**University of São Paulo
“Luiz de Queiroz” College of Agriculture**

**Dissecting the genetic architecture of quantitative traits in common bean
(*Phaseolus vulgaris* L.): response to root-knot nematode infection and seed
morphology**

Willian Giordani

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

**Piracicaba
2021**

Willian Giordani
Bachelor in Agronomy

Dissecting the genetic architecture of quantitative traits in common bean (*Phaseolus vulgaris* L.): response to root-knot nematode infection and seed morphology

Advisor:
Profª Drª **MARIA LUCIA CARNEIRO VIEIRA**

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Giordani, Willian

Dissecting the genetic architecture of quantitative traits in common bean (*Phaseolus vulgaris* L.): response to root-knot nematode infection and seed morphology / Willian Giordani. -- Piracicaba, 2021.

106 p.

Tese (Doutorado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. *Meloidogyne incognita* 2. Tamanho de sementes 3. Mapeamento associativo 4. Estudos de associação genômica 5. GWAS 6. Mapeamento genético 7. QTL I. Título

DEDICATORY

To my beloved parents, Dirceu and Maristela

ACKNOWLEDGEMENTS

Completing this thesis was challenging but still much more rewarding than I could ever have imagined. It has been a long and astonishing journey that I would not have been able to complete without the support of dozens of remarkable individuals and institutions, which I would like to acknowledge. Thus, it is with immense gratitude that I acknowledge the following:

The “Luiz de Queiroz” College of Agriculture, University of São Paulo (ESALQ/USP), and in particular the Department of Genetics and its entire staff for accepting me as their student, providing countless amazing opportunities and giving me solid academic assistance.

My advisor, Professor Maria Lucia Carneiro Vieira, for her ongoing support, guidance, confidence and attention over the last four years. I am extremely grateful to have had an advisor who not only provided me with all the conditions to complete this thesis, but also for her dedication to preparing me as a scientist. I could not have imagined having an advisor better able to bring out the best in me.

Professor Antonio Augusto Franco Garcia and Professor Roberto Fritsche-Neto, both of whom had an important academic influence on me, for all the valuable discussions and advice regarding both research and life. Their professional dedication served as a source of inspiration and their kind words during my qualification exam motivated me to give of my best and strive to attain their outstanding standards as scientists. I extend this acknowledgement to all the other amazing professors I was privileged to work with during my life and whom I will not be able to include here by name.

Carlos Alberto Oliveira and my fellow labmates from the Laboratory of Molecular Genetics of Crop Plants, for the stimulating conversations, the nights and weekends that we spent together in the lab and for all the unforgettable moments we have shared during the last four years.

The Genetics and Plant Breeding Group “Professor Roland Vencovsky” (GVENCK) and Professor José Baldin Pinheiro for all the exceptional opportunities and for giving me the chance to work with a team of talented people and develop other aspects of my career.

Dr. Alisson Fernando Chiorato and all the staff and friends from Agronomic Institute of Campinas (IAC) for agreeing to collaborate on this work, providing the structure to carry out the experiments, welcoming me to their team and being willing to help whenever was necessary.

Professor Heqiang Huo (Alfred), from the University of Florida, for receiving me in his lab and agreeing to collaborate on this research. His contribution to the drafts were extremely valuable.

My beloved parents, Dirceu Antônio Giordani and Maristela Mara Mertz Giordani, two singular human beings who deeply believe in the value of education and have devoted their lives to providing their son with every opportunity to pursue his dreams. For their boundless love, support, integrity, trust and encouragement throughout my life.

My lovely girlfriend Marina Carvalho Peruzzolo, such a special woman, who has been truly and wonderfully supportive and has loved me unconditionally. She has been my rock and my fortress, and I simply could not have made it through without her continued encouragement. She has shared this entire

journey with me through the peaks and troughs and has constantly shown huge faith in my success. I will always be thankful for all her patience, affection and companionship, which have made me a better person.

My friends, Carolina, Emanuel, Elesandro, Gabriel, Getúlio, Gleicy, Henrique, Jéssica, João Vitor, Maiara, Patrícia, Pedro, Rafael, and Zirlane for making life in Piracicaba happier and problems easier to solve. I have been blessed to be part of their lives and grow with them. I will never forget their willingness to help me whenever I needed them.

My greatest role models, Liliane Marcia Mertz Henning and Fernando Augusto Henning, exemplary in their ethics, tenacity, commitment, resilience and wisdom, and an inspiration to me since I was a child. Thanks for the guidance on all the key decisions I have taken and for being present in every moment.

Finally, I would like to acknowledge the financial support given by the São Paulo Research Foundation (FAPESP, grant number 2018/09069-7); the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant number 454329/2014-8 and 140588/2018-3) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, finance code 001).

SUMMARY

RESUMO.....	8
ABSTRACT.....	9
LIST OF FIGURES.....	10
LIST OF TABLES	11
1. INTRODUCTION	13
References.....	15
2. GENETIC MAPPING REVEALS COMPLEX ARCHITECTURE AND CANDIDATE GENES INVOLVED IN COMMON BEAN RESPONSE TO MELOIDOGYNE INCOGNITA INFECTION.....	19
ABSTRACT.....	19
2.1. Introduction.....	19
2.2. Materials and methods	21
2.2.1. Plant material	21
2.2.2. Phenotypic data	22
2.2.3. Genotypic data.....	23
2.2.4. GWAS.....	24
2.2.5. QTL mapping.....	24
2.2.6. Searching for candidate genes	25
2.2.7. Histochemical analysis of H ₂ O ₂	26
2.3. Results	27
2.3.1. Phenotypic analysis – IAC panel	27
2.3.2. Phenotypic analysis – BC ₂ F ₄ Population.....	29
2.3.3. Genome-Wide Association Studies.....	30
2.3.4. Validation by QTL mapping	32
2.3.5. Candidate genes.....	32
2.3.6. Histochemical detection of H ₂ O ₂	36
2.4. Discussion.....	37
References	41
3. GENOME-WIDE ASSOCIATION STUDIES DISSECT THE GENETIC ARCHITECTURE OF SEED SHAPE AND SIZE IN COMMON BEAN	47
ABSTRACT.....	47
3.1. Introduction.....	47
3.2. Materials and methods	49

3.2.1. Plant material and growing conditions.....	49
3.2.2. Image acquisition and processing	50
3.2.3. Phenotypic data analysis	51
3.2.4. Genotypic data.....	52
3.2.5. GWAS and candidate genes search	52
3.3. Results.....	53
3.3.1. Phenotypic analysis	53
3.3.2. Genotypic correlation and principal component analysis	56
3.3.3. Genome-Wide Association Studies	58
3.3.4. Candidate genes	60
3.4. Discussion	62
References	66
4. CONCLUDING REMARKS	73
APPENDICES	75

RESUMO

Dissecando a arquitetura genética de características quantitativas em feijoeiro (*Phaseolus vulgaris* L.): resposta a infecção por nematoide de galhas e morfologia de sementes

O feijão comum (*Phaseolus vulgaris* L.) é o mais importante grão de espécies leguminosas para o consumo humano e um alimento básico na dieta de populações em diversas regiões do planeta. Uma vez que a maior parte do cultivo ocorre nos trópicos, sob baixo nível de investimento, problemas fitossanitários costumam causar severos impactos sobre a produtividade do feijoeiro. Entre os principais patógenos que atacam *Phaseolus* spp. estão os nematoides das galhas (RKN), particularmente da espécie *Meloidogyne incognita*, que são tidos como os mais dispersos e impactantes. Apesar da importância, ainda não existem estratégias efetivas para o controle de RKN e, devido à imprecisão e morosidade das abordagens de fenotipagem tradicionais, os programas de melhoramento do feijoeiro têm enfrentado enormes desafios para avaliar a resistência a nematoides nas atividades de rotina. Além disso, outro gargalo enfrentado pelos melhoristas é a necessidade de buscar características específicas que atendam não só os agricultores, mas também a demanda da indústria e dos consumidores finais, como ocorre para tamanho, peso e forma das sementes. Com o advento e a redução no custo do sequenciamento de nova geração, as abordagens de mapeamento genético, como a análise de locos envolvidos em caracteres quantitativos (QTL) e estudos de associação genômica (GWAS), se consolidaram como métodos proeminentes para estudar a arquitetura genética de características quantitativas. O Laboratório de Genética Molecular de Plantas Cultivadas tem como uma de suas principais atividades de pesquisa o desenvolvimento de estudos fundamentais sobre diversidade e mapeamento genético em *P. vulgaris*. Para isso, foi desenvolvido um painel de uma coleção de 180 acessos que foram genotipados com 10.362 SNPs e recomendados para uso em análises de mapeamento associativo. Nesta tese, duas investigações distintas são descritas, cada uma detalhando um estudo individual, mas que convergem para atingir o objetivo geral de expandir a compreensão sobre os fatores genéticos envolvidos no controle de características de importância para o feijoeiro. O primeiro estudo combina duas abordagens de mapeamento genético (GWAS e mapeamento de QTLs), bioinformática e análise histoquímica para dissecar a complexa arquitetura genética da resposta do feijão comum ao RKN. O GWAS identificou oito regiões genômicas independentes associadas ao número de massas de ovos de RKN (Pv06, Pv07, Pv08 e Pv11) e índice de galhas (Pv01, Pv02, Pv05 e Pv10), além de indicar vários genes como promissores candidatos para o envolvimento nos mecanismos de resistência do feijoeiro. No segundo estudo, GWAS e fenotipagem de alto rendimento baseada em imagens foram aplicados para localizar locos associados à morfologia de sementes de feijão. Sete associações explicando considerável porção da variabilidade fenotípica foram identificadas para comprimento, largura, área projetada, perímetro e circularidade das sementes, em quatro regiões genômicas distintas dos cromossomos Pv02, Pv08 e Pv11, e pelo menos 13 genes foram considerados candidatos promissores para regular a morfologia das sementes. Do ponto de vista prático, os estudos documentados nesta tese fornecem marcadores SNP que podem auxiliar os melhoristas na seleção assistida por marcadores, bem como os genes aqui sugeridos são proeminentes candidatos para uso em análises funcionais em feijoeiro ou eventualmente outras culturas relacionadas.

Palavras-chave: *Meloidogyne incognita*, Tamanho de sementes, Mapeamento associativo, Estudos de associação genômica, GWAS, Mapeamento genético, QTL

ABSTRACT

Dissecting the genetic architecture of quantitative traits in common bean (*Phaseolus vulgaris* L.): response to root-knot nematode infection and seed morphology

Common bean (*Phaseolus vulgaris* L.) is the leading legume grain for human consumption and a staple food in many regions of the world. Since most common bean is produced in the tropics under low-input systems, phytosanitary issues often severely impact yield. Among the principal *Phaseolus* spp. pathogens, root-knot nematodes (RKN), particularly *Meloidogyne incognita*, are probably the most widespread and damaging. Despite their importance, effective control is still not available and due to the imprecision and slowness of the traditional phenotyping procedures, common bean breeding programs have faced huge challenges in developing routine nematode resistance screening. Likewise, another complex task tackled by bean breeders is the need to focus on specific traits to satisfy the requirements of farmers, industry and end consumers. This applies to seed size, weight and shape. With the advent of low-cost next-generation sequencing, genetic mapping approaches such as quantitative trait loci (QTL) analysis and genome-wide association studies (GWAS) have proved to be remarkable approaches to quantitative trait genetic architecture studies. The Laboratory of Molecular Genetics of Crop Plants has, as one of its main research activities, the development of fundamental studies on diversity and genetic mapping in *P. vulgaris*. This has entailed developing a core collection panel of 180 accessions, genotyped using 10,362 SNPs and recommended for use in association mapping. In this thesis, two distinct investigations are described, each detailing an individual study, converging to achieve the overarching goal of expanding the comprehension about the genetic factors underlying common bean traits. The first study combines two genetic mapping approaches (GWAS and QTL mapping), bioinformatics and histochemical analysis to dissect the complex genetic architecture of the common bean response to RKN. GWAS identified independent genomic regions associated with the number of RKN egg masses (Pv06, Pv07, Pv08 and Pv11) and the root-galling index (Pv01, Pv02, Pv05 and Pv10), and several genes were highlighted as prominent candidates in common bean response to the pathogen. In the second study, GWAS and high-throughput image-based phenotyping were applied to pinpoint loci associated with common bean seed morphology. Seven marker-trait associations, explaining a considerable amount of phenotypic variation, were discovered for seed length, width, projected area, perimeter and circularity in four distinct genomic regions of chromosomes Pv02, Pv08 and Pv11, and at least 13 genes were considered promising candidates for regulating seed morphology. From a practical standpoint, the studies documented in this thesis provide SNP markers that may help breeders in marker-assisted selection, and the genes suggested herein are promising candidates for functional analysis in common bean and eventually other related crops.

Keywords: *Meloidogyne incognita*, Seed size, Association mapping, Genome-wide association studies, GWAS, Genetic mapping, QTL

LIST OF FIGURES

- Figure 1.** Frequency distribution of the number of egg masses (a) and root-galling index (b) for a common bean core collection (IAC panel) inoculated with *M. incognita*28
- Figure 2.** Caterpillar plot of the adjusted means with 95% confidence intervals for the number of egg masses (a) and root-galling index (b) in a BC₂F₄ population (green) and its donor (AND-277; dark red) and recurrent genitors (iac-milênio; pink) inoculated with *M. incognita*. Two resistant (blue) and two susceptible (red) controls are also shown30
- Figure 3.** GWAS of number of egg masses and root-galling index for the IAC panel inoculated with *M. incognita*. In the manhattan plots (a, c), the red dashed line indicates the bonferroni threshold ($p < 4.82e-06$) and the black dashed line denotes the permutation test threshold ($p < 2.29e-05$ for egg masses and $p < 1.50e-05$ for root-galling index). In the quantile-quantile plots (b, d) the shadow area denotes the 95% confidence interval.....31
- Figure 4.** Singular enrichment analysis of gene ontology (GO) for genomic regions associated with the number of egg masses (blue; a) and root-galling index (blue; b) and their representation in the complete genome annotation background (orange; a, b)34
- Figure 5.** Histochemical detection of hydrogen peroxide (H₂O₂) in roots of common bean genotypes stained over 12 h with 3,3'-diaminobenzidine (DAB). The mock column shows non-inoculated roots and the RKN column roots 24 h after inoculation. Based on the phenotypic analysis, Puebla-152-CIAT and IAC-Tybatã had the lowest number of egg masses while Jamapa-CNF-1671 and TB-01-13 exhibited the highest values. IPR Saracura showed an intermediate response.....37
- Figure 6.** High-throughput image-based phenotyping of common bean seed size and shape. Representation of the phenotypic diversity of the IAC core collection panel (a). Top (a) and front view (b) of the image acquisition station. Images of the seeds on a white background with a crossmark of 10 mm (d). Image after processing by the SmartGrain software (e). Single seed being measured for shape parameters (f)50
- Figure 7.** Frequency distribution of seven seed traits, namely seed length (a; LENGTH); width (b; WIDTH); perimeter (c; PERIM); projected area (d; AREA); length-to-width ratio (e; LWR); circularity (f; CIRC) and hundred seed mass (g; hsm) for the 180 genotypes of the iac core collection panel.....54
- Figure 8.** Boxplot of seven seed traits, namely seed length (a; LENGTH); width (b; WIDTH); perimeter (c; PERIM); projected area (d; AREA); length-to-width ratio (e; LWR); circularity (f; CIRC) and hundred seed mass (g; HSM), in 180 genotypes of the Andean and Mesoamerican gene pools56
- Figure 9.** Heatmap of pairwise pearson correlation coefficients (a) and network plot (b) of seven seed traits: seed length (LENGTH); width (WIDTH); perimeter (PERIM); projected area (AREA); length-to-width ratio (LWR); circularity (CIRC) and hundred seed mass (HSM) in the 180 genotypes of the IAC core collection panel57
- Figure 10.** Principal component analysis (PCA) of the IAC common bean core collection panel based on the phenotypic variability of seven seed traits: seed length (LENGTH); width (WIDTH); perimeter (PERIM); projected area (AREA); length-to-width ratio (LWR); circularity (CIRC) and hundred seed mass (HSM). Orange and blue dots indicate accessions of Andean and Mesoamerican origin, respectively.....58
- Figure 11.** GWAS on seven common bean seed traits for the 180 genotypes of the IAC core collection panel. Seed length (LENGTH) is shown in red; width (WIDTH) in green; perimeter (PERIM) in blue; projected area (AREA) in light blue; length-to-width ratio (LWR) in pink; circularity (CIRC) in yellow and hundred seed mass (HSM) in grey. The outer ring in the circular manhattan plot (a) and the heatmap bars in the joint manhattan plot (c) depict SNP density along the eleven *P. vulgaris* chromosomes, and the dashed line indicates the Bonferroni threshold ($p < 4.82e-06$). In the quantile-quantile (QQ) plot (b) the shadow area indicates a 95% confidence interval59

LIST OF TABLES

Table 1.	Descriptive summary and variance components of number of egg masses and root-galling index in a common bean core collection and a BC ₂ F ₄ population inoculated with <i>M. incognita</i>	27
Table 2.	Likelihood ratio test (LRT) and Wald test of random and fixed effects for the number of egg masses and root-galling index of two common bean genetic resources inoculated with <i>M. incognita</i>	28
Table 3.	Single-nucleotide polymorphism (SNP) significantly associated with the number of egg masses and root-galling index, according to the GWAS results for IAC panel genotypes inoculated with <i>M. incognita</i>	32
Table 4.	QTL identified for root-galling index in a BC ₂ F ₄ population inoculated with <i>M. incognita</i> ...	32
Table 5.	Candidate genes identified by the GWAS-associated genomic regions and differentially expressed in common bean roots at 4 and 10 days after inoculation with <i>M. incognita</i> according to Santini et al. (2016).....	35
Table 6.	Resistance gene analogs (RGA) predicted in the GWAS-associated genomic regions, their domains, functional annotation and genomic location	36
Table 7.	Descriptive statistics and variance components, genetic coefficient of variation and broad-sense heritability of seven seed traits in the iac common bean core collection panel.....	55
Table 8.	Likelihood ratio test (LRT) and Wald test of random and fixed effects for seven seed traits in the iac common bean core collection panel.....	55
Table 9.	Single-nucleotide polymorphisms (SNP) significantly associated with seed length (length), width (WIDTH), area (AREA), perimeter (PERIM) and circularity (CIRC), according to the GWAS results for the IAC core collection panel.	60
Table 10.	Candidate genes identified by the GWAS-associated genomic regions able to modulate seed shape and size in common bean	62

1. INTRODUCTION

In developed countries, consumer preference for healthy food is growing fast (Beverland 2014), leading to a transition towards an increase in the consumption of plant-based protein, to provide a good source of nutrition and respond to environmental concerns over meat production (Graça et al. 2019; Petrusán et al. 2016; Kumar et al. 2017). Similarly, protein resources of vegetable origin are of great importance in the developing world, but their appeal is justified mainly by the need to tackle undernourishment and to address demand boosted by significant population growth (Henchion et al. 2017). This scenario has highlighted pulse grains, such as beans, lentils, peas and chickpeas, and has generated great demand all around the world (Leterme 2002; Boye et al. 2010; Singh 2017), drawing the attention of well-known international organizations such as the United Nations (see UN 2014).

The Food and Agriculture Organization (FAO) describes pulses as annual legume crops harvested exceptionally for dry mature grain use and consumed directly in human or animal diets; definition that excludes those used for oil extraction, as soybean, and the ones consumed as green vegetable crops (FAO 1994). Rich in high-quality protein, energy, fiber and micronutrients (e.g. vitamins, iron, zinc), common bean (*Phaseolus vulgaris* L., $2n = 2x = 22$) is the main pulse crop for human consumption, and grown worldwide with current annual global production of around 30 million metric tons (FAOSTAT 2019). In several areas of the world (e.g. Latin America, Asia and Africa), the common bean crop is even more important, since beans supplement the deficiency of proteins from cereals, roots, and tubers, and are sometimes the only dietary protein nutrient (Phillips 1993; Gepts et al. 2008).

Brazil is a major consumer and producer of common bean, with an annual harvest of around 3 million metric tons per year (FAOSTAT 2019). Despite the expressive production, average yield has been around 1.0 t.ha⁻¹ (CONAB 2020), well below the yield potential of up to 5.0 t.ha⁻¹ (Graham and Ranalli 1997). The factors that affect common bean productivity vary according to the region and crop conditions, usually very heterogeneous. Most global cultivated area is in the tropics and the result of small-scale agriculture and low-input systems (Gepts et al. 2008). Due to poor management practices and the typically humid and hot weather, phytosanitary losses tend to be expressive (Graham and Ranalli 1997; Miklas et al. 2006).

As for the majority of crops, common bean breeding programs are primarily aimed at releasing high-yield cultivars with resistance to biotic and abiotic stresses (Miklas et al. 2006; Beaver and Osorno 2009; Singh and Schwartz 2010; Assefa et al. 2019). However, bean breeders face additional challenges since they need to pay close attention to the end consumers' needs (Kelly and Miklas 1998; Santalla et al. 2004; Beaver and Osorno 2009; Assefa et al. 2019). The success of a cultivar depends on good performance in terms of quality traits and market class features, such as cooking time, nutritional aspects, and seed size, shape and color (Nienhuis and Singh 1986; Voysest et al. 1994; Assefa et al. 2019). For instance, bigger seeds are popular on the principal markets in Latin America, Africa, and Asia (Nienhuis and Singh 1986), while seed shape is not only important to farmers and end-users, but is also a key commercial standard for the

processing industry (Kara et al. 2013; FAO and AfricaSeeds 2018; Shaw et al. 2020). Therefore, understanding the genetic architecture of key seed morphological aspects is a critical step towards the development of cultivars combining high yield, industrial compatibility and consumer acceptance (White and González 1990; Park et al. 2000; Blair et al. 2009; Lei et al. 2020).

Regarding disease attacks, root-knot nematodes (*Meloidogyne* spp.) are among the most damaging pathogens for many crops, including *P. vulgaris* (Trudgill and Blok 2001; Santos et al. 2012; Singh et al. 2013). They typically take hold in root tissues and modify a set of cells, inducing root swelling that originates the symptoms of galls or root-knots, therefore the common designation (Moens et al. 2009). The impact on yield is usually heavier in tropical rather than temperate areas, and therefore affects the main bean production regions (Moens et al. 2009; Beebe 2012; Assefa et al. 2019). The control initiatives to reduce nematode population size and mitigate plant damage involve management practices that are not always feasible under field conditions. In addition, nematicides are often inefficient, expensive and adversely impact the environment and human health (Nyczepir and Thomas 2009). Thus, finding resistant cultivars is considered the best control strategy, requiring breeding programs to develop resistant cultivars, along with studies aimed at dissecting the genes and genetic mechanisms involved in the host response (Singh and Schwartz 2011).

One of the main research programs of the Laboratory of Molecular Genetics of Crop Plants (LMGCP) at the “Luiz de Queiroz” College of Agriculture (ESALQ) is centered on fundamental studies of diversity and genetic mapping in *P. vulgaris* (Hanai et al. 2007, 2010; Diniz et al. 2014, 2019; Santini et al. 2016). For over a decade, the laboratory has been working in partnership on the Common Bean Breeding Program run by the Agronomic Institute of Campinas (IAC), responsible for the development of many cultivars currently used in Brazil (Chiorato et al. 2010). As a result of ESALQ - IAC cooperation, a core collection was proposed to contemplate the great genetic diversity of the IAC common bean germplasm, which was recommended for use in association mapping (Perseguini et al. 2015). The collection of 180 accessions was initially genotyped using a few number of microsatellite markers and single nucleotide polymorphisms (SNPs), which were used to analyze the linkage disequilibrium (LD) decay and identify loci associated with resistance to angular spot and anthracnose diseases (Perseguini 2013; Perseguini et al. 2016). Later, Diniz (2016) used this collection to pinpoint genomic regions associated with macro and micronutrient content in grains, as well as several agronomic traits. In that study, to increase the number of SNPs and better represent the genome, a new genotyping method was performed by genotyping by sequencing (GBS) approach. The procedures were realized at the Genomic Diversity Facility of Cornell University, USA. This genotypic data was recently used to measure the influence of population structure and kinship on common bean LD (Diniz et al. 2019).

Focusing specifically on the response to root-knot nematode infection, the LMGCP pioneered the use of transcriptomic data to scrutinize the *P. vulgaris* – *M. incognita* interaction. This involved building RNA-Seq libraries at 4 and 10 days after the inoculation of a moderately susceptible cultivar, allowing the detection of around 800 differentially expressed genes (Santini et al. 2016). In addition, although RNA-Seq studies are able to provide hints of genes and pathways underlying the host response to the pathogen, it

cannot be used to measure the effect of each gene on phenotypic variation. With the advent of low-cost next generation sequencing, genetic mapping approaches such as quantitative trait loci (QTL) analysis and genome-wide association studies (GWAS) provided remarkable, and complementary approaches to reveal DNA polymorphisms associated with phenotypes, facilitating genetic architecture studies on quantitative traits (Zargar et al. 2015).

This thesis is organized into two main parts, each detailing an individual study but that converge to achieve the overarching goal of elucidating the genetic factors underlying common bean traits. The first part describes a study that combined two genetic mapping strategies (QTL mapping and GWAS), bioinformatics and histochemical analysis to dissect the genetic architecture and identify genomic regions and candidate genes controlling the common bean response to *M. incognita*. In the second part, GWAS and high-throughput image-based phenotyping were applied to identify loci associated with the high variation in common bean seed morphology as well as suggest candidate genes putatively responsible for that variation.

References

- Assefa T, Mahama AA, Brown AV, et al (2019) A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). *Mol Breed* 39:1–23. <https://doi.org/10.1007/s11032-018-0920-0>
- Beaver JS, Osorno JM (2009) Achievements and limitations of contemporary common bean breeding using conventional and molecular approaches. *Euphytica* 168:145–175. <https://doi.org/10.1007/s10681-009-9911-x>
- Beebe S (2012) Common Bean Breeding in the Tropics. In: Junes, J (ed) *Plant Breeding Reviews*. Wiley, New York, pp 357–426. <https://doi.org/10.1002/9781118358566.ch5>
- Beverland MB (2014) Sustainable eating: mainstreaming plant-based diets in developed economies. *J Macromarketing* 34:369–382. <https://doi.org/10.1177/0276146714526410>
- Blair MW, Díaz LM, Buendía HF, Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). 6:955–972. <https://doi.org/10.1007/s00122-009-1064-8>
- Boye J, Zare F, Pletch A (2010) Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Res Int* 43:414–431. <https://doi.org/https://doi.org/10.1016/j.foodres.2009.09.003>
- Chiorato AF, Carbonell SAM, Vencovsky R, et al (2010) Genetic gain in the breeding program of common beans at IAC from 1989 to 2007. *Crop Breed Appl Biotechnol* 10:329–336. <http://dx.doi.org/10.1590/S1984-70332010000400007>
- CONAB (2021) Acompanhamento da safra brasileira grãos - Safra 2020/21. <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos> Accessed 07 February 2021.
- Diniz AL (2016) Regiões genômicas envolvidas no controle de caracteres agrônômicos e no teor de macro e micronutrientes em grãos de feijão comum, via mapeamento associativo. Tese, University of São Paulo - Escola Superior de Agricultura “Luiz de Queiroz”
- Diniz AL, Giordani W, Costa ZP, et al (2019) Evidence for strong kinship influence on the extent of linkage disequilibrium in cultivated. *Genes* 10:1–16. <https://doi.org/10.3390/genes10010005>
- Diniz AL, Zucchi MI, Santini L, et al (2014) Nucleotide diversity based on phaseolin and iron reductase genes in common bean accessions of different geographical origins. *Genome* 57:69–77. <http://dx.doi.org/10.1139/gen-2013-0183>
- FAO (1994) Definition and classification of commodities. *Defin Classif Commod* <http://www.fao.org/es/faodef/fdef04e.htm#4.01>. Accessed 10 Dec 2020
- FAO, AfricaSeeds (2018) Seeds Toolkit-Module 2: Seed quality assurance, 2nd edn. Food & Agriculture Org., Rome
- FAOSTAT (2019) Food and agriculture organization of the united nations statistics division. In: *Crop. Prod.* <http://www.fao.org/faostat/en/#data/QC>. Accessed 14 May 2020
- Gepts P, Aragão FJL, Barros E de, et al (2008) Genomics of phaseolus beans, a major source of dietary protein and micronutrients in the Tropics. In: Moore PH, Ming R (eds) *Genomics of Tropical Crop Plants*. Springer New York, New York, pp 113–143. https://doi.org/10.1007/978-0-387-71219-2_5

- Graham PH, Ranalli P (1997) Common bean (*Phaseolus vulgaris* L.). Field Crop Res 53:131–146. [https://doi.org/10.1016/S0378-4290\(97\)00112-3](https://doi.org/10.1016/S0378-4290(97)00112-3)
- Hanai L, Santini L, Camargo L, et al (2010) Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. Mol Breed 25:25–45. <https://doi.org/10.1007/s11032-009-9306-7>
- Hanai LR, De Campos T, Camargo LEA, et al (2007) Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. Genome 50:266–277. <https://doi.org/10.1139/g07-007>
- Henchion M, Hayes M, Mullen A, et al (2017) Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. Foods 6:53–74. <https://doi.org/10.3390/foods6070053>
- Kara M, Sayinci B, Elkoca E, et al (2013) Seed size and shape analysis of registered common bean (*Phaseolus vulgaris* L.) cultivars in turkey using digital photography. J Agric Sci 19:219–234
- Kelly JD, Miklas PN (1998) The role of RAPD markers in breeding for disease resistance in common bean. Mol Breed 4:1–11. <https://doi.org/10.1023/A:1009612002144>
- Kumar P, Chatli MK, Mehta N, et al (2017) Meat analogues: Health promising sustainable meat substitutes. Crit Rev Food Sci Nutr 57:923–932. <https://doi.org/10.1080/10408398.2014.939739>
- Lei L, Wang L, Wang S, Wu J (2020) Marker-trait association analysis of seed traits in accessions of common bean (*Phaseolus vulgaris* L.) in China. Front Genet 11:698–710. <https://doi.org/10.3389/fgene.2020.00698>
- Leterme P (2002) Recommendations by health organizations for pulse consumption. Br J Nutr 88:239–242. <https://doi.org/10.1079/BJN2002712>
- Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: From classical to MAS breeding. Euphytica 147:105–131. <https://doi.org/10.1007/s10681-006-4600-5>
- Moens M, Perry RN, Starr JL (2009) *Meloidogyne* species—a diverse group of novel and important plant parasites. In: Perry RN, Moens M, Starr JL (ed) Root-knot nematodes, 1st edn. CABI International, Cambridge, MA, USA, p 483. <https://doi.org/10.1079/9781845934927.0001>
- Nienhuis J, Singh SP (1986) Combining ability analyses and relationships among yield, yield components, and architectural traits in dry bean. Crop breeding genetics & cytology 1:21–27
- Nyczepir AP, Thomas SH (2009) Management strategies in resource-poor farming. In: Perry RN, Moens M, Starr JL (eds) Root-knot nematodes, 1st edn. CABI International, p 444
- Park SO, Coyne DP, Jung G, et al (2000) Mapping of qtl for seed size and shape traits in common bean. J Amer Soc Hort Sci 125:466–475. <https://doi.org/10.21273/JASHS.125.4.466>
- Perseguini JMCK, Silva GSB, Rosa JRBF, et al (2015) Developing a common bean core collection suitable for association mapping studies. Genet Mol Biol 38:67–78. <https://doi.org/10.1590/S1415-475738120140126>
- Perseguini JMCK (2013) Estudo do desequilíbrio de ligação em *Phaseolus vulgaris* L. visando o mapeamento associativo de fatores bióticos e abióticos. Tese, Universidade Estadual de Campinas
- Perseguini JMCK, Oblessuc PR, Rosa JRBF, et al (2016) Genome-wide association studies of anthracnose and angular leaf spot resistance in common bean (*Phaseolus vulgaris* L.). PLoS One 11:1–19. <https://doi.org/10.1371/journal.pone.0150506>
- Petrusán JI, Rawel H, Huschek G (2016) Protein-rich vegetal sources and trends in human nutrition: A review. Curr Top Pept Protein Res 17:1–19
- Phillips RD (1993) Starchy legumes in human nutrition, health and culture. Plant Foods Hum Nutr 44:195–211. <https://doi.org/10.1007/BF01088314>
- Santalla M, Monteagudo AB, González AM, De Ron AM (2004) Agronomical and quality traits of runner bean germplasm and implications for breeding. Euphytica 135:205–215. <https://doi.org/10.1023/B:EUPH.0000014912.07993.e7>
- Santini L, Munhoz CDF, Bonfim-Jr MF, et al (2016) Host transcriptional profiling at early and later stages of the compatible interaction between *Phaseolus vulgaris* and *Meloidogyne incognita*. Nematology 106:282–294. <https://doi.org/10.1094/PHYTO-07-15-0160-R>
- Santos LNS, Alves FR, Belan LL, et al (2012) Damage quantification and reaction of bean genotypes (*Phaseolus vulgaris* L.) to *Meloidogyne incognita* race 3 and *M. javanica*. Summa Phytopathol 38:24–29. <https://doi.org/10.1590/S0100-54052012000100004>
- Shaw N, Barak RS, Campbell RE, et al (2020) Seed use in the field: Delivering seeds for restoration success. Restor Ecol 28:276–285. <https://doi.org/10.1111/rec.13210>
- Singh N (2017) Pulses: an overview. J Food Sci Technol 54:853–857. <https://doi.org/10.1007/s13197-017-2537-4>
- Singh SK, Hodda M, Ash GJ (2013) Plant-parasitic nematodes of potential phytosanitary importance, their main hosts and reported yield losses. Eppo Bull 43:334–374. <https://doi.org/10.1111/epp.12050>
- Singh SP, Schwartz HF (2010) Breeding common bean for resistance to diseases : A review. Crop Sci 50:2199–2223. <https://doi.org/10.2135/cropsci2009.03.0163>
- Singh SP, Schwartz HF (2011) Review: Breeding common bean for resistance to insect pests and nematodes. Can J plant Sci 91:239–250. <https://doi.org/10.4141/CJPS10002>
- Trudgill DL, Blok VC (2001) Apomictic, polyphagous, root-knot nematodes: Exceptionally successful and damaging biotrophic root pathogens. Annu Rev Phytopathol 39:53–77. <https://doi.org/0066-4286/01/0901-0053>
- UN (2014) Resolution adopted by the general assembly on 20 December 2013, 68/231. General Assembly.

