UNIVERSITY OF CALIFORNIA
RIVERSIDE

Using Genomic Resources to Breed Cowpeas With Larger Seeds

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

Doctor of Philosophy

in

Genetics, Genomics, and Bioinformatics

by

Mitchell Ryan Lucas

December 2014

Dissertation Committee:
Dr. Timothy J. Close, Chairperson
Dr. Philip A. Roberts
Dr. Mikeal L. Roose

The Dissertation of Mitchell Ryan Lucas is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

Several individuals at the University of California Riverside deserve special recognition. My advisor Dr. Timothy J. Close and other members of our immediate cowpea team including Dr. Philip A. Roberts, Dr. Jeffrey D. Ehlers, Dr. Bao-Lam Huynh, Dr. Ndeye Ndack Diop, Dr. Jansen Rod Pereira Santos, Dr. Maria Munoz-Amatriain, Arsenio Ndeve, Sassoum Lo, Yi-ning Guo, and Savanah St. Clair. I appreciate the dedication of several of my instructors including Dr. Mikeal Roose, Dr. J. Giles Waines, Dr. Adam Lukaszewski, and Dr. Julia Bailey-Serres who also served on either my guidance committee, qualifying exam committee, or dissertation committee.

The Genetics, Genomics, and Bioinformatics graduate program provided an atmosphere for success and I value the flexibility offered to its students. Being a student at a land-grant institution meant I was able to access resources that became vital to my research and education including the Citrus Agricultural Research Station, the Coachella Valley Agricultural Research Station, and the University of California Kearney Agricultural Research Station.

Support for my graduate studies and research projects was provided by intramural and extramural sources including the University of California Riverside Chancellor's Distinguished Fellowship, the University of California Riverside Dissertation Year Program, the United States Department of Agriculture National Institute of Food and Agriculture Agriculturally Funded Research Initiative, the California Dry Bean Advisory Board, the Generation Challenge Program by the Consultative Group on International Agricultural Research, and the United States Agency for International Development.

The text of this dissertation, in part or in full, is a reprint of the material as it appears in: The Plant Genome [Cowpea-Soybean Synteny Clarified through an Improved Genetic Map, 2011]; Frontiers in Plant Science [Association Studies and Legume Synteny Reveal Haplotypes Determining Seed Size in Cowpea, 2013]; Frontiers in Plant Science [Introgression of Rare Haplotype from Southeastern Africa to Breed Cowpeas with Larger Seeds, 2014]; The Plant Genome [High-resolution SNP Genotyping Reveals a Substantial Problem Among Breeder Resources, 2013]; Molecular Breeding [Markers for Breeding Heat Tolerant Cowpea, 2013]; and Crop Science [Markers for Quantitative Inheritance of Resistance to Foliar Thrips in Cowpea, 2012]. The co-author Timothy J. Close listed in those publications directed and supervised the research which forms the basis for this dissertation. The co-authors Philip A. Roberts, Jeffrey D. Ehlers, and Bao-Lam Huynh provided technical expertise and helped in writing the manuscripts in which they were co-authors. The co-authors Ndiaga Cisse, Issa Drabo, Patricia d. Vinholes, Steve Wanamaker, and Ndeye-Ndack Diop provided technical expertise in the manuscripts in which they were co-authors.

Dedication

This dissertation is dedicated to my family.

ABSTRACT OF THE DISSERTATION

Using Genomic Resources to Breed Cowpeas With Larger Seeds

by

Mitchell Ryan Lucas

Doctor of Philosophy, Graduate Program in Genetics, Genomics, and Bioinformatics
University of California, Riverside, December 2014
Dr. Timothy J. Close, Chairperson

Cowpea (*Vigna unguiculata*) is a warm-season legume that is primarily cultivated for protein rich grain. Seed size is an important breeding target that distinguishes most domesticated crops from their wild relatives and is a particularly important trait for the grain legumes. This dissertation describes efforts to breed cowpea varieties with larger seeds using marker-assisted approaches to breeding. The first chapter of this dissertation describes the development of a consensus genetic map of 1,107 molecular markers which was constructed by analyzing bead-assay genotype data from 13 experimental populations. The content and organization of the cowpea genome was also compared to the genome of soybean (*Glycine max*) to describe regions of synteny. The second chapter utilizes the genetic map, legume synteny, and phenotypic information collected from field and greenhouse trials to develop associations between allelic variation and the inheritance of seed size. Several regions of the cowpea genome important for seed size

were found to be syntenic with regions of the soybean genome that were previously associated with the inheritance of seed size. These marker-trait associations are applied in the third chapter to breed cowpea varieties with up to 52% larger seeds which was accomplished by introgressing a haplotype from Southeastern Africa into the genetic background of a California blackeyed pea. Preliminary field screening identified introgression lines that also performed well for other important agronomic traits including yield, maturity, and plant architecture. The introgression lines developed in this work could be used as parents for deploying large seed size in other pedigrees and could be studied to better understand the impact of seed size on nutritional content and agronomic performance.

Table of Contents

List of Tables

List of Figures

Legumes: A Diverse Family of Important Plant Species

Legumes are a diverse group of flowering eudicots in the family Fabaceae. With more than 18,000 species, legumes represent the third largest plant family and are second only to cereals in terms of agricultural importance (Gepts et al., 2005; Lewis et al., 2005). There are three subfamilies within Fabaceae: Caesalpinioideae, Mimosoideae, and Papilionoideae. Based on fossil, morphological, and molecular evidence the Caesalpinioideae represent a paraphyletic group from which the Mimosoideae and Papilionoideae are derived (Kajita et al., 2001; Lavin et al., 2005; Wojciechowski et al., 2004). A few thousand tropical and subtropical tree and shrub species are members of the Caesalpinioideae genera (Pettigrew and Watson, 1977), including several important timber, ornamental, and food species like the Kentucky Coffee tree (*Gymnocladus dioicus*), flame tree (*Delonix regia*), and tamarind (*Tamarindus indica*), respectively. Plants in the Mimosoideae, specifically the tribes Acaciae and Ingeae, commonly have flowers with small petals and many stamens (Luckow et al., 2003). One example of the Mimosoideae genera is the sensitive plant (*Mimosa pudica*) which is interesting because it responds with rapid closing of leaves when it is touched or shaken. The most economically important subfamily of legumes is the Papilionoideae which can be divided into four important clades (Doyle and Luckow, 2003; Gepts et al., 2005). Genus Lupinus are members of the genistoids, while peanut (*Arachis hypogaea*) is the most economically important aeschynomenoid/dalbergioid clade member. Cool-season legumes including alfalfa (*Medicago sativa*), chickpea (*Cicer arietinum*), lentil (*Lens culinaris*), and pea (*Pisum sativum*) belong to the Hologalegina clade. Most important to

this dissertation are the warm-season legumes which are members of the Millettioids/Phaseoloids. Among the most widely grown grain legumes are warm-season annuals which include soybean (*Glycine max*), common bean (*Phaseolus vulgaris*), and the subject of this dissertation, cowpea (*Vigna unguiculata*). Other important Vigna species include adzuki bean (*V. angularis*), bambara groundnut (*V. subterranea*), mung bean (*V. radiata*), moth bean (*V. aconitifolia*), rice bean (*V. umbellata*), and urd bean (*V. mungo*). In comparison to their closest relatives, Phaseolous, they perform well in hot, arid environments with poor soils (Fery, 2002).

The sheer number of legume species represents a challenge for comparative genomics (Gepts et al., 2005) which would allow knowledge obtained from studying one species to be informative to relative species. Based on cytogenetic studies of root tips, most legumes are diploid (Goldblatt and Davidse, 1977). Reconstruction of legume phylogeny using molecular evidence suggests the Millettioids/Phaseoloids share a last common ancestor with other Papilionoideae ~59 million years ago (MYA) (Doyle and Luckow, 2003; Kajita et al., 2001; Wojciechowski et al., 2003). Comparisons of sequence data from Medicago, Lotus, and Glycine supports the division of the cool and warm season legumes (Choi et al., 2004). Knowledge of synteny has been important for developing and characterizing reference genomes for important legume species including soybean (Schmutz et al., 2009) and common bean (Schmutz et al., 2014). These comparisons have been less commonly studied among legumes that are of lesser importance to the Western world. However there are recent studies that utilize genome-

wide approaches (Chankaew et al., 2014; Tomooka et al., 2014) which update preliminary comparisons within Vigna (Menancio-Hautea et al., 1993), for instance.

There is increased attention on legumes due to their importance in food security and sustainable agriculture. Legumes are most notable for their protein rich grain and for their ability to fix atmospheric nitrogen through symbiotic relationships with soil dwelling bacteria which can improve soil health and can be utilized by subsequent crops, typically cereals. Legume and cereal nutritional profiles complement each other quite well which may explain why the two appear together in most centers of plant domestication (Gepts, 2004). Grain legumes are good sources of protein, complex carbohydrates, minerals, and B-vitamins (Gupta, 1987). They also provide sufficient sources of all essential amino acids except for sulfur-containing amino acids and tryptophan and are rich in lysine which is deficient in the nutritional profile of cereal grain (Iqbal et al., 2006). Unfortunately many grain legumes also produce anti-nutritional compounds including proteolytic inhibitors, phytohemagglutinins, lathyrogens, and cyanogenetic compounds (Gupta, 1987).

Cowpea: Prominence and Diversity

Cowpea crops are mainly harvested for dry grain, however, it is common for leaves and immature fruit to be eaten which are available early in the growing season. Significant cowpea growing regions include West Africa, South and East Africa, Brazil, Asia and the United States. Most abundant in Asia is the subspecies yard-long bean (*Vigna unguiculata* ssp. *sesquipedalis*) which is cultivated for its unusually long pods that are harvested before grain filling and cooked like a fresh vegetable. West African

countries including Nigeria, Burkina Faso, Ghana, Senegal, and Niger produce the most cowpea grain in terms of quantity and diversity of seed types which they depend upon as a staple food source. In these regions it is common for the grain to be cooked and consumed whole or to be processed into a flour for culinary preparations that include frying or steaming (i.e. akara and moin-moin). For crops destined to be used for production of flour, varieties with large white seeds and rough seed coats are desired because they produce pure colored flour and the seed coats are easily removed (Egbadzor et al., 2013; Singh et al., 2000). This is in contrast to the Western world where cowpea is relatively less important. Cowpea production peaked in 1937 with approximately 2.4 million hectares in cultivation and has subsequently declined due to the availability of other forage crops and mechanized harvesting equipment available to harvest these newer crops (Fery, 1990, 2002). Blackeyed peas, crowder, and cream types dominate dry grain production and have unique flavors and visual appeal. Vegetable types are popular among home gardeners which include fresh-shelled purple-hull pinkeye, yard-long, and ram's-horn varieties.

As a species cowpea is well adapted to production in hot and drought-prone environments. This is expected since the primary region of cowpea production spans the Sudano-Sahelian regions of West Africa and because the two primary genepools of domesticated cowpea are centered in arid regions of West and South Eastern Africa (Huynh et al., 2013). Because cowpea evolved in Africa it has many co-evolved pests including bacterial, fungal, insect, nematode, parasitic plant, and viral pathogens that constrain production in areas of the world where it is in high demand. Drought adaption

and pest and disease resistance continues to be primary foci of most cowpea research. It is important for new varieties to couple these exotic traits with seed qualities desired by consumers. Fortunately there is a substantial amount of genetic variation available to develop improved cultivars. This provides opportunities to understand mechanisms behind stress responses and to improve a food source that is critical to developing countries and capable of production in marginal environments. In terms of germplasm collections the International Institute of Tropical Agriculture maintains the most comprehensive collection, followed by the United States Department of Agriculture, and the University of California, Riverside (UCR) (Varshney et al., 2009).

Genomic Resources to Support Cowpea Breeding

A legacy of more than 38-years of collaborative cowpea research between UCR and national agricultural research stations in Africa began with commitments between emeritus faculty member Dr. Anthony E. Hall of UCR and Dr. Ndiaga Cisse of the Insititut Senegalais de Recherches Agricoles in Senegal. The team has expanded considerably and support from the United States Agency for International Development (USAID) has been continuous since 1980 which has resulted in the development of improved cultivars and training in approaches to crop improvement (Hall et al., 2003). Most recently, cowpea initiatives supported by the USAID and the Generation Challenge Program, administered by the Consultative Group on International Agricultural Research, have invested in molecular approaches to cowpea breeding.

Advances in genomic technologies have changed the way breeding is done. In private industry, knowledge provided by molecular markers is routinely accessed to

enhance breeding initiatives and to protect intellectual property, especially for crops that are intensively bred. This isn't true for many 'orphan' crops which have, until recently, lagged behind in terms of genomic technology development and application (Varshney et al., 2009).

The genetic system of cowpea makes it simple to study. Cowpea is diploid with 11 unique nuclear chromosomes (2n=2x=22) and has a relatively small genome (Arumuganathan and Earle, 1991). Most cowpea varieties reproduce quickly (some in as little as 55 days), it grows well in greenhouses, its flowers are large and easy to manipulate, and produce many seeds per pollination. These characteristics facilitated the timely development of experimental tools and populations. Cowpea research was propelled into the genomics era following the development of a single-nucleotide polymorphism (SNP) genotyping assay and its application to many experimental populations of bi-parental design (Muchero et al., 2009). Originally developed using a 1,536 Illumina GoldenGate SNP genotyping platform, this technology has been used to describe genetic diversity (Egbdazor et al., 2014; Huynh et al., 2013; Xu et al., 2012), the inheritance of traits (Agbicodo et al., 2010; Egbadzor et al., 2013b; Lucas et al., 2012, 2013a; Massimo, 2011; Muchero et al., 2011, 2013; Pottorff et al., 2012a, 2012b, 2014; Xu et al., 2011a, 2013), to validate pedigrees (Lucas et al., 2013b), for comparative genomics (Muchero et al., 2009; Xu et al., 2011b), and has been applied in marker-assisted approaches to breeding. Other important genomic resources include databases that provides access to a physical map (http://phymap.ucdavis.edu/cowpea), 183,000 EST sequences, sequences for a minimal tiling path of 4,300 bacterial artificial chromosomes,

and whole genome shotgun assemblies all available at HarvEST:Cowpea (http://harvest.ucr.edu).

Trajectory of the Dissertation

The aims of this dissertation includes the development and application of genomic resources for researching cowpea genetic diversity and breeding. Specifically the development of an improved consensus genetic map and characterization of synteny between cowpea and soybean, associations between SNPs and the inheritance of seed size, and their application in marker-assisted breeding in an attempt to breed cowpea varieties with larger seeds. The appendices of this dissertation describe other applications of genotyping to validate pedigrees, study the inheritance of heat tolerance during reproductive development, and studies concerning the inheritance of resistance to feeding damage caused by foliar thrips.

References

Agbicodo EM, Fatokun CA, Bandyopadhyay R, Wydra K, Diop NN, Muchero W, Ehlers JD, Roberts PA, Close TJ, Visser RGF, and van der Linden VG (2010) Identification of markers associated with bacterial blight resistance loci in cowpea [Vigna unguiculata (L.) Walp.]. Euphytica 175 (2): 215 - 226.

Arumuganathan K, and Earle ED (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9 (3): 208 - 218.

Chankaew S, Isemura T, Isobe S, Kaga A, Tomooka N, Somta P, Hirakawa H, Shirasawa K, Vaughan DA, and Srinives P (2014) Detection of genome donor species of neglected tetraploid crop Vigna reflexo-pilosa (créole bean), and genetic structure of diploid species based on newly developed EST-SSR markers from azuki bean (Vigna angularis). PloS one 9 (8): e104990.

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, and Cook DR (2004) Estimating genome conservation between crop and model legume species. Proceedings of the National Academy of Sciences 101 (43): 15289 - 15294.

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiology 131**:** 900 - 91.

Egbadzor KF, Yeboah M, Offei SK, Ofori K, and Danquah EY (2013a) Farmers key production constraints and traits desired in cowpea in Ghana. Journal of Agricultural Extension and Rural Development 5: 14 - 20.

Egbadzor KF, Yeboah M, Danquah EY, Ofori K, and Offei SK (2013b) Identification of SNP markers associated with seed size in cowpea [Vigna unguiculata (L.) Walp]. International Journal of Plant Breeding and Genetics 7 (2): 115 - 123.

Egbadzor KF, Ofori K, Yeboah M, Aboagye LM, Opoku-Agyeman MO, Danquah EY, and Offei SK (2014) Diversity in 113 cowpea [Vigna unguiculata (L) Walp] accessions assessed with 458 SNP markers. Springer Plus 3 (1): 541.

Fery, RL (1990) The cowpea: Production, utilization, and research in the United States. Horticultural Reviews 12: 197 - 222.

Fery, RL (2002) New opportunities in Vigna. Trends in New Crops and New Uses: 424 - 428.

Gepts P, Beavis WD, Brummer CE, Shoemaker RC, Stalker HT, Weeden NF, and Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiology 137 (4): 1228 - 1235.

Gepts P (2004) Domestication as a long-term selection experiment. Plant Breeding Reviews 24: 1 – 44.

Goldblatt P, and Davidse G (1977) Chromosome numbers in legumes. Annals of the Missouri Botanical Garden: 121 - 128.

Gupta,YP (1987) Anti-nutritional and toxic factors in food legumes: a review. Plant Foods for Human Nutrition 37(3): 201 - 228.

Hall AE, Cisse N, Thiaw S, Elawad HOA, Ehlers JD, Ismail AM, Fery RL, Roberts PA, Kitch LW, Murdock LL, Boukar O, Phillips RD, and McWatters KH (2003)

Development of cowpea cultivars and germplasm by the Bean/Cowpea CRSP. Field Crops Research 82 (2): 103 - 134.

Huynh BL, Close TJ, Roberts PA, Hu Z, Wanamaker S, Lucas MR, Chiulele R, Cisse N, David A, Hearne S, Fatokun C, Diop N, and Ehlers JD (2013) Gene pools and the genetic architecture of domesticated cowpea. The Plant Genome 6 (3).

Iqbal A, Khalil IA, Ateeq N, and Khan MS (2006) Nutritional quality of important food legumes. Food Chemistry 97 (2): 331 - 335.

Kajita, T, Ohashi H, Tateishi Y, Bailey CD, and Doyle JJ (2001) rbcL and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. Systems. Botany 26: 515 - 536.

Lavin M, Herendeen PS, and Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Systematic Biology 54 (4): 575 - 594.

Lewis G, Schrire B, Mackinder B, and Lock M (eds.) (2005) In: Legumes of the world. Royal Botanic Gardens, Kew 592.

Lucas MR, Ehlers JD, Roberts PA, and Close TJ (2012) Markers for quantitative inheritance of resistance to foliar thrips in cowpea. Crop Science 52: 2075 - 81.

Lucas MR, Ehlers JD, Huynh BL, Diop NN, Roberts PA, and Close TJ (2013a) Markers for breeding heat tolerant cowpea. Molecular Breeding 31: 529 - 36.

Lucas MR, Huynh BL, Ehlers JD, Roberts PA, and Close TJ (2013b) High-resolution single nucleotide polymorphism genotyping reveals a significant problem among breeder resources. The Plant Genome 6 (1).

Luckow M, Miller JT, Murphy DJ, and Livshultz T (2003) A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. Advances in legume systematics 10: 197 - 220.

Massimino C (2011) Exploring the Genetic Basis of Root Mucilage in Cowpea (Vigna unguiculata (L.) Walp.). The University of Maryland McNair Scholars Undergraduate Research Journal 3: 157 - 160.

Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D, and Young ND (1993) Comparative genome analysis of mungbean (Vigna radiata L. Wilczek) and cowpea (V. unguiculata L. Walpers) using RFLP mapping data. Theoretical and Applied Genetics 86 (7): 797 - 810.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, and 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.

Muchero W, Ehlers JD, Close TJ, and Roberts PA (2011) Genic SNP markers and legume synteny reveal candidate genes underlying QTL for Macrophoina phaseolina resistance and maturity in cowpea [Vigna unguiculata (L) Walp.]. BMC Genomics 12.

Muchero W, Roberts PA, Diop NN, Drabo I, Cisse N, Close TJ, Muranaka S, Boukar O, and Ehlers JD (2013) Genetic architecture of delayed senescence, biomass, and grain yield under drought stress in cowpea. PLoS One 8 (7): e70041.

Pettigrew CJ, and Watson L (1977) On the classification of Caesalpinioideae. Taxon: 57 - 64.

Pottorff M, Wanamaker S, Ma YQ, Ehlers JD, Roberts PA, and Close TJ (2012a) Genetic and physical mapping of candidate genes for resistance to Fusarium oxysporum f.sp. tracheiphilum race 3 in cowpea [Vigna unguiculata (L.) Walp]. PLoS One 7.

Pottorff M, Ehlers JD, Fatokun C, Roberts PA, and Close TJ (2012b) Leaf morphology in cowpea [Vigna unguiculata (L.)Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. BMC Genomics 13: doi:10.1186/1471-2164-13-234.

Pottorff M, Roberts PA, Close TJ, Lonardi S, Wanamaker S, and Ehlers JD (2014) Identification of candidate genes and molecular markers for heat-induced brown discoloration of seed coats in cowpea [Vigna unguiculata (L.) Walp]. BMC genomics 15 (1): 328.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht, J, Grimwood, J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178 - 183.

Schmutz, J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MMS,

Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, and Jackson SA (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genetics 46: 707 - 713.

Singh BB, and Ishiyaku MF (2000) Brief communication. Genetics of rough seed coat texture in cowpea. Journal of Heredity 91 (2): 170 - 174.

Tomooka N, Isemura T, Naito K, Kaga A, and Vaughan D (2014) Vigna Species. Broadening the Genetic Base of Grain Legumes: 175 - 208.

Varshney RK, Close TJ, Singh NK, Hoisington DA, and Cook DR (2009) Orphan legume crops enter the genomics era! Current Opinion in Plant Biology 12 (2): 202 - 210.

Wojciechowski MF, Lavin M, and Sanderson MJ (2004) A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. American Journal of Botany 91: 1846 - 1862.

Wojciechowski MF (2003) Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective. Advances in Legume Systematics 10: 5 - 35.

Xu P, Hu T, Yang Y, Wu X, Wang B, Liu Y, Qin D, Ehlers JD, Close TJ, Lu Z, and Li G (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. Horticultural Science 46 (8): 1102 - 1104.

Xu P, Wu X, Wang B, Hu T, Lu Z, Liu Y, Qin D, Wang S, and Li G (2013) QTL mapping and epistatic interaction analysis in asparagus bean for several characterized and novel horticulturally important traits. BMC Genetics 14 (1): 4.

Xu P, Wu X, Wang B, Luo J, Liu Y, Ehlers JD, Close TJ, Roberts PA, Lu Z, Wang S, and Li G (2012) Genome wide linkage disequilibrium in Chinese asparagus bean (Vigna. unguiculata ssp. sesquipedialis) germplasm: implications for domestication history and genome wide association studies. Heredity 109 (1): 34 - 40.

Xu P, Wu X, Wang B, Liu Y, Ehlers JD, Close TJ, Roberts PA, Diop NN, Qin D, Hu T, Lu Z, and Li G (2011b) A SNP and SSR based genetic map of asparagus bean (Vigna unguiculata ssp. sesquipedialis) and comparison with the broader species. PLoS One 6: e15952.

Chapter 1

Cowpea-Soybean Synteny Clarified through an Improved Genetic Map

Abstract

Linkage mapping is relevant to modern plant biology and provides a framework for downstream analyses including quantitative trait loci identification, map based cloning, assessment of diversity, association mapping, and molecular breeding. Here, we report a consensus genetic map of cowpea, *Vigna unguiculata*, and synteny to other legumes based on EST-derived SNPs. In total, 1,293 individuals representing 13 mapping populations were genotyped using an Illumina 1536 Golden Gate Assay. A consensus map containing 1107 EST-derived SNP markers (856 bins) on 11 linkage groups (680cM) was constructed from 13 population-specific maps. This effort combined six new population specific maps and seven revised population specific maps to construct an improved consensus map with 33% more bins, 19% more markers, and improved marker order when compared to the previous cowpea SNP consensus map. Comparative and whole genome visualizations are presented as a framework for discussing map quality and synteny with soybean.

Introduction

Cowpea, *Vigna unguiculata*, (2n=2x=22) is a leguminous crop cultivated for fresh and dry grains, leaves, and fodder. The crop is a valuable component of rotations and intercrops due to symbiotic nitrogen fixation (Quaye et al., 2009). Important cowpea producing regions span the globe; however, it is an especially valuable component of low-input farming systems in sub-Saharan Africa, South America, and Asia. Cowpea is popular in resource-poor farming because of its consistent production under abiotic stresses (drought, heat, low soil fertility) and in many regions it is a protein-rich component of an otherwise protein-poor diet (Ehlers et al., 1997; Hall, 2004).

Research and development of improved crop varieties encompasses advances in genomics and biotechnology. Consensus genetic maps are available for many model and important crop and animal species including soybean (Song et al., 2004), wheat (Somers et al., 2004), barley (Wenzl et al., 2006), and chicken (Groenen et al., 2000), and are central to breeding and diversity initiatives. Recently, technological advances that have substantially reduced costs of sequencing and genotyping promoted the development of genome resources for many non-model species (Varshney et al., 2009). Linkage mapping in cowpea has progressed with marker technology to yield informative and increasingly dense genetic maps (Menendez et al., 1997; Muchero et al., 2009; Ouedraego et al., 2002).

Prior to this work an Illumina 1536 Golden Gate SNP assay was developed and implemented to map 928 EST-derived SNPs in cowpea (Muchero et al., 2009). This map represented a substantial improvement over a previous but population-specific cowpea

map, which utilized 441 AFLP, RFLP, and RAPD markers (Ouedraego et al., 2002). The 2009 cowpea SNP map contained 645 bins with 928 markers arranged on 11 linkage groups (680cM) and was constructed by genotyping 632 individuals from six recombinant inbred line (RIL) populations. Both the 2009 consensus map and the improved map reported here are based entirely on EST-derived SNPs and all populations were genotyped for the same 1536 loci.

This SNP assay was also recently utilized in conjunction with SSR genotyping to develop the first genetic map of 'yard-long' or 'asparagus' bean, *Vigna unguiculata* ssp *sesquipedialis* (Xu et al., 2011). Comparative analyses between subspecies revealed macro-synteny across most linkage groups and demonstrated the utility of the previous consensus map.

Here we report a new consensus map containing 1107 EST-derived SNP markers (856 bins), which was developed by integrating thirteen population-specific maps. Improved methods of data analysis are realized in map characteristics and are apparent when surveying synteny of cowpea with soybean, *Medicago*, and *Arabidopsis* using HarvEST: Cowpea 1.27 (http://harvest.ucr.edu) and Circos (Krzywinski et al., 2009). Development of this highly robust genetic map is of value to ongoing projects including genome assembly, marker assisted breeding, QTL analysis, map-based cloning, and comparative genomics.

Materials and Methods

DNA Sources: The parents and progeny of 13 mapping populations were genotyped for 1536 SNPs using the Golden Gate Assay as previously described (Muchero et al., 2009). DNA isolation and preparation for genotyping also followed the methods described in Muchero et al. (2009). Detailed marker information and a plethora of other resources relevant to this work can be found online at harvest.ucr.edu.

Data Processing: Raw data from the assay were imported into an Illumina GenomeStudio V2010.3 Workspace for analysis using genotype module V1.8.4. Custom workspaces were created for each mapping population to optimize cluster positions. Genotype calls were exported from GenomeStudio as spreadsheets for further data processing.

Data processing before mapping included the removal of apparently rogue individuals which exhibited excessive heterozygosity, non-parental genotypes, or no-call data points. Standards for these parameters were determined empirically for each population by obvious break-points within the distribution of data. The parental phase of markers for which the parental genotype was uncertain was determined using the "Suspect Linkages" function within JoinMap4. Genotypically identical individuals among the mapping populations were identified using the "Similarity of Individuals" function in JoinMap4 and were removed prior to mapping. Only SNPs with minor allele frequency (MAF) > 0.25 and > 95% good calls were included for mapping.

