Characterization of a Putative Yield-related Gene in Common Bean

*(Phaseolus vulgaris L.)*

By

Yanzhou Qi

A Thesis
presented to
The University of Guelph

in partial fulfillment of the requirements
for the degree of
Master of Science
in
Plant Agriculture

Guelph, Ontario, Canada
© Yanzhou Qi, January, 2015
ABSTRACT

Characterization of a Putative Yield-related Gene in Common Bean (Phaseolus vulgaris L.)

Yanzhou Qi                                                                                                                          Advisor:
University of Guelph, 2015                                                                        Professor K. Peter Pauls
Co-Advisor:
Professor Steven Rothstein

Yield is a complex trait for breeders to improve, but the identification of yield-related genes and markers for yield genes may help breeders to select for high-yielding bean varieties. BnMicEmUp (a homolog of AT1G74730 from Arabidopsis) is a gene that is upregulated during microspore embryogenesis in Brassica napus and has been characterized as a transcription factor (bZIP) whose expression is negatively related to seed yield in Arabidopsis. The current study characterized a BnMicEmUp/AT1G74730 homolog (named Phvul.009G190100) in common bean (Phaseolus vulgaris. L), an important food crop throughout the world. The gene was isolated and sequenced from P. vulgaris variety “Lightning” by PCR and RT-PCR using primers designed from the sequence in the bean genome (Phytozome v10) annotated as Phvul.009G190100. The high similarity of gene sequence and protein structure between Phvul.009G190100 and BnMicEmUps suggests that Phvul.009G190100 encodes a transcription factor (cbZIP) that is localized in chloroplasts. Expression levels of Phvul.009G190100 in leaves of a number of bean varieties tended to be negatively related with the seed yield, but the correlation was not statistically significant. Homologs for Phvul.009G190100 were also identified in soybean, which is closely related to common bean. The two soybean homologs (Glyma04g34330 and Glyma06g20240) were found to be in close vicinity of molecular markers.
that are linked to a large number of yield-related QTLs. It appears that Phvul.009G190100 from *P. vulgaris*, AT1G74730 from *Arabidopsis*, BnMicEmUp from *B. napus*, and Glyma04g34330 and Glyma06g20240 from *Glycine max* all belong to a gene family that may affect seed yield in plants.
ACKNOWLEDGEMENTS

I would like to acknowledge that I have had the privilege of being Professor Peter Pauls’ student during my master program. I am truly thankful to him for his patient guidance, enthusiastic encouragement and his valuable and constructive suggestions on this research work which are really helpful and important for a non-native English speaker. My special thanks go to my advisory committee: co-advisor Dr. Steven Rothstein and Dr. Ali Navabi for their professional suggestions and invariable assistance throughout my thesis work. I would like to thank my examination committee, including: Dr. Clarence Swanton, Dr. Steven Rothstein and Dr. Istvan Rajcan for their valuable time and efforts. I’m also thankful to Tom Smith for his advice on planting and harvesting beans in the field and all the summer students have been involved in helping me with the field work. I am also grateful for Dr. Navabi for his advice on statistical analyses. Thanks to all lab mates particularly, thanks to Dr. Yarmilla Reinprecht, Dr. Loo-sar Chia, Dr. Weilong Xie, Dr. Mahbuba Siddiqua, PhD student Fariba Shamir and Jerlene Nessia for their technical help, advice, discussion and encouragement. Finally, I would like extend my appreciation to my family and friends, for their great encouragement and constant support throughout the entire thesis work.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS.......................................................... iv

TABLE OF CONTENTS.......................................................... v

LIST OF TABLES....................................................................... viii

LIST OF FIGURES..................................................................... ix

CHAPTER 1 LITERATURE REVIEW.................................................. 1

1.1 Introduction to Common Bean.............................................. 1

1.1.1 History............................................................................. 1

1.1.2 Growth Habit ............................................................... 2

1.1.3 Market Classes............................................................ 3

1.1.4 Nutritional Importance.................................................... 3

1.1.5 Production................................................................. 5

1.2 Yield of Common Bean....................................................... 5

1.3 Breeding for Increased Yield of Common Bean..................... 8

1.3.1 Conventional Breeding of Beans..................................... 8

1.3.2 Molecular Breeding of Beans........................................ 11

1.3.3 QTL Studies of Yield..................................................... 12

1.3.4 Historical Yield Trial..................................................... 14

1.4 Genomics of Common Bean............................................... 15

1.4.1 Available Common Bean Genomic Resources.................. 15

1.4.2 Linkage Maps in Common Bean.................................... 17

1.5 Comparison of Genome in Legumes................................. 18

1.6 Yield QTL Studies in Soybean............................................ 20

1.7 Characterization of a Candidate Yield Gene in *Phaseolus vulgaris* based on the Identification of its Effects on Seed Yield in *Arabidopsis* and *Brassica napus* ........................................ 23

1.8 Plant Heat Stress Response................................................ 25

1.9 Objectives and Hypotheses............................................... 27

CHAPTER 2 MATERIALS AND METHODOLOGY.......................... 29

2.1 Field Trial........................................................................... 29

2.1.1 Site Descriptions.......................................................... 29

2.1.2 Experiment Design....................................................... 29

2.1.3 Grain Yield................................................................. 30

2.1.4 Plant Tissue Harvesting............................................... 31

2.2 Nucleic Acid Extraction..................................................... 31

2.2.1 DNA Extraction........................................................... 31

2.2.2 RNA Extraction.......................................................... 32

2.2.3 RNA Purification and Analysis...................................... 33

2.3 Primer Design.................................................................... 33

2.4 Polymerase Chain Reaction............................................. 34

2.5 Agarose Gel Electrophoresis............................................. 35

2.6 Gene Cloning..................................................................... 36

2.6.1 PCR Product Purification............................................. 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6.2 Cloning into TOPO® Easy Vector</td>
<td>36</td>
</tr>
<tr>
<td>2.6.3 PCR Confirmation for Positive Clones</td>
<td>37</td>
</tr>
<tr>
<td>2.6.4 Plasmid Extraction</td>
<td>37</td>
</tr>
<tr>
<td>2.6.5 Plasmid DNA Sequencing</td>
<td>38</td>
</tr>
<tr>
<td>2.7 Real-time PCR (qPCR)</td>
<td>38</td>
</tr>
<tr>
<td>2.7.1 First Strand cDNA Synthesis (RT-PCR)</td>
<td>39</td>
</tr>
<tr>
<td>2.7.2 Primers</td>
<td>40</td>
</tr>
<tr>
<td>2.7.3 Relative Standard Curve</td>
<td>40</td>
</tr>
<tr>
<td>2.7.4 Data Analysis for Quantification of Gene Expression</td>
<td>41</td>
</tr>
<tr>
<td>2.7.5 Comparative Ct Method</td>
<td>41</td>
</tr>
<tr>
<td>2.7.6 Validation Experiments</td>
<td>41</td>
</tr>
<tr>
<td>2.8 Sequence Analysis</td>
<td>42</td>
</tr>
<tr>
<td>2.9 Phylogenetic Analysis</td>
<td>42</td>
</tr>
<tr>
<td>2.10 Statistical Analysis</td>
<td>43</td>
</tr>
<tr>
<td>CHAPTER 3 RESULTS</td>
<td>45</td>
</tr>
<tr>
<td>3.1 Environmental Variability</td>
<td>45</td>
</tr>
<tr>
<td>3.2 Analysis of Yield Data</td>
<td>47</td>
</tr>
<tr>
<td>3.2.1 Total Yield</td>
<td>47</td>
</tr>
<tr>
<td>3.2.2 Daily Yield Gain (kg/day/ha)</td>
<td>49</td>
</tr>
<tr>
<td>3.2.3 Days to Maturity</td>
<td>51</td>
</tr>
<tr>
<td>3.2.4 Relationship between Yield (kg/ha) and Yield Gain (kg/ha/day)</td>
<td>53</td>
</tr>
<tr>
<td>3.2.5 Yield Ranks Correlation and Regression</td>
<td>54</td>
</tr>
<tr>
<td>3.3 Annotation of AT1G74730 Homolog from Common Bean</td>
<td>55</td>
</tr>
<tr>
<td>3.4 Isolation of Phvul.009G190100 from P. vulgaris</td>
<td>56</td>
</tr>
<tr>
<td>3.5 Cloning of Phvul.009G190100</td>
<td>57</td>
</tr>
<tr>
<td>3.6 Gene Structure</td>
<td>59</td>
</tr>
<tr>
<td>3.7 Analysis of Elements in the Phvul.009G190100 Promoter</td>
<td>61</td>
</tr>
<tr>
<td>3.8 Protein Domain Characterization</td>
<td>64</td>
</tr>
<tr>
<td>3.8.1 Chloroplast Transit Peptide</td>
<td>65</td>
</tr>
<tr>
<td>3.8.2 bZip Region</td>
<td>67</td>
</tr>
<tr>
<td>3.8.3 Transmembrane Domain</td>
<td>69</td>
</tr>
<tr>
<td>3.8.4 Phosphorylation Sites (PS)</td>
<td>69</td>
</tr>
<tr>
<td>3.8.5 Protein Secondary Structure</td>
<td>70</td>
</tr>
<tr>
<td>3.9 Real-time PCR Results</td>
<td>72</td>
</tr>
<tr>
<td>3.9.1 Standard Curve Construction</td>
<td>72</td>
</tr>
<tr>
<td>3.9.2 Validation Experiment</td>
<td>74</td>
</tr>
<tr>
<td>3.10 Gene Expression of Phvul.009G190100</td>
<td>76</td>
</tr>
<tr>
<td>3.11 Correlation between Gene Expression Levels and Yield Data</td>
<td>79</td>
</tr>
<tr>
<td>3.11.1 Correlation between Gene Expression Levels and Historical Data</td>
<td>80</td>
</tr>
<tr>
<td>3.11.2 Correlation between Gene Expression Levels with 2013 Yield Data</td>
<td>83</td>
</tr>
<tr>
<td>3.12 Homolog of Phvul.009G190100 Localization in Soybean</td>
<td>86</td>
</tr>
<tr>
<td>3.12.1 Gene Structure</td>
<td>86</td>
</tr>
<tr>
<td>3.12.2 Protein Characterization</td>
<td>89</td>
</tr>
<tr>
<td>3.12.3 Expression of Glyma04g34330 and Glyma06g20240 in soybean</td>
<td>92</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2. 1 The twelve navy bean varieties used for the field trials at Elora and Woodstock, ON during 2013 growing season ................................................................. 30
Table 2. 2 Summary of Nucleotide sequences of PCR primers ........................................ 34
Table 2. 3 PCR cycles and time of each step used to amplify Phvul.009G190100 .................. 35
Table 2. 4 A 3-step real-time PCR protocol with melting curve .................................... 39
Table 2. 5 Thermal cycler programmed reverse transcription to produce cDNA ............... 40
Table 3. 1 Specifics of Arabidopsis AT1G74730, P. vulgaris Phvul.009G190100 and B. napus BnMicEmUP genes .............................................................. 60
Table 3. 2 cis regulatory elements found in upstream sequence (1000 bp) of Phvul.009G19010 62
Table 3. 3 Predicted localization of Phvul.009G190100 protein ......................................... 66
Table 3. 4 Normalized fold change expression of PV09 ................................................. 78
Table 3. 5 Predicted localization of Glyma04g34330 protein ........................................ 91
Table 3. 6 Predicted localization of Glyma06g20240 protein ........................................ 92
Table A- 1 Analyses of variance for Pearson correlations and covariance parameter estimates of yield (kg/ha) of twelve common bean varieties determined at Elora, ON in 2013 .......... 145
Table A- 2 Comparison of yield (kg/ha) between three groups of common bean varieties determined at Elora, ON in 2013 ................................................................. 145
Table A- 3 Analyses of variance for Pearson correlations and covariance parameter estimates of daily yield gain (kg/ha/day) of twelve common bean varieties determined at Elora, ON in 2013 ................................................................................................................. 145
Table A- 4 Comparison of the daily yield gain (kg/ha/day) of twelve common bean varieties determined at Elora, ON in 2013 and in the historical trial ................................................. 145
Table A- 5 Comparison of daily yield gain (kg/ha/day) between three groups of common bean varieties determined at Elora, ON in 2013............................................................................................................. 146
Table A- 6 Analysis of variance of the relative gene expression level of twelve common bean varieties determined at Elora, ON in 2013 ................................................................. 146
LIST OF FIGURES

Figure 1.1 Average yield of navy bean in Central Canada since 1930s.................................................. 15
Figure 1.2 Syntentic relationships of soybean relative to the common bean genetic map ........... 20
Figure 1.3 Comparison of four forms of B. napus BnMicEmUP protein sequences with
Arabidopsis AT1G74730........................................................................................................ 26
Figure 1.4 Comparison of final plant heights in Arabidopsis seedlings........................................ 26
Figure 1.5 A proposed model for a dual targeting of BnMicEmUP protein and its involvement in
ABA-mediated transcriptional regulation............................................................................. 27
Figure 3.1 Planting locations for 2013 yield trial ........................................................................... 46
Figure 3.2 Daily average temperature and accumulated precipitation in Elora, ON....................... 46
Figure 3.3 Daily average temperatures and accumulated precipitation in Woodstock, ON........... 47
Figure 3.4 Correlation between yield (kg/ha) for twelve common bean varieties determined at
Elora, ON in 2013 and in the historical trial........................................................................... 48
Figure 3.5 Average yield gains (kg/ha/day) for twelve genotypes at Elora ON in 2013 and in the
historical trial ..................................................................................................................... 50
Figure 3.6 Correlation between average yield gain (kg/ha/day) for twelve common bean varieties
determined at Elora, ON in 2013 and in the historical trial.................................................... 50
Figure 3.7 Maturity time (days) for twelve genotypes at Elora and Woodstock ON in 2013 and in the
historical trial ..................................................................................................................... 52
Figure 3.8 Correlation between maturity time (days) for twelve common bean varieties
determined at Elora, Woodstock, ON in 2013 and in the historical trial............................... 52
Figure 3.9 Correlation between total yield (kg/ha) and Average yield gains (kg/ha/day) for yield
trials at Elora ON in 2013 and in the historical trial............................................................ 53
Figure 3.10 Rank correlation of the total yields (kg/ha) for the twelve genotypes in the historical
trial and the yield trial in Elora in 2013.............................................................................. 54
Figure 3.11 Rank correlation of average yield gain (kg/ha/day) for the twelve genotypes in
historical trial and the yield trial in Elora in 2013............................................................... 55
Figure 3.12 Amplification of Phvul.009G190100 from P. vulgaris (variety Lightning) genomic
DNA ..................................................................................................................................... 56
Figure 3.13 Comparison of the gene sequences for Phvul.009G190100 and AT1G74730.. ....... 57
Figure 3.14 Colony PCR to confirm E.coli uptake of Phvul.009G190100 clone......................... 58
Figure 3.15 04-01-FC2 sequence aligned with genomic DNA and cDNA sequences of
Phvul.009G190100 from Phytozome.................................................................................. 59
Figure 3.16 Comparison of the P. vulgaris Phvul.009G190100 gene structure with the
Arabidopsis gene AT1G74730. ......................................................................................... 61
Figure 3.17 Cis regulatory elements found in upstream sequence (1000 bp) of
Phvul.009G190100 ........................................................................................................ 63
Figure 3.18 Comparison of P. vulgaris Phvul.009G190100 putative protein sequence with B.
napus BnMicEmUP3 and Arabidopsis AT1G74730 protein sequences............................. 64
Figure 3.19 Alignment of amino acid sequences of the predicted transit peptides of
Phvul.009G19010, AT1G74730, BnMicEmUP3, Glyma04g34330 and Glyma06g20240 with
the chloroplast targeted proteins in other species. .............................................................. 66
Figure 3. 20 Structural features of Phvul.009G190100 compared to plastid envelope DNA binding PEND proteins. ................................................................. 68
Figure 3. 21 The predicted secondary α-helix and β-strand structure of Phvul.009G190100. 71
Figure 3. 22 3D view of Phvul.009G190100 protein .............................................................. 72
Figure 3. 23 Expression standard curve of Phvul.009G190100 and reference gene (SKIP16) .... 75
Figure 3. 24 Expression distribution of Phvul.009G190100 in different tissues of a common bean inbred landrace line G19833. ................................................................. 76
Figure 3. 25 Transcript profile of Phvul.009G190100 in the leaves of twelve field grown common bean varieties. ................................................................. 75
Figure 3. 26 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and yield (kg/ha) in the historical trial. ................................................................. 81
Figure 3. 27 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and average gained yield (kg/ha/day) in the historical trial. ........................................ 82
Figure 3. 28 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and maturity time (days) in the historical trial. ............................................................ 83
Figure 3. 29 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and yield (kg/ha) in 2013 yield trial ................................................................. 84
Figure 3. 30 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and average gained yield (kg/ha/day) in 2013 trial ................................................ 85
Figure 3. 31 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and maturity time (days) in 2013 trial ................................................................. 86
Figure 3. 32 Comparison of the nucleotide sequences for Phvul.009G190100, Glyma04g34330 and Glyma06g20240 ................................................................. 89
Figure 3. 33 Gene structure comparison of Phvul.009G190100 with Glyma04g3433 and Glyma06g20240 ................................................................. 89
Figure 3. 34 Comparison of putative protein sequences encoded by Phvul.009G190100, Glyma04g34330 and Glyma06g20240 ................................................................. 91
Figure 3. 35 Expression of G. max Glyma04g34330 and Glyma06g20240 in different tissues of the soybean ................................................................. 93
Figure 3. 36 Expression proportion of Glyma04g34330, Glyma06g20240 in soybean and Phvul.009G190100 expression proportion in beans ................................................................. 93
Figure 3. 37 Relationships among AT1G74730, Phvul.009G190100, BnMicEmUPS, Glyma06g20240 and Glyma04g34330 genes. ................................................................. 95
Figure 3. 38 Secondary structures of putative proteins encoded by AT1G74730, BnMicEmUPS, Phvul.009G190100, Glyma04g34330 and Glyma06g20240 genes ................................................................. 96
Figure 3. 39 Sequence-based common bean-soybean comparative map ................................................................. 98
Figure 3. 40 Soybean yield related QTLs occurring in genomic regions of the soybean composite linkage map containing Glyma04g34330 and Glyma06g20240 ................................................................. 101
CHAPTER 1 LITERATURE REVIEW

1.1 Introduction to Common Bean

1.1.1 History

Common bean (*Phaseolus vulgaris* L), also known as dry bean, is an important leguminous species that belongs to the family Fabaceae, genus *Phaseolus*. "Dry beans" refers to the dry edible seeds that are the commercial product from field-grown bean varieties. As the most important form in which *P. vulgaris* is consumed, dry beans are important for human nutrition (Broughton et al., 2003) and an interesting crop for scientific research.

Common bean is believed to have been domesticated in two geographically-isolated and genetically differentiated wild gene pools (Mesoamerican and Andean), that diverged from a common ancestral wild population more than 100,000 years ago (Gepts, 1988). The two gene pools subsequently subdivided into the Mesoamerican (Mexican and Central American races) (Singh et al., 1991), and the Andean located on the western coast of South America, in Ecuador and Peru (Gepts et al., 1998). A recent study (Bitocchi et al., 2012) presented clear evidence of a Mesoamerican origin of *P. vulgaris* as the common ancestor, which was most likely located in Mexico. The Mesoamerican origin is supported by the analysis of the population structure that showed a much higher diversity in the Mesoamerican wild accessions compared with the diversity seen in South America (Bitocchi et al., 2012). Additional evidence for this model of diversity in bean includes: the finding that the closest relatives of wild *P. vulgaris* in the *Phaseolus* genus are distributed throughout Mesoamerica (Schmit et al., 1993; Freytag et al., 2002; Delgado-Salinas et al., 1999), the much higher diversity of Mesoamerican compared with
the diversity from South America, evidence for the occurrence of a severe bottleneck in the Andes before domestication (Rossi et al., 2009), and the occurrence in Mesoamerica of wild beans that are closely related to those beans found in South America (Andean and northern Peru–Ecuador gene pool).

Beans with small to medium seed sizes were domesticated from the Mesoamerican gene pool of Central America while the Andean gene pool of South America was domesticated to give larger seeded beans (Gepts et al., 1998). Among the available market classes of dry beans in North America, navy and black beans with small seeds belong to race Mesoamerica, while the pinto, great northern and small red beans belong to the race Durango of the Mesoamerican gene pool. Kidney and cranberry beans, with large seeds, belong to the race Nueva Granada of the Andean gene pool.