- Voysest O, Valencia MC, Amezquita MC (1994) Genetic diversity among Latin american andean and mesoamerican common bean cultivars. *Crop Sci* 34: 1100-1110. <https://doi.org/https://doi.org/10.2135/cropsci1994.0011183X003400040049x>
- White JW, Gonçalves A (1990) Characterization of the negative association between seed yield and seed size among genotypes of common bean. *Fields Crop Res* 23:159–175. [https://doi.org/10.1016/0378-4290\(90\)90052-D](https://doi.org/10.1016/0378-4290(90)90052-D)
- Zargar SM, Raatz B, Sonah H, et al (2015) Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *J Crop Sci Biotechnol* 18:293–308. <https://doi.org/10.1007/s12892-015-0037-5>

2. GENETIC MAPPING REVEALS COMPLEX ARCHITECTURE AND CANDIDATE GENES INVOLVED IN COMMON BEAN RESPONSE TO *Meloidogyne incognita* INFECTION

ABSTRACT

Root-knot nematodes (RKN), particularly *Meloidogyne incognita*, are among the most damaging and prevalent agricultural pathogens due to their ability to infect roots of almost all crop species, including common bean. The best strategy for their control is through the use of resistant cultivars. However, laborious phenotyping procedures make it difficult to assess nematode resistance in breeding programs. For common bean, this task is especially challenging since little has been done to discover resistance genes or find markers to assist selection. In this study, we performed genome-wide association studies and QTL mapping to explore the genetic architecture and genomic regions underlying the resistance to *M. incognita* and to identify candidate resistance genes. Phenotypic data were collected by a high-throughput assay, and the number of egg masses and root-galling index were evaluated 30 days after inoculation. Complex genetic architecture and independent genomic regions were associated with each trait according to the Fixed and random model Circulating Probability Unification. SNPs located on chromosomes Pv06, Pv07, Pv08 and Pv11 were associated with the number of egg masses, and on Pv01, Pv02, Pv05 and Pv10 with root-galling. A total of 215 candidate genes were identified, including 14 resistance gene analogs and five differentially expressed in a previous RNA-seq analysis. The histochemical analysis indicated that the reactive oxygen species might play a role in the resistance response. Our findings open new perspectives to improve selection efficiency for RKN resistance in common bean, and the candidate genes are valuable targets for functional investigation and gene editing approaches.

Keywords: Root-knot nematodes; *Phaseolus vulgaris* L.; GWAS; QTL mapping; Resistance gene analogs; Reactive oxygen species.

2.1. Introduction

The common bean (*Phaseolus vulgaris* L.) is the main legume grain consumed directly by humans (Broughton et al. 2003; Bellucci et al. 2014), and global annual production stands at 31 million metric tons (Faostat 2018). It is currently a staple food for more than 300 million people in Eastern Africa and Latin America, accounting for up to 65% of protein and 32% of energy intake (Gepts et al. 2008; Blair et al. 2010; Petry et al. 2015).

The evolutionary history and domestication process of *P. vulgaris* are particularly unique and well-studied. Originating in Mexico (Bitocchi et al. 2012, 2013), the species spread to South America, resulting in the formation of two gene pools: the Mesoamerican and the Andean (Gepts et al. 1986; Koinange and Gepts 1992). Domestication then occurred in at least two independent events that have been defined according to the seed phaseolin type (Gepts et al. 1986), allozymes (Koenig and Gepts 1989; Singh et al. 1991), plant morphology (Gepts and Debouck 1991; Singh et al. 1991), chloroplast (Chacón et al. 2005) and nuclear DNA polymorphisms (Hanai et al. 2007; Kwak and Gepts 2009; Rossi et al. 2009; Bitocchi et al.

2013; Rodriguez et al. 2016). Importantly, these findings provided a knowledge base helping in the conservation and use of the available germplasm (Bitocchi et al. 2017).

Root-knot nematodes (RKN; *Meloidogyne* spp.) are among the most devastating pathogens for many crop species, including common bean. Among more than 90 species, *M. incognita* is extremely widespread, especially in tropical regions. It is considered perhaps the single most damaging pathogen for global agriculture (Trudgill and Blok 2001). As biotrophic pathogens, RKN feed only on living cells, requiring an intimate relationship with their hosts. After invading the root vascular cylinder, second-stage juveniles (J2s) induce root cell differentiation into specialized, giant nurse cells, engendering the typical symptom of root-galling (Abad et al. 2009).

Although they rarely lead to plant death, attacks by RKN cause severe damage to the root system, resulting in reduced nutrient and water uptake and symptoms of stunted growth, wilting, and chlorosis (Lilley et al. 2011). Additionally, the injury caused by nematode infection might increase plant susceptibility to other biotic and abiotic stresses (Caromel and Gebhardt 2011). As a consequence, yield loss in susceptible crops can exceed 20%, especially in developing countries where subsistence farmers crop common bean under low-input systems and can suffer yield losses of up to 90% (Atkinson et al. 1995; Koenning et al. 1999; Abawi et al. 2005).

RKN control strategies are usually impracticable and ineffective under field conditions. The nematode attacks a wide range of hosts, and therefore practices such as crop rotation are not effective methods for suppressing *Meloidogyne* spp. (Nyczepir and Thomas 2009; Coyne and Affokpon 2018). Furthermore, the development of an alternative control approach has become even more pressing due to the ban on the use of nematicides in the interests of environmental protection (Lilley et al. 2011). Thus, the efforts of breeders to develop resistant cultivars are essential from economical, practical and environmental perspectives.

Anyway, the complex and laborious phenotyping procedures have made it difficult to consolidate an effective and reliable protocol for assessing nematode resistance in breeding programs. Over the past few years several common bean genotypes have been evaluated, but none has proved fully resistant to *M. incognita* (Alves et al. 2011; Santos et al. 2012; Bozbuga et al. 2015; Costa et al. 2019). Moreover, in contrast to what has been achieved in the Solanaceae, where some *Meloidogyne* resistance genes are extensively used by breeders and farmers (Dropkin 1969; Messeguer et al. 1991; Yaghoobi et al. 1995; Williamson 1998), for most crop species, the identification and functional characterization of the genetic factors involved into controlling RKN resistance remain elusive, hence the importance of conducting studies aimed at dissecting the genetic architecture of host responses and identifying the genes underlying the resistance (Singh and Schwartz 2011).

In one of the pioneering efforts towards elucidating the *P. vulgaris*–*M. incognita* interaction, Santini et al. (2015) carried out a transcriptome analysis of a moderately susceptible cultivar at 4 and 10 days after inoculation (DAI). The authors reported 797 root differentially expressed genes, evidencing the complex common bean response mechanisms. Although RNA-seq technology can provide insights regarding the

genes and pathways involved in the host response to the pathogen (Westermann et al. 2012), it does not allow measurement of the effect of each gene on the phenotypic variation, which is essential for breeding.

Since the development of genetic markers based on DNA, genes and quantitative trait loci (QTL) involved in the response to root-knot nematodes have been mapped in different plant species (Caromel and Gebhardt 2011). In the traditional QTL mapping approaches, DNA polymorphisms are identified in a biparental segregating population, and the recombination generated in each meiotic event allows detection of loci segregating jointly with the phenotypic variation of a given trait. The advent of low-cost next generation sequencing methods has given rise to genome-wide association studies (GWAS), also known as association mapping, as powerful tools for identifying genotype-phenotype associations. Since it is not necessary to create a segregating population and due to its higher resolution, this technique has been successfully applied in plant breeding, dissecting the complex architecture of quantitative traits as well as identifying genomic regions underlying phenotypic variation (Zhu et al. 2008; Korte and Farlow 2013; Huang and Han 2014).

In this study, we performed GWAS on a core collection to dissect the genetic architecture and identify genomic regions and candidate genes underlying the common bean response to *M. incognita* infection. Additionally, QTL mapping was conducted in a biparental BC₂F₄ population with the aim of validating the SNP-trait associations. Finally, we carried out preliminary histochemical analysis to indicate the potential involvement of reactive oxygen species (ROS) in the resistance response.

2.2. Materials and methods

2.2.1. Plant material

The GWAS were based on a panel of 180 common bean genotypes representative of the genetic diversity in the common bean germplasm repository (1800 accessions) at the Agronomic Institute of Campinas (IAC), Brazil (Perseguini et al. 2015). The IAC common bean gene bank is one of the largest in Brazil and its core collection panel is renowned for its variability in terms of seed morphology, resistance to biotic and abiotic stresses and grain nutrient composition (Perseguini et al. 2015; Diniz et al. 2019). Still, it was previously classified according to the institution of origin, type of phaseolin (27 T-type for Andean and 153 S-type for Mesoamerican gene pools), seed size and commercial group (Diniz et al. 2019).

For validation by QTL mapping, we used a BC₂F₄ population of 91 individuals. This population was previously generated by crossing donor accession AND-277 and recurrent cultivar IAC-Milênio. After two backcrosses, the plants were self-pollinated for three generations to obtain a set of BC₂F₄ lines. Originally designed to map QTLs to the angular leaf spot pathogen (*Pseudocercospora griseola*), the phenotypic contrast between the genitors in terms of the number of egg masses and root-galling index (Appendix A) suggested that they could be used for QTL analysis related to the common bean reaction to *M. incognita*.

2.2.2. Phenotypic data

Phenotypic data were collected by a throughput assay, in the absence of soil, using the seedling growth pouches method described in Atamian et al. (2012) with minor adaptations. Briefly, seeds were pre-germinated in a biochemical oxygen demand (BOD) incubator at 28 °C for 6 days and the seedlings transferred to plastic zipper bags (24 x 17 cm) containing a germination paper. The pouches were placed in a controlled-environment chamber in an upright position and kept at 25-28 °C under a 16/8 (light/dark) photoperiod. The plants were watered daily with autoclaved distilled water and inoculated with 1,500 J2s when the tertiary root tips emerged.

The population of *M. incognita* race 3 was initially collected from a cotton field in the Midwest region of Brazil (15°33'6"S 55°10'0"W), providing an inoculum bank. The inoculum was multiplied in infected roots of a susceptible line of tomato (*Solanum lycopersicum* - Santa ClaraVF5600, Sakata) that were grown under greenhouse conditions for 60 days. To extract nematode eggs, the tomato roots were well-washed, cut and shaken for 2 min in 500 mL of 10% NaOCl solution. The solution was poured through a set of 425, 90 and 25 µm sieves in order to retain the eggs. After removing the excess NaOCl with distilled water, the eggs were collected and placed onto double layer disposable wipes (Kimwipe, Sigma-Aldrich) fitted over a metal basket inside a Petri dish. After 5 to 8 days, the freshly hatched J2s that passed through the paper were collected. The juveniles were counted using a Peters slide under an optical microscope and the inoculation solution prepared by dilution, resulting in a final concentration of 300 J2s per mL.

To perform the inoculation, the plants were placed horizontally and the roots evenly infused with 1500 J2s in 5 mL of solution. After 24 h in the dark, the plants were moved to the growth chamber, being irrigated daily with water, and weekly with half-strength Hoagland's nutritive solution (Hoagland and Arnon 1950). Thirty DAI, each pouch was infused with 15 mL of erioglaucine dye (75 mg L⁻¹) for 12 h to facilitate counting the number of egg masses per root system under a stereoscope (10 x magnifier). Root-galling symptoms were indexed based on the Bridge and Page (1980) scale. The experiments were performed in randomized complete block design with five replicates for the IAC core collection panel (GWAS) and four replicates for the BC₂F₄ population (QTL mapping).

Linear mixed models were fitted to estimate the generalized measurement of heritability, find the adjusted means to map QTLs, obtain genetic correlations among traits, and rank genotypes for selection. For the IAC core collection panel, the phenotypic data acquired for number of egg masses and root-galling index were analyzed according to the following statistical model:

$$Y_{ijk} = \mu + \beta_j + \tau_{i(k)} + \rho_k + \varepsilon_{ijk}$$

where: Y_{ijk} is the phenotypic value of genotype i within gene pool k in block j ; μ is an intercept; β_j is the random effect of block j , where $\beta_j \sim N(0, \sigma_\beta^2)$; $\tau_{i(k)}$ is the random effect of genotype i within gene pool k , where $\tau_{i(k)} \sim N(0, \sigma_\tau^2)$; ρ_k is the fixed effect of gene pool k ; and ε_{ijk} is the experimental error associated with genotype i in block j within gene pool k , where $\varepsilon_{ijk} \sim N(0, \sigma^2)$. For the BC₂F₄ population, the model was similar but with no modeling for the gene pool effect.

The ASReml-R package (Butler et al. 2009) was used to carry out linear mixed-model analysis. Wald and likelihood ratio tests (LRT) were performed to verify the respective statistical significance of random and fixed effects. The best linear unbiased predictors (BLUP) obtained for the genotypic factor were used to calculate the genotypic correlations using Pearson's coefficient. Because of the unbalanced condition, that is typical when evaluating a large number of genotypes and repetitions, the generalized measurement of heritability was computed according to Cullis et al. (2006), using the following expression:

$$H_c^2 = 1 - \frac{vBLUP}{2\hat{\sigma}_\tau^2}$$

where: H_c^2 is the heritability; $vBLUP$ the variance of the average difference between two BLUPs, and $\hat{\sigma}_\tau^2$ the genetic variance. Coefficient of genetic variation (CV_g) for each trait was calculated as $CV_g = \frac{\sigma_\tau}{\bar{x}}$, where σ_τ is the square root of the estimated genetic variance and $\bar{x}$ the respective mean value.

2.2.3. Genotypic data

The IAC core collection panel was previously genotyped by the genotyping by sequencing (GBS) approach at the Genomic Diversity Facility of the Cornell University's Institute of Biotechnology, as described by Diniz et al. (2019). In brief, DNA extracted from young leaves was digested with *ApeKI* enzyme and barcoded adapters were ligated to digested fragments, which were pooled and amplified by PCR as described by Elshire et al. (2011). Single-end sequencing was performed with the HiSeq 2500 platform (Illumina, San Diego, CA, USA). The sequence reads with ~100nt from two DNA libraries are available in the GenBank database, BioSample SAMN05513252 and SAMN05513251, both included in BioProject PRJNA336556. SNP calling was performed using the TASSEL-GBS bioinformatics pipeline and the SNP data were curated based on minor allele frequency (MAF) ≥ 0.05 ; minimum coefficient of inbreeding of 0.9 and call rate of < 0.9 . SNPs. Heterozygous loci were set as missing data. Finally, a total of 10,362 SNPs were identified and can be accessed in Diniz et al. (2019).

The genotypic data used for QTL mapping were obtained from the BC₂F₃ generation. To do this, DNA was extracted from young leaves and converted into NExtRAD (Nextera-tagmented reductively-amplified DNA) libraries by SNPsaurus LLC (Eugene, Oregon, USA). The libraries were sequenced on the HiSeq4000 platform with 150 bp paired-end sequencing. Reads were demultiplexed and pre-processed by SNPsaurus, followed by assembly, mapping and SNP identification according to the standard workflow. SNPs were further filtered based on 4 x minimum covered, MAF ≥ 0.01 and a call rate < 0.8 using vcftools software v.0.1.15. Heterozygous loci were eliminated to allow the use of phenotypic data from advanced generations without altering the expected recombination frequency.

2.2.4. GWAS

To reveal SNPs associated with the number of egg masses and root-galling index, GWAS were conducted according to the Fixed and random model Circulating Probability Unification (FarmCPU; Liu et al. 2016) and implemented in the GAPIT - R package (Lipka et al. 2012). The strategy used by the FarmCPU model is to separate the Multiple Loci Linear Mixed Model (MLMM) into two iterative stages: a Fixed Effect Model (FEM) and a Random Effect Model (REM). In the FEM stage, each marker is tested and pseudo-QTNs (quantitative trait nucleotides) included as covariates to control false-positives and false-negatives. In the REM stage, multiple associated markers are used to obtain kinship and optimize the selection of pseudo-QTNs, preventing model over-fitting. In each iteration, the p-values of individual testing markers and multiple associated markers are unified (Liu et al. 2016). Thus, FEM and REM can be written as:

$$\text{FEM: } y_i = M_{i1}b_1 + M_{i2}b_2 + \dots + M_{it}b_t + S_{ij}d_j + \varepsilon_i$$

$$\text{REM: } y_i = u_i + \varepsilon_i$$

In both models, y_i is the trait adjusted mean for the i th common bean accession and ε_i corresponds to the residual $\sim N(0, \sigma_e^2)$, where N is a normal distribution. In the FEM, $M_{i1}, M_{i2}, \dots, M_{it}$ and $b_1, b_2, \dots, b_j$ are the i th and t th pseudo-QTNs corresponding respectively to genotype and effect (in the first interaction there was no pseudo-QTN); S_{ij} is the genotype of the i th common bean accession and the j th SNP marker, and d_j is the effect of the j th SNP marker. In the REM, u_i is the total genetic effect of the i th common bean accession, in which the variance and covariance matrix of the genetic effects is given by $G = 2K\sigma_a^2$, where K is the kinship matrix based on the pseudo-QTNs and σ_a^2 the genetic variance.

The associations were tested for the 10,362 SNP markers and the adjusted means of each trait were obtained as described in the previous phenotypic model, but treating genotype as a fixed effect to obtain the best linear unbiased estimators (BLUES) in order to avoid a double-shrinkage. The threshold for detecting significant SNPs was determined by Bonferroni multiple test correction ($p < 4.82e-06$) and by permutation tests using 1000 iterations, both at 5% of global significance for type I error.

The proportion of variance explained (PVE) by each SNP–trait association was estimated according to $VQTL = V_{\text{pheno}}$, where $VQTL = 2freq(1 - freq)effect^2$ and V_{pheno} is the variance of the adjusted means for each trait.

2.2.5. QTL mapping

To perform QTL mapping, a linkage map previously constructed for a BC₂F₃ population was used. Importantly, this map was built using the software OneMap v2.1.3 (Margarido et al. 2007), which was modified to fit the genetic structure of this specific population (Almeida et al. 2020; unpublished). In brief, markers were tested for the expected Mendelian frequencies (1/8:7/8), and those exhibiting segregation distortion were excluded. The *grid search* algorithm was adapted to compute the recombination fraction expected for the population. Linkage groups were defined according to the recombination fractions and

physical position on the reference genome. Ordination was established using the *compare*, *try* and *mds* algorithms in several runs. Markers were also manually adjusted based on heatmap analysis. After ordination, the recombination fractions were re-estimated using the Multipoint Hidden Markov models. We also attempted to insert SNPs with no chromosomal position using the *try* algorithm and in case of success, the Multipoint Hidden Markov models were recalculated. Finally, a map of 1,093 SNPs assigned to 11 linkage groups, totaling 1114 cM, was obtained (Appendix B).

QTL analysis of the number of egg masses and root-galling index was carried out using the R/qtl package (Broman et al. 2003). Initial interval mapping was performed with the *scanone* function, using Haley-Knott regression to find an adequate number of cofactors. Composite interval mapping (CIM; Zeng 1994) was then applied to scan for QTLs at a step size of 1 cM and window size of 25 cM using the *cim* function with 4 cofactors by both Haley-Knott and Expectation-Maximization (EM) regression algorithms. An empirical genome-wide threshold was determined based on the 90th percentile of 1,000 permutations for each trait (Churchill and Doerge 1994). To estimate the PVE for significant QTLs and the allelic effects, analysis of variance was fitted to the data using the *fitqtl* function.

2.2.6. Searching for candidate genes

The search for host candidate genes putatively involved in the response to nematode infection was carried out within the genomic loci in linkage disequilibrium (LD) with the significant SNPs according to the GWAS results. LD was calculated as the squared coefficient of correlation (r^2) between each pair of markers using the LDheatmap R-package (Shin et al. 2006), and the following expression:

$$r^2 = \frac{(P_{AB} - P_A P_B)^2}{P_A(1 - P_A)P_B(1 - P_B)}$$

where: P_A is the allele frequency of A; P_B is the allele frequency of B, and P_{AB} is the allele frequency of haplotype AB. The genomic region browsed for candidate genes was defined based on the LD block in which the significant SNP was located ($r^2 > 0.2$), similar to the approach proposed by Assefa et al. (2019).

The *P. vulgaris* genome sequence (Schmutz et al. 2014) available on the NCBI Assembly database was used to obtain the protein sequences of the respective coding genes at the genomic intervals. These sequences were examined for similarity with sequences deposited in the RefSeq database (O’Leary et al. 2015) using the BLASTp tool in Blast2GO v.5 (Conesa et al. 2005), with an e-value cutoff of 1×10^{-5} . Gene ontology (GO) mapping and annotation were run using the same software with default parameters. In addition, GO terms enrichment was performed employing singular enrichment analysis, implemented on the AgriGO v.2.0 platform (Tian et al. 2017) according to Fisher’s exact test ($\alpha = 0.05$).

To narrow down the number of candidate genes, we initially took into consideration the genes expressed in roots of the Jamapa cultivar according to the common bean expression atlas available on the PvGEA platform (O’Rourke et al. 2014). The expression profile of whole roots collected at the second

trifoliate stage was used as a reference. Furthermore, candidate genes were also selected among those previously identified by Santini et al. (2016) as differentially expressed during the *P. vulgaris* – *M. incognita* interaction by RNA-seq analysis. Briefly, this study assessed the root tissue transcriptome of the Brazilian cultivar IPR-Saracura at 4 and 10 DAI, respectively equivalent to the stages of nematode penetration and giant nurse cell development. Sequencing libraries were constructed following the TruSeq RNA Sample Preparation v2 protocols (Illumina, San Diego, USA) and paired-end sequencing was performed on the MiSeq platform (Illumina, San Diego, USA), generating 101 bp reads. After alignment with the common bean reference genome (Schmutz et al. 2014), around 327 million reads were mapped to 27,195 unigenes, which were tested for differential expression. The raw data can be accessed in the NCBI database, Bioproject PRJNA288189.

Finally, aiming to predict the presence of resistance gene analogs (RGA), the protein sequences and their isoforms were subjected to a search for conserved protein domains, using the Hidden Markov Model profile, through the PFAM platform (El-Gebali et al. 2019) and a threshold e-value $\leq 1 \times 10^{-5}$. The following conserved domains were considered: Nucleotide binding-ARC (NB-ARC), Leucine-rich repeats (LRR), Transmembrane (TM), Coiled-coil (CC), Toll/interleukin-1 receptor (TIR) and Protein kinase (Pkinase), related to NBS-LRR and TM-LRR classes and their variations, which are often related to molecular mechanisms of pathogen recognition in plants.

2.2.7. Histochemical analysis of H₂O₂

Aiming to move forward and provide preliminary evidence of pathways involved in the common bean response to *M. incognita*, histochemical analysis was conducted to examine if inoculation triggered any sign of root ROS accumulation, particularly hydrogen peroxide (H₂O₂), in roots. Next, we investigated how *M. incognita* inoculation affected ROS status of five genotypes: two moderately resistant (IAC-Tybatã and Puebla-152-CIAT); two highly susceptible (TB-01-13 and Jamapa-CIAT-1671) and one previously classified as moderately susceptible (IPR-Saracura), formerly used to transcriptome analysis by Santini et al. (2015).

Plants were inoculated in growth pouches as described above, but at a higher inoculum concentration (6.000 J2s). Control samples were mock-inoculated with distilled water. At 24 h after inoculation, the roots were immersed in 50 mL of 1 mg L⁻¹ 3,3-diaminobenzidine (DAB) staining solution with a pH 3.0, and kept for 12 h in the dark. Finally, root systems were visualized using a 3D digital microscope (Hirox KH-8700, JP).

2.3. Results

2.3.1. Phenotypic analysis – IAC panel

Out of 180 accessions from IAC panel, 175 have been successfully evaluated for their sensitivity to RKN. Extensive phenotypic variation was observed within the IAC panel for number of egg masses and root-galling index. The number of egg masses ranged from 1 to 470 with a mean of 142.07, while the mean root-galling index was 2.19 with a range from 0 to 5 (Table 1).

Table 1. Descriptive summary and variance components of number of egg masses and root-galling index in a common bean core collection and a BC₂F₄ population inoculated with *M. incognita*

		Amplitude	Mean	CV _g (%)	Variance Components			$H_c^{2^a}$ (%)
					Genetic	Block	Residual	
IAC panel	Egg masses	1-470	142.1	30.71	1904.3	874.8	6178.6	48.9
	Root-galling Index	0-5	2.19	23.63	2.7E-01	7.1E-08	7.0E-01	55.3
BC ₂ F ₄ population	Egg masses	7-305	90.9	36.36	1092.6	912.6	2197.2	49.5
	Root-galling Index	0-4	2.02	24.75	2.5E-01	2.5E-02	4.3E-01	53.4

^aGeneralized heritability of Cullis et al. (2006)

The adjusted means of both traits for the core collection (IAC panel) followed continuous distributions and indicated the absence of extreme outliers. In terms of egg masses, 80 genotypes were more susceptible than the average for the panel, and 95 more resistant. However, only three genotypes had adjusted means lower than 100 egg masses. Similarly, in comparison to the mean, 94 genotypes presented weaker symptoms of root-galling, but only 27 had indexes lower than 2.0 and just one scored less than 1.5 (Figure 1).

Figure 1. Frequency distribution of the number of egg masses (a) and root-galling index (b) for a common bean core collection (IAC panel) inoculated with *M. incognita*

Respective heritability estimates of 48.9 and 55.3% were found for egg masses and root-galling (Table 1), indicating that a significant part of the observed variance of response to *M. incognita* might be explained by genetic factors, corroborating the suitability of the core collection for association studies.

A significant genotype effect was confirmed for both traits, according to the LRT results (Table 2), providing further evidence of the existence of a genetic component controlling the response to RKN. Moderate resistance was found in some genotypes, such as IAC-Tybatã and Puebla-152-CIAT with respective adjusted means of 71.6 and 75 for the number of egg masses, whereas other genotypes were highly susceptible, including TB-01-13 and Jamapa-CNF-1671 with respective adjusted means of 246 and 226.1 for the same trait (Appendix A).

Table 2. Likelihood ratio test (LRT) and Wald test of random and fixed effects for the number of egg masses and root-galling index of two common bean genetic resources inoculated with *M. incognita*

Genetic Resource	Source of Variation	Egg masses		Root-galling Index	
		Wald test	p-value	Wald test	p-value
IAC panel	Fixed Effects				
	Gene Pool	3.57	0.058	10.06	1.0E-02*
	Random Effects	LRT	p-value	LRT	p-value
	Genotype	39.64	3.1E-10	49.17	2.3E-12*
BC ₂ F ₄ population	Block	42.87	5.8E-11	1.0E-05	0.99
	Random Effects	LRT	p-value	LRT	p-value
	Genotype	9.67	1.9E-03	14.36	1.5E-04*
	Block	33.21	8.E-09	2.56	0.109

The Wald test detected a significant gene pool effect for both number of egg masses and root-galling index ($p = 0.058$ and $p < 0.01$, respectively; Table 2). Although each gene pool exhibited considerable

variation, the Andean genotypes had higher adjusted means than the Mesoamericans (Appendix C). Pearson coefficients revealed that the genotypic correlation between the number of egg masses and root-galling index was 0.61, suggesting that there may be a relation between the common bean's ability to minimize nematode multiplication and the intensity of observable symptoms in the roots. Root morphology was also examined herein, but no significant correlations were detected between root size and nematode symptoms and signs (Appendix D).

2.3.2. Phenotypic analysis – BC₂F₄ Population

The BC₂F₄ population also exhibited noticeable phenotypic variability, although not as extensive as that observed in the core collection. The number of egg masses and the root-galling index ranged from 7 to 305 and 0 to 4 respectively, with means of 90.9 and 2.02 (Table 1). Moreover, a significant genotype effect was observed for both traits and the generalized measurement of heritabilities for egg masses (49.5%) and root-galling index (53.4%) were also consistent with the results obtained for the IAC panel (Table 1).

Figure 2 shows the adjusted means for the BC₂F₄ progeny, donor and recurrent genitors, and two resistant and two susceptible controls ranked according to the IAC panel evaluation (Appendix A). The recurrent genitor and resistant controls exhibited similar responses to RKN, especially in terms of the number of egg masses. In addition, the donor genitor also did not differ from the susceptible controls with regards to root-galling symptoms, the confidence intervals of several BC₂F₄ individuals did not intersect the mean of the recurrent genitor, indicating transgressive segregation.

Figure 2. Caterpillar plot of the adjusted means with 95% confidence intervals for the number of egg masses (a) and root-galling index (b) in a BC₂F₄ population (green) and its donor (And-277; dark red) and recurrent genitors (IAC-Milênio; pink) inoculated with *M. incognita*. Two resistant (blue) and two susceptible (red) controls are also shown

2.3.3. Genome-Wide Association Studies

GWAS were conducted with a set of 10,362 SNPs, which were tested against the response to RKN infection taking into account the severity of root-galling and the crop's ability to suppress the production of nematode eggs. According to the quantile-quantile (Q-Q) plots, except for the significant associations that stand out at the high end of the plots, the models used for both number of egg masses and root-galling index presented observed p-values very similar to those that would be expected under the null hypothesis, indicating effective control for false positive associations (Figure 3b, d).

Figure 3. GWAS of number of egg masses and root-galling index for the IAC panel inoculated with *M. incognita*. In the Manhattan plots (a, c), the red dashed line indicates the Bonferroni threshold ($p < 4.82 \times 10^{-6}$) and the black dashed line denotes the permutation test threshold ($p < 2.29 \times 10^{-5}$ for egg masses and $p < 1.50 \times 10^{-5}$ for root-galling index). In the quantile-quantile plots (b, d) the shadow area denotes the 95% confidence interval

Based on the highly conservative Bonferroni correction, GWAS were able to detect six SNPs significantly associated to RKN resistance, including two for the number of egg masses and four for the root-galling index. Additionally, when the threshold was defined according to the permutation test, two further SNPs were found associated with the number of egg masses, giving a total of eight significant marker-trait associations (Figure 3a, c; Table 3). Noteworthy, there were no coincident associations. The four SNPs associated with the number of egg masses (S1_257146517, S1_320960286, S1_380326043 and S1_510326192) were located on chromosomes Pv06, Pv07, Pv08 and Pv11 and their effects were -7.6 , 13.9 , 14.7 and 6.6 , respectively. From these, the two SNPs significant according to the Bonferroni threshold (S1_257146517 and S1_320960286) had the highest PVEs (5.37 and 4.57% ; Table 3) and were located at position $48,672,347$ bp of Pv07, and $56,279,482$ bp of Pv08.

With regards to the root-galling index, the four significant SNPs (S1_46835797, S1_98931885, S1_234697928 and S1_425572688) were located on chromosomes Pv01, Pv02, Pv05, and Pv10. Their effects on the scored index varied from 0.16 (Pv01) to 0.29 (Pv05) and thus explained a higher proportion of variance than those associated with the egg masses, i.e. S1_234697928 and S1_425572688, with PVEs of 7.21 and 9.97% .

Table 3. Single-nucleotide polymorphism (SNP) significantly associated with the number of egg masses and root-galling index, according to the GWAS results for IAC panel genotypes inoculated with *M. incognita*

Egg masses								
Marker name	Chr	Position (bp)	p-value	MAF	Allelic reference	Allelic variant	Effect of the variant allele	PVE (%)
S1_257146517	Pv06	16835934	8.38E-06	0.22	T	A	−7.61	2.20
S1_320960286	Pv07	48672347	3.87E-06	0.15	A	G	13.92	5.37
S1_380326043	Pv08	56279482	4.64E-06	0.11	C	G	14.67	4.57
S1_510326192	Pv11	45872040	1.28E-05	0.34	T	A	6.57	2.13
Root-galling Index								
Marker name	Chr	Position (bp)	p-value	MAF	Allelic reference	Allelic variant	Effect of the variant allele	PVE (%)
S1_46835797	Pv01	46835797	7.66E-07	0.19	A	G	0.16	5.52
S1_98931885	Pv02	46726254	6.82E-07	0.13	T	C	0.19	5.80
S1_234697928	Pv05	35206731	2.90E-11	0.06	C	T	0.29	7.21
S1_425572688	Pv10	4393787	7.82E-15	0.48	T	G	−0.17	9.97

SNPs significantly associated according to the permutation test threshold ($p < 2.29 \times 10^{-5}$ for egg masses and $p < 1.50 \times 10^{-5}$ for root-galling index) or the Bonferroni multiple test correction ($p < 4.82 \times 10^{-6}$); Chr: Chromosome; MAF: Minor allele frequency; PVE: Proportion of the variance explained by the SNP–trait association

2.3.4. Validation by QTL mapping

After evaluating the IAC panel, the phenotypic contrast between genotypes AND-277 and IAC-Milênio was confirmed. These two genotypes were previously used as donor and recurrent genitors for producing a BC₂F₄ population. This population was therefore used to further confirm the loci identified via GWAS. Composite interval mapping was performed using the permutation-based threshold. For the number of egg masses, no QTL was detected. However, one QTL was found controlling the root-galling on chromosome Pv05 at map position 18.6 cM (Table 4; Appendix E). This QTL explained 9.24% of the total phenotypic variance and had an additive effect of −0.17 on the galling index.

