JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Identification of Two Anthocyanidin Reductase Genes and Three Red-Brown Soybean Accessions with Reduced *Anthocyanidin Reductase 1* mRNA, Activity, and Seed Coat Proanthocyanidin Amounts

Nik Kovinich,[†,‡] Ammar Saleem,[§] John T. Arnason,*[,§] and Brian Miki[†]

[†]Bioproducts and Bioprocesses, Research Branch, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada K1A 0C6

[‡]Ottawa–Carleton Institute of Biology, Department of Biology, Carleton University, Ottawa, Ontario, Canada K1S 5B6

[§]Laboratory for the analysis of Natural and Environmental Toxins (LANSET), Department of Biology and Center for Research in Biopharmaceuticals and Biotechnology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

Supporting Information

ABSTRACT: Anthocyanidin reductase (ANR; EC 1.3.1.77) catalyzes a key step in the biosynthesis of proanthocyanidins (PAs; also known as condensed tannins), flavonoid metabolites responsible for the brown pigmentation of seeds. Here, two ANR genes (*ANR1* and *ANR2*) from the seed coat of brown soybean (*Glycine max* (L.) Merr.) have been isolated and their enzymatic function confirmed for the reduction of cyanidin to (−)-epicatechin in vitro. Biochemical and genetic comparisons of soybean lines differing in seed coat color revealed three red-brown lines to exhibit major reductions in the amounts of soluble PAs in the seed coat compared to brown soybean lines. Two spontaneous mutants with red-brown grain color had reduced *ANR1* gene expression in the seed coat, and an EMS-mutagenized red-brown mutant had nonsynonymous substitutions that resulted in slightly reduced *ANR1* activity in vitro. These results suggest that defects in the *ANR1* gene can be associated with red-brown soybean grain color. These results suggest that suppressing *ANR1* gene expression or activity may be a rational approach toward engineering seed coat color to enable the visual identification of genetically modified soybean grains.

KEYWORDS: *anthocyanidin reductase, proanthocyanidin, anthocyanin, red-brown seed coat color, soybean (Glycine max (L.) Merr.)*

INTRODUCTION

Flavonoids, including proanthocyanidin (PA, also known as condensed tannin) and anthocyanin pigments, are plant secondary metabolites that have a variety of ecophysiological functions including interorganism signaling, pathogen defense, and protection from abiotic stresses.[1−3] Furthermore, metabolic engineering of flavonoids has become an important target for plant biotechnology as they provide favorable agronomic traits to crops and health benefits to foods and may be used to color commercial genetically modified materials to facilitate their identification and monitoring.[4−6]

In the soybean (*Glycine max* (L.) Merr.), flavonoids including PA and anthocyanin pigments can be synthesized in the seed coat (Figure 1). Six seed coat colors have been identified to be associated with spontaneous mutations in genetic loci by classical genetics (*I*, *R*, *T*, *Wp*, *W1*, and *O*).[7] The presence or absence of flavonoid-based color in the seed coat has been associated with alleles of the *I* locus coding tandem genomic repeats of *CHS* genes.[7−9] Black (*iRT*) or brown (*irT*) soybean isolines that differ in alleles of the *R* locus differentially accumulate specific anthocyanins, PAs, isoflavones, and phenylpropanoids in the seed coat and differentially express 20 flavonoid/phenylpropanoid isogenes, suggesting that *R* encodes a regulatory gene.[10] Black (*iRT*) and imperfect black (*iRt*) seed coat color are associated with different alleles of the pleiotropic *T* locus coding flavonoid 3′-hydroxylase (*F3′H1*).[11−15] Imperfect black (*W1*) and buff (*w1*) seed coat colors in an *iRt* background are associated with alleles of the *W1* locus, coding flavonoid-3′,5′-hydroxylase (*F3′5′H*).[7,16] Pigmented soybeans carrying the *Wp* allele are black (*iRTWp*), whereas *wp* gives a lighter grayish (*iRTwp*) grain color.[17] Microarray analysis demonstrated that high levels of transcripts of the flavonone 3-hydroxylase gene (*F3H1*) were associated with purple (*Wp*) flower color and low levels with pink (*wp*) color.[17] Furthermore, *Wp* has black and *wp* light gray grain color in an *iRT* background. Brown (*O*) and red-brown (*o*) grain color are controlled by the *O* locus in an *irT* background.[7]

Anthocyanidin reductase (*ANR*) catalyzes the first committed biochemical step to PA biosynthesis in *Arabidopsis thaliana* and *Medicago truncatula*.[18] Recently, a partial soybean sequence with 86% sequence identity to the *M. truncatula ANR* gene was mapped to the genomic region coding the *O* locus.[19] However, the function of the putative *ANR* sequence was not investigated, nor was it determined whether there existed biochemical differences relating to PA biosynthesis in the brown (*O*) and red-brown (*o*) seed coats. In *Arabidopsis*, knockout mutations in the *ANR* (also known as *BANYULS*) gene result in the absence of PAs in the seed coat, the temporary accumulation of cyanic pigments during early seed development, and a

Received: August 31, 2011
Revised: November 16, 2011
Accepted: November 22, 2011
Published: November 22, 2011

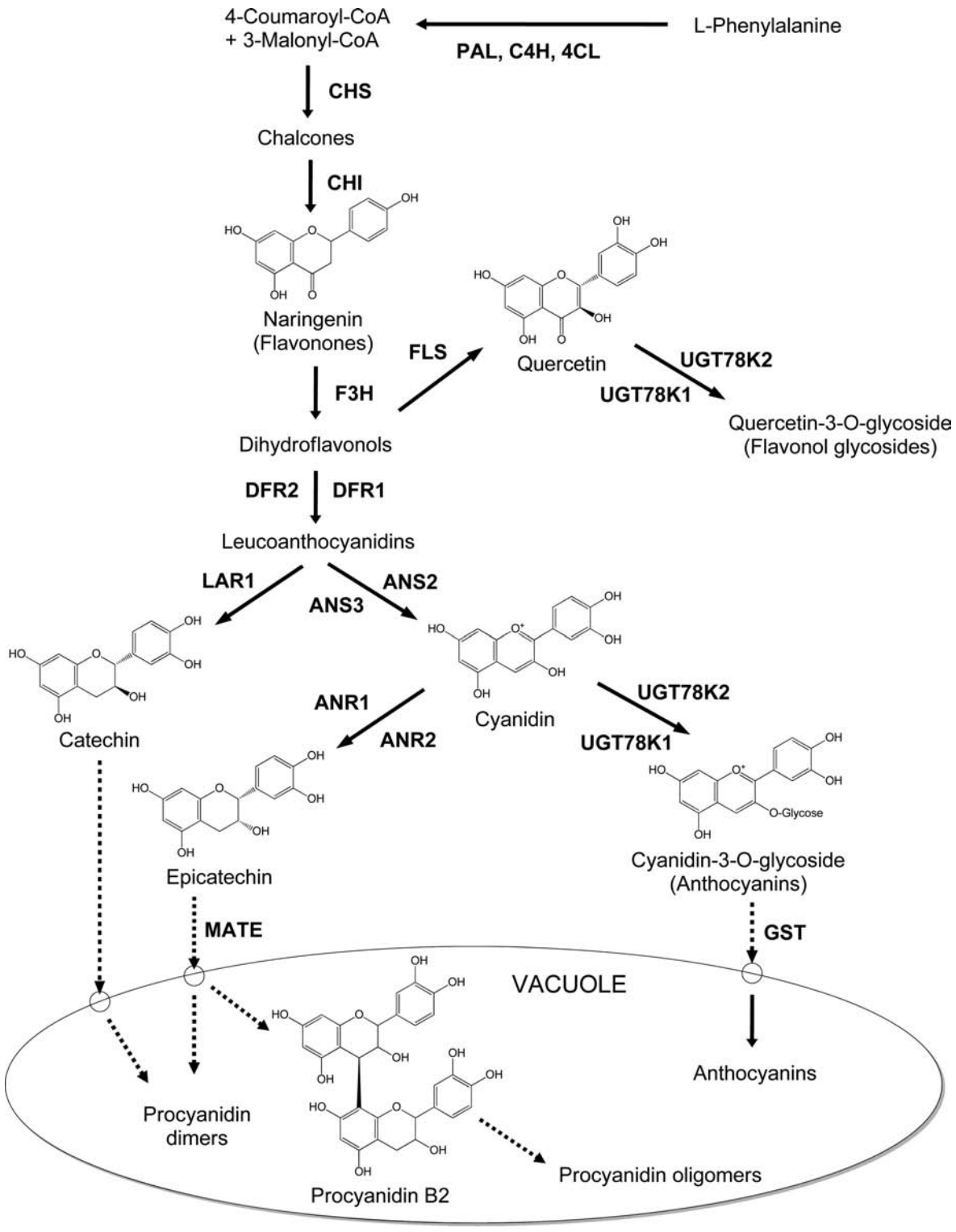

Figure 1. Flavonoid biosynthesis in the soybean seed coat.

transparent testa (*tt*) of the dried grain.[20] The similar accumulation of cyanic pigments in red-brown soybean may suggest defects in the *ANR* gene. Furthermore, association of red-brown soybean grain color with defects in an *ANR* gene may suggest gene suppression as an approach toward engineering cyanic color to mark genetically modified (GM) soybean grains.

