Genetic Analyses of Important Agronomic Traits in Soybean by Using Wild Soybean Chromosome Segment Substitution Lines

Genetic Analyses of Important Agronomic Traits in Soybean by Using Wild Soybean Chromosome Segment Substitution Lines

July 2018

Dequan LIU

Genetic Analyses of Important Agronomic Traits in Soybean by Using Wild Soybean Chromosome Segment Substitution Lines

A Dissertation Submitted to

the Graduate School of Life and Environmental Sciences,

the University of Tsukuba

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Agricultural Science

(Doctoral Program in Biosphere Resource Science and Technology)

Dequan LIU

Table of Contents

Abbreviations

AB-QTL	Advanced backcrossing Quantitative trait locus
ADD	Estimated additive effect of a QTL
ANOVA	Analysis of variance
cDNA	Complementary DNA
CHS	Chalcone synthase
cM	Centimorgan
CSSLs	Chromosome segment substitution lines
CTAB	Cetyl trimethylammonium bromide
DAE	Days after emergence
DHR	Donor homozygous allele
EDTA	Ethylenediaminetetraacetic acid
FPKM	Fragments per kilobase million
JAZ	Jasmonate-ZIM domain
LG	Linkage group
LOD	Logarithm (base 10) of odds score calculated from additive QTL with chromosome segment substitution
M (QQ)	Mean value of the QTL genotype of donor parent
M (qq)	Mean value of the QTL genotype of recurrent parent
NILs	Near-isogenic lines
PCR	Polymerase chain reaction
PRR	Pseudo response regulators
PVE (%)	Phenotypic variation explained by a QTL
qRT-PCR	Quantitative real-time polymerase chain reaction
QTL	Quantitative trait locus
RAD-seq	Restriction site associated DNA sequence

RHL	Residual heterozygous line
RILs	Recombinant inbred lines
SAM	Shoot apical meristem
SCC	Seed coat cracking
SD	Standard deviation
SSR	Simple sequence repeats
TBE	Tris borate EDTA
TFL	Terminal flower gene
TOE	Target of eat (delay of flowering time protein)

List of Tables

List of Figures

Figure 5. 5 Frequency distribution of seed coat cracking (SCC) of the ‘Tachiyutaka’ × ‘JWS156-1’ BC_3F_5 population in 2017.

CHAPTER 1

General introduction

1.1 Wild soybean (*Glycine soja* Sieb. & Zucc.)

The genus *Glycine* subgenus *Soja* includes two species: cultivated soybean *(Glycine max* (L.) Merr.) and wild soybean (*Glycine soja* Sieb. & Zucc.). Wild soybean is believed to be the progenitor of cultivated soybean, which was domesticated in China (Hymowitz 1970). Wild soybean accessions are mainly distributed in East Asia, such as China, Korea, and Japan. Wild soybean has a higher genetic diversity than cultivated soybean (Hyten *et al*. 2006; Lam *et al*. 2010; Li *et al*. 2010; Xu *et al*. 2002; Zhou *et al*. 2015a). There are big differences between wild and cultivated soybeans with respect to morphological and physiological characteristics as the result of divergent natural and artificial selection (Liu *et al*. 2007). In the process of wild soybeans being domesticated into cultivated soybeans, the morphological and physiological traits were changed by human selection and adaptation different planting environment (Harlan 1992). However, in this process, a large of genetic diversity has been lost. The previous study suggested that approximately half of the genetic diversity has been lost during soybean domestication from wild soybean to landraces (Zhou *et al*. 2015a). Most of the domestication-related traits in soybeans were conditioned by one or two major QTLs and some genotype-dependent minor QTLs, and thus introgression of useful genes from wild to cultivated soybeans can be carried out without large obstacles. (Liu *et al.* 2007). It is believed that the wild soybean is still rich in novel and/or useful variants for soybean breeding in various traits. Therefore, understanding the genetic basis underlying domestication-related traits is important for improving yield, adaptability for different cultivation of latitude, and seed

quality in soybean.

Several quantitative trait loci (QTLs) or genes have been identified in wild soybean and introduced into cultivated soybean (Concibido *et al*. 2003; Do *et al*. 2016; Hu *et al*. 2013; Sebolt *et al*. 2000; Tuyen *et al*. 2010; Wang *et al*. 2016b).A salt tolerance *Ncl* was identified from a salt-tolerant wild soybean accession and was used to develop an improved salt tolerance line. *Ncl* could increase the grain yield of soybean by 3.5-5.5 times in saline field conditions (Do *et al.* 2016); Two novel aphid resistance genes, *Rag6* and *Rag3*c were identified and validated on chromosome 8 and chromosome 16 from a wild soybean (*Glycine soja* 85-32), respectively (Zhang *et al*. 2017a; Zhang *et al*. 2017b). *Rag6* has shown strong resistance, while *Rag3c* has shown moderate resistance over the years; In addition, as soybean weight/size, which is an important agronomic trait for soybean yield, a greater seed weight gene (*PP2C-1*) from a wild soybean was isolated and was used to improve seed weight in cultivated soybean (Lu *et al.* 2017b). In addition, wild soybean was also used to identify QTLs for yield, and a positive allele from wild soybean has been detected (Concibido *et al*. 2003; Wang *et al*. 2004). These findings suggest that wild soybean is an important genetic resource pool for the improvement of cultivated soybean.

In my study, I used a Japanese wild accession, ‘JWS156-1’, from the Kinki area of Japan. This wild accession has been demonstrated to have high salt tolerance (Hamwieh and Xu 2008) and alkali salt tolerance (Tuyen *et al* 2010; Xu *et al*. 2012). The wild soybean accession also has some special morphological traits, such as it has short internode length. These characteristics make ‘JWS156-1’ a special wild soybean germplasm.

1.2 Chromosome segment substitution lines (CSSLs)

Most agronomic trait QTLs have been detected using biparental mapping populations, such as recombinant inbred lines (RILs), which are particularly effective for detecting QTLs with large effects. Recently, a more effective strategy known as an advanced backcross QTL (AB-QTL) analysis (Grandillo and Tanksley 2005) has been used to detect QTLs associated with traits of interest. Through continuous backcrossing followed by successive self-pollination, different genomic regions from the donor parent are introgressed into the genetic background of the recurrent parent to produce chromosome segment substitution lines (CSSLs). In CSSLs, more than 90% of the genome was recovered by the recurrent parent after three times of backcrossing; each CSSL carries one or more donor parent segments in the genetic background. The main characteristic of CSSLs is that each CSSL has the similar genetic background, which can reduce genetic background noise and improve the accuracy of gene mapping (Grandillo and Tanksley 2005). Since CSSLs have similar genetic background, once a particular QTL allele of interest is identified in a CSSL, near-isogenic lines (NILs) for the trait can be easily developed fine mapping and positional cloning of the QTL. Until now, many CSSL populations have been used in genetic studies of different crops. In tomato, the fruit size trait QTL were identified using CSSLs (Eshed and Zamir 1995). Two major loci affecting flag leaf size with functional candidate genes were identified and validated in a rice CSSL population (Tang *et al*. 2018). Wang and his colleagues used two CSSL populations, which were examined over 2 years in two locations, and revealed that overdominance was the main contributor to effects of heterosis on grain yield in maize (Wang *et al*. 2016a). These studies demonstrated that CSSL was powerful for detecting QTLs.

In soybean, several CSSL mapping populations, including two wild-soybean and one cultivated soybean CSSL populations, have been created and used for mapping QTLs controlling agronomic traits (Wang *et al*. 2013; Watanabe *et al.* 2018; Xin *et al*. 2016). Wang and colleagues (2013) developed a CSSL population from a cross between a Chinese wild soybean accession 'N24852' and a Chinese elite soybean cultivar 'NN1138-2', and detected three plant height QTLs and one number of nodes on main stem QTLs. Another wild soybean CSSL population was also developed by using a Chinses wild soybean accession 'ZYD00006' (Xin *et al*. 2016), which was employed for detecting QTLs controlling soybean seed weight. Recently, a high-density CSSLs was developed for a cross between Japanese soybean 'Enrei' and Chinese soybean 'Peking', and about 50 reproducible QTLs related to flowering, maturity, and yield were detected (Watanabe *et al.* 2018).

No CSSL population constructed with a Japanese wild soybean has been reported in spite of the fact that Japan has been demonstrated as an important soybean gene pool (Abe *et al.* 2003; Xu *et al.* 2002). In the present study, a CSSL population was developed from crosses between the cultivated soybean variety 'Jackson', which is a famous US soybean variety, and the wild soybean accession 'JWS156-1'. The developed CSSLs were employed for identifying QTLs of some important agronomic traits related with grain yield, variety adaptability, and grain quality. For grain yield, I have focused on seed weight QTLs. This was mainly due to the fact that there was about 10 times of difference between wild and cultivated soybeans and had a high possibility to discover new flowering time QTL or gene. For variety adaptability, I analyzed flowering time QTLs because wild soybean was more sensitive to photoperiod, and flowering time was regarded as a domestication trait (Liu *et al*. 2007). As to the grain quality, I studied seed

coat cracking since almost all wild soybean had no seed coat cracking phenomena.

1.3 100-seed weight (100SW)

The seed weight, generally expressed as 100-seed weight (100SW), is one of the most important yield components in soybean. Several studies have shown a positive association between seed weight/size and seed yield (Burris *et al.* 1973; Smith and Camper 1975). Seed weight is also reported to be associated with seed germination and vigor (Edwards and Hartwing 1971). The soybean varieties developed in the tropical or subtropical countries, such as India and Indonesia, tend to have a small seed size compared with varieties from the temperate regions, such as USA, China, and Japan. In addition, seed size is also an important factor for the end-use of soybeans. While a small seed size is preferred for natto and soy sprouts, a large seed size is preferred for vegetable soybean and tofu. Therefore, depending on their end-use or location of growth, different soybean varieties with different seed sizes have been established.

In the past decades, many studies have focused on genetic studies of seed weight/size in the soybean, particularly using the DNA marker for QTL analyses. To date, many QTLs of seed weight/size on all the 20 chromosomes have been reported in SoyBase (http://www.soybase.org/). However, due to the nature of multi-gene effects for controlling seed weight/size, most of the QTLs were minor, and not validated. Examples of major seed weight QTLs include one mapped to on chromosome 17, *qSw17-1* (Kato *et al.* 2014), and another, *swHCC2-1*, associated with Satt460 on chromosome 6 (Han *et al.* 2012). Both QTLs showed major and stable effects and a high level of phenotypic variation in multiple populations and environments.

Compared with other major crops, such as rice (Duan *et al.* 2015; Hu *et al.* 2015;

Wang *et al.* 2015), knowledge about the molecular mechanisms of seed weight control in soybean is very limited. To date, only two seed weight/size related genes are isolated from soybean. The *ln* (narrow leaf) gene has a large effect on the number of seeds per pod and seed size (Jeong *et al.* 2012), and recently, the *PP2C-1* (protein phosphatase type-2C) allele was found to enhance the seed size in a wild soybean accession ZYD7 (Lu *et al.* 2017b). Identification of a major and stable QTL for seed weight is very important for facilitating map-based cloning of seed weight gene and understanding of the molecular mechanisms of seed weight control in soybean.

1.4 Flowering time

Soybean is cultivated across the world in a wide range of latitudes extending from 50°N to 22°S. Because soybean is highly sensitive to photoperiod, each cultivar can only be cultivated in a very narrow latitudinal zone. Flowering time is a critical trait affecting soybean adaptability, growth, maturity, and grain yield. Understanding the genetic and molecular mechanisms regulating soybean flowering time is extremely important for soybean breeding.

To date, 11 genes controlling flowering time have been reported in soybean. These include *E1* (Bernard *et al*. 1971), *E2* (Bernard *et al*. 1971), *E3* (Buzzell *et al*. 1971), *E4* (Buzzell and Voldeng 1980), *E5* (McBlain and Bernard 1987), *E6* (Bonato and Vello 1999), *E7* (Cober and Voldeng 2001), *E8* (Cober *et al*. 2010), *E9* (Kong *et al*. 2015), *E10* (Samanfar *et al*. 2017), and *J* (Ray *et al*. 1995). Of these, *E1*, *E2*, *E3*, *E4*, *E9*, and *J* have been cloned, or the causal gene has been identified at the molecular level. *E1*, which is located on chromosome 6, encodes a transcription factor with a putative nuclear localization signal and a B3 DNA-binding domain (Xia *et al*. 2012). *E2* is a co-ortholog

of *Arabidopsis* flowering gene *GIGANTEA* (*GI*) located on chromosome 10 (Watanabe *et al*. 2011). *E3* and *E4* are homologs of *phytochrome A* (*PHYA*) located on chromosomes 19 and 20, respectively (Liu *et al*. 2008; Watanabe *et al*. 2009). *E9* is a FLOWERING LOCUS T ortholog gene (*FT2a*) located on chromosome 16 (Zhao *et al*. 2016). Recently, *J*, located on chromosome 4, was isolated using map-based cloning (Lu *et al*. 2017a; Yue *et al*. 2016). *J* is one of the orthologs of the *Arabidopsis* flowering time gene *ELF3* and associates with the *E1* promoter to suppress its transcription (Lu *et al*. 2017a).

QTL analysis using molecular markers has been widely used to identify flowering time genes in soybean. To date, more than a hundred flowering time QTLs, distributed on 17 chromosomes, have been reported and registered in Soybase (https://soybase.org/sbt/). However, many flowering time QTLs were repeatedly mapped on linkage group C2 (chromosome 6) coinciding with *E1*. Although some flowering time genes are identified and cloned from soybean, knowledge about gene or gene system regulating flowering time is still limited.

1.5 Seed coat cracking (SCC)

Seed coat cracking (SCC) has many negative effects, such as increasing the rate of virus infection, reducing soybean quality and their commercial value (Burchett *et al*. 1985; Duke *et al*. 1986; Yaklich and Barla-Szabo 1993). It has been reported that seed coat cracking results from the separation of epidermal and hypodermal tissues that exposes the underlying parenchyma tissue (Wolf and Baker 1972; Yaklich and Barla-Szabo 1993).

SCC can be genetically classified into two types: Type I with irregular cracks and Type II with net-like cracks (Liu 1949). Type I cracking is controlled by allelic combinations

of *I* and *T* loci (Nicholas *et al*. 1993). In addition, maturity genes also influence cracking. Dominant alleles *E1*, *E5*, *E7* and recessive alleles *e3* and *e4* suppress cracking (Benitez *et al*. 2004; Takahashi and Abe 1999). Oyoo and colleagues (2010a) identified two quantitative trait loci (QTLs) *cr1* on chromosome 2 and *cr2* on chromosome 7 associated with Type I SCC. Type II was not associated with *I* or *T* loci (Liu 1949). The *SoyPRP1* locus for proline-rich cell wall protein on chromosome 9 had a minor effect on the net-like cracking Type II (Nakamura *et al*. 2003). Two major QTLs, *ncr1* and *ncr2*, were located on chromosome 4 with approximately 62 cM and 16.0 and 32.4 % of phenotypic variation, respectively (Oyoo *et al*. 2010b).

Although many studies have focused on identifying QTLs associated with SCC in the past decades, the mechanism of cracking in soybean is unclear. Thus, identification of major and stable QTLs for SCC is very important for understanding of the molecular mechanisms of cracking control in soybean.

1.6 Objectives of this thesis

A serial of studies has demonstrated that wild soybean has a higher genetic diversity than cultivated soybean (Hyten et al. 2006; Lam *et al*. 2010; Li *et al*. 2010; Xu *et al*. 2002; Zhou *et al*. 2015a). Wild soybean thus is an important genetic resource pool for the improvement of cultivated soybean. However, it is not easy to uncover the favorable genes hidden in wild soybean because in the primary mapping population (such as RILs) due to that some wild traits, such as twining growth habit, are remained. These wild traits make accurate evaluation of the agronomic traits very difficult. Wild soybean CCSL population is ideal materials for discovering the favorable genes in wild soybean. To develop such a wild soybean CSSL population, my colleagues initiated the crossing

between a US cultivated soybean variety ‘Jackson’ and a Japanese wild soybean accession ‘JWS156-1’, and then backcrossed with ‘Jackson’ three times. I joined this work in the latter stage of this work and finally completed construction of a BC_3F_5 population with 120 CSSLs.

Wild soybean differs from cultivated soybean in many traits. Using the wild CSSL population, I initiated my research by focusing on the most important traits that affect soybean grain yield, adaptability, and seed quality. The seed size, a grain yield component trait, has about 10 times different between wild and cultivated soybean. Flowering time has been demonstrated an important domestication trait determining the adaptability and influencing soybean grain yield. Seed coat cracking can greatly reduce soybean quality and their commercial value. Since ‘Jackson’ and ‘JWS156-1’ have a large phenotypic difference for the three traits, QTL analyses using the CSSL population derived from these two genotypes can enable us to discover new QTLs or genes and to deepen understanding for the genetic basis of the three important agronomic traits.

The objectives of my study were:

(1) To construct a wild soybean CSSL population in a cultivated soybean background for exploring and utilizing the favorable alleles in wild soybean.

(2) To identify QTLs or candidate genes associated with important agronomic traits of seed weight, flowering time, seed coat cracking using the CSSL population.

(3) To validate the QTLs identified in the CSSL population using other genetic materials.

(4) To understand the genetic basis for the three important traits to facilitate their use in soybean breeding for improving soybean adaptability and productivity.

CHAPTER 2

Construction and characterization of a BC_3F_5 wild soybean chromosome segment substitution line (CSSL) population

2.1 Introduction

Wild soybean is believed to be the progenitor of cultivated soybean, which was domesticated in China (Hymowitz 1970). They are big differences between wild and cultivated soybeans with respect to morphological and physiological characteristics as the result of divergent natural and artificial selection (Liu *et al*. 2007). A serial of studies has demonstrated that wild soybean has a higher genetic diversity than cultivated soybean (Xu *et al.* 2002; Zhou *et al*. 2015a). . Several quantitative trait loci (QTLs) or genes have been identified in wild soybean and introduced into cultivated soybean (Concibido *et al*. 2003; Do *et al*. 2016; Hu *et al*. 2013; Sebolt *et al*. 2000; Tuyen *et al*. 2010; Wang *et al*. 2016b). These studies demonstrated that wild soybean is a potential genetic resource pool for the improvement of cultivated soybean.

To explore and utilize the favorable alleles in wild soybean, a more effective strategy known as an advanced backcross QTL (AB-QTL) analysis (Grandillo and Tanksley 2005) has been used to detect QTLs associated with traits of interest. Through continuous backcrossing followed by successive self-pollination, different genomic regions from the donor parent are introgressed into the genetic background of the recurrent parent to produce chromosome segment substitution lines (CSSLs). Until now, CSSL populations have been used in genetic studies of many crops, such as tomato (Eshed and Zamir 1995), rice (Tang *et al*. 2018), and maize (Wang *et al*. 2016a). In soybean, several CSSL

mapping populations, including two wild soybean CSSL populations, have been created and used for mapping QTLs controlling agronomic traits (Wang *et al*. 2013; Watanabe *et al.* 2018; Xin *et al*. 2016). However, no wild soybean CSSL population constructed with a Japanese wild soybean has been reported.

