

**Symbiotic Nitrogen Fixation in Heirloom, Landrace, and Modern  
Participatory-Bred and Conventionally-Bred Common Bean  
(*Phaseolus vulgaris* L.)**

by

Jennifer Leanne Wilker

A Thesis

presented to

The University of Guelph

In partial fulfilment of requirements

for the degree of

Doctor of Philosophy

in

Plant Agriculture

Guelph, Ontario, Canada

© Jennifer Leanne Wilker, February, 2021

## ABSTRACT

### SYMBIOTIC NITROGEN FIXATION IN HEIRLOOM, LANDRACE, AND MODERN PARTICIPATORY-BRED AND CONVENTIONALLY-BRED COMMON BEAN (*PHASEOLUS VULGARIS* L.)

Jennifer Leanne Wilker  
University of Guelph, 2021

Advisor(s):  
K. Peter Pauls  
Alireza Navabi (deceased March, 2019)

Among legumes, common bean (*Phaseolus vulgaris* L.) is generally thought to have poor capacity for symbiotic nitrogen fixation (SNF). To reduce the need for expensive and environmentally non-sustainable fertilizer amendments to achieve high yields, bean growers worldwide would benefit from the development of bean varieties with stable, yield-sustaining nitrogen fixing capacity. In this thesis, 400 common bean genotypes were tested in inoculated, low-nitrogen field conditions to observe the diversity for SNF-related and agronomic traits. SNF was quantified using the natural abundance method in which isotope values were calculated from bean seeds, and genotypes that fixed large proportions of N were considered to have superior SNF capacity. In a panel of heirloom and conventional varieties it was confirmed that SNF capacity is greater in Middle American genotypes than in Andean genotypes, and that heirloom bean varieties have a wider range of SNF capacities than conventional varieties. In collaboration with an NGO in Honduras, a panel of Honduran bean varieties was curated, which included landraces, conventional varieties, and varieties developed through participatory plant breeding (PPB). It was found that genetic diversity was greater among Honduran landraces, and that these genotypes had higher SNF

capacities than the other breeding categories. Candidate genes associated with SNF were identified in regions of the genome where nucleotide diversity was higher in the landraces and where population genetic differentiation was evident between the landraces and the PPB varieties. With the Middle American diversity panel (MDP) assembled by the Bean Coordinated Agricultural Project (BeanCAP), a genome wide association study (GWAS) was carried out to identify quantitative trait loci (QTL) associated with SNF traits in the entire panel and within each panel subpopulation. The largest number of QTL were found for seed nitrogen content (%N), followed by percent nitrogen derived from the atmosphere (%Ndfa) and leaf chlorophyll content. For most traits, QTL identified for the entire panel were not repeated in the panel subpopulations, nor were QTL shared between the subpopulations. Additionally, the QTL identified in the MDP are novel and do not co-localize to regions of the genome where SNF-related QTL were reported previously. Overall, it can be concluded that diversity for SNF in common bean is wide-ranging, and is found in germplasm as diverse as landraces, heirloom varieties, and varieties bred through conventional and participatory methods. This diversity can be employed in breeding efforts to generate varieties with improved nitrogen fixation for growers worldwide.

## **DEDICATION**

For my dad, who instilled in me his love of agriculture  
and who would have been so proud to see me achieve my PhD.

## **ACKNOWLEDGEMENTS**

I would like to offer heart-felt gratitude to my supervisor, Dr. K. Peter Pauls, for his support and supervision of this work. I also wish to extend thanks to my co-supervisor, Dr. Alireza Navabi, who helped me envision the studies presented here and supported me through the first years of this endeavor. The support and contributions provided by the members of my advisory committee were also invaluable – thank you, Dr. Sally Humphries, Dr. Timothy Porch, and Dr. Istvan Rajcan.

Field work for this thesis was carefully managed by three crews, one in Ontario, one in Honduras, and one in Puerto Rico. For the Ontario trials, I wish to gratefully acknowledge the technical assistance of Tom Smith and Lyndsay Schram. Their solid field management skills and joyful approach to the work were so appreciated. Thanks also to the many summer students who helped with my Ontario field trials. Sincere gratitude to Dr. Sally Humphries who helped facilitate the field study in Yorito, Honduras and to Marvin Gomez and Fredy Sierra who managed the Honduran trial. Thanks to Dr. Juan Carlos Rosas for his input and his support of the Honduran study. Much thanks to Dr. Timothy Porch who facilitated the field study at Isabela, Puerto Rico along with his dedicated field team.

Thanks to Chris Grainger for his help with the DNA work; to Dr. Michelle Edwards for her assistance with the statistical analyses; and to Dr. Davoud Torkamaneh for his assistance with the genomic analyses. Your instruction gave me the confidence to tackle more than I thought possible. Thanks to Lyndsay Schram for her help with

grinding seed samples and weighing seed powder – easily the most tedious aspect of this research. Gratitude to everyone in the Pauls lab, especially my cohort of grad students, who shared in the joys and the frustrations all along this journey.

Without the financial support of many agencies the work presented here would not have been possible. I gratefully acknowledge the support of: the Ontario Bean Growers, Agriculture Agri-Food Canada (Cluster Program), the Ontario Ministry for Food and Rural Affairs, the Ontario Ministry for Research and Innovation, and the Food from Thought Program (Canada First Research Excellence Fund). Additional support for the Honduran bean study was provided by SeedChange/USC Canada, Global Affairs Canada, International Development Research Centre, World Accord, Development Fund (Norway), and USAID.

I am also very grateful for the scholarships I received during the course of this PhD and wish to thank the donors, including: funders of the Dr. O. M. McConkey, the Schneller and Summers, and the Bullick scholarships; the Arrell Food Institute (Food from Thought) and the Kasha Scientific travel grants for supporting travel to conferences to present my research; and finally the Bean Improvement Cooperative and the Canadian Pulse Research Workshop for multiple presentation awards.

Lastly, I have so much love and gratitude for my family. They cheered me on, kept me motivated, and reminded me of the importance of the work-life balance. I would not have attempted this feat, nor could I have achieved this, without them.

## TABLE OF CONTENTS

Abstract .....	ii
Dedication .....	iv
Acknowledgements .....	v
Table of Contents .....	vii
List of Tables .....	xiii
List of Figures.....	xv
List of Abbreviations and Symbols .....	xvii
List of Appendices .....	xix
1 General Introduction.....	1
1.1 Introduction .....	2
1.2 Hypothesis .....	5
1.3 Research Objectives .....	5
2 Literature Review .....	6
2.1 Introduction .....	7
2.2 Soil-acquired nitrogen .....	7
2.3 Symbiotic nitrogen fixation .....	10
2.4 Common bean.....	11
2.5 Importance of genetic diversity.....	13
2.5.1 Germplasm developed through informal means .....	13
2.5.2 Germplasm developed through conventional plant breeding.....	16
2.5.3 Quantitative genetics for SNF characterization.....	20

2.6	Recent studies on SNF in beans.....	23
2.7	Measuring SNF .....	27
3	Agronomic Performance and Nitrogen Fixation of Heirloom and Conventional Dry Bean Varieties Under Low-Nitrogen Field Conditions .....	32
3.1	Abstract.....	33
3.2	Introduction .....	34
3.3	Materials and Methods.....	39
3.3.1	Plant material.....	39
3.3.2	Field experimental design and maintenance .....	43
3.3.3	Phenotyping.....	45
3.3.4	Seed isotope analysis.....	47
3.3.5	Genotyping .....	49
3.3.6	Identity by state analysis.....	50
3.3.7	Nucleotide diversity analysis.....	50
3.3.8	Statistical analysis .....	51
3.4	Results .....	52
3.4.1	Origins and phenotypic characteristics of selected beans .....	52
3.4.2	Field conditions.....	53
3.4.3	Genetic analysis of relatedness.....	54
3.4.4	Nucleotide diversity among genotype categories .....	58
3.4.5	Diversity for seed isotope traits.....	59
3.4.6	Diversity for agronomic traits .....	62
3.4.7	Diversity for leaf chlorophyll content .....	66
3.4.8	Nitrogen fixation in the HCP .....	67

3.4.9	Trait correlation.....	70
3.5	Discussion.....	74
3.5.1	Genetic diversity is greater in the Middle American than in the Andean genepool .....	74
3.5.2	Nitrogen fixation capacity in the Middle American genepool exceeds that found in the Andean genepool .....	76
3.5.3	Diversity for Nitrogen fixation in conventional bean genotypes similar to other studies .....	77
3.5.4	Modern breeding has not reduced SNF capacity.....	78
3.5.5	Incorporating heirloom genotypes into breeding for improved SNF holds potential .....	79
3.5.6	Heirloom beans may be particularly suited to breeding for organic agriculture .....	82
3.6	Conclusion .....	83
4	Genetic Diversity, Nitrogen Fixation, and Water Use Efficiency in a Panel of Honduran Common Bean ( <i>Phaseolus vulgaris</i> L.) Landraces and Modern Genotypes	85
4.1	Abstract.....	86
4.2	Introduction .....	87
4.3	Materials and Methods.....	94
4.3.1	Plant material.....	94
4.3.2	Field experimental design and maintenance .....	107
4.3.3	Phenotyping.....	110
4.3.4	Isotope analysis.....	111
4.3.5	Genotyping .....	113
4.3.6	Population structure.....	113
4.3.7	Genetic diversity .....	114

4.3.8	Candidate gene investigation .....	115
4.3.9	Statistical analysis .....	115
4.4	Results .....	117
4.4.1	Analysis of genetic relatedness .....	117
4.4.2	Nucleotide diversity and population differentiation: Landrace and PPB categories .....	121
4.4.3	Identification of candidate genes .....	126
4.4.4	Diversity for symbiotic nitrogen fixation.....	129
4.4.5	Diversity for agronomic traits .....	133
4.4.6	Trait correlation and genotype by trait biplot analyses.....	136
4.5	Discussion.....	142
4.5.1	Genotype origins and pedigree explain Honduran panel structure .....	142
4.5.2	Optimizing use of genetic diversity of Honduran landraces and PPB varieties.....	147
4.5.3	Regions of high genetic differentiation indicate regions impacted by selection.....	149
4.5.4	Landraces are superior nitrogen fixers .....	151
4.5.5	PPB genotypes have superior water use efficiency values.....	154
4.5.6	Conventional genotypes have superior yields .....	156
4.5.7	Utility of panel genotypes for breeding .....	157
4.6	Conclusion .....	159
5	Symbiotic Nitrogen Fixation and related QTL in a <i>Phaseolus vulgaris</i> (L.) Middle American Diversity Panel .....	161
5.1	Abstract.....	162
5.2	Introduction .....	163

5.3	Materials and Methods.....	167
5.3.1	Germplasm .....	167
5.3.2	Field conditions.....	171
5.3.3	Experimental design .....	172
5.3.4	Rhizobia Inoculation .....	172
5.3.5	Field trial maintenance.....	173
5.3.6	Phenotyping at Elora and Belwood, Ontario .....	174
5.3.7	Phenotyping at Isabela, Puerto Rico .....	174
5.3.8	Seed isotope analysis.....	175
5.3.9	Statistical analysis .....	175
5.3.10	Genotyping and SNP dataset development .....	176
5.3.11	Population structure .....	177
5.3.12	Genome-wide association study.....	177
5.3.13	Candidate gene identification .....	178
5.4	Results .....	178
5.4.1	Field conditions.....	178
5.4.2	Population structure.....	179
5.4.3	Trait diversity .....	182
5.4.4	Trait correlation.....	194
5.4.5	Marker-trait associations.....	198
5.4.6	SNP haplotype effects and candidate genes .....	213
5.5	Discussion.....	220
5.5.1	Marker-trait associations.....	222

5.5.2	QTL for SNF traits: MDP vs. ADP .....	225
5.5.3	Candidate genes associated with significant SNPs .....	227
5.6	Conclusion .....	233
6	General Conclusion.....	235
	References.....	239
	Appendix A.....	267
	Appendix B.....	275
	Appendix C.....	320

## LIST OF TABLES

Table 3.1 Market class, seed size, growth habit, gene pool and race for 42 dry bean ( <i>Phaseolus vulgaris</i> L.) genotypes of the heirloom-conventional panel.....	42
Table 3.2. Nitrogen derived from the atmosphere (%Ndfa) and differential ranking of common bean genotypes at three locations (Elora, Belwood) and in two seasons (2014, 2015).....	68
Table 4.1. Genotypes of the North American check and Honduran conventional breeding categories which were included in the HON panel tested at 3 field locations, 2014-15.....	98
Table 4.2. Genotypes of the Landrace breeding category which were included in the HON panel tested at 3 field locations, 2014-15. ....	101
Table 4.3. Genotypes of PPB breeding category which were included in the HON panel tested at 3 field locations, 2014-15.....	103
Table 4.4. Regions of the <i>P. vulgaris</i> (2.0) genome where high nucleotide diversity ( $\pi$ ) was discovered in landrace genotypes compared to PPB genotypes. ....	122
Table 4.5. Regions of the <i>P. vulgaris</i> (2.0) genome where SNPs with significantly high weighted $F_{ST}$ values ( $> 0.5$ ) were found. ....	125
Table 4.6. F-test of fixed effect breeding category in the GLIMMIX analysis, and the breeding category LSmeans comparisons of genotypes in the HON panel grown at Elora, 2014.....	130
Table 4.7. F-test of fixed effect breeding category in the GLIMMIX analysis, and the breeding category LSmeans comparisons of genotypes in the HON panel grown at Elora, 2015.....	131
Table 4.8. F-test of fixed effect breeding category in the GLIMMIX analysis, and the breeding category LSmeans comparisons of genotypes in the HON panel grown at Yorito, 2014-2015.....	132
Table 5.1. Middle American Diversity Panel (MDP) genotypes (n = 98) belonging to race Mesoamerica including navy, black and carioca market classes. ....	168
Table 5.2. Middle American Diversity Panel (MDP) genotypes (n = 182) belonging to race Durango-Jalisco including great northern, pinto, small red, and pink market classes. ....	169

Table 5.3. The top 20 MDP genotypes ranked by nitrogen fixation capacity (%Ndfa) at the two Ontario field trial locations (Elora, 2014 and Belwood, 2015). .....	186
Table 5.4. The worst 20 MDP genotypes ranked by nitrogen fixation capacity (%Ndfa) at the two Ontario field trial locations (Elora, 2014 and Belwood, 2015). .....	188
Table 5.5. The best 10 genotypes for nitrogen fixation capacity according to market class within the Durango-Jalisco subpopulation of the MDP. ....	189
Table 5.6. The best 10 genotypes for nitrogen fixation capacity according to market class within the Mesoamerican subpopulation of the MDP. ....	190
Table 5.7. SNPs marking QTL significantly associated with Nitrogen fixation traits in the entire MDP, by location (5% FDR). .....	200
Table 5.8. QTL significantly associated with nitrogen fixation traits in the MA subpopulation of the MDP, by location (5% FDR). .....	203
Table 5.9. QTL significantly associated with nitrogen fixation traits in the DJ subpopulation of the MDP, by location (5% FDR). .....	211

## LIST OF FIGURES

Figure 3.1. Images of <i>Phaseolus vulgaris</i> genotypes included in the heirloom-conventional panel. ....	40
Figure 3.2. Analysis of genetic structure and relatedness of thirty-nine genotypes of the heirloom-conventional panel. ....	55
Figure 3.3. Means for seed composition traits measured from seed harvested at three field locations from genotypes of the heirloom-conventional panel. ....	60
Figure 3.4. Means for agronomic traits measured at 3 field locations for the genotypes of the heirloom-conventional panel.....	63
Figure 3.5. Biplot analysis of traits for genotypes of the heirloom-conventional panel in 3 location-years.....	72
Figure 4.1. Map of west–central Honduras, outlining the six departments from which landrace and participatory bred (PPB) bean genotypes were sourced for the Honduran Panel. ....	97
Figure 4.2. Analysis of genetic structure and relatedness of 72 genotypes of the Honduran panel.....	118
Figure 4.3. Nucleotide diversity and population differentiation. ....	124
Figure 4.4. Biplot analysis of traits for genotypes of the Honduran panel in three location years. ....	137
Figure 4.5. Genotype performance for nitrogen fixation (x axis) and yield (y axis) at Yorito.....	141
Figure 5.1. STRUCTURE analysis of the Middle American diversity panel (MDP) from $K = 2$ to $K = 7$ .....	180
Figure 5.2. Phylogenetic tree of 280 genotypes of the Middle American Diversity Panel inferred using the neighbor-joining method in MEGA X.....	181
Figure 5.3. Data distributions of LSmeans for seed composition SNF-related traits in the MDP, and each subpopulation (DJ and MA) at field trial locations in Elora (2014), Belwood (2015), and Isabela (2016). ....	184
Figure 5.4. Data distributions of LSmeans for physiological and morphological SNF-related traits in the MDP, and each subpopulation (DJ and MA) at field trial locations in Elora (2014), Belwood (2015), and Isabela (2016).....	192

Figure 5.5. Significant Pearson correlations ( $p > 0.05$ ) between nitrogen fixation-related traits and locations based on LSmeans for the entire MDP (A), the Mesoamerican subpopulation (B), and the Durango-Jalisco subpopulation (C). .....	195
Figure 5.6. Genotype by trait bi-plots of principle components 1 and 2 for Elora2014 (A), Belwood 2015 (B), and Isabela 2016 (C). .....	197
Figure 5.7. Manhattan QQ plots for GWAS of 280 MDP genotypes and $\delta^{15}\text{N}$ values. ....	199
Figure 5.8. Manhattan QQ plots for GWAS of 82 race MA MDP genotypes and seed nitrogen discrimination ( $\delta^{15}\text{N}$ ) values. ....	202
Figure 5.9. Manhattan QQ plots for GWAS of 280 MDP genotypes and %Ndfa values. ....	204
Figure 5.10. Manhattan QQ plots for GWAS of 98 MDP genotypes belonging to the MA subpopulation and %Ndfa values. ....	206
Figure 5.11. Manhattan QQ plots for GWAS of 280 MDP genotypes and seed nitrogen content (%N). ....	208
Figure 5.12. Manhattan QQ plots for GWAS of 182 MDP genotypes belonging to the DJ subpopulation and %N values. ....	210
Figure 5.13. Manhattan QQ plots for GWAS of 280 MDP genotypes and leaf chlorophyll content at flowering values. ....	212
Figure 5.14. Manhattan QQ plots for GWAS of 98 MDP genotypes belonging to the MA subpopulation and leaf chlorophyll content at flowering values. ....	214

## LIST OF ABBREVIATIONS AND SYMBOLS

AAFC Agriculture Agri-Food Canada

CIAT International Center for Tropical Agriculture

CDC Crop Development Center

$\delta^{13}\text{C}$  carbon discrimination

$\Delta\text{C}$  carbon difference

$\delta^{15}\text{N}$  nitrogen discrimination

$\Delta\text{N}$  nitrogen difference

DTF days to flowering

DTM days to maturity

FIPAH Fundación para la Investigación Participativa con Agricultores de Honduras

GCMS gas chromatography mass spectrometry

HSW hundred seed weight

IBS identity by state

MSU Michigan State University

MYA million years ago

NDSU North Dakota State University

OMAFRA Ontario Ministry of Agriculture, Food and Rural Affairs

PIF Programa Investigación de Frijol

PPB participatory plant breeding

PVS participatory varietal selection

%N percent nitrogen

%Ndfa percent nitrogen derived from the atmosphere

SNF symbiotic nitrogen fixation

UG University of Guelph

USDA-ARS United States Department of Agriculture – Agriculture Research Service

WUE water use efficiency

## **LIST OF APPENDICES**

Appendix A. Supplementary tables associated with Chapter 3.

Appendix B. Supplementary tables and figures associated with Chapter 4.

Appendix C. Supplementary tables and figures associated with Chapter 5.

# 1 General Introduction

## 1.1 Introduction

Two looming crises are facing the world today: the hunger crisis and the climate crisis. And these crises are inextricably connected. Food insecurity and malnutrition are multifaceted problems which cannot be solved without addressing climate change. Common bean (*Phaseolus vulgaris* L.) is the world's most important grain legume for human consumption (Broughton et al., 2003). This, combined with its capacity for symbiotic nitrogen fixation, makes common bean a key crop which will contribute to attaining satiety and nutrition goals in an environmentally sustainable way.

Common bean is a member of the Fabaceae and is among the approximately 80 wild and domesticated species of the *Phaseolus* genus. Other *Phaseolus* species which are cultivated for human consumption are year bean (*P. dumosus*), scarlet runner bean (*P. coccineus*), tepary bean (*P. acutifolius*) and lima bean (*P. lunatus*) (Bellucci et al., 2014). All *Phaseolus* species are diploid and common bean has 11 chromosomes ( $2n = 2x = 22$ ) and a small genome at 587 Mbp (Schmutz et al., 2014). Genomic discoveries made in common bean also have relevance to closely related species including soybean (*Glycine max* L. Merr.), which shares a common ancestor approximately 19 MYA (McClellan et al., 2010).

The center of origin of common bean is in Mesoamerica and it has two centers of domestication, the Oaxaca valley of present-day Mexico and a region spanning the border between present-day Bolivia and Argentina (Bitocchi et al., 2013). Beans domesticated in these regions belong to the Middle American gene pool and the Andean

genepool, respectively. Bean domestication began approximately 8,000 years ago (Mamidi et al., 2011). A number of agronomic characteristics have been improved in bean through domestication and breeding including, indeterminate and bush growth types with reduced vining, increased seed size, and loss of pod shattering and seed dormancy. A characteristic of members of the legume family is their ability to form symbiotic relationships with nitrogen fixing bacteria. Common bean has retained the capacity for symbiotic nitrogen fixation (SNF) throughout its domestication.

Common bean is consumed worldwide, primarily as dried seeds or fresh (unripe) pods, and in some locations the leaves are also used in cuisine. Beans contribute to balanced nutrition as an excellent source of dietary protein (Baudoin and Maquet, 1999), fibre (Kutoš et al., 2003), folate and essential micronutrients (Duranti, 2006; Hefni et al., 2010). In addition, bean consumption improves gut microbiota and gut health (Monk et al., 2016).

Global common bean production in 2018 totaled  $55.2 \times 10^6$  tonnes grown on  $36.1 \times 10^6$  ha; dry beans make up the larger proportion of the total ( $30.4 \times 10^6$  tonnes,  $24.8 \times 10^6$  ha) with fresh (green) beans making up the remainder (FAO, 2018). Large-scale bean producers dominate production in the highest-producing countries (Brazil, Mexico, China, the USA), however small-scale and often subsistence production is common in Mexico, Central and South America, Africa and Asia. Dry bean production in Canada was 341 100 tonnes in 2018 produced primarily in Ontario, Manitoba, Saskatchewan and Alberta on 137 100 ha (FAO, 2018). In the coming decades

common bean will continue to be an important commodity worldwide and production demand will increase approximately 30% with population growth as 2050 approaches (Porch et al., 2013).

Dry bean production around the world is carried out under diverse production conditions including varying fertilizer regimes. Rhizobia inoculation, animal waste and synthetic nitrogen fertilizers are used, while synthetic fertilizers have been found to return the largest yield increases (Otieno et al., 2009). In Ontario, nitrogen fertilizer rates ranging from 10 to 100 kg ha<sup>-1</sup>, are recommended for economical dry bean yields (OMAFRA, 2009). Despite the capacity of common bean to obtain nitrogen through SNF, beans are routinely grown using synthetic fertilizer where it is accessible and affordable. Unfortunately, environmental pollution occurs through fertilizer runoff and volatilization, but the potential to reduce this occurrence exists for common bean production. The diversity for SNF capacity recently reported in common bean (Farid and Navabi, 2015; Kamfwa et al., 2015; Heilig et al., 2017b) and the indication that gains can be made in selecting for SNF under optimal conditions (Farid et al., 2017) suggests that bean breeders could improve SNF in the crop leading to a reduction in fertilizer use and to sustainable bean production in the future.

## **1.2 Hypothesis**

The hypotheses that were investigated in this thesis were that:

1. Agronomically useful genetic diversity for SNF is present in dry bean varieties and breeding lines;
2. Genomic regions contributing to SNF and related traits can be identified; and
3. Genes affecting variability for SNF capacity can be identified.

The overarching hypothesis of this thesis was that if genetic variation exists for SNF potential in common bean, then it should be feasible to improve the SNF potential of high yielding common bean genotypes.

## **1.3 Research Objectives**

The objectives of this research were to:

1. Compare and contrast the phenotypic diversity for SNF and related traits of modern, heirloom and landrace dry bean genotypes. (Chapter 3, Chapter 4)
2. Identify genomic regions associated with SNF through a GWAS of a panel composed of modern dry bean genotypes. (Chapter 5); and
3. Identify potential candidate gene(s) conditioning SNF in common bean. (Chapter 5)

## **2 Literature Review**

## 2.1 Introduction

Nitrogen ( $N_2$ ) is one of the most abundant elements on earth, constituting approximately 78% of the earth's atmosphere (Williams, 2020). Natural organic forms of nitrogen also exist in soils, including ammonia, nitrates and nitrites. Nitrogen is a critical limiting macronutrient required for plant metabolism; it is a component of chlorophyll, amino acids, nucleic acids, proteins and enzymes, including RuBisCO. Plant nitrogen availability impacts water-use efficiency, plant biomass accumulation, and seed protein content. World crop production in 2010 was estimated to require 174 Tg N yr<sup>-1</sup> (Zhang et al., 2015b). Plants require N to be in its reactive (Nr), or fixed, biologically-available form, and both non-biological and biological processes are responsible for this conversion. Atmospheric N-fixation through lightning storms and anthropogenic activities generate approximately 35 Tg N yr<sup>-1</sup>, and synthetic fertilizer production (Haber-Bosch process) generates 120 Tg N yr<sup>-1</sup> (Fowler et al., 2013). Biological nitrogen fixation (BNF), also known as symbiotic nitrogen fixation (SNF), comprises an estimated 140 Tg N yr<sup>-1</sup> in aquatic environments, and approximately 120 Tg N yr<sup>-1</sup> in terrestrial ecosystems where agricultural SNF accounts for half of this value (Fowler et al., 2013). The primary sources of N for crop production are soil-available N and that generated through SNF, which will be discussed in the following sections.

## 2.2 Soil-acquired nitrogen

Soil-acquired nitrogen is derived from organic or synthetic fertilizers and from the decomposition of organic matter. During the time of the Industrial Revolution, nitrogen

for agricultural production came primarily from nitrogen recycling through animal and green manures, and from the application of mined nitrogen from natural fossil sources (Sutton et al., 2013). These organic sources were insufficient to meet growing agricultural and industrial demands at the turn of the 20th century. This deficit drove the research that led to the discovery of a chemical process that combines atmospheric N<sub>2</sub> with H<sub>2</sub> from natural gas to form synthetic ammonia, known as the industrial Haber-Bosch process. Chemical fertilizer production expanded throughout the first half of the 1900s, surpassing the amount of organic nitrogen used in agriculture, and fueling the Green Revolution which began in the 1960s (Sutton et al., 2013). The greatest gains during the Green Revolution are attributed to the development and adoption of high-yielding plant varieties (Vance, 2001). Yield gains in developing countries from 1960 to 2000 in cereals, for example, were over 200% for wheat, over 150% for maize, and over 100% for rice (FAO, 2004). High-yielding plant varieties respond favorably to N fertilizer amendments and synthetic fertilizer use increased from 10 Tg in the 1960s to 88 Tg in 2000, and N demands are projected to reach 120 Tg by 2030-2040 (Vance, 2001). However, continued reliance on synthetic fertilizer poses challenges.

The Food and Agriculture Organization of the United Nations projects that food production will need to increase by 49% from current levels to meet the needs of the more than 9 billion projected world population of 2050 (FAO, 2019). If current production practices are sustained, world fertilizer use in the next four decades will continue to rise. Crop production may increase enough to keep in step with human population growth,

however, the consequences of environmental pollution connected with synthetic fertilizer use and dependence on the finite fossil fuel resources are not sustainable.

Nitrogen use efficiency (NUE) is a measure of agricultural N efficiency which estimates the amount of N harvested in the crop from an agricultural system versus the amount of N fertilizer applied. For most major crops the estimated NUE is less than 50% (Zhang et al., 2015b). For example, the NUE estimates of cereal crops wheat, maize, and rice are 0.42, 0.46 and 0.39 respectively, while soybean NUE estimates are much higher at 0.8 and fruits and vegetables are much lower at 0.14 (Zhang et al., 2015b). The nitrogen which is not taken up by the crop plant remains in the soil and is susceptible to loss. Apart from the economic losses of inefficient N use, environmental concern is also created when the loss of excess nitrogen through runoff and leaching pollutes water resources, and the loss through volatilization and denitrification pollutes the atmosphere. Crop production is the source of more than 60% of N pollution (Bodirsky et al., 2014).

Certain practices could be put in place that would lead to sustainable agricultural production and to a reduction in fertilizer use. One approach is to increase NUE in agroecosystems. This could include cultivating plants that are efficient nitrogen users or modifying production practices to create less fertilizer waste. By reducing the amount of N required by a crop and improving the remobilization of N to the economic yield within crop plants, NUE could be increased resulting in less N released to the environment (Zhang et al., 2015b). Conservation tillage practices such as direct seeding, mulch-

based cropping systems, and continuous cover cropping improve NUE through enhancing mycorrhizal N scavenging activity and reducing soil N loss (Verzeaux et al., 2017). Another solution is to modify cropping rotations to include plants that can fix nitrogen from the atmosphere and are less reliant on external sources of nitrogen.

### **2.3 Symbiotic nitrogen fixation**

The capacity to fix atmospheric nitrogen is a characteristic exhibited solely by bacteria and Archaea which possess the nitrogenase enzyme which fixes atmospheric nitrogen. Legumes and rhizobia have evolved symbiotic relationships where the bacteria inhabit symbiosis-initiated plant nodules on the roots (or stems) which provide an ideal environment for bacteroid multiplication and N fixation. The ammonia generated through SNF is converted to ureide nitrogen compounds which are easily transported through the plant and used in metabolic processes.

It has been estimated that the contribution of N to agroecosystems by grain legumes alone is about 21 Tg annually, representing at least 30% of all the N fixed in agroecosystems (Herridge et al., 2008). Soybeans (*Glycine max* (L.) Merr.) are the worlds' leading grain legume crop and are among the most efficient nitrogen fixers, deriving an average of 68% of their nitrogen from the atmosphere (Herridge et al., 2008). Planted to 93.4 Mha in 2005, soybeans were estimated to have fixed 16.44 Tg of N that year (Herridge et al., 2008). In contrast, common beans (*Phaseolus vulgaris* L.), which are the most important grain legume grown worldwide for human consumption,

were planted on 25.1 Mha in 2005 and were estimated to have fixed only 0.58 Tg N (Herridge et al., 2008).

Research has shown that common beans are less efficient than other grain legumes at SNF, deriving only about 40% of their nitrogen from the atmosphere (Herridge et al, 2008). In North American dry bean production, reliance on SNF to meet N demands of the crop is uncommon except in organic cropping systems. Instead, synthetic fertilizer is routinely applied to ensure high yields. For example, the Ontario Ministry of Food and Rural Affairs recommends that farmers apply 10-100 kg N ha<sup>-1</sup> before planting, depending on bronzing and disease pressure (OMAFRA, 2009). However, in less developed regions of the world, such as in Africa, Asia, and Central and South America, where dry beans are an important crop for farmers and an integral part of people's diets, access to synthetic fertilizers is limited by poor infrastructure, availability and affordability, and soils are often tired and nitrogen-depleted. Breeding dry bean varieties which are efficient at SNF would reduce the need for the use of synthetic fertilizers in developed regions while providing a much needed tool to improve yields in developing regions of the world.

## **2.4 Common bean**

The most important grain legume grown worldwide for human consumption is the common bean. According to the Food and Agriculture Organization (FAO), beans were grown on 34.5 Mha worldwide in 2018 with 30.4 M tonnes produced (FAO, 2018).

Beans are a dietary staple, which provide an affordable form of protein and micronutrients as well as a stable income to small-holder growers in developing regions.

Bean is a diploid species with a 587 Mbp genome separated into 11 chromosomes ( $2n = 2X = 22$ ) (Schmutz et al., 2014). Common bean belongs to the Fabaceae (or Leguminosae) plant family. The center of origin of common bean is in present-day central Mexico, and the species underwent two independent domestications in Mesoamerica and the Andes resulting in two genepools (Bitocchi et al., 2012). Divergence of *P. vulgaris* from the most closely related species (*P. coccineus*, *P. dumosus*, *P. acutifolius*, and *P. lunatus*) occurred relatively recently around 2-4 MYA and high levels of synteny still remain between the genomes of these cultivated species (Bitocchi et al., 2017). In some cases, successful inter-species crosses have been made among *Phaseolus* species to introduce favorable traits to *P. vulgaris* (Mejía-Jiménez et al., 1994; Gurusamy et al., 2007; Kelly, 2010; Singh et al., 2014). For example, a recently released navy bean from the University of Guelph, OAC Mist, has plant introductions from *P. acutifolius* and *P. coccinius* in its pedigree (Khanal et al., 2017a). The extent of genetic divergence between the Andean and Middle American genepools is not insignificant, and inter-genepool crosses are not always successful (Singh, 1995). The morphological diversity exhibited in common bean is large, with indeterminate and determinate growth habits, a range of seed sizes and shapes, seed coat colours, patterns, and brilliance, among other traits. The FAO recognizes 62 market classes, 15 of which are recognized internationally (Jones, 1999). The International Center for Tropical Agriculture (CIAT), which is responsible for bean

germplasm conservation and international research, has nearly 38,000 accessions in its gene bank (CIAT, 2019), and the Germplasm Resource Information Network (GRIN) contains has over 17,000 accessions in the US National Plant Germplasm System.

## **2.5 Importance of genetic diversity**

Crop plant diversity is essential for securing food and nutritional security, reducing poverty, adapting to climate change, reducing environmental degradation, and contributing to sustainable agriculture. Landraces, or traditional varieties, represent the majority of the genetic diversity of domesticated crop species, whereas modern varieties are considered less diverse (Villa et al., 2005). Preserving the genetic diversity within a plant species ensures its availability for future generations, and the characterization of diverse germplasm enables it to be used in future breeding efforts.

### **2.5.1 Germplasm developed through informal means**

Landraces are heterogeneous populations of cultivated plants which lack formal genetic improvement (Villa et al., 2005). Landraces are derived from wild species that became important to indigenous groups and have been around since the beginning of organized agriculture. Despite being genetically diverse, landraces have a distinct identity and are well-adapted to the agroecosystems where they are cultivated (Villa et al., 2005).

Dry bean cultivation in North America had its beginning with indigenous peoples who grew maize, squash and beans in an intercropping system known as the “Three Sisters”. This system shares similarities with the Milpa cropping system used throughout

Mexico and Central America to the present day (Ebel et al., 2017). The combination of these crops provides a complete diet of carbohydrates, proteins, lipids and vitamins. The diverse plant architectures of these crops ensure they occupy complementary rather than competing niches within the agroecosystem enabling efficient interception of sunlight for photosynthesis and exploration of the soil for water and nutrients. At the same time, the intercropping of these plants is environmentally sustainable, in that reincorporating the unharvested plant biomass into the soil provides soil nutrition and improves soil structure in subsequent years of cultivation. As a result of geographic isolation and the independent evolution of two gene pools, bean landraces grown throughout South America belong primarily to the Andean gene pool while those grown in North America belong to the Middle American gene pool. Many bean landraces cultivated by early indigenous groups are still grown today by First Nations peoples and are an important part of their cultural heritage. Bean landraces were moved from the New World along trading routes across Europe and to Africa where they became an important part of traditional farming systems.

Bean landraces continue to be grown around the world for many reasons including cultural heritage and culinary preference, local adaptability and reliable performance. In less developed regions of the world, mechanized harvesting and high-input production is not economically feasible for many farmers. While bean varieties bred for modern agricultural production are sometimes grown by farmers using traditional means, these varieties often perform poorly because they are not adapted to adverse growing conditions and the unique microclimates where poor farmers grow

them. Marginalized farmers often rely on adapted landraces which are lower-yielding but provide a reliable harvest in uncertain growing conditions.

A second category of germplasm, which has not undergone formal genetic improvement through plant breeding are heirloom varieties. Heirloom crop varieties have been developed from landraces by farmers through intentional or unintentional selection, and like landraces they are regionally adapted and possess desirable agronomic and culinary qualities (De Ron et al., 2016). Often heirloom varieties are associated with a particular ethnic group or geographic region (De Ron et al., 2016). Cultivated landraces sometimes lack agronomic characteristics which make them suited to modern agricultural production, such as vining indeterminate growth habit and disease susceptibility. Domesticated heirloom varieties have improved agronomic characteristics because they have undergone simple selection for desirable traits and ease of cultivation. For example, pod shattering, daylight sensitivity, seed size, plant architecture, and yield may have been actively or passively improved in heirlooms. Heirloom bean varieties often have seed coat colours and patterns, as well as unique flavours and textures, which distinguish them from the common market classes.

Many heirloom bean varieties available in North America are of European origin. Middle American beans were introduced to Western Europe in the early 1500s and Andean beans arrived in the 1530s (De Ron et al., 2016). Landraces adapted to cultivation across Europe, and some crossing between the genepools resulted in germplasm that was uniquely adapted to European growing conditions (De Ron et al.,

2016). Such an extent of recombination and diversification occurred in Europe that the Iberian Peninsula is considered a secondary center of diversity for common bean (De Ron et al., 2016). European heirloom varieties were introduced to North America with European migration and in many cases are still associated with their originating localities or ethnic groups.

The SNF capacity of heirloom and landrace germplasm has not been extensively explored. The successful cultivation of landraces on marginal soils without fertilizer amendments by traditional farmers would suggest that landraces form efficient symbioses with rhizobia and have inherent nitrogen fixing capacity. In the case of European heirlooms, where efficient rhizobia may not have been introduced with the seed to Europe, SNF capacity is more in question. The diversity for SNF capacity in landrace and heirloom genotypes is explored in this thesis. A panel of select European heirloom and indigenous landraces, as well as conventionally-bred varieties is described in Chapter 3. A Honduran panel representative of the diverse landraces, participatory-bred lines, and conventional bean varieties grown by hillside farmers in the West-Central region of that country is described in Chapter 4.

### **2.5.2 Germplasm developed through conventional plant breeding**

Landraces and heirlooms formed the basis of germplasm used in early bean breeding programs. Agricultural production of dry beans was driven throughout the 1900s as the demand for inexpensive protein increased domestically and internationally, and scientific research began uncovering the health benefits of diets

containing beans. With the increase in demand for dry beans came scientific innovation in plant breeding and changes in agronomic practices which increased yield and improved harvestability (Kelly, 2001). Taking plant architecture, or growth habit, as an example, varieties with determinant flowering habit and upright stature are favored by modern producers because of their even seed ripening and their suitability for single-pass mechanical harvest, and considerable improvements were made for these traits in the early 2000s. The beginning of the Green Revolution in the 1960s and the continued use of synthetic nitrogen fertilizer in bean production to date has also led to the development of varieties which yield well under intensive production conditions. Bean breeding has focused on improvements in nutritional composition, resistance to biotic and abiotic stresses, short-season production, among other traits (Kelly, 2010; Assefa et al., 2019). Some breeding between market classes occurs, however maintenance of distinguishing characteristics such as seed size, shape, colour and pattern are essential for consumer marketing, therefore, breeding within market classes is more common (De Ron et al., 2016).

Traditional breeding methods and new molecular breeding technologies have been used in bean variety development. Traditional methods include single seed descent, pedigree and inbred backcrossing, recurrent selection and conical crossing. The innovation of new molecular techniques led to the release of the common bean reference genome (Schmutz et al., 2014), the development of a single nucleotide polymorphism (SNP) genotyping chip (Song et al., 2015), and effective genotyping-by-

sequencing protocols (Schröder et al., 2016), which have enabled genetic characterization of germplasm and the implementation of molecular breeding methods.

The adoption of modern varieties has led to a reduction in landrace cultivation, which has resulted in the narrowing of genetic diversity (Haudry et al., 2007; Maron et al., 2013; Bellucci et al., 2014; Zhou et al., 2015). Bellucci et al (2014) report that domesticated bean has seen a ~60% reduction in nucleotide diversity and 18% reduction of gene expression compared to wild accessions. Arguably, most bean breeding programs incorporate some accessions from outside their programs in crosses, which improves genetic diversity, however modern bean production is less diversified than smallholder agriculture (Brush, 2000; Kull et al., 2013; Bellon et al., 2020). The Middle American Diversity Panel (MDP) is a collection of 280 modern dry bean varieties and breeding lines representing the diversity of genotypes which have descended from landraces domesticated in the Mesoamerican region of the Americas, from Mexico through Central America to northern South America (McClellan, 2008). Collaborators of the American Common Bean Coordinated Agricultural Project (BeanCAP) assembled the panel such that a good representation of the Middle American market classes was included; navy (small white), small black, small red, pink, great northern, pinto and carioca. The MDP was used in studies outlined in Chapter 5 of this thesis to explore the diversity of SNF capacity in modern bean genotypes, to identify genomic regions associated with the trait, and to find putative candidate genes associated with the traits under study.

Another method of formal plant breeding is termed participatory plant breeding (PPB). PPB typically involves collaboration between a formal breeding program and growers to generate varieties to meet their needs. PPB breeding is successfully applied in organic and agroecological scenarios and has seen extensive application in international agriculture development programs. Two general breeding methodologies are applied in PPB; participatory breeding and participatory varietal selection (PVS). In PPB, farmer-grown varieties with desirable traits are crossed with modern varieties to generate new lines with improved characteristics. For example, in Honduras, a red, seda-type bean variety, PM2-Don Rey, was developed through a cross between Carrizalito (conventional) and Paraisito (landrace) using conventional breeding techniques and marker assisted selection in a collaboration between the EAP-Zamorano bean breeding program and marginalized hillside farmers (Rosas and Escoto, 2015). PM2-Don Rey is adapted to hillside production and has the preferred seed quality traits of its landrace parent, but also has the improved characteristics of drought tolerance and disease resistance of its conventional parent (Rosas and Escoto, 2015). In PVS, the formal plant breeder disseminates breeding lines generated in their conventional breeding program to growers for on-farm testing. For example, in order to increase bean yields in a number of Honduran marginal communities, on-farm variety trials are established to compare breeder materials with traditional farmer landraces. In one such study, PVS varieties Amilcar and Campechano had superior yields compared to farmer landrace varieties in favorable and unfavorable environments (Maradiaga et

al., 2016). In marginal or niche production systems, PPB varieties are often found to be superior to conventional varieties.

Taken together, the germplasm contained in the three panels used for studies in this thesis represent both the Andean and Middle American gene pools, large-, medium- and small-seeded market classes, and materials developed by modern breeding and participatory breeding programs as well as landraces and heirloom varieties which have had no formal breeding at all. An expansive study of SNF capacity in this germplasm was undertaken contributing to the body of knowledge for this defining trait in this essential crop.

### **2.5.3 Quantitative genetics for SNF characterization**

Underlying the phenotypic variation seen in all species is genetic variation. The expression of a particular phenotype is controlled by one gene (simple trait) or a number of genes (complex trait) and the interaction of those genes with the environment (G×E interaction). Complex quantitative traits, such as symbiotic nitrogen fixation, are controlled by many genes, with each gene having a different level of contribution to the phenotype. Through quantitative trait loci (QTL) mapping, genomic regions can be linked to phenotypic expression, creating a genetic marker for the trait of interest. In linkage mapping studies a bi-parental cross is made between two homozygous genotypes ( $P_1$  and  $P_2$ ), which have differing phenotypes for the trait of interest, creating an  $F_1$  population of recombinant inbred lines (RILs). The  $F_1$  lines are selfed for several generations to create a population of homozygous RILs which have a high level of

linkage disequilibrium enabling the identification of markers associated with phenotypic traits (Pasam et al., 2012). A study by Ramaekers et al. (2013) of SNF in a RIL population developed from a cross between a Mexican landrace climbing bean and a Peruvian bush bean reported relationships between SNF and additional phenotypic traits. There are two important limitations of QTL mapping; only the relatively small amount of allelic diversity that segregates in a bi-parental population can be evaluated, and the limited number of recombination events generated through a bi-parental cross results in lower QTL mapping resolution (Pasam et al., 2012; Korte and Farlow, 2013). Genome wide association studies (GWAS) are a powerful method that enable the researcher to connect a phenotypic trait of interest that has been scored across a large number of unrelated genotypes to its underlying genetics.

Initially developed as a research tool in human genetics studies, GWAS are used to determine relationships between regions of a genome and a trait of interest. Single nucleotide polymorphism (SNP) markers are favored for use in GWAS because of their abundance and ease of identification using high-throughput genotyping technologies (Huang and Han, 2014). GWAS has been applied successfully in many plant species to elucidate the genes behind many complex traits, from eating quality in apples (Kumar et al., 2013), drought tolerance in maize (Xue et al., 2013), agronomic traits in barley (Pasam et al., 2012), and domestication traits in common bean (Schmutz et al., 2014).

In GWAS, population size, the degree of linkage disequilibrium in the population and the quality of phenotypic data collected for those genotypes all factor into the

success of the study. The use of designed panels of genotypes, such as diversity panels, provides many more recombination events than does a bi-parental cross. Rather than relying on the limited number of recombination events generated in a bi-parental cross, a diversity panel enables the exploration of the large number of ancestral recombination events to identify significant QTL associated with a trait of interest (Pasam et al., 2012). The resolution of QTL determined using an association panel and GWAS is higher (within 2 cM or less) than with a RIL population and association mapping (10-30 cM) (Price, 2006; Pasam et al., 2012). In addition to greater mapping resolution, GWAS assesses more alleles through using diverse germplasm panels, and eliminates the time required to develop a RIL population, reducing overall research time (Yu and Buckler, 2006).

GWAS also have limitations including the large computational power and amount of time required to perform the analyses, rare causative variants and those with small phenotypic effects are difficult to detect, and population structure in the panel can lead to false associations (Korte and Farlow, 2013). Further, in association mapping studies, linkage disequilibrium (LD) is an important consideration because its rate of decay determines the power of the study (Pasam et al., 2012). When LD decays too quickly the number of markers needed to identify functional polymorphisms increases, whereas when LD decays slowly haplotype blocks become too large and resolution is lost (Ersoz et al., 2007). Rossi et al. (2009) report that LD in *P. vulgaris* is high, with a much higher level found in domesticated compared to wild populations. For association studies, a lower level of LD is preferable. One way to minimize LD is to take population structure

into account in an analysis. In a study of population structure and genetic diversity of a panel of bean genotypes from CIAT, Kwak and Gepts (2009) concluded that LD could be reduced in association mapping studies in bean by performing studies within the Andean or Middle American gene pools separately. When population structure is accounted for in association mapping studies, the risk of type I and type II errors occurring is minimized (Pasam et al., 2012). Ultimately, QTL and GWAS are complementary analyses where, for example, the choice of varieties for a bi-parental cross could be informed by GWAS results, and significant associations revealed by either of these genetic studies could be confirmed in the other.

As described above, SNF is a complex biological process and thousands of genes are expressed in legume nodules (Roy et al., 2020), and multiple metabolic processes throughout the plant are directly or indirectly related to nitrogen fixation. Forward and reverse genetic studies have identified nearly 200 genes that are needed for effective symbiosis (Roy et al., 2020). The completion of robust genetic studies on SNF in bean will help reveal the genetic controls for this important characteristic and contribute to the knowledge needed to establish a successful program to breed for the trait.

## **2.6 Recent studies on SNF in beans**

Common beans have historically been considered among the poorest nitrogen fixing legume species (Graham, 1981; Herridge et al., 2008). Rodino et al. (2011) observed that bean landraces have poor SNF capacity and hypothesized this was a

result of low selection pressure for efficient Rhizobial symbiosis throughout its domestication. Conventional bean breeding programs carry out trials using conventional production methods, which use synthetic nitrogen fertilizer and rarely use Rhizobial inoculants. SNF is downregulated by the plant in the presence of adequate exogenous nitrogen (Jiang et al., 2020), therefore conventional breeding methods would not favor the selection of high-SNF genotypes.

SNF is a quantitative trait, affected by many factors, each contributing to the complexity of breeding for this trait. The plant-rhizobia relationship is species-specific; certain Rhizobia form relationships with certain legume crops more effectively than others. Dry beans are promiscuous (Michiels et al., 1998) and are able to form relationships with 27 Rhizobia species, including *R. leguminosarum*, *R. tropici* and *R. etli* (Dall’Agnol et al., 2013), however some interactions produce more fixed nitrogen than others (Rodino et al., 2011). Different environmental factors can affect nitrogen fixation, including the availability of synthetic nitrogen sources. If other forms of nitrogen are available, SNF will be down-regulated (Rondon et al., 2007; Jiang et al., 2020). Soil structure, pH, moisture and salinity can affect nitrogen availability and the symbiotic relationship between plant and Rhizobia (Bohloul et al., 1992). Plant growth habit has an impact, with indeterminate and vining types fixing more nitrogen (Ramaekers et al., 2013).

Early studies of SNF in common bean addressed agronomic questions comparing phenotypic performance of diverse germplasm under varying nutrient

regimes (Munns et al., 1977; Graham and Rosas, 1979; Pereira and Bliss, 1987), uncovering interactions between photosynthesis and SNF (Bethlenfalvai and Phillips, 1977), testing the efficiency of different Rhizobial symbionts (Pacovsky et al., 1984), exploring the differences in SNF capacity between genotypes differing for growth habit and plant density, and in monoculture versus polyculture (Graham and Rosas, 1977, 1978b, 1978a). In his critical review of nodulation and nitrogen fixation in bean, Graham (1981) stated that previous research had revealed little about the genetic variability for SNF in bean, however, by the early 1990s Bliss (1993) commented that “field data [indicated] considerable genetic variability for total N<sub>2</sub> fixation and traits associated with fixation.” Bliss and colleagues released some of the first bean varieties with improved SNF efficiency in the early 1990s by selecting for high yield in N deficient soils (Bliss et al., 1989; Bliss, 1993). A faster and more accurate method of selection in variety development is through the use of marker-assisted selection (MAS).

Genetic linkage maps using molecular markers such as restriction fragment length polymorphisms (RFLPs) and random amplified polymorphic DNA (RAPDs) came into use in the 1990s in bean, accelerating the speed of breeding efforts by enabling breeders to tag genomic regions associated with traits (Tanksley et al., 1989; Bassett, 1991; Freyre et al., 1998). These regions are termed quantitative trait loci (QTL). Bean breeders identified QTL associated with agronomic traits such as disease resistance (Johnson et al., 1995) and drought tolerance (Schneider et al., 1997), as well as numerous QTL associated with SNF related traits. For example, Tsai et al. (1998) used an RFLP linkage map to identify regions associated with nodule number in the BAT 93 x

Jalo EEP558 mapping population, Diaz et al. (2017) used RAPDs, amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), and gene-based markers in a recombinant inbred line (RIL) population of DOR 364 × BAT 477 to discover QTL associated with SNF under two levels of phosphorus supply, and Ramaekers et al. (2013) used RAPDs, SSRs, sequence-characterized amplified region (SCAR), and single nucleotide polymorphism (SNP) markers in an inter-genepool biparental mapping population between G2333 × G19839 to identify QTL for SNF. QTL discovered in these studies identified general regions, however, fine mapping of QTL was not possible due to the relatively low marker density of early marker technologies.

QTL detection improved with the use of SNP markers, which are abundant across genomes and can be used in high-throughput detection platforms enabling fine mapping of traits (Mammadov et al., 2012). Further, the mapping of SNPs to the bean reference genome (Schmutz et al., 2014) enables direct comparison between studies facilitating the identification of common QTL. Farid (2015) reported QTL for percent nitrogen derived from the atmosphere (%Ndfa) on Pv01, Pv05, Pv07, and Pv08 in a RIL population (Mist × Sanilac). Heilig et al. (2017a) used a RIL population (Puebla 152 × Zorro) and reported QTL for traits such as Ndfa on Pv01, Pv06, Pv08 and Pv11, and for %N in seed on Pv05, Pv08 and Pv11. Kamfwa et al. (2019) used a RIL population (Solwezi × AO-1012-29-3-3A) and discovered QTL for %Ndfa on Pv01, Pv04 and Pv09.

The affordable cost and relative ease of SNP genotyping has led to large numbers of varieties being genotyped with this method. Two large diversity panels have

been created in bean that enable the large-scale implementation of genome discovery in bean to improve this crop. A Middle American Diversity Panel (MDP) was assembled for the Bean Coordinated Agricultural Project (McClellan, 2008), and an Andean Diversity Panel (ADP) was assembled for the USAID ARS Grain Legume Project (Cichy et al., 2015). Kamfwa et al. (2015) used the 259-genotype ADP in a GWAS to discover QTL for SNF traits and Chapter 5 of this thesis presents a GWAS analysis of QTL for SNF in the MDP. In their GWAS study, Kamfwa et al. (2015) did not identify any QTL for %Ndfa in the seed but did find QTL for %Ndfa in the shoot on Pv02, Pv03, Pv04, Pv09, Pv10 and Pv11. The discovery of efficient nitrogen fixing varieties and the identification QTL for SNF traits in the ADP and the MDP contributes to the body of knowledge on SNF in bean and may facilitate the application of marker-assisted selection for SNF.

## **2.7 Measuring SNF**

Numerous methodologies have been developed to measure nitrogen fixation in plants. Peoples et al. (2009) provide a comprehensive review of the different approaches applied in legumes. Since legume total N is comprised of soil-acquired N and fixation-derived N<sub>2</sub>, the most effective techniques for measuring SNF distinguish between these components to accurately estimate SNF capacity. Here methods relevant to field trials will be described.

Two methods of measuring SNF in the field are the nitrogen balance method and the nitrogen difference method. The nitrogen balance method involves measuring all N fertilizer inputs added to a system and estimating all N outputs from the system, then

the N surplus above the known inputs is an estimation of fixed N fixed by the legumes in the system. When implemented in the field, inaccuracies are introduced in this method because measuring N volatilization and leaching is difficult. In the nitrogen difference method the legume nitrogen content is compared to a non-leguminous plant grown in close proximity. Because non-leguminous plants cannot fix atmospheric nitrogen, and surplus N detected in the legume can be attributed to nitrogen fixation. This method assumes that the legume and non-legume uptake available soil N at the same rate, and may be more accurate in N-limited growing conditions. Ideally, the non-fixing control plant is a non-nodulating variety of the species under study because this enables precise SNF quantification. In bean, a number of non-fixing varieties have been generated, including a non-nodulating isolate of the CIAT variety BAT 477 (Rondon et al., 2007) and the non-nodulating mutant R99, which was derived from the University of Guelph variety OAC Rico (Park and Buttery, 1992). The nitrogen balance and nitrogen difference methods of field-based N fixation measurement can be carried out with little specialized equipment. The following three methods require more sophisticated lab equipment, and are more easily carried out in controlled environments.

The acetylene reduction assay takes advantage of the dual functionality of the nitrogenase enzyme. Nitrogenase reduces atmospheric N to ammonia, and it is also able to reduce acetylene to ethylene. Gas chromatography is used to measure ethylene generation from excised nodulated legume roots while they are exposed to acetylene in an air-tight container. This method can be used to estimate N fixation capacity over short periods of time, multiple times per day, but cannot provide estimates of total N

fixed over the plant life cycle and is challenging to scale-up to field plot application. A second method which measures nitrogenase activity is termed hydrogen evolution. Nitrogen fixation in nodules generates  $H_2$  as a by-product and using gas chromatography to measure hydrogen evolution provides an estimate of nitrogenase activity. The method is challenging to implement because accurate measurements of nitrogenase activity are possible to obtain only in an atmosphere free of  $N_2$ . Like acetylene reduction, this method is difficult to scale up to field experiments. The third methodology uses xylem sap ureide components, allantoin and allantoic acid, to estimate SNF. When a plant is actively fixing nitrogen the proportion of allantoin and allantoic acid increases in the xylem sap, whereas in plants relying on soil-derived nitrogen, nitrates and amino acids are more abundant. Xylem sap can be collected from controlled environment- or field grown-plants, making it an easier technique to use than acetylene reduction and hydrogen evolution, but like those techniques it provides only a real-time estimate of SNF and not a measurement of total SNF activity over the season. The remaining two SNF measurement technologies provide a time-integrated measure of SNF, and are based on measuring N isotopes in plant material.

The  $^{15}N$  isotope techniques, natural abundance and N enrichment, capitalize on the difference in relative concentration of  $^{14}N$  and  $^{15}N$  isotopes in the atmosphere and the soil. The lighter  $N_2$  isotope  $^{14}N$  is more abundant in the atmosphere than the heavier isotope  $^{15}N$ , which results in soil N containing higher relative proportions of  $^{15}N$ . Plants which carry out active nitrogen fixation incorporate higher levels of atmospheric  $N_2$  ( $^{14}N$ ) into the plant tissue than non-fixing plants. Mass spectrometry is used to measure the

levels of each isotope in plant tissue samples, and the  $^{14}\text{N}:^{15}\text{N}$  ( $\delta^{15}\text{N}$ , ‰) values obtained for fixing and non-fixing reference plants incapable of fixing  $^{14}\text{N}$  are compared; any level of  $^{14}\text{N}$  found in SNF-capable plant tissue above the levels found in non-SNF-capable plant tissue is attributable to nitrogen fixation. A plant that does not fix nitrogen will have a  $^{14}\text{N}:^{15}\text{N}$  ratio that closely reflects the nitrogen ratio of the soil, but a plant that fixes nitrogen will have a  $^{14}\text{N}:^{15}\text{N}$  ratio that more closely reflects the atmospheric N ratio. The percent nitrogen derived from the atmosphere (%Ndfa) can be calculated by comparing the nitrogen ratios of fixing and non-fixing plants using the following formula (Shearer and Kohl, 1988):

$$\%Ndfa = \frac{\delta^{15}\text{N reference plant} - \delta^{15}\text{N N fixing plant}}{\delta^{15}\text{N reference plant} - B}$$

where B is an average  $\delta^{15}\text{N}$  reference value calculated using legume plants grown in N-free media such as coarse sand or turface. The B-value accounts for N isotopic fractionation resulting from the SNF process.

The N-15 enrichment technique differs from the natural abundance technique in that  $^{15}\text{N}$  labeled fertilizer is used to create a pool of labeled N in the soil in addition to the naturally occurring soil N. Plots of a SNF-capable varieties are paired with plots of a control, non-fixing variety. SNF is quantified using the following formula (McAuliffe et al., 1958):

$$\%Ndfa = 100 \times \frac{\text{atom percent excess N15 in N fixing plant}}{\text{atom percent excess N15 in reference plant}}$$

Both the natural abundance and nitrogen enrichment techniques require access to sophisticated mass spectrometry facilities to measure N in the plant tissue, making these methods inaccessible if funds and facilities are not available. The nitrogen enrichment technique has the added input cost of labeled fertilizer. Overall, the natural abundance method is the most easily implemented field technique.

In recent studies on SNF in bean, the favored method of measuring nitrogen fixation has been the natural abundance technique (Ramaekers et al., 2013; Farid and Navabi, 2015; Kamfwa et al., 2015, 2019; Diaz et al., 2017; Farid et al., 2017; Heilig et al., 2017b, 2017a; Barbosa et al., 2018; Wilker et al., 2019, 2020; Reinprecht et al., 2020; Jiang et al., 2020). Either above-ground biomass or seed samples can be used to obtain a time-integrated measure of SNF performance. For experiments presented in this thesis, seed samples were used to quantify SNF because seed N at maturity represents the total N accumulated over the growing season. Shoot N, in contrast, is transitory and fluctuates over the plant life cycle making coordination of sampling times challenging in studies with multiple genotypes (Masclaux-Daubresse et al., 2010). Furthermore, collecting and processing seed samples is faster and less expensive than shoot samples, and a recent study carried out at CIAT found that %Ndfa levels measured in shoot and seed samples are highly correlated (Barbosa et al., 2018).

### **3 Agronomic Performance and Nitrogen Fixation of Heirloom and Conventional Dry Bean Varieties Under Low-Nitrogen Field Conditions**

This chapter was published as:

Wilker, Jennifer; Navabi, Alireza; Rajcan, Istvan; Marsolais, Frederic; Hill, Brett; Torkamaneh, Davoud; Pauls, K. Peter (2019). Agronomic Performance and Nitrogen Fixation of Heirloom and Conventional Dry Bean Varieties Under Low-Nitrogen Field Conditions. *Frontiers in plant science*, 2019-07-26, Vol.10. DOI: 10.3389/fpls.2019.00952

The data generated for this study is publically available and can be found here: <https://doi.org/10.5683/SP2/NZY3W5>.

Author contributions:

Jennifer Wilker, Alireza Navabi and K. Peter Pauls designed the project.

Jennifer Wilker performed the experiments and carried out the phenotypic analysis.

Brett Hill and Frederic Marsolais helped in the seed analysis experiments.

Davoud Torkamaneh carried out the genetic diversity analysis.

Jennifer Wilker and Alireza Navabi analyzed the data.

Alireza Navabi, K. Peter Pauls, and Istvan Rajcan advised the experimental work.

Jennifer Wilker wrote the manuscript with comprehensive editing by K. Peter Pauls.

### 3.1 Abstract

Common beans (*Phaseolus vulgaris* L.) form a relationship with nitrogen-fixing Rhizobia and, through a process termed symbiotic nitrogen fixation (SNF), they are provided with a source of nitrogen. However, beans are considered poor nitrogen fixers, and modern production practices involve routine use of N fertilizer, which leads to the downregulation of SNF. High-yielding, conventionally-bred bean varieties are developed using conventional production practices and selection criteria, typically not including SNF efficiency, and may have lost this trait over decades of modern breeding. In contrast, heirloom bean genotypes were developed before the advent of modern production practices and may represent an underutilized pool of genetics which could be used to improve SNF. This study compared the SNF capacity under low-N field conditions of a collection of heirloom varieties and conventionally-bred dry bean varieties. The heirloom-conventional panel (HCP) consisted of 42 genotypes from various online seed retailers or from the University of Guelph Bean Breeding program seedbank. The HCP was genotyped using a single nucleotide polymorphism (SNP) array to investigate genetic relatedness within the panel. Field trials were conducted at three locations in ON, Canada from 2014 to 2015 and various agronomic and seed composition traits were measured, including capacity for nitrogen fixation (using the natural abundance method to measure seed N isotope ratios). Significant variation for SNF was found in the panel. However, on average, heirloom genotypes did not fix significantly more nitrogen than conventionally bred varieties. However, five heirloom genotypes fixed >60% of their nitrogen from the atmosphere. Yield ( $\text{kg ha}^{-1}$ ) was not

significantly different between heirloom and conventional genotypes, suggesting that incorporating heirloom genotypes into a modern breeding program would not negatively impact yield. Nitrogen fixation was significantly higher among Middle American genotypes than among Andean genotypes, confirming previous findings. The best nitrogen fixing line was Coco Sophie, a European heirloom white bean whose genetic makeup is admixed between the Andean and Middle American gene pools. Heirloom genotypes represent a useful source of genetics to improve SNF in modern bean breeding.

### **3.2 Introduction**

Since its origin in central Mexico some 2 MYA, common bean (*Phaseolus vulgaris* L.) has diverged into two gene pools in Central America and South America, been domesticated and spread throughout the world (Gepts et al., 1999; Kaplan and Lynch, 1999; Bitocchi et al., 2017). First Nations' ancestral groups gathered wild beans and cultivated them with other crops, including maize (*Zea* spp.) and squash (*Cucurbita* spp.). Beans were among the crops which explorers brought back to Europe after they visited the Americas. Centuries of cultivation and movement of seed through human migration and trade led to beans becoming staples in diets around the world, and inseparable parts of numerous cultural heritages. Recent years have seen increases in heirloom bean popularity, stretching beyond farmers' markets and seed exchanges to specialty grocers, culinary circles and mainstream culture.

Before the establishment of formal bean breeding programs, landraces maintained by First Nations groups and European settlers were grown throughout North America (Kelly, 2010). Aside from their historical origin and association with early farming systems, bean landraces are characterized by having local genetic adaptation, high genetic diversity and a lack of formal genetic improvement (Villa et al., 2005). Similarly, heirloom beans have not been developed through formal plant breeding programs. Heirloom bean varieties derive from bean landraces introduced to Europe by explorers to the Americas. European farmers informally selected beans adapted to their local environment and passed the seeds from generation to generation as family heirlooms. As a result of the informal selection process heirloom beans are less genetically diverse than landraces. In many instances, heirloom beans have distinctive characteristics such as unique seed coat colours/patterns, and desirable flavors or cooking traits. However, yield, disease resistance, and growth habit may be poor compared to conventionally-bred, relatively modern, bean varieties. In contrast, modern bean varieties conform to standard requirements for size and colour particular to a few market classes, and are bred to produce high yields under conventional production practices (Kelly, 2010). Market demands and producer requirements are believed to have led to narrow breeding objectives and reduced genetic diversity in modern bean varieties (Singh, 1988). This reduction in genetic diversity may have also led to a reduction in diversity and capacity for nitrogen fixation in modern bean genotypes.

Between the two gene pools of common bean, the Andean gene pool is much less diverse than the Middle American gene pool. This reduced diversity is a result of a

bottleneck created when founder populations established the Andean gene pool at a distance from the center of origin of bean, in present-day central Mexico (Bitocchi et al., 2012). The independent and parallel domestication of beans beginning some 8000 years ago in the Andean and Middle American regions resulted in separate gene pools of domesticated bean (Papa and Gepts, 2003; Chacon et al., 2005; Kwak and Gepts, 2009; Rossi et al., 2009; Mamidi et al., 2011; Nanni et al., 2011; Bitocchi et al., 2013, 2017; Schmutz et al., 2014; Rendón-Anaya et al., 2017). The divergence has led to some difficulties in hybridization between Andean and Middle American genotypes (Johnson and Gepts, 1999). Nevertheless, introgression between gene pools has been found in bean collections throughout the world (Gioia et al., 2013). In particular, introgression has influenced the diversity of the bean germplasm grown across Europe, where 40.2% of accessions show introgression compared to the much lower level of introgression in North American genotypes, which is 12.3% (Gioia et al., 2013).

Symbiotic nitrogen fixation (SNF) is an ancient trait, characteristic of the Fabaceae family. In bean, various *Rhizobium* species (including *R. leguminosarum*, *R. tropici* and *R. etli*) inhabit root nodules and fix atmospheric nitrogen, which is utilized by the plant in exchange for carbohydrates. However, among modern leguminous crops, beans are considered to be poor nitrogen fixers (Hardarson et al., 1993). In the latter half of the 20th century, research largely concluded that the rates of nitrogen fixation in bean were low, at 25 to 71 kg N<sup>2</sup> fixed ha<sup>-1</sup> for mid- to long-season varieties (Graham, 1981). These values are considerably lower than rates for soybean at the time, which ranged from 78-161 fixed ha<sup>-1</sup> in one study (Muldoon et al., 1980). LaRue and Patterson

reviewed multiple studies of nitrogen fixation in legume species and calculated that soybean fixed 75 kg N<sup>2</sup> ha<sup>-1</sup> on average while dry beans fixed just 10 kg N<sup>2</sup> ha<sup>-1</sup> (Larue and Patterson, 1981). However, recent studies have examined hundreds of bean genotypes for traits related to nitrogen fixation and reported wide-ranging capacity for these traits (Ramaekers et al., 2013; Kamfwa et al., 2015; Diaz et al., 2017; Farid et al., 2017; Heilig et al., 2017a; Wilker et al., 2020), indicating genotypic and genetic diversity, which could be exploited to enhance this trait through breeding. For example, it was found among twelve modern genotypes tested, nitrogen fixing capacity ranged from 2.7 to 69.7 kg N<sup>2</sup> fixed ha<sup>-1</sup>, which represents a range of 5.2% to 78.5% nitrogen derived from the atmosphere (%Ndfa) (Farid, 2015). Another study examined 79 navy and black commercial varieties and advanced breeding lines under organic production and found a similar range for nitrogen fixing capacity (16 to 94 kg N<sup>2</sup> ha<sup>-1</sup>) and for %Ndfa (9.8% to 71.7%) (Heilig, 2015).

Nitrogen fixation and root nodule traits are controlled by multiple genes. They are affected by environmental conditions, and are difficult to measure. As a result, modern bean breeding programs do not focus on breeding genotypes efficient at nitrogen fixation but rather release high-yielding genotypes, which perform consistently under conventional production practices, including the application of 33-67 kg ha<sup>-1</sup> of nitrogen fertilizer and crop protection chemicals (OMAFRA, 2009). In contrast, many heirloom varieties were developed and are maintained under natural growing conditions where fertility is managed using crop rotation and non-synthetic fertilizer and symbiosis with appropriate Rhizobia species occurs naturally or is enhanced by the use of inoculants.

Therefore, heirloom genotypes may be a genetic resource for modern breeding programs that contain genetic diversity for nitrogen fixation and other traits that have not been eroded by modern breeding practices.

Nitrogen fixation capacity of modern dry bean varieties needs to be improved, while the discovery of additional diversity for the trait will provide genetic resources for breeding programs. The current study tests the hypothesis that heirloom beans have a greater capacity for nitrogen fixation than conventionally-bred bean varieties and examines whether they could be useful sources of germplasm to improve this trait. The objectives of this study were to compare heirloom and conventionally-bred bean genotypes from both the Andean and Middle American gene pools for their capacity for symbiotic nitrogen fixation, to assess whether genetic diversity has been lost over years of modern breeding, and to assess agronomic characteristics to determine the suitability of using heirloom varieties in modern breeding programs.

### **3.3 Materials and Methods**

#### **3.3.1 Plant material**

The heirloom-conventional panel (HCP) was assembled in 2014 and contained 25 heirloom and 17 conventionally-bred dry bean genotypes. In the first growing season, six genotypes failed to reach physiological maturity and were removed from the panel. For the second growing season, six new genotypes were added, resulting in the HCP consisting of 23 heirloom and 19 conventional genotypes. Only genotypes for which two or three location years of data was collected are included in the analyses in this chapter. Seed images of the genotypes in the HCP are displayed in Figure 3.1.

Heirloom seeds were purchased as pure line varieties from Canadian seed retailers (Heritage Harvest Seed [www.heritageharvestseed.com](http://www.heritageharvestseed.com), Assiniboine Tipis [www.assiniboinetipis.com](http://www.assiniboinetipis.com), and Annapolis Seeds [www.annapolisseeds.com](http://www.annapolisseeds.com)) with the intent of including a wide representation of seed coat patterns, seed sizes and plant growth habits. Heirloom seed coat patterns ranged from uniform, to bi-colour spotted/speckled/striped, or tri-colour; often very different in appearance compared to conventional market classes. In this study, the term “heirloom” refers to genotypes of the HCP that were not derived from a conventional bean breeding program. Given the limited information available for each heirloom genotype in this panel (see compiled variety summaries at <https://doi.org/10.5683/SP2/NZY3W5>), it was impossible to further categorize these genotypes into groupings such as “improved landrace” or “vintage cultivar”.



**Figure 3.1.** Images of *Phaseolus vulgaris* genotypes included in the heirloom-conventional panel. Twenty-three heirloom bean genotypes and nineteen conventionally bred bean genotypes grown at Elora and Belwood, Ontario, 2015 are shown. White bar = 1 cm.

Seed of conventional bean genotypes was sourced from the University of Guelph Bean Breeding program's seed stores. Germplasm was chosen to represent a range of market classes, seed sizes and growth habits, mirroring the diversity found among the heirloom genotypes, where possible. Conventional genotypes were registered with the Canadian Food Inspection Agency between 1938 and 2016 and were developed by modern breeding programs and institutions [including: University of Guelph (UG), Michigan State University (MSU), United States Department of Agriculture-Agriculture Research Station (USDA-ARS), Crop Development Centre (CDC) in Saskatchewan, International Center for Tropical Agriculture (CIAT), Instituto Colombiano Agropecuario (ICA), and Agriculture Agri-Food Canada (AAFC) in Ontario and Alberta]. Descriptions of the genotypes, including market class, origin, seed size, plant growth habit, and gene pool membership are presented for the HCP in Table 3.1.

**Table 3.1** Market class, seed size, growth habit, genepool and race for 42 dry bean (*Phaseolus vulgaris* L.) genotypes of the heirloom-conventional panel.

Variety Code	Variety Name	Market class	Origin, year of CFIA registration	Seed Size†	Growth habit¶	Gene pool‡	Reference
<b>Heirloom</b>							
2	Annie Jackson	red calypso	Russian heirloom, na	medium	III	Andean	Heritage Harvest Seed
3	Arikara Yellow	Canario mexicano	Arikara FN, 2002	medium	I	Andean	Mündel <i>et al.</i> , 2004
5	Canadian Wonder	red kidney	unknown, na	large	I	Andean	Heritage Harvest Seed
7	Deseronto Potato	white kidney	Mohawk FN, na	large	II	Andean	Heritage Harvest Seed
8	Early Mohawk	Cranberry	Iroquois FN, na	large	I	Andean	Assiniboine Tipis
13	Hidatsa Shield Figure	Unknown	Hidatsa FN, na	large	II	Andean	Heritage Harvest Seed
15	Iroquois Cornbread	speckled red kidney	Iroquois FN, na	large	I	Andean	Heritage Harvest Seed
17	Jacob's Cattle	Unknown	unknown, na	large	I	Andean	Heritage Harvest Seed
21	Snowcap	Unknown	unknown, na	large	II	Andean	Heritage Harvest Seed
22	Speckled Algonquin	Cranberry	Algonquin FN, na	medium	I	Andean	Heritage Harvest Seed
23	Sweeney Family	speckled red kidney	Canadian heirloom, na	large	I	Andean	Heritage Harvest Seed
25	Worcester Indian	Tan	unknown, na	medium	I	Andean	Heritage Harvest Seed
46	Coco Sophie	Navy	French heirloom, na	medium	III	Andean	Heritage Harvest Seed
47	Fisher	Tan	Algonquin FN, na	medium	I	Andean	Assiniboine Tipis
1	Amish Gnuttle	Unknown	Seneca FN, na	small	III	MA	Annapolis Seeds
4	Canadian Wild Goose	grey speckle	unknown, na	small	II	MA*	Heritage Harvest Seed
10	Flagg	speckled black kidney	Iroquois FN, na	large	III	MA	Assiniboine Tipis
11	Ga Ga Hut Pinto	Pinto	Seneca FN, na	medium	II	MA	Heritage Harvest Seed
12	Hidatsa Red	small red	Hidatsa FN, na	medium	II	MA	Heritage Harvest Seed
16	Kahnawake Mohawk	Pinto	Mohawk FN, na	large	III	MA*	Annapolis Seeds
18	Mandan black	Black	Mandan FN, na	small	II	MA	Heritage Harvest Seed
20	Roja de Seda	small red	Central American heirloom, na	small	III	MA	Heritage Harvest Seed
36	PI207262	Tan	Gene bank plant introduction, na	small	II	MA	Coyne and Schuster, 1974
<b>Conventional</b>							
26	Red Rider	Cranberry	AAFC, 2008	large	I	Andean	Park <i>et al.</i> , 2009
27	Majesty	red kidney	AAFC, 2005	large	II	Andean	Park, 2006
28	CDC Sol	Yellow	CDC, 2010	medium	I	Andean	Vandenberg and Bett, 2013
29	Yeti	white kidney	UG, 2013	large	I	Andean	Khanal <i>et al.</i> , 2016
43	OAC Inferno	light red kidney	UG, 2011	large	I	Andean	Smith <i>et al.</i> , 2012
9	Hi N line	Black	UG, na	small	II	MA	breeding line
30	Zorro	Black	MSU, 2012	small	II	MA	Kelly <i>et al.</i> , 2009

Variety Code	Variety Name	Market class	Origin, year of CFIA registration	Seed Size†	Growth habit¶	Gene pool‡	Reference
32	R99	Navy	AAFC, na	small	II	MA*	Park and Buttery, 2006
33	OAC Rico	Navy	UG, 1983	small	II	MA	Beversdorf, 1984
34	Mist	Navy	UG, 2013	small	II	MA	Khanal et al., 2017
35	ICA Pijao	Black	ICA, na	small	II	MA	Voysest, 2000
37	ICB-10	Black	USDA-ARS, na	small	II	MA	Miklas et al., 1999
38	VAX 4	Tan	CIAT, na	small	II	MA	Singh et al., 2001
39	OAC Speedvale	Navy	UG, 1991	small	II	MA	Beattie et al., 2003
41	OAC Spark	navy	UG, 2012	small	I	MA	Khanal et al., 2017b
42	OAC Rex	navy	UG, 2002	small	II	MA	Michaels et al., 2006
44	Michelite	navy	MSU, 1940	small	II	MA	Kelly, 2010
45	Corvette	navy	AAFC, 1943	small	II	MA	McGregor et al., 1956
48	Limelight	navy/wt kidney	AAFC, 1972	medium	I	MA	Sears, 1986

† Small = 13 to 29 g per 100 seeds; medium = 30 to 45 g per 100 seeds; large = 46 to 63 g per 100 seeds

¶ Growth habit according to Singh (1982)

‡ Genepool assigned according to STRUCTURE analysis. Threshold genetic contribution from assigned genepool was > 50%. Note: genotypes marked with (\*) were assigned to genepool according to market class appearance - these genotypes were not SNP genotyped MA = Middle American

### 3.3.2 Field experimental design and maintenance

Field trial locations were selected based on low soil nitrogen levels as measured by pre-planting soil tests which showed that NO<sub>3</sub>- concentrations were under 5 ppm (“very low”) or 5-10 ppm (“low”) and by site crop rotation histories that indicated that no dry bean crops had been produced at the sites for the previous decade, at a minimum. Soil nitrogen and growing season details can be found in Appendix A, Table S3.1.

Clean seed of each genotype was coated with commercially available Nodulator® (Becker-Underwood) *Rhizobium leguminosarum* bv *phaseoli* inoculant prior to planting. The day before planting, 1/8 teaspoon (approximately 0.2 g) of inoculant powder was added to each seed envelope and the contents were shaken to coat the seeds. Inoculated seed was stored at the Elora Research Station (ERS) at 4 °C until

planting to maintain inoculant viability. The entire contents of each envelope (coated seed + loose inoculant powder) was planted.

The HCP was grown in three low-nitrogen field location-years using a rectangular lattice design (6x7) with two replications. At the ERS in 2014, 100 seeds of each genotype were grown in single-row plots 6 m in length with approximately 6 cm between plants and 60 cm spacing between entry rows. In 2015, the HCP was grown in another field at the ERS and at an offsite location near Belwood, Ontario. Increased seed availability enabled planting of 135 seeds in 4-row plots (150 cm x 90 cm, 37.5 cm between rows) with approximately 5 cm between plants within rows.

Throughout the growing season, plots were maintained with standard practices, except no-nitrogen fertilizer was used. Pre-plant fertilizer (0:20:20) at a rate of 200 kg ha<sup>-1</sup> was applied approximately one week prior to planting. Pre-plant herbicides [200 ml ha<sup>-1</sup> Pursuit (BASF) and 1.5 L ha<sup>-1</sup> Frontier (BASF)] were applied to control broadleaf and grass weeds. At Elora 2014, insecticides against leaf hoppers were applied July 11 [1.0 L ha<sup>-1</sup> Lagon (Loveland products), 40 ml ha<sup>-1</sup> Matador (Syngenta)], and fungicides against Anthracnose and root rot were applied July 11 [0.5 L ha<sup>-1</sup> Quadris (Syngenta), 1.0 L ha<sup>-1</sup> Allegro (Syngenta)] and again against Anthracnose on August 7 [400 ml ha<sup>-1</sup> Headline (BASF), 1 L ha<sup>-1</sup> Allegro]. At Elora 2015, herbicides were applied on July 15 [2.25 L ha<sup>-1</sup> Basagran (BASF), 0.67 L ha<sup>-1</sup> Excel Super (Excel Crop Care)], 1 L ha<sup>-1</sup> Assist (BASF)] followed by insecticides against leaf hoppers [1.0 L ha<sup>-1</sup> Cygon (FMC Corporation), 40 ml ha<sup>-1</sup> Matador] and fungicides against Anthracnose [400 ml ha<sup>-1</sup>

Headline (BASF), 1 L ha<sup>-1</sup> Allegro] on July 16. Fungicide against Anthracnose (0.5 L ha<sup>-1</sup> Quadris) and insecticide against leaf hoppers [200 ml ha<sup>-1</sup> Admire (Bayer)] were again applied August 6. At Belwood 2015, insecticides (1.0 L ha<sup>-1</sup> Cygon, 40 ml ha<sup>-1</sup> Matador) and fungicides (400 ml ha<sup>-1</sup> Headline, 1 L ha<sup>-1</sup> Allegro) were applied on July 16. The Belwood plots were treated against Anthracnose (0.5 L ha<sup>-1</sup> Quadris) and leaf hoppers (200 ml ha<sup>-1</sup> Admire) again on August 6. Plots at all locations were manually weeded once before canopy closure each year.

### **3.3.3 Phenotyping**

Days to flowering was observed throughout July and August and was recorded as the date when 50% of the plants in a plot had one flower open. The days to flowering data were converted into growing degree days to flowering (GDD<sub>f</sub>) by summing the calculated GDD temperature from daily max and min temperatures. Hourly temperatures were recorded at the ERS by the University of Guelph School of Environmental Sciences Agrometeorology group (available at: <https://www.uoguelph.ca/ses/service/weather-records>). For the Belwood site, temperature data from the nearest Government of Canada weather station data was used (Fergus Shand Dam, available at: [http://climate.weather.gc.ca/index\\_e.html](http://climate.weather.gc.ca/index_e.html)).

Relative leaf chlorophyll content was measured twice during the growing season (when the mean number of plots had reached (1) the second trifoliate stage, and (2) at 100% flowering) using a SPAD 502 Plus Chlorophyll Meter (Konica-Minolta). The meter was calibrated according to the manufacturers' instructions each time the unit was

powered-on ([https://www.specmeters.com/assets/1/22/2900P\\_SPAD\\_502.pdf](https://www.specmeters.com/assets/1/22/2900P_SPAD_502.pdf)). The middle leaflet in the top-most, fully expanded trifoliate leaf was used for the measurements and three plants were sampled per plot.

Plots were rated for days to maturity throughout September and early October. Plots were considered to have reached maturity when they reached harvest maturity (>50% of plants were dry and pods were brown). Days to maturity measurements were converted into growing degree days to maturity (GDDm) in the same way as for GDDf (see above).

For plant biomass data collection, three plants were randomly sampled from mature plots, placed in large paper bags, and dried in a re-purposed tobacco kiln (De Cloet Bulk Curing Systems, model TPG-360, Tillsonburg, Ontario) at 33 °C at the Elora Research Station for 24-48 hours. Prior to weighing, roots were cut from each plant and above-ground biomass was measured. Plants were then threshed using an indoor belt thresher (Agriculex SPT-1A, Guelph, ON), their seed collected, weighed and counted. Harvest index (biomass/seed weight) as well as 100 seed weight (HSW) were calculated.

At Elora 2014, the harvest was staggered according to maturity. The plots were pulled by hand at maturity and threshed at the side of the field using a Wintersteiger plot combine (Wintersteiger AG, Upper Austria, Austria) with a Classic Seed-Gauge weighing system by Harvest-Master (Juniper Systems Inc., Utah, USA) and plot seed

weight and moisture content were recorded. In 2015, plot harvest took place after all plots reached maturity with the same Wintersteiger combine.

### **3.3.4 Seed isotope analysis**

The natural abundance method was used to calculate percent nitrogen derived from the atmosphere (%Ndfa) for each genotype (Shearer and Kohl, 1988). Seed was used for this assessment because seed N at maturity represents the total N accumulated over the growing season, whereas shoot N is transitory and fluctuates over the plant life cycle making coordination of sampling times challenging in studies with multiple genotypes (Masclaux-Daubresse et al., 2010). Additionally, %Ndfa levels measured in shoot and seed samples are highly correlated, and processing of seed samples is faster and less expensive than shoot tissue (Barbosa et al., 2018).

Nodule traits (number and size), as an indicator of nitrogen fixing capacity, were not measured in this study. Numerous studies in dry bean have found that nodule traits are not correlated with nitrogen fixation capacity. For example, Farid (2015) found no correlation between nodule numbers and SNF, and our study of SNF in the Middle American Diversity Panel (see Chapter 5) found no correlation between SNF and nodule size or nodule number. An in-field ureide assay was not feasible and a controlled environment study was not initiated for this panel.

To prepare for gas-chromatography mass-spectrometry (GCMS) analysis, a 5 g subsample of seed from each plot was oven-dried (Blue M Electric, SPX Corporation) at 60 °C at the University of Guelph for 24 hours prior to being ground to a coarse powder

in a coffee grinder (various models used). The coarse seed powder was further processed into a fine powder suitable for gas chromatography mass spectrometry (GCMS) analysis by grinding a sub-sample in a small Eppendorf tube along with a steel bead in a bead mill (Beadruptor 12, Omni International Inc). Samples (5 mg) of bean powder were measured into small tin capsules (8 x 5 mm, standard weight, Elemental Microanalysis) using an analytical balance (Quintix 65-1S, Sartorius Lab Instruments GmbH & Co), enveloped and compressed into a tiny pellet so that no atmosphere remained in the capsule. The bean powder pellets were collected in 96-well plates and sent to the Agriculture and Agri-food Canada (AAFC) GCMS facility in Lethbridge, Alberta for analysis. The samples were analyzed with a Finnigan Delta V Plus (Thermo Electron, Bremen, Germany) Isotope Ratio Mass Spectrometer (IRMS) fitted with a Flash 2000 Elemental Analyzer (Thermo Fisher Scientific, Voltaweg, The Netherlands) and Conflo IV (Thermo Fisher Scientific, Bremen, Germany) interface between the IRMS and the analyzer. A standardized curve for nitrogen content was created using an alfalfa standard provided by the AAFC GCMS facility. Further isotope standards L-glutamic acid USGS40 and USGS41 (United States Geological Survey) were included with each plate of samples processed to normalize isotope values and enable inter-lab comparison. Samples were analyzed for %N,  $\delta^{15}\text{N}$  (‰), and  $\delta^{13}\text{C}$  (‰).

The natural abundance method uses the following equation,

$$\%N_{dfa} = \frac{\delta^{15}\text{N}_{reference\ plant} - \delta^{15}\text{N}_{N - fixing\ plant}}{\delta^{15}\text{N}_{reference\ plant} - B}$$

where,  $\delta^{15}\text{N}$  reference plant is the rate of  $\delta^{15}\text{N}$  in the reference genotype (R99),  $\delta^{15}\text{N}$ -fixing plant is the  $\delta^{15}\text{N}$  of the N-fixing bean genotype and B is the average  $\delta^{15}\text{N}$  of beans grown in an environment where its entire N source is from fixation (Peoples et al., 2009). The B-value was obtained for this experiment as described by Farid (2015). Briefly,  $\delta^{15}\text{N}$  was measured and averaged for 20 bean genotypes from both the Andean and Middle American gene pools, which were grown in a growth room in N-free media. Normalized  $\delta^{15}\text{N}$  values were used for all genotypes and an average of  $\delta^{15}\text{N}$  values for R99 were used in the %Ndfa calculations.

### **3.3.5 Genotyping**

Leaf tissue samples were collected from young plants of 42 genotypes grown in a controlled environment (16 h photoperiod, 22 °C) at the University of Guelph. For 29 genotypes, DNA was extracted using the manufacturer's instructions for the NucleoSpin Plant II kit (Macherey-Nagel, Germany), and for the remaining 13 genotypes the DNeasy Plant Mini Kit (Qiagen, Canada) was used. DNA quality was tested using a spectrophotometer (ND-1000, Nanodrop) and a fluorometer (Qubit 2.0, Invitrogen by Life Technologies), and DNA of 39 genotypes was determined to be of sufficient quality to send for genotyping. Genomic DNA was analyzed at the Genome Quebec Innovation Centre (McGill University, Montreal) for single nucleotide polymorphisms (SNPs) using the Illumina Infinium iSelect Custom Genotyping BeadChip (BARCBEAN6K\_3) containing 5,398 SNPs (Song et al., 2015).

### 3.3.6 Identity by state analysis

SNP data from the above analysis was imported to TASSEL (Bradbury et al., 2007) for filtering such that the retained SNPs were present in 95% of the panel and the minor allele frequency (MAF) was 0.05. This resulted in 39 genotypes and 4,704 SNPs retained for further analysis. TASSEL was used to generate a genotype distance matrix and R software (R Core Team, 2013) was used to create a dendrogram using the *dendextend* package (Galili, 2015). The hierarchical clustering function, *hclust* (Müllner, 2013) was used to perform the cluster analysis using the UPGMA method. The *as.dendrogram* function was used to create dendrograms, which were then modified in R using the *dendextend* package and the *circlize* package (Gu et al., 2014). STRUCTURE (Pritchard et al., 2000) was used to determine the population genetic structure of the HCP. The analysis (20 replications) was performed with the length of burn-in set at 5,000 and the number of MCMC replications after burn-in set at 50,000. A range of genetic groups (2K to 9K) were tested and the number that best fit the data was determined by visualizing the STRUCTURE results and using the  $\Delta K$  statistic in STRUCTURE HARVESTER online (<http://taylor0.biology.ucla.edu/structureHarvester/>) (Evanno et al., 2005; Earl and VanHoldt, 2012).

### 3.3.7 Nucleotide diversity analysis

The levels of genetic diversity in the heirloom vs. conventional categories and the Andean vs. Middle American categories of the HCP were assessed. The  $\pi$  statistic provides an indication of polymorphism within a population as measured by nucleotide diversity (Nei and Li, 1979), and Tajima's D provides an indication of selection pressure

(Tajima, 1989). The 5K SNP dataset was used to calculate  $\pi$  and Tajima's D with VCFtools 0.1.12b (Danecek et al., 2011), and a MAF cutoff of  $\geq 0.01$  and a window of 1000 bp was used. Genome-wide averages of  $\pi$  and Tajima's D for each germplasm category were generated by taking the average across all windowed calculations. A t-test (GraphPad Prism8) was used to determine differences in both  $\pi$  and Tajima's D values between heirloom and conventional categories within each gene pool.

### **3.3.8 Statistical analysis**

Analysis of variance (ANOVA) tests were performed on the data collected from each environment and the environments combined using the MIXED procedure in SAS (version 9.4, SAS Institute, Cary, NC, USA, 2012). In each ANOVA, genotypes were considered fixed effects while all other effects and the interaction effects were considered random. The Shapiro-Wilks test (Shapiro and Wilk, 1965) was performed on the residuals in the UNIVARIATE procedure to test their normality. Random and independent distributions of the residuals were visually examined by plotting the studentized residuals against the predicted values. Data that generated outlier residuals were removed from the data set. Further, single degree of freedom contrasts were conducted in ANOVA between genotype categories, heirloom vs. conventional and Middle American vs. Andean. Repeated measures of leaf chlorophyll content (SPAD) were taken, and a separate ANOVA test was used to compare SPAD values at each time point. In each ANOVA, the genotype least squared means (LSmeans) were computed using the LSMEANS statement in the MIXED procedure.

The pair-wise Pearson's coefficients of correlation were computed for all traits measured using the CORR procedure in SAS. The RINCOMP and PRINQUAL procedures were used in SAS to generate the principal component (PC) values, to estimate the proportion of variance accounted for by each PC, and to plot PC1 against PC2 to generate a genotype × trait (GT) biplot (Yan and Rajcan, 2002) to determine genotype and trait interactions overall and in each environment.

