A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Genetic Diversity of Ethiopian *Capsicum* spp and Molecular Mapping of the *Up* Gene in Pepper (*Capsicum spp*.)

에티오피아 고추의 유전적 다양성 분석 및 고추에서의 *Up* 유전자 연관 유전자지도 작성

FEBRUARY 2020

SOLOMON ABATE MEKONNEN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Genetic Diversity of Ethiopian *Capsicum* spp and Molecular Mapping of the *Up* Gene in Pepper (*Capsicum spp*.)

에티오피아 고추의 유전적 다양성 분석 및 고추에서의 *Up* 유전자 연관 유전자지도 작성

FEBRUARY, 2020

SOLOMON ABATE MEKONNEN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Genetic Diversity of Ethiopian *Capsicum* spp and Molecular Mapping of the *Up* Gene 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

SOLOMON ABATE MEKONNEN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY DEPARTMENT OF PLANT SCIENCE

FEBRUARY 2020

APPROVED AS A QUALIFIED DISSERTATION OF SOLOMON ABATE MEKONNEN FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY THE COMMITTEE MEMBERS

CHAIRMAN	**Doil Choi, Ph.D.**
VICE-CHAIRMAN	**Byoung-Cheorl Kang, Ph.D.**
MEMBER	**Cecile Segonzac, Ph.D.**
MEMBER	**Jin Kyung Kwon, Ph.D.**
MEMBER	**Yeong Deuk Jo, Ph.D.**

Genetic Diversity of Ethiopian *Capsicum* spp and

Molecular Mapping of the *Up* Gene in Pepper (*Capsicum spp*.)

SOLOMON ABATE MEKONNEN

Department of Plant Science, Seoul National University

ABSTRACT

Diversity study of *Capsicum* germplasms preserved in the Ethiopian Gene bank and understanding genetic factors that controls the fruit orientation in pepper can provide important information for breeding.

Investigating the extent and structure of crop germplasm diversity is essential for the conservation and utilization of the available genetic resources. Despite the great economic and social importance of pepper in Ethiopia, detailed studies of the genetic diversity using single nucleotide polymorphism (SNP) markers is limited. Thus, with the objective of investigating the variability and genetic structure of Ethiopian pepper germplasms, a total of 142 accessions which were collected and maintained by the Ethiopian Biodiversity Institute were evaluated. We identified and validated 53,284

genome wide SNP molecular markers using genotyping-by-sequencing (GBS). Employing model based population structure, phylogenetic tree and principal coordinate analysis, we identified *C. annuum* and *C. frutescens* as two distinct genetic populations with 132 and 9 accessions, respectively. Besides this, genome wide association (GWAS) analysis detected 509 SNP markers that were significantly associated with fruit, stem and leaf-related traits. Overall, this report is useful to understand the genetic variability existed in Ethiopian *Capsicum* species, for its conservation and breeding.

There are specific known markets for the two major fruit orientation in pepper, i.e. pendant and upright. Pendant and upright position with rare occurrence of intermediate orientation of pepper fruit is a distinguishing morphological variability that can be easily recognized before or after the fruit is set, depending on the population types. Identification of genes from three different bi-parental populations and highly significant SNP positions from GWAS analysis of *Capsicum* core collection enable us to narrow down the mapping region of fruit orientation controlling the locus in chromosome 12. To further develop molecular markers and isolate the candidate gene underlying fruit orientation, fine mapping was performed using *Capsicum annuum* LA F_2 mapping population. The LA F_2 mapping population consisting of 450 individuals was developed from the cross between parental lines *C. annuum* LP97 and A79. Genetic analysis revealed that the phenotype of the up fruit orientation was controlled by a single recessive gene, *Up*. One

hundred fifty SNP markers were used for fine mapping of the *Up* locus. High-resolution genetic mapping of these markers in Karia F_2 mapping population placed Redu0119 and SAR201-1386 at genetic distances of 0.8 and 0.6 cM, respectively, on either side of the *Up* locus. These two marker sequences were aligned to the CM334 genome and the *Up* locus was delimited to a 101 kb genomic region. Fifteen candidate genes were predicted in the target region. Overall, there were a total of 31 conservative amino acid substitutions, due to 18 non-synonymous and 13 synonymous nucleotide switches. The first CDS region of the candidate gene, Zinc finger MYM-type protein 1-like has displayed one non-synonymous nucleotide substitution at 106 bp and one nucleotide deletion at 104 bp. Therefore, Zinc finger MYM-type protein 1-like was the most likely candidate gene for the *Up* locus. The information obtained here will facilitate further research on the trait and breeding of pepper varieties with the early determination of fruit orientation through marker-assisted selection.

Keywords: Ethiopia, fruit orientation, genetic diversity, pepper, pendant, population structure, *Up* locus, upright positon

Student number: 2016-31189

CONTENTS

LIST OF TABLES

Chapter I

Chapter II

LIST OF FIGURES

Chapter I

Chapter II

LIST OF ABBREVIATION

AMOVA	Analysis of molecular variance
AN	Allele number
CDS	Coding sequence
cM	CentiMorghan
COSII	Conserved Ortholog Set II
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
Fst	Fixation index
GA	Genetic advance
GAPIT	Genome association and prediction integrated tool
GBS	Genotyping-by-sequencing
GCV	Genotypic coefficients of variation
GWAS	Genome wide association
HRM	High resolution melting analysis
LD	Linkage disequilibrium

MAF	Major allele frequency
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PCV	Phenotypic coefficients of variation
PIC	Polymorphic information content
QTL	Quantitative trait loci
RIL	Recombinant inbred line
SNP	Single nucleotide polymorphism
ZF	Zinc finger

GENERAL INTRODUCTION

Pepper [*Capsicum* spp. (2x=2n=24)], a genus native to Latin Americas, is one of the oldest cultivated crops having various uses. The most cultivated pepper typically characterized as annual, herbaceous, seed propagated and shrub (CABI, 2017). Pepper fruits are good sources of vitamins A, C and E, folic acid and potassium. As a result of evolution, domestication, and natural and artificial selection, pepper is known to encompass broad diversity (Pereira-Dias et al., 2019). Depending on the variation in pungency, fruit morphology, color and use, peppers can be named as chili peppers, red peppers, green peppers or bell peppers. The morphological, physiological and functional variation in pepper reflects the underlying genetic diversity, which facilitates the adaptation of the five known species of pepper, i.e. *C. annuum*, *C. frutescens*, *C. baccatum*, *C. baccatum* and *C. pubescens* to various environments. Currently, growing in 151 countries in seven continents, pepper is known to be cosmopolitan (CABI, 2017; Ribes-Moya et al., 2018). Understanding the extent and structure of pepper genetic diversity in germplasm collections is essential for the conservation and utilization of the available resources to develop new and improved cultivars with desirable characteristics (Lv et al., 2012).

Numerous detailed studies characterize the genetic, morphological,

phenotypic and cytological diversity as well as the description of genetic structure of *Capsicum* in various parts of the world (Pereira-Dias et al., 2019; Castilla et al., 2019; Lee et al., 2016; Costa et al., 2016; Zhang et al., 2016). For example, Xiao-Zhen and colleagues used 29 simple sequence repeat (SSR) markers to study the genetic diversity of 1,904 pepper accessions conserved at Gene bank located in Beijing, China. Their result identified two different groups of pepper germplasms which are varied in fruit type and geographical distribution (Xiao-Zhen et al., 2019). In another study, using genotyping-by-sequencing (GBS) as a large scale single nucleotide polymorphism (SNP) detection tool, Taranto and colleagues have assessed the level of genetic diversity of 222 cultivated pepper genotypes (Taranto et al., 2016). This group was able to classify the pepper collection into three geographical origin and fruit related features based clusters using Bayesian and Hierarchical clustering, which is derived from 32,950 high quality SNPs. In 2016, 48 transcriptome-based SNP markers were used to explore genetic diversity and population structure of a large pepper germplasm collection comprising 3,821 accessions. In this study *Capsicum* germplasm collection was divided into 10 clusters using population structure and five groups using phylogenetic analysis (Lee et al., 2016), that all based on their taxonomic distinctness and geographical origins. The group identified 240 accessions (~6.3%) as a core collection considering high genetic variation (I=0.95), genetic evenness (J'=0.80) and representativeness of a wider range of phenotypic variation

(MD=9.45%, CR=98.40). A comprehensive study by Naegele and colleagues using 23 SSR markers on the genetic diversity and population structure of fruit traits of *Capsicum annuum* encompassing 114 lines drawn from 44 various countries revealed the existence of significant differences within all fruit phenotype categories (Naegele et al., 2016).

In line with this, so far, few molecular marker based studies were performed to investigate Ethiopian pepper genetic diversity using different germplasm collections (Shiferaw et al., 2018; Geleta et al., 2005). The first molecular study by Geleta et al. (2005) employed amplified fragment length polymorphism markers (AFLP) on 39 pepper genotypes obtained from different countries including seven Ethiopian accessions. Based on the limited genetic resources they used, the research group has come to the conclusion of having narrow genetic base for the pepper fruit morphology in Ethiopia. The other study with 73 representative accessions from four administrative regions in the country, vis. Amhara, Southern Nation, Nationalities and Peoples State (SNNPS), Benishangul Gumuz and Oromia, used inter simple sequence repeat (ISSR) marker to evaluate the genetic diversity (Shiferaw et al., 2018). According to this study, high genetic diversity was observed in some parts of pepper growing regions. Similarly other studies in Ethiopia based on either cultivar collections (Alemu et al., 2016), or mixtures of cultivars and landraces (Alayachew et al., 2017) or landrace panels from distinct geographical regions only (Habtie et al., 2017; Aklilu et al., 2016;

Zegeye et al., 2018) were reported. None of these studies, however, used high resolution SNP marker and performed GWAS. To provide detailed genetic description so that the information can be used for the sustainable use, management and conservation of Ethiopian *Capsicum* species, we evaluated the structure, diversity and genetic relationships of 142 accessions collected from various growing regions of the country using high resolution SNP markers.

Pendant, intermediate and upright flower/fruit positons are the three major fruit orientation related variability of pepper germplasms (IPGRI, 1995). All ornamental peppers, most *C. frutescens* and some of other species display upright phenotype, which is thought to be characteristic of the wild progenitor for the species (Costa et al., 2019; Pessoa et al., 2018: Paran and van der Knaap, 2007). It has been mentioned that erect habit of the fruit is a useful trait for machine-aided mass harvest (Lee et al., 2008). Fruits primarily tend to be pendant when the pedicel carrying them grows curved. Conversely, fruits are in upright position when pedicels grows strait upward from the point of attachment on the node. In some cases, however, due to the plants' loose branching habit, bigger fruit size and extended pedicel length, the fruits can behave laterally pendant or horizontal, irrespective of the pedicel curvature. The general adaxial-abaxial polarity of pedicel and its impact on fruit orientation was documented (Cheng et al., 2016). Hence factors that controls the growth and development of pedicels could directly governs the

fruit orientation.

In a study by Yunandra et al. (2018), pedicel length showed moderate heritability and correlation with days to flowering, fruit length and number of fruit per plants. In Arabidopsis, *CORYMBOSA1*/auxin transport protein BIG (CRM1/BIG) encodes a membrane-associated protein which is required for auxin transport and is important for the developmental changes in pedicel cell growth (Yamaguchi and Komeda, 2013). These researchers also showed the role of *CYCLIN B1* expression in cell proliferation of *Arabidopsis* pedicel, and described member of class I knotted1-like homeobox gene family controls pedicel orientation. In another study, knockdown of SVP family MADS-box gene, *NtSVP* by RNA interference (RNAi) caused elongated pedicel, while overexpression resulted in compact inflorescences with much shortened pedicels in tobacco (Wang et al., 2015).

In case of pepper, Setiamihardja and Knavei (1990) reported the positive correlation of pedicel length with fruit length; and pedicel diameter with fruit diameter. They also showed that the pendant types have longer and narrower pedicel than upright pepper types. The mapping study of Chaim et al. (2001) using F_3 population derived from a cross between bell type and small fruited lines by employing restriction fragment linked polymorphism (RFLP), amplified fragment linked polymorphism (AFLP), random amplification of polymorphic DNA (RAPD) and morphological markers revealed the presence of three and four QTLs for length and diameter of pedicels, respectively.

Using Saengryeog 211 (pendant), Saengryeog 213 (erect), and their F_1 and BC_1 progeny, Lee et al. (2008) demonstrated that *Up* is a recessive gene. In this study, bulked segregant analysis (BSA) and AFLP were employed using 108 $F_{2:3}$ individuals and an *Up* marker was developed at a genetic distance of 1.7 cM and 4.3 cM from the locus, respectively. Similarly an ultra-high density bin mapping using RIL population detected a major QTL controlling fruit orientation *FP-12.2* located at 199.6 Mb on chromosome 12 in *Capsicum annuum* cv. CM334v.1.55 reference genome (Han et al., 2018). In 2016 group reported a high density interspecific genetic map with 5,569 SNPs which was constructed using 297 F_2 individuals. Based on this genetic map, they detected one major QTL, named *Up*12.1 for fruit orientation and predicted a total of 65 protein-coding genes within the QTL region based on Zunla-1 reference genome (Cheng et al., 2016). Following these results, by using segregating populations and diversity panels, we aimed at fine-mapping of the fruit orientation controlling the *Up* locus to identify candidate genes and developing closely linked markers for breeding applications.

Therefore, in this study I have included two topics which can provide useful information for pepper breeders; morpho-molecular diversity of Ethiopian peppers germplasms and molecular mapping of fruit orientation controlling gene.

REFERENCE

Aklilu S, Abebie B, Wogari D, Adefris TW (2016) Genetic variability and association of characters in Ethiopian hot pepper (*Capsicum annuum* L.) landraces. J Agric Sci 61(1): 19-36.

Alayachew SA, Atnafu DM, Gedefa SA (2017) Genetic diversity study of Ethiopian hot pepper cultivars (*Capsicum* spp.) using inter simple sequence repeat (ISSR) marker. J Plant Mol Breed 5(2): 27-37

Alemu A, Wodajo A, Chuntal K (2016) Performance evaluation of elite hot pepper (*Capsicum annuum*) varieties for yield and yield components at Derashea, South Eastern Ethiopia. Int J Res 4(12): 95-100

CABI (2017). *Capsicum annuum* (bell pepper). In: Invasive species compendium, CABI Publishing, Wallingford, UK.

Castilla LC, Hernández RG, Aguilar CCC, Martínez-Hernández A, et al (2019) Structure and genetic diversity of nine important landraces of *Capsicum* species cultivated in the Yucatan Peninsula, Mexico. Agron J 9(376): 1-11

Chaim AB, Paran I, Grube RC, Jahn M (2001) QTL mapping of fruit related traits in pepper (*Capsicum annuum*). Theor Appl Genet 102: 1016-1028

Cheng J, Win C, Tang X, Zhou H, et al (2016) Development of a SNP array and its application to genetic mapping and diversity assessment in pepper

(*Capsicum* spp.). Sci Rep 6: 1-11

Costa GDN, Da Silva BMP, Lopes ACD, Carvalho LCB, et al (2019) Selection of pepper accessions with ornamental potential. Rev Cattinga 32(2): 566-574

Costa MP, do Rego MM, da Silva AP, do Rego ER, et al (2016) Characterization and genetic diversity of pepper (*Capsicum* spp) parents and interspecific hybrids. Genet Mol Res 15(2): 1-12

Geleta LF, Labuschagne M, Viljoen CD (2005) Genetic variability in pepper (*Capsicum annuum* L.) estimated by morphological data and amplified fragment length polymorphism markers. Biodivers Conserv 14(10): 2361-2375

Habtie B, Dejen T, Dessalegn Y (2017) Morphological characterization of hot pepper (*Capsicum annuum* L.) land races of Ethiopia for qualitative characters. Int J Res Stud Sci Eng Tech 4(9): 4-9

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

IPGRI, AVRDC, CATIE (1995) Description for *Capsicum* (*Capsicum* spp.). International Plant Genetic Resources Institute, Rome, Italy; the Asian Vegetable Research and Development Center, Taipei, Taiwan, and the Centro Agronomico Tropical de Investigacion y Ensenanza, Turrialba, Costarica

Lee HR, Cho MC, Kim HJ, Park SW, et al (2008) Marker development for erect versus pendant-oriented fruit in *Capsicum annuum* L. Mol Cells 26: 548-553

Lee HY, Ro NY, Jeong HJ, Kwon JK, et al (2016) Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. BMC Genet 17(1): 142

Lv Q, Qi J, Shi Q, Shen D, et al (2012) Genetic diversity and population structure of cucumber (*Cucumis satuvus* L.) PloS One 7(10): e46919. https://doi.org/10.1371/journal.pone.0046919

Naegele RP, Mitchell J, Hausbeck MK (2016) Genetic diversity, population structure, and heritability of fruit traits in *Capsicum annuum*. PloS One 11(7): e0156969. https://doi.org/10.1371/journal.pone.0156969

Paran I, der Knaap E (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. J Exp Bot 58(14): 3841-3852

Pereira-Dias L, Vilanova S, Fita A, Prohens J, et al (2019) Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Hortic Res 6(54): 1-13

Pessoa AMS, do Rêgo ER, de Carvalho MG, dos Santos CAP, et al (2018) Genetic diversity among accessions of *Capsicum annuum* L. through morphoagronomic characters. Genet Mol 17(1): 1-14

Ribes-Moya A, Raigon MD, Moreno-Peris E, Fita A, Rodriguz-Burruezo A (2018) Response to organic cultivation of heirloom *Capsicum* peppers: Variation in the level of bioactive compounds and effect of ripening. PloS One 13(11): e0207888. https://doi.org/10.1371/journal.pone.0207888

Setiamihardja R, Knavei DE (1990) Association of pedicel length and diameter with fruit length and diameter and ease of fruit detachment in pepper. J Amer Soc Hort Sci 115(4): 677-681

Shiferaw AA, Destaw MA, Samuel AG (2018) Genetic diversity study of Ethiopian hot pepper cultivars (*Capsicum* species*)* using inter simple sequence repeat (ISSR) marker. Int J AgriSci 13(2): 65-73

Sun J, Xiao T, Nie J, Chen Y, et al (2019) Mapping and identification of *CsUp*, a gene encoding an Auxilin-like protein, as a putative candidate gene for the upward-pedicel mutation (*up*) in cucumber. BMC Plant Biol 19(157): 1-12

Taranto F, D'Agostino N, Greco B, Cardi T, et al (2016) Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. BMC Genomics 17(943): 1-13

Wang D, Chen X, Zhang Z, Liu, D, et al (2015) A MADS-box gene *NtSVP* regulates pedicel elongation by directly suppressing a KNAT1-like KNOX gene *NtBPL* in tobacco (*Nicotiana tabacum* L.). J Exp Bot 66(20):6233-6244

Xiao-zhen Gu, Ya-cong CAO, Zheng-hai Z, Bao-xi Z, et al (2019) Genetic diversity and population structure analysis of *Capsicum* germplasm accessions. J Integr Agric 18(5): 970-981

Yamaguchi N, Komeda Y (2013) The role of CORYMBOSA1/BIG and auxin in the growth of *Arabidopsis* pedicel and internode. Plant Sci 209: 64-74

Yunandra, Syukur M, Maharijaya A (2018) Inheritance study for yield components of pepper (*Capsicum annuum* L.). Earth Environ Sci 196: 1-8

Zegeye BH, Dejene T, Dessalegn Y (2018) Genetic variability and heritability of pod yield and related traits of hot pepper (*Capsicum annuum* L.) landraces in West and North West Ethiopia. Int J Plant Breed and Crop Sci 5(3): 408-414

Zhang X, Zhang Z, Gu X, Mao S, et al (2016) Genetic diversity of pepper (*Capsicum* spp.) germplasm resource in China reflects selection for cultivar types and spatial distribution. J Integr Agric 15(9): 1991-2001

CHAPTER I

Genetic Diversity of Ethiopian *Capsicum* spp.

The research described in this chapter has been published in *Plos One*.

Solomon AM, Han K, Lee JH, Lee HY, Jang S, Kang BC (2019) Genetic diversity and population structure of Ethiopian *Capsicum* germplasms. PloS One 14(5): e0216886. https://doi.org/10.1371/journal.pone.0216886

ABSTRACT

We established a collection of 142 *Capsicum* genotypes from different geographical areas of Ethiopia with the aim of capturing genetic diversity. Morphological traits and high-resolution melting analysis distinguished 132 *Capsicum annuum*, nine *Capsicum frutescens* and one *Capsicum baccatum* accessions in the collection. Measurement of plant growth parameters revealed variation between germplasms in plant height, stem thickness, internode length, number of branches, fruit width, and fruit length. Broad-sense heritability was maximum for fruit weight, followed by length and width of leaves. We used genotyping by sequencing (GBS) to identify single-nucleotide polymorphisms (SNPs) in the panel of 142 *Capsicum* germplasms and found 2,831,791 genome-wide SNP markers. Among these, we selected 53,284 high-quality SNPs and used them to estimate the level of genetic diversity, population structure, and phylogenetic relationships. From model-based ancestry analysis, the phylogenetic tree, and principal-coordinate analysis (PCoA), we identified two distinct genetic populations: one comprising 132 *C. annuum* accessions and the other comprising the nine *C. frutescens* accessions. GWAS analysis detected 509 SNP markers that were significantly associated with fruit-, stem- and leaf-related traits. This is the first comprehensive report of the analysis of genetic variation in Ethiopian

Capsicum species involving a large number of accessions. The results will help breeders utilize the germplasm collection to improve existing commercial cultivars.

INTRODUCTION

Members of the genus *Capsicum* in the Solanaceae family, commonly known as chilli peppers, are major crop plants and are almost cosmopolitan in distribution (Panda et al., 2004). Chilli pepper fruits are used as spices, as vegetables and for medicinal purpose (Zonneveld et al., 2015) and are a significant source of Vitamins A and C. They are also used as natural coloring agents, cosmetics and active ingredient in host defense repellents. Some are also used as ornamentals (Geleta et al., 2005; Taranto et al., 2016). The genus includes 27 species, of which five are known to be domesticated (Ince et al., 2010). The five cultivated species of *Capsicum*, namely *C. annuum* L., *C. chinense* Jacq., *C. frutescens* L., *C. baccatum* L., and *C. pubescens* Ruiz et Pav., represent the most economically important vegetables worldwide (Gonzalez-Perez et al., 2014).

The independent diffusion and subsequent domestication of the five cultivated species of *Capsicum* from its primary center of cultivation, Bolivia and Peru (Zonneveld et al., 2015), to Europe by Columbus, and from there to Africa, India and China resulted in the crop's diversification due to human activity and in the development of non-deciduous, pendant, larger and non-pungent fruits with greater shape variation and increased fruit mass in several *Capsicum* growing areas (Gonzalez-Perez et al., 2014). Characterization of genetic diversity in various *Capsicum* species growing in different regions

has been reported (Lee et al., 2016). It has also been shown that ecological distribution has a significant influence on the genetic diversity of plants, including *Capsicum* (Albrecht et al., 2012).

According to Geleta et al. (2005), *Capsicum* was introduced to Ethiopia first by the Portuguese in the 17th century and subsequently from all over the world; they have since been cultivated for centuries and adapted to various agro-ecological zones. Such situations contributed to the evolution of local Ethiopian genotypes with different fruit types, pungency levels and disease resistance (Marame et al., 2009). Adaptation and cultivation in the wide range of agro-ecologies of Ethiopia may have led to the development of accessions with great variation in many important traits. In this regard, an Ethiopian origin small fruited, pungent *C. annuum* inbred line, H3, was reported to be the most important and persistent source of powdery mildew (*Leveillula taurica*) resistance and has been used to breed resistant varieties (Venkatesh et al., 2017). Different pepper cultivars have been produced and used as spice and vegetable crops ever since the first introduction of the genus. Pepper is consumed in many different forms in Ethiopia. The green fruit, known locally as “karia”, is eaten raw as a salad, and the dried red fruit is ground into powder and added to a sauce known as “wot”. Eating hot pepper is a deeply rooted Ethiopian food habit, and hot peppers are cultivated over more than 246,000 ha in Ethiopia (Zewdie, 1998). Owing to its centrality in the daily diet of most

Ethiopian societies, hot pepper plays an important role in the national economy. Though the number of local collections was small and investigations were based mainly on morphological assessment, with only limited use of genetic markers, previous work by Shimelis et al. (2016), Shumbulo et al. (2017), Marame et al. (2009) and Geleta et al. (2005) on Ethiopian *Capsicum* species indicated the existence of considerable genetic variability.

Germplasm diversity is crucial to successful breeding programs. Such diversity is important for broadening the genetic base, as it increases the probability of finding more unique genes for which two parents have different alleles (that is, the genetic distance). Numerous methods have been used to estimate genetic diversity among *Capsicum* genotypes, including multivariate analyses of large numbers of phenotypic descriptors (Geleta et al., 2005; Lee et al., 2016; Jeong et al., 2010) as well as of cytological (Panda et al., 2004), biochemical (Popovsky-Sarid et al., 2017) and molecular variations (Lee et al., 2016; Tsaballa et al., 2015). Since most morphological traits are polygenic and their expression depends on environmental factors, among others, the use of molecular markers is the most suitable method for estimating genetic diversity due to its ability to recognize specific DNA sequences in closely related genotypes, irrespective of growth stage, time, place and agronomic practices. Various types of DNA markers, such as restriction fragment length

polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphism (AFLPs) and microsatellite repeats (or simple sequence repeats, SSRs), can be used to determine relationships and genetic variation levels in wild and domesticated crops, including *Capsicum* spp. (Gonzalez-Perez et al., 2014). Next-generation sequencing (NGS)-based genotyping methods have recently been used for whole-genome sequencing and for re-sequencing projects. In NGS, the genomes of several specimens are sequenced to discover large numbers of single-nucleotide polymorphisms (SNPs) that can be used to explore within-species diversity, construct haplotype maps and perform genome-wide association studies (GWAS) (Elshire et al., 2011). NGS has made routine screening of plant germplasm feasible and cost-effective (Wong et al., 2015). The use of SNP markers provides the most useful information for detecting genetic diversity and determining genetic relationships between lines, owing to the abundance of SNPs in plants, to the technique's flexibility, low error rate, high speed of detection and cost-effectiveness and to the ease with which the resulting data can be converted to universal genotype information from different technological sources (Taranto et al., 2016). Genotyping by sequencing (GBS) is a genome-wide reduced representation of SNPs obtained using Illumina sequencing technology (Elshire et al., 2011). The use of restriction enzymes in GBS reduces genome complexity by avoiding the sequencing of repetitive regions, resulting in more straightforward bioinformatics analysis for large

genomes (Wong et al., 2015; Han et al., 2018). It is thus a rapid, high-throughput, genome-wide and cost-effective tool for SNP discovery (Elshire et al., 2011). It is helpful for genotyping without prior knowledge about the genome of the species and is useful for exploring plant genetic diversity on a genome-wide scale (Taranto et al., 2016). In the last few years, GBS has been used to investigate the genetic diversity of many crop species, including maize, rice, barley, tomato, wheat, sorghum, soybean, watermelon and *Capsicum* (Ertiro et al., 2017; Tang et al., 2016; Poland et al., 2012; Pailles et al., 2017; Zhang et al., 2015; Iquira et al., 2015; Nimmakayala et al., 2014; Nimmakayala et al., 2016).

