Canadian Journal of Plant Science
Revue canadienne de phytotechnie

Population structure analysis and association mapping of seed antioxidant content in USDA cowpea (*Vigna unguiculata* L. Walp.) core collection using SNPs

Journal:	<i>Canadian Journal of Plant Science</i>
Manuscript ID	CJPS-2016-0090.R2
Manuscript Type:	Article
Date Submitted by the Author:	13-May-2016
Complete List of Authors:	Qin, Jun Shi, Ainong Xiong, Haizheng Mou, Bei-quan Motes, Dennis R. Lu, weiguo Miller, Jr., Creighton Scheuring, Douglas C. Nzaramba, M. Ndambe Weng, Yuejin Yang, Wei
Keywords:	Cowpea, <i>Vigna unguiculata</i> , seed antioxidant content, seed color, single nucleotide polymorphism (SNP)

SCHOLARONE™
Manuscripts

Population structure analysis and association mapping of seed antioxidant content in USDA cowpea (*Vigna unguiculata* L. Walp.) core collection using SNPs

Jun Qin^{1,2}, Ainong Shi^{1,7}, Haizheng Xiong¹, Beiquan Mou³, Dennis Motes⁴, Weiguo Lu¹, J. Creighton Miller Jr.⁵, Douglas C. Scheuring⁵, M. Ndambe Nzaramba⁶, Yuejin Weng¹, Wei Yang¹

¹Department of Horticulture, University of Arkansas, Fayetteville, AR 72701, USA; ²National Soybean Improvement Center Shijiazhuang Sub-Center. North China Key Laboratory of Biology and Genetic Improvement of Soybean Ministry of Agriculture. Cereal & Oil Crop Institute, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang 050031, P.R. China; ³Crop Improvement and Protection Research Unit, US Department of Agriculture, Agricultural Research Service (USDA-ARS), 1636 E. Alisal Street, Salinas, CA 93905, USA; ⁴Vegetable Research Center, University of Arkansas, Alma, AR 72921, USA; ⁵Department of Horticultural Sciences, Texas A&M University, College Station, TX 77843, USA; ⁶Ministry of Agriculture and Animal Resources, BP. 621 Kigali, Rwanda; ⁷Corresponding author, Email: ashi@uark.edu

With 3 figures and 4 tables; 6 supplement tables.

Abstract

Cowpea (*Vigna unguiculata* L. Walp.) is an important legume, and the antioxidant content in cowpea seeds has been recognized as a health-promoting compound for humans. The objectives of this study were to analyze the population structure of cowpea collections and to identify single nucleotide polymorphism (SNP) markers associated with the seed antioxidant content and seed coat color. A set of 1,047 SNPs were used to analyze 369 cowpea core collection from 47 countries. Results indicated that: (1) there were three clusters in the 369 entries; and the germplasm collected from India, South Africa, and the US showed broader genetic diversity; (2) Scaffold7139_14363 and Scaffold29110_4657 were strongly associated with antioxidant content, and C35063613_1497, Scaffold81493_886, and Scaffold84620_6785 were strongly associated with seed coat color across three models; (3) significant correlations were detected between the seed antioxidant content and black seed color ($r = 0.45$), between seed antioxidant content and red seed coat color ($r = 0.50$); and (4) Scaffold42008_191 and C35082838_2258 were associated with both seed antioxidant content and seed coat color. The SNP markers identified could potentially be used in marker-assisted breeding to accelerate genetic improvement of cowpea for higher seed antioxidant content.

Key words: Cowpea – *Vigna unguiculata* – seed antioxidant content – seed color – single nucleotide polymorphism (SNP) – association analysis

Introduction

Cowpea (*Vigna unguiculata* L. Walp.) is an annual legume and often referred to as southern pea, black-eye pea, crowder pea, lubia, niebe, coupe or frijole (Oplinger et al. 1991). Cowpea originated in Africa (Singh et al. 2003; Ehlers 1997; Ehlers et al. 2007). Cowpea is one of the

most ancient food sources and has probably been used as a crop plant since Neolithic times (Summerfield et al. 1985). The precise center of origin of cowpea has been difficult to determine as it is broadly-adapted and cultivated around the world as a pulse, especially in tropical and subtropical areas. Today, cowpea is widely grown in Africa, Latin America, Southeast Asia and the southern US (Muchero et al. 2009; Tan et al. 2012). Cowpea is called “poor man’s meat” because the seed protein contents range from 23% to 32% of seed weight rich in lysine and tryptophan, and a substantial amount of mineral and vitamins (folic acid and vitamin B) necessary for preventing birth defect during the pregnancy stage (Tan et al. 2012). Cowpea was found high in methionine and threonine compared with chickpea, lentil, cowpea and green pea. The total essential amino acids were maximum in cowpea (Iqbal et al. 2006). Cowpea is low in fat, but rich in dietary fiber and a variety of micronutrients and phytochemicals (Kadam et al. 1985; Siddhuraju and Becker 2007). Recent evidence suggests that cowpea is effective at binding cholesterol and lowering blood cholesterol in hamsters, thus contributing to chronic disease prevention (Frota et al. 2008). Research shows that regular consumption of dry beans and other legumes may reduce serum cholesterol, improve diabetic therapy, and provide metabolic benefits that aid in weight control (Anderson et al. 1999; Winham et al. 2007), as well as reduce the risk of coronary heart disease (Bazzano et al. 2001; Winham et al. 2007), and cancer (Abraído-Lanza et al. 2006).