### **3.4 Results**

#### **3.4.1 Origins and phenotypic characteristics of selected beans**

The germplasm comprising the HCP includes genotypes with a wide diversity of seed traits (colours, patterns, shapes, and sizes) found in dry bean. According to the descriptions from the source seed retailers, 16 of the heirloom genotypes are part of the cultural heritage of North American First Nations communities (the Algonquin, the Iroquois, the Seneca, and the Mohawk from the Great Lakes region of North America; the Arikara, the Hidatsa, and the Mandan from the Plains region in present-day United States). Genotype descriptions for the remaining 9 heirloom genotypes suggest the varieties were passed down through communities or families from as far back as colonial times. For example, Sweeney Family Heirloom was first grown by the Sweeney family in Nova Scotia and has moved with the family and been grown in Alberta (Heritage Harvest Seeds). Further, while Sweeney Family Heirloom shows similarities to other heirloom genotypes, it is considered a unique variety by heirloom seed growers. Coco Sophie is a European variety from the 1700s (Heritage Harvest Seed). Amish Gnuttie (Amish Nuttie; also known as Cornhill Bean or Mayflower) is described by some

retailers as a variety that was introduced to America with the early settlers and has been grown by Amish communities for generations, while other variety descriptions suggest that Amish Gnuttle originated with the Seneca First Nation.

The heirloom category was equally split between Andean and Middle American types (Table 3.1) and a variety of seed coat colour patterns are represented, including bi-colour, yellow eye, pinto/cranberry, and uncommon solid colours (Figure 3.1), which make them unique and difficult to categorize using conventional market classes. The conventional category was equally split between Andean and Middle American types and could mostly be categorized as kidney (dark red, light red, white), cranberry, yellow, white or black market class beans (Table 3.1, Figure 3.1).

### **3.4.2 Field conditions**

Fields with low nitrogen levels were used in this study to maximize the potential for SNF activity. In the growing seasons prior to 2014 and 2015, fields at the ERS had been planted with high-N demanding cereal crops to remove as much available nitrogen from the soil as possible. At the Belwood location, the field had been used to produce mixed hay with minimal inputs in the growing seasons previous to our trial. Soil test results showed that nitrate ( $\text{NO}_3^-$ ) levels ranged between 3.7 ppm and 8.6 ppm and ammonium ( $\text{NH}_4$ ) levels ranged between 2.6 ppm and 6.1 ppm in the bean root zone. Soil analysis laboratory guidelines indicate that levels of  $\text{NO}_3^-$  below 10 ppm are considered low (A & L Canada Laboratories).

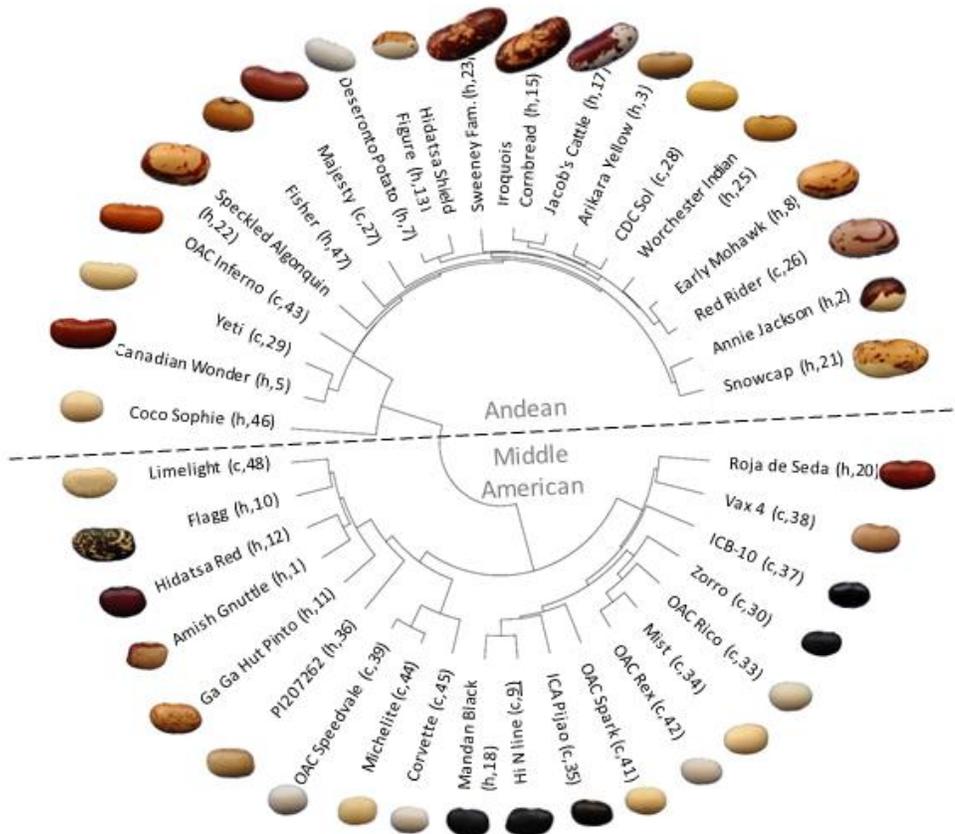
Planting in 2015 occurred two weeks later than in 2014 as a result of wet spring weather. Despite the late start to the 2015 season, accumulated growing degree days (GDD) over the growing season were similar for all three locations (Elora 2014 - 1912.8, Elora 2015 - 1862.6, Belwood 2015 - 2012.3). A summary of pre-plant soil test results, precipitation and total growing degree days for all location-years is provided in Appendix A, Table S3.1.

### **3.4.3 Genetic analysis of relatedness**

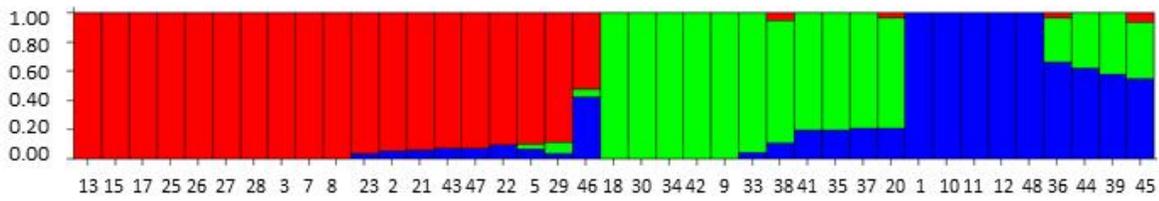
The HCP was composed of genotypes from both the Middle American and Andean gene pools, however the gene pool composition and genetic relatedness of the genotypes was unknown. An identity-by-state (IBS) analysis on SNP genetic data from 39 genotypes of the HCP was undertaken to confirm genotype membership in either gene pool and to determine the genetic relationships among them. The IBS analysis found that the panel is composed of three sub-groupings, with 19 genotypes belonging to the Andean gene pool and 20 belonging to the Middle American gene pool (11 race Mesoamerica, 9 race Durango-Jalisco). In the dendrogram (Figure 3.2A), large-seeded genotypes generally sorted into the Andean grouping while smaller-seeded genotypes sorted into the Middle American grouping. STRUCTURE analysis (Figure 3.2B) and determination of the best-fit  $\Delta K$  value for the panel (Figure 3.2C) using STRUCTURE HARVESTER confirmed that there were three genetic groupings in the panel, corresponding to the Andean gene pool and two sub-groupings in the Middle American gene pool. The IBS analysis revealed the degree of genetic relatedness between modern and heirloom genotypes. For example, all of the black seed coat genotypes

**Figure 3.2.** Analysis of genetic structure and relatedness of thirty-nine genotypes of the heirloom-conventional panel. A. Dendrogram of genetic relatedness generated in R. Andean genotypes above and Middle American genotypes below the mid-line. Heirloom or conventional category membership is denoted by an “h” or “c”, respectively, along with the genotype code number; B. STRUCTURE plot indicating the division of the panel into three genetic sub-groupings, Andean (red), Mesoamerica (green) and Durango-Jalisco (blue); C. Delta K plot from fastSTRUCTURE indicating that the most appropriate sub-division of the panel is into 3 genetic groupings; D. Principle component analysis plot confirming three genetic groupings in the panel.

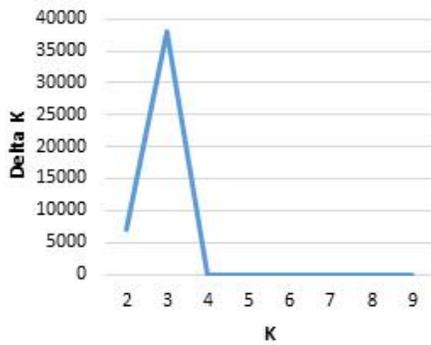
**A**



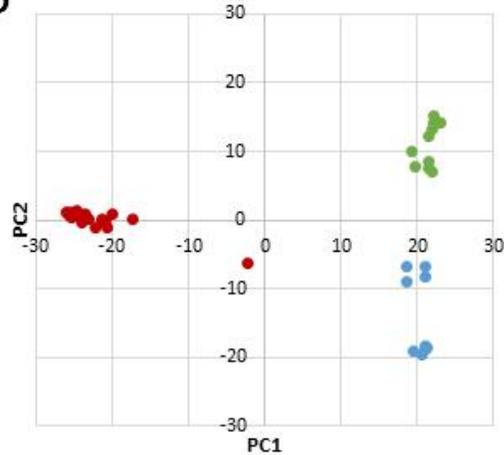
**B**



**C**



**D**



belong to the race Mesoamerica grouping of the Middle American genepool, and the University of Guelph breeding line, “Hi N” (Figure 3.2A and 3.2B, #9), is most closely related to the heirloom genotype Mandan Black (#18) and the conventional genotype ICA Pijao (#35), but it is less similar to Zorro (#30) and ICB-10 (#37). Assignment of varieties to either genepool based on genetic composition was generally in agreement with genepool assignments using seed characteristics, except for a few cases. For example, the large, flat-seeded Limelight (# 48) and Flagg (# 10) genotypes, which appear to be of Andean origin, belong by genetic analysis, to the Middle American genepool.

Evidence of admixture is apparent for a number of genotypes in the panel. Within the Middle American genepool, 5 of the genotypes are of entirely Durango-Jalisco and 5 are of entirely Mesoamerican ancestry. The remaining 10 Middle American genotypes are admixed between Durango-Jalisco and Mesoamerican races with 4 genotypes also containing < 10% genetic material from the Andean genepool. Less admixture is evident within the Andean genepool, where 10 genotypes are entirely Andean and 8 genotypes contain < 10% Middle American genetic material. Coco Sophie (Figure 3.2A and 3.2B, #46), a round, white bean of European heritage is unique in that it is approximately 50% Andean and 50% Middle American. In the principle component analysis (Figure 3.2D) Coco Sophie falls midway between the three genepool/race clusters. Repeated iterations of the STRUCTURE analysis of the panel assigned Coco Sophie to the Andean genepool 60% of the time, whereas on the basis of its seed colour and shape this genotype would have been assigned to the Middle American genepool.

#### 3.4.4 Nucleotide diversity among genotype categories

Nucleotide diversity was measured in the HCP to ascertain whether genotypes comprising the heirloom category are more diverse than those in the conventional category, and similarly whether genotypes belonging to the Middle American gene pool are more diverse than those belonging to the Andean gene pool. According to the  $\pi$  and Tajima's  $D$  statistics, nucleotide diversity for the heirloom category overall ( $\pi = 3.64 \times 10^{-4}$ ,  $D = 7.262 \times 10^{-3}$ ) was very similar to that found in the conventional category overall ( $\pi = 3.88 \times 10^{-4}$ ,  $D = 7.908 \times 10^{-3}$ ).

The number of SNPs among the Middle American genotypes in the HCP was 3294 compared to 2696 for the Andean genotypes. Nucleotide diversity using  $\pi$ , for the Middle American group ( $\pi = 3.64 \times 10^{-4}$ ) was significantly ( $p=0.0014$ ) larger than for the Andean group ( $\pi = 2.13 \times 10^{-4}$ ). Similarly, Tajima's  $D$  statistic for the Middle American gene pool ( $D = 0.79$ ) was significantly higher ( $p=0.0009$ ) than for the Andean gene pool ( $D = -0.18$ ).

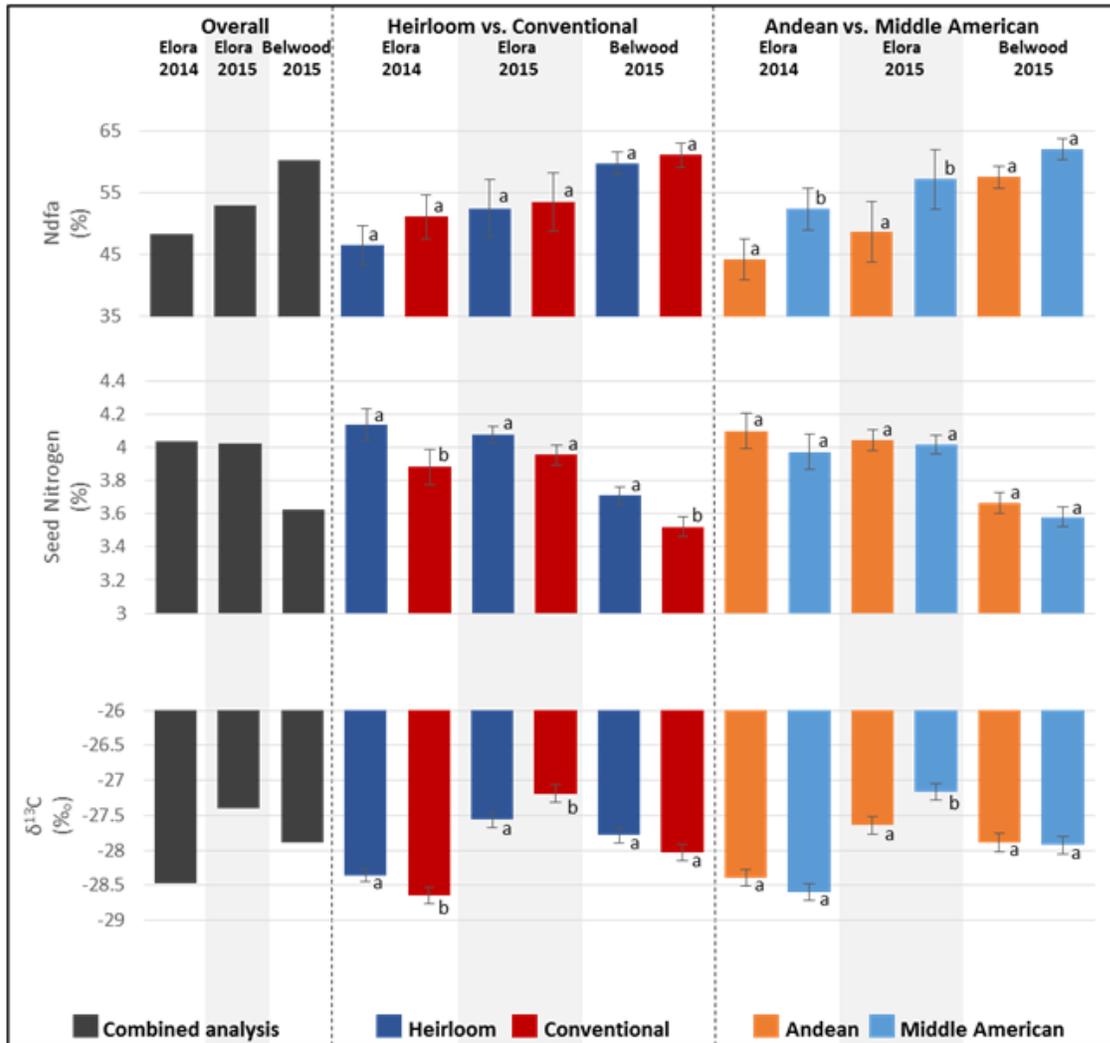
Nucleotide diversity between heirloom and conventional categories was further analyzed within the gene pools. In the Middle American gene pool, nucleotide diversity was not significantly different ( $\pi$ :  $p=0.4137$ ;  $D$ :  $p=0.9783$ ) between the heirloom ( $\pi = 4.08 \times 10^{-4}$ ,  $D = 0.63$ ) and the conventional genotypes ( $\pi = 3.61 \times 10^{-4}$ ,  $D = 0.64$ ). However, within the Andean gene pool, heirloom nucleotide diversity was significantly higher ( $p=0.0082$ ) in conventional genotypes ( $\pi = 3.98 \times 10^{-4}$ ) than heirloom genotypes

( $\pi = 2.35 \times 10^{-4}$ ), but Tajima's *D* values were not significantly different ( $p=0.1310$ ) between heirloom ( $D = -0.09$ ) and conventional genotypes ( $D = 0.47$ ).

### 3.4.5 Diversity for seed isotope traits

Significant differences were seen among the genotypes for the seed traits analyzed by GCMS, including: nitrogen derived from the atmosphere (%Ndfa;  $p=0.0002$ ), seed nitrogen content (%N;  $p<0.0001$ ), and carbon discrimination ( $\delta^{13}\text{C}$ ;  $p<0.0001$ ) (Appendix A, Table S3.2; Figure 3.3). Among the categories overall, significant differences were found for %Ndfa ( $p<0.0001$ ), where Middle American genotypes (mean 62.16%) outperformed Andean (mean 54.82%) genotypes, and for seed nitrogen content ( $p<0.0001$ ), where heirloom genotypes (mean 3.97 %N) contained higher levels of N than conventional (mean 3.79 %N) genotypes. Significant differences were not found for other category comparisons of seed composition traits. While the effect of environment alone was not significant, the environment by genotype interaction effect (env\*ENTRY) was significant for all seed composition traits (Appendix A, Table S3.2), and warranted further exploration.

When seed composition traits are analyzed for each location, significant genotype effects were found. At Elora 2014 (Appendix A, Table S3.3; Figure 3.3), significant differences were found between genotypes for %Ndfa ( $p=0.0072$ ), seed N content ( $p=0.0105$ ), and carbon discrimination ( $p=0.0031$ ). A comparison of genotype categories found significantly higher levels for %Ndfa ( $p=0.0144$ ) in Middle American



**Figure 3.3.** Means for seed composition traits measured from seed harvested at three field locations from genotypes of the heirloom-conventional panel. Comparisons within each year and subcategory ± standard error are presented. Means labeled with different letters within categories are significantly different according to ANOVA,  $p = 0.05$ .

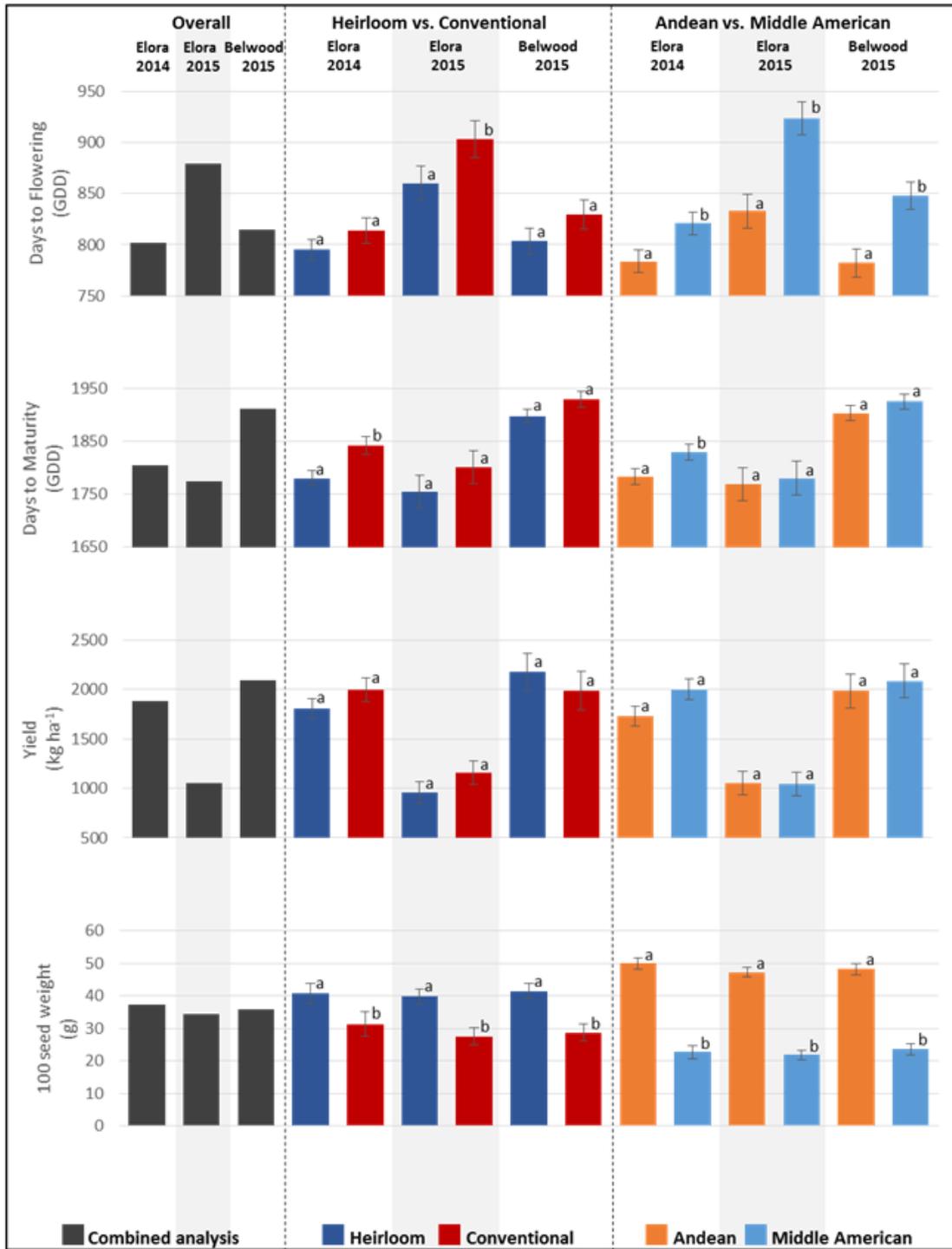
genotypes (mean 54.37%) compared to Andean genotypes (mean 45.94%); and conventional genotypes (mean 53.06%) fixed more nitrogen than heirloom genotypes (mean 48.23%), although this difference was not statistically significant. For seed N

content ( $p=0.0070$ ), significant differences were seen at Elora 2014 where the heirloom category (mean 4.14 %N) had higher seed N content than the conventional category (mean 3.88 %N), however no significant differences were seen between Andean (mean 4.1 %N) and Middle American (mean 3.97 %N) genotypes. For carbon discrimination ( $\delta^{13}\text{C}$ ), significant differences ( $p=0.0452$ ) were found between heirloom (mean -27.5) and conventional (mean -27.8) genotypes, but not between Andean (mean -27.54) and Middle American (mean -27.75). Although significant differences were found among genotypes for %Ndfa ( $p=0.0049$ ), seed N content ( $p=0.0126$ ), and carbon discrimination ( $p=0.0001$ ) at Belwood in 2015 (Appendix A, Table S3.4; Figure 3.3), the only genotype category comparison where significant differences were found was for seed N content ( $p=0.0251$ ), where heirloom genotypes had higher %N (mean 3.71) than conventional genotypes (mean 3.52). At Elora 2015 (Appendix A, Table S3.5; Figure 3.3), significant differences were found between genotypes for %Ndfa ( $p=0.0026$ ), seed N content ( $p<0.0001$ ), and carbon discrimination ( $p=0.0078$ ), and comparisons of genotype categories found further significant differences. Similar to results for 2014, at Elora 2015 Middle American genotypes (mean 63.54%) fixed significantly ( $p=0.0020$ ) more nitrogen than the Andean genotypes (mean 54.19%), while the difference between heirloom (mean 58.36) and conventional (mean 59.59) was not significant ( $p=0.6980$ ). For seed N content at Elora 2015, no significant differences were seen between heirloom (mean 4.08 %N) vs conventional (mean 3.95 %N) or Andean (mean 4.04 %N) vs Middle American (mean 4.02 %N) categories. For carbon discrimination ( $\delta^{13}\text{C}$ ), significant differences ( $p=0.0233$ ) were found between heirloom (mean -27.8) and conventional

(mean -27.41) genotypes. Additionally, significant differences ( $p=0.0049$ ) between Andean (mean -27.9) and Middle American (mean -27.34) were found.

### **3.4.6 Diversity for agronomic traits**

For agronomic traits, significant differences in the combined environments analysis were found among genotypes for days to flowering (GDD;  $p<0.0001$ ), days to maturity (GDD;  $p<0.0001$ ), yield ( $\text{kg ha}^{-1}$ ;  $p=0.0003$ ), and hundred seed weight (g;  $<0.0001$ ) (Appendix A, Table S3.2; Figure 3.4). Among categories overall, significant differences were found for days to flowering, where heirloom genotypes (mean 819.44 GDD) flowered significantly earlier than conventional genotypes (mean 849.83 GDD), and Andean genotypes (mean 798.80 GDD) flowered significantly earlier than Middle American genotypes (mean 865.44 GDD). Similarly, for days to maturity, heirloom genotypes reached maturity significantly earlier (mean 1811.24 GDD) than conventional genotypes (mean 1857.28 GDD). Significant differences were not found for either genotype category comparison for yield ( $\text{kg ha}^{-1}$ ), however significant differences were found for 100 seed weight, where heirloom genotypes (mean 40.7 g) were larger than conventional genotypes (mean 28.8 g), and Andean (mean 48.35 g) genotypes were larger than Middle American genotypes (mean 22.70 g). While the effect of environment alone was not significant, the environment by genotype interaction effect (env\*ENTRY) was significant for days to flowering, yield and 100 seed weight (Appendix A, Table S3.2), and warranted further exploration.



**Figure 3.4.** Means for agronomic traits measured at 3 field locations for the genotypes of the heirloom-conventional panel. Comparisons within each year and subcategory  $\pm$  standard error are presented. Means labelled with different letters within categories are significantly different according to ANOVA,  $p=0.05$ .

When agronomic traits are analyzed for each location, significant genotype effects were found. At Elora 2014 (Appendix A, Table S3.3; Figure 3.4), significant differences were found between genotypes for days to flowering (GDD;  $p=0.0485$ ), days to maturity (GDD;  $p<0.0001$ ), yield ( $\text{kg ha}^{-1}$ ;  $p=0.0033$ ) and 100 seed weight (g;  $p<0.0001$ ), and comparisons of genotype categories found further significant differences. For days to flowering, Middle American genotypes (mean 820.86 GDD) flowered significantly earlier than Andean genotypes (mean 783.60 GDD); and heirloom genotypes (mean 795.02 GDD) flowered earlier than conventional genotypes (mean 813.74 GDD), although this difference was not statistically significant. For days to maturity, heirloom genotypes (mean 1780.40 GDD) matured significantly earlier than conventional genotypes (mean 1842.20 GDD), and Andean genotypes (mean 1783.29 GDD) matured significantly earlier than Middle American genotypes (mean 1829.42 GDD). For yield, no significant differences were found between heirloom and conventional genotypes nor between Andean and Middle American genotypes. For 100 seed weight (HSW), heirloom genotypes had significantly higher weights (mean 40.82 g) than conventional genotypes (mean 31.35 g), and Andean genotypes (mean 49.97 g) were significantly heavier than Middle American genotypes (mean 22.69 g). At Belwood 2015 (Appendix A, Table S3.4; Figure 3.4), significant differences were found between genotypes for days to flowering ( $p<0.0001$ ), days to maturity ( $p<0.0001$ ), HSW ( $p<0.0001$ ). No significant differences were found among genotypes for yield. When category comparisons were performed, significant differences were found for days to flowering, with Andean genotypes (mean 782.02 GDD) flowering earlier than Middle

American genotypes (mean 848.38 GDD). For HSW, heirloom genotypes (mean 41.51 g) were significantly heavier than conventional genotypes (mean 28.68 g), and Andean genotypes (mean 48.23 g) were significantly heavier than Middle American genotypes (mean 23.59 g). At Elora 2015 (Appendix A, Table S3.5; Figure 3.4), significant differences were found between genotypes for days to flowering ( $p < 0.0001$ ), days to maturity ( $p = 0.0002$ ), yield ( $p < 0.0001$ ), and comparisons of genotype categories found further significant differences for days to flowering and HSW. In particular, heirloom genotypes (mean 860.49 GDD) flowered significantly earlier than conventional genotypes (mean 903.76 GDD), and Andean genotypes (mean 833.02 GDD) flowered significantly earlier than Middle American genotypes (mean 924.12 GDD). For HSW, it was found that heirloom genotypes (mean 39.87 g) were significantly heavier than conventional genotypes (mean 27.53 g), and Andean genotypes (mean 47.24 g) were significantly heavier than Middle American genotypes (mean 21.80 g).

When random effects in the combined ANOVA are considered, the effect of environment is not significant for any trait, however the genotype by environment interaction was significant for all traits, except Days to Maturity (Appendix A, Table S3.2), indicating that genotype performance for most traits was affected by the growing environment. The block within environment interaction was not significant at any location, however the incomplete block within the environment by block interaction was significant for %Ndfa, yield, and days to flowering (Appendix A, Table S3.2), indicating some variation in performance across the field sites.

### 3.4.7 Diversity for leaf chlorophyll content

As a repeated measure, leaf chlorophyll content (SPAD) was analyzed in separate F-tests. Overall, SPAD values differed significantly by genotype during each field season ( $p < 0.0001$ , Appendix A, Table S3.6) and at all locations, significant differences were found among genotypes for leaf chlorophyll content ( $p < 0.0001$ , Appendix A, Table S3.6). In 2015, at both locations, significant differences were seen between SPAD measurements taken at different growth stages (early vegetative stage vs. reproductive stage) (Appendix A, Table S3.6). Furthermore, at each location the growth stage at which leaf chlorophyll content was measured had a significant effect on genotype SPAD performance (significant SPADT\*G interaction; Appendix A, Table S3.6). The observation within block by genotype by SPAD time interaction was significant in all environments (Appendix A, Table S3.6).

Leaf chlorophyll content rating comparisons were also made between genotype categories using ANOVA. In 2014, no significant difference was found between heirloom and conventional genotypes ( $p = 0.7372$ ), whereas Middle American genotypes had significantly higher SPAD ratings (mean SPAD value 37.19) than Andean genotype ratings (mean SPAD value 34.39). At Belwood 2015, significant differences ( $p = 0.0121$ ) were found between heirloom (mean SPAD value 37.15) and conventional (mean SPAD value 38.90) genotypes, and further SPAD sampling time ( $p = 0.0002$ ) and category\*SPADT interaction ( $p = 0.0164$ ) were significant for the heirloom vs. conventional comparison. When genotypes were categorized according to genepool membership, significant differences ( $p = 0.0013$ ) were found between Middle American

(mean SPAD value 39.24) and Andean (mean SPAD value 36.43) genotypes. In addition, SPAD sampling time was significant ( $p=0.0007$ ), as was the interaction between genepool category and SPAD sampling time ( $p=0.0222$ ). At Elora 2015, no significant difference was found between heirloom and conventional genotypes ( $p=0.7840$ ), nor SPAD sampling time or the interaction (SPADT\*breeding category). When genotypes were compared according to genepool membership, significant differences ( $p<0.0001$ ) were found, where Middle American genotypes had significantly higher SPAD ratings (mean SPAD value 35.07) than Andean genotypes (mean SPAD value 31.86). Neither the SPADT nor the genepool\*SPADT interaction was significant at Elora in 2015.

#### **3.4.8 Nitrogen fixation in the HCP**

Table 3.2 ranks all genotypes in the panel for nitrogen fixing capacity as measured by %Ndfa. At Elora 2014, the %Ndfa range was between 20.8% (Jacob's Cattle, heirloom, Andean) and 76.4% (Flagg, heirloom, Middle America) with an average value of 48.3%. At Elora 2015, the %Ndfa range was from 19.9% (Fisher, heirloom, Andean) to 70.9% (Coco Sophie, heirloom, Middle America) with an average value of 53.0%. At Belwood 2015, the %Ndfa range was from 43.5% (Limelight, conventional, Andean) to 76.3% (Hi N line, conventional, Middle American) with an average value of 60.3%.

**Table 3.2.** Nitrogen derived from the atmosphere (%Ndfa) and differential ranking of common bean genotypes at three locations (Elora, Belwood) and in two seasons (2014, 2015).

Code	Genotype	Category	Genepool‡	Elora 2014		Elora 2015		Belwood 2015		Combined	
				%Ndfa	Rank	%Ndfa	Rank	%Ndfa	Rank	%Ndfa	Rank
46	Coco Sophie	Heirloom	Andean	NA	NA	70.9	1	75.7	3	69.0	1
9	Hi N line	Conventional	MA	67.2	2	61.2	8	76.3	1	66.9	2
18	Mandan black	Heirloom	MA	45.7	22	66.0	3	75.9	2	63.7	3
20	Roja de Seda	Heirloom	MA	60.4	6	54.0	21	75.0	4	62.7	4
36	PI207262	Heirloom	MA	51.2	15	60.0	12	74.7	5	62.4	5
10	Flagg	Heirloom	MA	76.4	1	59.4	14	61.8	18	62.4	6
38	Vax4	Conventional	MA	62.0	4	59.0	16	58.9	22	60.3	7
1	Amish Gnuttie	Heirloom	MA	63.4	3	66.3	2	48.8	37	59.8	8
43	OAC Inferno	Conventional	Andean	NA	NA	54.2	20	68.3	11	59.7	9
30	Zorro	Conventional	MA	45.1	24	59.2	15	69.5	8	59.3	10
11	Ga Ga Hut Pinto	Heirloom	MA	52.9	11	64.9	4	58.9	23	58.9	11
13	Hidatsa Shield Figure	Heirloom	Andean	53.1	10	59.7	13	63.0	15	58.7	12
23	Sweeney Family Heirloom	Heirloom	Andean	49.3	19	57.9	17	66.7	12	58.0	13
26	RedRider	Conventional	Andean	60.6	5	52.5	22	62.1	16	57.4	14
37	ICB-10	Conventional	MA	42.4	27	61.7	6	64.6	14	57.1	15
42	OAC Rex	Conventional	MA	50.2	18	51.1	26	69.7	6	56.9	16
35	ICA Pijao	Conventional	MA	52.4	12	62.9	5	53.1	31	56.8	17
4	Canadian Wild Goose	Heirloom	MA	51.1	16	48.4	30	69.5	7	56.2	18
27	Majesty	Conventional	Andean	56.6	8	51.4	25	60.8	21	55.8	19
12	Hidatsa Red	Heirloom	MA	52.1	13	60.1	11	57.3	27	55.4	20
16	Kahnawake Mohawk	Heirloom	MA	48.2	20	48.3	31	69.2	9	55.2	21
34	Mist	Conventional	MA	35.2	30	61.2	7	68.7	10	54.5	22
44	Michelite	Conventional	MA	NA	NA	48.6	29	60.9	20	53.9	23
29	Yeti	Conventional	Andean	45.0	25	54.9	19	62.0	17	53.7	24

Code	Genotype	Category	Genepool‡	Elora 2014		Elora 2015		Belwood 2015		Combined	
				%Ndfa	Rank	%Ndfa	Rank	%Ndfa	Rank	%Ndfa	Rank
33	OAC Rico	Conventional	MA	45.2	23	51.8	24	58.2	25	53.3	25
15	Iroquois Cornbread	Heirloom	Andean	56.6	7	48.1	32	49.2	36	52.2	26
39	OAC Speedvale	Conventional	MA	36.4	29	60.2	10	61.4	19	52.1	27
5	Canadian Wonder	Heirloom	Andean	31.3	32	61.1	9	56.2	29	51.8	28
7	Deseronto Potato	Heirloom	Andean	51.1	17	48.7	28	57.8	26	50.9	29
28	CDC Sol	Conventional	Andean	55.4	9	37.7	38	56.8	28	50.3	30
41	OAC Spark	Conventional	MA	51.6	14	45.2	34	53.0	32	48.7	31
45	Corvette	Conventional	MA	NA	NA	56.3	18	50.2	35	48.4	32
21	Snowcap	Heirloom	Andean	47.8	21	44.6	35	46.2	39	48.1	33
22	Speckled Algonquin	Heirloom	Andean	29.9	33	44.3	36	66.0	13	47.6	34
2	Annie Jackson	Heirloom	Andean	35.0	31	52.3	23	58.7	24	47.6	35
25	Worcester Indian	Heirloom	Andean	39.8	28	49.0	27	50.7	34	46.6	36
3	Arikara Yellow	Heirloom	Andean	42.8	26	41.4	37	43.6	40	42.3	37
8	Early Mohawk	Heirloom	Andean	22.2	34	46.6	33	51.2	33	40.8	38
17	Jacob's Cattle	Heirloom	Andean	20.8	35	37.0	39	54.1	30	38.3	39
48	Limelight	Conventional	Andean	NA	NA	35.8	40	43.5	41	35.8	40
47	Fisher	Heirloom	Andean	NA	NA	19.9	41	48.7	38	32.1	41
				<b>LSmean</b>	<b>se</b>	<b>LSmean</b>	<b>se</b>	<b>LSmean</b>	<b>se</b>	<b>LSmean</b>	<b>se</b>
<b>Heirloom</b>				46.5	3.26	52.5	4.62	59.8	1.8	53.0	3.55
<b>Conventional</b>				51.1	3.54	53.6	4.73	61.1	1.9	54.4	3.6
<b>Andean</b>				44.3	3.34	48.6	4.9	57.6	1.8	50.3	3.45
<b>Middle American</b>				52.4	3.35	57.1	4.86	62.1	1.7	57.2	3.44

# Code for genotypes shown in Figure 3.1

‡ Genepool according to Gepts (1988) MA = Middle American.

Although no significant differences were found in nitrogen fixing capacity between the heirloom and conventional genotype categories (Table S3.2), when ranked overall, four of the top five genotypes for nitrogen fixation capacity in this study were heirloom genotypes (including: Coco Sophie, Mandan Black, Roja de Seda, and PI2017262; Table 3.2). The conventional genotypes which ranked in the top ten for nitrogen fixation consist of two breeding lines (Hi N, Vax 4) and two recently released varieties (OAC Inferno and Zorro).

In addition to desirable growth habit, the modern varieties also possess disease resistance; the cream-colored Vax 4 is resistant to Common Bacterial Blight (CBB) and Bean Common Mosaic (BCM) virus (Singh et al., 2001), the light red kidney bean OAC Inferno is BCM and Anthracnose resistant (Smith et al., 2012), and the black bean Zorro is resistant to rust and Anthracnose and partially resistant to CBB (Kelly et al., 2009). Disease resistance and good nitrogen fixing performance make these genotypes desirable candidates for breeding programs. Nitrogen fixing capacity was consistently higher in Middle American than Andean genotypes, and four of the top five nitrogen fixing genotypes belong to the Middle American gene pool (Mandan Black, Roja de Seda, PI207262, and Hi N line).

#### **3.4.9 Trait correlation**

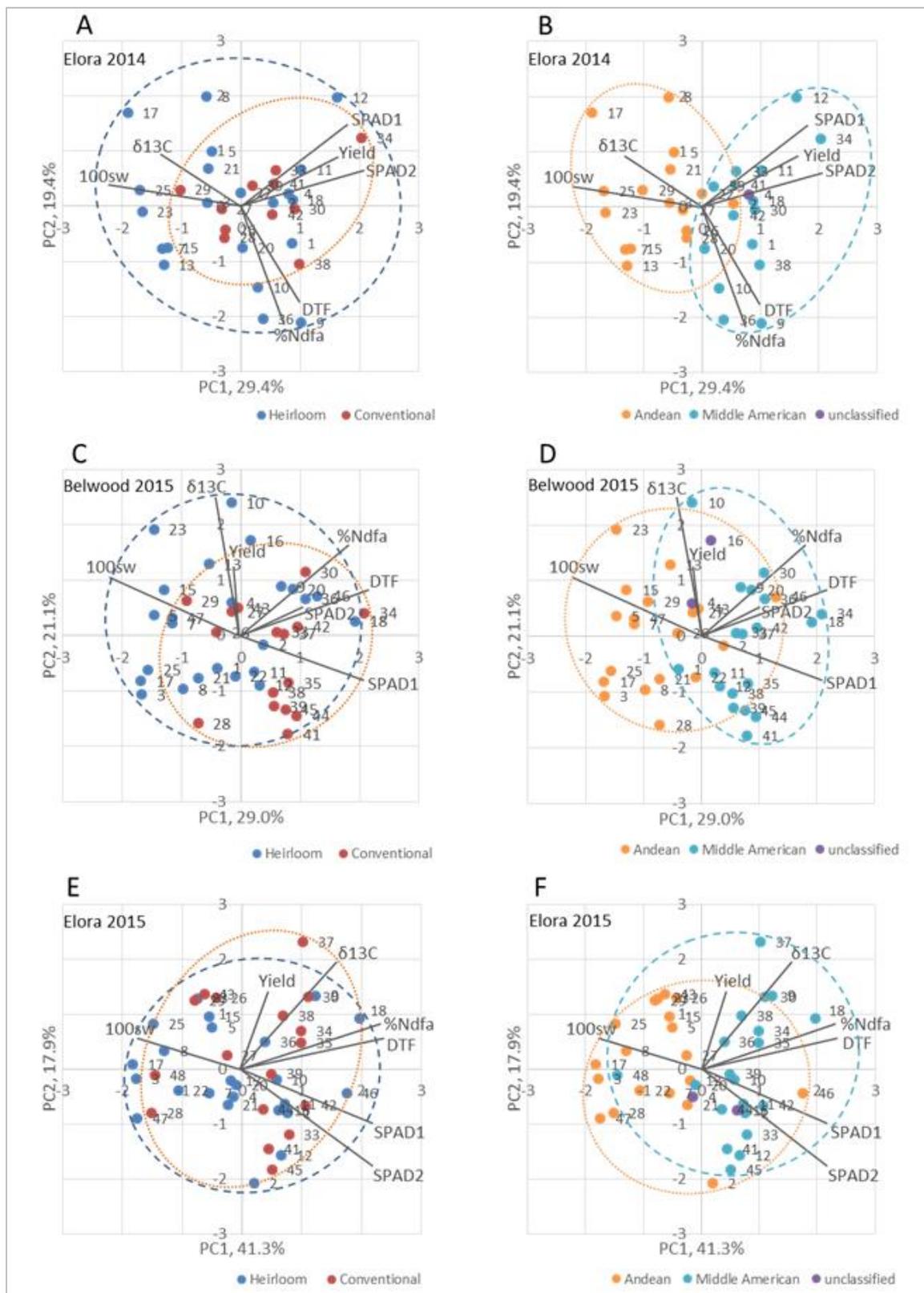
At Elora 2014, the correlation between days to flowering and days to maturity and the correlation between the first and second SPAD measurement time were positive and significant (Appendix A, Table S3.7). At Elora 2015, significant, positive

correlations were found between %Ndfa and all traits except yield; a significant, negative correlation was seen that year between seed N and yield (Appendix A, Table S3.8). Similarly, at Belwood 2015, significant, positive correlations were seen for %Ndfa and all traits except yield and  $\delta^{13}\text{C}$  (Appendix A, Table S3.9). Yield was not found to be significantly correlated with any trait in 2015 at either location (Appendix A, Tables S3.8 and S3.9).

The first two principle components in trait biplots (Figure 3.5) accounted for 49.9% of the variation in Elora 2014 (panels A, B), 64.9% in Elora 2015 (panels C, D), and 51.3% in Belwood 2015 (panels E, F). The positive relationships between days to flowering and %Ndfa at each location-year are indicated by the acute angle formed by the vectors for these traits. The near-right angles formed by the %Ndfa and SPAD vectors at each location-year indicate that no relationship exists between these traits. The obtuse angle formed by the carbon discrimination ( $\delta^{13}\text{C}$ ) and %Ndfa vectors in Elora 2014 indicates a negative relationship between these traits, while in 2015 the vectors are closer together forming a smaller angle and indicating a closer relationship.

When genotypes are categorized according to breeding history (Figure 3.5, panels A, C, E), the conventional and heirloom genotypes occupy largely overlapping areas of the plot. However, when the genotypes are categorized according to genepool membership (Andean vs Middle America, Figure 3.5, panels B, D, F), a significant fraction of the Andean population falls exclusively into areas defined by PC1. In these representations the MA genotypes are clustered in the direction of the %Ndfa vector.

**Figure 3.5.** Biplot analysis of traits for genotypes of the heirloom-conventional panel in 3 location-years. Panels A, C and E divide the panel based on heirloom/conventional-bred categories. Panels B, D and F divide the panel based on Andean and Middle American gene pool membership. The ellipses encompass all the genotypes of a particular category. DTF, days to flowering; Yield, yield ( $\text{kg ha}^{-1}$ ); 100sw, hundred seed weight (g);  $\delta^{13}\text{C}$ , carbon discrimination; %Ndfa, percent nitrogen derived from the atmosphere; SPAD1, leaf chlorophyll content, second trifoliolate stage; SPAD 2, leaf chlorophyll content, 100% flowering.



## **3.5 Discussion**

### **3.5.1 Genetic diversity is greater in the Middle American than in the Andean gene pool**

The IBS and nucleotide diversity analyses of the HCP were in accordance with the often-observed higher level of genetic diversity within the Middle American gene pool compared to the Andean gene pool. Multiple studies have found higher levels of diversity in the Middle American than in the Andean gene pool (Koenig and Gepts, 1989; Beebe et al., 2000, 2001; Papa and Gepts, 2003; McClean et al., 2004; Mamidi et al., 2011, 2013; Bellucci et al., 2014; Schmutz et al., 2014). In a study of AFLP and SSR marker diversity in domesticated and wild bean populations, Rossi et al. (2009) found evidence of a bottleneck event before domestication in the Andean gene pool. Bitocchi et al. (2012) also found significant differences in genetic diversity between wild Middle American and Andean genotypes lending support to the occurrence of a genetic bottleneck prior to domestication of the Andean gene pool. Therefore, the current low level of diversity among domesticated Andean genotypes was caused by bottlenecks during the establishment of the wild progenitor bean populations and during domestication (Bitocchi et al., 2013). The HCP has similar nucleotide diversity and no genetic differentiation between heirloom and conventional genotypes in the Andean gene pool.

Decades of breeding, based on the use of a limited pool of elite varieties has generated concern that this practice has led to a narrowing of crop genetic diversity in modern crop varieties (Plucknett et al., 1987; Gepts, 2006). However, the perception

that heirloom genotypes are more genetically diverse than varieties from modern breeding programs was not supported by the genetic diversity analysis of the HCP. The interspersed nature of heirloom and conventional genotypes around the dendrogram (Figure 3.2) suggests that decades of isolated development of these two germplasm categories has not led to genetic divergence. Furthermore, genetic diversity measurements with the  $\pi$  and Tajima's D statistics were not significantly lower for conventional genotypes than heirloom genotypes in this study. This was true for the overall comparison and the comparison within the Middle American gene pool. Within the Andean gene pool, greater nucleotide diversity was indicated by  $\pi$  and Tajima's D within the conventional genotypes compared to the heirloom genotypes.

While this finding is in accordance with analyses performed in other crop species which concluded that modern breeding practices have not reduced genetic diversity (van de Wouw and Hintum, 2010), it contradicts a recent comprehensive study in bean based on SSR marker diversity among wild, landrace and modern American genotypes of each gene pool that concluded that genetic diversity has been lost as a result of breeding practices (Gioia et al., 2019). The contradictory conclusions may be related to differences in the marker systems and number of markers that were used in the studies; 24 SSR markers were used in the Gioia et al. (2019) study versus more than 4700 SNP markers in the current study. In addition, the number of individuals that were analyzed differed, with 192 advanced bean varieties plus 349 accessions of wild plus domesticated beans used in the Gioia et al. (2019) study versus 25 heirloom and 17 conventionally-bred dry bean genotypes in the present study. However, it is likely the

case that the difference is related to the fact that both the heirloom and conventional varieties used in the present study were selected materials (informally in the case of heirlooms and formally in the case of conventional varieties) that have both been subjected to a domestication bottleneck. Our results suggest that modern practices have not introduced another significant loss in genetic diversity.

### **3.5.2 Nitrogen fixation capacity in the Middle American gene pool exceeds that found in the Andean gene pool**

Although the range for nitrogen fixation among genotypes in the Middle American gene pool (Mist, 35.2 %Ndfa to Hi N, 76.3 %Ndfa) was narrower than in the Andean gene pool (Fisher, 19.9 %Ndfa to Coco Sophie, 75.7 %Ndfa), average nitrogen fixation among Middle American genotypes (average = 62.2 %Ndfa) was significantly higher than among the Andean genotypes (average = 54.8 %Ndfa). This suggests that the genes controlling nitrogen fixation capacity may differ between the gene pools, perhaps both in the number of loci and in their diversity. However, few studies exist that compare the nitrogen fixing capacities of Middle American genotypes with Andean genotypes. Ramaekers (2011) identified a few quantitative trait loci (QTL) associated with SNF-capacity using a recombinant inbred line (RIL) population created from a cross between an Andean and a Middle American genotype. Other studies have used sets of either Middle American or Andean genotypes. For example, Kamfwa et al. (2015) studied 259 genotypes belonging to the Andean Diversity Panel (Cichy et al., 2015), and a 188 F4:5 RIL population derived from two Andean parents (Kamfwa et al., 2019) and found a number of QTL associated with nitrogen fixation. Similar studies with Middle American

germplasm have identified similar as well as unique QTL associated with nitrogen fixation (Farid, 2015; Diaz et al., 2017; Heilig et al., 2017a; this thesis, chapter 5). Further research to identify QTL associated with nitrogen fixation in a panel comprised of genotypes from each gene pool followed by an assessment of haplotype diversity at the QTL would provide information on whether Middle American genotypes contain a greater number of active sites for N fixation than Andean genotypes or unique, more effective alleles. A higher level of genetic diversity was confirmed in the Middle American gene pool in this study, which may bring a higher number of diverse alleles and more genes which positively impact nitrogen fixation in the Middle American gene pool. Alternatively, the Middle American genotypes may have performed better with the Rhizobia inoculant and/or strains present in the soil.

### **3.5.3 Diversity for Nitrogen fixation in conventional bean genotypes similar to other studies**

Nitrogen fixation (%Ndfa) among the 18 conventional genotypes in the HCP (excluding R99) ranged from the lowest overall ranked AAFC-bred Limelight historic variety at 35.8% to the highest overall ranked University of Guelph breeding line Hi N at 66.9% (Table 3.2). These results fall generally within the range of %Ndfa reported for beans in contemporary research studies using conventional genotypes but other studies of nitrogen fixation, using conventional genotypes, have reported a broader range for this trait. For example Kamfwa et al. (2015) found a range from 3.6-98.2 %Ndfa in their study of the 259-genotype Andean Diversity Panel. A study with 79 Middle American genotypes under organic production (Heilig et al., 2017b) reported a range of 9.8 to 71.1

%Ndfa. Early studies of nitrogen fixation in bean (Graham and Rosas, 1977; Graham, 1981) reported that fixation varied according to plant architecture, where determinate bush types had poorer performance than indeterminate climbing types. Economically viable seed yields (1000-2000 kg ha<sup>-1</sup>) were not attainable when plant %Ndfa levels were low, although variation for nitrogen fixation was acknowledged (Bliss, 1993). Therefore, the 18 conventional genotypes in the HCP, spanning decades of variety releases by breeding programs across North America, likely represent the mid-range of nitrogen fixing capacity among conventional bean genotypes.

#### **3.5.4 Modern breeding has not reduced SNF capacity**

This study showed that despite decades of modern production and breeding practices, which include the use of nitrogen fertilizer that downregulates SNF activity, SNF capacity has not been lost from conventional genotypes. Recently released varieties such as Zorro, a black bean developed at Michigan State University (Kelly et al., 2009), and OAC Inferno, a light red kidney bean developed at the University of Guelph (Smith et al., 2012), showed good performance for nitrogen fixation in our study. OAC Inferno also performed well in a study examining SNF in the Andean Diversity Panel in Michigan (Kamfwa et al., 2015). The breeding methodologies used to develop Zorro and OAC Inferno are representative of modern breeding practices. Zorro was developed by pedigree and pure line selection from a backcross population generated from a bi-parental cross of Michigan State University black bean breeding lines (B00103 and X00822), with an emphasis on selection for disease resistance, plant architecture and yield. OAC Inferno was derived from a conical cross of diverse kidney bean variety

parentage (HR85-1885/Montcalm//USWA-39/AC Litekid///Foxfire/AC Elk//Sacramento/AC Calmont) sourced from across North America, using disease resistance and yield as selection criteria. Kamfwa et al. (2015) found that OAC Inferno was the only genotype in that study to contain major effect alleles for Ndfa at three loci. The complex pedigree of OAC Inferno may have contributed to its genetic diversity and higher than usual capacity for nitrogen fixation in this Andean genotype.

The finding that SNF in the heirloom category overall was not superior to the conventional category did not support the hypothesis on which the study was based and may be attributable to the composition of the HCP. The panel is small and was designed to include a broad representation of bean genotypes; the heirloom varieties come from wide geographic origins and are of unspecified breeding heritage (landraces and vintage varieties), and the modern genotypes include those released across recent decades as well as recent, elite modern varieties. Different results may have been achieved had the study included wild bean germplasm and landraces and more-recently registered modern varieties.

### **3.5.5 Incorporating heirloom genotypes into breeding for improved SNF holds potential**

Previous to this study, there was no indication that nitrogen fixation capacity would be superior in heirloom bean genotypes. The discovery of the diversity in capacity for nitrogen fixation among the 23 heirloom genotypes in the HCP [ranging from the lowest overall ranked genotype (Fisher at 32.1%) to the highest overall ranked (Coco Sophie at 69.0%, Table 3.2)] suggests that heirloom varieties may be an excellent

germplasm resource for studying this trait. Furthermore, we found a wide range in capacity for nitrogen fixation and yield performance among the heirloom genotypes of the HCP that was on par with conventional genotypes, indicating the suitability of heirloom beans for incorporation into breeding programs. In addition, the ranked panel for SNF performance (%Ndfa), was dominated by heirloom genotypes. Heirloom bean landraces are not routinely used to breed conventional varieties. For example, Navabi et al. (2014) undertook a pedigree analysis of Canadian dry bean varieties since the 1930s, and while a few introgressions of *P. coccinius* and *P. acutifolius* were made, the use of heirloom genotypes were not evident, except among the oldest crosses. Heirloom beans possess diversity that could be exploited without the challenges encountered when breeding with wild relatives, such as infertile crosses and reintroduction of 'wild' traits. In addition, heirloom varieties grown by First Nations groups for centuries in the Great Lakes region of North America are well-adapted to the climate and soils and perhaps the Rhizobium of this region.

Additionally, Coco Sophie (# 46), which is unique in the HCP for its admixture between the genepools and is representative of European bean germplasm (Gioia et al., 2013), might be used as a bridging parent to transfer desirable traits from one genepool to the other (Duc et al., 2015). In particular, because Coco Sophie already possesses good SNF capacity it could be useful to introgress SNF traits from higher-fixing Middle American germplasm to lower-fixing Andean germplasm.

The similar yield of heirloom and conventional categories indicates heirloom genotypes have breeding potential in modern programs. When genotypes of the HCP were compared based on breeding history, no significant difference was found in yield between heirloom (1651 kg ha<sup>-1</sup>) and conventional (1714 kg ha<sup>-1</sup>) groups. A number of explanations for the similar yield performance of heirloom and conventional genotypes in the present study are plausible. Firstly, heirloom varieties were sourced from commercial seed suppliers and the HCP may have been enriched in heirloom lines that had reasonable performance characteristics. Secondly, low soil nitrogen levels may have limited the yield performance of conventional genotypes, which have been bred to perform under intensive management regimes. And finally, the conventional genotypes were not chosen for the panel based on superior yield potential but on market class similarity to heirloom genotypes in the panel. Some of the conventional genotypes were registered as long ago as the 1940s, and yields in bean crops grown in Ontario have increased by 1000 kg ha<sup>-1</sup> in three decades (OMAFRA, 2016) and comparisons of bean varieties released over 40 years produced under conventional conditions show that breeding has increased their yield potential by more than 1% per year (A. Navabi, personal communication). Overall, the yield performance of the heirloom genotypes in our study would suggest that incorporation of these genotypes into a modern breeding program for organic production would not introduce significant yield drag. Singh et al. (2011) suggested that the use of well-adapted heirloom genotypes in bean breeding could be “crucial for developing high-yielding broadly adapted varieties for sustainable

organic and conventional production systems, thus reducing research and production costs.”

### **3.5.6 Heirloom beans may be particularly suited to breeding for organic agriculture**

The rise in demand for organic food has broadened societal interest in heirloom varieties. Heirloom genotypes may be inherently well suited to organic production practices where growing conditions share similarities with the environments in which First Nations peoples grew them (Singh et al., 2011). Heirloom beans often possess characteristics such as attractive seed coat colours and patterns, desirable texture and flavor, and heritage value, which increase their marketability and make them attractive to organic growers (Boyhan and Stone, 2016). Culinary characteristics were found to be of particular importance to heirloom bean growers in one study (Brouwer et al., 2016), while unique seed coat patterns as well as flavor and texture characteristics were emphasized by growers in another study (Swegarden et al., 2016).

Conventional varieties lack traits which give them a competitive advantage in low-input production systems and may hamper their yield performance. However, modern, conventionally-bred crop varieties account for more than 95% of varieties grown in organic production (Lammerts van Bueren et al., 2011). Direct comparisons of the yield performance of heirloom and conventional genotypes under organic production show mixed results. Miles et al. (2015) found that yield did not differ significantly between heirloom (1852 kg ha<sup>-1</sup>) and conventional (1983 kg ha<sup>-1</sup>) groups, whereas, Swegarden et al. (2016) found that heirloom genotypes (1362 kg ha<sup>-1</sup>) yielded

significantly less than the conventional genotypes (2447 kg ha<sup>-1</sup>). In an evaluation of a large panel of conventional black and navy bean genotypes under organic production the yields ranged from 1228-1762 kg ha<sup>-1</sup> (Heilig et al., 2017b), which is similar to the range found in the current study (1160-2002 kg ha<sup>-1</sup>) of heirloom and conventional genotypes under low nitrogen management.

In the present study, weed growth was difficult to manage, and lesions symptomatic of Common Bacterial Blight or Anthracnose were found on various genotypes (disease notes not recorded). Therefore, the development of genotypes exhibiting early canopy closure and disease resistance might be particularly advantageous for organic production systems. Studies in bean comparing the outcome of selection under organic and conventional growth conditions resulted in different genotypes being chosen based on yield performance (Singh et al., 2011). Similarly in soybean, Boyle (2016) found that selection performed under organic production favored genotypes with improved performance for resource acquisition traits (early canopy development, nodule mass, and root length).

### **3.6 Conclusion**

This study represents the first comparison of symbiotic nitrogen fixation in a panel of heirloom and conventional dry beans and will serve as a starting point for further research on promising heirloom genotypes. The finding that genetic diversity is similar between heirloom and conventional categories is consistent with the finding that %Ndfa in heirloom and conventional categories is not significantly different. This result does not

support the hypothesis that genetic diversity for nitrogen fixation has been eroded over years of modern breeding practices. The heirloom genotypes, as a group, had similar yield performance to the conventional genotypes under low-input field conditions, and although their capacity for nitrogen fixation was not significantly better than the conventional genotypes, they dominate the list of the best nitrogen fixers. Considering these characteristics, heirloom genotypes hold some promise for breeding to improve nitrogen fixation capacity in modern bean varieties. Heirloom beans represent an underutilized resource, which could be exploited to improve nitrogen fixation in breeding for organic production and conventional production where reduction of synthetic inputs and improved environmental stewardship are of growing concern.

## **4 Genetic Diversity, Nitrogen Fixation, and Water Use Efficiency in a Panel of Honduran Common Bean (*Phaseolus vulgaris* L.) Landraces and Modern Genotypes**

This chapter was published as:

Wilker, Jennifer; Humphries, Sally; Rosas-Sotomayor, Juan Carlos; Gómez Cerna, Marvin; Torkamaneh, Davoud; Edwards, Michelle; Navabi, Alireza; Pauls, K. Peter (2020). Genetic Diversity, Nitrogen Fixation, and Water Use Efficiency in a Panel of Honduran Common Bean (*Phaseolus vulgaris* L.) Landraces and Modern Genotypes. MDPI Plants, 2020-09-19, Vol.9 (9), DOI: 10.3390/plants9091238

The data generated for this study is publically available and can be found here: <https://doi.org/10.5683/SP2/DJB7VY>

Author contributions:

Jennifer Wilker, Alireza Navabi and K. Peter Pauls designed the project with input from Sally Humphries, Juan Carlos Rosas-Sotomayor and Marvin Gómez Cerna.

Jennifer Wilker performed the experiments and carried out the phenotypic analysis in Elora and Guelph.

Marvin Gómez Cerna facilitated the trial in Yorito, Honduras.

Davoud Torkamaneh carried out the genetic diversity analysis.

Jennifer Wilker and Michelle Edwards performed the statistical analyses.

Jennifer Wilker wrote the manuscript with comprehensive editing by K. Peter Pauls.

## 4.1 Abstract

Common bean (*Phaseolus vulgaris* L.) provides critical nutrition and a livelihood for millions of smallholder farmers worldwide. Beans engage in symbiotic nitrogen fixation (SNF) with Rhizobia. Honduran hillside farmers farm marginal land and utilize few production inputs; therefore, bean varieties with high SNF capacity and environmental resiliency would be of benefit to them. We explored the diversity for SNF, agronomic traits, and water use efficiency (WUE) among 70 Honduran landrace, participatory bred (PPB), and conventionally bred bean varieties (HON panel) and 6 North American check varieties in 3 low-N field trials in Ontario, Canada and Honduras. Genetic diversity was measured with a 6K single nucleotide polymorphism (SNP) array, and phenotyping for agronomic, SNF, and WUE traits was carried out. STRUCTURE analysis revealed two subpopulations with admixture between the subpopulations. Nucleotide diversity was greater in the landraces than the PPB varieties across the genome, and multiple genomic regions were identified where population genetic differentiation between the landraces and PPB varieties was evident. Significant differences were found between varieties and breeding categories for agronomic traits, SNF, and WUE. Landraces had above average SNF capacity, conventional varieties showed higher yields, and PPB varieties performed well for WUE. Varieties with the best SNF capacity could be used in further participatory breeding efforts.

## 4.2 Introduction

The common bean (*Phaseolus vulgaris* L.) is the most important food legume grown and consumed worldwide. High in protein, fiber, and essential nutrients, the nutritional profile and affordability of beans relative to other protein sources make beans a dietary staple in developing economies.

A member of the Fabaceae family, common bean is a predominantly self-pollinating species with a genome size of 587 Mbp and ploidy of  $2n = 2x = 22$  (Bitocchi et al., 2017). The center of origin for common bean is in present-day Central Mexico (Bitocchi et al., 2012). The species consists of two domesticated gene pools. Larger-seeded market classes evolved and were domesticated in South America and belong to the Andean gene pool, while smaller-seeded market classes evolved and were domesticated in Central America and belong to the Middle American gene pool (Gepts et al., 1986; Mamidi et al., 2011, 2013; Bitocchi et al., 2012, 2013; Schmutz et al., 2014).

Beans are traditionally grown as monocrops or with maize in either a relay cropping or an intercropping system in Honduras. There are two growing seasons—the rainy *Primera* (May–September) and the traditionally dry *Postre* (October–January)—although climate change causes more fluctuation in precipitation levels and the duration of these seasons. Food insecurity is an issue in hillside communities and is a particular problem during the summer months before the *Primera* harvest. In some locations, this hungry period is termed *los juniros* after the month during which food becomes scarce. More than 50% of bean production in Honduras takes place on steep hillside slopes

(15–30° and greater) (Bernsten, nd). In addition, the country's infrastructure is poor, and less than 30% of the bean growing area is located within an hour's travel to a road, restricting market access (Bernsten, nd).

Bean production in Honduras is affected by various biotic and abiotic stresses, and productivity is low and averaged 785 kg ha<sup>-1</sup> in 2018 compared to yields in Canada, which averaged 2888 kg ha<sup>-1</sup> in 2018 (FAO). Bean diseases and insect pests comprise the primary biotic stresses of Honduran bean production. The most impactful diseases are Bean Golden Mosaic Virus (BGMV), rust (caused by *Uromyces phaseoli*), web blight (caused by *Rhizoctonia solani*), anthracnose (caused by *Colletotrichum lindemuthianum*), and angular leaf spot (ALS; caused by *Pseudoscercospora griseola*) (Bernsten, nd). The whitefly (*Bemisia tabaci*) is the vector for BGMV and is the most important insect pest of bean in Honduras. Weevils (*Acanthoscelides obtectus* and *Zabrotes subfasciatus*) are serious pests of stored beans, reducing marketability and damaging seed for planting. Climate change is expected to affect the impact of these biotic stresses in bean production and may lead to a shift in the complex of pests and diseases involved (Beebe et al., 2011).

Extreme weather, such as high temperatures and flooding, including from hurricanes, reduces bean production. Climate models for Honduras predict higher temperatures and reduced overall rainfall but more extreme weather events increasing floods in the coming decades (World Bank Group, Climate Change Knowledge Portal, Honduras). Another abiotic stress impacting bean production is soil health. Soils across

Central America are deficient in available phosphorus (P), nitrogen (N), calcium (Ca), and organic matter, and aluminum (Al) and manganese (Mn) toxicity are exacerbated by low soil pH levels. Bean productivity is limited by soil nutrient availability, particularly N and P (Beebe, 2012). Nitrogen deficiency reduces grain yield because N is a structural component of various essential molecules, including chlorophyll, amino acids, nucleic acids, and lipids, required for the production of storage carbohydrates and proteins. Soils can be supplemented with nitrogen fertilizer throughout the growing season to avert yield losses; however, synthetic amendments are expensive, difficult to access, and generally not used by bean growers in Honduras. Instead, bean growers rely on organic forms of N, including that derived through symbiotic nitrogen fixation (SNF), as a nutrient source for their crop.

Beans are capable of generating their own organic nitrogen through SNF where nitrogen-fixing bacteria infect root nodules and reduce atmospheric nitrogen into forms useable by the host plant in exchange for carbohydrates (Lindstrom and Mousavi, 2020). SNF is a complex biological process and its efficiency is impacted by abiotic, biotic, and genetic factors, including soil nutrient levels, environmental conditions, the presence of efficient *Rhizobium* strains, and genetic variation in the host plant (Beebe, 2012). Recent studies have confirmed that SNF capacity in beans has a wide range and can reach high, yield-sustaining levels under optimal conditions (Farid and Navabi, 2015; Kamfwa et al., 2015; Diaz et al., 2017; Heilig et al., 2017b; Barbosa et al., 2018; Wilker et al., 2019; Aserse et al., 2020; Reinprecht et al., 2020; Sanyal et al., 2020).

Smallholder hillside farmers (0.5–5 ha) comprise approximately 70% of Honduran bean growers, and the remainder of production occurs in foothill and valley regions by larger-scale producers. Hillside farmers cultivate marginal land with steep slopes and low soil fertility, they tend towards subsistence production, and produce primarily for household consumption following traditional practices and planting traditional crop varieties. These smallholder farmers have limited access to markets, which has a two-fold impact, reducing the influence of market demands on what growers produce and limiting access to modern bean varieties and production inputs. These constraints notwithstanding, hillside farmers market approximately 50% of their bean harvest.

Landraces, known locally as *criollos*, comprise the majority of varieties traditionally grown by hillside farmers in Honduras (Humphries et al., 2005). Ninety-five percent of beans produced across the country are small, light red beans, which are preferred by Hondurans (Bernsten, nd), and are also exported to El Salvador and to the United States to meet the needs of Central Americans who have emigrated there. Some black beans are grown in Honduras and are primarily exported to neighboring Guatemala where that market class is favored (Bernsten, nd). Landraces have local genetic adaptation, high genetic diversity, and lack formal genetic improvement (Villa et al., 2005). The genetic heterogeneity of bean landraces lends resilience and makes them able to adapt to the changeable growing conditions of mountain hillsides and other marginal areas where they are grown. Among the preferred traits of the landraces included in this study are adaptation to cultivation at a range of altitudes, more

marketable seed coat color and appealing kinesthetic properties, and yield stability in a changeable climate. In addition to traditional landraces, hillside bean farmers also grow conventionally bred and participatory bred varieties.

Conventional bean breeding in Honduras has been primarily the responsibility of the *Programa de Investigaciones en Frijol (PIF)* in the Department of Agronomy at the *Escuela Agrícola Pan-Americana (Zamorano)* since the late 1980s when government funding to the agricultural research department (*Dirección de Ciencia y Tecnología Agropecuaria, DICTA*) was reduced restricting agricultural research and extension services (Mather et al., 2003). The International Center for Tropical Agriculture (CIAT), the Bean/Cowpea Collaborative Research Support Program (CRSP), and the Regional Cooperative Bean Program (*Programa Cooperativo Regional de Frijol-PROFRIJOL*) have also been involved in variety development and/or providing funding for bean research in Honduras. Zamorano's early breeding focus was on developing conventional varieties with BGMV resistance and improved heat and drought tolerance for the lowland and valley production regions of Honduras (Mather et al., 2003). By the late 1990s, Zamorano took leadership in bean breeding for the region, developing small red varieties for Honduras, Guatemala, El Salvador, Nicaragua, and Costa Rica, as well as black bean varieties for Guatemala and Haiti (Reyes, 2012). Conventionally bred commercial varieties are adapted for cultivation across a wide geographic region and have disease resistance and agronomic traits, which can bolster their yields. Adoption of commercial varieties among hillside growers is limited for a number of reasons, including darker seed coat color and other culinary traits, which reduce their

marketability, as well as poor yield performance compared to landraces (Humphries et al., 2005). In the mid-1990s, Zamorano embarked on collaborative research with social scientists from CIAT to explore the social and economic factors that impact adoption of conventional varieties (Humphries et al., 2005). As a result, participatory research has become an important method used in the Zamorano bean breeding program.

The term participatory plant breeding encompasses two main methods of plant variety development, 'participatory varietal selection' and 'participatory plant breeding'. Participatory varietal selection (PVS) involves farmers locally testing varieties or advanced breeding lines provided by a formal plant breeder and making selections based on their needs (Ceccarelli et al., 2000). Participatory plant breeding (PPB) involves farmers locally testing early stage (F<sub>2</sub>-F<sub>3</sub>) breeding material and can further involve the farmers actively participating in choosing parents and driving variety development by selecting progeny that meet local needs and preferences. As in other reports on the subject, the term PPB will be used to refer to both PPB and PVS in this study (Almekinders and Elings, 2001; Humphries et al., 2015).

Participatory bean breeding at Zamorano has been facilitated through collaboration with CIAT-initiated *comités de investigación agrícola local* (CIALs), which are village-level farmer research teams that create a space where applied agricultural research can be carried out. For this study, we collaborated with the *Fundación para la Investigación Participativa con Agricultores de Honduras* (FIPAH) and *Programa de Reconstrucción Rural* (PRR). FIPAH supports over 100 CIALs, backstopped by regional

offices across the country (<https://fipah-hn.org/>). PRR is an NGO that works with smallholder farmers in Santa Barbara and Lempira and supports approximately 60 CIALs (<https://www.iaf.gov/grants/honduras/2017-prr/>) (Gomez et al., 2020). CIAL members are trained in the scientific method, and most CIALs focus their research on obtaining higher-yielding and climate-resilient corn and bean varieties. The relationship between the CIALs and the bean breeding program at Zamorano is collegial and formal, responding to the needs of the farmers while the research is carried out with scientific rigor (Humphries et al., 2005). Traits of interest to the farmers are emphasized, and trials are performed using statistically valid designs and research methods. Landraces, conventional varieties, and germplasm from across the region are used in PPB efforts. PPB generates varieties that combine the local adaptation of landraces with improved traits from conventional genotypes such as disease resistance and higher yields. Other traits that factor into selection by farmer-researchers include seed color, appearance and size, pod length, plant architecture, even ripening, early maturity, and various culinary qualities (Humphries et al., 2005). Zamorano has developed some PPB varieties using landraces as parents in Honduras, Costa Rica, and Nicaragua (Rosas, 2013). Between 1994 and 2015, 24 PPB varieties were developed by Zamorano in collaboration with CIAL groups using participatory research methods, and one of these varieties, 'PM2-Don Rey', has been supported for national registration (Humphries et al., 2015). Adoption of the PPBs among CIAL members is above 60%, and PPBs are gaining wide acceptance among other farmers in communities where participatory research is carried out (Humphries et al., 2015). Extensive discussion of the

development of ‘Macuzalito’, ‘Cedron’, ‘Amilcar’, ‘Esperanceño’, ‘Chepe’, and ‘PM2-Don Rey’ (representing both PPV and PPB methods of variety development) can be found in Humphries et al. (Humphries et al., 2015).

Due to limited production resources and the threat of climate change, farmers in remote hillside communities would benefit from growing high-yielding common bean varieties that are climate resilient and have high nitrogen fixing capacity. To examine Honduran bean germplasm for these traits of interest, we curated a panel of Honduran bean genotypes representative of the traditional landraces and the participatory bred varieties grown by hillside bean farmers, as well as Honduran conventional and North American checks. The current study tests the hypothesis that bean landraces are a good source of germplasm with a high capacity for nitrogen fixation. The objectives of this study were to determine the genotypic and phenotypic diversity of the Honduran panel and to identify germplasm sources for breeding improved varieties suited to hillside production in Honduras.

## **4.3 Materials and Methods**

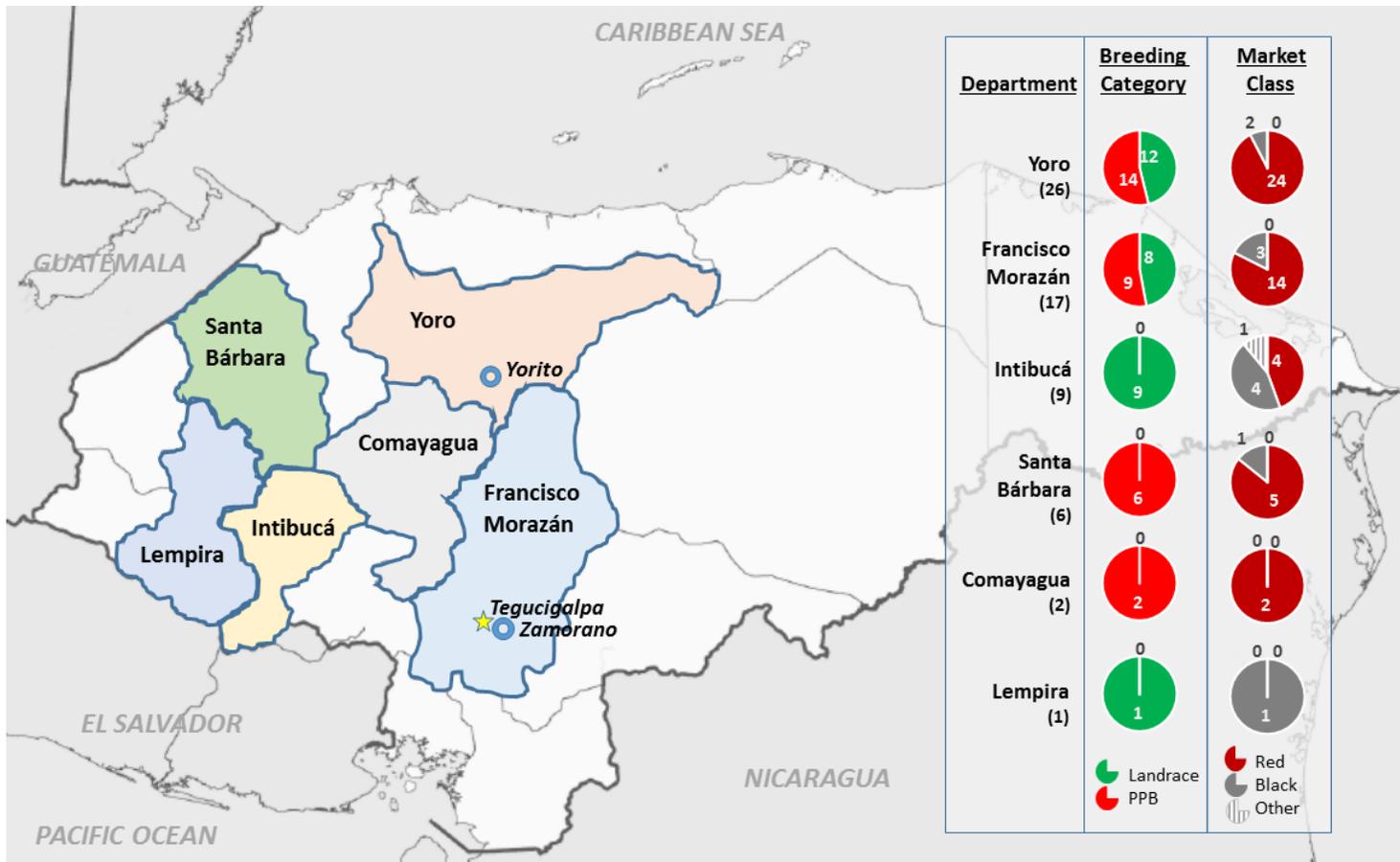
### **4.3.1 Plant material**

The Honduran panel was assembled in 2014 at the University of Guelph in collaboration with agronomists at FIPAH. The initial panel contained 27 landraces, 26 PPB varieties, and 5 Honduran conventional checks provided by FIPAH, as well as 6 North American checks sourced from the University of Guelph bean breeding program.

The landraces consisted of traditional inbred varieties unimproved by modern plant breeding, which are grown by subsistence farmers in hillside communities. The PPB varieties were generated either through participatory varietal selection (PVS) or participatory plant breeding (PPB) through a collaboration between the bean breeding program at Zamorano and CIALs associated with FIPAH. The landraces and PPB varieties were sourced by M. Gomez (FIPAH) from six departments in west-central Honduras. The landraces were from Yoro, Intibucá, Francisco Morazán, and Lempira, and the PPB varieties were from Yoro, Francisco Morazán, Santa Bárbara, and Comayagua (Figure 4.1). Seed was either collected directly from farmers in their communities or sourced from central seed banks maintained by FIPAH and PRR. The five Honduran conventional checks were developed for production in lower to mid-altitude, hillside and valley commercial-production regions of the country. The six North American varieties consisted of Merlot and OAC Rosito as small red market class checks, Zorro as a black market class check, and three navy beans: OAC Mist, a high-nitrogen-fixing genotype, R99, a non-fixing mutant, and its parent line OAC Rico. All genotypes in the panel belong to race Mesoamerica (Singh et al., 1991).

In the first trial location (Elora 2014), 10 Honduran genotypes were found to be daylight sensitive and were not grown at the subsequent locations. Additionally, seed of 16 genotypes that exhibited uneven maturity in Elora 2014 were sent to Puerto Rico for seed increase over the winter of 2015. For the second trial location (Yorito, 2014–2015), 12 new genotypes (6 landraces, 5 improved, and 1 Honduran conventional check) were added to the panel. For the third trial location (Elora 2015), seed harvested from Elora

2014, from the Puerto Rican seed increase, and from Yorito was used, as available. An additional Honduran conventional variety was grown that year to fill in the experimental design. Overall, a total of 77 genotypes were tested in the Honduran panel, 50 genotypes of which were grown at all 3 locations. A summary of the genotypes included in the Honduran Panel, including trial year, market class, seed source, and pedigree information is provided in Tables 4.1-4.3 according to breeding history.



**Figure 4.1.** Map of west–central Honduras, outlining the six departments from which landrace and participatory bred (PPB) bean genotypes were sourced for the Honduran Panel. The chart at the right describes the number of landraces and PPB genotypes obtained from each department and the market classes to which those genotypes belong. The location of Yorito where the Honduran field trial was carried out, Zamorano where the Escuela Agrícola Panamericana is located, as well as the capital of Honduras, Tegucigalpa, are shown.

**Table 4.1.** Genotypes of the North American check and Honduran conventional breeding categories, which were included in the HON panel tested at 3 field locations, 2014-15. The trials in which each genotype was included and whether the entry was SNP genotyped are indicated. Breeding category, market class, genealogy and origin, are provided where available.

HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Market class	Genealogy	Institution or organization	Seed origin	Notes
59	OAC Rico	x		x	X	Check	Navy	(Ex Rico 23 / Narda)/Ex Rico 23 See (Beversdorf, 1984)	University of Guelph		Resistant to BCMV and Anthracnose. Tolerant to white mold. Unremarkable SNF capacity. In other studies it fixed approximately 53% of N (chapter 3 - Wilker et al., 2019).
60	R99 <sup>1</sup>	x	x	x	X	Check	Navy	See (Park and Buttery, 2006)	Agriculture Agri-Food Canada (AAFC)		Non-nodulating experimental line. Derived from OAC Rico through ethyl methanesulphonate (EMS) mutagenesis (Park and Buttery, 1992). Used in natural abundance method to establish a baseline nitrogen level in seed measured for <sup>14</sup> N and <sup>15</sup> N accumulation.
61	OAC Mist	x	x	x	X	Check	Navy	See (Khanal et al., 2017a)	University of Guelph		High-yielding, late season. Resistant to BCMV and CBB. Generally high SNF capacity. Farid and Navabi (2015) reported that OAC Mist fixed as much as 78.5% of N (Farid and Navabi, 2015). Wilker et al. (chapter 3 - Wilker et al., 2019) reported that OAC Mist fixed an average of 55% N.
62	Merlot	x	x	x	X	Check	Small red	See (Hosfield et al., 2004)	United States Department of Agriculture – Agriculture Research Service (USDA-ARS)		Intense red seed colour. Larger seed size than Honduran beans (mean HSW 39.2 g). Resistant to rust, BCMV and BCMNV. Susceptible to anthracnose. Moderate SNF capacity. Wilker (chapter 5) found Merlot fixed as much as 64.9% of its N.
63	OAC Rosito	x	x	x	X	Check	Small red	See (Aguilera et al., 2019)	University of Guelph		Developed from a diverse landrace originating in El Salvador (Aguilera et al., 2019). Dark red seed colour. Similar seed size to Honduran beans (mean HSW 21.7 g). Resistant to BCMV. Susceptible to Anthracnose and CBB. The SNF capacity of OAC Rosito has not been examined previous to the current study.

											Seed origin	
HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Market class	Genealogy		Institution or organization	Notes	
64	Zorro	x	x	x	X	Check	Black	See (Kelly et al., 2009)		Michigan State University	Resistant to rust and anthracnose, and is less affected by white mold. Moderate SNF capacity. Wilker et al. (chapter 3 - Wilker et al., 2019) reported that Zorro fixed an average of 59%, and Wilker et al (chapter 5) found that Zorro fixed as much as 46.9% of its N.	
80	CENTA San Andrés <sup>3</sup>			x	X	Conventional	Small red	EAP 9510-77, [MD 30-75 / DICTA 105]		PIF/Zamorano, UPR, CENTA, EI Salvador; 2003	Same breeding line as Amadeus-77 (Rosas et al., 2004). Resistant to BGYMV and BCMV. Heat tolerant and adapted for production in lower-altitude coastal areas (Rosas et al., 2004).	
52	Amadeus-77 <sup>3</sup>	x	x	x	X	Conventional	Small red	EAP 9510-77, [MD 30-75 / DICTA 105]		PIF/Zamorano, UPR, DICTA, Honduras; 2003	Same breeding line as CENTA San Andres (Rosas et al., 2004). Resistant to BGYMV and BCMV. Heat tolerant and performs well in low altitude, coastal areas. Widely adopted across Central America, and in 2010 accounted for around 50% of commercial production in the region (Reyes, 2012).	
56	Dorado <sup>3</sup>	x	x	x	X	Conventional	Small red	DOR 364, [BAT 1215 x (RAB 166 x DOR 125)]		Profrijol, DICTA, Honduras; 1990	Also known as 'DOR 364' Resistant BGYMV and BCMV (Rosas et al., 2004). Yields well across environments and has mid-range maturity, however it has a dark red seed coat (Rosas et al., 2004).	
53	DEORHO <sup>2,3</sup>	x	x	x		Conventional	Small red	SRC 2-18-1, [Milenio / MD 30-75]		PIF/Zamorano, UPR, DICTA, Honduras; 2007	Also known as 'DEHORO' and 'INTA Matagalpa'. Resistant to BGYMV and BCMV. Higher yielding and desirable light red seed coat colour (Jamora and Maredia, 2010). Popular with Honduran growers accounting for 23% of the red bean acreage in 2010 (Reyes, 2012). DEORHO was not grown for DNA extraction and consequently was not included in the genetic analyses carried out for this study.	
55	Tio Canela 75 <sup>2,3</sup>	x	x	x	X	Conventional	Small red	MD 30-75, [DOR 483 // DOR 391 / Pompadour J]		PIF/Zamorano, UPR, Honduras; 1996	Resistant to BGYMV and BCMV (Rosas et al., 2004). Yields well across environments, has mid-range maturity, and has a shiny red seed (Rosas et al., 2004). Tio Canela 75 is a parent line of Amadeus-77 and Carrizalito.	

											Seed origin	
HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Market class	Genealogy	Institution or organization	Notes		
77	Carrizalito <sup>3</sup>		x	x	x	Conventional	Small red	EAP 9510-1, [MD 30-75 / DICTA 105]	PIF/Zamorano, UPR, DICTA, Honduras; 2003	Resistant to BGYMV and BCMV. Early maturity (68-70 DAP) and upright plant architecture. High-yielding variety, adapted to mid-altitude production (Jamora and Maredia, 2010).		
54	Aifi Wuriti <sup>2</sup>	x	x	x	X	Conventional	Black	EAP 9712-13, Tio Canela 75/DICTA 105/BG12WB12//Tio Canela 75/DICTA 105/AL12	PIF/Zamorano, UPR, SNS, Haiti; 2008	Also known as 'Negro Olfirwit'. Resistant to BGYMV, BCMV, is tolerant of low soil fertility and is early-maturing (Beaver et al., 2003). Popular in Haiti and the Dominican Republic, and was successfully adopted by growers in southeast Guatemala (Beaver et al., 2003).		

<sup>1</sup> R99 was genotyped but not included in the genetic analyses

<sup>2</sup> genotypes exhibiting uneven maturity at Elora 2014 and sent to Puerto Rico for seed increase in winter 2015

<sup>3</sup> varieties developed using conventional breeding methods according to J.C. Rosas, pers. comm.

**Table 4.2.** Genotypes of the Landrace breeding category, which were included in the HON panel tested at 3 field locations, 2014-15. The trials in which each genotype was included and whether the entry was SNP genotyped are indicated. Breeding category, market class, and origin details, are provided where available.

HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Seed origin				
							Market class	Institution, farmer or organization	Locality	Municipality	Department
02	Concha Rosada <sup>2</sup>	x	x	x	x	Landrace	Small red	FIPAH	Yorito	Yorito	Yoro
06	Negro Pedreño	x	x	x	x	Landrace	Black	Odir Palma	La Esperanza	Yorito	Yoro
07	Negro Concha Blanca <sup>2</sup>	x	x	x	x	Landrace	Black	Odir Palma	La Esperanza	Yorito	Yoro
08	Balin Rojo <sup>2</sup>	x	x	x	x	Landrace	Small red	Edy Hernandez	La Patastera	Yorito	Yoro
09	Carmelita <sup>1</sup>	x			x	Landrace	Small red	Francisco Murillo	La Esperanza	Yorito	Yoro
11	Estica	x		x	x	Landrace	Small red	Irene Hernandez	La Esperanza	Yorito	Yoro
22	Rosado <sup>2</sup>	x	x	x	x	Landrace	Small red	Odir Palma	La Esperanza	Yorito	Yoro
27	Chapin Rojo	x	x	x	x	Landrace	Small red	Daniel Vargas	El Injerto	Comayagua	Francisco Morazán
29	Uva	x	x	x		Landrace	Black	Alonso Gutierrez	San Jose	Vallecillo	Francisco Morazán
30	Chapin Negro <sup>2</sup>	x	x	x	x	Landrace	Black	Ovidio Valeriano	Nocoro	Vallecillo	Francisco Morazán
34	Vaina Rosada	x	x	x	x	Landrace	Small red	CIAL San Jose	San Jose	Vallecillo	Francisco Morazán
36	Milpero Negro <sup>1</sup>	x			x	Landrace	Black	Bertilio Antonio Rodriguez	San Pedrito	Opalaca	Intibucá
37	Negro Vaina Blanca	x	x	x	x	Landrace	Small red	Carmen Azucenaa Giron	Guayabal	Jesus de Otoro	Intibucá
38	Mano de Piedra	x	x	x	x	Landrace	Small red	Maria Laines	Maye	Jesus de Otoro	Intibucá
39	Milpero Rojo <sup>1</sup>	x			x	Landrace	Small red	Bertilio Antonio Rodriguez	San Pedrito	Opalaca	Intibucá
40	Vaina Blanca	x	x	x	x	Landrace	Black	Tiburcio Dias	Pueblo Viejo	NA	NA
41	Negro Arbolito	x	x	x	x	Landrace	Black	Armando Pineda	Crucita Oriente	Jesus de Otoro	Intibucá
42	Negro Cuarenteño	x	x	x	x	Landrace	Black	Maria Laines	Maye	Jesus de Otoro	Intibucá
43	Negro <sup>2</sup>	x	x	x	x	Landrace	Black	Antonio Espinosa	Iguala	Lempira	Lempira

Seed origin												
HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Market class	Institution, farmer or organization	Locality	Municipality	Department	
44	Milpero Gatiador <sup>1</sup>	x			x	Landrace		Carioca	Evelino Sanchez	La Vegas	NA	NA
45	Ponga la Olla	x	x	x	x	Landrace		Black	Estalin Diaz	Pueblo Viejo	NA	NA
46	Madura Parejo	x	x	x	x	Landrace		Small red	Maria Lainez	Maye	Jesus de Otoro	Intibucá
47	Milpero Blanco <sup>1</sup>	x			x	Landrace		White	Maria Juana Gutierrez	Monte Verde	Opalaca	Intibucá
48	Cincuentaño	x	x	x	x	Landrace		Small red	FIPAH	Yorito	Yorito	Yoro
49	Paraísito	x	x	x	x	Landrace		Small red	FIPAH	Yorito	Yorito	Yoro
50	Rojo de Seda	x	x	x	x	Landrace		Small red	FIPAH	Yorito	Yorito	Yoro
51	Marciano	x	x	x	x	Landrace		Small red	FIPAH	Yorito	Yorito	Yoro
65	Olanchano Negro		x	x	x	Landrace		Black	CIAL San Jose de la Mora	San Jose de la Mora	Vallecillo	Francisco Morazán
66	Seda-Vallecillo		x	x	x	Landrace		Small red	CIAL San José de la Mora	San Jose de La Mora	Vallecillo	Francisco Morazán
67	Chirineño		x	x	x	Landrace		Small red	CIAL Chirinos	Chirinos	Cedros	Francisco Morazán
68	Roseño		x	x	x	Landrace		Small red	Adan Bustillo	La Fortuna	Victoria	Yoro
70	Negro Opalaca		x	x	x	Landrace		Black	NA	Monte Verde	San Francisco de Opalaca	Intibucá
82	Olanchano Rojo		x			Landrace		Small red	San Jose de la Mora	San Jose de la Mora	Vallecillo	Francisco Morazán

<sup>1</sup> genotypes exhibiting day-light sensitivity at Elora 2014

<sup>2</sup> genotypes exhibiting uneven maturity at Elora 2014 and sent to Puerto Rico for seed increase in winter 2015

<sup>NA</sup> information not available

**Table 4.3.** Genotypes of PPB breeding category, which were included in the HON panel tested at 3 field locations, 2014-15. The trials in which each genotype was included and whether the entry was SNP genotyped are indicated. Breeding category, market class, and origin details, are provided where available.

HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Market class	Genealogy	Seed origin			
									Institution, farmer or organization	Locality	Municipality	Department
01	Macuzalito	x	x	x	x	PPB (PPB) <sup>3</sup>	Small red	PPB-9911-44-5-13M, [Concha Rosada // SRC 1-1-18 / SRC 1-12-1]	PIF/Zamorano, FIPAH, Honduras; 2004	Yorito	Yorito	Yoro
13	FPY-722-53 <sup>1</sup>	x			x	PPB (PPB)	Small red	FPY-722-53, Tio Canela 75/ Estica	PIF/Zamorano, FIPAH, CIAL Santa Cruz	Santa Cruz	Yorito	Yoro
14	FPY-722-38	x	x	x	x	PPB (PPB)	Small red	FPY-722-38, Tio Canela 75/Estica	PIF/Zamorano, FIPAH, CIAL Santa Cruz	Santa Cruz	Yorito	Yoro
15	FPY-722-13 <sup>1</sup>	x	x		x	PPB (PPB)	Small red	FPY-722-13, Tio Canela 75/Estica	PIF/Zamorano, FIPAH, CIAL Santa Cruz	Santa Cruz	Yorito	Yoro
16	FPY-724-43 <sup>1</sup>	x	x	x	x	PPB (PPB)	Small red	FPY-724-43, Macuzalito/Estica	PIF/Zamorano, FIPAH, CIAL Santa Cruz	Santa Cruz	Yorito	Yoro
18	FPY-721-16 <sup>1</sup>	x			x	PPB (PPB)	Small red	FPY-721-16, Amadeus 77/Estica	PIF/Zamorano, FIPAH, CIAL Santa Cruz	Yorito	Yorito	Yoro
19	FPY-722-41 <sup>2</sup>	x	x	x	x	PPB (PPB)	Small red	FPY-722-41, Tio Canela 75/Estica	PIF/Zamorano, FIPAH, CIAL Santa Cruz, Amilcar Orellana	La Esperanza	Yorito	Yoro
23	PM2-Don Rey <sup>2</sup>	x	x	x	x	PPB (PPB)	Small red	IBC-302-29, Carrizalito//Carrizalito/Paraíso	PIF/Zamorano, UPR, DICTA, ASOCIAL Vallecillo, Reinaldo Funez; 2014	San Isidro	Vallecillo	Francisco Morazán
25	FPV-921-4	x	x	x	x	PPB (PPB)	Small red	FPV-921-4, Vaina Rosada/Amadeus 77	CIAL San Isidro	San Isidro	Vallecillo	Francisco Morazán

HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Market class	Genealogy	Seed origin			
									Institution, farmer or organization	Locality	Municipality	Department
26	Quebradeño	x	x		x	PPB (PPB)	Small red	IBC-307-7, [TC75//TC75/Cincuentaño]	CIAL Quebrada	Trinidad de Quebrada	Vallecillo	Francisco Morazán
28	FPV-921-61 <sup>1</sup>	x				PPB (PPB)	Small red	FPV-921-61, Vaina Rosada/Amadeus 77	CIAL San Isidro	San Isidro	Vallecillo	Francisco Morazán
31	FPV-923-25 <sup>2</sup>	x	x	x	x	PPB (PPB)	Small red	FPV-923-25, Vaina Rosada/Conan 33	CIAL San Isidro	San Isidro	Vallecillo	Francisco Morazán
32	FPV-923-21 <sup>1</sup>	x				PPB (PPB)	Small red	FPV-923-21, Vaina Rosada/Conan 33	CIAL San Isidro	San Isidro	Vallecillo	Francisco Morazán
33	FPV 921-65 <sup>1</sup>	x				PPB (PPB)	Small red	FPV 921-65, Vaina Rosada/Amadeus 77	CIAL San Isidro	San Isidro	Vallecillo	Francisco Morazán
03	Cedron	x	x	x	x	PPB (PVS) <sup>4</sup>	Small red	EAP 9508-93, [Bribri // MD 30-37//PR 9177-214-1/Tio Canela 75]	PIF/Zamorano, FIPAH, Honduras; 2005	Yorito	Yorito	Yoro
04	Chepe	x	x	x	x	PPB (PVS)	Small red	703-SM15216-11-5	PIF/Zamorano, CIAT, FIPAH, Honduras; 2012	Yorito	Yorito	Yoro
05	Amilcar <sup>2</sup>	x	x	x	x	PPB (PVS)	Small red	IBC 308-24, Amadeus 77//Amadeus 77/Cincuentaño	PIF/Zamorano, FIPAH, Honduras; 2012	Yorito	Yorito	Yoro
10	Marcelino <sup>2</sup>	x	x	x	x	PPB (PVS)	Small red	EAP 9508-41, Bribri/MD 30-37//PR 9177-214-1/Tio Canela 75	PIF/Zamorano, FIPAH, Edy Hernandez; 2012	La Patastera	Yorito	Yoro
12	Esperanceño <sup>2</sup>	x	x	x	x	PPB (PVS)	Small red	PR 0310-26-3-3, VAX 6	PIF/Zamorano, UPR, CIAT, FIPAH, Pablo Orellana; 2011	La Esperanza	Yorito	Yoro
17	Paisano PF	x	x	x	x	PPB (PVS)	Small red	MER-2212-28, Milenio/Amadeus 77	PIF/Zamorano, PRR, CIAL Palmicha Fatima; 2011	Palmichal Fatima	Siguatopeque	Comayagua
20	523-DFBS 15092-04-4	x	x	x	x	PPB (PVS)	Small red	523-DFBS 15092-04-4	PIF/Zamorano, CIAT, FIPAH	Yorito	Yorito	Yoro

HON Entry	Genotype	Elora '14	Yorito '14-'15	Elora '15	SNP genotyped	Breeding category	Market class	Genealogy	Seed origin			
									Institution, farmer or organization	Locality	Municipality	Department
21	519-DFBZ 15094-39-4 <sup>2</sup>	x	x	x	x	PPB (PVS)	Small red	519-DFBZ 15094-39-4	PIF/Zamorano, CIAT, FIPAH	Yorito	Yorito	Yoro
24	Conan 33	x	x	x	x	PPB (PVS)	Small red	PRF-9659-25B-1, [EAP 9503 / RS3 // Bribri / MD 30-37 //// EAP 9503 / RS3 // A429 / K2 /// V8025 / XR 16492 // APN83 / CNC]	PIF/Zamorano, FIPAH, Asocial Vallecillo; 2005	Trinidad de Quebrada	Vallecillo	Francisco Morazán
35	San Jose	x	x	x	x	PPB (PVS)	Small red	X0-233-171-4, VAX 3	PIF/Zamorano, UPR, CIAT, FIPAH, CIAL San José; no date	San Jose	Vallecillo	Francisco Morazán
57	Campechano	x	x	x	x	PPB (PVS)	Small red	SX14825-7-1	PIF/Zamorano, CIAT, PRR, ASOCIALAYO; 2012	La Buena Fe	Zacapa	Santa Bárbara
58	Don Kike	x	x	x	x	PPB (PVS)	Small red	MDSX14797-6-1	PIF/Zamorano, CIAT, PRR, ASOCIALAYO; 2012	La Buena Fe	Zacapa	Santa Bárbara
72	Arbolito Negro		x	x	x	PPB (PVS)	Black	SJC 729-89, Negro Vaina Blanca/PRF 9924-50N	PIF/Zamorano, PRR, ASOCIALAYO; 2013	La Buena Fe	Zacapa	Santa Bárbara
74	Rojo Delicia		x	x	x	PPB (PVS)	Small red	703-SM15216-11-4-VR	PIF/Zamorano, CIAT, PRR, ASOCIALAYO; 2015	La Buena Fe	Zacapa	Santa Bárbara
75	Don Cristóbal		x	x	x	PPB (PVS)	Small red	SRC1-12-1-8, [DOR476//XAN155/DOR364]	PIF/Zamorano, CIAT, PRR, CIAL Laguna Seca; 2015	Laguna Seca	Taulabe	Comayagua
76	Victoria		x	x	x	PPB (PVS)	Small red	SRS 56-3, [Amadeus77/SEA5]	PIF/Zamorano, PRR, CIAL Nueva Esperanza; 2015	Nueva Esperanza	Concepción Sur	Santa Bárbara
78	Nueva Esperanza		x	x	x	PPB (PVS)	Small red	DICTA 9801, UPR9606-2-2/MD 30-37	PIF/Zamorano, PRR, CIAL Nueva Esperanza; 2005	Nueva Esperanza	Concepción Sur	Santa Bárbara

<sup>1</sup> genotypes exhibiting day-light sensitivity at Elora 2014

<sup>2</sup> genotypes exhibiting uneven maturity at Elora 2014 and sent to Puerto Rico for seed increase in winter 2015

<sup>3</sup> PPB are participatory plant bred varieties derived from a cross between a landrace and a breeding line; classified by FIPAH

<sup>4</sup> PVS are varieties derived from participatory varietal selection where breeding lines are selected through generations of testing by CIALs; classified by FIPAH

### 4.3.2 Field experimental design and maintenance

Field trials were carried out at the University of Guelph Elora research station (ERS) in summer 2014 and 2015 and at Yorito, Honduras in the *Postrera* season (planted December, 2014).

#### 4.3.2.1 Elora

The Elora fields were selected based on low soil nitrogen levels as measured by preplant soil tests and by site crop rotation histories that indicated that no dry bean crops had been produced in those fields for the previous decade at a minimum. In 2014, nitrates ( $\text{NO}_3^-$ ) were low at 7.1 ppm, and ammonium ( $\text{NH}_4^+$ ) levels were 3.2 ppm. In 2015, nitrates ( $\text{NO}_3^-$ ) were very low at 4.8 ppm, and ammonium ( $\text{NH}_4^+$ ) levels were 4.5 ppm.

A square lattice design (8 × 8) with two replications was used for each trial. At the ERS in 2014, 135 seeds of each genotype were grown in 4-row plots (150 cm × 190 cm, 37.5 cm between rows) with approximately 6 cm between plants. At the ERS in 2015, 60 seeds of each genotype were grown in 4-row plots (150 cm × 90 cm, 37.5 cm between rows) with approximately 5 cm between plants within rows.

Methods followed for seed inoculation and field trial maintenance were the same as described in Chapter 3. Briefly, bean seed was inoculated with Nodulator® (Becker-Underwood) prior to planting (1/4 teaspoon, approximately 0.4 g in 2014; 1/8 teaspoon, approximately 0.2 g in 2015). Successful inoculation was confirmed each year by

observing pink (active) nodules on a few plants chosen at random throughout the trial at flowering time.

At the ERS, plots were maintained with standard practices throughout the growing season, except no-nitrogen fertilizer was used. Pre-plant fertilizer (0:20:20) at a rate of 200 kg ha<sup>-1</sup> was applied approximately one week prior to planting. Pre-plant herbicide was applied to control broadleaf weeds, and pesticides were applied as needed throughout the growing season at standard rates to control leafhoppers, anthracnose, and root rot (see Chapter 3 for details). Plots were manually weeded once before canopy closure each year.

At Elora 2014, the harvest was staggered according to maturity. The plots were pulled by hand at maturity and threshed at the side of the field using a Wintersteiger plot combine (Wintersteiger AG, Upper Austria, Austria) with a Classic Seed-Gauge weighing system by Harvest-Master (Juniper Systems Inc., Utah, USA), and plot seed weight and moisture content were recorded. Plots that did not reach reproductive/physiological maturity were not harvested. In 2015, plot harvest took place after all plots reached maturity with an SPC20 Almaco plot combine (ALMACO, Nevada, IO, USA), which automatically recorded moisture and weight (kg ha<sup>-1</sup>) for each plot at 13% moisture.

#### **4.3.2.2 Yorito**

The Yorito trial site was chosen based on field uniformity, access to irrigation, and proximity to the FIPAH regional office. Soil NO<sub>3</sub><sup>-</sup> levels were 18 ppm (“moderate” to

“high”) at Yorito, and the field had been used for bean and maize production in previous seasons. The FIPAH agronomist, M. Gomez, indicated that the trial site conditions were representative of the bean production areas around Yorito.

Seed of each North American check variety were provided by the University of Guelph, and seed of all Honduran bean genotypes in the trial were sourced in Honduras. A square lattice design (8 × 8) with two replications was used for the trial. At Yorito, 100 seeds of each genotype were grown in 2-row plots (100 cm × 500 cm, 50 cm between rows) with approximately 30 cm between plants and 3 seeds sown per hole, as per the traditional planting system.

Inoculant for the trial, a mixture of *Rhizobium etli* (CIAT 632) and *R. tropici* (CIAT 899) strains, was provided by J.C. Rosas (EAP-Zamorano) and was applied according to a protocol provided by PIF at a rate of 500 g ha<sup>-1</sup> (Rosas and Vargas, nd). Briefly, the peat-based inoculant powder was applied to slightly moistened seed to ensure it adhered well to the seeds, and once sown, the seeds were covered with soil to protect the inoculant from the sun. CIAT 899 has high symbiotic stability and efficient N-fixation characteristics (Diaz et al., 2017).

Plots were maintained with standard practices throughout the growing season. A preplant fertilizer of 12:24:12 NPK was used at a rate of 64.81 kg ha<sup>-1</sup> as a formulation without N was not available in Honduras. Carbendazim (DEROSAL, Bayer) at a rate of 400 mL ha<sup>-1</sup> was used to protect the trial from angular leaf spot (*Pseudoscercospora*

*griseola*), common bacterial blight (*Xanthomonas axonopodis* pv. *phaseoli*), and rust (*Uromyces appendiculatus*).

At Yorito, plots were harvested by hand over a number of days as each plot matured. Yield ( $\text{kg ha}^{-1}$ ) was measured at harvest, the two plots of each genotype were bulked, and a subsample of seed was sent to Guelph for further analyses.

### **4.3.3 Phenotyping**

#### **4.3.3.1 Elora**

As plots initiated the reproductive phase, DTF was recorded as the date when 50% of the plants in a plot had one flower open. Relative leaf chlorophyll content was measured twice during the growing season (when the mean number of plots had reached (1) the second trifoliate stage and (2) at 100% flowering) using an SPAD 502 Plus Chlorophyll Meter (Konica-Minolta). The meter was calibrated according to manufacturers' instructions each time the unit was powered-on ([https://www.specmeters.com/assets/1/22/2900P\\_SPAD\\_502.pdf](https://www.specmeters.com/assets/1/22/2900P_SPAD_502.pdf)). The middle leaflet in the top-most, fully expanded trifoliate leaf was used for the measurements, and three plants were sampled at random per plot.

As plots reached physiological maturity, DTM was recorded as the date when plots were ready to harvest. Three plants were randomly sampled from mature plots, placed in large paper bags, and dried in a repurposed tobacco kiln (De Cloet Bulk Curing Systems, model TPG-360, Tillsonburg, ON, Canada) at 33 °C at the ERS for 24–48 h. Roots were cut from each plant and the above-ground biomass was weighed.

Plants were then threshed using an indoor belt thresher (Agriculex SPT-1A, Guelph, ON, Canada), their seed collected, weighed, and counted. Hundred seed weights (HSW) were calculated.

#### **4.3.3.2 Yorito**

As plots initiated the reproductive phase, DTF was recorded as the date when 50% of the plants in a plot had one flower open. Disease ratings and agronomic scores were collected throughout the growing season; however, statistical analyses revealed no significant differences between genotypes, and these traits are not further reported here. DTM was recorded as the date when 90% of the pods in a plot had changed color.

#### **4.3.4 Isotope analysis**

Seed from each plot was processed and analyzed as detailed in chapter 3. Briefly, seed was finely ground, precisely measured, and the isotope analyses were carried out using mass spectrometry at the Agriculture and Agri-Food Canada (AAFC) gas chromatography mass spectrometry facility in Lethbridge, Alberta. Samples were analyzed for  $\delta^{15}N$  (‰) and  $\delta^{13}C$  (‰).

To calculate the percent nitrogen derived from the atmosphere (%Ndfa), the natural abundance method was used on seed samples in this study (Shearer and Kohl, 1988). Seed N represents the total N accumulated by a plant over the course of the growing season, and seed N values are representative of whole-plant N values (Barbosa et al., 2018). Additionally, processing seed samples is more efficient than shoot tissue.

The natural abundance method uses the following equation,

$$\%Ndfa = \frac{\delta^{15}N_{reference\ plant} - \delta^{15}N_{fixing\ plant}}{\delta^{15}N_{reference\ plant} - B}$$

where  $\delta^{15}N_{reference\ plant}$  is the rate of  $\delta^{15}N$  in the reference genotype (R99),  $\delta^{15}N_{fixing\ plant}$  is the  $\delta^{15}N$  of the N-fixing bean genotype, and  $B$  is the average  $\delta^{15}N$  of beans grown in an environment where its entire N source is from fixation (Peoples et al., 2009). The  $B$ -value was obtained for this experiment as described by Farid (2015). Briefly,  $\delta^{15}N$  was measured and averaged for 20 bean genotypes from both the Andean and Middle American gene pools, which were grown in a growth room in N-free media. Normalized  $\delta^{15}N$  values were used for all genotypes, and an average of  $\delta^{15}N$  values for R99 were used in %Ndfa calculations.

To estimate water use efficiency,  $\delta^{13}C$  values obtained from GCMS seed analysis were used following the methods proposed by Farquhar et al. (1989). Because the current WUE discussion utilizes  $\Delta^{13}C$  values, the raw  $\delta^{13}C$  values were converted to  $\Delta^{13}C$  using the following equation:

$$\Delta C = \frac{\delta^{13}C_{air} - \delta^{13}C_{plant}}{1 + \delta^{13}C_{plant}}$$

where  $\delta^{13}C_{air}$  is the current free atmospheric level of approximately  $-8\text{‰}$  and  $\delta^{13}C_{plant}$  is calculated per plant seed sample using appropriate C isotope standards. For example,

a plant with a  $\delta^{13}\text{C}$  value of  $-28.2\text{‰}$  yields  $\Delta^{13}\text{C} = (-0.008 + 0.0282)/(1 - 0.0282) = 0.0202/0.9718 = 20.7\text{‰}$ .

#### **4.3.5 Genotyping**

To enable discovery of the genetic structure of the HON panel, 73 genotypes (see Tables 4.1-4.3) were genotyped for single nucleotide polymorphisms (SNPs). DNA was extracted following the same methods outlined in chapter 3. Briefly, plants were grown in a controlled environment (16 h photoperiod, 22 °C) at the University of Guelph, and young-leaf tissue samples were harvested, freeze dried, and the DNA extracted according to manufacturer's instructions using the NucleoSpin Plant II kit (Macherey-Nagel, Dueren, Germany). DNA of adequate quality was sent to the Genome Quebec Innovation Centre (McGill University, Montreal) for SNP genotyping using the Illumina Infinium iSelect Custom Genotyping BeadChip (BARCBEAN6K\_3) containing 5398 SNPs (Song et al., 2015). TASSEL was used to filter the SNP data ( $\text{MAF} > 0.01$ ) to a set representing 72 individuals and containing 4314 polymorphic SNPs (Bradbury et al., 2007). Missing data comprised less than 3% of the data, and these were subsequently imputed using Beagle v4.1 (Browning and Browning, 2016) as described by Torkamaneh and Belzile (2015).

#### **4.3.6 Population structure**

The population structure of the Honduran panel was determined using a number of methods, as follows. First, the population structure was estimated using variational Bayesian inference implemented in fastSTRUCTURE (Raj et al., 2014). Five runs were

performed for each number of populations ( $K$ ) set from 1 to 10 using the 4.3K genome-wide SNPs identified in this study. A ChooseK analysis was conducted to determine the number of subpopulations that maximize the marginal likelihood. Then, a principal component analysis (PCA) was conducted in TASSEL and plotted using PCASHINY in R (R Core Team, 2013). Finally, the evolutionary relationships among the genotypes of the panel were inferred using the Neighbor-Joining method with the genome-wide SNP data (Saitou and Nei, 1987). The taxa were clustered together using the bootstrap test (1000 replicates) (Felsenstein, 1985). The tree was drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and the units correspond to the number of base substitutions per site. Evolutionary analyses were conducted in MEGA7 (Tamura et al., 2004; Kumar et al., 2016).

#### **4.3.7 Genetic diversity**

The levels of genetic diversity in the landrace and PPB breeding history categories of the HON panel were assessed using the 4.3K SNP dataset and VCFtools (Danecek et al., 2011). The  $\pi$  statistic provides an indication of polymorphism within a population as measured by nucleotide diversity, and Tajima's  $D$  ( $D$ ) provides an indication of selection pressure (Nei and Li, 1979; Tajima, 1989). Both  $\pi$  and  $D$  were measured in sliding windows of 1 Mb across the genome using the—window-pi and—TajimaD options in VCFtools (Danecek et al., 2011), which resulted in an average of 6 SNPs per window. The pairwise  $\pi$  and  $D$  values were also calculated among different

subpopulations. Genome-wide averages of  $\pi$  and  $D$  for each breeding history category were generated by taking the average across all windowed calculations. Landrace and PPB  $\pi$  values were compared across the genome, and regions where landrace  $\pi$  exceeded PPB  $\pi$  by more than 3 times were considered highly differentiated, and the regions that were at least 25,000 bp in length were considered significant. To investigate the level of differentiation between the landrace and PPB genotypes, the  $F_{ST}$  statistic was computed.  $F_{ST}$  was calculated using the—weir-fst-pop option in VCFtools in sliding windows of 100 bp across the genome (Danecek et al., 2011). Weighted  $F_{ST}$  values range from 0 with no genetic differentiation to 1 where fixation of alleles has occurred.  $F_{ST}$  values exceeding 0.5 were considered significant in our analysis.

#### **4.3.8 Candidate gene investigation**

A literature search was carried out to identify previously reported QTL and genes that co-localize with regions where  $\pi$  values were highly differentiated between the Landrace and PPB categories. JBrowse (<https://legumeinfo.org/genomes/jbrowse/>) was used to explore the bean genome around SNPs with significant weighted  $F_{ST}$  values in order to identify candidate genes. A 100 kb region centered on each significant marker was searched.

#### **4.3.9 Statistical analysis**

Analysis of variance (ANOVA) tests were performed on the phenotypic data collected from each environment and all environments combined, using the GLIMMIX procedure in SAS (version 9.4, SAS Institute, Cary, NC, USA, 2012). In the combined

analysis, genotype, environment, and the genotype-by-environment interaction were considered fixed effects, while all other effects and their interactions were considered random. In the separate environment analyses, genotypes were considered fixed effects, while all other effects and the interaction effects were considered random. The Shapiro–Wilk test was performed on the residuals in the UNIVARIATE procedure to test their normality (Shapiro and Wilk, 1965). Random and independent distributions of the residuals were visually examined by plotting the studentized residuals against the predicted values. Data that generated outlier residuals were removed from the data set. Further, single degree of freedom contrasts were conducted in GLIMMIX between breeding history categories—landraces, PPBs, and conventional and check genotypes—contrasting each category to each of the others. Repeated measures of leaf chlorophyll content (SPAD) were taken, and separate ANOVA tests were used to compare SPAD values at each time point in a combined analysis and by environment. The least squared means (LSmeans) for each trait were computed using the LSMEANS statement in the GLIMMIX procedure for each genotype.

Using the LSmeans calculated above, the pair-wise Pearson’s coefficients of correlation were computed for all traits in the CORR procedure in SAS. The PRINCOMP and PRINQUAL procedures were used in SAS to generate the principal component (PC) values, to estimate the proportion of variance accounted for by each PC, and to plot PC1 against PC2 to generate genotype x trait (GT) biplots to determine genotype and trait interactions in each environment (Yan and Rajcan, 2002).

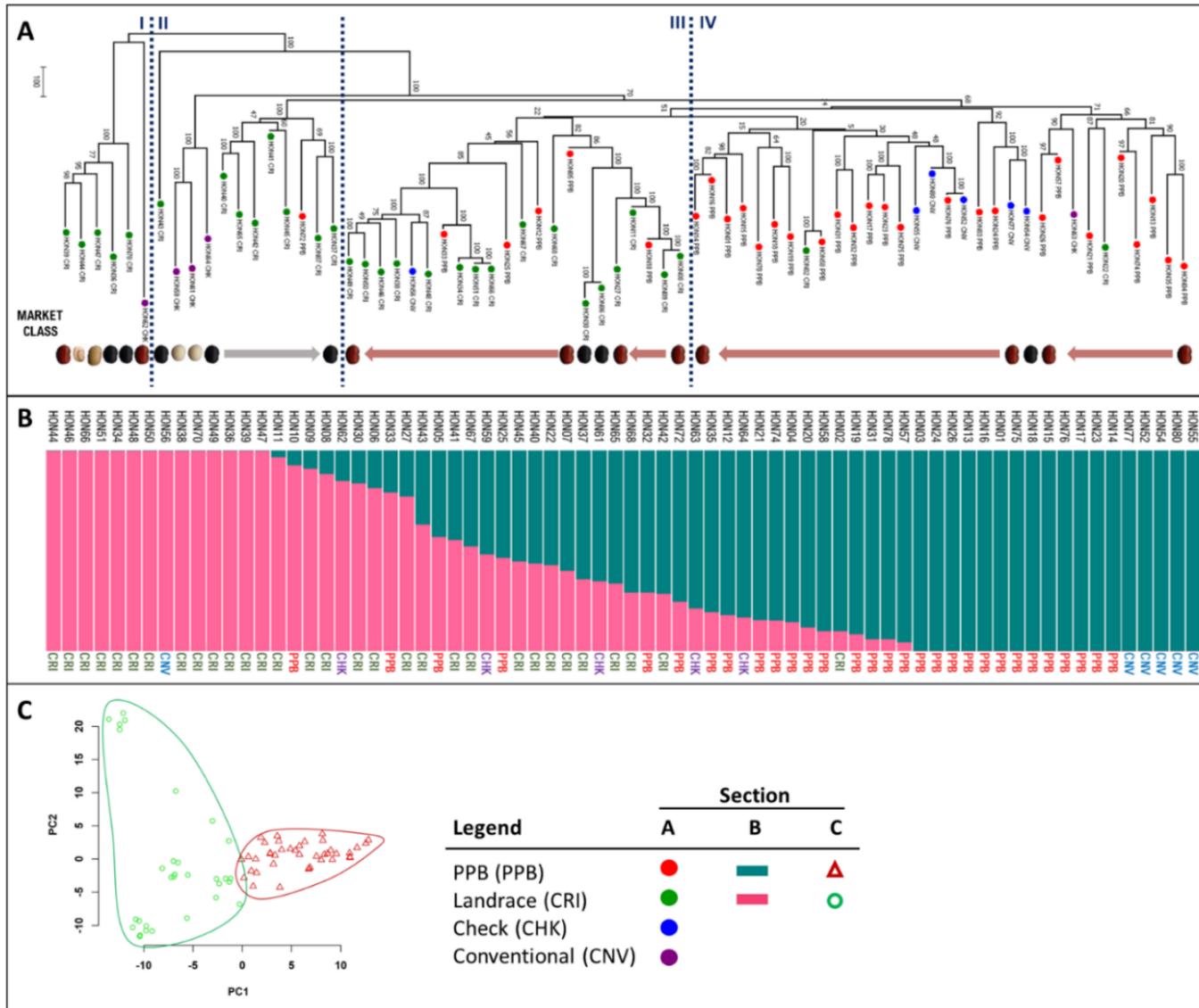
## 4.4 Results

### 4.4.1 Analysis of genetic relatedness

Landrace and PPB plant material for the Honduran panel were sourced from six municipalities across west–central Honduras. The majority of genotypes came from Yoro (26) and Francisco Morazán (17), with less than 20 genotypes coming from Intibucá, Santa Bárbara, Comayagua, and Lempira combined (Figure 4.1).

The genetic structure of the Honduran panel was explored to determine the evolutionary relatedness of the genotypes in the panel and the genetic composition of the genotypes. It is apparent from the topology of the phylogenetic tree (Figure 4.2A) that the landrace genotypes ('CRI') generally group into clusters of connected branches in the tree structure that are positioned in the left half of the figure, denoted as groupings I, II, and III (Figure 4.2A). The PPB genotypes ('PPB') grouped into separate clusters that are positioned in the right half of the figure, denoted as grouping IV (Figure 4.2A). Grouping I at the left of the tree, is comprised of the 'Milpero' genotypes, two landraces (HON70 and HON43), and Merlot (HON62). The Milpero landraces belong to diverse market classes, including black, small red, white, and carioca, and they included genotypes that did not flower at Elora in 2014. The remaining landrace genotype clusters were generally delineated by market class membership, with black genotypes comprising grouping II (including HON07, HON45, HON41, HON42, HON65, HON40, and HON43) and small red genotypes comprising grouping III (including HON08, HON09, HON11, HON27, HON68, HON67, HON66, HON51, HON34, HON48, HON38,

**Figure 4.2.** Analysis of genetic structure and relatedness of 72 genotypes of the Honduran panel. A. Dendrogram of evolutionary genetic relatedness. PPB (PPB), Landrace (CRI), North American check (CHK), Honduran conventional (CNV). Market classes are denoted by representative beans. B. Genetic structure plot using 2 genetic groupings ( $\Delta K=2$ ). C. Principal component analysis indicating two genetic groupings in the panel. HON #s in A and B correspond to those listed in Tables 6-8. (Note: In section A, grouping names I-IV are assigned to natural subsections of the tree for descriptive purposes and do not correspond to the genetic groups presented in section B.)



HON46, HON50, and HON49). The landraces 'Concha Rosada' (HON02) and 'Rosado' (HON22) are displaced and found among the PPB branches of the tree. The North American check genotypes ('CHK'; including HON64, HON61, and HON59) formed a separate cluster that branched off between the Milpero landraces and the black landraces in grouping II. 'OAC Rosito' (HON63), clustered with the Honduran PPBs. All Honduran conventional genotypes ('CNV'; including black HON54, and small red HON77, HON52, HON80, and HON55) grouped with the PPBs (grouping IV), except 'Dorado' (HON56), which is found among the landraces (grouping III). Six PPB genotypes (including HON10, HON05, HON12, HON25, HON33, and HON72) were found within the landrace clusters of the tree (groupings II and III).

The genetic similarity of genotypes in the panel is depicted in a STRUCTURE plot using two subpopulations ( $K = 2$ ) (Figure 4.2B). Fourteen of the landraces (including all of the Milpero types) belong to one genetic subgroup at the left of the plot and the PPB varieties belong to the other subgroup at the right of the plot, with an intermediary admixed group (Figure 4.2B). The Honduran conventional genotypes, except 'Dorado' (HON56), group with the PPBs. The North American check genotypes are found among the admixed genotypes, along with some PPBs and landraces. The principle component analysis of the panel also indicates the relatedness of the genotypes using two genetic groupings (Figure 4.2C). PC1 divides the genotypes into PPB (green triangles) and landrace (red circles) categories. Along the PC2 axis, the landraces show wide dispersion, with the Milpero group forming a small cluster near the

axis at the top of the plot, and the North American check genotypes are scattered among the landraces. In contrast, PC2 generates a denser cluster of PPB genotypes, and the Honduran conventional genotypes are at the right edge of the plot.

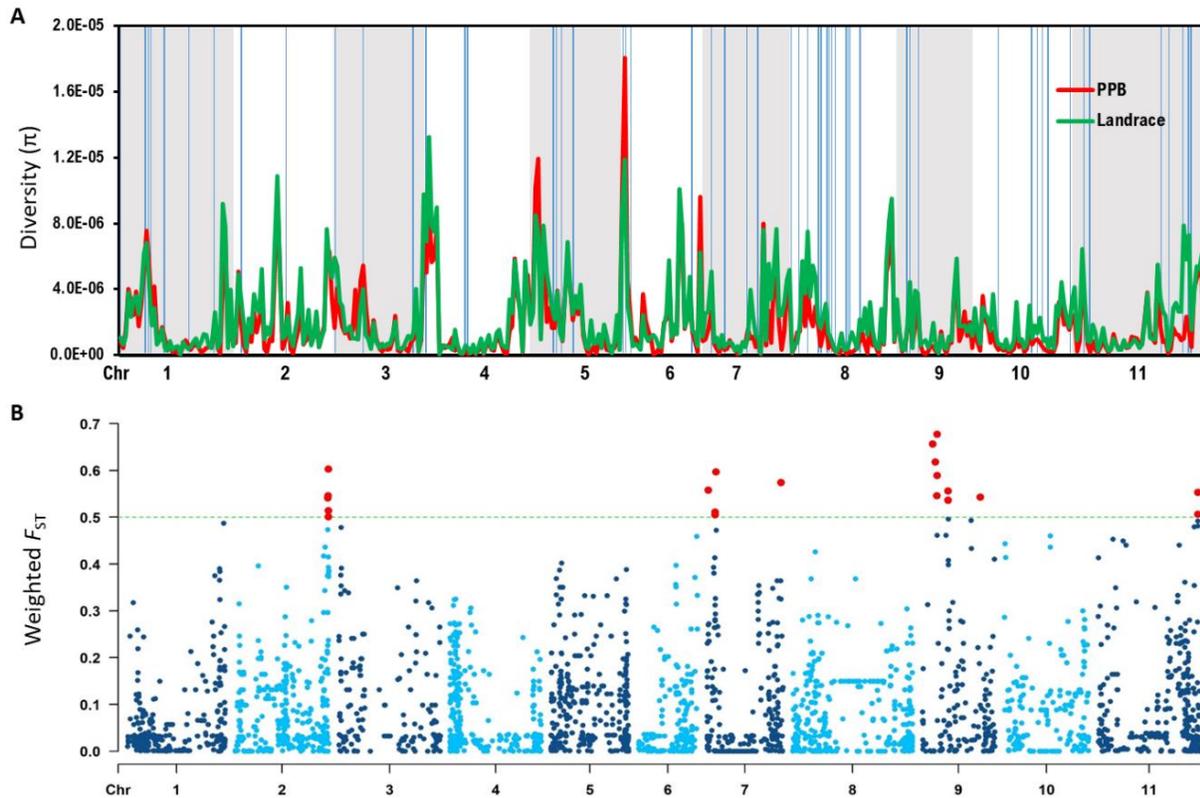
#### **4.4.2 Nucleotide diversity and population differentiation: Landrace and PPB categories**

Nucleotide diversity was measured in the two largest groupings within the Honduran panel, the landraces and the PPBs, to ascertain the genetic diversity of these groups. According to the  $\pi$  statistic, nucleotide diversity for the landrace category overall ( $\pi = 3.20 \times 10^{-4}$ ) was significantly greater ( $P = 0.04$ , Welch two-sample  $t$ -test) than that found in the PPB category overall ( $\pi = 2.89 \times 10^{-4}$ ). Additionally, according to the  $D$  statistic, the overall nucleotide diversity for the landrace category ( $D = 0.669$ ) was significantly greater ( $P = 0.02$ , Welch two-sample  $t$ -test) than that found in the PPB category ( $D = 0.476$ ). The positive Tajima's  $D$  value indicates that both landraces and PPBs are under balancing selection and implies that both categories are probably experiencing different selective pressure. Fifty-six subregions (>100 Mbp long) across the genome were identified where landrace  $\pi$  values exceeded PPB  $\pi$  values by more than 3 times (Table 4.4, Figure 4.3A). These regions, identified on all 11 chromosomes, may contain loci related to traits favored by selection associated with formal plant breeding (Figure 4.3A)

**Table 4.4.** Regions of the *P. vulgaris* (2.0) genome where high nucleotide diversity ( $\pi$ ) was discovered in landrace genotypes compared to PPB genotypes. A literature search was performed to identify candidate genes within these regions. See Appendix B, Table S4.1 for candidate gene annotation. (Chr – Chromosome)

Chr	Region of high diversity		$\pi$ value		Candidate genes
	Start (Mbp)	End (Mbp)	Landrace	PPB	Number
1	23.0	24.0	5.50E-07	1.31E-07	2
1	40.0	41.0	7.71E-07	1.83E-07	4
1	41.0	42.0	1.24E-06	2.49E-07	4
1	42.0	43.0	1.22E-06	3.98E-07	1
1	47.0	48.0	8.60E-07	9.66E-08	3
2	4.0	5.0	1.95E-06	6.51E-07	0
2	22.0	23.0	1.71E-06	1.64E-07	8
2	32.0	33.0	5.53E-07	1.22E-07	5
2	48.0	49.0	5.92E-06	1.62E-06	18
3	34.0	35.0	6.73E-07	1.97E-07	8
3	48.0	49.0	1.17E-06	1.89E-07	8
4	12.0	13.0	1.53E-06	2.62E-07	0
4	17.0	18.0	4.05E-07	6.55E-08	0
4	38.0	39.0	4.74E-07	9.66E-08	6
4	40.0	41.0	4.32E-07	1.27E-07	0
5	24.0	25.0	3.55E-07	6.55E-08	1
5	25.0	26.0	2.10E-06	1.92E-07	0
5	27.0	28.0	7.46E-07	1.31E-07	1
5	32.0	33.0	1.16E-06	3.28E-07	11
6	9.0	10.0	1.21E-06	6.55E-08	7
6	13.0	14.0	1.02E-06	1.31E-07	14
6	15.0	16.0	4.82E-07	1.55E-07	6
7	8.0	9.0	4.32E-07	9.66E-08	1
7	20.0	21.0	5.27E-07	6.55E-08	1
7	26.0	27.0	3.98E-06	1.01E-06	0
7	35.0	36.0	3.36E-06	7.38E-07	9
7	39.0	40.0	2.43E-06	3.32E-07	8
8	11.0	12.0	5.44E-06	1.74E-06	1
8	15.0	16.0	1.63E-06	5.32E-07	2
8	18.0	19.0	1.30E-06	3.98E-07	1

Chr	<u>Region of high diversity</u>		<u><math>\pi</math> value</u>		<u>Candidate genes</u>
	Start (Mbp)	End (Mbp)	Landrace	PPB	Number
8	23.0	24.0	5.27E-07	6.55E-08	0
8	29.0	30.0	1.35E-06	3.32E-07	0
8	38.0	39.0	3.29E-07	3.33E-08	0
8	41.0	42.0	7.62E-07	2.26E-07	2
8	44.0	45.0	1.35E-06	1.62E-07	0
8	45.0	46.0	9.15E-07	2.90E-07	0
8	51.0	52.0	2.86E-06	6.33E-07	6
8	52.0	53.0	1.21E-06	9.66E-08	12
8	57.0	58.0	3.45E-07	6.55E-08	0
9	16.0	17.0	7.81E-07	1.60E-07	0
9	18.0	19.0	9.04E-07	3.33E-08	3
9	21.0	22.0	1.94E-06	4.99E-07	0
10	17.0	18.0	2.97E-07	9.32E-08	0
10	33.0	34.0	2.26E-06	7.41E-07	0
10	35.0	36.0	9.78E-07	2.62E-07	2
10	37.0	38.0	1.07E-06	1.94E-07	7
10	39.0	40.0	3.00E-06	7.34E-07	1
11	2.0	3.0	8.74E-07	1.93E-07	3
11	7.0	8.0	8.00E-07	1.55E-07	23
11	10.0	11.0	1.37E-06	3.01E-07	7
11	37.0	38.0	2.64E-06	7.84E-07	1
11	40.0	41.0	9.07E-07	3.01E-07	3
11	45.0	46.0	7.85E-06	9.80E-07	9
11	47.0	48.0	7.31E-06	1.53E-06	0
11	48.0	49.0	1.52E-06	4.99E-07	0
11	53.0	54.0	6.90E-07	2.28E-07	11



**Figure 4.3.** Nucleotide diversity and population differentiation. A. Patterns of nucleotide diversity ( $\pi$ ) across the genome between *P. vulgaris* Landrace (green) and PPB (red) genotypes. Blue vertical bars show the strongly differentiated (3x) regions. B. Weighted  $F_{ST}$  plot of genetic variance differentiation among Landrace and PPB categories. Significant SNPs are red. Significance threshold  $F_{ST} > 0.5$ .

Calculation of population genetic differentiation ( $F_{ST}$ ) between landrace and PPB beans enabled identification of loci under selection between landrace and PPB genotypes. Twenty-six single nucleotide polymorphisms (SNPs) with significant weighted  $F_{ST}$  values ( $>0.5$ ) were found on Pv02, Pv07, Pv09, and Pv11 (Table 4.5, Figure 4.3B). These SNPs do not fall within the regions of high nucleotide diversity identified in the  $\pi$  comparison above.

**Table 4.5.** Regions of the *P. vulgaris* (2.0) genome where SNPs with significantly high weighted  $F_{ST}$  values ( $> 0.5$ ) were found. JBrowse was used to search for candidate genes within 100 Kb of significant SNPs. Candidate gene descriptions are listed in Appendix B, Table S4.2. (<https://legumeinfo.org/genomes/jbrowse/>). (Chr – chromosome)

<u>Chr</u>	<u>SNP Position (Mbp)</u>	<u><math>F_{ST}</math> value</u>	<u>Number of candidate genes</u>
2	48.9	0.541	8
2	49.1	0.546	11
2	49.1	0.603	13
2	49.2	0.514	10
	49.2	0.501	
7	0.6	0.558	15
	0.7	0.558	
7	4.2	0.505	13
	4.2	0.511	
7	4.7	0.597	7
7	38.9	0.574	10
9	5.5	0.656	8
9	6.9	0.618	3
9	7.7	0.546	5
9	7.8	0.677	9
	7.9	0.589	
9	13.5	0.536	11
	13.6	0.536	
9	13.7	0.556	11
9	30.6	0.543	6
11	52.4	0.553	10
	52.4	0.506	
	52.4	0.553	
	52.4	0.553	
	52.5	0.553	
	52.5	0.553	

#### 4.4.3 Identification of candidate genes

Two approaches were used to identify candidate genes associated with regions of significantly high (3x) nucleotide diversity ( $\pi$ ) in landraces and high population differentiation ( $F_{ST}$ ) between landraces and PPBs, including: exploring the recent bean literature for reported quantitative trait loci (QTL) and searching the bean genome using JBrowse.

QTL associated with various traits have been reported in the literature, including those related to agronomic traits (Schmutz et al., 2014; Moghaddam et al., 2016) and nitrogen fixation (Cichy et al., 2015; Farid, 2015; Diaz et al., 2017; Heilig et al., 2017a; Kamfwa et al., 2019). Our literature search revealed 10 QTL that fall within regions of significantly high landrace  $\pi$  values, 8 of which are associated with agronomic traits and 2 of which are associated with SNF-related traits. In a GWAS study of agronomic traits in the Middle American Diversity panel (MDP) (chapter 5), we found QTL for days to flowering on Pv01 (23.2 Mbp), Pv02 (48.6 Mbp), and Pv06 (13.9 Mbp); QTL for days to maturity on Pv07 (35.6 Mbp) and Pv11 (40.3 Mbp); and QTL for hundred seed weight on Pv01 (23.2 Mbp), Pv05 (32.5 Mbp), and Pv11 (53.5 Mbp). Various candidate genes were found underlying these agronomic QTL and more detailed information is available in Appendix B, Table S4.1. In a separate GWAS study of agronomic traits in the MDP, Moghaddam et al. (2016) found a QTL on Pv01 (42.9 Mbp) associated with growth habit, which contained an RNA polymerase-associated protein RTF1 homolog (Phvul.001G167200). For SNF-related traits, reported QTL that fall within regions of high landrace  $\pi$  values are associated with seed %N content and plant biomass. In a

GWAS study of SNF related traits in the MDP (chapter 5) we found a QTL associated with seed %N content at 22.8 Mbp on Pv02. The QTL contains seven putative candidate genes, including a Ras homologous (Rho)/Rho of plants (Rop) family GTPase (Phvul.002G106600). These genes play a role in the symbiotic interaction between the host plant and rhizobia (Flores et al., 2018). Two separate studies investigating SNF and related traits in the Andean and the Middle American gene pools identified a QTL associated with shoot biomass at 45.1 Mbp on Pv11 (Heilig et al., 2017a; Kamfwa et al., 2019). Shoot biomass supports root symbiosis through carbohydrates generated through photosynthesis as well as serving as a sink for N generated through SNF, which is a source of N ultimately stored in the seed (Kamfwa et al., 2019). Beyond searching the recent literature for QTL associated with SNF and agronomic traits, we also examined the study by Schmutz et al. (2014), which identified >1800 domestication candidate genes in the Middle American gene pool. Over 140 of the domestication genes identified by Schmutz et al. (2014) fall within regions of high nucleotide diversity discovered in our study (see Appendix B, Table S4.1). Two of these genes have a role in symbiosis: Phvul.008G217100 is a short open reading frame (sORF) small protein of the glycerin rich protein family and is expressed during nodule ontology (Guillén et al., 2013); and Phvul.010G102300 belongs to the plant nuclear factor Y (NF Y) gene family, whose members are involved in nodule ontology, epidermal infection, and rhizobia discrimination (Rípodas et al., 2015).

The bean genome was explored using JBrowse in 100 kb segments centered on SNPs with significant genetic differentiation ( $F_{ST}$ ) to identify putative candidate genes.

All genes found within these regions are listed in Appendix B, Table S4.2. A diverse range of gene types and functions were seen, including plant defense and stress response genes, enzymes, and transcription factors. The PubMed Central database of NCBI (<https://www.ncbi.nlm.nih.gov/pmc/>) was used to search for published research on putative candidate genes, and some of those findings are discussed here. The region flanking the significant  $F_{ST}$  SNP on Pv02 (48.9 Mb) contains two leucine rich repeat disease resistance proteins, Phvul.002G323708 and Phvul.002G323712. This region was identified by Oladzad et al. (2019) as a major QTL peak in their GWAS study of *Rhizoctonia solani* resistance in Andean beans. A second region on Pv02 (49.0 Mb) contains a disease resistance gene and one associated with nodulation. Tock et al. (2017) found that the pentatricopeptide repeat superfamily protein (Phvul.002G326200) at 49.0 Mb was associated with halo blight damage, while Nova-Franco et al. (2015) found that a 1-aminocyclopropane-1-carboxylate oxidase gene (Phvul.002G326600) in this region was associated with nodule senescence. A third region on Pv02 (49.2 Mb) contains a protein kinase superfamily protein (Phvul.002G328300) that Zuiderveen et al. (2016) found to be significantly associated with Anthracnose resistance in a GWAS of Andean beans. The region centered at 38.9 Mb on Pv07 contains a protein kinase superfamily protein (Phvul.007G268200), which was downregulated in a slow darkening pinto bean study (Duwadi et al., 2018). On Pv09, the region located at 7.8 Mb contains a GATA transcription factor (Phvul.009G035400), which belongs to a family of transcription factors that have been studied in soybean under nitrogen stress and may play a role in nitrogen metabolism (Zhang et al., 2015a). In the region centered on 13.5

Mb on Pv09, a cytokinin oxidase/dehydrogenase 1 gene (Phvul.009G081800) is located that was found to be upregulated in bean root cortical cells inoculated with arbuscular mycorrhizal fungi under drought stress, compared to non-inoculated roots (Recchia et al., 2018).

#### **4.4.4 Diversity for symbiotic nitrogen fixation**

The influences of genotype, environment, and the genotype by environment interaction were significant for the combined locations ANOVA for %Ndfa (Appendix B, Table S4.3); therefore, each environment was analyzed separately for this trait. At each location, significant differences were found between genotypes for %Ndfa (Appendix B, Tables S4.4-6). At Elora 2014 ( $N = 49$ ), the average nitrogen fixation capacity was 49.3% and ranged from 21.0% to 64.4%, a difference of 43.4% between the least and most effective genotypes. At Elora 2015 ( $N = 62$ ), the average nitrogen fixation capacity was higher at 55.8%, yet the range for this trait was narrower with a low of 40.5% and a high of 67.3%, a difference of 26.8% between the least and most effective genotypes. At Yorito ( $N = 53$ ), the average nitrogen fixation capacity was 49.0% with a range of 14.0% to 66.4%, a difference of 52.4% between the least and most effective genotypes, which was the greatest range in performance of all locations.

Further, in a separate ANOVA for each location, the genotypes were divided into breeding history categories and their means were compared. In these analyses, significant differences were found among breeding categories at two of the three trial locations. At Elora 2014, the landrace genotypes ( $N = 20$ ,  $M = 52.5$  %Ndfa) performed

better than all other breeding categories, although the difference between categories was not significant (Table 4.6). It is evident from the Elora 2014 %Ndfa histogram (Appendix B, Figure S4.1A) that many landrace genotypes had above average nitrogen fixation performances. At Elora 2015, the Honduran conventional genotypes ( $N = 7$ ,  $M = 59.0$  %Ndfa) and the landraces ( $N = 26$ ,  $M = 58.3$  %Ndfa) exhibited the best nitrogen fixing capacities, but they were not significantly different from each other (Table 4.6).

**Table 4.6.** F-test of fixed effect breeding category in the GLIMMIX analysis, and the breeding category LSmeans comparisons of genotypes in the HON panel grown at Elora, 2014.

Breeding category <sup>1</sup>	N derived from the atmosphere (%)		Carbon discrimination (Δ) (‰)		Flowering (days)		Maturity (days)		Yield (kg ha <sup>-1</sup> )	
	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>
	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>
CHK	1.86	0.1441	1.51	0.2195	6.13	0.0007	3.06	0.0368	7.15	0.0004
CNV	51.4 <sup>a</sup>	0.04	20.2 <sup>a</sup>	0.23	48.6 <sup>ab</sup>	1.44	114.9 <sup>a</sup>	2.0	933.7 <sup>a</sup>	53.18
LDR	43.3 <sup>a</sup>	0.04	19.6 <sup>a</sup>	0.24	51.0 <sup>ab</sup>	1.44	108.6 <sup>a</sup>	2.82	700.9 <sup>ab</sup>	77.10
PPB	52.5 <sup>a</sup>	0.02	20.1 <sup>a</sup>	0.14	48.2 <sup>b</sup>	0.87	109.9 <sup>a</sup>	1.19	912.8 <sup>a</sup>	28.39
PPB	46.7 <sup>a</sup>	0.02	19.9 <sup>a</sup>	0.14	51.6 <sup>a</sup>	0.83	114.0 <sup>a</sup>	1.26	721.6 <sup>b</sup>	30.51

<sup>1</sup> CHK – check; CNV – Honduran conventional; LDR – landrace; PPB – participatory plant bred

\*Means labeled with different letters within trait categories are significantly different according to ANOVA,  $p = 0.05$ .

The average nitrogen fixing capacities of the North American check genotypes ( $N = 5$ ,  $M = 50.0$  %Ndfa) and the PPBs ( $N = 24$ ,  $M = 53.2$  %Ndfa) were similar at Elora 2015 and significantly lower than the values for the Honduran conventional and landrace genotypes (Table 4.7). Of particular note, Merlot (HON62) fixed the most N at Yorito (66.4%), almost 6% more than the next best genotype. This genotype, bred for Northern US production, also performed well at Elora in 2014 (64.3 %Ndfa), but had the worst

performance among conventional genotypes at Elora 2015 (43.5 %Ndfa). Merlot has very dark green leaves, an indicator of plant N status, and consistently had high leaf chlorophyll content when measured with the SPAD meter in the Elora trials (SPAD was not measured at Yorito). As with the Elora 2014 trial, many landrace genotypes had above average nitrogen fixation performances at Elora 2015 (Appendix B, Figure S4.1B).

**Table 4.7.** F-test of fixed effect breeding category in the GLIMMIX analysis, and the breeding category LSmeans comparisons of genotypes in the HON panel grown at Elora, 2015.

Breeding category <sup>1</sup>	N derived from the atmosphere (%)		Carbon discrimination (Δ) (‰)		Flowering (days)		Maturity (days)		Yield (kg ha <sup>-1</sup> )	
	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>
	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>
<b>CHK</b>	6.69	0.0004	0.88	0.4535	0.94	0.4225	3.27	0.0251	3.66	0.0148
<b>CNV</b>	50.0 <sup>a</sup>	0.03	19.8 <sup>a</sup>	0.32	50.8 <sup>a</sup>	0.83	113.1 <sup>ab</sup>	1.65	1454.7 <sup>ab</sup>	145.59
<b>LDR</b>	59.0 <sup>bc</sup>	0.02	19.8 <sup>a</sup>	0.29	49.6 <sup>a</sup>	0.75	109.9 <sup>ab</sup>	1.42	1613.2 <sup>ab</sup>	121.38
<b>PPB</b>	58.3 <sup>b</sup>	0.01	19.5 <sup>a</sup>	0.24	49.3 <sup>a</sup>	0.36	109.0 <sup>a</sup>	0.63	1396.5 <sup>a</sup>	58.55
<b>PPB</b>	53.2 <sup>ac</sup>	0.01	19.6 <sup>a</sup>	0.24	49.3 <sup>a</sup>	0.36	112.1 <sup>b</sup>	0.65	1686.2 <sup>b</sup>	62.74

<sup>1</sup> CHK – check; CNV – Honduran conventional; LDR – landrace; PPB – participatory plant bred

\*Means labeled with different letters within trait categories are significantly different according to ANOVA,  $p = 0.05$

At Yorito, the landrace genotypes ( $N = 22$ ,  $M = 46.4$  %Ndfa) showed significantly higher nitrogen fixing capacities than the PPBs ( $N = 22$ ,  $M = 40.1$  %Ndfa), whereas, the check and Honduran conventional genotypes had intermediary SNF means and did not have significantly different nitrogen fixing performance values when compared to each other nor the other breeding categories (Table 4.8). As we found at the other trial

locations, the landraces showed above average nitrogen fixing performance at Yorito (Appendix B, Figure S4.1C).

**Table 4.8.** F-test of fixed effect breeding category in the GLIMMIX analysis, and the breeding category LSmeans comparisons of genotypes in the HON panel grown at Yorito, 2014-2015.

Breeding category <sup>1</sup>	N derived from the atmosphere (%)		Carbon discrimination (Δ) (‰)		Days to Flowering (days)		Yield (kg ha <sup>-1</sup> )	
	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>	<i>F</i>	<i>Pr &gt; F</i>
	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>	<i>LSmean*</i>	<i>SE</i>
CHK	3.72	0.0143	3.18	0.0280	3.60	0.0163	1.34	0.2647
CNV	49.5 <sup>ab</sup>	0.04	18.0 <sup>ab</sup>	0.25	39.0 <sup>ab</sup>	1.28	669.9 <sup>a</sup>	135.22
LDR	40.1 <sup>ab</sup>	0.03	18.0 <sup>ab</sup>	0.22	36.8 <sup>ab</sup>	1.06	956.5 <sup>a</sup>	108.87
PPB	46.4 <sup>a</sup>	0.01	18.2 <sup>a</sup>	0.10	36.2 <sup>a</sup>	0.53	745.2 <sup>a</sup>	52.16
	40.1 <sup>b</sup>	0.01	17.7 <sup>b</sup>	0.10	38.3 <sup>b</sup>	0.54	823.2 <sup>a</sup>	52.62

<sup>1</sup> CHK – check; CNV – Honduran conventional; LDR – landrace; PPB – participatory plant bred

\*Means labeled with different letters within trait categories are significantly different according to ANOVA,  $p = 0.05$ .

The top five landraces with the highest SNF performance at Yorito were Vaina Rosada (60.6 %Ndfa; HON34), Cincuentaño (59.5 %Ndfa; HON48), Negro Cuarenteño (57.0 %Ndfa; HON42), Olanchano Negro (56.4 %Ndfa; HON65), and Paraíso (53.6 %Ndfa; HON49). These landraces represent both small red and black market classes and are from three different departments (Yoro, Francisco Morazán, and Intibucá). Vaina Rosada, Cincuentaño, and Paraíso are already used in participatory breeding efforts between Zamorano and FIPAH, and a number of the resulting PPB varieties were included in our panel (including HON05, HON23, HON25, HON26, HON28, HON31, HON32, and HON33). The PPB progeny of these landraces ranged in SNF capacity from 26.6 to 53.3 %N at Yorito, which is noteworthy considering SNF was not a

breeding objective. Amilcar (53.3 %Ndfa; HON05) is among the top five SNF performing PPB varieties at Yorito, which also included Conan 33 (55.7 %Ndfa; HON24), Campechano (54.5 %Ndfa; HON57), San Jose (51.1 %Ndfa; HON35), and Arbolito Negro (50.8 %Ndfa; HON72). Both small red and black beans are represented in this list, and they come from three departments (Yoro, Francisco Morazán, and Santa Barbara).

Leaf chlorophyll content was measured at Elora in 2014 and 2015, and these values were analyzed in separate ANOVAs because of the repeated-measure nature of trait data collection. The combined ANOVA indicated significant differences for the fixed effects of genotype, environment, and the genotype by environment interaction at both the early vegetative and reproductive stages (Appendix B, Table S4.7). When this trait was analyzed by location, significant differences were found between genotypes at both locations and for both growth stages (Appendix B, Table S4.7).

#### **4.4.5 Diversity for agronomic traits**

A series of agronomic traits were measured for this study, including carbon discrimination ( $\Delta^{13}\text{C}$ ) as an indicator of water use efficiency, plant growth stages (days to flowering and maturity), yield ( $\text{kg ha}^{-1}$ ), and hundred seed weight. Significant differences were found for the fixed effects of genotype, environment, and the genotype by environment interaction for the agronomic traits carbon discrimination ( $\Delta^{13}\text{C}$ ), days to flowering (DTF) and days to maturity (DTM), yield ( $\text{kg ha}^{-1}$ ), and hundred seed weight

(HSW) in a combined ANOVA (Appendix B, Table S4.3). These traits were therefore analyzed further within locations.

Significant differences were found between genotypes at all locations (Appendix B, Tables S4.4–6) for carbon discrimination (in  $\Delta^{13}\text{C}$  units) calculated according to the method of Farquhar et al. (1989) from seed carbon discrimination ( $\delta^{13}\text{C}$ ) values obtained from isotope analysis of seed samples. At Elora 2014 ( $N = 48$ ),  $\Delta^{13}\text{C}$  values ranged from 18.4‰ to 21.4‰ (Appendix B, Figure S4.2A), and at Elora 2015 ( $N = 62$ ), the range was similar, from 17.4‰ to 21.1‰ (Appendix B, Figure S4.2B). At Yorito ( $N = 53$ ), the  $\Delta^{13}\text{C}$  values were lower, ranging from 16.4‰ to 20.0‰ (Appendix B, Figure S4.2C). When genotypes were divided into breeding categories and compared, significant differences were only found at Yorito (Table 4.8), where the average PPB  $\Delta^{13}\text{C}$  value ( $N = 22$ ,  $M = 17.7\text{‰}$ ) was significantly lower than the average landrace  $\Delta^{13}\text{C}$  value ( $N = 22$ ,  $M = 18.2\text{‰}$ ) (Table 4.8).

Significant differences were found between genotypes in DTF measured at Elora in 2014 and 2015 (Appendix B, Tables S4.4–6). At Elora 2014 ( $N = 58$ ), the average was 50 DTF with a range of 42–62 DTF (Appendix B, Figure S4.3A). Some Honduran genotypes did not flower at that first trial location and were replaced by different genotypes at the subsequent locations (Tables 4.2 and 4.3). At Elora 2015 ( $N = 57$ ), the average was 50 DTF with a range of 42–55 DTF (Appendix B, Figure S4.3B). When genotypes were divided into breeding categories and compared, significant differences were found at Elora 2014 only (Table 4.6). Overall, the landrace genotypes flowered the

earliest ( $N = 22$ ,  $M = 48$  days) and they were significantly earlier than the PPB genotypes ( $N = 26$ ,  $M = 52$  days) (Table 4.6).

Significant differences were found among genotypes for DTM measured at Elora in 2014 and 2015 (Appendix B, Tables S4.4–6). At Elora 2014 ( $N = 35$ ), the average was 112 DTM with a range of 97–120 DTM (Appendix B, Figure S4.4A). At Elora 2015 ( $N = 56$ ), the average was 111 DTM with a range of 94–115 DTM (Appendix B, Figure S4.4B). Significant differences in DTM were found only at Elora 2015 (Table 4.7), when genotypes were grouped by breeding history and contrasted. At Elora 2015, landraces ( $N = 23$ ,  $M = 109$  days) matured significantly earlier than PPBs ( $N = 22$ ,  $M = 112$  days).

Significant differences were found among the yields of genotypes in the Elora trials only (Appendix B, Tables S4.4 and S4.5). At Elora 2014 ( $N = 35$ ), the average yield was 828.4 kg ha<sup>-1</sup> with a range from 325.2–1124.2 kg ha<sup>-1</sup> (Appendix B, Figure S4.5A). At Elora 2015 ( $N = 62$ ), the average yield was 1558.6 kg ha<sup>-1</sup> with a range from 600.5–2263.4 kg ha<sup>-1</sup> (Appendix B, Figure S4.5B). At Yorito ( $N = 58$ ), the average yield was 791.1 kg ha<sup>-1</sup> with a range from 299–1471 kg ha<sup>-1</sup> (Appendix B, Figure S4.5C). Significant differences were found only in the Elora trials among breeding categories (Tables 4.6 and 4.7). At Elora 2014, the North American check ( $N = 5$ ,  $M = 933.7$  kg ha<sup>-1</sup>) and the Honduran landrace ( $N = 14$ ,  $M = 912.8$  kg ha<sup>-1</sup>) genotypes yielded significantly more than the other categories (Table 4.6), and at Elora 2015, the PPBs ( $N = 23$ ,  $M = 1686.2$  kg ha<sup>-1</sup>) yielded significantly more than the landraces ( $N = 27$ ,  $M = 1396.5$  kg ha<sup>-1</sup>; Table 4.7). At Yorito, the Honduran conventional genotypes ( $N = 6$ ,  $M =$

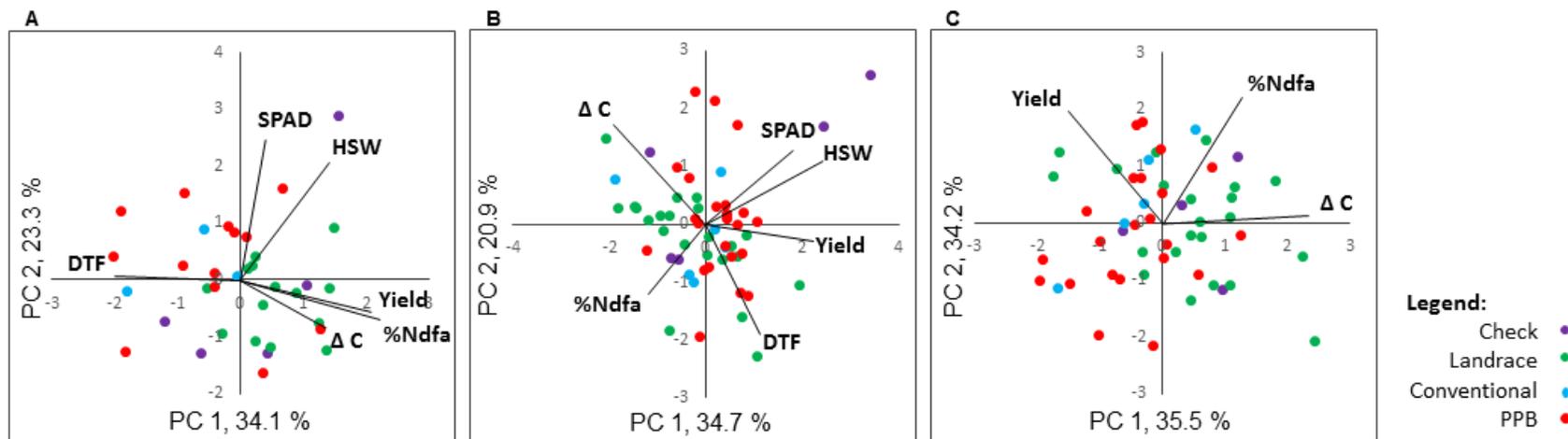
956.5 kg ha<sup>-1</sup>) returned the highest yields, followed by the PPBs ( $N = 24$ ,  $M = 823.2$  kg ha<sup>-1</sup>), while the landraces ( $N = 24$ ,  $M = 745.2$  kg ha<sup>-1</sup>) and the North American checks ( $N = 4$ ,  $M = 669.9$  kg ha<sup>-1</sup>) had lower yields (Table 4.8).

Significant differences were found among hundred seed weight (HSW) calculated for samples from the Elora trials (Appendix B, Tables S4.4 and S4.5). At Elora in 2014 ( $N = 49$ ), the average HSW was 20.2 g with a range from 13.3–29.4 g. At Elora in 2015 ( $N = 62$ ), the average HSW was 21.4 g with a range from 14.5–31.9 g. No significant differences were found among genotypes grouped by breeding history (data not shown).

Plant height was measured only at the Yorito location ( $N = 60$ ), and significant differences were found between genotypes for this trait (Appendix B, Table S4.6). The average height was 35.3 cm with a range of 4 cm to 47 cm. Significant differences were not found when breeding history categories were contrasted (data not shown).

#### **4.4.6 Trait correlation and genotype by trait biplot analyses**

Pearson correlation analyses were performed on LSmeans for each trial environment to determine trait interactions (Appendix B, Table S4.8). In addition, trait correlations and genotype performance were visualized using genotype  $\times$  trait biplots for each location (Figure 4.4). In the biplots, positive correlations between traits are evidenced by vectors forming acute angles, for example between SPAD and HSW at Elora in 2014 and 2015 (Figure 4.4B), whereas negative correlations between traits are



**Figure 4.4.** Biplot analysis of traits for genotypes of the Honduran panel in three location years. A. Elora 2014; B. Elora 2015; and C. Yorito 2014–15. DTF, days to flowering; Yield, yield ( $\text{kg ha}^{-1}$ ); HSW, 100 seed weight (g);  $\Delta C$ , carbon discrimination; %Ndfa, percent nitrogen derived from the atmosphere; and SPAD, leaf chlorophyll content at 100% flowering.

evidenced by obtuse angles formed between vectors, such as that formed by DTF and yield at Elora in 2014 (Figure 4.4A). A right-angle formed between trait vectors indicates a weak or lack of association between those traits. The results of our correlation and biplot analyses support each other.

#### **4.4.6.1 Percent nitrogen derived from the atmosphere (%Ndfa)**

At Elora in 2014, %Ndfa was negatively correlated with DTF ( $r = -0.31$ ) but was positively correlated with  $\Delta^{13}\text{C}$  ( $r = 0.45$ ) and with yield ( $r = 0.38$ ) (Appendix B, Table S4.8). In the Elora 2014 biplot (Figure 4.4A), the landrace genotypes are clustered towards the yield and the %Ndfa vectors. The Honduran conventional genotypes are more closely associated with the DTF vector, as are the majority of the Honduran conventional genotypes (Figure 4.4A).

At Elora in 2015, %Ndfa was not significantly associated with any other trait (Appendix B, Table S4.8). The biplot analysis showed that DTF and  $\Delta^{13}\text{C}$  are not associated with %Ndfa, and leaf chlorophyll content at flowering (SPAD) and HSW have a negative relationship with %Ndfa (Figure 4.4B). As in the Elora 2014 biplot, the landrace genotypes cluster towards the %Ndfa vector (Figure 4.4B).

At Yorito, %Ndfa was not significantly correlated with other traits (Appendix B, Table S4.8). The biplot analysis for Yorito shows landrace genotypes cluster more towards the %Ndfa and  $\Delta^{13}\text{C}$  vectors, whereas the PPB genotypes cluster away from the %Ndfa vector and are more closely associated with the yield vector (Figure 4.4C).

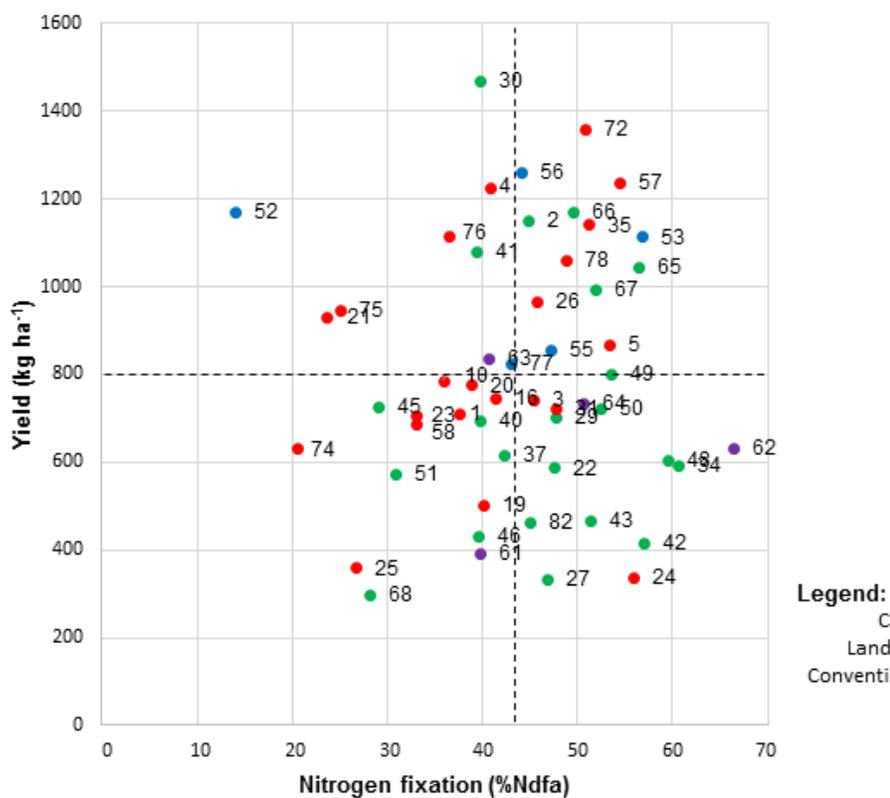
#### 4.4.6.2 Agronomic traits

Leaf chlorophyll content at the early reproductive stage (SPAD) was positively correlated with hundred seed weight (HSW) at both Elora 2014 ( $r = 0.36$ ) and Elora 2015 ( $r = 0.44$ ) (Appendix B, Table S4.8). DTF was negatively associated with yield at Elora 2014 ( $r = -0.48$ ), but no association was found between these traits at Elora 2015. DTF was negatively associated with  $\Delta^{13}\text{C}$  at Elora 2015 ( $r = -0.37$ ), but no association was found between these traits at Elora 2014 (Appendix B, Table S4.8). Yield was positively associated with HSW ( $r = 0.49$ ) and negatively associated with  $\Delta^{13}\text{C}$  ( $r = -0.33$ ) at Elora 2015; however, these associations were not repeated at the other trial locations (Appendix B, Table S4.8).

#### 4.4.6.3 High-yielding and high-fixing genotypes

The aim of any breeding program is to generate high-yielding genotypes, and in this study an additional goal was to identify genotypes that were also high-N fixing. It is particularly useful to examine genotype performance at Yorito, where growing conditions are representative of the marginal production regions in Honduras. At Yorito, 14 genotypes had above-average yields coupled with above-average SNF performance (Figure 4.5). These included four Honduran conventional varieties (HON56, HON53, HON55, and HON77), four landraces (HON2, HON49, HON65, and HON66) and six PPB varieties (HON5, HON26, HON35, HON57, HON72, and HON78). Of the six PPB varieties, five were developed through participatory varietal selection, and one was developed through participatory plant breeding. These genotypes are dispersed throughout the phylogenetic tree (Figure 4.2A), suggesting a lack of close genetic

relatedness; however, three of the high-yielding high-fixing PPB varieties share common genotypes in their pedigrees: HON5 and HON26 have a common landrace parent, Cincuentaño (HON48); HON26 shares a conventional parent, Tio Canela 75 (HON55), with HON78; and, HON5 has Amadeus 77 as a parent, which is a daughter of Tio Canela 75. Although the release-date information is incomplete for the PPBs (Table 4.3), there was no apparent relationship between release date and higher yields, nor was there a temporal trend for SNF performance.



**Figure 4.5.** Genotype performance for nitrogen fixation (x axis) and yield (y axis) at Yorito. For this trial, average yield was 799 kg ha<sup>-1</sup> and average nitrogen fixation was 43.3 %Ndfa, and these values are shown on the plot in dashed lines, dividing the plot into quadrants. Genotypes in the upper right quadrant are higher yielding and higher nitrogen-fixing.

## 4.5 Discussion

### 4.5.1 Genotype origins and pedigree explain Honduran panel structure

The patterns observed in the phylogenetic tree and STRUCTURE diagrams derived from the SNP compositions of the lines in the Honduran panel largely agree with what is known about their geographic origins and their pedigrees, but there are also a few exceptions. The two large groupings in the dendrogram based on SNP similarities (groups I–III and IV) show a clear separation (with some exceptions) between landraces (CRI) and material that has been conventionally bred or is the product of participatory plant breeding (PPB). Among the landraces, the small red beans that came from Yoro, Francisco Morazán, and Intibucá, were randomly interspersed throughout group III of the phylogenetic tree with no particular pattern, based on genotype origin. In contrast, clustering of genotypes by region of origin is found among the black bean landraces (groups I and II). Group I consists of the black bean landraces, which came from Intibucá, Yoro, Francisco Morazán, and Lempira, and contains the Milpero landraces (HON36, HON47, HON44, and HON39) and ‘Negro Opalaca’ (HON70), which are all from Intibucá, and ‘Negro’ (HON43), which is found alone on the next branch is from the neighboring department, Lempira [‘Merlot’ (HON62) is also found in this region of the tree and is discussed below]. These landraces are the most distantly related genotypes with respect to the rest of the panel. The Milpero landraces were daylight sensitive when grown at Elora in 2014, and their photoperiod sensitivity was likely inherited from a common ancestor. All photoperiod sensitive varieties in the panel may carry the dominant *ppd* gene responsible for control of this trait (Wallace et al., 1993). The

remaining black landraces from Yoro, Francisco Morazán, and Intibucá are found without any particular pattern throughout the next branch of the tree, and two are found among the small red landrace branches (HON06 and HON30). The STRUCTURE analysis shows that most of the black genotypes are admixed, suggesting a closer genetic relationship to the conventional and PPB germplasm. No small red landraces were found within the black landrace branches; however, two small red landraces, 'Concha Rosada' (HON02) and 'Rosado' (HON22), are displaced and found in group IV among the PPBs. The STRUCTURE analysis shows that Concha Rosada contains only ~10% genetic material from the landrace subpopulation, and Rosado is almost equally admixed between the landrace and PPB subpopulations. In the case of Concha Rosada, this may be explained by the fact that this landrace is widely grown and has been used as a parent in participatory breeding efforts and thus shares ancestry with many PPB varieties. Alternatively, Concha Rosada may not be a traditional landrace but instead a creolized variety derived from a formal-sector variety introduced to the Yorito region in the early 1980s (Humphries et al., 2005). The reason that Rosado is found among the PPB varieties in the tree is less apparent. Rosado has not been used as a parent for any of the PPB varieties in the panel according to the pedigree information available. Rosado recently arrived in the Yorito region, and a survey of bean farmers revealed that its origin is unknown (Humphries et al., 2015). According to M. Gomez (FIPAH), the initial population of Rosado showed phenotypic heterogeneity, and some selection has been made to create a uniform line for PPB breeding in collaboration with Zamorano. At Elora 2014, Rosado had uneven maturity, which may indicate that the

seed planted that year, and the seed grown for DNA extraction, was not a fixed homogeneous line, and this heterogeneity may have resulted in misplacement of this genotype in the phylogenetic tree.

The organization of the PPB branches of the phylogenetic tree can be explained in part by common ancestry. The Honduran conventional varieties, 'Tio Canela 75' (HON55) and 'Amadeus-77' (HON52), and the landraces, 'Estica' (HON11) and 'Vaina Rosada' (HON34), have been used frequently in generating the PPB genotypes in this panel. For example, five of the seven genotypes in the left-most PPB branch (HON19, HON18, HON15, HON16, and HON14) are derived from crosses with Tio Canela 75 or Estica, or both. The STRUCTURE plot indicates that these genotypes contain a greater proportion of conventional/PPB than landrace genetic material, and therefore, they are found within the PPB branches of the tree. However, six PPB varieties (HON10, HON05, HON12, HON25, HON33, and HON72) are found among the landrace branches of the tree, which indicates that they share greater genetic similarity to their landrace parent than the other genotypes in their pedigrees. For example, FPV 921-65 (HON33) has the landrace Vaina Rosada (HON34) and Amadeus 77 in its pedigree. FPV 921-65 is found in the same branch as Vaina Rosada among the landraces in the tree, and the STRUCTURE analysis shows that FPV 921-65 has more similarity to the landrace genetic subgroup. Marcelino (HON10) and Amilcar (HON05) also contain >50% landrace genetic material (Figure 4.2B) and are found among the landrace branches. Fourteen PPB genotypes have between 5% and 50% admixture with the landrace subgroup. Alleles favoring agronomic characteristics for local adaptation and

culinary traits contributed by landrace parents were likely prioritized under selection among PPB progeny from crosses between landrace and conventional genotypes, resulting in PPB varieties consisting of a large proportion of landrace genetic material.

Common parentage may explain the clustering of the five Honduran conventional genotypes (Amadeus-77, CENTA San Andrés, Tio Canela 75, Aifi Wuriti, and Carrizalito) in the panel. CENTA San Andrés, Amadeus-77, and Carrizalito have common ancestry. Tio Canela 75 is a parent of breeding line EAP 9510-77, which was released in the early 2000s in countries across Central America, including in El Salvador, as CENTA San Andrés, and in Honduras, as Amadeus-77. Tio Canela 75 is also a parent of EAP 9510-1, a sister line of EAP 9510-77, which was released in Honduras as Carrizalito in 2003 and as Telire in Costa Rica in 2004 (Rosas et al., 2004). Accordingly, these genotypes are closely associated in the phylogenetic tree, although some genetic differentiation has occurred between CENTA San Andres and Amadeus-77. The line 'MD 30-75' (released as Tio Canela 75) was used as a parent in generating four of the five conventional genotypes' mentioned above. This is likely a result of the effort to introduce BGMV resistance into Central American germplasm, as MD 30-75 is a highly resistant line, which carries the *bgm-1* resistance gene (Singh, 2005). DEORHO (HON53) was not genotyped in our study; however, it also has MD 30-75 in its pedigree, and it is reasonable to suppose that it would also appear in this region of the dendrogram. Aifi Wuriti was the only black conventional genotype included in the panel. Interestingly, it is most closely related to three small red genotypes (Carrizalito, Conan 33, and Cedrón) rather than the small black genotypes, although Aifi

Wuriti does not appear in the pedigrees of any of the small black PPB genotypes in the panel. The final Honduran conventional genotype, 'Dorado', is unusual because it is found among the landrace genotypes in the phylogenetic tree and the STRUCTURE analysis. This may be explained by the lack of a common genotype in its pedigree compared to the other conventional genotypes.

The placement of the North American check varieties (except OAC Rosito) as a separate group within the dendrogram is consistent with the unique alleles that they would be expected to have relative to the genotypes in this panel. The location of the check varieties in the landrace portion of the tree (groups I and II) may reflect the genetic diversity and wide-ranging geographic origins of the germplasm used in the University of Guelph/AAFC and Michigan State University/USDA-ARS bean breeding programs, which are aimed at introducing resistance to abiotic and biotic stresses and improving various agronomic traits. The locations of the North American checks interspersed throughout the admixed portion of the STRUCTURE plot also indicate that genetic diversity has been retained in these modern North American genotypes. OAC Rosito is a special case, since it was derived from El Salvadoran germplasm, and it is found among the PPBs in the phylogenetic analysis where it is most closely related to two PPB genotypes, Campechano (HON57) and Quebradeño (HON26). According to the STRUCTURE analysis, OAC Rosito is more genetically similar to the Honduran conventional/PPB sub-group than the landraces. This suggests either that El Salvadoran landraces are not similar to Honduran landraces, or more likely that the landrace population from which OAC Rosito was developed was actually a creolized

conventional variety. Germplasm provided by Zamorano has been used in El Salvadoran variety development since the early 2000s (Reyes, 2012), and this could explain the genetic similarity of OAC Rosito to the PPB varieties in our panel, which were developed in collaboration with Zamorano.

#### **4.5.2 Optimizing use of genetic diversity of Honduran landraces and PPB varieties**

The larger nucleotide diversity among the landraces ( $\pi = 3.20 \times 10^{-4}$ ) than observed in the PPBs ( $\pi = 2.89 \times 10^{-4}$ ) in the current study is consistent with general observations that landraces are more diverse than materials that are products of selection (Haudry et al., 2007; Maron et al., 2013; Bellucci et al., 2014; Zhou et al., 2015). However, other studies that compared diversity in wild to domesticated bean accessions found wider diversification between those groups than we found between landraces and PPBs in our panel. Nanni et al. (2011) reported that within the Mesoamerican gene pool, nucleotide diversity was 3.2 times higher among wild genotypes ( $\pi = 17.34 \times 10^{-3}$ ) than domesticated genotypes ( $\pi = 5.43 \times 10^{-3}$ ). The difference between landraces and PPBs in the current study was only 1.1 times, probably because of the small population size and because these genotype groups do not represent extremes of the diversity continuum that was sampled in the previous study. Both landraces and PPBs are selections from wild accessions, and the PPBs have probably not been as strongly selected as conventionally derived varieties.

The high level of diversity in Honduran landraces suggests they could be a source of novel alleles that could be used in breeding to improve various traits.

Landraces are adapted to the environmental conditions of the locations where they were maintained, in some cases, over thousands of years. Landraces that were grown in fluctuating environments and in low-input agricultural systems may be enriched for rare alleles enabling phenotypic plasticity and inherent responsiveness to diverse abiotic and biotic stresses (Dwivedi et al., 2016). Landraces, in regions where they are still grown, have often been pushed to marginal production environments where their performance often exceeds that of modern varieties (Ceccarelli, 1994, 1996; Humphries et al., 2015).

Dry bean landrace germplasm across Mesoamerica is genetically diverse (Beebe et al., 2000). Soil conditions across this region are poor, and the terrain ranges from low to high altitude with steep slopes, leading to certain trait adaptations in the landraces. For example, 'Common Red Mexican', a red-seeded landrace from Mexico, has been found to be drought resistant (Muñoz-Perea et al., 2006), while 'Puebla 152', a black-seeded landrace also from Mexico, has superior SNF capacity (Bliss, 1993). Originating in the Andean region, G19833, a 'Chaucha Chuga' landrace from Perú, has tolerance to high levels of soil aluminum and low levels of phosphorus (Dwivedi et al., 2016) and resistance to a number of bean diseases (Broughton et al., 2003; Hurtado-Gonzales et al., 2017). Our survey of the literature found genes in high landrace  $\pi$  regions associated with abiotic stress tolerance, phosphorus use efficiency, and nitrogen fixation (Appendix B, Table S4.1). Conservation of landraces and mobilization of the unique genetic diversity they contain through plant breeding can help address the future need for higher yielding and climate resilient varieties.

#### **4.5.3 Regions of high genetic differentiation indicate regions impacted by selection**

Genetic divergence between the PPB and landrace subpopulations in the panel is indicated by regions of high genetic differentiation ( $F_{ST}$ ), and these genomic regions may contain loci that have been subjected to selection pressure. We identified several regions on chromosomes Pv02, Pv07, Pv09, and Pv11 where  $F_{ST}$  values exceeded 0.5. Similarly, in a study of genetic diversity of Italian bean landraces, Lioi et al. (2019) reported that genomic regions related to domestication were concentrated on Pv02, Pv07, and Pv09 for Mesoamerican types. For comparisons between wild and domesticated bean landraces, Papa et al. (2005) also reported significantly larger levels of  $F_{ST}$  differentiation around genomic regions associated with domestication. While the genetic distance between the Honduran landraces and PPB genotypes included in our study is not likely as wide as that between the wild and landrace genotypes investigated by Papa et al. (2005), similar trends towards genetic differentiation between landrace and PPB genotypes developed with modern breeding objectives and germplasm could be expected.

In particular, the genomic regions with large  $F_{ST}$  differences may be associated with traits that were a focus of selection in PPB breeding. However, an extensive search of the recent bean literature did not reveal any known QTL associated with agronomic, SNF, or WUE traits that are located within the regions of large  $F_{ST}$  differentiation identified in this study. This may be because this is the first genomic survey study of Honduran material, and the distinguishing traits between landraces and PPB materials

are specific to materials from that region or expressed in that location. In particular, because we are comparing two domesticated groups of genotypes, namely farmer traditional landraces and PPB varieties, the genes underlying the regions of large differentiation found in our study could be those responsible for local adaptation, culinary qualities, and favorable plant traits, rather than traits associated with domestication (Schmutz et al., 2014). Additionally, the conventional germplasm used to generate PPB genotypes, either through crosses with landraces or through varietal selection, is largely limited to material in the Zamorano breeding program, which may have a specific genetic architecture.

In spite of the lack of previous QTL evidence for selection for domestication in the high  $F_{ST}$  regions, several genes in those regions that have been studied for various reasons may be associated with domestication. For example, disease resistance genes, such as those found on Pv02 for Rhizoctonia resistance (Phvul.002G323708, Phvul.002G323712; Oladzad et al., 2019), Halo blight resistance (Phvul002G326200; Tock et al., 2017), and Anthracnose resistance (Phvul.002G328300; Zuiderveen et al., 2016), and on Pv08 for Anthracnose resistance (Phvul.008G019600; Padder et al., 2016), and for the bean-rust interaction (Phvul.008G270500; Ayyappan et al., 2015), have been associated with domestication in several crops (Zheng et al., 2016; Ma et al., 2019). Genes controlling agronomic traits have been identified in domestication studies in other crops (Varshney et al., 2017, 2019; Zhang et al., 2017). Additionally, genes that affect survival in diverse growing conditions may have also been favored over the course of domestication. Two such genes are located in a region of high genetic

diversity on Pv08; the ethylene-responsive transcription factor (Phvul.008G019600; Soltani et al., 2018), an ortholog of an Arabidopsis gene known to be involved in flooding tolerance (Liu et al., 2012), and the transcription factor IIIA (Phvul.008G270400), which is upregulated in phosphorus-restricted conditions (Silva et al., 2019). Soltani et al. (2018) suggest that further studies are needed to understand the process of local adaptation and allelic selection using bean landraces and wild populations. Insight to develop climate-resilient crops can be drawn from the study of crop adaptation under natural selection and domestication (Henry and Nevo, 2014; Henry, 2019).

#### **4.5.4 Landraces are superior nitrogen fixers**

Although some genotype- and environment-influenced variability was seen in our study, our examination of symbiotic nitrogen fixation in the Honduran panel revealed a wide range of capacity for this trait. The superiority of the landraces for SNF capacity at all trial locations may be the consequence of the continual selections of these materials under conditions of low soil fertility endemic to Central America. Even today, these materials continue to be grown by small scale farmers who do not have access to fertilizer inputs. Under these conditions, bean genotypes that have developed efficient associations with nitrogen-fixing bacteria would have a larger source of nitrogen for metabolic processes and better phenotypic fitness compared to poor nitrogen-fixing genotypes. Strong nitrogen-fixers would have a competitive advantage in the low input environments and would likely be preferentially selected over time. There may be parallels between the selection pressures during landrace evolution and the selection of

heirloom bean varieties, which have also been shown to have strong SNF capacity (Wilker et al., 2019).

There are few studies that have investigated SNF capacity of bean landrace genotypes. Heilig et al. (2017a) used 'Puebla 152', a black-seeded Mexican landrace known for its nitrogen fixing capacity (Chaverra and Graham, 1992), in a cross with conventional genotype 'Zorro' to create a RIL population to study SNF. The authors found that Puebla 152 fixed between 13.0 to 45.5 % of the nitrogen in samples (seed + biomass), which was slightly more than Zorro, which fixed between 5.4 to 44.4 % (Heilig et al., 2017a). Many landraces in our study fixed more N than Puebla 152, indicating that Honduran landraces may be a useful source of SNF capacity. The SNF performance of Zorro ranged from 47.7 to 53.0 %Ndfa in our study, a mid-range performance among our check genotypes, and overall better than its performance in Heilig's study (Heilig et al., 2017a).

The SNF performance of the progeny of the cross between the conventional genotype Zorro and the landrace Puebla 152 may be predictive of the performance of PPB varieties that are crosses between Honduran conventional varieties and landraces. The SNF performance of Honduran PPB varieties ranged between 20.5 to 55.7 %Ndfa at Yorito. Although the focus of the participatory breeding program between FIPAH and Zamorano has been to generate higher-yielding genotypes, rather than on improving SNF performance, the SNF capacity seen for the PPB varieties falls within the upper range of that found for the RILs in Heilig's study (Heilig et al., 2017a).

For the Honduran panel, the insights gleaned from Yorito are of particular interest because this location has growing conditions representative of small-scale growers across the region, and as much as possible, local growing practices were employed in the trial. At Yorito, there was a range in SNF performance in the landraces, and overall genotypes belonging to this breeding history group performed better than the others studied. In addition, PPB genotypes derived from crosses with the best SNF-performing landraces had strong SNF capacity. Zamorano used the methods of participatory varietal selection to develop these PPB varieties with CIALs, enabling local growers to evaluate genotype performance on their farms. For example, Amilcar and San Jose were tested by various CIALs through the regional adaptation nursery (VIDAC, Vivero de Adaptación Centroamericano) in the mid-2000s. Amilcar has the high-SNF performing landrace Cincuentaño in its pedigree. Generally, native Rhizobia inhabit tropical soils and farmers do not use Rhizobia inoculants, although Zamorano disseminates SNF knowledge through the CIALs, including effective Rhizobia inoculants and protocols for use. In addition, the breeding program has the capacity to test SNF performance in '*bancales*' where soil N levels are low and SNF-related traits, such as nodulation, can be observed. PPB breeding for enhanced SNF capacity could be expanded if grower demand and the threat of climate change and resulting raising input costs warrant it.

In addition, the range in SNF performance among the conventionally bred North American checks and Honduran conventional varieties was wide. The superior SNF capacity Merlot exhibited at Yorito would suggest it has value as a breeding parent for

this trait in Honduras; however, it has a larger seed size and a dark red seed coat; traits that are less preferred by Honduran consumers and could be challenging to select against in a breeding program. Of the Honduran conventional genotypes, DEORHO (HON53) fixed the highest amount of N (56.8%), but it performed poorly in Elora fixing 34.4% (2014) and 52.6% (2015) of its N. It is not found in the pedigrees of any PPB genotype included in our panel, but DEORHO has been a popular variety in commercial growing regions of the country. It has disease resistance, high yield, and the preferred light red seed coat color, making it a good candidate for future PPB breeding efforts.

