UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Mapping Domestication-Related Traits and QTL Pyramiding in Cowpea [Vigna unguiculata (L.) Walp]

Permalink

https://escholarship.org/uc/item/59g7c9vw

Author

Lo, Sassoum

Publication Date

2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mapping Domestication-Related Traits and QTL Pyramiding in Cowpea [*Vigna unguiculata* (L.) Walp]

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Sassoum Lo

June 2019

Dissertation Committee:
Dr. Timothy J. Close, Chairperson
Dr. Philip A. Roberts
Dr. Shizhong Xu

Copyright by
Sassoum Lo
2019

The Dissertation of Sassoum Lo is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgement

I am pleased to thank all the people that helped me for the realization of this document. First, I would like to thank my supervisor Dr. Timothy Close who allowed me to work on this interesting topic. His guidance allowed me to acquire scientific knowledge. I would like to thank Dr. Philip A. Roberts and Dr. Shizhong Xu for their guidance and evaluating this work. I express my gratitude to Dr. María Muñoz-Amatriaín for serving as a great mentor.

I address my gratitude to Dr. Ndiaga Cisse and through him the Institut Senegalais de la Recherche Agricole and the West Africa Agricultural Productivity Program for believing in me.

Special recognition to current and former members of the UCR cowpea team Dr. Stefano Lonardi, Dr. Bao-Lam Huynh, Ira Herniter, Yi-Ning Guo, Qihao Liang, Steve Wanamaker, Dr. Arsenio Ndeve, Dr. Mitchell Lucas, Dr. Jansen Santos, Audrey Pongs, Eric Castillo, Sabrina Phengsy, Julia Valdepena and Mia Rochford for useful discussions, feedback and help in the lab.

My thanks go to Dr. Christian Fatokun, Dr. Ousmane Boukar, Dr. Paul Gepts, Dr. Andrew Farmer, Dr. Samuel Hokin and Travis Parker for their collaboration in this work.

I would like also to thank Dr. Mikeal Roose, Dr. Giles Waines, Dr. Adam Lukaszewski, and Dr. Juan Pablo Giraldo for serving on either my guidance committee, or qualifying exam committee.

Great thanks to members of the Plant Biology Graduate Program specially Dr. Linda Walling, Dr. Amy Litt, Dr. Dawn Nagel, Dr. John Chater, Dr. Brandon Le, Yoko Hiraoka and Eric Focht.

I owe great thanks to my family, friends and some special persons Dr. Sheila Close and Mary Camin.

Thanks to all those who have supported me in any way during this stage of my life.

The text of chapter 1 is a reprint with some edits of the material as it appears in Scientific Reports [Identification of QTL controlling domestication-related traits in cowpea (Vigna unguiculata L. Walp)]

Merci beaucoup Dieureudieuf

Dedication

This dissertation is dedicated to my family, friends and my late parents.

ABSTRACT OF THE DISSERTATION

Mapping Domestication-Related Traits and QTL
Pyramiding in Cowpea [*Vigna unguiculata* (L.) Walp]

by

Sassoum Lo

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, June 2019
Dr. Timothy J. Close, Chairperson

Cowpea (*Vigna unguiculata* [L.] Walp) is one of the most important food and nutritional security crops. It is one of the main sources of dietary protein and folic acid for millions of people in sub-Saharan Africa and other parts of the developing world. Cowpea is a diploid (2n = 22) with a genetically diverse gene-pool composed of wild and cultivated forms. Cowpea was domesticated in Africa, from where it spread into all continents and now is commonly grown in many parts of Asia, Europe, United States and Central and South America. Domestication of cowpea has, in general, resulted in a determinate growth habit, increased pod and seed size, early flowering, and reduction of pod shattering. However, the genetic control of these traits is largely unknown. This lack of domestication-related knowledge in cowpea has limited the utilization of broad germplasm for crop improvement. This dissertation investigates the genetic basis of domestication traits using recently developed genetic and genomic resources available for cowpea and studies the effect of two seed size loci in a different genetic background. In the first chapter, regions of the cowpea genome that played an important role in cowpea domestication were identified. A

total of 215 recombinant inbred lines derived from a cross between a cultivated and a wild cowpea accession were used to evaluate nine domestication-related traits (pod shattering, peduncle length, flower color, days to flowering, 100-seed weight, pod length, leaf length, leaf width and seed number per pod). A high-density genetic map containing 17,739 single nucleotide polymorphisms was constructed and used to identify 16 quantitative trait loci (QTLs) for these nine traits. Four regions were identified showing QTL clustering for these traits, including one region on Vu08 where four QTLs related to increased organ size (seed weight, pod length, leaf length and leaf width) were mapped. Using sequence homology comparison with common bean, a candidate gene (*Vigun08g217000*) for increased organ size was identified. This gene codes for a histidine kinase 2 and the Arabidopsis ortholog AHK2 (*AT5G35750.1*) has been shown to regulate, among other things, plant organ size. The second chapter investigates the genetic basis of seed size, which is a main domestication target and one of the key yield determinants. A "mini-core" panel of 368 genetically diverse cowpea accessions, mainly landraces, was used to evaluate four seed size-related traits (seed weight, length, width, and density). A genome-wide association study (GWAS) and meta-analysis identified 17 loci associated with seed size. One locus was common to weight, width and length, suggesting pleiotropy. By integrating synteny based-analysis with common bean, six candidate genes (*Vigun05g036000*, *Vigun05g039600*, *Vigun05g204200*, *Vigun08g217000*, *Vigun11g187000*, and *Vigun11g191300*) which are implicated in multiple functional categories related to seed size were identified. In the third chapter, two seed size QTLs were targeted for introgression in the background of a popular cultivar from Senegal "Pakau". Four

combinations of the positive (large seed size) alleles of the two QTL were observed in the backcross progenies and were analyzed to test the effect on seed phenotype of these QTL in the Pakau genetic background: no positive QTL alleles, positive allele of *Css-1* only, positive allele of *Css-4* only, and positive alleles of both QTL.

The results of this dissertation provide a basis for further fine mapping of genes controlling domestication traits and a foundation for the utilization and exploitation of diverse landraces and wild relatives in cowpea breeding programs.

Table of Contents

Contents

List of Tables

List of Figures

Introduction

An overview on the origin and domestication of cowpea

Ten thousand years ago human societies around the globe began to transition from hunting and gathering to agriculture. By this time, ancient peoples had already domesticated the major crop species (Harlan 1992) upon which human survival is dependent, including rice, wheat, and maize. Compared to their wild progenitors, food crops typically have larger fruits or grains, more robust plants overall, more determinate growth, and a loss of natural seed dispersal. These characters are collectively called the domestication syndrome (Hammer 1984). The genetic basis of domestication traits has come from research focusing mainly on major grains. This focus has left gaps in our knowledge on the genetics of domestication traits in other crop species including cowpea.

Cowpea [*Vigna unguiculata* (L.) Walp] is one of the highest sources of protein, folic acid and several vitamins for many people in sub-Saharan Africa where it is grown for fresh and dry seeds, fresh pods and leaves. Cowpea is also grown in some parts of Asia, Latin America, and the United States (Dadson et al. 2005). It contributes substantially to global food and nutritional security. In addition to providing food for humans cowpea also provides fodder for livestock. Among its advantages are: a general drought tolerance, the ability to improve soil conditions through nitrogen fixation, and the beneficial impact in households by diversifying their sources of income. Cowpea belongs to the family Fabaceae and genus *Vigna* and shares a high degree of genetic collinearity with common bean (*Phaseolus vulgaris* L.) (Muñoz-Amatriaín et al. 2017; Vasconcelos et al. 2015). It

is an annual, diploid (2n = 22) crop with a morphologically and genetically diverse gene-pool composed of cultivated forms and several wild taxa (Pasquet 1996). Cultivated cowpeas are generally divided into four cultivar groups: *unguiculata*, *biflora*, *sesquipedalis*, and *textilis* (Maréchal et al. 1978). Cultivar groups *unguiculata* and *sesquipedalis* are grown as pulse and vegetable respectively, while *biflora* is grown for forage and *textilis* cultivated for the fibres of its floral peduncles. Later, Pasquet (1998) proposed the addition of another cultivar group, *melanophthalmus*. Among the cultivated cowpeas, subsp. *unguiculata* and subsp. *sesquipedalis* are the most widely cultivated.

The history of cowpea dates to ancient West African cereal farming, five to six thousand years ago, where it was closely associated with the cultivation of sorghum (*Sorghum bicolor*) and pearl millet (*Pennisetum glaucum*) (Steele and Mehra 1980). Although many details on cowpea domestication are debated, the wild annual *V. unguiculata* subsp. *dekindtiana* is believed to be the probable progenitor of cultivated cowpeas (Lush 1979; Lush and Evans 1981; Ng and Marechal 1985; Rawal 1975; Steele and Mehra 1980). On the other hand, Pasquet (1999) reported that annual cowpea (subsp. *unguiculata* var. *spontanea*) could be the most likely progenitor of cultivated cowpeas. In this study, Pasquet investigate the cowpea gene-pool organization and evolution by characterizing a panel comprising 91 accessions from eight perennial subspecies, 95 wild annual accessions and 13 cultivated accessions with 21 allozyme markers. Based on the genetic distances, he found that the closest group to cultivated cowpea among the subspecies studied is var. *spontanea*.

The precise location of the center of domestication of cowpea is still controversial. Previous speculation on the origin and domestication of cowpea had been based on botanical and cytological evidence, information on its geographical distribution and historical records (Faris 1965; Steele and Mehra 1980) and severales domestication locales have been proposed including West and Northeast Africa (Faris 1965; Pasquet 2000; Rawal 1975; Steele and Mehra 1980; Vaillancourt and Weeden 1992). Carbon dating (c. 3500 BP) of cowpea (or wild cowpea) remains from the Kintampo Rock Shelter in central Ghana has been carried out and this indicates that wild cowpeas could have been gathered as fodder to feed cattle and later domesticated as early as 4000 BP in West Africa (Flight 1976). Based on the oldest archeological evidence for cowpea in Ghana, the identification of a wild and cultivated accession with an identical cpDNA in Nigeria, the existence of weedy intermediates between wild and cultivated cowpeas, and the highest level of morphological diversity between cultivated cowpea, the western African center of domestication has been proposed (D'Andrea et al. 2007; Ng 1995; Ng and Marechal 1985; Vaillancourt and Weeden 1992). An alternative center of domestication has been proposed in northeastern Africa based on detailed studies on morphological diversity and ecogeographical information (Baudoin and Maréchal 1985), and isozyme studies (Pasquet 1996). Also, Coulibaly et al. (2002) provided evidence based on molecular markers that early domestication occurred in northeastern Africa. Cowpea in these regions could have been domesticated together with sorghum and pearl millet in the third millennium BC (Steele 1976). However, some investigations showed that the highest genetic diversity and most primitive wild varieties occur in southern

Africa in the region encompassing Namibia from the west, across Botswana, Zambia, Zimbabwe and Mozambique to the east, and the Transvaal region of the Republic of South Africa (Ng 1995). Based on this as well as some characteristics such as perenniality, hairiness, small size of the pods and seeds, Padulosi and Ng (1997) suggested that southern Africa may be the site of origin of cowpea. All these studies showed divergence on a narrowly defined center of origin but agreed that cowpea was domesticated somewhere on the African continent where wild forms are encountered.

Mapping domestication-related traits

Determining the genetic basis of key domestication traits is important to better understand the genetics of important traits. This information can be used for crop improvement using marker assisted selection. Studies have been conducted to identify quantitative trait loci (QTL) associated with domestication-related traits in many crop species. Several QTLs and specific genes controlling the differences between maize and its wild progenitor teosinte have been reported (Briggs et al. 2007; Clark et al. 2004; Doebley 2004; Weber et al. 2008). A major locus *teosinte branched 1* (*tb1*) was identified controlling the growth habit differences between maize and teosinte (Doebley and Stec 1991). The major effect gene, *tb1*, is a transcription factor that acts as a repressor of shoot branching. Fine mapping later determined that a transposable element Hopscotch inserted in the regulatory region of *tb1* is the causative polymorphism controlling morphological differences between maize and teosinte (Studer et al. 2011).

Domestication of tomato has been studied focusing on the dramatic increase in fruit size. The wild progenitor *Lycopersicon esculentum* has small fruits compared to cultivated tomatoes. Many QTL have been identified with a major locus, *fw2.2*, explained up to 30% of the phenotypic variation. The cause of *fw2.2* effect is a single gene, ORFX, that is expressed in floral development and controls carpel cell number (Frary et al. 2000).

In soybean, studies have been conducted to identify loci associated with domestication related traits including pod shattering. A major locus (*qPdh1*) associated with pod shattering has been identified consistently in different studies (Bailey et al. 1997; Liu et al. 2007; Suzuki et al. 2009). A dirigent-like protein (*Glyma16g25580*; *Pdh1*) has been identified in the QTL region as the putative trait-determining gene (Funatsuki et al. 2014). *Pdh1* transcript was found abundant in the inner sclerenchyma of pod walls and likely promotes shattering by increasing torsion forces through lignin accumulation (Funatsuki et al. 2014). Another gene associated with pod shattering was cloned and named *SHAT1-5* (Dong et al. 2014). Authors reported that shattering resistance lies in the excessively lignified fiber cap cells (FCC). *SHAT1-5* activates secondary wall biosynthesis and promotes significant thickening of FCC secondary walls (Dong et al. 2014).

In cowpea, the genetic basis underlying many domestication-related traits remain unknown. Relatively little research has been conducted to identify QTLs associated with domestication-related traits. No candidate genes have been reported prior to the work in this thesis. An earlier study was conducted to mapped QTLs for 10 domestication-related traits (number of ovules, seed coat thickness, seed weight, pod fiber layer thickness, days

to flowering, seed germination, root architecture, pod color, growth habit, and pod shattering) (Andargie et al. 2014). A biparental mapping population of 159 F7 recombinant inbreeds lines (RILs) derived from a cross between cultivated and wild cowpeas was used. Using 202 simple sequence repeat (SSR) and 4 morphological markers, they detected 15 QTLs for the 10 traits, with the QTLs explaining 8-20% of the phenotypic variation. All the 15 QTLs were distributed on 6 of the 11 chromosomes of cowpea.

Larger seed size is one of the major traits that was selected during domestication. Seed size has been studied based on seed weight. Classical genetic analysis suggested that at least eight factors influenced seed size (Aryeetey and Laing 1973; Drabo et al. 1984). Later Fatokun et al. (1992) published the first report of seed weight QTLs using molecular markers. Using 58 F2 lines derived from a cross between cultivated and wild cowpeas, and 188 restriction fragment length polymorphism (RFLP) markers they identified two major QTLs. These QTLs explained 32-36% of the phenotypic variation and are orthologous to seed weight QTLs in mung bean (*Vigna radiata*). Another study was conducted using 94 F8 RILs derived from the intersubspecific cross of an improved line and a wild accession (Ubi et al. 2000). They used 77 randomly amplified polymorphic DNA (RAPD) markers and identified five loci for seed weight. These QTLs explained between 7-15% of the phenotypic variation. The relationship between the seed weight QTLs in the study of Ubi et al. (2000) and those identified in the study of Fatokun et al. (1992) was not clarified. Since then, efforts have been made to identify more loci with different types of markers and genetic background. Most of them used RILs derived

from crosses between cultivated accessions. For example, Lucas et al. (2013) identified 10 QTLs for seed weight using eight biparental mapping populations and 1,536 SNPs. Seven of these 10 QTLs were syntenic to regions associated with seed size in soybean (*Glycine max*). Recently, two of these QTL were also mapped in a multi-parent advanced generation inter-cross (MAGIC) population (Huynh et al. 2018)

Studies have been conducted to map domestication traits in yardlong bean (*V. unguiculata* subsp. *sesquipedalis*). Kongjaimun et al. (2012) analyzed 24 domestication-related traits including seed dormancy, plant type, earliness, pigmentation in 188 F2 and 190 BC1F1populations derived from a cross between a cultivated yardlong bean and wild cowpea. Using 226 SSR markers for the BC1F1 population and 113 SSRs for the F2, they identified 153 QTLs for 24 traits. One to 11 QTLs were detected for each trait. Results show co-localization of many domestication-related QTLs. Another study conducted by Suanum et al. (2016) reported co-localization of QTLs for pod fiber content and pod shattering. They used the same populations as Kongjaimun et al. (2012) and identified two QTLs for pod shattering, three, two, and one QTLs for cellulose, hemicellulose, and lignin content, respectively. The QTLs explained 2-75% of the phenotypic variation of the trait. Based on comparative genome analysis with adzuki bean (*Vigna angularis*), they suggested that the QTL region for cellulose, hemicellulose, lignin and pod shattering in yardlong bean contains genes encoding a MYB transcription factor, MYB83, regulating biosynthesis of the three fibers.

Other domestication-related traits such as flower color and pod length have also been mapped in yardlong bean. Xu et al. (2011) conducted a study to map flower and seed coat

color using 209 F7:8 RILs derived from a cross between a landrace and a commercial cultivar of yardlong bean, with 184 SSRs and 191 single nucleotide polymorphism (SNP) markers. They identified one locus for each trait, and both loci are tightly linked with a genetic distance of 0.4 cM. Synteny analysis with soybean showed that the locus for seed coat is syntenic the soybean T locus controlling seed coat color. For pod length, Xu et al. (2017) detected 72 SNPs by genome-wide association study (GWAS) using a diversity panel of 299 landraces and breeding lines. Phenotypic variation explained by a single SNP varied from 4.6-7.1%. Transcriptomic analysis in this study suggested the involvement of sugar, gibberellin and nutritional signaling in regulation of pod length.

Knowledge of the genetic basis of domestication traits is valuable to utilize beneficial alleles from broad germplasm efficiently for further improvement of complex traits, which may be missing in cultivated accessions because of the genetic bottleneck induced by domestication.