1.1.2 Growth Habit

Common bean plants exhibit two distinct stem growth habits: indeterminate and determinate. Plants with indeterminate growth habit have a terminal meristem that remains vegetative throughout vegetative and reproductive growth (Ojehomon and Morgan, 1969). In contrast, in determinate plants, the terminal shoot meristem switches from a vegetative to reproductive state, thus, producing a terminal meristem (Ojehomon and Morgan, 1969). A notable difference between wild and domesticated common bean is the larger variability of growth habit types that exists among domesticated classes (Smartt et al., 1990). While only indeterminate growth is observed in wild-type common bean populations, both growth habits can be observed in domesticated varieties. The determinate growth habit has been artificially
selected, in combination with photoperiod insensitivity, to obtain varieties with a shortened flowering period and earlier maturation, allowing a shorter period until harvest and tolerance of mechanical harvesting (Cober and Tanner, 1995; Koinange et al., 1996).

### 1.1.3 Market Classes

Dry beans can be classified into several market classes by seed size and seed coat color. In Canada, the major market classes are navy, pinto, dark red kidney, light red kidney, cranberry, black, small red, brown, pink and great northern (Goodwin, 2003; Kelly et al., 2009). Niche market classes such as otebo, kintoki, Dutch brown, yellow eye and Jacob's cattle bean are also grown on a smaller scale. The major bean production areas are Manitoba and Ontario, as well as niche areas in Saskatchewan, Alberta, and Quebec (Goodwin, 2003). The navy, black and pinto bean classes are of the Mesoamerican ancestry and are commonly grown in all bean producing areas of Canada, except Alberta (Goodwin, 2003; Mamidi et al., 2011). The remaining colored beans have Andean ancestry and are grown in all Canadian bean producing regions (Goodwin, 2003; Mamidi et al., 2011). In Ontario, the navy, colored, and Japanese classes (Kintoki and Otebo) are commonly grown. The Japanese classes are specific to Ontario and not produced in Manitoba, whereas navy, black and a variety of colored beans are more commonly grown.

### 1.1.4 Nutritional Importance

Common bean is one of the most nutritious crops for humans because it is a good source of fiber and dietary protein and contains little or no fat, sodium or cholesterol (Duranti, 2006). Moreover, dry bean is also an excellent source of vitamins, minerals and other nutrients, but contain only moderate levels of calories (Hefni et al., 2010; Ruggeri et al., 1999). Dry beans are
very high in fiber, containing both soluble and insoluble fibers. While soluble fiber helps to decrease blood cholesterol levels and control blood sugar levels, insoluble fiber helps with digestion and regularity. On average, beans provide 7 or more grams of total dietary fiber per ½-cup serving. Beans also provide high levels of protein for humans and represent an important complement to animal proteins, especially for people in developing areas of the world, where beans are available at lower costs than meats.

Moreover, beans provide abundant key minerals including: iron, calcium, phosphorus, magnesium, and in lesser degree, zinc, copper and calcium (Broughton et al., 2003). Folate, commonly known as vitamin B9 (Arcot and Shrestha, 2005), is found in relatively high levels in beans (Bekaert et al., 2008). As an essential vitamin for the production of red blood cells and development of an embryo’s nervous system during pregnancy, folates also help to reduce the risk of neural tube defect in newborns (Geisel, 2003). Currently, micronutrient deficiencies (especially iron) have become common even in developed countries. As a source of micronutrients, beans are much superior to the cereals grains that normally make up the diets (Welch et al., 2000). This is because cereals are normally polished before they are eaten (to produce white rice or wheat flour etc.), thus eliminating the seed coats (or bran) which contain a significant portion of the minerals. Beans, on the other hand, are consumed whole so that their mineral content is conserved. To ensure a balanced diet, it is advisable to consume beans with cereals (Bressani, 1983)) because they are nutritionally complementary. Beans are deficient in sulphur-containing amino acids, such as methionine, but are sufficient in lysine, whereas, cereals generally contain more methionine but are deficient in lysine (Gepts and Bliss, 1985).
1.1.5 Production

The total global production of dry beans is approximately 20 million tonnes, which are harvested from ~30 million hectares, annually, in ~150 countries (FAOSTAT, 2012). The majority of production occurs in developing countries in Asia, Africa, and South America, as well as some developed countries in Oceania, Europe, and North America (Gepts et al., 2008). Asia and the Americas contribute approximately 81% of the total bean production each year and include the top five producers in the world, namely: Brazil, India, Myanmar, China and USA (FAOSTAT, 2012).

Canadian bean production is 250 thousand tonnes annually and it constitutes 1.4% of the total bean production worldwide (Stats Canada, 2012; FAOSTAT, 2012). In Canada, Ontario is the largest producer and accounts for approximately 60% of Canada's dry bean production (Stats Canada, 2012). Despite the low percentage that Canada contributes to the total bean production worldwide, dry bean serves as an important export commodity. Of the annual dry bean production in Canada, only 10% is used domestically, whereas, 90% was exported to the USA and Europe. Other major exporting countries are Myanmar, China, USA, and Argentina (Agriculture and Agri-Food Canada, 2012; FAOSTAT, 2012). In the last ten years the majority of exports from these countries go to India and the European Union, who have annually imported over 500 thousand tonnes and 1.8 million tonnes, respectively (FAOSTAT, 2012).

1.2 Yield of Common Bean

Yield is the most complex trait from a crop physiology perspective because a number of morphological characteristics and physiological processes contribute to yield. Common bean
yield is often described as a functional product of many yield-related components such as number of pods per plant, number of seeds per pod and 100 seed weight. To achieve maximum yield, all these yield components should be at an optimum balance. The number of pods per plant is an important yield component that accounts for 67% of the variation in grain yield of dry beans. Bennet et al. (1977) reported that among the yield components, the number of pods per plant has often been recommended as an indirect selection criterion for increasing bean yield because of its high and consistent correlation with grain yield. The number of seeds per pod and 100-seeds weight are also important yield components. Tanaka and Fujita (1979) reported that the number of seeds per pod (X) were significantly linearly related to grain yield (Y) \(Y = -4.6121 + 4.1534X, R^2 = 0.2116\). Similarly, 100-seed weight (X) was found to be significantly positively correlated with grain yield \(r = 166.9762, R^2 = 0.4051\) (Tanaka and Fujita, 1979). Therefore, grain yield may be influenced by altering the yield-component traits (Chapman et al., 2003), which are genetically controlled although they are also influenced by environmental stress factors.

The loss of bean yield is usually caused by both biotic and abiotic stresses. Among biotic stresses are crop infestation with diseases, insects, weeds, and lower nitrogen (N) fixation capacity of bean plants (Fageria, et al., 1998; Fageria et al., 2002). Disease is the biggest production problem that dry bean growers face in Canada. Most commonly seen diseases are seed-borne (such as anthracnose and bacterial blight) while other disease problems over-winter in the soil (such as root rot, rust and white mould/sclerotinia). The risk of disease can be decreased by preventative practices (e.g., using certified seed, growing disease resistant bean varieties, increasing the airflow under the canopy to stop the disease flourishing, and adhering to
a proper rotation). However, if weather conditions are ideal for the disease, it can still flourish despite all the producer’s best efforts. Reactive measures, such as seed treatments (for example a mixture of diazinon, captan and thiophanate methyl for anthracnose control) can decrease disease infection levels, however, it is ineffective if seeds are severely infected (Goodwin, 2003). In contrast, genetic resistance is a valuable disease and pest management tool for bean production. Therefore, the development of cultivars with greater levels of disease and pest resistance is a primary objective for most bean breeding programs.

Moreover, bean production is also threatened by abiotic stresses such as drought, low soil fertility, soil acidity and temperatures unfavorable for the growth and development of the crop (Lynch, 2007). Abiotic stress resistance is typically subject to large environmental effects and may be the key to improving yields of common bean in both stressed and unstressed environments (Ishitani et al., 2004). Compared to pest and disease resistance, much less is known about genetics of resistance to abiotic constraints or physiological stress because it is typically governed by polygenic inheritance and may be conditioned by multiple, interacting mechanisms. Cultivars with improved stress resistance can reduce reliance on pesticides, the risk of yield loss from pests, and enable more stable bean production across diverse and adverse environments (low precipitation, high humidity, etc.) and poor soil conditions (low fertility, hillsides, etc.). Therefore, the development of cultivars with improved resistance to biotic and abiotic stresses is a primary goal of high-yield bean breeding programs throughout the world.
1.3 Breeding for Increased Yield of Common Bean

Increased seed yield is a universal objective of common bean breeding programs. In Canada, the common bean breeding has a history of more than 120 years. In 1886, Central Experimental Farm (CEF) was established in Ottawa, ON and early breeding efforts started by testing of dry and garden bean introductions. In 1910s, ‘Norwegian Brown’ and ‘Beauty’ were developed with the pure line selection in the introduced land races. Later on, in the 1920s, cross-breeding and selection in segregating populations were initiated in Ottawa. Agriculture and Agri-Food Canada (AAFC) research centers in Harrow, ON, and Lethbridge, AB, began bean breeding activities in 1950s. The Bean Breeding Programs at the University of Guelph in Guelph, ON, was established in 1977 and at the University of Saskatchewan in Saskatoon, SK, in 1990 (Park and Buzzell, 1995). Since 2004, the dry bean breeding activities in Manitoba have been centered at the AAFC Morden Research Station and the two publicly funded bean breeding programs in Ontario merged and formed the collaborative AAFC–University of Guelph Bean Breeding Program in 2008. To increase dry bean yield in Canada, it is generally accepted that the long-term bean breeding practices have played an essential role to develop well-adapted, high-yielding varieties. In Canada, dry bean production has reached 1.8 tonnes ha⁻¹ in the 1990s and 2000s (20-yr average), which is a 63% increase compared to an average of 1.1 tonnes ha⁻¹ during the 1910s and 1920s (20-yr average) (Navabi et al., 2014).

1.3.1 Conventional Breeding of Beans

Common bean (Phaseolus vulgaris L.) improvement programs have been successful using conventional plant breeding techniques in the development and release of dry bean
cultivars with greater seed yield (Kelly et al., 1998; Singh, 1991). Conventional breeding programs are initiated involved in selection among genetically variable individuals and populations, crosses among diverse gene pools to select for the elite genotype and also usage of the wild germplasm for the backcross.

Recurrent selection, one of the vital classical breeding techniques, permits the accumulation of favorable alleles as the result of recombination in each cycle of selection. Beaver and Kelly (1994) used inter-gene pool crosses and recurrent selection to develop indeterminate, large-seeded bean lines that had greater seed yield than determinate check cultivars. Singh et al. (1999) used recurrent selection based on S1 evaluations to increase seed yield of inter-racial (Middle American) and inter-gene pool (Middle American x Andean) bean populations. As a result, breeding lines were developed with greater yield than the parents and recurrent selection is suggested to be effective for recombining and forming new desirable gene complexes in lines from inter-gene pool crosses. However, the limitations of this method include: difficulties in applying to self-pollinating crops like beans; long-term efforts required for generations.

Backcrossing is another important classical breeding technique. It is often utilized by plant breeders to incorporate simply inherited traits by crossing of an F1 hybrid with the recurrent parent (a superior variety which needs improvement for a special character) (Miklas 2007; Miklas et al., 2003). The donor parent (non-recurrent parent) acts as a source of gene(s) to be added to the recurrent parent. However, this method is limited by the long time required to recover the backcross desirable traits and because it does not result in the improvement of other characters. Although backcross breeding is not useful for the improvement of quantitatively
inherited traits such as yield, it is well suited for marker-assisted selection of traits such as disease resistance (Miklas, 2007; Miklas et al., 2003).

Wild common beans are very diverse and can be a useful source for enhancing the variability of cultivated beans (Kelly et al., 1999; Singh. 2001; Johnson et al., 2002). Singh et al. (1995) made the attempt to use wild common beans in crosses with high-yielding small seeded cultivated beans for improvement of seed yield. Several authors (Singh et al., 1995; Gepts 2002) pointed out some of the limitations to using wild common beans for breeding including: the undesirable agronomic characteristics of wild beans, such as poor architecture, small-seededness, and long growing period; the practical difficulties in crossing wild beans to cultivated beans; and the lack of breeding methodologies in the past to efficiently undertake wild × cultivated common bean crosses. Therefore, the advanced backcross method suggested by Tanksley and Nelson (1996) has been applied overcome some of the limitations of using wild relatives in breeding programs. This technique integrated QTL analysis with variety development by identifying and transferring favorable QTL alleles from wild donor parent into cultivated lines (Tanksley and McCouch, 1997) and these can be tracked in further crosses via marker-assisted selection (Tanksley, 1996). Blair et al. (2006) tested the advanced backcross method in common bean and determined positive QTLs for yield and yield components within an advanced backcross population derived from a wild common bean (G24404) donor parent and an Andean (‘ICA Cerinza’) recurrent parent.

In summary, although conventional breeding techniques have resulted in significant genetic improvements in common bean, an important challenge for dry bean breeders working
with seed yield potential (Kelly et al., 1998) is a lack of desirable alleles for seed yield which may limit breeding progress.

1.3.2 Molecular Breeding of Beans

Sorrells and Wilson (1997) noted that molecular plant breeding techniques can generate, characterize and utilize genetic variation to identify desirable alleles or permit the use of indirect selection to reduce negative associations between these economically important traits. Recent developments with molecular markers help plant breeding for selection of major agronomical important traits such as yield (Tanksley, 1991). Computer simulations by Zhang and Smith (1992), Gimelfarb and Lande (1994) and Knapp (1998) suggested that marker-assisted selection (MAS) could be more efficient than phenotypic selection for identifying superior genotypes in a segregating population. MAS has become a common tool used in common bean breeding programs (Miklas et al., 2006; Kelly et al., 2003) especially for yield trait and permits the selection in absence of selection pressure. Schneider et al. (1997) found yield improvement of 11% and 8% under stress and non-stress conditions respectively. Five RAPD (Random Amplified Polymorphic DNA) markers were used for MAS of yield under a stress environment, however, conventional selection for yield failed to improve the performance. As a consequence, the use of MAS for more complex traits like yield and abiotic stress tolerance require a better understanding of the genetic basis of these traits. Incorporation of genomic information may provide an appropriate platform to study the yield trait and may also help bean breeders devise more effective selection strategies. In bean breeding, MAS was commonly relied on the results of a number of quantitative trait loci (QTL) analyses, which were identified through selection with molecular markers among populations, based on the genomic mapping of the yield trait.
constructed for common bean (Freyre et al., 1998). Molecular markers, used to measure genetic
distances between parents and to estimate genetic variances in derived populations, have been
sought for construction of linkages maps and eventual application in MAS for complex traits
such as yield in beans (Bai et al., 1998; Kidwell et al., 1999; Kisha et al., 1997). A simple and
effective procedure was developed for MAS of complex traits for common bean (*Phaseolus
vulgaris* L.) including seed yield, plant height, harvest index, etc. With the information obtained
from QTL studies, the QTL-based-index in conjunction with the ultrametric genetic distance to
the target parent were used as selection criteria for identifying superior lines in a breeding
program (Beattie et al., 2003). Another selection method in common bean (*Phaseolus vulgaris*
L.) used marker-based cluster analysis (MBCA) to assist phenotypic selection by directing a
breeder’s attention to a subsample of the population containing a high proportion of superior
lines. With the molecular markers developed for certain traits such as yield, a target line can be
selected and the clustering around this line could be used as a “cut-off” point for the subsequent
selection (Tar’an et al., 2003).

### 1.3.3 QTL Studies of Yield

In the past two decades, several QTL studies in beans have examined agronomic traits of
yield and yield-related traits. A relatively large number of linkage maps have been developed
basically from inter-gene pool crosses, to identify single-locus QTLs for yield (Koinange et al.,
1996; Johnson and Gepts 2002; Tar’an et al., 2002; Beattie et al., 2003; Blair et al., 2006, Wright
and Kelly 2011) based on molecular markers. Holland (2001) pointed out that, in autogamous
plants, epistasis is to be expected in traits that are controlled by several genes/QTLs. Strong
interactions between QTLs have been detected in common bean for seed yield (Johnson and
Gepts, 2002). Recently, a multi-environment Quantitative Trait Loci (QTL) analysis has identified and mapped a total of 59 QTLs including single-locus QTLs, epistatic QTLs (E-QTL) and their environment interaction effects (QTLs $\times$ Environment, QE; and E-QTLs $\times$ Environment, E-QE) for the genetic control of seed quality traits in the Andean common bean (Yuste-Lisbona et al., 2014). It indicated that digenic epistatic interactions may play an important role in the genetic control of seed quality and seed yield traits of common bean.

Beattie et al. (2003) identified three yield QTLs based on the closest markers (UBC267.1650, UBC388.1200, UBC464.2100) and mapped to linkage group B03 (Pv03) and B05 (Pv05) total accounted for 39.4% of the phenotypic variation for this trait. Two of the alleles contributed by ‘OAC Speedvale’ increased yield, whereas the third QTL contributed to a decrease in yield. Moreover, Tar’an et al. (2002) mapped three QTLs on linkage groups B05 (Pv05), B09 (Pv09) and B11 (Pv11) associated with yield and collectively accounted for 27.8% of the phenotypic variation. The two QTLs from B05 (Pv05) and B11 (Pv11) from OAC95-4 contributed to the increased yield. Blair et al. (2006) also identified a total of nine yield QTLs ($yld_{2.1}$, $yld_{3.1}$, $yld_{3.2}$, $yld_{4.1}$, $yld_{4.2}$, $yld_{4.3}$, $yld_{4.4}$, $yld_{9.1}$, $yld_{9.2}$) on linkage group B02 (Pv02), B03 (Pv03), B04 (Pv04), and B09 (Pv09), from an Andean bean population and nine nearest markers are found (BM142, BM172, BM98, PVag004, BMd8, BM199, BM161, BM169, BM141). Wright and Kelly (2011) discovered a single QTL designated SY10.2J115 on linkage group B10 (Pv10) that accounted for up to 28% of the variation for yield trait.

In addition to the yield trait, studies were also conducted on identification QTLs of 100-seed weight (SW), which is the most important factor affecting the yield. A total of 11 QTL were
identified across eight linkage groups [B02 (Pv02), B03 (Pv03), B06 (Pv06), B07 (Pv07), B08 (Pv08), B09 (Pv09), B10 (Pv10) and B11 (Pv11)] for 100-seed weight (Blair et al., 2006) while three QTLs were discovered by Blair et al. (2011) on B02 (Pv02), B03 (Pv03), B11 (Pv11), from Andean common beans. The presence of 100-seed weight QTLs on B07 (Pv07), and B11 (Pv11), has also been observed before in reports of intra- and inter-gene pool crosses (Tar’an et al., 2002; Park et al., 2000) and the cultivated x wild combination studied by Koinange et al. (1996). In most of these reports, the maps used to identify QTLs differed in the parents used, the number of molecular markers present, and had a relatively low degree of alignment with other linkage maps. Different number and relative positions of the QTLs involved in the genetic control of the yield trait have been obtained in most of these cases.

1.3.4 Historical Yield Trial

The Dry Bean Breeding Program of University of Guelph collaborative with AAFC–in 2008.was established to develop white and colored beans with superior traits, for instance, productivity, adaptability, diseases resistance and cooking properties. Identification, characterization and utilization genes for bean improvement are also important aspects covered by the program. Since the highly improved agronomic practices occurred in the past years, the yield of the varieties might be affected. In order to examine the contribution of the genetic improvement in yield potential and to discover the association between yield and other agronomic traits, a historical yield trial has been conducted by Navabi et al. (2012) in Ontario, Canada. Twenty-nine white navy beans which released since 1930s were selected and planted in seven location years (St. Thomas, ON in 2010 and St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012) (Navabi et al., 2012). The result showed a significant difference among the
cultivars for yield in all location years and the slope of the regression line of yield (P < 0.01) over year of release was significant and indicated an annual gain of 18 kg ha⁻¹ (Figure 1.1) (Navabi et al., 2012).

Figure 1.1 Average yield, combined over years, of 29 navy beans which released in Central Canada since 1930s. The historical trial was conducted at St. Thomas, ON in 2010 and St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively. All of the trials were in Randomized Complete Block Design with 4 replications each (Navabi et al., 2012).

1.4 Genomics of Common Bean

1.4.1 Available Common Bean Genomic Resources

The genome sequence of common bean is available from Phytozome V1.0 (http://phytozome.jgi.doe.gov/), the first chromosome scale version of *Phaseolus vulgaris*. L. This release includes BAC and fosmid end sequence and a dense marker map (7,015 markers)
contributed by the USDA-NIFA project (A sequence map of the common bean genome for bean improvement) combined with DOE-JGI and ARRA funded whole genome shotgun data sets and RNA-seq data. Utilizing an inbred landrace line of *P. vulgaris* (G19833) derived from the Andean pool (Race Peru), a whole-genome shotgun sequencing strategy was performed to obtain a reference genome of common bean. The resulting assembled sequences were organized into 11 chromosomal pseudomolecules including 473 Mb of the ~587-Mb genome and genetically anchored 98% of this sequence, with half of the assembled nucleotides in contigs longer than 39.5 kb (contig N50) (Schmutz et al., 2014). To annotate the chromosomal assembly, the Sanger-derived EST resources was combined with a substantial amount of RNA sequencing (RNA-seq) reads (727 million reads from 11 tissues and developmental stages) with homology-based and *de novo* gene prediction approaches. The chromosomal assembly annotation includes 27,197 protein-coding loci, including 4,491 alternative spliced transcripts, an underestimate that will increase with additional transcriptomes and analyses (Schmutz et al., 2014).