#### **4.5.5 PPB genotypes have superior water use efficiency values**

Plants that have higher water use efficiency (WUE) are more drought tolerant, and WUE can be estimated using carbon differentiation ( $\Delta^{13}\text{C}$ ) values measured from plant biomass. During photosynthesis, plants discriminate against the incorporation of the heavy C isotope ( $^{13}\text{C}$ ), depleting  $^{13}\text{C}$  in plant biomass and driving lower  $\Delta^{13}\text{C}$  values (Moghaddam et al., 2013). Plants with comparatively low biomass  $\Delta^{13}\text{C}$  values can be considered more drought tolerant. WUE has been studied in beans, including landraces. A study by Muñoz-Perea et al. (2007) of the WUE of 16 dry bean genotypes in drought-stressed and non-stressed environments found that the two landraces included differed in their responses to drought stress, but Common Red Mexican was among the best performers under drought stress conditions. In contrast, in our study, the significantly lower  $\Delta^{13}\text{C}$  values measured in Yorito for the PPB genotypes than the landraces indicate that the PPB varieties in our panel may be more resilient to drought conditions than the landraces.

The drought resistant characteristics of the PPBs were likely contributed by the conventional parents. For example, PM2-Don Rey (HON23) was the most WUE PPB genotype at Yorito, with a  $\Delta^{13}\text{C}$  value of 16.36 ‰. PM2-Don Rey was developed through PPB methods from a cross between the landrace, Paraísito, and the Honduran conventional variety, Carrizalito. It was released as a drought-resistant variety in 2016 (Drought-Resistant Bean Variety Offers Hope to Central American Communities, 2016), and it is the first variety from the EAP-Zamorano-CIAL PPB collaborations to be released at the national level. A second PPB genotype, Marcelino (HON10), was developed through participatory varietal selection, and it had similar WUE (16.43 ‰) to PM2-Don Rey. The PPBs FPY-724-43 (HON16; 17.03 ‰), Cedrón (HON03; 17.41 ‰), and Amilcar (HON05; 17.42 ‰) have the next best WUEs.

Three landraces had WUE values below 17.5 ‰, including Concha Rosada (HON02; 16.85 ‰), Chapin Rojo (HON27; 16.86 ‰), and Chirineño (HON67; 17.30 ‰). Concha Rosada is of particular note because it is favored by poor farmers for its early maturity, which allows it to escape drought conditions late in the growing season (Humphries et al., 2015; Gomez et al., 2020). Our study indicates that Concha Rosada not only has drought resistance through ‘drought escape’ but also has WUE characteristics that enable it to survive drought.

Among the conventionally bred genotypes, Carrizalito (HON77) and OAC Rosito (HON63) were the most water use efficient with low  $\Delta^{13}\text{C}$  values of 17.47 ‰ and 17.54 ‰ at Yorito. Carrizalito is a commercial Honduran variety, and it was used as a parent

contributing disease resistance, agronomic, and likely WUE traits, to PM2-Don Rey. Among the check genotypes, OAC Rosito, recently developed at the University of Guelph from an El Salvadoran plant introduction, had the best WUE performance. This characteristic is likely retained from its domestication in Central America, and this enabled it to outperform the other check genotypes that have been developed for production in the Great Lakes region of North America.

In the coming decades, the effects of climate change are predicted to bring drier conditions to Honduras, and drought-resistant crops will help protect yields through periods of minimal rainfall. It has been proposed that WUE can be improved through selection and breeding. In alfalfa, evaluating genotypes for  $\Delta^{13}\text{C}$  and selecting for lower  $\Delta^{13}\text{C}$  values has been used to improve WUE in this important forage species (Moghaddam et al., 2013). The current results indicate that there is variation among the Honduran PPB and conventional bean germplasm in WUE traits, and selecting for lower  $\Delta^{13}\text{C}$  values could be applied to beans in Honduras to generate improved varieties that are more resilient to drought conditions.

#### **4.5.6 Conventional genotypes have superior yields**

Releasing varieties with higher yields is the objective of modern breeding programs, and our trial results suggest that improvements impacting yield have been made along the breeding history continuum from landraces to PPBs to conventional varieties. Considering the Honduran germplasm, the landraces were the lowest-yielding group at Yorito, followed by PPBs and the conventional varieties. This result is counter

to the findings of early experiments performed by CIALs, where landraces out-yielded conventional materials (Humphries et al., 2005). However, our trial was conducted at the FIPAH office in Yorito, where soil fertility is less restrictive and the plot is flat, whereas the early CIAL trials were conducted in farmers' fields, which have low-fertility soils and sloped land; conditions for which conventional materials were not developed. The superior performance of the conventional materials in our trial at Yorito is consistent with the aim of modern breeding practices in generating higher yielding varieties. The North American checks were the highest yielding at Elora 2014, as they were bred for performance in this region, whereas the Honduran conventional and PPBs performed poorly at Elora 2014. The landraces also performed well at Elora 2014, and this may be attributed to phenotypic plasticity resulting from retention of useful nucleotide diversity enabling them to perform well in a new environment.

#### **4.5.7 Utility of panel genotypes for breeding**

The different breeding and/or selection histories for the materials contributing to the phenotypic diversity present in the Honduran panel may provide opportunities for improving different traits in beans in the same way that a number of studies with different crops have found unique benefits from the use of landraces. In wheat, for example, cultivation of landraces in low-input systems has led to the conservation of traits that increase the duration of photosynthesis, which can lead to an increase in grain yield (Lopes et al., 2015). In a study comparing barley landrace and modern variety performances under stress conditions, the landraces were higher yielding and were less likely to fail outright (Ceccarelli, 1994). The advantage of using landraces as

parents in breeding programs has also been explored. In a study examining barley yields under drought conditions, progeny from crosses using landrace genotypes were found to be higher yielding than progeny from crosses without landraces in their pedigrees (Dwivedi et al., 2016). The authors concluded that breeding crops for vulnerable environments could be enhanced by identifying landrace alleles associated with yield performance and abiotic stress adaptation and employing these in breeding programs (Dwivedi et al., 2016).

In the current study, landraces, which had superior symbiotic nitrogen fixation characteristics could be excellent sources of novel alleles for this trait. Similarly, PPB materials, which had superior WUE, and conventional varieties, which had superior yields within their target environments, might be exploited, respectively, for these purposes. In general, all the germplasm types that were tested represent useful resources for breeding for important traits in the face of climate change and increasing production costs/demands.

The diversity for SNF capacity inherent in Honduran bean landraces, and their unique adaptation to the microclimates where they are grown, leads us to conclude that the inclusion of landrace germplasm in breeding for enhanced SNF would produce high fixing genotypes with growth and culinary characteristics already accepted by small-scale bean growers.

## 4.6 Conclusion

The aim of our study was to evaluate a large set of Honduran landraces and varieties generated through participatory plant breeding, as well as check conventional genotypes, to ascertain their value in future breeding efforts. We used simple genomics and phenotyping to characterize the panel. Our genetic analyses found that the panel is divided into predominantly-landrace and predominantly-PPB groupings, with Honduran conventional genotypes sharing most similarity to the PPBs. Breeding history and pedigrees account for this division. The genetic diversity analysis revealed that landraces have retained a higher level of nucleotide diversity than PPB genotypes, which we attribute to selection pressure imposed by breeding for different production environments/objectives, and the use of a small number of conventional/elite parents in breeding efforts. The nucleotide diversity inherent in landraces can be used to increase the frequency of rare alleles in breeding programs. Beyond genetic characterization, it is important to classify germplasm for trait phenotypic diversity, which could be employed in breeding, because landraces that have evolved in adverse environments contribute adaptive traits to variety development.

Two traits that contribute to climate resiliency, nitrogen fixation capacity, and water use efficiency were evaluated in our study. Genotypes with good nitrogen fixation capacity are an asset for remote hillside growers who have limited funds and limited opportunity to purchase inputs because of poor market access. Landraces were shown to have superior SNF capacity and are already favored by hillside producers. Genotypes with enhanced water use efficiency will also be an asset to hillside growers

in a future with drier and more changeable growing conditions. The PPB and conventional varieties in our study show promising characteristics for drought resilience. Further evaluation of PPB varieties under drought conditions is warranted.

Honduran bean production will continue to be carried out predominantly by small-scale hillside producers. The widely grown farmer landraces are locally adapted and accepted by consumers, and future breeding efforts should deliver varieties that maintain the inherent SNF capacity of landraces, while enhancing drought resilience and producing high yields. PPB methods employed in the breeding efforts between EAP-Zamorano and FIPAH- and PRR-supported CIALs have succeeded in generating promising PPB varieties. One variety that combines these characteristics is Amilcar (HON05). It is a small red PPB variety selected among germplasm provided by Zamorano for improvement by a CIAL in Yoro. Ultimately, Amilcar has been widely accepted because of its commercial value, culinary qualities, and disease resistance, but its climate resiliency traits and yield potential are of critical importance. Landrace characterization, conservation, and employment in breeding programs will bring continued benefits.

**5 Symbiotic Nitrogen Fixation and related QTL in a  
*Phaseolus vulgaris* (L.) Middle American Diversity Panel**

## 5.1 Abstract

Common bean (*Phaseolus vulgaris* L.) is one of the most important legumes grown world-wide for human consumption. The symbiotic nitrogen fixation (SNF) capacity of legumes reduces the need for agricultural N<sub>2</sub> fertilizer inputs, however common bean is widely thought to have poor SNF capacity. To examine the range in SNF capacity and understand the genetic architecture of the trait, a genome-wide association study (GWAS) was carried out using the 280-genotype Middle American Diversity panel (MDP). The MDP was evaluated for SNF and related traits in three low nitrogen field trials, and separate GWAS were performed on the entire panel and the two panel subpopulations using >54 K SNPs. Significant SNPs were identified on chromosomes Pv07, Pv08 and Pv10 for %Ndfa in the seed in the entire MDP. Twenty, SNPs were identified across the genome in the MDP for seed N content. Significant SNPs for %Ndfa were found in race Mesoamerica (MA) panel subpopulation, while significant SNPs for seed N content were found in race Durango-Jalisco (DJ). Two protein kinase family genes (Phvul.008G021001 and Phvul.008G084600) on Pv08 were identified as candidate genes for %Ndfa. Protein kinases are a large family of plant enzymes involved in diverse cellular regulation and metabolic processes, and some play roles in establishing effective rhizobia-host plant interactions. Within the MA subpopulation at Elora a TGACG-sequence-specific DNA-binding protein gene (Phvul.008G023800) was identified as a candidate gene associated with %Ndfa. Candidate genes associated with seed N content included six cytochrome P450 genes (Phvul.002G126000, Phvul.002G126100, Phvul.002G126200, Phvul.007G257300,

Phvul.007G257400, Phvul.007G257500), which are known to play roles in diverse metabolic processes, potentially including symbiosis. This study provides important insights into the genetic control of SNF in the Middle American gene pool of common bean.

## 5.2 Introduction

Dry beans (*Phaseolus vulgaris* L.) are a staple food crop worldwide, providing an affordable source of dietary protein, a source of income to growers and processors, and diverse ecosystem services such as the maintenance of soil fertility and pollinator attraction. Dry beans belong to Andean (AN) or Middle American (MA) gene pools (Papa and Gepts, 2003; Chacon et al., 2005; Kwak and Gepts, 2009; Rossi et al., 2009; Mamidi et al., 2011; Nanni et al., 2011; Bitocchi et al., 2012, 2013, 2017; Schmutz et al., 2014; Rendón-Anaya et al., 2017). According to the Food and Agriculture Organization, in 2018 Canada was among the top 20 bean-producing countries in the world and produced 341,100 tonnes of dry beans on 137,100 ha, primarily in the provinces of Ontario, Manitoba, Saskatchewan, Alberta and Quebec (Pulse Canada, nd; FAO, 2018). The navy bean market class comprises the largest proportion of the Canadian bean market, and pinto, great northern, kidney, cranberry, yellow, red and black beans are also grown for commercial production. More than 80% of the beans produced in Ontario are exported (Ontario Bean Growers, nd).

Symbiotic nitrogen fixation (SNF), also termed biological nitrogen fixation, occurs in beans when nitrogen fixing Rhizobia enter bean roots, stimulating nodule formation

and the differentiation of the bacteria into bacteroids that convert atmospheric nitrogen to ammonia, which is transferred to the plant for use in biological processes. Beans have a range of nitrogen fixing capacities and in some production environments SNF capacity can reach yield-sustaining levels (Farid and Navabi, 2015; Kamfwa et al., 2015; Diaz et al., 2017; Heilig et al., 2017b; Barbosa et al., 2018; Wilker et al., 2019; Aserse et al., 2020; Reinprecht et al., 2020; Sanyal et al., 2020). However, conventional production practices in North America routinely include fertilization with 50-100 kg ha<sup>-1</sup> of nitrogen fertilizer to ensure optimal yields (OMAFRA, 2009). SNF is inhibited by excessively low soil N availability, phosphorus and micronutrient deficiencies, excessively hot or cool temperatures, drought or flooding conditions and symbioses with inefficient bacteria.

As a complex biological relationship, low SNF rates in beans have been attributed to environmental factors, low bacterial efficacy, plant deficiencies or a combination of these factors (Unkovich and Pate, 2000). A possible explanation for the relatively low SNF rates in dry bean compared to other crop legumes is that breeding under nitrogen-replete, SNF-dampening conditions has led to the passive loss of SNF capacity in elite varieties; beans that lack the need to form symbiotic relationships with bacteria to provide them nitrogen lose the capacity to do so over time. However, recent studies using Mesoamerican and Andean germplasm have reported genotypes that fix more than 75% of their nitrogen from the atmosphere under favorable conditions (Farid and Navabi, 2015; Kamfwa et al., 2015; Diaz et al., 2017; Heilig et al., 2017b; Barbosa et al., 2018; Wilker et al., 2019; Aserse et al., 2020; Reinprecht et al., 2020; Sanyal et

al., 2020). The presence of diversity for SNF suggests that breeding for enhanced SNF performance in bean is possible, and it enables the discovery of genomic regions associated with SNF traits.

Complex biological traits are controlled by multiple genetic loci, termed quantitative trait loci (QTL). Two approaches, which have been used to identify QTL, are linkage mapping and genome-wide association studies (GWAS). Linkage mapping uses recombinant inbred line (RIL) populations generated from a cross between parent genotypes, which differ significantly for the trait of interest. Multiple studies using QTL mapping to investigate SNF traits in dry bean have been reported (Ramaekers et al., 2013; Farid, 2015; Diaz et al., 2017; Heilig et al., 2017a; Kamfwa et al., 2019; Souza et al., 2019). They have identified QTL on almost every bean chromosome associated with SNF, but in most cases the QTL discovered are not common between studies. First, this shows the genetic complexity and genotype by environment interaction that impacts SNF capacity, and second, it highlights a significant limitation of QTL mapping wherein the results are most relevant to the specific parents and RILs studied and may not account for genetic variation for the trait in other germplasm. In contrast, the diversity panels used in GWAS studies are comprised of >100 unrelated individuals, and QTL associated with traits identified in these studies find utility in a wider range of germplasm.

For GWAS, a large panel of germplasm is genotyped and investigated for QTL associated with a trait. SNP genotyping can be carried out using a SNP-chip or by

performing genotyping by sequencing (GBS). The SNP chip developed for dry bean (Illumina BARCBear6K\_3 BeadChip) contains 5,398 SNPs (Song et al., 2015). In contrast to genotyping with a SNP chip, GBS generates a large number of SNPs, which increases the marker density coverage across the genome, lending more precision to QTL discovery and candidate gene identification. The 259-genotype Andean Diversity panel was studied for QTL associated with SNF using GWAS by Kamfwa and colleagues (Kamfwa et al., 2015). In that study, the BARCBear6K\_3 chip was used for genotyping and QTL associated with SNF capacity measured in the greenhouse and the field were identified on Pv02, Pv03, Pv07, Pv09 and Pv10 (Kamfwa et al., 2015).

To expand the knowledge of the genetic basis of SNF in market classes primarily grown by North American producers, the present study was carried out using the 280-genotype Middle American Diversity Panel (MDP). The MDP is a subset of a large association mapping panel assembled by the Bean Coordinated Agricultural Project (BeanCAP, 2009-13) that consists of diverse modern bean genotypes representative of bean market classes grown in the major production regions of the United States (<http://arsfiftbean.uprm.edu/beancap/>). Moghaddam et al. (2016) determined that at  $K = 2$  the MDP was divided into race subpopulations, consisting of 182 race Durango-Jalisco (DJ) and 98 race Mesoamerica (MA) genotypes, and at  $K = 7$  clustering according to market classes was apparent with some admixture observed. The MDP was genotyped using the genotyping-by-sequencing (GBS) approach at North Dakota State University as described by Oladzad et al. (2019). Briefly, a two-enzyme protocol (MseI and Taqα1; Schröder et al., 2016) was used to generate SNP libraries from 482

Middle American genotypes from multiple bean diversity panels. SNPs were processed for quality and mapped to the *Phaseolus vulgaris* v2.1 reference genome ([https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Pvulgaris](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvulgaris)). Imputation was carried out for SNPs with <50% missing data (fastPHASE; Scheet and Stephens, 2006). The MDP has been used to investigate agronomic traits, nutritional traits, seed coat colour traits, and abiotic stress response (Hoyos-Villegas et al., 2016; Moghaddam et al., 2016, 2018; McClean et al., 2017, 2018; Soltani et al., 2017). The objective of this study was to use the GWAS approach combined with GBS genotyping to identify QTL associated with SNF in the 280-genotype Middle American Diversity Panel.

## **5.3 Materials and Methods**

### **5.3.1 Germplasm**

The association panel used in this study consisted of the 280-genotype Middle American Diversity panel (MDP) and a non-nodulating mutant R99. R99 was derived from navy bean OAC Rico by EMS mutagenesis (Beversdorf, 1984; Park and Buttery, 2006). R99 provided a baseline for comparison of N-fixation capacity of N-fixing genotypes. R99 was included in the trial design and was grown in plots bordering the trial. Genotype information including variety name and market class is presented in Tables 5.1 and 5.2. Seed for the majority of genotypes in the panel was sourced from the University of Saskatchewan's bean breeding program (Dr. K. Bett), and seed for a small proportion of the panel was provided by the bean breeding programs at the University of Nebraska (Dr. C. Urrea), Agriculture Agri-Food Canada (Harrow, ON; T. Rupert), and from the University of Guelph.

**Table 5.1.** Middle American Diversity Panel (MDP) genotypes (n = 98) belonging to race Mesoamerica including navy, black and carioca market classes.

Entry	Variety	Market Class	Entry	Variety	Market Class	Entry	Variety	Market Class	Entry	Variety	Market Class
4	AC Black Diamond	Black	31	Mackinac	Navy	56	Morden 003	Navy	82	CDC Whitecap	Navy
5	AC Polaris	Navy	32	Phantom	Black	57	A-55	Black	83	CDC Espresso	Black
6	Xan 176	Black	33	Jaguar	Black	58	I9365-31	Black	84	CDC Jet	Black
7	Morales	Small White	34	Seahawk	Navy	59	92BG-7	Black	85	OAC Rex	Navy
8	Verano	Small White	35	Condor	Black	60	ICB-10	Black	86	Nautica	Navy
9	DPC-4	Black	36	Zorro	Black	61	ICB-3	Black	87	Fleetwood	Navy
10	PR 0443-151	Black	37	B05055	Black	62	NW-395	Small White	88	OAC Laser	Navy
11	Aifi Wuriti	Black	38	T-39	Black	63	Hyden	Small White	89	OAC Gryphon	Navy
12	F07-004-9-1	Navy	39	Cornell 49-242	Black	64	USWA-50	Small White	90	Lightning	Navy
13	F04-2801-4-6-6	Black	40	N05324	Navy	65	Orca	Black mottle	91	Harrowhawk	Black
14	F04-2801-4-5-1	Black	41	115M (Black Rhino)	Black	66	A285	Carioca	92	AC Harblack	Black
15	F04-2801-4-1-2	Black	42	Puebla 152	Black	67	A801	Carioca	94	AC Compass	Navy
17	Sanilac	Navy	43	Shania	Black	68	BAT 477	Tan	95	T9905	Navy
18	Seafarer	Navy	44	Bandit	Black	69	SEA 10	Tan	96	T9903	Navy
20	Neptune	Navy	45	Loreto	Black	71	Norstar	Navy	97	HY 4181	Navy
21	Domino	Black	46	Schooner	Navy	72	Avalanche	Navy	98	UCD 96114	Small White
22	Black Magic	Black	47	Ensign	Navy	73	Eclipse	black	99	UCD 9634	Black
23	Bunsi	Navy	48	Voyager	Navy	74	ND021717	Black	100	Albion	Navy
24	C-20	Navy	49	Seabiskit	Navy	75	ND021574	Black	101	Avanti	Navy
25	Laker	Navy	50	Medalist	Navy	76	Crestwood	Navy	102	Midland	Navy
26	Mayflower	Navy	51	Navigator	Navy	77	Reliant	Navy	103	Black Velvet	Black
27	Blackhawk	Black	52	Midnight	Black	78	Vista	Navy	104	McHale	Navy
28	Huron	Navy	53	Black Knight	Black	79	Envoy	Navy	120	Shiny Crow	Black
29	Raven	Black	54	UI-906	Black	80	Blackjack	Black			
30	Newport	Navy	55	UI-911	Black	81	CDC Nighthawk	Black			

**Table 5.2.** Middle American Diversity Panel (MDP) genotypes (n = 182) belonging to race Durango-Jalisco including great northern, pinto, small red, and pink market classes.

Entry	Variety	Market Class	Entry	Variety	Market Class	Entry	Variety	Market Class	Entry	Variety	Market Class
1	BelMiNeb-RMR-7	Great Northern	129	CENTA Pupil	Small Red	154	Remington	Pinto	182	UI-196	Pinto
2	BelMiNeb-RMR-4	Great Northern	130	INTA Precoz	Small Red	155	La Paz	Pinto	183	UI-228	Small Red
3	BelMiNeb-RMR-8	Great Northern	131	Dehoro	Small Red	156	Baja	Pinto	184	UI-239	Small Red
105	BelMiNeb-RR-2	Great Northern	132	TARS09-RR004	Small Red	157	Durango	Pinto	186	UI-59	Great Northern
106	BelMiNeb-RMR-3	Great Northern	133	TARS09-RR007	Small Red	158	Sonora	Pinto	187	UI-111	Pinto
107	BelDakMi-RR-5	Pinto	134	TARS09-RR023	Pinto	159	Pink Floyd	Pink	188	UI-114	Pinto
108	BelNeb-RR-1	Great Northern	135	TARS09-RR029	Small Red	160	Red Ryder	Small Red	191	BelMiNeb 1	Great Northern
109	AC Redbond	Small Red	136	F07-449-9-3	Small Red	162	Beryl	Great Northern	192	BelMiNeb 2	Great Northern
110	AC Island	Pinto	137	F07-014-22-2	Small Red	163	Beryl R	Great Northern	193	BelMiNeb 5	Great Northern
111	AC Early Rose	Pink	138	I06-2575-17	Pinto	164	Marquis	Great Northern	194	GN#1Sel27	Great Northern
112	AC Resolute	Great Northern	139	Sierra	Pinto	165	Sapphire	Great Northern	195	GN Harris	Great Northern
113	AC Earlired	Small Red	140	Aztec	Pinto	167	Garnet	Small Red	196	GN Star	Great Northern
114	Bill Z	Pinto	141	Kodiak	Pinto	168	ROG 312	Pink	197	Tara	Great Northern
115	Ouray	Pinto	142	Matterhorn	Great Northern	169	Desert Rose	Flor de Mayo	198	Starlight	Great Northern
116	Grand Mesa	Pinto	143	Santa Fe	Pinto	170	SDPI-1	Pinto	199	Emerson	Great Northern
117	Fisher	Pinto	144	P07863	Pinto	171	Shoshone	Pinto	201	ABC-Wei hing	Great Northern
118	Montrose	Pinto	145	Merlot	Small Red	172	UI-3	Small Red	203	ABCP-8	Pinto
119	Olathe	Pinto	146	Sedona	Pink	173	UI-37	Small Red	204	Chase	Pinto
122	Croissant	Pinto	147	S08418	Pink	174	UI-537	Pink	205	ABCP-15	Pinto
123	Arapaho	Pinto	148	Poncho	Pinto	175	Common Pinto	Pinto	206	ABCP-17	Pinto
124	DOR 364	Small Red	149	Topaz	Pinto	177	IP08-2	Pinto	207	NE1-09-13	Great Northern
125	PR 0340-3-3-1	Small Red	150	Buckskin	Pinto	178	Kimberly	Pinto	208	NE1-09-19	Great Northern
126	Amadeus 77	Small Red	151	Flint	Pinto	179	Sawtooth	Great Northern	209	NE1-09-20	Great Northern
127	PR 0401-259	Pink	152	Fargo	Pinto	180	UI-123	Great Northern	210	NE1-09-22	Great Northern
128	IBC 301-204	Small Red	153	Agassiz	Pinto	181	UI-126	Pinto	211	NE2-09-1	Pinto

Entry	Variety	Market Class	Entry	Variety	Market Class	Entry	Variety	Market Class	Entry	Variety	Market Class
212	NE2-09-3	Pinto	238	USPT-CBB-3	Pinto	267	Lariat	Pinto	293	Fiesta	Pinto
213	NE2-09-4	Pinto	239	USPT-ANT-1	Pinto	268	Stampede	Pinto	295	Buster	Pinto
214	NE2-09-8	Pinto	240	USPT-CBB-5	Pinto	269	ND-307	Pinto	296	Medicine Hat	Pinto
215	NE2-09-10	Pinto	242	Le Baron	Small Red	270	Frontier	Pinto	297	Windbreaker	Pinto
216	NE1-09-9	Great Northern	243	NW-63	Small Red	271	NDZ06249	Small Red	298	Mariah	Pinto
217	AC Pintoba	Pinto	244	USRM-20	Small Red	272	ND040494-4	Pinto	299	Focus	Pinto
219	Win Mor	Pinto	246	USWA-13	Great Northern	273	ND041062-1	Pinto	300	Vision	Pinto
220	AC Scarlet	Small Red	247	6R-42	Pink	274	ND060197	Pinto			
221	ICB-12	Pinto	248	Victor	Pink	275	Hatton	Pinto			
222	JM-24	Great Northern	249	USWA-61	Pink	276	GTS-900	Pinto			
223	USWA-12	Great Northern	250	I9365-25	Pink	277	CDCWM-2	Pinto			
224	Quincy	Pinto	251	I9365-5	Pink	278	CDC Pinnacle	Pinto			
225	Burke	Pinto	252	Rojo Chiquito	Small Red	279	CDC Nordic	Great Northern			
226	TARS-VCI-4B	Pinto	254	Viva	Pink	281	CDC Pintium	Pinto			
227	JM-126	Pinto	255	Roza	Pink	282	CDC Rosalee	Pink			
228	Pindak	Pinto	256	Harold	Pink	283	CDC Camino	Pinto			
229	Nodak	Pinto	257	Gloria	Pink	284	Gemini	Great Northern			
230	Holberg	Pinto	258	URS-117	Pink	285	Orion	Great Northern			
231	92US-1006	Pinto	259	PK9-7	Pink	286	UCD 9623	Pink			
232	Othello	Pinto	260	PK915	Pink	287	Yolano	Pink			
233	NW-590	Pinto	261	SR7-3	Small Red	288	Max	Pinto			
234	NW-410	Pinto	262	SR9-4	Small Red	289	Jackpot	Pinto			
235	PT7-2	Pinto	264	GN9-1	Great Northern	290	Gala	Pinto			
236	USPT-WM-1	Pinto	265	PT9-17	Pinto	291	Sequoia	Pinto			
237	USPT-CBB-1	Pinto	266	Maverick	Pinto	292	Apache	Pinto			

### 5.3.2 Field conditions

Field trials were conducted in a total of three location-years during 2014 to 2016. In 2014, a low-nitrogen field at the University of Guelph's Elora Research Station (ERS) near Elora, Ontario (43°38'27.8"N 80°24'20.4"W, 379 m elev.) was used. In 2015, a low-nitrogen farmer's field near Belwood, Ontario (43°40'16"N, 80°11'34"W, 430 m elev.) was used. In the growing seasons prior to 2014 the field at the ERS had been planted with high-N demand cereal crops to remove as much available nitrogen from the soil as possible. At the Belwood location, the field had been used to produce mixed hay with minimal inputs in the growing seasons previous to our trial. For 2014 and 2015 soil tests were performed by SGS Agrifood Laboratories (Guelph, ON) on composite samples constructed by mixing at least 12 soil core samples taken across the field, from the top (0-15 cm) and mid (15-30 cm) layers. Nitrate nitrogen ( $\text{NO}_3^-$ ) and ammonium nitrogen ( $\text{NH}_4^+$ ) were reported in parts per million (ppm). Wet weather conditions in 2015 prevented planting a second location at the ERS and the trial was sent to Puerto Rico later that year and grown over the winter of 2016 at the Isabela Agricultural Research Station in collaboration with Dr. T. Porch (USDA-ARS, Tropical Agriculture Research Station, Mayaguez, PR). The soil test for 2016 was performed by Waypoint Analytical (Richmond, VA) on a composite field sample. The soil at the Ontario locations is a Grey-Brown Podzol while the soil at Isabela is an Oxisol series Coto. Site soil test results are presented in Appendix C, Table S5.1.

Rainfall and temperature data for the Elora Research Station (Elora) were collected by the School of Environmental Sciences, University of Guelph; rainfall and

temperature data for the Belwood field site are as reported from the nearest Canada Weather Service location at the Fergus Shand Dam. The Isabela site was drip irrigated. Growing degree days (GDD calculated with base '0') were calculated based on the total growing days for each location-year. Site temperature and rainfall data are presented in Appendix C, Table S5.1.

### **5.3.3 Experimental design**

In 2014 at the ERS, 60 seeds of each of 281 genotypes (280 MDP and R99) were grown in four-row plots (150 cm × 90 cm, 37.5 cm between-row spacing and approximately 5 cm within row spacing) arranged in an  $\alpha$ -lattice design (2 replications). Border plots of R99 were also sown along the outer edge of the trial. In 2015 at Belwood, 135 seeds were sown in four-row plots (150 cm × 190 cm, 37.5 cm between-row spacing and approximately 5 cm within row spacing) and the genotypes were assigned to sub-lattices according to their days to maturity scores from 2014 to enable machine-harvesting as blocks of plots reached maturity. Plots of R99 were sown between each sub-lattice and along the outer edge of the trial. In the 2016 trial, 45 seeds were sown in single-row plots (300 cm) with 76 cm spacing between rows, using an  $\alpha$ -lattice design (2 replications).

### **5.3.4 Rhizobia Inoculation**

Rhizobial inoculation for Elora 2014 and Belwood 2015 followed the same methodologies presented in Chapter 3. Briefly, Nodulator® (Becker-Underwood) *Rhizobium leguminosarum* bv *phaseoli* inoculant was applied to the seed prior to

planting. To account for the differing plot sizes and seed numbers used at each location, the amount of inoculant powder used per seed envelope at Elora in 2014 was approximately 0.2 g and at Belwood in 2015 approximately 0.4 g was used. At Isabela (2016), the seeds were inoculated with a peat-based formulation of CIAT 1899 Rhizobia provided by Dr. C. Estevez de Jensen (University of Puerto Rico, Mayagüez). The inoculant contained about 25 g of Rhizobium inoculant (at a concentration of  $1 \times 10^9$  rhizobia  $\text{g}^{-1}$  peat) and PREMAX (bacterial protector) per 1 kg of seeds. Seed was mixed with the inoculant in a weigh boat container and placed in an envelope for planting. Inoculation was performed the day before planting and the seed was stored in an air-conditioned facility until planting occurred.

### **5.3.5 Field trial maintenance**

The plots were maintained with standard practices throughout each growing season. The Elora and Belwood field trials were maintained following the same methods described in Chapter 3. For the Isabela 2016 site, triple phosphate fertilizer was applied at planting and pre-emergent herbicide [Dual (S-metolachlor; Syngenta)] was applied to control broadleaf and grass weeds. Throughout the growing season, insecticides [Lannate (Dupont) and Admire (Bayer)] were applied twice. Foliar fertilizer [Nutrileaf 20:20:20 with micronutrients (Miller)] was applied with each pesticide application. The Isabela trial was drip irrigated approximately twice weekly to avoid water stress when rainfall was inadequate. Plots at Elora and Belwood were manually weeded once before canopy closure, and at Isabela as needed throughout the growing season.

### **5.3.6 Phenotyping at Elora and Belwood, Ontario**

Relative leaf chlorophyll content was measured at Elora and Belwood as described in Chapter 3. In the first growing season (Elora 2014), the plots were harvested by hand over a number of weeks, as plots reached maturity. The plants were threshed at the edge of the field using a Wintersteiger AG, (Upper Austria, Austria) with a Classic Seed-Gauge weighing system by Harvest-Master (Juniper Systems Inc., Utah, USA) and plot seed weight and moisture content were recorded. Seeds were placed in a low-temperature dryer for 24-48 hours at the ERS.

In the second growing season (Belwood 2015), the plots were machine-harvested using the Wintersteiger combine on three separate dates as the early, mid, and late-season varieties reached maturity. The seed was transported to the ERS and dried in a low-temperature dryer as in 2014.

### **5.3.7 Phenotyping at Isabela, Puerto Rico**

Different traits were evaluated at the Puerto Rico site, including leaf chlorophyll content and nodule traits. Relative leaf chlorophyll content was measured 49-50 days after planting (when all plots had reached the reproductive stage) using a SPAD 502 Plus Chlorophyll Meter (Konica-Minolta). The middle leaflet in the top-most, fully expanded trifoliate leaf was used for the measurements and three randomly-chosen plants were sampled per plot.

The number of nodules and their sizes were evaluated 48-49 days after planting (when all plots had reached the reproductive stage). Three plants per plot were carefully

dug from the ground and their root systems rinsed off in a bucket of water. Images of the root systems of each plant were taken using a DSLR camera (Nikon D70). Root nodules were counted and their sizes were rated using a relative scale (1 – small, 2 – medium, 3 – large).

### **5.3.8 Seed isotope analysis**

Seeds were prepared for isotope analysis using GCMS for the following measurements %N,  $\delta^{15}\text{N}$  (‰), and  $\delta^{13}\text{C}$  (‰). A description of the seed preparation and GCMS methodologies are found in Chapter 3. The natural abundance method was used to quantify the nitrogen-fixing capacity of each genotype as previously described in Chapter 3.

### **5.3.9 Statistical analysis**

Analysis of variance (ANOVA) tests were performed on the data collected from each environment and the environments combined using the MIXED procedure in SAS (version 9.4, SAS Institute, Cary, NC, USA, 2012). In each ANOVA, genotypes were considered fixed effects while all other effects and the interaction effects were considered random. The Shapiro-Wilk test (Shapiro and Wilk, 1965) was performed on the residuals in the UNIVARIATE procedure to test their normality. Random and independent distributions of the residuals were visually examined by plotting the studentized residuals against the predicted values. Data that generated outlier residuals was removed from the data set. Further, single degree of freedom contrasts were conducted in ANOVA between Middle American and Durango-Jalisco subpopulations.

Repeated measures of leaf chlorophyll content (SPAD), nodule count and nodule size were taken, and separate ANOVA tests were used to compare these values at each time point. In each ANOVA, the genotype least squared means (LSmeans) were computed using the LSMEANS statement in the MIXED procedure.

The LSmeans were used in correlation analyses where the pair-wise Pearson's coefficients of correlation ( $r$ ) were computed for all traits measured using the CORR procedure in SAS. The LSmeans were further used in the PRINCOMP and PRINQUAL SAS procedures to generate the principal component (PC) values and to estimate the proportion of variance accounted for by each PC. Excel was used to plot PC1 against PC2 to generating genotype  $\times$  trait (GT) biplots (Yan and Rajcan, 2002) to show genotype and trait interactions overall and in each environment.

### **5.3.10 Genotyping and SNP dataset development**

North Dakota State University (Dr. P. McClean, Dr. A. Oladzad) provided the SNP dataset used in this study. Development of the SNP dataset was previously described (Oladzad et al., 2019). The imputed dataset contained 278,760 SNPs and was further filtered for use in our genome-wide association study (GWAS) according to methods previously described (Torkamaneh and Belzile, 2015; Torkamaneh et al., 2018). Briefly, the SNP dataset was filtered using TASSEL v5 (Bradbury et al., 2007) and VCFtools (Danecek et al., 2011) to remove variants with  $\geq 50\%$  heterozygosity and a minor allele frequency (MAF)  $\leq 0.02$ . A further linkage disequilibrium (LD) filter using the *--indep-pairwise* command in PLINK (v1.90b6.5 64-bit) (Chang et al., 2015) with

parameters “20000 10 0.99” was conducted to remove redundant SNPs. After these filtering steps were carried out a set of 54,185 SNPs for the entire MDP, and sets of 48,414 and 43,520 SNPs for the DJ and MA subpopulations remained, respectively.

### **5.3.11 Population structure**

We used fastSTRUCTURE (Raj et al., 2014) to characterize population structure and the number of tested subpopulations (K) ranged from 1 to 15 with 5 independent runs of each. A python script (ChooseK.py) was used to determine the most likely K value based on the rate of change in LnP between successive K values.

Analysis of the evolutionary relationship of genotypes in the panel was performed in MEGA X (Kumar et al., 2018). A phylogenetic tree was constructed using the Neighbor-Joining method (Saitou and Nei, 1987) and distances were calculated using the Tamura-Nei method (Tamura and Nei, 1993) with 1000 bootstraps (Felsenstein, 1985).

### **5.3.12 Genome-wide association study**

GWAS analyses were performed using the rMVP package in R (Yin et al., 2020) using the Fixed and random model Circulating Probability Unification (FarmCPU) model (Liu et al., 2016) and Mixed Linear Model (MLM) (Zhang et al., 2010). Two different approaches (PCA (covariate P) and fastStructure (covariate Q)) were used to capture population structure. Two kinship matrices provided estimates of the relatedness among individuals (covariates K and K\*) (Kang et al., 2008; Li et al., 2014). Based on the cumulative distribution of p-values for different traits, models that took into account

kinship and PCA (P+K\*) were found to provide the best fit. The FDR-adjusted p-value (q-value) below 0.05 was used to establish a significance threshold (Wang et al., 2012).

### **5.3.13 Candidate gene identification**

The common bean genome version 2.0 (Schmutz et al., 2014) was searched using JBrowse (Buels et al., 2016) on Phytosome v12 (Goodstein et al., 2012) to explore the genome near significant SNPs and identify candidate genes. Any gene containing a significant SNP or within  $\pm 50$  kb of the SNP was considered a candidate gene. JBrowse on the Legume Information System website (<https://legumeinfo.org/>) was used to obtain annotation information of the candidate bean genes.

## **5.4 Results**

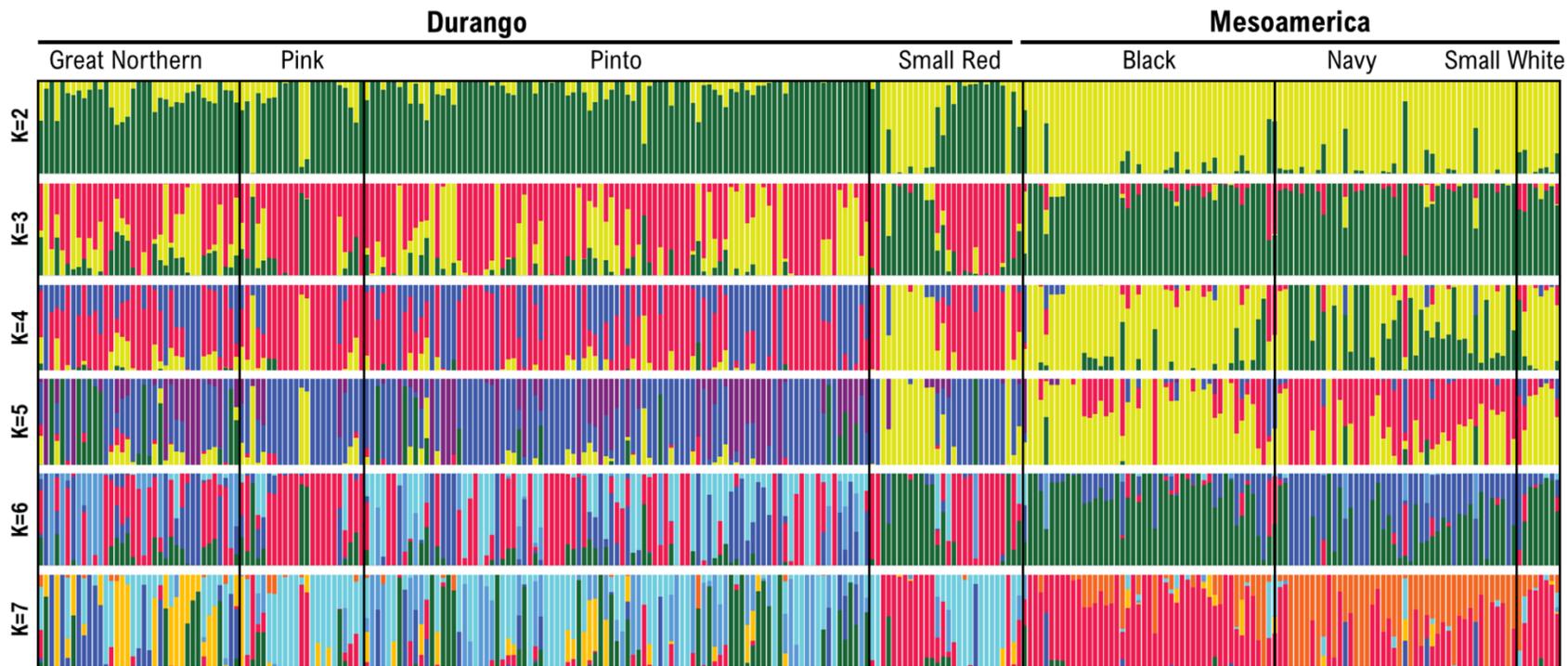
### **5.4.1 Field conditions**

A summary of pre-plant soil test results, precipitation and total growing degree days for all location-years is provided in Appendix C, Table S5.1. The soil test results show low ( $\leq 8.6$  ppm) levels of nitrates ( $\text{NO}_3^-$ ) and ammonium ( $\text{NH}_4^+$ ) in the bean root zone. Planting in 2015 occurred two weeks later than in 2014 as a result of wet spring weather, and the fall harvest was impacted by heavy precipitation. Despite the late start to the 2015 season, the number of accumulated growing degree days (GDD) over the 2015 growing season (1862 GDD) was similar to that observed in 2014 (1913 GDD). At Isabela, the trial growing period was 30-40 days shorter than the Elora and Belwood trials and moisture stress was mitigated by drip irrigation.

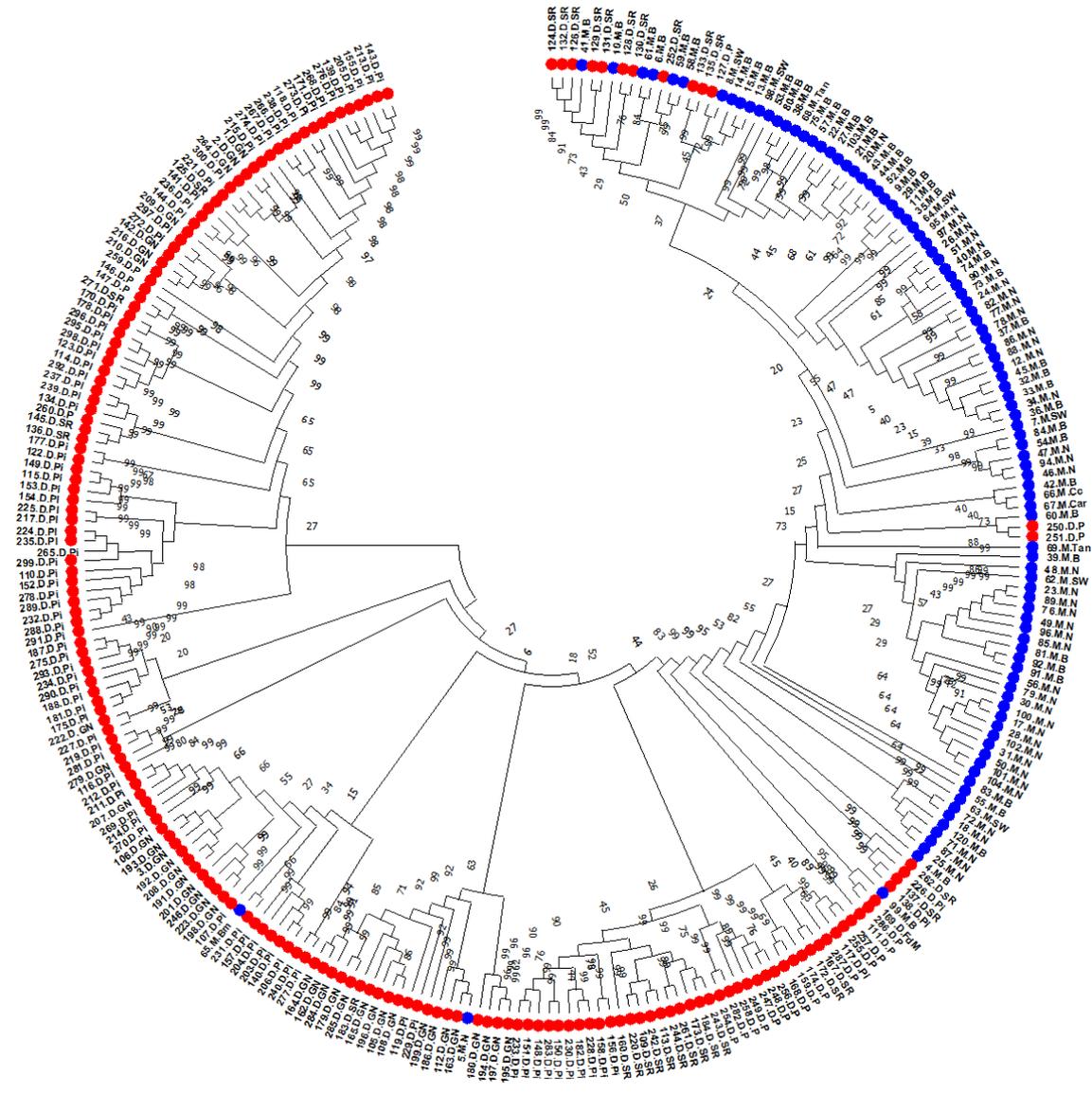
### 5.4.2 Population structure

The MDP is known to consist of two subpopulations ( $K = 2$ ), race Mesoamerica and race Durango-Jalisco, that are further subdivided into market classes when STRUCTURE analysis is carried out at the  $K = 3$  to  $K = 7$  levels (Moghaddam et al., 2016). Our STRUCTURE analysis, using the GBS-derived SNP dataset for the first time, confirms the previously published (Moghaddam et al., 2016) two subpopulation structure of the panel. However strong genetic differentiation among market classes at higher  $K$  levels is not apparent for all market classes in our analysis (Figure 5.1). For each  $K$  analysis, the small red market class is comprised of genotypes from each subpopulation and contains admixed genotypes. Distinction between the black and white Mesoamerican market classes is evident at  $K = 3$ , whereas distinction among the pinto, pink and great northern Durango-Jalisco market classes is not apparent until  $K = 7$ . The pink and pinto market classes remained similar through most of the levels of the STRUCTURE analysis.

The evolutionary relationships among genotypes in the MDP are presented in Figure 5.2 as a neighbor-joining phylogenetic tree. Two subpopulations are apparent in the tree representing race Mesoamerica and race Durango-Jalisco, with a small number of genotypes appearing among the branches of the other subpopulation.



**Figure 5.1.** STRUCTURE analysis of the Middle American diversity panel (MDP) from  $K = 2$  to  $K = 7$ . The  $K = 2$  plot divides the panel into two subpopulations, Durango (Durango-Jalisco) and Mesoamerica. Market classes are denoted for each subpopulation.



**Figure 5.2.** Phylogenetic tree of 280 genotypes of the Middle American Diversity Panel inferred using the neighbor-joining method in MEGA X. Two race subpopulations are depicted, Mesoamerican (blue) and Durango-Jalisco (red). Each genotype is labeled with the MDP number code, race subpopulation assignment (M – Mesoamerican, D – Durango-Jalisco), and market class (B – black, N – navy, SW – small white, P – pink, Pi – pinto, GN – great northern, SR – small red, FdM – flor de mayo, Bm – black mottle, Car - Carioca, CC – cream carioca, Tan – tan).

### 5.4.3 Trait diversity

In the combined mixed model ANOVA significant differences were found for SNF-related traits among both the fixed and random effects categories. For the fixed effects of genotype and subpopulation categories (race MA vs. DJ), significant differences were seen among the genotypes for nitrogen derived from the atmosphere (%N<sub>d</sub>;  $p = 0.0008$ ), nitrogen discrimination ( $\delta^{15}N$ ;  $p = 0.0015$ ), and seed nitrogen content (%N;  $p < 0.0001$ ) (Appendix C, Table S5.2). Between the two subpopulations, significant differences were found for %N ( $p < 0.0001$ ) where MA genotypes ( $M = 4.0\%$ ) outperformed DJ ( $M = 3.9\%$ ) genotypes.

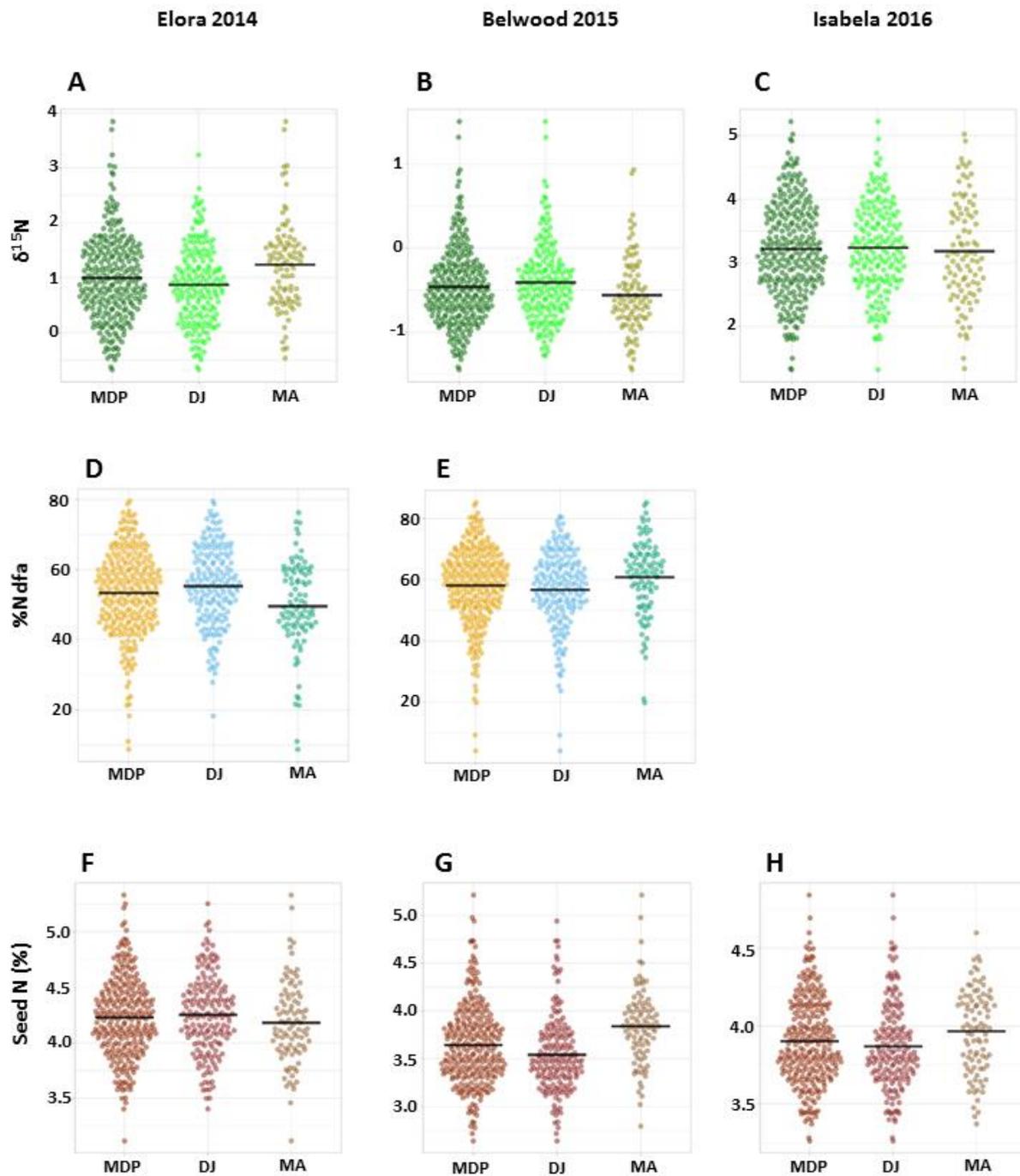
When the random effects in the combined ANOVA were considered, the effect of environment was not significant for any trait, however the genotype by environment interaction (ENV\*genotype) was significant for all traits (Appendix C, Table S5.2), indicating that genotype performance was impacted by the growing environment. The block within environment interaction was not significant at any location, however the incomplete block within the environment by block interaction was significant for all seed composition traits (Appendix C, Table S5.2), indicating some variation in performance across the field sites. Because genotype performance was significantly affected by growing environment, all traits were further analyzed by environment.

When SNF-related traits were analyzed for each location, significant genotype effects were apparent (Appendix C, Tables S5.3-5) and are explored by location in the following sections.

#### 5.4.3.1 Nitrogen discrimination ( $\delta^{15}\text{N}$ )

Nitrogen discrimination ( $\delta^{15}\text{N}$ ) saw significant variation in the panel at each trial location (Appendix C, Tables S5.3-5) and ranged from -0.65 ‰ to 3.85 ‰ at Elora (Figure 5.3A), from -1.45 ‰ to 1.51 ‰ at Belwood (Figure 5.3B), and from 1.35 ‰ to 5.23 ‰ at Isabela (Figure 5.3C). For the non-nodulating, non N-fixing reference genotype R99, average  $\delta^{15}\text{N}$  values were 4.40 ‰ ( $n = 47$ ) at Elora, 1.66 ‰ ( $n = 99$ ) at Belwood, and 3.56 ‰ ( $n = 6$ ) at Isabela. Whereas at Elora and Belwood the MDP genotypes had lower  $\delta^{15}\text{N}$  values than R99, at Isabela a number of MDP genotypes had higher  $\delta^{15}\text{N}$  values than the reference, indicating that those genotypes fixed less N than R99. This prevented accurate calculation of %Ndfa values for the Isabela trial and consequently %Ndfa is not presented for Isabela.

Comparisons between race subpopulations for nitrogen discrimination revealed further significant differences. When nitrogen discrimination was compared by subpopulation at Elora (Figure 5.3A), DJ genotypes ( $M = 0.869$ ) had significantly lower ( $p < 0.0001$ )  $\delta^{15}\text{N}$  values than MA genotypes ( $M = 1.237$ ) (Appendix C, Table S5.3). At Belwood (Figure 5.3B) the opposite was found, where DJ genotypes ( $M = -0.41$ ) had significantly higher  $\delta^{15}\text{N}$  values ( $p = 0.0022$ ) than MA genotypes ( $M = -0.56$ ) (Appendix C, Table S5.4). Similarly, when nitrogen discrimination was compared between race subpopulations at Isabela (Figure 5.3C), DJ genotypes ( $M = 3.23$ ) had higher levels than MA genotypes ( $M = 3.19$ ), but these values were not significantly different ( $p = 0.7109$ ) (Appendix C, Table S5.5).



**Figure 5.3.** Data distributions of LSmeans for seed composition SNF-related traits in the MDP, and each subpopulation (DJ and MA) at field trial locations in Elora (2014), Belwood (2015), and Isabela (2016). Distribution means are indicated with a black line. A-C. Nitrogen discrimination (%), (A) Elora 2014, (B) Belwood 2015, (C) Isabela 2016. D-E. Percent nitrogen derived from the atmosphere (%Ndfa), (D) Elora 2014, (E) Belwood 2015. F-H. Seed nitrogen (%), (F) Elora 2014, (G) Belwood 2015, (H) Isabela 2016.

#### 5.4.3.2 Percent nitrogen derived from the atmosphere (%Ndfa)

The capacity for nitrogen fixation, quantified by the calculation of the percentage of nitrogen derived from the atmosphere (%Ndfa), was significantly different between locations (Appendix C, Tables S5.3-5). Nitrogen fixation ranged from 8.7% to 79.1% at Elora (Figure 5.3D), and from 4.2% to 85.2% at Belwood (Figure 5.3E).

When %Ndfa was compared by race subpopulations, nitrogen fixing capacity was found to differ significantly at Elora ( $p < 0.0001$ ; Appendix C, Table S5.3) where DJ genotypes ( $M = 55.3\%$ ) fixed higher levels of nitrogen than MA genotypes ( $M = 49.5\%$ ) (Figure 5.3D). At Belwood, nitrogen fixing capacity differed significantly between race subpopulations ( $p = 0.0022$ ; Appendix C, Table S5.4), and MA genotypes ( $M = 60.8\%$ ) fixed significantly more nitrogen than DJ genotypes ( $M = 56.7\%$ ) (Figure 5.3E). This is the reverse of what was observed at Elora.

The top 20 genotypes at Elora were primarily Durango-Jalisco and fixed between 70.5% and 79.1% nitrogen, while at Belwood the top 20 genotypes were primarily Mesoamerican and fixed between 74.7% and 85.2% nitrogen (Table 5.3). The bottom 20 genotypes at Elora were equally divided between Durango-Jalisco and Mesoamerican races and fixed from 35.1% to as little as 8% nitrogen, while at Belwood Durango-Jalisco genotypes dominated the bottom 20 genotypes and fixed from 37.7% to as little as 4% nitrogen (Table 5.4). The environmental effect on nitrogen fixation is apparent in this study when variety performance is compared across locations: (1) no

**Table 5.3.** The top 20 MDP genotypes ranked by nitrogen fixation capacity (%Ndfa) at the two Ontario field trial locations (Elora, 2014 and Belwood, 2015).

Rank	Elora					Belwood				
	MDP No.	Variety	Race	Market class	%Ndfa	MDP No.	Variety	Race	Market class	%Ndfa
1	203	ABCP-8	DJ	Pinto	79.1	69	SEA 10*	MA	Tan	85.2
2	172	UI-3	DJ	Small Red	78.0	21	Domino	MA	Black	84.3
3	120	Shiny Crow	MA	Black	76.8	60	ICB-10	MA	Black	81.8
4	181	UI-126	DJ	Pinto	76.5	132	TARS09-RR004	DJ	Small Red	80.5
5	230	Holberg	DJ	Pinto	75.9	219	Win Mor	DJ	Pinto	80.4
6	139	Sierra	DJ	Pinto	75.4	6	Xan 176	MA	Black	80.3
7	158	Sonora	DJ	Pinto	74.7	58	I9365-31	MA	Black	79.5
8	182	UI-196	DJ	Pinto	74.5	275	Hatton	DJ	Pinto	79.2
9	117	Fisher	DJ	Pinto	74.2	41	115M (Black Rhino)	MA	Black	79.1
10	114	Bill Z	DJ	Pinto	73.9	276	GTS-900	DJ	Pinto	78.3
11	258	URS-117	DJ	Pink	73.4	255	Roza	DJ	Pink	77.7
12	25	Laker	MA	Navy	73.1	9	DPC-4	MA	Black	77.1
13	26	Mayflower	MA	Navy	73.0	67	A801	MA	Carioca	77.0
14	151	Flint*	DJ	Pinto	71.4	123	Arapaho	DJ	Pinto	76.2
15	124	DOR 364	DJ	Small Red	71.3	24	C-20	MA	Navy	76.2
16	122	Croissant	DJ	Pinto	71.2	90	Lightning*	MA	Navy	76.0
17	175	Common Pinto	DJ	Pinto	71.1	131	Dehoro	DJ	Small Red	75.5
18	246	USWA-13	DJ	Great Northern	70.9	18	Seafarer	MA	Navy	75.2
19	68	BAT 477	MA	Tan	70.7	52	Midnight	MA	Black	75.0
20	289	Jackpot	DJ	Pinto	70.5	215	NE2-09-10	DJ	Pinto	74.8

\* Variety %Ndfa performance was among worst 20 in previous/successive location

varieties ranked in the top 20 at both locations (Table 5.3), and (2) a few genotypes ranked among the top 20 at Elora but among the worst 20 at Belwood and vice versa (Tables 5.3 and 5.4). When the panel is examined by market class, more consistent patterns of nitrogen fixing capacity are apparent. Among DJ market classes grown in Ontario, the pinto genotypes showed the best nitrogen fixation capacity at Elora with the top 10 genotypes fixing between 71.2% and 79.1% nitrogen, followed by the top 10 small red genotypes, which fixed between 55.2% and 78% nitrogen, and the top 10 great northern genotypes, which fixed between 61.9% and 70.9% nitrogen (Table 5.5). Ontario DJ market classes performed similarly at Belwood where the top 10 pinto genotypes fixed between 70.8% and 80.4% nitrogen, followed by small red where the top ten genotypes fixed between 62.7% and 80.5% nitrogen and the great northern where the top 10 fixed between 63.6% and 74% nitrogen (Table 5.5). Nitrogen fixation performance was most consistent within the small red market class with five genotypes ranking in the top 10 at each location, whereas the great northern and pinto Durango-Jalisco market classes had only one genotype ranked in the top 10 at each location. Among the MA market classes grown in Ontario, the navy/small white market class and the black market class showed similar nitrogen fixing capacity at Elora with the top 10 navy/small white genotypes fixing between 60.8% to 73.1% nitrogen and the top 10 black genotypes fixing between 59.1% and 76.8% nitrogen (Table 5.6).

**Table 5.4.** The worst 20 MDP genotypes ranked by nitrogen fixation capacity (%Ndfa) at the two Ontario field trial locations (Elora, 2014 and Belwood, 2015). Italicized MDP No. and variety name indicates genotype appearance in list at each location.

Rank	Elora					Belwood				
	MDP No.	Variety	Race	Market class	%Ndfa	MDP No.	Variety	Race	Market class	%Ndfa
1	74	ND021717	MA	Black	35.1	27	Blackhawk	MA	Black	37.7
2	127	PR 0401-259	DJ	Pink	34.9	284	Gemini	DJ	Great Northern	37.3
3	160	Red Ryder	DJ	Small Red	34.2	273	ND041062-1	DJ	Pinto	37.3
4	69	SEA 10*	MA	Tan	34.2	5	AC Polaris	MA	Navy	36.5
5	<i>130</i>	<i>INTA Precoz</i>	DJ	Small Red	34.0	271	NDZ06249	DJ	Small Red	36.1
6	112	AC Resolute	DJ	Great Northern	33.0	150	Buckskin	DJ	Pinto	36.1
7	96	T9903	MA	Navy	32.9	240	USPT-CBB-5	DJ	Pinto	35.6
8	173	UI-37	DJ	Small Red	32.5	100	Albion	MA	Navy	34.5
9	236	USPT-WM-1	DJ	Pinto	32.2	224	Quincy	DJ	Pinto	34.2
10	252	Rojo Chiquito	DJ	Small Red	32.0	292	Apache	DJ	Pinto	32.2
11	169	Desert Rose	DJ	Flor de Mayo	30.1	186	UI-59	DJ	Great Northern	31.9
12	259	PK9-7	DJ	Pink	27.2	295	Buster	DJ	Pinto	30.5
13	23	Bunsi	MA	Navy	26.8	143	Santa Fe	DJ	Pinto	29.2
14	33	Jaguar	MA	Black	24.1	152	Fargo	DJ	Pinto	28.7
15	90	Lightning*	MA	Navy	23.9	233	NW-590	DJ	Pinto	25.3
16	65	Orca	MA	Black mottle	23.5	151	Flint*	DJ	Pinto	23.7
17	5	AC Polaris	MA	Navy	21.1	73	Eclipse	MA	black	21.1
18	113	AC Earlired	DJ	Small Red	18.5	17	Sanilac	MA	Navy	19.9
19	30	Newport	MA	Navy	11.4	<i>130</i>	<i>INTA Precoz</i>	DJ	Small Red	9.3
20	44	Bandit	MA	Black	8.7	287	Yolano	DJ	Pink	4.2

\* Variety %Ndfa performance was among top 20 at previous/successive location

**Table 5.5.** The best 10 genotypes for nitrogen fixation capacity according to market class within the Durango-Jalisco subpopulation of the MDP. Genotypes are ranked by nitrogen fixation capacity (%Ndfa) at the two Ontario field trial locations (Elora, 2014 and Belwood, 2015). Italicized MDP No. and variety name indicates genotype appearance in list at each location.

Elora					Belwood			
Rank	<u>MDP No.</u>	<u>Variety</u>	<u>Market class</u>	<u>%Ndfa</u>	<u>MDP No.</u>	<u>Variety</u>	<u>Market class</u>	<u>%Ndfa</u>
1	172	<i>UI-3</i>	Small Red	78.0	132	TARS09-RR004	Small Red	80.5
2	124	<i>DOR 364</i>	Small Red	71.3	131	<i>Dehoro</i>	Small Red	75.5
3	184	UI-239	Small Red	66.6	124	<i>DOR 364</i>	Small Red	72.7
4	133	<i>TARS09-RR007</i>	Small Red	61.6	136	F07-449-9-3	Small Red	69.8
5	220	AC Scarlet	Small Red	58.5	133	<i>TARS09-RR007</i>	Small Red	69.6
6	109	AC Redbond	Small Red	57.5	128	<i>IBC 301-204</i>	Small Red	69.4
7	128	<i>IBC 301-204</i>	Small Red	55.8	172	<i>UI-3</i>	Small Red	66.7
8	145	Merlot	Small Red	55.8	113	AC Earlired	Small Red	63.5
9	126	Amadeus 77	Small Red	55.6	160	Red Ryder	Small Red	63.4
10	131	<i>Dehoro</i>	Small Red	55.2	262	SR9-4	Small Red	62.7
1	203	ABCP-8	Pinto	79.1	219	Win Mor	Pinto	80.4
2	181	UI-126	Pinto	76.5	275	Hatton	Pinto	79.2
3	230	<i>Holberg</i>	Pinto	75.9	276	GTS-900	Pinto	78.3
4	139	Sierra	Pinto	75.4	123	Arapaho	Pinto	76.2
5	158	Sonora	Pinto	74.7	215	NE2-09-10	Pinto	74.8
6	182	UI-196	Pinto	74.5	230	<i>Holberg</i>	Pinto	72.4
7	117	Fisher	Pinto	74.2	265	PT9-17	Pinto	72.1
8	114	Bill Z	Pinto	73.9	227	JM-126	Pinto	71.3
9	151	Flint	Pinto	71.4	298	Mariah	Pinto	71.3
10	122	Croissant	Pinto	71.2	239	USPT-ANT-1	Pinto	70.8
1	246	USWA-13	Great Northern	70.9	223	USWA-12	Great Northern	74.0
2	3	BelMiNeb-RMR-8	Great Northern	69.0	105	BelMiNeb-RR-2	Great Northern	73.8
3	142	Matterhorn	Great Northern	68.8	180	UI-123	Great Northern	73.6
4	165	Sapphire	Great Northern	68.1	195	<i>GN Harris</i>	Great Northern	72.4
5	195	<i>GN Harris</i>	Great Northern	68.1	2	BelMiNeb-RMR-4	Great Northern	70.3
6	162	Beryl	Great Northern	67.3	209	NE1-09-20	Great Northern	69.8
7	196	GN Star	Great Northern	67.2	197	Tara	Great Northern	69.5
8	108	BelNeb-RR-1	Great Northern	64.4	264	GN9-1	Great Northern	65.3
9	279	CDC Nordic	Great Northern	63.9	198	Starlight	Great Northern	64.7
10	285	Orion	Great Northern	61.9	164	Marquis	Great Northern	63.6

At Belwood the black genotypes had better %Ndfa performance than the navy/small white genotypes, with the top 10 black genotypes fixing between 71.4% and 84.3% nitrogen and the navy/small white top 10 genotypes fixing between 68.3% and 76.2% nitrogen (Table 5.6). Navy bean nitrogen fixation performance was slightly more consistent than black bean performance among the Mesoamerican market classes, with 4 navy varieties versus 3 black varieties appearing in the top 10 both years.

**Table 5.6.** The best 10 genotypes for nitrogen fixation capacity according to market class within the Mesoamerican subpopulation of the MDP. Genotypes are ranked by nitrogen fixation capacity (%Ndfa) at the two Ontario field trial locations (Elora, 2014 and Belwood, 2015). Italicized MDP No. and variety name indicates genotype appearance in list at each location.

Rank	Elora				Belwood			
	MDP No.	Variety	Market class	%Ndfa	MDP No.	Variety	Market class	%Ndfa
1	25	<i>Laker</i>	Navy	73.1	24	C-20	Navy	76.2
2	26	Mayflower	Navy	73.0	90	Lightning	Navy	76.0
3	63	Hyden	Small White	69.9	18	Seafarer	Navy	75.2
4	76	Crestwood	Navy	67.3	87	<i>Fleetwood</i>	Navy	71.5
5	48	Voyager	Navy	64.8	51	Navigator	Navy	70.6
6	46	Schooner	Navy	63.4	25	<i>Laker</i>	Navy	70.5
7	87	<i>Fleetwood</i>	Navy	63.3	7	<i>Morales</i>	Small White	70.5
8	7	<i>Morales</i>	Small White	62.7	64	<i>USWA-50</i>	Small White	68.9
9	20	Neptune	Navy	60.9	98	UCD 96114	Small White	68.7
10	64	<i>USWA-50</i>	Small White	60.8	12	F07-004-9-1	Navy	68.3
1	120	<i>Shiny Crow</i>	Black	76.8	21	Domino	Black	84.3
2	43	Shania	Black	61.8	60	<i>ICB-10</i>	Black	81.8
3	15	F04-2801-4-1-2	Black	61.6	6	Xan 176	Black	80.3
4	61	ICB-3	Black	61.3	58	I9365-31 115M (Black Rhino)	Black	79.5
5	9	<i>DPC-4</i>	Black	61.3	41	<i>DPC-4</i>	Black	79.1
6	80	Blackjack CDC	Black	61.0	9	<i>DPC-4</i>	Black	77.1
7	81	Nighthawk	Black	60.4	52	Midnight	Black	75.0
8	99	UCD 9634	Black	60.0	45	Loreto	Black	72.8
9	83	CDC Espresso	Black	59.6	120	<i>Shiny Crow</i>	Black	71.9
10	60	<i>ICB-10</i>	Black	59.1	59	92BG-7	Black	71.4

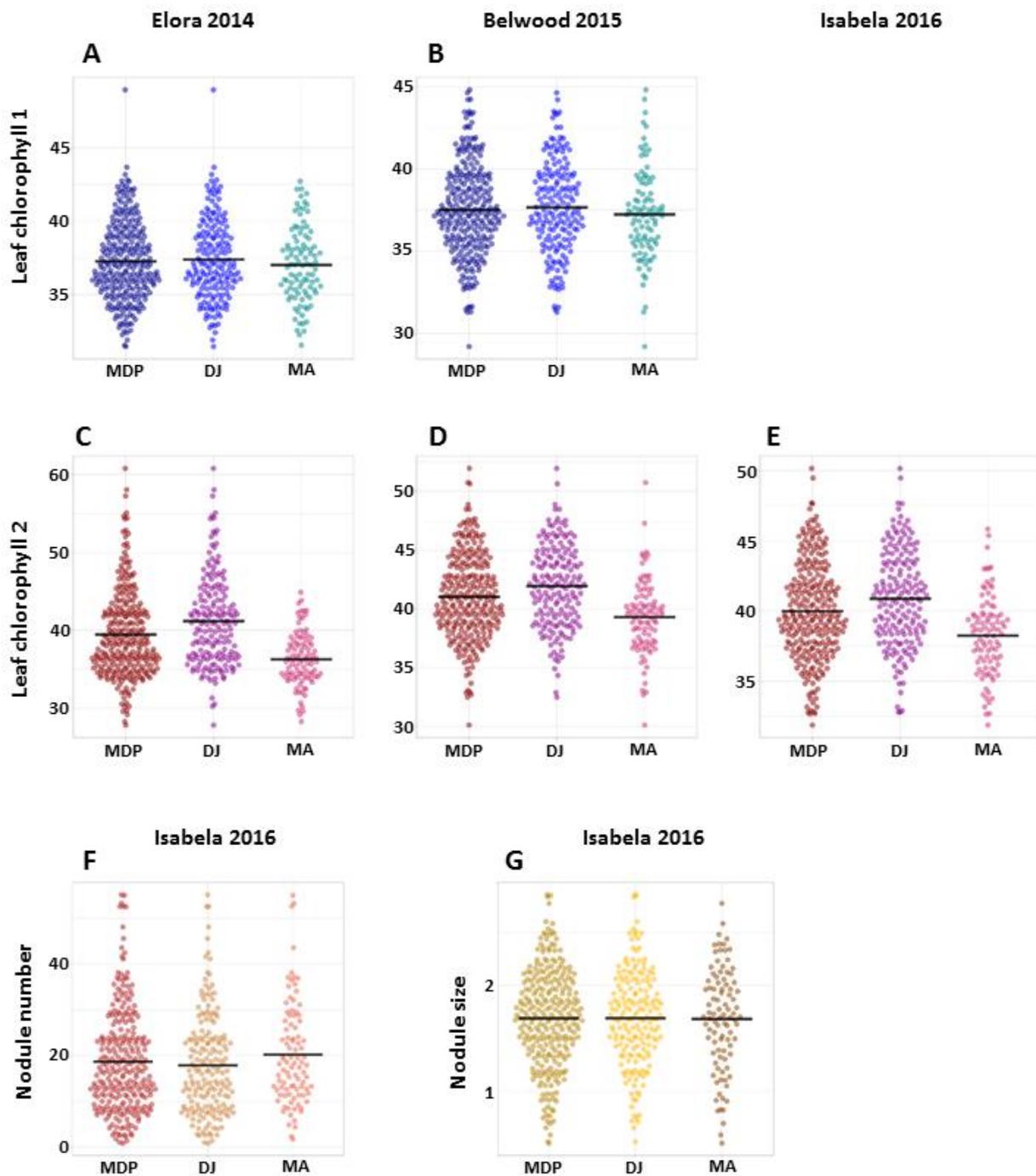
#### 5.4.3.3 Seed nitrogen content (%N)

Significant differences were found between genotypes for seed N content (%) and ranged from 3.1% to 5.3% at Elora ( $p < 0.0001$ , Appendix C, Table S5.3; Figure 5.3F), between 2.6% and 5.2% at Belwood ( $p < 0.0001$ , Appendix C, Table S5.4; Figure 5.3G), and from 3.3% to 4.8% at Isabela ( $p < 0.0001$ , Appendix C, Table S5.5; Figure 5.3H). The %N values found at Isabela were similar although narrower in range than the values found at Elora and Belwood.

When seed %N was compared by subpopulation at Elora (Figure 5.3F), DJ genotypes ( $M = 4.24\%$ ) contained slightly more seed N than MA genotypes ( $M = 4.19\%$ ), which was not a significant difference (Appendix C, Table S5.3). At Belwood (Figure 5.3G), DJ genotypes ( $M = 3.5\%$ ) had significantly lower ( $p < 0.0001$ ) seed N proportions than MA genotypes ( $M = 3.8\%$ ) (Appendix C, Table S5.4). The same trend was seen at Isabela (Figure 5.3H) where DJ genotypes ( $M = 3.9\%$ ) had significantly lower ( $p < 0.0010$ ) seed N proportions than MA genotypes ( $M = 4.0\%$ ) (Appendix C, Table S5.5).

#### 5.4.3.4 Leaf chlorophyll content

As a repeated measure, leaf chlorophyll content (SPAD) was analyzed in separate F-tests. For the genotype effect, SPAD values differed significantly at each location (Appendix C, Table S5.6). When leaf chlorophyll content was measured at the early vegetative stage (SPAD1), population and subpopulation averages were similar between Elora (2014) and Belwood (2015) (Figure 5.4 A & B). Similarly, when



**Figure 5.4.** Data distributions of LSmeans for physiological and morphological SNF-related traits in the MDP, and each subpopulation (DJ and MA) at field trial locations in Elora (2014), Belwood (2015), and Isabela (2016). Distribution means are indicated with a black line. A-C. Leaf chlorophyll content at vegetative stage, (A) Elora 2014, (B) Belwood 2015. C-E. Leaf chlorophyll content at reproductive stage, (C) Elora 2014, (D) Belwood 2015, (E) Isabela 2016. F. Nodule number, Isabela 2016. G. Nodule size, Isabela 2016.

leaf chlorophyll content was measured at the reproductive stage (SPAD2), the population and subpopulation averages were similar at all locations (Figure 5.4 C-E). Significant differences were seen when SPAD measurements taken at different growth stages were compared, with SPAD1 values significantly lower than SPAD2 values (Appendix C, Table S5.6). Furthermore, at Elora and Belwood the growth stage at which leaf chlorophyll content was measured had a significant effect on genotype performance (significant SPADT\*G interaction; Appendix C, Table S5.6). The observation within block by genotype by SPAD time interaction (SPADT(BLOC\*genotype)) was significant in all environments (Appendix C, Table S5.6).

When SPAD scores were compared for genotypes according to race subpopulation membership, leaf chlorophyll content was higher for the DJ subpopulation than the MA subpopulation at both measurement times and all locations (Figure 5.4). The difference between the DJ and MA subpopulation was not significant at the early vegetative stage (SPAD1) (Figure 4A-B), but at the reproductive stage (SPAD2) (Figure 4C-E), significant differences were seen at each location where DJ genotypes had higher SPAD readings than MA genotypes (Table 5.8). At Elora, DJ genotypes ( $n = 186$ ) had a significantly higher ( $p < 0.0001$ ) average SPAD2 value of 41.2 units compared to the MA genotypes ( $n = 112$ ), which had an average of 36.6 units (Figure 4C). At Belwood, DJ genotypes ( $n = 186$ ) had a significantly higher ( $p < 0.0001$ ) average SPAD2 value of 41.9 units compared to the MA genotypes ( $n = 109$ ), which had an average of 39.4 units (Figure 4D). At Isabela, the SPAD2 value difference between DJ ( $n = 186$ ) and MA ( $n = 108$ ) genotypes was more narrow but still significant

( $p < 0.0001$ ), with the DJ average at 40.9 units and the MA average at 38.2 units (Figure 4E).

#### **5.4.3.5 Nodule number and size**

As a repeated measure, nodule size ranking and nodule number scored at Isabela were analyzed in separate F-tests. For the fixed effect genotype significant differences were observed for nodule size ( $p < 0.0001$ ) and nodule count ( $p < 0.0001$ ) (Appendix C, Table S5.6). Variation across the trial site for nodule size was also observed (Appendix C, Table S5.6). When nodule traits were compared between the two subpopulations, no significant differences were found for nodule size, however nodule counts were significantly higher ( $p = 0.0048$ ) in MA genotypes ( $M = 20.5$  nodules) than DJ genotypes ( $M = 17.7$  nodules). Trait distributions for nodule count and nodule size are presented in Figures 5.4F and 5.4G, respectively.

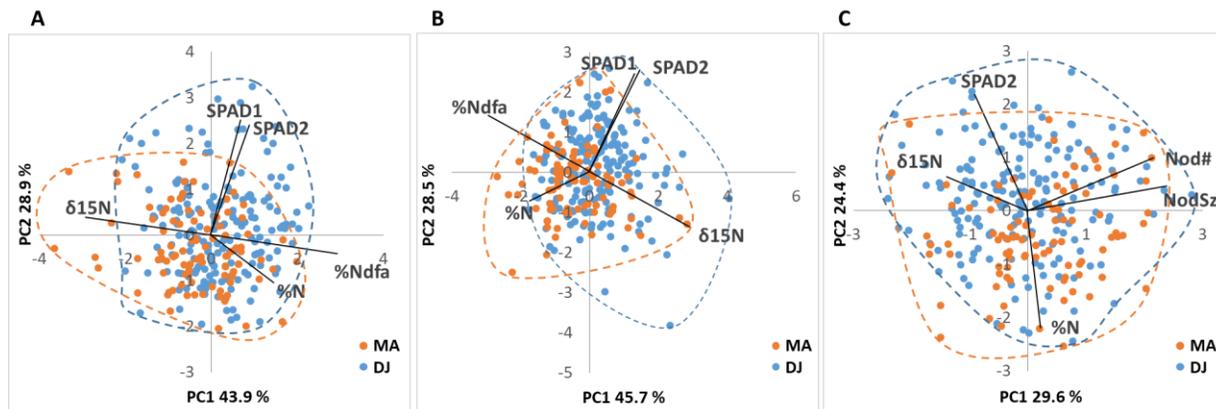
#### **5.4.4 Trait correlation**

The Pearson correlation (Figure 5.5) analyses revealed a number of trait and location relationships. In the MDP (Figure 5.5A), %Ndfa was positively correlated with seed %N at Elora and at Belwood, and conversely  $\delta^{15}\text{N}$  was negatively correlated with %N at Elora and Belwood and the trend between these traits was not significant at Isabela. %Ndfa and  $\delta^{15}\text{N}$  were not correlated to leaf chlorophyll content measured at the second trifoliolate stage (SPAD1). Chlorophyll measurements taken at flowering time



(SPAD2) were positively associated with %Ndfa at Elora, but negatively associated at Belwood. The opposite pattern was seen with  $\delta^{15}\text{N}$  and SPAD2, where these traits were negatively correlated at Elora, and positively correlated at Belwood and Isabela. Nitrogen fixation (%Ndfa) also had positive correlations with some agronomic traits such as days to maturity and yield, yet was negatively correlated with days to flowering and harvest index, depending on the trial location (data not shown). Leaf chlorophyll content (SPAD2) was negatively correlated to seed nitrogen (%N) at Belwood and Isabela. Leaf chlorophyll measurements taken at two developmental stages were positively correlated with each other at all locations. Nodule size was positively associated with nodule count at Isabela.

Overall similar correlation patterns were seen at each location when the panel was analyzed by Mesoamerican (Figure 5.5B) and Durango-Jalisco (Figure 5.5C) subpopulations, although there were a few different trends apparent. Correlations between %Ndfa and SPAD1 were revealed, where negative associations between these traits were found for individuals of the MA subpopulation at Belwood and positive associations were found among the traits for individuals in the DJ subpopulation at Elora. The negative correlations between %N and SPAD that were seen in the MDP overall were repeated in the DJ subpopulation but disappeared in the MA subpopulation. At Isabela, seed %N was negatively associated with nodule number among the MA subpopulation, and nodule size was negatively associated with  $\delta^{15}\text{N}$  among the DJ subpopulation.



**Figure 5.6.** Genotype by trait bi-plots of principle components 1 and 2 for Elora2014 (A), Belwood 2015 (B), and Isabela 2016 (C). %Ndfa, % N derived from the atmosphere;  $\delta^{15}\text{N}$ , N discrimination; %N, seed nitrogen (%); SPAD1, SPAD meter reading at second trifoliolate stage; SPAD2, SPAD meter reading at flowering; Nod#, nodule number; NodSz, nodule size.

The genotype  $\times$  trait biplot analyses (Figure 5.6) supported the relationships we observed in the correlation analyses. Principle component 1 explained approximately 45% of the variation observed at Elora and Belwood (Figure 5.6A and B), whereas it explained under 30% of the variation at Isabela (Figure 5.6C). The amount of variation explained by principle component 2 was more similar between locations with approximately 29% explained at Elora and Belwood, and 24% explained at Isabela. The distribution pattern of varieties across the Elora and Belwood biplots are more dense than the distribution observed at Isabela. Additionally, the MA and DJ subpopulations occupy overlapping as well as unique areas of the Elora biplot, whereas the two subpopulations occupy largely the same regions of the Belwood and Isabela biplots, except for a few outliers. At Belwood and Isabela the separation between the two subpopulations appears to be explained by variation in %N values. The near-90° angle formed between the leaf chlorophyll content vectors and the % Ndfa and  $\delta^{15}\text{N}$  vectors

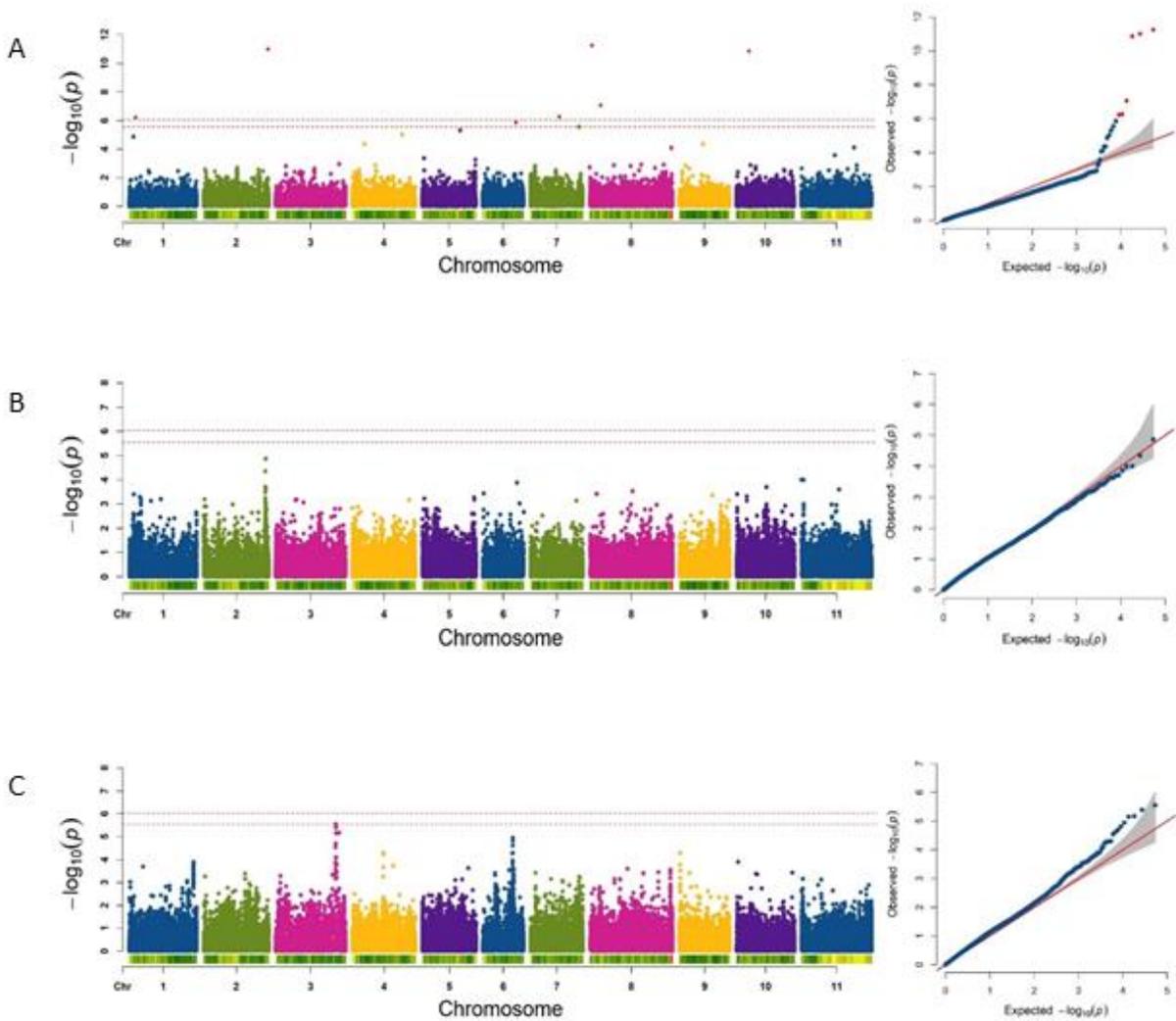
indicates that these traits are not related. At Isabela, the nodule traits are associated with each other, but have little to no relationship with %N and SPAD2, and have a negative relationship with  $\delta^{15}\text{N}$ . Although %Ndfa values are not presented for Isabela, following the pattern observed at Elora and Belwood, the vector would plot directly opposite  $\delta^{15}\text{N}$  and would create an acute angle with the nodule trait vectors indicating a positive association between %Ndfa and nodule number and nodule size.

#### **5.4.5 Marker-trait associations**

GWAS was used to determine the genomic regions associated with each phenotypic trait in the MDP overall and within each race subpopulation. The SNPs that plot above the 5% false discovery rate (FDR) cutoff were considered significant, and in some cases when a 10% FDR was used further significant SNPs were identified.

##### **5.4.5.1 Nitrogen discrimination**

For nitrogen discrimination ( $\delta^{15}\text{N}$ ) in the MDP at Elora, significant associations were found for SNP markers on Pv01, Pv02, Pv07, Pv08, and Pv10 (Figure 5.7, A). The most significant SNP occurred at 1.8 Mb (C/T) on Pv08 and it accounted for a decrease of 1.8 ‰ in  $\delta^{15}\text{N}$  (Table 5.7). SNPs on Pv07 at 22.1 Mb (G/A), on Pv08 at 1.8 Mb (C/T) and 8.4 Mb (G/A) and on Pv10 at 9.8 Mb (A/T) marked peaks for %Ndfa found in the entire panel. No significant associations were found for the entire panel at Belwood nor Isabela (Figure 5.7, B and C).