The present study was undertaken to characterize *Capsicum* germplasms collected from different localities of the six regions of Ethiopia using morphological and molecular markers to explore the genetic diversity available in a wide collection of germplasm. The data presented herein may be useful to understand the diversity of *Capsicum* in Ethiopia and use the information for the breeding purpose. Similarly, our finding may give additional insight into the quantitative trait loci controlling fruit weight.

MATERIALS AND METHODS

Plant materials

The germplasm collection of 142 genotypes used in this study was obtained from the Ethiopian Biodiversity Institute (EBI). These germplasms were collected from different pepper-growing areas of the country: 47 from eight zones of Amhara (11° 39′ 38.88″ N, 37° 57′ 28.08″ E); five from the Metekel zone of Benishangul Gumuz (10° 20′ 0″ N, 34° 40′ 0″ E); 38 from eight zones of Oromia (7° 59′ 20.62″ N, 39° 22′ 52.25″ E); 40 from five zones of the Southern Nations, Nationalities and Peoples (SNNPs) region (6° 3′ 31.03″ N, 36° 43′ 38.28″ E); one from the Jigjiga area of Somali (7° 26′ 19.43″ N, 44° 17′ 48.75″ E); and two from different weredas (districts) of Tigray (14° 8′ 11.68″ N, 38° 18′ 33.58″ E) (Fig 1). Nine germplasms used in this experiment have no accession passport data and are thought to be recent introductions. Among the germplasms, 42 were classified into particular species of *Capsicum* by the EBI based on the descriptor: four as *C. frutescens* and 38 as *C. annuum*. According to the germplasm descriptors, the EBI collected peppers in Ethiopia during 16 different years since the first germplasm record from the Limu wereda of the Oromia region in 1978. The majority of *Capsicum* germplasm collection (59%) was done between 1984 and 1990, during which most geographical areas with significant pepper cultivation were covered. Seeds were disinfected using 2% sodium chlorate and 10% trisodium phosphate. In this experiment, three plants of each of the

142 germplasms were grown under greenhouse conditions at Biotong Seed Co. Ltd., Anseong, Republic of Korea, in 2017.

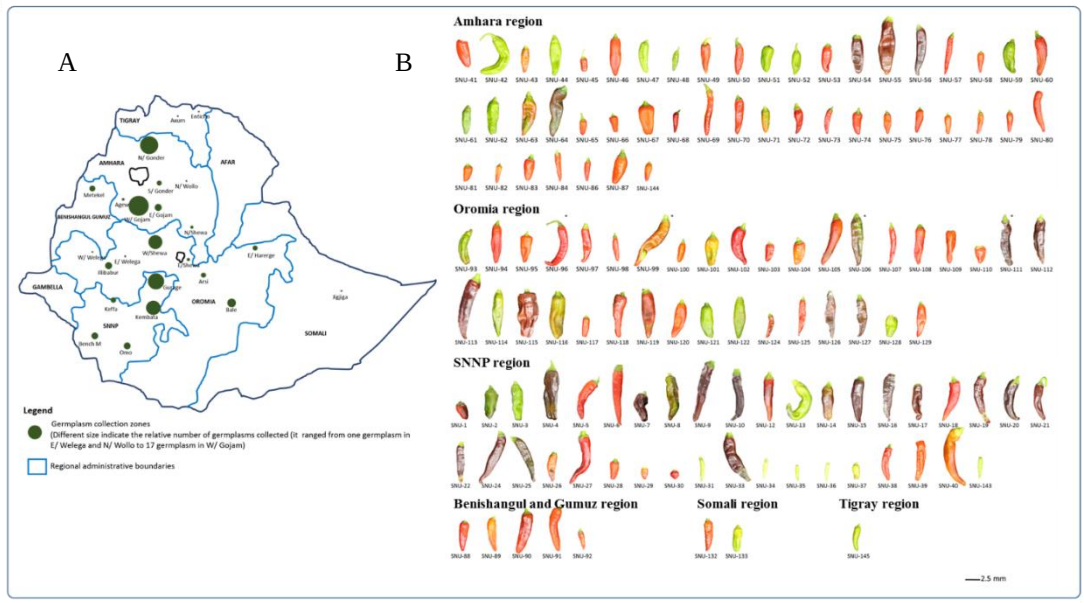


Figure 1. Map of germplasm collection areas of Ethiopian *Capsicum* and associated fruit variation. (**A**) Germplasm collection regions were Amhara (47 accessions); Benishangul Gumuz (5 accessions); Oromia (38 accessions); Southern Nations, Nationalities and Peoples (SNNPs) (40 accessions); Somali (1 accession); Tigray (2 accessions) and unknown (9 accessions). Green circles indicate the relative number of accessions. (**B**) Morphology of representative fruits of pepper accessions collected from different regions.

Morphological characterization and statistical analysis

For each germplasm, 12 growth, nine flower-, 17 fruit and two seed-related traits were evaluated according to the descriptions used by the Rural Development Administration (RDA) gene bank, South Korea, with some modifications. Morphological traits such as plant type (PT), plant height (PH), plant width (PW), main stem length (MSL), internode length (INL), number of side branches (NSB), stem thickness (ST), stem color (SC), leaf color (LC), and leaf length and width (LL and LW) were recorded 114 days after sowing. One representative flower from each plant was assessed for stamen number (SN), filament color (FC), anther color (AC), petal color (PC), petal length (PL), petal width (PW), petal number (PN) and calyx shape (CS).

Tomato analyzer, a software application to collect morphological and colorimetric data from two dimensional objects, version 3.0 (Rodriguez et al., 2010) was used to measure fruit perimeter (FP), fruit area (FA), fruit width mid height (FWMH), maximum fruit width (MFW), fruit height mid-width (FHMW) and fruit curved height (FCH) as previously described (Naegele et al., 2016). A Spearman's rank correlation coefficient was calculated among all the variables, including the altitude at which the germplasms were collected. Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were performed using SPSS software 22.0 to measure sampling adequacy and sphericity, respectively (Sacco et al., 2015).

The phenotypic (PCV) and genotypic (GCV) coefficients of variation were estimated as percentages of the corresponding phenotypic and genotypic standard deviations from the trait grand means as used by Khan et al. (2016). Estimates of broad-sense heritability in percent were obtained using the formula suggested by Burthon and de Vane (1953).

DNA extraction and library construction for genotyping by sequencing

Two or three young leaves from each germplasm were used as source of DNA. Total genetic DNA was extracted using the modified cetyl trimethylammonium bromide (CTAB) method as described previously (Clarke, 2009). The concentration and purity of DNA samples were determined with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). DNA samples with absorbance ratios above 1.8 at 260/280 nm were used for analysis (Jeong et al., 2010), and gel electrophoresis was conducted on a 0.8% agarose gel.

Two GBS libraries were constructed based on a modified protocol as used previously (Wong et al., 2015; Annicchiarico et al., 2017) using a two-enzyme system, *Pst*I (rare cutter) and *Mse*I (frequent cutter).

HRM-PCR amplification and data analysis

To the identify the species of 142 *Capsicum* germplasms, high-resolution melting (HRM) was performed as described previously (Jeong et al., 2010; Song et al., 2016). A Rotor-Gene 6000 real-time PCR thermocycler (Corbett Research, Sydney, Australia) was used with the following PCR amplification conditions: 95°C for 10 min; 50 cycles of 94°C for 20 s, 55°C for 20 s, and 72°C for 40 s; 95°C for 60 s; and 40°C for 60 s. For HRM analysis, an increase of 0.1°C temperature per minute from 65°C to 90°C was used. A combination of five markers (Table 1) developed previously was implemented for species identification (Jeong et al., 2010). Five reference materials, viz. *C. annuum*, *C. chinense*, *C. frutescens*, *C. chacoense* and *C. baccatum*, were used.

Table 1 HRM markers used for species identification (Jeong et al., 2010)

Marker name	GeneBank Acc no.	Primer	Sequence	Tm	Product size
C2_At5g04590	GU90688	Forward	TTTGGAGTTCATGGTTATGG	55.9	457
		Reverse	ACACAACAAATGGGGTAGG	55.7	
C2_At1g50020	GU906107	Forward	AGTTAGCGGGCTAAATATGC	56.3	279
		Reverse	GACCACAAGGATCTATTGTACG	56.2	
C2_At2g19560	GU906118	Forward	TGGAATGTCAGAAGGGTTTGTTG	55.5	125
		Reverse	TTTCATGAGAAATGAGCCAGCAC	55.8	
PepTrn	HM209440	Forward	GAGCAAGGAATCCCTAGTTG	55.0	451
	HM209444	Reverse	GGATTTTCAGGGGTATACCAA	55.0	
HRM Waxy	AF397131	Forward	GAACTTAGGTTCAGCTTGTTGTGTC	55.0	286
		Reverse	ATGAAGAGAACATCCTCTCCTGC	55.0	

Marker name	**GeneBank Acc no.**	**Primer**	**Sequence**	**Tm**	**Pr siz**
C2_At5g04590	GU90688	Forward	TTTGGAGTTCATGGTTATGG	55.9	
		Reverse	ACACAACAAATGGGGTAGG	55.7	
C2_At1g50020	GU906107	Forward	AGTTAGCGGGCTAAATATGC	56.3	
		Reverse	GACCACAAGGATCTATTGTACG	56.2	
C2_At2g19560	GU906118	Forward	TGGAATGTCAGAAGGGTTTGTTG	55.5	
		Reverse	TTTCATGAGAAATGAGCCAGCAC	55.8	
PepTrn	HM209440	Forward	GAGCAAGGAATCCCTAGTTG	55.0	
	HM209444	Reverse	GGATTTTCAGGGGTATACCAA	55.0	
HRM Waxy	AF397131	Forward	GAACTTAGGTTCAGCTTGTTGTGTC	55.0	
		Reverse	ATGAAGAGAACATCCTCTCCTGC	55.0	

Sequencing data analysis, SNP identification and genome-wide association analysis

Sequencing was performed with an Illumina HiSeq 2500 (Macrogen Inc., Seoul, Korea). Data analysis and SNP identification were performed as described previously (Lee et al., 2017). Raw reads were de-multiplexed in accordance with individual barcodes, and the adapter and barcode sequences were removed using commercially available CLC genomic workbench software (version 6.5). Trimmed reads were mapped to CM334 chromosome version 1.6 (Pepper.v.1.6.total.chr.fa) by Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2010). The SAMtools program was used to group and sort the reads by chromosomal order (Li et al., 2009). The Genome Analysis Toolkit (GATK) program was used to call SNPs over whole chromosomes McKenna et al., 2010). From the 142 total germplasms, 12 were omitted from SNP-based analysis due to significant loss of reads.

The 53,284 filtered SNPs based on criteria, MAF >0.05, SNP coverage >0.6 and IF>0.8, were used for genome-wide association (GWAS) mapping. The default settings of Genomic Association and Prediction Integrated Tool of the R package were used to estimate GWAS based on the compressed mixed linear model (Lipka et al., 2012). SNPs with a calling rate of more than 0.1 were retained and FILLIN in TASSEL was used for imputation. R^2 values and imputed ratio of minor and major alleles were used to select suitable imputed quality. A final filtering was performed based on minor allele

frequency of more than 0.05, SNP coverage >0.6 and inbreeding coefficient >0.8. The P values of SNPs from GWAS were subjected to a false-discovery rate (FDR) analysis, and Bonferroni correction was done to reduce false-negative results from the GWAS analysis. A significance threshold level at a P value of 0.05 was set after Bonferroni multiple-test correction.

Genetic diversity and population structure analysis

For each SNP, polymorphic information content (PIC), heterozygosity (H2), gene diversity, genotype number, allele number and allele frequency were calculated using Power Marker software (Tsaballa et al., 2015), and the genetic diversity for the entire set of *Capsicum* genotypes as well as the geographically based subpopulations were also identified by PowerMaker version 3.25. To investigate the population structure, assess genetic diversity and remove near-duplicates (i.e., highly similar genotypes), both parametric and non-parametric approaches were used. Pairwise geographic distances between accessions, pairwise FST between accessions in the different groups and analysis of molecular variances (AMOVAs) were calculated using GenAlEx 6.503, with 999 permutations for testing variance components (Peakall and Smouse, 2012).

Population structure was estimated from 13,998 representative SNPs from the 53,284 polymorphic SNPs used in GWAS analysis as used previously (Kumbhar et al., 2015). Population structure was determined using

STRUCTURE software (http://pritch.bsd.uchicago.edu/structure.html) (Excoffier and Lischer, 2010), which was run from the command line using the admixture model, a burn-in period length of 10,000 and 10,000 Markov-chain Monte Carlo (MCMC) iterations after burn-in. Ten independent runs were performed for each K from K = 1 to K = 5. The best number of K was chosen with the DeltaK method (Pailles et al., 2017) by running the STRUCTURE HARVESTER software (Earl and vonHoldt, 2012).

Phylogenetic and principal-coordinate analyses

Phylogenetic trees were produced using genotyping data with 53,284 SNP markers using both the unweighted neighbor-joining method and the hierarchical cladding method based on the dissimilarity matrix calculated with Manhattan index, as implemented in the DARwin software (version 6.0.9) (Perrier and Jacquemound, 2006), and was visualized with Dendroscope (version 3.5.9) (Huson et al., 2007). Inkscape 0.92 was used to make annotations and to apply visual effects to the phylogenetic tree. Principal-coordinate analyses (PCoA) were performed with GenAlEx version 6.503 (Peakall and Smouse, 2012).

RESULTS

Species identification based on HRM genotyping

In addition to the preexisting identification of some of the germplasms by EBI, and the assessment of morphological features (Jeong et al., 2010) and SNP information (Fig 1 and 4), we obtained further confirmation of species through a real-time HRM-PCR protocol as described earlier (Song et al., 2016). We used five high-resolution melting markers (HRMs) to assign germplasms to different species (Table 1) (Jeong et al., 2010). The 142 germplasms were classified into three species, *C. annuum*, *C. frutescens* and *C. baccatum*. The four markers C2_At5g19560, C2_At5g50020, Waxy and PepTrn showed a polymorphism that was highly specific for the three species (Fig 2). Eighty-six *C. annuum* accessions were identified by the C2_At5g50020 marker, of which 12 were further confirmed by both Waxy and PepTrn, 28 by Waxy and 33 by PepTrn. The remaining 46 *C. annuum* accessions were identified by the single markers C2_At5g19560 (two), C2_At5g50020 (13), Waxy (eight) and PepTrn (32). Confirmation of *C. baccatum* status was made on the basis of the specific melting curve shape using PepTrn and C2_At5g50020 markers. *C. frutescens* accessions were likewise identified on the basis of melting curve shape using the markers Waxy and PepTrn. C2_At5g50020 also identified two *C. frutescens* accessions (SNU-142 and SNU-143).

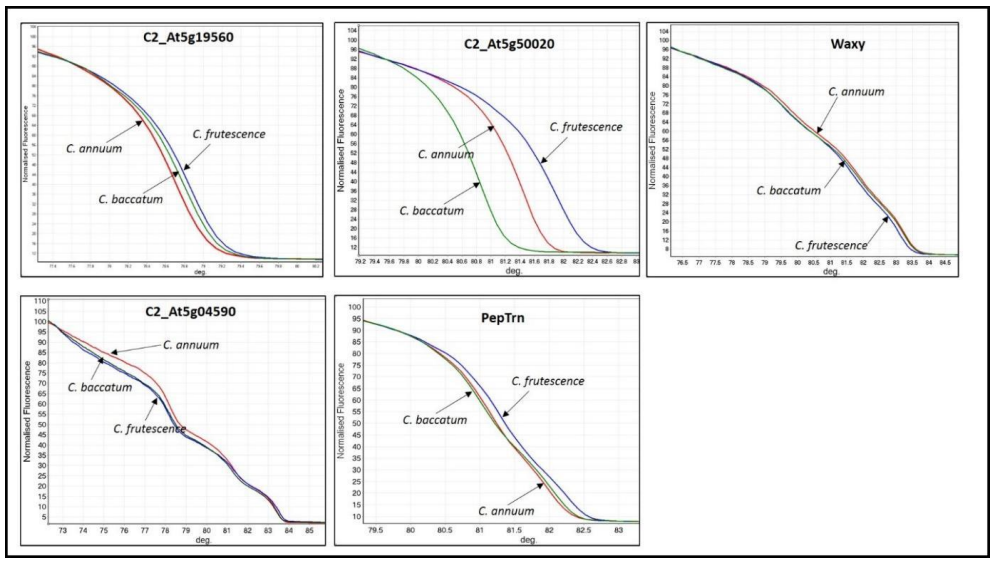


Figure 2. Melting curve analysis (HRM) developed from Conserved Ortholog Set II (COSII). The COSII markers used are: (a) C2_At5g1956, (b) C2_At5g50020, (c) C2_At5g04590, (d) Waxy and (e) PepTrn.

Qualitative and quantitative morphological characterization

We evaluated the qualitative properties of plants, flowers, leaves and fruits based on descriptions used by Rural Development Administration (RDA) gene bank of South Korea. Plant habits were spreading (12.9%), half-spreading (59.7%), erect (26.6%) and fasciculate (0.7%). Spreading types were collected from Amhara, Oromia and SNNPs at altitudes ranging between 1,000 and 2,570 meters above sea level (m.a.s.l.), while the altitude range for half-spreading plant types was 1,150–2,780 m.a.s.l. distributed in different locations of Amhara, Benishangul Gumuz, Oromia, SNNPs, Somalia and Tigray. A relatively narrower altitude range (1,200–2,060 m.a.s.l.) was observed for the erect *Capsicum* types from Amhara, Oromia, SNNPs and Tigray. The only fasciculate *Capsicum* type was obtained from the Bibugn wereda of Amhara at an altitude of 1,850 m.a.s.l.

Measurement of plant growth parameters revealed variation between germplasms. Plant height (48–175 cm), stem thickness (6–33 mm), internode length (4–18.5 cm) and number of side branches (4–30) all evidenced variation. The seven tallest germplasms (140–175 cm in height) were collected from SNNPs at a mean altitude of 1,300 m.a.s.l. Short to medium-height germplasms were collected from all *Capsicum* growing regions, with a wider altitude range (1,000–2,570 m.a.s.l.). Wider variation in stem thickness was observed in *C. annuum* (6.32–33 mm) than in *C.*

frutescens (7.2–15.7 mm). Mean stem thickness for *C. baccatum* was 23.3 mm. The maximum mean internode length (11.4 cm) was measured for *C. frutescens*, while the minimum (9.23 cm) average internode length was observed for *C. annuum*. Similarly, the maximum average number of side branches below first node (17, with a range of 12–21) was counted in *C. frutescens*, followed by *C. annuum* (15, with a range of 4–30) (Table 2 and Fig 3).

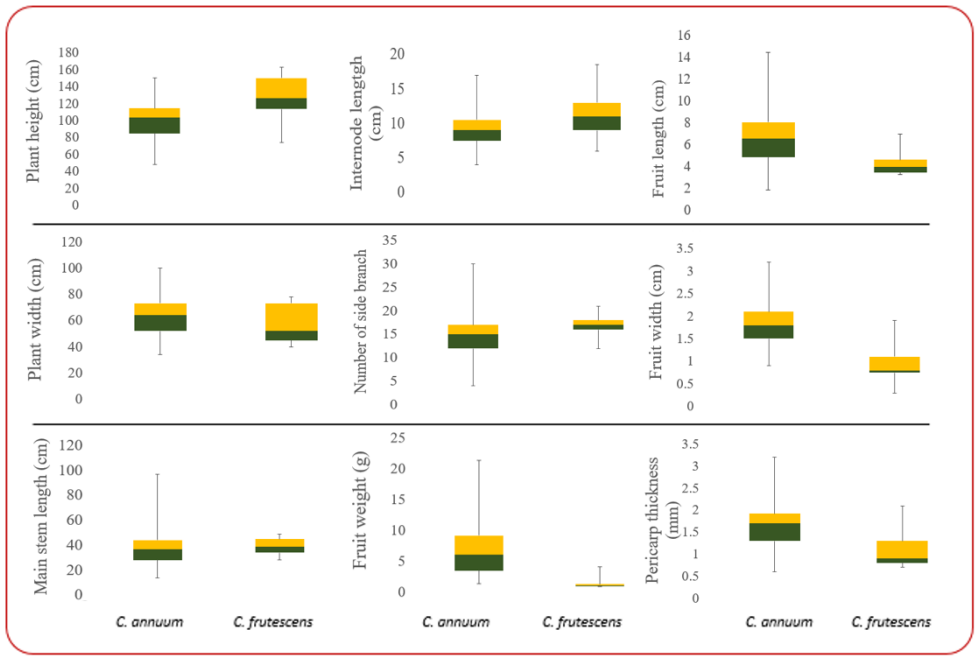


Figure 3. Box plot of nine selected phenotypes of *C. annuum* and *C. frutescens.* While the green boxes represent first quartile (Q1), yellow boxes show the third quartile (Q3). The intersection between the two colored boxes in each figure is median. The minimum and maximum values for each trait is shown by the error bar.

Table 2. Summary of agronomic and morphological traits exhibiting high correlation with the estimated ancestry membership coefficients (Q) in the STRUCTURE software

Species	No. of Acc.	Range (min – max)				Corolla color	Stem color
		Plant height (cm)	Internode length (cm)	Fruit length (cm)	Fruit width		
C. annuum	121	48-175	4-17	1.8-14.4	0.9-3.2	White, yellowish and purple	Green, green with purple stripe & with purple dot; and purple
C. frutescens	9	74-163	6-18.5	3.2-6.9	0.3-1.9	White and greenish yellow	Green and green with purple stripe
C. baccatum	1	175	13	7.7	1.5	White with yellow spots in the center	Green

Other morphological traits similarly showed variation (Table 2; Fig 1 and Fig 4). Most germplasms (97.2%) had one flower per axil, while the remaining four germplasms, which were collected from Amhara and SNNPs (SNU17-2, SNU17-13 and SNU17-76 from *C. annuum* and SNU17-34 from *C. frutescens*), had two. The maximum measured fruit length was 14.4 cm and the minimum was 1.8 cm, each found for one accession. Respectively, 24%, 22%, 21%, 19%, 18% and 12% of germplasms had fruit lengths measuring 8, 5, 4, 7, 6 and 9 cm. The minimum fruit length score was 3 cm (seen for seven germplasms). Fruit width ranged between 0.3 and 3.2 cm. The majority of germplasms (39%) had fruit widths measuring 2 cm, with 35% measuring 1.5 and 22% measuring 2.5 cm. The fruit shape index, which is the ratio of fruit length to width, ranged between 1.5 and 13. Whereas the fruits of 73.5% of germplasms have two locules, 24.3% had three locules. Only three germplasms (SNU17-67, SNU17-133 and SNU17-140) displayed four locules. Fresh weights of fruits varied between 1 and 22 g. Seed number per fruit ranged from zero for one germplasm (SNU17-40, *C. annuum*) to over 175 for two other germplasms of *C. annuum* (SNU17-102 and SNU17-116).

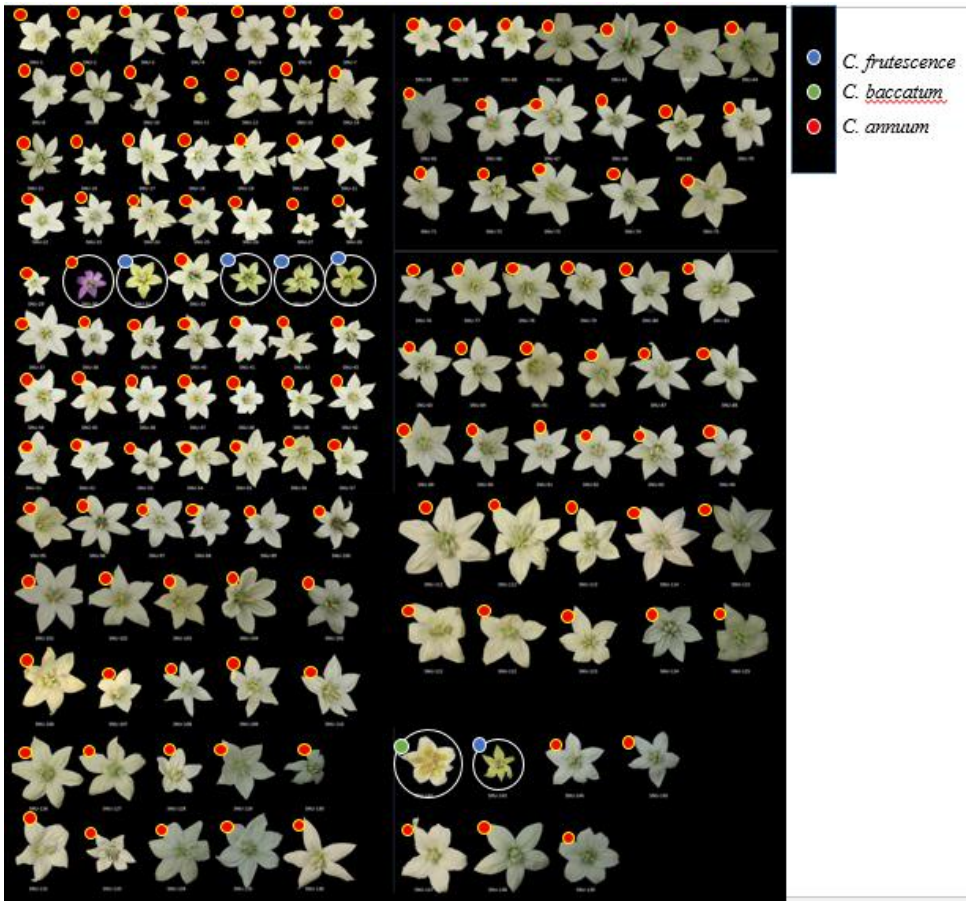


Figure 4. Flower morphology of germplasm collections. The encircled flowers are representatives of the different flower types as explained on the key on the right up.

