Functional Characterization of Proanthocyanidin Pathway Enzymes from Tea and Their Application for Metabolic Engineering[1][W][OA]

Yongzhen Pang[2], I. Sarath B. Abeysinghe, Ji He, Xianzhi He, David Huhman, K. Mudith Mewan, Lloyd W. Sumner, Jianfei Yun, and Richard A. Dixon[3]*

Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73401 (Y.P., J.H., X.H., D.H., L.W.S., J.Y., R.A.D.); and Biochemistry Division, Tea Research Institute of Sri Lanka, Talawakelle 22100, Sri Lanka (I.S.B.A., K.M.M.)

Tea (*Camellia sinensis*) is rich in specialized metabolites, especially polyphenolic proanthocyanidins (PAs) and their precursors. To better understand the PA pathway in tea, we generated a complementary DNA library from leaf tissue of the blister blight-resistant tea cultivar TRI2043 and functionally characterized key enzymes responsible for the biosynthesis of PA precursors. Structural genes encoding enzymes involved in the general phenylpropanoid/flavonoid pathway and the PA-specific branch pathway were well represented in the library. Recombinant tea leucoanthocyanidin reductase (CsLAR) expressed in *Escherichia coli* was active with leucocyanidin as substrate to produce the 2R,3S-trans-flavan-ol (+)-catechin in vitro. Two genes encoding anthocyanidin reductase, *CsANR1* and *CsANR2*, were also expressed in *E. coli*, and the recombinant proteins exhibited similar kinetic properties. Both converted cyanidin to a mixture of (+)-epicatechin and (−)-catechin, although in different proportions, indicating that both enzymes possess epimerase activity. These epimers were unexpected based on the belief that tea PAs are made from (−)-epicatechin and (+)-catechin. Ectopic expression of *CsANR2* or *CsLAR* led to the accumulation of low levels of PA precursors and their conjugates in *Medicago truncatula* hairy roots and anthocyanin-overproducing tobacco (*Nicotiana tabacum*), but levels of oligomeric PAs were very low. Surprisingly, the expression of *CsLAR* in tobacco overproducing anthocyanin led to the accumulation of higher levels of epicatechin and its glucoside than of catechin, again highlighting the potential importance of epimerization in flavan-3-ol biosynthesis. These data provide a resource for understanding tea PA biosynthesis and tools for the bioengineering of flavanols.

Tea (*Camellia sinensis*) originated in southern China. It is the most widely produced and consumed beverage in the world after water and has received a great deal of attention with respect to its numerous health benefits to humans (Lin et al., 2003). Although the composition of tea varies with cultivar, age of leaf, season, and processing method (Lin et al., 1996), tea contains uniquely high concentrations of polyphenolic compounds, which consist mainly of six flavan-3-ols: epigallocatechin gallate (EGCG; 9%–13% of dry weight), epigallocatechin (EGC; 3%–6%), epicatechin gallate (ECG; 3%–6%), epicatechin (1%–3%), gallocatechin (GC; 1%–2%), and catechin (1%–2%; Balentine et al., 1997; Fig. 1); together, these can constitute more than 30% of the dry weight. A further 3% of the dry weight is composed of proanthocyanidins (PAs; oligomeric or polymeric flavan-3-ols).

Increasing evidence suggests that tea flavan-3-ols are beneficial for human health as a result of their antioxidant capacity (Luximon-Ramma et al., 2006), chemopreventive activities against prostate and ovarian cancers (Bemis et al., 2006; Ravindranath et al., 2006), antiobesity and antidiabetic effects (Yang and Koo, 2000; Kao et al., 2006; Wolfram et al., 2006), and cardiovascular disease prevention properties (Yang and Koo, 2000). Tea flavanols, especially catechins, may also play roles in the defense of tea against infection/infestation (Punyasiri et al., 2004).

The blister blight disease caused by the fungal pathogen *Exobasidium vexans* is the major disease affecting tea quality and yield in Asia (Arulpragasam, 1992). Previous investigations have indicated that flavan-3-ol content and composition differ in different blister blight-resistant and -susceptible tea cultivars and at different stages during infection. The levels of (−)-E (for epicatechin) in resistant cultivars are significantly higher than in susceptible ones, whereas

Figure 1. Structures of the flavan-3-ol monomers.

the levels of EGCG in resistant cultivars are significantly lower than in susceptible ones. Epicatechin and EGC levels declined significantly in the infested areas of the leaf during blister blight infection, whereas the levels of ECG and EGCG increased at the translucent stage of the infection and declined on maturation of the blister (Punyasiri et al., 2005). Furthermore, infection of tea leaves susceptible to blister blight gave rise to a shift in the PA stereochemistry away from 2,3-trans (53% and 61% of the total starter and extension units of the PAs, respectively) toward 2,3-cis (26% and 40%, respectively; Nimal Punyasiri et al., 2004), and infection also resulted in increased gallic acid esterification of catechin and epicatechin.

Flavan-3-ol monomers are synthesized via two distinct branches of the general flavonoid pathway, which share the same upstream biosynthetic pathway to leucoanthocyanidin (Supplemental Fig. S1). Leucoanthocyanidins can be converted either to 2*R*,3*S*-trans-flavan-3-ols [(+)-catechins] by the action of leucoanthocyanidin reductase (LAR; Tanner et al., 2003) or to anthocyanidins by anthocyanidin synthase (Saito et al., 1999). Subsequently, anthocyanidins can be converted to 2*R*,3*R*-cis-flavan-3-ols [(−)-Es] by anthocyanidin reductase (ANR; Xie et al., 2003; Supplemental Fig. S1). Therefore, the two monomer types [(+)-catechin and (−)-E] are produced from two distinct PA-specific pathways despite their only differing in the stereochemistry at the C3 position. ANRs from Arabidopsis (*Arabidopsis thaliana*) and *Medicago truncatula* were shown to produce minor amounts of an epimer of epicatechin identified as (−)-catechin, and it was suggested that

this may be formed by nonenzymatic epimerization (Xie et al., 2003, 2004). However, it has recently been shown that grapevine (*Vitis vinifera*) ANR possesses intrinsic epimerase activity and, produces (+)-E rather than (−)-E as the major product (Gargouri et al., 2009, 2010). The stereochemical outcome of the ANR reaction, therefore, is quite complex, and the implications of this for subsequent downstream steps in PA biosynthesis are unclear. (−)-E can be further glycosylated by the UDP-glycosyltransferase UGT72L1 to form *E*-3′-*O*-glucoside in vitro, the substrate for vacuolar transport by the multidrug and toxic compound extrusion (MATE) family transporters TT12 (Arabidopsis) and MATE1 (*M. truncatula*; Pang et al., 2008; Zhao and Dixon, 2009). Following transport to the vacuole, PA precursors undergo polymerization by mechanisms yet to be fully determined (Pourcel et al., 2005; Zhao et al., 2010).

EST sequencing approaches have been applied to elucidate the (differential) biosynthesis of flavonoid classes in tea. Based on the observation that the level of total catechins in young leaves was higher than in mature leaves, Park et al. (2004) sequenced and analyzed 588 ESTs from a suppression subtractive library of young versus old leaves and compared the transcript levels of several putative flavonoid biosynthetic pathway genes that were annotated based only on sequence similarity. Zhao et al. (2005) sequenced 4,320 clones from a tea library and analyzed the sequences of a total of 2,963 ESTs. Recently, the transcriptome of tea was analyzed using a high-throughput Illumina RNA-seq technique (Shi et al., 2011); however, no detailed functional characterization of gene products was attempted beyond preliminary biochemical assays (Singh et al., 2009; Zhang et al., 2012). Biochemical studies using leaf protein extracts revealed high activities of ANR in tea leaves and linked this to the formation of E and EGC (Punyasiri et al., 2004); the combined enzymatic reactions of dihydroflavonol 4-reductase (DFR) and LAR confirmed that the formation of catechin and GC proceeded from leucoanthocyanidins.

In this study, we sequenced a complementary DNA (cDNA) library from a blister blight-resistant cultivar of tea and functionally annotated the ESTs. Focusing on PA-specific pathway genes, we have biochemically characterized *LAR* and *ANR* gene products and shown that tea contains two *ANR* genes that encode proteins with different levels of epimerase activity, leading to the formation of the less common isomers (+)-E and (−)-catechin. The potential use of these enzymes for pathway engineering in vivo has been addressed through genetic transformation in *M. truncatula* and anthocyanin-overproducing tobacco (*Nicotiana tabacum*).

RESULTS

Generation and Classification of a Tea Leaf cDNA Library

To characterize genes involved in PA biosynthesis in tea, we used an EST approach. The library was generated from RNA isolated from leaves of the blister

blight-resistant cv TRI2043 harvested at three different stages after infection, as described in "Materials and Methods." In total, 6,829 clones were randomly picked for plasmid isolation and sequenced from their 5′ ends. These clones clustered into 668 contigs and 3,293 singletons. In total, 3,931 unigenes were obtained for further analysis. These were annotated through BLASTX search against GenBank's nonredundant database and were then classified into three groups based on the "bit score" (Degtyarenko and Archakov, 1993) of the target unigene sequence segments compared with the top BLAST hit. A total of 1,434 unigenes (36%) had a bit score of equal or greater than 200, 1,794 unigenes (46%) had a bit score of less than 200, and the remaining 703 unigenes (18%) had no score (i.e. no hit when compared with the National Center for Biotechnology Information [NCBI] database). We assume that these latter unigenes represent either currently unknown proteins or noncoding RNAs.

The relative frequency of tea leaf unigene hits in each Gene Ontology (GO) category is presented in Supplemental Figure S2. It should be noted that when a unigene is described as unknown or no hit to the GO database, this indicates that it has not been assigned to a GO term. Some unigenes may be classed into overlapping subcategories. Of the top 20 most abundant unigenes from the tea leaf library, 19 have similarity to genes encoding proteins from other plant species (Supplemental Table S1). Among the most abundant ESTs were Rubisco small subunit, the most abundant protein in plants, two chlorophyll *a/b*-binding proteins, a chloroplast hypothetical protein, a PSII 10-kD polypeptide, and a PSI reaction center subunit XI, all associated with/involved in photosynthesis, as would be predicted for a library made from leaf material. It is interesting that a unigene annotated as LAR was among the top 20 most highly expressed unigenes from the tea leaf library. Several ESTs, encoding putative Blight-Associated Protein12 precursors and chitinases, might be involved in fungal resistance in tea (Supplemental Table S2).