Gene predictions were made with the standard JGI plant gene annotation pipeline. RNA-seq transcript assemblies (43,627) and published ESTs (79,630) from NCBI were aligned to the genome by PASA (Haas et al., 2003). Proteins from soybean, *Arabidopsis thaliana*, poplar, *Medicago truncatula* and wine grape were BLASTXed to the repeatmasked genome and the alignments of proteins were refined by EXONERATE.

Loci were determined from BLAT alignments of PASA EST assemblies and EXONERATE alignments of homologous proteins described above with 2kb wiggle room added. Each locus genomic sequence and homologous proteins and EST ORF in the locus were fed into GenomeScan (Yeh, 2001), FGENESH+ and FGENESH_EST (Salamov et al., 2000) for gene
prediction. A best gene prediction per locus was selected based on EST assemblies and homologous proteins alignment. The selected gene predictions were then fed into PASA pipeline where the EST assemblies were obtained for gene model improvement including adding UTRs.

1.4.2 Linkage Maps in Common Bean

Molecular linkage maps are important tools for candidate gene discovery and cloning, applications in marker-assisted selection (MAS), diversity and evolutionary history studies and cross-species comparisons. As with any species, common bean benefits from the availability of molecular linkage maps which differ in characteristics such as the parents used, the segregating generation in which these were established, traits evaluated in each population, the type and number of markers used, the degree of saturation and total length of the map (Broughton et al., 2003). Most of these genetic maps have been used to identify genes and QTL for disease resistance and morphological traits (reviewed in Miklas et al., 2006). Random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP) markers are often used to create molecular maps. More recently, simple sequence repeat (SSR), also called microsatellites markers have also become a prominently used type of molecular marker to increase the density of existing maps.

Of all the populations used, a collection of over 25 linkage maps have been developed from crosses both within and between the two common bean gene pools: Mesoamerican and Andean (Kelly et al., 2003). Particularly, Freyre et al. (1998) developed a core linkage map by developing recombinant inbred (RI) population based on a cross between BAT93, a Middle American cultivar, and Jalo EEP558, an Andean landrace (hereafter referred to as the BJ
population) and used it to extend an RFLP map (Nodari et al., 1993a). This map has a total length of 1226 cM and comprises 563 markers, including 120 RFLP and 430 RAPD markers, in addition to a few isozyme and phenotypic marker loci. The BJ population is also generally considered to be a community-wide mapping population that is used by multiple research groups (Nodari et al., 1993; Miklas et al., 2000, 2001, 2003; Park et al., 2001; Schneider et al., 1997; Tar’an et al., 2001; McClean et al., 2002; Kelly et al., 2003; Kolkman and Kelly 2003; López et al., 2001; Kelly and Vallejos 2004; Murray et al., 2004; Papa et al., 2005; Papa et al., 2007) and this BJ map was subsequently used to integrate other linkage maps (Freyre et al., 1998).

1.5 Comparison of Genome in Legumes

Soybean (*Glycine max* L.) can be considered a model legume plant (Vodkin et al., 2008) because a complete genome sequence has been available for more than 5 years. The size of the soybean genome is estimated to be 1,115 Mb. It consists of 978 Mb arranged in 1190 scaffolds in 20 chromosome pseudomolecule sequences with a small additional amount of mostly repetitive sequence in unmapped scaffolds (Phytozome V1.0). One hundred and twenty-three scaffolds (99.2% of the genome) are larger than 50 Kbp. In addition, approximately 1.6 million ESTs, some 454 ESTs and 1.5 billion paired-end Illumina RNA-seq reads with homology-based gene predictions. There are 56,044 protein-coding loci and 88,647 transcripts that have been predicted (Phytozome V1.0, http://www.phytozome.net/soybean.php).

It is generally believed that *P. vulgaris* (common bean) and *G. max* (soybean) diverged ~19.2 Mya but shared a whole-genome duplication (WGD) event ~56.5 Mya (Paterson et al., 2009). *G. max* experienced an independent WGD ~10 Mya (Schmutz et al., 2010). As a
consequence of the new duplication, any given genomic region in common bean (*Phaseolus vulgaris* L.), are likely to correspond to two genomic regions in soybean (Schmutz et al., 2010). Typically, most of the *P. vulgaris* genes (91%; 24,861) were in identifiable synteny blocks compared to *G. max*, and 57% were in synteny blocks within *P. vulgaris* itself, as a result of the ancient WGD event ~55 Mya (Schmutz et al., 2014). Furthermore, McClean et al. (2010) compared 300 mapped gene-based loci from *P. vulgaris* against all of the scaffold sequences from the initial build of *G. max* genome. Their results showed that not only nearly all of the bean genes are duplicated in soybean but also a single linkage group in bean is syntenic with multiple soybean chromosomes (Figure 1.2). Galeano et al. (2011) found the same pattern of macrosyteny in comparisons between common bean and soybean genomes. This indicates that the soybean genome was fractionated and the fragments rearranged following the duplication event, to construct the modern soybean genome with 20 chromosomes.
1.6 Yield QTL Studies in Soybean

In the past two decades, many studies have focused on mapping of QTLs influencing soybean yield. Due to the small number of ancestors and subsequent breeding and selection practices, the genetic base of soybean cultivars in North America is narrow. The primary gene pools including elite soybean germplasm, exotic soybean lines and *Glycine soja* that (Hymowitz and Singh, 1987) were used to improve the current soybean germplasm. Yield QTLs of soybean were mapped by Orf et al. (1999) from the northern elite cultivar Archer and in Minsoy and Noir
1 on linkage group D2, F and accounted for 8% and 13% of the phenotypic yield variation, respectively (Orf et al., 1999).

However, the limited diversity in elite soybean germplasm has resulted in evaluating *G. soja* as a new source of genetic diversity for improving the crop. With widespread genetic markers, researchers identified positive transgressive segregants and mapped useful QTLs from *G. soja* (Ertl and Fehr 1985; Graef et al., 1989; Suarez et al., 1991). Useful QTL from *G. soja* were discovered in a population of F2-derived lines from a cross between the *G. max* experimental line A81–356022 and the *G. soja* accession PI468916. In this population, QTL controlling hard seededness, plant morphology, maturity (Keim et al., 1990), and protein and oil concentration (Diers et al., 1992) were mapped. The *G. soja* alleles for the protein QTL were associated with greater protein concentrations than the *G. max* alleles (Diers et al., 1992). These *G. soja* alleles have been backcrossed into the background of the *G. max* parent and one region continued to increase protein concentration in a BC3 population, although it was associated with reduced yield (Sebolt et al., 2000). Four significant yield QTLs were identified on linkage groups C2, E, K and M by using a series of backcross populations developed from *G. soja* soybean (donor parent) and soybean cultivar (recurrent parent) (Wang et al., 2003).

One yield QTL identified by Concibido et al. (2003) was reported to be associated with increased seed yield from *G. soja* marked on LG B2 and provided a significant 9.3% yield increase across the test environments which suggest the potential of using exotic germplasm to improve soybean yield. Kabelka et al. (2004) identified nine QTLs for the increase of seed yield in North American cultivars. One positive yield QTL was identified on LG A1 explaining 14% of the phenotypic variation and increased seed yield 2.1% across 12 environments. Li et al. (2008)
reported two positive yield QTLs from *G. soja* PI245331 in an elite soybean genetic background marked on LG A1 with an additive yield effect of 191 to 235 kg ha$^{-1}$, and on LG J with increased seed yield of 374 to 430 kg ha$^{-1}$ and accounting for 10% to 13% of the phenotypic variance. Mansur and Orf (1996) identified a positive yield QTL (A060) from Minsoy on LG J by developing a recombinant inbred line (RIL) population from crossing Noir 1 and Minsoy. Specht et al. (2001) developed a F7:11-derived RIL population from the cross of Minsoy and Noir 1 and identified a positive yield QTL from Noir 1 marked by G815 on LG J. This QTL explained 2% of the phenotypic variation. Wang et al. (2004) also reported four positive yield QTL alleles from *G. soja* PI 468916 on LG K, C2, E and M.

Additionally, studies on QTL alleles from exotic soybean germplasm that significantly increase seed yield have been reported previously. Kabelka et al. (2004) identified nine positive yield QTL alleles that trace to the exotic soybean germplasm accessions FC 04007B and PI 68508. Li et al. (2008) reported one positive yield QTL allele from *G. soja* and the QTL mapped to the same region on chromosome 5 where Kabelka et al. (2004) also reported a yield QTL. Guzman et al. (2007) identified eight positive yield QTL alleles from PIs but all of them mapped to the same regions where yield QTL were reported previously. Although these results suggest that it may be difficult to identify new positive yield QTL from exotic germplasm, there is a need to identify these positive alleles to help increase the rate of yield improvement of future cultivars. Kim et al. (2012) identified QTL and epistatic interactions associated with important agronomic traits in soybean using two backcross populations that each had a different PI as the donor parent. They confirmed the yield QTL in the E and W population. In the E population, three yield QTLs were identified on chromosome 4, 14 and 18 across the six selected environments. 12.2%, 7.8%
and 10.6% of the phenotypic variance for yield were accounted by the three QTLs over the two years, respectively (Kim et al., 2012). On the other hand, in the W population, two yield QTLs were identified on chromosome 3 and 16 (Kim et al., 2012). Although in most of the reports, yield QTLs are marked on different linkage groups and positions, several yield QTLs have been reported on chromosome 3 (Smalley et al., 2004; Kassem et al., 2006; Kabelka et al., 2004; Specht et al. 2001; Wang et al., 2004). Furthermore, most of the phenotypic values of yield traits were examined in single genetic background for QTL analysis while few of these studies have used common parents among their multiple genetic backgrounds (Orf et al., 1999; Hyten et al., 2004), which is a very important factor influencing quantitative traits like yield components (Specht et al., 1999; Liao et al., 2001; Hyten et al., 2004; Zhang et al., 2004). One exception is that Han et al. 2012 identified stable QTL for seed weight of soybean in different genetic backgrounds and environments, using three populations with one common male parent, ‘Hefeng 25’. Besides, many results indicated that yield of soybean is controlled by a series of QTLs with different expression in different genetic populations or environments.

1.7 Characterization of a Candidate Yield Gene in Phaseolus vulgaris based on the Identification of its Effects on Seed Yield in Arabidopsis and Brassica napus

The subject of this thesis is a gene in P. vulgaris that is homologous to a gene called BnMicEmUP that was found to be involved in microspore culture in Brassica napus (rapeseed, canola) (Shahmir and Pauls, 2014). The B. napus gene was shown to be homologous to AT1G74730 in Arabidopsis and both genes were shown to affect seed production (Shahmir and Pauls, 2014). In particular, BnMicEmUP codes for protein sequences that contain Domain of Unknown Function 1118 (DUF1118) (Figure 1.4) consisting of a DNA-binding domain and a
chloroplast targeting region. BnMicEmUP was proposed to be a possible gene homolog of PEND proteins which is a member of the bZIP (leucine zipper) family because they shared the same features. Furthermore, a more complete analysis has been conducted by Shahmir and Pauls (2014) to investigate the functions of AT1G74730 and BnMicEmUP by constructing the transgenic Arabidopsis plants with over expressed AT1G74730 or reduced expression. Transgenic plants phenotypes showed that the over-expressing AtBnMicEmUP lines had a reduced seed yield while the knock down lines, including the RNAi silencing group and the mutant lines, showed higher yield. The results lead to the conclusion that the BnMicEmUP gene functions as a suppressor to the plant yield.

From its structure, BnMicEmUP appears to be a regulatory factor gene with dual targeting, first to the chloroplast and later, to the nucleus (Shahmir, 2014). It is proposed that BnMicEmUP is stored in chloroplasts as an inactive transcription factor and is released from the plastid envelope and translocated to the nucleus when induced by stress signals, including increased ABA levels (Figure 1.5). A model for ABA signaling that includes BnMicEmUP functioning as a transcription factor like AREB/ABI5 was suggested by Shahmir (2014, PhD thesis). In the model, ABA binds to ABA receptor (PYR/PYL/RCAR; Ma et al. 2009, Park et al. 2009) resulting in a type 2C protein phosphatase (PP2C) inhibition while activation of a SNF1-related protein kinase (SnRK2) that can phosphorylate transcription factors such as ABI5 (Nakashima et al., 2009) (ABSCISIC ACID-INSENSITIVE MUTANT 5) and BnMicEmUP. Then, the activated ABI5, BnMicEmUP and other transcription factors may induce the expression of an inhibitory protein; thus resulting in rapid responses in cells to stresses during seed germination and vegetative growth.
1.8 Plant Heat Stress Response

When faced with the slight increase of the environmental temperatures, plants may respond to the stress in many ways. The structure of organelles, membrane, chloroplast and cytoskeleton might be affected by the stress, leading to the protein activity metabolism changes and plant growth variation. Plant genomes have therefore evolved to meet environmental challenges and many plant genes are dedicated to stress protective mechanisms. Examples of adverse environmental factors include drought, salinity, solar radiation (excess light or high light intensities, UV-light), extreme temperatures (heat and low temperature) (Tsanko S. Gechev 2012). The membrane plays important roles in heat shock tolerance by regulating the sensor signaling ability. The sensors are able to prospect and transfer a heat shock into a signal that simulated the heat stress genes to be active (Plieth, 1999). The membrane alters its permeability and remolds membrane lipid to strengthen the membrane rigidity and leads to Ca\(^{2+}\)-influx variation, resulting in a modification of the heat-stress-related gene expression (Saidi et al., 2010). The chloroplast, also a sensor of heat stress, may be structurally modified by reorganization of chloroplasts in thylakoid membranes, leading to grana stacking loss (McDonald and Paulsen, 1997). As a result, the photosynthetic activity of chloroplasts might be reduced (Zhang et al., 2005).
Figure 1.3 Comparison of four forms of \textit{B. napus} BnMicEmUP protein sequences with \textit{Arabidopsis} AT1G74730 (Shahmir, 2014)

Figure 1.4 Comparison of final plant heights in \textit{Arabidopsis} seedlings (Shahmir, 2014)
1.9 Objectives and Hypotheses

The overall goal of the thesis is to isolate and characterize the function a homolog of AT1G74730 and BnMicEmUP from common bean (*P. vulgaris* L).

**Hypotheses**

1) It is hypothesized that seed yields among different varieties of *P. vulgaris* are negatively associated with the levels of expression of a *P. vulgaris* homolog of AT1G74730 and BnMicEmUP.

2) It is anticipated that homologs of bean gene homologous to AT1G74730 and BnMicEmUP can be identified in syntenic regions of the soybean genome that have previously been observed to contain genetic markers for seed yield.
The objectives of the work were to:

1. examine possible associations between the transcripts levels of a homolog of AT1G74730/BnMicEmUP and yield component parameters in *P. vulgaris* L.

2. clone and isolate the AT1G74730/BnMicEmUP homolog from *P. vulgaris* L.

3. perform an *in silico* analysis of AT1G74730/BnMicEmUP homolog DNA and amino acid sequences to identify potential functional domains.

4. characterize specific elements within the AT1G74730/BnMicEmUP homolog promoter region.

5. examine syntenic relationships between bean and soybean in regions surrounding AT1G74730/BnMicEmUP homolog
CHAPTER 2 MATERIALS AND METHODOLOGY

2.1 Field Trial

2.1.1 Site Descriptions

In 2013 growing season, planting sites were located in Elora, ON (43°38'N, 80°24'W) and Woodstock, ON (43°08'N, 80°47'W). The precipitation and temperature data used in the current work was collected at the Elora Research Farm (Environment Canada, 2014). The daily temperature was calculated averaged maximum and minimum temperatures. Further descriptions of each field can be found in Chapter 3.

2.1.2 Experiment Design

A collection of twelve *P. vulgaris* L varieties listed in Table 2.1 were used for the 2013 yield trial. Their selection was based on their performance in yield trials of varieties released in Canada since the 1930s (Navabi et al., 2012). These varieties were classified into high-yielding, medium-yielding and low-yielding classes. Seeds of the selected varieties were obtained from the yield trial from Elora Research Station of the 2012 growing season and were cleaned and packed in envelopes of 350 seeds for each plot. The twelve varieties were grown in a randomized complete block design with four replications. Experiment was grown as 36 cm rows, 6 m row length, 4 row plots with 50 plants (cropping density of 200,000 plant/ha). The fields were sprayed pesticides used in conventional agricultural practices.
2.1.3 Grain Yield

Harvest occurred in mid-October at the Elora and Woodstock sites. A Wintersteiger (Innkreis, Austria) small plot combine was used to harvest four rows of each plot. Harvested grain from individual plots was collected in cloth bags weighed and a subsample was taken for moisture determination. The plot area was calculated using the width of three bean rows and the recorded length from each individual plot. Maturity was recorded when 80% of the bean plants in each plot were dry.

In addition to measuring the total yield (kg/ha), which was the mean seed weight adjusted to 18% moisture for each variety, the average yield gain was also evaluated. Average yield gain (kg/ha/day) describes the seed weight accumulated per growing day from a unit area for each variety and is calculated by the formula:

\[
\text{Average yield gain (kg/ha/day)} = \frac{\text{Total yield (kg/ha)}}{\text{Maturity time (days)}}
\]

Table 2.1 The twelve navy bean varieties used for the field trials at Elora and Woodstock, ON during 2013 growing season (L stands for low-yielding, M stands for medium-yielding, H stands for high-yielding).
2.1.4 Plant Tissue Harvesting

Plant tissues were harvested directly from the field grown plants. Sharp, clean razor blades were sterilized with 97% ethanol prior to using them to harvest tissue from the plants. To harvest leaves, 10 to 20 fresh young trifoliate leaves were randomly taken from each plot and immediately flash-frozen in liquid nitrogen in a 5 ml Nuclease-free tube. The tubes were placed into the liquid nitrogen tank and brought back to the molecular lab at the University of Guelph. For long-term storage, plant tissues were ground to fine powder with a mortar and pestle in liquid nitrogen and stored at -80 °C.

2.2 Nucleic Acid Extraction

2.2.1 DNA Extraction

Fresh leaf tissues were harvested, immediately flash-frozen in liquid nitrogen and ground with a ceramic mortar and pestle. A DNeasy Plant Mini Kit (QIAGEN Inc., Mississauga) was used for DNA extraction, and the DNA was isolated according to the manufacturer's instructions. Approximately 100 mg of the fine powder were placed into a 2 ml safe-lock microcentrifuge tube and 400 μl Lysis Buffer (AP1) and 4 μl RNaseA stock solution (100 mg/ml) were added to the tube. The mixture was incubated for 10 min at 65 °C, after vigorous vortexing. 130 μl Neutralization Buffer (P3) were added to the lysate, mixed, and incubated for 5 min on ice. After incubation, the lysate were pipetted into the QIAshredder Mini spin column (DNeasy Plant Mini kit, Qiagen) in a 2 ml collection tube, and centrifuged for 2 min at 20,000 × g. The flow-through was transferred to a new tube with 1.5 volumes of Wash Buffer (AW1) to the cleared lysate. 650 μl of the flow-through mixture was pipetted into the DNeasy Mini spin column placed in a 2 ml
collection tube. The flow-through was discarded after 1 min centrifugation at 13,000 \times g. The DNeasy Mini spin column was placed into a new 2 ml collection tube with 500 μl Wash Buffer (AW2), and centrifuged for 1 min at 20,000 \times g. The flow-through was discarded. Another 500 μl of Buffer AW2 was added to the DNeasy Mini spin column, and centrifuged for 2 min at 20,000 \times g to dry the membrane. The DNeasy Mini spin column was transferred to a new 2 ml microcentrifuge tube and 100 μl Elution Buffer (AE) was pipetted directly onto the DNeasy membrane. To elute the DNA, the tube was incubated for 5 min at room temperature (15-25 °C), and then centrifuged for 1 min at 20,000 \times g.