Estimates of phenotypic (PCV) and genotypic coefficients of variation (GCV), broad-sense heritability and genetic advance are shown in Table 3. Across the traits studied, the PCV values ranged from 39.6% for plant width to 99.6% for fruit weight. Similar to the latter, PCV values were high for number of seeds per fruit, fruit length, fruit width, pericarp thickness, fruit perimeter, internode length, leaf length and main stem length, with respective values of 81.5%, 65.6%, 58.7%, 58.3%, 53.3%, 53.0%, 52.4% and 50.6%. In contrast, number of side branches, leaf length, pedicel length, plant height, stem thickness and plant width showed comparatively lower PCV values (<50%). The GCV estimates were lowest (31.9%) for plant width and highest (98.6%) for fruit weight. High GCV values were also recorded for seeds per fruit, leaf width and fruit length. However, relatively low GCV values (<50%) were recorded for fruit width, pericarp thickness, fruit perimeter, internode length, number of side branch, leaf length, main stem length, pedicel length, plant height and stem thickness. Broad-sense heritability was greatest for fruit weight (98.1%), followed by length and width of leaves with respective values of 77.7% and 77.1%. The heritability values of the remaining traits ranged from 62.7 to 75.9%. Genetic advance (GA) as percentage of the mean ranged from 1.7% to 107.7% for plant height and pericarp thickness, respectively (Table 3).

Table 3 Estimation of genetic parameters of different traits

Characters	Range	Mean ± SE	CV	Genetic variance	Phenotypic variance	Grand mean	Heritability (%)	GCV (%)	PCV (%)	Genetic advance (%)
Plant height (cm)	43–175	100.83 ± 1.29	21.4	1205.7	1672.0	100.8	72.1	34.4	40.6	1.74
Plant width (cm)	26.67–90.67	62 ± 0.8	23.5	391.2	604.2	62.0	64.8	31.9	39.6	2.68
Main stem length (cm)	17.67–64	37.42 ± 0.61	29.4	237.2	357.9	37.4	66.3	41.2	50.6	4.49
Internode length (cm)	3.67–18.17	9.32 ± 0.16	28.9	17.1	24.4	9.3	70.1	44.4	53.0	18.52
Number of side branch (count)	4–23.7	14.61 ± 0.22	24.2	39.5	52.1	14.6	75.9	43.0	49.4	12.31
Stem thickness (mm)	5.83–24.37	13.76 ± 0.18	22.6	21.4	31.0	13.8	68.8	33.6	40.5	12.44
Leaf length (cm)	8.33–136.67	84.4 ± 1.28	23.4	1307.9	1696.4	84.4	77.1	42.9	48.8	2.15
Leaf width (cm)	4.2–80	41.59 ± 0.67	24.8	368.9	474.8	41.6	77.7	64.2	52.4	4.37
Pedicel length (cm)	1.03–6.23	3.52 ± 0.05	27.4	1.7	2.6	3.5	64.4	36.8	45.8	47.03
Fruit length (cm)	1.53–10.93	6.42 ± 0.13	34.6	12.8	17.8	6.4	72.2	55.8	65.6	27.30
Fruit width	0.27–3.53	1.71 ± 0.03	31.7	0.72	1.01	1.7	70.9	49.4	58.7	101.34
Fruit weight (g)	0.73–15.4	6.8 ± 0.18	13.6	39.59	40.34	6.8	98.1	98.6	99.6	2.04
Pericarp thickness (mm)	0.3–3.23	1.59 ± 0.03	32.7	0.59	0.86	1.6	68.6	48.3	58.3	107.65
Seed number/fruit (count)	12–167.67	78.98 ± 2.12	49.7	2596.55	4138.6	79.0	62.7	64.5	81.5	2.07
Fruit perimeter (cm)	2.71–19.94	12.37 ± 0.21	28.7	30.8	43.5	12.4	70.9	44.9	53.3	14.04

GBS and single-nucleotide polymorphisms

We identified 2,831,791 genome-wide SNPs in our germplasm panel. We filtered these by removing rare alleles (with prevalence less than 5%), alleles with high missing ratios (absent from more than 30% of the germplasms) and alleles with high heterozygosity (more than 80%). To explore the genetic diversity of the panel, we analyzed all the germplasms using 53,284 high-quality SNPs. The chromosomal distribution and proportion of polymorphic markers used for the competition is shown in Table 4. We mapped the SNP density (number of SNPs per Mbp) and their distribution across the 12 chromosomes (Fig 5) and found that SNP densities varied across chromosomes. There was relatively high uniformity on chromosomes 3, 4, 5, 6, 8, 9, 10 and 12. On chromosome 2, a higher SNP distribution was found towards one end, and it was higher on both arms of chromosome 11. Relatively more SNPs were also recorded around the middle and one end of chromosome 7.

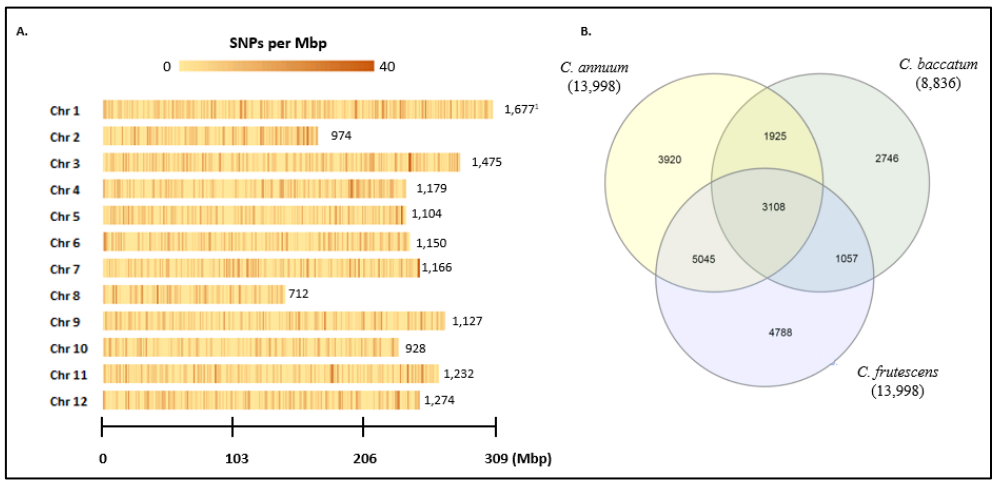


Figure 5. Single-nucleotide polymorphism (SNP) density and distribution across the 12 chromosomes of the *Capsicum* germplasms.

(A) SNP density (number of SNPs per Mbp) of Ethiopian *Capsicum* germplasms DNA digested using *Pst*I and *Mse*I. (B) Venn diagram of unique and shared SNPs of *Capsicum* germplasms, kept in each dataset after SNP filtering. The figure was drawn using the interactiVenn website.

[1]Number of SNPs.

Table 4 The chromosomal distribution and proportion of polymorphic markers used for computing heterogeneity, genetic distance and principal coordinate analysis and population structure

Chromosome	No. of markers	Proportion
1	1,677	12.0%
2	974	7.0%
3	1,475	10.5%
4	1,179	8.4%
5	1,104	7.9%
6	1,150	8.2%
7	1,166	8.3%
8	712	5.1%
9	1,127	8.1%
10	928	6.6%
11	1,232	8.8%
12	1,274	9.1%
Total	13,998	100.0%

Genetic diversity

The amount and organization of genetic diversity among the model-based populations is presented in Table 5. In this analysis *C. baccatum* was not included. A high allele number was recorded in *C. frutescens* (5.00), and the mean major allele frequency was greater in *C. annuum* (0.30) than in *C. frutescens* (0.28). The average expected heterozygosity (a measure of genetic diversity), observed heterozygosity and polymorphic information content (PIC; denoting allelic diversity and frequency) values were 0.74, 0.02 and 0.69 for *C. annuum* and 0.75, 0.03 and 0.70 for *C. frutescens*, respectively. The PIC value was 0.692 in 35 *C. annuum* and 0.701 in six *C. frutescens* germplasms. Categorically, the average *C. annuum* PIC value for the major growing region (SNNPs) was 0.693 per marker, with a range running from 0.692 in six germplasms collected from Gurage and Kembata zones to 0.693 in 24 germplasms collected from Gurage, Kembata, Bench Maji, Keficho and Semen Omo zones. A similar trend was observed in the PIC values of *C. annuum* for the Amhara and Oromia regions, representing respectively 39 and 24 germplasms. For the *C. frutescens* germplasms of SNNPs, the average PIC value was 0.701 per marker, and for Oromia it was 0.964. The overall gene diversity and mean heterozygosity were 0.74 and 0.02, respectively.

Table 5 Genetic diversity analysis of *Capsicum* germplasms

Subpop.	*N*	AN	MAF	H_e	H_o	PIC
C. annuum	121	4.46	0.30 (0.29–0.30)	0.74 (0.740–0.744)	0.02 (0.004–0.104)	0.69 (0.69–0.70)
C. frutescens	9	5.00	0.28 (0.28–0.30)	0.75 (0.74–0.75)	0.03 (0.01–0.06)	0.70 (0.69–0.70)
Total	130	4.50	0.3 (0.28–0.30)	0.74 (0.74–0.75)	0.02 (0.004–0.104)	0.69 (0.69–0.70)

N, number of germplasms; AN, number of allele per locus; H_e, expected heterozygosity (gene diversity); H_o, observed heterozygosity; PIC, polymorphism information content.

Analysis of molecular variance

To quantify the genetic diversity within and among subpopulations, we partitioned the total molecular variance into two clades according to the STRUCTURE simulation result. Averaged across the 130 germplasms, 92% of the total genetic diversity was partitioned between germplasms within the subpopulations, and only 8% was attributed to differences at the individual level (Table 6).

Table 6 AMOVA for the two subpopulations suggested by STRUCTURE for all Ethiopian *Capsicum* germplasms

Source	Df	SS	MS	Var.	%
Among subpopulations	1	0.027	0.027	0.000	0%
Among germplasms	128	11.915	0.093	0.045	92%
Within germplasms	130	0.500	0.004	0.004	8%
Total	259	12.442		0.048	100%

$P < 0.001$. Df, degrees of freedom; SS, sum of squares; MS, mean square; Var., estimated variation.

Population structure

We inferred the population structure of the 130 Ethiopian *Capsicum* germplasms using the program STRUCTURE 2.3.4 (Prichard and Stephens, 2000). We carried out admixture model-based simulations by varying K from 1 to 5 with 10 iterations using 130 germplasms. The estimated likelihood (lnP(D)) was greatest for K = 2 (Fig 6), suggesting the presence of two main populations in the *Capsicum* germplasm panel (Earl and vonHoldt, 2012). The classification of germplasms into populations based on the model-based structure from STRUCTURE 2.3.4 (Fig 6) showed that subpopulation *C. annuum* comprised 123 germplasms and subpopulation *C. frutescens* comprised the remaining 7 germplasms. We tested the genetic variation within the subpopulations using the fixation index (Fst) statistic for genetic differentiation. We observed a low average distance (HE) between individuals in the same clade in subpopulation *C. annuum* (0.05) and a higher HE in subpopulation *C. frutescens* (0.07). The FST values for subpopulations *C. annuum* and *C. frutescens* were 0.713 and 0.850, respectively.

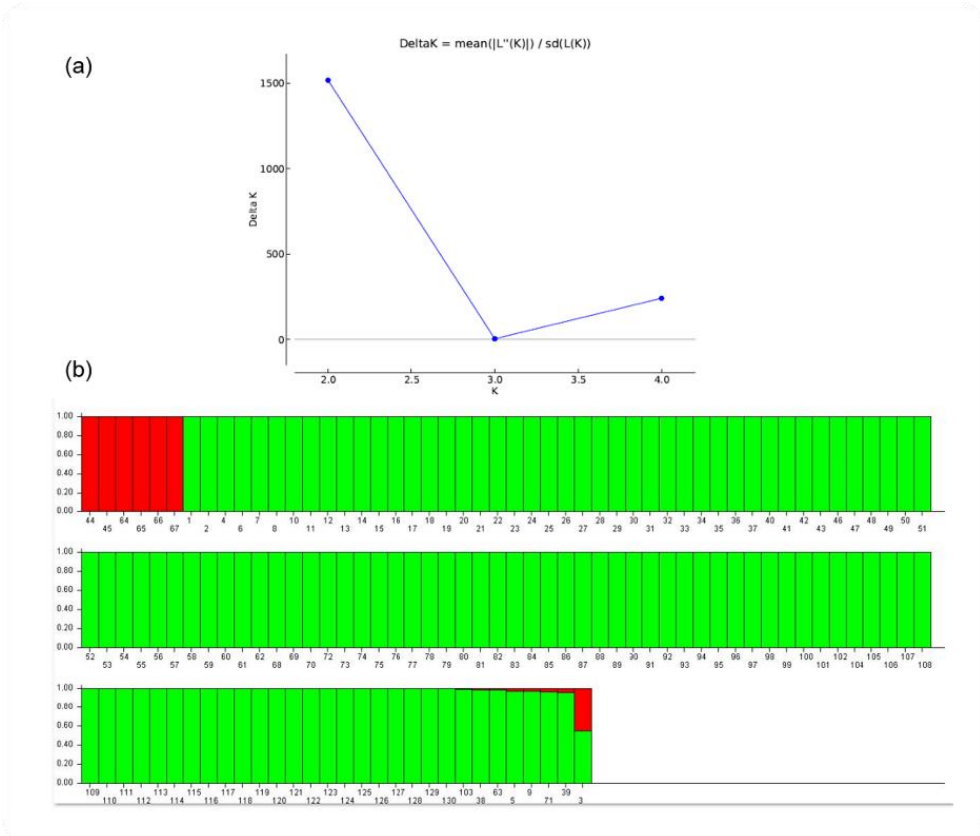


Figure 6. Population structure of 139 *Capsicum* germplasms.

(a) STRUCTURE estimation of the number of subgroups from K values ranging from 1 to 10, by delta K (K) values. (b) Population structure from K = 2. The two colors each represent one subpopulation (green, *C. annuum*; red, *C. frutescens*), and the lengths of the colored segments shows the estimated membership proportion of each germplasm in the designated group.

Molecular phylogenetic and principal-coordinate analysis

The unrooted phylogenetic tree with two clades is consistent with the model-based population structure, in which *C. frutescens* germplasms were grouped separately from *C. annuum* (Fig 7). Clade 1 contained *C. annuum* accessions growing in an altitude range between 1,000 and 2,780 m.a.s.l. and consisting mainly of germplasms collected from different growing localities of the SNNPs region (35 germplasms), Amhara (47 germplasms), Oromia (37 germplasms), Benishangul (5 germplasms), Somali (1 germplasm) and Tigray (2 germplasms), which accounted for 24%, 33%, 26%, 3.5%, 0.7% and 1.4% of germplasms, respectively. The growing altitude range of *C. frutescens* (1,200–1,310 m.a.s.l.) was narrow compared with that of *C. annuum* (Fig 7).

We also performed PCoA on 130 germplasms (Fig 7). This analysis largely supported the separation of the germplasms into two subpopulations fairly well distributed on the axes, with one variation as indicated by an arrow. Cluster A consisted mainly of *C. annuum*, a pattern also evidenced in the model-based genetic clustering using STRUCTURE and the phylogenetic tree. Germplasms in Cluster B were all from *C. frutescens*.

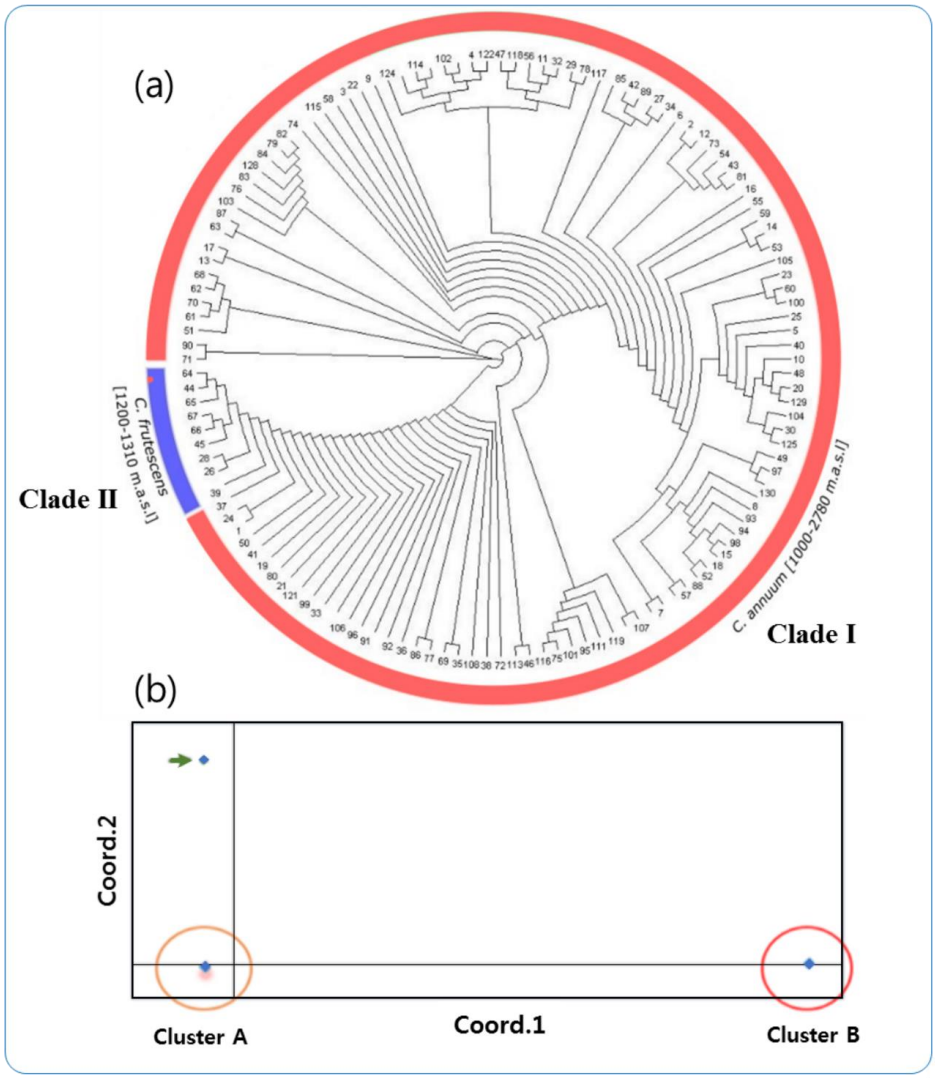


Figure 7. Phylogenic tree and principal-coordinate analysis results confirming the presence of two groups.

(a) Unrooted neighbor-joining (NJ) tree of the 142 *Capsicum* germplasms. (b) Principal-coordinate analysis results showing diversity of *Capsicum* species clustered in the two subpopulations.

GWAS for selected traits

Genome-wide association result on fruit weight is summarized by Manhattan plots in Fig 8. With a Bonferroni correction threshold of 5% (–log10(P > 6.03), the number of markers linked to various traits varied from a maximum of 187 for fruit length to a minimum of one for fruit shape index, leaf color, petal length, petal width and stem color. A total of 509 significant SNPs were identified, 81.53% of which were for fruit traits, 10.61% for leaf traits, 0.39% for petal length and width and 7.47% for stem-related traits. The largest fraction of significant SNPs (26.68%) was detected on chromosome 3, followed by chromosomes 8 and 9 with 16.11% and 13.56%, respectively. The smallest concentration of significant SNP markers was observed on chromosome 12 for fruit length, fruit number, fruit weight and petal width. SNP markers related to fruit traits (area, color, length, width, number, weight, shape-index, number of locules, pericarp thickness and perimeter) were distributed across all 12 chromosomes. While SNP markers for stem-related traits (internode length, hairiness, thickness, color and branching) are distributed on chromosomes 1, 2, 3, 4, 6, 9, 10 and 11, those for leaf-related traits (length, width and color) are localized on chromosomes 2, 3, 6, 7 and 9 (Fig 9).

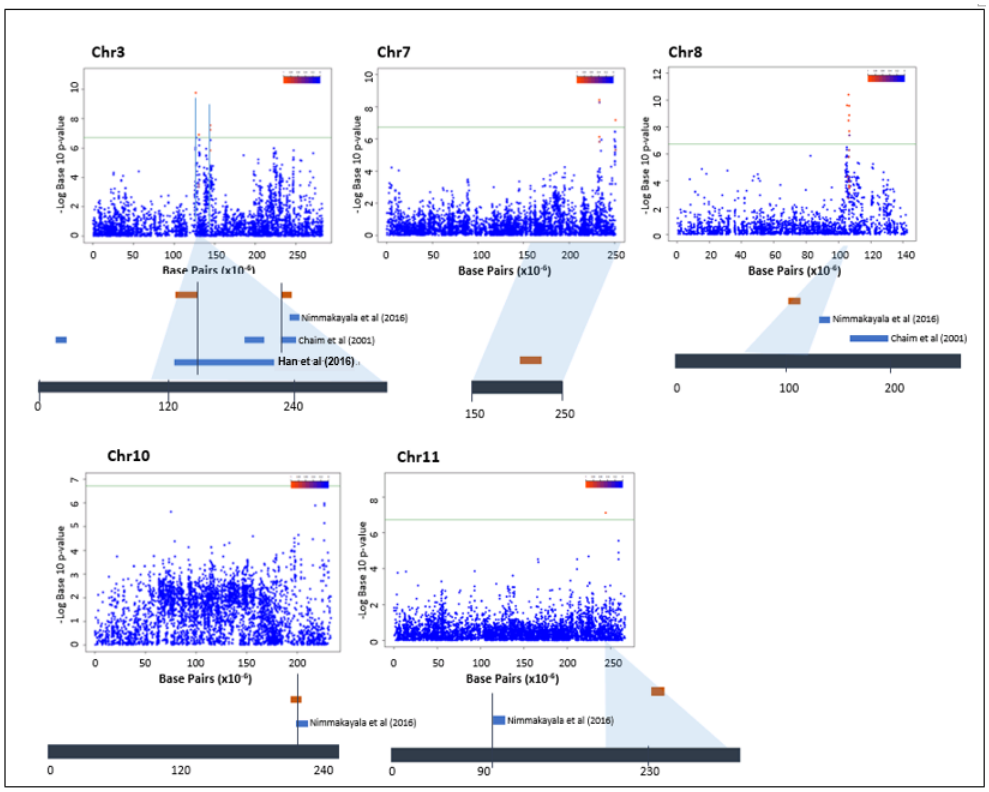


Figure 8. Manhattan plots of the genome-wide association study for fruit weight. Chromosome coordinates are displayed along the X axis with the –log10 of the association P value for each single nucleotide polymorphism displayed on the Y axis. A greater –log10 indicates stronger association with the trait. The green line denotes the significance threshold.

The threshold for –log10(P) was 6.0. Blue thick horizontal bars denote QTL positions reported previously, while brown horizontal thick bars are our results.

For fruit weight (Fig 8), two regions containing 12 SNPs were detected on chromosome 3, from 126.3 to 144.9 Mbps and from 223.02 to 223.04 Mbps, and one region was detected on chromosome 7 with seven SNPs in the region between 233.4 and 251 Mbps. There were ten SNP marker positions on chromosome 8 between 106.28 and 106.6 Mbps, and one each on chromosomes 10 and 11 at 226.8 and 244.4 Mbps, respectively. From this result, our research demonstrated the existence of three new regions of significant SNP markers for fruit weight on chromosomes 7, 8 and 11 in their specific mentioned regions. A total of 27 protein-coding genes were predicted within the significant SNP regions based on the annotation of the CM334.v.1.6. genome (Table 7).

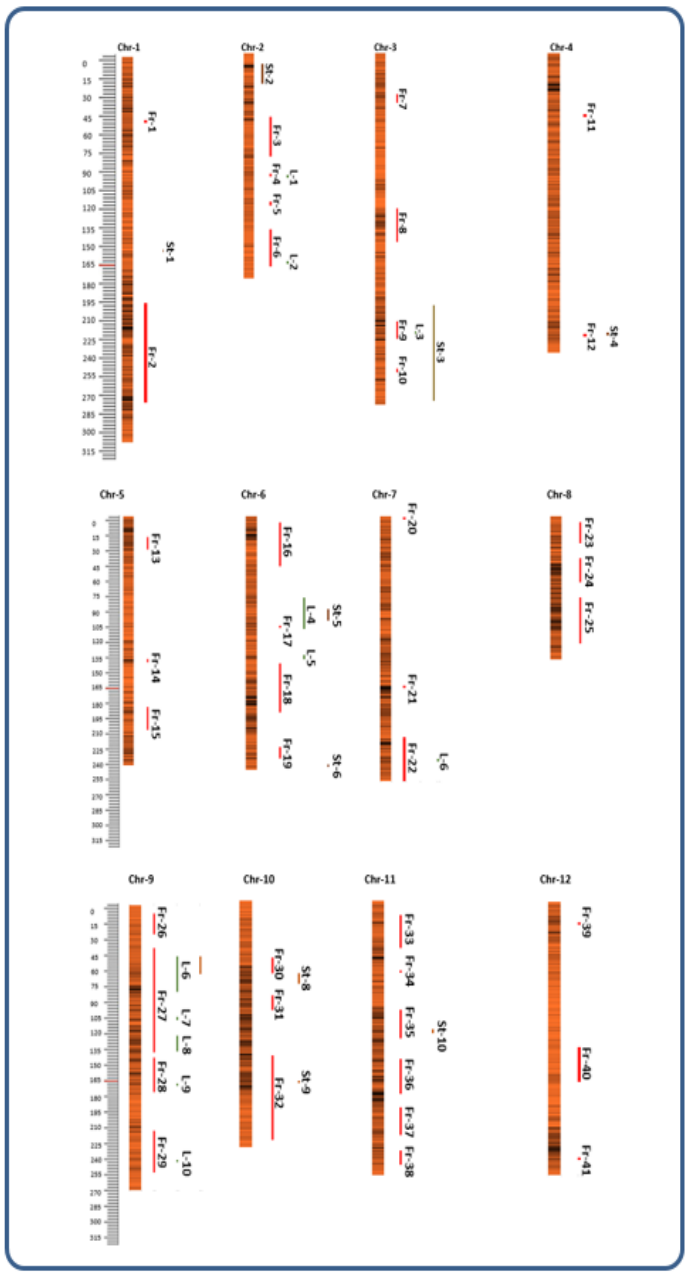


Figure 9. Significant SNP marker distribution map across 12 chromosomes of *Capsicum* for fruit (red bars), leaf (green bars) and stem (orange bars) traits based on GWAS result. The scale shows the physical distance of SNPs in the chromosome (Mb).