Genes of Specialized Metabolism Are Well Represented in the Tea Leaf cDNA Library

To better interrogate specialized metabolism in tea, the BLASTX annotation of the unigenes was searched for genes encoding enzymes involved in phenylpropanoid/flavonoid, terpenoid, and alkaloid biosynthesis. In addition to the above-mentioned LAR, the most highly abundant unigenes potentially involved in specialized metabolism encoded chalcone synthase (11 tandem contigs and 38 ESTs) and flavonol synthase (seven tandem contigs and 19 ESTs). ANR was represented by a single EST in the present tea library (*CsANR2*). Except for cinnamate 4-hydroxylase, all the other phenylpropanoid/flavonoid pathway genes shown in Supplemental Figure S1 were represented in the tea leaf cDNA library (Supplemental Table S3).

The alkaloid caffeine is an important specialized metabolite from tea, accounting for approximately 2.8% of the dry weight. The main biosynthetic pathway to caffeine consists of four sequential steps with three methylations and one nucleosidase reaction. The four enzymes involved are xanthosine *N*-methylxanthosine synthase (or 7-methylxanthosine synthase), *N*-methylnucleosidase, theobromine synthase (monomethylxanthine *N*-methyltransferase), and caffeine synthase (dimethylxanthine *N*-methyltransferase; Kato et al., 1999; Ashihara and Crozier, 2001; Yoneyama et al., 2006; Ashihara et al., 2008). Three unigenes encoding theobromine synthase and three encoding caffeine synthase were represented in the tea cDNA library, and these have been shown to be functional (Kato et al., 1999; Yoneyama et al., 2006); however, the other two enzymes were not represented.

Sequence Analysis of Tea *ANR* and *LAR* Genes

By using 3′ and 5′ RACE approaches, the full-length cDNA of the *CsLAR* gene was obtained (GenBank accession no. GU992401), corresponding to the longest and most highly expressed of the LAR unigenes in the library (Supplemental Table S3). It is 1,282 bp in length with a 5′ untranslated region (UTR) of 100 bp and a 3′ UTR of 190 bp. The open reading frame (ORF) of the *CsLAR* gene is 1,014 bp long and encodes a protein of 342 amino acids with a calculated molecular mass of 38 kD and a pI of 5.43. It shows 63%, 65%, 66%, and 71% amino acid identity to LARs from *M. truncatula*, *Lotus corniculatus*, *Desmodium uncinatum*, and grape (*VvLAR1*), respectively (Supplemental Fig. S3). The additional LAR unigenes in the library represent different portions of the above gene with minor nucleotide differences.

An available tea EST sequence similar to ANR (GenBank accession no. AAO13092; released on November 1, 2004), originally annotated as LAR but without functional information in NCBI, was used for the design of primers to amplify its ORF from tea 'TRI2043' in order to compare with the *ANR* gene from the present cDNA library. The ORF of this previously reported gene, designated as *CsANR1* here (GenBank accession no. GU992402), was 1,044 bp in length and encoded a protein of 347 amino acids with a molecular mass of 37 kD and a pI of 5.37. It showed 60%, 72%, 73%, and 79% amino acid identity to ANRs from Arabidopsis, *M. truncatula*, *L. corniculatus*, and grape at the amino acid level, respectively. The *CsANR1* amplified from TR12043 differed from the sequence of AAO13092 by 10 nucleotides and three amino acid substitutions.

The full-length cDNA of the second tea *ANR* gene (designated as *CsANR2*; GenBank accession no. GU992400) was obtained by 3′ and 5′ RACE after sequence analysis of the partial EST from the present tea cDNA library. It is 1,265 bp in length, with a 5′ UTR of 114 bp and a 3′ UTR of 137 bp. The ORF of the *CsANR2* gene is 1,014 bp long and encodes a protein of 337 amino acids with a calculated molecular mass of 36

kD and a pI of 6.54. It shows 64%, 75%, 76%, and 83% amino acid identity to ANRs from Arabidopsis, *M. truncatula*, *L. corniculatus*, and grape, respectively (Supplemental Fig. S4). *CsANR2* shows 83% identity to *CsANR1* at the amino acid level.

LAR and ANR proteins are related to members of the reductase-epimerase-dehydrogenase protein superfamily. Phylogenetic analysis with the closely related ANR, DFR, and isoflavone reductase (IFR) protein sequences showed that *CsLAR*, *CsANR1*, and *CsANR2* group into the corresponding LAR and ANR clusters and are clearly separated from the DFR or IFR clusters (Fig. 2).

Functional Characterization of Recombinant Tea LAR and ANR in Vitro

The ORFs of *CsLAR*, *CsANR1*, and *CsANR2* were cloned into the bacterial expression vector pQE30 containing a six-His tag at the N terminus and expressed in *Escherichia coli* strain M15. For functional characterization of the recombinant *CsLAR* protein, [3H]leucocyanidin was used as the substrate. After incubation with the purified *CsLAR* protein, a small peak corresponding to authentic (+)-catechin in retention time and UV spectrum was observed by HPLC analysis (Fig. 3A, bottom panel), and this peak contained the peak of the radioactivity as determined by scintillation counting of fractions (Fig. 3B). No compound peak or radioactivity at this retention time were observed from control incubations with boiled enzyme (Fig. 3A, top panel). These data are consistent with the functional identification of LAR from other species (Bogs et al., 2005; Pang et al., 2007; Paolocci et al., 2007).

After incubation of purified recombinant CsANR1 protein with cyanidin and analysis of the products by reverse-phase HPLC, a major peak with the same
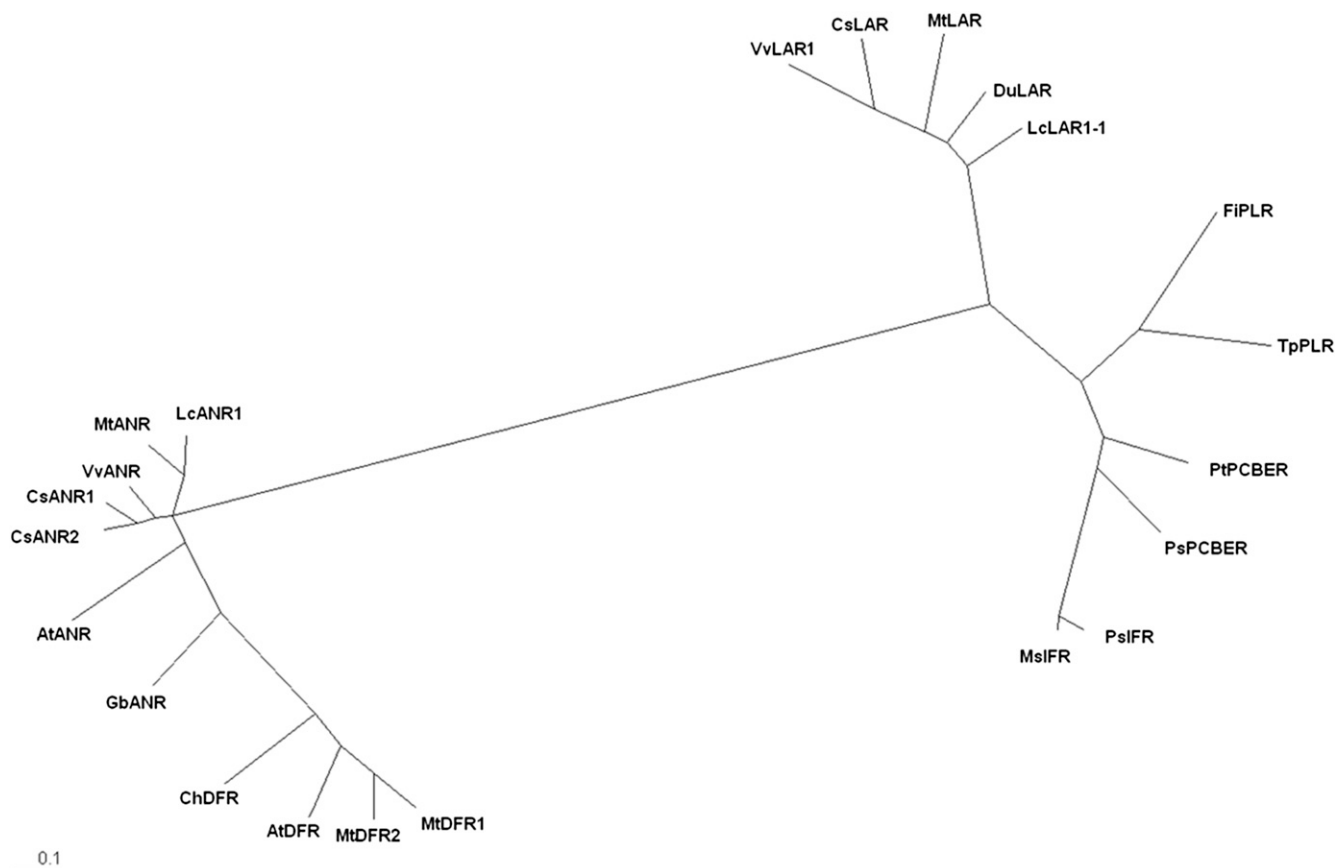
Figure 2. Unrooted phylogram comparison of LAR, ANR, DFR, and related proteins from the reductase-epimerase-dehydrogenase superfamily. The tree was constructed from the ClustalW alignment using the neighbor-joining method and MrBayes software. The proteins are labeled according to the species followed by the name of the protein (e.g. CsDFR is DFR from tea; AAT66505). The protein names not listed in Supplemental Figures S3 or S4 are as follows: ANR from *Ginkgo biloba* (AAU95082); DFR from Arabidopsis (P51102); DFR1 and DFR2 from *M. truncatula* (AAR27014 and AAR27015); DFR from *Cymbidium hybrid* (AAC17843); IFR from *Pisum sativum* (AAB31368); IFR from alfalfa (P52575); pinoresinol:lariciresinol reductase (PLR) from *Forsythia × intermedia* (AAC49608); PLR from *Thuja plicata* (AAF63507); phenylcoumaran benzylic ether reductase (PCBER) from *P. trichocarpa* (CAA06707); and PCBER from *Pinus taeda* (AAF64173). The scale bar indicates the estimated number of amino acid substitutions per site.

retention time and identical UV spectrum as epicatechin, and a minor peak possibly corresponding to catechin, were observed (Fig. 3C, top panel). Mass spectrometry (MS) analysis confirmed that both compounds had the same molecular ion. Control incubations with boiled protein gave no conversion of cyanidin (Fig. 3C, bottom panel). *CsANR1* was also able to convert the anthocyanidins delphinidin and pelargonidin into the corresponding flavan-3-ols EGC and epiafzelechin, with putative epimers GC and afzelechin as minor products (Fig. 3, D and E).