2.2.2 RNA Extraction

Total RNA was extracted from the young trifoliate leaves of the common bean using the RNeasy plant mini kit (QIAGEN Inc., Mississauga). The leaf tissues were immediately placed into liquid nitrogen and ground thoroughly with a chilled mortar and pestle. Less than 100 mg of the fine powder was weighed and transferred to an RNase-free, liquid-nitrogen–cooled, 2 ml microcentrifuge tube with 450 μl RNeasy Lysis Buffer (RLT) (RNeasy Plant Mini Kit, QIAGEN) and vortexed vigorously to disrupt the tissue. The lysate was transferred into a 2 ml collection tube with a QIAshredder spin column and centrifuged for 2 min at 20,000 \times g to remove cell debris. The supernatant of the flow-through was transferred into a new microcentrifuge tube and 0.5 volumes of ethanol (96-100%) were added into the cleared lysate. After mixing, 650 μl of the solution was transferred to a 2 ml collection tube with an RNeasy spin column and centrifuged for 15 s at 8,000 \times g, to bind the total RNA to the RNeasy spin column. To wash the spin column membrane, 700 μl of Wash Buffer (RW1) (RNeasy Plant Mini Kit, QIAGEN) were added to the RNeasy spin column and it was centrifuged for 15 s at 8,000 \times g. After discarding the flow-
through, the RNA was washed twice by adding 500 μl of Wash Buffer (RPE) (RNeasy Plant Mini Kit, QIAGEN) to the RNeasy spin column and centrifuging for 15 s at 8,000 × g. 30-50 μl RNase-free water was added directly to the spin column membrane and centrifuged for 1 min at 8,000 × g to elute the RNA.

2.2.3 RNA Purification and Analysis

To facilitate removal of the genomic DNA, DNase treatment was applied to all the RNA samples according to the instructions of the DNA-free kit (Ambion, Austin, TX). For each reaction, 5 μl (0.1 volumes) of 10× DNase 1 buffer and 1 μl of rDNase I were added to the 50 μl RNA sample. The tube was mixed gently and incubated at 37 °C for 20–30 min. Then, 5 μl of the resuspended DNase Inactivation Reagent (0.1 volumes) was added into the solution. After mixing, the tube was incubated at room temperature for 2 min and centrifuged at 10,000 × g for 1.5 min. The supernatant, which contained the RNA, was transferred to a new DNase-free tube and stored at -80 °C.

The amount of RNA was estimated by measuring its absorbance at 260/280 nm using the Nanodrop-1000 (Thermo Fisher Scientific, Wilmington, DE). The RNA quality was confirmed by loading 5-10 μl of the sample and analyzing the ribosomal RNA (rRNA) components on a 1% agarose gel stained with EtBr (Ethidium bromide); at 100V for 1 h and visualizing the two bands of 28S and 18S rRNA with a BIO-Rad Gel Doc System.

2.3 Primer Design

To amplify the Phvul.009G190100 gene from P. vulgaris L, PCR primer pairs were designed to anneal to the sequence for Phvul.009G190100 in the Phytozome database.
(http://www.phytozome.net/) using the online version of the Primer3 software (Rozen and Skaletsky 2000;http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The oligonucleotides were produced by Molecular Biology Unit in Laboratory Services of the University of Guelph (Table 2.1). The optimal annealing temperatures for the primer sets were determined using the BioRad thermo cycler, and a gradient PCR program.

Table 2.2 Summary of Nucleotide sequences of PCR primers

<table>
<thead>
<tr>
<th>Name</th>
<th>Designation</th>
<th>Oligo sequence</th>
<th>Annealing temperature</th>
<th>Product size (bp)</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC-2</td>
<td>Forward</td>
<td>CTGCAACATCCTCCGCTT</td>
<td>60 °C</td>
<td>genomic DNA: 638</td>
<td>Partial isolation of Phvul.009G190100 genomic and cDNA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>AGCCACGAAGAGCCCCAA</td>
<td></td>
<td>cDNA: 554</td>
<td></td>
</tr>
<tr>
<td>PV09</td>
<td>Forward</td>
<td>CAGTTCTTCCACGCCTCTCC</td>
<td>61 °C</td>
<td>272</td>
<td>Real time PCR amplification of Phvul.009G190100</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CCCTTCGTAATAAGCCCCCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKIP16</td>
<td>Forward</td>
<td>CACCAGGATGCAAAAAGTGG</td>
<td>58 °C</td>
<td>163</td>
<td>Monitoring housekeeping gene expression</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ATCCGCTTTGTCCTTGAAC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.4 Polymerase Chain Reaction

The 25 μl PCR mix contained the following components: 2.5 μl of 10× PCR buffer (500 mM potassium chloride, 15 mM magnesium chloride, 10 mM Tris-HCl pH 9.0, 1% Triton X-100), 3 μl MgCl₂ (25 mM), 1 μl of dNTP (Invitrogen® 10 mM dNTP), 1 μl of Taq DNA Polymerase (Sigma®), 1 μl of the forward and reverse primers (10 μM) and 50 ng DNA. The components were added to a 0.2 ml PCR tube for each reaction and loaded into the PCR machine. PCR amplification was performed in a programmable thermocycler (Bio-Rad®)
iCycler™) with the cycles shown in Table 2.3. A variable number of reaction cycles and annealing temperatures are used, depending on the target gene (Table 2.2).

Table 2. 3 PCR cycles and time of each step used to amplify Phvul.009G190100

<table>
<thead>
<tr>
<th>Cycling Step</th>
<th>Temperature</th>
<th>Hold Time (min: sec)</th>
<th>Cycle #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiation denaturation</td>
<td>94 °C</td>
<td>5:00</td>
<td>1</td>
</tr>
<tr>
<td>Denaturing</td>
<td>94 °C</td>
<td>0:30</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>50-70 °C</td>
<td>1:00</td>
<td>35</td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>1:00</td>
<td></td>
</tr>
<tr>
<td>Elongate DNA</td>
<td>72 °C</td>
<td>10:00</td>
<td>1</td>
</tr>
<tr>
<td>Hold</td>
<td>4 °C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.5 Agarose Gel Electrophoresis

The PCR products were analyzed by gel electrophoresis. The 1% agarose gel was made by dissolving the agarose in 1× TBE running buffer (Tris base, boric acid and EDTA) by heating in a microwave for 1-1.5 min. When the solution cooled down to approximately 45 °C, 60 μl/ml of 10 mg/ml EtBr (ethidium bromide) was added, mixed well, and poured into a tray with an appropriate comb to create sample wells after cooling for approximately 30 min. 15 μl samples were loaded into the wells after mixing with 2 μl of loading buffer. All the gels were run with a 1 kb DNA ladder (Invitrogen) as a size reference. Gels were run at 100V for 40 min to an hour and visualized under UV light using a Gel Doc 1000 (BioRad, Hercules, CA, USA).
2.6 Gene Cloning

2.6.1 PCR Product Purification

PCR products were isolated from the 1% agarose gel using the PureLink® Quick Gel Extraction Kit (Invitrogen, Carlsbad, CA). A gel slice containing the desired DNA fragment was excised using a clean sharp razor blade under UV light and transferred into 1.5 ml microcentrifuge tube. The gel slice was weighed with the tube using a scale sensitive to 0.001 g. To dissolve the gel, the tube(s) containing the gel slice and Gel Solubilization Buffer (L3) was placed into a 50 °C water bath for at least 10 min. The tube was inverted by hand every 3 min to mix and incubated for an additional 5 min after the gel slice dissolved. The dissolved gel piece containing the target DNA fragment was placed onto the center of a Quick Gel Extraction Column inside a Wash Tube to centrifuge for 1 min at 20,000 × \textit{g} in order to bind the DNA. 500 μl Wash Buffer (W1) containing ethanol was added to the Quick Gel Extraction Column and centrifuged at 20,000 × \textit{g} for 1 min. To remove any residual Wash Buffer and ethanol, the column was centrifuged again at maximum speed for 1-2 min. 50 μl Elution Buffer (E5) was added to the center of the Quick Gel Extraction Column on a Recovery Tube. After 1 min incubation at room temperature, the column was centrifuged at 20,000 × \textit{g} for 1 min and the purified DNA was collected in the recovery tube. The purified DNA was stored at 4 °C for immediate use or store at -20 °C for long-term storage.

2.6.2 Cloning into TOPO® Easy Vector

Purified PCR product was ligated with the pCR®2.1-TOPO® vector and cloned to the chemically competent cells with the TOPO TA Cloning® Kit (Invitrogen, Carlsbad, CA). A 6 μl
reaction mix containing 4 μl of DNA (50-60 ng), 1 μl of salt solution (200 mM NaCl and 10 mM MgCl$_2$) and 1 μl of TOPO® vector (50 ng/μl), which were mixed gently and incubated for 5 min at room temperature (22-23 ºC). To perform the rapid chemical transformation protocol, 2 μl of the TOPO® Cloning reaction were added into a vial of One Shot ® E. coli, DH5-T1 competent cells. The reaction were mixed gently and incubated on ice for 30 min. Then, another quick heat-shock was performed by incubating the tube at 42 ºC for 30 s and then transferred to the ice by adding 250 μl room temperature S.O.C medium. The plasmids were grown by shaking the tube horizontally (200 rpm) at 37 ºC for 1 h. 50 μl and 80 μl of the transformation were spread onto a prewarmed plate with the LB (Lucia-Bertani) selective medium (containing 5 mg/ml yeast extract, 10 mg/ml bactotryptone, 10 mg/ml NaCl, 15 mg/ml bactoagar and 50 μg/ml ampicillin) to culture at high and low concentration, respectively. The plates were incubated overnight (18-20 h) at 37 ºC.

2.6.3 PCR Confirmation for Positive Clones

Colony PCR was performed to confirm the positively transformed colonies with target inserts. All the procedures of the colony PCR are the same as the regular PCR protocol except for the template used are 5 to 10 light blue or white colonies picked from the LB plate using a yellow tip and dipped in a PCR reaction vial. The tip was placed into a LB liquid culture medium containing 50 μg/ml ampicillin for overnight shaking at 37 ºC.

2.6.4 Plasmid Extraction

The plasmid DNA was isolated from the LB liquid culture by PureLink Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA). According to the manuscript, 1 to 5 ml of an overnight
culture \(E. coli\) in LB medium) was transferred into a new tube and resuspended the pellet in 250 µl Resuspension Buffer (R3) with RNase A. Then, 250 µl Lysis Buffer (L7) was added to the cells and the tube was incubated at room temperature for 5 min. 350 µl Precipitation Buffer (N4) was added to the cells and mixed immediately until the solution is homogeneous. To clarify the lysate from lysis debris, the mixture was centrifuged at \(12,000 \times g\) for 10 min at room temperature using a microcentrifuge. The supernatant was loaded onto a Spin Column with a 2-ml Wash Tube. To purify the plasmid DNA, 700 µl Wash Buffer (W9) with ethanol to the column and the tube was centrifuged twice at \(12,000 \times g\) for 1 min to remove the residual. To elute the DNA, the Spin Column was placed in a clean 1.5 ml Recovery Tube with 75 µl of preheated TE Buffer (TE) to the center of the column. The column was incubated at room temperature for 1 min and centrifuged at \(12,000 \times g\) for 2 min. The purified plasmid DNA was collected in the Recovery Tube and stored at -20 °C.

2.6.5 Plasmid DNA Sequencing

Plasmid DNA containing the PCR product were sent to the University of Guelph Genomics Facility where samples were sequenced by the Sanger method (Sanger et al., 1997) using an Applied Biosystem BioDye Terminator V3.1 and an Applied Biosystems 3730 sequencer. The clones were sequenced in both orientations using forward and reverse FC-2 primer.

2.7 Real-time PCR (qPCR)

Real-time PCR was carried out using iQ™ SYBR Green PCR supermix (Bio-Rad Laboratories). Each 20 µL reaction containing the following contents: Nuclease-free water,
forward and reverse primers (300 nM), 10 μl iQ™ SYBR Green PCR supermix, and genomic DNA or cDNA template (100 ng). The reaction replicates were loaded onto the 96 microplates and the vessels were sealed. The BioRad iCycler iQ thermal cycler was used for performing the real-time PCR. The programme of the real-time PCR was set with a melting curve as described in Table 2.4. The sealed reaction vessels were placed in the block of Bio-Rad Multicolor Real-Time Detection System and the analysis was performed using the iCycler iQ Optical System Software Version 3.1 (Bio-Rad laboratories, Mississauga, Canada). All qRT-PCRs were performed using 3 technical replicates for each sample and each block included one positive control (genomic DNA) and a negative control with no template.

Table 2.4 A 3-step real-time PCR protocol with melting curve

<table>
<thead>
<tr>
<th>Cycling Step</th>
<th>Temperature</th>
<th>Hold Time (min: sec)</th>
<th># of cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiation denaturation and enzyme</td>
<td>95 °C</td>
<td>2:00-3:00</td>
<td>1</td>
</tr>
<tr>
<td>activation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturing</td>
<td>95 °C</td>
<td>0:10-0:15</td>
<td>40</td>
</tr>
<tr>
<td>Annealing</td>
<td>55-60 °C</td>
<td>0:15-0:30</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>0:30</td>
<td></td>
</tr>
<tr>
<td>Melting curve</td>
<td>55-95 °C (in 0.5 increments)</td>
<td>0:10-0:30</td>
<td>1</td>
</tr>
</tbody>
</table>

2.7.1 First Strand cDNA Synthesis (RT-PCR)

Complementary DNA (cDNA) was synthesized according to the manufactures instructions for qScript cDNA SuperMix Kit (Quanta Biosciences, Inc., Gaithersburg, USA). Each single 20 μl reaction contained the following components: 4 μl qScript cDNA SuperMix (5×), and 800 ng template RNA, variable nuclease-free water was added to fill the reaction to 20
μl. To ensure that all signals came from the cDNA, a negative RT control (no addition of the qScript Reverse transcriptase) and a no RNA control (no addition of the RNA) were included as a template for the RT-PCR. RT reactions were placed in a thermal cycler programmed as in Table 2.4. After process completion, 4 μl of this RT reaction were used as template for downstream PCR amplification to confirm the cDNA fragments. 25 μl of the final PCR product were run out on a 1% agarose gel, viewed and blotted for band verification. The remainder of the positive PCR product was used as the template for downstream cloning. The successfully reverse transcription cDNA were stored at -20 ºC for real-time PCR experiment and at -80 ºC for long-term usage.

Table 2. 5 Thermal cycler programmed reverse transcription to produce cDNA

<table>
<thead>
<tr>
<th>Cycle #</th>
<th>Temperature</th>
<th>Hold Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25 ºC</td>
<td>5 min</td>
</tr>
<tr>
<td>1</td>
<td>42 ºC</td>
<td>30 min</td>
</tr>
<tr>
<td>1</td>
<td>85 ºC</td>
<td>5 min</td>
</tr>
<tr>
<td></td>
<td>4 ºC</td>
<td>hold</td>
</tr>
</tbody>
</table>

2.7.2 **Primers**

Primer pairs PV09 were designed to amplify Phvul.009G190100 fragment with a smaller size (272 bp) (Figure 3.15).

2.7.3 **Relative Standard Curve**

SKIP16 (SKP1/ACK-interacting protein 16) was used as the internal control due to its consistently high expression in all tissues in common bean (Borges et al., 2012). Relative standard curves for Phvul.009G190100 and the housekeeping gene (SKIP16) were constructed
using plasmid DNA containing the target sequences (standard curves are shown in Chapter 3). Each plasmid vector was ten-fold serially diluted six times and the dilutions were used throughout the experiment. These serial dilutions or gene copies start from 4.07e+7, 4.07e+6, 4.07e+5, 4.07e+4, 4.07e+3, 4.07e+2, 4.07e+1.

2.7.4 Data Analysis for Quantification of Gene Expression

Real time PCR results were expressed as Ct (cycle threshold) values. This value corresponded to the cycle at which the fluorescence of the SYBR Green dye reached above the threshold or background fluorescence value. In quantifying expression level of the target gene (Phvul.009G190100) in different samples, the relative expression level was obtained by the cycle differences of the target gene (Phvul.009G190100) and the reference gene (SKIP16).

2.7.5 Comparative Ct Method

The comparative Ct method, also referred to as the ∆∆CT Method, was used to analyze the relative quantitation of the samples by arithmetic formulas. The prerequisite for the ∆∆CT Method is that the PCR efficiencies between the target(s) and endogenous control(s) should be relatively equivalent.

2.7.6 Validation Experiments

For a valid ∆∆Ct calculation, the efficiency of the target amplification and the efficiency of the reference amplification must be approximately equal. To validate the efficiency, the ∆Ct (Ct target–Ct reference) was calculated for the template dilution by running standard curves for each amplicon utilizing the same sample. These ∆Ct values were plotted with log input amount to create a semi-log regression line. The slope of the resulting semi-log regression line can be
used as a general criterion for passing a validation experiment. In a validation experiment that passes, the absolute value of the slope of ΔCt vs. log input would be < 0.1.

2.8 Sequence Analysis

Common bean sequence results were obtained from Phytozome v10 (http://www.phytozome.net/) by BLAST with the Arabidopsis sequence AT1G74730 acquired from TAIR (http://www.arabidopsis.org/). Soybean homologous genes and the expression profile were BLASTed against Soybase (http://www.soybase.org/). Other orthologous protein and nucleotide sequences were retrieved from NCBI (http://www.ncbi.nlm.nih.gov) and aligned with common bean sequences using CLUSTALW version 1.81 multiple sequence alignment program (Chenna et al., 2003; Higgins et al., 1992) (http://workbench.sdsc.edu/) using default parameters and with manual modifications. To predict protein structure, the Phyre2 server was used (Kelley and Sternberg, 2009).

2.9 Phylogenetic Analysis

Phylogenetic and molecular evolutionary analyses of the Phvul.009G190100 gene family were conducted using Workbench, version 3.0.2 (CLC bio, Aarhus, Denmark) software. Translated protein sequences were aligned and 60–amino acid regions containing putative homeodomains were chosen for phylogenetic analysis. They were imported into the CLC Workbench (CLC Bio, Aarhus, Denmark) program and globally aligned by CLUSTALW. A phylogenetic tree was generated with neighbor-joining methods, and bootstrap values were calculated.
The current syntenic map was developed by Mapchart (Voorrips et al., 2002) in silico based on the McClean et al. (2010) common bean-soybean comparative map and the Phenylpropanoid pathway gene sequence-based common bean-soybean comparative map (Reinprecht et al., 2013). Synteny blocks were defined by the presence of three or more sequence features shared by the common bean and soybean genomes. Molecular markers used included common bean gene-specific (McConnell et al., 2010) and random RAPD, STS and simple sequence repeats (SSRs). Marker sequences and were retrieved from the Legume Information System (LIS, http://www.comparative-legumes.org/) and/or Phaseolus Genes database (http://phaseolusgenes.bioinformatics.ucdavis.edu/) using the default settings.

2.10 Statistical Analysis

All statistical analyses were performed using SAS 9.4 (SAS Institute Cary, NC) and Microsoft Excel 2007 (Microsoft Corporation). Data were analyzed with either paired Student’s t-test or one-way ANOVA to assess the variations in expression level of genes in a series of varieties of common bean. The analysis was conducted in a randomized complete block design with four replications. Analysis of variance (ANOVA) of total yield, gene expression level was performed in the Proc Mixed procedure to assess the effect of varieties on the yield and levels of gene expression to evaluate significant differences.

Least square means were generated and compared pair-wise using Tukey’s test with a Type I error rate of 0.05. Group means and differences between the high-yielding, medium-yielding and low-yielding groups of common bean varieties were determined using estimate statements within the Proc Mixed analysis.
To determine the association between the gene expression level and yield, correlation and regression analyses of variance were performed using PROC CORR and PROC GLM in ANOVA.

Residuals of each model were tested through PROC UNIVARIATE and PROC PLOT to determine if error assumption criteria were achieved. Normal distribution of the residuals was evaluated with the Shapiro-Wilk statistic and a histogram of residual values. Examination for patterns of a residual by predicted scatter plot was used to verify random distribution. Finally, variance homogeneity was tested by evaluating uniformity of the range of residual data points across each level of the fixed effects. Outliers were detected using Lund’s test of outliers with a Type I error rate of 0.05 (Lund, 1975).
CHAPTER 3 RESULTS

3.1 Environmental Variability

In the 2013 growing season, planting sites were located in Elora, ON (43°38'N, 80°24'W) and Woodstock, ON (43°08'N 80°47'W) (Figure 3.1). Henceforth, the names of the locations will be referred to as Elora 2013, Woodstock 2013 and historical to distinguish the location of studies.

Environmental variation was observed, during the growing season (June to October), between the experimental locations (Figure 3.2, Figure 3.3). For example, although some temperature values in Woodstock were missing, the overall average temperature during the growing time in Woodstock was observed (17.4 °C), which was 0.7 °C higher than the average temperature in Elora (16.7 °C). There were two peak days that the temperatures were above 25 °C on planting days 36 and 91 in Elora, and the same situation was found in Woodstock for planting days 27 and 81, respectively. Moreover, the cumulative amount of rainfall in Elora (500 mm) was nearly twice more than that observed in Woodstock (226 mm). Additionally, in Elora the rainfall accumulated steadily over the full season, whereas Woodstock had a dry period from 46 to 90 days after planting, but a heavy rainfall following that brought the cumulative precipitation up to 226 mm (Figure 3.3). Unfortunately, the plots in Woodstock could not be harvested because of the heavy rainfall during harvesting time.
Figure 3.1 Planting locations for 2013 yield trial. Red spots indicate the locations of Elora, ON (43°38' N, 80°24'W) and Woodstock, ON (43°08'N, 80°47'W) planting sites.

