A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Identification of Genetic Factors Controlling Multiple-flower per node in Pepper (*Capsicum* spp.)

고추의 복화방 형질을 조절하는
유전인자 탐색

AUGUST, 2019

YOUNGIN KIM

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Identification of Genetic Factors Controlling Multiple-flower per node in Pepper (*Capsicum* spp.)

UNDER THE DIRECTION OF DR. BYOUNG-CHEORL KANG
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF SEOUL NATIONAL UNIVERSITY

BY
YOUNGIN KIM

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

AUGUST, 2019

APPROVED AS A QUALIFIED DISSERTATION OF YOUNGIN KIM
FOR THE DEGREE OF MASTER OF SCIENCE
BY THE COMMITTEE MEMBERS

CHAIRMAN _____
Jin Hoe Huh, Ph.D.

VICE-CHAIRMAN _____
Byoung-Cheorl Kang, Ph.D.

MEMBER _____
Doil Choi, Ph.D.

Identification of Genetic Factors Controlling Multiple-flower per node in Pepper (*Capsicum* spp.)

YOUNGIN KIM

DEPARTMENT OF PLANT SCIENCE
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSRACT

Flowering is enormously important in all crops, serving as the foundation for yield and increased profits. *Capsicum annuum* has a sympodial shoot structure with a solitary flower. In contrast, *C. chinense* produce multiple-flower per node. *C. annuum* is the most widely cultivated species which account for 80% of world pepper production. Therefore, identifying genes controlling multiple flowers and transferring the multiple flowering trait from *C. chinense* to *C. annuum* may be potentially useful to increase fruit yield. In this study, we performed two experiments to identify the genetic factors controlling the multiple-flower per node character in *Capsicum*. To find the quantitative trait loci (QTL) controlling multiple flowering, 85 recombinant inbred lines (RILs) between *C. annuum* 'TF68' and *C. chinense* 'Habanero' were used. Average flower numbers of the first to sixth nodes with three

replicates of each line was collected. A high density molecular genetic map was constructed by using genotyping by sequencing (GBS) technique. A total of 10,851 SNP markers on 12 chromosomes were converted to bin markers to construct a high-density linkage map. The map covered a total length of 1,713 cM with a mean bin marker distance of 0.96 cM. QTL analysis identified four novel QTLs on chromosome 1, 2, 7 and 11 for multiple-flower per node trait, accounting for 65% total phenotypic variation. To validate and clarify the detected QTLs, genome-wide association study was performed. A total of 276 *C. annuum*-clade accessions, including 98 *C. annuum*, 66 *C. chinense*, 67 *C. frutescens* and 45 *Capsicum* spp. were used. Genotyping was performed with GBS method and after filtering the SNPs, a total of 156,589 highly reliable SNPs were selected for association study. Genome-wide association analysis revealed that a total of 28 QTL regions were significantly associated with multiple-flower per node. Among the QTLs, three were collocated with the QTL region detected in the biparental population. In the QTL regions, we identified five candidate genes involved in the development of shoot and flower meristem for controlling multiple-flower per node in pepper. These results will contribute to understand multiple-flower per node character in *Capsicum* and will be useful for developing high yield cultivars.

Keywords: Pepper, Flower production, Yield, Quantitative trait locus (QTL), Genome-wide association study (GWAS), Genotyping-by-sequencing (GBS)

Student number: 2017-23343

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BLAST	Basic local alignment search tool
CAPS	Cleaved Amplified Polymorphic Sequences
cM	centi Morgan (the unit of genetic distance)
CTAB	Cetyl trimethylammonium bromide
SCAR	Sequence-characterized amplified region
GAPIT	Genome Association and Prediction Integrated Tool
GATK	GenomeAnalysisTK
GBS	Genotyping-by-sequencing
GWAS	Genome wide association study
HRM	High resolution melting analysis
Kasp	Kompetitive Allele Specific PCR
LG	Linkage group
LOD	Logarithm of the odds
MAF	Minor allele frequency
MF	Multiple flower
NGS	Next generation sequencing
NIL	Near isogenic line

PCA	Principal component analysis
Q-Score	Phred quality score
QTL	Quantitative trait locus
RIL	Recombinant inbred line
RT-PCR	Reverse transcription PCR
SAM	Shoot apical meristem
SCAR	Sequence-characterized amplified region
SNP	Single nucleotide polymorphism

INTRODUCTION

Pepper (*Capsicum* spp.) is considered as one of the most agriculturally and economically important vegetable crops in the world. Fresh and dried pepper production was estimated to be 53 million tons in the top 20 pepper-producing countries according to the United Nations Food and Agriculture Organization (FAO) statistics in 2017. Pepper is esteemed in various cuisines due to capsaicinoids, various nutritional benefits such as pro-vitamin A (carotene), E (a-tocopherol), and vitamin C (ascorbic acid) (Penella and Calatayud 2018). Pepper is also grown for use as an ornament, for chemical industries, and for their pain-killing and medicinal properties (Fraenkel et al., 2004).

Yield, stress resistance and fruit quality are three most important traits in crop breeding. Among the traits, improvement of crop yield is an ever-most important objective. Since the total number of flowers determines the total fruit yield per plant, flower number per node (or per inflorescence) is an essential element for yield increase. To improve crop yield, genetic factors determining flower number per inflorescence have been studied in cereal and horticultural crops such as rice and tomato (Bai et al., 2012; Soyk et al., 2017a; Sasaki et al., 2017).

C. annuum is the most widely cultivated in the world among five domesticated *Capsicum* species including *C. annuum*, *C. chinense*, *C. frutescens*, *C. baccatum*, and *C. pubescens*. *C. annuum* has a sympodial shoot structure with a solitary flower

whereas *C. chinense* produces multiple flowers per node. Therefore, identification of genes controlling multiple-flower per node and transferring the genes from *C. chinense* to *C. annuum* may be potentially useful to increase yield. Tanksley and Oliva (1984) found that multiple-flower per node trait in *Capsicum* are quantitatively inherited. However, genetic mechanisms that control multiple-flower per node in pepper is still poorly understood.

Identification of genetic factors underlying quantitative trait loci is commonly performed with biparental mapping populations. Various populations have been used for QTL mapping in crop plants such as F2, backcross, recombinant inbreed lines (RIL), and diverse inbred populations (Szalma et al., 2007). With recent advances in high-throughput genotyping, high-density genetic linkage map is used for precise detection and characterization of QTLs. Even high-density genetic maps are used for QTL study, however, the conventional QTL mapping with biparental population has limitations to identify underlying genes due to low mapping resolution. Furthermore, only two allelic variations could be analyzed in biparental populations when there are many other alleles available in natural populations.

The limitations of traditional QTL analysis can be supplemented using genome-wide association studies (GWAS), which test associations between nucleotide polymorphisms and phenotypic variations using natural populations (Yano et al., 2016). GWAS could generates false positive associations between the phenotype and unlinked markers due to strong population structure. Although many statistically advanced models have been developed, spurious associations peak arising from

population structure in crops cannot be easily controlled. Therefore, the combination of GWAS and QTL analyses can compensate the defect of each approach, allowing the identification of genes controlling agronomically useful QTL traits.

To verifying detected QTLs and fine-mapping the candidate region, near-isogenic lines (NIL) can be used. Fine-mapping QTL with NIL population involves creating a set of lines in which each NIL carries only a small trait linked region of the donor parent genome. NIL population was successfully used to fine mapping several QTLs linked with total spikelet number per panicle in rice (*Oryza sativa*) (Sasaki et al., 2017) and major QTL for seed coat color in *Brassica rapa* (Zhao et al., 2019). In addition, NILs containing major QTL can be an intermediate set of material itself that breeders can utilize further for improved varieties.

In this study, we performed QTL analysis for multiple-flower per node in one interspecific RIL population of *Capsicum*. A high-density genetic map was used to ensure an accurate linkage analysis. In addition, a total of 276 *C. annuum*-clade accessions, including *C. annuum*, *C. chinense*, *C. frutescens* and *Capsicum* spp. were genotyped by genotype-by-sequencing (GBS) and analyzed for GWAS. By comparing the physical locations of the QTLs from traditional QTL mapping and GWAS, we identified three common QTLs in both analysis, and three candidate genes that are known to control inflorescence meristem development. These results will be useful to understand multiple-flower per node development in pepper and breeding pepper cultivars with high yield potential.

LITERATURE REVIEW

1. Multiple-flower per node trait in *Capsicum*

Pepper (*Capsicum* spp.) has been an important vegetable crop with a food additive for their unique pigment, aroma and pungency. Among various traits of plants, yield of fruit is their most important trait for fruit vegetable crops and flower production is the key factor. *C. annuum* sets a single flower at each node, whereas *C. chinense* normally sets two to four flowers at each node.

Subramanya (1983), Watson et al. (1986) and Shuh et al. (1990) revealed that three to seven genetic factors determine the multiple-flower per node in *C. chinense* by genetic inheritance study. Tanksley and Oliva (1984) found that a minimum of five independently segregating chromosomal regions control the difference in multiple-flower per node trait through mapping trait with *C. annuum* and *C. chinense*. However, only twelve markers were used for study which was too low density to detect specific location of QTL. Recently, multiple-flower per node trait was analyzed with high density molecular genetic linkage map. Zhu et al. (2019) identified three QTLs on chromosome 2, 7 and 10 that are associated with the multiple-flower per node trait with high density genetic map using 150 *C. annuum* and *C. chinense* interspecific F_2 population. However, detailed QTL locations and candidate genes information were not shown in this study.

2. Inflorescence development in Solanaceous species

In Solanacease family, flowering marks the end of main shoot growth, and vegetative aerial growth is renewed from axillary meristems in a perennial growth system known as “sympodial” (Lippman et al., 2008). During vegetative growth in pepper, stems and leaves that are arranged in an alternate spiral pattern are derived from the growth of dome-shaped groups of pluripotent cells called shoot apical meristem (SAM). SAM first produce leaves, and upon flowering induction vegetative meristems gradually mature to a reproductive state, and they produce transition meristems that transition to floral meristems, which produce flowers.

Depending on variations in the activity of meristem development, flower number per inflorescence could be different. Since flower production is important to increase yield, continual meristem development and floral meristem formation from previous phase has been widely studied to manipulate crop yield in tomato. Tomato (*Solanum lycopersicum*) generates a few-flowered inflorescence organized in a zigzag branch, but there are some mutants called *compound inflorescence* (*s*), *falsiflora* (*fa*) and *anantha* (*an*) that bear highly branched inflorescences that produce more flower per inflorescence (Lippman et al., 2008). Pepper, on the other hand, mutation of *Capsicum annuum S* (*CaS*) ortholog of *S* in tomato shows stem fasciation, lack of flowers and leafy appearance (Cohen et al., 2014). In case of *ca-anantha* (*Ca-an*) mutant in pepper, single-flower inflorescence of pepper convert to a compound inflorescence but shows much less organ identity that fail to initiate normal flowers (Lippman et al., 2008). This may be due to the fact that even pepper and tomato

shows close phylogenetic relationship, several architectural and inflorescence development pattern is so different.

3. High-throughput genotyping by sequencing

The rapidly developing sequencing technology provides a new way of genotyping. Next-generation sequencing (NGS) technology makes it possible to detect genome-widely located single-nucleotide polymorphism (SNP) that can be used for genotyping individual plants. Genotyping-by-sequencing (GBS) with NGS platform is a representative method of cost-effective genome-wide genotyping technology. It provides thousands to millions of SNPs that generate high resolution genetic map.

Restriction enzymes is used to reduce the complexity of genome on GBS library preparation. Choice of enzymes used in GBS library construction depends on several elements including how many markers are needed, the desiring number of multiplexing samples, and whether the enrichment of genic SNP is preferred. Four cutter enzymes can produce a large amount of small fragments on the other hands, rare cutting restriction endonucleases like six cutters with methylation sensitivity can reduce complexity by targeting fewer sites (Pootakham et al., 2016). Precedent studies suggested to use several combinations of common cutter and methylation sensitive restriction enzyme sets like *PstI* and *MseI* or *EcoRI* and *MspI* (Poland et al., 2012; Truong et al., 2012; Sonah et al., 2013; Han et al., 2018).

Recently, GBS is widely used for high density QTL mapping and Genome-wide association study (GWAS). It has been successfully applied to the identification of candidate genes controlling quantitative traits by QTL mapping or GWAS in plant including soybean (*Glycine max*), grape (*Vitis vinifera*), maize (*Zea mays*) and melon (*Cucumis melo*) (Navarro et al., 2017; Guo et al., 2019; Su et al., 2017; Pereira et al., 2018).

4. Quantitative trait analysis

Most agricultural traits are controlled by quantitative trait loci (QTL). During the last few decades, many methods was developed to identify quantitative inherited traits in plants. With several assembled pepper reference genomes, numerous SNP markers can be simply developed by aligning sequencing data to the reference genomes. Furthermore, according to several recent studies, cost efficient low-coverage sequencing could be utilized in complicate trait mapping or association study and find candidate genes (Su et al., 2017; Han et al., 2018; Wang et al., 2018; Zhu et al., 2019).

Identification of genetic factors underlying agronomic quantitative trait loci (QTL) is commonly performed with biparental mapping populations. F_2 population has a big advantage because it's easy to develop. Recombinant inbred lines (RILs) can genotyping once and phenotype multiple individuals in multiple environments. Advanced backcross populations like near isogenic lines (NILs), and double haploid lines (DHs) can also serve as mapping material for QTL analysis. The biparental

mapping population has limitations, however, due to restricted recombination event occurred and limited genetic diversity between the parents. Two allelic variations from the parental lines could be analyzed in a biparental population while several alleles occurring in other germplasms are missed.

