Identification of seed coat pattern trait QTL and a description of seed coat development in cowpea (*Vigna unguiculata* [L.] Walp.)

Ira A. Herniter[1], Ryan Lo[1], María Muñoz-Amatraín[1], Sassoum Lo[1], Yi-Ning Guo[1], Bao-Lam, Huynh[1], Mitchell Lucas[1,2], Zhenyu Jia[1], Philip A. Roberts[3], Timothy J. Close[1]

[1]Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92521, USA

[2]Syngenta Seeds Inc., 10290 Greenway Road, Naples, Florida 34114, USA

[3]Department of Nematology, University of California, Riverside, Riverside, CA 92521, USA

Genetic control of cowpea seed coat traits

Key words: *Vigna unguiculata*, seed coat pattern, development, QTL analysis, SNP genotyping

Corresponding author:

Ira A. Herniter

4161 Batchelor Hall

900 University Avenue, Riverside, CA, 92507

ihern014@ucr.edu

ABSTRACT

The appearance of the seed is an important aspect of consumer preferences for cowpea (*Vigna unguiculata* [L.] Walp). Seed coat pattern in cowpea has been a subject of study for over a century. This study makes use of newly available resources, including mapping populations, a reference genome, and a high-density single nucleotide polymorphism genotyping platform, to map various seed coat pattern traits to three loci, concurrent with the Color Factor (C), Watson (W), and Holstein (H) factors identified previously. Several gene models encoding proteins involved in the regulating the anthocyanin biosynthesis pathway have been identified as candidate genes, including a basic helix-loop-helix protein (Vigun07g110700), a WD repeat gene (Vigun09g139900), and an E3 ubiquitin ligase gene (Vigun10g163900). Additionally, a model of seed coat development, involving six distinct stages, has been described which explains some of the observed seed coat pattern phenotypes.

INTRODUCTION

Cowpea (*Vigna unguiculata* [L.] Walp.) is a diploid (2n = 22) warm season legume which is primarily grown and serves as a major source of protein and calories in sub-Saharan Africa. Further production occurs in the Mediterranean Basin, southeast Asia, Brazil, and the United States. The most common form of consumption is as a dry grain. The seeds are used either whole or ground into flour (Singh 2014; Tijjani, Nabinta, and Muntaka 2015). Most cowpea production is by smallholder farmers in marginal conditions in sub-Saharan Africa, often as an intercrop with maize (Ehlers and Hall 1997). Due to its high adaptability to both heat and drought and its association with nitrogen fixing bacteria, cowpea is a versatile crop (Ehlers and Hall 1997). Over

eight million metric tonnes of dry cowpeas were produced worldwide in 2013 (FAOSTAT 2014).

Seed coat pattern is an important consumer-related quality in cowpea. Previous research has shown that consumers make qualitative decisions about the acceptability, quality, and taste of a product based on appearance (Jaeger *et al.* 2018; Kostyla, Clydesdale, and McDaniel 1978). As such, determination of the genetic control of seed coat pattern traits would be advantageous for breeding programs developing new varieties with a likelihood of market viability. Cowpea displays a variety of patterns, including varied eye shapes, Holstein, Watson, and Full Coat pigmentation, among others (Figure 1). Each region has its own preferred varieties, valuing certain color and pattern traits above others for determining quality and use of the seeds. For example, in West Africa, consumers pay a premium for seeds exhibiting specific characteristics specific to the locality, such as lack of color for use as flour or solid brown for use as whole beans (Langyintuo *et al.* 2003; Mishili *et al.* 2009), while in the United States, consumers prefer varieties with tight black eyes, so much so that the crop is often referred to as "black-eyed peas," (Fery 1985). It should be noted that color and pattern are two sets of distinct traits controlled by independent loci, but which are often confused, even by breeders. Color is defined as the pigmentation present on the seed coat while pattern is defined by how the pigment is distributed. The black-eye pattern preferred by consumers in the United States is often referred to in the literature as a "white seed" (Faye et al. 2004; Langyintuo et al. 2004). It would be more correct genetically to refer to such seeds as black with the pigmentation restricted to an eye pattern.

Seed coat traits in cowpea have been studied in detail since the early 20th century, when Spillman (1911) and Harland (1919, 1920), reviewed by Fery (1980), identified genetic factors controlling traits such as seed coat color and pattern. In a series of F2 populations Spillman (1911) and

Harland (1919) identified genetic factors responsible for color expression, including "Color factor" (C), "Watson" (W), "Holstein-1" (H-1), and "Holstein-2" (H-2).

A genotyping array for 51,128 single nucleotide polymorphisms (SNP) was recently developed for cowpea (Muñoz-Amatriaín *et al.* 2017), which offers opportunities to improve the precision of genetic mapping of traits. Numerous biparental populations have been used to map major quantitative trait loci (QTL) for various traits, including heat tolerance (Lucas *et al.* 2013), leaf shape (Pottorff *et al*. 2012), and black seed coat color (Herniter *et al.* 2018) and to develop consensus genetic maps of cowpea (Lucas, Diop, and Wanamaker 2011; Muchero *et al.* 2009; Muñoz-Amatriaín *et al.* 2017). In addition, new populations have been developed for higher-resolution mapping including a minicore containing 367 diverse accessions (Muñoz-Amatriaín, unpublished) and an 8-Parent Multi-parent Advanced Generation InterCross (MAGIC) population containing 305 lines (Huynh *et al.* 2018). A reference genome sequence of cowpea has been recently produced (available in Phytozome, phytozome.jgi.doe). Here, we make use of genetic and genomic resources available for cowpea to map a variety of seed coat pattern traits, determine candidate genes, and develop a model for genetic control of seed coat pattern. Additionally, we posit a developmental pattern for the cowpea seed coat which explains some of the observed phenotypic variation.

MATERIALS AND METHODS

Plant Materials

Eleven populations were used for mapping: a minicore representing worldwide diversity of cultivated cowpea containing 367 accessions (Muñoz-Amatriaín, *et al.* unpublished), an eight parent Multiparent Advanced Generation InterCross (MAGIC) population containing 305 lines

(Huynh *et al.* 2018), four biparental recombinant inbred line (RIL) populations, and five F2 populations. One biparental population consists of 93 RILs developed at the University of California, Riverside (UCR), derived from a cross between California Blackeye 27 (CB27), which has a medium-sized black eye seed coat, and IT82E-18, also known as "Big Buff" (BB), which has a solid brown coat (Muchero *et al.* 2009). The second biparental RIL population consists of 84 RILs developed at UCR derived from a cross between CB27 and IT97K-556-6 (556), which has a solid brown coat (Huynh et al. 2015). The third biparental RIL population consists of 105 RILs developed at UCR derived from a cross between California Blackeye 46 (CB46), which has a medium sized black eye pattern, and IT93K-503-1 (503), which has a medium sized brown eye pattern (Pottorff et al. 2014). The fourth biparental RIL population consists of 79 RILs developed at UCR and in Nigeria derived from a cross between 524B, which has a medium sized black eye pattern, and IT84S-2049 (2049), which has a medium sized brown eye pattern (Menéndez, Hall, and Gepts 1997). The F2 populations were developed at UCR. Two of the F2 populations, consisting of 176 and 132 individuals, resulted from independent crosses between CB27 and Bambey 21, which has no coloration in the seed coat. One of the F2 populations, consisting of 143 individuals, resulted from a cross between Bambey 21 and California Blackeye 50 (CB50) which has a medium-sized black eye seed coat. Two of the F2 populations, consisting of 175 and 119 individuals, resulted from independent crosses between Tvu15426, which has a solid purple coat, and MAGIC014, a line developed as part of the MAGIC population but not included in the final population, which has a black Watson patterned coat.

To describe the seed coat development four accessions were examined: CB27, MAGIC059, Sanzi, and Sasaque. CB27, Sanzi, and Sasaque are described above. MAGIC059 has the starry night pattern in black and purple and is one of the lines included in the MAGIC population.

SNP genotyping and data curation

DNA was extracted from young leaf tissue using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germany). 51,128 SNPs were assayed in each sample using the Illumina Cowpea iSelect Consortium Array (Illumina Inc., California, USA) (Muñoz-Amatriaín *et al.* 2017). Genotyping was performed at the University of Southern California Molecular Genomics Core facility (Los Angeles, California, USA). The same custom cluster file as in Muñoz-Amatriaín *et al.* (2017) was used for SNP calling.