Figure 3.2 Daily average temperature and accumulated precipitation in Elora, ON from June 9th, 2013 to October 3rd, 2013. June 9th was the planting date at Elora ON; Oct 3rd is the day of harvest.
3.2 Analysis of Yield Data

3.2.1 Total Yield

The 12 navy bean genotypes were significantly different (P < 0.001) for yield in the Elora research station trial during the summer of 2013 (Table A-1). The yields in the 2013 trial were positively correlated with the yields in the historical trial (r = 0.4454, P = 0.0081), with 52% of the yield data accounted for by the regression model (Figure 3.5). Among the 12 genotypes used in the study, Sanilac and Saginaw are the only two determinate-type varieties. Their total yields were significantly different and lower than the 10 indeterminate-type lines. All varieties, except for Sanilac and Saginaw, had higher yields in the historical trial than in the summer 2013 trial at the Elora research station (Figure 3.4).
In the contrast analysis (Table A-2) the high yielding varieties, including Lightning, Apex, Rexeter and HR200 had significantly different seed yields than the low yielding varieties, including Midland, Sanilac, Hensal and Saginaw. Moreover, the yield of the medium yielding group including AC Compass, OAC Thunder, Vista and Michelite, was significantly different from the low yielding group. Apex had the highest yield among all the selected varieties while Midland was the lowest yielding variety among the indeterminate-type lines in both the historical trial and the Elora 2013 trial (Figure 3.4).

![Graph showing the correlation between yield (kg/ha) for twelve common bean varieties determined at Elora, ON in 2013 and in the historical trial conducted at St. Thomas, ON in 2010 and in St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown, n = 4 for both the historical and the 2013 Elora yield data. Full names of the shown varieties are: Lightning, Apex, Rexeter, HR200, Midland, Sanilac, Hensal, Saginaw, AC Compass, OAC Thunder, Vista and Michelite.](image-url)

\[
y = 0.4454x + 1549.4
\]

\[
R^2 = 0.52 \quad P = 0.0081
\]
3.2.2 Daily Yield Gain (kg/day/ha)

The least square mean values for the daily yield gain (kg/day/ha) were significantly different (P < 0.0001) among the 12 genotypes tested in Elora in 2013 (Table A-3). From the contrast result between historical trial and Elora 2013 yield data, significant differences (P < 0.0001) were observed between the daily yield gains for all the genotypes (Table A-4). However, the Elora 2013 and historical trial average yields were positively correlated (r = 0.4402, P = 0.01), with 44% of the yield data accounted for by the regression model (Figure 3.7). When considering the different growth habits, all the indeterminate-type lines grown in Elora, except for Midland and Michelite, had lower daily yield gains (kg/ha/day) than the historical trial. However, the determinate-type varieties (Sanilac and Saginaw) gained considerably less in Elora 2013 than in the historical trial (Figure 3.6).

A significant difference (P < 0.0001) was also observed between the high yielding group and low yielding group, the same pattern was found between medium yielding group and low yielding group (Table A-5). However, no significant difference (P = 0.7692) was observed between the high yielding group and medium yielding group for the total yield and the average yield (Table A-5).
Figure 3.5 Average yield gains (kg/ha/day) for twelve genotypes at Elora ON in 2013 and in the historical trial; Historical yield trials conducted at St. Thomas, ON in 2010 and St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown, n = 4 for the historical data and the 2013 Elora yield data.

Figure 3.6 Correlation between average yield gain (kg/ha/day) for twelve common bean varieties determined at Elora, ON in 2013 and in the historical trial conducted at St. Thomas, ON in 2010 and in St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown, n = 4 for both the historical data and the 2013 Elora yield data. Full names of the shown varieties are: Lightning, Apex, Rexeter, HR200, Midland, Sanilac, Hensal, Saginaw, AC Compass, OAC Thunder, Vista and Michelite.
3.2.3 Days to Maturity

All the varieties had a significantly different days to maturity in the yield trials of Elora 2013, Woodstock 2013 and the historical. Plants in Elora needed the longest time (~120 days) to mature compared to those in Woodstock which required 100 days to mature (on average) while lines in the historical data set had the shortest maturity times ranging from 90 to 100 days (Figure 3.7).

The ranks of the varieties, with respect to their maturities in the 2013 field trials were positively correlated with their ranks in the historical trial ($r = 0.4502$, $P = 0.03$), with 39% of the yield data accounted for by the regression model (Figure 3.8). Specifically, Midland, which had the shortest days to maturation in Elora and Woodstock, was also the fastest to mature in the historical trial measurement. The two determinate-type lines (Sanilac and Saginaw), which had shorter maturity time than the other indeterminate-type lines (Midland, AC Compass, Lightning, OAC Thunder, Vista, Michelite, Hensal, Apex, Rexeter, HR200) in the historical trial, required longer times to mature in the 2013 trials (Figure 3.9).
Figure 3. 7 Maturity time (days) for twelve genotypes at Elora and Woodstock ON in 2013 and in the historical yield trial conducted at St. Thomas, ON in 2010, St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown, n = 4 for both the historical data and the 2013 maturity data.

Figure 3. 8 Correlation between maturity time (days) for twelve common bean varieties determined at Elora, Woodstock, ON in 2013 and in the historical trial conducted at St. Thomas, ON in 2010, St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown, n = 4 for both the historical data and the 2013 maturity data.

\[ y = 0.450x + 44.99 \]

\[ R^2 = 0.39 \quad P = 0.03 \]
3.2.4 Relationship between Yield (kg/ha) and Yield Gain (kg/ha/day)

Across all varieties, the total yield (kg/ha) was positively correlated with the yield gain (kg/ha/day) for both the historical trial and the Elora 2013 trial ($r = 0.0096$, $P < 0.0001$; $r = 0.0090$, $P = 0.0001$). The Elora 2013 trial showed a wider spread for both measures of yield, namely the total yield (kg/ha) and the average yield gain (kg/ha/day), than the historical trial (Figure 3.10). Sanilac displayed the lowest yield at 1045.7 kg/ha and 9.09 kg/ha/day while Apex showed the highest, reaching at 4191.4 kg/ha and 36.46 kg/ha/day for total yield and daily yield gain in the Elora 2013 trial, respectively.

Figure 3. 9 Correlation between total yield (kg/ha) and Average yield gains (kg/ha/day) for yield trials at Elora ON in 2013 and in the historical trial conducted at St. Thomas, ON in 2010, St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown, $n = 4$ for both the historical data and the 2013 yield data.
3.2.5 Yield Ranks Correlation and Regression

The total yield ranks for the 12 varieties were positively correlated with the historical trial and Elora 2013 trial \( (r = 0.7972, \ P = 0.0019) \), with Apex as the highest yielding variety in both trials (Figure 3.10). Although the maturity times were diverse across the genotypes, the ranks of varieties with respect to yield gain (kg/ha/day) in the Elora 2013 trial was positively correlated with the historical trial \( (r = 0.7552, \ P = 0.0045) \) (Figure 3.11). Therefore, although absolute yield values were variable in different environments, the relative yields were consistent among trials.

![Graph showing rank correlation of total yields in historical and Elora 2013 trials.](image)

Figure 3.10 Rank correlation of the total yields (kg/ha) for the twelve genotypes in the historical trial and the yield trial in Elora in 2013. Yields are ranked from 1 to 12 representing the lowest to highest yield values. Each point indicates the rank of the mean value from 4 replicated plots.

\[
y = 0.7972x + 1.3182 \\
R^2 = 0.6355 \ P = 0.0019
\]
Figure 3. 11 Rank correlation of average yield gain (kg/ha/day) for the twelve genotypes in historical trial and the yield trial in Elora in 2013. Yields are ranked from 1 to 12 representing the lowest to highest yield values. Each point indicates rank of the mean value from 4 replicated plots.

### 3.3 Annotation of AT1G74730 Homolog from Common Bean

BLAST searches with *Arabidopsis* AT1G74730 against the Phytozome *P. vulgaris* sequence database (http://www.phytozome.net) suggested that the closest nucleotide match was to Phvul.009G190100, located on *P. vulgaris* chromosome 9 at 28,180,739–28,181,741 bp, which is 74.3% (179 of 241) homologous to the query. A search of the protein database yielded a single homology to Pfam: 06549, a protein of unknown function (DUF1118).
3.4 Isolation of Phvul.009G190100 from *P. vulgaris*

A Phvul.009G190100 gene sequence (homologous to the *Arabidopsis* gene AT1G74730) was amplified from *P. vulgaris* genomic DNA of young Lightning leaves by PCR, using a specific primer pair (FC-2F and FC-2R) (Figure 3.13). The primers were designed to amplify the whole gene based on the available sequence in the Phytozome (http://www.phytozome.net) (Table 2.1). The fragment amplified from the genomic DNA was 671 bp long and contained a partial exon 1 (143 bp, 10 nt downstream from start codon), the intron (84 bp) and the second exon (444 bp) (Figure 3.12).

![Figure 3.12 Amplification of Phvul.009G190100 from P. vulgaris (variety Lightning) genomic DNA with the FC-2 primer pairs. The expected 671 bp fragment is indicated with an arrow.](image)

56
Figure 3. 13 Comparison of the gene sequences for Phvul.009G190100 and AT1G74730. Primers used for genomic DNA amplification are indicated with horizontal black arrows (FC-2). Primers used for qRT–PCR experiments are indicated with horizontal red arrows (PV09). Exons are highlighted in blue and the intron is highlighted in yellow. 5’ un-translated and 3’ un-translated regions are not highlighted. Start and stop codons are shaded in red.

3.5 Cloning of Phvul.009G190100

A Phvul.009G190100 gene sequence (homologous to the Arabidopsis gene AT1G74730) was amplified from P. vulgaris complementary DNA obtained by RT-PCR from young Lightning
leaves, using the primer pair (FC-2F and FC-2R) (Figure 3.14). Figure 3.15 shows an alignment of the cloned Phvul.009G190100 cDNA sequence from bean “Lightning” with the annotated sequence from the Phytozome database. The total cloning length was 587 bp with two partial exons (143 bp, 444 bp) and excluded the intron (84 bp) (Figure 3.14). The cloned fragment was used as a template for constructing the standard curve for validation experiment of the real time PCR.

Figure 3.14 Colony PCR to confirm E. coli uptake of Phvul.009G190100 clone. A single colony of E. coli transfected with purified PCR product of Phvul.009G190100 amplified from cDNA produced from RNA isolated from Lightning leaves. The bacterial colonies grown on LB medium were picked directly from the plates; FC-2 primer pairs were used to amplify the gene from 10 nt downstream from start codon to stop codon. Lanes ladder: DNA ladder 100 bp; Land 1: positive colony 04-01-FC2 with Phvul.009G190100
Figure 3. 15 04-01-FC2 sequence aligned with genomic DNA and cDNA sequences of Phvul.009G190100 from Phytozome. Positions of FC-2 primers were used to amplify the cDNA are indicated by red arrows. Cloning region is highlighted in yellow. Start codon and stop codons are shaded in blue and the intron is shown in grey.

3.6 Gene Structure

The structure of the *P. vulgaris* gene Phvul.009G190100 is very similar to the *Arabidopsis* gene AT1G74730, which consists of two exons (coding region) and one intron (Table 3.1, Figure 3.16). In fact, the two exons in Phvul.009G190100 and AT1G74730 are the same lengths, namely: 153 bp and 444 bp, respectively. In addition, the nucleotide similarity of the exons for Phvul.009G190100 and AT1G74730 are high. The first exon (153 bp) of Phvul.009G190100 has 62.09% identity with the AT1G74730 (153 bp) and the second exon (444 bp) is 67.12% identity to the AT1G74730 second exon (444 bp) (Figure 3.16). Most of the
differences between the exons are single nucleotide polymorphisms which occurred more frequently in the first exon. Phvul.009G190100 contains a shorter (84 bp) intron differed from AT1G74730 intron (240 bp). Phvul.009G190100 has different size of the upstream and downstream nucleotide sequence compared to AT1G74730 (Figure 3.16). The degree of identity between the 5'UTR of Phvul.009G190100 (138 bp) and AT1G74730 (185 bp) was 44.93%. More similarity (56.2%) was found in 3' UTR with 156 bp of Phvul.009G190100 and 137 bp of AT1G74730.

The gene AT1G74730 model was based on the TAIR *Arabidopsis* database where gene models were built using a combination of cDNA and EST sequences from GenBank. The 5' UTR and 3' UTR of each gene was determined on EST alignments. Similarly, as mentioned in Chapter 1, the Phvul.009G190100 gene model from Phytozome was predicted based on the RNA-seq transcript and EST assemblies and homologous proteins alignment. The UTRs were added when the EST assemblies were obtained for gene model improvement.

Table 3.1 Specifics of *Arabidopsis* AT1G74730, *P. vulgaris* Phvul.009G190100 and *B. napus* BnMicEmUP genes

<table>
<thead>
<tr>
<th>Name</th>
<th>Size gDNA (bp)</th>
<th>Size cDNA (bp)</th>
<th>Size CDS1 (bp)</th>
<th>Size CDS2 (bp)</th>
<th>Size Intron (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G74730</td>
<td>1131</td>
<td>891</td>
<td>153</td>
<td>444</td>
<td>240</td>
</tr>
<tr>
<td>Phvul.009G190100</td>
<td>1003</td>
<td>919</td>
<td>153</td>
<td>444</td>
<td>84</td>
</tr>
<tr>
<td>BnMicEmUP1</td>
<td>709</td>
<td>588</td>
<td>146</td>
<td>442</td>
<td>121</td>
</tr>
<tr>
<td>BnMicEmUP2</td>
<td>803</td>
<td>594</td>
<td>152</td>
<td>442</td>
<td>209</td>
</tr>
<tr>
<td>BnMicEmUP3</td>
<td>863</td>
<td>594</td>
<td>152</td>
<td>442</td>
<td>269</td>
</tr>
<tr>
<td>BnMicEmUP4</td>
<td></td>
<td>588</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**3.7 Analysis of Elements in the Phvul.009G190100 Promoter**

To characterize specific elements within the Phvul.009G190100 promoter, a 1185 bp sequence upstream of the coding region was obtained from Phytozome *Phaseolus vulgaris* v1.0 (http://phytozome.jgi.doe.gov/) and analyzed with PlantCARE and PLACE (http://www.dna.affrc.go.jp/PLACE/) (Higo et al., 1999). The analysis identified 22 types of potential promoter sequences in this region. In particular, the analysis of the upstream sequence of Phvul.009G190100 identified core promoter elements, including CAAT-boxes and TATA-boxes, as well as a number of *cis*-regulatory elements that are associated with gene regulation by light, stress, hormone, fungal elicitors, and growth regulators. The functions of these predicted *cis*-acting elements are categorized in Table 3.2 and their arrangement in the upstream region is shown in Figure 3.17. The most frequent category of *cis*-acting regulatory elements involved light responsiveness and included 9 members namely, G-box which is a *cis*-acting regulatory element and other partial light responsive elements such as I-box, P-box, Gap-box (Table 3.2).
The Phvul.009G190100 promoter also contains elements that might be involved in responsiveness to physiological and environmental factors, including: fungal elicitors (ERE), heat shock (HSE), pathogen or abiotic stresses (TC-rich repeats), abscisic acid (ABRE), auxin (ARE), salicylic acid (TCA-element) and low temperatures (LTR). In addition, the promoter region contains a regulatory element named Skn-1-like motif required for endosperm expression and a TGACG site, involved in the methyl jasmonate (MeJA)-responsiveness.

Table 3. 2 cis regulatory elements found in upstream sequence (1000 bp) of Phvul.009G190100 with their functional annotations, according to their putative function assigned from Plant CARE and PLACE

<table>
<thead>
<tr>
<th>Name of the cis element</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Light element</strong></td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td>part of a conserved DNA module involved in light responsiveness</td>
</tr>
<tr>
<td>Box I</td>
<td>part of a light responsive element</td>
</tr>
<tr>
<td>Box II</td>
<td>part of a light responsive element</td>
</tr>
<tr>
<td>Box III</td>
<td>part of a light responsive element</td>
</tr>
<tr>
<td>G-box</td>
<td>cis-acting regulatory element involved in light responsiveness</td>
</tr>
<tr>
<td>GA-motif</td>
<td>part of a light responsive element</td>
</tr>
<tr>
<td>Gap-box</td>
<td>part of a light responsive element</td>
</tr>
<tr>
<td>I-box</td>
<td>part of a light responsive element</td>
</tr>
<tr>
<td>P-box</td>
<td>gibberellin-responsive element and part of a light responsive element</td>
</tr>
<tr>
<td><strong>Stress element</strong></td>
<td></td>
</tr>
<tr>
<td>ABRE</td>
<td>cis-acting element involved in the abscisic acid responsiveness</td>
</tr>
<tr>
<td>ARE</td>
<td>cis-acting element involved in salicylic acid responsiveness</td>
</tr>
<tr>
<td>ERE</td>
<td>ethylene-responsive element</td>
</tr>
<tr>
<td>TATC-box</td>
<td>cis-acting element involved in gibberellin-responsive</td>
</tr>
<tr>
<td>TCA-element</td>
<td>cis-acting element involved in salicylic acid responsiveness</td>
</tr>
<tr>
<td>CGTCA-motif</td>
<td>cis-acting regulatory element involved in the MeJA-responsiveness</td>
</tr>
<tr>
<td>TGACG-motif</td>
<td>cis-acting regulatory element involved in the MeJA-responsiveness</td>
</tr>
<tr>
<td>HSE</td>
<td>cis-acting element involved in heat stress responsiveness</td>
</tr>
<tr>
<td>LTR</td>
<td>cis-acting element involved in low-temperature responsiveness</td>
</tr>
<tr>
<td>TC-rich repeats</td>
<td>cis-acting element involved in defense and stress responsiveness</td>
</tr>
<tr>
<td><strong>Regulation element</strong></td>
<td></td>
</tr>
<tr>
<td>Skn-1-like motif</td>
<td>cis-acting regulatory element required for endosperm expression</td>
</tr>
<tr>
<td><strong>Core element</strong></td>
<td></td>
</tr>
<tr>
<td>CAAT-Box</td>
<td>common cis-acting element in promoter and enhancer regions</td>
</tr>
<tr>
<td>TATA-box</td>
<td>core promoter element around -30 of transcription start</td>
</tr>
</tbody>
</table>
Figure 3. 17 Cis regulatory elements found in upstream sequence (1000 bp) of Phvul.009G190100 derived from Plant CARE and PLACE, red boxes indicate the overlapped regions. ABRE: cis-acting element involved in abscisic acid responsiveness; ARE: cis-acting element involved in salicylic acid responsiveness; Box 4/Box I: light responsive element; Box III: protein binding site; ERE: ethylene-responsive element; G-box: cis-acting regulatory element involved in light responsiveness; GA-motif/Gap-box: light-responsive element; HSE: cis-acting element involved in heat stress responsiveness; I-box: light responsive element; LTR: cis-acting element involved in low-temperature responsiveness; P-box: gibberellin-responsive element; Skn-1 motif: cis-acting regulatory element required for endosperm expression; TC-rich repeats: cis-acting element involved in defence and stress responsiveness; TCCACCT-motif: TCA-element: cis-acting element involved in salicylic acid responsiveness; TGACT-motif: cis-acting regulatory element involved in the MeJA-responsiveness CAAT-box: common cis-acting element in promoter and enhancer regions; TATA-box: core promoter element around -30 of transcription start.
3.8 Protein Domain Characterization

The Phvul.009G190100 ORF encodes a protein of 198 amino acids with an estimated molecular mass of 20.88 kDa. Four amino acids are prominent in Phvul.009G190100 protein: Alanine (Ala, 14.4%), Leucine (Leu, 13.1%), Ser (Serine (Ser, 11.1%) and Valine (Val, 11.1%). The calculated pI (Isoelectric point) is 9.44 (SDSC Workbench Biology). A comparison of the proteins encoded by Phvul.009G190100 and AT1G74730 showed that they are 64% identical (Figure 3.18). A Domain of Unknown Function 1118 (DUF1118), which was identified in the Phvul.009G190100 protein sequence, is also present in the proteins encoded by AT1G74730 and BnMicEmUP (Shahmir, 2014) (Figure 3.18).

![Figure 3.18 Comparison of P. vulgaris Phvul.009G190100 putative protein sequence with B. napus BnMicEmUP3 and Arabidopsis AT1G74730 protein sequences. The DUF1118 (Domain unknown function) region is shown in the large red box. Putative chloroplast transit peptides are highlighted in black and putative motifs (A/A) for chloroplast transit peptide cleavage sites are shaded in green. Basic DNA binding and leucine zipper (Dimerization domain) are shaded in yellow and red, respectively. Potential casein kinase II phosphorylation sites (S/TxxD/E) are shaded in blue. Putative transmembrane helix topologies are shaded in grey.](image)

64
3.8.1 Chloroplast Transit Peptide

Further searches for motifs in the putative Phvul.009G190100 protein sequence using the domain alignments in protein prediction programs including PFAM (Bateman et al., 2002), SMART (Letunic et al., 2002), PRINTS (Attwood et al., 2002) and PROSITE showed that Phvul.009G190100 contains a putative cTP (chloroplast transit peptide) in its N-terminus (Table 3.3).