The following chapters are the results of my PhD thesis investigating the genetics of domestication-related traits in cowpea. The first chapter focuses on identifying genomic regions associated with domestication-related traits including pod shattering, seed size, days to flowering, using a biparental mapping population. The second chapter investigates the genetic basis of seed size more broadly using diverse cowpea accessions. The remaining chapter extends these results by studying the effect of two seed size QTLs in a different genetic background.

References

Andargie M, Pasquet RS, Gowda BS, Muluvi GM, Timko MP (2014) Molecular mapping of QTLs for domestication-related traits in cowpea (*V. unguiculata* (L.) Walp.). *Euphytica* 200:401-412

Aryeetey A, Laing E (1973) Inheritance of yield components and their correlation with yield in cowpea (*Vigna unguiculata* (L.) Walp.). *Euphytica* 22:386-392

Bailey M, Mian M, Carter T, Ashley D, Boerma H (1997) Pod dehiscence of soybean: identification of quantitative trait loci. *Journal of Heredity* 88:152-154

Baudoin J, Maréchal R (1985) Genetic diversity in Vigna. Cowpea research, production and utilization. *Wiley New York*, pp 1-9

Briggs WH, McMullen MD, Gaut BS, Doebley J (2007) Linkage mapping of domestication loci in a large maize–teosinte backcross resource. *Genetics* 177:1915-1928

Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene tb1. *Proceedings of the National Academy of Sciences* 101:700-707

Coulibaly S, Pasquet R, Papa R, Gepts P (2002) AFLP analysis of the phenetic organization and genetic diversity of *Vigna unguiculata* L. Walp. reveals extensive gene flow between wild and domesticated types. *Theoretical and Applied Genetics* 104:358-366

D'Andrea AC, Kahlheber S, Logan AL, Watson DJ (2007) Early domesticated cowpea (*Vigna unguiculata*) from Central Ghana. *Antiquity* 81:686-698

Dadson R, Hashem F, Javaid I, Joshi J, Allen A, Devine T (2005) Effect of water stress on the yield of cowpea (*Vigna unguiculata* L. Walp.) genotypes in the Delmarva region of the United States. *Journal of agronomy and crop science* 191:210-217

Doebley J (2004) The genetics of maize evolution. *Annu Rev Genet* 38:37-59

Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. *Genetics* 129:285-295

Dong Y, Yang X, Liu J, Wang B-H, Liu B-L, Wang Y-Z (2014) Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nature communications* 5

Drabo I, Redden R, Smithson J, Aggarwal V (1984) Inheritance of seed size in cowpea (*Vigna unguiculata* (L.) Walp.). *Euphytica* 33:929-934
Faris D (1965) The origin and evolution of the cultivated forms of *Vigna sinensis*. *Canadian journal of genetics and cytology* 7:433-452

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841-846

Flight C (1976) The Kintampo Culture and Its Place inthe Economic Prehistory of West Africa. *Origins of African plant domestication*:211

Frary A, Nesbitt TC, Frary A, Grandillo S, Van Der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB (2000) *fw2. 2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85-88

Funatsuki H, Suzuki M, Hirose A, Inaba H, Yamada T, Hajika M, Komatsu K, Katayama T, Sayama T, Ishimoto M (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proceedings of the National Academy of Sciences* 111:17797-17802

Hammer K (1984) Das domestikationssyndrom. *Die Kulturpflanze* 32:11-34

Harlan JR (1992) Crops and man. *American Society of Agronomy*

Huynh BL, Ehlers JD, Huang BE, Muñoz-Amatriaín M, Lonardi S, Santos JR, Ndeve A, Batieno BJ, Boukar O, Cisse N (2018) A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *The Plant Journal* 93:1129-1142

Kongjaimun A, Kaga A, Tomooka N, Somta P, Vaughan DA, Srinives P (2012) The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. unguiculata cv.-gr. sesquipedalis. *Annals of botany* 109:1185-1200

Liu B, Fujita T, Yan Z-H, Sakamoto S, Xu D, Abe J (2007) QTL mapping of domestication-related traits in soybean (*Glycine max*). *Annals of Botany* 100:1027-1038

Lucas MR, Huynh BL, da Silva Vinholes P, Cisse N, Drabo I, Ehlers JD, Roberts PA, Close TJ (2013) Association Studies and Legume Synteny Reveal Haplotypes Determining Seed Size in *Vigna unguiculata*. *Frontiers in plant science* 4:95

Lush W (1979) Floral morphology of wild and cultivated cowpeas. *Economic Botany* 33:442-447

Lush W, Evans L (1981) The domestication and improvement of cowpeas (*Vigna unguiculata* (L.) Walp.). *Euphytica* 30:579-587

Maréchal R, Mascherpa J-M, Stainier F (1978) Etude taxonomique d'un groupe complexe d'especes des genres Phaseolus et Vigna (Papilionaceae) sur la base donnees morphologiques et polliniques, traitees par l'analyse informatique. *Boissiera* 28:2-7

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O (2017) Genome resources for climate-resilient cowpea, an essential crop for food security. *The Plant Journal* 89:1042-1054

Ng N (1995) Cowpea, Vigna unguiculata (Leguminosae-Papil-ionoideae). *Evolution of crop plants*:326-332

Ng N, Marechal R (1985) Cowpea taxonomy, origin and germplasm. Cowpea research, production, and utilization Wiley, Chichester, UK:11-21

Padulosi S, Ng N (1997) Origin, taxonomy, and morphology of *Vigna unguiculata* (L.) Walp. Advances in cowpea research:1-12

Pasquet R (1996) Wild cowpea (*Vigna unguiculata*) evolution. Advances in legume systematics 8:95-100

Pasquet R (2000) Allozyme diversity of cultivated cowpea *Vigna unguiculata* (L.) Walp. *Theoretical and Applied Genetics* 101:211-219

Pasquet RS (1998) Morphological study of cultivated cowpea *Vigna unguiculata* (L) Walp: importance of ovule number and definition of cv gr *Melanophthalmus*. A*gronomie-Sciences des Productions Vegetales et de l'Environnement* 18:61-70

Pasquet R (1999) Genetic relationships among subspecies of *Vigna unguiculata* (L.) Walp. based on allozyme variation. *Theoretical and Applied Genetics* 98:1104-1119

Rawal KM (1975) Natural hybridization among wild, weedy and cultivated *Vigna unguiculata* (L.) Walp. *Euphytica* 24:699-707

Steele W (1976) Cowpeas. In “Evolution of Crop Plants”.(NW Simmonds, ed.). Longman Group. London

Steele W, Mehra K (1980) Structure, evolution and adaptation to farming systems and environments in Vigna. Advance in Legume Sci Royal Botanical Gardens UK
Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. *Nature genetics* 43:1160

Suanum W, Somta P, Kongjaimun A, Yimram T, Kaga A, Tomooka N, Takahashi Y, Srinives P (2016) Co-localization of QTLs for pod fiber content and pod shattering in F2 and backcross populations between yardlong bean and wild cowpea. *Molecular Breeding* 36:1-11

Suzuki M, Fujino K, Funatsuki H (2009) A major soybean QTL, qPDH1, controls pod dehiscence without marked morphological change. *Plant production science* 12:217-223

Ubi BE, Mignouna H, Thottappilly G (2000) Construction of a Genetic Linkage Map and QTL Analysis Using a Recombinant Inbred Population Derived from an Intersubspecific Cross of Cowpea: *Vigna unguiculta* (L.) Walp. *Breeding Science* 50:161-172

Vaillancourt R, Weeden N (1992) Chloroplast DNA polymorphism suggests Nigerian center of domestication for the cowpea, *Vigna unguiculata* (Leguminosae). *American Journal of Botany*:1194-1199

Vasconcelos EV, de Andrade Fonsêca AF, Pedrosa-Harand A, de Andrade Bortoleti KC, Benko-Iseppon AM, Da Costa AF, Brasileiro-Vidal AC (2015) Intra-and interchromosomal rearrangements between cowpea [*Vigna unguiculata* (L.) Walp.] and common bean (*Phaseolus vulgaris* L.) revealed by BAC-FISH. *Chromosome Research* 23:253-266

Weber AL, Briggs WH, Rucker J, Baltazar BM, de Jesús Sánchez-Gonzalez J, Feng P, Buckler ES, Doebley J (2008) The genetic architecture of complex traits in teosinte (*Zea mays* ssp. parviglumis): new evidence from association mapping. *Genetics* 180:1221-1232

Xu P, Hu T, Yang Y, Wu X, Wang B, Liu Y, Qin D, Ehlers J, Close T, Lu Z (2011) Mapping genes governing flower and seedcoat color in asparagus bean (*Vigna unguiculata* ssp. sesquipedalis) based on single nucleotide polymorphism and simple sequence repeat markers. *HortScience* 46:1102-1104

Xu P, Wu X, Muñoz-Amatriaín M, Wang B, Wu X, Hu Y, Huynh BL, Close TJ, Roberts PA, Zhou W (2017) Genomic regions, cellular components and gene regulatory basis underlying pod length variations in cowpea (*V. unguiculata* L. Walp). *Plant biotechnology journal* 15:547-557

Chapter 1: Identification of QTL Controlling Domestication-Related Traits in Cowpea [*Vigna unguiculata* (L.) Walp]

Abstract

Cowpea (*Vigna unguiculata* L. Walp) is a warm-season legume with a genetically diverse gene-pool composed of wild and cultivated forms. Cowpea domestication involved considerable phenotypic changes from the wild progenitor, including reduction of pod shattering, increased organ size, and changes in flowering time. Little is known about the genetic basis underlying these changes. In this study, 215 recombinant inbred lines derived from a cross between a cultivated and a wild cowpea accession were used to evaluate nine domestication-related traits (pod shattering, peduncle length, flower color, days to flowering, 100-seed weight, pod length, leaf length, leaf width and seed number per pod). A high-density genetic map containing 17,739 single nucleotide polymorphisms was constructed and used to identify 16 quantitative trait loci (QTL) for these nine traits. Based on annotations of the cowpea reference genome, genes within these regions are reported. Four regions with clusters of QTLs were identified, including one on chromosome 8 related to increased organ size. This study provides new knowledge of the genomic regions controlling domestication-related traits in cowpea as well as candidate genes underlying those QTLs. This information can help to exploit wild relatives in cowpea breeding programs.

Introduction

Plant domestication is considered to be the key technological element of a transition to agriculture (Gepts 2014). Domesticated crop plants differ from their wild progenitors in various characters collectively termed the "domestication syndrome" (Hammer 1984). Some of the most important syndrome traits include loss of seed dispersal, early flowering, increased size of the edible organs, and changes in plant architecture and organ coloration (Abbo et al. 2014), and are the outcome of selection for adaptation to cultivated environments (Gepts 2010). The genetic control of domestication-related traits (DRTs) has been investigated in numerous crop species, including legumes, mainly by quantitative trait loci (QTL) mapping, using crosses between cultivated and wild accessions (Doebley et al. 2006; Frary et al. 2000; Isemura et al. 2007; Kaga et al. 2008; Koinange et al. 1996; Liu et al. 2007). Increased knowledge of the genomic regions controlling DRTs is valuable to exploit wild germplasm efficiently for improving resistance to biotic and abiotic stresses and overall yield.

Cowpea (*Vigna unguiculata* L. Walp) is one of the main sources of dietary protein and folic acid for millions of people in sub-Saharan Africa. Cowpea is a diploid (2n= 22) with a genetically diverse gene-pool. The wild gene-pool is composed of both perennial and annual types, while cultivated forms are all annual. There are several cultivar groups, of which the most important are subsp. *unguiculata* (grain-type cowpea) and subsp. *sesquipedalis* (yardlong bean) (Maréchal et al. 1978). *V. unguiculata* subsp. *dekindtiana*, a perennial form that only exists in Africa, is considered the wild progenitor of cowpea (Lush

and Evans 1981). However, there is a lack of consensus on where in Africa cowpea domestication occurred. Several domestication locales have been proposed including West and Southeast Africa (Ba et al. 2004; Faris 1965; Rawal 1975; Steele and Mehra 1980; Vaillancourt and Weeden 1992). Domestication of cowpea has, in general, resulted in a determinate growth habit, increased pod and seed size, early flowering, and reduction of pod shattering. Cultivated cowpea also shows a wide range of flower and seed coat colors, whereas wild cowpeas typically have purple flowers and dark mottled seed coats. Only a few studies have identified QTL controlling DRTs in cowpea. Andargie et al. (2014) reported QTL for six DRTs including pod shattering, days to flowering, ovule number and growth habit. In subsp. *sesquipedalis*, QTL for DRTs including pod shattering, pod length and flower color have been also reported (Kongjaimun et al. 2012; Suanum et al. 2016; Xu et al. 2011; Xu et al. 2017). However, to date no specific gene underlying any domestication trait has been reported in cowpea.

Currently available genomic resources, which include an annotated reference genome (Lonardi et al. 2019) and a Cowpea iSelect Consortium Array (Muñoz-Amatriaín et al. 2017), have enabled new insights into the genetics of domestication of cowpea. In the present study, we investigated the genetic control of nine DRTs in cowpea including pod shattering, seed size, and flowering time using a population of 215 recombinant inbred lines (RILs) derived from a cross between a cultivated and a wild cowpea. The Cowpea iSelect Consortium Array, including 51,128 single nucleotide polymorphism (SNPs) markers (Muñoz-Amatriaín et al. 2017), was used to genotype this population, which allowed for high density genetic mapping. Using the annotated reference genome, we

report genes within QTL regions that control DRTs. This study provides new insights into the genetic control of domestication-related traits in cowpea and can help to utilize more efficiently wild germplasm in breeding programs.

Materials and Methods

Plant material and growth conditions

A biparental mapping population of 215 F8 RILs derived from a cross between IT99K-573-1-1 and TVNu-1158, cultivated and wild cowpea accessions, respectively, was used for linkage mapping and QTL analysis. IT99K-573-1-1 is an early-maturing, white-seeded, high yielding and Striga resistant variety that was released in Nigeria under the name SAMPEA 14. TVNu-1158 is small seeded and has a perennial growth habit. It is cross-compatible with cultivated cowpea, although the F1 and subsequent generations showed partial sterility.

The parents and the F8 population were grown in pots filled with 5.0 kg topsoil and placed in a screen house at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria (latitude 7° 30'N and longitude 3° 54'E, elevation 240 masl). Five seeds of each RIL were sown per pot and then thinned to two plant per pot when they were well established. The pots were watered regularly.

Phenotypic data collection and analyses

A total of nine DRTs (pod shattering, peduncle length, flower color, days to flowering, seed weight, pod length, leaf length, leaf width, and seed number per pod) segregating in the population were evaluated. Phenotypic data for pod shattering were collected by visual scoring, with scores of 0 = "no shattering", and 1 = "pods opened and twisted". Peduncle length was determined by measuring the distance from the point of peduncle attachment to

the node on the stem to where the first flower bud emerged. Five peduncles were measured on each plant beginning with the lowest peduncle. Flower color was evaluated by visual scoring with 0 = “white”, and 1= “purple”. Flowering time was scored as the number of days from date of planting to when the first flower opened. Seed weight was determined by the weight in grams of 100 seeds. The length (from base to the tip) and width (at the widest point) of the first leaf (opposite in arrangement) following seedling emergence was determined using a measuring tape. Pod length was determined by measuring the length of the first five pods per plant. Seed number per pod was evaluated by counting the number of seeds per 10 pods for each plant.

For the seven non-binary traits, the mean, standard error, and range were calculated as well as the degrees of skewness. The frequency distribution of the seven traits was determined. The segregation pattern for pod shattering and flower color was analyzed by chi-square test for goodness-of-fit.

Genotyping

Young leaves from each RIL and the parents were collected and placed into a ziplock bag containing silica gel packs for desiccation. Total genomic DNA of each line and parents was extracted from dried leaves using Plant DNeasy (Qiagen, Germany), quantified using Quant-IT dsDNA Assay Kit (Thermo Fisher Scientific, USA), and the concentration adjusted to 80 ng/µl. Genotyping was performed at the University of Southern California (Los Angeles, CA, USA) using the Cowpea iSelect Consortium Array, which includes 51,128 SNPs (Muñoz-Amatriaín et al. 2017).

Linkage map construction

SNPs were called using the GenomeStudio software V.2011.1 (Illumina, Inc., San Diego, CA, USA). Data curation was performed by removing SNPs with more than 20% missing or heterozygous calls. Individuals were then plotted by missing calls and by heterozygous calls, and outliers were eliminated from further analysis. Lines carrying non-parental alleles and duplicated lines were also removed. Only SNPs that were polymorphic in both the parents and the RIL population and had minor allele frequencies (MAFs) > 0.25 were used for linkage mapping. MSTmap (Wu et al. 2008; http://www.mstmap.org/) was used for genetic map construction, with the following parameters: grouping LOD criteria = 10; population type = DH (doubled haploid); no mapping size threshold = 2; no mapping distance threshold: 10 cM; try to detect genotyping errors = no; and genetic mapping function = kosambi. The chromosomes were numbered and oriented according to the recently developed cowpea pseudomolecules (Lonardi et al. 2019). Since the use of DH function inflated the cM distance for a RIL population, cM map distances were divided by two to correct for the extra round of effective recombination occurring in a RIL population compared to a DH population.