Table 7 List of 27 candidate genes located in chromosomes 3, 7, 8, 10 and 11 for the fruit weight

SNP ID	Chr	Position	maf	P.value	Predicted protein	GO Name
S03_126637882	3	126,637,882	0.27	1.76E-10	S-locus-specific glycoprotein S6	F:binding; F:electron transfer activity; C:membrane
S03_144571942	3	144,571,942	0.25	5.60E-08	hypothetical protein BC332_25369	
S03_130472754	3	130,472,754	0.22	1.26E-07	DNA/RNA polymerases superfamily protein	F:binding; P:DNA metabolic process
S03_127229794	3	127,229,794	0.26	1.94E-07	Lysine-specific demethylase	P:phosphatidylserine biosynthetic process; C:integral component of membrane; F:transferase activity
S03_131045471	3	131,045,471	0.26	2.71E-07	hypothetical protein T459_16762	F:mannosyl-oligosaccharide 1,2-alpha-mannosidase activity; F:calcium ion binding; C:membrane
S03_144926307	3	144,926,307	0.24	2.90E-07	Putative gag-pol polyprotein, identical	F:nucleic acid binding; F:catalytic activity; F:zinc ion binding; F:1-deoxy-D-xylulose-5-phosphate synthase activity; P:DNA integration; P:terpenoid biosynthetic process
S03_130816018	3	130,816,018	0.22	6.81E-07	PREDICTED: uncharacterized protein LOC107844328	F:ion binding; F:organic cyclic compound binding; F:heterocyclic compound binding

SNP ID	Chr	Position	maf	P.value	Predicted protein	GO Name
S03_126408053	3	126408053	0.27	9.28E-07	Lysine-specific demethylase	F:nucleic acid binding; P:gene expression; F:transferase activity; P:cellular nitrogen compound metabolic process; P:cellular protein metabolic process
S03_223037785	3	223,037,785	0.48	1.05E-06	Malate dehydrogenase, glyoxysomal	F:catalytic activity
S03_223037803	3	223,037,803	0.48	1.05E-06	Malate dehydrogenase, glyoxysomal	F:catalytic activity
S03_139442074	3	139,442,074	0.26	1.11E-06	ATP-dependent DNA helicase SRS2-like protein	C:integral component of membrane; P:oxidation-reduction process
S07_233424834	7	233,424,834	0.15	3.94E-09	ATPase ASNA1 -like protein	F:protein serine/threonine kinase activity; F:ATP binding; C:endoplasmic reticulum; P:generation of precursor metabolites and energy; P:protein phosphorylation; P:proteolysis; P:transport; F:cysteine-type peptidase activity; C:integral component of membrane; F:ATPase activity; P:protein insertion into ER membrane; P:negative regulation of transcription, DNA-templated

SNP ID	Chr	Position	maf	P.value	Predicted protein	GO Name
S07_233499205	7	233,499,205	0.19	5.50E-09	Potassium transporter 5	P:cellular protein modification process; P:phosphate-containing compound metabolic process; F:methyltransferase activity; C:membrane; F:hydrolase activity; P:methylation; F:adenyl ribonucleotide binding; F:anion binding; C:intracellular part; F:catalytic activity, acting on a protein
S07_251125166	7	251,125,166	0.37	6.81E-08	putative copia-type protein	C:retrotransposon nucleocapsid; F:nucleic acid binding; F:motor activity; F:protein tyrosine phosphatase activity; F:ubiquitin-protein transferase activity; F:ATP binding; P:actin filament organization; F:protein tyrosine/serine/threonine phosphatase activity; P:DNA integration; C:integral component of membrane; C:myosin complex; P:protein ubiquitination; P:peptidyl-tyrosine dephosphorylation; F:ADP binding; F:actin filament binding
S07_250424475	7	250,424,475	0.29	3.48E-07	uncharacterized protein LOC114075241	F:nucleic acid binding; F:RNA-DNA hybrid ribonuclease activity; P:DNA integration; P:RNA phosphodiester bond hydrolysis, endonucleolytic
S07_250605815	7	250,605,815	0.38	3.75E-07	TPR repeat-containing thioredoxin TDX	C:cell; F:transferase activity

SNP ID	Chr	Position	maf	P.value	Predicted protein	GO Name
S07_233424790	7	233,424,790	0.16	7.40E-07	ATPase ASNA1 -like protein	F:protein serine/threonine kinase activity; F:ATP binding; C:endoplasmic reticulum; P:generation of precursor metabolites and energy; P:protein phosphorylation; P:protein dephosphorylation; P:proteolysis; F:protein tyrosine/serine/threonine phosphatase activity; F:cysteine-type peptidase activity; C:integral component of membrane; F:ATPase activity; P:protein insertion into ER membrane; P:negative regulation of transcription, DNA-templated
S07_249749737	7	249,749,737	0.33	1.07E-06	ADP,ATP carrier protein, mitochondrial	F:catalytic activity; C:membrane
S08_106280443	8	106,280,443	0.33	3.85E-11	Agamous-like MADS-box protein AGL16	F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated
S08_105353008	8	105,353,008	0.29	2.54E-10	Putative retrotransposon protein, identical	F:carboxypeptidase activity; F:binding; P:nitrogen compound metabolic process; P:macromolecule metabolic process; P:primary metabolic process
S08_106308670	8	106,308,670	0.31	3.35E-09	[Pyruvate dehydrogenase (acetyl-transferring)] kinase, mitochondrial	F:binding
S08_106726114	8	106,726,114	0.27	2.10E-08	PREDICTED: uncharacterized	F:nucleic acid binding; P:DNA integration

SNP ID	Chr	Position	maf	P.value	Predicted protein	GO Name
					protein LOC107865331	
S08_105142620	8	105,142,620	0.44	3.22E-07	PREDICTED: uncharacterized protein LOC107861928	F:zinc ion binding
S08_105183429	8	105,183,429	0.29	4.79E-07	keratin, type I cytoskeletal 12-like	F:protein tyrosine phosphatase activity; F:binding; P:proteolysis; P:transport; F:protein tyrosine/serine/threonine phosphatase activity; F:cysteine-type peptidase activity; C:integral component of membrane; P:peptidyl-tyrosine dephosphorylation; C:cytoplasmic part; C:membrane protein complex
S08_106609467	8	106,609,467	0.32	5.13E-07	Putative polyprotein, identical	F:nucleic acid binding; F:zinc ion binding; P:DNA integration
S10_226847725	10	226,847,725	0.46	1.05E-06	hypothetical protein CQW23_01337	F:copper ion binding; F:oxidoreductase activity; P:oxidation-reduction process
S11_244407143	11	244,407,143	0.19	7.57E-08	PREDICTED: uncharacterized protein LOC104231117	

DISCUSSION

The greater the genetic diversity of germplasm, the greater is the chance of success in breeding desirable strains. Knowledge of population structure and genetic diversity is essential for association mapping studies, genomic selection and the classification of individual genotypes into different groups. In the present study, we classified Ethiopian *Capsicum* germplasms into different species and analyzed their genetic diversity. According to our classification, the majority of Ethiopian pepper germplasms collected from diverse agro-ecologies are *C. annuum*, whose mature fruit is an integral ingredient of the local spice mixture called berbere, used to season many Ethiopian dishes. The green fruit of *C. annuum* is also a very important component in the daily diet. The brown chilli pepper type (*C. annuum*) is especially highly valued for its high pungency for flavoring and coloring. Work by Berhanu et al. (2017), Shimelis et al. (2016) and Abrham et al. (2017) had demonstrated the prevalence and variability of *C. annuum*. We also recognized the presence of some *C. frutescens* accessions, known locally as "mitmita", growing on some part of the country. They are known locally for being highly pungent. Yayeh (1998) had previously described the existence of *C. frutescens* in Ethiopia (Zewdie, 1998). In addition, the distribution of these two *Capsicum* species across Africa was described by Eshbaugh (1983),

who summarized the evolutionary history of peppers and described how the genus was introduced to East Africa (Eshbaugh, 1983). Similarly, Dagnoko et al. (2013) mentioned the importance of *C. annuum* and *C. frutescens* in West African countries (Dagnoko et al., 2013).

Although morphological traits are important in the study of genetic diversity, because of their mostly polygenic nature and their dependence on various environmental factors, they may not always reflect real genetic variation (Kumbhar et al., 2015). Owing to their ability to recognize specific DNA sequences in the closely related genotypes, SNP markers have been used successfully to estimate genetic diversity among different plants. GBS is a preferred high-throughput genotyping method involving targeted complexity reduction and multiplex sequencing to produce high-quality polymorphism data at a relatively low cost per sample (Annicchiarico et al., 2017). The GBS method uses restriction enzymes coupled with DNA-barcoded adapters and can simultaneously perform SNP discovery and genotyping with or without reference genome sequences. GBS have been applied to various approaches for plant breeding and plant genetic studies, including linkage maps (Han et al., 2018; Lee et al., 2017; Han et al., 2016), genome-wide association studies (Han et al., 2018; Nimmakayala et al., 2016), genomic selection (Biazzi et al., 2017) and genomic diversity studies (Lee et al., 2016).

We performed GBS for genotyping 142 germplasms of *Capsicum* species. Two enzymes, *Pst*I and *Mse*I, were used to reduce genome complexity,

consistent with previous studies by Han et al. (2018). Using the CM334 genomic reference, SNP calling generated 53,284 high-quality SNPs. Transitions (72.45%) were more frequent than transversions (27.55%). The percentages of each SNP type in our study were 36.08%, 36.37%, 8.66%, 5.49%, 8.49% and 4.91% for AG, CT, GT, AT, AC and CG, respectively. This observation agrees with the report of Taranto et al. (2016), which also found a higher frequency of transitions than transversions.

Heritability values are helpful in predicting the expected progress to be achieved through the process of selection; high heritability coupled with high genetic advance (GA) is an indicator of a high proportion of additivity in the genetic variance, and consequently suggests that a high genetic gain can be expected from selection (Khan et al., 2016). The heritability values of the traits assessed in this analysis fell into two categories—very high (98.1%) for fruit weight and moderately high for the remainder—as illustrated in previous studies (Marame et al., 2009; Khan et al., 2016). High GA (>20%) with moderate heritability values were observed for pericarp thickness, fruit width, pedicel length and fruit length. This pattern is partly supported by the findings of Usman et al. (2014). High values of PCV and GCV values for some of fruit-related traits were also reported previously (Shimeles et al., 2016) and indicated the existence of substantial variability, ensuring ample scope for improvement of these traits through selection.

Our panel of germplasms exhibited a wide range of genetic diversity for

different agro-morphological traits. Plant habits as a measure of plant architecture; growth parameters including plant height, stem thickness, internode length, number of side branches; and morphological traits such as number of flowers per axil, length and thickness of fruits, fruit weight and number of locules all showed variation among germplasms.

Results of neighbor-joining clustering with model-based STRUCTURE, phylogenetic and PCoA all similarly suggested that there are two genetically distinct subclasses among the Ethiopian *Capsicum* germplasms investigated in this study. The PCoA showed tight clustering within the first clade, composed of *C. annuum*, and the second clade, in which all *C. frutescens* are grouped. A previous diversity study of 39 cultivated Ethiopian *C. annuum* strains using AFLP distance estimation, however, showed four major clusters (Geleta et al., 2005). Although *C. frutescens* has been shown to have close affinity to *C. annuum*, they were grouped separately in this study (Ince et al., 2010). Based on our data, the high diversity values in the two subpopulations suggests the existence of excessive genetic variation within them. The lower value of Ho (observed heterozygosity) as compared to He (expected heterozygosity) in the subpopulations indicated the presence of inbreeding in the majority of Ethiopian *Capsicum* germplasms. The average distance (HE) between individuals in same clade value was lower in subpopulation *C. annuum* (0.05) than in subpopulation *C. frutescens* (0.07), indicating that *C. annuum* contained less variation. Genetic differentiation (FST) value for

subpopulations *C. annuum* and *C. frutescens* were 0.713 and 0.85, respectively, predicting that the germplasms in the two clades have several genotype patterns. There was a significant correlation between some of the morphological traits, such as plant height and fruit width, as indicated by SNP-marker-based matrices. In summary, the model-based ancestry analysis, the phylogenetic tree and the PCoA strongly supported the possibility that the collection of Ethiopian *Capsicum* germplasms has two well-differentiated genetic populations and some admixtures.

In our research, additional information was provided by the GWAS analysis. As was shown earlier by Chaim et al. (2001) and Han et al. (2016), the highest numbers of SNP markers for various agronomic traits were detected on chromosome 3. From the total of 398 significant SNPs of selected traits, 10 SNPs from chromosome 8, 6 SNPs from chromosome 7 and 1 SNP from chromosome 3 were common for the traits of fruit weight and fruit length. The remaining major SNP marker (10) observed for fruit weight in our study was detected by Chaim et al. (2001). A recent report by Chunthawodtiporn et al. (2018) demonstrated the distribution of fruit-trait QTL; the authors considered transverse and longitudinal section, fruit shape and blossom-end shape on all chromosomes except 4, 5 and 7. However, our study included more fruit-related traits, and our results suggest that significant SNP markers are found on all chromosomes. For leaf length, for which we identified various significant SNP markers on chromosomes 2, 3, 6, 7 and 9,

co-localized QTL results were reported on chromosomes 6, 8, 9 and 11 by Han et al. (2016) and on chromosomes 1, 2 and 3 by Chunthawodtiporn et al. (2018). We detected SNP markers for stem traits on seven of the same chromosomes reported by Han et al. (2016).

GWAS identified 31 SNP markers for fruit weight in this study. On chromosome 3, the SNP markers were co-localized at two different locations with the QTL reports of Han et al. (2016), Chaim et al. (2001) and Nimmakayala et al. (2016), at 126 Mbps and 222 Mbps. In a pattern similar to that of our results, Nimmakayala et al. (2016) reported a QTL region between 220 and 237 Mbps on chromosome 10. Although both that study and Chaim et al. (2001) reported QTL positions on chromosome 8, our results for chromosome 8 indicated the presence of a SNP concentration at a different location. Similarly, the SNP position we identified on chromosome 11 at 239 Mbp is different from that reported by Nimmakayala et al. (2016) of a QTL elsewhere on the same chromosome.

This study is the first detailed characterization of a large sample of the Ethiopian Capsicum germplasms, representing the six administrative regions covering all *Capsicum*-growing agro-ecologies in the country. Our morphological and molecular characterization provides insight into the genetic variability of *Capsicum* in Ethiopia.

REFERENCES

Albrecht E, Zhang DP, Mays AD, Saftner RA, et al (2012) Genetic diversity in *Capsicum baccatum* is significantly influenced by its ecogeographical distribution. BMC Genet 13(68): 1-15

Annicchiarico P, Nazzicari N, Wei Y, Pecetti L, et al (2017) Genotyping-by-sequencing and its exploitation for forage and cool-season grain legume breeding. Front Plant Sci 8(679): 1-13

Biazzi E, Nazzicari N, Pecetti L, Brummer EC, et al (2017) Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. PloS One 12(1): e0169234. https://doi.org/10.1371/journal.pone.0169234

Birhanu H, Tiegist DDY (2017) Morphological characterization of hot pepper (*Capsicum annuum*.L) land races of Ethiopia for qualitative characters. Int J Res Stud in Sci Eng & Tech 4(9): 4-9

Burton GW, DeVane EH (1953) Estimating heritability in tall fescue (Festzjcu Av-undinuceu) from replicated clonal material. Agron J 45: 478-81

Chaim AB, Paran I, Grube RC, Jahn M, et al (2001) QTL mapping of fruit-related traits in pepper (*Capsicum annuum*). Theor Appl Genet 102(6): 1016-28

Chunthawodtiporn J. Hill T, Stoffel K, Deynze AV (2018) Quantitative

trait loci controlling fruit size and other horticultural traits in bell pepper (*Capsicum annuum*). Plant Genome 11(1): 1-11

Clarke J (2009) Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. Cold Spring Harbor protocols. 4(3): 1-2

Dagnoko S, Yaro-DDN, Sanogo PN, Adetula O, et al (2013) Overview of pepper (*Capsicum* spp.) breeding in West Africa. Afr J Agric Res 8(13): 1108-14

Earl DA, von Holdt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2): 359-61

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One 6(5): e19379. https://doi.org/10.1371/journal.pone.0019379

Ertiro BT, Semagn K, Das B, Olsen M, et al (2017) Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. BMC Genomics 18(1): 777: 1-11

Eshbaugh WH (1983) The genus *Capsicum* (Solanaceae) in Africa. Bothalia 14(3&4): 845-8

Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and

Windows. Mol Ecol Resour 10(3): 564-7

Fekaku M, LD, Chemeda F, Roland S (2009) Genetic analysis for some plant and fruit traits, and its implication for a breeding program of hot pepper (*Capsicum annuum* var. *annuum* L.). Hereditas 146: 131-40

Geleta LF, Labuschagne MT, Viljoen C (2005) Genetic variability in pepper (*Capsicum annuum* L.) estimated by morphological data and amplified fragment length polymorphism markers. Biodivers Conserv 14(10): 2361-75

González-Pérez S, Garcés-Claver A, Mallor C, Sáenz de Miera LE, et al (2014) New insights into *Capsicum* spp relatedness and the diversification process of *Capsicum annuum* in Spain (analysis of a *Capsicum* collection from Spain). PloS One 9(12): e116276. https://doi.org/10.1371/journal.pone.0116276

Han K, Jeong HJ, Yang HB, Kang SM, et al (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, et al (2018) QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. Plant Biotechnol J 16: 1546-1558

Huson DH, Richter DC, Rausch C, Dezulian T, et al (2007) Dendroscope: An interactive viewer for large phylogenetic trees. BMC

Bioinformatics 8(460): 1-6

Ince AG, Karaca M, Onus A (2010) Genetic relationships within and between *Capsicum* Species. Biochem Genet 48(1-2): 83-95

Iquira E, Humira S, Francois B (2015) Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. BMC Plant Bio 1-12

Jeong HJ, Jo YD, Park SW, Kang B (2010) Identification of *Capsicum* species using SNP markers based on high resolution melting analysis. Genome 53(12): 1029-40

Jo J, Venkatesh J, Han K, Lee HY, et al (2017) Molecular mapping of PMR1, a novel locus conferring resistance to powdery mildew in pepper (*Capsicum annuum*). Front Plant Sci 8: 2090

Khan A, Eyasmin R, Rashid MH, Ishtiaque S, et al (2016) Variability, heritability, character association, path analysis and morphological diversity in snake gourd. Agric Nat Resour 50(6): 483-489

Kumbhar SD, Kulwal PL, Patil JV, Sarawate CD, et al (2015) Genetic diversity and population structure in landraces and improved rice varieties from India Rice Sci 22(3): 99-107

Lee HY, Ro NY, Jeong HJ, Kwon JK, et al (2016) Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. BMC Genet 17(1): 142

Lee JH, An JT, Siddique MI, Han K, et al (2017) Identification and molecular genetic mapping of Chili veinal mottle virus (ChiVMV) resistance genes in pepper (*Capsicum annuum*). Mol Breeding 37(121): 1-10

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26(5): 589-595

Li H, Handsaker B, Wysoker A, Fennell T, et al (2009) The sequence alignment /map format and SAMtools. Bioinformatics 25(16): 2078-2079

Lipka AE, Tian F, Wang Q, Peiffer J, et al (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28(18): 2397–2399

Marame F, Dessalegne L, Fininsa C, Sigvald R (2009) Heterosis and heritability in crosses among Asian and Ethiopian parents of hot pepper genotypes. Int J Plant Bree 168(2): 235-247

McKenna A, Hanna M, Banks E, Sivachenko A, et al (2010) The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. Gen Res 20(9): 1297-303

Nimmakayala P, Abburi VL, Saminathan T, Alaparthi SB, et al (2016) Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. Sci Rep 6: 1-14

Nimmakayala P, Levi A, Abburi L, Abburi VL, et al (2014) Single

nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. BMC Genomics 15(767):1-14

Pailles Y, Ho S, Pires IS, Tester M, et al (2017) Genetic diversity and population structure of two tomato species from the Galapagos Islands. Front Plant Sci 8(138): 1-11

Panda RC, Kumar OA, Rao K (2004) Cytogenetic studies of some F_1 hybrids between wild and cultivated taxa of *Capsicum* L. Cytologia 69(2): 203–208

Peakall R, Smouse P (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28(19): 2537-2539

Perrier X, Jacquemoud-Collet J (2006) DARwin software http://darwin.ciard,fr/.. Epub 6.

Poland JA, Brown PJ, Sorrells ME, Jannink J (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. https://doi.org/10.1371/journal.pone.0032253

Popovsky-Sarid S, Borovsky Y, Faigenboim A, Parsons E, et al (2017) Genetic and biochemical analysis reveals linked QTLs determining natural variation for fruit post-harvest water loss in pepper

(*Capsicum*). Int J Plant Bree Res 130(2): 445-459

Prichard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype Data. Genet 155: 945-959

Rodríguez GR, Moyseenko JB, Robbins MD, Morejón NH, et al (2010) Tomato Analyzer: A useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. J Vis Exp 37: 1-12

Sacco A, Ruggieri V, Parisi M, Festa G, et al (2015) Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. PloS One 10(9): e0137139. https://doi.org/10.1371/journal.pone.0137139

Shimeles A, Bekele A, Dagne W, TW Adefris (2016) Genetic variability and association of characters in Ethiopian hot pepper (*Capsicum annum* L.) landraces. J Agri Sci 61(1): 19-36

Shumbulo A, Nigussie M, Alamerew S (2017) Correlation and path coefficient analysis of hot pepper (*Capsicum annuum* L.) genotypes for yield and its components in Ethiopia. Adv Crop Sci Tech 5(3): 1-5

Song M, Li J, Xiong C, Liu H, et al (2016) Applying high-resolution melting (HRM) technology to identify five commonly used *Artemisia* species. Sci Rep 6: 1-7

Tang W, Wu T, Ye J, Sun J, et al (2016) SNP-based analysis of genetic

diversity reveals important alleles associated with seed size in rice. BMC Plant Biol 16(93): 1-11

Taranto F, D'Agostino N, Greco B, Cardi T, et al (2016) Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. BMC Genomics 17(943): 1-13

Tsaballa A, Ganopoulos I, Timplalexi A, Aliki X, et al (2015) Molecular characterization of Greek pepper (*Capsicum annuum* L) landraces with neutral (ISSR) and gene-based (SCoT and EST-SSR) molecular markers. Biochem Syst Ecol 59: 256-263

Usman MG, Rafii M, Ismail M, Malek MA, et al (2014) Heritability and genetic advance among chili pepper genotypes for heat tolerance and morphophysiological characteristics. Sci World J 2014: 1-14

van Zonneveld M, Ramirez M, Williams DE, Petz M, et al (2015) Screening genetic resources of *Capsicum* peppers in their primary center of diversity in Bolivia and Peru. PloS One 10(9): e0134663. https://doi.org/10.1371/journal.pone.0134663

Wong MML, Gujaria-Verma, N, Ramsay L, Yuan HY, et al (2015) Classification and characterization of species within the genus lens using genotyping-by-sequencing (GBS). PloS One 10(3): e0122025. https://doi.org/10.1371/journal.pone.0122025

Zewdie Y (1998) Ethiopian Chiles: A Deep-Rooted Tradition. The Chile

Pepper Institute

Zhang D, Kong W, Robertson J, Goff VH, et al (2015) Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). BMC Plant Biol 15(107): 1-15

CHAPTER II

Fine mapping of the *Up* gene controlling fruit orientation in pepper (*Capsicum* spp.)

