A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHYLOSOPHY

Genetic and Genomic Study of Shoot Branching in Soybean

BY

SANGREA SHIM

AUGUST, 2018

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Genetic and Genomic Study of Shoot Branching in Soybean

UNDER THE DIRECTION OF DR. SUK-HA LEE
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF
SEOUL NATIONAL UNIVERSITY

BY
SANGREA SHIM

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE

MAY, 2018

APPROVED AS A QUALIFIED DISSERTATION OF SANGREA SHIM
FOR THE DEGREE OF DOCTOR OF PHYLOSOPHY
BY THE COMMITTEE MEMBERS

JULY, 2018

CHAIRMAN _____
Hee-jong Koh, Ph.D.

VICE-CHAIRMAN _____
Suk-Ha Lee, Ph.D.

MEMBER _____
Nam-Chon Paek, Ph.D.

MEMBER _____
Tae-Jin Yang, Ph.D.

MEMBER _____
Soon-Chun Jeong, Ph.D.

Genetic and Genomic Study of Shoot Branching in Soybean

SANGREA SHIM

DEPARTMENT OF PLANT SCIENCE

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

GENERAL ABSTRACT

The number of soybean branch directly affected yield components of pod and seed number per plant by generating more pod bearing sites on the branches. The number of branches in soybean is immensely affected by various environmental factors, numerous genetic factors and their interactions. For this reason, a little genetic factors associated with branch number has been identified. The causal genes controlling branch development has not been identified in soybean. QTLs associated with number of branches and promising candidate genes including *BRANCHED1* (*BRC1*) and *REVOLUTA* (*REV*) for branch development were identified using bi-parental mapping population and high-resolution genetic map.

To validate the major QTL, genetic association within a set of 430 soybean germplasms under three geographic replications was analyzed. Resultantly, a total of six out of 45 markers, which were located in the major QTL for branch number, were significantly associated with branch numbers that measured in at least two geographical replications. The six markers located in exon of *BRANCHED1* (*BRC1*) gene, intron of gene encoding transcription factor TFIIE alpha subunit and intergenic regions. In the meantime, using a set of near-isogenic lines (NILs) derived from a F_6 residual heterozygous line (RHL) for the major QTL, expression of two candidate genes between the NILs were compared. The result showed in that the *BRC1* gene was significantly down-regulated in NIL that exhibited more branches (*p-value* < 0.001), and no significant expression difference was identified for the other gene encoding TFIIE alpha subunit. In addition, genetic association of missense and upstream SNP for *BRC1* gene were identified in 59 accessions.

RNA-seq analysis for the set of NILs was conducted to provide global view of comprehensive mechanism for branch development in soybean. As a result, a total 376 differentially expressed genes were identified by comparison of the set of NILs. These DEGs were mapped to biological pathways such as abiotic stress signaling, plant-hormonal, secondary metabolism and development. Hormonal pathways, including auxin, cytokinin, gibberellin and abscisic acid, which were

associated with branch development, were observed.

Through a series of study about soybean branch development, strong candidate gene (*BRC1*) controlling branch development in soybean was identified. This gene encodes TEOSINTE BRANCHED 1/CYCLODEA/PROLIFERATING CELL FACTORS (TCP) transcription factor 18. The TCP transcription factor (TF) family are involved in developmental growth, such as leaf development, flower symmetry and shoot branching. Based on bioinformatics analysis such as phylogenetic tree, gene structure, motif conservation, synteny and expression pattern analysis, we could construct the duplication network of soybean TCPs. Through the duplication network, it was revealed that soybean TCP family was duplicated by whole genome duplication (WGD) and tandem duplication. Thus, the retention and structural/transcriptional divergence which could be supported by gene balance hypothesis and sub-/neo-functionalization were observed.

These results will be helpful for understanding of branch development, breeding high yield soybean by increased branches and evolution of TCP TF genes.

Keywords: soybean, branch, QTL, association, *BRC1*, evolution

***Student Number*:** 2012-30984

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION USED

aa	Amino acid
ABA	Abscisic acid
AM	Axillary meristem
CHR	Chromosome
CK	Cytokinin
DEG	Differentially expressed gene
FPKM	Fragment per kilobase million
GA	Gibberellic acid
HMM	Hidden markov model
JTT	Jones-Taylor-Thornton model
LG	Linkage group
Mw	Molecular weight
NGS	Next generation sequencing
NIL	Near-isogenic line
pI	Isoelectric point
QTL	Quantitative trait loci
R:FR	Red to far-red light ratio
RHL	Residual heterozygous line
RIL	Recombinant inbred line
SAM	Shoot apical meristem
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
TD	Tandem duplication

TCP TEOSINTE BRANCHED 1/CYCLODEA/PROLIFERATING CELL FACTORS

TF Transcription factor

WGD Whole genome duplication

GENERAL INTRODUCTION

Generally, seed yield of soybean increases with planting density until reaching a plateau (Agudamu et al. 2016). To maximize seed yield, two cultivation practices are available for achieving optimal plant density: In the USA, narrow-row/high-density planting is widely applied, with a plant density >25 plants m^{-2} and inter-row spacing of 40–45 cm (Heatherly and Elmore 2004). In countries including Korea, due to the incidence of lodging and disease, lower plant densities <20 plants m^{-2} are practiced to decrease seed and labor costs (Cho and Kim 2010). Increased branch number on the main stem can compensate for decreased the sowing rate under lower planting density (Cox et al. 2010; Agudamu et al. 2016). On the other hand, branching plasticity can decrease number of branches under dense planting, increasing branch development in proportion to land space per plant (Agudamu et al. 2016). However, a substantial variation in yield arising from differences in branch development under low planting density has been reported among US soybean cultivars (Board and Kahlon 2013). This difference in number of branches has also been observed among Korean/Japanese soybean cultivars (with more branches) and American/Chinese ones (with fewer branches), based on phenotypic data released by the Germplasm Resources Information Network

(GRIN, www.ars-grin.gov/) (Sayama et al. 2010). These observations represent that genetic diversity in branching exists among soybean genotypes, and that each country has developed soybean varieties that generate the appropriate number of branches in response to commonly used cultivation practices.

It has also been reported that the branch development was affected by diverse environmental stimuli including light quality and drought, along with phyto-hormones, such as auxin, cytokinin (CK), gibberellin (GA) and abscisic acid (ABA) (Board 2000; Basuchaudhuri 2016; Toyota et al. 2017). Drought treatment on early reproductive stage inhibited branch outgrowth (Frederick et al. 2001), suggests that drought stress responsive gene may affect genes regulating branch development. Plant hormones auxin has been reported as negative regulator of axillary branch outgrowth (Shimizu-Sato et al. 2009). Cytokinin (CK) promoting shoot branching has been demonstrated in soybean (Leyser 2003; Shimizu-Sato et al. 2009). One of the gibberellin, GA3 promoted number of branches in soybean (Sarkar et al. 2002). Additionally, a possibility of ABA affecting branch development has been raised by the demonstration causing more branch outgrowth along with decreased ABA under the shading treatment (Zhang et al. 2011).

Thus far, several quantitative trait loci (QTLs) analyses using simple sequence repeat (SSR) marker–based genetic maps have led to

the detection of 14 loci associated with branching number (Chen et al. 2007; Li et al. 2008; Sayama et al. 2010; Yao et al. 2015). These loci, which are described in SoyBase (https://www.soybase.org/search/qtllist_by_symbol.php), are distributed among 10 different linkage groups (LGs): C1, A1, C2, O, B1, B2, E, D2, G, and L. Among them, LGs B1 and C2 harbor the major QTLs for branching, explaining more than 10% of phenotypic variations (Chen et al. 2007; Sayama et al. 2010). Although a considerable number of QTLs associated with branching have been identified, the QTL regions still contain large genomic regions, due to the low resolution of the markers flanking the QTLs. Consequently, the causal genes for controlling soybean branching have yet to be fully characterized.

The study in this manuscript attempted to identify the causal genes and to elucidate the mechanism of branch development. Firstly, we identified QTLs and promising candidate genes associated with branch development using a set of recombinant inbred lines (RILs) and high-density genetic map. Secondly, major QTL and candidate genes that identified in QTL study was validated using a set of 430 soybean core germplasms and near-isogenic lines (NILs) derived from a F_6 residual heterozygous line (RHL). Next, transcriptome analysis of the NILs was conducted in order to dissect the mechanism of branch development. Resultantly, *BRANCHED1* (*BRC1*) gene act as negative regulator of

branch development under auxin signaling pathway (Aguilar-Martínez et al. 2007) was identified. This gene encodes *TEOSINTE BRANCHED 1*/*CYCLOIDEA*/ *PROLIFERATING CELL FACTORS 1* and *2* (TCP) transcription factor (TF). Lastly, we analyzed evolution mechanism of soybean TCP TF family. These results will provide understanding of branch development mechanism and evolution mechanism of TCP TF genes, as well as information for breeding of high yield soybean by increasing branches.

LITERATURE REVIEW

Branch of soybean

Soybean yield is a complex trait which is affected by numerous genetic factors, environmental factors and interactions between two factors (Hamawaki et al. 2012). Broad sense heritability of soybean yield ranges 20 ~ 30% (Hamawaki et al. 2012). Correlations between the yield and agronomic traits, such as plant height, number of nodes per plant, number of pods per plant, number of seeds per plant and number of branches per plant, has been reported (Ghodrati 2013). Seed yield has low broad sense heritability of 0.24 and positive correlation with plant height, number of nodes per plant, number of pods per plant and number of seed per plant (Ghodrati 2013). Interestingly, number of branches per plant show strong positive correlation with number of pods per plant and number of seeds per plant (Ghodrati 2013). This suggest that the number of branches is a factor affecting yield of soybean.

However, number of branches per plant show weak correlation with seed yield (per ha) (Ghodrati 2013). This can be explained by the compensation of yield and branch numbers along with planting density. Soybean yield varies with planting density. Mostly, two cultivation practices are applied to maximize the yield of soybean. One is high-

density method with >25 plants per square meter, which is usually practiced in USA (Heatherly and Elmore 2004). Another is low-density method with <20 plants per square meter, which is applied in countries including Korea to prohibit lodging and disease outbreak (Cho and Kim 2010). Under the low density practice, branches outgrow from the main stem compensate for decrease from lower sowing rate (Cox et al. 2010; Agudamu et al. 2016). While, the branching plasticity decrease branch number under dense planting, increasing branch development in proportion to land space per plant (Agudamu et al. 2016). However, a considerable variation in yield triggered by differences of branch development under low plant density has been reported among US soybean cultivars (Board and Kahlon 2013). The differences in branch development have also been reported between Korean/Japanese cultivars which exhibit more branches and American/Chinese ones which have less branches, based on phenotype data released by the Germplasm Resources Information Network (GRIN, www.ars-grin.gov/) of USDA (Sayama et al. 2010). These indicate that genetic diversity in branching development among soybean cultivars, and soybean cultivars of each country have been developed for appropriate branch number for each cultivation methods. The genetic regulation of branch development is not only soybean breeding objective for yield increase, but also interesting subject for plant developmental biology.

QTL for branching

Numerous agronomic traits are regulated by multiple genomic loci known QTLs. Along with the advances in molecular markers and analytical techniques, QTL association study become facilitated in various plant species. In soybean, based on the soybean genetic map, QTLs associated with large numbers of agronomic traits, such as number of nodes, branches and seeds have been mapped and reported in Soybase (https://www.soybase.org/search/qtllist_by_symbol.php). Even though the branch development is fascinating traits affecting yield, relatively low number of QTLs for branching have been reported compared to other phenotypic traits. A total 14 loci associated with branch number have been detected using simple sequence repeat (SSR) marker and genetic maps constructed by the markers (Chen et al. 2007; Li et al. 2008; Sayama et al. 2010; Yao et al. 2015). These loci are distributed 10 different linkage groups: A1, B1, B2, C1, C2, D2, E, G, O and L. Although the numbers of QTLs for branching which have been identified, the QTLs include large number of genes due to the low number and resolution of markers. As a result, causative genes for controlling soybean branching have not been fully characterized in soybean.

Development of near-isogenic line (NIL) from residual heterozygous line (RHL)

Near-isogenic line represents genetically homozygous line excepting few alleles or regions of interested. Near-isogenic line which have only one different allele or locus is created and used in validation of interesting QTL (Osborn et al. 1987). In development of NIL, usually, repeated backcrossing procedure is used to introduce specific allele of interested from a donor genotype to recipient genotype (Frisch et al. 1999). Transferred allele from donor line, then, is selected based on the phenotypes or genotypes examined by molecular markers for the transferred allele and background. However, this procedure takes a lot of time and efforts.

Another way to develop NIL is using the residual heterozygous line (RHL) of specific allele or region of interested in proceeded generation (Tuinstra et al. 1997). This procedure is consisted by three step; i) select residual heterozygous line within the bi-parental mapping population identified in QTL analysis, ii) proceeding selected RH line to next generation, iii) select segregated lines for specific allele of interested based on the phenotypes or genotypes using molecular marker. This method is more time and labor efficient compared to backcrossing method.

Evolution of TCP transcription factors

BRANCHED1 (*BRC1*) gene, one of the strong candidate of branching QTL, is known as TEOSINTE-BRNACHED 1/CYCLODEA/PROLIFERATING CELL FACTOR 1 and 2 (TCP) transcription factor which exist only in higher plant species and involved in various regulatory processes of growth and development such as cell growth, proliferation and axillary branch outgrowth (Cubas et al. 1999; Aguilar-Martínez et al. 2007). TCP transcription factors (TFs) contains highly conserved DNA-binding motif, referred TCP domain which is characterized by 59 amino-acid non-canonical basic-Helix-Loop-Helix (bHLH) (Cubas et al. 1999). TCP TF family is classified into two classes (Navaud et al. 2007), Class I and II which are distinguished by presence of four amino acid in basic region of bHLH and R-domain (Broholm et al. 2008). Generally, the two classes of TCP TFs regulated growth and development antagonistically (Martín-Trillo and Cubas 2010; Danisman et al. 2012). TCP of class I have been reported to promote growth (Kosugi and Ohashi 2002), otherwise, class II TCP have been reported to negatively regulate growth and proliferation (Doebley et al. 1997; Aguilar-Martínez et al. 2007; Martín-Trillo et al. 2011; Braun et al. 2012).

Identification and comprehensive analysis of TCP TFs have been conducted in various plant species, such as Arabidopsis, cotton species, tomato and *Populus euphratica* (Parapunova et al. 2014; Li 2015; Ma

et al. 2016a, b).

REFERENCES

Agudamu, Yoshihira T, Shiraiwa T (2016) Branch development responses to planting density and yield stability in soybean cultivars. Plant Prod Sci 19:331–339. doi: 10.1080/1343943X.2016.1157443

Aguilar-Martínez JA, Poza-Carrión C, Cubas P (2007) *Arabidopsis BRANCHED1* Acts as an Integrator of Branching Signals within Axillary Buds. Plant Cell 19:458–472. doi: 10.1105/tpc.106.048934

Basuchaudhuri P (2016) INFLUENCES OF PLANT GROWTH REGULATORS ON YIELD OF SOYBEAN. Indian J Plant Sci 5:25–38

Board J (2000) Light Interception Efficiency and Light Quality Affect Yield Compensation of Soybean at Low Plant Populations. Crop Sci 40:1285–1294. doi: 10.2135/cropsci2000.4051285x

Board JE, Kahlon CS (2013) Morphological Responses to Low Plant Population Differ Between Soybean Genotypes. Crop Sci 53:1109–1119. doi: 10.2135/cropsci2012.04.0255

Braun N, Germain A de S, Pillot J-P, et al (2012) The Pea TCP Transcription Factor PsBRC1 Acts Downstream of Strigolactones to Control Shoot Branching. Plant Physiol 158:225–238. doi: 10.1104/pp.111.182725

Broholm SK, Tähtiharju S, Laitinen RAE, et al (2008) A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. Proc Natl

Acad Sci 105:9117–9122. doi: 10.1073/pnas.0801359105

Chen Q, Zhang Z, Liu C, et al (2007) QTL Analysis of Major Agronomic Traits in Soybean. Agric Sci China 6:399–405. doi: 10.1016/S1671-2927(07)60062-5

Cho Y, Kim S (2010) Growth parameters and seed yield compenets by seeding time and seed density of non-/few branching soybean cultivars in drained paddy field. Asian J Plant Sci 9:140–145

Cox WJ, Cherney JH, Shields E (2010) Soybeans Compensate at Low Seeding Rates but not at High Thinning Rates. Agron J 102:1238–1243. doi: 10.2134/agronj2010.0047

Cubas P, Lauter N, Doebley J, Coen E (1999) The TCP domain: a motif found in proteins regulating plant growth and development. Plant J 18:215–222. doi: 10.1046/j.1365-313X.1999.00444.x

Danisman S, Wal F van der, Dhondt S, et al (2012) Arabidopsis Class I and Class II TCP Transcription Factors Regulate Jasmonic Acid Metabolism and Leaf Development Antagonistically. Plant Physiol 159:1511–1523. doi: 10.1104/pp.112.200303

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nat Lond 386:485–8. doi: http://dx.doi.org/10.1038/386485a0

Frederick JR, Camp CR, Bauer PJ (2001) Drought-Stress Effects on Branch and Mainstem Seed Yield and Yield Components of Determinate Soybean. Crop Sci 41:759–763. doi: 10.2135/cropsci2001.413759x

Frisch M, Bohn M, Melchinger AE (1999) Comparison of Selection Strategies for Marker-Assisted Backcrossing of a Gene. Crop

Sci 39:1295–1301. doi: 10.2135/cropsci1999.3951295x

Ghodrati G (2013) Study of genetic variation and broad sense heritability for some qualitative and quantitative traits in soybean (*Glycine max* L.) genotypes. Curr Opin Agric 2:

Hamawaki OT, Sousa LB de, Romanato FN, et al (2012) Genetic parameters and variability in soybean genotypes. Comun Sci 3:76–83

Heatherly LG, Elmore RW (2004) Managing Inputs for Peak Production. Soybeans Improv Prod Uses agronomymonogra:451–536. doi: 10.2134/agronmonogr16.3ed.c10

Kosugi S, Ohashi Y (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J 30:337–348. doi: 10.1046/j.1365-313X.2002.01294.x

Leyser O (2003) Regulation of shoot branching by auxin. Trends Plant Sci 8:541–545. doi: 10.1016/j.tplants.2003.09.008

Li S (2015) The *Arabidopsis thaliana* TCP transcription factors: A broadening horizon beyond development. Plant Signal Behav 10:. doi: 10.1080/15592324.2015.1044192

Li W, Zheng D, Van K, Lee S (2008) QTL Mapping for Major Agronomic Traits across Two Years in Soybean (*Glycine max* L. Merr.). J Crop Sci Biotech 11:171

Ma J, Liu F, Wang Q, et al (2016a) Comprehensive analysis of TCP transcription factors and their expression during cotton (*Gossypium arboreum*) fiber early development. Sci Rep 6:21535. doi: 10.1038/srep21535

Ma X, Ma J, Fan D, et al (2016b) Genome-wide Identification of TCP Family Transcription Factors from *Populus euphratica* and Their Involvement in Leaf Shape Regulation. Sci Rep 6:32795. doi: 10.1038/srep32795

Martín-Trillo M, Cubas P (2010) TCP genes: a family snapshot ten years later. Trends Plant Sci 15:31–39. doi: 10.1016/j.tplants.2009.11.003

Martín-Trillo M, Grandío EG, Serra F, et al (2011) Role of tomato *BRANCHED1*-like genes in the control of shoot branching. Plant J 67:701–714. doi: 10.1111/j.1365-313X.2011.04629.x

Navaud O, Dabos P, Carnus E, et al (2007) TCP Transcription Factors Predate the Emergence of Land Plants. J Mol Evol 65:23–33. doi: 10.1007/s00239-006-0174-z

Osborn TC, Alexander DC, Fobes JF (1987) Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. TAG Theor Appl Genet Theor Angew Genet 73:350–356. doi: 10.1007/BF00262500

Parapunova V, Busscher M, Busscher-Lange J, et al (2014) Identification, cloning and characterization of the tomato TCP transcription factor family. BMC Plant Biol 14:157. doi: 10.1186/1471-2229-14-157

Sarkar PK, Haque MS, Karim MA (2002) Effects of GA3 and IAA and their Frequency of Application on Morphology, Yield Contributing Characters and Yield of Soybean. J Agron 1:119–122. doi: 10.3923/ja.2002.119.122

Sayama T, Hwang T-Y, Yamazaki H, et al (2010) Mapping and

comparison of quantitative trait loci for soybean branching phenotype in two locations. Breed Sci 60:380–389. doi: 10.1270/jsbbs.60.380

Shimizu-Sato S, Tanaka M, Mori H (2009) Auxin-cytokinin interactions in the control of shoot branching. Plant Mol Biol 69:429–435. doi: 10.1007/s11103-008-9416-3

Toyota M, Maitree L, Chomsang K (2017) Changes in radiation interception and R:FR over time and with canopy depth of two soybean cultivars with different branching characteristics. Plant Prod Sci 20:205–214. doi: 10.1080/1343943X.2017.1294464

Tuinstra MR, Ejeta G, Goldsbrough PB (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95:1005–1011. doi: 10.1007/s001220050654

Yao D, Liu ZZ, Zhang J, et al (2015) Analysis of quantitative trait loci for main plant traits in soybean. Genet Mol Res GMR 14:6101–6109. doi: 10.4238/2015.June.8.8

Zhang J, Smith DL, Liu W, et al (2011) Effects of shade and drought stress on soybean hormones and yield of main-stem and branch. Afr J Biotechnol 10:14392–14398. doi: 10.5897/AJB11.2143

Chapter 1.

Identification of QTLs for branching in soybean (*Glycine max* (L.) Merill)

doi: 10.1007/s10681-017-2016-z
Euphytica September 2017 213:225
Sangrea Shim, Moon Young Kim, Jungmin Ha, Yeong-Ho Lee, Suk-Ha Lee

Abstract

In soybean, the number of branches directly affects total pod number per plant. In this study, we sought to identify QTLs and candidate genes associated with branching in 200 F_6 recombinant inbred lines derived from a cross between Jiyu69 and SS0404-T5-76, which exhibit significant differences in branch number. Using a high-resolution genetic map constructed using the BARCSoySNP6K chip, we detected a novel QTL and confirmed three known QTLs related to branching, as well as two known QTLs for total pod number. Two of the QTLs conferring branching, including a major QTL on chromosome 6 with an R^2 value of 14.5%, were co-localized with QTLs associated with total pod number. Although several of the QTLs we identified for the two traits were located near identified QTLs, the high-resolution map enabled us to significantly narrow down the genomic regions for these

QTLs (from 26 Mb to 460 kb at most), facilitating identification of promising candidate genes. From the QTL regions we identified, we selected six candidate genes, mostly encoding transcription factors regulating expression of gene networks involved in axillary branching via interactions with the auxin hormone network, including a TEOSINTE-BRANCHED1/CYCLOIDEA/PCF (TCP) transcription factor (*BRANCHED1*: *BRC1*) and a homeobox-leucine zipper protein (*REVOLUTA*: *REV*). The results of this study will help breeders improve soybean yield by increasing the branch number using marker-assisted selection, and will facilitate identification of the causative genes for branching.

Keywords soybean; quantitative trait loci; branching; TEOSINTE-BRANCHED1/CYCLOIDEA/PCF (TCP) transcription factor; homeobox-leucine zipper protein; auxin

Introduction

Soybean seed yield generally increases with planting density until reaching a plateau (Agudamu et al. 2016). Two cultivation methods are available for achieving optimal plant density to maximize seed yield: In the USA, narrow-row/high-density planting is widely practiced, with a plant density >25 plants m^{-2} and inter-row spacing of 40–45 cm (Heatherly and Elmore 2004). In countries including Korea, due to the incidence of lodging and disease, lower plant densities <20 plants m^{-2} are used to decrease seed and labor costs (Cho and Kim 2010). Under lower plant density, increased branch number on the main stem can compensate for decreased the sowing rate (Agudamu et al. 2016; Cox et al. 2010). On the other hand, branching plasticity can decrease branch number under dense planting, increasing branch development in proportion to land space per plant (Agudamu et al. 2016). However, a substantial variation in yield arising from differences in branch development under low plant density has been reported among US soybean cultivars (Board and Kahlon 2013). This difference in branch development has also been observed among Korean/Japanese soybean varieties (with more branches) and American/Chinese ones (with fewer branches), based on phenotypic data from the Germplasm Resources Information Network (GRIN, www.ars-grin.gov/) (Sayama et al. 2010). These observations indicate that genetic diversity in branching exists

among soybean genotypes, and that each country has developed soybean cultivars that generate the appropriate branch number in response to commonly used cultivation practices. The genetic control of branch development is not the only soybean breeding strategy for increasing yield, but it remains an interesting topic in plant developmental biology.

In soybean, several quantitative trait loci (QTLs) analyses using simple sequence repeat (SSR) marker–based genetic maps have led to the detection of 14 loci associated with branching number (Chen et al. 2007; Li et al. 2008a; Sayama et al. 2010; Yao et al. 2015). These loci, which are described in SoyBase and the Soybean Breeder's Toolbox (https://www.soybase.org/search/qtllist_by_symbol.php), are distributed among 10 different linkage groups (LGs): C1, A1, C2, O, B1, B2, E, D2, G, and L. Among them, LGs B1 and C2 harbor the major QTLs for branching, explaining more than 10% of phenotypic variation each (Chen et al. 2007; Sayama et al. 2010). Although a considerable number of QTLs conferring branching have been identified, the QTL regions still contain a large number of genes, largely due to the low resolution of the markers flanking the QTLs. Consequently, the genes responsible for controlling soybean branching have yet to be fully characterized.

The objective of this study was to detect QTLs conditioning branching, with the ultimate goal of identifying candidate genes

involved in branch development. To this end, we constructed a high-density genetic map using the BARCSoySNP6K single-nucleotide polymorphism (SNP) chip on an F_6 recombinant inbred line (RIL) population derived from a cross between two soybean genotypes that differ in branch number. The genetic locations of the markers flanking the identified QTLs were converted into physical positions on the soybean reference genome to validate genes likely to be implicated in branching. In addition, we investigated the correlation between branching and total pod number, as well as the genomic proximity of QTLs for the two traits.

Materials and methods

Plant materials

A mapping population of 200 F_6 RILs was developed by single-seed descent in F_{2s} derived from a cross between Jiyu69 and SS0404-T5-76. Jiyu69 is an elite cultivar developed in China. SS0404-T5-76 is an elite high-yielding line selected from RILs derived from a cross between Pungsannamulkong and SS2-2 by Crop Genomics Laboratory, Seoul National University. Jiyu69, the maternal line, has fewer branches than SS0404-T5-76, the paternal line.

Phenotype assessment

The RIL population was planted on the experimental farm of Seoul National University, Suwon, Republic of Korea (N 37° 16' 12.094", E 126° 59' 20.756") in 2016 to investigate the number of branches and total pod number for QTL analysis. To maximize branching performance, the RILs were planted at relatively low planting density with a row spacing of 0.8 m and planting interval of 0.2 m. The plants were grown under a natural photoperiod of 11.5–14.5 h per day using standard farming practices. Phenotypes of each trait were examined in three replicates (three plants per replicate for parents, and two plants per replicate for each RIL) at development stage R8, when 95% of the pods had matured.

DNA extraction and SNP genotyping

Genomic DNA was extracted from healthy young leaves of the mapping parents and their RILs as previously described (Shure et al. 1983). Concentration of each DNA sample was measured on an ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA) and diluted to a working concentration with Tris-EDTA buffer (pH 8.0). The BARCSoySNP6K Illumina BeadChip, which consists of 5,403 SNPs well distributed across soybean reference genome (Lee et al. 2015), was used to genotype the RILs and their parents. Allele calling of each SNP locus was performed using the Illumina GenomeStudio software with default parameters (Illumina Inc., San Diego, CA, USA).

Genetic linkage map construction and QTL analysis

A soybean linkage map was constructed using JoinMap v.4.1 (Ooijen 2006) with SNP genotyping results from the 200 RILs and their parental lines. The option of "Assign Identical Loci to Their Groups" was used to include as many markers as possible and thus increase the marker resolution of the genetic map. Regression mapping with Kosambi's calculation algorithm was used to calculate genetic distances between markers. The QTL IciMapping software v.4.1.0.0 (Meng et al. 2015) was used to perform QTL analysis through inclusive

composite interval mapping with additive effects. Missing phenotypic measurements were considered invalid. Association analysis was conducted for every 1.0 cM step with a PIN threshold of 0.001. To determine the statistically significant threshold for the LOD score, a thousand-permutation test was applied at a type I error rate of 0.05.

Comparison of the identified QTLs with the previously reported QTLs

Marker information for previously reported QTLs for branching and total pod number was obtained from SoyBase and the Soybean Breeder's Toolbox (https://www.soybase.org/search/qtllist_by_symbol.php), where most QTLs were identified by single-marker ANOVA rather than by interval mapping. Therefore, the two nearest markers provided by SoyBase (https://www.soybase.org/SeqMapSearch/GbrowseSearch/SeqGenSearch4.php?qtlname=Branching%203-3) were used to locate the previously reported QTLs on the soybean reference genome.

Expression patterns of candidate genes

To investigate the expression patterns of candidate genes likely to be involved in branching and total pod number, the expression levels of these genes in nine tissues of soybean cv. Williams 82, including flower, leaf, nodule, pod, root, root hair, shoot apical meristem (SAM), seed

and stem, were obtained from the RNA-seq data (FPKM values) at Phytozome v10.0 (Schmutz et al. 2010). A heat map with hierarchical clustering of the genes was constructed using R package pheatmap, to visualize the expression levels in nine tissues based on the $\log_2$(FPKM + 1) values of the genes.

Results

Phenotypic evaluation for traits related to branching and total pod number

The parental cultivars Jiyu69 and SS0404-T5-76 exhibited significant differences in branching and total pod number (P=0.0002 and 0.0052, respectively) (Table 1-1). Jiyu69 and SS0404-T5-76 had 5.8 ± 0.8 and 11.3 ± 2.9 branches, respectively (Table 1-1; Fig. 1-1a). In the RIL population, branch numbers ranged from 0 to 18.5 (Table 1-1). Total pod number of Jiyu69 was 167.7 ± 76.9, whereas SS0404-T5-76 produced more than twice as many pods (346.1 ± 148.9) (Table 1-1; Fig. 1-1b). In the RIL population, total pod number ranged widely, from 36.6 to 557.5 (Table 1-1). The phenotypic frequencies of branching and total pod number in this population obeyed a normal distribution with transgressive variation (Fig. 1-1), implying control by multiple genes. The two traits were positively correlated (r=0.75, P<0.001, Fig. 1-2). Broad-sense heritability of branch number (H^2=0.52) was higher than that of total pod number (H^2=0.35) (Table 1-1).

Map construction and QTL analysis of branching and total pod number

BARCSoySNP6K chip analysis of the RIL population of Jiyu69 × SS0404-T5-76 revealed that 2,049 markers out of 5,128 genotyped SNP

loci exhibited parental polymorphisms (Table 1-2). Among these markers, 1,981 were integrated into 20 soybean chromosomes (Chrs) (Table 1-2; Fig. 1-3), spanning a total of 2,773.3 cM with an average inter-marker distance of 1.4 cM (Table 1-3).

Map positions and characteristics of the QTLs identified in this study are shown in Table 1-4. Four QTLs for branch number were detected on Chrs 6, 11, 12, and 19, together explaining 37.3% of phenotypic variations (Table 1-4; Fig. 1-4). The QTL *qBR6-1* in the interval between markers Gm06_20486758_C_A and Gm06_20943239_T_C on Chr 6 accounted for 14.5% of variation and had the highest LOD score, 10.3 (Table 1-4). The two remaining QTLs, *qBR19-1* and *qBR11-1*, had LOD scores of 7.5 and 6.1 and explained 9.5% and 7.6% of phenotypic variation, respectively (Table 1-4; Fig. 1-4). On Chr 12, a minor QTL for branching (*qBR12-1*) between Gm12_5610878_T_C and Gm12_6023395_T_C had the lowest R^2 value (5.7%) among all QTLs detected (Table 1-4). At one QTL, *qBR19-1*, the SS0404-T5-76 allele increased the number of branches, whereas the Jiyu69 alleles of the remaining QTLs (*qBR6-1*, *qBR11-1*, and *qBR12-1*) decreased branching (Table 1-4).

Two QTLs for total pod number, *qPN6-1* and *qPN11-1*, were identified on Chrs 6 and 11 (Fig. 1-4), where two QTLs controlling branching (*qBR6-1* and *qBR11-1*) were co-localized (Fig. 1-4). The QTL *qPN6-1*, with showed the LOD score of 4.8 and 7.9% of

variations, flanked by the same markers as *qBR6-1*, Gm06_20486758_C_A...Gm06_20943239_T_C (460 kb interval) (Table 1-4; Fig. 1-5). The 80 kb interval between Gm11_10847172_T_C and Gm11_10926986_T_C on Chr 11 harbored the second QTL, *qPN11-1*, and the marker Gm11_10847172_T_C was also linked with branching (Table 1-4). The LOD score of *qPN11-1* was 6.6, and the individual effect of this QTL on total pod number was 10.8% (Table 1-4). At *qPN6-1* and *qPN11-1*, Jiyu69 contributed deleterious alleles that decreased total pod number (Table 1-4).

Survey of candidate genes residing in QTL regions and their expression patterns

Because the sequence-based SNP markers associated with QTLs for branching and total pod number directly provide the chromosomal locations, these markers enabled us to survey candidate genes in the soybean reference genome that are likely to be involved in these traits. For *qBR6-1* and *qPN6-1*, 13 protein-coding genes were present within the 460 kb region flanked by the markers Gm06_20486758_C_A…Gm0620943239_T_C (Table 1-4; Table 1-5). The 130 kb genomic regions corresponding to *qBR11-1* (Gm11_10721006_A_G…Gm11_10847172_T_C) and *qBR19-1* (Gm19_44636089_T_C…44761515_G_A) on Chrs 11 and 19 harbored 18 and 13 genes, respectively (Table 1-4; Table 1-5). The 420 kb

genomic region of *qBR12-1* (Gm12_5610878_T_C...Gm12_6023395_T_C) contained 32 genes, the largest number among the regions examined (Table 1-4; Table 1-5). By contrast, the 79 kb marker interval corresponding to *qPN11-1* (Gm11_10847172_T_C…Gm11_10926986_T_C) contained only seven genes (Table 1-4; Table 1-5).

From this set of genes, we excluded housekeeping genes, and then selected six candidate genes that were likely to be relevant to branching and total pod number (Fig. 1-5) based on the functional annotations of their *Arabidopsis thaliana* homologs in TAIR (http://www.arabidopsis.org), which are involved in developmental growth (Fig. 1-5; Table 1-5). In the QTL region of *qBR6-1* and *qPN6-1*, a gene encoding a TEOSINTE-BRANCHED1/CYCLOIDEA/PCF (TCP) transcription factor (Glyma06g23410) is the homolog of *BRC1* (*BRANCHED1*), which regulates axillary bud formation in *A. thaliana* (Fig. 1-5; Table 1-5). The *qBR11-1* region, which co-localized with *qPN11-1*, includes a gene encoding a C2H2-type zinc-finger protein (Glyma11g15140) (Fig. 1-5; Table 1-5). The *qBR12-1* region harbors genes containing homeodomains, including *REV* (*REVOLUTA*; Glyma12g08080) and *BLH8/PNF* (*POUND-FOOLISH*; Glyma12g08270), along with *EXO70H7* (*exocyst subunit exo70 family protein H7*; Glyma12g08020) (Fig. 1-5; Table 1-5). In the *qBR19-1* region, we selected the candidate gene *PTL* (*PETAL LOSS*;

Glyma19g37660), which encodes a Trihelix transcription factor (Fig. 1-5; Table 1-5). Using the available RNA-Seq data of *G. max* (Schmutz et al. 2010), in silico expression profiling of the six selected candidate genes revealed that the TCP transcription factor gene (Glyma06g23410) showed higher expression level in SAM than other tissues (Fig. 1-6). For three genes encoding *REV* (Glyma12g08080), *BLH8/PNF* (Glyma12g08270) and *PTL* (Glyma11g15140), high levels of transcript abundances were also exhibited in SAM, though they were expressed at high levels in some other tissues simultaneously.

Discussion

Branching of soybean affects yield by producing additional reproductive organs and pods on branches (Carpenter and Board 1997). In this study, we identified QTLs for branching and total pod number based on a high-density SNP map generated using BARCSoySNP6K array analysis (Lee et al. 2015). The high-density genetic map, which comprises 99 SNP markers per chromosome (Table 1-3), allows us to achieve sufficiently high marker resolution to survey candidate genes. Four QTLs for branching and two for total pod number were identified on Chrs 6, 11, 12, and 19 (Table 1-4; Fig. 1-4). Among these, two QTLs for total pod number (*qPN6-1* and *qPN11-1*) on Chrs 6 and 11 overlapped with branching QTLs (*qBR6-1* and *qBR11-1*), in accordance with the positive correlation between branching and total pod number (Fig. 1-4).

A total of 14 previously identified QTLs for branching are distributed among 10 different chromosomes (Fig. 1-7); in SoyBase, these regions are designated *Branching 1-1* to *1-5*, *2-1*, *3-1* to *3-5*, and *4-1* to *4-3* (https://soybase.org/). For total pod number, 34 previously reported QTLs are named *Pod number 1-1* to *1-10*, *2-1* to *2-2*, *3-1* to *3-4*, *4-1* to *4-2*, *5-1* to *5-2*, *7-1* to *7-3*, *8-1* to *8-3*, *9-1* to *9-3*, *10-1*, and *11-1* to *11-4*. These QTLs are distributed among 17 different chromosomes, of which nine (Chrs 4, 5, 6, 10, 11, 15, 17, 18, and 19) carry QTLs for

both branching and total pod number (Fig. 1-7).

Positional comparison of our identified QTLs with the reported QTLs revealed that only one, *qBR12-1* on Chr 12, was novel, whereas the other three, on Chrs 6, 11, and 19, co-localized with the existing QTLs (Table 1-4; Fig. 1-5). The 460 kb marker interval (Gm06_20486758_C_A…Gm06_20943239_T_C) encompassing QTLs *qBR6-1* and *qPN6-1* on Chr 6 lies within the genomic regions of the previously reported QTLs *Branching 3-1* (BARC-020405-04602…Satt489; 5.5 Mb interval) and *Pod number 3-4* (BARC-064115-18558…Sat_238; 26.0 Mb interval) (Fig. 1-5). The QTLs *qBR11-1* (Gm11_10721006_A_G...Gm11_10847172_T_C) and *qPN11-1* map in the vicinity of other known QTLs, *Branching 3-3* (BARC-018099-02516...Sat_247; 3.8 Mb interval) and *Pod number 3-2* (Satt509...Sat_247; 2.9 Mb interval) (Fig. 1-5). The *qBR19-1* region (Gm19_44636089_T_C…Gm19_44761515_G_A; 130 kb) on Chr 19 is anchored to the region of *Branching 3-2* (BARC-065769-19741...BARC-040521-07773; 0.7 Mb interval) (Fig. 1-5). In this study, however, the intervals of the markers flanking the known QTLs were narrowed down significantly, enabling us to identify six candidate genes likely to be involved in controlling branching, and thus total pod number (Fig. 1-5; Table 1-5). Most of these genes encode transcription factors that regulate the expression of gene networks involved in axillary branching, such as the *MORE AXILLARY GROWTH* (*MAX*)-

dependent pathway, which interacts with auxin hormone networks (Domagalska and Leyser 2011).

Auxin exerts a major effect on formation of axillary meristem (AM) in leaf axils and the outgrowth of axillary buds, and consequently on shoot branching, but does not move into axillary buds to repress bud growth (Domagalska and Leyser 2011; Rameau et al. 2015). Other secondary messengers such as carotenoid-derived signal molecules, which are produced by the serial actions of four *MAX* genes (*MAX1* to *MAX4*) regulated by auxin, are required for branch suppression (Bennett et al. 2006; Domagalska and Leyser 2011; Schwartz et al. 2004). Downstream of this complex mechanism of branching control, the gene *BRC1* (Glyma06g23410), located in the major QTL *qBR6-1* (or *qPN6-1*) (Fig. 1-5; Table 1-5), encodes a TCP transcription factor that plays an important role as the integrator of endogenous (hormones) and environmental (planting density, shade, and low R:FR) stimuli within axillary buds to determine bud fate (Aguilar-Martinez et al. 2007). Loss of *BRC1* activity leads to higher levels of branching, and overexpression of this gene results in repression of branch outgrowth (Aguilar-Martinez et al. 2007).