The limitations of conventional QTL analysis can be overcome using genome-wide association studies (GWAS), which test a large number of genetic variants for association (Shaffer et al, 2012). GWAS can provide much higher resolution than QTL mapping, often to the gene level. GWAS also have weakness that can generates false positive associations between the phenotype and unlinked markers. Although use statistically advanced models, spurious associations peak arising from population structure cannot be easily controlled (Yano et al., 2016). Therefore, the combination of GWAS and QTL analyses can compensate the defect of each approach, allowing the identification of QTL controlling agronomically useful traits. He et al. (2017) successfully mapped branch number in *Brassica napus,* Zhao et al. (2018) reveal the genetic control of cadminum accumulation in maize (*Zea mays*) leaf, and Han et al. (2018) identified genetic factors controlling capsaicinoid content in *Capsicum* through QTL mapping and GWAS combined approaches.

MATERIALS AND METHODS

Plant materials

An intraspecific population of 85 RILs ($F_{10:12}$) was derived from a cross between *C. annuum* accession 'TF68' and *C. chinense* accession 'Habanero' (Haba) by single-seed descent (Kang et al., 2001; Kim et al., 2010). This population were referred to as 'TH' RILs following their parental names. 'TF68' sets a single flower at each node on the other hands, whereas the male parent, 'Haba', is normally sets two to four flowers at each node (**Figure 1**). Three replicates for each line of 'TH' RILs were grown in the green house of FarmHannong in Anseong, Republic of Korea (2017). Three plants were grown for each line in soil.

A CSHL (Cold Spring Harbor Laboratory) *Capsicum* core collections was used for GWAS. To reduce population structure *C. baccatum* and *C. pubescens* were excluded in this study. A total of 276 accessions including 98 of *C. annuum*, 66 of *C. chinense*, 67 of *C. frutescens* and 45 of *Capsicum* spp. were used for multiple-flower per node trait study. At least four plants of each accession were grown in upland farm, riverhead field or greenhouse of Cold Spring Harbor Laboratory in New York, United States of America (2015). (This experiment was done in Zacchary Lippman's lab, and all the CSHL population multiple-flower per node phenotype data related to GWAS analysis was used under permission)

Figure 1. Comparison of the parental lines. *Capsicum annuum* 'TF68' bears single flower per node (A), *Capsicum chinense* 'Haba' produce multiple flowers at one node (B), and it is able to become fruit (C).

Phenotypic data collection and shoot apical meristem imaging

Evaluation of flower number per node of plant was completed after the flower on each node (until sixth) was set and could be clearly distinguish. The mean flower number per node were calculated by averaging the flower numbers of first to sixth nodes with three replicates of each line in the 'TH' RIL population. In the case of CSHL population, the average of first to third nodes with the minimum four biological replicates of each accessions were collected.

To image shoot apical meristems (SAM) using a stereo microscope, shoot apices were dissected from seedlings of all stages and genotypes. Older leaf primordia (larger than 150 mm) were trimmed off under a stereo microscope (Discovery.V12, Carl Zeiss, Germany). The SAM images were taken immediately after dissection with an integrated digital camera (AxioCam MRc, Carl Zeiss, Germany) attached to the stereo microscope.

Genomic DNA and RNA extraction

Genomic DNA (gDNA) was extracted from health young leaves of 'TH' RILs along with the 'TF68', 'Haba' and F_1 using a modified cetyltrimethylammonium bromide (CTAB) method (Porebski *et al*., 1997). Leaf tissues were homogenized using 3 mm steel beads by TissueLyserII (Qiagen, Netherlands). The quantity and purity of gDNA were measured with Nanodrop spectrophotometer (BioTek, Winooski, VT, USA) and by electrophoresis in a 1.2% agarose gels. Confirmed DNA

was diluted to a final concentration of 50 ng/μL with distilled water.

Total RNA was extracted from target tissues including cotyledon and young leaves. Dissected tissues were immediately frozen in liquid nitrogen and ground to fine powder. Total RNA was then extracted using MG Total RNA Extraction kit (MGmed, Korea) following manufacturer's instruction. RNA samples were diluted in RNase-free water (MGmed, Korea). After isolation, the quantity and purity of RNA were measured by Nanodrop spectrophotometer and checked by performing electrophoresis in a 1.2% agarose gels. Complementary DNA (cDNA) was synthesized from 2 μg of RNA using EasyScript Reverse Transcriptase kit (TransGen, China) with oligo(dT) primers. Resulting cDNA were used for cDNA sequence variance confirmation.

Construction of genotyping-by-sequencing libraries

GBS libraries of 'TH' RILs were generated by digest gDNA with two restriction enzyme *PstI*/*MseI* using a SBG 100-Kit v2.0 (Keygene N.V., Wageningen, Netherlands) and gDNA of CSHL population were digested with *EcoRI*/*MseI* (Han et al., 2018). After digest gDNA with the restriction enzymes, adapters were ligated to it. The libraries were amplified with selective primers, which used 'GA' for 'TH' RIL and 'TA' for the CSHL population. Amplified libraries generated from 'TH' RILs and two biological replicates of parents were pooled in a single tube. The libraries of the CSHL population were pooled in three tubes maximum 96 samples each. All tubes were sequenced in separate lanes of a HiSeq 2000 (Illumina, San

Diego, CA, USA) at Macrogen (Seoul, Republic of Korea).

Data analysis for GBS and identification of SNPs

Raw reads get from HiSeq 2000 of the 'TH' RIL and CSHL population libraries were trimmed to a minimum length of 80bp and filtered for a minimum quality of Q20. The filtered reads were aligned to the *C. annuum* 'Dempsey' reference genome v0.1 (unpublished) using the Burrows-Wheeler Aligner program v0.7.12 (Li and Durbin, 2009). For SNP calling and filtering, UnifiedGenotyper of the GATK v.3.8-0 was used (DePristo et al., 2011). SNPs from the 'TH' RIL population were filtered for minimum genotype quality of Q30 and a minimum three read depth.

For the CSHL population, SNP calling was done by same procedure with 'TH' RILs SNP calling criteria. Mono or multiallelic SNPs and SNPs less than 0.6 calling rate from the whole population were removed. After filtering all SNPs, missing data was imputed by using the Beagle v4.1. After fill-in SNPs. Finally, minor allele frequency > 0.03, and an inbreeding coefficient (F) > 0.8 was additionally filtered.

Bin map construction for the biparental population

Missing data and non-polymorphic SNP for 'TH'RIL population parental lines were eliminated and the recombination breakpoints of the RILs were detected using a modified sliding-window approach (Han et al., 2016). The ratios of SNPs with both

parental genotypes was calculated for each window, defined as 20 linked SNPs, and the overall genotype of each window was decided. The SNP ratios of >0.7 and 0.3-0.7 were scored as maternal or paternal, and heterozygous genotypes, respectively. The linkage map of the bins was constructed by using the Carthagene software (De Givry et al., 2005). The criteria used to construct a linkage group were a 3.0 LOD threshold score and limited distance of 30 cM. The Kosambi genetic mapping function was used to calculate the distances between bin markers. The bin maps of 'TH' RILs were constructed based on the *C. annuum* 'Dempsey' reference genome, and were compared using physical locations on the reference genome by the R/script.

QTL analysis for multiple-flower per node trait

QTL controlling the multiple-flower per node trait were detected for 'TH' RILs. Using multiple-flower per node phenotype data and high-density genetic bin maps, composite interval mapping was performed with Windows QTL Cartographer v2.5 (Wang et al., 2012). The genome-wide significance LOD threshold level was determined by 1000 permutation tests with a $p < 0.05$ probability level for trait. Additive by additive effect was measured using a multiple-interval mapping with a Bayesian information criterion (BIC-X) model. The physical locations of the QTL were also compared with the genetic and physical position of bins linked to the QTL within 99% probability. Box plots of QTLs located with the major bin markers were calculated as quartiles of the multiple-flower per node level in the 'TH' RIL population.

Genome-wide association study for multiple-flower per node trait, population structure and haplotype block estimation

The 156,589 filtered SNPs detected from the 276 individuals of the CSHL population were used for association mapping. The population structure estimation Principal component analysis (PCA); Kinship matrixes and compressed mixed linear GWAS model were conducted using the R package Genomic Association and Prediction Integrated Tool (GAPIT) with the default options (Lipka et al., 2012). The *P*-values to select the significant SNPs that affecting on the phenotype variation were determined by Bonferroni correction.

To identify the CSHL population structure, Principal component analysis (PCA) was performed using the R package GAPIT with the default options and result data was regenerated in a plot with 'ggfortify' library in R. The haplotype block of the CSHL population was estimated using PLINK v1.9 (Tang et al., 2016; Chang et al., 2015) with the following settings: '--blocks no-pheno-req no-small-max-span --blocks-inform-frac 0.9 --blocks-max-kb 2000 --blocks-min-maf 0.05 --blocks-recomb-highci 0.80 --blocks-strong-highci 0.85 --blocks-strong-lowci 0.7 --no-parents --no-sex.

Candidate gene prediction and sequence variation analysis

Candidate genes were selected within closely located regions associated with detected SNPs exceeding a significant threshold. Furthermore, genes known for

inflorescence development were searched and selected as candidate genes if placed within QTL positions. CM334 v1.6 reference ('Annuum v2.0' file) was used for gene prediction instead (Kim et al., 2014). To classify the corresponding physical locations of significant SNPs in 'CM334' v1.6 reference genome, 450 kbp interval sequence of significant SNPs was obtained from 'Dempsey' and BLAST searched against 'CM334' v.1.6 reference. After find location of significant SNP from 'CM334' v1.6 reference, closely located significant SNPs within 2 Mbp interval was grouped at each locus and used to find annotated genes.

For candidate gene analysis, sequences of each candidate gene was obtained from NCBI database and blasted against *C. annuum* 'Dempsey' v0.1 reference. Primers were designed based on the candidate gene sequence of 'Dempsey' to amplify the genes. PCR was performed with 35 cycles of 98°C for 10 s, 60°C for 30 s, and 68°C for 1 to 3 min depending on the gene length by using PrimeSTAR GXL DNA polymerase (TaKaRa, Japan). PCR amplicons were separated on a 1.2% agarose gel and purified using LaboPass PCR clean-up kit (Cosmo Genetech, Korea). Standard sanger sequencing was performed at Macrogen (Seoul, Korea). Nucleotide sequences were analyzed by Lasergene's SeqMan program (DNASTAR, Madison, WI, USA). Already existing *C. annuum* 'Dempsey' v0.1, 'CM334' v1.55 and *C. chinense* 'PI159236' v1.2 reference genomes were used to check sequence variations of candidate genes (Kim et al., 2017) . *C. annuum* 'Dempsey' and 'CM334' has a single flower per node the same as 'TF68', whereas *C. chinense* 'PI159236' has multiple flowers like 'Haba'.

SCAR Marker development for candidate gene

Sequence-characterized amplified region (SCAR) marker for *FT* were designed based on the gene structural variations. Primers were designed for genotyping with Primer3web (v.4.1.0, http://bioinfo.ut.ee/primer3/) (**Table 10**). SCAR markers were used to genotype CSHL population to validate the trait association result. For SCAR marker analyses, PCR was performed with 32 cycles of 98°C for 15 s, 57°C for 15 s, and 72°C for 3 min. Resulting amplicons were separated on a 1.2% agarose gel.

RESULTS

Phenotypic variation of multiple-flower per node trait

Average flower number per node from the first to the sixth nodes in 'TF68' was 1.1, whereas, 'Haba' had 3.3 flowers per node. The average flower number per node of the 85 'TH' RILs used on the QTL mapping was 1.49 (**Table 1**). 'TF68' sometimes set more than one flower at the first node.

The multiple-flower per node phenotype of CSHL population was measured from first to third nodes of each accession. The flower number ranged from 1.0 to 3.4 with an average number of 1.32. The difference in flower number per node in each species was observed in that *C. annuum* had an average 1.02, *C. frutecense* 1.33 and *C. chinense* 1.87 flowers per node which can be characterized as a single, medium and multiple flower, respectively (**Table 1**). The distribution of flower number per node showed a positive skew in the both 'TH' RIL and CSHL populations (**Figure 2**).

Table 1. Multiple-flower per node phenotype evaluation result of 'TH' RIL and CSHL populations

Trait	TH-RIL population			CSHL population		
	TF68	Haba	RIL	*C. annuum*	*C. frutescens*	*C. chinense*
Multiflower	1.1 ± 0.2	3.3 ± 1.3	1.49 ± 0.45	1.02 ± 0.11	1.33 ± 0.33	1.87 ± 0.46

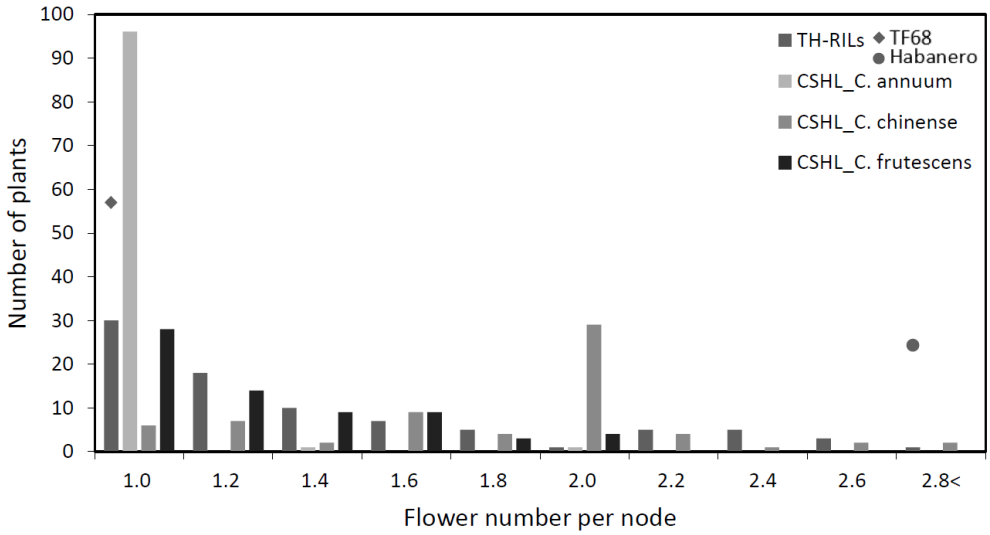


Figure 2. Flower number per node distribution of the 'TH' RIL and CSHL populations. Diamond and circle show the average flower number per node of the 'TF68' and 'Haba', respectively.