For the minicore collection, a total of 47,338 SNPs were used after removing those with high levels of missing data and/or heterozygous calls (>20%), with minor allele frequencies <0.05, and those not assigned to a chromosome. SNPs were ordered based on their physical position in cowpea pseudomolecules, which is available from Phytozome (phytozome.jgi.doe.gov). For the MAGIC population, SNP data and a genetic map were available from Huynh *et al.* (2018). The map included 32,130 SNPs in 1,568 genetic bins (Huynh *et al.* 2018). For the biparental RIL populations, SNP data and genetic maps for the CB27 by BB and the CB46 by 503 populations were available from Muñoz-Amatriaín *et al* (2017), SNP data and a genetic map were available for the 524B by 2049 population from (Huynh *et al.* 2016). The CB27 by 556 genetic map was created using MSTMap (Wu *et al.* 2008). The CB27 by BB genetic map included 16,566 polymorphic SNPs in 977 genetic bins (Muñoz-Amatriaín *et al.* 2017); the CB27 by 556 genetic map contained 16,284 SNPs in 2604 bins (Table S1); the CB46 by 503 genetic map contained 16,578 SNPs in 683 bins (Muñoz-Amatriaín *et al.* 2017); the 524B by 2049 genetic map

contained 14,202 SNPs in 933 bins (Huynh *et al.* 2016). For each F2 population, SNPs were filtered to remove non-polymorphic loci between the respective parents. As a result, the number of markers used for each population is as follows: the two CB27 by Bambey 21 populations, 8,551 SNPs; the Bambey 21 by CB50 population, 8,629 SNPs; the two Tvu15426 by MAGIC014 populations, 20,010 SNPs.

Seed coat phenotyping

Phenotype data for seed coat traits were collected by visual examination of the seeds. The phenotypic classes consisted of Color Factor, Eye 1, Eye 2, Blue-grey Ring, Holstein, Watson, and Full Coat. Examples of each phenotype can be seen in Figure 1. The two types of eye patterns observed segregated separately. Eye 1 consists of a tight eye in the shape of two wings. No pigment is observed outside the edge of the eye. Eye 2 consists of a loose eye in the shape of a teardrop with spots of color outside the eye on the wider side. Seeds with a paler brown color are often difficult to distinguish between the Eye 2 and Watson patterns. The minicore was scored for color factor, eye, blue-grey ring, and Full Coat patterns. The MAGIC population was scored for eye, Holstein, Watson, and Full Coat patterns. The biparental RIL populations were scored for Eye 1, Eye 2, Holstein, Watson, and Full Coat patterns. The CB27 by Bambey 21 and Bambey 21 by CB50 F2 populations were scored for color factor. The Tvu15426 by MAGIC014 F2 populations were scored for the Watson and Full coat patterns. The scores were assigned as "1" indicating presence of the trait and a "0" indicating absence. Pattern phenotypes are mutually exclusive, excepting blue-grey ring, which only appears on lines expressing the Eye 2 pattern.

Trait mapping

Trait mapping was achieved through different methods for each type of population. In the minicore population, genome wide association (GWA) was performed using the weighted mixed-linear model function (Zhang *et al.* 2010) in TASSEL v.5 (maizegenetics.net/tassel) with a principal component analysis (three components) accounting for population structure in the dataset. The $-\log_{10}(p)$ values were plotted against the physical coordinates of the SNPs, available from Phytozome (phytozome.jgi.doe.org). A Bonferroni correction was applied to correct for multiple testing the significance cut-off set at α/n, where α is 0.05 and n is the number of tested markers (46,055). This resulted in a cut-off of p = 1.08566E-06 [$-\log_{10}(p) = 5.92$].

In the MAGIC population, the R package "mpMap" (Huang and George 2011) was used as described by Huynh *et al.* (2018). The significance cutoff values were determined through 1000 permutations, resulting in a threshold of p = 8.096679e-05 [$-\log_{10}(p) = 4.09$]. Due to the high number of markers in the genotype data, imputed markers spaced at 1 cM intervals were used. The genetic map and descent data of the population was available from Huynh *et al.* (2018).

In the biparental RIL populations the R packages "qtl" (Broman et al. 2003) and "snow" (Tierney *et al.* 2015) were used as in Herniter *et al.* (2018). In short, probability values were assigned to each SNP using a Haley-Knott regression, tested for significance with 1000 permutations, and marker effects were determined using a hidden Markov model.

For the F2 populations, the genotype calls of each population were filtered to leave only the markers known to be polymorphic between the parents which were sorted based on physical position, available from Phytozome (phytozome.jgi.doe.gov). Each population's genotype was then examined visually in Microsoft Excel for areas where the recessive bulk was homozygous, and the dominant bulk was heterozygous. Duplicated populations were examined in conjunction.

Determining haplotype blocks

In each biparental population lines were grouped by phenotype and all 51,128 assayed SNPs from the iSelect SNP Genotyping Array were examined visually in Microsoft Excel for regions of identity within phenotypic groups. SNPs removed from analysis due to high levels of missing data were utilized as presence/absence variations which segregate similar to other variations. Once regions were established, the overlap between populations was examined to determine the minimal area where all four biparental populations overlapped. This area was determined to be the minimal haplotype block and examined for candidate genes.

Determining candidate genes

In each area of greatest interest, genes were examined. Relative expression level data available from the Legume Information System (legumeinfo.org) was examined for increased expression in the developing seeds. Further, genes encoding proteins known to be involved in the anthocyanin production pathway were prioritized. This led to the determination of a single candidate gene in each locus.

Determining allele series

Some dominance relationships were determined by examining the segregation ratios in the F2 populations. In addition, crosses were made between CB27 and three lines from the CB27 by BB population. CB27/BB-090 has a Watson pattern. CB27/BB-113 has a Holstein pattern. CB27/BB-074 has a Full Coat pattern. The seeds collected from the F1 plants were visually examined for pattern phenotype in the maternal (F1) seed coats.

Collecting development data

The four accessions for which pattern development was recorded (CB27, MAGIC059, Sanzi, and Sasaque) were grown in a greenhouse on the campus of the University of California, Riverside (Riverside, California) at a constant temperature of about 32°C from March through May 2018. For each accession, three plants were used. Upon flowering, each flower was tagged with the date it opened. The flowers were permitted to self-fertilize. For each day after the flower opened, beginning on the second day, on each of the three test plants a pod was collected until no more green pods were observed.

The seeds from each collected pod were photographed using a Canon EOS Rebel T6i at a 90º angle under good lighting conditions. The length of the most advanced seed within the pod was measured using ImageJ (imagej.nih.gov). A developmental scale was developed (see Results) and each photo was scored on the scale from 0 to 5.

Data and Material Availability

Parents of the F2 populations are available upon request. Genotype data for the minicore can be found in Munoz-Amatriain *et al* (unpublished). Genotype data for the biparental RIL populations of CB27 by BB and CB46 by 503 can be found in the supporting information Data S3 of Muñoz-Amatriaín, *et al.* (2017). Genotype data for the biparental 524B by 2049 population can be found in (Santos et al. 2018). Genotype data for the MAGIC population can be found in the supporting information Data S1 of *Huynh et al.* (2018). Genotype and genetic map data for the CB27 by 556 population can be found in Table S1. The full SNP array data for the CB27 by BB, CB27 by 556, CB46 by 503, and 524B by 2049 populations can be found in Tables S2, S3, S4, and S5, respectively. Phenotype data for the minicore can be found in Table S6. Phenotype data for the MAGIC population can be found in Table S7. Phenotype data for the CB27 by BB, CB27 by 556, CB46 by 503, and 524B by 2049 populations can be found in Tables S8, S9, S10, and S11,

respectively. Genotypic data for the bulked DNA from the CB27 by Bambey 21 F2 populations can be found in Table S12. Genotypic data for the bulked DNA from the Bambey 21 by CB50 F2 population can be found in Table S13. Genotypic data for the bulked DNA from the Tvu15426 by MAGIC014 F2 populations can be found in Table S14.

RESULTS

Segregation Ratios

MAGIC

The MAGIC population segregated for the following traits: eye, blue-grey ring, Holstein, and Full Coat patterns. Due to the nature of the MAGIC population as having eight parents, the expected ratio for each trait varied based on the ratio of parents expressing the trait. Eye, Holstein, and Full Coat are mutually exclusive patterns. In the population, 138 lines expressed the eye pattern, 16 lines expressed the Holstein pattern, and 147 lines expressed the Full Coat pattern. Phenotype data was not available for four lines. This indicates that the Holstein pattern is an epistatic trait, necessitating the presence of at least two loci controlling pattern in the MAGIC population. No lines in the MAGIC population expressed the Watson pattern. For a possible explanation see Discussion.

For the eye pattern, 4 parents expressed the trait and 4 did not. In the population, 157 lines expressed the trait while 140 lines did not. The chi-square value for a 1:1 ratio was 0.97 with a p-value of 0.32. This indicates a likely single locus for the eye trait in the MAGIC population.