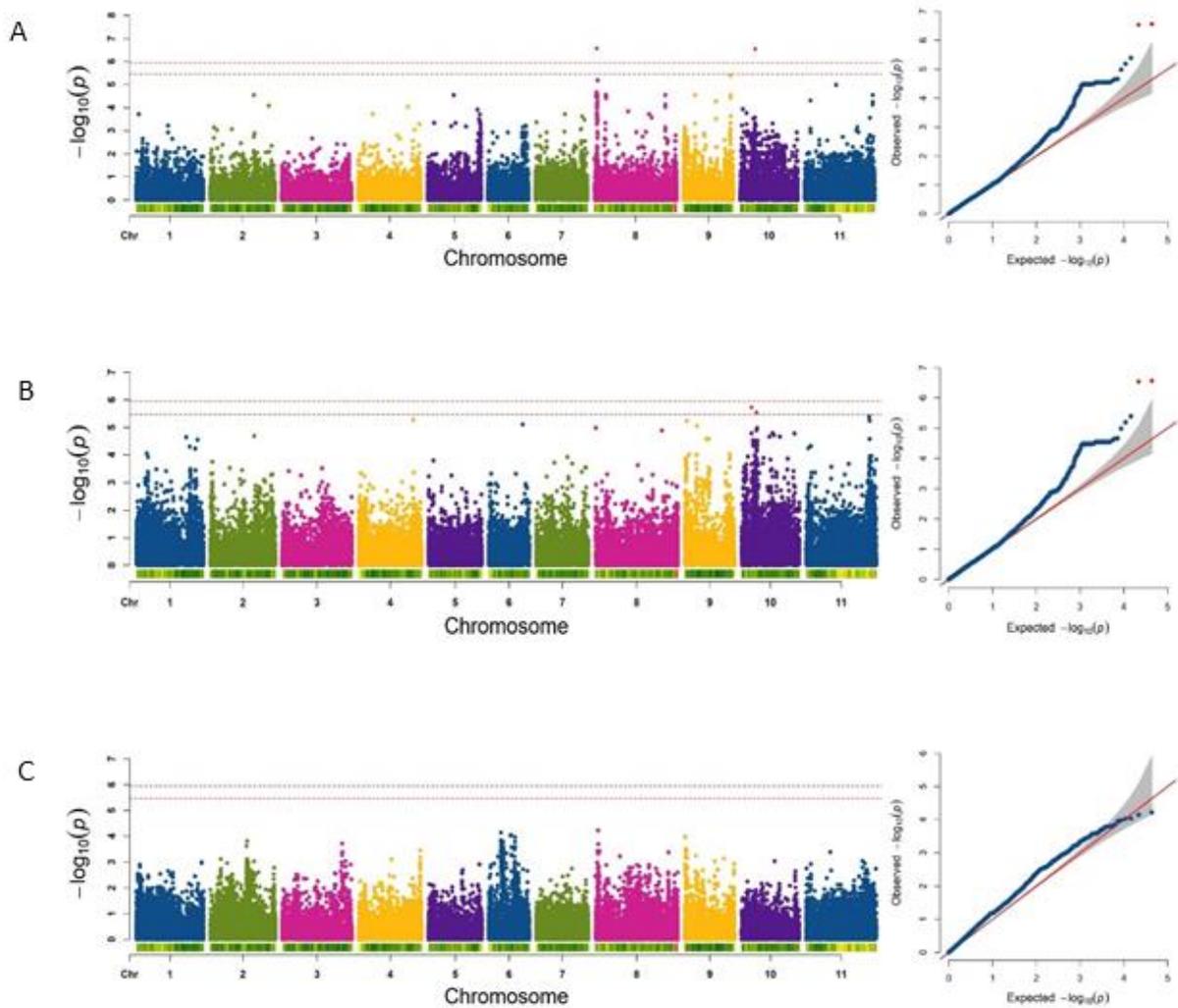
**Figure 5.7.** Manhattan and QQ plots for GWAS of 280 MDP genotypes and  $\delta^{15}\text{N}$  values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

**Table 5.7.** SNPs marking QTL significantly associated with Nitrogen fixation traits in the entire MDP, by location (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Chr	SNP rs#	SNP position (Mbp)	minus log <sub>10</sub> (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Number of candidate genes within 100 Kb centred on SNP
<b>Nitrogen difference (<math>\delta^{15}\text{N}</math>)</b>									
1	4,933,671	4.9	5.99E-07	A/G	1.0	0.4	-0.60	Elora	5
2	49,084,007	49.1	9.93E-12	A/G	0.6	1.1	0.50	Elora	15
7	22,082,166	22.1	5.50E-07	G/A	1.0	0.4	-0.63	Elora	1
8	1,759,458	1.8	5.57E-12	C/T	2.8	1.0	-1.80	Elora	11
8	8,377,081	8.4	8.66E-08	G/A	1.0	0.4	-0.63	Elora	6
10	9,818,036	9.8	1.37E-11	A/T	0.9	1.2	0.25	Elora	0
<b>%Ndfa</b>									
7	22,082,170	22.1	2.60E-07	A/G	52.6	62.4	9.82	Elora	1
8	1,759,458	1.8	7.52E-09	T/C	54.0	25.8	-28.18	Elora	13
8	8,377,081	8.4	7.22E-07	G/A	53.1	62.4	9.30	Elora	6
10	9,818,036	9.8	6.34E-11	A/T	54.8	50.8	-3.99	Elora	0
<b>Seed N content (%)</b>									
2	11,937,912	11.9	3.00E-10	A/G	4.2	3.7	-0.54	Elora	7
2	21,924,033	21.9	2.44E-07	G/A	4.2	4.5	0.24	Elora	2
5	10,781,169	10.8	1.45E-14	G/A	4.3	4.0	-0.29	Elora	4
7	32,785,603	32.8	2.20E-09	A/G	4.3	4.0	-0.25	Elora	12
11	30,570,090	30.6	3.31E-09	T/C	4.3	4.1	-0.13	Elora	1
2	26,691,470	26.7	2.41E-10	G/C	3.6	4.1	0.51	Belwood	5
2	34,858,759	34.9	7.49E-11	G/A	3.6	4.4	0.79	Belwood	7
4	28,756,110	28.8	1.91E-10	A/T	3.7	2.8	-0.84	Belwood	0
5	13,923,752	13.9	3.26E-12	C/T	3.7	3.5	-0.19	Belwood	0
5	28,555,185	28.6	6.21E-07	C/T	3.6	4.3	0.68	Belwood	0
7	14,622,857	14.6	4.76E-09	A/G	3.6	4.1	0.45	Belwood	3
7	21,374,917	21.4	9.67E-22	A/G	3.7	3.4	-0.34	Belwood	2
8	60,524,571	60.5	1.55E-08	T/G	3.7	3.4	-0.30	Belwood	9
10	3,745,502	3.7	8.25E-07	G/A	3.5	3.9	0.39	Belwood	7
10	24,982,398	25.0	7.74E-07	G/T	3.7	3.3	-0.42	Belwood	2
11	27,645,168	27.6	3.82E-08	C/T	3.6	4.0	0.39	Belwood	1
2	33,210,458	33.2	3.17E-07	G/A	3.9	4.1	0.26	Isabela	8
7	17,225,864	17.2	8.73E-07	C/T	3.9	3.8	-0.17	Isabela	1
9	14,695,454	14.7	1.64E-09	A/G	3.9	4.3	0.40	Isabela	7
11	39,678,873	39.7	7.99E-07	G/C	3.9	3.8	-0.08	Isabela	0

Chr	SNP rs#	SNP position (Mbp)	minus log <sub>10</sub> (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Number of candidate genes within 100 Kb centred on SNP
<b>Leaf chlorophyll content 2</b>									
2	21,626,534	21.6	1.44E-11	A/T	39.1	52.4	13.26	Elora	5
10	25,230,082	25.2	5.46E-11	C/T	38.3	45.5	7.21	Elora	0
7	972,126	1.0	6.73E-07	G/A	39.9	39.9	0.00	Isabela	17
7	38,392,817	38.4	2.39E-07	A/G	39.8	44.7	4.93	Isabela	10
8	55,420,128	55.4	6.42E-07	A/T	39.9	40.5	0.56	Isabela	7
11	25,870,481	25.9	2.03E-07	C/T	39.9	40.2	0.36	Isabela	0

For the MA subpopulation, significant associations were found at Elora on Pv08 and Pv10 (Figure 5.8, A). The most significant SNP occurred at 2.0 Mb (A/G) on Pv08 and it accounted for a decrease of 1.5 ‰  $\delta^{15}\text{N}$  (Table 5.8). This peak corresponded with that found in the MA subpopulation at Elora for %Ndfa (Table 5.8). When the FDR was adjusted to 10%, peaks on Pv10 (7.8 Mb and 11.6 Mb) were found for  $\delta^{15}\text{N}$  at Belwood (Figure 5.8, B), which corresponded to the same peaks found for %Ndfa in the MA subpopulation. No significant associations were found at Isabela for  $\delta^{15}\text{N}$  in the MA subpopulation (Figure 5.8, C). No significant associations were found at any location for  $\delta^{15}\text{N}$  in the DJ subpopulation (Appendix C, Figure S5.1).



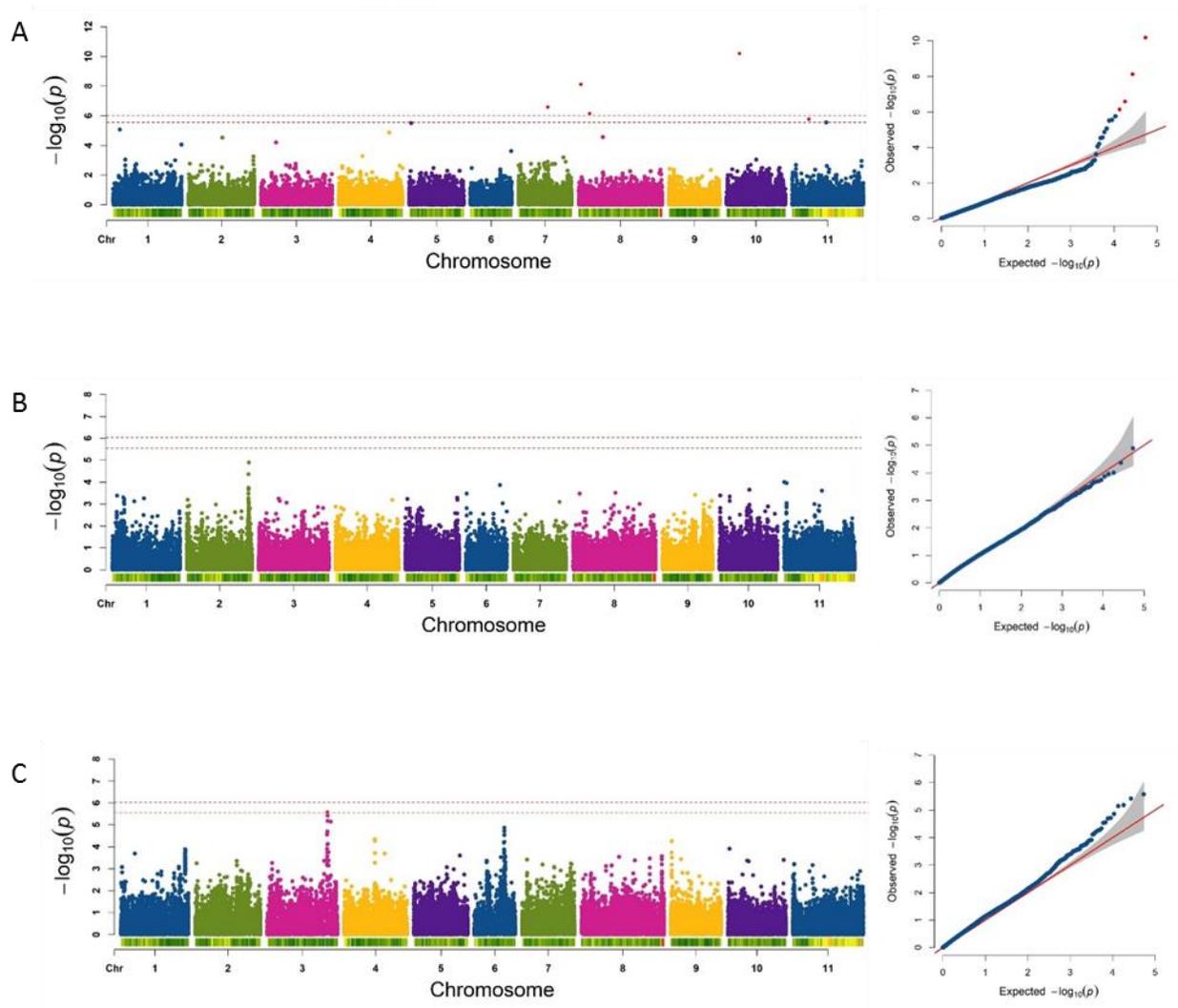
**Figure 5.8.** Manhattan and QQ plots for GWAS of 82 race MA MDP genotypes and seed nitrogen discrimination ( $\delta^{15}\text{N}$ ) values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceeded 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

**Table 5.8.** QTL significantly associated with nitrogen fixation traits in the MA subpopulation of the MDP, by location (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Chr	SNP rs#	SNP position (Mbp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location at which association significant	Number of candidate genes within 100 Kb centred on SNP
<b>Nitrogen difference (<math>\delta^{15}\text{N}</math>)</b>									
8	1,967,416	2.0	2.8748E-07	A/G	1.2	2.7	1.54	Elora	16
10	11,592,491	11.6	2.9066E-07	G/C	1.1	2.6	1.42	Elora	2
<b>%Ndfa</b>									
8	1,967,416	2.0	2.8748E-07	A/G	50.8	26.8	-24.04	Elora	16
10	11,592,491	11.6	3.0759E-07	G/C	51.0	28.8	-22.22	Elora	2
<b>Leaf chlorophyll content 2</b>									
9	33,332,147	33.3	6.79E-07	T/C	37.4	39.6	2.18	Isabela	8
9	33,332,084	33.3	7.08E-07	A/G	37.4	39.5	2.11	Isabela	8

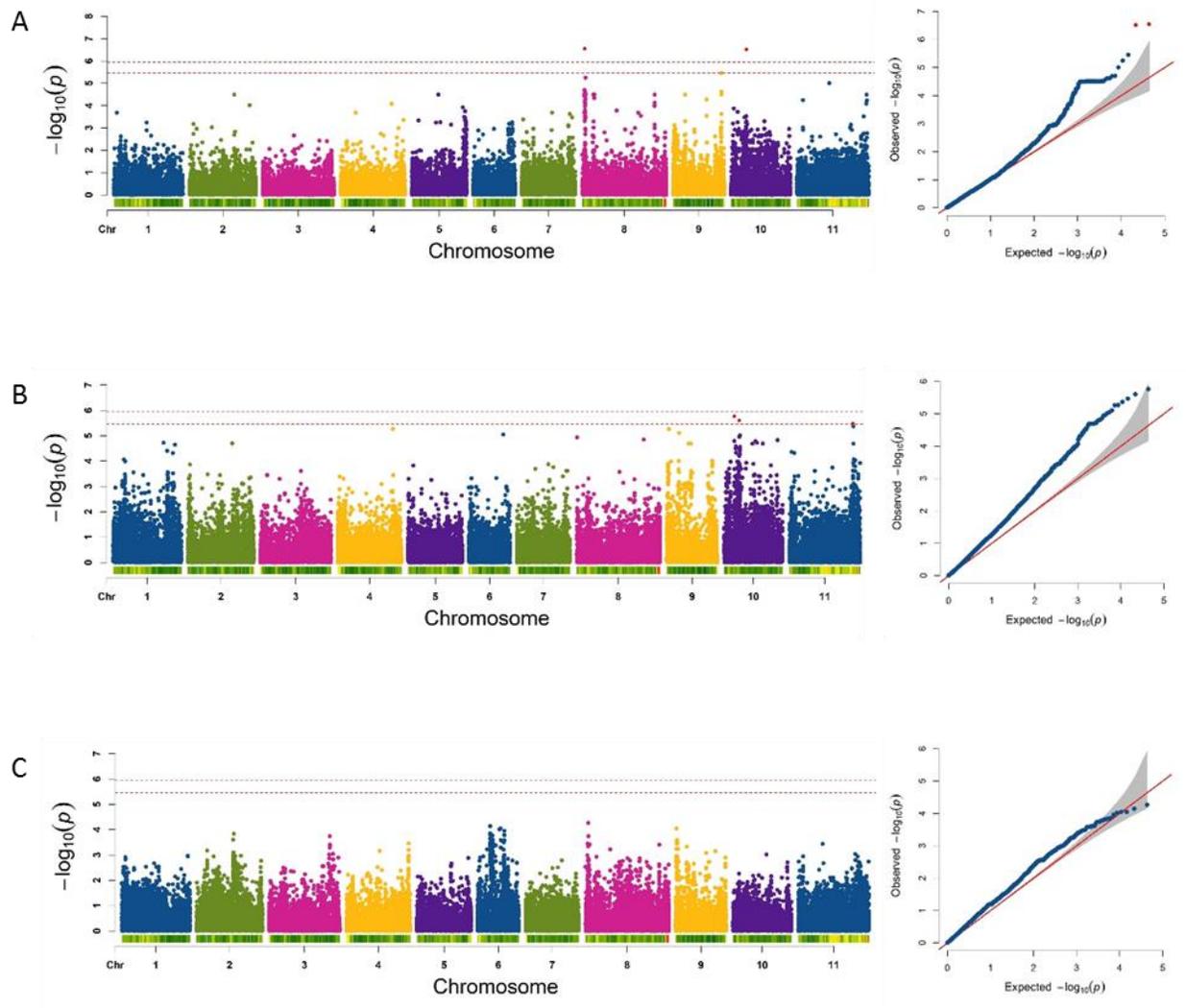
#### 5.4.5.2 Symbiotic nitrogen fixation performance

For %Ndfa in the MDP, significant associations were found at Elora on Pv07, Pv08, and Pv10 (Figure 5.9, A). The highest SNP peak occurred at Pv10 (9.8 Mb; A/T) and it accounts for a reduction of nitrogen fixation of approximately 4% (Table 5.7). The SNPs at 22.1 Mb (A/G) on Pv07 and at 8.4 Mb (G/A) on Pv08 each account for an increase of approximately 9 %Ndfa, while the SNP at 1.8 Mb (T/C) on Pv08 accounts for an approximate 28% reduction in %Ndfa (Table 5.7). No significant associations were found for the entire panel at Belwood nor Isabela at the 5% false discovery rate (FDR), however a significant SNP was identified at Isabela on Pv03 when the FDR was increased to 10% (Figure 5.9, B and C).



**Figure 5.9.** Manhattan and QQ plots for GWAS of 280 MDP genotypes and %Ndfa values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

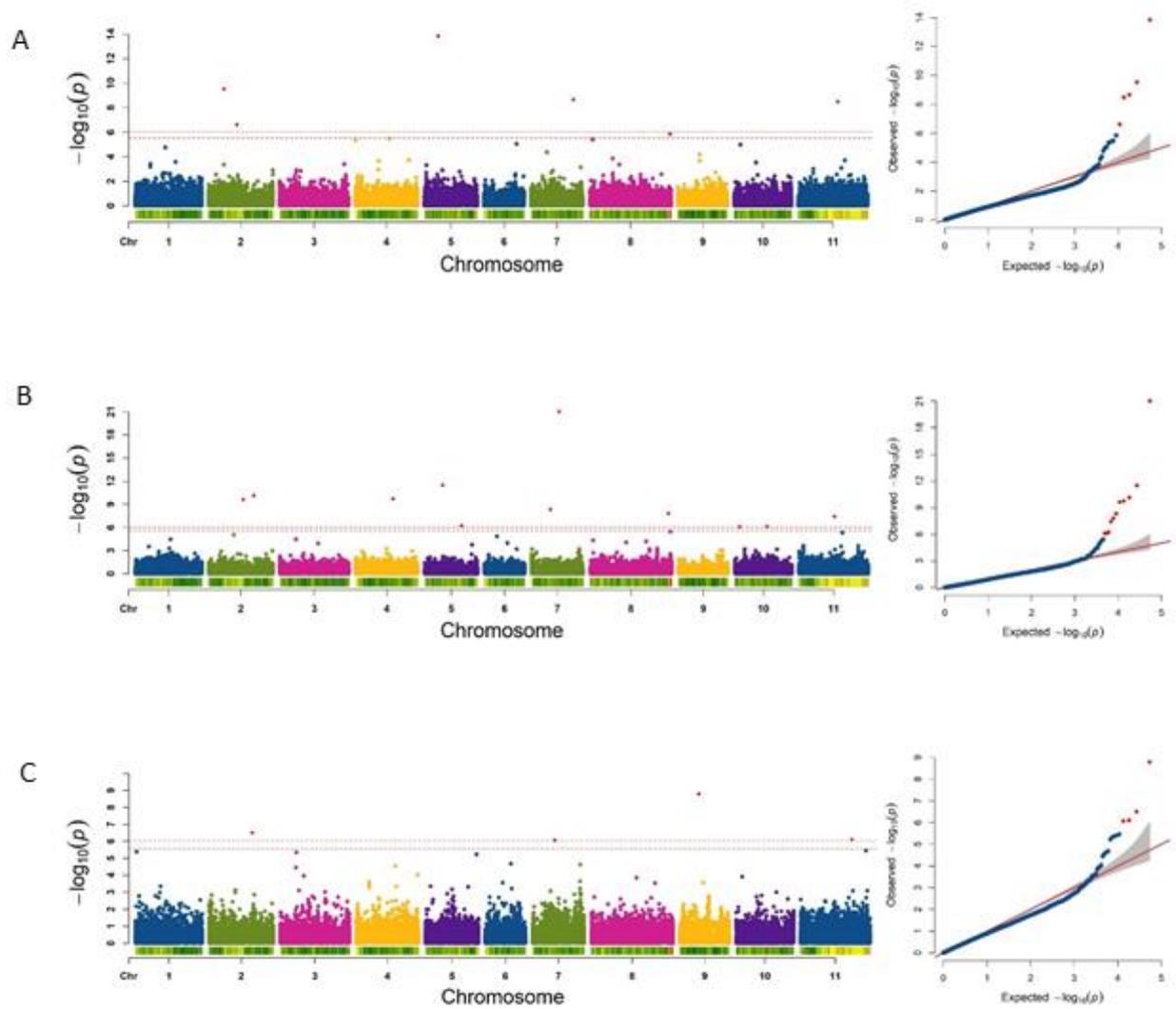
For the MA subpopulation, significant associations were found at Elora on Pv08 at 2.0 Mb (A/G) and on Pv10 at 11.6 Mb (G/C), and these SNPs accounted for 24% and 22.2% decreases in nitrogen fixation respectively (Figure 5.10, A; Table 5.8). The peak at 2.0 Mb on Pv08 for the MA subpopulation does not correspond to the peaks found for the entire panel at Elora on Pv08 (1.8 Mb and 8.4 Mb, Figure 5.14, A). No significant associations were found for the MA subpopulation at Belwood nor Isabela at 5% FDR, however two SNP peaks (Pv10 and Pv11) were identified at Belwood when the FDR was adjusted to 10% (Figure 5.10, B). No significant associations with %Ndfa were found for the DJ subpopulation at any location (Appendix C, Figure S5.2).



**Figure 5.10.** Manhattan and QQ plots for GWAS of 98 MDP genotypes belonging to the MA subpopulation and %Ndfa values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

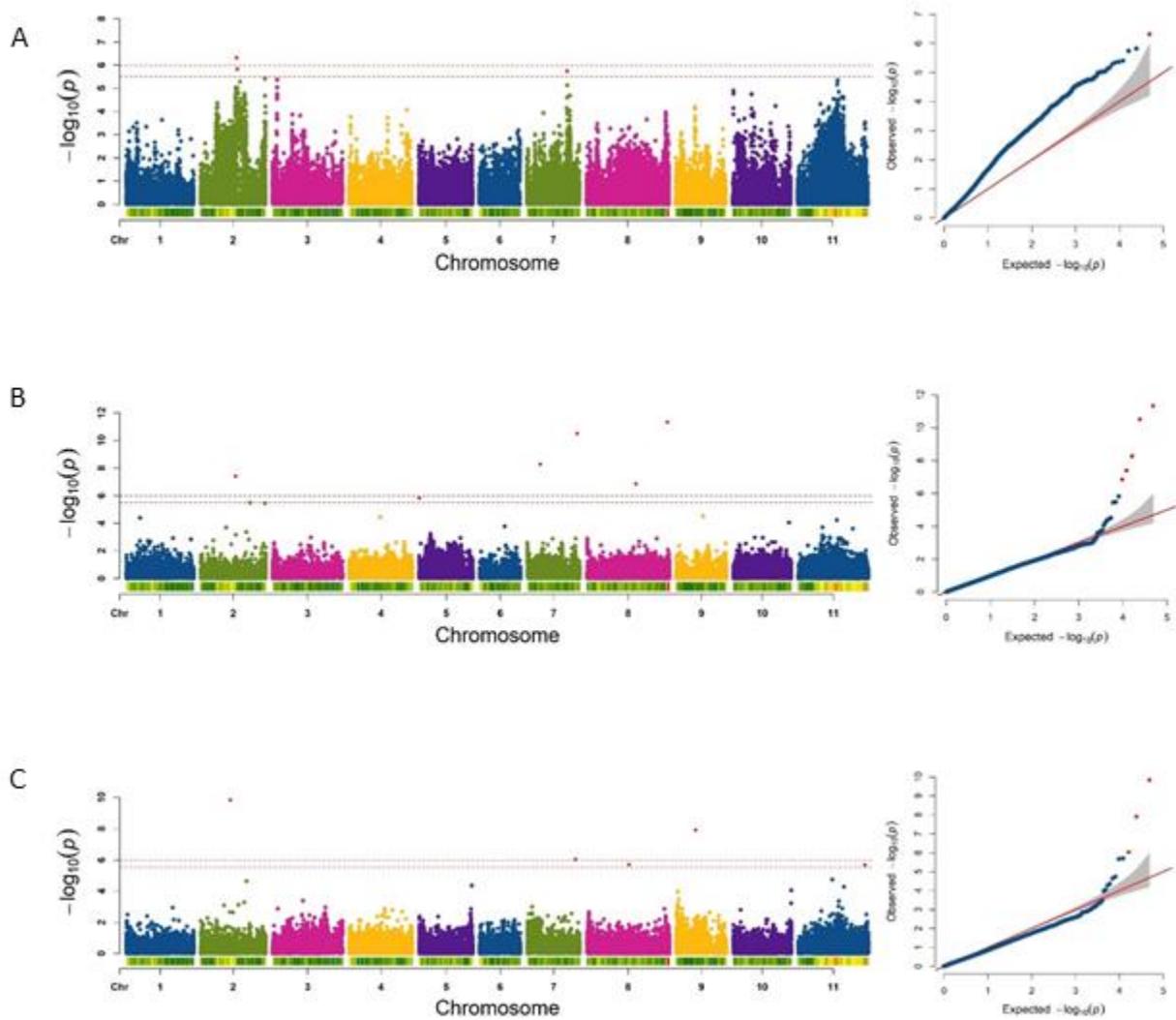
### 5.4.5.3 Seed nitrogen content

The greatest number of trait associations in this study were found for seed nitrogen content (%N). For %N in the MDP at Elora, significant associations were found on Pv02, Pv05, Pv07 and Pv11 (Figure 5.11, A). The SNP with the greatest effect on %N at Elora was at 11.9 Mb (A/G) on Pv02, which decreased seed N content by 0.54% (Table 5.7). At Belwood, significant associations were found on Pv02, Pv04, Pv05, Pv07, Pv08, Pv10 and Pv11 (Figure 5.11, B). The SNP with the greatest effect on %N at Belwood was at 28.8 Mb (A/T) on Pv04, which decreased seed N content by 0.84% (Table 5.7). For %N at Isabela significant trait associations were found on Pv02, Pv07, Pv09 and Pv11 (Figure 5.11, C). The SNP with the greatest effect on %N at Isabela was at 14.7 Mb (A/G) on Pv09, which increased seed N content by 0.40% (Table 5.7). None of the significant peaks for %N were shared between locations. No significant associations were found for %N for the MA subpopulation (Appendix C, Figure S5.3).



**Figure 5.11.** Manhattan and QQ plots for GWAS of 280 MDP genotypes and seed nitrogen content (%N). The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

For the DJ subpopulation, significant trait associations for %N were identified at each location (Figure 5.12). At Elora, a peak was found at 27.7 Mb (T/C) on Pv02 (Figure 5.12, A). This peak accounted for a small increase of seed N content of 0.16% (Table 5.9). At Belwood, peaks were found on Pv02, Pv07 and Pv08 (Figure 5.12, B). The peak at 26.7 Mb (G/C) on Pv02 had the greatest effect on seed N content, accounting for an increase of 0.62% (Table 5.9). In addition, this peak corresponds to that found for %N in the entire MDP panel at Belwood (Table 5.9). At Isabela, peaks were found on Pv02, Pv07 and Pv09 (Figure 5.12, C). The peak on Pv09 at 14.7 Mb (A/G) had the greatest effect on %N and accounted for an increase of 0.45% (Table 5.9). This peak corresponds to that found for %N in the entire panel at Isabela (Table 5.9).



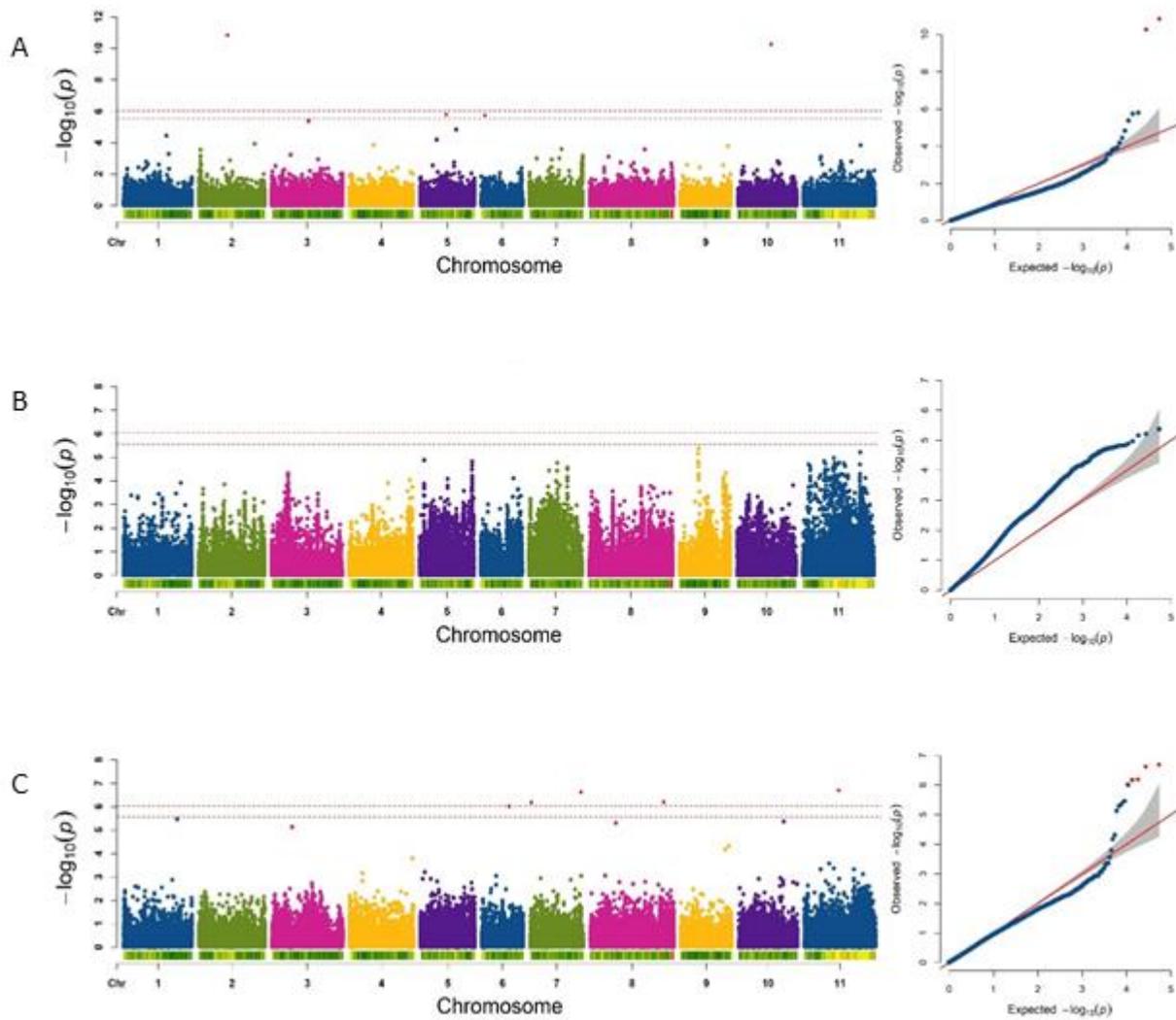
**Figure 5.12.** Manhattan and QQ plots for GWAS of 182 MDP genotypes belonging to the DJ subpopulation and %N values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

**Table 5.9.** QTL significantly associated with nitrogen fixation traits in the DJ subpopulation of the MDP, by location (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Chr	SNP rs#	SNP position (Mbp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Number of candidate genes within 100 Kb centred on SNP
<b>Seed N content (%N)</b>									
2	27,672,387	27.7	4.85274E-07	T/C	4.18	4.35	0.16	Elora	9
2	26,691,470	26.7	3.99E-08	G/C	3.52	4.14	0.62	Belwood	5
7	9,752,020	9.8	5.33E-09	C/T	3.43	3.89	0.46	Belwood	8
7	37,910,063	37.9	3.08E-11	A/G	3.61	3.46	-0.15	Belwood	9
8	37,482,649	37.5	1.43E-07	C/T	3.54	3.50	-0.04	Belwood	0
8	61,434,463	61.4	4.67E-12	G/A	3.56	3.43	-0.13	Belwood	13
2	22,757,513	22.8	1.43817E-10	T/C	3.76	4.00	0.24	Isabela	6
7	36,757,868	36.8	9.38584E-07	G/A	3.82	3.96	0.14	Isabela	11
9	14,695,454	14.7	1.23478E-08	A/G	3.84	4.28	0.45	Isabela	7

#### 5.4.5.4 Leaf chlorophyll content

No significant associations were found for leaf chlorophyll content at the early vegetative stage in the MDP nor the MA or DJ subpopulations (Appendix C, Figures S5.4-6). Significant associations were found for the entire panel for leaf chlorophyll content at flowering time on Pv02 and Pv10 at Elora (Figure 5.13, A). The SNP on Pv02 at 21.6 Mb (A/T) had the greatest effect, increasing SPAD ratings by 13.26 units (Table 5.7). At Isabela, significant associations were found on Pv07, Pv08 and Pv11 (Figure 5.13, C). The most significant SNP at Isabela was that at 38.4 Mb (A/G) on Pv07 that increased SPAD ratings by 4.93 (Table 5.7). No trait associations were found for leaf chlorophyll content at flowering time at Belwood for the entire panel (Figure 5.13, B).



**Figure 5.13.** Manhattan and QQ plots for GWAS of 280 MDP genotypes and leaf chlorophyll content at flowering values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

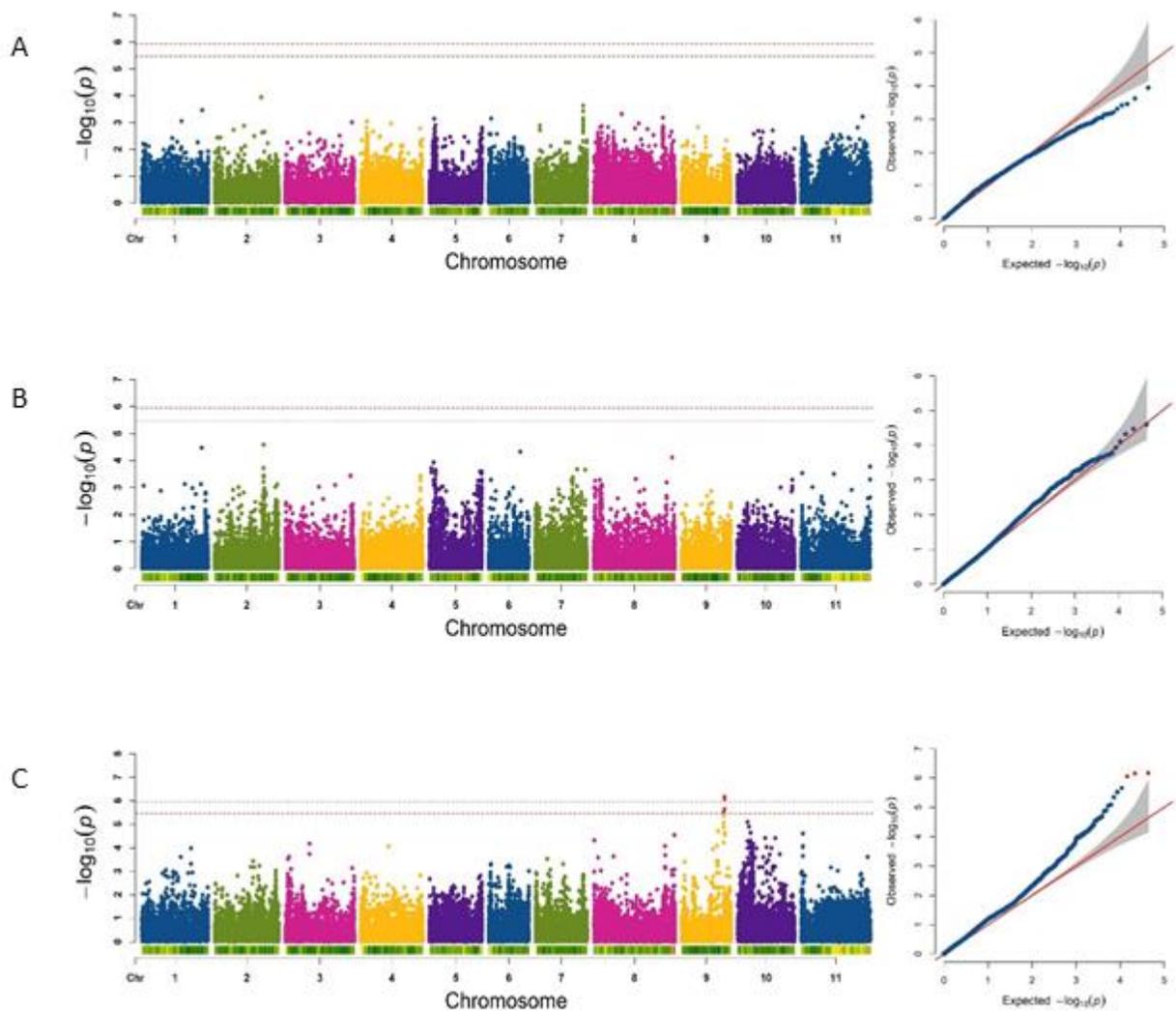
For the MA subpopulation, significant associations were found for leaf chlorophyll content at flowering only at Isabela (Figure 5.14, C). The two significant SNPs located were located at 33.3 Mb (T/C and A/G) on Pv09 each increased SPAD ratings slightly by just over 2 units (Table 5.8). No significant associations were found for SPAD for the DJ subpopulation (Appendix C, Figure S5.7).

#### **5.4.5.5 Nodule number and size**

No significant SNPs were identified for nodule number or nodule size at Isabela (Appendix C, Figures S5.8 and S5.9).

#### **5.4.6 SNP haplotype effects and candidate genes**

SNP haplotypes were determined for each significant SNP (5% FDR) discovered in the GWAS. Haplotype allelic effects were large in some instances and minor in others, and both positive and negative homozygous allelic effects were observed. JBrowse was used to search the bean genome within a 100 Kb region centered on significant SNPs to identify candidate genes, and those with apparent relevance to SNF are reported below. Certain candidate genes are described further in the Discussion section.



**Figure 5.14.** Manhattan QQ plots for GWAS of 98 MDP genotypes belonging to the MA subpopulation and leaf chlorophyll content at flowering values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.

#### 5.4.6.1 Symbiotic nitrogen fixation performance

For %Ndfa in the MDP at Elora, the SNP at 22.1 Mb (A/G) on Pv07 increased nitrogen fixation by 9.8 %Ndfa from 52.6 %Ndfa (AA) to 62.4 %Ndfa (GG) (Table 5.7). The nearest candidate gene to this SNP was a double Clp-N motif-containing P-loop nucleoside triphosphate hydrolase superfamily protein (Phvul.007G138900) (Appendix C, Table S5.8). The SNP at 8.4 Mb (G/A) on Pv08 increased nitrogen fixation by 9.3% from 53.1 %Ndfa (GG) to 62.4 %Ndfa (AA) (Table 5.7). There were six candidate genes located within the 100 Kb region centered on this SNP including a protein kinase family (Phvul.008G084600) (Appendix C, Table S5.8). The SNP at 1.8 Mb (T/C) on Pv08 reduced nitrogen fixation by 28.2% from 54 %Ndfa (TT) to 25.8 %Ndfa (CC) (Table 5.7). There were 13 candidate genes located within the 100 Kb region centered on this SNP. This SNP was located within a protein kinase superfamily gene (Phvul.008G021001) (Appendix C, Table S5.8).

Significant SNPs associated with %Ndfa were found in the MA subpopulation at Elora on Pv08 and Pv10 (Figure 14, A). The highest SNP peak occurred on Pv08 at 2.0 Mb (G/A), and does not correspond to the peak on Pv08 found for the entire panel. For the MA subpopulation the SNP on Pv08 had A as its major allele and G as the minor allele and it reduced nitrogen fixation by 24% from 50.8 %Ndfa (AA) to 26.8 %Ndfa (GG) (Table 5.8). The 100 Kb region centered on this SNP contains 16 candidate genes, including a TGACG-sequence-specific DNA-binding protein (Phvul.008G023800) (Appendix C, Table S5.9). The second SNP associated with

%Ndfa in the MA subpopulation at Elora was at 11.6 Mb (G/C) on Pv10. This SNP has G as its major allele and C as its minor allele and results in a reduction in nitrogen fixation of 22.2%, from 51.0 %Ndfa (GG) to 28.8 %Ndfa (CC) (Table 5.8). Two candidate genes are found within the 100 Kb region centered on this SNP and both are disease resistance-responsive (dirigent-like protein) family proteins (Phvul.010G063500 and Phvul.010G063600; Appendix C, Table S5.9).

#### **5.4.6.2 Seed nitrogen content**

The largest number of significant genomic associations found in our study were for seed nitrogen content. Multiple significant SNPs were identified for the entire panel at each location, however none of these were shared between locations. The SNP with the greatest effect on %N in the entire panel at Elora was at 11.9 Mb (A/G) on Pv02 and this SNP had A as its major allele and G as its minor allele, which resulted in a decrease of seed N content from 4.2% (AA) to 3.7% (GG) (Table 5.7). Seven candidate genes were found within a 100 Kb region centered on this SNP including, one encoding a germin-like protein (Phvul.002G077300), among other genes (Appendix C, Table S5.12). A second significant SNP on Pv02 at Elora located at 21.9 Mb had G as its major allele and A as its minor allele and increased seed N content from 4.2% to 4.5%. Underlying this SNP was a gene for UDP-glucose 6-dehydrogenase family protein (Phvul.02G102900). The SNP with the greatest positive effect on %N in the entire panel at Belwood was at 34.9 Mb (G/A) on Pv02 and this SNP had G as its major allele and A as its minor allele, which increased seed N content from 3.6% (GG) to 4.4% (AA) (Table 5.7). Seven candidate genes were found within the 100 Kb region centered on this SNP,

including three glutamate decarboxylases (Phvul.002G185800, Phvul.002G185900 and Phvul.002G186000) among other genes (Appendix C, Table S5.12). The SNP identified in the Belwood trial at 25 Mb (G/T) on Pv10 decreased seed N content by 0.42% from 3.7% (GG) to 3.3% (TT). Two non-specific lipid-transfer protein genes (Phvul.010G066400, Phvul.010G066500) were found within 100 Kb of that SNP. The SNP with the greatest effect on %N at Isabela was at 14.7 Mb (A/G) on Pv9 and this SNP had A as its major allele and G as its minor allele, which resulted in an increase in seed N content of 0.4% from 3.9% (AA) to 4.3% (GG) (Table 5.7). Seven candidate genes were located within the 100 Kb region centered on the SNP including gene for a bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein (Phvul.009G092000; Appendix C, Table S5.12).

For the DJ subpopulation at Elora one SNP was found on Pv02 at 27.7 Mb (T/C) and this peak had T as its major allele and C as its minor allele and it accounted for a small increase of seed N content of 0.16% from 4.18% (TT) to 4.35% (CC) (Table 5.9). There were nine candidate genes found within the 100 Kb region centered on this SNP including an isoflavone reductase homolog (Phvul.002G132400), three membrane protein genes (Phvul.002G132600, Phvul.002G132800 and Phvul.002G133200), and an early-responsive to dehydration stress (ERD) family protein gene (Phvul.002G133000), among other genes (Appendix C, Table S5.13). At Belwood the most significant SNP was found at 26.7 Mb (G/C) on Pv02 and had G as its major allele and C as its minor allele and it increased seed N content by 0.62% from 3.52% (GG) to 4.14% (CC) (Table 5.9). There were five candidate genes located within the 100 Kb

region centered on this SNP including three cytochrome P450 superfamily protein genes (Phvul.002G126000, Phvul.002G126100 and Phvul.002G126200) (Appendix C, Table S5.13). Another SNP identified at Belwood was found at 61.4 Mb (G/A) on Pv08 and was responsible of a 0.13% decrease in seed N content from 3.56% (GG) to 3.43% (AA). Among the 13 candidate genes underlying that SNP was an auxin-responsive family protein gene (Phvul.008G270000). At Isabela the SNP on Pv09 at 14.7 Mb had A as its major allele and G as its minor allele. This SNP corresponded to the most significant SNP found in the MDP overall at Isabela increasing seed N content by 0.4% from 3.9% (AA) to 4.3% (GG), but in the DJ subpopulation it had a greater effect on seed N content, increasing it by 0.45% from 3.52% (AA) to 4.28% (GG) (Table 5.9). Seven candidate genes were found within the 100 Kb region centered on this SNP including a bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein gene (Phvul.009G092000) among others (Appendix C, Table S5.13). A second significant SNP identified at Isabela in the DJ subpopulation was found on Pv02 at 22.7 Mb (T/C), which increased seed N content by 0.24% from 3.76% (TT) to 4.00% (CC) (Table 5.9). Six candidate genes are located within the 100 Kb region centered on this SNP including a rac-like GTP-binding protein gene (Phvul.002G106600) and an ascorbate peroxidase 3 gene (Phvul.002G107000), among other genes (Appendix C, Table S5.13). A third significant SNP identified at Isabela was found on Pv07 at 36.7 Mb (G/A), which had a small effect increasing seed N content by 0.14% from 3.82% (GG) to 3.96% (AA) (Table 5.9). Within the 100 Kb region centered

on this SNP 11 candidate genes were located including a nodulin-like major facilitator superfamily protein gene (Phvul.007G244600) (Appendix C, Table S5.13).

#### **5.4.6.3 Leaf chlorophyll content**

For the entire MDP, significant SNPs were identified for leaf chlorophyll content at flowering time at Elora and Isabela. The most significant SNP at Elora was located on Pv02 at 21.6 Mb and had A as its major allele and T as its minor allele and increased SPAD ratings by 13.26 units (Table 5.7). Five candidate genes were located within the 100 Kb region centered on this SNP, including an alpha/beta fold hydrolase gene within which the SNP is found (Phvul.002G102200; Appendix C, Table S5.14). The most significant SNP at Isabela was that at 38.4 Mb on Pv07 where A was the major allele and G was the minor allele and increased SPAD ratings by 4.93 units (Table 5.7). Ten candidate genes were located within the 100 Kb region centered on this SNP, including an F-box/LRR protein gene within which the SNP was found (Phvul.007G261900; Appendix C, Table S5.14).

When the panel was subdivided into the two race subpopulations, significant associations with leaf chlorophyll content at flowering were found only for the MA subpopulation and only at Isabela. Two significant SNPs, located near each other at 33.3 Mb on Pv09, were identified and each increased SPAD ratings by just over 2 units (Table 5.8). The first SNP had T as the major allele and C as the minor allele, and the second SNP had A as its major allele and G as its minor allele. Eight candidate genes were located within a 100 Kb region centered on these SNPs including a CHUP1

protein gene (chloroplastic-like isoform X2; Phvul.009G220500; Appendix C, Table S5.15).

## 5.5 Discussion

In order to increase the SNF capacity of modern dry bean varieties it is necessary to explore the diversity for the trait among currently used varieties and breeding lines and to gain an understanding of the genetic basis of SNF, which can be employed in future breeding endeavors. This study was initiated to characterize SNF in Middle American genotypes and to identify genomic regions associated with SNF using GWAS. Significant differences were seen among genotypes of the MDP for SNF-related traits, including %Ndfa and seed nitrogen content. The average nitrogen fixation capacity found in our study (Elora, 53.4%; Belwood, 58.1%) was higher than that reported for the Andean Diversity Panel (45.5%; Kamfwa et al., 2015), higher than that reported for the Bean Abiotic Stress Evaluation (BASE\_120) field trials comprised primarily of Middle American genotypes (50.5% and 29.0%; Oladzad et al., 2020), and within the range reported by other studies using Middle American genotypes (Farid et al., 2017; Heilig et al., 2017a, 2017b). Historically, common bean has been thought of as a poor nitrogen fixer compared to other legumes such as soybean (Herridge et al., 2008), however the average %Ndfa found in our trials (Elora 53.4 %Ndfa; Belwood 58.1 %Ndfa) was comparable to results from a meta-analysis of N traits in soybean where average N fixation was 58% in non-fertilized treatments (Salvagiotti et al., 2008).

Overall, genotype performance in our study was impacted by the growing environment and the top 20 genotypes differed each year, with Durango-Jalisco genotypes performing better at Elora in 2014 and Mesoamerican genotypes performing better at Belwood in 2015. This inconsistency cannot be easily explained by the environmental parameters we measured. Soil type and growing conditions were similar at Elora and Belwood, and although the Belwood season was more challenging with higher rainfall, delayed planting, and fewer growing degree days, these factors did not adversely affect nitrogen fixation averages. Soil NO<sub>3</sub>- levels were higher in the deeper soil profile at Elora, however, by soil testing standards they were not considered high. Some N fixation ranking stability was seen at the market class level particularly in small red and navy beans where a few varieties appeared in the top 10 each year. For each market class grown in Ontario there were varieties that fixed more than 70% of their nitrogen. The phenotypic diversity for SNF and the existence of high-fixing genotypes of different market classes indicates that there are genetic resources to be used in breeding to improve the trait in the Middle American genepool. Interestingly, a meta-analysis of soybean studies reported that when higher proportions of plant N are derived from %Ndfa, grain yields are driven relatively lower (Ciampitti and Salvagiotti, 2018), and N fixation is generally not high enough in soybean to support high-yielding varieties (>7 Mg ha<sup>-1</sup>) (Salvagiotti et al., 2008). The positive correlation between nitrogen fixation (%Ndfa) and seed N content (%N) shows that genotypes with a strong capacity for nitrogen fixation were also successful at transferring the fixed N to the seed.

### 5.5.1 Marker-trait associations

In the current study, a genetic characterization of SNF traits under low-nitrogen field conditions was conducted that adds to the several agronomic trait, abiotic stress tolerance, nutritional profile, and seed coat colour studies in dry beans (Hoyos-Villegas et al., 2016; Moghaddam et al., 2016, 2018; McClean et al., 2017, 2018; Soltani et al., 2017), as well as the wide body of work on SNF in common bean (including: Ramaekers et al., 2013; Farid and Navabi, 2015; Diaz et al., 2017; Heilig et al., 2017b, 2017a; Kamfwa et al., 2019; Oladzad et al., 2020). We identified QTL for %Ndfa in the MDP on Pv07, Pv08 and Pv10 at Elora 2014. These associations were not replicated at the other trial locations. Two of the QTL had a positive effect on %Ndfa, and for these QTL six MA genotypes (including Laker, Mayflower, BAT 477, Blackjack, UCD 96114 and UCD 9634) and three DJ genotypes (including Bill Z, Topaz and Gemini) carry the positive effect alleles, although no genotype possessed both. Topaz and BAT 477 have been investigated for SNF capacity in previous studies. The SNF performance of Topaz was largely unremarkable in a study of 22 dry bean genotypes for SNF traits (Akter et al., 2014). For example, nodulation and nitrogen fixation in Topaz were lower than the treatment means under both inoculants tested, and although Topaz had higher N<sub>2</sub> fixation than the treatment mean in the +N fertilizer treatment it was not the best performing genotype under that condition (Akter et al., 2014). BAT 477 has been well-characterized as a high-SNF performing genotype (Kipe-Nolt et al., 1993; Miklas et al., 2006; Diaz et al., 2017; Oladzad et al., 2020). Investigating SNF capacity under differing phosphorus levels using BAT 477 as a parent in a biparental population (DOR 364 ×

BAT 477), Diaz et al. (2017) reported a QTL for %Ndfa near the end of Pv07 at 154 cM (marker: g1947) that explained 21% of phenotypic variance for the trait. Examining SNF in the BASE\_120 field trials, Oladzad et al. (2020) found that BAT 477 had a positive allele for %Ndfa at 25.9 Mbp on Pv09. In our study, BAT 477 had the positive effect allele at 8.4 Mbp on Pv08. Miklas et al (2006) suggested that BAT 477 may possess multiple SNF genes, which contribute to its SNF capacity under differing abiotic stress conditions, and the results of our study and that of Diaz et al. (2017) support that conclusion.

In this work, nitrogen fixation and seed N content were positively correlated. This indicates that under conditions of low soil-available nitrogen, genotypes which are efficient at nitrogen fixation are also efficient at transporting N to the seed, thus increasing economic yield. A total of 20 QTL were identified in the MDP for seed N content. QTL were identified for this trait at each trial location although none were replicated between locations. A further nine QTL for %N were found in the DJ subpopulation, two of which were the same as QTL identified in the MDP; namely, at Belwood the QTL at 26.7 Mbp on Pv02 in the MDP was also identified in the DJ subpopulation, and at Isabela the QTL at 14.7 Mbp on Pv09 for the MDP was also identified in the DJ subpopulation.

Interestingly, for a few QTL the minor homozygous allelic state was the dominant condition in either one or the other subpopulation. For example, in the MDP at Belwood the positive effect QTL for seed %N, tagged by the SNP (G/A) at 3.7 Mbp on Pv10, was

predominantly homozygous for the minor allele (AA) in the MA subpopulation but it rarely had this state in the DJ subpopulation. A second example of this was seen for the significant SNP (A/G) at 21.4 Mbp on Pv07 at Belwood, which reduced %N. In this case, this SNP was predominantly homozygous for the minor allele (GG) in the DJ subpopulation but rarely so in the MA subpopulation. Were markers to be designed for %N using these QTL, their usability may be restricted to whichever subpopulation they were discovered in. This underscores the importance of the awareness of population structure in GWAS studies.

In our study, a positive correlation was also found at Elora between %Ndfa and leaf chlorophyll content at flowering time. The use of SPAD as a proxy trait to identify high-SNF genotypes has been explored by numerous studies with mixed results, and indeed we only found the positive association between these traits at one location. Topaz pinto bean was homozygous for the positive effect allele for %Ndfa on Pv07 and was homozygous for the positive effect allele for leaf chlorophyll content on Pv02. This genotype may be useful in further studies investigating SPAD as a proxy trait for SNF. Remote sensing using a mobile cart with canopy-height sensors was attempted at Isabela, however the results were deemed not useable after data processing. Further studies using a mobile cart or drone technology would be of interest to investigate nitrogen fixation throughout the growing season.

### 5.5.2 QTL for SNF traits: MDP vs. ADP

The results presented here on nitrogen fixation in the MDP, together with Kamfwa et al's (2015) study of nitrogen fixation in the ADP, represent a comprehensive analysis of nitrogen fixation capacity in modern common bean. These studies used over 500 genotypes and identified numerous QTL for nitrogen fixation and related traits that, once validated, can be employed in future breeding efforts. It is of interest to determine whether any QTL for SNF traits were identified in both studies and would have common use across the bean gene pools and market classes.

*%Ndfa in seed.* In the MDP we found QTL for %Ndfa (seed) on Pv07, Pv08 and Pv10. No significant SNPs for %Ndfa in the seed were identified in the ADP, however the study identified significant SNPs for %Ndfa in the shoot on Pv02, Pv03, Pv07, Pv09, Pv10 and Pv11 (Kamfwa et al, 2015). None of the %Ndfa QTL revealed in our study co-localized with those identified in the ADP.

*%N in seed.* In the ADP, Kamfwa et al (2015) identified 16 SNPs significantly associated with seed %N content on Pv09, one on Pv02 and one on Pv03. In the MDP, we found 29 QTL for seed %N across nine chromosomes, including eight on Pv02, one on Pv03, and two on Pv09. None of the QTL revealed in our study co-localized with the QTL identified in the ADP.

*Proxy traits.* Leaf chlorophyll content and nodule traits have been found to be associated with SNF capacity in some studies. In their field study, Kamfwa et al (2015) identified significant SNPs for leaf chlorophyll content on Pv01 and Pv09. QTL for leaf

chlorophyll content at flowering in the MDP were found in our study on Pv02, Pv07, Pv08, Pv09, Pv10 and Pv11. The QTL on Pv09 identified in each diversity panel were not located in the same region of the chromosome. Nodulation in the ADP was scored in the field and Kamfwa et al (2015) reported two SNPs on Pv09. Nodule size and nodule number were measured in our study of the MDP, however no significant SNPs were identified.

The identification of different QTL for SNF traits in the ADP and the MDP could be explained by a number of factors. Firstly, nitrogen fixation is a quantitative trait impacted by environmental conditions and dependent on the presence of an effective bacterial symbiont, among other growing conditions. The *Rhizobium* inoculant, in particular, was variable among the studies. Secondly, the germplasm used in each of these studies belongs to diverged genepools, which evolved in Central and South America. Although the genepools share a common wild ancestor, it is plausible that different genes and processes would evolve to influence aspects of SNF efficiency in each genepool. Thirdly, Kamfwa's study of the ADP used a smaller set of markers (Illumina BARC-Bean6K\_3 BeadChip, filtered to 5,326 SNPs), whereas we used a larger set of 54,185 SNPs generated by GBS. The greater density of the GBS markers may have enabled us to identify more QTL, including smaller effect QTL. Performing a new GWAS analysis with the ADP phenotypic data collected by Kamfwa et al. (2015) with the now-available ADP GBS SNP dataset (Oladzad et al., 2019) would be of interest. Beyond comparison with the ADP, it is also worth noting that there were no common QTL between those identified in our study and those found in recent studies

using biparental crosses within and between the Middle American and Andean gene pools (Farid, 2015; Heilig et al., 2017a; Kamfwa et al., 2019). Further, the recently published GWAS of nodulation and SNF in the BASE\_120 field trials (Oladzad et al., 2020) did not reveal the same QTL associated with %Ndfa as were identified in the present study. To reiterate, the quantitative nature of nitrogen fixing capacity and the sensitivity of this trait to environmental conditions, soil nutrient levels, and the efficacy of the relationship established with the Rhizobial symbiosis partner all potentially influence the suite of genes responsible for the SNF response, which in turn would impact discoverable QTL. While repeatable QTL for SNF would be desirable, identification of unique QTL is also valuable and provides further understanding of the genetic controls of SNF in common bean.

### **5.5.3 Candidate genes associated with significant SNPs**

We further examined the QTL identified in our study to determine which genes were likely candidates related to each SNF trait under investigation. Confirmation of which QTL/genes are related to SNF is necessary before they can be used for modern molecular breeding. Numerous candidate genes were found within close proximity to the QTL identified in our GWAS study. In most cases it was possible to make logical deductions which candidate genes might impact SNF traits. Further validation of these candidate genes using advanced methods such as RNA-seq to measure gene expression levels will be necessary.

### **5.5.3.1 Percent nitrogen derived from the atmosphere (%Ndfa)**

The candidate gene associated with the SNP on Pv07 for %Ndfa in the MDP at Elora encodes a triphosphate hydrolase superfamily protein. A transcriptome profiling study in chickpea inoculated with *Mesorhizobium* found that triphosphate hydrolase proteins were expressed during the nodulation process (Kaur et al., 2017). The two QTL identified on Pv08 (1.8 Mbp and 8.4 Mbp) contain protein kinase family genes (Phvul.008G021001 and Phvul.008G084600). Protein kinases are a large family of plant enzymes involved in diverse cellular regulation and metabolic processes (Stone and Walker, 1995). Certain protein kinases are involved in signaling, such as the calcium/calmodulin-dependent protein kinase (CCaMK), which perceives calcium oscillations activated by the host plant-microbe interaction (Oldroyd and Downie, 2004). Effective CCaMK genes are required for establishing a strong SNF relationship between host and rhizobia. Within the MA subpopulation at Elora a QTL was identified on Pv08 that was associated with a gene for a TGACG-sequence-specific DNA-binding protein (Phvul.008G023800). TGACG-binding transcription factors are involved in plant defense in *Arabidopsis* and some forms may play a role in symbiosis in legumes (Udvardi et al., 2007; Ullah et al., 2019).

### **5.5.3.2 Seed N content**

Candidate genes associated with seed nitrogen content were identified on multiple chromosomes, within the entire MDP and the DJ subpopulation and at each trial location. On Pv02 a series of candidate genes with different metabolic functions associated with nodules were found in the entire MDP and within the DJ subpopulation

at each trial location. Underlying the QTL at 11.9 Mbp on Pv02 at Elora was a germin-like protein gene (Phvul.002G077300). Germin-like proteins have been found to be associated with nodules in pea (*Pisum sativum*) (Gucciardo et al., 2007). Underlying the QTL at 21.9 Mb on Pv02 at Elora was a UDP-glucose 6-dehydrogenase family protein gene (Galvez et al., 2005; Marx et al., 2016). UDP-glucose levels were reduced in pea nodules that had been subjected to water stress, likely because the oxidative pentose phosphate pathway was inhibited reducing carbon supply to bacteroids (Galvez et al., 2005). UDP-glucose 6-dehydrogenase is also known to play a role in the *Medicago truncatula/Sinorhizobium meliloti* symbiosis (Marx et al., 2016).

Underlying the QTL at 32.8 Mbp on Pv07 at Elora were two MADS-box transcription factor family protein genes (Phvul.007G205500, Phvul.007G205700). In soybean (*Glycine max*), a MADS-box transcription factor (GmNMHC5) was able to sense sucrose and play a role in nodulation and root development (Liu et al., 2015). Underlying the QTL at 26.7 Mbp on Pv02 at Belwood were three Cytochrome P450 superfamily protein genes (Phvul.002G126000, Phvul.002G126100, Phvul.002G126200). This QTL was also identified in the DJ subpopulation at Belwood along with a second location on Pv07 (37.9 Mbp) at Belwood, which also had three Cytochrome P450 superfamily protein genes (Phvul.007G257300, Phvul.007G257400, Phvul.007G257500). Cytochrome P450 monooxygenases are membrane-bound proteins involved in diverse metabolic processes in plants. Guttikonda et al. (2010) carried out a whole-genome analysis of Cytochrome P450 genes in soybean and found four nodulation-specific genes, which suggests a role for this gene family in symbiosis.

In other studies, a transcriptome analysis study in bean found that the Phvul.002G126200 Cytochrome P450 protein was upregulated under phosphorus-deficient conditions (Silva et al., 2019), and a study in Rhizobia found a cytochrome P450 homolog had a role in symbiosis (Luyten et al., 2001).

Underlying the QTL at 34.9 Mbp on Pv02 at Belwood were three glutamate decarboxylase genes (Phvul.002G185800, Phvul.002G185900, Phvul.002G186000). The glutamate decarboxylase enzyme is responsible for converting glutamate to  $\gamma$ -aminobutyric acid (GABA) (Baum et al., 1996) and studies in *Medicago truncatula* have indicated that GABA may have role in increasing nodule activity (Sulieman and Schulze, 2010).

Underlying the QTL at 21.4 Mbp on Pv07 at Belwood was an ethylene-responsive transcription factor gene. Ethylene has a role in nodule regulation and has been shown to reduce nodulation in bean (Grobbelaar et al., 1971) and pea (Lee and LaRue, 1992), and recent transcriptomic studies in *Lotus japonicus* have indicated that a previously uncharacterized ethylene-responsive transcription factor may have a role in nodule function (Pérez-Delgado et al., 2020).

Underlying the QTL at 60.5 Mbp on Pv08 at Belwood was a ferritin 4 gene (Phvul.008G258100). Ferritins are iron storage proteins, and iron is an essential micronutrient for efficient SNF because it is required for building leghemoglobin and nitrogenase. Both iron deficiency and overabundance can have adverse effects on SNF. In *Lotus japonicus* a study examined the effect of varying ferritin expression levels on

nitrogen fixation and found that the co-expression of ferritin and oxidative stress response genes had a positive impact on SNF efficiency and duration (Chikoti et al., 2020).

Underlying the QTL at 25.0 Mbp on Pv10 at Belwood were two non-specific lipid transfer protein genes (Phvul.010G066400 and Phvul.010G066500), and underlying the QTL at 14.7 Mb on Pv09 at Isabela was a bifunctional inhibitor/lipid transfer protein/seed storage 2S albumin superfamily protein (Phvul.009G092000). In the *Medicago truncatula/Sinorhizobium meliloti* symbiosis a non-specific lipid transfer protein (MtN5) was essential for successful symbiosis, possibly interacting with the bacterial membrane and having a role in symbiosis partner signaling (Pii et al., 2013). In a study of Chinese Milk Vetch, a nodule-specific lipid transfer protein was identified, which plays an essential role in transporting plant lipids to the symbiosome membrane enabling effective nodule formation and nitrogen fixation (Lei et al., 2014).

Near the QTL on Pv08 at in the DJ subpopulation at Belwood an auxin-responsive family protein was found. During the nodulation process auxin is involved in cell cycle control, vascular tissue differentiation, and symbiont infection (Kohlen et al., 2018). In a study exploring genetic controls in the *M. truncatula/S. meliloti* symbiotic relationship, Breakspear et al. (2014) found four auxin-responsive genes that were induced in *M. truncatula* root hairs by *S. meliloti*. Near the QTL on Pv02 in the DJ subpopulation at Isabela a rac-like GTP-binding protein gene was found. Rac-like GTP-binding proteins belong to a group of small GTPases some of which have been shown

to function in infection and symbiosome formation in legume root nodules (Flores et al., 2018). A second putative candidate gene in this region of Pv02 was an ascorbate peroxidase 3 gene (Phvul.002G107000). Ascorbate peroxidase is found in legume nodules where it reduces oxygen levels to enable the efficient function of leghemoglobin and create conditions favorable for N-fixation (Ross et al., 1999). Underlying the QTL on Pv07 at 36.8 Mbp in the DJ subpopulation at Isabela was a nodulin-like major facilitator superfamily protein gene (Phvul.007G244600). Kamfwa et al. (2017) found this transporter protein to be upregulated in nodules of a strong nitrogen-fixing line (SA36) in their study.

### **5.5.3.3 Leaf chlorophyll content**

Leaf chlorophyll content is a potential indicator of nitrogen fixation efficiency and was correlated with %Ndfa in the Elora trial. Candidate genes associated with leaf chlorophyll content QTL were related to chlorophyll not specifically nitrogen. For example, one candidate gene (Phvul.002G102300) underlying the QTL on Pv02 at 21.6 Mbp was a trehalose phosphate synthase (TPS), which is involved in the trehalose biosynthetic pathway. The trehalose pathway coordinates carbon supply with metabolic processes including photosynthesis. In a study on tobacco grown under nitrogen-deficient conditions, Lin et al., 2017 found that exogenously applied trehalose (Tre) improved nitrogen metabolism and increased leaf chlorophyll and total nitrogen content compared to the control. It is possible that the expression of Tre genes in our bean study under low nitrogen field conditions had positive effects on leaf chlorophyll content.

At Isabela an F-box/LRR protein gene was located near the leaf chlorophyll content QTL on Pv07. LRR (leucine rich repeat) proteins belong to the resistance (R) class of plant proteins and some R genes are involved in legume-rhizobia symbiosis. For example, in the soybean/*Bradyrhizobium japonicum* and soybean/*Sinorhizobium fredii* symbioses two R genes were identified, which control host-symbiont recognition (Yang et al., 2010). In the GWAS of nitrogen fixation in the bean Andean Diversity panel, Kamfwa et al (2015) identified multiple LRR candidate genes associated with SNF-related QTL, although these were primarily of the receptor-like kinase (RLK) group.

Within the MA subpopulation, the QTL for leaf chlorophyll content on Pv09 may be caused by a CHUP1, chloroplastic-like isoform X2 gene. The CHUP1 protein is involved in chloroplast positioning within the cell and its structure prevents chloroplast aggregation (Oikawa et al., 2008).

## **5.6 Conclusion**

In this study, I explored the phenotypic diversity for SNF and related traits and broadened the understanding of the genetic architecture of SNF in the Middle American gene pool of common bean. Combined with SNF information previously-described for the Andean gene pool, bean breeders now have a broad understanding of SNF in modern germplasm. The diversity for nitrogen fixation capacity in all market classes represented in the MDP can be functionally employed in future breeding endeavors to develop varieties with improved SNF capacity. GWAS facilitated the identification of QTL associated with SNF and related candidate genes were described. Future studies

to confirm the functional role of the potential SNF candidate genes and to investigate the applicability of the QTL to other germplasm will broaden the impact of the results of this study on bean breeding and variety development.

## **6 General Conclusion**

Forty years ago, a defining review of SNF in common bean by P. H. Graham (1981) was published. In that paper, the complex factors impacting SNF in beans were outlined, acknowledging the importance of the Rhizobium symbiont, the impact of environmental and soil conditions, as well as the role of the plant, and the interaction of all of these factors. Despite the overall bleak outlook conveyed, Graham offered hope for making gains through plant breeding stating that, “the genetic diversity of the host ... offers possibilities not readily apparent in most other species” (Graham, 1981). The results presented in this thesis support the further improvement of SNF in common bean through plant breeding. Specifically, the studies explored the phenotypic and genetic diversity for SNF, found promising high-SNF capacity genotypes, identified QTL associated with SNF, and uncovered putative genes regulating SNF capacity.

Heirloom and Honduran landrace varieties were found to have useful phenotypic and genetic diversity for nitrogen fixation. Overall, heirloom and landrace varieties outperformed conventionally-bred and participatory-bred varieties for SNF. Additionally, yield performances were similar among germplasm categories in most environments. These findings support the use of diverse bean germplasm in breeding programs to improve SNF. Further studies characterizing heirloom germplasm compatibility for modern breeding could involve scoring disease susceptibility, growth habit suitability for mechanized cultivation and harvest, drought tolerance, and profiling nutritional compounds. A few heirloom genotypes evaluated here have been used in further research on SNF and in breeding at the University of Guelph. Similar characterization studies could be carried out on landrace varieties, with consideration of the needs of the

farmers or indigenous groups where the landraces originate. Using participatory and demand-led breeding objectives when developing varieties for marginalized growers will have the greatest impact on food- and economic security. Characterization and conservation of heirloom and landrace germplasm will preserve genetic diversity for use in the future to improve climate resiliency.

The commonly held belief that genetic diversity has been eroded in modern crops was challenged by the findings presented in Chapter 3. Heirloom and conventionally-bred beans were found to have similar levels of genetic diversity, and within the Middle American genepool nucleotide diversity was similar between the two breeding categories. These results suggest that modern breeding has not reduced genetic diversity in domesticated common bean. Expanding this study to a larger panel of germplasm would increase the confidence and applicability of these findings.

Characterizing the MDP for SNF capacity has contributed to the extensive phenotypic dataset already available for the panel, including agronomic performance and nutritional profile information. Additionally, the combined characterization of the ADP and the MDP for SNF provides a comprehensive picture of the diversity for this trait in modern common bean. An understanding of the current level of diversity present in bean provides a basis upon which to develop breeding initiatives to improve the trait.

The GWAS uncovered multiple novel QTL for SNF capacity and putative candidate genes were proposed. Many of the candidate gene types described here have functions apparently related to nitrogen fixation, however, further studies are

needed to confirm their relationship to SNF in bean. Functional genomics techniques such as gene-expression profiling, transcriptomics and proteomics have already been used in model legumes and soybean to elucidate gene functions in SNF and related metabolic processes (Guttikonda et al., 2010; Cañas and Beltrán, 2018), and these methodologies could be employed in common bean. Upon confirmation of the role of a gene in SNF, molecular markers could be developed, tested in germplasm of interest, and used for marker assisted selection (MAS) in breeding. The efficiency of selection for major- and minor-effect QTL should be improved by using MAS, resulting in the accelerated release of promising efficient SNF varieties.

Further research could capitalize on the genomic synteny which exists between common bean and soybean or model legumes to explore regions of the bean genome to find homologous genes already known to be associated with SNF in other species. Another study could examine the relationship between bean root hair traits such as density and length (Jochua et al., 2020) and nodulation.

Among researchers and growers, common bean has had a reputation for poor nitrogen fixation that has been difficult to counter. The studies presented in this thesis expand the body of knowledge on SNF in beans, and underscore the fact that useful diversity exists in the crop, which could be employed in breeding programs. Improving nitrogen fixation capacity in common bean has important implications for addressing world hunger and climate change.

## REFERENCES

- Aguilera, A., Smith, T. H., and Pauls, K. P. (2019). Cultivar Description: Small Red Bean 'OAC Rosito.' Available at: [https://www.gobeans.ca/downloads/OAC\\_Rosito-SmallRed.pdf](https://www.gobeans.ca/downloads/OAC_Rosito-SmallRed.pdf) .
- Akter, Z., Pageni, B. B., Lupwayi, N. Z., and Balasubramanian, P. M. (2014). Biological nitrogen fixation and nifH gene expression in dry beans (*Phaseolus vulgaris* L.). *Can. J. Plant Sci.* 94, 203–212. doi:10.4141/CJPS2013-200.
- Almekinders, C. J. M., and Elings, A. (2001). Collaboration of farmers and breeders: Participatory crop improvement in perspective. *Euphytica* 122, 425–438. doi:10.1023/A:1017968717875.
- Aserse, A. A., Markos, D., Getachew, G., and Yli-halla, M. (2020). Rhizobial inoculation improves drought tolerance, biomass and grain yields of common bean (*Phaseolus vulgaris* L.) and soybean (*Glycine max* L.) at Halaba and Boricha in Southern Ethiopia. *Arch. Agron. Soil Sci.* 66, 488–501. doi:10.1080/03650340.2019.1624724.
- Assefa, T., Mahama, A. A., Brown, A. V., Cannon, E. K. S., Rubyogo, J. C., Rao, I. M., et al. (2019). A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 39, 23. doi:10.1007/s11032-018-0920-0.
- Ayyappan, V., Kalavacharla, V., Thimmapuram, J., Bhide, K. P., Sripathi, V. R., Smolinski, T. G., et al. (2015). Genome-wide profiling of histone modifications (H3K9me2 and H4K12ac) and gene expression in rust (*Uromyces appendiculatus*) inoculated common bean (*Phaseolus vulgaris* L.). *PLoS One* 10, 1–29. doi:10.1371/journal.pone.0132176.
- Barbosa, N., Portilla, E., Buendia, H. F., Raatz, B., Beebe, S., and Rao, I. (2018). Genotypic differences in symbiotic nitrogen fixation ability and seed yield of climbing bean. *Plant Soil* 428, 223–239. doi:10.1007/s11104-018-3665-y.
- Bassett, M. J. (1991). A Revised Linkage Map of Common Bean. *HortScience* 26, 834–836.
- Baudoin, J.-P., and Maquet, A. (1999). Improvement of protein and amino acid contents in seeds of food legumes. A case study in *Phaseolus*. *Biotechnol. Agron. société Environ.* 3, 220–224. Available at: <https://popups.uliege.be/1780-4507/index.php?id=15504>.
- Baum, G., Lev-Yadun, S., Fridmann, Y., Arazi, T., Katsnelson, H., Zik, M., et al. (1996). Calmodulin binding to glutamate decarboxylase is required for regulation of glutamate and GABA metabolism and normal development in plants. *EMBO J.* 15,

2988–2996. doi:10.1002/j.1460-2075.1996.tb00662.x.

- Beattie, A. D., Larsen, J., Michaels, T. E., and Pauls, K. P. (2003). Mapping quantitative trait loci for a common bean (*Phaseolus vulgaris* L.) ideotype. *Genome* 46, 411–422. doi:10.1139/g03-015.
- Beaver, J. S., Rosas, J. C., Myers, J., Acosta, J., Kelly, J. D., Nchimbi-Msolla, S., et al. (2003). Contributions of the Bean/Cowpea CRSP to cultivar and germplasm development in common bean. *F. Crop. Res.* 82, 87–102. doi:10.1016/S0378-4290(03)00032-7.
- Beebe, S. (2012). “Common Bean Breeding in the Tropics,” in *Plant Breeding Reviews, Volume 36*, ed. J. Janick (Hoboken, New Jersey: Wiley-Blackwell), 357–426. doi:10.1002/9781118358566.ch2.
- Beebe, S., Ramirez, J., Jarvis, A., Rao, I. M., Mosquera, G., Bueno, J. M., et al. (2011). Genetic Improvement of Common Beans and the Challenges of Climate Change. *Crop Adapt. to Clim. Chang.*, 356–369. doi:10.1002/9780470960929.ch25.
- Beebe, S., Rengifo, J., Gaitan, E., Duque, M., and Tohme, J. (2001). Diversity and Origin of Andean Landraces of Common Bean. *Crop Sci.* 41, 854–862.
- Beebe, S., Skroch, P. W., Tohme, J., Duque, M. C., Pedraza, F., and Nienhuis, J. (2000). Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Sci.* 40, 264–273. doi:10.2135/cropsci2000.401264x.
- Bellon, M. R., Kotu, B. H., Azzarri, C., and Caracciolo, F. (2020). To diversify or not to diversify, that is the question. Pursuing agricultural development for smallholder farmers in marginal areas of Ghana. *World Dev.* 125, 104682. doi:10.1016/j.worlddev.2019.104682.
- Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., et al. (2014). Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell* 26, 1901–1912. doi:10.1105/tpc.114.124040.
- Bernsten, R. H. Bean-growing Environments in Honduras. *Common Bean Atlas Am.* Available at: [https://msu.edu/~bernsten/beanatlas/Country Pages--withGIS/Honduras/2.BeanGrowingEnvironment/hon\\_text\\_bean\\_growingenvironments.htm#:~:text=While beans are grown throughout,\(7%2C074 ha%2C 8%25\).](https://msu.edu/~bernsten/beanatlas/Country Pages--withGIS/Honduras/2.BeanGrowingEnvironment/hon_text_bean_growingenvironments.htm#:~:text=While beans are grown throughout,(7%2C074 ha%2C 8%25).)
- Bethlenfalvay, G. J., and Phillips, D. A. (1977). Ontogenetic Interactions between Photosynthesis and Symbiotic Nitrogen Fixation in Legumes. *Plant Physiol.* 60, 419–421. doi:10.1104/pp.60.3.419.

- Beversdorf, W. (1984). OAC Rico Field Bean. *Can. J. Plant Sci.* 64, 753–755.
- Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., 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. doi:10.1111/j.1469-8137.2012.04377.x.
- Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., et al. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 109. doi:10.1073/pnas.1108973109.
- Bitocchi, E., Rau, D., Bellucci, E., Rodriguez, M., Murgia, M. L., Gioia, T., et al. (2017). Beans (*Phaseolus* spp.) as a model for understanding crop evolution. *Front. Plant Sci.* 8, 1–21. doi:10.3389/fpls.2017.00722.
- Bliss, F. A. (1993). Breeding common bean for improved biological nitrogen fixation. *Plant Soil* 152, 71–79. doi:10.1007/BF00016334.
- Bliss, F. A., Pereira, P. A. A., Araujo, R. S., Henson, R. A., Kmiecik, K. A., McFerson, J. R., et al. (1989). Registration of Five High Nitrogen Fixing Common Bean Germplasm Lines. *Crop Sci.* 29, 240–241. doi:10.2135/cropsci1989.0011183X002900010065x.
- Bodirsky, B. L., Popp, A., Lotze-Campen, H., Dietrich, J. P., Rolinski, S., Biewald, A., et al. (2014). Reactive nitrogen requirements to feed the world in 2050 and potential to mitigate nitrogen pollution. *Nat. Commun.* 5, 1–7. doi:10.1038/ncomms4858.
- Bohlool, B. B., Ladha, J. K., Garrity, D. P., and George, T. (1992). Biological nitrogen fixation for sustainable agriculture: A perspective. *Plant Soil* 141, 1–11. doi:10.1007/bf00011307.
- Boyhan, G. E., and Stone, S. P. (2016). “Breeding for Organic and Sustainable Production,” in *Organic Farming for Sustainable Agriculture. Sustainable Development and Biodiversity, vol 9.*, ed. D. Nandwani (Springer, Cham), 123–136. doi:10.1007/978-3-319-26803-3\_6.
- Boyle, T. (2016). Relative Performance of Ontario Soybean Cultivars and Differential Selection of Early Soybean Breeding Lines in Organic versus Conventional Production Systems.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi:10.1093/bioinformatics/btm308.

- Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., et al. (2014). The Root Hair “Infectome” of *Medicago truncatula* Uncovers Changes in Cell Cycle Genes and Reveals a Requirement for Auxin Signaling in Rhizobial Infection. *Plant Cell* 26, 4680–4701. doi:10.1105/tpc.114.133496.
- Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) - Model food legumes. *Plant Soil* 252, 55–128. doi:10.1023/A:1024146710611.
- Brouwer, B., Winkler, L., Atterberry, K., Jones, S., and Miles, C. (2016). Exploring the role of local heirloom germplasm in expanding western Washington dry bean production. *Agroecol. Sustain. Food Syst.* 40, 319–332. doi:10.1080/21683565.2015.1138013.
- Browning, B. L., and Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* 98, 116–126. doi:10.1016/j.ajhg.2015.11.020.
- Brush, S. B. (2000). *Genes in the field: on-farm conservation of crop diversity.* , ed. S. B. Brush Boca Raton: Lewis Publishers.
- Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., et al. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 17, 1–12. doi:10.1186/s13059-016-0924-1.
- Canada, P. Growing Regions. Available at: <http://www.pulsecanada.com/producers-industry/about-pulses/growing-regions/> [Accessed September 23, 2020].
- Cañas, L. A., and Beltrán, J. P. (2018). “Model Legumes: Functional Genomics Tools in *Medicago truncatula*” in *Functional Genomics in Medicago truncatula. Methods in Molecular Biology, vol 1822*, eds. L. Cañas and J. Beltrán (New York: Humana Press). doi:10.1007/978-1-4939-8633-0\_2.
- Ceccarelli, S. (1994). Specific adaptation and breeding for marginal conditions. *Euphytica* 77, 205–219.
- Ceccarelli, S. (1996). Adaptation to low/high input cultivation. *Euphytica* 92, 203–214.
- Ceccarelli, S., Grando, S., Singh, M., Michael, M., Shikho, A., Al Issa, M., et al. (2000). A methodological study on participatory barley breeding I. Selection phase. *Euphytica* 133, 185–200. doi:10.1023/A:1025535609828.
- Chacon, M. I., Pickersgill, S. B., and Debouck, D. G. (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. doi:10.1007/s00122-

004-1842-2.

- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 1–16. doi:10.1186/s13742-015-0047-8.
- Chaverra, M. H., and Graham, P. H. (1992). Cultivar Variation in Traits Affecting Early Nodulation of Common Bean. *Crop Sci.* 32, 1432–1436. doi:10.2135/cropsci1992.0011183x003200060024x.
- Chikoti, Y. F., Supriadi, Duangkhet, M., Chungopast, S., Tajima, S., Ma, J. F., et al. (2020). Effect of ferritin on nitrogen fixation in *Lotus japonicus* nodules under various iron concentrations. *J. Plant Physiol.* 252, 153247–153247. doi:10.1016/j.jplph.2020.153247.
- Ciampitti, I. A., and Salvagiotti, F. (2018). New Insights into Soybean Biological Nitrogen Fixation. *Agron. J.* 110, 1185–1196. doi:10.2134/agronj2017.06.0348.
- CIAT (2019). Crop conservation and use. Available at: <https://ciat.cgiar.org/what-we-do/crop-conservation-and-use/>.
- Cichy, K. A., Porch, T. G., Beaver, J. S., Cregan, P., Fourie, D., Glahn, R. P., et al. (2015). A *Phaseolus vulgaris* Diversity Panel for Andean Bean Improvement. *Crop Sci.* 55, 2149–2160. doi:10.2135/cropsci2014.09.0653.
- Coyne, D. P., and Schuster, M. L. (1974). Breeding and genetic studies of tolerance to several bean (*Phaseolus vulgaris* L.) bacterial pathogens. *Euphytica* 23, 651–656. doi:10.1007/BF00022486.
- Dall’Agnol, R. F., Ribeiro, R. A., Ormeno-Orrillo, E., Rogel, M. A., Delamuta, J. R. M., Andrade, D. S., et al. (2013). *Rhizobium freirei* sp. nov., a symbiont of *Phaseolus vulgaris* that is very effective at fixing nitrogen. *Int. J. Syst. Evol. Microbiol.* 63, 4167–4173. doi:10.1099/ijs.0.052928-0.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330.
- De Ron, A. M., González, A. M., Godoy, L., Santalla, M., Godoy, L., and Papa, R. (2016). History of the common bean crop: Its evolution beyond its areas of origin and domestication. *Arbor Cienc.* 192, a317. doi:10.3989/arbor.2016.779n3007.
- Diaz, L. M., Ricaurte, J., Cajiao, C., Galeano, C. H., Rao, I., Beebe, S., et al. (2017). Phenotypic evaluation and QTL analysis of yield and symbiotic nitrogen fixation in a common bean population grown with two levels of phosphorus supply. *Mol. Breed.*

37. doi:10.1007/s11032-017-0673-1.

- Drought-Resistant Bean Variety Offers Hope to Central American Communities (2016). 1–2. Available at: [https://www.canr.msu.edu/legumelab/uploads/files/Success\\_Drought\\_resistant\\_beans\\_Central\\_America.pdf](https://www.canr.msu.edu/legumelab/uploads/files/Success_Drought_resistant_beans_Central_America.pdf) [Accessed September 9, 2020].
- Duc, G., Agrama, H., Bao, S., Berger, J., Bourion, V., De Ron, A. M., et al. (2015). Breeding Annual Grain Legumes for Sustainable Agriculture: New Methods to Approach Complex Traits and Target New Cultivar Ideotypes. *CRC. Crit. Rev. Plant Sci.* 34, 381–411. doi:10.1080/07352689.2014.898469.
- Duranti, M. (2006). Grain legume proteins and nutraceutical properties. *Fitoterapia* 77, 67–82. doi:10.1016/j.fitote.2005.11.008.
- Duwadi, K., Austin, R. S., Mainali, H. R., Bett, K., Marsolais, F., and Dhaubhadel, S. (2018). Slow darkening of pinto bean seed coat is associated with significant metabolite and transcript differences related to proanthocyanidin biosynthesis. *BMC Genomics* 19, 1–13. doi:10.1186/s12864-018-4550-z.
- Dwivedi, S. L., Ceccarelli, S., Blair, M. W., Upadhyaya, H. D., Are, A. K., and Ortiz, R. (2016). Landrace Germplasm for Improving Yield and Abiotic Stress Adaptation. *Trends Plant Sci.* 21, 31–42. doi:10.1016/j.tplants.2015.10.012.
- Earl, D. A., and VanHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi:10.1007/s12686-011-9548-7.
- Ebel, R., Aguilar, M. de J. M., and Putnam, H. R. (2017). Milpa: One Sister Got Climate-sick. The Impact of Climate Change on Traditional Maya Farming Systems. *Int. J. Sociol. Agric. Food* 24, 175–199.
- Ersoz, E. S., Yu, J., and Buckler, E. S. (2007). “Applications of linkage disequilibrium and association mapping in crop plants” in *Genomics-assisted crop improvement, volume 1, genomics approaches and platforms*, eds. R. K. Varshney and R. Tuberosa (Dordrecht: Springer), 97–119.
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi:10.1111/j.1365-294x.2005.02553.x.
- FAO FAOSTAT database, dry bean production statistics, 2018. *FAOSTAT*. Available at: <http://www.fao.org/faostat/en/#data>.
- FAO (2004). *The State of Food and Agriculture 2003–2004*. Food and Agriculture

Organization of the United Nations, Rome.

- FAO, F. and A. O. of the U. N. (2018). FAOSTAT, Dry Bean production values, 2018. Available at: <http://www.fao.org/faostat/en/#data> [Accessed September 23, 2020].
- FAO, F. and A. O. of the U. N. (2019). FAO's work on climate change; United Nations climate change conference 2019. Available at: <http://www.fao.org/3/ca7126en/ca7126en.pdf>.
- Farid, M. (2015). Symbiotic Nitrogen Fixation in Common Bean. PhD dissertation, University of Guelph.
- Farid, M., Earl, H. J., Pauls, K. P., and Navabi, A. (2017). Response to selection for improved nitrogen fixation in common bean (*Phaseolus vulgaris* L.). *Euphytica* 213, 1–13. doi:10.1007/s10681-017-1885-5.
- Farid, M., and Navabi, A. (2015). N<sub>2</sub> fixation ability of different dry bean genotypes. *Can. J. Plant Sci.* 95, 1243–1257. doi:10.4141/CJPS-2015-084.
- Farquhar, G. D., Ehleringer, J. R., and Hubick, K. T. (1989). Carbon Isotope Discrimination and Photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 503–337.
- Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution (N. Y.)* 39, 783. doi:10.2307/2408678.
- Flores, A. C., Via, V. D., Savy, V., Villagra, U. M., Zanetti, M. E., and Blanco, F. (2018). Comparative phylogenetic and expression analysis of small GTPases families in legume and non-legume plants. *Plant Signal. Behav.* 13. doi:10.1080/15592324.2018.1432956.
- Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Reis, S., et al. (2013). The global nitrogen cycle in the twenty-first century. *Philos. Trans. R. Soc. B* 368. doi:10.1098/rstb.2013.0164.
- Freyre, R., Skroch, P. W., Geffroy, V., Adam-Blondon, A.-F., Shirmohamadali, A., Johnson, W. C., et al. (1998). Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. *Theor. Appl. Genet.* 97, 847–856. doi:10.1007/s001220050964.
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720. doi:10.1093/bioinformatics/btv428.
- Galvez, L., Gonzalez, E. M., and Arrese-Igor, C. (2005). Evidence for carbon flux

shortage and strong carbon/nitrogen interactions in pea nodules at early stages of water stress. *J. Exp. Bot.* 56, 2551–2561. doi:10.1093/jxb/eri249.

- Gepts, P. (2006). Plant Genetic Resources Conservation and Utilization: The Accomplishments and Future of a Societal Insurance Policy. *Crop Adv. F. Crop Reports* 46, 2278–2292. doi:10.2135/cropsci2006.03.0169gas.
- Gepts, P., Osborn, T. C., Rashka, K., and Bliss, F. A. (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.
- Gepts, P., Papa, R., Coulibaly, S., Mejia, A. G., and Pasquet, R. (1999). Wild legume diversity and domestication - insights from molecular methods. in *The Seventh Ministry of Agriculture, Forestry and Fisheries (MAFF), Japan, International Workshop on Genetic Resources*, eds. K. Oono, D. Vaughan, N. Tomooka, A. Kaga, and S. Miyazaki (Ibaraki, Japan: National Institute of Agrobiological Resources), 19–31.
- Gioia, T., Logo, O., Logozzo, G., Marzario, S., Zeuli, P. S., and Gepts, P. (2019). Evolution of SSR diversity from wild types to U.S. advanced cultivars in the Andean and Mesoamerican domestications of common bean (*Phaseolus vulgaris*). *PLoS One* 14, p.e0211342. doi:10.1371/journal.pone.0211342.
- Gioia, T., Logozzo, G., Attene, G., Bellucci, E., Benedettelli, S., Papa, R., et al. (2013). Evidence for Introduction Bottleneck and Extensive Inter- Gene Pool (Mesoamerica x Andes) Hybridization in the European Common Bean (*Phaseolus vulgaris* L.) Germplasm. *PLoS One* 8, 1–14. doi:10.1371/journal.pone.0075974.
- Gomez, M., Rosas, J. C., Humphries, S., Jiménez, J., Barahona, M., Avila, C., et al. (2020). “Pushing back against bureaucracy: Farmers’ role in decentralizing plant breeding and seed production in Honduras” in *Farmers and Plant Breeding*, eds. O. T. Westengen and T. Winge (London: Routledge). doi:10.4324/9780429507335.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, 1178–1186. doi:10.1093/nar/gkr944.
- Graham, P. H. (1981). Some problems of nodulation and symbiotic nitrogen fixation in *Phaseolus vulgaris* L.: A Review. *F. Crop. Res.* 4, 93–112. doi:https://doi.org/10.1016/0378-4290(81)90060-5.
- Graham, P. H., and Rosas, J. C. (1977). Growth and Development of Indeterminate Bush and Climbing Cultivars of *Phaseolus vulgaris* L. Inoculated with Rhizobium. *J. Agric. Sci.* 88, 503–508. doi:10.1017/S0021859600035024.

