Theor Appl Genet
DOI 10.1007/s00122-016-2819-7

ORIGINAL ARTICLE

Mapping and identification of a potential candidate gene for a novel maturity locus, *E10*, in soybean

Bahram Samanfar[1] · Stephen J. Molnar[1] · Martin Charette[1] · Andrew Schoenrock[2] · Frank Dehne[2] · Ashkan Golshani[3] · François Belzile[4] · Elroy R. Cober[1]

Received: 18 July 2016 / Accepted: 27 October 2016

Abstract

Key message **E10 is a new maturity locus in soybean and FT4 is the predicted/potential functional gene underlying the locus.**

Abstract Flowering and maturity time traits play crucial roles in economic soybean production. Early maturity is critical for north and west expansion of soybean in Canada. To date, 11 genes/loci have been identified which control time to flowering and maturity; however, the molecular bases of almost half of them are not yet clear. We have identified a new maturity locus called "*E10*" located at the end of chromosome Gm08. The gene symbol *E10e10* has been approved by the Soybean Genetics Committee. The *e10e10* genotype results in 5–10 days earlier maturity than *E10E10*. A set of presumed *E10E10* and *e10e10* genotypes was used to identify contrasting SSR and SNP haplotypes. These haplotypes, and their association with maturity, were maintained through five backcross generations. A functional genomics approach using a predicted protein–protein interaction (PPI) approach (Protein–protein Interaction Prediction Engine, PIPE) was used to investigate approximately 75 genes located in the genomic region that SSR and SNP analyses identified as the location of the *E10* locus. The PPI analysis identified FT4 as the most likely candidate gene underlying the *E10* locus. Sequence analysis of the two FT4 alleles identified three SNPs, in the 5′UTR, 3′UTR and fourth exon in the coding region, which result in differential mRNA structures. Allele-specific markers were developed for this locus and are available for soybean breeders to efficiently develop earlier maturing cultivars using molecular marker assisted breeding.

Abbreviations

GO	Gene ontology
MP-PIPE	Massively parallel protein–protein interaction prediction engine
PIPE	Protein–protein interaction prediction engine
PPIs	Protein–protein interactions
QTLs	Quantitative trait loci
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat

Introduction

In order to expand soybean, *Glycine max* (L.) Merr., cultivation in western Canada and in northern regions, one must accumulate early maturity alleles to develop short-season adapted (early flowering and maturity) lines. Both time of flowering and maturity in soybean are controlled by photoperiod sensitivity genes (Garner and Allard 1927; Cober et al. 2014). These genes also have pleiotropic effects on seed yield

Communicated by V. Hahn.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-016-2819-7) contains supplementary material, which is available to authorized users.

✉ Elroy R. Cober
elroy.cober@agr.gc.ca

[1] Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, ON K1A 0C6, Canada

[2] School of Computer Science, Carleton University, Ottawa, ON K1S 5B6, Canada

[3] Department of Biology and Ottawa Institute of Systems Biology, Carleton University, Ottawa, ON K1S 5B6, Canada

[4] Département de Phytologie and Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec City, QC G1V 0A6, Canada

Published online: 10 November 2016

and other characteristics such as plant height (Cober and Morrison 2010). There are large numbers of quantitative trait loci (QTLs) [first flower—77 QTLs, pod maturity (R8)—160 QTLs, pod maturity beginning (R7)—5 QTLs, pod beginning (R3)—6 QTLs, etc.] in different linkage groups which have been identified to be involved in flowering and maturity and catalogued in different databases including SoyBase (http://soybase.org/). So far 11 major genes/loci [E1 and E2 (Bernard 1971), E3 (Buzzell 1971), E4 (Buzzel and Voldeng 1980; Saindon et al. 1989a, b), E5 (McBlain and Bernard 1987; McBlain et al. 1987), E6 (Bonato and Vello 1999), E7 (Cober and Voldeng 2012), E8 (Cober et al. 2010), E9 (Kong et al. 2014; Zhao et al. 2016), J (Ray et al. 1995) and Dt1 (Liu et al. 2010; Tian et al. 2010)] have been reported in soybean affecting time of flowering and maturity. The underlying genes in half of the loci (E1–E4, E9 and Dt1) have been identified (Liu et al. 2008; Watanabe et al. 2009, 2011; Xia et al. 2012; Zhao et al. 2016), but the candidate genes at other maturity loci have not been identified.

As mentioned previously, flowering and maturity processes in soybean are not fully understood. In many organisms, protein–protein interactions (PPI) have been used to identify and/or propose a shortlist of novel potential candidates in complex pathways. PPIs are vital physical interactions that define the functionality of an organism, its development and responses to various provocations. Most proteins realize their functions by interacting with each other and forming protein complexes. Understanding protein interactions in an organism helps us better understand the biology of that organism (Jessulat et al. 2011). Theoretically, if a protein interacts with a number of proteins involved in a specific pathway, the protein of interest might also be involved in the same specific pathway. This is a concept called "guilt by association" which is broadly used to assign function(s) for unknown protein(s) based on their interactions with known protein partners (Omidi et al. 2014; Vlasblom et al. 2015). Our knowledge of comprehensive PPI networks in complex organisms such as human and plants is technically limited (Fields and Song 1989;Jessulat et al. 2011). Traditional methods such as yeast two hybrid and tandem affinity purification have high false negative/positive rates and some proteins simply cannot be tested by these methods due to technical limitations. In this context, computational methods can provide efficient alternatives for studying novel PPIs. The first global PPI prediction was performed in the unicellular model organism, *Saccharomyces cerevisiae* (Pitre et al. 2008). With advancements in prediction algorithms and high-performance computing, prediction methods have recently been applied to more complex organisms such as human (Schoenrock et al. 2014).

In this study, we used a bioinformatics tool called Protein–Protein Interaction Prediction Engine (PIPE) to predict effectively (low rate of false negative/positive) the first semi-comprehensive PPI network in soybean. Soybean functional genomics data acquired through PIPE will help us to identify, to develop a shortlist and to confirm the involvement of candidate genes in various pathways including early flowering and maturity. PIPE has been used to generate proteome-wide, all proteins-against-all proteins (all-to-all) predicted interactomes in a range of organisms including yeast, *S. cerevisiae*, (Pitre et al. 2008) and human, *Homo sapiens*, (Schoenrock et al. 2014), among others (Pitre et al. 2012). PIPE can quickly produce individual PPI predictions and is typically tuned, for a given organism, to achieve a specificity of 99.95% (see Pitre et al. 2008; Schoenrock et al. 2014, for details on how PIPE's predictive performance is evaluated). 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 (Park 2009). PIPE (Pitre et al. 2012) has also been shown to efficiently produce cross-species predictions (i.e. use interaction data from one organism to make high-quality PPI predictions in another related organism).

Soybean maturity isolines have been developed in both 'Maple Presto' and 'Harosoy' backgrounds. Based on genetic and genotyping data, OT98-17 in a Maple Presto background and OT02-18 in a Harosoy background have the same known maturity genotype (Cober et al. 2010); however, OT98-17 was earlier maturing in the field in Ottawa, Canada (45.4214°N, 75.6919°W) compared to OT02-18. The hypothesis is that this maturity difference is controlled by a single gene and that earliness or lateness could be backcrossed into OT02-18 and OT98-17, respectively. The objective of this study was to understand the genetic control of maturity between these two isolines, to map the maturity locus, and to identify a candidate gene controlling maturity.