Coloring soybeans with endogenous pigments such as anthocyanins has been proposed as a strategy to enable effective detection and monitoring of the adventitious presence of genetically modified (GM) soybeans in non-GM grain shipments.[6] Recent studies aimed at elucidating the molecular mechanisms that control soybean seed coat color have identified late-stage pigmentation genes and biochemical differences associated with grain color. A UDP-glucose:flavonoid-3-*O*-glucosyltransferase (UF3GT; EC 2.4.1.115) gene (*UGT78K1*) was isolated from black soybean seed coat cDNA using homology-based methods, and its function was confirmed in vitro and by complementation of the *Arabidopsis*

Table 1. Soybean Lines, Alleles, and Phenotypic Properties of Their Seeds

variety	abbreviated name	allele[a]	seed coat color	seed coat luster	seed size	relative timing of seed maturation	source/origin
Clark	ir	*irTO*	light brown	dull	medium	late	i-mutation in L11, found in 1965 at Urbana, IL
Soysota	r17	*irTO*	midbrown	moderate	small	mid	collected in 1910 in Italy
Linia N145	M100	unknown	dark brown	shiny	large	early	EMS-mutagenized Linia N145 (PI 522192 A), collected in 2005 at Ottawa, Canada
Agate	Ag	i^k*rTo*	red-brown saddle	shiny	large	early	collected at Hokkaido, Japan; donated in 1951
Kuro Daizu	KD	*irTo*	red-brown	shiny	large	early	collected in 1929 at Hokkaido, Japan
Linia N145	M11	unknown	red-brown	shiny	large	early	EMS-mutagenized Linia N145 (PI 522192 A), collected in 2005 at Ottawa, Canada

[a]All lines used in this study were homozygous for the *I*, *R*, *T*, and *O* alleles listed.

mutation *ugt78d2*.[21] *UGT78K1*, a second UF3GT gene (*UGT78K2*), and an anthocyanidin 3′-*O*-methyltransferase (AOMT) gene (*OMT5*) were found to be up-regulated with anthocyanin biosynthesis in the seed coat of black soybean during seed development and to have low expression levels in brown soybean.[10] However, genetic and biochemical differences underlying brown versus red-brown grain color remained unexamined.

Here we employed nontargeted and targeted approaches to identify differences in metabolite compositions between brown and red-brown soybean seed coats. Liquid chromatography–tandem mass spectrometry (LC-MS/MS) in multiple reaction monitoring (MRM) mode revealed quantitative differences in flavan-3-ol, procyanidin, and other flavonoids. We also identified two ANR genes (*ANR1* and *ANR2*) from soybean seed coat tissue and compared their expressions and activity in relation to seed coat color. We show that the three publicly available red-brown soybean varieties exhibit alterations in seed coat PA content, *ANR1* transcript profiles, and/or recombinant enzyme activity. On the basis of our results, we suggest that the manipulation of *ANR1* gene expression or activity may be a rational approach for engineering red-brown soybean color, which would ultimately enable the visual identification of GM grains.

MATERIALS AND METHODS

Chemicals. Cyanidin was purchased from Indofine (Somerville, NJ) and NADPH from Fisher Scientific (Ottawa, ON, Canada). (−)-Epicatechin, procyanidin B2, cyanidin 3-*O*-galactoside, and the 3-*O*-glucosides of cyanidin, quercetin, and isorhamnetin were purchased from Extrasynthese (Lyon, France). All chemicals were of the highest available purity.

Plant Materials, Growth Conditions, and Nucleic Acid Isolation. *G. max* (L.) Merr.) varieties M100 and M11 were obtained from the Agriculture and Agri-Food Canada Soybean EMS-mutagenesis collection 2005 (M. Morrison, Eastern Ontario Oilseed Research Centre, Ottawa, Canada); varieties Clark (ir; PIS47438), Soysota (r17; PI 548417), Kuro Daizu (KD; PI 81041-1), and Agate (Ag; PI 548296) were obtained from the U.S. Department of Agriculture Soybean Germplasm Collection (Agricultural Research Service, University of Illinois at Champaign-Urbana). The three red-brown varieties (Ag, KD, and M11) had seed coat luster (shiny), seed size (large), and relative timing of seed maturation (early) similar to those of the brown variety M100 (Table 1). The light-brown variety Clark (ir) and midbrown variety Soysota (r17) were selected to serve as comparators to the red-brown varieties, in addition to M100, which has a dark-brown seed coat, to better represent the range of brown seed coat color intensities. Varieties Kuro Daizu (KD) and M11 (a line produced from EMS mutagenesis of Linia N145) had red-brown color distributed across the entire seed coat surface, whereas variety Agate (Ag) with a red-brown saddle (i^k*rTo*) had pigmentation confined to a saddle-shaped region that encompassed 60–71% of the seed coat surface area (not shown).

Seeds were germinated in vermiculite and transferred to soil after 1 week. Plants were grown in a Conviron E15 cabinet under 16/8 h light/dark at 25/20 °C for 1 month, following 12/12 h light/dark at 25/20 °C to promote reproductive development. Pods were harvested into crushed ice. Seed coats were dissected from pod and embryo material, immediately frozen in liquid nitrogen, lyophilized, and stored at −80 °C. DNA was isolated from trifoliate unexpanded leaves using a kit (Qiagen). RNA was isolated from seed coats as described[22] and then washed using the RNeasy Plant Mini Kit (Qiagen) to remove impurities.

Metabolite Analysis and LC-MS/MS. To quantify total soluble and solvent-insoluble PAs, lyophilized seed coats at the 400–500 mg fresh seed weight (FSW) stage of development were ground to a fine powder in liquid nitrogen, extracted (10 mg mL^{-1}) with 70% acetone (v/v) and 0.1% ascorbic acid (w/v) in HPLC grade water, and analyzed as described previously.[23] Total anthocyanins were quantified by photospectroscopy in comparison to a standard curve of cyanidin-3-*O*-glucoside using the formula $A_{530} - 0.25A_{657}$ as described previously.[10] For compound identification, 10–12 mg of lyophilized seed coats was extracted (20 μL mg^{-1}) with MeOH/H_2O/HCO_2H (80:15:5, v/v) on ice with shaking (250 rpm) for 16 h. The supernatant was filtered through Teflon (0.2 μm; Chromatographic Specialties), and 5 μL aliquots were analyzed by LC-MS/MS in scanning mode as described previously.[10] For compound quantification, 2 μL aliquots were analyzed by LC-MS/MS in MRM mode in positive polarity at unit resolution with a scanning speed of 4000. The system consisted of a 1200 series Agilent liquid chromatograph with a high-performance autosampler (model G1376B), binary pump (model G1312A), column thermostat (model G1316A), and PEEK 0.12 mm i.d. lines from column inlet onward to a three-way splitter of a 3200 QTRAP with PEEK 0.17 mm i.d. × 60 cm lines on the splitter outlet.

For MRM, two Q1 (parent) and Q3 (product) transitions were developed per authentic standard (Supplemental Table 1 in the Supporting Information). The most sensitive transition was used for quantification and the second for authentication of identity. For quantitative optimization, authentic standards (10–20 μg mL^{-1} dissolved in methanol) were infused into the turbo ion source using an electrospary probe at a flow rate of 10 μg mL^{-1}. Flow injection analysis (FIA) was applied to optimize source conditions. The calibration curves were generated by diluting authentic standard mixes serially in methanol. Calibration curves were linear fit ($R^2 \geq 0.998$). LC conditions consisted of a linear gradient of 32–70% MeOH in 5% HCO_2H over 22 min at a flow rate of 1 mL min^{-1}. The column (Luna C18 (2), 150 × 4.6 mm, 5 μm particle size (Phenomenex Inc.)) was washed by ramping the MeOH percentage from 70 to 100% for 8 min. The initial conditions were brought back in 0.1 min, and the column was equilibrated for 6 min before the next injection.