Homeodomain transcription factors are also implicated in AM initiation under the control of auxin (Sablowski 2015). The homeobox-leucine zipper protein encoded by *REV*, positioned in QTL *qBR12-1* (Fig. 1-5; Table 1-5), directly up-regulates *STM*

(*SHOOTMERISTEMLESS*) gene expression at a minimum level of auxin in leaf axils, which is required for AM development (Shi et al. 2016). *STM*, a member of the KNOX (Knotted-like homeobox) protein family, heterodimerizes with BEL1-like homeodomain (BLH) transcription factors such as *PNF*, in order to be targeted to the nuclear compartment (Cole et al. 2006). Hence, in *A. thaliana BLH/PNF* is critical for the initiation, maintenance, and development of shoot meristem (Kanrar et al. 2008; Rutjens et al. 2009; Ung et al. 2011), and a gene encoding a soybean homolog of this protein (Glyma12g08270) was found in QTL *qBR12-1* (Fig. 1-5; Table 1-5). In soybean, moreover, *REV* and *BLH/PNF* were highly expressed in SAM, in company with *BRC1* (Fig. 1-6), indicating that these candidate genes may function in SAM directly or indirectly to mediate branch development.

A gene encoding a C2H2 zinc-finger protein (Glyma11g15140) located in QTL *qBR11-1* is a candidate for branching regulation (Fig. 1-5; Table 1-5). Overexpression of this protein, generally referred to as the TFIIIA-type zinc finger, in transgenic petunia, tobacco, and *Arabidopsis* plants results in a dramatic increase in lateral shoots (Nakagawa et al. 2005). The Trihelix transcription factor gene *PTL* was selected as the candidate for QTL *qBR19-1* (Fig. 1-5) because its gain-of-function mutant exhibits a defect in auxin action and altered auxin distribution (Li et al. 2008b), probably affecting branch development in

some way. Furthermore, we also put *EXO70H7* (Glyma12g08020) in *qBR12-1* on the candidate list (Fig. 1-5) because T-DNA insertional mutants of the *AtEXO70A1* gene (an *EXO70H7* homolog) exhibit a loss of apical dominance, leading them to initiate highly branched inflorescence (Synek et al. 2006).

In conclusion, the high-density genetic map constructed using the BARCSoySNP6K SNP chip enabled us to identify a new QTL controlling branching and remarkably narrowed down the flanking regions of three QTLs. These loci will help breeders to develop soybean genotypes with appropriate branch number through marker-assisted selection to increase seed yield. Furthermore, in the four QTLs we identified, we noted six candidate genes, including *BRC1*, *REV*, and *BLH/PNF*, that are very likely to be involved in branch development; most of these genes encode transcription factors. To fully characterize the candidate genes, the identification of genetic variations, its association with phenotypic variations and gene transformation (overexpression or knockout) are required. These functional validation of the candidate genes should shed new light on the molecular regulation of soybean branch development, and thus total pod number.

References

Agudamu, Yoshihira T, Shiraiwa T (2016) Branch development responses to planting density and yield stability in soybean cultivars. Plant Production Science 19:331-339. doi:10.1080/1343943x.2016.1157443

Aguilar-Martinez JA, Poza-Carrion C, Cubas P (2007) *Arabidopsis BRANCHED1* acts as an integrator of branching signals within axillary buds. Plant Cell 19:458-472. doi:10.1105/tpc.106.048934

Bennett T, Sieberer T, Willett B, Booker J, Luschnig C, Leyser O (2006) The *Arabidopsis MAX* pathway controls shoot branching by regulating auxin transport. Curr Biol 16:553-563. doi:10.1016/j.cub.2006.01.058

Board JE, Kahlon CS (2013) Morphological responses to low plant population differ between soybean genotypes. Crop Science 53:1109. doi:10.2135/cropsci2012.04.0255

Carpenter AC, Board JE (1997) Branch yield components controlling soybean yield stability across plant populations. Crop Science 37:885-891. doi:10.2135/cropsci1997.0011183X003700030031x

Chen Q-s, Zhang Z-c, Liu C-y, Xin D-w, Qiu H-m, Shan D-p, Shan C-y, Hu G-h (2007) QTL analysis of major agronomic traits in soybean. Agricultural Sciences in China 6:399-405. doi:10.1016/s1671-2927(07)60062-5

Cho Y-S, Kim S-D (2010) Growth parameters and seed yield compenets by seeding time and seed density of non-/few branching soybean cultivars in drained paddy field. Asian Journal of Plant Science 9:140-145

Cole M, Nolte C, Werr W (2006) Nuclear import of the transcription factor SHOOT MERISTEMLESS depends on

heterodimerization with BLH proteins expressed in discrete sub-domains of the shoot apical meristem of *Arabidopsis thaliana*. Nucleic Acids Res 34:1281-1292. doi:10.1093/nar/gkl016

Cox WJ, Cherney JH, Shields E (2010) Soybeans compensate at low seeding rates but not at high thinning rates. Agronomy Journal 102:1238. doi:10.2134/agronj2010.0047

Domagalska MA, Leyser O (2011) Signal integration in the control of shoot branching. Nat Rev Mol Cell Biol 12:211-221. doi:10.1038/nrm3088

Heatherly LG, Elmore RW (2004) Managing inputs for peak production. In: Boerma HR, Specht JE (eds) Soybean: Improvement, production, and uses. ASA-CSSA-SSSA, Madison, WI, pp 451-536

Kanrar S, Bhattacharya M, Arthur B, Courtier J, Smith HM (2008) Regulatory networks that function to specify flower meristems require the function of homeobox genes *PENNYWISE* and *POUND-FOOLISH* in *Arabidopsis*. Plant J 54:924-937. doi:10.1111/j.1365-313X.2008.03458.x

Lee S, Freewalt KR, McHale LK, Song Q, Jun T-H, Michel AP, Dorrance AE, Mian MAR (2015) A high-resolution genetic linkage map of soybean based on 357 recombinant inbred lines genotyped with BARCSoySNP6K. Mol Breeding 35. doi:10.1007/s11032-015-0209-5

Li W, Zheng D-H, Van K, Lee S-H (2008a) QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. Merr.). J Crop Sci Biotech 11:171 ~ 190. doi:10.1007/s00122-003-1527-2

Li X, Qin G, Chen Z, Gu H, Qu L-J (2008b) A gain-of-function mutation of transcriptional factor PTL results in curly leaves,

dwarfism and male sterility by affecting auxin homeostasis. Plant Mol Biol 66:315-327. doi:10.1007/s11103-007-9272-6

Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal 3:269-283. doi:10.1016/j.cj.2015.01.001

Nakagawa H, Jiang C-J, Sakakibara H, Kojima M, Honda I, Ajisaka H, Nishijima T, Koshioka M, Homma T, Mander LN, Takatsuji H (2005) Overexpression of a petunia zinc-finger gene alters cytokinin metabolism and plant forms. Plant J 41:512-523. doi:10.1111/j.1365-313X.2004.02316.x

van Ooijen JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, The Netherlands

Rameau C, Bertheloot J, Leduc N, Andrieu B, Foucher F, Sakr S (2015) Multiple pathways regulate shoot branching. Frontiers in Plant Science 5. doi:10.3389/fpls.2014.00741

Rutjens B, Bao D, van Eck-Stouten E, Brand M, Smeekens S, Proveniers M (2009) Shoot apical meristem function in *Arabidopsis* requires the combined activities of three BEL1-like homeodomain proteins. Plant J 58:641-654. doi:10.1111/j.1365-313X.2009.03809.x

Sablowski R (2015) Control of patterning, growth, and differentiation by floral organ identity genes. J Exp Bot 66:1065-1073. doi:10.1093/jxb/eru514

Sayama T, Hwang T-Y, Yamazaki H, Yamaguchi N, Komatsu K, Takahashi M, Suzuki C, Miyoshi T, Tanaka Y, Xia Z, Tsubokura Y, Watanabe S, Harada K, Funatsuki H, Ishimoto M (2010) Mapping and comparison of quantitative trait loci for soybean branching phenotype in two locations. Breeding Science

60:380-389. doi:10.1270/jsbbs.60.380

Schmutz J., Cannon S. B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D. L., Song Q., Thelen J. J., Cheng J., Xu D., Hellsten U., May G. D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M. K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X. C., Shinozaki K., Nguyen H. T., Wing R. A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R. C., Jackson S. A. et al. (2010) Genome sequence of the palaeopolyploid soybean Nature 463:178-183 doi:10.1038/nature08670

Schwartz SH, Qin X, Loewen MC (2004) The biochemical characterization of two carotenoid cleavage enzymes from *Arabidopsis* indicates that a carotenoid-derived compound inhibits lateral branching. J Biol Chem 279:46940-46945. doi:10.1074/jbc.M409004200

Shi B, Zhang C, Tian C, Wang J, Wang Q, Xu T, Xu Y, Ohno C, Sablowski R, Heisler MG, Theres K, Wang Y, Jiao Y (2016) Two-step regulation of a meristematic cell population acting in shoot branching in *Arabidopsis*. PLoS Genet 12:e1006168. doi:10.1371/journal.pgen.1006168

Shure M, Wessler S, Fedoroff N (1983) Molecular identification and isolation of the *Waxy* locus in maize. Cell 35:225-233. doi:10.1016/0092-8674(83)90225-8

Synek L, Schlager N, Elias M, Quentin M, Hauser MT, Zarsky V (2006) AtEXO70A1, a member of a family of putative exocyst subunits specifically expanded in land plants, is important for polar growth and plant development. Plant J 48:54-72. doi:10.1111/j.1365-313X.2006.02854.x

Ung N, Lal S, Smith HM (2011) The role of PENNYWISE and POUND-FOOLISH in the maintenance of the shoot apical meristem in Arabidopsis. Plant Physiol 156:605-614. doi:10.1104/pp.110.171462

Yao D, Liu ZZ, Zhang J, Liu SY, Qu J, Guan SY, Pan LD, Wang D, Liu JW, Wang PW (2015) Analysis of quantitative trait loci for main plant traits in soybean. Genet Mol Res 14:6101-6109. doi:10.4238/2015.June.8.8

Table 1-1. Summary statistics of branching and total pod number in the RIL population of Jiyu69 × SS0404-T5-76

Traits	Mapping parents			RIL population			
	Jiyu69	SS0404-T5-76	*P*-value	Avg ± Std Dev	Min	Max	H^2
Branching	5.8 ± 0.8	11.3 ± 2.9	0.0002***	7.6 ± 3.0	0.0	18.5	0.52
Total pod number	167.7 ± 76.9	346.1 ± 148.9	0.0052**	200.2 ± 96.0	36.6	577.5	0.35

Table 1-2. Statistics of SNP marker usage

	No. of SNP loci
SoySNP 6k iSelect BeadChip	5,403
Genotyped loci	5,128
Loci of missed genotype rate exceed 10%	38
Loci of heterozygous genotype in the parents	6
Loci of monomorphic genotype in the parents	3,035
Polymorphic loci	2,049
Loci integrated into linkage map	1,981

Table 1-3. Summary statistics of each linkage group and their markers

Chr.	No. of markers	Genetic dist. (cM)	Avg. marker distance (cM)
Gm01	56	64.6	1.2
Gm02	96	174.5	1.8
Gm03	102	145.6	1.4
Gm04	85	142.4	1.7
Gm05	83	134.0	1.6
Gm06	120	173.7	1.5
Gm07	116	146.7	1.3
Gm08	141	135.3	1.0
Gm09	60	137.1	2.3
Gm10	91	117.2	1.3
Gm11	97	151.0	1.6
Gm12	91	134.2	1.5
Gm13	125	167.4	1.3
Gm14	121	119.6	1.0
Gm15	92	156.7	1.7
Gm16	91	109.7	1.2
Gm17	97	144.9	1.5
Gm18	142	141.8	1.0
Gm19	97	139.4	1.4
Gm20	78	137.5	1.8
Total	1981	2773.3	-
Average	99.1	138.7	1.4

Table 1-4. QTLs for branching and total pod number identified in the RIL population of Jiyu69 × SS0404-T5-76

Trait	Locus	Chr	Pos. (cM)[a]	Marker interval[b]	LOD[c]	Add[d]	PVE (%)[e]	No. of genes[f]	Known QTL[g]
Branching	*qBR6-1*	6	122	*Gm06_20486758_C_A* *Gm06_20943239_T_C*	10.3	-1.3	14.5	13	*Branching 3-1 (FT1)*
	qBR11-1	11	72	*Gm11_10721006_A_G* *Gm11_10847172_T_C*	6.1	-0.9	7.6	18	*Branching 3-3 (Satt251)*
	qBR12-1	12	98	*Gm12_5610878_T_C* *Gm12_6023395_T_C*	4.4	-0.8	5.7	32	- -
	qBR19-1	19	97	*Gm19_44636089_T_C* *Gm19_44761515_G_A*	7.5	1.0	9.5	13	*Branching 3-2 (Sat_286)*
Total pod number	*qPN6-1*	6	122	*Gm06_20486758_C_A* *Gm06_20943239_T_C*	4.8	-29.0	7.9	13	*Pod number 3-4 (Satt509)*
	qPN11-1	11	73	*Gm11_10847172_T_C* *Gm11_10926986_T_C*	6.6	-33.8	10.8	7	*Pod number 3-2 (Satt277)*

[a] Pos. represents genetic position of a QTL peak in the genetic linkage map constructed in this study.
[b] The prefix 'BARC_1.01_' is omitted from the marker names.
[c] LOD represents the maximum-likelihood LOD score of each QTL.
[d] Add represents the allelic additive effect.
[e] PVE represents the percent of phenotypic variance explained by the QTL.
[f] The number of protein-coding genes present within marker intervals, on the basis of *Glycine max* gene models ver. 1.1.
[g] Parentheses indicate markers associated with the known QTLs.

Table 1-5. Anchored genes and their functional annotations of identified QTL region

QTL loci	*G.max* gene ID	*A.th* ortholog	Gene symbol	Gene function	Description	Reference
qBR6-1, qPN6-1	Glyma06g23340	AT3G18570		Oleosin family protein	Increased branches	Aguilar-Martinez et al. 2007
	Glyma06g23380	AT1G49000				
	Glyma06g23400	AT1G03280		Transcription factor TFIIE, alpha subunit		
	Glyma06g23405					
	Glyma06g23410	AT3G18550	*ATTCP18,BRC1, TCP18*	TCP family transcription factor		
	Glyma06g23420	AT1G68310		Protein of unknown function (DUF59)		
	Glyma06g23447	AT1G73930				
	Glyma06g23474	AT3G24330		O-Glycosyl hydrolases family 17 protein		
	Glyma06g23502	AT3G51690		PIF1 helicase		
	Glyma06g23530	AT1G03390		HXXXD-type acyl-transferase family protein		
	Glyma06g23560	AT1G08510	*FATB*	fatty acyl-ACP thioesterases B		
	Glyma06g23570	AT5G22950	*VPS24.1*			
	Glyma06g23580	AT5G19950		Domain of unknown function (DUF1767)		
qBR11-1	Glyma11g14980	AT5G59300	*ATUBC7,UBC7*	ubiquitin carrier protein 7		
	Glyma11g14990					
	Glyma11g15000	AT3G46450		SEC14 cytosolic factor family protein / phosphoglyceride transfer family protein		
	Glyma11g15010	AT2G28760	*UXS6*	UDP-XYL synthase 6		
	Glyma11g15020	AT3G46440	*UXS5*	UDP-XYL synthase 5		
	Glyma11g15030	AT5G59960				
	Glyma11g15040	AT5G59950		RNA-binding (RRM/RBD/RNP motifs) family protein		
	Glyma11g15050	AT3G45980	*H2B,HTB9*	Histone superfamily protein		

	Glyma11g15060	AT3G45980	*H2B,HTB9*	Histone superfamily protein		
	Glyma11g15070	AT3G45980	*H2B,HTB9*	Histone superfamily protein		
	Glyma11g15075					
	Glyma11g15080					
	Glyma11g15090	AT3G45980	*H2B,HTB9*	Histone superfamily protein		
	Glyma11g15105					
qBR11-1	Glyma11g15120	AT5G59840		Ras-related small GTP-binding family protein		
	Glyma11g15130	AT5G59830				
	Glyma11g15140	AT3G53600		C2H2-type zinc finger family protein	Increase in lateral shoots	Nakagawa et al. 2005
	Glyma11g15150	AT5G59790		Domain of unknown function (DUF966)		
	Glyma12g07960	AT3G46290	*HERK1*	hercules receptor kinase 1		
	Glyma12g07970	AT1G07705		NOT2 / NOT3 / NOT5 family		
	Glyma12g07980	AT3G46220				
	Glyma12g07990	AT1G07710		Ankyrin repeat family protein		
	Glyma12g08000	AT3G46200	*aTNUDT9, NUDT9*	nudix hydrolase homolog 9		
	Glyma12g08010	AT2G28630	*KCS12*	3-ketoacyl-CoA synthase 12		
qBR12-1	Glyma12g08020	AT5G59730	*ATEXO70H7, EXO70H7*	exocyst subunit exo70 family protein H7	Highly branched inflorescence in EXO70A1	Synek et al. 2006
	Glyma12g08030	AT5G02600		Heavy metal transport/detoxification superfamily protein		
	Glyma12g08041	AT2G37170	*PIP2;2,PIP2B*	plasma membrane intrinsic protein 2		
	Glyma12g08050	AT2G37190		Ribosomal protein L11 family protein		
	Glyma12g08060	AT3G45210		Protein of unknown function, DUF584		
	Glyma12g08070	AT2G28380	*DRB2*	dsRNA-binding protein 2		

	Glyma12g08080	AT5G60690	*IFL,IFL1,REV*	Homeobox-leucine zipper family protein / lipid-binding START domain-containing protein	Axillary bud formation	Shi et al. 2016
	Glyma12g08090	AT2G28370		Uncharacterised protein family (UPF0497)		
	Glyma12g08100	AT1G07990		SIT4 phosphatase-associated family protein		
	Glyma12g08110	AT2G28350	*ARF10*	auxin response factor 10		
	Glyma12g08121					
	Glyma12g08131	AT1G08010	*GATA11*	GATA transcription factor 11		
	Glyma12g08140	AT5G60700		glycosyltransferase family protein 2		
	Glyma12g08150	AT2G28305	*ATLOG1,LOG1*	Putative lysine decarboxylase family protein		
	Glyma12g08160	AT2G28260	*ATCNGC15 ,CNGC15*	cyclic nucleotide-gated channel 15		
	Glyma12g08170	AT4G27490		3\'-5\'-exoribonuclease family protein		
	Glyma12g08180	AT5G45890	*SAG12*	senescence-associated gene 12		
	Glyma12g08186	AT5G22000	*RHF2A*	RING-H2 group F2A		
	Glyma12g08193	AT4G00730	*AHDP,ANL2*	Homeobox-leucine zipper family protein / lipid-binding START domain-containing protein		
	Glyma12g08200	AT5G45890	*SAG12*	senescence-associated gene 12		
qBR12-1	Glyma12g08210	AT2G28250	*NCRK*	Protein kinase superfamily protein		
	Glyma12g08230	AT1G08060	*MOM,MOM1*	ATP-dependent helicase family protein		
	Glyma12g08240	AT2G23770		protein kinase family protein / peptidoglycan-binding LysM domain-containing protein		
	Glyma12g08251					
	Glyma12g08260					
	Glyma12g08270	AT2G27990	*BLH8,PNF*	BEL1-like homeodomain 8	Initiation, maintenance and development of shoot	Kanrar et al. 2008; Rutjens et al. 2009;

					meristem	Ung et al. 2011
	Glyma19g37520	AT2G36530	*ENO2,LOS2*	Enolase		
	Glyma19g37530	AT5G42820	*ATU2AF35B, U2AF35B*	Zinc finger C-x8-C-x5-C-x3-H type family protein		
	Glyma19g37540	AT5G03740	*HD2C,HDT3*	histone deacetylase 2C		
	Glyma19g37550	AT2G04842	*EMB2761*	threonyl-tRNA synthetase, putative / threonine--tRNA ligase, putative		
	Glyma19g37561	AT2G36490	*DML1,ROS1*	demeter-like 1		
	Glyma19g37570	AT5G03730	*AtCTR1,CTR1,SIS1*	Protein kinase superfamily protein		
	Glyma19g37585	AT5G03720	*AT-HSFA3,HSFA3*	heat shock transcription factor A3		
qBR19-1	Glyma19g37600	AT2G36480		ENTH/VHS family protein		
	Glyma19g37600	AT4G04885	*PCFS4*	PCF11P-similar protein 4		
	Glyma19g37610	AT2G36480		ENTH/VHS family protein		
	Glyma19g37621	AT5G03700		D-mannose binding lectin protein with Apple-like carbohydrate-binding domain		
	Glyma19g37630	AT2G27770		Plant protein of unknown function (DUF868)		
	Glyma19g37640	AT2G36460		Aldolase superfamily protein		
	Glyma19g37660	AT5G03680	*PTL*	Duplicated homeodomain-like superfamily protein	Defect in auxin action and the alteration of auxin distribution	Li et al. 2008b
	Glyma11g15160	AT1G43860		sequence-specific DNA binding transcription factors		
	Glyma11g15180	AT3G46130	*ATMYB48*	myb domain protein 48		
	Glyma11g15180	AT5G59780	*ATMYB59*	myb domain protein 59		
qPN11-1	Glyma11g15190	AT5G59770		Protein-tyrosine phosphatase-like, PTPLA		
	Glyma11g15200	AT4G01470	*ATTIP1*	3,GAMMA-TIP3,TIP1;3		
	Glyma11g15210	AT3G46430				
	Glyma11g15220	AT2G28690		Protein of unknown function (DUF1635)		

Glyma11g15230	AT3G55280	*RPL23AB*	ribosomal protein L23AB

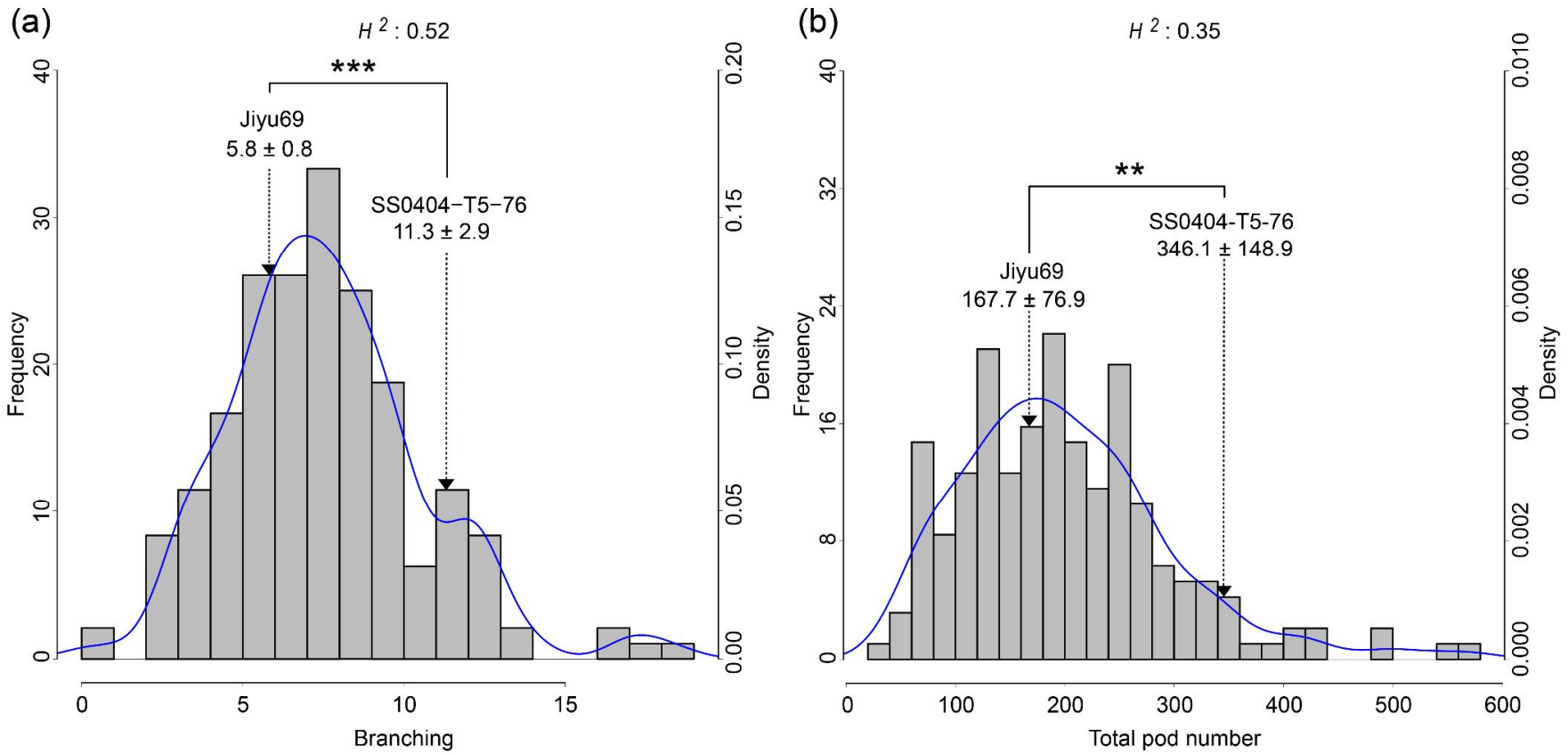


Figure 1-1 Phenotypic frequency distribution of branching (a) and total pod number (b) in the RIL population of Jiyu69 × SS0404-T5-76. Broad-sense heritability (H^2) of each phenotypic trait is represented on the upper side of each panel. Mean value and standard deviation of parental genotypes are indicated by vertical dotted lines. Blue line represents density plot. Significance level for the two parental genotypes, based on a one-tailed Student's T test, are indicated by asterisks: *** and ** indicate $P<0.001$ and $P<0.01$, respectivey.

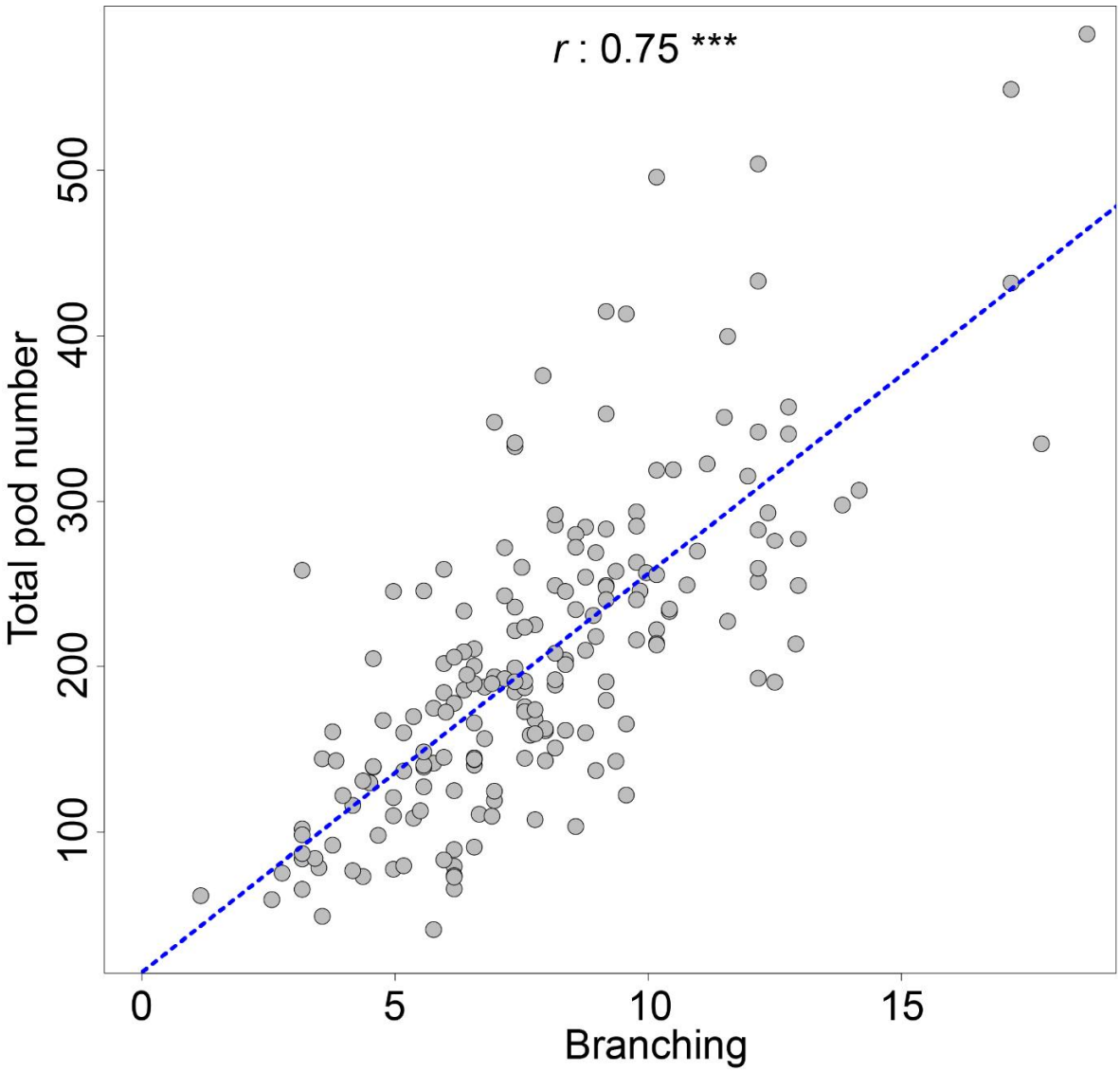


Figure 1-2 Correlation between branching and total pod number
Trend line based on the linear model is expressed as blue dashed line.

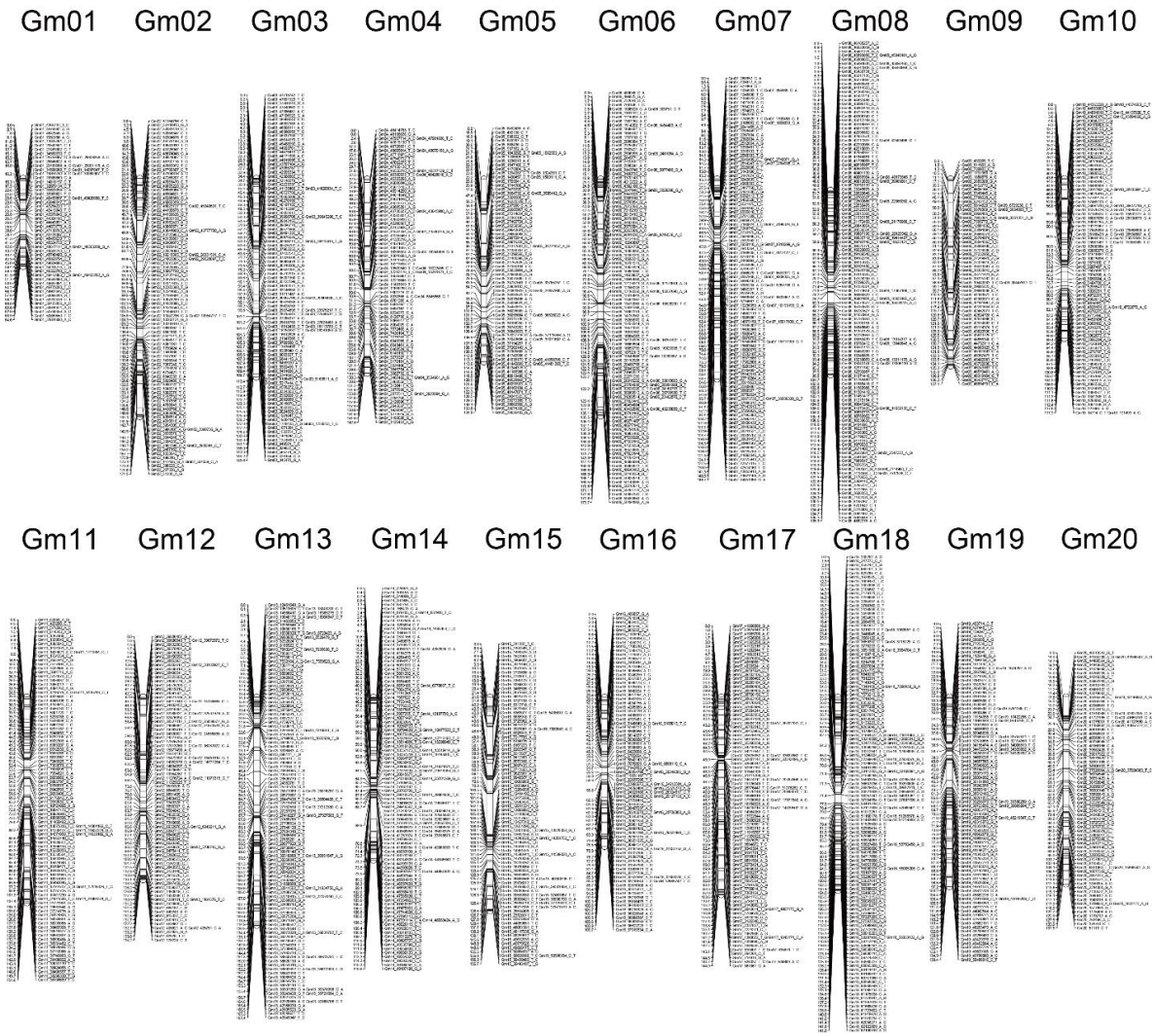


Figure 1-3 A soybean genetic linkage map using 1,981 polymorphic SNP markers of RIL population of Jiyu69 x SS0404-T5-76.

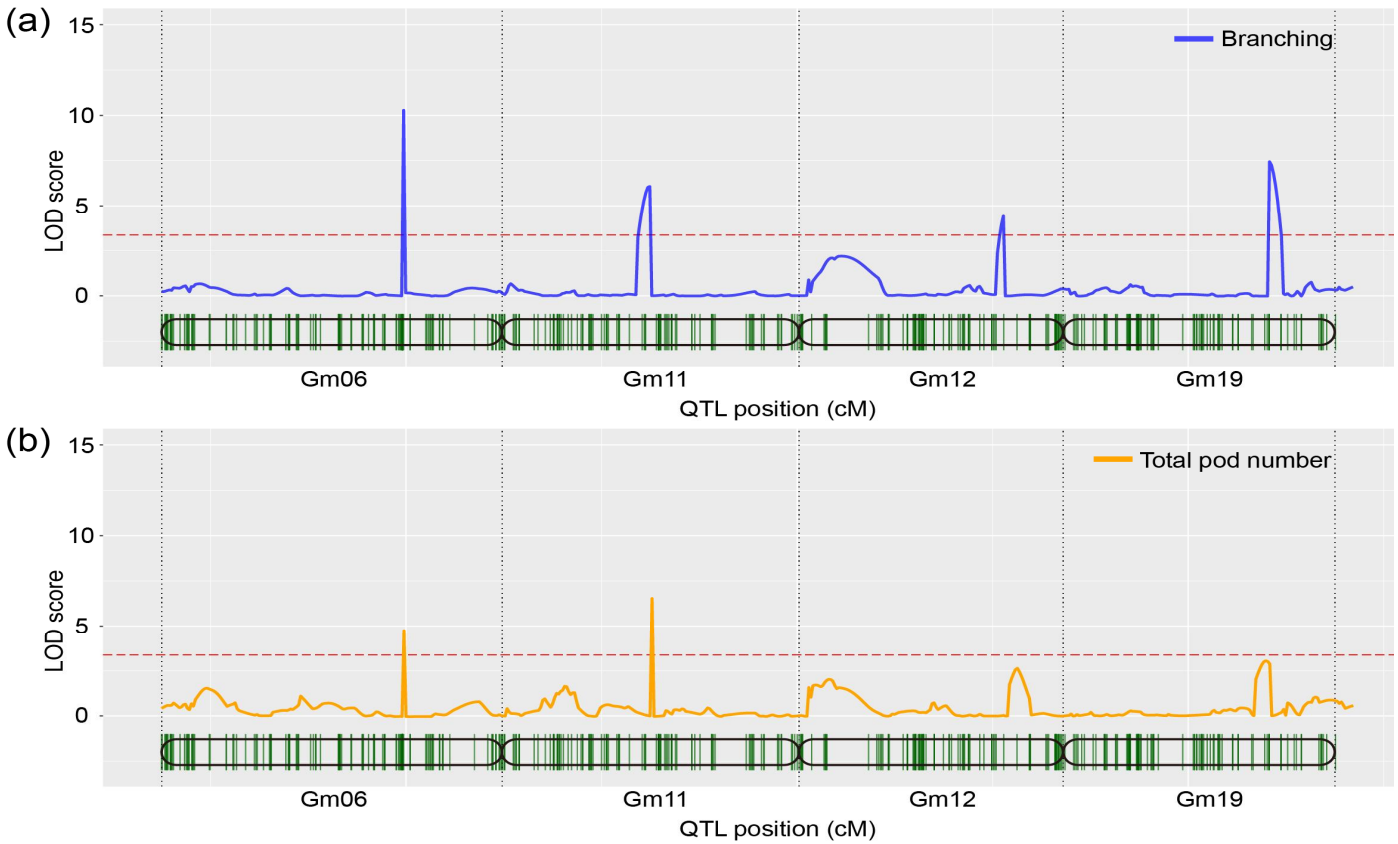


Figure 1-4 LOD score distributions of QTLs conferring branching (a) and total pod number (b) on soybean chromosomes 6, 11, 12, and 19. Red horizontal dotted lines indicate a LOD threshold at the 5% significance level under a thousand-permutation test for each trait. Green vertical bars on the chromosomes under the LOD score curves represent the positions of the SNP markers.

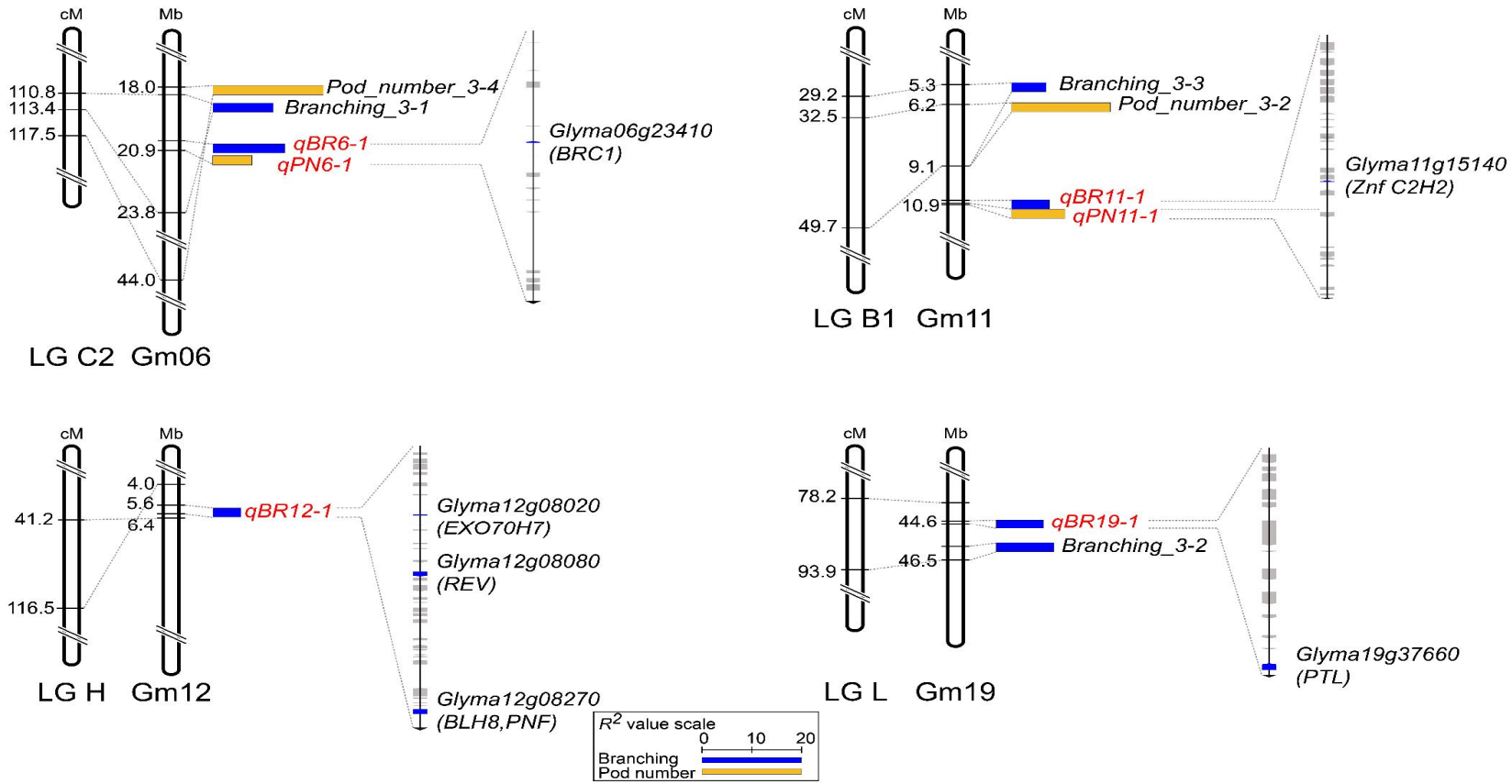


Figure 1-5 Map showing comparable regions between the QTLs identified in this study (red font) and previously reported QTLs (black font) for branching and total pod number on soybean Chrs 6, 11, 12, and 19. Phenotypic R^2 values of each QTL for branching and total pod number are indicated on the right side of the map by blue and yellow bars, respectively. The nearest sequence-based genetic markers flanking the known QTLs are used as bridges (dotted lines) to link the genetic map (GmComposite2003) and the physical map

(*Glycine max* reference genome version 1.01). Protein-coding genes between the markers flanking the QTLs are shown as small gray rectangles, and the six selected candidate genes are indicated by blue rectangles.