Materials and methods

DNA extraction, PCR, and sequencing

DNA was extracted from frozen leaves of plants grown in the greenhouse or the field using a modified urea extraction technique described by Molnar et al. (2003).

Sequencing primers were designed using Primer3 online software (www.bioinfo.ut.ee/primer3) for FT4 and FTL7. Soybean SSR primers designed by Cregan et al. (1999) were used in this study for DNA amplification. All primers were purchased from IDT Inc. (Integrated DNA Technology, Coralville, Iowa, United States). A list of primers used in this study is presented in supplementary file 1.

DNA amplification was performed as described by Cregan et al. (1999) with some modifications described by Molnar et al. (2003).

Extracted DNA was amplified using gene-specific primers (supplementary file 1) to produce high-quality PCR amplicons and purified using BigDye® Terminator v3.1 Cycle Sequencing Kit (Life technologies, Carlsbad, California, USA) as per the manufacturer's instructions. DNA sequencing of the PCR amplicons was performed using ABI BigDye Terminator chemistry v3.1 (Applied Biosystems, Foster City, California, USA) and run on an ABI 3130*xl* automated sequencer (Applied Biosystems/Hitachi, Foster City, California, USA) at in-house facilities located at AAFC-ORDC, Ottawa.

Bioinformatics analyses

A database of SNP genotypes of 300 Canadian soybean varieties created by Sonah et al. (2013, 2015) was used as a source for SNP haplotype analysis (Table 5).

Gene ontology analysis was performed using online tools including SoyBase (http://soybase.org/) for Soybean and GeneMANIA (http://www.genemania.org/) for *Arabidopsis*.

Two-dimensional structures of RNA for FT4 and FTL7 coding sequences isolated from *E10* and *e10* lines were determined using phiSITE (http://www.phisite.org/main/) and RNA fold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

The Protein–Protein Interaction Prediction Engine (PIPE) works based on re-occurring short amino acid sequences between known interacting protein pairs (similarity matrix); simply, it predicts interactions based on protein sequence information and a database of known interacting pairs (Pitre et al. 2006).

PIPE requires a set of known interacting protein pairs as well as their primary (amino acid) sequences to be able to make its predictions. For a given pair of proteins *A* and *B*, PIPE performs the following steps to predict if they interact or not:

(i) First PIPE converts the protein pairs known to interact into an interaction graph G, where every protein is represented as a node in G and a known interaction between two proteins is represented as an edge in G.
(ii) Protein A is investigated to find similar proteins in the interaction graph G. To do this, protein A is split up into overlapping windows of size 20 and each one of these amino acid segments is compared to all segments of size 20 in all of the proteins that make up the graph G. To evaluate how similar these segments are, the PAM120 matrix is used. If the similarity score is above a tunable threshold, the segments are said to match. If protein X in G contains a segment which matches a segment at position i in protein A, all neighbours of X in G are put into a list R.
(iii) Next, protein B is investigated to see if it contains similar regions to those proteins now stored in list R in the same manner as step 2. If protein Y (known to interact with X) in R contains a region similar to protein B at position *j*, the result matrix of size $n*m$ (where $|A| = n$ and $|B| = m$) which is initially set up to contain only zeroes, is incremented at position I, *j*.
(iv) The above steps are repeated for all possible protein window combinations on proteins A, B and those proteins which make up the graph G. Each cell a, b of the result matrix now stores how many segments of proteins known to interact share similar regions to query proteins A (at position a) and B (at position b). These co-occurring segments (in our query pair and our known interacting pairs) represent evidence towards potential interaction. The result matrix is then visualized as a 3D landscape and, based on this landscape, proteins A and B are predicted to interact or not.
(v) A subset of 44,140 proteins (this subset included all the confirmed genes involved in the flowering and maturity pathway of soybean, plus the 75 genes located in the *E10* region, plus a subset of approximately 44,000 random soybean genes, approximately 60% of the total soybean gene complement) was derived for PIPE analysis. PIPE evaluated all possible protein pairs between the 44,140 proteins studied. The results were then sorted by their predicted interaction score and only the top scoring pairs were considered as potentially interacting for gene ontology analysis.

Massively Parallel PIPE (MP-PIPE) is the latest implementation of PIPE and computer scientists have improved PIPE efficiency. MP-PIPE is capable of producing extremely high-quality predicted interactomes (specificity of 99.95%) for organisms as complex as human and plants in a reasonable time using a variety of compute clusters (Schoenrock et al. 2014). We used MP-PIPE to investigate protein–protein interactions in a reduced set of soybean proteins due to current programing limits.

Results

Genetic stocks and field phenotyping for maturity

Isolines of Maple Presto and Harosoy have been grown in the field in Ottawa over multiple years (Table 1). OT98-17 (Maple Presto isoline) and OT02-18 (Harosoy isoline) have the same genotype at the major known maturity loci (e1e1 e2e2 e3e3 e4e4 e7e7 e8e8) and yet OT98-17 matured 6 days

Table 1 Agronomic characteristics of Maple Presto and Harosoy maturity near-isogenic soybean lines grown at Ottawa, ON from 2007 to 2011

Isolines	Background (substituted alleles)	Seed yield (kg/ha)	Maturity (days)	Plant height (cm)
OT98-17[a,b,c]	Maple Presto (e8)	917	90	61
Maple Presto[b]	MP[b,e]	2048	100	76
OT02-18[d]	Harosoy[f] (e3 e4 e7 e8)	1365	96	61
OT94-47	Harosoy (e3 e4 e7)	2096	103	72
OT89-5	Harosoy (e3 e4)	2387	105	83
OT99-17	Harosoy (E2 e3 e4)	2609	112	85

[a] OT98-17 = X824A-ve/7*Maple Presto. X824A-ve = PI 438477/2*Evans/2/L62-667

[b] (Cober et al. 2010)

[c] OT98-17 has known maturity alleles e1-nl e2 e3-tr e4-SORE

[d] OT02-18 has known maturity alleles e1-fs e2 e3-tr e4-SORE

[e] Maple Presto (Amsoy/Portage/2/PI 438477) has the genotype e1 e2 e3 e4 e7 E8 E9

[f] The recurrent parent Harosoy has the genotype (e1-as e2 E3 E4 E7 E8 E9). For locus E9, E9E9 is the early genotype

Table 2 Days to maturity of BC_4 soybean parents and BC_5 bulks, from the introgression of *E10* into OT98-17 and *e10* into OT02-18, and maturity check cultivar Maple Presto, grown at Ottawa, ON in 2015

		Background	Days to maturity	Standard error
BC_4 parents				
Late maturity introgressed	Maple Presto		94.3	1.3
Early maturity introgressed	Harosoy		89.6	1.3
Maple Presto			98.5	2.1
BC_5 bulks				
Late maturity		Maple Presto	103.2	0.6
		Harosoy	109.6	0.5
Early maturity		Maple Presto	94.8	0.8
		Harosoy	95.2	0.8
Maple Presto			105.0	1.1

The BC_4 parents and Maple Presto and the BC_5 bulks and Maple Presto were grown in different fields

earlier than OT02-18 in the field (Table 1). To introgress early maturity into OT02-18 and late maturity into OT98-17, an OT02-18/OT98-17 cross was made in 2002. The F_2 population was grown in 2003 and F_3 progeny rows were grown in the field at Ottawa in 2004. The three earliest maturing lines were selected to backcross to OT02-18 and the three latest maturing lines were selected for backcrossing to OT98-17. Parental selection for early or late maturity during backcrossing was carried out to the backcross five generation. BC_1 F_3 progeny rows were grown in 2006. BC_2 F_3 progeny rows were grown in 2008. BC_3 F_3 progeny rows were grown in 2010. BC_4 F_3 progeny rows were grown in 2013. BC_5 F_3 progeny rows were grown in 2015 and date of maturity was observed. BC_4 lines used as parents were also grown in 2015 and date of maturity was recorded (Table 2).