In contrast to CsANR1, purified recombinant CsANR2 protein used delphinidin and cyanidin as substrates to produce major products with the same retention times and UV spectra as GC and catechin and with lower levels of products with the retention times of EGC and epicatechin (Fig. 3, F and G, top panels). No products were formed with boiled protein (Fig. 3, F and G, bottom panels). Interestingly, the reaction with pelargonidin as substrate gave epiafzelechin as the major product with afzelechin as a minor product; again, no products were generated in incubations with boiled enzyme (Fig. 3H).

To determine the chirality of the different products generated from cyanidin, we analyzed the reaction products by HPLC on a chiral column, along with authentic standards. The results indicated that both CsANR1 and CsANR2 produce (+)-E and (−)-catechin, but in varying proportions (Fig. 3, I and J).

Kinetic analyses revealed similar V_{max} values for CsANR1 and CsANR2 with cyanidin and delphinidin. Although the K_m values for cyanidin, delphinidin, and pelargonidin were somewhat lower for CsANR2 than for CsANR1, the K_m for NADPH was higher and the K_{cat}/K_m values (where K_{cat} is the unimolecular rate constant or turnover number) were similar (Table I). The optimal pH and temperature for CsANR1 and CsANR2 with cyanidin as substrate were 7.5 and 6.5, 37°C and 55°C, respectively (Table I).

Expression of Tea LAR and ANR in *M. truncatula* Hairy Roots

Previously, we have shown that transformation of *M. truncatula* hairy roots with the MYB transcription
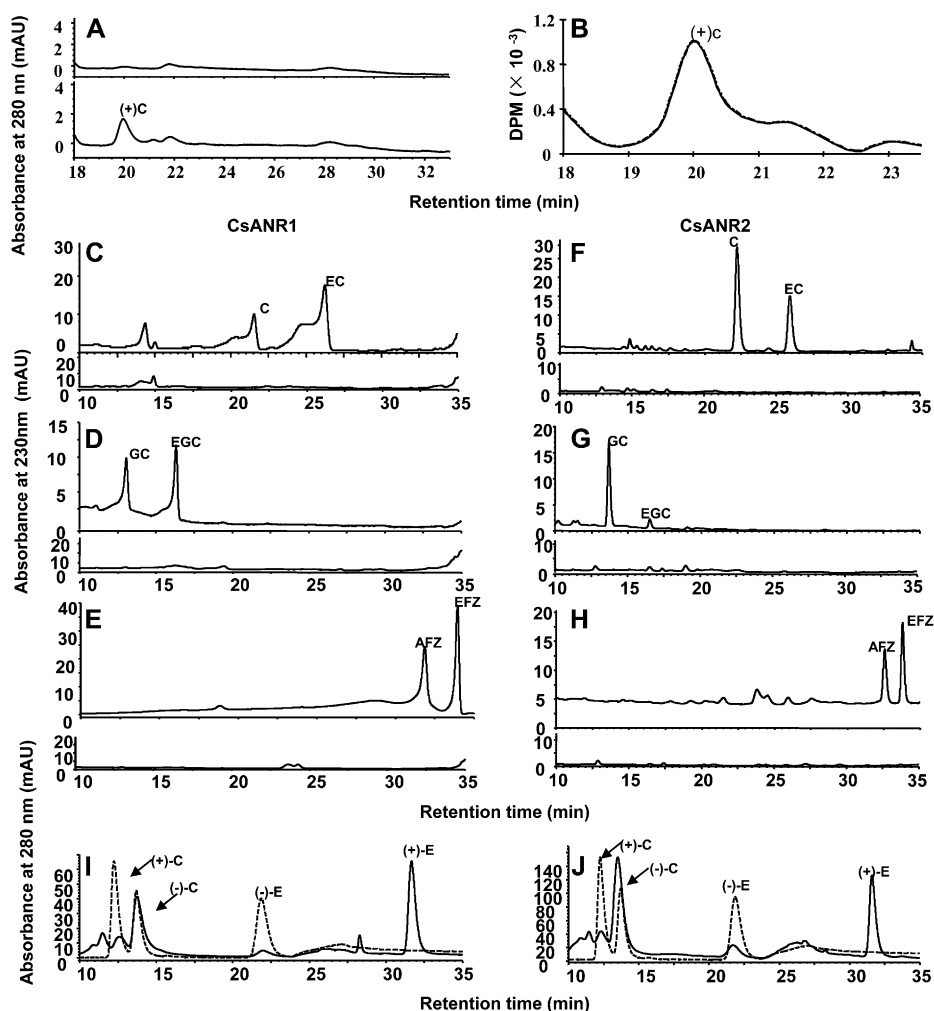
Figure 3. HPLC chromatograms from assays of recombinant tea LAR and ANR proteins. A, Chromatograms of products from assay of recombinant CsLAR protein (bottom panel) and boiled protein as a control (top panel). B, Profile of radioactivity in the peaks shown in A, bottom panel. The main product peaks in A and B elute at the same retention time as an authentic standard of (+)-catechin (for comparison of assays with *M. truncatula* LAR, see Pang et al., 2007). C to E, Chromatograms of products from assay of recombinant CsANR1 (top panels) and boiled protein as a control (bottom panels) incubated with the substrates cyanidin (C), delphinidin (D), and pelargonidin (E). F to H, Chromatograms of products from assay of recombinant CsANR2 protein (top panels) and boiled protein as a control (bottom panels) incubated with the substrates cyanidin (F), delphinidin (G), and pelargonidin (H). I and J, Chromatograms of products from assay of recombinant CsANR1 (I) and CsANR2 (J) on a chiral column. Solid lines represent enzyme products, and dashed lines represent authentic standards. An authentic standard of (+)-E was not available; this peak is surmised based on its retention time, UV spectrum, and the fact that the other three isomers had been identified. AFZ, Afzelechin; C, catechin; EC, epicatechin; EFZ, epiafzelechin (Fig. 1).

factor TT2 from Arabidopsis results in a greater than 10-fold increase in PA levels, accompanied by strong induction of ANR (Pang et al., 2008). Therefore, we asked whether simply expressing tea LAR or ANR in *M. truncatula* hairy roots would lead to the accumulation of PA monomers or oligomeric/polymeric PAs. To facilitate the screening of transgenic events, we used a vector with GFP as selection marker for the *CsANR2* transgene, in parallel with phosphinothricin selection (Fig. 4A; Supplemental Fig. S5A). More than 100 independent hairy root lines were generated from *M. truncatula* A17 with the four constructs, and four highly expressing lines (based on reverse transcription [RT]-PCR analysis of the transgene transcripts) were selected for further analysis (Supplemental Fig. S5A). Anthocyanin levels were reduced in all the *CsANR1*- and *CsANR2*-expressing lines, whereas three of the four *CsLAR*-expressing lines had clearly increased levels (Fig. 4B). Based on an assay with dimethlyaminocinnamaldehyde (DMACA) reagent, total soluble PA-like material was increased in all *CsANR2*-expressing lines, but no consistent changes were observed in *CsANR1* or *CsLAR*-expressing lines (Fig. 4C). Furthermore, total insoluble PA content was not significantly affected in the transgenic lines compared with the wild-type control (Fig. 4D).

The DMACA method detects flavan-3-ol monomers as well as oligomers and polymers. To more specifically quantify PA oligomers/polymers, soluble PA extracts were assayed by the butanol/HCl hydrolysis method (Pang et al., 2007). The results (Supplemental Fig. S6) confirmed increased levels of PAs in individual transformants with all four transgene constructs, although the overall levels were very low. In contrast to the results obtained by DMACA assay, the highest oligomeric PA levels were in *CsLAR* and *CsANR1* transformants. This suggests that these lines possess a higher ratio of oligomers to monomers than the *CsANR2*-expressing lines.

Analysis by normal-phase HPLC revealed that the *CsLAR*-expressing line 5 produced a range of PA-like compounds of short chain length, with a small peak eluting at the same retention time as the catechin/epicatechin standard (14–15 min) and a larger, unidentified peak eluting earlier (10 min; Fig. 5A). In contrast, extracts from hairy roots of both *CsANR1* (line 82) and *CsANR2* (line 9) as well as *CsANR2*(GFP) (line 3) expressing lines exhibited a strong peak by normal-phase HPLC at the same retention time as the

epicatechin standard (Fig. 5, B–D). No PA-like compounds were seen on normal-phase HPLC analysis of extracts from the vector control line (Fig. 5E).

Expression of Tea *LAR* and *ANR* in *PAP1* Tobacco

To determine whether tea *ANR* or *LAR* genes have better potential for metabolic engineering of PAs in a system that can potentially supply excess upstream precursors, they were next transformed into purple tobacco plants that accumulate massive levels of anthocyanins as a result of expression of the Arabidopsis *Production of Anthocyanin Pigment1* (*PAP1*) gene (Borevitz et al., 2000; Xie et al., 2006). The *PAP1*-expressing tobacco can theoretically provide leucoanthocyanidin and anthocyanidin as endogenous substrates for *LAR* and *ANR*, respectively. Thirty phosphinothricin-resistant lines expressing each of the three target genes in *PAP1* tobacco were detected by PCR and RT-PCR screening (Supplemental Fig. S5B), and, for each tea gene, the three lines with highest transcript level were selected for further analysis, together with *PAP1*-expressing and wild-type tobacco lines and a previously generated transgenic line coexpressing PAP1 and *M. truncatula* ANR (Xie et al., 2006) as a positive control. Anthocyanin levels decreased significantly in all *CsLAR*- and *CsANR*-expressing lines compared with PAP1 tobacco, and the same was observed for the PAP1 × MtANR line (Fig. 6, A and B). Soluble PA-like materials, based on colorimetry after reaction with DMACA reagent, were detected in all the transgenic lines, with highest values in several of the ANR-expressing lines (Fig. 6C). It should be pointed out that anthocyanins are weakly reactive with DMACA reagent and that the bulk of the signal in lines expressing PAP1 may be due to nonspecific detection of anthocyanins. Furthermore, analysis of the soluble PA extracts by butanol/HCl hydrolysis indicated that levels of oligomeric/polymeric PAs were low and that none of the transgenic lines accumulated significantly more oligomeric/polymeric PAs than the original PAP1-overexpressing line (Supplemental Fig. S7). Insoluble PA levels were also extremely low in all plants analyzed; therefore, the apparent increase in insoluble PA levels in plants expressing *CsANR2* may not be of significance (Fig. 6D).