Individual and Consensus Map Construction: Individual linkage maps representing 13 populations were constructed using JoinMap4. Eleven of these

populations were F_8 to F_{10} RIL populations developed by inbreeding and single seed descent, while the remaining two (IT84S-2246 x IT93K-503 and IT84S-2246 x Mouride) were F_3-derived F_4 families. These populations were selected on the basis of relevance to modern breeding programs, parental polymorphism, and segregation of agronomic traits. The mapping data sets used in Muchero et al. (2009) and Xu et al. (2011) were re-used in the present work following the additional clean-up steps summarized above. Marker grouping was determined using LOD thresholds ≥ 4 while marker order was determined using LOD thresholds ≥ 6. These individual maps were compared to each other and to the Muchero et al. (2009) map to determine spurious linkages. When referenced to the consensus map, only two markers were found among the individual maps that corresponded to different consensus linkage groups. In these two scenarios the ultimate linkage group assignment agreed with the most popular assignment among the individual maps. The confounding marker among the population specific maps in which the assignment disagreed was removed, and the population was re-mapped.

The consensus map was constructed by first comparing the 13 population-specific maps generated using JoinMap4 to define 11consensus linkage groups, each of which consisted of at least one linkage group from each population. Consensus linkage groups (VuLGs) were constructed one at a time using MergeMap (Wu et al., 2008). A coefficient was applied to all map coordinates to correct for the inflation of map distances introduced by MergeMap and to normalize the map to 680cM. Linkage group orientation was aligned to that of the Muchero et al. (2009) map for consistency and to facilitate comparisons between the two maps. MapChart (Voorrips, 2002) was used to align

linkage maps and Circos (Krzywinski et al., 2009) was used for additional visualization of map characteristics.

Synteny: Cowpea-soybean and cowpea-*Medicago* synteny was visualized using HarvEST: Cowpea 1.27 (http://harvest.ucr.edu) which determines synteny based on BLASTX scores (<e-10) between cowpea unigenes containing a mapped SNP and translated gene models from reference genomes. For soybean the JGI Glyma1database (http://phytozome.net) was used while for *Medicago* the *Medicago trunculata* HAPMAP Mt3.5 database was used. TAIR 10 was utilized in HarvEST 1.27 to determine cowpea-*Arabidopsis* synteny. Consensus map coordinates of cowpea unigenes were compared with chromosomal positions of soybean only if they contained at least five markers in common. Information was extracted from HarvEST to develop Circos diagrams.

Results

Population Specific Maps: Characteristics of the 13 individual maps used in the construction of the consensus map are given in Table 1. The number of RILs in each population ranged from 56 in the CB27 x UCR 779 population to 160 in the CB27 x IT82E-18 population, with an average of 99 individuals. The number of SNPs mapped per population ranged from 155 in the IT84S-2246 x IT93K-503 population to 560 in the CB27 x UCR 779 population, with an average of 364 SNPs mapped per population. Population-specific map sizes ranged from 302cM for IT84S-2246 x IT93K-503 to 710cM for 524B x IT84S-2049 with an average size of 576cM. The number of linkage groups for each population ranged from 14 (IT84S-2246 x IT93K-503) to 23 (CB27 x IT97K-556-6) with an average number of linkage groups per map of 17. Supplemental File 1 provides a pairwise comparison of SNPs common among the 13 mapping populations. On average 136 SNPs are shared between a pair of mapping populations, ranging from IT84S-2246 x IT93K-503 and LB30#1 x LB1162 #7 sharing 27 markers to CB27 x IT82E-18 and CB27 x UCR 779 sharing 290 markers.

The phases of 333 SNPs among all 13 mapping populations in which the parental genotypes were uncertain was inferred using the "Suspect Linkages" function within JoinMap4. For 159 of these loci it was determined that the original phase designation was incorrect. Phases were subsequently reversed and the linkage group containing them was re-mapped. The phases of the remaining 174 SNPs were determined to have been called correctly (by chance) and were not inverted. Sixty-three duplicated pairs of individuals, which were genotypically identical for the SNPs under consideration, were solved by

removing one of the contributing individuals. In addition, 120 individuals among the 13 mapping populations contained excessive numbers of non-parental genotypes and were removed prior to mapping. A synopsis of data processing for each mapping population can be found in the supplementary information, Supplemental File 2. Coordinates for the consensus map and the thirteen population-specific maps can be found online at harvest.ucr.edu.

Consensus Map: Eleven linkage groups were constructed spanning 680cM. Table 2 summarizes the number of markers mapped, the length (cM), and the number of bins for each VuLG. In total, 1107 markers were mapped across all 11 linkage groups with a range of 72 markers on VuLG8 to 203 markers on VuLG3. Linkage group length ranged from 45.2cM on VuLG9 to 92.4cM on VuLG3. Eight-hundred and forty-five bins were mapped on the 11 linkage groups with an average distance between bins of 0.79cM. This translates to an average of one bin per 733Kb of the cowpea genome.

Figure 1 provides a graphical view of the cowpea map using Circos (Krzywinkski et al., 2009) which includes the depiction of parameters characterizing map quality. In this figure five data tracks are drawn whose parameters are oriented so that maximum and minimum values are distal and proximal respectively to the ideogram's center. The first data track, a green histogram, symbolizes the average distance between bins for each VuLG where each white grid line represents a distance of 0.25cM. The second data track, a purple histogram, displays the average number of markers per bin for each VuLG where each white grid line represents a value of 0.5 markers per bin. The third and fourth data tracks share the same radial position of initiation and thus share the same grid axis

where the blue histogram overlaps the underlying red histogram and each white grid line represents 25 units. The blue and red histograms visualize the number of bins per VuLG and the number of markers per VuLG respectively The fifth data track, bands, reside within the VuLGs and depict the relative location of bins in each linkage group.

Synteny: In the soybean and *Medicago* genome sequences, homeologous genes were identified for 85% and 80% of the SNPs mapped in cowpea, respectively. Supplemental File 3 lists the soybean and *Medicago* chromosomes that are most syntenic with VuLGs based on the number of cowpea homeologs detected on syntenic chromosomes. All cowpea consensus linkage groups had syntenic regions on multiple soybean and *Medicago* chromosomes. VuLG1 and VuLG3 displayed synteny with seven of the twenty soybean chromosomes while VuLG8 was syntenic with three soybean chromosomes. When compared to the eight chromosomes of *Medicago*, VuLG2 and VuLG3 were syntenic with seven chromosomes while VuLG4, VuLG5, and VuLG7 were syntenic with only three chromosomes. All 11 VuLGs had regions with similarity to the five *Arabidopsis* chromosomes but co-linearity was nearly non-existent and major genome rearrangement was obvious. The current cowpea consensus map accounts for all but three of the 191 SNPs mapped by Xu et al. (2011) in *V. unguiculata* ssp. *sesquipedalis* (population LB30#1 x LB1162 #7) and the two maps are highly syntenic across all VuLGs. Figure 2 uses Circos to display a genome-wide comparative view of the cowpea consensus map and the 20 chromosomes of soybean. This figure colors links based on cowpea VuLG origin and traces more than 2200 relationships between the two legume species. Expanded views of synteny with soybean are available in Supplemental

File 4 for each of the 11 VuLGs. Figure 3 juxtaposes syntenic views of VuLG3 with soybean chromosome 17 using both the new and previous (Muchero et al., 2009) versions of the cowpea consensus map.

Discussion

The 13 populations included in the genotyping and consensus map construction (Table 1) were chosen because of their relevance to modern breeding programs and diversity, especially considering the broad range of important traits for which they segregate. These traits include aphid resistance, bacterial blight resistance, cowpea weevil resistance, drought tolerance, *Fusarium* wilt resistance, flower thrips and foliar thrips resistance, heat tolerance, individual grain weight, maturity, *Macrophomina* resistance, nematode resistance, Striga resistance, seedling cold tolerance, virus resistance, and yield components.

The LB30#1 x LB1162 #7 population is derived from a cross of two *V. unguiculata* ssp. *sesquipedalis* accessions from China, and not only shared the least amount of markers with the other maps but also contained a relatively low number of mapped markers (180) when compared to the average of all 13 populations (364). This is consistent with the initial mapping of this population recently published by Xu et al. (2011) where SSR and SNP markers were employed to construct the first genetic map of 'asparagus bean'. Their effort mapped many new loci which provided an additional perspective to the EST-derived SNP marker framework. Consensus maps constructed with and without the LB30#1 x LB1162#7 map were compared to determine the influence this population had on the consensus map. This map was included in the reported consensus map because it contributed three unique markers and did not impose conflicting marker assignments. It is important to consider the fact that the 1536-SNP assay used for genotyping was mainly developed for *V. unguiculata* ssp. *unguiculata* and

this ascertainment bias may influence the interpretation of diversity for a distinct subspecies for which it was not optimized. This consideration is also important in interpreting marker number comparisons in Supplemental File 1 because the number of SNPs mapped heavily influences the number of SNPs shared between populations. CB27 x IT82E-18 and CB27 x UCR 779 may have been expected to share the most markers not only because they share a parent (CB27), but the number of mapped SNPs for each of these populations was relatively high (430 and 560 respectively) when compared to the average (364).

High-throughput genotyping may have unexpected benefits to breeding programs as a quality filter. Data processing and cleanup prior to linkage mapping provides a map with better marker order which ultimately affects downstream analyses such as QTL mapping or map-based cloning. In our analysis we identified a number of rogue individuals. Rogues are defined as those individuals with excessive no-call rates, heterozygosity, or non-parental alleles. Highly homozygous individuals with significant numbers of non-parental alleles are likely to have arisen from labeling or contamination errors during the eight or more cycles of planting/ harvesting/seed cleaning and packaging operations required to develop an advanced RIL population. Individuals with heterozygosity greater than expected based on the level of inbreeding and significant number of non-parental alleles likely arose from out-crossing during the advancement of the RIL populations. Individuals with excessive heterozygosity, no-calls, and non-parental genotypes were excluded prior to importing data into JoinMap4. Another quality filter was to address genotypically identical individuals, which were removed prior to

mapping, and thus circumvented the possibility of biasing the map towards recombination events unique to those duplicated individuals. Supplemental File 2 summarizes these quality control parameters for each of the 13 populations used in genotyping and mapping.

Genotyping a relatively large number of loci, 1536, allows for conservative thresholds to be applied when deciding which markers to include. In addition to thinning the populations, we only included SNPs with a minor allele frequency greater than 0.25 and less than 5% no-calls. The "Suspect Linkages" tab in JoinMap4 was utilized to identify SNPs with incorrect parental phase, which was subsequently corrected by inverting the genotype calls (A to b and B to a). These situations may have arisen due to a no-call, low quality call, incorrect call, or monomorphic call for the parents and the marker is otherwise successful among the progeny. A monomorphic SNP among the parents that is polymorphic in the population can be attributed to the inability to genotype individuals genotypically identical to the true parents. After inversion, those SNPs for which the genotypes were inverted did not appear as "Suspect Linkages". This genotype inversion was implemented on 159 of 333 SNPs for which the parental genotype was previously unknown. This ratio matches expectations that the parental phase was called correctly by chance for approximately 50% of the 333 SNPs. Supplemental File 2 summarizes the results of these marker clean-up efforts. These improvements are reflected in marker order as indicated by improved co-linearity when observing synteny with soybean, Figure 3. However, this may be an incomplete comparison of map quality between the two consensus maps because other aspects besides data processing also were

variable (number of mapping populations, mapped SNPs, etc…); nonetheless improvements such as marker and bin density are also apparent.

This improved cowpea consensus map contains 1107 markers, a 19% increase in marker density compared to the previously reported 928 SNP consensus map (Muchero et al., 2009). Not only were 179 new markers mapped but the number of informative positions, bins, also increased. In total 211 bins, an increase of 33%, were added over all 11 VuLGs with a range of 12 bins on VuLG4 and 34 bins on VuLG3, and an average increase of 19 bins per linkage group. The average distance between bins was reduced from 1.05cM to 0.79cM. The total number of bins on a linkage map is a function of the number of individuals genotyped. By increasing the number of individuals the probability of observing unique meiotic crossover also increases. Traditionally the number of markers on a genetic map is the most popular statistic to report; however, when discussing map resolution bin statistics are more relevant. This map and the Muchero et al. (2009) map are somewhat unique when compared to other consensus genetic maps. Our approach was able to not only map expressed genes but also to be confident in their placement. Prior to the availability of high-throughput SNP genotyping typical consensus maps integrated different marker types (SSRs, RAPDs, RFLPs, AFLPs, SNPs) even when only a few were shared among populations. This approach may yield more markers but the accuracy of marker order and distances, critical aspects of map quality, were compromised. Any one of the 13 populations used in the construction of this new cowpea consensus map shared on average 37% of its markers with any other map. This simplified approach to consensus mapping may yield more robust consensus maps when compared

to maps that were constructed from only a few populations or those that used multiple marker systems.

Because they are derived from gene transcripts, the mapped SNPs can be used for explorations of structural and functional synteny across species. Ancestral soybean genome duplication is evident due to the visualization of cowpea haplotype blocks being syntenic with more than one soybean chromosome, Supplemental File 4. Chromosomal rearrangements were also observed between cowpea and soybean and can be easily visualized using the HarvEST:Cowpea comparative genome viewer or the Circos diagrams included in the Supplemental File 4. Of the 1107 total mapped SNPs, 941 SNPs representing 85% of the genome had homeologs and exhibited synteny and co-linearity with soybean. Supplemental File 4 VuLG4 provides an example of a common pattern observed when scanning synteny with soybean one linkage group at a time. Syntenic blocks between cowpea and soybean often span an entire linkage group of cowpea, while in soybean the homeologous region is most often found on a single arm of a chromosome. This observation may indicate a similar mechanism of genome evolution to that which was recently described in grasses (Murat et al., 2011) where centomeric/telomeric recombination led to nested chromosome fusions and synteny break points. Current HarvEST:Cowpea resources indicate that from a genome perspective cowpea is more similar to soybean than to *Medicago.* However, the relationship is complex in that the ancestral genome of modern soybean underwent various modifications including duplication and diversification. As expected and previously observed (Muchero et al., 2009) *Arabidopsis*-cowpea synteny is complicated by extensive

chromosomal rearrangements, however, microsynteny with this model dicot is still informative. When comparing the current cowpea consensus map to the 'asparagus bean' map (Xu et al., 2011) large scale genome conservation was observed. All 11 VuLGs were syntenic with this subspecies specific map and these relationships conveyed the very close evolutionary relatedness of the two subspecies.

The utility of this new cowpea consensus map complements ongoing genome sequencing and map-based cloning efforts. Whole genome sequence information regarding the context of these ESTs could provide insight into regulatory regions and splice junctions. In many genome sequencing projects the use of genetic maps is a popular tool for accurate assembly and genome finishing. This consensus map will help place sequence scaffolds to make a whole-genome assembly of cowpea. Analyses dependent upon accurate and dense genetic maps including marker-assisted breeding, QTL analysis, and map-based cloning should consider map quality an important factor. In future work the utilities of Circos could be exploited to help infer an ancestral legume genome, which could promote discussion concerning the molecular mechanisms involved during the evolution of this important plant family.

References

Ehlers JD and Hall AE (1997) Cowpea (*Vigna unguiculata* L. Walp). Field Crops Research 53: 187 - 204.

Groenen MA, Cheng HH, Bumstead N, Benkel BF, Briles WE, Burke T, Burt DW, Crittenden LB, Dodgson J, Hillel J, Lamont S, Ponce de Leon A, Soller M, Takahashi H, and Vignal A (2000) A consensus linkage map of the chicken genome. Genome Research 10: 137 - 147.

Hall AE (2004) Breeding for adaptation to drought and heat in cowpea. European Journal of Agronomy 21: 447 - 454.

Krzywinski M, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, and Marra MA (2009) Circos: An information aesthetic for comparative genomics. Genome Research 19: 1639 - 1645.

Menendez CM, Hall AE, and Gepts P (1997) A genetic linkage map of cowpea (*Vigna unguiculata*) developed from a cross between two inbred, domesticated lines. Theoretical and Applied Genetics 95: 1210 - 1217.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, and 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.

Murat F, Xu JH, Tannier E, Abrouck M, Guilhot N, Pont C, Messing J, and Salse J (2011) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. Genome Research 20: 1545 - 1557.

Ouedraego JT, Gowda BS, Jean M, Close TJ, Ehlers JD, Hall AE, Gillaspie AG, Roberts PA, Ismail AM, Bruening G, Gepts P, Timko MP, and Belzile FJ (2002) An improved genetic linkage map for cowpea (Vigna unguiculata L.) Combining AFLP, RFLP, RAPD, biochemical markers, and biological resistance traits. Genome 45: 175 - 188.

Somers DJ, Isaac P, and Edwards K (2007) A high density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics 109: 1105 - 1114.

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theoretical and Applied Genetics 109: 122 - 128.

Quaye W, Adofo K, Madode YE, and Abizari AR (2009) Exploratory and multidisciplinary survey of the cowpea network in Tolon-Kumbungu district of Ghana: A food sovereignty prospective. African Journal of Agricultural Research 4: 311 - 320.

Van Ooijen JW (2006) JoinMap4: Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V. Wageningen, Netherlands.

Varshney RK, Close TJ, Singh NK, Hoisington DA, and Cook DR (2009) Orphan legume crops enter the genomics era. Current Opinion in Plant Biology 12: 202 - 210.

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93: 77 - 78.

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesna J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ,

Chalmers KJ, Kleinhofs A, Huttner E, and Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics **7**: 206.

Wu Y, Close TJ, and Lonardi S (2008) On the accurate construction of consensus genetic maps. Computational Systems Bioinformatics 7: 285 - 296.

Xu Pei, Wu X, Wang B, Liu Y, Ehlers JD, Close TJ, Roberts PA, Diop NN, Qin D, Hu T, Lu Z, and Li G (2011) A SNP and SSR based genetic map of asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*) and comparison with the broader species. PLoS One 6: e15952.

Table 1: 13 Mapping populations used in the construction of the consensus map.

Population	Individuals Genotyped	Individuals used for Mapping	Mapped SNPs	Linkage Groups	Map Size (cM)
CB27 x IT97K-566-6	95	92	438	23	505.56
CB27 x IT82E-18	166	160	430	23	701.15
CB27 x UCR 779	58	56	560	22	489.40
CB46 x IT93K-503-1	130	114	374	17	639.59
524B x IT84S-2049	91	85	438	22	710.09
Dan Ila x TVu-7778	113	79	288	22	549.56
Yacine x 58-77	141	97	435	22	650.98
Sanzi x Vita 7	142	122	413	19	753.22
IT84S-2246 x IT93K-503	93	88	155	14	302.46
IT84S-2246 x Mouride	92	87	347	15	595.33
TVu14676 x IT84S-2246-4	147	136	345	14	666.89
CB27 x 24-125B-1	108	87	329	23	526.75
LB30#1 x LB1162 #7	95	90	180	20	409.94

Table 2: Comparison of the current cowpea consensus map and the Muchero et al (2009) map.

Consensus VuLG	Markers		Length (cM)		Bins	
	Current map	Muchero et al 2009	Current map	Muchero et al 2009	Current map	Muchero et al 2009
1	81	69	64.7	85.2	65	50
2	142	116	74.4	84.0	98	74
3	204	168	92.4	81.8	154	120
4	83	68	45.7	66.4	67	55
5	93	75	58.2	62.6	71	56
6	109	93	79.1	59.1	84	60
7	82	72	48.1	52.9	64	43
8	72	65	63.9	49.0	65	48
9	78	66	45.2	48.4	60	47
10	90	77	62.4	46.0	73	53
11	73	59	46.0	44.8	55	39
Total	1107	928	680	680	856	645

Figure 1: Consensus genetic map of cowpea and parameters depicting map characteristics. (A) Average distance between bins (0.25cM). (B) Average number of markers per bin (0.5units). (C) Number of bins (25units). (D) Number of markers (25units). (E) Bin locations. (C and D) begin at the same radial position.

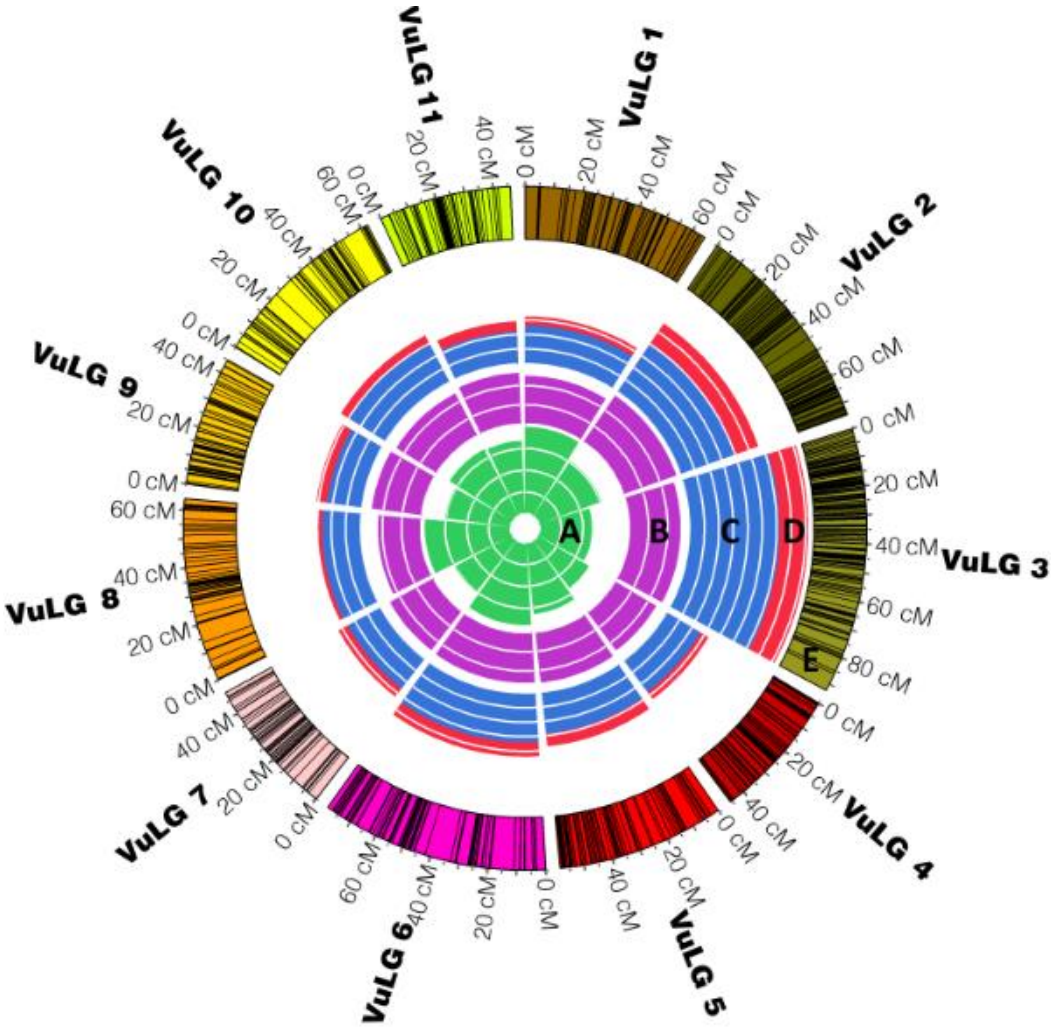


Figure 2: Genome view of synteny between cowpea linkage groups (VuLG) and soybean chromosomes (Gm). Links connect locations of homeologs between genomes.

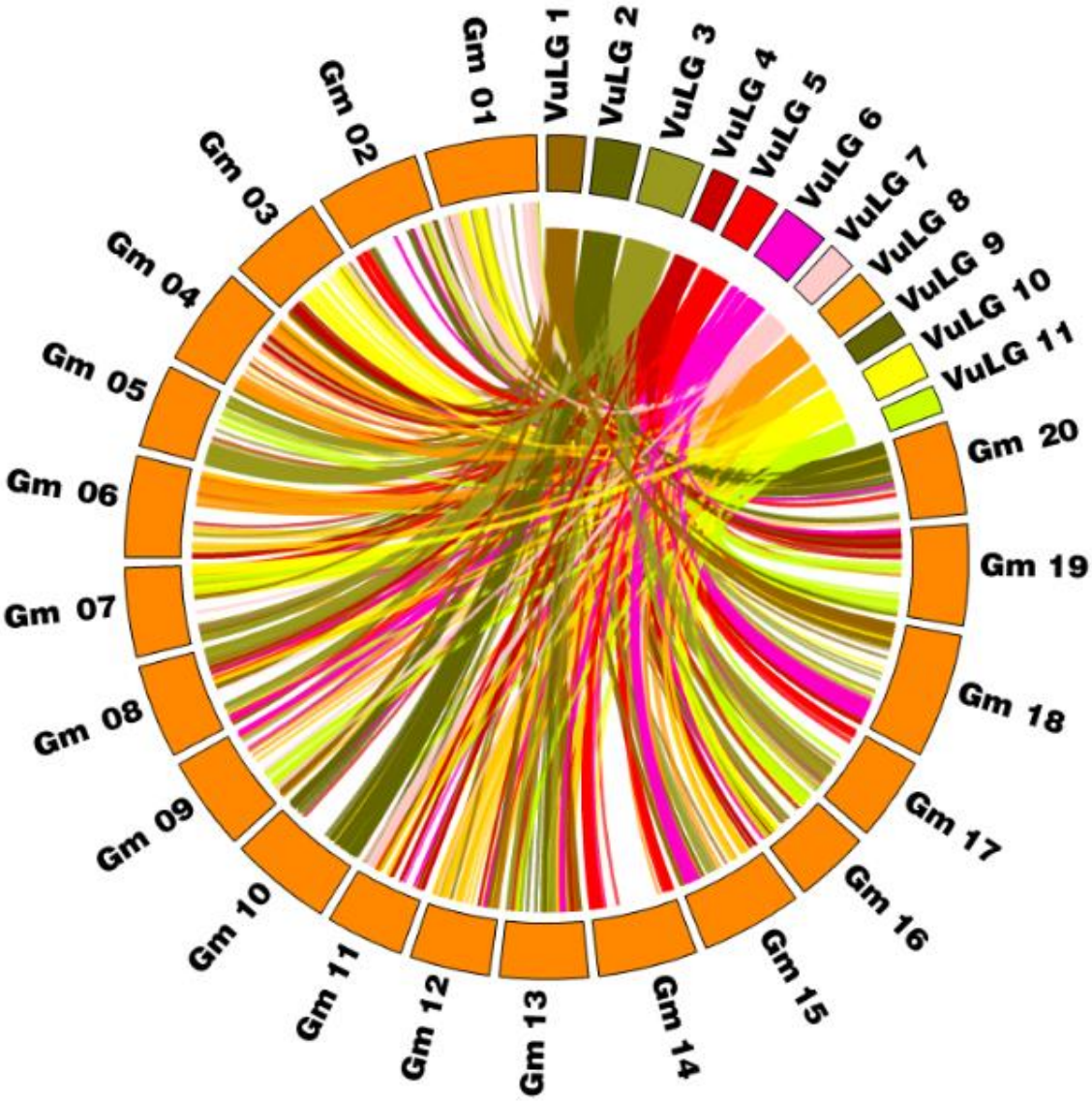
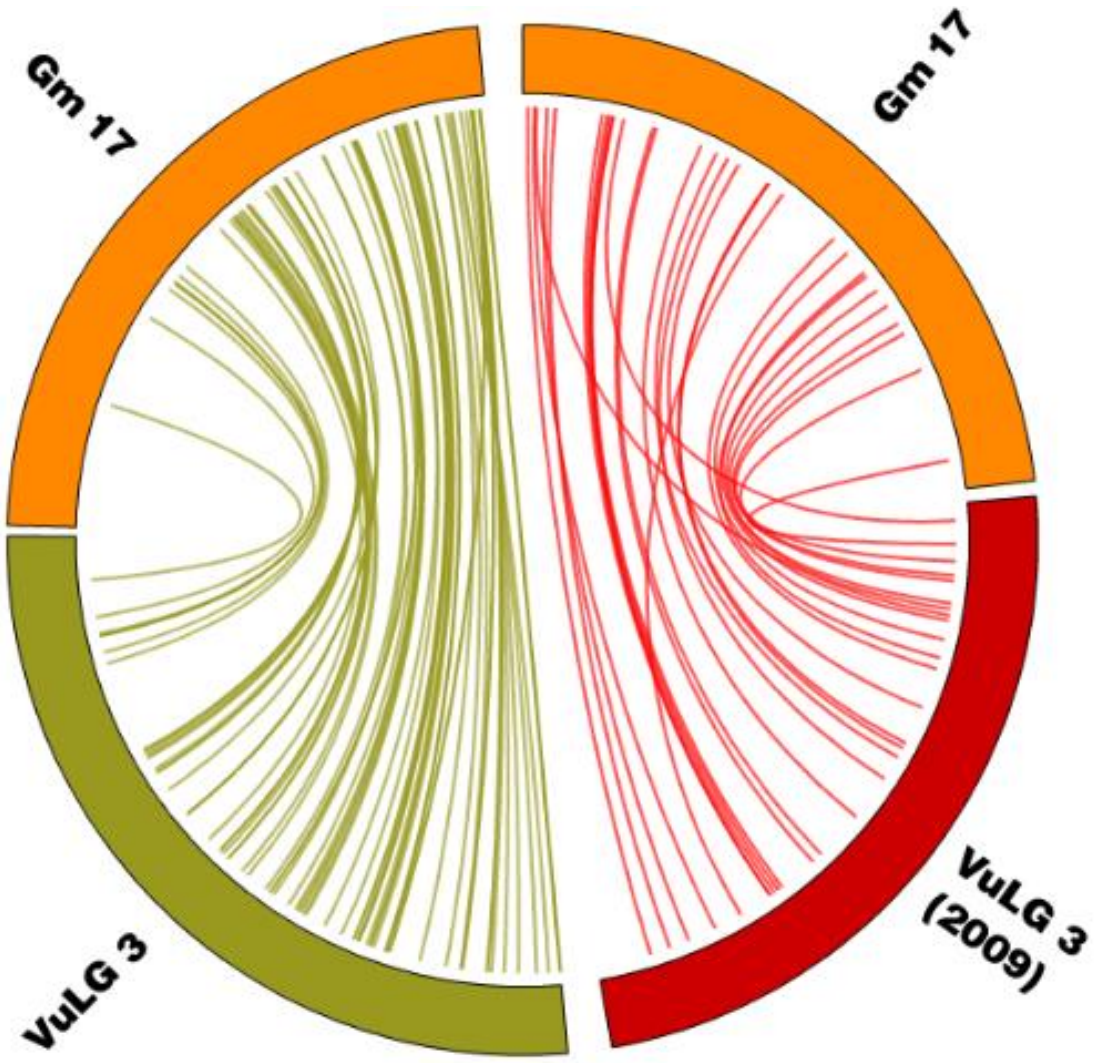


Figure 3: Current (green) and previous (red) cowpea consensus linkage group 3 (VuLG3), and synteny to soybean chromosome 17 (Gm17). Links connect locations of homeologs between genomes.