SSR markers linked to maturity

Both OT02-18 and OT98-17, as well as related germplasm, were genotyped with 78 SSR markers as part of a previous study of the E8 maturity locus (Cober et al. 2010). These SSR markers are widely distributed across the whole genome. To search for the chromosomal location of *E10*, these SSR genotyping data were investigated for a contrast between OT98-17 (*e10*) and the presumed *E10* lines OT02-18, Harosoy, Maple Presto, and OT94-47. Only one genomic region of three adjacent SSR markers (Satt429, Satt538 and Satt378) on chromosome Gm 08 (Linkage Group A2) correlated perfectly with this genetic contrast (Table 3). The correlation was maintained when these lines were genotyped with the adjacent marker Sat_347 (Table 3).

Part way through the development of new maturity isolines, genotyping of five early maturing putative *e10* BC3 lines with the four SSR markers produced SSR banding patterns identical with the *e10* donor line OT98-17 (Table 3). Genotyping four late-maturing putative *E10* BC3 lines gave banding patterns identical with the *E10* donor line OT02-18 at three of the four SSR markers (Table 3).

SNP markers linked to maturity

A number of genetic lines, presented in Table 3, were part of a collection of approximately 300 Canadian breeding lines that have been independently genotyped with SNP markers (Sonah et al. 2013, 2015). The approximately 10,300 SNPs provide wide genome coverage. The SNP genotype database was examined for a contrast between *e10* and *E10* lines. Two lines, X5285 and X5288 (OT02-18*3/OT98-17, BC_2 *e10* lines), are presumed to be *e10* because they are the early maturity selections used as BC_1 parents in the next backcross to OT02-18. The lines OT02-18, Harosoy, OT94-47, X5289, and X5291 (the two later maturing lines have the same pedigree: OT02-18/3*OT98-17, *E10* BC_2) lines are presumed to be *E10*. The only genomic location where the SNP haplotype correlated perfectly with *e10*/*E10* status was a continuous block of 18 SNPs on the end of chromosome Gm08. This block also contained some SNPs that were monomorphic across this subset of *e10* and *E10* lines. The first of the correlating SNPs was at Gm08:46262303 (*E10E10*: A and *e10e10*: C) and the last at Gm08:46948064 (*E10E10*: C and *e10e10*: T), where positions are defined in *Glycine max* Genome Assembly Version Glyma.Wm82.a1 (Gmax1.01) at www.soybase.org (accessed March 7, 2016). These SNPs span the region of Gm08 which also contains the four SSR markers Sat_347 (46284254 … 46284485), Satt429 (46365992 … 46366258), Satt538 (46545795 … 46545904) and Satt378 (46760180 … 46760343) used previously. Therefore, the location of the *E10*/*e10* locus based on the SNP data is completely consistent with the location that was independently predicted by the SSR data (see Table 5).

Table 3 SSR screening for the *E10* soybean maturity locus (color figure online)

Name	Genetics	E10 status	SNP or SSR Marker[a] and its Location on recombination (cM) and sequence (bp) maps of chromosome Gm08			
			Sat_347	Satt429	Satt538	Satt378
			158.4 cM	162.0 cM	159.6 cM	165.7 cM
			46284254	46365992	46545795	46760180
OT98-17	-	*e10e10*	B	B	B	B
OT02-18	-	*E10E10*	C	C	C	A
X5494 (OT02-18/X5285-1-9-1)	Harosoy e10 (BC_3)	*e10e10*	B	B	B	B
X5495 (OT02-18/X5285-1-9-2)	Harosoy e10 (BC_3)	*e10e10*	B	B	B	B
X5496 (OT02-18/X5285-1-9-3)	Harosoy e10 (BC_3)	*e10e10*	B	B	B	B
X5497 (OT02-18/X5288-1-23-1)	Harosoy e10 (BC_3)	*e10e10*	B	B	B	B

Table 3 continued

Name	Genetics	E10 status	SNP or SSR Marker[a] and its Location on recombination (cM) and sequence (bp) maps of chromosome Gm08			
			Sat_347	Satt429	Satt538	Satt378
			158.4 cM	162.0 cM	159.6 cM	165.7 cM
			46284254	46365992	46545795	46760180
X5498 (OT02-18/X5288-1-23-2)	Harosoy e10 (BC_3)	*e10e10*	B	B	B	B
X5499 (OT98-17/X5289-1-10-1)	M. Presto E10 (BC_3)	*E10E10*	C	C	C	-
X5500 (OT98-17/X5289-1-10-2)	M. Presto E10 (BC_3)	*E10E10*	C	C	C	-
X5501 (OT98-17/X5291-1-14-1)	M. Presto E10 (BC_3)	*E10E10*	C	C	C	-
X5502 (OT98-17/X5291-1-14-2)	M. Presto E10 (BC_3)	*E10E10*	C	C	C	-
Harosoy	Recurrent parent of OT02-18	*E10E10*	D	C	C	-

Table 3 continued

Name	Genetics	E10 status	SNP or SSR Marker[a] and its Location on recombination (cM) and sequence (bp) maps of chromosome Gm08			
			Sat_347	Satt429	Satt538	Satt378
			158.4 cM	162.0 cM	159.6 cM	165.7 cM
			46284254	46365992	46545795	46760180
Maple Presto	Recurrent parent of OT98-17	*E10E10*	E	D	C	A
Amsoy 71[b]	Ancestor of OT98-17	*E10E10*	F	D	C	A
Portage	Ancestor of OT98-17	*E10E10*	CD	CD	AC	A
840-7-3	Ancestor of OT98-17	-	A	A	A	A
Evans	Ancestor of OT98-17	*E10E10*	D	A	C	A
L62-667	Ancestor of OT98-17	*E10E10*	C	C	C	A
OT94-47	-	*E10E10*	-	A	A	A

The table shows banding patterns for SSR markers. *Dark blue shading* shows agreement between *e10* lines and the *e10* donor (OT98-17) while light blue shading shows agreement between *E10* lines and OT02-18

[a] These four SSR markers are ordered in accordance with the physical genomic sequence map

[b] Amsoy is in the pedigree of Maple Presto and OT98-17. Amsoy 71 used in this study is Amsoy with a phytophthora resistance gene backcrossed in, and is an accurate representation of Amsoy at the *e10* locus, assuming the *e10* locus is genetically independent of the phytophthora resistance locus