Table 4. QTL identified for root-galling index in a BC₂F₄ population inoculated with *M. incognita*

Nearest marker	Chr	Map position (cM)	LOD	Effect of the variant allele	PVE (%)
Chr05_213377	Pv05	18.552	5.14	−0.17	9.24

Chr: Chromosome; LOD: Logarithm of odds ratio; PVE: Proportion of the variance explained by the SNP–trait association

2.3.5. Candidate genes

Functional mapping of the associated regions can provide valuable indications of which genes or pathways are causally related to the trait. Thus, to establish the size of the genomic regions for candidate gene identification, LD analysis was performed to define the loci in linkage disequilibrium with the SNP associated with egg masses and the root-galling index. These genomic intervals comprised between 89.5 kb

(Pv07) and 306 kb (Pv11) for egg masses and 76.7 kb (Pv02) and 677.3 kb for the root-galling index (Pv05; Appendix F). Of the 217 genes that we identified in these regions, 216 were functionally annotated, being 125 related to the number of egg masses and 91 to root-galling index (Appendix G).

Based on singular enrichment analysis, we contrasted the genome-wide associated regions for both traits and the complete common bean reference genome (Schmutz et al. 2014). In regions related to the number of egg masses, GO terms functionally enriched for biological processes included response to stress and biotic stimulus, defense response, and programmed cell death (PCD) and apoptosis, which were significantly overrepresented (17 to 28%,) when compared with genome-wide annotation (always below 5%). Similarly, 13 GO terms were enriched for molecular functions, notably those related to ATP- and nucleotide-binding, represented by up to 35% (Figure 4a). For the 91 genes annotated in the galling-associated regions, biological process terms were highlighted by protein metabolic processes (28%), such as proteolysis (11%) and translation (12%), while molecular functions related to hydrolase (24%), peptidase (14%), and serine-type peptidase activities (10%; Figure 4b).

Figure 4. Singular enrichment analysis of gene ontology (GO) for genomic regions associated with the number of egg masses (blue; a) and root-galling index (blue; b) and their representation in the complete genome annotation background (orange; a, b)

To narrow down the number of candidate genes regulating nematode resistance in common bean, we initially referred to a *P. vulgaris* expression atlas (O'Rourke et al. 2014) and verified that of the 216 annotated genes, 164 are expressed in root tissues (Appendix G). Five of these genes were previously reported as differentially expressed during the *P. vulgaris* - *M. incognita* interaction (Santini et al. 2016; Table 5), including four located on root-galling associated regions (Phvul.001G201300, Phvul.005G129600, Phvul.005G130300, Phvul.010G030100) and one on egg mass associated regions (Phvul.008G250700; Table 5). In particular, the genes encoding peroxidase 3-like (Phvul.001G201300), hypersensitive reaction associated Ca^{2+} binding protein (Phvul.005G129600) and serine threonine-protein kinase *OXI1*-like (Phvul.010G030100) were potentially implicated in defense responses, such as apoptosis, ROS pathways and hypertensive response, corroborating the GO enrichment results (Figure 4, Table 5).

Table 5. Candidate genes identified by the GWAS-associated genomic regions and differentially expressed in common bean roots at 4 and 10 days after inoculation with *M. incognita* according to Santini et al. (2016)

Trait	Associated SNP	Gene ID	Functional annotation	Exp. ^a 4 DAI	Exp. ^a 10 DAI	Genomic location
Egg masses	S1_380326043	Phvul.008G250700	trans-resveratrol di-o-methyltransferase-like	*	-1.23	Pv08:56528566.. 56529850
Root-galling	S1_46835797	Phvul.001G201300	peroxidase 3-like	2.638	*	Pv01:46687438.. 46688957
Root-galling	S1_234697928	Phvul.005G129600	hypersensitive reaction associated ca2+ binding protein	-1.274	*	Pv05:35461253.. 35462167
Root-galling	S1_234697928	Phvul.005G130300	zinc finger protein zat5-like	1.645	*	Pv05:35542896.. 35543786
Root-galling	S1_425572688	Phvul.010G030100	serine threonine-protein kinase <i>OXI1</i> -like	-1.87	*	Pv10:4404524.. 4406793

^a The expression values refer to those reported by Santini et al. (2016) according to log₂ fold-change (FC) ≥1.0 and false discovery rate ≤0.05

* Non-significant differential expression

Subsequently, trying to predict the existence of RGAs in the GWAS-associated regions, conserved domains such as NB-ARC, LRR, TM, CC, TIR, Pkinase, and variations were sought. We identified 14 RGAs in five regions, of which 12 were found regions associated with the number of egg masses and two with the root-galling index (Table 6). Among the 12 RGAs related to egg masses, there is a cluster with six genes encoding a disease resistance protein 1-like predicted on Pv06 (Phvul.006G051500, Phvul.006G052300, Phvul.006G052400, Phvul.006G052500, Phvul.006G052600, Phvul.006G052900). In addition, five putative disease resistance proteins were identified on Pv11 (Phvul.011G181400, Phvul.011G181500, Phvul.011G181600, Phvul.011G181700, Phvul.011G182300), including a classic TIR-NBS-LRR related to tobacco mosaic virus resistance (TMV Resistance protein N isoform X1). In the root-galling regions, Phvul.010G030100 stands out as a candidate both because encodes a serine threonine-protein kinase *OXI1*-like and because it was reported as differentially expressed at 4 DAI (log₂FC = -1.87; Table 5) by Santini et al. (2016).

Table 6. Resistance gene analogs (RGA) predicted in the GWAS-associated genomic regions, their domains, functional annotation and genomic location

Trait	RGA ID	Domains ^a	Functional annotation	Genomic location
Egg masses	Phvul.006G051500	Rx_N, NB-ARC	disease resistance protein 1-like	Pv06:16620561 ..16623632
Egg masses	Phvul.006G052300	Rx_N, NB-ARC	disease resistance protein 1-like	Pv06:16682591 ..16685935
Egg masses	Phvul.006G052400	Rx_N, NB-ARC	disease resistance protein 1-like	Pv06:16692493 ..16696743
Egg masses	Phvul.006G052500	Rx_N, NB-ARC	disease resistance protein 1	Pv06:16711094 ..16713766
Egg masses	Phvul.006G052600	Rx_N, NB-ARC	disease resistance protein 1-like	Pv06:16729789 ..16730533
Egg masses	Phvul.006G052900	LRR, TM, COR, CC	protein <i>TRN-1</i>	Pv06:16772649 ..16777588
Egg masses	Phvul.007G246600	N_T, LRR, TM	peceptor-like protein <i>EIX2</i>	Pv07:48600377 ..48603449
Egg masses	Phvul.011G181400	TIR, NB-ARC, LRR	TMV resistance protein N isoform X1	Pv11:45756765 ..45758261
Egg masses	Phvul.011G181500	Rx_N, CC, NB-ARC	putative disease resistance protein At3g14460	Pv11:45766765 ..45767029
Egg masses	Phvul.011G181600	Rx_N, NB-ARC, CC	putative disease resistance protein <i>RPP13</i> -like protein 1	Pv11:45768113 ..45772035
Egg masses	Phvul.011G181700	Rx_N, CC, NB-ARC	putative disease resistance protein At3g14460	Pv11:45773646 ..45777320
Egg masses	Phvul.011G182300	N_T, LRR, TM, Pkinase	leucine-rich repeat receptor-like protein kinase	Pv11:45895025 ..45898349
Root-galling	Phvul.001G202600	F_Box, TIR	transport Inhibitor response 1-like	Pv01:46818071 ..46823760
Root-galling	Phvul.010G030100	TIR, NB-ARC	serine threonine-protein kinase <i>OXI1</i> -like	Pv10: 4404524 ..4406793

^a Nucleotide binding-ARC (NB-ARC), Leucine-rich repeats (LRR), Transmembrane (TM), Coiled-coil (CC), Toll/interleukin-1 receptor (TIR), F-Box protein (F-Box), Resistance protein N-terminal (Rx_N), N-terminal (N_T) and Protein kinase (Pkinase)

2.3.6. Histochemical detection of H₂O₂

In many plant-pathogen interactions, ROS play a key role in the response mechanisms of the host. These molecules act reinforcing the cell wall barrier, inducing direct toxicity and triggering the activation of other defense responses, such as the hypersensitive response (Sato et al. 2019). Histochemical detection of hydrogen peroxide was performed in the roots of contrasting common bean genotypes by staining infected roots with 3,3'-diaminobenzidine (DAB).

In the histochemical analysis, no marked H₂O₂ accumulation was observed in the mock condition (no nematode inoculation). However, at 24 h after inoculation, we found evidence of greater H₂O₂ production in the roots of partially resistant genotypes IAC-Tybatã and Puebla-152-CIAT, compared to the highly susceptible TB-01-13 and Jamapa-CIAT-1671, which exhibited no obvious difference from the controls. Interestingly, the moderately susceptible IPR-Saracura also showed modest accumulation (Figure 5). This result suggests that H₂O₂ molecules may act in the common bean response to RKN, possibly involved in resistance mechanisms.

Figure 5. Histochemical detection of hydrogen peroxide (H_2O_2) in roots of common bean genotypes stained over 12 h with 3,3'-diaminobenzidine (DAB). The mock column shows non-inoculated roots and the RKN column roots 24 h after inoculation. Based on the phenotypic analysis, Puebla-152-CIAT and IAC-Tybatã had the lowest number of egg masses while Jamapa-CNF-1671 and TB-01-13 exhibited the highest values. IPR Saracura showed an intermediate response

2.4. Discussion

Constant screening of gene banks and an understanding of the genetic mechanisms underlying host responses to RKN are essential breakthroughs to identify and characterize sources of resistance and to develop tools to support breeding programs. In nematological studies, the crop responses to RKN are measured by the severity of the symptoms in the roots and the ability of the pathogen to reproduce in the host (Greco and Di Vito 2009). In this study, both responses were covered by assessing the number of egg masses and the root-galling index. We evaluated a core collection panel derived from one of the most extensively used common bean germplasms in Brazil and although complete immunity to nematode infection was not observed, a noticeable variation in RKN sensitivity was verified (Figure 1). For example, IAC-Tybatã and Puebla-152-CIAT showed moderate resistance, while TB-01-13 and Jamapa-CNF-1671 were highly susceptible to RKN (Appendix A). In addition to this panel, we also assessed a BC_2F_4 population, which despite its modest size, exhibited considerable phenotypic variability (Figure 2). Our results revealed that variance components and heritabilities were similar for both genetic resources, indicating that the genetic component explains a considerable portion of the phenotypic variance (Table 1).

A positive correlation was detected between egg mass production and severity of root-galling (Appendix D), but the interpretation at the biological level of this finding is not straightforward. When nematodes are not able to penetrate the roots there is no galling, but even though the occurrence of galls indicates that RKN have invaded the tissue, it does not necessarily imply that they will be capable of laying eggs and satisfactorily reproduce. Furthermore, no swellings are observed in some cases although significant reproduction occurs (Williamson and Roberts 2009). This complex interaction has been studied in many host plants and though there are species in which major resistance genes act to suppress both traits (Roberts 1995), in leguminous crops such as common bean (Fassuliotis et al. 1970; Mullin et al. 1991), soybean (Harris et al. 2003) and lima bean (Roberts et al. 2008) differential responses to root-galling and reproduction have been reported.

When undertaking GWAS, researchers explore LD particularly arising from the physical linkage between genomic markers and the QTLs controlling a given trait. Hence, relatedness among individuals and population structure are often taken into account in the association models in order to control biases and avoid false positive associations (Zhao et al. 2007; Zhang et al. 2010). In a previous study using the same core collection, Diniz et al. (2019) reported that this panel partially consists of improved lines, and therefore the kinship bias on the LD estimates is stronger than the one related to the population structure. The study also revealed that for this panel, population structure bias on the LD is already taken into account when modeling genetic relationships by the kinship matrix, especially since accessions from the same gene pool are closely related. Corroborating these findings, we confirmed herein that when fitting kinship to the association model, low general inflation of p-values was observed, demonstrating goodness of fit and therefore low false discovery rates (Figure 3b, d).

GWAS analysis identified eight genomic regions with small-to-moderate effects significantly associated with the RKN response, including four for egg masses and four for root-galling index. It denotes the existence of several genes controlling the traits and indicates the complex genetic architecture of the detected resistance. The GWAS results also suggested independent genetic control for each trait, since there were no coincident SNPs significantly associated with egg mass production and the root-galling index. For number of egg masses, significant associations were located on chromosomes Pv06, Pv07, Pv08 and Pv11, whereas for root-galling the associations were on Pv01, Pv02, Pv05 and Pv10 (Figure 3; Table 3). This result supports the view that the detected correlation among traits is consequence of indirect relationship and corroborates the findings of Roberts et al. (2008), who reported independent genetic control of resistance to root-galling and RKN reproduction in *Phaseolus lanatus*, which are probably unlinked based on their random assortment in segregating populations.

To validate the GWAS results, we conducted QTL mapping on a BC₂F₄ population. Using the CIM method, we detected one minor QTL on Pv05 for the root-galling index (Table 4). Interestingly, a SNP-galling association on the same chromosome was discovered by the GWAS, but although having comparable effects and PVE, the positions of the QTL and the associated SNPs did not overlap. While the nearest QTL marker was physically located at the beginning of Pv05 (213377 bp), the SNP-association was positioned at the end of the chromosome (35206731 bp). Although the results may suggest distinct QTLs,

it is also possible that they are the same, being the discrepancy in their locations ascribed to methodological aspects, such as the design scheme and the small BC₂F₄ population size, which may reflect in substantial within-class variance for QTL mapping and result in limited recombination events, implying low mapping resolution and detection power (Lynch and Walsh, 1998). It can also denote the existence of QTL x environment interaction, as the populations were evaluated independently, or even indicate the presence of a different QTL, as distinct alleles might be responsible for the detected effect in the GWAS panel and in the BC₂F₄ population. In some leguminous crops, such as peanut (*Arachis stenosperma*), soybean (*Glycine max*), and especially cowpea (*Vigna unguiculata*), major QTLs with PVEs higher than those found herein have been associated with resistance to root-galling and egg mass (Caromel and Gebhardt 2011; Huynh et al. 2016; Leal-bertioli et al. 2016; Passianotto et al. 2017; Santos et al. 2018; Ndeve et al. 2019). Nevertheless, in terms of aspects such as the complexity of these traits in *Phaseolus* spp. (Roberts et al. 2008), as well as the phenotyping challenges and the lack of complete immunity, our findings could help to increase selection accuracy in marker-assisted breeding.

Based on the GWAS, we delimited the genomic regions in LD with the SNP-trait associations and explored candidate genes putatively involved in the suppression of egg mass production and attenuation of root-galling severity. Singular enrichment analysis showed that for egg masses, the biological process GO terms were functionally enriched for response to stress and biotic stimulus, especially through PCD and apoptosis (Figure 4). PCD is a common response of plants to pathogens where resistance is conferred by R-genes, which recognize pathogen avirulence genes (*Avr*) initiating a cascade of signaling processes, that involves major changes in gene expression, oxygen burst and leads to a hypersensitivity response (Dropkin 1969; Blevé-Zacheo et al. 1998; Williamson and Roberts 2009). In tomato, for example, most of the available cultivars contain the R-gene *Mi*, which as a RGA encodes a classic cytoplasmatic nucleotide-binding leucine-rich repeat (NB-LRR) protein (Jones and Goto 2011) and confers resistance, though not immunity, to *M. incognita*, *M. javanica* and *M. arenaria* (Milligan et al. 1998; Williamson and Roberts 2009). Several studies have also reported that RGAs are unevenly spread, typically forming complex clusters in plant genomes (Sekhwal et al. 2015), including the *P. vulgaris* (Schmutz et al. 2014).

Interestingly, in line with other nematode resistance genes, presenting typical conserved RGA domains, as NB-LRR (Milligan et al. 1998), in this study 12 RGAs were found in the regions associated to egg mass (Table 6). These genes include a cluster with six genes encoding a disease resistance protein 1-like on Pv06 (Phvul.006G051500, Phvul.006G052300, Phvul.006G052400, Phvul.006G052500, Phvul.006G052600, Phvul.006G052900), a receptor-like protein *EIX2* on Pv07 (Phvul.007G246600) and a cluster of five putative disease resistance proteins on Pv11 (Phvul.011G181400, Phvul.011G181500, Phvul.011G181600, Phvul.011G181700, Phvul.011G182300), including a TMV resistance protein N isoform X1, a tobacco mosaic virus (TMV) resistance gene and member of the TIR-NBS-LRR class (Dinesh-Kumar and Baker 2000). Interestingly, two other analogs of the TMV resistance gene (N gene) have already been reported as differentially expressed during the *P. vulgaris*-*M. incognita* interaction (Santini et al. 2016) and mapped by homology on the same Pv11 chromosome (Santos et al. 2018).

For the root-galling associated regions, biological processes were enriched for terms such as protein metabolic processes, structural functions and organic acids (Figure 4). With regard to molecular functions, genes such as Phvul.001G202900, Phvul.002G304800, Phvul.005G130900 and Phvul.010G030600, encoding for hydrolases and peptidase (Appendix G), may play a role as membrane components and promote cell wall modifications influencing nematode infection and swelling symptoms. Shukla et al. (2018) were able to demonstrate differential regulation of genes encoding cell wall-degrading enzymes and peptidases in a RKN-susceptible tomato. By contrast, the absence of differential expression during the resistance response indicates that these genes are fundamentally involved in cell enlargement and cell wall rearrangement during gall formation, and additionally suggests that RKN effectors might modulate their expression as the disease progresses.

Other remarkable candidate genes located in root-galling regions include two up- and two down-regulated during an interaction of a susceptible cultivar at 4 DAI as reported by Santini et al. (2015), three of which are related to protection against oxidative damage. The hypersensitive reaction associated Ca^{2+} -binding protein (Phvul.005G129600) is an interesting candidate since a Ca^{2+} influx into the cytoplasm is an early stage response in the signaling pathway of ROS production (Dodd et al. 2010), as has been demonstrated in other *Meloidogyne* pathosystems (Davies et al. 2015). A second one codes to an enzyme similar to a peroxidase (peroxidase 3-like; Phvul.001G201300), involved in the antioxidant response; and lastly, a RGA annotated as serine threonine-protein kinase *OXI1*-like (Phvul.010G030100; Table 6), with correspondence to a well-studied gene necessary for oxidative burst-mediated signaling in plants, including defense against root pathogens (Rentel et al. 2004; Petersen et al. 2009).

Histochemical analysis provided preliminary evidence that the accumulation of ROS, specifically H_2O_2 , is involved in the common bean response to *M. incognita* and is higher in the more resistant genotypes (Figure 5). Our results corroborate those of Melillo et al. (2006), who first demonstrated that the phenotypic expression of complete resistance to RKN in tomato was characterized by a hypersensitive reaction induced by the accumulation of ROS. In a study of leguminous cowpea, Oliveira et al. (2012) reported a higher increase of antioxidant enzymes in resistant genotypes, suggesting higher production and accumulation of H_2O_2 . In any case, the genotypes investigated herein by histochemical analysis did not exhibit complete immunity, indicating that if the hypersensitivity reaction occurs, as suggested by GO enrichment analysis and gene candidate discovery, it cannot fully prevent nematode invasion and establishment. This absence of typical hypersensitive response was also observed in cowpea, and despite the fact that resistant genotypes present higher ROS accumulation 24 h after inoculation with *M. incognita*, there was no characteristic hypersensitive response (Das et al. 2008).

To the best of our knowledge, this is the first study of genetic mapping for RKN resistance in common bean. Results indicate distinct genomic regions associated with egg mass production and severity of root-galling, with complex response mechanisms. In addition, several candidate genes were indicated to putatively control the resistance, including 12 RGAs and five DE genes in common bean roots after inoculation. Thus, validation by fine mapping approaches might provide valuable information for marker-assisted selection and aid breeders to face the challenges imposed by the phenotyping bottleneck (Caromel

and Gebhardt 2011), as is already the case for peanut (Chu et al. 2011), cotton (Jenkins et al. 2012) and soybean crops (Dubiel et al. 2019). Finally, functional analysis of the candidate genes will help elucidate the molecular basis of the *P. vulgaris* - *M. incognita* interaction and could eventually be used for genetically modifying other related crops.

References

- Abad P, Castagnone-Sereno P, Rosso MN, et al (2009) Invasion, feeding and development. In: Perry RN, Moens CB, Starr JL (eds) Root-knot nematodes, 1st edn. CABI International, London, pp 163–181
- Abawi GS, Mullin BA, Mai WF (2005) Diseases caused by nematodes. In: Schwartz HF, Steadman JR, Hall R (eds) Compendium of bean diseases, 2nd edn. American Phytopathological Society Press, St. Paul, pp 54–57
- Alves FR, Santos LNS dos, Moraes WB, et al (2011) Reaction of common bean genotypes to *Meloidogyne incognita* Race 1. Idesia 29:95–98. <https://doi.org/10.4067/S0718-34292011000200012>
- Assefa T, Mahama AA, Brown A V, et al (2019) A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). Mol Breed 39:1–23. <https://doi.org/10.1007/s11032-018-0920-0>
- Atamian HS, Roberts PA, Kaloshian I, High I (2012) High and Low Throughput Screens with Root-knot Nematodes *Meloidogyne*. J Vis Exp 61:2–7. <https://doi.org/10.3791/3629>
- Atkinson HJ, Urwin PE, Hansen E, McPherson MJ (1995) Designs for engineered resistance to root-parasitic nematodes. Trends Biotechnol 13:369–374. [https://doi.org/10.1016/S0167-7799\(00\)88983-0](https://doi.org/10.1016/S0167-7799(00)88983-0)
- Bellucci E, Bitocchi E, Rau D, et al (2014) Genomics of Origin, Domestication and Evolution of *Phaseolus vulgaris*. In: Tuberosa R, Graner A, Frison E (eds) Genomics of Plant Genetic Resources, 1st edn. Springer Netherlands, Dordrecht, pp 483–507
- Bitocchi E, Bellucci E, Giardini A, et al (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. New Phytol 197:300–313. <https://doi.org/10.1111/j.1469-8137.2012.04377.x>
- Bitocchi E, Nanni L, Bellucci E, et al (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. PNAS 109:788–796. <https://doi.org/10.1073/pnas.1108973109>
- Bitocchi E, Rau D, Bellucci E, et al (2017) Beans (*Phaseolus* spp.) as a Model for Understanding Crop Evolution. Front. Plant Sci. 8:722–730. <https://doi.org/10.3389/fpls.2017.00722>
- Blair MW, Gonza LF, Kimani PM, Butare L (2010) Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. Theor Appl Genet 121:237–248. <https://doi.org/10.1007/s00122-010-1305-x>
- Bleve-Zacheo T, Bongiovanni M, Melillo MT, Castagnone-Sereno P (1998) The pepper resistance genes *Me1* and *Me3* induce differential penetration rates and temporal sequences of root cell ultrastructural changes upon nematode infection. Plant Sci 133:79–90. [https://doi.org/10.1016/S0168-9452\(98\)00021-1](https://doi.org/10.1016/S0168-9452(98)00021-1)
- Bozbuga R, Dasgan HY, Akhoundnejad Y, et al (2015) Identification of common bean (*Phaseolus vulgaris*) genotypes having resistance against root knot nematode *Meloidogyne incognita*. Legum Res 38:669–674. <https://doi.org/10.18805/lr.v38i5.5948>
- Bridge J, Page SLJ (1980) Estimation of root-knot nematode infestation levels on roots using a rating chart. Int J Pest Manag 26:296–298. <https://doi.org/10.1080/09670878009414416>
- Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890. <https://doi.org/10.1093/bioinformatics/btg112>
- Broughton WJ, Hern G, Blair M, et al (2003) Beans (*Phaseolus* spp.) – model food legumes. Plant Soil 55:55–128. <https://doi.org/10.1023/A:1024146710611>
- Butler D, Cullis B, Gilmour A, Gogel B (2009) Analysis of Mixed Models for S-language Environments: ASReml–R Reference Manual. Orange, Australia
- Caromel B, Gebhardt C (2011) Breeding for nematode resistance: use of genomic information. In: Jones J, Gheysen G, Fenoll C (eds) Genomics and Molecular Genetics of Plant-Nematode Interactions, 1st edn. Springer, Dordrecht, pp 465–492
- Chacón S MI, Pickersgill B, Debouck DG (2005) Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. Theor Appl Genet 110:432–444. <https://doi.org/10.1007/s00122-004-1842-2>
- Chu Y, Wu CL, Holbrook CC, et al (2011) Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. Plant Genome 4:110–117. <https://doi.org/10.3835/plantgenome2011.01.0001>

- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971. <https://doi.org/10.1534/genetics.107.080101>
- Conesa A, Götz S, García-Gómez JM, et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Costa JPG da, Soares PLM, Vidal RL, et al (2019) Reaction of common bean genotypes to the reproduction of *Meloidogyne javanica* and *Meloidogyne incognita*. *Pesqui. Agropecuária Trop.* 49:e54008
- Coyne D, Affokpon A (2018) Nematode parasites of tropical root and tuber crops (excluding potatoes). In: Sikora RA, Coyne D, Hallman J, Timpper P (eds) *Plant parasitic nematodes in subtropical and tropical agriculture*, 3rd edn. CAB International, Wallingford, pp 252–289
- Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *J Agric Biol Environ Stat* 11:381–393. <https://doi.org/10.1198/108571106X154443>
- Das S, DeMason DA, Ehlers JD, et al (2008) Histological characterization of root-knot nematode resistance in cowpea and its relation to reactive oxygen species modulation. *J Exp Bot* 59:1305–1313. <https://doi.org/10.1093/jxb/ern036>
- Davies LJ, Brown CR, Elling AA (2015) Calcium is involved in the *RMε1*(blb)-mediated hypersensitive response against *Meloidogyne chitwoodi* in potato. *Plant Cell Rep* 34:167–177. <https://doi.org/10.1007/s00299-014-1697-1>
- Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *PNAS* 97:1908–1913. <https://doi.org/10.1073/pnas.020367497>
- Diniz AL, Giordani W, Costa ZP, et al (2019) Evidence for Strong Kinship Influence on the Extent of Linkage Disequilibrium in Cultivated. *Genes* 10:1–16. <https://doi.org/10.3390/genes10010005>
- Dodd AN, Kudla J, Sanders D (2010) The Language of Calcium Signaling. *Annu Rev Plant Biol* 61:593–620. <https://doi.org/10.1146/annurev-arplant-070109-104628>
- Dropkin VHP (1969) The necrotic reaction of tomatoes and other hosts resistant to *Meloidogyne*: reversal by temperature. *Phytopathology* 59:1632–1637
- Dubiela C, Montecelli TDN, Lazzari F, et al (2019) Development and validation of SNP assays for the selection of resistance to *Meloidogyne incognita* in soybean. *Crop Breed Appl Biotechnol* 19:102–109. <https://doi.org/10.1590/1984-70332019v19n1a14>
- El-Gebali S, Mistry J, Bateman A, et al (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* 47:427–432. <https://doi.org/10.1093/nar/gky995>
- Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One* 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Faostat (2018) Food and agriculture organization of the united nations statistics division. *Crop. Prod.* <http://www.fao.org/faostat/en/#data/QC>. Accessed 14 May 2020
- Fassuliotis G, Deakin JR, Hoffman JC (1970) Root-knot nematode resistance in snap beans: breeding and nature of resistance. *J Am Soc Hortic Sci* 95:640–645
- Gepts P, Aragão FJL, Barros E de, et al (2008) Genomics of *Phaseolus* Beans, a Major Source of Dietary Protein and Micronutrients in the Tropics. In: Moore PH, Ming R (eds) *Genomics of Tropical Crop Plants*, 1st edn. Springer New York, New York, pp 113–143
- Gepts P, Debouck D (1991) Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). In: Voysest O, Schoonhoven A Van (eds) *Common beans: research for crop improvement*. CABI, Wallingford, pp 7–53
- Gepts P, Osborn TC, Rashka K, Bliss FA (1986) Phaseolin-protein Variability in Wild Forms and Landraces of the Common Bean (*Phaseolus vulgaris*): Evidence for Multiple Centers of Domestication. *Econ Bot* 40:451–468. <https://doi.org/10.1007/BF02859659>
- Greco N, Di Vito M (2009) Population Dynamics and Damage Levels. In: Perry RN, Moens M, Starr JL (eds) *Root-knot nematodes*, 1st edn. CABI International, p 246
- Hanai LR, De Campos T, Camargo LEA, et al (2007) Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. *Genome* 50:266–277. <https://doi.org/10.1139/g07-007>
- Harris DK, Boerma HR, Hussey RS, Finnerty SL (2003) Additional sources of soybean germplasm resistant to two species of root-knot nematode. *Crop Sci* 43:1848–1851. <https://doi.org/10.2135/cropsci2003.1848>
- Hoagland DR, Arnon DI (1950) The water-culture method for growing plants without soil. University of California, Berkeley
- Huang X, Han B (2014) Natural Variations and Genome-Wide Association Studies in Crop Plants. *Annu Rev Plant Biol* 65:531–551. <https://doi.org/10.1146/annurev-arplant-050213-035715>
- Huynh B-L, Matthews WC, Ehlers JD, et al (2016) A major QTL corresponding to the *Rk* locus for resistance to root-knot nematodes in cowpea (*Vigna unguiculata* L. Walp.). *Theor Appl Genet* 129:87–95. <https://doi.org/10.1007/s00122-015-2611-0>
- Jenkins JN, McCarty JC, Wubben MJ, et al (2012) SSR markers for marker assisted selection of root-knot nematode (*Meloidogyne incognita*) resistant plants in cotton (*Gossypium hirsutum* L.). *Euphytica* 183:49–54. <https://doi.org/10.1007/s10681-011-0470-6>

- Jones MGK, Goto DB (2011) Root-knot nematodes and giant cells. In: Jones J, Gheysen G, Fenoll C (eds) Genomics and molecular genetics of plant-nematode interactions, 1st edn. Springer, Dordrecht, pp 83–100
- Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. Theor Appl Genet 78:809–817. <https://doi.org/10.1007/BF00266663>
- Koenning SR, Overstreet C, Noling JW, et al (1999) Survey of Crop Losses in Response to Phytoparasitic Nematodes in the United States for 1994. J nema 31:587–618
- Koinange EMK, Gepts P (1992) Hybrid weakness in wild *Phaseolus vulgaris* L. J Hered 83:135–139. <https://doi.org/10.1093/oxfordjournals.jhered.a111173>
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29. <https://doi.org/10.1186/1746-4811-9-29>
- Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet 118:979–992. <https://doi.org/10.1007/s00122-008-0955-4>
- Leal-bertioli SCM, Moretzsohn MC, Roberts PA, et al (2016) Genetic Mapping of Resistance to *Meloidogyne arenaria* in *Arachis stenosperma*: A New Source of Nematode Resistance for Peanut. G3 6:377–390. <https://doi.org/10.1534/g3.115.023044>
- Lilley CJ, Kyndt T, Gheysen G (2011) Nematode Resistant GM Crops in Industrialised and Developing Countries. In: Jones J, Gheysen G, Fenoll C (eds) Genomics and Molecular Genetics of Plant-Nematode Interactions, 1st edn. Springer Netherlands, Dordrecht, pp 517–541
- Lipka AE, Tian F, Wang Q, et al (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Liu X, Huang M, Fan B, et al (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genet 12:e1005767. <https://doi.org/10.1186/1471-2156-13-100>
- Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, Massachusetts
- Margarido GRA, Souza AP, Garcia AAF (2007) OneMap: software for genetic mapping in outcrossing species. Hereditas 144:78–79. <https://doi.org/10.1111/j.2007.0018-0661.02000.x>
- Melillo MT, Leonetti P, Bongiovanni M, et al (2006) Modulation of reactive oxygen species activities and H₂O₂ accumulation during compatible and incompatible tomato–root-knot nematode interactions. New Phytol 170:501–512. <https://doi.org/10.1111/j.1469-8137.2006.01724.x>
- Messeguer R, Ganai M, de Vicente MC, et al (1991) High resolution RFLP map around the root knot nematode resistance gene (*Mi*) in tomato. Theor Appl Genet 82:529–536. <https://doi.org/10.1007/BF00226787>
- Milligan SB, Bodeau J, Yaghoobi J, et al (1998) The Root Knot Nematode Resistance Gene *Mi* from Tomato Is a Member of the Leucine Zipper, Nucleotide Binding, Leucine-Rich Repeat Family of Plant Genes. Plant Cell 10:1307–1319. <https://doi.org/10.1105/tpc.10.8.1307>
- Mullin BA, Abawi GS, Pastor-Corrales MA (1991) Modification of Resistance Expression of *Phaseolus vulgaris* to *Meloidogyne incognita* by Elevated Soil Temperatures. J Nematol 23:182–187
- Ndeve AD, Santos JRP, Matthews WC, et al (2019) A Novel Root-Knot Nematode Resistance QTL on Chromosome Vu01 in Cowpea. G3 9:1199–1209. <https://doi.org/10.1534/g3.118.200881>
- Nyczepir AP, Thomas SH (2009) Management Strategies in Resource-poor Farming. In: Perry RN, Moens M, Starr JL (eds) Root-knot nematodes, 1st edn. CABI International, p 444
- O’Leary NA, Wright MW, Brister JR, et al (2015) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:733–745. <https://doi.org/10.1093/nar/gkv1189>
- O’Rourke JA, Iniguez LP, Fu F, et al (2014) An RNA-Seq based gene expression atlas of the common bean. BMC Genomics 15:866. <https://doi.org/10.1186/1471-2164-15-866>
- Oliveira JTA, Andrade NC, Martins-Miranda AS, et al (2012) Differential expression of antioxidant enzymes and PR-proteins in compatible and incompatible interactions of cowpea (*Vigna unguiculata*) and the root-knot nematode *Meloidogyne incognita*. Plant Physiol Biochem 51:145–152. <https://doi.org/10.1016/j.plaphy.2011.10.008>
- Passianotto ALDL, Sonah H, Dias WP, Abdelnoor R V (2017) Genome-wide association study for resistance to the southern root-knot nematode (*Meloidogyne incognita*) in soybean. Mol Breed 37:148. <https://doi.org/10.1007/s11032-017-0744-3>
- Perseguini JMCK, Silva GSB, Rosa JRBF, et al (2015) Developing a common bean core collection suitable for association mapping studies. Genet Mol Biol 38:67–78. <https://doi.org/10.1590/S1415-475738120140126>
- Petersen LN, Ingle RA, Knight MR, Denby KJ (2009) *OXI1* protein kinase is required for plant immunity against *Pseudomonas syringae* in Arabidopsis. J Exp Bot 60:3727–3735. <https://doi.org/10.1093/jxb/erp219>
- Petry N, Boy E, Wirth JP, Hurrell RF (2015) Review: The Potential of the Common Bean (*Phaseolus vulgaris*) as a Vehicle for Iron Biofortification. Nutrients 7:1144–1173. <https://doi.org/10.3390/nu7021144>
- Rentel MC, Lecourieux D, Ouaked F, et al (2004) *OXI1* kinase is necessary for oxidative burst-mediated signalling in Arabidopsis. Nature 427:858–861. <https://doi.org/10.1038/nature02353>
- Roberts PA (1995) Conceptual and practical aspects of variability in root-knot nematodes related to host plant resistance. Annu Rev Phytopathol 33:199–221. <https://doi.org/10.1146/annurev.py.33.090195.001215>