Chapter 2

Association Studies and Legume Synteny Reveal Haplotypes Determining Seed Size in Cowpea

Abstract

Highly specific seed market classes for cowpea and other grain legumes exist because grain is most commonly cooked and consumed whole. Size, shape, color, and texture are critical features of these market classes and breeders target development of cultivars for market acceptance. Resistance to biotic and abiotic stresses that are absent from elite breeding material are often introgressed through crosses to landraces or wild relatives. When crosses are made between parents with different grain quality characteristics, recovery of progeny with acceptable or enhanced grain quality is problematic. Thus genetic markers for grain quality traits can help in pyramiding genes needed for specific market classes. Allelic variation dictating the inheritance of seed size can be tagged and used to assist the selection of large-seeded lines. In this work we applied 1,536-plex SNP genotyping and knowledge of legume synteny to characterize regions of the cowpea genome associated with seed size. These marker-trait associations will enable breeders to use marker-based selection approaches to increase the frequency of progeny with large seed. For 804 individuals derived from eight bi-parental populations, QTL analysis was used to identify markers linked to ten trait determinants. In addition, the population structure of 171 samples from the USDA core collection was identified and incorporated into a genome-wide association study which supported more than half of the trait-associated regions important in the bi-parental populations. Seven of the total ten QTLs were supported based on synteny to seed size associated regions identified in the related legume soybean. In addition to delivering markers linked to major trait determinants in the context of modern breeding, we provide an analysis of the

diversity of the USDA core collection of cowpea to identify genepools, migrants, admixture, and duplicates.

Introduction

Cowpea is a warm-season legume grown throughout the tropics and several areas of the subtropics. West African countries led by Nigeria and Niger produce 70% of the world's crop on 10 million ha (FAOSTAT, 2013). The North-Eastern region of Brazil is the second largest region of production, followed by Eastern and Southern Africa, South Asia and North America. As is the case for other grain legumes, farmers' and market acceptance of cowpea are driven by the visual appearance of the grain. In most markets, large seed size is desirable and this is reflected in price premiums for large cowpea grain. Across Africa a diversity of grain sizes and colors exist which have varying importance in local or regional contexts. In West Africa the two most important grain types are large white or brown with rough seed coat texture, while in East and Southern regions of Africa relatively smaller seeds with smooth texture and brown to red color predominate in markets. In the Western United States, Southern Europe and the Middle-East the 'blackeyed pea' cowpea predominates. This type of cowpea is characterized by a large grain and white seed coat with a pigmented 'eye' around the hilum. Figure 4 displays a diversity of cowpea seed types. "Fresh-shell" varieties are also desired which are harvested before maturity for their large seed that can be easily removed from green pods. Consumer preference primarily demands large seed when grown for grain; however, small seed is preferred when seed is sold by volume for use as a fodder or cover crop.

Seed size has several agronomically important impacts. Large seeded cowpea have enhanced emergence when planted deep (up to 5 cm), tend to emerge earlier, and

produce larger plants during early development (Lush and Wien, 1980). In contrast, while large seeds typically have advantages over small seeded competitors (Wulff, 1986), small seeds are desirable for early drought conditions because they are able to transpire less water relative to their ability to reach water supplies (Hendrix et al., 1991).This may be particularly important for semi-arid rain fed growing regions.

Seed size is a very stable component of grain yield with high heritability for many crop plants including wheat (Giura and Saulescu, 1996), soybean (Cober et al., 1997), cowpea (Drabo et al., 1984), and mung bean (Fery et al., 1980). Several genes are known to impact the inheritance of seed size in cowpea. Drabo et al. (1984) proposed that at least eight loci contribute to the quantitative inheritance of seed size and Fatokun et al. (1992) identified two major, unlinked genomic regions, one of which is orthologous to a seed size QTL in mung bean. The orthology of this locus was later confirmed by its identification and association to seed size in soybean (Maughan et al., 1996). Exploration of legume synteny for cowpea trait characterization continues to be a rewarding approach that has also been used to better describe resistance to fungal pathogens (Muchero et al., 2011; Pottorff et al., 2012a), tolerance to heat during reproductive development (Lucas et al., 2012a), and leaf morphology (Pottorff et al., 2012b).

The introgression of novel traits from diverse collections typically compromises seed size among progeny. Because of the importance of grain size in market appeal, recovery of adequate grain size is an important objective following elite x exotic crosses. Wide crosses are commonly pursued to help deliver new varieties with enhanced resistance to biotic and abiotic stress. Several cycles of backcrossing help recover elite

characteristics including seed size; however, this process can be cumbersome and inefficient due to possible linkage drag and the polygenic nature of the trait. To help improve the selection of desirable lines we developed associations between genic SNP markers and seed size using experimental populations, a diversity collection, and knowledge of legume synteny.

Materials and Methods

Phenotype Data: Seed size was calculated as weight per 100 seed. The seeds we measured were harvested from plants grown under favorable conditions whether in the field or in the greenhouse. This means that plants were well watered and treated with pesticides as needed. The populations which were used are presented in Table 3 and are among those used to develop the consensus genetic map of cowpea (Lucas et al., 2011). All populations were at least at the F_8, except the IT84S-2246 x Mouride population which was phenotyped and genotyped at the F_4 generation. All eight populations were grown in the greenhouse, while the CB27 x IT82E-18 and CB46 x IT93K-503 populations were also grown in field trials. The CB27 x IT82E-18 population was grown during the summers of 2010 and 2011 at the University of California Riverside Citrus Experiment Station in Riverside, CA. The CB46 x IT93K-503 population was grown during the summer of 2008 at two field stations led by 1) the Senegalese Institute of Agricultural Research (ISRA/CNRA) in Bambey, Senegal and 2) the Insitut de l'Environnement et de Recherches Agricoles (INERA) at Kamboinse, Burkina Faso. In field trials ~100 seeds per 6-meter plot were planted in four replicates for each sample. In all greenhouse and field trials mature pods were harvested and dried for storage (< 15% moisture). Seeds were subsequently cleaned from the pods, counted, and weighed to determine the weight of a random sample of one hundred seeds. Seed size data provided online by Germplasm Resources Information Network (USDA-ARS, 2013) was used for genome-wide association mapping.

Genotype Data: The 1,536-plex EST-derived SNP genotype data used to build the consensus genetic map of cowpea (Lucas et al., 2011) was also used to perform QTL analyses of the eight bi-parental populations. Genotype data for 171 individuals of the USDA core collection of cowpea (USDA Core) (Gillaspie et al., 1996) were also obtained using the 1,536-plex genotyping platform developed by our group (Muchero et al., 2009). SNP calls were exported for further processing from the Illumina GenomeStudio software (Illumina, 2010). Rogue individuals among the bi-parental populations which were described in Lucas et al. (2013) were removed prior to QTL analysis. Similarly, genotype data for the USDA Core were used to identify and remove duplicate individuals using ParentChecker (Hu et al., 2012). ParentChecker was also helpful for formatting files for downstream analyses. SNPs were filtered on the basis of minor allele frequency (>0.20 for QTL and >0.10 for GWAS and analysis of population structure) to develop a set of polymorphic markers appropriate for analyses. The genotype data for the USDA core are provided in Supplemental File 5.

Marker-Trait Associations: QTL IciMapping (Li et al., 2008) was used to perform inclusive composite interval mapping for seed size based on one-hundred seed weight data from eight bi-parental mapping populations. In a method similar to Lucas et al. (2012a), the genetic map used for QTL analyses was a composition of population specific map marker orders and distances, and consensus linkage groups assignments. Regions of the genome contributing major QTL were identified after considering 1) regions with LOD scores > 3.0; 2) effect size >15% of phenotypic variance explained; 3) marker density; 4) span of the trait-associated region; 5) discovery in multiple

populations or via GWAS; 6) haplotype consistency when QTL were discovered in multiple populations; and 7) homology with trait-associated regions in soybean. The potential effect of stacking favorable alleles for multiple QTL was also investigated by grouping lines based on their QTL composition. This was done for populations in which multiple QTLs were discovered. Individuals with no, one, or several favorable alleles underlying the seed size QTLs we report were grouped and the average seed size was determined for that group and compared to the population average. A single factor analysis of variance was performed to determine if differences in seed size were due to QTL content. The ICIM-EPI function within QTL IciMapping (Li et al., 2008) was used to search for QTL interactions.

Six-hundred and sixty-five EST-derived SNP markers with minor-allele frequency >0.10 that were located among unique bins (one marker per bin) of the cowpea consensus genetic map were used to identify population structure of the subset of the USDA core. STRUCTURE (Pritchard et al., 2000) was used with BURNIN = 10,000 and NUMREPS = 50,000, with five runs of K = 1 – 15. The Evanno method (Evanno et al., 2005) facilitated by STRUCTURE HARVESTER (Earl and von Holdt, 2012) was used in addition to CLUMPP (Jakobsson and Rosenberg, 2007) and DISTRUCT (Rosenberg, 2004) to reconcile genepools on the basis of geographic collection information provided online by GRIN (USDA-ARS, 2013). Whole genome ancestry estimates (Q-matrix) computed from multiple STRUCTURE runs by CLUMPP were used as a covariate in the generalized linear model of association mapping provided by TASSEL 3.0 (Bradbury et

al., 2007). Markers that showed –Log (P-Values) > 3.0 that were also identified using the bi-parental populations were considered significantly associated.

Synteny Analysis: Regions of the soybean genome syntenic with the cowpea seed size QTLs reported here were searched for seed size QTLs. HarvEST:Cowpea (Wanamaker and Close, 2011) was used to identify synteny based on BLASTX scores ($<10^{-10}$) between cowpea unigenes containing mapped SNPs and translated gene models from soybean (Schmutz et al., 2010). Soybean genomic locations homeologous to cowpea seed size QTL were reconciled with an abundance of soybean seed size QTL inventoried and integrated with the physical genome by SoyBase (Grant et al., 2010). Only soybean QTL that were within or tightly linked to the syntenic region (< 3 million base pairs) were considered orthologous.

Results

Field and Greenhouse Trials: The two field trials using the CB27 x IT82E-18 population produced seed size data that were strongly correlated to each other (Pearson's r = 0.83) and similar to that of the greenhouse trial (r = 0.59 and 0.63 for 2010 and 2011 respectively). This was also the case for the multiple trials of the CB46 x IT93K-503 population where field trials were correlated to each other (r = 0.30) and more so to the greenhouse trial (r = 0.52 and 0.31 for ISRA and INERA trials respectively). Figure 5 provides the phenotypic distribution of seed size among all eight bi-parental populations. The smallest seeded line had a one-hundred seed weight of 3.26 grams and was produced by the parents Dan Ila and TVu-7778. The largest seeded line was produced by CB46 and IT93K-503 and had a one-hundred seed weight of 34.06 grams. The average seed of an individual from the eight bi-parental populations had a one-hundred seed weight of 15.50 grams. Phenotypic distributions of seed size for each trial are provided in Supplemental File 5. Seed sizes for the parents of the mapping populations are provided in Supplemental File 5 which ranged from 11.60 to 26.41 grams per one-hundred seed. Phenotypic and genotypic characteristics of the mapping populations are provided in Table 3.

Association Studies: Ten QTL for seed size, representing ~10% of the mapped cowpea genome were identified among the eight bi-parental populations (Table 4). Most had narrow spans (<5 cM), accounted for a substantial proportion of the phenotypic variance (average of 30%), and were associated with multiple SNP markers (average LOD > 8.5). LOD score traces for each QTL discovery are included in Supplemental File

5. Haplotypes associated with large and small seed were consistent among discovery populations when QTL were detected in multiple populations (Supplemental File 5). This is the situation for *Css – 1* where markers 1_0974 and 1_0078 were detected among different experiments. Allelic variation important for seed size can be found among all parents of the bi-parental populations except for Dan Ila and IT84S-2049. The additive allelic effect of *Css – 1* was similar (1.77 and 2.18 grams) between multiple trials of the CB27 x IT82E-18 population. This is also true for the multi-trial detection of *Css – 2* using the CB46 x IT93K-503 population (1.97 grams for both experiments).

Supplemental File 5 displays the potential for genetic gain by combining favorable alleles for multiple QTLs. QTL content has the most significant effect on seed size for the CB27 x IT82E-18 population $F(3, 149) = 28.51$, $p = 1.25E\text{-}14$, $\eta^2 = 0.36$). This is also true for all other populations except for the CB46 x IT93K-503 population where groups based on QTL content are mainly different due to chance $F(4, 86) = 1.29$, $p = 0.28$, $\eta^2 = 0.06$. See the Supplemental File 5 for test statistics for all populations. No significant QTL interactions were found among the discovery experiments.

Six-hundred and sixty-five SNPs which were polymorphic among 171 accessions of the USDA core were used to identify 27 duplicated accessions (Supplemental File 5). An additional ten accessions were excluded from further analysis due to a lack of geographic collection information. This filter yielded 134 accessions appropriate for population structure and association analyses. Geographic collection information and the Evanno method (Supplemental File 5) supported four subpopulations which accounted for a substantial proportion of population structure underlying the USDA core (Figure 6).

Genepool 1 was the most common and was comprised of a majority of the samples collected in Eastern and Southern Africa. Samples collected in Asia were categorized primarily in genepool 2, and West Africa and Turkey were identified as genepool 3 and genepool 4, respectively. The genomes of 46 samples were primarily derived from genepool 1 and up to 87 samples contained a substantial proportion originating from genepool 1 (Supplemental File 5). While only 8 samples could be attributed entirely to genepool 3, 43 samples were admixed with genepool 3. Samples collected in South America were almost always an admixture of genepools 1 and 3. Most of the migrants were collected in West Africa and Asia.

Thirty-six SNP loci used in the GWAS of the USDA core surpassed –Log (P-value) thresholds and confirmed six of the ten QTL proposed by the bi-parental populations (Figure 7). This information is incorporated into Table 4 and is more comprehensively provided in the Supplemental File 5.

Synteny: Based on the syntenic relationships described by Lucas et al. (2011), seven out of the ten QTL identified in the bi-parental populations were supported by knowledge of seed size in soybean (Table 4 and Table 5). A total of 19 associations between markers and seed size developed in soybean (Chen et al., 2007; Csanadi et al., 2001; Gai et al., 2007; Hoeck et al., 2003; Hyten et al., 2004; Orf et al., 1999; Panthee et al., 2005; Reinprecht et al., 2006; Specht et al., 2001; Zhang et al., 2004) were in regions homoeologous to the cowpea seed size QTL reported here. More details of the synteny analysis are presented in Supplemental File 5.

Discussion

Cowpea with specific seed size can be predicted on the basis of marker-trait associations. These associations provide a foundation for marker-assisted breeding and can be developed through QTL analysis and association mapping which couple phenotypes, genotypes, and a genetic map (Figure 8). DNA markers tagging allelic variation underlying seed size QTL can be used to track trait determinants among breeding cycles. This approach facilitates the simultaneous improvement of a variety for different traits of interest.

From the standpoint of breeding, the most applicable association studies assess broad pedigrees and tag associated genomic regions with dense markers. Marker-trait associations identified in one population may not segregate or contribute to the inheritance of the trait in a different population. To support new marker-trait associations we used multiple populations, two popular methodologies (QTL and GWAS), and knowledge of seed size in soybean. The intent of this study was to assess allelic variation important for the inheritance of seed size in cowpea primarily in the context of marker-assisted breeding and comparative genomics. The associations developed in this study would be best validated after years of using them in breeding; however, we feel this work provides an important framework for future breeding initiatives and explores the potential of genomics to help deliver new varieties of cowpea. The accuracy of these marker-trait associations could be assessed by comparing the estimated additive allelic effects reported here with realized gains after using these markers for selection.

Based on our analyses there is a large potential to produce larger-seeded lines by combining favorable alleles for multiple QTLs. However, our analysis of this potential is limited because our study lacked recombinants for all possible QTL combinations. A more complete view could be provided by studying the behavior of QTLs outside of their discovery pedigree. This could be accomplished by pursuing a mating scheme which used lines from different pedigrees and with different QTL content.

Breeders interested in using marker-trait associations would benefit from knowledge of linkage between trait determinants. The locations of the QTLs reported here are mainly unlinked or distantly linked to other traits characterized using the consensus genetic map of cowpea, including heat tolerance during reproductive development (Lucas et al., 2012a), leaf morphology (Pottorff et al., 2012b), and resistance to *Fusarium oxysporum* f.sp. *tracheiphilum* race 3 (Pottorff et al., 2012a). One (*Thr – 1*) of the three QTL known to impact resistance to feeding damage caused by foliar thrips (*Thrips tabaci* and *Frankliniella schultzei)* (Lucas et al., 2012b) overlaps with a seed size QTL (*Css – 3*) reported in the current work. The markers within this overlapping region include 1_0164 and 1_0589 where genotypes homozygous for AA at these were associated with large seed and thrips resistance. This means it would require a rare recombination to break the linkage between resistance to foliar thrips conferred by *Thr – 1* and a small seed conferred by *Css – 3*.Other overlaps can be found between seed size QTL and regions associated with *Macrophomina phaseolina* resistance (Muchero et al., 2011) (*Css – 9* with *Mac – 6*, and *Css – 4* with *Mac – 8*). Therefore using markers linked to these overlapping regions may simultaneously affect seed size and resistance to

Macrophomina. In such regions a higher density of markers would be useful for marker-assisted breeding.

SoyBase (Grant et al., 2010) is an excellent resource for legume researchers. The integration of QTL studies with the physical map made it possible for us to survey commonalities among association studies performed in plants of different genera. Such knowledge may provide paths for mechanistic studies aiming to pinpoint trait determinants. From the standpoint of this study, the co-localization of seed size QTL in soybean and cowpea provides a level of validation for new marker-trait associations. Knowledge of legume synteny and trait-determinants would be enhanced by developing resources similar in density to the soybean community for other legumes (i.e. common bean, cowpea, mung bean, peanut, chickpea, etc.). An agricultural project that would be coordinated among groups with expertise in different legumes could greatly enhance comparative resources and the efficiency of new initiatives.

The fact that our study uses approaches capable of clarifying the domestication history and dispersal of modern cowpea does not escape our attention; however, due to sample size we advocate a conservative interpretation of the diversity analysis using the USDA core as presented here. Rather than focusing on potential insight concerning cowpea domestication or proposing new marker-trait associations, we present the results of the genome-wide association study only to provide a modest assessment of collection diversity and to help support QTL identified among the bi-parental populations. The International Institute of Tropical Agriculture maintains a diverse collection which has been previously characterized on the basis of geographic, agronomic, and botanical

descriptors (Mahalakshmi et al., 2007), but no collection of cowpea has been viewed in light of dense genotype data. The financial costs required for genotyping the rest of the USDA core is inexpensive relative to the value of the insight that can be gained. From our analysis of a small subset (171 samples) of the entire USDA core (720 samples) we were able to identify many duplicated accessions (~17%), overrepresentation of the South/East African genepool, and we were able to perform an association study which mainly agreed with the QTL studies stemming from the bi-parental recombinant-inbred populations. SNP data from the entire core collection could be used to improve the diversity collection and its impact on the cowpea community. Phenotypic data for a number of traits are available on GRIN and could be combined with genotype data, similar to this work, to facilitate the discovery of numerous marker-trait associations. The use of historical data would be a cost effective approach to improve knowledge of cowpea genetic diversity and allelic variation contributing to the inheritance of agronomically important traits. The feasibility of this approach was recently supported within the barley community (Wang et al., 2011). That work helped demonstrate the utility of historical data after careful consideration of population size and experimental design. Furthermore, a comparative analysis of the diversity among core collections (i.e. IITA, USDA, and UCR) would be valuable for identifying instances of ascertainment bias and duplicated accessions possibly known by different names. This is a documented issue for U.S. collections (Vigna Crop Germplasm Committee, 1996), and continued application of genotype data to identify duplicates would be particularly helpful in cutting costs associated with the maintenance of collections and for designing new experiments.

References

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, and Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633 - 2635.

Chen Q, Zhang Z, Liu C, Xin D, Qiu H, Shan D, Shan C, and Hu G (2007) QTL analysis of major agronomic traits in soybean. Agricultural Science in China 6: 399 - 405.

Cober ER, Voldeng HD, and Fregeau-Reid JA (1997) Heritability of seed shape and seed size in soybean. Crop Science 37: 1767 - 1769.

Csanadi G, Vollmann J, Stift G, and Lelley T (2001) Seed quality QTL identified in a molecular map of early maturing soybean. Theoretical and Applied Genetics 103: 912 - 919.

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

Earl DA, and von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359 - 361.

Evanno G, Regnaut S, and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611 - 2620.

FAOSTAT F (2013) Agriculture Organization of the United Nations. Statistical Database.

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

Fery RL (1980) Genetics of Vigna, In Horticultural Reviews, ed. J. Janick (Westport, Conn. AVI publishing): 311 - 394.

Gai J, Wang Y, Wu X, and Chen S (2007) A comparative study on segregation analysis and QTL mapping of quantitative traits in plants-with a case in soybean. Frontiers of Agricultural in China 1: 1 - 7.

Gillaspie AG, Chambliss OL, Fery RL, Hall AE, Miller Jr, JC, and Morelock TE (1996) A core subset established for the USDA cowpea [Vigna unguiculata (L.) Walp.] germplasm collection. Horticultural Science 31: 762.

Giura A, and Saulescu NN (1996) Chromosomal location of genes controlling grain size in a large-grained selection of wheat (*Triticum aestivum* L.). Euphytica 89: 77 - 80.

Grant D, Nelson RT, Cannon SB, and Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research 38: 843 - 846.

Hendrix SD, Nielsen E, Nielsen T, and Schutt M (1992) Are seedlings from small seeds always inferior to seedlings from large seeds? Effects of seed biomass on seedling growth in Pastinaca sativa L. New Phytologist 119: 299 - 305.

Hoeck JA, Fehr WR, Shoemaker RC, Welke GA, Johnson SL, and Cianzio SR (2003) Molecular marker analysis of seed size in soybean. Crop Science 43: 68 - 74.

Hu Z, Ehlers JD, Roberts PA, Close TJ, Lucas MR, Wanamaker S, and Xu S (2012) ParentChecker: a computer program for automated inference of missing parental genotype calls and linkage phase correction. BMC Genetics 13.

Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, and Schmidt ME (2004) Seed quality QTL in a prominent soybean population. Theoretical and Applied Genetics 109: 552 - 561.

Jakobsson M, and Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801 - 1806.

Illumina (2010) GenomeStudio genotyping module v.2010.3. Illumina, Inc., San Diego, CA. http://www.illumina.com/software/genomestudio_software/ilmn.

Li H, Ribaut JM, Li Z, and Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theoretical and Applied Genetics 116: 243 - 260.

Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, and Close TJ (2011) Cowpea-soybean synteny clarified through an improved genetic map. The Plant Genome 4: 218 - 225.

Lucas MR, Ehlers JD, Huynh BL, Diop NN, Roberts PA, and Close TJ (2013) Markers for breeding heat tolerant cowpea. Molecular Breeding 31: 529 - 536.

Lucas MR, Ehlers JD, Roberts PA, and Close TJ (2012b) Markers for quantitative inheritance of resistance to foliar thrips in cowpea. Crop Science 52: 2075 - 2081.

Lucas MR, Hunyh BL, Ehlers JD, Roberts PA, and Close TJ (2013) High-resolution single nucleotide polymorphism genotyping reveals a significant problem among breeder resources. The Plant Genome 6: doi:10.3835/plantgenome2012.08.0020.

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

Mahalakshmi V, Ng Q, Lawson M, and Ortiz R (2007) Cowpea [Vigna unguiculata (L.)Walp.] core collection defined by geographical, agronomical, and botanical descriptors. Plant Genetic Resources: Characterization and Utilization 5: 113 - 119.

Maughan PJ, Saghai-Maroof MA, and Buss GR (1996) Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for ortholgous evolution among three legume species. Theoretical and Applied Genetics 93: 574 - 579.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, and 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.

Muchero W, Ehlers JD, Close TJ, and Roberts PA (2011) Genic SNP markers and legume synteny reveal candidate genes underlying QTL for *Macrophoina phaseolina* resistance and maturity in cowpea [*Vigna unguiculata* (L) Walp.]. BMC Genomics 12: doi:10.1186/1471-2164-12-8.

Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, and Lark KG (1999) Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Science 39: 1642 - 1651.

Panthee DR, Pantalone VR, West DR, Saxton AM, and Sams CE (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Science 45: 2015 - 2022.

Pottorff M, Wanamaker S, Ma YQ, Ehlers JD, Roberts PA, and Close TJ (2012a) Genetic and physical mapping of candidate genes for resistance to *Fusarium oxysporum* f.sp. *tracheiphilum* race 3 in cowpea [*Vigna unguiculata* (L.) Walp]. PLoS One 7.

Pottorff M, Ehlers JD, Fatokun C, Roberts PA, and Close TJ (2012b) Leaf morphology in cowpea [Vigna unguiculata (L.)Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. BMC Genomics 13: doi:10.1186/1471-2164-13-234.

Pritchard JK, Stephens M, and Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945 - 959.

Reinprecht Y, Poysa V, Yu K, Rajcan I, Ablett G, and Pauls K (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (Glycine max (L.)Merrill) germplasm. Genome 4: 1510 - 1527.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137 - 138.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T,

Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht, J, Grimwood, J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178 - 183.

Specht, JE, Chase K, Macrander M, Graef GL, Chung J, Markwell JP, Germann M., Orf JH, and Lark KG (2001) Soybean response to water: a QTL analysis of drought tolerance. Crop Science 41: 493 - 509.