Figure 3.19 shows that the putative Phvul.009G190100 cTP contains three regions of homology that are well conserved in transit peptides sequences among diverse proteins targeted to the chloroplast in Brassica and Arabidopsis species. The three homology blocks are separated by different lengths of non-conserved sequences. All the sequences start with methionine and alanine (MA), followed by hydrophobic amino acids, like alanine or valine (A or V). A pair of serines (S) was found to be highly conserved in the middle part and a proline (P) was found as a terminator (Zybailov, 2008). A cleavage motif, consisting of two alanines (A/A) (Zhang and Glaser, 2002) was identified at the N-terminal sequences of the AT1G74730, BnMicEmUP2 and BnMicEmUP3 putative protein sequences (Figure 3.19). However, Phvul.009G190100 has an A/S instead of A/A for the cleavage site of the cTPs.
Table 3. Predicted localization of Phvul.009G190100 protein

<table>
<thead>
<tr>
<th>Program description</th>
<th>Possible compartments</th>
<th>Prediction</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSORT</td>
<td>Chloroplast thylakoid membrane: 0.948</td>
<td>Chloroplast thylakoid membrane</td>
<td>Plants</td>
</tr>
<tr>
<td></td>
<td>Mitochondrial inner membrane: 0.884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ChloroP</td>
<td>cTP:0.558</td>
<td>Chloroplast thylakoid membrane</td>
<td>Plants</td>
</tr>
<tr>
<td>TargetP</td>
<td>cTP:0.854, mTP:0.408</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
<tr>
<td>ESL Pred</td>
<td>Mitochondria: 1.18</td>
<td>Mitochondria</td>
<td>Plants</td>
</tr>
<tr>
<td>Plant-PLoc</td>
<td>Chloroplast: 0.33</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
<tr>
<td>BaCelLo</td>
<td>Chloroplast</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
</tbody>
</table>

Figure 3.19 Alignment of amino acid sequences of the predicted transit peptides Phvul.009G19010, AT1G74730, BnMicEmUP3, Glyma04g34330 and Glyma06g20240 with the chloroplast targeted proteins in other species. Species represented and the genes used to generate the sequence alignments are as follows: B. napus (BRANA), B. rapa subsp. campestris (BRACM), A. thaliana (ARATH), FER1 (Ferritin-1 Q96540.1), PSBR (Photosystem II 10 kDa polypeptide P49108.1), LEU3 (isopropylmalate dehydrogenase P29102.1), FABI (NADH-dependent enoyl-ACP reductase P80030.2), ILVB1 (Acetolactate synthase 1 AAA62705.1), GRS16 (Monothiol glutaredoxin-S16 AEC09515.1), F16P1 (Fructose-1,6-bisphosphatase
AAA82750.1), TRXM (Thioredoxin M-type Q9XGS0.1), RBCS (Rubisco small subunit), PAP1(Plastid lipid-associated protein 1), APR2 (5’-adenylylsulfate reductase 2). Conserved residues are highlighted in yellow (block 1), grey (block 2), green (block 3), blue (block 4) and red (block 5). The consensus non-shared residues are colored according to: red, hydrophobic; bold blue, basic; black, hydrophilic; bold yellow; acidic. (Adapted from Shahmir PhD thesis 2014)

3.8.2 bZip Region

A basic DNA binding region was found in the putative protein of Phvul.009G190100 when it was aligned to typical bZIP amino acid sequences in the plant transcription database (planttfdb.cbi.edu.cn) (Figure 3.20A). Highly conserved sequences that are characteristic features of bZIP transcription factors were identified including a basic DNA binding domain and a zipper region. In particular, the DNA-binding domain found in the putative Phvul.009G190100 protein is most similar to those found in the basic leucine zipper (bZIP) family (Figure 3.20A). Moreover, the alignment of the putative Phvul.009G190100, AT1G74730 and BnMicEmUP3 proteins shows that they are also highly conserved for the bZip region (Figure 3.18). They especially share two of the essential parts of the bZIP structure, namely: a N-X\textsubscript{7}-R motif of the basic DNA-binding region and a heptad repeat of 7 leucines or other hydrophobic amino acids such as isoleucine (I), alanine and valine (V), or tryptophan (W).
Figure 3. 20 Structural features of Phvul.009G190100 compared to plastid envelope DNA binding PEND proteins. A) Alignment of the Phvul.009G190100 bZIP domain with selected bZIP domains of PEND proteins. Highlighted residues show identical basic region (yellow) or similar zipper region (red) amino acid residues in Phvul.009G190100 bZIP domains and bZIP domains of PEND proteins. Sources of sequences: *P. sat* (*Pisum sativa* PEND, cDNA: X98740, genome: AB189736); *B. nap* (*B. napus* GSBF1 cDNA: X91138); *B. nap*1 (genome: AB189734); *B. nap*2 (genome: AB189735); *A. tha* (*A. thaliana* EST: BX825084, genome: AL094711); *G. max* (*G. max* EST from somatic embryos: BM308592; EST from flowers: BE807880); *M. tru*
(Medicago truncatula ESTs from seedling roots: BG644842, AJ500077); L. esc (Lycopersicon esculentum ESTs from root: BE450861, AW218782); H. vul (Hordeum vulgare overlapping ESTs from germinating shoots: AV833513, BJ463210); O. sat (Oryza sativa genome: CDS matching an EST AK106548: P0010C01.1 in AP004768); C. sat (Cucumis sativus genome: AB189737); P. yed (Prunus yedoensis genome: AB189738). B) Schematic diagram of the Phvul.009G190100 protein showing the locations of the chloroplast transit peptide (dark blue), the basic DNA binding (black), zipper region (orange), predicted transmembrane domains (yellow) and casein kinase II phosphorylation sites (blue). C) Schematic diagram of the PEND protein includes a chloroplast transit peptide (dark blue), a basic DNA binding region (black), a zipper region (orange), predicted transmembrane domains (yellow) and casein kinase II phosphorylation sites (blue). (Adapted from Shahmir PhD thesis 2014)

3.8.3 Transmembrane Domain

Two transmembrane helix regions are detected by the TMHMM v2.0 program (SMART: http://smart.embl-heidelberg.de/smart/set_mode.cgi?GENOMIC=1) in the putative Phvul.009G190100 protein sequence. The first transmembrane domain contains 20 amino acids “LLASAALPAFVAAIAIVL”, starting at position 140 and ends at position 159 and the second region consist of 23 amino acids “VQVVVAAALGVGAVGLFVGSVVL” and starts at position 169 and ends at position 191 (Figure 3.21). The Phvul.009G190100, AT1G74730 and BnMicEmUP proteins are very similar in their two transmembrane regions (Figure 3.18).

3.8.4 Phosphorylation Sites (PS)

A protein analysis by NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) of the amino acid sequence for Phvul.009G190100 showed that it contains several sequences that correspond to casein kinase II phosphorylation sites consensus sequences (S/TXXD/E), in the 89 to 129 amino acid region (Figure 3.18). The same analysis of AT1G74730 and BnMicEmUP3 showed that they contain the same casein kinase II phosphorylation sites (Figure 3.18).
3.8.5 Protein Secondary Structure

The secondary structure of the protein predicted by phyre2 (Figure 3.21) (Kelley and Sternberg, 2009) showed that the putative Phvul.009G190100 protein has a characteristic α-helical structure in the region that contains the leucine-rich repeats (Figure 3.21), which is the main structural motif found in bZIP proteins (Jakoby et al., 2002). In addition, two putative transmembrane domains are also modeled by phyre2. The three-dimensional structure of the Phvul.009G190100 protein predicted by phyre2 (Figure 3.22), contains an extended C terminal region with some helical structure and an N terminal with two regions of strongly folded (alpha helix or beta strand) amino acids that are associated with the transmembrane regions.

Therefore, the essential features of the Phvul.009G190100 protein (DUF1118) include a chloroplast transit peptide region, a basic DNA binding region and a repeat region containing 11 complete leucine-rich repeats and two transmembrane domains. These features have been observed in members of a bZIP family encoding plastid envelope DNA binding proteins (PENDs). This family of genes encodes a DNA binding protein found in the inner envelope membrane of developing chloroplast (Terasawa and Sato, 2009). A multi alignment was performed between Glyma04g34330, Glyma06g20240, Phvul.009G190100, BnMicEmUP, AT1G74730 and PEND proteins of different species of angiosperms including: *Pisum sativa*, *B. napus*, *A. thaliana*, *G. max*, *Medicago truncatula*, *Lycopersicon esculentum*, *Hordeum vulgare*, *Oryza sativa*, *Cucumis sativus* and *Prunus yedoensis* (Figure 3.20a). The result reveals that the basic region, zipper region and transmembrane domains are conserved, but the central region is highly variable. It also shows that Phvul.009G190100 can be considered as a possible gene homolog of a PEND protein gene because it encodes a protein with a possible chloroplast transit
peptide, a DNA binding site, a leucine rich repeat region and transmembrane domains (Figure 3.20A,B,C).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Secondary structure</th>
<th>Disorder confidence</th>
<th>Disorder confidence</th>
<th>Disorder confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VFAHAPYKKNVKYDEKWK</td>
<td>VFAHAPYKKNVKYDEKWK</td>
<td>? ? ?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.21 The predicted secondary α-helix and β-strand structure of Phvul.009G190100. The basic DNA binding area is shown in a grey box and the leucine zipper region is highlighted by black boxes. Potential casein kinase II phosphorylation sites (PS) are indicated in blue boxes (S/TxxD/E). Putative transmembrane helixes are shown in red boxes.
3.9 Real-time PCR Results

3.9.1 Standard Curve Construction

A serial dilution of linearized plasmid DNA was used to generate a standard curve for qPCR. Knowing the size of the plasmid (3930 bp) that contained the gene of interest (587 bp), the molecular weight (grams/molecule) of the plasmid was calculated as follows:

\[
\text{Molecular weight for a plasmid (g/molecule) =}
\]
\[
\frac{(\text{Plasmid bp size} + \text{insert bp size}) + (330 \text{ Da} \times 2 \text{ nucleotide/bp})}{\text{Avogadro's number } 6.02E+23}
\]

Knowing the molecular weight of the plasmid (4.91E-18 g/molecule) and the concentration of the plasmid (2.00E-07 g/ul) that was added to each PCR reaction, the precise number of molecules (copy number) in each reaction was determined as follows:
Having calculated the numbers of molecules per microliter of linearized plasmid solution (4.07E+10 molecules/ul), a series of dilutions was made for subsequent amplification to generate a standard curve. The stock solution was diluted to generate series of solutions containing 4.07E+7, 4.07E+6, 4.07E+5, 4.07E+4, 4.07E+3, 4.07E+2 and 4.07E+1 (molecules/ul). For the purpose of a clear display, the copy numbers of the plasmid DNA was converted to the Log values# (log# molecules/ul): 7.61, 6.61, 5.61, 4.61, 3.61, 2.61 and 1.61. Relative standard curves were constructed by plotting the Log copy number (Log#) of each diluted sample with the correspondent Ct (cycle threshold which is the number of cycles required for the fluorescent signal to cross the threshold) value.

A real-time PCR standard curve was graphically represented as a semi-log regression line plot of CT value versus log of input nucleic acid. PCR amplification efficiency was the rate at which a PCR amplicon was generated commonly estimated with the slope of a standard curve as follows:

\[
\text{Efficiency} = \left(10^{-1/-\text{slope}} - 1\right) \times 100
\]

If a particular PCR amplicon doubles in quantity during the geometric phase of its PCR amplification then the PCR assay had 100% efficiency. A standard curve slope of –3.32 indicates a PCR reaction with 100% efficiency which yielded a 10-fold increase in PCR amplicon every
3.32 cycles during the exponential phase of amplification ($\log_{10} = 3.3219$). The amplification efficiencies of the target gene and the endogenous control (SKIP16) were 109.18 and 106.76, respectively, which were non-significant being greater than 100%.

### 3.9.2 Validation Experiment

For a valid comparative ($\Delta\Delta Ct$) calculation, the efficiency of the target DNA fragment amplification and the efficiency of the reference amplification must be approximately equal. To determine if the two amplification reactions had the same PCR efficiencies, the $\Delta Ct$ ($Ct$ target – $Ct$ reference) variation of a series of template dilutions was assessed by running standard curves for both the target gene and the endogenous reference gene (SKIP16). The $Ct$ values generated from equivalent standard curve mass points (target VS reference) were used in the $\Delta Ct$ calculation ($Ct$ target – $Ct$ reference). These $\Delta Ct$ values were plotted VS log copy number (molecules/ul) to create a semi-log regression line. The slope of the resulting semi-log regression line was equal to 0 ($P = 0.7123$), which indicated equivalent amplification efficiencies for both the target gene and reference gene.
Figure 3. 23 Real time PCR standard curves for the target gene and reference gene and their relative amplification efficiencies A) standard curve for the target gene Phvul.009G190100; B) standard curve for the internal control gene SKIP16; C) the ΔCT value of Phvul.009G190100 and the reference gene SKIP16. Standard errors are provided which represent three PCR replicates of one biological trial.
3.10 Gene Expression of Phvul.009G190100

The RNA-seq data obtained from the common bean Phytozome (http://phytozome.jgi.doe.gov/) database showed an expression distribution of Phvul.009G190100 in a number of common bean tissues of an inbred landrace G19833 (Figure 3.24). It was found that the highest expression occurred in leaves, which had approximately 700 FPKM (fragments per kilo base per millions of the mapped reads) and has almost no expression occurred in nodules and roots (Figure 3.24). It provided evidence that, for Phvul.009G190100 expression analysis, young leaves were the best part of the bean plant to sample.

![Figure 3.24](http://www.phytozome.net/) Expression of Phvul.009G190100 in different tissues of a common bean inbred landrace line G19833. FPKM stands for fragments per kilo base per millions of mapped reads. Expression data available from Phytozone v9.1 (http://www.phytozome.net/)
Real time PCR results were expressed as Ct (cycle threshold) values. This value corresponded to the cycle at which the fluorescence of the SYBR Green dye reached above the threshold or background fluorescence value. To quantifying gene expression, the mRNA levels of the gene of interest were compared to a reference sample (Sanilac was designated as the calibrator).

Table 3.4 presents data from an experiment where the expression levels of a target (PV09) and an endogenous control (SKIP16) were evaluated. The levels of these amplicons in a series samples were compared to the lowest expression sample (Sanilac). A validation experiment previously demonstrated that the efficiency of the target amplification and the efficiency of the reference (endogenous control gene) amplification were approximately equal. Therefore, the ΔΔCT method can be used to calculate the fold-differences in PV09 between samples. Since each sample ran with triplicated qPCR reactions, least square Means of Ct values were used in the ΔΔCT calculations. The resulting Ct levels were statistically significant different among the twelve varieties (Table 3.4). When converting the ΔCT to a comparative CT value, relative to Sanilac (calibrator), the fold numbers were calculated as shown in the Table 3.4. There was a significant difference of the expression fold numbers among the varieties (P < 0.0001, Table A-6). Midland showed the highest gene expression which was 2.47 fold more than Sanilac. This conclusion was supported by a visual examination of the levels of amplicons in the PCR reactions by agarose gel electrophoresis (Figure 3.25), since it had the brightest band of all the samples and Sanilac, which had the lowest Ct value and a very faint band in the gel.
Table 3. 4 Normalized fold change expression of PV09, calculated by $\Delta\Delta C_t$

<table>
<thead>
<tr>
<th>Number</th>
<th>Varieties</th>
<th>$\Delta C_t=C_t$ target-$C_t$ skip16</th>
<th>$\Delta\Delta C_t=\Delta C_t$ sample-$\Delta C_t$ Sanilac</th>
<th>Fold differences in PV09 relative to Sanilac</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Apex</td>
<td>-3.363</td>
<td>0.791</td>
<td>1.730</td>
</tr>
<tr>
<td>2</td>
<td>Midland</td>
<td>-2.850</td>
<td>1.304</td>
<td>2.469</td>
</tr>
<tr>
<td>3</td>
<td>Hensal</td>
<td>-2.981</td>
<td>1.173</td>
<td>2.255</td>
</tr>
<tr>
<td>4</td>
<td>Sanilac</td>
<td>-4.158</td>
<td>-0.003</td>
<td>0.998</td>
</tr>
<tr>
<td>5</td>
<td>HR200</td>
<td>-3.538</td>
<td>0.617</td>
<td>1.533</td>
</tr>
<tr>
<td>6</td>
<td>Michelite</td>
<td>-3.552</td>
<td>0.602</td>
<td>1.518</td>
</tr>
<tr>
<td>7</td>
<td>Rexeter</td>
<td>-3.288</td>
<td>0.867</td>
<td>1.823</td>
</tr>
<tr>
<td>8</td>
<td>Lightning</td>
<td>-3.239</td>
<td>0.915</td>
<td>1.885</td>
</tr>
<tr>
<td>9</td>
<td>OAC Thunder</td>
<td>-3.165</td>
<td>0.989</td>
<td>1.985</td>
</tr>
<tr>
<td>10</td>
<td>AC Compass</td>
<td>-3.067</td>
<td>1.087</td>
<td>2.125</td>
</tr>
<tr>
<td>11</td>
<td>Vista</td>
<td>-3.809</td>
<td>0.345</td>
<td>1.270</td>
</tr>
<tr>
<td>12</td>
<td>Saginaw</td>
<td>-3.982</td>
<td>0.172</td>
<td>1.127</td>
</tr>
</tbody>
</table>

Apex  Midland  Hensal  Sanilac  HR200  Michelite  Rexeter  Lightning  OAC Thunder  AC Compass  Vista  Saginaw

![Image of gel electrophoresis with band intensities](image)

![Bar graph showing fold numbers relative to Sanilac for various varieties](graph)
Figure 3. 25 Transcript profile of Phvul.009G190100 in the leaves of twelve field grown common bean varieties. The transcript of the gene from each variety with 587 bp was indicated by the arrow. Vertical bar represents the average fold numbers of 3 real-time PCR replicates from 4 biological replications from the Elora research station in 2013. Error bars show standard errors among four biological replicates.

3.11 Correlation between Gene Expression Levels and Yield Data

The comparison of the historical and 2013 yield profiles revealed that Sanilac and Saginaw showed a different yield pattern than the other ten genotypes (Figure 3.5). The possible reason of the difference is that they are the only determinate varieties evaluated in the field study. Therefore, in the comparisons between Phvul.009G190100 expression and the yield data, it was decided to consider only in determinate lines (thus excluding Sanilac and Saginaw) in order to remove the growth habit effect on the gene expression. However, since Sanilac was used as a calibrator variety, because it had the lowest absolute expression level, when the expression levels of all varieties were determined (Table 3.4), it was placed in the correlation graphs as a reference. It has been calibrated to 1 and all other varieties were presented as fold numbers relative to it (See Materials and Methods).

In addition to measuring the total yield (kg/ha), which was the mean seed weight adjusted to 18% moisture for each variety, the average yield gain was also considered to be an important parameter for evaluating crop productivity. Average yield gain (kg/ha/day) describes the seed weight accumulated per growing day from a unit area for each variety and is calculated by the formula:
3.11.1 Correlation between Gene Expression Levels and Historical Data

The trend in the relationship between expression of the Phvul.009G190100 gene in leaf tissues collected from indeterminate plants grown at Elora in 2013 and the yield measurements made for the varieties over a number of years (2010, 2011, 2012) in different locations (St. Thomas, Woodstock, and Thorndale, ON), was negative, with an $r = -0.0233$ and an $R^2 = 0.21$ but it was not statistically significant ($P = 0.189$; Figure 3.26). A negative trend ($r = -0.0233$) was also found in the relationship between average yield gain (kg/ha/day) against the expression levels ($R^2 = 0.14$) but it was not statistically significant ($P = 0.29$) (Figure 3.27). The time to maturity (days) showed a slightly stronger correlation ($r = -0.0631$ and $R^2 = 0.38$) with the expression levels of Phvul.009G190100 in the indeterminate genotypes (Figure 3.28). This trend was statistically significant at the 10% level but not the 5% level ($P = 0.0571$).
Figure 3. Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and yield (kg/ha) in the historical trial. The historical trial was conducted at St. Thomas, ON in 2010 and in St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown, n = 3 for the gene expression data and n = 4 for the yield data. Full names of the shown varieties are: Lightning, Apex, Rexeter, HR200, Midland, Sanilac, Hensal, Saginaw, AC Compass, OAC Thunder, Vista and Michelite.

\[ y = -0.0003x + 2.7125 \]
\[ R^2 = 0.21 \quad P = 0.1888 \]
Figure 3. 27 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and average gained yield (kg/ha/day) in the historical trial. The historical trial was conducted at St. Thomas, ON in 2010 and in St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi et al., 2012). Means ± standard errors are shown. n = 3 for the gene expression data and n = 4 for the yield data. Full names of the shown varieties are: Lightning, Apex, Rexeter, HR200, Midland, Sanilac, Hensal, Saginaw, AC Compass, OAC Thunder, Vista and Michelite.
3.11.2 Correlation between Gene Expression Levels with 2013 Yield Data

The trend in the relationship between expression of the Phvul.009G190100 gene in leaf tissues collected from indeterminate plants grown at Elora in 2013 and the yield measurements made for the varieties at Elora in 2013, was negative, with an $r = -0.0004$ and an $R^2 = 0.19$ but it was not statistically significant ($P = 0.2046$; Figure 3.29). A negative trend ($r = -0.021$) was also found in the relationship between average yield gain (kg/ha/day) against the expression levels ($R^2 = 0.05$) but it was not statistically significant ($P = 0.5173$) (Figure 3.30). The time to maturity
(days) also showed a statistically non-significant ($P = 0.2951$) negative trend ($r = -0.026$ and $R^2 = 0.14$) with the expression levels of Phvul.009G190100 in the indeterminate genotypes (Figure 3.31).