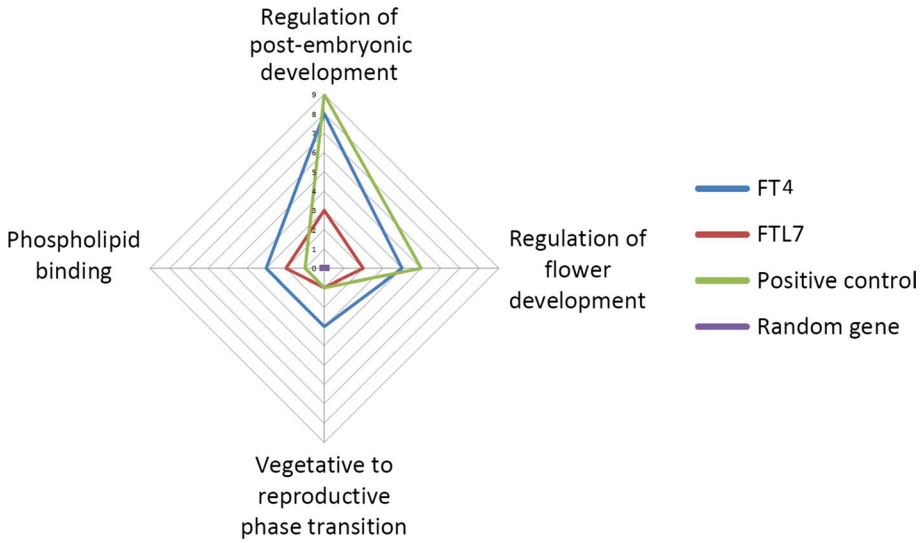

Fig. 1 Representative GO analysis of soybean *E10* candidate genes FT4 and FTL7, a positive control maturity gene E3 (PhyA3), and a random negative control gene (zero interaction with genes involved in flowering and maturity), identified through the semi-comprehensive run of soybean MS-PIPE. Detailed information is presented in supplementary file 3

Identifying a candidate gene for *E10*

In order to identify the underlying gene responsible for the *E10* locus, a list of annotated genes within the *E10* chromosomal region (between Sat_347 and Satt378) was compiled using the SoyBase browser (http://soybase.org/) and approximately 75 candidate genes were identified. To identify which of these genes corresponds to the *E10* locus, a bioinformatics approach called PIPE (Protein–protein Interaction Prediction Engine), which predicts PPIs, was used.

Given that there is great homology between the flowering model plant *Arabidopsis* and soybean (Grant et al. 2000; Wickland and Hanzawa 2015), and due to the lack of known, experimentally validated soybean PPIs, PPI data from *Arabidopsis thaliana* were used to supplement the approximately 15 confirmed PPIs of soybean reported in Uniprot (http://www.uniprot.org/). The PIPE reference database for this project was constructed using 27,450 confirmed protein sequences sourced from the *Arabidopsis* genome database at TAIR (https://www.Arabidopsis.org/) and 14,264 experimentally validated high confidence *Arabidopsis* PPIs taken from TAIR, GeneMANIA, and Uniprot. First, an all-to-all predicted interactome was produced on all possible 376,764,975 *Arabidopsis* protein pairs. Leave-one-out cross-validation experiments (as described by Pitre et al. 2008; Schoenrock et al. 2014) were carried out on the resulting interactome and PIPE achieved a sensitivity of 42.65% at its normal operating point of a specificity of 99.95%. Given this good predictive performance, the known *Arabidopsis* interactions plus the 15 confirmed PPIs of soybean were used to predict PPIs in soybean. As shown by Pitre et al. (2012), PIPE can successfully predict PPIs in one organism given good interaction data from another related organism. PIPE was shown to produce quality PPI predictions using data from organisms as distant as human, *Homo sapiens*, and yeast, *S. cerevisiae*. A total of 75,778 confirmed soybean protein sequences were retrieved from SoyBase (http://soybase.org/). This high number of protein sequences results in a significant increase in the number of combinations of potential PPIs, exceeding the current computational capability of MP-PIPE. Consequently, a subset of 44,140 proteins was derived for PIPE analysis. This subset included all the confirmed genes involved in the flowering and maturity pathway of soybean, plus the 75 genes located in the *E10* region, plus a subset of approximately 44,000 soybean genes chosen with a random number generator for approximately 60% of the total soybean gene complement. This subset was used to run a semi-comprehensive PPI analysis of soybean. MP-PIPE evaluated all 974,191,870 possible protein pairs between the 44,140 proteins studied. The results were then sorted by their predicted interaction score and only the top scoring pairs were considered as potentially interacting. More than 250,000 soybean PPIs had high specificity. Here we have performed the first semi-comprehensive functional genomics analysis of soybean through MP-PIPE. While a sequence analysis of the 75 putative candidates and soybean and *Arabidopsis* flowering time genes might also have been used to try to identify candidate genes for *E10*, PIPE was developed for soybean since the predicted PPIs can be used in the future to address questions beyond our immediate identification of a candidate gene for *E10*.

In order to identify the potential candidate gene corresponding to *E10* from the approximately 75 genes in the *E10* chromosomal region, PPIs were predicted for all of the genes located in the *E10* region. Gene ontology (GO) analyses were then performed on the basis of interacting partners; that means predicting whether or not genes in the *E10* region have interactions with genes involved in flowering and maturity. GO analysis represents a systematic

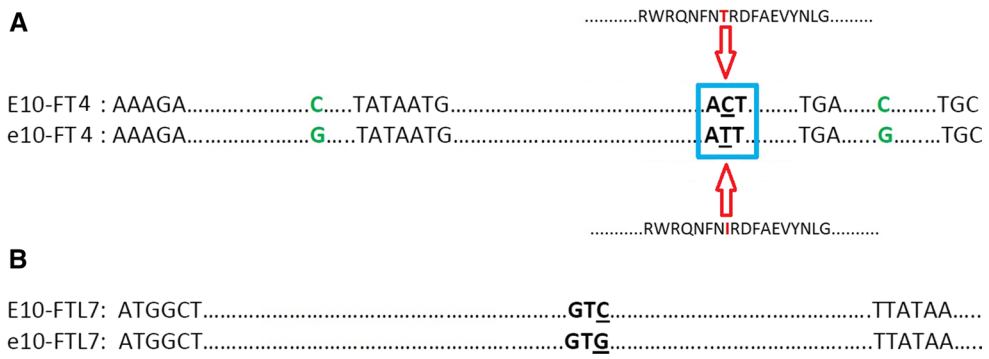

Fig. 2 Representative sequencing data for the soybean FT4 (*panel* **a**) and FTL7 genes (*panel* **b**) for the lines presumed to be *E10* (OT02-18, Harosoy, OT94-47, X5289, and X5291) and *e10* (X5285, X5288 and X5495). **a** Showing three SNPs: at the 5′UTR (position 33), within the 4th exon (position 126) and the 3′UTR (position 56). **b** Showing one SNP in the coding region (position 176 in fourth exon)

investigation of the functional property of a gene or group of selected genes; in other words, if a given gene of interest interacts with groups of genes which are involved in a specific pathway, that gene of interest might also be involved in the same particular pathway (guilt by association). The complete list of genes investigated in the *E10* chromosomal region is presented in supplementary file 2.

