in leaves, have been described previously (Carron et al., 1994). S50 and S41 plants transformed with the maize (*Zea mays*) bHLH regulatory gene *Sn* or with the *GUS* reporter gene were obtained and characterized as reported previously (Paolocci et al., 1999, 2005; Robbins et al., 2003).

#### Isolation of RNA

For RNA isolation, the following organs were collected from S41 clonal plants at late-flowering stage: flowers, trifoliate unexpanded leaves, and flower pods deprived of the immature seeds (seedless fruits). For RNA isolation from leaves of *Sn* and *GUS* transgenic lines in both S50 and S41 genetic backgrounds and from control nontransformed S50 and S41 lines, plants were grown contemporaneously in a glasshouse under identical outdoor conditions and only healthy young (unexpanded) trifoliate leaves were harvested. RNA was isolated using the NucleoSpin RNA plant isolation kit (Macherey-Nagel) according to the supplier’s instructions, after which a DNase treatment was added. The quality and quantity of RNA was verified by agarose gel electrophoresis and spectrophotometric analysis.

#### Gene Cloning

**Cloning of ANR Gene Sequences from *Lotus corniculatus***

The primer pair Bn1F (5’-aagagcaagttgtatgmggtggmaStg-3’) and Bn1R (5’-gtcgtcaacactccagyaTg-3’) and the nested primer pair Bn2F (5’-gtctctsa-maagggattgtctg-3’) and Bn2R (5’-ctcctttsyaagaagttctgccagattgatg-3’) were designed and used in a 25-μL volume of PCR reaction to amplify a partial ANR cDNA fragment on first-strand cDNA from S41 leaves obtained using the SMART RACE cDNA amplification kit (CLONTECH). The resulting RT-PCR fragment was sequenced and then extended toward the 5’ and 3’ ends using the above-mentioned kit by using the adaptor primers provided with the kit in combination with ANR gene-specific primers 5’-ggagcactgaaaggtgagaagaacagacaagcaagtctg-3’ and 5’-tggtggggagagaactaaccgcgtcgg-3’ for the 5’ and 3’ RACE, respectively, according to the manufacturer’s instructions. Primers were then designed proximally to the 5’ (5’-acagcagaaaaaacactgtg-3’) and 3’ (5’-gctgtctttgtaattttacctgtg-3’) sequence ends to amplify the full-length cDNA, as well as the ANR gene from *Lotus corniculatus*, which were then cloned into the pGEM T-Easy vector (Promega).

**Cloning of LAR Genes from *Lotus corniculatus***

By aligning the *Desmodium uncinatum* LAR sequence with the *Medicago truncatula* bacterial artificial chromosome clone mth2-1037 (AC146570.4; see “Results”), the primer pairs 1LarF (5’-ggagcaactgctttcaagtg-3’) and 4LarR (5’-catctatgctacacgaaagtcagtcgg-3’) and the nested primer pairs 2LarF (5’-gtctcttactcggctcggctgg-3’) and 5LarF (5’-gtctcttcactcggctcggctgg-3’) and 6Lar (5’-gctggcgtcgtgagaactaatg-3’) were used to amplify LAR sequences from the S41 leaf cDNA (as above). The resulting partial cDNA fragments were sequenced and then extended toward the 5’ and 3’ ends using the SMART 5’-3’ RACE kit (CLONTECH) by using gene-specific primer either the reverse primer 7lar
Southern Analysis

Isolation, blotting, and Southern hybridization of genomic DNA were performed as in Damiani et al. (1999). Aliquots of approximately 8 μg (3.5 × 10^7 genomes) of genomic DNA were restricted independently with the corresponding endonucleases: XbaI and ScaI (New England Biolabs) according to the supplier’s instructions. Digestions were run in parallel to obtain replicated filters. Each one was hybridized with XbaI- and ScaI-digested DNA. Each filter was hybridized with the ANR, LAR1, or LAR2 probes. The LARI (970 bp long) and LAR2 (650 bp long) probes were obtained by amplifying the cloned LARI-1 and LAR2 genes (see “Results”) with the primer pairs 5′-ggggttggttg-gagaacctt-3′ / 5′-agccaaacttagggagca-3′ and 5′-tggtggtggtggattgtgctg-3′ / 5′-agttcataaactcttggc-3′, respectively. Similarly the ANR genomic probe was obtained by amplifying the longest ANR gene (see “Results”) with the 5′- and 3′-extreme primers (see above).

Expression of ANR and LAR Genes in Escherichia coli

The coding regions of LeANR1-1, LeLAR1-1, LeLAR2-1, and LeLARN2-2 were amplified from the clones harboring the corresponding full-length cDNAs with Pfx50 DNA polymerase (Invitrogen Life Technology) in the presence of the primer pairs exANRF (5′-catacggcaagcataccagacctttat); exANRR (5′-catacggcaagcataccagacctttat) / exLAR1F (5′-catacggcaagcataccagacctttat) / exLAR1R (5′-catacggcaagcataccagacctttat); and exLAR2F (5′-catacggcaagcataccagacctttat) / exLAR2R (5′-catacggcaagcataccagacctttat), respectively. The resulting amplicons were cloned into the pET160/D/W vector (Invitrogen) and sequenced to confirm their identity. Expression and protein isolation was done according to the manufacturer’s instructions (Invitrogen).