A number of studies have demonstrated that cowpeas have high antioxidant content (Nzaramba et al. 2005; Siddhuraju and Becker 2007). Antioxidants are widely used in dietary supplements and have been investigated for the prevention of diseases such as cancer, coronary heart disease and even altitude sickness (Brown et al. 2003; Rafieian-Kopaei et al. 2013). An antioxidant is a molecule that inhibits the oxidation of other molecules, removes free radical intermediates, and

inhibits other oxidation reactions, thus reducing agents such as thiols, ascorbic acid, or polyphenols to prevent diseases (Rahman 2007; Valko et al. 2007). Zia-Ul-Haq *et al* (2013) determined the antioxidant activity of the methanolic extracts obtained from seeds of four cowpea cultivars. Their results suggested that cowpea seeds maintained greater antioxidant activity than chickpea. Therefore, cowpea can improve nutrient utilization and provide potential nutraceuticals for human health.

It was reported that phenolic antioxidants were associated with the color pigments of food plants in many studies. In earlier studies, while evaluating variability for AOA in the Regional Southernpea Cooperative Trials, it was observed that high AOA appeared to be associated with dark seed coat color (Warrington et al. 2002). The antioxidant activity from five species of legumes with colored seed coat, pea (*pisum sativum* L), faba bean (*Vicia faba* L.), Lentil (*Lens culinaris* Medik) and everlasting pea (*Lathyrus latifolius* L.) was investigated. The result showed that antioxidant substances in legumes were present mainly in the seed coat (Amarowicz et al. 1996). The inheritance of antioxidant activity (AOA) and its association with seed coat color was investigated in cowpea. There was a very strong relationship between AOA and seed coat color. The inheritance pattern of factors governing AOA was similar to that of factors governing seed coat. Therefore, higher level of activity could be obtained utilizing colored cowpea varieties as parents in crossing for breeding selections (Nzaramba et al. 2005).

Association mapping aims to measure deviation from the random occurrence of alleles in a haplotype among unrelated individuals or unrelated families. However, linkage analysis depends on the co-segregation of a genetic phenotype through a pedigree (March 1999). Linkage analysis in plants has been typically conducted with experimental populations that are derived from a bi-parental cross. Linkage analysis exploits the shared inheritance of functional polymorphisms and

the adjacent markers within families or pedigrees of known ancestry. Although based on the same fundamental principles of genetic recombination as linkage analysis, association mapping examines this shared inheritance for a collection of individuals often with unknown ancestry. As the unrelated ancestries extend thousands of generations, the shared inheritance will only persist for adjacent loci after recombination events occur in many generations (Yu and Buckler 2006). Essentially, association mapping exploits historical and evolutionary recombination at the population level. By exploring population genealogy rather than family pedigree, association mapping offers three advantages over linkage analysis: (1) much higher mapping resolution; (2) greater allele number and broader reference populations; and (3) less research time in establishing an association (Buckler and Thornsberry 2002).

Currently, single nucleotide polymorphism (SNP) markers are the most frequently employed tagging methods for studying genetic diversity (Courtois et al. 2013; Evans et al. 2015) and genome-wide association (Gurung et al. 2014; Kamfwa et al. 2015). The emergence of various types of SNP data processing software promotes the application of SNP markers in the analysis of association. In similar studies, SNP markers have been applied in linkage studies for several crops, including wheat (Jighly et al. 2015; Lopes et al. 2014), corn (Gowda et al. 2015; Zila et al. 2013), rice (Smith et al. 2011; Talukdar et al. 2015), and soybean (Kumar et al. 2014; Zhao et al. 2015). Genotyping by sequencing (GBS) is one of the next-generation sequencing platforms for genome-wide SNP discovery which can be used in association mapping (Elshire et al. 2011; Sonah et al. 2013).

Using four hundred and fifty eight SNPs, Egbadzor et al. (2014) characterize 113 cowpea accessions from Ghana and five from other countries. The results showed that the SNP markers were more efficient in discriminating among the cowpea germplasm than a morphological

strategy. The population structure of 171 samples from the USDA core collection was identified and incorporated into a genome wide association study which supported more than half of the trait-associated regions important in the bi-parental populations (Lucas et al. 2013).

The cowpea core collection was established by the USDA, and the detailed phenotypic data were collected and recorded. However, there was no systematic report on molecular studies and no seed antioxidant content and seed coat colors associated mapping was reported. Hence, the objectives of this study was to analyze the population structure in the USDA Cowpea Core Collection, simultaneously, to provide research information for effective and efficient germplasm conservation, to conduct association mapping for seed antioxidant content and seed coat colors, and to identify SNP markers associated with seed antioxidant content and seed coat colors for use in molecular breeding efforts in cowpea improvement.

Materials and Methods

Germplasm

A panel of 369 cowpea accessions, obtained from the United States Department of Agriculture Germplasm Resources Information Network (USDA-GRIN) was used for association analysis. The accessions were collected from 47 countries/regions from Africa (50.53%), Asia (25.00%), North America (14.63%), Europe (1.86%), Latin-America (6.38%), and Oceania (1.60%) (Supplementary Table S1 and Table 2).