As presented in Fig. 1, FT4, also known as FTL8, (Glyma08g47810, GenBank ID: FJ573234.1; Flowering Locus T-like Protein 8) has the most interacting partners known to be involved in flowering and maturity (19 genes, Gene ontology annotation *P* value ≤0.05); however, FTL7 (Glyma08g47820, GenBank ID: FJ573233.1; Heading Date 3A-like) also has interacting partners involved in maturity and flowering, fewer than FT4 but still significant compared to other genes in this region (8 genes, Gene ontology annotation *P* value ≤ 0.05). In order to confirm the accuracy of the data and the statistical analysis (*P* value ≤0.05), both a positive control (E3 maturity gene, PhyA3 gene Glyma19g41210) and a random gene control (Glyma08g47920) were included in the analysis. In contrast with FT4 and FTL7, the rest of the genes located in this chromosomal region had either no interactions or a very limited number of interactions (statistically not significant, *P* value ≥0.05, in terms of the GO analysis) with any genes confirmed to be involved in flowering and maturity. A list of interacting partners predicted by PIPE for FT4, FTL7, and the positive and negative (random gene) controls are presented in supplementary file 3.

Investigating FT4 and FTL7 as candidates for *E10*

Many pathways including flowering and maturity have been investigated in more depth in the model plant *Arabidopsis thaliana.* There is considerable conserved homology between *Arabidopsis* and soybean so therefore the *Arabidopsis* flowering pathway can be used as a guide to help better understand the flowering and maturity pathway in soybean (Zhai et al. 2014; Wang et al. 2015; Wickland and Hanzawa 2015).

To confirm the PPI data generated by PIPE for the *E10* candidate gene FT4, a series of functional genomics analyses were performed using the *Arabidopsis* conserved homolog of FT4 (Q9SXZ2) through GeneMANIA and *Arabidopsis*-PIPE (using PIPE to investigate PPI at specificity of 99.95% and sensitivity of 42.65%; *Arabidopsis*-PIPE, data is not publicly available). Using GeneMANIA (http://www.genemania.org/), an online gene ontology clustering software, and *Arabidopsis*-PIPE, the conserved homolog of FT4 in *Arabidopsis* (Q9SXZ2) has been investigated independently for its PPIs with genes involved in the maturity and flowering network of *Arabidopsis*. This investigation showed that the *Arabidopsis* homolog of soybean-FT4 plays an important role (making connections between different maturity genes including E1, E2, E3, E4 and Dt1 (O23659, Q9SYL8, Q9SQI2, P14712 and P93003), *P*-value ≤ 0.01) in flowering and maturity PPI pathways which is consistent with Soybean-PIPE data and GeneMANIA functional genomics investigations (Gene ontology annotation *P*-value: 1.78E−04) and is in accordance with Wickland and Hanzawa (2015).

PIPE data also suggest that FTL7 and FT4 interact with themselves and each other most likely by making a protein complex which interacts with other maturity and flowering genes. These data are consistent with our *Arabidopsis* functional genomics investigations indicating self-interaction of the conserved homolog of FT4 (Q9SXZ2) in *Arabidopsis* (GeneMANIA and *Arabidopsis*-PIPE).

To investigate the nature of allelic variation of FTL7 and FT4, these two genes were sequenced from soybean lines that are presumed *E10* [OT02-18, Harosoy, OT94-47, (X5289, and X5291 both are *E10* BC_2 lines)] and *e10* [(X5285, X5288 both are *e10* BC_2 lines) and X5495 an *e10* BC_3 line].

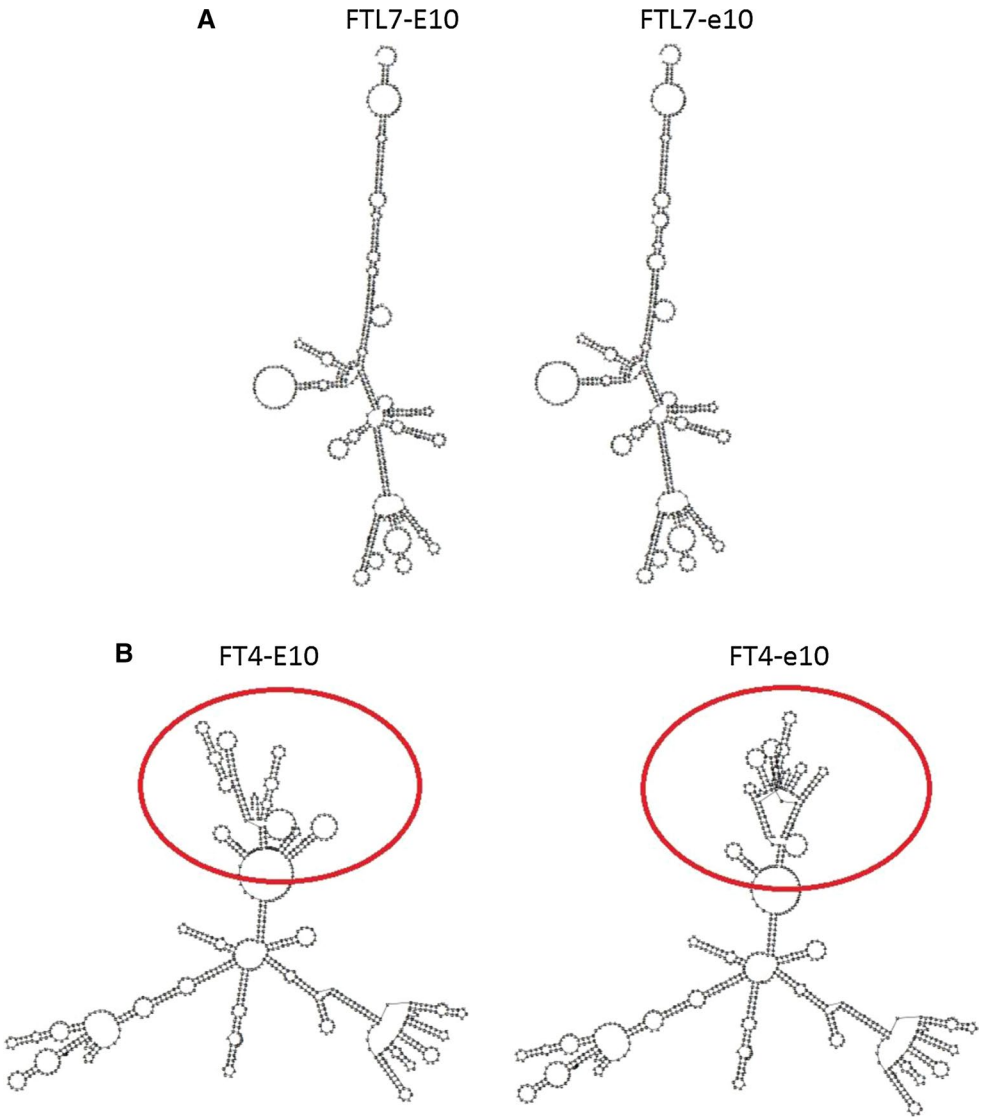

Fig. 3 Two-dimensional (2D) structure of mRNA in *E10* and *e10* soybean lines (this study). **a** 2D mRNA structure of FTL7 in *E10* and *e10* lines. **b** 2D mRNA structure of FT4 in *E10* and *e10* lines