- Roberts PA, Matthews WC, Ehlers JD, Helms D (2008) Genetic determinants of differential resistance to root-knot nematode reproduction and galling in lima bean. *Crop Sci* 48:553–561. <https://doi.org/10.2135/cropsci2007.07.0384>
- Rodriguez M, Rau D, Bitocchi E, et al (2016) Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytol* 209:1781–1794. <https://doi.org/10.1111/nph.13713>
- Rossi M, Bitocchi E, Bellucci E, et al (2009) Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol Appl* 2:504–522. <https://doi.org/10.1111/j.1752-4571.2009.00082.x>
- Santini L, Munhoz CDF, Bonfim-Jr MF, et al (2016) Host Transcriptional Profiling at Early and Later Stages of the Compatible Interaction Between *Phaseolus vulgaris* and *Meloidogyne incognita*. *Nematology* 106:282–294. <https://doi.org/10.1094/PHYTO-07-15-0160-R>
- Santos JRP, Ndeve AD, Huynh B-L, et al (2018) QTL mapping and transcriptome analysis of cowpea reveals candidate genes for root-knot nematode resistance. *PLoS One* 13:1–22. <https://doi.org/10.1371/journal.pone.0189185>
- Santos LNS, Alves FR, Belan LL, et al (2012) Damage quantification and reaction of bean genotypes (*Phaseolus vulgaris* L.) to *Meloidogyne incognita* race 3 and *M. javanica*. *Summa Phytopathol* 38:24–29. <https://doi.org/10.1590/S0100-54052012000100004>
- Sato K, Kadota Y, Shirasu K (2019) Plant Immune Responses to Parasitic Nematodes. *Front Plant Sci* 10:1165–1179. <https://doi.org/10.3389/fpls.2019.01165>
- Schmutz J, McClean PE, Mamidi S, et al (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet* 46:707–713. <https://doi.org/10.1038/ng.3008>
- Sekhwil MK, Li P, Lam I, et al (2015) Disease Resistance Gene Analogs (RGAs) in Plants. *Int J Mol Sci* 16:19248–19290. <https://doi.org/10.3390/ijms160819248>
- Shin JH, Blay S, McNeney B (2006) LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms. *J Stat Softw* 16:1–10. <https://doi.org/10.18637/jss.v016.c03>
- Shukla N, Yadav R, Kaur P, et al (2018) Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Mol Plant Pathol* 19:615–633. <https://doi.org/10.1111/mpp.12547>
- Singh SP, Gutierrez JA, Molina A, et al (1991) Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. *Crop Sci* 31:23–29. <https://doi.org/10.2135/cropsci1991.0011183X003100010005x>
- Singh SP, Schwartz HF (2011) Review: Breeding common bean for resistance to insect pests and nematodes. *Can J plant Sci* 91:239–250. <https://doi.org/10.4141/CJPS10002>
- Tian T, Liu Y, Yan H, et al (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* 45:122–129. <https://doi.org/10.1093/nar/gkx382>
- Trudgill DL, Blok VC (2001) Apomictic, polyphagous, root-knot nematodes: Exceptionally Successful and Damaging Biotrophic Root Pathogens. *Annu Rev Phytopathol* 39:53–77. <https://doi.org/10.1093/0066-4286/01/0901-0053>
- Westermann AJ, Gorski SA, Vogel J (2012) Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* 10:618–630. <https://doi.org/10.1038/nrmicro2852>
- Williamson VM (1998) Root-knot nematode resistance genes in tomato and their potential for future use. *Annu Rev Phytopathol* 36:277–293. <https://doi.org/10.1146/annurev.phyto.36.1.277>
- Williamson VM, Roberts PA (2009) Mechanisms and genetics of resistance. In: Perry RN, Moens M, Starr JL (eds) *Root-knot nematodes*, 1st edn. CAB International, Wallingford, pp 301–325
- Yaghoobi J, Kaloshian I, Wen Y, Williamson VM (1995) Mapping a new nematode resistance locus in *Lycopersicon peruvianum*. *Theor Appl Genet* 91:457–464. <https://doi.org/10.1007/BF00222973>
- Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468. <https://doi.org/10.1007/s00122-012-2032-2>
- Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42:355–360. <https://doi.org/10.1038/ng.546>
- Zhao K, Aranzana MJ, Kim S, et al (2007) An Arabidopsis example of association mapping in structured samples. *PLoS Genet* 3:e4. <https://doi.org/10.1371/journal.pgen.0030004>
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. *Plant Genome* 1:5–20. <https://doi.org/10.3835/plantgenome2008.02.0089>

Declarations**Funding**

This research was funded by the following Brazilian Institutions: Fundação de Amparo à Pesquisa do Estado de São Paulo, grant numbers 2014/06647-9, 2018/09069-7 and 2020/02755-2; Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant numbers 454329/2014-8 and 140588/2018-3; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, finance code 001.

Availability of data and material

GBS data can be accessed at the GenBank database, BioSamples SAMN05513252 and SAMN05513251, included in BioProject PRJNA336556. RNA-seq libraries can be obtained under accession number PRJNA288189. Phenotypic data are available by contacting the corresponding author.

Acknowledgments

The authors wish to thank the following Brazilian Institutions: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grant number 2014/06647-9, 2018/09069-7 and 2020/02755-2); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grant number 454329/2014-8); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Finance Code 001) for financial support. We also thank Carlos A. de Oliveira for technical assistance, Professor Mario Massayuki Inomoto (University of São Paulo) for providing the nematode inoculum and Mr. Steve Simmons for proofreading the manuscript.

3. GENOME-WIDE ASSOCIATION STUDIES DISSECT THE GENETIC ARCHITECTURE OF SEED SHAPE AND SIZE IN COMMON BEAN

ABSTRACT

Seed weight and size are important yield components and thus selecting for large seeds has been a key objective in crop domestication and breeding. For common bean, shape is also a noteworthy seed trait since it influences industrial processing, and affects the preferences of consumers and farmers. Hence, understanding the genetic mechanisms underlying seed morphology is a critical step towards the development and acceptance of common bean cultivars. In this study, we performed genome-wide association studies (GWAS) on a core collection of 180 common bean accessions to dissect the genetic architecture and identify genomic regions associated with seven seed morphological traits related to weight (hundred seed mass), size (width, length, projected area and perimeter) and shape (circularity and length-to-width ratio). Phenotypic data on seed shape and size was collected by high-throughput image-based approaches and used to test associations with 10,362 SNP markers by multi-locus mixed-models. Finally, we used expression studies and searched within the genome-associated regions for candidate genes putatively involved in seed phenotypic variation. The core collection panel exhibited high variability for the entire set of seed traits, and greater size and weight were observed for the Andean gene pool. Strong pairwise correlations were verified for most seed traits. GWAS identified seven marker-trait associations, explaining a considerable amount of phenotypic variation for length, width, projected area, perimeter and circularity on four distinct genomic regions of chromosomes Pv02, Pv08 and Pv11. In the associated regions, 188 genes were identified, including 156 expressed in seed development tissues according to previous studies. Thirteen genes were considered promising candidates for regulating seed morphology in common bean. From a practical standpoint, this study provides SNP markers that may help breeders in marker-assisted selection. Moreover, the genes that were pinpointed are good candidates for functional analysis in order to validate their influence on seed shape and size in common bean and eventually other related crops.

3.1. Introduction

Common bean (*Phaseolus vulgaris* L., $2n = 2x = 22$) is the most consumed legume grain in the human diet (Broughton et al. 2003) and a valuable source of protein, energy and micronutrients for more than 300 million people in the developing world (Gepts et al. 2008; Blair et al. 2010; Petry et al. 2015). Its relevance to global food security is becoming increasingly important due to markedly high population growth in the leading common bean consumption countries (Akibode and Maredia 2011; Bellucci et al. 2014) and the rising importance of protein resources of vegetable origin (Petrusán et al. 2016; Kumar et al. 2017). Brazil is one of the largest consumers and producers of common bean, harvesting around 3 million metric tons per year (FAOSTAT 2018).

In common bean, the occurrence of dual independent domestication events led to the establishment of two gene pools, the Mesoamerican and Andean. These gene pools display many morphological and genetic features (Singh et al. 1991b; Kwak and Gepts 2009) that make *P. vulgaris* an

exceptional model for studies on crop evolution and domestication (Bitocchi et al. 2017). Genotypes in the Mesoamerican gene pool exhibit greater genetic diversity (Mamidi et al. 2013) and are characterized by small to medium-sized seeds with “S” or “B” phaseolin types, while Andean seeds are larger with “T”, “C”, “H” and “A” phaseolin patterns (Gepts et al. 1986; Singh et al. 1991a).

Seed weight and size are important yield components and thus selecting for large seeds has been a key target in crop domestication (Gavazzi and Sangiorgio 2017). A superficial examination of the relationship between yield and seed size might suggest that these traits are positively correlated, since seed size should have positive influence on yield. However, for common bean, in which hundred seed mass may range from 15 to over 60 g, strong negative correlations have been reported (Adams 1967; Nienhuis and Singh 1986; White and Gonçalves 1990), since heavier seeds ought to reflect on a fewer seeds per plant. This negative correlation has been of great interest to bean breeders because selecting for higher yields can result in genotypes with low seed size and weight (White and Gonçalves 1990). Nevertheless, in the major common bean markets such as Latin America, Africa, and Asia, larger seeds are more attractive to consumers and drive up market prices (Nienhuis and Singh 1986).

Seed shape is also an important commercial criterion for the processing industry, consumers and farmers. For instance, information on bean morphological seed features such as length, width, thickness, projected area, perimeter, weight, circularity and length-to-width ratio are critical for designing seed drills and separation, sizing and handling systems (Kara et al. 2013; FAO and AfricaSeeds 2018; Shaw et al. 2020). Hence, bean breeders have become increasingly interested in understanding the genetic architecture of seed shape and size traits, mainly for identifying genotypes that combine high yield, large seeds, adaptation to industrial processing and consumer appeal (White and Gonçalves 1990; Park et al. 2000; Blair et al. 2009; Lei et al. 2020).

Since the early days of genetics as a science, common bean seed traits such as weight and size have been reported to be subject to quantitative inheritance (Johannsen 1911; Sax 1923), leading the botanist Wilhelm Johannsen to coin terms such as “genotype”, “phenotype” and “gene” (Johannsen 1909, 1911), and crucially providing evidence to resolve the classical biometric-Mendelian controversy (Roll-Hansen 1989). Further down the road, molecular markers have been to dissect the genetic architecture of complex traits in plants based on two main approaches, known as quantitative trait locus (QTL) mapping and genome-wide association studies (GWAS) (Mitchell-olds 2010; Xu et al. 2017). In family-based QTL mapping, designed crosses are used to unravel the genetic variations that differentiate progeny, in which meiotic cross-over events allow the identification of genetic loci segregating together with phenotypic variation (Lander and Botstein 1989). This methodology has been used to identify QTLs linked to many yield-related traits in common bean (González et al. 2017), including seed weight, size and shape (Koinange et al. 1996; Park et al. 2000; Guzman-Maldonado et al. 2001; Blair et al. 2009; Pañeda et al. 2010; Checa and Blair 2013; Yuste-Lisbona et al. 2014; Geng et al. 2017).

The development of next-generation sequencing technologies has significantly improved genotyping capacity and cut its costs (Varshney et al. 2009), boosting the popularity of GWAS, which takes

advantage of the historical accumulation of recombination events, obviating the need to synthesize segregating populations. This approach has enabled researchers to explore a wider genetic base and provided higher mapping resolution for studies focused on understanding the genetic underpinnings of quantitative traits (Zhu et al. 2008; Korte and Farlow 2013; Huang and Han 2014). Furthermore, high-throughput approaches involving image acquisition and analysis have also decisively favored GWAS, allowing the screening of large mapping populations, germplasm collections and other breeding resources with progressively greater precision and accuracy for several phenotypic traits, including seed attributes (Mir et al. 2019; Junker et al. 2015). Recently, Lei et al. (2020) successfully applied GWAS to search for associations between 116 simple sequence repeat (SSR) markers and four traits related to bean seed size. Nevertheless, there are still a lack of reports combining GWAS with high-density single nucleotide polymorphisms (SNPs) and high-throughput seed phenotyping to dissect common bean seed morphology including not only seed size and weight, but also other morphological aspects.

The aim of this study was to perform GWAS on a core collection of 180 common bean accessions image-based phenotyped, and genotyped with 10,362 SNPs to dissect the genetic architecture and identify genomic regions associated with seven seed morphological traits related to weight (hundred seed mass), size (width, length, projected area and perimeter) and shape (circularity and length-to-width ratio). Finally, we searched within the genome-associated regions for candidate genes putatively involved in seed phenotypic variation.

3.2. Materials and methods

3.2.1. Plant material and growing conditions

A core collection of one hundred eighty common bean genotypes was used in this study. This core collection represents the genetic diversity of over 1,800 accessions from the Agronomic Institute of Campinas (IAC), Brazil (Perseguini et al. 2015), one of the largest Brazilian *P. vulgaris* germplasm. It has already been highlighted for the variability of many traits of interest to breeders and consumers, such as seed morphology, biotic and abiotic stress resistance and nutritional value (Perseguini et al. 2015; Diniz et al. 2019) and were previously classified according to commercial group, origin and type of phaseolin (Diniz et al. 2019).

The seeds used for phenotypic evaluation were obtained from plants grown in the greenhouse in a completely randomized experimental design with two replicates, conducted in Piracicaba (22°42'30" S and 47°38'00" W), State of São Paulo, Brazil, during the 2017-2018 growing season (December 2017 to May 2018). Briefly, seeds were germinated in a biochemical oxygen demand (BOD) incubator at 26 °C for 5 days prior to transplanting. Three seedlings were transferred to 3L pots containing soil and sand (3:1). After establishment, only two plants were maintained per pot, being fertilized using 5N–10P–30K according to crop recommendations and irrigated periodically until harvesting.

3.2.2. Image acquisition and processing

In order to evaluate seed size, shape and weight, seven specific traits were measured on 50 seeds randomly selected from each plot. Seed size was determined by measuring seed length (LENGTH), width (WIDTH), perimeter (PERIM) and projected area (AREA). Regarding the shape aspect, we calculated the length-to-width ratio (LWR) and circularity (CIRC). Finally, in relation to seed weight, hundred seed mass (HSM) was estimated.

Seed size and shape properties were evaluated using high-throughput image-processing. Seeds were photographed in a dark room with supplementary light source using a digital SLR camera (EOS 1100D, Canon, Tokyo, Japan) fixed perpendicularly on an image acquisition station, 35 cm above the seeds. The background color was either white or blue, to contrast with the tegument color. A 10 mm crossmark was added to the background to convert from pixels to mm and indicate actual seed size on the image (Figure 6). HSM in grams was determined on an analytical balance (Mark 1300, Technal, Brazil). A total of 18,000 seeds were phenotyped.

Figure 6. High-throughput image-based phenotyping of common bean seed size and shape. Representation of the phenotypic diversity of the IAC core collection panel (a). Top (a) and front view (b) of the image acquisition station. Images of the seeds on a white background with a crossmark of 10 mm (d). Image after processing by the SmartGrain software (e). Single seed being measured for shape parameters (f)

Images were analyzed using the SmartGrain software (Tanabata et al. 2012), which can automatically detect every seed in the image and assess size and shape proprieties being recommended to high-throughput seed phenotyping in genetic mapping studies. The software processes images rendering

sequential coordinates along the seed perimeter. Briefly, the first step is to load the image and convert it from 24-bit (full color) to 1-bit (black and white) using a segmentation method, as detailed in Tanabata et al. (2010). Next, OpenCV functions “cvErode” and “cvDilate” are used to analyze seed morphology and exclude features such as awn, pedicel (for Poaceae) and funicle rest (for Fabaceae). It then detects the outlines, labels all seeds and computes AREA, PERIM and center of gravity using the “cvFindContour” function (Suzuki and Abe, 1985), and automatically finds a set of perimeter coordinates for each seed in the image using: $P_i = (x_i, y_i)$ for $i = 0, 1, \dots, n$; where the origin (0,0) for all P_i is the image top-left corner. The “cvContourArea” function then computes the area within the set of perimeter coordinates for each seed, and “cvArcLength” computes the PL (OpenCV Developers Team 2012). To calculate the LENGTH and WIDTH of each seed, the longitudinal and transversal axes are estimated from the set of perimeter points. For LENGTH, the algorithm detects the longest distance between two points on the perimeter coordinates and WIDTH is estimated based on the longest segment perpendicular to LENGTH. Finally, LWR and CIRC are calculated according to the following expressions:

$$LWR = \frac{LENGTH}{WIDTH}$$

and

$$CIRC = \frac{4\pi(AREA)}{PERIM^2}$$

In our study, all the JPEG images were stored in a dedicated folder and batch analyzed. Finally, output images were manually checked to correct possible errors.

3.2.3. Phenotypic data analysis

Linear mixed models were performed to estimate the variance components and obtain adjusted means using restricted maximum likelihood (REML), implemented in ASReml-R software (Butler et al. 2009):

$$Y_{ij} = \mu + g_{i(j)} + \rho_j + \varepsilon_{ij}$$

where Y_{ijk} is the phenotypic value of genotype i within gene pool j ; μ is an intercept; $g_{i(j)}$ is the random effect of genotype i within gene pool j , where $g_{i(j)} \sim N(0, \sigma_g^2)$; ρ_j is the fixed effect of gene pool j ; and ε_{ij} is the experimental error associated with genotype i within gene pool j , where $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Likelihood ratio tests (LRT) were accomplished to verify the statistical significance of the genotype factor and the Wald test were applied to verify the significance of gene pool. The best linear unbiased predictors (BLUPs) were applied to compute the genotypic correlations using Pearson's coefficient, and standardized to perform principal component analysis (PCA). Broad sense heritabilities for all seven seed traits were estimated as the ratio of genotypic to phenotypic variance, expressed as a

percentage using the formula: $H^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2} * 100$, where $\hat{\sigma}_g^2$ is the genotypic variance and $\hat{\sigma}_e^2$ is the residual variance.

3.2.4. Genotypic data

Genotyping by sequencing (GBS) approach has been previously accomplished to genotype the IAC core collection panel, as reported by Diniz et al. (2019). The procedures were carried out at the Genomic Diversity Facility of Cornell University's Institute of Biotechnology. Briefly, DNA was extracted from young leaves using the DNeasy® 96 Plant Kit (Quiagen, Hilden, Germany) following the manufacturer's protocol. Libraries were prepared using restriction enzyme ApeKI (5' CWGC 3'), and barcode adapters ligated to each individual sample. The samples were then pooled and amplified by PCR as described by Elshire et al. (2011), establishing two libraries in 95-plex. Single-end sequencing of up to 100 bp was carried out on the HiSeq 2500 platform (Illumina, San Diego, CA, USA) in a single lane. Sequences from the two libraries can be downloaded from the GenBank database, BioProject PRJNA336556, BioSamples SAMN05513252 and SAMN05513251. Bioinformatics procedures for SNP calling were performed by processing the GBS raw sequences through the TASSEL-GBS pipeline, with the *P. vulgaris* (G19833) genome as reference (Schmutz et al. 2014). SNP data were filtered using the following selection criterion: minimum coefficient of inbreeding of 0.9, minor allele frequency (MAF) ≥ 0.05 , call rate < 0.9 . SNPs, and setting heterozygous loci to missing data. A total of 10,362 SNPs were detected and are available in Diniz et al. (2019).

3.2.5. GWAS and candidate genes search

GWAS on the 10,362 SNP markers and the seed traits were performed by using the multi-locus mixed-model approach (MLMM; Segura et al. 2012), accounting for kinship matrix, implemented in the GAPIT-R package (Lipka et al. 2012). In short, this method comprises in the model significant effects via a forward-backward stepwise approach. At each step, the variance components are re-estimated. If the freshly included fixed effects really has a true influence, they might significantly reduce the residual variance and efficiently diminish the restrictions imposed by the mixed model on other markers correlated with population structure. This in turn improves statistical power and results in lower false discovery rate in compared with the single-marker scan and stepwise linear regression, especially if a conservative threshold such as Bonferroni (α /number of markers) is used. To avoid bias ascribed to double-shrinkage, best linear unbiased estimators (BLUEs) of the genotypic values for each trait were used in the association analysis.

The proportion of variance explained (PVE) by each association (SNP and seed trait) was calculated using the expression: $VQTL = V_{pheno}$, where $VQTL = 2freq(1 - freq)effect^2$ and V_{pheno} is the variance of the adjusted means for each trait.

The squared coefficient of correlation (r^2) between the significant SNPs and neighboring SNPs (1 MB up and downstream) was computed in order to establish the linkage disequilibrium (LD) block inherited together with the associated SNP or inherited more frequently than could be expected by randomness. For that, we used the LDheatmap R-package (Shin et al. 2006) following the expression:

$$r^2 = \frac{(P_{AB} - P_A P_B)^2}{P_A(1 - P_A)P_B(1 - P_B)}$$

where P_A is the allelic frequency of A; P_B is the allelic frequency of B, and P_{AB} is the allelic frequency of the AB haplotype. The genomic intervals investigated for candidate genes was then defined based on the LD block, as proposed by Assefa et al. (2019).

Protein sequences encoded by genes located in genomic intervals were obtained from the *P. vulgaris* genome (Schmutz et al. 2014), available in the NCBI Assembly database. For functional annotation, these sequences were subjected to a BLASTp search against the RefSeq database (O’Leary et al. 2015) by running Blast2GO v.5 (Conesa et al. 2005) with an e-value cutoff of 1×10^{-5} and a maximum of 20 hits. In addition, we used Gene Ontology (GO) mapping to assign GO terms by running Blast2GO with default parameters. Finally, aiming to assess the expression patterns of these genes during seed development, we analyzed the common bean expression atlas (O’Rourke et al. 2014). The normalized expression data (reads/Kb/Million; RPKM) were obtained from eight tissues: young flowers (FY), young pods (PY), pods approximately 9 cm long (PH), pods between 10 and 11 cm long (P1), pods between 12 and 13 cm long (P2), heart stage seeds (SH), stage 1 seeds (S1), and stage 2 seeds (S2). Moreover, in order to investigate the relative gene expression among these tissues, we calculated the Z-score according to the expression: $Z = (x - \mu)/\sigma$, where x is the RPKM value for each tissue, μ is the mean of RPKM values across tissues, and σ is the standard deviation. The application of this classical normalization allowed us to standardize RPKM values across a range of experiments, conditions and tissues, and contrast independently collected data (Cheadle et al. 2003).

3.3. Results

3.3.1. Phenotypic analysis

Phenotypic evaluation of the core collection panel revealed wide variations for the whole set of seed traits (Figure 7; Appendix H). The descriptive statistics of LENGTH, WIDTH, PERIM, AREA, LWR, CIRC and HSM are summarized in Table 7. Examining the dispersion of phenotypic data, broad amplitudes were verified, especially for AREA, PERIM and HSM. Coefficients of variation (CV) between different replicates for all the traits had small to medium values, ranging from 3.1% (CIRC) to 22.6% (AREA), indicating good experimental precision. The only traits with CVs $> 10\%$ were AREA and HSM, with respective values of 22.6% and 17.2%. Similarly, high broad-sense heritabilities (H^2) were verified for all traits, varying from 75.9% (CIRC) to 90.9% (HSM). These results show that much of the observed

phenotypic variation can be attributed to the genetic component, supporting the use of the core collection for GWAS purposes.

Figure 7. Frequency distribution of seven seed traits, namely seed length (a; LENGTH); width (b; WIDTH); perimeter (c; PERIM); projected area (d; AREA); length-to-width ratio (e; LWR); circularity (f; CIRC) and hundred seed mass (g; HSM) for the 180 genotypes of the IAC core collection panel

The suitability of the core collection for the GWAS was also supported by the LRT, which confirmed a significant genotype effect for the seven seed traits, reinforcing the indication of genetic factors underlying seed in common bean (Table 8). Of the 180 accessions evaluated, “Apetito Blanco”, “Branco Argentino” and “Jalo 110” were highlighted for large seed size and high weight, while “Sanilac”, “Rio Tibagi” and “IAC Carioca Arua” exhibited some of the lowest values. In terms of shape (CIRC and LWR), “Garbancillo”, “Sanilac” and “Vermelhinho” presented the most circular seeds and “Red Kidney”, “Jalo” and “Cal 143”, the most elongated (Appendix H).

Table 7. Descriptive statistics and variance components, genetic coefficient of variation and broad-sense heritability of seven seed traits in the IAC common bean core collection panel

Seed Trait ^a	Amplitude	General Mean	Mean Mesoamerican	Mean Andean	Genetic variance	Residual variance	CV ^b (%)	H ² ^c (%)
LENGTH	7.52-15.0	10.47	10.353	11.217	0.8956	0.1314	10.0	87.2
WIDTH	5.14-9.52	6.70	6.677	6.825	0.2246	0.0465	7.7	82.84
PERIM	21.52-40.04	29.44	29.163	31.252	6.0249	0.8724	9.2	87.35
AREA	31.45-100.39	55.74	54.825	61.723	77.1369	11.2812	22.6	87.24
LWR	1.33-2.02	1.57	1.555	1.637	0.0102	0.0019	7.2	84.29
CIRC	0.65-0.85	0.80	0.804	0.793	0.0005	0.0001	3.1	75.94
HSM	16.62-55	26.18	25.437	30.862	30.5645	3.0638	17.2	90.89

^a Seed length (LENGTH, in mm); width (WIDTH, in mm); perimeter (PERIM, in mm); projected area (AREA, in mm²); length-to-width ratio (LWR); circularity (CIRC); Hundred seed mass (HSM, in g). ^b CV: Coefficient of variation.

^c Broad-sense heritability

Significant gene pool effects were also detected for all traits ($p < 0.01$) except WIDTH ($p = 0.12$; Table 8). The means for the Andean gene pool were higher than those for the Mesoamerican pool for LENGTH, PERIM, AREA, LWR and HSM, indicating bigger and heavier seeds. On the other hand, the seeds of Mesoamerican genotypes were more circular and less elongated (Table 7; Figure 8). It is important to note that, despite significant differences were detected among gene pools, there was also considerable variation within the pools (Figure 8; Appendix H).

Table 8. Likelihood ratio test (LRT) and Wald test of random and fixed effects for seven seed traits in the IAC common bean core collection panel

Seed trait ^a	Gene pool effect		Genotype effect	
	Wald test	p-value	LRT ^b	p-value
LENGTH	20.284	<0.0001	248.69	<0.0001
WIDTH	2.448	0.12	200.71	<0.0001
PERIM	17.651	<0.0001	250.79	<0.0001
AREA	15.069	<0.0001	246.89	<0.0001
LWR	16.758	<0.0001	215.16	<0.0001
CIRC	7.279	0.00698	145.29	<0.0001
HSM	23.094	<0.0001	246.35	<0.0001

^a Seed length (LENGTH); width (WIDTH); perimeter (PERIM); projected area (AREA); length-to-width ratio (LWR); circularity (CIRC); Hundred seed mass (HSM). ^b LRT: likelihood ratio test

Figure 8. Boxplot of seven seed traits, namely seed length (a; LENGTH); width (b; WIDTH); perimeter (c; PERIM); projected area (d; AREA); length-to-width ratio (e; LWR); circularity (f; CIRC) and hundred seed mass (g; HSM), in 180 genotypes of the Andean and Mesoamerican gene pools

3.3.2. Genotypic correlation and principal component analysis

Pairwise genetic correlation analysis enabled us to detect at least three groups of correlated traits. Pearson coefficients were strongly positive between all pairs of seed size traits, including weight (Figure 9a and b). Among these traits, correlations varied from 0.69 (WIDTH-LENGTH) to 0.98 (AREA-PERIM). For instance, the strongest correlations were found between PERIM-AREA, PERIM-LENGTH and AREA-LENGTH (0.98, 0.98 and 0.95, respectively). In terms of seed shape (LWR and CIRC), distinct behavior was detected as even though they showed strong negative correlation each other (-0.77) their

relationship with the remaining traits differed, as shown by the network plot (Figure 9b). LWR showed a moderately positive correlation with LENGTH, PERIM, AREA and HSM, and a weak negative correlation with WIDTH. In contrast, CIRC was moderate negatively correlated with all seed traits except WIDTH, where a near zero value was found (Figure 9).

Figure 9. Heatmap of pairwise Pearson correlation coefficients (a) and network plot (b) of seven seed traits: seed length (LENGTH); width (WIDTH); perimeter (PERIM); projected area (AREA); length-to-width ratio (LWR); circularity (CIRC) and hundred seed mass (HSM) in the 180 genotypes of the IAC core collection panel

Principal component analysis was performed to ascertain the key sources of variation in the set of seed traits and to check whether the seven evaluated traits were consistent for differentiating the genotypes according to gene pool origin. The first two principal components (PC) accounted for 93.58 % of the variation, in which PC1 mainly described the variation in LENGTH, PERIMETER, AREA, WIDTH and HSM, and PC2 captured most of the variance in CIRC and LWR (Figure 10), corroborating the correlation clusters verified by the Pearson coefficient (Figure 9b). The strong negative correlation between CIRC and LWR can also be highlighted by PCA, since the loadings representing these traits point in opposite directions. Finally, although it was not possible to clearly cluster the gene pools, most of the genotypes that stood out in terms of AREA, HSM, PERIM, LENGTH and LWR, were of Andean origin.

Figure 10. Principal component analysis (PCA) of the IAC common bean core collection panel based on the phenotypic variability of seven seed traits: seed length (LENGTH); width (WIDTH); perimeter (PERIM); projected area (AREA); length-to-width ratio (LWR); circularity (CIRC) and hundred seed mass (HSM). Orange and blue dots indicate accessions of Andean and Mesoamerican origin, respectively

3.3.3. Genome-Wide Association Studies

GWAS were conducted on a total of 10,362 polymorphic SNPs for the seven seed traits. As shown in the quantile–quantile (QQ) plots (Figure 11b), according to the MLM model the observed p-values are closely adhered to the expected values under the null hypothesis for all traits, indicating that there were no systematic false positive associations.

Figure 11. GWAS on seven common bean seed traits for the 180 genotypes of the IAC core collection panel. Seed length (LENGTH) is shown in red; width (WIDTH) in green; perimeter (PERIM) in blue; projected area (AREA) in light blue; length-to-width ratio (LWR) in pink; circularity (CIRC) in yellow and hundred seed mass (HSM) in grey. The outer ring in the Circular Manhattan plot (a) and the heatmap bars in the Joint Manhattan plot (c) depict SNP density along the eleven *P. vulgaris* chromosomes, and the dashed line indicates the Bonferroni threshold ($p < 4.82E-06$). In the Quantile-Quantile (QQ) plot (b) the shadow area indicates a 95% confidence interval

Seven marker-trait associations with p -values lower than the Bonferroni threshold ($p < 4.82E-06$) were considered significant for LENGTH, WIDTH, AREA, PERIM and CIRC, and were located on four distinct genomic regions of three chromosomes (Pv02, Pv08 and Pv11; Figure 11). Of the seven associations, two SNPs were associated with more than one trait. For LENGTH, the two significantly associated SNPs (S1_325844152 and S1_507205534) were positioned on loci 1797591 bp and 42751382 bp of chromosomes Pv08 and Pv11, respectively. These SNPs had effects of similar magnitude (0.69 and -0.67 mm) and explained 5.94 and 6.39% of the phenotypic variation. The SNP S1_325844152 was also detected as associated with PERIM and CIRCULARITY. For PERIM, it accounted for 7.16% of the phenotypic variation, while for CIRC the PVE association was as high as 17.6% (Table 9; Figure 11; Appendix I).

Table 9. Single-nucleotide polymorphisms (SNP) significantly associated with seed length (LENGTH), width (WIDTH), area (AREA), perimeter (PERIM) and circularity (CIRC), according to the GWAS results for the IAC core collection panel

Seed Trait ^a	Marker name	Chr ^b	Position (bp)	p-value	MAF ^c	Allelic reference	Allelic variant	Effect of the variant allele	PVE ^d (%)
LENGTH	S1_325844152	Pv08	1797591	3.61E-06	0.07	T	C	0.69	5.94
	S1_507205534	Pv11	42751382	4.28E-06	0.08	G	A	-0.67	6.39
WIDTH	S1_53011276	Pv02	805645	6.64E-08	0.11	A	C	0.55	23.98
AREA	S1_53011276	Pv02	805645	1.22E-10	0.11	A	C	10.86	23.3
	S1_379992973	Pv08	55946412	7.91E-07	0.08	C	T	6.01	5.18
PERIM	S1_325844152	Pv08	1797591	3.36E-07	0.07	T	C	1.95	7.16
CIRC	S1_325844152	Pv08	1797591	5.11E-08	0.07	T	C	-0.023	17.6

SNPs significantly associated according to the Bonferroni multiple test correction ($p < 4.82E-06$); ^a Seed length (LENGTH); width (WIDTH); perimeter (PERIM); projected area (AREA); and circularity (CIRC). ^b Chr: Chromosome; ^c MAF: Minor allele frequency; ^d PVE: Proportion of variance explained by the SNP–trait association

AREA-associated SNPs (S1_53011276 and S1_379992973) were located on Pv02 and Pv08 in the respective 805645 bp and 55946412 bp positions. For this trait, the effect of S1_379992973 was 6.01 mm² with and PVE of 5.18%, while S1_53011276 explained even a higher portion of the phenotypic variance (23.3%), showing additive effect of 10.86 mm², mainly resulting from its influence on seed width. For WIDTH, this SNP also had large explanation of the total phenotypic variance (23.7%) and the variant allele affected the trait in 0.55 mm. No significant associations were detected for LWR and WEIGHT (Table 9; Figure 11).