Phenotypic evaluation

(i) Antioxidant activity evaluation: In order to evaluate variability for antioxidant activity (AOA) in USDA cowpea germplasm core set, a total of available 697 cowpea accessions were analyzed for AOA by Nzaramba et al. (2005) at Texas A&M University. There were about 7,000

cowpea accessions in the USDA collection in GRIN and 720 accessions was selected as cowpea core set based on geographic representation and lines suggested by a survey US cowpea breeders during the mid-1990s (Gillaspie et al. 1996; <https://npgsweb.ars-grin.gov/gringlobal/methodaccession.aspx?id1=188027&id2=490425>). Two grams of dry seed from each accession were ground into powder using a Braun KSM2 coffee grinder. The powder was placed in falcon tubes and 15ml of HPLC-grade methanol was added to each tube. The mixture was homogenized for 3 minutes with an IKA® ultra-turrax tissue mixer, and centrifuged for 15 min at 15,000 rpm using a Beckman J2-21 centrifuge. The seed extract was analyzed for AOA using the free radical, 2,2-Diphenyl-1-picrylhydrazyl (DPPH), method (Brand-Williams et al. 1995) and expressed as μg of trolox equivalents/gram of dry weight ($\mu\text{g/gdw}$). The data for AOA in the 697 cowpea accessions was not published in a scientific journal, but they can be viewed and downloaded at the USDA GRIN web sites at <https://npgsweb.ars-grin.gov/gringlobal/methodaccession.aspx?id1=188030&id2=493878> and <https://npgsweb.ars-grin.gov/gringlobal/descriptordetail.aspx?id=188030>. In this research, we used the AOA data in 369 cowpea accessions as the association mapping panel based on available genotypic data (SNP data).

(ii) Cowpea seed color evaluation: All 369 cowpea accessions were planted in the field at the Research and Extension Station, University of Arkansas, on May 26, during 2014 and 2015. Each accession was planted in a single row with 15 feet length and 3-ft row spacing with 4-inch distance between plants. The seed coat color was recorded as white (cream), black, red, brown, blue, green, grey, purple, tan, and variations immediately after harvest. The seed color data were also obtained from the USDA-GRIN website at <https://npgsweb.ars-grin.gov/gringlobal/descriptordetail.aspx?id=188024> where there is seed coat color data for 6325

cowpea accessions.

Phenotypic data of the seed antioxidant content across seed coat color were analyzed using Microsoft (MS) Excel 2013 and JMP Genomics 7 software (SAS Institute, Cary, NC, USA) (Sall et al. 2012) for the average, range, standard deviation (Stdev). The correlation coefficient (r) between the seed antioxidant content and each seed coat color were estimated by the “CORREL” function in Excel. The distributions of the seed antioxidant content crossed seed coat color were drawn using MS Excel and JMP. Analysis of variance (ANOVA) was conducted for the antioxidant content using JMP for all tested cowpea genotypes and grouped by seed coat color as fixed model. The least squared mean (LSM) of each cowpea genotype from JMP was used as the phenotypic data in the association mapping for both seed antioxidant content and seed coat color trait.

DNA extraction and genotyping-by-sequencing (GBS)

Genomic DNA was extracted following the CTAB (cetyl trimethylammonium bromide) method (Hulbert and Bennetzen 1991). DNA concentrations were determined with a NanoDrop 200c spectrophotometer (Thermo SCIENTIFIC, Wilmington, DE). DNA qualities were checked on 1% agarose gels with ethidium bromide gel stain. A DNA library was prepared using the restriction enzyme ApeKI following the GBS protocol described by Elshire *et al.* (2011). DNA sequencing was conducted using GBS (Elshire et al. 2011; Sonah et al. 2013) that was done by HiSeq 2000 in Beijing Genome Institute (BGI). Sequence assembly, mapping and SNP identification were done by BGI using SOAP family software (<http://soap.genomics.org.cn/>). The SOAPaligner/soap2 (<http://soap.genomics.org.cn/>) was used to align the short-read to the cowpea genome reference (cowpea_Genome_0.03.fa) and SOAPsnp v 1.05 was used for SNP calling (Li 2011; Li et al. 2009). The

cowpea_Genome_0.03.fa (6,750 scaffolds or contigs) (<http://harvest-blast.org/>) was kindly provided by Dr. Timothy J. Close, University of California, Riverside. The SNP markers with minor allele frequencies (MAF) lower than 5% were discarded from statistical analysis, where MAF refers to the frequency at which the least common allele occurs in a given population. SNPs with a MAF of 5% or greater were targeted by the HapMap project (The International HapMap Consortium, 2005). The remaining 1,047 high quality SNP markers were used for population structure and Mark-trait association analysis.

Population Genetic Diversity and Association analysis

Summary statistics including major allele frequency, heterozygosity, and polymorphism information content (PIC) were estimated using the software PowerMarker version 3.25 (Liu and Muse 2005). STRUCTURE is a program that uses Bayesian methods to analyze multilocus data in population genetics (Kaeuffer et al. 2007). This study used a hybrid model and an allelic variation occurrence non-correlative model to examine the population structure of cowpea germplasm. The number of the subpopulation (K) was assumed to be between 1 and 12. Thus, each K was run 10 times, the Markov Chain Monte Carlo (MCMC) length of the burn-in period was set to 20000 and the number of MCMC iterations after the burn-in was set to 50000. Delta K was used to screen for appropriate K values. Subsequently, CLUMPP was used to integrate the STRUCTURE-generated results with the “repeat 1000” parameter. In addition, Three different association mapping models were used to analyze the association between the molecular markers and seed coat color, the TASSEL GLM-Q (the general linear model), the mixed linear model (MLM) combining kinship with population structure (Q-matrix) (Bradbury et al. 2007; <http://www.maizegenetics.net/tassel>), MEGA (Tamura et al. 2013) was used to plot the phylogenetic tree of the cowpea germplasm, and JMP® Genomics 7 (Sall et al. 2012) was used

to analyze the variance, standard error and significance.