Sequencing data of FT4 identified some SNPs between *E10*, GenBank ID: KX138527, and *e10*, GenBank ID: KX138528, alleles within introns but more interestingly also identified a single SNP in each of the 5′untranslated region, UTR (GM08:44608620), the 3′UTR (GM08:46607056) and the fourth exon of the coding region. The point mutation in the fourth exon (nucleotides ACT in *E10* lines to ATT in *e10* lines, position 126, GM08:46607209) results in a threonine to isoleucine (T 140 I) amino acid substitution. This amino acid substitution leads to different mRNA structure and protein folding (representative sequencing data are presented in Fig. 2a). A threonine to isoleucine substitution has been reported to be involved in protein function modification in several studies (Pepys et al. 1993; Baraas et al. 2012). More interestingly, this amino acid substitution between *E10* and *e10* lines is located very close to the fourth exon external loop found in FT and FTL (FT-Like) genes, which is conserved across species (Wickland and Hanzawa 2015) and therefore can be expected to influence protein three-dimensional structure. Like FT4, sequencing data for FTL7 (Fig. 2b) also showed some SNPs between alleles from *E10* and *e10* lines within introns and one SNP in an exon. However, the latter SNP (GTC in *E10* to GTG in *e10*, position 176 in forth exon, GM08:46616083) is a silent mutation since both GTG and GTC code for the amino acid valine. SNPs could have an effect on RNA structures, which can play important roles in gene regulation. Therefore, the effects of SNPs between *E10* and *e10* lines on two-dimensional (2D) structures of mRNA for FT4 and FTL7 have been investigated. As presented in Fig. 3a, the occurrence of SNPs in the FTL7 sequence is predicted to have no significant effect on the 2D structure of the mRNA for lines carrying *E10* or *e10*. In contrast, SNPs in FT4 mRNA sequences are predicted to cause differential 2D mRNA folding for the lines carrying *E10* and *e10* alleles (Fig. 3b).

Confirmation of markers for *E10*

In order to provide both genetic knowledge and diagnostic markers for breeders to accelerate their breeding for early maturity soybeans, specific markers have been developed for *E10* and *e10* alleles. Generally speaking, three methods have been used for allele-specific marker development including: PCR based, PCR-sequencing based, and PCR-enzymatic digestion based analysis as described by Liu et al. (2008), Watanabe et al. (2009, 2011), Xu et al. (2013), and Tsubokura et al. (2014). As presented in supplementary file 1, three sets of diagnostic primers have been designed independently for the three SNPs identified in FT4 for *E10* and *e10* in our germplasm using the PCR-sequencing based method. In order to confirm the efficiency of these primers and to reconfirm the BC_3 data for *E10* and *e10* alleles, 47 BC_5 samples (early maturity *e10* introgressed into a Harosoy background as well as later maturity *E10* introgressed into a Maple Presto background) were used for allele-specific marker assessment using the PCR-sequence based technique. As presented in Table 4, allele-specific markers designed for the *E10* locus showed perfect correlation with the maturity field data and previously sequenced lines. Note that the microsatellites shown in Table 3 can also be used as breeding friendly markers to assist breeding programmes.

Discussion

OT98-17 (Maple Presto isoline) and OT02-18 (Harosoy isoline) have the same genotype at the major known maturity loci (e1e1 e2e2 e3e3 e4e4 e7e7 e8e8) and yet OT98-17 matured 6 days earlier than OT02-18 in the field (Table 1). The ability to introgress earlier maturity into OT02-18 and later maturity into OT98-17 (Table 2) simply by selecting un-replicated F_3 progeny rows as parents suggested that OT98-17 carries an additional, as yet unidentified, early maturity gene, at a locus labelled *E10*. A novel early maturity allele at a new locus could have significant potential to assist in breeding for early maturity soybeans for northern and western short-season regions of Canada and globally.

While both FT4 and FTL7 may be candidates for *E10*, we suggest that there is a much stronger case for FT4. First, a higher number of flowering time genes were predicted by PIPE to interact with FT4 compared to FTL7. Second, there was greater allelic sequence variation for FT4 compared to FTL7 when comparing late maturity versus early maturity mRNA 2D structure, and also when comparing amino acid substitutions. We have demonstrated a maturity phenotype for alleles at *E10* and propose that FT4 is the candidate gene for *E10*. FT4 has been shown to be expressed in parallel with E1 and was upregulated by long days, resulting in later flowering, and down regulated by short days, resulting in earlier flowering (Zhai et al. 2014). As a result FT4 was proposed to function downstream of E1 and to function as a repressor in soybean flowering time networks (Zhai et al. 2014; Xu et al. 2015). We have found allelic variation for FT4 in OT98-17 versus all other lines studied.

We surveyed the SNP database of approximately 300 short season Canadian lines in Table 5 (Sonah et al. 2013, 2015) for the OT98-17 *e10* haplotype at the six SNPs in the region of FT4 and FTL7 (149 kb). Six haplotypes were identified at these six SNPs; however, none match the OT98-17 *e10e10* haplotype. We hypothesize that the *e10* early maturity allele is rare in Canadian breeding germplasm.

The pedigree of OT98-17 is X824A-ve/7*Maple Presto, where X824A-ve is PI 438477/2*'Evans'/2/L62-667. Maple Presto's pedigree is 'Amsoy'/'Portage'/2/PI 438477 (Cober et al. 2010). PI 438477 is in the pedigree of both parents (X824A-ve and Maple Presto) of OT98-17. PI 438477 is also known as 840-7-3 and as Holmberg 840-7-3. It is a maturity group (MG) 00, very early maturing line from Sweden (about 60 degrees north) and a potential donor of *e10*. Portage is also MG 00 while Amsoy is MG II. In an effort to identify the source of the novel *e10* allele, five ancestral lines (Amsoy 71, Portage, PI438477, Evans and L62-667) identified in the pedigree of OT98-17 were genotyped with the four diagnostic SSR markers (Table 3). None, however, displayed alleles identical with OT98-17. Identical alleles would be expected if any of the five ancestral lines was the donor of *e10*, hence we cannot identify the ancestral donor of the *e10* allele in OT98-17 and the *e10* source may have been transferred via insect cross pollination from other lines in the nursery. Selection for early maturity in the development of OT98-17 possibly resulted in concurrent selection for *e10* in OT98-17.