3.3.4. Candidate genes

With the aim of delimitating the genomic window for candidate gene investigation, LD analyses were performed bordering the significant SNP–seed trait associations. The genomic intervals varied between 358.4 kb for S1_53011276, and 558.6 kb for S1_507205534 (Appendix J). Appendix K gives a complete list of 192 genes predicted in the set of associated regions, 188 of which were annotated; 50 in the genomic region harboring SNP S1_53011276; 64 in S1_325844152; 45 in S1_325844152; and 29 in S1_507205534. To give a clearer picture of the potential functions of these genes, functional groups were classified based on the Gene Ontology (GO) database. Among the inferred genes, 32 have no functional annotation or were classified as unknown function protein-coding genes.

In order to narrow down the number of candidate genes and obtain further evidence of their importance in regulating common bean seed size and shape, we consulted the *P. vulgaris* expression atlas for eight tissues and stages related to seed development (O’Rourke et al. 2014), extending from flowering to late phase seed formation. Among the 188 annotated genes, 156 were expressed in one or more tissues or

stages (Appendix L) and 13 were considered the most promising (Table 10) with homologs already reported as putatively involved in seed development or related traits in crops or model species.

In the LD block containing SNP S1_53011276, associated with WIDTH and AREA, several genes could be pinpointed as promising candidates (Table 10). S1_53011276 is positioned within an exon of Phvul.002G007100, which encodes a transcriptional regulator *SUPERMAN*, well-known for playing a role in Arabidopsis reproductive organ morphology (Sakai et al. 1995, 2000). In terms of the Z-score, substantial negative values were observed for this gene in most tissues, indicating that compared to the whole set of genes and tissues analyzed herein, Phvul.002G007100 was the one with lowest general expression (Appendix M). Note that the expression of this gene was verified in the three stages of seed development (S1 and S2), being more expressed stronger in the two later stages (S1; RPKM=69 and S2; RPKM=53). Likewise, in this genomic region, Phvul.002G004400 encodes for a pentatricopeptide repeat-containing protein, already reported as a prominent player in seed development in various crops (Gutiérrez-Marcos et al. 2007; Liu et al. 2013a; Li et al. 2014; Verma et al. 2015) and highly expressed at the initial stage of seed development (SH; RPKM=122) (Appendix L). Phvul.002G006100 is annotated as a protein phosphatase 2C 29-like gene and had remarkable expression in all the eight evaluated tissues, with RPKM varying from 111 in P1 to 2029 in SH, and particularly high Z-scores for SH (1.22), S1 (1.57), and S2 (1.66) (Appendix L and M). Finally, Phvul.002G006700 is a gene predicted as an AT-hook motif nuclear-localized protein 8 which has been shown to underlie seed size in other species (Sharma and Michael 2020), with high RPKM values for FY, PH and the three seed stage tissues (Appendix L).

Associated with LENGTH, PERIM and CIRC, SNP S1_325844152 is located within an exon of Phvul.008G020600, a gene that encodes a crowded nuclei 1 protein. Consistent with the expression atlas, this gene exhibited the greatest expression for FY, PV, PH, P1, P2 and SH, and also the highest Z-scores for most tissues, especially in the early developmental stages (Appendix M). In the same haplotype block, a homolog of Phvul.008G019500, a protein *MEI2*-like 4 isoform, was recently reported as a modulator of grain size and weight in *Oryza sativa* (Lyu et al. 2020) and was highly expressed in common bean during the early seed developmental stage (SH; RPKM=488). In addition, this region comprises several genes encoding for pentatricopeptide repeat-containing proteins (Phvul.008G022600, Phvul.008G023900, Phvul.008G024300 and Phvul.008G024400). Among them, Phvul.008G022600 and Phvul.008G024400 stand out due to the high RPKM values along all tissues checked herein (Appendix L). Furthermore, in Pv08, significant for AREA, the S1_379992973 LD region contains a putative B3 domain-containing protein (Phvul.008G244300) that is worthy of attention since the B3 domain has been found in transcription factors involved in seed development control (Swaminathan et al. 2008). Finally, SNP S1_507205534, significantly associated with LENGTH, is positioned in an exonic region of Phvul.011G163100, a ribosomal protein L11 methyltransferase whose expression is highlighted in SH (RPKM=59) and S1 (RPKM=48) tissues (Appendix L). This SNP has an LD with a genomic locus that exhibits other promising candidate genes, such as Phvul.011G160400, a protein cup-shaped cotyledon 2-like with high expression in FY (RPKM=205) and PH (RPKM= 275) tissues (Appendix L).

Table 10. Candidate genes identified by the GWAS-associated genomic regions able to modulate seed shape and size in common bean

Seed Trait ^a	Associated SNP	Gene ID	Functional annotation	Genomic location
WIDTH and AREA	S1_53011276	Phvul.002G004400	pentatricopeptide repeat-containing protein	Pv02:765463..766998
		Phvul.002G006100	phosphatase 2C 29-like protein	Pv02:516568..518904
		Phvul.002G006700	AT-hook motif nuclear-localized protein 8	Pv02:656431..661984
		Phvul.002G007100	transcriptional regulator SUPERMAN	Pv02:698486..703324
LENGTH, PERIM and CIRC	S1_325844152	Phvul.008G019500	<i>MEI2</i> -like 4 isoform X1 protein	Pv08:1727505..1734392
		Phvul.008G020600	crowded nuclei 1 protein	Pv08:1626867..1635158
		Phvul.008G022600	pentatricopeptide repeat-containing protein	Pv08:1854380..18556627
		Phvul.008G023900	pentatricopeptide repeat-containing protein	Pv08:1974646..1978175
		Phvul.008G024300	pentatricopeptide repeat-containing protein	Pv08:1999723..2003178
		Phvul.008G024400	pentatricopeptide repeat-containing protein	Pv08:2004322..2008280
AREA	S1_379992973	Phvul.008G244300	putative B3 domain-containing protein	Pv08:59274246..59279462
LENGTH	S1_507205534	Phvul.011G160400	cup-shaped cotyledon 2-like protein	Pv11:45614432..45616861
		Phvul.011G163100	ribosomal protein L11 methyltransferase	Pv11:46181028..46188144

SNPs significantly associated according to the Bonferroni multiple test correction ($p < 4.82E-06$); ^a Seed length (LENGTH); width (WIDTH); perimeter (PERIM); projected area (AREA); and circularity (CIRC)

3.4. Discussion

Plant domestication can be understood as long-term experimentation involving the multifaceted relationships between environmental, anthropological and evolutionary forces (Gepts 2004; Chacón-Sánchez 2018). The nature and number of domestication events, along with the magnitude of the domestication bottleneck and the genetic basis of the adaptive courses, are vital in breeding since they offer extraordinary possibilities for improving existing crops. In particular, the domestication process in *P. vulgaris* has been very widely studied, promising to shed light on these issues, especially those linked to the domestication syndrome. Originating in Mexico (Bitocchi et al. 2012, 2013), the species spread to South America, resulting in the formation of the Mesoamerican and Andean gene pools (Gepts et al. 1986; Koinange and Gepts 1992). Dual domestication had important impacts on traits related to the decrease in pod shattering and improvement in seed size. These traits merit singular attention because of their role in early adaptation and importance for plant breeding (Chacón-Sánchez 2018). In addition, common bean seed size combined with shape aspects are highlighted due to their correlation to yield (Adams 1967; Nienhuis and Singh 1986; White and González 1990), apparent relationship with seed vigor (Ambika et al. 2014), and attractiveness to consumers, industry and farmers (Nienhuis and Singh 1986; Kara et al. 2013; FAO and AfricaSeeds 2018; Shaw et al. 2020).

Herein we studied the genetic architecture of seven common bean seed traits including those related to size (LENGTH, WIDTH, PERIM and AREA), shape (LWR and CIRC) and weight (HSM). For that, we used a core collection panel of 180 genotypes derived from one of the largest Brazilian germplasms. We observed wide variability for all the traits analyzed and coefficients of variation were predominantly low

(<10%), indicating that image analysis of seed properties lead to slight experimental error, similarly as reported by Fıratlıgil-Durmuş et al. (2010). Significant genotype effects were observed for all traits, what jointly with the high broad-sense heritabilities (>75.9%) demonstrates that a substantial portion of phenotypic variation might be ascribed to the genetic component, highlighting its importance in the expression of these quantitative traits, and thus lending weight to the use of this core collection for GWAS (Table 7). These results corroborate the findings of several other studies on bean seed morphology, also reporting high heritabilities (Motto et al. 1978; Yuste-Lisbona et al. 2014; Rana et al. 2015; Singh et al. 2018; Wu et al. 2019).

Even though both domestication events have increased seed size and weight compared to their wild ancestors (Gepts 2004; Chacón-Sánchez 2018), Andean genotypes are acknowledged for producing larger seeds than Mesoamericans (Gepts and Bliss 1988; Singh et al. 1991b; Logozzo et al. 2007; Lei et al. 2020). Evaluating the IAC core collection panel, we once again confirmed the larger seed size of Andean genotypes according to LEGNTH, PERIM, AREA and HSM. Anyhow, there were no significant differences in WIDTH ($p = 0.12$) exhibited by the individual gene pools. These findings suggest that the main reasons for the seed size discrepancy between the pools are related to the shape aspect of longitudinal development, an observation supported by the higher LWR of the Andean seeds and the higher CIRC values for Mesoamerican genotypes.

Corroborating the other evidences, the common bean core collection herein evaluated presented strong positive pairwise correlations among seed size traits (Pañeda et al. 2010; Rana et al. 2015; Herron et al. 2020; Lei et al. 2020; Murube et al. 2020). We verified that the overall size, considering both weight and projected area, was influenced principally by an increase in the length dimension. Even though less pronounced, we observed a negative correlation between CIRC and all the other size traits except WIDTH, suggesting that small seeds are usually rounder while larger seeds tend to be elongated. Hence, our results show that an extension in length does not necessarily imply an increase in the distance between one side and the other at the hilum region. Consequently, since WIDTH shows lower genetic variance (Table 7), LENGTH is suggested to be the main seed size parameter affecting shape proprieties. PCA provided further insights into the relationships among seed traits. PC1 accounted for over 70% of total variability, explaining mainly the variation in seed size traits, while for PC2, the most prevalent traits were related to seed shape. Although PCA was not capable to clearly grouping the gene pools, Andeans represented the majority of the genotypes positioned away from the intersection of the biplot were found to be Andean.

The GWAS approach is fundamentally based on the LD since once its objective is to identify genetic markers physically linked to the actual causal variant of given trait. Nonetheless, several other factors over and above physical linkage may influence LD estimates. In the populations typically used for GWAS, related individuals share both causal and non-causal alleles, and thus LD between these loci can result in spurious associations (Korte and Farlow 2013). For this reason, kinship and population structure are usually modeled in the association analysis (Zhao et al. 2007; Zhang et al. 2010). For the core collection used herein, Diniz et al. (2019) studied LD and showed that kinship bias is heavier than the bias ascribed by population

structure. Additionally, these authors proved that when kinship is considered, population structure bias is no longer observed in the collection. Here, we confirmed these results and demonstrated that, for all seed traits, when fitting kinship to the MLM models, p-values exhibited low general inflation, since they were closely adhered to the expected results under null hypothesis, suggesting goodness of fit and low rates of false discovery (Figure 11b).

Regarding the genetic architecture of common bean seed size and shape, our results indicate that, although exhibiting quantitative behavior, some genomic regions do account for a considerable amount of variation. GWAS analysis allowed us to identify seven marker associations with moderate to high effects for five traits (LENGTH, WIDTH, AREA, PERIM and CIRC) on four regions of chromosomes Pv02, Pv08 (2) and Pv11. On Pv02 at 805645 bp, S1_53011276 was significantly associated with WIDTH and AREA and explained over 23% of the phenotypic variance. This SNP is a promising candidate for marker-assisted selection, especially since the beginning of chromosome Pv02 has also been reported as linked to seed width (Geng et al. 2017, Wu et al. 2019; Geravandi et al. 2020) and other seed traits (Blair et al. 2006; Pañeda et al. 2010; Lei et al. 2020; Murube et al. 2020). Our findings also endorse Pv08 as a key chromosome underlying the size and shape aspects of common bean seeds. At positions 1797591 and 55946412 bp, two SNPs (S1_325844152 and S1_379992973) were associated with four traits (LENGTH, AREA, PERIM and CIRC). Among these associations, it is worthwhile stressing the high PVE (17.6%) found between S1_325844152 and CIRC. Remarkably, this chromosome has already been reported to carry seed size QTLs (Park et al. 2000; Blair et al. 2006; Pañeda et al. 2010; Wright and Kelly 2011; Geravandi et al. 2020). Since there is still no genetic mapping studies for CIRC in common bean and considering that it exhibits a marked LENGTH influence, the high effect QTLs previously reported for seed length (Park et al. 2000; Pañeda et al. 2010) not only confirm the role of Pv08 in determining LENGTH but also point toward to its contribution to CIRC. Finally, at position 42751382 bp of Pv11, the association between LENGTH and SNP S1_507205534 corroborates previous findings that reported QTLs underlying seed length at a similar location in the same chromosome (Park et al. 2000; Murube et al. 2019).

In order to define the genomic window considered for the search for candidate genes, the loci spanning significant associations were investigated based on LD (Appendix J). Of the 192 genes predicted in those regions, 188 were annotated and 156 classified based on the Gene Ontology database (Appendix K) and expressed in one or more tissues related to seed development according to the *P. vulgaris* expression atlas (O'Rourke et al. 2014). Herein we have highlighted some of the most promising candidates that may influence WIDTH, AREA, LENGTH, PERIM and CIRC.

For WIDTH and AREA, homologs of Phvul.002G004400, Phvul.002G006100, Phvul.002G006700 and Phvul.002G007100 have been reported in the literature as involved in the morphology of seed and related tissues. Interestingly, the significantly-associated SNP S1_53011276 was positioned within an exon of Phvul.002G007100. Although this does not necessarily indicate a cause-effect relationship, this gene is a good candidate since it encodes a transcriptional regulator *SUPERMAN*, one of the best characterized plant *TFIIIA* zinc finger proteins associated with plant development, including leaf, shoot and floral organ morphogenesis and gametogenesis. According to the atlas, it was not verified strong

expression of this gene, which was noticeably expressed only in late seed development (S1 and S2) and thus had high negative Z-scores compared to the whole gene set (Appendix M). Similarly, Huang et al. (2006) indicated that a *SUPERMAN*-like gene, namely *GmZFP1*, was intronless and expressed principally in late developing seeds and reproductive organs of soybean. The authors further suggested that this gene might act through a *MADS*-box transcription factor mediated signal network, as occurs in *Arabidopsis*. In the same LD block, Phvul.002G004400, consistently expressed at the initial stage of seed development (SH), belongs to a protein family recognized as a pentatricopeptide repeat-containing protein, a modular RNA-binding protein which intermediates numerous aspects of gene expression, such as RNA processing, splicing, editing, stability and translation. Already associated with several phenotypes, this family is recognized as an important player in seed development in various crops (Gutiérrez-Marcos et al. 2007; Li et al. 2014; Verma et al. 2015; Lyu et al. 2020). In maize, empty pericarp5 and small kernel 1 mutants, both encoding pentatricopeptide repeat proteins, exhibited embryo abortion, small endosperm and stunted seed development in the early stages (Liu et al. 2013b; Li et al. 2014; Qi et al. 2017). Furthermore, many other related genes have also been reported as candidates for seed size and weight regulation in genetic mapping studies of many crops (Zhou et al. 2017; Du et al. 2019), including Fabaceae ones (Verma et al. 2015). Phvul.002G006100, predicted as a protein phosphatase 2C 29-like gene, presented remarkable expression in all the eight evaluated tissues. Phosphatase 2C proteins are crucial negative regulators of ABA signaling (Bai et al. 2013), and in soybean have been reported as important regulators for seed weight and size. Lu et al. (2017) used whole-genome sequencing of a recombinant inbred line population to map QTLs for seed weight, and noticed that a phosphatase 2C-1 allele from wild soybean had strong effect in increasing seed weight/size, what was also validated using transgenic plants. Finally, Phvul.002G006700, an AT-hook motif nuclear-localized protein 8, exhibited strong expression for FY, PH, and tissues at the three seed stages. For instance, Sharma and Michael (2020) showed that an AT-hook motif (*AtSOB3-D*) modulates seed size as well as hypocotyl length in brassica species *Camelina sativa* and *A. thaliana*.

Associated with LENGTH, PERIM and CIRC, the polymorphism of S1_325844152 is located inside an exon of Phvul.008G020600, a crowded nuclei 1 protein, which exhibited the strongest absolute expression and highest Z-scores in most tissues considered herein, especially in the flowering and early developmental stages. Although this gene has not yet been reported as a direct modulator of seed size, the control it exercises on nuclear size has been detailed described (Wang et al. 2013), as well as its role in ABA-controlled seed germination (Zhao et al. 2016). In the same genomic haplotype block, Phvul.008G019500, a protein *MEI2*-like 4 isoform, is noteworthy due to its strong expression in early stage seed development (SH). *MEI2*-like proteins with conserved RNA recognition motifs have been identified in several species of plant and seem to play a role in several functions during plant development (Anderson et al. 2004). In rice, an *MEI2*-like 4 protein (*OML4*) is phosphorylated by a glycogen synthase kinase 2 and negatively controls seed size and weight. Thus, loss of function of *oml4* mutants produces larger and heavier seeds, while the overexpression of this gene leads to smaller, lighter grains (Zhang 2020; Lyu et al. 2020). As observed in the association of S1_53011276 with WIDTH and AREA, this region comprises genes encoding for

pentatricopeptide repeat-containing proteins (Phvul.008G022600, Phvul.008G023900, Phvul.008G024300 and Phvul.008G024400), in which Phvul.008G022600 and Phvul.008G024400 were abundantly expressed in all the tissues investigated.

The region containing the association between AREA and S1_379992973, a putative B3 domain-containing protein (Phvul.008G244300), exhibited low transcriptional levels in the eight tissues considered herein. The plant-specific B3 superfamily encompasses well-studied protein families, including prominent transcriptional elements, such as auxin response factors (Swaminathan et al. 2008). Transcription factors associated with this domain have been proved to be involved in activating and repressing seed development and grain filling in several crops and model species (Peng and Weselake 2013; Guo et al. 2013; Ahmad et al. 2019; Yang et al. 2020). Finally, an association between SNP S1_507205534 and LENGTH was found in an exonic region of a ribosomal protein L11 methyltransferase gene (Phvul.011G163100). In this region a protein cup-shaped cotyledon 2-like gene (Phvul.011G160400) has a homolog previously reported to be involved in ovule development (Gonçalves et al. 2015). Aida et al. (1997) studied mutations in two homologs of this gene and found defects in the separation of cotyledons that can lead to notable modifications in seed shape.

As far as we know, this is the first study to combine high-density genotyping and high-throughput image-based seed analysis to perform association mapping and dissect the genetic architecture of both seed size and shape in common bean. Our findings indicate that even though the estimated genetic parameters for the core collection point to several genes controlling the traits, four genomic regions could explain a substantial amount of the phenotypic variation. From a practical standpoint, this study provides breeders with SNP markers that, given the concordance with the literature, are very good resources for use in marker-assisted selection. Moreover, several genes were pinpointed as interesting candidates underlying the phenotypic variation in LENGTH, WIDTH, AREA, PERIM and CIRC. These genes are promising choices for functional analysis, which could validate their influence on seed shape and size in common bean and eventually other related crops.

References

- Adams MW (1967) Basis of yield component compensation in crop plants with special reference to the field bean, *Phaseolus vulgaris*. Crop Sci 7:505-510. <https://doi.org/10.2135/cropsci1967.0011183X000700050030x>
- Ahmad B, Zhang S, Yao J, et al (2019) Genomic organization of the b3-domain transcription factor family in grapevine (*Vitis vinifera* L.) and expression during seed development in seedless and seeded cultivars. Int J Mol Sci. 18:4553. <https://doi.org/10.3390/ijms20184553>
- Aida M, Ishida T, Fukaki H, et al (1997) Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. Plant Cell 9:841-857. <https://doi.org/10.1105/tpc.9.6.841>
- Akibode S, Maredia M (2011) Global and regional trends in production, trade and consumption of food legume crops. Department of Agricultural, Food and Resource Economics: Michigan State University, 2011. <https://doi.org/10.22004/ag.econ.136293>
- Ambika S, Manonmani V, Somasundaram G (2014) Review on the effect of seed size on seedling vigor and yield. Res J Seed Sci 7:31-38. <https://doi.org/10.3923/rjss.2014.31.38>
- Anderson GH, Alvarez NDG, Gilman C, et al (2004) Diversification of genes encoding mei2-like RNA binding proteins in plants. Plant Mol Biol 54:653-670. <https://doi.org/10.1023/B:PLAN.0000040819.33383.b6>

- Assefa T, Mahama AA, Brown A V, et al (2019) A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). Mol Breed 39:1–23. <https://doi.org/10.1007/s11032-018-0920-0>
- Bai G, Yang D-H, Zhao Y, et al (2013) Interactions between soybean ABA receptors and type 2C protein phosphatases. Plant Mol Biol 83:651–664. <https://doi.org/10.1007/s11103-013-0114-4>
- Bellucci E, Bitocchi E, Rau D, et al (2014) Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Springer Netherlands, Dordrecht, pp 483–507. https://doi.org/10.1007/978-94-007-7572-5_20
- Bitocchi E, Bellucci E, Giardini A, et al (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. New Phytol 197:300–313. <https://doi.org/10.1111/j.1469-8137.2012.04377.x>
- Bitocchi E, Nanni L, Bellucci E, et al (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. PNAS 109:788–796. <https://doi.org/10.1073/pnas.1108973109>
- Bitocchi E, Rau D, Bellucci E, et al (2017) Beans (*Phaseolus* spp.) as a model for understanding crop evolution. Front Plant Sci 8:722–730. <https://doi.org/10.3389/fpls.2017.00722>
- Blair MW, Díaz LM, Buendía HF, Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). 955–972. <https://doi.org/10.1007/s00122-009-1064-8>
- Blair MW, Gonza LF, Kimani PM, Butare L (2010) Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. Theor Appl Genet 121:237–248. <https://doi.org/10.1007/s00122-010-1305-x>
- Blair MW, Iriarte G, Beebe S (2006) QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean X wild common bean (*Phaseolus vulgaris* L.) cross. Theor Appl Genet 112:1149–1163. <https://doi.org/10.1007/s00122-006-0217-2>
- Broughton WJ, Hern G, Blair M, et al (2003) Beans (*Phaseolus* spp.) – model food legumes. Plant Soil 55:55–128. <https://doi.org/10.1023/A:1024146710611>
- Butler D, Cullis B, Gilmour A, Gogel B (2009) Analysis of mixed models for S-language environments: Asreml–R reference manual. training and development series, No QE02001. QLD Department of Primary Industries and Fisheries, Brisbane.
- Chacón-Sánchez MI (2018) The domestication syndrome in phaseolus crop plants: A review of two key domestication traits. In: Pontarotti P (ed). Origin and evolution of biodiversity. 1st edn. Springer International Publishing, Cham, pp 37–59. https://doi.org/10.1007/978-3-319-95954-2_3
- Cheadle C, Vawter MP, Freed WJ, Becker KG (2003) Analysis of microarray data using Z score transformation. J Mol diagnostics 5:73–81. [https://doi.org/10.1016/S1525-1578\(10\)60455-2](https://doi.org/10.1016/S1525-1578(10)60455-2)
- Checa OE, Blair MW (2013) Inheritance of yield-related traits in climbing beans (*Phaseolus vulgaris* L.). Crop Sci 52:1998–2013. <https://doi.org/10.2135/cropsci2011.07.0368>
- Conesa A, Götz S, García-Gómez JM, et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Diniz AL, Giordani W, Costa ZP, et al (2019) Evidence for strong kinship influence on the extent of linkage disequilibrium in cultivated. Genes 10:1–16. <https://doi.org/10.3390/genes10010005>
- Du B, Wang Q, Sun G, et al (2019) Mapping dynamic QTL dissects the genetic architecture of grain size and grain filling rate at different grain-filling stages in barley. Sci Rep 9:1–16. <https://doi.org/10.1038/s41598-019-53620-5>
- Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- FAO, AfricaSeeds (2018) Seeds Toolkit-Module 2: Seed quality assurance, 2nd edn. Food & Agriculture Org., Rome
- FAOSTAT (2018) Food and agriculture organization of the united nations statistics division. Crop Prod. <http://www.fao.org/faostat/en/#data/QC>. Accessed 14 May 2020
- Fıratlıgil-Durmuş E, Šárka E, Bubník Z, et al (2010) Size properties of legume seeds of different varieties using image analysis. J Food Eng 99:445–451. <https://doi.org/10.1016/j.jfoodeng.2009.08.005>
- Gavazzi G, Sangiorgio S (2017) Seed size: an important yield component. In: Pulu R, Gavazzi G (eds) More food: Road survival 1st edn. Bentham booksm United Arab Emirates, pp 143–167.
- Geng Q-H, Wu J, Wang S-M, Wang L-F (2017) QTL Mapping for Seed Size and Shape in Common Bean. Acta Agron Sin 43:1149–1160. <https://doi.org/10.3724/SP.J.1006.2017.01149>
- Gepts P (2004) Crop domestication as a long-term selection experiment. Plant Breed Rev 24:1–44. <https://doi.org/10.1002/9780470650288>
- Gepts P, Aragão FJL, Barros E de, et al (2008) Genomics of phaseolus beans, a major source of dietary protein and micronutrients in the Tropics. In: Moore PH, Ming R (eds) Genomics of Tropical Crop Plants. Springer New York, New York, pp 113–143. https://doi.org/10.1007/978-0-387-71219-2_5
- Gepts P, Bliss FA (1988) Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. II. Europe and Africa. Econ Bot 42:86–104. <https://doi.org/10.1007/BF02859038>

- Gepts P, Osborn TC, Rashka K, Bliss FA (1986) Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): Evidence for Multiple Centers of Domestication. *Econ Bot* 40:451–468. <https://doi.org/10.1007/BF02859659>
- Geravandi M, Cheghamirza K, Farshadfar E, et al (2020). QTL analysis of seed size and yield-related traits in an inter-genepool population of common bean (*Phaseolus vulgaris* L.). *Scientia Horticulturae*, 274:109678. <https://doi.org/10.1016/j.scienta.2020.109678>
- Gonçalves B, Hasson A, Belcram K, et al (2015) A conserved role for CUP-SHAPED COTYLEDON genes during ovule development. *Plant J* 83:732–742. <https://doi.org/10.1111/tpj.12923>
- González AM, Yuste-Lisbona FJ, Fernández-Lozano A, et al (2017) Genetic mapping and QTL analysis in common bean. In: Pérez de la Vega M, Santalla M, Marsolais F (eds) *The common bean genome*. Springer International Publishing, Cham, pp 69–107. https://doi.org/10.1007/978-3-319-63526-2_4
- Guo X, Hou X, Fang J, et al (2013) The rice GERMINATION DEFECTIVE 1, encoding a B3 domain transcriptional repressor, regulates seed germination and seedling development by integrating GA and carbohydrate metabolism. *Plant J* 75:403–416. <https://doi.org/10.1111/tpj.12209>
- Gutiérrez-Marcos JF, Dal Prà M, Giulini A, et al (2007) Empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. *Plant Cell* 19:196–210. <https://doi.org/10.1105/tpc.105.039594>
- Guzman-Maldonado SH, Martinez O, Acosta-Gallegos JA, et al (2001) Putative quantitative trait loci for physical and chemical components of common bean. *Crop Sci* 43:1029–1035. <https://doi.org/10.2135/cropsci2003.1029>
- Herron SA, Rubin MJ, Ciotir C, et al (2020) Comparative analysis of early life stage traits in annual and perennial phaseolus crops and their wild relatives. *Front Plant Sci* 11:34–44. <https://doi.org/10.3389/fpls.2020.00034>
- Huang F, Chi Y, Meng Q, et al (2006) GmZFP1 encoding a single zinc finger protein is expressed with enhancement in reproductive organs and late seed development in soybean (*Glycine max*). *Mol Biol Rep* 33:279–285. <http://doi.org/10.1007/s11033-006-9012-z>
- Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol* 65:531–551. <https://doi.org/10.1146/annurev-arplant-050213-035715>
- Johannsen W (1911) The genotype conception of heredity. *Am Nat* 45:129–159
- Johannsen W (1909) *Elemente der exakten Erblchkeitslehre*. Gustav Fischer, Jena.
- Kara M, Sayinci B, Elkoca E, et al (2013) seed size and shape analysis of registered common bean (*Phaseolus vulgaris* L.) cultivars in turkey using digital photography. *J Agric Sci* 19:219–234
- Koinange EMK, Gepts P (1992) Hybrid weakness in wild *Phaseolus vulgaris* L. *J Hered* 83:135–139. <https://doi.org/10.1093/oxfordjournals.jhered.a111173>
- Koinange EMK, Singh SP, Gepts P (1996) Genetic control of the domestication syndrome in common bean. *Crop Sci* 1045:1037–1045. <https://doi.org/10.2135/cropsci1996.0011183X003600040037x>
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29. <https://doi.org/10.1186/1746-4811-9-29>
- Kumar P, Chatli MK, Mehta N, et al (2017) Meat analogues: Health promising sustainable meat substitutes. *Crit Rev Food Sci Nutr* 57:923–932. <https://doi.org/10.1080/10408398.2014.939739>
- Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor Appl Genet* 118:979–992. <https://doi.org/10.1007/s00122-008-0955-4>
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199. <https://doi.org/10.1093/genetics/121.1.185>
- Lei L, Wang L, Wang S, Wu J (2020) Marker-Trait Association Analysis of Seed Traits in Accessions of Common Bean (*Phaseolus vulgaris* L.) in China. *Front Genet* 11:698–710. <https://doi.org/10.3389/fgene.2020.00698>
- Li X-J, Zhang Y-F, Hou M, et al (2014) Small kernel 1 encodes a pentatricopeptide repeat protein required for mitochondrial nad7 transcript editing and seed development in maize (*Zea mays*) and rice (*Oryza sativa*). *Plant J* 79:797–809. <https://doi.org/10.1111/tpj.12584>
- Lipka AE, Tian F, Wang Q, et al (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Liu S, Wang X, Wang H, et al (2013a) Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of *Zea mays* L. *PLOS Genet* 9:e1003790. <https://doi.org/10.1371/journal.pgen.1003790>
- Liu Y-J, Xiu Z-H, Meeley R, Tan B-C (2013b) Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *Plant Cell* 25:868–883. <https://doi.org/10.1105/tpc.112.106781>
- Logozzo G, Donnoli R, Macaluso L, et al (2007) Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. *Genet Resour Crop Evol* 54:1763–1779. <https://doi.org/10.1007/s10722-006-9185-2>
- Lu X, Xiong Q, Cheng T, et al (2017) A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol Plant* 10:670–684. <https://doi.org/https://doi.org/10.1016/j.molp.2017.03.006>
- Lyu J, Wang D, Duan P, et al (2020) Control of Grain Size and Weight by the GSK2-LARGE1/OML4 pathway in rice. *Plant Cell* 32:1905–1918. <https://doi.org/10.1105/tpc.19.00468>