Figure 3.29 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and yield (kg/ha) in 2013 yield trial. 10 indeterminate bean varieties were determined at Elora, ON in 2013. Means ± standard errors are shown. $n = 3$ for the gene expression data and $n = 4$ for the yield data. Full names of the shown varieties are: Lightning, Apex, Rexeter, HR200, Midland, Sanilac, Hensal, Saginaw, AC Compass, OAC Thunder, Vista and Michelite.
Figure 3. 30 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and average gained yield (kg/ha/day) in 2013 trial. 10 indeterminate bean varieties were determined at Elora, ON in 2013. Means ± standard errors are shown. n = 3 for the gene expression data and n = 4 for the yield data. Full names of the shown varieties are: Lightning, Apex, Rexeter, HR200, Midland, Sanilac, Hensal, Saginaw, AC Compass, OAC Thunder, Vista and Michelite.
3.11 Correlation between relative expression levels (to Sanilac) of Phvul.009G190100 and maturity time (days) in 2013 trial. 10 indeterminate bean varieties were determined at Elora, ON in 2013. Means ± standard errors are shown. n = 3 for the gene expression data and n = 4 for the yield data. Full names of the shown varieties are: Lightning, Apex, Rexeter, HR200, Midland, Sanilac, Hensal, Saginaw, AC Compass, OAC Thunder, Vista and Michelite.

3.12 Homolog of Phvul.009G190100 Localization in Soybean

3.12.1 Gene Structure

Two orthologs of the Phvul.009G190100 were identified in soybean with a BLAST search on soybean chromosome 4 from 40,445,359 bp to 40,447,941 bp and on chromosome 6 from 16,663,873 bp to 16,665,393 bp annotated as Glyma04g34330 and Glyma06g20240, respectively (Glycine max v1.1 assembly, Phytozome). Gene models of soybean in Phytozome
were predicted by homology-based predictors, FGENESH+ (Salamov, 2000), FGENESH_EST, and GenomeScan (Yeh, 2001) with transcript assemblies including RNA-seq transcript assemblies above, as well as Sanger and Roche/454 ESTs (Schmutz J, et al., 2010).

Glyma04g34330 and Glyma06g20240 showed an overall similarity of 46% and 50% with Phvul.009G190100 at the nucleotide acid level, respectively (Figure 3.32). The Glyma04g34330 and Glyma06g20240 share 79% identity of the nucleotide acid. The Intron/exon structures and genomic sizes of the 3 sequences from common bean and soybean were similar that two exons and one intron were present (Figure 3.33). In addition, all three genes have the similar sized first exon region (144-153 bp) and the second exon region (444 bp) but the intron region varied to a greater extent (Figure 3.22).
Figure 3. 32 Comparison of the nucleotide sequences for Phvul.009G190100, Glyma04g34330 and Glyma06g20240; yellow highlighted regions identify the homologies in the introns; blue highlighted regions identify the homologies in the exons; un-highlighted regions identify the 5` un-translated and 3` un-translated regions.

![Gene Structure Comparison](image)

Figure 3. 33 Gene structure comparisons of *P. vulgaris* Phvul.009G190100 with *G. max* Glyma04g3433 and Glyma06g20240; blue boxes show the exons; green boxes show the 5' and 3' un-translated regions; introns are shown with black lines.

### 3.12.2 Protein Characterization

The proteins encoded by the soybean genes (Glyma04g34330 and Glyma06g20240) shared a domain of unknown function with the protein encoded by the bean gene (DUF1118), and were 87% and 85% similar to Phvul.009G190100 protein at the amino acid level, respectively. Similar to the Phvul.009G190100 encoded protein, the soybean homologs had several motifs, including the: chloroplast transit peptide region, the DNA binding region,
phosphorylation sites and a leucine zipper region, that are predicted from amino acid alignments with the putative Phvul.009G190100 protein (Figure 3.34). Searches for motifs in the putative Glyma04g34330 and Glyma06g20240 protein sequences using the domain alignments in protein prediction programs including PFAM (Bateman et al., 2002), SMART (Letunic et al., 2002), PRINTS (Attwood et al., 2002) and PROSITE showed that they were localized to chloroplasts. Therefore, they were predicted to contain a cTP (chloroplast transit peptides) in their N-termini (Table 3.5 and 3.6). In fact, the sequences all share the same sequence (MAVPATSSAP/SVL) for a putative chloroplast transit peptide region. For this sequence the putative bean protein has a serine (S) in the third last position instead of the proline (P) found in the soybean sequence.

For the basic DNA binding region, all of the putative proteins have an asparagine (N) and end with an arginine (R) with a motif N-X7-R. The leucine zipper region that follows shows a high level of conservation for a repeat containing of 7 Leucine (L), or other hydrophobic residues such as, isoleucine (I), Trypotophan (W) and valine (V). Although there are slight differences in the nucleotides of the phosphorylation sites, the four PS (phosphorylation sites) regions SxxE in the putative proteins for the two species are identical. At the N terminals, all the putative proteins have two transmembrane domains that have identical sequences in bean and soybean.
Figure 3. 34 Comparison of putative protein sequences encoded by Phvul.009G190100 and the G. max Glyma04g34330 and Glyma06g20240 sequences. The DUF1118 (Domain unknown function 1118) region is shown with a large red box. Putative chloroplast transit peptides and basic DNA binding regions are shaded in black and yellow, respectively. The conserved hydrophobic amino acids in the leucine zipper domain are highlighted in red and the potential casein kinase II phosphorylation sites (S/TxxD/E) are shown by blue highlighting. Putative transmembrane helix topologies are highlighted in grey.

Table 3. 5 Predicted localization of Glyma04g34330 protein

<table>
<thead>
<tr>
<th>Program description</th>
<th>Possible compartments</th>
<th>Prediction</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSORT</td>
<td>chloroplast thylakoid membrane:0.948</td>
<td>chloroplast thylakoid membrane</td>
<td>Plants</td>
</tr>
<tr>
<td>ChlороP</td>
<td>cTP:0.548</td>
<td>chloroplast thylakoid membrane</td>
<td>Plants</td>
</tr>
<tr>
<td>TargetP</td>
<td>cTP:0.766</td>
<td>mTP:0.515</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>ESL Pred</td>
<td>Mitochondria:1.30</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
<tr>
<td>Plant-PLoc</td>
<td>Chloroplast</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
<tr>
<td>BaCellLo</td>
<td>Chloroplast</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
</tbody>
</table>
Table 3.6 Predicted localization of Glyma06g20240 protein

<table>
<thead>
<tr>
<th>Program description</th>
<th>Possible compartments</th>
<th>Prediction</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSORT</td>
<td>Mitochondrial inner membrane:0.791</td>
<td>Mitochondrial inner membrane</td>
<td>Plants</td>
</tr>
<tr>
<td>ChloroP</td>
<td>cTP:0.565</td>
<td>Chloroplast thylakoid membrane</td>
<td>Plants</td>
</tr>
<tr>
<td>TargetP</td>
<td>cTP:0.830 mTP:0.335</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
<tr>
<td>ESL Pred</td>
<td>Mitochondria:1.07</td>
<td>Mitochondria</td>
<td>Plants</td>
</tr>
<tr>
<td>Plant-PLoc</td>
<td>Chloroplast</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
<tr>
<td>BaCelLo</td>
<td>Chloroplast</td>
<td>Chloroplast</td>
<td>Plants</td>
</tr>
</tbody>
</table>

3.12.3 Expression of Glyma04g34330 and Glyma06g20240 in soybean

Data for the expression of Glyma04g34330 and Glyma06g20240 in different tissues of soybean cultivar Williams 82 was obtained from soybean database (http://www.soybase.org/) (Figure 3.35). Similar to the bean gene, it was found that the young leaves had the highest levels of expression (with 60-70 reads per kilo base per millions of the mapped reads) for both Glyma04g34330 and Glyma06g20240 among all the sampled tissues (Figure 3.25). They had their lowest levels of expression in root and nodule tissues. In addition, the relative expression (# FPKM of each tissue/total # FPKM * 100%) levels of these two genes were highest in young leaves and accounted for 20-25% in all the sampled tissues for both the bean and soybean (Figure 3.36).
Figure 3. 35 Expression of G. max Glyma04g34330 and Glyma06g20240 in different tissues of the soybean. RPKM stands for reads per kilo base per millions of mapped reads. Expression data available from soybean database (http://www.soybase.org/).

Figure 3. 36 Expression proportion of Glyma04g34330 and Glyma06g20240 in different tissues of the soybean and Phvul.009G190100 expression proportion in bean. The X axis represents the proportion of FKPM in each tissue as a proportion of all the sampled tissues.
3.13 Phylogenetic Analyses

The genomic similarities of Phvul.009G19010 and homologs from *Arabidopsis*, *B. napus* and *G. max* are shown in Figure 3.37 (CLC genomic workbench), which shows two distinct groups. One group consists only of a *B. napus* gene (BnMicEmUp) and the other group contains AT1G74730, Phvul.009G19010, Glyma04g34330 and Glyma06g20240 (Figure 3.37A). The relationships among the putative proteins encoded by these genes are shown in the unrooted tree in Figure 3.37B. The tree contains 2 distinct clusters with the BnMicEmUps and AT1G74730 in one group and Phvul.009G19010, Glyma04g34330 and Glyma06g20240 in a separate cluster. Similar relationships were found on the basis of the predicted secondary protein structures for the putative proteins encoded by AT1G74730, Phvul.009G19010, BnMicEmUps, Glyma04g34330 and Glyma06g20240 (Figure 3.38).
Figure 3. Relationships among *Arabidopsis* AT1G74730, *P. vulgaris* Phvul.009G190100, *B. napus* BnMicEmUPs, and *G. max* Glyma06g20240 and Glyma04g34330 genes and their protein sequences they encode. The phylogenetic trees represent the relationships between A) genomic DNA B) predicted coding proteins. Phylogenetic trees were constructed by the neighbor-joining method with 1,000 bootstrap replicates using CLC software.
3.14 Synteny Analysis between Common Bean and Soybean

A map comparing bean and soybean genomes was developed *in silico* by aligning sections of soybean chromosomes (GM04 and GM06) related to common bean chromosome 9 (PV09) using the sequence positions of a number of common bean molecular markers (g
markers) that are shared between the two species (Figure 3.39). The genetic markers used to construct Figure 3.39 included phenylpropanoid pathway genes (GenBank); common bean seed coat color gene-based markers (McClean et al., 2002/LIS, Legume Information System http://www.comparative-legumes.org/); soybean seed coat color gene-based markers (Yang et al., 2010/GenBank); common bean “g” markers (LIS); common bean SSRs (LIS and/or PhaseolusGenes database (http://phaseolusgenes.biinformatics.ucdavis.edu/); and soybean SSRs (SoyBase, http://soybase.org/). The correspondence between common bean and soybean chromosomes agrees to a large extent with previously reported synteny patterns between the two species (McClean et al., 2010; Galeanoetal et al., 2011). The map identified gene markers F3H-CV, SN317-F, and g126 in positions close to Phvul.009G19010 on common bean chromosome PV09. The markers were used to identify corresponding syntenic regions in the soybean genome and confirmed that they were close to the related soybean homologs Glyma04g34330 (at 40,445,359 nt on chromosome 4) and Glyma06g20240 (at 16,663,873 nt on chromosome 6). The analysis was used to identify additional markers in soybean flanking the genes, including: markers Satt294 and Satt476 associated with Glyma04g34330; and Satt277 and BARC-025705-05001, associated with Glyma06g20240.
Figure 3. 39 Sequence-based common bean-soybean comparative map. Gene markers (abbreviated plus accession number) are shown on right of each chromosome. Feature position [bp, start nucleotide shown (Phytozome)] is indicated on left of each chromosome. Syntenic fragments of soybean chromosomes (each in different color) are shown on left (Chromosome 4, shaded in yellow) and right (Chromosome 6, shaded in green) of common bean chromosome 9 (in the center, black). Gene Phvul.009G1901 and its two homologous genes in soybean (Glyma04g34330, Glyma06g20240) are shown in red, bold, and underlined. SSR markers associated with the soybean genes are shown in bold blue and underlined.

3.15 Yield Related QTLs Associated with Soybean Genes

Using the markers associated with Glyma04g34330 on GM04 (Satt294 located at 78.65 cM and Satt476 located at 80.62 cM) and SNP marker BARC-025705-05001 and a SSR marker Satt277 associated with Glyma06g20240, a number of yield related QTLs were identified (i.e. seed yield 12-2, seed weight 26-15, seed yield 21-1 seed yield 19-2 et al.) and placed on the linkage map (Figure 3.40).
Figure 3. 40 Soybean yield related QTLs occurring in genomic regions of the soybean composite linkage map containing Glyma04g34330 and Glyma06g20240. Feature positions [cM, start point shown (Soybase)] are indicated on left side of each linkage map (cM). Molecular markers associated with Glyma04g34330 and Glyma06g20240 are highlighted in red, bold and underlined; QTLs related to markers are indicated by blue arrows; linkage map was constructed by Mapchart based on the data adapted from Soybase (http://www.soybase.org/). A) Molecular markers associated with Glyma04g34330 and yield QTLs related to the markers. B) Molecular markers and QTLs associated with Glyma06g2024.
CHAPTER 4 DISCUSSION

4.1 Yield Measurements in Bean

The current measurements of yield in beans support many studies that show that this trait is highly variable and fluctuates from year to year and among plots established in different locations (Hala et al., 2011). But the correlation between the yields in the current study with the yields for the same varieties in the previous “Historical Field Trials” in multiple years and locations (Navabi et al., 2012) indicates that, although the genotype × environment is often significant, the overall trends and ranks of the lines are relatively stable. In this study, a major confounding factor was the growth habit since the yields for the only two determinate-type varieties (Sanilac and Saginaw) were very much lower in 2013 than in the Historical Trials.

4.2 Isolation and Characterization of a Bean Homolog to BnMicEmUp/AT1G74730

Using a PCR cloning technique, the current study identified and characterized a bean gene from \textit{P. vulgaris} variety “Lightning”, annotated as Phvul.009G190100, with high homology to BnMicEmUp and AT1G74730. The homology to the \textit{B. napus} and \textit{Arabidopsis} genes was extensive and occurred at the genome structure level, gene expression level, putative protein sequence level and putative protein functional motif level. In particular, Phvul.009G190100 is characterized as a member of cbZIP transcription factor family which was the DNA-binding domain of the PEND polypeptide because; first, they contain a basic region at the N-terminus that could promote the transport of the proteins into plastids. Secondly, they both have a bZIP-like transcription factor motif.
4.3 Promoter Analysis

The multiple TATA (28) and CAAT (16) box sequences, which are defined as core promoter elements important for accurate initiation of transcription (Kozak, 1981), found upstream of the transcription start area in the Phvul.009G190100 gene, may function in promoting high levels of transcription of this gene and may also differentially regulate gene activity in response to different environmental conditions (Doyle and Han, 2001). In particular, the stress responsiveness of the Phvul.009G190100 gene is supported by the occurrence of different types of stress-related elements in the promoter region of this gene. These DNA sequences are thought to interact with transcription factors, which act as molecular switches in response to environmental stress signals (biotic and abiotic) in plants (Yamaguchi-Shinozaki and Shinozaki, 2005). In particular, the ABRE element found in the Phvul.009G190100 gene promoter is an important cis-acting element in ABA-responsive genes and regulates gene expression in response to abiotic stress (Mundy et al., 1990; Xu et al., 1996). In addition, the environmental sensitivity of Phvul.009G190100 is supported by the occurrence of the heat-shock cis regulatory elements (HSEs) in its promoter. Thermo-inducibility of the BnMicEmUP gene homolog of Phvul.009G190100 was observed during embryogenesis in induced B. napus microspore cultures (Shahmir, 2014).

When plants are faced with abiotic and biotic stresses, many stress-related genes can be coordinately expressed because they are connected by common signaling pathways involving molecules such as salicylic acid (SA), ethylene, Methyl jasmonate (MeJA) and absisic acid (ABA) (Apel and Hirt, 2004; Durrant and Dong, 2004; Mittler et al., 2004; Delledonne, 2005; Lorenzo and Solano, 2005; Torres and Dangl, 2005; van Loon et al., 2006). The cis elements
identified in the promoter of the Phvul.009G190100 gene, such as the TGACG-motif, ERE and the ABRE motifs, suggest that it is regulated by jasmonic acid, ethylene and abscisic acid. Stress-related and many defense-related genes respond to jasmonates (Creelman and Muller, 1997) and most ABA-inducible genes contain an ABRE in their promoter regions. Interestingly, transcription factors, such as bZIPs (Shinozaki et al., 2003), with a domain similar to the Phvul.009G190100 protein, are involved in regulating the expression of genes with stress related cis-acting elements in their promoters.

The current study also shows that the upstream sequence of Phvul.009G190100 contains many light responsive cis elements, such as the G-Box, may indicate regulation of its activity by light. Although light responsive elements (LREs) are present in many photoregulated genes, some are found in promoters of genes that are not light-regulated (Kuno and Furuya, 2000). Therefore, the presence of LREs in the promoter of Phvul.009G190100 is not enough to consider this gene as a photoregulated gene. Nevertheless, there may be a link between the light stress and the chloroplast localization of the protein since plants usually reduce the maximum capacity of plant photosynthesis under drought and high salinity conditions by closing stomata and decreasing the carbon dioxide availability in the chloroplasts (Keishi Osakabe and Yuriko Osakab, 2012).

A very interesting finding is that the cis-acting regulatory elements overlap in some regions. For example, an overlap occurs between the HSE and ERE elements at position 330. Other regions include: the overlapping of Box 4, ERE and Box 1 at position 525 and at position 932 with LTR and HSE. These regions may targets for competitive transcription factors.
Overall, these results suggest that it could be interesting to test the expression of Phvul.009G190100 under a variety of stress and light regimes and to examine the role of the gene product in mediating responses to hormones and stresses. Such studies might provide interesting links between development and stress responses.

4.4 Expression Pattern

For the current study, fresh leaf samples were used to measure Phvul.009G190100 gene expression in different bean varieties because it was reasoned that leaves would have relatively high RNA contents. The RNA-seq data available for bean on Phytozome confirmed that this was a good choice, since the Phvul.009G190100 gene expression profiles for different tissues showed that Phvul.009G190100 is highly expressed in green tissues, especially leaves, but is only weakly expressed in the stems, flowers and roots, and is not expressed in nodules.

Similarly, the soybean homologs (Glyma04g34330 and Glyma06g20240) were highly expressed in leaves, pods and flowers but are not much expressed in roots, seeds and nodules. The results indicate that green tissues have highest levels of expression, compared to other organs. Perhaps, because the genes encode putative chloroplast-targeted proteins, their high levels of expression in leaves are related to the relatively high numbers of chloroplasts that green tissues contain. Leaves have been shown to have the largest number of chloroplasts and large increases in chloroplast numbers have been reported for growing and expanding leaf cells (Possingham and Saurer, 1969).

Furthermore, the gene expression data for *Arabidopsis* (eFP browser; Winter et al., 2007) indicates that AT1G74730 (homolog of Phvul.009G190100) is highly expressed in *Arabidopsis*
expanded leaves, young leaves, and the globular stage of the embryo, but, is weakly expressed in the roots and old leaves, and is not expressed in mature siliques. The high levels of expression of AT1G74730 in the global stage of embryo development also supports the relationship between AT1G74730 expression and chloroplast development, because chloroplast biogenesis has been observed during early globular stage embryogenesis (Colombo et al., 2008; Tejos et al., 2010). The results from Tejos et al. (2010) also suggested that the ability of cells to develop chloroplasts is an embryonic factor, which allows chloroplast differentiation to be used as a marker for cell differentiation during embryogenesis.