QTL analysis and identification of candidate genes

QTL analysis was performed using a linear mixed model implemented in R and described by Xu (2013). In this approach, the background effect is captured via a polygene and a marker inferred kinship matrix (Wei and Xu 2016; Xu et al. 2013). To reserve a window around the test marker, markers around a ± 2 cM window at the tested marker were

excluded from the kinship matrix. The effect of each marker was estimated as a fixed effect and tested using the Wald test statistic (squared effect divided by the variance of the estimated effect). Under the null model, the Wald test follows a chi-square distribution with one degree of freedom, from which a p-value was calculated for each marker. SNP markers of the entire genome were scanned and a test statistic, $-\log_{10}$ (*P*-value), profile was plotted. A genome-wide critical value was calculated with a modified Bonferroni correction using a trait specific "effective number of markers or effective degrees of freedom" as the denominator (Wang et al. 2015). The effective degrees of freedom was defined as

$m0 = \frac{\sum(Wk-1)}{Wk}$, where Wk is the Wald test statistic for SNP k. The trait specific Bonferroni corrected critical value was $-\log_{10}(0.05/m0)$. A SNP was declared as significant if its $-\log_{10}$ (*P*-value) was larger than $-\log_{10}$ (0.05/m0). For each significant SNP, an estimate of the percentage of phenotypic variation was calculated. The proportion of phenotypic variance contributed by each SNP was calculated using the following formula:

$\frac{var(X)a^2}{var(y)}$, where X is a variable holding the genotype code (+1,-1) for SNP k, var(X) is the variance of variable X, a is the estimated effect for SNP k and var(y) is the total phenotypic variance of the trait under study.

The physical region of each QTL was determined from the cowpea genome V.1.0 (Lonardi et al. 2019) and used to identify candidate genes underlying the QTL interval. Then, we focused on one region containing four QTLs associated with increased organ size. The four QTLs span the genomic region Vu08:36035190-38248903, which contains 313 annotated genes. The corresponding syntenic segment in Phaseolus vulgaris (Chr08: 57594596-59622008) was determined using the legumeinfo.org instance of the Genome Context Viewer (GCV) (Cleary and Farmer 2017). This region contained 289 Phaseolus genes, of which only one (*Phvul.008G285800*) was present in the intersection with a list of genes associated with domestication reported in Schmutz et al. (2014) using functions of cowpeamine and legumemine (https://mines.legumeinfo.org). The cowpea syntelog of that gene is *Vigun08g217000*, according to the genomic segment alignment provided by the GCV.

Results

Phenotypic variation in the population

Nine traits (pod shattering, peduncle length, flower color, days to flowering, seed weight, pod length, leaf length, leaf width, and seed number per pod) related to domestication in cowpea that differed between the parents were evaluated in the population. The mean of 100-seed weight, seed number per pod, primary leaf length and width, and pod length of the cultivated parent (IT99K-573-1-1) were all higher than those of the wild parent (TVNu-1158), whereas the mean of peduncle length and days to flowering were higher in the wild parent than the cultivated parent (Table 1.1). The segregation pattern of pod shattering fit the expected Mendelian ratio of 1:1 (P value = 0.25), while significant segregation distortion was observed for flower color (P value = 0.0035). The frequency distribution of the seven remaining traits are shown in Supplementary Fig.S1. Transgressive segregation was observed for peduncle length, days to flowering, seed number per pod, primary leaf length, and pod length (Table 1.1).

Development of the cultivated x wild genetic map

A total of 17,739 polymorphic SNPs and 170 RILs which passed quality controls (see Materials and Methods) were used for genetic map construction. All SNPs were mapped into 1,825 bins on 12 linkage groups (LGs). LGs were named and oriented based on cowpea pseudomolecules (www.phytozome.net), from Vu01 to Vu11. Note that this numbering of LGs differs from the one used in previous cowpea genetic maps (Lucas et al. 2011; Muchero et al. 2009; Muñoz-Amatriaín et al. 2017). A cross reference to the previous

nomenclature is included in Table 1.2. One chromosome (Vu03) was separated into two LGs, which were joined based on the consensus genetic map of Muñoz-Amatriaín et al. (2017). Based on the consensus map, the gap was estimated at 18 cM and was attributed to the absence of polymorphisms between the parents and to the presence of highly distorted markers. The genetic map covered 1,026.03 cM, with an average density of one marker bin per 1.8 cM, and 9.72 SNPs per bin. The length of the LGs varied from 15.25 (Vu04) to 139.72 cM (Vu03) (Table 1.2). This genetic map contains the highest number of SNPs in an individual cowpea map to date. Compared to the previously published genetic maps of cowpea constructed with SNP markers (Lucas et al. 2011; Muchero et al. 2009; Muñoz-Amatriaín et al. 2017), we observed a much shorter Vu04 (old LG11). The difference may be due to genomic divergences between the cultivated and wild parents leading to disturbed chromosome pairing which reduces or suppresses recombination in affected regions and/or structural chromosomal rearrangement. The IT99K-573-1-1 x TVNu-1158 genetic map can be found in Supplementary Table S1.1.

Identification of domestication-related QTL and candidate genes

A total of 16 QTL were identified for the nine traits using a mixed model for QTL mapping (Xu 2013) implemented in R. The 16 domestication-related QTL were distributed on seven of the eleven chromosomes of cowpea (Figure 1.1; Table 1.3) and their $-\log_{10}$ (*P*-value) ranged from 5.15 for seed number per pod to 20.00 for peduncle length, flower color, and leaf width (Table 1.3). The significance levels ranged from 4.57 to 4.98 depending on the

trait (Supplementary Table S1.2). The percentage of phenotypic variation for the identified QTL ranged from 18.32% for seed number per pod to 85.65% for flower color (Table 1.3). We identified four regions showing QTL clustering for domestication traits, including one region on Vu08 where four QTL related to increased organ size (seed weight, pod length, leaf length and leaf width) were mapped. Further studies would be required to determine if this clustering of domestication-related QTL results from pleiotropic effects or tightly linked QTL.

For all the traits and QTL below, additional information on SNP positions within the pseudomolecules, including distance to genes, is available through phytozome (www.phytozome.net).

*Pod shattering***.** Two significant QTL were detected for pod shattering, one each on Vu03 and Vu05 (Figure 1.1; Table 1.3). These QTL, named *CPshat3* and *CPshat5*, explained 37.69 % and 30.27 % of the phenotypic variation, respectively. *CPshat3* spanned 6.1 cM corresponding to ~ 3.4 Mb on the cowpea pseudomolecules (Lonardi et al. 2019) and contained 268 genes, while *CPshat5* spanned 7.74 cM corresponding to ~ 1.60 Mb and contained 204 genes (Supplementary Table S1.3). Among the genes underlying the main QTL (*CPshat3*) two encoding transcription factors stand out in the context of pod shattering, *Vigun03g306000* and *Vigun03g302600* (see Discussion). Among the annotated genes underlying the *CPshat5* region, *Vigun05g273500*, annotated as Myb domain protein 26, is considered to be an especially interesting candidate (see Discussion).

*Peduncle length***.** QTL analysis detected one major QTL on Vu05, *CPedl5*. This QTL accounted for 71.83% of the phenotypic variation and it spanned a 15.37 cM region corresponding to ~ 4.53 Mb (Figure 1.1; Table 1.3). A total of 379 annotated genes were identified in the *CPedl5* interval (Supplementary Table S1.3). Notably, three genes encode auxin-responsive GH3 family proteins (*Vigun05g201700*, *Vigun05g217600* and *Vigun05g223100*) were located in the interval, and members of the GH3 family have been found to influence organ elongation (Nakazawa et al. 2001; Takase et al. 2004) (see Discussion).

*Flower pigmentation***.** A major-effect QTL controlling purple flower (*CFcol7*) was detected in a 64 cM region (~ 4.56 Mb) on Vu07 (Figure 1.1; Table 1.3). This QTL explained 85.65% of the phenotypic variation and contained 254 annotated genes (Supplementary Table S1.3). The transcription factor TRANSPARENT TESTA8 (TT8, *Vigun07g110700*), which contains a basic helix-loop-helix domain, was identified in this QTL region (see Discussion).

Days to flowering. Two significant QTL related to days to flowering were detected, one each on Vu05 and Vu09. The QTL on Vu05 (*CFt5*) explained 20% of the phenotypic variation, while *CFt9* on Vu09 explained 79.3% of the phenotypic variation (Figure 1.1; Table 1.3). *CFt5* spans 7 cM which correspond to ~ 6.64 Kb, while *CFt9* maps to a 16 cM region corresponding to ~ 3.86 Mb. A total of 86 genes were identified in the *CFt5* region, while 299 genes were identified underlying *CFt9* (Supplementary Table S1.3). Among the annotated genes in the *CFt9* region, a phytochrome E photoreceptor (*Vigun09g050600*)

and a transcription factor TCP 18 (*Vigun09g062200*) were found. No genes previously shown to have a role in flowering were identified in the *CFt5* region (see Discussion).

100-seed weight. Three QTL for seed weight were detected, one each on Vu01, Vu06 and Vu08. The QTL with the highest contribution to the trait (*CSw8*) was located on Vu08 and explained 36.87% of the phenotypic variation, while *CSw1* and *CSw6* explained 19.85% and 21.48% of the phenotypic variation, respectively (Table 1.3). *Csw8* spans 5.45 cM (~ 1.61 Mb) and contains 225 genes, while *Csw1* and *Csw6* span 4.3 cM (~1.60 Mb) and 4.04 cM (~1.12 Mb) and contain 206 and 160 annotated genes, respectively (Supplementary Table S1.3). Among the many genes in these QTL regions, several were involved in carbohydrate metabolism, including UDP-glycosyltransferases and cellulose synthases were identified (Supplementary Table S1.3).

Pod length. Pod length was analyzed as a measure of increase in organ size, and two QTL were identified, one each on Vu03 and Vu08. These QTL explained 23.98% and 46.08% of the phenotypic variation and spanned 0.47 cM (~ 84 Kb) and 6.92 cM (~ 2.1 Mb), respectively. Nine genes including an AGAMOUS-like8 gene (*Vigun03g343500*), which is known to mediate cell differentiation during fruit development in Arabidopsis (Gu et al. 1998), were identified underlying *CPodl3*, while 309 genes were found in the *CPodl8* interval. *CPodl8* maps to the same genomic region as *CSw8* (Figure 1.1; Table 1.3) and includes genes encoding cellulose synthases, UDP-glycosyltransferases and a cluster of pectin lyases (Supplementary Table S1.3), all involved in carbohydrate metabolism (see Discussion).

Leaf length and width. The leaf is the main organ for photosynthesis in cowpea and one QTL for leaf length (*CLl8*) and two QTL for leaf width (*CLw1*, *CLw8*) were identified in this population (Figure 1.1; Table 1.3). *CLl8* explained 38.24% of the phenotypic variation and spanned 1.85 cM (~ 9.9 Kb). *CLw1* accounted for 63.28% of the phenotypic variation and spanned 19.1 cM (~ 3.61 Mb), while *CLw8* explained 34.75% of the phenotypic variation and spanned 2.45 cM (~1.08 Mb). Regarding leaf length, a total of 142 genes were identified in the physical region of *CLl8*. For leaf width, over 300 genes were identified in *CLw1* interval, while 148 genes were identified underlying *CLw8* (Supplementary Table S1.3). *CLl8* and *CLw8* map to the same genomic region as *CSw8* and *CPodl8* (Figure 1.1; Table 1.3). Genes present in *Csw8* and *CPodl8* are also contained in these two QTL, including proteins of the pectin lyase-like superfamily and cellulose synthases (Supplementary Table S1.3 and Discussion).

Seed number per pod. Two QTL were detected, one each on Vu05 and Vu09, accounting for 18.32% and 21.09% of the phenotypic variation, respectively, spanned 1.54 cM (~3.71 Kb) and 2.81 cM (~4.94 Kb), respectively. The total number of genes underlying *CSp5* and *CSp9* was only 40 and 33, respectively. One of the genes underlying *CSp9* encodes a transcription factor TCP5, which has been associated with ovule development in Arabidopsis (Wei et al. 2015). No gene was identified in the *CSp5* region with known functions related to seed number per pod (Supplementary Table S1.3 and Discussion).

Discussion

Through the domestication process, cowpea underwent many phenotypic changes compared to its wild progenitor. In this study, we have focused on nine traits that differ between wild and cultivated cowpea and identified 16 QTL for trait determination distributed over seven chromosomes. For each trait, only one or two major QTL were identified except for seed weight, for which three QTL were detected. These results are consistent with previous studies suggesting that, in most crops, domestication related traits seem to be controlled by a small number of QTL with large phenotypic effects (Burger et al. 2008; Gross and Olsen 2010).

Among the domestication traits considered in this study, loss of pod shattering and time to flowering are the most relevant for cowpea breeding, together with increase of organ size. The pod shattering habit causes pre-harvest yield losses. Two major QTL were identified for pod shattering, one each on Vu03 and Vu05. In previous studies in cowpea, Suanum et al. (2016) reported one major QTL and one minor QTL for pod shattering, while Andargie et al. (2014) identified only one QTL explaining 10.8% of the phenotypic variation. While it was not possible to compare the QTL identified by Suanum et al. (2016) because of unavailability of marker sequences, BLAST searches of SSR markers from the Andargie et al. (2014) study to the reference genome sequence (Lonardi et al. 2019) revealed that the QTL identified in that study is located in a different chromosome compared to those reported here. This suggests the QTL identified in this study are novel.

Genes involved in seed dispersal have been cloned in dicots including soybean (Dong et al. 2014), Brassica (Tao et al. 2017) and Arabidopsis (Liljegren et al. 2000). For *CPshat3*, the major pod shattering QTL, we identified a gene *Vigun03g306000*, which encodes a NAC domain transcription factor (NAC007) involved in secondary cell wall biosynthesis (Wang et al. 2011). In soybean, the NAC family gene *SHATTERING1-5* has been found to activate secondary wall biosynthesis affecting pod shattering resistance (Dong et al. 2014). In addition, a C2H2 zinc finger protein (*Vigun03g302600*) was identified in this QTL region. A member of the C2H2 zinc finger family of proteins has been shown to enhance silique shattering resistance in Brassica (Tao et al. 2017). In the *CPshat5* region we identified *Vigun05g273500*, a gene annotated as Myb domain protein 26. *Vigun05g273500* is an orthologue of *AT3G13890.1*, which acts upstream of the lignin biosynthesis pathway and has been shown to be key for anther dehiscence by regulating the development of secondary thickening in the endothecium in Arabidopsis (Yang et al. 2007).

Time to flowering is one of the most important agronomic traits that plays a key role in the adaptation of a variety to specific agro-ecological areas. Two major QTL associated with days to flowering have been identified in this work. As with pod shattering, none of these QTL seem to coincide with the main QTL reported by Andargie et al. (2014). The main QTL for days to flowering (*CFt9*) in the present study mapped to Vu09, with the wild accession alleles conferring late flowering. There are two genes in the *CFt9* region with functions related to time to flowering. One is a phytochrome E (*Vigun09g050600*), one of the photoreceptors perciving red/far-red light ratio and influencing flowering time (Devlin et al. 1998). *Vigun09g050600* is an ortholog of the adzuki bean *Vigan.02G285600.01*,

which is one of the candidate genes for the major photoperiod QTL *Flowering Date1* (Yamamoto et al. 2016). The other gene is *Vigun09g062200*, encoding a transcription factor TCP 18 which has been shown in Arabidopsis to repress flowering by interacting with the florigen proteins FLOWRING LOCUS T (FT) and TWIN SISTER OF FT (TSF) (Niwa et al. 2013).

Larger seeds play a major role in consumer preference, and larger leaves provide more surface than smaller leaves for the production of photosynthate. Four traits including 100-seed weight, pod length, primary leaf length and width were analyzed as a measure of increased organ size in the population. QTL identified for seed weight (*CSw8*), leaf length (*CLl8*), leaf width (*CLw8*), and pod length (*CPodl8*) mapped to the same region on Vu08. Multiple QTL located in this region suggest potential pleiotropy or clustering of genes controlling increased organ size, a fundamental target during domestication. In rice bean, QTL for leaf size were detected close to QTL controlling seed- and pod-size related traits (Isemura et al. 2010). Syntenic regions in the common bean genome were identified, the largest of which is located on common bean chromosome 8 (Pv08). That region contains a total of 289 common bean syntelogs, which were then compared with the list of common bean genes associated with domestication available from Schmutz et al. (2014). The intersection of these two lists contained only a single gene, *Phvul.008G285800*, a *P. vulgaris* candidate gene for increased seed size that corresponds to cowpea *Vigun08g217000*. This gene codes for a histidine kinase 2 that is expressed in several cowpea tissues including root, seed, pod and leaf (https://legumeinfo.org). The Arabidopsis ortholog *AHK2* (*AT5G35750.1*) is a cytokinin receptor that has been shown to regulate,

among other things, plant organ size (Bartrina et al. 2017; Riefler et al. 2006). *Vigun08g217000* is thus a candidate gene for further investigation. In addition to that a cluster of pectin lyase-like superfamily proteins, which are known to be involved in many biological processes including cellular metabolism (Cao 2012), was found in this hotspot region. Also, clusters of other genes with functions in carbohydrate metabolism such as UDP-glycosyltransferases and cellulose synthases were identified in this region. Further work, including fine mapping of this hotspot region, would be needed to elucidate the genetic control of increased organ size in cowpea.

Two additional QTL related to seed weight were also detected outside this region, one each on Vu01 and Vu06. One of these (*CSw6*) mapped to the same region as C*ss-4*, a seed size QTL identified from a cultivated by cultivated RIL population[43]. Two other previous studies on cowpea reported QTL involved in seed weight using wild x cultivated population (Andargie et al. 2014; Fatokun et al. 1992). While it was not possible to compare QTL regions between this study and that of Fatokun et al. (1992) because of unavailability of marker sequences, none of these QTL seem to coincide with the ones reported by Andargie et al. (2014). As one of the major yield components, the number of seeds per pod is an important trait in cowpea breeding. Two QTL, *CSp5* and *CSp9*, were detected for this trait on Vu05 and Vu09, respectively. The main QTL (*CSp9*) accounted for 21.09% of the phenotypic variation and co-localized with the main QTL for days to flowering. The allele from the wild parent at *CSp9* conferred a higher number of seeds per pod, making it a good target for introgression into cultivated cowpea varieties. *Vigun09g060700* was identified in *Csp9* region. This gene is annotated as a transcription factor TCP5, which is involved in

ovule development in Arabidopsis (Wei et al. 2015). Since the number of seeds is determined among other things by the number of ovules (and fertile ovules) per ovary, this gene is a promising candidate for number of seeds per pod in cowpea.

