

S-08

Soybean Protein-protein Interaction Prediction Engine (Soybean-PIPE): A computational approach in soybean functional genomics

*Bahram Samanfar**, AAFC-Ottawa RDC, Ontario, Canada

Elroy Cober, AAFC-Ottawa RDC, Ontario, Canada

Brad Barnes, Carleton University, Ontario, Canada

James Green, Carleton University, Ontario, Canada

Soybean is one of the major Canadian grain crops and its production is expanding in Canada with the majority of the increase in short season areas (Western Canada and northern regions). So far, eleven maturity loci have been reported in soybean, however, the molecular basis of almost half of them is not yet clear. The list of novel factors affecting these pathways in soybean, and in model plants like Arabidopsis, continues to grow suggesting the presence of other novel players which are yet to be discovered. The soybean Protein-protein Interaction Prediction Engine (Soybean-PIPE) is a computational tool used to predict protein-protein interactions (PPI) in soybean. Protein-Protein Interactions (PPIs) are essential molecular interactions that define the biology of a cell, its development and its responses to various stimuli. Theoretically, if a gene interacts with groups of genes involved in one specific pathway, that gene might also be involved in that specific pathway (“guilt by association”). Our knowledge of global PPI networks in complex organisms such as human and plants is restricted by technical limitations of current methods. Briefly, PIPE searches for re-occurring short polypeptide sequences between known interacting protein pairs and novel proteins and predicts interactions based on protein sequence information and a database of known interacting protein pairs. PIPE has been used to produce proteome-wide, all-to-all predicted interactomes in a variety of organisms including yeast (*Saccharomyces cerevisiae*), human (*Homo sapiens*), Arabidopsis and others. PIPE is typically tuned, for a given organism, to achieve a specificity of 99.95%, i.e., to be conservative in predicting PPIs. PIPE has been independently evaluated and compared to other PPI prediction methods and has been shown to significantly outperform the others in terms of recall-precision across all of the datasets tested. It has also been shown that PIPE has the ability to produce cross-species predictions (i.e. use experimental interaction data from one organism to make high quality PPI predictions in another). Recently, PIPE has been redesigned to be able to computationally handle the large proteome of soybean (75,778 confirmed soybean protein sequences). Currently we are using PIPE towards predicting the first comprehensive protein-protein interaction network for soybean ever generated. In an independent study (Samanfar et al., 2016), we have used three different approaches; bioinformatics (Soybean-PIPE), classical plant breeding, and molecular biology (analysis of SSR and SNP haplotypes) to identify a novel gene involved in time of flowering and maturity in soybean. This strategy successfully identified a new maturity locus tentatively called “E10” and the underlying candidate gene (FT4). Identification of molecular markers tagging the PIPE-identified genes controlling flowering and maturity in soybean will allow soybean breeders to efficiently develop varieties using molecular marker assisted breeding. Allele specific markers will allow stacking of early maturity alleles to develop even earlier maturing

cultivars. This bioinformatics approach (soybean-PIPE) will also help to bridge the gap in knowledge of the flowering and maturity pathway in soybean and can be applied to other important traits such as seed protein content, oil quality and host-pathogen interactions.