Study of inheritance and identification of molecular markers for seed protein content in pigeonpea (*Cajanus cajan* (L.) Millsp.)

Jimmy Obala
MSc. Plant Breeding and Seed Systems
Makerere University

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Plant Breeding

School of Agricultural, Earth and Environmental Science
College of Agriculture and Engineering Science, University of KwaZulu-Natal, Pietermaritzburg, Republic of South Africa

August, 2017

Abstract

Pigeonpea is an important source of protein to the vegetarian and poor families around the globe, however, very little is known about the genetic control of seed protein content (SPC) and how it relates with other traits of agronomic importance in the crop. Availability of genomic resources such as a reference genome and whole genome resequencing data of germplasm lines in pigeonpea coupled with recent advances in next generation sequencing technologies provide opportunity to dissect the genetic architecture of SPC in the crop. The objectives of this study were to: (i) determine variation of SPC and its relationship with agronomic traits of importance in a set of breeding lines and landraces, (ii) study the inheritance of SPC and its relationship with seed weight and seed yield, (iii) identify quantitative trait loci (QTLs) conditioning SPC, and (iv) identify candidate genes involved in the accumulation of SPC using whole genome sequencing approach.

To determine variation in SPC and its relationship with some agronomic traits in pigeonpea, 23 pigeonpea genotypes were used. The genotypes are parents of different mapping populations presently being developed at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India. The 23 genotypes were evaluated under field conditions at ICRISAT in 2014-2015 growing season. The experiment was carried out in RCB design with two replications. Data were recorded on SPC, number of days to first flowering (DTF), plant height (PH) at maturity, number of pods per plant (NPP), number of seeds per pod (NSP), 100-seed weight (SW) and seed yield per plant (SY). Seed protein content ranged from 19.3 to 25.5%, DTF (48 to 156 days), PH (67.5 to 230 cm), NPP (31.7 to 582 pods), NSP (2.9 to 4.6 seeds/pod), SW (6.2 to 20.8 g) and SY (7.9 to 333.4 g). There were significant differences among genotypes for all traits. Broad-sense heritability was 0.693 for SPC and ranged from 0.517 to 0.999 among the agronomic traits. Genetic advance (GA) was 2.4 % for SPC but ranged from 1.2 % to 141. % among the agronomic traits. Genetic gain, which is GA expressed as a percentage of the trait's grand mean, was 11.0 % for SPC but ranged from 56.4 to 713.4 % among the agronomic traits. Simple correlation indicated that SPC is generally negatively associated with all measured traits but only significantly with SW. However, path coefficient analysis revealed that, in addition to SW, NPP also had a strong negative direct influence on SPC, whereas SY had strong positive direct effect on SPC. Indirect effects of the agronomic

traits on SPC were also noticeable with NPP and SW having strong negative and positive effects, respectively on SPC via SY.

To investigate inheritance pattern of SPC in pigeonpea, four elite germplasm lines of varying SPC were used to develop three crosses. Six generations (P_1, P_2, F_1, F_2, BC_1P_1 and BC_1P_2) were generated. Generation mean analysis (GMA) revealed the importance of dominance and epistatic effects for SPC. Duplicate and negative additive × additive epistasis were predominant. Transgressive segregation for SPC was conspicuous. Additive genetic variance component was higher than the environmental and dominance components. Broad-sense heritability ranged from 0.52 to 0.60. Predicted genetic gain after one cycle of selection was highest at 5% selection intensity. Seed weight and yield were positively and negatively correlated with SPC, respectively. The results suggests that careful selection of parents, and recurrent selection procedure targeting transgressive segregants should be effective for improving SPC in pigeonpea.

For the identification of QTLs associated with SPC and its relationship with some agronomic traits, five F_2 mapping populations segregating for SPC were developed, genotyped using genotyping-by-sequencing and phenotyped for SPC, 100-seed weight (SW), seed yield (SY), days to first flower (DTF) and growth habit (GH) under field conditions. The average inter-marker distance in the population-specific maps varied from 1.6 cM to 3.5 cM. On the basis of the population-specific and consensus linkage maps, a total of 196 main effect QTLs (M-QTLs) across all traits were detected that explained 0.7 to 91.3% of the phenotypic variation for the five traits across the five F_2 mapping populations. In the case of SPC as the core trait in the present study, a total of 48 main effect QTLs (M-QTLs) with phenotypic variance explained (PVE) ranging from 0.7 to 23.5% were detected across five populations of which 15 M-QTLs were major (PVE≥10). Twenty seven of the M-QTLs from the five F_2 mapping populations could be projected into six consensus M-QTL regions. Out of 573 epistatic QTLs (E-QTLs) detected with PVE ranging from 6.3 to 99.4% across traits and populations, 34 involved SPC with PVE ranging from 6.3 to 69.8%. Several co-localization of M-QTLs and E-QTLs affecting SPC and the agronomic traits were also detected and could explain the genetic basis of the significant ($P < 0.05$) correlations of SPC with SW ($r^2 = 0.22$ to 0.30), SY

(r^2 = -0.18 to -0.28), DTF (r^2 = -0.17 to -0.31) and GH (r^2 = 0.18 to 0.34). The quantitative nature of genetic control of SPC and its relationship with agronomic traits suggest that marker-assisted recurrent selection or genomic selection would be effective for the simultaneous improvement of SPC and other important traits.

To identify candidate variants and genes associated with SPC, whole genome resequencing (WGRS) data with an average of 12× coverage per genotype when compared to the Asha (ICPL 87119) reference genome was used. By combining a common variant (CV) filtering strategy with knowledge of gene functions in relation to SPC, 108 sequence variants whose presence lead to protein change were selected. The variants were found in 57 genes spread over all chromosomes except CcLG05. Identified genes were assigned to 19 categories based on gene ontology molecular function with fifty six percent of the identified genes belonging to only two functional categories. Sanger sequencing confirmed the presence of 52 (75.4%) sequence variants in 37 genes between low and high SPC genotypes. Fifty nine variants were converted into CAPS/dCAPS markers and assayed for polymorphism. Highest level of polymorphism was in low by high SPC parental pairs, while the lowest was in high by high parental pairs. Assay of 16 polymorphic CAPS/dCAPS markers on an F_2 segregating population of the cross ICP 5529 × ICP 11605 (high × low), resulted in 11 of the markers being incorporated into a GBS-derived SNPs genetic map. Single marker analysis (SMA) indicated four of the 16 CAPS/dCAPS markers to be significantly correlated with SPC. Three out of the four markers were positioned at <10.0 cM distance away from main effect SPC QTLs all on CcLG02. All the three markers found in close proximity to SPC QTL positions and those with significant association to SPC were derived from mutations in the same genes including NADH-GOGAT, copper transporter and BLISTER all on CcLG02. Results from this study provide a foundation for future basic research and marker-assisted breeding of pigeonpea for increased SPC.

In general, the complex nature of the genetic architecture of SPC as revealed by classical quantitative genetic analysis, QTL analysis and candidate gene analysis suggests that breeding approaches that target genome wide variations for crop improvement would be more appropriate in achieving larger genetic gains for SPC in shorter periods than using conventional phenotype-based selection.

Declaration

I, **Jimmy Obala**, declare that:

1. This thesis, except where otherwise indicated, is my original work.
2. It has not been submitted to any university for any degree or examination.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then their words have been paraphrased and summarised but the general information attributed to them referenced.
5. Finally the thesis presented herein does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and reference section

Sign _____ 18.08.2017

Jimmy Obala — Date

As the candidate's supervisors, we have approved the submission of this thesis:

Sign _____ 18|08|2017

Prof. Pangirayi Tongoona (Major supervisor) — Date

Sign _____ 21|08|2017

Dr. Rajeev K. Varshney (Co-supervisor) — Date

Sign _____ 31/08/2017

Dr. Julia Sibiya (Co-supervisor) — Date

Acknowledgement

Financial support for this study from the United States Agency for International Development (USAID) through ICRISAT is immensely appreciated. My sincere gratitude goes to Dr. Rajeev K. Varshney of ICRISAT for accepting me to work under this project.

I am greatly indebted to my supervisors: Prof. Pangirayi Tongoona, Dr. Rajeev K. Varshney and Dr. Julia Sibiya for their professional guidance and advice throughout this study. Their tireless effort and dedication are highly appreciated. I also owe special thanks to Dr. Rachit K. Saxena and Dr. Vikas K. Singh for the invaluable expert guidance provided from inception to completion of the study.

In a special way, I also greatly appreciate the efforts of Dr. Vinay Kumar, Mr Vinay Blue, Dr. Sandip Kale, Mr. Amir Khan, and Ms. Vanika Garg for the help in sequencing and processing of the genomic sequences used in this study. I do acknowledge the help rendered by Dr. K.B. Saxena and Dr. C.V. SameerKumar in recommending and providing some of the plant materials used in this work as well as for providing support during field trials.

Special thanks go to Ms. A. Chitekeni, Ms. B. Manjula and Mr B. Anjaiah for all excellent administrative work that made being part of 'Building No. 300' at ICRISAT easy and memorable. Thanks also go to K. Patel, S. Parupalli, S. Kaoneka, R.Y. Sashidar, D. Jaganathan, V. Garg, A. Muralli, F. Alo, A. Gafoor, Vinay Blue, M. NarsiReddy, P. Katiyar, V. Suryanarayana, and M. Brian as friends, colleagues and lab mates at the Applied Genomics Laboratory at ICRISAT.

I am deeply indebted to my employer, National Agricultural Research Organisation (NARO), Uganda, for granting me the leave of absence which made this work possible, and for accepting me to resume my duties after the study.

In a special way, I would also like to acknowledge the many constant words of encouragement from Dr. S.N. Said (formerly Director ICRISAT-ESA, Nairobi), Prof. P. Okori (Principal Scientist, ICRISAT Malawi) and Dr. R. Kajobe (Director of Research, NARO-Rwebitaba

ZARDI), which in the first place made me to take up the fellowship, and to a great extent, pushed me to complete the PhD study no matter what.

The patience and encouragement of my wife, Ann, and children, Elizabeth, Arthur and Vivianne, and the seemingly never-ending cry for help from nephew Ethan, and niece Adriel have been, in a unique way, of great motivation and inspiration towards completing this work, especially when I lacked these attributes.

To those whose names have not been mentioned rest assured that your contibutions have not been forgotten.

And, glory and praise to Lord God Almighty.

I thank you all.

Dedication

In memory of our sister Stella. May your soul forever rest in peace!

To Ann, Liz, Arthur, Viviane, Adriel, Ethan, David

And to my mother

Table of Contents

List of Tables

List of Figures

Introduction

Protein deficiency leads to developmental problems in children and their mothers but protein-rich plant foods may offer solutions particularly in areas of the world where intake of animal protein is low (Li et al., 2015). Among food plants, grain legumes are a major source of dietary protein in the developing world (Baudoin and Maquet, 1999; Iqbal et al., 2006). For sustained supply of dietary protein there is need not only to improve the agronomic practices but also to use crop cultivars which give reliable yields even under severe conditions (Foley et al., 2011). In the scenario mentioned above, pigeonpea (*Cajanus cajan* (L.) Millsp.) seems to be a promising crop as it is tolerant to heat and drought, and has the ability to give relatively better yields in marginal soils than any other food legume (Rao et al., 2010).

Pigeonpea is a sub-tropical and tropical grain legume that originated in the northern region of the Indian sub-continent, spreading to East Africa at least 4000 years BCE, and then to Southeast Asia, West Africa, Latin America, and the Caribbean (Khoury et al., 2015). It is an often cross pollinated species with 11 pairs of chromosomes (2n=2x=22) and a genome size of 833.07 Mbp (Varshney et al., 2012). It is the only cultivated food legume of the tribe Phaseoleae, sub-tribe Cajaninae, family Fabaceae (Leguminosae) and sub-family Papilionoideae (Greilhuber and Obermayer, 1988). Global area under pigeonpea cultivation continues to increase annually (Akibode and Maredia, 2011) standing at 5.6 million ha in the year 2013 with a production of ~4.0 million tons (FAOSTAT, 2015). Pigeonpea has diverse uses including being source of food, feed, fodder, building material and fuel wood, in addition to its contribution to biological nitrogen fixation (Rao et al., 2010). It is also a cash crop that supports the livelihoods of millions of resources-poor farmers in Asia and Africa (Mula and Saxena, 2010). As a source of food it provides dietary protein to more than a billion people globally.

Considering the importance of total seed protein content, hereafter referred to simply as seed protein content (SPC), in global food and nutritional security, there is need to produce more protein per unit area to meet the present and future dietary protein demands (Saxena and Sawargaonkar,

2015). However, breeding objectives in pigeonpea have for a long time, almost entirely focused on increasing yield and crop adaptability (Saxena, 2005; Odeny, 2007; Mligo and Craufurd, 2005; Upadhyaya et al., 2007). Very little or no attention has been given to the nutritional quality of the pigeonpea seed in terms of genetic enhancement, yet it has been reported that adequate genetic variability for SPC exists within the cultivated genepool that can be harnessed for trait improvement through breeding (Remanandan et al., 1988; Upadhyaya et al., 2007).

To improve SPC in pigeonpea through breeding requires a clear understanding of the genetic control of the trait. It is also essential to know the relationship of SPC with important agronomic traits such as seed yield and seed weight. Such information would allow designing cultivars with increased and stabilised SPC and acceptable agronomic characters (Burstin et al., 2007). There are, however, few documented studies on the genetic control of SPC (Dahiya et al., 1977; Vaghela et al., 2009; Baskaran and Muthiah, 2007) and its association with important agronomic characters in pigeonpea (Dahiya et al., 1977; Saxena et al., 1987; Rekha et al., 2013). However, the classical quantitative genetic approaches used in the reported studies are limited in power and resolution to dissect the genetic architecture of a quantitative trait like SPC and its relationship with important agronomic characters. Therefore, the available information is not only limited but also does not give a clear picture of the genetic architecture of SPC nor its relationship with important agronomic traits in the crop.

An earlier study using classical quantitative genetic approach reported that three to four genes condition SPC in pigeonpea (Dahiya et al., 1977). In other legumes, three to seven genes in cowpea (Santos et al., 2012), and one to 10 genes in common bean (Noubissié et al., 2012) have been reported. In soybean, quantitative trait locus mapping has revealed over 170 genomic loci to condition SPC (Soybase, 2016) while up to eight QTLs have been detected in garden pea (Burstin et al. 2007). These results indicate that SPC is a quantitative trait with complex molecular genetic mechanisms involving several biochemical pathways conditioning the trait (Burstin et al., 2007; 2011; Fauteux and Strömvik, 2009; Xu et al., 2012). It is therefore of interest to dissect the genetic factors that underlie SPC variation in pigeonpea.

The common approaches used for the understanding genetics of quantitative traits include classical quantitative genetic methods based on variance component analysis (VCA) and generation mean analysis (GMA) (Hallauer and Miranda, 1988) and genomics approaches (Frascaroli et al., 2007). Classical quantitative genetics facilitates estimation of heritability, trait correlations, and predicted responses to various selection schemes for practical plant breeding purposes (Holland and Cardinal, 2008). On the other hand genomics approaches provide means for in-depth analysis of the genetic architecture of quantitatively inherited traits through large-scale, high-throughput DNA sequencing, identification of quantitative trait loci (QTLs) and/or genes, gene expression analyses, and reverse genetics methods (Holland and Cardinal, 2008).

Large amount of genomic resources have become available in pigeonpea notably molecular markers, genetic maps, transcriptome assemblies, a draft genome sequence (Varshney et al., 2012) and whole genome resequencing (WGRS) data of several pigeonpea lines (Kumar et al., 2016; Varshney et al., 2017). These resources have facilitated construction of high-resolution genetic maps (Saxena et al., 2012) as well as rapid genetic analysis through molecular mapping of QTLs and genes controlling abiotic and biotic stresses and agronomic traits in pigeonpea (Saxena et al., 2011, 2012; Bohra et al., 2012; Kumawat et al., 2012; Mir et al., 2014; Sahu et al., 2015; Singh et al. 2016). However, the resources have not been applied for dissecting the genetic architecture of SPC in the crop.

The draft genome sequence, combined with the reduction in the sequencing cost and advances in sequencing technology opens up new avenues for employing massively paralleled (next-generation) sequencing (NGS) approaches for identification of genomic segments and candidate genes underlying traits of interest (Pazhamala et al., 2015). To this end, the application of sequence-based mapping (Varshney et al., 2014) has been successfully applied in pigeonpea to identify QTLs/genes for resistance to Fusarium wilt and sterility mosaic diseases (Singh et al., 2015). The availability of WGRS data in the presence of a reference genome sequence can facilitate the rapid detection of candidate genes through inexpensive bioinformatics and experimental assays and its potential has been demonstrated in both animals and plants (Sobreira et al., 2010; Silva et al., 2012; Gilissen et al., 2012; Xu et al., 2014). The whole genome sequence

based candidate gene identification is not limited by the need for prior knowledge of the function or position of a putative candidate gene as in the traditional candidate gene identification approach (Pflieger et al. 2001; Zhu and Zhao, 2007). The WGRS approach to candidate gene identification involves sequencing at greater depth two or a few individuals contrasting in the trait(s) of interest followed by a series of prioritization and validation steps to verify the candidacy of *in silico*-identified putative candidates and associated variants (Sobreira et al., 2010; Silva et al., 2012; Xu et al., 2014; Gilissen et al., 2012).

The use of genomics approaches in identification of quantitative trait loci (QTLs)/genes to understand the genetic control of SPC has been demonstrated in several crops including soybean (Lu et al., 2013; Hwang et al., 2014; Wang et al., 2014; Wang et al., 2015; Warrington et al., 2015; Soybase, 2016), chickpea (Jadhav et al., 2015; Upadhyaya et al., 2016), pea (Burstin et al., 2007; Krajewski et al., 2012) and wheat (Balyan et al., 2013). A few QTLs/genes associated with SPC have been successfully deployed in practical plant breeding through marker-assisted selection particularly in soybean (Sebolt et al., 2000; Chee et al., 2001) and wheat (Balyan et al., 2013; Zhang et al., 2015; Vishwakarma et al., 2014, 2016).

However, despite the importance of proteins stored in seed for human nutrition, and availability of a wealth of genomics resources in pigeonpea, SPC has remained untouched by the genomics revolution in this crop. As a result, the genetics of SPC in pigeonpea remain poorly understood, and no QTLs/genes for SPC have been identified that would otherwise facilitate marker assisted breeding (MAB) for the trait.

Therefore, the present study aimed at understanding the pattern of inheritance of SPC and identifying genomic segments/candidate genes associated with the trait that may facilitate MAB for SPC in pigeonpea. To do this, a combination of classical quantitative genetics and genomics approaches were applied. Further, the relationships between SPC and other important agronomic traits including seed yield, seed weight, days to flowering and growth habit were investigated.

The overall goal of this study was to contribute to the genetic improvement of SPC in pigeonpea by understanding its genetic control and identification of QTLs/genes controlling the trait. The specific objectives to achieve the goal were as follows:

1. Determine variation of seed protein content and its relationship with agronomic traits of importance in a set of pigeonpea genotypes, which parents of mapping populations
2. Study the inheritance of seed protein content and its relationship with seed weight and seed yield
3. Identify QTLs conditioning seed protein content and its association with agronomic traits in pigeonpea
4. Identify candidate genes involved in the accumulation of seed protein content in pigeonpea using whole genome sequencing approach

In light of the above-stated specific objectives, the following hypotheses were therefore tested:

1. There is no variation in seed protein content (SPC) among pigeonpea genotypes used as parents of mapping populations at ICRISAT. If the variation in SPC exists, then it is not related to variation in any of the agronomic traits of pigeonpea.
2. Seed protein content (SPC) is inherited in an additive-dominance manner, and the inheritance of SPC is not related with inheritance of seed yield or seed weight in pigeonpea
3. Quantitative trait loci (QTLs) for seed protein content (SPC) in pigeonpea are not associated with any single nucleotide polymorphism (SNP) markers, and the QTLs are only inherited in an additive-dominance manner. No QTL for SPC is co-inherited with QTLs conditioning agronomic traits in pigeonpea.
4. Candidate gene associated with seed protein content in pigeonpea cannot be identified using whole genome resequencing approach.

Thus, this thesis has been laid out as follows:

1. Introduction
2. Chapter One: Literature review

3. Chapter Two: Variation of seed protein content and its relationships with agronomic traits
4. Chapter Three: Inheritance of seed protein content and its association with seed weight and yield
5. Chapter Four: Quantitative trait loci analysis for seed protein content
6. Chapter Five: Identification of candidate genes for seed protein content using whole genome sequencing approach
7. Chapter Six: Overview of research findings

With the exception of the Introduction, Chapter One (review of literature) and Chapter Six (overview of research finding), all other chapters are written in the format: Introduction, Materials and Methods, Results, and Discussion (IMRAD). Each of the chapters has a reference list. There may also be few repetitions as well as overlapping content, especially among the introduction sections of the research chapters, the literature review chapter and also the reference sections of the chapters.

References

Akibode, C. S., and M. Maredia. 2011. Global and regional trends in production, trade and consumption of food legume crops. Report submitted to the Standing Panel on Impact Assessment (SPIA) of the CGIAR Science Council, FAO, Rome.

Balyan, H.S., P.K. Gupta, S. Kumar, R. Dhariwal, V. Jaiswal, S. Tyagi, P. Agarwal, V. Gahlaut, and S. Kumari. 2013. Genetic improvement of grain protein content and other health-related constituents of wheat grain. Plant Breeding 132:446-457.

Baskaran, K., and A.R. Muthiah. 2007. Studies on combining ability in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Legumes Research 30:67-69.

Baudoin, J.P., and A. Maquet. 1999. Improvement of protein and amino acid content in seeds of food legumes – a case study in *Phaseolus*. Biotechnology and Agronomical Society of Environment 3: 220-224.

Bohra, A., R.K. Saxena, B.N. Gnanesh, K.B. Saxena, M. Byregowda, A. Rathore, P.B. KaviKishor, D.R. Cook, and R.K. Varshney. 2012. An intra-specific consensus genetic map

of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theoretical and Applied Genetics 125:1325-1338.

Burstin, J., P. Marget, M. Huart, A. Moessner, B. Mangin, C. Duchene, B. Desprez, N. Munier-Jolain, and G. Duc. 2007. Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiology 144:768-781.

Burstin. J., K. Gallardo, R.R. Mir, R.K. Varshney, and G. Duc. 2011. Improving protein content and nutrition quality. In: Biology and breeding of food legumes, p. 314-328. Wallingford: CAB International.

Chee, P.W., E.M. Elias, J.A. Anderson, and S.F. Kianian. 2001. Evaluation of a high grain protein QTL from *Triticum turgidum* L. var. dicoccoides in an adapted durum wheat background. Crop Science 41:295-301.

Dahiya, B.S., J.S. Brar, and B.S. Bhullar. 1977. Inheritance of protein content and its correlation with grain yield in pigeonpea (*Cajanus cajan* (L.) MilIsp.). Qualitas Plantarum Plant Foods for Human Nutrition 27:327-334.

FAOSTAT. 2015. http://faostat3.fao.org. Accessed 20 April 2016.

Fauteux, F., and M.V. Strömvik. 2009. Seed storage protein gene promoters contain conserved DNA motifs in brassicaceae, fabaceae and poaceae. BMC Plant Biology 9:126.

Foley, J.A., N. Ramankutty, K.A. Brauman, E.S. Cassidy, J.S. Gerber, M. Johnston, N.D. Mueller, C. O'Connell, D.K. Ray, P.C. West, C. Balzer, E.M. Bennett, S.R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockström, J. Sheehan, S. Siebert, D. Tilman, and D.P.M. Zaks 2011. Solutions for a cultivated planet. Nature 478:337–342.

Frascaroli, E., M.A. Cane`, P. Landi, G. Pea, L. Gianfranceschi, M. Villa, M. Morgante, and M.E Pe`. 2007. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. Genetics 176:625–644.

Gilissen, C., A. Hoischen, H.G. Brunner, and J.A. Veltman. 2012. Disease gene identification strategies for exome sequencing. European Journal of Human Genetics 20:490–497.

Greilhuber, J., and R. Obermayer. 1998. Genome size variation in *Cajanus cajan* (Fabaceae): a reconsideration. Plant Systematics and Evolution 212:135-141.

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative genetics in maize breeding. Iowa State University Press, Ames, IA.

Holland, J.B., and A.J. Cardinal. 2008. Harnessing quantitative genetics and genomics for understanding and improving complex traits in crops. Drought Frontier Project Planning Workshop, North Carolina State University. p. 123-136.

Hwang, E.-Y., Q. Song, G. Jia., J.E. Specht, D.L. Hyten, J. Costa, and P.B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15:1.

Iqbal, A., I.A. Khalil, N. Ateeq, and M.S. Khan. 2006. Nutritional quality of important food legumes. Food Chemistry 97:331–335.

Jadhav, A.A., S.J. Rayate, L.B. Mhase, M. Thudi, A. Chitikineni, P.N. Harer, A.S. Jadhav, R.K. Varshney, and P.L. Kulwal. 2015. Marker-trait association study for protein content in chickpea (*Cicer arietinum* L.). Journal of Genetics 94:279-286.

Khoury, C.K., N.P. Castañeda-Alvarez, H.A. Achicanoy, C.C. Sosa, V. Bernau, M.T. Kassa, S.L. Norton, L.J.G. van der Maesen, H.D. Upadhyaya, J. Ramírez-Villegas, A. Jarvis, and P.C. Struik. 2015. Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) millsp.]: distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. Biological Conservation 184:259-270.

Krajewski, P., J. Bocianowski, M. Gawłowska, Z. Kaczmarek, T. Pniewski, W. Święcicki, and B. Wolko. 2012. QTL for yield components and protein content: a multi-environment study of two pea (*Pisum sativum* L.) populations. Euphytica 183:323-336.

Kumar, V., A.W. Khan, R.K. Saxena, V. Garg, and R.K. Varshney. 2016. First-generation HapMap in Cajanus spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnology Journal 14:1673-1681.

Kumawat, G., R.S. Raje, S. Bhutani, J.K. Pal, A.S.V.C.R. Mithra, K. Gaikwad, T.R.S. Sharma, and N.K. Singh. 2012. Molecular mapping of QTLs for plant type and earliness traits in pigeonpea (*Cajanus cajan* L. Millsp.). BMC Genetics 13:84.

Li, L., W. Zheng, Y. Zhu, H. Ye, B. Tang, Z.W. Arendsee, D. Jones, R. Li, D. Ortiz, X. Zhao, C. Du, D. Nettleton, M.P. Scott, M.P. Salas-Fernandez, Y. Yin, and E.S. Wurtele. 2015. QQS

orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. Proceedings of the National Academy Sciences, USA 112:14734-14739.

Lu, W., Z. Wen, H. Li, D. Yuan, J. Li, H. Zhang, Z. Huang, S. Cui, and W. Du. 2013. Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. Theoretical and Applied Genetics 126:425-433.

Mir, R.R., H. Kudapa, S. Srikanth, R.K. Saxena, A. Sharma, S. Azam, K.B. Saxena, R.V. Penmetsa, and R.K.Varshney. 2014. Candidate gene analysis for determinacy in pigeonpea (*Cajanus* spp.). Theoretical and Applied Genetics 127:2663-2678.

Mligo, J.K., and P.Q. Craufurd. 2005. Adaptation and yield of pigeonpea in different environments in Tanzania. Field Crops Research 94:43-53.

Mula, M.G., and K.B. Saxena. 2010. Lifting the level of awareness on pigeonpea – a global perspective. International Crops Research Institute for the Semi-Arid Tropics – Patancheru, India.

Noubissié, J.B.T., J.M. Bell, and H.T. Yadji. 2012. Studies on variability and gene effects of seed protein contents in Cameroonian bean (*Phaseolus vulgaris*) cultivars. Journal of Agriculture and Social Sciences 8:17–23.

Odeny, D.A. 2007. The potential of pigeonpea (*Cajanus cajan* (L.) Millsp.) in Africa. Natural Resources Forum 31:297-305.

Pazhamala, L., R.K. Saxena, V.K. Singh, C.V. Sameerkumar, V. Kumar, P. Sinha, K. Patel, J. Obala, S.R. Kaoneka, P. Tongoona, H.A. Shimelis, N.V.P.R. Gangarao, D. Odeny, A. Rathore, P.S. Dharmaraj, K.N. Yamini, and R.K .Varshney. 2015. Genomics-assisted breeding for boosting crop improvement in pigeonpea (*Cajanus cajan*). Frontiers in Plant Science 6:50.

Pflieger, S., V. Lefebvre, and M. Causse. 2001. The candidate gene approach in plant genetics: a review. Molecular Breeding 7:275–291.

Rao, P.P., P.S. Birthal, S. Bhagavatula, and M.C.S. Bantilan. 2010. Chickpea and pigeonpea economies in Asia: facts, trends and outlook. Patancheru 502 324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics.

Rekha, R., L. Prasanthi, M.R. Sekhar, and M.S. Priya. 2013. Studies on selection indices in pigeonpea (*Cajanus cajan* (L.) Millsp.). International Journal of Applied Biology and Pharmaceutical Technology 4:291-294.

Remanandan, P.D., V.S.S.R. Sastry, and H.M. Mengesha. 1988. ICRISAT pigeonpea germplasm catalog: passport information. Andra Pradesh (India): International Crops Research Institute for the Semi-arid Tropics.

Sahu, A.R., R.R. Mishra, R.S. Chandra, and J. Panigrahi. 2015. Construction of interspecific genetic linkage map of pigeonpea using SCoT, RAPD, ISSR markers and simple inherited trait loci. Nucleus 58:23-31.

Santos, C.A.F., D.C.C. da Costa, W.R. da Silva, and L.S. Boiteux. 2012. Genetic analysis of total seed protein content in two cowpea crosses. Crop Science 52:2501-2506.

Saxena, K.B. 2005. Pigeonpea (*Cajanus cajan* (L.) Millsp.). In: Singh RJ, and P.P. Jauhar, editor, Genetic resources, chromosome engineering and crop improvement: Vol. 1, Grain Legumes. CRC Press, Boca Raton. p. 86-110.

Saxena, K.B., and S.L. Sawargaonkar. 2015. Genetic enhancement of seed proteins in pigeonpea – methodologies, accomplishments and opportunities. International Journal of Scientific Research 4:254-258.

Saxena, K.B., D.G. Faris, U. Singh, and R.V. Kumar. 1987. Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition 36:335-340.

Saxena, R.K., R.V. Penmetsa, H.D. Upadhyaya, A. Kumar, N. Carrasquilla-Garcia, J.A. Schlueter, A. Farmer, A.M. Whaley, B.K. Sarma, G.D. May, D.R. Cook, and R.K. Varshney. 2012. Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeon pea and comparative mapping in legumes. DNA Research 19:449-461.

Saxena, R.K., X. Cui, V. Thakur, B. Walter, T.J. Close, and R.K. Varshney. 2011. Single feature polymorphisms (SFPs) for drought tolerance in pigeonpea (*Cajanus* spp.). Functional and Integrative Genomics 11:651-657.

Sebolt, A.M., R.C. Shoemaker, and B.W. Diers (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Science 40:1438-1444.

Silva, J., B. Scheffler, Y. Sanabria, C. de Guzman, D. Galam, A. Farmer, J. Woodward, G. May, and J. Oard. 2012. Identification of candidate genes in rice for resistance to sheath blight disease by whole genome sequencing. Theoretical and Applied Genetics 124:63–74.

Singh, V.K., A.W. Khan, R.K. Saxena, V. Kumar, S.M. Kale, P. Sinha, A. Chitikineni, L.T. Pazhamala, V. Garg, M. Sharma, C.V.S. Kumar, S. Parupalli, S. Vechalapu, S. Patil, S. Muniswamy, A. Ghanta, K.N. Yamini, P.S. Dharmaraj, and R.K. Varshney. 2016. Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). Plant Biotechnology Journal 14:1183-1194.

Sobreira, N.L.M., E.T. Cirulli, D. Avramopoulos, E. Wohler, G.L. Oswald, E.L. Stevens, D. Ge, K.V. Shianna, J.P. Smith, J.M. Maia, C.E. Gumbs, J. Pevsner, G. Thomas, D. Valle, J.E. Hoover-Fong, and D.B. Goldstein. 2010. Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genetics 6:e1000991.

Soybase. 2016. Soybean breeder's toolbox genetic map information. Available online at: http://www.soybase.org [Accessed on: 20 April 2016).

Upadhyaya, H.D., D. Bajaj, L. Narnoliya, S. Das, V. Kumar, C.L.L. Gowda, S. Sharma, A.K. Tyagi, and S.K. Parida. 2016. Genome-wide scans for delineation of candidate genes regulating seed-protein content in chickpea. Frontiers in Plant Science 7:302.

Upadhyaya, H.D., K.N. Reddy, D.V.S.S.R. Sastry, and C.L.L. Gowda. 2007. Identification of photoperiod insensitive sources in the world collection of pigeonpea at ICRISAT. Journal of Semi-Arid Tropical Agricultural Research 3:46-49.

Upadhyaya, H.D., S. Kumar, C.L.L. Gowda, and S. Singh. 2006. Two major genes for seed size in chickpea (*Cicer arietimum* L.). Euphytica 147: 311-315.

Vaghela, K.O., R.T. Desai, J.R. Nizama, J.D. Patel, and V. Sharma. 2009. Combining ability analysis in pigeonpea (C*ajanus cajan* (L.) Millsp.). Legume Research 32:274-277.

Varshney, R.K., R. Terauchi, and S.R. McCouch. 2014. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biology 2:e1001883.

Varshney, R.K., R.K. Saxena, H.D. Upadhyaya, A.W. Khan, Y. Yu, C. Kim, A. Rathore, D. Kim, J. Kim, S. An, V. Kumar, G. Anuradha, K.N. Yamini, W. Zhang, S. Muniswamy, J. Kim, R.V. Penmetsa, E. von Wettberg, S.K. Datta. 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nature Genetics 49:1082–1088.

Varshney, R.K., W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T.A. Donoghue, S. Azam, G. Fan, A.M. Whaley, A.D. Farmer, J. Sheridan, A. Iwata, R. Tuteja, R.V. Penmetsa, W. Wu, H.D. Upadhyaya, S. Yang, T. Shah, K.B. Saxena, T. Michael, W.R. McCombie, B. Yang, G. Zhang, H. Yang, J. Wang, C. Spillane, D.R. Cook, G.D. May, X. Xu, and S.A. Jackson. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology 30:83-89.

Vishwakarma, M.K., B. Arun, V.K. Mishra, P.S. Yadav, H. Kumar, and A.K. Josh. 2016. Marker-assisted improvement of grain protein content and grain weight in Indian bread wheat. Euphytica 208:313–321.

Vishwakarma, M.K., V.K. Mishra, P.K. Gupta, P.S. Yadav, H. Kumar, and A.K. Joshi. 2014. Introgression of the high grain protein gene Gpc-B1 in an elite wheat variety of Indo-Gangetic Plains through marker assisted backcross breeding. Current Plant Biology 1:60–67.

Wang, J., P. Chen, D. Wang, G. Shannon, A. Zeng, M. Orazaly, and C. Wu. 2015. Identification and mapping of stable QTL for protein content in soybean seeds. Molecular Breeding 35:92.

Wang, X., Jiang, G.-L, M. Green, R.A. Scott, Q. Song, D.L. Hyten, and P.B. Cregan. 2014. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. Molecular Genetics and Genomics 289:935–949.

Warrington, C.V., H. Abdel-Haleem, D.L. Hyten, P.B. Cregan, J.H. Orf, A.S. Killam, N. Bajjalieh, Z. Li, and H.R. Boerma. 2015. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. Theoretical and Applied Genetics 128:839-50.

Xu, G., X. Fan, and A.J. Miller. 2012. Plant nitrogen assimilation and use efficiency. Annual Review of Plant Biology 63:153-182.

Xu, J., Y. Yuan, Y. Xu, G. Zhang, X. Guo, F. Wu, Q. Wang, T. Rong, G. Pan, M. Cao, Q. Tang, S. Gao, Y. Liu, J. Wang, H. Lan, and Y. Lu. 2014. Identification of candidate genes for drought tolerance by whole-genome re-sequencing in maize. BMC Plant Biology 14:83.

Zhang, Y.H., M.F. Liu, J.B. He, Y.F. Wang, G.N. Xing, Y. Li, S.P. Yang, T.J. Zhao, and J.Y. Gai. 2015. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.]. Theoretical and Applied Genetics 128:1061-1072.

Zhu, M., and S. Zhao. 2007. Candidate gene identification approach: Progress and challenges. International Journal of Biological Science 3:420-427.

Chapter One
Review of literature

Review of literature provides insight into the manner in which a given research problem has previously been tackled, the nature of results obtained and the conclusions drawn. It covers research work that might have been carried out on the same or similar or another species, in different regions, and under different sets of environmental conditions. Nevertheless, information gathered from such previous studies is helpful in adopting appropriate study design including formulation of research hypotheses, applying suitable methodology and proper interpretation of results. This chapter, therefore, gives an overview of the taxonomy, origin, distribution and diversity of pigeonpea. It also briefly presents the importance of pigeonpea and composition of its seed. A deeper review of genetic variation of seed protein content, and its relationships with other traits of importance in pigeonpea and other legume crops have also been covered. The review also looked at the various methods used in deciphering factors that influence variations of quantitatively inherited traits which includes classical quantitative genetic methods and the "omics" technologies with an in-depth analysis of genomics approaches that have been used and/or available for pigeonpea genetic analysis. A summary of the review is provided which draws out gaps in the literature, and explains how the gaps are covered by the studies that have been accomplished in Chapter Two to Chapter Five. Lastly, a section on prospects for future research on seed protein content based on developments in the plant biology systems approaches was included. It is, however worth noting that while classifying the literature into different sections and sub-sections, information relevant to the specific objectives of the thesis remained the centre of focus.

1.1 Taxonomy, origin, distribution and diversity of pigeonpea

Pigeonpea (*Cajanus cajan* (L.) Millsp.) is the only cultivated food legume of the sub-tribe Cajaninae, which belongs to the tribe Phaseoleae. The tribe Phaseoleae also contains important crop species including soybean (*Glycine max* L.), common bean (*Phaseolus vulgaris* L.), mungbean (*Vigna radiata* L. Wilczek) and others (Young et al., 2003). The genus *Cajanus* comprises of 34 species (Lewis et al., 2005) most of which are found in Southern and South-

Eastern Asia, and Australia (Fortunato, 2000; van der Maesen, 1990) although one is native to West Africa (Purseglove, 1968; Tindall, 1988).

Pigeonpea is a sub-tropical and tropical grain legume that is believed to have originated from the northern region of the Indian sub-continent, spreading to East Africa at least 4000 years BCE, and then to Southeast Asia, West Africa, Latin America, and the Caribbean (Khoury et al., 2015). Pigeonpea is an often cross pollinated species with 11 pairs of chromosomes (2n=2x=22) and a genome size of 833.07 Mbp (Varshney et al., 2012).

Pigeonpea germplasm represents a diverse set of landraces and wild relatives that are adapted to different sets of environments (Saxena et al., 2008). Despite extensive phenotypic diversity, molecular evidence from diversity array technology (DArT) (Yang et al., 2006) and simple sequence repeats (SSRs) (Odeny et al., 2007) suggests very low genetic diversity within cultivated pigeonpea when compared to its wild relatives. Due to the low genetic diversity in the cultivated genepool attempts have been made to introgress a number of desirable traits from the secondary genepool in to the cultivated genepool (Saxena et al., 2008).

1.2 Importance of pigeonpea

Pigeonpea is a major legume crop, which accounts for 5% of total legume production in the world (Hillocks et al., 2000; Khoury et al., 2015). Pigeonpea has wide adaptation, producing desirable yields in humid as well as hot and dry environments although it is frost sensitive, it can still produce considerable yields of 0.2 to 2.5 t/ha across a broad range of environments (Snapp et al., 2003). The crop is largely grown by subsistence farmers in the warm semi-arid and sub-tropics, and often on marginal soils with low inputs (Adu-Gyamfi et al., 2007). Pigeonpea plays an important role in food security, balanced diet and alleviation of poverty because it can be used in diverse ways as a source of food, feed, fodder, fuel wood (Rao et al., 2002). As a leguminous plant, pigeonpea contributes as much as 40 kg of nitrogen per hectare per year to the soil (Emefiene et al., 2014).

Pigeonpea seed is a major source of protein to about 20% of the total world population (Thu et al., 2003) and it is an abundant source of minerals and vitamins (Saxena et al., 2002). The protein-rich

seed of pigeonpea makes it an ideal supplement to traditional cereal-, banana- or tuber-based diets of resource poor farmers that are generally protein-deficient (Odeny et al., 2007).

Typically, the mature dry pigeonpea seed may contain 13.2-26.5 % SPC, 56.3-64.1% starch, 4.7-5.7% soluble sugars, 1.2-2.2% fat, 1.0-1.2% crude fibre and 3.3-4.3% ash (Singh and Jambunathan, 1984; Saxena et al., 2002). The major storage proteins in pigeonpea seeds are the salt-soluble globulins (59.9% of the total protein), acid/alkali-soluble glutenin (17.4%), water-albumin (10.2%), and alcohol-soluble prolamin (3.0%) (Singh and Jambunathan, 1984). However, pigeonpea seed protein is rated inferior to that of other legumes in terms of sulphur-containing amino acids resulting from the low proportion of the albumin fraction (Saxena et al., 2002). Singh et al. (1990) assessed chemical composition of high SPC lines developed from crosses between cultivated pigeonpea and an accession of a wild relative. They reported large differences in SPC between high-SPC lines (28.7 to 31.1%) and control cultivars (23.1 to 24.8%), while the starch content was lower in the high-SPC lines (54.3 to 55.6 %) than in the control cultivars (58.7 to 59.3%). They also observed that the globulin protein fraction in the control cultivars (60.3 to 60.5%) was lower than that in the high-SPC lines (63.5 to 66.2%), and the reverse was true for glutelin. The accumulation of proteins in the seed is conditioned by genetic and environmental factors (Martre et al., 2002). Because of the importance of SPC of pigeonpea in human nutrition it is essential to understand the genetic basis of the trait for its genetic improvement.

1.3 Genetic variation and environmental effects for seed protein content

Substantial genetic variation has been observed for total SPC in the gene pools of grain legumes including primary and alien reservoirs (Baudoin and Maquet, 1999). In the secondary genepool of some wild relatives of pigeonpea such as *C. scarabaeoides*, *C. sericeus*, and *C. albicans* SPC ranges from 24.1 to 34.4% (Saxena et al., 1990; Reddy et al., 2000; Saxena et al., 2002). While in cultivated pigeonpea reports indicated SPC to range from 12.8 to 29.0% (Remanandan et al., 1988), 19.5 to 22.9% (Hedley, 2001) and 16.1 to 24.1% (Upadhyaya et al., 2007). Considering the lower SPC within the primary gene pool, a breeding program to transfer high SPC into pigeonpea from its wild relatives was undertaken at ICRISAT. However, these wild species have a number

of agronomically undesirable traits such as bushy or trailing plant type, small dark coloured seeds, hard seed coat, pod shattering, and low yield, which make the development of desirable high yielding, high SPC cultivars challenging (Saxena et al. 2008). This therefore implies that breeding within the cultivated genepool could be more beneficial even with only moderate expected genetic gain for SPC.

Besides genotypic differences, environmental effects on SPC in the legumes have been noted to be large but genotype × environment interactions (GEI) are often small, indicating that the relative differences between genotypes should be similar in several environments (Baudoin and Maquet, 1999). This is true in the case of pigeonpea for which it has been reported that although GEI was found statistically significant, it did not affect the ranking between high and low SPC lines tested over several seasons in two locations (Saxena et al., 1987; Saxena et al., 2002).

1.4 Relationships of seed protein content with agronomic characters

Increasing or maintaining yield is an overall ultimate objective of any breeding programme (Cromwell et al., 1992). Because selection for one character often leads to simultaneous change in other characters (Falconer, 1989), an understanding of the relationship of SPC with seed yield (SY) and yield-related characters is important for designing effective breeding strategy for genetic enhancement of SPC.

The relationship between SPC and other plant characters have been reported in pigeonpea as well as in other food legume crops. In an evaluation of 1,974 single F_7 plants from intergeneric crosses, highly significant correlations between SPC and seed weight (SW) was reported (Saxena et al., 1987). Two crosses in that study exhibited significant negative correlations and one showed a significant positive correlation while in two other crosses no significant association was detected. Based on all the selections, a highly significant negative correlation between SPC and SW was found. However, the extent of variation that could be attributed to this association was small, and it was concluded that, simultaneous improvement could be made for SPC and SW in pigeonpea.

Similarly, in an evaluation of 40 pigeonpea genotypes collected from different states of India, small, positive and non-significant correlation between SPC and SW, SPC and number of pods per plant, and between SPC and plant height (PH) were observed (Rekha et al., 2013). In the same study, correlation between SPC and SY, and between SPC and days to flowering were small, negative and non-significant. The only significant correlation was between SPC and number of seeds per pod, which was small and negative.

Reports in other legume crops such as common bean, cowpea, soybean and mung bean are similar to the mixed results reported in pigeonpea with an often negative and at times positive, and in some cases non-significant relationship of SPC with SY, SW and other plant traits. For example, significant and negative relationships between SPC and SY have been observed in common bean (Leleji et al., 1972), and cowpea (Bliss et al., 1973; Oluwatosin, 1997). While non-significant and negative relationships have been reported in lentil (Hamdi et al. 1991) and soybean (Cober and Voldeng, 2000). Similar conclusions were drawn on the effect of the correlation between SPC and SW in pigeonpea (Saxena et al., 1987). On the other hand, selection for high SPC has often but not always led to SY reduction (Leleji et al., 1972, Brim and Burton, 1978, Wilcox and Cavins, 1995).

Correlations of SPC with plant morphological and phenological traits have also been reported. For example, a significant positive correlation between SPC and duration of reproductive phase in mungbean (Lawn and Rebetzke, 2006) and between SPC and plant height in pea (Burstin et al., 2007) have been reported. However, relationships among traits are dependent upon the set of materials evaluated and the environment in which they are studied (Hamdi et al., 1993). Therefore, it is important to continuously assess the germplasms for SPC and its relationship with other traits and draw the conclusions before applying specific material in genetic improvement programs.

1.5 Genetics of seed protein content

The nature and magnitude of genetic effects controlling a character is important in the interpretation of quantitative genetic experiments. Such information also guides in designing a breeding methodology for cultivar development and determining cultivar types i.e. whether hybrid, pure line or synthetics (Lamkey and Edwards, 1999). There are two broad approaches that are commonly used for determining genetic control of quantitatively inherited traits. These approaches include: (1) classical quantitative genetics methods such as variance component analysis (VCA) and generation mean analysis (GMA) (Hallauer and Miranda, 1988), and (2) genomics approaches such as quantitative trait loci (QTL) analysis (Frascaroli et al., 2007) and candidate gene identification (Zhu and Zhou, 2007).

1.5.1 Classical quantitative genetic analysis

Variance component analysis and GMA have been used to varying extents in deriving inferences about genetic effects for agronomic as well as seed traits including SPC in the legume crops. VCA relies on mating designs such as the diallels, North Carolina (NCD), and line × tester (L×T) designs. Diallel analysis has been used to study the genetics of SPC, seed oil content and SY in peanut lines with results indicating the importance of additive gene action over non-additive effects for SPC (Layrisse et al., 1980). L×T design to determine gene action for SPC in faba bean indicated that both additive and non-additive effects were important in controlling the trait (Fillipetti et al., 1999). A combining ability study using a full diallel design indicated that non-additive effects were more important than additive effects for SPC in mung bean (Tiwari et al. 1993). Similarly, a full diallel analysis and L×T analysis revealed the importance of non-additive effects over additive effects for SPC in common bean (Mebrahtu and Mohamed, 2003; Iqbal et al., 2012; Ceyhan et al., 2014). Hazra et al. (1996) using a diallel analysis and Santos et al. (2012) using GMA reported additive effects to be more important than non-additive effects while

Tchiagam et al. (2012) reported non-additive effects to be more important than additive effects for SPC in cowpea when using a diallel analysis. Similarly, L×T design analysis and a diallel analysis revealed the importance of non-additive over additive gene action for SPC in pigeonpea (Baskaran and Muthiah, 2007; Vaghela et al., 2009).

The VCA-based methods have largely been used to reveal information required to guide selection of parental lines for hybridization as well as the relative importance of additive vs non-additive effects (Hallauer and Miranda, 1988). Variance component analysis, however, lacks the power to differentiate among the components of non-additive genetic effects, which include dominance and the different types of epistasis. Generation mean analysis on the other hand has been indicated to be more robust than the VCA methods, because it allows for simultaneous detection of the main additive-dominance effects and the epistatic effects (Hallauer and Miranda, 1988; Bernado, 2010). Generation mean analysis has been used for detecting genetic effects controlling SPC in cowpea (Santos et al., 2012; Tchiagam et al., 2011) and in common bean (Noubissié et al., 2012) However, despite the advantages of GMA in detecting genetic effects it has not been used for SPC in pigeonpea.

Besides estimation of variances and genetic effects, quantitative genetics analysis also makes it possible to predict changes in the structure of a breeding population as affected by selection and other forces of evolution (Geiger and Tomerius, 1997). A common measurement of quantitative traits in plant breeding programs is heritability, which is a quantification of the proportion of phenotypic variance that is attributable to genetic effects or, in other words, that is exploitable by selection (Holland et al., 2003). In pigeonpea an evaluation of two crosses between low and high SPC genotypes in one environment revealed broad-sense heritability (H^2) of 0.34 to 0.62 on plant mean basis. In soybean, H^2 of 0.83 in two populations evaluated for the identification of stable QTLs for SPC and oil concentration was reported (Lee et al., 1996). Similarly, Hwang et al. (2014) estimated H^2 of 77.9% on entry mean basis from 298 soybean accessions of genome-wide association panel evaluated in two locations for the identification SPC QTLs. Recently, Wang et al. (2014) estimated H^2 to be 87 to 94% for SPC on plot basis in two populations of F_5-derived recombinant inbred lines (RILs) in soybean evaluated in five environments. Wiggins (2012)

reported H^2 of 0.78 on entry basis and narrow-sense heritability (h^2) of 0.27 using data from 239 F_4-derived RILS in five environments. In cowpea, Tchiagam et al. (2011) obtained H^2 estimates of 0.77 to 0.78 and h^2 of 0.16 to 0.41 on plant mean basis in a generation mean analysis study conducted in one environment. Similarly, Emibiri (1991) estimated H^2 for SPC to be 0.74 in cowpea. With the exception of the study of Dayiha et al. (1977), estimates of heritability for SPC in pigeonpea are very limited

Estimates of heritability are used for calculating expected response to selection (Holland et al., 2003). There is no information on expected response to selection for SPC in pigeonpea. However, in cowpea, an increase of 3.6 to 10.16% in SPC has been predicted after one cycle of selection at 10% selection intensity (Noubissié et al., 2012). Similarly, gains of 19.6, 16.7, 14.8 and 13.4 % for SPC at 5, 10, 15 and 20% selection intensities have been predicted in soybean (Wiggins, 2012).

Response to selection is affected by the number of segregating genes in a population for the target trait. The higher the number of segregating genes, the higher the maximum population mean achievable through selection and the longer the duration to achieve the selection limit (Geiger and Heun, 1989). The number of genes controlling SPC has been estimated to be three to four in pigeonpea (Dahiya et al., 1977), three to seven in cowpea (Santos et al., 2012), and one to 10 in common bean (Noubissié et al., 2012). Although these estimates of number of genes controlling SPC based on means and variances have thrown some light on the probable number of effective factors controlling the trait, they can only estimate the minimum number of loci segregating in a population. This minimum number genes can be biased if the underlying assumptions are violated when not all alleles behave additively, there is linkage among loci, unequal effects of alleles the two parental strains are not diploid and are heterozygous for alternative alleles at all loci affecting the trait (Jones, 2001). With the developments in the field of genomics precision in estimating the number and effects of loci affecting a trait have increased. Genomics is one of the fields belonging to the systems biology approach, which is also generally referred to as the “omics” (Kaddurah-Daouk et al., 2008; Sheth and Thaker, 2014)

1.5.2 The omics

The omics include the study of entire metabolome (metabolics), proteome (proteomics), transcriptome (transcriptomics) and genome (genomics) of a system (Sheth and Thaker, 2014; Kaddurah-Daouk et al., 2008) in a specific biological sample in a non-targeted and non-biased manner (Horgan and Kenny, 2011).

Metabolomics can generally be defined as the study of global metabolite profiles in a system (cell, tissue or organism) under a given set of conditions (Goodracre et al., 2004). Metabolomics has a number of theoretical advantages over the other omic approaches. The metabolome is the final downstream product of gene transcription and, therefore, changes in the metabolome are amplified relative to changes in the transcriptome and the proteome (Urbanczyk-Wochniak et al., 2003). Additionally, as the downstream product, the metabolome is closest to the phenotype of the biological system studied. Although the metabolome contains the smallest domain among the 'omes,' it is more diverse, containing many different biological molecules, making it more physically and chemically complex than the other "omes" (Sheth and Thaker, 2014). Metabolomics based approach can provide a comprehensive understanding of seed metabolism and more generally of seed quality because seed traits are inherently associated with seed metabolism and plant-seed carbon-nitrogen allocation (Toubiana and Fait, 2012). This seems to have been recognized earlier. For example, by using a combination of metabolomics and genetic approaches, it was possible for Vigeolas et al. (2008) to reveal a link between the polyamine pathway and albumin 2 in pea (*Pisum sativum*). In a similar manner, Li et al. (2015) used an integrated metabolomics and transcriptomics data to understand seed composition in soybean. They concluded that that during soybean seed development, modulations in end-products of metabolism are affected by a small proportion of the soybean genome, and that the majority of gene transcripts showed a relatively constant level of expression. They also concluded that the metabolome is more sensitive to the developmental program than is the transcriptome. These conclusions suggest that as genetic information is expressed through the processes of transcription and translation, coupled with the catalytic properties of the proteome, subtle changes at the transcriptome level are amplified at the level of the metabolome.

The proteome is defined as the set of all expressed proteins in a cell, tissue or organism (Theodorescu and Mischak, 2007). Proteomics aims to characterize information flow within the cell and the organism, through protein pathways and networks (Petricoin et al., 2002) with the ultimate target of understanding the functional relevance of identified proteins (Vlahou and Fountoulakis, 2005). The proteome is a dynamic reflection of both genes and the environment, and it has great potential for discovery of biomarkers because proteins are most likely to be universally influenced under varied environmental conditions (Rifai et al., 2006). Information obtained through proteomic analysis are important for decoding of protein structure and complex mechanisms such as enzymatic and regulatory functions of proteins coded by specific genes (Ramalingam et al., 2015). However, many proteomics based publications that relate to plant development and other biological processes and events in the model plants including legumes and *Arabidopsis thaliana*, and also in crop plants such as rice (*Oryza sativa*), wheat (*Triticum aestivum*), maize (*Zea mays*), soybean, tomato (*Solanum lycopersicum*), tobacco (*Nicotiana tabacum*) (Jorrín-Novo et al., 2015; Ramalingam et al., 2015) and very recently in pigeonpea (Krishnan et al., 2017). Consistent with the high degree of synteny reported between the pigeonpea and soybean genomes, the pigeonpea seed proteome map revealed that a large number of pigeonpea seed proteins exhibited significant amino acid homology with soybean seed proteins (Krishnan et al., 2017). The pigeonpea seed proteomic analysis also identified a large number of stress-related proteins, presumably due to its adaptation to drought-prone environments (Krishnan et al., 2017). The availability of a pigeonpea seed proteome reference map should shed light on the roles of these identified proteins in various biological processes and facilitate the improvement of seed composition.

The transcriptome is the total mRNA in a cell or organism and the template for protein synthesis in a process called translation. The transcriptome reflects the genes that are actively expressed at any given moment. The application of transcriptomics to unravel underlying genetic factors in the legumes abound. For instance, Verdier et al. (2008) developed a gene expression profile of *M. truncatula* transcription factors and identified putative regulators of grain legume seed filling. While Weigelt et al. (2009) studied specific transcriptional and metabolic changes of ADP-glucose

pyrophosphorylase-deficient pea embryos to reveal changes of carbon-nitrogen metabolism and stress responses. In pigeonpea, Kudapa et al. (2012) developed a comprehensive transcriptome assembly based on a hybrid approach consisting of Sanger expressed sequence tags (ESTs) and mRNA sequence data. This pigeonpea transcriptome assembly as well as several transcriptome datasets have been used to develop functional markers (Dubey et al., 2011; Saxena et al., 2012). The combination of transcript profiles and genome variants also can help in the identification expression quantitative trait loci (eQTLs), i.e the discovery of genetic variants that explain variation in gene expression levels, as well as in mapping regions with *cis*- and *trans*-effects (Nica and Dermitzakis, 2013). This is an area that could be useful in pigeonpea and should be explored.

Genomics is the systematic study of an organism's genome, which is the total DNA of a cell or organism (Horgan and Kelly, 2011). It lies at the base of the complex plant systems' hierarchy and it provides an understanding toward the organisms' behavioural explanation (Sheth and Thaker, 2014). The era of single gene sequencing marked the beginning of plant genomics followed by whole genome sequencing, single nucleotide polymorphism (SNP) and medium density arrays, and eventually led to the current whole genome resequencing (WGRS) (Sheth and Thaker, 2014). Genomics resources include molecular markers, genetic maps, genome assemblies and mapping populations (Pazhamala et al., 2015)

1.5.3 Genomic resources in pigeonpea and analysis of quantitative traits

The recent availability of these genomic resources in pigeonpea has enabled plant geneticists and breeders to study germplasm diversity, and to recognize and tag novel genes and alleles in the crop (Varshney et al., 2010; Varshney et al., 2013; Pazhamala et al., 2015).

1.5.3.1 Molecular markers and genetic maps

Molecular markers are grouped into three categories based on the methods of their detection, namely; hybridization-based markers, polymerase chain reaction (PCR)-based markers; and sequence-based markers (Collard et al., 2005). Different types of molecular markers have been

developed and used in pigeonpea genetic studies. The markers have included restriction fragment length polymorphism (RFLP) (Nadimpali et al., 1993; Sivaramakrishnan et al., 1997; Sivaramakrishnan et al., 2002; Lakshmi et al., 2000), randomly amplified polymorphic DNA (RAPD) markers (Ratnaparkhe et al., 1995; Lohithaswa et al., 2003; Malviya and Yadav, 2010), amplified fragment length polymorphism (AFLP) (Panguluri et al., 2005; Wasike et al., 2005; Aruna et al., 2008), short codon targeted (SCoT) polymorphism (Sahu et al., 2015), simple sequence repeats (SSRs) developed from different sources including genome sequence (gSSRs), expressed sequence tags (ESTs-SSRs) and bacterial artificial chromosome (BAC)-end sequences (BES-SSRs) (Odeny et al., 2007; Aruna et al., 2008; Saxena et al., 2010; Songok et al., 2010; Upadhyaya et al., 2011). With development of high-throughput next generation sequencing (NGS) technologies, marker types such as diversity arrays technology (DArT) (Yang et al., 2006, Yang et al., 2011), SNPs and related assay platforms such as GoldenGate assay and competitive allele-specific polymerase chain reaction (KASPar) (Saxena et al., 2012). Single feature polymorphisms (SFPs) (Saxena et al., 2011) and intron spanning region (ISR) markers (Kudapa et al., 2012) have been developed and used to varying degrees in pigeonpea. The most common use of these markers in pigeonpea have been for the assessment of genetic diversity within crop germplasm in case of RFLPs, AFLPs, RAPDs and SSRs, and the construction of genetic maps for mapping QTLs and genes controlling economically important traits in case of RAPDs, SSRs, DArTs, single nucleotide polymorphism (SNPs) and its derivatives - SFPs and KASPars (PKAMs) (Pazhamala et al., 2015).