The blue-grey ring phenotype is expressed only in seeds with an eye pattern, so all lines which did not have an eye were considered missing data. Of those remaining, 3 parents expressed the trait and 1 did not. Since the other 4 parents may carry the trait but not express it a chi-square

analysis would not be a reliable method of predicting expected segregation ratios. In the population, 92 lines expressed the trait and 47 lines did not.

For the Full Coat pattern 4 parents expressed the trait and 4 did not. In the population 144 lines expressed the trait while 153 did not. The chi-square value for a 1:1 ratio was 0.27 with a p-value of 0.60. This indicates a likely single locus for the Full Coat trait in the MAGIC population.

None of the parents expressed the Holstein pattern, so expected segregation ratios for this trait could not be determined.

Biparental RILs

The CB27 by BB population segregated into four distinct patterns: Eye1, Holstein, Watson, and Full Coat. 21 lines expressed the Eye 1 pattern, 29 lines expressed the Holstein pattern, 19 lines expressed the Watson pattern, and 24 lines expressed the Full Coat pattern. The chi-square for a 1:1:1:1 ratio was 2.44 with a p-value of 0.49. This indicates that the traits are likely controlled by two independent loci.

The CB27 by 556 population segregated into four distinct patterns: Eye 1, Holstein, Watson, and Full Coat. 14 lines expressed the eye pattern, 32 lines expressed the Holstein pattern, 18 lines expressed the Watson pattern, and 20 lines expressed the Full Coat pattern. The chi-square for a 1:1:1:1 ratio was 8.57 with a p-value of 0.04. The deviation from the expected ration may be due to the small population size.

The CB46 by 503 population segregated into Eye 1, Eye 2, Holstein, Watson, and Full Coat patterns. 13 lines expressed the Eye 1 pattern, 50 lines expressed the Eye 2 pattern, 17 lines expressed the Holstein pattern, 9 lines expressed the Watson pattern, and 16 lines expressed the

Full Coat pattern. The chi-square for a 1:4:1:1:1 ratio was 3.19 with a p-value of 0.53. This indicates that the traits are likely controlled by three loci with epistatic effects.

The 524B by 2049 population segregated into Eye 1, Eye 2, Holstein, Watson, and Full Coat patterns. 5 lines expressed the Eye 1 pattern, 47 lines expressed the Eye 2 pattern, 8 lines expressed the Holstein pattern, 6 lines expressed the Watson pattern, and 13 lines expressed the Full Coat pattern. The chi-square for a 1:4:1:1:1 ratio was 6.70 with a p-value of 0.15. This indicates that the traits are likely controlled by three loci with epistatic effects.

F2 populations

Each of the F2 populations resulting from unique parents segregated for different traits.

The two CB27 by Bambey 21 populations segregated for Color Factor, expressing either the Eye 1 or no color phenotypes. The CB27 by Bambey 21 A population had 129 individuals expressing the Eye 1 pattern and 47 which had no color. The chi-square value for a 3:1 ratio was 0.27 with a p-value of 0.60. The chi-square for a 3:1 ratio was 1.61 with a p-value of 0.20. The CB27 by Bambey 21 B population had 88 individuals expressing the Eye 1 pattern and 44 which had no color. The chi-square value for a 3:1 ratio was 4.89 with a p-value of 0.027. The deviation of the second population from the expected 3:1 ratio is possibly due to the small population size. The segregation pattern indicates a likely single locus for color factor in these populations.

The Bambey 21 by CB50 population segregated for Color Factor, expressing either the Eye 1 or no color phenotypes. 112 individuals expressed Eye 1 pattern while 31 had no color. The chi-square value for a 3:1 ratio was 0.84 with a p-value of 0.36. This indicates a likely single locus for this trait in this population.

The two Tvu15426 by MAGIC014 populations segregated between the Full Coat and Watson patterns. The Tvu15426 by MAGIC014 A population had 131 individuals which expressed the Full Coat pattern and 44 which expressed the Watson pattern. The chi-square for a 3:1 ratio was 0.0019 with p-value of 0.97 The Tvu15426 by MAGIC014 B population had 93 individuals which expressed the Full Coat pattern and 27 which expressed the Full Coat pattern. The chi-square value for a 3:1 ratio was 0.4 with a p-value of 0.53. This indicates a likely single locus for the trait in these populations.

QTL mapping

A total of 42 QTL and regions of interest were identified using different methods for each population type (see Materials and Methods for details). The QTL and regions of interest were concentrated on three of the eleven chromosomes of cowpea, Vu07, Vu09, and Vu10 (Table 1). Of those that could be used to determine phenotypic variation explained the range was from 10.1% for the eye pattern on Vu10 in the CB46 by 503 population to 75.9% for the eye pattern on Vu10 in the MAGIC population. Through all methods, regions of interest were identified for the following traits: color factor, eye, blue-grey ring, Holstein, Watson, and Full Coat patterns.

Color factor

For the color factor trait one region was identified as the overlap in associated regions in two F2 populations. In the Bambey 21 by CB50 population an associated region on Vu07 between 2_00392 and 2_04168 (17,387,452 bp) was identified. In the CB27 by Bambey 21 population an associated region on Vu07 between 2_12054 and 2_46013 (15,863,644 bp) was identified.

Eye pattern

For the distinct eye pattern three regions were identified using the MAGIC, minicore, CB27 by BB, CB46 by 503, CB27 by 556, and 524B by 2049 populations.

On Vu07 a region was identified in the MAGIC population from 41 to 43 cM (2,348,152 bp) with a peak at 42 cM. The region had a $-\log_{10}(p)$ value of 189.76 and explained 75.9% of the variation. In the CB46 by 503 population a region was identified between 2_28580 and 2_09527 (18,979,752 bp) with a peak at 2_55172, which had a LOD score of 13.98, an effect of 0.32, and which explained 42.3% of the observed variation. In the 524B by 2049 population a region was identified between 2_38431 and 2_18173 (15,280,092 bp) with a peak at 2_12938, which had a LOD score of 10.65, an effect of 0.34, and which explained 39.6% of the observed variation.

On Vu09 a region was identified in the minicore between 2_52510 and 2_01961 (55,461 bp) with a peak at 2_12692, which had a $-\log_{10}(p)$ value of 7.71. In the CB27 by BB population a region was identified between 2_03581 and 2_50013 (25,219,754 bp) which had a peak at 2_00744, which had a LOD score of 6.13, an effect of -0.21, and which explained 29.4% of the observed variation. In the 524B by 2049 population a region between 2_09783 and 2_54450 (146,745 bp) was identified which had a peak at 2_54450, which had a LOD score of 3.76, an effect of -0.22, and which explained 18.7% of the observed variation.

On Vu10 a region was identified in the minicore between 2_13094 and 2_23834 (5,264 bp), with a peak at 2_31919, which had a $-\log_{10}(p)$ value of 9.24. In the MAGIC population a region from 48 to 54 cM (873,082 bp) with a peak at 50 cM which had a $-\log_{10}(p)$ value of 10.05 and which explained 16.0% of the variation. In the CB27 by BB population a region between 2_01336 and 2_27022 (7,050,573 bp) was identified, with a peak at 2_24359, which had a LOD score of 10.45, an effect of -0.27, and which explained 38.7% of the observed variation. In the CB46 by 503 population a region between 2_13304 and 2_01150 (3,337,074 bp), with a peak at 2_55033,

which had LOD score of 4.39, an effect of -0.21, and which explained 10.1% of the observed variation. In the CB27 by 556 population, a region between 2_01890 and 2_32596 (1,878,809 bp) was identified, with a peak at 2_05167, which had a LOD score of 5.05, an effect of -0.20, and which explained 21.8% of the observed variation. In the 524B by 2049 population a region between 2_23640 and 2_02794 (375,022 bp) was identified, with a peak at 2_51879, which had a LOD score of 4.55, an effect of -0.24, and which explained 21.8% of the observed variation.

Blue-grey ring pattern

On Vu07 a region was identified in the minicore between 2_45021 and 2_09113 (23,572,396 bp), with a peak at 1_0585, which had a $-\log_{10}(p)$ value of 9.20. In the MAGIC population a region between 39 and 44 cM was identified, with a peak at 41 cM, which had a $-\log_{10}(p)$ score of 12.52 and which explained 34.1% of the observed variation.

Holstein pattern

For the Holstein pattern, three significant regions were identified, one on Vu07, one of Vu09, and one of Vu10.