Results

Seed antioxidant content and seed coat color variations

A total of 369 accessions from the USDA Cowpea Core Collection were analyzed for seed antioxidant content with six seed coat colored, among which 27 cowpea accessions are red seed coat colors; 20 black; 24 tan; 134 brown; 16 blue; and 148 white/cream (Table 1, Fig.1). Overall, the seed antioxidant content of the 369 showed a biased normal distribution (Left top one of Fig. 2), ranged from 43.1 to 1838.3 $\mu\text{g/g DW}$, with an average antioxidant content of 641.6 $\mu\text{g/g DW}$. The large ratio between the maximum and minimum with 42.6, indicating there are large variation of seed antioxidant content among the 369 core collection (Table 1). Based on seed coat color, wider range of seed antioxidant content in each seed coat color were observed with high ratio (max/min) ranged from 6.6 to 39.9 (Table 1), indicating there were large variation of antioxidant content existed in each colored seeds. The distribution of antioxidant content in each colored group also showed a wider range (Fig. 2). The ANOVA and T-test showed that seed antioxidant content differed depending on seed coat colors; the red and black colored seeds contained higher antioxidant content than brown, blue and white/cream seeds. The white/cream seeds had significantly lower content than the other colored seeds (Table 1), indicating the red and black cowpea contain higher seed antioxidant content. The correlation between antioxidant content and seed coat colors were also observed between black/red seed coat and antioxidant content with 0.55; between red seed coat and seed antioxidant content, and black seed coat and antioxidant content, were 0.50 and 0.45, respectively. In this study, the cowpea seed with darker color coats showed higher antioxidant content, which agreed with the results of Nzaramba *et al.*

(2005). Therefore, in the breeding program, breeders should select darker color coated seed when breeding for high antioxidant content.

The 369 cowpea accessions were originally collected from 48 countries (Supplementary Table S1 and Table 2), which mainly came from seven countries: Botswana, United States, India, South Africa, Turkey, Nigeria, Senegal, and Mexico with 85.9% (317/369) of the total tested cowpea accessions. The overall mean of seed antioxidant content in the 369 accessions was 641.6 $\mu\text{g/g}$ DW. The average of seed antioxidant content was higher than the mean with 641.6 $\mu\text{g/g}$ DW from 15 countries: Uruguay, Japan, Thailand, Cuba, Sri Lanka, Argentina, Congo, Mozambique, Botswana, United States, Ghana, Hungary, Afghanistan, Kenya, and Tanzania (Table 2). Among the 15 countries, four countries had the majority number of cowpea accessions, where Botswana has 81; United States 36; Afghanistan 10; and Kenya 9, and other 11 countries had only 1, 2, or 3 accessions. The 36 India cowpea accessions had the averaged antioxidant close to the mean with 649.7 $\mu\text{g/g}$ DW. Other 31 countries had the averaged antioxidant lower than the overall mean.

Population structure and genetic diversity

Analyzing population structure and elucidating its sub-population composition were necessary for the subsequent trait-locus association analysis via linkage disequilibrium in order to avoid false positive results due to the infiltration of sub-populations. The 1047 SNPs were used to perform population structure analysis for the 369 core collections accessions using the STRUCTURE software (Kaeuffer et al. 2007). When $K = 3$, Delta K was maximal with a relatively stable α value. Thus, based on the population structure, the study materials were divided into three clusters. Cluster I was comprised of 116 accessions from which 43

germplasm were collected from Botswana, 11 from India, 21 from the US and six from Australia. Cluster II, included 159 accessions, with 24 from Botswana, 16 from the US, 14 from India, 12 from Nigeria, 11 from Senegal, eight from Turkey, seven from Mexico, and eight from South Africa. Cluster III contained 101 accessions from Botswana, the US, South Africa, Afghanistan, and Cameroon with the collection numbers of 20, 10, 10, 8, and 7, respectively. The largest number of accessions were from Botswana, and were also distributed in three clusters with the most accessions belonging to cluster I. The accessions from the US were evenly distributed in three clusters, with 9.48% ratio in cluster I (11/116), 10.06% ratio in cluster II (16/159), and 9.90% ratio in cluster III (10/101). The accessions from India were only distributed in clusters I and cluster II (Fig. 3B, Supplementary Table S1). Genetic diversity analysis showed germplasm from India, South Africa and the US had broader genetic diversity, with the PIC values being 0.26, 0.25 and 0.24, respectively (Table 4).

MEGA software was used to establish a phylogenetic tree from the 369 accessions core collection based on the clustering results from STRUCTURE, the population was divided into three major branches (Fig. 3C). Data from TASSLE was performed using the principal component analysis (PCA) program which also showed three groups, indicating three population structures in this association panel (Fig. 3A). Thus, the three different analysis software produced similar results.