Phenotypic variation of shoot apical meristem development

Shoot apical meristem (SAM) first give rise to leaves before transitioning to the floral meristem. Average number of leaves in 'TF68' until reproductive transition and flowering was 12.0 and 13.9 in spring and winter, respectively (**Figure 3; Table 2**). In 'Haba', flowering transition took longer and ranged from 18 to 22 leaves. Transition meristem (TM) showed broader and taller meristem than vegetative stages with a smaller last formed leaf (**Figure 4D, M**). As plants mature, transition meristem (TM) become floral meristem (FM) shape as apical dome and differentiated directly into flower (**Figure 4E, N**). As 'TF68' normally sets a single flower per node, TM developed a single FM and finally showed terminal single flower (**Figure 4E-J**). By contrast, TM gradually developed multiple FM in one node with a time series in 'Haba' (**Figure 4N-S**).

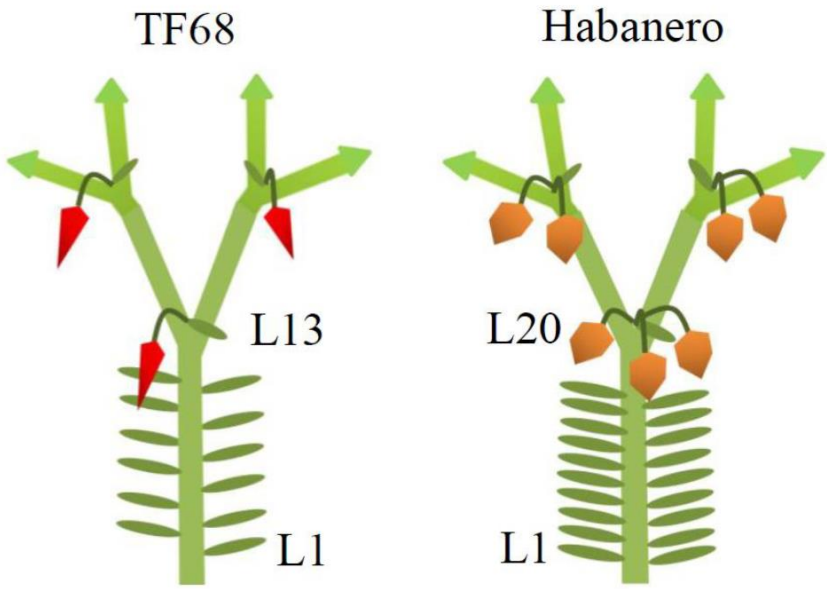


Figure 3. The number of leaves until the first flower emerges in 'TF68' and 'Haba'. Shoot apical meristem (SAM) of pepper produces stems and leaves arranged in an alternate spiral pattern. After generate several true leaves, SAM makes the transition from vegetative to reproductive growth that form first flower.

Table 2. The number of leaves until the first flower emerges in two different environments

	True leaf # until first node	
	2018 Spring	2018 Winter
TF68	12.0 ± 0.00	13.9 ± 0.83
Haba	18.4 ± 0.55	20.8 ± 1.03

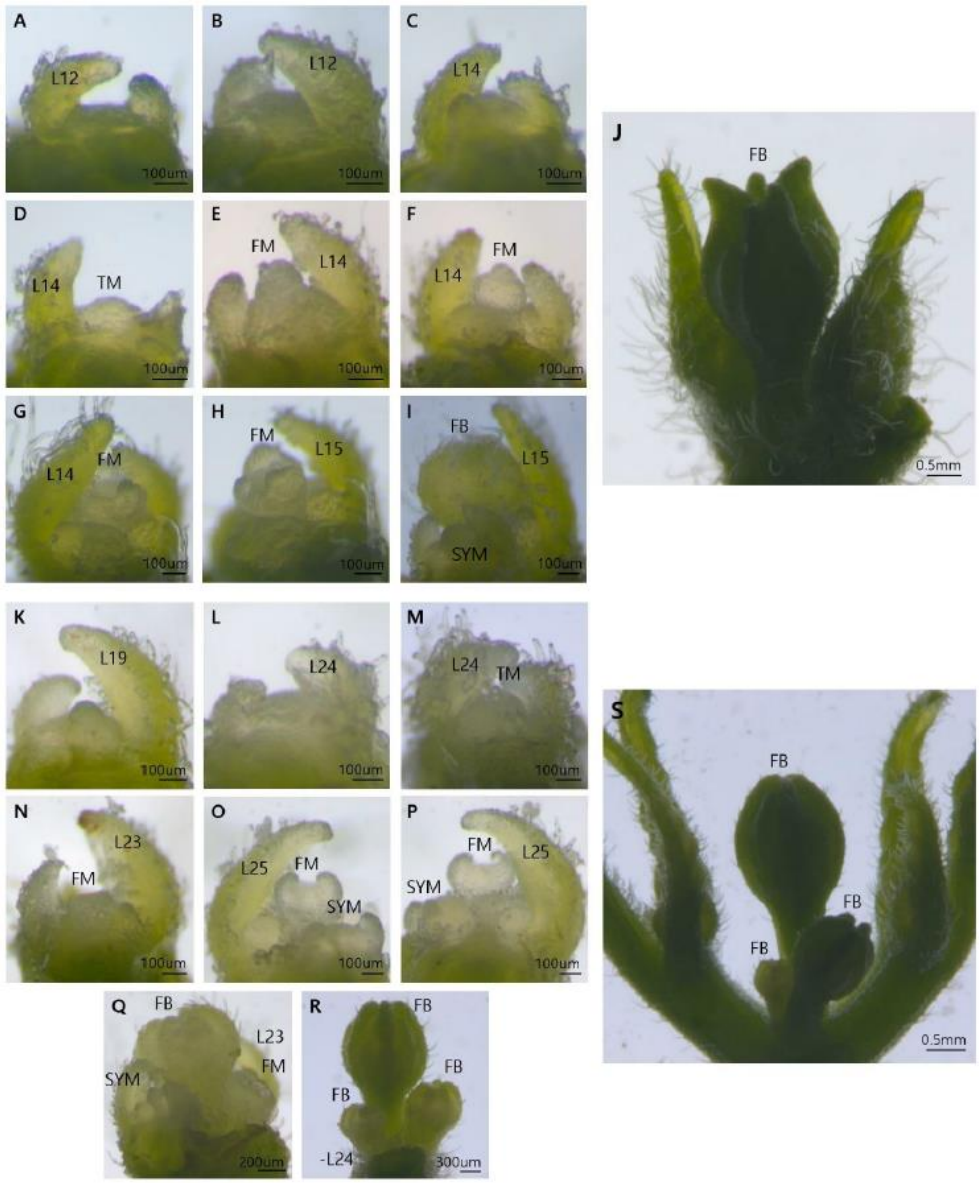


Figure 4. Shoot apical meristem and flowering pattern of 'TF68' (A-J) and 'Haba' (K-S). Leaf number are marked by L, transition meristem stage marked by TM, flower meristem is marked by FM, flower bud marked by FB and sympodial meristem marked by SYM.

Bin map of biparental population

Genotypes of 'TH' RILs were analyzed using GBS. GBS libraries were prepared from *PstI*/*MseI*-digested DNA. The average number of reads per sample was around 4 million, and a total of 10,851 SNPs were detected by aligning the sequences obtained from GBS to the *C. annuum* 'Dempsey' reference genome (**Table 3**). The SNPs were more densely distributed at the ends of the chromosomes than the centromeric region (**Figure 5A**). To correct missing data and genotyping error, a modified sliding window approach was used (Han et al., 2016). Recombination breakpoints were determined using 20 consecutive SNPs as one sliding window, and a high-density bin map of the 'TH' RIL population was constructed (**Figure 6)**. The map consisted of 1,789 bins with an average genetic distance of 0.96 cM (**Table 4**). Among the 12 linkage groups, the genetic distance of chromosome 12 was longest and chromosome 8 was shortest. The total genetic length of the linkage map was estimated to be 1,713 cM. The linkage map was compared with the *C. annuum* 'Dempsey' reference genome, and the physical position of each bin was determined (**Figure 7**). Overall genetic and physical positions of the bins were collinear.

Table 3. Number of sequencing reads generated from GBS and SNPs from GWAS and QTL mapping

	TH-RIL	CSHL population
# of accessions (lines)	85	276
Genotyping method	GBS (*PstI*/*MseI*)	GBS (*EcoRI*/*MseI*)
Avg. # of reads per sample	4,103,757	1,936,524
Total # of SNPs	10,851	156,589
Avg. distance b/w SNPs (bp)	278,746	19,351

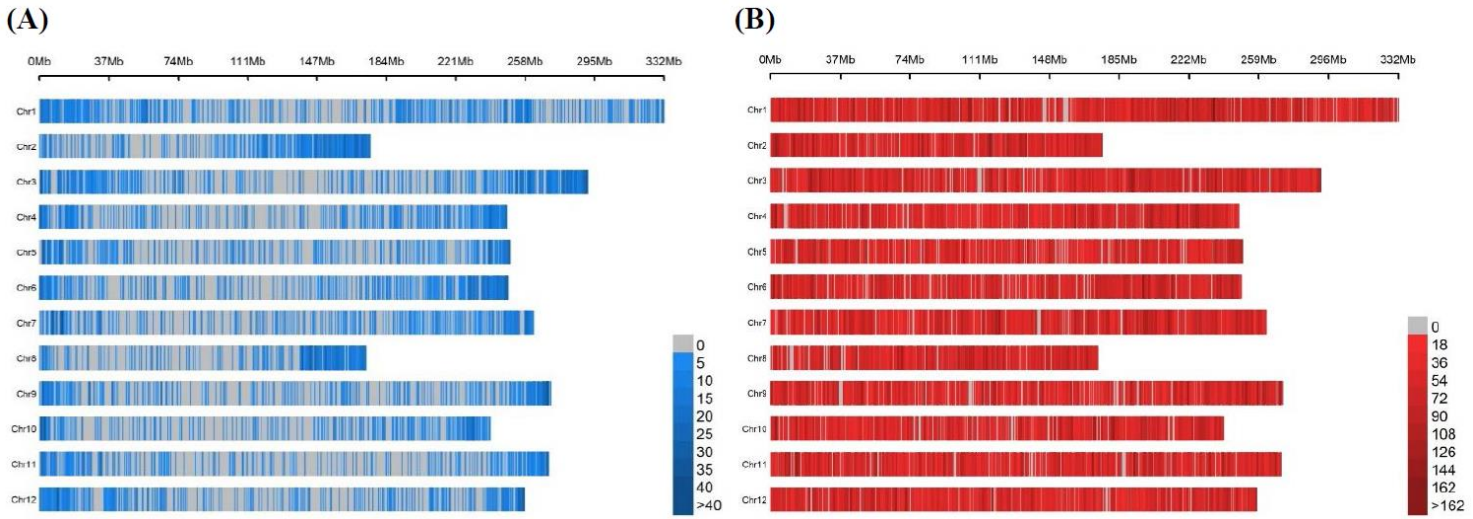


Figure 5. SNP density (number of SNPs within 1 Mb window size) of the 'TH' RIL (A) and CSHL (B) populations.

Table 4. Summary of bins on each chromosome of 'TH' RIL population by GBS

Chr.	Number of SNPs	Number of bins	Physical length of bin (Mb)		Genetic distance of bin (cM)	
			Mean	Total	Mean	Total
1	1,253	192	1.73	332.8	0.78	149.6
2	925	148	1.19	175.8	0.85	126.3
3	1,321	203	1.43	291.2	0.85	171.7
4	854	150	1.66	248.4	0.78	116.4
5	801	143	1.75	250.2	1.04	148.8
6	930	145	1.72	249.5	0.87	125.7
7	896	130	2.02	262.9	0.92	119.7
8	744	113	1.53	173.4	1.00	112.7
9	839	142	1.91	271.6	0.80	113.3
10	711	141	1.70	240.1	1.21	170.8
11	723	138	1.96	271.2	1.04	143.8
12	854	144	1.79	257.7	1.49	214.1
Total	10,851	1,789	1.69	3,024.7	0.96	1,712.9

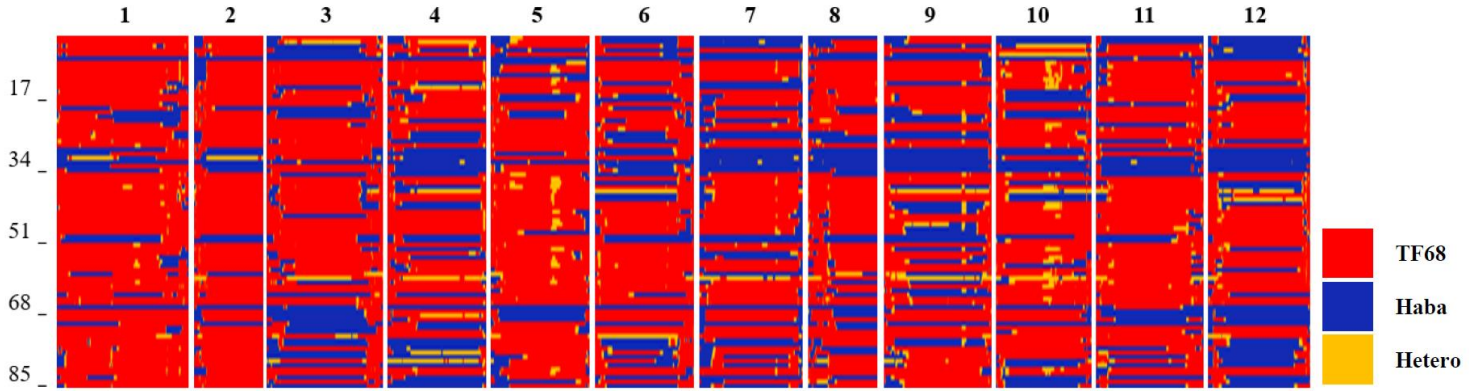


Figure 6. Bin map of the 'TH' RIL population. Red region indicates same genotype with 'TF68', blue means 'Haba' and yellow means heterozygous genotype.