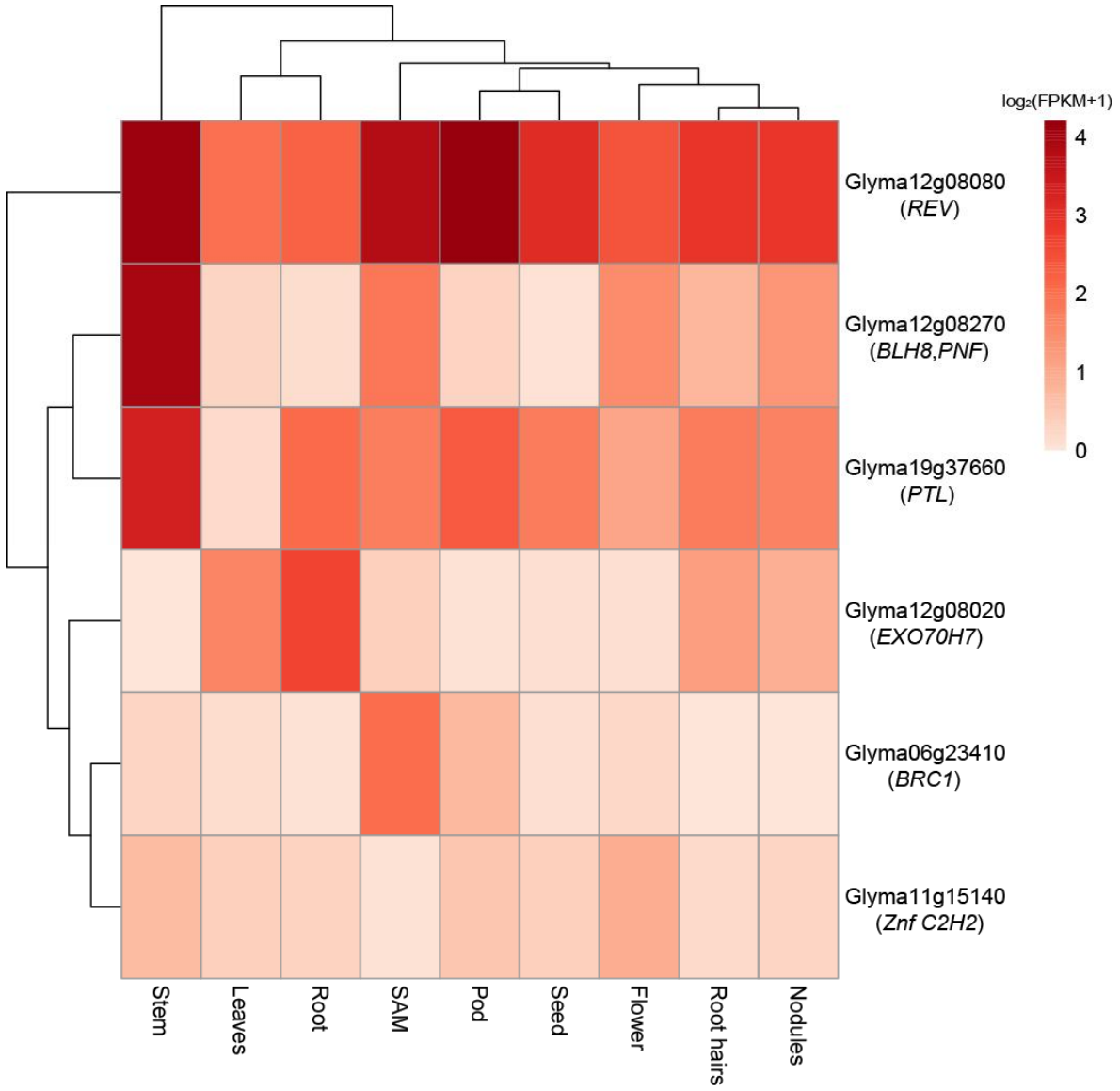


Figure 1-6 Heat map of the expression profiles of six candidate genes involved in the control of branch number in nine different soybean tissues. FPKM values of six candidate genes were retrieved from RNA-seq data (Schmutz et al. (2010) at phytozome v.10 (http://genome.jgi.doe.gov/). The heat map with hierarchical clustering of the genes was constructed based on $\log_2$(FPKM + 1) values using R package pheatmap. The color scale represents the relative transcript abundance of the genes in nine soybean tissues.

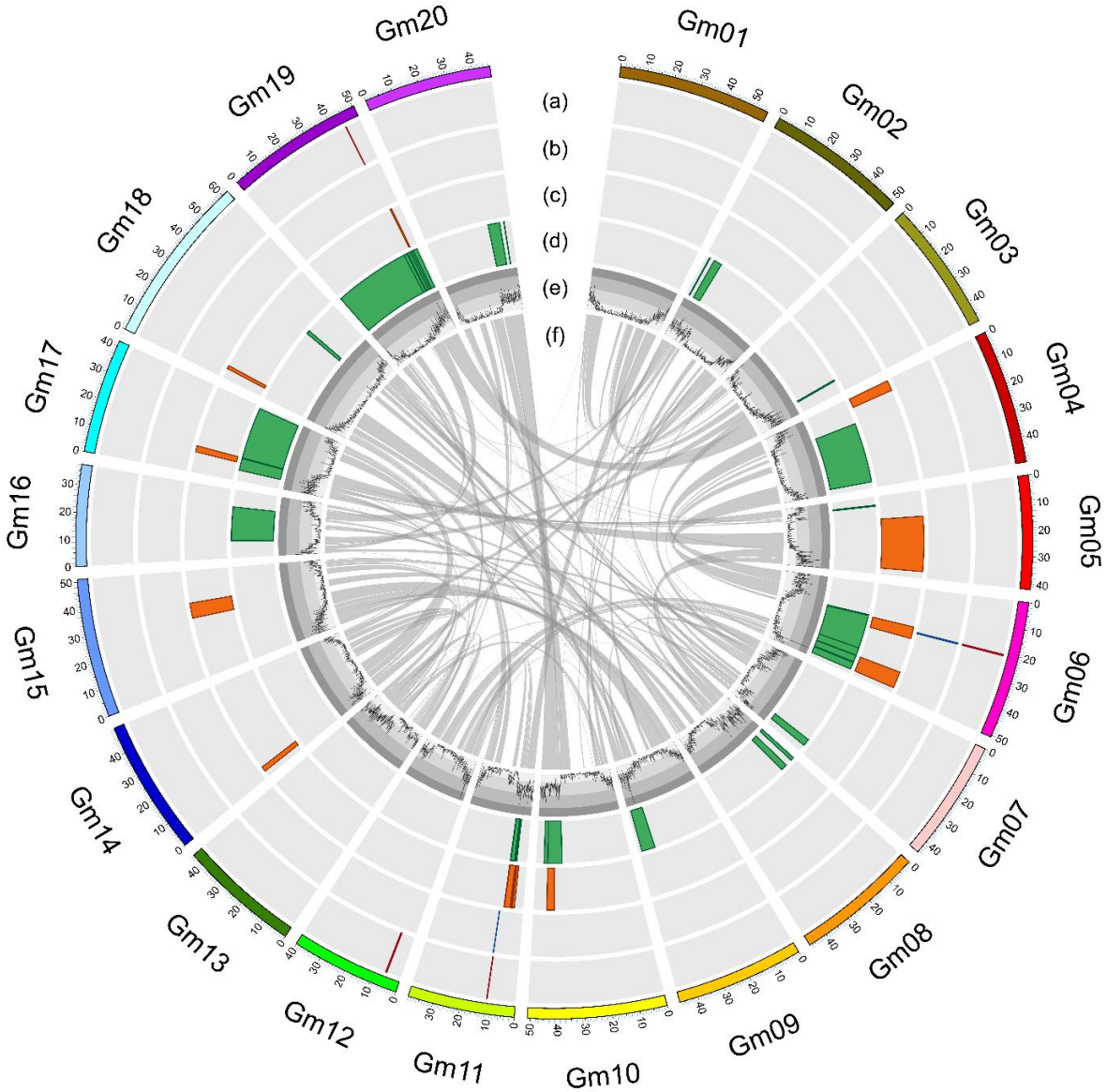


Figure 1-7 Genomic distributions of QTL for branching and total pod number QTLs of branching and total pod number which had identified in this study are presented in layer (a) and (b). To convenient compare the QTLs with previously reported QTLs, previous QTLs presented in layer (c) and (d). Gene density and duplication synteny blocks are displayed in layer (e) and (f).

1

Chapter 2.

Genetic variations underlying branching in soybean (*Glycine max* (L.) Merill)

Abstract

Number of branches in soybean is one of the important factor that directly affects yield components of pod or seed number per plant. There are numbers of QTLs associated with number of branches, but genes controlling number of branches have not been characterized. We recently identified a major QTL (*qBR6-1*) with a LOD score of 10.3, spanning 460 Kb on Chromosome 6 and including 13 protein-coding genes. To narrow down this QTL regions and identify a candidate gene, we analyzed association between genetic variations, located in *qBR6-1*, and number of branches that measured under three different environments using a set of 430 soybean core germplasms. Two markers in intergenic regions and four markers in two different genes were significantly associated with the number of branches. The two genes were *BRNACHED 1* (*BRC1*) gene and gene encoding transcription factor TFIIE alpha subunit. Meanwhile, we developed a

set of near isogenic lines (NILs) that showed significant difference in branch number and derived from a F_6 residual heterozygous line of mapping population used in previous study. Resequencing of the NILs resulted that 99.9% of background were homozygous. The gene *BRC1*, that acts as a negative regulator of branching in *Arabidopsis*, was up-regulated in shoot region of the less-branching NIL compared to more-branching NIL. The other candidate gene showed no expressional difference between the NILs. Moreover, genotype association of missense and upstream SNP of *BRC1* with branch number were identified in additional 59 soybean germplasms. Protein sequence alignment of *BRC1*, *G. max* paralogues and homologues in *A. thaliana* and *O. sativa* resulted that the missense mutation was not presented in conserved amino acid residue. Through the results, it was revealed that branch development was attributed to expression of *BRC1*. However, the functional validation using transformation of *BRC1* gene is still needed.

Keywords validation, QTL, number of branches, *BRNACHED1* (*BRC1*), soybean

Introduction

Soybean yield is a complex trait which is affected by numerous genetic factors, environmental factors and interactions between two factors (Hamawaki et al. 2012). Broad sense heritability of soybean yield ranges from 20 to 30 %, and correlations between soybean yield and yield related traits has been reported (Hamawaki et al. 2012; Ghodrati 2013). According to Ghodrati, number of seeds per plant shows strong positive correlation with number of branches per plant (Ghodrati 2013), suggests that the number of branches is a factor affecting yield of soybean.

The branch number of soybean interacts with various environmental factors including light intensity, soil water, and planting density (Linkemer et al. 1998; Board 2000; Agudamu et al. 2016). Among these environmental factors, effect of planting density on branch outgrowth is well-established (Cox et al. 2010; Agudamu et al. 2016). Branching plasticity can increase or decrease branch number depending on the planting densities (Agudamu et al. 2016), indicate that branch development is more likely to be controlled by expression of causal genes, not by functional mutation of protein sequence. Although the influences of various environmental factors, fundamental differences in number of branches and yield have been observed

between soybean cultivars in low planting density (Board and Kahlon 2013). This suggest that branch development is regulated by underlying genetic factors.

In soybean, a total of 18 loci associated with number of branches and anchored in the soybean chromosomes (or linkage map) have been identified using sets of simple sequence repeat (SSR) or single nucleotide polymorphism (SNP) markers (Chen et al. 2007; Li et al. 2008; Sayama et al. 2010; Yao et al. 2015; Chapter 1). Among the identified loci, nine QTLs anchored in soybean chromosome 4, 5, 6, 11, 14 and 19 are major one which explains more than 10% of phenotypic variations. A major QTL, *qBR6-1*, we identified in previous study, was overlapped with a major QTL, *qBR1* (Sayama et al. 2010) and had narrower region of 460 kb containing 13 genes (Chapter 1). The *BRC1* gene located in this QTL region was suggested as promising candidate gene (Chapter 1). However, the gene controlling branch number has not been fully characterized.

In this study, using a set of soybean core collection which was comprised by 430 soybean germplasms, genetic association was analyzed for markers located in *qBR6-1*. As a result, one marker that anchored in the exon of *BRANCHED1* (*BRC1*) showed significant association with branch numbers measured in two different environments, and three markers that located in intron of transcription

factor TFIIE alpha subunit coding gene were associated with branch numbers measured in three environments (ANOVA and regression analysis, *q-value* < 0.05). Meanwhile, a pair of near-isogenic line (NILs) for *qBR6-1* developed from a F_6 residual heterozygous line (RHL) of mapping population used in previous QTL study. This set of NILs showed 99.9% of homozygous genetic background and phenotypic difference (*p-value* < 0.05 for field experiment; *p-value* < 0.001 for green house experiment). Comparison of expression of two candidate genes between the two NILs resulted that only *BRC1* gene was significantly down-regulated in the NIL exhibited more branches (*p-value* < 0.001). Moreover, genetic associations of missense and upstream SNP for *BRC1* with number of branches were identified within a set of 59 soybean germplasms. Multiple protein sequence alignment of *BRC1*, its soybean paralogues and orthologues of *BRC1* in *A. thaliana* and *O. sativa* resulted that the amino acid residue mutated by missense SNP was located in functional domain of *BRC1*, but not conserved in all tested protein sequences. This represented that branch development was more likely attributed by the expression difference of *BRC1* gene.

In conclusion, *BRC1* gene seemed to be a strong candidate gene promoting branch outgrowth by lowering its expression level. Additionally, it was identified that the genetic factors associated with

branch numbers and located in upstream of *BRC1* gene seemed to control the expression of *BRC1* gene. This study will be the basis of high-yield soybean breeding and branch development in other plant species.

Materials and methods

Plant materials and phenotype assessment

Phenotypes of soybean core collection constituted by 430 soybean germplasms were investigated in three different geographical regions of Wanju (N 35° 50' 27.384", E 127° 2' 46.1826"), Cheonan (N 36° 49' 49.2816", E 127° 10' 1.9122") and Ochang (N 36° 43' 14.0982", E 127° 26' 1.1148") in Republic of Korea at R8 stage. Phenotypes of three biological replicates were collected. Planting density of three different field were different.

Near-isogenic line (NIL) for *qBR6-1* locus had derived from a F_6 residual heterozygous line (RHL) of mapping population which used in previous QTL study. Seven plants for each of the two NILs were selected by foreground selection for flanking marker of *qBR6-1* and planted in experimental farm of Seoul National University, under 60 cm planting and 80 cm row spacing to exhibit maximum branches. Phenotypes of all plants were assessed for the number of branches which outgrew from main stem and had more than two nodes in R8 stage.

Each NILs were also cultivated on a glass house located in experimental farm of Seoul National University to examine the phenotype and extract total RNA in winter season. Generally, in winter season, light intensity is quite lower than in regular cultivation period.

To compensate the light intensity and exhibit branch outgrowth, averagely 2779.3 lux of light was additionally irradiated for 12 hours per day using metal halide lamp in day time.

Additional 59 germplasms provided by Germplasm Resources Information Network (GRIN, www.ars-grin.gov) of United States Department of Agriculture (USDA) were used for validate genetic factors located in *BRANCHED1* (*BRC1*) gene and associated with number of branches. Phenotypes of these germplasms, which were surveyed and posted on the GRIN, were downloaded and used for analysis.

Genotypes of soybean core collection

Genotypes of soybean core collection that comprised by 430 germplasms were analyzed using Axiom Soybean Genotyping Array that containing 170,223 SNP markers. The data from genotype panel was provided by Dr. Soon-Chun Jeong.

Association analysis in germplasms

Genetic variations underlying branching were analyzed within numbers of selected markers which were included QTL, *qBR6-1*. A total of 60 markers were located in *qBR6-1* region. Among the 60 markers, 15 homozygous markers were excluded in following analysis. Each of

single markers were subjected to single marker ANOVA and regression analysis using python script implemented by scipy.stats module (Millman and Aivazis 2011). Resultant *p-value*, then, submitted for *post hoc* test to determine false positives using FDR-correction suggested by Benjamini and Hochberg (Benjamini and Hochberg 1995) implemented in FDR online calculator (https://www.sdmproject.com/utilities/?show=FDR).

For association analysis of variants located in *BRC1* gene using additional 59 accessions from GRIN, SNPs of non-synoymous and upstream upto 2kb from start codon were designed using Primer3 (Untergasser et al. 2012) based on the comparison of whole genome sequences of two NILs and its parental line of mapping population. Designed markers were listed in Table 2-1. Regions include variants were amplified by PCR and sequenced by ABI3730xl (Applied Biosystems, Carlsbad, CA, USA). The association of genotype and phenotype were analyzed using ANOVA.

Genome-wide association analysis

Genome-wide association study (GWAS) for branch number was conducted based on the genotypes of soybean core collection and phenotypes measured in three geographical regions in Korea using GAPIT (Zhang et al. 2010). Prior to the GWAS analysis, markers which

minor allele frequency below 3% for total number of alleles and anchored in the scaffolds, were eliminated. Kinship and structure model were selected under compressed MLM algorithm using group.from and group.to options. To confirm the result from automated GAPIT analysis, association analysis under manually specified kinship and population structure model which analyzed by STRUCTURE (Pritchard et al. 2000) was also conducted.

DNA extraction of NILs and germpalsms provided by GRIN for resequencing of NILs

DNA was extracted from young and fresh leaf tissues, lysed by bead and TissueLyser II (QIAGEN, GmbH, Hilden, Germany), using Exgene Plant SV mini DNA extraction kit (GeneAll, Cat No. 117-152, Seoul, South Korea) for NIL and germplasm from GRIN, USDA by following the instruction of kit. Intact DNA of NIL and two parental lines were sequenced by Illumina HiSeqX (Illumina, Inc., San Diego, CA, USA) for more than 30 times of soybean genome size.

All the NGS reads were aligned against v1.01 of reference genome sequences of *Glycine max* using Burrows-Wheeler aligner (BWA) (Li and Durbin 2009) with default parameter, and variants including SNPs and INDELs were called using Samtools (Li et al. 2009). After calling variants, annotation was conducted by SnpEFF (Cingolani et al. 2012).

As downstream analysis for annotated variants, filtering of QUAL < 50 and depth < 10 was conducted by home-made python codes or simple shell commands.

RNA extraction and qRT-PCR of candidate genes in NILs

According to Chapter 1, it has been shown that the candidate *BRC1* gene was expressed in shoot apical meristem using public available RNA-seq data (Chapter 1). Furthermore, it has been reported that the branch were produced between R1 to R5 stage (Board and Settimi 1986). Based on these findings, RNA was extracted from healthy tissue of shoot region (5 mm of shoot apical region) in R1 stage, lysed by TissueLyser II and bead, using Ribospin Plant (GeneAll, Cat No. 307-150, Seoul, South Korea) for NIL with a modification of adding 2% volume of 2-mercaptoethanol (CAS No. 60-24-2, Sigma-Aldrich, Co, St. Louis, MO) into lysis buffer. Gene expressions were quantified from intact RNA using primers listed in Table 2-1 by LightCycler96 (Roche, Indianapolis, IN, USA). Markers for each gene and reference gene (*ACT11*) were tested for three biological replications and three technical replications.

Miltiple protein sequence alignment of *BRC1* gene, *G. max* paralogues and homologues in *A. thaliana* and *O. sativa*

To validate effect of missense SNP identified in *BRC1* gene, protein sequences of *BRC1*, three soybean paralogues of *BRC1* and homologues of *BRC1* in *A. thaliana* and *O. sativa* were aligned using Muscle aligner with default parameter (Edgar 2004). Protein sequences of *BRC1* and three paralogues encoding TEOSINTE BRANCHED/CYCLODEA/PROLIFERATING CELL FACTORS 18 (TCP18) were identified based on the conservation analysis of TCP domain (PF03634) using HMMER software (Eddy 2009) and reference genome annotation information (Schmutz et al. 2010). TCP TF family in *A. thaliana* and *O. sativa* were retrieved from PlantTFDB (http://planttfdb.cbi.pku.edu.cn/) (Jin et al. 2017). Homologues of TCP18 in *A. thaliana* and *O. sativa* were filtered based on the protein sequence homology using BLASTP (Boratyn et al. 2013).

Results

Number of soybean branches in different geographical regions of Korea

The number of branches were observed for 430 soybean germplasms in three geographic replications, including Cheonan, Wanju and Ochang in Korea. All the phenotypic distributions measured in three geographic replications were pair-wisely compared to each other to validate correlation of branch number (Fig. 2-1). As a result, branch numbers were normally distributed in all three regions and positive correlation was observed in pair-wise comparison between the distributions measured in different geographic replication ($r < 0.5$, Fig. 2-1).

Association of genetic variations underlying branching using 430 germplasms

Association analysis within the genetic variations located in *qBR6-1* resulted that a total of six out of 45 markers were significantly associated with at least two phenotypic distributions measured in three different geographic regions (*q-value* < 0.05) (Table 2-2; Fig. 2-2). Among the six markers, three markers were consistently associated with the branch number measured in all three geographic replications, whereas, the other three markers showed significant association with

the branch number measured in two geographical regions (Table 2-2; Fig. 2-2b). In the genomic region, two out of six markers were located in the intergenic region of Glyma06g23340 - Glyma06g23380 and Glyma06g23380 - Glyma06g23400. Another three out of six markers were located in the intron of Glyma06g23400 encoding transcription factor TFIIE alpha subunit. The other one was anchored in exon of *BRANCHED 1* (*BRC1*, Glyma06g23410) (Table 2-2; Fig. 2-2b).

However, genome-wide association study (GWAS) showed that there was no significant genetic association for number of branches measured in all three geographic replications under both automatically and manually specified models (Fig. 2-3).

Foreground and background selection of NILs

A residual heterozygous line (RHL) was identified in QTL mapping population of F_6 recombinant inbred lines. Number of progeny of the RHL at F_7 generation were selected by foreground selection using flanking marker of *qBR6-1*. Genotypes of selected progenies were segregated into two different genotypes originated from different parent. However, due to the existence of the QTL in peri-centromeric region of soybean chromosome 6 (Fig. 2-2a), recombination between the flanking markers of *qBR6-1* was not identified.

Parental lines and two near-isogenic lines representing each

parental genotypes for *qBR6-1* were sequenced for background selection by Illumina HiSeqX (Illumina, Inc., San Diego, CA, USA) for >30 average depth. Average 223 million NGS reads which were 34 folds of soybean genome were generated and 99.5~99.6% of total reads were properly mapped to the reference genome (v.1.01) of soybean (Table 2-4). After filtration of each position which had below 50 for QUAL value and 10 for depth, total genotyped positions reached 837 million positions (Table 2-5). Among the positions only 225 thousand positions including QTL region were heterozygous for the two NILs. This indicated that 99.9% of genome sequence of two NILs were homozygous and it was more homozygous than theoretical homozygosity, respectively (Table 2-5). Based on the sequence comparison of NILs, numbers of variants including SNPs and INDELs were also identified for the candidate genes which were associated with branch numbers in soybean core collection (Table 2-6).

Difference of phenotype and candidate gene expression in NILs

Phenotype of selected NILs were evaluated by field and green house experiments (Fig. 2-4; Table 2-3). Phenotypic differences in number of branches between NILs were analyzed using ANOVA showed that statistically significant differences were exhibited by segregation of QTL, *qBR6-1*, in field experiment and green house experiment (*p-value*

< 0.05 for field experiment and *p-value* < 0.001 for green house experiment, Fig. 2-4; Table 2-3).

Comparison of expression level for the two candidate genes associated with branch number at shoot region in R1 stage, was conducted using qRT-PCR. As a result, *BRC1* gene was differentially expressed in NILs (*p-value* < 0.001, Fig. 2-2c; Fig. 2-5). Moreover, expression of *BRC1* gene was higher in NIL which exhibited less branches as shown in Aguilar-Martínez et al. (Aguilar-Martínez et al. 2007). On the contrary, the other candidate gene, which encodes transcription factor TFIIE alpha subunit, had no significant differences in expression between the two NILs (*p-value* > 0.2, Fig. 2-2c; Fig. 2-5).

Validation of genetic association of *BRC1* and its upstream sequences

A total 108 and 34 sequence variants including SNP and insertion/deletion (INDELs) were identified in genic region and up/downstream of candidate genes (Glyma06g23400 and Glyma06g23410) (Table 2-6). Among the variants, genetic variations of one missense SNP and 10 upstream SNPs which were located within 2kb upstream were analyzed. Using a set of 59 germplasms, association of genetic variations resulted missense variant causing amino acid change from glutamate to lysine in 199th amino acid and upstream

sequences located in 995 and 1,013 bp upstream from start codon of *BRC1* (Glyma06g23410) gene were associated with branch number (*p-value* < 0.01, Fig. 2-2d).

Protein sequence alignment of *BRC1* gene, *G. max* paralogues and orthologues in *A. thaliana* and *O. sativa*

Protein sequence alignment of candidate gene *BRC1*, *G. max* paralogues and *A. thaliana* and *O. sativa* orthologues of *BRC1* was conducted to elucidate the effect of missense mutation. As a result, the missense SNP causing mutation in functional domain of *BRC1* (TCP domain), however, changed protein residue was conserved neither in soybean paralogous genes nor in *A. thaliana* and *O. sativa*, whereas numbers of other amino acid residues were conserved over all tested protein sequences of *BRC1* genes (Fig. 2-6).

Discussion

In soybean, branch number affects yield components of pod and seed number per plant by producing additional flowers and pods on branches (Ghodrati 2013). Therefore, understanding of genetic factor that regulates branch development in soybean is critical for molecular breeding of yield increasing in soybean. Thus far, a total of 18 QTLs associated with branching were identified and reported in soybean (Chen et al. 2007; Li et al. 2008; Sayama et al. 2010; Yao et al. 2015; Chapter 1). Among these QTLs, recently reported major QTL, *qBR6-1*, was overlapped with major QTL, *qBR1*, and had narrower region of 430 kbp along with 13 genes (Sayama et al. 2010; Chapter 1). However, the major gene controlling number of branches have not been characterized in this QTL region. For successful application of molecular breeding, genetic variations underlying branching should be analyzed in various environments and genetic backgrounds.

In this study, association analysis for genetic variations located in *qBR6-1* region was conducted in three different geographic replications. Resultantly, six markers were commonly associated with branch numbers in at least two geographic replications (Fig. 2-2b). Among the six markers, one and three markers anchored in *BRANCHED1* (*BRC1*) gene and gene encoding transcription factor TFIIE alpha subunit (Fig. 2-2b). It has been reported that the *BRC1* gene acts as negative

regulator of branch development under auxin signaling pathway in *Arabidopsis*: When the gene expression is down regulated, branch outgrowth is facilitated, otherwise, branch development is inhibited (Aguilar-Martínez et al. 2007). Based on this finding, we compared gene expression level of the two candidate genes in a pair of NILs. As result, the gene expression of *BRC1* was significantly down regulated in more branching NIL compared to less branching one (Fig. 2-2c; Fig. 2-5). This suggested that the expression level of *BRC1* gene may regulate branch development in soybean. Genetic factors that causing missense mutation and located in upstream sequence of *BRC1* were associated with branch numbers (Fig. 2-2d). Protein sequence alignment of *BRC1*, soybean paralogs and orthologues in *A. thaliana* and *O. sativa* to elucidate the effect of missense SNP resulted that the amino acid residue presented in functional motif but was not conserved in all *BRC1* homologues (Fig. 2-6). Through the results, it could be concluded that the expression level of *BRC1* gene negatively regulated branch development in soybean. This conclusion could be also supported by the fact that the difference of branch number that caused by the different planting density in a same genotype (Board and Kahlon 2013; Agudamu et al. 2016).

However, in GWAS analysis, no significant association was identified (Fig. 2-3). The phenotypic distributions of tested population

showed normal distributions (Fig. 2-1) represented that the branch numbers were controlled by multiple genetic factors. The pairwise correlations of phenotype distributions indicated that branching of soybean were immensely influenced by environments (Fig. 2-1). The phenotypic variations that explained by genetic factors were ranged from 26.3 to 33.2% (Fig. 2-7), suggested that the residual variations of branch numbers were too large that identification of major genetic factors were obscured in this study.

In conclusion, *BRC1* gene appears to be a strong candidate gene that promoting branch development by lowered expression level. Thus, association of genetic factors that putatively control expression level of *BRC1* gene was identified in 1 kbp upstream of *BRC1*. However, the functional validation of *BRC1* gene using transformation or CRISPR system is still needed to reveal the relationship of the gene and soybean branch development. This study will be the basis for understanding of soybean branch development and helps in molecular breeding of high yield soybean.

References

Agudamu, Yoshihira T, Shiraiwa T (2016) Branch development responses to planting density and yield stability in soybean cultivars. Plant Prod Sci 19:331–339. doi: 10.1080/1343943X.2016.1157443

Aguilar-Martínez JA, Poza-Carrión C, Cubas P (2007) *Arabidopsis BRANCHED1* Acts as an Integrator of Branching Signals within Axillary Buds. Plant Cell 19:458–472. doi: 10.1105/tpc.106.048934

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol 57:289–300. doi: 10.2307/2346101

Board J (2000) Light Interception Efficiency and Light Quality Affect Yield Compensation of Soybean at Low Plant Populations. Crop Sci 40:1285–1294. doi: 10.2135/cropsci2000.4051285x

Board JE, Kahlon CS (2013) Morphological Responses to Low Plant Population Differ Between Soybean Genotypes. Crop Sci 53:1109–1119. doi: 10.2135/cropsci2012.04.0255

Board JE, Settimi JR (1986) Photoperiod Effect before and after Flowering on Branch Development in Determinate Soybean 1. Agron J 78:995–1002. doi: 10.2134/agronj1986.00021962007800060013x

Boratyn GM, Camacho C, Cooper PS, et al (2013) BLAST: a more efficient report with usability improvements. Nucleic Acids Res 41:W29–W33. doi: 10.1093/nar/gkt282

Chen Q, Zhang Z, Liu C, et al (2007) QTL Analysis of Major Agronomic Traits in Soybean. Agric Sci China 6:399–405. doi: 10.1016/S1671-2927(07)60062-5

Cingolani P, Platts A, Wang LL, et al (2012) A program for annotating

and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w1118*; *iso-2*; *iso-3*. Fly (Austin) 6:80–92. doi: 10.4161/fly.19695

Cox WJ, Cherney JH, Shields E (2010) Soybeans Compensate at Low Seeding Rates but not at High Thinning Rates. Agron J 102:1238–1243. doi: 10.2134/agronj2010.0047

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. In: Genome Informatics 2009. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO., pp 205–211

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

Ghodrati G (2013) Study of genetic variation and broad sense heritability for some qualitative and quantitative traits in soybean (*Glycine max* L.) genotypes. Curr Opin Agric 2:

Hamawaki OT, Sousa LB de, Romanato FN, et al (2012) Genetic parameters and variability in soybean genotypes. Comun Sci 3:76–83

Jin J, Tian F, Yang D-C, et al (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45:D1040–D1045. doi: 10.1093/nar/gkw982

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl 25:1754–1760. doi: 10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl

25:2078–2079. doi: 10.1093/bioinformatics/btp352

Li W, Zheng D, Van K, Lee S (2008) QTL Mapping for Major Agronomic Traits across Two Years in Soybean (*Glycine max* L. Merr.)

Linkemer G, Board JE, Musgrave ME (1998) Waterlogging effects on growth and yield components in late-planted soybean. Crop Sci 38:1576–1584

Millman KJ, Aivazis M (2011) Python for Scientists and Engineers. Comput Sci Eng 13:9–12. doi: 10.1109/MCSE.2011.36

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Sayama T, Hwang T-Y, Yamazaki H, et al (2010) Mapping and comparison of quantitative trait loci for soybean branching phenotype in two locations. Breed Sci 60:380–389. doi: 10.1270/jsbbs.60.380

Schmutz J, Cannon SB, Schlueter J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. doi: 10.1038/nature08670

Untergasser A, Cutcutache I, Koressaar T, et al (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115. doi: 10.1093/nar/gks596

Yao D, Liu ZZ, Zhang J, Liu SY, Qu J, Guan SY, Pan LD, Wang D, Liu JW, Wang PW (2015) Analysis of quantitative trait loci for main plant traits in soybean. Genet Mol Res 14:6101-6109. doi:10.4238/2015.June.8.8

Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360. doi: 10.1038/ng.546

Table 2-1. Primer used in this study

Forward primer	Reverse primer	Tm	Used in	Included variants/Genes
CTTGGCACCTGCAGATTCTT	CCTCTCACAATCGCAATGAC	67	Genomic region amplification	Gm06:20668622_C/T
CTACAGATTGAGAAAGACAGCACAA	CTGCAATCAGATTTGCTACAACTAA	60	Genomic region amplification	Gm06:20669831_C/G, Gm06:20669898_A/T, Gm06:20670000_G/T, Gm06:20670307_C/A
TCAAAATCGAATCAAACCGA	TTTTCAATCCAACCCAACTCA	60	Genomic region amplification	Gm06:20670427_C/T, Gm06:20670429_T/C, Gm06:20670445_G/A
TGATTGCAGCAAAACTTATCG	CACGTGAAGTGCCACAACTT	60	Genomic region amplification	Gm06:20670669_C/T, Gm06:20670809_C/T, Gm06:20670826_G/A
AATCAGTGCATTTGACCCTCTT	AGCTAGCACTCCACGATTTCTC	57	qRT-PCR	*BRC1*
CAAGTTGCACACGCACTCTT	TAATCGCAATGCATCCAGAG	57~59	qRT-PCR	TFIIE alpha subunit
CGGTGGTTCTATCTTGGCATC	GTCTTTCGCTTCAATAACCCTA	57~60	qRT-PCR	*ACT11*

Table 2-2. SNP markers associated with branch number in soybean

Marker name	Chromosome	Position	Geographic loc.	*p-value*	*q-value*	LOD	PVE(%)	Annotations
AX-90417900	Gm06	20511878	Wanju	0.0006	0.015	3.2	2.8	intergenic of Glyma06g23340-Glyma06g23380
			Cheonan	0.0025	0.028	2.6	2.2	
			Ochang	0.0002	0.004	3.7	3.3	
AX-90362777	Gm06	20562451	Wanju	0.0029	0.021	2.5	2.1	intergenic of Glyma06g23380-Glyma06g23400
			Ochang	0.0034	0.025	2.5	2.1	
AX-90403869	Gm06	20569695	Wanju	0.0009	0.015	3.0	2.7	intron of Glyma06g23400
			Cheonan	0.0021	0.028	2.7	2.3	
			Ochang	0.0003	0.004	3.5	3.3	
AX-90318984	Gm06	20570305	Wanju	0.0010	0.015	3.0	2.6	intron of Glyma06g23400
			Cheonan	0.0016	0.028	2.8	2.4	
			Ochang	0.0001	0.004	3.9	3.5	
AX-90457845	Gm06	20572237	Wanju	0.0033	0.021	2.5	2.1	intron of Glyma06g23400
			Ochang	0.0004	0.005	3.4	3.0	
AX-90520481	Gm06	20668226	Cheonan	0.0001	0.005	3.9	3.5	exon of Glyma06g23410
			Ochang	0.0040	0.026	2.4	2.0	

Table 2-3. Phenotypic differences of NILs in field and green house

Trait	NIL-less branches (Jiyu69 type)	NIL-more branches (SS0404-T5-76 type)	P-value
No. of branches in field	12.2 ± 2.1	14.5 ± 1.1	0.019
No. of branches in green house	3.4 ± 0.7	5.8 ± 1.7	0.0001

Table 2-4. Mapping statistics of NGS reads for parental lines and NILs derived from RHL.

	Jiyu69 (ratio)	SS0404-T5-76 (ratio)	NIL-less branches (ratio)	NIL-more branches (ratio)
No. of total reads	240,237,793 (N/A)	199,084,924 (N/A)	235,661,714 (N/A)	219,717,119 (N/A)
No. of mapped reads	239,183,154 (99.6%)	198,200,138 (99.6%)	234,415,474 (99.5%)	218,784,664 (99.6%)
No. of properly paired reads	216,944,472 (90.3%)	182,819,718 (91.8%)	212,816,170 (90.3%)	207,039,586 (94.2%)
No. of singltons	625,081 (0.3%)	513,794 (0.3%)	649,530 (0.3%)	498,627 (0.2%)
Average depth	35.8 (N/A)	30.5 (N/A)	35.2 (N/A)	34.4 (N/A)

Table 2-5. Statistics of background selection for NILs

	Min. depth of 10
Total genotyped position	837,223,446
No. of position which show different genotype in NILs	225,555
Homozygousity of NILs	99.97%
Theoretical homozygousity	98.44%

.

Table 2-6. Statistics of SNP/INDEL which was identified in NILs for candidate genes.

	UPSTREAM	5UTR	EXON(SYN)	EXON(NONSYN)	INTRON	3UTR	DOWNSTREAM
Glyma06g23400	36(5)*	3(0)	2(0)	-	20(7)	2(0)	31(2)
Glyma06g23410	16(2)	-	-	1(0)	0(1)	0(1)	11(2)

* value in ( ) represented number of INDELs.

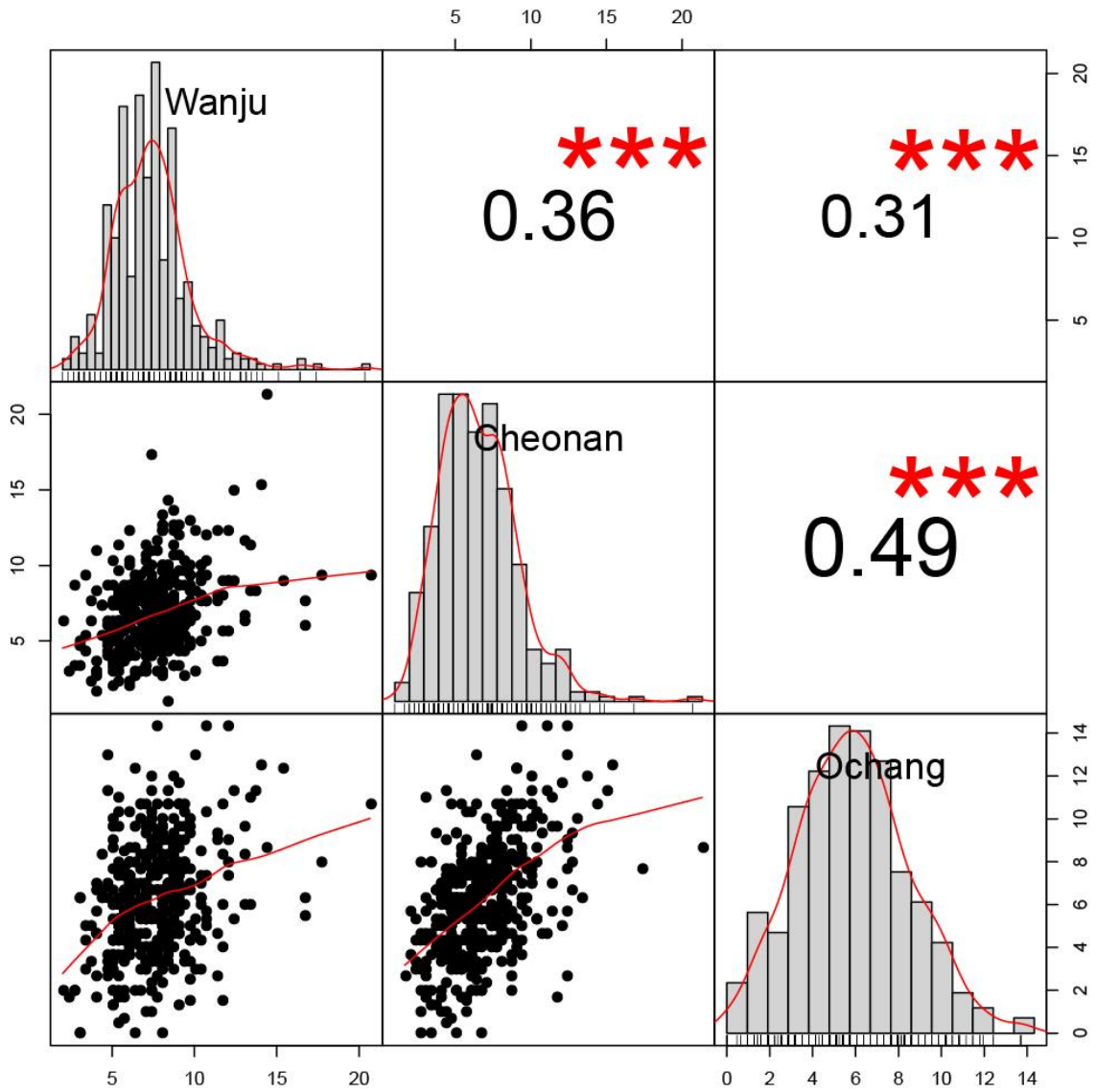
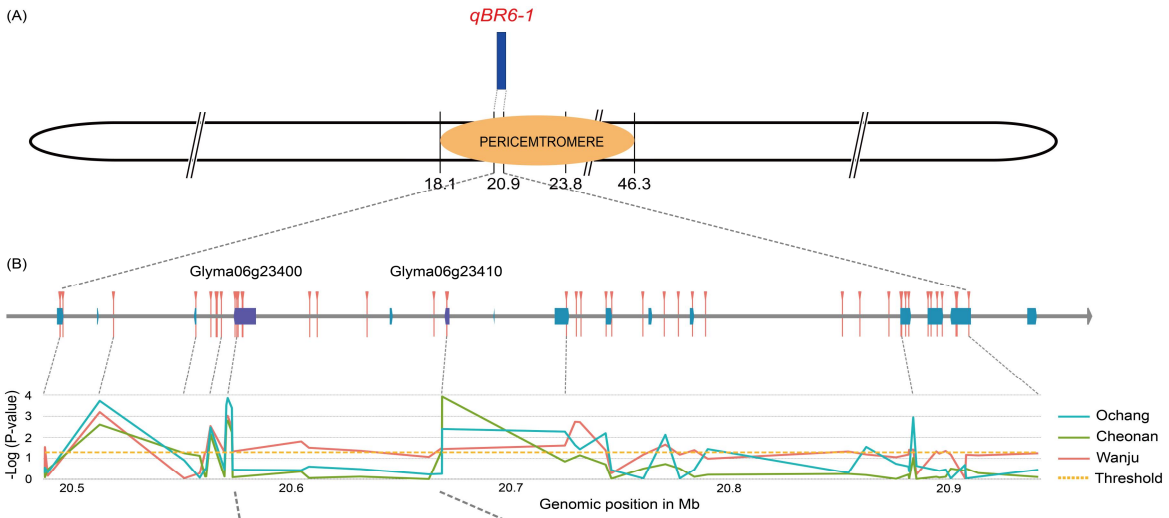


Figure 2-1. Distributions and correlations of phenotypes measured in three geographical regions. Distributions of branch numbers were expressed histogram in diagonally from left upper side to right lower side. Scatter plots explained correlation between environments were located in left lower side panels. Correlation coefficient and significant level were represented as value and asterisk marks in the right upper side panels.