The expression of *CsLAR* in *PAP1* tobacco produced a range of DMACA-reactive compounds as detected

Table I. *Kinetic properties of CsANR1 and CsANR2*

| Substrate	CsANR1				CsANR2			
	V_{max}	K_m	K_{cat}	K_{cat}/K_m	V_{max}	K_m	K_{cat}	K_{cat}/K_m
	nmol min^{-1} mg^{-1}	*μM*	*s^{-1}*	*M^{-1} s^{-1}*	*nmol min^{-1} mg^{-1}*	*μM*	*s^{-1}*	*M^{-1} s^{-1}*
Cyanidin	0.9 ± 0.4	26.7 ± 9.5	2.8×10^{-3}	104.5	0.6	13.1 ± 2.0	1.2×10^{-3}	92.4
Delphinidin	0.4 ± 0.1	24.4 ± 2.1	1.2×10^{-3}	196.8	1.3	14.2 ± 1.2	2.6×10^{-3}	186.6
Pelargonidin		20.1 ± 1.4				13.0 ± 0.5		
NADPH (with cyanidin)		97.2 ± 8.2				219.9 ± 56.3		
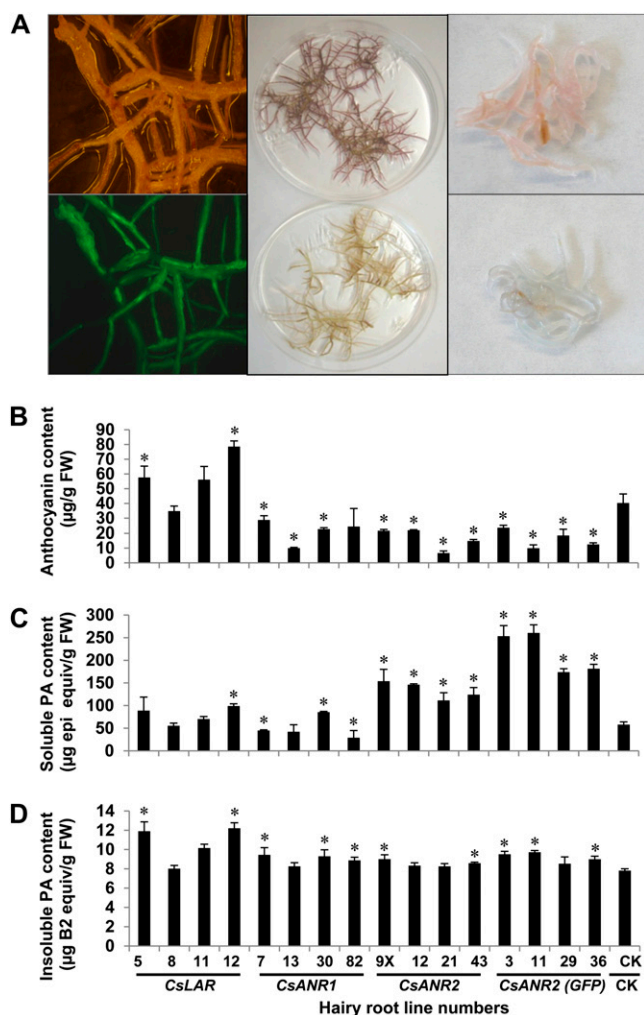
Figure 4. Ectopic expression of *CsLAR* and *CsANRs* in *M. truncatula* hairy roots. A, Anthocyanin pigmentation phenotype of *CsANR2*-expressing *M. truncatula* hairy roots under light microscopy for regular observation (top left), and GFP signal observed by fluorescence microscopy for screening of positive events (bottom left). The remaining photographs were taken under white light for the vector control line (top middle), roots expressing *CsANR2* (bottom middle), vector control roots stained with DMACA (top right), and a *CsANR2*-expressing line stained with DMACA (bottom right). B to D, Levels of anthocyanins (B), soluble PAs (C), and insoluble PAs (D) in transgenic *M. truncatula* hairy roots and a vector control (CK). Values show means and SD of triplicate analytical replicates from independent transgenic lines and controls that were pooled from several clonally propagated replicas of the same transgenic line. Asterisks indicate values that are significantly different from that of the vector control by Student's *t* test (*P* < 0.05). FW, Fresh weight.

extracts from PAP1 lines not expressing ANR or LAR (Fig. 7F) and likely represent cross-reactive anthocyanin derivatives that do not accumulate in wild-type control tobacco (Fig. 7F). The levels of these compounds were reduced in *CsLAR*, *CsANR*, and PAP1 × MtANR lines when compared with the PAP1 line, suggesting that pathway flux was shifted away from anthocyanin toward PAs.

Ultra-performance liquid chromatography (UPLC)-MS analysis was performed to confirm the presence of putative flavan-3-ol monomers and quantify other flavonoids produced in the transgenic tobacco lines. On this basis, the expression of *PAP1* led to the
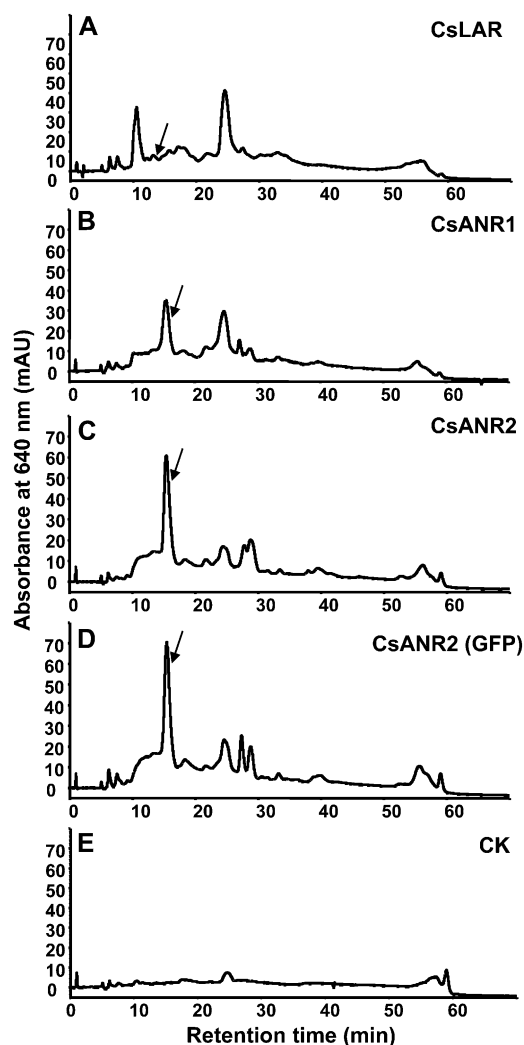
Figure 5. PA profiles of transgenic *M. truncatula* hairy roots expressing tea *ANR* and *LAR*. Normal-phase HPLC chromatograms show DMACA-reactive PA peaks (by postcolumn derivatization) in extracts from hairy roots expressing *CsLAR1* (line 5; A), *CsANR1* (line 82; B), *CsANR2* (line 9; C), *CsANR2*(GFP) (line 3; D), and wild-type control (CK; E) detected at 640 nm. Arrowheads indicate peaks that coelute with authentic standards of catechin and epicatechin. mAU, Milli-absorbance units.

by normal-phase HPLC (Fig. 7A), with a distinct broad peak at the retention times of the catechin and epicatechin standards (14–15 min). Both *CsANR1* (Fig. 7B) and *CsANR2* (Fig. 7C) expressing lines, together with PAP1 × MtANR lines (Fig. 7D), produced DMACA-reactive compounds at retention times from 25 to 32 min; however, such compounds are also present in
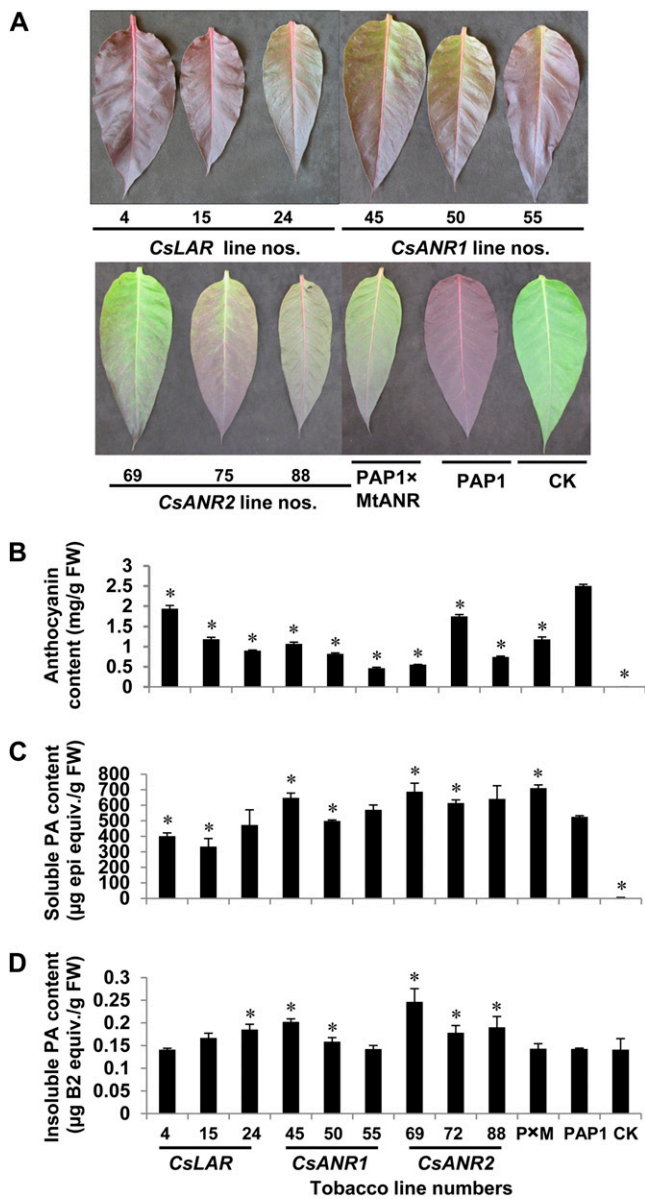
Figure 6. Expression of *CsLAR* and *CsANRs* in *PAP1*-expressing tobacco. A, Anthocyanin pigmentation phenotype of transgenic tobacco expressing *CsLAR*, *CsANR1*, or *CsANR2* in the PAP1 background, together with PAP1, PAP1 × MtANR, and wild-type (CK) plants. B to D, Levels of anthocyanins (B), soluble PAs (C), and insoluble PAs (D) in transgenic tobacco expressing *CsLAR*, *CsANR1*, or *CsANR2* in the PAP1 background, along with PAP1 × MtANR, PAP1 tobacco, and wild-type control (CK). Values show means and SD of triplicate analytical replicates from independent transgenic lines and controls that were pooled from several clonally propagated plants of the same transgenic line. Asterisks indicate values that are significantly different from that of the PAP1 line by Student's *t* test (*P* < 0.05). FW, Fresh weight.