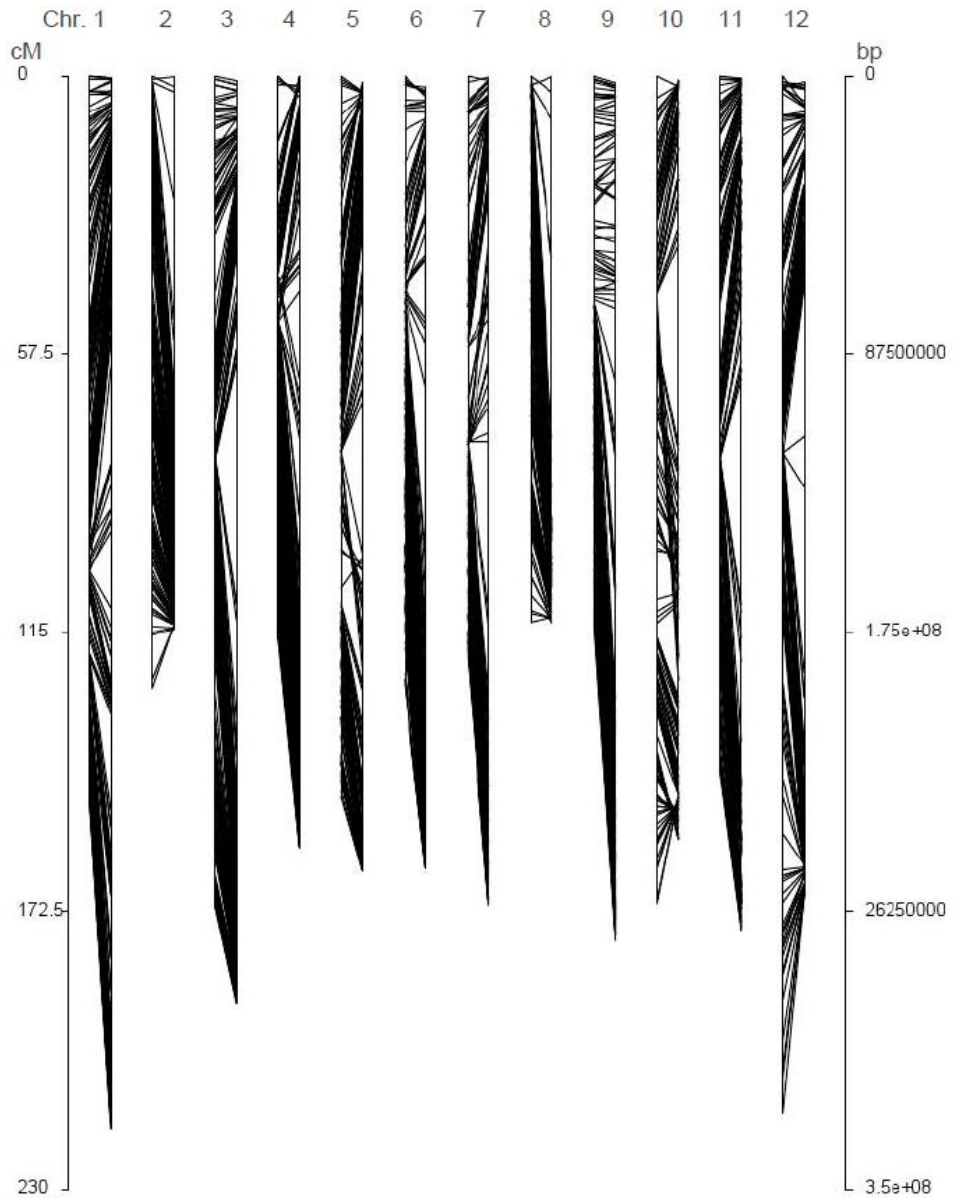


Figure 7. Comparison of the genetic map of 'TH' RILs with the physical map. Bars on the left shows the genetic map position (cM) and right represent the physical map position (bp). Left bars on each chromosome indicate linkage group on the other hand, right bars on each chromosome specify physical location of every marker.

QTL mapping for multiple-flower per node

QTLs controlling the multiple-flower per node were detected in 'TH' RILs (**Figure 8**). Multiple-flower per node phenotype data and an high-density bin map of 85 RILs were used to identify QTLs. QTLs were detected on chromosomes 1, 2, 7, and 11 (**Table 5**). *TH-mf11* showed the highest LOD score and explained 8.13% of total phenotypic variation. The physical locations of the QTLs detected in 'TH' RILs were compared using the *C. annuum* 'Dempsey' reference genome. *TH-mf1* were located at 172.9–191.4 Mbp on chromosome 1, *TH-mf2* was 128.6-139.6 Mbp on chromosome 2, *TH-mf7* was 16.0-32.1 Mbp on chromosome 7, and *TH-mf11* was 1.6-3.2 Mbp on chromosome 11.

Skewed distribution of multiple-flower per node in the 'TH' RILs indicates that there may be epistatic interactions between the QTL. Using multiple-interval mapping (MIM), epistasis effects between common QTL were measured. However, meaningful additive-by-additive epistasis between QTLs was not found. Using multiple-interval mapping (MIM) the effects of the all QTL were calculated (**Table 6**). Four QTL controlling multiple-flower per node trait could explain 65.0% of the phenotypic variation. To check phenotypic effect of detected QTL, box plot was drawn with most significantly associated bin markers with trait. Bin markers of TH1-104.6, TH2-56.1, TH7-50.7 and TH11-6.0 were used for plotting. Detected four QTL were significantly associated with the difference in the multiple-flower per node (**Figure 9**).

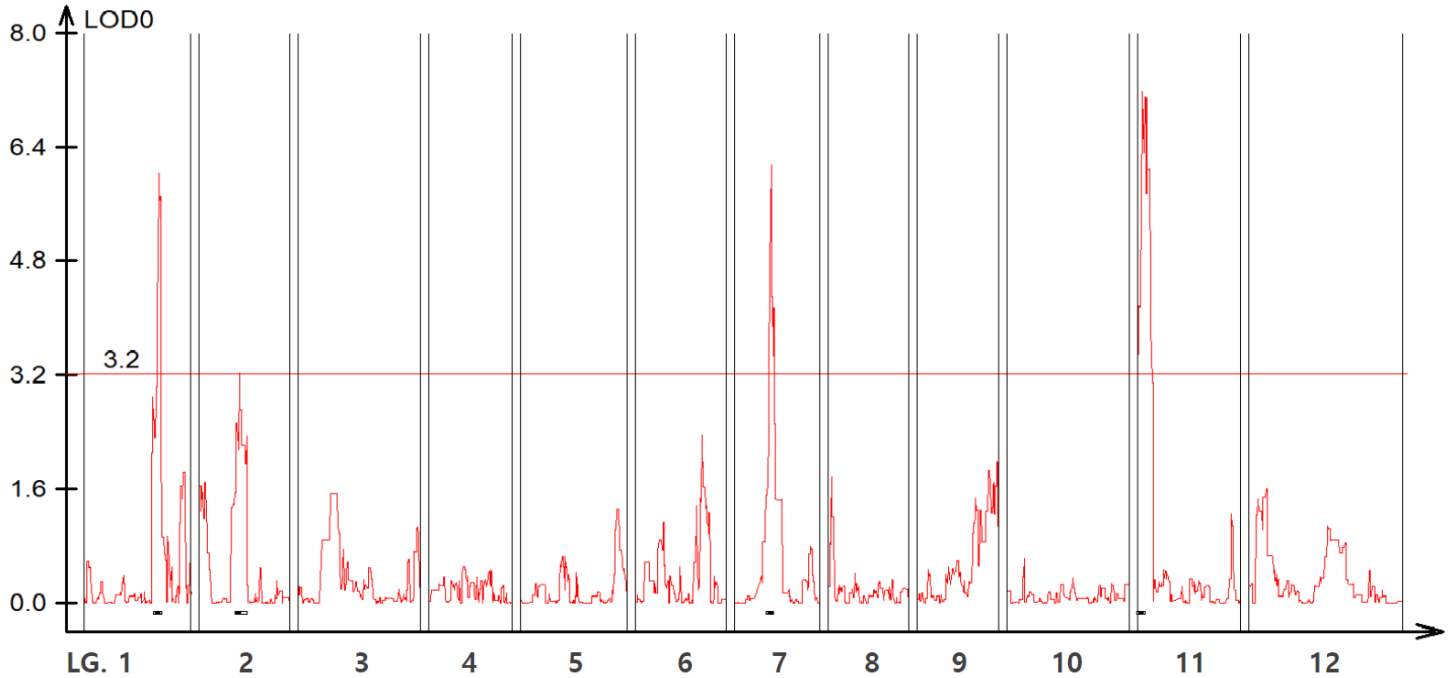


Figure 8. Genome-wide plot of QTL analysis controlling multiple-flower per node trait detected in 'TH' RILs. Red horizontal line determine threshold of the LOD value of 3.2. Small black bars above the X-axis are detected QTL.

Table 5. QTL controlling multiple-flower per node trait detected in 'TH' RILs

Trait	QTL	Chr.	Location (cM)	LOD	R^2 (%)	Direction*	Additive effect
Multiflower	*TH-mf1*	1	102.1-106.9	6.05	6.10	-	0.2
	TH-mf2	2	53.7-65.1	3.23	3.03	-	0.11
	TH-mf7	7	48.3-52.3	6.15	6.45	-	0.17
	TH-mf11	11	3.3-8.4	7.19	8.13	-	0.18

* Genotypes that increase the multiple-flower per node level. - means the genotype resembles that of 'Haba'.

Table 6. A total effect of all multiple-flower per node QTL in 'TH' RILs

Trait	QTL	R^2 (%)*	Total R^2 (%)
Multiflower	*TH-mf1*	26.9	65.0
	TH-mf2	3.6	
	TH-mf7	18.5	
	TH-mf11	16.0	

* R^2 value of individual QTL were evaluated by MIM analyses.

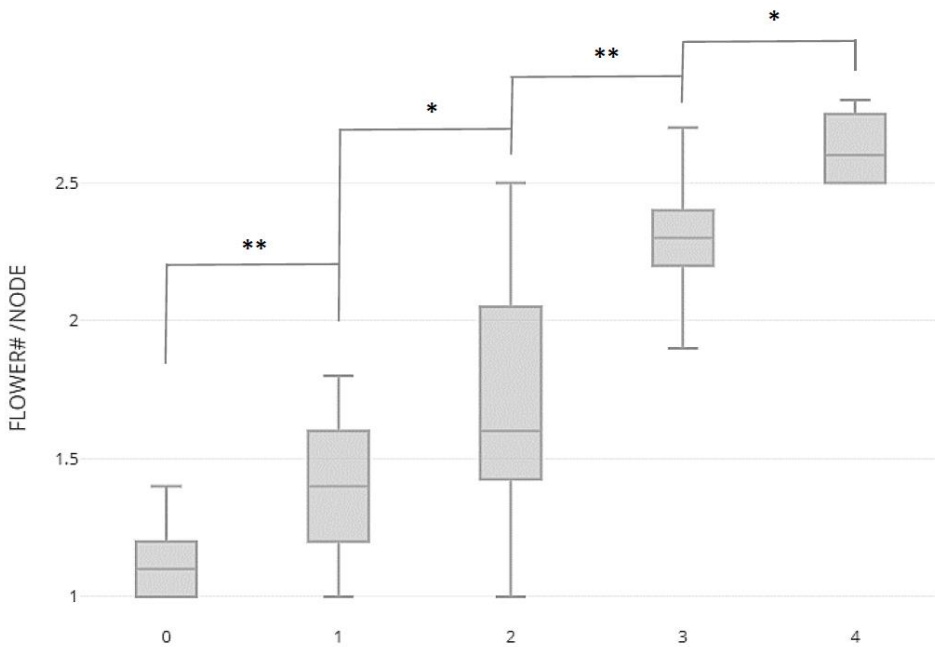


Figure 9. **Box plots of multiple-flower per node phenotype regulated by four *TH-mf* QTLs in plants of the 'TH' RILs.** Numbers on the X-axis indicate 'Haba' and/or heterozygous genotype number on QTL location. Asterisks indicate significantly different multiple-flower per node level from each different number of QTL containing lines (* $P < 0.05$, ** $P < 0.01$). *P*-values were determined by a two-tailed, two-sample *t*-test.

SNPs filtering and haplotype blocks analysis of GWAS population

To validate the QTL detected from the biparental population, a GWAS study for multiple-flower per node was performed using 276 *C. annuum*-clade accessions, including 98 *C. annuum*, 66 *C. chinense*, 67 *C. frutescens* and 45 species unidentified *Capsicum* spp.. The accessions were genotyped using the GBS method. GBS libraries were constructed using *EcoRI*/*MseI* restriction enzyme sets. The average number of reads per sample was around 2 million. SNPs were obtained by aligning the reads to the *C. annuum* 'Dempsey' reference genome. After filtering the SNPs for minor allele frequencies, inbreeding coefficient and calling rate, a total of 156,589 SNPs were selected for further study (**Table 7**). The SNPs of the GWAS population were evenly distributed, with an average distance between SNPs of 19,351 bp (**Figure 5B; Table 7**).