- Mamidi S, Rossi M, Moghaddam SM, et al (2013) Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity* 110:267–276. <https://doi.org/10.1038/hdy.2012.82>
- Mitchell-olds T (2010) Complex-trait analysis in plants. *Genome Biol* 11:113–115. <https://doi.org/10.1186/gb-2010-11-4-113>
- Motto M, Soressi GP, Salamini F (1978) Seed size inheritance in a cross between wild and cultivated common beans (*Phaseolus vulgaris* L.). *Genetica* 49:31–36.
- Murube E, Campa A, Song, Q, et al (2020). Toward validation of QTLs associated with pod and seed size in common bean using two nested recombinant inbred line populations. *Mol Breeding* 40:1–17. <https://doi.org/10.1007/s11032-019-1085-1>
- Nienhuis J, Singh SP (1986) Combining ability analyses and relationships among yield, yield components, and architectural traits in dry bean. *Crop breeding genetics & cytology* 1:21–27 <https://doi.org/10.2135/cropsci1986.0011183X002600010005x>
- O’Leary NA, Wright MW, Brister JR, et al (2015) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44:733–745. <https://doi.org/10.1093/nar/gkv1189>
- O’Rourke JA, Iniguez LP, Fu F, et al (2014) An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* 15:866. <https://doi.org/10.1186/1471-2164-15-866>
- OpenCV Developers Team (2012) OpenCV Reference Manual. OpenCV Ref. Man. <http://opencv.org/documentation.html>. Accessed 10 Sep 2020
- Pañeda EPA, Campa CRA, Giraldez R, José J (2010) Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 7:1367–1380. <https://doi.org/10.1007/s00122-010-1261-5>
- Park SO, Coyne DP, Jung G, et al (2000) Mapping of qtl for seed size and shape traits in common bean. *J Amer Soc Hort Sci* 125:466–475. <https://doi.org/10.21273/JASHS.125.4.466>
- Peng FY, Weselake RJ (2013) Genome-wide identification and analysis of the B3 superfamily of transcription factors in Brassicaceae and major crop plants. *Theor Appl Genet* 126:1305–1319. <https://doi.org/10.1007/s00122-013-2054-4>
- Perseguini JMCK, Silva GSB, Rosa JRBF, et al (2015) Developing a common bean core collection suitable for association mapping studies. *Genet Mol Biol* 38:67–78. <https://doi.org/10.1590/S1415-475738120140126>
- Petrusán JI, Rawel H, Huscsek G (2016) Protein-rich vegetal sources and trends in human nutrition: A review. *Curr Top Pept Protein Res* 17:1–19
- Petry N, Boy E, Wirth JP, Hurrell RF (2015) Review: The Potential of the Common Bean (*Phaseolus vulgaris*) as a Vehicle for Iron Biofortification. *Nutrients* 7:1144–1173. <https://doi.org/10.3390/nu7021144>
- Qi W, Yang Y, Feng X, et al (2017) Mitochondrial function and maize kernel development requires Dek2, a pentatricopeptide repeat protein involved in nad1 mRNA splicing. *Genetics* 205:239–249. <https://doi.org/10.1534/genetics.116.196105>
- Rana JC, Sharma TR, Tyagi RK, et al (2015) Characterisation of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. *Euphytica* 205:441–457. <https://doi.org/10.1007/s10681-015-1406-3>
- Roll-Hansen N (1989) The crucial experiment of Wilhelm Johannsen. *Biol Philos* 4:303–329. <https://doi.org/10.1007/BF02426630>
- Sakai H, Krizek BA, Jacobsen SE, Meyerowitz EM (2000) Regulation of SUP expression identifies multiple regulators involved in Arabidopsis floral meristem development. *Plant Cell* 12:1607–1618. <https://doi.org/10.1105/tpc.12.9.1607>
- Sakai H, Medrano LJ, Meyerowitz EM (1995) Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries. *Nature* 378:199–203. <https://doi.org/10.1038/378199a0>
- Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8:552
- Schmutz J, McClean PE, Mamidi S, et al (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet* 46:707–713. <https://doi.org/10.1038/ng.3008>
- Segura V, Vilhjálmsson BJ, Platt A, et al (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* 44:825–830. <https://doi.org/10.1038/ng.2314>
- Sharma P, Michael K (2020) Improving seed size, seed weight and seedling emergence in *Camelina sativa* by overexpressing the Atsob3 - 6 gene variant. *Transgenic Res* 29:409–418. <https://doi.org/10.1007/s11248-020-00208-9>
- Shaw N, Barak RS, Campbell RE, et al (2020) Seed use in the field: Delivering seeds for restoration success. *Restor Ecol* 28:276–285. <https://doi.org/10.1111/rec.13210>
- Shin JH, Blay S, McNeney B (2006) LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw* 16:1–10. <https://doi.org/10.18637/jss.v016.c03>
- Singh DK, Singh DP, Singh SS (2018) Studies of Genetic variability, heritability and genetic advance for yield and related traits in French bean (*Phaseolus vulgaris* L.). *J Pharmac Phytoch* 7:236–240

- Singh SP, Gepts P, Debouck DG (1991a) Races of common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot 45:379–396. <https://doi.org/10.1007/BF02887079>
- Singh SP, Gutierrez JA, Molina A, et al (1991b) Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci 31:23–29. <https://doi.org/10.2135/cropsci1991.0011183X003100010005x>
- Swaminathan K, Peterson K, Jack T (2008) The plant B3 superfamily. Trends Plant Sci 13:647–655. <https://doi.org/https://doi.org/10.1016/j.tplants.2008.09.006>
- Tanabata T, Shibaya T, Hori K, et al (2012) SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiol 160:1871–1880. <https://doi.org/10.1104/pp.112.205120>
- Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. <https://doi.org/10.1016/j.tibtech.2009.05.006>
- Verma S, Gupta S, Bandhiwal N, et al (2015) High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). Sci Rep 5:17512. <https://doi.org/10.1038/srep17512>
- Wang H, Dittmer TA, Richards EJ (2013) Arabidopsis CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. BMC Plant Biol 13:1–13. <https://doi.org/10.1186/1471-2229-13-200>
- White JW, González A (1990) Characterization of the negative association between seed yield and seed size among genotypes of common bean. Field Crops Research 23:159–175. [https://doi.org/10.1016/0378-4290\(90\)90052-D](https://doi.org/10.1016/0378-4290(90)90052-D)
- Wright EM, Kelly JD (2011) Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). Euphytica 179:471–484. <https://doi.org/10.1007/s10681-011-0369-2>
- Wu J, Wang L, Fu J, et al (2019). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. Nat Genet 52:118–125. <https://doi.org/10.1038/s41588-019-0546-0>
- Xu Y, Li P, Yang Z, Xu C (2017) Genetic mapping of quantitative trait loci in crops. Crop J 5:175–184. <https://doi.org/https://doi.org/10.1016/j.cj.2016.06.003>
- Yang T, Guo L, Ji C, et al (2020) The B3 domain-containing transcription factor ZmABI19 coordinates expression of key factors required for maize seed development and grain filling. Plant Cell 8:e1 <https://doi.org/10.1093/plcell/koaa008>
- Yuste-Lisbona FJ, González AM, Capel C, et al (2014) Genetic analysis of single-locus and epistatic QTLs for seed traits in an adapted × nuña RIL population of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 127:897–912. <https://doi.org/10.1007/s00122-014-2265-3>
- Zhang T (2020) When less is more: GSK2-OML4 module negatively regulates grain size in rice. Plant Cell 32:1–2. <https://doi.org/10.1105/tpc.20.00219>
- Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360. <https://doi.org/10.1038/ng.546>
- Zhao K, Aranzana MJ, Kim S, et al (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4. <https://doi.org/10.1371/journal.pgen.0030004>
- Zhao W, Guan C, Feng J, et al (2016) The Arabidopsis CROWDED NUCLEI genes regulate seed germination by modulating degradation of ABI5 protein. J Integr Plant Biol 58:669–678. <http://doi.org/10.1111/jipb.12448>
- Zhou Q, Dong Y, Shi Q, et al (2017) Verification and fine mapping of qGW1.05, a major QTL for grain weight in maize (*Zea mays* L.). Mol Genet Genomics 292:871–881. <http://doi.org/10.1007/s00438-017-1318-0>
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. Plant Genome 1:5–20. <https://doi.org/10.3835/plantgenome2008.02.0089>

Declarations**Funding**

This research was funded by the following Brazilian Institutions: Fundação de Amparo à Pesquisa do Estado de São Paulo, grant numbers 2014/06647-9, 2018/09069-7 and 2020/02755-2; Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant numbers 454329/2014-8 and 140588/2018-3; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Finance code 001.

Availability of data and material

GBS data can be accessed at the GenBank database, BioSample SAMN05513252 and SAMN05513251, included in BioProject PRJNA336556. Phenotypic data are available through direct contact with the corresponding author.

Acknowledgments

The authors wish to thank the following Brazilian Institutions: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grant number 2014/06647-9, 2018/09069-7 and 2020/02755-2); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grant number 454329/2014-8); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Finance Code 001) for financial support. We also thank Mr. Steve Simmons for proofreading the manuscript.

4. CONCLUDING REMARKS

When common bean is compared to soybean, its relative crop, it is evident the lack of genetic information to boost the crop improvement. It includes the discovery of molecular markers linked to traits of interest and the generation of a clearer picture about the mechanisms underlying quantitative traits. The outcome of the research herein satisfies both of these needs and provides a more comprehensive understanding of common bean root-knot nematode resistance and seed morphology.

In the first study, we assessed the response of both a common bean diversity panel and a mapping population to root-knot nematode infection by measuring the severity of root symptoms and the nematode's ability to reproduce in the host. We identified eight genomic regions associated to the common bean response to RKN, suggesting the existence of several genes controlling the quantity of nematode egg masses and root-galling index, and indicating the complex nature of the genetic architecture associated with the resistance detected. Interestingly, even though significant positive correlation was verified between these traits, there was no coincident associations, suggesting that egg masses production and root-galling index are under independent genetic control in common bean. In the associated regions, bioinformatics analysis detected great number of genes, among which, several candidates were indicated to putatively control the resistance, including 12 RGAs and five genes differentially expressed in *P. vulgaris* – *M. incognita* interaction. This study also presented preliminary evidence that H₂O₂ accumulation is involved in the common bean response to RKN, being further evident in the more resistant genotypes.

In the subsequent chapter, the IAC core collection panel was evaluated and exhibited high variability for seven seed traits related to size (length, width, perimeter and projected area), shape (length-to-width ratio and circularity) and weight (hundred seed mass). Andean gene pool seeds were observed to be larger and heavier, and less circular than those of the Mesoamerican pool, corroborating previous diversity studies. GWAS identified seven marker-trait associations, explaining a considerable amount of the phenotypic variation for seed length, width, projected area and circularity in four distinct genomic regions of chromosomes Pv02, Pv08 and Pv11. In the associated regions, 188 genes were annotated, 156 of which were expressed during seed development, according to the common bean expression atlas, with 13 promising candidates involved in seed morphology regulation in common bean.

In conclusion, the two studies documented herein explored high-density genotyping and novel phenotyping procedures to accomplish GWAS and dissect the genetic architecture of common bean quantitative traits, particularly response to RKN and seed morphology. From a practical standpoint, these works provide breeders with SNP markers that present great potential to be used in marker-assisted breeding. Furthermore, several genes were indicated as promising candidates to underlie the phenotypic variation of those traits and consequently were considered remarkable options for validation by functional analysis in common bean and eventually other related crops.

APPENDICES

APPENDIX A. Caterpillar plot of the adjusted means with 95% confidence intervals for number of egg masses (a) and root-galling index (b) in the IAC common bean core collection including Andean (red) and Mesoamerican (blue) gene pools

APPENDIX B. Linkage map file of a BC₂F₃ population generated by crossing the donor accession AND-277 and the recurrent IAC-Milênio

APPENDIX C. Number of egg masses (a) and root-galling index (b) in common bean genotypes from the Andean (red) and Mesoamerican (blue) gene pools

APPENDIX D. Histograms (diagonal), scatter charts (lower diagonal) and genetic correlations (upper diagonal) for egg masses, root-galling and root morphology traits

APPENDIX E. Logarithm of odds (LOD) scores profile for egg masses (a) and root-galling index (b) according to composite interval mapping using the EM algorithm (red) and Haley-Knot regression (blue) in a BC₂F₄ population. The horizontal dashed lines indicate the permutation threshold

APPENDIX F. Linkage disequilibrium (LD) analysis to define the genomic regions for candidate gene discovery. Significant SNPs and flanking markers in linkage disequilibrium associated with egg masses (a) and root-galling index (b). Heatmap indicates the degree of LD estimates

APPENDIX G. Functional annotation and ontology terms predicted for the candidate genes located in genomic regions significantly associated with egg masses and root-galling index

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Egg masses			
PhvuL006G051000	beta-(1,2)-xylosyltransferase	C:GO:0000139; C:GO:0005797; P:GO:0006487; C:GO:0016021; P:GO:0031204; P:GO:0048367; F:GO:0050513	C:Golgi membrane; C:Golgi medial cisterna; P:protein N-linked glycosylation; C:integral component of membrane; P:posttranslational protein targeting to membrane, translocation; P:shoot system development; F:glycoprotein 2-beta-D-xylosyltransferase activity
PhvuL006G051100	uncharacterized protein LOC115708834	C:GO:0016021	C:integral component of membrane
PhvuL006G051200	ankyrin repeat-containing protein At5g02620-like isoform X2	C:GO:0016021	C:integral component of membrane
PhvuL006G051300	40S ribosomal protein S8-like	P:GO:0000462; F:GO:0003735; P:GO:0006412; C:GO:0022627	P:maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); F:structural constituent of ribosome; P:translation; C:cytosolic small ribosomal subunit
PhvuL006G051400	probable cytokinin riboside 5'-monophosphate phosphoribohydrolase LOGL3	F:GO:0016787	F:hydrolase activity
PhvuL006G051500	disease resistance protein RPM1	F:GO:0043531	F:ADP binding
PhvuL006G051600	F-box protein At1g30790		
PhvuL006G051700	protein PRD1	P:GO:0042138	P:meiotic DNA double-strand break formation
PhvuL006G051800	F-box protein At1g30790		
PhvuL006G051900	putative F-box protein At1g47790		
PhvuL006G052000	protein transport protein SEC23-like	C:GO:0000139; F:GO:0005096; C:GO:0005789; P:GO:0006886; F:GO:0008270; C:GO:0030127; P:GO:0043547; C:GO:0070971; P:GO:0090110	C:Golgi membrane; F:GTPase activator activity; C:endoplasmic reticulum membrane; P:intracellular protein transport; F:zinc ion binding; C:COPII vesicle coat; P:positive regulation of GTPase activity; C:endoplasmic reticulum exit site; P:COPII-coated vesicle cargo loading
PhvuL006G052100	putative F-box protein At1g47790		
PhvuL006G052200	ABC transporter C family member 3 isoform X1	C:GO:0000325; F:GO:0005524; C:GO:0005774; C:GO:0016021; F:GO:0016887; F:GO:0042626; P:GO:0055085	C:plant-type vacuole; F:ATP binding; C:vacuolar membrane; C:integral component of membrane; F:ATPase activity; F:ATPase-coupled transmembrane transporter activity; P:transmembrane transport
PhvuL006G052300	disease resistance protein RPM1-like	F:GO:0043531	F:ADP binding
PhvuL006G052400	disease resistance protein RPM1-like	F:GO:0043531	F:ADP binding
PhvuL006G052500	disease resistance protein RPM1	F:GO:0043531	F:ADP binding
PhvuL006G052600	disease resistance protein RPM1-like isoform X2	F:GO:0043531	F:ADP binding
PhvuL006G052700	kinesin-like protein KIN-12F	F:GO:0005524; C:GO:0005871; C:GO:0005874; P:GO:0007018; F:GO:0008017; F:GO:0008574	F:ATP binding; C:kinesin complex; C:microtubule; P:microtubule-based movement; F:microtubule binding; F:ATP-dependent microtubule motor activity, plus-end-directed
PhvuL006G052800	hypothetical protein PHAVU_006G052800g		
PhvuL006G052900	protein TORNADO 1		
PhvuL006G053000	heterogeneous nuclear ribonucleoprotein 1-like	F:GO:0003723; C:GO:0005634; C:GO:1990904 F:GO:0004674; F:GO:0005524; C:GO:0005737; P:GO:0006468; P:GO:0035556	F:RNA binding; C:nucleus; C:ribonucleoprotein complex
PhvuL006G053100	probable serine/threonine-protein kinase WNK11		F:protein serine/threonine kinase activity; F:ATP binding; C:cytoplasm; P:protein phosphorylation; P:intracellular signal transduction

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.006G053200	FRIGIDA-like protein 1	P:GO:0009908; P:GO:0030154	P:flower development; P:cell differentiation
Phvul.006G053300	probable WRKY transcription factor 20 isoform X1	F:GO:0003700; C:GO:0005634; P:GO:0006355; F:GO:0043565	F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated; F:sequence-specific DNA binding
Phvul.007G246400	methyltransferase-like protein 13	F:GO:0008171; P:GO:0032259	F:O-methyltransferase activity; P:methylation
Phvul.007G246300	BLOC-1-related complex subunit 8 homolog isoform X3		
Phvul.007G246500	protein FLOWERING LOCUS D	F:GO:0003677; F:GO:0008168; P:GO:0032259; F:GO:0052894; F:GO:0052895; F:GO:0052901; P:GO:0055114	F:DNA binding; F:methyltransferase activity; P:methylation; F:norspermine:oxygen oxidoreductase activity; F:N1-acetylspermine:oxygen oxidoreductase (N1-acetylspermidine-forming) activity; F:spermine:oxygen oxidoreductase (spermidine-forming) activity; P:oxidation-reduction process
Phvul.007G246600	receptor-like protein EIX2	C:GO:0016021; F:GO:0016301; P:GO:0016310 F:GO:0030599; P:GO:0042545; P:GO:0043086; F:GO:0045330; P:GO:0045490; F:GO:0046910	C:integral component of membrane; F:kinase activity; P:phosphorylation
Phvul.007G246700	probable pectinesterase/pectinesterase inhibitor 41	F:GO:0030599; P:GO:0043086; F:GO:0046910	F:pectinesterase activity; P:cell wall modification; P:negative regulation of catalytic activity; F:aspartyl esterase activity; P:pectin catabolic process; F:pectinesterase inhibitor activity
Phvul.007G246800	pectinesterase inhibitor-like	F:GO:0030599; P:GO:0043086; F:GO:0046910	F:pectinesterase activity; P:negative regulation of catalytic activity; F:pectinesterase inhibitor activity
Phvul.007G246900	hypothetical protein PHAVU_007G246900g		
Phvul.007G247000	zinc finger protein ZOP1	P:GO:0000398; F:GO:0003690; F:GO:0003725; F:GO:0008270; P:GO:0009845; C:GO:0015030; P:GO:0080188	P:mRNA splicing, via spliceosome; F:double-stranded DNA binding; F:double-stranded RNA binding; F:zinc ion binding; P:seed germination; C:Cajal body; P:RNA-directed DNA methylation
Phvul.007G247100	hypothetical protein PHAVU_007G247100g		
Phvul.007G247200	pleiotropic drug resistance protein 2-like	F:GO:0005524; C:GO:0005886; C:GO:0016021; F:GO:0016887; F:GO:0042626; P:GO:0055085 F:GO:0004402; P:GO:0006357; P:GO:0016573; C:GO:0032777 C:GO:0000145; P:GO:0006893; P:GO:0006904	F:ATP binding; C:plasma membrane; C:integral component of membrane; F:ATPase activity; F:ATPase-coupled transmembrane transporter activity; P:transmembrane transport
Phvul.007G247300	hypothetical protein PHAVU_007G247300g		F:histone acetyltransferase activity; P:regulation of transcription by RNA polymerase II; P:histone acetylation; C:Piccolo NuA4 histone acetyltransferase complex
Phvul.007G247400	exocyst complex component SEC15A		C:exocyst; P:Golgi to plasma membrane transport; P:vesicle docking involved in exocytosis
Phvul.008G247700	extra-large guanine nucleotide-binding protein 3-like		
Phvul.008G247800	transcription factor RF2b-like	F:GO:0003700; C:GO:0005634; P:GO:0006355	F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated
Phvul.008G247900	hypothetical protein PHAVU_008G247900g		
Phvul.008G248000	expansin-A8	C:GO:0005576; C:GO:0005618; P:GO:0009664; C:GO:0016020	C:extracellular region; C:cell wall; P:plant-type cell wall organization; C:membrane
Phvul.008G248100	hypothetical protein PHAVU_008G248100g		
Phvul.008G248200	probable serine/threonine-protein kinase WNK4	F:GO:0004674; F:GO:0005524; C:GO:0005737; P:GO:0006468; P:GO:0035556	F:protein serine/threonine kinase activity; F:ATP binding; C:cytoplasm; P:protein phosphorylation; P:intracellular signal transduction
Phvul.008G248300	transcription repressor OFP4-like	F:GO:0003677; P:GO:0045892	F:DNA binding; P:negative regulation of transcription, DNA-templated

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
PhvuL008G248400	transcription repressor OFP17		
PhvuL008G248500	protein GRAVITROPIC IN THE LIGHT 1 isoform X2	P:GO:0009639; P:GO:0009959	P:response to red or far red light; P:negative gravitropism
PhvuL008G248600	premnaspirodien oxygenase-like	F:GO:0005506; C:GO:0016021; F:GO:0016709; F:GO:0020037; P:GO:0055114	F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen; F:heme binding; P:oxidation-reduction process
PhvuL008G248700	EPIDERMAL PATTERNING FACTOR-like protein 6	P:GO:0010052; C:GO:0016021 F:GO:0004674;	P:guard cell differentiation; C:integral component of membrane
PhvuL008G248800	CBL-interacting serine/threonine-protein kinase 11	F:GO:0005524; C:GO:0005634; C:GO:0005737; P:GO:0006468; P:GO:0035556	F:protein serine/threonine kinase activity; F:ATP binding; C:nucleus; C:cytoplasm; P:protein phosphorylation; P:intracellular signal transduction
PhvuL008G248900	CBL-interacting serine/threonine-protein kinase 10-like	F:GO:0004674; F:GO:0005524; P:GO:0006468; P:GO:0007165; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; P:signal transduction; C:integral component of membrane
PhvuL008G249000	serine/arginine repetitive matrix protein 2-like		
PhvuL008G249100	LOB domain-containing protein 13 isoform X2		
PhvuL008G249200	putative pentatricopeptide repeat-containing protein At1g12700, mitochondrial		
PhvuL008G249300	tyrosine--tRNA ligase, chloroplastic/mitochondrial	F:GO:0003723; F:GO:0004831; F:GO:0005524; C:GO:0005739; C:GO:0005829; P:GO:0006437; C:GO:0009570	F:RNA binding; F:tyrosine-tRNA ligase activity; F:ATP binding; C:mitochondrion; C:cytosol; P:tyrosyl-tRNA aminoacylation; C:chloroplast stroma
PhvuL008G249400	U-box domain-containing protein 4-like	F:GO:0004842; C:GO:0005634; C:GO:0005737; P:GO:0016567 F:GO:0004601; C:GO:0005576;	F:ubiquitin-protein transferase activity; C:nucleus; C:cytoplasm; P:protein ubiquitination
PhvuL008G249500	peroxidase 4	P:GO:0006979; F:GO:0020037; P:GO:0042744; F:GO:0046872; P:GO:0055114; P:GO:0098869	F:peroxidase activity; C:extracellular region; P:response to oxidative stress; F:heme binding; P:hydrogen peroxide catabolic process; F:metal ion binding; P:oxidation-reduction process; P:cellular oxidant detoxification
PhvuL008G249600	biogenesis of lysosome-related organelles complex 1 subunit 1	P:GO:0016197; C:GO:0031083 F:GO:0004601; C:GO:0005576;	P:endosomal transport; C:BLOC-1 complex
PhvuL008G249700	peroxidase P7-like	P:GO:0006979; C:GO:0016021; F:GO:0020037; P:GO:0042744; F:GO:0046872; P:GO:0055114; P:GO:0098869 F:GO:0004601; C:GO:0005576;	F:peroxidase activity; C:extracellular region; P:response to oxidative stress; C:integral component of membrane; F:heme binding; P:hydrogen peroxide catabolic process; F:metal ion binding; P:oxidation-reduction process; P:cellular oxidant detoxification
PhvuL008G249800	cationic peroxidase 1-like	P:GO:0006979; C:GO:0016021; F:GO:0020037; P:GO:0042744; F:GO:0046872; P:GO:0055114; P:GO:0098869 F:GO:0004601; C:GO:0005576;	F:peroxidase activity; C:extracellular region; P:response to oxidative stress; C:integral component of membrane; F:heme binding; P:hydrogen peroxide catabolic process; F:metal ion binding; P:oxidation-reduction process; P:cellular oxidant detoxification
PhvuL008G249900	cationic peroxidase 1-like	P:GO:0006979; F:GO:0020037; P:GO:0042744; F:GO:0046872;	F:peroxidase activity; C:extracellular region; P:response to oxidative stress; F:heme binding; P:hydrogen peroxide catabolic process; F:metal ion binding; P:oxidation-reduction process; P:cellular oxidant detoxification

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
		P:GO:0055114; P:GO:0098869	
Phvul.008G250000	protein TIC110, chloroplastic	P:GO:0009658; C:GO:0009941; C:GO:0016021; P:GO:0045037; C:GO:0061927	P:chloroplast organization; C:chloroplast envelope; C:integral component of membrane; P:protein import into chloroplast stroma; C:TOC-TIC supercomplex I
Phvul.008G250100	putative tRNA pseudouridine synthase isoform X1	F:GO:0003723; P:GO:0031119; F:GO:0106029	F:RNA binding; P:tRNA pseudouridine synthesis; F:tRNA pseudouridine synthase activity
Phvul.008G250200	putative tRNA pseudouridine synthase	F:GO:0003723; F:GO:0009982; P:GO:0031119	F:RNA binding; F:pseudouridine synthase activity; P:tRNA pseudouridine synthesis
Phvul.008G250300	hypothetical protein PHAVU_008G250300g		
Phvul.008G250400	hypothetical protein PHAVU_008G250400g		
Phvul.008G250500	protein NUCLEAR FUSION DEFECTIVE 4	C:GO:0016021	C:integral component of membrane
Phvul.008G250600	probable O-methyltransferase 3	F:GO:0008171; F:GO:0008757; P:GO:0019438; P:GO:0032259; F:GO:0046983	F:O-methyltransferase activity; F:S-adenosylmethionine-dependent methyltransferase activity; P:aromatic compound biosynthetic process; P:methylation; F:protein dimerization activity
Phvul.008G250700	probable O-methyltransferase 3	F:GO:0016206; P:GO:0019438; P:GO:0032259; F:GO:0046983; F:GO:0102084; F:GO:0102938	F:catechol O-methyltransferase activity; P:aromatic compound biosynthetic process; P:methylation; F:protein dimerization activity; F:L-dopa O-methyltransferase activity; F:orcinol O-methyltransferase activity
Phvul.008G250800	protein LURP-one-related 7 isoform X1	C:GO:0005886; C:GO:0016021 C:GO:0005829;	C:plasma membrane; C:integral component of membrane
Phvul.008G250900	E3 ubiquitin-protein ligase listerin	F:GO:0008270; P:GO:0016567; F:GO:0016874; F:GO:0043023; F:GO:0061630; P:GO:0072344; C:GO:1990112; P:GO:1990116	C:cytosol; F:zinc ion binding; P:protein ubiquitination; F:ligase activity; F:ribosomal large subunit binding; F:ubiquitin protein ligase activity; P:rescue of stalled ribosome; C:RQC complex; P:ribosome-associated ubiquitin-dependent protein catabolic process
Phvul.008G251000	wound-induced protein 1-like		
Phvul.011G181400	TMV resistance protein N isoform X1	P:GO:0007165; F:GO:0043531	P:signal transduction; F:ADP binding
Phvul.011G181500	putative disease resistance protein At3g14460	F:GO:0043531	F:ADP binding
Phvul.011G181600	putative disease resistance RPP13-like protein 1	F:GO:0043531	F:ADP binding
Phvul.011G181700	putative disease resistance protein At3g14460	F:GO:0043531	F:ADP binding
Phvul.011G181800	potassium channel KAT1	F:GO:0005249; C:GO:0016021; P:GO:0071805	F:voltage-gated potassium channel activity; C:integral component of membrane; P:potassium ion transmembrane transport
Phvul.011G181900	putative kinase-like protein TMKL1	F:GO:0004672; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.011G182000	diphthine methyltransferase homolog isoform X1	F:GO:0008168; P:GO:0032259	F:methyltransferase activity; P:methylation
Phvul.011G182100	60S ribosomal protein L11	P:GO:0000027; F:GO:0003735; P:GO:0006412; C:GO:0022625	P:ribosomal large subunit assembly; F:structural constituent of ribosome; P:translation; C:cytosolic large ribosomal subunit
Phvul.011G182200	embryonic abundant protein VF30.1-like		
Phvul.011G182300	probable leucine-rich repeat receptor-like protein kinase At1g35710	F:GO:0004672; F:GO:0005524; C:GO:0005886; P:GO:0006468; C:GO:0016021	F:protein kinase activity; F:ATP binding; C:plasma membrane; P:protein phosphorylation; C:integral component of membrane
Phvul.011G182400	acid phosphatase 1	F:GO:0003993; P:GO:0016311	F:acid phosphatase activity; P:dephosphorylation
Phvul.011G182500	MLP-like protein 43	P:GO:0006952; C:GO:0016021	P:defense response; C:integral component of membrane

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
PhvuL011G182600	MLP-like protein 43	P:GO:0006952; C:GO:0016021	P:defense response; C:integral component of membrane
PhvuL011G182700	MLP-like protein 43	P:GO:0006952	P:defense response
PhvuL011G182800	MLP-like protein 43	P:GO:0006952	P:defense response
PhvuL011G182900	MLP-like protein 43	P:GO:0006952	P:defense response
PhvuL011G183000	MLP-like protein 43	P:GO:0006952	P:defense response
PhvuL011G183100	MLP-like protein 43	P:GO:0006952	P:defense response
PhvuL011G183200	N-terminal acetyltransferase A complex catalytic subunit NAA10	P:GO:0017198; P:GO:0018002; C:GO:0031415; F:GO:1990189; F:GO:1990190	P:N-terminal peptidyl-serine acetylation; P:N-terminal peptidyl-glutamic acid acetylation; C:NatA complex; F:peptide-serine-N-acetyltransferase activity; F:peptide-glutamate-N-acetyltransferase activity
PhvuL011G183300	MLP-like protein 43	P:GO:0006952	P:defense response
PhvuL011G183400	MLP-like protein 43	P:GO:0006952; C:GO:0016021	P:defense response; C:integral component of membrane
PhvuL011G183500	MLP-like protein 43	P:GO:0006952; C:GO:0016021	P:defense response; C:integral component of membrane
Root-galling Index			
PhvuL001G199900	rhomboid-like protein 15	C:GO:0000839; F:GO:0004252; C:GO:0030176; P:GO:0030433; P:GO:0030968; F:GO:0051787; F:GO:1990381	C:Hrd1p ubiquitin ligase ERAD-L complex; F:serine-type endopeptidase activity; C:integral component of endoplasmic reticulum membrane; P:ubiquitin-dependent ERAD pathway; P:endoplasmic reticulum unfolded protein response; F:misfolded protein binding; F:ubiquitin-specific protease binding
PhvuL001G200000	eukaryotic translation initiation factor 3 subunit C	P:GO:0001732; F:GO:0003743; C:GO:0005852; C:GO:0016282; F:GO:0031369; C:GO:0033290	P:formation of cytoplasmic translation initiation complex; F:translation initiation factor activity; C:eukaryotic translation initiation factor 3 complex; C:eukaryotic 43S preinitiation complex; F:translation initiation factor binding; C:eukaryotic 48S preinitiation complex
PhvuL001G200100	TNF receptor-associated factor family protein DDB_G0290931	F:GO:0008270	F:zinc ion binding
PhvuL001G200200	splicing factor 3B subunit 3 isoform X1	F:GO:0003684; C:GO:0005634; P:GO:0043161; C:GO:0000159; F:GO:0004721; F:GO:0005515; C:GO:0005730; C:GO:0005829; P:GO:0006470; P:GO:0006952; F:GO:0046872	F:damaged DNA binding; C:nucleus; P:proteasome-mediated ubiquitin-dependent protein catabolic process
PhvuL001G200300	serine/threonine-protein phosphatase PP2A catalytic subunit	F:GO:0008195; C:GO:0016021; P:GO:0016311	C:protein phosphatase type 2A complex; F:phosphoprotein phosphatase activity; F:protein binding; C:nucleolus; C:cytosol; P:protein dephosphorylation; P:defense response; F:metal ion binding
PhvuL001G200400	lipid phosphate phosphatase delta		F:phosphatidate phosphatase activity; C:integral component of membrane; P:dephosphorylation
PhvuL001G200500	hypothetical protein PHAVU_001G200500g		
PhvuL001G200600	protein WHAT'S THIS FACTOR 9, mitochondrial		
PhvuL001G200700	carbon catabolite repressor protein 4 homolog 1	P:GO:0000289; F:GO:0004535; C:GO:0005634; C:GO:0005737; C:GO:0030014; P:GO:0090503	P:nuclear-transcribed mRNA poly(A) tail shortening; F:poly(A)-specific ribonuclease activity; C:nucleus; C:cytoplasm; C:CCR4-NOT complex; P:RNA phosphodiester bond hydrolysis, exonucleolytic
PhvuL001G200800	hypothetical protein PHAVU_001G200800g		
PhvuL001G200900	60S ribosomal protein L21-1	F:GO:0003735; P:GO:0006412; C:GO:0022625	F:structural constituent of ribosome; P:translation; C:cytosolic large ribosomal subunit
PhvuL001G201000	60S ribosomal protein L21-1	F:GO:0003735; P:GO:0006412; C:GO:0022625	F:structural constituent of ribosome; P:translation; C:cytosolic large ribosomal subunit
PhvuL001G201100	NPL4-like protein 1	C:GO:0005634; P:GO:0006511;	C:nucleus; P:ubiquitin-dependent protein catabolic process; F:ubiquitin protein ligase binding; F:ubiquitin binding