ABSTRACT

The orientation of fruits is one of a distinguishing morphological features of *Capsicum* varieties. The downward curved growth of fruit stalks is highly correlated with fruit weight and pedicel length. Genetic analysis revealed that the pendant fruit orientation is governed by a dominant gene while incomplete inheritance is also observed in some *Capsicum* accessions. To identify and localize the fruit orientation controlling gene, a single trait locus analysis involving one F_2 and two recombinant inbred line (RIL) populations following the procedure of quantitative trait loci (QTL), and a genome wide association (GWAS) using core collection were performed. Common QTL regions on chromosome 12 were detected for fruit orientation in the populations. A total of 187,966 genome wide SNPs were derived from genotyping-by-sequencing (GBS) for GWAS analysis that included 196 *C. annuum*, 25 *C. baccatum*, 21 *C. chinense* and 14 *C. frutescens* accessions, representing the worldwide germplasm collection. Based on QTL and GWAS results, two markers were developed in an F_2 population derived from a cross between upright and pendant accessions, which delimited the *Up* locus to a 101 kb region on chromosome 12. Fifteen candidate genes, nine of which encoding proteins with known functional domains and six encoding hypothetical protein were identified in the delimited region. ZF-MYM transcription factor displayed one nucleotide deletion and one transition

mutations in the first exon, which was predicted to display a frameshift mutation resulting in an amino acid sequence change in the entire protein part encoded by the two exons at the C-termini. Consistent alignment results of ZF-MYM confirms the existence of variation between the two orientation types. Therefore, Zinc finger MYM-type protein 1-like transcription factor is the most likely candidates for the *Up* gene. These results provide a base for cloning of *Up*, which has an important contribution to the upright and pendant fruit orientation in pepper

INTRODUCTION

Pepper (*Capsicum spp* L.), originated in the New World of Central and South America, is one of the most important cultivated crops in the Solanaceae family, with five species namely *C. annuum*, *C. frutescens*, *C. baccatum*, *C. chinense* and *C. pubescens* domesticated since more than 6,000 years (Ramchiary et al., 2013; Bai and Lindhout, 2007). Initially, peppers were used as food preservative and medicines but were subsequently consumed as a spice and a vegetable (Chunthawodtiporn et al., 2018). The wild forms of pepper are small and soft fruited, red colored, pungent and deciduous (fruits are easy to separate from the calyx), and leaves are small and pubescent. However, selection during domestication resulted in cultivated non-deciduous types with large and less pungent fruits of different color and hairless leaves. The change in fruit position from erect to pendant has also been described as one of these few important traits selected during pepper domestication (Paran and Knaap, 2007; Albrecht et al., 2012; Chunthawodiporn et al., 2018). The change in fruit orientation from erect in which the fruit is held in an upright position to pendant where the fruits are pendulous or hang freely may be associated with an increase in fruit size, increase in length and/or decrease in thickness of pedicel, better protection from sun exposure, and predation by birds (Paran and Knaap, 2007; Setiamihardja and Knavei, 1990). In capsule and follicle bearing plants, erect

fruit phenotype is required for the dispersal of seeds (Niu et al., 2016). In Arabidopsis, the role of *BREVIPEDICELLUS* (BP) gene in regulating the proliferation of cells in pedicels elongation and its effect on the curvature on the abaxial region of the pedicel was described (Wang et al., 2015). The role of a MADS-box gene *NtSVP (SHORT VEGETATIVE PHASE)* on the elongation and orientation of pedicel of tobacco is reported (Wang et al., 2015). Similarly, *ARGONAUTE7* (*S1AGO7*) in tomato was shown responsible for the upward-pointing growth of pedicels (Lin et al., 2016).

The straight and curved growth of pedicel, a short stem like organ that links the flower/ fruit to the inflorescence axis, governs the orientation of fruits thus become either erect or pendant. This phenomenon is known to be regulated by a controlled sequence of cell proliferation, differentiation and elongation (Bundy et al., 2012). A study on the pedicel growth of *Arabidopsis* has provided a clue on how the proximal constriction of pedicel along abaxial and lateral side led to the downward bending of the distal pedicel (Douglas and Riggs, 2005). In 2002, Douglas and colleagues described *KNAT1* and *ERECTA* having a role in bending of pedicels at nodes and downward orientation of flowers, and proposed the loss of chlorenchyma tissue at the node adjacent to lateral organs and in the abaxial regions of pedicels is the reason for (Douglas et al., 2002). Later, a detailed function of *LEAFY (LFY)* for the pendant growth of pedicel in *Arabidopsis* by reducing the cortical cell length in the abaxial domain and its relation with *BREVIPEDICELLUS* (BP)

was described (Yamaguchi et al., 2012). Consistent with these *KNAT6* and *KNAT2* were shown to have a role in downward pointing phenotype of *Arabidopsis* inflorescences (Ragni et al., 2008).

Over the years, genetic studies on the fruit orientation have been reported in pepper (Kaiser 1935; Cheng et al., 2016). These studies have revealed that fruit orientation is qualitative trait controlled by a single gene located on chromosome 12. The early study by Lee et al (2008) using Saengryeog 211 (pendant), Saengryeog 213 (erect), and their F_1 and BC_1 progeny demonstrated *Up* as a recessive gene and developed a CAPS marker which was mapped at the genetic distance of 4.3 cM from the locus. In 2016, an ultra high density bin mapping using RIL population detected a major QTL controlling fruit orientation *FP-12.2*. This reported locus, residing at 199.6 Mb on chromosome 12 in the CM334 reference genome, explained >40% of the phenotypic variation (Han et al., 2016). However, based on the available results of previous studies, dissection and identification of responsible genes underlying these QTLs is difficult due to large region in the chromosome.

In this study, we used two RIL and three F_2 populations for linkage analysis and combined them with GWAS using the core collection with the objective to re-evaluate the genetic effect of the *Up* gene in a new population, fine-map the *Up* locus and to infer potential candidate genes responsible for fruit orientation.

MATERIALS AND METHODS

Plant materials

This study included three F_2, originated from a cross between erect 'Micropep' (*Capsicum annuum*) and pendant 'Jeju' (*C. annuum*) with 219 F_2 plants, erect (*C. annuum*) 'Lp97' and pendant 'A79' (*C. annuum*) with 379 F_2 plants and erect (*C. annuum*) 'U92' and pendant (*C. annuum*) 'A106' with 63 F_2 plants and two RIL populations which belongs to *C. annuum* , vis erect 'Perennial' and pendant 'Dempsey' with 77 lines and erect '35001 (F)' and pendant '35009 (C)' with 174 lines. They are respectively coded hereafter as MJ, LA, UA, PD and FC from the initials of their respective parents` name. Parental lines Lp79, A79, U92 and A106 were provided by EcoSeed P.L.C, South Korea. A core collection composed of 196 *C. annuum*, 25 *C. baccatum*, 21 *C. chinense* and 14 *C. frutescens* was used for GWAS (Table 2; Lee et al., 2016).

Growing conditions and phenotyping

Five seeds of each line/accession of bi parental populations were sown either on the field or greenhouse at Seoul National University, Suwon, South Korea. While MJ was grown in the greenhouse during 2017, it was grown in the open field in 2018; FC and PD were grown in greenhouses for two consecutive years (2016 and 2017) and (2017 and 2018), respectively; and UA and LA populations were grown once in 2018 and 2019, respectively both

in greenhouses. The core collection was planted in a greenhouse of Biotong Seed Co. Ltd., Anseong, Republic of Korea during 2018. F_1 and F_2 of biparental lines were used for allelism test and data on fruit orientation was recorded from all plants included in the experiment. Quantitative traits included in this experiment were measured from five representative samples of each line. Length and width of pedicels and fruits were measured using caliper and fresh weight of fruits was determined using digital balance.

Light microscopic observation

Light microscopic analysis was used to observe the cross-sectional and longitudinal part of the pedicel at the point of curvature for pendant types and at near attachment point to the fruit for straight types. The cut part was stained with 0.05% toluidine blue O in 2.5% sodium carbonate solution and semi-thin sections were observed and photographed using an Axiophot photomicroscope (Zeiss) as described previously (Jeong 2014).

Genomic DNA extraction

Two to three young leaves from each plant were used for DNA extraction. Leaf tissues were grinded using TissueLyserII (Qiagen, Haan, Germany). The concentration and purity of DNA samples was measured using a NanoDrop (NanoDrop Technoogies, Wilmington, DE, USA) after DNA was extracted using cetyl trimethylammonium bromie (CTAB) method. DNA samples showing absorbance ratios at 260/280 nm above 1.8 were diluted to a final

concentration of 50 ng/μl with distilled water for downstream analysis.

Genotyping-by-sequencing

GBS libraries from the DNAs of FC and MJ populations were constructed based on two restriction enzymes, *Pst*I and *Mse*I, using the SBG 100 -kit v2.0 (KeyGene N.V., Wageningen, the Netherlands) as described in the manufacturer's protocol. Pooled libraries were sequenced using an Illumina HiSeq2000 sequencing system Illumina, Inc, San Diego, US) at Macrogen (Macrogen, Seoul, Korea).

Development of SNP markers and linkage analysis of molecular markers

Reanalysis of the sequencing data of PD (Han et al., 2018) and analysis of GBS results of FC and MJ was performed using the updated genome reference of *C. annuum* cv. CM334 ver 1.6 (Kim et al., 2014; http://peppergenome.snu.ac.kr/), "L_Zunla-1" (Qin et al., 2014; http://peppersequence.genomics.cn) and newly developed 'Dempsey' (unpublished) reference genomes. Quality control and GBS sequence data trimming were performed using CLC Genomics Workbench v6.5 (QIAGEN, Aarhus, Denmark) with Q20, and the minimum length of reads was set to 30 bp. The trimmed sequence reads were mapped to each of the reference genomes by Burrows-Wheeler Aligner version 0.7.12 (Li 2013). Picard Tools version 1.119 and SAMtools version 1.1 were used for read grouping

and sorting (Li et al., 2009). For genome wide SNP calling, Genome Analysis Toolkit Unified Genotyper version 3.3 was used. High quality SNPs with quality value larger than 30 and minimum depth 3 were selected for further analysis. Bin linkage map was constructed, and following sliding window approach, the missed data and genotyping errors were imputed as described previously (Han et al., 2016). Windows QTL Cartographer 2.5 was used for the analysis of possible fruit orientation related QTL on PD, FC and MJ populations using a powerful analytical technique, Composite Interval Mapping (CIM) (Zeng, 1994).

Settlement of MLM Under Progressively Exclusive Relationship (SUPER) GWAS was utilized by R package of Genomic Association and Prediction Integrated Tool (GAPIT) using default parameters (Wang et al. 2014). All the probabilities generated in the association runs were transformed by $\log10_P(0.05)$ following previous reports (Siddique et al,. 2019). Scores for chromosome 12 were then inspected in Manhattan plots to determine whether the SNPs reached the significance threshold. The $-\log_{10}P$ values of SNPs from the GWAS were adjusted by Bonferroni multiple test correction.

Genotyping was performed using high resolution melting (HRM) as described by Park et al. (2009). Specific to HRM analysis, PCR was performed in 20 µl of reaction mixture including 60 mM KCl, 10 mM Tris-HCl, 2.5 mM $MgCl_2$, 0.25 mM dNTP, 10 pmol forward primer, 10 pmole

reverse primer, 1 unit Taq polymerase, 1.25 µM Syto9 and 50 ng gDNA. The PCR program was 95^0C, for 4 min, 35 cycles of 95^0C for 20 s, 57^0C for 20 s, and 72^0C for 40 s, after 72^0C for 10 min for extension with a thermocycler (My Cycler, Bio-Rad). Fluorescence signals were measured for each 0.1^0C increase from 65^0C to 95^0C with LightScanner®.

Linkage analysis was performed using CarthaGene software (Schiex and Gaspain, 1997). To construct a linkage group, a logarithm of the odds threshold was set 3.0 score and a maximum distance was set 30 cM. Candidates of linkage groups was determined based on physical position of the markers in the reference genomes. Genetic distance between markers were determined in centi-Morgan (cM) using the Kosambi mapping function. The resulting genetic linkage maps were drawn using MapChart 2.3 software (Voorips 2002).

PCR amplification and localization of the *Up* gene

Throughout the experiment, polymerase chain reaction (PCR) amplification was carried out in a total volume of 50 µL reaction mixture containing 50 ng template DNA (3 µl), 10X PCR buffer, 0.1 mM dNTP mix, 10 pmoles/µl of each primer, 1 U Taq DNA polymerase (Takara Korea Biomedical Inc., Seoul, Republic of Korea). The PCR cycling program was as follows: 95°C for 10 min; 30 cycles at 95°C for 30 s, 56°C for 30 s, 72°C

for 30 s; and a final step of 72°C for 10 min. PCR products were analyzed on a 1% agarose gel in 1X TAE buffer and visualized using Biorad Universal Hood II Gel Doc System after staining with ethidium bromide.

Based on Basic Local Alignment Search Tool (BLAST), the physical positon of the *up*CAPS marker (Lee et al., 2008) was in the lower arm of chromosome 12. For localization of the *Up* gene, 48 randomly selected SNPs between 210-212 Mbs region of chromosome 12 from the whole genome sequenced PD (Han et al., 2016) were used (CM334 v.1.6), and applied on the segregating population of PD. PCR primers were designed using Primer3 software (http://web.bioneer.co.kr/cgi-bin/primer/primer3.cgi) (Rozen and Skaletsky, 2000). Co-segregation analysis of pheno-genotypes enabled us to develop closer SNP markers that can be used in FC, MJ, UA and LA populations. Polymorphic markers co-segregated with the phenotype were combined for linkage analysis using the MAPMAKER/EXP3.0 program and is used to develop genetic linkage map (Fig. 7).

Gene cloning and sequencing

To confirm SNPs detected by HRM and validate the nucleotide variation between the parental lines in the candidate regions, sequence analysis was performed by cloning the target region or directly using PCR products obtained from plants, which were ultimately found to contain different alleles

for the fruit orientation. The PCR products were purified by a Zymoclean PCR Purification Kit following the manufacturer's protocol (Invitrogen Korea, Seoul, Korea), and cloned into a pGEM-T vector (pGEM®-T Easy Vector Systems, Promega, Seoul, Korea). Purified plasmids were sequenced using universal primer, M13 at Macrogen. Nucleotide sequences were aligned using the MegAlign program (DNA Star, DNASTAR, Inc., Madison, WI, USA).

RESULTS

Fruit orientation in pepper and its temporal change

Fruit orientation, which is governed mainly by the curvature of pedicels, may vary depending on the relative position at which the bending occur on the pedicel; near the point of attachment with the branch or further away towards the fruit (Fig 1A). Those plants having a curvature at the base of the pedicel which is observed in all parental lines used in the experiment (such as Fig 1B, 1C), are strictly pendant types with no confusing phenotypes even during early flowering period. Some lines in the segregating populations, however, such in LA F_2 (LP97 X A97), fruit orientation cannot be determined during flowering and early fruiting stages as the phenotypes can be shifted from one state to the other maybe based on the variation in weather or growing stage. In others, however, lateral pendant (horizontally oriented fruits) and lateral upright are observed not only because of the curvature of the pedicels, instead the overall plants loose architecture, branch growth habit, fruit weight and pedicel length (Fig 2). Accordingly, loose branched peppers with long pedicel and heavy fruits tend to grow horizontally or become pendant, even though the pedicels are erect.

Pedicel morphology in segregation population and its correlation with other related traits

To establish a close understanding of pedicel morphology and evaluate its correlation with some fruit related traits towards the fruit growth orientation, we took observations and measurements of representative samples from PD in 2017 and MJ in 2019. Generally, the average pedicel length was higher for PD than MJ with respective values of 3.4 cm and 2.9 cm with a range between 1.8 to 6.5 cm, and 1.9 to 5.8 cm, respectively. There was no variation in average pedicel thickness between the two populations, where both measured 0.4 cm, while the majority (80%) are between 0.35 to 0.45 cm. The average pedicel length was found invariably longer for pendant oriented types than the upright ones in both populations, while there was no variation in pedicel thickness for both types and populations. Straight and curved pedicels of MJ parental lines were analyzed using light microscope by horizontal sectioning of mature tissues (Fig 3). It was shown that in curved pedicels, abaxial side collenchyma cells around the curvature of pedicels are smaller and denser than the collenchyma cells at the adaxial side, while other components of the internal tissue (pith, xylem and phloem) are unaffected (Fig 3A,3C). In the case of the straight pedicels (Fig 3B), there was no variation in cellular number and size for all components.

Pearson correlation matrix result from the PD population showed positive correlation among fruit weight, pedicel length and pendant

orientation (Table 3). The observation has revealed the negative correlation between length and thickness of pedicels; and thickness and pendant fruit orientation.

Inheritance analysis of fruit position

Two intraspecific segregating populations were used to understand if fruit orientation among different populations in pepper is similar, study its inheritance and confirm what has been reported previously (Lee et al., 2008; Han et al., 2016; Cheng et al., 2016). *C. annuum* parental lines A79 and Jeju showed pendant fruit orientation, whereas LP97 and Micropep are erected. Although fruit orientation seems to be a qualitative trait, in this study, we have observed four types of orientation; vertical upright in which all fruits have vertically in up position, vertical pendant where all fruits are vertically pendant, lateral pendants where majority tend to have pendant growth with some horizontal orientation; and finally lateral up in which majority tend to have vertically upward position with some horizontal growth (Fig 1A to 1D). For inheritance study we considered the lateral pendants as pendant and lateral up as up.

Accordingly, of 379 F_2 plants from LA population, 291 were pendant and 88 were erect, fitting a 3:1 ratio ($X^2 = 0.64$, $p = 0.50$) for a single dominant gene (Table 1). Similarly from 214 F_2 plants of MJ, 154 showed pendant phenotype and the remaining 60 erect following the Mendelian segregation

ratio (3:1, X^2 =1.05, p = 0.30). Overall, these results suggests that the pendant fruit orientation in pepper is dominant over erect types.

Fruit orientation was segregated in the four *Capsicum* species of the core collection included in this experiment. There were 57, 4, 1, and 12 accessions with upright fruit position in a total of 196, 25, 21 and 14 germplasms of *C, annuum, C. baccatum, C.chinense* and *C. frutescens*, respectively (Table 2).

Table 1 Segregation analysis of fruit orientation for LA and MJ populations

Population	Generation	No. of plants	Phenotype		Expected ratio		
			Pendant	Erect	Pendant: Erect	X^2	P-value
A79	Parent	20	20	-	20:0		
LP97	Parent	20	-	20	0:20		
LA	F_2	379	291	88	3:1	0.64	0.50
MJ	F_2	214	154	60	3:1	1.05	0.30

A79 and LP97 are parental lines for LA population; MJ is a population derived from Micropep and Jeju parental lines

Table 2 Segregation of different pepper species in the core collection for fruit orientation, used for GWAS

Species	Fruit orientation		Total
	Erect	Pendant	
C. annuum	57	139	196
C. baccatum	4	21	25
C. chinense	1	20	21
C. frutescens	12	2	14

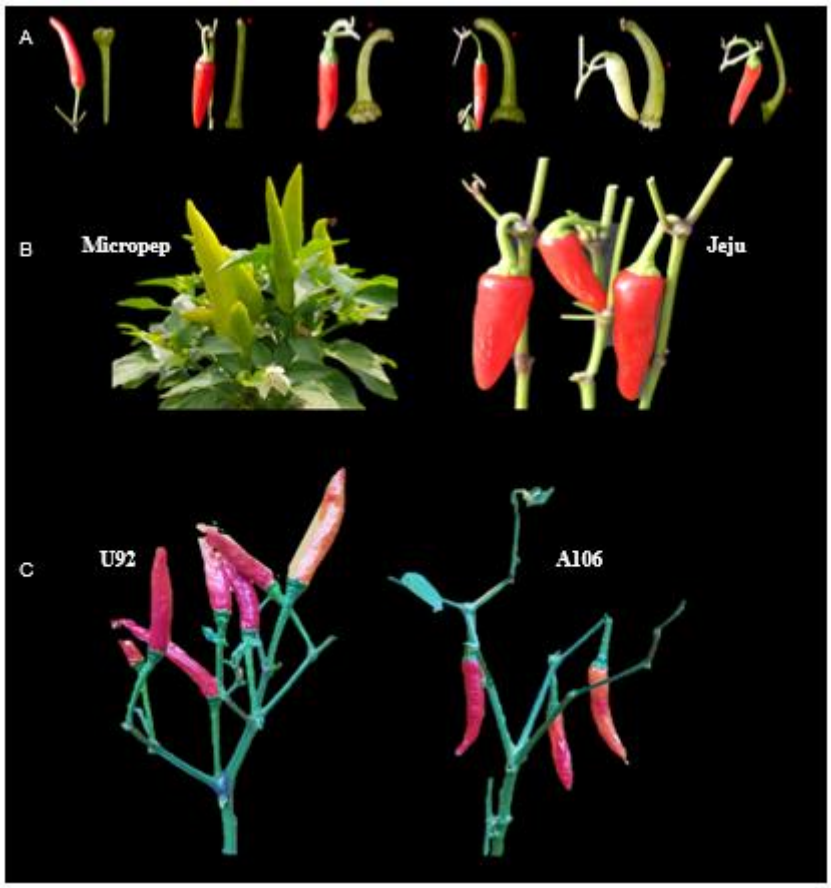


Figure 1 Pedicel curvature and fruit orientation phenotype. (**A)** Pedicels with different positions of curvature, marked with red stars. (**B and C)** Vertically upright and vertically pendant fruit type.

Figure 2 Variation in fruit orientation. (**A)** Lateral pendant phenotype; (**B)** Lateral upright phenotype

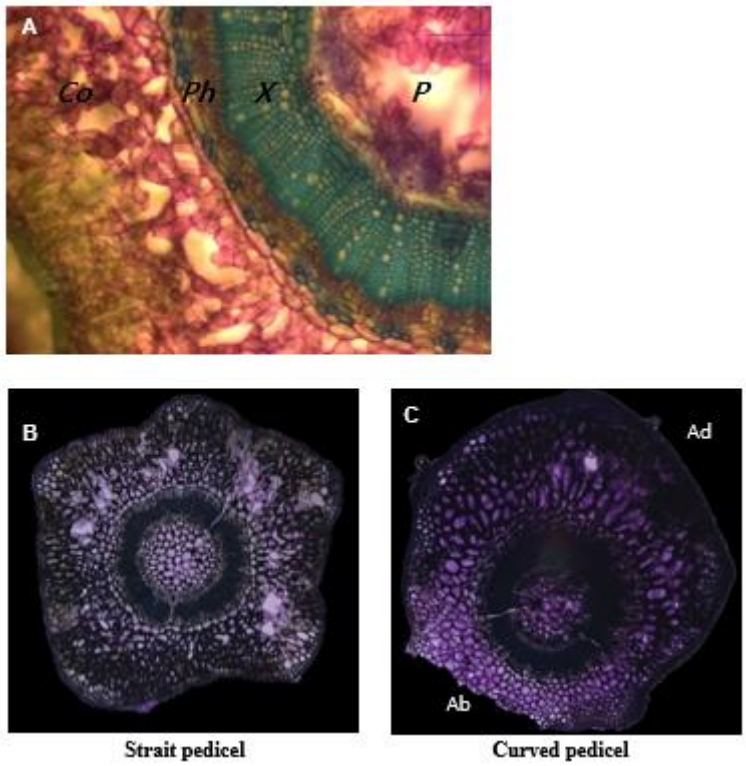


Figure 3. Cross sectional view of pedicel cells (**A**) Closed view of the pedicel cell types; collenchyma (*Co*), phloem (*Ph*), xylem (*X*) and pith (*P*); (**B**) Cross section view of erect pedicel, uniform distribution of cells across the surface; (**C**) Cross sectional view of pendant pedicel at the point of curvature, where abaxial cells (Ab) are av. 2.6X compacted at the curvature than the cells at the adaxial phase (Ad).

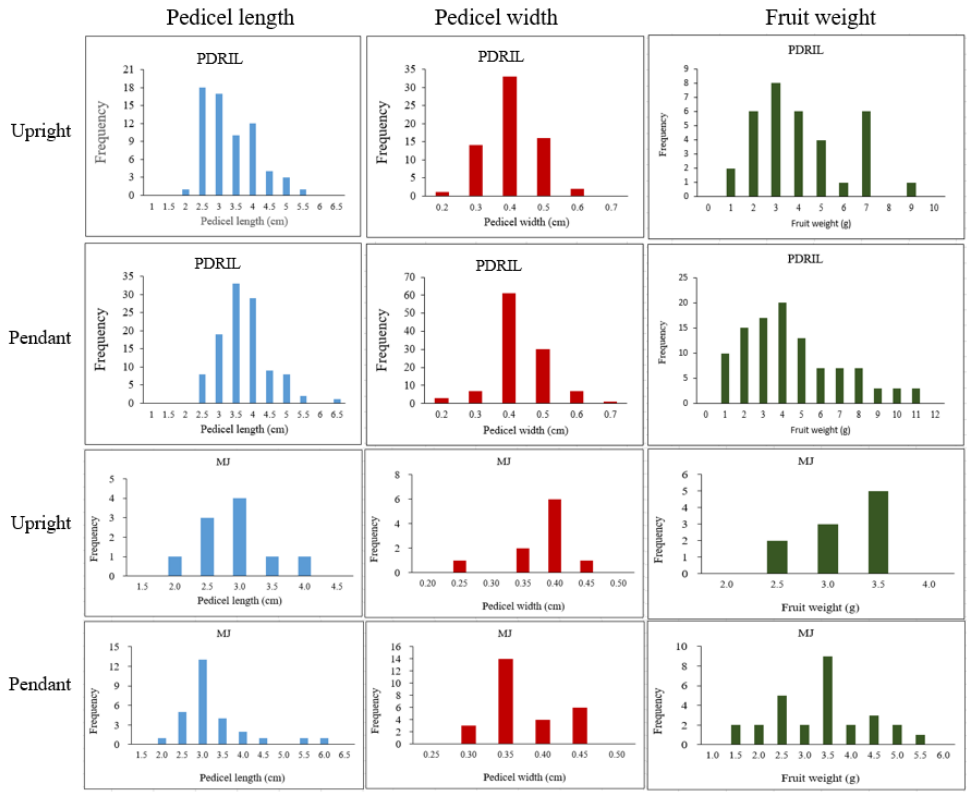


Figure 4 Frequency distribution of traits related to fruit orientation. Pedicel length, pedicel thickness (width) and fruit weight of PD and MJ F_2 populations associated with the upright and pendant fruit orientation.

Table 3 Pearson correlation matrix of fruit orientation related traits using PD.

Variables	Pedicel Length (cm)	Pedicel width (cm)	Fruit weight (g)	Pendant fruit orientation
Pedicel Length (cm)	1	-0.021	0.164	0.222
Pedicel width (cm)		1	0.137	-0.069
Fruit weight (g)			1	0.135
Pendant fruit orientation				1

Values in bold are different from 0 with a significance level alpha=0.05

Localization of the target region for *Up* locus

Considering the lateral pendant and lateral upright phenotypes as pendant and erect, respectively, our inheritance study suggested fruit orientation is controlled by single or one major gene. However, the fruit orientation is not strictly qualitative, as it was described by Cheng and colleagues (Cheng et al., 2016). Therefore, we wanted to perform analysis of responsible loci using QTL approach on different biparental populations and Super GWAS using the diversity panel, and find a colocalized region before fine mapping the gene. Trimmed sequencing reads of PD, FC, MJ and core collection which were obtained previously (Han et al., 2016; Lee et al., 2016), were aligned to the *C. annuum* cv. CM334 reference genome ver. 1.6. After removing SNP positions which are missed on the parental lines and those having more than 70% missing, we detected SNPs of 123,502 for PD, 76,952 for FC, 1,890 for MJ and 1,048,576 for core collection. For all populations, there was similarity in SNPs distribution pattern at each chromosome positions, with a slight difference in MJ (Fig 5).