USDA-ARS, National Genetic Resources Program (2013) Germplasm Resources Information Network – (GRIN). [online database] National Germplasm Resources Laboratory, Beltsville, Maryland. Available: http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?150001 (05 February 2013).

Vigna Crop Germplasm Committee (1996) Vigna germplasm current status and future needs. National Germplasm Resources Laboratory, Beltsville, Maryland.Available at http://www.ars-grin.gov/npgs/cgc_reports/vigna96.html.

Wanamaker S, and Close TJ (2011) HarvEST. HarvEST:Cowpea version 1.27. Available at http://harvest.ucr.edu (verified date). University of California, Riverside, CA.

Wang H, Smith KP, Combs E, Blake T, Horsley RD, and Muehlbauer GJ (2011) Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. Theoretical and Applied Genetics 124: 111 – 124.

Wulff RD (1986) Seed size variation in *Desmodium paniculatum* II. Effects on seedling growth and physiological performance. Journal of Ecology 74: 99 - 114.

Zhang W, Wang Y, Luo G, Zhang J, He C, Wu X, Gai J, and Chen S (2004) QTL mapping of ten agronomic traits on the soybean (Glycine max L. Merr) genetic map and their association with EST markers. Theoretical and Applied Genetics 108: 1131 - 1139.

Table 3: Characteristics of eight bi-parental populations of cowpea used to associate loci with seed size.

* Number of samples used for QTL analysis after eliminating rogues.

† Number of polymorphic SNPs out of 1,536 genotyped SNPs which could be mapped.

Parents		Population Size*	Number of Polymorphic SNPs†	Population 100 seed weight (grams)	
				Range	Average
CB27	IT82E-18	160	430	8.53 – 30.96	17.58
CB27	UCR 779	56	560	9.57 – 29.81	18.28
CB27	24-125B-1	87	329	9.98 – 28.50	17.54
CB46	IT93K-503	114	374	7.99 – 34.95	17.00
Dan Ila	TVu-7778	79	288	3.26 – 19.50	12.77
524B	IT84S-2049	85	438	12.55 – 24.15	17.29
TVu-14676	IT84S-2246	136	345	5.61 – 30.51	15.59
IT84S-2246	Mouride	87	347	12.55 – 22.35	16.70

Table 4: Ten seed-size QTL identified among eight bi-parental populations of cowpea. Statistical tests used to identify QTL from the discovery experiments are reported for the Log of Odds score (LOD), percent of phenotypic variation explained by the QTL (%Phe), and the absolute value of the additive effects (Additive). Linkage group and centi-Morgan (cM) positions were identified from the consensus genetic map of cowpea. The number of markers used to tag a QTL among the discovery experiments is also included. Underlined QTL were also identified in the genome-wide association study.

* Indicates the QTL was discovered by analyzing data from both field and greenhouse experiments.

† Indicates the QTL was discovered by analyzing data from greenhouse experiments.

‡ Indicates the QTL was discovered by analyzing data from field experiments.

§ Indicates the QTL is supported based on synteny to soybean seed size associated loci.

QTL Name	Discovery Population(s)	Large Seed Allele Donor(s)	LOD	%Phe	Additive (grams)	Linkage Group	QTL Location (cM)	Number of Markers
Css – 1 *	CB27 x IT82E-18 CB27 x UCR 779 TVu-14676 x IT84S-2246	IT82E-18 UCR 779 TVu-14676	3.82 – 34.43	24.6 – 45.0	1.77 – 3.34	5	53.23 – 57.30	5
Css – 2 †§	CB27 x IT82E-18 CB46 x IT93K-503 TVu-14676 x IT84S-2246	CB27 CB46 TVu-14676	4.64 – 7.94	4.2 – 17.2	1.03 – 1.97	7	18.70 – 22.68	6
Css – 3 †	524B x IT84S-2049 IT84S-2246 x Mouride TVu-14676 x IT84S-2246	524B IT84S-2246 TVu-14676	2.94 – 6.63	5.2 – 25.2	0.69 – 1.30	2	18.92 – 32.75	6
Css – 4 †§	CB27 x IT82E-18 IT84S-2246 x Mouride	CB27 Mouride	5.37 – 5.42	9.9 – 26.8	1.12 – 1.16	6	31.28 – 57.41	4

Css – 5 ‡§	CB46 x IT93K-503	IT93K-503	5.30	41.1	1.70	8	55.20 – 61.60	2
Css – 6 †§	Dan Ila x TVu-7778 IT84S-2246 x Mouride	TVu-7778 Mouride	3.38 – 3.69	20.3 – 40.8	1.24 – 1.38	10	0.00 – 4.90	3
Css – 7 †§	CB46 x IT93K-503	CB46	4.07	15.0	1.95	2	56.95 – 62.06	2
Css – 8 †§	CB27 x 24-125B-1	24-125B-1	5.18	26.7	1.80	6	3.82 – 4.55	2
Css – 9 †	CB27 x 24-125B-1	24-125B-1	3.24	16.6	1.42	5	18.50 – 21.57	2
Css – 10 ‡§	CB46 x IT93K-503	CB46	10.4	46.7	1.66	7	31.40 – 32.20	2

Table 5: Seven QTL controlling the inheritance of seed size in cowpea are syntenic to regions with known association to seed size in soybean. Chromosome (Chr) and base pair starting (Start) and ending (End) positions in the soybean genome are indicated for each syntenic relationship. Up to three soybean QTL can be found within or tightly linked (< 3 Mbp) to the syntenic span; except Sd wt 18-1.1, 21-1, 22-2 and 25-2 which are < 5Mbp.

Cowpea	Soybean			Soybean Seed Size Associations		
QTL	Chr	Start	End	QTL 1	QTL 2	QTL 3
Css - 2	1	1484033	3411009	Sd wt 18-1.2	-	-
	9	39678719	40435370	Sd wt 10-10	Sd wt 15-6	-
Css - 4	19	48637729	50058893	Sd wt 12-3	Sd wt 13-9	-
	9	39678719	40435370	Sd wt 10-10	Sd wt 15-6	-
Css - 5	4	1776391	1787115	Sd wt 13-4	-	-
Css - 6	7	185839	2305626	Sd wt 10-11	Sd wt 7-6	-
	8	15355701	17919999	Sd wt 22-1	-	-
Css - 7	10	45955504	47689454	Sd wt 25-4	-	-
	20	35530502	38505007	Sd wt 15-5	Sd wt 24-3	-
Css - 8	8	14338579	15148040	Sd wt 22-1	-	-
Css -10	11	5421236	5680908	Sd wt 21-1	Sd wt 22-2	Sd wt 25-2
	1	49378273	49834904	Sd wt 15-2	Sd wt 18-1.1	Sd wt 7-4

Figure 4: Popular cowpea seed types include ‘blackeyed’ and ‘buff’ represented by (A) California Blackeye 27 and (B) IT82E-18. However, a diversity of cowpea seed types exist (C).

A **B** **C**

Figure 5: Phenotypic distribution of seed size among eight bi-parental populations of cowpea.

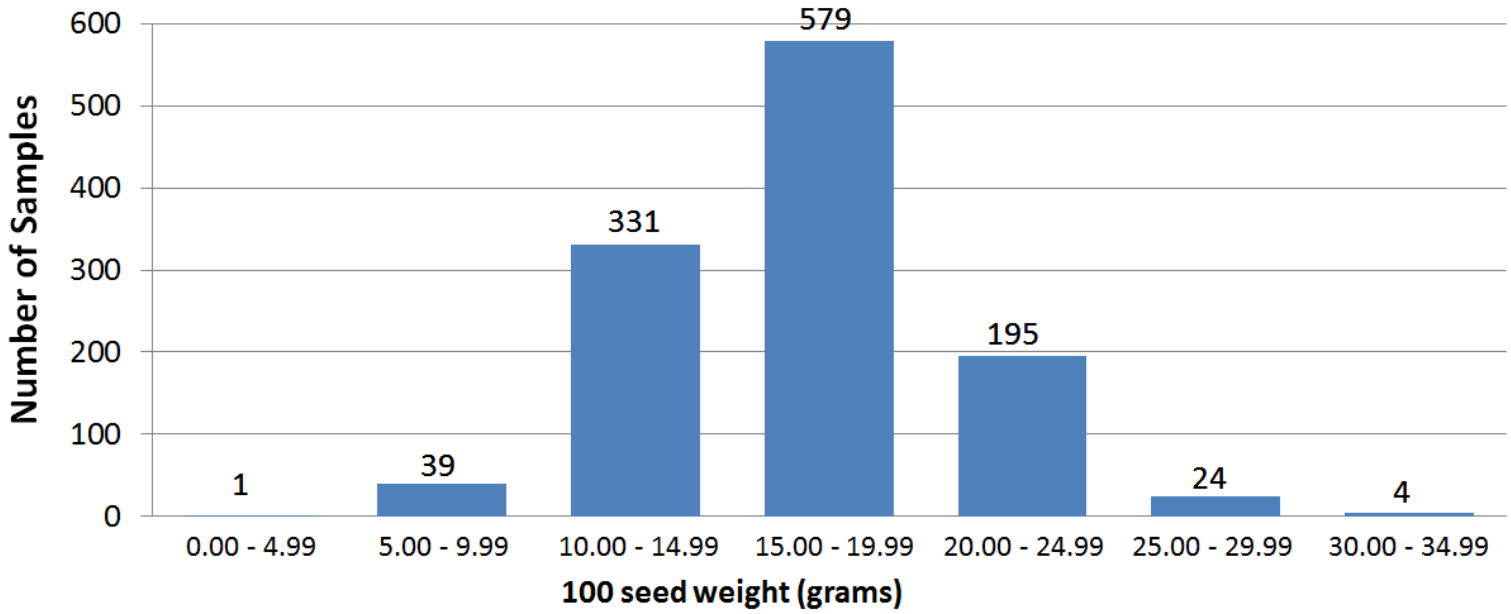


Figure 6: Population structure underlying a subset of the USDA core collection of cowpea. Samples are first sorted based on their geographic location of collection and then sorted based on a coancestry matrix with K = 4.

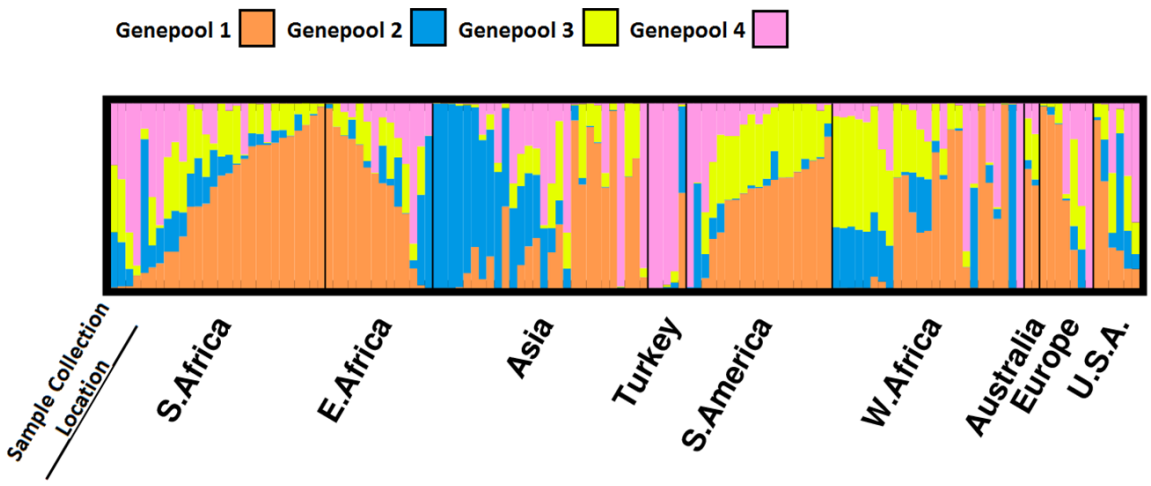


Figure 7: Genome wide association analysis of seed size using the USDA core collection of cowpea. Loci surpassing significance thresholds that were also associated with seed size among the bi-parental populations are boxed.

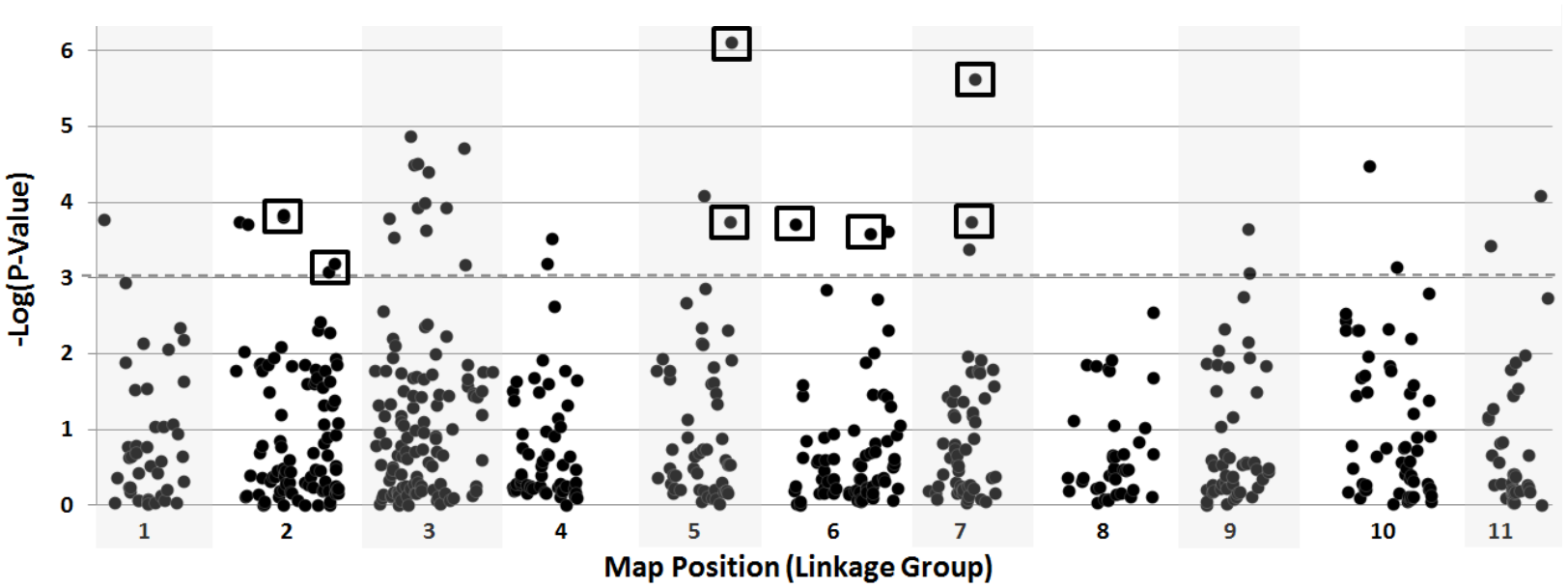
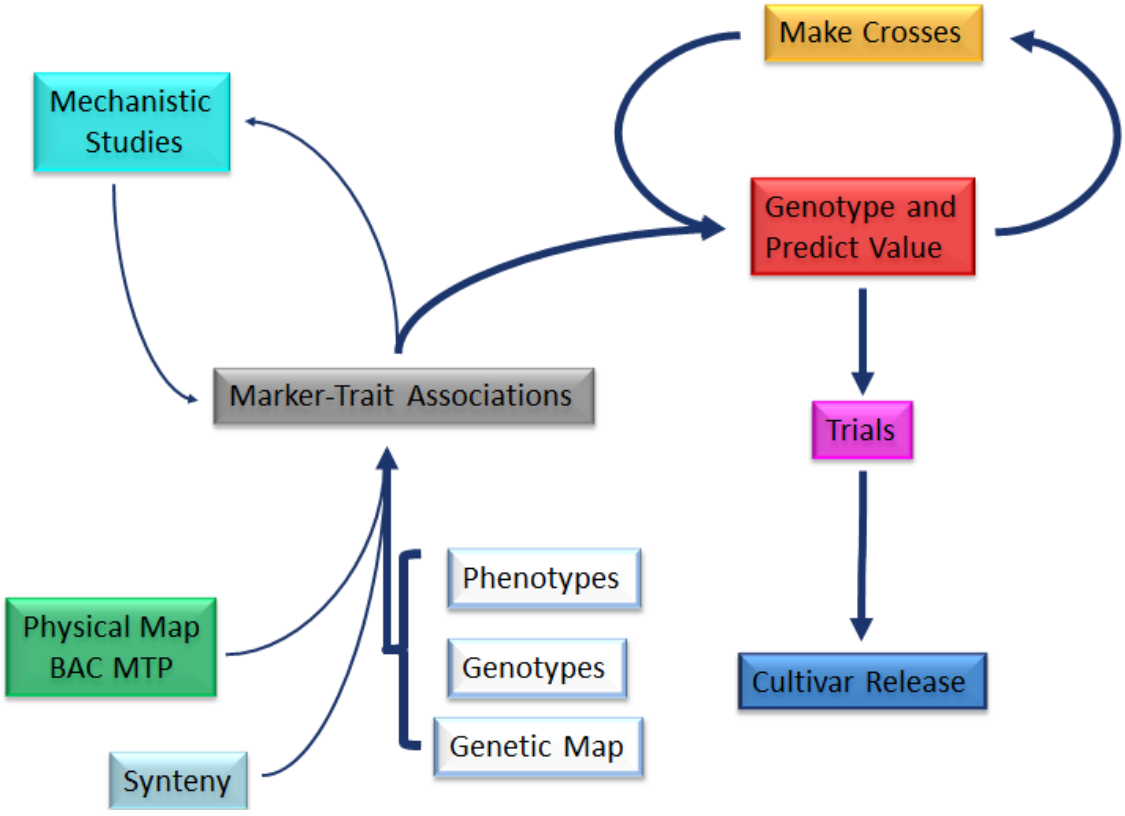


Figure 8: General pathway of marker-assisted breeding strategies which rely heavily on the development of marker-trait associations.



Chapter 3

Introgression of Rare Haplotype from Southeastern Africa to Breed Cowpeas with Larger Seeds

Abstract

Seed size distinguishes most crops from their wild relatives and is an important quality trait for the grain legume cowpea. In order to breed cowpea varieties with larger seeds we introgressed a rare haplotype associated with large seeds at the *Css-1* locus from an African buff seed type cultivar, IT82E-18 (18.5g/100 seeds), into a blackeye seed type cultivar, CB27 (22g/100 seed). Four RILs derived from these two parents were chosen for marker-assisted backcrossing based on SNP genotyping with a goal of stacking large seed haplotypes into a CB27 background. Foreground and background selection were performed during two cycles of backcrossing based on genome-wide SNP markers. The average seed size of introgression lines homozygous for haplotypes associated with large seeds was 28.7g/100 seed and 24.8g/100 seed for cycles 1 and 2, respectively. One cycle 1 introgression line with desirable seed quality was selfed for two generations to make families with very large seeds (28-35g/100 seeds). Field-based performance trials helped identify breeding lines that not only have large seeds but are also desirable in terms of yield, maturity, and plant architecture when compared to industry standards. A principal component analysis was used to explore the relationships between the parents relative to a core set of landraces and improved varieties based on high-density SNP data. The geographic distribution of haplotypes at the *Css-1* locus suggests the haplotype associated with large seeds is unique to accessions collected from Southeastern Africa. Therefore this QTL has a strong potential to develop larger seeded varieties for other growing regions which is demonstrated in this work using a California pedigree.

Introduction

Seed size is one of the most universal features that distinguishes domesticated plants from their wild relatives. Larger seeds produce more competitive seedlings under cultivated conditions (Purugganan et al., 2009) and are preferred for most culinary preparations of naked grain. This is true for cowpea (*Vigna unguiculata*) where the demand for large seeds continues for most market classes, especially blackeyes and rough seed types grown for flour production. Traders, farmers, food vendors, and consumers in West Africa prefer and are willing to pay a price premium for larger cowpea grains (Langyintuo et al., 2004; Mishili et al., 2009; Egbadzor et al., 2013a) so this trait has the potential to improve the income of cowpea growers in regions where ~125 million people live in poverty (IFAD, 2001). Cowpea breeders can help meet this demand by developing varieties with larger seeds.

Seed size in cowpea is highly heritable and quantitative, and small seeds are partially dominant to large seeds (Drabo et al., 1984; Egbadzor et al., 2013b; Egbadzor et al., 2013c). Genetic mapping using experimental populations has tagged a few seed size associated QTLs with markers that could be useful in breeding (Fatokun et al., 1992; Egbadzor et al., 2013b; Lucas et al., 2013a). Interestingly, two of these publications report on the orthology of seed size based on comparative mapping to known seed size associated loci in the genomes of cowpea relatives mung bean (*Vigna radiata*) and soy bean (*Glycine max*). Knowledge of marker-trait associations from these studies is an essential component of marker-assisted breeding strategies to help develop varieties with larger seeds.

California Blackeye 27 (CB27) and IT82E-18 are two cultivars that were used to develop many of the genomic resources in cowpea. A recombinant inbred population derived from the cross of these two individuals was used to help construct a consensus genetic map of 1,107 EST-derived SNP loci (Lucas et al., 2011). This population was also used to characterize the inheritance of heat tolerance during reproductive development (Lucas et al., 2013b), resistance to feeding damage caused by foliar thrips (Lucas et al., 2012), and seed size (Lucas et al., 2013a). The *Css-1* QTL described by Lucas et al. (2013a) provides an attractive breeding opportunity because it is known to be a major determinant of seed size and the haplotype associated with large seeds is absent from the California Blackeye pedigree. In this work we targeted the introgression of a 4.1 cM *Css-1* haplotype from an African buff seed type variety, IT82E-18, into a California Blackeye variety, CB27 using marker-assisted backcrossing.

Materials and Methods

Principal Component Analysis: The relatedness of CB27 and IT82E-18 was assessed through comparison with a core set of 212 individuals. The majority of the accessions in the core set are landraces which represent the West and South-East African genepools described by Huynh et al. (2013). To understand how CB27 compares, the core set also included other improved varieties from California and landraces representative of other geographic regions including Asia, Europe, the Middle-East, and North Africa. Genotype data for these samples were obtained from Lucas et al. (2011) and Huynh et al. (2013) which utilized the 1,536-plex EST-derived SNP genotyping platform of cowpea (Muchero et al., 2009). A principal component analysis was performed after filtering SNPs for MAF >0.01 and imputing missing genotype data using the software TASSEL (Bradbury et al., 2007). The first two principal components were plotted on a scatter plot and samples were colorized based on their geographic origin.

Distribution of *Css-1* Haplotypes: The *Css-1* locus comprises a 4.1 cM region on linkage group 5 of the cowpea consensus genetic map defined by SNP markers 1_0099, 1_0935, 1_0974, and 1_0078 (Lucas et al., 2013a). The haplotype of cowpea variety IT82E-18 characterized by the aforementioned SNPs is associated with the inheritance of large seed and the genotype calls are GG, GG, GG, and AA, respectively, in Illumina Top Strand format. This unique combination of genotype calls was queried against the diversity panel used in the principal component analysis and the results were tallied by country and by geographic region which helped understand the origin and distribution of the QTL in domesticated cowpea germplasm.

Introgression of the IT82E-18 *Css-1* Haplotype: Donor parents for both cycles of the backcrossing project were chosen based on four criteria: 1) Foreground selection for the haplotype associated with large seeds at *Css-1* (Lucas et al., 2013a); 2) Background selection for similarity to CB27 based on 1,536 SNP genotype data; 3) 100-seed weight (Lucas et al., 2013a); and 4) Seed coat type. Using these criteria RILs -62, -74, -90, and -113 were chosen as donor parents for cycle 1. These lines were crossed to one female CB27 plant to produce 4 types of F1s. One F1 plant of each type was selfed to produce four F2 families. After two weeks of growth, tissue was taken from 45 F2s using the LGC genomics tissue collection method. DNA extraction and genotyping for 49 SNPs was performed using the KASP technology of LGC Genomics. Four of these SNPs distinguish the *Css-1* haplotypes while the others were chosen based on their distribution in the consensus genetic map (Lucas et al., 2011), their linkage to two other QTL affecting seed size that are segregating in this pedigree (Lucas et al., 2013a), and polymorphism between the parents. The genotype calls were compared to *Css-1* haplotypes and to CB27 to generate a ranking of F2s. This analysis was completed before the F2 families finished flowering so another cycle of backcrossing could be immediately pursued. One cycle 1 F2 plant was chosen for additional trials and seed increase because it was homozygous for large seed alleles at *Css-1*, had very large seeds, and a blackeye seed coat. Cycle 2 began by crossing one F2 plant from cycle 1, -113 family, with a female CB27 while another F2 plant from cycle 1, -74 family, was crossed with CB27 as both a male and a female. The cycle 2 F1s that were produced from these crosses were selfed and 93 F2s comprised of three families (CB27 x Cycle 1 -113 F2, CB27 x Cycle 1

-74 F2, and Cycle 1 -74 F2 x CB27) were grown. Tissue from the cycle 2 F2s was collected using the LGC genomics collection method and the DNA was genotyped for 108 SNPs, 4 of which distinguish the *Css-1* haplotypes. A total of 19 of these SNPs were chosen to assess the content of heat tolerance associated QTLs (*Cht-1*, *Cht-4*, and *Cht-5*) that were characterized by Lucas et al. (2013b). Favorable alleles for heat tolerance QTL *Cht-2* and *Cht-3* are fixed in the cycle 1 parents so markers tagging these loci were not included for additional genotyping. All plants in this work were grown in temperature, irrigation, and pesticide controlled greenhouses. CB27 and IT82E-18 plants were included during each generation as a reference to a known seed size. Seeds were harvested from dried plants, counted, and weighed for each generation to determine the mass of 100 seeds (seed size). The seed type of each plant was observed and categorized as blackeye, browneye, or other.

Performance Testing: To assess the impact of *Css-1* introgression on other agronomic factors parents and breeding lines were planted on May 29th, 2014 at the University of California, Division of Agriculture and Natural Resources, Kearney Agricultural Research and Extension Center in Parlier, CA. Seventy-five seeds were sown in each experimental plot which was 6.70 meter long and separated from other plots by a 0.91 meter alley and 0.76 meter centered beds. All plots were treated with Temik as a precaution against insects and were well watered every ten days using furrow irrigation. The experiment was surrounded by buffer plots of the industry standard, CB46. Other California Blackeye cultivars including CB46, CB27, and CB50 in addition to the donor parent, IT82E-18, were grown in blocks of four adjacent plots to compare yield and

maturity. 26 cycle 1 F4 families derived from RIL-113 that are homozygous for favorable alleles at all three QTL reported to be segregating in this pedigree by Lucas et al. (2013a) were also tested for maturity, plant height, row closure, visual performance, yield, and 100 seed weight. Maturity was assigned based on six maturity groups including Early, Medium Early, Medium, Medium Late, Late, and Photoperiod Sensitive which was determined by inspecting plots during pod-filling (August 25th, 2014). Plant height, row closure, and visual performance were estimated using a number scale of 1-10 by inspecting plots following pod-set (August 7th, 2014). Yield was assessed by harvesting all pods in a 0.91 meter section in the middle of plots (September 25th, 2014). Photoperiod sensitive varieties which grew vegetative and covered neighboring plots prevented yield data collection for eight of the cycle 1 F4 families. These were omitted because they matured and were later over-grown by neighboring plots which prevented sufficient yield sampling (some pods may have been missed due to over-growth). One hundred representative seeds were weighed to determine 100-seed weight.

Results

Principal Component Analysis: A total of 1,073 out of 1,536 SNP markers have MAF > 0.01 for the core set of 214 accessions and were used in the principal component analysis (Supplemental File 6). The first principal component described 24% of the variation and distinguished West African landraces from South-East African landraces (Figure 9, Supplemental File 6). The second principal component explained 8% of the variation and distinguished these two centers of diversity from all other landraces and improved California varieties. In the principal component analysis IT82E-18 clustered with landraces from South-East Africa while CB27 formed a cluster with other California varieties, landraces from the Middle East, and North Africa that were only separated from the West-African landraces by principal component 2 (Supplemental File 6).