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.001G201200	sister-chromatid cohesion protein 3	F:GO:0031625; F:GO:0043130 C:GO:0000785; F:GO:0003677; F:GO:0003682; C:GO:0005634; P:GO:0007062; C:GO:0008278 F:GO:0004601; C:GO:0005576;	C:chromatin; F:DNA binding; F:chromatin binding; C:nucleus; P:sister chromatid cohesion; C:cohesin complex
Phvul.001G201300	peroxidase 3-like	P:GO:0006979; C:GO:0009505; C:GO:0009506; F:GO:0020037; P:GO:0042744; F:GO:0046872; P:GO:0055114; P:GO:0098869	F:peroxidase activity; C:extracellular region; P:response to oxidative stress; C:plant-type cell wall; C:plasmodesma; F:heme binding; P:hydrogen peroxide catabolic process; F:metal ion binding; P:oxidation-reduction process; P:cellular oxidant detoxification
Phvul.001G201400	rho GTPase-activating protein 2	P:GO:0007165	P:signal transduction
Phvul.001G201500	phosphopantothienoylcysteine decarboxylase subunit VHS3		
Phvul.001G201600	RNA-binding protein BRN1	F:GO:0003729; C:GO:0005634; C:GO:0005737; C:GO:1990904	F:mRNA binding; C:nucleus; C:cytoplasm; C:ribonucleoprotein complex
Phvul.001G201700	hypothetical protein PHAVU_001G201700g		
Phvul.001G201800	zerumbone synthase	P:GO:0055114; F:GO:0102069	P:oxidation-reduction process; F:zerumbone synthase activity
Phvul.001G201900	hypothetical protein PHAVU_001G201900g		
Phvul.001G202000	auxin response factor 23-like	F:GO:0003677; C:GO:0005634; P:GO:0006355; P:GO:0009734 F:GO:0004499;	F:DNA binding; C:nucleus; P:regulation of transcription, DNA-templated; P:auxin-activated signaling pathway
Phvul.001G202100	probable indole-3-pyruvate monooxygenase YUCCA10	F:GO:0050660; F:GO:0050661; P:GO:0055114; F:GO:0103075 F:GO:0004499;	F:N,N-dimethylaniline monooxygenase activity; F:flavin adenine dinucleotide binding; F:NADP binding; P:oxidation-reduction process; F:indole-3-pyruvate monooxygenase activity
Phvul.001G202200	probable indole-3-pyruvate monooxygenase YUCCA10	F:GO:0005524; F:GO:0050660; F:GO:0050661; P:GO:0055114; F:GO:0103075	F:N,N-dimethylaniline monooxygenase activity; F:ATP binding; F:flavin adenine dinucleotide binding; F:NADP binding; P:oxidation-reduction process; F:indole-3-pyruvate monooxygenase activity
Phvul.001G202300	classical arabinogalactan protein 26	C:GO:0016021	C:integral component of membrane
Phvul.001G202400	protein NETWORKED 3A isoform X2	C:GO:0005774; C:GO:0005886; F:GO:0016301; P:GO:0016310; F:GO:0051015	C:vacuolar membrane; C:plasma membrane; F:kinase activity; P:phosphorylation; F:actin filament binding
Phvul.001G202500	---NA---		
Phvul.001G202600	protein TRANSPORT INHIBITOR RESPONSE 1-like	F:GO:0000822; P:GO:0009734; F:GO:0010011; P:GO:0010152; P:GO:0010311; P:GO:0016036; C:GO:0019005; F:GO:0038198; P:GO:0048443	F:inositol hexakisphosphate binding; P:auxin-activated signaling pathway; F:auxin binding; P:pollen maturation; P:lateral root formation; P:cellular response to phosphate starvation; C:SCF ubiquitin ligase complex; F:auxin receptor activity; P:stamen development
Phvul.001G202700	hypothetical protein PHAVU_001G202700g	C:GO:0016021	C:integral component of membrane
Phvul.001G202800	stress response protein nst1	F:GO:0016874 C:GO:0005886; P:GO:0007043;	F:ligase activity
Phvul.001G202900	CASP-like protein 1E1	C:GO:0016021; P:GO:0042545; C:GO:0048226; F:GO:0051539	C:plasma membrane; P:cell-cell junction assembly; C:integral component of membrane; P:cell wall modification; C:Casparian strip; F:4 iron, 4 sulfur cluster binding
Phvul.001G203000	zinc finger protein CONSTANS-LIKE 13	C:GO:0005634; F:GO:0008270	C:nucleus; F:zinc ion binding

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
PhvuL001G203100	membrane-anchored ubiquitin-fold protein 3	C:GO:0005886	C:plasma membrane
PhvuL001G203200	monothiol glutaredoxin-S11	F:GO:0004791; C:GO:0005623; C:GO:0005634; C:GO:0005737; F:GO:0009055; F:GO:0015035; P:GO:0022900; P:GO:0045454; P:GO:0098869 C:GO:0005623; F:GO:0009055; F:GO:0015035; P:GO:0022900; P:GO:0045454	F:thioredoxin-disulfide reductase activity; C:cell; C:nucleus; C:cytoplasm; F:electron transfer activity; F:protein disulfide oxidoreductase activity; P:electron transport chain; P:cell redox homeostasis; P:cellular oxidant detoxification
PhvuL001G203300	glutaredoxin-C11	F:GO:0009055; F:GO:0015035; P:GO:0022900; P:GO:0045454	C:cell; F:electron transfer activity; F:protein disulfide oxidoreductase activity; P:electron transport chain; P:cell redox homeostasis
PhvuL001G203400	OTU domain-containing protein 6B	F:GO:0004843; P:GO:0006508	F:thiol-dependent ubiquitin-specific protease activity; P:proteolysis
PhvuL001G203500	monothiol glutaredoxin-S6	F:GO:0004791; C:GO:0005623; F:GO:0009055; F:GO:0015035; P:GO:0022900; P:GO:0045454; P:GO:0098869 C:GO:0005623; F:GO:0009055; F:GO:0015035; P:GO:0022900; P:GO:0045454	F:thioredoxin-disulfide reductase activity; C:cell; F:electron transfer activity; F:protein disulfide oxidoreductase activity; P:electron transport chain; P:cell redox homeostasis; P:cellular oxidant detoxification
PhvuL001G203600	monothiol glutaredoxin-S1-like	F:GO:0009055; F:GO:0015035; P:GO:0022900; P:GO:0045454	C:cell; F:electron transfer activity; F:protein disulfide oxidoreductase activity; P:electron transport chain; P:cell redox homeostasis
PhvuL001G203700	hypothetical protein PHAVU_001G203700g		
PhvuL001G203800	BTB/POZ domain-containing protein At1g03010 isoform X1	P:GO:0006415; C:GO:0009507; P:GO:0009658; P:GO:0010027; F:GO:0016149; P:GO:0032544; F:GO:0043022 F:GO:0003824; C:GO:0009507 F:GO:0003723; F:GO:0003735; P:GO:0006412; C:GO:0022625 F:GO:0003676; F:GO:0004150; F:GO:0008270; P:GO:0046654; P:GO:0046656	P:translational termination; C:chloroplast; P:chloroplast organization; P:thylakoid membrane organization; F:translation release factor activity, codon specific; P:plastid translation; F:ribosome binding
PhvuL001G203900	peptide chain release factor APG3, chloroplastic		
PhvuL001G204000	spermatogenesis-associated protein 20 isoform X1	F:GO:0003824; C:GO:0009507	F:catalytic activity; C:chloroplast
PhvuL001G204100	60S ribosomal protein L19-1-like	F:GO:0003723; F:GO:0003735; P:GO:0006412; C:GO:0022625 F:GO:0003676; F:GO:0004150; F:GO:0008270; P:GO:0046654; P:GO:0046656	F:RNA binding; F:structural constituent of ribosome; P:translation; C:cytosolic large ribosomal subunit
PhvuL001G204200	dihydroneopterin aldolase 2-like	F:GO:0003676; F:GO:0004150; F:GO:0008270; P:GO:0046654; P:GO:0046656	F:nucleic acid binding; F:dihydroneopterin aldolase activity; F:zinc ion binding; P:tetrahydrofolate biosynthetic process; P:folic acid biosynthetic process
PhvuL002G304700	hypothetical protein PHAVU_002G304700g	C:GO:0000276; P:GO:0015986; F:GO:0046933	C:mitochondrial proton-transporting ATP synthase complex, coupling factor F(o); P:ATP synthesis coupled proton transport; F:proton-transporting ATP synthase activity, rotational mechanism
PhvuL002G304800	pollen-specific protein SF21-like	C:GO:0016021; F:GO:0016787 F:GO:0004177; P:GO:0006508; F:GO:0030145; F:GO:0102009	C:integral component of membrane; F:hydrolase activity
PhvuL002G304900	xaa-Pro dipeptidase	F:GO:0006508; F:GO:0030145; F:GO:0102009	F:aminopeptidase activity; P:proteolysis; F:manganese ion binding; F:proline dipeptidase activity
PhvuL002G305000	1-aminocyclopropane-1-carboxylate synthase 7-like	F:GO:0003824; P:GO:0009058; F:GO:0030170	F:catalytic activity; P:biosynthetic process; F:pyridoxal phosphate binding
PhvuL002G305100	hypothetical protein PHAVU_002G305100g	C:GO:0000276; P:GO:0015986; F:GO:0046933	C:mitochondrial proton-transporting ATP synthase complex, coupling factor F(o); P:ATP synthesis coupled proton transport; F:proton-transporting ATP synthase activity, rotational mechanism
PhvuL002G305200	probable caffeoyl-CoA O-methyltransferase At4g26220	F:GO:0008171; P:GO:0032259	F:O-methyltransferase activity; P:methylation

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.002G305300	40S ribosomal protein S15a-5	F:GO:0003735; P:GO:0006412; C:GO:0022627	F:structural constituent of ribosome; P:translation; C:cytosolic small ribosomal subunit
Phvul.002G305400	F-box/LRR-repeat protein At4g29420		
Phvul.002G305500	zinc protease PQQL-like isoform X1	F:GO:0004222; P:GO:0006508; F:GO:0046872 P:GO:0002181;	F:metalloendopeptidase activity; P:proteolysis; F:metal ion binding P:cytoplasmic translation; F:structural
Phvul.002G305600	60S ribosomal protein L31	F:GO:0003735; C:GO:0022625	constituent of ribosome; C:cytosolic large ribosomal subunit
Phvul.002G305700	hypothetical protein PHAVU_002G305700g	C:GO:0016021	C:integral component of membrane
Phvul.002G305800	asparagine--tRNA ligase, cytoplasmic 1	F:GO:0003676; F:GO:0004816; F:GO:0005524; P:GO:0006421	F:nucleic acid binding; F:asparagine-tRNA ligase activity; F:ATP binding; P:asparaginyl-tRNA aminoacylation
Phvul.002G305900	vacuolar protein sorting-associated protein 32 homolog 2	P:GO:0007034	P:vacuolar transport
Phvul.002G306000	transcription factor MYB36-like	F:GO:0003700; C:GO:0005634; P:GO:0006355; C:GO:0016021; P:GO:0030154; F:GO:0043565; F:GO:0044212 F:GO:0003677;	F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated; C:integral component of membrane; P:cell differentiation; F:sequence-specific DNA binding; F:transcription regulatory region DNA binding
Phvul.005G126600	dehydration-responsive element-binding protein 1F-like	F:GO:0003700; C:GO:0005634; P:GO:0006355	F:DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated
Phvul.005G126700	hypothetical protein PHAVU_005G126700g		
Phvul.005G126800	probable ribosome biogenesis protein RLP24	P:GO:0000027; F:GO:0003723; F:GO:0003735; P:GO:0006412; C:GO:0022625; P:GO:1902626	P:ribosomal large subunit assembly; F:RNA binding; F:structural constituent of ribosome; P:translation; C:cytosolic large ribosomal subunit; P:assembly of large subunit precursor of preribosome
Phvul.005G126900	RNase P Rpr2/Rpp21 subunit domain-containing protein		
Phvul.005G127000	nucleobase-ascorbate transporter 12	C:GO:0005886; C:GO:0016021; F:GO:0022857; P:GO:0055085 F:GO:0004664;	C:plasma membrane; C:integral component of membrane; F:transmembrane transporter activity; P:transmembrane transport F:prephenate dehydratase activity; P:L-phenylalanine biosynthetic process;
Phvul.005G127100	arogenate dehydratase/prephenate dehydratase 6, chloroplastic-like	P:GO:0009094; C:GO:0009570; F:GO:0047769 C:GO:0005576;	C:chloroplast stroma; F:arogenate dehydratase activity C:extracellular region; P:signal transduction;
Phvul.005G127200	putative phytosulfokines 6 isoform X1	P:GO:0007165; F:GO:0008083; P:GO:0008283	F:growth factor activity; P:cell population proliferation
Phvul.005G127300	protein TIC 100		
Phvul.005G127400	actin-related protein 2/3 complex subunit 1B-like isoform X1	C:GO:0005737; C:GO:0005885; P:GO:0034314; F:GO:0051015	C:cytoplasm; C:Arp2/3 protein complex; P:Arp2/3 complex-mediated actin nucleation; F:actin filament binding
Phvul.005G127500	hypothetical protein PHAVU_005G127500g		
Phvul.005G127600	histone deacetylase HDT1	F:GO:0003676; C:GO:0005730; P:GO:0010162; P:GO:0045892 P:GO:0000398;	F:nucleic acid binding; C:nucleolus; P:seed dormancy process; P:negative regulation of transcription, DNA-templated P:mRNA splicing, via spliceosome;
Phvul.005G127700	splicing factor U2af small subunit B-like	C:GO:0005681; F:GO:0030628; F:GO:0046872; C:GO:0089701 P:GO:0000398;	C:spliceosomal complex; F:pre-mRNA 3'-splice site binding; F:metal ion binding; C:U2AF P:mRNA splicing, via spliceosome;
Phvul.005G127800	splicing factor U2af small subunit B-like	C:GO:0005681; F:GO:0030628; F:GO:0046872; C:GO:0089701	C:spliceosomal complex; F:pre-mRNA 3'-splice site binding; F:metal ion binding; C:U2AF
Phvul.005G127900	AT-hook motif nuclear-localized protein 9-like	F:GO:0003677; F:GO:0003680	F:DNA binding; F:AT DNA binding

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
PhvuL005G128000	DNA repair protein RAD5A isoform X1	F:GO:0003676; F:GO:0005524; F:GO:0008270; F:GO:0016818	F:nucleic acid binding; F:ATP binding; F:zinc ion binding; F:hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
PhvuL005G128100	DDT domain-containing protein PTM-like	F:GO:0046872	F:metal ion binding
PhvuL005G128200	AP-2 complex subunit alpha-1	P:GO:0006886; C:GO:0030122; F:GO:0035615; P:GO:0072583	P:intracellular protein transport; C:AP-2 adaptor complex; F:clathrin adaptor activity; P:clathrin-dependent endocytosis
PhvuL005G128300	protein RETICULATA-RELATED 1, chloroplastic	C:GO:0005739; C:GO:0009706; C:GO:0016021; P:GO:0099402	C:mitochondrion; C:chloroplast inner membrane; C:integral component of membrane; P:plant organ development
PhvuL005G128400	BTB/POZ domain-containing protein At3g44820		
PhvuL005G128500	UDP-D-apiose/UDP-D-xylose synthase 2	F:GO:0003824; F:GO:0050662	F:catalytic activity; F:coenzyme binding
PhvuL005G128600	GDSL esterase/lipase At5g22810	F:GO:0004806; P:GO:0006629	F:triglyceride lipase activity; P:lipid metabolic process
PhvuL005G128700	protein saal1 isoform X2		
PhvuL005G128800	magnesium transporter MRS2-11, chloroplastic	C:GO:0009941; P:GO:0010027; P:GO:0010117; P:GO:0010960; F:GO:0015095; C:GO:0016021; P:GO:1903830 P:GO:0000245; F:GO:0004674; F:GO:0005524; C:GO:0005634; C:GO:0005737; P:GO:0006468; P:GO:0035556; P:GO:0050684 F:GO:0004190; P:GO:0006508; C:GO:0016021; P:GO:0030163	C:chloroplast envelope; P:thylakoid membrane organization; P:photoprotection; P:magnesium ion homeostasis; F:magnesium ion transmembrane transporter activity; C:integral component of membrane; P:magnesium ion transmembrane transport
PhvuL005G128900	SRSF protein kinase 1-like	P:GO:0000245; F:GO:0004674; F:GO:0005524; C:GO:0005634; C:GO:0005737; P:GO:0006468; P:GO:0035556; P:GO:0050684 F:GO:0004190; P:GO:0006508; C:GO:0016021; P:GO:0030163	P:spliceosomal complex assembly; F:protein serine/threonine kinase activity; F:ATP binding; C:nucleus; C:cytoplasm; P:protein phosphorylation; P:intracellular signal transduction; P:regulation of mRNA processing
PhvuL005G129000	aspartic proteinase-like protein 2	F:GO:0004180; P:GO:0006508; F:GO:0008236; F:GO:0008239; C:GO:0016021	F:aspartic-type endopeptidase activity; P:proteolysis; C:integral component of membrane; P:protein catabolic process
PhvuL005G129100	lysosomal Pro-X carboxypeptidase-like	F:GO:0004180; P:GO:0006508; F:GO:0008236; F:GO:0008239; C:GO:0016021	F:carboxypeptidase activity; P:proteolysis; F:serine-type peptidase activity; F:dipeptidyl-peptidase activity; C:integral component of membrane
PhvuL005G129200	lysosomal Pro-X carboxypeptidase	F:GO:0004180; P:GO:0006508; F:GO:0008236; F:GO:0008239; C:GO:0016021	F:carboxypeptidase activity; P:proteolysis; F:serine-type peptidase activity; F:dipeptidyl-peptidase activity; C:integral component of membrane
PhvuL005G129300	lysosomal Pro-X carboxypeptidase	F:GO:0004180; P:GO:0006508; F:GO:0008236; F:GO:0008239; C:GO:0016021	F:carboxypeptidase activity; P:proteolysis; F:serine-type peptidase activity; F:dipeptidyl-peptidase activity; C:integral component of membrane
PhvuL005G129400	lysosomal Pro-X carboxypeptidase-like	F:GO:0004180; P:GO:0006508; F:GO:0008236; F:GO:0008239	F:carboxypeptidase activity; P:proteolysis; F:serine-type peptidase activity; F:dipeptidyl-peptidase activity
PhvuL005G129500	lysosomal Pro-X carboxypeptidase	F:GO:0004180; P:GO:0006508; F:GO:0008236; F:GO:0008239	F:carboxypeptidase activity; P:proteolysis; F:serine-type peptidase activity; F:dipeptidyl-peptidase activity
PhvuL005G129600	calcium-binding protein CP1	F:GO:0005509	F:calcium ion binding
PhvuL005G129700	RING finger protein PFF0165c isoform X2		
PhvuL005G129800	hypothetical protein PHAVU_005G129800g	F:GO:0000977; F:GO:0003682; C:GO:0005634; F:GO:0042393; P:GO:0045944; F:GO:0046872	F:RNA polymerase II regulatory region sequence-specific DNA binding; F:chromatin binding; C:nucleus; F:histone binding; P:positive regulation of transcription by RNA polymerase II; F:metal ion binding

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.005G129900	uncharacterized isomerase BH0283-like	P:GO:0009058; F:GO:0102943	P:biosynthetic process; F:trans-2,3-dihydro-3-hydroxy-anthranilate isomerase activity
Phvul.005G130000	serine carboxypeptidase-like 51	F:GO:0004185; P:GO:0051603	F:serine-type carboxypeptidase activity; P:proteolysis involved in cellular protein catabolic process
Phvul.005G130100	protein SENSITIVE TO PROTON RHIZOTOXICITY 2-like	F:GO:0003676	F:nucleic acid binding
Phvul.005G130200	hypothetical protein PHAVU_005G130200g, partial		
Phvul.005G130300	zinc finger protein ZAT5-like	F:GO:0003676	F:nucleic acid binding
Phvul.005G130400	hypothetical protein PHAVU_005G130400g	P:GO:0000184	P:nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
Phvul.005G130500	transcription factor FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR	F:GO:0046983	F:protein dimerization activity
Phvul.005G130600	protein UPSTREAM OF FLC isoform X1		
Phvul.005G130700	protein enabled homolog		
Phvul.005G130800	probable glucuronoxylan glucuronosyltransferase IRX7	P:GO:0006486; P:GO:0010417; C:GO:0016021; F:GO:0102983 F:GO:0004553; C:GO:0005618; P:GO:0010411; F:GO:0016762; P:GO:0042546; C:GO:0048046; P:GO:0071555	P:protein glycosylation; P:glucuronoxylan biosynthetic process; C:integral component of membrane; F:xylogalacturonan beta-1,3-xylosyltransferase activity
Phvul.005G130900	xyloglucan endotransglucosylase/hydrolase protein 31-like	F:GO:0004185; C:GO:0005773; P:GO:0051603	F:hydrolase activity, hydrolyzing O-glycosyl compounds; C:cell wall; P:xyloglucan metabolic process; F:xyloglucan:xyloglucosyl transferase activity; P:cell wall biogenesis; C:apoplast; P:cell wall organization
Phvul.005G131000	serine carboxypeptidase-like		F:serine-type carboxypeptidase activity; C:vacuole; P:proteolysis involved in cellular protein catabolic process
Phvul.005G131100	hypothetical protein PHAVU_005G131100g		
Phvul.005G131200	alpha-L-fucosidase 1	F:GO:0004560; C:GO:0005764; P:GO:0006004; C:GO:0016021; P:GO:0016139 F:GO:0003700; C:GO:0005634; P:GO:0006355; P:GO:0030154; F:GO:0043565; F:GO:0044212	F:alpha-L-fucosidase activity; C:llysosome; P:fucose metabolic process; C:integral component of membrane; P:glycoside catabolic process
Phvul.005G131300	transcription factor RAX2-like	C:GO:0016021; F:GO:0016491; F:GO:0016874; P:GO:0055114	F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated; P:cell differentiation; F:sequence-specific DNA binding; F:transcription regulatory region DNA binding
Phvul.005G131400	4-coumarate--CoA ligase-like 7	P:GO:0030001; F:GO:0046872	C:integral component of membrane; F:oxidoreductase activity; F:ligase activity; P:oxidation-reduction process
Phvul.005G131500	heavy metal-associated isoprenylated plant protein 3-like	F:GO:0005524; F:GO:0016887	P:metal ion transport; F:metal ion binding
Phvul.005G131600	ABC transporter F family member 1-like		F:ATP binding; F:ATPase activity
Phvul.005G131700	hypothetical protein PHAVU_005G131700g	F:GO:0003723	F:RNA binding
Phvul.005G131800	auxin-responsive protein SAUR50	P:GO:0009733	P:response to auxin
Phvul.005G131900	dolichol kinase EVAN isoform X2	F:GO:0004168; P:GO:0016310; C:GO:0030176; P:GO:0043048 F:GO:0005524; C:GO:0005886; F:GO:0015419; C:GO:0016021; F:GO:0016887; P:GO:1902358	F:dolichol kinase activity; P:phosphorylation; C:integral component of endoplasmic reticulum membrane; P:dolichyl monophosphate biosynthetic process
Phvul.005G132000	ABC transporter G family member 3	P:GO:1902358	F:ATP binding; C:plasma membrane; F:ATPase-coupled sulfate transmembrane transporter activity; C:integral component of membrane; F:ATPase activity; P:sulfate transmembrane transport
Phvul.010G029800	TMV resistance protein N-like	P:GO:0007165; F:GO:0043531	P:signal transduction; F:ADP binding
Phvul.010G029900	E3 ubiquitin-protein ligase CHIP	F:GO:0004839; F:GO:0004842; P:GO:0016567; F:GO:0016746	F:ubiquitin activating enzyme activity; F:ubiquitin-protein transferase activity; P:protein ubiquitination; F:transferase activity, transferring acyl groups

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
PhvuL010G030000	sugar transport protein 1-like	P:GO:0008643; C:GO:0016021; F:GO:0022857; P:GO:0055085 F:GO:0004674;	P:carbohydrate transport; C:integral component of membrane; F:transmembrane transporter activity; P:transmembrane transport
PhvuL010G030100	serine/threonine-protein kinase OXI1-like	F:GO:0005524; C:GO:0005634; C:GO:0005737; C:GO:0005886; P:GO:0006468	F:protein serine/threonine kinase activity; F:ATP binding; C:nucleus; C:cytoplasm; C:plasma membrane; P:protein phosphorylation
PhvuL010G030200	DNA-directed RNA polymerase II subunit 4	F:GO:0000166; P:GO:0000288; C:GO:0000932; F:GO:0003697; F:GO:0003727; F:GO:0003899; C:GO:0005665; P:GO:0006367; F:GO:0031369; P:GO:0031990; P:GO:0034402; P:GO:0045948	F:nucleotide binding; P:nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay; C:P-body; F:single-stranded DNA binding; F:single-stranded RNA binding; F:DNA-directed 5'-3' RNA polymerase activity; C:RNA polymerase II, core complex; P:transcription initiation from RNA polymerase II promoter; F:translation initiation factor binding; P:mRNA export from nucleus in response to heat stress; P:recruitment of 3'-end processing factors to RNA polymerase II holoenzyme complex; P:positive regulation of translational initiation
PhvuL010G030300	sugar transport protein 1-like	P:GO:0008643; C:GO:0016021; F:GO:0022857; P:GO:0055085	P:carbohydrate transport; C:integral component of membrane; F:transmembrane transporter activity; P:transmembrane transport
PhvuL010G030400	hypothetical protein PHAVU_011G0934000g, partial	F:GO:0008270	F:zinc ion binding
PhvuL010G030500	auxin-induced in root cultures protein 12	C:GO:0016021	C:integral component of membrane
PhvuL010G030600	PREDICTED: uncharacterized protein LOC108347394 isoform X1	F:GO:0005525; P:GO:0006629; C:GO:0016021; F:GO:0016787 C:GO:0000145; C:GO:0005829; P:GO:0006893; P:GO:0006904; P:GO:0009846; P:GO:0009860; P:GO:0060321 F:GO:0004816; F:GO:0005524; C:GO:0005737; P:GO:0006421; C:GO:0043231	F:GTP binding; P:lipid metabolic process; C:integral component of membrane; F:hydrolase activity
PhvuL010G030700	exocyst complex component SEC15A	C:GO:0005829; P:GO:0006893; P:GO:0006904; P:GO:0009846; P:GO:0009860; P:GO:0060321 F:GO:0004816; F:GO:0005524; C:GO:0005737; P:GO:0006421; C:GO:0043231	C:exocyst; C:cytosol; P:Golgi to plasma membrane transport; P:vesicle docking involved in exocytosis; P:pollen germination; P:pollen tube growth; P:acceptance of pollen
PhvuL010G030800	asparagine--tRNA ligase, cytoplasmic 2	F:GO:0004816; F:GO:0005524; C:GO:0005737; P:GO:0006421; C:GO:0043231	F:asparagine-tRNA ligase activity; F:ATP binding; C:cytoplasm; P:asparaginyl-tRNA aminoacylation; C:intracellular membrane-bounded organelle
PhvuL010G030900	hypothetical protein PHAVU_010G030900g	F:GO:0003677	F:DNA binding

APPENDIX H. Caterpillar plot of adjusted means with 95% confidence intervals for seven seed traits: seed length (a; LENGTH); width (b; WIDTH); perimeter (c; PERIM); projected area (d; AREA); length-to-width ratio (e; LWR); circularity (f; CIRC) and hundred seed mass (g; HSM), in 180 genotypes of the Andean (lighter dashes) and Mesoamerican (darker dashes) gene pools

APPENDIX I. Boxplots relating the adjusted means of common bean genotypes for SNPs significantly associated according to GWAS. Seed length (a and b; LENGTH) is represented in red; seed projected area (c and d; AREA) in light blue; seed width (e; WIDTH) in green; seed perimeter (f; PERIM) in blue; and circularity (g; CIRC) in yellow. No associations were detected for the remaining seed traits

APPENDIX J. Linkage disequilibrium (LD) analysis to define genomic regions for candidate gene discovery among significant SNPs: S1_325844152 (a), S1_507205534 (b), S1_53011276 (c) and S1_379992973 (d), and flanking markers in linkage disequilibrium with common bean seed traits. The heatmap indicates the degree of estimated LD

APPENDIX K. Functional annotation and ontology terms predicted for candidate genes located in genomic regions significantly associated with common bean seed shape and size

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
PhvuL002G004100.1	expansin-like B1	C:GO:0005576	C:extracellular region
PhvuL002G004300.2	haloacid dehalogenase-like hydrolase domain-containing protein Sgpp isoform X5	F:GO:0016787	F:hydrolase activity
PhvuL002G004300.1	haloacid dehalogenase-like hydrolase domain-containing protein Sgpp isoform X5	F:GO:0016787	F:hydrolase activity
PhvuL002G004400.1	pentatricopeptide repeat-containing protein Atlg32415, mitochondrial		
PhvuL002G004500.1	hypothetical protein PHAVU_002G004500g	C:GO:0009941	C:chloroplast envelope
PhvuL002G004600.2	vacuolar protein sorting-associated protein 55 homolog	C:GO:0005768; C:GO:0016021; P:GO:0032511; P:GO:2000009	C:endosome; C:integral component of membrane; P:late endosome to vacuole transport via multivesicular body sorting pathway; P:negative regulation of protein localization to cell surface
PhvuL002G004600.1	vacuolar protein sorting-associated protein 55 homolog	C:GO:0005768; C:GO:0016021; P:GO:0032511; P:GO:2000009	C:endosome; C:integral component of membrane; P:late endosome to vacuole transport via multivesicular body sorting pathway; P:negative regulation of protein localization to cell surface
PhvuL002G004700.1	hypothetical protein PHAVU_002G004700g	C:GO:0005783; C:GO:0016021; P:GO:0000160; F:GO:0000976;	C:endoplasmic reticulum; C:integral component of membrane
PhvuL002G004800.2	two-component response regulator-like APRR2 isoform X1	F:GO:0003700; C:GO:0005634; P:GO:0045893 P:GO:0000160;	P:phosphorelay signal transduction system; F:transcription regulatory region sequence-specific DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:positive regulation of transcription, DNA-templated
PhvuL002G004800.1	two-component response regulator-like APRR2 isoform X1	F:GO:0003700; C:GO:0005634; P:GO:0045893	P:phosphorelay signal transduction system; F:transcription regulatory region sequence-specific DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:positive regulation of transcription, DNA-templated
PhvuL002G004900.1	tobamovirus multiplication protein 2A-like	C:GO:0016021	C:integral component of membrane
PhvuL002G004900.2	tobamovirus multiplication protein 2A-like	C:GO:0016021	C:integral component of membrane
PhvuL002G005000.1	tobamovirus multiplication protein 2A-like	C:GO:0016021	C:integral component of membrane
PhvuL002G005100.1	tobamovirus multiplication protein 2A-like	C:GO:0016021	C:integral component of membrane
PhvuL002G005100.2	tobamovirus multiplication protein 2A-like	C:GO:0016021	C:integral component of membrane
PhvuL002G005200.1	putative ABC transporter C family member 15		
PhvuL002G005300.1	type I inositol polyphosphate 5-phosphatase 2-like	F:GO:0004439; F:GO:0004445; F:GO:0034485; P:GO:0046856; F:GO:0052658; F:GO:0052659	F:phosphatidylinositol-4,5-bisphosphate 5-phosphatase activity; F:inositol-polyphosphate 5-phosphatase activity; F:phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase activity; P:phosphatidylinositol dephosphorylation; F:inositol-1,4,5-trisphosphate 5-phosphatase activity; F:inositol-1,3,4,5-tetrakisphosphate 5-phosphatase activity
PhvuL002G005300.2	type I inositol polyphosphate 5-phosphatase 2-like	F:GO:0004439; F:GO:0004445; F:GO:0034485; P:GO:0046856; F:GO:0052658; F:GO:0052659	F:phosphatidylinositol-4,5-bisphosphate 5-phosphatase activity; F:inositol-polyphosphate 5-phosphatase activity; F:phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase activity; P:phosphatidylinositol dephosphorylation; F:inositol-1,4,5-trisphosphate 5-phosphatase activity; F:inositol-1,3,4,5-tetrakisphosphate 5-phosphatase activity
PhvuL002G005400.1	peroxidase 20	F:GO:0004601; C:GO:0005576; P:GO:0006979; F:GO:0020037; P:GO:0042744; F:GO:0046872; P:GO:0055114; P:GO:0098869	F:peroxidase activity; C:extracellular region; P:response to oxidative stress; F:heme binding; P:hydrogen peroxide catabolic process; F:metal ion binding; P:oxidation-reduction process; P:cellular oxidant detoxification
PhvuL002G005400.2	peroxidase 20	F:GO:0004601; C:GO:0005576; P:GO:0006979; F:GO:0020037; P:GO:0042744; F:GO:0046872; P:GO:0055114; P:GO:0098869	F:peroxidase activity; C:extracellular region; P:response to oxidative stress; F:heme binding; P:hydrogen peroxide catabolic process; F:metal ion binding; P:oxidation-reduction process; P:cellular oxidant detoxification
PhvuL002G005500.1	myosin-G heavy chain		