In the PD we identified two QTLs that explained 49.1% (LOD of 20.9) to 51.4% (LOD of 22.5) of phenotypic variation in the physical region between 203 and 208 Mbp of chromosome 12 (Fig 6B). There were two high peaks with LOD values of 5.7 and 6.4 (Fig 6C) and three QTL for FC population spanning between 97.6 to 229.8 Mbp in the same chromosome.

These three QTLs contributed 7.2-10.5% to the observed phenotypic variations (Fig 6C). Ten QTLs were detected on MJ in the range between 57.7 and 242 Mb, explaining 8.1 to 39.6% of the phenotypic variation (R^2) with the maximum LOD value of 13.8 (Fig 6D; Table 5).

In addition to the QTL mapping, we used SNP data of pepper core collection (Lee et al., 2016) of 256 plants (Table 2) to conduct genome wide association study (GWAS) for fruit orientation. Minor allele frequency of >0.05, SNP coverage of > 0.6 and inbreeding coefficient > 0.8 were used as a filtering criteria to obtain 176,951 high quality SNPs for the downstream analysis. Excluding all false positive results even above the Bonferroni correction, by randomly selecting –log10(p) values > 26, only on chromosome 12, we identified 14 highly significant SNPs associated with fruit orientation between 205 and 214 Mbps (Fig 6A).

Combining these results, a colocalized SNP area (212 to 214 Mbp) for the different populations and two approaches was taken as the candidate mapping region for the fruit orientation (Fig 7A). In this 2 Mb genomic region that surrounds the *Up* locus (Chr 12: 212,025,444-216,980,510 bp) we characterized the linkage disequilibrium (LD) and haplotype block structure. There were four blocks each containing 5, 3, 2 and 6 SNPs with the respective size of 102 kbp, b4 bp, 2 bp and 377 kbp (Fig 7B and 7C).

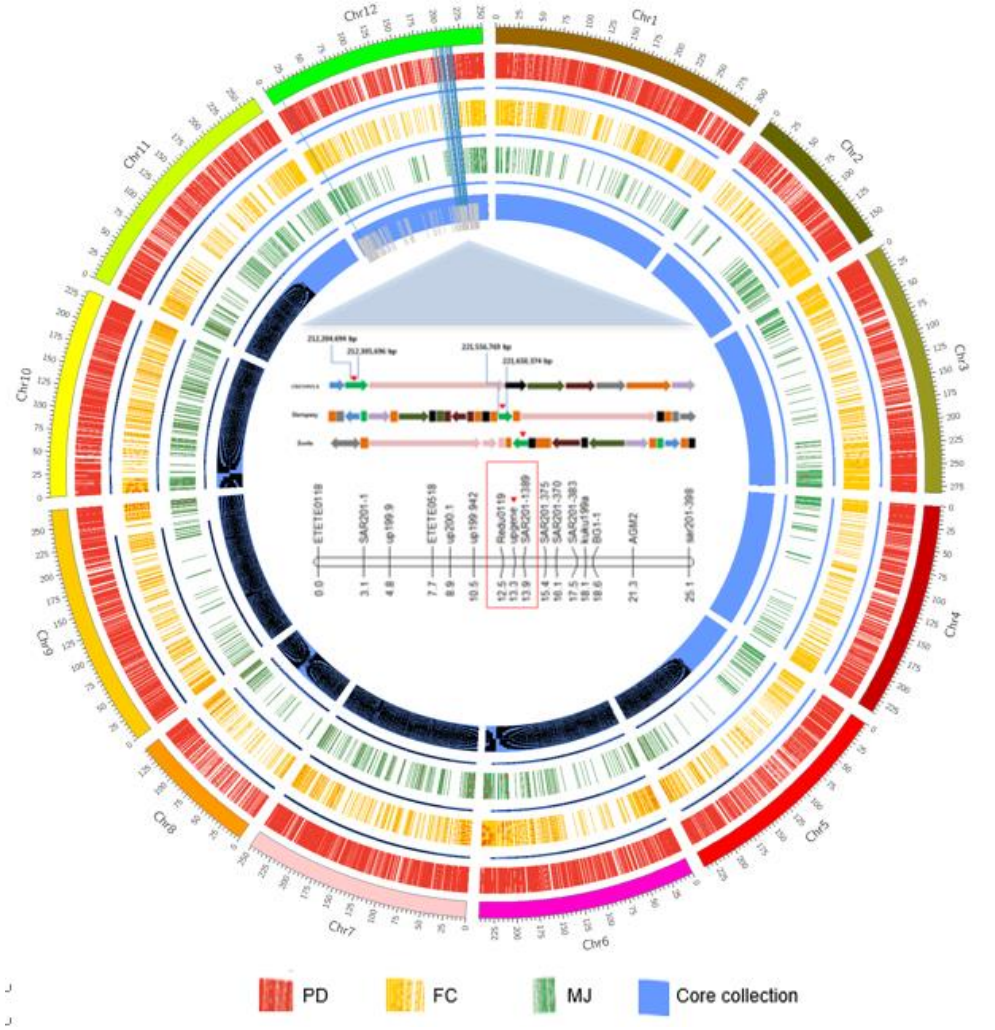


Figure 5 Distribution of SNPs and target region for the fruit orientation in chromosome 12. The outermost circle represents 12 pepper chromosomes in different colors; the middle circle represents SNP in PD, FC and MJ, from the outside to inside. The blue circle is representing a high SNP density of the core collection which is used for GWAS analysis. The most filtered SNPs in chromosome 12 is shown on the most inner circle with silver colored lines.

The picture inside the circle shows the delimited mapping region, and the *Up* locus (represented by red arrow head) flanked by the two markers on the three reference genomes (from top to down, CM334 v1.6, Dempsey (unpublished) and Zunla).

Table 4 Primers used in this study

Table 5. Detected QTL regions for fruit orientation of the three populations.

Population	Chromosome	QTL Position	LOD	Additive	Dominant	R2
PD	12	165.81	20.9	0.5004	0	49.1
PD	12	168.71	22.5	0.5156	0	51.4
FC	12	113.41	5.7	-0.1824	0	9.6
FC	12	119.71	4.2	-0.1634	0	7.2
FC	12	180.61	6.4	-0.2	0	10.5
MJ	12	78.21	4.9	0.1454	-0.4901	17.75
MJ	12	89.51	5.4	0.1735	-0.473	25.0
MJ	12	102.91	7.4	0.2187	-0.5277	38.7
MJ	12	116.81	8.4	0.0911	-0.6013	16.1
MJ	12	127.91	7.4	0.087	-0.6285	12.4
MJ	12	138.51	11.6	0.0343	-0.6746	8.1
MJ	12	146.31	13.8	0.0739	-0.7257	11.0
MJ	12	153.81	10.1	0.03	-0.6113	10.3
MJ	12	159.71	9.7	0.111	-0.5718	39.6
MJ	12	171.71	7.9	0.0722	-0.536	19.2

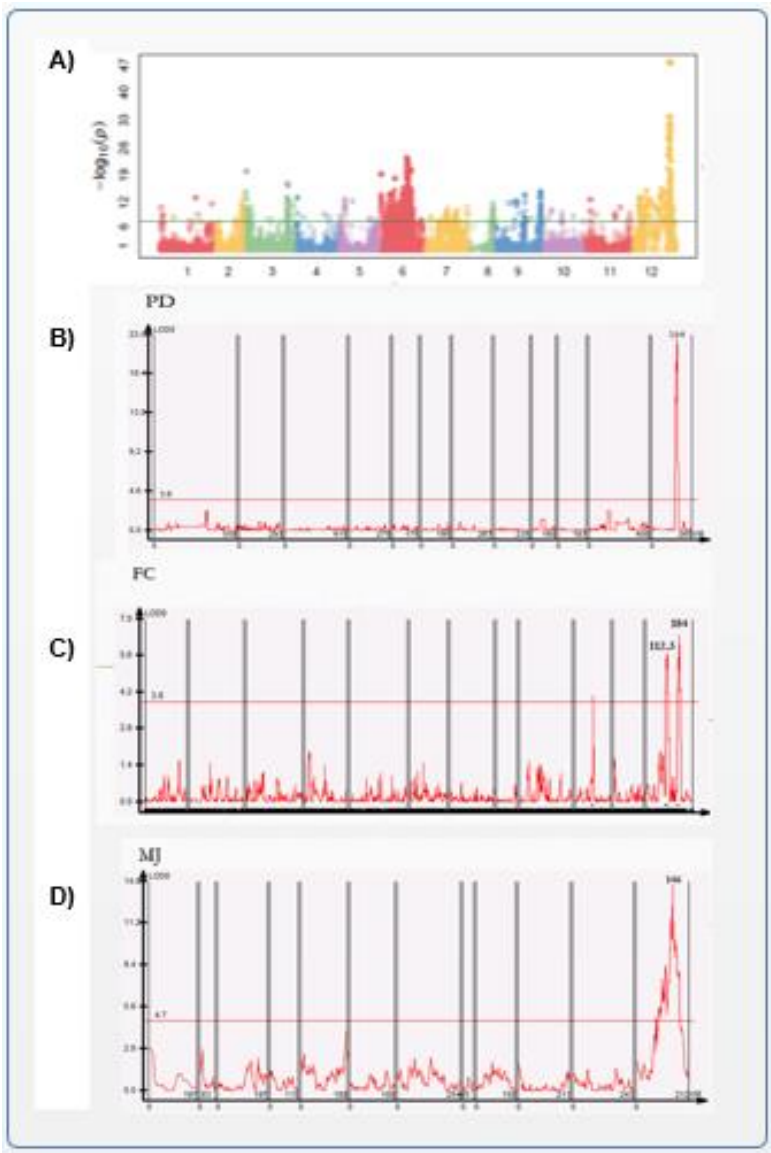


Figure 6 GWAS result of the core collection and QTL results of different population for fruit orientation. **A**) Manhattan plot of the core collection, showing significant peak on chromosome 12 (-LogP of 47) for the fruit orientation; **B**) to D) QTL results for PD, FC and MJ populations all showing similar significant QTL region on chromosome 12, with one high pick in PD, two picks for FC and wider range at the lower LOD score of MJ.

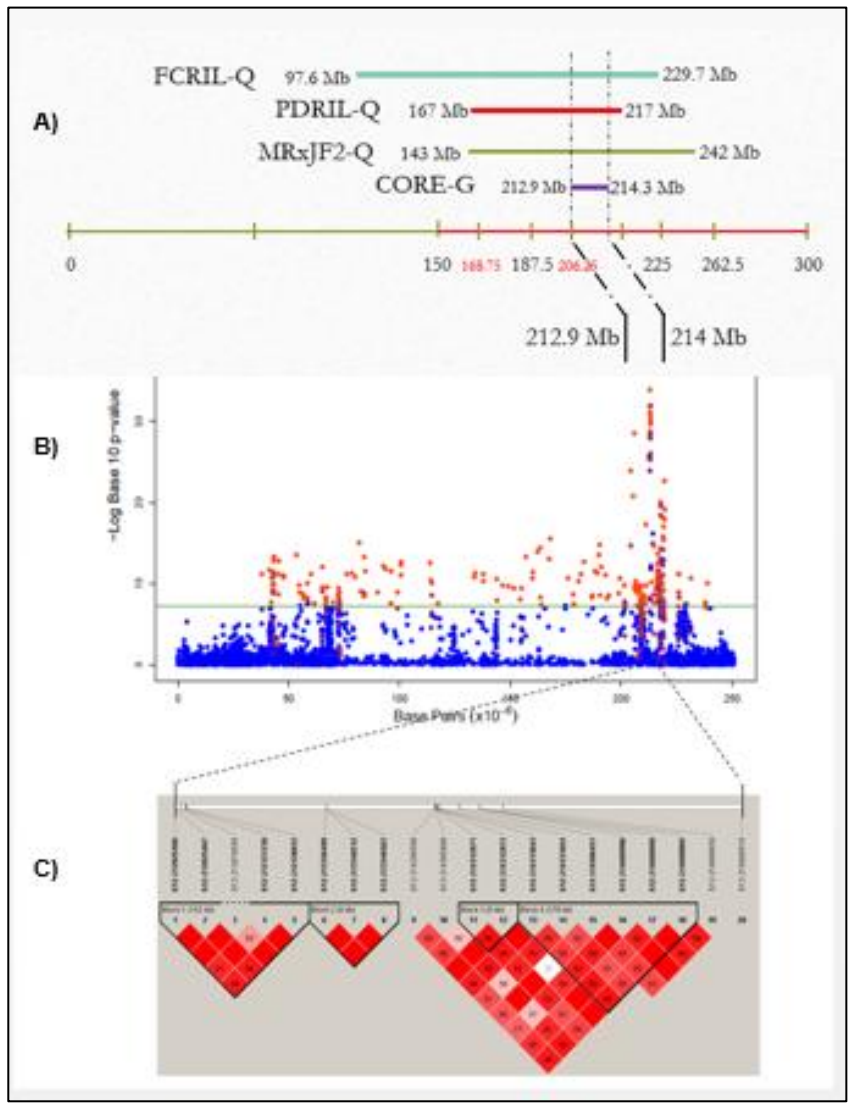


Figure 7 Co-localized region for the *Up* locus and LD block.

A) QTL regions controlling the fruit orientation for each population under the study (i.e., FC, PD, MJ and core collection) and GWAS result from the core collection. The co-localized region is indicated by a dotted vertical line; **B)** Manhattan plot of chromosome 12. **C)** Summary of local LD and haplotype blocks for a 2 Mb genomic region that surrounds *Up* (Chr 12: 212,025,444-216,980,510 bp). LD plot, generated in Haploview (Barrett et al., 2005), indicates r^2 values between pairs of SNPs multiplied by 100. Haplotype blocks (blocks 1-4) in the *Up* genomic region were defined with the confidence interval method (Gabriel et al., 2002).

Fine-mapping of *Up* locus and validation of markers

Initially we developed 48 different HRM markers around the *Up* locus reported by Lee et al. (2008) in a region between 210-212 Mbps and used for genotyping of PD population (Table 4). Two markers 'up2750_199' and 'up3450_200' were found close to the locus with their respective distance of 6.8 and 2.9 cM from the gene. With this marker information, additional SNPs within these regions were used to identify close markers, up199.73 and up199.942 at 2.2 cM and 1.4 cM distances from the gene. These two markers were located in the scaffold1243 and 214 Mbp regions, respectively. Based on co-segregation result of phenotype and genotype, additional markers were developed and evaluated on different populations (Fig 8). Four markers (Etete0318 with 17.9 cM, Kuku199a with 3.2 cM, Up199.942 with 4.8 cM and SARS0419 with 23 cM) for UA population were identified in the 205-216 Mbps physical positons. The closest markers, Kuku199a and Up199.942 were both localized in 214 Mbp region for this population. For MJ, 11 markers (Up200.1, up201, SAR201-398, SAR201-413, SAR201-462, SAR2011386, SAR201-1, SAR201-308/616, Up199.46 and SAR201-154) were located on each side of the *Up* locus in 33.8 and 32.6 cM region with the closest flanking markers'SAR201-413' and 'SAR201-462' situated at 5.2 cM and 7.3 cM distance from the locus both at 212 Mbs physical region (Fig 9). Summarizing these result, common flanking marker Up199.942 was closest for the UA and PD at a distance of 4.8 and 1.4 cM respectively and

Up201 identified as close marker for the three population, vis. PD, MJ and FC with their respective distance of 0.7, 4.1 and 0 cM.

Finally, a segregating population LA with 379 F_2 individuals was used to improve the accuracy of the fine-mapping in summer 2019. Accordingly, the developed close markers were used for genotyping and identification of recombinants. Using the same analytical method, we successfully narrowed the locus to the physical distance of 101 kb flanked by dominant and codominant HRM markers, Redu0119 (0.8 cM) and SAR201-1386 (0.6 cM) with three and two recombinants, respectively (Fig 10).

Consistent to the stable intron InDel observed throughout 10 parental lines and 10 accessions including six reference genomes for gene 10 (AGL31) we have developed HRM marker (PepUP) at 0 cM distance from the *Up* locus. This marker was able to effectively differentiate all segregating lines of LA and 35 *C. annuum* accessions randomly collected from the experimental station at Suwon.

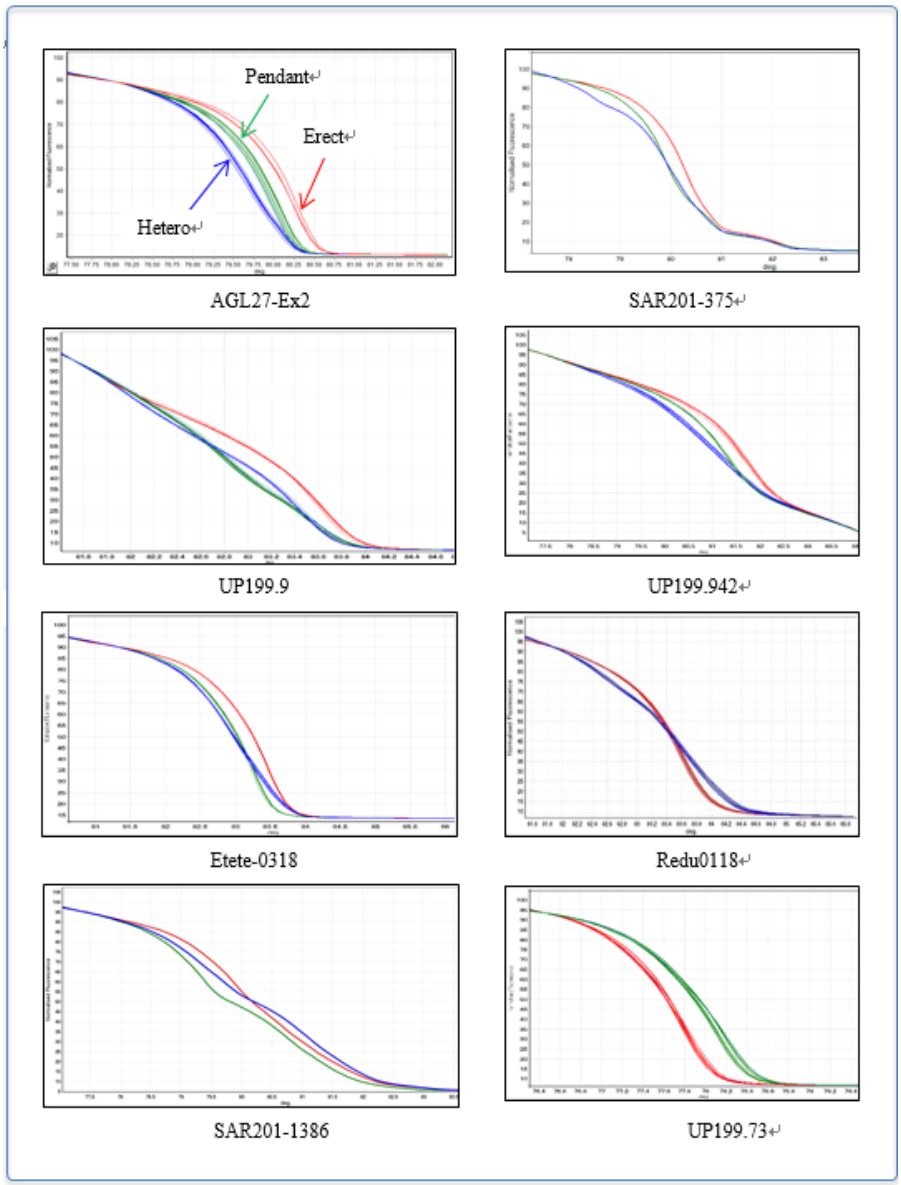


Figure 8 Codominant and dominant polymorphic markers for fruit orientation used on different populations. Markers AGL27-Ex2, SAR201-375, Up199.9, Up199.942, Etete0318, SAR201-1386 are codominant markers predicting the genotype of dominant pendant (green), recessive erect (red) and heterozygous (blue) types. Redu0118 and Up199.73 are dominant markers showing only dominant and recessive types.

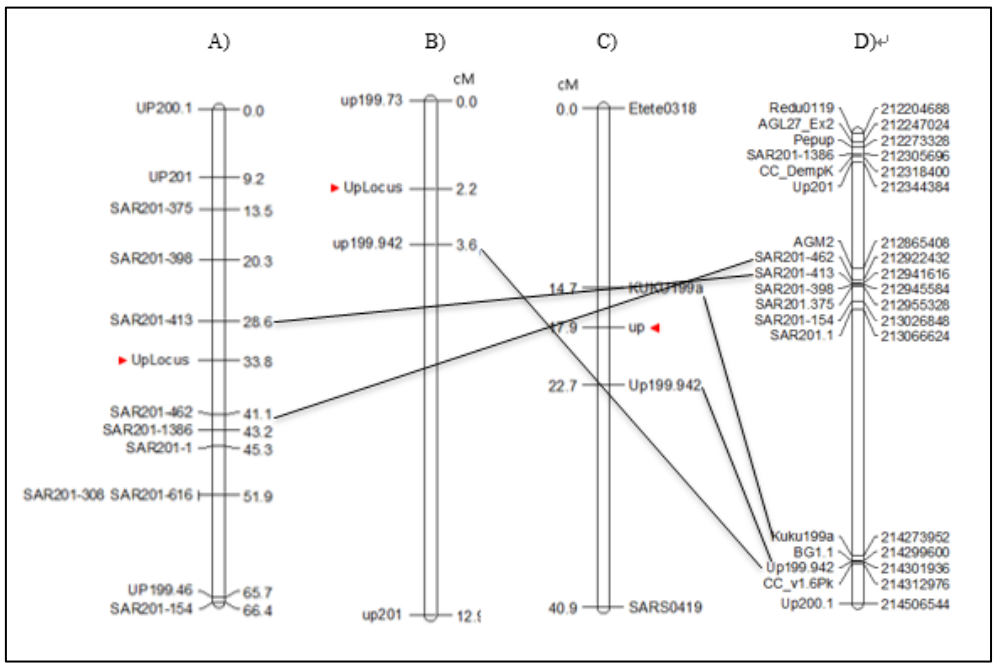


Figure 9 Genetic map of the *Up* locus on three populations. A) Localization of the *Up* locus between two flanking markers SAR201-413 and SAR201-462 using MJ; B) Genetic map showing the *Up* locus is situated between up199.73 and up199.941 in PD population; C) The *Up* locus was situated between Kuku199a and Up199.942 when using UA population was used; D) The physical distance of markers.

MJ with the bigger number of accession has shown similar location to the *Up* gene in the fine mapping region, while UA and PD with 63 and 77 lines were at 1.4 Mbp distant.

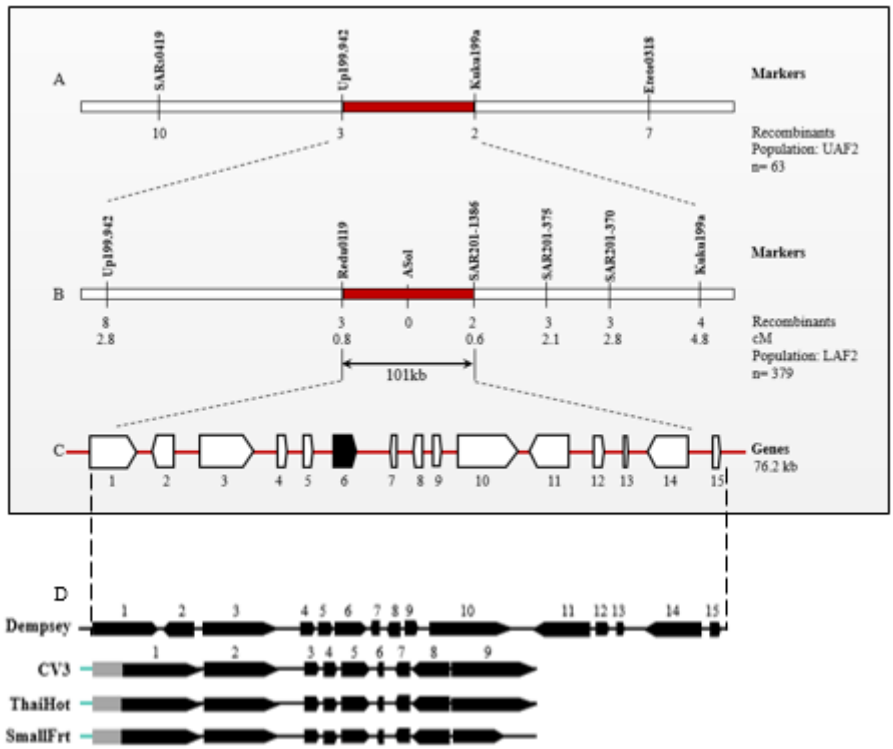


Figure 10 Fine mapping of *Up* locus and sequence alignments of different reference genomes. **A**) The locus was mapped to the interval between Up199.942 and Kuku199a of chromosome 12 using 63 UAF_2 population. Numbers below the markers indicate the number of recombinants that were detected using each marker. **B**) The *Up* locus was fine-mapped to the interval between Redu0119 and SAR201-1386, which are 0.8 cM and 0.6 cM far from the locus, respectively using 379 LAF_2 population. Physically the gene is located in 101 kb mapping region of CM334 v1.6, 93.6 kb of Dempsey and 107.7 kb of Zunla genome references. **C**) The 13 predicted genes in the mapping region. Numbers indicate the gene order, arrows indicate the direction and its size the relative size of each gene. The deep dark colored gene (gene number 5) is the most likely candidate gene. **D**) Alignment of candidate genes in four pepper reference genomes, vis. pendant Dempsey and erect CV3, ThaiHot and Small fruit.