4.5 Basic Characteristics of the bZIP Domain

The in-silico identification of a basic DNA binding region and a dimerization motif in the putative polypeptide of Phvul.009G190100 suggests that it encodes a bZIP protein, which is the largest transcription factor family in plant genomes (Jakoby et al., 2002). bZIPs contain basic DNA binding domains, consisting of invariant Asparagine (N) and Arginine (R) residues (Miller, 2009) that are believed to be involved in the specificity of DNA binding (Suckow et al., 1993; Suckow et al., 1994) and a dimerization domain consisting of leucine rich repeats (Jakoby et al., 2002).

In the dimerization region of the putative protein encoded by Phvul.009G190100, some of the leucine residues are replaced by other hydrophobic residues such as isoleucine or valine. This finding is in accordance with previous comparisons of the amino acid sequences of typical bZIP proteins, which found that their basic domains are more similar than their leucine zippers (Vettore et al., 1998). Also, in Phvul.009G190100, the periodic repetitions of leucine are located
14 amino acids downstream from the basic region, while the intervening region is nine amino acids long in classical bZIP proteins. However, variation in the number of amino acids between the basic region and the dimerization domain has been observed in many members of the bZIP family, such as the pea PEND (cbZIP) protein (Terasawa and Sato, 2005).

The amino acid sequence analysis of the putative Phvul.009G190100 protein also suggested that it contains sequences that match the consensus phosphorylation site for casein kinase II (CKII). They are similar to phosphorylation sites which occur in many bZIP proteins (Choi H et al., 2000) and are involved in regulating their function (Ciceri et al., 1997). Specifically in the conserved phosphorylation domains of AREB1/2 (AtbZIP38), a bZIP type of transcription factor induced by ABA or abiotic stresses (Choi H et al., 2000; Finkelstein and Lynch, 2000; Lopez-Molina et al., 2001), have been shown to be essential for AREB1/2 to induce downstream genes (Jakoby et al., 2002).

### 4.6 Chloroplast Localization

Several lines of evidence strongly support the conclusion that Phvul.009G190100 is targeted to, and functions within chloroplasts. Protein targeting to the chloroplast is usually due to an N-terminal pre-sequence (Cline and Henry, 1996) and a putative cTP (chloroplast transit peptide) was detected by various software programs for identifying plastid-targeting signals (such as TargetP and PSORT) in the N-terminals of the putative Phvul.009G190100 protein.

As Zhang and Glaser (2002) have reported, the length of typical cTPs varies greatly, from 13 to 146 amino acid residues. In addition, several studies have revealed that the amino acid compositions of all known transit peptides are rich in hydrophobic (i.e., Ala, Leu, Phe, Val) and
hydroxylated (i.e., Ser, Thr, and Pro) amino acid residues but deficient in acidic amino acids (von Heijne et al., 1989; Zhang and Glaser, 2002). Consistent with the above observations, the pre-sequence of the N-terminal in Phvul.009G190100 protein contains 19 aa residues and is rich in hydrophobic (52.6%) and hydrophilic (42.1%) aa but contains no acidic amino acids. Moreover, an \( \alpha \)-helix structure was observed in the cTP region of Phvul.009G190100, that is consistent with the previous report that transit peptides form helical structures during their association with chloroplast envelope membranes (Wienk et al., 1999, 2000).

Furthermore, chloroplast import pre-sequences are cleaved during the import process. Gavel and von Heijne (1990) suggested that a cleavage site motif (Ile/Val)-Xaa-(Ala/Cys) Ala is the characteristic of a majority of higher plant cTPs. Later, a conserved Val-Xaa-Ala-Ala motif (where Xaa represents any amino acid residue) was identified at the putative cleavage site in the N-terminal pre-sequence of the typical cTPs. Specifically, the A/A motif was suggested to be a unique cleavage site for cTPs (Zhang and Glaser, 2002), with the cleavage occurring between the two alanines. Although the cleavage motif, consisting of two alanines (A/A), is not present in the putative Phvul.009G190100 protein cTP, Gavel and von Heijne (1990) had reported that some cTPs, for example PSI V from spinach, that do not have a good match to the consensus at the proposed cleavage site, suggesting that it may be cleaved in more than one step by different proteases. However, it is still possible to predict the most likely cleavage site by locating it to a region where there is a drop in the frequency of serines and a corresponding increase in the frequency of acidic residues.

Evidence for a chloroplast location for the Phvul.009G190100 polypeptide also comes from studies of with similar proteins. For example, the AtbZIP28 (type II membrane protein)
contains an N-terminal bZIP DNA-binding domain at the cytosolic face (Liu et al., 2007; Tajima et al., 2008). A study to identify proteins targeted to the plastid (Koo and Ohlrogge, 2002) indicated the AT1G74730 was possibly localized to the chloroplast envelope. Recently, Shahmir (2014) showed directly that GFP fusion proteins with the homologs for the Phvul.009G190100 in Arabidopsis and B. napus (AtBnMicEmUP::GFP and BnMicEmUP::GFP), in transient expression studies with Nicotiana benthamiana and in stable Arabidopsis transgenics, accumulated in chloroplasts.

4.7 Transmembrane Domain (TMD)

The identification of TMDs in the C-terminal end of the Phvul.009G190100 polypeptide is based on the occurrence of stretches of ~20 hydrophobic residues. Such an arrangement of amino acids is necessary for the formation of a helix to span a biological membrane (Rost et al., 1995).

4.8 Plastid Envelope DNA Binding (PEND) Proteins

The combination of DNA binding motifs with a chloroplast localizing peptide sequence that occurs in the putative Phvul.009G190100 protein is a unique arrangement that is found in a number of proteins including: a chloroplast nucleotide DNA binding protein (CND41) from N. tabacum containing a zinc finger motif (Nakano et al., 1993); a protein designated as PTF1 (Plastid Transcription Factor 1) containing a basic helix-loop-helix (bHLH) motif from Arabidopsis which has been suggested as a plastid transcription factor (Baba et al., 2001); and a bHLH motif containing, wound-induced protein from N. tabacum (NtWIN4) localized in plastids. The NtWIN4 has also been reported to be a transcription repressor with a role in defense

110
against biotic and abiotic environmental stresses (Kodama and Sano, 2006). It has been suggested that these proteins have been functionally changed from eukaryotic nuclear transcription factors into plastid functional proteins during evolution (Kodama and Sano, 2007).

However, a chloroplast cbZIP DNA-binding protein, located in the inner envelope of the Pisum sativum chloroplast (Terasawa and Sato, 2005), called a plastid envelop DNA binding protein (PEND), likely has the most similar structure to the proposed Phvul.009G190100 protein. The PEND protein consists of a short pre-sequence, an N-terminal DNA-binding domain and a central repeat domain (cbZIP) (Sato et al., 1998), and a C-terminal transmembrane domain. Genes coding similar proteins have been identified in more than 6 plant species and although considerable variation in the size of the central region has been observed, the PEND proteins have a common general structure consisting of a chloroplast transit peptide, a basic DNA binding region, a leucine rich repeat and a transmembrane domain region (Sato et al., 1998). The structure of the putative Phvul.009G190100 polypeptide with these various motifs suggests that it is similar to a pea PEND protein (Terasawa and Sato, 2009).

The pre-sequence of the PEND subunit precursor consists of 15 aa, which is therefore the shortest of all known chloroplast protein precursors (which are typically ~30-100 residues). The pre-sequence aa sequence is not conserved in all homologs of PEND; however, a cleavage site L/A was found in most PEND homologs suggesting it may be a unique cleavage site for a chloroplast transit peptide in this protein family.

It has been suggested that PEND may act as a transcription factor that mediates chloroplast signaling to the nucleus (Wagner and Pfannschmidt, 2006). A model for the PEND
protein suggests that the PEND is first targeted to the chloroplast, and the N-terminal pre-sequence is cleaved during transit (Terasawa and Sato, 2009). The C-terminal TMDs are inserted into chloroplast inner membrane and the N-terminal DNA-binding domain is oriented towards the stroma so that it can bind to chloroplast DNA (Wagner and Pfannschmidt, 2006).

Alternatively, if the chloroplast envelope is degraded by stress, the mature PEND might be released and retargeted to the nucleus. The ability of PEND from B. napus to repress nuclear gene expression was found in a tobacco overexpressor line (Wycliffe et al., 2005). This effect could be the result of relocalization of the PEND protein to the nucleus, or dual localization. It is worth noting that the relocalization of transcription factors is not uncommon, as in the case of steroid hormone receptors (Tsuge et al., 2008). A GATA factor is retargeted from the endoplasmic reticulum (ER) to the nucleus upon signal perception (Tsuge et al., 2008). However, currently no evidence for the relocalization of PEND to the nucleus exists. This novel idea could be tested in future studies by transforming the plastid genome with a construct that contains a Phvul.009G190100 protein modified with a peptide tag that can be detected by the Immunohistochemistry analysis.

4.9 Regulated by Stress

The eFP browser revealed that the expression of AT1G74730 is regulated by biotic stresses, such as osmotic, heat and UV stress. For example, infection of Arabidopsis rosette leaves with a geminivirus led to a down-regulation of AT1G74730 (Ascencio-Ibáñez et al., 2008). Previous studies also found that different heat treatments (high temperatures or mild heat stress) in tissues (seedlings and shoots or embryogenic microspores) resulted in different
responses by up or down regulation of the BnMicEmUP expression in *B. napus* (Shahmir et al., 2014). That might be the case that AT1G74730 and BnMicEmUP are targeted to chloroplast and major modifications were reported to occur in chloroplast structure at high temperatures (Karim et al., 1997).

As mentioned earlier, the bean seed development is related to the embryogenesis process which is controlled by many genes and their malfunction can disrupt interspecific embryos formation (Raghavan, 1997). A vast number of genes, such as homeobox genes, heat shock protein genes and lipid transfer protein genes that play a significant role in zygotic embryo development have been identified (Laux et al., 2004). Especially, expression of HEAT-SHOCK genes occurs during the maturation stage of the seed, when cell division has ceased and seeds become tolerant to desiccation. Therefore, the stress might be a vital role in affecting the seed development in beans via gene expression regulation.

The putative undesirable effect of Phvul.009G190100 gene expression on yield and the gene structure analysis in present study might be supporting evidence that the gene Phvul.009G190100 might act as a stress responsive gene. It has been established that bean yield losses occur under high night temperature that negatively affect all stages of reproductive development. Microsporogenesis was found to be the most sensitive developmental stage and pollen viability is decreased at high temperatures and reduced yields have been associated with morphological abnormalities in pollen and anthers (Gross and Kigel, 1994; Porch and Jahn, 2001; Suzuki et al., 2001). High temperatures also inhibit the pollen tube growth and fertilization in studies of ovule positions in relation to seed set (Dickson and Boettger, 1984; Gross and Kigel, 1994; Halterlein et al., 1980; Omrod et al., 1967). Additionally, when under heat stress,
common beans have shown to produce more flower primordial but a high proportion of the abscise, either as primordial or more mature organs (Konsens et al., 1991). It is in agreement with the ABA (abscise acid) signaling cascades involved in regulation of the BnMicEmUP expression (Shahmir and Pauls, 2014). Together, the results suggest that expression of Phvul.009G190100 is a stress-related gene.

4.10 Phvul.009G190100’s Relationship to Yield

The current finding that Phvul.009G190100 is differentially expressed in bean genotypes with different yield potential supports the observations from other systems that members of this gene family are yield-related genes. In particular, the negative trends in the relationships between yield, or yield gain with gene expression levels seen in the current study are consistent with studies with transgenic Arabidopsis plants and mutant lines that showed a significant negative relationship between the level of gene expression from the homologous Arabidopsis gene and seed yield. However, the correlation between bean yield in 2013 and gene expression of Phvul.009G190100 was not statistically significant in the current study. This may be because of the 2013 study was only one year and one location data. But, the historical yield data, from 7 location years, was also not significantly correlated with the expression levels of the Phvul.009G190100. This study also had experimental limitations but it is clear that the work in bean needs to be repeated under more controlled conditions to properly test the hypothesis that expression from the Phvul.009G190100 gene is related to yield.
4.11 Syntenic Relationship of Beans and Soybean

Phvul.009G190100 in *P. vulgaris*, AT1G74730 in *Arabidopsis* and Glyma04g34330 and Glyma06g20240 from *G. max* are suggested to belong to the same gene family, because they share the high nucleotide homologies and have similar gene and protein structures. Furthermore, all the genes had their highest levels of expression in green tissues, where most chloroplasts are found, suggesting the common localization – the chloroplast. From the phylogenetic analysis, proteins of AT1G74730 from *Arabidopsis* and BnMicEmUP from *B. napus* are grouped together while Phvul.009G190100 from common bean and Glyma04g34330, Glyma06g20240 from soybean are grouped into another cluster. These relationships are consistent with the fact that *Arabidopsis* and *Brassica* species are phylogenetically-closely related because they are in the same family (Brassicaceae) while beans and soybean are from the same legume family, namely the Phaseoleae (Arondel et al., 1992; Brunel et al., 1999).

Previous comparative studies revealed conservation of synteny between bean and soybean utilizing marker positions (Galeano et al., 2009; McClean et al., 2010; Galeano et al., 2011). The current study identified the Pv09 of *P. vulgaris* shared syntenic regions with Gm04 and Gm06 of *G. max* which is in agreement with results obtained from Galeano et al. (2011) that Pv09 of bean is syntenic to Gm04 and Gm06 of soybean. In addition, the loci Phvul.009G190100 in bean have two counterparts in the soybean genome which fit well with the whole genome duplication event in soybean after the divergence of beans and soybeans (Schmutz et al., 2010).

In particular, Phvul.009G190100, Glyma04g34330 and Glyma06g20240 are more closely related also because they are yield-related genes that a large number of yield-related QTLs
associated with them. The Phvul.009G190100 physical location on chromosome 9 is in agreement with the previous yield QTL studies (Blair et al., 2006; Tar’an et al., 2002) that two yield QTLs ($yld_{9.1}$, $yld_{9.2}$) had been mapped on the chromosome 9 of the bean. Moreover, the soybean chromosome 6, where the Glyma06g20240 was positioned, has been reported to have four positive yield QTLs from $G. \textit{soja}$ PI 468916 (Wang et al., 2003; 2004). Three yield QTLs were also identified on chromosome 4 where the Glyma04g34330 was located (Kim et al., 2012).
CHAPTER 5 SUMMARY

5.1. Limitations and Future Directions

Even though the similar gene structure and amino acids comparison analysis with BnMicEmUP \((B.\ napus)\) as well as the identification of soybean yield-related QTLs strongly supported the yield-related gene function. In this study’s field yield experiment, there was no correlation of Phvul.009G190100 gene expression with the yield of the bean genotypes. Thus, a wider range of common bean genotypes with different yield ability would be required in the further research to have a better understanding of the association between the gene expression levels with the yield trait. Since there is a Genotype x Environment effect on the yield trait of the crop, multiple planting sites and years with more replications would be required to obtain more typical yield data for each variety. Moreover, a number of agronomic traits for example: 100-seed weight, seed number per pod, number of pods per plant would be helpful to analyze. Moreover, the gene expression levels Phvul.009G190100 could also be tested in the leaves over time during the plant growth to identify a potential relationship between the gene expressions with the plant development.

Although the chloroplast localization of the Phvul.009G190100 has been supported by firstly, software programs have predicted the chloroplast localization with high confidence; a conserved chloroplast transit peptide was detected from the amino acids alignment of Phvul.009G190100 with chloroplast proteins from other species. The localization can be further confirmed experimentally by constructing the Phvul.009G190100-GFP (green fluorescence
protein) fusion protein; the signals in the plastid can be detected under the microscope to see the localization of the Phvul.009G190100 protein.

In addition to identification of gene structure and comparison with already known functions of similar genes, characterization of unknown-function gene can further be determined through gain-of-function and loss-of-function approaches including over expression and RNA interference–based suppression. In the current study, the gene structure analysis and protein function assumption of Phvul.009G190100 suggested that it might function as a yield-related gene with suppressor activity. However, the molecular function remained uncharacterized. Therefore, the future study might be required to investigate functions for Phvul.009G190100 by manipulating AT1G74730 expression degree in Arabidopsis which is a replacement plant for the transformation study. By over-expressing AT1G74730 and reducing its expression by RNA silencing in Arabidopsis, the transgenic Arabidopsis plants can be accessed in terms of biomass yield and seed yield.

As a homolog of Phvul.009G190100, BnMicEmUP (bZIP) from Brassica napus has been suggested to be involved in ABA-mediated transcriptional regulation as a transcription factor induced by stress. The similar stress responsive model (Figure 1.5) might also be accessible for P. vulgaris. To test the direct effects of the gene in ABA response, the ABA sensitivities of Phvul.009G190100 mutants, overexpression and RNAi lines of Phvul.009G190100 can be determined. Gene expression of the Phvul.009G190100 can be compared in these lines when they are treated with ABA.
5.2. Conclusions

In the current study, it was suggested that Phvul.009G190100 functioning as a yield-related gene in common bean and characterized as a cbZIP transcription factor localized to chloroplast involved in stress response. This study play a significant role in uncovering functions of genes with unknown function that may lead to discoveries of genes that control agronomically important traits and that lead to the crop genetic improvement. Hence, the gene could be a potential selection marker for breeding bean varieties with improved yields and it will be important to analyze further populations to identify and confirm the major and minor genes involved in controlling this trait. Molecular genetics, specifically, manipulation of Phvul.009G190100 expression (by down regulation) might be used to produce crops with higher yielding ability and modify yield physiology in favor of higher yields in the future.
REFERENCES


Blair M, Hurtado N, Chavarro CM, Muñoz-Torres MC, Giraldo MC, Pedraza F, Tomkins, Wing R. "Gene-based SSR markers for common bean (Phaseolus vulgaris L.) derived from

121


129


APPENDIX

Table A-1 Analyses of variance for Pearson correlations and covariance parameter estimates of yield (kg/ha) of twelve common bean varieties determined at Elora, ON in 2013

<table>
<thead>
<tr>
<th>CovParm</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Z Value</th>
<th>Pr &gt; Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>6153.38</td>
<td>12292</td>
<td>0.50</td>
<td>0.3083</td>
</tr>
<tr>
<td>Residual</td>
<td>104072</td>
<td>25621</td>
<td>4.06</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Type 3 Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>11</td>
<td>33</td>
<td>37.75</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table A-2 Comparison of yield (kg/ha) between three groups of common bean varieties (high yielding, medium yielding and low yielding) determined at Elora, ON in 2013

<table>
<thead>
<tr>
<th>Contrasts</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>*HY vs *LY</td>
<td>1</td>
<td>33</td>
<td>243.91</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HY vs *MY</td>
<td>1</td>
<td>33</td>
<td>3.33</td>
<td>0.0771</td>
</tr>
<tr>
<td>MY vs LY</td>
<td>1</td>
<td>33</td>
<td>190.24</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*HY: High yielding
*M: Medium yielding
*LY: Low yielding

Table A-3 Analyses of variance for Pearson correlations and covariance parameter estimates of daily yield gain (kg/ha/day) of twelve common bean varieties determined at Elora, ON in 2013

<table>
<thead>
<tr>
<th>CovParm</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Z Value</th>
<th>Pr &gt; Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>1.6637</td>
<td>2.1765</td>
<td>0.76</td>
<td>0.2223</td>
</tr>
<tr>
<td>Residual</td>
<td>16.6302</td>
<td>3.5456</td>
<td>4.69</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Type 3 Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>11</td>
<td>44</td>
<td>17.35</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table A-4 Comparison of the daily yield gain (kg/ha/day) of twelve common bean varieties determined at Elora, ON in 2013 and in the historical trial; Historical yield trials conducted at St. Thomas, ON in 2010 and St. Thomas, Woodstock, and Thorndale, ON in 2011 and 2012, respectively (Navabi, et al. 2012).

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013 VS historical</td>
<td>1</td>
<td>6699092.667</td>
<td>6699092.667</td>
<td>33.36</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Table A- 5 Comparison of daily yield gain (kg/ha/day) between three groups of common bean varieties (high yielding, medium yielding, low yielding) determined at Elora, ON in 2013

<table>
<thead>
<tr>
<th>Label</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>HY vs LY</td>
<td>1</td>
<td>33</td>
<td>248.29</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>HY vs MY</td>
<td>1</td>
<td>33</td>
<td>0.09</td>
<td>0.7692</td>
</tr>
<tr>
<td>MY vs LY</td>
<td>1</td>
<td>33</td>
<td>239.05</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

*HY: High yielding
*MY: Medium yielding
*LY: Low yielding

Table A- 6 Analysis of variance (ANOVA) of the relative gene expression level (Ct target-Ct skip16) of twelve common bean varieties determined at Elora, ON in 2013

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>11</td>
<td>127</td>
<td>14.58</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>