Cloning ANR1 and ANR2 cDNAs. By searching the GenBank *Glycine max* EST collection using the BLASTn algorithm and the Lotus *corniculatus ANR1-1* sequence as a query, forward and reverse gene-specific primers (AEV1F, AEV1R, AB4F, AEB4R) were designed to amplify partial *ANR* sequences from seed coat cDNA using 5′ and 3′ RACE (Clontech), respectively. The resulting partial cDNA fragments were cloned into the TOPO-TA vector (Invitrogen) and sequenced. The full-length *ANR1* coding region was then amplified from seed coat

cDNA by end-to-end PCR using primers AHF/AHR and the Pfx50 DNA polymerase (Invitrogen). The *ANR2* cDNA was similarly cloned using primers designed on the basis of the Glyma08g06640 sequence. The resulting 1108 bp (*ANR1*) and 1017 bp (*ANR2*) amplicons were cloned into the *Nde*I and *Bam*HI sites of the pET-14b vector (Novagen) and sequenced to confirm their identities. Primers for this study are listed in Supplemental Table 2 of the Supporting Information.

Cloning ANR1 Genes. *ANR1* sequences were amplified from genomic DNA by PCR using primers OGF/OGR and the Pfx50 DNA polymerase (Invitrogen), cloned into the *Bam*HI and *Eco*RI sites of the pUC19 vector, and five clones per variety were sequenced.

Phylogenetic Analyses. A multiple alignment of the deduced amino acid sequences was performed using MAFFT.[24] Node support was estimated by parsimony bootstrap analysis (1000 bootstrap replicates, 10 random addition sequences per bootstrap replicate with tree bisection−reconnection branch swapping, limited to a maximum of 10000 trees) using PAUP 4.0b10 (Sinauer Associates).

RT-PCR and qRT-PCR. RT-PCR and qRT-PCR were performed essentially as described previously.[10] Briefly, for RT-PCR, RNA samples (3 μg) were treated with DNase I (amplification grade, Invitrogen), and first-strand cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. For quantitative RT-PCR (qRT-PCR), the cDNA template was prepared as described above and reactions (25 μL) consisted of 2 μL of first-strand cDNA (or untreated RNA), 250 nM of forward and reverse primers, and 12.5 μL of the iQ SYBR Green Supermix (Bio-Rad). qRT-PCR of each target gene for each seed coat sample was performed in triplicate on cDNA samples or untreated RNA samples. qRT-PCR data and PCR efficiencies were analyzed using Opticon Monitor 3 software (Bio-Rad). To verify the specificity of the RT-PCR reactions, melting curves were performed subsequent to each reaction in addition to fractionation of RT-PCR products on agarose gels. RT-PCR and qRT-PCR experiments were performed in triplicate. Primers used in this study can be found in Supplemental Table 2 of the Supporting Information.

Recombinant Protein Expression in *Escherichia coli*. The full-length *ANR1* and *ANR2* coding regions were fused in-frame with the pET-14b vector (above), transformed into the expression host *E. coli* BL21(DE3) pLysS (Novagen), and single colonies were selected for production of recombinant proteins. Growth, expression, and purification of soluble recombinant proteins by ion metal-affinity chromatography were performed essentially as described;[21] however, 1 L of *E. coli* was routinely cultured to have enough recombinant protein for kinetics analysis. To confirm the purity of the recombinant enzyme, the eluted fractions were visualized on 12.5% acrylamide gel stained with 0.25% Coomassie blue. The amount of purified recombinant enzyme was determined using a protein assay kit (Bio-Rad Laboratories Inc., http://www.bio-rad.com).

Recombinant Enzyme Assays and Kinetics. To measure recombinant enzyme activity for cyanidin, the concentrations of cyanidin were varied from 3 to 400 μM, and 2 mM NADPH was used as a hydride donor. Enzyme assays (total volume = 500 μL) consisted of 75 μg of recombinant enzyme, cyanidin, and 2 mM NADPH in buffer (50 mM MES, pH 5) and were incubated at 45 °C for 60 min. Gentle agitation (100 rpm) and occasional vortexing (15−20 min intervals) were provided to minimize precipitation. The amount of each reaction product was determined by comparison of integrated peak areas of epicatechin to the corresponding standard curve analyzed on the same day. Standards were processed in the same manner as enzyme reactions with the exception that the recombinant enzyme was boiled for 10 min prior to incubation. Standard curves were verified to be linear (R^2 > 0.99) over the range investigated using Chem32 software (Agilent Technologies Inc., Montreal, Quebec, Canada). All reaction solutions were brought to 45 °C prior to the addition of cyanidin. Cyanidin stocks were 10 mM in MeOH and stored at −80 °C. Reactions were stopped by vortexing in EtOAc (1 mL) for 60 s, followed by removal of the organic phase, which was then evaporated under a stream of nitrogen gas. The residue was resuspended in 100 μL of LC-MS grade MeOH. All reactions were filtered through Teflon syringe filters (0.2 μm; Chromatographic Specialties), and 20 μL injections were analyzed by HPLC-DAD as described previously.[21]

RESULTS

The Seed Coats of Three Red-Brown Soybean Varieties Have Reduced Amounts of Soluble Proanthocyanidins (PAs) Compared to Brown Soybean. Reaction of seed coat extracts with DMACA reagent showed that red-brown varieties KD and M11 had 46−71% of the amount of soluble PAs compared to the brown varieties (Figure 2A). The red-brown saddle variety Ag had 32−45% soluble PAs compared to brown varieties (significantly less than what may be attributed to the incomplete (saddle-shaped) distribution of pigments across 60−71% of the seed coat surface area) (Figure 2A). By contrast, the amounts of solvent-insoluble PAs measured following acid-catalyzed hydrolysis of seed coat residues were not found to differ significantly between red-brown and brown varieties (Figure 2B). Photospectroscopy revealed that red-brown seed coats had an increase in cyanic pigment amounts; however, the increase (0.5−1.4 mg g^{-1} lyophilized seed coat (LSC)) was much less than the decrease in soluble PA amounts (15.3−50.4 mg g^{-1} LSC) (compare panels C and A of Figure 2).

To identify and quantify differentially accumulated metabolites, we performed high-performance liquid chromatography−tandem mass spectroscopy (HPLC-MS/MS) in product ion scanning and MRM modes, respectively, in comparison to authentic standards. This approach led us to identify seven flavonoids that had accumulated to different extents in red-brown seed coats compared to the brown lines (Figure 2D). Red-brown seed coats had lower amounts of (−)-epicatechin, procyanidin B2 (line M11), and an unknown procyanidin dimer (Figure 2D, top panels) and elevated amounts of the anthocyanins cyanidin-3-*O*-glucoside (C3G) and cyanidin-3-*O*-galactoside (C3Ga) (Figure 2D, middle panels) and the flavonol quercetin-3-*O*-glucoside (Figure 2D, bottom left panel). The red-brown lines Ag and KD, but not the EMS mutagenized line M11, had elevated amounts of an unknown compound that had the same M^+ (m/z 479.0) but greater chromatographic retention (Rt, 16.3 min) compared to isorhamnetin-3-*O*-glucoside (Rt, 12.0 min) (Figure 2D, bottom right panel). Due to the high similarity of the fragments, the possibility exists that the unknown metabolite may be an isomeric form of isorhamnetin-3-*O*-glucoside. However, further structural elucidation would be needed to confirm the identity of this unknown metabolite. Overall, these results demonstrated that red-brown soybean varieties had major reductions in the amount of soluble PAs in their seed coat compared to the brown varieties, accompanied by minor increases in specific anthocyanin, flavonol, and unknown metabolites.

Identification of ANR1 and ANR2 Genes from the *G. max* Seed Coat. Yang et al.[19] recently mapped a putative anthocyanidin reductase (*ANR*) gene sequence to the genomic region coding the *O* locus[26] and speculated that *O* may correspond to *ANR*.[19] Our results show that red-brown soybean seed coats have reduced amounts of soluble PAs (Figure 2A), further suggesting the color phenotype may be associated with a defect in the PA pathway. These results prompted us to identify the functional *ANR* gene sequence(s) from soybean to determine whether red-brown seed coat color may correspond to defects in the *ANR* gene(s).