In the present study, a CSSL population was developed from crosses between a cultivated soybean variety 'Jackson' and a Japanese wild soybean accession 'JWS156-1' for identifying QTLs of important agronomic traits.

2.2 Materials and methods

2.2.1 Plant materials

The parental materials, 'Jackson' and 'JWS156-1' were used for developing the BC_3F_5 wild soybean CSSL population (Figure 2. 1). 'Jackson' is a soybean variety from the US (PI548657). This variety was extensively studied in previous studies (Hesler *et al*. 2012; Hufstetler *et al*. 2006; Nichols *et al*. 2007; Vadez and Sinclair 2002). 'JWS156-1' was selected from a wild soybean accession B06097, which was provided by the National BioResource Project-LegumeBase (https://www.legumebase.brc.miyazaki-u.ac.jp/top.jsp). 'JWS156-1' has been demonstrated to have high salt tolerance (Hamwieh and Xu 2008) and alkali salt tolerance (Tuyen *et al* 2010; Xu *et al.* 2012). The wild soybean accession also has some special morphological traits, such as it has short internode length, differing from other wild soybean.

2.2.2 Developing of the chromosome segment substitution line (CSSL) population

The wild soybean was the donor parent and the 'Jackson' cultivar was the recurrent parent. Different chromosomal segments of the wild soybean were introgressed into the

genetic background of ‘Jackson’. These lines were developed by backcrossing of {[((‘Jackson’ × ‘JWS156-1’) × ‘Jackson’) × ‘Jackson’] × ‘Jackson’} and five successive generations of self-pollination without any selection (Figure 2. 2). The crossing between the cultivated soybean variety ‘Jackson’ and the Japanese wild soybean accession ‘JWS156-1’ and the following backcross with ‘Jackson’ were done by my colleagues. I joined this work in the F_5 self-pollination stage and finally obtained 120 BC_3F_5 CSSLs.

2.2.3 SSR analysis

Genomic DNA was extracted from young leaf tissues using a modified CTAB method. A total of 235 SSR soybean markers from all 20 chromosomes (Song *et al.* 2004; 2010) were used for genotyping the 120 CSSLs. The 235 simple sequence repeats (SSR) makers covered 2,073.79 cM with an average distance of 9.5 cM between two neighbor makers (Table 2. 1). The SSR reaction mix contained 2 µl (10 pmol) of each primer, 10 µl Quick *Taq*TMHS DyeMix (Toyobo, Osaka, Japan), and 3 µl (10–50 ng) of genomic DNA template in a total volume of 20 µl. The polymerase chain reaction (PCR) was performed at 94°C for 30 s, followed by 32 cycles of 30 s at 94°C, 30 s at 55°C, 40 s at 72°C, and a final extension at 72°C for 10 min. After amplification, PCR products were separated on an 8% (w/v) denaturing polyacrylamide gel in 1 × Tris borate EDTA (TBE) running buffer at 250 V for 3.5 h and stained by ethidium bromide. To detect polymorphism in fragment size of PCR products, the gel was scanned in the pharos FX Molecular Imager (Bio-Rad Laboratories, Hercules, USA). Display Graphical Genotypes (2.0) software was used to draw the graphical genotypes schematic diagram of the CSSLs.

2.3 Result

2.3.1 Genotypic characteristics of the BC_3F_5 CSSLs

The graphical genotypes of the 120 CSSLs were shown in Figure 2. 3. Of the 235 SSR markers, 224 markers had at least one 'JWS156-1' allele in the 120 CSSLs, and 11 markers showed no wild soybean allele as indicated by small gaps in the graphical genotypes. The CSSLs were almost recovered by the recurrent parent 'Jackson' after three times of backcrosses and no line with abnormal growth were observed in the 120 CSSLs. The proportion of the recurrent parent 'Jackson' alleles in each CSSL ranged from 80.3% to 99.2% with an average of 92.9 ± 4.0%.

2.4 Discussion

The advanced QTL strategy is to develop BC_2 or BC_3 populations, typically BC_3; generations beyond the BC_3 are likely to have too few segments from wild soybean, lead to have low statistical power to detect most QTLs in BC_4 population (Grandillo and Tanksley 2005). After five generations of self-pollination, the CSSLs were almost homologous, however, some heterozygous segments still remained. The remained heterozygous segments might be used to produce residual heterozygous line, which can be used for validating and fine mapping QTLs.

In this study, I was involved in the construction of BC_3F_5 CSSL population from a cross between a cultivated soybean cultivar 'Jackson' and a wild soybean accession 'JWS156-1'. To my knowledge, this population was the first CSSL population in which a Japanese wild soybean accession was used as donor parent. This BC_3F_5 CSSL population was used to identify QTLs to gain a better understanding of the genetic basis of 100-seed weight, soybean flowering time, and seed coat cracking. Of course, this CSSL

population might also be used for studying other traits in soybean.

Table 2. 1 The 235 soybean SSR markers from the 20 chromosomes used for genotyping the 120 CSSLs.

Marker	Chr.	cM	Marker	Chr.	cM	Marker	Chr.	cM	Marker	Chr.	cM	Marker	Chr.	cM	Marker	Chr.	cM	Marker	Chr.	cM	Marker	Chr.	cM
Sat_413	1	5.93	Satt022	3	102.06	06_0409	6	42.94	Satt242	9	14.35	Satt666	12	0.59	Satt126	14	27.63	Satt582	17	53.85	19_1044	19	51.38
Satt184	1	17.52	Sat_337	4	32.10	Satt422	6	44.66	Sat_119	9	17.11	Satt635	12	4.88	Satt083	14	51.49	Satt447	17	66.27	Satt481	19	54.57
Sat_353	1	36.23	04_0415	4	40.86	Sat_336	6	51.84	Satt102	9	30.28	Satt353	12	8.48	14_0645	14	56.09	Satt669	17	67.71	Satt448	19	64.66
Satt321	1	50.16	Sat_140	4	41.43	Satt643	6	94.65	Satt178	9	40.86	12_0220	12	27.51	Satt474	14	75.34	Sat_292	17	75.29	Satt678	19	70.20
Satt283	1	60.97	Satt646	4	70.52	Sat_076	6	99.18	Sat_281	9	45.14	Satt568	12	27.64	Satt534	14	87.59	Sat_354	17	79.63	Sat_099	19	78.23
Satt370	1	60.99	Satt718	4	73.79	06_1162	6	103.42	Sat_111	9	55.70	Satt192	12	44.04	Satt726	14	100.55	Sat_209	17	85.64	Satt664	19	92.66
Satt436	1	70.69	Sat_416	4	76.41	Satt371	6	145.48	Sat_043	9	61.67	12_0343	12	46.00	Satt687	14	113.61	Satt528	17	86.34	Satt229	19	93.89
Sat_036	1	75.25	Sat_357	4	76.43	Satt636	7	5.00	Satt499	9	71.01	Satt442	12	46.95	Satt575	15	3.30	Sat_362	17	86.69	Satt513	19	106.37
Sat_414	1	85.48	Sat_322	4	79.27	Satt201	7	13.56	Sat_243	9	86.78	Satt541	12	53.35	Sat_112	15	8.67	Sat_300	17	87.56	Satt571	20	18.50
Satt129	1	109.67	Sat_042	4	82.51	Sat_316	7	21.00	Sat_126	9	108.20	Sat_158	12	73.46	Satt720	15	20.80	Satt615	17	91.21	Satt367	20	27.98
Satt216	2	9.80	Satt195	4	84.81	07_0262	7	34.27	Satt588	9	117.02	Satt142	12	86.49	Satt651	15	32.10	Sat_086	17	118.66	Satt614	20	31.94
Sat_351	2	20.61	Sat_207	4	87.31	Satt463	7	50.10	10_0087	10	9.04	Satt293	12	89.09	Satt685	15	56.70	Sat_022	17	120.30	Sat_219	20	36.03
Sat_373	2	35.75	Satt524	4	120.12	Satt702	7	61.04	Satt500	10	14.17	Satt317	12	89.52	15_1454	15	87.06	Satt386	17	125.00	Satt239	20	36.94
Sat_211	2	38.04	Satt682	4	127.06	Satt494	7	71.71	Sat_318	10	24.61	Sat_218	12	99.50	15_1583	15	96.42	Satt570	18	12.74	Satt354	20	46.22
Satt634	2	46.62	Satt625	5	26.16	Satt680	7	77.19	Satt653	10	38.09	Sat_180	12	104.37	Satt674	16	15.95	Satt235	18	21.98	Satt270	20	50.11
Sat_254	2	46.92	05_0724	5	31.29	Satt697	7	85.35	Satt259	10	39.82	Satt434	12	105.74	Sat_228	16	23.91	Sat_290	18	29.03	Sat_268	20	55.10
Satt542	2	53.02	Sat_356	5	42.80	Sat_359	7	139.80	Sat_193	10	61.11	13_0281	13	23.35	Satt693	16	33.88	Sat_308	18	43.09	Satt650	20	63.33
Satt266	2	59.61	Satt717	5	51.95	Sat_330	7	140.69	Satt581	10	106.03	Sat_298	13	32.32	Sat_370	16	37.40	Satt566	18	49.91	Sat_104	20	65.62
Sat_135	2	70.65	Satt648	5	59.18	Satt207	8	26.50	Sat_307	10	123.43	13_0517	13	35.01	Satt132	16	39.18	Satt504	18	59.83	Satt671	20	72.09
Sat_139	2	93.35	Satt385	5	64.74	Satt315	8	45.29	Sat_109	10	127.50	Sat_309	13	41.47	Satt215	16	44.08	Satt505	18	63.00	Satt330	20	77.84
Satt274	2	116.35	Sat_267	5	78.45	Sat_212	8	56.32	Sat_190	10	129.80	Sat_133	13	50.78	Satt621	16	53.68	18_1535	18	71.58	Sat_324	20	84.48
Satt271	2	137.06	Satt200	5	92.88	Satt424	8	60.59	11_0141	11	18.03	Satt510	13	71.41	16_0969	16	57.95	Satt288	18	76.76	Satt162	20	86.74
03_0020	3	2.81	Sat_271	5	97.76	Satt341	8	77.70	Satt426	11	28.33	Sat_375	13	88.09	Satt431	16	78.57	Satt472	18	94.84	Sat_155	20	98.06
Satt624	3	35.32	Sat_217	5	101.57	Sat_129	8	84.08	Sat_411	11	30.87	Satt490	13	97.97	Sat_394	16	89.43	Sat_117	18	100.00	Sat_299	20	99.83
Sat_084	3	36.86	06_0106	6	12.98	Sat_392	8	106.29	Sat_261	11	32.95	Satt522	13	119.19	Sat_393	16	90.33	Sat_372	18	107.75	Satt148	20	100.78
Sat_266	3	47.28	Satt227	6	26.65	Sat_382	8	116.41	Sat_247	11	49.73	Satt656	13	135.11	Sat_296	17	6.42	Sat_301	19	11.12			
Satt339	3	75.91	Satt640	6	30.47	Satt209	8	128.44	Satt583	11	84.19	Sat_074	13	142.35	Satt458	17	24.52	Satt523	19	27.92			
Satt255	3	76.49	Sat_062	6	30.80	Sat_294	8	131.97	Satt444	11	85.92	Satt395	13	146.42	Sat_277	17	28.79	Sat_405	19	29.62			
Sat_091	3	79.51	06_0282	6	31.78	Satt409	8	145.57	Sat_123	11	100.88	14_0050	14	8.00	Satt014	17	29.56	Sat_320	19	32.36			
Sat_285	3	81.92	Satt281	6	40.30	Satt228	8	154.11	Satt484	11	118.52	Sat_342	14	20.31	Satt486	17	34.09	Satt462	19	41.00			

Figure 2. 1 The parent materials, ‘Jackson’ and ‘JWS156-1’ used for developing the CSSL population. A: ‘JWS156-1’ grown in a field of JIRCAS farm. B: Comparison of the seed sizes between ‘JWS156-1’ and ‘Jackson’.

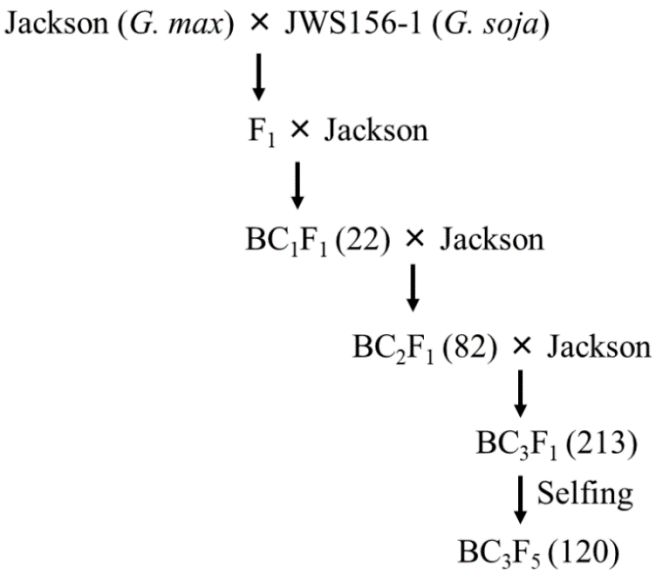

Figure 2. 2 Scheme for development of the wild soybean 'JWS156-1' chromosome segment substitution lines (CSSLs) in the cultivated soybean variety 'Jackson' background. The numbers of plants or lines are indicated in parentheses.

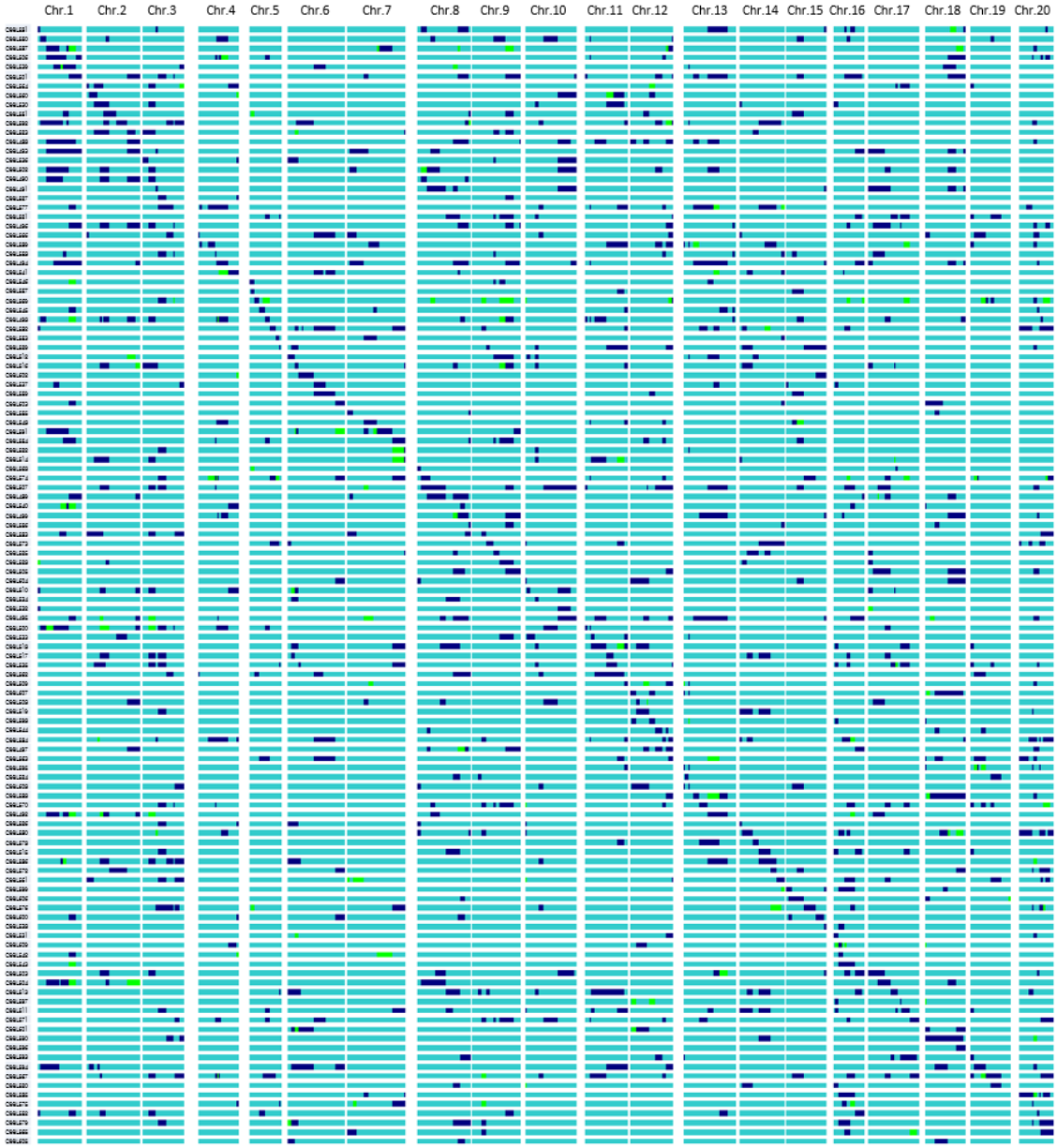

Figure 2. 3 Graphical genotypes of the 120 wild soybean 'JWS156-1' chromosome segment substitution lines (CSSLs) in the cultivated soybean variety 'Jackson' background. blue, 'JWS156-1' homozygous; cyan, 'Jackson' homozygous; green, heterozygous.

CHAPTER 3

Identification and validation of QTLs for 100-seed weight using the CSSL population

3.1 Introduction

The seed weight, generally expressed as 100-seed weight (100SW), is one of the most important yield components in soybean. Several studies have shown a positive association between seed weight/size and seed yield (Burris *et al.* 1973; Smith and Camper 1975). In the past decades, many studies have focused on genetic studies of seed weight/size in the soybean, particularly using the DNA marker for QTL analyses. To date, 318 QTLs of seed weight/size on all the 20 chromosomes have been reported in SoyBase (http://www.soybase.org/). However, compared with other major crops, such as rice (Duan *et al.* 2015; Hu *et al.* 2015; Wang *et al.* 2015), knowledge about the molecular mechanisms of seed weight control in soybean is very limited. To date, only two seed weight/size related genes are isolated from soybean (Jeong *et al.* 2012; Lu *et al.* 2017b). Identification of a major and stable QTL for seed weight is very important for facilitating map-based cloning of seed weight gene and understanding of the molecular mechanisms of seed weight control in soybean.

3.2 Materials and methods

3.2.1 Plant materials

A total of 120 BC_3F_5 CSSLs were used for detecting the 100SW QTL. The 100SW of 'Jackson' was around 22 g and that of 'JWS156-1' was about 2.2 g: a ten-fold

difference. To validate the QTL (*qSW12.1*) detected in the present study, a residual heterozygous line (RHL), RHL564, which was heterozygous at the *qSW12.1* QTL region, was selected from the BC_3F_5 population. Progenies produced by self-pollinating heterozygous plants of RHL564 were used to confirm the effect of *qSW12.1*.