Distribution of *Css-1*: Out of all 214 accessions studied in the principal component analysis only 19 carried the SNP haplotype of IT82E-18 and all of these are from countries in South and East Africa (Supplemental File 6). These include accessions from Botswana, Lesotho, Malawi, Mauritania, Mozambique, South Africa, Uganda, Zambia, and Zimbabwe. Mauritania is a country located in West Africa. However, the one accession (TVu-467) collected from this region that also contains the large seed haplotype at the *Css-1* locus is clearly a migrant or error in records based on the diversity study of Huynh et al. (2014) which associates a probability of 100% that this accession belongs to the Southeastern genepool (Genepool 2). The haplotype was absent from all other accessions including those from West Africa, North Africa, Europe, Middle-East, Asia, and California.

Introgression of *Css-1*: A total of 435 out of 1,536 SNPs are polymorphic between CB27 and IT82E-18 (Supplemental File 6). A genome-wide subset of 50 polymorphic SNPs were genotyped for the cycle 1 plants which helped categorize the breeding lines based on foreground and background selection (Supplemental File 6). The seed size for cycle 1 plants are plotted in Figure 10 where homozygotes carrying favorable alleles at *Css-1* had an average seed size of 28.70g/100 seed while heterozygotes averaged 26.33g/100 seed and homozygous unfavorable alleles had an average seed size of 22.14g/100 seed. Only ~16% of cycle 1 plants produced seeds with the targeted blackeye seed type. One cycle 1 homozygote, -113-2-6, was selfed for two generations and consistently made large seeds which averaged ~ 31.68g/100 seed and is photographed next to the parents in Figure 11 (Supplemental File 6). A total of 109 genome wide SNPs were genotyped on the cycle 2 plants (Supplemental File 6). Favorable alleles for heat tolerance QTLs *Cht-2*, *Cht-3*, *Cht-4*, and *Cht-5* are primarily fixed among the cycle 2 plants while *Cht-1* was still segregating (Supplemental File 6). The seed size of cycle 2 plants is also plotted in Figure 10 where homozygotes carrying favorable alleles at *Css-1* had an average seed size of 24.75g/100 seed, while heterozygotes averaged 24.70g/100 seed, and homozygous unfavorable alleles had an average seed size of 21.96g/100 seed. About 85% of cycle 2 plants produced the targeted blackeye seed type.

Performance Testing: In terms of maturity, no introgression lines were as early as CB27 or as late as CB46 because all were categorized in either Medium Early, Medium, or Medium Late maturity groups (Supplemental File 6). The donor parent

IT82E-18 was categorized as Medium Early while the other California Blackeye, CB50, matured at the same time as other Medium Late maturity group members.

The introgression lines varied for plant height, row closure, and visual estimate of performance (Supplemental File 6). For these traits CB46 received relatively high scores meaning it was the tallest, had the best row closure, and was visually estimated to perform well. Most introgression lines behaved similarly to CB27 and IT82E-18 which are earlier and shorter in plant height although there were a few introgression lines resembling the height and visual performance of CB46 and CB50.

The introgression lines that were tested in the field had much larger seeds than their parents and other industry standards (Supplemental File 6). Based on the trial of 26 cycle 1 F4 introgression lines, these lines had an average 100-seed weight of 32.83 grams with the largest family producing a 100-seed weight of 35.06 grams. CB50 is the largest California Blackeye and produced seeds with a 100-seed weight of 28.01 grams while CB46 is a smaller sized California Blackeye which made seeds with a 100-seed weight of 23.15 grams. The two parents of the introgression effort had 100-seed weights of 24.87 grams and 19.75 grams for CB27 and IT82E-18, respectively.

In terms of yield the introgression lines were variable (Supplemental File 6). Out of all the breeding lines, parents, and industry standards the two highest yielding lines were introgression lines which yielded ~600 grams of naked grain per 0.91 meter section. A few introgression lines yielded less than half this much. CB50 yielded the most grain (589.80 grams) of the registered California Blackeye varieties that were tested which

outcompeted CB27 (482.7 grams), CB46 (488.1 grams), and the African donor of *Css-1*, IT82E-18 (505.58 grams).

Discussion

Origin and Features of the Parents: CB27 and IT82E-18 represent two very different pedigrees which are grown for different market classes. CB27 produces medium-large blackeye seeds (22g/100 seed) with a rough seed coat and was bred for production in the San Joaquin valley of central California, USA. This contrasts with IT82E-18 which is an improved variety released in Mozambique, among other countries, and developed by the IITA in Africa which produces medium (18g/100 seed) light tan seeds that have a smooth seed-coat texture. Figure 11 provides an image of eight seeds of each parent separated by eight seeds of a cycle 1 F4 introgression line. In addition to these morphological and geographic differences these varieties can also be distinguished based on genotype data.

Out of the 13 bi-parental populations genotyped on the GoldenGate platform the population derived from CB27 and IT82E-18 had the second most polymorphic markers (437/~1200). However, this relatively high rate of polymorphism could be an artifact of ascertainment bias because these two cultivars were used for SNP discovery. The principal component analysis in this work also indicated major differences between the parents. CB27 met our expectations by clustering closer to West African varieties than to IT82E-18 because it was developed by breeding California blackeyes with two Nigerian varieties (Ehlers et al., 2000). It was also no surprise that IT82E-18 localized near the South-East African landraces which are separated from CB27 by both principal components.

Introgression of IT82E-18 *Css-1* Haplotype: The dramatic differences in seed type and pedigree between the parents used in this study may keep a breeder from wanting to cross the two. Given our goal of increasing the seed-size of an already large blackeye, it seems even less intuitive to use a moderate seed size variety like IT82E-18 as a parent. However, our study was informed by an association study which described the potential to breed a larger CB27 by incorporating a 4.1 cM haplotype from IT82E-18 (Lucas et al., 2013a). Furthermore, selection based on background markers provided a means to assess and recover all other features of CB27. The outcomes of this marker-assisted breeding project included new breeding lines that have up to 52% larger seeds, the targeted blackeye seed type, and perform well for other traits like yield, maturity and plant architecture under preliminary, early-generation field screens.

This introgression work still requires much more attention if these breeding lines are to be developed into registered varieties. Several important traits need to be assessed including pest resistance and multi-location yield testing. Furthermore, the preliminary field screening of F4 lines should be repeated on inbred materials. Issues for the lines developed in this work also concern seed quality. The introgression lines have very large seeds relative to the diversity found within domesticated cowpea germplasm collections. Their striking size makes them standout visually and they look odd when placed next to other cowpea varieties. When cleaning seeds from harvested pods special attention must be given to spacing the thresher drums to prevent large grains from splitting. This issue and perhaps fragility of the seed led to a substantial amount of split seeds during processing. Other noticeable features are slight discoloration and easily removable seed

coats. Future work should revisit cycle 1 and cycle 2 materials to advance lines that have non-splitting seeds and seed coats that are not discolored and do not crack because these traits are fixed among the introgression lines that were tested in the field (-113 family). However, if large seeded varieties are needed for culinary preparations where seed coats need to be removed (i.e. akara and moin moin) then the introgression lines from the -113 family could be desirable. While these preparations are most popular in West Africa these large seeded lines that lose their seed coat easily could be useful as a value-added supplement in processed foods in which cowpea flour is incorporated with other flour mixtures to enhance the nutritional profile.

Performance Trials: Seed size data collected from the field supported observations in greenhouse experiments. This was not a surprise because seed size is known to be one of the most highly heritable traits which is particularly true for the current work in well-watered and pest controlled environments. The most important trait to consider when breeding grain legumes is yield. This study only reports a single season and single site of field-based testing on one of several introgression families. The environment behaved expectedly and the trial was a success, however, we designed the experiment as a preliminary trial that was incorporated with planting a larger field experiment. This field trial was primarily conducted to assess variation in seed size and other traits including maturity, plant architecture and yield.

SNP Genotyping in Cowpea: One immediate impact of SNP genotyping is the ability to validate crossing records. This has been a particular valuable tool for identifying rogue lines (Lucas et al., 2013c), which is applicable to this work. We noticed

from genotype data that one cycle 2 family, 11327, may have arisen from a selfing event because all of the alleles for each line were contributed by CB27. Without genotype knowledge this error may have gone undetected. Future crosses and seed stocks can be validated using SNP genotyping to eliminate lines deviating from a designed pedigree.

Future Efforts: The *Css-1* haplotype associated with large seeds is unique to cowpea varieties originating from South and East Africa which means it is absent from the West African genepool and also has not been incorporated into varieties for other cowpea growing regions including the Americas and Asia. The distribution of this haplotype in cowpea diversity and its dramatic effect on seed size may be interesting to continue to study because it could relate to domestication. This work builds upon knowledge of the effect of *Css-1* in one pedigree and shouldn't be considered a diagnostic marker that can predict seed size in a random population, which could be explored through future introgression efforts.

There are a few warm season legumes that have a substantial amount of genomic resources and knowledge concerning the inheritance of traits, like soybean and common bean. For less intensively studied crops like cowpea knowledge of synteny to well-studied relatives provides opportunities to reconcile knowledge across plants from different genera. Lucas et al. (2013a) found that regions of the genome important for the inheritance of seed size are largely conserved between cowpea and soybean. Since that publication the genome sequence of common bean has been released (Schmutz et al., 2014). Cowpea and common bean shared a last common ancestor ~8 million years ago and are more closely related to each other than to soybean (Lavin et al., 2005). This

provides an enhanced framework for understanding the genetic mechanism dictated by the *Css-1* locus and for identifying orthologous factors determining seed size. Initial attempts could reconcile the syntenic location of the common bean genes controlling nitrogen metabolism and cytokinin synthesis which are important seed-size factors related to common bean domestication. Unfortunately there seems to be no common bean equivalent of SoyBase for soybean that catalogues literature findings into a searchable database and genome network.

Determining nutritional profiles of introgression lines and parents are important future experiments for at least two reasons. Comparison of nutritional profiles would allow us to quantify compositional differences in breeding lines from this work that may or may not be desirable for health or cooking characteristics. This should strongly influence the decision to deploy *Css-1* and change how to breed for seed size. Nutritional profiling could also suggest biochemical pathways that could help reveal the genetic mechanism underlying the *Css-1* locus, perhaps through protein variants or gene regulatory elements.

Acknowledgements

Support provided by the United States Department of Agriculture, National Institute of Food and Agriculture award: 2013-67011-21132, and the University of California Agricultural Experiment Station project: CA-R-BPS-5306-H. Thank you to the Genetics, Genomics, and Bioinformatics program at UC Riverside, the Citrus Agricultural Experiment Station at UC Riverside, and the UC Kearney Agricultural Research and Extension Center.

References

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, and Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633 - 2635.

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

Egbadzor KF, Yeboah M, Offei SK, Ofori K, and Danquah EY (2013a) Farmers key production constraints and traits desired in cowpea in Ghana. Journal of Agricultural Extension and Rural Development 5: 14 - 20.

Egbadzor KF, Yeboah M, Danquah EY, Ofori K, and Offei SK (2013b) Identification of SNP markers associated with seed size in cowpea [Vigna unguiculata (L.) Walp]. International Journal of Plant Breeding and Genetics 7 (2): 115 - 123.

Egbadzor KF, Dadoza M, Danquah EY, Yeboah M, Offei SK, and Ofori K (2013c) Genetic control of seed size in cowpea (Vigna unguiculata (L.) Walp). International Journal of Agriculture Sciences 5 (2): 367 - 371.

Fatokun CA, Menancio-Hautea DI, Danesh D, and Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132 (3): 841 - 846.

Huynh BL, Close TJ, Roberts PA, Hu Z, Wanamaker S, Lucas MR, Chiulele R, Cisse N, David A, Hearne S, Fatokun C, Diop N, and Ehlers JD (2013) Gene pools and the genetic architecture of domesticated cowpea. The Plant Genome 6 (3).

IFAD, International Fund for Agricultural Development (2001) Assessment of Rural Poverty: Western and Central Africa. http://www.ifad.org/poverty/region/pa/english.pdf.

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

Lavin M, Herendeen PS, and Wojciechowski, MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. System Biology 54: 530 - 549.

Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, and Close TJ (2011) Cowpea–soybean synteny clarified through an improved genetic map. The Plant Genome 4 (3): 218 - 225.

Lucas MR, Huynh BL, Silva PS, Cisse N, Drabo I, Ehlers JD, Roberts PA, and Close TJ (2013a) Association studies and legume synteny reveal haplotypes determining seed size in Vigna unguiculata. Frontiers in Plant Science 4.

Lucas MR, Ehlers JD, Huynh BL, Diop NN, Roberts PA, and Close TJ (2013b) Markers for breeding heat tolerant cowpea. Molecular Breeding 31 (3): 529 - 536.

Lucas MR, Huynh BL, Ehlers JD, Roberts PA, and Close TJ (2013c) High resolution single nucleotide polymorphism genotyping reveals a significant problem among breeder resources. The Plant Genome 6 (1).

Lucas MR, Ehlers JD, Roberts PA, and Close TJ (2012) Markers for quantitative inheritance of resistance to foliar thrips. Crop Science 52 (5): 2075 - 2081.

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

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, and 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.

Purugganan MD, and Fuller DQ (2009) The nature of selection during plant domestication. Nature 457 (7231): 843 - 848.

Schmutz, J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MMS, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, and Jackson SA (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genetics 46: 707 - 713.

Figure 9: Principle component analysis of 1,536 SNP data from 214 landraces and improved cultivars of cowpea. Samples are colorized based on their geographic origin and the two parents used in this study are labeled. Principle component 1 distinguishes cowpeas collected in West Africa from those collected in South or East Africa. Principle component 2 separates cowpeas primarily collected from outside Africa from those collected within Africa.

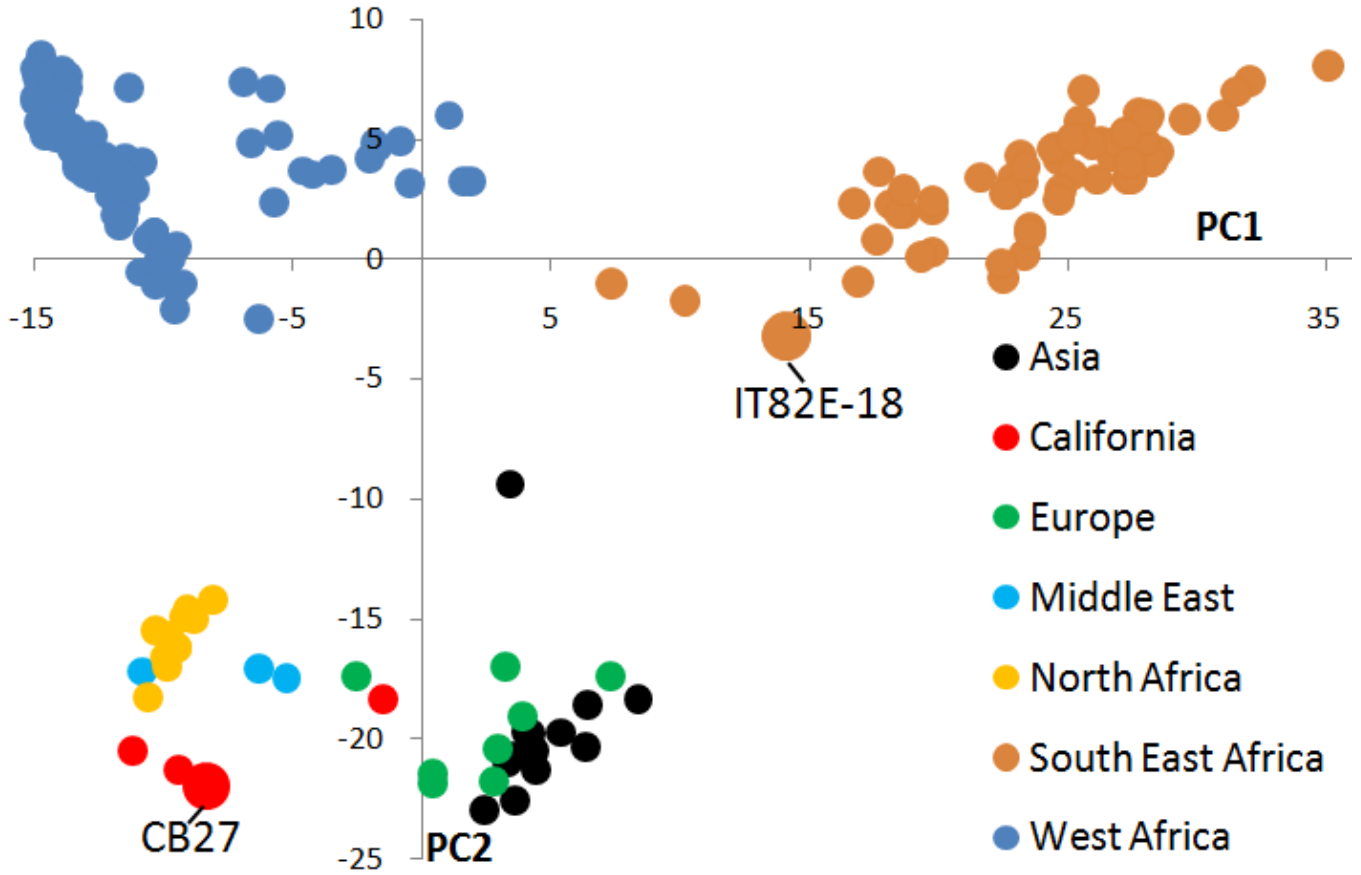


Figure 10: Marker-assisted introgression of the IT82E-18 haplotype at the Css-1 locus into a CB27 background. Samples are divided into recombinant inbred line population (RIL), cycle 1, cycle 2, and inbred selections. Samples are further divided based on their QTL content for three segregating QTL (i.e. 2.5 means samples are homozygous for large seed alleles at two QTL and heterozygous for large seed alleles at the third QTL).

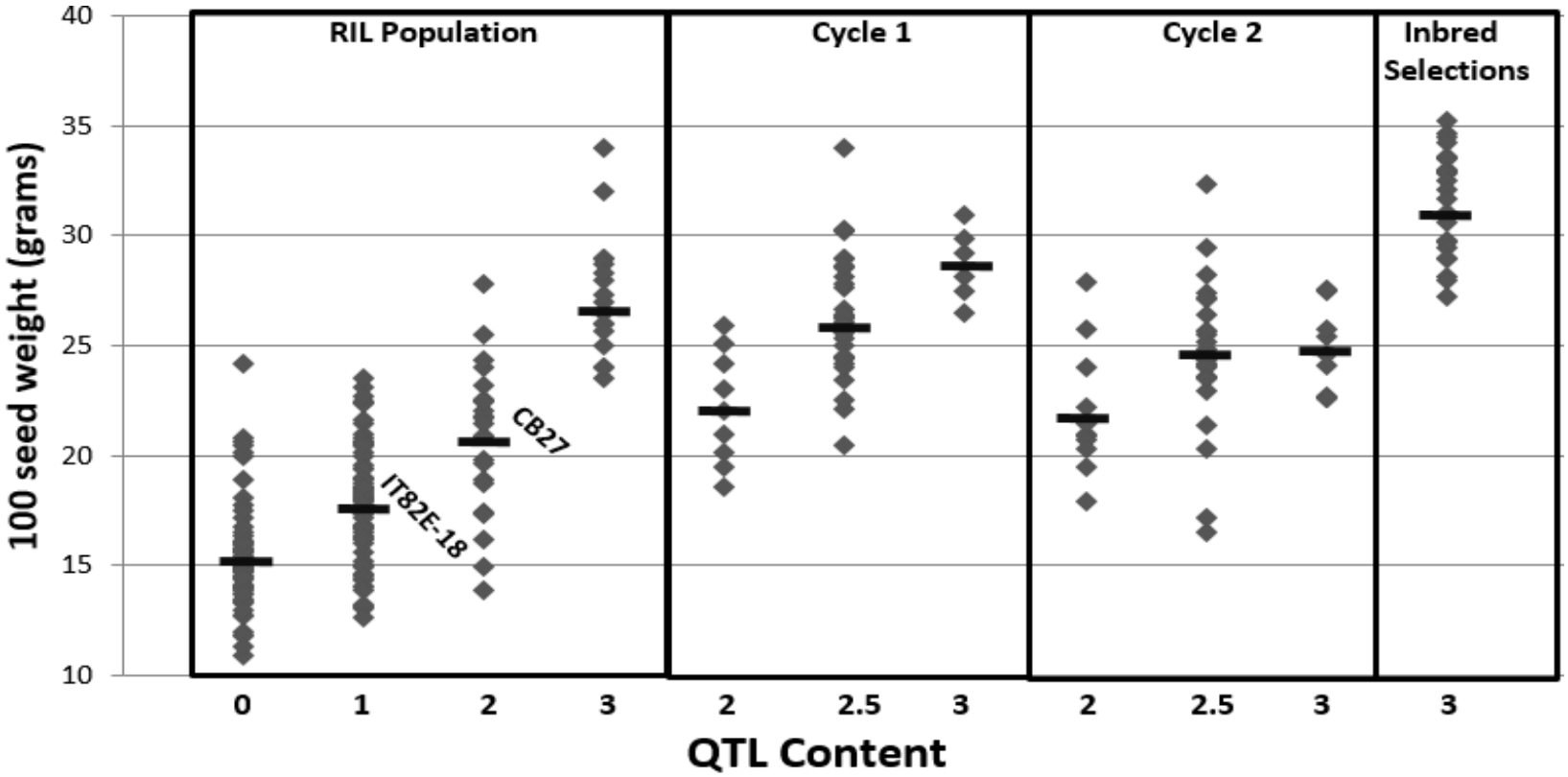
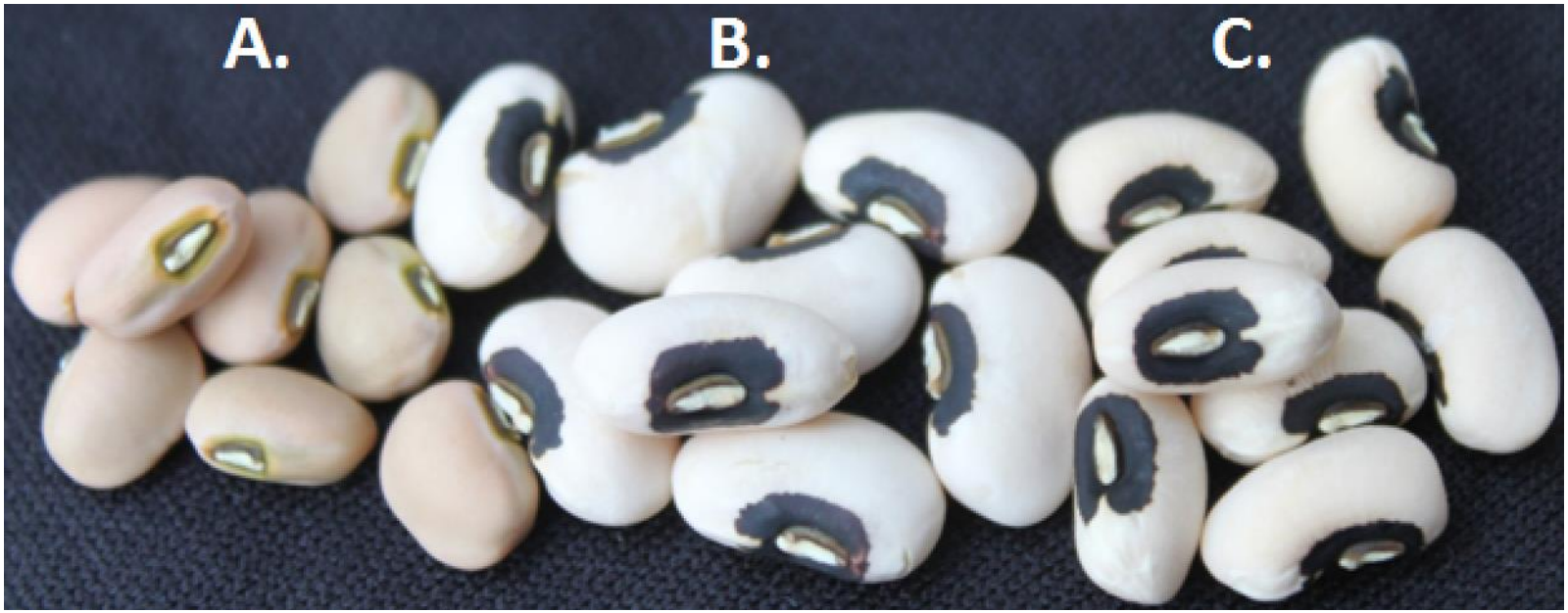


Figure 11: Cowpea seeds for the two parents A) IT82E-18 and C) CB27 used in this work. B) Cycle 1 F4 introgression line homozygous for large seed alleles at all three segregating QTL which has a 100-seed weight of 32 grams and a blackeye seed type like the recurrent parent CB27. Each sample has eight seeds represented in the picture.



Conclusion

Marker-assisted approaches to plant breeding require knowledge concerning the content and organization of the genome and how genetic factors relate to agriculturally important traits. To develop a better understanding of the cowpea genome an improved consensus genetic map of cowpea was created which is described in Chapter 1 of this dissertation. The approach applied 1,536 SNP genotyping to 11 important bi-parental recombinant inbred line populations. These populations were chosen for study based on their availability, their importance for breeding initiatives, and their genetic diversity. Genotype data from these populations was used to group and order loci on linkage groups using mapping by recombination frequency approaches. Because all populations were genotyped using the same markers the maps could be integrated into a consensus genetic map which represents a comprehensive description of the content and organization of the cowpea genome. A total of 1,107 SNP markers were placed on 11 consensus linkage groups and through comparisons with earlier genetic maps and the soybean genome the improved quality of the new map was demonstrated. This genetic map became an essential resource for Chapters 2 and 3, in addition to other genetic studies of cowpea. While SNP genotyping technologies can be considered expensive up front investments, knowledge provided by these approaches is essential to genetic studies and these methods continues to be broadly applied. Economy of scale and improved technologies continue to reduce the price of genomic studies. These factors, in addition to the simplicity and autonomy of SNP genotyping provided by outsourced service providers, have facilitated our work in the application of these technologies to breeding, particularly in developing countries of Africa. Subsidized and outsourced approaches to molecular breeding can

circumvent infrastructural deficiencies of developing nations and thereby allow scientists access to the latest technologies and approaches to crop improvement. Future work on genetic mapping of cowpea, specifically at UCR, will utilize much higher-density genotyping platforms and will be integrated with physical maps. Because legacy 'immortal' populations will be genotyped using newer technologies, improved association studies, possibly using the same phenotype data, can be pursued.

The genetic map established in Chapter 1 permitted the development of associations between allelic variation and the inheritance of seed size which is the topic of Chapter 2. This was accomplished by first observing several experimental populations in greenhouse and field trials for seed size, recorded as 100-seed weight. This phenotypic information was then combined with SNP genotype information using inclusive composite interval mapping to identify quantitative trait loci (QTL) associated with the inheritance of seed size. Seed size data provided by the United States Department of Agriculture for a diversity collection of cowpea samples was also utilized in a genome-wide association study. These approaches not only identified regions of the genome important for the inheritance of seed size, but also predicted pedigree-specific effects of allelic variation such that a strategy to pyramid favorable allelic variation through marker-assisted breeding could be pursued. One interesting observation documented in Chapter 2 is the orthology of seed size. Published seed size marker-trait associations for soybean were compared to cowpea loci associated with seed size using knowledge of synteny with other legumes established in Chapter 1. Interestingly several regions important for seed size in cowpea are also orthologous to regions of the soybean genome

that are important for seed size. This speaks to the orthology of genetic factors controlling seed size in legumes which can be exploited to translate knowledge from studies on plants of different genera. Future work, specifically at UCR, in the development and validation of association between allelic variation and the inheritance of seed size should explore the vast amount of genetic diversity available in germplasm collections. Seed collections of field and greenhouse grown materials have recently been inventoried which provides thousands of data points, including seed size data, which can be cross-referenced to genotyped accessions to pursue a genome-wide association study.