Another major difference observed between the wild and the cultivated parent used in this study was the length of the peduncles. A single QTL with a high contribution to the phenotypic variation of the trait (71.83%) was identified on Vu05. Long peduncles in cowpea are desirable as they allow pods to be positioned above the canopy, a characteristic that reduces damage to pods by the pod borer *Maruca vitrata* and are also advantageous for harvesting of pods. Some genes involved in plant growth and development were found underlying *CPedl5,* including genes belonging to the auxin-responsive GH3 family. In cucumber, two genes belonging to this family (*Csa6G492310* and *Csa6G493310*) were identified as candidates for the main QTL for fruit peduncle length (Song et al. 2016).

The presence or absence of purple pigment in the flower is controlled by a single major QTL in this population. This is similar to a previous study in soybean, where one QTL for flower color was identified (Josie et al. 2007). The QTL controlling flower color was mapped to Vu07 and explained 85.65% of the phenotypic variation for the trait. An earlier study of cowpea suggested that a single gene controls flower color with purple flower color being dominant (Sangwan and Lodhi 1998). *Vigun07g110700,* involved in flavonoid biosynthesis, was identified in the QTL region. *Vigun07g110700* has sequence similarity to the Arabidopsis *TT8* gene (*AT4G09820.1*) encoding a basic helix-loop-helix domain protein involved in the control of flavonoid biosynthesis, whose final compounds include anthocyanins (Nesi et al. 2000). *Medicago truncatula TT8* (*MtTT8*) has also been found to

regulate anthocyanin and proanthocyanidin biosynthesis by interacting with other transcription factors and forming regulatory complexes (Li et al. 2016). Hence, *Vigun07g110700* is a strong candidate for the control of flower pigmentation in cowpea.

In summary, this study provides novel QTL for many DRTs as well as a link between genetic and physical maps. Thanks to the new genomic resources available for cowpea, which include a reference genome sequence of IT97-499-35 (Lonardi et al. 2019), we were able to estimate physical sizes for all QTL identified and to determine, for the first time in cowpea, candidate genes underlying QTL controlling DRTs. The results of this study provide a basis for further fine mapping of genes involved in cowpea domestication and a genetic foundation for the utilization and exploitation of wild relatives in cowpea breeding programs.

References

Abbo S, van-Oss RP, Gopher A, Saranga Y, Ofner I, Peleg Z (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends in Plant Science* 19:351-360

Andargie M, Pasquet RS, Gowda BS, Muluvi GM, Timko MP (2014) Molecular mapping of QTLs for domestication-related traits in cowpea (*V. unguiculata* (L.) Walp.). *Euphytica* 200:401-412

Ba FS, Pasquet RS, Gepts P (2004) Genetic diversity in cowpea [*Vigna unguiculata* (L.) Walp.] as revealed by RAPD markers. *Genetic Resources and Crop Evolution* 51:539-550

Bartrina I, Jensen H, Novak O, Strnad M, Werner T, Schmülling T (2017) Gain-of-function mutants of the cytokinin receptors AHK2 and AHK3 regulate plant organ size, flowering time and plant longevity. *Plant physiology*:pp. 01903.02016

Burger JC, Chapman MA, Burke JM (2008) Molecular insights into the evolution of crop plants. *American Journal of Botany* 95:113-122

Cao J (2012) The pectin lyases in Arabidopsis thaliana: evolution, selection and expression profiles. *PloS one* 7:e46944

Cleary A, Farmer A (2017) Genome Context Viewer: visual exploration of multiple annotated genomes using microsynteny. *Bioinformatics* 34:1562-1564

Devlin PF, Patel SR, Whitelam GC (1998) Phytochrome E influences internode elongation and flowering time in Arabidopsis. *The Plant Cell* 10:1479-1487

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309-1321

Dong Y, Yang X, Liu J, Wang B-H, Liu B-L, Wang Y-Z (2014) Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nature communications* 5

Faris D (1965) The origin and evolution of the cultivated forms of Vigna sinensis. *Canadian journal of genetics and cytology* 7:433-452

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841-846

Frary A, Nesbitt TC, Frary A, Grandillo S, Van Der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB (2000) fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289:85-88
Gepts P (2010) Crop domestication as a long-term selection experiment. *Plant breeding reviews* 24:1-44

Gepts P (2014) The contribution of genetic and genomic approaches to plant domestication studies. *Current opinion in plant biology* 18:51-59

Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. *Trends in plant science* 15:529-537

Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R (1998) The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development* 125:1509-1517

Hammer K (1984) Das domestikationssyndrom. *Die Kulturpflanze* 32:11-34

Isemura T, Kaga A, Konishi S, Ando T, Tomooka N, Han OK, Vaughan DA (2007) Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. *Annals of Botany* 100:1053-1071

Isemura T, Kaga A, Tomooka N, Shimizu T, Vaughan DA (2010) The genetics of domestication of rice bean, *Vigna umbellata*. *Annals of botany* 106:927-944

Josie J, Alcivar A, Rainho J, Kassem MA (2007) Genomic regions containing QTL for plant height, internodes length, and flower color in soybean [*Glycine max* (L.) Merr]. *Bios* 78:119-126

Kaga A, Isemura T, Tomooka N, Vaughan DA (2008) The genetics of domestication of the azuki bean (*Vigna angularis*). *Genetics* 178:1013-1036

Koinange EM, Singh SP, Gepts P (1996) Genetic control of the domestication syndrome in common bean. *Crop Science* 36:1037-1045

Kongjaimun A, Kaga A, Tomooka N, Somta P, Vaughan DA, Srinives P (2012) The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. unguiculata cv.-gr. sesquipedalis. *Annals of botany* 109:1185-1200

Li P, Chen B, Zhang G, Chen L, Dong Q, Wen J, Mysore KS, Zhao J (2016) Regulation of anthocyanin and proanthocyanidin biosynthesis by Medicago truncatula bHLH transcription factor MtTT8. *New Phytologist* 210:905-921

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000) SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature* 404:766-770

Liu B, Fujita T, Yan Z-H, Sakamoto S, Xu D, Abe J (2007) QTL mapping of domestication-related traits in soybean (*Glycine max*). *Annals of Botany* 100:1027-1038

Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo M-C, Alhakami H, Ounit R, Hasan AM, Verdier J, Roberts PA, Santos JRP, Ndeve A, Doležel J, Vrána J, Hokin SA, Farmer AD, Cannon SB, Close TJ The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *The Plant Journal* https://doi.org/10.1111/tpj.14349

Lucas MR, Diop N-N, Wanamaker S, Ehlers JD, Roberts PA, Close TJ (2011) Cowpea–Soybean Synteny Clarified through an Improved Genetic Map. *The Plant Genome Journal* 4:218

Lush W, Evans L (1981) The domestication and improvement of cowpeas (*Vigna unguiculata* (L.) Walp.). *Euphytica* 30:579-587

Maréchal R, Mascherpa J-M, Stainier F (1978) Etude taxonomique d'un groupe complexe d'especes des genres Phaseolus et Vigna (Papilionaceae) sur la base donnees morphologiques et polliniques, traitees par l'analyse informatique. *Boissiera* 28:2-7

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ (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* 106:18159-18164

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O, Bozdag S, Cisse N, Drabo I, Ehlers JD, Farmer A, Fatokun C, Gu YQ, Guo Y-N, Huynh B-L, Jackson SA, Kusi F, Lawley CT, Lucas MR, Ma Y, Timko MP, Wu J, You F, Barkley NA, Roberts PA, Lonardi S, Close TJ (2017) Genome resources for climate-resilient cowpea, an essential crop for food security. *The Plant Journal* 89:1042-1054

Nakazawa M, Yabe N, Ichikawa T, Yamamoto YY, Yoshizumi T, Hasunuma K, Matsui M (2001) DFL1, an auxin-responsive GH3 gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. *The Plant Journal* 25:213-221

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. *The Plant Cell* 12:1863-1878

Niwa M, Daimon Y, Kurotani K-i, Higo A, Pruneda-Paz JL, Breton G, Mitsuda N, Kay SA, Ohme-Takagi M, Endo M (2013) BRANCHED1 interacts with FLOWERING LOCUS T to repress the floral transition of the axillary meristems in Arabidopsis. *The Plant Cell* 25:1228-1242

Rawal KM (1975) Natural hybridization among wild, weedy and cultivated *Vigna unguiculata* (L.) Walp. *Euphytica* 24:699-707

Riefler M, Novak O, Strnad M, Schmülling T (2006) Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *The Plant Cell* 18:40-54

Sangwan R, Lodhi G (1998) Inheritance of flower and pod colour in cowpea (Vigna unguiculata L. Walp.). *Euphytica* 102:191-193

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nature genetics* 46:707-713

Song Z-C, Miao H, Zhang S, Wang Y, Zhang S-P, Gu X-F (2016) Genetic Analysis and QTL Mapping of Fruit Peduncle Length in Cucumber (*Cucumis sativus* L.). *PloS one* 11:e0167845

Steele W, Mehra K (1980) Structure, evolution and adaptation to farming systems and environments in Vigna. Advance in Legume Sci Royal Botanical Gardens UK

Suanum W, Somta P, Kongjaimun A, Yimram T, Kaga A, Tomooka N, Takahashi Y, Srinives P (2016) Co-localization of QTLs for pod fiber content and pod shattering in F2 and backcross populations between yardlong bean and wild cowpea. *Molecular Breeding* 36:1-11

Takase T, Nakazawa M, Ishikawa A, Kawashima M, Ichikawa T, Takahashi N, Shimada H, Manabe K, Matsui M (2004) ydk1-D, an auxin-responsive GH3 mutant that is involved in hypocotyl and root elongation. *The Plant Journal* 37:471-483

Tao Z, Huang Y, Zhang L, Wang X, Liu G, Wang H (2017) BnLATE, a Cys2/His2-Type Zinc-Finger Protein, Enhances Silique Shattering Resistance by Negatively Regulating Lignin Accumulation in the Silique Walls of *Brassica napus*. *PloS one* 12:e0168046

Vaillancourt R, Weeden N (1992) Chloroplast DNA polymorphism suggests Nigerian center of domestication for the cowpea, *Vigna unguiculata* (Leguminosae). *American Journal of Botany*:1194-1199

Wang H, Zhao Q, Chen F, Wang M, Dixon RA (2011) NAC domain function and transcriptional control of a secondary cell wall master switch. *The Plant Journal* 68:1104-1114

Wang Q, Wei J, Pan Y, Xu S (2015) An efficient empirical Bayes method for genomewide association studies. *Journal of Animal Breeding and Genetics*

Wei B, Zhang J, Pang C, Yu H, Guo D, Jiang H, Ding M, Chen Z, Tao Q, Gu H (2015) The molecular mechanism of SPOROCYTELESS/NOZZLE in controlling Arabidopsis ovule development. *Cell research* 25:121

Wei J, Xu S (2016) A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. *Genetics* 202:471-486
Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* 4:e1000212

Xu P, Hu T, Yang Y, Wu X, Wang B, Liu Y, Qin D, Ehlers J, Close T, Lu Z (2011) Mapping genes governing flower and seedcoat color in asparagus bean (*Vigna unguiculata* ssp. sesquipedalis) based on single nucleotide polymorphism and simple sequence repeat markers. *HortScience* 46:1102-1104

Xu P, Wu X, Muñoz-Amatriaín M, Wang B, Wu X, Hu Y, Huynh BL, Close TJ, Roberts PA, Zhou W (2017) Genomic regions, cellular components and gene regulatory basis underlying pod length variations in cowpea (*V. unguiculata* L. Walp). *Plant biotechnology journal* 15:547-557

Xu P, Wu X, Wang B, Hu T, Lu Z, Liu Y, Qin D, Wang S, Li G (2013) QTL mapping and epistatic interaction analysis in asparagus bean for several characterized and novel horticulturally important traits. *BMC genetics* 14:4

Xu S (2013) Mapping quantitative trait loci by controlling polygenic background effects. *Genetics* 195:1209-1222

Yamamoto H, Horiuchi Y, Ogura R, Sakai H, Sato H, Kato K (2016) Identification and molecular mapping of Flowering Date1 (FD1), a major photoperiod insensitivity gene in the adzuki bean (*Vigna angularis*). *Plant Breeding* 135:714-720

Yang C, Xu Z, Song J, Conner K, Barrena GV, Wilson ZA (2007) Arabidopsis MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *The Plant Cell* 19:534-548

Table 1.1: Mean, standard error, range, and skewness values for the parental lines and the F8 RIL population

Trait	**IT99K-573-1-1**	**TVNu-1158**	**RIL population**		
			Mean +/- SE	*Range*	*Skewness*
Pod shattering	0	1	-	-	-
Peduncle length (cm)	15.70 +/- 0.98	20.73 +/- 1.33	16.76 +/- 0.41	5.54-33.8	0.70
Flower color	0	1	-	-	-
Days to flowering	41.00 ± 1.00	70.00 +/- 2.00	47 +/- 1	35-88	1.85
100-seed weight (g)	19.6 +/- 0.36	0.88 +/- 0.01	5.07 +/- 0.17	1.77-11.89	0.76
Pod length (cm)	17.89 +/- 1.13	6.82 +/- 0.36	16.75 +/- 0.40	5.07-20.9	0.88
Leaf length (cm)	7.2 +/- 0.33	2.76 +/- 0.12	4.85 +/- 0.08	2.8-8	0.23
Leaf width (cm)	3.94 +/- 0.26	1.39 +/- 0.11	2.27 +/- 0.05	1-4	0.65
Number of seeds/pod	11 +/- 1	9 +/- 1	9 +/- 1	2-16	0.06

Table 1.2: Number of SNPs, and length of each linkage group in the cultivated x wild cowpea genetic map

Linkage Group (LG, old)	Chromosome	SNPs	Bins	cM
4	Vu01	1742	132	65.31
7	Vu02	1368	140	98.62
3	Vu03	2780	280	157.61
11	Vu04	117	24	15.25
1	Vu05	1740	205	114.62
6	Vu06	1676	150	78.29
2	Vu07	1968	228	134.15
5	Vu08	1607	159	81.59
8	Vu09	1774	189	100.37
10	Vu10	1416	144	77.37
9	Vu11	1551	174	102.85
Total	**Total**	**17739**	**1825**	**1026.03**

Table 1.3: QTL for the domestication-related traits identified by the linear mixed model analysis and their map position [a] QTL name is designated as follow: "C" to indicate

Trait	**QTL[a]**	**Peak SNP**	**Chr. Vu**	**Position**	**(-) Log10(P)**	**QTL region (cM)**	**% Phenotypic variation**	**Effect[b]**
Pod shattering	*CPshat3*	2_45094	03	104.84	8.17	100.57–106.67	37.69	−0.31
	CPshat5	2_23044	05	100.02	7.28	96.11–103.85	30.27	−0.28
Peduncle length	*CPedl5*	2_00251	05	66.23	20	60.97–76.34	71.83	−4.50
Flower color	*CFcol7*	2_24259	07	65.89	20	64.39–72.43	85.65	−4.16
Days to flowering	*CFt5*	2_05332	05	5.77	5.51	0.93–7.93	20.00	4.13
	CFt9	2_03945	09	21.79	20	13.82–30.02	79.30	−8.46
100-seeds weight	*CSw1*	2_17042	01	46.62	6.85	45.15–49.45	19.85	1.03
	CSw6	2_55420	06	72.16	6.66	69.26–73.30	21.48	1.06
	CSw8	2_05809	08	78.57	11.86	76.14–81.59	36.87	1.39
Pod length	*CPodl3*	2_49508	03	116.49	5.19	116.02–116.49	23.98	1.08
	CPodl8	2_00195	08	78.17	13.04	74.67–81.59	46.08	1.52
Leaf length	*CLl8*	2_09764	08	78.87	6.98	78.17–80.02	38.24	0.69
Leaf width	*CLw1*	2_19941	01	34.13	20	30.35–49.45	63.28	0.54
	CLw8	2_21200	08	78.87	7.09	76.73–79.18	34.75	0.40
N° seed per pod	*CSp5*	2_07573	05	11.37	5.15	10.57–12.11	18.32	1.07
	CSp9	2_19332	09	24.73	5.77	22.51–25.32	21.09	−1.18

cowpea, followed by the trait code, then followed by the chromosome number [b] Positive or negative effect alleles, for which a positive value indicates allele of the cultivated parent is positive, and a negative value indicates the allele of the wild parent is positive

Table S1.1: Significance level of each trait evaluated in the F8 population

Trait	**Effective degrees of freedom (m0)**	**Critical value**	**Significance level**
Pod shattering	4144.08	1.21E-05	4.92
Peduncle length	4367.53	1.14E-05	4.94
Flower color	1852.01	2.70E-05	4.57
Days to flowering	3146.32	1.59E-05	4.80
100-seed weight	4195.80	1.19E-05	4.92
Pod length	3881.83	1.29E-05	4.89
Leaf length	3946.57	1.27E-05	4.90
Leaf width	3666.21	1.36E-05	4.87
N^{o} seeds per pod	4738.88	1.06E-05	4.98

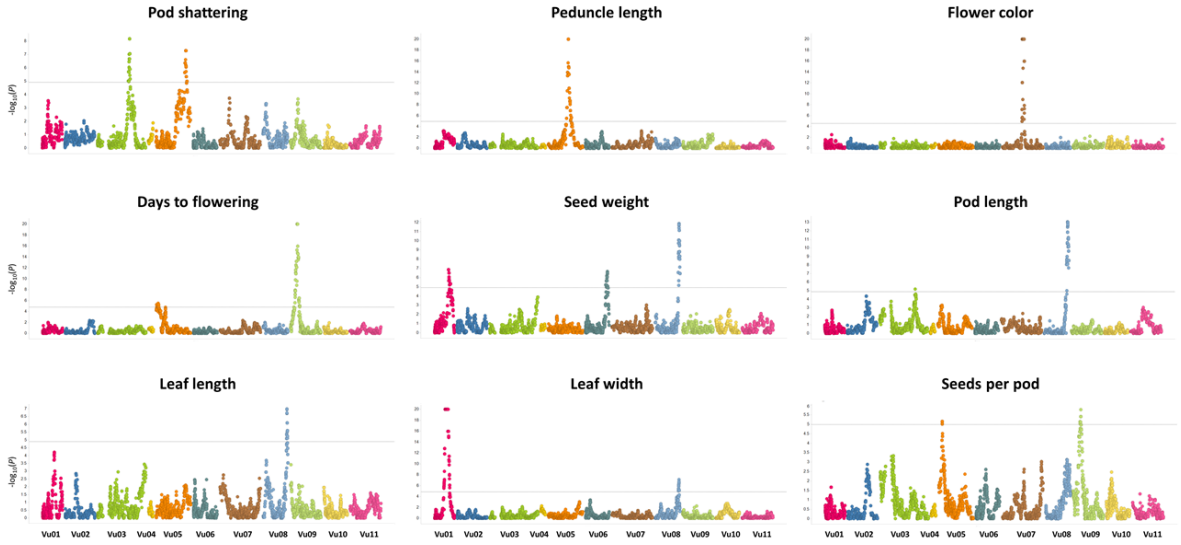

Pod shattering
Peduncle length
Flower color
Days to flowering
Seed weight
Pod length
Leaf length
Leaf width
Seeds per pod
Vu01 Vu02 Vu03 Vu04 Vu05 Vu06 Vu07 Vu08 Vu09 Vu10 Vu11

Figure 1.1*:* Genetic map of the RILs and QTL plots for the nine domestication-related traits. The horizontal axis indicate the chromosomes, the vertical axis indicates the –log10 of the probability (P-values). The dashed line indicates the significance threshold at 0.05.