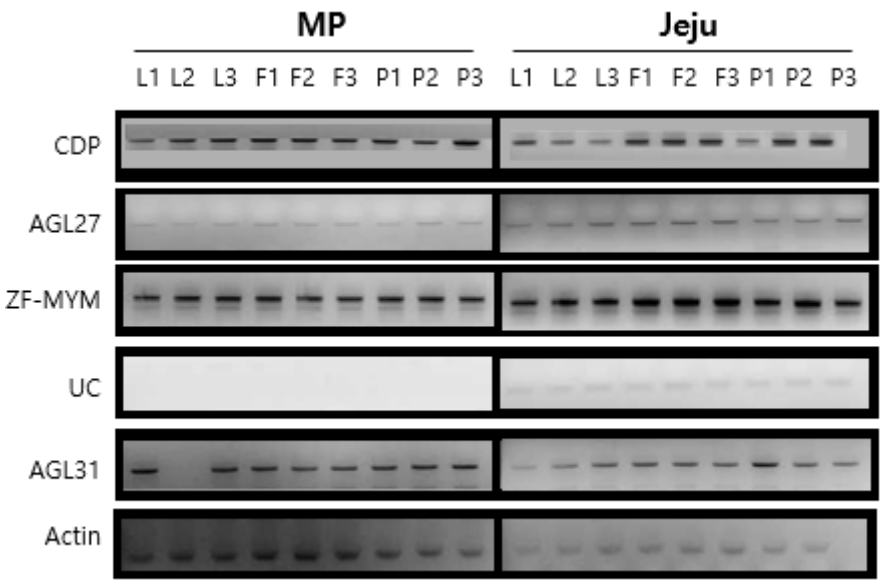


Figure 11 Expression analysis of five selected genes. Agarose gel electrophoresis of PCR products of cDNA from leaf, floral bud and pedicels of the MJ parental lines for five selected genes. No expression variation was evidenced.

Confirmation of sequence variations in candidate genes

According to the pepper genome annotation (Kim et al., 2014), a total of 15 hypothetical genes are predicted to reside in the mapping region with a predicted total gene size of 76.2 kb. Among them, two agamous like MADS-box proteins (AGL27 and AGL31), two zinc finger MYM type protein 1- like, Methyltransferase 1-like, F-box protein, CaMYBA TF, cell differentiation protein RCD 1-like and inversin-like proteins were predicted. Exon number was varied from one to 11, with a size variation between 123 and 1029 bp. The positions and putative functions of these predicted genes are shown in Table 6. Eighteen non-synonymous and thirteen synonymous SNPs that caused conservative amino acid substitutions were existed between Redu0119 and SAR201-1386 between pendant and erect lines. One of the non-synonymous nucleotide substitutions in gene 1 caused aliphatic amino acid change from alanine to valine with law coefficient of difference. The SNP in the second gene has changed a hydroxyl serine to aliphatic leucine with a corresponding high coefficient of difference with relatively low dissimilarity value based on Sneath's index. The third gene, cell differentiation protein RCD1-like protein displayed two SNPs at the second and fourth exons which resulted in the non-synonymous amino acid substitution from cysteine (hydroxyl) and leucine (aliphatic) to glycine (aliphatic) to tryptophan (aromatic), respectively. The fourth gene, agamous-like MADS-box protein AGL27 which presumed to be a peptide for its short amino acids content had

one conservative amino acid substitution from isoleucine (aliphatic) to leucine (aliphatic). There was no detected SNP in gene 5, whereas one non-synonymous nucleotide substitution and one InDel was observed in the six gene (Zinc finger MYM type protein 1- like gene). The two SNPs detected in the seventh gene (Clone BAC CaM303O04) caused two amino acid changes, vis. aliphatic leucine to acidic glutamine and basic lysine to basic arginine. Gene 8 which encodes for a hypothetical protein exhibited three non-synonymous nucleotide changes with associated non conservative amino acid substitutions from phenylalanine to serine, serine to leucine and from serine to proline with respective coefficient of difference values of 0.81, 0.8 and 0.8. Other important SNPs that caused amino acid substitutions include isoleucine to arginine and methionine to threonine changes which were detected in gene 13 (F-box/kelch-repeat protein At3g06240 like) and gene 15 (CaMYBA genes fro L1-EN), respectively. The only InDel in CDS was detected in gene 6, ZF-MYM (ZF-MYM) of erect plant types which was consistently observed by sequence comparison of upright (CV3, ThaiHot, Small fruit) and pendant (Dempsey, CM334, Zunla and UC Davis) genome references (four reference genomes used here are not published). This stable sequence result and non radical amino acid shifts in most other candidate genes suggested that transcription factor zinc finger MYM could be the candidate for the *Up* gene (Table 7).

The gene annotation of zinc finger MYM suggested the presence of 3 exons and 2 introns (Fig 12) and the predicted coding sequence (CDS) was 750 bp, with a predicted corresponding protein length of 249 amino acids. Alignment of ZF-MYM gene sequence between pendant and erect type revealed one non-synonymous G → A mutation at 106 bp and one InDel at 104 bp positions of the first exon (Figure 13). The observed InDel has resulted in a frameshift mutation and a truncated protein with complete amino acid change at the C-terminus (Figure 14). The gene expression analysis of ZF-MYM which was investigated along with other candidate genes, however, did not show expected variation (Fig 11).

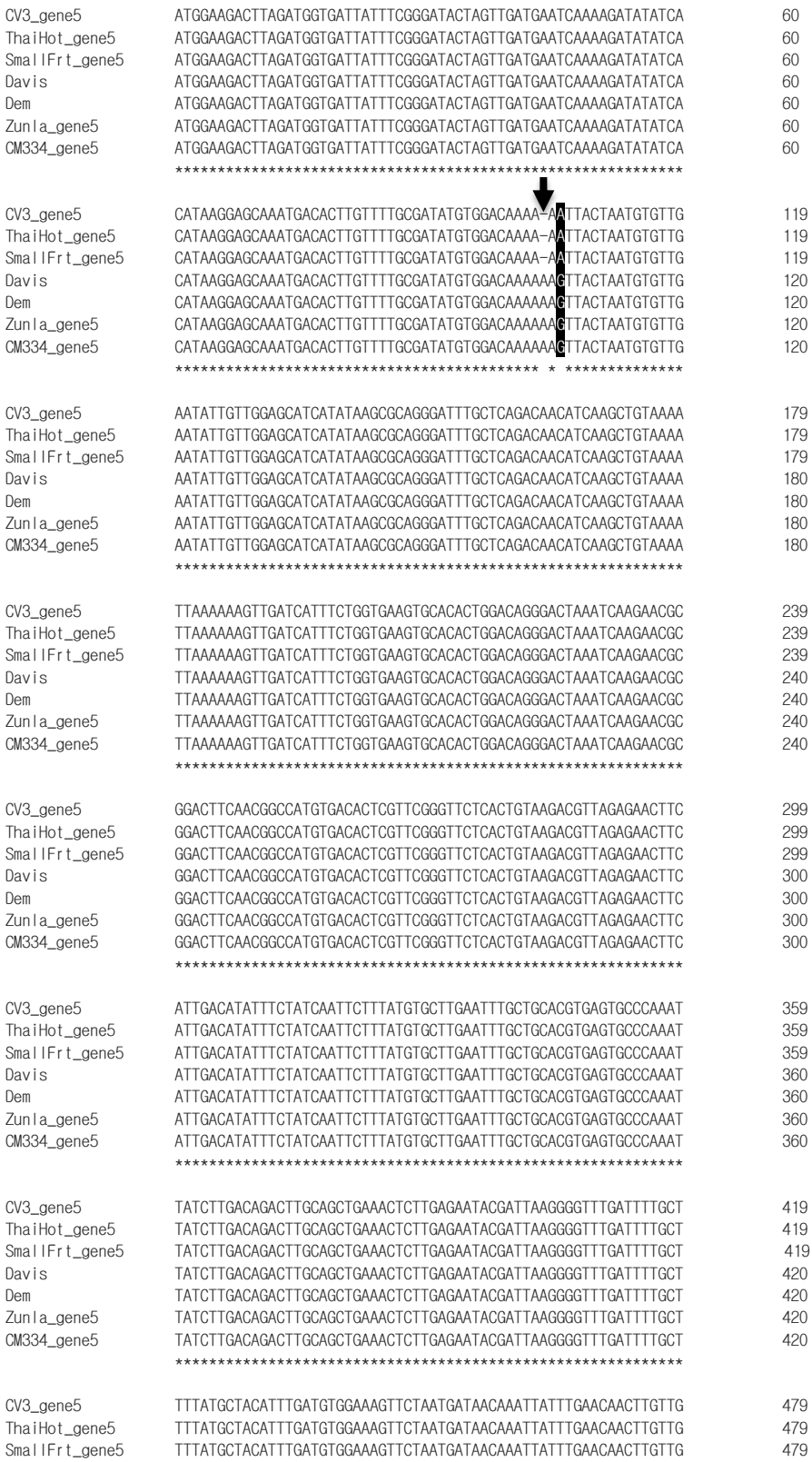
Table 6 Candidate genes in the *Up* mapping region

Candidate gene	Start position	End position	Gene size (bp)	Exon size	No of aa	strand	No of exon	Prediction (ncbi)
1	221,560,842	221,569,821	8,979	225 bp	74	+	3	Hypothetical protein
2	221,570,972	221,575,112	4,140	390 bp	129	-	3	Hypothetical protein
3	221,576,093	221,586,325	10,232	588 bp	195	+	5	Cell differentiation protein RCD1-like protein
4	221,589,750	221,591,626	1,876	225 bp	74	+	2	agamous-like MADS-box protein AGL27, transcript variant X1, mRNA
5	221,592,023	221,593,835	1,812	393 bp	130	+	1	zinc finger MYM-type protein 1-like mRNA
6	221,593,904	221,598,235	4,331	750 bp	249	+	3	zinc finger MYM-type protein 1-like (LOC107849038), mRNA
7	221,598,585	221,599,731	1,146	123 bp	40	-	2	Clone BAC CaCM303O04
8	221,601,977	221,603,819	1,842	444 bp	147	-	1	Hypothetical protein
9	221,605,841	221,607,398	1,557	240 bp	79	+	2	Uncharacterized
10	221,608,344	221,619,821	11,477	1029 bp	342	+	11	agamous-like MADS-box protein AGL31 transcript variant X1, mRNA
11	221,620,268	221,627,962	7,694	195 bp	64	-	3	Uncharacterized
12	221,628,558	221,630,383	1,825	216 bp	71	+	2	inversin-like mRNA
13	221,630,691	221,631,626	935	336 bp	111	+	1	F-box/kelch-repeat protein At3g06240-like
14	221,634,424	221,642,188	7,764	375 bp	124	-	5	Methyltransferase 1-like
15	221,643,166	221,644,576	1,410	162 bp	53	+	1	CaMYBA genes for L1-EN and RT_nLTR_like domain-containing, MYB transcription factor

Table 7 Amino acid substitutions in candidate genes based on reference genomes CM334 and CV3

Candidate gene	Gene name	Exon size	SNP in exon		InDel	Homolog in tomato	Amino acid change		Physico-chemical Distance		
			Non Syno	Syno			From	to	Difference	Dissimila	C.D
1	Hypothetical protein	225 bp	2	3	0	*S. pennellii* and *S. lycopersicum* Chr3/7/10	Alanine (Aliphatic)	Valine (Aliphatic)	64	12	0.4
2	Hypothetical protein	390 bp	1	1	0	*S. pennellii* and *S. lycopersicum* Chr3/7/10	Serine (Hydoxyl)	Leucine (Aliphatic)	145	23	0.8
3	Probable cyclic nucleotid gated ion channel or cell differentiation protein RCD1-like protein	588 bp	2	0	0	*S. lycopersicum* Chr11	Cystine (Hydroxyl)	Glycine (Aliphatic)	159	21	0.31
							Leucine (Aliphatic)	Trypthophan (Aromatic)	61	30	0.36
4	agamous-like MADS-box protein AGL27 (LOC107850711), transcript variant X1, mRNA	225 bp	1	-	0	*S. pennellii* agamous-like MADS-box protein AGL70	Isoleucine (Aliphatic)	Leucine (Aliphatic)	5	5	0
5	zinc finger MYM-type protein 1-like (LOC107849038), mRNA	393 bp	-	-	0	*S. pennellii* zinc finger MYM-type protein 1-like					
6	zinc finger MYM-type protein 1-like (LOC107849038), mRNA	750 bp	1	1	1	*S. pennellii* zinc finger MYM-type protein 1-like	Isoleucine (Aliphatic)	Methionine (Hydroxyl)	10	22	0.05
7	Clone BAC CaCM303O04	123 bp	2	0	0	*S. pennellii* Chr12	Leucine (Aliphatic)	Glutamine (Acidic)	138	22	1
							Lysine (Basic)	Argenine (Basic)	26	14	0.05
8	Hypothetical protein	444 bp	3	1	0	*S. pennellii*/*S. lycopersicum* Chr12	Phenylalanine (Aromatic)	Serine (Hydoxyl)	155	25	0.81
							Serine (Hydroxyl)	Leucine (Aliphatic)	145	23	0.8
							Serine (Hydroxyl)	Proline (Cyclic)	74	24	0.8
9	Uncharacterized	240 bp	0	0	0						
10	agamous-like MADS-box protein AGL31 (LOC107850791), transcript variant X1, mRNA	1029 bp	0	1	0	*S. pennellii* Chr12					
11	Uncharacterized	195 bp	1	0	0	*S. lycopersicum*/ *pennellii* Chr12	Alanine (Aliphatic)	Threonine (Hydroxyl)	58	20	0.41
12	inversin-like (LOC107869181), mRNA	216 bp	0	0	0	-					
13	F-box/kelch-repeat protein At3g06240-like	336 bp	4	1	0	*S. albornozii* voucher	Methionine (Hydroxyl)	Isoleucine (Aliphatic)	10	22	0.03
							Isoleucine (Aliphatic)	Arginine (Basic)	97	34	1.01
							Glutamic acid (Acidic)	Arginine (Basic)	54	31	0.05
14	Methyltransferase 1-like	375 bp	0	3	0	*S. lycopersicum*/ *pennellii* Chr12					
15	CaMYBA genes for L1-EN and RT_nLTR_like domain-containing, MYB transcription factor	162 bp	1	2	0	-	Methionine (Hydroxyl)	Threonine (Hydroxyl)	81	25	0.8

Figure 12 Structure of ZF-MYM coding gene with a total size of 4,331 bp, having three exons with coding sequence (CDS) of 750 bp encoding for 249 amino acids in the wild types. One non-synonymous transition at a positons of 106 bp from a dominant “G” to a recessive “A”; and deletion of one nucleotide at 104 bp shown.


CV3_gene5 ATGGAAGACTTAGATGGTGATTATTTCGGGATACTAGTTGATGAATCAAAAGATATATCA 60
ThaiHot_gene5 ATGGAAGACTTAGATGGTGATTATTTCGGGATACTAGTTGATGAATCAAAAGATATATCA 60
SmallFrt_gene5 ATGGAAGACTTAGATGGTGATTATTTCGGGATACTAGTTGATGAATCAAAAGATATATCA 60
Davis ATGGAAGACTTAGATGGTGATTATTTCGGGATACTAGTTGATGAATCAAAAGATATATCA 60
Dem ATGGAAGACTTAGATGGTGATTATTTCGGGATACTAGTTGATGAATCAAAAGATATATCA 60
Zunla_gene5 ATGGAAGACTTAGATGGTGATTATTTCGGGATACTAGTTGATGAATCAAAAGATATATCA 60
CM334_gene5 ATGGAAGACTTAGATGGTGATTATTTCGGGATACTAGTTGATGAATCAAAAGATATATCA 60
CV3_gene5 CATAAGGAGCAAATGACACTTGTTTTGCGATATGTGGACAAAA-AATTACTAATGTGTTG 119
ThaiHot_gene5 CATAAGGAGCAAATGACACTTGTTTTGCGATATGTGGACAAAA-AATTACTAATGTGTTG 119
SmallFrt_gene5 CATAAGGAGCAAATGACACTTGTTTTGCGATATGTGGACAAAA-AATTACTAATGTGTTG 119
Davis CATAAGGAGCAAATGACACTTGTTTTGCGATATGTGGACAAAAAAGTTACTAATGTGTTG 120
Dem CATAAGGAGCAAATGACACTTGTTTTGCGATATGTGGACAAAAAAGTTACTAATGTGTTG 120
Zunla_gene5 CATAAGGAGCAAATGACACTTGTTTTGCGATATGTGGACAAAAAAGTTACTAATGTGTTG 120
CM334_gene5 CATAAGGAGCAAATGACACTTGTTTTGCGATATGTGGACAAAAAAGTTACTAATGTGTTG 120
CV3_gene5 AATATTGTTGGAGCATCATATAAGCGCAGGGATTTGCTCAGACAACATCAAGCTGTAAAA 179
ThaiHot_gene5 AATATTGTTGGAGCATCATATAAGCGCAGGGATTTGCTCAGACAACATCAAGCTGTAAAA 179
SmallFrt_gene5 AATATTGTTGGAGCATCATATAAGCGCAGGGATTTGCTCAGACAACATCAAGCTGTAAAA 179
Davis AATATTGTTGGAGCATCATATAAGCGCAGGGATTTGCTCAGACAACATCAAGCTGTAAAA 180
Dem AATATTGTTGGAGCATCATATAAGCGCAGGGATTTGCTCAGACAACATCAAGCTGTAAAA 180
Zunla_gene5 AATATTGTTGGAGCATCATATAAGCGCAGGGATTTGCTCAGACAACATCAAGCTGTAAAA 180
CM334_gene5 AATATTGTTGGAGCATCATATAAGCGCAGGGATTTGCTCAGACAACATCAAGCTGTAAAA 180
CV3_gene5 TTAAAAAAGTTGATCATTTCTGGTGAAGTGCACACTGGACAGGGACTAAATCAAGAACGC 239
ThaiHot_gene5 TTAAAAAAGTTGATCATTTCTGGTGAAGTGCACACTGGACAGGGACTAAATCAAGAACGC 239
SmallFrt_gene5 TTAAAAAAGTTGATCATTTCTGGTGAAGTGCACACTGGACAGGGACTAAATCAAGAACGC 239
Davis TTAAAAAAGTTGATCATTTCTGGTGAAGTGCACACTGGACAGGGACTAAATCAAGAACGC 240
Dem TTAAAAAAGTTGATCATTTCTGGTGAAGTGCACACTGGACAGGGACTAAATCAAGAACGC 240
Zunla_gene5 TTAAAAAAGTTGATCATTTCTGGTGAAGTGCACACTGGACAGGGACTAAATCAAGAACGC 240
CM334_gene5 TTAAAAAAGTTGATCATTTCTGGTGAAGTGCACACTGGACAGGGACTAAATCAAGAACGC 240
CV3_gene5 GGACTTCAACGGCCATGTGACACTCGTTCGGGTTCTCACTGTAAGACGTTAGAGAACTTC 299
ThaiHot_gene5 GGACTTCAACGGCCATGTGACACTCGTTCGGGTTCTCACTGTAAGACGTTAGAGAACTTC 299
SmallFrt_gene5 GGACTTCAACGGCCATGTGACACTCGTTCGGGTTCTCACTGTAAGACGTTAGAGAACTTC 299
Davis GGACTTCAACGGCCATGTGACACTCGTTCGGGTTCTCACTGTAAGACGTTAGAGAACTTC 300
Dem GGACTTCAACGGCCATGTGACACTCGTTCGGGTTCTCACTGTAAGACGTTAGAGAACTTC 300
Zunla_gene5 GGACTTCAACGGCCATGTGACACTCGTTCGGGTTCTCACTGTAAGACGTTAGAGAACTTC 300
CM334_gene5 GGACTTCAACGGCCATGTGACACTCGTTCGGGTTCTCACTGTAAGACGTTAGAGAACTTC 300
CV3_gene5 ATTGACATATTTCTATCAATTCTTTATGTGCTTGAATTTGCTGCACGTGAGTGCCCAAAT 359
ThaiHot_gene5 ATTGACATATTTCTATCAATTCTTTATGTGCTTGAATTTGCTGCACGTGAGTGCCCAAAT 359
SmallFrt_gene5 ATTGACATATTTCTATCAATTCTTTATGTGCTTGAATTTGCTGCACGTGAGTGCCCAAAT 359
Davis ATTGACATATTTCTATCAATTCTTTATGTGCTTGAATTTGCTGCACGTGAGTGCCCAAAT 360
Dem ATTGACATATTTCTATCAATTCTTTATGTGCTTGAATTTGCTGCACGTGAGTGCCCAAAT 360
Zunla_gene5 ATTGACATATTTCTATCAATTCTTTATGTGCTTGAATTTGCTGCACGTGAGTGCCCAAAT 360
CM334_gene5 ATTGACATATTTCTATCAATTCTTTATGTGCTTGAATTTGCTGCACGTGAGTGCCCAAAT 360
CV3_gene5 TATCTTGACAGACTTGCAGCTGAAACTCTTGAGAATACGATTAAGGGGTTTGATTTTGCT 419
ThaiHot_gene5 TATCTTGACAGACTTGCAGCTGAAACTCTTGAGAATACGATTAAGGGGTTTGATTTTGCT 419
SmallFrt_gene5 TATCTTGACAGACTTGCAGCTGAAACTCTTGAGAATACGATTAAGGGGTTTGATTTTGCT 419
Davis TATCTTGACAGACTTGCAGCTGAAACTCTTGAGAATACGATTAAGGGGTTTGATTTTGCT 420
Dem TATCTTGACAGACTTGCAGCTGAAACTCTTGAGAATACGATTAAGGGGTTTGATTTTGCT 420
Zunla_gene5 TATCTTGACAGACTTGCAGCTGAAACTCTTGAGAATACGATTAAGGGGTTTGATTTTGCT 420
CM334_gene5 TATCTTGACAGACTTGCAGCTGAAACTCTTGAGAATACGATTAAGGGGTTTGATTTTGCT 420
CV3_gene5 TTTATGCTACATTTGATGTGGAAAGTTCTAATGATAACAAATTATTTGAACAACTTGTTG 479
ThaiHot_gene5 TTTATGCTACATTTGATGTGGAAAGTTCTAATGATAACAAATTATTTGAACAACTTGTTG 479
SmallFrt_gene5 TTTATGCTACATTTGATGTGGAAAGTTCTAATGATAACAAATTATTTGAACAACTTGTTG 479
Davis TTTATGCTACATTTGATGTGGAAAGTTCTAATGATAACAAATTATTTGAACAACTTGTTG 480

```
Dem             TTTATGCTACATTTGATGTGGAAAGTTCTAATGATAACAAATTATTTGAACAACTTGTTG    480
Zunla_gene5     TTTATGCTACATTTGATGTGGAAAGTTCTAATGATAACAAATTATTTGAACAACTTGTTG    480
CM334_gene5     TTTATGCTACATTTGATGTGGAAAGTTCTAATGATAACAAATTATTTGAACAACTTGTTG    480
                ************************************************************

CV3_gene5       CAAAAGATGGATCAAGATATTGTCAATGCTCTGAAACTGCTCAATACTGCAAAGCAAGAG    539
ThaiHot_gene5   CAAAAGATGGATCAAGATATTGTCAATGCTCTGAAACTGCTCAATACTGCAAAGCAAGAG    539
SmallFrt_gene5  CAAAAGATGGATCAAGATATTGTCAATGCTCTGAAACTGCTCAATACTGCAAAGCAAGAG    539
Davis           CAAAAGATGGATCAAGATATTGTCAATGCTCTGAAACTGCTCAATACTGCAAAGCAAGAG    540
Dem             CAAAAGATGGATCAAGATATTGTCAATGCTCTGAAACTGCTCAATACTGCAAAGCAAGAG    540
Zunla_gene5     CAAAAGATGGATCAAGATATTGTCAATGCTCTGAAACTGCTCAATACTGCAAAGCAAGAG    540
CM334_gene5     CAAAAGATGGATCAAGATATTGTCAATGCTCTGAAACTGCTCAATACTGCAAAGCAAGAG    540
                ************************************************************

CV3_gene5       TTGCAAAGGATGAGGGATAGTGGATGGAGATCATTACTGGATAATGTCTTTTCTTTTTGT    559
ThaiHot_gene5   TTGCAAAGGATGAGGGATAGTGGATGGAGATCATTACTGGATAATGTCTTTTCTTTTTGT    559
SmallFrt_gene5  TTGCAAAGGATGAGGGATAGTGGATGGAGATCATTACTGGATAATGTCTTTTCTTTTTGT    559
Davis           TTGCAAAGGATGAGGGATAGTGGATGGAGATCATTACTGGATAATGTCTTTTCTTTTTGT    600
Dem             TTGCAAAGGATGAGGGATAGTGGATGGAGATCATTACTGGATAATGTCTTTTCTTTTTGT    600
Zunla_gene5     TTGCAAAGGATGAGGGATAGTGGATGGAGATCATTACTGGATAATGTCTTTTCTTTTTGT    600
CM334_gene5     TTGCAAAGGATGAGGGATAGTGGATGGAGATCATTACTGGATAATGTCTTTTCTTTTTGT    600
                ************************************************************

CV3_gene5       GATAAATATGAGATAGCGATCCCAAAAATGGATGCTCGCTACATTCCTGTCATGAACACG    659
ThaiHot_gene5   GATAAATATGAGATAGCGATCCCAAAAATGGATGCTCGCTACATTCCTGTCATGAACACG    659
SmallFrt_gene5  GATAAATATGAGATAGCGATCCCAAAAATGGATGCTCGCTACATTCCTGTCATGAACACG    659
Davis           GATAAATATGAGATAGCGATCCCAAAAATGGATGCTCGCTACATTCCTGTCATGAACACG    660
Dem             GATAAATATGAGATAGCGATCCCAAAAATGGATGCTCGCTACATTCCTGTCATGAACACG    660
Zunla_gene5     GATAAATATGAGATAGCGATCCCAAAAATGGATGCTCGCTACATTCCTGTCATGAACACG    660
CM334_gene5     GATAAATATGAGATAGCGATCCCAAAAATGGATGCTCGCTACATTCCTGTCATGAACACG    660
                ************************************************************

CV3_gene5       TCCCGATTTCATGAGAGCTTAGAGAATAGGCCTTTACTAACATTCTCCTTGAAGCGGCTT    719
ThaiHot_gene5   TCCCGATTTCATGAGAGCTTAGAGAATAGGCCTTTACTAACATTCTCCTTGAAGCGGCTT    719
SmallFrt_gene5  TCCCGATTTCATGAGAGCTTAGAGAATAGGCCTTTACTAACATTCTCCTTGAAGCGGCTT    719
Davis           TCCCGATTTCATGAGAGCTTAGAGAATAGGCCTTTACTAACATTCTCCTTGAAGCGGCTT    720
Dem             TCCCGATTTCATGAGAGCTTAGAGAATAGGCCTTTACTAACATTCTCCTTGAAGCGGCTT    720
Zunla_gene5     TCCCGATTTCATGAGAGCTTAGAGAATAGGCCTTTACTAACATTCTCCTTGAAGCGTCTT    720
CM334_gene5     TCCCGATTTCATGAGAGCTTAGAGAATAGGCCTTTACTAACATTCTCCTTGAAGCGGCTT    720
                ************************************************************ ***

CV3_gene5       CCACTTTGCCCTCACATAGGTGATTTCTAA  749
ThaiHot_gene5   CCACTTTGCCCTCACATAGGTGATTTCTAA  749
SmallFrt_gene5  CCACTTTGCCCTCACATAGGTGATTTCTAA  749
Davis           CCACTTTGCCCTCACATAGGTGATTTCTAA  750
Dem             CCACTTTGCCCTCACATAGGTGATTTCTAA  750
Zunla_gene5     CCACTTTGTCCTCACATAGGTGATTTCTAA  750
CM334_gene5     CCACTTTGTCCTCACATAGGTGATTTCTAA  750
                ******** **********************
```

Figure 13 Alignment of coding sequence (CDS) for the pendant and erect types. *Arrow* indicate a single-base adenine deletion of ZF-MYM mRNA. The SNP position shown in shade. The consensus sequences are labeled "⋆" below the alignment.