3.2.2 Field trials and phenotypic evaluation of the BC_3F_5 CSSLs

All plant materials were grown in the experimental farm of the Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki, Japan; located at 36.05°N, 140.08°E. The 120 BC_3F_5 CSSLs were grown in 2014, 2015, and 2016. In 2014, a single BC_3F_5 plant for each CSSL was harvested and evaluated for seed weight. Seeds harvested from the single plant were used for phenotype evaluation in 2015, and the bulked seeds harvested in 2015 for each CSSL, excluding the individual plants at each end of the row, were used for phenotype evaluation in 2016.

Soybean seeds were sown on June 24, 2014; June 23, 2015; and June 21 2016. Each CSSL or cultivar was planted in a single row plot 6 m in length, 30 plants per row, a row spacing of 60 cm and plant separation of 20 cm. All plots were arranged with a completely randomized block design with one (2014, 2015) or two (2016) replicates. Soybeans were harvested on November 14, 2014; November 12, 2015; and November 16, 2016.

After harvesting, soybean plants were air-dried in natural conditions for at least four weeks and then threshed. When the seed desiccation was complete, the 100SW trait was measured for each BC_3F_5 CSSL as grams per 100 random healthy seeds.

3.2.3 SSR analysis

SSR analysis was the same as the methods described in 2.2.2 SSR.

3.2.4 QTL analysis

QTL analysis was conducted with the IciMapping software (Meng *et al.* 2015). Stepwise regression-based likelihood ratio tests of additive QTL (RSTEP-LRT-ADD) method was used for detecting 100SW QTL. A threshold of LOD score greater or equal to 3.0 was set to declare existence of a QTL.

3.2.5 Validation of the allelic effect of qSW12.1

To further verify the effect of the *qSW12.1* QTL detected in the present study, the residual heterozygous line RHL564 was selected from the BC_3F_5 population. Homozygous plants with the 'Jackson' and 'JWS156-1' genotypes were selected from the progenies produced by self-pollinating heterozygous plants of RHL564 using the SSR marker Sat_180 for genotyping. Because these two genotypes had a similar genetic background and differed only in the *qSW12.1* region, they can be regarded as near-isogenic lines (NILs). For the homozygous plants with the 'Jackson' genotype, I designated the lines as NILs564-C, whereas homozygous plants with the 'JWS156-1' genotype were designated as NILs564-W. Forty-three plants of NILs564-C and 44 plants of NILs564-W were evaluated in filed condition in 2016. The cultivation conditions were the same as those of the CSSLs in 2016. A two-tailed student's t test was performed to determine the statistical significance between NILs564-C and NILs564-W.

3.3 Results

3.3.1 The phenotypic variation of the BC_3F_5 CSSLs in the 3 years

The frequency distribution and statistics analysis of 100SW in the 120 BC_3F_5 CSSLs

over the 3 years was showed in Figure 3. 1 and Table 3. 1. The 100SW demonstrated a continuous distribution, and the seed size of most CSSLs is intermediate compared to that of the two parents (Figure 3. 2). However, in 2014, 2015, and 2016, one, three, and one CSSLs, respectively, showed a higher 100SW than the cultivated soybean 'Jackson'. ANOVA analyses showed significant effects ($P < 0.01$) of the different CSSLs and the different years. The 120 CSSLs showed a large variation in terms of 100SW, ranging approximately from 10 g to 22 g, with a two-fold difference. Over the 3 years, the highest average 100SW of all CSSLs was 19.84 g (2015), and the lowest was 16.89 g (2016), whereas the average in 2014 was 17.64 g. This result indicates different growth conditions over the 3 years. There was a strong positive correlation ($P < 0.01$) between different years (Figure 3. 3), suggesting that seed weight trait has high hereditability in soybean and it is suitable for QTL analysis.

3.3.2 QTLs associated with seed weight

Based on QTL analysis, four, five, and five SSR markers associated with 100SW were detected in 2014, 2015, and 2016, respectively (Table 3. 2). These markers were distributed on chromosomes 8, 9, 12, 13, 14, 16, 17, and 20. BARCSOYSSR_14_0645 ($3.00 < LOD < 6.63$) located on chromosome 14 and Sat_180 ($6.78 < LOD < 12.31$) located on chromosome 12 were detected in all 3 years. Satt582 ($7.16 < LOD < 8.99$) located on chromosome 17 was detected in both 2014 and 2015. Satt431 (LOD = 4.52, chromosome 16), Satt490 (LOD = 3.77, chromosome 13), Satt354 (LOD = 3.83, chromosome 20), Sat_129 (LOD = 7.66, chromosome 8), Sat_126 (LOD = 3.46, chromosome 9), and Sat_086 (LOD = 3.33, chromosome 17) were detected in only 1 year. Overall, a total of nine markers were detected over the 3 years. Sat_180 has the highest

LOD scores in 2015 and 2016 and the second highest LOD score in 2014. In all nine loci, the wild soybean allele contributed to decreasing seed size with effects ranging from −1.96 g to −0.87 g.

Based on their chromosome locations, nine 100SW QTLs, corresponding to the nine SSR markers associated with 100SW, were designated as *qSW8.1, qSW9.1, qSW12.1, qSW13.1, qSW14.1, qSW16.1, qSW17.1, qSW17.2,* and *qSW20.1* (Table 3. 2). For the three major and stable QTLs (*qSW12.1*, *qSW14.1*, and *qSW17.1*) detected in the present study, *qSW14.1* may correspond to the seed weight QTL reported by Hoeck *et al.* (2003) and Li *et al.* (2007a), whereas *qSW17.1* may correspond to the QTL (*qSw17-1*) detected by Kato *et al.* (2014). In contrast, no seed weight QTL has been reported for the *qSW12.1* region to date, suggesting that this is a novel major QTL in soybean.

3.3.3 Validating the allelic effect and determining the physical position of qSW12.1

To further verify the effect of the *qSW12.1* QTL, the residual heterozygous line RHL564 was selected from the BC_3F_5 population. Homozygous plants with the 'Jackson' and 'JWS156-1' genotypes were selected from the progenies of RHL564 to produce *qSW12.1* NILs (NILs564-C and NILs564-W). Field evaluation revealed that NILs564-C had 100SW of 18.64 ± 1.03 g, which was significantly ($P < 0.01$) higher than that of the NILs564-W plants (17.71 ± 1.25 g) (Figure 3. 4). In addition, the seed size of NILs564-C was also larger than that of the NILs564-W (Figure 3. 5). This result confirmed the effect of the *qSW12.1* QTL for controlling seed weight in soybean.

3.3.4 Determination of physical position and prediction of candidate gene of a major

100SW QTL qSW12.1

Based on the genotyping of NILs564-C and NILs564-W derived from RHL564, *qSW12.1* was delimited in a 1,348-kb interval between the BARCSOYSSR_12_1282 and BARCSOYSSR_12_1347 markers (Figure 3. 6). According to the soybean genome sequence of *Glycine max Wm82.a2.v1* at Phytozome 12 (https://phytozome.jgi.doe.gov), there are 129 predicted genes in this region.

Of these, 103 genes were expressed in the seed, 112 genes in the pod, and 99 genes in both the seed and pod tissues. The RNA Seq-Atlas of *Glycine max Wm82.a2.v1* revealed that 54 predicted genes had Fragments Per Kilobase Million (FPKM) values ranging from 5.01 to 232.40 in the seed, and 52 candidate genes had FPKM values from 5.53 to 99.89 in the pod. Of the 129 predicted genes, *Glyma.12G224200* encodes a putative ubiquitin-conjugating enzyme 25. Several factors involved in ubiquitin-related activities have been known to influence seed size in plants (Li and Li 2014). Because *Glyma.12G224200* has an FPKM value >5 in both pod and seed tissues. it may be regarded as a strong candidate gene for controlling seed size in the *qSW12.1* QTL.

3.4 Discussion

To identify the QTL for 100SW, 120 BC_3F_5 CSSLs of soybean were cultivated over 3 years. A total of nine QTLs (*qSW8.1*, *qSW9.1*, *qSW12.1*, *qSW13.1*, *qSW14.1*, *qSW16.1*, *qSW17.1*, *qSW17.2*, and *qSW20.1*) were detected on eight chromosomes. Of these, three (*qSW12.1*, *qSW17.1*, and *qSW14.1*) are major and stable additive QTLs. The *qSW14.1* QTL was detected over 3 successive years. The seed weight QTL has been previously detected in the same region by Hoeck *et al.* (2003) and Li *et al.* (2007a), indicating consistency across different environmental conditions and genetic backgrounds. The

qSW17.1 QTL was detected over 2 years (2014 and 2015) as a major QTL with LOD scores of 8.99 and 7.16, respectively, it was not detected in 2016. Recently, Lu *et al.* (2017b) identified a protein phosphatase type 2C-1 allele (PP2C-1) as the causative gene underlying a QTL that contributed a positive effect for increasing seed weight. This QTL located in the same region of *qSW17.1*.

qSW12.1 (LOD = 6.78-12.31) was detected over the 3 successive years on chromosome 12. There was no QTL was previously reported around this region. Among the two homologous progenies produced by self-pollination of RHL564, higher seed weight was observed in 'Jackson' genotype plants than that in 'JWS156-1' genotype plants. Analyses of RHL564 delimited *qSW12.1* in an interval of approximately 1,348 kb between the BARCSOYSSR_12_1282 and BARCSOYSSR_12_1347 markers on chromosome 12. Thus, *qSW12.1* was regarded as a novel, stable, and major QTL and delimited in an interval of chromosome 12. However, it is still difficult to identify the candidate gene underlying *qSW12.1* from the 129 predicted genes.

Of the 129 predicted genes, *Glyma.12G224200* encodes a putative ubiquitin-conjugating enzyme 25. Several factors involved in ubiquitin-related activities have been known to influence seed size in plants (Li and Li 2014). This predicted gene may be regarded as a strong candidate gene underlying *qSW12.1*. While the region of *qSW12.1* is currently too large for positional cloning of the casual gene underlying the QTL, In the future work, I will use a much larger segregation population to narrow down the *qSW12.1* QTL into a smaller chromosome region to identify the candidate causal gene for seed weight. Compared to genetic studies of seed weight in other plants, such studies are very limited in the soybean. Cloning genes controlling seed weight would not only contribute to a better understanding of the mechanism of seed development but also provide

important implications for the future breeding strategies for soybean improvement.

Table 3. 1 Statistical analysis of 100-seed weight in the BC_3F_5 chromosome segment substitution line (CSSL) population in 2014, 2015, and 2016

Year	Trait	Parents		BC_3F_5 CSSLs								
		Jackson	JWS156-1	SampleSize[a]	Mean	Variance[b]	Std[c]	Skewness[d]	Kurtosis[e]	Minimum	Maximum	Range
2014	100SW (g)	22.67	2.31	120	17.64	6.51	2.55	-0.38	-0.33	10.98	23.04	12.06
2015	100SW (g)	23.63	2.34	120	19.84	5.26	2.29	-0.46	-0.06	12.4	24.46	12.06
2016	100SW (g)	21.6	2.3	120	16.89	6.67	2.58	-0.47	-0.32	9.81	21.88	12.08

[a]SampleSize: Size of the mapping population.

[b]Variance: Variance of the phenotypic trait.

[c]Std: Standard deviation of the phenotypic trait.

[d]Skewness is a measure of the phenotypic trait of the probability distribution of a real-valued random variable about its mean.

[e]Kurtosis is a measure of the phenotypic trait of the probability distribution of a real-valued random variable.

Table 3. 2 SSR makers associated with 100-seed weights (100SW) in the BC_3F_5 chromosome segment substitution line (CSSL) population over different years

Year	Marker name	QTL	Chr. (LG)	Physical position (bp)[a]/Genetic Position (cM)[b]	LOD[c]	PVE (%)[d]	Add[e]	M (QQ)[f]	M (qq)[g]	DHR(%)[h]
2014	Sat_180	*qSW12.1*	Gm12 (H)	38,388,961/104.37	6.78	15.25	-1.21	12.26	14.69	19.17
	BARCSOYSSR_14_0645	*qSW14.1*	Gm14 (B2)	13,276,701/-	3.00	6.27	-1.11	12.37	14.59	8.33
	Satt431	*qSW16.1*	Gm16 (J)	35,718,476/78.57	4.52	9.72	-1.38	12.10	14.86	8.33
	Satt582	*qSW17.1*	Gm17 (D2)	9,949,907/53.84	8.99	21.16	-1.75	11.72	15.23	11.67
2015	Sat_180	*qSW12.1*	Gm12 (H)	38,388,961/104.37	10.07	21.06	-1.28	14.01	16.56	19.17
	Satt490	*qSW13.1*	Gm13 (F)	36,699,447/97.97	3.77	6.94	-0.87	14.41	16.16	14.17
	BARCSOYSSR_14_0645	*qSW14.1*	Gm14 (B2)	13,276,701/-	5.18	9.80	-1.24	14.05	16.53	8.33
	Satt582	*qSW17.1*	Gm17 (D2)	9,949,907/53.84	7.16	14.11	-1.28	14.00	16.57	11.67
	Satt354	*qSW20.1*	Gm20 (I)	34,569,381/46.22	3.83	7.07	-1.24	14.04	16.53	5.83
2016	Sat_129	*qSW8.1*	Gm8 (A2)	14,730,465/84.08	7.66	13.16	-1.95	7.74	11.63	6.67
	Sat_126	*qSW9.1*	Gm9 (K)	44,406,619/108.19	3.46	5.46	-1.43	8.25	11.12	5.00
	Sat_180	*qSW12.1*	Gm12 (H)	38,388,961/104.37	12.31	23.25	-1.64	8.04	11.32	19.17
	BARCSOYSSR_14_0645	*qSW14.1*	Gm14 (B2)	13,276,701/-	6.63	11.17	-1.62	8.07	11.30	8.33
	Sat_086	*qSW17.2*	Gm17 (D2)	39,875,727/118.66	3.33	5.25	-1.96	7.72	11.65	2.92

[a]Physical positions are based on Glycine max Wm82.a2.v1 (Schmutz *et al*. 2010).

[b]Genetic positions are based on SoyBase (https://soybase.org/).

[c]LOD: LOD score calculated from single marker analysis.

[d]PVE (%): Phenotypic variation explained by the marker.

[e]Add: Estimated additive effect of the marker.

[f]M (QQ): Mean value of the QTL genotype of donator parent.

[g]M (qq): Mean value of the QTL genotype of recurrent parent.

[h]DHR: Genotype ratio of the donor homozygous allele.

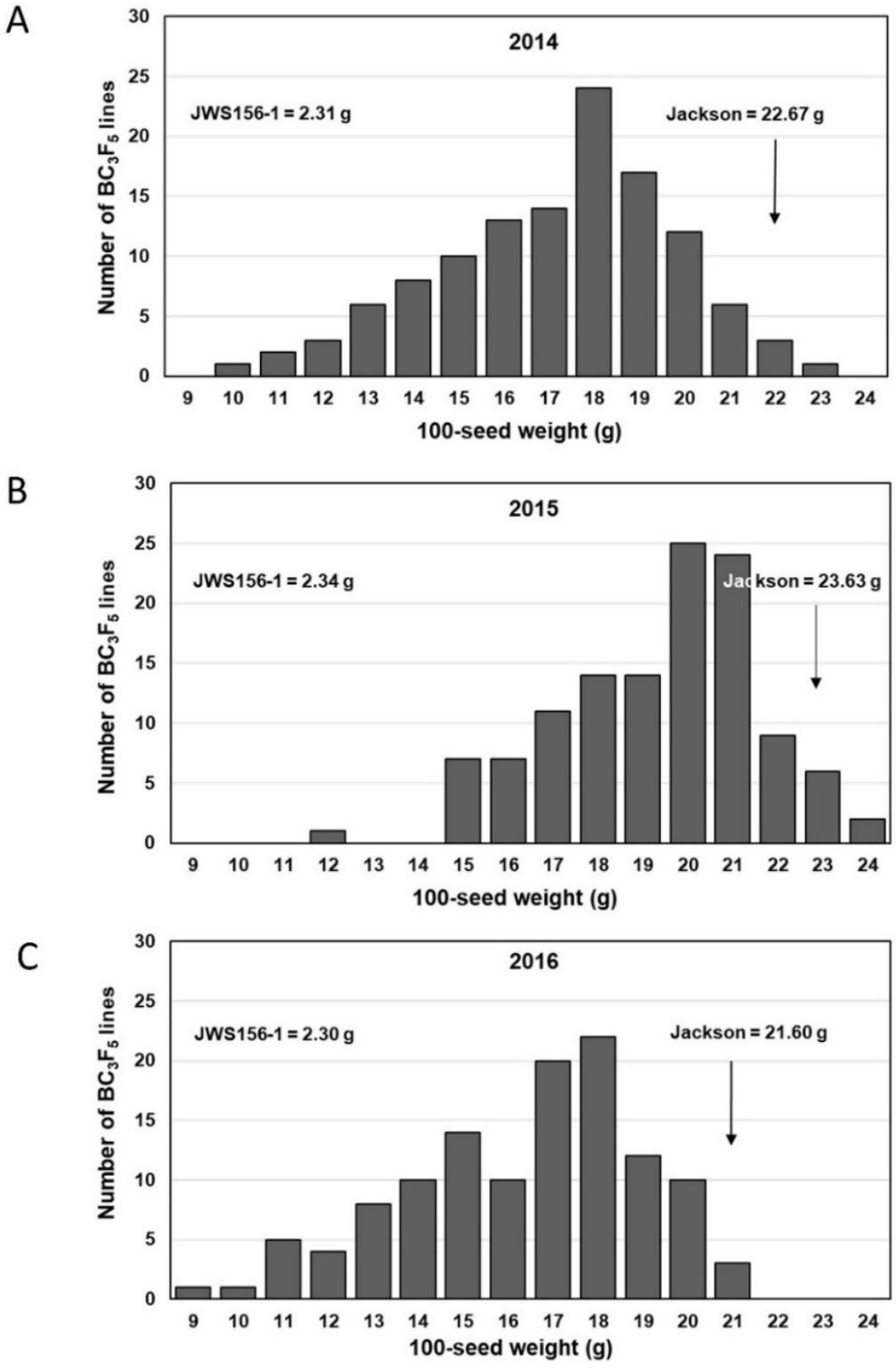

Figure 3. 1 Frequency distribution of 100-seed weight of the 120 BC_3F_5 chromosome segment substitution lines (CSSLs) in the years 2014 (A), 2015 (B), and 2016 (C).