accumulation of several phenylpropanoid-derived compounds that were not detectable in wild-type tobacco (Table II; Supplemental Data Set S1), including 7,3′,4′,5′-tetrahydroxyflavone, luteolin-7-*O*-

glucoside malonate, esculin, kaempferol-7-neohesperidoside, cyanidin-3-*O*-glucoside, 3,7,3′,4′,5′-pentahydroxyflavone, epicatechin-3′-*O*-glucoside, 3,5-dihydroxybenzoic acid, and gentisic acid. Wild-type plants accumulated low but detectable levels of epicatechin and catechin, which were increased on expression of *PAP1*. Epicatechin levels increased by 1.6- to 2.0-fold in *ANR*-expressing lines (CsANR1-45, CsANR1-55, and CsANR2-69) when compared with *PAP1* tobacco. However, the *CsLAR15*-expressing line accumulated 13.7- and 7.2-fold more epicatechin and catechin, respectively, and 21.8-fold more epicatechin glucoside than the PAP1 control (Table II; Supplemental Data Set S1).
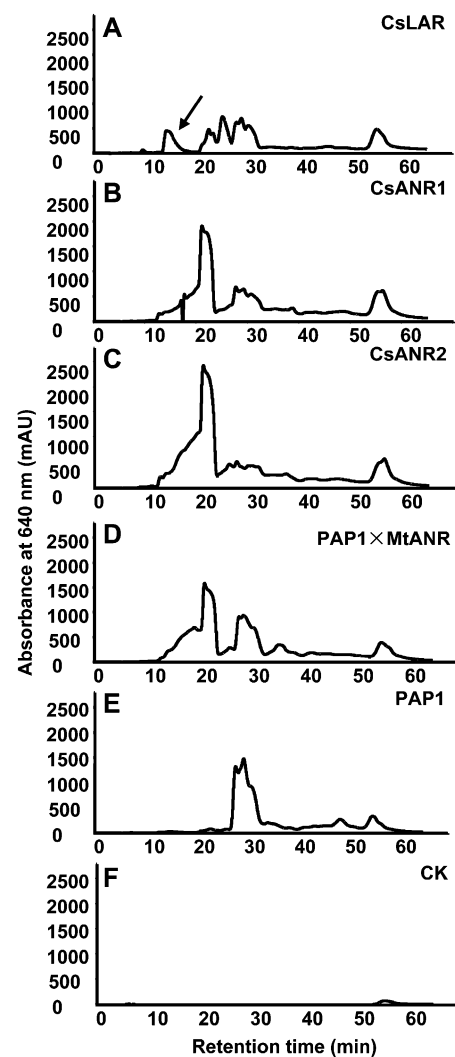
Figure 7. PA profiles from transgenic tobacco expressing the PAP1 transcription factor and tea ANR and LAR enzymes. Normal-phase HPLC chromatograms show DMACA-reactive PA peaks (by post-column derivatization) in extracts from *CsLAR* (line 15; A), *CsANR1* (line 45; B), *CsANR2* (line 69; C), PAP1 × MtANR (D), PAP1 (E), and wild-type tobacco (CK; F) detected at 640 nm. The arrow indicates the peak that coelutes with an authentic catechin standard.

Table II. *Levels of selected flavonoid compounds in leaf tissues of wild-type and transgenic (CsLAR- or CsANR-expressing) PAP1 tobacco lines determined by UPLC-MS analysis*

The data represent the peak area corresponding to each compound divided by that of the internal standard and multiplied by 1,000. The complete list of data from biological replicates with SD is provided in Supplemental Data Set S1. CK, Wild-type control.

Ion Used for Quantification	Compound Name	CK	PAP1	Ratio PAP1/ CK	PAP1 × MtANR	Ratio PAP1 × MtANR/ PAP1	CsANR2 Line 69	Ratio CsANR2 Line 69/ PAP1	CsANR1 Line 45	Ratio CsANR1 Line 45/ PAP1	CsLAR Line 15	Ratio CsLAR Line 15/ PAP1
161.0239	Umbelliferone	1,000	1,000	1.0	1,000	1.0	1,000	1.0	1,000	1.0	1,000	1.0
285.0407	7,3′,4′,5′-Tetrahydroxyflavone	0	26	∞	43	1.7	27	1.0	39	1.5	38	1.5
489.1040	Luteolin-7-O-glucoside malonate	0	18	∞	10	0.6	4	0.2	3	0.2	19	1.1
339.0712	Esculin	0	17	∞	14	0.8	13	0.8	5	0.3	10	0.6
593.1554	Kaempferol-7-neohesperidoside	0	15	∞	19	1.3	20	1.3	85	5.7	42	2.8
465.1035	Cyanidin-3-O-glucoside	0	12	∞	19	1.6	14	1.2	17	1.4	27	2.3
451.1235	Epicatechin-3-glucoside	0	5	∞	11	2.2	9	1.8	9	1.8	109	21.8
153.0168	3,5-Dihydroxybenzoic acid	5	186	37.2	234	1.3	235	1.3	110	0.6	211	1.1
301.0373	3,7,3′,4′,5′-Pentahydroxyflavone	5	110	22.0	49	0.4	46	0.4	33	0.3	43	0.4
153.0178	Gentisic acid	1	21	21.0	41	2.0	40	1.9	20	1.0	28	1.3
137.0226	4-Hydroxybenzoic acid	10	95	9.5	28	0.3	30	0.3	23	0.2	31	0.3
134.0358	1-Hydroxybenzotriazole	13	98	7.5	102	1.0	99	1.0	95	1.0	79	0.8
151.0025	3,4-Dihydroxybenzoic acid	1	7	7.0	49	7.0	43	6.1	6	0.9	13	1.9
301.0353	Dihydroquercetin	5	32	6.4	50	1.6	47	1.5	34	1.1	45	1.4
455.3562	α-Boswellic acid	7	44	6.3	39	0.9	22	0.5	20	0.5	28	0.6
299.0183	Quercetin	1	6	6.0	7	1.2	8	1.3	5	0.8	8	1.3
609.1505	Luteolin-3-7-di-O-glucoside	1,765	10,471	5.9	11,607	1.1	12,197	1.2	8,685	0.8	10,596	1.0
301.0335	Taxifolin	37	191	5.2	159	0.8	161	0.8	208	1.1	154	0.8
179.0367	Caffeic acid	569	2,915	5.1	3,478	1.2	3,414	1.2	2,582	0.9	3,175	1.1
351.0731	Cryptochlorogenic acid	365	1,837	5.0	1,778	1.0	1,659	0.9	1,738	0.9	1,680	0.9
285.0398	Luteolin	1	5	5.0	3	0.6	4	0.8	8	1.6	2	0.4
447.0954	Luteolin-4′-O-glucoside	1	5	5.0	8	1.6	7	1.4	5	1.0	6	1.2
191.0344	Scopoletin	4	18	4.5	10	0.6	8	0.4	5	0.3	13	0.7
289.0726	Epicatechin	2	9	4.5	17	1.9	14	1.6	18	2.0	123	13.7
285.0396	Kaempferol	4	16	4.0	14	0.9	17	1.1	17	1.1	15	0.9
447.0950	Kaempferol-7-O-glucoside	1	4	4.0	6	1.5	7	1.8	5	1.3	4	1.0
479.0830	Delphinidin-3-O-glucoside	1	4	4.0	2	0.5	4	1.0	2	0.5	7	1.8
607.1317	Rutin	375	1,333	3.6	1,333	1.0	1,272	1.0	948	0.7	1,093	0.8
193.0482	Ferulic acid	2	7	3.5	8	1.1	5	0.7	4	0.6	9	1.3
593.1505	Kempferol-3-O-rutinoside	393	1,130	2.9	1,523	1.3	1,568	1.4	933	0.8	556	0.5
151.0347	2,6-Dihydroxyacetophenone	6	15	2.5	15	1.0	21	1.4	12	0.8	14	0.9
351.0742	Neochlorogenic acid	89	220	2.5	208	0.9	195	0.9	150	0.7	143	0.7
285.0418	3,6,2′,4′-Tetrahydroxyflavone	5	12	2.4	12	1.0	13	1.1	14	1.2	17	1.4
289.0712	Catechin	10	14	1.4	14	1.0	15	1.1	16	1.1	101	7.2

DISCUSSION

PAs are important bioactive compounds for both human health and agriculture. Tea accumulates unusually high levels of PAs and related compounds in the leaves, which makes it an excellent subject in which to investigate the PA biosynthetic pathway. We here determined the biochemical properties of the two entry point enzymes into the PA pathway from tea and examined how these may be used for the engineering of flavanols and PAs in other species.

LAR and PA Engineering in Transgenic Plants

The LAR enzymes characterized to date from *D. uncinatum* (Tanner et al., 2003), grape (Bogs et al., 2005; Maugé et al., 2010), *M. truncatula* (Pang et al., 2007), and *L. corniculatus* (Paolocci et al., 2007) produce trans-flavan-3-ols from the corresponding leucoanthocyanidin in vitro. *CsLAR*, which shares more than 60% amino acid identity with the above LARs, likewise produces (+)-catechin from leucocyanidin in vitro. Although the LAR from *D. uncinatum* was confirmed to be active in vitro, transgenic tobacco and white clover (*Trifolium repens*) plants overexpressing this LAR did not produce detectable levels of (+)-catechin in vivo (Tanner et al., 2003), even though LAR activity could be detected in extracts from the leaves of these transgenic plants. Likewise, the expression of other LARs in transgenic plants also failed to result in the accumulation of catechin (Bogs et al., 2005; Pang et al., 2007; Paolocci et al., 2007).

M. truncatula hairy roots accumulate significant levels of anthocyanins (Pang et al., 2008), suggesting that the upstream leucocyanidin may be present as a substrate for LAR in these tissues. DMACA-reactive PA-like oligomers were found in transgenic hairy roots expressing *CsLAR*. Likewise, when *CsLAR* was overexpressed in the *PAP1*-expressing line of tobacco that accumulates massive amounts of anthocyanins (Borevitz et al., 2000), catechin monomer and DMACA-reactive PA oligomers accumulated. Surprisingly, however, a comparable amount of epicatechin and epicatechin glucoside also accumulated in PAP1 tobacco expressing *CsLAR*. The simplest hypothesis to explain this result is that the epicatechin arises from the epimerization of catechin and is then glucosylated. However, the specific stereochemistry of the compounds produced in transgenic plants remains to be determined, and it cannot be ruled out that *CsLAR* may use some other endogenous substrate to produce epicatechin monomers.