(D)

Acc.	Genomic pos.	20668622	20669831	20669898	20670000	20670307	20670427	20670429	20670445	Genomic pos.	No. of branches
PI232999		C	C	A	G	C	C	T	G		1
PI153294		C	C	A	G	C	C	T	G		1
PI180529		C	C	A	G	C	C	T	G		1.5
PI180952		C	C	A	C	C	C	T	C		1.5
PI257428		C	C	A	G	C	C	T	G		1.5
PI257432		C	C	A	G	C	C	T	G		1.5
PI257433		C	C	A	G	C	C	T	G		1.5
PI257436		C	C	A	G	C	C	T	G		1.5
PI548601		C	C	A	G	C	C	T	G		1.5
PI78242		C	C	A	G	C	C	N	G		2
PI78243		C	C	A	G	C	C	T	G		2
PI80822		C	C	A	G	C	C	T	G		2
PI86741		C	C	A	G	C	C	T	G		2
PI92465		C	C	A	G	C	C	T	G		2
PI92468		C	C	A	G	C	C	T	G		2
PI92470		C	C	A	G	C	C	T	G		2
PI132205		C	C	A	G	C	C	T	G		2
PI153226		C	C	A	G	C	C	T	G		2
PI153272		C	C	A	G	C	C	T	G		2
PI153278		C	C	A	G	C	C	T	G		2
PI153279		C	C	A	G	C	C	T	G		2
PI184042		C	C	A	G	C	C	T	G		2
PI194654		C	C	A	G	C	C	T	G		2
PI232099		C	C	A	G	C	C	T	G		2
PI248396		C	C	A	G	C	C	T	G		2
PI248398		C	C	A	G	C	C	T	G		2
PI251586		C	C	A	G	C	C	T	G		2
PI548506		C	C	A	G	C	C	T	G		2
PI548573		C	C	A	G	C	C	T	G		2
NIL-less branches		C	C	A	G	C	C	T	G		Less
Jiyu69		C	C	A	G	C	C	T	G		Less
SS0404-T5-76		T	G	T	T	A	T	C	A		More
NIL-more branches		T	G	T	T	A	T	C	A		More
PI88352		T	G	T	T	A	T	C	A		2
PI131531		T	G	T	T	A	T	C	A		2
PI132217		T	G	T	T	A	T	C	A		2
PI153212		T	G	T	T	A	T	C	A		2
PI153236		T	C	A	G	C	T	T	A		2
PI184045		T	C	A	G	C	T	T	A		2
PI189881		T	C	A	G	C	T	T	A		2
PI194634		T	G	T	T	A	T	C	A		2
PI248405		T	C	A	G	C	T	T	A		2
PI248409		T	C	A	G	C	T	T	A		2
PI251585		T	C	A	G	C	T	T	A		2
PI548356		T	G	T	T	A	T	C	A		2
PI548516		T	C	A	G	C	T	T	A		2
PI548572		T	C	A	G	C	T	T	A		2
PI83881		T	G	T	T	A	T	C	A		4
PI87525		T	G	T	T	A	T	C	A		4
PI87629		T	C	A	G	C	T	T	A		4
PI103091		T	G	T	T	A	T	C	A		4
PI154191		T	G	T	T	A	T	C	A		4
PI171652		T	G	T	T	A	T	C	A		4
PI180516		T	C	A	G	C	T	T	A		4
PI196525		T	G	T	T	A	T	C	A		4
PI248408		T	C	A	G	C	T	T	A		4
PI257429		T	C	A	G	C	T	T	A		4
PI548317		T	G	T	T	A	T	C	A		4
PI548415		T	C	A	G	C	T	T	A		4
PI548427		T	C	A	G	C	T	T	A		4
PI548429		T	C	A	G	C	T	T	A		4
PI81785		T	G	T	T	A	T	C	A		4.5
PI161431		T	C	A	G	C	T	T	A		4.5

Figure 2-2. Genomic region of *qBR6-1* and its association with branching phenotypes

measured in three different geographic regions in Korea. *P-value* derived from single marker association using ANOVA were transformed into log scale and expressed as different color by geographic regions where phenotypes measured.

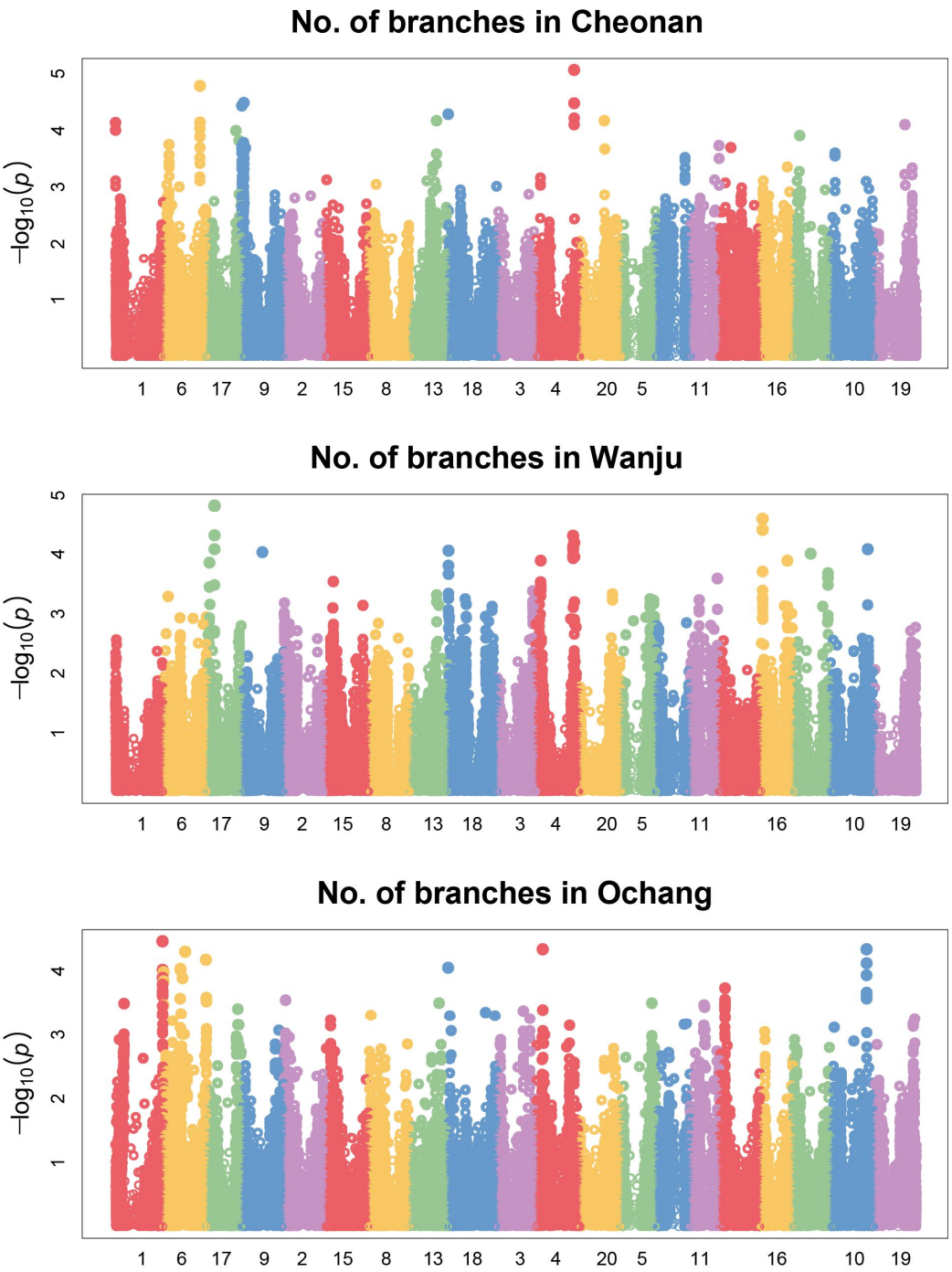


Figure 2-3. Manhattan plot for soybean branch numbers

Figure 2-4. Picture of NILs and its branch numbers. A set of NIL exhibited different number of branches were presented. Branch of each NIL was indicated by red arrow.

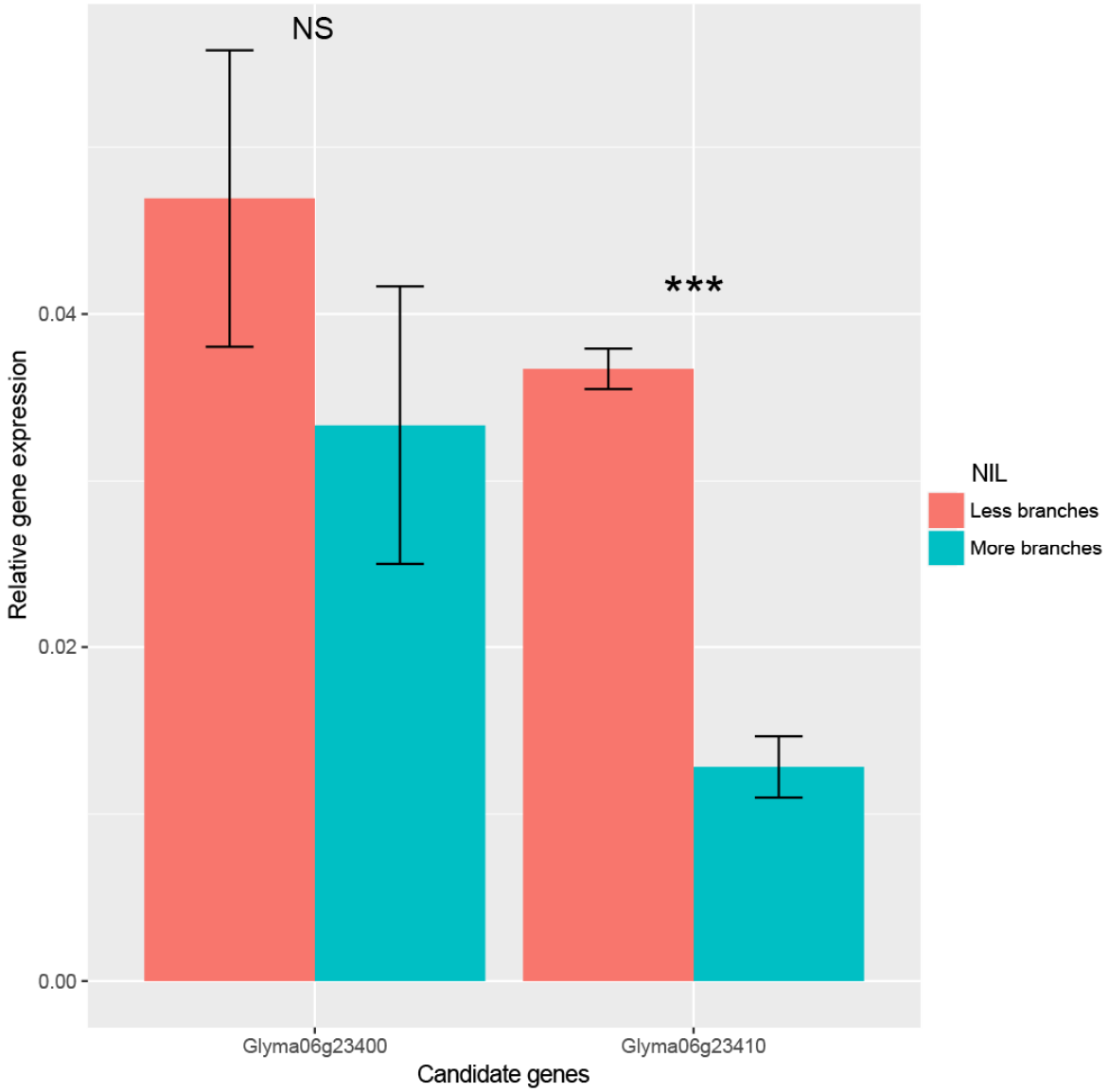


Figure 2-5. Comparison of candidate gene expression between NIL. Mean and standard error were expressed as bar and whisker. Statistical significance of relative expression was presented by asterisk mark or 'NS'.

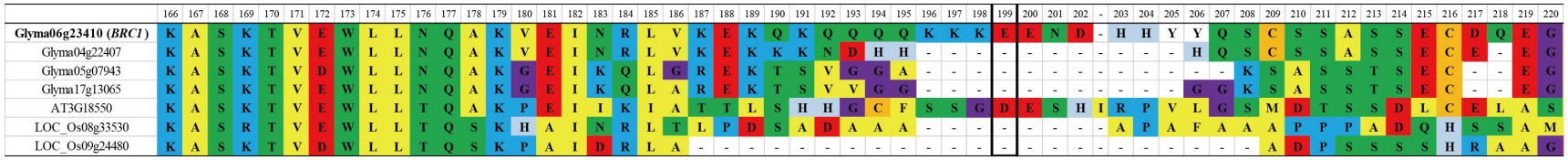


Figure 2-6. Protein sequence alignment of *BRANCHED1* (*BRC1*) gene, soybean paralogues and orthologues in *A. thaliana* and *O. sativa*. Amino acid residue changed by missense SNP at 199th of *BRC1* was highlighted by black box.

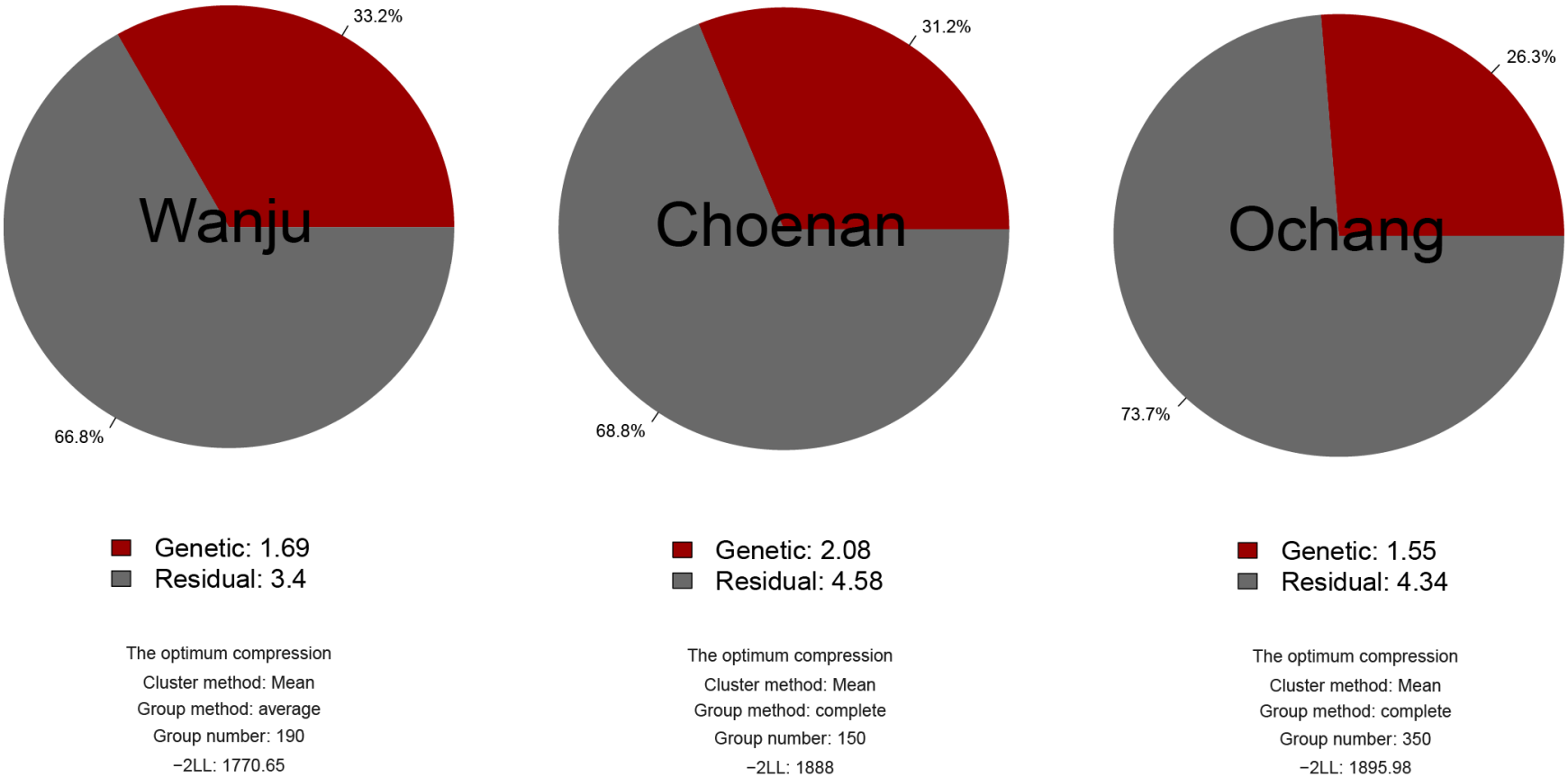


Figure 2-7. Phenotypic variations explained by genetic factors in three different geographic replications

Chapter 3.

Transcriptome profiling of soybean NILs for a major QTL of branching, *qBR6-1*

Abstract

Branching in soybean affects yield factors of pod and seed numbers per plant by increasing flowers and pods on branches. Thus far, a total of 18 QTLs associated with number of branches have been reported. Two major QTLs were overlapped in genomic regions of chromosome 6 of soybean. Among the two QTLs, recently reported major QTL, referred *qBR6-1*, have been narrowed down into a region spanning 460 kb that containing 13 genes. For that region, a set of near-isogenic lines (NIL) has been developed from a F_6 residual heterozygous line (RHL) of mapping population used in previous QTL study and showed 99.9% of background homozygosity. In this study, we profiled differences of whole transcriptome expression between the two NILs. A total of 387 statistically significant differentially expressed genes (DEGs) were identified and mapped to the pathway using Mapman software. The result showed that DEGs were mapped to pathways such as, flavonoids,

simple phenols synthesis, plant hormone, abiotic stress, drought stress signaling and development related pathways. Especially, some DEGs in development pathway were linked to the genes like *REVOLUTA* (*REV*) and *MORE AXILLARY GROWTH 4* (*MAX4*) that reported as causal genes for branch development. As conclusion, difference of branch development exhibited by a set of NIL, was attributed by orchestration of genes related to pathways including plant hormone, secondary metabolite and development pathway. This will help to understand the branch development mechanism in soybean.

Keywords transcriptome, profiling, QTL, branching, soybean

Introduction

Number of branches in soybean affects yield components of pod and seed number per plant. It has been reported that the branch development was affected by diverse environmental stimuli, such as light quality, planting density and drought, along with phytohormones, such as auxin, cytokinin (CK), gibberellin (GA) and abscisic acid (ABA) (Board 2000; Basuchaudhuri 2016; Toyota et al. 2017). It was demonstrated that drought treatment on early reproductive stage inhibited branch outgrowth (Frederick et al. 2001). This suggest drought stress responsive gene may affect genes regulating branch development. For plant hormones, auxin has been reported as negative regulator of axillary branch outgrowth (Shimizu-Sato et al. 2009). Cytokinin promoting shoot branching has been demonstrated in soybean (Leyser 2003; Shimizu-Sato et al. 2009). One of the gibberellin, GA3 promoted number of branches in soybean (Sarkar et al. 2002). Additionally, a possibility of ABA affecting branch development has been raised by the demonstration causing more branch outgrowth along with decreased ABA under the shading treatment (Zhang et al. 2011).

Due to the interactions between branch development and various environmental factors, it was obscured to identify genetic factors

affecting branch development. A little number of QTLs conferring branching has been identified in soybean (Chen et al. 2007; Li et al. 2008; Sayama et al. 2010; Yao et al. 2015; Chapter 1). Two QTLs out of nine major QTLs are overlapped in soybean chromosome 6 (Sayama et al. 2010; Chapter 1). Among the two QTLs, recently reported QTL, *qBR6-1* have been narrowed down into 450 kbp (Chapter 1). A set of near-isogenic lines (NIL) for the *qBR6-1* region have been developed from a F_6 residual heterozygous line (RHL) of mapping population used in previous QTL study (Chapter 1 and 2). Genetic background of the NILs showed 99.9% of homozygosity in whole genome resequencing of NILs (Chapter 2).

Recently, next-generation sequencing (NGS) technology has been highly applied. One of the derivate NGS technologies, RNA-seq method has been taken a spot light for detecting novel transcriptomes and quantifying the expression of genes in global view. The transcriptome analysis using RNA-seq has been most efficient method for genetic dissection and global view of comprehensive mechanism for specific trait of interest.

In this study, transcriptome analysis between the set of NILs that exhibited different branch numbers was conducted using RNA-seq method. A total 387 differentially expressed genes (DEGs) were mapped to biological pathways using Mapman software (Thimm et al.

2004). DEGs mapped to the various pathways related to branch development, such as abiotic stress signaling, hormones including auxin, cytokinin (CK), gibberellin (GA) and abscisic acid (ABA), secondary metabolites and development. Interestingly, two *NO APICAL MERISTEM* (*NAM*) domain transcription factors, four NAC domain containing protein genes, *AGRONAUT10* (*AGO10*) and nodulin MtN genes, which were related to *REVOULTA* (*REV*) and *MORE AXILLARY GROWTH 4* (*MAX4*), were identified in DEGs mapped to the development pathway.

As conclusion, it was revealed that difference of branch development in the set of NILs were orchestrated by number pathway including plant-hormone, secondary metabolite and development pathway under veiled interactions of candidate gene in *qBR6-1*. Especially, the possibility that auxin signaling related genes, *REV* and *MAX4*, played important roles in branch development in association with *qBR6-1* region. This study will provide better understanding of branch development mechanism in soybean.

Materials and methods

Plant materials and growth conditions

Near-isogenic lines (NILs) for *qBR6-1* locus were developed from a F_6 residual heterozygous line (RHL) of mapping population in previous QTL study, and selected by fore-ground selection using flaking markers of *qBR6-1* (Chapter 2). Phenotypic and genotypic variations for each NILs were tested in previous chapter, resulted that significant phenotypic difference was exhibited in filed and green house experiments and 99.9% of background homozygosity was observed in the background selection using whole genome re-sequencing method (Chapter 2).

Each NILs were grown on a glass house located in experimental farm of Seoul National University to extract RNA and validate the phenotype expression. Generally, in winter season, light intensity is lower than in regular cultivation period. To compensate the light intensity, averagely 2779.3 lux of light was additionally irradiated for 12 hours per day using metal halide lamp. Growth temperature was precisely controlled to keep the range from 19 to 35 degrees in Celsius.

RNA extraction and RNA seq analysis of NILs

Intact RNA was extracted from shoot region using Ribospin Plant (GeneAll, Cat No. 307-150, Seoul, South Korea) with a modification of adding 1~2% volume of 2-mercaptoethanol (CAS No. 60-24-2, Sigma-Aldrich, Co, St. Louis, MO) into lysis buffer. Three biological replications of RNA extract were submitted to NGS sequencing. Averagely 6 Gb of reads were generated for each replication by Illumina HiSeqX (Illumina, Inc., San Diego, CA, USA).

Transcriptome reads were, then, aligned using Tophat2 and Bowtie2 against reference genome of soybean along with the guidance of reference genomic features (Langmead and Salzberg 2012; Kim et al. 2013). All three replications of each NILs were treated as one conditions. The expression value (FPKM) were called under pooled dispersion method using cufflinks pipeline (Trapnell et al. 2010).

Validation of FPKM value was conducted by qRT-PCR using LightCycler96 (Roche, Indianapolis, IN, USA) and corresponded primers for tested genes and reference gene (*ACT11*) (Table 3-1). Each of genes were tested for three biological replications and three technical replications.

Pathway analysis of DEGs from RNA-seq result

Differentially expressed genes (DEGs) that statistically significant (*p-value* < 0.05) and the $\log_2$ fold change was over 1 or below -1 were selected. A total 387 DEGs were identified and analyzed by Mapman software (Thimm et al. 2004). To facilitate the Mapman analysis, database was constructed for 1.01 version of soybean reference genome using web version Mercator genome annotator (Lohse et al. 2014).

Results

RNA-seq mapping

Averagely 87.6 million NGS reads were generated by Illumina HiSeqX (Table 3-2). Among the reads, paired read were 75.1 million. The number of properly mapped reads to the genome were 60.9 million and averagely occupied 81.1% of total reads (Table 3-2).

Expression of candidate genes located in *qBR6-1*

Expression of 13 candidate genes that located in the major QTL, *qBR6-1*, were analyzed. Among the 13 candidate genes, three genes (Glyma06g23405, Glyma06g23502 and Glyma06g23420) were not expressed in both of NILs. Two genes, Glyma06g23474 encoding O-Glycosyl hydrolase family 17 proteins and Glyma06g23380 unknown protein coding gene, showed significantly down regulated in more branching NIL (Fig. 3-1A). Expression of *BRANCHED1* (*BRC1*) gene which was proposed as a promising candidate gene in QTL study was not significantly different, but, slightly down regulated in more branching NIL. However, validation conducted using qRT-PCR resulted that the *BRC1* gene was statistically down regulated in more branching NIL (Fig. 3-1B). Inconsistency between the RNA-seq and qRT-PCR seemed to be caused by sensitivity of RNA-seq for extremely

low or high abundant transcriptomes that raised by Shapiro et al. (Shapiro et al. 2013).

Differentially expressed gene and pathway mapping overview

A total of 4,785 genes that showed $\log_2$ fold change over 1 or below -1 were selected as differentially expressed genes. Among the 4,785 genes, a set of genes containing 387 genes which showed significantly different expression was identified and used in downstream analysis. Mapman pathway mapping results showed that all the DEGs were mapped to 21 different pathways, such as major/minor CHO metabolism, mitochondrial electron transport, lipid metabolism, cell wall associated, secondary metabolism, plant hormones, stress/signaling, cellular responses and development (Fig. 3-2 and Table 3-3).

Out of the 21 pathways mapped by DEGs, four pathways, including secondary metabolism, hormones, development and drought stress signaling, that associated with branch developments, were analyzed.

Plant hormone related DEGs

DEGs related to plant hormones, such as auxin, abscisic acid (ABA), cytokinin (CK) and gibberellin (GA) were identified (Fig. 3-3). Firstly,

auxin was known as suppressor of axillary bud outgrowth. A total three DEGs, Glyma02g03420 (*UGT74B1*) related to auxin synthesis/degradation and Glyma12g03920 and Glyma06g43411, SAUR-like auxin responsive genes which was positively regulated by auxin in *Arabidopsis* (Goda et al. 2004), were down regulated in more branching NIL. For gibberellin, it was noted that application of GA3, one of the gibberellins, enhanced in number of branches in soybean (Sarkar et al. 2002). GA20 oxidase (Glyma03g02260) associated with GA3 synthesis was up regulated in more branching NIL, otherwise, gene regulated by gibberellin, Glyma04g02660 (*GASA1*), was down regulated in more branching NIL (Fig. 3-3). Cytokinin (CK) controlling and promoting shoot branching in soybean has been demonstrated by (Leyser 2003; Shimizu-Sato et al. 2009). A gene Glyma07g33880 encode UDP-glycosyltransferase (*UGT73B4*), was down regulated in more branching NIL (Fig. 3-3). UDP-glycosyltransferase was known for deactivating enzyme of CK by its conjugation with sugar. Possibilities of abscisic acid affecting branching in soybean was demonstrated by shading treatment. Under the shading treatment, abscisic acid was decreased, while the branch numbers increased. In this study, *ABA2* gene synthesis precursor of abscisic acid (abscisic aldehyde) and its paralogue were down regulated in more branching NIL (Fig. 3-3).

DEGs associated with secondary metabolism

It has been reported that flavonoids are inhibitors of auxin transport which is known as shoot branching control mechanism (Brown et al. 2001; Buer and Muday 2004; Peer et al. 2004; Bennett et al. 2006). A total five genes out of 15 DEGs, were mapped to the flavonoids synthesis pathway (Fig. 3-4). A gene, Glyma07g33880 encoding *UGT73B4* and associated with flavonols synthesis and two genes, Glyma15g02140 (encoding dihydroflavonol 4-reductase-like1) and Glyma07g09970 (encoding *CYP71A22*) associated with dihydroflavonol pathway were down regulated in more branching NIL (Fig. 3-4). Moreover, a gene Glyma11g10380, which encodes chalcone and stilbene synthase, was also down regulated in more branching NIL (Fig. 3-4).

It has been reported that application of phenol as an antioxidant along with biotic agents, such as *B. subtillis* and *S. cerevisiae*, increased branch number and height in soybean (EL-Hai et al. 2016). A set of four DEGs including Glyma12g06480, Glyma18g38661, Glyma18g39440 and Glyma18g38710 related to phenol pathway, were up regulated in more branching NIL (Fig. 3-4). All these four gene encoded laccase enzyme (*LAC2* and *LAC17*) which affects formation of lignin by oxidation of monolignols, one of the phenols. Interestingly, a

gene Glyma07g09970 (*CYP71A22*) also related to lignin biosynthesis and dihydroflavonol synthesis.

Development associated DEGs.

A total 12 DEGs related to development were mapped to the pathway (Fig. 3-5). Among the 12 DEGs, six genes were NAC domain transcription factors. Three genes, Glyma03g35570, Glyma05g09110 and Glyma12g22880 were down regulated, otherwise, the other six gene, Glyma05g04250, Glyma05g36031 and 07g15180 were up regulated in more branching NIL (Fig. 3-5). It has been reported that the NAC domain transcription factor promoted shoot branching in rice (Mao et al. 2007). Two genes encoding nodulin MtN were mapped to the development pathway, one of them, Glyma08g01300 associated with seed development, was downregulated in more branching NIL (Fig. 3-5).

Another gene, Glyma10g33130 encoding EamA-like transporter family protein, was up regulated in more branching NIL (Fig. 3-5). Interestingly, Glyma10g33130 were connected to Glyma04g08910 encoding *MORE AXILLARY GROWTH 4* (*MAX4*) by network constructed using SoyNet.

A gene encodes *ARGONAUTE10/ZWILLE/PINHEAD* (*AGO10/ZLL/PNH*, Glyma17g12850), which was involved in the

negative regulation of auxin signaling under *REVOLUTA* (*REV*) mediation (Roodbarkelari et al. 2015), was also mapped to the pathway of development and up regulated in more branching NIL (Fig. 3-5). In addition, DEGs related to male sterility (Glyma02g26670, *MS2*), cell death (Glyma11g04360, *BAP2*) and cell differentiation (Glyma16g26080, *TET6*) were also identified and down regulated in more branching NIL (Fig. 3-5).

DEGs related to drought

Soybean branch development negatively interacted with drought stress between R1 to R5 growth stage (Frederick et al. 2001). Two DEGs mapped to drought stress pathway. One responsive gene to salt stress, Glyma17g23602, was turned off in more branching NIL, while expressed in less branching NIL (Fig. 3-5). Another gene encoding *HYP1* (Glyma18g49750) has been suggested as drought responsive gene in rice (Wang et al. 2011), was up regulated in more branching NIL (Fig. 3-5).

Discussion

Branch development affects yield components of pod and seed number per plant (Ghodrati 2013). Despite the importance of number of branches, the interactions to various environmental factors obscured identifying of genetic factor that controls branch development. So far, 18 QTLs were identified and reported in soybean (Chen et al. 2007; Li et al. 2008; Sayama et al. 2010; Yao et al. 2015; Chapter 1). Among these QTLs, recently identified major QTL, *qBR6-1*, spanned 450 kbp and contained only 13 genes (Chapter 1). In this QTL region, promising candidate gene of *BRANCHED1* (*BRC1*) which was act as a negative regulator in downstream of *MORE AXILLARY GROWTH* (*MAX*) pathway and *REVOLUTA* (*REV*) gene (Otsuga et al. 2001; Bennett et al. 2006; Aguilar-Martínez et al. 2007). For this QTL, a set of NILs was developed from a F_6 RHL in previous chapter (Chapter 2).

In this study, transcriptome analysis of a set of NILs that exhibited different number of branches and segregated into different parental genotypes in *qBR6-1* was conducted. As a result, a total of 387 DEGs were identified and mapped to the pathways, such as hormones, secondary metabolism and development pathway that related to branch development (Table 3-3 and Fig. 3-2).

Several plant hormones, such as auxin, abscisic acid (ABA), cytokinin (CK) and gibberellin (GA), affecting soybean branch

development has been reported (Basuchaudhuri 2016). In this study, DEGs for several plant hormonal pathways were identified and mapped (Fig. 3-3). Auxin, suppressor of branch outgrowth, related three DEGs were identified and down regulated in more branching NIL (Fig. 3-3). Two genes encoding SAUR-like auxin responsive genes and positively regulated by auxin in *Arabidopsis* (Goda et al. 2004), were also identified (Fig. 3-3). This suggested that the auxin level was lower in more branching NIL and IGs synthesis pathway was down regulated due to lowered level of auxin in more branching NIL. It has been noted that application of GA3, enhanced the number of branches in soybean (Sarkar et al. 2002). Two DEGs related to gibberellin pathway were also identified. GA20 oxidase (Glyma03g02260) which overexpression lead increasing of GA concentration, was up regulated in more branching NIL (Fig. 3-3), represented that GA concentration was up regulated by GA20 oxidase, and consequently, caused more branches exhibition. However, *GASA1* gene which was positively regulated by gibberellin concentration, was down regulated in more branching NIL (Fig. 3-3). It has been reported that Cytokinin promotes shoot branching in soybean (Leyser 2003; Shimizu-Sato et al. 2009). UDP-glycosyl transferase encoding gene known as deactivating enzyme of cytokinin was down regulated in more branching NIL (Fig. 3-3), represented that deactivation mechanism of cytokinin was inhibited to

exhibit more branches. Branch increasing along with abscisic acid decreasing has been observed when shading was treated in soybean (Zhang et al. 2011), represented that possibility of abscisic acid affecting branch development. DEGs coding *ABA2* and its homologs which synthesize abscisic aldehyde, precursor of abscisic acid, were down regulated in more branching NIL (Fig. 3-3). This suggested a possibility of more branch exhibition caused by lowering abscisic acid level.

Secondary metabolism of flavonoids whose product can prohibit auxin transport (Brown et al. 2001; Buer and Muday 2004; Peer et al. 2004; Bennett et al. 2006) and simple phenol that act as antioxidants and increased branch number when applied with biotic agents like *B. subtilis* or *S. cerevisiae* (EL-Hai et al. 2016) were identified as DEGs (Fig. 3-4). Five DEGs mapped to the flavonoid metabolism were identified and down regulated in more branching NIL (Fig. 3-4). The five DEGs were flavonoid synthsis gene *UGT73B4*, dihydroflavonol synthesis genes (dihydroflavonol 4-reductase and *CYP71A22*) and chalcone/stilbene synthase gene (Fig. 3-4). This indicated that lowered flavonoids synthesis may accelerate branch development. A set of four DEGs related to phenol metabolism were up regulated in more branching NIL (Fig. 3-4). All these four gene encoded laccase enzyme

(*LAC2* and *LAC17*) which affects synthesis of lignin, which is one of the phenols, by oxidation of monolignols.

A total 12 DEGs related to development were also identified (Fig. 3-5). Six DEGs were genes encoding NAC domain containing transcription factors, which has been reported that the gene promote shoot branching in rice (Mao et al. 2007). However, half of the genes were down regulated, otherwise, the other half were up regulated in more branching NIL (Fig. 3-5). Two genes encoding nodulin MtN were also identified (Fig. 3-5). One of the genes, interestingly, Glyma10g33130, was upregulated in more branching NIL (Fig. 3-5) and connected to *MAX4* gene in putative gene network constructed by SoyNet (Kim et al. 2017). It has been reported that *MAX4* gene promoted axillary branch outgrowth under auxin signaling in *Arabidopsis* (Bennett et al. 2006). Another gene encodes *ARGONAUTE10/ZWILLE/PINHEAD* (*AGO10/ZLL/PNH*, Glyma17g12850) involved in the negative regulation of auxin signaling under *REVOLUTA* (*REV*) mediation (Roodbarkelari et al. 2015) was also identified and up regulated in more branching NIL (Fig. 3-5). These suggested that difference of branch development between the NILs was attributed to auxin signaling mediated by these genes.

Soybean branch development negatively interacted with drought stress between R1 to R5 growth stage (Frederick et al. 2001),

represented that drought stress responsive genes may affecting genes regulating branch development. In this study, two drought or salt stress responsive genes were identified in DEGs, but, the two genes showed different expression pattern, Glyma17g23602 was down regulated and the other *HYP1* which has been suggested as drought responsive gene in rice (Wang et al. 2011), was up regulated in more branching NIL (Fig. 3-5). This result suggested possibility of relevance between the two drought stress responsive genes and branch development.

In conclusion, it was shown that the difference of branch development between two NILs was orchestrated by various biological process and molecular pathway such as plant hormone, secondary metabolite and development pathway under genotype difference of *qBR6-1*. All the results were integrated and presented as Fig. 3-6. However, the interlude between genes which can provide profound understanding are still unknown. Therefore, transcriptome analysis using the NIL set under more sophisticated experimental design seemed to be needed to get better understanding of branch development in soybean. This study will provide starting point in revealing the unknown properties of soybean branch development.

References

Aguilar-Martínez JA, Poza-Carrión C, Cubas P (2007) *Arabidopsis BRANCHED1* Acts as an Integrator of Branching Signals within Axillary Buds. Plant Cell 19:458–472. doi: 10.1105/tpc.106.048934

Basuchaudhuri P (2016) INFLUENCES OF PLANT GROWTH REGULATORS ON YIELD OF SOYBEAN. Indian J Plant Sci 5:25–38

Bennett T, Sieberer T, Willett B, et al (2006) The *Arabidopsis MAX* Pathway Controls Shoot Branching by Regulating Auxin Transport. Curr Biol 16:553–563. doi: 10.1016/j.cub.2006.01.058

Board J (2000) Light Interception Efficiency and Light Quality Affect Yield Compensation of Soybean at Low Plant Populations. Crop Sci 40:1285–1294. doi: 10.2135/cropsci2000.4051285x

Brown DE, Rashotte AM, Murphy AS, et al (2001) Flavonoids Act as Negative Regulators of Auxin Transport in Vivo in Arabidopsis. Plant Physiol 126:524–535. doi: 10.1104/pp.126.2.524

Buer CS, Muday GK (2004) The *transparent testa4* Mutation Prevents Flavonoid Synthesis and Alters Auxin Transport and the Response of Arabidopsis Roots to Gravity and Light. Plant Cell 16:1191–1205. doi: 10.1105/tpc.020313

Chen Q, Zhang Z, Liu C, et al (2007) QTL Analysis of Major Agronomic Traits in Soybean. Agric Sci China 6:399–405. doi: 10.1016/S1671-2927(07)60062-5

EL-Hai KMA, Elhersh MS, Mahmoud MK (2016) Incidence of Soybean Root and Stalk Rot Diseases as a Result of Antioxidant and Biotic Agents. Biotechnology(Faisalabad) 15:52–64. doi: 10.3923/biotech.2016.52.64

Frederick JR, Camp CR, Bauer PJ (2001) Drought-Stress Effects on Branch and Mainstem Seed Yield and Yield Components of Determinate Soybean. Crop Sci 41:759–763. doi: 10.2135/cropsci2001.413759x

Ghodrati G (2013) Study of genetic variation and broad sense heritability for some qualitative and quantitative traits in soybean (*Glycine max* L.) genotypes. Curr Opin Agric 2:

Goda H, Sawa S, Asami T, et al (2004) Comprehensive Comparison of Auxin-Regulated and Brassinosteroid-Regulated Genes in Arabidopsis. Plant Physiol 134:1555–1573. doi: 10.1104/pp.103.034736

Kim D, Pertea G, Trapnell C, et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36. doi: 10.1186/gb-2013-14-4-r36

Kim E, Hwang S, Lee I (2017) SoyNet: a database of co-functional networks for soybean *Glycine max*. Nucleic Acids Res 45:D1082–D1089. doi: 10.1093/nar/gkw704

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi: 10.1038/nmeth.1923

Leyser O (2003) Regulation of shoot branching by auxin. Trends Plant Sci 8:541–545. doi: 10.1016/j.tplants.2003.09.008

Li W, Zheng D, Van K, Lee S (2008) QTL Mapping for Major Agronomic Traits across Two Years in Soybean (*Glycine max* L. Merr.)