- Graham, P. H., and Rosas, J. C. (1978a). Nodule development and nitrogen fixation in cultivars of *Phaseolus vulgaris* L. as influenced by planting density. *J. Agric. Sci.* 90, 19–29. doi:10.1017/S0021859600048553.
- Graham, P. H., and Rosas, J. C. (1978b). Plant and nodule development and nitrogen fixation in climbing cultivars of *Phaseolus vulgaris* L. grown in monoculture, or associated with *Zea mays* L. *J. Agric. Sci.* 90, 311–317. doi:10.1017/S0021859600055398.
- Graham, P. H., and Rosas, J. C. (1979). Phosphorus Fertilization and Symbiotic Nitrogen Fixation in Common Bean. *Agron. J.* 71, 925–926. doi:10.2134/agronj1979.00021962007100060007x.
- Grobbelaar, N., Clarke, B., and Hough, M. (1971). The nodulation and nitrogen fixation of isolated roots of *Phaseolus vulgaris* L. III. The effect of carbon dioxide and ethylene. *Plant Soil* 35, 215–223. doi:10.1007/BF02661852.
- Growers, O. B. Ontario Beans - So Many to Choose From! Available at: <http://ontariobeans.on.ca/types-of-beans/> [Accessed September 23, 2020].
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi:10.1093/bioinformatics/btu393.
- Gucciardo, S., Wisniewski, J.-P., Brewin, N. J., and Bornemann, S. (2007). A germin-like protein with superoxide dismutase activity in pea nodules with high protein sequence identity to a putative rhicadhesin receptor. *J. Exp. Bot.* 58, 1161–1171. doi:10.1093/jxb/erl282.
- Guillén, G., Díaz-camino, C., Loyola-torres, C. A., Aparicio-fabre, R., Hernández-lópez, A., Díaz-sánchez, M., et al. (2013). Detailed analysis of putative genes encoding small proteins in legume genomes. 4, 1–12. doi:10.3389/fpls.2013.00208.
- Gurusamy, V., Vandenberg, A., and Bett, K. E. (2007). Manipulation of in vivo pollination techniques to improve the fertilization efficiency of interspecies crosses in the genus *Phaseolus*. *Plant Breed.* 126, 120–124. doi:10.1111/j.1439-0523.2007.01337.x.
- Guttikonda, S. K., Trupti, J., Bisht, N. C., Chen, H., An, Y.-Q. C., Pandey, S., et al. (2010). Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation- specific P450 monooxygenases. *BMC Plant Biol.* 10, 243. doi:10.1186/1471-2229-10-243.
- Hardarson, G., Bliss, F. A., Cigales-Rivero, M. R., Henson, R. A., Kipe-Nolt, J. A., Longeri, L., et al. (1993). Genotypic variation in biological nitrogen fixation by

- common bean. *Plant Soil* 152, 59–70. doi:<https://doi.org/10.1007/BF00016333>.
- Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., et al. (2007). Grinding up wheat: A massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* 24, 1506–1517. doi:[10.1093/molbev/msm077](https://doi.org/10.1093/molbev/msm077).
- Hefni, M., Öhrvik, V., Tabekha, M., and Witthöft, C. (2010). Folate content in foods commonly consumed in Egypt. *Food Chem.* 121, 540–545. doi:[10.1016/j.foodchem.2009.12.044](https://doi.org/10.1016/j.foodchem.2009.12.044).
- Heilig, J. A. (2015). QTL mapping of symbiotic nitrogen fixation in dry bean; dry bean performance under organic production systems.
- Heilig, J. A., Beaver, J. S., Wright, E. M., Song, Q., and Kelly, J. D. (2017a). QTL analysis of symbiotic nitrogen fixation in a black bean population. *Crop Sci.* 57, 118–129. doi:[10.2135/cropsci2016.05.0348](https://doi.org/10.2135/cropsci2016.05.0348).
- Heilig, J. A., Wright, E. M., and Kelly, J. D. (2017b). Symbiotic nitrogen fixation of black and navy bean under organic production systems. *Agron. J.* 109, 2223–2230. doi:[10.2134/agronj2017.01.0051](https://doi.org/10.2134/agronj2017.01.0051).
- Henry, R. J. (2019). Innovations in plant genetics adapting agriculture to climate change. *Curr. Opin. Plant Biol.* doi:[10.1016/j.pbi.2019.11.004](https://doi.org/10.1016/j.pbi.2019.11.004).
- Henry, R. J., and Nevo, E. (2014). Exploring natural selection to guide breeding for agriculture. *Plant Biotechnol. J.* 12, 655–662. doi:[10.1111/pbi.12215](https://doi.org/10.1111/pbi.12215).
- Herridge, D. F., Peoples, M. B., and Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 311, 1–18. doi:[10.1007/s11104-008-9668-3](https://doi.org/10.1007/s11104-008-9668-3).
- Hosfield, G. L., Varner, G. V., Uebersax, M. A., and Kelly, J. D. (2004). Registration of ‘Merlot’ Small Red Bean. *Crop Sci.* 44, 351–352. doi:[10.2135/cropsci2004.3510](https://doi.org/10.2135/cropsci2004.3510).
- Hoyos-Villegas, V., Wright, E. M., and Kelly, J. D. (2016). GGE biplot analysis of yield associations with root traits in a mesoamerican bean diversity panel. *Crop Sci.* 56, 1081–1094. doi:[10.2135/cropsci2015.10.0609](https://doi.org/10.2135/cropsci2015.10.0609).
- Huang, X., and Han, B. (2014). Natural Variations and Genome-Wide Association Studies in Crop Plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi:[10.1146/annurev-arplant-050213-035715](https://doi.org/10.1146/annurev-arplant-050213-035715).
- Humphries, S., Gallardo, O., Jiménez, J., and Sierra, F. (2005). Linking small farmers to the formal research sector: Lessons from a participatory bean breeding programme in Honduras. *Agric. Res. Ext. Netw.*, 20.

- Humphries, S., Rosas, J. C., Gómez, M., Jiménez, J., Sierra, F., Gallardo, O., et al. (2015). Synergies at the interface of farmer-scientist partnerships: Agricultural innovation through participatory research and plant breeding in Honduras. *Agric. Food Secur.* 4, 1–17. doi:10.1186/s40066-015-0046-0.
- Hurtado-Gonzales, O. P., Gilio, T. A. S., and Pastor-Corrales, M. A. (2017). Resistant Reaction of Andean Common Bean Landrace G19833, Reference Genome, to 13 Races of *Uromyces appendiculatus* Suggests Broad Spectrum Rust Resistance. in *Annual report of the Bean Improvement Cooperative, No. 60*, 3–5.
- Jamora, N., and Maredia, M. (2010). Database of improved beans and cowpea varietal releases in major bean and cowpea producing countries in Africa, Latin America and the U.S., 1980-2010. Available at: [https://www.canr.msu.edu/legumelab/uploads/files/CRSP\\_bean\\_varietal\\_outputs\\_1980-2010.pdf](https://www.canr.msu.edu/legumelab/uploads/files/CRSP_bean_varietal_outputs_1980-2010.pdf).
- Jiang, Y., MacLean, D. E., Perry, G. E., Marsolais, F., Hill, B., and Pauls, K. P. (2020). Evaluation of beneficial and inhibitory effects of nitrate on nodulation and nitrogen fixation in common bean (*Phaseolus vulgaris*). *Legum. Sci.* e45. doi:10.1002/leg3.45.
- Jochua, C. N., Strock, C. F., and Lynch, J. P. (2020). Root phenotypic diversity in common bean reveals contrasting strategies for soil resource acquisition among gene pools and races. *Crop Sci.* 60, 3261–3277. doi:10.1002/csc2.20312.
- Johnson, E., Miklas, P. N., Stavely, J. R., and Martinez-Cruzado, J. C. (1995). Coupling- and repulsion-phase RAPDs for marker-assisted selection of PI 181996 rust resistance in common bean. *Theor. Appl. Genet.* 90, 659–664. doi:10.1007/BF00222130.
- Johnson, W. C., and Gepts, P. (1999). Segregation for performance in recombinant inbred populations resulting from inter-gene pool crosses of common bean (*Phaseolus vulgaris* L.). *Euphytica* 106, 45–56.
- Jones, A. L. (1999). *Phaseolus* bean Post-harvest Operations. INPhO Post-harvest Compendium. Available at: <http://www.fao.org/3/a-av015e.pdf>.
- Kamfwa, K., Cichy, K. A., and Kelly, J. D. (2015). Genome-wide association analysis of symbiotic nitrogen fixation in common bean. *Theor. Appl. Genet.* 128, 1999–2017. doi:10.1007/s00122-015-2562-5.
- Kamfwa, K., Cichy, K. A., and Kelly, J. D. (2019). Identification of quantitative trait loci for symbiotic nitrogen fixation in common bean. *Theor. Appl. Genet.* 132, 1375–1387. doi:10.1007/s00122-019-03284-6.

- Kamfwa, K., Zhao, D., Kelly, J. D., and Cichy, K. A. (2017). Transcriptome analysis of two recombinant inbred lines of common bean contrasting for symbiotic nitrogen fixation. *PLoS One* 12, 1–19. doi:10.1371/journal.pone.0172141.
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics* 178, 1709–1723. doi:10.1534/genetics.107.080101.
- Kaplan, L., and Lynch, T. F. (1999). Phaseolus (Fabaceae) in Archaeology : AMS Radiocarbon Dates and Their Significance for Pre-Colombian Agriculture. 53, 261–272.
- Kaur, S., Singh, S., and Tandon, G. (2017). Effect of Mesorhizobium in chickpea root nodulation using transcriptome profiling approach. *Indian J. Agric. Sci.* 87, 1507–1514.
- Kelly, J. D. (2001). Remaking bean plant architecture for efficient production. *Adv. Agron.* 71, 109–143. doi:10.1016/s0065-2113(01)71013-9.
- Kelly, J. D. (2010). The Story of Bean Breeding: White Paper Prepared for BeanCAP and PBG Works on the topic of dry bean production and breeding research in the U.S. 30. Available at:  
[https://www.canr.msu.edu/beanbreeding/\\_pdf/Story\\_of\\_Bean\\_Breeding\\_in\\_the\\_US.pdf](https://www.canr.msu.edu/beanbreeding/_pdf/Story_of_Bean_Breeding_in_the_US.pdf).
- Kelly, J. D., Varner, G. V., O’Boyle, P., and Long, B. (2009). Registration of ‘Zorro’ Black Bean. *J. Plant Regist.* 3, 226–230. doi:10.3198/jpr2008.12.0730crc.
- Khanal, R., Rupert, T., Navabi, A., Smith, T. H., Michaels, T. E., Burt, A. J., et al. (2016). Yeti kidney bean. *Can. J. Plant Sci.* 96, 280–282. doi:10.1139/cjps-2015-0181.
- Khanal, R., Rupert, T., Navabi, A., Smith, T. H., Michaels, T. E., Burt, A. J., et al. (2017a). Cultivar Description: Mist Common Bean. *Can. J. Plant Sci.* 97, 192–195.
- Khanal, R., Smith, T. H., Michaels, T. E., and Pauls, K. P. (2017b). OAC Spark common bean. *Can. J. Plant Sci.* 97, 169–171. doi:10.1139/cjps-2016-0045.
- Kipe-Nolt, J. A., Vargas, H., and Giller, K. E. (1993). Nitrogen fixation in breeding lines of *Phaseolus vulgaris* L. *Plant Soil* 152, 103–106. doi:10.1007/BF00016338.
- Koenig, R., and Gepts, P. (1989). Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. *Theor. Appl. Genet.* 78, 809–817. doi:10.1007/BF00266663.
- Kohlen, W., Ng, J. L. P., Deinum, E. E., and Mathesius, U. (2018). Auxin transport,

- metabolism, and signalling during nodule initiation: indeterminate and determinate nodules. *J. Exp. Bot.* 69, 229–244. doi:10.1093/jxb/erx308.
- Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 1–9. doi:10.1186/1746-4811-9-29.
- Kull, C. A., Carrière, S. M., Moreau, S., Ramiarantsoa, H. R., Blanc-Pamard, C., and Tassin, J. (2013). Melting Pots of Biodiversity: Tropical Smallholder Farm Landscapes as Guarantors of Sustainability. *Environ. Sci. Policy Sustain. Dev.* 55, 6–16. doi:10.1080/00139157.2013.765307.
- Kumar, S., Garrick, D. J., Bink, M. C. A. M., Whitworth, C., Chagné, D., and Volz, R. K. (2013). Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* 14, 393. doi:10.1186/1471-2164-14-393.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054.
- Kutoš, T., Golob, T., Kač, M., and Plestenjak, A. (2003). Dietary fibre content of dry and processed beans. *Food Chem.* 80, 231–235. doi:10.1016/S0308-8146(02)00258-3.
- Kwak, M., and 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. doi:10.1007/s00122-008-0955-4.
- Lammerts van Bueren, E. T., Jones, S. S., Tamm, L., Murphy, K. M., Myers, J. R., Leifert, C., et al. (2011). The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. *NJAS - Wageningen J. Life Sci.* 58, 193–205. doi:10.1016/j.njas.2010.04.001.
- Larue, T. A., and Patterson, T. G. (1981). “How much Nitrogen do Legumes fix?,” in *Advances in Agronomy*, ed. N. C. Brady (New York: Academic Press, Inc.), 15–38.
- Lee, K. H., and LaRue, T. A. (1992). Exogenous Ethylene Inhibits Nodulation of *Pisum sativum* L. cv Sparkle. *Plant Physiol.* 100, 1759–1763. doi:10.1104/pp.100.4.1759.
- Lei, L., Chen, L., Shi, X., Li, Y., Wang, J., Chen, D., et al. (2014). A Nodule-Specific Lipid Transfer Protein AsE246 Participates in Transport of Plant-Synthesized Lipids to Symbiosome Membrane and Is Essential for Nodule Organogenesis in Chinese Milk Vetch. *Plant Physiol.* 164, 1045–1058. doi:10.1104/pp.113.232637.

- Li, H., Glusman, G., Huff, C., Caballero, J., and Roach, J. C. (2014). Accurate and Robust Prediction of Genetic Relationship from Whole-Genome Sequences. *PLoS One* 9, 1–6. doi:10.1371/journal.pone.0085437.
- Li, R., Jiang, X., Jin, D., Dhaubhadel, S., Bian, S., and Li, X. (2015). Identification of 14-3-3 Family in Common Bean and Their Response to Abiotic Stress. *PLoS One* 10, e0143280. doi:10.1371/journal.pone.0143280.
- Lin, Y., Zhang, J., Gao, W., Chen, Y., Li, H., Lawlor, D. W., et al. (2017). Exogenous trehalose improves growth under limiting nitrogen through upregulation of nitrogen metabolism. *BMC Plant Biol.* 17, 247. doi:10.1186/s12870-017-1207-z.
- Lindstrom, K., and Mousavi, S. A. (2020). Effectiveness of nitrogen fixation in rhizobia. *Microb. Biotechnol.* 13, 1314–1335. doi:10.1111/1751-7915.13517.
- Lioi, L., Zuluaga, D. L., Pavan, S., and Sonnante, G. (2019). Genotyping-by-sequencing reveals molecular genetic diversity in Italian common bean landraces. *Diversity* 11, 1–14. doi:10.3390/d11090154.
- Liu, P., Sun, F., Gao, R., and Dong, H. (2012). RAP2.6L overexpression delays waterlogging induced premature senescence by increasing stomatal closure more than antioxidant enzyme activity. *Plant Mol. Biol.* 79, 609–622. doi:10.1007/s11103-012-9936-8.
- Liu, W., Han, X., Zhan, G., Zhao, Z., Feng, Y., and Wu, C. (2015). A Novel Sucrose-Regulatory MADS-Box Transcription Factor GmNMHC5 Promotes Root Development and Nodulation in Soybean (*Glycine max* [L.] Merr.). *Int. J. Mol. Sci.* 16, 20657–20673. doi:10.3390/ijms160920657.
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome- Wide Association Studies. *PLoS One* 12, 1–24. doi:10.1186/1471-2156-13-100.
- Lopes, M. S., El-Basyoni, I., Baenziger, P. S., Singh, S., Royo, C., Ozbek, K., et al. (2015). Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* 66, 3477–3486. doi:10.1093/jxb/erv122.
- Luyten, E., Swinnen, E., Vlassak, K., Verreth, C., Dombrecht, B., and Vanderleyden, J. (2001). Analysis of a Symbiosis-Specific Cytochrome P450 Homolog in *Rhizobium* sp. BR816. *Mol. Plant-Microbe Interact.* 14, 918–924. doi:10.1094/MPMI.2001.14.7.918.
- Ma, Y., Liu, M., Stiller, J., and Liu, C. (2019). A pan-transcriptome analysis shows that disease resistance genes have undergone more selection pressure during barley domestication 06 Biological Sciences 0604 Genetics. *BMC Genomics* 20, 1–11.

doi:10.1186/s12864-018-5357-7.

- MacQueen, A. H., White, J. W., Lee, R., Osorno, J. M., Schmutz, J., Miklas, P. N., et al. (2020). Genetic Associations in Four Decades of Multi-Environment Trials Reveal Agronomic Trait Evolution in Common Bean. *Genetics* 215, 267–284.
- Mamidi, S., Rossi, M., Annam, D., Moghaddam, S., Lee, R., Papa, R., et al. (2011). Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38, 953–967. doi:10.1071/FP11124.
- Mamidi, S., Rossi, M., Moghaddam, S. M., Annam, D., Lee, R., Papa, R., et al. (2013). Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity (Edinb)*. 110, 267–276. doi:10.1038/hdy.2012.82.
- Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). Markers based on single nucleotide polymorphisms (SNPs) have rapidly gained the center stage of molecular genetics during the recent years due to their abundance in the genomes and their amenability for high-throughput detection formats and platforms. *Int. J. Plant Genomics*, 728398. doi:10.1155/2012/728398.
- Maradiaga, O. J. E., Urbina Algabas, R., Raudales, N., and Galeas, O. (2016). Validation of Improved Bean Varieties (*Phaseolus vulgaris* L.) in 28 sites of the departments of Choluteca and South of E Paraiso.
- Maron, L. G., Guimarães, C. T., Kirst, M., Albert, P. S., Birchler, J. A., Bradbury, P. J., et al. (2013). Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proc. Natl. Acad. Sci. U. S. A.* 110, 5241–5246. doi:10.1073/pnas.1220766110.
- Marx, H., Minogue, C. E., Jayaraman, D., Richards, A. L., Kwiecien, N. W., Siahpirani, A. F., et al. (2016). A proteomic atlas of the legume *Medicago truncatula* and its nitrogen-fixing endosymbiont *Sinorhizobium meliloti*. *Nat. Biotechnol.* 34. doi:10.1038/nbt.3681.
- Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., and Suzuki, A. (2010). Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. *Ann. Bot.* 105, 1141–1157. doi:10.1093/aob/mcq028.
- Mather, D. L., Bernsten, R., Rosas, J. C., Viana Ruano, A., and Escoto, D. (2003). The economic impact of bean disease resistance research in Honduras. *Agric. Econ.* 29, 343–352. doi:10.1016/S0169-5150(03)00061-6.
- McAuliffe, C., Chamblee, D. S., Uribe-Arango, H., and Woodhouse, W. W. J. (1958). Influence of Inorganic Nitrogen on Nitrogen Fixation by Legumes as Revealed by

- N15. *Agron. J.* 50, 334–337. doi:10.2134/agronj1958.00021962005000060014x.
- McClellan, P. E. (2008). Common Bean Coordinated Agricultural Project: Project Summary. Fargo Available at: [http://arsftfbean.uprm.edu/beancap/wp-content/uploads/2019/03/BeanCAP\\_Project\\_Description.pdf](http://arsftfbean.uprm.edu/beancap/wp-content/uploads/2019/03/BeanCAP_Project_Description.pdf).
- McClellan, P. E., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., et al. (2018). White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytol.* 219, 1112–1123. doi:10.1111/nph.15259.
- McClellan, P. E., Lee, R. K., and Miklas, P. N. (2004). Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. *Genome* 47, 266–280. doi:10.1139/G03-103.
- McClellan, P. E., Mamidi, S., McConnell, M., Chikara, S., and Lee, R. (2010). Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics* 11, 184. doi:10.1186/1471-2164-11-184.
- McClellan, P. E., Moghaddam, S. M., López-Millán, A. F., Brick, M. A., Kelly, J. D., Miklas, P. N., et al. (2017). Phenotypic diversity for seed mineral concentration in North American dry bean germplasm of middle american ancestry. *Crop Sci.* 57, 3129–3144. doi:10.2135/cropsci2017.04.0244.
- McGregor, W. G., MacLean, A. J., and Wallen, V. R. (1956). Field Beans in Canada. *pub.* 843, 1–16. Available at: [http://publications.gc.ca/collections/collection\\_2013/aac-aafc/agrhist/A53-843-1956-eng.pdf](http://publications.gc.ca/collections/collection_2013/aac-aafc/agrhist/A53-843-1956-eng.pdf).
- Mejía-Jiménez, A., Muñoz, C., Jacobsen, H. J., Roca, W. M., and Singh, S. P. (1994). Interspecific hybridization between common and tepary beans: increased hybrid embryo growth, fertility, and efficiency of hybridization through recurrent and congruity backcrossing. *Theor. Appl. Genet.* 88, 324–331. doi:10.1007/bf00223640.
- Michaels, T. E., Smith, T. H., Larsen, J., Beattie, A. D., and Pauls, K. P. (2006). OAC Rex common bean. *Can. J. Plant Sci.* 86, 733–736. doi:10.4141/P05-128.
- Michiels, J., Dombrecht, B., Vermeiren, N., Xi, C., Luyten, E., and Vanderleyden, J. (1998). *Phaseolus vulgaris* is a non-selective host for nodulation. *FEMS Microbiol. Ecol.* 26, 193–205. doi:10.1111/j.1574-6941.1998.tb00505.x.
- Miklas, P. N., Kelly, J. D., Beebe, S. E., and Blair, M. W. (2006). Common bean breeding for resistance against biotic and abiotic stresses: From classical to MAS breeding. *Euphytica* 147, 105–131. doi:10.1007/s10681-006-4600-5.

- Miklas, P. N., Zapata, M., Beaver, J. S., and Grafton, K. F. (1999). Registration of Four Dry Bean Germplasms Resistant to Common Bacterial Blight: ICB-3, ICB-6, ICB-8, and ICB-10. *Crop Sci.* 39, 594–595. doi:10.2135/cropsci1999.0011183X003900020065x.
- Moghaddam, A., Raza, A., Vollmann, J., Ardakani, M. R., Wanek, W., Gollner, G., et al. (2013). Carbon isotope discrimination and water use efficiency relationships of alfalfa genotypes under irrigated and rain-fed organic farming. *Eur. J. Agron.* 50, 82–89. doi:10.1016/j.eja.2013.05.010.
- Moghaddam, S. M., Brick, M. A., Echeverria, D., Thompson, H. J., Brick, L. A., Lee, R., et al. (2018). Genetic Architecture of Dietary Fiber and Oligosaccharide Content in a Middle American Panel of Edible Dry Bean. *Plant Genome* 11, 1–11. doi:10.3835/plantgenome2017.08.0074.
- Moghaddam, S. M., Mamidi, S., Osorno, J. M., Lee, R., Brick, M., Kelly, J., et al. (2016). Genome-Wide Association Study Identifies Candidate Loci Underlying Agronomic Traits in a Middle American Diversity Panel of Common Bean. *Plant Genome* 9, 1–21. doi:10.3835/plantgenome2016.02.0012.
- Monk, J. M., Lepp, D., Zhang, C. P., Wu, W., Zarepoor, L., Lu, J. T., et al. (2016). Diets enriched with cranberry beans alter the microbiota and mitigate colitis severity and associated inflammation. *J. Nutr. Biochem.* 28, 129–139. doi:10.1016/j.jnutbio.2015.10.014.
- Muldoon, J. F., Hume, D. J., and Beversdorf, W. D. (1980). Effects of seed- and soil-applied *Rhizobium japonicum* inoculants on soybeans in Ontario. *Can. J. Plant Sci.* 60, 399–409. doi:10.4141/cjps80-061.
- Mullner, D. (2013). fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *J. Stat. Softw.* 53, 1–18. doi:10.18637/jss.v053.i09.
- Munns, D. N., Fox, R. L., and Koch, B. L. (1977). Influence of lime on nitrogen fixation by tropical and temperate legumes. *Plant Soil* 46, 591–601. doi:10.1007/BF00015918.
- Muñoz-Perea, C. G., Allen, R. G., Westermann, D. T., Wright, J. L., and Singh, S. P. (2007). Water use efficiency among dry bean landraces and cultivars in drought-stressed and non-stressed environments. *Euphytica* 155, 393–402. doi:10.1007/s10681-006-9340-z.
- Muñoz-Perea, C. G., Terán, H., Allen, R. G., Wright, J. L., Westermann, D. T., and Singh, S. P. (2006). Selection for drought resistance in dry bean landraces and cultivars. *Crop Sci.* 46, 2111–2120. doi:10.2135/cropsci2006.01.0029.

- Nanni, L., Bitocchi, E., Bellucci, E., Rossi, M., Rau, D., Attene, G., et al. (2011). Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 123, 1341–1357. doi:10.1007/s00122-011-1671-z.
- Navabi, A., Balasubramanian, P., Pauls, K. P., Bett, K., and Hou, A. (2014). Genetic Diversity of the Canadian Dry Bean Varieties Released Since 1930: A Pedigree Analysis. *Crop Sci.* 54, 993–1003. doi:10.2135/cropsci2013.04.0210.
- Nei, M., and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 76, 5269–5273. doi:10.1073/pnas.76.10.5269.
- Nova-Franco, B., Íñiguez, L. P., Valdés-López, O., Alvarado-Affantranger, X., Leija, A., Fuentes, S. I., et al. (2015). The micro-RNA72c-APETALA2-1 node as a key regulator of the common bean-Rhizobium etli nitrogen fixation symbiosis. *Plant Physiol.* 168, 273–291. doi:10.1104/pp.114.255547.
- Oblessuc, P. R., Matioli, C. C., Chiorato, A. F., Camargo, L. E. A., Benchimol-Reis, L. L., and Melotto, M. (2015). Common bean reaction to angular leaf spot comprises transcriptional modulation of genes in the ALS10.1 QTL. *Front. Plant Sci.* 6, 1–11. doi:10.3389/fpls.2015.00152.
- Oikawa, K., Yamasato, A., Kong, S. G., Kasahara, M., Nakai, M., Takahashi, F., et al. (2008). Chloroplast outer envelope protein CHUP1 is essential for chloroplast anchorage to the plasma membrane and chloroplast movement. *Plant Physiol.* 148, 829–842. doi:10.1104/pp.108.123075.
- Oladzad, A., González, A., Macchiavelli, R., Estevez de Jensen, C., Beaver, J., Porch, T., et al. (2020). Genetic Factors Associated With Nodulation and Nitrogen Derived From Atmosphere in a Middle American Common Bean Panel. *Front. Plant Sci.* 11, 576078. doi:10.3389/fpls.2020.576078.
- Oladzad, A., Porch, T., Rosas, J. C., Moghaddam, S. M., Beaver, J., Beebe, S. E., et al. (2019). Single and multi-trait GWAS identify genetic factors associated with production traits in common bean under abiotic stress environments. *G3 Genes, Genomes, Genet.* 9, 1881–1892. doi:10.1534/g3.119.400072.
- Oldroyd, G. E. D., and Downie, J. A. (2004). Calcium, kinases and nodulation signalling in legumes. *Nat. Rev. Mol. Cell Biol.* 5, 566–576. doi:10.1038/nrm1424.
- OMAFRA (2016). Historical Provincial Estimates by Crop, 1981-2014 (Metric Units). Available at: [http://www.omafra.gov.on.ca/english/stats/crops/estimate\\_hist\\_metric.htm](http://www.omafra.gov.on.ca/english/stats/crops/estimate_hist_metric.htm) [Accessed March 10, 2019].

- OMAFRA, O. M. of F. A. and R. A. (2009). Nitrogen Rates for Edible Beans. *Crop Adv. F. Crop Reports*, 1–3. Available at: <http://www.ontariosoilcrop.org/wp-content/uploads/2015/07/v6bea1.pdf> [Accessed September 22, 2020].
- Otieno, P. E., Muthomi, J. W., Chemining'wa, G. N., and Nderitu, J. H. (2009). Effect of Rhizobia inoculation, farm yard manure and Nitrogen fertilizer on nodulation and yield of food grain legumes. *J. Biol. Sci.* 9, 326–332. doi:10.3923/jbs.2009.326.332.
- Pacovsky, R. S., Bayne, H. G., and Bethlenfalvay, G. J. (1984). Symbiotic Interactions between Strains of *Rhizobium phaseoli* and Cultivars of *Phaseolus vulgaris* L. *Crop Sci.* 24, 101–105. doi:10.2135/cropsci1984.0011183X002400010023x.
- Padder, B. A., Kamfwa, K., Awale, H. E., and Kelly, J. D. (2016). Transcriptome profiling of the *Phaseolus vulgaris* - *Colletotrichum lindemuthianum* pathosystem. *PLoS One* 11, 1–26. doi:10.1371/journal.pone.0165823.
- Papa, R., Acosta, J., Delgado-Salinas, A., and Gepts, P. (2005). A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theor. Appl. Genet.* 111, 1147–1158. doi:10.1007/s00122-005-0045-9.
- Papa, R., and Gepts, P. (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor. Appl. Genet.* 106, 239–250. doi:10.1007/s00122-002-1085-z.
- Park, S. J. (2006). Majesty. Canadian Food Inspection Agency variety registration. Available at: <https://www.inspection.gc.ca/english/plaveg/pbrpov/cropreport/bea/app00004941e.shtml>.
- Park, S. J., and Buttery, B. R. (1992). Ethyl-methane sulphonate (EMS) induced nodulation mutants of common bean (*Phaseolus vulgaris* L.) lacking effective nodules. *Plant Soil* 139, 295–298.
- Park, S. J., and Buttery, B. R. (2006). Registration of Ineffective Nodulation Mutant R69 And Nonnodulation Mutant R99 Common Bean Genetic Stocks. *Crop Sci.* 46, 1415–1417. doi:10.2135/cropsci2005.10-0350.
- Park, S. J., Rupert, T., Yu, K., and Navabi, A. (2009). Red Rider common bean. *Can. J. Plant Sci.* 89, 925–927. doi:10.4141/CJPS09016.
- Pasam, R. K., Sharma, R., Malosetti, M., Eeuwijk, F. A. Van, Haseneyer, G., Kilian, B., et al. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.* 12, 16. doi:10.1186/1471-2229-12-

16.

- Peoples, M. B., Unkovich, M. J., and Herridge, D. F. (2009). "Measuring Symbiotic Nitrogen Fixation in Legumes," in *Nitrogen Fixation in Crop Production*, eds. D. W. Emerich and H. B. Krishnan doi:doi.org/10.2134/agronmonogr52.c6.
- Pereira, P. A. A., and Bliss, F. A. (1987). Nitrogen fixation and plant growth of common bean (*Phaseolus vulgaris* L.) at different levels of phosphorus availability. *Plant Soil* 104, 79–84. doi:10.1007/BF02370628.
- Pérez-Delgado, C. M., García-Calderón, M., Monje-Rueda, M. D., Márquez, A. J., and Betti, M. (2020). Transcriptomic Analysis of *L. japonicus* Symbiosis Reveals New Candidate Genes for Local and Systemic Regulation of Nodule Function. *Agronomy* 10, 1–26. doi:10.3390/agronomy10060819.
- Pii, Y., Molesini, B., and Pandolfini, T. (2013). The involvement of *Medicago truncatula* nonspecific lipid transfer protein N5 in the control of rhizobial infection. *Plant Signal. Behav.* 8, e24836. doi:10.4161/psb.24836.
- Plucknett, D. L., Smith, N. J. H., Williams, J. T., and Anishetty, N. M. (1987). *Gene banks and the world's food*. Princeton, NJ: Princeton University Press.
- Porch, T., Beaver, J. S., Debouck, D. G., Jackson, S. A., Kelly, J. D., and Dempewolf, H. (2013). Use of Wild Relatives and Closely Related Species to Adapt Common Bean to Climate Change. *Agronomy* 3, 433–461. doi:10.3390/agronomy3020433.
- Price, A. H. (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci.* 11, 213–216. doi:10.1016/j.tplants.2006.03.006.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959.
- R Core Team (2013). R: A Language and Environment for Statistical Computing.
- Raj, A., Stephens, M., and Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi:10.1534/genetics.114.164350.
- Ramaekers, L. (2011). Climbing for Nitrogen: Genetic Analysis of Symbiotic Nitrogen Fixation Capacity and Adoption Analysis in Climbing Beans.
- Ramaekers, L., Galeano, C. H., and Blair, M. W. (2013). Identifying quantitative trait loci for symbiotic nitrogen fixation capacity and related traits in common bean. *Mol. Breed.* 31, 163–180. doi:10.1007/s11032-012-9780-1.

- Recchia, G. H., Konzen, E. R., Cassieri, F., Caldas, D. G. G., and Tsai, S. M. (2018). Arbuscular mycorrhizal symbiosis leads to differential regulation of drought-responsive genes in tissue-specific root cells of common bean. *Front. Microbiol.* 9, 1–24. doi:10.3389/fmicb.2018.01339.
- Reinprecht, Y., Schram, L., Hill, B., and Pauls, K. P. (2020). Effects of Nitrogen Application on Nitrogen Fixation in Common Bean Production. 11, 1–19. doi:10.3389/fpls.2020.01172.
- Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Alvarez, S., Vlasova, A., Capella-Gutierrez, S., Ordaz-Ortiz, J. J., et al. (2017). Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* 18, 1–17. doi:10.1186/s13059-017-1190-6.
- Reyes, B. A. (2012). The Economic Impact of Improved Bean Varieties and Determinants of Market Participation: Evidence from Latin America and Angola.
- Rípodas, C., Castaingts, M., Clúa, J., Blanco, F., and Zanetti, M. E. (2015). Annotation , phylogeny and expression analysis of the nuclear factor Y gene families in common bean (*Phaseolus vulgaris*). 5. doi:10.3389/fpls.2014.00761.
- Rodino, A. P., Fuente, M. D. La, De Ron, A. M., Lema, M. J., Drevon, J. J., and Santalla, M. (2011). Variation for nodulation and plant yield of common bean genotypes and environmental effects on the genotype expression. *Plant Soil* 346, 349–361. doi:10.1007/s11104-011-0823-x.
- Rondon, M. A., Lehmann, J., Ramírez, J., and Hurtado, M. (2007). Biological nitrogen fixation by common beans (*Phaseolus vulgaris* L.) increases with bio-char additions. *Biol. Fertil. Soils* 43, 699–708. doi:10.1007/s00374-006-0152-z.
- Rosas, J. C. (2013). Contribuciones del Programa de Investigaciones en Frijol en Centro América y El Caribe. *Ceiba* 52, 65–73. doi:10.5377/ceiba.v52i1.967.
- Rosas, J. C., Beaver, J. S., Beebe, S., and Viana, A. (2004). Nomenclatura de variedades de frijol común liberadas en Centro América y El Caribe. 14, 221–224.
- Rosas, J. C., and Escoto, D. (2015). Paraisito Mejorado 2 (PM2-Don Rey). Available at: <http://www.programafpma.com/PDF/variedades/honduras/06PM2DonRey.pdf>.
- Rosas, J. C., and Vargas, A. G. Rhizobium Biofertilizante para el frijol comun.
- Ross, E. J. H., Kramer, S. B., and Dalton, D. A. (1999). Effectiveness of ascorbate and ascorbate peroxidase in promoting nitrogen fixation in model systems. *Phytochemistry* 52, 1203–1210. doi:10.1016/S0031-9422(99)00407-0.

- Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G., et al. (2009). Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol. Appl.* 2, 504–522. doi:10.1111/j.1752-4571.2009.00082.x.
- Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., and Pislariu, C. I. (2020). Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *Plant Cell* 32, 15–41. doi:10.1105/tpc.19.00279.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454.
- Salvagiotti, F., Cassman, K. G., Specht, J. E., Walters, D. T., Weiss, A., and Dobermann, A. (2008). Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. *F. Crop. Res.* 108, 1–13. doi:10.1016/j.fcr.2008.03.001.
- Sanyal, D., Osorno, J. M., and Chatterjee, A. (2020). Influence of Rhizobium inoculation on dry bean yield and symbiotic nitrogen fixation potential and symbiotic nitrogen fixation potential. *J. Plant Nutr.* 43, 798–810. doi:10.1080/01904167.2020.1711946.
- Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi:10.1086/502802.
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi:10.1038/ng.3008.
- Schneider, K. A., Brothers, M. E., and Kelly, J. D. (1997). Marker-Assisted Selection to Improve Drought Resistance in Common Bean. *Crop Sci.* 37, 51–60. doi:10.2135/cropsci1997.0011183X003700010008x.
- Schröder, S., Mamidi, S., Lee, R., McKain, M. R., McClean, P. E., and Osorno, J. M. (2016). Optimization of genotyping by sequencing (GBS) data in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 36, 1–9. doi:10.1007/s11032-015-0431-1.
- Sears, L. J. (1986). Lethbridge Research Station, 1976-1986: the Eighth Decade. *Hist. Ser. No. 24*, 1–52. Available at: <https://archive.org/details/lethbridgeresear24sear/mode/2up>.
- Shapiro, A. S. S., and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52, 591–611.
- Shearer, G., and Kohl, D. H. (1988). Natural <sup>15</sup>N abundance as a method of estimating

- the contribution of biologically fixed nitrogen to N<sub>2</sub>-fixing systems: Potential for non-legumes. *Plant Soil* 110, 317–327. doi:10.1007/BF02226812.
- Silva, D. A. da, Tsai, S. M., Chiorato, A. F., da Silva Andrade, S. C., de Fatima Esteves, J. A., Recchia, G. H., et al. (2019). Analysis of the common bean (*Phaseolus vulgaris* L.) transcriptome regarding efficiency of phosphorus use. *PLoS One* 14, 1–27. doi:10.1371/journal.pone.0210428.
- Singh, S. P. (1988). Gene pools in cultivated Dry Bean. in *Annual report of the Bean Improvement Cooperative, No. 31* (Denver, Colorado), 180–182.
- Singh, S. P. (1995). Selection for seed yield in Middle American versus Andean x Middle American interracial common-bean populations. *Plant Breed.* 114, 269–271. doi:10.1111/j.1439-0523.1995.tb00811.x.
- Singh, S. P. (2005). “Chapter 2: Common Bean (*Phaseolus vulgaris* L.),” in *Genetic Resources, Chromosome Engineering, and Crop Improvement: Grain Legumes, Volume 1*, eds. R. J. Singh and P. P. Jauhar (Boca Raton: Taylor & Francis), 32. doi:10.1201/9780203489284.
- Singh, S. P., Gepts, P., and Debouck, D. G. (1991). Races of Common Bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 45, 379–396.
- Singh, S. P., Munoz, C. G., and Teran, H. (2001). Registration of Common Bacterial Blight Resistant Dry Bean Germplasm VAX 1, VAX 3, and VAX 4. *Crop Sci.* 41, 275–276. doi:10.2135/cropsci2001.411275x.
- Singh, S. P., Schwartz, H. F., Viteri, D., Terán, H., and Otto, K. (2014). Introgressing White Mold Resistance from *Phaseolus coccineus* PI 439534 to Common Pinto Bean. *Crop Sci.* 54, 1026–1032. doi:10.2135/cropsci2013.07.0489.
- Singh, S. P., Terán, H., Lema, M., and Hayes, R. (2011). Selection for Dry Bean Yield On-Station Versus On-Farm Conventional and Organic Production Systems. *Crop Sci.* 51, 621–630. doi:10.2135/cropsci2010.02.0063.
- Smith, T. H., Michaels, T. E., Navabi, A., and Pauls, K. P. (2012). OAC Inferno common bean. *Can. J. Plant Sci.* 92, 589–592. doi:10.4141/CJPS2011-186.
- Soltani, A., MafiMoghaddam, S., Oladzad-Abbasabadi, A., Walter, K., Kearns, P. J., Vasquez-Guzman, J., et al. (2018). Genetic analysis of flooding tolerance in an andean diversity panel of dry bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.00767.
- Soltani, A., MafiMoghaddam, S., Walter, K., Restrepo-Montoya, D., Mamidi, S., Schroder, S., et al. (2017). Genetic architecture of flooding tolerance in the dry

- bean middle-American diversity panel. *Front. Plant Sci.* 8, 1–15. doi:10.3389/fpls.2017.01183.
- Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E. Y., Schroeder, S. G., et al. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3 Genes, Genomes, Genet.* 5, 2285–2290. doi:10.1534/g3.115.020594.
- Souza, L. P. De, Scossa, F., Proost, S., Bitocchi, E., Papa, R., Tohge, T., et al. (2019). Multi-tissue integration of transcriptomic and specialized metabolite profiling provides tools for assessing the common bean (*Phaseolus vulgaris*) metabolome. 1132–1153. doi:10.1111/tpj.14178.
- Stone, J. M., and Walker, J. C. (1995). Plant protein kinase families and signal transduction. *Plant Physiol.* 108, 451–457. doi:10.1104/pp.108.2.451.
- Sulieman, S., and Schulze, J. (2010). Phloem-derived  $\gamma$ -aminobutyric acid (GABA) is involved in upregulating nodule N<sub>2</sub> fixation efficiency in the model legume *Medicago truncatula*. *Plant, Cell Environ.* 33, 2162–2172. doi:10.1111/j.1365-3040.2010.02214.x.
- Sutton, M. A., Bleeker, A., Howard, C. M., Bekunda, M., Grizzetti, B., Vries, W. de, et al. (2013). Our Nutrient World: The challenge to produce more food and energy with less pollution. Global Overview of Nutrient Management. Edinburgh Available at: <http://initrogen.org/index.phppublications/our-nutrient-world/>.
- Swegarden, H. R., Sheaffer, C. C., and Michaels, T. E. (2016). Yield Stability of Heirloom Dry Bean (*Phaseolus vulgaris* L.) Cultivars in Midwest Organic Production. *HortScience* 51, 8–14. doi:10.21273/HORTSCI.51.1.8.
- Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* 123, 585–595.
- Tamura, K., and Nei, M. (1993). Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. *Mol. Biol. Evol.* 10, 512–526. doi:10.1093/oxfordjournals.molbev.a040023.
- Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11030–11035. doi:10.1073/pnas.0404206101.
- Tanksley, S. D., Young, N. D., Paterson, A. H., and Bonierbale, M. W. (1989). RFLP Mapping in Plant Breeding: New Tools for an Old Science. *Nat. Biotechnol.* 7, 257–264. doi:10.1038/nbt0389-257.

- Tock, A. J., Fourie, D., Walley, P. G., Holub, E. B., Soler, A., Cichy, K. A., et al. (2017). Genome-wide linkage and association mapping of halo blight resistance in common bean to race 6 of the globally important bacterial pathogen. *Front. Plant Sci.* 8, 1–17. doi:10.3389/fpls.2017.01170.
- Torkamaneh, D., and Belzile, F. (2015). Scanning and Filling : Ultra-Dense SNP Genotyping Combining Genotyping-By- Sequencing, SNP Array and Whole-Genome Resequencing Data. *PLoS One* 10, 1–16. doi:10.1371/journal.pone.0131533.
- Torkamaneh, D., Boyle, B., and Belzile, F. (2018). Efficient genome wide genotyping strategies and data integration in crop plants. *Theor. Appl. Genet.* 131, 499–511. doi:10.1007/s00122-018-3056-z.
- Tsai, S.-M., Nodari, R. O., Moon, D. H., Camargo, L. E. A., Vencovsky, R., and Gepts, P. (1998). QTL mapping for nodule number and common bacterial blight in *Phaseolus vulgaris* L. *Plant Soil* 204, 135–145. doi:10.1023/A:1004347515127.
- Udvardi, M. K., Kakar, K., Wandrey, M., Montanari, O., Murray, J., Andriankaja, A., et al. (2007). Legume Transcription Factors: Global Regulators of Plant Development and Response to the Environment. *Plant Physiol.* 144, 538–549. doi:10.1104/pp.107.098061.
- Ullah, I., Magdy, M., Wang, L., Liu, M., and Li, X. (2019). Genome-wide identification and evolutionary analysis of TGA transcription factors in soybean. *Sci. Rep.* 9, 11186. doi:10.1038/s41598-019-47316-z.
- Unkovich, M. J., and Pate, J. S. (2000). An appraisal of recent field measurements of symbiotic N<sub>2</sub> fixation by annual legumes. *F. Crop. Res.* 65, 211–228. doi:10.1016/S0378-4290(99)00088-X.
- van de Wouw, M., and Hintum, T. Van (2010). Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theor. Appl. Genet.* 120, 1241–1252. doi:10.1007/s00122-009-1252-6.
- Vance, C. P. (2001). Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. *Plant Physiol.* 127, 390–397. doi:10.1104/pp.010331.
- Vandenberg, A., and Bett, K. (2013). CDC Sol. Canadian Food Inspection Agency variety registration. Available at: <https://www.inspection.gc.ca/english/plaveg/pbrpov/cropreport/bea/app00007688e.shtml>.
- Varshney, R. K., Saxena, R. K., Upadhyaya, H. D., Khan, A. W., Yu, Y., Kim, C., et al.

- (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49, 1082–1088. doi:10.1038/ng.3872.
- Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi:10.1038/s41588-019-0401-3.
- Verzeaux, J., Hirel, B., Dubois, F., Lea, P. J., and Tétu, T. (2017). Agricultural practices to improve nitrogen use efficiency through the use of arbuscular mycorrhizae: Basic and agronomic aspects. *Plant Sci.* 264, 48–56. doi:10.1016/j.plantsci.2017.08.004.
- Villa, T. C. C., Maxted, N., Scholten, M., and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. *Plant Genet. Resour.* 3, 373–384. doi:10.1079/pgr200591.
- Voysest, V. (2000). Mejoramiento genético del frijol (*Phaseolus vulgaris* L.): Legado de variedades de América Latina 1930-1999 (Publicación CIAT no. 321). Cali.
- Wallace, D. H., Yourstone, K. S., Masaya, P. N., and Zobel, R. W. (1993). Photoperiod gene control over partitioning between reproductive and vegetative growth. *Theor. Appl. Genet.* 86, 6–16. doi:10.1007/BF00223803.
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. Cambridge* 94, 73–83. doi:10.1017/S0016672312000274.
- Wilker, J., Humphries, S., Rosas-Sotomayor, J. C., Gomez Cerna, M., Torkamaneh, D., Edwards, M., et al. (2020). Genetic Diversity, Nitrogen Fixation, and Water Use Efficiency in a panel of Honduran common bean (*Phaseolus vulgaris* L.) Landraces and Modern Genotypes. *Plants* 9, 1238. doi:https://doi.org/10.3390/plants9091238.
- Wilker, J., Navabi, A., Rajcan, I., Marsolais, F., Hill, B., Torkamaneh, D., et al. (2019). Agronomic Performance and Nitrogen Fixation of Heirloom and Conventional Dry Bean Varieties Under Low-Nitrogen Field Conditions. *Front. Plant Sci.* 10, 1–21. doi:10.3389/fpls.2019.00952.
- Williams, D. R. (2020). NASA Earth Fact Sheet. Available at: <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>.
- World Bank Group, Climate Change Knowledge Portal, Honduras *Clim. Chang. Knowl. Portal, Proj. Honduras*. Available at: <https://climateknowledgeportal.worldbank.org/country/honduras/climate-data-projections>.

- Xue, Y., Warburton, M. L., Sawkins, M., Zhang, X., Setter, T., Xu, Y., et al. (2013). Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. *Theor. Appl. Genet.* 126, 2587–2596. doi:10.1007/s00122-013-2158-x.
- Yan, W., and Rajcan, I. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Sci.* 42, 11–20. doi:10.2135/cropsci2002.1100.
- Yang, S., Tang, F., Gao, M., Krishnan, H. B., and Zhu, H. (2010). R gene-controlled host specificity in the legume – rhizobia symbiosis. *PNAS* 107, 18735–18740. doi:10.1073/pnas.1011957107.
- Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2020). rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study. *bioRxiv*, 2020.08.20.258491. doi:10.1101/2020.08.20.258491.
- Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160. doi:10.1016/j.copbio.2006.02.003.
- Zhang, C., Hou, Y., Hao, Q., Chen, H., Chen, L., Yuan, S., et al. (2015a). Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. *PLoS One* 10, 1–25. doi:10.1371/journal.pone.0125174.
- Zhang, X., Davidson, E. A., Mauzerall, D. L., Searchinger, T. D., Dumas, P., and Shen, Y. (2015b). Managing nitrogen for sustainable development. *Nature* 528, 51–59. doi:10.1038/nature15743.
- Zhang, X., Zhang, J., He, X., Wang, Y., Ma, X., and Yin, D. (2017). Genome-wide association study of major agronomic traits related to domestication in peanut. *Front. Plant Sci.* 8, 1–10. doi:10.3389/fpls.2017.01611.
- Zhang, Z., Ersoz, E., Lai, C., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi:10.1038/ng.546.Mixed.
- Zheng, F., Wu, H., Zhang, R., Li, S., He, W., Wong, F. L., et al. (2016). Molecular phylogeny and dynamic evolution of disease resistance genes in the legume family. *BMC Genomics* 17, 1–13. doi:10.1186/s12864-016-2736-9.
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414. doi:10.1038/nbt.3096.

Zuiderveen, G. H., Padder, B. A., Kamfwa, K., Song, Q., and Kelly, J. D. (2016).  
Genome-Wide association study of anthracnose resistance in andean beans  
(*Phaseolus vulgaris*). *PLoS One* 11, 1–17. doi:10.1371/journal.pone.0156391.

## **APPENDIX A**

Supplementary (S) Tables for Chapter 3

**Table S3.1.** Description of trial sites of the 2014-15 low-nitrogen dry bean (*Phaseolus vulgaris* L.) field trials.

Soil test (ppm) <sup>†</sup>		Elora 2014		Elora 2015		Belwood 2015	
		NO <sub>3</sub> -	NH <sub>4</sub>	NO <sub>3</sub> -	NH <sub>4</sub>	NO <sub>3</sub> -	NH <sub>4</sub>
	<b>0-15cm</b>	5.7	3.4	5.9	6.1	6.4	3.5
	<b>15-30cm</b>	8.6	2.9	3.7	3.0	5.5	2.6
<b>Planting date</b>		14-Jun		26-Jun		25-Jun	
<b>First killing frost date</b>		11-Oct		17-Oct		17-Oct	
<b>Total growing days</b>		124		114		115	
<b>Total growing degree days (GDD)<sup>‡</sup></b>		1912.8		1862.6		2012.3	
<b>Precipitation (mm)<sup>§</sup></b>							
	<b>June</b>	68.9		204.7		198.3	
	<b>July</b>	133.7		68.6		41.5	
	<b>August</b>	51.0		104.8		70.2	
	<b>September</b>	164.8		47.9		26.8	
	<b>October</b>	74.3		108.2		117.3	

<sup>†</sup> Soil tests performed by SGS Agrifood Laboratories (Guelph) on composite samples across the field sites of the top layer (0-15 cm) and mid layer (15-30 cm) of soil. Nitrate nitrogen (NO<sub>3</sub>-) and Ammonium nitrogen (NH<sub>4</sub>) are reported in parts per million (ppm). NO<sub>3</sub>- levels <5ppm are considered very low, and between 5<10 are considered low according to soil analysis laboratory guidelines.

<sup>‡</sup> Growing degree days (GDD, base '0') were calculated based on the total growing days for each location-year.

<sup>§</sup> Rainfall and temperature data for the Elora Research Station (Elora) were collected by the School of Environmental Sciences; rainfall and temperature data for the Belwood field site are as reported from the nearest Canada Weather Service location at the Fergus Shand Dam.

**Table S3.2.** F-test of fixed effect of genotype, and variance component estimates ( $S^2$ ) and their standard error (Se) of random effects in the combined mixed-model analysis of 41 heirloom and modern genotypes tested in multiple field locations in Ontario, Canada, 2014-2015.

Fixed effect†	Ndfa§		Seed N content		$\delta^{13}C^\diamond$		Flowering		Maturity		Yield		Hundred seed weight	
	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value
	(%)		(%)		(‰)		(GDD)		(GDD)		(kg ha <sup>-1</sup> )		(g)	
<b>Genotype</b>	2.61	0.0002	3.0200	<.0001	2.82	<.0001	11.47	<.0001	18.50	<.0001	2.52	0.0003	129.23	<.0001
<b>H vs. C</b>	2.01	0.1576	15.8900	<.0001	0.04	0.8362	8.37	0.0043	14.42	0.0002	0.55	0.4594	34.86	<.0001
<b>A vs. M</b>	19.54	<.0001	2.5500	0.1122	1.51	0.2211	44.01	<.0001	4.36	0.0385	1.95	0.1644	385.8	<.0001
Random effects‡	S <sup>2</sup>	Se	S <sup>2</sup>	Se	S <sup>2</sup>	Se	S <sup>2</sup>	Se	S <sup>2</sup>	Se	S <sup>2</sup>	Se	S <sup>2</sup>	Se
<b>env</b>	31.06 <sup>ns</sup>	42.388	0.05432 <sup>ns</sup>	0.057	0.1028 <sup>ns</sup>	0.108	1796.49 <sup>ns</sup>	1914.22 <sub>0</sub>	3878.29 <sup>ns</sup>	4058.16 <sub>0</sub>	0.2828 <sup>ns</sup>	0.297	0.1571 <sup>ns</sup>	0.6633
<b>BLOC(env)</b>	14.64 <sup>ns</sup>	16.086	0.002709 <sup>ns</sup>	0.004	0.00	.	166.43 <sup>ns</sup>	162.100	279.56 <sup>ns</sup>	293.980	0.01150 <sub>ns</sub>	0.020	0.389 <sup>ns</sup>	0.6545
<b>IBLK(env* BLOC)</b>	22.51*	9.865	0.000	.	0.00	.	0 <sup>ns</sup>	.	185.1 <sup>ns</sup>	140.520	0.05259 <sub>*</sub>	0.027	1.5325*	0.8413
<b>env*ENTR Y</b>	25.56*	11.315	0.01879*	0.010	0.08428*	0.039	592.93**	220.630	129.03 <sup>ns</sup>	195.370	0.05933 <sub>*</sub>	0.035	1.6721*	1.0016
<b>Residual</b>	64.55***	10.053	0.0703***	0.010	0.2513***	0.034	1285.23** <sub>*</sub>	171.160	1349.99** <sub>*</sub>	227.500	0.2181** <sub>*</sub>	0.034	6.0486***	0.9841

† Genotype effects overall and by subcategory, heirloom (H) vs. conventional (S), and Andean (A) vs. Middle American (M) genepool

‡ Environment (env), block (BLOC), and incomplete block (IBLK)

§ Nitrogen derived from the atmosphere

◇ Carbon discrimination, normalized

ns Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S3.3.** F-test of fixed effect of genotype, and variance component estimates ( $S^2$ ) and their standard error (Se) of random effects in the combined mixed-model analysis of 36 heirloom and modern genotypes tested at Elora, Ontario, Canada, 2014.

Fixed effect†	Ndfa§		Seed N content		δ <sup>13</sup> C◇		Flowering		Maturity		Yield		Hundred seed weight	
	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value
<b>Genotype</b>	3.26	0.0072	2.9400	0.0105	3.66	0.0031	2.00	0.0485	9.38	<.0001	3.48	0.0033	49.19	<.0001
<b>H vs. C</b>	2.23	0.1420	7.6700	0.0079	4.22	0.0452	1.67	0.2015	11.36	0.0014	1.54	0.2196	4.86	0.0340
<b>A vs. M</b>	6.47	0.0144	1.7000	0.1983	2.03	0.1613	6.32	0.015	5.54	0.0227	3.28	0.0764	106.83	<.0001
<b>Random effects‡</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>
<b>BLOC</b>	17.0987 <sub>ns</sub>	27.802	0.007079 <sub>ns</sub>	0.019	0.000	.	0.000	.	0.00	.	0	.	0	.
<b>IBLK(BLOC)</b>	0.000	.	0.02754 <sub>ns</sub>	0.026	0.06848 <sub>ns</sub>	0.05	0.000	.	132.47 <sub>ns</sub>	249.07 <sub>0</sub>	0	.	7.3088 <sub>ns</sub>	6.7337
<b>Residual</b>	74.09 <sup>***</sup>	19.801	0.06497 <sup>**</sup>	0.022	0.1178 <sup>**</sup>	0.04	2458.65 <sup>**</sup> <sub>*</sub>	596.31 <sub>0</sub>	1149.42 <sup>**</sup> <sub>*</sub>	357.66 <sub>0</sub>	161564 <sup>**</sup> <sub>*</sub>	41037	6.0957 <sub>ns</sub>	3.8924

† Genotype effects overall and by subcategory, heirloom (H) vs. conventional (S), and Andean (A) vs. Middle American (M) genepool

‡ Block (BLOC), and incomplete block (IBLK)

§ Nitrogen derived from the atmosphere

◇ Carbon discrimination, normalized

ns Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S3.4.** F-test of fixed effect of genotype, and variance component estimates ( $S^2$ ) and their standard error (Se) of random effects in the combined mixed-model analysis of 41 heirloom and modern genotypes tested at Belwood, Ontario, Canada, 2015.

Fixed effect†	Ndfa§		Seed N content		$\delta^{13}C^\diamond$		Flowering		Maturity		Yield		Hundred seed weight	
	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value
<b>Genotype</b>	2.64	0.0049	2.3000	0.0126	4.19	0.0001	17.15	<.0001	7.64	<.0001	1.52	0.1296	68.24	<.0001
<b>H vs. C</b>	0.31	0.5779	5.2500	0.0251	2.96	0.0899	1.95	0.1672	2.50	0.1193	1.27	0.2639	15.14	0.0002
<b>A vs. M</b>	3.68	0.0596	0.9900	0.3228	0.06	0.8043	14.08	0.0000	1.24	0.2699	0.41	0.5224	113.99	<.0001
Random effects‡	$S^2$	Se	$S^2$	Se	$S^2$	Se	$S^2$	Se	$S^2$	Se	$S^2$	Se	$S^2$	Se
<b>BLOC</b>	0.000	.	0.00	.	0.01 <sup>ns</sup>	0.025	179.340 <sup>n</sup> <sub>s</sub>	279.91 0	0	.	27905 <sup>ns</sup>	84308	1.3469 <sup>ns</sup>	2.3289
<b>IBLK(BLOC)</b>	0.000	.	0.00	.	0.00	.	0.000	.	0	.	147548 <sup>ns</sup>	11537 8	0.82 <sup>ns</sup>	1.5484
<b>Residual</b>	63.3036 <sup>**</sup> *	14.155	0.08644 <sup>**</sup> *	0.019	0.1683 <sup>**</sup> *	0.038	771.6 <sup>***</sup>	172.53 0	1631.7 <sup>**</sup> *	390.05	373670 <sup>**</sup> *	10015 9	7.0451 <sup>***</sup>	1.9072

† Genotype effects overall and by subcategory, heirloom (H) vs. conventional (S), and Andean (A) vs. Middle American (M) genepool

‡ Block (BLOC), and incomplete block (IBLK)

§ Nitrogen derived from the atmosphere

◇ Carbon discrimination, normalized

ns Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S3.5.** F-test of fixed effect of genotype, and variance component estimates ( $S^2$ ) and their standard error (Se) of random effects in the combined mixed-model analysis of 41 heirloom and modern genotypes tested at Elora, Ontario, Canada, 2015.

Fixed effect†	Ndfa§		Seed N content		$\delta^{13}C^\diamond$		Flowering		Maturity		Yield		Hundred seed weight	
	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value	F-test	P-value
<b>Genotype</b>	2.91	0.0026	4.7000	<.0001	2.44	0.0078	19.4300	<.0001	11.2500	0.0002	5.45	<.0001	99.55	<.0001
<b>H vs. C</b>	0.15	0.6980	3.3200	0.0731	5.39	0.0233	4.780	0.0320	3.6100	0.0640	3.0700	0.0843	15.17	0.0002
<b>A vs. M</b>	10.45	0.0020	0.1300	0.7213	8.53	0.0049	24.920	<.0001	0.1700	0.6825	0.0100	0.9437	169.89	<.0001
<b>Random effects‡</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>	<b>S<sup>2</sup></b>	<b>Se</b>
<b>BLOC</b>	32.3229 <sup>n</sup> <sub>s</sub>	60.210	0.000	.	0.00	.	329.6 <sup>ns</sup> <sub>0</sub>	500.110	1253.11 <sup>n</sup> <sub>s</sub>	1974.85	0	.	0	.
<b>IBLK(BLOC)</b>	60.3893 <sup>*</sup>	34.019	0.01057 <sup>ns</sup>	0.008	0.00	.	32.1592 <sup>n</sup> <sub>s</sub>	123.940	606.39 <sup>ns</sup>	542.7	57235 <sup>n</sup> <sub>s</sub>	32589	0.234 <sup>ns</sup>	0.7947
<b>Residual</b>	59.968 <sup>**</sup> <sub>*</sub>	16.583	0.02858 <sup>**</sup> <sub>*</sub>	0.008	0.3193 <sup>**</sup> <sub>*</sub>	0.07	773 <sup>***</sup>	201.770	1072.99 <sup>*</sup> <sub>*</sub>	431.76	70697 <sup>n</sup> <sub>s</sub>	18948	4.643 <sup>***</sup>	1.246

† Genotype effects overall and by subcategory, heirloom (H) vs. conventional (S), and Andean (A) vs. Middle American (M) gene pool

‡ Block (BLOC), and incomplete block (IBLK)

§ Nitrogen derived from the atmosphere

◇ Carbon discrimination, normalized

ns Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S3.6.** F-test of fixed effects and random effects in the mixed-model analysis of leaf chlorophyll content (SPAD) for 48 heirloom and conventional genotypes tested at three field locations in Ontario, Canada, 2014-2015.

Fixed effects†	Elora 2014		Elora 2015		Belwood 2015	
	F-test	P-value	F-test	P-value	F-test	P-value
<b>G</b>	5.86	<.0001	7.93	<.0001	5.06	<.0001
<b>SPADT</b>	3.47	0.0637	4.35	0.0378	43.64	<.0001
<b>G*SPADT</b>	2.5	<.0001	1.65	0.0109	2.52	<.0001
Random effects‡	S <sup>2</sup>	Se	S <sup>2</sup>	Se	S <sup>2</sup>	Se
<b>BLOC</b>	0.03	0.346 <sup>ns</sup>	0.00	.	0.00	.
<b>IBLK(BLOC)</b>	0.62	0.718	0.30	0.471 <sup>ns</sup>	0.73	0.9199 <sup>ns</sup>
<b>observation(BLOC*G*SPADT)</b>	13.73	1.2784 <sup>***</sup>	14.00	1.1871 <sup>***</sup>	22.38	1.8618 <sup>***</sup>
<b>Residual</b>	0.94	0.000	0.94	0	0.96	0

† Genotype (G) main effect, SPAD timing (SPADT), and the interaction of G\*SPADT

‡ Block (BLOC), and incomplete block (IBLK),

ns Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S3.7.** Phenotypic ( $r_p$ ) correlations among %Ndfa and other traits estimated in panel of 35 heirloom or conventional genotypes grown in Elora, Ontario, 2014.

	Days to Maturity (GDD)	$\delta^{13}\text{C}$ (‰)	Yield (kg ha <sup>-1</sup> )	Seed N content (mg)	Ndfa (%)
<b>Days to flowering</b>	0.59**	-0.04	-0.28	-0.13	0.24
<b>Days to maturity</b>		0.12	-0.02	0.08	0.10
<b><math>\delta^{13}\text{C}</math></b>			-0.21	0.15	-0.21
<b>Yield</b>				-0.21	-0.03
<b>Seed N content</b>					-0.04

\* and \*\* are significant at 0.05 and 0.01, respectively. Significant  $r_p$  tested by the Student's t test.

**Table S3.8.** Phenotypic ( $r_p$ ) correlations among %Ndfa and other traits estimated in panel of heirloom or conventional genotypes grown in Elora, Ontario, 2015. The number of genotypes in each correlation is presented in the table.

	Days to Maturity (GDD)	$\delta^{13}\text{C}$ (‰)	Yield (kg ha <sup>-1</sup> )	Seed N content (mg)	Ndfa (%)
<b>Days to flowering</b>	0.76**	0.53**	-0.09	0.05	0.51**
number of genotypes	35	41	40	41	41
<b>Days to maturity</b>		0.37*	-0.13	-0.09	0.45**
number of genotypes		35	34	35	35
<b><math>\delta^{13}\text{C}</math></b>			0.11	0.07	0.45**
number of genotypes			40	41	41
<b>Yield</b>				-0.43763**	0.21
number of genotypes				40	40
<b>Seed N content</b>					-0.07
number of genotypes					41

\* and \*\* are significant at 0.05 and 0.01, respectively. Significant  $r_p$  tested by the Student's t test.

**Table S3.9.** Phenotypic ( $r_p$ ) correlations among %Ndfa and other traits estimated in panel of heirloom or conventional genotypes grown in Belwood, Ontario, 2015. The number of genotypes in each correlation is presented in the table.

	Days to Maturity (GDD)	$\delta^{13}\text{C}$ (‰)	Yield (kg ha <sup>-1</sup> )	Seed N content (mg)	Ndfa (%)
<b>Days to flowering</b>	0.59**	0.11	-0.06	0.38	0.55**
number of genotypes	40	41	40	41	41
<b>Days to maturity</b>		0.48**	0.06	0.56**	0.37
number of genotypes		40	40	40	40
<b><math>\delta^{13}\text{C}</math></b>			0.17	0.56**	0.26
number of genotypes			40	41	41
<b>Yield</b>				-0.05	0.04
number of genotypes				40	40
<b>Seed N content</b>					0.30
number of genotypes					41

\* and \*\* are significant at 0.05 and 0.01, respectively. Significant  $r_p$  tested by the Student's t test.

## **APPENDIX B**

Supplementary (S) Tables and Figures for Chapter 4.

**Table S4.1.** Candidate genes identified in regions of the *P. vulgaris* (2.0) genome where high nucleotide diversity ( $\pi$ ) was discovered in landrace genotypes compared to PPB genotypes. Candidate genes within or overlapping with these regions are listed with annotation from JBrowse (<https://legumeinfo.org/genomes/jbrowse/>). (Chr – Chromosome)

Region of high diversity		$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies	
Chr	Start (Mbp)	End (Mbp)	Landrace					PPB
1	23.0	24.0	5.4997E-07	1.3107E-07	2	<b>Phvul.001G100800</b>	<b>Phvul.001G108101</b> (23,191,560..23,195,119) MATE efflux family protein	QTL (23.2Mbp) associated with days to flowering and hundred seed weight in Middle American diversity panel (Chapter 5, this thesis).
1	40.0	41.0	7.7102E-07	1.8305E-07	4	<b>Phvul.001G148100</b> <b>Phvul.001G150700</b> <b>Phvul.001G150800</b> <b>Phvul.001G151700</b>		
1	41.0	42.0	1.2422E-06	2.4915E-07	4	<b>Phvul.001G156000</b> <b>Phvul.001G156200</b> <b>Phvul.001G156400</b> <b>Phvul.001G160000</b>		
1	42.0	43.0	1.2226E-06	3.9831E-07	1		<b>Phvul.001G167200</b> (42,856,008..42,859,005) RNA polymerase-associated protein RTF1 homolog [Glycine max]	QTL (42.9Mbp) associated with growth habit in Middle American diversity panel (Moghaddam et al, 2016).
1	47.0	48.0	8.6039E-07	9.661E-08	3	<b>Phvul.001G210000</b> <b>Phvul.001G210300</b> <b>Phvul.001G212400</b>		
2	4.0	5.0	1.9461E-06	6.5085E-07	0			
2	22.0	23.0	1.7113E-06	1.6441E-07	8	<b>Phvul.002G109400</b>	<b>Phvul.002G106500</b> (21,498,604..21,509,530) vacuolar sorting protein 35 <b>Phvul.002G106600</b> (21,510,899..21,513,855) RHO family GTPase <b>Phvul.002G106700</b> (21,516,894..21,520,358) dehydroquinase synthase	Genes underlying QTL (22.8Mbp) associated with %N in the Middle American diversity panel (Chapter 5, this thesis). GTPase gene (Phvul.002G106600), Flores et al, 2018.

Region of high diversity		$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace				
						<b>Phvul.002G106800</b> (21,521,806..21,525,244) Myb/SANT-like DNA-binding domain protein  <b>Phvul.002G106900</b> (21,526,081..21,530,585) uncharacterized protein  <b>Phvul.002G107000</b> (21,531,223..21,533,198) peroxidase  <b>Phvul.002G107100</b> (21,585,247..21,592,897) histidine kinase	
2	32.0	33.0	5.5262E-07	1.2203E-07	5	<b>Phvul.002G172300</b> <b>Phvul.002G172500</b> <b>Phvul.002G172600</b> <b>Phvul.002G174500</b> <b>Phvul.002G174600</b>	
2	48.0	49.0	5.9207E-06	1.6232E-06	18	<b>Phvul.002G320800</b>  <b>Phvul.002G320000</b> (47,985,266..47,989,823) Ubiquitin-conjugating enzyme E2  <b>Phvul.002G320100</b> (47,991,022..47,997,910) Histidyl-tRNA synthetase  <b>Phvul.002G320200</b> (47,999,699..48,001,370) Ribosomal protein S19 family protein  <b>Phvul.002G320300</b> (48,004,702..48,009,035) XAP5 family protein	Genes underlying QTL (48.6Mbp) associated with days to flowering in Middle American diversity panel (Chapter 5, this thesis). Calmodulin-like gene (Phvul.002G320800), Mohanta et al, 2017.

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
							<b>Phvul.002G320400</b> (48,014,427..48,017,026) XAP5 family protein (circadian clock regulator)	
							<b>Phvul.002G320500</b> (48,021,216..48,022,697) diphosphoinositol polyphosphate phosphohydrolase	
							<b>Phvul.002G320600</b> (48,029,075..48,030,117) unknown protein	
							<b>Phvul.002G320700</b> (48,030,317..48,033,276) uncharacterized putative methyltransferase	
							<b>Phvul.002G320800</b> (48,035,445..48,036,011) probable calcium-binding protein CML25-like	
							<b>Phvul.002G320900</b> (48,037,219..48,041,606) ubiquitin fusion degradation 1	
							<b>Phvul.002G321000</b> (48,043,674..48,047,394) Ankyrin repeat family protein	
							<b>Phvul.002G321100</b> (48,047,868..48,055,921) protein VAC14 homolog [Glycine max]	
							<b>Phvul.002G321200</b> (48,057,723..48,067,814) nuclear pore complex protein Nup155-like [Glycine max]	

Region of high diversity		$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace				
						<b>Phvul.002G321300</b> (48,070,793..48,074,236) RING-H2 finger protein [Glycine max] <b>Phvul.002G321400</b> (48,077,719..48,079,912) Ribosomal S5 family protein <b>Phvul.002G321500</b> (48,082,538..48,087,118) microtubule-associated proteins 70-2 <b>Phvul.002G321600</b> (48,087,876..48,089,363) Mitochondrial transcription termination factor family protein <b>Phvul.002G321700</b> (48,091,104..48,099,000) DYNAMIN-like 1C vacuolar sorting protein	
3	34.0	35.0	6.7266E-07	1.9661E-07	8	<b>Phvul.003G144500</b> <b>Phvul.003G145300</b> <b>Phvul.003G145400</b> <b>Phvul.003G145700</b> <b>Phvul.003G146000</b> <b>Phvul.003G146600</b> <b>Phvul.003G146800</b> <b>Phvul.003G148900</b>	
3	48.0	49.0	1.1682E-06	1.887E-07	8	<b>Phvul.003G253300</b> <b>Phvul.003G254000</b> <b>Phvul.003G254800</b> <b>Phvul.003G254900</b> <b>Phvul.003G258800</b> <b>Phvul.003G259500</b> <b>Phvul.003G259700</b> <b>Phvul.003G259800</b>	

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
4	12.0	13.0	1.5299E-06	2.6215E-07	0			
4	17.0	18.0	4.0455E-07	6.5537E-08	0			
4	38.0	39.0	4.7435E-07	9.661E-08	6	Phvul.004G116300 Phvul.004G116700 Phvul.004G117700 Phvul.004G118100 Phvul.004G119700 Phvul.004G120000		
4	40.0	41.0	4.3152E-07	1.2655E-07	0			
5	24.0	25.0	3.5537E-07	6.5537E-08	1	Phvul.005G089100		
5	25.0	26.0	2.1047E-06	1.9209E-07	0			
5	27.0	28.0	7.4564E-07	1.3107E-07	1	Phvul.005G092300		
5	32.0	33.0	1.1592E-06	3.2768E-07	11	Phvul.005G109700 Phvul.005G110000	Phvul.005G104700 (31,137,049..31,138,536) 60S acidic ribosomal protein family Phvul.005G104800 (31,142,070..31,143,525) Clathrin light chain protein  Phvul.005G104900 (31,158,921..31,164,749) pre-mRNA splicing factor-related  Phvul.005G105000 (31,165,180..31,167,685) sugar porter (SP) family MFS transporter  Phvul.005G105100 (31,190,419..31,191,384) hypothetical protein	Genes underlying QTL (32.5Mbp) associated with hundred seed weight in the Middle American diversity panel (Chapter 5, this thesis). Gene (Phvul.005G105200) involved in phosphorus use efficiency, da Silva et al, 2019, and involved in abiotic stress, Konzen et al, 2019.

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American gene pool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
							<p><b>Phvul.005G105200</b> (31,203,406..31,204,629) ethylene responsive element binding factor 1</p> <p><b>Phvul.005G105300</b> (31,226,716..31,230,233) Regulator of Vps4 activity in the MVB pathway protein</p> <p><b>Phvul.005G105400</b> (31,234,261..31,235,004) Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family</p> <p><b>Phvul.005G105500</b> (31,235,613..31,236,357) hypothetical protein</p>	
6	9.0	10.0	1.2057E-06	6.5537E-08	7	<p><b>Phvul.006G018100</b></p> <p><b>Phvul.006G018700</b></p> <p><b>Phvul.006G019300</b></p> <p><b>Phvul.006G019500</b></p> <p><b>Phvul.006G019600</b></p> <p><b>Phvul.006G019700</b></p> <p><b>Phvul.006G019900</b></p>	<b>Phvul.006G019300</b>	Gene (Phvul.006G019300) associated with phosphorus use efficiency, da Silva et al, 2019.
6	13.0	14.0	1.018E-06	1.3107E-07	14	<p><b>Phvul.006G032500</b></p> <p><b>Phvul.006G033700</b></p> <p><b>Phvul.006G034400</b></p>	<p><b>Phvul.006G039100</b> (14,865,891..14,874,127) DEAD-box ATP-dependent RNA helicase</p> <p><b>Phvul.006G039000</b> (14,862,108..14,863,739) LRR and NB-ARC domain disease resistance protein</p>	Genes underlying QTL (13.9Mbp) associated with days to flowering in the Middle American diversity panel (Chapter 5, this thesis). Gene (Phvul.006G038100) involved in bean common mosaic virus response, Martin et al, 2016.

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
							<b>Phvul.006G038800</b> (14,854,569..14,857,149) uncharacterized protein LOC100777222 isoform X2 [Glycine max]	
							<b>Phvul.006G038900</b> (14,855,007..14,855,339) uncharacterized protein LOC100803137 [Glycine max]	
							<b>Phvul.006G038700</b> (14,849,231..14,850,850) LRR and NB-ARC domain disease resistance protein	
							<b>Phvul.006G038600</b> (14,843,654..14,846,550) LHCP translocation defect protein, putative	
							<b>Phvul.006G038500</b> (14,834,906..14,839,087) serine/arginine repetitive matrix protein 2-like [Glycine max]	
							<b>Phvul.006G038400</b> (14,825,494..14,828,231) septum-promoting GTP-binding protein 1-like [Glycine max]	
							<b>Phvul.006G038300</b> (14,818,508..14,823,231) hypothetical protein	
							<b>Phvul.006G038200</b> (14,808,094..14,816,761) Nucleotide/sugar transporter family protein	

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
							<b>Phvul.006G038100</b> (14,805,826..14,807,523) disease resistance protein (CC-NBS-LRR class) family protein	
6	15.0	16.0	4.8229E-07	1.5537E-07	6	<b>Phvul.006G040400</b> <b>Phvul.006G040800</b> <b>Phvul.006G043100</b> <b>Phvul.006G044500</b> <b>Phvul.006G046100</b> <b>Phvul.006G046400</b>		
7	8.0	9.0	4.3205E-07	9.661E-08	1	<b>Phvul.007G011700</b>		
7	20.0	21.0	5.2723E-07	6.5537E-08	1	<b>Phvul.007G120500</b>		
7	26.0	27.0	3.9831E-06	1.0124E-06	0			
7	35.0	36.0	3.3644E-06	7.3785E-07	9	<b>Phvul.007G232000</b> (47,185,070..47,194,601) 3-isopropylmalate dehydratase, large subunit <b>Phvul.007G232100</b> (47,195,159..47,196,475) Calcium-dependent lipid-binding (CaLB domain) family protein <b>Phvul.007G232200</b> (47,197,987..47,204,888) ras GTPase-activating protein-binding protein 2-like isoform X1 [Glycine max] <b>Phvul.007G232300</b> (47,208,285..47,211,396) cation/H+ exchanger	Genes underlying QTL (35.6Mbp) associated with days to maturity in the Middle American diversity panel (Chapter 5, this thesis).	

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
							<b>Phvul.007G232400</b> (47,219,495..47,223,009) cation/H+ exchanger <b>Phvul.007G232500</b> (47,225,864..47,227,706) Pentatricopeptide repeat (PPR-like) superfamily protein <b>Phvul.007G232600</b> (47,242,795..47,243,289) small acidic protein 1-like [Glycine max] <b>Phvul.007G232800</b> (47,262,900..47,289,477) lipase-like isoform X2 [Glycine max] <b>Phvul.007G232900</b> (47,293,727..47,320,751) DNA mismatch repair MUTS family protein	
7	39.0	40.0	2.4331E-06	3.3164E-07	8	<b>Phvul.007G164500</b> <b>Phvul.007G165900</b> <b>Phvul.007G166100</b> <b>Phvul.007G166200</b> <b>Phvul.007G166300</b> <b>Phvul.007G166700</b> <b>Phvul.007G166800</b> <b>Phvul.007G166900</b>		
8	11.0	12.0	5.4389E-06	1.7424E-06	1	<b>Phvul.008G103600</b>		
8	15.0	16.0	1.6298E-06	5.322E-07	2	<b>Phvul.008G121000</b> <b>Phvul.008G121100</b>		
8	18.0	19.0	1.2977E-06	3.9774E-07	1	<b>Phvul.008G125300</b>		
8	23.0	24.0	5.2723E-07	6.5537E-08	0			
8	29.0	30.0	1.3453E-06	3.322E-07	0			

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
8	38.0	39.0	3.2893E-07	3.3333E-08	0			
8	41.0	42.0	7.6203E-07	2.2599E-07	2	Phvul.008G162600 Phvul.008G162700		
8	44.0	45.0	1.3533E-06	1.6215E-07	0			
8	45.0	46.0	9.1486E-07	2.8983E-07	0			
8	51.0	52.0	2.8593E-06	6.3333E-07	6	Phvul.008G202200 Phvul.008G202300 Phvul.008G202400 Phvul.008G204800 Phvul.008G208700 Phvul.008G208800		
8	52.0	53.0	1.211E-06	9.661E-08	12	Phvul.008G211100 Phvul.008G211200 Phvul.008G211600 Phvul.008G213000 Phvul.008G214400 Phvul.008G214600 Phvul.008G216100 Phvul.008G216200 Phvul.008G216300 Phvul.008G216600 Phvul.008G216700 Phvul.008G217100	Phvul.008G211200 Phvul.008G214400 Phvul.008G217100	Genes (Phvul.008G211200, Phvul.008G214400) involved in bean rust response, Ayyappan et al, 2015. Gene (Phvul.008G217100) involved in nodule ontogeny, Guillen et al, 2013.
8	57.0	58.0	3.4532E-07	6.5537E-08	0			
9	16.0	17.0	7.8107E-07	1.5989E-07	0			
9	18.0	19.0	9.0376E-07	3.3333E-08	3	Phvul.009G123600 Phvul.009G123900 Phvul.009G124700		
9	21.0	22.0	1.9434E-06	4.9944E-07	0			

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
10	17.0	18.0	2.9667E-07	9.322E-08	0			
10	33.0	34.0	2.2597E-06	7.4124E-07	0			
10	35.0	36.0	9.7832E-07	2.6215E-07	2	Phvul.010G101800 Phvul.010G102300	Phvul.010G101800 Phvul.010G102300	Gene (Phvul.010G101800) described in bean metabolomics study, Perez de Souza et al, 2019. NF-Y family gene (Phvul.010G102300), Ripodas et al, 2014.
10	37.0	38.0	1.0661E-06	1.9435E-07	7	Phvul.010G110900 Phvul.010G112300 Phvul.010G112400 Phvul.010G112500 Phvul.010G112600 Phvul.010G112800 Phvul.010G112900	Phvul.010G112400	Gene (Phvul.010G112400) involved in Pseudomonas response, Gonzalez et al, 2017.
10	39.0	40.0	2.9995E-06	7.339E-07	1	Phvul.010G121500	Phvul.010G121500	Gene (Phvul.010G121500) involved in phosphorus use efficiency, da Silva et al, 2019.
11	2.0	3.0	8.7414E-07	1.9322E-07	3	Phvul.011G026200 Phvul.011G026300 Phvul.011G026400	Phvul.011G026300	Gene (Phvul.011G026300) involved in phosphorus use efficiency, da Silva et al, 2019.

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American gene pool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
11	7.0	8.0	7.9958E-07	1.5537E-07	23	Phvul.011G077200 Phvul.011G077300 Phvul.011G078400 Phvul.011G078600 Phvul.011G078700 Phvul.011G078800 Phvul.011G079000 Phvul.011G079300 Phvul.011G079500 Phvul.011G079600 Phvul.011G079900 Phvul.011G080900 Phvul.011G081200 Phvul.011G081400 Phvul.011G081500 Phvul.011G081600 Phvul.011G081700 Phvul.011G081800 Phvul.011G082100 Phvul.011G082400 Phvul.011G082700 Phvul.011G083100 Phvul.011G083200	Phvul.011G079300	Gene (Phvul.011G079300) involved in drought response under AM symbiosis, Recchia et al, 2018.
11	10.0	11.0	1.3749E-06	3.0057E-07	7	Phvul.011G098500 Phvul.011G099300 Phvul.011G099700 Phvul.011G099900 Phvul.011G100000 Phvul.011G100800 Phvul.011G101000		
11	37.0	38.0	2.6367E-06	7.8418E-07	1	Phvul.011G144600		
11	40.0	41.0	9.0693E-07	3.0057E-07	3		Phvul.011G144600 (37,051,493..37,051,901) DEAD-box ATP-dependent RNA helicase	Genes underlying QTL (40.3Mbp) associated with days to maturity in the

Region of high diversity		$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace				
						<p><b>Phvul.011G144500</b> (37,048,250..37,049,231) UPF0481 protein At3g47200-like [Glycine max]</p> <p><b>Phvul.011G144700</b> (37,053,630..37,055,280) DEAD-box ATP-dependent RNA helicase 39-like [Glycine max]</p>	Middle American diversity panel (Chapter 5, this thesis).
11	45.0	46.0	7.8477E-06	9.8023E-07	9	<p><b>Phvul.011G175900</b> (45,068,805..45,071,512) putative ribonuclease H protein At1g65750-like [Glycine max]</p> <p><b>Phvul.011G176000</b> (45,075,074..45,076,054) RNA recognition motif, a.k.a. RRM, RBD protein</p> <p><b>Phvul.011G176200</b> (45,088,313..45,089,882) salicylic acid carboxyl methyltransferase</p> <p><b>Phvul.011G176300</b> (45,106,776..45,110,315) receptor kinase 3</p> <p><b>Phvul.011G176400</b> (45,118,181..45,121,444) Protein kinase superfamily protein</p> <p><b>Phvul.011G176500</b> (45,126,787..45,131,817) WD repeat-containing protein 5-like [Glycine max]</p> <p><b>Phvul.011G176600</b> (45,134,053..45,146,869) Transducin/WD40 repeat-like superfamily protein</p>	Genes underlying QTL (45.1Mbp) associated with shoot biomass (Kamfwa et al, 2019; Heilig et al, 2017).

Region of high diversity		$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace				
						<b>Phvul.011G176700</b> (45,151,411..45,155,628) Protein kinase superfamily protein  <b>Phvul.011G176800</b> (45,178,234..45,191,719) Protein kinase superfamily protein	
11	47.0	48.0	7.3062E-06	1.526E-06	0		
11	48.0	49.0	1.5156E-06	4.9944E-07	0		
11	53.0	54.0	6.8958E-07	2.2768E-07	11	<b>Phvul.011G215400</b> (50,112,467..50,116,453) RING/U-box superfamily protein  <b>Phvul.011G215500</b> (50,117,514..50,119,873) organic cation/carnitine transporter 2  <b>Phvul.011G215600</b> (50,122,884..50,125,029) organic cation/carnitine transporter  <b>Phvul.011G215700</b> (50,126,054..50,130,876) unknown protein  <b>Phvul.011G215800</b> (50,132,400..50,137,832) eukaryotic translation initiation factor SUI1 family protein  <b>Phvul.011G215900</b> (50,142,442..50,154,147) dynamin-like 3 (GTPase activity, GTP binding)  <b>Phvul.011G216000</b> (50,155,120..50,159,314) Galactose-binding protein	Genes underlying QTL (53.5Mbp) associated with hundred seed weight in the Middle American diversity panel (Chapter 5, this thesis).

Region of high diversity			$\pi$ value		Total number of genes per region	Middle American genepool domestication genes*	QTL and genes identified in other studies	Description of QTL and genes identified in other studies
Chr	Start (Mbp)	End (Mbp)	Landrace	PPB				
							<b>Phvul.011G216100</b> (50,160,731..50,171,829) Molybdenum cofactor sulfurase family protein  <b>Phvul.011G216200</b> (50,176,173..50,185,138) ATP binding protein, putative n=1 Tax=Ricinus communis RepID=B9SHL9_RICCO  <b>Phvul.011G216300</b> (50,188,296..50,190,525) Peroxidase superfamily protein  <b>Phvul.011G216400</b> (50,196,967..50,204,051) Signal transduction histidine kinase, hybrid-type, ethylene sensor	

\*see Schmutz et al, 2014, "Supplementary Table 16 Mesoamerican domestication candidates" for details.