Using these SNPs, *Capsicum* accessions were divided into three subgroups using a PCA (**Figure 10A**) and a phylogenetic analysis (**Figure 10B**). These analyses showed that the accessions of the CSHL population were grouped according to their expected species groups, *C. annuum*, *C. chinense*, and *C. frutescens*. Fifteen accessions were not included in any of the subgroups. The population structure determined from the PCA was applied for the GWAS. Haplotype blocks were calculated in each chromosome using PLINK v1.9 with less strict options than the default settings. About 60% of SNPs were grouped into 5,373 blocks, and each block

contained 3–112 SNPs, with an average of 17.8 SNPs (**Table 7**). The block size varied between 2 bp to 2 Mbp, with average block sizes of 726, 381, 511, 437, 377, 471, 454, 217, 476, 465, 440 and 418 kbp for the twelve chromosomes, respectively. Genome-wide, average haplotype block size was 450 kbp, which was larger than the average distance between the SNPs used for the GWAS.

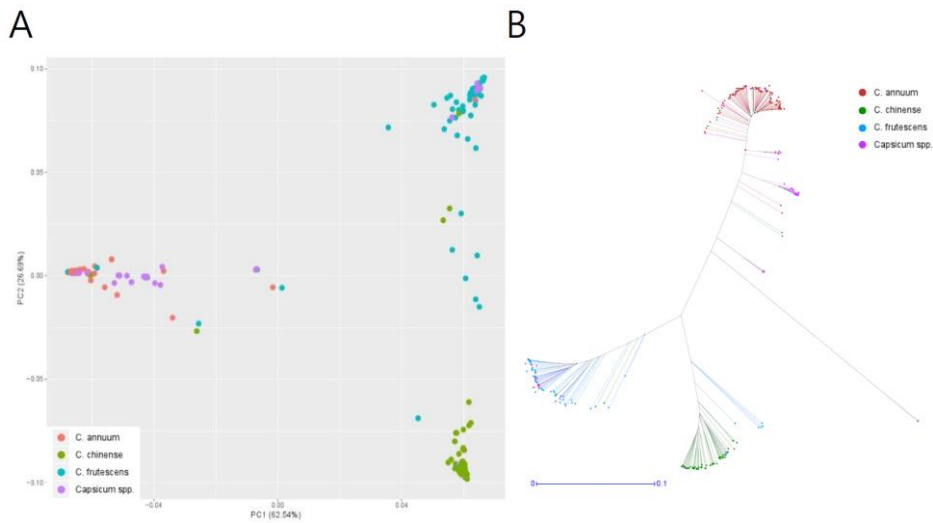


Figure 10. Population structure of the CSHL population, with a principal component analysis (A) and a phylogenetic tree (B) determined from 156,589 SNPs. Red, green, sky blue and purple color indicate *C. annuum*, *C. chinense*, *C. frutescens* and *Capsicum* spp., respectively.

Table 7. Haplotype block estimated by GBS of the CSHL population

Chr.	Number of SNPs*	Avg. distance between SNPs (bp)	Number of SNPs grouped into blocks	Number of blocks	Average block size (Kbp)	Avg. number of SNPs per LD block
1	18,640	17,854	10,665	726	365.78	14.69
2	9,904	17,749	5,817	381	373.85	15.27
3	15,780	18,453	9,137	511	430.63	17.88
4	12,267	20,247	7,225	437	442.35	16.53
5	11,681	21,418	6,906	377	466.22	18.32
6	12,891	19,354	7,760	471	462.03	16.48
7	13,842	18,991	8,475	454	466.66	18.67
8	9,581	18,102	5,630	217	483.45	25.94
9	13,084	20,758	8,062	476	481.5	16.94
10	12,068	19,899	7,498	465	474.03	16.12
11	13,695	19,800	8,209	440	477.24	18.66
12	13,156	19,584	7,639	418	481.58	18.28
Total	156,589	19,351	93,023	5,373	450.44	17.81

*Number of SNPs are specified only filtered SNPs and used for GWAS analysis

GWAS for multiple-flower per node trait

A total of 156,589 SNPs were used to GWAS for multiple-flower per node using 276 *C. annuum*-clade accessions. A total of 83 SNPs were exceed the false discovery rate (FDR) threshold that was considered significantly associated with multiple-flower per node trait (**Figure 11**). These SNPs were grouped into 41 genomic regions using a haplotype block estimation. To identify the corresponding physical locations of significant SNPs in 'CM334' v1.6 reference genome, an average LD block with a size of 450kb was obtained from 'Dempsey' and blast against 'CM334' v.1.6 reference. To avoid missing genes located near the trait-associated SNPs and to cover all significant SNPs which were not included in haplotype block, closely located significant SNPs within 2 Mbp interval was grouped at each locus (**Table 8**). Using gene annotation data ('Annuum.v.2.0.chromosome.gff3') and grouped SNP loci, 388 genes located at 28 associated regions were identified and their functions were predicted (data not shown).

Among 28 regions, three regions on chromosomes 1, 2 and 11 were collocated with QTLs detected in biparental QTL mapping (**Figure 12**). On chromosome 1, one locus physically located 4 Mbp away from *TH-mf1* which is relatively close to considering total 332.8 Mbp length of chromosome 1. Two significant SNPs was located at the *TH-mf2* QTL region. The locus detected on chromosome 11 was only 800 kbp away from *TH-mf11* region.

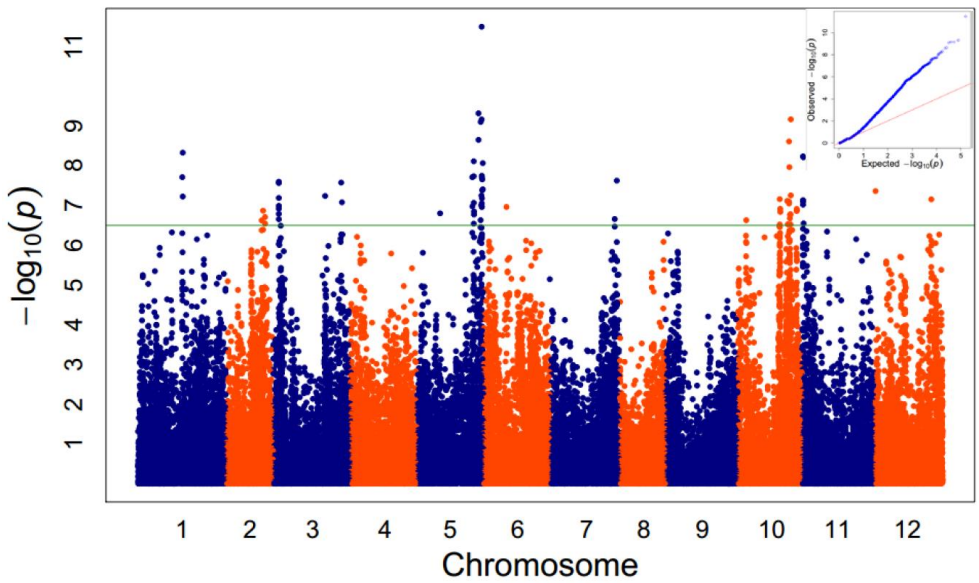


Figure 11. Manhattan plot of SNPs associated with multiple-flower per node trait in CSHL population. QQ plot on the top-right side of the Manhattan plot shows the expected distribution of association test statistics (X-axis) across the million SNPs compared to the observed values (Y-axis).

Table 8. Physical location of 28 SNPs associated the multiple flower traits detected from GWAS analysis

Chr.	Physical position (Mb)	Chr.	Physical position (Mb)
1	137.55-140.9	6	105.06-105.82
2	125.98-126.37	7	230.54-231.16
	129.54-129.78		240.15-240.62
	134.56-135.34	10	31.07-31.68
	142.23-142.35		89.42-89.84
3	15.56-16.17		103.71-104.06
	162.53-162.64		167.78-169.41
	251.63-254.11		183.16-183.78
5	126.92-127.61		186.95-197.02
	198.95-200.41		216.44-216.93
	207.88-207.94	11	3.79-8.56
	212.4-215.34	12	6.84-7.25
	221.15-221.76		206.77-210.47
	227.73-228.26		
	231.08-237.02		

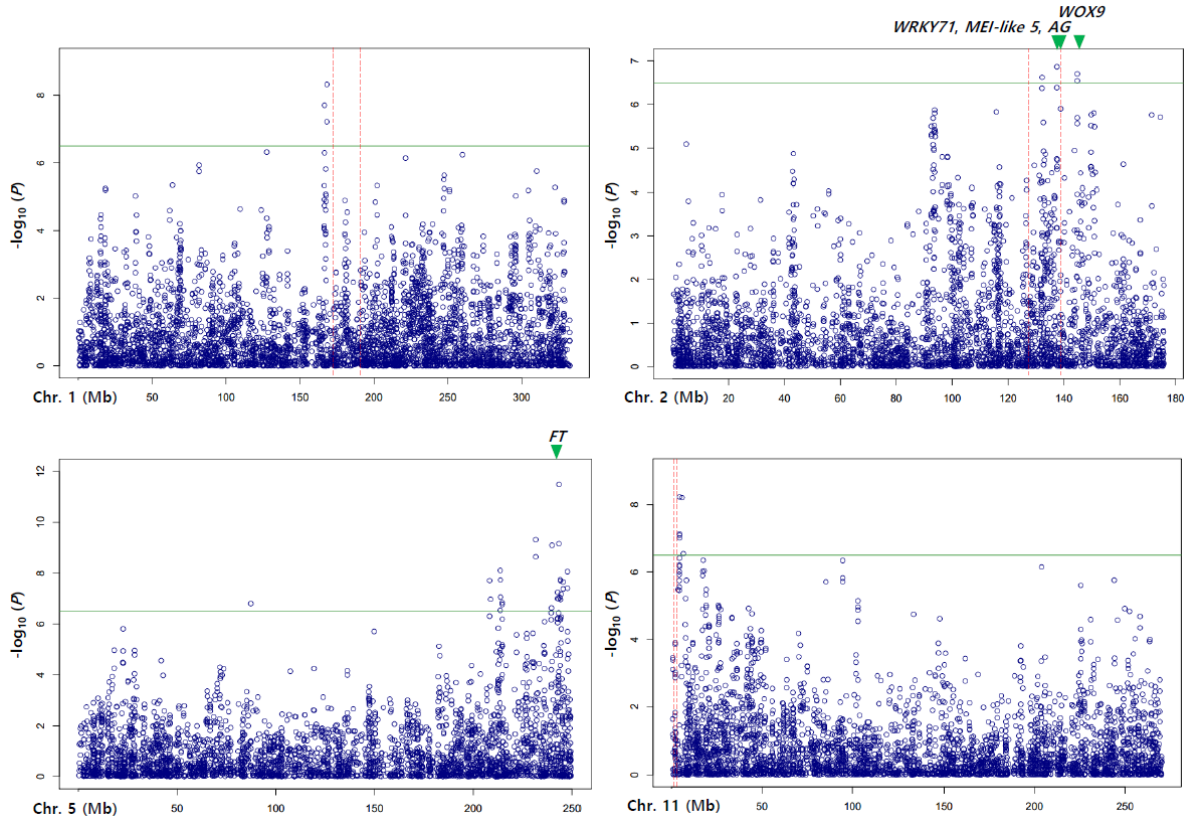


Figure 12. Comparison of QTL region from 'TH' RIL and Manhattan plot from GWAS with multiple-flower per node trait. The threshold of the –log(*P*) was 6.2. QTL region detected from QTL mapping was marked by red dash lines and green triangles indicate position of candidate.

Prediction of candidate genes control multiple-flower per node trait

We were able to identify candidate genes involved in the multiple-flower per node from the QTL mapping and GWAS (**Table 9**). Among 28 associated loci regions from GWAS, the most significant trait related region was on the chromosome 5 and corresponding to *FLOWERING LOCUS T* (*SELF PRUNING 5G* homolog) known as functions in control plant architecture and flower production (**Figure 12**). Another gene, *WUSCHEL*-related homeobox 9 (*WOX9*) (CA.PGAv.1.6.scaffold79.74) known as major determinant of inflorescence architecture in tomato was located around 600 kbp away from four significant SNPs located in 144.7 Mbp on Chromosome 2 (physical location of 'Dempsey' reference) (Lippman et al., 2008).

The comparison of the QTL mapping and GWAS results led to identify 63 genes on the collocated region (**Appendix 1**). Among them, the collocated region on chromosome 2 was associated with three genes encoding floral homeotic protein *AGAMOUS*, *MEI2-like 5*, and putative WRKY transcription factor 71. Floral homeotic protein *AGAMOUS* (CA.PGAv.1.6.scaffold411.1), *MEI2-like 5* (CA.PGAv.1.6.scaffold257.50) and putative WRKY transcription factor 71 (CA.PGAv.1.6.scaffold257.32) known as have a function of shoot apical meristem development and/or floral meristem initiation was detected at 139.4, 137.7 and 137.4 Mbp, respectively.

Table 9. Candidate genes of multiple-flower per node trait on reference genome

Gene annotation	Gene ID (CA.PGAv.1.6.)	Chr.	Physical location in 'CM334 v1.6' (bp)		Physical location in 'Dempsey' (bp)	
			Start	End	Start	End
WUSCHEL-related homeobox 9	scaffold79.74	2	129,246,544	129,248,628	144,102,139	144,104,229
Protein MEI2-like 5	scaffold257.50	2	134,623,870	134,633,408	137,762,619	137,771,713
putative WRKY transcription factor 71	scaffold257.32	2	134,965,138	134,966,803	137,481,716	137,483,298
Floral homeotic protein AGAMOUS	scaffold411.1	2	138,038,471	138,050,630	139,475,699	139,487,957
Protein FLOWERING LOCUS T	scaffold1310.12	5	234,808,906	234,812,234	241,815,139	241,818,453

Confirmation of QTL and GWAS analysis

To confirm the multiple-flower per node of plants to their genotypes at these candidate genes, individual plants of the 'TH' RILs were grouped by their genotypes at floral homeotic protein *AGAMOUS*, putative WRKY transcription factor 71 and Protein *MEI2-like 5*. To draw box plot, TH2-65.9, TH2-64.1 and TH2-62.9 bin markers located 100 kbp, 8 kbp and 27 kbp distance away from the floral homeotic protein *AGAMOUS,* putative WRKY transcription factor 71 and Protein *MEI2-like 5* respectively were used. In case of GWAS analysis, sequence-characterized amplified region (SCAR) marker for *FT* was designed based on 589 bp deletion in the first intron of 'Haba' which was also identified in *C. chinense* 'PI159236' v1.2 reference genome (**Figure 13A; Table 10**). Developed SCAR marker, FT_F2 and FT_R3 primer was used for genotyping CSHL population (**Figure 13B**). Individual plants in 'TH' RIL and CSHL populations were classified into three groups according to their genotypes: 'TF68', heterozygote, and 'Haba' genotype marked as A, H and B respectively.