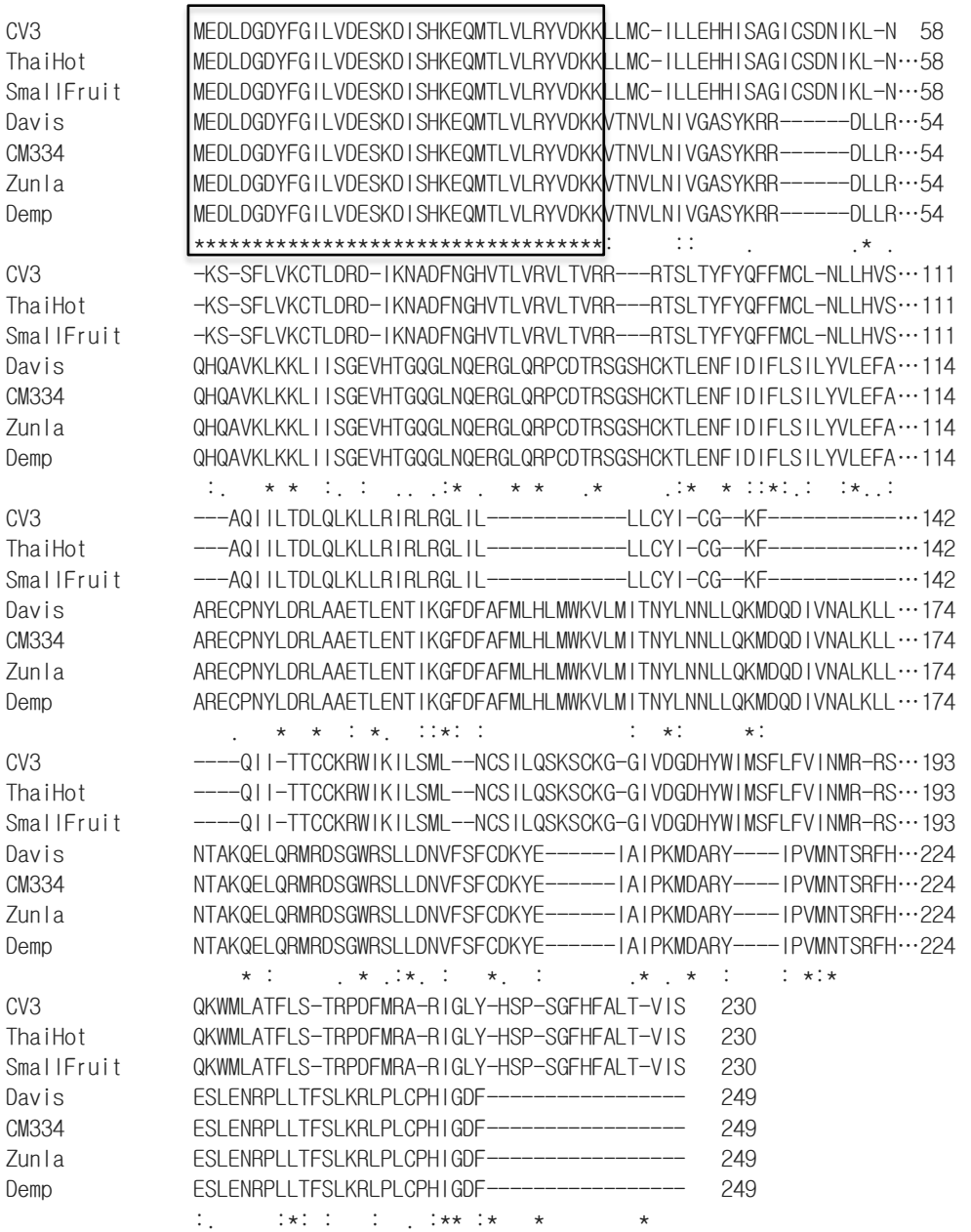


Figure 14 Sequence alignment of amino acids for the erect (CV3, SmallFrt and ThaiHot) and pendant (Dempsey, Zunla, CM334 and UC Davis) types, where frameshift mutation resulted after the first exon (boxed). The consensus sequences are labeled "*", conservative sequences ":" and similar types labeled "." below the alignment.

Phylogenic analysis

In order to understand the relationship between ZF-MYM transcription factor and its close homologs, we searched public databases NCBI using BLAST with the ZF-MYM amino acid sequence and used an alignment of the closest homologs from six plant species in *Solanaceae* family (Fig 15). The resulting neighbor-joining tree showed that *Capsicum annuum* ZF-MYM grouped together with homologs from *Nicotiana sylvestris*. We also found that the site where the mutation occurred was similar among the species compared. These presented data suggested that ZF-MYM is the most likely candidate gene for *Up*, and the non-synonymous SNP and nucleotide deletions existed in the first exon are responsible for the erect phenotype.

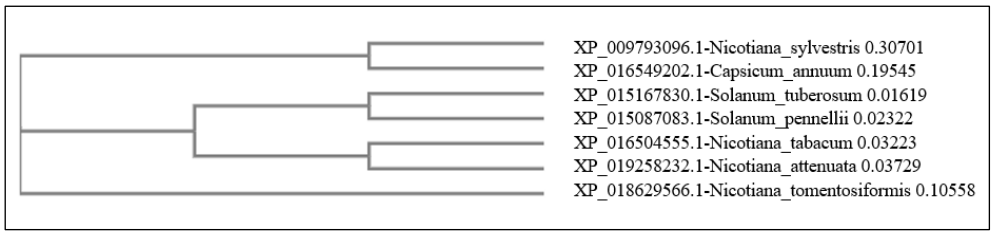


Figure 15 Phylogenetic analysis of ZF-MYM and its homologs in six other plant species in *Solanaceae* family. Evolutionary relationships were inferred using neighbor-joining method without distance corrections. The ID numbers refer to the gene IDs in the NCBI database and the names of the species are provided.

DISCUSSION

This study was designed to fine map the fruit orientation controlling gene of *Capsicum* and to identify responsible candidate gene(s), by using segregating populations and diverse accessions, employing the combination of QTL and GWAS. While the first provides two, three and ten major QTLs for PD, FC and MJ populations, respectively, the later identified 14 highly significant SNPs associated with fruit orientation using 256 diverse accessions. We combined these results and delimited the mapping region.

In most cases pepper pedicel curvature determines fruit orientation (Lin et al., 2016; Sun et al., 2019); and the relative position at which the bending occur on the pedicel can lead to have a vertical pendant phenotype, which is easy to distinguish even at early flowering stages and is a typical character for certain populations; or lateral pendant phenotype with a near-horizontally oriented fruit tip (Cheng et al., 2016). In some populations, we observed the difficulty of determining whether the fruit orientation is upright or pendant during flowering and early stages of fruiting as the phenotypes can be shifted from one state to the other through time. Similar observation on *C. annuum* was also reported by Munting (1974).

The inheritance of fruit orientation trait in *Capsicum* was investigated by crossing pendant and upright fruited parental lines of different genetic architectures. Categorizing the two intermediate types of fruit orientation into

the vertical types depending on the majority rule, our results suggests that the pendant fruit orientation in pepper is dominant over erect types. Similarly, Cheng et al. (2016) reported as erect phenotype is controlled by recessive gene with incomplete inheritance. Another experiment on cucumber confirmed that fruit orientation follows Mendelian segregation that *Up* was regulated by a single recessive locus (Sun et al., 2019).

Understanding the extent and type of relationships between pepper fruit orientation and some of fruit related traits is very important since change in one character can influence fully or partly the other one. Our result of positive correlation of fruit weight and pedicel length that contributed to the tendency of pendant growth agrees with the reports of Han et al. (2016). Since the fruits are set at the tip of non-wooden pedicel, it is theoretically logical to bend downward due to the gravitropism as the length of the fruit bearing pedicel get increased. Interestingly, we also have observed a negative correlation between pedicel thickness and length; and pendant growth and pedicel thickness. The negative correlation between length and thickness of pedicel can be explained by the invariable tissue bulk density of pedicels in the two types, which can also be substantiated by the observation we made on the reduced size of the collenchyma cells of abaxial region of pendant types; and this might be the reason why the opposite relation was evidenced between pedicel thickness and pendant fruit growth.

In order to understand the trend of fruit orientation in different pepper

populations, identify the possible minor alleles that contribute for the few intermediate phenotypes and determine the important locus that controls the upright and pendant fruit orientation, we performed QTL and GWAS. Linkage mapping, based on the analysis of the segregation of polymorphism between the parental lines and their progeny, has proved invaluable for detecting quantitative trait loci (QTLs) for traits of interest and to unravel their underlying genetic architecture (Sauvage et al., 2014; Jamann et al., 2015), while linkage disequilibrium (LD) mapping, also known as genome-wide association study (GWAS) is a useful tool to detect candidate loci responsible for the natural variations that can identify significant associations between polymorphic molecular markers and targeted traits in a large natural populations (Zhang et al., 2015).

Using composite interval mapping approach and involving three biparental lines, our mapping region was delimited to 208-214 Mbps in chromosome 12. This result is similar with previous reports. The first molecular study to develop AFLP and CAPS markers for the pendant orientation of *Capsicum annuum* by Lee et al. (2008) using 108 $F_{2:3}$ individuals localized the gene in chromosome 12. In 2016, a report by Han and colleagues identified two consistent QTL regions in the same chromosome (FP-12.1 and FP-12.2) for fruit positon in PDRIL in more than two different environments (Han et al., 2016). Of the two QTLs, they reported FP-12.2 as the possible loci for the existence of major gene explaining over

40% of phenotypic variation located at 199.6 Mbp using CM334 v.1.55 genome reference. Same year, in another study, one major QTL named *Up12.1* was detected in the same region using 297 F_2 lines obtained by interspecific crossing between BA3 (*Capsicum annuum*) and YNXML (*Capsicum frutescens*) (Cheng et al., 2016).

The pepper core collection (Lee et al., 2016) was used for genome wide association study. Super GWAS was used to identify 14 highly significant SNPs that were associated with the fruit orientation between the physical position 205 and 214 Mbps in chromosome 12 with –log10(p) value > 26 (Wang et al., 2014). By combining these results and previously published reports (Lee et al., 2008; Han et al., 2016; Cheng et al., 2016), we identified target region between 212 to 214 Mbp for further study. The combined use of QTL and GWAS was shown powerful approach for identification of loci in pepper (Han et al., 2018). Other studies have also shown the relevance of the combined use of QTL and GWAS approach for the effective identification of responsible locus and candidate genes in maize, pepper, brassica, cotton and cucumber (Zhao et al., 2018; Siddique et al., 2019; He et al., 2017; Liu et al., 2018; Bo et al., 2019).

After evaluation of different high resolution melting (HRM) markers, we were able to delimit the genomic region containing the *Up* locus to 101 kb of sequences on chromosome 12 based on two flanking HRM markers, Redu0119 and SAR201-1389, which are 0.8 and 0.6 cM distant from the gene,

respectively. The *Up* locus identified in this study is 1.5 Mb distant from previously reported *Up12.1* locus which was mapped in 4.52 Mb region by Cheng et al. (2016), and 1.3 Mb far from *FP-12.2* reported in another study (Han et al., 2016). Although the high sensitivity of the method hinders to easily develop reliable markers, use of HRM for fine mapping was reported previously (Jeong et al., 2018; Keller et al., 2015).

Using *Capsicum annuum* cv. CM334 reference genome ver. 1.6 (http://peppergenome.snu.ac.kr/), we identified a total of 15 hypothetical genes, 10 on the +ve and five on the –ve strands. The function of six of these candidate genes are not known, while others were described. BLAST search in GenBank using the National Center for Biotechnology Information (NCBI) database suggested that eight candidate genes have homology in chromosome 12 of tomato, either *Solanum pennellii* or *S. lycopersicum.* For those proteins whose functions are known, we identified 10 non-synonymous and nine synonymous SNPs, all resulted in a conservative amino acid substitution as described previously (Grantham 1974; Sneath 1966; Epstein 1967; Miyata et al., 1979). In ZF-MYM transcription factor, however, one nucleotide deletion followed by G → A SNP was detected at the end of the first exon. This InDel has caused a frameshift mutation and a truncated protein at the 3' end, which might affect the DNA binding domains of the gene. Takatsuji (1999) described the key roles of different Zinc fingers, SUPERMAN, $AtZFP_1$, $PetSPL_3$ and $BcZFP_1$ in the developmental regulation of various floral and

vegetative organs. Thus, ZF-MYM transcription factor can be a strong candidate gene underlying the erect phenotype in pepper.

REFERENCES

Albrecht E, Zhang D, Mays AD, Saftner RA, et al (2012). Genetic diversity in *Capsicum baccatum* is significantly influenced by its ecogeographical distribution. BMC Genet 13(68): 1-15

Bai Y, Lindhout P (2007) Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? Ann Bot 100: 1085-1094

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263-265

Bo K, Wei S, Wang W, Miao H, et al (2019) QTL mapping and genome-wide association study reveal two novel loci associated with green flesh color in cucumber. BMC Plant Bio 19(243)

Bundy MGR, Thompson OA, Sieger MT, Shpak ED (2012) Patterns of cell division, cell differentiation and cell elongation in epidermis and cortex of *Arabidopsis* pedicels in the wild type and in *erecta*. PLoS One 7(9): e46262. https://doi.org/10.1371/journal.pone.0046262

Cheng J, Win C, Tang X, Zhou H, et al (2016) Development of a SNP array and its application to genetic mapping and diversity assessment in pepper (*Capsicum* spp.). Sci Rep 6: 1-11

Chunthawodtiporn J, Hill T, Soffel K, Deynze AV (2018) Quantitative trait

loci controlling fruit size and other horticultural traits in bell pepper (*Capsicum annuum*). Plant Genome 11(1): 1-11

Clark S (2003) Plant development: A long way from its roots; Book review. Cell 113: 11-19

Douglas SJ, Chuck G, Dengler RE, Pelecanda L, et al (2002) KNAT1 and ERECTA regulate inflorescence architecture in Arabidopsis. Plant Cell 14(3): 547-558

Douglas SJ, Riggs CD (2005) Pedicel development in *Arabidopsis thaliana*: contribution of vascular positioning and the role of the brevipedicellus and erecta genes. Dev Biol 284: 451-463

Epstein CJ (1967) Non-randomness of ammo-acid changes in the evolution of homologous proteins. Nature 215: 355-359

Gabriel SB, Schaffiner SF, Nguyen H, Moore JM, et al (2002) The structure of haplotype blocks in the human genome. Sci 21(296)5576: 2225-2229

Grantham R (1974) Amino acid difference formula to help explain protein evolution. Science 185 (4154): 862–864

Kaiser S (1935) The factors governing shape and size in *Capsicum* fruits; a genetic and developmental analysis. Torrey Bot Club Bul 62:433-454

Kim S, Park M, Yeom SI, Kim YM, et al (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet 46:270-278

Han K, Jeong HJ, Yang HB, Kang SM, et al (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, et al (2018) QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. Plant Biotechnol J 16: 1546-1558

He Y, Wu D, Wei D, Fu Y, et al (2017) GWAS, QTL mapping and gene expression analyses in *Brassica napus* reveal genetic control of branching morphogenesis. Sci Rep 7(15971)

Jamann TM, Kurti JB, Holland JB (2015) QTL mapping using high-throughput sequencing. In: Alonso J., Stepanova A. (eds) Plant Functional Genomics. Methods in Molecular Biology, Vol 1284. Humana Press, New York, NY

Jeong HJ (2014) Development of pre-breeding materials in *Capsicum* spp. and map based cloning of a male sterile gene, *ms10*^{35} in tomato. Desertation. SNU

Jeong K, Choi D, Lee J (2018) Fine mapping of the genic male-sterile *ms1* gene in *Capsicum annuum* L. Theor Appl Genet 131: 183-191

Keller B, Manzanares C, Jara C, Lobaton JD, et al (2015) Fine-mapping of a major QTL controlling angular leaf spot resistance in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 128: 813-826

Lee HR, Cho MC, Kim HJ, Park SW, et al (2008) Marker development for

erect versus pendant-oriented fruit in *Capsicum annuum* L. Mol Cells 26: 548-553

Lee HY, Ro NY, Jeong HJ, Kwon JK, et al (2016) Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. BMC Genet 17(1): 142

Li H (2013) Aligning sequcene reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint* arXiv: 1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, et al (2009) The sequence alignment/ map (SAM) format and SAMtools. Bioinformatics 25: 2078-2079

Lin D, Xiang Y, Xian Z, Li Z (2016) Ectopic expression of SIAGO7 alters leaf pattern and inflorescence architecture and increases fruit in tomato. Physio Planta 157(4): 490-506

Liu R, Gong J, Xiao X, Zhang Z, et al (2018) GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high density SNP markers. Front Plant Sci 9(1067)

Leyser O (2016) Plant development: a special issue. Dev 143: 3223

Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitutions in protein evolution. J. Mol. 12: 219-236

Munting AJ (1974) Development of flower and fruit of *Capsicum annuum* L. Acta Bot Neerl 23(4): 415-432

Niu Y, Zhuo Z, Sha W, Sun H (2016) Post-floral Erection of Stalks Provides Insight into the Evolution of Fruit Orientation and Its Effects on Seed Dispersal. Sci Rep 6: 1-10

Paran I, der Knaap E (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. J Exp Bot 58(14): 3841-3852

Prunet N, Meyerowitz EM (2016) Genetics and plant development. CR Biologies 339: 240-246

Qin C, Yu C, Shen Y, Fang X, et al (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci USA 111:5135-5140

Ragni L, Belles-Boix E, Guni M, Pautot V (2008) Interaction of *KNAT6* and *KNAT2* with *BREVIPEDICELLUS* and *PENNYWISE* in *Arabidopsis* inflorescences. Plant Cell 20: 888-900

Ramchiary N, Kehie M, Brahma V, Kumaria S, et al (2013) Application of genetics and genomics towards *Capsicum* translational research. Plant Biotechnol Rep 8(2): 101-123

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol 132: 365-386

Sauvage C, Segura V, Bauchet G, Stevens R, et al (2014) Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits.

Plant Physiol 165(3): 1120-1132

Setiamihardja R, Knavei E (1990) Association of pedicel length and diameter with fruit length and diameter and ease for fruit detachment in pepper. J Am Soc Hortic Sci 115(4): 677-681

Schiex T, Gaspin C (1997) Carthagene: constructing and joining maximum likelihood genetic maps. Proceeding. AAAI 258-267

Siddique MI, Lee HY, Ro NY, Han K, et al (2019) Identifying candidate genes for *Phytophthora capsici* resistance in pepper (*Capsicum annuum*) via genotyping-by-sequencing based QTL mapping and genome wide association study. Sci Rep 9: 1-15

Sneath PH (1966) Relations between chemical structure and biological activity in peptides. J. Theor.Biol. 12(2): 157-195

Sun J, Xiao T, Nie J, Chen Y, et al (2019) Mapping and identification of *CsUp*, a gene encoding an Auxilin-like protein, as a putative candidate gene for the upward pedicel mutation (up) in cucumber. BMC Plant Biol 19(157): 1-12

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93(1): 77-78

Wang Q, Tian F, Pan Y, Buckler ES, et al (2014) A SUPER powerful method for genome wide association study. PLoS One 9(9): e107684. https://doi.org/10.1371/journal.pone.0107684

Wang D, Chen X, Zhang Z, Liu D, et al (2015) A MADS-box gene *NtSVP*

regulates pedicel elongation by directly suppressing a KNAT1-like KNOX gene *NtBPL* in tobacco (*Nicotiana tabacum* L.). J Exp Bot 66(20): 6233-6244

Yamaguchi N, Yamaguchi A, Abe M, Wagner D, et al (2012) Leafy controls Arabidopsis pedicel length and orientation by affecting adaxial-abaxial cell fate. Plant J 69(5): 844-856

Zhao X, Luo L, Cao Y, Liu Y, et al (2018) Genome-wide association analysis and QTL mapping reveal the genetic control of cadmium accumulation in maize leaf. BMC Genomic 19(91): 1-13

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136(4): 1457-1468

Zhang J, Zhao J, Xu Y, Liang J, et al (2015) Genome wide association mapping for tomato volatiles positively contributing to tomato flavor. Fornt Plant Sci 6(1042): 1-13

ABSTRACT IN KOREAN

에티오피아 유전자 은행에 보존된 고추 유전자원의 다양성 연구와 고추 과실의 기원을 조절하는 유전적 요인들을 이해하는 것은 육종에 중요한 정보를 제공할 수 있다.

작물 유전자원의 다양성의 구조와 깊이를 연구하는 것은 이용 가능 한 유전적 자원의 실용화와 보존에 매우 필수적이다. 에티오피아에서 고추는 매우 경제적이고 사회적 중요성을 가진 작물임에도 불구하고, 단일염기다형성(SNP) 마커를 이용한 유전적 다양성에 관련한 자세한 연구는 제한적이다. 따라서, 에티오피아 고추 유전자원의 다양성과 유전적 구조를 조사하기 위해, 에티오피아 생물다양성 연구소에서 수집하고 유지된 총 142 종에 대하여 평가하였다. 우리는 53,284 genome wide SNP 분자마커를 genotyping-by-sequencing (GBS)를 이용하여 발견하여 검증하였다. 우리는 모델 기반의 집단 구조, 계통도, 주성분 분석을 사용하여 *C. annuum* 과

*C. frutescens*를 각각 132종과 9종의 두 개의 뚜렷한 집단으로 확인하였다. 이 밖에도, 전장유전체 연관 분석 (GWAS) 결과로 과실, 줄기, 잎 관련 특성과 크게 연관된 509개의 SNP 마커를 발견하였다. 전반적으로, 이 연구는 보존과 육종이라는 관점에서 에티오피아 고추 종에 존재하는 유전적 다양성을 이해하는데 유용할 것이다.

고추의 과실방향에서 두개의 주요한 시장으로, pendant 와 upright 가 있다. 과실의 Pendant 와 upright 방향 그리고 드물게 발생하는 중간형은 집단의 구조에 따라 과실이 달리기 전과 후에 쉽게 구별되어지는 형태학적 가변성을 판별하는 것이다. 고추 핵심 집단 (core collection)에서 GWAS 분석을 통한 highly significant SNP 와 3 개의 양친형 집단 (bi-parental population)으로부터 유전자를 발견하는 것은 염색체 7 번에 존재하는 과실의 방향성을 조절하는 유전자좌의 맵핑 구역을 좁혀가는 것을 가능하게 한다. 추가로 분자 마커를 개발하고, 과실 방향에 기초하는 후보 유전자를 분리하기 위해, *Capsicum annuum* LA F2 mapping population 을 사용하여 fine mapping 을 수행하였다. 450 individuals 로 구성된 LA F2 mapping population 은 양친형 라인인 *C. annuum* LP97 and A79 의 교배를 통해 개발되었다. 유전자 분석 결과 상향 과실 방향의 포현형은 단일 열성 유전자 Up 에 의하여 조절됨을 나타냈다. Up 유전자좌의 fine mapping 을 위해 150 개의 SNP 분자마커가 사용되었다. 유전자 분석 결과 상향 과실 방향의 포현형은 단일 열성 유전자 Up 에 의하여 조절됨을 나타냈다. Up 유전자좌의 fine mapping 을 위해 150 개의 SNP 분자마커가 사용되었다. 이 두개의

마커 서열은 CM334 유전체에 정렬되었으며, Up 유전자좌는 101kb genomic region 으로 구분되었다. 대상 지역에서 13 개의 후보유전자가 예측되었다. 후보유전자의 CDS 영역인 Zinc finger MYM-type protein 1-like 는 66 과 75 위치의 두 개의 비 동의 염기치환과, 첫번째 뉴클레오타이드의 104bp 에서 104b 에서 한 개의 염기 결실을 나타내었다. 이러한 돌연변이는 12 개의 parental line 과 5 개의 종에서 일관되게 관찰되었다. 따라서 이 유전자는 Up 유전자좌에 매우 유력한 후보유전자였다. 이 연구를 통해 얻은 정보는 분자표지 기반 선발을 통하여 조기 과실 방향 결정으로 고추 품종의 특성과 육종에 대한 추가 연구를 촉진할 것이다.