The broad study on the genetics of seed size presented in Chapter 2 provided opportunity to select associations with the largest predicted impact. This is the theme of Chapter 3 which describes a marker-assisted breeding effort to pyramid favorable allelic variation into one cultivar. The QTL *Css-1* described in Chapter 2 has the largest predicted effect and can be considered rare because it is unique to landraces and improved cultivars collected/developed in Southeastern Africa. The variety 'California Blackeye 27' (CB27) was chosen as the recipient of allelic variation contributing larger grain size, donated by an African cultivar. Not only is large seed size a desired feature of blackeyed peas, but the pedigree of CB27 facilitated screening of new breeding lines. CB27 is resistant to several pathogens and flowers quickly, even under long days, which was not the case for all parents used in the association studies. Furthermore, improving a trait in an already intensively bred cultivar promotes the continuation of my work on seed size. The impact of *Css-1* introgression was assessed during two cycles of backcrossing. A field trial of a family of introgression lines points out the successes and potential

pitfalls of this introgression effort. Breeding lines with up to 52% larger seeds that also performed well in terms of yield and other important agronomic traits were identified, however, issues concerning seed quality and processing of such dramatically larger seeds were also observed. Future work should revisit other introgression families because it was noted they performed well in terms of these negative attributes in greenhouse screenings. Ongoing efforts could also pursue more backcrossing, deployment of *Css-1* in other pedigrees, examining the nutritional impact of seed size loci, identification and functional analyses of candidate genes, and the orthology of seed size among legumes.

Three additional publications that are presented in Appendices A, B, and C are relevant to marker-assisted breeding of cowpea. Appendix A describes the unforeseen impacts of genotype data. Among the most important knowledge provided by genotype data is validation of pedigrees which can remove accumulated errors in breeding programs and is demonstrated to improve germplasm collections and efficiency of genetic studies. Appendix B details a study of heat tolerance during reproductive development in a bi-parental population of cowpea and describes haplotypes associated with the inheritance of resistance to heat-induced injury during reproductive development. Appendix C describes a study of the inheritance of resistance to feeding damage caused by foliar thrips and provides haplotypes associated with resistance to this insect pest. In conjunction with the new knowledge provided in the first three chapters of this dissertation, and other association studies in cowpea, resources are being developed to take a realistic approach to genomic selection.

Appendix A

High Resolution Single Nucleotide Polymorphism Genotyping Reveals a Substantial Problem Among Breeder Resources

Abstract

Modern breeders work with information-rich genotype data to make selection decisions, develop marker-trait associations, and to assess diversity. One benefit stemming from such information includes verification of pedigree records and lower costs associated with marker-assisted selection. In this communication we provide a detailed example of errors accumulated during the development, maintenance, and distribution of recombinant-inbred populations. This example on cowpea will be of more general interest for other breeder communities.

Introduction

The logistics associated with a modern breeding program can be complex; relying on accuracy and communication between plant breeders, pathologists, quantitative geneticists, and support staff. International and academic facets may bring additional challenges to already error prone activities including the development, maintenance, and distribution of lines. Furthermore, practices such as bulking of seed and the maintenance of within-accession variation among landraces must be considered when pursuing marker-assisted approaches to breeding.

Cultivars, germplasm, and populations that have been bred by specific design have several expected characteristics including allelic diversity, heterozygosity, and individuality. The existence of rogues, individuals which violate these premises, is documented among important crop and model species. Authors often do not elaborate on the potential origin of rogues, but in some cases hypotheses have been formed with outlandish biological explanations. From a practical standpoint the unintentional use of rogues can be problematic when used for breeding or when developing breeder resources. Undetected rogues may also have financially costly impacts if they are included in field trials or are repeatedly genotyped.

Fortunately, insight provided by high-throughput genotyping can assess how well an individual matches its pedigree record. The original intent of genotyping resources was not for error detection "forensics"; however, in actual practice the benefits are immediate and significant. Here we discuss typical examples of rogues, their impacts, and detection using data from our work on cowpea (*Vigna unguiculata*).

Origin and Detection of Rogues: Unintentional outcrossing events are likely sources of rogues. Even species that normally have high frequencies of self-pollination will occasionally out-cross (Lloyd et al., 1992). These outcrosses can lead to heterozygosity and the presence of non-parental alleles. Organizational errors (i.e. harvesting seed from volunteer plants, mislabeling, different paths of single-seed descent, and/or mixing seed from different lines) could also create rogues, including duplicate lines in a collection.

All thirteen populations used to construct the consensus genetic map of cowpea (Lucas et al., 2011) contained at least some rogues (Table 6), which range from 3% to 31% providing an average of 11.3% rogue. In our experience, these errors have been more common than expected. But, a search of the literature provides a fair number of documented cases of such problems. Genotype information has been used to identify duplicated lines of apple (*Malus x domestica* Borkh.) (Hokanson et al., 1998), orange (*Poncirus trifoliata* L. Raf.) (Fang et al., 1997), and rice (*Oryza sativa* L.) (Virk et al., 1995). In our work, ten out of the eleven recombinant inbred populations genotyped for 1,536 SNPs were found to contain duplicate individuals (Table 6). In that work and in subsequent QTL analyses (Lucas et al., 2012a, Lucas et al., 2012b) duplicate lines and other rogues were omitted from data analysis. Interestingly, cowpea lines that are identical often have a sequential or similar name (Supplemental File 7). Seventeen out of 54 instances of duplications occurred between lines of sequential naming (i.e. line -036 identical to line -037), while fourteen duplications were found between lines with a similar name (i.e. line -049 identical to line -094, line -088-2 identical to line -002, line -

084 identical to line -184). Such duplications seem most likely to be the result of human error which may have occurred at a number of different stages during the creation, maintenance, or distribution of these inbred populations.

Within-accession variation has been diagnosed via molecular markers in rice (Olufowote et al., 1997), spanish melon (*Cucumis melo* L.) (Lopez-Sese et al., 2002), and cowpea (Hearne et al., 2010). A genome covering set of 80 SSRs was recently used to verify pedigree records in apple (*Malus* species) (Evans et al., 2011). In addition to the identification of 15 rogues, two plants known by the same name (Priscilla) were found to be different based on genotype information collected in that study. We assessed the possibility of cowpea within-accession variation by genotyping inbred stocks carrying the same name, but provided from different sources of seed. This was performed to capture one aspect of within-accession variation, different paths of single-seed descent rather than assessing heterogeneity of one seed stock. Although some accessions were identical between seed sources, based on 1536-plex SNP genotyping, six sets of inbreds were identified which were different at many loci (Table 7). These polymorphisms between lines with the same name were only dispersed among some linkage groups and tend to be localized to the ends of linkage groups. These haplotype blocks provide evidence for divergent descent from a common parent.

Molecular markers can also be used to assess the frequency of self-fertilization. SNPs are being considered for a molecular hybridity test in faba bean (*Vicia faba* L.) (Cottage et al., 2012). In that work 31 out of 32 plants self-fertilized in an over-winter glass house while one out-crossed. A similar situation was observed among the SNP

genotyped populations of cowpea. Six out of the eleven recombinant inbred populations of cowpea used to build the consensus map contained lines that were heterozygous far beyond expectation based on the number of inbreeding generations (Supplemental File 8), often correlating with the presence of non-parental alleles. This type of variation may arise from unintentional outcrossing, organizational errors, or when inherently variable land race accessions are used as parents.

As indicated above, genotype information can also identify carriers of non-parental alleles, similar to what was observed in wheat (*Triticum aestivum* L. and *Triticum turgidum* L.) (Khan et al., 2000), banana (*Musa*) (Crouch et al., 1999), and rapeseed (*Brassica napus*) (Trick et al., 2009). During the construction of the cowpea consensus map, nine out of the eleven recombinant inbred populations were found to have at least one individual carrying non-parental alleles (Lucas et al., 2011). In that work a fixed array genotyping platform was used which provided information for some markers which were fixed among the parents. Rogues were quickly identified when working with genotype information graphically (Figure 12). As a part of this communication we detail non-parental allele containing rogues among the populations used to construct the cowpea consensus map (Supplemental File 8).

A notorious example of apparent outcrossing in *Arabidopsis thaliana* has led to controversial discussion. Lolle et al. (2005) provide a biological explanation for the genome wide inheritance of non-parental alleles which involves a hypothetical cache of ancestral RNA, a notion that has no support from any prior work in any organism. Alternative explanations crafted on the basis of little experimental data are also being

considered (Ray, 2005; Chaudhury, 2005; and Comai, 2005; among many others); however we underline the elegant work of Peng et al. (2005) which provides the most parsimonious explanation concerning the origin of non-parental alleles, unintentional outcrossing.

Impact of Rogues: Rogues are particularly problematic when taking marker-assisted approaches to breeding. Accurate genetic maps and statistical estimates are required to associate markers with traits. Maps are often constructed by observing recombination frequencies among members of a population. Progeny provide a sample of possible recombination events occurring during meiosis and no two individuals should be identical. Additionally, rogues carrying non-parental alleles may contribute phenotypic variation disregarded by statistical models intended to operate on populations of bi-parental design. Lines carrying non-parental alleles are known to be significant obstacles for map construction (Ming et al., 1997; Lucas et al., 2011). Inaccurate estimations of genetic distances and associations may confound attempts aiming to utilize a marker-assisted approach to selection. To assess the impact of rogues on developing resources for breeding, we built genetic maps and performed QTL analysis with and without rogue individuals. Maps constructed with and without rogues are different (Figure 13) and often have greater distances between bins when rogues are included. QTL analyses are also affected (Figure 14), where the inclusion of rogues can suggest the existence of a QTL where none are known to exist. Furthermore, rogues with excessive heterozygosity or non-parental alleles which perform strong phenotypically may be incorrectly assumed to be inbred while these individuals are actually benefitting from hybrid vigor or alleles

thought to be absent from its pedigree. The unintentional use of these individuals for breeding or for the development of marker-trait associations may lead to unpredictable outcomes including multiform progeny and linkage drag.

Population size is a major constraint for marker-assisted breeding initiatives because of financial costs associated with genotyping and phenotyping. Selection decisions and marker-trait associations are typically developed by observing the performance of lines in replicated, multi-location field trials. The logistics associated with these operations demand a substantial proportion of financial resources available to a breeding program. Removing rogues before phenotyping trials and extensive genotyping would eliminate unnecessary expenses. Therefore, rogues are not only problematic, but also financially inefficient to maintain.

Geneticists should put a greater emphasis on developing high quality populations rather than relying on historical populations made by breeders who may be more permissive of rogues. Improved analyses, community resources, and financial efficiency could be realized by approaching genotype data from a forensics perspective. This approach could also be used as a quality control measure for the future development of lines. Our group is verifying pedigree records for members of a MAGIC population by genotyping the progeny for genome-wide markers known to segregate among the intended parents. This work has helped verify cross-pollination events and helps to ensure a quality community resource is being provided on the basis of genotype validated seed. Validation via genotyping would be particularly valuable for species which are difficult to cross or for lines/populations which will be heavily used.

References

Chaudhary A (2005) Plant genetics: Hothead healer and extragenomic information. Nature 437, E1.

Comai L (2005) A toxic mutator and selection alternative to the non-mendelian RNA cache hypothesis for hothead reversion. The Plant Cell 17: 2856 - 2858.

Cottage A, Gostkiewicz K, Thomas JE, Borrows R, Torres AM, and O'Sullivan DM (2012) Heterozygosity and diversity analysis using mapped single nucleotide polymorphism in a faba bean inbreeding programme. Molecular Breeding 30: 1799 - 1809.

Crouch JH, Crouch HK, Constandt H, Van Gysel A, Breyne P, Van Montagu M, Jarret RL, and Ortiz R (1999) Comparison of PCR-based molecular marker analyses of *Musa* breeding populations. Molecular Breeding 5: 233 - 244.

Evans KM, Patocchi A, Rezzonico F, Mathis F, Durel CE, Fernandez-Fernandez F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Gianfranceschi L, Komjanc M, Lateur M, Madduri M, Noordijk Y, and can de Weg WE (2011) Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: trueness-to-type of cultivars and their parentages. Molecular Breeding 28: 535 - 547.

Fang DQ, Roose ML, Krueger RR, and Federici CT (1997) Fingerprinting trifoliate orange germ plasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. Theoretical and Applied Genetics 95: 211 - 219.

Hearne S, Franco-Duran J, and Magembe E (2010) Optimal units of selection – how many plants are representative of diversity in collections of cowpea landrace

accessions? Proceedings of the World Cowpea Conference Hosted September 27th – October 1st in Saly, Senegal; Program and Book of Abstracts: 46 - 47.

Hokanson SC, Szewc-McFadden AK, Lamboy WF, and McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica borkh*. core subset collection. Theoretical and Applied Genetics 97: 671 - 683.

Khan IA, Procunier JD, Humphreys DG, Tranquilli G, Schlatter AR, Marcucci-Poltri S, Frohberg R, and Dubcovsky J (2000) Development of PCR-based markers for a high grain protein content gene from *Triticum turgidum ssp. dicoccoides* transferred to bread wheat. Crop Science 40: 518 - 524.

Lloyd DG, and Schoen DJ (1992) Self- and cross- fertilization in plants I functional dimensions. International Journal of Plant Sciences 153: 358 - 369.

Lolle SJ, Victor JL, Young JM, and Pruitt RE (2005) Genome-wide non-mendelian inheritance of extra-genomic information in *Arabidopsis*. Nature 434: 505 - 509.

Lopez-Sese AI, Staub J, Katzir N, and Gomez-Guillamon ML (2002) Estimation of between and within accession variation in selected Spanish melon germplasm using RAPD and SSR markers to assess strategies for large collection evaluation. Euphytica 127: 41 - 51.

Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, and Close TJ (2011) Cowpea - soybean synteny clarified through an improved genetic map. The Plant Genome 4: 218 - 225.

Lucas MR, Ehlers JD, Roberts PA, and Close TJ (2012a) Markers for Quantitative Inheritance of Resistance to Foliar Thrips in Cowpea. Crop Science 52: 2075 - 2081.

Lucas MR, Ehlers JD, Huynh BL, Diop NN, Roberts PA, and Close TJ (2012b) Markers for breeding heat tolerant cowpea. Molecular Breeding 31: 529 - 536.

Ming R, Brewbaker JL, Pratt RC, Musket TA, and McMullen MD (1997) Molecular mapping of a major gene conferring resistance to maize mosaic virus. Theoretical and Applied Genetics 95: 271 - 275.

Olufowote JO, Xu Y, Chen X, Park WD, Beachell HM, Dilday RH, Goto M, and McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40: 370 - 378.

Peng P, Chan SWL, Shah GA, and Jacobsen SE (2006) Plant Genetics: Increased outcrossing in hothead mutants. Nature 443: E8.

Ray, A (2005) Plant genetics: RNA cache or genome trash? Nature 437: E1 - E2.

Trick M, Long Y, Meng J, and Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploidy *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnology Journal 7: 334 - 346.

Virk PS, Newbury HJ, Jackson MT, and Ford-Lloyd BV (1995) The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theoretical and Applied Genetics 90: 1049 - 1055.

Table 6: Excerpt from Lucas et al. (2011) describing rogues among thirteen mapping populations of cowpea determined from the analysis of 1536-plex SNP genotype information.

*The number of individuals that are highly-heterozygous or non-parental in genotype (HNPG).

Population	Individuals Genotyped	Individuals used for Mapping	HNPG*	Genotypically Identical Sets of Individuals
CB27 x IT97K-556-6	95	92	1	2
CB27 x IT82E-18	166	160	2	4
CB27 x UCR 779	58	56	0	2
CB46 x IT93K-503-1	130	114	16	0
524B x IT84S-2049	91	85	5	1
Dan Ila x TVu-7778	113	79	11	23
Yacine x 58-77	141	97	43	1
Sanzi x Vita 7	142	122	11	9
IT84S-2246 x IT93K-503	93	88	5	0
IT84S-2246 x Mouride	92	87	5	0
TVu14676 x IT84S-2246-4	147	136	10	1
CB27 x 24-125B-1	108	87	18	3
LB30#1 x LB1162 #7	95	90	4	1

Table 7: Number of SNP polymorphisms found by genotyping two lines with the same name from six accessions of cowpea. The distribution of polymorphisms is indicated among the eleven linkage groups of the cowpea genome.

*With respect to a 1536-plex SNP genotyping platform.

Accession	Number of Polymorphic Loci*	Location of Polymorphisms on Linkage Groups		
		Beginning	End	Dispersed
Yacine	10	6	7	-
TVu-16722	63	2, 5	1	3, 9
TVu-15112	90	1, 2, 3, 10, 11	-	5, 7, 9
TVu-10513	103	1, 6, 9	3, 8	2, 4, 5, 10
58-77	113	3, 4, 5, 8	-	1, 2, 10, 11
TVu-14321	159	3, 8	4	1, 2, 5, 9, 10

Figures

Figure 12: Two individuals with non-parental genotype calls among a recombinant inbred population of cowpea. Individuals homozygous for allele A (A) and allele B (B) at SNP locus 1_0757 are shown using Illumina GenomeStudio genotype visualization software. The parents and 164 progeny are monomorphic and contain allele B, while two individuals are homozygous for allele A.

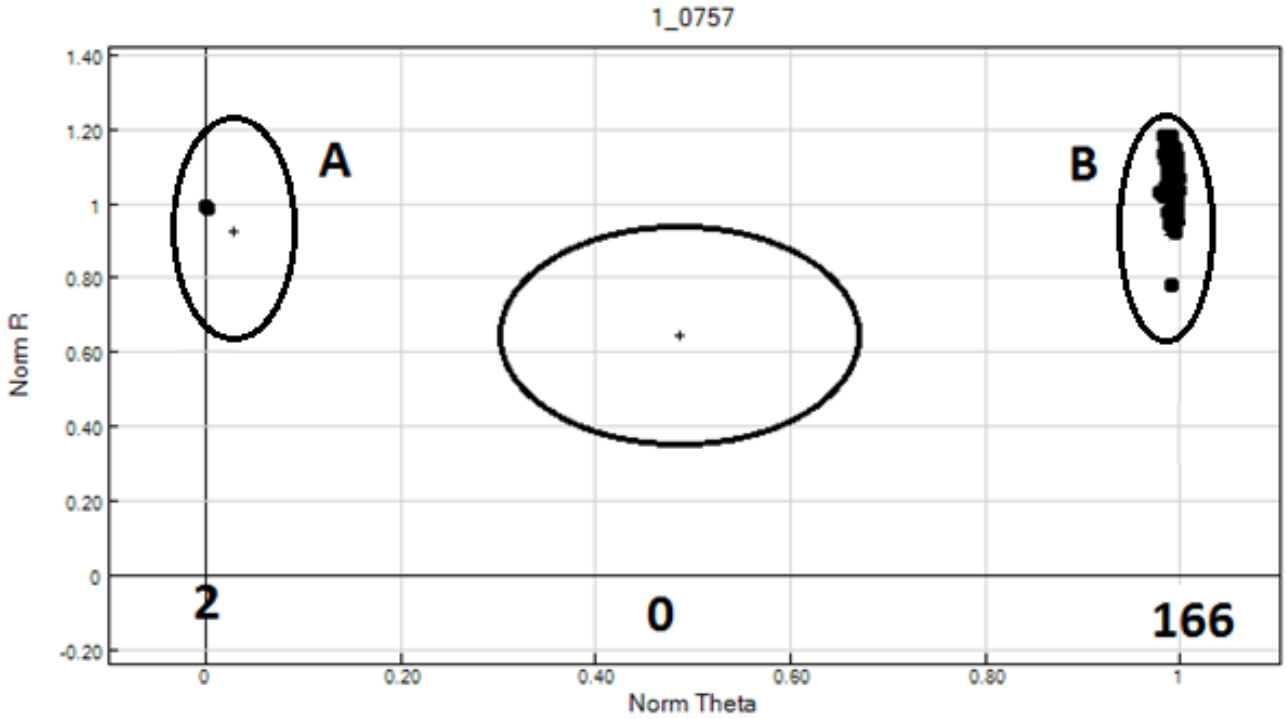


Figure 13: Linkage maps constructed with (A) and without (B) rogues.

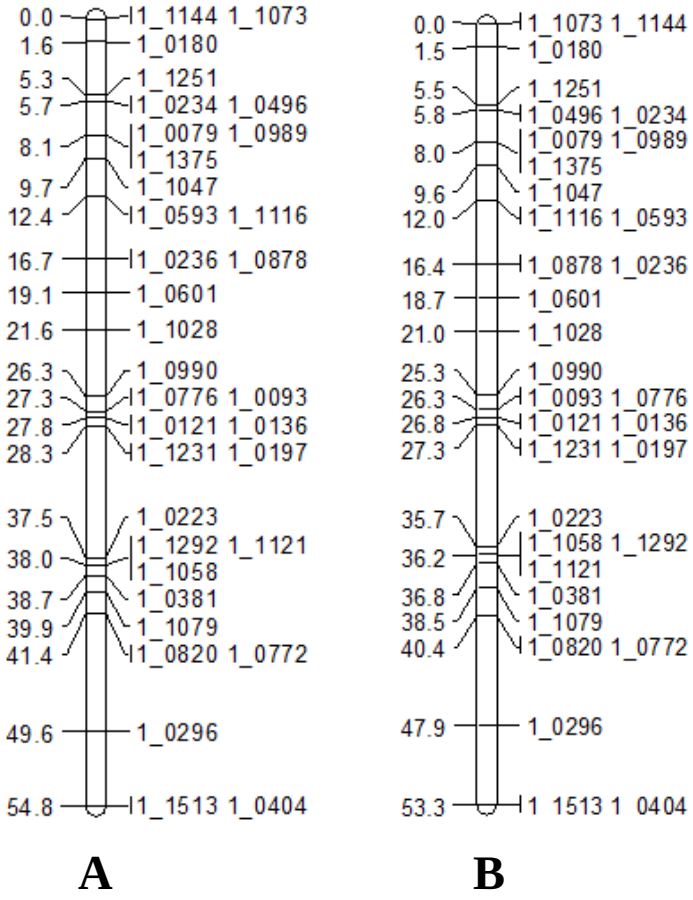
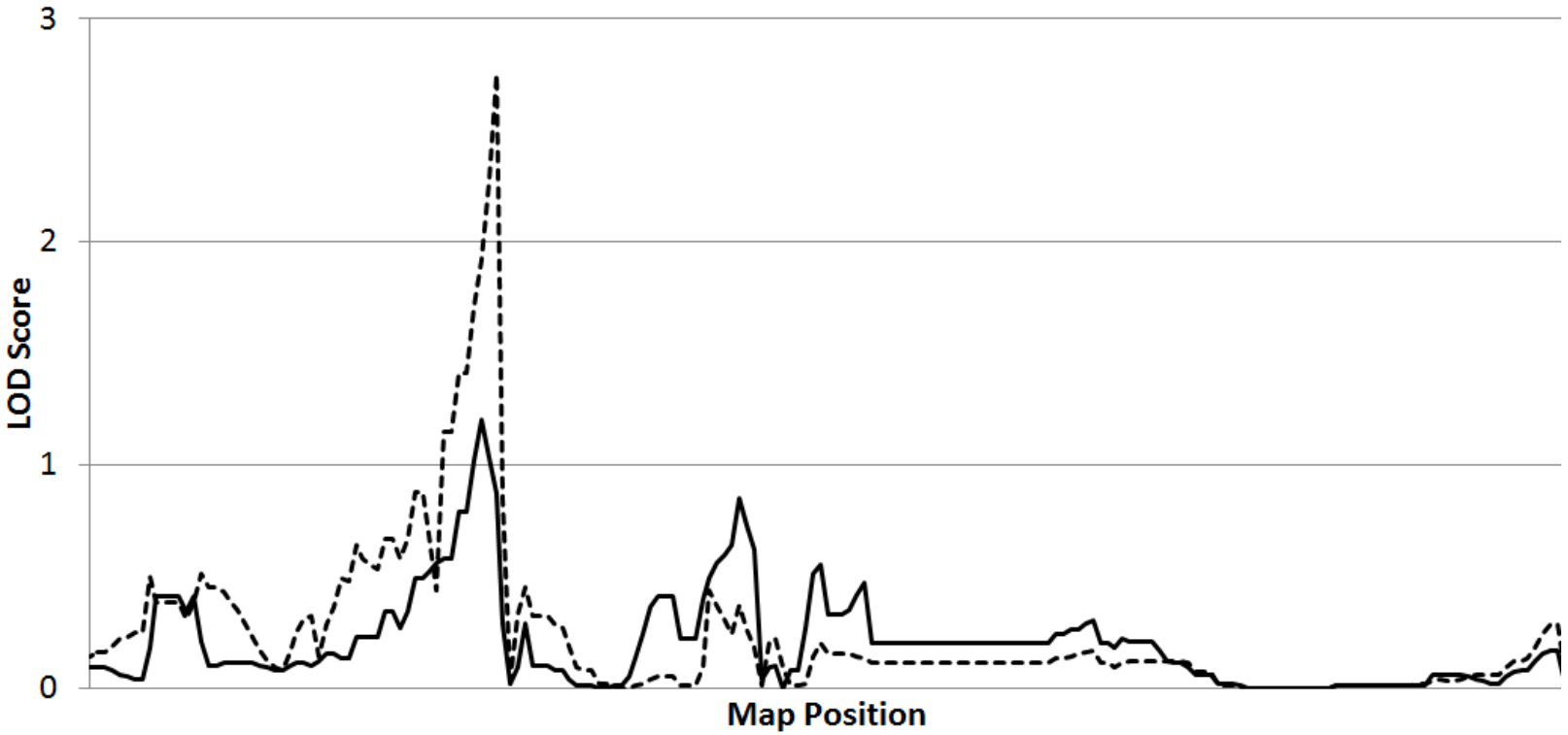


Figure 14: Logarithm of the odds (LOD) score traces for QTL analyses performed with (solid line) and without (dashed line) rogues.



Appendix B

Markers for Breeding Heat Tolerant Cowpea

Abstract

The warm season legume, cowpea (*Vigna unguiculata*), is an important crop that performs well in marginal environments. The effects of high temperature are among the most substantial challenges growers of cowpea face. Heat injury during late reproductive development sterilizes pollen such that no fruit is set. To study the inheritance of this trait and to deliver resources to breed cowpea with enhanced tolerance to heat we performed a QTL analysis using 141 individuals from a recombinant inbred population made from a cross between cowpea varieties CB27 and IT82E-18. Five regions which represent 9% of the cowpea genome explain 11.5 – 18.1% of the phenotypic variation and are tagged with 48 transcript derived SNP markers. Favorable haplotypes were donated by CB27 for four of these regions while IT82E-18 was the source of tolerance explained by the fifth QTL. Homeologous regions in soybean contain several genes important for tolerance to heat including heat shock proteins, heat shock transcription factors, and proline transporters. This work presents essential information for marker-assisted breeding and supports previous findings concerning heat induced male sterility in cowpea.

Introduction

Cowpea (*Vigna unguiculata*) is a warm season legume cultivated for grain, leaves, and fodder. It is closely related to mung bean (*Vigna radiata*) and shares more distant common ancestry with common bean (*Phaseolus vulgaris*), soybean (*Glycine max*), and pigeon pea (*Cajanus cajan*) (Choi et al., 2004). Primary cowpea growing regions include semi-arid zones of Sub-Saharan Africa, South-East Africa, India, the Americas, and parts of Asia where the subspecies "asparagus" or "yard-long" bean (*Vigna unguiculata* ssp. *sesquipedalis*) is cultivated for long immature pods. Cowpea germplasm is notably diverse, especially when considering tolerance to several biotic and abiotic stresses; however, the genetics of these traits are not sufficiently understood in the context of modern, marker-assisted, breeding. Many agronomically important traits display a continuous phenotypic distribution. These quantitatively inherited traits are typically influenced by several loci and the environment, and are difficult to breed using conventional methods reliant on phenotypic assessments.

From the standpoint of grain yield, deleterious effects of heat on reproductive development are the most damaging. In cowpea, damage from heat can manifest itself through inhibition of floral bud development (stage I heat effects) (Dow el-medina et al., 1986). Stage I effects are influenced by photoperiods and so this response is assumed to be influenced by phytochrome (Mutters et al., 1989a; Hall, 1992). Stage I effects only occur under long photoperiods and therefore may not be relevant to most tropical zones which exhibit only small changes in photoperiod. High temperatures during floral development, stage II, cause male sterility which results in fruit abortion (Warrag et al.,

1984). Anatomical studies of anther tissue development during high night temperatures describe distortions of microspore cells, the tapetal layer, and the endothecium (Ahmed et al., 1992). These reproductive tissue abnormalities may reduce translocation of proline from anthers to pollen (Ahmed et al., 1992), which has been associated with male sterility in cowpea (Mutters et al., 1989b). Readers interested in heat tolerance may find utility in reviews compiled by Hall (2012) and Wahid et al., (2007).