Lohse M, Nagel A, Herter T, et al (2014) Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ 37:1250–1258. doi: 10.1111/pce.12231

Mao C, Ding W, Wu Y, et al (2007) Overexpression of a NAC-domain

protein promotes shoot branching in rice. New Phytol 176:288–298. doi: 10.1111/j.1469-8137.2007.02177.x

Otsuga D, DeGuzman B, Prigge MJ, et al (2001) *REVOLUTA* regulates meristem initiation at lateral positions. Plant J Cell Mol Biol 25:223–236

Peer WA, Bandyopadhyay A, Blakeslee JJ, et al (2004) Variation in Expression and Protein Localization of the PIN Family of Auxin Efflux Facilitator Proteins in Flavonoid Mutants with Altered Auxin Transport in *Arabidopsis thaliana*. Plant Cell 16:1898–1911. doi: 10.1105/tpc.021501

Roodbarkelari F, Du F, Truernit E, Laux T (2015) *ZLL/AGO10* maintains shoot meristem stem cells during *Arabidopsis* embryogenesis by down-regulating ARF2-mediated auxin response. BMC Biol 13:. doi: 10.1186/s12915-015-0180-y

Sarkar PK, Haque MS, Karim MA (2002) Effects of GA3 and IAA and their Frequency of Application on Morphology, Yield Contributing Characters and Yield of Soybean. J Agron 1:119–122. doi: 10.3923/ja.2002.119.122

Sayama T, Hwang T-Y, Yamazaki H, et al (2010) Mapping and comparison of quantitative trait loci for soybean branching phenotype in two locations. Breed Sci 60:380–389. doi: 10.1270/jsbbs.60.380

Shapiro E, Biezuner T, Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet 14:618–630. doi: 10.1038/nrg3542

Shimizu-Sato S, Tanaka M, Mori H (2009) Auxin-cytokinin interactions in the control of shoot branching. Plant Mol Biol 69:429–435. doi: 10.1007/s11103-008-9416-3

Thimm O, Bläsing O, Gibon Y, et al (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic

pathways and other biological processes. Plant J Cell Mol Biol 37:914–939

Toyota M, Maitree L, Chomsang K (2017) Changes in radiation interception and R:FR over time and with canopy depth of two soybean cultivars with different branching characteristics. Plant Prod Sci 20:205–214. doi: 10.1080/1343943X.2017.1294464

Trapnell C, Williams BA, Pertea G, et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515. doi: 10.1038/nbt.1621

Wang D, Pan Y, Zhao X, et al (2011) Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. BMC Genomics 12:149. doi: 10.1186/1471-2164-12-149

Yao D, Liu ZZ, Zhang J, Liu SY, Qu J, Guan SY, Pan LD, Wang D, Liu JW, Wang PW (2015) Analysis of quantitative trait loci for main plant traits in soybean. Genet Mol Res 14:6101-6109. doi:10.4238/2015.June.8.8

Zhang J, Smith DL, Liu W, et al (2011) Effects of shade and drought stress on soybean hormones and yield of main-stem and branch. Afr J Biotechnol 10:14392–14398. doi: 10.5897/AJB11.2143

Table 3-1. Primer used in this study

Forward primer	Reverse primer	Tm (°C)	Target Genes
GTGAGGGAGCAGTGGAGGAT	TAAATGACGCCCGAGAAATC	59~60	Glyma06g23380
CACCAAGTGGCATTCCTTCT	CCCATCAGAGCCGTCATAAT	59~60	Glyma06g23474
CAGGCTGGTGGAATGTTTTT	TTTGAATGCGCAGTTCTGTC	59~60	Glyma06g23502
AGGTTACCCTTTTCCGCTGT	CAGGTTGGATCCCTCTTCAA	59~60	Glyma06g23530
ACGAGGTTGGTGTTTGGAAG	AAGCGTTGCGAAGAAGAGAG	59~60	Glyma06g23580
AATCAGTGCATTTGACCCTCTT	AGCTAGCACTCCACGATTTCTC	57	Glyma06g23410
CAAGTTGCACACGCACTCTT	TAATCGCAATGCATCCAGAG	57~59	Glyma06g23400
CGGTGGTTCTATCTTGGCATC	GTCTTTCGCTTCAATAACCCTA	57~60	*ACT11*

Table 3-2. Summary statistics for mapping of transcriptome reads

	Less branching NIL (rep.1)	Less branching NIL (rep.2)	Less branching NIL (rep.3)	More branching NIL (rep.1)	More branching NIL (rep.2)	More branching NIL (rep.3)	Average
No. of total reads	88,973,054	73,909,133	86,116,162	90,326,019	112,178,281	74,157,501	87,610,025
No. of mapped reads	88,973,054	73,909,133	86,116,162	90,326,019	112,178,281	74,157,501	87,610,025
No. of paired reads in sequencing	74,029,600	63,985,300	76,443,435	81,106,643	92,235,571	63,176,585	75,162,856
No. of properly paired	58,048,236	54,564,898	62,962,568	65,970,120	73,707,242	50,504,616	60,959,613
No. of singletons	2,625,434	1,947,864	2,294,909	2,901,649	3,093,765	2,817,589	2,613,535
Ratio of properly paired	78.4	85.3	82.4	81.3	79.9	79.9	81.1

Table 3-3. Summary table for mapped pathway and corresponded differentially expressed genes

Code	Pathway	Description of pathway	Mapped DEGs
2	Major CHO met.	degradation.sucrose.invertases.cell wall	Glyma08g20480
3	Minor CHO met.	others	Glyma08g20480
		galactose.alpha-galactosidases	Glyma20g35850
9	Mitochondrial electron transport / ATP synthesis	uncoupling protein	Glyma07g17380
10	Cell wall	cell wall proteins.LRR	Glyma17g35760
		modification	Glyma13g37395
11	Lipid met.	FA synthesis and FA elongation.acyl coa ligase	Glyma11g08890
		Phospholipid synthesis.phosphatidate cytidylyltransferase	Glyma02g46441
		lipid transfer proteins etc	Glyma03g04960
		lipid degradation.lipases.triacylglycerol lipase	Glyma01g01580
		lipid degradation.lysophospholipases.glycerophosphodiester phosphodiesterase	Glyma02g31764, Glyma07g10641, Glyma07g10561
13	Amino acid met.	synthesis.aspartate family.asparagine.asparagine synthetase	Glyma11g27480
15	Metal handling	metal handling	Glyma03g39600
		binding, chelation and storage	Glyma03g39600
16	Secondary met.	simple phenols	Glyma12g06480, Glyma18g38661, Glyma18g39440, Glyma18g38710
		phenylpropanoids	Glyma18g50290
		phenylpropanoids.lignin biosynthesis.F5H	Glyma07g09970
		sulfur-containing.glucosinolates.synthesis.shared.UDP-glycosyltransferase	Glyma02g03420

		sulfur-containing.glucosinolates.regulation.aliphatic	Glyma06g45520
		sulfur-containing.glucosinolates.regulation.indole	Glyma06g45520
		flavonoids.anthocyanins.anthocyanin 5-aromatic acyltransferase	Glyma08g27130
		flavonoids.chalcones	Glyma11g10380
		flavonoids.dihydroflavonols.dihydroflavonol 4-reductase	Glyma15g02140
		flavonoids.dihydroflavonols.flavonoid 3"-monooxygenase	Glyma07g09970
		flavonoids.flavonols	Glyma07g33880
		flavonoids.flavonols.flavonol 3-O-glycosyltransferase	Glyma07g33880
17	Hormone met.	abscisic acid.synthesis-degradation.synthesis.short chain alcohol dehydrogenmase (ABA2)	Glyma11g21180, Glyma03g26590
		auxin.synthesis-degradation	Glyma02g03420
		auxin.induced-regulated-responsive-activated	Glyma12g03920, Glyma06g43411
		cytokinin.synthesis-degradation	Glyma07g33880
		ethylene.synthesis-degradation.1-aminocyclopropane-1-carboxylate oxidase	Glyma14g05360, Glyma14g05355
		ethylene.signal transduction	Glyma20g34550
		ethylene.induced-regulated-responsive-activated	Glyma20g34550
		gibberelin.synthesis-degradation.GA20 oxidase	Glyma03g02260
		gibberelin.induced-regulated-responsive-activated	Glyma04g02660
20	Stress	stress	Glyma18g47250
		biotic	Glyma02g05320, Glyma13g01451, Glyma01g36480, Glyma01g37040, Glyma15g06790, Glyma15g06780, Glyma03g22050, Glyma07g33880, Glyma07g17370, Glyma07g17350, Glyma07g17291
		biotic.signalling	Glyma04g39270
		biotic.signalling.MLO-like	Glyma06g44023
		biotic.PR-proteins	Glyma12g34690, Glyma06g41430, Glyma15g06790, Glyma15g06780, Glyma03g22050, Glyma03g04260, Glyma16g28970, Glyma18g46101
		biotic.PR-proteins.proteinase inhibitors.trypsin inhibitor	Glyma08g25520, Glyma09g29360
		abiotic	Glyma05g38130, Glyma01g42661
		abiotic.heat	Glyma14g04970, Glyma02g01730, Glyma08g24710
		abiotic.drought/salt	Glyma17g23602, Glyma18g49750
		abiotic.unspecified	Glyma02g42990

21	Redox	ascorbate and glutathione.ascorbate	Glyma14g04530
23	Nucleotide met.	degradation	Glyma16g22830
		degradation.pyrimidine.uridine nucleosidase	Glyma20g00200
		degradation.pyrimidine.dihydropyrimidinase	Glyma08g18920
26	Misc.	cytochrome P450	Glyma08g48030, Glyma07g09970
		UDP glucosyl and glucoronyl transferases	Glyma07g33880
		gluco-, galacto- and mannosidases	Glyma13g17600, Glyma15g42570
		beta 1,3 glucan hydrolases	Glyma13g17600, Glyma15g23435, Glyma20g06250
		beta 1,3 glucan hydrolases.glucan endo-1,3-beta-glucosidase	Glyma06g23474
		O-methyl transferases	Glyma17g15031, Glyma18g50290
		oxidases - copper, flavone etc	Glyma03g24020
		nitrilases, nitrile lyases, berberine bridge enzymes, reticuline oxidases, troponine reductases	Glyma05g25460
		glutathione S transferases	Glyma02g45336, Glyma08g12510, Glyma08g18690, Glyma10g33650, Glyma07g16810
		alcohol dehydrogenases.cinnamyl alcohol dehydrogenase	Glyma12g02250
		peroxidases	Glyma06g15030
		acid and other phosphatases	Glyma13g05731, Glyma16g34720
		plastocyanin-like	Glyma12g34100, Glyma13g35100
		protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	Glyma08g43605, Glyma08g17190, Glyma13g11090, Glyma01g44540
		short chain dehydrogenase/reductase (SDR)	Glyma03g39881, Glyma03g38150, Glyma03g26590
		GCN5-related N-acetyltransferase	Glyma19g43040
27	RNA	transcription	Glyma02g31295
		regulation of transcription.AP2/EREBP, APETALA2/Ethylene-responsive element binding protein family	Glyma17g37350, Glyma13g21570, Glyma19g44580, Glyma15g19910, Glyma03g41910
		regulation of transcription.bHLH,Basic Helix-Loop-Helix family	Glyma17g06610
		regulation of transcription.C2C2(Zn) GATA transcription factor family	Glyma09g07090
		regulation of transcription.C2H2 zinc finger family	Glyma02g16280

		regulation of transcription.G2-like transcription factor family, GARP	Glyma02g45231, Glyma02g30714, Glyma13g39290, Glyma09g17452
		regulation of transcription.HB,Homeobox transcription factor family	Glyma14g14060, Glyma18g14921
		regulation of transcription.MYB domain transcription factor family	Glyma02g45231, Glyma02g01300, Glyma02g30714, Glyma06g45520, Glyma08g04670, Glyma13g39290, Glyma19g05080, Glyma09g17452, Glyma16g34490
		regulation of transcription.NAC domain transcription factor family	Glyma12g22880, Glyma05g09110, Glyma03g35570
		regulation of transcription.WRKY domain transcription factor family	Glyma14g11920, Glyma14g11960, Glyma06g15220, Glyma04g05700, Glyma17g33920, Glyma03g00460, Glyma16g34590
		regulation of transcription.Global transcription factor group	Glyma06g46930
		regulation of transcription.PHD finger transcription factor	Glyma02g11970
		regulation of transcription.putative transcription regulator	Glyma14g35980, Glyma02g45491, Glyma08g16346, Glyma07g11560
		regulation of transcription.unclassified	Glyma02g16280, Glyma11g01510, Glyma19g29820, Glyma18g38410
28	DNA	synthesis/chromatin structure	Glyma14g17800, Glyma06g46930, Glyma17g29140, Glyma07g38180
		synthesis/chromatin structure.retrotransposon/transposase.hat-like transposase	Glyma06g35964, Glyma06g35972
		unspecified	Glyma18g38410
29	Protein	aa activation.glycine-tRNA ligase	Glyma18g07790
		synthesis.ribosome biogenesis	Glyma16g00630
		synthesis.ribosome biogenesis.Pre-rRNA processing and modifications.snoRNPs	Glyma07g11560
		synthesis.initiation	Glyma17g12850
		postranslational modification	Glyma14g17675, Glyma06g23530, Glyma07g11201
		postranslational modification.kinase	Glyma04g35410
		postranslational modification.kinase.receptor like cytoplasmatic kinase VII	Glyma04g01890
		degradation.subtilases	Glyma14g06981
		degradation.ubiquitin.E3.RING	Glyma17g03155, Glyma05g30920, Glyma09g26100, Glyma09g32910, Glyma18g01720
		degradation.ubiquitin.E3.SCF.FBOX	Glyma14g17675, Glyma15g07550

		degradation.cysteine protease	Glyma06g42520, Glyma07g39590
		degradation.aspartate protease	Glyma11g01510
		degradation.AAA type	Glyma18g48910
		assembly and cofactor ligation	Glyma13g08940, Glyma16g07081
30	Signalling	in sugar and nutrient physiology	Glyma08g18980
		receptor kinases.leucine rich repeat XI	Glyma20g37010
		receptor kinases.leucine rich repeat XII	Glyma18g48600
		receptor kinases.thaumatin like	Glyma07g10508, Glyma07g10641, Glyma07g10561
		receptor kinases.Catharanthus roseus-like RLK1	Glyma02g31764, Glyma07g10641, Glyma07g10561
		receptor kinases.DUF 26	Glyma18g47250
		receptor kinases.wheat LRK10 like	Glyma02g31764, Glyma07g10641, Glyma07g10561
		receptor kinases.S-locus glycoprotein like	Glyma06g40050
		receptor kinases.wall associated kinase	Glyma09g03190, Glyma09g01750
		receptor kinases.leucine rich repeat VIII.VIII-2	Glyma12g36170
		receptor kinases.misc	Glyma06g45520, Glyma07g17370, Glyma07g17350, Glyma07g17291
		calcium	Glyma05g34640, Glyma09g31450, Glyma09g37150, Glyma10g28990
		phosphinositides	Glyma01g37136
		light	Glyma07g11790
31	Cell	organisation	Glyma06g36910, Glyma06g37040, Glyma15g17266, Glyma15g09316, Glyma03g88, Glyma18g48310
		cycle	Glyma12g32200, Glyma06g35940
		cycle.peptidylprolyl isomerase	Glyma06g00740
		vesicle transport	Glyma02g46110, Glyma17g26440
33	Development	unspecified	Glyma02g26670, Glyma12g22880, Glyma11g04360, Glyma08g01300, Glyma17g12850, Glyma05g04250, Glyma05g09110, Glyma05g36031, Glyma03g35570, Glyma10g33130, Glyma07g15180, Glyma16g26080
34	Transport	sugars	Glyma03g30550
		amino acids	Glyma02g10870, Glyma06g02210, Glyma05g32810
		metabolite transporters at the mitochondrial membrane	Glyma10g37370
		metal	Glyma07g17060
		peptides and oligopeptides	Glyma01g27490
		ABC transporters and multidrug resistance systems	Glyma19g01970
		Major Intrinsic Proteins.NIP	Glyma02g15870
35	Not assigned	no ontology	Glyma06g44320, Glyma08g23272, Glyma08g12910, Glyma17g36300, Glyma05g31300, Glyma15g03761, Glyma03g01790, Glyma07g33441
		no ontology.C2 domain-containing protein	Glyma07g36065

no ontology.formin homology 2 domain-containing protein	Glyma14g11883
no ontology.hydroxyproline rich proteins	Glyma05g31300
no ontology.pentatricopeptide (PPR) repeat-containing protein	Glyma07g11930
unknown	Glyma14g10030, Glyma14g05621, Glyma14g05780, Glyma14g36470, Glyma14g28090, Glyma14g39201, Glyma14g37710, Glyma02g17390, Glyma02g01250, Glyma02g38041, Glyma02g46311, Glyma02g38152, Glyma02g45160, Glyma02g01240, Glyma02g38620, Glyma02g41030, Glyma02g33830, Glyma12g34900, Glyma12g13680, Glyma12g36341, Glyma12g31280, Glyma12g06101, Glyma11g37695, Glyma11g02523, Glyma06g08150, Glyma06g08490, Glyma06g10951, Glyma06g24300, Glyma06g16040, Glyma06g21315, Glyma06g14571, Glyma06g34930, Glyma06g28100, Glyma06g46361, Glyma04g16880, Glyma04g12645, Glyma04g08020, Glyma04g09750, Glyma04g12156, Glyma04g37281, Glyma04g42130, Glyma08g13870, Glyma08g00260, Glyma08g46140, Glyma08g10435, Glyma08g38820, Glyma08g20970, Glyma0017s50, Glyma17g02410, Glyma17g36560, Glyma17g14201, Glyma17g03330, Glyma13g27510, Glyma13g26030, Glyma13g11860, Glyma13g02231, Glyma13g12210, Glyma13g07151, Glyma13g31931, Glyma13g32060, Glyma13g37320, Glyma13g31046, Glyma13g25981, Glyma05g35650, Glyma05g26131, Glyma05g29120, Glyma05g36956, Glyma01g34620, Glyma01g34400, Glyma01g35096, Glyma01g37690, Glyma01g17590, Glyma01g03281, Glyma01g02230, Glyma19g23453, Glyma19g33670, Glyma19g07274, Glyma19g34605, Glyma19g00700, Glyma15g08380, Glyma15g16990, Glyma15g13800, Glyma15g42500, Glyma15g18486, Glyma15g15171, Glyma15g41993, Glyma15g17131, Glyma15g37355, Glyma15g08310, Glyma03g01811, Glyma03g01800, Glyma03g30840, Glyma03g39220, Glyma03g01840, Glyma03g02440, Glyma03g01835, Glyma03g25651, Glyma03g01830, Glyma20g20355, Glyma20g26332, Glyma20g05630, Glyma20g05250, Glyma20g34771, Glyma20g00420, Glyma20g39191, Glyma20g11990, Glyma20g25080, Glyma09g01520, Glyma09g19800, Glyma09g04510, Glyma09g33744, Glyma09g23640, Glyma09g31740, Glyma09g32261, Glyma09g12570, Glyma10g27756, Glyma10g01300, Glyma10g44360, Glyma10g28150, Glyma10g09971, Glyma10g26395, Glyma10g32930, Glyma10g44370, Glyma10g33050, Glyma10g11471, Glyma10g29595, Glyma10g26411, Glyma10g23776, Glyma07g08311, Glyma07g23480, Glyma07g36210, Glyma07g36805, Glyma07g06930, Glyma07g26481, Glyma07g10502, Glyma07g14065, Glyma07g16760, Glyma07g04810, Glyma07g37240, Glyma07g11660, Glyma07g39540, Glyma07g37115, Glyma07g11075, Glyma07g11160, Glyma16g07750, Glyma16g10740, Glyma16g25270, Glyma18g46550, Glyma18g00886, Glyma18g46120, Glyma18g20146, Glyma18g45930, Glyma18g46091, Glyma18g04697, Glyma18g46111

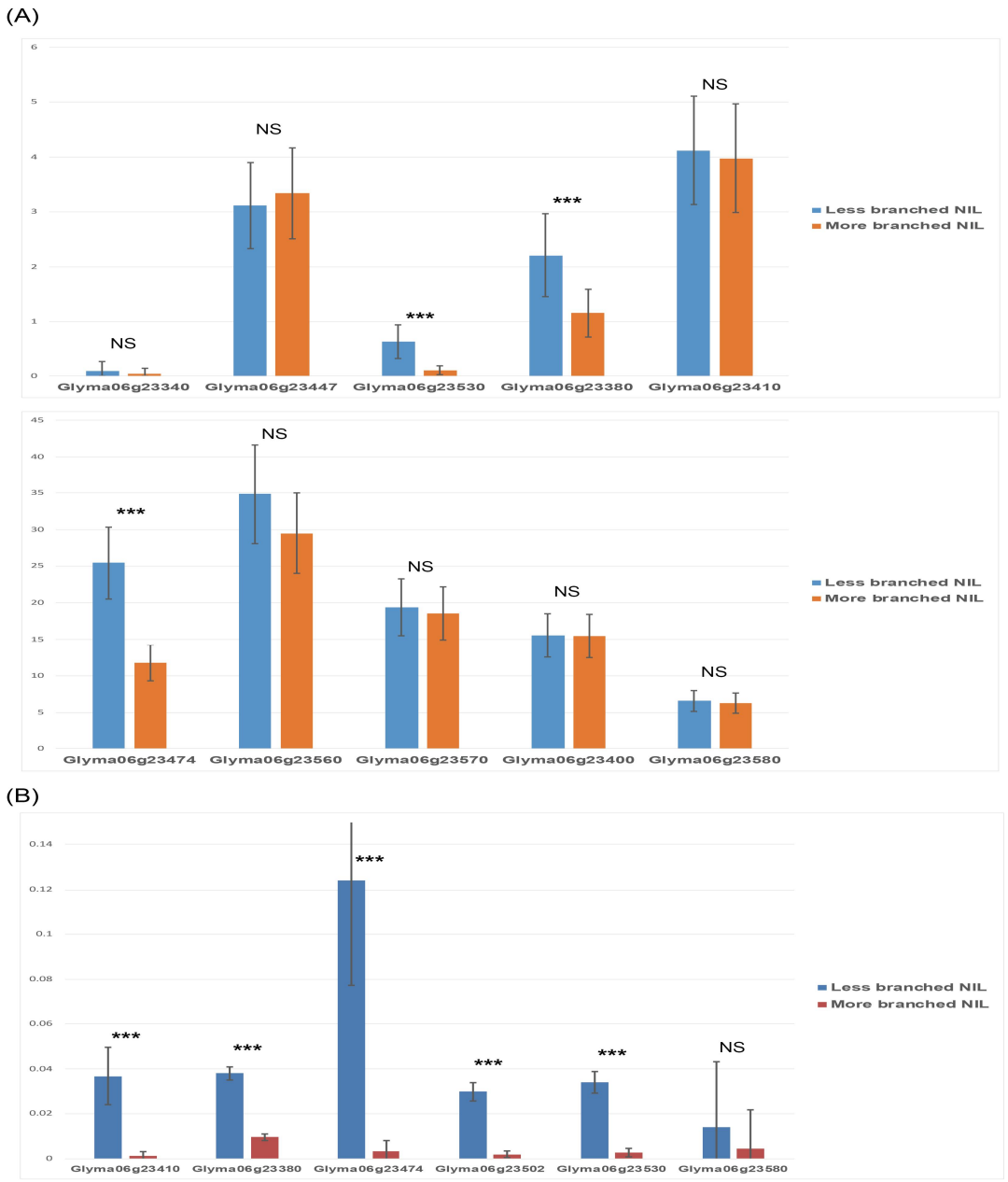


Figure 3-1. Relative expression of genes located in QTL, *qBR6-1*. Relative transcriptome expression of genes located in *qBR6-1*, which measured by RNA-seq, were presented as bar plot with error bar (A). Relative expression of selected gene measured by qRT-PCR were presented in (B)

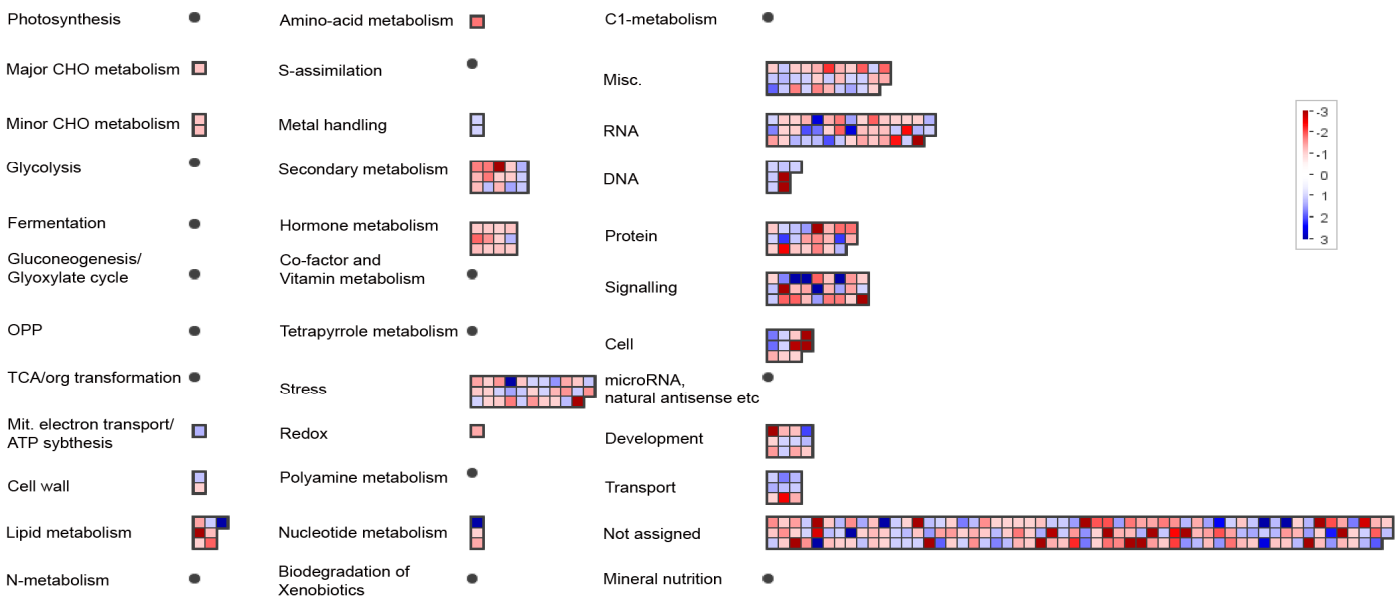


Figure 3-2. Overview of pathway mapping using differentially expressed genes between more and less branching NILs. Each box represented gene mapped to each pathway. The color of box represented log2 fold change. Negative values indicated by red color represented that the gene was down regulated in more branching NIL, otherwise, positive value indicated by blue color represented that it was up regulated in more branching NIL.

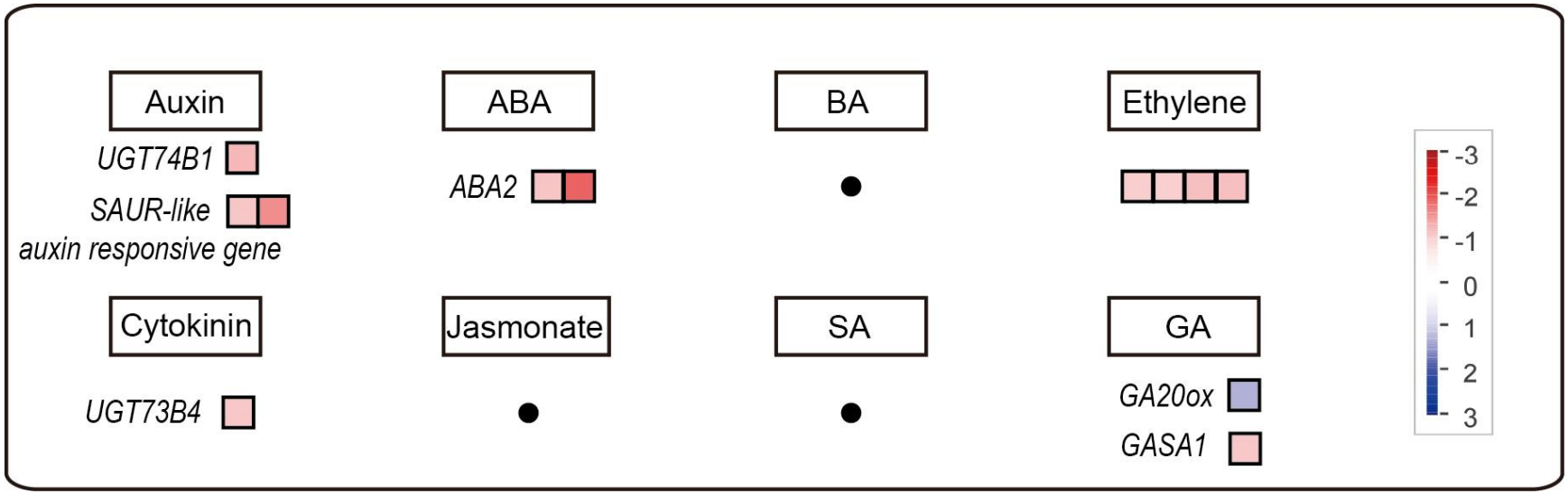


Figure 3-3. Expression of DEGs mapped to the pathways related to plant-hormones. Each box represented gene mapped to each pathway. The color of box represented log2 fold change. Negative values indicated by red color represented that the gene was down regulated in more branching NIL, otherwise, positive value indicated by blue color represented that it was up regulated in more branching NIL..

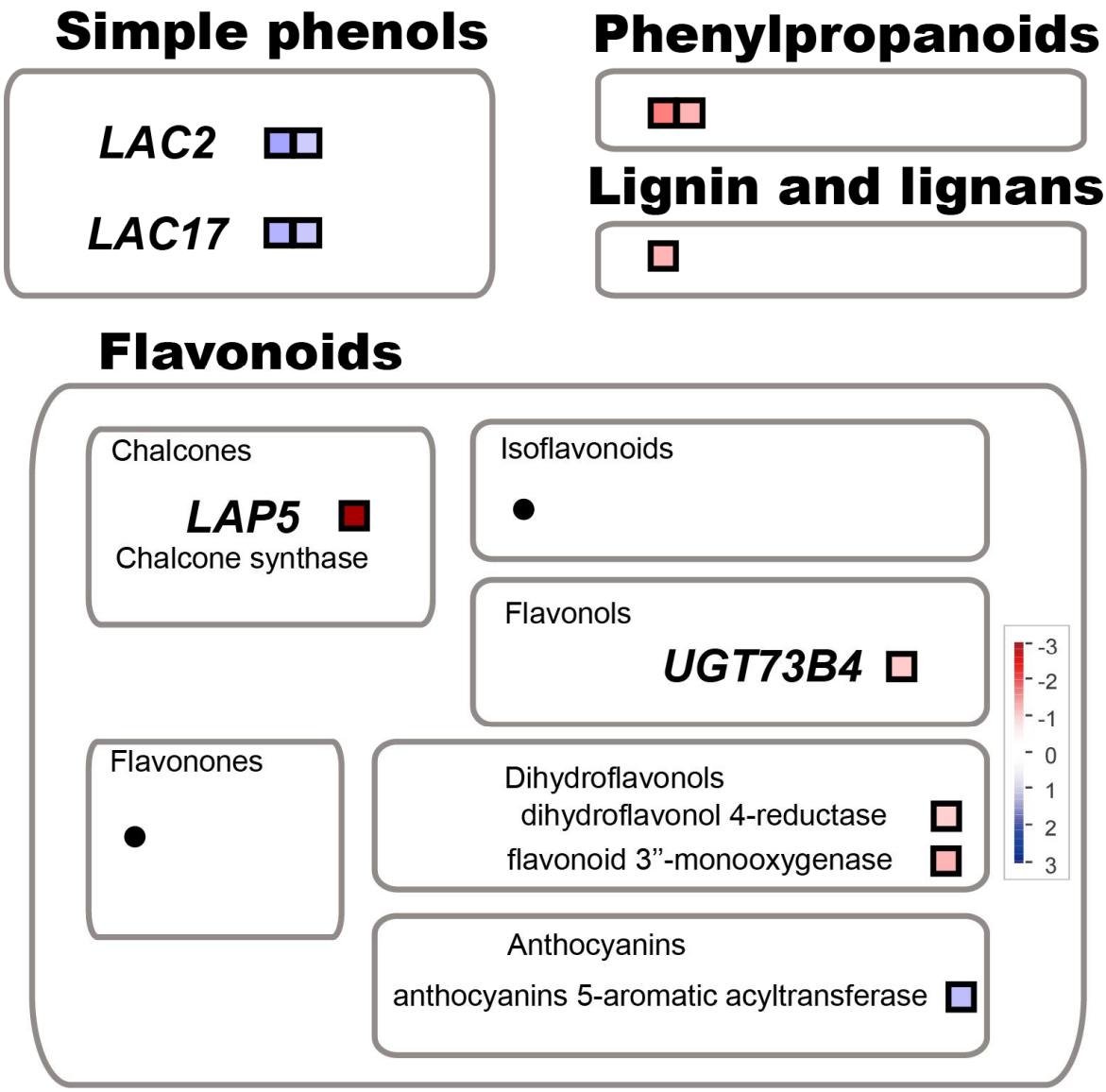


Figure 3-4. Expression of DEGs related to secondary metabolisms. Each box represented gene mapped to each pathway. The color of box represented log2 fold change. Negative values indicated by red color represented that the gene was down regulated in more branching NIL, otherwise, positive value indicated by blue color represented that it was up regulated in more branching NIL.

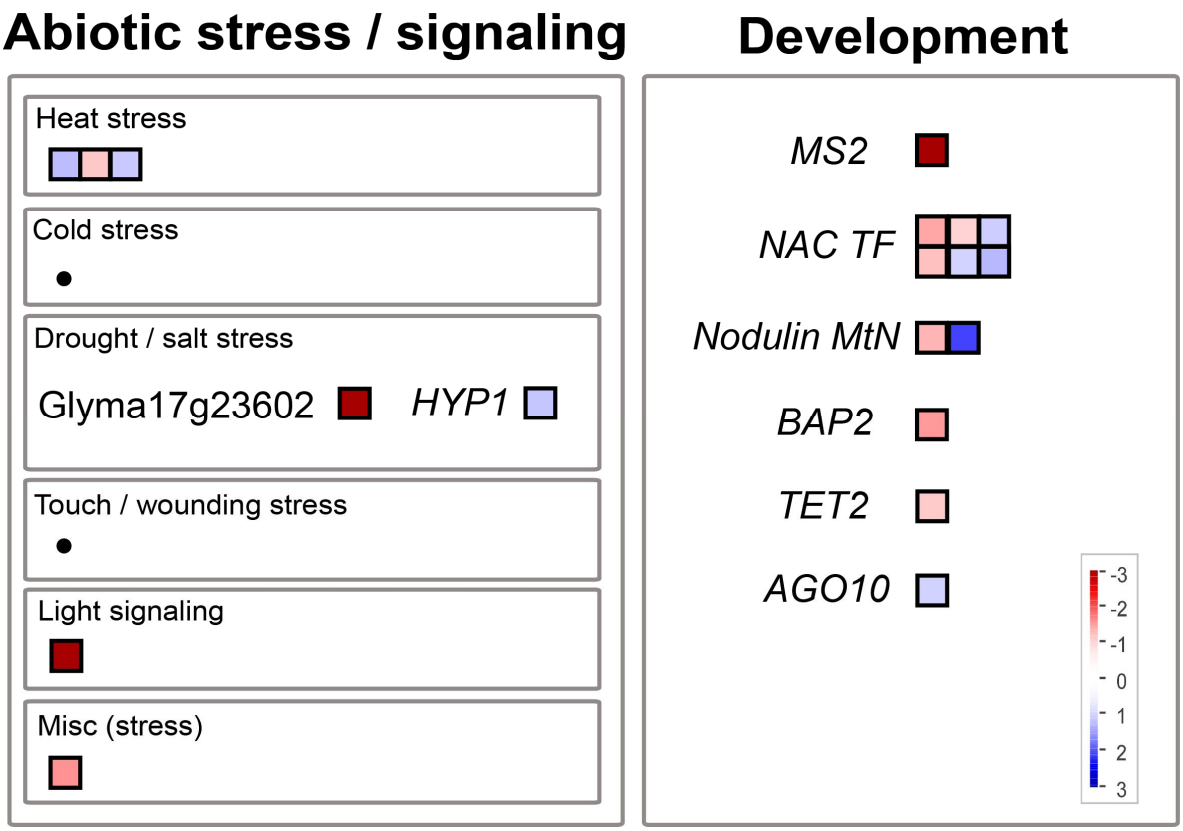


Figure 3-5. Expression of DEGs mapped to the drought stress and development pathway. Each box represented gene mapped to each pathway. The color of box represented log2 fold change. Negative values indicated by red color represented that the gene was down regulated in more branching NIL, otherwise, positive value indicated by blue color represented that it was up regulated in more branching NIL.

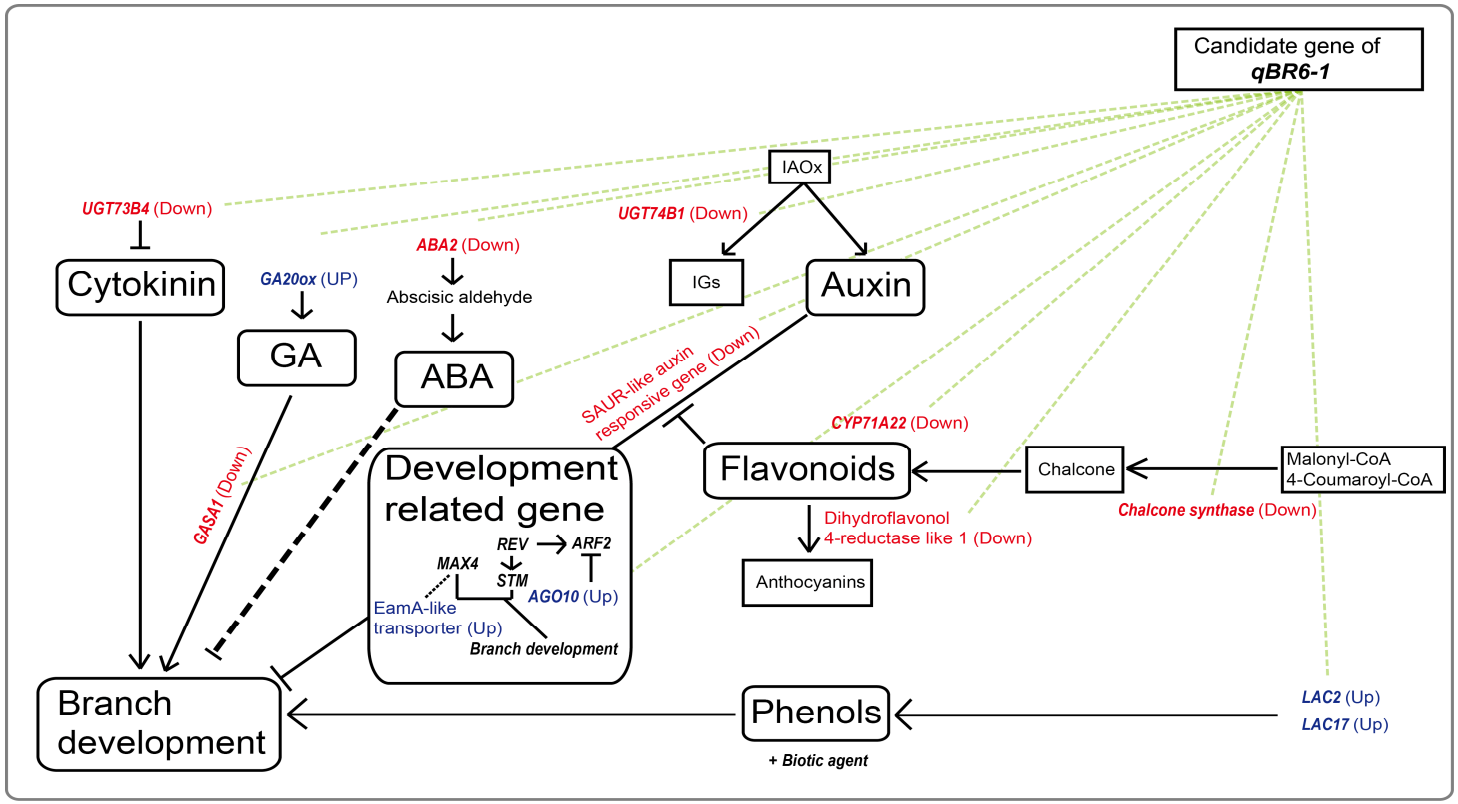


Figure 3-6. Summary schematics of gene expression pattern in more branching NIL. Mapped DEGs were expressed as gene symbol. Color of gene symbols represented up-/down-regulated genes in more branching NIL. Blue color represented up-regulated genes, otherwise, red represented down-regulated genes.