Recently, a LAR homolog, *PtrLAR3*, encoding an enzyme with yet unidentified in vitro activity, was isolated from *Populus trichocarpa* (Yuan et al., 2012); its overexpression in Chinese white poplar (*Populus tomentosa*) led to a significant increase of PA levels, although no flavan-3-ol monomer product was reported. Interestingly, the crude leaf extracts from *PtrLAR3*-expressing poplar were able to inhibit significantly the hyphal growth of the fungal pathogen *Marsonnina brunnea* f. sp. *multigermtubi* (Yuan et al., 2012), presumably as a result of the production of catechin or catechin-derived PAs. *CsLAR* is among the top 20 most abundantly expressed unigenes from our blister blight-resistant tea cDNA library, consistent with the observation that catechins may play important roles in blister blight resistance in tea. *CsLAR*, therefore, may be of value for engineering disease resistance in plant genetic backgrounds that provide the necessary leucocyanidin substrate. However, it is important to note that epicatechin, rather than catechin, derivatives are the major flavan-3-ols found in tea leaves, although LAR transcripts are expressed at a significantly higher level than ANR transcripts. Although relative transcript levels do not necessarily equate to in vivo enzyme activity, it is interesting to consider the possibility that LAR may function in the biosynthesis of both catechin and epicatechin in tea, based on the increased production of epicatechins in tobacco expressing tea LAR. The increase in anthocyanin levels in *M. truncatula* hairy roots expressing tea LAR is also interesting; the reason for this apparently counterintuitive observation is not currently obvious.

Tea ANRs and PA Engineering in Transgenic Plants

The two purified ANRs from tea produce (+)-E [along with (−)-catechin], in contrast to the more commonly observed (−)-E, which is the major product of ANR proteins from Arabidopsis and *M. truncatula* (Xie et al., 2003, 2004). Likewise, the ANR protein from grape also produced (+)-E (Gargouri et al., 2009, 2010), and the two CsANRs exhibit higher amino acid sequence identity to VvANR (79% and 83%) than to MtANR or AtANR (64%–73%). In contrast to the widely distributed and well-documented (+)-catechin and (−)-E isomers, there are only a few reports of the presence of (−)-catechin and (+)-E in plants, such as (−)-catechin from *Chamaebatia foliolosa* and *Centaurea maculosa* (Nahrstedt et al., 1987; Bais et al., 2003; Perry et al., 2005) and (+)-E from *Palmae* spp. and *Paullinia cupana* (Delle Monache et al., 1972; Yamaguti-Sasaki et al., 2007). (+)-Catechin, (−)-E, and their derivatives such as (−)-ECG are the major flavan-3-ols reported from tea (Tijburg et al., 1997; Henning et al., 2003; Hodgson, 2008). To the best of our knowledge, neither (−)-catechin nor (+)-E have been described from the tea plant itself, as is the case in grape, where the in vitro products of ANR appear to be the enantiomers of the most abundant in vivo metabolites (Gargouri et al., 2009). Except for the products of VvANR and the CsANRs reported here, the identification of flavan-3-ols in previous studies was achieved by comparison of UV spectra and HPLC retention times with authentic standards by reverse-phase HPLC, which can separate cis- and trans-isomers, such as epicatechin and catechin, but cannot distinguish enantiomers. Therefore, the lack of stereospecific assays could be one of the main reasons why the nature of the ANR products appears puzzling (Gargouri et al., 2009). In this respect, it will also be important to obtain a more detailed picture of the exact flavan-3-ol stereoisomers generated in transgenic plants expressing LARs and ANRs from other sources, including the tea enzymes reported here. Another ANR from a different cultivar of tea with only two amino acid substitutions (15 nucleotide substitutions) from *CsANR*2 (and therefore possibly an allelic variant) has been isolated and its in vitro product identified as (−)-E by reverse-phase HPLC (Singh et al., 2009; Zhang et al., 2012); this needs reevaluation in view of the criteria above.

The epimerase activity inherently associated with ANR was first observed for VvANR, and the epimerization was found to be at the C3 position (Gargouri et al., 2010). In addition to (+)-E, the two CsANRs also produced (−)-catechin, but in different proportions, indicating that the two CsANR proteins may have different degrees of epimerase activity. We previously reported that production of the minor product (−)-catechin by AtANR and MtANR was a result of the nonenzymatic epimerization of the major product (−)-E (Xie et al., 2003, 2004); this is distinct from the demonstrated epimerase activity of VvANR and CsANRs. Further evidence that this epimerization is an intrinsic activity of the tea enzymes is provided by the fact that the proportions of the two epimers produced by CsANR1 and CsANR2 are quite different.

When the two CsANRs were introduced into hairy roots of *M. truncatula* or PAP1 tobacco, DMACA-reactive compounds could be detected by normal-phase HPLC in all events, with a corresponding decrease of anthocyanins, suggesting diversion of anthocyanins to PAs or their precursors. However, except for putative flavan-3-ol-like compounds in the hairy root cultures, no definitively identified PAs were detected, as was also the case following the overexpression of *MtANR* and *VvANR*; although the levels of anthocyanins in petals were reduced at the same time as an increase of DMACA-reactive PA-like compounds, no solid evidence supported the appearance of a measurable pool of flavan-3-ol monomer (Xie et al., 2003, 2004; Bogs et al., 2005). This was again the case when *MtANR* and *CsANRs* were overexpressed in *PAP1* tobacco lines that accumulated massive levels of anthocyanins and clearly detectable levels of DMACA-reactive PA-like material (Xie et al., 2006). This apparent paradox may be explained if the monomers produced are used immediately (through coupled or channeled reactions) for the production of PA-like compounds using endogenous modification and polymerization enzymes or if ANR possesses additional functions for producing novel DMACA-reactive compounds in vivo.

Application of LAR and ANR for PA Biotechnology

Introduction of PAs into the foliage of forage crops has long been a goal of forage improvement programs for preventing lethal pasture bloat. So far, it has not been fully successful, although intensive efforts have been made via the introduction of positive regulators of the PA pathway into alfalfa (*Medicago sativa*) or white clover; these genes include the transcription factors *Lc* (Ray et al., 2003), *LAP1* (Peel et al., 2009), *MtWD40-1* (Pang et al., 2009), and *MtPAR* (Verdier et al., 2012). Although increased levels of anthocyanins and/or DMACA-reactive PA-like compounds could be detected by HPLC/UPLC, the level of PAs was far lower than desired, and the endogenous *ANR* genes were not greatly up-regulated (Pang et al., 2009; Peel

et al., 2009; Verdier et al., 2012). Although the MYB transcription factor TT2 can activate PA accumulation in *M. truncatula* hairy roots with very strong induction of ANR (Pang et al., 2008), expressing the same gene in alfalfa plants does not up-regulate ANR or lead to PA production in foliage (Peel et al., 2009). This indicates the importance of ANR for PA engineering. However, our observation here that provision of sufficient substrate and high expression of LAR and ANR are still insufficient to support high levels of PA accumulation in transgenic plants highlights our lack of understanding of the interface between PA and anthocyanin metabolism and the downstream mechanisms of PA polymerization. The recently discovered epimerase activity of ANR (and possibly other enzymes?) may be an important component of this puzzle and clearly requires further study in the context of the stereochemistry of naturally occurring flavanol-3-ols.

MATERIALS AND METHODS

Plant Materials and Chemicals

The tea plants (*Camellia sinensis*) used in this study were grown and plucked at the Tea Research Institute of Sri Lanka. Leaf discs were harvested from healthy leaves and blisters infected with *E. vexans* and were immediately frozen in liquid nitrogen, lyophilized, and stored for later use.

Cyanidin, delphinidin, and pelargonidin chlorides were purchased from Indofine. Ten millimolar methanol stock solutions were stored at −20°C. NADPH was from Sigma, and solutions were freshly prepared in cold reaction buffer when used for enzymatic assays. (+)-Catechin, (−)-catechin, and (−)-E were purchased from Sigma, and methanolic stock solutions were stored at −20°C.

Construction, Sequencing, and Statistical Analyses of a Leaf cDNA Library from Tea

Total RNAs were isolated separately from healthy leaves, leaf discs showing the translucent symptoms of blister blight disease, and leaf discs from healthy areas of infected leaves, all from the blister blight-resistant tea 'TRI2043', using the cetyl-trimethyl-ammonium bromide method (Jaakola et al., 2001). Equal amounts of total RNA from the three leaf samples were mixed together, and poly(A$^+$) RNA was isolated using an Oligotex mRNA Mini Kit (Qiagen). The Creator Smart cDNA Library Construction Kit (Clontech) and 3' and 5' RACE Kit (Clontech) were used for the cDNA and RACE library construction according to the manufacturer's instructions. Colonies were randomly picked for overnight culture in Terrific Broth liquid medium supplemented with 30 mg L^{-1} chloramphenicol. Biomek 2000 robots were used to isolate sequencing templates, and sequencing was performed on an ABI 3730 sequencer. An automated in-house NEST-PIPE (Noble internal EST-PIPline) system was used for the analysis of tea EST sequences. Features of the NEST-PIPE include sequencing error detection, sequence cleaning, vector removal, and sequence assembly. The unigenes were assembled using PLAN, a Web platform for automating high-throughput BLAST searches and for managing and mining results (bioinfo.noble.org/plan/). The unigene sequences were then annotated through BLAST search against GenBank's nonredundant sequences and further annotated using GO (http://www.geneontology.org/) based on sequence alignment. All EST sequences are deposited in NCBI GenBank under the library accession no. LIBEST_027432.

Sequence Alignment and Phylogeny Analysis

A multiple alignment of the deduced amino acid sequences of *CsLAR*, *CsANR1*, *CsANR2*, and other LAR and ANR proteins was constructed using ClustalW (Thompson et al., 1994). The resulting alignment was further used to generate an unrooted consensus tree by using MrBayes online software at

http://www.phylogeny.fr/version2_cgi/index.cgi (Huelsenbeck and Ronquist, 2001; Dereeper et al., 2008).