The individual lines in 'TH' RILs were separated by their genotypes at analyzed three candidate genes and all of the markers were associated with significant differences in multiple-flower per node index (**Figure 14A-C**). Also in the CSHL population, differences in the genotypes of *FT* led to highly significant differences in multiple-flower per node (**Figure 14D**).

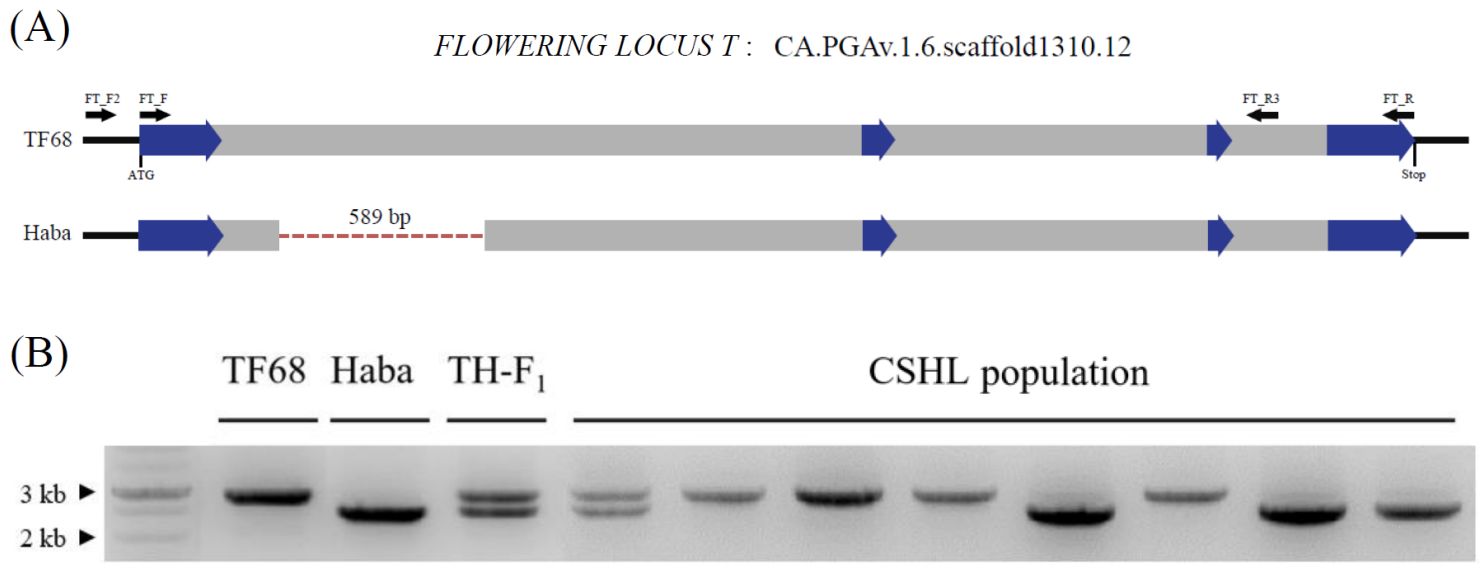


Figure 13. Structural variation of *FT* in 'Haba' (A) and SCAR marker genotyping of *FT* in CSHL population (B). Dark blue arrows indicate exon region and grey boxes signify intron region of *FT*. Red dotted line means sequence deletion on first intron of 'Haba'. Primers used for gene analysis was depicted as black arrow above the gene structure. A bigger band 2,737bp represent TF68-type allele, while a smaller band 2,148bp characterize Haba-type allele in *FT*.

Table 10. List of primers used in SCAR marker and sequence variation analysis for *FT*

Primer name	Primer sequence (5' to 3')	Position ('Dempsey' Chr.02)
FT_F	ATGCCAAGAGATCCTTTAATT	241,815,139 - 241,815,159
FT_R	TTATAGACGACGACCACCAGT	241,818,433 - 241,818,453
FT_F2	ATCGTAGCTAAACCCCAAACAC	241,814,926 - 241,814,947
FT_R3	CAAGCAAGTTAGGTGCCTTTTC	241,817,637 - 241,817,658

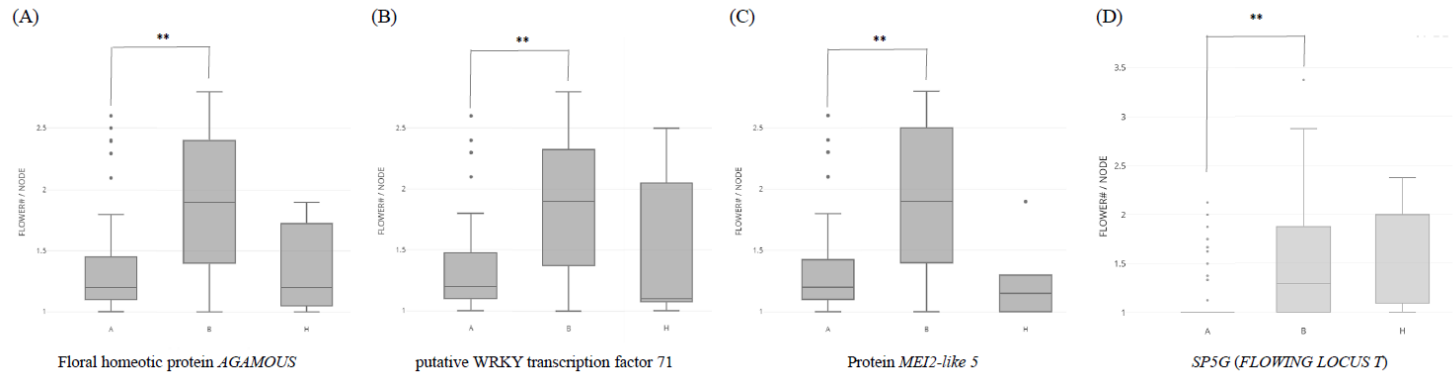


Figure 14. **Box plots of multiple-flower per node variation regulated by candidate genes in 'TH' RIL (A-C) and CSHL (D) populations.** 'TH' RILs were classified by genotype of the most closely linked bin marker to the target genes. CSHL *C. annuum* clade accessions were genotyped by *FT* gene based SCAR marker. Asterisks indicate significantly different phenotype (* $P < 0.05$, ** $P < 0.01$). *P*-values were determined by a two-tailed, two-sample *t*-test.

Sequence variation of candidate genes

To reveal sequence differences in candidate genes, target sequence regions were obtained from reference pepper genome. Gene coding regions were gained from single flowering per node accession 'Dempsey' or 'CM334' and multiple flowering plants 'PI159236'. In the predicted coding sequence (CDS) of Protein *MEI2-like 5*, and *WOX9*, amino acid sequence variations were detected (**Appendix 2**). On the CDS of *MEI2-like 5*, six amino acid difference was observed between 'CM334' and 'PI156236'. 'Dempsey' and 'PI159236' had three nonsynonymous substitutions in *WOX9*. Direct sequencing of *FT* from cDNA of 'TF68' and 'Haba' with FT_F and FT_R primer set revealed two SNPs in the coding region. However, both point mutation was synonymous mutation (**Table 10; Appendix 3**). Other candidate gene floral homeotic protein *AGAMOUS* also had SNPs on the coding region between 'Dempsey' and 'PI159236' reference genome but not in amino acid sequence. Putative WRKY transcription factor 71 had no sequence variation in the coding region.

DISCUSSION

Flower production is enormously important in all crops, serving as the foundation for yield and increased profits. Flower number can affect total yield of fruits in pepper. Therefore, flower number per node is an essential factor to evaluate fruit production and yield enhancement. Flower number per node in *Capsicum* varies depending on species. Identifying the genetic factor controlling multiple-flower per node and introgression of the trait into elite varieties could increase fruit yield potential. According to the previous report, multiple-flower per node trait is controlled by several loci (Tanksley and Oliva, 1984). Precise mapping of quantitative trait has been challenging due to difficulties of phenotyping in multiple year and environments and the need of high density genetic maps. Nowadays, rapidly developing high throughput next generation sequencing (NGS) based genotyping technologies and statistically advanced quantitative trait loci predict models provides unprecedented opportunities to identify genes underlying quantitative traits. In this study, we attempted to identify genetic factors associated with multiple-flower per node in *Capsicum* using QTL mapping and GWAS.

When the physical locations of QTL for multiple-flower per node characters detected in traditional QTL mapping and GWAS were compared, three collocated QTLs on chromosome 1, 2 and 11. Zhu et al. (2019) also showed that the multiple-flower per node associated QTLs on Chromosome 2, 7 and 10 using a high density

genetic map of 150 F_2 population derived from a cross between *C. chinense* '740' and *C. annuum* 'CA1'. Among them, QTL on chromosome 2 showed the largest effect and could explain approximately 40% of the multiple-flower per node variation. Through the BLAST of target region on 'Dempsey' reference genome it was found that located around 159 - 160 Mbp region which is 20 Mbp away from *TH-mf2* and 14 Mbp apart from significant SNPs in GWAS. They also mentioned that two novel minor QTLs on chromosome 8 and 11 were found by other interval mapping method. Due to lack of information for other QTLs except one on chromosome 2, we were not able to compare our data with theirs. Tanksley and Oliva (1984) suggests that epistasis plays a major role in the determination of multiple-flower per node in the segregating generations. However, significant epistatic interactions between QTL was not detected in this study.

By combining GWAS and QTL analyses, we nominated five candidate genes involved in the development of shoot and flower meristem for controlling multiple-flower per node in pepper: *WOX9*, *MEI2-like 5*, *WRKY71*, *AGAMOUS* and *FT*. All genes chosen as candidate are known to have functions in SAM development, determine plant and inflorescence architecture. Finding candidate genes from other model Solanaceous plant like tomato is a logical approach. *WOX9* shares homology with the meristem maintenance gene *WUSCHEL* (*WUS*) and are plant-specific transcription factors. Lippman et al. (2008) found that *WOX9* (*COMPOUND INFLORESCENCE*; *S* in tomato) is a major determinant of inflorescence architecture in tomato. Mutant alleles of *s* dramatically increase branch and flower number (Soyk

et al., 2017a; Lippman et al., 2008). *S* homolog in pepper was revealed by Cohen et al. (2014) and named *Capsicum annuum S* (*CaS*). They found that the *CaS* mutant shows delayed initiation of sympodial growth or in the termination of sympodial meristems and completely inhibits flower formation (Cohen et al., 2014). Since *WOX9* in pepper promotes meristem transition from vegetative to reproductive phase and is required for flower formation, we can conclude that it could be considered as one of the genetic factors control multiple-flower per node trait. We showed that there are three amino acid changes between 'Dempsey' and 'PI159236' from WOX9.

Function of *MEI2-like 5* was studied in maize (*Zea mays*). If *terminal ear 1* (*te1*) (*MEI2-like* gene) in maize has mutation, plants show smaller vegetative shoot apex and fewer leaf founder cells (Veit et al., 1998). They claimed that small size of mutant SAM in *te1* is associated with leaf initiation occurring more frequently than in the wild type (Anderson et al., 2004). Furthermore, long vegetative internodes of the main shoot that precede the tassel are abnormally short and the tassel could either less branched or feminized (Alvarez, 2002). There is a possibility that genetic change of *MEI2-like* gene could also change SAM activity in pepper. More recently, it was shown that *WRKY71* accelerates the initiation of the floral meristem (FM) via the direct activation of *FLOWERING LOCUS T* and *LEAFY* in *Arabidopsis thaliana* (Yu et al., 2016). *SP5G*, the *FLOWERING LOCUS T like* gene, in tomato are known as flowering repressor (Cao et al., 2016). In tomato, *SP5G* controls primary and canonical axillary shoot flowering time and is the major contributor to day-length sensitivity in wild tomato (Lemmon et al., 2018) Therefore, mutations in *SP5G* cause

rapid flowering and enhance the compact determinate growth habit of tomatoes (Soyk et al., 2017b). However, in other Solanaceae crop groundcherry (*Physalis pruinosa*), mutation in orthologue of *SP5G* (*Ppr-SP5G*) resulted in more fruits along each shoot not in early primary shoot flowering (Lemmon et al., 2018). They found *Ppr-sp5g*^{CR} showing 50% higher concentrations of fruits per each shoot. When the pepper, groundcherry and tomato are compared, groundcherry are more distantly related to the pepper than tomato (Wolff, 1991; Sarkinen et al., 2013). In addition, plant architecture and sympodial growth habit of pepper are much closer to groundcherry than those of tomato. As revealed in the previous studies, due to species specific sympodial growth patterns, phenotypic differences can occur from mutations in orthologous florigen gene family. Therefore, we suggest that putative WRKY transcription factor 71 and *FT* are strong candidates for multiple-flower per node trait. Another candidate gene on chromosome 2 is an *AGAMOUS* homolog which are known as a MADS domain transcription factor essential for the termination of floral stem cell fate. Mutations in this gene terminate floral stem cell maintenance in *Arabidopsis* through indirectly repressing *WUS* (Liu et al., 2011). In conclusion, we demonstrated that five candidate genes including *WOX9*, *MEI2-like 5*, *WRKY71*, *AGAMOUS* and *FT* could be strong candidate controlling multiple-flower per node in pepper. More studies are required to validate the roles of these gene.

