

Julia C. Hooker^{1,2}, Elroy Cober¹, Ashkan Golshani², Bahram Samanfar^{1,2}

¹Agriculture and Agri-Food Canada, Ottawa, Ontario, K1A 0C5; ²Carleton University, Ottawa, Ontario, K1S 5B6

Abstract

Soybean [*Glycine max* (L.) Merr.] is among the most important agronomic crops in Canada with widespread uses in human consumption, animal feed, and biotechnology. The capacity to fix atmospheric nitrogen into biologically available forms gives soybean an important role in sustainable agricultural practices (i.e., reducing the need for nitrogen fertilizers). Understanding the effect environmental variation has on gene expression in Canadian soybean is valuable for developing sustainable agricultural systems. The Canadian Grain Commission has reported lower seed protein content from soybeans grown in western Canada compared to eastern Canada, regardless of genotype. This project will uncover key genes responsible for differences in seed protein content across Canada.

Here we use a transcriptome-wide approach to identify differences in expression of genes which contribute to seed protein content, and to study the effect of environmental variation on geographically-dependent gene expression (West vs East). Ten soybean lines ranging low to high in seed protein content are growing in four locations over four years across western and eastern Canada. Using RNA sequencing, differential transcript analysis of each line is compared between West and East to determine key genes responsible for lower seed protein content in western soybeans. Analysis of first year data has identified three genes encoding cupins and 93 lipid-related genes. This research will provide novel information about the best geographically-fitting soybean cultivars to grow across Canada's growing regions. The findings of this research will be used to develop allele-specific markers, assisting breeding programs to develop high protein soybean cultivars for western Canada.

Introduction

Soybean seeds contain higher plant protein content than any other legume, making seed protein content a key factor in quality standards^{2,3}. The 2020 Quality of Canadian Soybean (oilseed-type) reports 38.8% protein content for soybean grown in eastern Canada and 37.6% protein for those in western Canada¹. Loss of interest in soybean production means losing a key nitrogen fixing partner from Canadian crop rotation practices. The Canadian Agricultural Policy Framework urges that initiatives be taken for environmentally sustainable agricultural practices to adapt to anticipated changes surrounding climate change and population growth. Adapting Canadian soybean agriculture to changing climate conditions is important to producing an adequate crop yield with acceptable levels of seed protein content. Strategic planting practices are fundamental to the future of agriculture as the impacts of climate change emerge, thus it is imperative to understand soybean seed protein gene expression patterns in eastern- and western-grown soybeans.

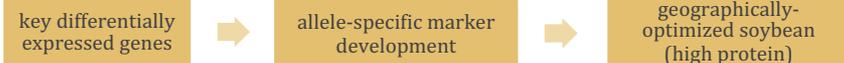
There is a knowledge gap in understanding comprehensive mechanisms of seed protein biosynthesis in soybean. Key genes need to be established as well as the regulatory mechanisms for expression/suppression. Further, the environmental factors affecting quality and quantity of seed protein production have yet to be established, posing another major inhibitory gap in knowledge on this subject.

Objectives & Hypothesis

The main objective of this study is to identify new genes contributing to seed protein and oil content, and to investigate the effect of environmental variation on geographically-dependent gene expression (West vs East) using transcriptome-wide differential expression analysis.

It is hypothesized that soybeans grown in western Canada differentially express some of their seed storage protein-related genes from those grown in eastern growing regions.

Future directions



The results of this project will be used for development of allele-specific markers to be used in marker-assisted selection (MAS) for strategic molecular breeding of soybean. Additionally, this project is generating a large RNA-seq data set (big data) that is incredibly valuable for subsequent investigation of DGE analysis of genes in other economically important pathways, such as time of flowering and maturity.

Acknowledgements

A special thank you to my fellow lab members, our collaborators at the University of Saskatchewan, the bioinformatics teams at Carleton University and McGill University, and the AAFC soybean field units in Brandon and Morden. Thank you to CFCRA, AAFC and Carleton University for financial support.

1. Canadian Grain Commission, 2020; 2. Natarajan et al., 2013; 3. Huang et al., 2019

Methods

Using transcriptomics to geographically-optimize soybean seed protein



Figure 1 Average soybean seed protein content is higher in soybeans grown in eastern growing regions. Ten soybean lines ranging in seed protein content (low to high) are growing in four different locations across eastern and western Canada. Tissue samples in triplicate are being collected annually and subjected to RNA-seq to establish a comprehensive data set for all ten lines in each location. × 4 years

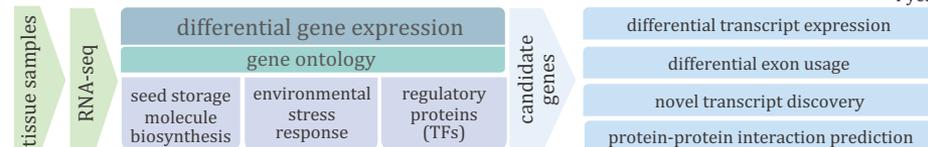


Figure 2 Workflow for RNA-sequencing and differential gene expression analysis and curation of candidate gene lists. Planting, sampling and analysis will be repeated over four years and cumulative data provides a high-quality comprehensive dataset.

Preliminary results

Table 1 Gene ontology and NCBI gene description of differentially expressed seed storage protein genes. Table values represent the log₂ fold change differences in expression of each gene, DESeq, adjusted p-value and edgeR adjusted p-value < 0.01, log₂ fold change of at least 3 in at least 3 of 10 lines in the West.

	Brandon										Morden										MF_GO_ID	MF_GO description	Gene Description (NCBI)
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10			
Glyma.16G*****		4.49		3.33			3.17														GO:0045735; GO:0030145	nutrient reservoir activity; manganese ion binding	germin-like protein 4
Glyma.19G*****												-8.98	-3.68		-4.83			-5.48	-5.01		GO:0045735; GO:0030145	nutrient reservoir activity; manganese ion binding	germin-like protein
Glyma.20G*****	-3.11	-4.94	-3.74		-3.26																GO:0030145; GO:0045735	manganese ion binding; nutrient reservoir activity	germin-like protein

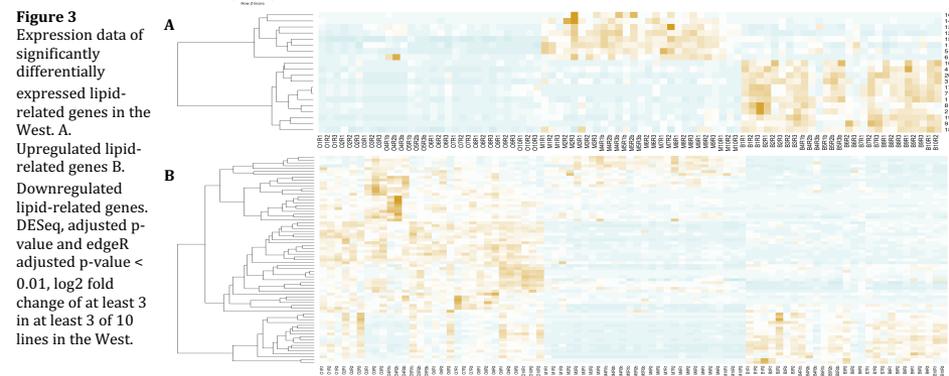


Figure 3 Expression data of significantly differentially expressed lipid-related genes in the West. A. Upregulated lipid-related genes. B. Downregulated lipid-related genes. DESeq, adjusted p-value and edgeR adjusted p-value < 0.01, log₂ fold change of at least 3 in at least 3 of 10 lines in the West.