As the Glyma1 soybean genome sequence[27] (for sequence, see www.phytozome.net/soybean) was not available at the time

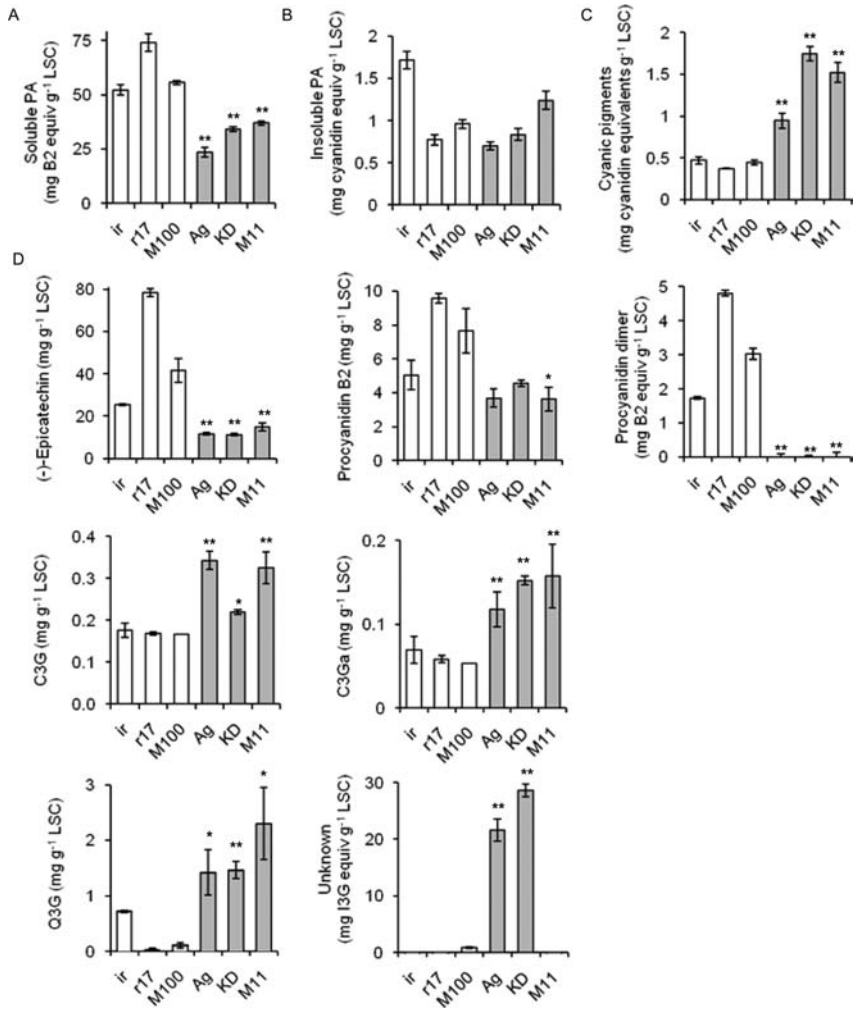

Figure 2. Flavonoid amounts from the seed coats of brown (*O*) and red-brown (*o*) soybean varieties expressed in units of milligrams of flavonoid per gram of LSC: (A) total soluble PAs by DMACA assay; (B) total solvent-insoluble PAs by acid-catalyzed hydrolysis; (C) cyanic pigments by photospectroscopy (Abs 520 nm); (D) individual flavonoids by LC-MS/MS in MRM mode. White bars represent brown (*O*) soybean varieties (ir, r17, and M100), and gray bars represent red-brown (*o*) varieties (Ag, KD, and M11). Values are the mean ± SD, $n = 3$ biological replicates. (−)-Epicatechin (Rt, 4.3 min; M^+, m/z 291.4; MS^2, m/z 139.1 and 123.1), cyanidin-3-*O*-galactoside (Rt, 3.8 min; M^+, m/z 449.4; MS^2, m/z 287.2), quercetin-3-*O*-glucoside (Rt, 9.5 min; M^+, m/z 465.0; MS^2, m/z 303.2), procyanidin B2 (Rt, 3.0 min; M^+, m/z 579.1; MS^2, m/z 291.1), cyanidin-3-*O*-glucoside (Rt, 4.2 min; M^+, m/z 449.3; MS^2, m/z 287.1), unknown (Rt, 16.3 min; M^+, m/z 479.0; MS^2, m/z 317.1).

our study was initiated, *G. max* ESTs from the GenBank collection were searched with the tBLASTn algorithm using the sequence of every reported catalytically characterized ANR protein as a query to identify *G. max* ANR candidates (September 2007). Two *G. max* ESTs (Genbank accession No. EV264997 and BM092813) were identified by tBLASTn that had high identity (>80%) to ANR1-1 from *Lotus corniculatus* (LcANR1-1; ABC71332). 5′- and 3′-RACE using seed coat cDNA from brown soybean variety Clark (ir) as a template was performed to extend the EST sequence information and revealed that both ESTs coded the same cDNA, which we named *ANR1* (JF433915). The full-length *ANR1* cDNA was 1324 bp, and the deduced amino acid sequence had high identity (83%) and length similar to LcANR1-1 (337 vs 338 residues for LcANR1-1).

Primers AHF and AHR designed to amplify from the 5-coding region to the 3′-UTR of the *ANR1* cDNA (Supplemental Table 2, Supporting Information) were used to amplify and clone the corresponding genomic regions from variety ir to gain insight into the structural organization of the *ANR1* gene in the *G. max* genome. The genomic *ANR1* sequence coded six exons and five introns (not shown), similar to the banyuls (*BAN*) gene from *Arabidopsis* (Genbank accession no. NC_003070). A BLASTn search of the Glyma1 soybean genome sequence using genomic *ANR1* as a query found it to be 100% identical to the predicted gene Glyma08g06630 (www.phytozome.net/soybean) and 70% identical to a second predicted gene, Glyma08g06640 (76% identical at the amino acid level), which we named *ANR2* (JN098512) due to its sequence similarity.

To determine the relationship of putative *G. max* ANR proteins with other catalytically characterized reductase epimerase dehydratase (RED) superfamily proteins, we performed a phylogenetic analysis. The phylogenetic tree demonstrated that ANR1 and ANR2 were more closely related to ANR proteins than to other RED superfamily proteins (Figure 3). This suggests that ANR1 and ANR2 are more likely to have anthocyanidin reductase function, rather than other RED protein functions. Gene expression analysis by RT-PCR demonstrated highly divergent expression patterns, with *ANR2* expressed at low levels in all organs and *ANR1* expressed preferentially and highly in the seed coat (Figure 4), consistent with the identification of

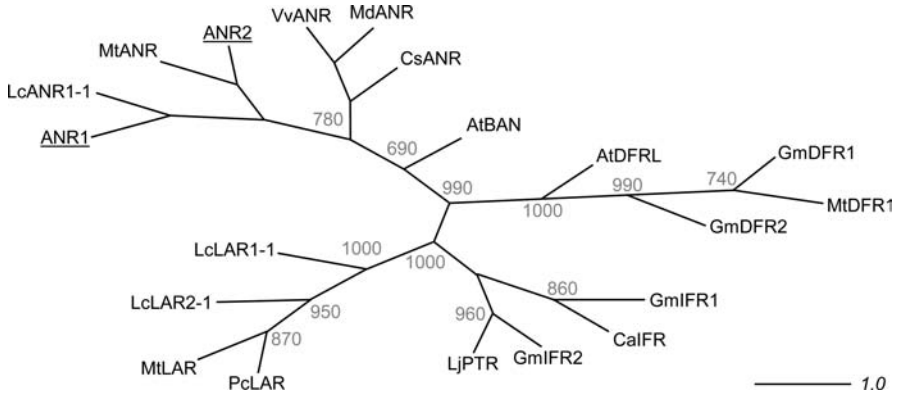

Figure 3. Phylogenetic tree of ANR and other RED superfamily proteins. Amino acid sequences of characterized enzymes from *Arabidopsis*, AtBAN (NP_176365), AtDFRL (CAP08805); *Camellia sinensis*, CsANR (AAT68773); *Cicer arietinum*, CaIFR (Q00016); *G. max*, ANR1, ANR2, GmDFR1 (AAD54273), GmDFR2 (ABM64803), GmIFR1 (AAF17577), GmIFR2 (AAF17578); *L. corniculatus*, ANR1-1 (ABC71332), LAR1-1 (ABC71324), LAR2-1 (ABC71328); *L. japonicus*, PTR (BAF34844); *Malus* × *domestica*, MdANR (AAX12184); *M. truncatula*, ANR (AAN77735), DFR1 (AAR27014), LAR (CAI56327); *Phaseolus coccineus*, LAR (CAI56322); and *Vitis vinifera* ANR (BAD89742) were aligned using MAFFT,[24] and the phylogenetic tree was produced using Dendroscope (Huson et al., 2007). The scale bar represents 1.0 substitutions per site.