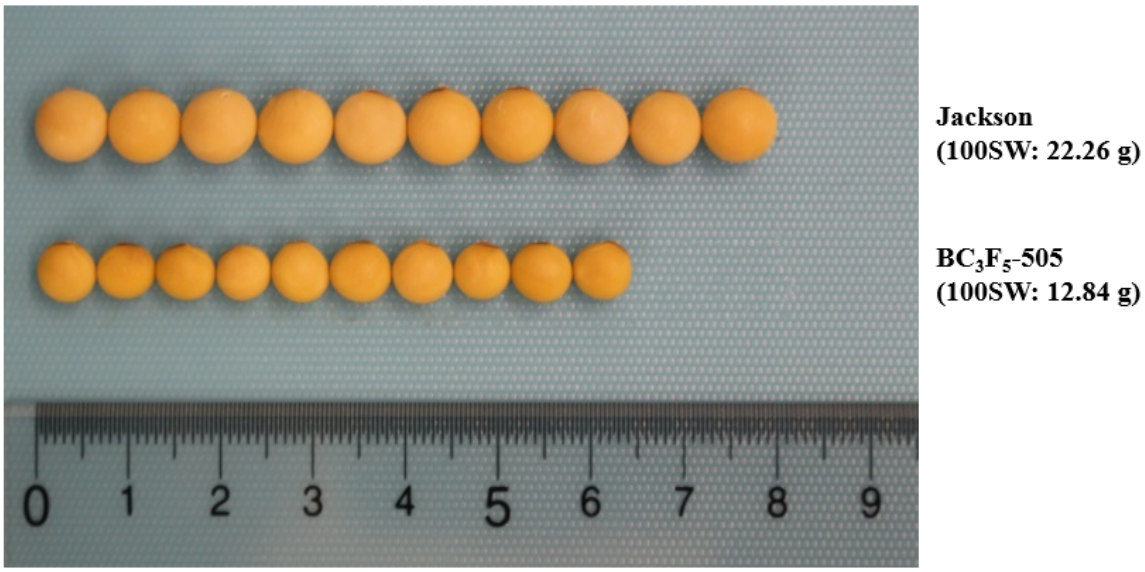

Figure 3. 2 Comparison of the seed sizes between the CSSL BC_3F_5-506 and 'Jackson'.

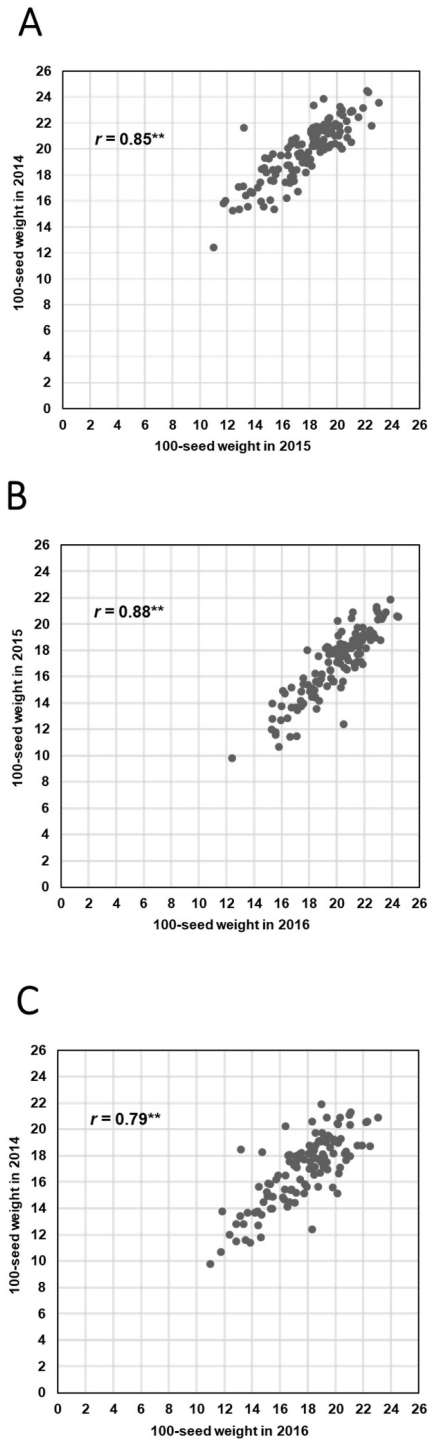

Figure 3. 3 The correlation of the 100-seed weight of the BC_3F_5 chromosome segment substitution lines (CSSLs) over different years. ** Indicates significant correlation at *P-value* < 0.01.

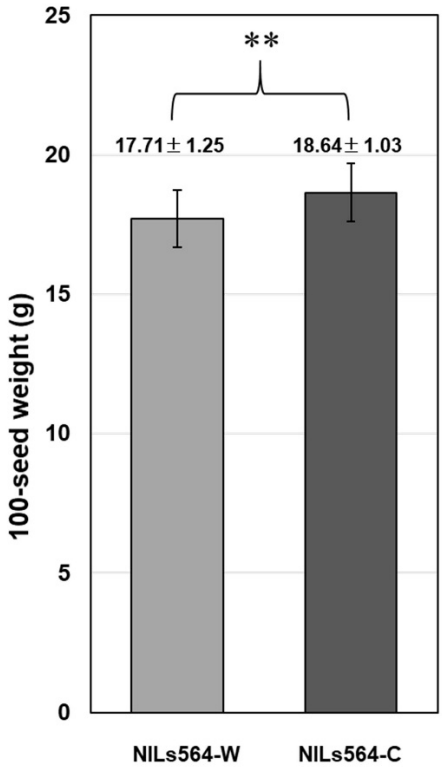

Figure 3. 4 The allele effect of *qSW12.1* in near-isogenic lines NILs564-C ('Jackson' genotype, n = 43) and NILs564-W ('JWS156-1' genotype, n = 44). Data are shown as means ± SD (standard deviation). ** Indicates significant difference between the means (*P-value* < 0.01).

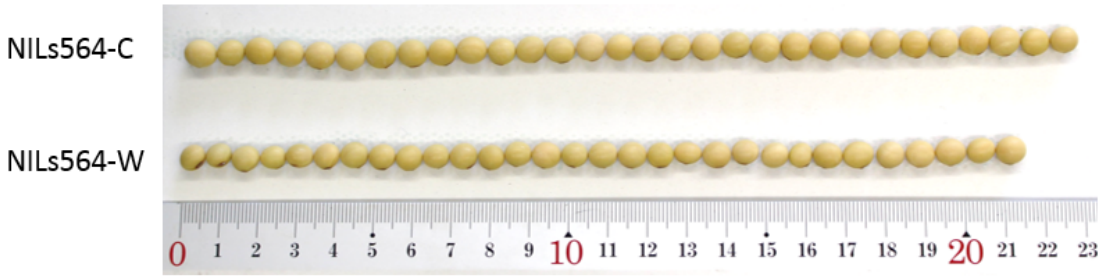

Figure 3. 5 Comparison of seeds sizes between near-isogenic lines NILs564-C and NILs564-W.

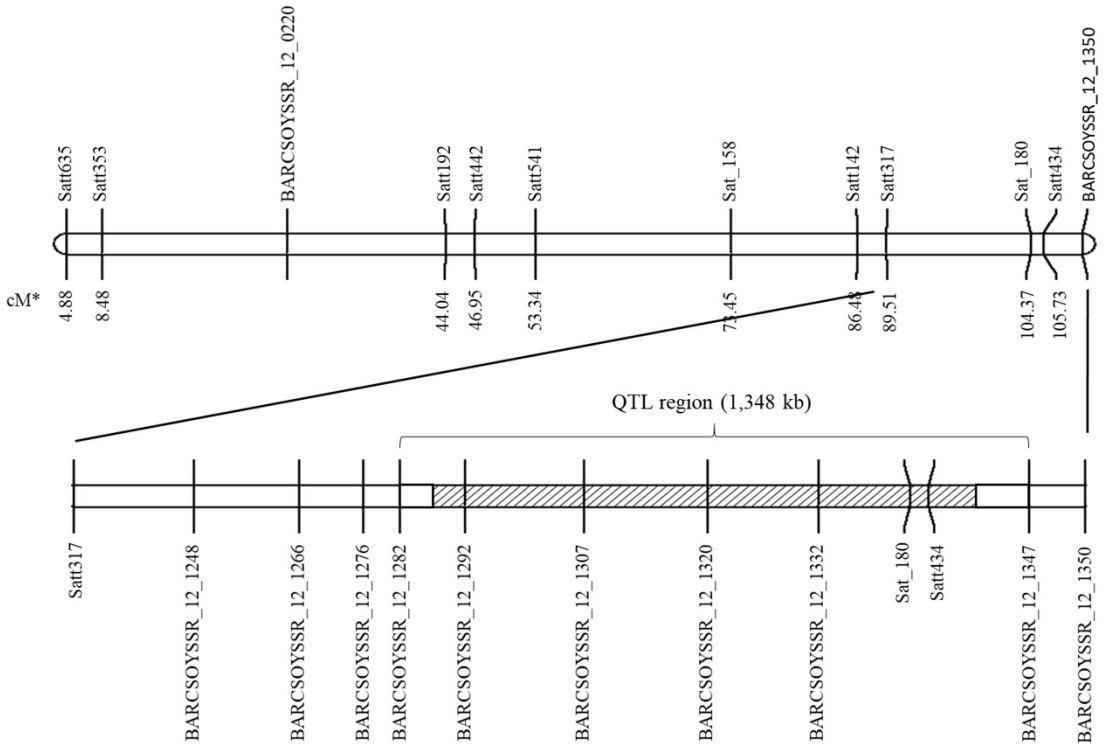

Figure 3. 6 The novel seed weight QTL (*qWS12.1*) was delimited in a 1,348-kb interval between the BARCSOYSSR_12_1282 and BARCSOYSSR_12_1347 markers on chromosome 12. * marker positions are based on SoyBase (https://soybase.org/). Shaded area indicates the segregating chromosomal region between near-isogenic lines NILs564-C and NILs564-W.

CHAPTER 4

A major QTL (*qFT12.1*) allele from wild soybean delays flowering time

4.1 Introduction

Soybean is highly sensitive to photoperiod. Each cultivar can only be cultivated in a very narrow latitudinal zone. Flowering time is a critical trait affecting soybean adaptability, growth, maturity, and grain yield. Understanding the genetic and molecular mechanisms regulating soybean flowering time is extremely important for soybean breeding. QTL analysis using molecular markers has been widely used to identify flowering time genes in soybean. To date, more than one hundred of flowering time QTLs, distributed on 17 chromosomes, have been reported and registered in Soybase (https://soybase.org/sbt/). In soybean, 11 genes controlling flowering time have been reported in soybean. These include *E1* (Bernard *et al*. 1971), *E2* (Bernard *et al*. 1971), *E3* (Buzzell *et al*. 1971), *E4* (Buzzell and Voldeng 1980), *E5* (McBlain and Bernard 1987), *E6* (Bonato and Vello 1999), *E7* (Cober and Voldeng 2001), *E8* (Cober *et al*. 2010), *E9* (Kong *et al*. 2015), *E10* (Samanfar *et al*. 2017), and *J* (Ray *et al*. 1995).

Although some flowering time genes are identified and cloned from soybean, knowledge about gene or gene system regulating flowering time is still limited. In the present study, I detected a major and stable QTLs (*qFT12.1*) for flowering time in the 120 BC_3F_5 CSSL population and further validated the allelic effect of the major QTL. This QTL may be a new flowering time gene in soybean.

4.2 Materials and methods

4.2.1 QTL analysis for flowering time using a CSSL population

The 'Jackson' × 'JWS156-1' BC_3F_5 CSSL population was used to for detecting flowering time QTL. The 120 BC_3F_5 CSSLs as well as the parental genotypes ('Jackson' and 'JWS156-1') were grown at the experimental farm of JIRCS, Tsukuba, Ibaraki, Japan (36.05°N, 140.08°E) over 2 years. Soybean seeds were sown on June 21, 2016, and June 27, 2017. Each CSSL or cultivar was planted in a single row plot 6 m in length, with 60 cm spacing between rows and 20 cm spacing between plants. All plots were arranged in a completely randomized block design with one or two replicates. Flowering time was recorded as the number of days from emergence to the opening of the first flower on 50% of plants in each line (Fehr *et al*. 1971).

Total DNA was extracted from leaf tissue of the CSSLs and the two parents using a modified CTAB method. A total of 235 soybean SSR markers spanning all 20 chromosomes (Song *et al*. 2004, 2010) were used for genotyping. The polymerase chain reaction (PCR) mixture comprised 3 µl (10–50 ng) of template DNA, 2 µl (10 pmol) of each primer, 10 µl of Quick *Taq* TMHS DyeMix (Toyobo, Osaka, Japan), and 5 µl of distilled deionized water in a total volume of 20 µl. PCR amplification in a TAdvanced 384 thermal cycler (Biometra, Göttingen, Germany) was performed with an initial activation step at 95°C for 30 s, followed by 32 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 40 s, and a final extension at 72°C for 10 min. After PCR, 12 µl of each PCR product was separated on an 8% denaturing polyacrylamide gel in 1 × TBE running buffer for 3.5 h at 250 V, followed by ethidium bromide staining. The fragments were visualized by scanning using the pharos FX Molecular Imager (Bio-Rad Laboratories, Hercules, CA, USA).

QTL analysis was conducted using the IciMapping software (Meng *et al*. 2015). Stepwise regression-based likelihood ratio tests of additive QTL (RSTEP-LRT-ADD) were used for detecting additive flowering time QTLs. A threshold of LOD score > 3 was set to declare the existence of QTL. Statistical significance of the QTLs was determined using a 1,000-permutation test at 5% significance level.

4.2.2 Confirmation of a flowering time QTL, qFT12.1, using RIL mapping population

An F_7 RIL population (n = 94) derived from the same cross as that of the CSSL population ('Jackson' × 'JWS156-1') was used to confirm the effect of the QTL *qFT12.1* detected in the CSSLs. RILs and their parents were sown on July 1, 2017, at the experimental farm of JIRCAS. Phenotypic and genotypic analyses of RILs were conducted as described for CSSLs. Inclusive composite interval mapping of additive QTLs (ICIM-ADD) was used to detect additive flowering time QTLs (Li *et al*. 2007b; Zhang *et al*. 2008).

4.2.3 Validation of the allelic effect of **qFT12.1** *using a residual heterozygous line (RHL)*

To further validate the effect of QTL *qFT12.1* detected in the CSSL population, I selected RHL509, which was heterozygous at the *qFT12.1* locus, from the BC_3F_5 population. Selfed progenies of the heterozygous RHL509 were grown in an LPH-411SP growth chamber (NKsystem, Osaka, Japan) at 25°C and 14-h light/10-h dark photoperiod. Flowering time was recorded for individual plants at the opening of the first flower. With the genotype of the SSR marker Satt192, which was the nearest marker to *qFT12.1*, these plants were grouped into three groups: the 'Jackson' homologous, the 'JWS156-1'

homologous, and the heterozygous groups. The allelic effect of *qFT12.1* was validated by comparing the flowering time of the three groups.

4.2.4 Investigation of the effect of **qFT12.1** *under different growth conditions*

To investigate the effect of *qFT12.1* under different growth conditions, I selected homozygous lines with the 'Jackson' and 'JWS156-1' genotypes from RHL509 progeny using the SSR marker Satt192. Because these two genotypes had a similar genetic background and differed only in the *qFT12.1* QTL region, they can be regarded as near-isogenic lines (NILs). The homozygous plants with the 'Jackson' genotype were designated as NILs509-C, whereas those with the 'JWS156-1' genotype were designated as NILs509-W.

The two *qFT12.1* NILs (NILs509-C and NILs509-W) as well as 'Jackson' were evaluated at the experimental farm of the Japan International Research Center for Agricultural Sciences in 2017. To investigate the effect of the *qFT12.1* QTL under different growth conditions, we sowed these three lines in the field at 1-week intervals on June 27, July 4, and July 11, 2017. The genotypes were planted in a single row plot 3 m in length, with 60 cm spacing between rows and 20 cm spacing between plants. All lines in the three replicates were arranged in a completely randomized block design. Flowering times were recorded for each line.

4.2.5 Determination of the physical position of the flowering time QTL, **qFT12.1**

The *qFT12.1* NILs (NILs509-C and NILs509-W) were genotyped for 16 SSR markers (Song *et al*. 2004, 2010) located in the QTL region. Genotypes of the NILs were compared to delimit the physical position of *qFT12.1* on chromosome 12. Candidate

genes underlying *qFT12.1* were selected by reference to the soybean genome sequence of *G. max Wm82.a2.v1* at Phytozome 12 (https://phytozome.jgi.doe.gov). The predicted genes and their annotations in the *qFT12.1* interval were based on Gmax_275_Wm82.a2.v1.annotation_info.txt and Gmax_275_Wm82.a2.v1.gene.gff3.gz files downloaded from Phytozome 12. Expression information (RNA Seq-Atlas) for these genes was obtained from the genes.fpkm_tracking file, which was also downloaded from *G. max Wm82.a2.v1* at Phytozome 12.

4.2.6 Gene expression analysis of candidate genes underlying qFT12.1

To identify candidate genes underlying *qFT12,* 1, RNA was extracted from the 40-day-old leaf of 'Jackson', 'JWS156-1', NILs509-C, and NILs509-W using RNeasy™ Plant Mini Kit (Takara, Tokyo Japan). The cDNA template synthesis for quantitative PCR was made using PrimeScript™ RT reagent Kit (Perfect Real Time) (Takara, Tokyo Japan). Quantitative PCR was performed using the SYBR Select Master Mix (Thermo Fisher Scientific, Massachusetts USA). Genes expression were quantified at the logarithmic phase using the expression of the housekeeping *Glyma.08G146500.1* (*GmACTIN)* RNA as an internal control in the StepOne Real-Time PCR System (Thermo Fisher Scientific, Massachusetts USA). Three biological replicates were performed for each genotype. The primers of *Glyma.08G146500.1* (*GmACTIN), Glyma.12G073300.1* (*TOE2-like* gene), and *Glyma.12G073900.1* (*PRR7-like* gene) was shown in Table 4. 3.

4.3 Results

4.3.1 QTL analysis of flowering time using the CSSL population

The frequency distribution and statistics analysis of flowering time in the 120 BC_3F_5

CSSLs in 2016 and 2017 was showed in Figure 4. 1 and Table 4. 1. The flowering time of the 120 BC_3F_5 CSSLs tended toward the early-flowering parent 'Jackson'. Most of the CSSLs were intermediate between 'Jackson' and 'JWS156-1'. However, three and five CSSLs flowered earlier than the recurrent parent 'Jackson' in 2016 and 2017, respectively. The flowering time of CSSLs ranged from 48.5 to 57.5 days after emergence (DAE), with an average of 52.43 DAE, in 2016 and from 46 to 57 DAE, with an average of 49.31 DAE, in 2017. A significant positive correlation was observed between the 2 years ($P < 0.01$; Figure 4. 2; Table 4. 1), suggesting that flowering time has high hereditability in the CSSL population.