The developments of molecular markers, have enhanced construction of genetic maps for the identification of underlying genetic variants associated with target traits. The first genetic map of pigeonpea was developed by Yang et al. (2011) using diversity arrays technology (DArT) markers in an F_2 mapping population of 72 individuals derived from an interspecific cross between ICP 28 (*Cajanus cajan*) and ICPW 94 (*C. scarabaoeides*). Maternal and paternal maps were generated based on 122 and 172 unique DArT loci with map lengths of 270.0 cM and 451.6 cM, respectively. An interspecific genetic map was constructed (Bohra et al., 2011) based on 239 SSR loci with a total map length of 930.9 cM. Saxena et al. (2012) developed a genetic map using 167 F_2 individuals derived from the same cross as in the Yang et al. (2011) and Bohra et al. (2011) studies using PKAMs (pigeonpea KASPar markers). A total 875 PKAMs were mapped with an average

intermarker distance of 1.11 cM. SSR markers were also integrated into the PKAM map and a total of 910 markers could be mapped with map distance of 996.21 cM. Sahu et al. (2015) using 116 F_2 interspecific cross constructed a genetic map with 191 markers spanning a total length of 1624.71 cM with average marker interval of 8.51 cM. The markers included 31 SCoT, 148 RAPD, and six inter-simple sequence repeats (ISSRs) markers and six simply inherited trait loci.

Gnanesh et al. (2011) generated two F_2 intraspecific genetic maps with 120 and 78 SSR markers spanning a distance of 534.89 cM and 466.97 cM, respectively. Bohra et al. (2012) constructed four genetic maps based on intraspecific F_2 populations comprising 59-140 simple sequence repeat (SSR) loci with map lengths ranging from 586.9 to 881.6 cM. In the same study, the four intraspecific maps together with two previous intraspecific maps were used to construct a consensus map comprised of 339 SSR loci spanning a distance of 1,059 cM. In a similar study, Kumawat et al. (2012) using a population of $F_{2:3}$ lines derived from an intraspecific cross between inbred lines 'Pusa Dwarf' and 'HDM04-1' constructed a genetic map of 296 genic SNP and SSR markers covering a map length of 1520.22 cM with average marker interval of 4.95 cM.

Availability of the pigeonpea draft genome sequence (Varshney et al., 2012) and next generation sequencing (NGS) technologies have led to the application of new genotyping methodologies such as genotyping-by-sequencing (GBS) and whole genome resequencing (WGRS). Genotyping-by-sequencing and WGRS approaches provide the possibility to generate high-density SNPs, and insertion and deletion (indel) genotyping data. These new technologies are being used in several crop species including soybean (Hyten et al., 2010; Hwang et al., 2014), grapevine (Lijavetzky et al., 2007), barley and wheat (Poland et al., 2012), chickpea (Jaganathan et al., 2014; Kale et al., 2015) as well as in pigeonpea (Singh et al., 2016; Saxena et al., 2017a,b; Varshney et al. 2017. However, the genomic technologies have not yet been specifically applied for unravelling the genetic architecture of SPC in pigeonpea.

1.5.3.2 Quantitative trait loci mapping for seed protein content

Quantitative trait loci mapping identifies regions of the genome that are contributing to variation in the trait of interest (Broman, 2001). The information obtained from QTL mapping provides a mechanism to track the co-segregation of genetic markers with the target trait in segregating populations. In pigeonpea, studies have been reported on molecular mapping of QTLs for agronomic characters, for example, Kumawat et al. (2012) evaluated $F_{2:3}$ mapped QTLs for six agronomic traits including plant height, number of primary branches, number of pods per plant, days to flowering and days to maturity. Bohra et al. (2012) reported mapping of QTLs for fertility restoration. However, in the case of SPC there are no reports on genetic mapping of QTLs. Examples, however, exist in other legume crops such as soybean and pea (Tar'an et al., 2004; Irzykowska and Wolko, 2004; Burstin et al., 2007; Qi et al., 2011; Soybase, 2015; Krajewski et al., 2012; Hwang et al., 2014; Wang et al., 2014; Zhang et al., 2015).

The first QTL detected for SPC in a legume crop was in soybean (Diers et al. 1992), and since then over 152 QTLs for SPC have been reported in soybean from about 30 different studies (Qi et al., 2011; Soybase, 2015; Hwang et al., 2014; Wang et al., 2014; Zhang et al., 2015). In pea, up to 31 QTLs have been reported (Tar'an et al., 2004; Irzykowska and Wolko, 2004; Burstin et al., 2007; Krajewski et al., 2012).

Different kinds of bi-parental populations, including F_2, $F_{2:3}$, $F_{2:5}$, F_4, F_5, F_6, BC_3F_4, and RILs from different crosses and association mapping panels have been used for SPC QTL detection in soybean and pea (Irzykowska and Wolko, 2004; Tar'an et al., 2004; Burstin et al., 2007; Qi et al., 2011; Krazewski et al., 2012; Wang et al., 2014; Zhang et al., 2015; Soybase, 2015; Hwang et al., 2014). The statistical methods used in detecting the soybean and pea SPC QTLs are also diverse including single marker ANOVA (SMA), interval mapping (IM), iterative QTL mapping (iQTLm), multiple interval mapping (MIM), composite interval mapping (CIM), and inclusive interval mapping (ICIM). (Irzykowska and Wolko, 2004; Tar'an et al., 2004; Burstin et al., 2007; Qi et al., 2011; Krazewski et al., 2012; Wang et al., 2014; Zhang et al., 2015; Soybase, 2015). In general, nearly all QTL mapping studies for SPC have incorporated mapping QTLs for other

important traits like seed yield and yield-related characters such as seed weight. Besides QTL localization, genomic resources presently available in pigeonpea, especially the draft genome sequence (Varshney et al., 2012) and whole genome resequenced data on parental lines (Kumar et al., 2016) can also facilitate the identification of candidate genes controlling important traits including SPC.

1.5.3.3 Candidate gene identification

Traditionally, candidate genes have been selected because they resemble genes associated with similar traits, or because the predicted protein function seems relevant to the physiology of the trait (functional candidate genes), or because a positional mapping approach pointed to these genes in a genomic region (positional candidate genes) (Pflieger et al., 2001; Gilissen et al., 2012; Eskandari et al., 2013; Mir et al., 2014). The traditional candidate gene approach is, however, limited by its dependence on the prior knowledge of physiological, biochemical and metabolic pathways, which is incomplete or sometimes completely unavailable (Hoehe et al., 2000; Zhu and Zhao, 2007).

Recent advances in NGS technologies have revolutionized the process of candidate identification in plants and animals through techniques such as whole-exome sequencing (WES) and whole genome resequencing (WGRS). Through WES and WGRS tens of thousands, if not, millions to billions of genomic variants can be identified in each exome or genome (Gilissen et al., 2012; Elkan-Miller and Avraham, 2013). In the presence of a draft genome sequence, WGRS approach on few individuals can be used to identify variants and the genes associated with the trait of interest. The success of such approach has been demonstrated for a number of traits in animals such as humans (Rios et al., 2010; Roach et al., 2010; Sobreiro et al., 2010) and chicken (Jang et al., 2014) and in crop species such as rice (Silva et al., 2012; Lim et al., 2014) and maize (Xu et al., 2014) but not for SPC in pigeonpea.

The major challenge to the use NGS-based techniques for candidate gene identification is the prioritization of putative variants from the thousands or millions of variations obtained (Gilissen

et al., 2012; Elkan-Miller and Avraham, 2013). Silva et al. (2012) developed two filtering strategies, namely, common variant (CV) and principal component-biplot (PB) prioritization strategies. They used the CV and PB prioritization procedures to identify non-synonymous (ns) SNPs and genes between two groups of known-resistant and known-susceptible rice inbred lines to sheath blight. They concluded that both prioritization strategies gave similar results. Since then, other research groups have adopted similar prioritization strategies. For example, Xu et al. (2014) used similar approaches to that of Silva et al. (2012) and successfully identified candidate genes for drought tolerance in maize.

A major assumption in filtering variants from NGS data for the purpose of candidate gene selection is that the causative variant likely leads to change on the protein level, so changes such as nonsense, missense, splicing, and frameshift variants are prioritized (Coonrod et al., 2013). Further prioritization may be based on information on gene function in relation to the phenotype (Gilissen et al., 2012).

Next generation sequencing produces short reads, which makes misalignments to the reference genome a more common occurrence (Church et al., 2011). Validation of variants identified from NGS-based approaches must therefore be done to determine analytical sensitivity and analytical specificity by comparing NGS test results to those obtained from independently validated method such as Sanger sequencing (Deschamps et al., 2010; Silva et al., 2012; Gilissen et al., 2012; Wong, 2013; Jang et al., 2014). The final testing of the role of a candidate gene can be carried out by conventional co-segregation analysis in structured population such as F_2, or by SNP-phenotype associations in germplasm collections or natural populations, or in functional experiments (Pflieger et al., 2001; Grattapaglia, 2008; Gilissen et al., 2012).

1.7 Summary of literature review

The variation in SPC and its relationships with other traits of importance are dependent upon genetic and environmental background, but with negligible genotype × environment interaction. Because of the effect of genotype and environment on SPC, it is essential to establish the level of

variation of the trait and its relationships with other traits of importance in genotypes of unknown SPC before including them in genetic studies or breeding programs. Only few studies have been published on the genetic control of SPC in pigeonpea. The few reports indicate that the trait is quantitative in nature with low to moderate heritability, is conditioned by a minimum of three to four genes with the non-additive gene action being more important than the additive component. No study, however, has reported on determining which of the components of the non-additive gene action, either dominance or epistasis, contributes more to protein accumulation in the seeds of pigeonpea. Two approaches that are robust for detecting both additive and non-additive components of genetic variation are GMA and QTL analysis. However, there is no report of the use of GMA and QTL analysis for dissecting genetic control of SPC in pigeonpea. Similarly, the availability of a pigeonpea reference genome sequence provides opportunity to generate high-density SNPs that could be used for QTL and candidate gene identification but no study has yet been reported on the use of such genomic techniques for dissecting the genetic architecture of SPC in the crop.

1.8 Prospects for future research

The study reported in this thesis used the classical quantitative genetics analysis in combination with genomics methods to gain insight into the genetic control of SPC in pigeonpea. While such approaches continue to be used and have led, and will continue to contribute, to the development of superior phenotypes through breeding, the structural variations detected at the genetic level are not always translated into the predicted phenotype, which leads to the so-called “missing heritability”. Also, mechanisms involved in seed storage protein accumulation can be complex due to involvements of multigene families, metabolites and post translational modifications for which classical quantitative genetics, genomics or transcriptomics have limitations in detecting. In this scenario, proteomics and metabolomics hold the promise to enhance the understanding of functional molecules on specific aspects of multigene families and post-translational modifications, instead of analysing only the genetic code or only the transcript abundance, which may not associate with their corresponding proteins. Opportunity also exists to take on a more systems biology approach to understanding seed behaviour in terms of seed development and seed

nutrient reserve accumulation, including SPC. Such approach may include the use of two or more combinations of the omics technologies.

References

Adu-Gyamfi, J.J., F.M. Myaka, W.D. Sakala, R. Odgaard, and H. Høgh-Jensen. 2007. Nitrogen and phosphorus budgets and nitrogen fixation in farmer-managed intercrops of maize-pigeonpea in semi-arid Southern and Eastern Africa. Plant and Soil 295:127-136.

Aruna, R., R.D. Manohar, S. Sivaramakrishnanan, R.L. Janardhan, B. Paula, and H. Upadhyaya. 2008. Efficiency of three DNA markers in revealing genetic variation among wild *Cajanus* species. Plant Genetic Resources: Characterization and Utilization 7:113-121.

Baskaran, K., and A.R. Muthiah. 2007. Studies on combining ability in pigeonpea [*Cajanus cajan* (L.) Millsp.] Legumes Research 30:67-69.

Baudoin, J.-P., and A. Maquet. 1999. Improvement of protein and amino acid contents in seeds of food legumes. A case study in Phaseolus. Biotechnology, Agronomy, Society and Environment 3:220–224.

Bliss, F.A., L.N. Barker, J.D. Franckowiak, and T.C. Hall. 1973. Genetic and environmental variation of seed yield, yield components, and seed protein quantity and quality of cowpea. Crop Science 13:656-660.

Bohra, A., A. Dubey, R.K. Saxena, R.V. Penmetsa, K.N. Poornima, N. Kumar, A.D. Farmer, G. Srivani, H.D. Upadhyaya, R. Gothalwal, S. Ramesh, D. Singh, K.B. Saxena, P.B. KaviKishor, N.K. Singh, C.D. Town, G.D. May, D.R. Cook, and R.K. Varshney. 2011. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea, (*Cajanus* spp.). BMC Plant Biology 11:56.

Bohra, A., R.K. Saxena, B.N. Gnanesh, K.B. Saxena, M. Byregowda, A. Rathore, P.B. KaviKishor, D. Cook, and R.K. Varshney. 2012. An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L) Millspaugh] derived from six mapping populations. Theoretical and Applied Genetics 125:1325-1338.

Brim, C.A., and J.W. Burton. 1978. Recurrent selection in soybeans. II Selection for increased percent protein in seeds. Crop Science 19:494-498.

Broman, K.W. 2001. Review of statistical methods for QTL mapping in experimental crosses. Lab Animal 307:44-52.

Burstin, J., Marget, P., M. Huart, A. Moessner, B. Mangin, C. Duchene, B. Desprez, N. Munier-Jolain, and G. Duc. 2007. Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiology 144:768-781.

Castle, W.E. 1921. An improved method of estimating number of genetic factors concerned in cases of blending inheritance. Proceedings of the National Academy of Sciences, USA 81:6904-6907.

Ceyhan, E., M. Harmankaya, and A. Kahraman. 2014. Combining ability and heterosis for concentration of mineral elements and protein in common bean (*Phaseolus vulgaris* L.). Turkish Journal of Agriculture and Forestry 38:581-590.

Church, D.M., V.A. Schneider, T. Graves, K. Auger, F. Cunningham, N. Bouk, H.-C. Chen, R. Agarwala, W.M. McLaren, G.R.S. Ritchie, D. Albracht, M. Kremitzki, S. Rock, H. Kotkiewicz, C. Kremitzki, A. Wollam, L. Trani, L. Fulton, R. Fulton, L. Matthews, S. Whitehead, W. Chow, J. Torrance, M. Dunn, G. Harden, G. Threadgold, J. Wood, J. Collins, Heath, P. Griffiths, S. Pelan, D. Grafham, E.E. Eichler, G. Weinstock, E.R. Mardis, R.K. Wilson, K. Howe, P. Flicek, and T. Hubbard. 2011. Modernizing reference genome assemblies. PLoS Biology 9: e1001091.

Cober, E.R., and H.D. Voldeng. 2000. Developing high-protein, high-yield soybean populations and lines. Crop Science 40:39-42.

Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196.

Coonrod, E.M., J.D. Durtschi, R.L. Margraf, and K.V. Voelkerding. 2013. Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. Archives of Pathology and Laboratory Medicine 137:415-433.

Cromwell E., E. Friis-Hansen, and M. Turner. 1992. The seed sector in developing countries: a framework for performance analysis. Working Paper 65. Overseas Development Institute, Reagent's College, Inner Circle, London NWI 4NS.

Dahiya, B.S., J.S. Brar, and B.S. Bhullar. 1977. Inheritance of protein content and its correlation with grain yield in pigeonpea (*Cajanus cajan* (L.) Millsp.). Qualitus Plantarium 27:327-334.

Deschamps, S., M. la Rota, J.P. Ratashak, P. Biddle, D. Thureen, A. Farmer, S. Luck, M. Beatty, N. Nagasawa, L. Michael, V. Llaca, H. Sakai, G. May, J. Lightner, and M.A. Campbell. 2010. Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina Genome Analyzer. The Plant Genome 3:53-68.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theoretical and Applied Genetics 83:608-612.

Dubey, A., A. Farmer, J. Schlueter, S.B. Cannon, B. Abernathy, R. Tuteja, J. Woodward, T. Shah, B. Mulasmanovic, H. Kudapa, NL. Raju, R. Gothalwal, S. Pande, Y. Xiao, C.D. Twon, N.K. Singh, G.D. May, S. Jackson, and R.K. Varshney. 2011. Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). DNA Research 18:153-164.

Elkan-Miller, T., and K.B. Avraham. 2013. Deafness gene discovery in the genomic era. In: Ahmed, Z., and S. Riazuddin, editors, Inner ear development and hearing loss. Nova Science Publishers, New York. p. 127-142.

Emebiri, L.C. 1991 Inheritance of protein content in seeds of selected crosses of cowpea (*Vigna unguiculata* (L.) Walp.). Journal of Science of Food and Agriculture 54:1-7.

Emefiene, M.E., V.I. Joshua, C. Nwadike, A.Y. Yaroson, and N.D.E. Zwalnan. 2014. Profitability analysis of Pigeonpea (*Cajanus cajan*) production in Riyom LGA of Plateau State. International Letters of Natural Sciences 18:73-88.

Eskandari, M., E.R .Cober, and I. Rajcan. 2013. Using the candidate gene approach for detecting genes underlying seed oil concentration and yield in soybean. Theoretical and Applied Genetics 126:1839-1850.

Falconer, D.S. 1989. Introduction to quantitative genetics, 3rd ed. Longman, London.

Fatokun, C.A., D. Menancio-Hautea, D. Danesh, and N.D. Young. 1992. Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132:841-846.

Filipetti, A., G.H. Azadegan, and C. de Pace. 1999. Breeding strategies for seed protein content and trypsin inhibitors inferred from combining ability and heterosis in test-crosses of *Vicia faba*. Plant Breeding 118:411-416.

Fortunato, R.H. 2000. Systematic relationships in *Rhynchosia* (Cajaninae-Phaseoleae-Papilionoideae-Fabaceae) from the neotropics. In: Herendeen, P.S., A. Bruneau, editors. Advances in legume systematics series No 9. The Royal Botanic Gardens, Kew, UK. 339–354.

Frascaroli, E., M.A. Cane`, P. Landi, G. Pea, L. Gianfranceschi, M. Villa, M. Morgante, and M.E. Pe` 2007. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. Genetics 176:625-644.

Geiger, H.H., and A. Tomerius. 1997. Quantitative genetics and optimum breeding plans. In: Krajewski, P., and Z. Kaczmarek, editors, Advances in biometrical genetics, Institute of Plant Genetics, Polish Academy of Sciences, Poland, p.15-26.

Gilissen, C., A. Hoischen, H.G. Brunner, and J.A. Veltman. 2012. Disease gene identification strategies for exome sequencing. European Journal of Human Genetics 20:490-497.

Gnanesh, B.N., A. Bohra, M. Sharma, M. Byregowda, S. Pande, V. Wesley, R.K. Saxena, K.B. Saxena, P.B. KaviKishor, and R.K. Varshney. 2011. Genetic mapping and quantitative trait locus analysis of resistance to sterility mosaic disease in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Field Crops Research 123:56-61.

Goodacre, R., S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, and D.B. Kell. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends in Biotechnology 22:245-524.

Grattapaglia, D. 2008. Genomics of Eucalyptus, a global tree for energy, paper and wood. In: Moore, P., and R. Ming, editors, Genomics of tropical crop plants. Springer Sciences. p. 257-295.

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative genetics in maize breeding. Iowa State University Press, Ames, IA.

Hamdi, A., W. Erskine, and P. Gates. 1991. Relationships among economic characters in lentil Euphytica 57:109-116.

Hazra, P., P.K. Das, and M.G. Som. 1996. Combining ability for pod yield and seed protein in cowpea (*Vigna unguiculata* (L) Walp.). Indian Journal of Genetics 56:553-555.

Hedley, C.L. 2001. Introduction. In: Carbohydrate in grain legume seeds. Hedley, C.L., editor. pp 1-11.

Hillocks, R.J., E. Minja, A. Mwaga, M.N. Silim, and P. Subrahmanyam. 2000. Diseases and pests of pigeonpea in eastern Africa: a review. International Journal of Pest Management 46:7-18.

Hoehe, M.R., K Kopke, B. Wendel, K. Rohde, C. Flachmeier, K.K. Kidd, W.H. Berrettini, and G.M. Church. 2000. Sequence variability and candidate gene analysis in complex disease: association of μ opioid receptor gene variation with substance dependence. Human Molecular Genetics 9:2895-2908

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martinez. 2003. Estimating and interpreting heritability for plant breeding: An update. Plant Breeding Reviews 22:9-112.

Horgan, R.P., and L.C. Kenny. 2011. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician and Gynaecologist 13:189-195.

Hwang, E.-Y., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, J. Costa, and P.B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15:1.

Hyten, D.L., I.Y. Choi, Q. Song, J.E. Specht, T.E. Carter, R.C.Jr. Shoemaker, E.Y. Hwang, L.K. Matukumalli, and P.B. Cregan. 2010. A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Science 50:960-968.

Iqbal, A.M., F.A. Nehvi, S.A. Wani, Z.A. Dar, A.A. Lone, and H. Qadri. 2012. Combining ability study over environments in dry beans (*Phaseolus vulgaris* L.). SAARC Journal of Agriculture 10:61-69.

Irzykowska, L., and B. Wolko. 2004. Interval mapping of QTLs controlling yield-related traits and seed protein content in *Pisum sativum*. Journal of Applied Genetics 45:297-306.

Jaganathan, D., M. Thudi, S. Kale, S. Azam, M. Roorkiwal, P. Gaur, P. KaviKishor, H. Nguyen, T. Sutton, and R.K. Varshney. 2015. Genotyping-by-sequencing based intra-specific genetic map refines a ''QTL hotspot" region for drought tolerance in chickpea. Molecular Genetics and Genomics 290:559-571.

Jang, H.-M., G.F. Erf, K.C. Rowland, and B.-W. Kong. 2014. Genome resequencing and bioinformatic analysis of SNP containing candidate genes in the autoimmune vitiligo Smyth line chicken model. BMC Genomics 15:707.

Jones, C.D. 2001. Extension of the Castle-Wright effective factor estimator to sex linkage and haplodiploidy. Journal of Heredity 92:2744-276.

Jorrín-Novo, J.V., A.M. Maldonado, S. Echevarría-Zomeño, L. Valledor, M.A. Castillejo, M. Curto, J. Valero, B. Sghaier, G. Donoso, I. Redondo. 2009. Plant proteomics update (2007-2008): second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. Journal of Proteomics 72:285-314.

Kaddurah-Daouk, R., B.S. Kristal, and R.M. Weinshilboum. 2008. Metabolomics: A global biochemical approach to drug response and disease. Annual Review of Pharmacology and Toxicology 48:653-683.

Kale, S.M., D. Jaganathan, P. Ruperao, C. Chen, R. Punna, H. Kudapa, M. Thudi, M. Roorkiwal, M.A.V.S.K. Katta, D. Doddamani, V. Garg, P.B. KaviKishor, P.M. Gaur, H.T. Nguyen, J. Batley, D. Edwards, T. Sutton, and R.K. Varshney. 2015. Prioritization of candidate genes in "QTL-hotspot" region for drought tolerance in chickpea (*Cicer arietinum* L.). Scientific Reports 5:15296.

Khoury, C.K., N.P. Castañeda-Alvarez, H.A. Achicanoy, C.C. Sosa, V. Bernau, M.T. Kassa, S.L. Norton, L.J.G. van der Maesen, H.D. Upadhyaya, J. Ramírez-Villegas, A. Jarvis, and P.C. Struik. 2015. Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Millsp.]: distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. Biological Conservation 184:259-270.

Krajewski, P., J. Bocianowski, M. Gawłowska, Z. Kaczmarek, T. Pniewski, W. Święcicki, and B. Wolko. 2012. QTL for yield components and protein content: a multi-environment study of two pea (*Pisum sativum* L.) populations. Euphytica 183:323-336.

Krishnan, H.B., S.S. Natarajan, N.W. Oehrle, W.M. Garrett, and O. Darwish. 2017. Proteomic analysis of pigeonpea (*Cajanus cajan*) seeds reveals the accumulation of numerous stress-related proteins. Journal of Agriculture and Food Chemistry 65:4572-4581.

Kudapa, H., A.K. Bharti, S.B. Cannon, A.D. Farmer, B. Mulaosmanovic, R. Kramer, A. Bohra, N.T. Weeks, J.A. Crow, R. Tuteja, T. Shah, S. Dutta, D.K. Gupta, A. Singh, K. Gaikwad, T.R. Sharma, G.D. May, N.K. Singh, and R.K. Varshney. 2012. A comprehensive transcriptome assembly of pigeonpea (*Cajanus cajan* L.) using Sanger and second-generation sequencing platforms. Molecular Plant 5:1020-1028.

Kumar, V., A.W. Khan, R.K. Saxena, V. Garg, and R.K. Varshney. 2016. First-generation HapMap in Cajanus spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnology Journal 14:1673-1181.

Kumawat, G., R.S. Raje, S. Bhutani, J.K. Pal, A.S. Mithra, K. Gaikwad, T.R. Sharma, and N.K. Singh. 2012. Molecular mapping of QTLs for plant type and earliness traits in pigeonpea (*Cajanus cajan* L. Millsp.). BMC Genetics 13:84.

Lakshmi M., P. Senthilkumar, M. Parani, M.N. Jithesh, and A.K. Parida. 2000. PCR-RFLP analysis of chloroplast gene regions in *Cajanus* (Leguminosae) and allied genera. Euphytica 116:243-250.

Lamkey, K.R., and J.W. Edwards. 1999. The quantitative genetics of heterosis. In: Coors, J.G., and S. Pandey, editors. The genetics and exploitation of heterosis in crops. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI. p. 31–48.

Lawn, R.J., and G.J. Rebetzke. 2006. Variation among Australian accessions of wild mung bean (*Vigna radiata* ssp. *sublobata*) for traits of agronomic, adaptive, or taxonomic interest. Australian Journal of Agricultural Research 57:119-132.

Layrisse, A., J.C. Wynne, and T.G. Isleib. 1980. Combining ability for yield, protein and oil of peanut lines from South American centres of diversity. Euphytica 29:561-570.

Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter., E.R. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey, and H.R. Boerma. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theoretical and Applied Genetics 93:649-657

Leleji, O.I., M.H. Dickson, L.V. Crowder, and J.B. Bourke. 1972. Inheritance of crude protein percentage and its correlation with seed yield in beans, *Phaseolus vulgaris L*. Crop Science 12:168-171.

Li, L., M. Hur, J.-Y. Lee, W. Zhou, Z. Song, N. Ransom, C.Y. Demirkale, D. Nettleton, M. Westgate, Z. Arendsee, V. Iyer, J. Shanks, B. Nikolau and E.S. Wurtele. 2015. A systems biology approach toward understanding seed composition in soybean. BMC Genomics 16:S9.

Lijavetzky, D., J.A. Cabezas, A. Ibanez, V. Rodriguez, and J.M. Martinez-Zapater. 2007. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics 8:424.

Lim, J.-H., H.-J. Yang, K.-H. Jung, S.-C. Yoo, and N.-C. Paek. 2014. Quantitative trait locus mapping and candidate gene analysis for plant architecture traits using whole genome re-sequencing in rice. Molecules and Cells 37:149-160.

Lohithaswa, H.C., S. Hittalmani, H.E. Shashidhar, and P.S. Dhanaraj. 2003. Assessment of genetic variability in some pigeonpea [*Cajanus cajan* (L.)Millsp.] genotypes using RAPD markers. Indian Journal of Genetics 63:329-330.

Malviya, N., and D. Yadav. 2010. RAPD analysis among pigeonpea [*Cajanus cajan* (L.) Millsp.] cultivars for their genetic diversity. Genetic Engineering and Biotechnology Journal 1:1-9.

Martre, P., J.R. Porter, P.D. Jamieson, and E. Triboï. 2003. Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. Plant Physiology 133:1959-1967.

Mebrahtu, T., and A. Mohamed. 2003. A seven-parental diallel analysis of nutritional composition of common beans. Plant Foods for Human Nutrition 58:1-11.

Mir, R.R., H. Kudapa, S. Srikanth, R.K. Saxena, A. Sharma, S. Azam, K.B. Saxena, R.V. Penmetsa, and R.K.Varshney. 2014. Candidate gene analysis for determinacy in pigeonpea (*Cajanus* spp.). Theoretical and Applied Genetics 127:2663-2678.

Nadimpalli, R.G., J.L. Jarret, S.C. Pathak, and G. Kochert. 1993. Phylogenetic relationships of pigeonpea (*Cajanus cajan*) based on nuclear restriction fragment length polymorphism. Genome 36: 216-223.

Nica, A.C., and E.T. Dermitzakis. 2013. Expression quantitative trait loci: present and future. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 368: 20120362.

Noubissié, J.B.T., J.M. Bell, and H.T. Yadji. 2012. Studies on variability and gene effects of seed protein contents in Cameroonian bean (*Phaseolus vulgaris*) cultivars. Journal of Agriculture and Social Sciences 8:17-23.

Odeny, D.A., B. Jayashree, M. Ferguson, D. Hoisington, L.J. Cry, and C. Gebhardt. 2007. Development, characterization and utilization of microsatellite markers in pigeonpea. Plant Breeding 126:130-136.

Oluwatosin, O.B. 1997. Genetic and environmental variation for seed yield, protein, lipid and amino acid composition in cowpea (*Vigna unguiculata* (L) Walp). Journal of the Science of Food and Agriculture 74:107-116.

Panguluri, S.K., J. Janaiah, J.N. Govil, P.A. Kumar, and P.C. Sharma. 2005. AFLP fingerprinting in pigeonpea (*Cajanus cajan* L. Millsp.) and its wild relatives. Genetic Resources and Crop Evolution 53,523–531.

Pazhamala, L., R.K. Saxena, V.K. Singh, C.V. Sameerkumar, V. Kumar, P. Sinha, K. Patel, J. Obala, S.R. Kaoneka, P. Tongoona, H.A. Shimelis, N.V.P.R. Gangarao, D. Odeny, A. Rathore, P.S. Dharmaraj, K.N. Yamini, and R.K .Varshney. 2015. Genomics-assisted breeding for boosting crop improvement in pigeonpea (*Cajanus cajan*). Frontiers in Plant Science 6:50.

Petricoin E., K. Zoon, E. Kohn, J. Barrett, and L. Liotta. 2002. Clinical proteomics: translating bench side promise into bedside reality. Nature Reviews 1:683-95.

Pflieger, S., V. Lefebvre, and M. Causse. 2001. The candidate gene approach in plant genetics: a review. Molecular Breeding 7:275-291.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253.

Purseglove, J.W. 1968. Tropical Crops: Dicotyledons, UK: Longman Group.

Qi, Z.-M., Y.-N. Sun, Q. Wu, C.-Y. Liu, G.-H. Hu, and Q.-S. Chen. 2011. A meta-analysis of seed protein concentration QTL in soybean. Canadian Journal of Plant Science 91:221-230.

Ramalingam, A., H. Kudapa, L.T. Pazhamala, W. Weckwerth, and R.K. Varshney. 2015. Proteomics and metabolomics: two emerging areas for legume improvement. Frontiers in Plant Science 6:1116.

Ratnaparkhe, M.B., V.S. Gupta, V.M.R. Murthy, and P.K. Ranjekar. 1995. Genetic fingerprinting of pigeonpea [*Cajanus cajan* (L.) Millsp.] and its wild relatives using RAPD markers. Theoretical and Applied Genetics 91:893-898.

Reddy, L.J., P. Bramel, K.B. Saxena, and R. Ortiz. 2000. Utilization of wild species in pigeonpea improvement at ICRISAT. Poster presented In: International conference on science and technology for managing plant genetic diversity in the 21st Century, Kuala Lumpur, Malaysia, 12-16 June, 2000.

Rekha, R., L. Prasanthi, M.R. Sekhar, and M.S. Priya. 2013. Studies on selection indices in pigeonpea [*Cajanus cajan* (L.) Millsp]. International Journal of Applied Biology and Pharmaceutical Technology 4:291-294.

Remanandan, P., and L. Singh. 1997. Pigeonpea. In: Fuccillo, D., L. Sears, and P. Stapleton, editors, Biodiversity in Trust: conservation and use of plant genetic resources in CGIAR centres. Cambridge University Press, New York. p. 156-167.

Remanandan, P.D., V.S.S.R. Sastry, and H.M. Mengesha. 1988. ICRISAT pigeonpea germplasm catalog: passport information. Andra Pradesh (India): International Crops Research Institute for the Semi-arid Tropics.

Rifai, N., M.A. Gillette, and S.A. Carr. 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nature Biotechnology 24:971-83.

Rios, J., E. Stein, J. Shendure, H.H. Hobbs, and J.C. Cohen. 2010. Identification by whole genome resequencing of gene defect responsible for severe hypercholesterolemia. Human Molecular Genetics 19:4313-4318.

Roach, J.C., G. Glusman, A.F.A. Smit, C.D. Huff, R. Hubley, R.T. Shannon, L. Rowen, K.P. Pant, N. Goodman, M. Bamshad, J. Shendure, R. Drmanac, L.B. Jorde, L. Hood, and D.J. Galas. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636-639.

Sahu, A.R., R.R. Mishra, R.S. Chandra, and J. Panigrahi. 2015. Construction of interspecific genetic linkage map of pigeonpea using SCoT, RAPD, ISSR markers and simple inherited trait loci. Nucleus 58:23-31.

Santos, C.A.F., D.C.C. da Costa, W.R. da Silva, and L.S. Boiteux. 2012. Genetic analysis of total seed protein content in two cowpea crosses. Crop Science 52:2501-2506.

Saxena, K.B. 2008. Genetic improvement of pigeonpea – a review. Tropical Plant Biology 1:159–178.

Saxena, K.B., and D. Sharma. 1990. Pigeonpea genetics. In: Nene YL, S.D. Hall, and V.K. Shiela, editors.The pigeonpea. CAB International UK. p. 137-157.

Saxena, K.B., D.G. Faris, U. Singh, and R.V. Kumar. 1987. Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition 36:335-340.

Saxena, K.B., R.V. Kumar, and P.V. Rao. 2002. Pigeonpea nutrition and its improvement. In: A. Basra, S., and I. S. Randhawa, editors, Quality improvement in field crops. Food Products Press. p. 227-260.

Saxena, R.K., C. Prathima, D.A. Hoisington, N.K. Singh, and R.K. Varshney. 2010. Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). Plant Breeding 129:142-148.

Saxena, R.K., J. Obala, A. Sinjushin, C.V. Sameer-Kumar, K.B. Saxena, and R.K. Varshney. 2017b. Characterization and mapping of Dt1 locus which co-segregates with CcTFL1 for growth habit in pigeonpea. Theoretical and Applied Genetics DOI 10.1007/s00122-017-2924-2.

Saxena, R.K., R.V. Penmetsa, H.D. Upadhyaya, A. Kumar, N. Carrasquilla-Garcia, J.A. Schlueter, A. Farmer, A.M. Whaley, B.K. Sarma, G.D. May, D.R. Cook, and R.K. Varshney. 2012. Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeon pea and comparative mapping in legumes. DNA Research 19:449-461.

Saxena, R.K., V.K. Singh, S.M. Kale, R. Tathineni, S. Parupalli, V. Kumar, V. Garg, R.R. Das, M. Sharma, K.N. Yamini, S. Muniswamy, A. Ghanta, A. Rathore, C.V. Sameer-Kumar, K.B. Saxena, P.B.K Kishor, and R.K. Varshney. 2017a. Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for fusarium wilt resistance in pigeonpea. Scientific Reports 7: 1911.

Saxena, R.K., X. Cui, V. Thakur, B. Walter, T.J. Close, and R.K. Varshney. 2011. Single feature polymorphisms (SFPs) for drought tolerance in pigeonpea (*Cajanus* spp.). Functional and Integrative Genomics 11:651-657.

Sheth, B.P., and V. S. Thaker. Plant systems biology: insights, advances and challenges. Planta 240:33-4.

Silva, J., B. Scheffler, Y. Sanabria, C. de Guzman, D. Galam, A. Farmer, J. Woodward, G. May, and J. Oard. 2012. Identification of candidate genes in rice for resistance to sheath blight disease by whole genome sequencing. Theoretical and Applied Genetics 124:63-74.

Singh, U., and R. Jambunathan. 1982. Distribution of seed protein fractions and amino acids in different anatomical parts of chickpea (*Cicer arietinum* L.) and pigeonpea (*Cajanus cajan* L.). Plant Foods for Human Nutrition 31:347-354.

Singh, U., R. Jambunathan, K.B. Saxena, and N. Subramaniam. 1990. Nutritional quality evaluation of newly developed high protein genotypes of pigeonpea (*Cajanus cajan* L.). Journal of the Science of Food and Agriculture 50:201-209.

Singh, V.K., A.W. Khan, R.K. Saxena, V. Kumar, S.M. Kale, P. Sinha, A. Chitikineni, L.T. Pazhamala, V. Garg, M. Sharma, C.V.S. Kumar, S. Parupalli, S. Vechalapu, S. Patil, S. Muniswamy, A. Ghanta, K.N. Yamini, P.S. Dharmaraj, and R.K. Varshney. 2015. Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). Plant Biotechnology Journal 14:1183-1194.

Sivaramakrishnan S., K. Seetha, A.N. Rao, and L. Singh. 1997. RFLP analysis of cytoplasmic male sterile lines in pigeonpea (*Cajanus cajan* L. Millsp.). Euphytica 126:293-299.

Sivaramakrishnan, S., K. Seetha, and L.J. Reddy. 2002. Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. Euphytica 125:21-28.

Snapp, S.S., R.B. Jones, E.M. Minja, J. Rusike, and S.N. Silim. 2003. Pigeonpea for Africa: a versatile vegetable - and more. HortScience 38:1073-1079.

Sobreira, N.L.M., E.T. Cirulli, D. Avramopoulos, E. Wohler, G.L. Oswald, E.L. Stevens, D. Ge, K.V. Shianna, J.P. Smith, J.M. Maia, C.E. Gumbs, J. Pevsner, G. Thomas, D. Valle, J.E. Hoover-Fong, and D.B. Goldstein. 2010. Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genetics 6:e1000991.

Soybase. 2015. Soybean breeder's toolbox genetic map information. [Online]. Available at: http://www.soybase.org/search/index.php?search=true&result=qtl&qtl=Seed+protein (Accessed on April 20, 2016).

Tar'an, B., T. Warkentin, D.J. Somers, D. Miranda, A. Vandenberg, S. Blade, S. Woods, and D. Bing. 2004. Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.). Euphytica 136:297-306.

Tchiagam J.-B.N., J.M. Bell, A.M. Nassourou, N.Y. Njintang, and E. Youmbi. 2011. Genetic analysis of seed proteins contents in cowpea (*Vigna unguiculata* L. Walp.). African Journal of Biotechnology 10:3077-3086.

Theodorescu, D., and H. Mischak. 2007. Mass spectrometry based proteomics in urine biomarker discovery. World Journal of Urology 25:435–43.

Thu, T.T., T.T.X. Mai, E. Dewaele, S. Farsi, Y. Tadesse, G. Angenon, and M. Jacobs. 2003. In vitro regeneration and transformation of pigeonpea (*Cajanus cajan* (L.) Millsp.). Molecular Breeding 11: 159-168.

Tindall, H.D. 1988. Vegetables in the Tropics. UK: Macmillan International edition.

Tiwari, D.S., V. Singh, and P.S. Shukla. 1993. Combining ability in mung bean [*Vigna radiata* (L.) Wilczek]. Indian Journal of Genetics 53:395-398.

Toubiana, D., and A. Fait. 2012. Metabolomics-assisted crop breeding towards improvement in seed quality and yield. In: Agrawal, G.K., and R. Rwakal, editors. Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield. Springer, pp 453-475.

Upadhyaya, H.D., and K.N. Reddy, D.V.S.S.R. Sastry, C.L.L. Gowda. 2007. Identification of photoperiod insensitive sources in the world collection of pigeonpea at ICRISAT. Journal of Semi-Arid Tropical Agricultural Research 3:46-49.

Upadhyaya, H.D., K.N. Reddy, S. Shivali, R.K. Varshney, R. Bhattacharjee, S. Singh, and C.L.L. Gowda. 2011. Pigeonpea composite collection for enhanced utilization of germplasm in crop improvement programs. Plant Genetic Resources 9:97-108.

Urbanczyk-Wochniak, E., A. Luedemann, J. Kopka, J. Selbig, T. Roessner-Tunali, L. Willmitzer, and A.R. Fernie. 2003. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Reports 4:989–993.

Vaghela, K.O., R.T. Desai, J.R. Nizama, J.D. Patel, and V. Sharma. 2009. Combining ability analysis in pigeonpea [C*ajanus cajan* (L.) Millsp.]. Legume Research 32:274-277.

van der Maesen, L.J.G. 1980. India is the native home of the pigeonpea. Liber gratulatorius in honorem HCD de Wit. Misc. Paper 19. Wageningen, the Netherlands. pp 257–262.

Varshney, R.K., R.K. Saxena, H.D. Upadhyaya, A.W. Khan, Y. Yu, C. Kim, A. Rathore, D. Kim, J. Kim, S. An, V. Kumar, G. Anuradha, K.N. Yamini, W. Zhang, S. Muniswamy, J. Kim, R.V. Penmetsa, E. von Wettberg, and S.K. Datta. 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nature Genetics 49:1082–1088.

Varshney, R.K., R.V. Penmetsa, S. Dutta, P.L. Kulwal, R.K. Saxena, S. Datta, T.R. Sharma, B. Rosen, N. Carrasquilla-Garcia, A.D. Farmer, A. Dubey, K.B. Saxena, J. Gao, B. Fakrudin, M.N. Singh, B.P. Singh, K.B. Wanjari, M. Yuan, R.K. Srivastava, A. Kilian, H.D. Upadhyaya, N. Mallikarjuna, C.D. Town, G.E. Bruening, G. He, G.D. May, R. McCombie, S.A. Jackson, N.K. Singh, and D.R. Cook. 2010. Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). Molecular Breeding 26:393-408.

Varshney, R.K., S.M. Mohan, P.M. Gaur, N.V.P.R. Gangarao, M.K. Pandey, A. Bohra, S.L. Sawargaonkar, A. Chitikineni, P.K. Kimurto, P. Janila, K.B. Saxena, A. Fikre, M. Sharma, A. Rathore, A. Pratap, S. Tripathi, S. Datta, S.K. Chaturvedi, N. Mallikarjuna, A.G. Anuradha, A. Babbar, A.K. Choudhary, M.B. Mhase, C. Bharadwaj, D.M. Mannur, P.N. Harer, B. Guo, X. Liang, N. Nadarajan, and C.L.L. Gowda. 2013. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnology Advances 31:1120-1134.

Varshney, R.K., W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T.A. Donoghue, S. Azam, G. Fan, A.M. Whaley, A.D. Farmer, J. Sheridan, A. Iwata, R. Tuteja, R.V. Penmetsa, W. Wu, H.D. Upadhyaya, S. Yang, T. Shah, K.B. Saxena, T. Michael, W.R. McCombie, B. Yang, G. Zhang, H. Yang, J. Wang, C. Spillane, D.R. Cook, G.D. May, X. Xu, and S.A. Jackson. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology 30:83-89.

Verdier, J., K. Kakar, K. Gallardo, C. Le-Signor, G. Aubert, A. Schlereth, C.D. Town, M.K. Udvardi, and R.D. Thompson. 2008. Gene expression profiling of *M. truncatula* transcription

factors identifies putative regulators of grain legume seed filling. Plant Molecular Biology 67: 567-580.

Vigeolas, H., C. Chinoy, E. Zuther, B. Blessington, P. Geigenberger, and C. Domoney. 2008. Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and albumin 2 in developing pea seeds. Plant Physiology 146 74–82.

Vlahou, A., and M. Fountoulakis. 2005. Proteomic approaches in the search for disease biomarkers. Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences 814:11–19.

Wang, X., G.-L. Jiang, M. Green, R.A. Scott, Q. Song, D.L. Hyten, and P.B. Cregan. 2014. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. Molecular Genetics and Genomics 289:935-949.

Wasike, S., P. Okori, and P.R. Rubaihayo. 2005. Genetic variability and relatedness of the Asian and African pigeonpea as revealed by AFLP. African Journal of Biotechnology 4:1228-1233

Weigelt K, H. Küster, T. Rutten, A. Fait, A.R. Fernie, O. Miersch, S. Wasternack, R.J.N. Emery, C. Desel, F. Hosein, M. Müller, I. Saalbach, and H. Weber. 2009. ADP-glucose pyrophosphorylase-deficient pea embryos reveal specific transcriptional and metabolic changes of carbon-nitrogen metabolism and stress responses. Plant Physiology 149:395-411.

Wiggins, B.T. 2012. Heritability and genetic gain of seed protein, oil, and yield among RIL of soybean. MSc. thesis, University of Tennessee.

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Science 35:1036-1041.

Wong, L.-J.C. 2013. Challenges of bringing next generation sequencing technologies to clinical molecular diagnostic laboratories. Neurotherapeutics 10:262-272.

Wu, R., H.D. Bradshaw, and R.F. Stettler. 1997. Molecular genetics of growth and development of populus (Salicaceae). V. Mapping quantitative trait loci affecting leaf variation. American Journal of Botany 84:143-153.

Xu, J., Y. Yuan, Y. Xu, G. Zhang, X. Guo, F. Wu, Q. Wang, T. Rong, G. Pan, M. Cao, Q. Tang, S. Gao, Y. Liu, J. Wang, H. Lan, and Y. Lu. 2014. Identification of candidate genes for drought tolerance by whole-genome re-sequencing in maize. BMC Plant Biology 14:83.

Yang, S., R.K. Saxena, P.L. Kulwal, G.J. Ash, A. Dubey, J.D. Harper, H.D. Upadhyaya, R. Gothalwal, A. Kilian, and R.K. Varshney. 2011. First genetic map of pigeonpea based on diversity array technology (DArT) markers. Journal of Genetics 90:103-109.

Yang, S., W. Pang, J. Harper, J. Carling, P. Wenzl, E. Huttner, X .Zong, and A. Kilian. 2006. Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology (DArT). Theoretical and Applied Genetics 113:585-595.

Young, N.D., J. Mudge, and T.N. Ellis. 2003. Legume genomes: more than peas in a pod. Current Opinion in Plant Biology 6:199-204.

Zhang, Y.H., M.F. Liu, J.B. He, Y.F. Wang, G.N. Xing, Y. Li, S.P. Yang, T.J. Zhao, and J.Y. Gai. 2015. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.]. Theoretical and Applied Genetics 128:1061-1072.

Zhu, M., and S. Zhao. 2007. Candidate gene identification approach: Progress and challenges. International Journal Biological Science 3:420-427.

Chapter Two

Variation in seed protein content and its relationship with some agronomic traits in pigeonpea

Abstract

Seed protein content (SPC) is an important trait of cultivated pigeonpea. This study was conducted to determine variation in SPC and its relationships with some agronomic traits in pigeonpea. Twenty three pigeonpea genotypes, which are parents of different types of mapping populations under development, were evaluated under field conditions in 2014-2015 growing season at ICRISAT, Patancheru, India. The experiment was carried out in a randomised complete block design with two replications. Data were recorded on SPC, number days to first flowering (DTF), plant height at maturity (PH), number of pods per plant (NPP), number of seeds per pod (NSP), 100-seed weight (SW) and seed yield per plant (SY). Seed protein content ranged from 19.3 to 25.5%, DTF (48 to 156 days), PH (67.5 to 230 cm), NPP (31.7 to 582 pods), NSP (2.9 to 4.6 seeds/pod), SW (6.2 to 20.8 g) and SY (7.9 to 333.4 g). There were significant differences among genotypes for all traits. Broad-sense heritability was 0.693 for SPC and ranged from 0.517 to 0.999 among the agronomic traits. Genetic advance (GA) was 2.4 % for SPC but ranged from 1.2 % to 141. % among the agronomic traits. Genetic gain, which is GA expressed as a percentage of the trait's grand mean, was 11.0 % for SPC but ranged from 56.4 to 713.4 % among the agronomic traits. Both favourable and unfavourable relationships exist between SPC and some of the agronomic traits with result showing strong negative relation of SPC with NPP and SW, which indicates that simultaneous selection for both high NPP and heavier seeds, or both NPP and high SY would lead to reduction in total SPC. However, simultaneous selection for high SY and high SPC, or for both high SW and high SY could result in increased SPC. It is therefore concluded that adequate variability for SPC exists among the parents of pigeonpea mapping populations being developed at ICRISAT that can be used for genetic studies including identification of marker-trait associations, and for breeding purposes.

2.1 Introduction

Earlier studies reported presence of variation for SPC in the cultivated pigeonpea (Salunkhe et al., 1986; Upadhyaya et al., 2007; Reddy et al., 1997; Saxena et al., 2002). Similarly, relationship of SPC with other traits of agronomic importance have been reported in the pigeonpea (Dahiya and Brar, 1976; Dahiya et al., 1977; Saxena et al., 1987; Rekha et al., 2013) but with mixed results. It is also known that a trait's variates including variance, heritability, and its relationships with other traits can change depending upon the set of materials evaluated and the environment in which they are studied (Hamdi et al., 1993; Wray and Visscher, 2008). This warrants continuous assessment of germplasms for variation in the targeted trait and its relationship with other traits, before drawing conclusions and applying specific materials in genetic improvement programs.

For studying trait variations, measures such as phenotypic and genotypic coefficients of variation (PCV and GCV, respectively) are often used in addition to estimates of heritability and genetic advance. On the other hand simple correlation and path coefficient analyses are used for studying interrelationships among traits. Simple correlation indicates how change in the variance of one trait affects the change in the variance of the other trait regardless of cause and effect relationship. Unlike simple correlation, path coefficient analysis helps to measure the direct effect of one trait on another by separating correlation coefficient into direct and indirect components, which enables detection of the most influential traits.

At ICRISAT, Patancheru, India, a number of different types of mapping populations including nested association mapping (NAM), multiparent advanced generation intercross (MAGIC), recombinant inbred lines (RIL) and introgression lines (IL) are being developed in pigeonpea for the identification of QTLs/molecular markers for various traits. These populations can also be used for dissecting genetic control of SPC, including marker-SPC associations, or even to directly select for lines with improved levels of the trait. However, the variability of SPC and its relationships with important agronomic traits among the parental lines are not known. The objectives of this study were to (i) characterize variability for SPC among 23 pigeonpea genotypes, and (ii)

determine interrelationships of SPC with some agronomic traits including seed yield, seed weight, pod characters, plant height and number of days to flowering.

2.2 Materials and methods

2.2.1 Plant material and data collection

The pigeonpea genotypes that were used in this study are presented in Table 1. Twenty one of these genotypes are parents of different types of mapping populations being developed at ICRISAT under the United States Agency for International Development (USAID) funded project "Pigeonpea Improvement Using Molecular Breeding". The remaining two genotypes, namely HPL 31 and HPL 26 are known high SPC breeding lines developed from the cross between *Cajanus cajan* variety 'Baigani' and *C. scarabaioedes* (a wild relative of pigeonpea) accessions. All the 23 genotypes were, at the initiation of the present study, preliminarily assessed to determine the level of SPC variation among them. To validate the level of variation obtained in the preliminary assessment and to determine relationship of SPC with other traits, all the 23 genotypes were planted under field conditions in 2014-2015 growing season.

To avoid insect pollinators, the materials were grown under nylon bee screens. A randomized complete block design in two replications. Each genotype was planted in a single 4 m long row with inter- and intra-row spacing of 75 cm and 30 cm, respectively. Agronomic practices included application of 100 kg/ha of diammonium phosphate as basal fertilizer without any top dressing, 2 and 4 L/ha of pendimethalin and paraquat dichloride pre-emergence herbicides, respectively, provision of two irrigations, one each at planting and pod filling stages, and two weedings one each at early vegetative and podding stages. Pod borers (*Maruca vitrata* Fab. and *Helicoverpa armigera* Hub.) were controlled by spraying with acephate and spinosad insecticides at rates of 1.0 kg/ha and 0.2 L/ha, respectively at 15 days intervals from flowering to podding stages. At maturity individual pods from individual plants were carefully hand-harvested leaving out plants at the beginning and at the end of each row and those at the field borders to avoid border effects.

Sun drying was done for one week before threshing and another one week after threshing to ensure uniform reduction in seed moisture content. All data were recorded on three plants per genotype.

To estimate SPC, 10 g of mature dry clean seeds of each plant were analysed at the Central Analytical Services (Charles Renard Analytical) Laboratory at ICRISAT, India. Before grinding, seeds were oven-dried at 60ºC for 48 hours. The dried seed samples were ground into powder in a mill with Teflon chambers. The ground samples were again kept in an oven at 60ºC overnight. Samples and appropriate blanks were digested simultaneously in duplicate (i.e. two independent analyses) using tri-acid digestion procedure as described in Upadhyaya et al. (2016).

Briefly, 1.0 g of ground sample was transferred to a 75 ml digestion tube containing 10 ml of tri-acid mixture of nitric, sulfuric and perchloric acids in the ratio of 10:0.5:2 (v/v). The contents were left overnight in the digestion chamber for cold digestion. In order to obtain clear and colourless digests, samples were initially digested at 120ºC for 1 hour followed by digestion at 230ºC for approximately 2 hours. After digests were cooled, the contents were dissolved in distilled water and volume made up to 75 ml and then mixed well by shaking. Aliquots were obtained from the digests and used to estimate the total nitrogen (N) using a San++Automated Wet Chemistry Analyzer (Skalar, Breda, The Netherlands). Seed protein of a sample was estimated by multiplying its N (%) content by factor 6.25. Besides SPC, data were also collected on number of days to first flowering (DTF), plant height (PH), number of pods per plant (NPP), number of seeds per pod, 100 seed weight (SW) and seed yield (SY) per plant. The DTF was scored daily as described in Craufurd et al. (2001). Plant height was recorded as height in cm from the base to the tip of the plant. NPP and NSP were recorded as counts of number of pods on a plant and number of seeds per plant, respectively. Hundred seed weight was recorded as weight of 100 dry, clean and healthy seeds in g, and SY was obtained by weighing all seeds from a plant.

Table 2.1: Features of pigeonpea genotypes evaluated for seed protein content and some agronomic characters

Accession	Features	Source population
HPL 28	High seed protein content breeding line	-†
HPL 31	High seed protein content breeding line	-
ICPL 87119 (Asha)	Genome sequence available, leading variety, resistant to Fusarium wilt (FW) and sterility mosaic disease (SMD)	IL‡, NAM§
ICP 7426	High pod numbers, medium duration	MAGIC¶
HPL 24	High protein content, medium duration, compact, susceptible to FW and resistant to SMD, inter-specific derivative	MAGIC, NAM
ICP 11605	Early flowering, germplasm line	MAGIC
ICP 14209	High number of pods, germplasm line	MAGIC
ICP 14486	Early flowering, germplasm line	MAGIC
ICP 5529	Medium duration, obcordate leaves, compact plant, poor yielding, modified flower	MAGIC
ICP 7035	Medium duration, SMD resistant to both Patancheru and Bangalore races, large purple seed, high sugar	MAGIC, NAM
ICP 8863	Erect, mid-late, highly resistant to FW and susceptible to SMD, red seeded genotype	MAGIC, NAM, RIL#
ICPL 87	Early duration, determinate, short, high combiner	NAM
ICPL 88039	Extra early maturity, indeterminate, good yield	NAM
ICP 85063 (Lakshmi)	Medium duration, indeterminate, good yield, more branching	NAM
MN-1	Super early, small seeded, determinate	NAM
ICP 28	Early maturity, local varieties	NAM
ICP 85010 (Sarita)	Early maturity, local varieties	NAM
UQ 50	Determinate, long podded, white seeded	NAM
ICPL 20096	Resistant to FW and SMD	RIL
ICPL 20097	Resistant to both SMD and FW	RIL
ICPL 332	Tolerant to pod borer, high yielding.	NAM
ICPB 2049	Susceptible to FW	RIL
ICPL 99050	Resistant to FW	NAM

† Not a parent in any population; ‡ Introgression line; § Nested association mapping population; ¶ Multiparent advanced generation intercross; # Pigeonpea recombinant inbred line population.

2.2.1 Data analysis

2.2.1.1 Genotypic and phenotypic variation

All statistical analyses were performed using SAS statistical software v9.4 (SAS Institute, 2015). Analysis of variance was carried out, and means were separated using Least Significance Difference (LSD) at 5%. Genotypic and phenotypic coefficients of variation were calculated as described in Singh and Chaudhary (1979) as follows:
$PCV(\%) = \left(\sqrt{\sigma^2 P}/\mu\right)/\times 100$, and $GCV(\%) = \left(\sqrt{\sigma^2 G}/\mu\right)/\times 100$, where PCV and GCV are the phenotypic and genotypic coefficients of variation, respectively, and $\sigma^2 P$ and $\sigma^2 G$ are the phenotypic and genotypic variances, respectively. Phenotypic and genotypic coefficients of variations were categorized as low (<10%), moderate (10-20%), and high (>20%) (Subramanian and Menon, 1973).

Broad-sense heritability (H^2) was estimated using the formula: $H^2 = (\sigma^2 G/\sigma^2 P)$, where $\sigma^2 G$ and $\sigma^2 P$ are genotypic and phenotypic variances respectively. The heritability was placed into three categories of low (0.0-0.3), moderate (0.3-0.6) and high (>0.6) (Johnson et al., 1955).

Genetic advance (GA) was obtained as: $GA = H^2 \times \sqrt{\sigma^2 P} \times K$, where H^2 is the broad-sense heritability, $\sigma^2 P$ is the phenotypic standard deviation and K is the selection differential (2.06 at 5%). Genetic advance was converted to percent genetic gain as: $Genetic\ gain = GA \times 100$, and categorized as low (0-10 %), moderate (10-20%) and high (>20%) (Johnson et al., 1955).

2.2.1.2 Genetic correlation and path coefficient analyses

Genotypic correlations were calculated according to Falconer and Mackay (1996) using the formula: $rG = \sigma Gxy/\left(\sqrt{\sigma^2 Gx \times \sigma^2 Gy}\right)$, where σGxy is genotypic covariance and $\sigma^2 Gx$ and $\sigma^2 Gy$ are genotypic variances of trait *x* and trait *y*, respectively.

Direct and indirect path coefficients were calculated using genotypic correlation coefficients following methods of Wright (1921). For path analysis, SPC was considered as a response variable and DTF, PH, NPP, NSP, SW and SY as causal variables.

2.3 Results

2.3.1 Performance of genotypes for seed protein content and agronomic traits

Mean square for each of the seven studied traits are presented in Table 2.2. Highly significant ($p \leq 0.01$) differences existed among the 23 genotypes for SPC and all measured agronomic traits, and therefore genotype means were compared to determine differences (Table 2.2).

Table 2.2: Analysis of variance for seed protein content and six agronomic traits in 23 pigeonpea genotypes

	Mean square	
Trait	Genotype (DF§ = 22)	Error (DF = 22)
Seed protein content (%)	4.9 ***	0.9
Number of days to first flowering	1754.6 ***	0.1
Plant height (cm)	5151.2 ***	10.4
Number of pods per plant	13782.0 ***	4359.0
Number of seeds per pod	1.1 ***	0.2
100-seed weight (g)	18.8 ***	0.6
Seed yield (g)	9534.2 ***	59.2

† Coefficient of variability. § Degrees of freedom.

Mean SPC in the present study ranged from 19.3% (ICPL 87119) to 25.5 % (HPL 31) with an overall mean value of 22.1% (Table 2.3). Of the 23 genotypes, HPL 24, ICP 14486, ICP 5529, HPL 28 and HPL 31 recorded relatively high SPC while genotypes ICPL 87, ICPL 20097, ICPL 85063, ICP 99050 and ICPL 87119 recorded low SPC in that order (Table 2.3). For the agronomic traits, DTF ranged from 48.0 days (MN 1) to 156.0 days with a mean of 100.0 days (ICPL 332), PH ranged from 67.5 cm (MN 1) to 230.0 cm (ICPL 20097) with an average of 179.7 cm (Table 2.3). Number of pods per plant ranged from 31.7 (MN 1) to 582.3 (HPL 24) with a mean of 229.1

while NSP ranged from 2.9 to 4.6 with a mean of 3.5 (Table 2.3). Hundred seed weight ranged from 6.2 g/plant (ICP 7426) to 20.8 g/plant (ICP 7035) with a mean of 10.1 g/plant while SY varied from 7.9 g/plant (MN 1) to 333.4 g/plant (ICP 7035) with a mean of 61.2 g/plant (Table 2.3). The relatively low CV values across traits (Table 2.3) is expected because the genotypes used in the study are highly inbred landraces or breeding lines.