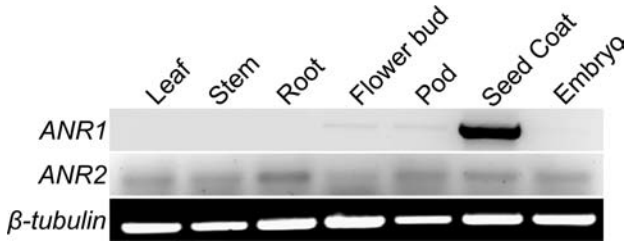

Figure 4. Transcript detection of ANR1 and ANR2 from Clark (ir) cDNA by RT-PCR with gene-specific primers.

(−)-epicatechin and (−)-epicatechin-based procyanidins in the seed coat of brown soybean.[10]

To confirm the putative catalytic function of ANR1 and ANR2 proteins for the reduction of cyanidin to (−)-epicatechin, their open reading frames were separately cloned from brown soybean (ir) seed coat cDNA into the pET-14b N-terminal hexahistidine fusion tag expression vector. The resultant plasmids were transformed into BL21(DE3) pLysS *E. coli*, expressed at 16 °C, and the recombinant his-tagged proteins were purified from the soluble fraction by ion metal affinity chromatography (IMAC). Separation of the IMAC fractions by SDS-PAGE followed by staining with Coomassie blue showed apparently pure recombinant enzymes with estimated molecular masses of ca. 39 kDa for both ANR1 and ANR2 (Figure 5A). These values matched well with their respective calculated molecular masses (38.72 kDa for ANR1 and 39.30 kDa for ANR2) with the inclusion of the 2.18 kDa N-terminal tag.

Recombinant enzyme assays were conducted using NADPH as a hydride donor and cyanidin as an acceptor. Incubation of ANR1 or ANR2 recombinant enzymes with cyanidin and NADPH resulted in the production of (−)-epicatechin, as identified by the identical retention time (Rt, 6.9 min), parent ion (M^+ m/z 291.4), and identical major fragments (m/z 139.1 and 123.1) compared to the authentic standard (Figure 5B). The boiled enzyme and bacteria expressing the corresponding empty vector could not catalyze this reaction (not shown).

In summary, two functional ANR genes were identified from the Glyma1 soybean genome sequence that exhibit different expression patterns, with *ANR2* being expressed at low levels in all organs and *ANR1* preferentially and highly expressed in the seed coat.

Red-Brown Soybean Varieties M11 and KD Have Nonsynonymous Substitutions in the ANR1 Gene That Confer Minor Reductions in Enzyme Activity. We previously reported the flavan-3-ol and PA contents of the brown soybean seed coat to be (−)-epicatechin and (−)-epicatechin-based, respectively.[10] Figure 2A demonstrates that the three publicly available red-brown soybean varieties have lower amounts of soluble PAs in the seed coat compared to brown varieties. To determine whether reduced soluble PA amounts in the red-brown seed coat may be associated with sequence-based defects in the *ANR1* gene, the 3 kb genomic region coding *ANR1* from variety M11 was cloned using primers OGF/OGR (Supplemental Table 2, Supporting Information) and compared to the *ANR1* sequences from the five other varieties (Table 1). To ensure accurate sequence information, five individual clones per variety were sequenced with at least 3 times coverage.

Genomic *ANR1* sequences from all brown varieties were identical and were the same as the predicted gene Glyma08g06630 from the reference soybean genome sequence Glyma1 (www.phytozome.net/soybean). By contrast, red-brown varieties had at least one sequence polymorphism compared to the *ANR1* sequence from brown genotypes (Table 2). Red-brown saddle variety Ag had a single base pair insertion in intron 1 at position 550. Red-brown variety KD had two nonsynonymous substitutions: T to C in exon 3 at position 1269 and G to C in exon 4 at position 1957. These mutations corresponded to the predicted amino acid changes of V122A and G214A (Table 2) in conserved and nonconserved residues, respectively (Figure 6). Red-brown EMS-mutagenized variety M11 had the same G to C substitution in exon 4 at position 1957 and had a unique substitution in exon 4 (A to T at position 1870) that conferred a predicted amino acid change of N185I (Table 2) in a conserved region (Figure 6).

To determine whether these amino acid substitutions altered enzyme activity, the respective *ANR1* ORFs from varieties KD and M11 were individually cloned into an N-terminal hexahistidine fusion tag vector for comparison of enzyme kinetics with ANR1 from brown soybean variety ir in vitro.

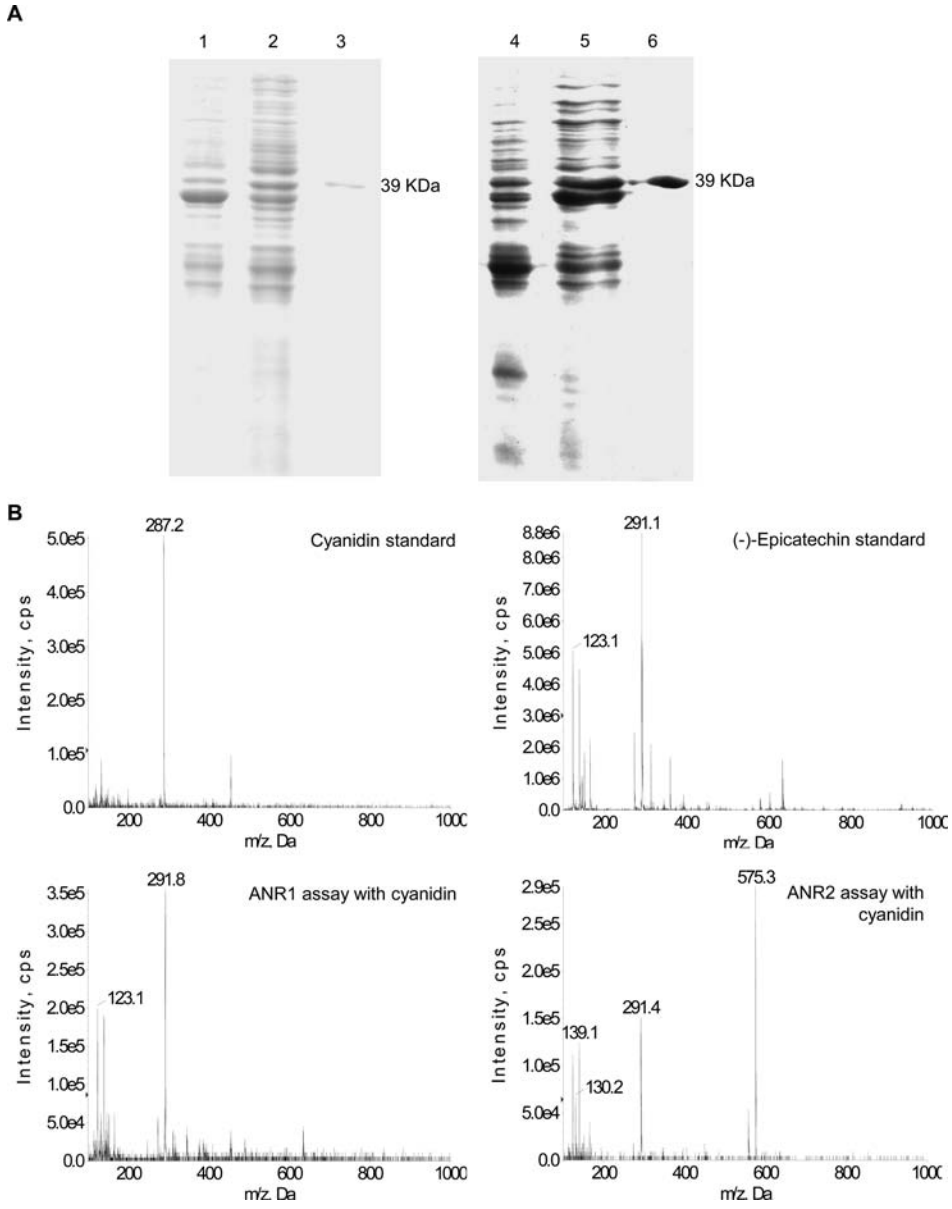

Figure 5. Purification of his-tagged ANR1 and ANR2 and identification of reaction products. (A) SDS-PAGE analysis of ANR1 and ANR2. Total soluble protein from *E. coli* expressing ANR1 prior to induction with 100 μM IPTG (lane 1), 24 h postinduction (lane 2), purified ANR1 (lane 3); total soluble protein from *E. coli* expressing ANR2 prior to induction with 100 μM IPTG (lane 4), 24 h postinduction (lane 5), and purified ANR2 (lane 6). (B) ANR1 and ANR2 enzyme reactions as revealed by LC-MS/MS. LC retention times: cyanidin (Rt, 9.0 min); (−)-epicatechin (Rt, 4.3). *m/z* 575.3 (lower right panel) corresponded to a cyanidin (Cy) dimer formed nonenzymatically, as indicated by its presence in control (boiled enzyme) and enzyme reactions at higher Cy concentrations.