Associations between genotype and phenotype can expedite development of improved varieties containing favorable alleles for several traits through streamlined approaches to breeding. Key genomic resources facilitating marker-trait associations are now available for cowpea. These include single nucleotide polymorphism (SNP) genotyping assays (Muchero et al., 2009), a consensus genetic map containing 1107 transcript-derived SNP markers (Lucas et al., 2011), and knowledge of synteny with crop and model species (Wanamaker et al., 2011).

Here we describe the inheritance of tolerance to heat at pod set (stage II) in a domesticated, bi-parental, F_8 recombinant inbred line (RIL) population of cowpea. Tolerance-associated haplotypes are tagged with several SNP markers and represent the basis for marker-assisted breeding. Breeding lines containing different combinations of QTLs are used to assess the stacking potential of favorable haplotypes and correlated with the ability to set pods under high temperatures. Syntenic regions of the soybean genome are surveyed for potential candidate genes.

Materials and Methods

Discovery Population and Experiments: All marker-trait associations were discovered in the domesticated, bi-parental, F_8-RIL population CB27 x IT82E-18 consisting of 166 individuals which were previously genotyped (Lucas et al., 2011) using a 1536 SNP assay (Muchero et al., 2009). Six individuals previously thought to be part of this RIL population but characterized as rogues by SNP genotyping (Lucas et al., 2011) were omitted from the present analysis.

This population was grown in four replicates while the parents were grown in sixteen replicates using a complete randomized block design during two experimental greenhouse and two control field trials. Phenotypes were gathered from two trials in a greenhouse at the Citrus Experiment Station at the University of California Riverside during the summers of 2010 and 2011 (long days) with day temperatures of 110°F and night temperatures of 82°F (Supplemental File 9). Both parents are day neutral and do not exhibit stage I sensitivity (J. Ehlers, unpublished data, verified in the present work). Temperatures were recorded every ten minutes during the course of the experiments with a data logger and confirmed periodically with alcohol/mercury thermometers. Plants were examined after maturity by removing and counting the number of pods and peduncles. The average number of pods per peduncle was used as an estimate of stage II heat tolerance.

Two field trials using the CB27 x IT82E-18 population were also planted at the Citrus Experiment Station at the University of California Riverside during the summers of 2010 (July 8th) and 2011 (July 7th) with 100 seeds planted per 18 foot plot and 3 foot

row spacing. The average high temperature for the months of July – September was 33°C for both 2010 and 2011 respectively. The average low temperature during these months was 17°C and 18°C for 2010 and 2011, respectively. These field trials which were conducted under long-day photoperiod, moderately-warm day time and cool night time temperature field trials allowed the identification of lines with sterility issues not caused by sensitivity to heat. These lines appeared parthenocarpic or stenospermocarpic such that pods were set but seeds were aborted early in development or were not found. These plants were easily identified upon visual inspection because their pods were often shorter, up-right, and lacked contours where seeds would normally be present. Data from these lines were excluded from QTL analysis.

Marker-Trait Associations: Genotype calls and phenotypic data collected from the CB27 x IT82E-18 population were combined for QTL analysis with a genetic map by interval mapping (Jansen et al., 1994) provided through MapQTL 5 (Van Ooijen, 2004). Marker order and distances from the CB27 x IT82E-18 population specific map were anchored to the previously published cowpea consensus map (Lucas et al., 2011). When more than one population specific linkage groups represented one consensus linkage group, the population specific linkage groups were merged by adding map distances, and by adding fifty centi Morgans to the group being merged. Only QTLs surpassing a LOD threshold of 3.0 are reported. Parent genotype calls for SNPs linked to heat tolerance were used to define a list of favorable and alternative haplotypes termed *Cht – 1* through *Cht – 5*. The lone marker linked to *Cht – 2* was omitted when calculating distance between haplotype marker bins.

Percent similarity to favorable haplotypes was determined for members of CB27 x IT82E-18 by comparing genotype calls for markers linked to the five QTLs. These values were then compared to phenotypic performance to determine correlations and to evaluate QTL stacking potential among the population. This was performed for all pairwise combinations of QTLs in addition to independent QTLs.

Synteny to Soybean: Synteny information presented in Lucas et al. (2011) was extracted from HarvEST: Cowpea 1.27 (Wanamaker et al., 2011) which contains *Arabidopsis* information from TAIR 10 (Lamesch et al., 2011), soybean information from JGI Glyma1 (Schmutz et al., 2010), and *Medicago truncatula* information from Mt3.5 (*Medicago truncatula* HapMap Project, 2010). Six gene ontology functions which play roles in tolerance and response to heat were surveyed for abundance among the soybean candidate regions. These include heat shock proteins, heat shock transcription factors, DNA J heat shock proteins, late embryogenesis abundant hydroxyproline rich glycoproteins, hydroxyproline rich glycoproteins, and proline transporters.

Results

Based on Pearson correlation analysis the number of pods per peduncle among lines were strongly correlated across greenhouse trials (r = 0.91). CB27 averaged 1.84 (SE = 0.08) and 1.99 (SE = 0.11) pods/peduncle during the 2010 and 2011 greenhouse trials, respectively, resulting in an average of 1.91 pods/peduncle when considering performance across trials. The more heat susceptible parent IT82E-18 averaged 0.86 (SE = 0.07) and 1.21 (SE = 0.07) pods/peduncle during the 2010 and 2011 greenhouse trials respectively, resulting in an average of 1.04 pods/peduncle across trials. Figure 15 depicts the range of phenotypes for the CB27 x IT82E-18 discovery population which was 0.00 – 2.22 pods/peduncle, yielding a population mean of 0.64 and 1.02 for the 2010 and 2011 trials, respectively. Images of this trait were captured during the 2010 greenhouse trial and examples of tolerance and susceptibility phenotypes are presented in Figure 16. Nineteen lines with sterility issues not caused by sensitivity to heat were consistently identified among both field trials.

Table 8 summarizes significant findings for five regions, *Cht – 1* through *Cht – 5*, spanning a total of 61.42 cM (9% of the total genome) of the cowpea genome contributing to the inheritance of stage II heat tolerance. LOD score traces from QTL analysis of both greenhouse experiments are presented in Figure 17. Favorable alleles of *Cht – 5* were donated from the more heat sensitive parent, IT82E-18, and explain approximately 11.5 percent of the heritable tolerance. CB27 was the source of the favorable alleles for the majority of QTLs, *Cht – 1* through *Cht – 4*, which independently explain between 16 and 18.1 percent of the heritable tolerance.

Favorable and alternative allele haplotypes using 48 SNP markers linked to the five cowpea stage II heat tolerance QTLs are presented in Supplemental File 10. Half of these are linked to *Cht – 4*, whereas, only one is linked to *Cht – 2*. The average distance between unique bins within these regions is 1.23 cM. Supplemental File 11 presents correlations between phenotypic performance and haplotypes among the discovery population for all five QTLs, all possible pairwise combinations of QTLs, and joint consideration of all QTLs. Favorable haplotypes are positively correlated with phenotypic performance for all five QTLs independently ($0.34 \leq r \leq 0.42$). Joint consideration of all QTLs and all pairwise combinations of QTLs are also positively correlated with pods/peduncle among the discovery population ($r = 0.49$, and $0.23 \leq r \leq 0.43$, respectively).

These five QTLs were syntenic with 77.6 Mb (7%) of the 1.1 Gb soybean genome which are summarized in Table 9. Six potentially interesting functional annotations are abundant within these regions including DNA J Heat Shock Proteins (DNAJ HSP), HydroxyProline-Rich glycoproteins (HPR), Heat Shock Proteins (HSP), Heat Shock Transcription Factors (HSTF), Late Embryogenesis Abundant HydroxyProline-Rich glycoproteins (LEA HPR), and Proline Transporters (PT). At least one of these annotations can be found within the syntenic regions of soybean except for the region syntenic with *Cht – 2*.

Discussion

In order to minimalize the impacts of unpredictable weather on agriculture, new varieties of crops should be developed with enhanced tolerance to drought and temperature. Sub-Saharan is a critical growing region for cowpea and the amount of productive land is forecasted to decrease up to 2.6% due to climate change (Lane et al., 2007). Heat tolerant varieties of cowpea are important for maintaining agriculture in such marginal environments. A better understanding of the inheritance of this important trait may also be useful for other crop species.

Contributions from studying the molecular responses plants evoke in response to environmental stimuli should aid efforts in developing new varieties of plants. Surveying candidate regions for genes known to play roles in stress response may support associations between phenotype and genotype, and provide a framework for cloning and characterizing underlying genetic factors. For crops lacking high-quality reference genomes information can be attained by evaluating synteny with more comprehensively studied species. Relationships between the genomes of cowpea and important model and crop species including *Arabidopsis*, *Medicago truncatula*, and *Glycine max* translate an abundance of gene annotations across species (Lucas et al., 2011). Reference genome resources for the more closely related commonbean, *Phaseolus vulgaris*, or Mung bean (*Vigna radiata*) will further improve legume comparative genomics and are expected in the near future.

Observations of heat tolerance are different between trials. Although the results from the two experiments are strongly correlated, heat susceptibility was more

pronounced in the 2010 trial. We believe this is most likely due to differences in calibration of temperature control. This is the most likely cause of variation between trials and may explain the skew towards susceptibility for the 2010 trial. As stated in the results, phenotypic performance within lines is strongly correlated between trials and analysis of data from the two trials consistently identifies the same QTLs.

Based on inheritance studies, Marfo et al. (1992) provided evidence for two dominant genes controlling the majority of the heritable tolerance to heat at pod set in cowpea. In that study heat tolerant cowpeas 'Prima' and TVu 4552 were used to develop F1, F2, and backcross populations with heat sensitive cowpeas. However, their results also present evidence for QTLs controlling heat tolerance including: a) F1 of tolerant x sensitive, on average, are less tolerant than either of the tolerant parents, b) F1 of sensitive x sensitive is more tolerant than either sensitive parent, c) the F1 of two tolerant parents were more tolerant than either parent in one trial, d) some F2 and backcross populations did not match ratios expected from one or two dominant genes. Observations from the present study provide more evidence for a complicated inheritance of heat tolerance during pod set. A typical 1:1 segregation ratio indicative of the effects of a single gene was not observed among the inbred progeny (Figure 15). Furthermore, transgressive segregation was observed among the progeny, indicating that neither parent is a donor of all favorable alleles. The results from the QTL analyses are consistent with this idea; while CB27 contributed most of the favorable alleles the more susceptible parent IT82E-18 contains allelic variation at Cht - 5 which contributes enhanced tolerance among the offspring. QTLs are known to control tolerance to heat in several

other important crops including rice (Ye et al., 2012), maize (Frova et al., 1994), and wheat (Hays et al., 2007). While it would have been more convenient to tag the single dominant gene observed by Marfo et al., (1992) the current study used a population for which this gene was likely of lower impact or was not segregating. It is important to note the heat tolerant cowpeas used in Marfo et al., (1992) contribute to the pedigree of the heat tolerant parent used in this study, CB27 (Ehlers et al. 2000). The inbred population derived from the cross of CB27 and IT82E-18 would not segregate for the loci described in 1992 if only one allele existed among both parents. Alternatively, if the allelic compositions at these loci are different between the parents their effects may have been less dramatic in the population due to the influence of other loci (QTLs). This is the first report we are aware of which links QTLs/SNPs to heat tolerance in cowpea.

The performance of QTLs in a breeding program is influenced by how well they are defined. The ultimate definition of a QTL is a DNA sequence of a resistance-associated region from the donor parent; however, technology capable of delivering this resolution is underdeveloped and may be impractical for current breeding initiatives. Characterization of multiple SNPs within resistance-associated regions can be used to develop favorable haplotypes which can be evaluated using a variety of platforms. The more markers that are used to define a QTL the greater the probability that the trait determinant will remain linked to markers during selection. Breeders wishing to employ a marker-assisted approach to breeding require user-friendly, trait-associated markers. Such markers can be used as the basis for making breeding decisions by determining how similar the material genotyped is to favorable genotypes.

The primary limitation of any association study is the broader applicability of the results. This study, which is based on one population in a greenhouse environment, identified five regions of the cowpea genome associated with tolerance to heat. CB27 contributes favorable alleles for four of the regions while the more sensitive parent, IT82E-18, contributes favorable alleles at *Cht – 5*. Combinations of *Cht – 5* with any of the other four QTLs have more positive correlations with tolerance to heat than combinations using any other two QTLs. Furthermore, lines carrying favorable haplotypes for all five QTLs have the most positive correlation with heat tolerance. This means the joint consideration of markers linked to all five QTLs is the best predictor of tolerance to heat. These results indicate pyramiding favorable haplotypes comprising these QTLs will result in heat tolerant cowpea varieties. The five regions we report represent a small proportion of the total mapped genome (9%). Additional allelic variation important for tolerance to heat might exist at other loci and may be different from those we describe depending on the environment and pedigree. However, our results do support previous knowledge of heat tolerance. The 7% of the soybean genome which is syntenic with the cowpea QTLs we report contain several interesting candidate genes including proline transporters, heat shock transcription factors, and heat shock proteins. Interestingly, all three loci annotated as proline transporters in soybean are in regions syntenic with *Cht – 5*. These findings provide some support for the hypothesis and findings of Ahmed et al. (1992) and Mutters et al. (1989b) who also associated cowpea reproductive stage heat tolerance to proline translocation. Because IT82E-18 is the source of favorable haplotypes comprising *Cht – 5*, this parent may carry alleles encoding

proline transporters which act to enhance cowpea's tolerance to heat when compared to CB27. Alternatively, CB27 is the donor of the favorable alleles for the majority of the QTLs which are syntenic with regions of the soybean genome abundant in heat shock proteins, heat shock transcription factors, and proline rich proteins.

Future researchers could aim at developing markers to tag the dominant gene identified by Marfo et al. (1992). Allelic variation at this locus and the regions we report should be surveyed among broader pedigrees. While it has been previously observed that the number of pods per peduncle is correlated with yield under hot environments (Nielsen et al., 1985), the relationship of these QTLs with yield should be determined. This may be accomplished by determining the yield of lines carrying different combinations of QTLs. The environmental aspect of QTL performance is also important. Are the QTLs we report environment or pedigree specific such that they can only be used to predict performance in a certain pedigree, environment, or pedigree-by-environment scenario? Future experiments could also develop markers for the inheritance of leaf electrolyte-leakage which has been used to estimate tolerance to heat in cowpea (Thiaw et al., 1998). It may be useful to understand whether or not the genes controlling this trait map to regions similar to the ones identified in the present manuscript.

Acknowledgements

This work was supported by the Generation Challenge Program (GCP) through a grant from the Bill and Melinda Gates Foundation and U.S. Agency for International Development Collaborative Research Support Program GDG-G-00-02-00012-00 and EDH-A-00-07-00005. We acknowledge the collaborative effort led by members of the

GCP, the Centro International de Mejoramiento de Maiz y Trigo within the Integrated Breeding Platform, and members of the University of California Riverside cowpea team for the development of the cowpea Kbiosciences genotyping platform and for the implementation of Genotyping Support Services which can be accessed by following the link: http://www.generationcp.org/gss_homepage.

References

Ahmed FE, Hall AE, and DeMason DA (1992) Heat injury during floral development in cowpea (Vigna unguiculata, Fabaceae). American Journal of Botany 79: 784 - 791.

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, and Cook DR (2004) Estimating genome conservation between crop and model legume species. Proceedings of the National Academy of Science 101: 15289 - 15294.

Dow El-Medina IM, and Hall AE (1986) Flowering of contrasting cowpea (*Vigna unguiculata* [L.] Walp.) genotypes under different temperatures and photoperiods. Field Crop Research 14: 87 - 104.

Ehlers JD, Hall AE, Patel PN, Roberts PA, and Matthews WC (2000) Registration of 'California Blackeye 27' cowpea. Crop Science 40: 854 - 855.

Frova C and Sari-Gorla M (1994) Quantitative trait loci (QTLs) for pollen thermotolerance detected in maize. Molecular Genetics and Genomics 245: 424 - 430.

Hall AE (2012) Heat Stress. In Shabala S (Ed), Plant Stress Physiology 118 - 131. Cambridge, MA: CABI.

Hall AE (1992) Breeding for Heat Tolerance. Plant Breeding Reviews 10: 129 - 168.

Hays D, Mason E, Hwa Do J, Menz M, and Reynolds M (2007) Expression quantitative trait loci mapping heat tolerance during reproductive development in wheat

(Triticum Aestivum). In Buck HT, Nisi JE, and Salomon N (Eds), Developments in Plant Breeding Vol 12, Wheat Production in Stressed Environments 12: 373 - 382.

Jansen RC and Stam P (1994) High Resolution of Quantitative Traits Into Multiple Loci via Interval Mapping. Genetics 136: 1447 - 1455.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, and Huala E (2011) The Arabidopsis information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research 40: 1202 - 1210.

Lane A and Jarvis A (2007) Changes in climate will modify the geography of crop suitability: agricultural biodiversity can help with adaptation. Journal of Semi-Arid Tropical Agricultural Research 4: 1 - 12.

Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, and Close TJ (2011) Cowpea-soybean synteny clarified through an improved genetic map. The Plant Genome 4: 218 - 225.

Marfo KO and Hall AE (1992) Inheritance of heat tolerance during pod set in cowpea. Crop Science 32: 912 - 918.

Medicago truncatula HapMap Project (2010) Mt3.5 genome assembly annotation November 2010. Available at http://medicagohamap.org/. Nevin Young, University of Minnesota, St. Paul, MN.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, and 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.

Mutters RG, Ferreira LGR, and Hall AE (1989a) Proline content of the anthers and pollen of heat-tolerant and heat-sensitive cowpea subjected to different temperatures. Crop Science 29: 1497 - 1500.

Mutters RG, Hall AE, and Patel PN (1989b) Photoperiod and light quality effects on cowpea floral development at high temperatures. Crop Science 29: 1501 - 1505.

Nielsen CL and Hall AE (1985) Responses of cowpea (Vigna unguiculata (L.) Walp.) in the field to high night air temperature during flowering. II. Plant Responses. Field Crops Research 10: 181 - 196.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, GRimwood J, Rokhsar D, Stacey G, Shoemaker RC, and Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178 - 183.

Thiaw S and Hall AE (2004) Comparison of selection for either leaf electrolyte-leakage or pod set in enhancing heat tolerance and grain yield of cowpea. Field Crops Research 86: 239 - 253.

Van Ooijen JW (2004) MapQTL5. Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, Netherlands.

Wahid A, Gelani S, Ashraf M, and Foolad MR (2007) Heat tolerance in plants: An overview. Environmental Experimental Botany 61: 199 - 223.

Wanamaker S and Close TJ (2011) HarvEST. HarvEST:Cowpea version 1.27. Available at http://harvest.ucr.edu; University of California Riverside.

Warrag MOA, and Hall AE (1984) Reproductive responses of cowpea (Vigna unguiculata [L.] Walp.) to heat stress. II. Responses to night air temperature. Field Crop Research 8: 17 - 33.

Ye C, Argayoso MA, Redona ED, Sierra SN, Laza MA, Dilla CJ, Mo Y, Thomson MJ, Chin J, Delavina CB, Diaz GQ, and Hernandez JE (2012) Mapping QTL for heat tolerance at flowering stage in rice using SNP markers. Plant Breeding 131: 33 - 41.

Table 8: Cowpea heat tolerance at pod set QTLs, map location, percent of heritable phenotypic variance explained, LOD score, parent donating tolerance, and discovery trial.

QTL Name	Linkage Group (cM)	Percent Phenotype Explained	LOD Score	Resistance Donor	Discovery Experiment
Cht - 1	2 (92.99 - 124.22)	18.1	5.11	CB27	2010 + 2011
Cht - 2	7 (63.16 - 66.65)	17.1	5.75	CB27	2010 + 2011
Cht - 3	6 (13.45 - 15.97)	16.2	5.39	CB27	2010 + 2011
Cht - 4	10 (11.38 - 30.02)	16	4.49	CB27	2010 + 2011
Cht - 5	3 (33.61 - 39.15)	11.5	3.73	IT82E-18	2010 + 2011

Table 9: Regions of the soybean genome displaying synteny with cowpea heat tolerance QTL. Annotations of candidate genes within the syntenic regions of soybean based on homology with *A. thaliana* of genes. Parentheses in the annotation column describe the number of times a given annotation was present. DNA J HSP (DNA J Heat Shock family Protein), LEA HPR (Late Embryogenesis Abundant HyrdoxyProline-Rich glycoprotein family), HPR (HydroxyProline-Rich glycoprotein family), HSP (Heat Shock Protein family), HSTF (Heat Shock Transcription Factor), PT (Proline Transporter).

QTL Name	Homeologous Soybean Chromosome(s) (Region)	Candidate Gene Annotations
Cht - 1	2 (0.13 - 1.54 Mb) 10 (1.17 – 1.61 Mb; 36.88 – 40.23 Mb) 20 (23.23 – 33.01 Mb; 43.51 – 45.81 Mb)	DNA J HSP (4), HSP (12), HPR (1)
Cht - 2	3 (2.99 Mb)	-
Cht - 3	3 (45.79 – 46.47 Mb) 15 (7.42 – 11.45 Mb) 19 (48.42 – 50.05 Mb)	DNA J HSP (3), HSP (4), HPR (4)
Cht - 4	1 (32.80 – 47.09) 3 (3.43 – 31.80 Mb) 7 (5.03 – 6.27Mb; 12.81 – 13.70 Mb) 16 (2.90 – 3.29 Mb; 7.97 – 9.59 Mb)	DNA J (6), HSP (9), HSTF (1), HPR (2), LEA HPR (2)
Cht - 5	5 (0.22 – 2.13 Mb; 37.16 – 38.05 Mb) 8 (0.14 – 0.68 Mb) 17 (6.38 – 10.30 Mb)	HSP (1), HPR (1), PT (3)

Figure 15: Phenotypic distribution of the number of pods per peduncle for 2010 (dark) and 2011 (light) greenhouse experiments using the CB27 x IT82E-18 discovery population of cowpea.

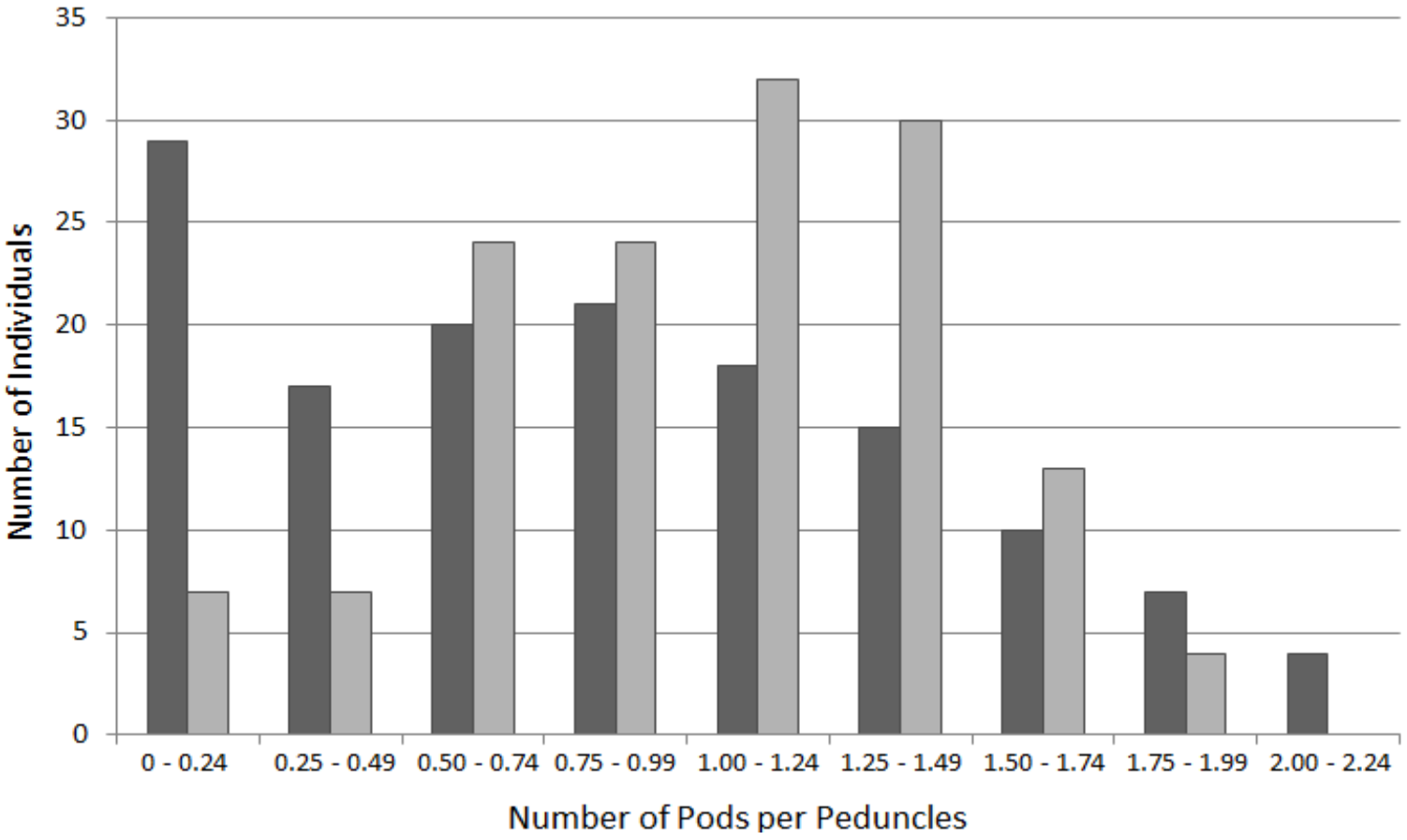
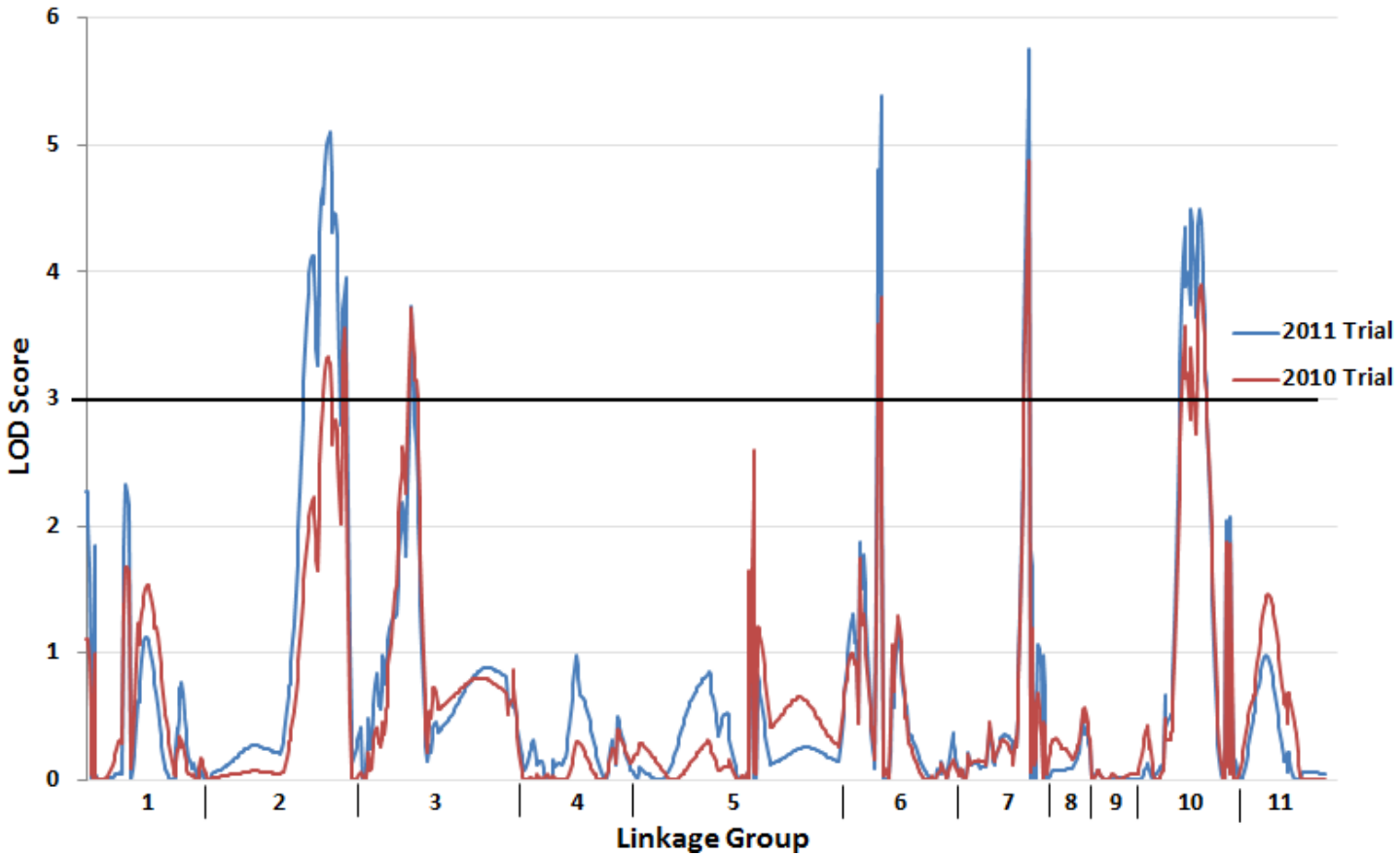


Figure 16: Stage II heat tolerance (**A**) and susceptibility (**B**) phenotypes in cowpea.