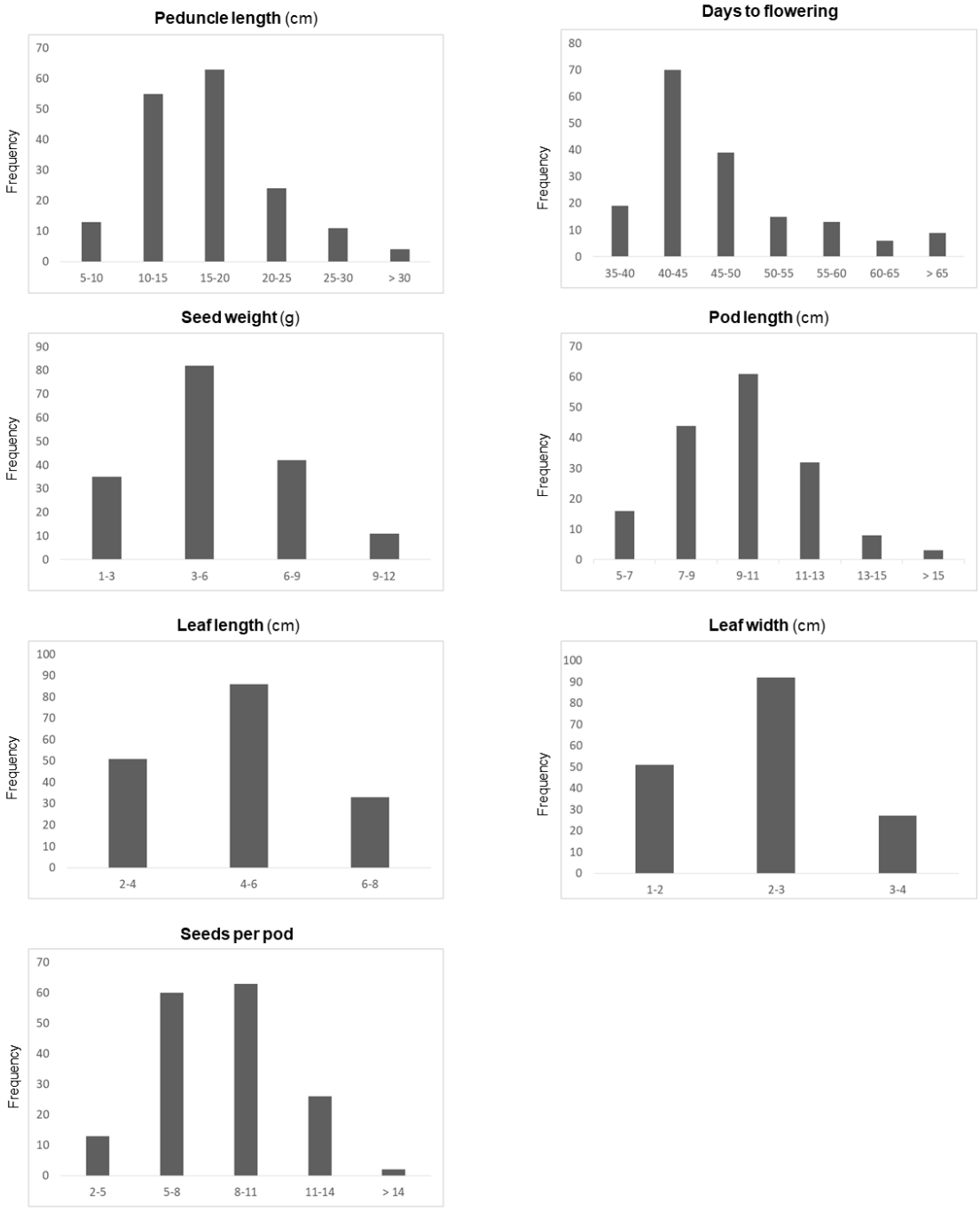


Figure S1.1: Phenotypic distribution of seven out of the nine domestication-related traits evaluated in the F8 population

Chapter 2: A Genome-Wide Association and Meta-Analysis Reveal Regions Associated with Seed Size in Cowpea [*Vigna unguiculata* (L.) Walp]

Abstract

Seed size is an important trait for yield and commercial value in dry-grain cowpea. Seed size varies widely among different cowpea accessions and the genetic basis of such variation is not yet well understood. To better decipher the genetic basis of seed size, a genome-wide association study (GWAS) and meta-analysis were conducted on a panel of 368 cowpea diverse accessions from 51 countries. Four traits, including seed weight, length, width and density were evaluated across three locations. Using 51,128 single nucleotide polymorphisms covering the cowpea genome, 17 loci were identified for these traits. One locus was common to weight, width and length, suggesting pleiotropy. By integrating synteny-based analysis with common bean, six candidate genes (*Vigun05g036000*, *Vigun05g039600*, *Vigun05g204200*, *Vigun08g217000*, *Vigun11g187000*, and *Vigun11g191300*) which are implicated in multiple functional categories related to seed size such as endosperm development, embryo development, and cell elongation were identified. These results suggest that a combination of GWAS meta-analysis with synteny comparison in a related plant is an efficient approach to identify candidate gene (s) for complex traits in cowpea. The identified loci and candidate genes provide useful information for improving cowpea varieties and for molecular investigation of seed size.

Introduction

Cowpea [*Vigna unguiculata* (L.) Walp] is a multipurpose crop and a major source of dietary protein, fiber, vitamins and minerals for millions of people and livestock in sub-Saharan Africa. Eaten in the form of fresh seeds, dry seeds, and fresh pods and as forage, cowpea is also an important crop in some parts of Asia, Latin America and the USA (Dadson et al. 2005; Singh 2014). Cowpea is also an additional source of income for smallholder farmers in Africa, and the impact on household economies depends largely on seed appearance including seed size.

Since the beginning of agriculture, increased seed size has been a main domestication target as an important component of grain yield. Domesticated crops produce larger seeds compared to their wild ancestors. Seed size has several agronomically important impacts. Lush and Wien (1980) reported that large seeded cowpea emerged earlier than small seeded types when planted deeply (up to 5 cm) and produce larger plants during early development. Cowpea seed size is an essential market trait in present day Africa and other parts of the world. Consumers tend to prefer larger seeds (Mishili et al. 2009). Understanding the genetic basis of cowpea seed size is fundamental for breeding for this complex trait. Classical inheritance analysis suggested that at least eight loci control cowpea seed size (Drabo et al. 1984). Since the first publication of quantitative traits loci (QTLs) of seed weight by molecular markers (Fatokun et al. 1992), efforts have been made to identify QTLs associated with seed size based on seed mass using mapping populations with different genetic backgrounds (Andargie et al. 2014; Huynh et al. 2018; Lo et al. 2018; Lucas et al. 2013; Pan et al. 2017). All of these QTLs were discovered

through linkage mapping using small populations. As a result, these QTLs are quite wide making the identification of causative genes challenging. Recently developed high-density genotyping tools and diverse germplasm subsets have made it possible to explore the genetic basis of this complex trait at finer resolution. To this aim, a genome-wide association study (GWAS) was conducted to investigate the genetic basis of seed size-related traits including seed length, width, and density using the recently developed genetic and genomic resources. Association mapping as described by Zhu et al. (2008) is a valuable tool to better understand the genetic basis of complex traits in plants. It has been widely adopted in several crop species to identify QTLs and to find candidate genes (Zhu et al. 2008). This approach enables the identification of genomic regions with finer resolution because of the smaller linkage disequilibrium in an association panel (Nordborg and Weigel 2008) as well as a larger population size.

A few GWA studies have been reported in cowpea for root architecture (Burridge et al. 2017), pod length (Xu et al. 2017), and black seed coat color (Herniter et al. 2018). No association mapping study has been reported on seed weight, length, width, and density in cowpea to date. Here, we conducted GWAS and meta-analysis to enhance our knowledge of the genetic architecture of seed size in cowpea. Meta-analysis (Rudner et al. 2002) is a useful tool through which GWAS results from different environments can be statistically pooled. This technique has been widely used in medical and social sciences. Seed size is known from studies on other plants to be controlled by various genes involved in different mechanisms including embryo and endosperm growth (Venglat et al. 2014). Genes affecting seed size have been cloned in *Arabidopsis*

including *APETALA2* (*AP2*), *SHB1* and *IKU1* (Jofuku et al. 2005; Wang et al. 2010; Zhou et al. 2009a). In soybean, *GmCYP78A10* and *GmCYP78A72* have been shown to play an important role in controlling seed size (Wang et al. 2015; Zhao et al. 2016). Here, we studied the genetic basis of seed size in domesticated cowpeas and identified single nucleotide polymorphisms (SNPs) significantly associated with seed size-related traits as well as promising candidate genes for which syntelogs in common bean (*Phaseolus vulgaris*) have been reported as seed weight candidate genes.

Materials and Methods

Plant materials

A mini-core collection consisting of 368 accessions (Muñoz-Amatriaín et al. 2017) was used for association mapping. The 368 accessions included landraces and breeding materials from 51 countries (Muñoz-Amatriaín et al. 2017). It also included members of both the subspecies *unguiculata* and *sesquipedalis* (yardlong bean). The 368 lines were classified into six subgroups based on population structure (Muñoz-Amatriaín et al. unpublished). All accessions were grown under favorable conditions in the greenhouse and field in 2016 and 2017. Greenhouse growth was carried out at the University of California Riverside campus (33.57°N; 117.20W) as follows: seeds from each accession were grown in 3.8 liter pots filled with UCmix3 with temperatures at 35 °C day and 23°C night. Mature, dried pods were harvested from each plant. For field experiments, each accession was planted in a single-row plot at the UCR Citrus Experiment Station, California, USA (UCR-CES, 33.97°N, 117.34°W; Field11) and at the Coachella Valley Agricultural Research Station, California, USA (CVARS, 33.52°N, 116.15°W). In field trials mature, dried pods were harvested randomly as to minimize bias.

Phenotypic evaluation and statistical analysis

Seed size-related traits were evaluated as 100-seed weight (g), seed length (mm), seed width (mm), and seed density (g/cm^3) in greenhouse and field experiments. The phenotyping for seed length and width was based on the average length and width of 10 seeds of each accession measured using a digital caliper. Seed density was calculated for

each accession based on seed mass and its volume following the formula:

$$\textbf{seed density} = \frac{\boldsymbol{seed\ mass}}{\frac{\boldsymbol{\pi}}{\boldsymbol{6}} * \textbf{length} * \textbf{width\^2}}$$

Among the seed size traits, seed weight was measured in three environments. Broad sense heritability was estimated based on the data across the three environments. The variance components were estimated with the PROC VARCOMP procedure in SAS, and broad sense heritability calculated as follows:

$$\boldsymbol{H^2} = \frac{\boldsymbol{V_G}}{\boldsymbol{V_G + V_E}}$$

where V_G is the variance of genotype and V_E is the variance of the experimental error.

Pearson's correlation coefficients were calculated using the cor.test () function in R (Ihaka and Gentleman 1996). The correlation coefficient was calculated between seed mass, seed length, seed width, and seed density.

SNP genotyping

Total genomic DNA was extracted from dried leaves collected from one plant of each accession using Plant DNeasy (Qiagen, Germany). Total DNA was quantified using a Quant-IT dsDNA Assay Kit (Thermo Fisher Scientific, USA). The 368 accessions were genotyped using the Cowpea iSelect Consortium Array containing 51,128 SNPs (Muñoz-Amatriaín et al. 2017). Genotyping was conducted at the University of Southern California Molecular Genomics Core facility (Los Angeles, California, USA). SNPs were called using GenomeStudio software V.2011.1 (Illumina, Inc. San Diego, CA). For

GWAS, data were filtered by removing SNPs with more than 25% missing calls, and minor allele frequency (MAF) less than 0.05. The physical positions of these SNPs were determined using the IT97K-499-35 reference genome (Lonardi et al. 2019).

Genome-wide association study and meta-analysis

GWAS was conducted on 100-seed weight (measured in three locations), length, width, and density (each measured in one location; CVARS). GWAS was performed using the mixed linear model (Zhang et al. 2010) in TASSEL V5.0 (http://www.maizegenetics.net/tassel). Principal Component Analysis (PCA) and a kinship coefficient matrix (*K*) were generated in TASSEL. PCA was used to account for population structure (*Q*) and *K* was used to correct for relatedness of accessions. The percentage contribution of each SNP to the total phenotypic variation was calculated using marker R2 values (computed by TASSEL) multiplied by 100.

Seed weight was measured in three environments. GWAS was first performed separately for each environment and then results were combined for a meta-analysis. A simple meta-analysis procedure recommended by Kang et al. (2014) was used to increase power of association and detect GxE interaction loci. This method essentially treats multiple environments as multiple populations. Loci that have different effects across different environments are G×E interaction loci. The method has been used in detecting interaction loci in a multiple environment mouse experiment (Kang et al. 2014). Among several specialized algorithms, Fisher's method was adopted because the method only requires the p-values across multiple environments. However, this method requires a

comprehensive tool to prove the distribution. Therefore, we developed a similar method that does not need the proof of the distribution. This method is called the probit method.

Let p_k be the p-value from environment k for $k = 1, \cdots, m$, where m is the number of environments. The test statistic is defined as

$$\boldsymbol{X = \sum_{k=1}^{m} [\Phi^{-1}(1 - 0.5 p_k)]^2}$$

where $\Phi^{-1}()$ is called the probit function, which is the inverse function of the standardized normal distribution. In other words, if $p = \Phi(z)$ and p is a standardized uniform variable, then $z = \Phi^{-1}(p)$ is a standardized normal variable. The square of a standardized normal variable is a Chi-square variable with one degree of freedom. The sum of m one-degree Chi-square variables is a Chi-square variable with m degrees of freedom. Under the null model (p_k is a standardized uniform variable between 0 and 1), this test statistic follows a Chi-square distribution with m degrees of freedom (no proof is needed). The new p-value was calculated as follows:

$$\boldsymbol{P_{probit} = 1 - Pr(X_m^2 \leq X^2)}$$

The result is much the same as in Fisher's method; the only difference is that the Chi-square distribution is given without need for proof. The threshold for genome-wide significance cutoff was applied based on Bonferroni correction at $\alpha = 0.05$

Candidate genes and syntelogs

The regions significantly associated with the traits were localized on the IT97K-499-35 reference genome (Lonardi et al. 2019) to determine the underlying candidate genes. The significant regions were also compared to syntenic regions on the common bean genome (Schmutz et al. 2014) using the legumeinfo.org instance of the Genome Context Viewer (GCV) (Cleary and Farmer 2017) to draw a list of common bean genes (syntelogs). Syntelogs were further intersected with a list of genes associated with seed weight reported in Schmutz et al. (2014) using functions of cowpeamine and legumemine (https://mines.legumeinfo.org). Cowpea genes homologous to the common bean genes present in the intersection were considered to be strong candidate genes associated with seed weight.

Results

Phenotypic variation and seed size trait correlations

The phenotyping panel contained 368 cowpea accessions comprising both landraces and breeding materials, classified into six subgroups based on population structure (Muñoz-Amatriaín et al. 2017). Seed size-related traits were evaluated based on seed weight, length, width and density. Broad sense heritability of seed weight was calculated, and the estimated value of 0.61 suggests relatively high heritability. The 368 accessions show a variation of grain sizes ranging from 5 to 31g per 100 seeds. Principal component analysis (PCA) showed the distribution of "seed weight" (Figure 2.1) in the diversity panel. To see the relationship between seed size traits, Pearson's correlation coefficients were calculated between seed mass, length, width, and density. Seed mass is highly correlated (positive) with seed length (0.84) and width (0.89). Furthermore, seed width and length showed moderate correlation (0.67), possibly because of the yardlong bean accessions in the diversity set. Negative correlations were observed between density and length (-0.67), and density and width (-0.85) indicating that as seeds are larger they are also less dense.