Here, we found genetic factors associated with multiple-flower per node in *Capsicum* using QTL mapping and GWAS. We identified five candidate genes

involved in the development of shoot and flower meristem within the QTLs in this study. A deep knowledge of the genetic element of multiple-flower per node will be helpful to further breeding for high yield varieties in pepper.

REFERENCES

Alvarez NDG (2002) Molecular genetic analysis of plant *Mei2-like* genes. Department of Plant Biology, Massey University

Anderson GH, Alvarez ND, Gilman C, Jeffares DC, Trainor VC, Hanson MR, Veit B (2004) Diversification of genes encoding *Mei2-like* RNA binding proteins in plants. Plant Mol Biol 54(5):653-70

Bai X, Wu B, Xing Y (2012) Yield-related QTLs and their applications in rice genetic improvement. J of Int Plant Bio 54:300–311

Cao K, Cui L, Zhou X, Ye L, Zou Z, Deng S (2015) Four tomato FLOWERING LOCUS T-Like proteins act antagonistically to regulate floral initiation. Front Plant Sci 6:1213

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7

Cohen O, Borovsky Y, David-Schwartz R, Paran I (2014) *Capsicum annuum S* (*CaS*) promotes reproductive transition and is required for flower formation in pepper (*Capsicum annuum*). New Phytol 202(3):1014-23

de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T (2004) CARHTAGENE: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21(8):1703-4

DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5): 491–498

Fraenkel L, Bogardus ST Jr, Concato J, Wittink DR (2004) Treatment options in knee osteoarthritis: the patient's perspective. Arch Intern Med 164(12):1299-

304

Guo DL, Zhao HL, Li Q, Zhang GH, Jiang JF, Liu CH, Yu YH (2019) Genome-wide association study of berry related traits in grape [*Vitis vinifera* L.] based on genotyping-by-sequencing markers. Hortic Res 6:11

Han K, Jeong HJ, Yang HB, Kang SM, Kwon JK, Kim S, Choi D, Kang BC (2016) An ultra-high-density bin map facilitates high-throughput QTL mapping of horticultural traits in pepper (*Capsicum annuum*). DNA Res 23(2):81-91

Han K, Lee HY, Ro NY, Hur OS, Lee JH, Kwon JK, Kang BC (2018) QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. Plant Biotechnol J 16(9):1546-1558

He Y, Wu D, Wei D, Fu Y, Cui Y, Dong H, Tan C, Qian W (2017) GWAS, QTL mapping and gene expression analyses in *Brassica napus* reveal genetic control of branching morphogenesis. Sci Rep 7(1):15971

Kang BC, Nahm SH, Hhu JH, Yoo HS, Yu UW, Lee MH, Kim BD (2001) An interspecific (*Capsicum annuum* × *C. chinense*) F_2 linkage map in pepper using RFLP and AFLP markers. Theor Appl Genet 102:531–539

Kim OR, Cho MC, Kim BD, Huh JH (2010) A splicing mutation in the gene encoding phytoene synthase causes orange coloration in Habanero pepper fruits Mol Cells 30(6):569-74

Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, Kim SB, Han K, Lee J, Park M, Lee HA, Lee HY, Lee Y, Oh S, Lee JH, Choi E, Choi E, Lee SE, Jeon J, Kim H, Choi G, Song H, Lee J, Lee SC, Kwon JK, Lee HY, Koo N, Hong Y, Kim RW, Kang WH, Huh JH, Kang BC, Yang TJ, Lee YH, Bennetzen JL, Choi D (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol 18:210

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K, Lee GW, Oh SK, Bae C, Kim SB, Lee HY, Kim SY, Kim MS, Kang BC, Jo YD, Yang HB, Jeong HJ, Kang WH, Kwon JK, Shin C,

Lim JY, Park JH, Huh JH, Kim JS, Kim BD, Cohen O, Paran I, Suh MC, Lee SB, Kim YK, Shin Y, Noh SJ, Park J, Seo YS, Kwon SY, Kim HA, Park JM, Kim HJ, Choi SB, Bosland PW, Reeves G, Jo SH, Lee BW, Cho HT, Choi HS, Lee MS, Yu Y, Do Choi, Y, Park BS, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai HS, Ahn HK, Yeam I, Giovannoni JJ, Rose JK, Sorensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim, JS, Lee YH, Choi D (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet 46:270-278

Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, Rodriguez-Leal D, Van Eck J, Lippman ZB (2018) Rapid improvement of domestication traits in an orphan crop by genome editing. Nat Plants 4(10):766-770

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754-60

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28(18):2397-9

Lippman ZB, Cohen O, Alvarez JP, Abu-Abied M, Pekker I, Paran I, Eshed Y, Zamir D (2008) The making of a compound inflorescence in tomato and related nightshades. PLoS Biol 6(11):e288

Liu X, Kim YJ, Müller R, Yumul RE, Liu C, Pan Y, Cao X, Goodrich J, Chen X (2011) *AGAMOUS* terminates floral stem cell maintenance in *Arabidopsis* by directly repressing *WUSCHEL* through recruitment of Polycomb Group proteins. Plant Cell 23(10):3654-70

Navarro JAR, Willcox M, Burgueno J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, Ortega A, Banda AE, Montiel MOG, Ortiz-Monasterio I, Vicente FS, Espinoza AG, Atlin G, Wenzl P, Hearne S (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet 49:476–480

Penella C, Calatayud A (2018) Pepper Crop under Climate Change: Grafting as an

Environmental Friendly Strategy. Climate Resilient Agriculture - Strategies and Perspectives. London: IntechOpen.

Pereira L, Ruggieri V, Pérez S, Alexiou KG, Fernández M, Jahrmann Y, Pujol M, Garcia-Mas J (2018) QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. BMC Plant Biol 18:324

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7(2):e32253

Pootakham W, Sonthirod C, Naktang C, Jomchai N, Sangsrakru D, Tangphatsornruang S (2016) Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: a case study in oil palm (*Elaeis guineensis*). Mol Breed 36(11):154

Porebski S, Bailey G, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep 15:8-15

Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree BMC Evol Biol 13:214

Sasaki K, Fujita D, Koide Y, Lumanglas PD, Gannaban RB, Tagle AG, Obara M, Fukuta Y, Kobayashi N, Ishimaru T (2017) Fine mapping of a quantitative trait locus for spikelet number per panicle in a new plant type rice and evaluation of a near-isogenic line for grain productivity. J Exp Bot 68(11):2693-2702

Shaffer JR, Feingold E, Marazita ML (2012) Genome-wide association studies prospects and challenges for oral health. J Dent Res 91(7): 637-641

Shuh DM, Fontenot JF (1990) Gene transfer of multiple flowers and pubescent leaf from *Capsicum chinense* into *Capsicum annuum* backgrounds. J Am Soc Hortic Sci 155:499-502

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

Soyk S, Lemmon ZH, Oved M, Fisher J, Liberatore KL, Park SJ, Goren A, Jiang K, Ramos A, van der Knaap E, Van Eck J, Zamir D, Eshed Y, Lippman ZB (2017a) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. Cell 169(6):1142-1155

Soyk S, Müller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jiménez-Gómez JM, Lippman ZB (2017b) Variation in the flowering gene *SELF PRUNING 5G* promotes day-neutrality and early yield in tomato. Nat Genet 49(1):162-168

Su C, Wang W, Gong S, Zuo J, Li S, Xu S (2017) High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. Front Plant Sci 8:706

Subramanya R (1983) Transfer of genes for multiple flowers from *Capsicum chinense* to *Capsicum annuum*. HortScience 18(5):747-749

Szalma SJ, Hostert BM, Ledeaux JR, Stuber CW, Holland JB (2007) QTL mapping with near-isogenic lines in maize. Theor Appl Genet114(7):1211-28

Tang Y, Horikoshi M, Li W (2016) ggfortify: unified interface to visualize statistical result of popular R packages. The R Journal 8(2):478-489

Tanksley SD, Olivas JI (1984) Inheritance and transfer of multiple-flower character from *Capsicum chinense* into *Capsicum annuum*. Euphytica 33:769–777

Truong HT, Ramos AM, Yalcin F, de Ruiter M, van der Poel HJ, Huvenaars KH, Hogers RC, van Enckevort LJ, Janssen A, van Orsouw NJ, van Eijk MJ (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLoS ONE 7:e37565

Veit B, Briggs SP, Schmidt RJ, Yanofsky MF, Hake S (1998) Regulation of leaf initiation by the *terminal ear 1* gene of maize. Nature 393:166–168

Wang S., C. J. Basten, and Z.-B. Zeng (2012). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University.

Wang X, Chen Z, Li Q, Zhang J, Liu S, Duan D (2018) High-density SNP-based QTL mapping and candidate gene screening for yield-related blade length and width in *Saccharina japonica* (Laminariales, Phaeophyta). Sci Rep 8:13591

Watson JE, Greenleaf WH (1986) Pepper breeding, M.J. Bassett; Breeding vegetable crops. AVI, Westport, Conn. 67-134

Wolff XY (1991) Species, cultivar, and soil amendments influence fruit production of two *Physalis* species. HortScience 26:1558-1559

Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, Matsuoka M (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat Genet 48(8):927-34

Yu Y, Liu Z, Wang L, Kim SG, Seo PJ, Qiao M, Wang N, Li S, Cao X, Park CM, Xiang F (2016) *WRKY71* accelerates flowering via the direct activation of *FLOWERING LOCUS T* and *LEAFY* in *Arabidopsis thaliana*. Plant J 85(1):96-106

Zhao H, Basu U, Kebede B, Qu C, Li J, Rahman H (2019) Fine mapping of the major QTL for seed coat color in *Brassica rapa* var. Yellow Sarson by use of NIL populations and transcriptome sequencing for identification of the candidate genes. PLoS ONE 14(2):e0209982

Zhao X, Luo L, Cao Y, Liu Y, Li Y, Wu W, Lan Y, Jiang Y, Gao S, Zhang Z, Shen Y, Pan G, Lin H (2018) Genome-wide association analysis and QTL mapping reveal the genetic control of cadmium accumulation in maize leaf. BMC Genomics 19(1):91

Zhu Z, Sun B, Wei J, Cai W, Huang Z, Chen C, Cao B, Chen G, Lei J (2019) Construction of a high density genetic map of an interspecific cross of *Capsicum chinense* and *Capsicum annuum* and QTL analysis of floral traits. Sci Rep 9:1054

ABSTRACT IN KOREAN

꽃 생산은 모든 작물에서 중요한 요소이며 이는 수확량 및 수익증대를 위한 기본적인 요소이다. 고추내 *Capsicum annuum*종은 가축성 분지 구조에 한 개의 꽃을 가지고 있는데 반해 *C. chinense*종은 마디 당 여러 개의 꽃을 생산한다. *C. annuum* 은 전세계에서 가장 널리 재배되는 고추종이며, 이는 총 고추 과실 생산량의 80%를 차지하는 대표적인 종이다. 따라서 마디 당 화방수(복화방)를 조절하는 유전인자를 탐색하여 이를 *C. annuum* 종으로 이입하면 수확량을 증가시킬 수도 있다. 따라서 본 연구에서는 두 가지 연구를 통해 고추의 복화방을 조절하는 유전적 요인을 규명하고자 하였다. 복화방을 조절하는 양적형질유전자좌를 탐색하기 위해 *C. annuum* 'TF68'과 *C. chinense* 'Habanero' 로 구축한 85 개의 근동질유전자계통 집단을 활용하였다. 계통 당 세 반복 씩 1~6 마디내 평균 꽃 수를 조사하였고, 차세대유전체분석 기술을 기반으로 다수의 단일염기다형성 마커를 탐색하는 방법을 이용하여 고밀도 유전자지도를 작성하였다. 12 개의 염색체에 총 10,851 개의 마커를 bin 마커로 변환하여 고밀도 유전적지도를 만들었다. 이 유전적 지도의 총 길이는 1,713cM 이며 평균 bin 마커간 거리는 0.96cM 이다. QTL 분석을 통해 복화방 형질을 조절하는 네 개의 QTL 을 1, 2, 7, 11 번 염색체에서 발견하였으며, 이는 전체 복화방 표현형 변이의 65%를 설명할 수 있었다. 탐색된 QTL 을 입증하여 명확히 하고자 전장유전체상관성분석 (GWAS)를 실시하였다. *C. annuum* 속에 포함되는 *C.*

annuum 98 계통, *C. chinense* 66 계통, *C. frutescens* 67 계통과 *Capsicum* spp. 45 계통, 총 276 계통이 분석에 사용되었다. 차세대유전체분석 기술을 기반으로 다수의 단일염기다형성 마커를 탐색하는 방식을 통해 SNP 마커를 탐색하였고, 필터링 후 남은 총 156,589 개의 SNP 마커가 상관분석에 사용되었다. 그 결과 총 28 개의 유전적지역이 복화방 형질에 연관되어 있는 것으로 밝혀졌으며 세 지역은 QTL 분석과 공통적으로 확인되었다. 해당 연구결과를 바탕으로 정단 분열조직 및 화기 분열조직 발달에 관여하는 다섯 개의 후보 유전자를 확인하였다. 이 결과는 고추종에서의 복화방특성을 이해하는데 기여할 것으로 기대되며, 생산량을 증대하는 품종개량에 도움될 것이라 판단된다.