Table 2.3: Means for seed protein content and six agronomic characters studied in 23 genotypes of pigeonpea

Genotype	SPC§ (%)	DTF‡	PH ¶ (cm)	NPP#	NSP††	SW §§ (g)	SY ‡‡ (g)
HPL 31	25.5 a	100.0 m	188.3 d	167.8 hi	2.95 ef	9.9 ghijk	38.6 fghi
HPL 28	25.2 a	101.0 l	191.7 d	310.5 de	3.51 cd	10.0 ghijk	41.2 gfh
ICP 5529	24.6 ab	103.7 j	210.0 cb	152.8 hi	4.37 a	8.6 klm	23.3 ijkl
ICP 14486	24.1 ab	86.0 p	133.3 f	31.7 l	4.10 ab	8.6 klm	10.2 lm
ICP 14209	23.1 bcd	138.0 c	208.0 c	212.2 fgh	3.57 bcd	8.7 jklm	17.3 klm
HPL 24	23.0 bcd	111.8 g	208.3 c	582.3 a	2.93 ef	8.1 lm	152.7 b
ICP 8863	22.3 cde	90.2 o	210.0 cb	124.8 ijk	3.47 cdef	9.9 hijk	24.5 ijkl
ICPL 85010	22.2 cde	50.3 u	82.5 i	76.5 jkl	3.41 cdef	9.0 jkl	18.0 klm
ICPL 88039	22.2 cde	60.5 t	149.2 e	178.8 ghi	3.62 bc	11.5 defgh	54.3 ef
MN 1	22.2 cde	48.0 v	67.5 j	37.7 l	3.02 def	7.2 mn	7.9 m
ICP 7426	22.1 def	120.0 f	205.0 c	363.3 d	3.48 cde	6.2 n	70.2 ed
ICPL 20096	22.1 def	131.3 d	215.8 b	162.0 hi	3.22 cdef	13.1 cd	48.2 fg
UQ 50	21.9 def	106.8 h	204.2 c	352.5 d	3.57 bcd	13.6 c	127.4 c
ICP 28	21.6 def	79.8 q	128.3 f	132.5 ij	2.91 f	8.6 klm	26.2 hijk
ICP 11605	21.5 def	66.0 s	93.3 h	53.2 kl	3.38 cdef	12.2 cde	22.5 jklm
ICP 7035	21.3 def	129.0 e	226.7 a	517.7 ab	4.60 a	20.8 a	333.4 a
ICPL 332	21.3 def	156.0 a	228.3 a	124.2 ijk	3.03 def	15.6 b	36.9 ghij
ICPB 2049	20.8 efg	102.0 k	206.7 c	131.8 ij	3.37 cdef	9.9 ghijk	37.4 ghij
ICPL 87	20.8 efg	68.7 r	116.7 g	219.8 fgh	3.77 bc	11.1 efghi	69.2 ed
ICPL 20097	20.7 efg	151.3 b	230.0 a	251.3 efg	3.35 cdef	11.8 def	78.0 d
ICPL 85063	20.4 efg	92.5 n	228.3 a	375.0 cd	3.38 cdef	9.6 ijkl	67.3 ed
ICPL 99050	20.2 fg	104.8 i	210.0 cb	264.2 ef	3.35 cdef	10.3 fghij	64.1 ed
ICPL 87119	19.3 g	103.2 j	191.7 d	445.7 bc	3.75 bc	11.5 defg	38.5 fghi
Grand mean	22.1	100.0	179.7	229.1	3.50	10.7	61.2
S.e.m†	0.5	0.5	3.1	26.9	0.2	0.4	8.8
Range	19.3-25.5	48.0-156.0	67.5-230.0	31.7-582.3	2.9-4.6	6.2-20.8	7.9-333.4
CV (%)	4.3	1.1	4.8	28.8	13.9	7.8	12.6
LSD 5%	1.4	1.3	8.5	75.5	0.55	1.1	24.6

Means followed by the same letter are not significantly different at $P \leq 0.05$. † Standard error of the mean; § Seed protein content; ‡ Number of days to first flowering; ¶ Plant height; # Number of pods per plant; †† Number of seeds per pod; §§ 100-seed weight; ‡‡ Seed yield.

2.3.2 Heritability, genotypic and phenotypic coefficients of variation and genetic gain

Whereas the mean, range and CV can suggest the extent to which improvement can be made for a given trait they, however, depict nothing about effect of genotype on trait variation. Hence, in the present study, parameters such as genotypic, environmental and phenotypic variances, genotypic and phenotypic coefficients of variation, heritability, genetic advance and genetic gain were estimated (Table 2.4).

Table 2.4: Estimates of broad-sense heritability, genotypic and phenotypic coefficients of variation, and genetic gain for seven traits in 23 pigeonpea genotypes

Trait	σ^2G †	σ^2E §	σ^2P ‡	H^2¶	GCV# (%)	PCV †† (%)	GA‡‡ (%)	GG§§ (%)
DTF	877.25	0.07	877.32	0.999	29.6	29.6	61.0	60.9
PH	2570.40	10.44	2580.84	0.996	28.2	28.3	104.2	58.2
SW	9.10	0.61	9.71	0.937	28.3	29.2	6.0	56.4
SY	4737.52	59.20	4796.72	0.988	112.5	113.2	141.0	230.4
SPC	2.01	0.89	2.90	0.693	6.4	7.7	2.4	11.0
NPP	4711.50	4359.00	9070.50	0.519	30.0	41.6	101.8	713.4
NSP	0.50	0.20	0.70	0.712	19.2	23.6	1.2	65.5

DTF, Number of days to first flowering; PH, Plant height; SW, Hundred seed weight; SY, Seed yield; SPC, Seed protein content; NPP, Number of pods per plant; NSP, Number of seeds per pod; † Genetic variance; § Environmental variance; ‡ Phenotypic variance; ¶ Broad-sense heritability; # Genotypic coefficient of variability; †† Phenotypic coefficient of variability; ‡‡ Genetic advance; §§ Genetic genetic gain.

In general, σ^2G and GCV were always close to σ^2P and PCV, respectively, with σ^2G always larger than σ^2E for all traits. This was also consistent with the generally high H^2 ranging from 0.519 for NPP to 0.999 for DTF (Table 2.4). There were small differences between PCV and GCV values for SPC and most of the other traits except NPP (Table 2.4). Although SPC showed high H^2 estimate (>0.60), the GCV and genetic advance (GA) were low resulting in a relatively low genetic gain estimate for the trait. High H^2 with high GCV and high or moderate GA estimates for DTF, PH, SW, SY, NPP and NSP resulted in >50 % genetic gain (Table 2.4).

2.3.3 Relationships of total SPC with agronomic characters

Results of simple genotypic correlations between SPC and agronomic traits are presented in Table 2.5. Generally, SPC had negative correlations with all traits although significant only with SW (Table 2.5).

Table 2.5: Genotypic correlation coefficients for pair-wise association of SPC with agronomic traits

	PH†	NPP§	NSP¶	SW#	SY‡	SPC††
DTF	0.85***	0.41**	0.06^{NS}	0.39**	0.33*	-0.07^{NS}
PH		0.56***	0.15^{NS}	0.32*	0.38**	-0.07^{NS}
NPP			0.20^{NS}	0.27^{NS}	0.73***	-0.27^{NS}
NSP				0.43***	0.46**	0.00^{NS}
SW					0.68***	-0.30*
SY						-0.20^{NS}

+, *, **, and *** significant at the 0.1, 0.05, 0.01, 0.001 probability levels, respectively; NS, not significant at 0.05 probability level. † Number of days to first flowering; § Plant height; ¶ Number of pods per plant; # 100-seed weight; ‡ Seed yield; †† Seed protein content.

On the basis of the path coefficient analysis results (Table 2.6), all values of direct effects were below one, showing that increments resulting from multi-collinearity were marginal. The values of direct path coefficient were relatively large and negative between SPC and NPP (-0.73) and SPC and SW (-0.68). It was positive and large between SPC and SY (0.63) but small between SPC and DTF (0.08), SPC and PH, and SPC and NSP. Indirect effects of agronomic traits on SPC were large and negative for NPP via SY, but positive for SW also via SY.

Table 2.6: Direct (boldfaced main diagonals) and alternate/indirect path coefficient values of seed protein content against agronomic traits of pigeonpea

	Trait					
	DTF†	PH§	NPP¶	NSP#	SW‡	SY§§
DTF	**0.08**	0.19	-0.30	0.01	-0.27	0.21
PH	0.07	**0.23**	-0.41	0.02	-0.22	0.24
NPP	0.03	0.13	**-0.73**	0.02	-0.18	0.46
NSP	0.00	0.04	-0.15	**0.11**	-0.29	0.29
SW	0.03	0.07	-0.20	0.05	**-0.68**	0.43
SY	0.03	0.09	-0.53	0.05	-0.47	**0.63**

†Number of days to first flowering, § Plant height; ¶ Number of pods per plant; # Number of seeds per pod; ‡ 100-seed weight; §§ Seed yield.

2.4 Discussion

The knowledge of genetic variation for a trait and trait correlations are important components of any breeding objective. Seed protein content in pigeonpea is an important grain quality trait, and it impacts the nutritional importance of pigeonpea in the human diet. The range of SPC values obtained in the present study is within 12.0 to 29.0 % reported earlier among 1,974 germplasm genotypes at ICRISAT (Remanandan et al., 1988). It is also close to 15.9 to 24.1% reported recently among 310 germplasm collection from different altitudes of Kenya (Upadhyaya et al., 2007). Among the genotypes tested in this study, interspecific derivatives (HPL 24, HPL 28, HPL 31) from the cross between wild (*C. scarabaoiedes*) and cultivated (*C. cajan*) pigeonpea (Saxena et al., 2002) showed the highest SPC. This suggests that the wild genotypes or their interspecific progenies could provide the needed source of high SPC genes for trait improvement, but the use of the wild relatives is associated with undesirable agronomic characters due to linkage drag (Saxena, 2008). Landrace cultivars that showed comparable level of SPC included ICP 5529 (24.6%) and ICP 14486 (24.1%), and they are equally potential sources of desirable genes for improving SPC.

The significant differences among pigeonpea genotypes in the present study indicates presence of variability for all traits measured. This is supported by the generally high H^2 indicating influence

of genetic factors on phenotype. Whereas heritability estimates can be used to predict the reliability of the phenotypic value as a guide to breeding value (Falconer and Mackay, 1996), heritability alone does not reveal the extent of response to selection. Broad-sense heritability along with GCV and GA provide reliable estimates of the amount of genetic gain to be expected through phenotypic selection (Burton, 1952). The combination of high H^2, GCV, GA and genetic gain (%) for DTF, PH, SY and NPP indicates that the variation in these traits is largely due to genetic factors, and selection would be effective for these traits. However, SPC as a core trait in this study had high H^2 but low GCV and low genetic gain estimates, depicting a low response to selection. Similarly, SW and NSP with high and moderate H^2, respectively, had low genetic gain values indicative of a poor response to selection. Given the poor predicted response to selection based on SPC alone, determining the relationship of SPC with agronomic traits could provide an indication of which of the agronomic traits could be used to indirectly select for improved SPC. It could also pinpoint which of the agronomic traits affect SPC either positively or negatively, which in turn could help in deciding on appropriate selection/breeding strategy.

Few studies have been conducted on the relationships of SPC with agronomic traits in pigeonpea. Results of simple genotypic correlations in the present study indicated that SW was the major trait that negatively influenced SPC in the set of genotypes tested. This observation is in agreement with that of earlier studies in pigeonpea (Saxena et al. 1987), soybean (Filho et al. 2001), mung bean (Afzal et al. 2003), and cowpea (Asante et al. 2004) who reported significant negative correlations between SPC and SW.

If only simple genotypic correlations were considered in the present study, SW would be the only agronomic trait that influences SPC, but negatively, in the set of pigeonpea genotypes tested. However, path-coefficient analysis allocated the strongest negative direct effects on SPC to NPP and SW indicating that selection for increased NPP or SW would lead to reduced SPC. On the other hand the strong positive direct effect due to SY indicates that simultaneous selection for high SPC and high SY is possible, and is in agreement with conclusions from previous studies that selection for high SPC does not always lead to SY reduction in the grain legumes (Leleji et al., 1972; Brim and Burton, 1978; Wilcox and Cavins, 1995). Similarly, through path coefficient

analysis, a large negative indirect effect of NPP on SPC via SY was detected indicating that simultaneous selection for high NPP and SY would lead to reduced SPC. In a similar manner, SW had a large positive indirect effect on SPC also via SY indicating that simultaneous selection for increased SW and SY would lead to increased SPC.

Because relationships among traits is dependent upon the set of materials evaluated and the environment in which they are studied (Hamdi et al., 1991), future re-evaluation of the 23 and other potentially useful genotypes for SPC and agronomic traits in multiple sets of environments may be necessary.

2.5 Conclusions

There is variation for SPC among the pigeonpea genotypes used as parents of the mapping populations at ICRISAT although no large differences were detected, which is a possible reflection of the low genetic diversity that has repeatedly been reported within the cultivated pigeonpea gene pool. Although the H^2 and GCV for SPC were large, the genetic advance estimate was low resulting in low expected genetic gain. Nonetheless there is possibility of generating desirable recombinants through biparental mating. Both favourable and unfavourable relationships exist between SPC and some of the agronomic traits with strong negative relationships of SPC with NPP and SW, which indicates that simultaneous selection for both high NPP and heavier seeds, or both NPP and high SY would lead to reduction in total SPC. However, simultaneous selection for high SY and high SPC, or for both high SW and high SY could result in increased SPC. An understanding of the genetic basis of the observed variation in SPC and its relationships with agronomic traits will facilitate the designing of efficient breeding strategies for improving SPC while maintaining other desirable agronomic attributes such SY and SW in pigeonpea.

References

Afzal, M.A., M.M. Bashir, N.K. Luna, M.A. Bakr, and M.M. Haque. 2003. Relationship between seed size, protein content and cooking time of mung bean [*Vigna radiata* (L.) Wilczek] Seeds. Asian Journal of Plant Science 2:1008-1009.

Asante, I.K., H. Adu-Dapaah, and P. Addison. 2004. Seed weight and protein and tannin contents of 32 cowpea genotypes in Ghana. Tropical Science 44:77-19.

Brim, C.A., and J.W. Burton. 1978. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Science 19:494-498.

Burton, G.W. 1952. Quantitative inheritance in grasses. Proceeding of the 6th International Grassland Congress 1:277-283.

Dahiya, B.S., and J.S Brar. 1976. The relationship between seed size and protein content in pigeonpea (*Cajanus cajan* (L.) Millsp.) Tropical Grain Legume Bulletin. 3:18.

Dahiya, B.S., J.S. Brar, and B.S. Bhullar. 1977. Inheritance of protein content and its correlation with grain yield in pigeonpea (*Cajanus cajan* (L.) MilIsp.). Qualitas Plantarum Plant Foods for Human Nutrition 27:327-334.

Craufurd, P.Q., H.S. Soko, H.K. Jones, and R.J. Summerfield. 2001. Inheritance of duration from sowing to first flowering in pigeonpea. Euphytica 119:323-333.

Falconer, D.S., and F.C. Mackey. 1996. Introduction to Quantitative Genetics, 4th ed. Longman, London.

FAOSTAT (2015) http://faostat3.fao.org. Accessed 20 April 2016.

Filho, M.M., D. Destro, L.A. Miranda, W.A. Spinosa, M.C. Carrão-Panizzi, and R. Montalván. 2001. Relationships among oil content, protein content and seed size in soybeans. Brazilian Archives of Biology and Technology 44:23-32.

Hamdi, A., W. Erskine, and P. Gates. 1991. Relationships among economic characters in lentil. Euphytica 57:109-116.

Iqbal, A., I.A. Khalil, N. Ateeq, and M.S. Khan. 2006. Nutritional quality of important food legumes. Food Chemistry 97:331–335.

Johnson, H.W., H.E. Robinson, and R.E. Comstock. 1955. Estimates of genetic and environmental variability in soybean. Agronomy Journal 47:314-318.

Leleji, O.I., M.H. Dickson, L.V. Crowder, and J.B. Bourke. 1972. Inheritance of crude protein percentage and its correlation with seed yield in beans, *Phaseolus vulgaris L*. Crop Scieince 12:168-171.

Reddy, L.J., K.B. Saxena, K.C. Jain, U. Singh, J.M. Green, D. Sharma, D.G. Faris, A.N. Rao, R.V. Kumar, and Y.L. Nene. 1997. Registration of high-protein pigeonpea elite germplasm ICPL 87162. Crop Science 37:294.

Rekha, R., L. Prasanthi, M.R. Sekhar, and M.S. Priya. 2013. Studies on selection indices in pigeonpea [*Cajanus cajan* (L.) Millsp]. International Journal of Applied Biology and Pharmaceutical Technology 4:291-294.

Remanandan, P., D.V.S.S.R. Sastry, and M.H. Mengesa. 1988. ICRISAT pigeonpea germplasm catalog: evaluation and analysis. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, A.P. 502324, India.

Salunkhe, D.K., J.K. Chavan, and S.S. Kadam. 1986. Pigeonpea as important food source. Critical Reviews in Food Science and Nutrition 23:103-141.

Samonte, S.O.P.B., L.T. Wilson, and A.M. McClung. 1998. Path analysis of yield and yield related traits of fifteen diverse rice genotypes. Crop Science 38:1130-1136.

SAS Institute (2015) User's Guide: SAS/STAT® 14.1, SAS Institute, Cary, NC.

Saxena, K.B. 2008. Genetic improvement of pigeonpea-a review. Tropical Plant Biology 1:159-178.

Saxena, K.B., D.G. Faris, U. Singh, and R.V. Kumar. 1987. Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition 36: 335-340.

Saxena, K.B., R.V. Kumar, and P.V. Rao. 2002. Pigeonpea nutrition and its improvement. In: A.S. Basra, and I.S. Randhawa, editors. Quality Improvement in Field Crops. Food Products Press, p. 227-260.

Saxena, R.K., E. von Wettberg, H.D. Upadhyaya, V. Sanchez, S. Songok, K.B. Saxena, P. Kimurto, R.K. Varshney. 2014. Genetic diversity and demographic history of *Cajanus* spp. illustrated from genome-wide SNPs. PLoS ONE 9:e88568.

Singh, R.K., and B.D. Chaudhary. 1979. Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi.

Subramanian, S., and M. Menon. 1973. Heterosis and inbreeding depression in rice. The Madras Agricultural Journal 60:1139.

Upadhyaya, H.D., D. Bajaj, L. Narnoliya, S. Das, V. Kumar, C.L.L. Gowda, S. Sharma, A.K. Tyagi, and S.K. Parida. 2016. Genome-wide scans for delineation of candidate genes regulating seed protein content in chickpea. Frontiers in Plant Science 7:302.

Upadhyaya, H.D., K.N. Reddy, C.L.L. Gowda, and S.N. Silim. 2007. Patterns of diversity in pigeonpea (*Cajanus cajan* (L.) Millsp.) germplasm collected from different elevations in Kenya. Genetic Resources and Crop Evolution 54:1787-1795.

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Science 35:1036-1041.

Williams, W.A., M.B. Fones and M.W. Demment. 1990. A concise table for path analysis statistics. Agronomy Journal 82:1022-1024.

Wright, S. 1921. Correlation and causation. Journal of Agricultural Research. 20:557-585

Chapter Three

Inheritance of seed protein content and its association with seed weight and yield in pigeonpea

Abstract

Pigeonpea is an important source of dietary protein and it is widely consumed in the tropics and sub-tropics. To investigate inheritance of seed protein content (SPC) in pigeonpea, four elite germplasm lines with varying SPC were used to develop three crosses. Each cross consisted of six generations (P_1, P_2, F_1, F_2, BC_1P_1 and BC_1P_2). Generation mean analysis revealed the importance of dominance and epistatic effects for SPC. Duplicate and negative additive × additive epistasis were predominant and associated with transgressive segregants for SPC. Additive genetic variance component was higher than the environmental and dominance components. Broad-sense heritability ranged from 0.52 to 0.60. Predicted genetic gain after one cycle of selection was highest at 5% selection intensity. Seed weight and yield were positively and negatively correlated with SPC, respectively. Careful selection of parents, and reciprocal recurrent selection could be effective for improving SPC in pigeonpea.

3.1 Introduction

Undernutrition kills, disables or prevents millions of children from reaching their full intellectual and productive potential (Morris et al., 2008). One of the major forms of undernutrition is the inadequate intake of dietary protein. In communities where intake of animal protein is difficult or not affordable, crops, especially food legumes provide the bulk of needed dietary protein (Santos et al., 2012). Pigeonpea (*Cajanus cajan* (L.) Millsp.) is one of the major legume crops cultivated as a source of food protein for over a billion people in the developing world especially in the semi-arid tropics of Africa and Southern Asia (Mula and Saxena, 2010). The area cultivated with pigeonpea continues to increase annually and this can be attributed to its drought tolerance and ability to give relatively better yields in marginal soils than any other cultivated food legume

(Akibode and Maredia, 2011). Pigeonpea is cultivated for diversity of uses but in its major areas of cultivation its core importance is as a source of food protein.

Differences among genotypes for SPC have been reported in cultivated pigeonpea (Remanandan et al., 1988; Upadhyaya et al., 2007), which suggests there is potential for genetic improvement through hybridisation and selection. Genetic studies of SPC (Saxena et al., 2002; Baskaran and Muthiah, 2007; Vaghela et al., 2009) as well as correlations of SPC with important traits such as seed weight (Saxena et al., 1987) and seed yield (Dahiya et al., 1977) in pigeonpea have been reported. However, the available information is not only limited but also does not give a clear picture of the genetic control of SPC nor its relationship with important agronomic traits in the crop. The scarce information impedes the effective use of the available genetic variability for improvement of the trait.

The study of quantitative traits in plants can be accomplished using specific techniques such as generation mean analysis (GMA) and partitioning of genetic variance components (Mather and Jinks, 1974). Variance component analysis is based on mating schemes such as diallel, North Carolina and line × tester designs, which provide information about combining ability of the parental lines. However, because the parent populations used in such designs are usually selected for some desirable characteristics, the parents do not represent a random population resulting in biased estimates of variances and heritability (Araújo et al., 2005). Generation mean analysis combines the study of population means and variances (Mather and Jinks, 1974) and is believed to be more robust than the genetic variance component analysis (Bernado, 2010). Generation mean analysis also allows simultaneous detection of additive-dominance, and epistatic effects (Hallauer and Miranda, 1988), which is essential in designing appropriate breeding strategies.

Estimation of genetic effects and trait correlations in breeding populations are essential for designing selection strategies for the development of cultivars with improved SPC, acceptable seed weight and high seed yield. In this context, this study was conducted with the objective of investigating the inheritance pattern for SPC and its relationship with seed weight and seed yield in segregating populations of pigeonpea.

3.2 Materials and methods

3.2.1 Plant materials

Four cultivars, which included ICP 8863, ICP 14209, ICP 11605 and ICPL 87119 were selected on the basis of their SPC and diverse genetic background. ICP 8863 is a selection from landrace ICP 7626 (P-15-3-3) from Maharashtra, India (ICRISAT, 1993a). It is high yielding (1.5 t/ha) with 100-seed weight of 9.5 g and matures in 150 - 180 days. ICP 8863 has moderate SPC of 22.0%. ICP 11605 (ICPL 151) was selected from the cross ICP 6997 × Prabhat. It is a determinate cultivar, yielding 1.03 t/ha with 100-seed weight of 10 g and matures in 120-130 days (ICRISAT, 1993b; Remanandan and Singh, 1997) and has a low SPC of 20.9%. ICP 14209 is a landrace from India with moderate SPC (23.0%) and 100 seed weight of 5.5 g. ICPL 87119 was developed at ICRISAT from the cross ICP 1-6-W3–Wl × C 11. It matures in 160 – 202 days, yields ~1.5 t/ha with 100-seed weight of 10.2 g and it is widely adapted (ICRISAT, 1993c). It is low in SPC (19.3 %).

Pure seeds of the parental genotypes were obtained from ICRISAT's Genebank and used to develop three crosses: ICP 11605 × ICP 14209, ICP 8863 × ICP 11605 and ICP 8863 × ICPL 87119, hereafter referred to as Cross 1, Cross 2 and Cross 3, respectively. F_1 seeds of the three crosses were generated in 2012 rainy season. In the subsequent rainy season of 2013, the F_1 populations were grown and selfed to generate F_2 seed. The F_1 plants were also crossed to the parents to generate backcross one to parent one (BC_1P_1) and backcross one to parent two (BC_1P_2) with the parents as the seed plants. Additional F_1 seeds were also generated. At most two F_1 and two parental plants were used to generate the crosses. The hybridization was done manually after emasculating the unopened floral buds.

3.2.2 Field trials

All six generations (P_1, P_2, F_1, F_2, BC_1P_1 and BC_1P_2) in each cross were sown in the field at ICRISAT, Patancheru, India (545 meters above sea level, 17°32'N and 78°16'E). Preparation of

land and control of pests were carried out following standard agronomic practices for pigeonpea to ensure suitable conditions for plant growth and development. To avoid insect pollinators, the materials were grown under nylon bee screens. Trial design was a randomised complete block design with two replications. Sowing was done in 4 m long rows with inter- and intra-row spacing of 75 cm and 30 cm, respectively. Plot sizes in the experiment ranged from single row for F_1 to 13 rows for F_2. This gave rise to the variable population size in different entries which ranged from 13 (F_1) to 140 (F_2) plants/plot. At maturity individual plants were hand-harvested. To ensure proper representation of genotypes between the non-segregating and the segregating generations, three plants per replication for each of the P_1, P_2 and F_1 generations, 20 plants for each of the BC_1P_1 and BC_1P_2 generations and 100 to 130 F_2 plants were harvested. Pods from individual plants were carefully harvested leaving out plants at the beginning and at the end of each row and those at the field borders to avoid border effects. Sun drying was done for one week before threshing and another one week after threshing to ensure uniform seed moisture content.

3.2.3 Data collection

The total number of plants evaluated for SPC, SW and SY ranged from six in F_1 of Cross 3 to 253 F_2 in the same cross. To estimate SPC, ten grams of mature dry clean seeds per plant in each of the three crosses were analysed at the Central Analytical Services (Charles Renard Analytical) Laboratory at ICRISAT, India. Before grinding, seeds were oven-dried at 60ºC for 48 hours. The dried seed samples were ground into powder in a mill with Teflon chambers. The ground samples were again kept in an oven at 60ºC overnight. Samples and appropriate blanks were digested simultaneously in duplicate (i.e. two independent analyses) using the tri-acid procedure as described in Upadhyaya et al. (2016). Briefly, one g of ground sample was transferred to a 75 ml digestion tube containing 10 ml of tri-acid mixture of nitric, sulfuric and perchloric acids in the ratio of 10:0.5:2 (v/v). The contents were left overnight in the digestion chamber for cold digestion. Samples were initially digested at 120ºC for 1 hour followed by digestion at 230ºC for approximately 2 hours in order to obtain clear and colourless digests. After digests were cooled, the contents were dissolved in distilled water and volume made up to 75 ml and then mixed well

by shaking. Aliquots were obtained from the digests and used to estimate the total nitrogen (N) using a San++Automated Wet Chemistry Analyzer (Skalar, Breda, The Netherlands). SPC of a sample was estimated by multiplying its N (%) content by factor 6.25. Besides SPC, data were also recorded for 100-seed weight (SW) and seed yield (SY) in grams per plant.

3.2.4 Data analysis

Frequency distributions were constructed using F_2 SPC data. Proportions of transgressive segregants in each cross were obtained as the number of F_2 plants whose SPC fall outside the range of either parents. All other analyses were carried out using SAS v9.4 (SAS Institute, 2015). The General Linear Model procedure was used to obtain the best linear unbiased estimates of generation means. The analysis also included testing for significance of two a priori linear contrast parameters: (i) deviation of F_1 mean from the MPV as a measure of mid-parent heterosis (Holland, 2001), and (ii) deviation of F_2 mean from the average of F_1 and MPV as a measure of the overall effect of epistasis (Fenster and Galloway, 2000). The means for the six generations were separated using Fisher's Protected Least Significance Difference (LSD) at 5% probability. ABCD scaling test was performed to assess the adequacy of additive-dominance model in explaining the observed phenotypic variation (Pooni et al., 1987). Where the additive-dominance model was inadequate to explain the observed variation, the additive ([a]), dominance ([d]) and their interactions [aa], [ad] and [dd] were estimated using a six-parameter mean separation analysis procedure (Hayman, 1958). Three selection intensities of 5, 10 and 20% in order of increasing stringency were used to predict genetic gain for SPC from one cycle of selection using the model described by Hallauer and Miranda (1988). Phenotypic, environmental and genotypic correlations were calculated as described in Searle (1961) using SAS CORR procedure.

3.3 Results

3.3.1 Frequency distributions, treatment means and variances

Frequency distributions of F_2 SPC data (Fig. 3.1) were continuous in all three crosses, suggesting the involvement of a number of genes, each with a little or minor effect. Transgressive segregation in the F_2 was observed in all three crosses (Fig. 3.1; Table 3.1).

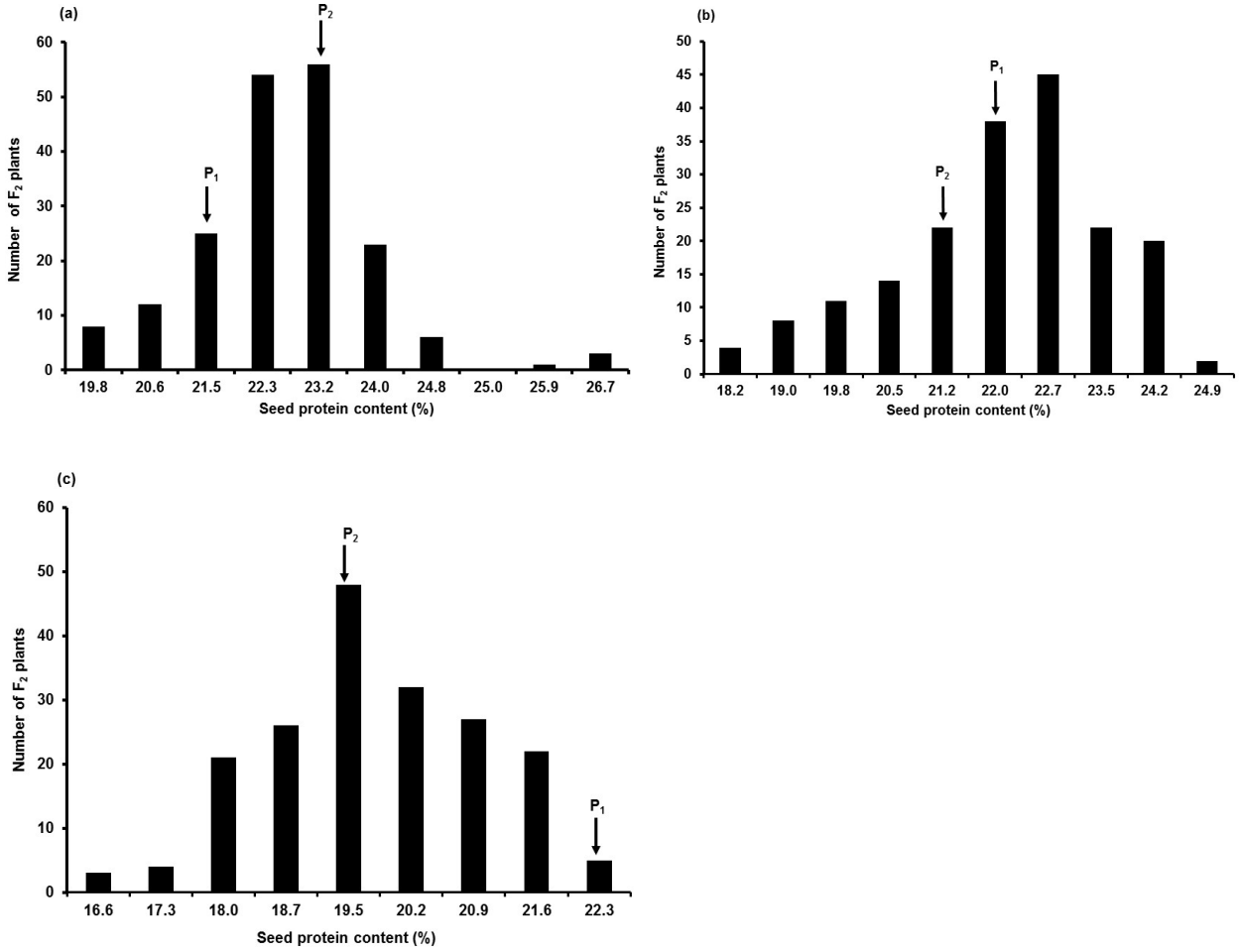


Fig. 3.1: F_2 frequency distribution for seed protein content in three crosses of pigeonpea
(a) Cross 1, ICP 11605 (P_1) × ICP 14209 (P_2) with P_1 and P_2 being low and high seed protein content parents, respectively; **(b)** Cross 2, ICP 8863 (P_1) × ICP 11605 (P_2) with P_1 and P_2 being high and low seed protein content parents, respectively; **(c)** Cross 3, ICP 8863 (P_1) × ICPL 87119 (P_2) with P_1 and P_2 being high and low SPC parents, respectively

Table 3.1: Generation, sample size, least squares mean, variance, deviation of F_1 from mid-parent value (MPV), deviation of F_2 from average of F_1 and MPV and percentage of transgressive segregants for seed protein content in three crosses of pigeonpea

	Cross 1			Cross 2			Cross 3		
Generation	n	LsMean	σ^2	n	LsMean	σ^2	n	LsMean	σ^2
P_1	18	21.42 bc	0.73	18	22.21 a	0.44	18	21.98 c	0.50
P_2	18	23.04 a	0.84	18	20.86 bc	0.65	18	19.34 a	0.07
F_1	12	21.60 bc	0.69	12	20.11 cd	1.90	6	20.86 b	1.08
F_2	237	21.97 b	1.86	236	21.36 b	2.73	253	19.58 a	1.42
BC_1P_1	40	20.99 c	1.76	40	20.23 cd	0.95	40	19.92 a	0.45
BC_1P_2	40	21.35 c	0.95	40	19.90 d	2.47	39	19.71 a	0.78
MP§		22.23			21.54			20.66	
Deviations									
F_1-MPV		-0.63			-1.43^{***}			0.20	
$F_2 – [(F_1+MPV)/2]$		0.05			0.53^{*}			-1.17^{**}	
Transgressive segregants (%)									
Lower than low parent		32.1			35.6			41.9	
Higher than high parent		20.7			36.0			1.2	
Pooled		52.7			71.6			43.1	

Means followed by the same letter within a column are not significantly different at 0.05 probability level; *, ** and *** significant at 0.05, 0.01 and 0.001 probability levels, respectively. Cross 1, ICP 11605 (P_1) × ICP 14209 (P_2) with P_1 and P_2 being low and high seed protein content parents, respectively; Cross 2, ICP 8863 (P_1) × ICP 11605 (P_2) where P_1 and P_2 are high and low seed protein content parents, respectively; Cross 3, ICP 8863 (P_1) × ICPL 87119 (P_2) with P_1 and P_2 being high and low seed protein content parents, respectively. n, population size; LsMean, least squares mean; σ^2, variance.

Proportions of transgressive segregants with SPC higher than that of the high parent were 20.7%, 36.0%, and 1.2% in Cross 1, Cross 2 and Cross 3, respectively (Table 3.1). Pooled over both extremes in each cross, Cross 1 and Cross 2 had over 50% of the F_2 individuals being transgressive, suggesting genes combined from the parents in each cross conferred both low and high SPC. Strong transgression (41.9%) only towards low SPC in Cross 3 indicated genes combined from the two parents resulted in progenies with lower SPC than that of the low SPC parent, a possible effect of accumulation of genes for low SPC from the two parents.

Parents of each of the three crosses significantly differed in SPC from each other (Table 3.1). Significant difference between parents of a cross is a pre-requisite for accurate determination of genetic effects controlling a trait (Mather and Jinks, 1982). Among the parental lines, ICP 14209 had the highest SPC (23.04 %) followed by ICP 8863 (22.09 %), ICP 11605 (21.14%) and ICP 87119 (19.3%) (Table 3.1). Generally, non-segregating parental and F_1 generations had lower variances than the segregating F_2 and backcross generations, suggesting more or less similar environmental influence on SPC accumulation in all generations.

Deviation of F_1 from MPV represents a measure of mid-parent heterosis (Holland, 2001). Hence, Cross 1 and Cross 2 had negative heterosis though non-significant in Cross 1 but highly significant in Cross 2 ($P \leq 0.01$) (Table 3.1). Cross 2 had positive but non-significant ($P > 0.05$) heterosis. Similarly, deviation of the F_2 mean from the average of MPV and F_1 represents a measure of the effect of epistasis, in which case F_2 mean greater or less than the average of MPV and F_1 indicates that genes combined and interacting for a given trait have either a positive or a negative effect on trait expression (Fenster and Galloway, 2000). The deviation of F_2 mean from average of F_1 and MPV was positive but not significant ($P < 0.05$) in Cross 1 and positive but significant ($P = 0.05$) in Cross 2. Cross 3 had a negative, highly significant ($P \leq 0.01$) deviation of the F_2 from the average of MPV and F_1 (Table 3.1). Therefore, Cross 1 and Cross 2 had more favourable combination of genes from the parents for SPC than Cross 3.

3.3.2 Genetic effects

The ABCD scaling test showed significance ($P = 0.05$) of at least one of the individual A, B, C, and D scales (Table 3.2) indicating the inadequacy of the three-parameter model in explaining the observed variations for SPC (Mather, 1949 and Pooni et al., 1987). Thus, Hayman's (1958) six parameter model was used to estimate the main and epistatic gene effects (Table 3.3).

Table 3.2: ABCD scaling test for seed protein content in three crosses of pigeonpea

Cross	Scale			
	A	B	C	D
1	$-1.05^{NS} \pm 0.87$	$-1.94^{*} \pm 0.82$	$0.21^{NS} \pm 1.44$	$1.60^{**} \pm 0.61$
2	$-1.86^{NS} \pm 1.43$	$-1.18^{NS} \pm 1.53$	$2.06^{NS} \pm 3.12$	$2.55^{*} \pm 1.13$
3	$-3.0^{**} \pm 0.77$	$-0.81^{NS} \pm 1.27$	$-4.70^{**} \pm 1.48$	$-0.44^{NS} \pm 0.54$

NS, not significant at 0.05 probability level; *, ** and *** significant at 0.05, 0.01 and 0.001 probability levels, respectively. Cross 1, ICP 11605 × ICP 14209; Cross 2, ICP 8863 × ICP 11605; Cross 3, ICP 8863 × ICPL 87119.

Table 3.3: Hayman's main and epistatic gene effect estimates, priori linear contrasts and proportion of transgressive segregation for seed protein content in three crosses of pigeonpea

Gene model	Cross 1	Cross 2	Cross 3
[m]†	$21.97^{***} \pm 0.09$	$21.36^{***} \pm 0.11$	$19.58^{***} \pm 0.07$
[a]§	$-0.37^{NS} \pm 0.36$	$0.33^{NS} \pm 0.4$	$0.22^{NS} \pm 0.25$
[d]¶	$-3.84^{*} \pm 1.53$	$-6.64^{***} \pm 1.81$	$1.15^{NS} \pm 1.34$
[aa]‡	$-3.21^{**} \pm 1.08$	$-5.21^{***} \pm 1.24$	$0.94^{NS} \pm 0.8$
[ad]††	$0.44^{NS} \pm 0.57$	$-0.34^{NS} \pm 0.58$	$-1.10^{**} \pm 0.36$
[dd]§§	$6.20^{*} \pm 2.71$	$8.25^{*} \pm 3.18$	$2.84^{NS} \pm 2.37$

NS, not significant at 0.05 probability level; *, ** and *** significant at 0.05, 0.01 and 0.001 probability levels, respectively. Cross 1, ICP 11605 × ICP 14209; Cross 2, ICP 8863 × ICP 11605; Cross 3, ICP 8863 × ICPL 87119. †parental mean effect; § main additive genetic effect; ¶ main dominance genetic effects; ‡additive × additive, †† additive × dominance and §§ dominance × dominance epistatic genetic effects, respectively.

Parental mean effect [m] was significant and larger than all genetic effects measured in all three crosses (Table 3.3), while [a] was non-significant ($P > 0.05$) in all three crosses. The [d], [aa] and [dd] were significant ($P \leq 0.05$) in Cross 1 and Cross 2 while [ad] was significant ($P = 0.01$) only in Cross 3. Of the significant genetic effects, [d], [ad] and [aa] were negative while [dd] was consistently positive. The magnitude of the genetic effects in absolute terms were in the order of [dd] > [d] > [aa] > [ad] except in Cross 3 where [ad] was larger than [aa]. The significant [d] and [dd] in Cross 1 and Cross 2 were of opposite sign (Table 3.3), and therefore the epistatic effects can be categorised as duplicate epistasis (Mather and Jinks, 1982).

3.3.3 Estimates of variance components, heritability and genetic gain

The genetic variance component (σ^2_G) had consistently larger effects than the environmental variance component (σ^2_E) in all three crosses (Table 3.4).

Table 3.4: Variance components and heritability estimates for seed protein content in three crosses of pigeonpea

	Variance components					Heritability		Genetic gain		
Population	σ^2_P	σ^2_E	σ^2_G	σ^2_A	σ^2_D	H^2	h^2	5%	10%	20%
Cross 1	1.86	0.74	1.12	1.02	0.10	0.60	0.55	1.54	1.31	1.04
Cross 2	2.74	1.22	1.51	2.05	-0.54	0.55	<u>0.75</u>	-‡	-‡	-‡
Cross 3	1.42	0.68	0.74	1.61	-0.87	0.52	<u>1.13</u>	-‡	-‡	-‡

Cross 1, ICP 11605 × ICP 14209; Cross 2, ICP 8863 × ICP 11605; Cross 3, ICP 8863 × ICPL 87119; σ^2_P, σ^2_E, σ^2_G, σ^2_A and σ^2_D are the phenotypic, environmental, genotypic, additive and dominance variances, respectively; H^2 and h^2 are the broad- and narrow-sense heritability estimates, respectively. Values of h^2 underlined are greater than their corresponding H^2 values as result of high σ^2_A. 5%, 10% and 20% are selection intensities used for estimating genetic gain after one cycle of selection. ‡ could not be estimated due uncertainty into the estimation of narrow-sense heritability

The partitioning of the σ^2_G provided larger additive variance (σ^2_A) than the dominance variance (σ^2_D) in all three crosses (Table 3.4). Negative σ^2_D in Cross 2 and Cross 3 led to higher than expected σ^2_A, which in turn led to uncertainty into the estimation of narrow-sense heritability (h^2) with h^2 greater than broad-sense (H^2) (Table 3.4). Broad-sense heritability (H^2) is the maximum value for h^2, and h^2 must always be less or equal to the H^2 (Hartl and Jones, 2011). Therefore h^2 values greater than the H^2 in Cross 2 and Cross 3 (underlined in Table 3.4) were omitted from further discussions but only presented here for future references. Thus, the heritability estimates for SPC were 0.55 for h^2 and 0.52 to 0.60 for H^2 (Table 3.4). The predicted gain from selection was estimated for only one population (Cross 1) due to difficulty in estimating h^2 in Cross 2 and Cross 3. The genetic gain estimated in Cross 1 ranged from 1.51% at 5% SI to 1.04 at 20.0% SI in (Table 3.5).

3.3.4 Correlations of seed protein content with seed weight and seed yield

Phenotypic correlations (r_P) between SPC and 100-seed weight were moderate, positive and significant ($P \leq 0.01$) in Cross 1 and Cross 2 but very weak, negative and not significant ($P > 0.05$) in Cross 3 (Table 3.5). It was negative between SPC and seed yield in all three crosses and only significant ($P = 0.01$) in Cross 2.

Environmental correlations (r_E) were larger than r_P in all crosses except in Cross 2. The r_E between SPC and 100-seed weight were negative and highly significant ($P = 0.01$) in all three crosses. It was non-significant between SPC and seed yield in Cross 1 and Cross 2 but significant ($P = 0.05$) in Cross 3. Genotypic correlations (r_G) (Table 3.5), whenever calculable, were larger than either r_P or r_E in absolute terms, being large, positive and highly significant ($r_G = 0.87$; $P = 0.01$) between SPC and 100-seed weight in Cross 2, and negative and highly significant ($r_G = -0.33$; $P = 0.01$) between SPC and seed yield in Cross 1.

Table 3.5: Phenotypic, environmental and genotypic correlation coefficients between seed protein content and 100-seed weight, and between seed protein content and seed yield in three crosses of pigeonpea

Correlated traits (X × Y)		Cross 1	Cross 2	Cross 3
SPC† × SW§	r_P	0.23^{**}	0.20^{**}	-0.07^{NS}
	r_E	-0.44^{**}	-0.39^{**}	-0.76^{**}
	r_G	n/a¶	0.87^{**}	n/a
SPC × SY‡	r_P	-0.07^{NS}	-0.27^{**}	-0.04^{NS}
	r_E	0.20^{NS}	-0.21^{NS}	-0.31^{*}
	r_G	-0.33^{**}	n/a	n/a

NS, not significant at 0.05 probability level; *, ** and *** significantly different from zero at 0.05, 0.01 and 0.001 probability levels, respectively; Cross 1, ICP 11605 × ICP 14209; Cross 2, ICP 8863 × ICP 11605; Cross 3, ICP 8863 × ICPL 87119; † Seed protein content (%) per plant; § 100 seed weight in g per plant; ‡ Seed yield in g per plant; ¶ Could not be estimated due to excessively high σ^2_A as a result of negative σ^2_D; r_P, r_E and r_G are the phenotypic, environmental and genotypic correlation coefficients, respectively.

3.4 Discussion

3.4.1 Mean seed protein content

The mean SPC of currently popular pigeonpea cultivars is approximately 22%. Results of the present study suggest that some specific crosses involving low (19%) to moderate (23%) SPC cultivars can yield transgressive segregants with SPC as high as 25 to 27% which is a significant improvement over the parental values. Such increments in SPC within the already well-adapted cultivars can lead to significant protein yield on a sustainable basis (Saxena et al., 2002).

Although the appearance of transgressive segregants for enhanced SPC among F_2 populations used in the present study was not expected, the phenomenon has been observed for SPC in cowpea (Santos et al., 2012) and soybean (Zhang et al. 2015). In both the cowpea and soybean studies, transgressive segregants were selected at the F_2 generation, and lines derived from individual transgressive F_2 plants maintained their superior SPC in subsequent generations (Santos et al. 2012; Zhang et al. 2015). Zhang et al. (2015) also reported even more transgressive effect among $F_{2:5:6,}$ lines after following a marker-assisted recurrent selection procedure. This suggests that selection for high SPC transgressive segregants could start as early as in the F_2 generation, and more beneficial effects of transgression can be harnessed by following a recurrent selection procedure, which allows accumulation of beneficial alleles into a single genetic background.

The mean SPC of the F_1 was always lower than the mid-parent value, representing negative mid-parent heterosis, and closer to the low protein parent suggesting partial dominance of low SPC. This confirms the observations made earlier in pigeonpea by Dahiya et al. (1977), and in common beans by Noubissié et al. (2012). It is apparent that some parental combinations result in lower than expected progeny SPC compared to parental values. This is evident in the segregation of Cross 3, which had an overall transgression towards very low SPC. Cross 3 involves two popular cultivars in India that is, ICP 8863 and ICPL 87119. Selections from such a cross based on yield *per se*, without due consideration to SPC, may result in new cultivars with much lower SPC than

either parents individually. Such low SPC cultivars would exacerbate the low protein intake in areas where they are cultivated and/or consumed.

3.4.2 Genetic effects

The present study indicated that non-additive effects are more important in controlling SPC in the pigeonpea crosses studied. The results are consistent with those of earlier studies in pigeonpea and other food legume crops. For example, Baskaran and Muthiah (2007) and Vaghela et al., (2009) using line × tester and full diallel analyses, respectively reported the importance of non-additive over additive gene action for SPC in pigeonpea but they did not partition out the non-additive effects into its components. Similarly, in using full diallel analysis (Iqbal et al., 2012; Mebrahtu and Mohamed, 2004) and line × tester analysis (Ceyhan et al., 2014) reported the importance of non-additive effects over additive effects for SPC in common bean. While Tiwari et al. (1993) using a full diallel analysis also found non-additive effects to be more important than additive effects for SPC in mung bean. In the present study, duplicate epistasis and negative [aa] were the most common type of non-additive effects and their presence in the same cross was associated with increased SPC and presence of desirable transgressive segregants. Complementary epistasis, on the other hand, resulted in low SPC and negligible number of desirable transgressive segregants. Therefore, epistasis likely contributed, to a large extent, to the observed segregation pattern for SPC among the F_2 progeny in the studied populations. This observation is consistent with the fact that epistasis is considered one of the major causes of transgression in intraspecific crosses (Rieseberg et al., 1999).

3.4.3 Variance components, heritability and genetic gain

Despite the predominance of non-additive effects, the σ^2_A for SPC in all three crosses were higher than σ^2_D. The high σ^2_A amidst pervasive epistatic effects suggests that epistasis contributed to σ^2_A which is in agreement with earlier studies that epistasis contributes to, and increases σ^2_A (Cheverud and Routman, 1995; Walsh, 2005; Monnahan and Kelly, 2015). According to Mannohan and Kelly (2015) the long-held opinion that additive genetic variance component is synonymous with

additive gene effect only arises due to confusion from the simultaneous use of the terms to describe both the effects of individual genes as well as the genetic variance components of populations.

Heritability estimates on plant mean basis were moderate, with the h^2 of 0.55 being close to 0.65 previously reported in pigeonpea (Saxena et al., 2002), 0.47 in cowpea (Santos et al., 2012) and 0.63 to 0.73 in common beans (Kelly and Bliss, 1975). The H^2 of 0.52 to 0.60 is within the range of 0.34 to 0.62 reported earlier in pigeonpea (Dahiya et al., 1977). Estimates of genetic gain after one cycle of selection suggests that selection at 5% intensity would be highest genetic gain for SPC, however, comparison with results of previous similar studies in pigeonpea is not possible as none has been reported.

3.4.4 Correlation of seed protein content with seed weight and seed yield

The inconsistency in strength, sign and significance of r_P and r_E among the crosses indicates that phenotypic relationship of SPC with 100-seed weight and of both r_P and r_E between SPC and seed yield in pigeonpea is genetic-background dependent. In the case of SPC and 100-seed weight, the consistency of r_E in strength, sign and significance among all crosses indicate that genetic factors controlling the two traits responded similarly to the environment (Scully et al., 1991). However, r_G between SPC and 100-seed weight, and between SPC and seed yield could not be estimated in all crosses. Accurate estimation of r_G requires large sample sizes (Hébert et al., 1994), which were difficult to obtain in our study due to difficulty in obtaining large number of seeds in early generations (F_1, F_2 and backcross F_1). Waitt and Levin (1998) suggests that where r_G is not obtainable, r_P may be a good reflection of genetic correlations in plants but Hébert et al. (1994) found no similarity between r_P and r_G and concluded that r_P cannot be substituted for r_G. We therefore base our discussion of genetic correlations on r_G obtained in Cross 2 for SPC and 100-seed weight and in Cross 1 for SPC and seed yield.

The strong, positive and highly significant genetic correlation (r_G = 0.87; P = 0.01) between SPC and 100-seed weight in Cross 2 indicates the two traits can be simultaneously improved. Saxena et al. (1987) found positive, negative and in some cases non-significant correlations between SPC

and 100-seed weight in pigeonpea. They, however, noted that the correlations are small and therefore simultaneous selection for both high SPC and seed weight would be possible. The moderate, negative and significant genetic correlation (r_G = -0.33; P = 0.01) between SPC and seed yield in our study agrees with the report of Dahiya et al. (1977). However, given that estimates of r_G between SPC and 100-seed weight, and between SPC and seed yield in the present study could be obtained only from one cross each, generalization of these observations may not be very appropriate.

3.5 Conclusions

It may be possible to derive genotypes with SPC as high as 25 to 27 % from crosses between the well-adapted but low to moderate SPC parents. Selection of high SPC transgressive segregants would be effective at later generations followed by cycles of intermating and selfing using reciprocal recurrent selection procedures. Continuous distribution of the F_2 data, the presence of transgressive segregants and epistasis point to polygenic control of SPC in pigeonpea. Genomics approaches such as quantitative trait loci (QTL) analysis can facilitate estimating of the number of genetic loci and in-depth investigation of the pattern of epistasis. The negative genotypic association of SPC with SY suggests that selection for yield *per se* without considering SPC may lead to reduction in SPC, but the correlation is weak indicating selection for both high SPC and high SY is possible. The high positive genotypic correlation between SPC and SW implies that simultaneous selection for both traits is possible. It also suggests that SW could be used to indirectly select for enhanced SPC levels.

References

Akibode, C. S., and M. Maredia. 2011. Global and regional trends in production, trade and consumption of food legume crops. Report submitted to the Standing Panel on Impact Assessment (SPIA) of the CGIAR Science Council, FAO, Rome.

Araújo, A.P., I.F. Antunes, and M.G. Teixeira. 2005. Inheritance of root traits and phosphorus uptake in common bean (*Phaseolus vulgaris* L.) under limited soil phosphorus supply. Euphytica 145:33-40.

Baskaran, K., and A.R. Muthiah. 2007. Studies on combining ability in pigeonpea (*Cajanus cajan* (L.) Millsp.). Legume Research 30:67-69.

Bernado, R. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press, Minnesota.

Ceyhan, E., M. Harmankaya, and A. Kahraman. 2014. Combining ability and heterosis for concentration of mineral elements and protein in common bean (*Phaseolus vulgaris* L.). Turkish Journal of Agriculture and Forestry 38:581-590.

Cheverud, J.M., and E.J. Routman. 1995. Epistasis and its contribution to genetic variance components. Genetics 139:1455-1461.

Dahiya, B.S., J.S.Brar, and B.S. Bhullar. 1977. Inheritance of protein content and its correlation with grain yield in pigeonpea (*Cajanus cajan* (L.) Millsp.). Qualitus Plantarum 27:327-334.

Fenster, C.B., and L.F. Galloway. 2000. Population differentiation in an annual legume: genetic architecture. Evolution 54:1157-1172.

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative genetics in maize breeding, Iowa State University Press, Ames, IA.

Hartl, L., and J.W. Jones. 2011. Genetics: analysis of genes and genomes, 7th ed. Jones and Bartlett's Publishers International, London.

Hayman, B.I. 1958. The separation of epistatic from additive and dominance variation in generation means. Heredity 12:371-390.

Hébert, D., S Fauré, and I Olivieri. 1994. Genetic, phenotypic and environmental correlations in black medic, *Medicago lupulina* L., grown in three environments. Theoretical Applied and Genetics 88:604-613.

Holland, J.B. 2001. Epistasis and plant breeding. Plant Breeding Reviews 21:27-92.

ICRISAT. 1993a. Pigeonpea variety ICP 8863. Plant material description no. 44. Andhra Pradesh (India): International Crops Research Institute for the Semi-Arid tropics.

ICRISAT. 1993b. Pigeonpea variety ICPL 151. Plant material description no. 41. Andhra Pradesh (India) International Crops Research Institute for the Semi-Arid tropics.

ICRISAT. 1993c. Pigeonpea variety ICPL 87119. Plant material description no. 43. Andhra Pradesh (India): International Crops Research Institute for the Semi-Arid tropics.

Iqbal, A.M., F.A. Nehvi, S.A. Wani, Z.A. Dar, A.A. Lone, and H. Qadri. 2012. Combining ability study over environments in dry beans (*Phaseolus vulgaris* L.). SAARC Journal of Agriculture 10:61-69.

Kelly, J.D., and F.A. Bliss. 1975. Heritability estimates of percentage seed protein and available methionine and correlations with yield in dry beans. Crop Science 15:753-757.

Mather, K., and Jinks, J.L. 1982. Biometrical Genetics, 3rd ed. Chapman and Hall, New York.

Mebrahtu, T., and A. Mohamed. 2003. A seven-parental diallel analysis of nutritional composition of common beans. Plant Foods for Human Nutrition 58:1-11.

Monnahan, P.J., and J.K. Kelly (2015) Epistasis is a major determinant of the additive genetic variance in *Mimulus guttatus*. PLoS Genetics 11:e1005201.

Morris, S.S., B. Cogill, and R .Uauy. 2008. Effective international action against undernutrition: why has it proven so difficult and what can be done to accelerate progress? The Lancet 371:608-621.

Mula, M.G., and K.B. Saxena. 2010. Lifting the level of awareness on pigeonpea - a global perspective. Andhra Pradesh (India): International Crops Research Institute for the Semi-Arid Tropics.

Noubissié, J.B.T., J.M. Bell, and H.T. Yadji. 2012. Studies on variability and gene effects of seed protein contents in Cameroonian bean (*Phaseolus vulgaris* L.) cultivars. Journal of Agriculture, Forestry and the Social Sciences 8:17-23.

Pooni, H.S., D.T. Coombs, P.S. Virk, and J.L. Jinks. 1987. Detection of epistasis and linkage of interacting genes in the presence of reciprocal differences. Heredity 58:257-266.

Remanandan, P., and L. Singh. 1997. Pigeonpea. In: Fuccillo, D., L. Sears, and P. Stapleton, editors, Biodiversity in trust: Conservation and use of plant genetic resources in CGIAR centres. Cambridge University Press, New York, p 156-167.

Remanandan, P.D., V.S.S.R. Sastry, and H.M. Mengesha. 1988. ICRISAT pigeonpea germplasm catalog: passport information. Andra Pradesh (India): International Crops Research Institute for the Semi-arid Tropics.

Rieseberg, L.H., M.A. Archer, and R.K. Wayne. 1999. Transgressive segregation, adaptation and speciation. Heredity 83: 363-372.

Santos, C.A.F., D.C.C. da Costa, and W.R. da Silva, L.S. Boiteux. 2012. Genetic analysis of total seed protein content in two cowpea crosses. Crop Science 52:2501-2506.

SAS Institute. 2015. User's Guide, SAS/STAT® 14.1, Cary, NC.

Saxena, K.B., D.G. Faris, U. Singh, and R.V. Kumar. 1987. Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition 36: 335-340.

Saxena, K.B., R.V. Kumar, and P.V. Rao. 2002. Pigeonpea nutrition and its improvement. In: Basra, A.S., and I.S. Randhawa, editors, Quality improvement in field crops. Food Products Press, p. 227-260.

Scully, B.T., D.H. Wallace, and D.R. Viands. 1991. Heritability and correlation of biomass, growth rates, harvest index, and phenology to the yield of common beans. Journal of the American Society for Horticultural Science 116:127-130.

Searle, S. 1961. Phenotypic, genetic and environmental correlations. Biometrics 17:474-480.

Tiwari, D.S., V. Singh, and P.S. Shukla. 1993. Combining ability in mung bean (*Vigna radiata* (L.) Wilczek). The Indian Journal of Genetics and Plant Breeding 53:395-398.

Upadhyaya, H.D., K.N. Reddy, C.L.L. Gowda, and S.N. Silim. 2007. Patterns of diversity in pigeonpea (*Cajanus cajan* (L.) Millsp.) germplasm collected from different elevations in Kenya. Genetic Resources and Crop Evolution 54:1787-1795.

Vaghela, K.O., R.T. Desai, J.R. Nizama, J.D. Patel, and V. Sharma. 2009. Combining ability analysis in pigeonpea (C*ajanus cajan* (L.) Millsp.). Legume Research 32:274-277.

Waitt, D.E., and D.A. Levin (1998) Genetic and phenotypic correlations in plants: a botanical test of Cheverud's conjecture. Heredity 80:310-319.

Walsh, B. 2005. The struggle to exploit non-additive variation. Australian Journal of Agricultural Research 56:873-881.

Chapter Four

Mapping quantitative trait loci for seed protein content in pigeonpea

Abstract

The objective of the present study was to use high-density genetic maps to identify quantitative trait loci (QTL) associated with SPC and its relationship with 100-seed weight (SW), seed yield (SY), days to first flower (DTF) and growth habit (GH). Five F_2 mapping populations segregating for SPC were genotyped using genotyping-by-sequencing and phenotyped for the traits. The average inter-marker distance on the population-specific maps varied from 1.6 cM to 3.5 cM. On the basis of the population-specific and consensus linkage maps, a total of 196 main effect QTLs (M-QTLs) were detected that explained 0.7 to 91.3% of the phenotypic variation for the five traits across the five F_2 mapping populations. In the case of SPC, a core trait in the present study, a total of 48 main effect QTLs (M-QTLs) with phenotypic variance explained (PVE) ranging from 0.7 to 23.5% were detected across five populations of which 15 M-QTLs were major (PVE≥10%). Twenty-seven of the M-QTLs from the five F_2 mapping populations could be projected into six consensus M-QTL regions. Out of 573 epistatic QTLs (E-QTLs) detected with PVE ranging from 6.3 to 99.4% across traits and populations, 34 involved SPC with PVE ranging from 6.3 to 69.8%. Several co-localization of M-QTLs and E-QTLs affecting SPC and the agronomic traits were also detected and could explain the genetic basis of the significant ($P < 0.05$) correlations of SPC with SW (r^2 = 0.22 to 0.30), SY (r^2 = -0.18 to -0.28), DTF (r^2 = -0.17 to -0.31) and GH (r^2 = 0.18 to 0.34). The quantitative nature of genetic control of SPC and its relationship with agronomic traits suggest that marker-assisted recurrent selection or genomic selection would be effective for the simultaneous improvement of SPC and other important traits.