Association mapping

A total of 42,711 polymorphic SNPs were used for GWAS. A total of 17 loci were identified for the seed size-related traits (Figure 2.2; Table 2.1; Table 2.2). A hotspot was identified on chromosome Vu03 where QTLs for seed weight (*Sw3.2*), width (*Swi3*) and

length coincided. Colocalization of seed size-related trait QTLs suggests pleiotropy or physical linkage of genes controlling seed size in cowpea.

Seed weight. GWAS was performed separately for the three environments and results were combined for meta-analysis. The meta-analysis identified 13 significant regions on chromosomes 3, 4, 5, 6, 8, 10, and 11 (Figure 2.2; Table 2.1; Supplemental Table 2.1). Those loci included all the loci identified in each single environment (Supplemental Table 2.1), plus four loci that were not detected in any single environment. Four loci were consistent across two environments (UCR-CES and CVARS), while three loci were common to all three environments (GH, UCR-CES and CVARS) suggesting a presence of a significant G x E interaction. Effects and R^2 values of significant SNPs were calculated for each single environment and are reported in Supplemental Table 2.1.

Seed length, width and density. One significant region was associated with seed width on Vu03 and three regions were associated with seed density on Vu02, Vu08, and Vu11 (Figure 2.2; Table 2.2; Supplemental Table 2.1). The strongest QTL (*Sd11*) was localized on Vu11 and explained the highest genetic variance (21.51%; Table 2.2; Supplemental Table 2.1). Significant loci for seed length were not identified, although three loci were detected slightly below the significance threshold. Since the Bonferroni-corrected threshold is very conservative, loci with levels slightly lower than the significance threshold are listed in Supplemental Table 2.1. Of the loci considered for seed length, one was common with the locus detected for seed weight (*Sw3.2*) and seed width (*Swi3*) on Vu03, and another one was common with the locus detected for seed density (*Sd8*) on Vu08.

Candidate genes and syntelogs

All identified loci were aligned to the cowpea reference genome (Lonardi et al. 2019) and the underlying genes are listed in Supplemental Table 2.2. Synteny-based analysis performed with common bean using GCV (https://mines.legumeinfo.org) identified six syntelogs from the intersection with seed weight candidate genes reported by Schmutz et al. (2014). The corresponding cowpea genes of the syntelogs include two for *Sw5.1* (*Vigun05g036000, Vigun05g039600*), one for *Sw5.2* (*Vigun05g204200*), one for *Sw8.2* (*Vigun08g217000*), and two for *Sw11* (*Vigun11g187000*, *Vigun11g191300*). The gene *Vigun05g036000* is annotated as cell wall protein while *Vigun05g039600* encodes a phosphate transporter PHO1. *Vigun05g204200* is annotated as encoding a polycomb group protein FERTILIZATION-INDEPENDENT ENDOSPERM and *Vigun08g217000* codes for a histidine kinase 2. The genes *Vigun11g187000* and *Vigun11g191300* are annotated as WD repeat-containing protein 61-like isoform 1 and delta (24)-sterol reductase-like protein respectively. As comparative genomics is an efficient approach, we draw attention to these six interesting candidate genes.

Discussion

Seed size is one of the key yield determinants. Understanding the underlying genetic factors of cowpea seed size can help breeders develop improved varieties with a range of seed sizes. Also, identifying genes that could control seed size variation in domesticated cowpea can provide important insights into the genetic basis of adaptation, as seed size has been recognized as an important contributor to adaptation in plants (Chapin III et al. 1993; Igea et al. 2017). Previous studies using mapping populations from different genetic backgrounds have identified QTLs for seed size based on seed mass (Andargie et al. 2014; Fatokun et al. 1992; Huynh et al. 2018; Lo et al. 2018; Lucas et al. 2013; Pan et al. 2017). However, these studies have provided limited information on potential candidate genes for seed size. In cowpea, determining the genetic basis of complex traits has become increasingly effective since the availability of the genome sequence (Lonardi et al. 2019) and high-density genotyping (Muñoz-Amatriaín et al. 2017). As the next step to gain new knowledge on the genetic control of seed size, we used GWAS and meta-analysis to identify loci and candidate genes for seed size-related traits at a higher resolution that has been possible previously.

The meta-analysis identified 13 loci for seed weight (Figure 2.2; Table 2.1) including four loci that were not detected in any single location (Table 2.1). The power of meta-analysis to identify additional loci in the combined data set that were not detected in the single environment due to insufficient power provides support for its utility. As has been shown in many studies in medical and social sciences, meta-analysis can overcome the

limits of an individual environment by increasing the resolution power and reducing false-positive findings (Evangelou and Ioannidis 2013). This result also revealed a presence of a significant G x E interaction with some loci being location specific. The presence of G x E interaction loci noted here confirms the complexity of seed weight. Overlap between our meta-analysis result and previous studies was found. BLAST searches of RLFP probe sequences from Fatokun et al. (1992) against the reference genome sequence (Lonardi et al. 2019) revealed that the two QTLs identified by Fatokun et al. (1992) seem to correspond to *Sw3.2* and *Sw6.1*. Furthermore, Fatokun et al. (1992) reported that one of the QTLs (the one corresponding to *Sw3.2* here) is orthologous to a seed weight QTL in mung bean (*Vigna radiata*), suggesting that the genomic region has remained conserved through evolution. Also, Pan et al. (2017) reported a locus for grain weight which overlapped with *Sw3.2*. The location of QTL *Sw8.2* on Vu08 overlapped with that identified from several independent studies (Huynh et al. 2018; Lo et al. 2018; Lucas et al. 2013). However, *Sw6.2* on Vu06 is very close to a locus identified by Lo et al. (2018) for seed weight. This study may serve to refine this locus, or may indicate a novel locus. *Sw3.1*, *Sw4.1*, *Sw4.2*, *Sw5.1*, *Sw5.2*, *Sw8.1*, *Sw10.1*, *Sw10.2* and *Sw11* represent novel associated loci, further suggesting the complexity of seed size. Consistent with results from (Drabo et al. 1984), our study demonstrated that at least eight loci control cowpea seed size. Our meta-analysis result can guide the choice of QTL targeted for marker assisted selection.

In this study, we also identified significant regions for seed width and density. The locus for seed width (*Swi3*) was in the same region as *Sw3.2*. Also, a clear peak for seed length

was noted on the same genomic region. We note also that this region overlapped with a region identified for grain weight by Pan et al. (2017). Taken together with the high correlation between seed mass, width and length, we hypothesize that this QTL has a pleiotropic effect. However, the possibility of coexistence of multiple genes should not be excluded due to the complexity of these traits. Additional studies are necessary to further explore any of these hypotheses. In addition to their contribution to yield, seed width, density and length are important traits that have impact on the market value of cowpea. These QTLs will be a valuable resource for the improvement of cowpea.

We identified genes within these QTLs regions (Supplemental Table 2.2) using the annotated cowpea reference genome (Lonardi et al. 2019). Among the associated loci for seed weight, six particular candidate genes were identified based on comparative genomic analysis with common bean. Two interesting candidate genes for *Sw5.1* were *Vigun05g036000* and *Vigun05g039600*. The former encodes a cell wall protein, which was reportedly associated with seed size (Cheng et al. 1996; Jin et al. 2009; Weber et al. 1996). The latter, *Vigun05g039600*, encodes a phosphate transporter PHO1. In Arabidopsis, a PHO1 gene has been reported to be a positive regulator of seed development that affects both cell size and cell number (Zhou et al. 2009). Since several genetic pathways are known to control seed size, this gene is a promising candidate. Synteny-based analysis suggested that *Vigun05g204200*, which is a potential candidate for *Sw5.2* is annotated as encoding a polycomb group protein FERTILIZATION-INDEPENDENT ENDOSPERM (FIE). FIE genes are involved in endosperm development (Ohad et al. 1996) and have been shown to regulate seed size (Folsom et al.

2014). The candidate gene for *Sw8.2* was *Vigun08g217000* which codes for a histidine kinase 2. Interestingly, *Vigun08g217000* has been identified as potential candidate gene for increased organ size (Lonardi et al. 2019) during cowpea domestication and its Arabidopsis ortholog *AHK2* has been shown to regulate seed size (Bartrina et al. 2017; Riefler et al. 2006). Similarly, two candidate genes for *Sw11* were determined: *Vigun11g187000* and *Vigun11g191300*. The former is annotated as WD repeat-containing protein 61-like isoform 1. WD repeat proteins are mainly involved in cellular processes including cell division (van Nocker and Ludwig 2003). In addition, the Arabidopsis gene *AT2G34260* which encodes a WD repeat protein, is required for embryo and endosperm development (Bjerkan et al. 2012). *Vigun11g191300* encodes a delta (24)-sterol reductase protein and is an ortholog of the Arabidopsis *DIMINUTO* gene which has been shown to regulate cell elongation (Takahashi et al. 1995). *Vigun11g191300* is a strong candidate as seed size is influenced by multiple pathways.

In summary, combined GWAS meta-analysis and comparative genomics have led to a better understanding of the genetic basis of seed size-related traits. QTLs harboring candidate genes have been identified which deserve further in-depth studies to explore their possible roles in cowpea seed size. A direct result of the present study has been to establish genetic markers of these variants, which now are available to facilitate cowpea breeding for improved varieties with a range of seed sizes.

References

Andargie M, Pasquet RS, Gowda BS, Muluvi GM, Timko MP (2014) Molecular mapping of QTLs for domestication-related traits in cowpea (*Vigna unguiculata* (L.) Walp.). *Euphytica* 200:401-412

Bartrina I, Jensen H, Novak O, Strnad M, Werner T, Schmülling T (2017) Gain-of-function mutants of the cytokinin receptors AHK2 and AHK3 regulate plant organ size, flowering time and plant longevity. *Plant physiology*:pp. 01903.02016

Bjerkan KN, Jung-Roméo S, Jürgens G, Genschik P, Grini PE (2012) Arabidopsis WD repeat domain55 Interacts with DNA damaged binding protein1 and is required for apical patterning in the embryo. *The Plant Cell*:tpc. 111.089425

Burridge JD, Schneider HM, Huynh B-L, Roberts PA, Bucksch A, Lynch JP (2017) Genome-wide association mapping and agronomic impact of cowpea root architecture. *Theoretical and Applied Genetics* 130:419-431

Chapin III FS, Autumn K, Pugnaire F (1993) Evolution of suites of traits in response to environmental stress. *The American Naturalist* 142:S78-S92

Cheng W-H, Taliercio EW, Chourey PS (1996) The miniature1 seed locus of maize encodes a cell wall invertase required for normal development of endosperm and maternal cells in the pedicel. *The Plant Cell* 8:971-983

Cleary A, Farmer A (2017) Genome Context Viewer: visual exploration of multiple annotated genomes using microsynteny. *Bioinformatics* 34:1562-1564

Dadson R, Hashem F, Javaid I, Joshi J, Allen A, Devine T (2005) Effect of water stress on the yield of cowpea (*Vigna unguiculata* L. Walp.) genotypes in the Delmarva region of the United States. *Journal of agronomy and crop science* 191:210-217

Drabo I, Redden R, Smithson J, Aggarwal V (1984) Inheritance of seed size in cowpea (*Vigna unguiculata* (L.) Walp.). *Euphytica* 33:929-934

Evangelou E, Ioannidis JPA (2013) Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics* 14:379

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841-846

Folsom JJ, Begcy K, Hao X, Wang D, Walia H (2014) Rice Fertilization-Independent Endosperm1 Regulates Seed Size under Heat Stress by Controlling Early Endosperm Development. *Plant Physiology* 165:238-248
Herniter IA, Muñoz-Amatriaín M, Lo S, Guo Y-N, Close TJ (2018) Identification of Candidate Genes Controlling Black Seed Coat and Pod Tip Color in Cowpea (*Vigna unguiculata* [L.] Walp). *Genes, Genomes, Genetics*. 200521.202018

Huynh BL, Ehlers JD, Huang BE, Muñoz-Amatriaín M, Lonardi S, Santos JR, Ndeve A, Batieno BJ, Boukar O, Cisse N (2018) A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *The Plant Journal* 93:1129-1142

Igea J, Miller EF, Papadopulos AS, Tanentzap AJ (2017) Seed size and its rate of evolution correlate with species diversification across angiosperms. *PLoS biology* 15:e2002792
Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of computational and graphical statistics* 5:299-314

Jin Y, Ni D-A, Ruan Y-L (2009) Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. *The Plant Cell* 21:2072-2089

Jofuku KD, Omidyar PK, Gee Z, Okamuro JK (2005) Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proceedings of the National Academy of Sciences of the United States of America* 102:3117-3122

Kang EY, Han B, Furlotte N, Joo JWJ, Shih D, Davis RC, Lusis AJ, Eskin E (2014) Meta-analysis identifies gene-by-environment interactions as demonstrated in a study of 4,965 mice. *PLoS genetics* 10:e1004022

Lo S, Muñoz-Amatriaín M, Boukar O, Herniter I, Cisse N, Guo Y-N, Roberts PA, Xu S, Fatokun C, Close TJ (2018) Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp). *Scientific reports* 8:6261

Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo M-C, Alhakami H, Ounit R, Hasan AM, Verdier J, Roberts PA, Santos JRP, Ndeve A, Doležel J, Vrána J, Hokin SA, Farmer AD, Cannon SB, Close TJ The genome of cowpea (Vigna unguiculata [L.] Walp.). The Plant Journal https://doi.org/10.1111/tpj.14349

Lucas MR, Huynh BL, da Silva Vinholes P, Cisse N, Drabo I, Ehlers JD, Roberts PA, Close TJ (2013) Association Studies and Legume Synteny Reveal Haplotypes Determining Seed Size in *Vigna unguiculata*. *Frontiers in plant science* 4:95

Lush W, Wien H (1980) The importance of seed size in early growth of wild and domesticated cowpeas. *The Journal of Agricultural Science* 94:177-182

Mishili FJ, Fulton J, Shehu M, Kushwaha S, Marfo K, Jamal M, Kergna A, Lowenberg-DeBoer J (2009) Consumer preferences for quality characteristics along the cowpea value chain in Nigeria, Ghana, and Mali. *Agribusiness* 25:16-35

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O, Bozdag S, Cisse N, Drabo I, Ehlers JD, Farmer A, Fatokun C, Gu YQ, Guo Y-N, Huynh B-L, Jackson SA, Kusi F, Lawley CT, Lucas MR, Ma Y, Timko MP, Wu J, You F, Barkley NA, Roberts PA, Lonardi S, Close TJ (2017) Genome resources for climate-resilient cowpea, an essential crop for food security. *The Plant Journal* 89:1042-1054

Nordborg M, Weigel D (2008) Next-generation genetics in plants. *Nature* 456:720-723

Ohad N, Margossian L, Hsu Y-C, Williams C, Repetti P, Fischer RL (1996) A mutation that allows endosperm development without fertilization. *Proceedings of the National Academy of Sciences* 93:5319-5324

Pan L, Wang N, Wu Z, Guo R, Yu X, Zheng Y, Xia Q, Gui S, Chen C (2017) A High Density Genetic Map Derived from RAD Sequencing and Its Application in QTL Analysis of Yield-Related Traits in *Vigna unguiculata*. *Frontiers in plant science* 8:1544

Riefler M, Novak O, Strnad M, Schmülling T (2006) Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *The Plant Cell* 18:40-54

Rudner L, Glass GV, Evartt DL, Emery PJ (2002) A user's guide to the meta-analysis of research studies. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park 37

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nature genetics* 46:707

Singh B (2014) Cowpea: the food legume of the 21st century. *Crop Science Society of America*

Takahashi T, Gasch A, Nishizawa N, Chua N-H (1995) The DIMINUTO gene of Arabidopsis is involved in regulating cell elongation. *Genes & Development* 9:97-107

van Nocker S, Ludwig P (2003) The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC genomics* 4:50-50

Venglat P, Xiang D, Wang E, Datla R (2014) Genomics of seed development: Challenges and opportunities for genetic improvement of seed traits in crop plants. *Biocatalysis and Agricultural Biotechnology* 3:24-30

Wang A, Garcia D, Zhang H, Feng K, Chaudhury A, Berger F, Peacock WJ, Dennis ES, Luo M (2010) The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. *The Plant Journal* 63:670-679

Wang X, Li Y, Zhang H, Sun G, Zhang W, Qiu L (2015) Evolution and association analysis of GmCYP78A10 gene with seed size/weight and pod number in soybean. *Molecular biology reports* 42:489-496

Weber H, Borisjuk L, Wobus U (1996) Controlling seed development and seed size in Vicia faba: a role for seed coat-associated invertases and carbohydrate state. *The Plant Journal* 10:823-834

Xu P, Wu X, Muñoz-Amatriaín M, Wang B, Wu X, Hu Y, Huynh BL, Close TJ, Roberts PA, Zhou W (2017) Genomic regions, cellular components and gene regulatory basis underlying pod length variations in cowpea (*Vigna unguiculata* L. Walp). *Plant biotechnology journal* 15:547-557

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature genetics* 42:355-360

Zhao B, Dai A, Wei H, Yang S, Wang B, Jiang N, Feng X (2016) Arabidopsis KLU homologue GmCYP78A72 regulates seed size in soybean. *Plant molecular biology* 90:33-47

Zhou Y, Zhang X, Kang X, Zhao X, Zhang X, Ni M (2009) SHORT HYPOCOTYL UNDER BLUE1 associates with MINISEED3 and HAIKU2 promoters in vivo to regulate Arabidopsis seed development. *The Plant Cell* 21:106-117

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *The plant genome* 1:5-20

Table 2.1: List of significant QTLs for seed weight from GWAS meta-analysis results.