주요어: 고추 (*Capsicum annuum*), 꽃 생산량, 수확량, 양적형질유전자좌 (QTL), 전장유전체상관성분석 (GWAS), Genotyping-by-sequencing (GBS)

학번: 2017-23343

Appendix 1. List of 63 genes detected from collocated region from QTL mapping and GWAS

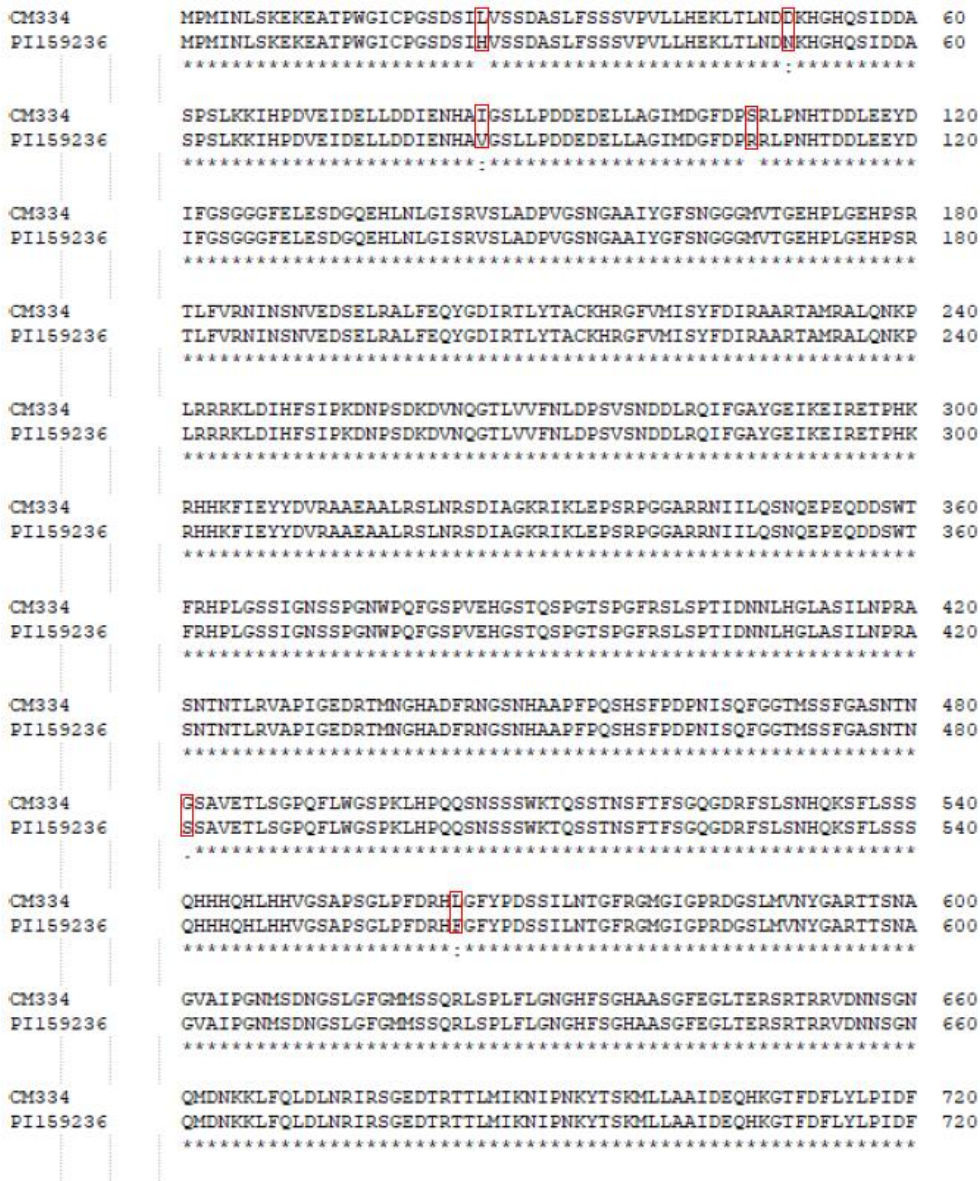
Chr.	Start	End	Protein feature
1	59,261,797	59,262,309	Zinc finger A20 and AN1 domain-containing stress-associated protein 8
1	126,467,683	126,472,417	Ethylene-responsive transcription factor
1	127,788,154	127,789,815	Pectinesterase
1	185,767,025	185,769,226	Pentatricopeptide repeat-containing protein, mitochondrial
1	189,773,684	189,774,094	histone H3.2
1	193,273,645	193,277,582	GDSL esterase/lipase
1	194,933,350	194,934,303	Photosystem II CP43 reaction center protein
2	125,900,657	125,907,494	DEAD-box ATP-dependent RNA helicase 16
2	125,930,191	125,931,792	Reticuline oxidase-like protein
2	126,041,161	126,044,631	Fatty acid 2-hydroxylase 2
2	126,117,344	126,118,837	Pentatricopeptide repeat-containing protein
2	126,146,936	126,150,956	putative LRR receptor-like serine/threonine-protein kinase
2	126,226,912	126,229,521	putative receptor-like protein kinase
2	126,268,465	126,274,467	putative NAC domain-containing protein 94
2	126,282,902	126,288,240	WPP domain-associated protein
2	126,308,111	126,312,250	2-succinylbenzoate--CoA ligase, chloroplastic/peroxisomal
2	126,369,456	126,372,218	60S acidic ribosomal protein P0
2	126,394,984	126,399,350	Ferredoxin--NADP reductase, root isozyme, chloroplastic
2	126,415,541	126,419,312	E3 ubiquitin-protein ligase SINAT2
2	126,424,586	126,428,939	Nuclear transcription factor Y subunit A-7
2	126,445,277	126,445,712	Elongation factor 1-alpha
2	126,467,053	126,470,434	40S ribosomal protein S25
2	126,471,690	126,475,791	50S ribosomal protein L20
2	129,157,277	129,158,026	Ethylene-responsive transcription factor 2
2	129,246,544	129,248,628	WUSCHEL-related homeobox 9
2	129,299,197	129,301,694	Rac-like GTP-binding protein 7
2	129,423,287	129,429,765	28 kDa heat- and acid-stable phosphoprotein
2	129,456,784	129,458,134	Phospholipase A1-II 1
2	129,508,627	129,510,011	Phospholipase A1-II 3
2	129,539,348	129,539,509	Non-specific lipid-transfer protein 1
2	129,626,578	129,628,112	Transcription elongation factor 1 -like protein
2	129,695,551	129,696,736	Trans-resveratrol di-O-methyltransferase
2	129,813,288	129,819,038	Auxin response factor 14
2	129,821,638	129,826,500	MLO-like protein 2
2	129,878,673	129,880,370	Calcium-binding protein CML19
2	129,884,088	129,887,238	Heat stress transcription factor B-1
2	129,956,324	129,961,125	Zeaxanthin epoxidase, chloroplastic
2	130,099,986	130,111,348	Senescence-specific cysteine protease SAG39
2	130,981,862	130,983,639	60S ribosomal protein L32
2	133,238,665	133,242,150	Pentatricopeptide repeat-containing protein, chloroplastic
2	133,707,820	133,708,633	putative 2-oxoglutarate/Fe(II)-dependent dioxygenase
2	133,735,921	133,737,167	Protein SRG1
2	134,274,920	134,275,126	Histone H2B
2	134,301,379	134,307,907	ABC transporter B family member 9
2	134,350,134	134,352,184	Expansin-B15
2	134,358,949	134,362,640	ABC transporter B family member 3
2	134,623,870	134,633,408	Protein MEI2-like 5
2	134,634,802	134,640,517	Phytochrome B
2	134,745,852	134,746,497	Two-component response regulator ARR9
2	134,769,574	134,772,255	Cold-regulated inner membrane protein 2, chloroplastic
2	134,777,566	134,785,295	Carboxyl-terminal-processing peptidase 1, chloroplastic

Chr.	Start	End	Protein feature
2	134,906,279	134,909,750	Coatomer subunit zeta-2
2	134,932,324	134,940,291	Two-pore potassium channel 3
2	134,965,138	134,966,803	putative WRKY transcription factor 71
2	135,036,952	135,041,184	putative purine permease 9
2	135,081,137	135,082,162	putative purine permease 8
2	135,225,525	135,230,147	Cation/H(+) antiporter 16
2	135,358,663	135,369,783	putative starch synthase 4, chloroplastic/amyloplastic
2	135,370,438	135,371,235	Chlorophyll a-b binding protein 3C, chloroplastic
2	135,463,081	135,467,003	LRR receptor-like serine/threonine-protein kinase FLS2
2	135,511,142	135,519,250	B2 protein
2	136,441,694	136,446,849	Protein NSP-INTERACTING KINASE 1
11	2,449,535	2,449,822	Non-specific lipid-transfer protein 2

Appendix 2. Alignment of amino acid sequences of *MEI2-like 5* and *WOX9* between 'CM334' or 'Dempsey' and 'PI159236'

/Nonsynonymous substitution positions were highlighted with red box

(A) *MEI2-like 5*



(A) ***MEI2-like 5***

```
CM334          KNKCNVGYAFINMLSPSLIIPFYEAFNGKKWEKFNSEKVAALAYARIQGKTALVAHFQNS  780
PI159236       KNKCNVGYAFINMLSPSLIIPFYEAFNGKKWEKFNSEKVAALAYARIQGKTALVAHFQNS  780
               ************************************************************

CM334          SLMNEDKRCRPILFHSESSELGDQIVQEHLSSGCVHIQVSQSNESDLVGSQGSPPEEDPV  840
PI159236       SLMNEDKRCRPILFHSESSELGDQIVQEHLSSGCVHIQVSQSNESDLVGSQGSPPEEDPV  840
               ************************************************************

CM334          DKLEKT    846
PI159236       DKLEKT    846
               ******
```

(B) *WOX9*

```
Dempsey       MASSNRHWPSMFKSKPCNSHHHQWQHDINSSLIQPRPPCNQEERSPEPKPRWNPRPEQIR  60
PI159236      MASSNRHWPSMFKSKPCNSHHHQWQHDINSSLIQPRPPCNEEERSPEPKPRWNPRPEQIR  60
              **************************************** ********************

Dempsey       ILETIFNSGMVNPPRDEIRKIRARLQEYGQVGDANVFYWFQNRKSRSKHKQRHVMNLLNS  120
PI159236      ILETIFNSGMVNPPRDEIRKIRARLQEYGQVGDANVFYWFQNRKSRSKHKQRHVMNLLNS  120
              ************************************************************

Dempsey       PTASVNQQNYNNDQFFTTTEQPFFFTVQQPVQTHDNSAMTQGFCFPDSTSSSSGLALSEL  180
PI159236      PTASVNQQNYNNDQFFTTTEQPFFFTVQQPVQTHDNSAMTQGFCFPDSTSSSSGLALSEL  180
              ************************************************************

Dempsey       MGISQTHSSSKKAENEKMNLQSQLMSYTVTSTPTTVSPLISTTTIPTISHIQGASLDPNE  240
PI159236      MGISQTESSSKKAENEKMNLQSQLMSYTVTSTPTTVSPLISTTTIPTISHIQGASLDPNE  240
              ****** ***************************************************

Dempsey       AVGPTRSTVFINDFAFEVGLGPFNVREVFGEDAVLFHSSGEPLITNEWGLTLQPLQHGAF  300
PI159236      AVGPTRSTVFINDVAFEVGLGPFNVREVFGEDAVLFHSSGEPLITNEWGLTLQPLQHGAF  300
              ************.***********************************************

Dempsey       YYLVRTSTPSTHDI    314
PI159236      YYLVRTSTPSTHDI    314
              **************
```

Appendix 3. Genic region sequences of *FT* in 'TF68' and 'Haba'

/Mutated nucleotide sequence were highlighted with red box

```
TF68      ATGCCAAGAGATCCTTTAATTGTTTCTGGAGTTGTTGGAGATGTTGTGGATCCATTCACT  60
Haba      ATGCCAAGAGATCCTTTAATTGTTTCTGGAGTTGTTGGAGATGTTGTGGATCCATTCACT  60
          ************************************************************

TF68      AGGTGTGTGGATTTTGGTGTTGTTTACAACAATAGGGTTGTGGTCTACAATGGATGTGCC  120
Haba      AGGTGTGTGGATTTTGGTGTTGTTTACAACAATAGGGTTGTGGTCTACAATGGATGTGCC  120
          ************************************************************

TF68      TTGAGGCCTTCACAAGTTGTCAATCAACCTAGGGTTGAAATTGGTGGCGGACGATCTTCGC  180
Haba      TTGAGGCCTTCACAAGTTGTCAATCAACCTAGGGTTGAAATTGGTGGTGGACGATCTTCGC  180
          ********************************************** ************

TF68      ACTTTTTACACGCTGGTTATGGTAGACCCTGATGCTCCAAACCCTAGCAACCCAAACCTA  240
Haba      ACTTTTTACACGCTGGTTATGGTAGACCCTGATGCTCCAAACCCTAGCAACCCAAACCTA  240
          ************************************************************

TF68      AGGGAGTATCTACACTGGTTGGTCACAGATATCCCCGCAACTACAGGAGCAAACTTTGGC  300
Haba      AGGGAGTATCTACACTGGTTGGTCACAGATATCCCCGCAACTACAGGAGCAAACTTTGGC  300
          ************************************************************

TF68      AATGAAGTTGTAAGCTACGAGAGCCCACGTCCATCAATGGGAATCCATCGCTATATCTTC  360
Haba      AATGAAGTTGTAAGCTACGAGAGCCCACGTCCATCAATGGGAATCCATCGCTATATCTTC  360
          ************************************************************

TF68      GTGTTGTATCGACAATTGGGCCGCGAGGCGATCAATGCGCCAGACATAATTGATTCACGC  420
Haba      GTGTTGTATCGACAATTGGGCCGCGAGGCGATCAATGCGCCAGACATAATCGATTCACGC  420
          ************************************************** *********

TF68      CAGAATTTTAACACCAGAGATTTTGCTAGGTTTCATAATCTTGGTGTTCCTGTTGCTGCT  480
Haba      CAGAATTTTAACACCAGAGATTTTGCTAGGTTTCATAATCTTGGTGTTCCTGTTGCTGCT  480
          ************************************************************

TF68      GTTTACTTCAATTGCAATAGGGAAGGTGGTACTGGTGGTCGTCGTCTATAA   531
Haba      GTTTACTTCAATTGCAATAGGGAAGGTGGTACTGGTGGTCGTCGTCTATAA   531
          ***************************************************
```