Recombinant Enzyme Assay

Catechin, epicatechin, and Cy were obtained from Roth. [14C]DHQ was synthesized using recombinant CHS, CHI, FHT, and flavonoid 3′-hydroxylase from various plant sources as described (Martens et al., 2003). Expression of Gerbera DFR for first-step synthesis of LCy in coupled assays with LAR was carried out as described (Martens et al., 2002). Chemical synthesis of unlabelled LCy from the respective dihydroflavonol, DHQ (TransMIT GmbH Flavonoidsforschung), was performed basically according to Tanner and Kristiansen (1993).

The coupled DFR/LAR assay (500 μL in total; 0.1 μL Tris-HCl buffer, pH 7.5) contained either 1C-labeled (0.03 nmol) or unlabelled DHQ (1 nmol), 25 μL 200 mM NADPH, and recombinant DFR and LAR (approximately 200–500 μg total protein of each enzyme according to Bradford, 1976). ANR assay (500 μL in total; 0.1 μL KPi buffer, pH 6.0) contained 50 mM saturated Cy solution (1 mg/mL in KP buffer), 25 μL 200 mM NADPH, recombinant ANR (approximately 300 μg total protein). After incubation at 30°C (LAR) or 45°C (ANR) for 60 min the reaction was stopped and extracted twice with 500 μL ethylacetate. The pooled and concentrated extractions were spotted on cellulose thin-layer chromatography plates (Merck) and developed in n-butanol:H2O:acetic acid:water:2:1:2:1:2. After drying, the plates were either exposed for autoradiography (Phosphor Imager; Fuji) software package AIDA, Raytest) or sprayed with DMACA solution (0.1% in ethanol/6% HCl; 1:1 [v/v]) to visualize the flavonoids. The identity of all substrates and products were done by cochromatography with authentic standards as described (Pfeiffer et al., 2006).

Cell Counts and Tissue Location of PA-Containing Cells in Leaves of L. corniculatus

Lotus PCa leaf counts were carried out on leaves of S41, S50, S50/10, and S50/9 genotypes essentially as described by Robbins et al. (2003). Trifoliate leaves were harvested randomly from selected lines and decolorized overnight in ethanol. After discarding the ethanol, samples were stained for 1 h using 0.3% (w/v) DMACA (Li et al., 1996) and then subjected to three changes of deionized water. Cells containing PA polymers stained blue and cell counts were performed by focusing on adaxial and abaxial surfaces in turn using an Olympus BH light microscope (10× magnification). Results are from 10 leaves per line and expressed ±SEM.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers D3439110 to D3439112 and D3439114.

Phylogenetic Analysis

Full-length LAR cDNA sequences from lege species were retrieved from public databases and two grapevine (Vitis vinifera) LAR sequences (Bogs et al., 2005) were used as outgroups to build a phylogenetic tree. The Clustal IV multiple sequence alignment was formed using the default parameters of the MEGA package (Kumar et al., 2004). The tree was constructed from the Clustal IV alignment using the neighbor-joining method by the MEGA program. One thousand bootstrapped datasets were used to estimate the confidence of each tree clade.

Expression Analysis of ANR, LAR1, and LAR2 in Organs of S41 Genotype and in Leaves of L. corniculatus Lines Showing Contrasting PA Phenotypes

About 20 μg of total RNA from a number of different organs of the L. corniculatus S41 genotype were run and blotted on Hybond N membrane (Amersham Biosciences) according to the supplier’s instructions. The ANR probe (about 1200 bp long) was obtained by amplifying the clone containing LeANR1-1 cDNA using the 5′ and 3′ extreme primer pairs (see above). Similarly, both the LARI (about 239 bp) and LAR2 (128 bp) probes were obtained by amplifying the respective cDNAs from plasmids carrying them with the primer pairs 5′Lar1cl/lar1qRn (5′-catcaagctggctcacttac-3′) and 5′Lar1cl (5′-gaccctggccttagaagcagat-3′) / 5′-tagacctggctccacaggatct-3′ for LARI-1 and LARI-2 genes (see “Results”) with the primer pairs 5′-ggggttggttg-gagaacctt-3′ / 5′-agccaaacttagggagca-3′ and 5′-tggtggtggtggattgtgctg-3′ / 5′-agttcataaactcttggc-3′, respectively. Similarly the ANR genomic probe was obtained by amplifying the longest ANR gene (see “Results”) with the 5′ and 3′ extreme primers (see above).
(ANR cDNAs), DQ349113 and DQ349115 (ANR genes), DQ349100 and DQ349101 (LeLAR1 cDNAs), DQ349102 and DQ349103 (LeLAR1 genes), DQ349104 and DQ349105 (LeLAR1 cDNAs), and DQ349106 and DQ349107 (LeLAR2 genes).

Supplemental Data
The following materials are available in the online version of this article.
Supplemental Figure S1. Analysis of L. corniculatus ANR amino acid sequences.
Supplemental Figure S2. Southern analysis of LcANR, LeLAR1, and LeLAR2.
Supplemental Table S1. Primers used in real-time RT-PCR analysis.

ACKNOWLEDGMENTS
We thank Dr. Andrea Rubini (Plant Genetics Institute) and Professor Phil Morris and colleagues in the Plant Cell Biology Team at the Institute of Grassland and Environmental Research. We also acknowledge Dr. John-Wayne Hughes for access to unpublished data and Dr. Greg Tanner and colleagues (Commonwealth Scientific and Industrial Research Organization, Canberra) and Professor Clint Chapple (Purdue University) for their assistance in the work of the Institute of Grassland and Environmental Research laboratory on PAs in legumes and Arabidopsis, respectively.

Received October 6, 2006; accepted November 2, 2006; published November 10, 2006.

LITERATURE CITED


as key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 13: 2099–2114