Chapter 4.

Duplication and diversification mechanisms of TCP family of transcription factors in soybean (*Glycine max* [L.] Merrill)

Abstract

TCP protein family is one of the plant-specific transcription factors (TFs), which are involved in developmental growth, such as leaf development, flower symmetry and shoot branching. Several comprehensive analysis of the TCP transcription factors have been reported in *Arabidopsis thaliana*, *Citrullus lanatus*, *Gossypium* species, and *Solanum lycopersicum*. However, few studies of the TCP protein family have been conducted in *Glycine max* (L.) Merrill despite of its importance. A total of 55 non-redundant TCP TFs were identified over soybean chromosomes except chromosome 14. Phylogenetic analysis showed TCP genes were distinctly clustered into two main classes (Class I and Class II) consisting of 11 subgroups. In this study, evolutionary conservations of TCP were analyzed. The TCP protein domain and gene structure were highly conserved in soybean. Several

subgroup specific conservation in motif and gene structures were also identified. Expression patterns of *G. max* TCP genes over nine different tissues were also investigated using public RNA-seq data. As an evolutionary study, whole genome duplication and tandem duplication which affected TCP gene duplication events were identified. Some duplication events which were resulted divergence into different TCP subclasses were observed. The results from phylogenetic analysis and expression pattern analysis were integrated into *G. max* TCP gene duplication network. Both retention and diversification of gene expression patterns and protein structures between the duplicated gene pairs, which can explain gene balance hypothesis and neo-/sub-functionalization, were identified through the duplication network analysis. As a conclusion, the evolution of soybean TCP genes were comprehensively analyzed and the function of soybean TCP genes could be predicted based on *Arabidopsis* orthologues already characterized. This study will be helpful for functional characterization of soybean TCP TFs and evolutionary study of TCP family in other plant species.

Keywords TCP (*TEOSINTE BRANCHED 1*/ *CYCLOIDEA*/ *PROLIFERATING CELL FACTORS 1* and *2*), Transcription factors, evolution, soybean

Introduction

TCP transcription factor (TF) family exists only in higher plant species and is involved in various regulatory processes associated with growth and development, such as cell growth, cell proliferation and axillary bud formation (Cubas et al. 1999; Aguilar-Martínez et al. 2007). Its name, TCP, is originated from *TEOSINTE BRANCHED 1* (*TB1*) in maize (*Zea mays*), *CYCLOIDEA* (*CYC*) in *Antirrhinum majus* and *PROLIFERATING CELL FACTORS 1* and *2* (*PCF1* and *PCF2*) in *Oryza sativa* (Luo et al. 1996; Doebley et al. 1997; Kosugi and Ohashi 1997). This TF family contains highly conserved DNA-binding motif, named TCP domain, characterized by 59-amino-acid non-canonical basic-Helix-Loop-Helix (bHLH) involved in DNA-binding and protein-protein interaction (Cubas et al. 1999). TCP TFs are classified into two major classes referred to as Class I and II (Navaud et al. 2007). Unlike Class I, Class II TCPs carry three-amino-acid [T(R)A(SLIVP)K(RQ)] in basic region of bHLH motif and functionally unknown arginine-rich motif named R domain next to the C-terminal of TCP domain (Broholm et al. 2008). DNA-binding specificity of two classes of TCP TFs is also slightly different: GGNCCCAC for Class I and GYGGNCCC for Class II (Kosugi and Ohashi 2002). Class II is divided into CIN (CINCINNATA)-type and CYC/TB1

(CYCLODEA/TEOSINTE BRANCHED 1)-type (Martín-Trillo and Cubas 2010).

In general, two major classes of TCP TF family regulate growth and development antagonistically (Martín-Trillo and Cubas 2010; Danisman et al. 2012). The Class I TCP are supposed to promote growth based on the finding of AtTCP20, OsPCF1 and 2 (Kosugi and Ohashi 1997). Double mutants of another Class I TCP genes of *A. thaliana*, AtTCP14 and AtTCP15 repressed replication of nuclear genome without mitosis (referring to as endoreplication) during leaf development (Peng et al. 2015). Similarly, mutants of AtTCP9, AtTCP19, and AtTCP20 promoted the size of pavement cells in the outmost epidermal layer and senescence (Danisman et al. 2012, 2013). CIN- and CYC/TB1-type of Class II TCP have been reported to regulate lateral organ development and axillary bud formation, respectively (Poza-Carrión et al. 2007; Aguilar-Martínez et al. 2007; Finlayson 2007). Unlike Class I type, TCP members of Class II have negative regulatory roles in cell growth and proliferation. For example, defect of the genes encoding TCP18 of CYC/TB1-type caused more lateral branches in *Arabidopsis thaliana*, *Zea mays*, *Pisum sativum*, and *Solanum lycopersicum*, those of which are *AtBRC1(AtBRANCHED1)*, *TB1*, *PsBRC1*, and *SlBRC1*, respectively (Doebley et al. 1997; Aguilar-Martínez et al. 2007; Martín-Trillo et al. 2011; Braun et al. 2012).

Similarly, a TCP18 homolog of soybean (*Glycine max* L.), *GmBRC1* (Glyma06g23410; Glyma.06G210600), was suggested as a candidate gene for a major QTL associated with branching (Chapter 1).

Soybean is one of the important crops cultivated in worldwide for food materials of human, animal feed, and industrial materials. In spite of the importance of TCP TFs in development and growth, only a few studies for soybean TCP TFs have been conducted; in a study referred to population structure and domestication of Korean soybeans, three soybean TCP genes were identified as candidate genes for artificial selection (Chung et al. 2014). In another study for genome-wide identification of soybean TCP, comprehensive analysis for phylogenetic analysis and motif conservation were conducted (Feng et al. 2018). However, the number of TCP genes identified and used in those study was different; 52 genes presented in first study, whereas, 54 genes in later one. Furthermore, one of the paralog of TCP18 (Glyma.06G210600) named as *BRANCHED1* was absent in Feng et al. Comprehensive evolution study of TCP transcription factor family is prominent to understand the context of a transcription factor family profoundly. However, study regarding duplication and diversification mechanism of soybean TCP transcription factor has not been conducted. Meanwhile, the well-refined genome reference sequence of soybean had been already released

(https://phytozome.jgi.doe.gov/pz/portal.html) and it provides a powerful resource for genome-wide identification of genes of interest. In other plant species with a reference genome, such as *Citrullus lanatus*, *Gossypium* species, and *Solanum lycopersicum* comprehensive analyses of TCP TF genes in terms of gene classification, gene structure, domain composition and expression pattern have been performed in the genome scale based on the *Arabidopsis* classification system (Parapunova et al. 2014; Li 2015; Ma et al. 2016a; Shi et al. 2016; Li et al. 2017).

The present study was performed to analyze the mechanism of duplication and divergence of TCP TF family in soybean. For the analysis, non-redundant TCP TF genes were identified in genome scale, and bioinformatics analysis including phylogenetic, motif conservation and expression pattern analysis were conducted. Furthermore, duplication mode of TCP genes in soybean genome was explored through analysis of synteny derived from whole genome duplication. Our finding will provide profound understanding of the evolutionary mechanism of soybean TCP TF genes and help characterize the function of TCP TF genes in growth and development of soybean and improve plant type for high yielding.

Materials and methods

TCP TFs identification in soybean genome

To search TCP TFs in soybean, the Hidden Markov Model (HMM) protein consensus sequence of TCP DNA-binding domain (PF03634) was obtained from Pfam protein family database (http://pfam.sanger.ac.uk/). The Pfam domain PF03634 was searched against soybean protein sequences from Phytozome (http://www.phytozome.net/) using HMMER software version 3.1b1 with 0.01 of e-value threshold and default parameters (Eddy 2009). The redundancy of protein sequences was manually inspected based on the sequence similarity using ClustalW (Larkin et al. 2007) and physical locations of the candidate genes coding TCP proteins were investigated (Schmutz et al. 2010). The protein sequences of *G. max* TCP genes were additionally validated by checking the presence of TCP-DNA binding domain using InterProScan software (Jones et al. 2014). The protein sequences of TCP TF genes of *Arabidopsis* and rice were retrieved from the PlantTFDB database (http://planttfdb.cbi.pku.edu.cn/) (Jin et al. 2017). *G. max* TCP genes homologous to each *Arabidopsis* TCP genes were determined by BLASTP and manually inspected based on the phylogenetic and motif conservation analysis.

Phylogentic analysis

To construct phylogeny tree of TCP TFs, the protein sequences of TCP TFs of *G. max*, *A. thaliana* and *O. sativa* were aligned using ClustalW alignment algorithm (command line version of ClustalX) (Larkin et al. 2007) with default parameter. Unrooted phylogenetic tree of TCP TFs for the three plant species was constructed using MEGA 7.0 software (Kumar et al. 2016) based on the Neighbor-joining algorithm with complete deletion option, Jones-Taylor-Thornton (JTT) model and 1,000 times of bootstraps test.

Additional phylogenetic tree of conserved TCP domains was constructed using same method described above. Conserved TCP domains of *G. max* TCP protein were retrieved using InterProScan (Jones et al. 2014) and manual python script.

Gene structure and conserved motif identification of *G. max* TCP TFs

The structures of *G. max* TCP genes were compared by Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015) using the gene features retrieved from genomic feature annotation file (GFF3) of the *G. max* reference genome available in Phytozome (Schmutz et al. 2010). Identification of conserved protein motifs for *G. max* TCP was

conducted by online Multiple Expectation maximization for Motif Elicitation (MEME; http://meme-suite.org/tools/meme) program (Bailey et al. 2009) based on the protein sequences of *G. max* TCP with maximum number of motifs: 30. The identified motifs were then functionally annotated using ProSiteProfiles included in InterProScan (Jones et al. 2014).

Expression analysis of TCP TF genes using soybean RNA-seq Atlas database

To analyze expression patterns of TCP TF genes, FPKM (Fragments per kilobase million) expressions matrix from nine different tissues in soybean were downloaded from Phytozome 10 (Schmutz et al. 2010). According to Schmutz et al.(Schmutz et al. 2010), RNA-seq reads were mapped against the second version of *G. max* reference genome using Tophat software v.2.0.13 with options "—no-discordant –no-mixed –max-multihits=2." To visualize the expression patterns of TCP genes, heatmap and histogram were constructed using an R package, 'Pheatmap'. For convenient comparison, log(FPKM+1) value was used for heatmap. The expression patterns were clustered based on the hierarchical clustering method. The number of clusters were determined by Silhouette method (Rousseeuw 1987) using R packages 'Vegan' and 'Cluster'.

Chromosomal location and duplication of *G. max* TCP gene

Physical locations of *G. max* TCP genes were obtained from *G. max* genome annotation in Phytozome (http://www.phytozome.net/). To grasp the duplication modes of TCP genes in soybean genome, duplicated gene pairs were identified by BLASTP (Boratyn et al. 2013) with e-value threshold of 1e-50. Soybean genome had experienced two times of whole genome duplications (WGD), referred as ancient and recent WGD (Schmutz et al. 2010). To detect TCP paralogous gene pairs duplicated by WGD we used information of the synteny blocks from recent WGD provided by Soybase (https://soybase.org/). Additionally, we also conducted synteny analysis using MCscanX software (Wang et al. 2012) based on BLASTP (Boratyn et al. 2013) alignments to find another duplicated genomic regions which were not included in the Soybase. Tandem duplication relationship was manually identified. The rate of synonymous substitution (Ks) and non-synonymous substitution (Ka) were calculated using add_ka_and_ks_to_collinearity.pl provided by MCScanX software (Wang et al. 2012) with some modification. Median Ks values for synteny blocks were calculated by homemade python script to date precise duplication date. The data were visualized by Circos program (Krzywinski et al. 2009). To visualize systemically paralogous relationships and expression patterns of *G. max* TCP TF gene, we

annotated phylogenetic subgroups and expression clusters to corresponding genes using home-made python codes and used Cytoscape program to visualize the data (Shannon et al. 2003).

Results

Genome-wide identification of the TCP TF gene family in *G. max* and their chromosomal distribution

We identified 88 putative TCP TF protein sequences of *G. max* based on the HMM Models of TCP domain (PF03634) downloaded from Pfam database. Among them, 33 redundant protein sequences were eliminated based on the result of multiple sequence alignment. The remaining 55 *G. max* TCP proteins were non-redundant and were confirmed to have TCP domains using InterProScan. These 55 non-redundant TCP TFs identified in soybean were used in further analysis. The 55 TCP TFs were classified following *Arabidopsis* classification system (Tatematsu et al. 2008): they had orthologous relationships with 17 out of 24 *Arabidopsis* TCP genes. The lengths of 55 *G. max* TCP genes range from 157 to 533 amino acids (aa) with an average of 357 aa. *A. thaliana* homologs, molecular weights (Mw), isoelectric points (pI) and physical positions of *G. max* TCP genes were listed with gene identifiers (Table 4-1).

The 55 *G. max* TCP genes are distributed over all chromosomes except chromosome 14 (Fig. 4-1). The number of TCP genes were not evenly distributed ranged from 0 to 5 genes per chromosome (Fig. 4-1). The highest number of five TCP genes were on chromosome 5, 6, 12,

and 13, whereas, only one TCP gene was located on chromosome 1, 2, 3, 7, 11, and 15 (Fig. 4-1).

Phylogenetic analysis of TCP TF gene family

An unrooted phylogenetic tree of soybean TCP TFs along with 24 *Arabidopsis* TCP TFs and 21 rice TCP TFs was constructed using Neighbor-Joining algorithm to elucidate phylogenetic relationship of soybean TCP TFs. The TCP TFs were classified into 11 subgroups designated to Subgroup A to K (Fig. 4-2). Subgroups A to G belonged to Class I subfamily, while the other Subgroup I, J, and K belonged to CIN-type of Class II subfamily and Subgroup H belonged to CYC/TB1-type of Class II subfamily (Fig. 4-2). The soybean TCP proteins were interspersed in all subgroups, except Subgroup G. The numbers of TCP proteins were not evenly distributed over the subgroups (Fig. 4-2). The largest Subgroup A contained 19 members of TCP proteins, comprised of five *A. thaliana*, four *O. sativa* and ten *G. max*. Subgroup A, B, E and J contained twice as many *G. max* TCP homologs as *Arabidopsis* TCP genes and there are three times or more soybean TCP homologs than those of *A. thaliana* in Subgroup C and H (Fig. 4-2). No orthologues of AtTCP16 in Subgroup G were identified in soybean (Fig. 4-2).

To verify the phylogenetic tree of TCP TF family, we constructed

additional phylogenetic tree of conserved TCP domain (Fig. 4-3). Comparison of the two phylogenetic trees resulted that the genes included in same subgroup were classified as same subgroup in the other phylogenetic tree (Fig. 4-2; Fig. 4-3).

Gene structure and conserved motifs

Gene structure and conserved motifs of soybean TCP TFs were detected to validate the phylogenetic tree constructed by using the alignment of protein sequences (Fig 4-4). We observed some variations in numbers and lengths of exon/intron/untranslated region (UTR) of soybean TCP genes among the subgroups (Fig. 4-4b). Particularly, Subgroup J and K had multiple numbers of UTR at N terminal region and only Subgroup H had two exons with intron (Fig. 4-4b).

A total of 30 conserved motifs were identified in the soybean TCP gene family (Fig. 4-4c, Fig. 4-5). Among the 30 conserved motifs, only two motifs (Motif 1 and 4) were annotated as TCP domain and R domain, whereas function of the other motifs still remained unknown (data not shown). TCP domain was conserved in all *G. max* TCP TFs but R domain was identified only in the TCP member of Subgroup J and H (Fig. 4-4c), which was consistent with fact that the R domain were conserved in some subgroups of Class II type TCP (Broholm et al. 2008). Subgroup-specific motifs were also observed. Motif 5, 9 and 13

were specific to Subgroup D (TCP7), A (TCP14), and I (TCP3 and TCP4), respectively (Fig. 4-4c). Motif 2 was identified only next to the C-terminal of soybean TCP motif in Class I ($-\log_{10}$ *P-value* = 23.5; Fig. 4-4c). The PCA analysis based on the identified motifs showed similar results with the subgroups classification in the phylogenetic tree (Fig. 4-6).

Expression analysis of TCP gene

We analyzed expression profiles of *G. max* TCP genes from nine different tissues, including stem, root, root hairs, nodules, pod, flower, seed, leaves, and shoot apical meristem (SAM). Based on expression patterns, the TCP genes were clustered into 19 clusters, designated Cluster 1 to 19, by silhouette method (Rousseeuw 1987) (Fig. 4-7). Overall, expression of genes from 29 out of 58 duplicated gene pairs having close phylogenetic relationship showed similar expression patterns and clustered as same expression cluster, otherwise, remnant showed diverged expression pattern (Table 4-2; Fig. 4-4a; Fig. 4-7). For examples of retained expression pattern, three genes in Cluster 3, which were homologous to TCP13 showed higher expression in leaves than other tissues (Fig. 4-7; Fig. 4-8). Two TCP4 genes in Cluster 9 were up-regulated in leaf and pod tissues (Fig. 4-7; Fig. 4-8). Five TCP genes in Cluster 5 are highly expressed in SAM (Fig. 4-7; Fig. 4-8), of which

three are TCP18 (Fig. 4-7; Fig. 4-8). Three TCP2 genes belonging to Cluster 11 showed also higher expression in SAM (Fig. 4-7; Fig. 4-8).

On the contrary, half of the duplicated gene pairs showed different expression patterns, resulting in different clusters, even though the genes had close phylogenetic relationship; for examples, Glyma.05G142000|TCP2 in Cluster 15 vs TCP2 genes in Cluster 11, Glyma.18G280700|TCP1 in Cluster 8 vs TCP1 genes in Cluster 4 (Fig. 4-7; Fig. 4-8). This indicated that these genes had diverged to play fine-tuned functions in specific tissues during soybean evolution after WGD. Interestingly, a gene annotated as TCP11/TCP21 in Cluster 4 and Subgroup F, which was not duplicated and had no paralog showed no gene expression in all nine tested tissues (Fig. 4-4; Fig. 4-7; Fig. 4-8).

Duplication of TCP TF genes in *G. max*

We analyzed paralogous relationships among *G. max* TCP genes to reveal the TCP gene expansion mechanisms during soybean evolution process characterized by two rounds of whole genome duplication events (Schmutz et al. 2010). Pair-wise BLASTP analysis with stringency of 1e-50 revealed 106 TCP gene pairs having homologous relationships (Table 4-2), in which 54 out of the 55 TCP genes were involved. Only one gene (Glyma.11G196000|TCP11TCP21) exists as a single copy in the soybean genome (Fig. 4-1; Table 4-2).

For deep understanding modes of TCP gene duplications in the soybean genome, we first used information about the WGD blocks in the soybean reference genome provided by Soybase (https://soybase.org/). In addition, we ourselves conducted synteny analysis to find other duplicated genome blocks which were not included in Soybase using MCscanX program. Forty-four gene pairs, constituted by 46 TCP genes, positioned on the WGD blocks in Soybase (Table 4-2). On the synteny blocks detected in our study, 15 gene pairs consisting of 23 TCP genes were located (Table 4-2). Median Ks values for synteny blocks harboring TCP genes were predominant in two ranges of 0.12–0.18 and 0.55-1.86 (Table 4-2; Fig. 4-9), indicating recent or ancient WGD events, respectively, as reported by Schmutz et al. (Schmutz et al. 2010). From these results, 30 and 29 out of 59 homologous pairs were duplicated by recent and ancient WGD (Table 4-2), respectively. The divergence time of the gene pairs showed similar Ks distribution to those by WGD of Soybase (https://soybase.org/) (Fig. 4-9). Notably, a duplicate pair (Glyma.09G284300–Glyma.09G284500) was tandemly located on the same chromosome, regarded as single gene duplication by tandem duplication, of which the Ks value was 0.01 (Table 4-2). However, remaining 46 gene pairs were irrelevant with synteny and multiple cross-links among these TCP genes in all directions were observed in

further networking by Cytoscape (Supplementary Fig. 4-10). It maybe indicated they have false paralogous relationship caused by homology of TCP proteins.

To illustrate duplication mode and divergence of TCP genes, we visualized networking of *G. max* TCP gene duplicates using Cytoscape (Fig. 4-11b). Considering twice WGD in soybean, a single interconnection consists of four duplicated genes which are cross-linked based on synteny conservation. After WGD, sequence divergence leads to a loss of one (or a few) of the links among the genes owing to decay of synteny blocks (Fig. 4-11a). We classified the 16 interconnections of TCP gene duplicates into four categories (Fig. 4-11).

The first category represented five complete interconnections (representing closed tetragons) including four TCP genes (representing four vertexes) on the four synteny blocks (representing four sides and diagonals) which were still highly conserved each other (Fig. 4-11a-1; Fig. 4-11b-1). Synteny blocks duplicated by ancient WGD were shown in dark orange and the other blocks duplicated by recent WGD were in yellow (Fig. 4-11b). The second category was five incomplete interconnections representing open tetragons where a side was missing (Fig. 4-11b-2). This was attributed to synteny decay after ancient WGD, resulting in a lack of preservation of the precise order of genes between

a donor block (A-B) for ancient WGD and a duplicated block (A') from its counterpart (B') by recent WGD (Fig. 4-11a-2). In this category, we observed two incomplete interconnections linked by two unexpected syntenys of Glyma.12G208800|TCP4 - Glyma.13G271700|TCP3 and Glyma.12G228300|TCP3 - Glyma.13G292500|TCP4 (Fig. 4-12; Fig. 4-11b-2). These two synteny relationships had median Ks value of 0.13 and were identified by DAGchainer and posted in Soybase (https://soybase.org/) (Table 4-2). However, the duplication relationship between the four directly interconnected genes (Glyma.12G208800|TCP4, Glyma.13G271700|TCP3, Glyma.12G228300|TCP3 and Glyma.13G292500|TCP4) seemed to do not fitted with the characterized two rounds of soybean whole genome duplication (Fig. 4-12; Fig. 4-11b-2; Table 4-2).

The third category was the incomplete interconnection including the tandemly duplicated TCP gene pair (Fig. 4-11a-3; Fig. 4-11b-3). A paralogous pair expanded by ancient WGD, Glyma.09G284300-Glyma.10G285900, and two pairs by recent WGD, Glyma.10G285900-Glyma.20G103400 and Glyma.09G284300-Glyma.20G001600 formed the incomplete interconnection (Fig. 4-11b-3). On this interconnection, Glyma.09G284300|TCP7 and Glyma.09G284500|TCP7 were tandemly duplicated, representing a dotted line segment attached to a tetragon (Fig. 4-11b-3). The last category contained five gene pairs duplicated

only by recent WGD, which reside on the highly conserved synteny blocks. They represented closed line interconnections including both endpoints (Fig. 4-11b-4).

Interestingly, structurally diverged gene pairs were observed. For example, the ancient duplicates of Glyma.01G045500|TCP14 (alternatively, Glyma.02G105900|TCP14) vs Glyma.18G121400|TCP15 (alternatively, Glyma.08G299400|TCP23) in the first category were diverged in both motif structure and expression pattern (Fig. 4-4; Fig. 4-11b). In addition, a gene (Glyma.13G056000) lost functional TCP domain in protein structure were connected by three TCP transcription factors (Fig. 4-11b-1). The gene without TCP domain was putatively duplicated from Glyma.19G030900 by recent WGD. Protein sequence of the non-TCP domain containing gene (Glyma.13G056000) had deletion in N-terminal region containing TCP domain in TCP paralogs and consequently had shorter protein sequence (Fig. 4-13). Furthermore, the remaining protein sequence of Glyma.13G056000 showed motif conservation with a copy of recent whole genome duplicates (Fig. 4-13). The nucleotide sequence alignment of the gene pair (Glyma.13G056000 – Glyma.19G030900) including upstream sequence resulted that gap filled with multiple unknown nucleotide sequence (N) to anchor scaffold in genome assembly was attributed to

loss of gene sequence coding TCP domain, and consequently, misannotation (data not shown).

Meanwhile, we explored alterations or retention of the expression pattern among the gene duplicates on the networks. Overall, 19 out of 29 gene pairs duplicated by ancient WGD (linked by dark orange bars) in total interconnections had diverged expression patterns (Table 4-2; Fig. 4-11b); for example, four soybean paralog genes (Glyma.04G161400, Glyma.05G019900, Glyma.06G204300 and Glyma.17G079900) for TCP5 belonging to Subgroup K exhibited the incomplete interconnection of Category 2, showing diverged expression patterns in three interconnections duplicated by ancient WGD (Fig. 4-4; Fig. 4-7 and Fig. 4-11b). On the other hand, expression patterns of recently duplicated genes (linked by yellow bars) was less diverged (36%) compared to gene expression pattern changes of gene pairs duplicated by ancient WGD, when the two suspicious syntenys were excluded (Fisher's exact test, *p-value* = 0.035, Table 4-2; Fig. 4-11b). For example, expression patterns of two TCP5 genes Glyma.04G161400 and Glyma.06G204300 which were duplicated by recent WGD were retained (Cluster 8) and showed high expression in flower and shoot apical meristem (SAM) tissues (Fig. 4-7), was consistent with previous report that the TCP5 was regulated by light in shoot apices (López-Juez et al. 2008). Otherwise, two paralogues

(Glyma.05G019900 and Glyma.17G079900) which were expanded by recent WGD, showed diverged expression patterns (Cluster 12 and 6) and exhibited low expression in flower tissue (Fig. 4-4; Fig. 4-7; Fig. 4-11). The tandemly duplicated gene pair, Glyma.09G284300 and Glyma.09G284500, displayed no change in the expression pattern (Fig. 4-11b-3). In three out of five single connects, duplicated by recent WGD showed the same expression patterns (Fig. 4-11b-4).

Discussion

In this study, we identified a total of 55 TCP coding genes in *G. max* genome (Table 4-1). They were classified into 11 subgroups (A to K) by phylogenetic analysis based on the TCP protein sequences (Fig. 4-2). Hieratical clustering analysis based on expression patterns grouped the soybean TCP genes into 19 clusters (1 to 19) (Fig. 4-7). Analysis of protein domain and gene structure revealed a motif specific to Class I TCP TFs. It was motif 2 identified in our study. The major difference between class I and II types of TCP has been reported to be the presence of R domain in some Class II TCP in *Arabidopsis* (Cubas et al. 1999; Broholm et al. 2008). Similar to this, R domain, designated as Motif 4 in this study, was only conserved in Subgroup H and J of class II TCP in soybean ($-\log_{10}$ *P-value* = 4.9; Fig. 4-4c). On the other hand, the presence of Motif 2 was observed in most of Class I TCP TFs, showing high association with the soybean TCP genes of Class I ($-\log_{10}$ *P-value* = 23.5; Fig. 4-4c). The antagonistic regulatory function of class I and II TCP had been reported in growth of *Arabidopsis* (Danisman et al. 2012). Even though it has not been functionally characterized, this result indicates that Motif 2 is likely involved in the antagonistic regulation of TCP genes between class I and II in soybean, along with R domain. In addition, we identified subgroup-specific motifs. Motif 5, 9 and 13 were only conserved in Subgroup D (TCP7), A (TCP14), and I

(TCP3 and TCP4) (Fig. 4-4c), respectively. These motifs might be involved in specific functions of each of the corresponding TCPs.

Overall, total number of soybean TCP genes were 2.3 times as much as that of *A. thaliana* but the number of soybean TCP genes in each subgroup were diverse. Six subgroups (A, B, E, J, C and H) had twice or three times as much or more TCP genes than *Arabidopsis* orthologues (Fig. 4-2). Subgroup F contained only one soybean TCP gene (Glyma.11G96000|TCP11/TCP21) and there was no soybean orthologue of AtTCP16 in Subgroup G (Fig. 4-2). These results indicated that the soybean TCP genes in each subgroup had been differentially expanded over the evolutionary time. TCP16 in *A. thaliana* has been demonstrated to play a crucial role in early processes in pollen development using transgenic plants harboring a TCP16 RNA interference (Takeda et al. 2006). TCP11 has been also reported to show similar functions to TCP16, as developmental regulators that influences the growth of leaves, stems and petioles, and pollen development (Takeda et al. 2006; Viola et al. 2011). In spite of functional redundancy of these two TCP genes, multiple orthologues of AtTCP11 and AtTCP16 has been identified in several plant species such as *Gossypium hirsutum*, *Gossypium raimondii*, *Populus euphratica*, *Populus trichorcarpa* and *Citrullus lanatus* (Ma et al. 2014, 2016b; Shi et al. 2016; Li et al. 2017). In the case of *G. max*, however,

all duplicates but one copy of TCP11/TCP21 and all TCP16 gene copies are deemed to be delete or pseudogenized, eventually causing gene loss of duplicate genes (Rutter et al. 2012; Panchy et al. 2016). Furthermore, the observation of no gene expression of TCP11/TCP21 gene (Glyma.11G196000) all nine tested tissues (Fig. 4-7; Fig. 4-8), also suggested that the gene probably is under non-functionalization procedure which has been previously suggested (Lynch and Conery 2000; Duarte et al. 2006; Yang et al. 2011).

Interconnecting of 54 duplicated genes, except the single copy TCP gene (Glyma.11G196000), were schematically represented using Cytoscape fed information of synteny and gene expression pattern (Fig. 4-4; Fig. 4-7; Fig. 4-11b). From these results, we found that expansion modes of soybean TCP genes are WGD and tandem duplication. The Ks distribution of TCP gene pairs showed two peaks of 0.12-0.18 and 0.55-1.86 (Table 4-2; Fig. 4-9), which indicates two rounds of WGD that occurred 13 and 59 million years ago (mya) in soybean, respectively (Schmutz et al. 2010). This result is contrast with the fact that segmental duplication affected the expansion of TCP in other plant species such as *S. lycopersicum*, *P. euphratica* and *Gossypium* spp. (Parapunova et al. 2014; Ma et al. 2014, 2016b; Li et al. 2017).

The fate of duplicated genes is explained by numbers of theories, including gene balance hypothesis (Birchler and Veitia 2007), gain-of-

function hypothesis (Ohno 1970), subfunctionalization (Force et al. 1999), increased gene dosage hypothesis (Conant and Wolfe 2008) and functional buffering model (Chapman et al. 2006). Within each interconnection, we observed both retention and diversification of paralogous genes after the duplication events in terms of protein structure and gene expression pattern which were supported by gene balance hypothesis, gain-of-function (neo-functionalization) and sub-functionalization. Among the interconnections, a total 18 out of 28 paralogous TCP gene pairs duplicated by recent WGD, excluding two suspicious interconnections, showed the similar expression patterns (Table 4-2; Fig. 4-11b), indicating functional retention. There are two opinions why both copies of a duplicated pair persist without functional changes: either because insufficient time has passed for a deleterious duplicate or there is selection pressure to retain redundant functions and expression patterns to maintain proper balance which could be supported by gene balance hypothesis (Edger and Pires 2009; Panchy et al. 2016).

On the other hand, we also observed examples of neo-/sub-functionalized gene pairs which exhibited different expression pattern compared to the counterpart duplicated by WGD (Fig. 4-11b). It has been suggested that divergence of gene expression may play an important role in preservation of duplicated genes by neo-/sub-

functionalization (Duarte et al. 2006; Ganko et al. 2007). Interestingly diversification ratio in expression pattern of the paralogous pairs duplicated by ancient WGD were much higher than that in young duplicates by recent WGD, showing 66% vs 36%, respectively (Table 4-2; Fig. 4-11b). This result could be explained by the age of gene duplication. Negative correlation between the synonymous substitution rate (Ks) and correlation coefficient of expression pattern of duplicates has been observed under limited nonsynonymous substitution rate (Ka) of less than 0.3 (Gu et al. 2002). The similar correlation has also been reported in *Oryza sativa* when Ks less than 2.0 (Li et al. 2009). Higher frequency of diverged expression patterns among the TCP duplicates expanded by ancient WGD compared to ones duplicated by recent WGD could be explained by accumulation of neutral mutation on genomic regions affecting expression of a gene over evolutionary time (Birchler et al. 2007). Meanwhile, the structural diversification, which was represented by motif composition and annotation, along with expression pattern diversification was also identified in ancient WG duplicated gene pairs of three interconnections represented by red box in Fig. 4-11. This indicated that coupled divergence of protein structure and gene expression occurred in a part of TCP duplicated by ancient WGD and could be supported by the observation of coupled divergence

in protein sequence and gene expression in *Drosophila* species (Nuzhdin et al. 2004; Lemos et al. 2005).

Intriguingly, incomplete interconnection linked by tandem duplicates was observed in Category 3 (Fig. 4-11b-3). Mechanisms of tandem duplication mediated by unequal cross-over and duplicative transposition has been proposed (Zhang 2003; Taylor and Raes 2005; Casneuf et al. 2006). Based on these mechanisms, possible scenario of tandem duplication mechanism of soybean TCP genes was depicted using annotation of repetitive sequences and sequence comparison of tandemly duplicated blocks (Fig. 4-14). In this scenario, promising unequal cross-over was occurred by misalignment between inverted donor sequence (containing T-rich repetitive sequence with LTR retrotransposon at each side of flanking sequence) and pre-existed recipient sequence (containing reverse complementary sequence of T-rich repeat and LTR retrotransposon) (Fig. 4-14). On the contrary, the positive correlation between the number of retro-transposable elements and tandem duplications has been reported (Thomas and Schneider 2011; Kono et al. 2017). In this study, nine LTR retro-transposable elements were identified in the adjacent intergenic region of tandem duplicates blocks (data not shown). This result indicated that the possibility of duplicative transposition resulted from a transposition mediated by LTR retro-transposable element cannot be excluded.

The recipient expanded by tandem duplication (Glyma.09G284500), was most recently duplicated from its donor (Glyma.09G284300, Ks = 0.01; Table 4-2), otherwise, the recipient (Glyma.09G283500) did not have duplication relationship with the gene recently duplicated by WGD from same donor (Fig. 4-11b-3). For these genes, similar expression pattern was displayed between the tandem duplicated gene pairs, otherwise, more diverged expression patterns observed between the gene pairs duplicated by recent WGD (Fig. 4-8), suggested that the tandemly duplicated soybean TCP gene expanded for specific objective like dosage effect (Panchy et al. 2016) or expression of the duplicates were balanced after tandem duplication to maintain proper stoichiometric balance (Edger and Pires 2009). On the contrary, it has been reported that gene pair duplicated by small-scale duplication including tandem and dispersed duplication are more diverged expression patterns compared to that of large-scale duplication (Casneuf et al. 2006). It has been suggested that disruption of promoter sequence caused by unequal cross-over and duplicative transposition may cause rapid divergence of expression pattern (Casneuf et al. 2006). In tandem duplicates of this study, duplicated block contained not only TCP gene along with 1.3 kbp upstream sequence but also adjacent gene located in downstream sequence of TCP gene (Fig. 4-14). This indicated that the similar expression patterns which was retained in the

tandemly duplicated TCP gene pair may be attributed to conservation of regulatory sequence located in up-/downstream of TCP duplicates.

In summary, soybean TCP genes were identified and classified through the protein sequence comparison and validated by phylogenetic analysis, conservation study on subgroup specific motifs and gene structures. Through the network of soybean TCP gene duplication based on the expression and duplication analysis, comprehensive mechanism of duplication and diversification of soybean TCP TF family was analyzed. Based on these analysis, it was observed that functional divergence, which can explain sub-/neo-functionalization, and retention of dosage-sensitive TF genes, which can explain gene balance hypothesis, had been arisen along with the TCP TF gene expansions accelerated by two rounds of WGD and tandem duplication. This study provides a basis for functional characterization of soybean TCP genes using transformation or CRISPR system and evolution of TCP genes in other plant species.