On the basis of QTL analysis, several SSR markers were detected to be associated with flowering time in the CSSL population. Two SSR markers, including Satt192 on chromosome 12 and Satt678 on chromosome 19, were detected in 2016, whereas four SSR markers, including BARCSOYSSR_07_0262, Satt192, Satt293, and Satt678 located on chromosomes 7, 12, 12, and 19, respectively, were detected in 2017. The markers Satt192 and Satt678 were detected in both years, whereas Satt293 and BARCSOYSSR_07_0262 were detected only in 2017. Of these, Satt192 showed the highest PVE values in the 2 years (38.27% in 2016 and 36.37% in 2017). The wild soybean alleles of all QTLs contributed to the delayed flowering phenotype with positive effects ranging from 1.05 to 2.39. On the basis of their chromosome locations, four QTLs, corresponding to the four SSR markers, were designated as *qFT07.1*, *qFT12.1*, *qFT12.2*, and *qFT19.1* (Table 4. 2).

4.3.2 Confirming the major flowering time QTL, qFT12.1, *using the F_7 RIL mapping*

population

The frequency distribution of flowering time in the RIL mapping population is shown in Figure 4. 3. QTL analysis detected a major QTL on chromosome 12, accounting for 39.13% of the total phenotypic variation (Figure 4. 4). The QTL was estimated to be located between the SSR markers Satt442 and BARCSOYSSR_12_0254 on chromosome 12. Satt192 was the most closely associated marker with the QTL, which confirmed *qFT12.1* as the major QTL in the RIL mapping population. As in the CSSL population, the wild soybean allele at *qFT12.1* contributed to delayed flowering in the RIL mapping population.

4.3.3 Validating the allelic effect of qFT12.1 *using RHL509*

In total, 49 segregating plants produced by self-pollinating RHL509 were evaluated for flowering time in the growth chamber. On the basis of the genotypes of the SSR marker Satt192, these plants were classified into three groups: the 'Jackson' homologous (n = 12), the 'JWS156-1' homologous (n = 12), and the heterozygous groups (n = 25). The average flowering times of the 'Jackson' homologous, 'JWS156-1' homologous, and heterozygous groups were 43.3, 49.4, and 45.2 DAE, respectively (Figure 4. 5). The flowering time of the 'JWS156-1' homologous plants was significantly longer than those of the 'Jackson' homozygous and the heterozygous plants. However, no significant differences in flowering time were observed between the 'Jackson' homozygous and the heterozygous plants, suggesting that the recessive allele of wild soybean at *qFT12.1* causes the late flowering phenotype.

4.3.4 Effect of **qFT12.1** *under different growth conditions*

To verify the effect of *qFT12.1* under different environmental conditions, I grew NILs509-W, NILs509-C, and 'Jackson' under field conditions with different sowing dates. The time to flowering decreased with delayed sowing. The average flowering times for the three genotypes were 50.6, 47.3, and 43.0 DAE for June 27, July 4, and July 11 sowing, respectively. In all three growth conditions, NILs509-W flowered later than NILs509-C and 'Jackson' (Figure 4. 6). The wild soybean allele at *qFT12.1* delayed flowering time by 2–4 days. This confirmed the allelic effect of *qFT12.1* under different growing conditions.

4.3.5 Determining the physical position of **qFT12.1**

Genotyping NILs509-C and NILs509-W with 16 SSR markers flanking the QTL region localized *qFT12.1* to a 2,664-kb interval between BARCSOYSSR_12_0220 and BARCSOYSSR_12_0367 (Figure 4. 7). According to the reference genome sequence of soybean William 82, 275 annotated genes were present in this region. The RNA Seq-Atlas of *G. max Wm82.a2.v1* showed that 231 genes were expressed in the shoot apical meristem (SAM), 231 genes were expressed in the leaf, and 221 genes were overapply expressed in both. Of these, 111 predicted genes showed fragments per kilobase million (FPKM) values > 5 in SAM and 103 candidate genes exhibited FPKM values > 5 in leaves. Of the predicted genes, *Glyma.12G073300* was identified as the best BLAST hit to an *Arabidopsis* flowering gene *TARGET OF EARLY ACTIVATION TAGGED 2* (*TOE2)*. It has been known that *AtTOE2* delays flowering time in *Arabidopsis*. The TOE2 protein interacts with JASMONATE-ZIM DOMAIN (JAZ) proteins and represses the expression of *FLOWERING LOCUS T* (*FT*) (Zhai *et al*. 2015). These data suggest

Glyma.12G073300 as the candidate gene underlying *qFT12.1*. Another annotated gene, *Glyma.12G073900*, homologous to *Arabidopsis PRR7* that controls flowering time in *Arabidopsis* (Yamamoto *et al.* 2003), was also a potential candidate for the gene underlying *qFT12.1*.

4.3.6 Analyses of candidate genes expressions among the parents and near isogenic lines

The *Glyma.12G073900* (*PRR7-like* gene) showed significant different relative expression between 'Jackson' and 'JWS156-1' in 40-day-old leaf (*P-value* < 0.01) (Figure 4. 8), suggesting that *Glyma.12G073900* might be the candidate gene underlying *qFT12.1*. However, no significant difference was observed between NILs509-C and NILs509-W in 40-day-old leaf (Figure 4. 8). *Glyma.12G073300* (*TOE2-like* gene) showed no significance between 'Jackson' and 'JWS156-1' and between NILs509-C and NILs509-W (data not shown).

4.4 Discussion

Four QTLs (*qFT07.1*, *qFT12.1*, *qFT12.2*, and *qFT19.1*) were detected on three chromosomes. Of these, *qFT12.1* showed the highest effect of the phenotypic variation over 2 years. This QTL was further confirmed in the F_7 recombinant inbred line population. Analysis of the *qFT12.1* BC_3F_5 residual heterozygous line RHL509 validated the allele effect of *qFT12.1* and revealed that the recessive allele of *qFT12.1* resulted in delayed flowering.

qFT12.1 was delimited to a 2703-kb interval between the markers BARCSOYSSR_12_0220 and BARCSOYSSR_12_0368 on chromosome 12. Kuroda

and colleagues identified a flowering time QTL in the proximal region of *qFT12.1* using F_2 and BC_1F_1 mapping populations derived from crosses between a wild soybean accession (JP110755) and a cultivated soybean variety 'Fukuyutaka' (Kuroda *et al*. 2013). According to the Genetic Resources Center, National Agriculture and Food Research Organization, Japan (http://www.gene.affrc.go.jp/index_en.php), the wild soybean accession used in the study of Kuroda *et al*. (2013) was from Hiroshima prefecture, Japan, about 300 km away from Kinki, the origin of the wild soybean accession JWS156-1 used in the present study. It will be very interesting to verify whether the QTL detected by Kuroda *et al*. (2013) coincides with *qFT12.1*.

Among the other QTLs detected in the present study, *qFT19.1* accounted for 10.54% and 10.46% of the total phenotypic variation in flowering time of the CSSL population in 2016 and 2017, respectively.. Flowering time QTLs in the same region of *qFT19.1* has previously been detected in many studies (Githiri *et al*. 2007; Mansur *et al*. 1993; Orf *et al*. 1999), indicating consistency across different populations and environmental conditions.

Since the expression of *Glyma.12G073900* (*PRR7-like* gene) gene showed significant difference in 40-day-old leaves between 'Jackson' and 'JWS156-1', *Glyma.12G073900* was a potential candidate gene underlying flowering time. *Glyma.12G073900* is orthologous to the *Arabidopsis APRR7* gene, which belongs to the APRR1/TOC1 quintet and regulates circadian clock in *Arabidopsis*. However, it remains unclear how the APRR1/TOC1 to regulate the mechanisms underlying the circadian rhythm (Yamamoto *et al*. 2003). Further study will be required to identify the causal gene and to reveal the mechanism of flowering time in soybean to contribute to the future breeding strategies for genetic improvement of soybean.

Table 4. 1 Statistics analysis of flowering time (FT) in the BC_3F_5 CSSL populations in 2016 and 2017

Year	Population	Trait	Parents		BC_3F_5 CSSLs								
			Jackson	JWS156-1	SampleSize	Mean	Variance	StdError	Skewness	Kurtosis	Minimum	Maximum	Range
2016	BC_3F_5	FT (d)	50	67	120	52.43	3.72	1.93	0.83	0.23	48.5	57.5	9
2017	BC_3F_5	FT (d)	47	65	120	49.31	4.82	2.2	1.15	1.39	46	57	11

SampleSize: Size of the mapping population.

Variance: Variance of the phenotypic trait.

StdError: Standard deviation of the phenotypic trait.

Skewness is a measure of the phenotypic trait of the probability distribution of a real-valued random variable about its mean.

Kurtosis is a measure of the phenotypic trait of the probability distribution of a real-valued random variable.

Table 4. 2 SSR markers associated with flowering time in the BC_3F_5 CSSL population in 2016 and 2017

Year	Marker Name	QTL	Chr. (LG)	Position (bp/cM)	LOD[a]	PVE (%)[b]	Add[c]	M (QQ)[d]	M (qq)[e]
2016	Satt192	*qFT12.1*	Gm12 (H)	5,965,123/44.04	15.16	38.27	2.12	57.73	53.5
	Satt678	*qFT19.1*	Gm19 (L)	43,032,409/70.19	5.09	10.46	1.6	57.22	54.02
2017	BARCSOYSSR_07_0262	*qFT07.1*	Gm7 (M)	4,993,961	4.12	6.39	2.26	58.3	53.78
	Satt192	*qFT12.1*	Gm12 (H)	5,965,123/44.04	17.73	36.37	2.39	58.43	53.65
	Satt293	*qFT12.2*	Gm12 (H)	36,045,835/80.08	3.78	5.83	1.05	57.09	54.99
	Satt678	*qFT19.1*	Gm19 (L)	43,032,409/70.19	6.48	10.54	1.86	57.9	54.18

[a]LOD: LOD score calculated from single marker analysis.

[b]PVE (%): Phenotypic variation explained by the marker.

[c]Add: Estimated additive effect of the marker.

[d]M (QQ): Mean value of the QTL genotype of donor parent.

[e]M (qq): Mean value of the QTL genotype of recurrent parent.

Table 4. 3 Primer sequences for qRT-PCR of the relative expression of *Glyma.12G073900.1* and *Glyma.12G073300.1*

Primer Name	Sequence 5' to 3'
GmActin-qRT-F	GAGCTATGAATTGCCTGATGG
GmActin-qRT-R	CGTTTCATGAATTCCAGTAGC
GmTOE2-qRT-F	GGAGGGGACAGTGAGATGAA
GmTOE2-qRT-R	GTATTGCGAGCTTCGAGACC
GmPRR7-qRT-F	GCTCAGCTGCACCAAATACA
GmPRR7-qRT-R	GAGGCCCTGTCATTGATGTT

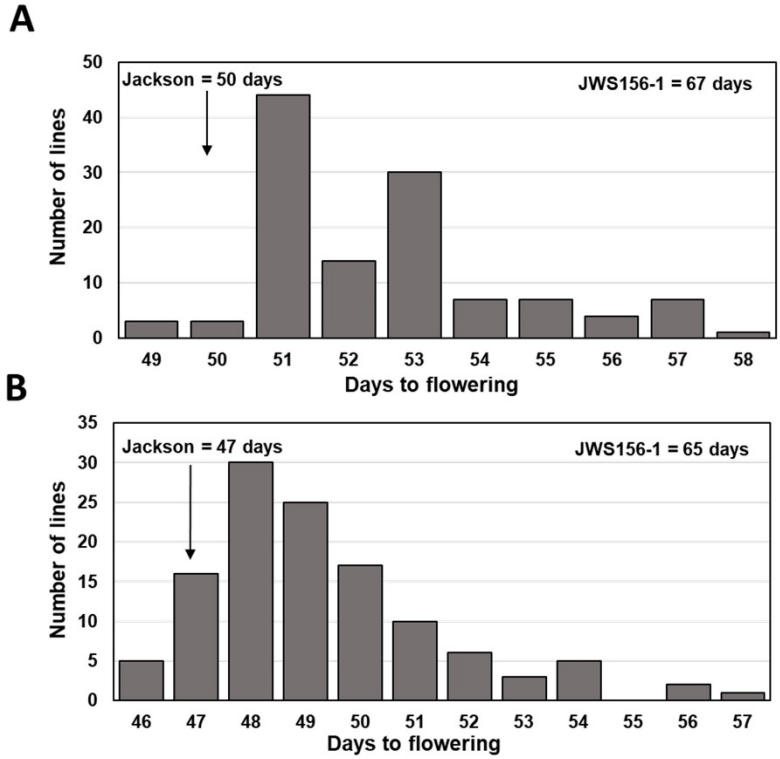

Figure 4. 1 Frequency distribution of flowering time of BC_3F_5 chromosome segment substitution lines (CSSLs; *n* = 120) in 2016 (A) and 2017 (B).

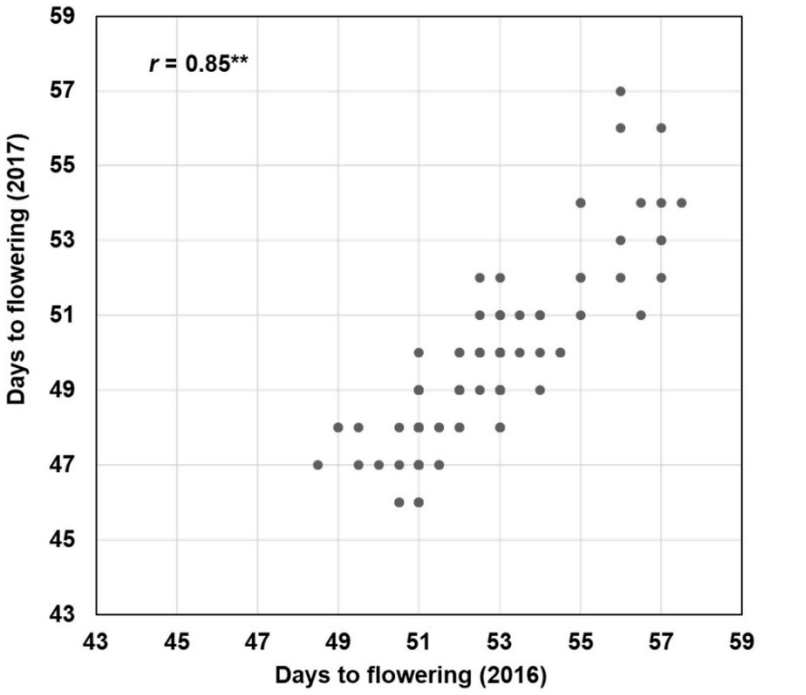

Figure 4. 2 The correlation of the flowering time of BC_3F_5 chromosome segment substitution lines (CSSLs) in 2016 and 2017 years. ** Indicates statistically significant correlation at ($P < 0.01$).

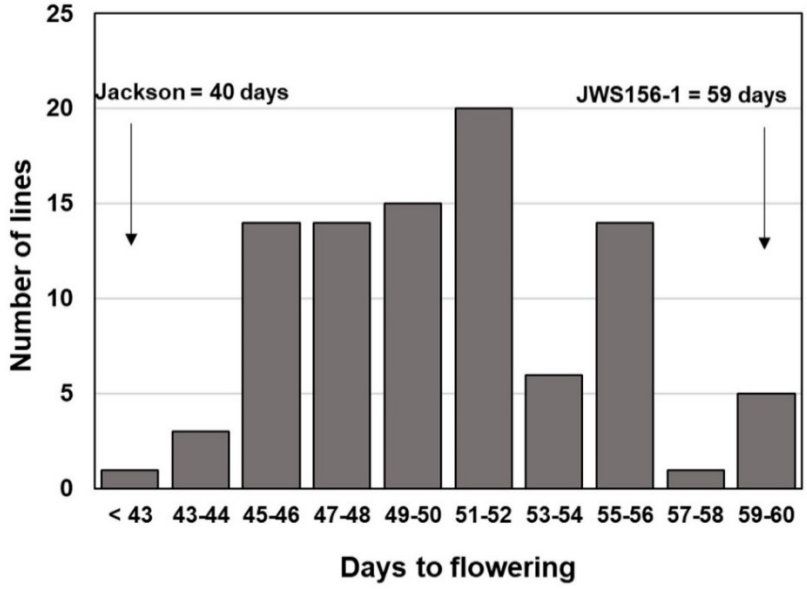

Figure 4. 3 Frequency distribution of flowering time of 94 recombinant inbred lines (RILs) ('Jackson' × 'JWS156-1') in 2017.

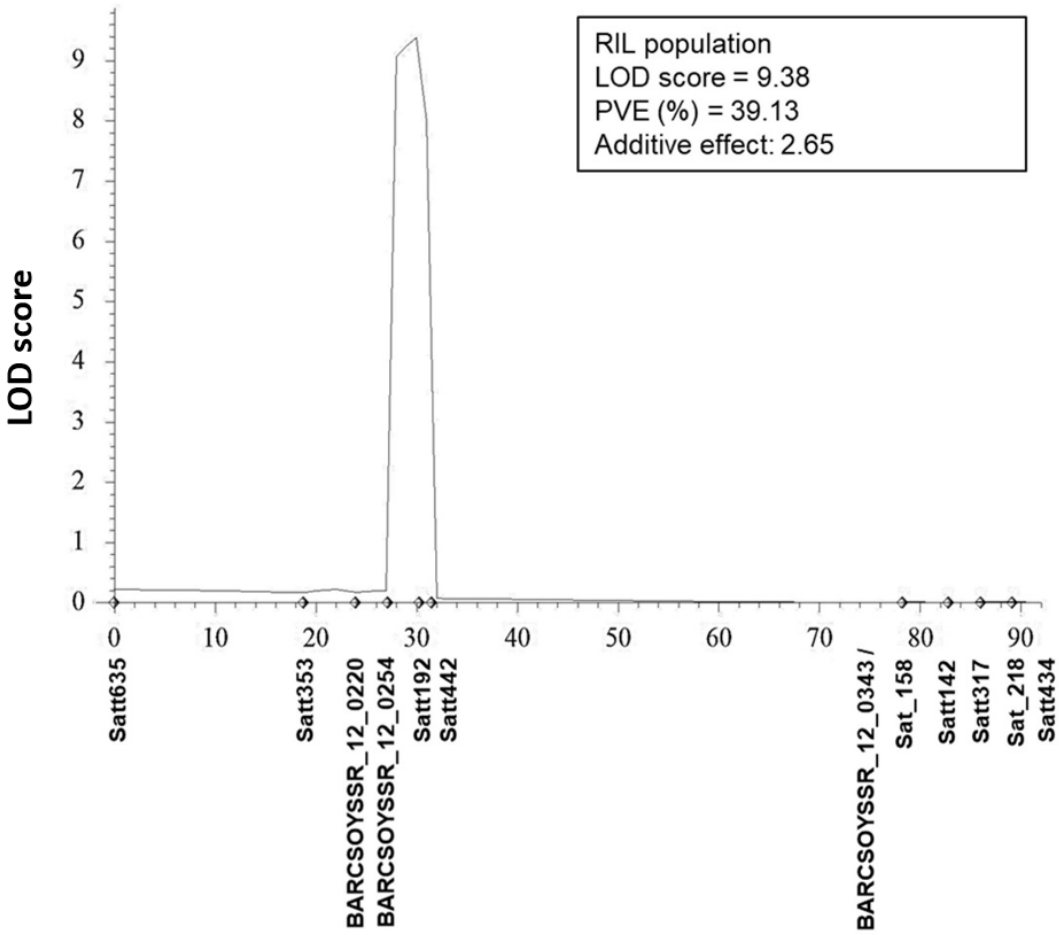

Figure 4. 4 A QTL for flowering time mapped on chromosome 12 in a RIL population derived from 'Jackson' × 'JWS156-1'.