In conclusion, we identified phenotypic differences in maturity between parents, and also in backcross lines selected for early or late maturity in Ottawa, Canada. *E10E10* results in late, and *e10e10* results in early maturity. The *E10* gene symbol has been approved by Soybean Genetics Committee. SSR and SNPs analysis identified a region on chromosome Gm08 (Linkage Group A2) correlating with this new maturity locus. Using soybean functional genomics (through PPI analysis using PIPE) followed by molecular biology techniques including sequencing, SNP analysis, and 2D RNA structure investigations, we have identified FT4 to be the candidate gene for the novel maturity locus, *E10*. Due to the large proteome size of soybean (75,778 confirmed soybean protein sequences are reported in Soybase http://soybase.org/) and the computational limits of the PIPE, an all-to-all analysis was not possible. In dealing with this limitation, we performed a partial run of PIPE using a subset of soybean data. Please note that

Table 4 Alleles at three SNPs differentiating the two soybean FT4 allelic forms

Line	Maturity Late/early	Background	Days to maturity	5′ UTR C/G	Coding region C/T	3′UTR C/G	Genotyping
X5905-42	Early	Har[a]	96	G	T	G	*e10*
X5908-61	Early	Har	94	G	T	G	*e10*
X5908-64	Early	Har	94	G	T	G	*e10*
X5908-86	Early	Har	96	G	T	G	*e10*
X5908-101	Early	Har	95	G	T	G	*e10*
X5911-33	Early	Har	96	G	T	G	*e10*
X5911-97	Early	Har	95	G	T	G	*e10*
X5911-103	Early	Har	96	G	T	G	*e10*
X5911-104	Early	Har	96	G	T	G	*e10*
X5912-2	Early	Har	95	G	T	G	*e10*
X5912-21	Early	Har	94	G	T	G	*e10*
X5912-25	Early	Har	95	G	T	G	*e10*
X5905-13	Late	Har	110	C	C	C	*E10*
X5905-52	Late	Har	109	C	C	C	*E10*
X5905-81	Late	Har	110	C	C	C	*E10*
X5905-87	Late	Har	110	C	C	C	*E10*
X5905-88	Late	Har	109	C	C	C	*E10*
X5908-6	Late	Har	110	C	C	C	*E10*
X5908-7	Late	Har	110	C	C	C	*E10*
X5908-10	Late	Har	109	C	C	C	*E10*
X5908-14	Late	Har	110	C	C	C	*E10*
X5908-30	Late	Har	109	C	C	C	*E10*
X5911-23	Late	Har	110	C	C	C	*E10*
X5915-4	Early	MP[b]	94	G	T	G	*e10*
X5915-50	Early	MP	94	G	T	G	*e10*
X5915-51	Early	MP	94	G	T	G	*e10*
X5915-54	Early	MP	94	G	T	G	*e10*
X5915-80	Early	MP	94	G	T	G	*e10*
X5918-23	Early	MP	95	G	T	G	*e10*
X5918-25	Early	MP	95	G	T	G	*e10*
X5918-62	Early	MP	95	G	T	G	*e10*
X5918-93	Early	MP	95	G	T	G	*e10*
X5918-101	Early	MP	94	G	T	G	*e10*
X5921-12	Early	MP	96	G	T	G	*e10*
X5921-14	Early	MP	96	G	T	G	*e10*
X5921-39	Early	MP	96	G	T	G	*e10*
X5915-103	Late	MP	105	C	C	C	*E10*
X5915-104	Late	MP	103	C	C	C	*E10*
X5918-36	Late	MP	103	C	C	C	*E10*
X5918-67	Late	MP	103	C	C	C	*E10*
X5918-79	Late	MP	103	C	C	C	*E10*
X5918-80	Late	MP	103	C	C	C	*E10*
X5918-87	Late	MP	103	C	C	C	*E10*
X5918-90	Late	MP	103	C	C	C	*E10*
X5918-96	Late	MP	103	C	C	C	*E10*
X5921-30	Late	MP	103	C	C	C	*E10*
X5921-51	Late	MP	103	C	C	C	*E10*

These function as allele-specific markers for the *E10* locus in backcross 5 (BC5) lines resulting from the introgression of late maturity into OT98-17 and early maturity into OT02-18. Also shown is the corresponding maturity phenotypes observed in the field at Ottawa in 2015

[a] Represents Harosoy

[b] Represents Maple Presto

Table 5 Allele frequency at SNPs flanking the FT4 locus among 300 Canadian soybean lines Re-analysis of SNP data reported by Sonah et al. (2013, 2015)

Chromosomal position of SNP	*E10* allele	*e10* allele	No. of lines with *e10* allele
Gm08:46528223	A	G	0
Gm08:46530935	T	A	1
Gm08:46530994	A	G	2
Gm08:46531017	T	G	10
Gm08:46631888	T	C	16
Gm08:46678423	T	C	44

higher capacity version is under development (our collaborators at Carleton University) and will be extremely useful for future studies. A comprehensive (all-protein-to-all-protein) functional genomics investigation of soybean through PIPE will give breeders a unique opportunity to understand interactions in the soybean proteome. SNPs were identified in sequences in the candidate FT4 between *E10* and *e10* alleles, one of which results in an amino acid substitution near an FT/FTL region which is conserved across species (Wickland and Hanzawa 2015). These SNPs also result in a predicted 2D conformation change in mRNA.

Threonine to isoleucine substitution has been reported to be involved in protein function modification in several studies (Pepys et al. 1993; Baraas et al. 2012). More interestingly, this amino acid substitution between *E10* and *e10* lines is located in the 4th exon very close to the external conserved loop, found in FT and FTL (FT-Like) genes, which is conserved across species (Wickland and Hanzawa 2015). Additionally, this amino acid substitution is close to the conserved position reported by Wang et al. (2015) for the FT gene family. As reported by Wang et al. (2015), other amino acid residues in addition to conserved locations (such as H88Y) play an important role in regulation of flowering. It has been reported that a tyrosine to leucine or isoleucine amino acid substitution could lead to flowering enhancement. By considering the importance of the 4th exon, amino acid substitution (T 140 I), SNPs located at 3′ and 5′ UTRs, bioinformatics data and phenotypic analysis on time of flowering and maturity, we believe that FT4 is a potential candidate for the gene underlying this maturity locus.

Identification of molecular markers tagging genes controlling flowering time and maturity in soybean, a prerequisite for molecular marker assisted breeding, will assist soybean breeders to resourcefully develop new early maturing varieties. Allele-specific markers will allow stacking of early maturity alleles to develop even earlier maturing cultivars. Here we have identified FT4 as a candidate for *E10* and developed sequence-based allele-specific markers for the *E10* maturity locus.

Author contribution statement E.C. developed the genetic populations and phenotyped the material. S.M. and M.C. conducted the SSR genotyping. B.S., A.S., F.D., and A.G. developed Soy-PIPE and carried out the PIPE analysis. F.B. conducted the SNP genotyping. B.S. conducted sequencing and bioinformatics analyses. B.S., S.M., and E.C. wrote the paper.

Acknowledgements This work was funded in part by the Canadian Field Crop Research Alliance (CFCRA) and Agriculture and Agri-Food Canada through the *Growing Forward 2* Agri-Innovation Program and also by the Natural Sciences and Engineering Council of Canada (NSERC).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Baraas RC, Hagen LA, Dees EW, Neitz M (2012) Substitution of isoleucine for threonine at position 190 of S-opsin causes S-cone-function abnormalities. Vision Res 73:1–9

Bernard RL (1971) Two major genes for time of flowering and maturity in soybeans. Crop Sci 11:242–247

Bonato ER, Vello NA (1999) E6 a dominant gene conditioning early flowering and maturity in soybeans. Genet Mol Biol 22:229–232

Buzzel RI, Voldeng HD (1980) Inheritance of insensitivity to long day length. Soybean Genet Newsl 7:26–29

Buzzell RI (1971) Inheritance of a soybean flowering response to fluorescent day length conditions. Can J Genet Cytol 13:703–707

Cober ER, Morrison MJ (2010) Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. Theor Appl Genet 120:1005–1012

Cober ER, Voldeng HD (2012) A retrospective look at short-season soybean cultivar development in Ontario. Can J Plant Sci 92:1239–1243