References

Aguilar-Martínez JA, Poza-Carrión C, Cubas P (2007) *Arabidopsis BRANCHED1* Acts as an Integrator of Branching Signals within Axillary Buds. Plant Cell 19:458–472. doi: 10.1105/tpc.106.048934

Bailey TL, Boden M, Buske FA, et al (2009) MEME Suite: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208. doi: 10.1093/nar/gkp335

Birchler JA, Veitia RA (2007) The Gene Balance Hypothesis: From Classical Genetics to Modern Genomics. Plant Cell 19:395–402. doi: 10.1105/tpc.106.049338

Birchler JA, Yao H, Chudalayandi S (2007) Biological consequences of dosage dependent gene regulatory systems. Biochim Biophys Acta 1769:422–428. doi: 10.1016/j.bbaexp.2006.12.002

Boratyn GM, Camacho C, Cooper PS, et al (2013) BLAST: a more efficient report with usability improvements. Nucleic Acids Res 41:W29–W33. doi: 10.1093/nar/gkt282

Braun N, Germain A de S, Pillot J-P, et al (2012) The Pea TCP Transcription Factor PsBRC1 Acts Downstream of Strigolactones to Control Shoot Branching. Plant Physiol 158:225–238. doi: 10.1104/pp.111.182725

Broholm SK, Tähtiharju S, Laitinen RAE, et al (2008) A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. Proc Natl Acad Sci 105:9117–9122. doi: 10.1073/pnas.0801359105

Casneuf T, De Bodt S, Raes J, et al (2006) Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. Genome Biol 7:R13. doi: 10.1186/gb-2006-7-2-r13

Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006) Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proc Natl Acad Sci 103:2730–2735. doi: 10.1073/pnas.0507782103

Chung W-H, Jeong N, Kim J, et al (2014) Population Structure and Domestication Revealed by High-Depth Resequencing of Korean Cultivated and Wild Soybean Genomes. DNA Res Int J Rapid Publ Rep Genes Genomes 21:153–167. doi: 10.1093/dnares/dst047

Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. Nat Rev Genet 9:938–950. doi: 10.1038/nrg2482

Cubas P, Lauter N, Doebley J, Coen E (1999) The TCP domain: a motif found in proteins regulating plant growth and development. Plant J 18:215–222. doi: 10.1046/j.1365-313X.1999.00444.x

Danisman S, Dijk V, J AD, et al (2013) Analysis of functional redundancies within the *Arabidopsis* TCP transcription factor family. J Exp Bot 64:5673–5685. doi: 10.1093/jxb/ert337

Danisman S, Wal F van der, Dhondt S, et al (2012) Arabidopsis Class I and Class II TCP Transcription Factors Regulate Jasmonic Acid Metabolism and Leaf Development Antagonistically. Plant Physiol 159:1511–1523. doi: 10.1104/pp.112.200303

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nat Lond 386:485–8. doi: http://dx.doi.org/10.1038/386485a0

Duarte JM, Cui L, Wall PK, et al (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. Mol Biol Evol 23:469–478. doi: 10.1093/molbev/msj051

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. In: Genome Informatics 2009. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO., pp 205–211

Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol 17:699–717. doi: 10.1007/s10577-009-9055-9

Feng Z-J, Xu S-C, Liu N, et al (2018) Soybean TCP transcription factors: Evolution, classification, protein interaction and stress and hormone responsiveness. Plant Physiol Biochem 127:129–142. doi: 10.1016/j.plaphy.2018.03.020

Finlayson SA (2007) Arabidopsis TEOSINTE BRANCHED1-LIKE 1 Regulates Axillary Bud Outgrowth and is Homologous to Monocot TEOSINTE BRANCHED1. Plant Cell Physiol 48:667–677. doi: 10.1093/pcp/pcm044

Force A, Lynch M, Pickett FB, et al (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Ganko EW, Meyers BC, Vision TJ (2007) Divergence in Expression between Duplicated Genes in Arabidopsis. Mol Biol Evol 24:2298–2309. doi: 10.1093/molbev/msm158

Gu Z, Nicolae D, Lu HH-S, Li W-H (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet 18:609–613. doi: 10.1016/S0168-9525(02)02837-8

Hu B, Jin J, Guo A-Y, et al (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31:1296–1297. doi: 10.1093/bioinformatics/btu817

Jin J, Tian F, Yang D-C, et al (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45:D1040–D1045. doi: 10.1093/nar/gkw982

Jones P, Binns D, Chang H-Y, et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. doi: 10.1093/bioinformatics/btu031

Kono TJY, Brohammer AB, McGaugh SE, Hirsch C (2017) Tandem Duplicate Genes in Maize are Abundant and Date to Two Distinct Periods of Time. doi: 10.1101/238121

Kosugi S, Ohashi Y (1997) PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. Plant Cell Online 9:1607–1619. doi: 10.1105/tpc.9.9.1607

Kosugi S, Ohashi Y (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J 30:337–348. doi: 10.1046/j.1365-313X.2002.01294.x

Krzywinski MI, Schein JE, Birol I, et al (2009) Circos: An information aesthetic for comparative genomics. Genome Res. doi: 10.1101/gr.092759.109

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874. doi: 10.1093/molbev/msw054

Larkin MA, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. doi: 10.1093/bioinformatics/btm404

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL (2005) Evolution of Proteins and Gene Expression Levels are Coupled in *Drosophila* and are Independently Associated with mRNA Abundance, Protein Length, and Number of Protein-Protein Interactions. Mol Biol Evol 22:1345–1354. doi:

10.1093/molbev/msi122

Li S (2015) The *Arabidopsis thaliana* TCP transcription factors: A broadening horizon beyond development. Plant Signal Behav 10:. doi: 10.1080/15592324.2015.1044192

Li W, Li D-D, Han L-H, et al (2017) Genome-wide identification and characterization of TCP transcription factor genes in upland cotton (*Gossypium hirsutum*). Sci Rep 7:10118. doi: 10.1038/s41598-017-10609-2

Li Z, Zhang H, Ge S, et al (2009) Expression pattern divergence of duplicated genes in rice. BMC Bioinformatics 10:S8. doi: 10.1186/1471-2105-10-S6-S8

López-Juez E, Dillon E, Magyar Z, et al (2008) Distinct Light-Initiated Gene Expression and Cell Cycle Programs in the Shoot Apex and Cotyledons of *Arabidopsis*. Plant Cell 20:947–968. doi: 10.1105/tpc.107.057075

Luo D, Carpenter R, Vincent C, et al (1996) Origin of floral asymmetry in *Antirrhinum*. Nat Lond 383:794–9. doi: http://dx.doi.org/10.1038/383794a0

Lynch M, Conery JS (2000) The Evolutionary Fate and Consequences of Duplicate Genes. Science 290:1151–1155. doi: 10.1126/science.290.5494.1151

Ma J, Liu F, Wang Q, et al (2016a) Comprehensive analysis of TCP transcription factors and their expression during cotton (*Gossypium arboreum*) fiber early development. Sci Rep 6:21535. doi: 10.1038/srep21535

Ma J, Wang Q, Sun R, et al (2014) Genome-wide identification and expression analysis of TCP transcription factors in *Gossypium raimondii*. Sci Rep 4:. doi: 10.1038/srep06645

Ma X, Ma J, Fan D, et al (2016b) Genome-wide Identification of TCP Family Transcription Factors from *Populus euphratica* and

Their Involvement in Leaf Shape Regulation. Sci Rep 6:32795. doi: 10.1038/srep32795

Martín-Trillo M, Cubas P (2010) TCP genes: a family snapshot ten years later. Trends Plant Sci 15:31–39. doi: 10.1016/j.tplants.2009.11.003

Martín-Trillo M, Grandío EG, Serra F, et al (2011) Role of tomato *BRANCHED1*-like genes in the control of shoot branching. Plant J 67:701–714. doi: 10.1111/j.1365-313X.2011.04629.x

Navaud O, Dabos P, Carnus E, et al (2007) TCP Transcription Factors Predate the Emergence of Land Plants. J Mol Evol 65:23–33. doi: 10.1007/s00239-006-0174-z

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM (2004) Common Pattern of Evolution of Gene Expression Level and Protein Sequence in *Drosophila*. Mol Biol Evol 21:1308–1317. doi: 10.1093/molbev/msh128

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, Berlin Heidelberg

Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of Gene Duplication in Plants. Plant Physiol 171:2294–2316. doi: 10.1104/pp.16.00523

Parapunova V, Busscher M, Busscher-Lange J, et al (2014) Identification, cloning and characterization of the tomato TCP transcription factor family. BMC Plant Biol 14:157. doi: 10.1186/1471-2229-14-157

Peng Y, Chen L, Lu Y, et al (2015) The Ubiquitin Receptors DA1, DAR1, and DAR2 Redundantly Regulate Endoreduplication by Modulating the Stability of TCP14/15 in Arabidopsis. Plant Cell Online 27:649–662. doi: 10.1105/tpc.114.132274

Poza-Carrión C, Aguilar-Martínez JA, Cubas P (2007) Role of TCP Gene *BRANCHED1* in the Control of Shoot Branching in

Arabidopsis. Plant Signal Behav 2:551–552

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. doi: 10.1016/0377-0427(87)90125-7

Rutter MT, Cross KV, Van Woert PA (2012) Birth, death and subfunctionalization in the *Arabidopsis* genome. Trends Plant Sci 17:204–212. doi: 10.1016/j.tplants.2012.01.006

Schmutz J, Cannon SB, Schlueter J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. doi: 10.1038/nature08670

Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. doi: 10.1101/gr.1239303

Shi P, Guy KM, Wu W, et al (2016) Genome-wide identification and expression analysis of the *ClTCP* transcription factors in *Citrullus lanatus*. BMC Plant Biol 16:. doi: 10.1186/s12870-016-0765-9

Takeda T, Amano K, Ohto M, et al (2006) RNA Interference of the *Arabidopsis* Putative Transcription Factor *TCP16* Gene Results in Abortion of Early Pollen Development. Plant Mol Biol 61:165–177. doi: 10.1007/s11103-006-6265-9

Tatematsu K, Nakabayashi K, Kamiya Y, Nambara E (2008) Transcription factor AtTCP14 regulates embryonic growth potential during seed germination in *Arabidopsis thaliana*. Plant J 53:42–52. doi: 10.1111/j.1365-313X.2007.03308.x

Taylor JS, Raes J (2005) CHAPTER 5 - Small-Scale Gene Duplications. In: Gregory TR (ed) The Evolution of the Genome. Academic Press, Burlington, pp 289–327

Thomas JH, Schneider S (2011) Coevolution of retroelements and tandem zinc finger genes. Genome Res 21:1800–1812. doi:

10.1101/gr.121749.111

Viola IL, Uberti Manassero NG, Ripoll R, Gonzalez DH (2011) The *Arabidopsis* class I TCP transcription factor AtTCP11 is a developmental regulator with distinct DNA-binding properties due to the presence of a threonine residue at position 15 of the TCP domain. Biochem J 435:143–155. doi: 10.1042/BJ20101019

Wang Y, Tang H, DeBarry JD, et al (2012) *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:e49–e49. doi: 10.1093/nar/gkr1293

Yang L, Takuno S, Waters ER, Gaut BS (2011) Lowly expressed genes in *Arabidopsis thaliana* bear the signature of possible pseudogenization by promoter degradation. Mol Biol Evol 28:1193–1203. doi: 10.1093/molbev/msq298

Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18:292–298. doi: 10.1016/S0169-5347(03)00033-8

Table 4-1. TCP transcription factor gene family in *Glycine max*.

Gene ID	AtTCP homolog	Length(aa)	MW(kD)	pI(pH)	Physical position
Glyma.01G045500.1.p	TCP14	386	40.17	6.79	Chr01:5226449..5228845
Glyma.02G105900.1.p	TCP14	425	44.19	6.44	Chr02:10090282..10092450
Glyma.03G018800.1.p	TCP9	337	35.88	9.55	Chr03:1875182..1876855
Glyma.04G152400.1.p	TCP18	386	44.19	6.30	Chr04:34221031..34225294
Glyma.04G161400.1.p	TCP5	388	42.98	9.44	Chr04:40001447..40003715
Glyma.04G170600.1.p	TCP14	399	43.73	7.25	Chr04:42715265..42717152
Glyma.05G013300.1.p	TCP18	377	42.50	6.64	Chr05:1241512..1244692
Glyma.05G019900.1.p	TCP5	392	43.20	9.59	Chr05:1779847..1782008
Glyma.05G027400.1.p	TCP14	417	44.90	6.85	Chr05:2388743..2390840
Glyma.05G050400.1.p	TCP9	347	36.47	7.14	Chr05:4505731..4507432
Glyma.05G142000.1.p	TCP2	479	52.81	7.39	Chr05:33530449..33539471
Glyma.06G193000.1.p	TCP14	409	44.50	6.77	Chr06:17124047..17126015
Glyma.06G204300.1.p	TCP5	391	43.12	9.23	Chr06:19210586..19213448
Glyma.06G210600.1.p	TCP18	397	46.02	6.01	Chr06:20867484..20869778
Glyma.06G232300.1.p	TCP3	392	42.49	6.41	Chr06:36693772..36696765
Glyma.06G284500.1.p	TCP4	334	37.38	6.04	Chr06:47294337..47297404
Glyma.07G080300.1.p	TCP9	336	35.59	9.41	Chr07:7300143..7301150
Glyma.08G097900.1.p	TCP2	469	51.57	7.62	Chr08:7489720..7494791
Glyma.08G247300.1.p	TCP13	356	39.40	8.48	Chr08:21414152..21416586
Glyma.08G256400.1.p	TCP1	371	41.21	8.82	Chr08:22923150..22925057
Glyma.08G299400.1.p	TCP23	287	31.26	8.90	Chr08:41743129..41746657
Glyma.09G284300.1.p	TCP7	241	25.78	9.21	Chr09:49956726..49959624
Glyma.09G284500.1.p	TCP7	240	25.68	9.21	Chr09:49968138..49971046
Glyma.10G057400.1.p	TCP11	191	20.20	6.57	Chr10:5241972..5242880
Glyma.10G240200.1.p	TCP19	312	32.97	6.33	Chr10:46871673..46872747
Glyma.10G246200.1.p	TCP1	381	42.97	6.94	Chr10:47469659..47471353
Glyma.10G285900.1.p	TCP7	247	26.04	9.90	Chr10:50569503..50574849
Glyma.11G196000.1.p	TCP11,TCP21	157	17.02	5.81	Chr11:26998156..26998643
Glyma.12G121500.1.p	TCP4	292	32.09	6.47	Chr12:13053141..13054090
Glyma.12G158900.1.p	TCP3	398	43.15	6.40	Chr12:26907587..26910452
Glyma.12G168300.1.p	TCP8	435	46.04	6.61	Chr12:32320419..32323195
Glyma.12G208800.1.p	TCP4	344	37.17	6.22	Chr12:36808660..36812062
Glyma.12G228300.1.p	TCP3	378	39.91	6.22	Chr12:38813527..38816237
Glyma.13G047400.1.p	TCP1	420	46.57	9.21	Chr13:14255642..14257802
Glyma.13G144100.1.p	TCP11	186	19.62	7.14	Chr13:25678350..25679495
Glyma.13G219900.1.p	TCP2	533	58.38	7.37	Chr13:33290348..33300721
Glyma.13G271700.1.p	TCP3	394	41.78	6.72	Chr13:37375211..37377285
Glyma.13G292500.1.p	TCP4	345	37.38	6.27	Chr13:39246642..39249825
Glyma.15G092500.1.p	TCP2	512	56.04	6.63	Chr15:7156448..7167239
Glyma.16G004300.1.p	TCP8	362	38.26	6.80	Chr16:237659..239566
Glyma.16G053900.1.p	TCP20	346	36.81	9.02	Chr16:5255082..5257112
Glyma.17G079900.1.p	TCP5	390	42.55	8.90	Chr17:6186387..6188530
Glyma.17G099100.1.p	TCP14	414	44.85	8.05	Chr17:7811940..7814942
Glyma.17G121500.1.p	TCP18	381	43.11	6.64	Chr17:9698419..9700535
Glyma.17G132400.1.p	TCP9	341	36.39	6.55	Chr17:10636008..10637777
Glyma.18G121400.1.p	TCP15	274	30.21	9.37	Chr18:15253731..15256736
Glyma.18G268700.1.p	TCP13	385	43.01	9.38	Chr18:55247696..55250091
Glyma.18G280700.1.p	TCP1	371	41.29	9.21	Chr18:56158545..56161978
Glyma.19G030900.1.p	TCP13	375	41.37	7.90	Chr19:3829902..3832827
Glyma.19G044400.1.p	TCP1	417	46.44	9.24	Chr19:6584768..6586927

Glyma.19G095300.1.p	TCP20	331	35.33	9.08	Chr19:33696059..33697935
Glyma.20G001600.1.p	TCP7	233	25.12	9.51	Chr20:168241..170379
Glyma.20G103400.1.p	TCP7	243	25.78	9.56	Chr20:34617424..34619196
Glyma.20G148500.1.p	TCP1	376	42.49	8.18	Chr20:38695595..38696722
Glyma.20G154400.1.p	TCP19	327	34.25	6.33	Chr20:39340912..39342744

Table 4-2. Duplication mode, phylogenetic subgroup, expression cluster and Ks substitution rate of duplicated TCP transcription factor gene pairs in *G. max*.

Duplication mode	Phylogenetic subgroup	Paralogous A	Arabidopsis ortholog A	Expression Cluster A	Paralogous B	Arabidopsis ortholog B	Expression Cluster B	Location of synteny block A	Locatio of synteny block B	No. of genes	Median Ks[A]	Source[B]
	A	Glyma.01G045500	TCP14	19	Glyma.08G299400	TCP23	5	Gm01:3946315..5486622	Gm08:40623058..41933603	81	0.74	Soybase
	A	Glyma.01G045500	TCP14	19	Glyma.18G121400	TCP15	5	Gm01:5059854..5450706	Gm18:14754240..15467459	37	0.82	Soybase
	A	Glyma.02G105900	TCP14	19	Glyma.08G299400	TCP23	5	Gm02:10042956..10209946	Gm08:41683076..41933603	15	0.78	Soybase
	A	Glyma.04G170600	TCP14	17	Glyma.17G099100	TCP14	17	Gm04:41326473..42834580	Gm17:7779339..8122203	64	0.74	Soybase
	A	Glyma.04G170600	TCP14	17	Glyma.05G027400	TCP14	17	Gm04:42191074..42834580	Gm05:2261858..2421928	95	0.96	Soybase
	A	Glyma.05G027400	TCP14	17	Glyma.06G193000	TCP14	17	Gm05:2049641..3970440	Gm06:14651782..18004233	91	0.68	Soybase
	A	Glyma.06G193000	TCP14	17	Glyma.17G099100	TCP14	17	Gm06:15064479..17962068	Gm17:6551300..8122203	69	0.74	Soybase
	D	Glyma.09G284300	TCP7	18	Glyma.10G285900	TCP7	2	Gm09:49938347..50138017	Gm10:50519029..50752452	6	0.55	Soybase
	D	Glyma.09G284500	TCP7	18	Glyma.10G285900	TCP7	2	Gm09:49938347..50138017	Gm10:50519029..50752452	6	0.55	Soybase
	H	Glyma.13G047400	TCP1	4	Glyma.18G280700	TCP1	8	Gm13:14162698..16600260	Gm18:54062899..56205303	16	0.70	Soybase
ANCIENT WGD	H	Glyma.18G280700	TCP1	8	Glyma.19G044400	TCP1	4	Gm18:56141403..56208314	Gm19:6526761..6767979	6	0.67	Soybase
	H	Glyma.05G013300	TCP18	5	Glyma.06G210600	TCP18	5	Gm05:1043861..1476161	Gm06:20286288..21145572	28	1.28	Soybase
	H	Glyma.06G210600	TCP18	5	Glyma.17G121500	TCP18	5	Gm06:20515068..21145572	Gm17:9486346..9885937	28	0.95	Soybase
	I	Glyma.06G284500	TCP4	12	Glyma.13G292500	TCP4	9	Gm06:46642915..47343650	Gm13:38973746..39281224	10	0.84	Soybase
	J	Glyma.05G142000	TCP2	15	Glyma.13G219900	TCP2	11	Gm05:33359443..33698314	Gm13:33181476..33485516	7	0.68	Soybase
	J	Glyma.05G142000	TCP2	15	Glyma.15G092500	TCP2	11	Gm05:33369730..33698314	Gm15:6957804..7258440	26	0.71	Soybase
	J	Glyma.08G097900	TCP2	11	Glyma.15G092500	TCP2	11	Gm08:7333883..7494791	Gm15:7156448..7258440	12	0.60	Soybase
	J	Glyma.08G097900	TCP2	11	Glyma.13G219900	TCP2	11	Gm08:7331414..7494791	Gm13:33181476..33300721	16	0.77	Soybase
	K	Glyma.08G247300	TCP13	3	Glyma.19G030900	TCP13	3	Gm08:21276337..21451007	Gm19:3752277..3845784	24	1.17	Soybase
	K	Glyma.18G268700	TCP13	3	Glyma.19G030900	TCP13	3	Gm18:54062899..55282369	Gm19:1925641..3987198	56	0.74	Soybase
	K	Glyma.05G019900	TCP5	12	Glyma.06G204300	TCP5	8	Gm05:1576538..1862854	Gm06:19081671..19882736	30	0.72	Soybase

	K	Glyma.06G204300	TCP5	8	Glyma.17G079900	TCP5	6	Gm06:19090158..19969940	Gm17:6118030..6432971	18	0.65	Soybase
	K	Glyma.04G161400	TCP5	8	Glyma.05G019900	TCP5	12	Gm04:40001447..46789897	Gm05:4340318..1779847	95	0.58	This study
	I	Glyma.12G158900	TCP3	13	Glyma.13G271700	TCP3	10	Gm12:26744877..29680486	Gm13:37464824..37275486	6	0.63	This study
	I	Glyma.06G232300	TCP3	13	Glyma.12G228300	TCP3	10	Gm06:32960675..38361292	Gm12:39169878..38629959	12	0.66	This study
	I	Glyma.06G232300	TCP3	13	Glyma.13G271700	TCP3	10	Gm06:34424166..38361292	Gm13:37583227..37091919	15	0.67	This study
	I	Glyma.06G284500	TCP4	12	Glyma.12G208800	TCP4	9	Gm06:47114654..47395432	Gm12:36915279..36745148	11	0.77	This study
	H	Glyma.04G152400	TCP18	6	Glyma.17G121500	TCP18	5	Gm04:28750437..34225294	Gm17:9700535..9011720	13	0.79	This study
	C	Glyma.03G018800	TCP9	19	Glyma.20G154400	TCP19	7	Gm03:1716204..2246512	Gm20:39454164..39117272	7	1.86	This study
	A	Glyma.01G045500	TCP14	19	Glyma.02G105900	TCP14	19	Gm01:5023297..7815312	Gm02:10019490..11362220	143	0.13	Soybase
	A	Glyma.05G027400	TCP14	17	Glyma.17G099100	TCP14	17	Gm05:2085852..3738846	Gm17:6530347..8122203	144	0.13	Soybase
	A	Glyma.08G299400	TCP23	5	Glyma.18G121400	TCP15	5	Gm08:41683076..42340832	Gm18:13579833..15467459	275	0.14	Soybase
	A	Glyma.12G168300	TCP8	1	Glyma.16G004300	TCP8	18	Gm12:31744153..32323195	Gm16:13845..239566	32	0.14	Soybase
	B	Glyma.16G053900	TCP20	2	Glyma.19G095300	TCP20	2	Gm16:4976323..5954298	Gm19:30721228..34220308	91	0.14	Soybase
	C	Glyma.10G240200	TCP19	7	Glyma.20G154400	TCP19	7	Gm10:39982780..50873313	Gm20:34179794..45820661	917	0.13	Soybase
	D	Glyma.09G284300	TCP7	18	Glyma.20G103400	TCP7	1	Gm09:49439843..50148711	Gm20:2658..671704	31	0.15	Soybase
	D	Glyma.09G284500	TCP7	18	Glyma.20G103400	TCP7	1	Gm09:49439843..50148711	Gm20:2658..671704	31	0.15	Soybase
RECENT WGD	D	Glyma.10G285900	TCP7	2	Glyma.20G103400	TCP7	1	Gm10:39982780..50873313	Gm20:34179794..45820661	917	0.13	Soybase
	E	Glyma.10G057400	TCP11	7	Glyma.13G144100	TCP11	7	Gm10:2926746..5548244	Gm13:23298362..25960846	206	0.14	Soybase
	H	Glyma.08G256400	TCP1	4	Glyma.18G280700	TCP1	8	Gm08:20487083..23451765	Gm18:54835660..56460708	127	0.13	Soybase
	H	Glyma.10G246200	TCP1	4	Glyma.20G148500	TCP1	4	Gm10:39982780..50873313	Gm20:34179794..45820661	917	0.13	Soybase
	H	Glyma.13G047400	TCP1	4	Glyma.19G044400	TCP1	4	Gm13:14021814..14556762	Gm19:6236487..7434576	26	0.13	Soybase
	I	Glyma.06G284500	TCP4	12	Glyma.12G121500	TCP4	6	Gm06:47114822..47716034	Gm12:12366452..13458474	210	0.18	Soybase
	I	Glyma.12G228300	TCP3	10	Glyma.13G271700	TCP3	10	Gm12:32628128..39433993	Gm13:36698350..42283834	406	0.13	Soybase
	I	Glyma.12G228300	TCP3	10	Glyma.13G292500	TCP4	9	Gm12:32628128..39442363	Gm13:36698350..42283834	406	0.13	Soybase
	I	Glyma.12G208800	TCP4	9	Glyma.13G271700	TCP3	10	Gm12:32628128..39433993	Gm13:36698350..42283834	406	0.13	Soybase

	I	Glyma.12G208800	TCP4	9	Glyma.13G292500	TCP4	9	Gm12:32628128..39433993	Gm13:36698350..42283834	406	0.13	Soybase
	J	Glyma.13G219900	TCP2	11	Glyma.15G092500	TCP2	11	Gm13:31862953..36299903	Gm15:4397004..8505582	155	0.14	Soybase
	K	Glyma.08G247300	TCP13	3	Glyma.18G268700	TCP13	3	Gm08:20487083..23451765	Gm18:54835660..56460708	127	0.13	Soybase
	K	Glyma.04G161400	TCP5	8	Glyma.06G204300	TCP5	8	Gm04:39578853..40437993	Gm06:18979288..19470656	705	0.14	Soybase
	K	Glyma.05G019900	TCP5	12	Glyma.17G079900	TCP5	6	Gm05:1447802..2048973	Gm17:5964874..6523434	63	0.14	Soybase
	D	Glyma.09G284300	TCP7	18	Glyma.20G001600	TCP7	18	Gm09:49589498..50148711	Gm20:511459..2658	31	0.12	This study
	J	Glyma.05G142000	TCP2	15	Glyma.08G097900	TCP2	11	Gm05:30930427..37018237	Gm08:10709578..5476587	518	0.12	This study
	H	Glyma.05G013300	TCP18	5	Glyma.17G121500	TCP18	5	Gm05:749837..1437920	Gm17:9905746..9181066	63	0.12	This study
	C	Glyma.03G018800	TCP9	19	Glyma.07G080300	TCP9	19	Gm03:790311..2238630	Gm07:7668139..6304097	90	0.12	This study
	A	Glyma.04G170600	TCP14	17	Glyma.06G193000	TCP14	17	Gm04:37047731..52137231	Gm06:20310840..8629834	705	0.13	This study
	C	Glyma.05G050400	TCP9	14	Glyma.17G132400	TCP9	16	Gm05:3744640..7479534	Gm17:12880103..9928270	221	0.13	This study
	H	Glyma.04G152400	TCP18	6	Glyma.06G210600	TCP18	5	Gm04:28603342..34225294	Gm06:23377225..20867484	18	0.15	This study
	I	Glyma.06G232300	TCP3	13	Glyma.12G158900	TCP3	13	Gm06:35317058..38082796	Gm12:30744285..26371720	15	0.16	This study
TANDEM	D	Glyma.09G284300	TCP7	18	Glyma.09G284500	TCP7	18	Gm09:3364..11670	Gm09:17464..25763	2	0.01	This study

[a] Median Ks represented median Ks values of corresponded synteny block.
[b] Synteny analysis in soybase was performed using DAGchainer. In this study, we used MCScanX to predict synteny.

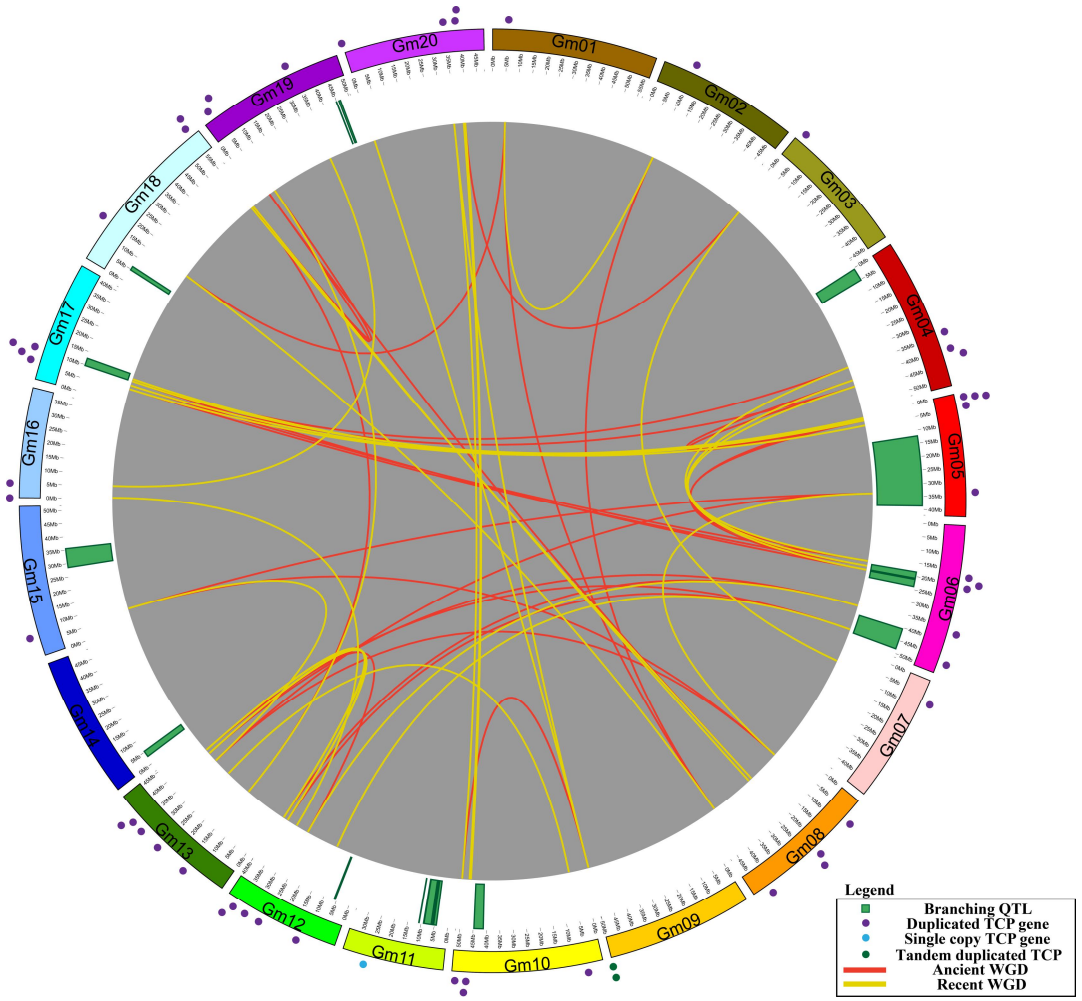


Figure 4-1. Chromosomal distribution and duplication relationship of TCP genes in *G. max*. The location of TCP genes is indicated as dot positioned at the outside of chromosome. The purple and green dot represented duplicated TCP gene by WGD and tandem duplication. The blue dot indicated single TCP gene which was not duplicated. The QTLs associated with branching are presented in first inner layer with green box. The whole genome duplication events are presented in second inner layer with distinguishable color. The red and yellow lines are represented duplication event which are duplicated by ancient whole genome duplication and recent WGD.

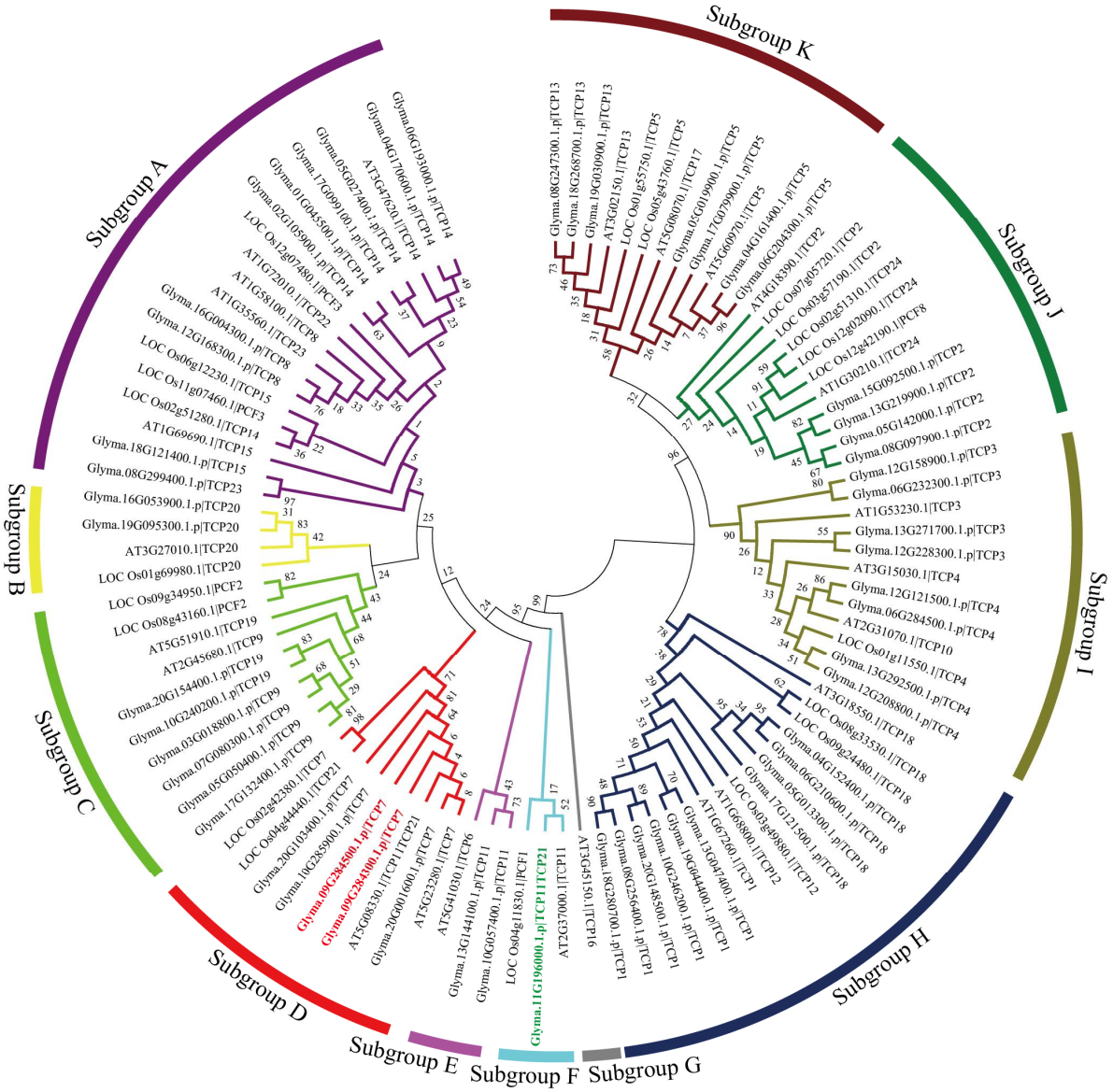


Figure 4-2. Phylogenetic relationships of TCP transcription factors from *Glycine max*, *Arabidopsis thaliana*, and *Oryza sativa*. The unrooted tree was constructed using MEGA 7.0 using NJ method with JTT model, complete deletion options and 1000 times bootstrap test. In total 11 subclades designated Group A to K were distinguished by color. The name of genes duplicated by tandem duplication and not duplicated were presented by red and green color.

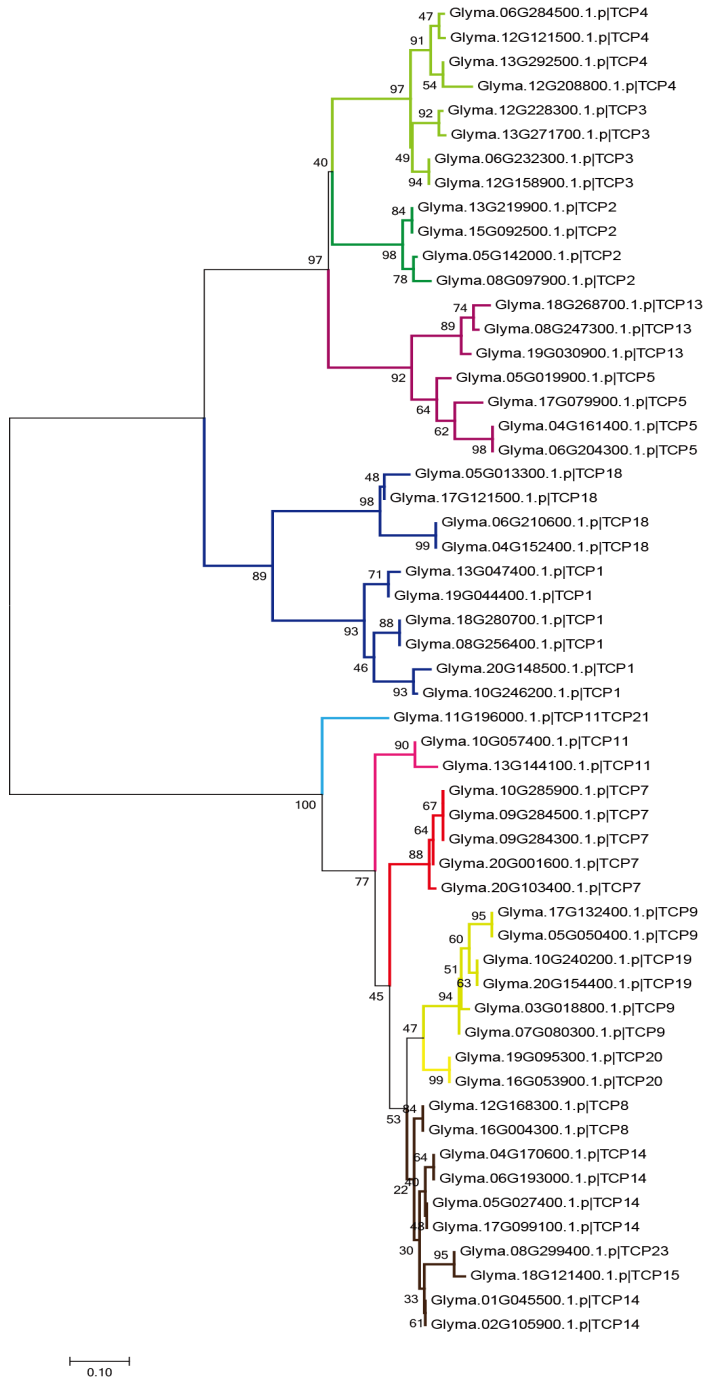


Figure 4-3. Phylogenetic tree of TCP domain. Phylogenetic tree of TCP domain was analyzed based on the motif region identified by InterProScan analysis.

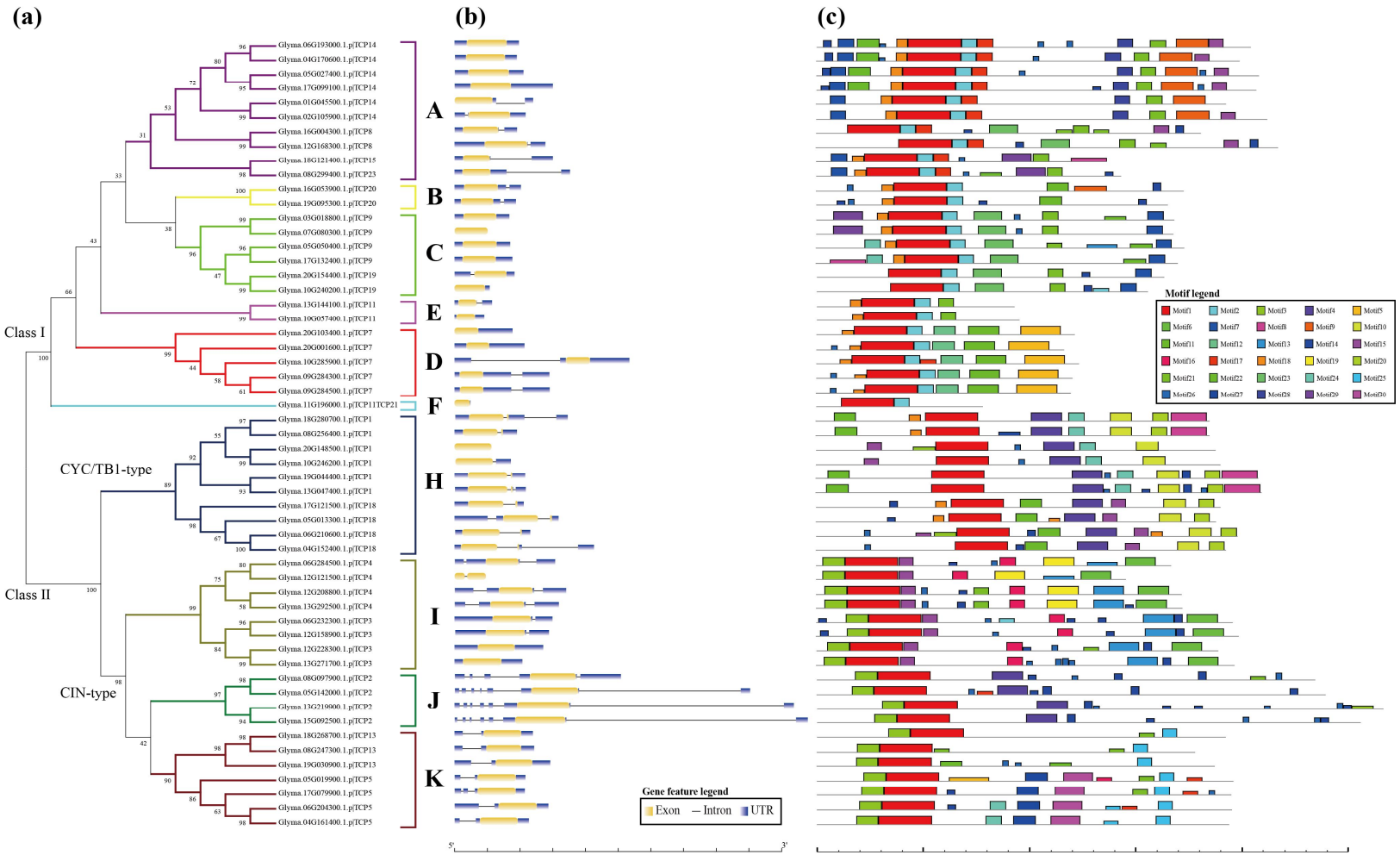


Figure 4-4. Phylogenetic analysis, gene structure and conserved motifs of TCP family in *Glycine max*. (a) The phylogenetic tree of *Glycine max* TCP transcription factors was constructed by NJ method in MEGA 7.0 with JTT model, complete deletion options and

1000 times bootstrap test. (b) Structure of *G. max* TCP genes were displayed. Exon, intron and UTR region were represented by blue box, yellow box and black line. (c) The conserved motif of TCP transcription factor proteins was analyzed using MEME program. The identified motifs distinguished by colors represented in motif legend.

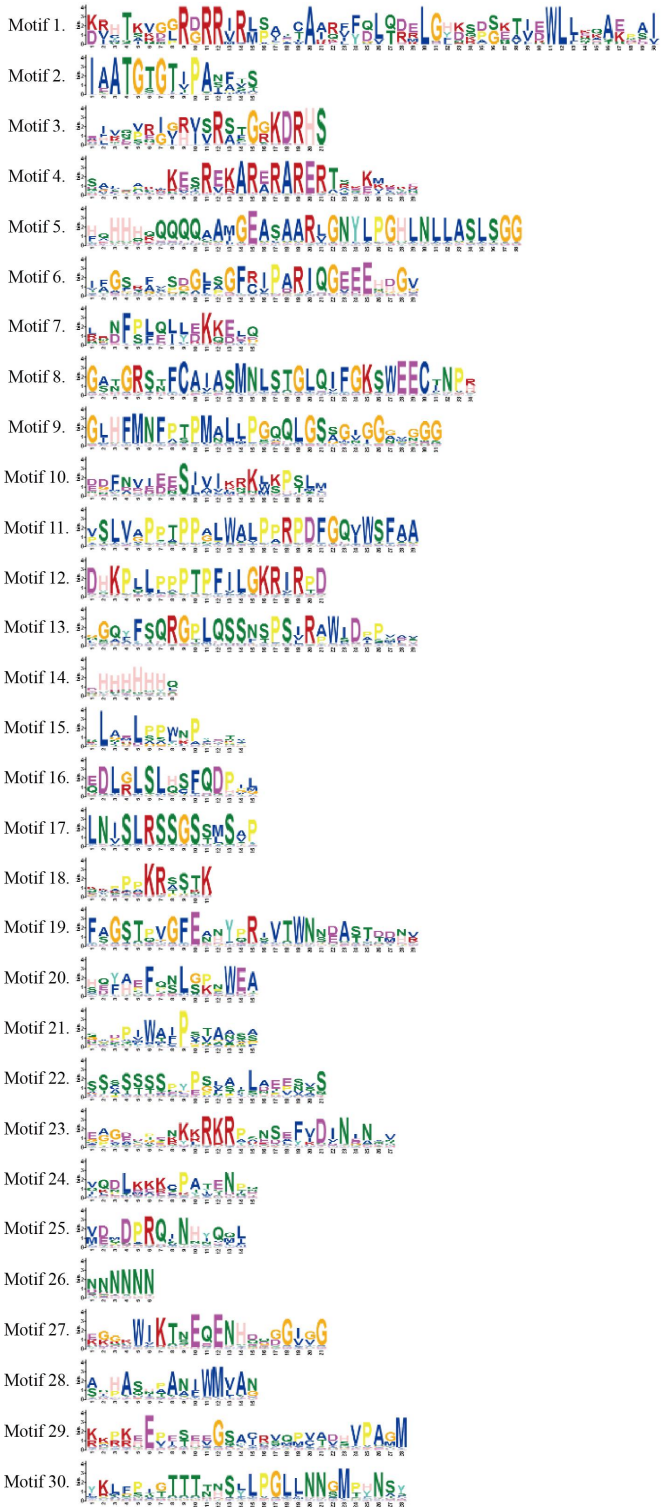


Figure 4-5. The consensus protein sequence of identified motifs.