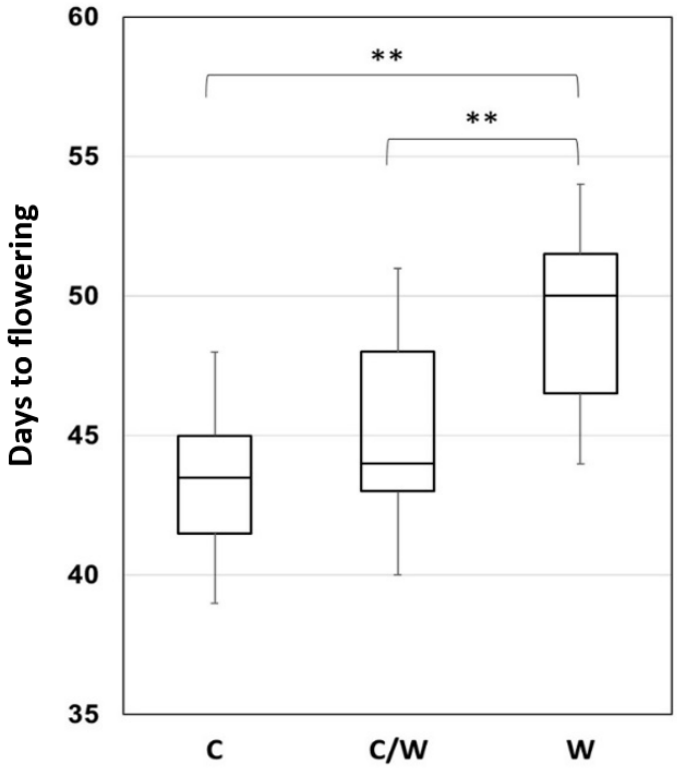

Figure 4. 5 Box plots showing flowering time of three plant genotype groups derived from the residual heterozygous line RHL509. C, 'Jackson' homologous group; W, 'JWS156-1' homologous group; C/W, heterozygous group. The interquartile region, median, and range are indicated by the box, bold horizontal line, and vertical line, respectively. ** indicates statistically significant differences between means ($P < 0.01$).

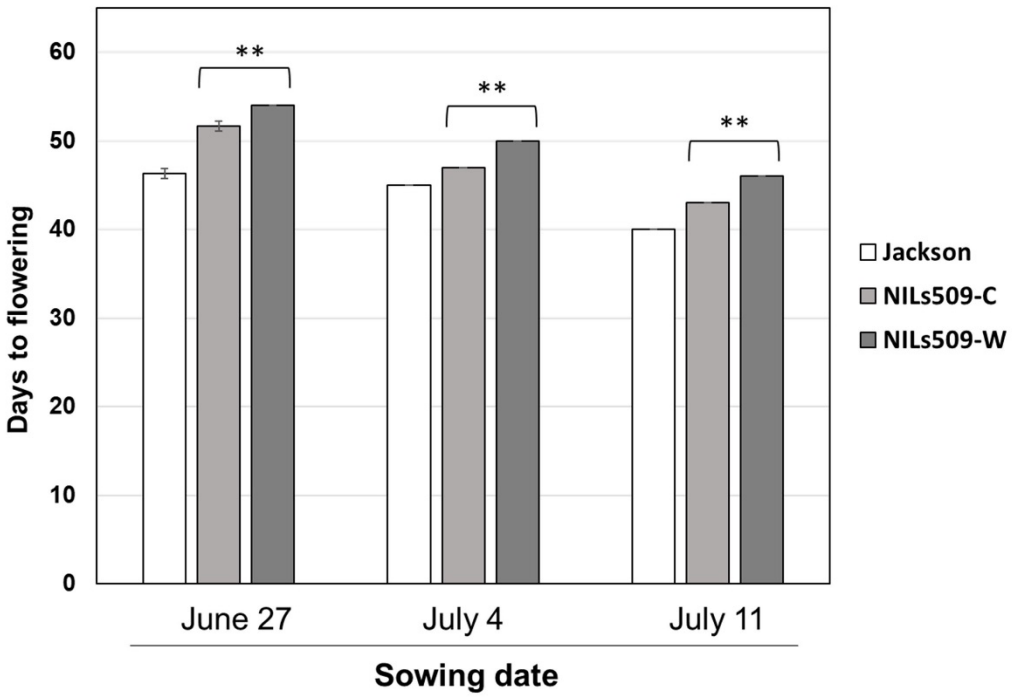

Figure 4. 6 Allelic effects of *qFT12.1* in near-isogenic lines (NILs) sown in the field on different dates in 2017. Data represent means ± standard deviation. ** indicates statistically significant differences between means ($P < 0.01$).

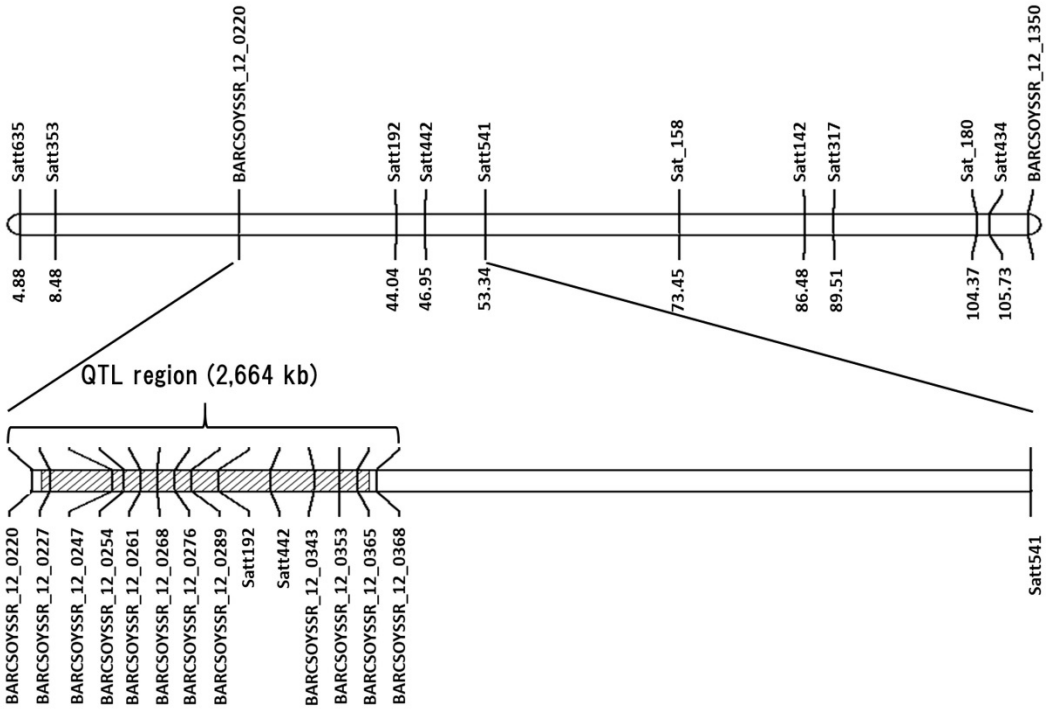

Figure 4. 7 Schematic showing the genomic location of the flowering time QTL *qFT12.1* on soybean chromosome 12. The QTL was delimited to a 2,703-kb interval between the SSR markers BARCSOYSSR_12_0220 and BARCSOYSSR_12_0368.

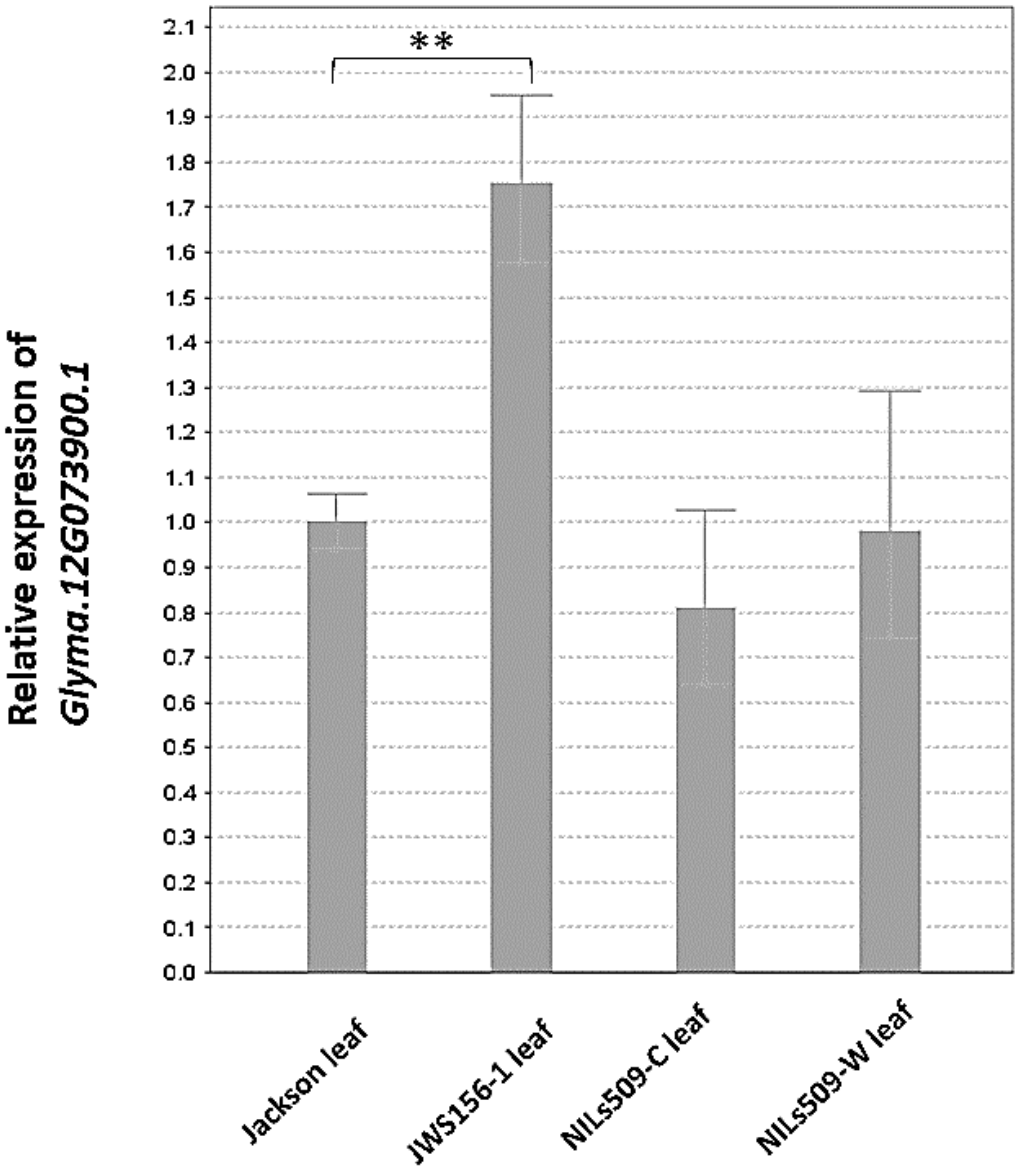

Figure 4. 8 The relative expression of *Glyma.12G073900.1* (*PRR7-like* gene) in the parental lines 'Jackson' and 'JWS156-1' and the *qFT12.1* near-isogenic lines (NILs) NILs509-C and NILs509-W. The RNA was extracted from the 40-day-old leaf of 'Jackson', 'JWS156-1', NILs509-C, and NILs509-W. ** Indicates significant difference between the means (*P-value* < 0.01).

CHAPTER 5

Identification and validation of QTLs for seed coat cracking (SCC) using two CSSL populations

Seed coat cracking (SCC) brings serious negative effects on the seed quality and their commercial value in soybean. The SCC can be caused by adverse environmental stresses. However, genetic mechanisms for controlling SCC in soybean is still remain unclear. To identify new QTLs controlling SCC, the 120 BC_3F_5 CSSLs were cultivated over 3 years under field conditions. SCC for each line was investigated by counting the cracked seeds in 200 health seeds harvested from each CSSLs. A total of four QTLs (*SCC07.1*, *SCC08.1*, *SCC13.1*, and *SCC20.1*) were identified on four chromosomes. Of these, one QTL (*SCC08.1*) located on chromosomes 8 showed major effects by explaining 20.59–53.13% of phenotypic variances in three years. To further confirm the major QTL SCC08.1, SCC was investigated in a BC_3F_5 population (n = 88) derived from a cross between a Japanese cultivar 'Tachiyutaka' and the wild soybean 'JWS156-1', followed by continuous backcrossing using ' Tachiyutaka ' as a recurrent parent. A QTL, explaining 19.86 % of the total phenotypic variances, was detected in the same position of *SCC08.1* in the Tachiyutaka × JWS156-1 BC_3F_5 population, suggesting that *SCC08.1* has effect in different genetic backgrounds.

CHAPTER 6

General discussion

6.1 The wild soybean CSSLs are potential for exploring and utilizing the favorable alleles in wild soybean for soybean improvement

Serials of studies have revealed that wild soybean has a higher genetic diversity than cultivated soybean (Hyten et al. 2006; Lam *et al*. 2010; Li *et al*. 2010; Xu *et al*. 2002; Zhou *et al*. 2015a), implying that wild soybean is potential for enlarging the genetic diversity of cultivated soybean. Until now, several QTLs alleles or genes controlling seed quality, yield traits, and tolerances to biotic and abiotic stresses have been identified in wild soybean and then introduced into cultivated soybean (Concibido *et al*. 2003; Do *et al*. 2016; Hu *et al*. 2013; Sebolt *et al*. 2000; Tuyen *et al*. 2010; Wang *et al*. 2016c), demonstrating that wild soybean is an important genetic resource for the improvement of cultivated soybean. Advanced backcross QTL (AB-QTL) analysis provided a more effective strategy for detecting detect QTLs associated with traits of interest, particularly for mining useful gene from wild species (Grandillo and Tanksley 2005). So far, numbers of wild species CSSL populations have been developed in rice, tomato, and other crops (Eshed and Zamir 1995; Tang *et al*. 2018; Wang *et al*. 2016a). However, only three CSSL populations, including two wild soybean CSSL populations, have been reported for in soybean (Wang *et al*. 2013; Watanabe *et al*. 2018; Xin *et al*. 2016). In this study, I was involved in the construction of 'Jackson' × 'JWS156-1' BC_3F_5 CSSL populations from the stage of fifth self-pollination, and identified the genetic background of each line using DNA marker analysis, and successfully completed the third wild soybean CSSL population in the world. Based on the analysis of a total of 235 SSR markers from the 20

chromosomes, the CSSLs genomic region were 92.9% recovered by the recurrent parent 'Jackson' after 3 times of backcrosses. By using the CSSLs, major and stable QTLs for some important agronomic traits (100 seed weight, flowering time, and seed coat cracking) were identified, demonstrating that the CSSLs are potential germplasm for soybean genetic studies. The most interesting results are the findings of a novel 100SW and a major flowering time QTL, which may correspond to a new flowering time gene. During the domestication process from wild soybean to cultivated soybean, many elite genes have been lost due to natural or artificial selections. The wild soybean CSSLs provide an efficient strategy for mining such elite genes in wild soybean and intruded and utilized them in cultivated. Just like the QTL analysis for the traits of 100SW, flowering time, and SCC performed in my current study, our ongoing researches with the CSSLs focusing on other important agronomic traits, such as biomass, biotic and abiotic stresses, and other yield related traits, will enable us to find and utilize more favorable alleles or genes in wild soybean.

6.2 QTLs analysis for 100-seed weight: not only confirmed previous reported QTLs but also discovered a new one

In agreement with the previous studies (Liu *et al.* 2007; Wang *et al.* 2016c; Xin *et al.* 2016), alleles from wild soybean contribute to a reduction in seed weight in all the nine 100SW QTLs. Soybean breeders are not simply interested in increasing the seed size, but require an optimized combination for each yield factor such as seed size, the number of seeds per pod, and the number of pods per plant. The soybean varieties developed in the tropical and subtropical countries, such as India and Indonesia, tend to have a small seed size compared with varieties from the temperate regions, such as USA, China, and

Japan. In addition, seed size is also an important factor for soybean end-use. Although a small seed size is preferred for natto and soy sprouts, a large seed size is preferred for vegetable soybean and tofu. Therefore, depending on the end-use or location of growth, different soybean varieties with different seed sizes are needed. The QTLs detected in the present study would be useful to control seed size via genomic breeding by design and positional cloning of relevant genes.

I detected three major and stable QTLs (*qSW12.1*, *qSW17.1*, and *qSW14.1*). There QTLs are good candidates for cloning the genes responsible for 100SW because a major QTL can be steadily detected in different environmental conditions. *qSW12.1* was identified in 3 successive years, accounting for 15.25%–23.25% of the total phenotypic variation of 100SW and presenting a novel locus associated with this trait. This QTL might differ from a seed weight QTL previously detected on chromosome 12 (Han *et al.* 2012), occurring in an interval between Satt181 (91.12 cM, genetic position based on SoyBase and so forth) and Satt302 (81.04 cM), where the proximal marker is approximately 10 cM away from *qSW12.1*.

The *qSW14.1* QTL was detected over 3 successive years, accounting for 6.27%–11.17% of the total phenotypic variation of 100SW. Seed weight QTL has been previously detected in the same region by Hoeck *et al.* (2003) and Li *et al.* (2007a), indicating consistency across different environmental conditions and genetic backgrounds.

Although the *qSW17.1* QTL was detected over 2 years (2014 and 2015) as a major QTL with LOD scores of 8.99 and 7.16, respectively, it was not detected in 2016. As it was shown in Table 3. 2, the 120 CSSLs as well as the two parents showed the lowest 100 seed weight in 2016 among the three years. The special growth condition in 2016

might affect the effect of *qSW17.1*. The other researchers working on different populations have consistently detected QTLs for seed weight in this region (Liu *et al.* 2007; Kato *et al.* 2014; Wang *et al.* 2016c; Xin *et al.* 2016). Recently, Lu *et al.* (2017b) identified a protein phosphatase type 2C-1 (PP2C-1) allele as the causative gene underlying a QTL that contributed a positive effect for increasing seed weight. This QTL located in the same region of *qSW17.1*. However, the result in Lu *et al.* (2017b) differs from my study in which the wild soybean contributed a negative effect, decreasing seed weight. It would be very interesting to determine whether these conflicting results are due to different genes or different alleles within the same locus of the wild soybean genome.