4.1 Introduction

Protein deficiency affects the health of millions of children and their mothers, but protein-rich plant foods may offer solutions particularly in areas of the world where intake of animal protein is low (Li et al., 2015). One such crop is pigeonpea (*Cajanus cajan* (L.) Millsp), which is an important source of dietary protein to nearly one billion people globally (Varshney et al., 2012). It is widely cultivated in the tropics and semi-arid tropics of Asia and Africa. Pigeonpea maintains better yields than other legume crops under environmental extremes such as heat and drought conditions, and low soil fertility (Rao et al., 2010; Akibode and Maredia, 2011). These attributes position pigeonpea as the crop for the marginal environments, and resources-poor farmers (Khoury et al., 2015). Increasing SPC of pigeonpea is, therefore, an important contribution towards alleviating malnutrition among the poor. To improve SPC requires an understanding of its genetic architecture and how it relates to traits of agronomic importance.

Few studies have been reported on genetic control of SPC in pigeonpea with results suggesting quantitative inheritance (Dahiya et al., 1977; Saxena et al., 2002; Baskaran and Muthiah, 2007; Vaghela et al., 2009). However, the classical quantitative genetic approaches used in the reported studies are limited in power and resolution to dissect the genetic architecture of a quantitative trait like SPC. Similarly, information is limited on the genetic basis of the often positive or negative or no relationships of SPC with seed yield (SY) and yield-related traits in the crop (Saxena et al., 1987, 2002; Rekha et al., 2015). Determining the genetic basis of trait correlations in pigeonpea is essential in designing breeding strategies that aim at improving and stabilizing SPC while maintaining yield and other desirable agronomic attributes. The availability of genomic resources in pigeonpea coupled with advances in high-throughput genotyping technologies provide the opportunity to dissect the genetic architecture of several quantitative traits in the crop (Bohra et al., 2012; Kumawat et al., 2012; Varshney et al., 2012; Singh et al.,2015; Kumar et al., 2016).

However, genetic architecture of SPC in pigeonpea, and the basis of its relationship with other traits of importance has remained untouched by genomic revolution in the crop. A common genomics approach to understand the genetic architecture of quantitative traits involves whole

genome scans to identify quantitative trait loci (QTLs) (Abiola et al., 2003). Through QTL analysis, genetic parameters such as number of loci, the types and size of their effects and epistasis, which constitutes the genetic architecture underlying quantitative phenotypic variation can be estimated (Simon et al., 2008). However, these parameters are mostly population specific (Lynch and Walsh, 1998). As a result, QTLs identified in one population may not necessarily be present in another population (Cui et al., 2014). Thus, any description of a trait's genetic architecture made from a single population, likely describes only a small part of all the loci, their effects, and potential interactions that contribute to phenotypic variation for a trait within a species (Symonds et al., 2005; Simon et al., 2008). To this end, the use of two or more segregating mapping populations in a single study have become common (Symonds et al., 2005; Lou et al., 2007; Simon et al., 2008; Borha et al., 2012; Li et al., 2015). Regardless of the number of segregating populations, QTL analysis is preceded by the development of appropriate mapping populations and anchoring of markers on a genetic linkage map.

Population types used for QTL mapping include F_2, backcross (BC), doubled haploid, recombinant inbred lines (RILs) and near-isogenic lines (NILs). The RILs and NILs are the most common because they are immortal, and have higher frequency of recombination events, which can reduce cost by genotyping within a limited population size (Chen et al., 2014). However, disadvantages of RILs and NILs are that their development is time-consuming and costly (Chen et al., 2014). This is true for a crop like pigeonpea in which most of the traditionally cultivated germplasm are from the medium and long duration maturity groups that take five to nine months to mature (Vales et al., 2012). This is likely the reason for the wide use of F_2 and the F_2-derived F_3 generations for both genetic map construction and QTL analysis in pigeonpea. For instance, the majority of genetic linkage maps constructed to date in pigeonpea are F_2-derived intraspecific maps.

The first intraspecific genetic map of pigeonpea was constructed using two F_2 populations comprising 120 and 78 simple sequence repeats (SSR) loci with map lengths of 534.89 cM and 466.97 cM, respectively (Gnanesh et al., 2011). The two genetic maps were used for QTL analysis for sterility mosaic disease using $F_{2:3}$ phenotype data (Gnanesh et al., 2011). This was followed by construction of four other F_2 intraspecific maps comprising 59 to 140 SSR loci spanning 586

to 881.6 cM (Borha et al., 2012). The four maps together with the two previous maps from Gnanesh et al. (2011) were used to create a consensus map comprising 339 SSR loci spanning a distance of 1,059 cM (Borha et al., 2012). Three of the mapping populations were used to map QTLs for fertility restoration (Borha et al., 2012). Another intraspecific genetic map was constructed with 296 genic single nucleotide polymorphisms (SNPs) and SSR markers covering a map length of 1520.22 cM with average marker interval of 5.1 cM (Kumawat et al., 2012). This map was used to map QTLs for plant type and earliness using phenotypic data derived from $F_{2:3}$ generation. The present study used genotyping-by-sequencing (GBS)-derived SNPs to construct genetic maps and to map QTLs for SPC as well as QTLs for other traits in five F_2 mapping populations. Genotyping-by-sequencing was proposed by Elshire et al. (2011), and since then it has been widely used for genotyping of different crop species such as barley and wheat (Poland et al., 2012), maize (Romay et al., 2013) and chickpea (Jaganathan et al., 2015; Kale et al., 2015).

The aim of the present study was to examine the genetic basis of phenotypic variation in SPC and its relationship with agronomic traits such as seed yield, seed weight, number of days to first flowering and growth habit. To achieve this, five F_2 mapping populations segregating for SPC were phenotyped and genotyped, high-density population-specific and consensus genetic maps constructed and QTL analysis performed.

4.2 Materials and methods

4.2.1 Plant material and field trial

Six pigeonpea genotypes that included ICP 11605, ICP 8863, ICP 14209, HPL 24, ICP 5529 and ICPL 88719 were used in the present study. ICP 8863 was selected from landrace ICP 7626 (P-15-3-3) and it is widely cultivated in India. It is high yielding with 100-seed weight of ~9.5 g and matures in 150-160 days. It is resistant to Fusarium wilt but susceptible to sterility mosaic virus (ICRISAT, 1993a) and it has a moderate SPC of ~22.0%. ICP 11605 (ICPL 151) was selected from the cross ICP 6997 × Prabhat. It is a determinate cultivar, yielding ~1.03 t/ha with 100-seed weight of 10 g and matures in 120-130 days (ICRISAT, 1993b) and has a low SPC of ~20.9%.

ICP 14209 is a landrace variety with moderate SPC (23.0%). ICPL 87119 was developed from the cross ICP 1-6-W3–Wl × C 11 and it is widely adapted and cultivated in India. It matures in 160-180 days, is high yielding and has resistance to Fusarium wilt and sterility mosaic virus diseases (ICRISAT, 1993c). It is low in SPC (~19.3%). HPL 24 is an advanced breeding line derived from the cross of cultivar *C. cajan* cv Baigani × *C. scarabaeoides* previously reported to have ~30% SPC (Saxena et al., 2002). It is indeterminate and of medium maturity duration. ICP 5529 with pedigree P-4864-1 originated from India. It is indeterminate with medium maturity duration and with SPC indicated to be 27%.

Seeds of the parental genotypes were obtained from ICRISAT's genebank. The following crosses were developed: ICP 11605 × ICP 14209, ICP 8863 × ICP 11605, HPL 24 × ICP 11605, ICP 5529 × ICP 11605 and ICP 8863 × ICPL 87119. For brevity, the populations are hereafter referred to as Pop1, Pop2, Pop3, Pop4 and Pop5, respectively. One F_1 plant was selfed to generate F_2 seeds. For trait evaluation, the parents and F_2 generations from each mapping population were sown under field conditions. Each population was sown with 350 to 400 F_2 seeds to ensure an adequate number of plants were established. Sowing was done in 4 m long rows spaced 75 cm apart and 30 cm within a row. Plot sizes were two rows for each of the two parents and 25 to 28 rows in the F_2.

Agronomic practices included application of 100 kg/ha of diammonium phosphate as basal fertilizer without any top dressing, 2 and 4 L/ha of pendimethalin and paraquat dichloride pre-emergence herbicides, respectively, provision of two irrigations, one each at planting and pod filling stages, and two weedings one each at early vegetative and podding stages. Pod borers (*Maruca vitrata* Fab. and *Helicoverpa armigera* Hub.) were controlled by spraying with acephate and spinosad insecticides at rates of 1.0 kg/ha and 0.2 L/ha, respectively at 15 days intervals from flowering to podding stages. At maturity individual pods from individual plants were carefully hand-harvested leaving out plants at the beginning and at the end of each row and those at the field borders to avoid border effects. Sun drying was done for one week before threshing and another one week after threshing to ensure uniform reduction in seed moisture content.

4.2.2 Phenotypic measurements

To estimate SPC, 10 g of mature dry clean seeds of each of 188 F_2 plants in each of the five mapping populations were analyzed at the Central Analytical Services (Charles Renard Analytical) Laboratory at ICRISAT, India using tri-acid total N digestion method as described in Upadhyaya et al. (2016). Seeds were oven-dried at 60ºC for 48 hours. The dried seed samples were ground into powder in a mill with Teflon chambers. The ground samples were again kept in an oven at 60ºC overnight. The samples and appropriate blanks were digested simultaneously in duplicate (i.e. two independent analyses). Briefly, 1.0 g of ground sample was transferred to a 75 ml digestion tube containing 10 ml of tri-acid mixture of nitric, sulfuric and perchloric acids in the ratio of 10:0.5:2 (v/v). The contents were left overnight in the digestion chamber for cold digestion. Samples were initially digested at 120ºC for 1 hour followed by digestion at 230ºC for approximately 2 hours in order to obtain clear and colourless digests. After digests were cooled, the contents were dissolved in distilled water and volume made up to 75 ml and then mixed well by shaking. Aliquots were obtained from the digests and used to estimate the total nitrogen (N) using a San++Automated Wet Chemistry Analyzer (Skalar, Breda, The Netherlands). Seed protein content of a sample was estimated by multiplying its N (%) content by factor 6.25. Besides SPC, data were also recorded for 100-seed weight (SW) in grams, seed yield (SY) in grams, number of days to first flowering (DTF), and growth habit (GH; determinate vs indeterminate).

4.2.3 DNA isolation and genotyping-by-sequencing

Total genomic DNA (gDNA) from 188 F_2 plants and the parents from each mapping population were isolated from young trifoliate leaves of one-month old plants using MN-NucleoSpin®96 Plant II DNA Kit (MN Ltd., Germany) following the manufacturer's specifications. The quantity and quality of gDNA was checked on 0.8% agarose gel. Genotyping-by-sequencing was used to obtain the genotypic data. Briefly, gDNA libraries of each of the two parents and F_2 individuals in each mapping population were fragmented using ApeKI endonuclease (recognition site: G/CWCG) for 2 hr at 75°C. Barcode adapters, which have unique multiplex sequence index, were ligated to the sticky ends of DNA fragments using T4 ligase in ligase buffer containing adenosine

triphosphate (ATP). The ligated samples were incubated at 22°C for 1 hr and heated to 65°C for 30 min to inactivate the T4 ligase. The libraries for 96 individuals were pooled at a rate of five µl per sample. The sample multiplex was purified to remove excess adapters, followed by elution in a final volume of 50 µl. The multiplexes were PCR amplified and cleaned using magnetic beads and evaluated for fragment sizes using a DNA analyser. Libraries without adapter dimers were subjected to sequencing at 5× depth using the Illumina HiSeq 2500 platform (Illumina Inc, San Diego, CA, USA).

4.2.4 Single nucleotide polymorphism (SNPs) identification and genotyping

The sequence reads obtained from the Illumina HiSeq 2500 platform were used for SNP identification and genotyping using GBS analysis pipeline implemented in TASSEL v4.02 (TASSEL-GBS) (Glaubitz et al., 2014). Firstly, the sequence reads were sorted, demultiplexed according to the sample barcodes and trimmed to first 64 bases starting from the enzyme cut site. Reads containing 'N' within the first 64 bases and reads having more than 50% of low quality base pairs (Phred <5%) were discarded. The filtered, high-quality reads from each sample were aligned to the pigeonpea draft genome sequence (*C. cajan* v1.0) (Varshney et al., 2012) using Bowtie 2 sequence alignment software. The alignment file was processed through TASSEL-GBS pipeline for SNP calling and genotyping. The quality of SNPs called in each F_2 individual was compared with the SNPs identified in parental lines. The parental line SNPs were obtained from existing whole genome resequencing (WGRS) data (Kumar et al., 2016). SNPs having confident parental calls were considered for further analysis. SNPs and F_2 individuals having more than 30% and 70% missing data, respectively, were filtered out. The quality SNP data was used for genetic map construction and QTL analysis.

4.2.5 Construction of population specific genetic maps

The SNPs obtained from GBS of the F_2 populations were used as markers for constructing the genetic maps using JoinMap v4.1 software (https://www.kyazma.nl/index.php/JoinMap/). SNPs which significantly deviated from the expected F_2 segregation pattern of 1:2:1 at Chi-square

goodness-fit-test (χ^2) probability 10^{-9}, and SNPs with similarity index of 1 were excluded. This was followed by creating base or anchor maps with only markers segregating in 1:2:1 pattern at χ^2 P > 0.05. The base maps were created using the following settings in JoinMap v4.1: independence LOD value 2.5 to 10.0, recombination frequency ≥0.49 and a χ^2 jump threshold for removal of loci at 5. A ''Ripple'' was performed after adding a marker into the map, depending on the number of markers in a given group (van Ooijen, 2006). Map distance was calculated using Kosambi mapping function (Kosambi, 1944) and a third round was set to allow mapping of an optimum number of loci in the genetic map. Placement of markers into different linkage groups (LGs) was done with ''LOD groupings,'' and ''Create group using the mapping tree'' commands. Mean χ^2 contributions or average contributions to the goodness-of-fit of each locus were also checked to determine the best fitting position for markers in genetic maps. The markers showing negative map distances or a large jump in mean χ^2 values were discarded. The markers in anchor maps were then used as fixed order markers onto which markers with χ^2 P>10^{-9} but <0.05 were added to create the final map. In doing this, the entire step used in creating the anchor maps were repeated. The final maps were drawn with the help of Map-Chart version 2.2 (Voorrips, 2002). Marker interval was estimated as map length divided by the number of mapped markers.

4.2.6 Construction of consensus genetic map

Genotype data from the five F_2 genetic maps were used for developing a consensus genetic map using JoinMap v4.1. In this approach, segregation data from all mapping populations on all individuals were used to achieve a consensus order of loci to be used to develop the consensus map (Wenzl et al., 2006). Map integration was accomplished by following three steps (Truco et al., 2007): (i) Common loci among different mapping populations were carried out *a priori* (ii) The "Combine groups for map integration" function from the "Join" menu of Joinmap was used to synthesise an integrated linkage group for each chromosome. To assess the amount of collinearity in marker orders between consensus and component genetic maps, correlation coefficients (*r*) were calculated from marker positions in consensus and individual genetic maps and their significance were tested. To further visualize the extent of correlation between consensus and component maps, scatter plots were generated between each of the consensus linkage groups

and corresponding component linkage groups from all populations. To visually assess the congruency of marker orders, all the developed maps were aligned together using a comparative mapping program CMap version 1.01 (Youens-Clarks et al., 2009).

4.2.7 QTL mapping in individual populations

Composite interval mapping (CIM) (Zeng, 1994) and inclusive composite interval mapping (ICIM) (Li et al., 2007) were used to detect main effect QTLs (M-QTLs) while epistatic QTLs (E-QTLs) were detected using ICIM. Composite interval mapping and ICIM are implemented in Windows QTL Cartographer v2.5 (Wang et al., 2010) and QTL Icimapping v4.0 (Wang et al., 2015), respectively. The advantage of both CIM and ICIM is that they are regression-based and therefore robust against non-Gaussian trait distribution (Rebai, 1997).

For CIM, the Standard Model 6, walk speed of 1.0 cM, and forward-backward stepwise regression for setting number of marker cofactors for background control were used to identify M-QTLs. To leave out signals within 10.0 cM distance on either side of the flanking markers or QTL test site a window size of 10 cM was used. Thresholds for declaring QTLs were determined by 1000 permutations at significance of 0.05.

In using ICIM to detect M-QTLs, marker selection was performed just once using stepwise regression and considering all marker information simultaneously. Phenotypic values were then adjusted by all markers retained in the regression equation, except the two markers flanking the current mapping interval (Li et al., 2007; Li et al., 2008; Li et al., 2011). Permutation tests were conducted using SPC in the five F_2 mapping populations to determine the criteria for model selection in the first step of ICIM. For all five F_2 populations, the probability of a marker moving into the model corresponding to the overall type I error $\alpha = 0.05$ was approximately 10^{-5}. The probability of a marker moving out of the model was set at twice the probability of a marker moving into the model. The LOD threshold to declare the existence of a QTL was calculated by permutation tests as well. However, because of the always conservative nature of thresholds

retained from permutation tests (Anderson et al., 2003), a default LOD threshold of 2.5 was used to report QTLs and determine common (consensus) QTLs across populations.

Furthermore, where M-QTL identified by CIM was also detected by ICIM, the region was considered as one QTL. Similarly, where an M-QTL for SPC identified by either CIM or ICIM co-localises with M-QTL(s) of another trait detected by either of the two methods, the region was treated as a region of co-localisation.

For E-QTL mapping, all possible pairs of scanning positions were tested by ICIM (Li et al., 2008). In other words, digenic interactions may be detected regardless of whether the two interacting QTLs have significant additive effects or not (Li et al., 2015). Due to the large amount of variables in digenic QTL mapping, a much stricter probability (10^{-6}) of a marker moving into the model was used. The probability of a marker moving out of the model was set at twice the probability of a marker moving into the model. The default QTL-Icimapping LOD threshold of 5.0 was used to declare the existence of E-QTLs.

4.2.8 Common or consensus QTLs across five F_2 populations

Due to differences among the population-specific genetic maps, it was difficult to directly find common QTLs across the five F_2 populations on the basis of the QTL or marker position in each genetic map. Therefore, QTLs obtained in each of the five individual populations were projected onto the consensus map by using either QTL peak- or flanking-marker positions indicated in the individual population maps using a procedure adopted from Schweizer and Stein (2011) as described below.

If only peak-marker positions from the individual map were available, the QTL region was assumed by default to extend 5 cM north and south from peak-marker position, resulting in a confidence interval of 10 cM. If only one flanking marker could be projected onto the consensus map, a QTL interval of 10 cM extension north or south from the lower or upper flanking marker, respectively, was assumed by default. If neither peak nor flanking markers were included in the

consensus map, nearby tightly linked markers (maximum of 5 cM from the peak or flanking markers) were searched on the consensus map. If no replacement markers could be identified within this distance, the QTL was excluded from the analysis. Based on these projections, two types of common QTLs were defined. Firstly, a 'Consensus QTL' was defined as any region of the consensus genetic linkage map with overlapping M-QTL intervals for SPC from more than one population. Secondly, a region of consensus linkage map at which M-QTL intervals for SPC overlap with that of one or more of the other traits was considered a 'QTL Cluster'

4.2.9 QTL nomenclature

For individual population, a specific identifier was assigned to each QTL, whereby "q" stands for QTL, followed by a set of upper case letters indicating the trait, followed by linkage group (LG) name, then a hyphen, method of QTL detection, and lastly, the QTL number on that LG in ascending order. For example, the designation "*qPROT-cim-3.1*" stands for "QTL for SPC" on LG "CcLG03" and it is the first QTL for SPC on that LG. For QTLs projected onto the consensus map, a prefix is added to the QTL name indicating the source population. For example "*Pop1qPROT-cim-3.1*" indicates a QTL for SPC from Pop1.

4.3 Results

4.3.1 Variation in seed protein content and agronomic traits in five F_2 mapping populations

The descriptive statistics for all traits is presented in Table 4.1. The mean SPC of the parents ranged from 19.3 to 21.5% among the low SPC parents and from 22.3 to 24.6% among the high SPC parents. The lowest SPC difference between parents of a pair was 0.8% (Pop2) and highest was 3.1% (Pop4). Among F_2 plants, the differences between the smallest and largest SPC individuals ranged from 5.8 (Pop4) to 10.3 % (Pop3) with mean range of 19.44±1.28 in Pop5 to 23.06±1.08 in Pop4 (Table 4.1). Similar statistics for the other traits are presented in Table 4.1. Shapiro-Wilk test showed that distributions for SPC in Pop1, Pop3 and Pop4 were not significantly

($P > 0.05$) different from a Gaussian distribution while Pop2 and Pop5 differed significantly ($P \leq 0.05$) from a normal distribution. Such non-Gaussian distributions were also noted for most of the other traits in some or all of five populations such as DTF (Pop1, Pop2, Pop3 and Pop4), SW (Pop2 and Pop4) and SY in all the five populations (Table 4.1).

Table 4.1: Population size, mean, variance, skewness, kurtosis, minimum and maximum values, and w-test for seed protein content and four agronomic traits in five F_2 mapping populations of pigeonpea

Population/Trait	P_1	P_2	$\|P_1-P_2\|$	n	Mean	Variance	CV (%)	S	K	F_2-range	W-test
Pop1 (ICP 11605 × ICP 14209)											
SPC†	21.5	23.1	1.6	178	22.2±1.2	1.6	5.6	0.2	1.1	19.1-26.5	1.0^{NS}
DTF§	138.0	66.0	72.0	178	101.1±12.6	158.1	12.4	-0.4	0.1	69.0-133.0	0.9***
SW¶	8.7	12.2	3.5	178	9.1±1.1	1.2	12.0	0.0	0.1	6.2-12.2	1.0^{NS}
SY‡	17.3	22.5	5.2	178	53.4±32.5	1056.5	60.8	1.1	1.6	8.7-186.9	0.9***
Pop2 (ICP 8863 × ICP 11605)											
SPC	22.3	21.5	0.8	175	21.7±1.5	2.2	6.9	-0.5	-0.2	17.5-24.8	1.0**
DTF	90.0	66.0	24.0	175	83.5±11.1	124.0	13.3	0.4	0.5	58.0-117.0	1.0***
SW	9.9	12.2	2.3	175	11.3±1.4	2.0	12.4	1.4	9.8	7.5-20.6	0.9***
SY	24.5	22.5	2.0	175	37.3±27.6	761.4	73.9	1.5	1.7	8.0-127.5	0.8***
Pop3 (HPL 24 × ICP 11605)											
SPC	23.0	21.5	1.5	157	22.4±1.7	2.8	7.5	0.3	0.5	17.7-28.0	1.0^{NS}
DTF	112.0	66.0	46.0	157	93.4±15.2	231.7	16.3	-0.6	-0.8	66.0-123.0	0.9***
SW	8.1	12.2	4.1	157	10.4±1.3	1.8	12.8	-0.3	0.2	5.7-13.7	1.0^{NS}
SY	152.7	22.5	130.2	157	33.6±21.6	466.9	64.4	1.1	0.8	5.7-106.5	0.9***

^{NS}, not significantly different from a Gaussian distribution at P = 0.05; *, ** and *** significantly different from a Gaussian distribution at 0.05, 0.01 and 0.001 probability levels, respectively. †seed protein content; §days to first flower; ¶ 100-seed weight; ‡ seed yield; P_1, parent 1; P_2, parent 2; $|P_1-P_2|$, absolute difference in trait value between two parents of a cross; S, skewness; K, kurtosis; W-test, Shapiro-Wilk test.

Table 4.1: (continued)

Population/Trait	P_1	P_2	$\|P_1-P_2\|$	n	Mean	Variance	CV (%)	S	K	F_2-range	W-test
Pop4 (ICP 5529 × ICP 11605)											
SPC	24.6	21.5	3.1	179	23.0±1.1	1.2	4.7	0.0	0.2	20.2-26.6	1.0^{NS}
DTF	104.0	66.0	38.0	156	81.2±9.2	84.2	11.3	-0.2	-0.5	65.0-102.0	0.9***
SW	8.6	12.2	3.6	179	10.3±1.4	1.8	13.1	-0.6	1.2	5.3-13.4	1.0*
SY	23.3	22.5	0.8	179	47.8±38.7	1497.9	80.9	1.8	3.5	5.3-203.1	0.8***
Pop5 (ICP 8863 × ICP 11605)											
SPC	22.3	19.3	3.0	137	19.4±1.3	1.63	6.6	-0.4	-0.3	16.0-21.8	1.0*
DTF	90.0	103.0	13.0	137	95.5±8.7	75.71	9.1	-0.4	0.5	62.0-116.0	1.0^{NS}
SW	9.9	11.1	1.2	136	11.6±1.1	1.11	9.1	-0.3	0.5	8.6-14.1	1.0^{NS}
SY	24.5	38.9	14.4	137	50.9±32.6	1062.36	64.0	1.8	3.9	7.9-192	0.8***

4.3.2 Correlation of seed protein content with other traits

Correlation analysis guides in developing selection criteria for accumulating optimum combination of desirable traits in a single genotype. With a focus on how SPC relates with other agronomic traits, simple Pearson's correlation analysis was conducted and the results are presented in Table 4.2

Table 4.2: Correlation coefficient of seed protein content with 100-seed weight, seed yield, days to first flower and growth habit in five F_2 mapping populations of pigeonpea

	Correlated traits (SPC† × AT§)			
Population	SPC × DTF¶	SPC × GH#	SPC × SW‡	SPC × SY††
Pop1 (ICP 11605 × ICP 14209)	-0.17*	0.20**	0.22**	-0.18*
Pop2 (ICP 8863 × ICP 11605)	-0.11^{NS}	0.18*	0.30***	-0.23**
Pop3 (HPL 24 × ICP 11605)	-0.31***	0.34***	0.13^{NS}	-0.28***
Pop4 (ICP 5529 × ICP 11605)	-0.01^{NS}	0.16^{NS}	-0.02^{NS}	-0.24**
Pop5 (ICP 8863 × ICPL 87119)	-0.00^{NS}	-	0.06^{NS}	-0.06^{NS}

^{NS}, not significantly different from zero at P = 0.05; *, ** and *** significantly different from zero at 0.05, 0.01 and 0.001 probability levels, respectively. † Seed protein content; § Agronomic traits; ¶ Days to first flower and # Growth habit ‡ 100-seed weight; †† Seed yield.

Correlations were negative between SPC and DTF in all mapping populations but significant (P≤0.05) in only two populations (Pop1 and Pop3). Similarly, correlations between SPC and SY were negative and significant in all populations except in Pop5. In contrast, positive significant correlations were noted between SPC and GH in three of the five populations (Table 4.2). While correlations between SPC and SW were positive in all populations except Pop3, although only significant in two populations (Pop1 and Pop2) (Table 4.2). In case of SPC and SY, correlations were negative in all mapping populations but significant in only four (Table 4.2). The genetic basis of these correlations were investigated through QTL linkage analysis.

4.3.3 Sequence data and SNPs discovery

The results of GBS of four F_2 mapping populations using Illumina HiSeq 2500 platform is presented in Table 4.3 and Appendix 1 to Appendix 4. Information on the 5th population (Pop5) is not presented because the genotyping and genetic map construction were performed under a separate project (S. Parupalli, Pers. Comm., 2016). Parental genotypes, together with 178, 175, 157 and 179 out of 188 F_2 individuals in Pop1, Pop2, Pop3 and Pop4 were sequenced. However, DNA from 10 F_2 plants in Pop1, 13 (Pop2), 31 (Pop3) and 9 (Pop4) were not used for GBS library preparation due to low concentration and quality.

Table 4.3: Number of reads and data size in gigabytes (Gb) generated in four F_2 mapping populations of pigeonpea

Data features / generation	Pop1	Pop2	Pop3	Pop4
Number of reads (millions)				
P_1	1.1	3.0	3.0	5.37
P_2	2.6	7.6	3.3	1.61
F_2 - range	0.8-5.8	0.5-9.5	0.7-6.8	0.41-5.26
F_2 - average	2.3	1.9	2.1	1.67
Data size (Gb)				
P_1	0.114	0.303	0.299	0543
P_2	0.263	0.766	0.335	0.163
F_2 - range	0.079-0.587	0.049-0.962	0.074-0.691	0.041-0.531
F_2 - average	0.267	0.192	0.212	0.168

P_1, parent 1; P_2, parent 2; Pop1, ICP 11605 (P_1) × ICP 14209 (P_2); Pop2, ICP 8863 (P_1) × ICP 11605 (P_2); Pop3, HPL 24 (P_1) × ICP 11605 (P_2); Pop4, ICP 5529 (P_1) × ICP 11605 (P_2)

In total, 403.66, 343.26, 345.52 and 305.54 million reads found in 40.77, 34.67, 34.90 and 30.86 Gb of GBS data were generated in Pop1, Pop2, Pop3 and Pop4, respectively. In the case of parental line data, ICP 11605, a common parent in all four F_2 mapping populations, had an average of 3.41 million reads and a range of 1.13 million (Pop1) to 7.59 million (Pop2). The reads in ICP 11605 were found in an average of 0.34 Gb of GBS data with a range of 0.11 Gb (Pop 1) to 0.77 Gb (Pop 2) in the four populations. For the remaining four parents, 2.60, 3.00, 2.96 and 5.37 million reads in 0.26, 0.30, 0.30 and 0.54 Gb of GBS data in ICP 14209 (Pop1), ICP 8863 (Pop2), HPL 24

(Pop3) and ICP 5529 (Pop4), respectively were generated. In the F_2, the average number of reads generated per individual varied from 1.67 million in Pop4 (with a range of 0.41 to 5.26 million reads) to 2.25 million reads in Pop1 (with a range of 0.79 to 5.82 million). The reads were found in 0.17 Gb GBS data in Pop4 (with a range of 0.04 to 0.53 Gb per plant) to 0.67 Gb in Pop1 (with a range of 0.08 to 0.59 Gb).

The generated sequencing data on F_2 plants was used for SNP identification and subsequently filtered to remove heterozygotes. The final number of good quality SNPs produced were 15728, 7494, 12030 and 12654 in Pop1, Pop2, Pop3 and Pop4, respectively (Table 4.4). In the case of Pop5, a total of 11526 SNPs were generated using same procedure as described for the four populations (S. Parupalli, Pers. Comm., 2016).

Table 4.4: Features of individual genetic maps from five F_2 mapping populations of pigeonpea

	Individual genetic maps				
Features	Pop1	Pop2	Pop3	Pop4	Pop5†
No. total SNPs	15728	7494	12030	12662	11526
No. SNPs showing severe segregation distortion ($P<1.0x10^{-9}$)	12121	6075	9129	9727	7585
No. markers segregating at 1:2:1 at $P \geq 1.0 \times 10^{-9}$	3607	1419	2901	2935	3941
No. markers in anchor maps	82	90	94	140	29
Length of anchor maps	561.9	696.2	578.2	584.2	374.5
No. total mapped loci	662	363	607	787	996
- Mapped non-distorted loci	160	132	178	262	182
- Mapped distorted loci	502	248	517	525	814
Total map length (cM)	1419.1	1327.6	1546.8	1454.0	1599.8
Average inter-marker distance (cM)	2.1	3.5	2.3	1.8	1.6
Number of gaps >10.0 cM	13	33	29	21	15
Largest gap (cM)	22.3	40.0	26.0	25.4	29.0

Pop1, ICP 11605 × ICP 14209; Pop2, ICP 8863 × ICP 11605; Pop3, HPL 24 × ICP 11605; Pop4, ICP 5529 × ICP 11605; Pop5, ICP8863 × ICPL 87119; † Information obtained from a separate project (S. Parupalli, Pers. Comm., 2016).

4.3.4 Construction of individual genetic maps

From a total of 15728, 7494, 12030 and 12654 SNPs identified, 3607, 1419, 2901 and 2935 SNPs in Pop1, Pop2, Pop3 and Pop4, respectively, segregated in 1:2:1 F_2 genotypic ratio at a χ^2 cutoff $P \geq 10^{-9}$, and were retained for genetic mapping (Table 4.4). Owing to high segregation distortion from the expected F_2 segregation ratio, markers segregating in a 1:2:1 ratio at $P>0.05$ were used as base or anchor markers for initial genetic map construction. As a result, a total of 82, 90, 94 and 142 markers in Pop1, Pop2, Pop3 and Pop4, respectively, which segregated in a 1:2:1 expected F_2 ratio at $P \geq 0.05$ could be mapped in the base or anchor genetic map (Table 4.4). A further 580, 273, 513 and 647 markers, which segregated in 1:2:1 ratio at $P<0.05 \geq 10^{-9}$ could be added to the base map resulting in 662, 363, 607 and 787 markers mapped, with map lengths of 1419.1, 127.6, 1546.8 and 1454.0 in Pop1, Pop2, Pop3 and Pop4, respectively. The average inter-marker distance, respectively, were 2.1, 3.5, 2.3 and 1.8 cM (Table 4.4). The number of gaps larger than 10.0 cM ranged from 13 in Pop1 to 33 (Pop2). The largest gaps on the maps ranged from 22.3 cM in Pop1 to 40 cM in Pop2 (Table 4.4). The graphical representations of the individual genetic maps are presented in Appendix 5 to Appendix 9.

4.3.5 Consensus genetic map

Combining maps from multiple independent crosses has the advantage of increasing the genetic diversity that is captured in the map, increasing support for marker order and position, and allowing markers from a single map to be placed relative to other markers (ICGMC, 2015). All markers used in the construction of the consensus map in the present study were SNPs except for one deletion marker on CcLG03 in Pop4. As a result, there was no disagreement in marker names among the individual maps. Segregation data for 3400 markers from five mapping populations was used to integrate the multiple genetic maps into a consensus map (Table 4.5).

Table 4.5: Number of common markers among individual mapping populations

			Number of markers common to 'n' mapping pop				Total common	
Population	Size	Total	n = 0	n=1	n=2	n=3	Number	%
Pop1	178	647	413	141	50	43	234	36.2
Pop2	175	363	170	107	44	42	193	53.2
Pop3	157	607	356	150	58	43	251	41.4
Pop4	179	787	532	157	58	40	255	32.4
Pop5	137	996	915	62	17	2	81	8.1
Total	826	3400	2386	617	227	170	1014	29.8

Pop1, ICP 11605 × ICP 14209; Pop2, ICP 8863 × ICP 11605; Pop3, HPL 24 × ICP 11605; Pop4, ICP 5529 × ICP 11605; Pop5, ICP 8863 × ICPL 87119.

Among the markers, 2386 were unique to individual mapping populations, 617 were common between two, 227 among three and 170 among four mapping populations (Table 4.5). The common markers were used as anchor points for integration of the individual maps. Most of the genetic linkage groups of the individual maps were integrated into the consensus map. Details of the consensus map and markers contributed from the different individual maps are given in Table 4.6. All common markers together led to the production of a consensus map comprising 984 marker loci on 11 linkage groups covering a map distance of 1609.5 cM with an average inter-marker distance of 1.6 cM (Table 4.6; Appendix 10).

4.3.6 Collinearity of component maps with consensus map

All maps were, to a large extent, collinear with the consensus map (Table 4.6, Fig. 4.1; Fig. 4.2). However, component LGs from Pop1 (CcLG02 and CcLG09) and Pop3 (CcLG02, CcLG04, CcLG06, CcLG07 and CcLG09) showed a reversal of marker order between component map and consensus map as revealed by the negative correlation coefficients ("*r*"; Table 4.6). Similarly, LGs from Pop5 that contributed any markers to the consensus map displayed poor collinearity with the consensus map. Finally, genome-wide, there were 13 gaps larger than 10 cM (one each on CcLG02 and CcLG11, two each on CcLG05, CcLG09 and CcLG10, and three each on CcLG03 and

CcLG07). Such gaps have been thought to result from recombination hotspots or regions that are identical-by-descent and thus lack polymorphisms (ICGMC, 2015).

Table 4. 6 Summary of a pigeonpea consensus genetic map constructed from five component genetic maps

Consensus map				Number of markers contributed from component genetic maps and their correlation with consensus map									
				Pop1		Pop2		Pop3		Pop4		Pop5	
LG†	n	ML§ (cM)	AID‡ (cM)	n	"*r*"	n	"*r*"	n	"*r*"	n	"*r*"	n	"*r*"
CcLG01	52	136.8	2.6	-	-	11	0.97***	13	0.81***	20	0.91***	1	-
CcLG02	219	224.3	1.0	9	0.80**	12	0.97***	13	0.99***	30	0.95***	172	0.23**
CcLG03	46	162.0	3.5	25	0.95***	15	0.97***	13	0.97***	22	0.95**	-	-
CcLG04	29	49.6	1.7	11	0.89***	4	0.99***	3	1.00*	18	0.57*	-	-
CcLG05	24	140.1	5.8	7	1.00***	4	0.99**	10	1.00***	13	0.87***	-	-
CcLG06	76	139.6	1.8	36	0.34*	23	0.98***	27	0.95***	48	0.94***	-	-
CcLG07	26	133.1	5.1	10	0.84**	8	0.95***	5	1.00***	8	0.73*	5	0.14^{NS}
CcLG08	34	119.3	3.5	16	0.98***	-	-	13	0.97***	24	0.98***	-	-
CcLG09	19	96.0	5.1	12	0.42^{NS}	6	0.99***	9	0.91***	8	0.95***	-	-
CcLG10	95	205.1	2.2	11	0.99***	3	1.00***	16	0.93***	8	0.99***	2	-
CcLG11	364	203.8	0.6	55	0.66***	47	0.90***	82	0.37***	102	0.12^{NS}	173	0.00^{NS}
Total	984	1609.5	1.6	192		133		204		301		352	

^{NS}, not significantly different from zero at 0.05 probability level; *, ** and *** significantly different from zero at 0.05, 0.01 and 0.001 probability levels, respectively. Pop1, ICP 11605 × ICP 14209; Pop2, ICP 8863 × ICP 11605; Pop3, HPL 24 x ICP 11605; Pop4, ICP 5529 × ICP 11605; Pop5, ICP 8863 × ICPL 87119; LG, linkage group; n, number of markers; ML, map length; AID, average inter-marker distance; "*r*", correlation coefficient.

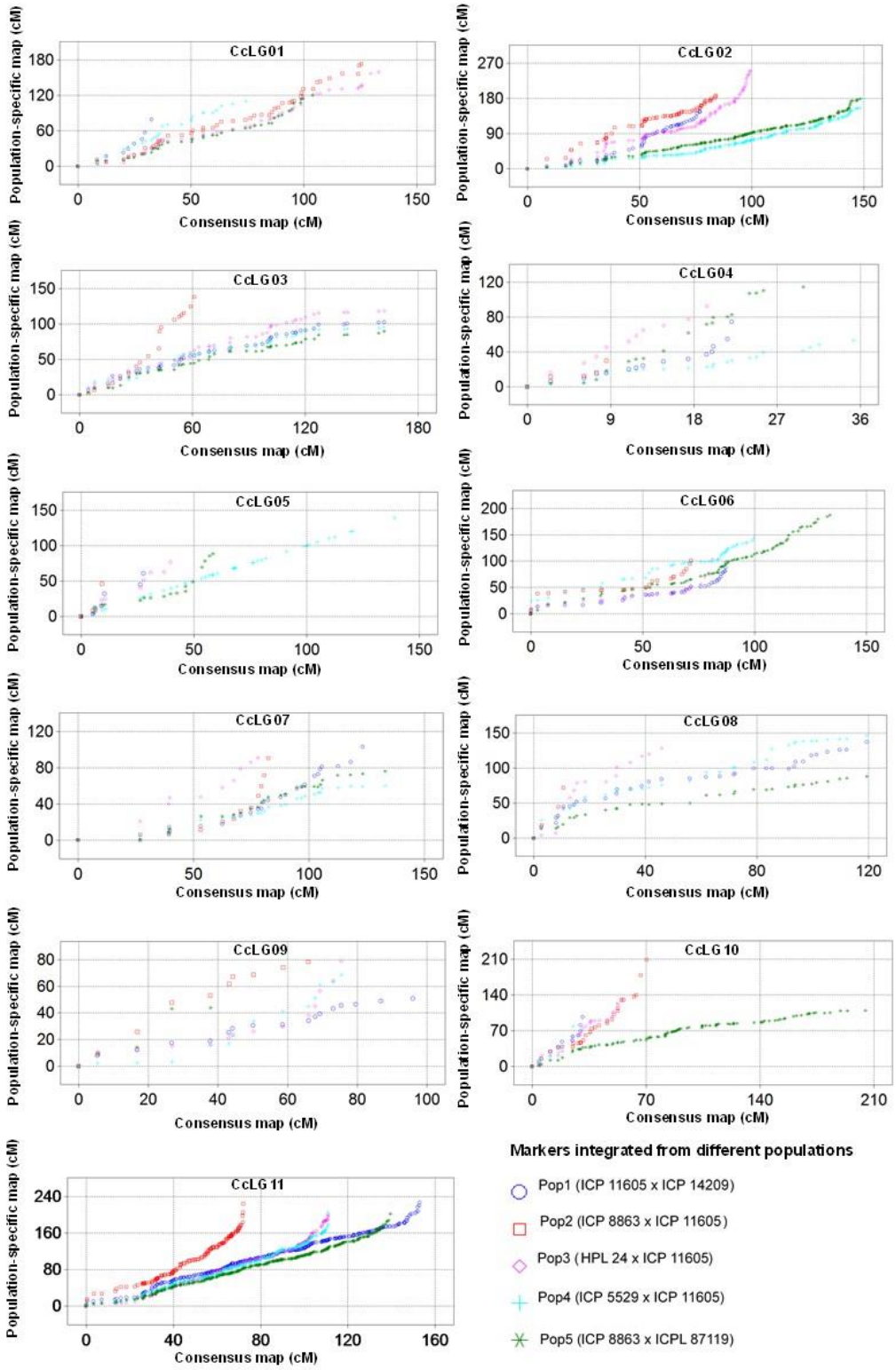


Fig. 4.1: Scatter plots showing the extent of correlations among population-specific and consensus genetic maps of pigeonpea

.

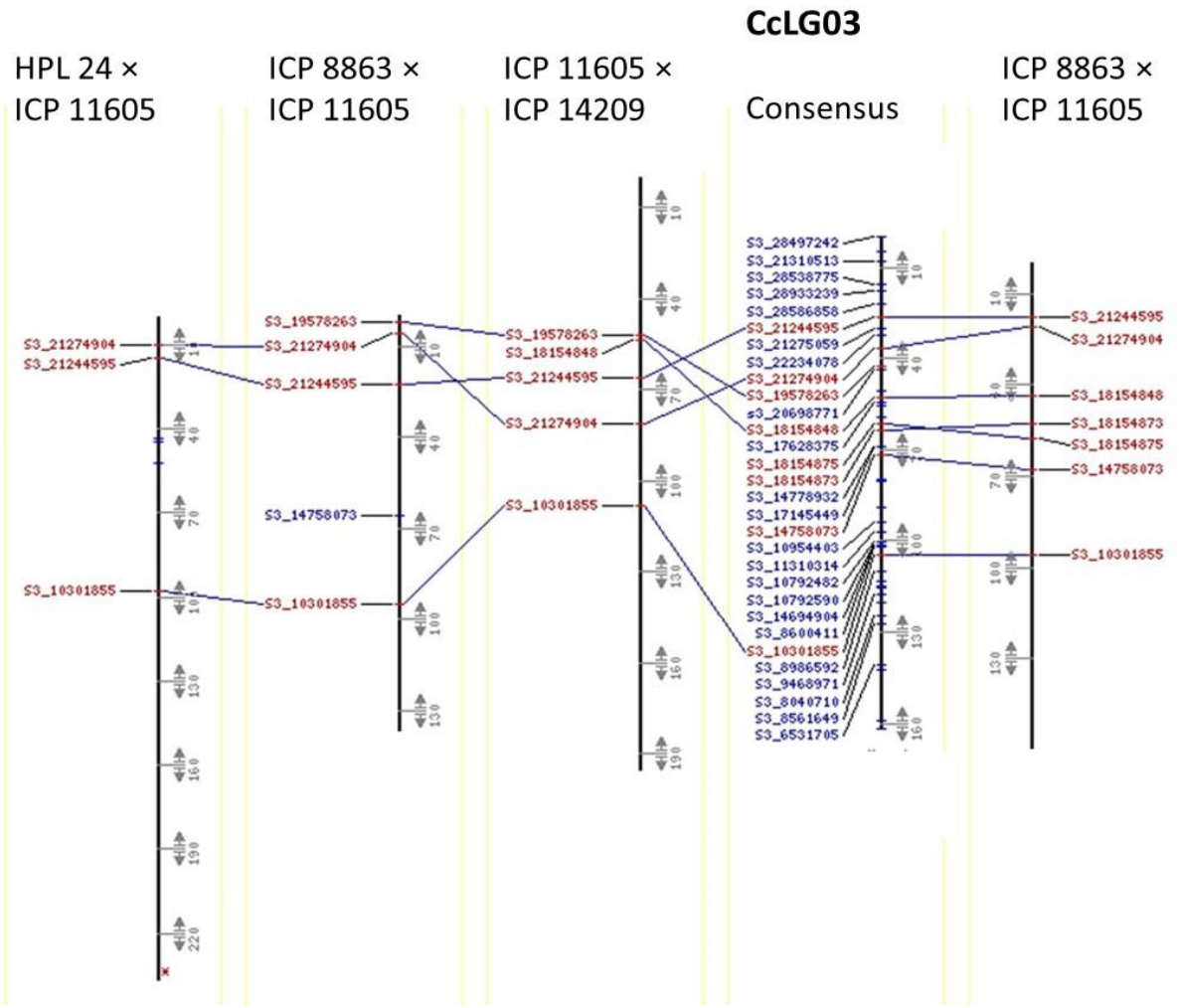


Fig. 4.2: A chart depicting marker-based correspondences of consensus with individual genetic maps, a case of CcLG03. Only common markers are included to visually asses the collinearity of marker orders and marker positions. Linkage groups are aligned together using comparative mapping programme CMap version 1.01. This figure and for all the other genetic linkage groups are presented as Appendix 11.

4.3.7 Main effect QTLs (M-QTLs) for seed protein content and colocalisation with M-QTLs for other traits

Main effect QTLs (M-QTLs) for SPC detected by both CIM and ICIM for all traits and individual populations are summarized in Table 4.7. Details of all M-QTLs detected for SPC are presented in Table 4.9 and Appendix 5 to Appendix 9. Similarly, M-QTLs for DTF, GH, SW and SY which co-localise with that of SPC are also presented in Table 4.9 and Appendix 5 to Appendix 9. It is worth noting at this point that the coding of parental alleles was done with respect to SPC as the core trait in the present investigation. Accordingly, positive additive

effect represents the difference in direction of SPC of the 'ICP 11605' parent in Pop1 and Pop4, 'ICP 8863' in Pop2, 'HPL 24' in Pop3 and 'ICPL 87119' in Pop5. In contrast, a negative additive effect represents a difference in the direction of 'ICP 14209' parent in Pop1, 'ICP 11605' in Pop2 and Pop3, 'ICP 5529' in Pop4 and 'ICP 8863' in Pop5.

In Pop1, five M-QTLs were detected by CIM and one by ICIM with PVE ranging from 7.8% to 16.6% resulting in a total of six M-QTLs (Table 4.7; Table 4.8; Appendix 5). Among the M-QTLs, *qPROT-cim-3.1*, *qPROT-cim-11.1* and *qPROT-cim-11.2* were major, explaining 10.3%, 13.8% and 16.6%, respectively, of the total within-population SPC variance (Table 4.7; Table 4.8). Based on the coding of parental alleles, the positive additive effect of qPROT-cim-3.1 indicated it was contributed by the low SPC parent (ICP 11605). All other M-QTLs showed negative additive effects, indicating they were contributed by the high SPC parent (ICP 14209). Two SPC M-QTLs (*qPROT-cim-3.1*; PVE = 7.8% and *qPROT-cim-3.2*; PVE = 10.3%) overlapped with M-QTLs for GH (*qGH-icim-3.2*; PVE = 8.6%) and DTF (*qDTFF-cim-3.3*; PVE = 7.4%), respectively (Table 4.9; Appendix 5). In addition, one minor M-QTL for SY (*qSY-cim-11.1*, PVE = 7.5%) was found 6.4 cM away from a major M-QTL for SPC (*qPROT-cim-11.1*, PVE = 16.6%) (Table 4.9; Appendix 5).

Table 4.7: Summary of main effect QTLs detected by composite interval mapping (CIM) and inclusive composite interval mapping (ICIM) for seed protein content, 100-seed weight, seed yield, days to first flower and growth habit in five F_2 mapping populations

		SPC†		SW§		SY¶		DTF‡		GH††	
Population	QTL features	CIM	ICIM	CIM	ICIM	CIM	ICIM	CIM	ICIM	CIM	ICIM
Pop1	No. QTLs‡‡	5(3)	1(0)	4(2)	4 (1)	6(1)	1(1)	6(2)	1(1)	3(1)	3(1)
	LOD	2.6-3.8	2.7	2.6-4.3	2.9-6.8	2.6-3.0	4.5	2.5-9.5	11.1	3.2-13.9	3.2-17.3
	PVE (%)	7.8-16.6	8.6	3.6-12.3	6.4-15.0	5.2-15.4	10.2	5.7-20.3	25.4	10.9-91.3	6.1-41.3
	Total No. QTLs#	6		8		6		6		4	
Pop2	No. QTLs	4(1)	4(1)	2(0)	2 (1)	4(2)	6 (2)	6(4)	3 (1)	5(4)	2(1)
	LOD	2.6-3.8	2.7-2.9	3.4-4.6	2.7-9.9	3.0-5.3	3.5-8.0	2.9-9.9	2.6-11.7	2.7- 16.1	12.4-15.0
	PVE (%)	0.7-12.8	6.9-12.3	6.1-7.5	8.4-29.1	1.7 -11.8	5.9-16.0	4.0-36.3	4.5-26.6	4.0-64.7	23.9-25.4
	Total No. QTLs	6		3		8		7		5	
Pop3	No. QTLs	6(1)	8(1)	5(3)	2(2)	8(4)	6 (1)	3(3)	6 (2)	5(1)	4(2)
	LOD	3.0-4.6	2.5-4.2	2.5-13.6	7.7-13.7	2.5-5.4	2.5-6.5	4.4-16.0	2.6-20.0	3.0-25.3	3.0-31.4
	PVE (%)	3.8-23.5	5.1-10.0	5.4-44.4	16.3-5.7	4.8-40.2	4.5-20.3	13.2-40.3	3.4-31.9	5.3-13.3	4.0-54.4
	Total No. QTLs	13		5		12		9		9	

Pop1, ICP 11605 × ICP 14209; Pop2, ICP 8863 × ICP 11605; Pop3, HPL 24 × ICP 11605, Pop4, ICP 5529 × ICP 11605, Pop5, ICP 8863 × ICPL 87119. † Seed protein content; § 100-seed weight; ¶ Seed yield; ‡ Number of days to first flowering; †† Growth habit. PVE; phenotypic variation explained by a QTL; Number in parenthesis represents numbers of major (PVE% ≥ 10.0%) M-QTLs. ‡‡ Number of QTLs detected by either CIM or ICIM; # Number of unique QTLs detected by either or both CIM and ICIM; LOD, Logarithm of odds ratio

Table 4.7: (continued)

		SPC		SW		SY		DTF		GH††	
Population	QTL features	CIM	ICIM	CIM	ICIM	CIM	ICIM	CIM	ICIM	CIM	ICIM
Pop4	No. QTLs	7(2)	5(2)	2(1)	4(1)	1(0)	4(1)	4(3)	3 (1)	8(5)	5(4)
	LOD	2.6-5.1	3.5-7.2	3.4-14.7	2.7-15.0	2.9	3.2-4.2	4.0-7.8	2.9-6.6	2.8-22.1	2.7-29.1
	PVE (%)	3.3-17.5	7.7-16.5	8.3-10.4	6.3-31.5	8.2	6.6-14.8	4.6-47.6	6.1-12.6	3.4-47.0	3.9-61.6
	Total No. QTLs	11		5		5		5		11	
Pop5	No. QTLs	10(2)	4(2)	6(5)	4(2)	5(4)	6(3)	16(9)	7(5)	-	-
	LOD	2.5-4.2	2.7-7.5	2.6-4.8	2.5-4.4	3.0-4.0	2.6-3.7	2.6-4.5	2.9-6.8	-	-
	PVE (%)	1.7-16.3	8.2-18.9	8.7-26.7	4.9-13.1	6.7-53.0	5.8-10.7	2.1-43.8	6.3-15.2	-	-
	Total No. QTLs	12		9		10		21			

Table 4.8: QTLs for seed protein content identified using composite interval mapping (CIM) and inclusive composite interval mapping (ICIM) in five F_2 populations of pigeonpea

QTL name	Position (cM)	Flanking markers	QTL interval (cM)	LOD	PVE (%)	[a]†	[d]§
Pop1 (ICP 11605 × ICP 1409)							
qPROT-cim-3.1	81.4	S3_12113347 - S3_21274904	1.7	2.8	7.8	0.4	-0.2
qPROT-cim-3.2	**89.7**	**S3_23699007 - S3_18226407**	**2.6**	**2.6**	**10.3**	**-0.5**	**0.4**
qPROT-icim-3.1	99.0	S3_25565937 - S3_16199983	0.4	2.7	8.6	-0.5	0.4
qPROT-cim-4.1	32.5	S4_1586438 - S4_5314034	5.1	2.9	9.0	-0.3	0.5
qPROT-cim-11.1	**43.0**	**S11_11249294 - S11_9768899**	**2.5**	**3.5**	**16.6**	**-0.6**	**0.5**
qPROT-cim-11.2	**170.1**	**S11_20646423 - S11_24857528**	**1.4**	**3.8**	**13.8**	**-0.5**	**0.3**
Pop2 (ICP 8863 × ICP 11605)							
qPROT-cim-1.1	20.2	S1_15372966 - S1_5944791	1.1	2.6	7.7	0.4	-0.3
qPROT-icim-3.1	2.0	S3_22234078 - S3_19578263	2.2	2.7	6.9	0.3	-0.5
qPROT-cim-3.1	**55.9**	**S3_17193829 - S3_14758073**	**11.8**	**3.8**	**12.8**	**0.7**	**-0.3**
qPROT-icim-3.2	56.0	S3_17193829 - S3_14758073	11.8	2.9	7.4	0.6	-0.3
qPROT-cim-9.1	67.4	S9_10003418 - S9_10229309	1.4	3.6	5.7	0.1	-0.8
qPROT-icim-11.1	**46.0**	**S11_2019429 - S11_22353396**	**2.0**	**2.7**	**12.3**	**-0.7**	**0.5**
qPROT-icim-11.2	126.0	S11_21940736 - S11_30337876	1.0	2.9	9.4	0.6	0.4
qPROT-cim-11.1	126.1	S11_30337876 - S11_45761666	1.6	3.4	0.7	0.4	0.5
Pop3 (HPL 24 × ICP 11605)							
qPROT-icim-2.1	20.0	S2_17395609 - S2_17836619	5.7	3.7	7.9	0.6	-0.1

LOD, logarithm of the odds ratio; PVE, phenotypic variation explained by a QTL. † Additive effect; § Dominance effect. Major effect QTLs (PVE ≥ 10.0%) are in bold font.

Table 4.8: (continued)

QTL name	Position (cM)	Flanking markers	QTL interval (cM)	LOD	PVE (%)	[a]†	[d]§
qPROT-cim-2.1	**28.6**	**S2_18621223 - S2_5077845**	**4.1**	**4.1**	**23.5**	**0.9**	**-0.6**
qPROT-icim-2.2	**67.0**	**S2_17642300 - S2_27324059**	**1.6**	**3.4**	**10.0**	**0.8**	**0.0**
qPROT-cim-2.2	67.5	S2_27324059 - S2_27324056	0.1	3.1	6.0	0.6	0.0
qPROT-icim-3.1	10.0	S3_28538775 - S3_21274904	10.0	3.0	5.8	-0.6	-0.1
qPROT-cim-3.1	40.8	S3_17145449 - S3_18154873	3.4	4.4	4.1	-0.6	-0.5
qPROT-cim-3.2	46.7	S3_18154848 - S3_17193829	3.4	4.6	3.8	-0.6	-0.6
qPROT-cim-3.3	55.1	S3_18154875 - S3_14813065	4.7	3.2	5.6	-0.7	-0.2
qPROT-icim-3.2	102.0	S3_6531705 - S3_24127268	1.8	2.5	6.7	-0.4	-0.6
qPROT-icim-6.1	75.0	S6_14548839 - S6_6094182	7.0	3.8	9.2	0.6	-0.5
qPROT-icim-10.1	21.0	S10_6745618 - S10_18754549	20.7	2.8	5.1	-0.2	-0.7
qPROT-cim-11.1	82.6	S11_17781645 - S11_39391791	0.3	3.0	7.4	0.4	-0.6
qPROT-icim-11.1	119.0	S11_45315652 - S11_32081128	0.7	2.7	9.5	0.7	0.7
qPROT-icim-11.2	123.0	S11_7540489 - S11_21960241	0.4	4.2	8.5	0.8	1.0
Pop4 (ICP 5529 × ICP 11605)							
qPROT-cim-1.1	46.6	S1_1798648 - S1_1798766	2.6	2.6	4.4	0.3	0.1
qPROT-cim-2.1	27.8	S2_6930418 - S2_16133939	0.3	3.0	9.3	-0.5	0.2
qPROT-icim-2.1	**34.0**	**S2_10279728 - S2_32698493**	**0.2**	**6.6**	**16.5**	**-0.4**	**0.4**
qPROT-cim-2.2	**34.0**	**S2_10279728 - S2_32698493**	**0.2**	**5.1**	**17.5**	**-0.7**	**0.3**
qPROT-cim-2.3	38.3	S2_28049603 - S2_9984747	0.1	3.1	9.0	-0.5	0.1

Table 4.8 (continued)

QTL name	Position (cM)	Flanking markers	QTL interval (cM)	LOD	PVE (%)	[a]†	[d]§
qPROT-cim-2.4	**102.9**	**S2_4297468 - S2_13394656**	**0.4**	**4.0**	**11.8**	**-0.5**	**0.4**
qPROT-icim-2.2	**121.0**	**S2_16519107 - S2_16348673**	**4.9**	**7.2**	**11.5**	**-0.2**	**-0.5**
qPROT-icim-3.1	77.0	S3_23614170 - S3_8195933	2.3	4.9	9.0	0.6	0.1
qPROT-icim-6.1	85.0	S6_3630897 - S6_11140261	0.5	3.5	7.7	-0.6	0.2
qPROT-cim-11.1	82.2	S11_21017392 - S11_9883313	0.2	3.2	7.2	-0.5	0.6
qPROT-icim-11.1	91.0	S11_38887609 - S11_29943293	0.7	3.8	7.9	-0.5	0.2
qPROT-cim-11.2	95.4	S11_31519207 - S11_30807675	0.5	3.0	3.3	-0.4	0.5
Pop5 (ICP 8863 × ICPL 87119)							
qPROT-cim-2.1	41.6	S2_16460899 - S2_2144739	1.3	3.0	1.7	0.3	0.5
qPROT-icim-2.1	42.0	S2_16460899 - S2_2144739	1.3	2.7	8.5	0.4	0.5
qPROT-icim-2.2	**178.0**	**S2_9426717 - S2_24073225**	**2.2**	**7.5**	**18.9**	**0.2**	**1.3**
qPROT-icim-6.1	**83.0**	**S6_20608121 - S6_12302413**	**5.3**	**4.3**	**10.6**	**0.1**	**-0.8**
qPROT-cim-6.1	**83.8**	**S6_12302413 - S6_1292942**	**4.4**	**4.2**	**16.3**	**0.5**	**-0.6**
qPROT-cim-6.2	154.4	S6_11344426 - S6_1641606	7.3	3.3	8.3	-0.1	0.8
qPROT-cim-7.1	**122.7**	**S7_462935 - S7_1601723**	**1.6**	**3.3**	**11.3**	**-0.2**	**0.9**
qPROT-cim-8.1	61.6	S8_1988786 - S8_19001660	1.9	2.5	6.7	0.2	-0.7
qPROT-icim-8.1	70.0	S8_9578163 -S8_4817489	0.9	3.8	8.2	-0.1	0.7
qPROT-cim-10.1	95.8	S10_17537652 - S10_632595	3.2	2.6	1.7	0.4	1.1
qPROT-cim-11.1	2.0	S11_38211354- S11_36725317	5.4	3.0	8.1	-0.2	0.9
qPROT-cim-11.2	84.4	S11_20139207 - S11_25774244	0.2	4.1	9.1	-0.1	0.9
qPROT-cim-11.3	100.4	S11_24067221 - S11_26654392	0.2	2.6	1.8	0.2	0.7
qPROT-cim-11.4	137.2	S11_18198760 - S11_11799702	1.0	2.7	2.0	0.5	0.3

Table 4.9: Colocalised main effect QTLs for seed protein content with that of days to first flowering, growth habit, 100-seed weight and seed yield in four F_2 mapping populations of pigeonpea

QTL-name	Position (cM)	Flanking markers	QTL interval (cM)	LOD	PVE (%)	[a]†	[d]§
Pop1 (ICP 11605 × ICP 14209)							
qPROT-cim-3.1	**81.4**	**S3_12113347 - S3_21274904**	**1.7**	**2.8**	**7.8**	**0.4**	**-0.2**
qGHT-icim-3.2	79.0	S3_14778845 - S3_12113347	1.0	2.9	8.6	0.1	-0.2
qDTF-cim-3.2	75.5	S3_14813065 - 3_14778845	4.2	3.1	7.4	-4.4	-0.3
qGHT-cim-3.2	78.5	S3_14813065 - S3_14778845	4.2	4.1	12.9	0.1	-0.1
qPROT-cim-3.2	**89.7**	**S3_23699007 - S3_18226407**	**2.6**	**2.6**	**10.3**	**-0.5**	**0.4**
qDTF-cim-3.3	91.0	S3_18226407 - S3_5582712	0.8	2.9	5.7	4.3	-0.4
Pop2 (ICP 8863 × ICP 11605)							
qPROT-cim-1.1	**20.2**	**S1_15372966- S1_5944791**	**1.1**	**2.6**	**7.7**	**0.4**	**-0.3**
qHSW-cim-1.1	22.2	S1_5944791 - S1_9033631	9.9	3.4	7.5	0.7	0.2
qPROT-cim-3.1	**55.9**	**S3_17193829 - S3_14758073**	**11.8**	**3.8**	**12.8**	**0.7**	**-0.3**
qGHT-cim-3.1	51.4	S3_18929378- S3_17193829	8.5	9.8	64.7	0.2	-0.4
qPROT-cim-11.2	**126.1**	**S11_30337876 - S11_45761666**	**1.6**	**3.4**	**0.7**	**0.4**	**0.5**
qDTF-cim-11.1	128.3	S11_45761666 - S11_18137395	0.6	3.0	8.1	2.9	-3.6
qPROT-icim-3.1	**2.0**	**S3_22234078 - S3_19578263**	**2.2**	**2.7**	**6.9**	**0.3**	**-0.5**
qSY-cim-3.1	2.0	S3_22234078 - S3_19578263	2.2	8.0	16.0	-1.6	21.1

LOD, logarithm of the odds ratio; PVE, phenotypic variation explained by a QTL. † Additive effect; § Dominance effect. QTLs in bold face are for seed protein content which co-localise with QTLs for the other traits.