Table 2. Comparison of ANR1 Genomic and Amino Acid Sequences among Brown (*O*) and Red-Brown (*o*) Soybean Varieties[a]

variety	% identity to ir (genomic DNA)	mutation	position	amino acid substitution
ir	100	none	N/A	N/A
r17	100	none	N/A	N/A
M100	100	none	N/A	N/A
Ag	99	A insertion	550 (intron 1)	N/A
KD	99	T → C	1269 (exon 3)	V122A
		G → C	1957 (exon 4)	G214A
M11	99	A → T	1870 (exon 4)	N185I
		G → C	1957 (exon 4)	G214A

[a]See Table 1 for a description of varieties.

Recombinant enzyme assays using NADPH as a hydride donor and cyanidin as an acceptor revealed minor reductions (1.8−26.2 fmol epicatechins μg^{-1}) in the activity of the ANR1 proteins from varieties KD and M11 compared to ANR1 from variety ir (Figure 7). Calculated K_m and V_{max} values are not given as the enzymes were intensely inhibited by cyanidin concentrations above ∼100 μM (not shown).

Red-Brown Soybean Seed Coats of Varieties Ag and KD Have Reduced ANR1 Transcript Levels. To determine whether reduced soluble PA amounts may be associated with transcript levels of the ANR genes, we compared *ANR1* and *ANR2* transcript profiles between red-brown and brown lines by quantitative RT-PCR (qRT-PCR) using seed coat cDNA as a template. The 25−50 mg FSW stage of seed development was

```
                                NADPH
ANR1-ir      MATIKPTGKKACVIGGSGFMASLLIKQLLEKGYAVNTTVRDPDNTKKIPHLLALQSLGEL 60
ANR1-r17     MATIKPTGKKACVIGGSGFMASLLIKQLLEKGYAVNTTVRDPDNTKKIPHLLALQSLGEL 60
ANR1-M100    MATIKPTGKKACVIGGSGFMASLLIKQLLEKGYAVNTTVRDPDNTKKIPHLLALQSLGEL 60
ANR1-Ag      MATIKPTGKKACVIGGSGFMASLLIKQLLEKGYAVNTTVRDPDNTKKIPHLLALQSLGEL 60
ANR1-KD      MATIKPTGKKACVIGGSGFMASLLIKQLLEKGYAVNTTVRDPDNTKKIPHLLALQSLGEL 60
ANR1-M11     MATIKPTGKKACVIGGSGFMASLLIKQLLEKGYAVNTTVRDPDNTKKIPHLLALQSLGEL 60
             ************************************************************

ANR1-ir      NIFGADLTGEKDFDAPIAGCELVFQLATPVNFASEDPENDMIKPAITGVLNVLKACVRAK 12C
ANR1-r17     NIFGADLTGEKDFDAPIAGCELVFQLATPVNFASEDPENDMIKPAITGVLNVLKACVRAK 12C
ANR1-M100    NIFGADLTGEKDFDAPIAGCELVFQLATPVNFASEDPENDMIKPAITGVLNVLKACVRAK 12C
ANR1-Ag      NIFGADLTGEKDFDAPIAGCELVFQLATPVNFASEDPENDMIKPAITGVLNVLKACVRAK 12C
ANR1-KD      NIFGADLTGEKDFDAPIAGCELVFQLATPVNFASEDPENDMIKPAITGVLNVLKACVRAK 12C
ANR1-M11     NIFGADLTGEKDFDAPIAGCELVFQLATPVNFASEDPENDMIKPAITGVLNVLKACVRAK 12C
             ************************************************************

ANR1-ir      GVKRVILTSSAAAVTINQLKGTDLVMDESNWTDVEYLSTAKPPTWGYPASKALAEKAAWK 18C
ANR1-r17     GVKRVILTSSAAAVTINQLKGTDLVMDESNWTDVEYLSTAKPPTWGYPASKALAEKAAWK 18C
ANR1-M100    GVKRVILTSSAAAVTINQLKGTDLVMDESNWTDVEYLSTAKPPTWGYPASKALAEKAAWK 18C
ANR1-Ag      GVKRVILTSSAAAVTINQLKGTDLVMDESNWTDVEYLSTAKPPTWGYPASKALAEKAAWK 18C
ANR1-KD      GAKRVILTSSAAAVTINQLKGTDLVMDESNWTDVEYLSTAKPPTWGYPASKALAEKAAWK 18C
ANR1-M11     GVKRVILTSSAAAVTINQLKGTDLVMDESNWTDVEYLSTAKPPTWGYPASKALAEKAAWK 18C
             *.************************************************************

ANR1-ir      FAEENHIDLITVIPTLTTGPSVTTDIPSSVGMAGSLITGNDFLINALKGMQLLSGSISIT 24C
ANR1-r17     FAEENHIDLITVIPTLTTGPSVTTDIPSSVGMAGSLITGNDFLINALKGMQLLSGSISIT 24C
ANR1-M100    FAEENHIDLITVIPTLTTGPSVTTDIPSSVGMAGSLITGNDFLINALKGMQLLSGSISIT 24C
ANR1-Ag      FAEENHIDLITVIPTLTTGPSVTTDIPSSVGMAGSLITGNDFLINALKGMQLLSGSISIT 24C
ANR1-KD      FAEENHIDLITVIPTLTTGPSVTTDIPSSVGMAASLITGNDFLINALKGMQLLSGSISIT 24C
ANR1-M11     FAEEIHIDLITVIPTLTTGPSVTTDIPSSVGMAASLITGNDFLINALKGMQLLSGSISIT 24C
             **** ****************************.************************

ANR1-ir      HVEDICRAQIFVAEKESASGRYICCAHNTSVPELAKFLSKRYPQYKIPTEFDDCPSKAKL 30C
ANR1-r17     HVEDICRAQIFVAEKESASGRYICCAHNTSVPELAKFLSKRYPQYKIPTEFDDCPSKAKL 30C
ANR1-M100    HVEDICRAQIFVAEKESASGRYICCAHNTSVPELAKFLSKRYPQYKIPTEFDDCPSKAKL 30C
ANR1-Ag      HVEDICRAQIFVAEKESASGRYICCAHNTSVPELAKFLSKRYPQYKIPTEFDDCPSKAKL 30C
ANR1-KD      HVEDICRAQIFVAEKESASGRYICCAHNTSVPELAKFLSKRYPQYKIPTEFDDCPSKAKL 30C
ANR1-M11     HVEDICRAQIFVAEKESASGRYICCAHNTSVPELAKFLSKRYPQYKIPTEFDDCPSKAKL 30C
             ************************************************************

ANR1-ir      IISSEKLVKEGFSFKYGIEEIYDQTLEYLKSKGALNN 337
ANR1-r17     IISSEKLVKEGFSFKYGIEEIYDQTLEYLKSKGALNN 337
ANR1-M100    IISSEKLVKEGFSFKYGIEEIYDQTLEYLKSKGALNN 337
ANR1-Ag      IISSEKLVKEGFSFKYGIEEIYDQTLEYLKSKGALNN 337
ANR1-KD      IISSEKLVKEGFSFKYGIEEIYDQTLEYLKSKGALNN 337
ANR1-M11     IISSEKLVKEGFSFKYGIEEIYDQTLEYLKSKGALNN 337
             *************************************
```

Figure 6. Alignment of *G. max* ANR1 proteins from brown (*O*) seed coat varieties (ir, r17, M100) and red-brown (*o*) seed coat varieties (Ag, KD, and M11) using the ClustalW program with default parameters. Mutations are shown with black background. Gray background represents amino acid residues that are conserved among all catalytically validated ANR proteins: (AtBAN (NP_176365), CsANR (AAT68773); *G. max* ANR1-ir, LcANR1-1 (ABC71332MdANR (AAX12184); *M. truncatula*, ANR (AAN77735); and *Vitis vinifera* ANR (BAD89742). The putative NADPH binding site and the catalytic triad (Gomez et al., 2009) are represented by a thick black line above the sequences and boxed amino acids, respectively.