Cober ER, Molnar SJ, Charette M, Voldeng HD (2010) A new locus for early maturity in soybean. Crop Sci 50:524–527

Cober ER, Curtis DF, Stewart DW, Morrison MJ (2014) Quantifying the effects of photoperiod, temperature and daily irradiance on flowering time of soybean isolines. Plants 3:476–497

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, Van Toai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature 340:245–246

Garner WW, Allard HA (1927) Effect of short alternating periods of light and darkness on plant growth. Science 66:40–42

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. Proc Natl Acad Sci USA (PNAS) 97:4168–4173

Jessulat M, Pitre S, Gui Y, Hooshyar M, Omidi K, Samanfar B, Tan LH, Alamgir M, Green J, Dehne F, Golshani A (2011) Recent advances in protein-protein interaction prediction: experimental and computational methods. Expert Opin Drug Discov 6:921–935

Kong F, Nan H, Cao D, Li Y, Wu F, Wang J, Lu S, Yuan X, Cober ER, Abe J, Liu B (2014) A new dominant gene E9 conditions early flowering and maturity in soybean. Crop Sci 54:2529–2535

Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J (2008) Genetic redundancy in soybean photo responses associated with duplication of the phytochrome A gene. Genetics 180:995–1007

Liu B, Watanabe S, Uchiyama T, Kong F, Kanazawa A, Xia Z, Nagamatsu A, Arai M, Yamada T, Kitamura K, Masuta C, Harada K, Abe J (2010) The soybean stem growth habit gene Dt1 is an ortholog of Arabidopsis TERMINAL FLOWER 1. Plant Physiol 153:198–210

McBlain BA, Bernard RL (1987) A new gene affecting the time of flowering-maturity in soybeans. J Hered 178:68–70

McBlain BA, Hesketh JD, Bernard RL (1987) Genetic effects on reproductive phenology in soybean isolines differing in maturity genes. Can J Plant Sci 67:105–116

Molnar SJ, Rai S, Charette M, Cober ER (2003) Simple sequence repeat (SSR) markers linked to E1, E3, E4, and E7 maturity genes in soybean. Genome 46:1024–1036

Omidi K, Hooshyar M, Jessulat M, Samanfar B, Sanders M, Burnside D, Pitre S, Schoenrock A, Xu J, Babu M, Golshani A (2014) Phosphatase complex Pph3/Psy2 is involved in regulation of efficient non-homologous end-joining pathway in the yeast Saccharomyces cerevisiae. PLoS One 9(1):e87248

Park Y (2009) Critical assessment of sequence-based protein-protein interaction prediction methods that do not require homologous protein sequences. BMC Bioinform 10:419

Pepys MB, Hawkins PN, Booth DR, Vigushin DM, Tennent GA, Soutar AK, Totty N, Nguyen O, Blake CC, Terry CJ (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis. Nature 362:553–557

Pitre S, Dehne F, Chan A, Cheetham J, Duong A, Emili A, Gebbia M, Greenblatt J, Jessulat M, Krogan N, Luo X, Golshani A (2006) PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. BMC Bioinform 7:365

Pitre S, Alamgir M, Green JR, Dumontier M, Dehne F, Golshani A (2008) Computational methods for predicting protein-protein interactions. Adv Biochem Eng/Biotechnol 110:247–267

Pitre S, Hooshyar M, Schoenrock A, Samanfar B, Jessulat M, Green JR, Dehne F, Golshani A (2012) Short co-occurring polypeptide regions can predict global protein interaction maps. Sci Rep 2:239

Ray JD, Hinson K, Mankono JEB, Malo MF (1995) Genetic control of a long-juvenile trait in soybean. Crop Sci 35:1001–1006

Saindon G, Beversdorf WD, Voldeng HD (1989a) Adjusting of the soybean phenology using the E4 loci. Crop Sci 29:1361–1365

Saindon G, Voldeng HD, Beversdorf WD, Buzzell RI (1989b) Genetic control of long daylength response in soybean. Crop Sci 29:1436–1439

Schoenrock A, Samanfar B, Pitre S, Hooshyar M, Jin K, Phillips CA, Wang H, Phanse S, Omidi K, Gui Y, Alamgir M, Wong A, Barrenäs F, Babu M, Benson M, Langston MA, Green JR, Dehne F, Golshani A (2014) Efficient prediction of human protein-protein interactions at a global scale. BMC Bioinform 15:383

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau É, Laroche J, Larose S, Jean M, Belzile F (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One 8(1):e54603

Sonah H, O'Donoughue L, Cober ER, Rajcan I, Belzile F (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol J 13:211–221

Tian Z, Wang X, Lee R, Li Y, Specht JE, Nelson RL, McClean PE, Qiu L, Ma J (2010) Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci (PNAS) 107:8563–8568

Tsubokura Y, Watanabe S, Xia Z, Kanamori H, Yamagata H, Kaga A, Katayose Y, Abe J, Ishimoto M, Harada K (2014) Natural variation in the genes responsible for maturity loci E1, E2, E3 and E4 in soybean. Ann Bot 113:429–441

Vlasblom J, Zuberi K, Rodriguez H, Arnold R, Gagarinova A, Deineko V, Kumar A, Leung E, Rizzolo K, Samanfar B, Chang L, Phanse S, Golshani A, Greenblatt JF, Houry WA, Emili A, Morris Q, Bader G, Babu M (2015) Novel function discovery with GeneMANIA: a new integrated resource for gene function prediction in Escherichia coli. Bioinformatics 31:306–310

Wang Z, Zhou Z, Liu Y, Liu T, Li Q, Ji Y, Li C, Fang C, Wang M, Wu M, Shen Y, Tang T, Ma J, Tian Z (2015) Functional evolution of phosphatidylethanolamine binding proteins in soybean and *Arabidopsis*. Plant Cell 27:323–336

Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T, Tabata S, Harada K (2009) Map-based cloning of the gene associated with the soybean maturity locus E3. Genetics 182:1251–1262

Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, Harada K (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genetics 188:395–407

Wickland DP, Hanzawa Y (2015) The Flowering Locus T/Terminal Flower 1 gene family: functional evolution and molecular mechanisms. Mol Plant 8:983–997

Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lü S, Wu H, Tabata S, Harada K (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proc Natl Acad Sci USA 109:E2155–E2164

Xu M, Xu Z, Liu B, Kong F, Tsubokura Y, Watanabe S, Xia Z, Harada K, Kanazawa A, Yamada T, Abe J (2013) Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biol 13:91

Xu M, Yamagishi N, Zhao C, Takeshima R, Kasai M, Watanabe S, Kanazawa A, Yoshikawa N, Liu B, Yamada T, Abe J (2015) The soybean-specific maturity gene E1 family of floral repressors controls night-break responses through down-regulation of flowering locus t orthologues. Plant Physiol 168:1735–1746

Zhai H, Lü S, Liang S, Wu H, Zhang X, Liu B, Kong F, Yuan X, Li J, Xia Z (2014) GmFT4, a homolog of FLOWERING LOCUS T, is positively regulated by E1 and functions as a flowering repressor in soybean. PLoS One 9(2):e89030

Zhao C, Takeshima R, Zhu J, Xu M, Sato M, Watanabe S, Kanazawa A, Liu B, Kong F, Yamada T, Abe J (2016) A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a flowering locus T orthologue. BMC Plant Biol 16:20