Table 4.9: (continued)

QTL-name	Position (cM)	Flanking markers	QTL interval (cM)	LOD[b]	PVE (%)[c]	[a][d]	[d][e]
qPROT-cim-3.2	**46.7**	**S3_18154848 - S3_17193829**	**3.4**	**4.6**	**3.8**	**-0.6**	**-0.6**
qDTFF-cim-3.1	46.7	S3_18154848 - S3_17193829	3.4	6.9	40.3	3.8	11.1
qGH-cim-3.1	46.7	S3_18154848 - S3_17193829	3.4	13.1	13.3	-0.3	-0.4
qPROT-cim-3.3	**55.1**	**S3_18154875 - S3_14813065**	**4.7**	**3.2**	**5.6**	**-0.7**	**-0.2**
qDTFF-cim-3.2	54.1	S3_18154875 - S3_14813065	4.7	4.4	39.7	-1.8	11.3
qGH-cim-3.2	55.1	S3_18154875 - S3_14813065	4.7	7.8	5.8	-0.1	-0.4
qDTFF-icim-3.2	36.0	S3_17628375 - S3_17145449	6.4	20.0	31.9	8.9	11.0
qPROT-cim-3.1	**40.8**	**S3_17145449 - S3_18154873**	**3.4**	**4.4**	**4.1**	**-0.6**	**-0.5**
qPROT-cim-11.1	**82.6**	**S11_17781645 - S11_39391791**	**0.3**	**3.0**	**7.4**	**0.4**	**-0.6**
qSY-cim-11.2	82.5	S11_6081367 - S11_45330880	0.1	3.0	11.3	6.6	-7.2
Pop5 (ICP 8863 × ICPL 87119)							
qPROT-cim-2.1	**41.6**	**S2_16460899- S2_2144739**	**1.3**	**3.0**	**1.7**	**0.3**	**0.5**
qDTFF-cim-2.1	41.0	S2_2989918-S2_16460899	0.6	4.4	4.2	-0.1	6.3
qPROT-cim-11.4	**137.2**	**S11_18198760 - S11_11799702**	**1.0**	**2.7**	**2.0**	**0.5**	**0.3**
qDTFF-cim-11.1	137.4	S11_22689703 - S11_19044341	0.9	3.3	5.0	0.2	5.4

In Pop2, CIM and ICIM detected four M-QTLs each for SPC with two of the M-QTLs (*qPROT-cim-3.1*/*qPROT-icim-3.2* and *qPROT-cim-11.1*/*qPROT-icim-11.2*) being common between the two methods resulting in a total of six M-QTLs (Table 4.7; Table 4.8; Appendix 6). Each method detected one major effect (PVE ≥ 10.0%) M-QTL but on different LGs (Table 4.7; Table 4.8). The major M-QTLs were *qPROT-cim-3.1* and *qPROT-icim-11.1* with PVE of 12.8% and 12.3%, respectively (Table 4.8). Of the M-QTLs, *qQPROT-cim-3.1* with PVE of 12.8% was also identified as *qPROT-icim-3.2* with PVE = 7.4% (Table 4.8; Appendix 6). All M-QTLs for SPC in this population showed positive additive effects except for *qPROT-icim-3.1*, which was negative (Table 4.8) indicating that all SPC increasing alleles were contributed by the high SPC parent, ICP 8863.

There were three regions of co-localisation between M-QTLs for SPC and M-QTLs of agronomic traits in Pop2 (Table 4.9; Appendix 6). One M-QTL for SPC (*qPROT-cim-1.1*; PVE = 7.7%) co-localised with M-QTL for SW (*qHSW-cim-1.1*; PVE = 6.1%) on CcLG01. Two M-QTLs for SPC on CcLG03 (*qPROT-icim-3.1*; PVE = 6.9% and *qPROT-cim-3.1*/*qPROT-icim-3.2*, PVE = 12.8%) co-localised with M-QTLs for SY and GH (*qSY-icim-3.1*, PVE = 16.0% and *qGH-cim-3.4*, PVE = 64.7%), respectively. One M-QTL for SPC (*qPROT-cim-11.1* and *qPROT-icim-11.2* with PVE of 0.7% and 9.4%, respectively) also overlapped with a minor M-QTL for DTF (*qDTFF-cim-11.1*, PVE = 4.0%) on CcLG11.

In the case of Pop3, CIM and ICIM detected six and eight M-QTLs for SPC, respectively. One M-QTL (*qPROT-cim-2.2*) was common but with different PVE values between the methods resulting in a total of 13 M-QTLs (Table 4.7; Table 4.8; Appendix 7). There were two major effect M-QTLs with PVE of 23.5% (*qPROT-cim-2.1*) and 10.0% (*qPROT-icim-2.2*), which, together with one minor M-QTL (*qPROT-cim-11.1*, PVE = 7.4%), showed positive additive effects indicating that the high SPC parent (HPL 24) allele contributed to increased SPC (Table 4.8). The remaining ten minor effect SPC M-QTLs showed negative additive effects indicating the low SPC parent (ICP 11605) allele contributed to decreased SPC (Table 4.8).

Most of the M-QTL overlaps between SPC and the agronomic traits in Pop3 were on CcLG03 involving mainly minor effect M-QTLs for SPC (PVE = 3.8 to 5.6%) (Table 4.9; Appendix 7). Among these are *qPROT-cim-3.2* (PVE = 3.8%) which co-localised with *qDTFF-cim-3.1* (PVE = 40.3%) and *qGH-cim-3.1* (PVE = 13.3%). Another M-QTL for SPC (*qPROT-cim-3.3*, PVE = 5.6%) also co-localised with M-QTLs for DTF (*qDTFF-cim-3.2*, PVE = 39.7%) and GH (*qGH-cim-3.2*, PVE = 5.8%). A minor M-QTL for SPC on CcLG11 (*qPROT-cim-11.1*, PVE = 7.4%) was 0.1 cM away from a major M-QTL for SY (qSY-cim-11.2, PVE = 11.3%).

For Pop4, CIM and ICIM detected seven and five M-QTLs each for SPC with PVE ranging from 3.3 to 17.5% and 7.7 to 16.5%, respectively (Table 4.7; Table 4.8; Appendix 8). One major M-QTL was detected by both CIM (*qPROT-cim-2.2*, PVE = 16.5) and ICIM (qPROT-icim-2.1, PVE = 17.5) (Table 4.7 Appendix 8), thus the total number of M-QTLs was 11 between the two methods (Table 4.7). Each method detected at least two major effect M-QTLs with PVE ranging from 11.5 to 17.5% (Table 4.7; Table 4.8). The additive effects were negative for all except one minor M-QTL (*qPROT-cim-2.1*, PVE = 4.4%). The negative M-QTLs' additive effects in this population indicates that the SPC increasing alleles were contributed by the high protein parent (ICP 5529) while positive effects indicates the SPC increasing QTL allele was contributed by the low protein parent (ICP 11605). Neither CIM nor ICIM detected any overlap between M-QTLs for SPC and that of any of the other traits.

For Pop5, CIM and ICIM detected 10 and four M-QTLs with PVE ranging from 1.7 to 16.3% and 8.2 to 18.9%, respectively (Table 4.7; Table 4.8; Appendix 9). Two of the M-QTLs were common between CIM and ICIM giving a total of 12 M-QTLs (Table 4.7; Table 4.8). In total three M-QTLs (*qPROT-icim-2.2*, q*PROT-cim-6.1*/*qPROT-icim-6.1*, and *qPROT-cim-7.1*) were of major effect with PVE ≥ 10% (Table 4.8). Seven of the M-QTLs had positive additive effects indicating the SPC increasing alleles was contributed by the high SPC parent (ICP 8863) whereas the remaining five had negative additive effects likely contributed by low SPC parent ICPL 87119 (Table 4.8). There was one overlap between M-QTL for SPC (*qPROT-cim-2.1*) and an M-QTL for DTF (*qDTFF-cim-2.1*), and one tight linkage (0.1 cM distance) between an M-QTL for SPC (*qPROT-cim-11.4*) and M-QTL for DTF (*qDTFF-cim-11.1*) (Table 4.9; Appendix 9).

4.3.8 Consensus genetic and main effect QTL maps across populations

Forty two out of a total 48 M-QTLs for SPC from the five mapping populations could be projected onto the consensus linkage map (Figure 4.3). Twenty seven of the projected SPC M-QTLs could be collapsed into six consensus QTL regions. The consensus QTLs contained M-QTLs from two populations (*Consensus QTL 1*, *Consensus QTL 2* and *Consensus QTL 5*), three populations (*Consensus QTL 3*) and four populations (*Consensus QTL* 4 and *Consensus QTL* 6).

In a similar manner, five QTL clusters could be recognized (*QTL Cluster 1* to *5*) (Figure 4.3). The five clusters included *QTL Cluster 1* on CcLG01 which harboured a QTL each for SPC and SW, *QTL Cluster 2* on CcLG03 (SPC and SY), and *QTL Cluster 4* and *QTL Cluster 5* on CcLG11 (SPC and DTF). The QTL cluster intervals ranged from 16.5 (*QTL Cluster 4*) to 35.9 cM (*QTL Cluster 3*). The most crowded cluster (*QTL Cluster 3*), contained QTLs for SPC from four of the five populations in addition to M-QTLs for DTF from one population and QTL for GH from two populations. However, individual M-QTLs for SPC in *QTL Cluster 3* accounted for only small proportions (4.1 to 9.0%) of the within-population SPC variations except for *Pop2qPROT-cim-3.1* with a PVE of 12.3%. While M-QTLs for DTF and GH that clustered with SPC QTLs in *QTL Cluster 3* accounted for 6.4 to 40.3% of the phenotypic variations (Figure 4.3). Most of the major (PVE ≥ 10%) M-QTLs for SPC could only be projected individually except *Pop1qPROT-cim-11.2* (PVE = 13.8%) found clustered with other M-QTLs in *QTL Cluster 5* on CcLG11 (Figure 4.3).

4.3.9 Epistatic QTLs (E-QTLs)

To gain some insight into the complexity of the genetic control of SPC and its relationship with other traits, epistatic QTLs (E-QTLs) were mapped in each of the five F2 populations using QTL Icimapping software v4.0. E-QTLs detected for all traits in each population are summarized in Table 4.10 while E-QTLs specific to SPC are presented in Table 4.11.

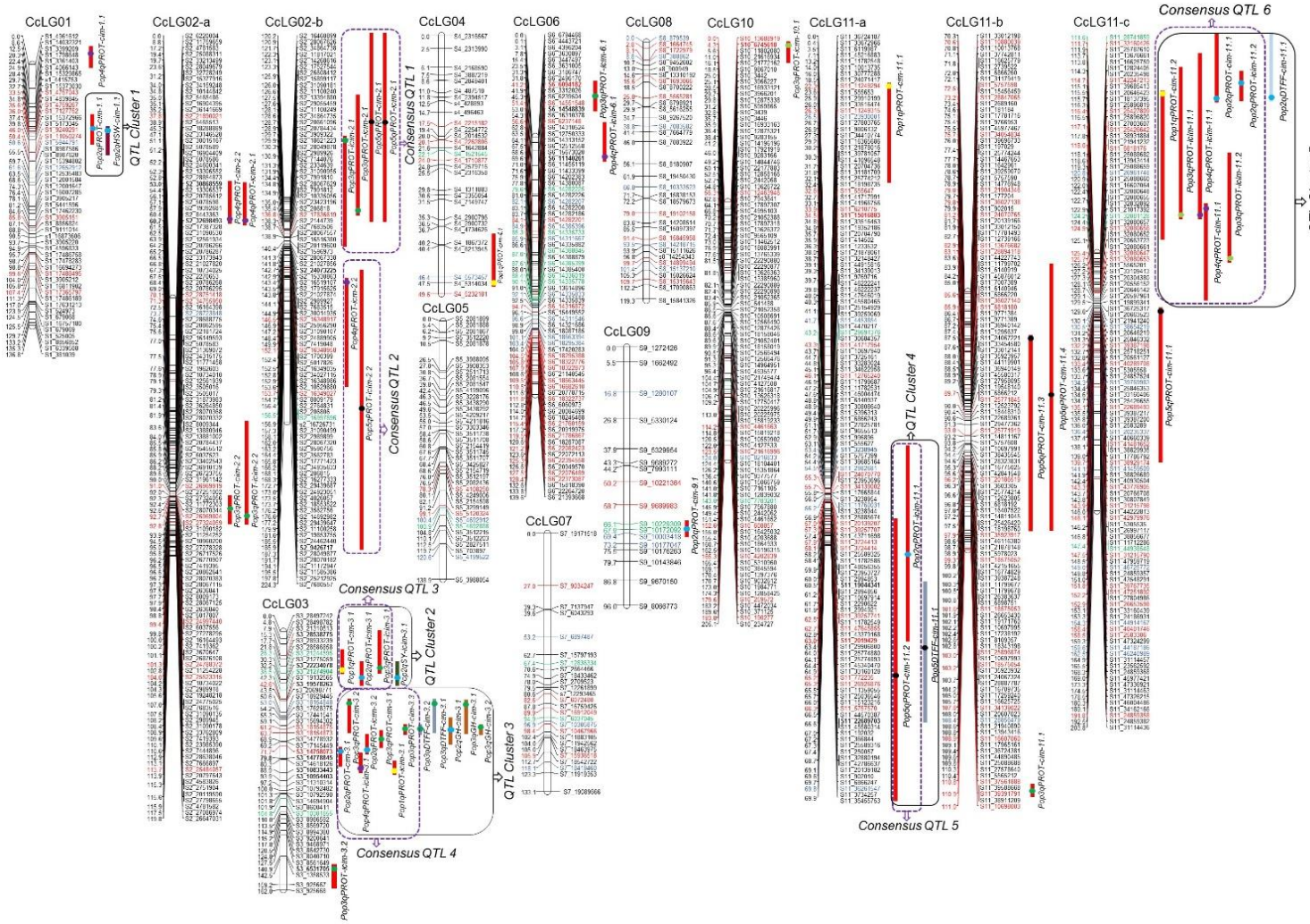


Fig. 4.3: Consensus genetic and QTL maps.

Markers are shown on right side of the linkage group while map distances are indicated on left side. Markers unique to mapping populations, common between two, three and four mapping populations have been shown by black, red, blue and green colours, respectively. QTLs for the different traits

are indicated by coloured bars with brown, green, grey, purple and red showing QTLs for GH, SY, DTF, SW and SPC, respectively. Source populations of projected QTLs are indicated by dots of different colours where yellow, blue, green, purple and black represent Pop1, Pop2, Pop3, Pop4 and Pop5, respectively. The position of the dot on the QTL bar indicates whether both or only one of the flanking markers were projected on to the consensus map. Where the dot lies at the centre of the QTL bar it indicates that both flanking markers of the QTL are present, while if the dot lies either on the upper or lower part of the QTL bar it shows that only one marker closest to the dot was projected.

Pop2 had the highest number E-QTLs (173) while Pop5 had the lowest number (52) across traits. Among traits, SPC on average had the lowest number of E-QTLs ranging from two in Pop2 to 11 in Pop1 while GH had the highest number ranging from 40 in Pop4 to 56 in Pop2 (Table 4.10). The E-QTLs were detected on all LGs in each population. None of the loci involved in the epistatic interactions showed independent effects. Overall, E-QTLs made large contributions to the phenotypic variations of the measured traits ranging from 6.3% for DTF in Pop1 to 99.4% for GH in Pop2 (Table 4.10). In case of SPC as the core trait in this study, E-QTLs accounted for up to 31.2% (Pop1), 69.8% (Pop2), 21.2% (Pop3 and Pop4) and 30.5% (Pop5) of the within-population SPC variations (Table 4.10; Table 4.11).

Table 4.10: Summary of epistatic QTLs detected for seed protein content, and some agronomic traits in five F_2 mapping populations of pigeonpea

	E-QTL features	SPC†	SW§	SY¶	DTF‡	GH#
Pop1	Number of E-QTLs	11(11)	8(8)	29(29)	5(5)	40(1)
	LOD	5.1 - 6.6	5.0-5.6	5.1 - 8.2	5.0-5.5	9.4 - 79.6
	PVE (%)	12.8 - 31.2	14.6 - 25.3	12.9-38.5	10.4 - 33.4	10.9 - 91 .3
Pop2	Number of E-QTLs	9(9)	63(63)	19(19)	26(26)	56(56)
	LOD	5.2 - 7.5	6.6–17.1	5.0–7.3	5.0 – 8.9	5.0 - 1132.5
	PVE (%)	55.0 – 69.8	29.8-41.8	10.6–36.4	14.8 – 44.3	10.4 - 99.4
Pop3	Number of E-QTLs	2(2)	53(53)	30(30)	10(6)	50(50)
	LOD	5.2 - 5.3	5.0–9.9	5.0-8.5	5.0 - 6.2	5.1 –41.8
	PVE (%)	19.3 - 21.2	14.6–39.8	14.6-37.1	6.3 - 14.6	14.1 – 96.0
Pop4	Number of E-QTLs	4(3)	20(20)	39(39)	5(5)	42(42)
	LOD	5.2-6.9	5.1-7.3	5.0-8.7	5.1-7.0	5.1-16.6
	PVE (%)	9.5-21.2	11.5-30.0	15.9-35.1	14.7-23.6	10.6-74.8
Pop5	Number of E-QTLs	8(7)	20(20)	12(12)	12(12)	-
	LOD	5.0-6.3	5.1-7.0	5.1-7.2	5.1-6.2	-
	PVE (%)	9.8-30.5	14.2-25.6	12.0-23.8	12.0-30.1	-

Pop1, ICP 11605 × ICP 14209; Pop2, ICP 8863 × ICP 11605; Pop3, HPL 24 × ICP 11605; Pop4, ICP 5529 × ICP 11605; Pop5, ICP 8863 × ICPL 87119. E-QTLs, epistatic QTLs; PVE, phenotypic variation explained. Number in parenthesis represents numbers of major (PVE% ≥ 10.0%) E-QTLs. † Seed protein content, § 100-seed weight, ¶ Seed yield; ‡Days to first flower; # Growth habit.

Table 4.11: Epistatic QTLs for seed protein content detected using inclusive composite interval mapping in five F_2 mapping populations of pigeonpea

Chr.1†	Pos.1§	Flanking markers (QTL1)	Chr.2†	Pos.2§	Flanking markers (QTL2)	LOD	PVE (%)	[aa]¶	[ad]‡	[dd]‡‡
Pop1 (ICP 11605 × ICP 14209)										
CcLG01	30	S1_2693194-S1_4757043	2	75	S2_7491873-S2_36672875	5.0	18.0	-0.4	-0.3	1.5
1	45	S1_4757043-S1_1575466	7	25	S7_3522458-S7_12010754	5.1	15.5	0.1	-0.6	1.5
4	55	S4_4734626-S4_9854357	6	95	S6_21393668-S6_3159471	5.1	17.3	0.0	0.9	1.0
3	115	S3_11310314-S3_25423101	11	155	S11_12420322-S11_38211354	5.2	31.2	-1.5	-0.7	2.1
7	30	S7_18419460-S7_15005750	8	85	S8_10556549-S8_19075759	5.2	12.8	0.1	-1.1	-1.0
6	15	S6_14310524-S6_14313152	8	185	S8_14218715-S8_15026624	5.2	16.7	0.3	0.2	0.5
6	60	S6_21760159-S6_11433399	6	85	S6_22264720-S6_21393668	5.4	25.7	-1.0	-1.2	0.2
5	20	S5_4692888-S5_312017	6	35	S6_14335733-S6_14282201	5.5	19.0	0.9	0.1	1.4
8	30	S8_11972776-S8_2049156	11	195	S11_8456046-S11_39507811	5.5	24.4	-0.1	-0.7	-1.7
7	75	S7_6012566-S7_9034247	11	90	S11_23469021-S11_46615058	5.6	22.3	0.8	0.0	1.4
1	45	S1_4757043-S1_1575466	8	100	S8_13310192-S8_4675310	6.6	23.8	0.5	-1.1	0.9
Pop2 (ICP 8863 × ICP 11605)										
9	35	S9_7212583-S9_7755937	10	170	S10_19001995-S10_7783201	5.2	56.6	0.0	0.4	5.1
3	130	S3_8772530-S3_1358533	10	155	S10_19001995-S10_7783201	5.5	67.4	-1.1	0.5	-0.6
1	55	S1_5552696-S1_3905151	10	155	S10_19001995-S10_7783201	5.5	60.5	-1.1	1.1	1.9
5	30	S5_4692912-S5_4199522	6	5	S6_2496170-S6_6237148	6.0	55.0	0.0	0.4	3.6
5	25	S5_4692912-S5_4199522	10	155	S10_19001995-S10_7783201	6.1	64.8	-0.2	0.8	4.1
10	160	S10_19001995- S10_7783201	10	195	S10_7783201-S10_5097784	6.1	69.8	0.8	0.8	-0.6

† Chromosomes harbouring interacting loci; § Positions of interacting loci; LOD, logarithm of odds ratio; phenotypic variation explained; ¶ Additive-by-additive epistatic effect; ‡ Additive-by-dominance epistatic effect; ‡‡ Dominance-by-dominance epistatic effect

Table 4.11: (continued)

Chr.1	Pos.1	Flanking markers (QTL1)	Chr.2	Pos.2	Flanking markers (QTL2)	LOD	PVE (%)	[aa]	[ad]	[dd]
2	80	S2_22473129-S2_18386711	10	155	S10_19001995-S10_7783201	6.3	67.0	0.0	-0.6	2.5
11	20	S11_27612418-S11_32832892	11	215	S11_10013681-S11_32879360	7.1	59.8	0.4	-1.0	3.1
10	155	S10_19001995-S10_7783201	11	5	S11_2882386-S11_27612418	7.5	66.4	0.3	-0.7	2.7
Pop3 (HPL 24 × ICP 11605)										
3	190	S3_23950418-S3_23900756	11	30	S11_34926164- S11_14777000	5.3	21.2	-0.9	1.0	1.4
1	20	S1_887236-S1_3399209	3	235	S3_11414215- S3_19102565	5.5	19.3	0.1	-0.3	-2.3
Pop4 (ICP 5529 × ICP 11605)										
5	85	S5_2154598-S5_3299149	11	55	S11_4453854 S11_4725362	5.2	21.2	0.3	0.5	1.9
4	40	S4_3887219-S4_3355054	7	175	S7_14683829- S7_14588865	5.2	15.0	0.2	0.1	-1.7
5	105	S5_3512215-S5_3512203	7	65	S7_19133038- S7_19133012	5.8	9.5	0.2	-0.1	-1.2
4	10	S4_1867372-S4_2222488	5	90	S5_2154598- S5_3299149	6.9	15.4	-0.2	0.2	0.7
Pop5 (ICP 8863 × ICPL 87119)										
6	145	S6_4528744-S6_11344426	10	25	S10_8682310- S10_22432012	5.0	19.1	0.4	-0.7	0.7
10	130	S10_12572056-S10_11360684	11	175	S11_41966756- S11_2739522	5.2	21.4	-0.3	-0.3	1.7
11	10	S11_9655513-S11_11782549	11	25	S11_39259070- S11_5870873	5.2	28.9	-0.3	0.4	2.4
2	5	S2_31090530-S2_11172947	6	120	S6_8998640- S6_22745176	5.3	17.0	0.1	-0.4	0.7
1	0	S1_3518364-S1_8912598	10	165	S10_11797822- S10_22177616	5.3	9.8	0.5	0.1	0.5
6	120	S6_8998640-S6_22745176	8	75	S8_4817492- S8_18097152	5.5	18.6	0.3	-0.5	1.5
10	115	S10_16591318-S10_19313155	10	165	S10_11797822- S10_22177616	5.5	12.8	-0.1	0.6	0.7
6	95	S6_21912913-S6_11368997	7	40	S7_5257435-S7_19515938	6.3	30.5	0.0	0.5	1.5

Table 4. 12 Colocalised epistatic QTLs (E-QTLs) for seed protein content with that for other traits in five F_2 mapping populations of pigeonpea

Trait	Chr.1†	Pos.1§	QTL1‡	Chr.2†	Pos.2§	QTL2‡	LOD	PVE (%)	[aa]††	[ad]§§	[dd]‡‡
	Pop1 (ICP 11605 × ICP 14209)										
SPC	**1**	**30**	**S1_2693194-S1_4757043**	**2**	**75**	**S2_7491873-S2_36672875**	**5**	**18**	**-0.4**	**-0.3**	**1.5**
GH	1	35	S1_2693194-S1_4757043	2	80	S2_24059431-S2_24764841	5	27.2	-0.2	0.1	-0.3
SPC	**1**	**45**	**S1_4757043-S1_1575466**	**7**	**25**	**S7_3522458-S7_12010754**	**5.1**	**15.5**	**0.1**	**-0.6**	**1.5**
SPC	**1**	**45**	**S1_4757043-S1_1575466**	**8**	**100**	**S8_13310192-S8_4675310**	**6.6**	**23.8**	**0.5**	**-1.1**	**0.9**
SW	1	40	S1_4757043-S1_1575466	2	85	S2_36121093-S2_36167974	5.3	14.7	0.1	-0.4	1.2
SW	1	40	S1_4757043-S1_1575466	6	95	S6_21393668-S6_3159471	5.5	18.2	-0.5	-0.1	0.7
SY	1	45	S1_4757043-S1_1575466	3	20	S3_12603960-S3_23502392	6.1	22.7	0.8	-28.4	47.1
SPC	**3**	**115**	**S3_11310314-S3_25423101**	**11**	**155**	**S11_12420322-S11_38211354**	**5.2**	**31.2**	**-1.5**	**-0.7**	**2.1**
SY	3	35	S3_16632580-S3_22917964	3	115	S3_11310314-S3_25423101	5.4	23.9	-19	-0.1	20.7
SPC	**5**	**20**	**S5_4692888-S5_312017**	**6**	**35**	**S6_14335733-S6_14282201**	**5.5**	**19**	**0.9**	**0.1**	**1.4**
GH	5	20	S5_4692888-S5_312017	5	50	S5_1264439-S5_624899	7.8	37.9	0	0	-0.8
SPC	**6**	**60**	**S6_21760159-S6_11433399**	**6**	**85**	**S6_22264720-S6_21393668**	**5.4**	**25.7**	**-1**	**-1.2**	**0.2**
SY	6	60	S6_21760159-S6_11433399	8	35	S8_2049156-S8_5791461	6.4	23.9	-9.9	-6.8	36.1
SY	6	60	S6_21760159-S6_11433399	11	80	S11_9768472-S11_24859358	5.3	17.8	-8.3	-6.7	43.9
SPC	**8**	**30**	**S8_11972776-S8_2049156**	**11**	**195**	**S11_8456046-S11_39507811**	**5.5**	**24.4**	**-0.1**	**-0.7**	**-1.7**
SY	8	25	S8_11972776-S8_2049156	8	95	S8_6388803-S8_9452602	6.4	31	-16.6	-21.1	56.8
GH	8	30	S8_11972776-S8_2049156	10	20	S10_14964979-S10_7161105	7.8	29.3	0.2	0.4	-0.3

†Chromosomes harbouring interacting loci; § Positions of interacting loci, ‡ Interacting QTLs; †† Additive-by-additive epistatic effect; §§ Additive-by-dominance epistatic effect; ‡‡ Dominance-by-dominance epistatic effect. LOD, logarithm of odds ratio; PVE, phenotypic variation explained. E-QTLs in bold font are for SPC which co-localize with E-QTLs for the other traits.

Table 4.12: (continued)

Trait	Chr.1	Pos.1	QTL1	Chr.2	Pos.2	QTL2	LOD	PVE (%)	[aa]	[ad]	[dd]
Pop2 (ICP 8863 × ICP 14209)											
SY	1	10	S1_4759267-S1_15329865	10	200	S10_7783201-S10_5097784	6.2	35.6	14.3	15.8	-13.1
SPC	**2**	**80**	**S2_22473129-S2_18386711**	**10**	**155**	**S10_19001995-S10_7783201**	**6.3**	**67**	**0**	**-0.6**	**2.5**
SW	2	80	S2_22473129-S2_18386711	3	0	S3_22234078-S3_19578263	14.6	31.9	-0.2	2.7	-4.5
SW	2	80	S2_22473129-S2_18386711	4	25	S4_1710877-S4_839628	15.3	33.7	2.7	-2.1	2.1
SW	2	80	S2_22473129-S2_18386711	2	115	S2_21890021-S2_7683449	15.5	34.1	-2.5	0.2	2.2
SPC	**3**	**130**	**S3_8772530-S3_1358533**	**10**	**155**	**S10_19001995-S10_7783201**	**5.5**	**67.4**	**-1.1**	**0.5**	**-0.6**
DTF	3	135	S3_8772530-S3_1358533	11	150	S11_22893601-S11_10881649	5.8	22.9	-7.8	-3.8	1.7
GH	5	10	S5_4692912-S5_4199522	11	90	S11_39309852-S11_28894118	8.8	15.1	-0.2	0	0.2
SW	5	15	S5_4692912-S5_4199522	8	40	S8_11776408-S8_11838449	15.6	37.5	-2.6	0.4	3.5
SW	5	20	S5_4692912-S5_4199522	10	115	S10_13626722-S10_18083991	15	32.5	-2.5	-0.2	3.2
SW	5	20	S5_4692912-S5_4199522	6	0	S6_2496170-S6_6237148	12.8	32.8	-2.5	0.1	3.2
GH	5	20	S5_4692912-S5_4199522	5	35	S5_4692912-S5_4199522	8.8	27.6	0.1	0.1	-0.7
SPC	**5**	**25**	**S5_4692912-S5_4199522**	**10**	**155**	**S10_19001995-S10_7783201**	**6.1**	**64.8**	**-0.2**	**0.8**	**4.1**
SW	5	25	S5_4692912-S5_4199522	7	5	S7_18419460-S7_6037045	15.3	36.9	-2.7	-0.2	3.5
SW	5	25	S5_4692912-S5_4199522	11	50	S11_3238944-S11_3724414	16.4	38.7	-2.4	-0.2	1.6
SPC	**5**	**30**	**S5_4692912-S5_4199522**	**6**	**5**	**S6_2496170-S6_6237148**	**6**	**55**	**0**	**0.4**	**3.6**
GH	5	30	S5_4692912-S5_4199522	10	150	S10_19001995-S10_7783201	12	40.7	-0.2	0	0.1
SW	6	60	S6_15449552-S6_18295388	10	205	S10_7783201-S10_5097784	13.1	33	-2.5	-0.2	3.2

Table 4.12: (continued)

Trait	Chr.1	Pos.1	QTL1	Chr.2	Pos.2	QTL2	LOD	PVE (%)	[aa]	[ad]	[dd]
SY	7	5	S7_18419460-S7_6037045	10	205	S10_7783201-S10_5097784	6.5	31.5	-17.3	0.2	-17.4
SY	9	15	S9_1280000-S9_7212583	10	150	S10_19001995-S10_7783201	5.3	30.1	9.6	21.3	3.1
SW	9	20	S9_1280000-S9_7212583	10	165	S10_19001995-S10_7783201	13.7	41.4	-2.4	0.6	3.9
SPC	**9**	**35**	**S9_7212583-S9_7755937**	**10**	**170**	**S10_19001995-S10_7783201**	**5.2**	**56.6**	**0**	**0.4**	**5.1**
GH	9	35	S9_7212583-S9_7755937	10	135	S10_17099642-S10_6765628	8.3	21.2	-0.2	0	0.1
DTF	9	45	S9_7212583-S9_7755937	10	100	S10_21365631-S10_16278049	5.8	18.8	-3.4	-1.7	18.0
SPC	**10**	**155**	**S10_19001995-S10_7783201**	**11**	**5**	**S11_2882386-S11_27612418**	**7.5**	**66.4**	**0.3**	**-0.7**	**2.7**
SPC	**10**	**160**	**S10_19001995-S10_7783201**	**10**	**195**	**S10_7783201-S10_5097784**	**6.1**	**69.8**	**0.8**	**0.8**	**-0.6**
GH	10	140	S10_19001995-S10_7783201	10	150	S10_19001995-S10_7783201	12.7	36.6	-0.1	-0.5	-0.6
SPC	**11**	**20**	**S11_27612418-S11_32832892**	**11**	**215**	**S11_10013681-S11_32879360**	**7.1**	**59.8**	**0.4**	**-1**	**3.1**
DTF	11	55	S11_33954110-S11_42065843	11	215	S11_10013681-S11_32879360	5.7	29.1	11.8	-4.4	12.2
SY	11	165	S11_41096347-S11_44938548	11	215	S11_10013681-S11_32879360	6.2	34	-14.9	21.9	-13.6
Pop3 (HPL 24 × ICP 11605)											
GH	1	20	S1_887236-S1_3399209	2	80	S2_28723848-S2_6405369	6.5	15.7	0.1	-0.3	-0.4
SW	1	20	S1_887236-S1_3399209	2	75	S2_23068242-S2_36264850	9	24.6	-0.9	1.1	2.7
SPC	**1**	**20**	**S1_887236-S1_3399209**	**3**	**235**	**S3_11414215-S3_19102565**	**5.5**	**19.3**	**0.1**	**-0.3**	**-2.3**
Pop4 (ICP 5529 × ICP 11605)											
SPC	**4**	**10**	**S4_1867372-S4_2222488**	**5**	**90**	**S5_2154598-S5_3299149**	**6.9**	**15.4**	**-0.2**	**0.2**	**0.7**
SY	4	5	S4_2313990-S4_1867372	11	165	S11_35533615-S11_24152330	5.3	20.8	-22.7	7	39.9
SPC	**4**	**40**	**S4_3887219-S4_3355054**	**7**	**175**	**S7_14683829-S7_14588865**	**5.2**	**15**	**0.2**	**0.1**	**-1.7**

Table 4.12: (continued)

Trait	Chr.1	Pos.1	QTL1	Chr.2	Pos.2	QTL2	LOD	PVE (%)	[aa]	[ad]	[dd]
DTF	4	40	S4_3887219-S4_3355054	11	80	S11_18575052-S11_38654219	7	22.2	4.8	0.8	10.8
SPC	**5**	**105**	**S5_3512215-S5_3512203**	**7**	**65**	**S7_19133038-S7_19133012**	**5.8**	**9.5**	**0.2**	**-0.1**	**-1.2**
SY	5	105	S5_3512215-S5_3512203	11	160	S11_38144723-S11_27887800	5.5	16.9	-6	23.7	-31.5
Pop5 (ICP 8863 × ICPL 87119)											
SPC	**6**	**95**	**S6_21912913-S6_11368997**	**7**	**40**	**S7_5257435-S7_19515938**	**6.3**	**30.5**	**0**	**0.5**	**1.5**
SW	6	95	S6_21912913-S6_11368997	7	40	S7_5257435-S7_19515938	6.1	18.6	0.4	-0.1	-1.1
SPC	**1**	**0**	**S1_3518364-S1_8912598**	**10**	**165**	**S10_11797822-S10_22177616**	**5.3**	**9.8**	**0.5**	**0.1**	**0.5**
SY	1	5	S1_3518364-S1_8912598	11	200	S11_10379800-S11_39387203	6.3	23.8	1.8	-1.8	-63.8
SW	2	40	S2_10335056-S2_2989918	11	175	S11_41966756-S11_2739522	6.2	24.4	0.2	-0.7	-0.7
SPC	**10**	**130**	**S10_12572056-S10_11360684**	**11**	**175**	**S11_41966756-S11_2739522**	**5.2**	**21.4**	**-0.3**	**-0.3**	**1.7**

4.3.10 Epistatic QTLs shared between seed protein content and the other traits

The number of E-QTL pairs shared between SPC and the other traits were variable depending on the population (Table 4.12). In Pop1, SPC shared E-QTLs with SW, SY and GH (Table 4.12). In Pop2, SPC shared E-QTLs with SW, SY, DTF and GH, while in Pop3, SPC shared E-QTL markers with SW and GH (Table 4.12). In Pop4, SPC shared two E-QTLs with SY, and one E-QTL with DTF. In Pop5, two E-QTLs for SPC were shared with SW, and one with SY (Table 4.12).

4.4 Discussion

4.4.1 Variation in seed protein content and four agronomic traits in five F_2 mapping populations

The existence of valuable phenotypic variation for SPC between the parental lines and within the corresponding genetic populations allowed the effective dissection of their genetic basis and identification of genomic regions for genetic improvements. Despite the moderate differences in SPC (0.8 to 3.5%) between parents of a cross, wide segregation was observed in the F_2 populations with differences between min and max SPC of F_2 plants ranging from 5.8% (Pop5) to 10.3% (Pop3). The range of segregation in the F_2 populations signifies transgressive segregation, a phenomenon that has commonly been observed for SPC in other legume crops such as soybean (Lee et al., 2010; Zhang et al., 2015; Soybase, 2016 and references therein) and pea (Irzykowska and Wolko, 2004; Taran et al., 2004; Burstin et al., 2007; Krajewski et al., 2012). The notable phenotypic variation in SPC and the other traits facilitated the detection of the underlying genomic segments.

4.4.2 Genetic maps

Molecular markers and genetic maps are fundamental in analyzing genetic architecture of a trait and for molecular breeding in any crop or animal species. Significant progress has been made in developing genetic maps in pigeonpea, and both interspecific and intraspecific maps have been constructed using different types of markers in pigeonpea. For breeding applications, intraspecific genetic maps are more useful (Borha et al., 2012), and so far seven have been reported in pigeonpea (Gnanesh et al., 2011; Borha et al., 2012; Kumawat et al., 2012). Nearly

all pigeonpea intraspecific genetic maps reported have used SSR markers, except one by Kumawat el al. (2012) which included both SSR and SNP markers. Single nucleotide polymorphisms have potential for high genotyping efficiency, automation, data quality, genome-wide coverage and analytical simplicity (Subramanian et al., 2003). Because of such attributes, SNPs have rapidly become the marker of choice for many applications in genetics and genomics (Varshney et al., 2010). The availability of a pigeonpea reference genome sequence (Varshney et al., 2012) has facilitated the application of new genotyping methodologies such as GBS in the crop. Genotyping-by-sequencing approach (Elshire et al., 2011) provides the possibility to generate high-density SNP genotyping data.

The present study focuses on construction of genetic maps based on four intraspecific mapping populations segregating for SPC using GBS-derived SNP markers. The four maps contain 363 (Pop2) to 787 (Pop4) SNP loci with an average of 1.8 (Pop1) to 3.5 cM (Pop2) marker spacing, respectively. In addition, a fifth genetic map constructed from F_2 mapping of the cross ICP 8863 × ICPL 87119 (Pop5) under another project was included in the present study. The map contain 996 GBS-derived SNPs with a total length of 1599.8 cM and average marker spacing of 1.6 cM. The SNP marker spacing in each of the five populations in the present study provides adequate power to detect a QTL, given that a 5-cM SNP spacing is considered sufficiently dense for optimized QTL detection power in a population of size of 200 individuals (Stange et al., 2013).

Marker segregation distortion was observed in all the five crosses with similar proportion of markers showing deviation from expectation. Segregation distortion may result from various factors such as residual heterozygosity, gametic or zygotic selections and genotyping errors (Liang et al., 2006). It is a common phenomenon observed in both intraspecific and interspecific crosses and has been reported in several crops including pigeonpea (Bohra et al., 2011; 2012) and chickpea (Gaur et al., 2011). Although distorted markers have generally been discarded in earlier studies, recent evidence indicate that distorted markers can be potentially helpful in the detection of QTLs (Xu, 2008). It has also been noted that discarding distorted markers could possibly remove substantial amounts of information and reduce genome coverage (Luo et al., 2005). Thus, in the present study distorted markers segregating in 1:2:1 Mendelian ratio with χ^2 cutoff $P \geq 10^{-9}$ were retained for genetic map construction. Similar χ^2 cutoff P-values have been used in other crop species, for example, Chen et al. (2014) used a χ^2

cutoff $P<10^{-10}$ in maize, while Eduardo et al. (2013) used χ^2 cutoff $P<10^{-6}$ to declare distorted markers in peach.

By integrating five component maps into a consensus map, conserved marker orders were observed among the five maps that could be attributed to use of relatively similar population size (137 to 179), same type of mapping populations (all F_2s) and same type of marker system (GBS-derived SNPs). The constructed genetic maps were then used for QTL analysis for SPC and how it relates to four agronomic traits.

4.4.3 Main effect QTLs for seed protein content

QTL mapping in experimental populations has become a common method by which many parameters of genetic architecture of a trait are characterized (Abiola et al., 2003). In the case of pigeonpea, no reports on the investigation of the genetic architecture of SPC yet exists, and therefore nothing is known about how the specific parameters of the genetic architecture vary within and across populations. The present study examined the genetic basis of the variation in SPC in pigeonpea using five partially connected F_2 mapping populations.

Estimating the consequences of both significant main and interaction effects makes it possible to better explain the total phenotypic variation in terms of main effect loci and combinations of loci (Carlborg et al., 2004). The QTL mapping results in the present study revealed a minimum of two and a maximum of three major effect M-QTLs and several modifier/minor effect M-QTLs/genes to be involved in SPC variation in the populations studied. These results are similar to that in soybean (Zhang et al., 2015; SoyBase, 2016), where only a few major M-QTLs were identified in the presence of several minor effect QTLs for SPC. However, it contrasts with results of QTL mapping for SPC in pea (Tar'an et al., 2004; Irzykowska and Wolko, 2004; Burstin et al., 2007) where the lowest PVE by an M-QTL has been reported as 9.0%.

All SPC major effect M-QTLs show population specificity with the exception of one major M-QTL (*qPROT-cim-3.1*) in Pop2 which shared one of the flanking markers with a minor M-QTL (*qPROT-cim-3.2*) in Pop3. A possible explanation of population specificity of the M-QTLs for SPC could be that population size or marker coverage contributed to the mapping of different

QTLs in the different populations (Symonds et al., 2005). However, this is unlikely because population-specific M-QTLs of relatively minor effects ranging from 0.7 (Pop2) to 7.8% (Pop1) were mapped in all five populations. Rather, it is possible that a QTL detected in a certain cross may not be detected in another cross because the parents of the second cross carry identical alleles at the same locus (Elberse et al., 2004; Symonds et al., 2005; Simon et al., 2008).

4.4.4 The role of epistasis

Besides the M-QTLs, E-QTLs were detected that explained additional phenotypic variation for SPC and the other traits. In general, the PVE of the individual E-QTLs were relatively larger than those of individual M-QTLs on average. Furthermore, the large number of E-QTLs for SPC and for the other traits identified in present study indicates that QTLs with minor effects or no effect interact with each other to influence expression of the traits. For instance, the PVE by E-QTLs for SPC were in general higher than the PVE by M-QTLs in all except one population (Pop2). The relatively low marker density in Pop2 likely contributed to the high PVE of the E-QTLs. This pattern of contributions of M-QTLs vs E-QTLs to phenotypic variation for SPC is similar to that made to the phenotypic variance of the other traits (DTF, GH, SW and SY).

4.4.5 Phenotypic correlation and QTL colocalization between seed protein content and agronomic traits

The existence of a genetic relationship between SPC and other plant traits could make it essential for breeders to consider the selection criterion to improve SPC while maintaining other desirable plant traits. In this study, two lines of evidence revealed the associations between SPC and the other plant traits, and that the nature of the associations is genetic background-dependent.

First, the phenotypic correlation analysis showed that SPC associates positively with GH and SW and negatively with DTF and SY. The association of SPC with DTF was significant in two of the five populations while that of SPC and SW was significant in three populations. In the case of SPC with SY and GH, the association was significant in four populations while no

significant correlations were found between SPC with any of the agronomic traits in Pop5. The pattern of correlation of SPC with SW is consistent with results of earlier studies which showed that the two traits associate either positively or negatively and sometimes non-significantly depending on genetic material used (Saxena et al., 1987). In the case of SPC with DTF and SY, negative though small and none significant relationship was reported in pigeonpea (Rekha et al., 2013). No relationship between SPC and GH has been reported in pigeonpea before. However, significant correlation of SPC with morphological and growth-related traits have been reported in pea (Burstin et al., 2007).

Second, co-localization of M-QTLs and shared E-QTLs for SPC with that of the other traits were found that possibly explains trait correlations. For instance, the co-localisation of M-QTLs for SPC with M-QTLs for DTF with opposite allelic effects could explain the negative correlations between SPC and DTF in Pop1, Pop2, Pop3 and Pop5 though the correlations were non-significant in Pop2 and Pop5. Similarly, the co-localisation of M-QTLs for SPC and M-QTLs for GH with allelic effects in the same direction in Pop1, Pop2 and Pop3 explains positive correlation between the two traits. Likewise, correlation of SPC with SW in Pop2 could be explained by the overlapping M-QTLs on CcLG02 with allelic effects in the same direction. While the negative correlation of SPC with SY could be attributed to opposing effect of co-localised M-QTLs for the two traits such as in Pop2.

However, not all correlations of SPC with agronomic traits could be explained by co-localisation of M-QTLs, for instance, GH and SY showed relatively strong correlation with SPC in Pop4 but no M-QTL overlaps were present. Therefore, presence of E-QTLs shared between SPC and the agronomic traits were searched that could explain correlations that are not explained by the M-QTLs. The phenomenon where one E-QTL affects expression of more than one trait have been termed ‘epistatic pleiotropy’ (Wolf et al., 2005). In this regard, the majority of epistatic pleiotropy involving SPC and other traits in the present study are the type in which the effects of a given pleiotropic locus are dependent upon the alleles present at the other loci (Cheverud, 2004). For example, in Pop1 a QTL on CcLG01 flanked by markers *S1_4757043* and *S1_1575466*, affected (i) SPC when it interacted with other QTLs on CcLG07 and CcLG08, (ii) SW when it interacted with QTLs on CcLG02 and CcLG06, and (iii) SY when it interacted with a QTL on CcLG03.

Similarly, a single EP-QTL on CcLG01 (*S1_887236* and *S1_3399209*) in Pop3 influenced the expression of SPC, SW and GH when it interacted with other QTLs on CcLG02 and CcLG03 and possibly contributed to the significant covariance between SPC and SW, and SPC and GH. Such epistatically pleiotropic QTLs (EP-QTLs) involving SPC were widespread among populations, and in some cases provided the only explanation to phenotypic correlation between SPC and other traits. For instance, the significant correlation between SPC and SY in Pop4 in the absence of overlaps in their M-QTLs could be explained by EP-QTL on CcLG07 flanked by markers *S7_14683829* and *S7_14588865*. The same EP-QTL also influenced expression of SW and DTF although the two traits show weak and non-significant correlation with SPC. In Pop5, three EP-QTLs were detected, two of which influenced SPC and SY, and one influenced SPC and SW even though no significant relationships of SPC with SW and SY were found.

4.5 Conclusions

The present study reports the first attempt to dissect the genetic architecture of SPC in pigeonpea in a manner that incorporates an investigation of the genetic basis of its correlations with important seed, phenological and morphological traits. High-density intraspecific genetic linkage maps of pigeonpea were assembled with map lengths ranging from 1327.6 cM to 1599.8 cM and an average marker to marker distance of 1.6 cM to 3.5 cM, respectively. Similarly, a high-density consensus genetic map was assembled from five component maps. The consensus map contains 984 markers, with an average marker to marker distance of 1.6 cM. Two to three major effect QTLs in the presence of several modifier/minor effect QTLs and epistatic QTLs control the expression of SPC in the study populations. Overlaps of main effect and epistatically pleiotropic QTLs explain the correlations between SPC and other traits. Projection of M-QTLs for SPC and agronomic traits onto the consensus map revealed common genomic regions governing SPC and its relationship with agronomic traits across different genetic backgrounds. Such integrated QTL information provide a valuable resource that can potentially contribute to genomic and genetic studies in the crop. To harness both the main effect and epistatic QTLs, genomic selection that targets the genome wide variations for crop improvement will be the best alternative in achieving larger genetic gains in shorter periods (Varshney et al., 2012).

Reference

Abiola, O., J.M. Angel, P. Avner, A.A. Bachmanov, J.K. Belknap, B. Bennett, E.P. Blankenhorn, D.A. Blizard, V. Bolivar, G.A. Brockmann, K.J. Buck, J.-F. Bureau, W.L. Casley, E.J. Chesler, J.M. Cheverud, G.A. Churchill, M. Cook, J.C. Crabbe, W. Crusio, A. Darvasi, G. de Haan, P. Demant, R.W. Doerge, R.W. Elliott, C.R. Farber, L. Flaherty, J. Flint et al. 2003. The identification of quantitative trait loci: a community's view. Nature Review Genetics 4:911-916.

Akibode, C. S., and M. Maredia 2011. Global and regional trends in production, trade and consumption of food legume crops. Report submitted to the Standing Panel on Impact Assessment (SPIA) of the CGIAR Science Council, FAO, Rome.

Anderson, M.J., and C.J.F. Ter-Braak. 2003. Permutation tests for multi-factorial analysis of variance. Journal of Statistical Computation and Simulation 73:85-113.

Baskaran, K., and A.R. Muthiah. 2007. Studies on combining ability in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Legumes Research 30:67-69.

Bohra, A., A. Dubey, R.K. Saxena, R.V. Penmetsa, K.N. Poornima, N. Kumar, A.D. Farmer, G. Srivani, H.D. Upadhyaya, R. Gothalwal, S. Ramesh, D. Singh, K.B. Saxena, P.B. KaviKishor, N.K. Singh, C.D. Town, G.D. May, D.R. Cook, and R.K. Varshney. 2011. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea, (*Cajanus* spp.). BMC Plant Biology 11:56.

Bohra, A., R.K. Saxena, B.N. Gnanesh, K.B. Saxena, M. Byregowda, A. Rathore, P.B. KaviKishor, D. Cook, and R.K. Varshney. 2012. An intraspecific consensus genetic map of pigeonpea [*Cajanus cajan* (L) Millspaugh] derived from six mapping populations. Theoretical and Applied Genetics 125:1325-1338.

Burstin, J., K. Gallardo, R.R. Mir, R.K. Varshney, G. Duc, J. Burstin, K. Gallardo, R.R. Mir, and R.K. Varshney. 2011. Chapter 20: Improving protein content and nutrition quality. In: Pratap, A., and J. Kumar, editors, Biology and breeding of food legumes. Wallingford: CABI, p. 314-328.

Burstin, J., Marget, P., M. Huart, A. Moessner, B. Mangin, C. Duchene, B. Desprez, N. Munier-Jolain, and G. Duc. 2007. Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiology 144:768-781.

Carlborg, O., and C.S. Haley. 2004. Epistasis: Too often neglected in complex trait studies? Nature Reviews Genetics 5:618-625.

Chen, Z., B. Wang, X. Dong, H. Liu, L. Ren, J. Chen, A. Hauck, W. Song, and J. Lai. 2014. An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F_2 maize population. BMC Genomics 15:433.

Cheverud, J.M. 2004. Modular pleiotropic effects of quantitative trait loci on morphological traits. In: Schlosser, G., and G.P.Wagner, editors, Modularity in Development and Evolution, University of Chicago Press, Chicago. p. 132-153.

Cui, Y., F. Zhang, J. Xu, Z. Li, and S. Xu. 2014. Mapping quantitative trait loci in selected breeding populations: A segregation distortion approach. Heredity (Edinb) 115:538-546.

Dahiya, B.S., J.S. Brar, and B.S. Bhullar. 1977. Inheritance of protein content and its correlation with grain yield in pigeonpea (*Cajanus cajan* (L.) Millsp.). Qualitus Plantarum 27:327-334.

Eduardo, I., G. Chietera, R. Pirona, I. Pacheco, M. Troggio, E. Banchi, D. Bassi, L. Rossini, A. Vecchietti, and C. Pozzi. 2013. Genetic dissection of aroma volatile compounds from the essential oil of peach fruit: QTL analysis and identification of candidate genes using dense SNP maps. Tree Genetics and Genomes 9:189-204.

Elberse, I.A.M., T.K. Vanhala, J.H.B. Turin, P. Stam, J.M.M. van Damme, and P.H. van Tienderen. 2004. Quantitative trait loci affecting growth-related traits in wild barley (*Hordeum spontaneum*) grown under different levels of nutrient supply. Heredity 93:22-33.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379.

Gaur, R., N.K. Sethy, S. Choudhary, B. Shokeen, V. Gupta, and S. Bhatia. 2011. Advancing the STMS genomic resources for defining new locations on the intraspecific genetic linkage map of chickpea (*Cicer arietinum* L.). BMC Genomics 12:117.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: a high capacity genotyping-by-sequencing analysis pipeline. PLoS One 9:e90346.

Gnanesh, B.N., A. Bohra, M. Sharma, M. Byregowda, S. Pande, V. Wesley, R.K. Saxena, K.B. Saxena, P.B. KaviKishor, and R.K. Varshney. 2011. Genetic mapping and quantitative

trait locus analysis of resistance to sterility mosaic disease in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Field Crops Research 123:56-61.

International Cassava Genetic Map Consortium (ICGMC). 2015. High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. Genes/Genomes/Genetics 5:133-144.

ICRISAT. 1993a. Pigeonpea variety ICP 8863. Plant material description no. 44. International Crops Research Institute for the Semi-Arid Tropics, Andhra Pradesh, India.

ICRISAT. 1993b. Pigeonpea variety ICPL 151. Plant material description no. 41. International Crops Research Institute for the Semi-Arid Tropics, Andhra Pradesh, India.

ICRISAT. 1993c. Pigeonpea variety ICPL 87119. Plant material description no. 43. International Crops Research Institute for the Semi-Arid Tropics, Andhra Pradesh, India.

Irzykowska, L., and B. Wolko. 2004. Interval mapping of QTLs controlling yield-related traits and seed protein content in *Pisum sativum*. Journal of Applied Genetics 45:297-306.

Jaganathan, D., M. Thudi, S. Kale, S. Azam, M. Roorkiwal, P. Gaur, P. KaviKishor, H. Nguyen, T. Sutton, and R.K. Varshney. 2015. Genotyping-by-sequencing based intra-specific genetic map refines a ‘‘QTL hotspot” region for drought tolerance in chickpea. Molecular Genetics and Genomics 290:559-571.

Kale, S.M., D. Jaganathan, P. Ruperao, C. Chen, R. Punna, H. Kudapa, M. Thudi, M. Roorkiwal, M.A.V.S.K. Katta, D. Doddamani, V. Garg, P.B. KaviKishor, P.M. Gaur, H.T. Nguyen, J. Batley, D. Edwards, T. Sutton, and R.K. Varshney. 2015. Prioritization of candidate genes in “QTL-hotspot” region for drought tolerance in chickpea (*Cicer arietinum* L.). Scientific Reports 5:15296.

Khoury, C.K., N.P. Castañeda-Alvarez, H.A. Achicanoy, C.C. Sosa, V. Bernau, M.T. Kassa, S.L. Norton, L.J.G. van der Maesen, H.D. Upadhyaya, J. Ramírez-Villegas, A. Jarvis, and P.C. Struik. 2015. Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) millsp.]: distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. Biological Conservation 184:259-270.

Kosambi, D.D. 1944. The estimation of map distances from recombination values. Annals of Eugenics 12:172-175.

Krajewski, P., J. Bocianowski, M. Gawłowska, Z. Kaczmarek, T. Pniewski, W. Święcicki, and B. Wolko. 2012. QTL for yield components and protein content: a multi-environment study of two pea (*Pisum sativum* L.) populations. Euphytica 183:323-336.

Kumar, V., A.W. Khan, R.K. Saxena, V. Garg, and R.K. Varshney. 2016. First-generation HapMap in *Cajanus* spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnology Journal 14:1673-81.

Kumawat, G., R.S. Raje, S. Bhutani, J.K. Pal, A.S. Mithra, K. Gaikwad, T.R. Sharma, and N.K. Singh. 2012. Molecular mapping of QTLs for plant type and earliness traits in pigeonpea (*Cajanus cajan* L. Millsp.). BMC Genetics 13:84.

Lee, J., J.G. Shannon, and M. Choung. 2010. Selection for protein content in soybean from single F_2 seed by near infrared reflectance spectroscopy. Euphytica 172:117-123.

Li, H., Bradbury, P., E. Ersoz, E.S. Buckler, and J. Wang. 2011. Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. PLoS One 6:e17573.

Li, H., J.-M. Ribaut, Z. Li, and J. Wang. 2008. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theoretical and Applied Genetics 116:243-60.

Li, H., P. Vikram, R.P. Singh, A. Kilian, J. Carling, J. Song, J.A. Burgueno-Ferreira, S. Bhavani, J. Huerta-Espino, T. Payne, D. Sehgal, P. Wenzl, and S. Singh. 2015. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. BMC Genomics 16:216.