**Table S4.2.** Candidate genes identified in regions of the *P. vulgaris* (2.0) genome where SNPs with significantly high weighted  $F_{ST}$  values (> 0.5) were found. Candidate genes within 100 Kb of the significant SNP are listed with annotation from JBrowse (<https://legumeinfo.org/genomes/jbrowse/>). (Chr – chromosome)

Chr	Position (Mbp)	$F_{ST}$ value	Candidate genes annotated
2	48,908,956	0.541	<p><b>Phvul.002G323708</b> Disease resistance protein (TIR-NBS-LRR class) family; IPR025875 (Leucine rich repeat 4) Position Pv02: 48,873,058..48,884,543 (reference: Oladzad et al 2019, [40])</p> <p><b>Phvul.002G323712</b> Disease resistance protein (TIR-NBS-LRR class), putative; IPR000157 (Toll/interleukin-1 receptor homology (TIR) domain), IPR000767 (Disease resistance protein), IPR025875 (Leucine rich repeat 4), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0000166 (nucleotide binding), GO:0005515 (protein binding), GO:0006952 (defense response), GO:0007165 (signal transduction), GO:0017111 (nucleoside-triphosphatase activity), GO:0043531 (ADP binding) Position Pv02: 48,888,578..48,895,789 (reference: Oladzad et al 2019, [40])</p> <p><b>Phvul.002G323900</b> Aspartate-semialdehyde dehydrogenase; IPR012080 (Aspartate-semialdehyde dehydrogenase), IPR016040 (NAD(P)-binding domain); GO:0003942 (N-acetyl-gamma-glutamyl-phosphate reductase activity), GO:0004073 (aspartate-semialdehyde dehydrogenase activity), GO:0005737 (cytoplasm), GO:0008652 (cellular amino acid biosynthetic process), GO:0009086 (methionine biosynthetic process), GO:0009088 (threonine biosynthetic process), GO:0009089 (lysine biosynthetic process via diaminopimelate), GO:0009097 (isoleucine biosynthetic process), GO:0046983 (protein dimerization activity), GO:0050661 (NADP binding), GO:0051287 (NAD binding), GO:0055114 (oxidation-reduction process) Position Pv02: 48,895,240..48,898,436</p> <p><b>Phvul.002G324100</b> PLATZ transcription factor family protein; IPR000315 (Zinc finger, B-box), IPR006734 (Protein of unknown function DUF597); GO:0005622 (intracellular), GO:0008270 (zinc ion binding) Position Pv02: 48,906,246..48,908,726</p> <p><b>Phvul.002G324200</b> Mitochondrial import receptor subunit TOM40-1-like [Glycine max]; IPR023614 (Porin domain), IPR027246 (Eukaryotic porin/Tom40); GO:0005741 (mitochondrial outer membrane), GO:0055085 (transmembrane transport) Position Pv02: 48,913,224..48,918,020</p> <p><b>Phvul.002G324300</b> 40S ribosomal protein S16-like [Glycine max]; IPR000754 (Ribosomal protein S9), IPR020568 (Ribosomal protein S5 domain 2-type fold); GO:0003735 (structural constituent of ribosome), GO:0005840 (ribosome), GO:0006412 (translation) Position Pv02: 48,919,930..48,920,791</p> <p><b>Phvul.002G324400</b> Clathrin coat assembly protein AP180-like [Glycine max]; IPR008942 (ENTH/VHS), IPR011417 (AP180 N-terminal homology (ANTH) domain); GO:0005543 (phospholipid binding), GO:0005545 (1-phosphatidylinositol binding), GO:0030118 (clathrin coat), GO:0030276 (clathrin binding), GO:0048268 (clathrin coat assembly) Position Pv02: 48,924,528..48,925,847</p> <p><b>Phvul.002G324500</b> THO complex subunit 2; IPR021418 (THO complex, subunitTHOC2, C-terminal), IPR021726 (THO complex, subunitTHOC2, N-terminal) Position Pv02: 48,926,349..48,950,645</p>
2	49,075,854	0.546	<p><b>Phvul.002G325500</b> Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)</p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			Position Pv02: 49,039,240..49,039,764 <b>Phvul.002G325600</b> Uncharacterized protein LOC100820443 [Glycine max]; IPR006747 (Protein of unknown function DUF599) Position Pv02: 49,053,148..49,053,874 <b>Phvul.002G325700</b> Uncharacterized protein LOC100820443 [Glycine max]; IPR006747 (Protein of unknown function DUF599) Position Pv02: 49,055,984..49,058,744 <b>Phvul.002G325800</b> Uncharacterized protein LOC100820443 [Glycine max]; IPR006747 (Protein of unknown function DUF599) Position Pv02: 49,062,024..49,064,981 <b>Phvul.002G326000</b> Uncharacterized protein LOC100787767 [Glycine max]; IPR004864 (Late embryogenesis abundant protein, LEA-14) Position Pv02: 49,075,423..49,076,898 <b>Phvul.002G326100</b> 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; IPR005123 (Oxoglutarate/iron-dependent dioxygenase); GO:0005506 (iron ion binding), GO:0016491 (oxidoreductase activity), GO:0031418 (L-ascorbic acid binding), GO:0055114 (oxidation-reduction process) Position Pv02: 49,081,592..49,085,315 <b>Phvul.002G326200</b> Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding) Position Pv02: 49,086,617..49,089,190 (reference: <i>MacQueen et al 2020, [34]</i> ) <b>Phvul.002G326300</b> T-complex protein 1 subunit delta, putative; IPR002423 (Chaperonin Cpn60/TCP-1), IPR027409 (GroEL-like apical domain), IPR027410 (TCP-1-like chaperonin intermediate domain), IPR027413 (GroEL-like equatorial domain); GO:0005524 (ATP binding), GO:0006457 (protein folding), GO:0044267 (cellular protein metabolic process), GO:0051082 (unfolded protein binding) Position Pv02: 49,090,552..49,092,942 <b>Phvul.002G326400</b> ATP-dependent RNA helicase, putative; IPR001650 (Helicase, C-terminal), IPR007502 (Helicase-associated domain), IPR011709 (Domain of unknown function DUF1605), IPR014001 (Helicase, superfamily 1/2, ATP-binding domain), IPR014720 (Double-stranded RNA-binding domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003676 (nucleic acid binding), GO:0004386 (helicase activity), GO:0005524 (ATP binding), GO:0008026 (ATP-dependent helicase activity) Position Pv02: 49,094,816..49,105,934 <b>Phvul.002G326500</b> Glycosyltransferase family 61 protein; IPR007657 (Glycosyltransferase AER61, uncharacterized) Position Pv02: 49,107,649..49,109,932 <b>Phvul.002G326600</b> 1-aminocyclopropane-1-carboxylate oxidase; IPR005123 (Oxoglutarate/iron-dependent dioxygenase), IPR026992 (Non-haem dioxygenase N-terminal domain), IPR027443 (Isopenicillin N synthase-like); GO:0016491 (oxidoreductase activity), GO:0055114 (oxidation-reduction process) Position Pv02: 49,114,990..49,116,625 (reference: <i>Nova-Franco et al 2015, [42]</i> )

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
2	49,136,945	0.603	<p><b>Phvul.002G326300</b>  T-complex protein 1 subunit delta, putative; IPR002423 (Chaperonin Cpn60/TCP-1), IPR027409 (GroEL-like apical domain), IPR027410 (TCP-1-like chaperonin intermediate domain), IPR027413 (GroEL-like equatorial domain); GO:0005524 (ATP binding), GO:0006457 (protein folding), GO:0044267 (cellular protein metabolic process), GO:0051082 (unfolded protein binding)  Position Pv02: 49,090,552..49,092,942</p> <p><b>Phvul.002G326400</b>  ATP-dependent RNA helicase, putative; IPR001650 (Helicase, C-terminal), IPR007502 (Helicase-associated domain), IPR011709 (Domain of unknown function DUF1605), IPR014001 (Helicase, superfamily 1/2, ATP-binding domain), IPR014720 (Double-stranded RNA-binding domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003676 (nucleic acid binding), GO:0004386 (helicase activity), GO:0005524 (ATP binding), GO:0008026 (ATP-dependent helicase activity)  Position Pv02: 49,094,816..49,105,934</p> <p><b>Phvul.002G326500</b>  Glycosyltransferase family 61 protein; IPR007657 (Glycosyltransferase AER61, uncharacterized)  Position Pv02: 49,107,649..49,109,932</p> <p><b>Phvul.002G326600</b>  1-aminocyclopropane-1-carboxylate oxidase; IPR005123 (Oxoglutarate/iron-dependent dioxygenase), IPR026992 (Non-haem dioxygenase N-terminal domain), IPR027443 (Isopenicillin N synthase-like); GO:0016491 (oxidoreductase activity), GO:0055114 (oxidation-reduction process)  Position Pv02: 49,114,990..49,116,625  (reference: Nova-Franco et al 2015, [42])</p> <p><b>Phvul.002G326800</b>  Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)  Position Pv02: 49,125,308..49,129,702</p> <p><b>Phvul.002G326900</b>  uncharacterized protein LOC100785837 [Glycine max]; IPR019448 (EEIG1/EHBP1 N-terminal domain)  Position Pv02: 49,131,317..49,134,457</p> <p><b>Phvul.002G327000</b>  uncharacterized protein LOC100785302 isoform X1 [Glycine max]  Position Pv02: 49,135,836..49,138,843</p> <p><b>Phvul.002G327100</b>  GTP binding Elongation factor Tu family protein; IPR000795 (Elongation factor, GTP-binding domain), IPR009000 (Translation protein, beta-barrel domain), IPR009001 (Translation elongation factor EF1A/initiation factor IF2gamma, C-terminal), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003924 (GTPase activity), GO:0005525 (GTP binding)  Position Pv02: 49,141,959..49,151,857</p> <p><b>Phvul.002G327200</b>  Glycosyltransferase family 29 (sialyltransferase) family protein; IPR001675 (Glycosyl transferase, family 29); GO:0006486 (protein glycosylation), GO:0008373 (sialyltransferase activity)  Position Pv02: 49,154,815..49,159,069</p> <p><b>Phvul.002G327300</b>  pre-mRNA-splicing factor SPF27 homolog [Glycine max]; IPR008409 (Pre-mRNA-splicing factor SPF27); GO:0006397 (mRNA processing)  Position Pv02: 49,161,117..49,164,161</p> <p><b>Phvul.002G327400</b></p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p>peroxin 3; IPR006966 (Peroxin-3); GO:0005779 (integral component of peroxisomal membrane), GO:0007031 (peroxisome organization)  Position Pv02: 49,166,882..49,171,778  <b>Phvul.002G327500</b>  guanine nucleotide-binding protein subunit beta-like protein [Glycine max]; IPR015943 (WD40/YVTN repeat-like-containing domain), IPR020472 (G-protein beta WD-40 repeat); GO:0005515 (protein binding)  Position Pv02: 49,176,943..49,178,861  <b>Phvul.002G327600</b>  WD repeat-containing protein 5-like [Glycine max]; IPR011047 (Quinonprotein alcohol dehydrogenase-like superfamily), IPR015943 (WD40/YVTN repeat-like-containing domain); GO:0005515 (protein binding)  Position Pv02: 49,180,858..49,185,410</p>
2	49,236,682 49,239,546	0.514 0.501	<p><b>Phvul.002G327700</b>  flowering locus protein T; IPR008914 (Phosphatidylethanolamine-binding protein PEBP)  Position Pv02: 49,193,717..49,195,868  <b>Phvul.002G327800</b>  uncharacterized GPI-anchored protein [Glycine max]  Position Pv02: 49,203,167..49,205,710  <b>Phvul.002G327900</b>  uncharacterized protein At5g39865-like [Glycine max]; IPR012336 (Thioredoxin-like fold); GO:0009055 (electron carrier activity), GO:0015035 (protein disulfide oxidoreductase activity), GO:0045454 (cell redox homeostasis)  Position Pv02: 49,209,233..49,210,093  <b>Phvul.002G328000</b>  pentatricopeptide (PPR) repeat-containing protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)  Position Pv02: 49,211,691..49,214,375  <b>Phvul.002G328100</b>  pentatricopeptide (PPR) repeat-containing protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)  Position Pv02: 49,216,072..49,219,019  <b>Phvul.002G328104</b>  Eukaryotic aspartyl protease family protein; IPR021109 (Aspartic peptidase domain)  Position Pv02: 49,221,401..49,223,395  <b>Phvul.002G328300</b>  Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain); GO:0004672 (protein kinase activity), GO:0004707 (MAP kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)  Position Pv02: 49,224,547..49,229,755  (reference: Zuiderveen et al 2016, [36])  <b>Phvul.002G328400</b>  Winged-helix DNA-binding transcription factor family protein, putative isoform 1 n=2 Tax=Theobroma cacao  RepID=UPI00042B60CA; IPR011991 (Winged helix-turn-helix DNA-binding domain), IPR020478 (AT hook-like); GO:0000786 (nucleosome), GO:0003677 (DNA binding), GO:0005634 (nucleus), GO:0006334 (nucleosome assembly)  Position Pv02: 49,236,361..49,239,509  <b>Phvul.002G234300</b>  SMAD/FHA domain-containing protein; IPR008984 (SMAD/FHA domain); GO:0005515 (protein binding)</p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			Position Pv02: 49,253,244..49,254,133 <b>Phvul.002G234304</b> uncharacterized protein LOC100783352 isoform X1 [Glycine max]; IPR026847 (Vacuolar protein sorting-associated protein 13) Position Pv02: 49,268,291..49,303,912
7	638,614 651,195	0.558 0.558	<b>Phvul.007G008200</b> ATP-binding ABC transporter; IPR013525 (ABC-2 type transporter), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0000166 (nucleotide binding), GO:0005524 (ATP binding), GO:0016020 (membrane), GO:0016887 (ATPase activity), GO:0017111 (nucleoside-triphosphatase activity) Position Pv07: 592,430..594,703 <b>Phvul.007G008300</b> U4/U6 small nuclear ribonucleoprotein Prp3-like isoform X5 [Glycine max]; IPR010541 (Domain of unknown function DUF1115), IPR013881 (Pre-mRNA-splicing factor 3), IPR027104 (U4/U6 small nuclear ribonucleoprotein Prp3); GO:0046540 (U4/U6 x U5 tri-snRNP complex) Position Pv07: 601,126..604,961 <b>Phvul.007G008400</b> Peroxidase superfamily protein; IPR010255 (Haem peroxidase); GO:0004601 (peroxidase activity), GO:0006979 (response to oxidative stress), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process) Position Pv07: 607,208..609,910 <b>Phvul.007G008500</b> chalcone-flavanone isomerase family protein; IPR016087 (Chalcone isomerase); GO:0009813 (flavonoid biosynthetic process), GO:0016872 (intramolecular lyase activity), GO:0045430 (chalcone isomerase activity) Position Pv07: 612,607..614,403 (reference: Przysiecka et al 2015, [120]) <b>Phvul.007G008600</b> chalcone-flavanone isomerase family protein; IPR016087 (Chalcone isomerase); GO:0009813 (flavonoid biosynthetic process), GO:0016872 (intramolecular lyase activity), GO:0045430 (chalcone isomerase activity) Position Pv07: 616,509..618,799 <b>Phvul.007G008700</b> hypothetical protein Position Pv07: 623,768..635,464 <b>Phvul.007G008800</b> probable membrane-associated kinase regulator 4-like [Glycine max] Position Pv07: 637,987..639,284 <b>Phvul.007G008900</b> Ankyrin repeat family protein; IPR020683 (Ankyrin repeat-containing domain), IPR026961 (PGG domain), IPR027001 (Caskin/Ankyrin repeat-containing protein); GO:0005515 (protein binding) Position Pv07: 643,030..647,444 <b>Phvul.007G009000</b> Catalytic/ protein phosphatase type 2C/ protein serine/threonine phosphatase n=6 Tax=Panicoidae RepID=B6TEB8_MAIZE; IPR001932 (Protein phosphatase 2C (PP2C)-like domain), IPR015655 (Protein phosphatase 2C); GO:0003824 (catalytic activity), GO:0004722 (protein serine/threonine phosphatase activity), GO:0006470 (protein dephosphorylation) Position Pv07: 649,915..654,480 <b>Phvul.007G009100</b>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p>3-ketoacyl-CoA synthase 19; IPR012392 (Very-long-chain 3-ketoacyl-CoA synthase), IPR016039 (Thiolase-like); GO:0003824 (catalytic activity), GO:0006633 (fatty acid biosynthetic process), GO:0008152 (metabolic process), GO:0008610 (lipid biosynthetic process), GO:0016020 (membrane)            Position Pv07: 657,680..659,317  <b>Phvul.007G009200</b></p> <p>3-ketoacyl-CoA synthase 19; IPR012392 (Very-long-chain 3-ketoacyl-CoA synthase), IPR016039 (Thiolase-like); GO:0003824 (catalytic activity), GO:0006633 (fatty acid biosynthetic process), GO:0008152 (metabolic process), GO:0008610 (lipid biosynthetic process), GO:0016020 (membrane)            Position Pv07: 661,883..663,334  <b>Phvul.007G009300</b></p> <p>Pectate lyase family protein; IPR011050 (Pectin lyase fold/virulence factor), IPR018082 (AmbAllergen)            Position Pv07: 673,746..677,076  <b>Phvul.007G009400</b></p> <p>60S ribosomal L35-like protein; IPR001854 (Ribosomal protein L29); GO:0003735 (structural constituent of ribosome), GO:0005622 (intracellular), GO:0005840 (ribosome), GO:0006412 (translation)            Position Pv07: 678,839..680,686  <b>Phvul.007G009500</b></p> <p>caffeoylshikimate esterase-like isoform X1 [Glycine max]; IPR000073 (Alpha/beta hydrolase fold-1), IPR022742 (Putative lysophospholipase)            Position Pv07: 686,105..689,556  <b>Phvul.007G009600</b></p> <p>splicing factor 3B subunit-like protein; IPR004871 (Cleavage/polyadenylation specificity factor, A subunit, C-terminal), IPR011047 (Quinonprotein alcohol dehydrogenase-like superfamily), IPR015943 (WD40/YVTN repeat-like-containing domain); GO:0003676 (nucleic acid binding), GO:0005515 (protein binding), GO:0005634 (nucleus)            Position Pv07: 690,051..695,717</p>
7	4,174,714 4,209,419	0.505 0.511	<p><b>Phvul.007G050800</b>            cysteine-rich receptor-like protein kinase 25-like [Glycine max]; IPR002902 (Gnk2-homologous domain)            Position Pv07: 4,130,561..4,131,931  <b>Phvul.007G050900</b>            cysteine-rich receptor-like protein kinase 25-like [Glycine max]; IPR002902 (Gnk2-homologous domain)            Position Pv07: 4,133,493..4,135,225  <b>Phvul.007G051000</b>            receptor kinase 2; IPR002902 (Gnk2-homologous domain), IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)            Position Pv07: 4,137,152..4,140,925  <b>Phvul.007G051100</b>            cysteine-rich RLK (RECEPTOR-like protein kinase) 25; IPR002902 (Gnk2-homologous domain), IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)            Position Pv07: 4,146,360..4,150,390  <b>Phvul.007G051200</b></p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p>cysteine-rich RLK (RECEPTOR-like protein kinase) 25; IPR002902 (Gnk2-homologous domain), IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)  Position Pv07: 4,151,954..4,155,957  <b>PhvuI.007G051300</b></p> <p>cysteine-rich RLK (RECEPTOR-like protein kinase) 25; IPR002902 (Gnk2-homologous domain), IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)  Position Pv07: 4,160,841..4,165,305  <b>PhvuI.007G051401</b></p> <p>putative cysteine-rich receptor-like protein kinase 23-like [Glycine max]; IPR002902 (Gnk2-homologous domain)  Position Pv07: 4,180,350..4,182,754  <b>PhvuI.007G051500</b></p> <p>receptor kinase 2; IPR002902 (Gnk2-homologous domain), IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)  Position Pv07: 4,198,645..4,203,117  <b>PhvuI.007G051600</b></p> <p>GDSL esterase/lipase At5g33370-like [Glycine max]; IPR013831 (SGNH hydrolase-type esterase domain), IPR015939 (Fumarate reductase/succinate dehydrogenase flavoprotein-like, C-terminal); GO:0016491 (oxidoreductase activity), GO:0016787 (hydrolase activity), GO:0055114 (oxidation-reduction process)  Position Pv07: 4,209,700..4,213,437  <b>PhvuI.007G051700</b></p> <p>Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family; IPR004864 (Late embryogenesis abundant protein, LEA-14)  Position Pv07: 4,216,491..4,217,619  <b>PhvuI.007G051800</b></p> <p>transmembrane protein, putative  Position Pv07: 4,234,889..4,238,457  <b>PhvuI.007G051900</b></p> <p>pyruvate, phosphate dikinase regulatory protein, putative; IPR005177 (Bifunctional kinase-pyrophosphorylase); GO:0005524 (ATP binding)  Position Pv07: 4,239,917..4,241,740  <b>PhvuI.007G052000</b></p> <p>gibberellin 2-beta-dioxygenase 8-like [Glycine max]; IPR005123 (Oxoglutarate/iron-dependent dioxygenase), IPR026992 (Non-haem dioxygenase N-terminal domain), IPR027443 (Isopenicillin N synthase-like); GO:0016491 (oxidoreductase activity), GO:0055114 (oxidation-reduction process)  Position Pv07: 4,242,844..4,245,600</p>
7	4,702,892	0.597	<p><b>PhvuI.007G055800</b></p> <p>seed linoleate 9S-lipoxygenase; IPR000907 (Lipoxygenase), IPR008976 (Lipase/lipoxygenase, PLAT/LH2), IPR027433 (Lipoxygenase, domain 3); GO:0005506 (iron ion binding), GO:0005515 (protein binding), GO:0016165 (linoleate 13S-lipoxygenase activity), GO:0046872 (metal ion binding), GO:0055114 (oxidation-reduction process)  Position Pv07: 4,657,074..4,663,363  <b>PhvuI.007G055900</b></p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p>photosystem I reaction center subunit II; IPR003685 (Photosystem I PsuD); GO:0009522 (photosystem I), GO:0009538 (photosystem I reaction center), GO:0015979 (photosynthesis)            Position Pv07: 4,666,084..4,667,007  <b>Phvul.007G056000</b>            haloacid dehalogenase-like hydrolase; IPR006439 (HAD hydrolase, subfamily IA), IPR010237 (Pyrimidine 5-nucleotidase), IPR023214 (HAD-like domain); GO:0008152 (metabolic process), GO:0016787 (hydrolase activity)            Position Pv07: 4,670,049..4,672,373  <b>Phvul.007G056100</b>            spermidine synthase 1; IPR001045 (Spermidine/spermine synthases family); GO:0003824 (catalytic activity)            Position Pv07: 4,712,606..4,715,663  <b>Phvul.007G056200</b>            Copper amine oxidase family protein; IPR000269 (Copper amine oxidase); GO:0005507 (copper ion binding), GO:0008131 (primary amine oxidase activity), GO:0009308 (amine metabolic process), GO:0048038 (quinone binding), GO:0055114 (oxidation-reduction process)            Position Pv07: 4,726,727..4,730,374  <b>Phvul.007G056300</b>            hypothetical protein            Position Pv07: 4,738,183..4,738,785  <b>Phvul.007G056400</b>            Copper amine oxidase family protein; IPR000269 (Copper amine oxidase); GO:0005507 (copper ion binding), GO:0008131 (primary amine oxidase activity), GO:0009308 (amine metabolic process), GO:0048038 (quinone binding), GO:0055114 (oxidation-reduction process)            Position Pv07: 4,742,026..4,744,904</p>
7	38,944,990	0.574	<p><b>Phvul.007G267900</b>            BEL1-like homeodomain protein 1-like isoform X2 [Glycine max]; IPR006563 (POX domain), IPR009057 (Homeodomain-like); GO:0003677 (DNA binding), GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)            Position Pv07: 38,901,727..38,906,220  <b>Phvul.007G268000</b>            NC domain-containing protein-related; IPR000064 (Endopeptidase, NLPC/P60 domain), IPR007053 (LRAT-like domain)            Position Pv07: 38,922,613..38,924,853  <b>Phvul.007G268100</b>            TGF beta-inducible nuclear protein 1 n=21 Tax=Endopterygota RepID=Q1W759_BOMMO; IPR022309 (Ribosomal protein S8e/ribosomal biogenesis NSA2)            Position Pv07: 38,926,720..38,929,382  <b>Phvul.007G268200</b>            Protein kinase superfamily protein; IPR001611 (Leucine-rich repeat), IPR003591 (Leucine-rich repeat, typical subtype), IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005515 (protein binding), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)            Position Pv07: 38,944,235..38,948,252  <i>(reference: Duwadi et al 2018, [43])</i>  <b>Phvul.007G268300</b>            uncharacterized protein LOC100790193 [Glycine max]</p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			Position Pv07: 38,950,910..38,952,376 <b>Phvul.007G268400</b> Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family Position Pv07: 38,960,563..38,961,721 <b>Phvul.007G268500</b> uncharacterized protein [Glycine max] Position Pv07: 38970861..38,971,710 <b>Phvul.007G268600</b> Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family; IPR004864 (Late embryogenesis abundant protein, LEA-14) Position Pv07: 38,981,117..38,982,594 <b>Phvul.007G268700</b> protein YLS9 [Glycine max]; IPR004864 (Late embryogenesis abundant protein, LEA-14) Position Pv07: 38,985,430..38,986,354 <b>Phvul.007G268800</b> protein YLS9-like [Glycine max] Position Pv07: 38,986,531..38,987,745
9	5,546,515	0.656	<b>Phvul.009G022600</b> Ribosomal protein S25 family protein; IPR004977 (Ribosomal protein S25) Position Pv09: 5,527,580..5,529,548 <b>Phvul.009G022700</b> DUF241 domain protein; IPR004320 (Protein of unknown function DUF241, plant) Position Pv09: 5,531,980..5,533,006 <b>Phvul.009G022800</b> uncharacterized protein LOC100809365 [Glycine max]; IPR004320 (Protein of unknown function DUF241, plant) Position Pv09: 5,536,083..5,537,006 <b>Phvul.009G022900</b> DUF241 domain protein; IPR004320 (Protein of unknown function DUF241, plant) Position Pv09: 5,538,931..5,539,818 <b>Phvul.009G023000</b> uncharacterized protein LOC100809365 [Glycine max]; IPR004320 (Protein of unknown function DUF241, plant) Position Pv09:5,549,891..5,550,781 <b>Phvul.009G023100</b> Glucose-1-phosphate adenylyltransferase family protein; IPR011831 (Glucose-1-phosphate adenylyltransferase); GO:0005978 (glycogen biosynthetic process), GO:0008878 (glucose-1-phosphate adenylyltransferase activity), GO:0009058 (biosynthetic process), GO:0016779 (nucleotidyltransferase activity) Position Pv09: 5,557,792..5,562,086 <b>Phvul.009G023200</b> SNF1-related protein kinase regulatory subunit gamma 1; IPR000644 (CBS domain); GO:0030554 (adenyl nucleotide binding) Position Pv09: 5,562,406..5,564,703 <b>Phvul.009G023300</b> Rer1 family protein; IPR004932 (Retrieval of early ER protein Rer1); GO:0016021 (integral component of membrane) Position Pv09: 5,567,146..5,569,659

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
9	6,874,190	0.618	<b>PhvuI.009G029800</b> trehalose-6-phosphate phosphatase; IPR006379 (HAD-superfamily hydrolase, subfamily IIB), IPR023214 (HAD-like domain); GO:0003824 (catalytic activity), GO:0005992 (trehalose biosynthetic process), GO:0008152 (metabolic process) Position Pv09: 6,838,960..6,841,520
			<b>PhvuI.009G029900</b> unknown protein; LOCATED IN: chloroplast Position Pv09: 6,878,373..6,880,001
			<b>PhvuI.009G030000</b> TPR repeat-containing thioredoxin TTL1-like [Glycine max]; IPR011990 (Tetratricopeptide-like helical), IPR012336 (Thioredoxin-like fold); GO:0005515 (protein binding), GO:0045454 (cell redox homeostasis) Position Pv09: 6,896,081..6,899,881
9	7,701,485	0.546	<b>PhvuI.009G034800</b> Target SNARE coiled-coil domain protein; IPR000727 (Target SNARE coiled-coil domain); GO:0005515 (protein binding) Position Pv09: 7,655,743..7,658,686
			<b>PhvuI.009G034900</b> ubiquitin-protein ligase 1; IPR000569 (HECT), IPR003903 (Ubiquitin interacting motif), IPR009060 (UBA-like), IPR010309 (E3 ubiquitin ligase, domain of unknown function DUF908), IPR016024 (Armadillo-type fold), IPR025527 (Domain of unknown function DUF4414); GO:0004842 (ubiquitin-protein ligase activity), GO:0005488 (binding), GO:0005515 (protein binding) Position Pv09: 7,665,586..7,683,468
			<b>PhvuI.009G035000</b> CRT (chloroquine-resistance transporter)-like transporter 2 Position Pv09: 7,690,678..7,698,131
			<b>PhvuI.009G035100</b> neuroguidin-like isoform X3 [Glycine max]; IPR007146 (Sas10/Utp3/C1D) Position Pv09: 7,703,487..7,706,835
			<b>PhvuI.009G035200</b> Golgi-body localisation protein domain ; RNA pol II promoter Fmp27 protein domain; IPR019441 (FMP27, GFWDK domain), IPR019443 (FMP27, C-terminal) Position Pv09: 7,722,501..7,749,619
9	7,807,599 7,872,442	0.677 0.589	<b>PhvuI.009G035400</b> GATA transcription factor 15; IPR010399 (Tify), IPR010402 (CCT domain), IPR013088 (Zinc finger, NHR/GATA-type); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0005515 (protein binding), GO:0008270 (zinc ion binding), GO:0043565 (sequence-specific DNA binding) Position Pv09: 7,764,191..7,768,361
			<b>PhvuI.009G035500</b> BEACH domain-containing protein lvsC-like isoform X4 [Glycine max]; IPR000409 (BEACH domain), IPR008985 (Concanavalin A-like lectin/glucanases superfamily), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup), IPR015943 (WD40/YVTN repeat-like-containing domain), IPR016024 (Armadillo-type fold), IPR023362 (PH-BEACH domain); GO:0005488 (binding), GO:0005515 (protein binding) Position Pv09: 7,785,251..7,852,293
			<b>PhvuI.009G035700</b> Unknown protein Position Pv09: 7,862,516..7,863,493
			<b>PhvuI.009G035800</b>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p>Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain), IPR028324 (Serine/threonine-protein kinase CTR1/EDR1); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)  Position Pv09: 7,865,582..7,873,144  <b>Phvul.009G035900</b></p> <p>Tetratricopeptide repeat (TPR)-like superfamily protein; IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)  Position Pv09: 7,877,920..7,879,060  <b>Phvul.009G036000</b></p> <p>RAB geranylgeranyl transferase alpha subunit 1; IPR001611 (Leucine-rich repeat), IPR002088 (Protein prenyltransferase, alpha subunit), IPR025875 (Leucine rich repeat 4); GO:0005515 (protein binding), GO:0008318 (protein prenyltransferase activity), GO:0018342 (protein prenylation)  Position Pv09: 7,881,919..7,886,744  <b>Phvul.009G036100</b></p> <p>double-stranded-RNA-binding protein 4; IPR011907 (Ribonuclease III); GO:0003723 (RNA binding), GO:0004525 (ribonuclease III activity), GO:0016075 (rRNA catabolic process)  Position Pv09: 7,888,094..7,891,100  <b>Phvul.009G036200</b></p> <p>NADPH-cytochrome P450 family 2 reductase; IPR001094 (Flavodoxin), IPR023173 (NADPH-cytochrome p450 reductase, FAD-binding, alpha-helical domain-3); GO:0003958 (NADPH-hemoprotein reductase activity), GO:0005506 (iron ion binding), GO:0010181 (FMN binding), GO:0016491 (oxidoreductase activity), GO:0055114 (oxidation-reduction process)  Position Pv09: 7,905,486..7,913,740  <b>Phvul.009G036300</b></p> <p>myb-like DNA-binding domain protein; IPR009057 (Homeodomain-like); GO:0003677 (DNA binding), GO:0003682 (chromatin binding)  Position Pv09: 7,916,029..7,924,420</p>
9	13,540,452 13,555,680	0.536 0.536	<p><b>Phvul.009G081700</b>  Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)  Position Pv09: 13,507,606..13,510,042  <b>Phvul.009G081800</b>  cytokinin oxidase/dehydrogenase 1; IPR016164 (FAD-linked oxidase-like, C-terminal), IPR016166 (FAD-binding, type 2), IPR016170 (Vanillyl-alcohol oxidase/Cytokinin dehydrogenase C-terminal domain); GO:0003824 (catalytic activity), GO:0008762 (UDP-N-acetylmuramate dehydrogenase activity), GO:0009690 (cytokinin metabolic process), GO:0016491 (oxidoreductase activity), GO:0019139 (cytokinin dehydrogenase activity), GO:0050660 (flavin adenine dinucleotide binding), GO:0055114 (oxidation-reduction process)  Position Pv09: 13,517,847..13,521,865  <i>(reference: Recchia et al 2018, [45])</i>  <b>Phvul.009G081900</b>  F-box/LRR-repeat protein 15-like [Glycine max]; IPR001810 (F-box domain), IPR006553 (Leucine-rich repeat, cysteine-containing subtype); GO:0005515 (protein binding)  Position Pv09: 13,539,823..13,545,847  <b>Phvul.009G082000</b>  uncharacterized protein LOC100787776 [Glycine max]  Position Pv09: 13,547,643..13,550,505</p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p><b>Phvul.009G082100</b>  ferredoxin-related; IPR014044 (CAP domain)  Position Pv09: 13,553,843..13,556,650</p> <p><b>Phvul.009G082200</b>  uncharacterized protein LOC102659395 isoform X3 [Glycine max]  Position Pv09: 13,557,884..13,558,507</p> <p><b>Phvul.009G082300</b>  YGGT family protein; IPR003425 (Uncharacterized protein family Ycf19); GO:0016020 (membrane)  Position Pv09: 13,560,871..13,562,756</p> <p><b>Phvul.009G082400</b>  P-ATPase family transporter: copper ion; heavy metal transporting P-type ATPase-like protein n=1 Tax=Ostreococcus lucimarinus (strain CCE9901) ReplID=A4S4X5_OSTLU; IPR001757 (Cation-transporting P-type ATPase), IPR023214 (HAD-like domain); GO:0000166 (nucleotide binding), GO:0006812 (cation transport), GO:0016021 (integral component of membrane), GO:0019829 (cation-transporting ATPase activity), GO:0030001 (metal ion transport), GO:0046872 (metal ion binding)  Position Pv09: 13,563,555..13,573,390</p> <p><b>Phvul.009G082466</b>  receptor-like protein kinase 2; IPR001611 (Leucine-rich repeat), IPR003591 (Leucine-rich repeat, typical subtype), IPR013210 (Leucine-rich repeat-containing N-terminal, type 2); GO:0005515 (protein binding)  Position Pv09: 13,588,429..13,593,654</p> <p><b>Phvul.009G082532</b>  ATP binding/protein serine/threonine kinase [Glycine max]; IPR011009 (Protein kinase-like domain); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)  Position Pv09: 13,593,710..13,595,345</p> <p><b>Phvul.009G082600</b>  myosin, putative; IPR000048 (IQ motif, EF-hand binding site), IPR001609 (Myosin head, motor domain), IPR002710 (Dilute), IPR004009 (Myosin, N-terminal, SH3-like), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003774 (motor activity), GO:0005515 (protein binding), GO:0005524 (ATP binding), GO:0016459 (myosin complex)  Position Pv09: 13,597,488..13,621,793</p>
9	13,667,257	0.556	<p><b>Phvul.009G082800</b>  endoglucanase 16-like [Glycine max]; IPR001701 (Glycoside hydrolase, family 9), IPR008928 (Six-hairpin glycosidase-like); GO:0003824 (catalytic activity), GO:0005975 (carbohydrate metabolic process)  Position Pv09: 13,623,739..13,626,641</p> <p><b>Phvul.009G082900</b>  guanine nucleotide-binding protein alpha-1 subunit isoform X3 [Glycine max]; IPR001019 (Guanine nucleotide binding protein (G-protein), alpha subunit), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003924 (GTPase activity), GO:0004871 (signal transducer activity), GO:0005525 (GTP binding), GO:0006184 (GTP catabolic process), GO:0007165 (signal transduction), GO:0007186 (G-protein coupled receptor signaling pathway), GO:0019001 (guanyl nucleotide binding), GO:0031683 (G-protein beta/gamma-subunit complex binding)  Position Pv09: 13,627,456..13,632,983</p> <p><b>Phvul.009G083000</b>  Unknown protein  Position Pv09: 13,632,025..13,632,978</p> <p><b>Phvul.009G083100</b>  alpha/beta-Hydrolases superfamily protein; IPR000073 (Alpha/beta hydrolase fold-1)</p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p>Position Pv09: 13,649,605..13,651,144</p> <p><b>Phvul.009G083200</b> alpha/beta-Hydrolases superfamily protein; IPR000073 (Alpha/beta hydrolase fold-1)</p> <p>Position Pv09: 13,652,896..13,655,387</p> <p><b>Phvul.009G083300</b> Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)</p> <p>Position Pv09: 13,660,303..13,662,564</p> <p><b>Phvul.009G083400</b> copper transporter 5; IPR007274 (Ctr copper transporter); GO:0005375 (copper ion transmembrane transporter activity), GO:0016021 (integral component of membrane), GO:0035434 (copper ion transmembrane transport)</p> <p>Position Pv09: 13,663,149..13,664,162</p> <p><b>Phvul.009G083500</b> uncharacterized protein LOC100813662 [Glycine max]; IPR012876 (Protein of unknown function DUF1677, plant)</p> <p>Position Pv09: 13,666,413..13,667,873</p> <p><b>Phvul.009G083600</b> Smr (small MutS-related) domain protein; IPR002625 (Smr protein/MutS2 C-terminal), IPR013899 (Domain of unknown function DUF1771)</p> <p>Position Pv09: 13,677,996..13,684,415</p> <p><b>Phvul.009G083700</b> Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)</p> <p>Position Pv09: 13,686,702..13,690,344</p> <p><b>Phvul.009G083800</b> DEAD-box ATP-dependent RNA helicase 21-like isoform 1 [Glycine max]; IPR000999 (Ribonuclease III domain), IPR001650 (Helicase, C-terminal), IPR003100 (Argonaute/Dicer protein, PAZ domain), IPR005034 (Dicer dimerisation domain), IPR014001 (Helicase, superfamily 1/2, ATP-binding domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003676 (nucleic acid binding), GO:0003723 (RNA binding), GO:0004386 (helicase activity), GO:0004525 (ribonuclease III activity), GO:0005515 (protein binding), GO:0005524 (ATP binding), GO:0006396 (RNA processing), GO:0008026 (ATP-dependent helicase activity)</p> <p>Position Pv09: 13,693,775..13,712,211</p>
9	30,555,204	0.543	<p><b>Phvul.009G200800</b> uncharacterized protein LOC100775409 [Glycine max]</p> <p>Position Pv09: 30,528,511..30,529,161</p> <p><b>Phvul.009G200900</b> protein CHUP1, chloroplastic-like [Glycine max]</p> <p>Position Pv09: 30,538,426..30,540,922</p> <p><b>Phvul.009G201000</b> Adenine nucleotide alpha hydrolases-like superfamily protein; IPR014729 (Rossmann-like alpha/beta/alpha sandwich fold); GO:0006950 (response to stress)</p> <p>Position Pv09: 30,543,139..30,545,216</p> <p><b>Phvul.009G201100</b></p>

Chr	Position (Mbp)	F <sub>ST</sub> value	Candidate genes annotated
			<p>Eukaryotic aspartyl protease family protein; IPR001461 (Aspartic peptidase), IPR021109 (Aspartic peptidase domain); GO:0004190 (aspartic-type endopeptidase activity), GO:0006508 (proteolysis)            Position Pv09: 30,558,554..30,559,720  <b>Phvul.009G201200</b>            DEAD-box ATP-dependent RNA helicase-like protein; IPR001650 (Helicase, C-terminal), IPR014001 (Helicase, superfamily 1/2, ATP-binding domain), IPR014014 (RNA helicase, DEAD-box type, Q motif), IPR025313 (Domain of unknown function DUF4217), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003676 (nucleic acid binding), GO:0004386 (helicase activity), GO:0005524 (ATP binding), GO:0008026 (ATP-dependent helicase activity)            Position Pv09: 30,560,152..30,566,190  <b>Phvul.009G201300</b>            ATP binding microtubule motor family protein n=1 Tax=Theobroma cacao RepID=UPI00042B89EE; IPR001752 (Kinesin, motor domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase), IPR027640 (Kinesin-like protein); GO:0003777 (microtubule motor activity), GO:0005524 (ATP binding), GO:0005871 (kinesin complex), GO:0007018 (microtubule-based movement), GO:0008017 (microtubule binding)            Position Pv09: 30,591,988..30,596,555</p>
11	52,372,623	0.553	<b>Phvul.011G207000</b>
	52,413,349	0.506	Pentatricopeptide repeat (PPR-like) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like
	52,419,673	0.553	helic); GO:0005515 (protein binding)
	52,445,931	0.553	Position Pv11: 52,348,900..52,357,148
	52,474,138	0.553	<b>Phvul.011G206900</b>
	52,487,507	0.553	<p>DNA-directed RNA polymerase; IPR006592 (RNA polymerase, N-terminal), IPR007066 (RNA polymerase Rpb1, domain 3), IPR007080 (RNA polymerase Rpb1, domain 1), IPR007081 (RNA polymerase Rpb1, domain 5), IPR015801 (Copper amine oxidase, N2/N3-terminal), IPR021602 (Protein of unknown function DUF3223); GO:0003677 (DNA binding), GO:0003899 (DNA-directed RNA polymerase activity), GO:0005507 (copper ion binding), GO:0009308 (amine metabolic process), GO:0048038 (quinone binding)            Position Pv11: 52,366,172..52,383,595  <b>Phvul.011G206800</b>            beta-galactosidase-like [Glycine max]; IPR006101 (Glycoside hydrolase, family 2), IPR008979 (Galactose-binding domain-like), IPR011013 (Galactose mutarotase-like domain), IPR013812 (Glycoside hydrolase, family 2/20, immunoglobulin-like beta-sandwich domain), IPR017853 (Glycoside hydrolase, superfamily), IPR023230 (Glycoside hydrolase, family 2, conserved site), IPR023232 (Glycoside hydrolase, family 2, active site); GO:0003824 (catalytic activity), GO:0004565 (beta-galactosidase activity), GO:0005975 (carbohydrate metabolic process), GO:0009341 (beta-galactosidase complex), GO:0030246 (carbohydrate binding)            Position Pv11: 52,386,468..52,395,605  <b>Phvul.011G206700</b>            PREFOLDIN 1; IPR009053 (Prefoldin); GO:0006457 (protein folding), GO:0016272 (prefoldin complex), GO:0051082 (unfolded protein binding)            Position Pv11: 52,404,482..52,407,154  <b>Phvul.011G206600</b>            uncharacterized protein LOC100777329 isoform X1 [Glycine max]            Position Pv11: 52,408,489..52,410,925  <b>Phvul.011G206500</b>            Amidohydrolase family; IPR011059 (Metal-dependent hydrolase, composite domain), IPR013108 (Amidohydrolase 3)            Position Pv11: 52,411,332..52,421,199  <b>Phvul.011G206400</b></p>

Chr	Position (Mbp)	$F_{ST}$ value	Candidate genes annotated
			<p>protein LATERAL ROOT PRIMORDIUM 1-like isoform X1 [Glycine max]; IPR007818 (Protein of unknown function DUF702)            Position Pv1 1: 52,434,964..52,440,240</p> <p><b>Phvul.011G206300</b>            Phosphatidylinositol-4-phosphate 5-kinase family protein; IPR023610 (Phosphatidylinositol-4-phosphate 5-kinase), IPR027483 (Phosphatidylinositol-4-phosphate 5-kinase, C-terminal), IPR027484 (Phosphatidylinositol-4-phosphate 5-kinase, N-terminal domain); GO:0005524 (ATP binding), GO:0016307 (phosphatidylinositol phosphate kinase activity), GO:0016308 (1-phosphatidylinositol-4-phosphate 5-kinase activity), GO:0046488 (phosphatidylinositol metabolic process)            Position Pv1 1: 52,442,526..52,447,566</p> <p><b>Phvul.011G206200</b>            unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system            Position Pv1 1: 52,485,285..52,489,990</p> <p><b>Phvul.011G206100</b>            Eukaryotic aspartyl protease family protein; IPR001461 (Aspartic peptidase), IPR021109 (Aspartic peptidase domain); GO:0004190 (aspartic-type endopeptidase activity), GO:0006508 (proteolysis)            Position Pv1 1: 52,498,725..52,500,499</p>

**Table S4.3.** F-test of fixed and Pearson's  $\chi^2$  test of random effects in the combined GLIMMIX mixed model analysis of the Honduran panel genotypes tested in multiple field locations in Ontario, Canada and Yorito, Honduras, 2014-2015.

	N derived from the atmosphere		Carbon discrimination ( $\Delta$ )		Flowering		Yield		Hundred seed weight	
	(%)		(‰)		(days)		(kg ha <sup>-1</sup> )		(g)	
<b>Fixed effect†</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>
<b>Genotype (G)</b>	3.09	<.0001	4.18	<.0001	11.85	<.0001	1.99	0.0008	10.37	<.0001
<b>Environment (E)</b>	49.31	0.0051	54.87	0.0043	349.67	0.0003	180.93	0.0007	13.74	0.0657
<b>G x E</b>	3.79	<.0001	2.34	<.0001	3.52	<.0001	2.18	<.0001	2.20	0.0010
<b>Random effects‡</b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>
<b>BLOC</b>	0.41	0.5207	6.26	0.0123	6.38	0.0115	0.77	0.3802	0.62	0.4321
<b>IBLK(BLOC)</b>	12.86	0.0003	0.87	0.3506	1.13	0.2880	7.30	0.0069	0.52	0.4692
<b>Residual</b>										

DTF, and HSW were measured at Elora 2014 and 2015 only.

† Genotype (G), Environment (E), and G x E interaction

‡ Block (BLOC), and incomplete block (IBLK)

**Table S4.4.** F-test of fixed effect of genotype overall and by breeding history category, and the Pearson's  $\chi^2$  test of random effects in the GLIMMIX analysis of 63 genotypes tested at Elora, Ontario, Canada, 2014.

	N derived from the atmosphere		Carbon discrimination ( $\Delta$ )		Flowering		Maturity		Yield		Hundred seed weight	
	(%)		(‰)		(days)		(days)		(kg ha <sup>-1</sup> )		(g)	
<b>Fixed effect†</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>
<b>Genotype</b>	2.52	0.0048	1.8	0.0496	5.89	<.0001	8.08	<.0001	10.61	<.0001	5.44	<.0001
<b>Random effects‡</b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>
<b>BLOC</b>	0.0	1.0	0.22	0.3192	2.4	0.1211	0.0	1.0	0.0	1.0	0.0	1.0
<b>IBLK(BLOC)</b>	2.73	0.0983	0.41	0.2620	0.0	1.0	0.0	0.9758	5.79	0.0161	0.06	0.7990
<b>Residual</b>												

† Genotype effects overall and by breeding history subcategory

‡ Block (BLOC), and incomplete block (IBLK)

**Table S4.5.** F-test of fixed effect of genotype overall and by breeding history category, and the Pearson's  $\chi^2$  test of random effects in the GLIMMIX analysis of 63 genotypes tested at Elora, Ontario, Canada, 2015.

	N derived from the atmosphere		Carbon discrimination ( $\Delta$ )		Flowering		Maturity		Yield		Hundred seed weight	
	(%)		(‰)		(days)		(days)		(kg ha <sup>-1</sup> )		(g)	
<b>Fixed effect†</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>
<b>Genotype</b>	1.61	0.0449	2.38	0.0012	3.99	<.0001	3.7	<.0001	3.69	<.0001	7.34	<.0001
<b>Random effects‡</b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>
<b>BLOC</b>	0.0	1.0	3.83	0.0251	0.0	1.0	0.0	1.0	0.0	1.0	0.53	0.4656
<b>IBLK(BLOC)</b>	5.76	0.0164	0.83	0.1811	0.0	1.0	8.09	0.0045	9.99	0.0016	0.0	1.0
<b>Residual</b>												

† Genotype effects overall and by breeding history subcategory

‡ Block (BLOC), and incomplete block (IBLK)

**Table S4.6.** F-test of fixed effect of genotype overall and by breeding history category, and the Pearson's  $\chi^2$  test of random effects in the GLIMMIX analysis of 63 genotypes tested at Yorito, Honduras, 2014-2015.

	N derived from the atmosphere		Carbon discrimination ( $\Delta$ )		Yield		Height	
	(%)		(‰)		(kg ha <sup>-1</sup> )		(cm)	
<b>Fixed effect†</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>	<b>F-test</b>	<b>P-value</b>
<b>Genotype</b>	18.82	<.0001	7.19	<.0001	0.94	0.5936	5.42	<.0001
<b>Random effects‡</b>	$\chi^2$	Pr> $\chi^2$	$\chi^2$	Pr> $\chi^2$	$\chi^2$	Pr> $\chi^2$	$\chi^2$	Pr> $\chi^2$
<b>BLOC</b>	1.19	0.1375	1.12	0.2890	0.0	0.4919	0.0	1.0
<b>IBLK(BLOC)</b>	0.14	0.3531	0.0	1.0	0.05	0.4117	2.85	0.0911
<b>Residual</b>	186.47	<.0001	.	.	.	.	.	.

† Genotype effects overall and by breeding history subcategory

‡ Block (BLOC), and incomplete block (IBLK)

**Table S4.7.** F-test of fixed effects and Pearson's  $\chi^2$  test of random effects in the GLIMMIX mixed-model analysis of leaf chlorophyll content (SPAD) at early vegetative (SPAD1) and reproductive (SPAD2) stages in the Honduran panel overall (Combined) and at Elora in 2014 and 2015.

	Combined				Elora 2014				Elora 2015			
	SPAD1		SPAD2		SPAD1		SPAD2		SPAD1		SPAD2	
<b>Fixed effect†</b>	<b>F-test</b>	<b>P-value</b>										
<b>Genotype (G)</b>	6.19	<.0001	6.44	<.0001	2.59	<.0001	3.66	<.0001	10.22	<.0001	8.39	<.0001
<b>Environment (E)</b>	32.21	<.0001	0.98	0.3237	-	-	-	-	-	-	-	-
<b>G x E</b>	3.11	<.0001	4.59	<.0001	-	-	-	-	-	-	-	-
<b>Random effects‡</b>	<b><math>\chi^2</math></b>	<b>Pr&gt; <math>\chi^2</math></b>										
<b>BLOC</b>	.	.	3.72	0.0539	1.60	0.2055	0.21	0.6461	0.95	0.3306	1.82	0.1767
<b>IBLK(BLOC)</b>	.	.	3.70	0.0543	2.80	0.0941	0.0	0.9619	0.95	0.3306	1.82	0.1767
<b>Residual</b>	3.76	0.0526	.	.	.	.	.	.	.	.	.	.

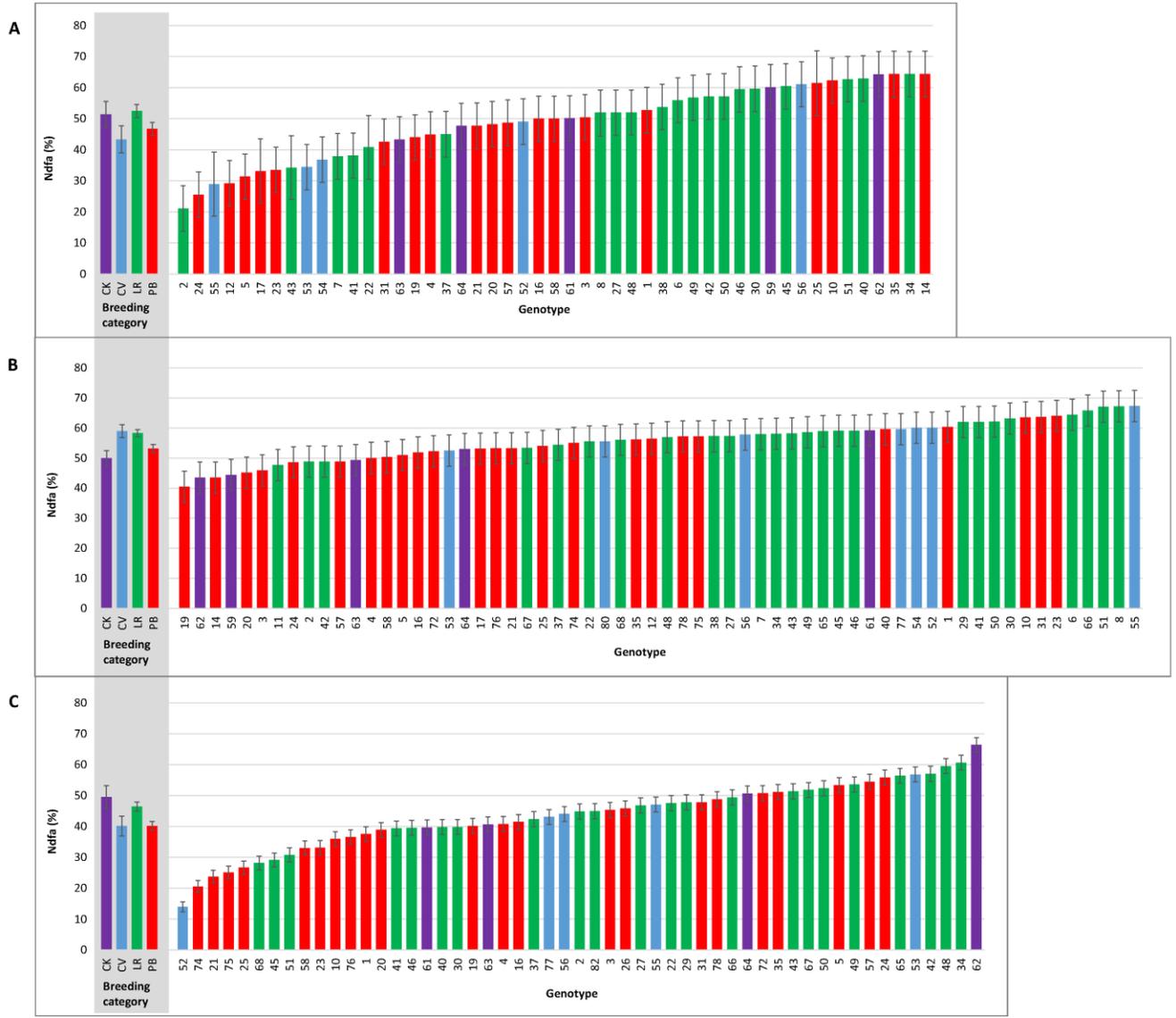
SPAD1 and SPAD2 were measured at Elora 2014 and 2015 only.

† Genotype (G), Environment (E), and G x E interaction

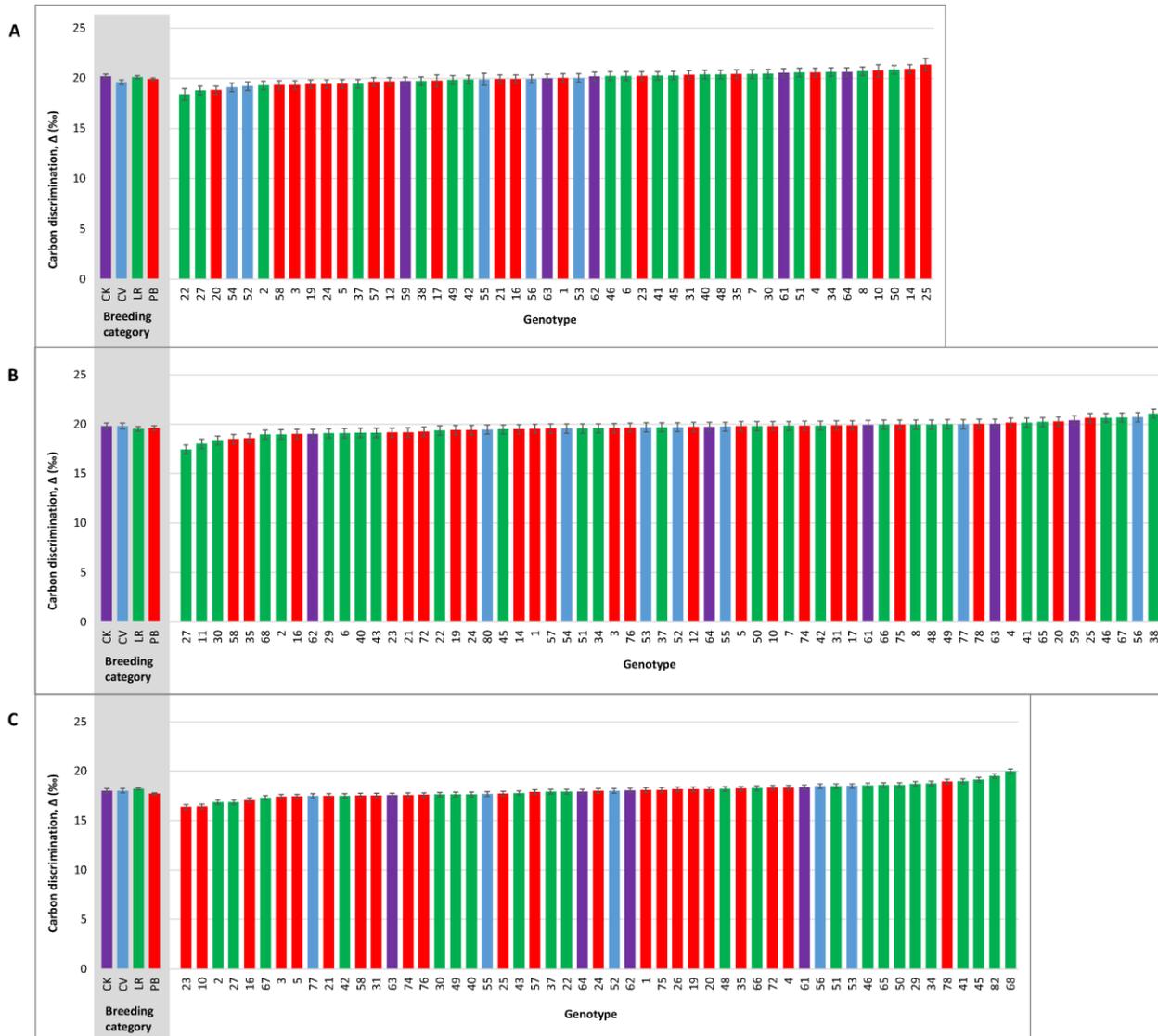
‡ Block (BLOC), and incomplete block (IBLK)

**Table S4.8.** Phenotypic ( $r_p$ ) correlations among %Ndfa and other traits estimated in the Honduran panel grown in 3 locations in 2014-2015. Values shown for traits within environments are significantly correlated at the 5% level. The number of genotypes analyzed in each correlation is presented in brackets for each significant correlation.

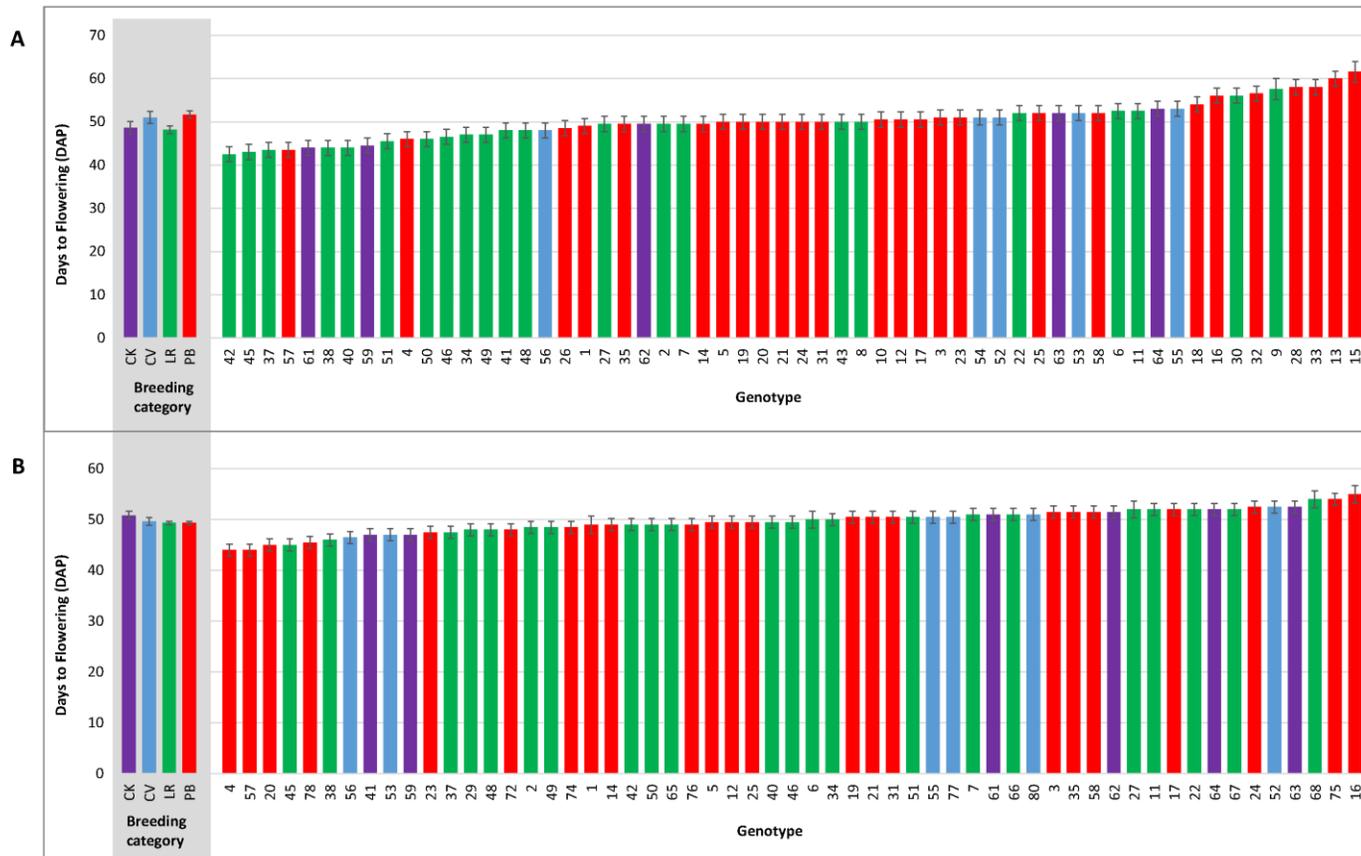
Trait Location	%Ndfa E14	%Ndfa E15	%Ndfa YOR	SPAD2 E14	SPAD2 E15	DTF E14	DTF E15	HSW E14	HSW E15	$\Delta C$ E14	$\Delta C$ E15	$\Delta C$ YOR	Yield E14	Yield E15
%Ndfa E15														
%Ndfa YOR														
SPAD2 E14														
SPAD2 E15														
DTF E14	-0.30704 (49)													
DTF E15														
HSW E14				0.36357 (49)										
HSW E15					0.43931 (62)									
$\Delta C$ E14	0.45399 (48)													
$\Delta C$ E15							-0.36797 (57)							
$\Delta C$ YOR														
Yield E14	0.3803 (35)					-0.48231 (35)								
Yield E15								0.48942 (62)	-0.32591 (62)					
Yield YOR														



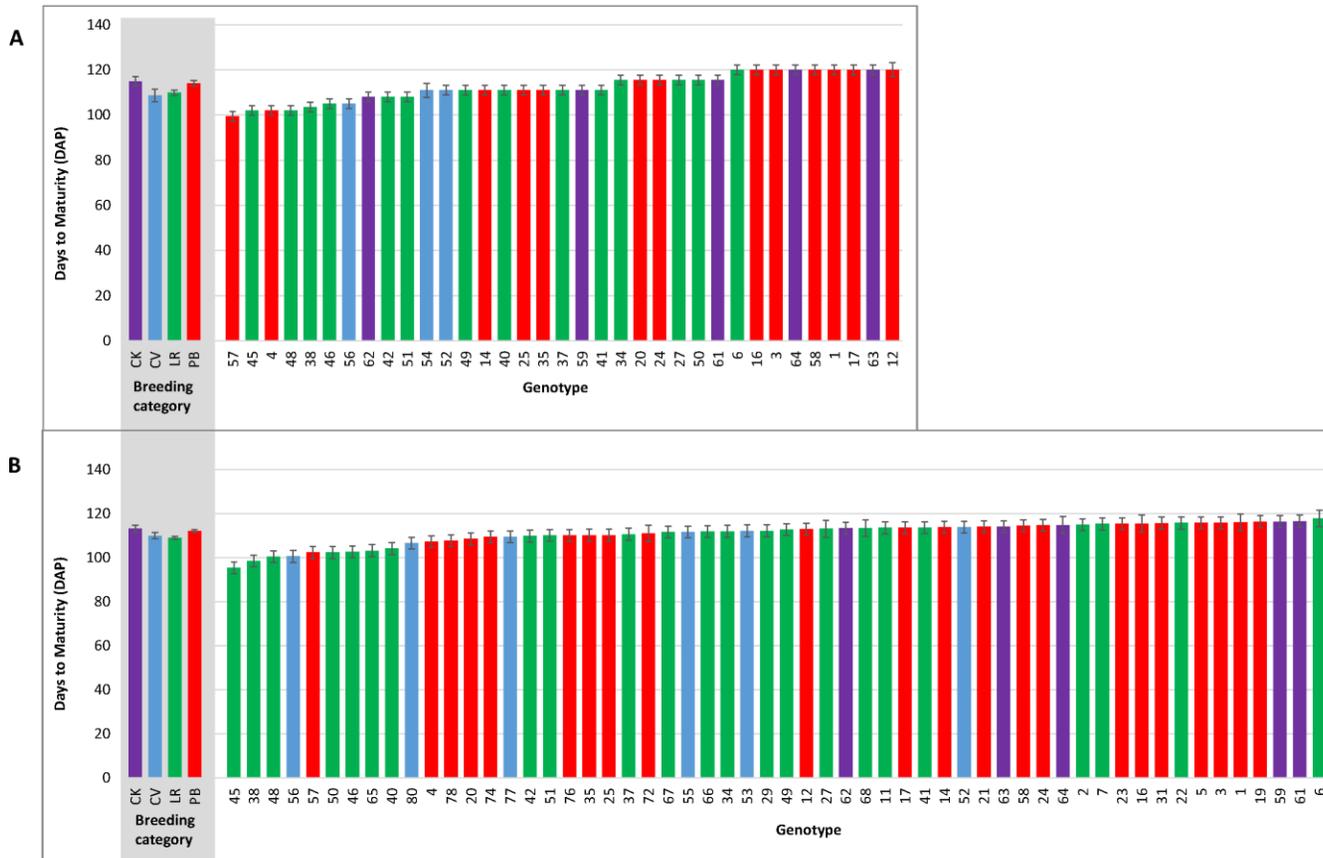
**Figure S4.1.** Histograms of nitrogen derived from the atmosphere (%) values of genotypes comprising the HON panel tested at three field locations from 2014-2015. A. Elora 2014, B. Elora 2015 and C. Yorito. Breeding history category averages with standard errors are presented, followed by individual genotype LSmeans with standard errors. North American check genotypes (CK; purple), Honduran conventional genotypes (CV; blue), landraces (LR; green), and PPB varieties (PB; red). Genotype numbers correspond to those listed in Tables 4.1-4.3.



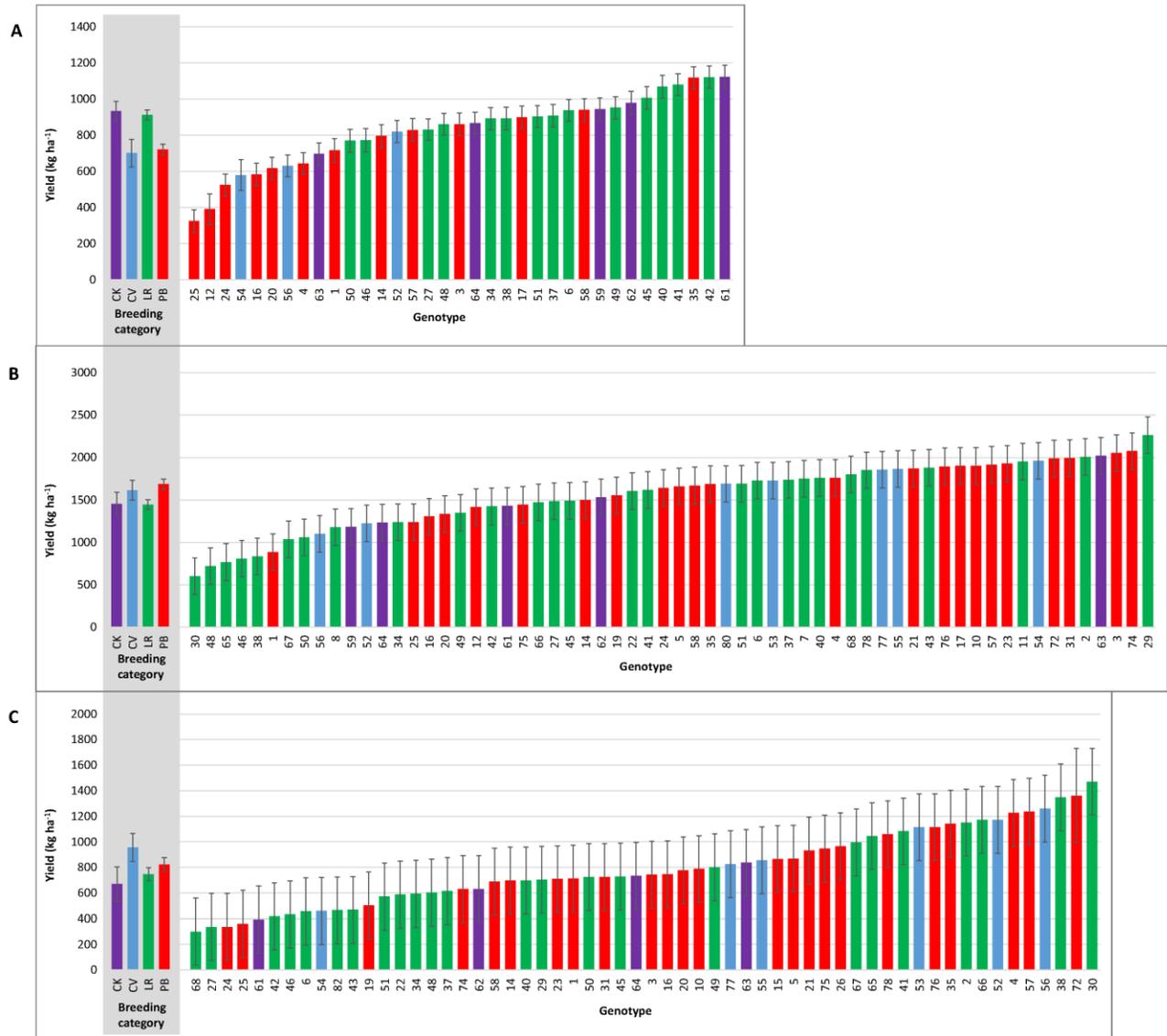
**Figure S4.2.** Histograms of carbon discrimination ( $\Delta$ ) values of genotypes comprising the HON panel tested at three field locations from 2014-2015. A. Elora 2014, B. Elora 2015 and C. Yorito. Breeding history category averages with standard errors are presented, followed by individual genotype LSmeans with standard errors. North American check genotypes (CK; purple), Honduran conventional genotypes (CV; blue), landraces (LR; green), and PPB varieties (PB; red). Genotype numbers correspond to those listed in Tables 4.1-4.3.



**Figure S4.3.** Histograms of days to flowering of genotypes comprising the HON panel tested at two field locations from 2014-2015. A. Elora 2014, B. Elora 2015. Breeding history category averages with standard errors are presented, followed by individual genotype LSmeans with standard errors. North American check genotypes (CK; purple), Honduran conventional genotypes (CV; blue), landraces (LR; green), and PPB varieties (PB; red). Genotype numbers correspond to those listed in Tables 4.1-4.3.



**Figure S4.4.** Histograms of days to maturity of genotypes comprising the HON panel tested at two field locations from 2014-2015. A. Elora 2014, B. Elora 2015. Breeding history category averages with standard errors are presented, followed by individual genotype LSmeans with standard errors. North American check genotypes (CK; purple), Honduran conventional genotypes (CV; blue), landraces (LR; green), and PPB varieties (PB; red). Genotype numbers correspond to those listed in Tables 4.1-4.3.



**Figure S4.5.** Histograms of yield ( $\text{kg ha}^{-1}$ ) values of genotypes comprising the HON panel tested at three field locations from 2014-2015. A. Elora 2014, B. Elora 2015 and C. Yorito. Breeding history category averages with standard errors are presented, followed by individual genotype LSmeans with standard errors. North American check genotypes (CK; purple), Honduran conventional genotypes (CV; blue), landraces (LR; green), and PPB varieties (PB; red). Genotype numbers correspond to those listed in Tables 4.1-4.3.

## **APPENDIX C**

Supplementary (S) Tables and Figures for Chapter 5.

**Table S5.1.** Trial site conditions for 2014-2016 low-nitrogen dry bean (*Phaseolus vulgaris* L.) field trials. The soil tests for 2014 and 2015 were performed on composite samples across the field sites of the top layer (0-15 cm) and mid layer (15-30 cm) of soil. The soil test for 2016 was performed on a composite field sample. Rainfall and temperature data for 2014 for the Elora Research Station (Elora). Rainfall and temperature data for the Belwood field site area as reported from the nearest Canada Weather Service location at the Fergus Shand Dam. Growing degree days (GDD calculated with base '0') were calculated based on the total growing days for each location-year. (na – not available; \* the Isabela site was drip irrigated.)

Soil test (ppm)	<b>Elora 2014</b>		<b>Belwood 2015</b>		<b>Isabela 2016</b>	
	NO <sub>3</sub> <sup>-</sup>	NH <sub>4</sub> <sup>+</sup>	NO <sub>3</sub> <sup>-</sup>	NH <sub>4</sub> <sup>+</sup>	NO <sub>3</sub> <sup>-</sup>	NH <sub>4</sub> <sup>+</sup>
<b>0-15 cm</b>	5.7	3.4	6.4	3.5	2	na
<b>15-30 cm</b>	8.6	2.9	5.5	2.6		na
<b>Planting date</b>	14-Jun		26-Jun		22-Dec	
<b>First killing frost date (E14, B15)/harvest date (I16)</b>	11-Oct		17-Oct		15-Mar	
<b>Total growing days</b>	124		114		84	
<b>Total growing degree days (GDD)</b>	1912.8		1862.6		na	
<b>Precipitation (mm)</b>						
<b>June</b>	68.9		204.7		--	
<b>July</b>	133.7		68.6		--	
<b>August</b>	51		104.8		--	
<b>September</b>	164.8		47.9		--	
<b>October</b>	74.3		108.2		--	
<b>December</b>	--		--		0	
<b>January</b>	--		--		0	
<b>February</b>	--		--		0	
<b>March</b>	--		--		0	
<b>Total</b>	492.7		534.2		0*	

**Table S5.2.** Combined mixed-model analysis of fixed and random effects on seed composition traits in 280 MDP genotypes tested at three field locations (Elora 2014, Belwood 2015, and Isabela 2016).

Fixed effect <sup>†</sup>	%Ndfa <sup>§</sup>		δ <sup>15</sup> N <sup>¶</sup>		Seed N (%) <sup>#</sup>	
	F value	Pr>F	F value	Pr>F	F value	Pr>F
<b>Genotype</b>	1.38	0.0008	1.35	0.0015	2.92	<.0001
<b>MA vs. DJ</b>	0.1	0.7516	0.34	0.5626	19.92	<.0001
Random effects <sup>‡</sup>	s <sup>2</sup>	Se	s <sup>2</sup>	se	s <sup>2</sup>	se
<b>ENV</b>	0.07812 <sup>ns</sup>	0.08271	3.2979 <sup>ns</sup>	3.4332	0.08116 <sup>ns</sup>	0.08688
<b>BLOC(ENV)</b>	0.008846 <sup>ns</sup>	0.00733	0.2619 <sup>ns</sup>	0.2174	0.0104 <sup>ns</sup>	0.008854
<b>IBLK(ENV*BLOC)</b>	0.000733*	0.00033	0.02652**	0.0107	0.003184**	0.001361
<b>ENV*Genotype</b>	0.004611***	0.00096	0.1204***	0.02714	0.04545***	0.004896
<b>Residual</b>	0.0178***	0.00093	0.5084***	0.02667	0.05941***	0.003076

<sup>†</sup> Genotype effects overall and by race, Mesoamerica (MA) and Durango-Jalisco (DJ)

<sup>‡</sup> Variance component estimates (s<sup>2</sup>) and their standard errors (se) of environment (ENV), block (BLOC), and incomplete block (IBLK)

<sup>§</sup> Percentage of nitrogen derived from the atmosphere

<sup>¶</sup> Nitrogen discrimination value indicates the relative amount of elemental (<sup>15</sup>N) nitrogen to atmospheric (<sup>14</sup>N) present in the seed (<sup>15</sup>N:<sup>14</sup>N) compared to a reference standard to obtain nitrogen discrimination values (δ<sup>15</sup>N, ‰)

<sup>#</sup> % nitrogen in the seed

<sup>ns</sup> Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S5.3.** Mixed model analysis of fixed and random effects on seed composition traits in 280 MDP genotypes tested at Elora, Ontario, Canada, 2014.

Fixed effect <sup>†</sup>	%Ndfa <sup>§</sup>		δ <sup>15</sup> N <sup>¶</sup>		Seed N (%) <sup>#</sup>	
	F value	Pr>F	F value	Pr>F	F value	Pr>F
<b>Genotype</b>	1.56	0.0003	1.56	0.0003	2.83	<.0001
<b>MA vs DJ</b>	17.92	<.0001	17.92	<.0001	1.86	0.1727
Random effects <sup>‡</sup>	s <sup>2</sup>	se	s <sup>2</sup>	se	s <sup>2</sup>	se
<b>BLOC</b>	0.001727 <sup>ns</sup>	0.00268	0.07055 <sup>ns</sup>	0.1096	0.02995 <sup>ns</sup>	0.04325
<b>IBLK(BLOC)</b>	0.001652*	0.00095	0.06748*	0.03898	0.004965 <sup>ns</sup>	0.003715
<b>Residual</b>	0.01633***	0.00156	0.6669***	0.06371	0.08028***	0.00761

<sup>†</sup> Genotype effects overall and by race, Mesoamerica (MA) and Durango-Jalisco (DJ)

<sup>‡</sup> Variance component estimates (s<sup>2</sup>) and their standard errors (se) of environment (ENV), block (BLOC), and incomplete block (IBLK)

<sup>§</sup> Percentage of nitrogen derived from the atmosphere

<sup>¶</sup> Nitrogen discrimination value indicates the relative amount of elemental (<sup>15</sup>N) nitrogen to atmospheric (<sup>14</sup>N) present in the seed (<sup>15</sup>N:<sup>14</sup>N) compared to a reference standard to obtain nitrogen discrimination values (δ<sup>15</sup>N, ‰)

<sup>#</sup> % nitrogen in the seed

<sup>ns</sup> Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S5.4.** Mixed model analysis of fixed and random effects on seed composition traits in 280 MDP genotypes tested at Belwood, Ontario, Canada, 2015.

Fixed effect <sup>†</sup>	%Ndfa <sup>§</sup>		δ <sup>15</sup> N <sup>¶</sup>		Seed N (%) <sup>#</sup>	
	F value	Pr>F	F value	Pr>F	F value	Pr>F
<b>Genotype</b>	2.11	<.0001	2.11	<.0001	6.37	<.0001
<b>MA vs DJ</b>	9.45	0.0022	9.45	0.0022	61.3	<.0001
Random effects <sup>‡</sup>	s <sup>2</sup>	se	s <sup>2</sup>	se	s <sup>2</sup>	se
<b>BLOC</b>	0.002876 <sup>ns</sup>	0.004125	0.03822 <sup>ns</sup>	0.05483	0.001886 <sup>ns</sup>	0.003149
<b>IBLK(BLOC)</b>	-0.00015	0.000314	-0.00204	0.004172	0.00138 <sup>ns</sup>	0.001706
<b>Residual</b>	0.01543 <sup>***</sup>	0.001360	0.2050 <sup>***</sup>	0.01808	0.05399 <sup>***</sup>	0.004762

<sup>†</sup> Genotype effects overall and by race, Mesoamerica (MA) and Durango-Jalisco (DJ)

<sup>‡</sup> Variance component estimates (s<sup>2</sup>) and their standard errors (se) of environment (ENV), block (BLOC), and incomplete block (IBLK)

<sup>§</sup> Percentage of nitrogen derived from the atmosphere

<sup>¶</sup> Nitrogen discrimination value indicates the relative amount of elemental (<sup>15</sup>N) nitrogen to atmospheric (<sup>14</sup>N) present in the seed (<sup>15</sup>N:<sup>14</sup>N) compared to a reference standard to obtain nitrogen discrimination values (δ<sup>15</sup>N, ‰)

<sup>#</sup> % nitrogen in the seed

<sup>ns</sup> Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S5.5.** Mixed model analysis of fixed and random effects on seed composition traits in 280 MDP genotypes tested at Isabela, Puerto Rico, 2016.

Fixed effect <sup>†</sup>	%Ndfa <sup>§</sup>		δ <sup>15</sup> N <sup>¶</sup>		Seed N (%) <sup>#</sup>	
	F value	Pr>F	F value	Pr>F	F value	Pr>F
<b>Genotype</b>	1.61	<.0001	1.61	<.0001	3.25	<.0001
<b>MA vs DJ</b>	0.14	0.7109	0.14	0.7109	10.96	0.001
Random effects <sup>‡</sup>	s <sup>2</sup>	se	s <sup>2</sup>	se	s <sup>2</sup>	se
<b>BLOC</b>	0.0219 <sup>ns</sup>	0.03138	0.6731 <sup>ns</sup>	0.9644	0.001096 <sup>ns</sup>	0.002123
<b>IBLK(BLOC)</b>	0.002117*	0.00125	0.06505*	0.03843	0.002338 <sup>ns</sup>	0.00199
<b>Residual</b>	0.02077 <sup>***</sup>	0.00184	0.6384 <sup>***</sup>	0.05639	0.04632 <sup>***</sup>	0.004095

<sup>†</sup> Genotype effects overall and by race, Mesoamerica (MA) and Durango-Jalisco (DJ)

<sup>‡</sup> Variance component estimates (s<sup>2</sup>) and their standard errors (se) of environment (ENV), block (BLOC), and incomplete block (IBLK)

<sup>§</sup> Percentage of nitrogen derived from the atmosphere

<sup>¶</sup> Nitrogen discrimination value indicates the relative amount of elemental (<sup>15</sup>N) nitrogen to atmospheric (<sup>14</sup>N) present in the seed (<sup>15</sup>N:<sup>14</sup>N) compared to a reference standard to obtain nitrogen discrimination values (δ<sup>15</sup>N, ‰)

<sup>#</sup> % nitrogen in the seed

<sup>ns</sup> Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S5.6.** Mixed model analysis of fixed and random effects on leaf chlorophyll content at two time-points in 280 MDP genotypes tested at Elora (2014) and Belwood (2015) Ontario, Canada, and tested at one time-point at Isabela (2016), Puerto Rico.

	Elora				Belwood				Isabela	
	<u>SPAD1</u> <sup>§</sup>		<u>SPAD2</u> <sup>¶</sup>		<u>SPAD1</u> <sup>§</sup>		<u>SPAD2</u> <sup>¶</sup>		<u>SPAD2</u> <sup>¶</sup>	
<b>Fixed effect</b> <sup>†</sup>	<b>F value</b>	<b>Pr&gt;F</b>								
<b>Genotype</b>	4.08	<.0001	3.99	<.0001	1.67	<.0001	2.88	<.0001	3.51	<.0001
<b>MA vs DJ</b>	3.14	0.0768	141.34	<.0001	3.31	<.0001	73.82	<.0001	93.33	<.0001
<b>Random effects</b> <sup>‡</sup>	<b>s<sup>2</sup></b>	<b>se</b>								
<b>BLOC</b>	0.2487 <sup>ns</sup>	0.3713	0.1911 <sup>ns</sup>	0.3287	1.2028 <sup>ns</sup>	1.7587	0.8584 <sup>ns</sup>	1.4321	1.5433 <sup>ns</sup>	2.2249
<b>IBLK(BLOC)</b>	-0.02047	0.1150	-0.2554	0.3924	0.2306 <sup>ns</sup>	0.1854	1.2189*	0.6182	0.08269 <sup>ns</sup>	0.1939
<b>SPADT(BLOC*genotype)</b>	9.1863***	0.4000	46.5057***	1.8140	-12.9232	1.5674	16.5428*	9.1559	19.1085	0.7692
<b>Residual</b>	0.9113	0	0.9796	0	27.3573***	1.8351	9.3486 <sup>ns</sup>	9.1004	0.952	0

<sup>†</sup> Genotype effects overall and by race, Mesoamerica (MA) and Durango-Jalisco (DJ)

<sup>‡</sup> block (BLOC), incomplete block (IBLK), and SPAD measurement time (SPADT)

<sup>§</sup> SPAD1 (leaf chlorophyll content) at first trifoliolate stage

<sup>¶</sup> SPAD2 (leaf chlorophyll content) at 100% flowering stage

<sup>ns</sup> Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S5.7.** Mixed model analysis of fixed and random effects on nodulation for 280 MDP genotypes tested at Isabela, Puerto Rico (2016).

<b>Fixed effect<sup>†</sup></b>	<b>Nodule count<sup>§</sup></b>		<b>Nodule size<sup>¶</sup></b>	
	<b>F value</b>	<b>Pr&gt;F</b>	<b>F value</b>	<b>Pr&gt;F</b>
<b>Genotype</b>	2.47	<.0001	2.48	<.0001
<b>MA vs DJ</b>	7.99	0.0048	0.07	0.7952
<b>Random effects<sup>‡</sup></b>	<b>s<sup>2</sup></b>	<b>se</b>	<b>s<sup>2</sup></b>	<b>se</b>
<b>BLOC</b>	9.3649 <sup>ns</sup>	15.555	0.007074 <sup>ns</sup>	0.01459
<b>IBLK(BLOC)</b>	13.0545*	6.7802	0.02575*	0.0128
<b>trait(BLOC*genotype)</b>	247.29***	10.3665	0.04068*	0.01815

<sup>†</sup> Genotype effects overall and by race, Mesoamerica (MA) and Durango-Jalisco (DJ)

<sup>‡</sup> block (BLOC), incomplete block (IBLK), and nodule trait interaction (trait)

<sup>§</sup> Nodule count, number of nodules per root system

<sup>¶</sup> Nodule size, size of nodules on a relative scale (1 – small, 2 – medium, 3 – large).