In Vitro Expression of *LAR* and *ANR* Genes in *Escherichia coli*

ESTs encoding CsLAR and two CsANRs were obtained from the cv TRI2043 leaf cDNA library, and their UTRs were obtained from the RACE library according to the manufacturer's instructions. The ORFs were amplified by *pfu* PCR with cDNA from total RNA using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. The primers used for *CsLAR* were *CsLAR*-BF (5′-GGATCCATGACTGTGTTGGAATCTGTGTC-3′, forward) and *CsLAR*-SR (5′-GAGCTCTCAAGCACACATTGTGAT-GGGAAG-3′, reverse) with introduced *Bam*HI and *Sac*I restriction enzymes sites (underlined) at the 5′ end of the forward and reverse primers, respectively. The primers for *CsANR1* were *CsLAR1*-BF (5′-GGATCCATGG-CAATGGCAATGGCAACAACAAC-3′, forward) and *CsANR1*-SR (5′-GAGCTCTCAGTTCTGCAAAAGCCCCTTAG-3′, reverse) with the same introduced restriction enzymes sites as above. The primers for *CsANR2* were *CsANR2*-SF (5′-GAGCTCATGGAAGCCCAACCGACAGCTC-3′, forward) and *CsANR2*-SR (5′-GTCGACTCAATTCTTCAAAATCCCCTTAG-3′, reverse) with introduced *Sac*I and *Sal*I sites at the 5′ end of the forward and reverse primers, respectively. PCR was carried out at 94°C for 5 min, 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min, followed by a 7-min extension at 72°C. The PCR products were gel purified using a Qiagen Gel Extraction Kit, ligated into pGEM-T Easy vector (Promega), and transformed into *E. coli* strain DH5α competent cells for sequencing. The construct containing the correct ORF of each gene was excised by digestion with the corresponding restriction enzymes and ligated into the *E. coli* expression vector pQE-30 (Qiagen) digested with the same enzymes to generate protein expression constructs pQE30-CsALAR and pQE30-CsANR1/2, in which the ORFs have a 6× His tag at their N termini.

pQE30-*CsLAR*, pQE30-*CsANR1*, pQE30-*CsANR2*, and empty pQE30 vector (control vector) were transformed into M15 competent cells (Qiagen) for protein expression. For each, a single colony harboring the construct was inoculated into 10 mL of Luria-Bertani broth containing 100 mg L^{-1} ampicillin and kanamycin, and cells were grown to an optical density at 600 nm of 0.6 at 37°C with shaking at 250 rpm. The cells were then harvested, and isopropyl-1-thio-β-D-galactopyranoside was added to a final concentration of 1 mM. The cells were harvested 3 h later by centrifugation at 2,900g at 4°C for 15 min. The cell pellets were then lysed, and the recombinant proteins were purified using the MagnetHis Protein Purification System (Promega). Protein concentrations were measured with the Bio-Rad Protein Assay system using the Bradford method (Bradford, 1976). For large-scale protein purification, the QIAexpressionist protocol for protein expression and purification in *E. coli* was used (Qiagen).

Enzyme Assays and HPLC Analysis of Products

Assay of CsLAR was carried out for 60 min in a total volume of 100 μL containing 20 mM potassium phosphate, pH 7.0, 5 mM [^{3}H]leucocyanidin, 2.5 mM NADPH, and 20 μg of CsLAR, as described previously (Pang et al., 2007). The enzyme reactions were initiated by adding enzyme and stopped by adding 100 μL of methanol, one-half of which was then injected for HPLC analysis as described previously (Pang et al., 2007). HPLC fractions with retention times ranging from 20 to 26 min were collected at 0.5 min per vial for scintillation counting. CsANR assays were carried out for 2 h in a total volume of 100 μL containing 20 mM potassium phosphate, pH 6.5, 100 μM cyanidin (delphinidin and pelargonidin) chloride, 2.5 mM NADPH, and 30 μg of CsANR1 or CsANR2 recombinant protein. Reactions were stopped with methanol, and 100 μL was injected for HPLC analysis.

Products from ANR enzyme assay were analyzed by reverse-phase HPLC on an Agilent HP1100 HPLC device with the following gradient: solvent A (1% phosphoric acid) and B (acetonitrile) at 1 mL min^{-1} flow rate: 0 to 5 min, 6% B; 5 to 10 min, 6% to 10% B; 10 to 20 min, 10% to 11% B; 20 to 25 min, 11% to 12% B; 25 to 45 min, 12% to 37% B; 45 to 48 min, 37% to 100% B; 48 to 58 min, 100% B; 58 to 60 min, 100% to 6% B. Data were collected at 280 and 230 nm (the latter for enzyme kinetic calculation only). Identifications were based on chromatographic behavior and UV spectra.

Enzyme kinetic studies for ANR proteins were as described by Xie et al. (2004). Kinetic constants were calculated by Lineweaver-Burk plot using Hyper32 software (http://hyper32.software.informer.com/).

The products of ANR reactions were further distinguished on an analytical chiral column (catalog no. 80325; Chiral Technologies) with a guard cartridge (catalog no. 80311; Chiral Technologies) on the same HPLC device with the following gradient: solvent A (hexanes with 0.5% acetic acid) and B (ethanol with 0.5% acetic acid) at 1 mL min^{-1} flow rate: 0 to 20 min, 20% B; 20 to 23 min, 20% to 50% B; 23 to 38 min, 50% B; 38 to 40 min, 50% to 20% B. UV absorption data were collected at 280 nm. Identifications were based on comparison of chromatographic behavior and UV spectra with authentic standards.

Construction of Binary Vectors for Plant Transformation

The ORFs of the *CsLAR*, *CsANR1*, and *CsANR2* genes were amplified from their corresponding sequenced *E. coli* expression vector, using DNA polymerase with proofreading activity, with the following primer pairs: *CsLAR*CF (5′-CACCATGACTGTGTTGGAATCTGTGTCCG-3′) and *CsLAR*R (5′-TCAAGCACACATTGTGATGGGAAG-3′) for the *CsLAR* gene; *CsANR1*CF (5′-CACCATGGCAATGGCAATGGCAACAACAAC-3′) and *CsANR1*R (5′-TCAGTTCTGCAAAAGCCCCTTAG-3′) for the *CsANR1* gene; and *CsANR2*CF (5′-CACCATGGAAGCCCAACCGACAGCTC-3′) and *CsANR2*R (5′-TCAATTCTTCAAAATCCCCTTAG-3′) for the *CsANR2* gene. PCR products were purified and cloned into the Gateway Entry vector pENTR/D-TOPO (Invitrogen) and confirmed by sequencing. Entry vectors were then transferred into the Gateway plant transformation destination vector pB2GW7 (Karimi et al., 2002) using Gateway LR Clonase enzyme mix according to the manufacturer's instructions (Invitrogen). The ORFs of the resulting vectors, pB2GW7-*CsLAR*, pB2GW7-*CsANR1*, and pB2GW7-*CsANR2*, were confirmed by sequencing, followed by transformation into *Agrobacterium tumefaciens* strain AGL1 by electroporation. A single colony containing each target construct was confirmed by PCR and used for genetic transformation of tobacco (*Nicotiana tabacum*) overexpressing the *PAP1* gene (Borevitz et al., 2000) using the leaf disc transformation method (Horsch et al., 1988) with 10 mg L^{-1} phosphinothricin selection. pENTR-*CsANR2* was also used to transfer the *CsANR2* ORF in pB7WG2D with phosphinothricin and GFP selection; the resulting vector, pB7WG2D-*CsANR2*, together with pB2GW7-*CsLAR*, pB2GW7-*CsANR1*, and pB2GW7-*CsANR2*, were introduced into ARqual1 for *Medicago truncatula* hairy root transformation as reported previously (Pang et al., 2008).

Transcript Analysis by RT-PCR

Total RNA from leaves of *M. truncatula* hairy roots and PAP1 tobacco lines was isolated using the RNeasy Plant Mini Kit (Qiagen). cDNA was synthesized from 3 μg of total RNA in a total volume of 20 μL using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions; 4 μL was used in subsequent RT-PCRs. The primers used were *CsLAR*F (5′-ATGACTGTGTTGGAATCTGTGTCCG-3′) and *CsLAR*R for the *CsLAR* transgene sequence; *CsANR1*F (5′-ATGGCAATGGCAATGGCAACAACA-AC-3′) and *CsANR1*R for *CsANR1*; and *CsANR2*F (5′-ATGGAAGCCCAA-CCGACAGCTC-3′) and *CsANR2*R for *CsANR2*. Optimized PCR conditions were 94°C for 5 min, 25 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min, followed by a final extension of 72°C for 7 min. Actin was used as an internal standard with the primers actin-F and actin-R for tobacco (Pang et al., 2007). PCR products were analyzed by electrophoresis of 10-μL aliquots of reactions on 1% agarose gels in Tris-acetic acid-EDTA buffer and visualized with SYBR DNA Stain Dye (Molecular Probes).

Extraction, Quantification, and Analysis of Anthocyanins, PAs, and Other Flavonoids

Anthocyanins were extracted with methanol and 0.1% HCl, followed by the addition of water and chloroform to remove chlorophyll, and then measured spectrophotometrically at 530 nm as described previously (Pang et al., 2007). Soluble PAs were extracted three times with 70% acetone and 0.5% acetic acid, three times with chloroform, and twice with hexanes; the aqueous phase was then freeze dried and resuspended in 70% acetone and 0.5% acetic acid, and PA levels were determined by reaction with 0.2% DMACA (Peel and Dixon, 2007) or by heating with butanol/HCl (95:5) and determining the A_{550} (Pang et al., 2007). The size distribution profile of soluble PAs was determined by normal-phase HPLC with postcolumn DMACA derivatization (Peel and Dixon, 2007). Insoluble PAs were measured by heating with butanol/HCl and determining the A_{550}, and the released anthocyanidins were further monitored by reverse-phase HPLC as described previously (Pang et al., 2007, 2008; Peel

and Dixon, 2007). For UPLC-MS analysis, 10 ± 0.6 mg of dried material was extracted with 500 μL of 80% methanol containing 18 mg mL^{-1} umbelliferone as an internal standard, and 5 μL of the clear supernatant was injected into a Waters Acquit UPLC system fitted with a hybrid quadrupole time-of-flight Premier mass spectrometer (Waters). Details of flavonoid identification and data analysis by this method have been given elsewhere (Pang et al., 2009).

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers LIBEST 027423 (all tea ESTs), GU992400 (CsANR2), GU992401 (CsLAR), and GU992402 (CsANR1).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Biosynthetic pathway to anthocyanins and PAs.

Supplemental Figure S2. Functional classification of the tea leaf library unigene data set.

Supplemental Figure S3. Alignment of the deduced amino acid sequences of *LAR* genes.

Supplemental Figure S4. Alignment of the deduced amino acid sequences of *ANR* genes.

Supplemental Figure S5. Expression of transgenes in genetically engineering plants as determined by RT-PCR.

Supplemental Figure S6. Soluble PA content of *M. truncatula* hairy roots expressing CsLAR and CsANRs as determined by butanol/HCl assay.

Supplemental Figure S7. Soluble PA content of tobacco expressing CsLAR and CsANRs as determined by butanol/HCl assay.

Supplemental Table S1. The top 20 most abundantly expressed unigenes in the tea 'TRI2043' leaf cDNA library.

Supplemental Table S2. Putative fungal resistance-related unigenes from the tea leaf cDNA library.

Supplemental Table S3. Phenylpropanoid biosynthetic pathway unigenes from the tea leaf cDNA library.

Supplemental Data Set S1. Levels of selected flavonoid compounds in leaf tissues of replicated wild-type and transgenic (CsLAR- or CsANR-expressing) PAP1 tobacco lines determined by UPLC-MS analysis.