Association analysis and SNP markers identification

Based on the population structural analysis, an association between 1047 SNPs and the seed antioxidant content was tested using three different models: single marker regression (SMR),

general linear model (GLM-Q) and mixed linear models (MLM-Q+K), respectively. The results indicated that, under different significance levels, three different models detected different numbers of SNPs that were strongly associated with the antioxidant content. Eleven SNPs associated with antioxidant content were identified using a single marker test at LOD ($-\log_{10}(P)$) > 4.0 (We define $-\log_{10}(P)$ as the LOD value, where P is the P value from TASSEL, and used LOD in the text and Tables instead of $-\log_{10}(P)$). Eight and six SNPs associated with antioxidant activity were identified using GLM-Q at LOD > 3.5 , and using MLM-Q+K at LOD > 2.0 , respectively (Supplementary Table S3). Two SNPs, Scaffold7139_14363 and scaffold29110_4657, were strongly associated with the seed antioxidant content across three models (Table 3).

The association analysis between 1047 SNPs and seed coat colors was also conducted using the three models. For the red coat color, nine associated loci were identified using MLM-Q+K at LOD > 2.0 ; thirteen associated loci were identified using GLM-Q at LOD > 3.0 ; twenty-one associated loci were identified using single marker regression at LOD > 3.5 (Supplementary Table S4). The SNPs, C35063613_1497, scaffold88685_3135 and scaffold81493_886 were identified crossing three models (Table 3). For black coat color, the detected loci were twenty one, nineteen and 1, when using single marker test at LOD > 4.5 , GLM-Q at $-\log(P) > 3.0$, and MLM-Q+K at LOD > 2.0 , respectively (Supplementary Table S5). The SNP, C35063613_1497, showed association with the black coat color across three models (Table 3). The numbers of loci associated with black and red coat color were twenty-one, eight and twelve, respectively by using single marker regression at LOD > 5.0 , MLM-Q+K at LOD > 2.0 and GLM-Q at LOD > 3.0 (Supplementary Table S6). Three SNPs, C35063613_1497, scaffold81493_886 and scaffold84620_6785, were identified using three different models (Table 3). Regarding seed coat

colors, 17 associated loci were identified using single marker regression at $LOD > 7.0$. Eight associated loci were identified using MLM-Q+K at $LOD > 2.0$; five associated loci were identified using GLM-Q at $LOD > 3.0$ (Supplementary Table S7); Two SNPs, scaffold81493_886 and scaffold40268_5600, were associated with seed coat colors with in three different models (Table 3). The SNP, C35063613_1497, was pleiotropic and associated with red and black coat color. It was also associated with coat colors when single marker regression was used. The SNP, Scaffold42008_191, related to antioxidant traits and also associated with black and red coat color (Supplementary Table S2, S3, S4, and Table 2). Similarly, the SNP, C35082838_2258, related to antioxidant traits was associated with coat color as well.

Discussion

Population structure analysis of core collections is valuable for breeding

This study analyzed the population structure of accessions from the USDA cowpea core collection. Based on population structure analysis by STRUCTURE, TASSEL, and MEGA software, 369 accessions were divided into three clusters. Accessions from the same country were likely clustered together, however, some accessions from the same country were distributed into different groups. Genetic diversity analysis showed germplasm from India, South Africa, and the US had broader genetic diversity. Egbadzor *et al.* (2014) assessed the genetic structure and diversity in 113 cowpea accessions, 48 of which are very diverse morphologically. The core collection is a small-scale sampling, but the collection is strongly representative of total cowpea genetic diversity. Therefore, the core collection is easier to utilize in evaluating cowpea genetic diversity, thus also has great practical use in breeding. Before introducing a suitable germplasm for breeding, it is necessary to clarify the relationship to existing germplasms before it can be

used efficiently. The results of population structure analysis can provide detailed information on genetic similarity of germplasm for breeding. This study provides a theoretical basis for the utilization of the core collection in cowpea breeding. To date, no reports on population structure which could facilitate cowpea breeding have been found in the literature, therefore, we hope this study on the core collection will enhance efficient utilization of the cowpea germplasm.

SNP markers for both antioxidant content and seed coat color

In this study, 369 cowpea accessions showed a varying levels of antioxidant content. There was correlation between antioxidant content and seed coat colors. The cowpea seed with darker color coats had higher antioxidant content. The top ten high antioxidant content accessions ranged from 1715.79 to 1838.31 $\mu\text{g/g DW}$, the seed coat colors of these ten accessions had darker colors, and five of them were red and black. Although the genetics of cowpea seed antioxidant content and seed coat colors are not very clear, Nzaramba et al. 2005 reported four advanced cowpea lines and their F1, F2, BC1, BC2 populations to elucidate the antioxidant content in cowpea seed coats by estimating heritability and number of genes involved, and investigating the relationship between antioxidant content and cowpea seed coat colors. The result also showed there is a very strong relationship between antioxidant content and seed coat color. Pigmented varieties of cowpea possess favorable factors that enhance antioxidant content. Factors governing high antioxidant activity in cowpea seeds appear to be the same factors responsible for seed coat color, with apparent pleiotropic effects. The minimum number of genes responsible for antioxidant content was estimated at five (Nzaramba et al. 2005). In this study, the similar results were observed that significant correlations between the seed antioxidant content and seed black color (correlation coefficient = 0.45), between seed antioxidant content and seed red coat color ($r = 0.50$). Several SNP markers located at different cowpea genome

contigs were identified to be strongly associated with cowpea seed antioxidant content and seed coat color, indicating both seed antioxidant content and seed coat color are quantitative traits. Although different SNP markers were identified to be associated with seed antioxidant content and seed coat colors with different models, the two SNP markers, Scaffold7139_14363 and scaffold29110_4657 were strongly associated with antioxidant content, and three SNP markers, C35063613_1497, scaffold81493_886, and scaffold84620_6785 were strongly associated with seed coat color across three models; and two SNP marker, Scaffold42008_191 and C35082838_2258 were associated with both seed antioxidant content and seed coat color, simultaneously. These results indicated these SNP markers identified in this study could potentially be used in marker-assisted breeding to accelerate genetic improvement of cowpea for higher seed antioxidant content.