On Vu07 a region was identified in two populations. In the CB46 by 503 population a region was identified between 2_54739 and 2_09638 (6,573,907 bp) with a peak at 2_54104, which has a LOD score of 3.58, an effect of -0.13, and which explains 17.9% of the observed variation. In the 524B by 2049 population a region was identified between 2_55393 and 2_33573 (9,900,686 bp) with a peak at 2_24259, which has a LOD score of 3.42, an effect of -0.13, and which explains 19.4% of the observed variation.

On Vu09 a region was identified in four populations. In the MAGIC population a region was identified between 49 and 51 cM with a peak at 50 cM, which had a LOD score of 18.92 and

which explained 25.5% of the observed variation. In the CB46 by 503 population a region was identified between 2_22742 and 2_18700 (23,073,614 bp) with a peak at 2_55300, which had a LOD score of 4.29, an effect of -0.14, and which explained 20.3% of the observed variation. In the CB27 by 556 population a region was identified between 2_34093 and 2_12652 (22,248,484 bp) which had a peak at 2_49918, which had a LOD score of 10.41, an effect of -0.31, and which explained 40.0% of the observed variation. In the CB27 by BB population a region was identified between 2_01594 and 2_31900 (11,959,155 bp) with a peak at 2_17818, which had a LOD score of 9.97, an effect of -0.29, and which explained 39.9% of the observed variation.

On Vu10 a region was identified in two populations. In the CB27 by 556 population a region was identified between 2_16408 and 2_27022 (3,800,751 bp) with a peak at 2_51879, which had a LOD score of 6.42, an effect of 0.27, and which explained 29.8% of the observed variation. In the CB27 by BB population a region was identified between 2_05564 and 2_51116 (5,649,732 bp) with a peak at 2_04106, which had a LOD score of 8.01, an effect of 0.27, and which explained 35.6% of the observed variation.

Watson pattern

For the Watson pattern two regions of interest were identified, one on Vu09 and one on Vu10.

On Vu09 a region was identified in the CB27 by 556 population between 2_03230 and 2_09379 (16,400,172 bp) with a peak at 2_52510, which had a LOD score of 6.46, an effect of 0.22 and which explained 32.0% of the observed variation. In the 524B by 2049 population a region was identified between 2_33047 and 2_54450 (3,756,209 bp) which had a peak at 2_51640, which had a LOD score of 3.60, an effect of 0.16, and which explained 22.2% of the observed variation. In the CB27 by BB population a region was identified between 2_11603 and 2_52180

(12,105,709 bp), which had a peak at 2_12692, which had a LOD score of 7.16, an effect of 0.22, and which explained 31.5% of the observed variation.

On Vu10 regions were identified in three populations. In the CB27 by 556 population a region was identified between 2_47559 and 2_27022 (7,581,028 bp) with a peak at 2_45879, which has a LOD score of 10.57, an effect of -0.27, and which explains 40.4% of the observed variation. In the CB27 by BB population a region was identified between 2_18936 and 2_27022 (2,974,292 bp) with a peak at 2_31919, which had a LOD score of 7.45, an effect of -0.23, and which explained 28.8% of the observed variation. In the F2 population Tvu15426 by MAGIC014. The region was identified on Vu10 between the SNPs 2_18935 and 2_06138 (2,973,174 bp).

Full Coat pattern

For the Full Coat pattern, three areas of interest were identified, one on each Vu07, Vu09, and Vu10.

On Vu07, a region was identified in three populations. In the MAGIC population, a region between 41 and 43 cM with a peak at 42 cM, which had a $-\log_{10}(p)$ value of 93.61 which explained 61.2% of the observed variation. In the 524B by 2049 population a region was identified between 2_29015 and 2_02375 (2,979,493 bp) with a peak at 2_10045, which had a LOD score of 4.44, an effect of -0.18, and which explained 21.9% of the observed variation. In the CB46 by 503 population a region was identified between 2_53213 and 2_46465 (3,742,070 bp) with a peak at 2_13278, which had a LOD score of 3.46, an effect of -0.13, and which explained 15.1% of the observed variation.

On Vu09, a region was identified in five populations. In the minicore, a region was identified between 2_52510 and 2_08647 (200,821 bp) with a peak at 2_01960, which had a $-\log_{10}(p)$ value

of 8.54. In the MAGIC population a region was identified between 48 and 55 cM with a peak at 52 cM, which had a $-\log_{10}(p)$ value of 11.93 and which explained 18.2% of the observed variation. In the CB46 by 503 population a region was identified between 2_10361 and 2_33606 (2,555,332 bp) with a peak at 2_23814, which had a LOD score of 3.43, an effect of 0.13, and which explained 15.1% of the observed variation. In the CB27 by 556 population a region was identified between 2_11696 and 2_22716 (7,599,881 bp) with a peak at 2_01196, which had a LOD score of 5.04, an effect of 0.20, and which explained 23.3% of the observed variation. In the CB27 by BB population a region was identified between 2_36202 and 2_54420 (24,434,889 bp) with a peak at 1_0643, which had a LOD score of 10.47, an effect of 0.28, and which explained 39.8% of the observed variation.

On Vu10 a region was identified in four populations. In the minicore a region was identified between 2_13094 and 2_23834 (5,264 bp) with a peak at 2_31919, which had a $-\log_{10}(p)$ value of 10.48. In the MAGIC population a region was identified between 48 and 54 cM (873,082 bp) with a peak at 51 cM, which had a $-\log_{10}(p)$ value of 8.24 and which explained 13.6% of the observed variation. In the CB27 by 556 population a region was identified between 2_23558 and 2_27022 (2,415,861 bp) which had a peak at 2_14509, which had a LOD score of 4.74, an effect of 0.20, and which explained 27.9% of the observed variation. In the CB27 by BB population, a region was identified between 1_0077 and 2_27022 (2,646,909 bp) with a peak at 2_54474, which had a LOD score of 5.84, an effect of 0.22, and which explained 29.1% of the observed variation.

Haplotype blocks

On chromosome Vu07 a block was identified in the CB46 by 503 population between 2_53213 and 2_09638 (693,350 bp) and in the 524B by 2049 population between 2_12939 and 2_09638 (228,331 bp). The overlapping area was between 2_12939 and 2_09638 and contained ten genes.

On chromosome Vu09 a block was identified in all four biparental RIL populations. In the CB27 by BB population the block was identified between 2_33224 and 2_12692 (166,724 bp). In the CB27 by 556 population a block was identified between 2_30507 and 2_32747 (469,513 bp). In the CB46 by 503 population a block was identified between 2_10361 and 2_33606 (2,555,332 bp). In the 524B by 2049 population a block was identified between 2_15975 and 2_32747 (19,065,948 bp). The overlapping area was between 2_33224 and 2_12692 and contained seventeen genes.

On chromosome Vu10 a block was identified in all four biparental RIL populations. In the CB27 by BB population a block was identified between 2_55265 and 2_51879 (285,530 bp). In the CB27 by 556 population a block was identified between 2_12467 and 2_15325 (120,513 bp). In the CB46 by 503 population a block was identified between 2_07619 and 2_32426 (278,708 bp). In the 524B by 2049 population a block was identified between 2_18515 and 2_02794 (523,607 bp). The overlapping area was between 2_12467 and 2_15325 and contained eleven genes.

Candidate genes

A predominant candidate gene was identified at each locus based on high relative expression in the seeds and a review of the literature on anthocyanin biosynthesis (see Discussion for details). On Vu07 the gene Vigun07g110700 has emerged as a strong candidate gene. This gene encodes a TT-8 protein, a basic helix-loop-helix protein. On Vu09 the gene Vigun09g139900 has emerged as a strong candidate gene. This gene encodes WD-repeat gene. On Vu10 the gene

Vigun10g163900 has emerged as a strong candidate gene. The gene encodes an E3 Ubiquitin ligase with a zinc finger. Expression level data can be found in Figure S1.

Allelic series

The segregation ratios in the Tvu15426 by MAGIC014 populations were in accord with a 3:1 ratio, indicating the dominance of the presence of the Holstein factor over its absence (Figure 2C). The segregation ratios of the CB27 by Bambey 21 and Bambey 21 by CB50 F2 populations were in accord with a 3:1 ratio, indicating the dominance of the presence of Color Factor over its absence (Figure 2D). The seeds collected from the F1 plant produced by a cross of CB27 with CB27/BB-090 indicated that the Watson pattern is dominant towards the Eye 1 pattern (Figure 2Ei). The seeds collected from the F1 plant produced by a cross of CB27 with CB27/BB-113 indicated that the Holstein pattern is dominant towards the Eye 1 pattern (Figure 2Eii). The seeds collected from the F1 plant produced by a cross of CB27 with CB27/BB-074 indicated that the Full Coat pattern is dominant towards the Eye 1 pattern (Figure 2Eiii).

