Genetic fine mapping and genotype-phenotype associations of anthocyanin biosynthetic genes in soybean

Soon-Chun Jeong

Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongwon, Chungbuk 363-883, Republic of Korea, Email scjeong@kribb.re.kr

Abstract

Proanthocyanidins and anthocyanins derived from the phenylpropanoid pathway most likely play a protective role from pathogens and UV light exposure within the plant, but are now attracting attention because of the medicinal and nutritional values due to their antioxidant properties and flavors. Three independent loci (*I*, *R*, and *T*) distinct from flower color-controlling loci control pigmentation of the seed coats in soybean (*Glycine max*) and several loci including *W1*, *W3*, and *Wp* control flower colors. The objectives of this study were to develop PCR-based molecular markers cosegregating with the genetic loci controlling the anthocyanin biosynthesis using public soybean EST and genomic sequence data. We have developed molecular markers from all major soybean anthocyanin biosynthetic genes and have shown their associations with the above mentioned genetic loci.

Key Words

Anthocyanin, flower color, proanthocyanidin, seed coat color, soybean

Introduction

Colors are one of the earliest genetically studied characters ever since Mendel. All genetic studies made on soybean (*Glycine max* (L.) Merr.) early 1900s after rediscovery of Mendel's law had been concerned with the color characters (Woodworth, 1921). Flower, seed coat and pubescence colors of soybean have provided little practical implications except the use as markers to ascertain a success in crossing. Several recent studies have revolved on partial seed coat pigmentation as results of chilling stress or viral diseases, because the pigmentation degrade the external appearance of soybean seeds and reduce their commercial value (e.g. Gore et al., 2002; Senda et al., 2004). But natural products that cause colors including flavonoid and anthocyanins are now attracting attention because of their medicinal and nutritional values due to their antioxidant properties and flavors (reviewed in Dixon and Sumner, 2003). Most soybean cultivars grown and consumed throughout the world today has yellow (also called colorless) seed coat while most known accessions of the wild progenitor, *G. soja*, have black or, rarely, brown seed coat. Thus, although domestication history of soybean have not been well documented, the appearance of seed coat color apparently suggests that ancient farmers preferred yellow soybean seed coats over the course of soybean domestication, as recently documented to the very strong, positive selection by ancient farmers for a single

white mutant allele for the change in pericarp color from red to white over the course of rice domestication (Sweeney et al., 2007). Thus, studies of color in practical aspects should direct to both ways that remove or add color. Pigmentation of flower and seed coat is caused by the deposition of various flavonoids in the respective tissues in soybean. So far, alleles of at least 5 genetic loci (*I*, *T*, *W1*, *R*, and *O*) are known to act epistatically to control seed coat pigmentation and 5 genetic loci (*W1*, *W3*, *W4*, *Wm*, and *Wp*) flower pigmentation (reviewed by Palmer et al., 2004). The objective of this study was to develop high throughput molecular markers cosegregating with genes or loci affecting biosyntheses of anthocyanins in soybean. The different flower and seed coat colors may reflect mutations affecting enzymes at different steps of the anthocyanin biosynthetic pathway. Therefore, we have developed markers using sequence information from genes involved in the anthocyanin metabolic pathway. The markers will facilitate for the marker-assisted selection of the seed coat color traits in a breeding program.

Materials and methods

A population of 112 F_{12} recombinant inbred lines generated by an interspecific cross between a *Glycine max* line 'Hwangkeum' and a *G. soja* Siebold & Zucc. line 'IT182932' (HI population) was used to construct a framework map consisting of 20 soybean linkage groups and subsequently to determine a genetic map location of anthocyanin biosynthetic genes. Sequence-based markers were developed from various sources of gene sequence information.

Results and discussion

Three independent loci (I, R, and T) control biosynthesis of the pigments determining the seed coat colors in soybean and are distinct from those controlling flower color. The *I* locus locates at a region containing a cluster of chalcone synthase (CHS) genes on soybean linkage group A2 (Todd and Vodkin, 1995) and controls distribution of anthocyanin and proanthocyanidin pigments. The dominant *I* allele completely exhibits colorless seed coat phenotype due to dominance inhibition. The soybean BAC clone 104J7 (BAC104J7) spanning the *I* locus was completely sequenced and annotated (Clough et al., 2004). The BAC clone (BAC104J7) sequence that harbors an inverted perfect repeat cluster of CHS genes comprising the *I* locus was used to generate three microsatellite markers. Linkage mapping confirmed that these markers are located at the same position as the *I* locus in molecular linkage group A2.

The *R* and *T* loci control specific seed coat color such as black (*i, R, T*), imperfect black (*i, R, t*), brown (*i, r, T*), or buff (*i, r, t*) by determining types of the anthocyanin and proanthocyanidin pigments. The *R* locus locates on soybean linkage group K (Lark et al., 1993). However, molecular nature of the *R* locus has not been determined to date. The *T* locus locates on soybean linkage group C2. Cloning and mapping of the soybean *flavonoid 3'-hydroxylase* (*F3'H*) gene revealed that the *F3'H* gene is cosegregating with the *T* locus (Toda et al., 2002; Zabala and Vodkin, 2003). We used the genomic sequences of the *F3'H* genome sequence for the development of markers. Linkage mapping confirmed that the markers are located at the same position as the *T* locus in molecular linkage group C2.

Most cultivars have either purple (*W1W1*) or white (*w1w1*) flowers. Biochemical genetics pointed to

flavonoid 3'5'-hydroxylase (*F3'5'H*) as a likely candidate of *W1*-encoding gene. In this study, we used the *F3'5'H* cDNA sequence with GenBank accession no. AY117551 deposited by Liao et al. in 2003, which is the same sequence used by Zabala and Vodkin (2007). The genomic DNA sequences corresponded to parts of first and second exons and first intron of the *F3'5'H* genomic DNA sequence reported by Zabala and Vodkin (2007) and revealed three polymorphic loci.