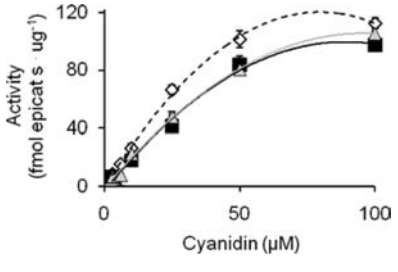

Figure 7. In vitro kinetics of recombinant ANR1 enzymes from brown soybean variety ir (◇) and red-brown soybean varieties M11 (■) and KD (△). Activity was measured as amount of epicatechin produced per second per microgram of enzyme.

chosen as the developmental stage for qRT-PCR analysis because *ANR1* transcripts were previously shown to be at maximum levels at this stage in brown (*irTO*) variety Clark (ir) and its black (*iRTO*) isoline.[10] Phosphenol pyruvate carboxylase (*PEPC*) was used as an endogenous reference for normalization of the qRT-PCR measurements of target genes across seed coat samples, as *PEPC* has been used previously as a reference for gene expressions from pigmented and nonpigmented soybean seed coats and is expressed at similar levels in a wide range of soybean tissues.[9,10,28] The expressions of other genes situated at the branchpoint between PA and anthocyanin biosynthesis, (namely, *UGT78K2* (HM591298) coding UF3GT,[10] *LAR1* (Genbank Accession No. JF433916) putatively coding leucoanthocyanidin

reductase,[10] and *ANS2/ANS3* (Genbank Accession No. AY382829/AY382830) coding anthocyanidin synthases[29] were also measured to investigate a relationship with seed coat color.

qRT-PCR analysis demonstrated that only *ANR1* and *ANS2/ANS3* were expressed at consistently high levels among all brown varieties, with the other genes showing highly variant expression levels with no consistent correlation with seed color (Figure 8). The EMS-mutagenized variety M11 had similar

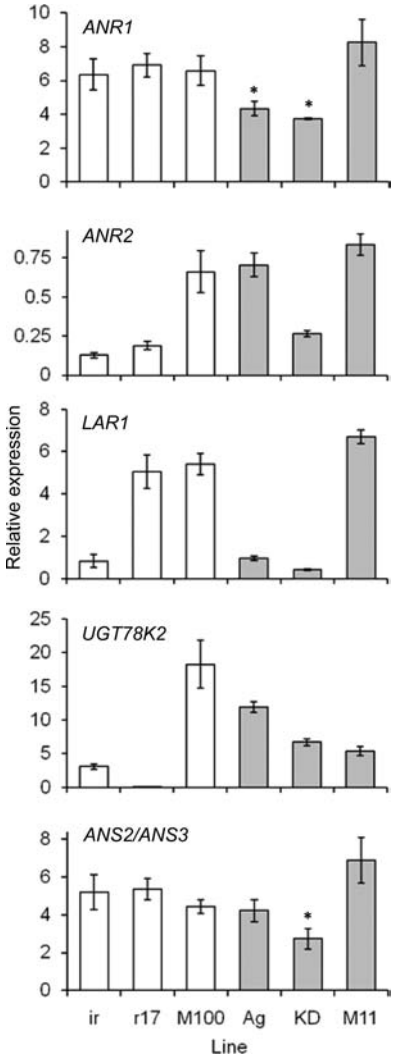

Figure 8. Seed coat transcript profiles of PA and anthocyanin branchpoint genes measured by qRT-PCR: PA genes *ANR1*, *ANR2*, and *LAR1*; anthocyanin genes *UGT78K2* and *ANS2/ANS3*. Student's *t* test indicated significance at $P < 0.05$ (*).

ANR1 expression levels compared to those of the brown varieties; however, of all the genes investigated, only *ANR1* expressions were significantly reduced in spontaneous red-brown mutants KD and Ag (Figure 8).

DISCUSSION

The seed coats of yellow (*I*) soybean have reduced amounts of flavonoids compared to those of pigmented (*i*) genotypes due to differences in the expressions of chalcone synthase genes mediated by a post-transcriptional gene silencing mechanism.[7] The seed coats of brown (*r*) soybean have reduced amounts of specific anthocyanins, PAs, isoflavones, and phenylpropanoids, compared to black (*R*) soybean, due to differences in the expressions of a putative regulatory gene.[10] Here we have shown that three red-brown soybean accessions have major reductions in the amounts of soluble PAs compared to brown soybean lines (Figure 2A). A previous study demonstrated that red fava beans (*Vicia faba*) have reduced PA content compared to the brown variety with normal flower color.[30] Similarly, red to orange-brown and dark red varieties of cowpea (*Vigna unguiculata*) had reduced PA content compared to their beige comparators.[31] These results have significant implications for the metabolic engineering of seed color for the purposes of marking commercial transgenic legume grains; they suggest that cyanic (e.g., red-brown) color may be engineered to mark GM grains by the suppression of PA biosynthesis in the seed coat. Furthermore, the reduction of PA amounts would be favorable for postharvest processing of the GM grain, as PAs bind proteins and other metabolites, interfering with their extraction.

Our LC-MS/MS analysis of red-brown seed coat metabolites demonstrated that the major reduction in soluble PA content was accompanied by minor increases in the amounts of anthocyanin pigments (C3G and C3Ga) (Figure 2). However, red color in beans may not always be associated with increased anthocyanin content, as commercial dark red kidney beans (*Phaseolus vulgaris* L.) mainly contained flavonol-glycosides and PAs with no anthocyanins identified by HPLC-DAD.[32] Other pigments identified from soybean seed coats that could provide a cyanic color include epiflavan-3-ol−anthocyanin conjugates, such as catechin−cyanidin-3-*O*-glucoside identified from the seed coat of black soybean cultivar Cheongja 3,[33] or 3-deoxyanthocyanidins, such as diosmetinidin identified from the seed coat of black soybean variety Clark.[10] However, these compounds were not identified as major pigments from the red-brown soybean seed coat by HPLC-DAD measuring absorbance at 520 nm (not shown) or as metabolites that have accumulated to different amounts compared to brown soybean by LC-MS/MS. Our LC-MS/MS results demonstrated that 3-*O*-glycosylated anthocyanins (C3G and C3Ga) were the main cyanic pigments present at significantly higher levels in the seed coat of red-brown soybean compared to the brown lines (Figure 2D).

Anthocyanin and PA biosynthesis are linked by the requirement for common flavonoid precursors; the PA enzyme anthocyanidin reductase (ANR) and the anthocyanin enzyme UDP-glycose:flavonoid-3-*O*-glycosyltransferase (UF3GT) both utilized cyanidin as a substrate (Figure 1). This suggests the possibility that the inhibition of PA biosynthesis by defect in ANR would render more cyanidin substrate available to UF3GT for anthocyanin biosynthesis and could result in a red-brown seed coat phenotype. In *Arabidopsis*, mutations causing loss of function of the *ANR* gene (also known as *BANYULS*) resulted in the loss of PAs in the seed coat and the temporary accumulation of red pigments (not formally identified as anthocyanins) during seed development.[25] In soybean, our results demonstrated that defects in the *ANR1* gene were accompanied by major reductions in PA content and minor elevations in the amounts of anthocyanins in the red-brown seed coat (Figure 2D). Although gene expression analysis demonstrated that *ANR1* was highly expressed in the seed coat compared to *ANR2* (Figure 4), it remains unclear whether the differences observed in gene expression or enzyme activity are sufficient to explain the phenotypic variation between brown and red-brown seed coats. Recently, a partial soybean sequence with 86% sequence identity to the *M. truncatula ANR* gene was

mapped to the genomic region coding the *O* locus.[19] The genes coding *ANR1* (Glyma08g06630) and *ANR2* (Glyma08g06640) were located in the same region of the Glyma1 soybean genome sequence, further supporting that either may encode the *O* gene. However, further analysis, such as gene complementation or gene suppression, will be required to substantiate that defects in the *ANR1* gene are responsible for red-brown seed coat color and that *ANR1* may be the *O* gene. The possibility exists that other regulatory, biosynthetic, or trafficking differences in the PA pathway not related to *ANR1* may be responsible for reduced PA content in red-brown seed coats.

In conclusion, a better understanding of the biochemistry of red-brown seed coat color and the role of the soybean *ANR1* gene in PA biosynthesis has been achieved in this work. We confirmed the in vitro catalytic function of two ANR genes isolated from the seed coat of brown soybean. Of the two genes, ANR1 was preferentially and highly expressed in the seed coat, suggesting a principal role in PA biosynthesis in this tissue. ANR1 exhibited reduced gene expression profiles in the seed coat of red-brown soybean relative to brown soybean and/or minor reductions in recombinant enzyme activity. Further analysis will be required to determine whether the differences observed in *ANR1* gene expression or enzyme activity are sufficient to explain the phenotypic variation between red-brown and brown seed coats. On the basis of these results, we are presently attempting to suppress the *ANR1* gene in brown soybean using RNA interference (RNAi) to determine whether such a mechanism will be sufficient to engineer a red-brown soybean grain color. We predict that the suppression of *ANR1* transcript profiles will result in the redirection of metabolic flux into the anthocyanin (and possibly other metabolite) pathways to produce a red-brown soybean grain color. If successful, the RNA interference construct could be linked to commercial transgenic traits to enable the visual detection of GM soybeans in the marketplace and abroad.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

MRM parameters (Supplemental Table 1) and primers used in this study (Supplemental Table 2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: John.Arnason@uottawa.ca.