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.002G005600.1.1	pyruvate kinase isozyme G, chloroplastic	F:GO:0000287; F:GO:0004743; P:GO:0006096; C:GO:0009570; F:GO:0016301; F:GO:0030955	F:magnesium ion binding; F:pyruvate kinase activity; P:glycolytic process; C:chloroplast stroma; F:kinase activity; F:potassium ion binding
Phvul.002G005600.0.1	pyruvate kinase isozyme G, chloroplastic	F:GO:0000287; F:GO:0004743; P:GO:0006096; C:GO:0009570; F:GO:0016301; F:GO:0030955	F:magnesium ion binding; F:pyruvate kinase activity; P:glycolytic process; C:chloroplast stroma; F:kinase activity; F:potassium ion binding
Phvul.002G005700.1	protein NRT1/ PTR FAMILY 7.3	C:GO:0016021; F:GO:0022857; P:GO:0055085	C:integral component of membrane; F:transmembrane transporter activity; P:transmembrane transport
Phvul.002G005800.1	hypothetical protein PHAVU_002G005800g		
Phvul.002G005800.2	hypothetical protein PHAVU_002G005800g		
Phvul.002G005900.1	beta-amylin 11-oxidase-like	F:GO:0005506; C:GO:0005783; P:GO:0007275; P:GO:0010268; C:GO:0016021; P:GO:0016125; P:GO:0016132; F:GO:0020037; F:GO:0051777; P:GO:0055114; F:GO:0102289; F:GO:0102290; F:GO:0102291	F:iron ion binding; C:endoplasmic reticulum; P:multicellular organism development; P:brassinosteroid homeostasis; C:integral component of membrane; P:sterol metabolic process; P:brassinosteroid biosynthetic process; F:heme binding; F:ent-kaurenoate oxidase activity; P:oxidation-reduction process; F:beta-amylin 11-oxidase activity; F:beta-amylin monooxygenase activity; F:11alpha-hydroxy-beta-amylin dehydrogenase activity
Phvul.002G006000.1	glycine cleavage system H protein, mitochondrial	C:GO:0005739; C:GO:0005960; P:GO:0019464	C:mitochondrion; C:glycine cleavage complex; P:glycine decarboxylation via glycine cleavage system
Phvul.002G006000.2	glycine cleavage system H protein, mitochondrial	C:GO:0005739; C:GO:0005960; P:GO:0019464	C:mitochondrion; C:glycine cleavage complex; P:glycine decarboxylation via glycine cleavage system
Phvul.002G006100.1	protein phosphatase 2C 29-like	F:GO:0004724; P:GO:0006470; P:GO:0009653; P:GO:0048507	F:magnesium-dependent protein serine/threonine phosphatase activity; P:protein dephosphorylation; P:anatomical structure morphogenesis; P:meristem development
Phvul.002G006200.1	protein ABCI7, chloroplastic	P:GO:0016226	P:iron-sulfur cluster assembly
Phvul.002G006300.1	aluminum-activated malate transporter 12-like	C:GO:0009705; P:GO:0015743; C:GO:0016021	C:plant-type vacuole membrane; P:malate transport; C:integral component of membrane
Phvul.002G006400.2	MND1-interacting protein 1		
Phvul.002G006400.1	MND1-interacting protein 1		
Phvul.002G006600.1	hypothetical protein PHAVU_002G006600g		
Phvul.002G006700.1	AT-hook motif nuclear-localized protein 8	F:GO:0003680; C:GO:0005634	F:minor groove of adenine-thymine-rich DNA binding; C:nucleus
Phvul.002G006800.1	auxin-responsive protein SAUR77-like		
Phvul.002G006900.1	AT-hook motif nuclear-localized protein 23-like	F:GO:0003680; F:GO:0003700; C:GO:0005634; P:GO:0006355	F:minor groove of adenine-thymine-rich DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated
Phvul.002G007000.1	hypothetical protein PHAVU_002G007000g		
Phvul.002G007100.1	transcriptional regulator SUPERMAN		
Phvul.002G007200.1	acetylornithine deacetylase	F:GO:0008777	F:acetylornithine deacetylase activity
Phvul.002G007300.1	protein STAY-GREEN	F:GO:0004222; C:GO:0016021; P:GO:0071586	F:metalloendopeptidase activity; C:integral component of membrane; P:CAAX-box protein processing
Phvul.002G007300.2	protein STAY-GREEN	F:GO:0004222; C:GO:0016021; P:GO:0071586	F:metalloendopeptidase activity; C:integral component of membrane; P:CAAX-box protein processing
Phvul.002G007400.1	transcription factor bHLH18-like	F:GO:0046983	F:protein dimerization activity
Phvul.002G007500.1	transcription factor bHLH18-like	F:GO:0046983	F:protein dimerization activity

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.002G007600.1	leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1	F:GO:0004672; F:GO:0005524; C:GO:0005886; P:GO:0006468; P:GO:0009755; C:GO:0016021	F:protein kinase activity; F:ATP binding; C:plasma membrane; P:protein phosphorylation; P:hormone-mediated signaling pathway; C:integral component of membrane
Phvul.002G007700.1	hypothetical protein PHAVU_002G007700g	F:GO:0004674; F:GO:0005524; C:GO:0005886; P:GO:0006468; C:GO:0016021; F:GO:0016491; P:GO:0050794 P:GO:0000226; P:GO:0000278;	F:protein serine/threonine kinase activity; F:ATP binding; C:plasma membrane; P:protein phosphorylation; C:integral component of membrane; F:oxidoreductase activity; P:regulation of cellular process
Phvul.002G007900.1	tubulin alpha-1 chain	F:GO:0003924; F:GO:0005200; F:GO:0005525; C:GO:0005737; C:GO:0005874; P:GO:0090378	P:microtubule cytoskeleton organization; P:mitotic cell cycle; F:GTPase activity; F:structural constituent of cytoskeleton; F:GTP binding; C:cytoplasm; C:microtubule; P:seed trichome elongation
Phvul.002G008000.1	hypothetical protein PHAVU_002G008000g		
Phvul.008G018900.1	protein bicaudal C homolog 1-A-like		
Phvul.008G019000.1	RNA polymerase II C-terminal domain phosphatase-like 3	C:GO:0005634; F:GO:0008420; P:GO:0070940 F:GO:0004185; C:GO:0005773; P:GO:0006508 P:GO:0002229; F:GO:0004675; F:GO:0005524; C:GO:0005886; P:GO:0007178; C:GO:0009506; C:GO:0016021; P:GO:0042742; P:GO:0046777 P:GO:0002229;	C:nucleus; F:RNA polymerase II CTD heptapeptide repeat phosphatase activity; P:dephosphorylation of RNA polymerase II C-terminal domain
Phvul.008G019100.1	serine carboxypeptidase-like 45	F:GO:0004185; C:GO:0005773; P:GO:0006508 P:GO:0002229; F:GO:0004675; F:GO:0005524; C:GO:0005886; P:GO:0007178; C:GO:0009506; C:GO:0016021; P:GO:0042742; P:GO:0046777 P:GO:0002229;	F:serine-type carboxypeptidase activity; C:vacuole; P:proteolysis
Phvul.008G019200.2	inactive leucine-rich repeat receptor-like protein kinase CORYNE isoform X2	F:GO:0004675; F:GO:0005524; C:GO:0005886; P:GO:0007178; C:GO:0009506; C:GO:0016021; P:GO:0042742; P:GO:0046777 P:GO:0002229;	P:defense response to oomycetes; F:transmembrane receptor protein serine/threonine kinase activity; F:ATP binding; C:plasma membrane; P:transmembrane receptor protein serine/threonine kinase signaling pathway; C:plasmodesma; C:integral component of membrane; P:defense response to bacterium; P:protein autophosphorylation
Phvul.008G019200.1	inactive leucine-rich repeat receptor-like protein kinase CORYNE isoform X1	F:GO:0004675; F:GO:0005524; C:GO:0005886; P:GO:0007178; C:GO:0009506; C:GO:0016021; P:GO:0042742; P:GO:0046777 F:GO:0005096; C:GO:0005829; C:GO:0005886; P:GO:0006897; P:GO:0009733; P:GO:0009965; P:GO:0010051; P:GO:0010087; C:GO:0030140; F:GO:0035091; P:GO:0043547; F:GO:0046872 F:GO:0005096; C:GO:0005829; C:GO:0005886; P:GO:0006897; P:GO:0009733; P:GO:0009965; P:GO:0010051; P:GO:0010087; C:GO:0030140; F:GO:0035091; P:GO:0043547; F:GO:0046872	P:defense response to oomycetes; F:transmembrane receptor protein serine/threonine kinase activity; F:ATP binding; C:plasma membrane; P:transmembrane receptor protein serine/threonine kinase signaling pathway; C:plasmodesma; C:integral component of membrane; P:defense response to bacterium; P:protein autophosphorylation
Phvul.008G019300.2	ADP-ribosylation factor GTPase-activating protein AGD3	F:GO:0005096; C:GO:0005829; C:GO:0005886; P:GO:0006897; P:GO:0009733; P:GO:0009965; P:GO:0010051; P:GO:0010087; C:GO:0030140; F:GO:0035091; P:GO:0043547; F:GO:0046872 F:GO:0005096; C:GO:0005829; C:GO:0005886; P:GO:0006897; P:GO:0009733; P:GO:0009965; P:GO:0010051; P:GO:0010087; C:GO:0030140; F:GO:0035091; P:GO:0043547; F:GO:0046872	F:GTPase activator activity; C:cytosol; C:plasma membrane; P:endocytosis; P:response to auxin; P:leaf morphogenesis; P:xylem and phloem pattern formation; P:phloem or xylem histogenesis; C:trans-Golgi network transport vesicle; F:phosphatidylinositol binding; P:positive regulation of GTPase activity; F:metal ion binding
Phvul.008G019300.1	ADP-ribosylation factor GTPase-activating protein AGD3	F:GO:0005096; C:GO:0005829; C:GO:0005886; P:GO:0006897; P:GO:0009733; P:GO:0009965; P:GO:0010051; P:GO:0010087; C:GO:0030140; F:GO:0035091; P:GO:0043547; F:GO:0046872	F:GTPase activator activity; C:cytosol; C:plasma membrane; P:endocytosis; P:response to auxin; P:leaf morphogenesis; P:xylem and phloem pattern formation; P:phloem or xylem histogenesis; C:trans-Golgi network transport vesicle; F:phosphatidylinositol binding; P:positive regulation of GTPase activity; F:metal ion binding

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.008G019400.1	trihelix transcription factor ASR3	F:GO:0003677	F:DNA binding
Phvul.008G019500.1	protein MEI2-like 4 isoform X1	P:GO:0000398; F:GO:0003723; C:GO:0005634	P:mRNA splicing, via spliceosome; F:RNA binding; C:nucleus
Phvul.008G019600.1	ethylene-responsive transcription factor ERF113-like	F:GO:0003677; F:GO:0003700; C:GO:0005634; P:GO:0006355	F:DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated
Phvul.008G019700.1	N-alpha-acetyltransferase MAK3	F:GO:0004596; P:GO:0017196; C:GO:0031417	F:peptide alpha-N-acetyltransferase activity; P:N-terminal peptidyl-methionine acetylation; C:NatC complex
Phvul.008G019800.1	60S ribosomal protein L30	F:GO:0003723; F:GO:0003735; C:GO:0022625	F:RNA binding; F:structural constituent of ribosome; C:cytosolic large ribosomal subunit
Phvul.008G019900.1	hypothetical protein PHAVU_008G019900g	C:GO:0005783; C:GO:0016021	C:endoplasmic reticulum; C:integral component of membrane
Phvul.008G020000.1	DNA-directed RNA polymerases I and III subunit RPAC2	F:GO:0003677; F:GO:0003899; C:GO:0005666; C:GO:0005736; P:GO:0006351; F:GO:0046983	F:DNA binding; F:DNA-directed 5'-3' RNA polymerase activity; C:RNA polymerase III complex; C:RNA polymerase I complex; P:transcription, DNA-templated; F:protein dimerization activity
Phvul.008G020100.1	hypothetical protein PHAVU_008G020100g	F:GO:0005044; P:GO:0006897; C:GO:0016020	F:scavenger receptor activity; P:endocytosis; C:membrane
Phvul.008G020200.1	allene oxide cyclase, chloroplastic-like	P:GO:0007623; P:GO:0009611; P:GO:0009625; P:GO:0009637; P:GO:0009646; P:GO:0009651; P:GO:0009695; P:GO:0009723; P:GO:0009734; P:GO:0009737; P:GO:0009751; P:GO:0009864; P:GO:0009908; P:GO:0010038; P:GO:0010114; P:GO:0010218; P:GO:0033274; P:GO:0042542; F:GO:0046423; P:GO:0048573; P:GO:0050832; P:GO:0080186; P:GO:1900367	P:circadian rhythm; P:response to wounding; P:response to insect; P:response to blue light; P:response to absence of light; P:response to salt stress; P:jasmonic acid biosynthetic process; P:response to ethylene; P:auxin-activated signaling pathway; P:response to abscisic acid; P:response to salicylic acid; P:induced systemic resistance, jasmonic acid mediated signaling pathway; P:flower development; P:response to metal ion; P:response to red light; P:response to far red light; P:response to vitamin B2; P:response to hydrogen peroxide; F:allene-oxide cyclase activity; P:photoperiodism, flowering; P:defense response to fungus; P:developmental vegetative growth; P:positive regulation of defense response to insect
Phvul.008G020400.1	transcription factor DICHOTOMA-like	F:GO:0003700; C:GO:0005634; P:GO:0006355; F:GO:0043565; P:GO:2000032	F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated; F:sequence-specific DNA binding; P:regulation of secondary shoot formation
Phvul.008G020500.1	translation initiation factor eIF-2B subunit delta	F:GO:0003743; P:GO:0006413	F:translation initiation factor activity; P:translational initiation
Phvul.008G020600.1	protein CROWDED NUCLEI 1	C:GO:0005634; P:GO:0006997	C:nucleus; P:nucleus organization
Phvul.008G020700.1	probable disease resistance protein At4g27220	F:GO:0043531	F:ADP binding
Phvul.008G020700.2	probable disease resistance protein At4g27220	F:GO:0043531	F:ADP binding
Phvul.008G020800.1	protein NRT1/ PTR FAMILY 7.1-like	C:GO:0016021; F:GO:0022857; P:GO:0055085	C:integral component of membrane; F:transmembrane transporter activity; P:transmembrane transport
Phvul.008G020900.1	probable disease resistance protein At4g27220 isoform X1	F:GO:0043531	F:ADP binding
Phvul.008G021000.1	proline-rich receptor-like protein kinase PERK8	F:GO:0004672; F:GO:0005524; C:GO:0005886; P:GO:0006468; C:GO:0016021	F:protein kinase activity; F:ATP binding; C:plasma membrane; P:protein phosphorylation; C:integral component of membrane
Phvul.008G021100.1	protein trichome birefringence-like 19	C:GO:0005794; C:GO:0016021; F:GO:0016413	C:Golgi apparatus; C:integral component of membrane; F:O-acetyltransferase activity

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.008G021200.1	protein trichome birefringence-like 19	C:GO:0005794; C:GO:0016021; F:GO:0016413	C:Golgi apparatus; C:integral component of membrane; F:O-acetyltransferase activity
Phvul.008G021400.1	protein FLX-like 2		
Phvul.008G021500.1	cell division cycle protein 27 homolog B-like isoform X1	C:GO:0005680; C:GO:0005737; P:GO:0016567; P:GO:0031145; P:GO:0045842; P:GO:0051301	C:anaphase-promoting complex; C:cytoplasm; P:protein ubiquitination; P:anaphase-promoting complex-dependent catabolic process; P:positive regulation of mitotic metaphase/anaphase transition; P:cell division
Phvul.008G021600.1	probable 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	F:GO:0046537; F:GO:0046872	F:2,3-bisphosphoglycerate-independent phosphoglycerate mutase activity; F:metal ion binding
Phvul.008G021700.1	DNA-directed RNA polymerase III subunit RPC6	F:GO:0003899; C:GO:0005666; P:GO:0006383	F:DNA-directed 5'-3' RNA polymerase activity; C:RNA polymerase III complex; P:transcription by RNA polymerase III
Phvul.008G021800.1	stress-induced protein KIN2-like		
Phvul.008G021900.1	protein YIF1B-like	C:GO:0005789; C:GO:0005793; P:GO:0006888; C:GO:0030134; C:GO:0030173	C:endoplasmic reticulum membrane; C:endoplasmic reticulum-Golgi intermediate compartment; P:endoplasmic reticulum to Golgi vesicle-mediated transport; C:COPII-coated ER to Golgi transport vesicle; C:integral component of Golgi membrane
Phvul.008G021900.2	protein YIF1B-like	C:GO:0005789; C:GO:0005793; P:GO:0006888; C:GO:0030134; C:GO:0030173	C:endoplasmic reticulum membrane; C:endoplasmic reticulum-Golgi intermediate compartment; P:endoplasmic reticulum to Golgi vesicle-mediated transport; C:COPII-coated ER to Golgi transport vesicle; C:integral component of Golgi membrane
Phvul.008G022000.1	proline dehydrogenase 2, mitochondrial	F:GO:0004657; C:GO:0005739; P:GO:0010133; P:GO:0055114; F:GO:0071949	F:proline dehydrogenase activity; C:mitochondrion; P:proline catabolic process to glutamate; P:oxidation-reduction process; F:FAD binding
Phvul.008G022200.1	S-adenosylmethionine decarboxylase proenzyme-like	F:GO:0004014; P:GO:0006597; P:GO:0008295	F:adenosylmethionine decarboxylase activity; P:spermine biosynthetic process; P:spermidine biosynthetic process
Phvul.008G022100.1	S-adenosylmethionine decarboxylase proenzyme-like	F:GO:0004014; C:GO:0005829; P:GO:0006557; P:GO:0006597; P:GO:0008295	F:adenosylmethionine decarboxylase activity; C:cytosol; P:S-adenosylmethionine biosynthetic process; P:spermine biosynthetic process; P:spermidine biosynthetic process
Phvul.008G022300.1	stress-induced protein KIN2-like		
Phvul.008G022400.1	hypothetical protein PHAVU_008G022400g		
Phvul.008G022500.1	short-chain dehydrogenase reductase ATA1	F:GO:0004316; F:GO:0010301; P:GO:0055114; F:GO:0102131; F:GO:0102132	F:3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity; F:xanthoxin dehydrogenase activity; P:oxidation-reduction process; F:3-oxo-glutaryl-[acp] methyl ester reductase activity; F:3-oxo-pimeloyl-[acp] methyl ester reductase activity
Phvul.008G022600.1	pentatricopeptide repeat-containing protein At3g02490, mitochondrial		
Phvul.008G022700.1	protein NSP-INTERACTING KINASE 1 isoform X1	F:GO:0004674; F:GO:0005524; C:GO:0005886; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; C:plasma membrane; P:protein phosphorylation; C:integral component of membrane
Phvul.008G022700.2	protein NSP-INTERACTING KINASE 1 isoform X1	F:GO:0004674; F:GO:0005524; C:GO:0005886; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; C:plasma membrane; P:protein phosphorylation; C:integral component of membrane
Phvul.008G022800.1	zinc finger protein CONSTANS-LIKE 2	C:GO:0005634; F:GO:0008270; P:GO:0009909	C:nucleus; F:zinc ion binding; P:regulation of flower development
Phvul.008G022900.1	ribosome biogenesis protein BRX1-like	P:GO:0000027; C:GO:0005730	P:ribosomal large subunit assembly; C:nucleolus
Phvul.008G022900.2	ribosome biogenesis protein BRX1-like	P:GO:0000027; C:GO:0005730	P:ribosomal large subunit assembly; C:nucleolus
Phvul.008G023000_1.1	ribosome biogenesis protein BRX1-like	P:GO:0000027; C:GO:0005730	P:ribosomal large subunit assembly; C:nucleolus
Phvul.008G023000_0.1	---NA---		
Phvul.008G023100.1	defensin-like protein 242	P:GO:0007165; C:GO:0016020	P:signal transduction; C:membrane

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.008G023200.1	putative defensin-like protein 244	P:GO:0007165; C:GO:0016021	P:signal transduction; C:integral component of membrane
Phvul.008G023300.1	putative defensin-like protein 244	P:GO:0007165; C:GO:0016021	P:signal transduction; C:integral component of membrane
Phvul.008G023400.1	protein GLUTAMINE DUMPER 6-like	C:GO:0016021; P:GO:0080143	C:integral component of membrane; P:regulation of amino acid export
Phvul.008G023500.1	hypothetical protein PHAVU_008G023500g		
Phvul.008G023600.1	hypothetical protein PHAVU_008G023600g, partial		
Phvul.008G023700.1	6-phosphogluconate dehydrogenase, decarboxylating 3	F:GO:0004616; C:GO:0005829; P:GO:0009051; P:GO:0046177; F:GO:0050661	F:phosphogluconate dehydrogenase (decarboxylating) activity; C:cytosol; P:pentose-phosphate shunt, oxidative branch; P:D-gluconate catabolic process; F:NADP binding
Phvul.008G023700.2	6-phosphogluconate dehydrogenase, decarboxylating 3	F:GO:0004616; C:GO:0005829; P:GO:0009051; P:GO:0046177; F:GO:0050661	F:phosphogluconate dehydrogenase (decarboxylating) activity; C:cytosol; P:pentose-phosphate shunt, oxidative branch; P:D-gluconate catabolic process; F:NADP binding
Phvul.008G023800.1	basic leucine zipper 43-like	F:GO:0003700; P:GO:0006355	F:DNA-binding transcription factor activity; P:regulation of transcription, DNA-templated
Phvul.008G023900.1	pentatricopeptide repeat-containing protein At3g02330, mitochondrial	F:GO:0008270; C:GO:0016021	F:zinc ion binding; C:integral component of membrane
Phvul.008G024000.2	WPP domain-interacting tail-anchored protein 1	C:GO:0016021	C:integral component of membrane
Phvul.008G024000.1	WPP domain-interacting tail-anchored protein 1	C:GO:0016021	C:integral component of membrane
Phvul.008G024100.1	probable sodium-coupled neutral amino acid transporter 6	P:GO:0003333; F:GO:0015171; C:GO:0016021	P:amino acid transmembrane transport; F:amino acid transmembrane transporter activity; C:integral component of membrane
Phvul.008G024200.1	putative UPF0481 protein At3g02645	C:GO:0016021	C:integral component of membrane
Phvul.008G024300.1	pentatricopeptide repeat-containing protein At3g29290 isoform X1		
Phvul.008G024300.2	pentatricopeptide repeat-containing protein At3g29290 isoform X2		
Phvul.008G024400.1	pentatricopeptide repeat-containing protein At2g31400, chloroplastic	C:GO:0009507	C:chloroplast
Phvul.008G024500.1	kanadaplin	F:GO:0003729	F:mRNA binding
Phvul.008G243800.1	ABC transporter B family member 25, mitochondrial	F:GO:0005524; C:GO:0005743; P:GO:0006879; C:GO:0016021; F:GO:0016887; F:GO:0042626; P:GO:0055085	F:ATP binding; C:mitochondrial inner membrane; P:cellular iron ion homeostasis; C:integral component of membrane; F:ATPase activity; F:ATPase-coupled transmembrane transporter activity; P:transmembrane transport
Phvul.008G243900.1	hypothetical protein PHAVU_008G243900g	C:GO:0016021	C:integral component of membrane
Phvul.008G244000.1	KH domain-containing protein At3g08620-like isoform X2	P:GO:0000381; F:GO:0003729; C:GO:0005634 C:GO:0005743; F:GO:0008121; P:GO:0022900; C:GO:0070469	P:regulation of alternative mRNA splicing, via spliceosome; F:mRNA binding; C:nucleus
Phvul.008G244100.1	cytochrome b-c1 complex subunit 8	C:GO:0005743; F:GO:0008121; P:GO:0022900; C:GO:0070469	C:mitochondrial inner membrane; F:ubiquinol-cytochrome-c reductase activity; P:electron transport chain; C:respirasome
Phvul.008G244100.2	cytochrome b-c1 complex subunit 8	C:GO:0005743; F:GO:0008121; P:GO:0022900; C:GO:0070469	C:mitochondrial inner membrane; F:ubiquinol-cytochrome-c reductase activity; P:electron transport chain; C:respirasome
Phvul.008G244200.1	metal tolerance protein 4-like	F:GO:0005384; C:GO:0016021; P:GO:0071421	F:manganese ion transmembrane transporter activity; C:integral component of membrane; P:manganese ion transmembrane transport
Phvul.008G244300.1	putative B3 domain-containing protein At5g58280	F:GO:0003677; C:GO:0005634	F:DNA binding; C:nucleus
Phvul.008G244400.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G244400.2	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G244500.1	probable disease resistance protein At4g27220	C:GO:0016020; F:GO:0043531	C:membrane; F:ADP binding

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.008G244600.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X2	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G244700.1	disease resistance protein At4g27190-like	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G244800_1.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X2	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G244800_0.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G244900.1	probable disease resistance protein At4g27220	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G245000.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G245100.1	probable disease resistance protein At4g27220	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G245200.1	probable disease resistance protein At4g27220	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G245300.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G245400_0.1	disease resistance protein At4g27190-like	C:GO:0016020; F:GO:0043531	C:membrane; F:ADP binding
Phvul.008G245500.1	hypothetical protein PHAVU_008G245500g		
Phvul.008G245600.2	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G245600.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G245700.1	root phototropism protein 2-like	P:GO:0009638	P:phototropism
Phvul.008G245800.1	zinc finger CCHC domain-containing protein 10-like	F:GO:0003676; F:GO:0008270	F:nucleic acid binding; F:zinc ion binding
Phvul.008G245800.2	zinc finger CCHC domain-containing protein 10-like	F:GO:0003676; F:GO:0008270	F:nucleic acid binding; F:zinc ion binding
Phvul.008G245900.2	probable inactive receptor kinase At5g58300	F:GO:0004672; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G245900.1	probable inactive receptor kinase At5g58300	F:GO:0004672; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G246000.1	pathogenesis-related genes transcriptional activator PTI6-like	F:GO:0003677; F:GO:0003700; C:GO:0005634; P:GO:0006355	F:DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated
Phvul.008G246000.2	pathogenesis-related genes transcriptional activator PTI6-like	F:GO:0003677; F:GO:0003700; C:GO:0005634; P:GO:0006355	F:DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated
Phvul.008G246100.1	eukaryotic translation initiation factor 2 subunit beta	P:GO:0001731; P:GO:0001732; F:GO:0003729; F:GO:0003743; C:GO:0005850; C:GO:0005886; F:GO:0031369	P:formation of translation preinitiation complex; P:formation of cytoplasmic translation initiation complex; F:mRNA binding; F:translation initiation factor activity; C:eukaryotic translation initiation factor 2 complex; C:plasma membrane; F:translation initiation factor binding
Phvul.008G246200.1	hypothetical protein PHAVU_008G246200g	C:GO:0016020; C:GO:0016021	C:membrane; C:integral component of membrane
Phvul.008G246300.1	protein NETWORKED 4A-like	F:GO:0003779; C:GO:0005774; F:GO:0016301; P:GO:0016310	F:actin binding; C:vacuolar membrane; F:kinase activity; P:phosphorylation

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.008G246400.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0000166; F:GO:0004672; C:GO:0016020; P:GO:0016310	F:nucleotide binding; F:protein kinase activity; C:membrane; P:phosphorylation
Phvul.008G246500.1	probable disease resistance protein At4g27220	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G246600.1	hypothetical protein PHAVU_008G246600g, partial	C:GO:0009536	C:plastid
Phvul.008G246700.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G246800_1.1	disease resistance protein At4g27190-like	C:GO:0016020	C:membrane
Phvul.008G246800_0.1	disease resistance protein At4g27190-like	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G246900.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X1	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.008G247000.1	probable disease resistance protein At4g27220	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G247000.2	probable disease resistance protein At4g27220	C:GO:0016021; F:GO:0043531	C:integral component of membrane; F:ADP binding
Phvul.008G247100.1	probable LRR receptor-like serine/threonine-protein kinase At1g06840 isoform X2	F:GO:0004672; F:GO:0005524; P:GO:0006468; C:GO:0016020	F:protein kinase activity; F:ATP binding; P:protein phosphorylation; C:membrane
Phvul.008G247200.1	probable disease resistance protein At4g27220	C:GO:0016020; F:GO:0043531	C:membrane; F:ADP binding
Phvul.008G247300.1	probable disease resistance protein At4g27220	F:GO:0043531	F:ADP binding
Phvul.011G160400.1	protein CUP-SHAPED COTYLEDON 2-like	F:GO:0000976; C:GO:0005634; P:GO:0006355; P:GO:0010072; P:GO:0010223; P:GO:0048366; P:GO:0090691; P:GO:0090709	F:transcription regulatory region sequence-specific DNA binding; C:nucleus; P:regulation of transcription, DNA-templated; P:primary shoot apical meristem specification; P:secondary shoot formation; P:leaf development; P:formation of plant organ boundary; P:regulation of timing of plant organ formation
Phvul.011G160500.1	protein SEEDLING PLASTID DEVELOPMENT 1	F:GO:0005524; C:GO:0009507; F:GO:0016787	F:ATP binding; C:chloroplast; F:hydrolase activity
Phvul.011G160500.2	hypothetical protein PHAVU_011G160500g	F:GO:0005524; C:GO:0009507; F:GO:0016787	F:ATP binding; C:chloroplast; F:hydrolase activity
Phvul.011G160600.1	CLAVATA3/ESR (CLE)-related protein 5-like		
Phvul.011G160700.1	CLAVATA3/ESR (CLE)-related protein 5-like		
Phvul.011G160800.1	putative threonine aspartase	F:GO:0004298; C:GO:0005737; P:GO:0006508; P:GO:0051604	F:threonine-type endopeptidase activity; C:cytoplasm; P:proteolysis; P:protein maturation
Phvul.011G160800.2	putative threonine aspartase	F:GO:0004298; C:GO:0005737; P:GO:0006508; P:GO:0051604	F:threonine-type endopeptidase activity; C:cytoplasm; P:proteolysis; P:protein maturation
Phvul.011G160900.1	3-hydroxybutyryl-CoA dehydrogenase	F:GO:0003857; C:GO:0005777; P:GO:0006631; P:GO:0055114; F:GO:0070403	F:3-hydroxyacyl-CoA dehydrogenase activity; C:peroxisome; P:fatty acid metabolic process; P:oxidation-reduction process; F:NAD+ binding
Phvul.011G161000.1	phytosulfokine receptor 2-like	F:GO:0004674; F:GO:0005524; P:GO:0006468; C:GO:0016021	F:protein serine/threonine kinase activity; F:ATP binding; P:protein phosphorylation; C:integral component of membrane
Phvul.011G161100.1	11-oxo-beta-amyrin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.011G161100.2	11-oxo-beta-amyirin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G161200.1	11-oxo-beta-amyirin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G161200.2	11-oxo-beta-amyirin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G161300_0.1	hypothetical protein PHAVU_011G1613000g, partial		
Phvul.011G161400.1	11-oxo-beta-amyirin 30-oxidase	F:GO:0005506; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:iron ion binding; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G161500.1	11-oxo-beta-amyirin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G161600.1	11-oxo-beta-amyirin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G161700.1	11-oxo-beta-amyirin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G161800.1	hypothetical protein PHAVU_011G161800g	F:GO:0003677	F:DNA binding
Phvul.011G161900.1	11-oxo-beta-amyirin 30-oxidase-like	F:GO:0004497; F:GO:0005506; C:GO:0016021; F:GO:0016705; F:GO:0020037; P:GO:0055114	F:monooxygenase activity; F:iron ion binding; C:integral component of membrane; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; P:oxidation-reduction process
Phvul.011G162000.1	ATP-dependent zinc metalloprotease FTSH 10, mitochondrial isoform X1	F:GO:0004222; F:GO:0005524; C:GO:0005743; P:GO:0008053; F:GO:0008270; C:GO:0016021; P:GO:0034982; P:GO:0042407; P:GO:0065003	F:metalloendopeptidase activity; F:ATP binding; C:mitochondrial inner membrane; P:mitochondrial fusion; F:zinc ion binding; C:integral component of membrane; P:mitochondrial protein processing; P:cristae formation; P:protein-containing complex assembly
Phvul.011G162100.1	peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	C:GO:0005773	C:vacuole
Phvul.011G162200.1	rab3 GTPase-activating protein non-catalytic subunit	P:GO:0043087	P:regulation of GTPase activity
Phvul.011G162300.1	polynucleotide 3'-phosphatase ZDP isoform X1	F:GO:0003677; F:GO:0008270	F:DNA binding; F:zinc ion binding
Phvul.011G162400.1			
Phvul.011G162500.1			
Phvul.011G162600.1	GEM-like protein 2 isoform X1		
Phvul.011G162700.1	protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 10	C:GO:0005634; P:GO:0009299; P:GO:0009416	C:nucleus; P:mRNA transcription; P:response to light stimulus

Gene ID	Functional Annotation	Gene Ontology ID	Gene Ontology Terms
Phvul.011G162800.1	30S ribosomal protein S20, chloroplastic	F:GO:0003735; P:GO:0006412; C:GO:0009570; C:GO:0009941; C:GO:0015935; F:GO:0070181	F:structural constituent of ribosome; P:translation; C:chloroplast stroma; C:chloroplast envelope; C:small ribosomal subunit; F:small ribosomal subunit rRNA binding
Phvul.011G162900.1	ethylene-responsive transcription factor 4	F:GO:0003677; F:GO:0003700; C:GO:0005634; P:GO:0006355	F:DNA binding; F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA- templated
Phvul.011G163000.1	---NA---		
Phvul.011G163100.1	ribosomal protein L11 methyltransferase	C:GO:0005840; P:GO:0006479; F:GO:0008276	C:ribosome; P:protein methylation; F:protein methyltransferase activity

APPENDIX L. RPKM values obtained from the Common Bean Expression Atlas (O'Rourke et al. 2014), for 13 common bean candidate genes for regulating common bean seed shape and size in eight tissues related to seed development: young flowers (FY), young pods (PY), pods approximately 9 cm long (PH), pods between 10 and 11 cm long (P1), pods between 12 and 13 cm long (P2), heart stage seeds (SH), stage 1 seeds (S1), and stage 2 seeds (S2)

Gene ID	PVFY	PVPY	PVPH	PVP1	PVP2	PVSH	PVS1	PVS2
Phvul.002G004400	49	16	40	15	31	122	40	11
Phvul.002G006100	599	185	696	111	213	2029	1575	1496
Phvul.002G006700	350	143	316	79	133	825	439	312
Phvul.002G007100	0	0	0	0	0	3	69	53
Phvul.008G019500	132	39	279	109	73	488	101	60
Phvul.008G020600	2163	719	2105	594	1205	3439	735	503
Phvul.008G022600	489	184	644	188	323	1096	773	467
Phvul.008G023900	34	12	13	5	11	47	15	10
Phvul.008G024300	81	45	108	43	77	91	50	37
Phvul.008G244300	45	20	12	4	2	88	9	11
Phvul.008G024400	612	213	803	270	498	834	609	510
Phvul.011G163100	22	11	19	4	7	59	48	15
Phvul.011G160400	205	91	275	44	130	132	84	127

APPENDIX M. Z-scores calculated from the absolute expressions of 13 common bean candidate genes for regulating common bean seed shape and size in eight tissues related to seed development according to the Common Bean Expression Atlas (O'Rourke et al. 2014). Young flowers (a; FY), pods between 10 and 11 cm long (b; P1), pods between 12 and 13 cm long (c; P2), pods approximately 9 cm long (d; PH), young pods (e; PY), stage 1 seeds (f; S1), stage 2 seeds (g; S2) and heart stage seeds (h; SH)