Magenta and pink flower colors results from the *W1_wmwm* and from the *W1_wpwp* genotypes, respectively (Palmer et al., 2004). The *Wp* locus, which changes a purple-flowered phenotype to pink in soybean, was reported to encode *flavonoid 3-hydroxylase* (*F3H*) 1 gene on the bases of cDNA microarrays and the analyses of purple and pink flower isolines and a *Tgm-Express1*-insertion line (Zabala and Vodkin, 2005). However, genetic linkage relationship of the *F3H* gene with public molecular markers has not been determined. The genomic DNA sequences of the *F3H1* and *F3H2* genes were used to generate molecular markers, respectively. Both markers cosegregated in the HI population.

The genotype *W1_W3_w4w4* produces the purple-throat phenotype (also known as dilute purple), and the genotype *W1_w3w3w4w4* produces the near-white phenotype. The *W3* locus, which controls the production of the purple-throat phenotype in the genotype *W1_W3_w4w4* and of the near-white phenotype in the genotype *W1_w3_w4w4*, was reported to cosegregate with the pDFR200, which contains part of the *dihydroflavonol 4-reductase 1* (*DFR1*) sequence, on the bases of RFLP analysis. (Fasoula et al., 1995). In an effort to develop markers from DFR genes, we found three paralogous DFR genes and developed markers from all three genes. The three genes mapped on three different LGs. Interestingly, markers developed from the DFR2 sequence mapped between Satt386 and Sct_137 on MLG D2, which is corresponding to the genetic location of the W4 locus (Xu and Palmer, 2005).

The markers presented in this report should facilitate markers-assisted selection of seed coat colors in molecular breeding programs and, in addition, play an important role in further elucidation of anthocyanin biosynthetic mechanism in soybean. The soybean genome, which is widely regarded as an ancient diploidized tetraploid (Shoemaker et al. 1996), has undergone extensive rearrangement and additional duplications since the initial polyploidization (Lee et al. 2001). Thus, even in case that a gene encoding a genetic locus was cloned in other plants, mapping of a candidate gene cosegregating with a genetic locus in soybean was not straightforward and required development of markers from highly homologous but unlinked gene family members consisting of 2 or more genes. Future research will be directed to the development of markers cosegregating with the *R* locus so that breeders can design which soybean seed coat color their new soybean cultivars have.

References

Clough, SJ, JH Tuteja, M Li, LF Marek, RC Shoemaker and LO Vodkin. 2004. Features of a 103-kb gene-rich region in soybean include an inverted perfect repeat cluster of CHS genes comprising the *I* locus. Genome 47: 819 - 831

Dixon, RA and LW Summer. 2003. Legume natural products: Understanding and manipulating complex pathways for human and animal health. Plant Physiol. 131: 878-885.

Fasoula, DA, PA Stephens, CD Nickell, and LO Vodkin. 1995. Cosegregation of purple-throat flower color

with dihydroflavonol reductase polymorphism in soybean. Crop Sci. 35: 1028-1031.

Gore, MA, AJ Hayes, SC Jeong, YG Yue, GR Buss, MA Saghai Maroof. 2002. Mapping tightly linked genes controlling potyvirus infection at the Rsv1 and Rpv1 region in soybean. Genome. 45: 592-599.

Lark, KG, JM Weisemann, BF Matthews, R Palmer, K Chase, and T Macalma. 1993. A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. Theor. Appl. Genet. 86: 901-906.

Lee, JH, D Grant, CE Vallejos, and RC Shoemaker. 2001. Genome organization in dicots. II. Arabidopsis as a "bridging species" to resolve genome evolution events among legumes. Theor. Appl. Genet. 103: 765–773.

Palmer, RG, TW Pfeiffer, GR Buss, and TC Kilen. 2004. Qualitative genetics. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. ASA, CSSA, SSSA, Madison, pp 137-214.

Senda, M., C Masuta, S Ohnishi, K Goto, A Kasai, T Sano, JS Hong and S MacFarlane. 2004. Patterning of virus-infected *Glycine max* seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. Plant Cell 16: 807-818.

Shoemaker, RC, K Polzin, J Labate, J Specht, EC Brummer, T Olson, N Young, V Concibido, J Wilcox, JP Tamulonis, G Kochert, and HR Boerma. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144: 329–338.

Sweeney, MT, MJ Thomson, YG Cho, YJ Park, SH Williamson, CD Bustamante, and SR McCouch. 2007. Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genetics 3: 1418-1424.

Toda, K, D Yang, N Yamanaka, S Watanabe, K Harada and R Takahashi. 2002. A single-base deletion in soybean flavonoid 3'-hydroxylase gene is associated with gray pubescence color. Plant Mol. Biol. 50: 187-196.

Todd, JJ and EO Vodkin. 1996. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell 8: 687-699.

Tuteja, JH, SJ Clough, WC Chan and LO Vodkin. 2004. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. Plant Cell 16: 819-835.

Woodworth, CM. 1921. Inheritance of cotyledon, seed-coat, hilum, and pubescence colors in soy-beans. Genetics 6: 487-553.

Xu, M and RG Palmer. 2005. Genetic analysis and molecular mapping of a pale flower allele at the *W4* locus in soybean. Genome 48: 334-340.

Zabala, G and LO Vodkin. 2003. 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 163: 295-309.

Zabala, G. and LO Vodkin. 2005. The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell 17: 2619-2632.

Zabala, G. and LO Vodkin. 2007. A rearrangement resulting in small tandem repeats in the F3'5'H gene of white flower genotypes is associated with the soybean *W1* locus. Crop Sci. 47: S113-S124.