Li, H., G. Ye, and J. Wang. 2007. A modified algorithm for the improvement of composite interval mapping. Genetics 175:361-74.

Li, L., W. Zheng, Y. Zhu, H. Ye, B. Tang, Z.W. Arendsee, D. Jones, R. Li, D. Ortiz, X. Zhao, C. Du, D. Nettleton, M.P. Scott, M.G. Salas-Fernandez, Y. Yin, and E.S. Wurtele. 2015. QQS orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. Proceedings of the National Academy Sciences, USA 112:14734-14739.

Liang, S.X., S.X. Zhen, and Z.T. Zhen. 2006. Segregation distortion and its effect on genetic mapping in plants. Chinese Journal of Agricultural Biotechnology 3:163-169.

Lou, P., J. Zhao, J.S. Kim, S. Shen, D.P.D. Carpio, X. Song, M. Jin, D. Vreugdenhil, X. Wang, M. Koormneef, and G. Bonnema. 2007. Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. Journal of Experimental Botany 58:4005-4016.

Luo, L., Y.-M. Zhang, and S. Xu. 2005. A quantitative genetics model for viability selection. Heredity 94:347-355.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7: e32253.

Rao, P., P.S. Birthal, S. Bhagavatula, and M.C.S. Bantilan. 2010. Chickpea and pigeonpea economies in Asia: facts, trends and outlook. International Crops Research Institute for the Semi-Arid Tropics. Patancheru 502 324, Andhra Pradesh, India.

Rekha, R., L. Prasanthi, M.R. Sekhar, and M.S. Priya. 2013. Studies on selection indices in pigeonpea [*Cajanus cajan* (L.) Millsp]. International Journal of Applied Biology and Pharmaceutical Technology 4:291-294.

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, R.J. Elshire, C.B. Acharya, S.E. Mitchell, S.A. Flint-Garcia, M.D. McMullen, J.B. Holland, E.S. Buckler, and C.A. Gardner. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biology 14:R55.

Sahu, A.R., R.R. Mishra, R.S. Chandra, and J. Panigrahi. 2015. Construction of interspecific genetic linkage map of pigeonpea using SCoT, RAPD, ISSR markers and simple inherited trait loci. Nucleus 58:23-31.

Saxena, K.B, D.G. Faris, U. Singh, and R.V. Kumar. 1987. Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition 36:335-340.

Saxena, K.B., R.V. Kumar, and P.V. Rao. 2002. Pigeonpea nutrition and its improvement. In: Basra, A. S., and I. S. Randhawa, editors. Quality improvement in field crops. Food Products Press. p. 227-260.

Saxena, R.K., R.V. Penmetsa, H.D. Upadhyaya, A. Kumar, N. Carrasquilla-Garcia, J.A. Schlueter, A. Farmer, A.M. Whaley, B.K. Sarma, G.D. May, D.R. Cook, and R.K. Varshney. 2012. Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeon pea and comparative mapping in legumes. DNA Research 19:449-461.

Schweizer, P., and N. Stein. 2011. Large-scale data integration reveals colocalization of gene functional groups with meta-QTL for multiple disease resistance in barley. Molecular Plant-Microbe Interactions 24:1492-1501.

Simon, M., O. Loudet, S. Durand, A. Beŕard, D. Brunel, F.-X. Sennesal, M. Durand-Tardif, G. Pelletier, and C. Camilleri. 2008. Quantitative trait loci mapping in five new large

recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. Genetics 178:2253-2264.

Singh, V.K., A.W. Khan, R.K. Saxena, V. Kumar, S.M. Kale, P. Sinha, A. Chitikineni, L.T. Pazhamala, V. Garg, M. Sharma, C.V.S. Kumar, S. Parupalli, S. Vechalapu, S. Patil, S. Muniswamy, A. Ghanta, K.N. Yamini, P.S. Dharmaraj, and R.K. Varshney. 2015. Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). Plant Biotechnology Journal 14:1183-94.

Soybase. 2016. Soybean breeder's toolbox genetic map information. [Online]. Available at: http://www.soybase.org (Accessed on October 20, 2016).

Stange, M., H.F. Utz, T.A. Schrag, A.E. Melchinger, and T. Würschum. 2013. High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. Theoretical and Applied Genetics 126:2563-2574.

Symonds, V.V., A.V. Godoy, T. Alconada, J.F. Botto, T.E. Juenger, J.J. Casal, and A.M. Lloyd. 2005. Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. Genetics 168:1649-1658.

Tar'an, B., T. Warkentin, D.J. Somers, D. Miranda, A. Vandenberg, S. Blade, and D. Bing. 2004. Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.) Euphytica 136:297-306.

Truco, M.J., R. Antonise, D. Lavelle, O. Ochoa, A. Kozik, H. Witsenboer, S.B. Fort, M.J. Jeuken, R.V. Kesseli, P. Lindhout, R.W. Michelmore, and J. Peleman. 2007. A high-density integrated genetic linkage map of lettuce (*Lactuca* spp.). Theoretical and Applied Genetics 115:735-746.

Vaghela, K.O., R.T. Desai, J.R. Nizama, J.D. Patel, and V. Sharma. 2009. Combining ability analysis in pigeonpea [C*ajanus cajan* (L.) Millsp.] Legume Research 32:274-277.

Vales, M.I., R.K. Srivastava, R. Sultana, S. Singh, I. Singh, G. Singh, S.B. Patil, and K.B. Saxena. 2012. Breeding for earliness in pigeonpea: development of new determinate and non determinate lines. Crop Science 52:2507-2516.

van Ooijen, J.W. 2006. JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.

Voorrips, R.E. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93:77-78.

Wang, J., H. Li, L, Zhang, and L. Meng. 2015. Users' manual of QTL IciMapping. The Quantitative Genetics Group, Institute of Crop Science, Chinese Academy of

Agricultural Sciences (CAAS), Beijing 100081, China, and Genetic Resources Program, International Maize and Wheat Improvement Center (CIMMYT), Apdo, Mexico.

Wang, S, C.J. Basten, and Z.-B. Zeng. 2012. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, North Carolina.

Wenzl, P., H. Li, J. Carling, M. Zhou, H. Raman, E. Paul, P. Hearnden, C. Maier, L. Xia, V. Caig, J. Ovesná, M. Cakir, D. Poulsen, J. Wang, R. Raman, K.P. Smith, G.J. Muehlbauer, K.J. Chalmers, A. Kleinhofs, E. Huttner, and A. Kilian. 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206.

Wolf, J.B., L.J. Leamy, E.J. Routman, and J.M. Cheverud. 2005. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. Genetics 171:683-694.

Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457-1468

Xu, S. 2008. Quantitative trait locus mapping can benefit from segregation distortion. Genetics 180:2201-2208.

Youens-Clark, K., B. Faga, I.V. Yap, L. Stein, and D. Ware. 2009. CMap 1.01: a comparative mapping application for the internet. Bioinformatics Applications Note 25:3040-3042.

Zhang, Y.H., M.F. Liu, J.B. He, Y.F. Wang, G.N. Xing, Y. Li, S.P. Yang, T.J. Zhao, and J.Y. Gai. 2015. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.]. Theoretical and Applied Genetics 128:1061-1072.

Chapter Five
Identification of candidate genes conditioning seed protein content in pigeonpea

Abstract

Recent developments in whole genome sequencing technologies provide rapid and cost effective methods to identify candidate genes and variants underlying qualitative and quantitative traits. The objective of the present study was to exploit whole genome resequencing (WGRS) data of four pigeonpea genotypes (~ 12 × coverage) to identify variants and candidate genes for seed protein content (SPC), an important nutritional attribute of pigeonpea seed. By combining a common variant (CV) filtering strategy with knowledge of gene functions in relation to SPC, 108 candidate sequence variants whose presence lead to protein change were selected. The variants were found in 57 genes spread over all chromosomes except CcLG05. All 57 genes have proteins related to SPC. Identified genes were assigned to 19 categories based on gene ontology molecular function with 56% of the identified genes belonging to only two functional categories. Sanger sequencing confirmed presence of 52 (75.4%) candidate sequence variants in 37 genes between low and high SPC genotypes. Fifty-nine variants were converted into CAPS/dCAPS markers and assayed for polymorphism. Highest number of polymorphic markers was in low by high SPC parental pairs, while the lowest was in high by high parental pairs. Assay of 16 polymorphic CAPS/dCAPS markers on an F_2 segregating population of the cross ICP 5529 × ICP 11605 (high × low), resulted in 11 of the markers incorporated into a genotyping-by-sequencing (GBS)-derived SNPs genetic map. Three of the CAPS/dCAPS markers were positioned at <10.0 cM distance away from main effect QTLs all on CcLG02. Single marker analysis (SMA) indicated four of the 16 CAPS/dCAPS markers to significantly correlate with SPC in the same population. The three markers found in close proximity to SPC QTL positions, and those showing significant correlation to SPC through SMA were derived from mutations in same three genes including NADH-GOGAT, copper transporter and BLISTER all on CcLG02. Results from this study provide a foundation for future basic research and marker-assisted breeding of pigeonpea for increased SPC.

5.1 Introduction

The available genomic resources in pigeonpea such as a reference genome (Varshney et al., 2012) and whole genome re-sequencing (WGRS) data (Kumar et al., 2016) provide an opportunity to improve productivity and quality traits in this crop through modern/molecular breeding approaches. A major quality trait in pigeonpea is its seed protein content (SPC), but the trait has remained untouched by the genomics revolution. The trait could benefit from breeding approaches such as genomic assisted breeding (GAB) for accelerated genetic gain. However, the first step in GAB is the identification of candidate genes or genetic markers associated with the trait(s) of interest (Feng et al., 2014), which in turn provides the breeder with a critical tool to modify those traits (Janninks, 2001).

The recent developments in next generation sequencing (NGS) technologies provide rapid and cost-effective methods to identify candidate genes and variants underlying qualitative and quantitative traits (Silva et al., 2012; Xu et al., 2014). In the presence of a reference genome sequence, WGRS data of one or a few individuals can be used to identify variants associated with phenotype of interest as demonstrated in humans (Rios et al., 2010; Roach et al., 2010; Sobreiro et al., 2010), chicken (Jang et al. 2014) and in crop plants such as rice (Lim et al., 2014; Silva et al., 2012), maize (Xu et al., 2014) and recently in pigeonpea (Varshney et al. 2017). To identify candidate genes using WGRS data, common variant (CV) and clustering analyses have been proposed and used (Silva et al., 2012; Xu et al., 2014). However, Silva et al. (2012) did not find any significant difference between the two analysis methods while Xu et al. (2014) found the CV analysis to be more efficient than the clustering approach.

Next generation sequencing as used for generating WGRS data of the parental lines produces short reads, which may result in to misalignments to the reference genome (Church et al., 2011). Thus, validation of sequence variants identified from NGS-based approaches must be done to determine the analytical sensitivity and analytical specificity by comparing NGS test results to those obtained from an independently validated method such as Sanger sequencing. Sanger sequencing is less prone to sequencing errors than NGS (Machado et al., 2011) and has preferentially been used to validate the presence of SNPs by sequencing the fragments containing the candidate SNPs.

The final testing of the role of candidate gene mutations can be carried out by conventional co-segregation analysis in structured population such as F_2, or by SNP-phenotype associations in germplasm collections or natural populations, or in functional experiments (Pflieger et al., 2001; Grattapaglia, 2008; Gilissen et al., 2012). In addition, combining WGRS candidate gene search with linkage analysis provides a more efficient approach for pinpointing the casual variants affecting a phenotype (Sobreira et al., 2010).

In view of the above, this study has been designed to identify candidate genes for SPC in pigeonpea by (i) identifying SNPs from WGRS data that play roles in seed storage protein accumulation, (ii) identifying corresponding candidate genes to identified SNPs, (iii) validating presence of the SNPs in candidate genes through Sanger sequencing, and (iv) determining the association of the SNPs/candidate genes with SPC in segregating mapping population.

5.2 Materials and methods

5.2.1 Plant material and seed protein content estimation

The plant materials included five pigeonpea (*C. cajan*) cultivars and one wild relative species (*C. scarabaoiedes*) (Table 5.1). The WGRS data of HPL 24, ICP 5529 and ICP 11605, and the draft genome of ICPL 87119 were used for the identification of putative candidate nsSNPs and genes. HPL 24 and ICP 11605 were used to validate presence of nsSNPs through Sanger sequencing. UQ 50 and ICPW 90 were included as independent genetic background for checking amplification of the primers, and also to facilitate comparison of read alignments across multiple individuals, which have the potential to filter out nsSNPs that are an artefact of inaccurate read alignments (Bansal et al., 2010). The genotypes were chosen based on historical information, discussion with the pigeonpea breeders and results of field evaluation as reported in Chapter Two of this thesis. To assess the co-segregation of the identified nsSNPs with SPC two parental lines (ICP 5529 and ICP 11605) with contrasting SPC values, and their segregating F_2 population were used.

Table 5.1: Pigeonpea lines and segregating populations used for the identification and validation of candidate genes for seed protein content

Pedigree	Description
HPL 24	Breeding line with high SPC. WGRS data available (Kumar et al. 2016)
ICPW 90	*C. scarabaoiedes* (a wild relative of *C. cajanus*). Presumably previously used to develop high SPC breeding lines
UG 50	Breeding line with moderate SPC. WGRS data available (Kumar et al. 2016)
ICP 5529	Landrace with high SPC. WGRS data available (Kumar et al. 2016)
ICP 11605	Germplasm line with low SPC. WGRS data available (Kumar et al. 2016)
ICPL 87119	Germplasm line with low SPC. Reference genome available (Varshney et al. 2016)
ICP 5529 × ICP 11605	F_2 mapping population segregating for SPC

SPC, seed protein content; WGRS, whole genome resequencing;

5.2.2 Seed protein content phenotyping

The five pigeonpea genotypes as well as one wild relative and 188 F_2 progeny of the cross between ICP 5529 × ICP 11605 were grown under under field conditions. Pigeonpea genotypes and wild relative accession were sown in single rows each while the F_2s were in 19 rows. Each row was 4 m long with row to row and plant to plant spacing of 75 cm and 30 cm, respectively. All cultural practices were carried out as described in sub-section 4.2.1 Chapter Four of this thesis.

Ten grams of mature dry clean seeds of three plants each per parental line and 188 F_2 plants were analyzed at the Central Analytical Services (Charles Renard Analytical) Laboratory at ICRISAT, India. Before grinding, seeds were oven-dried at 60ºC for 48 hours. The dried seed samples were ground into powder in a mill with Teflon chambers. The ground samples were again kept in an oven at 60ºC overnight. Samples and appropriate blanks were digested simultaneously in duplicate (i.e. two independent analyses) using tri-acid digestion procedure as described in Upadhyaya et al. (2016). Briefly, 1.0 g of the ground seed sample was transferred to a 75 ml digestion tube containing 10 ml of tri-acid mixture of nitric, sulfuric and perchloric acids in the ratio of 10:0.5:2 (v/v). The contents were cold-digested overnight in a digestion chamber. Colourless and clear digest were obtained by keeping the samples at 120ºC for 1 hour followed by digestion at 230ºC for 2 hours. After cooling, the digests were dissolved

in distilled water and volume topped up to 75 ml and then mixed well by shaking. Aliquots were obtained from the digests and used to estimate the total nitrogen (N) using a San++Automated Wet Chemistry Analyzer (Skalar, Breda, The Netherlands). Seed protein of a sample was estimated by multiplying its N (%) content by factor 6.25.

5.2.3 Whole genome resequencing data and sequence variant identification

Existing WGRS data of each of HPL 24, ICP 5529, and ICP 11605 (Kumar et al., 2016), were cleaned and trimmed to remove poor quality bases using Sickle (Joshi and Fass, 2011). The cleaned data were aligned on to version 1.0 of the pigeonpea reference genome (Varshney et al. 2012) using Bowtie 2 version 2.0 (Langmead and Salzberg, 2012) and the unique hits were retained for further analyses in the Binary Alignment/Map (BAM) (Li et al., 2009) files. The BAM files were processed using the IndelRealigner component of the genome analysis tool kit (GATK) version 4.0 suite (DePristo et al. 2011) and sequence variants were detected using the UnifiedGenotyper of GATK version 4.0 (DePristo et al., 2011). A position in a genotype was reported as a sequence variant if the Phred quality score for the base was ≥ 30 and if the number of sequence reads aligned in each of the lines against the reference genome was ≥ 5. Only one sequence variant was retained and reported if two or more sequence variants were present in a 5 bp window. The sequence variants obtained in the last step were then subjected to the common variant analysis (CV) (Silva et al., 2012) to identify candidate SNPs and genes.

5.2.4 Common variant analysis

The CV analysis was performed as follows; the sequence variations within the high and within the low SPC genotypes were compared. SNPs for which the allelic calls in HPL 24 was the same as in ICP 5529 but contrasting with that in ICP 11605 and ICPL 87119 (in which the calls in ICP 11605 was the same as that in ICPL 87119 as a genotype) were retained for further analysis. The SNPs were subjected to their effects using snpEff program (Cingolani et al. 2012). Annotation of the genes containing SNPs was carried out using BLASTX against SWISS-PROT and TREMBL databases. Corresponding gene ontologies for the genes were extracted using UniprotKB database (UniProt Consortium, 2008). Where UniprotKB database returns an uncharacterized protein, the *C. cajan* gene identifier (ID) was submitted to LegumeIP v2.0 (Li et al., 2012) to search for gene/protein function category within the integrated legume database. Potential causal SNPs that result in non-synonymous changes in

the coding DNA sequence (CDS) regions were identified by filtering out intergenic, intronic, and synonymous SNPs (sSNPs). Heterozygous calls were also removed from the list of sequence variants. A final selection of the candidates was based on information on gene function in relation to the SPC, resulting in nsSNPs, stop gains, frame-shifts, small insertion/deletion and splice-site mutations) to be retained as selected putative candidates for further analysis.

5.2.5 Sanger sequencing-based SNP validation

Genomic DNA (gDNA) was isolated from young trifoliate leaves using CTAB method (Mace et al., 2003) and then column purified using MN plant DNA purification kit (MN-Germany) following the manufacturer's instructions.

Sequences of approximately 350 bps flanking either side of the identified SNP sites were extracted using the pigeonpea reference genome. PCR primers of length 21-24 bp and Tm of 56–59.5°C were designed from each 601 bp sequence using BatchPrimer3 v1.0 primer design software tool (You et al., 2008).

Polymerase chain reaction (PCR) was performed for each of the selected variants in a total volume of 30 µL containing 21.9 µL of ddH_20, 10× Taq polymerase buffer, 2.0 µL of 2 mM dNTPs, 10 pmol/µL of each of the forward and reverse primers, 0.06 µL of Taq polymerase and 2.0 µL of 20 ng/µL gDNA. A touch-down PCR (Korbie and Martick, 2008) was used as follows: initial denaturation at 95 ^{0}C for 5 min followed by (1) 5 cycles consisting of (i) 94 ^{0}C for 15 sec, (ii) 62 ^{0}C for 20 sec and (iii) 72 ^{0}C for 30 sec, (2) 35 cycles consisting of (i) 94 ^{0}C for 15 sec, (ii) 54 ^{0}C for 30 sec and (iii) 72 ^{0}C for 30 sec, and a final extension of 72 ^{0}C for 20 min. PCR products were run in 3.5% Nusieve agarose gel. Gels were stained with ethidium bromide and visualized under UV light in a transilluminator.

Only PCR products showing single clear bands across the four genotypes were further processed for Sanger sequencing. PCR cleanup reactions were then performed by mixing 20 µL of PCR products with 1.1 µL of ExoSAP-IT (USB Corporation, Cleveland, OH) and incubating the mixture for 45 min at 37 ^{0}C followed by 15 min at 80 ^{0}C. Ten µL of each of the cleaned PCR products were vacuum dried and end-sequenced using forward and reverse

primers at Macrogen Korea (https://dna.macrogen.com/eng/). The two sequences generated by the forward and reverse primers from each genotype were combined into genotype-specific contigs. The genotype-specific contigs from all the four genotypes were compared with the reference sequence of Asha at the originally targeted SNP position using DNA Baser Sequence Assembler v4.23 (Heracle BioSoft, http://www.DnaBaser.com).

5.2.5 Cleaved amplified polymorphic sequence and derived-cleaved amplified polymorphic sequence primer design, PCR amplification and restriction digestion

Cleaved amplified polymorphic sequence (CAPS) and derived-CAPS (dCAPS) primers were designed by submitting 22-24 bp sequences flanking the nsSNP position for both 'wild-type' and 'mutant-type' alleles using online software dCAPS Finder 2.0 (Neff et al., 2002). In the case of dCAPS, suitable primers picked were those having single mismatches at position 1 or 2 from the 3' end. In case of nsSNPs for which neither CAPS nor suitable dCAPS (with mismatches at 3'end positions 1 or 2) were found, dCAPS with two mismatches at positions 4,5 or 6,7 or 7,9 were selected as suggested by Micheals and Amasimo (1998). Because the dCAPS Finder software generates only either a forward or reverse primer sequence in the case of dCAPS, the complementary strand of any chosen dCAPS primer was designed by submitting the 601 bp long reference fragment containing the appropriate SNPS allele (either wild type or mutant type) to Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) with the default settings.

PCR amplification and gel visualisation for the CAPS and dCAPS markers were performed as described under Sanger sequencing-based SNP validation in sub-section 5.2.5 of this chapter. Restriction digestion was performed in 30 µL reaction volume containing 17 µL of ddH_20, 1.0 µL restriction enzyme (RE), 2.0 µL RE buffer, and 10 µL PCR product. The digestion mixture was incubated at 37 to 50 ºC for 2 to 3 hr, and held at 0 to 80 ºC for 20 min depending on RE and the manufacturer's instructions.

5.2.6 Integration of CAPS/dCAPS markers in to genetic map and marker-trait association analysis

The CAPS/dCAPS genotyping data generated from 188 F_2 plants derived from cross ICP 5529 × ICP 11605 were combined with a GBS-derived SNP data already available on the same population (ICP 5529 × ICP 11605). The GBS data was obtained as described in sections 4.2.4 on page 79 and section 4.2.5 on page 80 and genetic map was constructed as described in sub-section 4.2.6 in Chapter Four on of this thesis.

Positions of the candidate gene CAPS/dCAPS markers on the genetic map were compared with that of main effect (M-QTLs) for SPC to assess their co-localisation. To assess co-segregation of the CAPS/dCAPS markers with SPC, single marker regression analysis (SMA) was carried out in Excel 2013 (Microsoft) using the F_2 CAPS/dCAPS marker genotypes as independent variables and the F_2 phenotypes as dependent variables. The F_2 phenotypic data were generated on single F_2 plants of the population ICP 5529 × ICP 11605 as described in section 4.2.3 in Chapter Four of this thesis.

5.3 Results

5.3.1 Sequence variants between high and low seed protein content groups

Sequencing data on genotypes obtained from Kumar et al. (2016) were used for alignment the draft genome and sequence variants detection. All the detected sequence variants were subjected to CV analysis as mentioned in material and method section. As a result, a total of 32,964 sequence variants were found between the high (HPL 24, ICP 5529) and low (ICP 11605, ICPL 87119) SPC groups. (Table 5.2). Intergenic region had the highest number of variants (83.4%) followed by the sequence variants present in intronic (12.6%) and the exonic (3.8%) regions. There were 485 synonymous SNPs (sSNPs), 718 nsSNPs, 26 stop-gains and one each of stop-loss and start-loss mutations in the exonic regions. Other sequence variant types identified in the exons included splice-sites (0.003%), indels (0.003%) and frameshifts (0.07%). Non-synonymous SNPs were more abundant with an average nsSNPs to sSNPs (Nonsyn/Syn) ratio of 1.48 (Table 5.2), which is close to 1.46 estimated previously (Kumar et al., 2016). To identify potential causal mutations that induce protein coding alterations, the

present study focused on mutations that lead to changes in amino acid sequences. Such changes included nsSNPs, stop-gains and splice-sites in the coding regions, and frameshift- and indel-mutations in coding and/or noncoding regions.

Table 5.2: Summary of type and number of detected variants, and genes and their distribution in different genomic regions of pigeonpea

		Exonic region										
Chr.	Total SNPs	sSNPs	nsSNPs	Stop-gain	Stop-loss	Start- loss	Splice sites	Intronic	Indels	Frame shifts	Intergenic	Het
CcLG01	1721	35	46	1	0	0	2	166	0	1	1470	342
CcLG02	2430	42	51	1	0	0	0	420	1	2	1913	692
CcLG03	1425	18	30	0	0	0	0	196	2	3	1176	405
CcLG04	925	15	22	1	0	0	0	141	0	1	744	168
CcLG05	171	3	6	0	0	0	0	34	1	0	127	75
CcLG06	726	17	18	1	0	0	0	108	1	1	580	341
CcLG07	1105	15	25	0	0	0	1	147	0	3	914	306
CcLG08	1436	16	31	0	0	0	0	202	0	0	1187	245
CcLG09	514	6	14	0	1	0	0	79	0	0	414	178
CcLG10	1016	11	23	0	0	0	0	106	0	0	876	526
CcLG11	2564	40	57	1	0	0	0	251	0	0	2215	692
Scaffolds†	18931	268	395	21	0	1	7	2244	5	11	15364	6091
Total	32964	485	718	26	1	1	10	4086	10	22	26979	10061
Distribution (%)	100	1.47	2.18	0.08	0.00	0.00	0.03	12.40	0.03	0.07	81.84	30.52

Chr., chromosome; nsSNPs, nonsynonymous SNPs; sSNPs, synonymous SNPs; Het, variants coded in the high protein group as Het, K, M, S, R, Y and W; † Includes variants without variant effects thus resulting in unbalanced totals for variants from scaffolds

5.3.2 Candidate genes for seed protein content

From sequence variants detected in the exonic regions, a total of 108 sequence variants were identified using the heuristic CV analysis in combination with prior knowledge of gene functions from literature and gene annotations (Table 5.3). These sequence variants were present in 57 pigeonpea genes spread over all chromosomes/pseudomolecules and several scaffolds except CcLG05 (Table 5.3). The distribution of selected sequence variants, and corresponding genes across chromosomes was not uniform. For example, a maximum of 25 sequence variants and nine genes were found on CcLG01 whereas 1, 3, 5, 17, and 19 sequence variants and 1, 3, 4, 9 and 4 genes were detected on CcLG09, CcLG 11, CcLG03, CcLG02 and CcLG07, respectively (Table 5.3). A considerable number of sequence variants and genes (14 and 9, respectively) were present in nine unanchored scaffolds. For simplicity, and owing to the abundance of nsSNPs among selected variant types, the terms sequence variant and nsSNP are invariable used hereafter.

Table 5.3: The 108 variants with ≥5 read depths that induce amino acid changes related to seed protein content metabolism

Chr.	Position	Gene ID	Ref. allele	Alt. allele	Effect	AA change	Uniprot_ID
CcLG01	4414830	*C.cajan_19617*	A	C	N-Syn†	gaT/gaG	Q9FX32
CcLG01	4415753	*C.cajan_19617*	T	C	N-Syn	aAa/aGa	Q9FX32
CcLG01	4992638	*C.cajan_19670*	T	C	N-Syn	tAc/tGc	C6SY88
CcLG01	16162501	*C.cajan_20717*	A	C	N-Syn	Aaa/Caa	Q39110
CcLG01	16163064	*C.cajan_20717*	C	G	N-Syn	Caa/Gaa	Q39110
CcLG01	16855461	*C.cajan_20775*	T	C	N-Syn	Aag/Gag	O04609
CcLG01	16855539	*C.cajan_20775*	A	G	N-Syn	Tcg/Ccg	O04609
CcLG01	16873384	*C.cajan_20776*	C	T	N-Syn	gCa/gTa	K7MEZ9
CcLG01	16873606	*C.cajan_20776*	T	A	N-Syn	ttT/ttA	K7MEZ9
CcLG01	16873818	*C.cajan_20776*	T	C	N-Syn	gTt/gCt	K7MEZ9
CcLG01	16873826	*C.cajan_20776*	A	G	N-Syn	Aac/Gac	K7MEZ9
CcLG01	16874033	*C.cajan_20776*	C	T	N-Syn	Ctt/Ttt	K7MEZ9
CcLG01	17475532	*C.cajan_20841*	A	T	N-Syn	tAt/tTt	O82345
CcLG01	17476096	*C.cajan_20841*	T	C	N-Syn	tTt/tCt	O82345
CcLG01	17479259	*C.cajan_20841*	C	T	N-Syn	tCc/tTc	O82345

Scaf_, Scaffold. † Nonsynonymous SNP; § Splice site acceptor; §§ Codon insertion; ‡ Stop-gained; ¶ Codon change + Codon deletion; †† Frameshift; ‡‡ Splice site donor.

Table 5.3: (continued)

Chr.	Position	Gene ID	Ref. allele	Alt. allele	Effect	AA change	Uniprot_ID
CcLG01	17482682	*C.cajan_20842*	C	T	N-Syn	Ctt/Ttt	P29060
CcLG01	17486133	*C.cajan_20843*	A	T	N-Syn	Acc/Tcc	P29060
CcLG01	17486917	*C.cajan_20843*	A	T	SS-Acc§	-	P29060
CcLG01	205461	*C.cajan_19199*	G	T	SS-Acc	-	P10978
CcLG01	205580	*C.cajan_19199*	A	G	N-Syn	Acc/Gcc	P10978
CcLG01	205708	*C.cajan_19199*	A	G	N-Syn	atA/atG	P10978
CcLG01	16876689	*C.cajan_20777*	C	T	N-Syn	aCa/aTa	P10978
CcLG01	16876698	*C.cajan_20777*	G	C	N-Syn	aGt/aCt	P10978
CcLG01	16876707	*C.cajan_20777*	A	G	N-Syn	aAa/aGa	P10978
CcLG01	16876742	*C.cajan_20777*	G	A	N-Syn	Gga/Aga	P10978
CcLG02	1201138	*C.cajan_04622*	C	T	N-Syn	cGt/cAt	Q03460
CcLG02	1204754	*C.cajan_04622*	G	C	N-Syn	atC/atG	Q03460
CcLG02	3074263	*C.cajan_04797*	G	GCTC	Insertion§§	-/GAG	I1M9S9
CcLG02	8895098	*C.cajan_05310*	C	G	N-Syn	gCg/gGg	P13917
CcLG02	12525321	*C.cajan_05664*	A	G	N-Syn	aTg/aCg	O24606
CcLG02	16750258	*C.cajan_06087*	T	A	N-Syn	Tcc/Acc	Q8GS60
CcLG02	16756557	*C.cajan_06087*	T	G	N-Syn	ttT/ttG	Q8GS60
CcLG02	16756576	*C.cajan_06087*	A	T	N-Syn	Atg/Ttg	Q8GS60
CcLG02	24499331	*C.cajan_06764*	G	A	N-Syn	Cgt/Tgt	Q9S7G7
CcLG02	36162648	*C.cajan_07942*	C	T	N-Syn	Gag/Aag	Q54JE4
CcLG02	16726667	*C.cajan_06086*	A	C	N-Syn	agA/agC	K7LJ30
CcLG02	16726668	*C.cajan_06086*	A	G	N-Syn	Att/Gtt	K7LJ30
CcLG02	16726731	*C.cajan_06086*	T	G	N-Syn	Tgg/Ggg	K7LJ30
CcLG02	16727058	*C.cajan_06086*	A	T	N-Syn	Atc/Ttc	K7LJ30
CcLG02	16727494	*C.cajan_06086*	T	C	N-Syn	cTg/cCg	K7LJ30
CcLG02	16736680	*C.cajan_06086*	A	C	N-Syn	tAt/tCt	K7LJ30
CcLG02	11940360	*C.cajan_05609*	T	G	Stp_Gained‡	tTa/tGa	I1M7K5
CcLG03	460462	*C.cajan_08095*	C	G	N-Syn	ttC/ttG	Q8L5R3
CcLG03	20453445	*C.cajan_10047*	C	T	N-Syn	Gag/Aag	Q7PC84
CcLG03	20477859	*C.cajan_10048*	G	C	N-Syn	Cta/Gta	Q7PC87
CcLG03	8244468	*C.cajan_08817*	G	A	N-Syn	tCa/tTa	K7KUS1
CcLG03	8243892	*C.cajan_08817*	CCTT	C	Cod deletion¶	gaaggc/ggc	K7KUS1
CcLG04	428893	*C.cajan_20905*	A	G	N-Syn	atA/atG	Q9FII5

Table 5.3: (continued)

Chr.	Position	Gene ID	Ref. allele	Alt. allele	Effect	AA change	Uniprot_ID
CcLG04	496463	*C.cajan_20914*	C	T	N-Syn	Ctt/Ttt	Q8U4K7
CcLG04	892871	*C.cajan_20963*	TTCCCC	T	FS††	-	B9RXW0
CcLG04	10982012	*C.cajan_21987*	G	A	Stp_Gained	tGg/tAg	Q9UR07
CcLG06	6651663	*C.cajan_11650*	A	C	N-Syn	Aat/Cat	Q41706
CcLG06	14406147	*C.cajan_12368*	A	T	N-Syn	Agg/Tgg	Q7XA39
CcLG07	9512986	*C.cajan_18213*	C	T	N-Syn	aGg/aAg	Q41385
CcLG07	9689686	*C.cajan_18233*	T	G	N-Syn	gaA/gaC	Q9S9K4
CcLG07	12293438	*C.cajan_18443*	C	G	N-Syn	Gat/Cat	Q9T048
CcLG07	12293448	*C.cajan_18443*	T	A	N-Syn	aaA/aaT	Q9T048
CcLG07	12293465	*C.cajan_18443*	C	A	N-Syn	Gat/Tat	Q9T048
CcLG07	12293471	*C.cajan_18443*	T	C	N-Syn	Aag/Gag	Q9T048
CcLG07	12293474	*C.cajan_18443*	T	C	N-Syn	Aag/Gag	Q9T048
CcLG07	12293485	*C.cajan_18443*	T	A	N-Syn	gAa/gTa	Q9T048
CcLG07	12293486	*C.cajan_18443*	C	T	N-Syn	Gaa/Aaa	Q9T048
CcLG07	12293498	*C.cajan_18443*	T	A	N-Syn	Ata/Tta	Q9T048
CcLG07	12293515	*C.cajan_18443*	G	A	N-Syn	aCa/aTa	Q9T048
CcLG07	12293541	*C.cajan_18443*	A	C	N-Syn	aaT/aaG	Q9T048
CcLG07	12293560	*C.cajan_18443*	C	G,T	N-Syn	aGg/aAg	Q9T048
CcLG07	12299616	*C.cajan_18443*	T	A	N-Syn	agA/agT	Q9T048
CcLG07	12300283	*C.cajan_18443*	T	A	N-Syn	gAg/gTg	Q9T048
CcLG07	12300597	*C.cajan_18443*	A	C	N-Syn	atT/atG	Q9T048
CcLG07	12301081	*C.cajan_18443*	C	T	SS-donor‡‡	-	Q9T048
CcLG07	12301614	*C.cajan_18443*	T	G	N-Syn	aAc/aCc	Q9T048
CcLG07	10222142	*C.cajan_18280*	G	A	N-Syn	Gcg/Acg	P04323
CcLG08	900949	*C.cajan_15445*	T	G	N-Syn	aaA/aaC	C0LGR6
CcLG08	901014	*C.cajan_15445*	T	G	N-Syn	Atg/Ctg	C0LGR6
CcLG08	902979	*C.cajan_15445*	C	A	N-Syn	Gtt/Ttt	C0LGR6
CcLG08	1729778	*C.cajan_15499*	T	C	N-Syn	Tcc/Ccc	Q6AX33
CcLG08	1797807	*C.cajan_15508*	A	G	N-Syn	cAt/cGt	P15170
CcLG08	1797810	*C.cajan_15508*	T	C	N-Syn	gTt/gCt	P15170
CcLG08	5286893	*C.cajan_15863*	A	T	N-Syn	tTc/tAc	Q9M0X5
CcLG08	5287380	*C.cajan_15863*	C	T	N-Syn	Gct/Act	Q9M0X5

Table 5.3: (continued)

Chr.	Position	Gene ID	Ref. allele	Alt. allele	Effect	AA change	Uniprot_ID
CcLG08	14254343	*C.cajan_16632*	A	G	N-Syn	Tcc/Ccc	Q9LT96
CcLG08	15677727	*C.cajan_16775*	A	G	N-Syn	Aca/Gca	Q8TX03
CcLG08	15858055	*C.cajan_16795*	C	G	N-Syn	aCa/aGa	O65388
CcLG08	17542631	*C.cajan_16978*	T	G	N-Syn	Acg/Ccg	D3DFG8
CcLG08	8702222	*C.cajan_16102*	A	G	N-Syn	Tgc/Cgc	Q7LHG5
CcLG08	4923892	*C.cajan_15828*	A	G	N-Syn	gAa/gGa	P10978
CcLG09	3758497	*C.cajan_22447*	G	T	N-Syn	gCt/gAt	P93194
CcLG10	7560344	*C.cajan_14054*	T	C	N-Syn	Atg/Gtg	Q99315
CcLG10	7560364	*C.cajan_14054*	T	G	N-Syn	gAa/gCa	Q99315
CcLG10	7560516	*C.cajan_14054*	A	C	N-Syn	aaT/aaG	Q99315
CcLG10	7560536	*C.cajan_14054*	T	C	N-Syn	Atg/Gtg	Q99315
CcLG11	25475840	*C.cajan_02320*	G	C	N-Syn	Gag/Cag	P81391
CcLG11	7967151	*C.cajan_00758*	A	G	N-Syn	cTg/cCg	P04323
CcLG11	8413398	*C.cajan_00788*	C	T	N-Syn	Gca/Aca	P10978
Scaf_000059	549145	*C.cajan_28359*	C	A	N-Syn	gCa/gAa	P93841
Scaf_000059	551244	*C.cajan_28359*	C	A	N-Syn	gaC/gaA	P93841
Scaf_000213	6068	*C.cajan_35646*	T	G	N-Syn	ttT/ttG	P10978
Scaf_000213	6136	*C.cajan_35646*	T	C	N-Syn	tTa/tCa	P10978
Scaf_000379	105750	*C.cajan_36235*	A	T	N-Syn	aAt/aTt	Q33282
Scaf_117710	1241	*C.cajan_47566*	C	A	Stp-Gained	Gaa/Taa	P15629
Scaf_117710	3147	*C.cajan_47566*	A	G	N-Syn	tTg/tCg	P15629
Scaf_132767	14801	*C.cajan_44248*	T	C	N-Syn	aAt/aGt	P49092
Scaf_135136	45544	*C.cajan_35089*	T	A	N-Syn	tTt/tAt	Q07423
Scaf_135136	45624	*C.cajan_35089*	A	C	N-Syn	Aaa/Caa	Q07423
Scaf_135136	45693	*C.cajan_35089*	G	T	N-Syn	Gtc/Ttc	Q07423
Scaf_135140	150820	*C.cajan_38328*	G	T	Stp-Gained	Gaa/Taa	P10978
Scaf_135147	20168	*C.cajan_45411*	A	G	N-Syn	Act/Gct	P10978
Scaf_136850	11420	*C.cajan_46295*	G	A	Stp-Gained	Cag/Tag	A2Q2J0

5.3.3 Functional groups and roles of selected candidate genes

The 57 identified candidate genes could be placed in 19 functional categories based on GO molecular function (Fig. 5.1). The functional groups which were highly represented in terms of selected genes include aspartic-type endopeptidase (protease), ATP binding/ATPase, DNA binding, iron ion binding, metal iron binding and chitinase activity with 17, 15, four, three, three and two genes, respectively (Fig. 5.1). The remaining functional categories contained one gene each (Fig. 5.1). These selected genes have also been identified in other plant species playing important roles in SPC accumulation (Table 5.4).

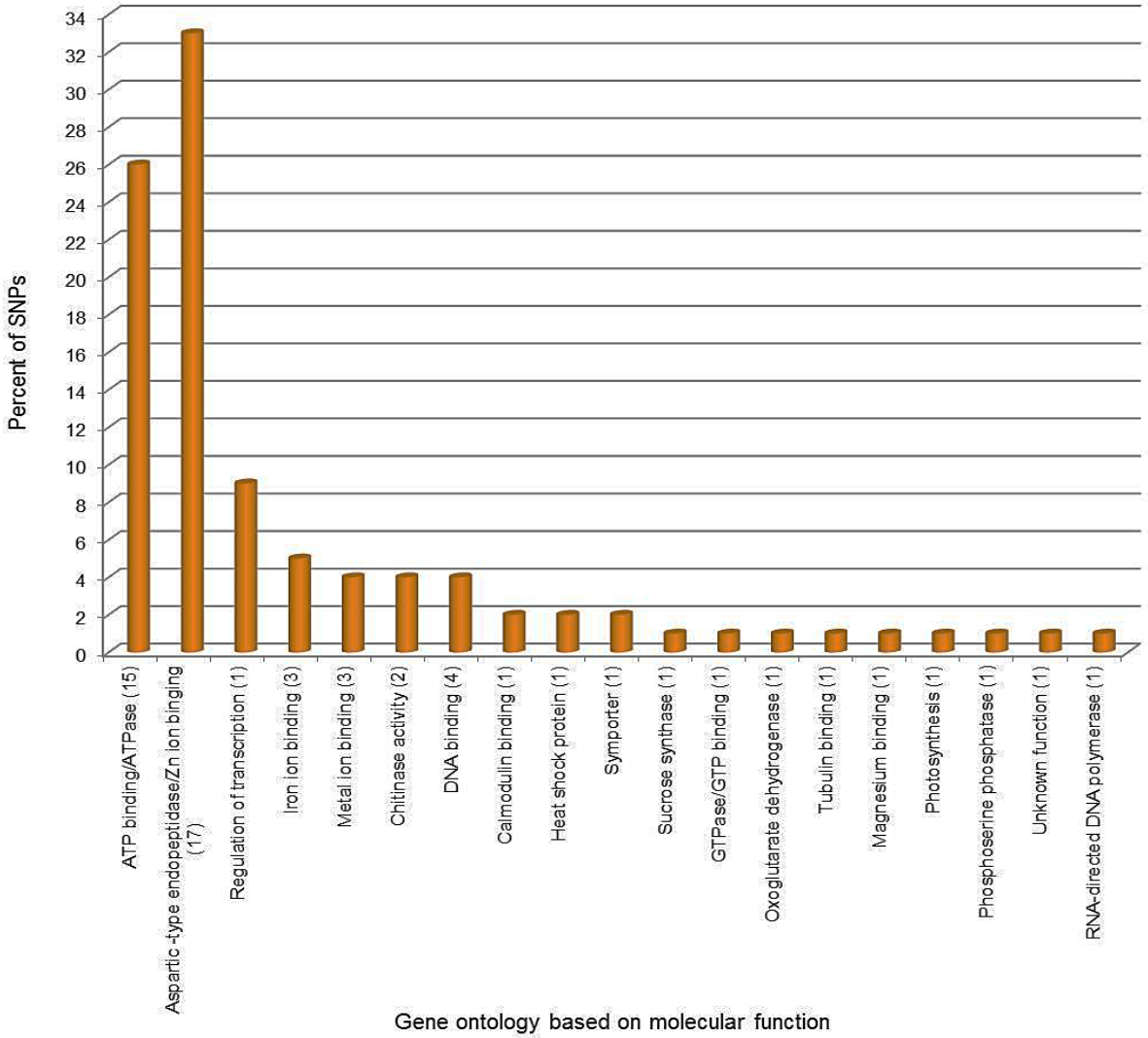


Fig. 5.1: Grouping of common variant-selected candidate genes based on GO molecular function. Number in parenthesis on horizontal axis represents the number of genes in the category.

Table 5.4: Function of proteins of selected candidate genes containing protein-changing variants related to seed protein content

Protein name	*Cajanus cajan* Gene ID	Role in (or relationship to) seed protein content
MYB transcription factor	*C.cajan_06764*	Remarkable up-regulation of many genes including seed maturation, dehydration, desiccation and storage protein genes in AtMYB118-over-expressed transgenic plants (Zhang et al., 2008).
2-oxoglutarate dehydrogenase (OGDH) (EC 1.2.4.2)	*C.cajan_07942*	Catalyses formation of 2-oxoglutarate (2OG), which provides the carbon skeleton required by GOGAT for net glutamate production (Hodges et al., 2002). Strong perturbation in the relative abundance of amino acids due to the OGDH inhibition was accompanied by decreased protein content (Araújo et al., 2013)
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148)	*C.cajan_28359*	Involved in monoterpene biosynthesis pathway (Nieuwenhuizen et al. 2015). Monoterpenes are involved in leaf senescence (Liu et al., 2016). Senescence leads to remobilized N from the senescing leaf tissues to the seeds for storage protein synthesis (Cohen et al., 2014).
7-hydroxymethyl chlorophyll a reductase, chloroplastic (EC 1.17.7.2)	*C.cajan_06087*	It is an important component of the interconversion pathway of Chlorophyll *a* to Chlorophyll *b* known as the chlorophyll cycle, which plays a crucial role in the processes of greening, acclimation to light intensity, and senescence (Meguro et al., 2011). Senescence leads to remobilized N from the senescing leaf tissues to the seeds for storage protein synthesis (Cohen et al., 2014).
ABC transporter ABCG.39	*C.cajan_10047, C.cajan_10048*	Belongs to the group of solute transporters involved in the transport of a broad range of substances, including sugars, peptides, alkaloids, and inorganic anions (Marty, 1999). A synonymous SNP in an ABC transporter gene was found strongly correlated (R^2 = 20%) with SPC in chickpea (Upadhyaya et al., 2016).
BLISTER	*C.cajan_06086*†	Found on *Glycine max* chromosome 20, a region harbouring a major QTL for seed protein content (Lestari et al., 2013).

† Information on gene functional annotation obtained from LegumeIP v2.0 (http://plantgrn.noble.org/LegumeIP/)

Table 5.4: (continued)

Protein name	*Cajanus cajan* Gene ID	Role in (or relationship to) seed protein content
Alpha-farnesene synthase (EC 4.2.3.46) ((E,E)-alpha-farnesene synthase)	*C.cajan_20963*	Are involved in terpenoid biosynthesis (Rattanakon et al., 2016). Terpenoid derivatives promoted senescing process of barley leaf segments (Jung and Grossmann, 1985). Senescence leads to remobilized N from the senescing tissues to the seeds for storage protein synthesis (Cohen et al., 2014).
Asparagine synthetase [glutamine-hydrolyzing] 1 (EC 6.3.5.4)	*C.cajan_44248*	Asparagine metabolic genes asparagine synthetase (AS) together with asparaginase (ASPG) associated with increased SPC in soybean (Pandurangan et al., 2012). AS enhances nitrogen status in seeds of plants (Lam et al., 2003).
Basic 7S globulin (SBg7S) | Nutrient reservoir activity	*C.cajan_05310*	Plays roles in seed storage albumin protein processing (Gruis et al., 2002). Associated with increased SPC in soybean (Krishnan et al., 2012) and found highly upregulated in high SPC genotypes of soybean (Bolon et al., 2010).
Disease resistance protein At4g27190	*C.cajan_18443*	Several disease resistance response protein genes found within major QTL for SPC on soybean chromosome 20 and its duplicated region in chromosome 10 (Lestari et al., 2013)
Gibberellin 20 oxidase 1 (EC 1.14.11.)	*C.cajan_20717*	In Arabidopsis, GA20ox overexpression leads to an increase in GA concentration (Huang et al., 1998; Coles et al., 1999). Exogenous application of GA3 increased total SPC during late seed filling stage in oil seed rape (Huang et al., 2014).
Tubulin-folding cofactor D (AtTFCD) (Protein TITAN 1)	*C.cajan_08095*	Play a role in protein or vesicular traffic inside the cell (Lopez-Fanarraga et al., 2001).
Proteasome-activating nucleotidase (PAN)	*C.cajan_16775*	PANs are involved in a wide range of biological processes including proteolysis, protein folding, transcriptional regulation, membrane trafficking and microtubule regulations (Santos et al., 2006).

Table 5.4: (continued)

Protein name	*Cajanus cajan* gene ID	Role in (or relationship to) seed protein content
Glutamate synthase (NADH-GOGAT) (EC 1.4.1.14)	*C.cajan_04622*	Catalyses the first step in nitrogen assimilation in plants resulting in the assimilation of NH_4^+ into glutamine (Gaur et al., 2012; Guan et al., 2015). One of the potential candidate genes involved in the control of the complex grain protein content in wheat (Nigro et al., 2013).
GTP-binding subunit	*C.cajan_15508*	Plays a role in diverse biological processes including protein synthesis, intracellular transport of proteins and membrane trafficking (Jha et al., 2012) Found on *Glycine max* chromosome 20, a region harbouring a major QTL for SPC (Lestari et al., 2013).
Heat Shock protein DnaJ (Hsp40)	*C.cajan_08817†*	DnaJ (Hsp40s) are involved in various cellular processes, including de novo protein folding, translocation of polypeptides across cellular membranes, and degradation of misfolded proteins (Ohta and Takaiwa, 2014) through substrate recognition and delivery to Hsp70 at the early stages of chaperone-mediated protein folding (Ohta et al., 2013). Heat shock protein genes were found to be expressed at higher levels in the low protein line of a near-isogenic line pair in barley (Jukanti et al., 2008) and in soybean (Bolon et al., 2010)
Photosystem I reaction centre subunit XI, chloroplastic (PSI-L) (PSI subunit V)	*C.cajan_18213*	A photochemical system containing P700, the chlorophyll a dimer that functions as a primary electron donor (Sobieszczuk-Nowicka et al., 2015). Chlorophyll a is part of chlorophyll cycle, which plays a crucial role in the processes of greening, acclimation to light intensity, and senescence (Meguro et al., 2011). Senescence leads to remobilized N from the senescing leaf tissues to the seeds for storage protein synthesis (Cohen et al., 2014).

Table 5.4: (continued)

Protein name	*Cajanus cajan* gene ID	Role in (or relationship to) seed protein content
Hexose carrier protein HEX6	*C.cajan_35089*	A hexose transporter gene located in the vicinity of a QTL for SPC in pea (Burstin et al., 2007).
Myb-related protein 305	*C.cajan_02320*	MYB induces the expression of several genes including proteinases through direct binding to gibberellic acid-responsive element (Gubler and Jacobsen, 1992; Gubler et al., 1999).
Photosystem II (PSII)/Photosystem Q(B) protein (EC 1.10.3.9)	*C.cajan_36235*	Highly significant positive correlation between PSII and SPC in soybean (Carrera et al., 2015).
Probable ureide permease A3	*C.cajan_11650*	Transport of ureides allantoin and allantoic acid forms of nitrogen from nodules, after biological nitrogen fixation (BNF), to the shoot (Rentsch et al., 2007; Collier and Tegeder, 2012).
Proteases (EC 3.4.23.-)	*C.cajan_00758, C.cajan_00788, C.cajan_14054, C.cajan_15828, C.cajan_18233, C.cajan_18280, C.cajan_19199, C.cajan_20777, C.cajan_35646, C.cajan_38328, C.cajan_47566, C.cajan_45411, C.cajan_46295, C.cajan_16102†*	Known to occur in seeds where they are involved in the processing of storage proteins during ripening and in their degradation during germination (Asakura et al., 2001; Pereira et al., 2008; Mazorra-Manzano et al., 2010). Involved in the maturation of seed storage proteins in castor bean (Hiraiwa et al., 1997)
Protein ETHYLENE INSENSITIVE-3 (EIN3)	*C.cajan_05664*	Transcription factor that initiates downstream transcriptional cascades for ethylene responses. Ethylene induces leaf senescence, which leads to remobilized N from senescing leaf tissues to the seeds for storage protein synthesis (Cohen et al., 2014; Schipper et al., 2015).
Cu transport protein	*C.cajan_05609*†	Increase in copper concentration leads to increase in total protein content of roots and shoots in *Phaseolus vulgaris* (Singh et al., 2014). Copper is found in several proteins (Bittencourt et al., 2014).

Table 5.4: (continued)

Protein name	Cajanus cajan gene ID	Role in (or relationship to) seed protein content
Putative pectate lyase 2 (EC 4.2.2.2)	*C.cajan_16795*	Legume pectate lyase is required for degradation of plant cell walls for root infection by rhizobia during initiation of the symbiotic interaction between rhizobia and legumes for BNF (Xie et al., 2011). Rapid BNF during pod fill (stages R5–R6) contributes to increased seed yield and seed protein content (Imsande, 1992; Fabre and Planchon, 2000).
Receptor-like protein kinases (EC 2.7.11.1)	*C.cajan_15445, C.cajan_15499, C.cajan_15863, C.cajan_16632, C.cajan_20905, C.cajan_22447*	Differentially expressed between low and high SPC near isogenic lines (NILs) of soybean (Bolon et al., 2010)
RING/U-box superfamily protein|C3HC4 –type Zinc finger protein	*C.cajan_04797*†, *C.cajan_20776*	Found on soybean chromosome 20, a region harbouring a major QTL for seed protein content (Lestari et al., 2013).
WRKY transcription factor	*C.cajan_20775*	A barley WRKY transcription factor was found to be upregulated in flag leaves during general senescence processes (Hollman et al., 2014). Senescence leads to remobilized N from the senescing leaf tissues to the seeds for storage protein synthesis (Cohen et al., 2014).
CCCH-type Zinc finger	*C.cajan_21987*†	Down-regulation of a CCHC-type zinc finer OsGZF1 by an RNAi approach increased grain nitrogen concentration in rice (Chen et al., 2014).
Acidic endochitinase (EC 3.2.1.14)	*C.cajan_20842, C.cajan_20843*	Abundant class III acidic chitinase homologue in tamarind (*Tamarindus indica*) seed serves as the major storage protein (Rao and Gowda, 2008).

Table 5.4 (continued)

Protein name	*Cajanus cajan* gene ID	Role in (or relationship to) seed protein content
Oxidative stress 3	*C.cajan_19670*	Related to leaf senescence (Blanvillain et al., 2009). May lead to remobilisation of N from leaf tissue for storage protein synthesis (Cohen et al., 2014; Schipper et al., 2015).
Sucrose synthase 6 (AtSUS6) (EC 2.4.1.13)	*C.cajan_19617*	One epistatic quantitative trait nucleotide (QTN) interaction between sucrose synthase (Sus) SNPs GhSus1At-A918G and GhSus1Dt-G2709C was found associated with increased SPC in cotton (Zeng et al., 2016).
2-phosphoglycerate kinase (2PGK) (EC 2.7.2.-)	*C.cajan_20914*	A gene encoding a phosphoglycerate kinase was found co-localised with a QTL for seed weight (SW) in pea. Owing to often negative correlation between SPC and SW, it is possible such genes may influence both traits (Burstin et al., 2007).
Putative disease resistance protein RGA4	*C.cajan_12368*	SDS-PAGE analysis in oat detected three storage protein (avenin) loci closely linked to resistance gene Pc68, which in turn, was found linked in repulsion to oat resistance gene analog (RGA) Orga2 (Satheeskumar et al., 2011).
BAG family molecular chaperone regulator 6	*C.cajan_20841*	AtBAG5 regulates leaf senescence by controlling the production of ROS and the expression of senescence-associated genes in Arabidopsis (Li et al., 2016). Meta QTL analysis in Arabidopsis showed co-localisation of plant senescence QTLs with that of seed N content (Chardon et al., 2014).
Phosphoserine phosphatase 1 (EC 3.1.3.3)	*C.cajan_16978*	Plastidial phosphoserine phosphatase, catalyzes the terminal step in Serine biosynthesis (Weigelt et al., 2009). Serine and glycine are biosynthetically linked, and together provide the essential precursors for the synthesis of proteins, nucleic acids, and lipids (Amelio et al., 2014).

5.3.4 Sanger sequencing of gene fragments containing candidate nsSNPs

Primer pairs were designed to amplify 108 nsSNP-containing fragments from 57 genes. A total of 86 nsSNP-containing gene fragments could be amplified and further processed for Sanger sequencing. Sixty-nine fragments from 42 genes were successfully sequenced (no missing genotype data) across the validation panel of two genotypes namely ICP 11605 (with low SPC) and HPL 24 (with high SPC) (Table 5.5). The ICP 11605 allele would be expected to match with the reference assembly allele of Asha (ICPL 87119) since ICPL 87119 is a low SPC genotype itself while the HPL 24 allele should match to the alternative allele (Table 5.5). Accordingly, not all PCR-generated nsSNP-specific alleles for the test genotypes were consistent with those from the WGRS data and the reference genome sequence (Table 5.5). By comparing ICP 11605 (low SPC) and ICP 5529 (high SPC) alleles with the reference genome and the WGRS-derived alternative alleles, respectively, presence of a total of 52 (75.36%) of the SNPs could be confirmed. However, a SNP locus at position 17486133 on CcLG01 had a different alternative SNP allele i.e. A to C instead of A to T (Table 5.5).

Table 5.5: Verification of 85 SNPs using PCR and Sanger sequencing

Chr.	Position	Ref_allele	Alt_allele	Gene_id	ICP11605	HPL24
CcLG01	4414830	A	C	*C.cajan_19617*	**A**	**C**
CcLG01	4415753	T	C	*C.cajan_19617*	**T**	**C**
CcLG01	4992638	T	C	*C.cajan_19670*	**T**	**C**
CcLG01	16162501	A	C	*C.cajan_20717*	**A**	**C**
CcLG01	16873384	C	T	*C.cajan_20776*	**C**	**T**
CcLG01	16873606	T	A	*C.cajan_20776*	**T**	**A**
CcLG01	16873826	A	G	*C.cajan_20776*	**A**	**G**
CcLG01	17475532	A	T	*C.cajan_20841*	**A**	**T**
CcLG01	17482682	C	T	*C.cajan_20842*	**C**	**T**
CcLG01	205708	A	G	*C.cajan_19199*	**A**	**G**
CcLG02	1204754	G	C	*C.cajan_04622*	**G**	**C**

Chr., chromosome; Confirmed SNPs in bold font; † Locus has a different SNP allele in HPL 24 as expected from the whole genome resequencing data.