Utilization of cowpea accessions with high seed antioxidant content

From the 369 cowpea accessions in this research, black and red coat colored cowpea accession had higher seed antioxidant with a high correlation, and the result is similar to those from Nzaramba et al. (2005a, 2005b). The USDA cowpea accessions with red and black seed coat color can be used as parents in cowpea breeding program to release new cowpea cultivars with high seed antioxidant content. Besides the red and black colored cowpea accessions, some accessions with other colored had also high antioxidant, such as PI 186386 with tan coat color and PI 582818 with white/cream coat color (Supplementary Table S1). The top highest antioxidant cowpea among the 369 accessions in this research are PI293587, PI339597, PI194207, PI211642, PI194211, PI186386, PI582818, PI209971, PI582881, and PI582821 (Supplementary Table S1), which will be excellent resources in cowpea breeding for high seed

antioxidant content in various seed coat colors, and efficient materials on future germplasm preservation, germplasm utilization and discovery of favorable genes.

Acknowledgments

This work was supported by USDA-ARS GRIN GERMPLASM EVALUATION PROPOSAL for National Plant Germplasm System (NPGS) by Crop Germplasm Committee (CGC) with Project Number: 6046-21000-011-15. Cowpea germplasm accessions were provide by USDA-ARS at GRIFFIN, GA Station. The cowpea_Genome_0.03.fa (6,750 scaffolds or contigs) (<http://harvest-blast.org/>) was kindly provided by Dr. Timothy J. Close at University of California Riverside, CA, USA.

Primary Sources

Secondary Sources

Uncategorized References

- Abraído-Lanza, A. F., Armbrister, A. N., Flórez, K. R. and Aguirre, A. N. 2006. Toward a theory-driven model of acculturation in public health research. *American Journal of Public Health* **96**(8):1342-1346.
- Amarowicz, D. R., Wanasundara, U. N. , Karamać, M, Shahidi, F. 1996. Antioxidant activity of ethanolic extract of mustard seed. *Nahrung* **40**(5) : 261–263.
- Anderson, J. W., Smith, B. M. and Washnock, C. S. 1999. Cardiovascular and renal benefits of dry bean and soybean intake. *The American Journal of Clinical Nutrition* **70**(3):464s-474s.
- Bazzano, L. A., He, J., Ogden, L. G., Loria, C., Vupputuri, S., Myers, L. and Whelton, P. K. 2001. Legume consumption and risk of coronary heart disease in US men and women: NHANES I Epidemiologic Follow-up Study. *Archives of Internal Medicine* **161**(21):2573-2578.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633–2635
- Brand-Williams, W., Cuvelier, M. E., and Berset, C. 1995. Use of a free radical method to evaluate antioxidant activity. *LWT-Food Science and Technology* **28**(1):25-30.
- Brown, J. K., Byers, T., Doyle, C., Courneya, K. S., Demark - Wahnefried, W., Kushi, L. H., McTiernan, A., Rock, C. L., Aziz, N. and Bloch, A. S. 2003. Nutrition and physical activity during and after cancer treatment: an American Cancer Society guide for informed choices. *CA: a cancer journal for clinicians* **53**(5):268-291.
- Buckler, E. S. and Thornsberry, J. M. 2002. Plant molecular diversity and applications to genomics. *Current opinion in plant biology* **5**(2):107-111.
- Courtois, B., Audebert, A., Dardou, A., Roques, S., Ghneim-Herrera, T., Droc, G., Frouin, J., Rouan, L., Gozé, E. and Kilian, A. 2013. Genome-wide association mapping of root traits in a japonica rice panel. *PLoS ONE* **8**(11): e78037. doi:10.1371/journal.pone.0078037
- Egbadzor, K. F., Ofori, K., Yeboah, M., Aboagye, L. M., Opoku-Agyeman, M. O., Danquah, E. Y. and Offei, S. K. 2014. Diversity in 113 cowpea [*Vigna unguiculata* (L) Walp] accessions assessed with 458 SNP markers. *SpringerPlus* **3**(1):541.
- Ehlers, J. D. and Hall, A. E. 1997. Cowpea (*Vigna unguiculata* L Walp). *Field Crops Res* **53**:187–204
- Ehlers, J. D., Fery, R. L. and Hall, A. E. 2007. Cowpea breeding in the USA: new varieties and improved germplasm. In: IITA Conference Proceedings
- Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S. and Sharon E. M. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **6**(5): e19379. doi:10.1371/journal.pone.0019379
- Evans, J., Crisovan, E., Barry, K., Daum, C., Jenkins, J., Kunde - Ramamoorthy, G., Nandety, A., Ngan, C. Y., Vaillancourt, B. and Wei, C. L. 2015. Diversity and Population Structure of Northern Switchgrass as Revealed Through Exome Capture Sequencing. *The Plant Journal* **84**(4):800–815
- Frota, K. d. M. G., Soares, R. A. M. and Arêas, J. A. G. 2008. Composição química do feijão caupi (*Vigna unguiculata* L. Walp), cultivar BRS-Milênio. *Ciência e Tecnologia de Alimentos* **28**(2):470-476.