LITERATURE CITED

Arulpragsam PV (1992) Disease control in Asia. *In* KC Willson, MN Clifford, eds, Tea: Cultivation to Consumption. Chapman & Hall, London, pp 353–374

Ashihara H, Crozier A (2001) Caffeine: a well known but little mentioned compound in plant science. Trends Plant Sci **6:** 407–413

Ashihara H, Sano H, Crozier A (2008) Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. Phytochemistry **69:** 841–856

Bais HP, Walker TS, Kennan AJ, Stermitz FR, Vivanco JM (2003) Structure-dependent phytotoxicity of catechins and other flavonoids: flavonoid conversions by cell-free protein extracts of *Centaurea maculosa* (spotted knapweed) roots. J Agric Food Chem **51:** 897–901

Balentine DA, Wiseman SA, Bouwens LC (1997) The chemistry of tea flavonoids. Crit Rev Food Sci Nutr **37:** 693–704

Bemis DL, Katz AE, Buttyan R (2006) Clinical trials of natural products as chemopreventive agents for prostate cancer. Expert Opin Investig Drugs **15:** 1191–1200

Bogs J, Downey MO, Harvey JS, Ashton AR, Tanner GJ, Robinson SP (2005) Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. Plant Physiol **139:** 652–663

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12:** 2383–2394

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:** 248–254

Degtyarenko KN, Archakov AI (1993) Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. FEBS Lett **332:** 1–8

Delle Monache F, Ferrari F, Poce-Tucci A, Marini-Bettolo GB (1972) Catechins with (+)-epi-configuration in nature. Phytochemistry **11:** 2333–2335

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res **36:** W465–W469

Gargouri M, Chaudière J, Manigand C, Maugé C, Bathany K, Schmitter JM, Gallois B (2010) The epimerase activity of anthocyanidin reductase from *Vitis vinifera* and its regiospecific hydride transfers. Biol Chem **391:** 219–227

Gargouri M, Manigand C, Maugé C, Granier T, Langlois d'Estaintot B, Cala O, Pianet I, Bathany K, Chaudière J, Gallois B (2009) Structure and epimerase activity of anthocyanidin reductase from *Vitis vinifera*. Acta Crystallogr D Biol Crystallogr **65:** 989–1000

Henning SM, Fajardo-Lira C, Lee HW, Youssefian AA, Go VL, Heber D (2003) Catechin content of 18 teas and a green tea extract supplement correlates with the antioxidant capacity. Nutr Cancer **45:** 226–235

Hodgson JM (2008) Tea flavonoids and cardiovascular disease. Asia Pac J Clin Nutr (Suppl 1) **17:** 288–290

Horsch RB, Fry J, Hoffmann N, Neidermeyer J, Rogers SG, Fraley RT (1988) Leaf disc transformation. *In* Plant Molecular Biology Manual. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 1–9

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:** 754–755

Jaakola L, Pirttilä AM, Halonen M, Hohtola A (2001) Isolation of high quality RNA from bilberry (*Vaccinium myrtillus* L.) fruit. Mol Biotechnol **19:** 201–203

Kao YH, Chang HH, Lee MJ, Chen CL (2006) Tea, obesity, and diabetes. Mol Nutr Food Res **50:** 188–210

Karimi M, Inzé D, Depicker A (2002) Gateway vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci **7:** 193–195

Kato M, Mizuno K, Fujimura T, Iwama M, Irie M, Crozier A, Ashihara H (1999) Purification and characterization of caffeine synthase from tea leaves. Plant Physiol **120:** 579–586

Lin Y, Juan I, Chen Y, Liang Y, Lin J (1996) Composition of polyphenols in fresh tea leaves and associations of their oxygen-radical-absorbing capacity with antiproliferative actions in fibroblast cells. J Agric Food Chem **44:** 1387–1394

Lin YS, Tsai YJ, Tsay JS, Lin JK (2003) Factors affecting the levels of tea polyphenols and caffeine in tea leaves. J Agric Food Chem **51:** 1864–1873

Luximon-Ramma A, Neergheen VS, Bahorun T, Crozier A, Zbarsky V, Datla KP, Dexter DT, Aruoma OI (2006) Assessment of the polyphenolic composition of the organic extracts of Mauritian black teas: a potential contributor to their antioxidant functions. Biofactors **27:** 79–91

Maugé C, Granier T, d'Estaintot BL, Gargouri M, Manigand C, Schmitter JM, Chaudière J, Gallois B (2010) Crystal structure and catalytic mechanism of leucoanthocyanidin reductase from *Vitis vinifera*. J Mol Biol **397:** 1079–1091

Nahrstedt A, Proksch P, Conn EE (1987) Dhurrin, (−)-catechin, flavonol glycosides and flavones from *Chamaebatia foliolosa*. Phytochemistry **26:** 1546

Nimal Punyasiri PA, Tanner GJ, Abeysinghe IS, Kumar V, Campbell PM, Pradeepa NH (2004) *Exobasidium vexans* infection of *Camellia sinensis* increased 2,3-cis isomerisation and gallate esterification of proanthocyanidins. Phytochemistry **65:** 2987–2994

Pang Y, Peel GJ, Sharma SB, Tang Y, Dixon RA (2008) A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. Proc Natl Acad Sci USA **105:** 14210–14215

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

Pang Y, Wenger JP, Saathoff K, Peel GJ, Wen J, Huhman D, Allen SN, Tang Y, Cheng X, Tadege M, 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, Arcioni S, Martens S, Damiani F (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 corniculatus.* Plant Physiol **143:** 504–516

Park JS, Kim JB, Hahn BS, Kim KH, Ha SH, Kim JB, Kim YH (2004) EST analysis of genes involved in secondary metabolism in *Camellia sinensis* (tea), using suppression subtractive hybridization. Plant Sci **166:** 953–961

Peel GJ, Dixon RA (2007) Detection and quantification of engineered proanthocyanidin in trangenic plants. Nat Prod Commun **2:** 1009–1014

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

Perry LG, Thelen GC, Ridenour WM, Weir TL, Callaway RM, Paschke MW, Vivanco JM (2005) Dual role for an allelochemical: (+/−)-catechin from *Centaurea maculosa* root exudates regulates conspecific seedling establishment. J Ecol **93:** 1126–1135

Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I (2005) *TRANSPARENT TESTA10* encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat. Plant Cell **17:** 2966–2980

Punyasiri PA, Abeysinghe IS, Kumar V, Treutter D, Duy D, Gosch C, Martens S, Forkmann G, Fischer TC (2004) Flavonoid biosynthesis in the tea plant *Camellia sinensis:* properties of enzymes of the prominent epicatechin and catechin pathways. Arch Biochem Biophys **431:** 22–30

Punyasiri PAN, Abeysinghe SB, Kumar V (2005) Preformed and induced chemical resistance of tea leaf against *Exobasidium vexans* infection. J Chem Ecol **31:** 1315–1324

Ravindranath MH, Saravanan TS, Monteclaro CC, Presser N, Ye X, Selvan SR, Brosman S (2006) Epicatechins purified from green tea (*Camellia sinensis*) differentially suppress growth of gender-dependent human cancer cell lines. Evid Based Complement Alternat Med **3:** 237–247

Ray H, Yu M, Auser P, Blahut-Beatty L, McKersie B, Bowley S, Westcott N, Coulman B, Lloyd A, Gruber MY (2003) Expression of anthocyanins and proanthocyanidins after transformation of alfalfa with maize Lc. Plant Physiol **132:** 1448–1463

Saito K, Kobayashi M, Gong Z, Tanaka Y, Yamazaki M (1999) Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red forma of *Perilla frutescens.* Plant J **17:** 181–189

Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Chen Q, Xia T, et al (2011) Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics **12:** 131

Singh K, Rani A, Paul A, Dutt S, Joshi R, Gulati A, Ahuja PS, Kumar S (2009) Differential display mediated cloning of anthocyanidin reductase gene from tea (*Camellia sinensis*) and its relationship with the concentration of epicatechins. Tree Physiol **29:** 837–846

Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR (2003) Proanthocyanidin biosynthesis in plants: purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. J Biol Chem **278:** 31647–31656

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:** 4673–4680

Tijburg LB, Mattern T, Folts JD, Weisgerber UM, Katan MB (1997) Tea flavonoids and cardiovascular disease: a review. Crit Rev Food Sci Nutr **37:** 771–785

Verdier J, Zhao J, Torres-Jerez I, Ge S, Liu C, He X, Mysore KS, Dixon RA, Udvardi MK (2012) MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula.* Proc Natl Acad Sci USA **109:** 1766–1771

Wolfram S, Wang Y, Thielecke F (2006) Anti-obesity effects of green tea: from bedside to bench. Mol Nutr Food Res **50:** 176–187

Xie DY, Sharma SB, Dixon RA (2004) Anthocyanidin reductases from *Medicago truncatula* and *Arabidopsis thaliana.* Arch Biochem Biophys **422:** 91–102

Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2003) Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. Science **299:** 396–399

Xie DY, Sharma SB, Wright E, Wang ZY, Dixon RA (2006) Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. Plant J **45:** 895–907

Yamaguti-Sasaki E, Ito LA, Canteli VC, Ushirobira TM, Ueda-Nakamura T, Dias Filho BP, Nakamura CV, de Mello JC (2007) Antioxidant capacity and in vitro prevention of dental plaque formation by extracts and condensed tannins of *Paullinia cupana.* Molecules **12:** 1950–1963

Yang TT, Koo MW (2000) Inhibitory effect of Chinese green tea on endothelial cell-induced LDL oxidation. Atherosclerosis **148:** 67–73

Yoneyama N, Morimoto H, Ye C-X, Ashihara H, Mizuno K, Kato M (2006) Substrate specificity of N-methyltransferase involved in purine alkaloids synthesis is dependent upon one amino acid residue of the enzyme. Mol Genet Genomics **275:** 125–135

Yuan L, Wang L, Han Z, Jiang Y, Zhao L, Liu H, Yang L, Luo K (2012) Molecular cloning and characterization of PtrLAR3, a gene encoding leucoanthocyanidin reductase from *Populus trichocarpa,* and its constitutive expression enhances fungal resistance in transgenic plants. J Exp Bot **63:** 2513–2524

Zhang X, Liu Y, Gao K, Zhao L, Liu L, Wang Y, Sun M, Gao L, Xia T (2012) Characterisation of anthocyanidin reductase from Shuchazao green tea. J Sci Food Agric **92:** 1533–1539

Zhao J, Dixon RA (2009) 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, Pang Y, Dixon RA (2010) The mysteries of proanthocyanidin transport and polymerization. Plant Physiol **153:** 437–443

Zhao LP, Gao QK, Chen L (2005) Sequencing of cDNA clones and analysis of the expressed sequence tag (ESTs) of tea plant [*Camellia sinensis* (L.) O. Kuntze] young shoots. Chin J Agric Biotechnol **2:** 137–141