Table 5.5: (continued)

Chr	Position	Ref_allele	Alt_allele	Gene_id	ICP11605	HPL24
CcLG02	3074263	G	GCTC	*C.cajan_04797*	**G**	**GCTC**
CcLG02	8895098	C	G	*C.cajan_05310*	**C**	**G**
CcLG02	24499331	G	A	*C.cajan_06764*	**G**	**A**
CcLG02	36162648	C	T	*C.cajan_07942*	**C**	**T**
CcLG02	16726731	T	G	*C.cajan_06086*	**T**	**G**
CcLG02	16727058	A	T	*C.cajan_06086*	**A**	**T**
CcLG02	16727494	T	C	*C.cajan_06086*	**T**	**C**
CcLG03	460462	C	G	*C.cajan_08095*	**C**	**G**
CcLG03	20453445	C	T	*C.cajan_10047*	**C**	**T**
CcLG03	20477859	G	C	*C.cajan_10048*	**G**	**C**
CcLG04	428893	A	G	*C.cajan_20905*	**A**	**G**
CcLG04	496463	C	T	*C.cajan_20914*	**C**	**T**
CcLG04	10982012	G	A	*C.cajan_21987*	**G**	**A**
CcLG06	14406147	A	T	*C.cajan_12368*	**A**	**T**
CcLG07	12293465	C	A	*C.cajan_18443*	**C**	**A**
CcLG07	12293471	T	C	*C.cajan_18443*	**T**	**C**
CcLG07	12293474	T	C	*C.cajan_18443*	**T**	**C**
CcLG07	12293498	T	A	*C.cajan_18443*	**T**	**C**
CcLG07	12293515	G	A	*C.cajan_18443*	**G**	**A**
CcLG07	12293541	A	C	*C.cajan_18443*	**A**	**C**
CcLG07	12293560	C	G,T	*C.cajan_18443*	**C**	**T**
CcLG07	12301081	C	T	*C.cajan_18443*	**C**	**T**
CcLG07	12301614	T	G	*C.cajan_18443*	**T**	**G**
CcLG08	900949	T	G	*C.cajan_15445*	**T**	**G**
CcLG08	5286893	A	T	*C.cajan_15863*	**A**	**T**
CcLG08	5287380	C	T	*C.cajan_15863*	**C**	**T**
CcLG08	14254343	A	G	*C.cajan_16632*	**A**	**G**
CcLG08	15677727	A	G	*C.cajan_16775*	**A**	**G**

Table 5.5: (continued)

Chr	Position	Ref_allele	Alt_allele	Gene_id	ICP11605	HPL24
CcLG08	8702222	A	G	*C.cajan_16102*	**A**	**G**
CcLG09	3758497	G	T	*C.cajan_22447*	**G**	**T**
CcLG10	7560364	T	G	*C.cajan_14054*	**T**	**G**
CcLG11	25475840	G	C	*C.cajan_02320*	**G**	**C**
Scaf000379	105750	A	T	*C.cajan_36235*	**A**	**T**
Scaf117710	1241	C	A	*C.cajan_47566*	**C**	**A**
Scaf132767	14801	T	C	*C.cajan_44248*	**T**	**C**
Scaf136850	11420	G	A	*C.cajan_46295*	**G**	**A**
CcLG01	17486133†	A	T	*C.cajan_20843*	**A**	**C**
CcLG07	9512986	C	T	*C.cajan_18213*	**C**	**T**
CcLG07	12299616	T	A	*C.cajan_18443*	**T**	**A**
CcLG11	7967151	A	G	*C.cajan_00758*	**A**	**G**
CcLG11	8413398	C	T	*C.cajan_00788*	**C**	**T**
CcLG01	16874033	C	T	*C.cajan_20776*	C	C
CcLG01	17486917	A	T	*C.cajan_20843*	A	A
CcLG01	16876698	G	C	*C.cajan_20777*	G	G
CcLG01	16163064	C	G	*C.cajan_20717*	C	C
CcLG01	16855461	T	C	*C.cajan_20775*	T	T
CcLG01	16855539	A	G	*C.cajan_20775*	A	A
CcLG01	16873818	T	C	*C.cajan_20776*	T	T
CcLG01	17476096	T	C	*C.cajan_20841*	T	T
CcLG01	17479259	C	T	*C.cajan_20841*	C	C
CcLG01	16876707	A	G	*C.cajan_20777*	A	A
CcLG01	16876742	G	A	*C.cajan_20777*	G	G
CcLG01	16876689	C	T	*C.cajan_20777*	no read	no read
CcLG02	1201138	C	T	*C.cajan_04622*	no read	no read
CcLG02	12525321	A	G	*C.cajan_05664*	A	A
CcLG02	16750258	T	A	*C.cajan_06087*	T	T

Table 5.5: (continued)

Chr	Position	Ref_allele	Alt_allele	Gene_id	ICP11605	HPL24
CcLG02	16756557	T	G	*C.cajan_06087*	T	no read
CcLG02	16756576	A	T	*C.cajan_06087*	A	no read
CcLG02	16726667	A	C	*C.cajan_06086*	A	A
CcLG02	16736680	A	C	*C.cajan_06086*	A	A
CcLG02	11940360	T	G	*C.cajan_05609*	no read	no read
CcLG06	6651663	A	C	*C.cajan_11650*	A	A
CcLG07	12293486	C	T	*C.cajan_18443*	C	C
CcLG08	901014	T	G	*C.cajan_15445*	T	T
CcLG08	902979	C	A	*C.cajan_15445*	C	no read
CcLG08	1729778	T	C	*C.cajan_15499*	no read	A
CcLG08	4923892	A	G	*C.cajan_15828*	A	A
CcLG10	7560516	A	C	*C.cajan_14054*	A	A
CcLG10	7560344	T	C	*C.cajan_14054*	T	T
CcLG10	7560536	T	C	*C.cajan_14054*	T	T
Scaf000059	551244	C	A	*C.cajan_28359*	A	A
Scaf000213	6068	T	G	*C.cajan_35646*	G	G
Scaf135136	45544	T	A	*C.cajan_35089*	T	no read
Scaf135136	45624	A	C	*C.cajan_35089*	A	no read
Scaf135136	45693	G	T	*C.cajan_35089*	G	no read

5.3.5 Conversion of nsSNPs to CAPS) and dCAPS markers

To verify the association of selected SNPs with SPC and further confirm their presence, a sample of 59 SNPs from 40 genes were converted into CAPS/dCAPS markers (Table 5.6).

Table 5.6: Cleaved amplified polymorphic sequence (CAPS) and derived-CAPS primers developed from nsSNPs from candidate genes for seed protein content

Chr.	Gene ID	Primer ID	Marker type	Forward sequence (5'-3')	Reverse sequence (3'-5')	Enzyme name
1	*C.cajan_19617*	spc047	dCAPS	TCCCCATCCTCTCTCAAAACCCCG	TCTCTCACCACTTCCCAAGG	AciI
1	*C.cajan_19617*	spc048	CAPS	AAGGGGATACTGAGAAAAGTTGT	AGCTGCTGTTGATCCACCTT	Hpy188I
1	*C.cajan_19670*	spc049	dCAPS	TTGCACACTCCTACATAGCTCTTT	TAGGAGGGGGCTTTCAAGAT	MseI
1	*C.cajan_20717*	spc050	dCAPS	AGTTCTGAAGACATTCTCTTTGAT	GAAAAGCTTATTTTATTTTTCCTTCAA	MboI
1	*C.cajan_20776*	spc054	dCAPS	AATTGATGAAGATGATTCAGAAGC	TAATGTTTCGGTCCGTATGC	AluI
1	*C.cajan_20776*	spc055	dCAPS	GGCATACGGACCGAAACAT	CATGGGTTCTTTAACATTGGTTTA	MseI
1	*C.cajan_20776*	spc057	dCAPS	CAGGTATGCTGATATTTGAGTTTTG	GCAATCCACCAGAAAAGCATATGA	MboI
1	*C.cajan_20841*	spc059	dCAPS	CCCTCAAATGAAAGTTGATCC	GGATATGAAGGAATGTGGCTATAT	MseI
1	*C.cajan_20842*	spc062	dCAPS	ACTGGGATGAGTTAGTGAAAGC	AGAACTGCACCCAAACGAAATCAT	NlaIII
1	*C.cajan_19199*	spc089	dCAPS	GGCAGCTGTTATGCACAAAT	AAATGCATTAAATGAGAAGACCTT	PsiI
2	*C.cajan_04622*	spc003	CAPS	AATAACTTTGCTGCGCCATC	TAGAAGCCCTTGGGTCAATG	NlaIII
2	*C.cajan_04797*	spc004	dCAPS	CAAAGACCATGATATTAAATACAT	CAGGCACTAGCAGTGAGCAG	NlaIII
2	*C.cajan_05310*	spc005	CAPS	CGCACCACTCTCTCAGATCC	ACCCAAAACAGAGCCCAAAG	AciI
2	*C.cajan_06764*	spc010	dCAPS	TAGGTGACAACGTTGGACTCATAT	TGGAACATCTTCGCAAGAAA	NlaIII
2	*C.cajan_07942*	spc011	dCAPS	TGTTGCTAACGGTAAACATCTCCG	GTTCGGCTAAGTGGTCAGGA	AciI
2	*C.cajan_06086*	spc100	CAPS	GAATTCATGCCGTTGGTGT	TCATTTGCTGTAGAGCATCCA	NlaIII
2	*C.cajan_06086*	spc101	dCAPS	TTTTCCTTTGCTTCTATTGTTTCA	ACTTGGAACTGAAAGAATCATTTA	MseI
2	*C.cajan_06086*	spc102	dCAPS	CGCAAGCCATCTGAAGAGAC	CATTTTCCTTGGCACTTATCATG	BspHI
3	*C.cajan_08095*	spc012	dCAPS	TAAGATAGACAAATCAGTTTGTTC	CCATCTTCTCCACAGCTTGC	AciI

Restriction enzyme cut site in color. S79, Scaffold000379; S10, Scaffold117710; S67, Scaffold132767; S36, Scaffold135136; S50, Scaffold136850.

Table 5.6: (continued)

Chr	Gene ID	Primer ID	Marker type	Forward sequence (5'-3')	Reverse sequence (3'-5')	Enzyme name
3	*C.cajan_10047*	spc013	dCAPS	TATGAACCATTTGCTTCATGAACC	TGACTTAGAACTTTGCAACTAACAA	AciI
3	*C.cajan_10048*	spc014	dCAPS	AGCGTCATCCTAAGGTTACCA	TTCTCAGAAACATTAGTGGAATT	ApoI
4	*C.cajan_20905*	spc065	CAPS	CTGGCATTGCTGTCTAACCTC	ACTCTCCGATTCCCTCTGGT	NlaIII
4	*C.cajan_20914*	spc066	dCAPS	GGCCTCATTGAAAGCATCTC	GTTCCCTAAGAGAAGCACCTACAT	NlaIII
4	*C.cajan_21987*	spc106	dCAPS	CATATCACCCTAAAATCGATGG	GTTGTATGGTAATAAGTGTTGTGA	Tsp45I
6	*C.cajan_12368*	spc016	dCAPS	AAGAACTACCACATGGGAGGTCTG	TCAAGGTGCTTTTTATCTCTCAAA	Hpy188I
7	*C.cajan_18213*	spc029	dCAPS	AACCAATTACTCGGGTGTATGT	CTCTCCAAATCTTGCCGCCTAGAA	BglI
7	*C.cajan_18443*	spc033	CAPS	GCACGGACAAGACCATTTTA	TGCAATTCCCATGTTTGTTG	Hpy188I
7	*C.cajan_18443*	spc034	dCAPS	TTTCAAGGGAATATCATCAGAAT	CACTTTGCAATTCCCATGTTT	ApoI
7	*C.cajan_18443*	spc035	dCAPS	TCAAGGGAATATCATCAGACTTCA	CAGGGAATGCCACTTTGC	NlaIII
7	*C.cajan_18443*	spc039	dCAPS	CCAATTATCGATGGTGTATTTATC	TTGGACGACTTGAATACTTTTCC	NlaIII
7	*C.cajan_18443*	spc040	dCAPS	CCTCCTTCAGAGAAAGTTTTCAG	AGATCATATTTGGACGACTTGAA	Hpy188I
7	*C.cajan_18443*	spc041	dCAPS	TTCATATTGGGGCATTCAGCTACG	AAGGAATTGTGTATCAATAAAAGAAAT	NlaIV
7	*C.cajan_18443*	spc042	dCAPS	CAACTCAGAAAGACTAGCATTTCC	GGTGAGTTAGACAAGCTGCAAG	MnlI
7	*C.cajan_18443*	spc045	dCAPS	AATCTCATTCCTATAACATCTCCG	AGTTTATGGGGTTGGTGGTG	AciI
7	*C.cajan_18443*	spc046	dCAPS	TCTTTGAATGGCGTTCATCA	TTAAGATGATGGAAAATGAAAGAT	MboI
8	*C.cajan_15445*	spc017	dCAPS	GGAAAGGATAGGTGATAGATTCACT	GAATGTTGTGCTAGCAGGTTTAAA	PmeI
8	*C.cajan_15863*	spc023	dCAPS	GAAGTTCCCCTATCGCATCA	TATCAACTACTGATAACAACAAGA	MboI
8	*C.cajan_15863*	spc024	dCAPS	AGAAAGGGTCGTCTTGTGCCTCCG	GAATGTTTTATTTGTTGCAAACC	AciI
8	*C.cajan_16632*	spc025	dCAPS	TGGAATATATCAGAGGGATTGTGA	TTAGCAGAAAACCAGCTTGAAGGA	MboI
8	*C.cajan_16775*	spc026	dCAPS	GGTTTAGGGAATGTTAGCAAGC	TAAGGGTTAGTTGCCAAGCAACCG	AciI

Table 5.6: (continued)

Chr.	Gene ID	Primer ID	Marker type	Forward sequence (5'-3')	Reverse sequence (3'-5')	Enzyme name
8	*C.cajan_16102*	spc082	dCAPS	GACGAAAAATCTTGTCCATGAGGA	GGTACAACCAGGTACCCATGT	AluI
9	*C.cajan_22447*	spc068	dCAPS	TCCATTAGCACTTAGATTTAGCTCAT	ATCATTTTAGTGGTGGTATCCCCG	AciI
10	*C.cajan_14054*	spc079	dCAPS	ATCATTGAGCACTTTTCAGGATTA	GATCTGCAGGAATTATTCGACA	AluI
11	*C.cajan_02320*	spc001	dCAPS	AACAAGAATTAGGTGTTGTAATATT	TGACATAGTTGATCAGAACGGAAT	Hpy188I
11	*C.cajan_00758*	spc083	dCAPS	CACCAAATTTGACGGGATG	GGGCGGACATACTCGGATCGTTCA	NlaIII
11	*C.cajan_00788*	spc084	dCAPS	GGCAACTCTTCTAGAATTCTTTTAATC	CAAGGAATAAAAGTGTTTGCTGCC	AciI
S79	*C.cajan_36235*	spc074	dCAPS	GAGTGGGACAAATGGTTCGT	CTTTTGATATTTTTTCCAAATTTA	MseI
S10	*C.cajan_47566*	spc076	dCAPS	CCTTGGGGACAATTCACTCA	CAGGAAGAAAATGGTAGTTTTCCT	MseI
S67	*C.cajan_44248*	spc075	dCAPS	TGTGCCTTGAGGCCATCAATCCAG	GCCAAAGGTACTATGTAATTTGCTG	AluI
S50	*C.cajan_46295*	spc108	dCAPS	GGGCTTCTATGTCTAGTAGCCTCC	TCCCTCCTTCTTTATGCATTC	AciI
SNPs not confirmed by Sanger sequencing						
1	*C.cajan_20775*	spc053	dCAPS	TCATCTGTTACCACCGTTTCTG	CAAACCGGAGAAACTGCCACAAGA	MboI
1	*C.cajan_20717*	spc051	dCAPS	TTCTAAAGCTTGTGTTGATGAGC	TGTGCATTCCAAGCCCCTCAACAT	NlaIII
1	*C.cajan_20776*	spc056	dCAPS	ATTTTCTTCTTTCTCCATCTTAAA	CCCATTACTTGAACGTTTTGC	ApoI
S36	*C.cajan_35089*	spc071	dCAPS	ACACAGCAAGGACCATCACA	TACTTTCTTTGCAAGATGTTCTTT	MseI
1	*C.cajan_20777*	spc091	dCAPS	AAGTTAAAACTAACAAGGGAAT	CATCCTTATAAATGACAAACATATGAC	Hpy188I
1	*C.cajan_20775*	spc052	dCAPS	TTAAGCCAAATTCATTTGGTTCCA	CACGCGACACAAACCTCA	NlaIII
SNPs with poor quality Sanger sequencing results						
2	*C.cajan_04622*	spc002	CAPS	ATGCGGTAGCCAATTTCTTG	GCGTTGAAGATGAAGAGGACA	NlaIII
2	*C.cajan_05609*	spc107	CAPS	ACAGGTGTGCATTCGGTGT	CATGATATAAAGGAAAAGGTGGA	MseI
1	*C.cajan_20777*	spc090	dCAPS	TTTCCTTGGTATTGAGGTTGTTT	ACCTTTTTGTCCTCACTTTCCCAT	NlaIII

The 59 SNPs included fifty Sanger confirmed SNPs between low and high SPC, six loci for which no SNP could be confirmed via Sanger, and three putative SNPs for which all genotypes had a missing Sanger sequence read (Table 5.6; Table 5.7).

5.3.6 Polymorphism of cleaved amplified polymorphic sequences (CAPS) and derived-CAPS markers among six diverse pigeonpea genotypes

Of the 59 CAPS/dCAPS markers, 30 were successfully amplified and digested, four were amplified but not digested, 19 had no amplification, three were not amplified, and four showed poor amplification (Table 5.7). Of the successfully amplified, one was poorly digested making it difficult to score for polymorphism. The highest number of polymorphic markers was found between the high/low parental pairs such as HPL 24/ICP 11605 with 17 markers, HPL 24/ICPL87119 (16) and ICP 5529/1CP 11605 (16) (Table 5.7). The lowest number of polymorphic markers was between high/high such as in HPL 24/ICP 5529 (01), moderate/moderate e.g. in ICP 8863/ICP 14209 (03) and low/low e.g. in ICP 11605/ICPL 87119 (03) (Table 5.7).

Among the six CAPS/dCAPS markers derived from WGRS SNPs but with the SNPs absent in the Sanger sequences, only two amplified and none was polymorphic in any of the tested parental pairs (Table 5.7). This is a further confirmation of the absence of the SNPs between low and high SPC genotypes. In the case of the three putative SNP-containing fragments with poor/no Sanger sequencing reads, two showed polymorphism in eight of the parental pairs involving low/high SPC (Table 5.7).

Table 5.7: Polymorphism of cleaved amplified polymorphic sequences (CAPS) and derived-CAPS markers among six diverse pigeonpea genotypes

			High/Low				High/Moderate				Moderate/Low				H/H‡	M/M§	L/L¶
Chr.	Primer ID	Amp status	1	2	3†	4	5	6	7	8	9	10	11	12	13	14	15
SNPs confirmed by Sanger sequencing																	
CcLG03	**spc012**	A	P	P	P	P	P	M	P	M	M	P	M	P	M	P	M
CcLG03	**spc013**	A	P	P	M	M	P	P	M	M	M	M	M	M	P	M	M
CcLG02	**spc100**	A	P	P	P	P	P	P	P	M	M	M	M	M	M	M	M
CcLG02	spc005	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG07	spc035	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG01	spc062	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG02	spc102	AND	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG03	spc014	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sc.000379	spc074	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG01	spc089	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG02	spc101	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Markers in bold are polymorphic in one or more parental pairs. , respectively. Amp status; amplification status; A, amplified; AND, amplified but not digested; APD, amplified but poorly digested; NYA, not yet amplified. 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14 and 15 refer to parental pairs HPL 24/ICP 11605, HPL 24/ICPL 87119, ICP 5529/ICP 11605, ICP 5529/ICPL 87119, HPL 24/ICP 8863, HPL 24/ICP 14209, ICP 5529/ICP 8863, ICP 5529/ICP 14209, ICP 11605/ICP 8863, ICP 11605/ICP 14209, ICPL 87119/ICP 8863, ICPL 87119/ICP 14209, HPL 24/ICP5529, ICP8863/ICP 14209 and ICP 11605/ICPL 87119, respectively.
† Population used for genetic mapping and marker-trait co-segregation analysis. ‡ High/High, § Moderate/Moderate and ¶ Low/Low parental pairs.

Table 5.7: (continued)

			High/Low				High/Moderate				Moderate/Low				H/H	M/M	L/L
Chr.	Primer ID	Amp status	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CcLG02	spc003	A	P	P	P	P	P	P	P	P	M	M	M	M	M	M	M
CcLG08	spc017	A	P	P	P	P	M	M	M	M	P	P	P	P	M	M	M
CcLG08	spc025	A	P	P	P	P	M	M	M	M	P	P	P	P	M	M	M
CcLG07	spc033	A	P	P	P	P	P	P	M	M	P	P	P	P	M	M	P
CcLG01	spc048	A	P	P	P	P	M	M	M	M	P	P	P	P	M	M	M
CcLG01	spc049	A	P	P	P	P	P	P	P	P	M	M	M	M	M	M	M
CcLG01	spc055	A	P	P	P	P	M	P	M	P	P	M	P	M	M	P	M
CcLG01	spc059	A	P	P	P	P	P	P	P	P	P	P	P	P	M	P	P
CcLG04	spc065	A	P	P	P	P	P	P	P	P	M	M	M	M	M	M	M
CcLG04	spc066	A	P	P	P	P	P	P	P	P	M	M	M	M	M	M	M
CcLG09	spc068	A	P	P	P	P	P	P	P	P	M	M	M	M	M	M	M
CcLG02	spc004	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG02	spc010	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG08	spc024	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG08	spc026	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG01	spc057	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
Sc.117710	spc076	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M

Table 5.7: (continued)

			High/Low				High/Moderate				Moderate/Low				H/H	M/M	L/L
Chr.	Primer ID	Amp status	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CcLG07	spc034	A		-	-	-	-		-	-	-	-		-	-	-	-	-
CcLG07	spc041	AND		-	-	-	-		-	-	-	-		-	-	-	-	-
Sc.136850	spc108	AND	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG02	spc011	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG08	spc023	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG07	spc039	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG07	spc045	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG01	spc047	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG01	spc050	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG01	spc054	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG08	spc082	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG04	spc106	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG07	spc046	NYA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG11	spc001	PA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG06	spc016	PA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Sc.132767	spc075	PA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CcLG10	spc079	PA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Table 5.7: (continued)

			High/Low				High/Moderate				Moderate/Low				H/H	M/M	L/L
Chr.	Primer ID	Amp status	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CcLG11	**spc084**	A	P	M	P	M	M	M	M	M	M	M	P	P	M	M	P
CcLG07	spc042	AND	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG07	spc029	APD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG07	spc040	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG11	spc083	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SNPs not confirmed by sanger sequencing																	
CcLG01	spc053	A	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
CcLG01	spc051	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG01	spc056	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sc.135136	spc071	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG01	spc091	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CcLG01	spc052	NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SNPs with poor quality Sanger sequencing results																	
CcLG02	**spc002**	A	P	P	P	P	P	P	P	P	M	M	M	M	M	M	M
CcLG02	**spc107**	A	P	P	P	P	P	P	P	P	M	M	M	M	M	M	M
CcLG01	spc090	NYA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total polymorphic markers			**17**	**16**	**16**	**15**	**13**	**12**	**10**	**9**	**6**	**6**	**7**	**7**	**1**	**3**	**3**

5.3.7 Genetic mapping of candidate gene-based CAPS and dCAPS markers

Sixteen polymorphic CAPS/dCAPS markers in parental pair ICP 5529 and ICP 11605 (Table 5.7) were combined with GBS-derived SNPs data in the population to construct an F_2 genetic map as described in sections Chapter 5. Eleven of the 16 markers could be mapped on to the genetic map with two markers each on CcLG01 and CcLG04, four (CcLG02), one (CcLG07), and two (CcLG08) (Fig. 5.2).

QTL analysis using both CIM and ICIM did not find any of the mapped markers within the vicinity of any QTL identified in this population (Fig. 5.2). However, there were three markers within <10 cM distance from main effect QTLs. Two of the markers, spc002 and spc107, derived from derived from mutations in the NADH-GOGAT (*C.cajan_04622*) and a copper transporter gene (*C.cajan_05609*) on CcLG02 were found 2.7 and 7.6 cM distances away, respectively from a QTL explaining 9.0% of the phenotypic variation in SPC. Another marker (spc100) derived from a mutation in a BLISTER gene (*C.cajan_06086*) on the same CcLG02 was 7.8 cM away from a major QTL explaining 11.5% of the phenotypic variation for SPC.

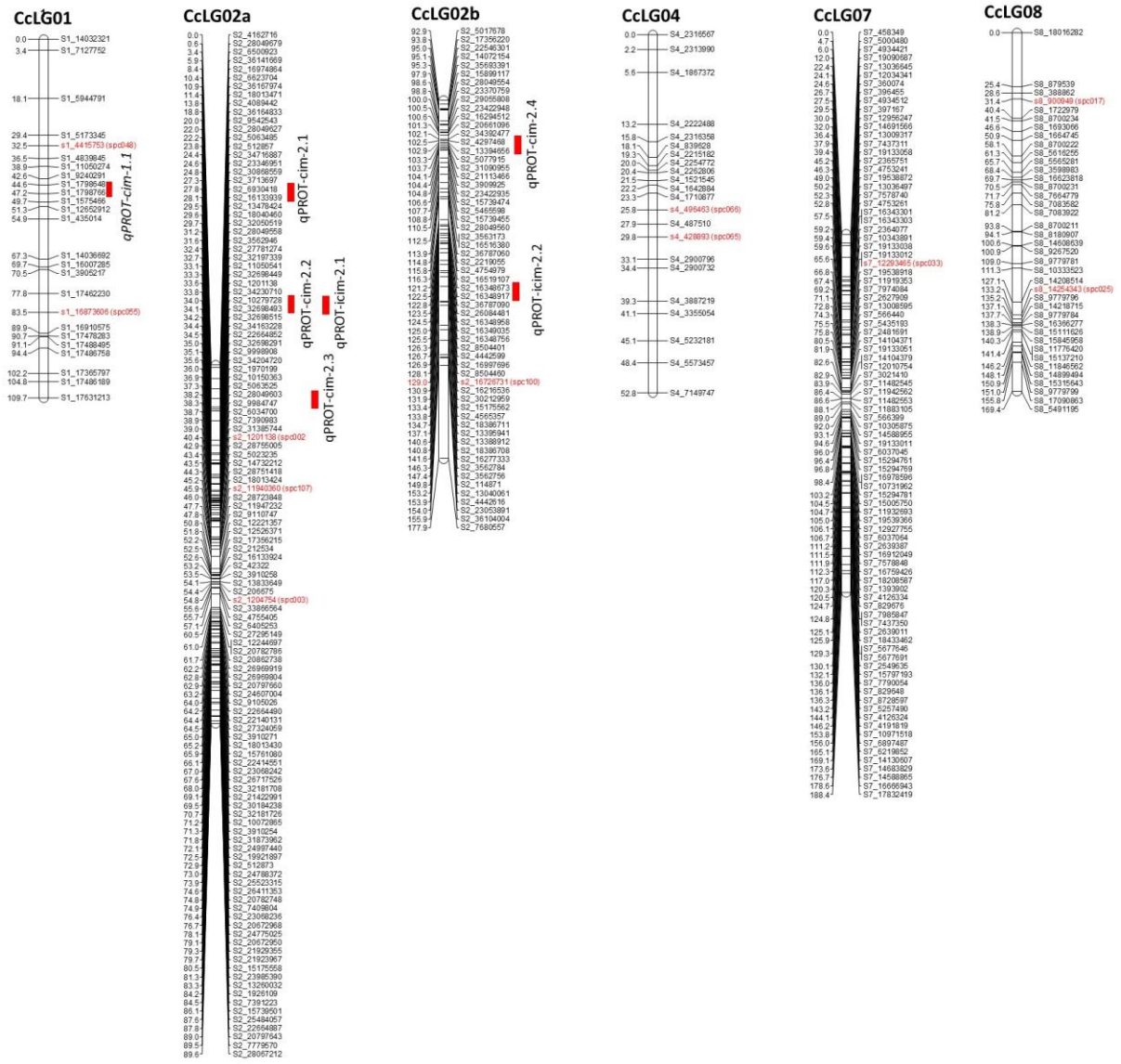


Fig. 5.2: Genetic map from 157 F_2 individuals of the pigeonpea mapping population ICP 5529 × ICP 11605 developed from 787 markers. Only chromosomes containing at least one mapped CAPS/dCAPS markers are presented. Markers in red are candidate CAPS/dCAPS markers for seed protein content. Red bars are QTLs for seed protein content.

5.3.8 Single marker analysis

Single marker analysis (SMA) using regression of F_2 genotype and phenotype data found four of the 11 assayed CAPS/dCAPs to have significant association with SPC (Table 5.8). Three of the markers were on CcLG02 and included spc003 (R^2 = 3.5%, P = 0.011) derived from a mutation in the NADH-GOGAT gene, spc107 (R^2 = 3.7, P = 0.008) derived from a mutation in a copper transport gene, and spc100 (R^2 = 2.8%, P = 0.023) derived from a BLISTER gene. The fourth marker, spc017 (R^2 = 2.2%, P = 0.043) was derived from a probable LRR receptor-like serine/threonine-protein kinas*e* gene on CcLG08.

Table 5.8: Cleaved amplified polymorphic sequences (CAPS)/derived-CAPS markers with significant association with seed protein content in an F_2 mapping population of the cross ICP 5529 × ICP 11605

Chr.	Gene ID	Marker (Type)	Enzyme	R^2 (%)	F-prob	Gene name
CcLG02	*C.cajan_04622*	spc003 (CAPS)	NIaIII	3.5	0.011	NADH-GOGAT
CcLG02	*C.cajan_05609*	spc107 (CAPS)	MseI	3.7	0.008	Copper transporter
CcLG08	*C.cajan_15445*	spc017 (dCAPS)	PmeI	2.2	0.043	Protein kinase
CcLG02	*C.cajan_06086*	spc100 (CAPS)	NIaIII	2.8	0.023	BLISTER
			Total	12.3		

5.4 Discussion

Modern NGS based genomics approaches are found to be efficient for rapid identification of potential candidate genes controlling complex traits in pigeonpea (Singh et al., 2016; Saxena et al., 2017; Varshney et al., 2017). The results obtained from previous NGS-based trait mapping studies have encouraged the use of similar approach for identification of candidate genes/markers associated with SPC in pigeonpea. In the present study, NGS-based sequencing data was integrated with candidate gene information and Sanger sequencing, marker development, genetic mapping and trait association analysis was performed to scale-down the natural allelic variants of candidate genes regulating SPC in pigeopea. The observed phenotypic variation of SPC among pigeonpea genotypes reflect the complex quantitative nature. Considering the efficacy for quantitative dissection of complex traits in diverse crop plants including pigeonpea (Silva et al., 2012; Lim et al., 2014; Xu et al., 2014; Singh et al.,

2016; Saxena et al., 2017; Varshney et al., 2017), the sequencing-based mapping strategy can be deployed for understanding the complex genetic architecture of SPC trait.

5.4.1 Functional roles of selected genes

The initial candidate variant/gene prioritization in this study was on the basis of predicted impact of the variants on protein function. This led to selection of nonsynonymous, stop, frame-shift, splice-site and indel mutations. A final selection of the candidates was based on information on gene function in relation to the SPC, eventually resulting in 108 variants which included nsSNPs, stop gains, frame-shifts, an insertion and splice-site mutations) in 57 genes considered for further analysis.

The 57 genes belong to 19 GO-molecular function categories. A number of the genes or their homologues have been implicated in the control of SPC. Such genes include sucrose synthase gene (Zeng et al., 2016) on CcLG01 at position 4415753 bp, glutamate synthase gene (NADH-GOGAT) (Shoenbeck et al., 2000; Nigro et al., 2013) on CcLG02 at position 1204754 bp, basic 7S globulin gene on CcLG02 at position 8895098 bp (Yamada et al., 2014), 2-oxoglutarate dehydrogenase (Araújo et al., 2013) on CcLG02 at position 36162648 bp, ABC transporter gene (Upadhyaya et al., 2016) on CcLG03 at position 20453445 and 20477859 bp, and asparagine synthetase gene (Lam et al., 2003; Pandurangan et al., 2012) at position.

Several of the putative candidate genes detected in the present study, although with no known proof that they increase or decrease SPC accumulation, have been reported to play a role in storage protein biosynthesis through various metabolic pathways. For example, genes of the proteolytic pathway such as the aspartic-type endopeptidase (proteases) (EC 3.4.23.-) and RNA-directed DNA polymerase (Reverse transcriptase; EC 2.7.7.49) genes are reported to play a role in proteolysis and processing of seed storage proteins (Pereira et al., 2008). Similarly, a number of transcription factors such as Heat shock proteins e.g Hsp 40 (Ohta et al., 2013, Bolon et al., 2010), Protein ETHYLENE INSENSITIVE-3 (EIN3) (Cohen et al., 2014), GTP-binding subunit (Lestari et al., 2013), WRKY transcription factor and Myb related proteins have been implicated as broad-range regulators of gene expression (Rahaie et al., 2013).

That a considerable number of genes identified from the pigeonpea WGRS data had been previously reported in literature to play roles in SPC in several crops underscores the probable role of these genes in conditioning SPC in pigeonpea. It also indicates a correct selection and grouping of the genotypes used for the detection of the candidate variants and genes in the present study.

5.4.2 SNP validation

To ensure certainty in the existence of the variants detected in the genes, a validation through Sanger sequencing to eliminate any false positive variant calls was done. Up to 75.4% of tested SNPs were found to be correct between one low (ICP 11605) and one high (ICP 5529) SPC genotypes. Both ICP 11605 and ICP 5529 were originally used for SNP prediction from the WGRS data (see Materials and Methods). In comparing results of the present study with that of earlier similar studies, the SNP prediction rate from the Illumina WGRS data as verified by Sanger sequencing is lower than 83% in *Eucalyptus grandis* (Novaes et al., 2008), 96.4% in *Oryza sativa* (Deschamps et al., 2010), 79 – 97% *Glycine max* (Hyten et al., 2010a; Deschamps et al., 2010), 86% *Phaseolus vulgaris* (Hyten et al., 2010b), and >80% in diploid wheat (*Aegilops tauchi*) (You et al., 2011). It is, however, higher than the 35.3% in *Cicer arietinum* (Azam et al., 2012).

Factors that may contribute to the low SNP prediction accuracy in the present study include narrow genetic base of pigeonpea (Sharma et al., 2013), genome assembly and errors associated with sequence alignment, genotype and variant calling (Olson et al., 2015) and use of small datasets (Azam et al., 2012). In addition, the read depths of 9.68 to 14.03 of the WGRS datasets (Kumar et al., 2016) used for the identification of putative variants may be considered to be relatively low and may also have contributed to the realised SNP prediction accuracy.

Nonetheless, with an accuracy of 75.4%, 81 out 108 final selected nsSNPs or 24,855 out of 32,964 total variants could be expected to be valid and may be useful in genetic studies and breeding applications aimed at improving SPC in pigeonpea. To test this hypothesis and further verify the presence of the SNPs, a set of 59 nsSNP loci comprising 50 Sanger confirmed, six unconfirmed nsSNPs and three doubtful (Sanger sequencing with poor/no reads) were

converted into CAPS/dCAPS markers and assayed for polymorphism on six (two low, two high and two moderate SPC) genotypes. The highest number of polymorphic markers observed in the high by low than in the high by moderate or high by high SPC parental pairs provided confirmation of the potential usefulness of the genic SNP-derived CAPS/dCAPS markers. It also further validated the presence of the SNPs particularly for two of the three markers for which Sanger sequencing results were poor.

With an objective to test for co-segregation of the markers with SPC, 16 polymorphic CAPS/dCAPS markers between parents ICP 5529 and ICP 11605 were assayed on an F_2 mapping population of the two parents. By comparing genetic map positions of 11 markers with that of SPC QTLs all the three markers that are in close proximity (<10 cM) to any main effect SPC QTLs were found on CcLG02. The three markers (spc003, spc107 and spc100) were derived from mutations in NADH-GOGAT, copper transporter and BLISTER genes, respectively. Similarly, three (spc002, spc107 and spc100) of the four markers that showed significant association with SPC through SMA are on the same CcLG02 and derive from mutations in the same three genes (NADH-GOGAT, copper transporter and BLISTER). Of the three genes, a higher expression of NADH-GOGAT in two durum wheats has been associated with higher grain protein content (Nigro et al., 2013). While the BLISTER gene is reported to localise within a major SPC QTL on chromosome 20 (Lestari et al., 2013). However, in the case of the copper transporter gene, no report exists that indicates its functional or positional relationship to SPC in any plant, and may therefore be considered novel.

5.5 Limitations to the study

While whole genome resequencing have allowed the identification and prioritization of candidate variants that could underlie or contribute to the SPC variability observed in the set pigeonpea genotypes studied, the approach does have limitations. It is possible that other causative variants/mutations have been overlooked in the panel of selected genes as a result of the (i) strategies used to prioritize the candidate SNPs and (ii) small differences between the high and low SPC genotypes used. Causative variants in the non-coding regions of the targeted putative SPC genes, which are not included in the exon, or indeed in other genes that are not in the panel of putative SPC candidate genes, would also be overlooked by this approach. The potential for other genetic mechanisms, such as copy number variation, large indels, or structural genomic variants, to contribute to the underlying mutations also cannot be discounted. Gene-gene interactions is another area that may add insight in gene functions related to SPC, especially that the QTL-QTL interactions already revealed strong influence of epistasis on SPC as noted in Chapter Five of this thesis. Although further studies are required to confirm the role of the candidate SNPs identified in this study, the approach used has highlighted plausible candidate genes for targeted analysis.

5.6 Conclusions

A total of 108 candidate variants putatively associated with variation in SPC were selected. The variants are associated with 57 genes spread over all the 11 chromosomes except CcLG05. Validation through Sanger sequencing confirmed presence of 72.5% of the identified variants. GO based function analysis and comparison with identified candidate genes reported in literature indicated that a number of these candidate genes play roles in SPC. The nsSNPs converted into CAPS/dCAPs markers are highly polymorphic between low and high, and less so between high and high, moderate and moderate, and low and low SPC genotypes pairs. A considerable number of the CAPS/dCAPS were integrated onto a genetic linkage map and their co-localisation with QTLs for SPC determined. Marker-phenotype co-segregation analysis in a segregating population revealed association of the CAPS/dCAPS markers with SPC. The method used in the study is an efficient approach for detecting candidate genes underlying complex traits, including SPC. Results from this study also provide a foundation for future basic research and marker-assisted breeding for improving SPC in piegeonpea.

References

Amelio, I., F. Cutruzzola´, A. Antonov, M. Agostini, and G. Melino. Serine and glycine metabolism in cancer. Trends in Biochemical Sciences 39:191-198.

Araújo, W.L., L. Trofimova, G. Mkrtchyan, D. Steinhauser, L. Krall, A. Graf, A.R. Fernie, and V.I. Bunik. 2013. On the role of the mitochondrial 2-oxoglutarate dehydrogenase complex in amino acid metabolism. Amino Acids 44:683-700.

Asakura, T., I. Matsumoto, J. Funaki, and S.A.K. Abe. 2001. The plant aspartic proteinase-specific polypeptide insert is not directly related to the activity of oryzasin. European Journal of Biochemistry 267:5115-5122.

Azam, S., V. Thakur, P. Ruperao, T. Shah, J Balaji, B. Amindala, A.D. Farmer, D.J. Studholme, G.D. May, D. Edwards, J.D. Jones, and R.K. Varshney. 2012. Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome. American Journal of Botany 99:186-192.

Bansal, V., O. Harismendy, R. Tewhey, S.S. Murray, N.J. Schork, E.J. Topol, and K.A. Frazer. 2010. Accurate detection and genotyping of SNPs utilizing population sequencing data. Genome Research 20:537-545.

Bittencourt, L.M., D.A.P.D. Lana, A.M. de C. Pimenta, A.V. dos Santos, A.P.F. Gonçalves, R. Augustic, and L.M. Costa. Determination of metal associated with proteins of wheat seed samples after sequential extraction procedure. Journal of the Brazilian Chemical Society 25:264-270.

Blanvillain, R., J.H. Kim, S. Wu, A. Lima, and D.W. Ow. 2009. OXIDATIVE STRESS 3 is a chromatin-associated factor involved in tolerance to heavy metals and oxidative stress. The Plant Journal 57:654-65.

Bolon, Y.T., B. Joseph, S.B. Cannon, M.A. Graham, B.W. Diers, A.D. Farmer, G.D. May, G.J. Muehlbauer, J.E. Specht, Z.J. Tu, N. Weeks, W.W. Xu, R.C. Shoemaker, and P.C. Vance 2010. Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Biology 10:41.

Burstin, J., P. Marget, M. Huart, A. Moessner, B. Mangin, C. Duchene, B. Desprez, N. Munier-Jolain, and G. Duc. 2007. Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiology 144:768-781.

Carrera, C.S., M.C. Carrão-Panizzi, J.M. Gontijo-Mandarino, R.S. Leite, V.V. Ergo, F. Andrade, R. Parola, R.H. Lascano, and C.R.C. Vega 2015. Relationship between soybean industrial-nutritional quality and the assimilate source under heat and water stress during seed filling. VII Brazilian Soy Congress and the Mercosoja 2015, 22-25 June 2015, Florianopolis, Brazil.

Chardon, F., S. Jasinski, M. Durandet, A. Lécureuil, F. Soulay, M. Bedu, P. Guerche, and C. Masclaux-Daubresse. 2014. QTL meta-analysis in Arabidopsis reveals an interaction between leaf senescence and resource allocation to seeds. Journal of Experimental Botany 65:3949-3962.

Chen, Y., A. Sun, M. Wang, Z. Zhu, and P.B.F. Ouwerkerk. 2014. Functions of the CCCH-type zinc finger protein OsGZF1in regulation of the seed storage protein GluB-1 from rice. Plant Molecular Biology 84:621–634.

Church D.M., V.A. Schneider, T. Graves, K. Auger, F. Cunningham, N. Bouk, H.-C. Chen, R. Agarwala, W.M. McLaren, G.R.S. Ritchie, D. Albracht, M. Kremitzki, S. Rock, H.

Kotkiewicz, C. Kremitzki, A. Wollam, L. Trani, L. Fulton, R. Fulton, L. Matthews, S. Whitehead, W. Chow, J. Torrance, M. Dunn, G. Harden, G. Threadgold, J. Wood, J. Collins, P. Heath, G. Griffiths, S. Pelan, D. Grafham, E.E. Eichler, G. Weinstock, E.R. Mardis, R.K. Wilson, K. Howe, P. Flicek, and T. Hubbard. 2011. Modernizing reference genome assemblies. PLoS Biology 9:e1001091.

Cingolani, P., A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, D.M. Ruden, and X. Lu. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w. Fly 6:1-13.

Cohen, H., H. Israeli, I. Matityahu, and R. Amir. 2014. Seed-specific expression of a feedback-insensitive form of cystathionine-g-synthase in Arabidopsis stimulates metabolic and transcriptomic responses associated with desiccation stress. Plant Physiology 166:1575-1592.

Coles, J.P., A.L. Phillips, S.J. Croker, R. Garcia-Lepe, M.J. Lewis, and P. Hedden. 1999. Modification of gibberellin production and plant development in Arabidopsis by sense and antisense expression of gibberellin 20-oxidase gene. The Plant Journal 17:547–556.

Collier, R., and M. Tegeder. 2012. Soybean ureide transporters play a critical role in nodule development, function and nitrogen export. The Plant Journal 72:355-367.

DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, A. McKenna, T.J. Fennell, A.M. Kernytsky, A.Y. Sivachenko, K.S. Cibulskis, B. Gabriel, D. Altshuler, and M.J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491-498.

Deschamps, S., M. la Rota, J.P. Ratashak, P. Biddle, D. Thureen, A. Farmer, S. Luck, M. Beatty, N. Nagasawa, L. Michael, V. Llaca, H. Sakai, G. May, J. Lightner, and M.A. Campbell. 2010. Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina Genome Analyzer. The Plant Genome 3:53-68.

Fabre, F., and C. Planchon. 2000. Nitrogen nutrition, yield and protein content in soybean. Plant Science 152:51-58.

Feng, X., X. Yu, and J. Tong. 2014. Novel single nucleotide polymorphisms of the insulin-like growth factor-i gene and their associations with growth traits in common carp (*Cyprinus carpio* L.). International Journal of Molecular Science 15: 22471-22482.

Gaur, V.S., U.S. Singh, A.K. Gupta, and A. Kumar. 2012. Influence of different nitrogen inputs on the members of ammonium transporter and glutamine synthetase genes in two rice genotypes having differential responsiveness to nitrogen. Molecular Biology Reporter 39:8035-8044.

Gilissen. C., A. Hoischen, H.G. Brunner, and J.A. Veltman. 2012. Disease gene identification strategies for exome sequencing. European Journal of Human Genetics 20:490-497.

Grattapaglia, D. 2008. Genomics of Eucalyptus, a global tree for energy, paper and wood. In: Moore, P., and R. Ming, editors, Genomics of tropical crop plants. Springer Sciences. pp 257-295.

Guan, M., I.S. Moller, and J.K. Schjoerring. 2015. Two cytosolic glutamine synthetase isoforms play specific roles for seed germination and seed yield structure in Arabidopsis. Journal of Experimental Botany 66:203-212.

Gubler, F., and J.V. Jacobsen. 1992. Gibberellin-responsive elements in the promoter of a barley high-pI α-amylase gene. The Plant Cell 4:1435-1441.

Gubler, F., D. Raventos, M. Keys, R. Watts, J. Mundy, and J.V. Jacobsen. 1999. Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. The Plant Journal 17:1-9.

Hiraiwa, N., M. Kondo, M. Nishimura, and I. Hara-Nishimura. 1997. An aspartic endopeptidase is involved in the breakdown of propeptides of storage proteins in protein-storage vacuoles of plants. European Journal of Biochemistry 246:133-141.

Hodges, M. 2001. Enzyme redundancy and the importance of 2-oxoglutarate in plant ammonium assimilation. Journal of Experimental Botany 53:905-916.

Hollmann, J., P.L. Gregersen, and K. Krupinska. 2014. Identification of predominant genes involved in regulation and execution of senescence-associated nitrogen remobilization in flag leaves of field grown barley. Journal of Experimental Botany 65:3963-3973.

Huang, S., A.S. Raman, J.E. Ream, H. Fujiwara, R.E. Cerny, and S.M. Brown. 1998. Overexpression of 20-oxidase confers a gibberellin-overproduction phenotype in Arabidopsis. Plant Physiology 118:773-781.

Huang, X.Q., R.Q. He, X.Y. Liao, B. Zhou, W.S. Peng, J.Z. Lin, D.Y. Tang, Y.H. Zhu, X.Y. Zhao, and X.M. Liu. 2014. Effect of exogenous gibberellin on reserve accumulation during the seed filling stage of oilseed rape. Genetics and Molecular Research 13:2827-2839.

Hyten, D.L., S.B. Cannon, Q. Song, and N. Weeks. 2010a. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38.

Hyten, D.L., Q. Song, E.W. Fickus, and C.V. Quigley. 2010b. High throughput SNP discovery and assay development in common bean. BMC Genomics 11:475.

Imsande, J. 1992. Agronomic characteristics that identify high yield, high protein soybean genotypes. Agronomy Journal 84:409-414.

Janninks, J. 2001. Using interconnected populations to find quantitative trait loci. [Online] Available at: http://www.reeis.usda.gov/web/crisprojectpages/0189427-using-interconnected-populations-to-find-quantitative-trait-loci.html. Accessed on February 24th 2015.

Jha, B., N.P. Singh, and A. Mishra. 2012. Proteome profiling of seed storage proteins reveals the nutritional potential of *Salicornia brachiata* Roxb., an extreme halophyte. Journal of Agricultural and Food Chemistry 60:4320−4326.

Joshi, N.A., and J.N. Fass. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. [Online] Available at: https://github.com/najoshi/sickle. Accessed January 20th 2015.

Jukanti, A., N. Heidlebaugh, D. Parrott, I. Fischer, K. McInnerney, and A. Fischer. 2008. Comparative transcriptome profiling of near-isogenic barley (*Hordeum vulgare*) lines differing in the allelic state of a major grain protein content locus identifies genes with possible roles in leaf senescence and nitrogen reallocation. New Phytologist 77:333-349.

Junk, J., and K. Grossmann. 1985. Effectiveness of new terpenoid derivatives, abscisic acid and its methyl ester on transpiration and leaf senescence of barley. Journal of Plant Physiology 121:361-367.

Krishnan, H.B., S.S. Natarajan, A.A. Mahmoud, and R.L. Nelson. 2007. Identification of glycinin and β-conglycinin subunits that contribute to the increased protein content of high-protein soybean lines. Journal of Agricultural and Food Chemistry 55:1839-1845.

Kumar, V., A.W. Khan, R.K. Saxena, V. Garg, and R.K. Varshney. 2016. First-generation HapMap in *Cajanus* spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnology Journal 14:1673-1681.

Lam, H.-M., P. Wong, H.-K. Chan, K.-M. Yam, L. Chen, C.-M. Chow, and G.-M. Coruzzi. 2003. Overexpression of the ASN1 gene enhances nitrogen status in seeds of Arabidopsis. Plant Physiology 132:926-935.

Langmead, B., and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.

Lestari, P., K. Van, J. Lee, Y.J. Kang, and S.-H. Lee. 2013. Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. Frontiers in Plant Science 4:176.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Li, J., X. Dai, T. Liu, and P.X. Zhao. 2012. LegumeIP: an integrative database for comparative genomics and transcriptomics of model legumes. Nucleic Acids Research 40(Database issue): D1221-D1229.

Li, L., Y. Xing, D. Chang, S. Fang, B. Cui, Q. Li, X. Wang, S. Guo, X. Yang, S. Men, and Y. Shen. 2016. CaM/BAG5/Hsc70 signaling complex dynamically regulates leaf senescence. Scientific Reports 6:31889.

Lim, J.-H., H.-J. Yang, K.-H. Jung, S.-C. Yoo, and N.-C. Paek. 2014. Quantitative trait locus mapping and candidate gene analysis for plant architecture traits using whole genome re-sequencing in rice. Molecules and Cells 37:149-160.

Liu, Y., L. Wang, H. Liu, R. Zhao, B. Liu, Q. Fu, and Y. Zhang. 2016. The antioxidative defense system is involved in the premature senescence in transgenic tobacco (*Nicotiana tabacum* NC89). Biological Research 49:30.

Lopez-Fanarraga, M., Avila, J., A. Guasch, M. Coll, and J.C. Zabala. 2001. Review: postchaperonin tubulin folding cofactors and their role in microtubule dynamics. Journal of Structural Biology 135:219-229.

Mace, E.S., H.K. Buhariwalla, and J.H. Crouch. 2003. A high throughput DNA extraction protocol for molecular breeding programs. Plant Molecular Biology Reporter 21:459-459.

Machado, M., W.C.Magalhães, A. Sene, B. Araújo, A.C. Faria-Campos, S.J. Chanock, L. Scott, G. Oliveira, E. Tarazona-Santos, and M.R. Rodrigues. 2011. Phred-Phrap package to analyses tools: a pipeline to facilitate population genetics re-sequencing studies. Investigative Genetics 2:3.

Marty, F. 1999. Plant Vacuoles. The Plant Cell 11:587-599.

Mazorra-Manzano, M.A., T. Tanaka, D.R. Dee, and R.Y. Yada. 2010. Structure-function characterization of the recombinant aspartic proteinase A1 from *Arabidopsis thaliana*. Phytochemistry 71:515-523.

Meguro, M., H. Ito, A. Takabayashi, R. Tanaka, and A. Tanaka. 2011. Identification of the 7-hydroxymethyl chlorophyll a reductase of the chlorophyll cycle in Arabidopsis. The Plant Cell 23:3442-3453.

Michaels, S.D., and R.M. Amasino. 1998. A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. The Plant Journal 14:381-385.

Neff, M.M., E. Turk, and M. Kalishman. 2002. Web-based primer design for single nucleotide polymorphism analysis. Trends in Genetics 18:613-61.

Nigro, D., Y.Q. Gu, N. Huo, I. Marcotuli, A. Blanco, A. Gadaleta, and O.D. Anderson. 2013. Structural analysis of the wheat genes encoding NADH-dependent glutamine-2-oxoglutarate amidotransferases and correlation with grain protein content. PLoS One 8: e73751.

Novaes, E., D.R. Drost, W.G. Farmerie, G.J. Pappas, D. Grattapaglia, R.R. Sederoff, and M. Kirst 2008. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics 9:312.

Ohta, M., and F. Takaiwa. 2014. Emerging features of ER resident J-proteins in plants. Plant Signaling and Behavior 9:e28194.

Ohta, M., Y. Wakasa, H. Takahashi, S. Hayashi, K. Kudo, and F. Takaiwa. 2013. Analysis of rice ER-resident J-proteins reveals diversity and functional differentiation of the ER-resident Hsp70 system in plants. Journal of Experimental Botany 64:5429-41.

Olson, N.D., S.P. Lund, R.E. Colman, J.T. Foster, J.W. Sahl, J.M. Schupp, P. Keim, J.B. Morrow, M.L. Salit, and J.M. Zook. 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Frontiers in Genetics 6:235.

Pandurangan, S., A. Pajak, S.J. Molnar, E.R. Cober, S. Dhaubhadel, C. Hernández-Sebastiá, W.M. Kaiser, R.L. Nelson, S.C. Huber, and F. Marsolais. 2012. Relationship between

asparagine metabolism and protein concentration in soybean seed. Journal of Experimental Botany 63:3173-3184.

Pereira, C.S., D.S. da Costa, S. Pereira, F.D. Nogueira, P.M. Albuquerque, J. Teixeira, C. Faro, and J. Pissarra. 2008. Cardosins in postembryonic development of cardoon: towards an elucidation of the biological function of plant aspartic proteinases. Protoplasma 232:203-213.

Pflieger, S., V. Lefebvre, and M. Causse. 2001. The candidate gene approach in plant genetics: a review. Molecular Breeding 7: 275-291.

Rahaie, M., G.P. Xue, and P.M. Schenk. 2013. The role of transcription factors in wheat under different abiotic stress. In: Vahdati K, and C. Leslie, editors, Abiotic stress-plant responses and applications in agriculture. InTechOpen. doi:10.5772/54795

Rao, D.H., and L.R. Gowda. 2008. Abundant class III acidic chitinase homologue in tamarind (*Tamarindus indica*) seed serves as the major storage protein. Journal of Agricultural and Food Chemistry 56:2175-2182.

Rattanakon, S., R. Ghan, G.A. Gambetta, L.G. Deluc, K.A. Schlauch, and G.R. Cramer. 2016. Abscisic acid transcriptomic signaling varies with grapevine organ. BMC Plant Biology 16:72.

Rentsch, D., S. Schmidt, and M. Tegeder. 2007. Transporters for uptake and allocation of organic nitrogen compounds in plants. Federation of European Biochemical Societies Letters 581:2281-2289.

Rios, J., E. Stein, J. Shendure, H.H. Hobbs, and J.C. Cohen. 2010. Identification by whole genome resequencing of gene defect responsible for severe hypercholesterolemia. Human Molecular Genetics 19:4313-4318.

Roach, J.C., G. Glusman, A.F.A. Smit, V.C.D. Hu, R. Hubley, R.T. Shannon, L. Rowen, K.P. Pant, N. Goodman, M. Bamshad, J. Shendure, R. Drmanac, L.B. Jorde, L. Hood, and D.J. Galas. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636-639.

Santos, L. 2006. Molecular mechanisms of the AAA proteins in plants. In: Guevara-González R.G., and I. Torres-Pacheco, editors. Advances in Agricultural and Food Biotechnology. Research Signpost, Trivandrum. p. 1–15.

Satheeskumar, S., P.J. Sharp, E.S. Lagudah, R.A. McIntosh, and S.J. Molnar. 2011. Genetic association of crown rust gene Pc68, storage protein loci, and resistance gene analogues in oats. Genome 54:484-497.

Saxena, R.K., V.K. Singh, S.M. Kale, R. Tathineni, S. Parupalli, V. Kumar, V. Garg, R.R. Das, M. Sharma, K.N. Yamini, S. Muniswamy, A. Ghanta, A. Rathore, C.V. Sameer-Kumar, K.B. Saxena, P.B.K Kishor, and R.K. Varshney. 2017. Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for fusarium wilt resistance in pigeonpea. Scientific Reports 7: 1911.

Schippers, J.H.M., R. Schmidt, C. Wagstaff, and H.-C. Jing. 2015. Living to die and dying to live: the survival strategy behind leaf senescence. Plant Physiology 169:914-930.

Schoenbeck, M.A., S.J. Temple, G.B. Trepp, J.M. Blumenthal, D.A. Samac, S.J. Gantt, G Hernandez, and C.P. Vance. 2000. Decreased NADH-glutamate synthase activity in nodules and flowers of alfalfa (*Medicago sativa* L.) transformed with an antisense glutamate synthase transgene. Journal of Experimental Botany 51:29-39.

Sharma, S., H.D. Upadhyaya, R.K. Varshney, and C.L.L. Gowda. 2013. Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. Frontiers in Plant Science 4:309.

Silva, J., B. Scheffler, Y. Sanabria, C. de Guzman, D. Galam, A. Farmer, J. Woodward, G. May, and J. Oard. 2012. Identification of candidate genes in rice for resistance to sheath blight disease by whole genome sequencing. Theoretical and Applied Genetics 124:63-74.

Singh, S., A.K. Singh, M. Kumar, A.K. Pandey, P.K. Pandey, and M.C. Singh. 2014. Effect of copper on germination, protein content and peroxidase activity of *Phaseolus vulgaris*. Elixir Food Science 73:26001-26003.

Singh, V.K., A.W. Khan, R.K. Saxena, V. Kumar, S.M. Kale, P. Sinha, A. Chitikineni, L.T. Pazhamala, V. Garg, M. Sharma, C.V.S. Kumar, S. Parupalli, S. Vechalapu, S. Patil, S. Muniswamy, A. Ghanta, K.N. Yamini, P.S. Dharmaraj, and R.K. Varshney. 2015. Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). Plant Biotechnology Journal 14:1183-94.Sobreira, N.L.M., E.T. Cirulli, D. Avramopoulos, E. Wohler, G.L. Oswald, E.L. Stevens, D. Ge, K.V. Shianna, J.P. Smith, J.M. Maia, C.E. Gumbs, J. Pevsner, G. Thomas, D. Valle, J.E. Hoover-Fong, and D.B. Goldstein. 2010. Whole-genome

sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genetics 6:e1000991.

Sobieszczuk-Nowicka, E., A. Zmienko, A. Samelak-Czajka, M. Łuczak, M. Pietrowska-Borek, R. Iorio, S. Del-Duca, M. Figlerowicz, and J. Legocka. 2015. Dark-induced senescence of barley leaves involves activation of plastid transglutaminases. Amino Acids 47:825-838

The UniProt Consortium. 2008. The universal protein resource (UniProt). Nucleic Acids Research 36(Database issue), D190–D195. doi:10.1093/nar/gkm895

Varshney, R.K., R.K. Saxena, H.D. Upadhyaya, A.W. Khan, Y. Yu, C. Kim, A. Rathore, D. Kim, J. Kim, S. An, V. Kumar, G. Anuradha, K.N. Yamini, W. Zhang, S. Muniswamy, J. Kim, R.V. Penmetsa, E. von Wettberg, S.K. Datta. 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nature Genetics 49:1082–1088.

Varshney, R.K., W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T.A. Donoghue, S. Azam, G. Fan, A.M. Whaley, A.D. Farmer, J. Sheridan, A. Iwata, R. Tuteja, R.V. Penmetsa, W. Wu, H.D. Upadhyaya, S. Yang, T. Shah, K.B. Saxena, T. Michael, W.R. McCombie, B. Yang, G. Zhang, H. Yang, J. Wang, C. Spillane, D.R. Cook, G.D. May, X. Xu, and S.A. Jackson. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology 30:83-89.

Weigelt. K., H. Küster, T. Rutten, A. Fait, A.R. Fernie, O. Miersch, C. Wasternack, R.J.N. Emery, C. Desel, F. Hosein, M. Müller, I. Saalbach, and H. Weber. ADP-glucose pyrophosphorylase-deficient pea embryos reveal specific transcriptional and metabolic changes of carbon-nitrogen metabolism and stress responses. Plant Physiology 149:395-411.

Upadhyaya, H.D., D. Bajaj, L. Narnoliya, S. Das, V. Kumar, C.L.L. Gowda, S. Sharma, A.K. Tyagi, and S.K. Parida. 2016. Genome-wide scans for delineation of candidate genes regulating seed protein content in chickpea. Frontiers in Plant Science 7:302.

Xu, J., Y., Yuan, Y. Xu, G. Zhang, X Guo, F Wu, Q Wang, T Rong, G Pan, M Cao, Q Tang, S. Gao, Y Liu, J Wang, H. Lan, and Y. Lu. 2014. Identification of candidate genes for drought tolerance by whole-genome re-sequencing in maize. BMC Plant Biology 14:83.

Yamada, T., Y. Mori, K. Yasue, N. Maruyama, K. Kitamura, and J. Abe. 2014. Knockdown of the 7S globulin subunits shifts distribution of nitrogen sources to the residual protein fraction in transgenic soybean seeds. Plant Cell Reports 33:1963-1976.

You, F.M., N. Huo, K.R. Deal, Y.Q. Gu, M.-C. Luo, P.E. McGuire, J. Dvorak, and O.D. Anderson. 2011. Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. BMC Genomics 12:59.

You, F.M., N. Huo, Y.Q. Gu, M.-C. Luo, Y. Ma, D. Hane, G.R. Lazo, J. Dvorak, and O.D. Anderson. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253.

Zeng, Y.-D., J.-L. Sun, S.-H. Bu, K.-S. Deng, T. Tao, Y.-M. Zhang, T.-Z. Zhang, X.-M. Du, and B.-L. Zhou. 2016. EcoTILLING revealed SNPs in GhSus genes that are associated with fiber- and seed-related traits in upland cotton. Scientific Reports 6:29250.

Zhang, Y., G. Cao, L.-J. Qu, and H. Gu. 2008. Involvement of an R2r3-Myb transcription factor gene AtMYB118 in embryogenesis in Arabidopsis. Plant Cell Reports 28:337-346.