Stages of color development

Color development in the seed coat has six distinct stages. In Stage 0, there is no color on the seed coat. In Stage 1, color appears at the base of the hilum. In Stage 2, color appears around the hilum. In Stage 3, color begins to spread along the outside edges of the seed. In Stage 4 color begins to fill in on the edges of the testa. In Stage 5, the color has completely developed to the mature level. After Stage 5 the pod and seeds begin to desiccate. Of the observed varieties, only Sasaque and Sanzi completed the full development program. MAGIC059 reached Stage 4, while CB27 only reached Stage 2. No seeds in Stage 0 were observed for Sasaque. Images of each tested variety at various stages can be seen in Figure 3. Color development was associated with

seed size, with larger seeds at more advanced stages of development. The range of seed sizes that were observed in a given stage can be seen in Figure 4.

DISCUSSION

Pattern traits QTL overlap

Several regions of the genome show hotspots for seed coat pattern traits where several traits mapped to overlapping and sometimes identical regions. These regions correspond to genetic factors identified by Harland (1919 and Spillman (1911). Harland (1919) and Spillman (1911) identified four factors controlling seed coat patterning termed Color Factor (C), Watson (W), Holstein-1 (H-1), and Holstein-2 (H-2). On Vu07 six traits, blue-grey ring, eye, Full Coat, color factor, and Holstein mapped to the same region, between 20,314,635 and 20,542,966 bp (228,331 bp). On Vu09, the eye, Holstein, Watson, and Full Coat traits mapped to the same region, between 30,374,388 and 30,429,849 bp (55,461 bp).On Vu10, the eye, Full Coat, Watson, and Holstein traits mapped to an overlapping region between 38,454,572 and 38,459,836 bp (5,264 bp).

Trait segregation efforts from the early twentieth century support the epistatic effects observed in this study. Harland (1919) and Spillman (1911) both produced crosses between varieties with eye and Full Coat patterns, similar to the CB27 by BB and CB27 by 556 populations, and observed segregation ratios between the eye, Holstein, Watson, and Full Coat patterns suggesting two independent loci with epistatic effects. Harland (1919) further noted a segregation ratio indicating a third interacting locus. However, Harland (1919) noted the presence of two H loci, which he termed “H-1” and “H-2.” The present data suggests the presence of only one H locus in

the tested populations. To avoid possible confusion, the Holstein locus discussed here is simply termed "H."

The MAGIC population did not contain any lines expressing the Watson pattern. This could have been due to removal of lines expressing the phenotype early in selection as the Watson pattern appears "dirty" and so may have been rejected.

Multiple QTL clustering in the regions noted above could indicate allelism, especially considering that these traits are mutually exclusive. It may be that the size of the area covered in pigmentation on the seed coat proceeds in stages, each controlled by a different gene, and that various failures to obtain normal gene function cause the observed variation in patterning. Further evidence for this is furnished by the noted developmental pattern of the seed coats. Further work would be required to determine the roles played by the candidate genes in seed coat development.

Epistatic interaction of seed coat pattern loci

In the biparental RIL populations, two distinct segregation patterns were observed. In the CB27 by BB and CB27 by 556 populations the lines segregated in a 1:1:1:1 ratio between the Eye 1, Holstein, Watson, and Full Coat patterns, indicating segregation at two loci. The CB46 by 503 and 524B by 2049 populations segregated in a 1:4:1:1:1 ratio between the Eye 1, Eye 2, Holstein, Watson, and Full Coat patterns, indicating segregation at three loci with epistasis. Considering these data a possible explanation is that three loci are responsible for the observed patterns and that these are concurrent with the C, W, and one of the H factors identified by Harland (1919) and Spillman (1911). The C locus contains a "constriction" factor while the W and H contain distinct "expansion" factors. Four alleles were identified at the locus: a C_0 allele

completely constricts color, resulting in a seed coat with no color; a C_1 allele results in an eye pattern; a C_2 allele allows expansion past the eye, dependent on alleles at loci W and H; a C_3 allele results in an eye with a blue-grey ring around it. An individual with a C_1C_1 genotype, regardless of the genotypes at the W and H loci, expresses an eye pattern, specifically the Eye 2 pattern. However, when the "constriction" factor is missing, in an C_2C_2 genotype, the "expansion" factors can be observed. An individual with the $C_2C_2W_0W_0H_1H_1$ genotype expresses the Holstein pattern, while and individual with the $C_2C_2W_1W_1H_0H_0$ genotype expresses the Watson pattern. An individual with the $C_2C_2W_1W_1H_1H_1$ genotype, expressing both "expansion" factors, expresses the Full Coat phenotype. In the case of an $C_2C_2W_0W_0H_0H_0$ genotype, an Eye 1 pattern would be expressed. The eye pattern would be observed despite the lack of a "constriction" factor due to the absence of the "expansion" factors. Based on this model, the CB27 by BB and CB27 by 556 populations segregate at the W and H loci (Figure 4A), while the CB46 by 503 and 524B by 2049 populations segregate at all three loci (Figure 4B).

The F2 population developed by crossing Tvu15426 and MAGIC014 segregated between the Watson and Full Coat patterns (Figure 4C). The segregation ratio of 3:1 in favor of Full Coat patterning indicates that the Full Coat trait is dominant. Based on the above proposed model, this means that the Tvu15426 by MAGIC014 populations segregate for the at locus H. Bulked segregant analysis showed that the area of interest in the populations as being on Vu10. This indicates that H is on Vu10.

In the MAGIC population, the eye pattern mapped to a major locus on Vu07, with minor loci on Vu09 and Vu10. This result supports the above hypothesis and further, suggests that the "constriction" factor at locus C is on Vu07. This is further supported by the segregation for color factor in the CB27 by Bambey 21 and Bambey 21 by CB50 population also involving a

restriction of color, in this case, between an eye pattern and a lack of color. The region of interest for these population can also be found on Vu07. It also indicates that Color Factor is merely an allele of the constriction locus, termed here C_0, which is recessive to C_2 (Figures 4D, 4G). While C_0 is recessive to C_2, it is unknown at present whether C_0 is epistatic towards the W and H loci, though evidence put forward by Harland (1919) and Spillman (1911) suggest that it is. Additional evidence is provided by the mapping of the blue-grey ring trait to Vu07 in both the MAGIC and minicore population. This phenotype may be due to an additional allele at locus C, C_3.

This would indicate that the Watson "expansion" locus must be on Vu09. Further studies would be required to determine if the Watson "expansion" locus is indeed on Vu09, as well as to determine relative dominance and possible additional epistatic interactions.

Evidence indicating that $W_1>W_0$ and that $H_1>H_0$ comes from the observed segregation ratio of the Tvu15426 by MAGIC014 F2 population (Figure 4C) and the F2 seeds collected from the crosses between CB27 and lines from the CB27 by BB population (Figure 4E).

Observed seed coat pattern as a result of failure to complete the normal developmental program

In examination of the developing seed coats of the various accessions it was noted that the varieties with Full Coat coloration at maturity completely followed the developmental pattern described above. However, those varieties which do not display Full Coat coloration appear to have had color development arrested at points. This is most obvious in CB27, where color development does not proceed past Stage 2. It is likely that other varieties which have distinct eye sizes proceed to varied stages of development. For example, varieties with no seed coat color

at all would not proceed past Stage 0. However, this does not explain the pattern observed in mature Sanzi seed, which exhibits a speckled black and purple seed coat. According to this analysis, Sanzi proceeds through to complete all five stages of seed coat development, indicating that the speckling pattern is controlled separately. A biparental RIL population, consisting of lines derived from a cross between Sanzi and Vita 7, which has a brown colored seed with a Full Coat, which was used for mapping the black seed coat color, showed a perfect correlation between black seed coat color and the speckling pattern (Herniter et al. 2018). This indicates that the speckling pattern is colocalized with black seed coat color and is likely an allele at the B locus. Further research should be conducted to determine if all cowpea accessions follow the pattern observed in the four tested lines. Additionally, transcriptome data could be gathered for the seed coat at each developmental stage. The currently available transcriptome data, (Yao *et al.* 2016, [https://legumeinfo.org/]) uses whole seeds at specific days post flowering, and does not distinguish between transcripts present in the seed coat and those in the embryo or cotyledons and further does not separate them by stage of development. Future studies could involve transcriptome profiling of the developing seed coat.