<sup>ns</sup> Not significant; \*, \*\*, and \*\*\* are significant at 0.05, 0.01, and 0.001, respectively

**Table S5.8.** QTL significantly associated with %Ndfa in the entire MDP (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
7	22,082,170	2.60125E-07	A/G	52.6	62.4	9.8	Elora	<b>Phvul.007G138900</b> (22,046,579..22,049,807) Double Clp-N motif-containing P-loop nucleoside triphosphate hydrolases superfamily protein; IPR004176 (Clp, N-terminal), IPR023150 (Double Clp-N motif), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0019538 (protein metabolic process)	NA
8	1,759,458	7.51713E-09	T/C	54.0	25.8	-28.2	Elora	<b>Phvul.008G020500</b> (1,715,225..1,721,101) translation initiation factor eIF-2B delta subunit; IPR000649 (Initiation factor 2B-related); GO:0044237 (cellular metabolic process)	NA
								<b>Phvul.008G020600</b> (1,727,505..1,734,392) nuclear matrix constituent protein-related	NA
								<b>Phvul.008G020700</b> (1,735,768..1,747,788) Disease resistance protein (TIR-NBS-LRR class) family; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)	NA
								<b>Phvul.008G020750</b> (1,743,459..1,745,204) disease resistance family protein/LRR protein, putative; IPR002182 (NB-ARC); GO:0043531 (ADP binding)	NA
								<b>Phvul.008G020800</b> (1,751,096..1,753,598) peptide transporter 1; IPR000109 (Proton-dependent oligopeptide transporter family), IPR016196 (Major facilitator superfamily domain, general substrate transporter); GO:0005215 (transporter activity), GO:0006810 (transport), GO:0016020 (membrane)	NA
								<b>Phvul.008G020850</b> (1,755,309..1,755,635) Disease resistance protein (CC-NBS-LRR class) family	NA
								<b>Phvul.008G020900</b> (1,755,656..1,757,267) LRR and NB-ARC domain disease resistance protein; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
								<b>Phvul.008G021001</b> (1,759,069..1,764,686) Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0	NA
								<b>Phvul.008G021100</b> (1,769,720..1,771,906) protein YLS7-like [Glycine max]; IPR025846 (PMR5 N-terminal domain), IPR026057 (PC-Esterase)	NA
								<b>Phvul.008G021200</b> (1,777,460..1,782,534) protein YLS7-like [Glycine max]; IPR025846 (PMR5 N-terminal domain), IPR026057 (PC-Esterase)	NA
								<b>Phvul.008G021300</b> (1,790,014..1,790,672) hypothetical protein	NA
								<b>Phvul.008G021400</b> (1,797,269..1,803,893) protein FLX-like 1-like isoform X1 [Glycine max]	NA
								<b>Phvul.008G021500</b> (1,804,913..1,814,885) cell division cycle protein 27 homolog B-like isoform X1 [Glycine max]; IPR011990 (Tetratricopeptide-like helical), IPR026819 (Cell division cycle protein 27A/B, plant); GO:0005515 (protein binding), GO:0005680 (anaphase-promoting complex), GO:0030154 (cell differentiation), GO:0051302 (regulation of cell division)	NA
<b>8</b>	8,377,081	7.21591E-07	G/A	53.1	62.4	9.3	Elora	<b>Phvul.008G084200</b> (8,338,902..8,339,693) unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: inflorescence meristem, hypocotyl, root, flower; EXPRESSED DURING: petal differentiation and expansion stage; Has 23 Blast hits to 23 proteins in 9 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 22; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink).	NA
								<b>Phvul.008G084300</b> (8,345,405..8,347,333) Inner membrane protein yicO n=3 Tax=Magnaporthe oryzae RepID=G4MX99_MAGO7; IPR006043 (Xanthine/uracil/vitamin C permease); GO:0005215 (transporter activity), GO:0006810 (transport),	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
								GO:0016020 (membrane), GO:0055085 (transmembrane transport)	
								<b>Phvul.008G084400</b> (8,377,154..8,382,533) stress response protein NST1-like [Glycine max]	NA
								<b>Phvul.008G084500</b> (8,387,795..8,394,133) stress response protein NST1-like [Glycine max]	NA
								<b>Phvul.008G084600</b> (8,396,993..8,398,378) protein kinase family protein; IPR020636 (Calcium/calmodulin-dependent/calcium-dependent protein kinase); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation), GO:0007165 (signal transduction)	NA
								<b>Phvul.008G084700</b> (8,421,134..8,425,801) saposin B domain-containing protein; IPR011001 (Saposin-like); GO:0006629 (lipid metabolic process)	NA
<b>10</b>	9,818,036	6.34466E-11	A/T	54.8	50.8	-4.0	Elora	NA	NA

**Table S5.9.** QTL significantly associated with %Ndfa in the MA subpopulation (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Pv	SNP rs# (position, bp)	minus log <sub>10</sub> (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location at which association significant	Candidate gene, location (bp), annotation	References
8	1,967,416	2.8748E-07	A/G	50.8	26.8	24.04	Elora	<b>Phvul.008G023000</b> (1,915,107..1,917,523) Ribosomal RNA processing Brix domain protein; IPR007109 (Brix domain), IPR026532 (Ribosome biogenesis protein BRX1)	NA
								<b>Phvul.008G023100</b> (1,926,485..1,926,986) SCR; IPR010682 (Plant self-incompatibility response); GO:0007165 (signal transduction)	NA
								<b>Phvul.008G023200</b> (1,929,549..1,930,171) SCR related; IPR010682 (Plant self-incompatibility response); GO:0007165 (signal transduction)	NA
								<b>Phvul.008G023300</b> (1,933,941..1,934,453) SCR; IPR010682 (Plant self-incompatibility response); GO:0007165 (signal transduction)	NA
								<b>Phvul.008G023400</b> (1,944,141..1,945,058) glutamine dumper 2	NA
								<b>Phvul.008G023501</b> (1,950,828..1,951,977) uncharacterized protein LOC102660012 isoform X2 [Glycine max]	NA
								<b>Phvul.008G023600</b> (1,952,432..1,952,880) Unknown protein	NA
								<b>Phvul.008G023700</b> (1,954,462..1,957,229) 6-phosphogluconate dehydrogenase family protein; IPR006113 (6-phosphogluconate dehydrogenase, decarboxylating), IPR008927 (6-phosphogluconate dehydrogenase, C-terminal-like), IPR016040 (NAD(P)-binding domain); GO:0004616 (phosphogluconate dehydrogenase (decarboxylating) activity), GO:0006098 (pentose-phosphate shunt), GO:0016491 (oxidoreductase activity), GO:0050661 (NADP binding), GO:0050662 (coenzyme binding), GO:0055114 (oxidation-reduction process)	NA
<b>Phvul.008G023800</b> (1,960,551..1,961,414) TGACG-sequence-specific DNA-binding protein TGA-1B-like [Glycine max]; IPR004827 (Basic-leucine zipper domain); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)	NA								

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location at which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.008G023900</b> (1,974,646..1,978,175) Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)	NA
								<b>Phvul.008G024000</b> (1,978,629..1,983,581) WPP domain-interacting tail-anchored protein 1-like isoform X2 [Glycine max]	NA
								<b>Phvul.008G024100</b> (1,992,355..1,995,948) Transmembrane amino acid transporter family protein; IPR013057 (Amino acid transporter, transmembrane)	NA
								<b>Phvul.008G024200</b> (1,996,339..1,998,018) UPF0481 protein [Glycine max]; IPR004158 (Protein of unknown function DUF247, plant)	NA
								<b>Phvul.008G024300</b> (1,999,683..2,003,178) Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)	NA
								<b>Phvul.008G024400</b> (2,004,322..2,008,280) Pentatricopeptide repeat (PPR) superfamily protein; IPR002625 (Smr protein/MutS2 C-terminal), IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)	NA
								<b>Phvul.008G024500</b> (2,009,363..2,017,455) kanadaplin-like isoform X1 [Glycine max]; IPR008984 (SMAD/FHA domain); GO:0005515 (protein binding)	NA
<b>10</b>	11,592,491	3.0759E-07	G/C	51.0	28.8	22.22	Elora	<b>Phvul.010G063500</b> (11,550,385..11,550,948) Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)	NA
								<b>Phvul.010G063600</b> (11,613,891..11,614,460) Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)	NA

**Table S5.10.** QTL significantly associated with nitrogen difference ( $\delta^{15}\text{N}$ ) in the entire MDP (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Pv	SNP rs# (position, bp)	minus log <sub>10</sub> (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
1	4,933,671	5.98732E-07	A/G	1.0	0.4	-0.60	Elora	<b>Phvul.001G046900</b> (4,924,736..4,926,129) ethylene-responsive transcription factor 3-like [Glycine max]; IPR016177 (DNA-binding domain); GO:0003677 (DNA binding), GO:0003700 (sequence-specific DNA binding transcription factor activity)	NA
								<b>Phvul.001G047001</b> (4,950,443..4,955,184) aspartate aminotransferase 5; IPR000796 (Aspartate/other aminotransferase), IPR015424 (Pyridoxal phosphate-dependent transferase); GO:0003824 (catalytic activity), GO:0006520 (cellular amino acid metabolic process), GO:0008483 (transaminase activity), GO:0009058 (biosynthetic process), GO:0030170 (pyridoxal phosphate binding)	NA
								<b>Phvul.001G047100</b> (4,955,650..4,961,025) Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)	NA
								<b>Phvul.001G047200</b> (4,966,056..4,972,213) aberrant lateral root formation 4; IPR013877 (YAP-binding/Alf4/Glomulin)	NA
								<b>Phvul.001G047300</b> (4,972,376..4,981,168) trafficking protein particle complex subunit 9-like [Glycine max]; IPR011990 (Tetratricopeptide-like helical), IPR013935 (TRAPP II complex, Trs120); GO:0005515 (protein binding)	NA
								<b>Phvul.002G325500</b> (49,039,240..49,039,764) Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)	NA
2	49,084,007	9.92802E-12	A/G	0.6	1.1	0.50	Elora	<b>Phvul.002G325600</b> (49,053,148..49,053,874) uncharacterized protein LOC100820443 [Glycine max]; IPR006747 (Protein of unknown function DUF599)	NA
								<b>Phvul.002G325700</b> (49,055,984..49,058,744) uncharacterized protein LOC100820443 [Glycine max]; IPR006747 (Protein of unknown function DUF599)	NA
								<b>Phvul.002G325800</b> (49,062,024..49,064,981) uncharacterized protein LOC100820443 [Glycine max]; IPR006747 (Protein of unknown function DUF599)	NA
								<b>Phvul.002G325900</b> (49,064,981..49,067,741) uncharacterized protein LOC100820443 [Glycine max]; IPR006747 (Protein of unknown function DUF599)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
								<b>Phvul.002G325900</b> (49,067,511..49,073,596) poly(A) polymerase 1; IPR014492 (Poly(A) polymerase); GO:0003723 (RNA binding), GO:0004652 (polynucleotide adenylyltransferase activity), GO:0005634 (nucleus), GO:0016779 (nucleotidyltransferase activity), GO:0031123 (RNA 3'-end processing), GO:0043631 (RNA polyadenylation)	NA
								<b>Phvul.002G326000</b> (49,075,423..49,076,898) uncharacterized protein LOC100787767 [Glycine max]; IPR004864 (Late embryogenesis abundant protein, LEA-14)	NA
								<b>Phvul.002G326100</b> (49,081,592..49,085,315) 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; IPR005123 (Oxoglutarate/iron-dependent dioxygenase); GO:0005506 (iron ion binding), GO:0016491 (oxidoreductase activity), GO:0031418 (L-ascorbic acid binding)	NA
								<b>Phvul.002G326200</b> (49,086,617..49,089,190) Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)	(MacQueen et al., 2020)
								<b>Phvul.002G326300</b> (49,090,552..49,092,942) T-complex protein 1 subunit delta, putative; IPR002423 (Chaperonin Cpn60/TCP-1), IPR027409 (GroEL-like apical domain), IPR027410 (TCP-1-like chaperonin intermediate domain), IPR027413 (GroEL-like equatorial domain); GO:0005524 (ATP binding), GO:0006457 (protein folding), GO:0044267 (cellular protein metabolic process), GO:0051082 (unfolded protein binding)	NA
								<b>Phvul.002G326400</b> (49,094,816..49,105,934) ATP-dependent RNA helicase, putative; IPR001650 (Helicase, C-terminal), IPR007502 (Helicase-associated domain), IPR011709 (Domain of unknown function DUF1605), IPR014001 (Helicase, superfamily 1/2, ATP-binding domain), IPR014720 (Double-stranded RNA-binding domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003676 (nucleic acid binding), GO:0004386 (helicase activity), GO:0005524 (ATP binding), GO:0008026 (ATP-dependent helicase activity)	NA
								<b>Phvul.002G326500</b> (49,107,649..49,109,932) Glycosyltransferase family 61 protein; IPR007657 (Glycosyltransferase AER61, uncharacterised)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
								<b>Phvul.002G326600</b> (49,114,990..49,116,625) 1-aminocyclopropane-1-carboxylate oxidase; IPR005123 (Oxoglutarate/iron-dependent dioxygenase), IPR026992 (Non-haem dioxygenase N-terminal domain), IPR027443 (Isopenicillin N synthase-like); GO:0016491 (oxidoreductase activity), GO:0055114 (oxidation-reduction process)	(Nova-Franco et al., 2015)
								<b>Phvul.002G326700</b> (49,120,944..49,123,956) MATE efflux family protein; IPR002528 (Multi antimicrobial extrusion protein); GO:0006855 (drug transmembrane transport), GO:0015238 (drug transmembrane transporter activity), GO:0015297 (antiporter activity), GO:0016020 (membrane), GO:0055085 (transmembrane transport)	NA
								<b>Phvul.002G326800</b>	NA
								<b>Phvul.002G326900</b> (49,131,317..49,134,457) uncharacterized protein LOC100785837 [Glycine max]; IPR019448 (EEIG1/EHBP1 N-terminal domain)	NA
7	22,082,166	5.49613E-07	G/A	1.0	0.4	-0.63	Elora	<b>Phvul.007G138900</b> (22,046,579..22,049,807) Double Clp-N motif-containing P-loop nucleoside triphosphate hydrolases superfamily protein; IPR004176 (Clp, N-terminal), IPR023150 (Double Clp-N motif), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0019538 (protein metabolic process)	NA
8	1,759,458	5.57324E-12	C/T	2.8	1.0	-1.80	Elora	<b>Phvul.008G020500</b> (1,715,225..1,721,101) translation initiation factor eIF-2B delta subunit; IPR000649 (Initiation factor 2B-related); GO:0044237 (cellular metabolic process)	NA
								<b>Phvul.008G020600</b> (1,727,505..1,734,392) nuclear matrix constituent protein-related	NA
								<b>Phvul.008G020700</b> (1,735,768..1,747,788) Disease resistance protein (TIR-NBS-LRR class) family; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)	NA
								<b>Phvul.008G020750</b> (1,743,459..1,745,204) disease resistance family protein/LRR protein, putative; IPR002182 (NB-ARC); GO:0043531 (ADP binding)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
								<b>Phvul.008G020800</b> (1,751,096..1,753,598) peptide transporter 1; IPR000109 (Proton-dependent oligopeptide transporter family), IPR016196 (Major facilitator superfamily domain, general substrate transporter); GO:0005215 (transporter activity), GO:0006810 (transport), GO:0016020 (membrane)	NA
								<b>Phvul.008G020850</b> (1,755,309..1,755,635) Disease resistance protein (CC-NBS-LRR class) family	NA
								<b>Phvul.008G020900</b> (1,755,656..1,757,267) LRR and NB-ARC domain disease resistance protein; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)	NA
								<b>Phvul.008G021001</b> (1,759,069..1,764,686) Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain), IPR013320 (Concanavalin A-like lectin/glucanase, subgroup); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0	NA
								<b>Phvul.008G021100</b> (1,769,720..1,771,906) protein YLS7-like [Glycine max]; IPR025846 (PMR5 N-terminal domain), IPR026057 (PC-Esterase)	NA
								<b>Phvul.008G021200</b> (1,777,460..1,782,534) protein YLS7-like [Glycine max]; IPR025846 (PMR5 N-terminal domain), IPR026057 (PC-Esterase)	NA
								<b>Phvul.008G021300</b> (1,790,014..1,790,672) hypothetical protein	NA
<b>8</b>	8,377,081	8.6646E-08	G/A	1.0	0.4	-0.63	Elora	<b>Phvul.008G084200</b> (8,338,902..8,339,693) unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: inflorescence meristem, hypocotyl, root, flower; EXPRESSED DURING: petal differentiation and expansion stage; Has 23 Blast hits to 23 proteins in 9 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 22; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink).	NA
								<b>Phvul.008G084300</b> (8,345,405..8,347,333) Inner membrane protein yicO n=3 Tax=Magnaporthe oryzae RepID=G4MX99_MAGO7; IPR006043 (Xanthine/uracil/vitamin C permease); GO:0005215 (transporter activity), GO:0006810 (transport), GO:0016020 (membrane), GO:0055085 (transmembrane transport)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	Reference
								<b>Phvul.008G084400</b> (8,377,154..8,382,533) stress response protein NST1-like [Glycine max]	NA
								<b>Phvul.008G084500</b> (8,387,795..8,394,133) stress response protein NST1-like [Glycine max]	NA
								<b>Phvul.008G084600</b> (8,396,993..8,398,378) protein kinase family protein; IPR020636 (Calcium/calmodulin-dependent/calcium-dependent protein kinase); GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation), GO:0007165 (signal transduction)	NA
								<b>Phvul.008G084700</b> (8,421,134..8,425,801) saposin B domain-containing protein; IPR011001 (Saposin-like); GO:0006629 (lipid metabolic process)	NA
<b>10</b>	9,818,036	1.37069E-11	A/T	0.9	1.2	0.25	Elora	NA	NA

**Table S5.11.** QTL significantly associated with nitrogen difference ( $\delta^{15}\text{N}$ ) in the MA subpopulation (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Pv	SNP rs# (position, bp)	minus log <sub>10</sub> (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location at which association significant	Candidate gene, location (bp), annotation	References
8	1,967,416	2.8748E-07	A/G	1.2	2.7	1.54	Elora	Phvul.008G023000 (1,915,107..1,917,523) Ribosomal RNA processing Brix domain protein; IPR007109 (Brix domain), IPR026532 (Ribosome biogenesis protein BRX1)	NA
								Phvul.008G023100 (1,926,485..1,926,986) SCR; IPR010682 (Plant self-incompatibility response); GO:0007165 (signal transduction)	NA
								Phvul.008G023200 (1,929,549..1,930,171) SCR related; IPR010682 (Plant self-incompatibility response); GO:0007165 (signal transduction)	NA
								Phvul.008G023300 (1,933,941..1,934,453) SCR; IPR010682 (Plant self-incompatibility response); GO:0007165 (signal transduction)	NA
								Phvul.008G023400 (1,944,141..1,945,058) glutamine dumper 2	NA
								Phvul.008G023501 (1,950,828..1,951,977) uncharacterized protein LOC102660012 isoform X2 [Glycine max]	NA
								Phvul.008G023600 (1,952,432..1,952,880) Unknown protein	NA
								Phvul.008G023700 (1,954,462..1,957,229) 6-phosphogluconate dehydrogenase family protein; IPR006113 (6-phosphogluconate dehydrogenase, decarboxylating), IPR008927 (6-phosphogluconate dehydrogenase, C-terminal-like), IPR016040 (NAD(P)-binding domain); GO:0004616 (phosphogluconate dehydrogenase (decarboxylating) activity), GO:0006098 (pentose-phosphate shunt), GO:0016491 (oxidoreductase activity), GO:0050661 (NADP binding), GO:0050662 (coenzyme binding), GO:0055114 (oxidation-reduction process)	NA
Phvul.008G023800 (1,960,551..1,961,414) TGACG-sequence-specific DNA-binding protein TGA-1B-like [Glycine max]; IPR004827 (Basic-leucine zipper domain); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)	NA								

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location at which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.008G023900</b> (1,974,646..1,978,175) Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)	NA
								<b>Phvul.008G024000</b> (1,978,629..1,983,581) WPP domain-interacting tail-anchored protein 1-like isoform X2 [Glycine max]	NA
								<b>Phvul.008G024100</b> (1,992,355..1,995,948) Transmembrane amino acid transporter family protein; IPR013057 (Amino acid transporter, transmembrane)	NA
								<b>Phvul.008G024200</b> (1,996,339..1,998,018) UPF0481 protein [Glycine max]; IPR004158 (Protein of unknown function DUF247, plant)	NA
								<b>Phvul.008G024300</b> (1,999,683..2,003,178) Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)	NA
								<b>Phvul.008G024400</b> (2,004,322..2,008,280) Pentatricopeptide repeat (PPR) superfamily protein; IPR002625 (Smr protein/MutS2 C-terminal), IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)	NA
								<b>Phvul.008G024500</b> (2,009,363..2,017,455) kanadapin-like isoform X1 [Glycine max]; IPR008984 (SMAD/FHA domain); GO:0005515 (protein binding)	NA
<b>10</b>	11,592,491	2.9066E-07	G/C	1.1	2.6	1.42	Elora	<b>Phvul.010G063500</b> (11,550,385..11,550,948) Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)	NA
								<b>Phvul.010G063600</b> (11,613,891..11,614,460) Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)	NA

**Table S5.12.** QTL significantly associated with seed nitrogen content (%N) in the entire MDP (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
2	11,937,912	2.99758E-10	A/G	4.2	3.7	-0.54	Elora	<b>Phvul.002G077200</b> (11,887,655..11,892,280) microtubule-associated protein futsch isoform X8 [Glycine max]	NA
								<b>Phvul.002G077300</b> (11,893,313..11,894,000) germin-like protein 7; IPR001929 (Germin); GO:0030145 (manganese ion binding), GO:0045735 (nutrient reservoir activity)	NA
								<b>Phvul.002G077400</b> (11,895,060..11,897,385) probable galacturonosyltransferase-like 9-like [Glycine max]; IPR002495 (Glycosyl transferase, family 8)	NA
								<b>Phvul.002G077500</b> (11,917,404..11,917,858) Unknown protein	NA
								<b>Phvul.002G077600</b> (11,928,817..11,935,492) pyruvate dehydrogenase E1 component, alpha subunit; IPR017597 (Pyruvate dehydrogenase (acetyl-transferring) E1 component, alpha subunit, subgroup y); GO:0004739 (pyruvate dehydrogenase (acetyl-transferring) activity), GO:0006096 (glycolysis), GO:0008152 (metabolic process), GO:0043231 (intracellular membrane-bounded organelle), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.002G077700</b> (11,945,222..11,953,149) helicases; ATP-dependent helicases; nucleic acid binding; ATP binding; DNA-directed DNA polymerases; DNA binding; IPR001650 (Helicase, C-terminal), IPR012961 (DSH, C-terminal), IPR014001 (Helicase, superfamily 1/2, ATP-binding domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003676 (nucleic acid binding), GO:0004386 (helicase activity), GO:0005524 (ATP binding), GO:0008026 (ATP-dependent helicase activity)	NA
								<b>Phvul.002G077800</b> (11,967,265..11,979,414) paired amphipathic helix protein Sin3-like 4-like isoform X1 [Glycine max]; IPR003822 (Paired amphipathic helix), IPR013194 (Histone deacetylase interacting); GO:0005634 (nucleus)	NA
2	21,924,033	2.44359E-07	G/A	4.2	4.5	0.24	Elora	<b>Phvul.002G102900</b> (21,898,467..21,900,647) UDP-glucose 6-dehydrogenase family protein; IPR017476 (UDP-glucose/GDP-mannose dehydrogenase); GO:0003979 (UDP-glucose 6-dehydrogenase activity), GO:0051287 (NAD binding), GO:0055114 (oxidation-reduction process)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.002G103000</b> (21,971,465..21,975,414) uncharacterized protein LOC102668171 isoform X3 [Glycine max]	NA
5	10,781,169	1.45012E-14	G/A	4.3	4.0	-0.29	Elora	<b>Phvul.005G065100</b> (10,768,480..10,770,453) F-box/RNI-like superfamily protein; IPR001810 (F-box domain), IPR006566 (FBD domain); GO:0005515 (protein binding)	NA
								<b>Phvul.005G065200</b> (10,802,709..10,803,179) RING-H2 finger protein 2B; IPR013083 (Zinc finger, RING/FYVE/PHD-type); GO:0005515 (protein binding), GO:0008270 (zinc ion binding)	NA
								<b>Phvul.005G065300</b> (10,820,735..10,823,500) uncharacterized protein LOC102664055 [Glycine max]; IPR021319 (Protein of unknown function DUF2921)	NA
								<b>Phvul.005G065400</b> (10,835,946..10,840,116) hypothetical protein	NA
7	32,785,603	2.19503E-09	A/G	4.3	4.0	-0.25	Elora	<b>Phvul.007G204700</b> (32,739,247..32,740,616) uncharacterized protein LOC102660270 [Glycine max]; IPR019786 (Zinc finger, PHD-type, conserved site)	NA
								<b>Phvul.007G204900</b> (32,748,473..32,749,795) isopenicillin N epimerase-like protein; IPR015424 (Pyridoxal phosphate-dependent transferase); GO:0003824 (catalytic activity), GO:0008152 (metabolic process), GO:0030170 (pyridoxal phosphate binding)	NA
								<b>Phvul.007G205000</b> (32,755,598..32,759,037) uncharacterized protein LOC100793929 isoform X3 [Glycine max]	NA
								<b>Phvul.007G205100</b> (32,759,933..32,764,966) aspartic proteinase A1; IPR001461 (Aspartic peptidase), IPR011001 (Saposin-like), IPR021109 (Aspartic peptidase domain); GO:0004190 (aspartic-type endopeptidase activity), GO:0006508 (proteolysis), GO:0006629 (lipid metabolic process)	NA
								<b>Phvul.007G205200</b> (32,776,415..32,776,920) hypothetical protein	NA
								<b>Phvul.007G205300</b> (32,782,674..32,783,116) hypothetical protein	NA
								<b>Phvul.007G205400</b> (32,799,930..32,800,355) hypothetical protein	NA
								<b>Phvul.007G205500</b> (32,804,527..32,805,000) MADS-box transcription factor family protein; IPR002100 (Transcription factor, MADS-box); GO:0003677 (DNA binding), GO:0046983 (protein dimerization activity)	NA
<b>Phvul.007G205600</b> (32,807,621..32,807,960) hypothetical protein	NA								

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.007G205700</b> (32,821,913..32,822,392) MADS-box transcription factor family protein; IPR002100 (Transcription factor, MADS-box), IPR002487 (Transcription factor, K-box); GO:0003677 (DNA binding), GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0005634 (nucleus), GO:0046983 (protein dimerization activity)	NA
								<b>Phvul.007G205800</b> (32,824,588..32,826,812) anaphase-promoting complex subunit 11-like [Glycine max]; IPR013083 (Zinc finger, RING/FYVE/PHD-type), IPR024991 (Anaphase-promoting complex subunit 11); GO:0004842 (ubiquitin-protein ligase activity), GO:0005515 (protein binding), GO:0005680 (anaphase-promoting complex), GO:0008270 (zinc ion binding)	NA
								<b>Phvul.007G205901</b> (32,828,086..32,829,220) Low temperature and salt responsive protein family; IPR000612 (Proteolipid membrane potential modulator); GO:0016021 (integral component of membrane)	NA
11	30,570,090	3.31281E-09	T/C	4.3	4.1	-0.13	Elora	<b>Phvul.011G137200</b> (30,519,900..30,528,790) cyclic nucleotide-gated channel 15; IPR014710 (RmlC-like jelly roll fold)	NA
2	26,691,470	2.41272E-10	G/C	3.6	4.1	0.51	Belwood	<b>Phvul.002G125900</b> (26,654,301..26,654,821) uncharacterized protein LOC102662259 [Glycine max]	NA
								<b>Phvul.002G126000</b> (26,724,113..26,724,667) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.002G126100</b> (26,726,180..26,728,487) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.002G126200</b> (26,728,908..26,731,648) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.002G126300</b> (26,738,204..26,738,911) Unknown protein; IPR001810 (F-box domain); GO:0005515 (protein binding)	NA

P v	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
2	34,858,759	7.49118E-11	G/A	3.6	4.4	0.79	Belwood	<b>Phvul.002G185800</b> (34,817,130..34,819,288) glutamate decarboxylase; IPR002129 (Pyridoxal phosphate-dependent decarboxylase), IPR015424 (Pyridoxal phosphate-dependent transferase); GO:0003824 (catalytic activity), GO:0004351 (glutamate decarboxylase activity), GO:0006536 (glutamate metabolic process), GO:0016831 (carboxy-lyase activity), GO:0019752 (carboxylic acid metabolic process), GO:0030170 (pyridoxal phosphate binding)	NA
								<b>Phvul.002G185900</b> (34,830,314..34,832,505) glutamate decarboxylase; IPR002129 (Pyridoxal phosphate-dependent decarboxylase), IPR015424 (Pyridoxal phosphate-dependent transferase); GO:0003824 (catalytic activity), GO:0004351 (glutamate decarboxylase activity), GO:0006536 (glutamate metabolic process), GO:0016831 (carboxy-lyase activity), GO:0019752 (carboxylic acid metabolic process), GO:0030170 (pyridoxal phosphate binding)	NA
								<b>Phvul.002G186000</b> (34,836,879..34,838,695) glutamate decarboxylase; IPR002129 (Pyridoxal phosphate-dependent decarboxylase), IPR015424 (Pyridoxal phosphate-dependent transferase); GO:0003824 (catalytic activity), GO:0004351 (glutamate decarboxylase activity), GO:0006536 (glutamate metabolic process), GO:0016831 (carboxy-lyase activity), GO:0019752 (carboxylic acid metabolic process), GO:0030170 (pyridoxal phosphate binding)	NA
								<b>Phvul.002G186100</b> (34,841,422..34,848,885) sequence-specific DNA binding transcription factors; DNA binding; DNA binding; IPR008895 (YL1 nuclear), IPR013272 (YL1 nuclear, C-terminal); GO:0005634 (nucleus)	NA
								<b>Phvul.002G186200</b> (34,858,350..34,864,576) tobamovirus multiplication protein 3; IPR009457 (Domain of unknown function DUF1084)	NA
								<b>Phvul.002G186300</b> (34,873,102..34,877,094) zinc finger CCCH domain-containing protein 17-like [Glycine max]; IPR000571 (Zinc finger, CCCH-type); GO:0046872 (metal ion binding)	NA
								<b>Phvul.002G186400</b> (34,887,473..34,897,791) Remorin family protein; IPR005516 (Remorin, C-terminal)	NA

P v	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
4	28,756,110	1.91238E-10	A/T	3.7	2.8	-0.84	Belwood	NA	NA
5	13,923,752	3.26437E-12	C/T	3.7	3.5	-0.19	Belwood	NA	NA
5	28,555,185	6.20753E-07	C/T	3.6	4.3	0.68	Belwood	NA	NA
7	14,622,857	4.76202E-09	A/G	3.6	4.1	0.45	Belwood	<b>Phvul.007G128300</b> (14,654,286..14,657,642) electron transfer flavoprotein beta; IPR012255 (Electron transfer flavoprotein, beta subunit); GO:0009055 (electron carrier activity)	NA
								<b>Phvul.007G128400</b> (14,613,444..14,614,383) C2 and GRAM domain-containing protein At1g03370-like isoform X2 [Glycine max]; IPR000008 (C2 domain); GO:0005515 (protein binding)	NA
								<b>Phvul.007G128500</b> (14,579,583..14,581,469) Unknown protein	NA
7	21,374,917	9.67267E-22	A/G	3.7	3.4	-0.34	Belwood	<b>Phvul.007G135800</b> (21,378,776..21,380,409) UDP-Glycosyltransferase superfamily protein; IPR02213 (UDP-glucuronosyl/UDP-glucosyltransferase); GO:0008152 (metabolic process)	(Souza et al., 2019)
								<b>Phvul.007G135900</b> (21,419,806..21,420,569) ethylene-responsive transcription factor 1B; IPR016177 (DNA-binding domain); GO:0003677 (DNA binding), GO:0003700 (sequence-specific DNA binding transcription factor activity)	(Ayyappan et al., 2015)
8	60,524,571	1.54867E-08	T/G	3.7	3.4	-0.30	Belwood	<b>Phvul.008G257800</b> (60,477,863..60,480,589) cinnamoyl coa reductase 1; IPR001509 (NAD-dependent epimerase/dehydratase), IPR002347 (Glucose/ribitol dehydrogenase); GO:0003824 (catalytic activity), GO:0044237 (cellular metabolic process), GO:0050662 (coenzyme binding)	NA
								<b>Phvul.008G257900</b> (60,483,888..60,486,558) cinnamoyl coa reductase 1; IPR001509 (NAD-dependent epimerase/dehydratase), IPR016040 (NAD(P)-binding domain); GO:0003824 (catalytic activity), GO:0044237 (cellular metabolic process), GO:0050662 (coenzyme binding)	NA
								<b>Phvul.008G258000</b> (60,491,019..60,493,275) cinnamoyl coa reductase 1; IPR001509 (NAD-dependent epimerase/dehydratase), IPR016040 (NAD(P)-binding domain); GO:0003824 (catalytic activity), GO:0044237 (cellular metabolic process), GO:0050662 (coenzyme binding)	NA

P v	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.008G258100</b> (60,503,234..60,505,349) ferritin 4; IPR001519 (Ferritin), IPR008331 (Ferritin/DPS protein domain), IPR009078 (Ferritin-like superfamily); GO:0006826 (iron ion transport), GO:0006879 (cellular iron ion homeostasis), GO:0008199 (ferric iron binding)	NA
								<b>Phvul.008G258200</b> (60,526,637..60,527,287) uncharacterized protein LOC100775798 [Glycine max]; IPR006936 (Domain of unknown function DUF640)	NA
								<b>Phvul.008G258300</b> (60,542,133..60,546,276) FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages ; IPR011687 (P60-like)	NA
								<b>Phvul.008G258400</b> (60,549,093..60,553,420) Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain); GO:0004672 (protein kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)	NA
								<b>Phvul.008G258500</b> (60,556,953..60,564,128) ubiquitin activating enzyme 2; IPR000011 (Ubiquitin/SUMO-activating enzyme E1), IPR018075 (Ubiquitin-activating enzyme, E1), IPR018965 (Ubiquitin-activating enzyme e1, C-terminal), IPR023280 (Ubiquitin-like 1 activating enzyme, catalytic cysteine domain); GO:0003824 (catalytic activity), GO:0005524 (ATP binding), GO:0006464 (cellular protein modification process), GO:0008641 (small protein activating enzyme activity)	NA
								<b>Phvul.008G258600</b> (60,565,710..60,567,192) uncharacterized protein At5g39865-like [Glycine max]; IPR012336 (Thioredoxin-like fold); GO:0009055 (electron carrier activity), GO:0015035 (protein disulfide oxidoreductase activity), GO:0045454 (cell redox homeostasis)	NA
<b>10</b>	3,745,502	8.24788E-07	G/A	3.5	3.9	0.39	Belwood	<b>Phvul.010G025300</b> (3,700,659..3,703,172) Disease resistance protein (TIR-NBS-LRR class) family	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.010G025400</b> (3,703,928..3,707,893) Disease resistance protein (TIR-NBS-LRR class) family; IPR000157 (Toll/interleukin-1 receptor homology (TIR) domain), IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0000166 (nucleotide binding), GO:0005515 (protein binding), GO:0006952 (defense response), GO:0007165 (signal transduction), GO:0017111 (nucleoside-triphosphatase activity), GO:0043531 (ADP binding)	NA
								<b>Phvul.010G025500</b> (3,724,245..3,728,360) Disease resistance protein (TIR-NBS-LRR class) family; IPR000157 (Toll/interleukin-1 receptor homology (TIR) domain), IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0000166 (nucleotide binding), GO:0005515 (protein binding), GO:0006952 (defense response), GO:0007165 (signal transduction), GO:0017111 (nucleoside-triphosphatase activity), GO:0043531 (ADP binding)	(Zuiderveen et al., 2016)
								<b>Phvul.010G025600</b> (3,729,489..3,729,743) uncharacterized protein LOC102660652 [Glycine max]	NA
								<b>Phvul.010G025700</b> (3,749,905..3,754,066) Disease resistance protein (TIR-NBS-LRR class) family; IPR000157 (Toll/interleukin-1 receptor homology (TIR) domain), IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0005515	NA
								<b>Phvul.010G026300</b> (3,776,788..3,779,101) Disease resistance protein (TIR-NBS-LRR class) family; IPR000157 (Toll/interleukin-1 receptor homology (TIR) domain); GO:0005515 (protein binding), GO:0007165 (signal transduction)	(Oblessuc et al., 2015)
								<b>Phvul.010G026350</b> (3,783,871..3,785,659) disease resistance protein (TIR-NBS-LRR class)	NA
10	24,982,398	7.73573E-07	G/T	3.7	3.3	-0.42	Belwood	<b>Phvul.010G066400</b> (25,019,438..25,021,034) Non-specific lipid-transfer protein, putative; IPR000528 (Plant lipid transfer protein/Par allergen), IPR016140 (Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain); GO:0006869 (lipid transport), GO:0008289 (lipid binding)	NA

P v	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.010G066500</b> (24,968,137..24,969,974) Non-specific lipid-transfer protein, putative; IPR000528 (Plant lipid transfer protein/Par allergen), IPR016140 (Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain); GO:0006869 (lipid transport), GO:0008289 (lipid binding)	NA
11	27,645,168	3.81806E-08	C/T	3.6	4.0	0.39	Belwood	<b>Phvul.011G122100</b> (27,670,943..27,672,231) O-methyltransferase family protein; IPR016461 (Caffeate O-methyltransferase (COMT) family); GO:0008168 (methyltransferase activity), GO:0008171 (O-methyltransferase activity), GO:0046983 (protein dimerization activity)	NA
2	33,210,458	3.17201E-07	G/A	3.9	4.1	0.26	Isabela	<b>Phvul.002G174200</b> (33,173,791..33,177,832) Ribosomal protein L17 component of cytosolic 80S ribosome and 60S large subunit n=1 Tax=Coccomyxa subellipsoidea C-169 RepID=I0YUE4_9CHLO; IPR001063 (Ribosomal protein L22/L17); GO:0003735 (structural constituent of ribosome), GO:0005622 (intracellular), GO:0005840 (ribosome), GO:0006412 (translation), GO:0015934 (large ribosomal subunit)	NA
								<b>Phvul.002G174300</b> (33,174,880..33,175,610) Unknown protein	NA
								<b>Phvul.002G174400</b> (33,190,913..33,191,194) probable signal peptidase complex subunit 1-like isoform X2 [Glycine max]; IPR009542 (Microsomal signal peptidase 12kDa subunit); GO:0005787 (signal peptidase complex), GO:0006465 (signal peptide processing), GO:0008233 (peptidase activity), GO:0016021 (integral component of membrane)	NA
								<b>Phvul.002G174500</b> (33,211,391..33,211,732) Unknown protein	NA
								<b>Phvul.002G174600</b> (33,211,782..33,213,945) kelch repeat F-box protein; IPR001810 (F-box domain), IPR015916 (Galactose oxidase, beta-propeller); GO:0005515 (protein binding)	NA
								<b>Phvul.002G174700</b> (33,218,592..33,229,088) endoplasmic reticulum metallopeptidase-like protein; IPR007484 (Peptidase M28); GO:0006508 (proteolysis), GO:0008233 (peptidase activity)	NA
								<b>Phvul.002G174800</b> (33,233,109..33,244,898) mannosyl-oligosaccharide glucosidase; IPR004888 (Glycoside hydrolase, family 63); GO:0003824 (catalytic activity), GO:0004573 (mannosyl-oligosaccharide glucosidase activity), GO:0009311 (oligosaccharide metabolic process)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.002G174901</b> (33,252,520..33,270,312) mannosyl-oligosaccharide glucosidase; IPR004888 (Glycoside hydrolase, family 63); GO:0003824 (catalytic activity), GO:0004573 (mannosyl-oligosaccharide glucosidase activity), GO:0009311 (oligosaccharide metabolic process)	NA
7	17,225,864	8.73085E-07	C/T	3.9	3.8	-0.17	Isabela	<b>Phvul.007G118300</b> (17,196,574..17,197,901) Tyrosine phosphatase family protein; IPR000387 (Protein-tyrosine/Dual specificity phosphatase), IPR004861 (Protein-tyrosine phosphatase, SIW14-like); GO:0004725 (protein tyrosine phosphatase activity), GO:0016311 (dephosphorylation), GO:0016791 (phosphatase activity)	(Silva et al., 2019)
9	14,695,454	1.63893E-09	A/G	3.9	4.3	0.40	Isabela	<b>Phvul.009G091500</b> (14,644,649..14,645,764) hypothetical protein; IPR016024 (Armadillo-type fold); GO:0005488 (binding)	NA
								<b>Phvul.009G091600</b> (14,648,278..14,660,094) Calcium-dependent lipid-binding (CaLB domain) family protein; IPR000008 (C2 domain); GO:0005515 (protein binding)	NA
								<b>Phvul.009G091800</b> (14,663,209..14,666,950) probable glycosyltransferase isoform X4 [Glycine max]; IPR004263 (Exostosin-like)	NA
								<b>Phvul.009G091900</b> (14,668,774..14,678,371) ubiquitin carboxyl-terminal hydrolase 12-like [Glycine max]; IPR008974 (TRAF-like); GO:0005515 (protein binding)	NA
								<b>Phvul.009G092000</b> (14,698,282..14,699,074) Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; IPR016140 (Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain)	NA
								<b>Phvul.009G092100</b> (14,711,726..14,721,299) Phox (PX) domain-containing protein; IPR001683 (Phox homologous domain); GO:0035091 (phosphatidylinositol binding)	NA
								<b>Phvul.009G092200</b> (14,723,594..14,727,563) adenine phosphoribosyltransferase 5; IPR000836 (Phosphoribosyltransferase domain), IPR005764 (Adenine phosphoribosyl transferase); GO:0003999 (adenine phosphoribosyltransferase activity), GO:0005737 (cytoplasm), GO:0006168 (adenine salvage), GO:0009116 (nucleoside metabolic process)	NA
11	39,678,873	7.99244E-07	G/C	3.9	3.8	-0.08	Isabela	NA	NA

**Table S5.13.** QTL significantly associated with seed nitrogen content (%N) in the DJ subpopulation (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
2	27,672,387	4.85274E-07	T/C	4.18	4.35	0.16	Elora	<b>Phvul.002G132400</b> (27,631,692..27,634,555) isoflavone reductase homolog [Glycine max]; IPR008030 (NmrA-like), IPR016040 (NAD(P)-binding domain)	NA
								<b>Phvul.002G132500</b> (27,634,530..27,637,330) myb family transcription factor APL-like isoform X1 [Glycine max]; IPR009057 (Homeodomain-like), IPR025756 (MYB-CC type transcription factor, LHEQLE-containing domain); GO:0003677 (DNA binding), GO:0003682 (chromatin binding)	NA
								<b>Phvul.002G132600</b> (27,653,160..27,656,746) transmembrane protein, putative; IPR016196 (Major facilitator superfamily domain, general substrate transporter)	NA
								<b>Phvul.002G132700</b> (27,657,989..27,662,222) Structural constituent of ribosome n=3 Tax=Camelineae RepID=D7MTD9_ARALL; IPR021137 (Ribosomal protein L35); GO:0003735 (structural constituent of ribosome), GO:0005622 (intracellular), GO:0005840 (ribosome), GO:0006412 (translation)	NA
								<b>Phvul.002G132800</b> (27,666,600..27,670,376) membrane protein, putative; IPR007300 (CidB/LrgB family)	NA
								<b>Phvul.002G132900</b> (27,670,660..27,675,562) 4-coumarate:CoA ligase 1; IPR000873 (AMP-dependent synthetase/ligase), IPR025110 (AMP-binding enzyme C-terminal domain); GO:0003824 (catalytic activity), GO:0008152 (metabolic process)	NA
								<b>Phvul.002G133000</b> (27,678,020..27,684,786) ERD (early-responsive to dehydration stress) family protein; IPR003864 (Domain of unknown function DUF221), IPR027815 (Domain of unknown function DUF4463); GO:0016020 (membrane)	NA
								<b>Phvul.002G133100</b> (27,698,191..27,701,446) phosphoribulokinase; IPR006082 (Phosphoribulokinase), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0005524 (ATP binding), GO:0005975 (carbohydrate metabolic process), GO:0008152 (metabolic process), GO:0008974 (phosphoribulokinase activity), GO:0016301 (kinase activity)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.002G133200</b> (27,705,799..27,713,779) Secretory carrier membrane protein (SCAMP) family protein; IPR007273 (SCAMP); GO:0015031 (protein transport), GO:0016021 (integral component of membrane)	NA
2	26,691,470	3.99E-08	G/C	3.52	4.14	0.62	Belwood	<b>Phvul.002G125900</b> (26,654,301..26,654,821) uncharacterized protein LOC102662259 [Glycine max]	NA
								<b>Phvul.002G126000</b> (26724113..26724667) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.002G126100</b> (26,726,180..26,728,487) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.002G126200</b> (26,728,908..26,731,648) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	(Silva et al., 2019)
								<b>Phvul.002G126300</b> (26,738,204..26,738,911) Unknown protein; IPR001810 (F-box domain); GO:0005515 (protein binding)	NA
7	9,752,020	5.33E-09	C/T	3.43	3.89	0.46	Belwood	<b>Phvul.007G093200</b> (9,702,355..9,703,579) laccase 10; IPR008972 (Cupredoxin); GO:0016491 (oxidoreductase activity), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.007G093300</b> (9,722,963..9,727,649) serine carboxypeptidase-like 27; IPR001563 (Peptidase S10, serine carboxypeptidase); GO:0004185 (serine-type carboxypeptidase activity), GO:0006508 (proteolysis)	NA
								<b>Phvul.007G093400</b> (9,728,969..9,731,364) tetraspanin-6 [Glycine max]; IPR018499 (Tetraspanin/Peripherin); GO:0016021 (integral component of membrane)	NA
								<b>Phvul.007G093500</b> (9743839..9746684) uncharacterized protein LOC102663792 [Glycine max]; IPR025558 (Domain of unknown function DUF4283)	NA
								<b>Phvul.007G093600</b> (9,765,831..9,770,789) MLO-like protein 12-like [Glycine max]; IPR004326 (Mlo-related protein);	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								GO:0006952 (defense response), GO:0016021 (integral component of membrane)	
								<b>Phvul.007G093700</b> (9,771,326..9,773,695) ubiquitin-conjugating enzyme 20; IPR016135 (Ubiquitin-conjugating enzyme/RWD-like), IPR023313 (Ubiquitin-conjugating enzyme, active site); GO:0016881 (acid-amino acid ligase activity)	NA
								<b>Phvul.007G093800</b> (9,774,382..9,777,710) transmembrane protein 115-like isoform X2 [Glycine max]; IPR013861 (Protein of unknown function DUF1751, integral membrane, eukaryotic), IPR022764 (Peptidase S54, rhomboid domain); GO:0004252 (serine-type endopeptidase activity), GO:0016021 (integral component of membrane)	NA
								<b>Phvul.007G093900</b> (9,778,743..9,788,355) Protein kinase superfamily protein; IPR011009 (Protein kinase-like domain); GO:0004672 (protein kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)	NA
7	37,910,063	3.08E-11	A/G	3.61	3.46	-0.15	Belwood	<b>Phvul.007G257000</b> (37,856,987..37,861,258) OTU-like cysteine protease family protein; IPR003323 (Ovarian tumour, otubain)	NA
								<b>Phvul.007G257100</b> (37,868,875..37,872,419) transcription factor bHLH130-like [Glycine max]; IPR011598 (Myc-type, basic helix-loop-helix (bHLH) domain); GO:0046983 (protein dimerization activity)	(Silva et al., 2019)
								<b>Phvul.007G257200</b> (37,874,864..37,877,377) GTP-binding nuclear protein Ran-3 [Glycine max]; IPR001806 (Small GTPase superfamily), IPR002041 (Ran GTPase), IPR005225 (Small GTP-binding protein domain), IPR024156 (Small GTPase superfamily, ARF type), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003924 (GTPase activity), GO:0005525 (GTP binding), GO:0005622 (intracellular), GO:0006184 (GTP catabolic process), GO:0006886 (intracellular protein transport), GO:0006913 (nucleocytoplasmic transport), GO:0007165 (signal transduction), GO:0007264 (small GTPase mediated signal transduction), GO:0015031 (protein transport), GO:0016020 (membrane)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.007G257300</b> (37,880,156..37,882,107) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.007G257400</b> (37,897,703..37,899,480) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.007G257500</b> (37,905,935..37,909,386) Cytochrome P450 superfamily protein; IPR001128 (Cytochrome P450); GO:0005506 (iron ion binding), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
								<b>Phvul.007G257600</b> (37,912,397..37,915,146) multiple C2 and transmembrane domain-containing protein 2-like [Glycine max]; IPR000008 (C2 domain), IPR013583 (Phosphoribosyltransferase C-terminal); GO:0005515 (protein binding)	NA
								<b>Phvul.007G257700</b> (37,918,571..37,926,187) TGACG-sequence-specific DNA-binding protein TGA-2.1-like isoform X1 [Glycine max]; IPR004827 (Basic-leucine zipper domain), IPR025422 (Transcription factor TGA like domain); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)	NA
								<b>Phvul.007G257800</b> (37,941,819..37,945,562) multiple C2 and transmembrane domain-containing protein 1-like [Glycine max]; IPR000008 (C2 domain), IPR013583 (Phosphoribosyltransferase C-terminal); GO:0005515 (protein binding)	NA
<b>8</b>	37,482,649	1.43E-07	C/T	3.54	3.50	-0.04	Belwood	NA	NA
<b>8</b>	61,434,463	4.67E-12	G/A	3.56	3.43	-0.13	Belwood	<b>Phvul.008G269000</b> (61,382,015..61,384,556) Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein; IPR004038 (Ribosomal protein L7Ae/L30e/S12e/Gadd45), IPR018492 (Ribosomal protein L7Ae/L8/Nhp2 family); GO:0003723 (RNA binding), GO:0005730 (nucleolus), GO:0030529 (ribonucleoprotein complex), GO:0042254 (ribosome biogenesis)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.008G269100</b> (61,386,577..61,394,005) rab3 GTPase-activating protein catalytic subunit-like isoform X1 [Glycine max]; IPR026147 (Rab3 GTPase-activating protein catalytic subunit); GO:0005097 (Rab GTPase activator activity)	NA
								<b>Phvul.008G269200</b> (61,394,369..61,395,392) uncharacterized protein LOC100798628 [Glycine max]	NA
								<b>Phvul.008G269300</b> (61,396,932..61,403,435) dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex 2, mitochondrial-like [Glycine max]; IPR006255 (Dihydrolipoamide succinyltransferase), IPR023213 (Chloramphenicol acetyltransferase-like domain); GO:0004149 (dihydrolipoyllysine-residue succinyltransferase activity), GO:0006099 (tricarboxylic acid cycle), GO:0008152 (metabolic process), GO:0045252 (oxoglutarate dehydrogenase complex)	NA
								<b>Phvul.008G269400</b> (61,404,275..61,405,419) B-cell receptor-associated 31-like; IPR008417 (B-cell receptor-associated 31-like); GO:0005783 (endoplasmic reticulum), GO:0006886 (intracellular protein transport), GO:0016021 (integral component of membrane)	NA
								<b>Phvul.008G269500</b> (61,405,530..61,406,740) Stigma-specific Stig1 family protein; IPR006969 (Stigma-specific protein Stig1)	NA
								<b>Phvul.008G269600</b> (61,408,688..61,412,350) GDP-L-galactose phosphorylase 1-like [Glycine max]	NA
								<b>Phvul.008G269700</b> (61,421,738..61,422,507) DUF4408 domain protein; IPR008480 (Protein of unknown function DUF761, plant), IPR025520 (Domain of unknown function DUF4408)	NA
								<b>Phvul.008G269800</b> (61,427,310..61,429,007) double-stranded RNA-binding motif protein; IPR011907 (Ribonuclease III); GO:0003723 (RNA binding), GO:0004525 (ribonuclease III activity), GO:0006396 (RNA processing), GO:0016075 (rRNA catabolic process)	NA
								<b>Phvul.008G269900</b> (61,429,629..61,434,687) Cell differentiation, Rcd1-like protein; IPR007216 (Rcd1), IPR016024 (Armadillo-type fold); GO:0005488 (binding)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.008G270000</b> (61,436,394..61,439,858) Auxin-responsive family protein; IPR005018 (DOMON domain), IPR017214 (Uncharacterised conserved protein UCP037471)	NA
								<b>Phvul.008G270200</b> (61,449,460..61,465,109) esophageal cancer associated protein, putative	NA
								<b>Phvul.008G270500</b> (61,477,364..61,479,640) WRKY family transcription factor; IPR003657 (DNA-binding WRKY); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)	(Ayyappan et al., 2015)
2	22,757,513	1.43817E-10	T/C	3.76	4.00	0.24	Isabela	<b>Phvul.002G106500</b> (22,720,823..22,731,749) Subunit of Retromer complex n=1 Tax=Volvox carteri RepID=D8U9T4_VOLCA; IPR005378 (Vacuolar protein sorting-associated protein 35, Vps35), IPR016024 (Armadillo-type fold); GO:0005488 (binding), GO:0015031 (protein transport), GO:0030904 (retromer complex)	NA
								<b>Phvul.002G106600</b> (22,733,118..22,736,074) rac-like GTP-binding protein 7-like [Glycine max]; IPR001806 (Small GTPase superfamily), IPR005225 (Small GTP-binding protein domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0005525 (GTP binding), GO:0005622 (intracellular), GO:0006184 (GTP catabolic process), GO:0007165 (signal transduction), GO:0007264 (small GTPase mediated signal transduction), GO:0015031 (protein transport), GO:0016020 (membrane)	(Flores et al., 2018)
								<b>Phvul.002G106700</b> (22,739,113..22,742,577) 3-dehydroquinase synthase, putative; IPR016037 (3-dehydroquinase synthase AroB); GO:0003856 (3-dehydroquinase synthase activity), GO:0005737 (cytoplasm), GO:0009073 (aromatic amino acid family biosynthetic process)	NA
								<b>Phvul.002G106800</b> (22,744,026..22,747,460) Myb/SANT-like DNA-binding domain protein; IPR024752 (Myb/SANT-like domain)	NA
								<b>Phvul.002G106900</b> (22,748,300..22,752,804) uncharacterized protein LOC102662030 [Glycine max]; IPR006867 (Domain of unknown function DUF632), IPR006868 (Domain of unknown function DUF630)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.002G107000</b> (22,753,442..22,755,417) ascorbate peroxidase 3; IPR010255 (Haem peroxidase); GO:0004601 (peroxidase activity), GO:0006979 (response to oxidative stress), GO:0020037 (heme binding), GO:0055114 (oxidation-reduction process)	NA
7	36,757,868	9.38584E-07	G/A	3.82	3.96	0.14	Isabela	<b>Phvul.007G243800</b> (36,717,412..36,723,284) uncharacterized protein LOC100818532 isoform X1 [Glycine max]	NA
								<b>Phvul.007G244000</b> (36,730,587..36,736,485) uncharacterized protein LOC100797206 isoform X7 [Glycine max]; IPR018971 (Protein of unknown function DUF1997)	NA
								<b>Phvul.007G244033</b> (36,741,509..36,742,084) transmembrane protein, putative	NA
								<b>Phvul.007G244066</b> (36,745,593..36,746,239) transmembrane protein, putative	NA
								<b>Phvul.007G244100</b> (36,757,553..36,758,622) two-component response regulator ARR2-like isoform X1 [Glycine max]; IPR009057 (Homeodomain-like); GO:0003677 (DNA binding), GO:0003682 (chromatin binding)	NA
								<b>Phvul.007G244200</b> (36,762,504..36,766,437) Single-stranded nucleic acid binding R3H protein; IPR001374 (Single-stranded nucleic acid binding R3H), IPR024771 (SUZ domain), IPR024773 (R3H domain-containing protein, viridiplantae); GO:0003676 (nucleic acid binding)	NA
								<b>Phvul.007G244300</b> (36,767,770..36,776,959) serine/threonine-protein kinase HT1-like [Glycine max]; IPR001932 (Protein phosphatase 2C (PP2C)-like domain), IPR011009 (Protein kinase-like domain); GO:0003824 (catalytic activity), GO:0004672 (protein kinase activity), GO:0004674 (protein serine/threonine kinase activity), GO:0005524 (ATP binding), GO:0006468 (protein phosphorylation)	NA
								<b>Phvul.007G244400</b> (36,778,207..36,778,854) zinc finger protein 4-like [Glycine max]; IPR013087 (Zinc finger C2H2-type/integrase DNA-binding domain); GO:0003676 (nucleic acid binding), GO:0046872 (metal ion binding)	NA
								<b>Phvul.007G244500</b> (36,781,647..36,784,147) Ribosomal protein L12 family protein; IPR000206 (Ribosomal protein L7/L12); GO:0003735 (structural constituent of ribosome), GO:0005622 (intracellular), GO:0005840 (ribosome), GO:0006412 (translation)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.007G244600</b> (36,784,809..36,786,912) Major facilitator superfamily protein; IPR010658 (Nodulin-like), IPR016196 (Major facilitator superfamily domain, general substrate transporter)	(Kamfwa et al., 2017)
								<b>Phvul.007G244700</b> (36,798,403..36,811,427) transcriptional corepressor LEUNIG-like isoform X1 [Glycine max]; IPR006594 (LisH dimerisation motif), IPR015943 (WD40/YVTN repeat-like-containing domain), IPR020472 (G-protein beta WD-40 repeat); GO:0005515 (protein binding)	NA
9	14,695,454	1.23478E-08	A/G	3.84	4.28	0.45	Isabela	<b>Phvul.009G091500</b> (14,644,649..14,645,764) hypothetical protein; IPR016024 (Armadillo-type fold); GO:0005488 (binding)	NA
								<b>Phvul.009G091600</b> (14,648,278..14,660,094) Calcium-dependent lipid-binding (CaLB domain) family protein; IPR000008 (C2 domain); GO:0005515 (protein binding)	NA
								<b>Phvul.009G091800</b> (14,663,209..14,666,950) probable glycosyltransferase isoform X4 [Glycine max]; IPR004263 (Exostosin-like)	NA
								<b>Phvul.009G091900</b> (14,668,774..14,678,371) ubiquitin carboxyl-terminal hydrolase 12-like [Glycine max]; IPR008974 (TRAF-like); GO:0005515 (protein binding)	NA
								<b>Phvul.009G092000</b> (14,698,282..14,699,074) Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; IPR016140 (Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain)	NA
								<b>Phvul.009G092100</b> (14,711,726..14,721,299) Phox (PX) domain-containing protein; IPR001683 (Phox homologous domain); GO:0035091 (phosphatidylinositol binding)	NA
								<b>Phvul.009G092200</b> (14,723,594..14,727,563) adenine phosphoribosyltransferase 5; IPR000836 (Phosphoribosyltransferase domain), IPR005764 (Adenine phosphoribosyl transferase); GO:0003999 (adenine phosphoribosyltransferase activity), GO:0005737 (cytoplasm), GO:0006168 (adenine salvage), GO:0009116 (nucleoside metabolic process)	NA

**Table S5.14.** QTL significantly associated with leaf chlorophyll content at flowering time in the entire MDP (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers. (Pv – *Phaseolus vulgaris* chromosome. NA – not available).

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
2	21,626,534	1.44477E-11	A/T	39.1	52.4	13.26	Elora	<b>Phvul.002G102200</b> (21,616,926..21,626,939) alpha/beta fold hydrolase; IPR000073 (Alpha/beta hydrolase fold-1), IPR000639 (Epoxide hydrolase-like); GO:0003824 (catalytic activity)	NA
								<b>Phvul.002G102300</b> (21,621,335..21,625,818) trehalose phosphate synthase; IPR001830 (Glycosyl transferase, family 20), IPR006379 (HAD-superfamily hydrolase, subfamily IIB), IPR023214 (HAD-like domain); GO:0003824 (catalytic activity), GO:0005992 (trehalose biosynthetic process), GO:0008152 (metabolic process)	NA
								<b>Phvul.002G102400</b> (21,647,580..21,649,708) glucan endo-1,3-beta-glucosidase-like protein 2-like [Glycine max]; IPR012946 (X8)	NA
								<b>Phvul.002G102500</b> (21,659,524..21,663,748) general regulatory factor 12; IPR000308 (14-3-3 protein), IPR023409 (14-3-3 protein, conserved site), IPR023410 (14-3-3 domain); GO:0019904 (protein domain specific binding)	(Li et al., 2015)
								<b>Phvul.002G102600</b> (21,670,579..21,676,059) probable methyltransferase PMT5-like [Glycine max]; IPR004159 (Putative S-adenosyl-L-methionine-dependent methyltransferase); GO:0008168 (methyltransferase activity)	NA
10	25,230,082	5.4613E-11	C/T	38.3	45.5	7.21	Elora	NA	NA
7	972,126	6.72854E-07	G/A	39.9	39.9	0.00	Isabela	<b>Phvul.007G013100</b> (929,432..929,942) adenine nucleotide alpha hydrolase-like domain kinase	NA
								<b>Phvul.007G013200</b> (930,387..931,521) Glutathione S-transferase family protein; IPR010987 (Glutathione S-transferase, C-terminal-like), IPR012336 (Thioredoxin-like fold); GO:0005515 (protein binding)	NA
								<b>Phvul.007G013300</b> (933,598..936,793) aluminum-activated, malate transporter 12; IPR020966 (Aluminum-activated malate transporter); GO:0015743 (malate transport)	NA
								<b>Phvul.007G013400</b> (939,889..942,157) Coiled-coil domain-containing protein 55 (DUF2040); IPR018612 (Domain of unknown function DUF2040)	NA

P v	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.007G013500</b> (942,925..947,241) aluminum-activated, malate transporter 12; IPR020966 (Aluminum-activated malate transporter); GO:0015743 (malate transport)	NA
								<b>Phvul.007G013600</b> (948,959..956,402) Calcium-binding EF hand family protein; IPR011992 (EF-hand domain pair); GO:0005509 (calcium ion binding)	NA
								<b>Phvul.007G013700</b> (960,405..961,318) GRAM domain-containing protein / ABA-responsive protein-related; IPR004182 (GRAM domain)	NA
								<b>Phvul.007G013800</b> (963,495..965,050) GRAM domain-containing protein / ABA-responsive protein-related; IPR004182 (GRAM domain)	NA
								<b>Phvul.007G013900</b> (967,973..969,291) GRAM domain-containing protein / ABA-responsive protein-related; IPR004182 (GRAM domain), IPR011993 (Pleckstrin homology-like domain)	NA
								<b>Phvul.007G014000</b> (973,494..974,918) GRAM domain-containing protein / ABA-responsive protein-related; IPR004182 (GRAM domain), IPR011993 (Pleckstrin homology-like domain)	NA
								<b>Phvul.007G014100</b> (980,956..982,483) RING finger protein 126-A-like [Glycine max]; IPR013083 (Zinc finger, RING/FYVE/PHD-type); GO:0005515 (protein binding), GO:0008270 (zinc ion binding)	NA
								<b>Phvul.007G014300</b> (987,814..990,978) transmembrane protein 53 [Glycine max]; IPR008547 (Protein of unknown function DUF829, TMEM53)	NA
								<b>Phvul.007G014400</b> (992,067..994,435) Cyclic pyranopterin monophosphate synthase accessory protein n=2 Tax=Clostridium ReplD=D8GLY9_CLOLD; IPR002820 (Molybdopterin cofactor biosynthesis C (MoaC) domain), IPR023045 (Molybdenum cofactor biosynthesis C); GO:0006777 (Mo-molybdopterin cofactor biosynthetic process)	NA
								<b>Phvul.007G014500</b> (995,205..999,262) Riboflavin biosynthesis protein ribF (Riboflavin kinase / FMN adenylyltransferase) n=1 Tax=Synechococcus sp. (strain RCC307) ReplD=A5GUT2_SYNR3; IPR014729 (Rossmann-like alpha/beta/alpha sandwich fold); GO:0003919 (FMN adenylyltransferase activity), GO:0009231 (riboflavin biosynthetic process)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.007G014600</b> (999,729..1,011,442) myosin 1; IPR000048 (IQ motif, EF-hand binding site), IPR001609 (Myosin head, motor domain), IPR027401 (Myosin-like IQ motif-containing domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003774 (motor activity), GO:0005515 (protein binding), GO:0005524 (ATP binding), GO:0016459 (myosin complex)	NA
								<b>Phvul.007G014700</b> (1,015,891..1,019,531) dihydroorotate dehydrogenase (quinone); IPR012135 (Dihydroorotate dehydrogenase, class 1/2), IPR013785 (Aldolase-type TIM barrel); GO:0003824 (catalytic activity), GO:0004152 (dihydroorotate dehydrogenase activity), GO:0004158 (dihydroorotate oxidase activity), GO:0006207 ('de novo' pyrimidine nucleobase biosynthetic process), GO:0006222 (UMP biosynthetic process), GO:0016020 (membrane), GO:0055114 (oxidation-reduction process)	NA
7	38,392,817	2.39191E-07	A/G	39.8	44.7	4.93	Isabela	<b>Phvul.007G261500</b> (38,351,561..38,353,215) zinc finger protein 8; IPR007087 (Zinc finger, C2H2); GO:0046872 (metal ion binding)	NA
								<b>Phvul.007G261600</b> (38,367,385..38,371,647) unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: membrane	NA
								<b>Phvul.007G261700</b> (38,373,507..38,374,058) Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)	NA
								<b>Phvul.007G261800</b> (38,382,553..38,390,625) carbon catabolite repressor-like protein; IPR005135 (Endonuclease/exonuclease/phosphatase)	NA
								<b>Phvul.007G261900</b> (38,392,523..38,397,577) F-box/LRR protein; IPR006553 (Leucine-rich repeat, cysteine-containing subtype)	NA
								<b>Phvul.007G262000</b> (38,400,117..38,405,834) DEAD-box ATP-dependent RNA helicase-like protein; IPR001650 (Helicase, C-terminal), IPR014001 (Helicase, superfamily 1/2, ATP-binding domain), IPR014014 (RNA helicase, DEAD-box type, Q motif), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0003676 (nucleic acid binding), GO:0004386 (helicase activity), GO:0005524 (ATP binding), GO:0008026 (ATP-dependent helicase activity)	NA

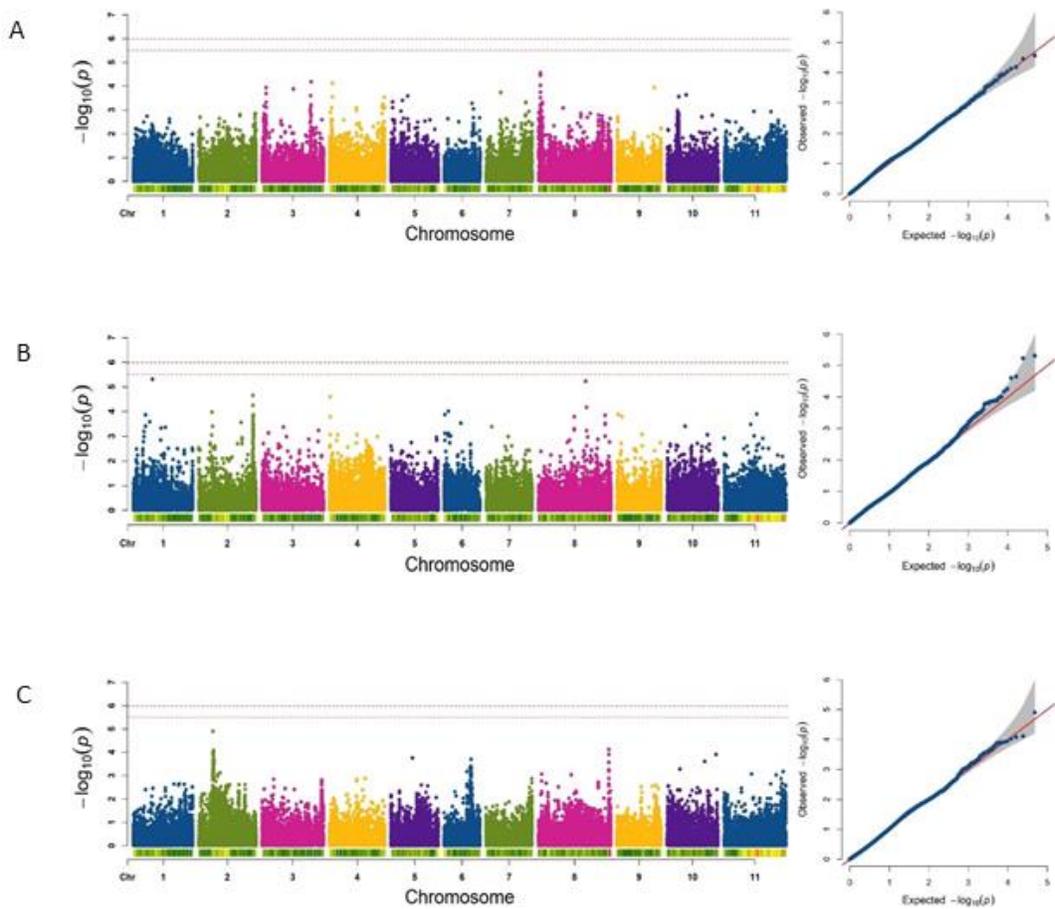
Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.007G262101</b> (38,408,012..38,408,920) Dihydropterin pyrophosphokinase / Dihydropteroate synthase; IPR011005 (Dihydropteroate synthase-like); GO:0004156 (dihydropteroate synthase activity), GO:0009396 (folic acid-containing compound biosynthetic process), GO:0042558 (pteridine-containing compound metabolic process), GO:0044237 (cellular metabolic process)	NA
								<b>Phvul.007G262200</b> (38,413,456..38,420,515) serine/threonine protein phosphatase 2A; IPR004843 (Phosphoesterase domain); GO:0016787 (hydrolase activity)	NA
								<b>Phvul.007G262300</b> (38,423,486..38,428,139) fatty acyl-CoA reductase 2-like [Glycine max]; IPR016040 (NAD(P)-binding domain), IPR026055 (Fatty acyl-CoA reductase); GO:0080019 (fatty-acyl-CoA reductase (alcohol-forming) activity)	NA
								<b>Phvul.007G262400</b> (38,432,084..38,437,914) histone-lysine N-methyltransferase ATXR4-like isoform X1 [Glycine max]; IPR001214 (SET domain); GO:0005515 (protein binding)	NA
8	55,420,128	6.42256E-07	A/T	39.9	40.5	0.56	Isabela	<b>Phvul.008G206000</b> (55,373,942..55,376,044) probable calcium-binding protein CML20 [Glycine max]; IPR011992 (EF-hand domain pair); GO:0005509 (calcium ion binding)	NA
								<b>Phvul.008G206100</b> (55,381,672..55,386,848) SNF1-related kinase regulatory subunit beta-2; IPR006828 (5-AMP-activated protein kinase, beta subunit, interaction domain), IPR014756 (Immunoglobulin E-set); GO:0005515 (protein binding)	NA
								<b>Phvul.008G206200</b> (55,407,463..55,410,092) RNA-binding protein 24-B-like isoform X1 [Glycine max]; IPR012677 (Nucleotide-binding, alpha-beta plait); GO:0000166 (nucleotide binding), GO:0003676 (nucleic acid binding)	NA
								<b>Phvul.008G206300</b> (55,413,647..55,422,607) Ia-related protein 1B-like [Glycine max]; IPR006607 (Protein of unknown function DM15), IPR011991 (Winged helix-turn-helix DNA-binding domain)	NA
								<b>Phvul.008G206400</b> (55,425,721..55,427,719) MaoC domain protein dehydratase n=2 Tax=Clostridium RepID=A6LV26_CLOB8; IPR002539 (MaoC-like domain); GO:0008152 (metabolic process), GO:0016491 (oxidoreductase activity)	NA

Pv	SNP rs# (position, bp)	minus log10 (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location in which association significant	Candidate gene, location (bp), annotation	References
								<p><b>Phvul.008G206500</b> (55,428,446..55,435,190) Argonaute family protein; IPR003100 (Argonaute/Dicer protein, PAZ domain), IPR012337 (Ribonuclease H-like domain), IPR014811 (Domain of unknown function DUF1785); GO:0003676 (nucleic acid binding), GO:0005515 (protein binding)</p>	NA
								<p><b>Phvul.008G206600</b> (55,451,810..55,459,628) Argonaute family protein; IPR003100 (Argonaute/Dicer protein, PAZ domain), IPR012337 (Ribonuclease H-like domain), IPR014811 (Domain of unknown function DUF1785); GO:0003676 (nucleic acid binding), GO:0005515 (protein binding)</p>	NA
11	25,870,481	2.03006E-07	C/T	39.9	40.2	0.36	Isabela	NA	NA

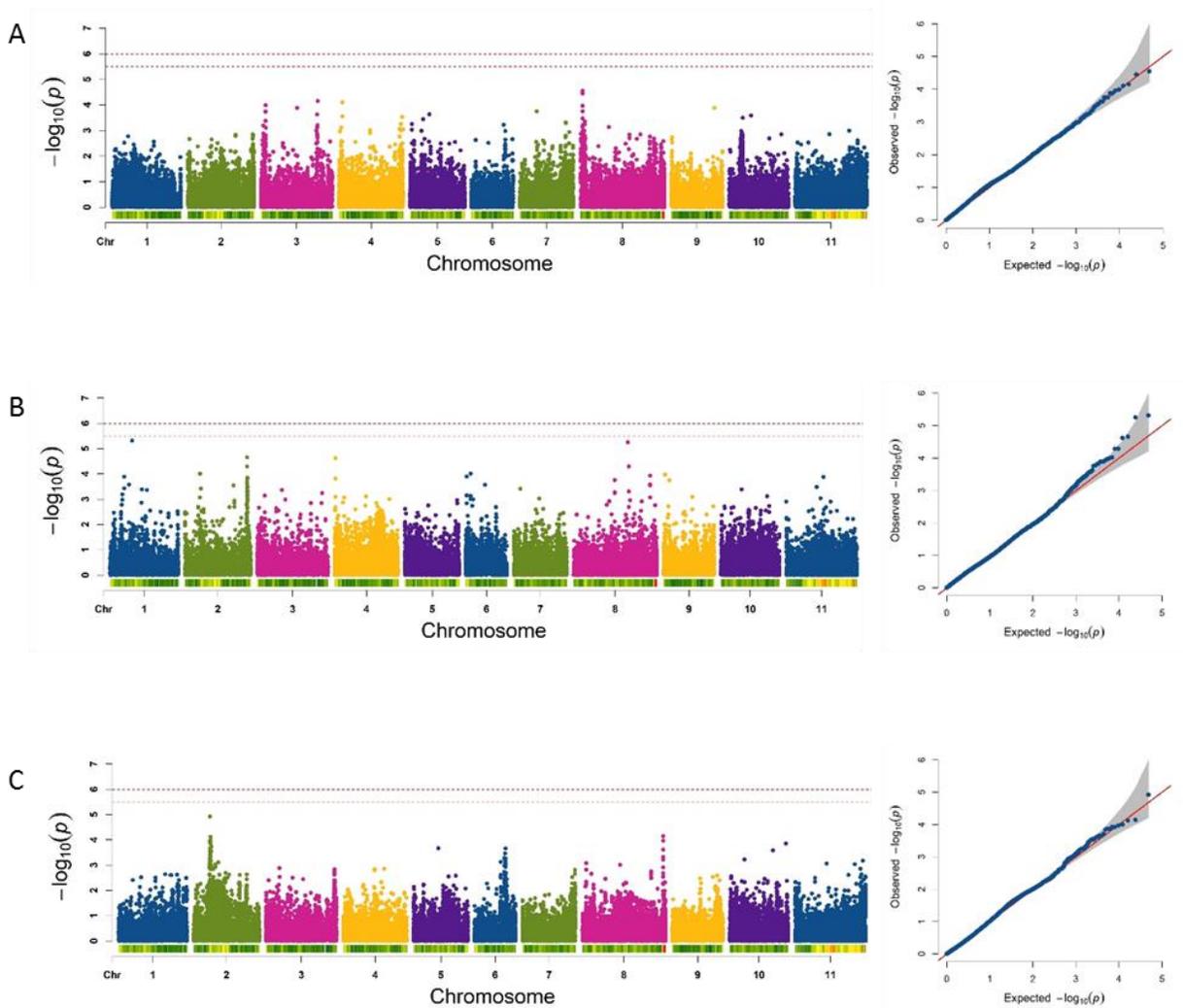
**Table S5.15.** QTL significantly associated with leaf chlorophyll content at flowering time (SPAD2) in the MA subpopulation (5% FDR). JBrowse was used to identify candidate genes in 100 Kb regions centered on significant markers.

Pv	SNP rs# (position, bp)	minus log <sub>10</sub> (P-value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location at which association significant	Candidate gene, location (bp), annotation	References
9	33,332,147	6.79E-07	T/C	37.4	39.6	-2.18	Isabela	<b>Phvul.009G220400</b> (33,254,933..33,299,183) HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase; IPR008380 (HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase), IPR023214 (HAD-like domain)	NA
								<b>Phvul.009G220500</b> (33,303,370..33,311,383) protein CHUP1, chloroplastic-like isoform X2 [Glycine max]	NA
								<b>Phvul.009G220600</b> (33,312,030..33,316,336) aspartic proteinase A1; IPR001461 (Aspartic peptidase), IPR011001 (Saposin-like), IPR021109 (Aspartic peptidase domain); GO:0004190 (aspartic-type endopeptidase activity), GO:0006508 (proteolysis), GO:0006629 (lipid metabolic process)	NA
								<b>Phvul.009G220700</b> (33,317,956..33,322,044) WRKY family transcription factor; IPR003657 (DNA-binding WRKY); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)	NA
								<b>Phvul.009G220800</b> (33,334,065..33,337,835) CMP/dCMP deaminase zinc-binding protein n=2 Tax=Paenibacillus RepID=E5YZ31_9BACL; IPR016193 (Cytidine deaminase-like); GO:0003824 (catalytic activity), GO:0008270 (zinc ion binding), GO:0016787 (hydrolase activity)	NA
								<b>Phvul.009G220900</b> (33,369,569..33,371,883) uncharacterized protein LOC100781058 isoform X2 [Glycine max]	NA
								<b>Phvul.009G221000</b> (33,371,392..33,373,158) Thioesterase superfamily protein; IPR006683 (Thioesterase superfamily)	NA
								<b>Phvul.009G221100</b> (33,377,281..33,377,681) uncharacterized protein LOC102668721 [Glycine max]	NA
9	33,332,084	7.08E-07	A/G	37.4	39.5	-2.11	Isabela	<b>Phvul.009G220400</b> (33,254,933..33,299,183) HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase; IPR008380 (HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase), IPR023214 (HAD-like domain)	NA

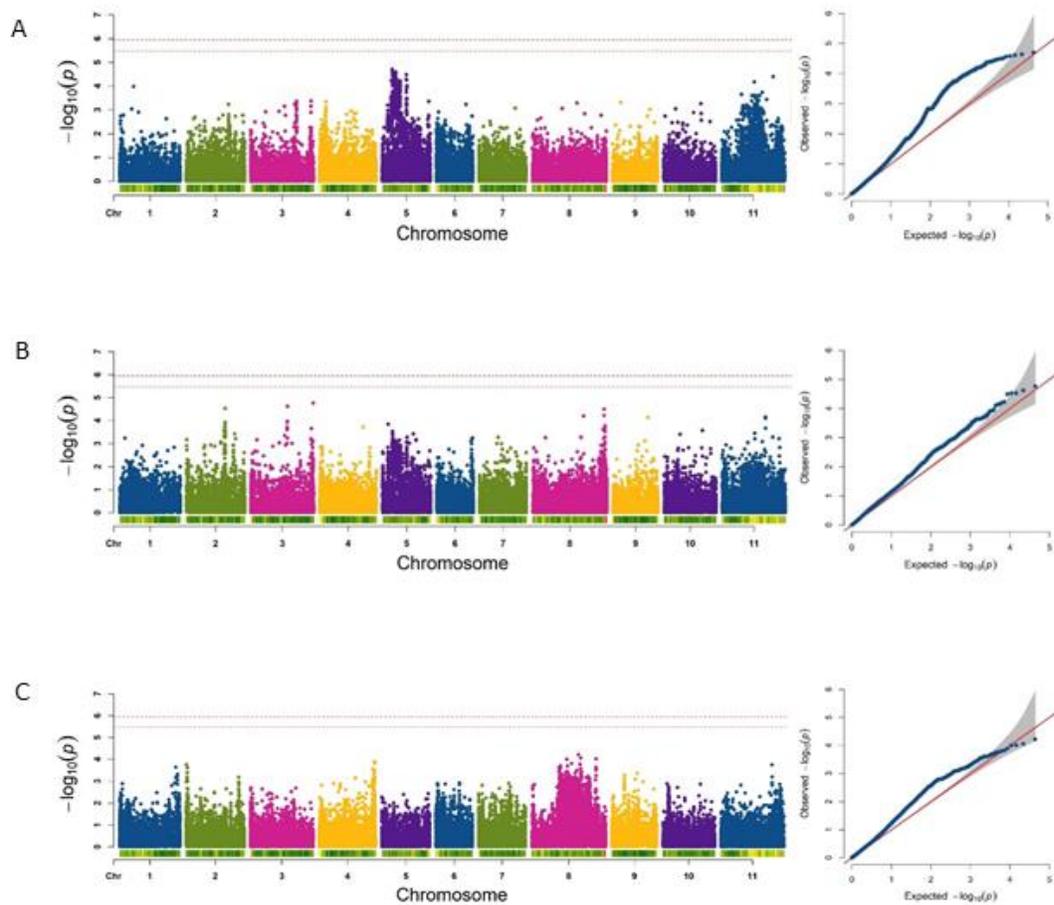
Pv	SNP rs# (position, bp)	minus log10 (P- value)	SNP	First allele homozygous trait average	Second allele homozygous trait average	SNP effect	Location at which association significant	Candidate gene, location (bp), annotation	References
								<b>Phvul.009G220500</b> (33,303,370..33,311,383) protein CHUP1, chloroplastic-like isoform X2 [Glycine max]	NA
								<b>Phvul.009G220600</b> (33,312,030..33,316,336) aspartic proteinase A1; IPR001461 (Aspartic peptidase), IPR011001 (Saposin-like), IPR021109 (Aspartic peptidase domain); GO:0004190 (aspartic-type endopeptidase activity), GO:0006508 (proteolysis), GO:0006629 (lipid metabolic process)	NA
								<b>Phvul.009G220700</b> (33,317,956..33,322,044) WRKY family transcription factor; IPR003657 (DNA-binding WRKY); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)	NA
								<b>Phvul.009G220800</b> (33,334,065..33,337,835) CMP/dCMP deaminase zinc-binding protein n=2 Tax=Paenibacillus RepID=E5YZ31_9BACL; IPR016193 (Cytidine deaminase-like); GO:0003824 (catalytic activity), GO:0008270 (zinc ion binding), GO:0016787 (hydrolase activity)	NA
								<b>Phvul.009G220900</b> (33,369,569..33,371,883) uncharacterized protein LOC100781058 isoform X2 [Glycine max]	NA
								<b>Phvul.009G221000</b> (33,371,392..33,373,158) Thioesterase superfamily protein; IPR006683 (Thioesterase superfamily)	NA
								<b>Phvul.009G221100</b> (33,377,281..33,377,681) uncharacterized protein LOC102668721 [Glycine max]	NA



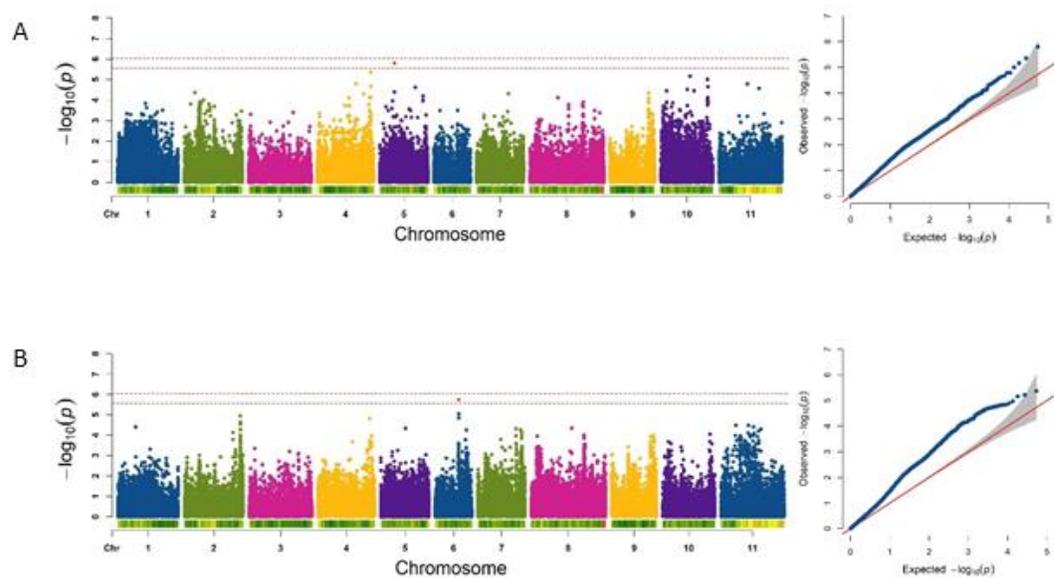
**Figure S5.1.** Manhattan and QQ plots for GWAS of 198 race DJ MDP genotypes and seed nitrogen discrimination ( $\delta^{15}\text{N}$ ) values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.



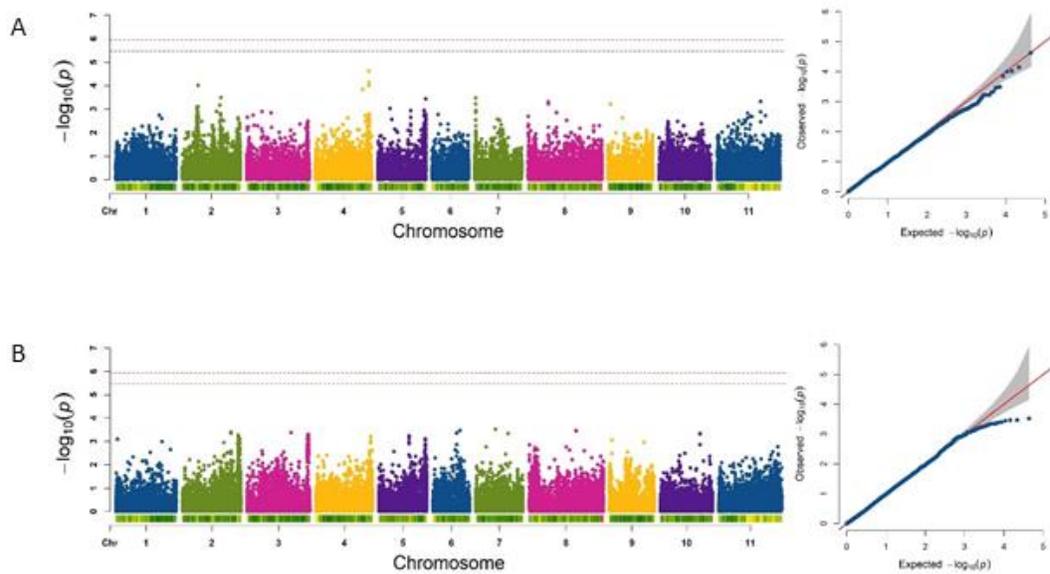
**Figure S5.2.** Manhattan and QQ plots for GWAS of 182 MDP genotypes belonging to the DJ subpopulation and %Ndfa values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.



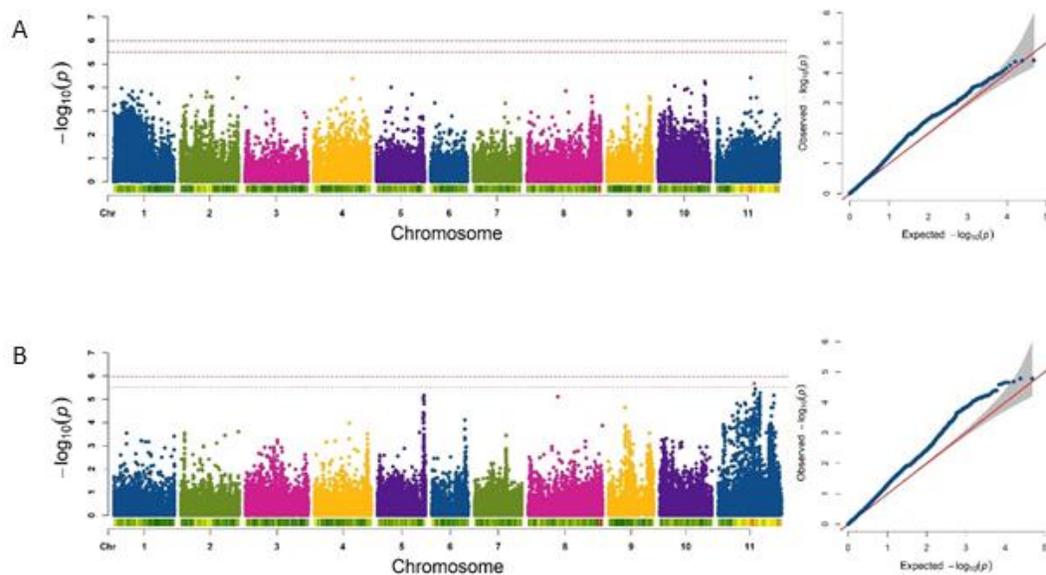
**Figure S5.3.** Manhattan and QQ plots for GWAS of 82 race MA genotypes and seed nitrogen content (%N). The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.



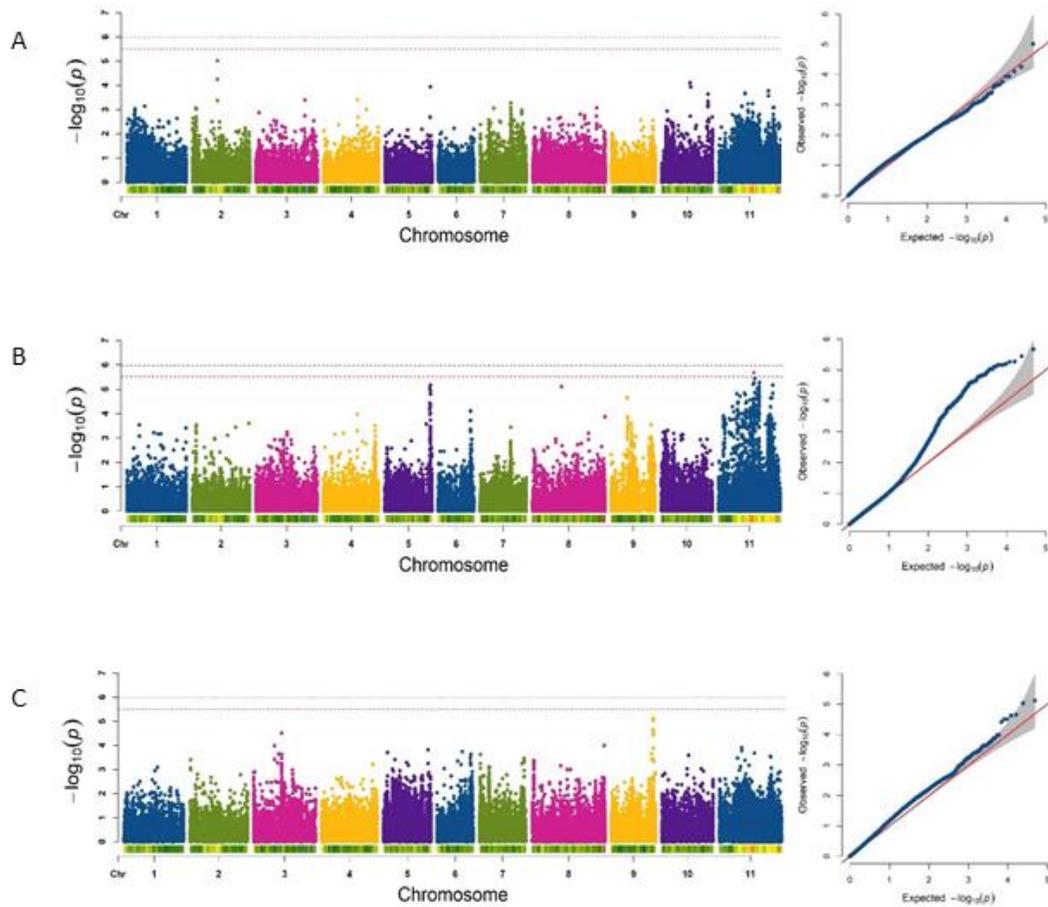
**Figure S5.4.** Manhattan and QQ plots for GWAS of 280 MDP genotypes and leaf chlorophyll content at first trifoliolate stage values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON.



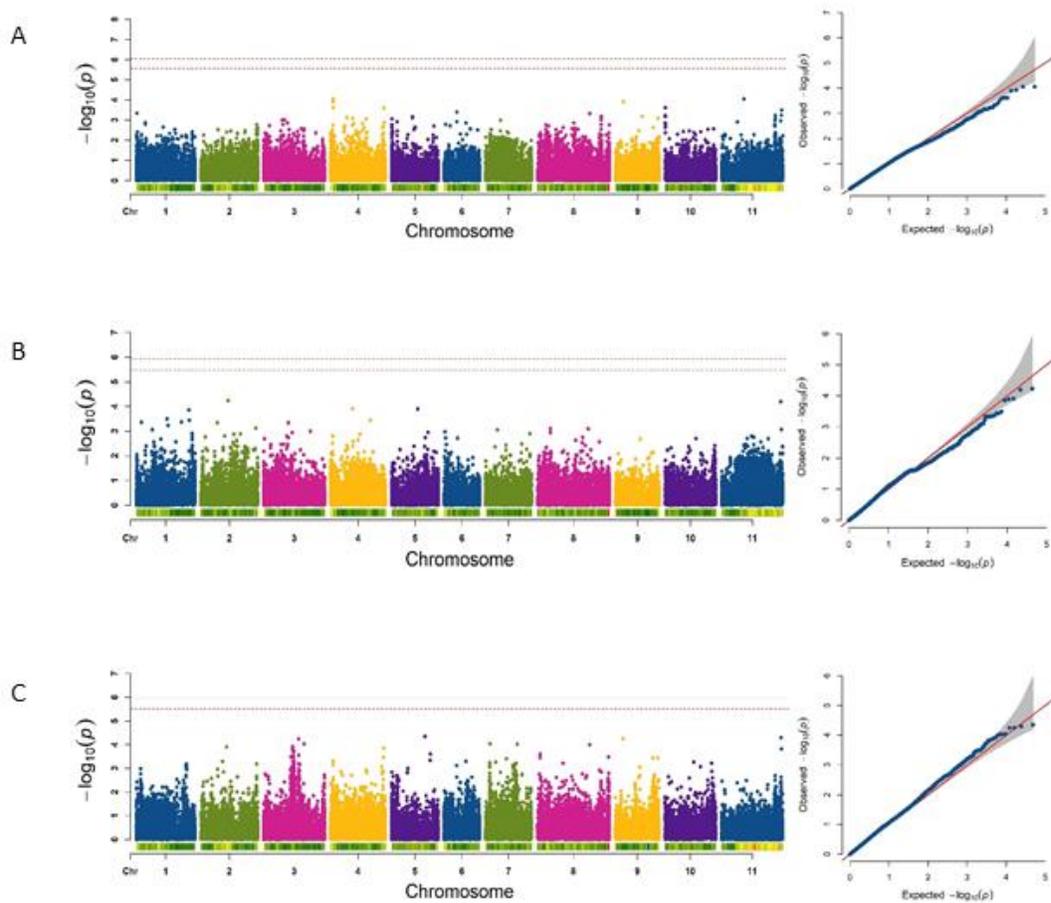
**Figure S5.5.** Manhattan and QQ plots for GWAS of 98 MDP genotypes belonging to the MA subpopulation and leaf chlorophyll content at first trifoliolate stage values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON.



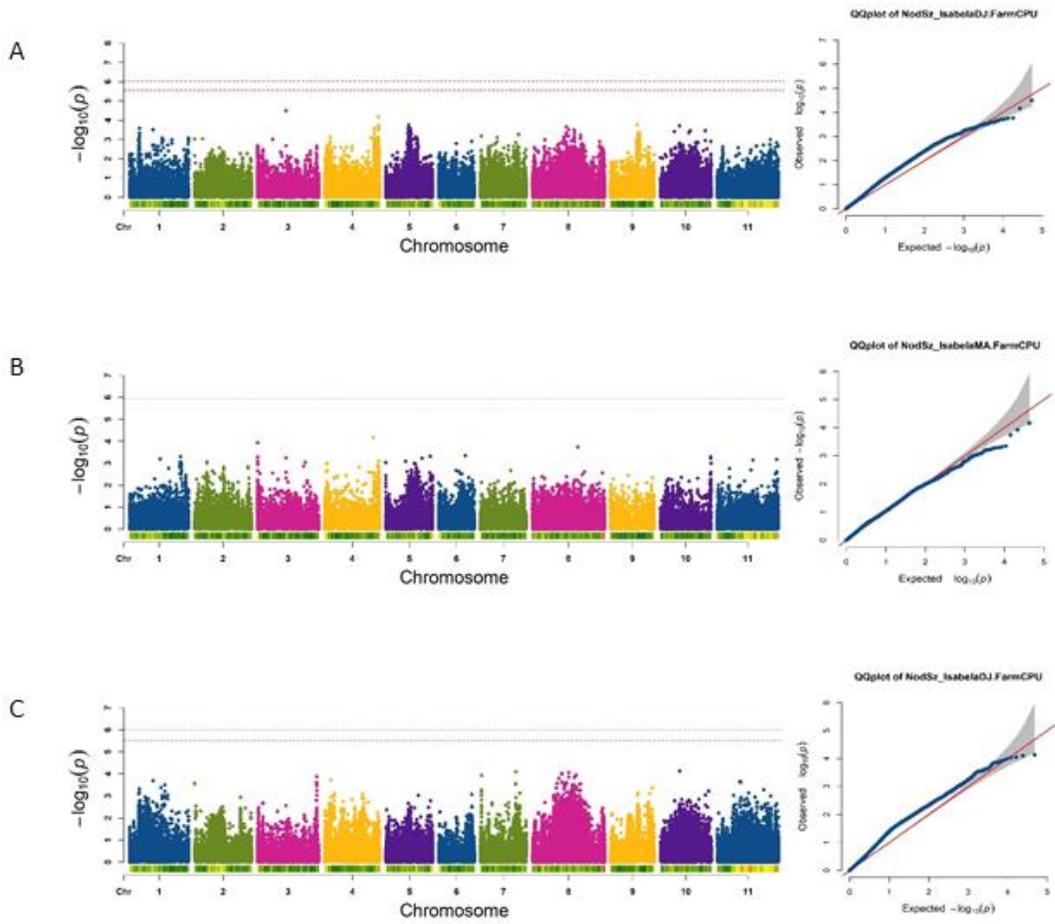
**Figure S5.6.** Manhattan and QQ plots for GWAS of 182 MDP genotypes belonging to the DJ subpopulation and leaf chlorophyll content at first trifoliolate stage values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON.



**Figure S5.7.** Manhattan and QQ plots for GWAS of 182 MDP genotypes belonging to the DJ subpopulation and leaf chlorophyll content at flowering values. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots. (A) Elora and (B) Belwood, ON, and (C) Isabela, PR.



**Figure S5.8.** Manhattan and QQ plots for GWAS of nodule number at Isabela, PR for (A) the MDP genotypes, (B) the MA subpopulation, and (C) the DJ subpopulation. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots.



**Figure S5.9.** Manhattan and QQ plots for GWAS of nodule size at Isabela, PR for (A) the MDP genotypes, (B) the MA subpopulation, and (C) the DJ subpopulation. The FarmCPU model was used in rMVP. Significance thresholds of 5% and 10% FDR are indicated by the upper and lower dashed lines across each plot. SNPs that exceed 5% FDR are red in the QQ plots.