- Gillaspie, A. G., Chambliss, O. L., Fery, R. L., Hall, A. E., Miller, Jr., J. C., and Morelock, T. E. 1996. A core subset established for the USDA cowpea (*Vigna unguiculata* (L.)) germplasm collection. *HortScience* **31**:762
- Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M. S., Bright, J. M., Beyene, Y. and Prasanna, B. M. 2015. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theoretical and Applied Genetics* **128**(10):1957-1968.
- Gurung, S., Mamidi, S., Bonman, J. M., Xiong, M., Brown-Guedira, G. and Adhikari, T. B. 2014. Genome-wide association study reveals novel quantitative trait Loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS ONE* **9**(9): e108179. doi:10.1371/journal.pone.0108179
- Hulbert, S. H. and Bennetzen, J. L. 1991. Recombination at the Rp1 locus of maize. *Molecular and General Genetics MGG* **226**(3):377-382.
- Iqbal, A., Khalil, I. A., Ateeq, N. and Khan, M. S. 2006. Nutritional quality of important food legumes. *Food Chemistry* **97**(2):331-335.
- Jighly, A., Oyiga, B. C., Makdis, F., Nazari, K., Youssef, O., Tadesse, W., Abdalla, O. and Ogbonnaya, F. C. 2015. Genome-wide DArT and SNP scan for QTL associated with resistance to stripe rust (*Puccinia striiformis* f. sp. tritici) in elite ICARDA wheat (*Triticum aestivum* L.) germplasm. *Theoretical and Applied Genetics* **128**(7): 1277-1295.
- Kadam, S., Salunkhe, D. and Maga, J. A. 1985. Nutritional composition, processing, and utilization of horse gram and moth bean. *Critical Reviews in Food Science & Nutrition* **22**(1):1-26.
- Kaeuffer, R., Réale, D., Coltman, D. and Pontier, D. 2007. Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity* **99**(4):374-380.
- Kamfwa, K., Cichy, K. A. and Kelly, J. D. 2015. Genome-wide association analysis of symbiotic nitrogen fixation in common bean. *Theoretical and Applied Genetics* **128**(10):1999-2017.
- Liu, K. and Muse, S.V. 2005. PowerMarker: an integrated analysis environment for geneticmaker analysis. *Oxford Journal Science & Mathematics Bioinformatics* **21**(9):2128-2129
- Kumar, B., Talukdar, A., Bala, I., Verma, K., Lal, S. K., Sapra, R. L., Namita, B., Chander, S. and Tiwari, R. 2014. Population structure and association mapping studies for important agronomic traits in soybean. *Journal of genetics* **93**(3):775-784.
- Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993
- Li, R., Yu, C., Li, Y., Lam, T., Yiu, S., Kristiansen, K. and Wnag, J. 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**(15):1966–1967
- Lopes, M., Dreisigacker, S., Peña, R., Sukumaran, S. and Reynolds, M. 2014. Genetic characterization of the Wheat Association Mapping Initiative (WAMI) panel for dissection of complex traits in spring wheat. *Theoretical and Applied Genetics* **128**(3):453-464.
- Lucas, M. R., Huynh, B. L., da Silva Vinholes, P., Cisse, N., Drabo, I., Ehlers, J. D., Roberts, P. A. and Close, T. J. 2013. Association studies and legume synteny reveal haplotypes determining seed size in *Vigna unguiculata*. *Frontiers in plant science* doi: 10.3389/fpls.2013.00095
- March, R. E. 1999. Gene mapping by linkage and association analysis. *Molecular biotechnology* **13**(2):113-122.
- Muchero, W., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., Hearne, S.,