As for the remaining significant QTLs, six were detected in this study (*qSW8.1, qSW9.1, qSW13.1, qSW16.1, qSW17.2,* and *qSW20.1*). Of these, *qSW20.1* might correspond to a region associated with seed weight reported by Han *et al.* (2012) and Liu *et al.* (2011). However, these six QTLs require further corroboration because they were detected in 1 year. Although some CSSLs demonstrated higher 100SW than that by the cultivated parent 'Jackson', no positive effect allele was detected from the wild soybean, suggesting the presence of undetected small effect QTLs or epistasis.

Despite the fact that, on average, more than 90% of the genome was recovered by the recurrent parent in the BC3-derived CSSLs, there were several CSSLs with quite low 100SW. For example, CSSLs507 and CSSLs505 had an average of 11.06 ± 1.30 g and 12.74 ± 2.71 g for 100SW over the three years. Genotyping analysis showed that CSSLs507 had the 'JWS156-1' genotypes at four 100SW QTLs (*qSW8.1*, *qSW12.1*, *qSW13.1*, and *qSW17.1*) and CSSLs505 had the 'JWS156-1' genotypes at three 100SW QTLs (*qSW9.1*, *qSW12.1*, and *qSW17.1*), both including the two major QTLs, *qSW12.1* and *qSW17.1*. Further genetic analysis for 100SW using a segregation population (BC_4)

derived from crosses between these CSSLs and 'Jackson' could enable us to understand the effects and relationships of the 100SW QTLs.

Analysis of the *qSW12.1* near-isogenic lines, NILs564-C and NILs564-W, not only confirmed the effect of *qSW12.1* but also delimited it in a 1,348-kb interval between the BARCSOYSSR_12_1282 and BARCSOYSSR_12_1347 markers on chromosome 12. However, it is still difficult to identify the candidate gene underlying *qSW12.1* from the 129 predicted genes.

My study has detected the major QTLs again, *qSW17.1* and *qSW20.1*, and identified a novel major QTL, *qSW12.1*, within a 1,348-kb interval region on chromosome 12. This mapping is sufficiently accurate for DNA marker assisted selection in a breeding program. While the region of *qSW12.1* is currently too large for positional cloning of the casual gene underlying the QTL, my ongoing fine-mapping work using a much larger segregation population will enable to narrow down the *qSW12.1* QTL into a smaller chromosome region for identifying the candidate gene causal to the seed weight trait. Compared to genetic studies of seed weight in other plants, such studies are very limited in the soybean. Cloning genes controlling seed weight will not only contribute to a better understanding of the mechanism of seed development but also provide important implications for breeding strategies of soybean.

Although positive correlations between 100SW and grain yield have been confirmed in previous studies (Burris *et al.* 1973; Smith and Camper 1975), the low correlations between grain yield and seed weight were observed in several segregating populations (Burton 1987). The grain yield decreased by 100SW might be compensated for by increasing the number of pods or seeds per plant. For example, *ln* locus associated decreased 100SW had function to increase the number of pods per plant (Jeong *et al*.

2012). Further studies are needed to confirm the effects of *qSW12.1* QTL on other grain yield traits, such as number of seeds per pod, number of pod per plant, and number of seeds per plant.

6.3 A major flower time QTL, *qFT12.1,* may correspond to a new flowering time gene in soybean

In the present study, I detected, validated, and characterized the effect of a major flowering time QTL, *qFT12.1*, in soybean. This new flowering time gene will contribute toward a better understanding of the regulation of flowering time in soybean. This gene could also be introduced into cultivated soybean, particularly for soybean breeding at lower latitudes to develop varieties adapted to the local environmental conditions.

The *qFT12.1* QTL was detected in 2 successive years as a major and stable QTL on chromosome 12 in the present study. Of the 107 flowering time QTLs deposited in Soybase (https://soybase.org/sbt/), only one QTL has been reported on chromosome 12 (Kuroda *et al*. 2013). Kuroda and colleagues identified a flowering time QTL in the proximal region of *qFT12.1* using F_2 and BC_1F_1 mapping populations derived from crosses between a wild soybean accession (JP110755) and a cultivated soybean variety 'Fukuyutaka' (Kuroda *et al*. 2013). According to the Genetic Resources Center, National Agriculture and Food Research Organization, Japan (http://www.gene.affrc.go.jp/index_en.php), the wild soybean accession used in the study of Kuroda *et al*. (2013) was from Hiroshima prefecture, Japan, about 300 km away from Kinki, the origin of the wild soybean accession 'JWS156-1' used in the present study. It would be very interesting to verify whether the QTL detected by Kuroda *et al*. (2013) coincides with *qFT12.1*.

Among the other QTLs detected in the present study, *qFT19.1* accounted for 10.54% and 10.46% of the total phenotypic variation in flowering time of the CSSL population in 2016 and 2017, respectively. Flowering time QTLs in the same region of *qFT19.1* has previously been detected in many studies (Githiri *et al*. 2007; Mansur *et al*. 1993; Orf *et al*. 1999), indicating consistency across different populations and environmental conditions. The region near *qFT19.1* also harbors the soybean stem growth habit gene, *Dt1* (Liu *et al*. 2010; Tian *et al*. 2010). *Dt1* is an ortholog of the *Arabidopsis Terminal Flower 1* (*TFL1*) gene. One of the *TFL1* gene family members, *FLOWERING LOCUS T* (*FT*) is a key regulator of flowering time (for review, see Benlloch *et al*. 2007). Given that the soybean cultivar 'Jackson' has a determinate stem growth habit and wild soybean has a typical indeterminate stem growth habit, *qFT19.1* most likely corresponds to the *Dt1* locus. Two additional flowering time QTLs, *qFT12.2* and *qFT07.1*, were detected in the present study. Several QTLs affecting flowering time, maturity, and yield-related traits have previously been reported around these two QTL regions (Han *et al*. 2012; Orf *et al*. 1999; Panthee *et al*. 2007; Wang *et al*. 2004). However, *qFT12.2* and *qFT07.1* showed relatively small effects and were only detected in 2017. Further research is needed to confirm the effects of *qFT12.2* and *qFT07.1* on flowering time.

Glyma.12G073300 was selected as a candidate because it was the best BLAST hit to *AtTOE1*. TOE1 and TOE2 have been shown to delay flowering in *Arabidopsis* by interacting with a subset of JAZ proteins and repressing the expression of the *FLOWERING LOCUS* (*FT*) gene (Zhai *et al*. 2015). To examine the genes regulating flowering time in the soybean genome, Jung and colleagues conducted BLAST searches of the soybean genes against the *Arabidopsis* genome and identified *Glyma12g07800* (old name of *Glyma.12G073300* in *G. max 1.09* version) as the soybean ortholog of *AtTOE1*

(Jung *et al*. 2012). Another annotated gene, *Glyma.12G073900*, was also considered a potential candidate gene underlying the *qFT12.2* QTL. *Glyma.12G073900* is orthologous to the *Arabidopsis APRR7* gene, which belongs to the APRR1/TOC1 quintet and regulates circadian clock in *Arabidopsis*. APRR7 is a member of a small protein family, *Arabidopsis* PSEUDO-RESPONSE REGULATORS. However, it remains unclear how the APRR1/TOC1 quintet regulate the mechanisms underlying the circadian rhythm (Yamamoto *et al*. 2003).

Although *Glyma.12G073300* and *Glyma.12G073900* were selected as candidate genes underlying *qFT12.2*, the target gene is not necessarily limited to these two genes. For instance, *Glyma.12G074100*, which shows homology to the *Arabidopsis* phototropin gene, may also be considered a candidate gene underlying the *qFT12.2* QTL. The phototropins (*phot1* and *phot2*) represent a family of genes encoding blue light receptors that control a range of responses to light. Recent genetic analyses show that *phot1* and *phot2* exhibit partially overlapping functions in mediating phototropism, chloroplast migration, and stomatal opening in *Arabidopsis* (for review, see Briggs & Christie 2002). On the basis of RAD-seq genotyping of a natural soybean population comprising 286 accessions, Zhou *et al*. (2015b) identified four genes (*Glyma.12G076700*, *Glyma.12G076800*, *Glyma.12G077300*, and *Glyma.12G077400*) on chromosome 12 that were associated with flowering time. These genes are also located in the *qFT12.1* QTL region, making them potential candidates for the effect of *qFT12.1*. My ongoing fine mapping experiments will assist in further narrowing down the QTL to a smaller region to enable better prediction, and subsequently cloning of the gene underlying *qFT12.1*.

In the present study, I identified, validated, and characterized a major QTL, *qFT12.1*, controlling flowering time. Further studies are needed to determine the effect of *qFT12.1*

in other genetic backgrounds and its interaction with other flowering time genes and with environmental conditions, such as temperature and light quality. To date, 11 genes controlling flowering time have been reported in soybean. *qFT12.1* may correspond to a new flowering time gene in soybean. On the basis of the series of results in this field, I propose that the flowering time gene corresponding to *qFT12.1* be named *E11*. Nonetheless, the identification of the *qFT12.1* QTL will not only contribute to a better understanding of the molecular basis of the photoperiod-dependent regulation of flowering time but will also be potentially useful in improving the adaptability and productivity of cultivated soybean by properly combining with other flowering time genes.

Summary

Wild soybean (*Glycine soja* Sieb. & Zucc.) has been found to have the potential to improve cultivated soybean (*Glycine max* (L.) Merr.) in terms of seed quality, yield traits, and tolerances to biotic and abiotic stress. Several QTLs (quantitative trait loci) or genes have been identified in wild soybean and introduced into cultivated soybean. Previous QTL analysis in soybean were mainly performed by primary mapping populations (F_2, $F_{2:3}$, RILs). However, the advanced mapping populations, such as chromosome segment substitution lines (CSSLs), are more powerful for identifying favorable alleles from wild soybean. In this study, I developed a BC_3F_5 CSSL population (n = 120) from a cross between a cultivated soybean cultivar 'Jackson' and a wild soybean accession 'JWS156-1', followed by continuous backcrossing using 'Jackson' variety as a recurrent parent. Based on analysis of a total of 235 SSR markers form the 20 chromosomes, the CSSLs were almost recovered by the recurrent parent 'Jackson' after three times of backcrosses and no line with abnormal growth were observed in the 120 CSSLs. The proportion of the recurrent parent 'Jackson' alleles in each CSSL ranged from 80.3% to 99.2% with an average of 92.9 ± 4.0%. This BC_3F_5 CSSL population was used to identify QTLs associated with several important agronomic traits, such as seed weight, flowering time, and seed coat cracking to gain a better understanding of the genetic basis of these traits and their applications in soybean breeding practice.

The seed weight, generally expressed as 100-seed weight (100SW), is one of the most important traits controlling yield in the soybean. To identify QTL for 100SW, the 120 BC_3F_5 CSSLs of soybean were cultivated over 3 years. A total of nine QTLs (*qSW8.1*, *qSW9.1*, *qSW12.1*, *qSW13.1*, *qSW14.1*, *qSW16.1*, *qSW17.1*, *qSW17.2*, and *qSW20.1*) were

detected on eight chromosomes. Of these, *qSW12.1* (LOD = 6.78-12.31) was detected over the 3 successive years on chromosome 12 as a novel, stable, and major QTL. To validate the effect of *qSW12.1*, a residual heterozygous line (RHL) RHL564, which showed heterozygous at the *qSW12.1* region, was selected from the BC_3F_5 population. Among the two homologous progenies produced by self-pollination of RHL564, higher seed weight was observed in 'Jackson' genotype plants than that in 'JWS156-1' genotype plants. Analyses of RHL564 delimited *qSW12.1* in an interval of approximately 1,348 kb between the BARCSOYSSR_12_1282 and BARCSOYSSR_12_1347 markers on chromosome 12. Out study not only confirmed previous reported 100SW QTLs, such as *qSW14.1* and *qSW17.1,* but also discovered a new and stable 100SW QTL (*qSW12.1*).

Soybean is highly sensitive to photoperiod. To improve the adaptability and productivity of soybean, it is essential to understand the molecular mechanisms regulating flowering time. To identify new flowering time QTLs, I evaluated the BC_3F_5 CSSL population over 2 years under field conditions. Four QTLs (*qFT07.1*, *qFT12.1*, *qFT12.2*, and *qFT19.1*) were detected on three chromosomes. Of these, *qFT12.1* showed the highest effect, accounting for 36.37–38.27% of the total phenotypic variation over 2 years. This QTL was further confirmed in the F_7 recombinant inbred line population (n = 94) derived from the same cross ('Jackson' × 'JWS156-1'). Analysis of the *qFT12.1* BC_3F_5 residual heterozygous line RHL509 validated the allele effect of *qFT12.1* and revealed that the recessive allele of *qFT12.1* resulted in delayed flowering. Evaluating the *qFT12.1* near-isogenic lines (NILs) under different growth conditions showed that NILs with the wild soybean genotype always showed later flowering than those with the cultivated soybean genotype. *qFT12.1* was delimited to a 2703-kb interval between the markers BARCSOYSSR_12_0220 and BARCSOYSSR_12_0368 on chromosome 12. According

to the reference genome sequence of soybean William 82, a total of 267 annotated genes were present in this region. Several genes from this region, such as *Glyma.12G073300* and *Glyma.12G073900*, were potential candidate gene underlying *qFT12.1*. To date, 11 genes controlling flowering time have been reported in soybean. *qFT12.1* detected in the present study may correspond to a new flowering time gene in soybean.

In this study, a wild soybean CSSL population was developed and characterized. These materials are very useful to exploit and utilize the favorable alleles in wild soybean. By using the CSSL population, QTLs or candidate genes controlling some important agronomic traits were identified. These results will not only contribute to a better understanding of the molecular basis of these traits, but will also be potentially in improving the adaptability and productivity of soybean.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Yasunari Fujita, Professor of University of Tsukuba and Senior researcher in JIRCAS, for providing me opportunity to study in Japan. In particular, I would like to express my grateful acknowledges to Dr. Donghe Xu, Project Leader in JIRCAS. His guidance and encouragement of the experiments really helped me a lot in these three years. I would really appreciate it. I also would like to express my gratitude to Drs. Ryo Ohsawa, Junichi Tanaka, and Yosuke Yoshioka, Professor, Associated Professor, and Assistant Professor of University of Tsukuba, respectively, for providing me with so many useful advices in this thesis. I would like to thank Ms. Kazuko Suzuki, Ms. Mieko Noda, Ms. Tomoko Kobayashi, Dr. Dong Cao, and Mr. Yongliang Yan for their supports for the experiments, Ms. Masami Toyoshima, Ms. Kumiko Ozawa, and Ms. Hiroko Saitou for editorial technical assistance. I also would like to thank Drs. Takuya Ogata, Chihiro Yamamizo, and Yukari Nagatoshi, Researchers in JIRCAS, for critical review of this thesis. It is a great pleasure to acknowledge to all of the members supporting me in University of Tsukuba and JIRCAS. A very special thanks to Professor. Qingyu Wang of Jilin University, China, for providing me support of my application of China Scholarship Council (CSC). Finally, I want to express thanks to my parents, my wife and my friends for their kindly supports and encouragements. This study was financially supported by a scholarship from the China Scholarship Council to D. Q. Liu.

References

Abe, I., H. Morita (2010) Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural Product Reports*, 27, 809–838.

Abe, J., D. H. Xu, Y. Suzuki, A. Kanazawa, Y. Shimamoto (2003) Soybean germplasm pools in Asia revealed by nuclear SSRs. *Theoretical and Applied Genetics*, 106, 445-453.

Benitez E. R., H. Funatsuki, Y. Kaneko, Y. Matsuzawa, S. W. Bang, R. Takahashi (2004) Soybean maturity gene effects on seed coat pigmentation and cracking in response to low temperatures. *Crop Science* 44, 2038–2042.

Bernard, R. L. (1971) Two major genes for time of flowering and maturity in soybeans. *Crop Science*, 11, 242–244.

Bernard, R. L., M. G. Weiss (1973) Qualitative genetics. In: Caldwell BE (ed) Soybean: improvement, production, and uses. American Society of Agronomy, Madison, pp 117–154.

Bonato, E. R., N. A. Vello (1999) *E6*, a dominant gene conditioning early flowering and maturity in soybeans. *Genetics Molecular Biology*, 22, 229–232.

Briggs, W. R., J. M. Christie (2002). Phototropins 1 and 2: versatile plant blue-light receptors. *Trends in Plant Science*, 7, 1360–1385.

Burchett, C. A., W. T. Schapaugh, C. B. Overley, T. L. Walter (1985) Influence of etched seed coats and environmental conditions on soybean seed quality. *Crop Science*, 25, 655–660.

Burris, J.S., O.T. Edje, A. H. Wahab (1973) Effects of seed size on seedling performance in soybeans: II. seedling growth and photosynthesis and field performance. *Crop*

Science, 13, 207–210.

Burton, J.W. (1987) Quantitative genetics: Results relevant to soybean breeding. *In*: Wilcox, J.R. (ed.) Soybeans: improvement, production, and uses, 2nd edn. American Society of Agronomy, Madison, 211–247.

Buzzell, R. I. (1971) Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Canadian Journal of Genetics and Cytology*, 13, 703–707.

Buzzell, R. I., H. D. Voldeng (1980) Inheritance of insensitivity to long daylength. *Soybean Genetics Newsletter*, 7, 26–29.

Cober, E. R., H. D. Voldeng (2001) A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Science*, 41, 698–701.

Cober, E. R., S. J. Molnar, M. Charette, H. D. Voldeng (2010) A new locus for early maturity in soybean. *Crop Science*, 50, 524–527.

Concibido, V. C., B. La Vallee, P. Mclaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, X. Delannay (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theoretical and Applied Genetics*, 106, 575–582.

Do, T. D., H. T. Chen, V. T. Hien, A. Hamwieh, T. Yamada, T. Sato, Y. L. Yan, H. Cong, M. Shono, K. Suenaga, D. H. Xu (2016) *Ncl* synchronously regulates Na^+, K^+, and Cl^- in soybean and greatly increases the grain yield in saline field conditions. *Scientific Reports*, 6, 19147.

Duan, P., S. Ni, J. M. Wang, B. L. Zhang, R. Xu, Y. X. Wang, H. Q. Chen, X. D. Zhu, Y. H. Li (2015) Regulation of *OsGRF4* by *OsmiR396* controls grain size and yield in rice. *Nature Plants,* 2, 15203.

Duke S. H., G. Kakefuda, C. A. Henson, N. L. Loeffler, N. M. Van Hulle (1986) Role of

the testa epidermis in the leakage of intracellular substances from imbibing soybean seeds and its implications for seedling survival. *Physiology Plant,* 68, 625–631.

Edwards, C. J., E. E. Hartwing (1971) Effect of seed size upon rate of germination in soybeans. *Agronomy Journal,* 63, 429–430.

Eshed, Y., D. Zamir (1995) An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, 141, 1147–1162.

Fehr, W. R., C. E. Caviness, D. T. Burmood, J. S. Pennington (1971) Stage of development descriptions for soybean, *Glycine max* (L.) Merrill. *Crop Science*, 11, 929–931.