QTL	**Peak SNP**	**Chr.**	**Position**	**-Log$_{10}$P**	**Comment**
Sw3.1	2_26016	3	2223359	18.36	GH, Field 11, CVARS, Meta
Sw3.2	2_40922	3	51562014	9.67	Field 11, Meta
Sw4.1	2_14347	4	3483218	6.51	Meta
Sw4.2	2_25894	4	39757606	10.41	CVARS, Meta
Sw5.1	2_07617	5	1783613	11.08	Field 11, CVARS, Meta
Sw5.2	2_08022	5	39230104	6.63	Field 11, Meta
Sw6.1	2_07078	6	19698161	14.84	GH, Field 11, CVARS, Meta
Sw6.2	2_07358	6	32476947	7.92	Meta
Sw8.1	2_50483	8	23709814	8.61	Meta
Sw8.2	2_17605	8	37710852	6.32	Meta

Sw10.1	2_50509	10	29351777	10.73	CVARS, Meta
Sw10.2	2_08400	10	35932948	6.87	Field 11, Meta
Sw11	2_28551	11	38568662	12.69	GH, Field 11, CVARS, Meta

Table 2.2: List of significant QTLs for seed width and density from GWAS results.

Trait	QTL	Peak SNP	Chr.	Position	-$Log_{10}P$	R^2 (%)	Alleles	Effect
Seed width	*Swi3*	2_29499	3	52931236	6.35	7.09	A/C	-0.59
Seed density	*Sd2*	2_09625	2	20857395	11.82	14.84	G/T	0.06
	Sd8	2_27144	8	36984438	6.22	7.06	A/C	-0.63
	Sd11	2_41753	11	24148832	17.27	21.93	C/T	0.12

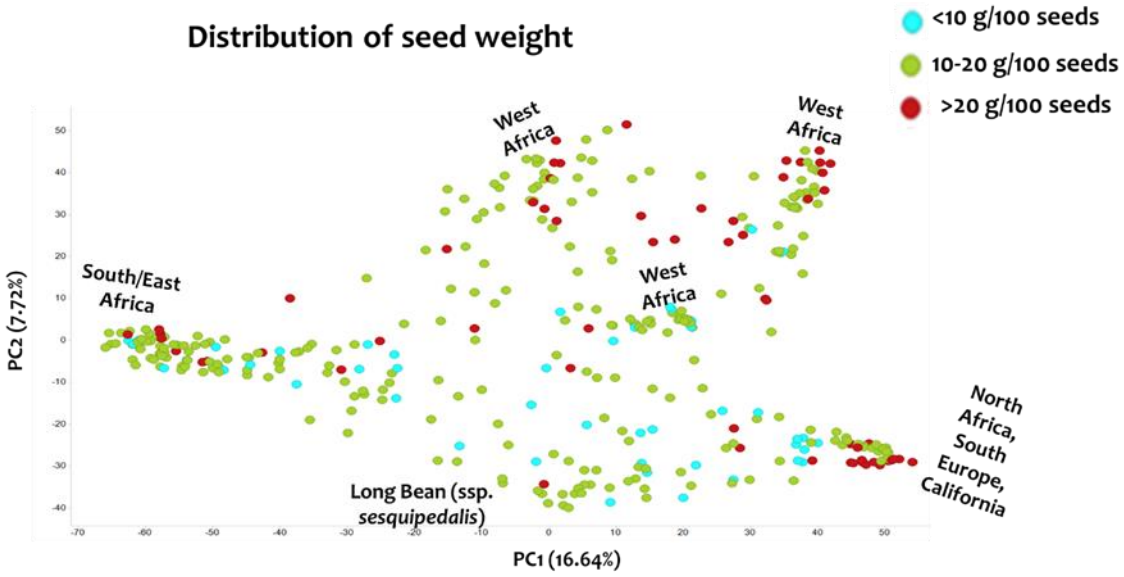

Distribution of seed weight
<10 g/100 seeds
10-20 g/100 seeds
>20 g/100 seeds
West Africa
West Africa
West Africa
South/East Africa
North Africa, South Europe, California
Long Bean (ssp. sesquipedalis)
PC2 (7.72%)
PC1 (16.64%)

Figure 2.1: Principal component analysis of the 368 accessions colored by seed weight range

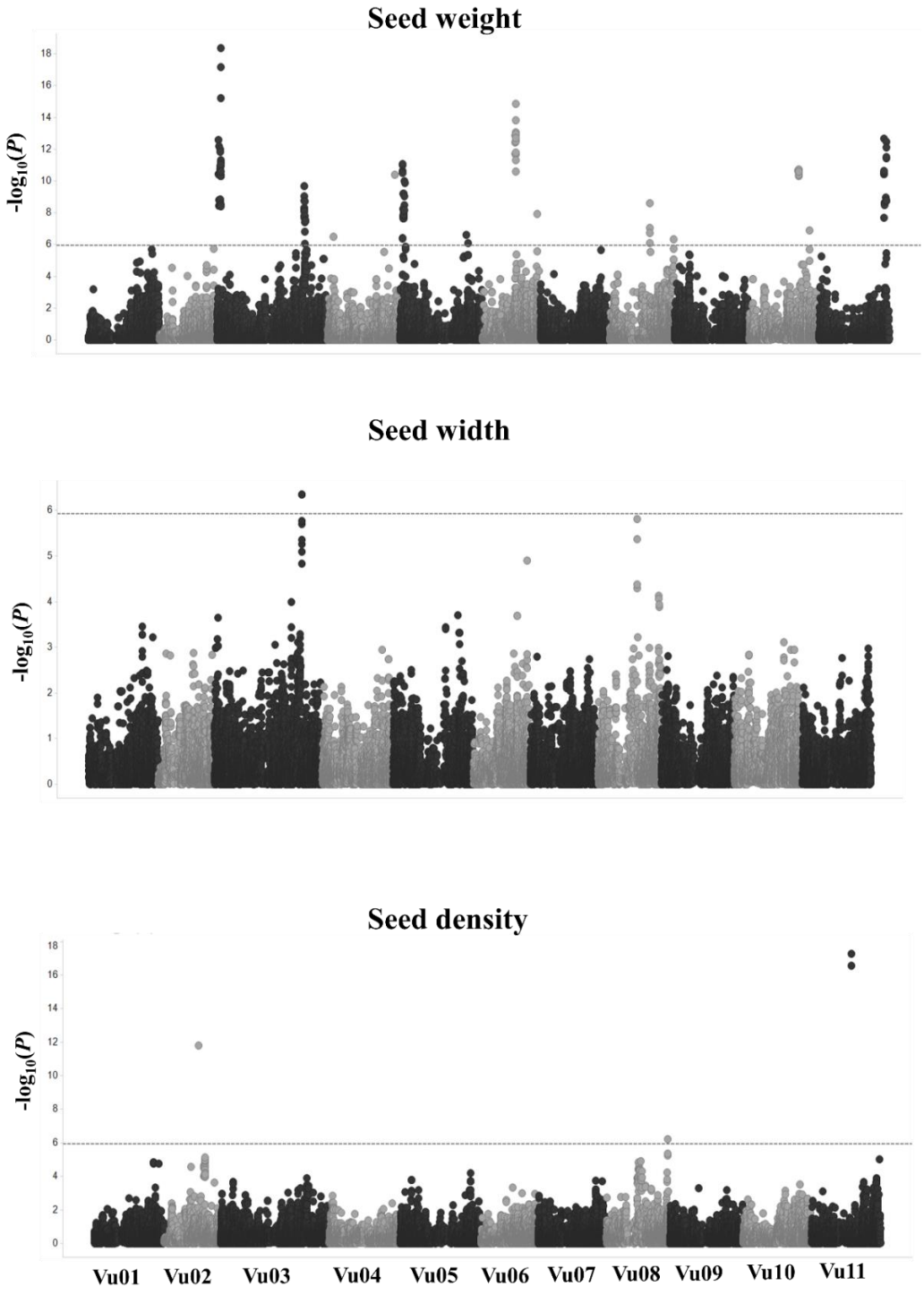

Seed weight
-log10(P)
Seed width
-log10(P)
Seed density
-log10(P)
Vu01
Vu02
Vu03
Vu04
Vu05
Vu06
Vu07
Vu08
Vu09
Vu10
Vu11

Figure 2.2: Manhattan plot of GWAS meta-analysis for seed weight and GWAS for seed width and density. Negative log 10 P-values are plotted against physical position on each of the 11 chromosomes. Dashed line indicates Bonferroni-corrected genome-wide significance threshold.

Chapter 3: Marker Assisted Pyramiding of Seed Size QTLs into a Popular Cowpea Cultivar from Senegal

Abstract

Seed size is an important target trait of cowpea breeding. Various surveys reported that consumers are willing to pay a premium for large cowpea grains. Therefore, it will be desirable to breed cowpea with a range of seed sizes. In this study, two seed size QTLs (*Css-1* and *Css-4*) were targeted for introgression in the background of a popular cultivar from Senegal, "Pakau". A set of SNPs linked to the seed size QTL regions was identified and used to develop introgression lines. Background and foreground selection identified promising lines with 90.44-98.01% of the recurrent parent genome carrying large seed alleles. The effect of each QTL has been determined. The addition of *Css-1* resulted in increased seed weight, seed weight decreased in the introgression lines carrying *Css-4*. Our results indicate that the potential of these QTLs in increasing seed size is determined by the genetic background of the recipient.

Introduction

Seed size has been targeted by human selection during the domestication of crop species. It is an important trait for yield. Seed size governs market classes in several crops, including cowpea [*Vigna unguiculata* (L.) Walp]. Cowpea is one of the highest sources of protein, folic acid and several vitamins for many people in sub-Saharan Africa. It contributes substantially to global food and nutritional security. Incorporating information on consumer preferences is important for decisions on developing improved cowpea cultivars, and increased seed size is one of the targeted preferred traits. Various surveys reported that consumers are willing to pay a premium for large cowpea grains (Langyintuo et al. 2004; Mishili et al. 2009). Improving cowpea cultivars by increasing grains size will increase market demand and have a big impact on household economies. Therefore, it will be desirable to breed cowpeas with a range of seed sizes.

Cowpea seed size is governed by at least eight loci (Drabo et al. 1984). Several quantitative trait loci (QTL) associated with seed size/mass have been mapped in populations with different genetic backgrounds (Andargie et al. 2014; Fatokun et al. 1992; Huynh et al. 2018; Lo et al. 2018; Lucas et al. 2013; Pan et al. 2017). Two seed size loci, *Css-1* and *Css-4*, were described in Lucas et al. (2013) as having a potential to develop large seeds. The *Css-1* locus was identified also in different genetic backgrounds including a multi-parent advanced generation inter-cross (MAGIC) population (Huynh et al. 2018) and a wild x cultivated population (Lo et al. 2018). *Css-4* mapped on the same chromosome as *CSw6* (Lo et al. 2018) but at ~5Mb physical distance based on the

cowpea reference genome (Lonardi et al. 2019). Recently, a Cowpea iSelect Consortium Array which contains 51,128 single nucleotide polymorphism (SNP) markers (Muñoz-Amatriaín et al. 2017) has been developed and used for genotyping the CB27/IT82E-18 population where the two loci have been mapped (Lucas et al. 2013). The same data were used in QTL mapping with denser SNP markers to refine the QTL intervals. Here, the newly identify SNPs were used to pyramid the two QTLs into a popular cowpea variety from Senegal "Pakau".

QTL pyramiding is the introgression of QTLs for a trait or multiple traits into an elite cultivar that is deficient for these traits. QTL pyramiding has been widely used in several crops to transfer multiple disease resistance loci. Pyramiding of loci for complex traits through conventional backcrossing is difficult and time-consuming. With the availability of high-density SNP markers in cowpea (Muñoz-Amatriaín et al. 2017), it has become much easier to combine desirable alleles from different loci. This study was undertaken to introgress beneficial seed size QTLs alleles into a popular cultivar from Senegal "Pakau". After three backcross cycles and selfing the backcross progenies, the effect of these QTLs on seed size in the "Pakau" genetic background was analyzed.

Materials and Methods

Plant materials

For the QTL analysis a RIL population derived from a cross between a California Blackeye (CB27) and an African buff type (IT82E-18) was used. Two parents were also used to develop introgression lines with a varying number of QTL for seed size. A breeding line "113-4-6-14-1" (29.75g/100seeds) developed from the study of Lucas et al. (2015) was selected as the donor parent for QTL introgression. It derived from a cross between CB27 and RIL113 (CB27/IT82E-18/CB27) and carried two larger seed size QTLs. The recurrent parent "Pakau" (17.66g/100seeds) is a popular Senegalese cultivar with desirable traits including bacterial blight resistance.

QTL analysis and identification of markers

The seed size QTLs (*Css-1* and *Css-4*) were mapped previously in the CB27 x IT82E-18 RIL population in the study of Lucas et al. (2013). The same data were reanalyzed using data from the Cowpea iSelect Consortium Array (51,128 SNPs; Muñoz-Amatriaín et al. 2017). A total of 16,566 high quality SNPs segregate in this population and have been used to identify additional SNPs linked to the seed size QTL regions. QTL mapping was performed using a modified version of the linear mixed model from Xu (2013) and described by Lo et al. (2018). The newly identified SNP markers were used to analyze the QTL content of the introgression lines at each generation.

SNP genotyping and QTL pyramiding

Total genomic DNA was extracted from dried leaves of the parental genotypes, F1s and backcross (BC) progenies using Plant DNeasy (Qiagen, Germany). Total DNA was quantified using Quant-IT dsDNA Assay Kit (Thermo Fisher Scientific, USA). The parental genotypes, F1s and BC progenies were genotyped with 51,128 SNPs using the Cowpea iSelect Consortium Array (Muñoz-Amatriaín et al. 2017). Ten SNP markers were selected within the QTL regions and converted to Kompetitive Allele Specific PCR (KASP) by LGC genomics (United Kingdom) to categorize the BC3F2 lines.

The crosses were performed by hand pollination and Pakau was used as female parent for each cross. All crosses were made in a greenhouse on the campus of the University of California, Riverside. As a first step, the seed size donor line 113-4-6-14-1 was crossed with Pakau to create F1 individuals. An F1 line was then used as donor parent for the first backcross, and four combinations were observed in the progenies. Donor parents for the second and third backcross cycles were chosen from each combination after background selection, based on all the polymorphic markers between the two parents, and foreground selection based on the QTL content. Several BC3F1 were selfed and homozygotes lines of the positive alleles from each combination were selected based on the KASP markers within the QTL regions and analyzed to test the effect on seed phenotype of these QTL in the “Pakau” genetic background.

Trait evaluation and data analysis

Phenotypic evaluations (100-seed weight) of the parents and introgression lines were conducted in the greenhouse at the University of California Riverside campus. F tests were conducted to analyze the association between the KASP markers and seed weight. The means and standard deviation of seed weight of the BC3F2 were calculated to test the effect on seed phenotype of *Css-1* and *Css-4* in the Pakau genetic background using the 10 KASP markers.

Results and Discussion

Seed size is a consumer trait of cowpea in Senegal and many parts of the world. In this study two seed size QTLs from a previous study of Lucas et al. (2013) were pyramided in the genetic background of a popular cultivar from Senegal Pakau. QTL pyramiding is commonly used to introgress desirable alleles of multiple QTLs for one or more traits into an elite cultivar. Here, QTL analysis was performed in the CB27 x IT82E-18 RIL population using the same phenotypic data and genotypic data from the Cowpea iSelect Consortium Array (Muñoz-Amatriaín et al. 2017). Two significant QTLs were mapped in this population: *Css-4* on chromosome 6 and *Css-1* on chromosome 8. From our results, *Css-1* explained 19.02% of the phenotypic variation with the favorable allele from IT82E-18, while *Css-4* explained 17.05% of the phenotypic variation and the positive allele contributed from CB27. For validation, *Css-1* was also mapped in other populations from different genetic background (Huynh et al. 2018; Lo et al. 2018). *Css-1* has been successfully introgressed to a California black-eyed type "CB27" (Lucas et al. 2015). Here, both QTLs, *Css-1* and *Css-4*, were selected for introgression into a different genetic background, Pakau.

A total of 16,052 SNPs were polymorphic between Pakau and line 113-4-6-14-1 and were used to analyze the percentage of Pakau background for each backcross progeny. Based on QTL analysis and the cowpea consensus map (Muñoz-Amatriaín et al. 2017), 40 and 51 SNPs were identified in the *Css-1* and *Css-4* QTL regions, respectively. Those SNPs were used to assess the QTL content of the introgression lines (foreground

selection). A total of 70 BC3F1 lines were produced, and four combinations were observed: lines with no QTL, with *Css-1*, with *Css-4*, or with *Css-1* and *Css-4* (Figure 3.1). Figure 3.2 provides an image of seeds of each parent and introgression lines (BC3F3 seeds) carrying positive alleles of both QTL, positive allele of *Css-1* only, and positive allele of *Css-4* only. The percentage of Pakau background and QTL content of donor lines chosen from each backcross generation are given in Table 3.1. The use of high-density SNPs markers enables high recovery of the recurrent parent in the backcross progenies allowing the selection of promising lines with "Pakau" genome ranging from 90-98% and carrying large seed alleles.

A total of ten SNPs were selected within *Css-1* and *Css-4* regions and transformed to KASP assays. The KASP markers were used to categorize the BC3F2, and F tests were conducted to determine their association with seed weight in the introgression lines. All the five KASP markers in *Css-1* region showed highly significant association with seed weight, while only one marker in the *Css-4* region was significant (Table 3.2). Allelic effect analysis of these QTLs in the Pakau genetic background was also carried out; the means with standard deviations of seed weight for each allele are presented in table 3.3. Homozygous lines carrying favorable alleles at *Css-1* had an average seed weight of >20g/100 seed, while heterozygotes averaged >19g/100 seed, and homozygous unfavorable alleles had an average seed weight of >17g/100 seed. An increase of ~ 3g/100 seed was apparent in the introgression lines compared to Pakau, which has an average weight of 17.66g/100 seed. This study validated the effect of the *Css-1* QTL in the potential to develop large seeded cowpea. In contrast, seed weight decreased

somewhat with *Css-4*. Homozygous lines with donor alleles for *Css-4* averaged >15g/100 seed while heterozygotes had an average seed weight of >17g/100, and homozygous lines with "Pakau" alleles averaged >18g/100 seed. Introgression of *Css-4* in a different genetic background is needed to determine the potential of this QTL in increasing seed weight. Nevertheless, the introgression of *Css-1* has enabled the development of promising lines. Future work includes testing the selected lines for yield performance in their target environment and seed composition to determine their utility as breeding materials.