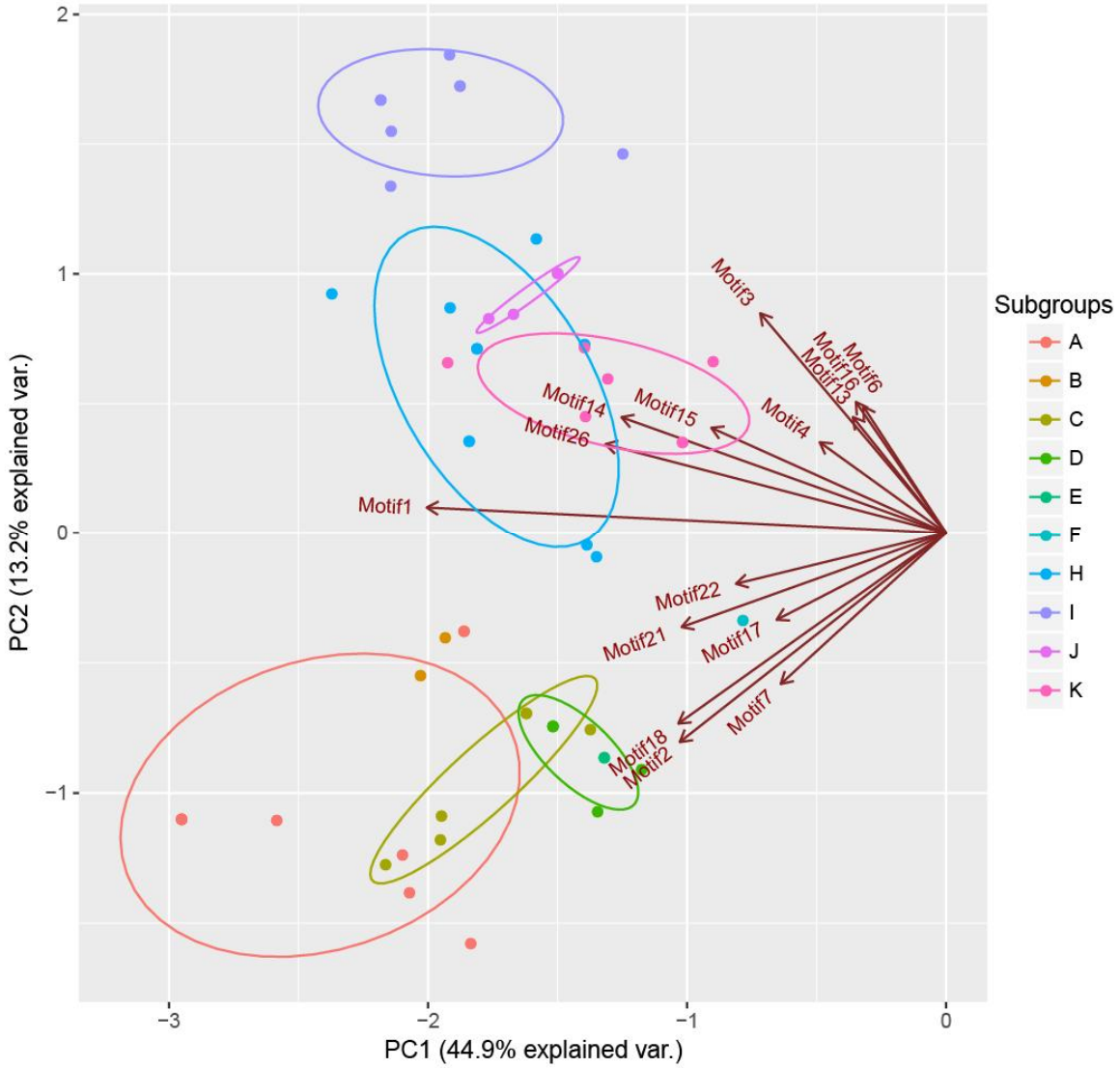


Figure 4-6. PCA analysis of motif composition in soybean TCP proteins.

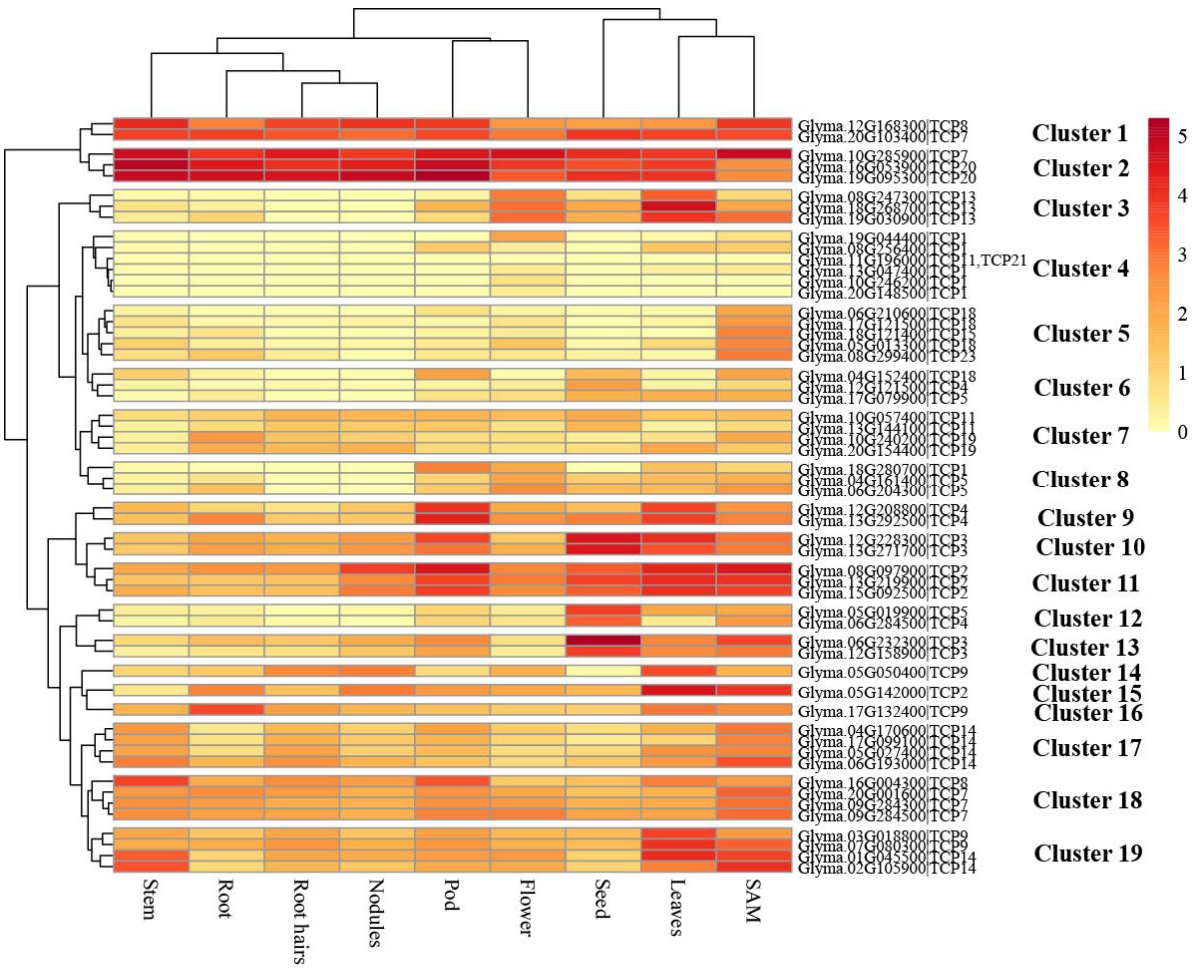


Figure 4-7. Expression patterns of *G. max* TCP transcription factor. The log2(FPKM+1) of TCP genes were quantified by RNA-seq of nine different tissues. A total 19 clusters designated by Cluster 1 to 19 were identified.

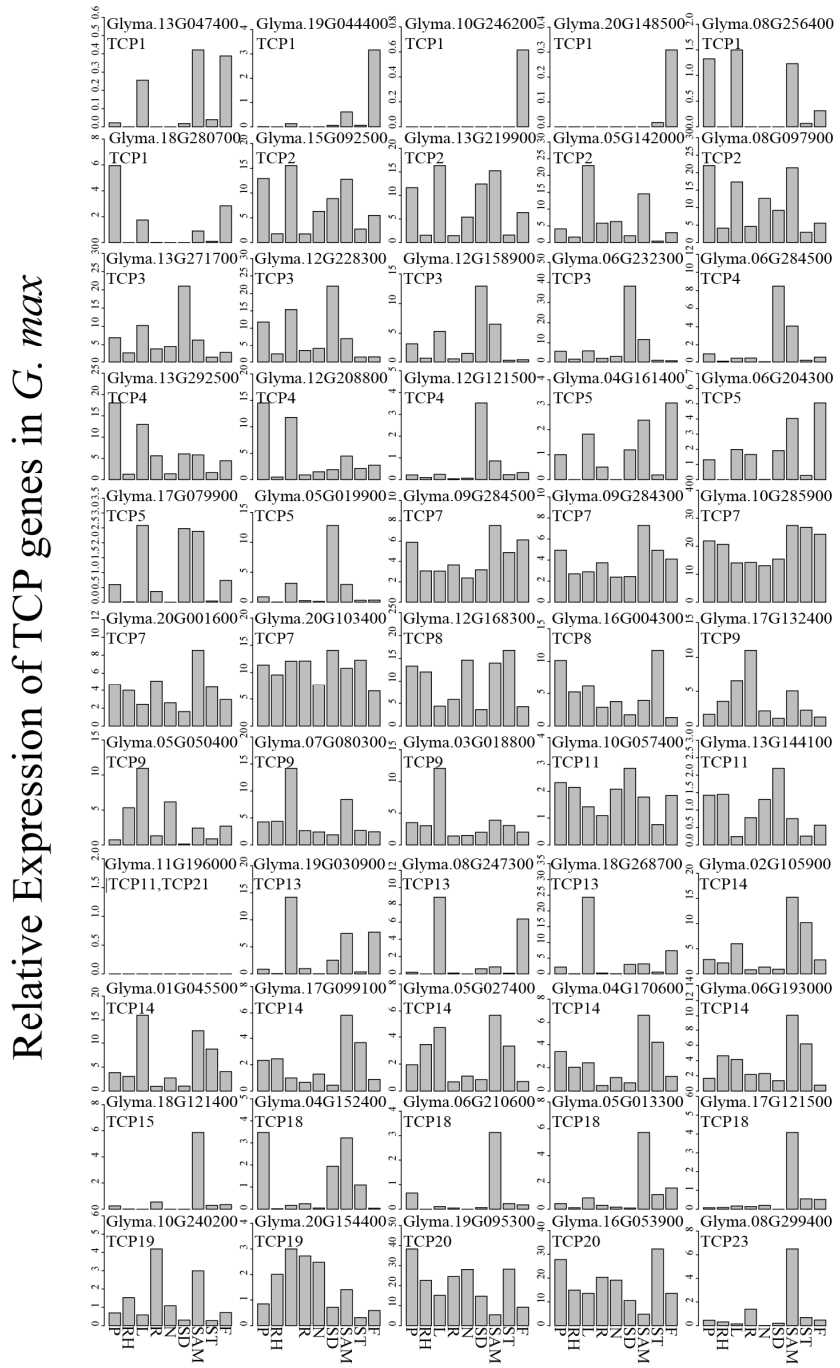


Figure 4-8. Expression profiles of the *G. max* TCP genes in nine different tissues. The order of x-axis was pod (P), root hairs (RH), leaf (L), root (R), nodules (N), seed (SD), shoot apical meristem (SAM), stem (ST), and flower (F).

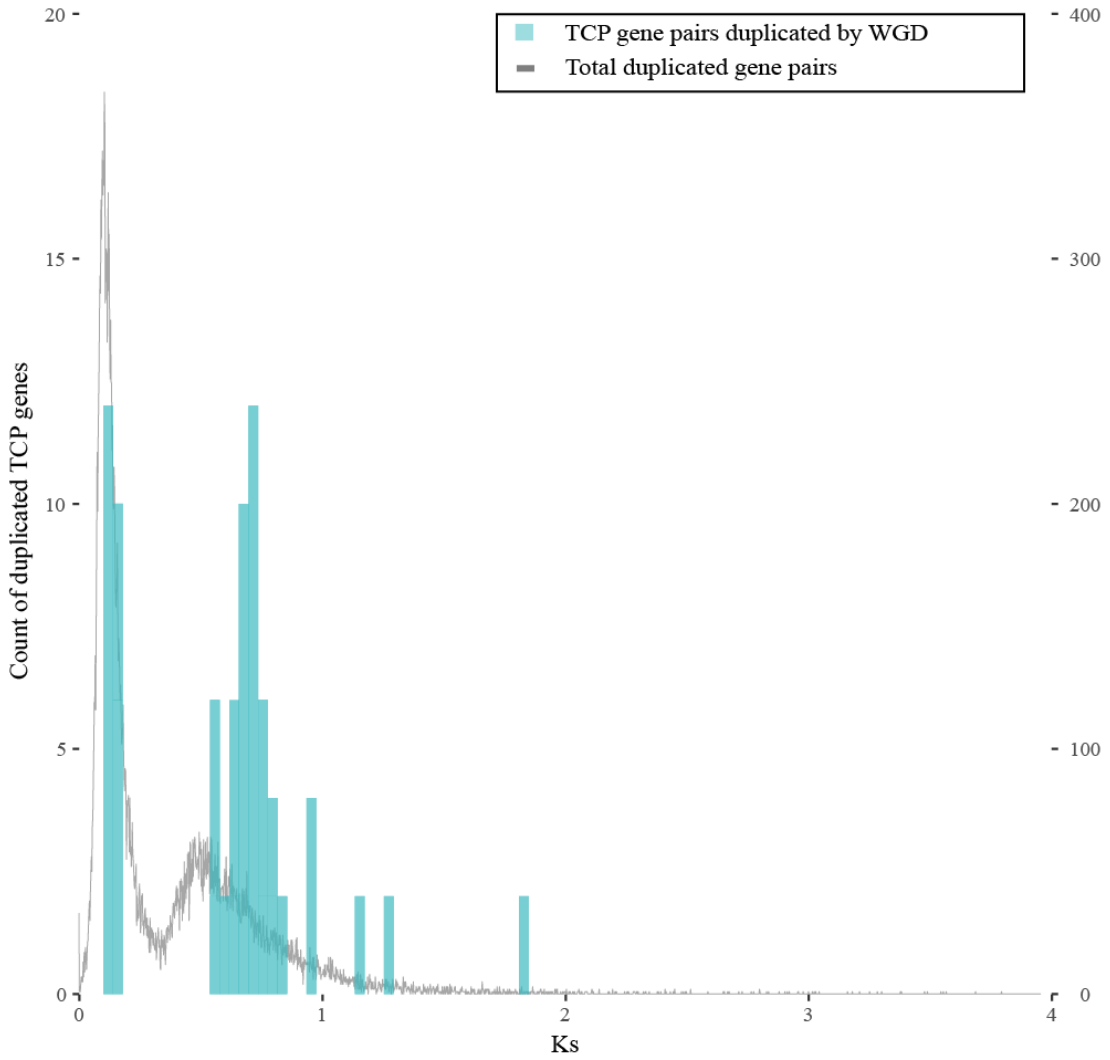


Figure 4-9. Ks distribution of TCP transcription factor genes duplicated by whole genome duplication in soybean. Frequency of soybean TCP genes was expressed as blue histogram. The Ks distribution of whole soybean genes was presented as grey line plot in background.

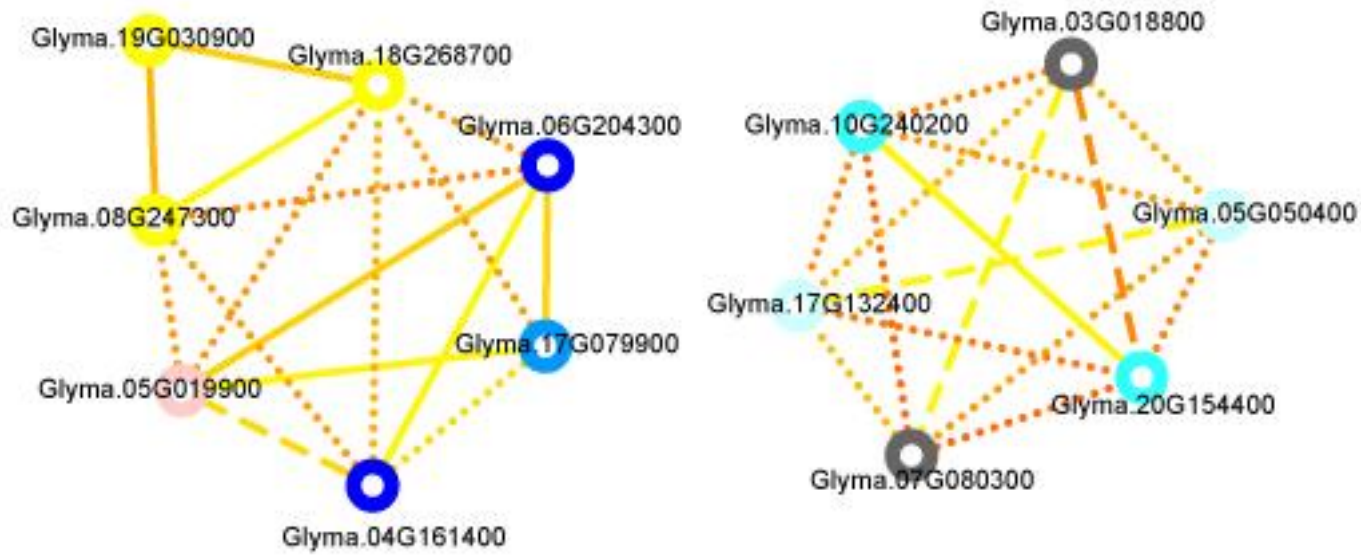


Figure 4-10. Networks of possible paralogous gene pair including 46 gene pairs which were irrelevant with synteny blocks.

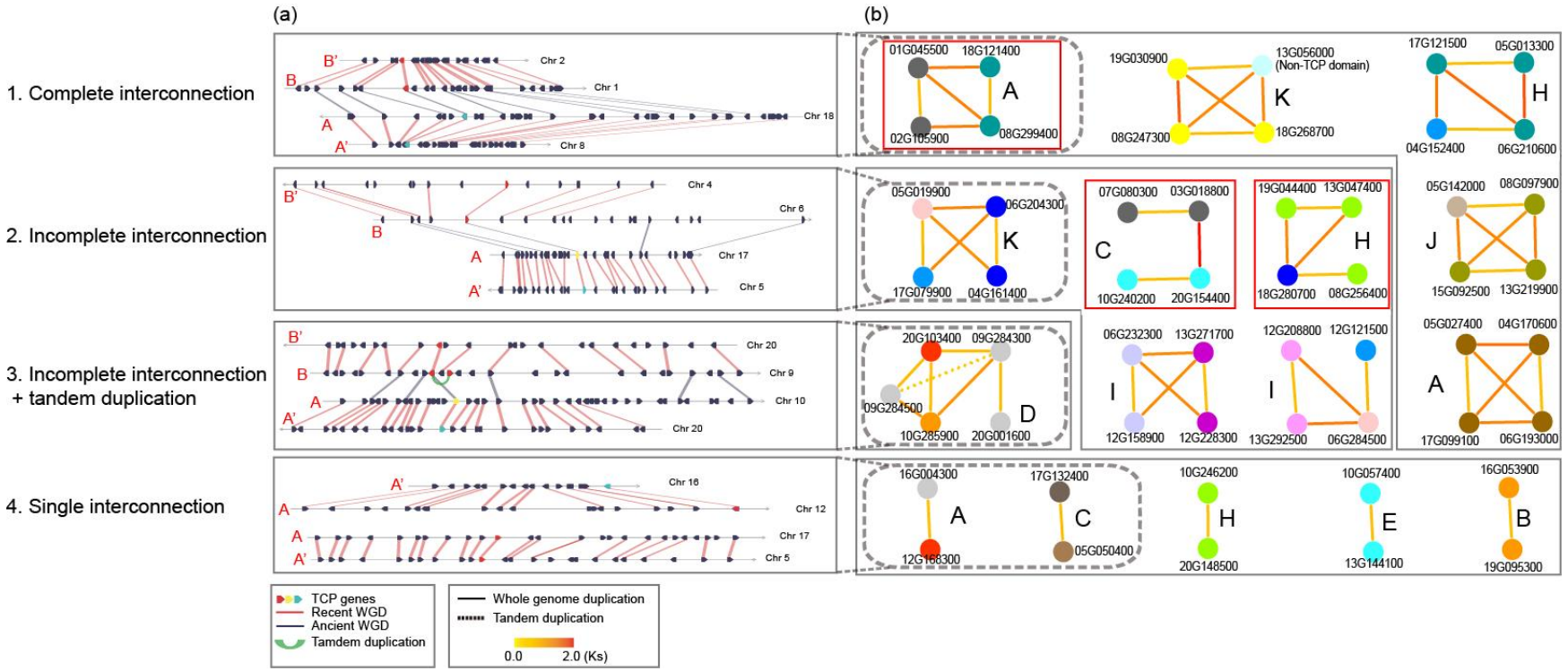


Figure 4-11. Soybean TCP gene duplication networks. The network was categorized as four categories designated one to four and presented in the most left side of figure. (a) Among the categories, possible gene duplication schemes of representatives were visualized. Ancient WGD were positioned in center of each figures and recent WGD were positioned at the upper and lower side of ancient duplication. TCP gene expression patterns were represented by different color of TCP genes. (b) All of soybean TCP gene duplication

relationships were visualized by Cytoscape and categorized. Each of genes were presented as node and the gene expression pattern clusters were presented as color of node. The modes of gene duplications were distinguished by types of line of edge. The solid line represented synteny relationship duplicated by WGD, otherwise dotted line indicated tandem duplicated relationship. The date of duplication was represented by colors of edges. Yellow line represented recent duplicates and dark orange (or red) line indicated ancient duplicates. Subgroup of phylogenetic tree were annotated as alphabet in adjacent space of each network. Red box represented examples of interconnections of duplicated gene pair which had structural divergence in protein sequence and different annotations.

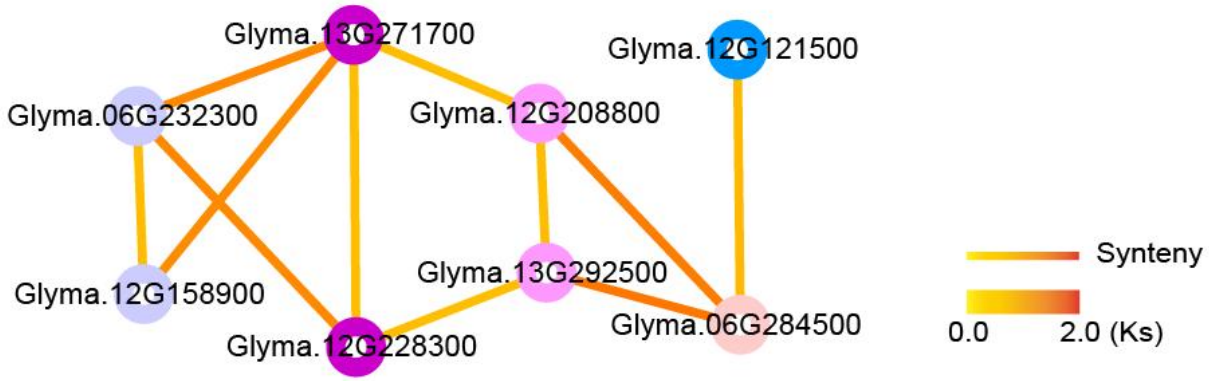


Figure 4-12. Suspicious syntenic relationships. The interconnections of four genes, Glyma.13G271700, Glyma.12G228300, Glyma.12G208800 and Glyma.13G292500, were duplicated by recent WGD. It was not fitted with the date of two rounds of WGD.

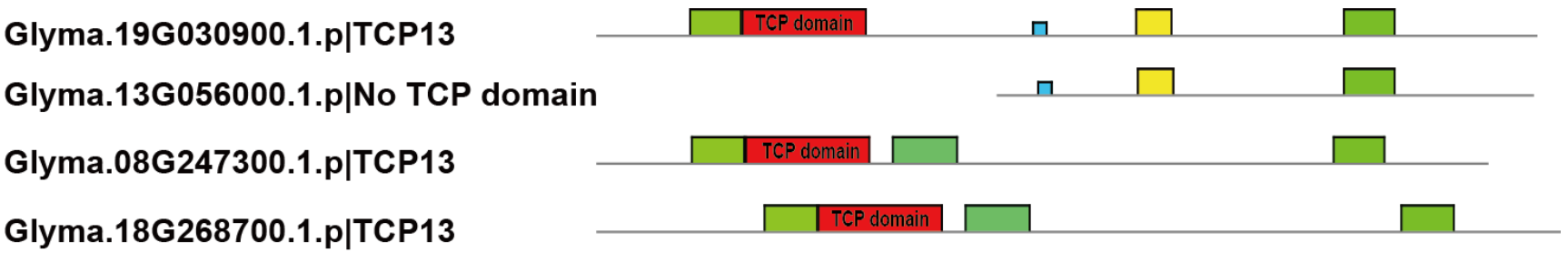


Figure 4-13. Protein sequence alignment of three soybean TCP13 and its paralogous gene which does not containing TCP domain.

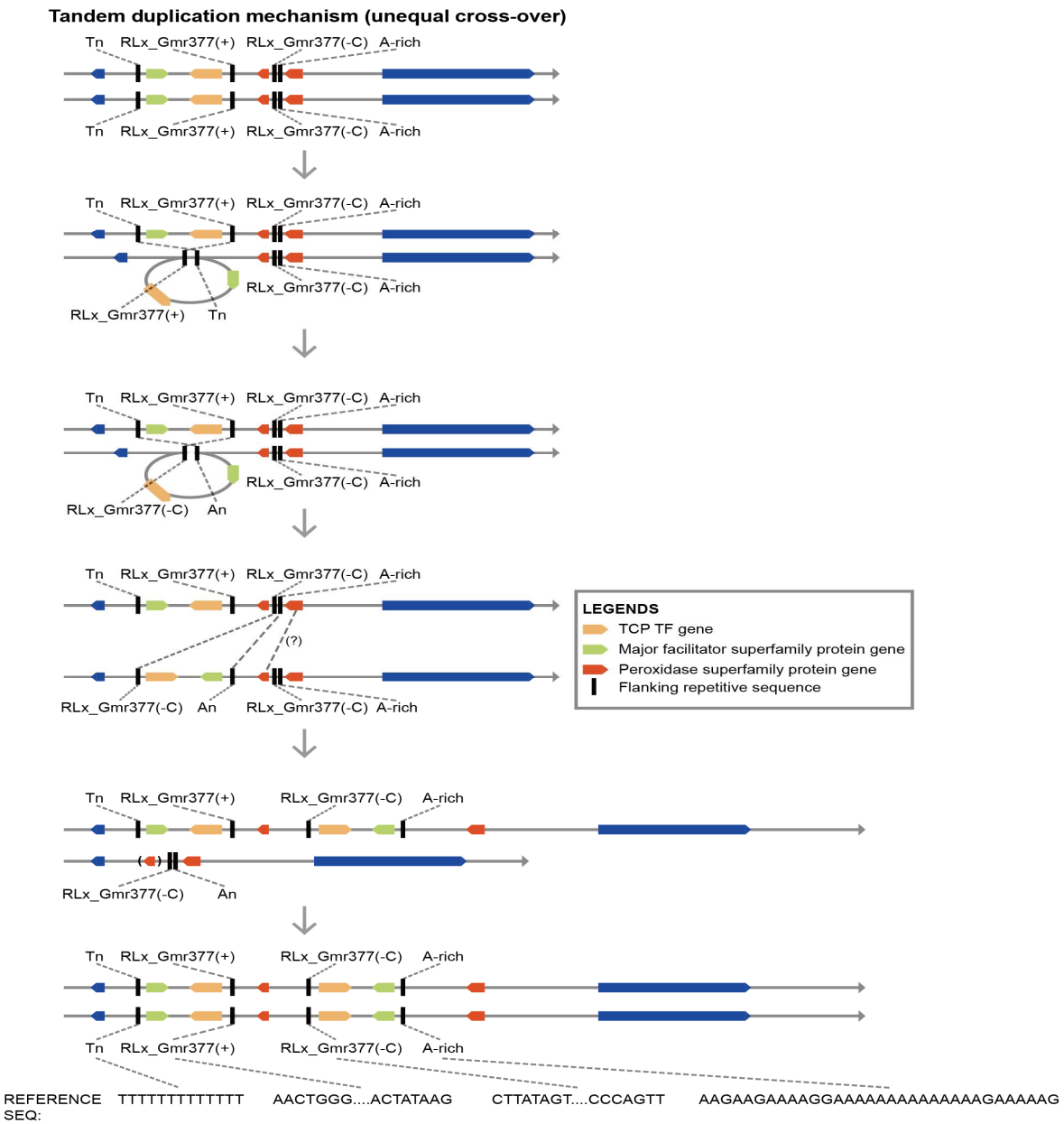


Figure 4-14. Mechanism of tandem duplication of TCP transcription factor in soybean. Putative scenario of inverted tandem duplication caused by unequal cross-over was presented. Genes and repetitive sequences were presented as pentagon with colors and black vertical lines. Direction and complementary sequence of repetitive sequence were presented by (+) or (-C).

CONCLUSION

Branch development affects yield components of pod and seed number per plant. However, due to the environmental influences, low number of genetic elements conferring branch development has been reported in soybean. To identify the genetic factors controlling branches, preliminarily, QTL association analysis based on the bi-parental population was conducted. A total four loci significantly associated with branch number were identified. Due to the high-resolution of genetic map, the loci were significantly narrowed down and candidate genes could be identified. As candidate genes, genes regulating branch development under auxin hormone signaling pathway in *Arabidopsis thaliana*, including *BRANCHED 1* (*BRC1*; encoding TCP transcription factor 18) and *REVOLUTA* (*REV*), were identified.

Using a set of 430 soybean core germplasms, associations analysis for genetic markers presented in the major QTL, *qBR6-1*, resulted six out of 45 markers were associated with branch numbers. Among these markers, one was anchored in the exon of *BRC1* gene and three were in intron of gene encoding transcription factor TFIIE alpha subunit. Comparison of gene expression between a set of NILs developed from a F_6 RHL for *qBR6-1*, resulted that the *BRC1* gene was significantly down regulated in NIL that exhibited more branches. SNPs at upstream and exon (missense variant) of *BRC1* gene were significantly

associated with branch numbers within an additional set of 59 soybean germplasms. Followed protein sequence alignment of orthologues of *BRC1* gene showed that changed amino acid residue was not conserved over orthologues. This suggested that the branch development is more likely to be controlled by gene expression of *BRC1*, not by functional change.

In RNA-seq analysis, global expression difference between the set of NILs was analyzed. The results indicated that the difference of branch numbers between the NILs was orchestrated by various molecular pathways including hormone, secondary metabolite and development pathway under genotype segregation into two different parental genotypes in *qBR6-1*.

Series of study showed that the *BRC1* gene encoding TCP transcription factor 18 played important role in branch development. To understand evolution mechanism of TCP TF family, comprehensive bioinformatics analysis was conducted. From this study, 55 TCP genes were identified and classified. Duplication and retention/divergence of soybean TCP genes were analyzed. As results, it was revealed that the soybean TCP genes expanded and functionally diverged along with two rounds of whole genome duplications and tandem duplication. Examples of functional retention/divergence which could be explained by gene balance hypothesis and neo-/sub-functionalization were

observed.

These findings will contribute to better understanding of soybean branch development and gene expansion/evolution mechanism, as well as breeding of high-yield soybean variety based on the more branch exhibition.

ABSTRACT IN KOREAN

콩의 분지는 수량 구성 요소인 주당 협 수와 주당 총립수에 영향을 미치는 요소이다. 기존의 보고에 의하면, 콩의 분지 수는 광질, 토양 수분, 재식 밀도 등 다양한 환경 요소에 의해 크게 영향을 받는 것으로 알려져 있다. 때문에, 다른 형질들과 비교하여 상대적으로 적은 양적 형질 유전자 좌 (QTL)가 보고되어왔다. 이번 연구에서는 콩의 분지 발생과 분지 수를 결정하는 유전적 원인을 찾고자 교배 집단을 활용하여 양적 형질 유전자 좌를 조사하였다. 그 결과 총 네 개의 유전자 좌가 콩의 분지 형성과 연관되어 있음을 밝힐 수 있었으며, 새로이 작성된 고밀도 유전자지도를 바탕으로 네 개의 유전자 좌를 가능한 한 많이 좁힐 수 있었다. 그 결과 아기장대에서 식물생장호르몬 중 하나인 옥신의 영향을 받아 분지 발생에 영향을 주는 *BRANCHED 1* (*BRC1*) 유전자와 *REVOLUTA* (*REV*) 유전자 등을 유전자좌에서 동정하였다.

이 연구를 통해 찾은 유전자 좌 중 효과가 큰 것으로 예측된 유전자 좌 (*qBR6-1*) 에 대해, 세계 각국에서 수집한 430개의 콩 유전자원을 활용하여 연관 분석을 수행하였고, 그 결과 해당 유전자 좌 내에 존재하는 45개의 SNP 분자표지

중 여섯 개의 분자 표지에 대해 표현형과의 연관성이 분석되었다. 그 중 하나의 분자 표지는 *BRC1* 유전자의 exon 영역에 위치하였고, 세개의 분자 표지는 인접한 전자조절인자 TFIIE의 alpha subunit을 코딩하는 유전자의 intron 영역에 위치하였으며, 나머지 두개는 유전자 사이의 영역에 위치하였다. 이와 별개로 *qBR6-1* 유전자 좌에 대한 F_6 residual heterozygous line으로부터 분리된 근동질 계통을 육성하였으며, 유전적 배경 분석을 통해 해당 근동질 계통이 99.9% 이상 동형 접합성을 보이는 것으로 확인되었다. 이를 이용하여 앞서 좁혀진 두개의 후보유전자에 대한 발현 량을 살펴본 결과, *BRC1* 유전자의 발현 량이 분지가 많이 발생하는 근동질 계통에서 유의하게 줄어드는 것을 확인하였다. 동 유전자의 발현 량 차이의 유전적 원인을 찾고자 미국 농무부 (USDA) 산하 유전자원 및 정보 네트워크 (GRIN)에서 분양 받은 콩 품종 59개를 사용하여 비동의 (non-synonymous) 단일염기변이와 프로모터 지역에 위치한 단일염기변이에 대한 연관성을 분석하였고, 결과 비동의 단일 염기 변이와 프로모터 (1kbp upstream) 지역에 위치한 단일 염기 변이가 콩의 분지 수와 연관되어 있음을 확인할 수 있었다. 하지만, 비동의 단일 염기 변이로부터 유래되는 아미노산 서열은 상동유전자들의 아미노

산 서열에서 보존되어 있지 않았으며, 이는 해당 아미노산이 단백질의 기능에 중요한 영향을 미치지 않는 것임을 간접적으로 보여주었다. 이를 통해 *BRC1* 유전자의 발현 량 차이로 콩의 분지 수 발생이 조절될 가능성이 농후함을 확인하였으며, 이는 동일 유전형 콩의 다양한 재식밀도에서의 서로 다른 수의 분지 수를 발생시키는 현상에 대한 설명도 가능한 것임을 밝힐 수 있었다.

앞서 개발된 근동질 계통을 활용하여 전장 유전체 내에서의 전장 전사체의 발현 량 변화를 살펴본 결과, 근동질체의 분지 수 차이가 식물생장호르몬, 이차대사산물 그리고 발달과 관련된 다양한 생물학적 경로의 향연에 기인하는 것임을 알 수 있게 되었다.

일련의 연구를 통해 콩의 분지 발생에 대한 유망한 후보 유전자인 *BRC1* 유전자는 TEOSINTE BRANCHED 1/ CYCLODEA/PROLIFERATING CELL FACTOR (TCP) 전사인자 18을 인코딩하는 것으로 알려져 있다. TCP 전사인자는 식물체에만 존재하며 식물의 기관 발달에 중요한 영향을 끼치는 것으로 알려져 있다. 마지막 장에서는 콩의 TCP 전사인자들에 대한 동정, 분류, 복제 양상, 발현 양상 등을 비교분석하였으며, 이 결과를 통합하여 콩의 TCP 전사 인자의 진화와

기능적 분기를 살펴보았다. 그 결과 콩의 TCP 전사인자는 두 번의 전장 유전체 복제 현상과 tandem 복제 현상에 의해 복제되었으며, 특히 전장 유전체 복제과정에서 발현 양상이 분화되어 기능적 차이를 보이게 되었음을 알 수 있었다.

일련의 연구를 통해 콩의 분지 형성에 대해 조금 더 이해를 할 수 있었으며, 이를 활용하여 다수성 콩 품종 개발에 이용할 수 있을 것으로 기대된다. 또한 콩의 TCP 전사 인자에 대한 복제와 분화에 관한 연구 결과를 바탕으로 콩의 유전자 복제/진화와 관련된 연구 및 다른 식물에서의 TCP 전사인자에 대한 연구에 밑거름이 될 것으로 사료된다.

감사의 글

학문에 뜻을 두고 시작했던 박사과정 생활을 이제 마무리 지으려고 합니다. 쉽지 않았기에, 그래서 포기하고도 싶었던 순간도 수 없이 많았던 것 같습니다. 하지만 끝까지 응원해주신 분들 덕분에 무사히 여기까지 올 수 있었습니다.

먼저, 학문적으로 물심양면 지원해 주시고, 지도해 주신 스승님, 지도교수 이석하 교수님께 깊이 감사드립니다. 또한 이 논문에 대해 아낌없는 지도와 조언을 해주신 작물생명과학 전공 교수님들께도 감사드립니다. 많이 부족한 저에게 많은 가르침을 주신 김문영 박사님, 하정민 박사님, 이영호 박사님, 박금룡 박사님께 감사의 마음을 전합니다. 수년간 함께 지내며 곁에서 힘이 되어준 실험실 식구들 현주 누나, 윤민영 박사님, 태영, 재언, 은수, Xuefei, Areena에게도 감사의 말을 전합니다. 도움이 필요할 땐 언제든 도와주었던 석사 후배들, 준희, 은비, 하경, 은샘, John에게 고마움을 전합니다. 아낌없는 조언을 주신 선배님들, 길현이형, 양제형, 경도형, 성우 선배에게도 감사드립니다. 두번째 챕터와 관련된 연구를 진행할 수 있도록 지원해주신 생공연의 정순천 박사님, 단국대학교의 강성택 교수님, 농진청의 문준경 박사님과 최만수 박사님께 감사드립니다.

이제 곧 20년 지기를 바라보는 강릉고등학교 동창들, 대용, 백환, 연달, 준기에게 힘들 때마다 곁에서 응원해 주고 함께 해줘서 고맙다는 말 전하고 싶습니다. 부족한 사위를 응원해주신 장인, 장모님, 저에 대한 지분을 4분의 1 가지고 계신다고 말씀하시는 할아버지, 할머니, 항상 응원해주신 이모님들과 고모님들, 그리고 친인척들에게 이 글을 빌어 감사의 말씀

을 전합니다.

마지막으로, 낳아주시고 길러주신, 그러고도 모자라 여태껏 물심양면 지원해주신 사랑하는 부모님, 하나밖에 없는 내 소중한 동생 상희, 부족한 나한테 시집와서 고생하고 있는 사랑하는 아내 인혜 그리고 눈에 넣어도 아프지 않을 우리 연우에게 지난 6년간의 결실을 바칩니다.

앞으로 더 좋은 연구자가 될 수 있도록 더욱 더 노력하겠습니다. 감사합니다.