Githiri, S. M., D. Yang, N. A. Khan, D. H. Xu, T. Komatsuda, R. Takahashi (2007) QTL analysis of low temperature induced browning in soybean seed coats. *Journal of Heredity*, 98, 360–366.

Grandillo, S., S. D. Tanksley (2005) Advanced backcross QTL analysis: results and perspectives. In: Tuberosa R., R.L. Phillips and M. Gale (eds.) Proceedings of the international congress “In the wake of the double helix: from the green revolution to the gene revolution”. Bologna, Italy, pp. 115–132.

Hamwieh, A., D. H. Xu (2008) Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans. *Breeding Science*, 58, 355-359.

Han, Y. P., D. M. Li, D. Zhu, H. Y. Li, X. P. Li, W. L. Teng, W. B. Li (2012) QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theoretical and Applied Genetics*, 125, 671–683.

Harlan, J. R. (1992) Crops and man, 2nd edn. Madison, WI: American Society of Agronomy, Inc. and Crop Science Society of America, Inc.

Hesler, L. S., D. A. Prischmann, K. E. Dashiell (2012) Field and laboratory evaluations of soybean lines against soybean aphid (hemipteran: aphididae). *Journal of Economic Entomology*, 105, 608-615.

Hoeck, J. A., W. R. Fehr, R. C. Shoemaker, G. A. Welke, S. L. Johnson and S. R. Cianzio (2003) Molecular marker analysis of seed size in soybean. *Crop Science*, 43, 68–74.

Hu, J., Y. X. Wang, Y. Fang, L. Zeng, J. Xu, H. Yu, Z. Shi, J. Pan, D. Zhang, S. Kang (2015) A rare allele of *GS2* enhances grain size and grain yield in rice. *Molecular Plant,* 8, 1455–1465.

Hu, Z. B., D. Zhang, G. Z. Zhang, G. Z. Kan, D. L. Hong, D. Y. Yu (2013) Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). *Breeding Science*, 63, 441–449.

Hufstetler, E. V., H. R. Boerma, T. E. Cater, H. J. Eari (2006) Genotypic variation for three physiological traits affecting drought tolerance in soybean. *Crop Science*, 47, 25-35.

Hymowitz, T. (1970) On the domestication of the soybean. *Economic Botany*, 24, 408-421.

Hyten, D. L., Q. J. Song, Y. L. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, R. C. Shoemaker, P. B. Cregan (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 45, 16666-16671.

Jeong, N., S. J. Suh, M. H. Kim, S. Lee, J. K. Moon, H. S. Kim, S. C. Jeong (2012) *Ln* is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell*, 24, 4807–4818.

Jung, C. H., C. E. Wong, M. B. Singh, P. L. Bhalla (2012) Comparative genomic analysis

of soybean flowering genes. *PloS One*, 6, e38250.

Kasai, A., K. Kasai, S. Yumoto, M. Senda (2007) Structural features of *GmIRCHS*, candidate of the *I* gene inhibiting seed coat pigmentation in soybean: implications for inducing endogenous RNA silencing of chalcone synthase genes. *Plant Molecular Biology*, 64, 467–479.

Kasai, A., S. Ohnishi, H. Yamazaki, H. Funatsuki, T. Kurauchi, T. Matsumoto, S. Yumoto, M. Senda (2009) Molecular mechanism of seed coat discoloration induced by low temperature in yellow soybean. *Plant & Cell Physiology,* 50, 1090–1098.

Kato, S., T. Sayama, K. Fuji, S. Yumoto, Y. Kono, T. Y. Hwang, A. Kikuchi, Y. Takada, Y. Tanaka, T. Shiraiwa, M. Ishimoto (2014) A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theoretical and Applied Genetics,* 127, 1365–1374.

Kong, F. J., H. Y. Nan, D. Cao, Y. Li, F. F. Wu, J. L. Wang, S. J. Lu, X. H. Yuan, E. R. Cober, J. Abe, B. H. Liu (2015) A new dominant gene *E9* conditions early flowering and maturity in soybean. *Crop Science*, 54, 2529–2535.

Kurauchi, T., A. Kasai, M. Tougou, M. Senda (2011) Endogenous RNA interference of chalcone synthase genes in soybean: formation of double-stranded RNA of *GmIRCHS* transcripts and structure of the 5’ and 3’ ends of short interfering RNAs. *Journal of Plant Physiology,* 168, 1264–1270.

Kuroda, Y., A Kaga., N. Tomooka, H. Yano, Y. Takada, S. Kato, D. Vaughan (2013) QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields. *Ecology and Evolution*, 3, 2150–2168.

Lam, H. M., X. Xu, X. Liu, W. B. Chen, G. H. Yang, F. L. Wong, M. W. Li, W. M. He, N. Qin, B. Wang, J. Li, M. Jian, J. Wang, G. H. Shao, J. Wang, S. S. M. Sun, G. Y.

Zhang (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature genetics*, 42, 1053-1059.

Li N, Li Y (2014) Ubiquitin-mediated control of seed size in plants. *Frontiers in Plant Science*, 5, 1–6.

Li, D. D., T. W. Pfeiffer, P. L. Cornelius (2007a) Soybean QTL for yield and yield components associated with *Glycine soja* alleles. *Crop Science*, 48, 571–581.

Li, Y. H., W. Li, C. Zhang, L. Yang, R. Z. Chang, B. S. Gaut, L. J. Qiu (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytologist*, 188, 242-253.

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

Liu, B. H., S. Watanabe, T. Uchiyama, F. J. Kong, A. Kanazawa, Z. J. Xia, A. Nagamatsu, M. Arai, T. Yamada, K. Kitamura, C. Masuta (2010) The soybean stem growth habit gene *Dt1* is an ortholog of *Arabidopsis TERMINAL FLOWER1*. *Plant Physiology*, 153, 198–210.

Liu, B. H., T. Fujita, Z. H. Yan, S. Sakamoto, D. H. Xu, J. Abe (2007) QTL mapping of domestication-related traits in soybean (*Glycine max*). *Annals of Botany*, 100, 1027–1038.

Liu, H. L. (1949) Inheritance of defective seed coat in soybean. *Journal Heredity,* 40, 317–322.

Liu, W. X., M.Y. Kim, K. J. Van, Y. H. Lee, H. L. Li, X. H. Liu, S. H. Lee (2011) QTL

identification of yield-related traits and their association with flowering and maturity in soybean. *Journal of Crop Science and Biotechnology*, 14, 65–70.

Lu, S. J., X. H. Zhao, Y. L. Hu, S. L. Liu, H. Y. Nan, X. M. Li, C. Fang, D. Cao, X. Y. Shi, L. P. Kong, T. Su, F. G. Zhang, S. C. Li, Z. Wang, X. H. Yuan, E. R. Cober, J. L. Weller, B. H. Liu, X. L. Hou, Z. X. Tian, F. J. Kong (2017a) Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nature Genetics*, 49, 773–779.

Lu, X., Q. Xiong, T. Cheng, Q. T. Li, X. L. Liu, Y. D. Bi, W. Li, W. K. Zhang, B. Ma, Y. C. Lai (2017b) A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Molecular Plant*, 10, 670–684.

Mansur, L. M., K. G. Lark, H. Kross, A. Oliveira (1993) Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theoretical and Applied Genetics*, 86, 907–913.

McBlain, B. A., R. L. Bernard (1987) A new gene affecting the time of flowering and maturity in soybeans. *Journal of Heredity*, 78, 160–162.

Meng, L., H. H. Li, L.Y. Zhang, J. K. Wang, (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal*, 3, 269–283.

Nakamura, T., D. Yang, S. Kalaiselvi, Y. Uematsu, R. Takahashi (2003) Genetic analysis of net-like cracking in soybean seed coats. *Euphytica,* 133, 179–184.

Nicholas C. D., J. T. Lindstrom, L. O. Vodkin (1993) Variation of proline rich cell wall proteins in soybean lines with anthocyanin mutations. *Plant Molecular Biology,* 21, 145–156.

Nichols, D. M., L. Z. Wang, Y. L. Pei, K. D. Glover, B. W. Diers (2006) Variability

among Chinese *Glycine soja* and Chinese and north American soybean genotypes. *Crop Science*, 47, 1289-1298.

Ohnishi, S., H. Funatsuki, A. Kasai, T. Kurauchi, N. Yamaguchi, T. Takeuchi, H. Yamazaki, H. Kurosaki, S. Shirai, T. Miyoshi, H. Horita, M. Senda (2011) Variation of *GmIRCHS* (*Glycine max* inverted-repeat *CHS* pseudogene) is related to tolerance of low temperature-induced seed coat discoloration in yellow soybean. *Theoretical and Applied Genetics*, 122, 633-642.

Okabe, A. (1996) Inheritance of seed coat cracking and effective selection method for the resistance in soybean. *Japan Agricultural Research Quarterly*, 30, 15-20.

Orf, J. H., K. Chase, T. Jarvik, L. M. Mansur, P. B. Cregan, F. R. Adler, K. G. Lark (1999) Genetics of soybean agronomic traits: I. comparison of three related recombinant inbred populations. *Crop Science*, 39, 1642–1651.

Oyoo, M. E., E. R. Benitez, H. Matsumura, R. Takahashi (2010a) QTL analysis of seed coat cracking in soybean. *Crop Science,* 50, 1230–1235.

Oyoo, M.E., S.M. Githiri, E.R. Benitez, R. Takahashi (2010b) QTL analysis of net-like cracking in soybean seed coats. *Breeding Science,* 60, 28–33.

Panthee, D. R., V. R. Pantalone, A.M. Saxton, D. R. West, C. E. Sams (2007) Quantitative trait loci for agronomic traits in soybean. *Plant Breeding*, 126, 51–57.

Percy, J. D., R. Philip, L. O. Vodkin (1999) A defective seed coat pattern (Net) is correlated with the post-transcriptional abundance of soluble proline-rich cell wall proteins. *Plant Molecular Biology*, 40, 603–613.

Qutob, D., F. Ma, C. A. Peterson, M. A. Bernards, M. Gijzen (2008) Structural and permeability of the soybean seed coat. *Botany*, 86, 219–227.

Ray, J.D., K. Hinson, J. E. B. Mankono, M. F. Malo (1995) Genetic control of a long-

juvenile trait in soybean. *Crop Science*, 35, 1001–1006.

Samanfar, B., S. J. Molnar, M. Charette, A. Schoenrock, F. Dehne, A. Golshani, F. Belzile, E. R. Cober (2017) Mapping and identification of a potential candidate gene for a novel maturity locus, *E10*, in soybean. *Theoretical and Applied Genetics*, 130, 377-390.

Schmutz, J., S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Son, J. J. Thelen, J. Cheng *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, 463, 178–183.

Sebolt, A. M., R. C. Shoemaker, B. W. Diers (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Science*, 40, 1438–1444.

Senda, M., N. Yamaguchi, M. Hiraoka, S. Kawada, R. Iiyoshi, K. Yamshita, S. Tomonori, H. Maeda, M. Kawasaki (2017) Accumulation of proanthocyanidins and/or lignin deposition in buff-pigmented soybean seed coats may lead to frequent defective cracking. *Planta*, 245, 659-670

Senda, M., T. Kurauchi, A. Kasai, S. Ohnishi (2012) Suppressive mechanism of seed coat pigmentation in yellow soybean. *Breeding Science*, 61, 523–530.

Smith, T., H. Camper (1975) Effects of seed size on soybean performance. *Agronomy Journal*, 67, 681–684.

Song, Q. J., G. F. Jia, Y. L. Zhu, D. Grant, R. T. Nelson, E. Y. Hwang, D. L. Hyten, P. B. Cregan (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. *Crop Science*, 50, 1950–1960.

Song, Q. J., L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido, X. Delannay, J.

E. Specht, P. B. Cregan (2004) A new integrated genetic linkage map of the soybean. *Theoretical and Applied Genetics*, 109, 122–128.

Stewart, R.T., J. B. Wentz (1930) A defective seed-coat character in soybeans. *Journal of American Society of Agronomy*, 22, 658–662.

Takahashi, R., J. Abe (1999) Soybean maturity genes associated with seed coat pigmentation and cracking in response to low temperatures. *Crop Science,* 39, 1657–1662.

Tang, X. X., R. Gong, W. Q. Sun, C. P. Zhang, S. B. Yu (2018) Genetic dissection and validation of candidate genes for flag leaf size in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 131, 801-815.

Tian, Z. X., X. B. Wang, R. Lee, Y. H. Li, J. E. Specht, R. L. Nelson, P. E. McClean, L. J. Qiu, J. X. Ma (2010) Artificial selection for determinate growth habit in soybean. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 8563–8568.

Tuyen, D. D., S. K. Lal, D. H. Xu (2010) Identification of a major QTL allele from wild soybean (*Glycine soja* Sieb. & Zucc.) for increasing alkaline salt tolerance in soybean. *Theoretical and Applied Genetics*, 121, 229–236.

Vadez, V., T. R. Sinclair (2002) Sensitivity of N2 fixation traits in soybean cultivar Jackson to Manganese. *Crop Science,* 42, 791-796.

Wang, D., G. L. Graef, A. M. Procopiuk, B. W. Diers (2004) Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theoretical and Applied Genetics*, 108, 458–467.

Wang, H. Q., X. G. Zhang, H. L. Yang, X. Y. Liu, H. M. Li, L. Yuan, W. H. Li, Z.Y. Fu, J. H. Tang, D. M. Kang (2016a) Identification of heterotic loci associated with

grain yield and its components using two CSSL test populations in maize. *Scientific Reports*, 6, 38205.

Wang, J., S. S. Chu, H. R. Zhang, Y. Zhu, H. Cheng, D. Y. Yu (2016b) Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Scientific reports*, 6, 20728.

Wang, S.K., S. Li, Q. Liu, K. Wu, J. Q. Zhang, S. S. Wang, Y. Wang, X. B. Chen, Y. Zhang, C. X. Gao *et al.* (2015) The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nature Genetics*, 47, 949–954.

Wang, W. B., Q. Y. He, H. Y. Yang, S. H. Xiang, T. J. Zhao, J. Y. Gai (2013) Development of a chromosome segment substitution line population with wild soybean (*Glycine soja* Sieb. & Zucc.) as donor parent. *Euphytica*, 189, 293–307.

Wang, W. B., X. L. Li, S. X. Chen, S. Y. Song, J. Y. Gai, T. J. Zhao (2016c) Using presence/absence variation markers to identify the QTL/allele system that confers the small seed trait in wild soybean (*Glycine soja* Sieb. & Zucc.). *Euphytica*, 208, 101–111.

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

Watanabe, S., T. Shinizu, K. Mchita, Y. Tsubokura, Z. J. Xia, T. Yamada, M. Hajika, M. Ishimoto, Y. Katayose, K. Harada, A. Kaga (2018) Development of a high-density linkagemap and chromosome segment substitution lines for Japanese soybean cultivar Enrei. *DNA Research*, 25, 123-136.

Watanabe, S., Z. J. Xia, R. Hideshima, Y. Tsubokura, S. Sato, K. Harada (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.

Wolf, W.J., F. L. Baker (1972) Scanning electron microscopy of soybeans. *Cereal Science Today,* 17, 125–130.

Xia, Z. J., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, T. Anai, S. Sato, T. Yamazaki, S. X. Lü, H. Y. Wu, S. Tabata, K. Harada (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proceedings of the National Academy of Sciences of the United States of America*, 109, E2155–E2164.

Xin, D. W., Z. M. Qi, H. W. Jiang, Z. B. Hu, R. S. Zhu, J. H. Hu, H. Y. Han, G. H. Hu, C. Y. Liu, Q. S. Chen (2016) QTL location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. & Zucc.) chromosome segment substitution lines. *PLoS One*, 11, e0149380.

Xu, D. H., D. D. Tuyen (2012) Genetic studies on saline and sodic tolerances in soybean. *Breeding Science*, 61, 599-565.

Xu, D. H., J. Abe, Y. Gai, Y. Shimamoto (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theoretical and Applied Genetics*, 105, 645-653.

Yaklich, R.W., G. Barla-Szabo (1993) Seed coat cracking in soybean. *Crop science,* 33, 1016–1019.

Yamamoto, Y., E. Sato, T. Shimizu, N. Nakamich, S. Sato, T. Kato, S. Tabata, A. Nagatani, T. Yamashino, T. Mizuno (2003) Comparative genetic studies on the *APRR5* and

APRR7 genes belonging to the *APRR1/TOC1* quintet implicated in circadian rhythm, control of flowering time, and early photomorphogenesis. *Plant & Cell Physiology*, 44, 1119–1130.

Yue Y. L., N. X. Liu, B. J. Jiang, M. Li, H. J. Wang, Z. Jiang, H. T. Pan, Q. J. Xia, Q. B. Ma, T. F. Han, H. Nian (2016) A single nucleotide deletion in *J* encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Molecular Plant*, 10, 656–658.

Zhai, Q. Z., Zhang, X., Wu, F. M., Feng, H. L., Deng, L., Xu, L., Zhang, M., Wang, Q. M., & Li, C. Y. (2015) Transcriptional mechanism of jasmonate receptor COI1-mediated delay of flowering time in *Arabidopsis. Plant Cell*, 27, 2814–2828.

Zhang, L., H. Li, Z. Li, J. Wang (2008) Interactions between markers can be caused by the dominance effect of QTL. *Genetics*, 180, 1177-1190.

Zhang, S. C., Z. G. Zhang, C. Bales, C. H. Gu, C. DiFonzo, M. Li, Q. J. Song, P. Cregan, Z. Y. Yang, D. C. Wang (2017a) Mapping novel aphid resistance QTL from wild soybean, *Glycine soja* 85-32, *Theoretical and Applied Genetics*, 130, 1941-1952.

Zhang, S. C., Z. G. Zhang, Z. X. Wen, C. H. Gu, Y. Q. Charles An, C. Bales, C. DiFonzo, Q. J. Song, D. C. Wang (2017b) Fine mapping of the soybean aphid-resistance genes *Rag6* and *Rag3c* from *Glycine soja* 85-32, *Theoretical and Applied Genetics*, 130, 2601-2615.

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

Zhou, L., S. B. Wang, J. B. Jian, Q. C. Geng, Q. J. Wen, Z. Z. Wu, G. J. Li, Y. Q. Liu, J.

M. Dunwell, J. Zhang, J. Y. Feng, Y. Niu, L. Zhang, W. L. Ren, Y. M. Zhang (2015b) Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Scientific Reports*, 5, 9350.

Zhou, Z. K., Y. Jiang, Z. Wang, Z. H. Gou, J. Lyu, W. Y. Li, Y. J. Yu, L. P. Shu, Y. J. Zhao, Y. M. Ma, C. Fang, Y. T. Shen, T. F. Liu, C. C. Li, Q. Li, M. Wu, Y. S. Wu, Y. Dong, W. T. Wan, X. Wang, Z. L. Ding, Y. D. Gao, H. Xiang, B. G. Zhu, S. H. Lee, W. Wang, Z. X. Tian (2015a) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, 33, 408–414.