Chapter Six
Overview of research finding

6.1 Introduction

Pigeonpea is an important source of dietary protein to nearly a billion people worldwide. However, breeding objectives in pigeonpea have for a long time, almost entirely focused on increasing yield and crop adaptability (Saxena, 2005; Odeny, 2007; Mligo and Craufurd, 2005; Upadhyaya et al., 2007). Very little or no attention has been given to the nutritional quality of the pigeonpea seed in terms of genetic enhancement. To improve SPC in pigeonpea through breeding requires a clear understanding of the genetic control of the trait. It is also essential to know the relationship of SPC with important agronomic traits such as seed yield and seed weight. Such information would allow designing cultivars with increased and stabilised SPC and acceptable agronomic characters (Burstin et al., 2007). There are, however, few documented studies on the genetic control of SPC (Dahiya et al., 1977; Vaghela et al., 2009; Baskaran and Muthiah, 2007) and its association with important agronomic characters in pigeonpea (Dahiya et al., 1977; Saxena et al., 1987; Rekha et al., 2013). Thus, the available information is not only limited but also does not give a clear picture of the genetic architecture of SPC nor its relationship with important agronomic traits in pigeonpea.

The specific objectives of this study were to:

1. Determine the variation of SPC and its relationship with agronomic traits of importance in a set of pigeonpea breeding lines and landraces.
2. Study the inheritance of seed protein content and its relationship with seed weight and seed yield.
3. Identify QTL conditioning seed protein content and its relationship with agronomic traits.
4. Identify candidate genes involved in the accumulation of seed protein content using whole genome sequencing approach.

6.2 Summary of major findings

The first study focused on establishing the level of variation in SPC and its relationship to some agronomic traits among diverse set of cultivated pigeonpea genotypes. The majority of the genotypes are parents of different types of mapping populations currently being developed at ICRISAT, India. The specific outcome of the study were that:

- A considerable variability for SPC and selected agronomic exists among the cultivated pigeonpea germplasm used in this study that can allow for selection and hybridisation of parental genotypes for further genetic studies and breeding purposes.
- Among the 23 genotypes studied, HPL 24, ICP 14486, ICP 5529, HPL 28 and HPL 31 recorded high SPC while genotypes ICPL 87, ICPL 20097, ICPL 85063, ICP 99050 and ICPL 87119 recorded low SPC in that order.
- Correlation and path analysis revealed the importance of 100-seed weight, number of seed per pods, number of pods per plant and seed yield in influencing SPC.

The second study examined the inheritance pattern of SPC in three crosses of pigeonpea developed from four elite germplasm lines of varying SPC. It also validated correlation of SPC with 100-seed weight and seed yield. The main findings of this study were:

- Some specific crosses involving low (19%) to moderate (23%) SPC cultivars can yield transgressive segregants with SPC as high as 25 to 27% which is a significant improvement over the parental values.
- Narrow-sense heritability of 0.55 was moderate and close to 0.65 previously reported in pigeonpea.
- Additive genetic variance was larger than non-additive component although non-additive genetic effect was more predominant compared to additive effects.
- Correlations of SPC with 100-seed weight and seed yield were generally negative.

The third study investigated the genetic architecture of SPC and its relationship with DTF, GH, SW and SY through quantitative trait loci (QTL) analysis. The main outcome of the study were:

- Five intraspecific genetic maps of pigeonpea with map lengths ranging from 1327.6 cM to 1599.8 cM and an average marker to marker distance of 1.6 cM to 3.5 cM, respectively, were constructed.
- The individual maps were integrated into a single consensus map containing 984 markers, with an average inter-marker distance of 1.6 cM.
- QTL analysis revealed the complex nature of the genetic architecture of SPC with two to three major effect QTLs in the presence of several modifier/minor effect and epistatic QTLs conditioning SPC in pigeonpea.
- Overlaps of main effect and epistatically pleiotropic QTLs explained nearly all correlations between SPC and the other traits.

The fourth study exploited available whole genome resequencing (WGRS) data of four pigeonpea genotypes to identify candidate variants and genes for SPC. The main findings of the study were that:

- A total of 108 candidate variants putatively associated with variation in SPC were selected.
- Selected variants are associated with 57 genes spread over all the 11 chromosomes except CcLG05.
- Sanger sequencing confirmed presence of 72.5% of the identified variants.
- Gene ontology based function analysis and comparison of identified candidate genes with those reported in literature indicated that the selected candidate genes play roles in SPC.
- Single nucleotide polymorphisms converted into CAPS/dCAPs markers showed high polymorphism between low and high, and less so between high and high, moderate and moderate, and low and low SPC genotypes.
- Eleven of the CAPS/dCAPS markers were incorporated into a genetic map of an F_2 mapping population (ICP 5529 × ICP 11605).
- Comparison of position of the 11 candidate SNP-derived CAPS/dCAPS markers in the genetic linkage map with positions of SPC QTLs in the population found three CAPS/dCAPS markers to fall within <10.0 cM distance of QTL positions.

- Marker-phenotype co-segregation analysis confirmed association of four CAPS/dCAPS markers with SPC in one F_2 mapping population of the cross ICP 5529 x ICP 11605. Three of the markers were located <10 cM from main effect SPC QTLs.

6.3 Implifications of the research findings

- Based on the findings of the present study;

- A considerable variability for SPC and other important agronomic traits exists within the elite cultivated pigeonpea germplasm used. Most of the germplasm used in the study are parents of different types of mapping populations. Therefore, available variability within and among the mapping populations can allow direct selection of enhanced SPC lines and/or hybridisation of parental genotypes for genetic studies and breeding purposes for SPC and agronomic traits of interest.
- Owing to the complex nature of the genetic architecture of SPC as revealed by classical quantitative genetic analysis, (Chapter Three), QTL analysis (Chapter 4) and candidate gene analysis (Chapter Five), breeding approaches that target genome wide variations for crop improvement would be more appropriate in achieving larger genetic gains for SPC in shorter periods than using conventional phenotype-based selection.

- Future work on genetics of SPC in pigeonpea could focus on use of;

i. Recombinant inbred line mapping populations

The F_2 populations used in the present study facilitated detection of QTLs of major and minor effects as well as digenically interacting QTLs. However, determining stability of the QTLs based on single environment and single plant phenotype data is not possible. Recombinant inbred lines, which are developed through single seed descent (SSD) populations are more attractive for QTL discovery, particularly because they (a) allow for more precise phenotype measurements as trials can be replicated over locations and seasons, and (b) usually involve more meiotic generations leading to many more cross-over events that will facilitate fine mapping of regions of interest leading to map-based cloning.

ii. Near-isogenic lines to improve current understanding of the genetics of seed protein content

A strategy to further understand the genetics of SPC could involve the production of a set of near-isogenic lines (NILs) that differ only for particular QTL regions of interest. The initial step in the NILs strategy would be the selection of suitable recurrent parents, which would be either a highly desirable cultivar or breeding line that would benefit from increased SPC. NILs can be produced by conducting several rounds of backcrossing using markers to select individuals that are heterozygous for the QTL region at each backcross. Background selection may also be conducted in order to increase the rate at which the recurrent parent is fixed. Near-isogenic lines, like RILs, would also allow for replicated, multi-environment phenotypic data to be obtained.

iii. Prioritization strategy that includes variants in both coding and non-coding regions as well as variant types such as copy number variations, large indels, structural genomic variants when conducting candidate gene analysis

While whole genome resequencing have allowed the identification and prioritization of candidate variants that could underlie or contribute to the SPC variability observed in the set pigeonpea genotypes studied, the approach does have limitations. It is possible that other causative variants/mutations have been overlooked in the panel of selected genes as a result of the strategies used to prioritize the candidate SNPs. Causative variants in the non-coding regions of the targeted putative SPC genes, which are not included in the exon, or indeed in other genes that are not in the panel of putative SPC candidate genes, would also be overlooked by this approach. The potential for other genetic mechanisms, such as copy number variation, large indels, or structural genomic variants, to contribute to the underlying mutations also cannot be discounted. Gene-gene interactions is another area that may add insight into gene functions related to SPC, especially that the QTL-QTL interactions already revealed strong influence of epistasis on SPC as noted in Chapter Five of this thesis.

References

Baskaran, K., and A.R. Muthiah. 2007. Studies on combining ability in pigeonpea [*Cajanus cajan* (L.) Millsp.] Legumes Research 30:67-69.

Burstin, J., P. Marget, M. Huart, A. Moessner, B. Mangin, C. Duchene, B. Desprez, N. Munier-Jolain, and G. Duc. 2007. Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiology 144:768-781.

Dahiya, B.S., J.S.Brar, and B.S. Bhullar. 1977. Inheritance of protein content and its correlation with grain yield in pigeonpea (*Cajanus cajan* (L.) Millsp.). Qualitus Plantarum 27:327-334.

Mligo, J.K., and P.Q. Craufurd. 2005. Adaptation and yield of pigeonpea in different environments in Tanzania. Field Crops Research 94:43-53.

Odeny, D.A. 2007. The potential of pigeonpea (*Cajanus cajan* (L.) Millsp.) in Africa. Natural Resources Forum 31:297-305.

Rekha, R., L. Prasanthi, M.R. Sekhar, and M.S. Priya (2013) Studies on selection indices in pigeonpea [*Cajanus cajan* (L.) Millsp]. International Journal of Applied Biology and Pharmaceutical Technology 4:291-294.

Saxena, K.B. 2005. Pigeonpea (*Cajanus cajan* (L.) Millsp.). In: Singh R.J., and P.P. Jauhar, editors, Genetic Resources, Chromosome Engineering and Crop Improvement: Vol. 1, Grain Legumes. CRC Press, Boca Raton, p. 86–110.

Saxena, K.B., D.G. Faris, U. Singh, and R.V. Kumar. 1987. Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition 36:335-340.

Upadhyaya, H.D., K.N. Reddy, D.V.S.S.R. Sastry, and C.L.L. Gowda. 2007. Identification of photoperiod insensitive sources in the world collection of pigeonpea at ICRISAT. Journal of Semi-Arid Tropical Agricultural Research 3:46-49.

Vaghela, K.O., R.T. Desai, J.R. Nizama, J.D. Patel, and V. Sharma. 2009. Combining ability analysis in pigeonpea (C*ajanus cajan* (L.) Millsp.). Legume Research 32:274-277.

Appendices

Appendix 1: Sequence data generated using genotyping by sequencing of parental lines and 178 F_2 individuals of pigeonpea mapping population Pop1 (ICP 114605 x ICP 14209)

Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
Parents			31	2.16	0.22
ICP 11605 (P_1)	1.13	0.11	33	2.93	0.30
ICP 14209 (P_2)	2.60	0.26	34	2.28	0.23
F_2			35	1.53	0.15
1	3.57	0.36	36	1.76	0.18
2	3.36	0.34	37	3.03	0.31
3	5.82	0.59	38	1.73	0.17
4	4.03	0.41	39	3.00	0.30
5	2.57	0.26	40	2.14	0.22
6	2.98	0.30	41	2.53	0.26
7	3.76	0.38	42	1.80	0.18
8	3.42	0.35	43	2.74	0.28
9	1.99	0.20	44	1.69	0.17
10	2.28	0.23	45	2.76	0.28
11	3.74	0.38	46	2.96	0.30
12	2.26	0.23	47	2.49	0.25
13	2.96	0.30	48	2.78	0.28
14	2.34	0.24	49	2.22	0.22
15	1.92	0.19	50	4.17	0.42
16	3.04	0.31	51	2.31	0.23
17	2.30	0.23	52	2.69	0.27
18	5.17	0.52	53	3.28	0.33
19	2.95	0.30	54	2.50	0.25
20	3.95	0.40	55	2.43	0.25
21	2.22	0.22	56	1.97	0.20
22	2.30	0.23	57	2.40	0.24
23	3.10	0.31	58	1.64	0.17
24	2.91	0.29	59	1.45	0.15
25	3.30	0.33	60	2.27	0.23
26	1.82	0.18	61	4.74	0.48
27	1.93	0.19	62	1.82	0.18
28	3.01	0.30	63	2.39	0.24
29	1.94	0.20	64	2.11	0.21
30	3.67	0.37	65	2.12	0.21

Appendix 1: (continued)

Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
66	2.14	0.22	109	1.15	0.12
67	1.89	0.19	110	2.24	0.23
68	2.47	0.25	111	2.53	0.26
69	3.17	0.32	112	2.01	0.20
70	1.92	0.19	113	1.77	0.18
71	2.45	0.25	114	2.85	0.29
72	2.28	0.23	115	2.41	0.24
73	2.13	0.22	116	1.90	0.19
74	1.20	0.12	117	2.46	0.25
75	1.57	0.16	118	1.76	0.18
76	2.11	0.21	119	2.02	0.20
77	1.94	0.20	120	1.27	0.13
78	1.54	0.16	121	1.30	0.13
79	1.68	0.17	122	1.65	0.17
80	1.62	0.16	123	1.11	0.11
81	3.00	0.30	124	2.28	0.23
82	3.31	0.33	125	1.06	0.11
83	2.44	0.25	126	1.03	0.10
84	2.29	0.23	127	1.82	0.18
85	0.79	0.08	128	1.82	0.18
86	1.21	0.12	129	1.11	0.11
87	1.01	0.10	130	1.26	0.13
88	1.52	0.15	131	2.25	0.23
89	1.11	0.11	132	1.89	0.19
95	2.32	0.23	133	2.32	0.23
96	1.68	0.17	134	2.54	0.26
97	3.13	0.32	135	3.58	0.36
98	2.90	0.29	136	2.13	0.21
99	1.90	0.19	137	2.59	0.26
100	2.55	0.26	138	3.29	0.33
101	2.38	0.24	139	2.96	0.30
102	3.18	0.32	140	3.09	0.31
103	1.51	0.15	141	1.66	0.17
104	2.34	0.24	142	2.89	0.29
105	1.59	0.16	143	1.66	0.17
106	1.43	0.14	144	3.28	0.33
107	1.92	0.19	145	2.18	0.22
108	1.74	0.18	146	4.01	0.40

Appendix 1 (continued)

Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
147	2.93	0.30	178	1.57	0.16
148	2.10	0.21	180	1.23	0.12
149	2.41	0.24	181	1.48	0.15
150	2.17	0.22	182	2.82	0.29
151	1.64	0.17	183	1.33	0.13
152	1.19	0.12	184	0.83	0.08
153	1.10	0.11	185	1.41	0.14
154	2.19	0.22	186	2.18	0.22
155	1.78	0.18	187	0.92	0.09
156	1.57	0.16	188	1.58	0.16
157	2.39	0.24	F_2 total	399.93	40.4
158	3.96	0.40	F_2 average	2.25	0.23
159	1.90	0.19			
160	1.13	0.11			
161	1.68	0.17			
162	1.72	0.17			
163	2.52	0.25			
164	1.14	0.11			
165	2.17	0.22			
166	2.29	0.23			
167	1.27	0.13			
168	1.11	0.11			
169	1.48	0.15			
170	2.88	0.29			
172	1.40	0.14			
175	1.46	0.15			
176	1.70	0.17			
177	1.35	0.14			

Appendix 2: Sequence data generated using genotyping by sequencing of parental lines and 178 F_2 individuals of pigeonpea mapping population Pop2 (ICP 8863 × ICP 11605)

Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
Parents			33	1.95	0.20
ICP 8863 (P_1)	3.00	0.30	34	2.13	0.21
ICP 11605 (P_2)	7.59	0.77	35	1.85	0.19
F_2			36	1.82	0.18
1	3.68	0.37	37	2.32	0.23
2	1.53	0.15	39	0.69	0.07
3	3.75	0.38	40	1.33	0.13
4	3.33	0.34	42	0.98	0.10
5	1.80	0.18	44	1.24	0.13
6	3.12	0.31	45	0.79	0.08
7	3.57	0.36	46	0.81	0.08
8	5.44	0.55	47	0.74	0.07
9	1.44	0.15	49	0.76	0.08
10	1.36	0.14	52	0.80	0.08
11	3.66	0.37	53	1.40	0.14
12	1.74	0.18	54	1.50	0.15
13	6.71	0.68	55	1.48	0.15
14	5.22	0.53	56	1.62	0.16
15	0.65	0.07	57	2.76	0.28
16	1.84	0.19	58	1.52	0.15
17	2.54	0.26	59	1.14	0.11
18	3.44	0.35	60	1.62	0.16
19	2.04	0.21	61	2.47	0.25
20	9.52	0.96	62	1.69	0.17
21	1.25	0.13	63	1.29	0.13
22	2.41	0.24	64	0.75	0.08
23	2.17	0.22	65	1.95	0.20
24	1.98	0.20	66	2.24	0.23
25	1.43	0.14	67	1.91	0.19
26	1.40	0.14	68	0.90	0.09
27	2.38	0.24	69	2.47	0.25
28	3.04	0.31	70	4.31	0.43
29	0.96	0.10	71	1.44	0.14
30	1.34	0.14	72	1.74	0.18
31	1.09	0.11	73	0.86	0.09

Appendix 2 (continued)

Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
74	4.69	0.47	111	1.37	0.14
75	0.78	0.08	112	1.36	0.14
76	3.37	0.34	113	1.01	0.10
77	1.58	0.16	114	1.09	0.11
78	2.97	0.30	115	1.41	0.14
79	0.86	0.09	117	6.78	0.68
80	1.15	0.12	118	0.72	0.07
81	4.03	0.41	119	3.80	0.38
82	2.58	0.26	121	1.53	0.15
83	1.04	0.11	122	3.23	0.33
84	0.70	0.07	123	1.98	0.20
85	0.56	0.06	124	2.98	0.30
86	0.76	0.08	125	2.74	0.28
87	0.68	0.07	126	2.19	0.22
88	1.57	0.16	127	3.94	0.40
89	0.86	0.09	128	3.99	0.40
90	0.63	0.06	129	3.93	0.40
91	1.44	0.15	130	3.93	0.40
92	1.21	0.12	131	4.75	0.48
93	0.79	0.08	132	3.65	0.37
94	1.22	0.12	133	1.63	0.16
95	2.38	0.24	135	1.97	0.20
96	2.21	0.22	136	2.52	0.25
97	3.77	0.38	137	1.94	0.20
98	3.42	0.34	138	2.14	0.22
99	0.95	0.10	139	1.37	0.14
100	3.19	0.32	140	2.30	0.23
101	1.74	0.18	141	1.89	0.19
102	4.22	0.43	142	1.39	0.14
103	2.28	0.23	143	1.55	0.16
104	1.46	0.15	144	1.62	0.16
105	2.90	0.29	145	0.94	0.09
106	1.13	0.11	146	0.85	0.09
107	1.77	0.18	147	2.03	0.20
108	1.48	0.15	148	1.26	0.13
109	1.13	0.11	149	0.92	0.09
110	0.75	0.08	150	0.87	0.09

Appendix 2 (continued)

Sample ID	No. reads (Million)	Data (Gb)
152	0.83	0.08
154	1.14	0.12
155	0.90	0.09
156	2.23	0.22
157	0.74	0.07
159	1.04	0.11
160	1.34	0.14
161	1.90	0.19
162	1.03	0.10
163	1.48	0.15
164	1.25	0.13
165	2.12	0.21
166	2.25	0.23
167	1.28	0.13
168	0.94	0.09
169	1.26	0.13
170	1.68	0.17
171	0.73	0.07
172	0.94	0.09
173	0.92	0.09
174	1.09	0.11
175	1.89	0.19
176	1.16	0.12
177	0.68	0.07
178	1.43	0.14
179	0.58	0.06
180	0.63	0.06
181	0.61	0.06
182	0.63	0.06
183	0.87	0.09
184	0.49	0.05
185	0.98	0.10
186	1.09	0.11
187	0.83	0.08
188	0.78	0.08
F_2 total	332.68	34.67
F_2 average	1.90	0.20

Appendix 3: Sequence data generated using genotyping by sequencing of parental lines and 178 F_2 individuals of pigeonpea mapping population Pop3 (HPL 24 × ICP 11605)

Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
Parents			51	1.36	0.14
HPL 24	2.96	0.30	52	1.84	0.19
ICP 11605	3.31	0.33	53	1.44	0.15
F_2			54	1.32	0.13
2	1.04	0.10	55	2.62	0.26
3	1.29	0.13	56	1.20	0.12
7	4.36	0.44	57	4.88	0.49
8	0.85	0.09	58	1.00	0.10
9	1.96	0.20	59	2.01	0.20
11	1.90	0.19	60	1.64	0.17
13	0.89	0.09	61	1.92	0.19
16	3.65	0.37	62	1.10	0.11
17	1.08	0.11	63	0.80	0.08
18	2.97	0.30	64	0.92	0.09
19	0.73	0.07	65	0.92	0.09
20	1.08	0.11	66	2.66	0.27
21	1.80	0.18	67	1.65	0.17
22	0.79	0.08	69	1.16	0.12
23	2.29	0.23	71	1.45	0.15
24	0.86	0.09	73	2.80	0.28
25	4.75	0.48	74	3.03	0.31
26	0.81	0.08	75	2.57	0.26
27	1.95	0.20	76	1.27	0.13
30	0.75	0.08	77	1.31	0.13
31	1.24	0.12	78	0.97	0.10
32	1.03	0.10	79	3.45	0.35
33	3.82	0.39	80	1.02	0.10
34	2.28	0.23	81	0.74	0.08
35	2.07	0.21	83	1.13	0.11
37	0.87	0.09	84	1.02	0.10
39	5.04	0.51	85	2.53	0.26
40	6.84	0.69	86	4.84	0.49
42	1.53	0.15	87	2.20	0.22
43	1.15	0.12	88	1.37	0.14
45	1.10	0.11	89	2.58	0.26
47	3.08	0.31	90	3.73	0.38
49	1.96	0.20	91	1.28	0.13

Appendix 3: (continued)

Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
92	2.57	0.26	134	2.72	0.27
93	3.12	0.32	135	1.80	0.18
94	3.75	0.38	136	2.19	0.22
95	3.45	0.35	137	1.63	0.16
96	2.95	0.30	138	2.54	0.26
97	4.42	0.45	139	1.75	0.18
98	4.99	0.50	140	1.52	0.15
99	1.13	0.11	141	0.80	0.08
100	1.80	0.18	142	1.54	0.16
101	2.75	0.28	143	0.77	0.08
102	2.62	0.26	144	3.40	0.34
103	2.42	0.24	145	1.73	0.18
104	2.61	0.26	146	1.74	0.18
105	3.13	0.32	147	1.17	0.12
106	1.79	0.18	148	1.23	0.12
107	3.14	0.32	149	3.12	0.32
108	1.79	0.18	150	2.04	0.21
109	1.26	0.13	151	1.80	0.18
110	2.40	0.24	152	1.64	0.17
111	3.17	0.32	153	2.17	0.22
112	3.18	0.32	154	2.52	0.25
113	1.33	0.13	155	2.47	0.25
114	1.87	0.19	156	3.05	0.31
115	1.26	0.13	157	2.47	0.25
117	1.44	0.15	158	1.80	0.18
118	0.77	0.08	159	2.64	0.27
119	1.83	0.18	160	1.86	0.19
122	2.25	0.23	161	2.01	0.20
124	1.85	0.19	162	3.07	0.31
125	0.89	0.09	163	3.07	0.31
126	1.64	0.17	164	2.34	0.24
127	2.02	0.20	165	2.79	0.28
128	2.21	0.22	166	2.56	0.26
129	1.08	0.11	167	4.54	0.46
130	1.33	0.13	168	2.01	0.20
131	1.26	0.13	169	2.65	0.27
132	2.06	0.21	170	2.63	0.27
133	1.42	0.14	171	1.91	0.19

Appendix 3: (continued)

Sample ID	No. reads (Million)	Data (Gb)
172	1.96	0.20
173	2.52	0.25
174	3.71	0.37
175	2.68	0.27
176	3.38	0.34
177	2.20	0.22
178	2.45	0.25
179	1.01	0.10
180	1.35	0.14
181	2.49	0.25
182	1.70	0.17
183	1.33	0.13
185	0.96	0.10
186	1.29	0.13
187	0.97	0.10
188	3.10	0.31
F_2 total	339.25	34.26
F_2 average	2.09	0.21

Appendix 4: Sequence data generated using genotyping by sequencing of parental lines and 178 F_2 individuals of pigeonpea mapping population Pop4 (ICP 5529 × ICP 11605)

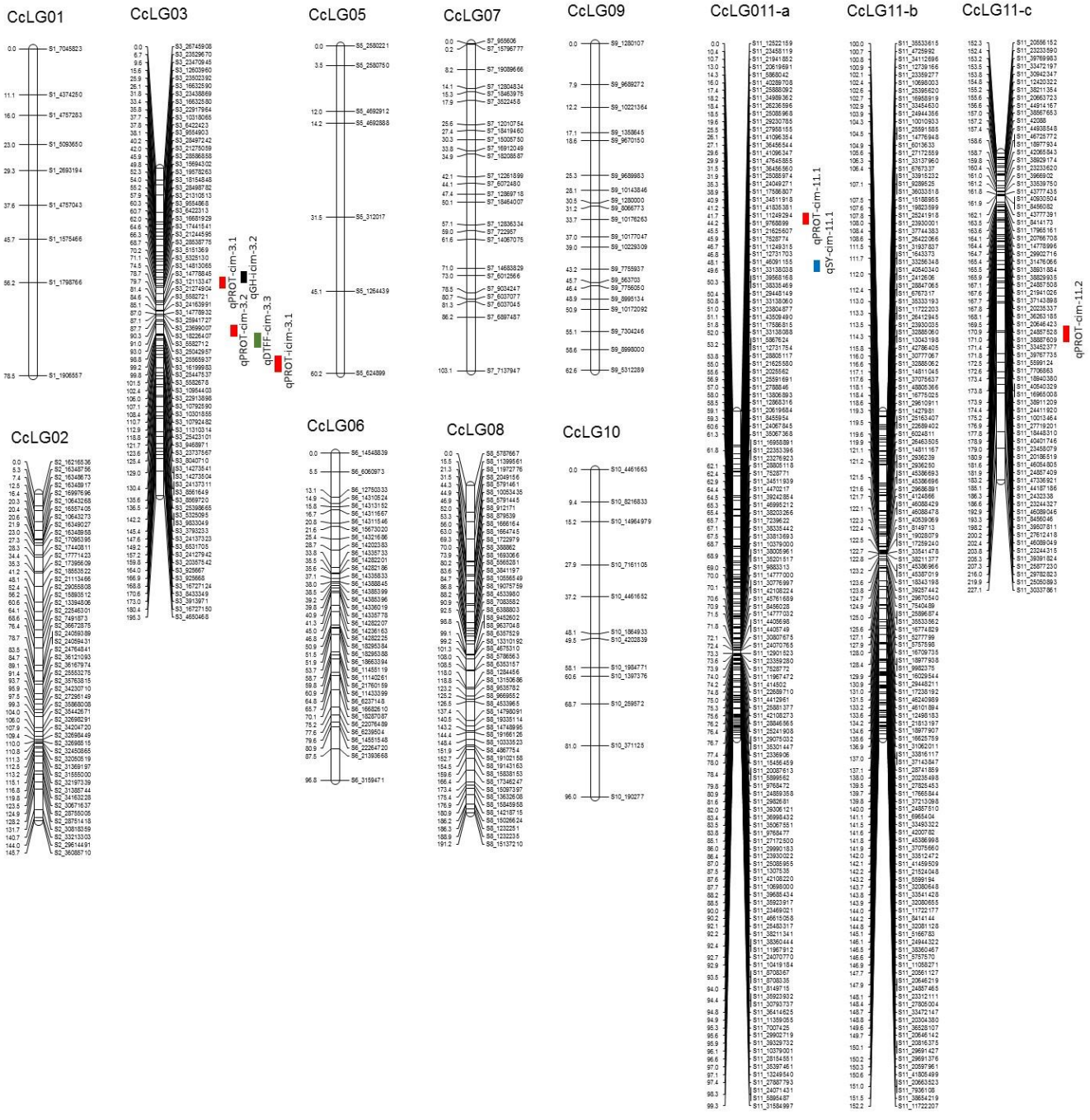
Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
Parents			38	0.64	0.06
ICP 5529	5.37	0.54	39	1.09	0.11
ICP 11605	1.61	0.16	40	1.28	0.13
F_2			41	2.43	0.25
2	2.25	0.23	42	0.89	0.09
3	0.97	0.10	43	2.19	0.22
4	1.82	0.18	44	0.91	0.09
6	3.40	0.34	45	0.86	0.09
7	1.91	0.19	46	0.96	0.10
8	5.26	0.53	48	2.22	0.22
9	2.57	0.26	49	2.66	0.27
10	1.00	0.10	50	2.02	0.20
12	1.43	0.14	52	0.71	0.07
13	1.14	0.12	53	0.96	0.10
14	1.97	0.20	54	0.64	0.06
15	0.62	0.06	55	0.80	0.08
16	1.00	0.10	56	0.99	0.10
17	1.33	0.13	57	3.05	0.31
18	0.77	0.08	58	1.30	0.13
19	2.59	0.26	59	2.02	0.20
20	0.87	0.09	60	2.04	0.21
21	0.68	0.07	61	1.92	0.19
22	1.02	0.10	62	1.73	0.17
23	4.11	0.41	63	1.05	0.11
24	1.82	0.18	64	1.26	0.13
25	1.98	0.20	65	3.57	0.36
26	1.22	0.12	66	2.52	0.25
27	0.96	0.10	67	4.01	0.41
28	2.82	0.28	68	1.46	0.15
29	0.67	0.07	69	2.18	0.22
30	1.15	0.12	70	2.20	0.22
31	1.26	0.13	71	1.19	0.12
32	0.72	0.07	72	1.83	0.18
33	0.89	0.09	73	2.04	0.21
35	1.94	0.20	74	2.26	0.23
36	1.58	0.16	75	2.70	0.27
37	1.29	0.13	76	1.91	0.19

Appendix 4 (continued)

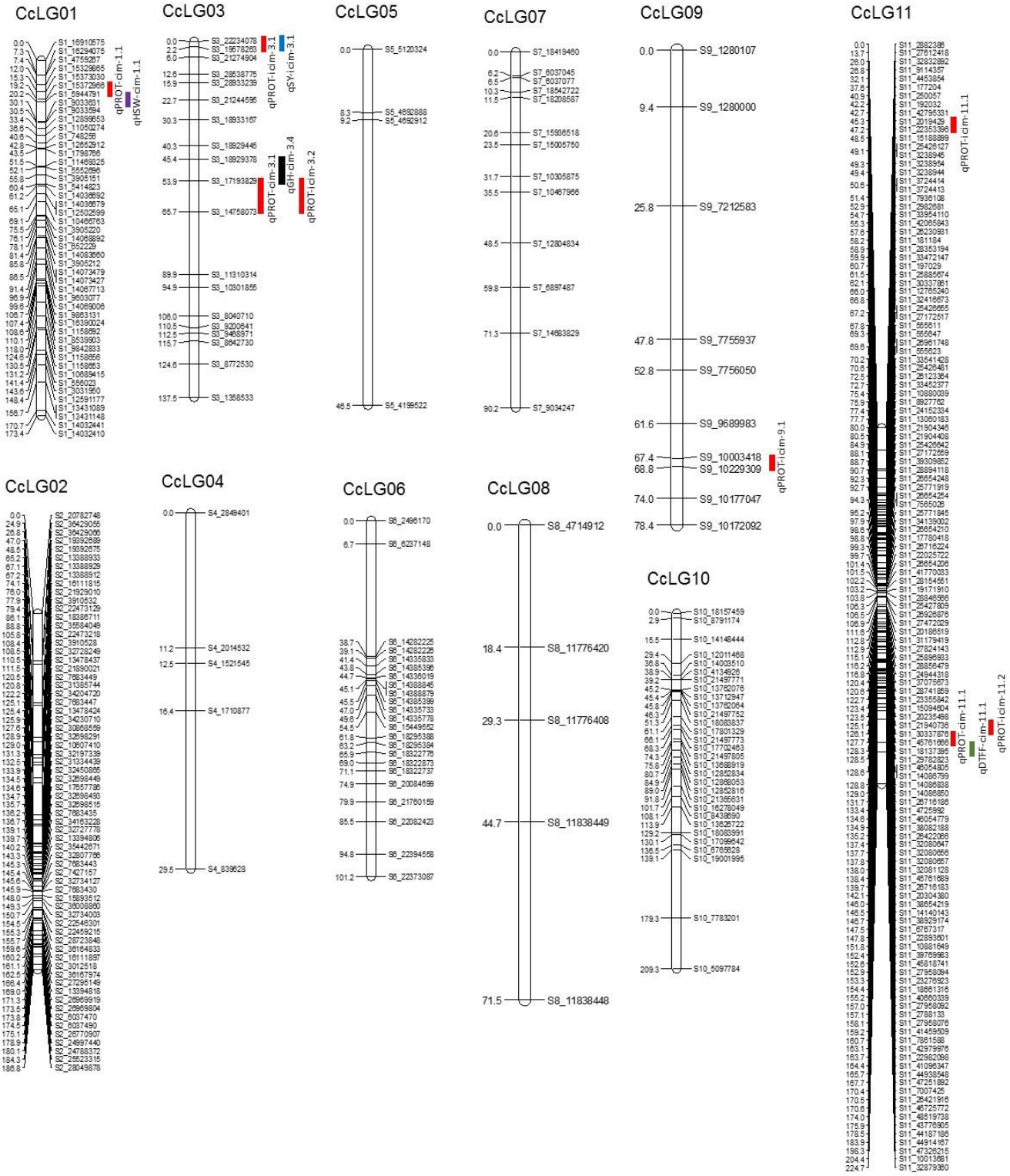
Sample ID	No. reads (Million)	Data (Gb)	Sample ID	No. reads (Million)	Data (Gb)
77	3.47	0.35	117	1.68	0.17
78	1.89	0.19	118	0.51	0.05
79	2.06	0.21	119	0.91	0.09
80	2.14	0.22	120	1.13	0.11
81	4.08	0.41	121	1.56	0.16
82	1.43	0.14	122	1.62	0.16
83	2.48	0.25	123	1.21	0.12
84	3.11	0.31	124	1.91	0.19
85	0.92	0.09	125	0.96	0.10
86	1.64	0.17	126	0.52	0.05
87	1.89	0.19	127	2.93	0.30
89	1.24	0.12	128	1.06	0.11
90	1.22	0.12	129	0.91	0.09
91	1.37	0.14	130	2.32	0.23
92	1.18	0.12	131	1.33	0.13
93	2.90	0.29	132	1.32	0.13
94	1.92	0.19	133	1.00	0.10
95	1.31	0.13	134	0.59	0.06
96	1.23	0.12	135	1.16	0.12
97	1.91	0.19	136	1.43	0.14
98	1.39	0.14	137	2.05	0.21
99	0.64	0.06	138	2.05	0.21
100	1.06	0.11	139	1.53	0.15
101	0.58	0.06	140	1.23	0.12
102	0.76	0.08	142	0.66	0.07
103	0.74	0.07	143	1.29	0.13
104	0.70	0.07	144	0.89	0.09
105	1.89	0.19	145	0.80	0.08
106	1.04	0.11	146	1.65	0.17
107	0.79	0.08	147	1.26	0.13
108	0.47	0.05	148	1.12	0.11
109	0.41	0.04	149	0.74	0.07
111	0.83	0.08	150	1.70	0.17
112	1.44	0.15	151	1.99	0.20
113	1.61	0.16	152	1.56	0.16
114	2.52	0.25	153	1.87	0.19
115	1.17	0.12	154	2.18	0.22
116	0.64	0.07	155	2.81	0.28

Appendix 4 (continued)

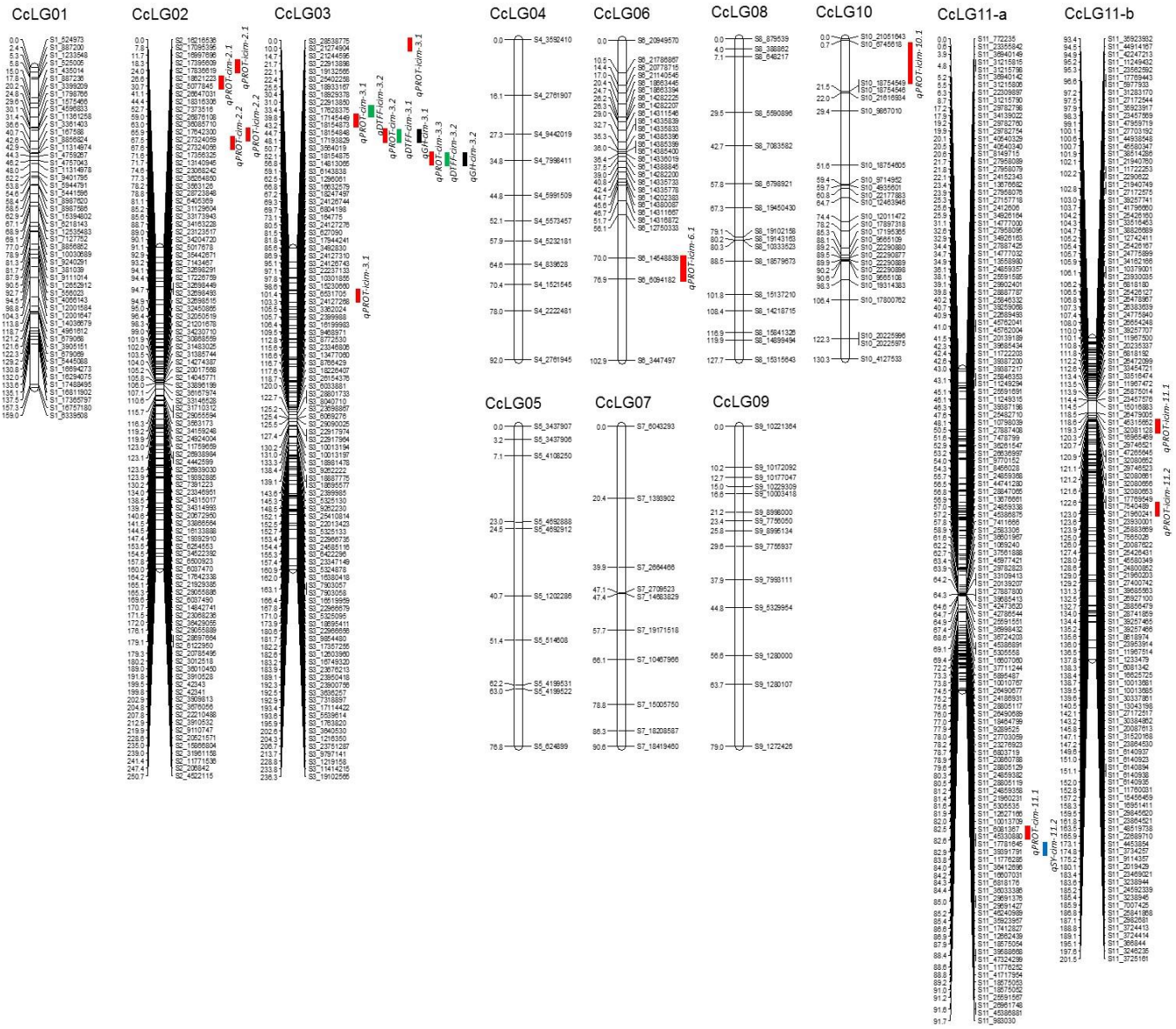
Sample ID	No. reads (Million)	Data (Gb)
156	1.90	0.19
157	2.28	0.23
158	2.73	0.28
159	1.27	0.13
160	2.25	0.23
161	2.04	0.21
162	1.38	0.14
163	3.36	0.34
164	1.86	0.19
165	3.77	0.38
166	2.36	0.24
167	3.08	0.31
168	2.32	0.23
169	1.55	0.16
170	1.98	0.20
171	1.83	0.19
172	2.36	0.24
173	2.96	0.30
174	1.53	0.15
175	3.02	0.31
176	1.95	0.20
177	1.63	0.16
178	2.35	0.24
179	2.11	0.21
180	2.65	0.27
181	2.13	0.22
182	0.69	0.07
183	2.16	0.22
184	0.62	0.06
185	0.92	0.09
186	0.61	0.06
187	1.96	0.20
188	1.94	0.20
F_2 total	298.56	30.15
F_2 average	1.67	0.17



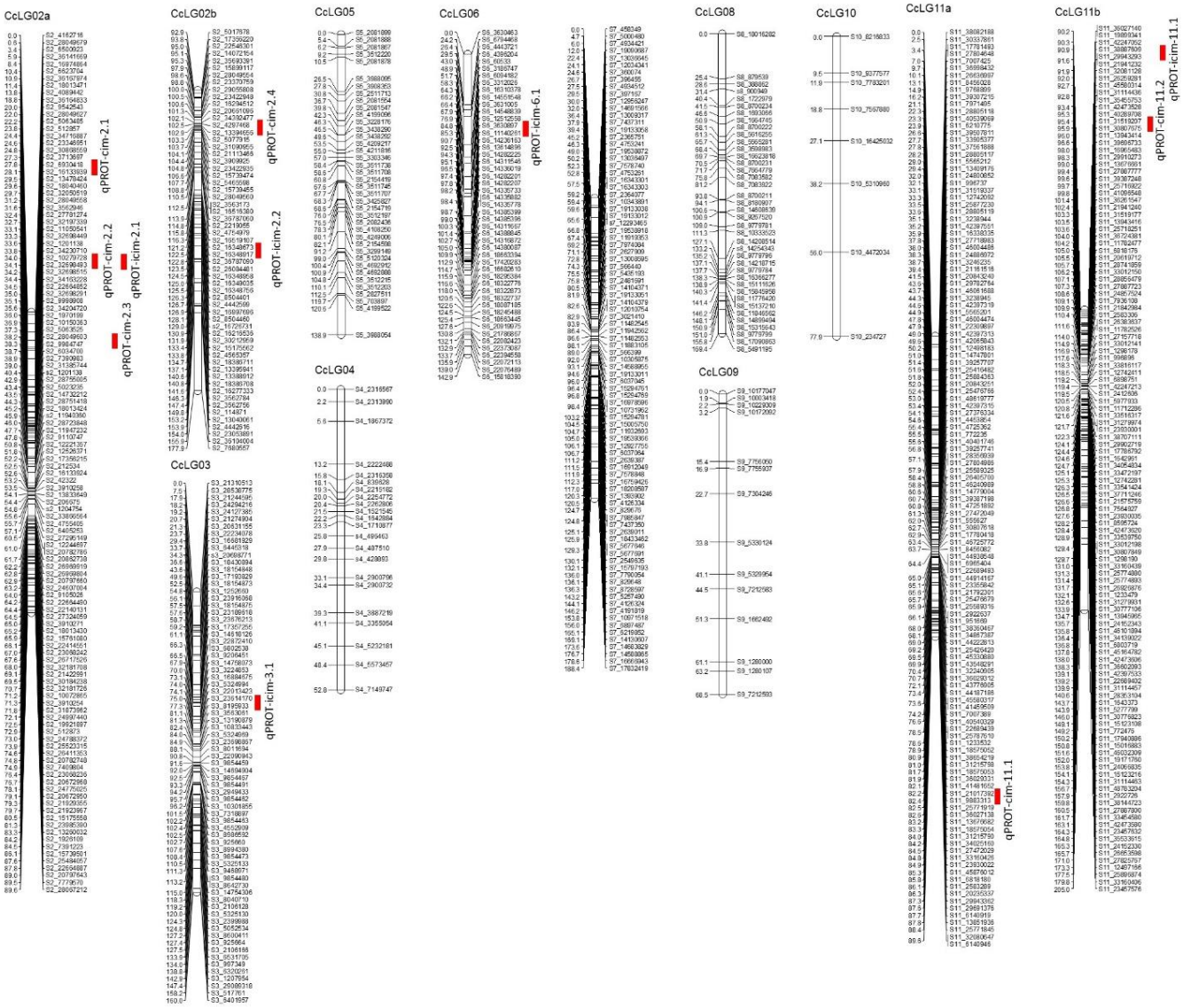
Appendix 5: Positions of markers and QTLs in linkage map of Pop1 (ICP 11605 × ICP 14209). Markers are shown on right side of the linkage group while map distances are indicated on left side. QTLs for the different traits are indicated by bars with different colours with black, green and red showing QTLs for GH, DTF and SPC, respectively.



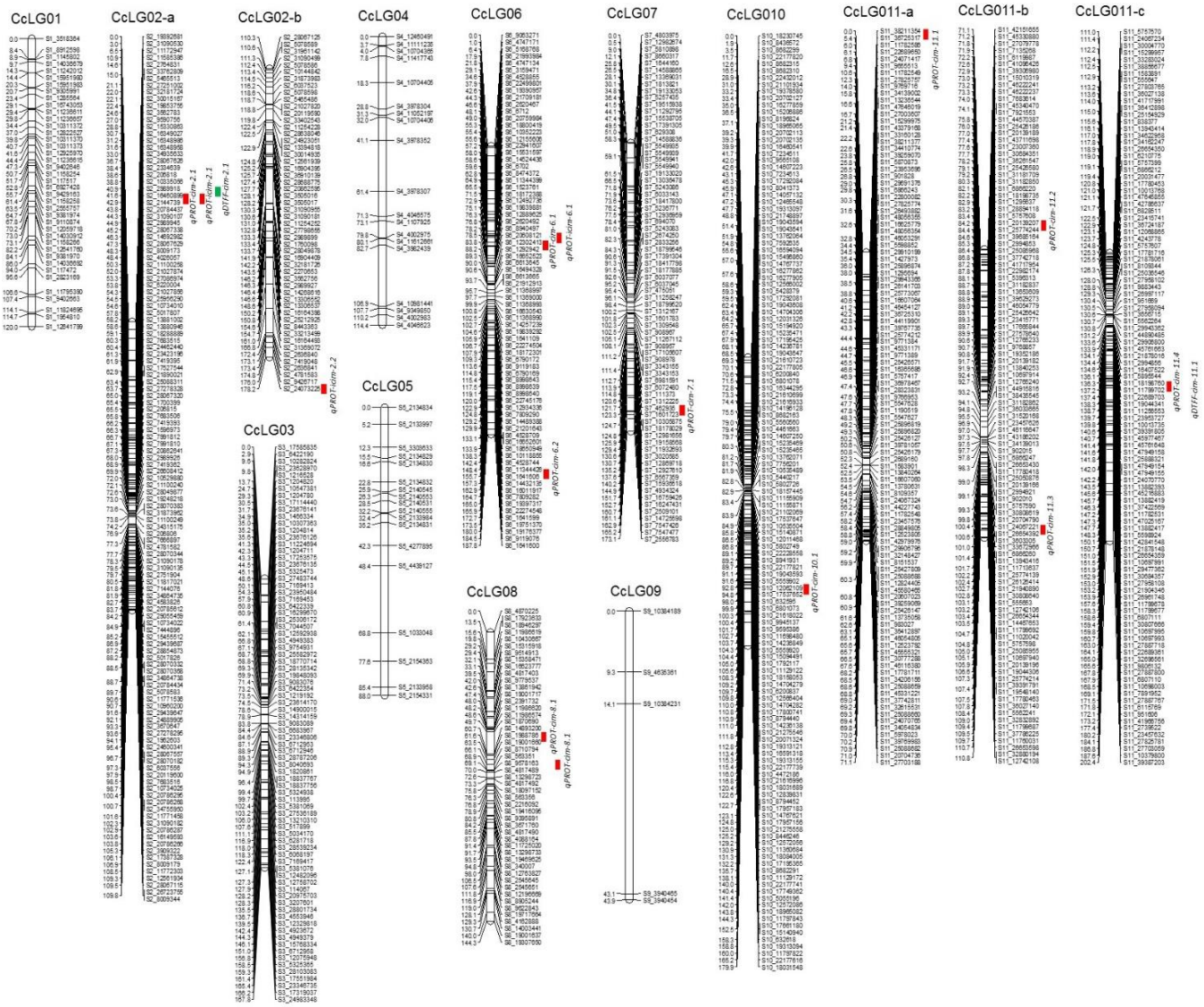
Appendix 6: Positions of markers and QTLs in linkage map of Pop2 (ICP 8863 × ICP 11605). Markers are shown on right side of the linkage group while map distances are indicated on left side. QTLs for the different traits are indicated by bars with different colours with black, blue, green, purple and red showing QTLs for GH, SY, DTF, SW and SPC, respectively.



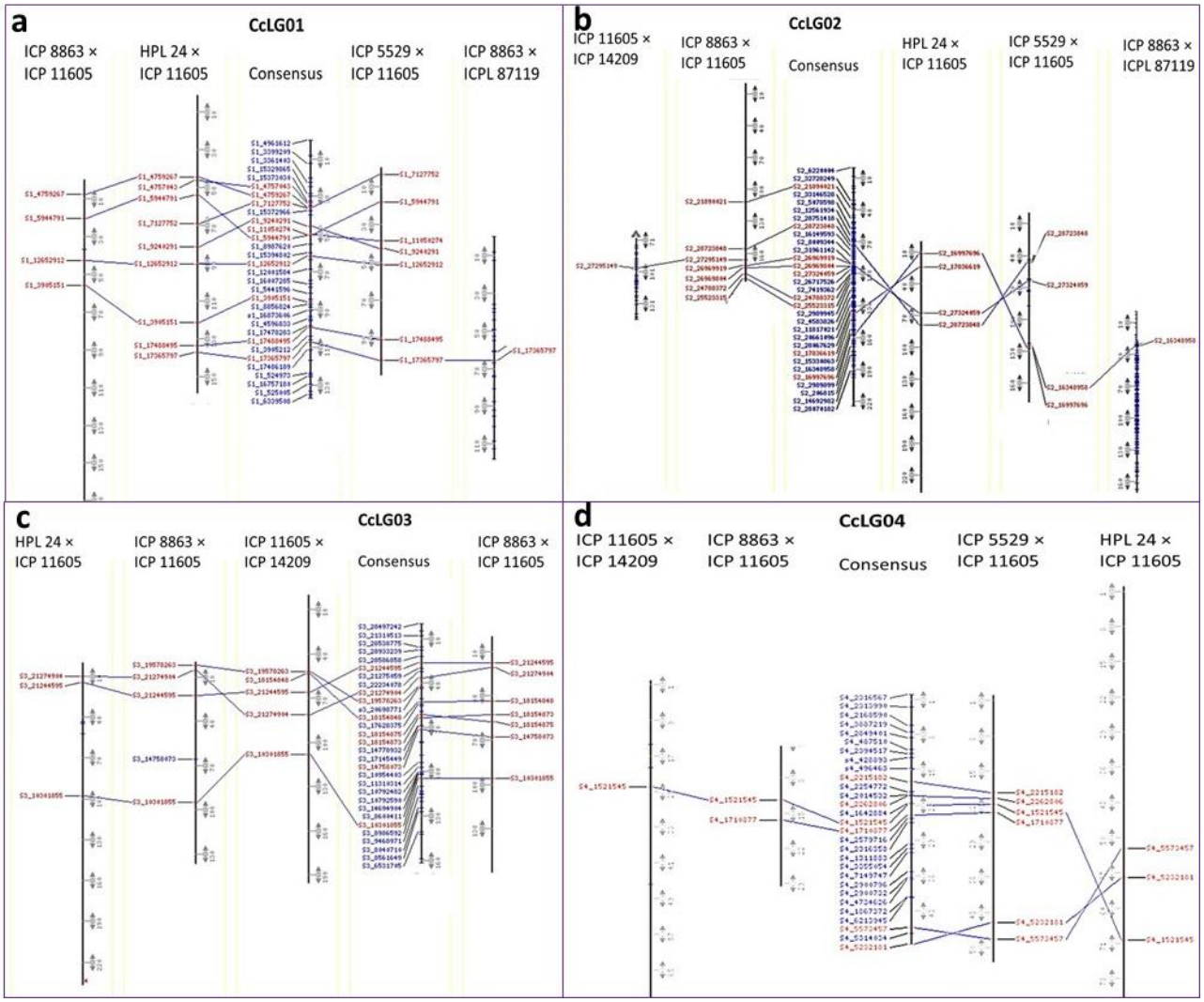
Appendix 7: Positions of markers and QTLs in linkage map of Pop3 (HPL 24 × ICP 11605). Markers are shown on right side of the linkage group while map distances are indicated on left side. QTLs for the different traits are indicated by bars with different colours with black, blue, green and red showing QTLs for GH, SY, DTF and SPC, respectively.



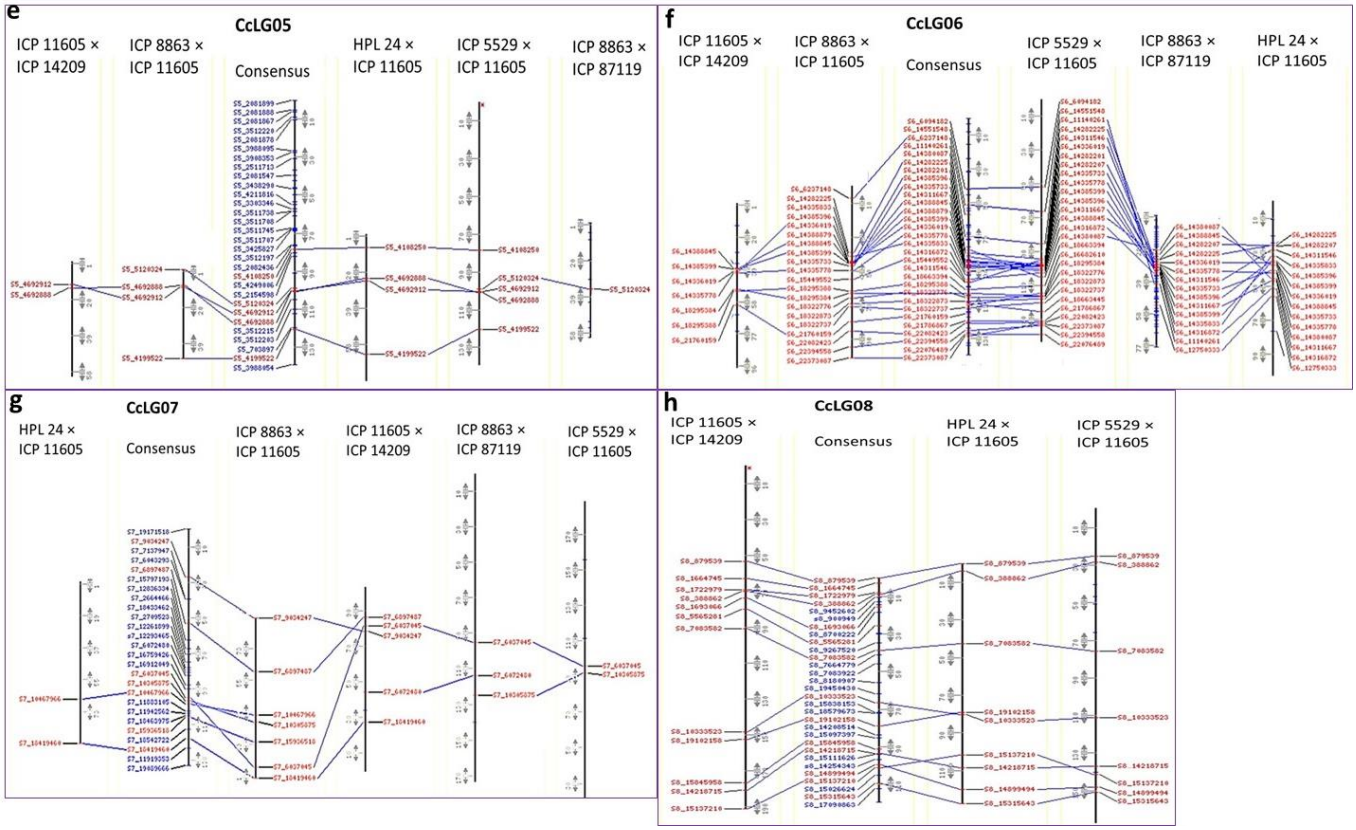
Appendix 8: Positions of markers and QTLs in linkage map of Pop4 (ICP 5529 × ICP 11605). Markers are shown on right side of the linkage group while map distances are indicated on left side. QTLs for SPC is represented with red colour bars.



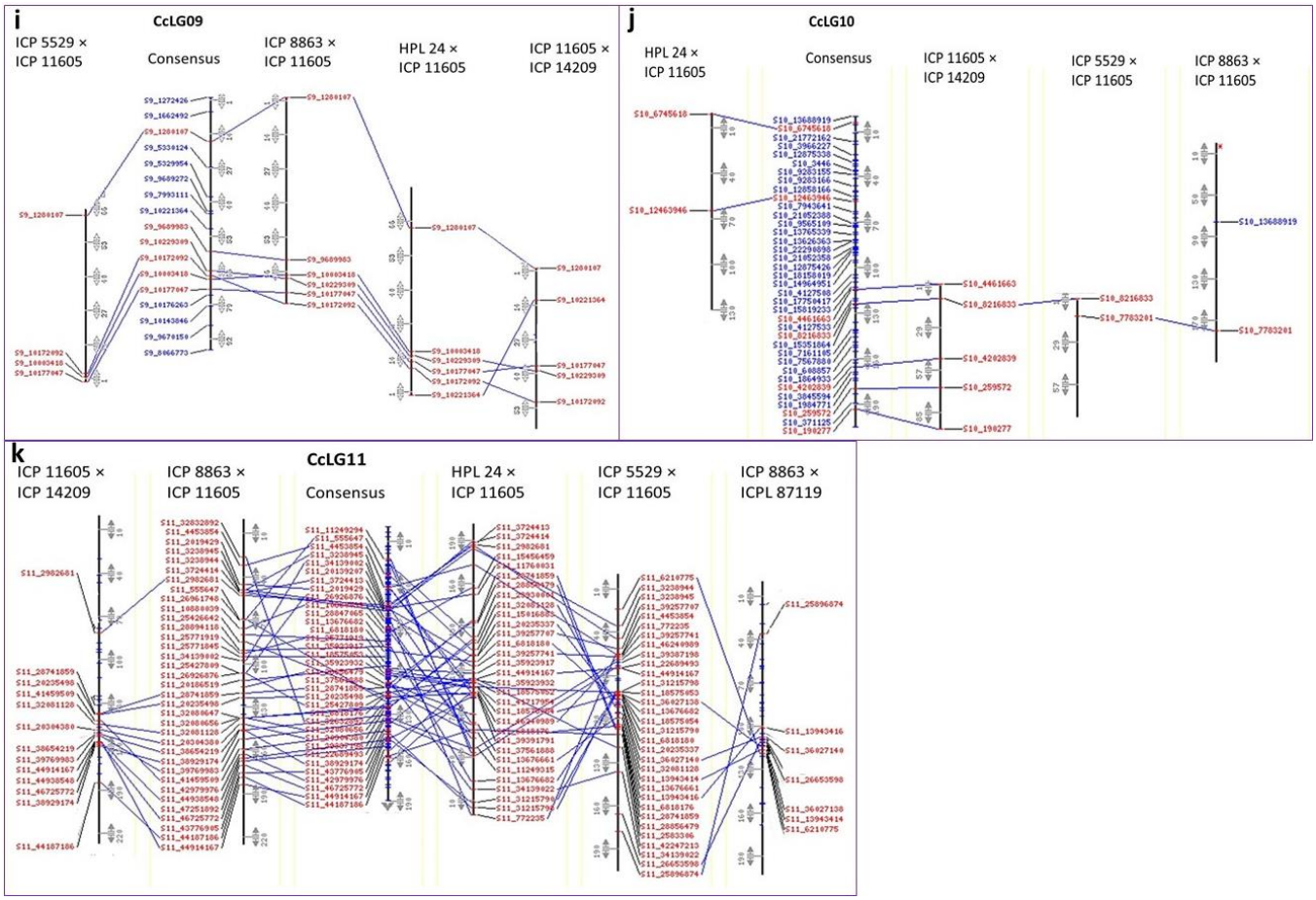
Appendix 9: Positions of markers and QTLs in linkage map of Pop5 (ICP 8863 × ICPL 87119). Markers are shown on right side of the linkage group while map distances are indicated on left side. QTLs for the different traits are indicated by bars with different colours with green and red showing QTLs for DTF and SPC, respectively.



Appendix 11: Charts depicting marker-based correspondences of consensus with individual genetic maps. a, b, c, d, e, f, g, h, i, j and k are CcLG01, CcLG02, CcLG03, CcLG04, CcLG05, CcLG06, CcLG07, CcLG08, CcL09, CcLG10 and CcLG11, respectively. Only common markers are included to visually asses the collinearity of marker orders and marker positions. Genetic linkage groups were aligned together using comparative mapping programme CMap version 1.01



Appendix 11: (continued)



Appendix 11 (continued)