Funding

This work was funded by the A-base (project inventory 126) from Agriculture and Agri-Food Canada to B.M. and NSERC Discovery Grants to B.M. and J.T.A.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the USDA-ARS and Malcolm Morrison of the AAFC for providing the seeds used in this study.

■ REFERENCES

(1) Gould, K. S. Nature's Swiss Army Knife: the diverse protective roles of anthocyanins in leaves. *J. Biomed. Biotechnol.* **2004**, *2004* (5), 314−320.

(2) Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* **2001**, *126* (2), 485−493.

(3) Grotewold, E. The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* **2006**, *57*, 761−780.

(4) Butelli, E.; Titta, L.; Giorgio, M.; Mock, H. P.; Matros, A.; Peterek, S.; Schijlen, E. G.; Hall, R. D.; Bovy, A. G.; Luo, J.; Martin, C. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* **2008**, *26* (11), 1301−1308.

(5) Dixon, R. A. Engineering of plant natural product pathways. *Curr. Opin. Plant Biol.* **2005**, *8* (3), 329−336.

(6) Kovinich, N.; Arnason, J. T.; DeLuca, V.; Miki, B. Coloring soybeans with anthocyanins? In *Recent Advances in Phytochemistry*; Gang, D. R., Ed.; Springer: New York, 2011; Vol. *41*, pp 47−57.

(7) Bernard, R.; Weiss, M. Qualitative genetics. In *Soybeans: Improvement, Production, and Uses*, 1st ed.; American Society of Agronomy: Madison, WI, 1973; pp 117−154.

(8) Todd, J.; Vodkin, L. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant Cell* **1996**, *8* (4), 687−699.

(9) Tuteja, J. H.; Clough, S. J.; Chan, W. C.; Vodkin, L. O. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. *Plant Cell* **2004**, *16* (4), 819−835.

(10) Kovinich, N.; Saleem, A.; Arnason, J. T.; Miki, B. Combined analysis of transcriptome and metabolite data reveals extensive differences between black and brown nearly-isogenic soybean (*Glycine max*) seed coats enabling the identification of pigment isogenes. *BMC Genomics* **2011**, *12*, 381.

(11) Buzzell, R.; Buttery, B.; MacTavish, D. Biochemical genetics of black pigmentation of soybean seed. *J. Heredity* **1987**, *78* (1), 53−54.

(12) Todd, J. J.; Vodkin, L. O. Pigmented soybean (*Glycine max*) seed coats accumulate proanthocyanidins during development. *Plant Physiol.* **1993**, *102* (2), 663−670.

(13) Woodworth, C. M. Inheritance of cotyledon, seed-coat, hilum and pubescence colors in soy-beans. *Genetics* **1921**, *6* (6), 487−553.

(14) Toda, K.; Yang, D.; Yamanaka, N.; Watanabe, S.; Harada, K.; Takahashi, R. A single-base deletion in soybean flavonoid 3′-hydroxylase gene is associated with gray pubescence color. *Plant Mol. Biol.* **2002**, *50* (2), 187−196.

(15) Zabala, G.; Vodkin, L. Cloning of the pleiotropic T locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3′ hydroxylase. *Genetics* **2003**, *163* (1), 295−309.

(16) Zabala, G.; Vodkin, L. O. A rearrangement resulting in small tandem repeats in the F3′5′H gene of white flower genotypes is associated with the soybean W1 locus. *Plant Genome* **2007**, *47* (S2), S113−S124.

(17) Zabala, G.; Vodkin, L. O. The wp mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. *Plant Cell* **2005**, *17* (10), 2619−2632.

(18) Xie, D. Y.; Sharma, S. B.; Paiva, N. L.; Ferreira, D.; Dixon, R. A. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science* **2003**, *299* (5605), 396−399.

(19) Yang, K.; Jeong, N.; Moon, J. K.; Lee, Y. H.; Lee, S. H.; Kim, H. M.; Hwang, C. H.; Back, K.; Palmer, R. G.; Jeong, S. C. Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. *J. Heredity* **2010**, *101* (6), 757−768.

(20) Albert, S.; Delseny, M.; Devic, M. BANYULS, a novel negative regulator of flavonoid biosynthesis in the *Arabidopsis* seed coat. *Plant J.: Cell Mol. Biol.* **1997**, *11* (2), 289−299.

(21) Kovinich, N.; Saleem, A.; Arnason, J. T.; Miki, B. Functional characterization of a UDP-glucose:flavonoid 3-*O*-glucosyltransferase from the seed coat of black soybean (*Glycine max* (L.) Merr.). *Phytochemistry* **2010**, *71* (11−12), 1253−1263.

(22) Wang, C.-S.; Vodkin, L. O. Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. *Plant Mol. Biol. Rep.* **1994**, *12* (2), 132−145.

(23) Li, Y.-G.; Tanner, G.; Larkin, P. The DMACA-HC1 protocol and the threshold proanthocyanidin content for bloat safety in forage legumes. *J. Sci. Food Agric.* **1996**, *70*, 89−101.

(24) Katoh, K.; Asimenos, G.; Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **2009**, *537*, 39−64.
(25) Devic, M.; Guilleminot, J.; Debeaujon, I.; Bechtold, N.; Bensaude, E.; Koornneef, M.; Pelletier, G.; Delseny, M. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. *Plant J.* **1999**, *19* (4), 387−398.
(26) Weiss, M. Genetic linkage in soybeans. Linkage group VII. *Crop Sci.* **1970**, *10*, 627−629.
(27) Schmutz, J.; Cannon, S. B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D. L.; Song, Q.; Thelen, J. J.; Cheng, J.; Xu, D.; Hellsten, U.; May, G. D.; Yu, Y.; Sakurai, T.; Umezawa, T.; Bhattacharyya, M. K.; Sandhu, D.; Valliyodan, B.; Lindquist, E.; Peto, M.; Grant, D.; Shu, S.; Goodstein, D.; Barry, K.; Futrell-Griggs, M.; Abernathy, B.; Du, J.; Tian, Z.; Zhu, L.; Gill, N.; Joshi, T.; Libault, M.; Sethuraman, A.; Zhang, X. C.; Shinozaki, K.; Nguyen, H. T.; Wing, R. A.; Cregan, P.; Specht, J.; Grimwood, J.; Rokhsar, D.; Stacey, G.; Shoemaker, R. C.; Jackson, S. A. Genome sequence of the palaeopolyploid soybean. *Nature* **2010**, *463* (7278), 178−183.
(28) Sugimoto, T.; Kawasaki, T.; Kato, T.; Whittier, R. F.; Shibata, D.; Kawamura, Y. cDNA sequence and expression of a phosphoenolpyruvate carboxylase gene from soybean. *Plant Mol. Biol.* **1992**, *20* (4), 743−747.
(29) Ku, H.-M. *Molecular Cloning and Expression of ANS (Anthocyanidin Synthase) and GST (Glutathione S-Transferase) in Seed Coats of Soybean*; National Chung Hsing University: Taichung, Taiwan, 2009.
(30) Martin, A.; Cabrera, A.; Lopez-Medina, J. Tannin content in *Vicia faba*: possibilities for plant breeding. *Options Mediterr.* **1995**, *10*, 105−110.
(31) Price, M. L.; Hagerman, A. E.; Butler, L. G. Tannin content of cowpeas, chickpeas, pigeon peas, and mung beans. *J. Agric. Food Chem.* **1980**, *28* (2), 459−461.
(32) Beninger, C. W.; Hosfield, G. L. Flavonol glycosides from Montcalm dark red kidney bean: implications for the genetics of seed coat color in *Phaseolus vulgaris* L. *J. Agric. Food Chem.* **1999**, *47* (10), 4079−4082.
(33) Lee, J. H.; Kang, N. S.; Shin, S.-O.; Shin, S.-H.; Lim, S.-G.; Suh, D.-Y.; Baek, I.-Y.; Park, K.-Y.; Ha, T. J. Characterization of anthocyanins in the black soybean (*Glycine max* L.) by HPLC-DAD-ESI/MS analysis. *Food Chem.* **2009**, *112*, 226−231.