Candidate gene function

Anthocyanin production is known to be controlled by a transcriptional factor complex composed of an R2-R3 MYB protein, a helix-loop-helix protein, and a WD-repeat protein (Xu, Dubos, and Lepiniec 2015). E3 Ubiquitin ligases are believed negatively regulate this complex (Shin *et al.* 2015). The color and location (such as leaf, seed pod, or seed coat) of the pigmentation are determined by expression patterns. Candidate genes on Vu07 and Vu09 encode a basic helix-loop-helix protein and a WD-repeat protein, respectively. A candidate gene on Vu10 encodes an E3 Ubiquitin ligase. This information lends itself to a model in which Vu07g110700 serves as a

“master switch” while Vigun09g139900 and Vigun10g163900 act as “modulating switches,” altering the effect of the pathway to result in the observed Holstein and Watson patterns (Figure 5).

CONCLUSION

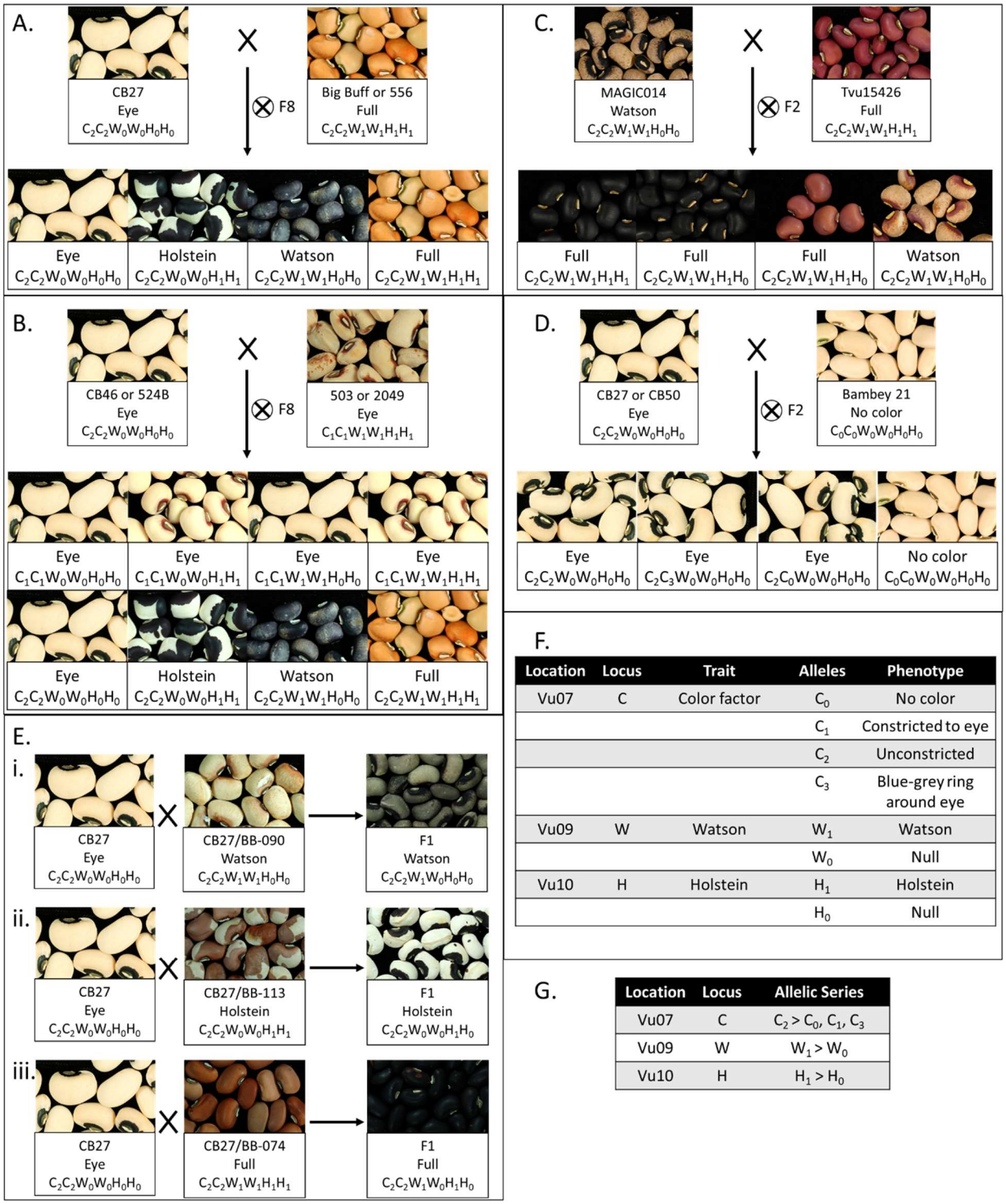
In this study QTL controlling the pattern of seed coat pigmentation in cowpea were identified. Three loci, one on each Vu07, Vu09, and Vu10 were found and a candidate gene identified at each locus. Further, the stages of seed coat development, as determined by the spread of pigmentation were defined. These findings can be utilized in future studies to determine the mechanisms of pigment development and for the development of molecular markers for use in marker assisted selection.

ACKNOWLEDGEMENTS

The authors would like to thank Amy Litt for helpful discussion and guidance; Eric Castillo and Sabrina Phengsy for assistance in seed photography. This study was supported by the Feed the Future Innovation Lab for Climate Resilient Cowpea (USAID Co- operative Agreement AID-OAA-A-13-00070), the National Science Foundation BREAD project “Advancing the Cowpea Genome for Food Security” (NSF IOS-1543963) and Hatch Project CA-R-BPS-5306-H.

Figure 1. Seed coat pattern traits. Images of lines from various populations demonstrating the phenotypes which were scored as part of this study.

Figure 2. Interaction of seed coat pattern loci. Images show the offspring of certain crosses. Number of images indicate relative prevalence of traits and proposed genotypes. (A) Segregation patterns for the CB27 by BB and CB27 by 556 populations. The populations segregate at the Watson (W) and Holstein (H) loci in a 1:1:1:1 ratio of eye, Holstein, Watson, and Full Coat patterns. (B) Segregation patterns for the CB46 by 503 and 524B by 2049 populations. The populations segregate at the Color factor (C) locus as well as the W and H loci. (C) Segregation pattern for the Tvu15426 by MAGIC014 F2 populations. The populations segregate at the H locus in a 3:1 ratio of Full Coat to Watson patterns. (D) Segregation pattern for the CB27 by Bambey 21 and Bambey 21 by CB50 populations. The populations segregate at the C locus in a 3:1 ratio of eye to no color. (E) Phenotype of seeds from the F1 plants resulting from crosses of CB27 with lines from the CB27 by BB. (i) Cross between CB27 and line with a Watson pattern resulting in Watson pattern in the F1. (ii) Cross between CB27 and a line a Holstein pattern resulting in Holstein pattern in the F1. (iii) Cross between CB27 and a line with a Full Coat pattern resulting in a Full Coat pattern in the F1. (F) Table displaying the pattern loci identified in mapping, their locations, the trait encoded, alleles identified, and phenotypes. (G) Table displaying the allelic series and relative dominance of alleles.



F.

Location	Locus	Trait	Alleles	Phenotype
Vu07	C	Color factor	C_0	No color
			C_1	Constricted to eye
			C_2	Unconstricted
			C_3	Blue-grey ring around eye
Vu09	W	Watson	W_1	Watson
			W_0	Null
Vu10	H	Holstein	H_1	Holstein
			H_0	Null

G.

Location	Locus	Allelic Series
Vu07	C	$C_2 > C_0, C_1, C_3$
Vu09	W	$W_1 > W_0$
Vu10	H	$H_1 > H_0$

Figure 3. Seed coat development. Images demonstrating the development the seed and the spread of pigmentation during that process. In Stage 0, there is no pigmentation. In Stage 1, pigmentation appears in a dot at the bottom of the hilum. In Stage 2, pigmentation surround the hilum to form an eye pattern. In Stage 3, pigmentation extends along the edges of the cotyledons. In Stage 4, pigmentation begins to fill in the sides of the seed. In Stage 5, pigment development is completed. Sasaque and Sanzi complete the normal developmental program, but MAGIC059 is arrested at Stage 4 and CB27 is arrested at Stage 2. No Sasaque seeds were observed to be in Stage 0.

Figure 4. Seed size at observed stages of color development. Clockwise from top left: CB27, Color development is arrested at Stage 2; MAGIC059. Color development is arrested at Stage 4; Sasaque, no seeds were observed in Stage 0; Sanzi.