Figure 17: Cowpea stage II heat tolerance QTL LOD score traces among eleven linkage groups of the cowpea genome. The 2010 and 2011 discovery trials are colored red and blue, respectively.



Appendix C

Markers for Quantitative Inheritance of Resistance to Foliar Thrips in Cowpea

Abstract

Molecular assisted breeding is currently constrained by the lack of breeder friendly trait-associated markers, especially among lesser studied crops. Recent advances in genomic technology are being applied to many important crop species, promoting the development of robust marker-trait associations. Regions of the cowpea (*Vigna unguiculata*) genome controlling the quantitative resistance to feeding by foliar thrips (*Thrips tabaci* and *Frankliniella schultzei*) were tagged by coupling phenotypic observations from two recombinant inbred line populations generated from domesticated parents. Three regions (*Thr - 1*, *Thr - 2*, and *Thr - 3*) explaining approximately 32%, 24%, and 9%, respectively, of the phenotypic variation were tagged with several EST-derived SNP markers and are presented here as haplotypes, composed of multiple SNP markers. Regions within the soybean genome which are syntenic to cowpea *Thr - 1*, *Thr - 2*, and *Thr - 3* are also reported.

Introduction

Feeding by thrips (Thysanoptera: Thripidae) can cause severe damage in many plant species. This insect pest is known to vector pathogenic viruses (Todd et al., 1996) and reduce yield (Rummel et al., 1979). Genetic resistance to multiple insecticides by foliar thrips (*Thrips tabaci*) has been reported (Allen et al., 2005) warranting the development of innately resistant cultivars. The application of molecular markers to breeding is an approach to develop varieties with favorable alleles for several traits, even those which are inherited quantitatively (Tanksley, 1993).

In cowpea (*Vigna unguiculata*), an important legume crop, genetic maps have been periodically improved as new marker technologies have been implemented (Menendez et al., 1997; Ouedraego et al., 2002; Muchero et al., 2009; Lucas et al., 2011). The impact of thrips damage, especially flower thrips (*Megalurothrips sjostedti*), represents a substantial obstacle for growers and breeders of cowpea (Ehlers et al., 1997). Initiatives to breed thrips resistant varieties would benefit from markers associated with resistance loci, which can then be combined with other markers in selection schemes to develop ideal varieties (ideotypes) (see Xu et al., 2008; for a review and interpretation of marker-assisted breeding).

The quantitative inheritance of resistance to foliar thrips was recently associated with AFLP (Amplified Fragment Length Polymorphism) markers in cowpea (Muchero et al., 2010). This work identified three regions of the cowpea genome explaining some of the heritable resistance in seven separate field trials across two environments. Recent developments in cowpea genotyping and genetic mapping (Muchero et al., 2009; Lucas et

al., 2011) provide a framework for discovering SNP-trait associations. Here we describe the association of EST (Expressed Sequence Tag) - derived SNP markers with resistance to feeding from foliar thrips, accomplished by re-analyzing phenotypic data from Muchero et al. (2010) and combining it with new data collected from two trials involving a different discovery population. Regions of the soybean genome displaying synteny to cowpea thrips resistance associated loci are also reported.

Materials and Methods

Discovery Populations and Trials: Phenotypic observations of foliar thrips damage (0 – 10, resistant-susceptible) were collected during nine separate trials of two eighth filial generation (F_8) recombinant inbred line (RIL) discovery populations CB46 x IT93K-503-1 and CB27 x IT82E-18. Scores reflected visual observations of the severity of leaf curling from feeding damage. Observations from CB46 x IT93K-503-1 were previously used (Muchero et al., 2010) to associate AFLP markers with foliar thrips resistance and these scores were used again in this work. Resistance scores for the CB27 x IT82E-18 population were collected during two separate trials of 160 individuals using a completely randomized block design with four replications. In a greenhouse during the spring of 2011, at the Citrus Experiment Station at the University of California Riverside, observations were collected one week prior to maturity. The second trial of the CB27 x IT82E-18 population also occurred at the University of California but was conducted as a summer field trial with a plot length of 5.5 meters and width of 0.9 meters where 100 seed were planted per plot. Phenotype scores for the field trial were also gathered one week prior to maturity. Please refer to Muchero et al. (2010) for images of resistant and susceptible phenotypes. All infestations were naturally occurring and evenly distributed among the greenhouse and field trials which required no artificial introductions.

The CB27 x IT82E-18 and CB46 x IT93K-503-1 discovery populations were filtered for quality by omitting data from known rogue individuals (Lucas et al., 2011). Data from six rogue individuals from the CB27 x IT82E-18 population, and from sixteen rogue individuals from the CB46 x IT93K-503-1 population, were removed yielding

population sizes of 160 and 114, respectively. These rogues which were initially defined in Lucas et al. (2011) are inappropriate for data analysis and have characteristics of excessive heterozygosity, non-parental genotypes, missing data, or are genetically identical to another member of the population.

Trait-Associated Marker Discovery: Phenotypic data reported in Muchero et al. (2010) (i.e. seven experiments with CB46 x 503-1 population) were analyzed with the Illumina 1536 SNP (Illumina, 2010) platform and consensus map of Lucas et al. (2011) using the interval mapping functions provided by MapQTL5 (Van Ooijen, 2004). Parental linkage phase was determined using the approach described in Lucas et al. (2011) which relied on utilities of JoinMap4 (Van Ooijen, 2006). Top strand (Illumina GenomeStudio data export option) genotype calls (ATGC) for SNPs linked to QTLs were used to report resistance-associated cowpea haplotypes. The QTL naming conventions established in Muchero et al. (2010) (*Thr - 1, Thr - 2,* and *Thr - 3*) are preserved. Genotype calls within *Thr - 1, Thr - 2,* and *Thr - 3* for all four parents were compared and used to determine which trait-associated markers were most appropriate for selection. Additionally, only trait-associated markers which were polymorphic among the parents of each discovery population were reported.

Synteny with Soybean: Positions within the soybean genome harboring cowpea mapped SNPs were extracted from HarvEST:Cowpea 1.27 (Wanamaker et al., 2011). Soybean translated gene models and genomic sequences were retrieved from the JGI Glyma1 database (Schmutz et al., 2010). Cowpea unigenes (Muchero et al., 2009) and consensus map coordinates (Lucas et al., 2011) were used to search the soybean genome.

When queried with mapped cowpea unigenes, only soybean loci yielding BLASTX scores less than 10^{-10} were used to report synteny. Visualization software Circos (Krzywinski et al., 2009) was used to develop Figure 19 and Supplemental Files 15 and 16 which depict homeologous loci based on BLASTX scores ($<10^{-10}$) between cowpea unigenes containing a mapped SNP and translated gene models from reference genomes. HarvEST:Cowpea (Wanamaker et al., 2011) provides an interactive interface constructed to explore synteny and export images (tiff format).

Results

Trait Segregation among Discovery Populations: CB27 received phenotypic scores of zero for all replicates in both field and greenhouse trials conducted using the CB27 x IT82E-18 population while IT82E-18 received an average score of 6.00 (SE = 0.51) for the field trial and a score of 6.06 (SE = 0.49) for the greenhouse trial. Frequency distribution histograms of phenotypic scores for the two trials using the CB27 x IT82E-18 population are presented in Supplemental File 12 and conveys positive skew (high frequency of resistance) and transgressive segregation towards susceptibility when considering the susceptible parent (IT82E-18) phenotypic averages.

Marker-Trait Associations: LOD traces for QTL discovery experiments are presented for *Thr - 1*, *Thr - 2*, and *Thr - 3* in Figure 18 and Supplemental Files 13 and 14, respectively. Table 10 characterizes two major (*Thr - 1*, *Thr - 2)* and one minor QTL (*Thr - 3*) explaining approximately 32%, 22%, and 9%, respectively, of the phenotypic variation. These QTL were identified among three linkage groups reported in Lucas et al. (2011). All three regions were identified from the analysis of multiple trials. *Thr - 1* was identified from all nine experimental data sets representing both discovery populations (CB46 x IT93K-503-1 and CB27 x IT82E-18) while *Thr - 2* and *Thr - 3* were identified in the analysis of 5 and 2 trials of differing populations (CB46 x IT93K-503-1 and CB27 x IT82E-18, respectively). All three QTLs surpassed significance thresholds calculated using the permutation test provided in MapQTL5 (Van Ooijen, 2004). Genotype calls (ATGC) for favorable and alternative haplotypes associated with resistance within *Thr - 1*, *Thr - 2,* and *Thr - 3* are presented in Table 11. Favorable haplotypes spanning *Thr - 1*

were contributed by CB46 and CB27 genotypes and were identical with respect to current cowpea genotyping resolution.

Thr - 1 has been tagged with 14 SNP markers (12 unique bins) spanning 10.86 cM on cowpea consensus linkage group 2. *Thr - 2* and *Thr - 3* have been tagged with 4 (3 unique bins) and 7 (6 unique bins) markers spanning 2.34 cM and 3.40 cM of cowpea consensus linkage groups 4 and 10, respectively. These markers are presently available on two SNP genotyping platforms (Illumina's 1536 GoldenGate Assay and Kbioscience's KASPAR, described in Robinson et al., 2010) and those which are polymorphic are reported in the Illumina top strand format.

Synteny with Soybean: Cowpea unigenes containing a mapped SNP within *Thr - 1*, *Thr - 2*, and *Thr - 3* found several significant BLASTX hits against the soybean genome. Figure 19 displays corresponding locations of homeologs between cowpea *Thr - 2* (a segment of cowpea consensus linkage group 4) and soybean chromosomes 3 and 19. Eight SNP markers (6 bins) within *Thr - 2* found hits to soybean chromosomes 3 (37.0 Mb – 39.0Mb) or 19 (39.5 Mb – 41.5 Mb). Regions of the soybean genome displaying synteny with cowpea *Thr - 1* are depicted in Supplemental File 15 which include chromosomes 2 (7 Mb – 18 Mb), two regions of 10 (2.5 Mb – 6.8 Mb and 30 Mb – 34 Mb), and 20 (23.4 Mb – 30 Mb). For cowpea *Thr - 3*, soybean chromosomes 1 (35.1 Mb – 44.3 Mb) and 3 (5.1 Mb – 8.6 Mb) are most similar and are presented in Supplemental File 16. Interactive synteny displays for the whole genome including the three regions presented in this manuscript can be accessed via HarvEST:Cowpea 1.27 (Wanamaker et

al., 2011) and provide additional information including gene annotations and E-values for several related species.

Discussion

Trait-associated AFLP markers previously developed for resistance to foliar thrips were used to tag three QTLs (Muchero et al., 2010), however, superior cowpea genetic markers (SNPs) are now available (Muchero et al., 2009; Lucas et al., 2011). Here we attempted to improve the understanding of quantitative inheritance of thrips resistance in cowpea and deliver useful markers to plant breeders.

The inheritance of thrips resistance behaved similarly among the two discovery populations used in this work. One parent from each population (CB46 and CB27) was completely resistant while the other parents (IT82E-18 and IT93K-503-1) were partially susceptible with similar phenotypic averages of approximately 6.00. The identification and co-localization of *Thr - 1* among the two discovery populations may be expected due to the similar pedigree of CB46 and CB27. In this study we failed to identify region(s) of the cowpea genome contributing to the inheritance of partial resistance in IT82E-18. However we do report a favorable haplotype (*Thr - 2*) from the partially susceptible parent IT93K-503-1. *Thr - 3* was contributed by CB27 which may have inherited this region from parents unique to its pedigree including CB3, Prima, TVu4552, and/or breeding line 7977 (Helms et al., 1991; Ehlers et al., 2000). Unfortunately, genotype data has not been collected for all of the discovery population grandparents, making it impossible to pinpoint an earlier source of these QTLs.

The diversity within cowpea germplasm cannot be comprehensively studied using only these two bi-parental populations, especially since two of the parents (CB46 and CB27) have a similar pedigree. The potential for application of *Thr - 1*, *Thr - 2*, and *Thr -*

3 outside the discovery population pedigrees could be partially assessed by phenotyping and genotyping a diversity panel for SNPs comprising the three QTLs we describe. If individuals maintaining the favorable haplotypes we describe display resistance and those with a different haplotype are more susceptible, then some level of confidence could be assigned to using these markers for breeding cowpeas more distantly related to the parents of the discovery populations.

Thr - 1 and *Thr - 2* were previously thought to be linked to the same cowpea linkage group (Muchero et al., 2010), however, new insight provided by the SNP consensus genetic map (Lucas et al., 2011) proposes independent segregation of three resistance-associated regions. Discrepancies between the two cowpea genotyping models and maps (AFLP and SNP) either identified two distinct regions of the genome or identified the same regions but placed them in different locations. Only speculations concerning the conservation of these regions can be made until a bridge between the cowpea AFLP and SNP maps has been reported.

The lack of breeder-friendly trait associated molecular markers is a limitation for many current marker-assisted breeding initiatives. While several marker-trait associations have been described among crop plants, many remain anchored to the technology in which they were discovered. Breeders wishing to employ marker-assisted selection should be provided with the simplest, most useful, and most cost-effective genotyping services. SNP genotyping is a promising for delivering these characteristics. Recent efforts in cowpea have developed “bead-assay” genotyping which has been essential in genetic mapping, discovering marker-trait associations, and comparative genomics.

Cooperative efforts led by the Generation Challenge Program (GCP), the Centro International de Mejoramiento de Maiz y Trigo (CIMMYT), and members of the University of California Riverside cowpea team have translated SNP probes from the fixed array Illumina GoldenGate Assay to a flexible genotyping platform (Kbiosciences KASPAR) that allows for user definition of which and how many loci will be genotyped. These genotyping support services, which include all aspects of DNA extraction and genotyping, are also established and essential in applying technological advances across disciplines (see http://www.generationcp.org/gss_homepage for more information). DNA extraction and genotyping services have several advantages and potentially remove the burden of wet lab activities from the breeder's agenda.

SNP assays provide a method for directly querying a single nucleotide in the genome. In contrast to AFLP technology which reports information about restriction fragment sizes, SNPs, report A, T, G, or C genotype calls. This permits a more definitive description of DNA sequence haplotypes. Understanding genotype variation at the resolution of a nucleotide also can potentially provide insight into the functional roles and context of the variation underlying a molecular marker. Mapped loci annotated with sequence information are also useful for characterizing synteny, especially when comparisons are made with species lacking information for common DNA markers, such as COS markers (Fulton et al., 2002). Perhaps most important to a plant breeder is the convenience of SNP assays when genotyping a large number of markers and/or a large number of individuals. Additionally, information used to design one SNP genotyping platform can potentially be translated to develop SNP genotyping on a different platform.

This allows researchers dependent on SNP genotyping the flexibility to choose the most appropriate service providers.

Synteny provides the ability to translate knowledge among closely related species. Studies conveying marker-trait associations in one crop therefore have potential applications in other crops. In this work we identify specific regions of the soybean genome with similarity to thrips resistance associated regions of the cowpea genome. Cowpea *Thr - 2* localizes to two approximately 2 Mb regions within the soybean genome. This equates to less than 0.5% of the total genome size of soybean, 1.1 Gb – 1.15 Gb (Arumuganathan et al., 1991). Preliminary examination of these regions among soybean genetic maps using SoyBase (Grant et al., 2010), where marker-trait associations exist, yielded no promising leads, such as the co-localization of other insect resistance QTLs. However, a friendlier environment for translating knowledge between soybean and cowpea would help to smooth this transition for other important traits which may be governed by homologous loci. A tool constructed specifically for the purpose of translating genomic information among the legumes would ease future comparative approaches.

Future work should determine the efficacy of these markers in developing resistant varieties. Additionally, attempts to reconcile the quantitative inheritance of resistance to both foliar and flower thrips should be explored. The discovery of a conserved mechanism of resistance or the discovery of tightly linked loci governing the resistance to both pests would make selection strategies aiming to pyramid traits for resistance to both foliar and flower thrips simpler. Approaches to develop trait-associated

markers for the breeding of varieties possessing tolerance to other continuously inherited traits (yield, heat, drought, or flowering time) could progress using the methods employed in this work. Finally, we advocate the use of genotypic data to filter out rogue individuals from experimental populations.

Acknowledgements

We would like to acknowledge the collaborative effort led by members of the Generation Challenge Programme (GCP) and the Centro International de Mejoramiento de Maiz y Trigo (CIMMYT) within the Integrated Breeding Platform for the development of the Kbiosciences genotyping platform and for the implementation of Genotyping Support Services (GSS) which can be accessed by following the link: http://www.generationcp.org/gss_homepage

References

Allen JKM, Scott-Dupree CD, Tolman JH, and Harris CR (2005) Resistance of Thrips tabaci to pyrethroid and organophosphorous insecticides in Ontario, Canada. Pest Management 61: 809 - 815.

Arumuganathan K, and Earle ED (1991) Estimation of nuclear DNA content of plants by flow cytometry. Plant Molecular Biology Reports 9: 229 - 241.

Ehlers JD and Hall AE (1997) Cowpea (Vigna unguiculata L. Walp.). Field Crops Research 53: 187 - 204.

Ehlers JD, Hall AE, Patel PN, Roberts PA, and Matthews WC (2000) Registration of California Blackeye 27 cowpea. Crop Science 40: 854 - 855.

Fulton TM, Van der Hoeven R, Eannetta NT, and Tanskley SD (2002) Identification, analysis, and utilization of conserved orthologs set markers for comparative genomics in higher plants. The Plant Cell 14: 1457 - 1467.

Grant D, Nelson RT, Cannon SB, and Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research 38: 843 - 846.

Helms D, Panella L, Buddenhagen IW, Tucker CL, and Gepts PL (1991) Registration of California Blackeye 46 cowpea. Crop Science 31: 1703.

Illumina (2010) GenomeStudio genotyping module v.2010.3.Available at http://www.illumina.com/software/genomestudio_software.ilmn. Illumina, Inc. San Diego, Ca.

Krzywinski M, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, and Marra MA (2009) Circos: An information aesthetic for comparative genomics. Genome Research 19: 1639 - 1645.

Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, and Close TJ (2011) Cowpea-soybean synteny clarified through an improved genetic map. The Plant Genome 4: 218 - 215.

Menendez CM, Hall AE, and Gepts P (1997) A genetic linkage map of cowpea (Vigna unguiculata) developed from a cross between two inbred, domesticated lines. Theoretical and Applied Genetics 95: 1210 - 1217.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, and 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.

Muchero W, Ehlers JD, and Roberts PA (2010) QTL analysis for resistance to foliar damage caused by Thrips tabaci and Frankliniella schultzei (Thysanoptera : Thripidae) feeding in cowpea [Vigna unguiculata (L.) Walp.]. Molecular Breeding 25: 47 - 56.

Ouedraego JT, Gowda BS, Jean M, Close TJ, Ehlers JD, Hall AE, Gillaspie AG, Roberts PA, Ismail AM, Bruening G, Gepts P, Timko MP, and Belzile FJ (2002) An improved genetic linkage map for cowpea (Vigna unguiculata L.) Combining AFLP, RFLP, RAPD, biochemical markers, and biological resistance traits. Genome 45: 175 - 188.

Robinson P and Holme J (2011) The KASP SNP genotyping manual v3.0. Available at http://www.kbioscience.co.uk/reagents/KASP_manual.pdf. Kbiosciences Ltd, LGC Genomics. Hoddesdon, Herts, England.

Rummel DR and Quisenberry JE (1979) Influence of Thrips Injury on Leaf Development and Yield of Various Cotton Genotypes. Journal of Economic Entomology 72: 706 - 709.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht, J, Grimwood, J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178 - 183.

Todd JW, Culbreath AK, and Brown SL (1996) Dynamics of Vecto Populations and Progress of Spotted Wilt Disease Relative to Insecticide Use in Peanuts. Acta Horticulture 431: 483 - 490.

Van Ooijen JW (2004) MapQTL5. Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, Netherlands.

Van Ooijen JW (2006) JoinMap4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.

Wanamaker S and Close TJ (2011) HarvEST. HarvEST:Cowpea version 1.27. Available at http://harvest.ucr.edu; University of California Riverside.

Xu Y and Crouch JH (2008) Marker-Assisted Selection in Plant Breeding: From Publications to Practice. Crop Science 48: 391 - 407.

Table 10: Three foliar thrips resistance associated QTL (*Thr - 1*, *Thr - 2*, and *Thr - 3*), the populations in which they were discovered, parents donating the resistance at these regions, the map positions of the QTL including linkage group (LG) and centi-morgan (cM), range of peak LOD scores, the percent of the phenotypic variation accounted for by the QTL, and the number of experiments in which the QTL was detected among the discovery experiments.

QTL	Discovery Population(s)	Resistance Donor Parent	LG	cM	Peak LOD (Average)	Percent of Phenotype Explained (Average)	Number of Discovery Experiments
Thr - 1	CB46 x IT93K-503-1	CB46	2	17.00 – 29.00	4.16 – 8.29 (5.65)	22.7 – 40.5 (31.5)	7
	CB27 x IT82E-18	CB27	2	13.00 – 25.00	4.64 – 12.69 (8.67)	14.1 – 33.5 (23.8)	2
Thr - 2	CB46 x IT93K-503-1	IT93K-503-1	4	16.50 – 20.70	2.57 – 6.82 (5.16)	19 – 31.1 (22.1)	5
Thr - 3	CB27 x IT82E-18	CB27	10	43.00 – 45.70	2.26 – 3.37 (2.81)	8.6 -9.6 (9.1)	2

Table 11: SNP markers, map positions including linkage group (LG) and centi-morgan (cM), favorable and alternative genotype calls for the cowpea haplotype spanning the foliar thrips resistance associated QTLs *Thr - 1*, *Thr - 2*, and *Thr - 3*.

QTL	Harvest Unigene Position	1536-plex	Map Position LG(cM)	Favorable Genotype	Alternative Genotype
Thr - 1	14508_128	1_0277	2(18.92)	AA	GG
	8190_327	1_0589	2(18.92)	AA	GG
	7857_1368	1_0698	2(18.92)	AA	GG
	2046_754	1_0829	2(19.45)	GG	AA
	458_1330	1_0492	2(19.53)	GG	AA
	5026_672	1_0253	2(20.16)	AA	TT
	16914_262	1_0337	2(20.35)	GG	AA
	3044_1453	1_0164	2(21.00)	AA	GG
	8395_1157	1_1086	2(22.96)	AA	TT
	411_247	1_0284	2(25.15)	AA	GG
	12996_239	1_1406	2(26.12)	AA	CC
	4402_623	1_1139	2(26.86)	AA	GG
	6561_1160	1_1061	2(29.22)	CC	AA
	11598_527	1_1048	2(29.78)	AA	TT
Thr - 2	1078_282	1_1413	4(18.38)	CC	AA
	16646_118	1_0774	4(20.16)	GG	AA
	1202_1215	1_1221	4(20.16)	GG	AA
	4217_685	1_1242	4(20.72)	GG	AA
Thr - 3	2597_339	1_0840	10(43.86)	GG	CC
	10780_756	1_0754	10(45.04)	AA	GG
	12439_253	1_0281	10(46.25)	AA	GG
	11054_889	1_1453	10(46.55)	GG	AA
	15786_379	1_0354	10(46.7)	AA	GG
	12584_1346	1_0952	10(47.26)	AA	GG
	8889_547	1_1062	10(47.26)	CC	AA

Figure 18: *Thr - 1* LOD score traces among cowpea linkage group two for nine discovery experiments. Traces contributed by discovery population CB46 x IT93K-503-1 are colored blue while CB27 x IT82E-18 are colored red.

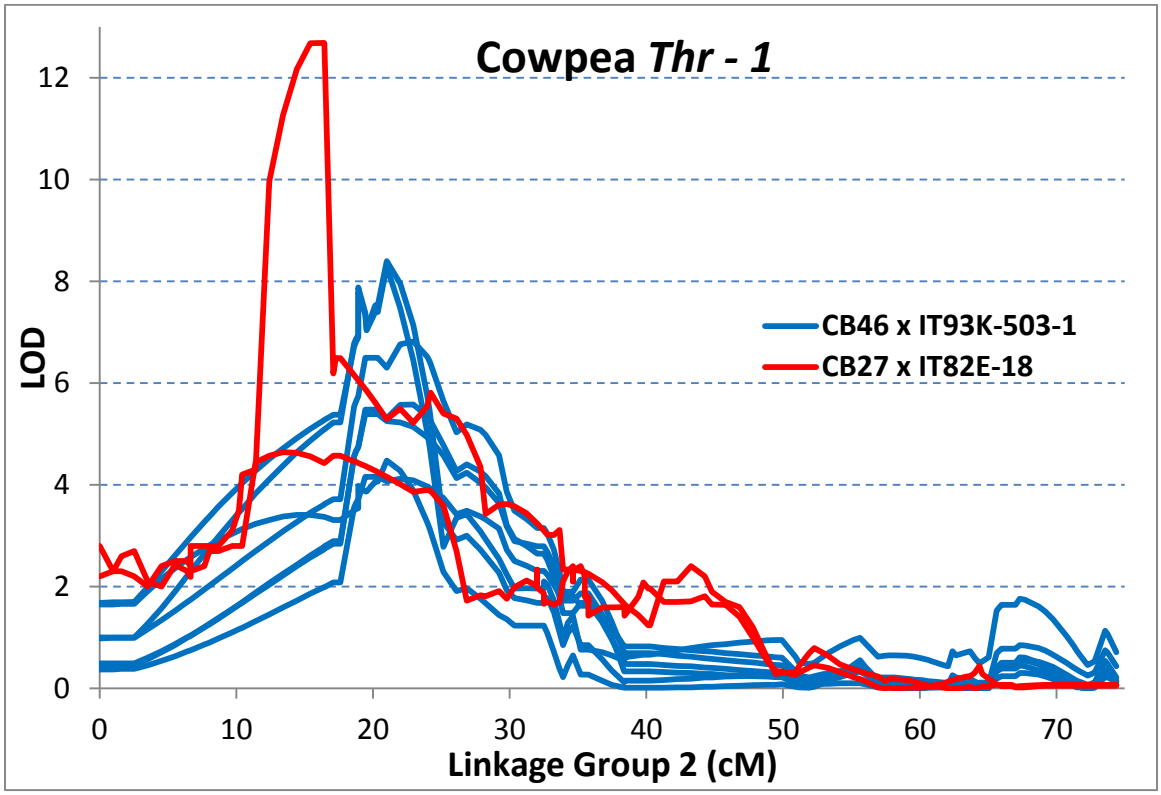


Figure 19: Consensus genetic map position of cowpea foliar thrips resistance region, *Thr - 2*, synteny with soybean chromosomes three and nineteen. Links connect the locations of cowpea SNP markers with the locations of soybean translated gene models.