Seed weight is a complex trait and other QTLs have been mapped in a broader germplasm (Lo et al. under review). Three major QTLs (Sw3.1, Sw6.1 and Sw11) were detected in three different locations and it is likely that these QTLs may have larger effect than *Css-1*. Therefore, these QTLs may be considered when breeding for large seeded cowpea.

References

Andargie M, Pasquet RS, Gowda BS, Muluvi GM, Timko MP (2014) Molecular mapping of QTLs for domestication-related traits in cowpea (*V. unguiculata* (L.) Walp.). *Euphytica* 200:401-412

Drabo I, Redden R, Smithson J, Aggarwal V (1984) Inheritance of seed size in cowpea (*Vigna unguiculata* (L.) Walp.). *Euphytica* 33:929-934

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841-846

Huynh BL, Ehlers JD, Huang BE, Muñoz-Amatriaín M, Lonardi S, Santos JR, Ndeve A, Batieno BJ, Boukar O, Cisse N (2018) A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *The Plant Journal* 93:1129-1142

Langyintuo A, Ntoukam G, Murdock L, Lowenberg-DeBoer J, Miller D (2004) Consumer preferences for cowpea in Cameroon and Ghana. *Agricultural Economics* 30:203-213

Lo S, Muñoz-Amatriaín M, Boukar O, Herniter I, Cisse N, Guo Y-N, Roberts PA, Xu S, Fatokun C, Close TJ (2018) Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp). *Scientific reports* 8:6261

Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo M-C, Alhakami H, Ounit R, Hasan AM, Verdier J, Roberts PA, Santos JRP, Ndeve A, Doležel J, Vrána J, Hokin SA, Farmer AD, Cannon SB, Close TJ The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *The Plant Journal* https://doi.org/10.1111/tpj.14349

Lucas MR, Huynh BL, da Silva Vinholes P, Cisse N, Drabo I, Ehlers JD, Roberts PA, Close TJ (2013) Association Studies and Legume Synteny Reveal Haplotypes Determining Seed Size in *Vigna unguiculata*. *Frontiers in plant science* 4:95

Lucas MR, Huynh BL, Roberts PA, Close TJ (2015) Introgression of a rare haplotype from Southeastern Africa to breed California blackeyes with larger seeds. *Frontiers in plant science* 6:126

Mishili FJ, Fulton J, Shehu M, Kushwaha S, Marfo K, Jamal M, Kergna A, Lowenberg-DeBoer J (2009) Consumer preferences for quality characteristics along the cowpea value chain in Nigeria, Ghana, and Mali. *Agribusiness: An International Journal* 25:16-35

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O, Bozdag S, Cisse N, Drabo I, Ehlers JD, Farmer A, Fatokun C, Gu YQ, Guo Y-N, Huynh B-L, Jackson SA, Kusi F, Lawley CT, Lucas MR, Ma Y, Timko MP, Wu J, You F, Barkley NA, Roberts PA, Lonardi S, Close TJ (2017) Genome resources for climate-resilient cowpea, an essential crop for food security. *The Plant Journal* 89:1042-1054

Pan L, Wang N, Wu Z, Guo R, Yu X, Zheng Y, Xia Q, Gui S, Chen C (2017) A High Density Genetic Map Derived from RAD Sequencing and Its Application in QTL Analysis of Yield-Related Traits in *Vigna unguiculata*. *Frontiers in plant science* 8:1544

Xu S (2013) Mapping quantitative trait loci by controlling polygenic background effects. Genetics 195:1209-1222

Table 3.1: Backcross progenies with their percentage Pakau background and QTL content

Generation	**Highest % Pakau**	**Range % Pakau**	**QTL content**
BC1F1	79.30	32.99 - 79.30	*Css-1*
BC1F1	64.02	25.26 - 64.02	*Css-4*
BC1F1	58.98	15.43 - 58.98	*Css-1* + *Css-4*
BC2F1	93.08	68.80 - 93.08	*Css-1*
BC2F1	84.64	75.16 - 84.64	*Css-4*
BC2F1	79.45	79.45	*Css-1* + *Css-4*
BC3F1	97.85	83.64 - 97.85	*Css-1*
BC3F1	93.28	81.4 - 93.28	*Css-4*
BC3F1	89.58	82.5 - 89.58	*Css-1* + *Css-4*
BC3F2	98	98	*Css-1*
BC3F2	92.47	91.36 - 92.47	*Css-4*
BC3F2	92.48	90.44 - 92.48	*Css-1* + *Css-4*

Table 3.2: F test results showing association between the KASP markers and seed weight

QTL	SNP	Df [a]	Sum Sq [b]	Mean Sq [c]	F value	Pr(>F)	Significance
Css-1	1_1130	2	178.626	89.313	20.953	7.05E-08	***
	2_02645	2	175.652	87.826	20.404	9.98E-08	***
	2_29527	2	178.626	89.313	20.953	7.05E-08	***
	1_0078	2	178.626	89.313	20.953	7.05E-08	***
	1_0387	2	187.650	93.825	22.530	2.94E-08	***
Css-4	1_0426	2	51.987	25.994	4.328	0.016898	*
	1_0794	2	25.312	12.656	1.942	0.151166	
	1_0860	2	23.667	11.833	1.814	0.170614	
	1_0911	2	37.164	18.582	2.979	0.057341	
	2_12854	2	21.225	10.612	1.617	0.205967	

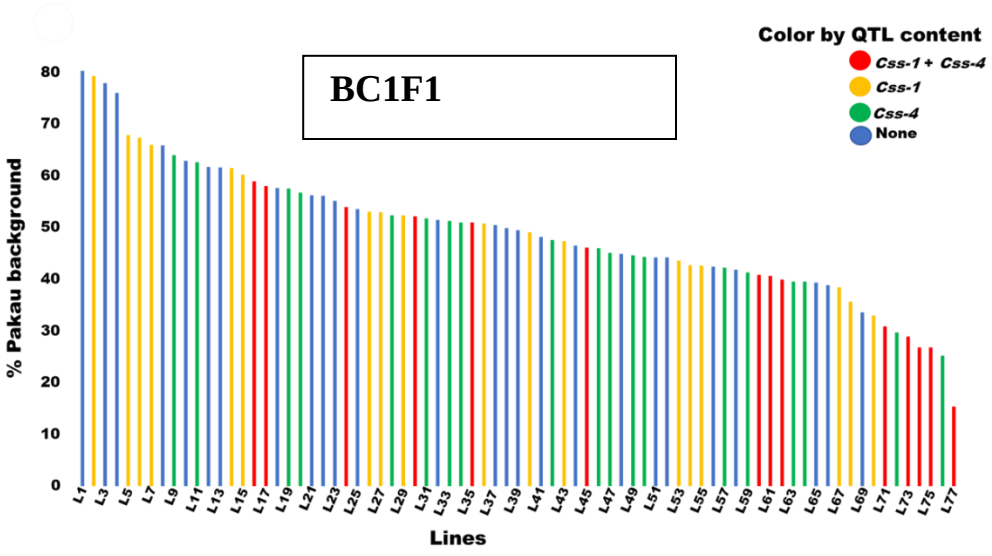
a = degree of freedom

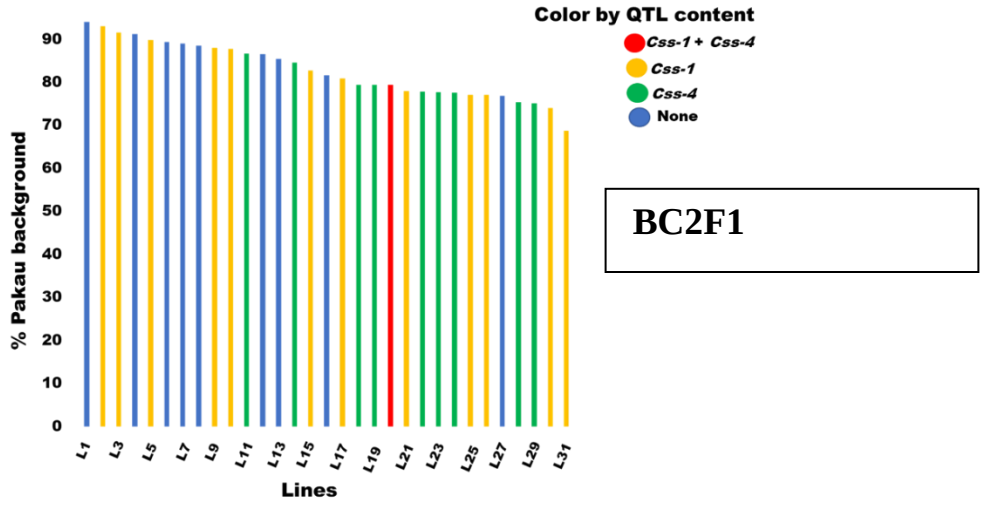
b = sum of squares

c = mean square

Table 3.3: QTL allele effects on seed weight (g) in the Pakau genetic background. Seed weight data from BC3F3

QTL	SNP	100-seed weight (g)	Alleles		
Css-1	1_1130		GG	AA	GA
		Range	17.65 - 23.96	14.42 - 20.6	13.16 - 21.84
		Mean (Stdev)	20.72 (2.02)	17.07 (1.64)	19.24 (2)
	2_02645		GG	TT	GT
		Range	17.65 - 23.96	14.42 - 20.6	13.16 - 21.84
		Mean (Stdev)	20.51 (2.08)	17.02 (1.64)	19.18 (1.95)
	2_29527		AA	GG	GA
		Range	17.65 - 23.96	14.42 - 20.6	13.16 - 21.84
		Mean (Stdev)	20.72 (20.3)	17.07 (1.64)	19.24 (2)
	1_0078		AA	GG	GA
		Range	17.65 - 23.96	14.42 - 20.6	13.16 - 21.84
		Mean (Stdev)	20.72 (2.03)	17.07 (1.64)	19.24 (2)
	1_0387		AA	GG	GA
		Range	14.42 - 20.6	17.65 - 23.96	13.16 - 21.84
		Mean (Stdev)	17.02 (1.63)	20.86 (1.92)	19.13 (1.96)
Css-4	1_0426		AA	GG	GA
		Range	15.42 - 19.5	13.16 - 23.96	14.42 - 20.6
		Mean (Stdev)	17.66 (2.07)	18.72 (2.18)	17.01 (1.94)
	1_0794		AA	GG	GA
		Range	14.89 - 15.48	13.16 - 23.96	14.42 - 20.6
		Mean (Stdev)	15.18 (0.41)	18.7 (2.21)	17.41 (1.92)
	1_0860		CC	GG	CG
		Range	15.11 - 23.96	13.16 - 17.05	14.42 - 20.68
		Mean (Stdev)	18.88 (2.13)	15.18 (1.22)	18.35 (1.86)
	1_0911		AA	CC	CA
		Range	15.11 - 23.96	13.16 - 17.05	14.95 - 20.68
		Mean (Stdev)	18.96 (2.03)	15.08 (1.16)	18.33 (1.81)
	2_12854		AA	GG	GA
		Range	13.16 - 17.05	15.11 - 23.96	14.42 - 20.68
		Mean (Stdev)	15.18 (1.22)	18.8 (2.13)	18.35 (1.87)


BC1F1
Color by QTL content
Css-1 + Css-4
Css-1
Css-4
None
% Pakau background
Lines


BC2F1
Color by QTL content
Css-1 + Css-4
Css-1
Css-4
None
% Pakau background
Lines

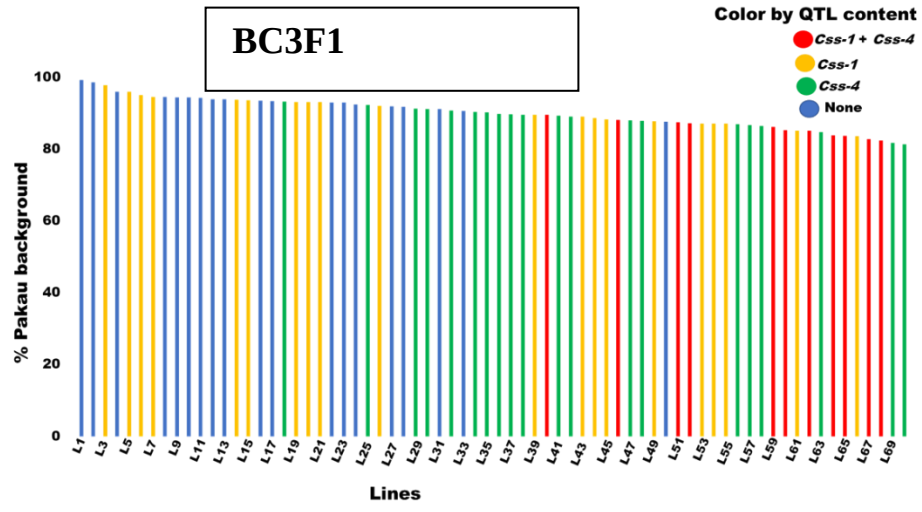

BC3F1
Color by QTL content
Css-1 + Css-4
Css-1
Css-4
None
% Pakau background
Lines

Figure 3.1: Graph of the BC progenies with their percentage Pakau background and QTL content


A
B
C
D
E

Figure 3.2: Seeds of the two parents and introgression lines (BC3F3). A) 113-4-6-14-1 (donor parent); B) Pakau (recurrent parent); C) line with both QTLs; D) line with Css-1 only; E) line with Css-4 only

Conclusion

During thousands of years, crop plants have been domesticated and cultivated for useful traits to fulfill our needs. Domestication history differs between crop plants but follows common patterns. Knowledge on the genetics of domestication traits will enhance our ability to use broad germplasm in an effective way in breeding programs. Compared to other major crops, little is known about the genetics of domestication traits in cowpea. The investigations included in my thesis identified loci and trait-associated SNP that can be used for cowpea genetic improvement as well as candidate genes that serve as promising targets for studies of molecular mechanisms underlying the related traits. Furthermore, marker assisted backcrossing can efficiently be used to pyramid QTLs into elite cultivars.

While we observed technological advancements in molecular biology in model organisms and several major crop plants over the past two decades, some essential resources for genome research were missing in cowpea until very recently. For example, a high-quality reference genome is essential for functional genomics studies. Recently, a reference genome has been developed (Lonardi et al. 2019) which enabled the identification of candidate genes for domestication traits in this thesis. Functional studies will help to validate the candidate genes and ascertain their roles on the related traits. Genetic transformation is of crucial importance for functional studies. In cowpea, successful transformation has been reported using cotyledonary node with *Agrobacterium*, but with

a low success rate of just over 1% (Bett et al. 2019). Still, this is a promising starting point for further improvement of transformation in cowpea.

Functions of candidate genes can be assessed by the traditional methods of functional characteristics of genes. In this case, it can be done by investigating the effect on the phenotype of gene knockout or overexpression in different cowpea accessions. The gene can be inactivated by virus-induced gene silencing (Baulcombe 1999) or by RNA interference (Smith et al. 2000). Both methods can interrupt the functions of specific genes by repressing the corresponding mRNAs and have been used successfully to investigate gene function. Another strategy to investigate the function of the candidate genes identified here is to explore the resulting phenotypes by gene overexpression. In the same way that gene knockout result in a loss of function phenotype, overexpression can produce new phenotypes. In this case, the target gene is expressed at a much higher level than normal as it is integrated in multiple copies in the genome.

Recently, the development of genome editing (Ma et al. 2016) provides great opportunities to study the function of a gene. Different genome editing methods including zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been used in functional studies. ZFNs and TALENs require complicated constructs and have been mostly replaced by a simpler technology, CRISPR/Cas9 (Ma et al. 2016). The CRISPR/Cas9 system is comprised of a guide RNA that directs the Cas9 nuclease to create a double strand break (Rodriguez-Rodriguez et al. 2019). CRISPR/Cas9 performs sequence specific cleavage at the target site of the genome. This method could be used to test the functions of the candidate genes identified in this study.

For that, different CRISPR/Cas9 mutants can be generated with different knockout mutations and then we analyze the resulting phenotypes. Another interesting use of the CRISPR/Cas9 system would be the introduction of traits of interested into cultivated cowpeas by taking only the gene of interest from the donor parent. Genome editing will have an important role in studying the molecular mechanisms of domestication traits in cowpea.

This work presents major advances in cowpea research, future studies can benefit from the resources and results presented here.

References

Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. *Current opinion in plant biology* 2:109-113

Bett B, Gollasch S, Moore A, Harding R, Higgins TJV (2019) An Improved Transformation System for Cowpea (*Vigna unguiculata* L. Walp) via Sonication and a Kanamycin-Geneticin Selection Regime. *Frontiers in plant science* 10

Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo M-C, Alhakami H, Ounit R, Hasan AM, Verdier J, Roberts PA, Santos JRP, Ndeve A, Doležel J, Vrána J, Hokin SA, Farmer AD, Cannon SB, Close TJ The genome of cowpea (Vigna unguiculata [L.] Walp.). The Plant Journal https://doi.org/10.1111/tpj.14349

Ma X, Zhu Q, Chen Y, Liu Y-G (2016) CRISPR/Cas9 platforms for genome editing in plants: developments and applications. *Molecular plant* 9:961-974

Smith NA, Singh SP, Wang M-B, Stoutjesdijk PA, Green AG, Waterhouse PM (2000) Gene expression: total silencing by intron-spliced hairpin RNAs. *Nature* 407:319