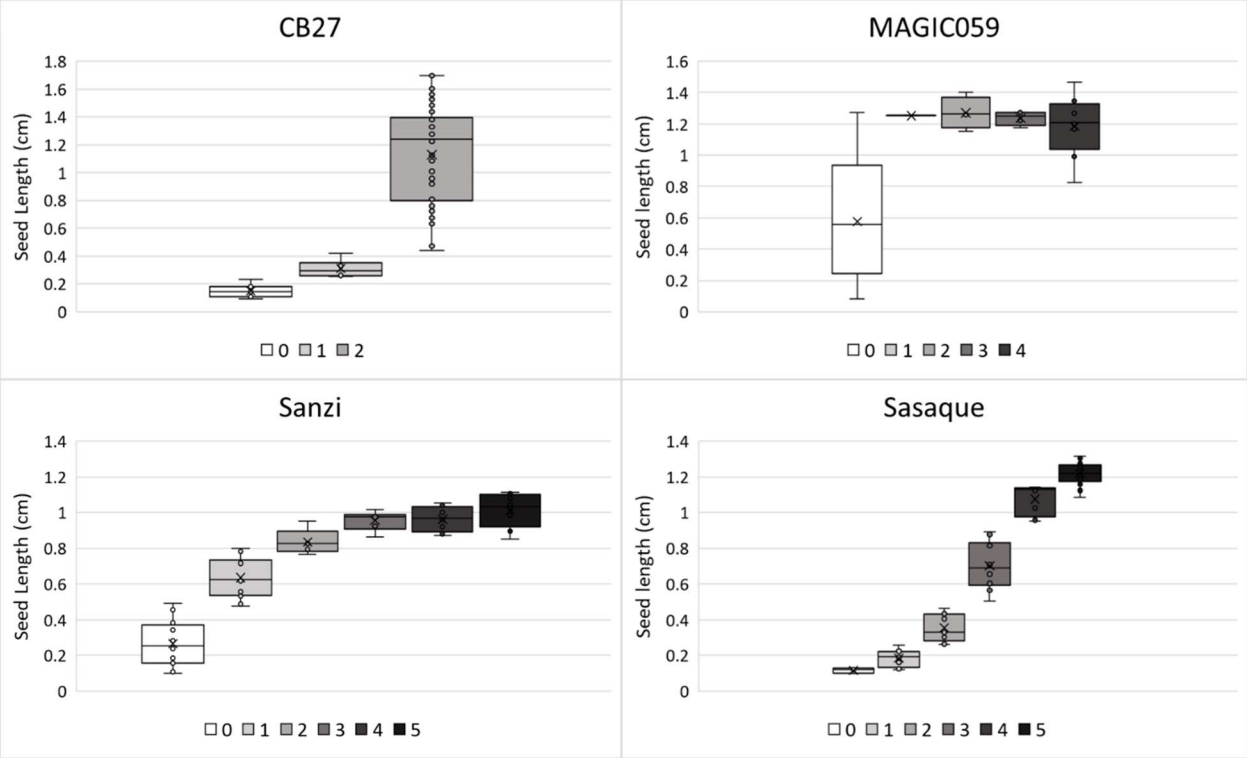


Figure 5. Model of the genetic control of pattern. Transcription of anthocyanin biosynthesis pathway genes are controlled by a complex made of three types of proteins, a basic helix-loop-helix protein (eg. Vigun07g110700), a WD-repeat gene (eg. Vigun09g139900), and an R2R3 MYB transcription factor. This complex is in turn negatively regulated by an E3 Ubiquitin ligase (eg. Vigun10g163900). The observed seed coat pattern phenotypes are a result of different alleles and expression patterns.

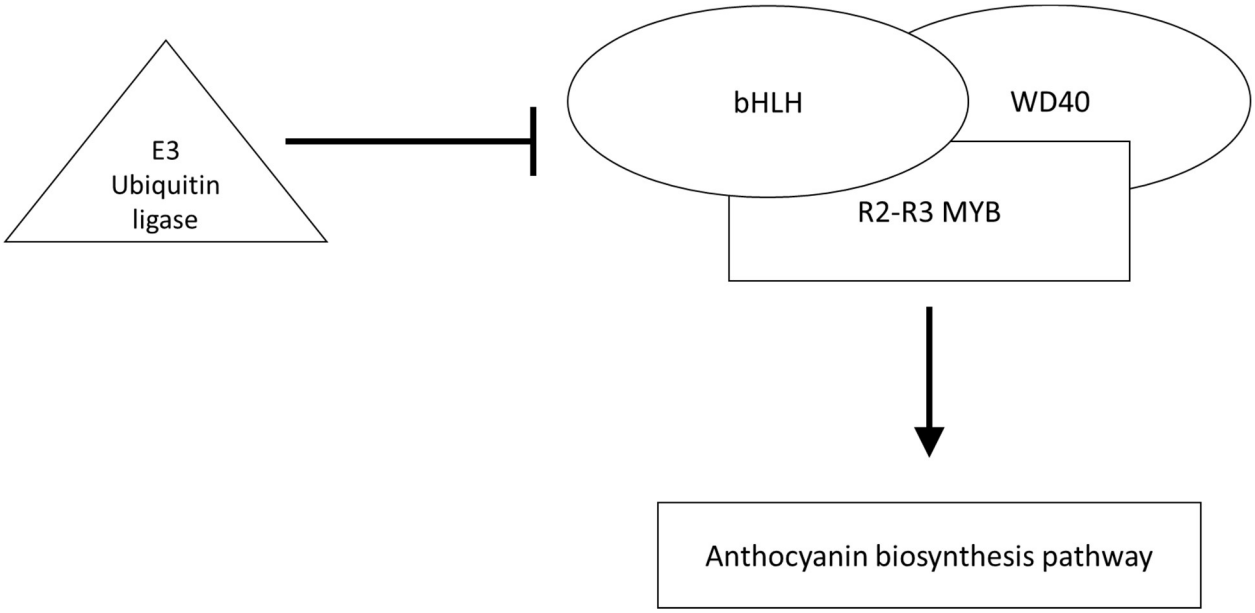
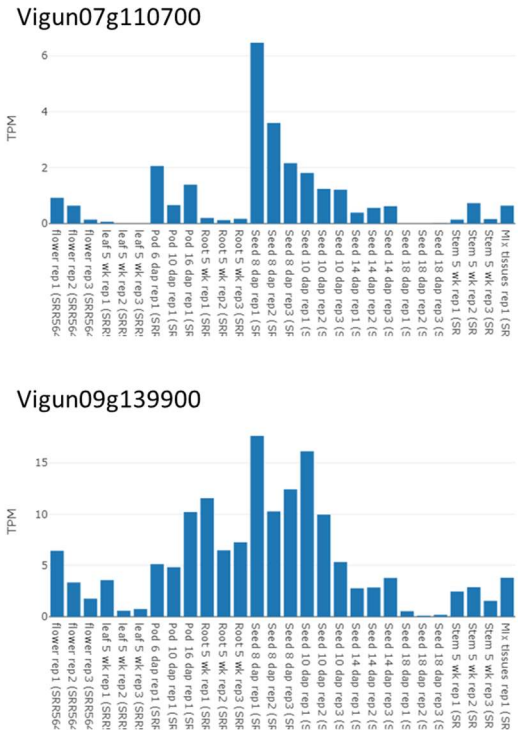
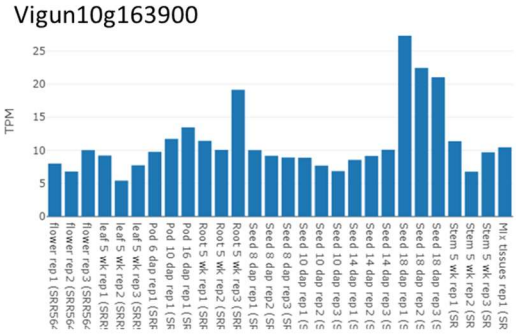


Figure S1. Relative expression levels of the candidate genes. Data retrieved from legumeinfo.org.





LITERATURE CITED

Ehlers, J. D. and A. E. Hall. 1997. "Cowpea (Vigna Unguiculata L. Walp.)." *Field Crops Research* 53(1–3):187–204.

FAOSTAT. 2014. *Crops*. Retrieved (http://www.fao.org/faostat/en/#data/QC).

Faye, M., A. Jooste, J. Lowenberg-DeBoer, and J. Fulton. 2004. "The Influence of Cowpea Characteristics on Cowpea Prices in Senegal." *Agrekon* 43(4):418–29.

Fery, Richard L. 1980. "Genetics of Vigna." Pp. 311–94 in *Horticultural Reviews*, vol. 2. Hoboken, NJ, USA: John Wiley & Sons, Inc. Retrieved (http://doi.wiley.com/10.1002/9781118060759.ch7).

Fery, Richard L. 1985. "Improved Cowpea Cultivars for the Horticultural Industry in the USA." Pp. 129–35 in *Cowpea Research, Production and Utilization*, edited by S. R. Singh and K. O. Rachie. John Wiley & Sons, Inc.