- Cisse, N., Fatokun, C. and Ehlers, J. D. 2009. A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. *Proceedings of the national academy of sciences* **106**(43):18159-18164.
- Nzaramba, M. N., Hale, A. L., Scheuring, D. C. and Miller, J. C. 2005. Inheritance of Antioxidant Activity and its Association with Seedcoat Color in Cowpea. *J. AMER. SOC. HORT. Sci.* **130**(3):386-391
- Nzaramba, M. N. , Scheuring D. C., and Miller J. C. 2005. Variability for antioxidant activity in the U.S. cowpea [*Vigna unguiculata* (L) WALP.] cowpeacore collection. *HORT. Sci.* **40**(3):881
- Oplinger, E., Oelke, E., Kaminski, A., Putnam, D., Teynor, T., Doll, J., Kelling, K., Durgan, B. and Noetzel, D. 1991. *Crambe: alternative field crops manual*. St Paul: University of Minnesota Available from: September:3.
- Rafieian-Kopaei, M., Baradaran, A. and Rafieian, M. 2013. Oxidative stress and the paradoxical effects of antioxidants. *Journal of Research in Medical Sciences* **18**(7):628.
- Rahman, K. 2007. Studies on free radicals, antioxidants, and co-factors. *Clinical interventions in aging* **2**(2):219.
- Sall, J., Lehman, A., Stephens, M. L. and Creighton, L. 2012. *JMP start statistics: a guide to statistics and data analysis using JMP*. SAS Institute.
- Siddhuraju, P. and Becker, K. 2007. The antioxidant and free radical scavenging activities of processed cowpea (*Vigna unguiculata* (L.) Walp.) seed extracts. *Food Chemistry* **101**(1):10-19.
- Singh, B. B., Ajeigbe, H. A., Tarawali, S. A., Fernandez-Rivera, S. and Abubakar, M. 2003. Improving the production and utilization of cowpea as food and fodder. *Field Crops Res* **84**(1-2):169-177
- Smith, E. N., Koller, D. L., Panganiban, C., Szelinger, S., Zhang, P., Badner, J. A., Barrett, T. B., Berrettini, W. H., Bloss, C. S. and Byerley, W. 2011. Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes. *PLoS Genet* **7**(6):e1002134.
- Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., Éric, N., Jérôme, L., Stéphane, L., Martine, J. and François B. 2013. An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. *PLoS ONE* **8**(1): e54603. doi:10.1371/journal.pone.0054603
- Summerfield, R. J., Fate, J. S., Roberts, E. H. and Wien, H. C. 1985. *Physiology of Cowpea*. In *Cowpea Research, Production and Utilization* Singh and Rachie Eds John Wiley and Sons, New York. pp. 66 -101
- Talukdar, P., Douglas, A., Price, A. H. and Norton, G. J. 2015. Biallelic and Genome Wide Association Mapping of Germanium Tolerant Loci in Rice (*Oryza sativa* L.). *PloS one* **10**(9):e0137577.
- Tamura, K., Stecher, G., Peterson, D., Ailipski, A. and Kumar, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**:2725-2729
- Tan, H., Tie, M., Luo, Q., Zhu, Y., Lai, J. and Li, H. 2012. A review of molecular makers applied in Cowpea (*Vigna unguiculata* L. Walp.) Breeding. *J Life Sci* **6**(2012):1190-1199.
- Valko, M., Leibfritz, D., Moncol, J., Cronin, M. T., Mazur, M. and Telser, J. 2007. Free radicals and antioxidants in normal physiological functions and human disease. *The international journal of biochemistry & cell biology* **39**(1):44-84.
- Warrington, R. T., Hale, A. L., Scheuring, D. C., Whitaker, D. W., Blessington, T. and Miller, J. C. 2002. Variability for antioxidant activity in cowpea [*Vigna unguiculata* (L) Walp.] as influenced by genotype and postharvest rehydration. *HortScience* **37**:738. (Abstr.)

- Winham, D. M., Hutchins, A. M. and Johnston, C. S. 2007. Pinto bean consumption reduces biomarkers for heart disease risk. *Journal of the American College of Nutrition* **26**(3):243-249.
- Yu, J. M. and Buckler E. S. 2006. Genetic association mapping and genome organization of maize 2006. *Current Opinion in Biotechnology* **17**:155–160
- Zia-Ul-Haq, M., Ahmad, S., Amarowicz, R. and De Feo, V. 2013. Antioxidant activity of the extracts of some cowpea (*Vigna unguiculata* (L) Walp.) cultivars commonly consumed in Pakistan. *Molecules* **18**(2):2005-2017.
- Zila, C. T., Samayoa, L. F., Santiago, R., Butrón, A. and Holland, J. B. 2013. A genome-wide association study reveals genes associated with *Fusarium* ear rot resistance in a maize core diversity panel. *G3: Genes| Genomes| Genetics* **3**(11):2095-2104.
- Zhao X., Han Y., Li Y., Liu D., Sun M., Zhao Y., Lv C., Li D., Yang Z., Huang L., Teng W., Qiu L., Zheng H., Li W. 2015. Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. *Plant J.* **82**(2):245-55.

Legends of Tables

Table 1. Seed antioxidant content information crossed various seed coat color among 369 cowpea USDA accessions.

Table 2. Country distribution of cowpea amount and the average seed antioxidant content among the 369 core collection.

Table 3. SNP markers associated with seed antioxidant content and seed coat colors using three statistical models, single marker regression (SMR), general linear model GLM (Q), and mixture linear model MLM (Q+K).

Table 4. Table. 4 Allelic analysis of cowpea core collections from 11 geographic regions and genotyped with 1047 SNP markers.

Legends of Figures

Fig. 1 The distribution of cowpea accessions based on seed coat color.

Fig. 2 The distributions of antioxidant content in cowpea seeds based on seed coat color: all color, black, blue, brown, red, tan, and white, respectively. X-axis signifies number of accessions, and y-axis signifies seed antioxidant content ($\mu\text{g/g DW}$).

Fig. 3 Model-based populations in the cowpea seed antioxidant association panel of 369 cowpea accessions of core collection: (A) Two dimension distribution analyzed by principal component analysis (PCA), (B) Classification of three populations using STRUCTURE 2.3.4, and (C) phylogenetic tree constructed by neighbor-joining (NJ) of genetic distance by MEGA 6. The

distribution of the accessions to different populations is indicated by the color code (Q1: red, Q2: green, Q3: blue), consistent in the figure A, B and C.

Supporting Information

Supplementary Table S1, S2, S3, S4, S5, and S6 - Qin_Supporting_information_Tables.xlsx

For Review Only