Harland, S. C. 1920. "Inheritance of Certain Characters in the Cowpea (Vigna Sinensis). II." *Journal of Genetics* 10(3):193–205. Retrieved (http://link.springer.com/10.1007/BF03007981).

Harland, S. C. 1919. "Inheritance of Certain Characters in the Cowpea (Vigna Sinensis)." *Journal of Genetics* 8(2):101–32. Retrieved (http://link.springer.com/10.1007/BF02983490).

Herniter, Ira A. et al. 2018. "Identification of Candidate Genes Controlling Black Seed Coat and Pod Tip Color in Cowpea (Vigna Unguiculata [L.] Walp)." *G3; Genes|Genomes|Genetics Genes|Genomes|Genetics* 8(10):3347–55. Retrieved June 29, 2018 (https://www.biorxiv.org/content/early/2018/06/26/355586).

Huang, B. Emma et al. 2015. "MAGIC Populations in Crops: Current Status and Future Prospects." *Theoretical and Applied Genetics* 128(6):999–1017. Retrieved September 6, 2017 (http://link.springer.com/10.1007/s00122-015-2506-0).

Huang, B. Emma and Andrew W. George. 2011. "R/MpMap: A Computational Platform for the Genetic Analysis of Multiparent Recombinant Inbred Lines." *Bioinformatics* 27(5):727–29. Retrieved March 12, 2018 (https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq719).

Huynh, Bao-Lam et al. 2016. "A Major QTL Corresponding to the Rk Locus for Resistance to Root-Knot Nematodes in Cowpea (Vigna Unguiculata L. Walp.)." *Theoretical and Applied Genetics* 129(1):87–95. Retrieved May 14, 2016 (http://link.springer.com/10.1007/s00122-015-2611-0).

Huynh, Bao-Lam et al. 2018. "A Multi-Parent Advanced Generation Inter-Cross (MAGIC) Population for Genetic Analysis and Improvement of Cowpea (Vigna Unguiculata L. Walp.)." *The Plant Journal* 93(6):1129–42. Retrieved March 13, 2018 (http://doi.wiley.com/10.1111/tpj.13827).

Huynh, Bao-Lam Lam et al. 2015. "Genetic Mapping and Legume Synteny of Aphid Resistance

in African Cowpea (Vigna Unguiculata L. Walp.) Grown in California." *Molecular Breeding* 35(1):36. Retrieved October 22, 2018 (http://link.springer.com/10.1007/s11032-015-0254-0).

Jaeger, Sara R. et al. 2018. "Quality Perceptions Regarding External Appearance of Apples: Insights from Experts and Consumers in Four Countries." *Postharvest Biology and Technology* 146:99–107. Retrieved October 22, 2018 (https://www.sciencedirect.com/science/article/pii/S0925521418307488).

Kostyla, Aurora S., Fergus M. Clydesdale, and Mina R. McDaniel. 1978. "The Psychophysical Relationships between Color and Flavor." *Food Science and Nutrition* 10(3):303–21. Retrieved (http://www.tandfonline.com/doi/abs/10.1080/10408397809527253).

Langyintuo, A. S. et al. 2003. "Cowpea Supply and Demand in West and Central Africa." *Field Crops Research* 82(2):215–31.

Langyintuo, A. S., G. Ntoukam, L. Murdock, J. Lowenberg-DeBoer, and D. J. Miller. 2004. "Consumer Preferences for Cowpea in Cameroon and Ghana." *Agricultural Economics* 30(3):203–13. Retrieved July 26, 2016 (http://doi.wiley.com/10.1111/j.1574-0862.2004.tb00189.x).

Lucas, Mitchell R. et al. 2013. "Markers for Breeding Heat-Tolerant Cowpea." *Molecular Breeding* 31(3):529–36.

Lucas, MR, NN Diop, and S. Wanamaker. 2011. "Cowpea–soybean Synteny Clarified through an Improved Genetic Map." *The Plant Genome* 4(3):218–25. Retrieved July 29, 2016 (https://dl.sciencesocieties.org/publications/tpg/abstracts/4/3/218).

Menéndez, C. M., A. E. Hall, and P. Gepts. 1997. "A Genetic Linkage Map of Cowpea (Vigna Unguiculata) Developed from a Cross between Two Inbred, Domesticated Lines." *TAG Theoretical and Applied Genetics* 95(8):1210–17. Retrieved March 28, 2017 (http://link.springer.com/10.1007/s001220050683).

Mishili, Fulgence Joseph et al. 2009. "Consumer Preferences for Quality Characteristics along the Cowpea Value Chain in Nigeria, Ghana, and Mali." *Agribusiness* 25(1):16–35. Retrieved July 28, 2016 (http://doi.wiley.com/10.1002/agr.20184).

Muchero, W. et al. 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):18159–64. Retrieved July 12, 2016 (http://www.pnas.org/cgi/doi/10.1073/pnas.0905886106).

Muñoz-Amatriaín, María et al. 2017. "Genome Resources for Climate-Resilient Cowpea, an Essential Crop for Food Security." *The Plant Journal* 89(5):1042–54. Retrieved (http://biorxiv.org/lookup/doi/10.1101/059261).

Pottorff, Marti et al. 2014. "Identification of Candidate Genes and Molecular Markers for Heat-Induced Brown Discoloration of Seed Coats in Cowpea [Vigna Unguiculata (L.) Walp]." *BMC Genomics* 15(1):328. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4035059&tool=pmcentrez&rendertype=abstract).

Pottorff, Marti, Jeffrey D. Ehlers, Christian Fatokun, Philip A. Roberts, and Timothy J. Close. 2012. "Leaf Morphology in Cowpea [Vigna Unguiculata (L.) Walp]: QTL Analysis, Physical Mapping and Identifying a Candidate Gene Using Synteny with Model Legume Species." *BMC Genomics* 13(1):234. Retrieved August 3, 2015 (http://www.biomedcentral.com/1471-2164/13/234).

Santos, Jansen Rodrigo Pereira, Arsenio Daniel Ndeve, Bao Lam Huynh, William Charles Matthews, and Philip Alan Roberts. 2018. "QTL Mapping and Transcriptome Analysis of Cowpea Reveals Candidate Genes for Root-Knot Nematode Resistance." *PLoS ONE* 13(1):1–22.

Shin, Dong Ho et al. 2015. "Identification of Genes That May Regulate the Expression of the Transcription Factor Production of Anthocyanin Pigment 1 (PAP1)/MYB75 Involved in Arabidopsis Anthocyanin Biosynthesis." *Plant Cell Reports* 34(5):805–15. Retrieved December 20, 2018 (http://link.springer.com/10.1007/s00299-015-1743-7).

Singh, B. B. 2014. *Cowpea: The Food Legume of the 21st Century*. Retrieved May 11, 2016 (https://dl.sciencesocieties.org/publications/books/tocs/acsesspublicati/cowpeathefoodle).

Spillman, W. J. 1911. "Inheritance of the 'Eye' in Vigna." *The American Naturalist* XLV(53):513–23.

Tijjani, A. R., R. T. Nabinta, and M. Muntaka. 2015. "Adoption of Innovative Cowpea Production Practices in a Rural Area of Katsina State , Nigeria." *Journal of Agricultural and Crop Research* 3(June):53–58.

Wu, Yonghui, Prasanna R. Bhat, Timothy J. Close, and Stefano Lonardi. 2008. "Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph" edited by L. Kruglyak. *PLoS Genetics* 4(10):e1000212. Retrieved January 2, 2019 (https://dx.plos.org/10.1371/journal.pgen.1000212).

Xu, Wenjia, Christian Dubos, and Loïc Lepiniec. 2015. "Transcriptional Control of Flavonoid Biosynthesis by MYB–bHLH–WDR Complexes." *Trends in Plant Science* 20(3):176–85. Retrieved December 27, 2018 (https://www.sciencedirect.com/science/article/pii/S1360138514003124).

Yao, Shaolun et al. 2016. "The Vigna Unguiculata Gene Expression Atlas (VuGEA) from de Novo Assembly and Quantification of RNA-Seq Data Provides Insights into Seed Maturation Mechanisms." *The Plant Journal* 88(2):318–27. Retrieved June 11, 2018 (http://doi.wiley.com/10.1111/tpj.13279).

Zhang, Zhiwu et al. 2010. "Mixed Linear Model Approach Adapted for Genome-Wide Association Studies." *Nature Genetics* 42(4):355–60. Retrieved March 20, 2017 (http://www.ncbi.nlm.nih.gov/pubmed/20208535).