A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Identification of Genetic Factors Controlling Capsaicinoid Content in Pepper (*Capsicum* spp.)

고추의 캡사이시노이드 함량을 조절하는 유전인자 탐색

FEBRUARY, 2018

KOEUN HAN

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

Identification of Genetic Factors Controlling Capsaicinoid Content 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
KOEUN HAN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE

FEBRUARY, 2018

APPROVED AS A QUALIFIED DISSERTATION OF KOEUN HAN
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN	**Doil Choi, Ph.D.**
VICE-CHAIRMAN	**Byoung-Cheorl Kang, Ph.D.**
MEMBER	**Tea-Jin Yang, Ph.D.**
MEMBER	**Jin-Ho Kang, Ph.D.**
MEMBER	**Sung-Chur Sim, Ph.D.**

Identification of Genetic Factors Controlling Capsaicinoid Content in Pepper (*Capsicum* spp.)

KOEUN HAN

Department of Plant Science, Seoul National University

ABSTRACTS

Peppers (*Capsicum* spp.) synthesize a pungent metabolite, capsaicinoid. Capsaicinoid biosynthesis pathway has been predicted based on primary and secondary metabolite synthesis pathway of other plants like *Arabidopsis*. From gene expression studies, multiple genes showed correlation with capsaicinoid content and designated as candidate genes controlling pungency. However, effects of the capsaicinoid biosynthetic genes for capsaicinoid content still unclear. In this study, we conducted two experiments to figure out the quantitative and qualitative genetic factors

controlling the level of capsaicinoid in *Capsicum*.

To find the quantitative trait loci (QTL) controlling capsaicinoid content, two recombinant inbred lines (RILs) were used. Capsaicinoid content was evaluated from placental tissue, where capsaicinoid mostly synthesized and accumulated. Furthermore, high-density genetic map using single-nucleotide polymorphism (SNPs) from next-generation sequencing (NGS) based genotyping methods. Based on the *C. annuum* 'CM334' reference genome, we found five common QTL located on chromosome 1, 2, 3, 4, and 10 which were detected from both RIL populations. To identify the capsaicinoid biosynthetic genes located on QTL, physical location of QTL were compared with the genome-wide association study (GWAS) conducted from *Capsicum* core collection. A total of ten regions were commonly detected, and five genes involved in capsaicinoid biosynthesis pathway were proposed to regulate pungency level.

Since non-pungent accessions has non-functional capsaicinoid biosynthetic gene, they were widely used to identify the genes. RIL and F_2 population derived from a cross between non-pungent accession *C. annuum* 'YCM334' and pungent accession *C. annuum* 'Tean' were used to find a novel locus controlling presence of capsaicinoid. High-density genetic map was constructed using genotyping-by-sequencing (GBS) method, and the

novel locus designated as *Pun3* was genetically mapped on chromosome 7.

To find the candidate genes on the *Pun3* locus, High-resolution melting (HRM) markers were designed and the locus was mapped between 192.2-193.1 Mbp and 199.3-200.1 Mbp on ‘Zunla-1’ and ‘PI159236’ reference genome, respectively. Additionally RNA-Seq of parental lines was performed in three developmental stages, and expressions of the genes in fatty-acid biosynthetic pathway were extremely reduced in non-pungent parent ‘YCM334’. Three genes including Capana07g001603, Capana07g001604, and Capana07g001614 showed significant difference in expression level. They were only expressed in pungent parent ‘Tean’, and all three genes were predicted to encode MYB transcription factors. One of the genes, Capana07g001604, was reported to be related with capsaicinoid level in previous research. Therefore, we proposed that one of the predicted MYB transcription factors will be the candidate gene for *Pun3*.

Taken together, we identified capsaicinoid biosynthetic genes controlling capsaicinoid content and a transcription factor regulates presence of pungency. Although the exact function of genes should be validated in multiple accessions, this result could be helpful to understand the capsaicinoid biosynthetic pathway as well as understanding the genetic control and breeding of pungent pepper.

Keywords: capsaicin, dihydrocapsaicin, pepper, pungency, quantitative trait locus (QTL), genotyping-by-sequencing (GBS)

Student number: 2014-30376

CONTENTS

LIST OF TABLES

Chapter I

Chapter II

LIST OF FIGURES

Chapter I

Chapter II

GENERAL INTRODUCTION

Pepper (*Capsicum* spp.) is one of the first domesticated vegetable crops (Paran and van der Knaap, 2007), which is grown for fresh fruit, spices, and sauces worldwide. Capsaicinoid having pungent taste is only synthesized in *Capsicum* species. Researchers tried to figure out the mechanism how mammals recognize pungency and the function in pepper and human (Caterina et al., 1997). Additionally, multiple genetic studies to figure out the heritability of pungency in pepper and to find the genes controlling pungency have been carried out.

Capsaicinoid protects pepper from mammals that can harm seeds and fruits (Tewksbury and Nabhan, 2001), and reduce the *Fusarium* growth inside of the fruit (Tewksbury et al., 2008b). In the habitat of wild-type pepper *C. chacoense*, the ratio of pungent and non-pungent plants was different by environmental conditions including altitude and humidity (Haak et al., 2012; Tewksbury et al., 2008a). Capsaicinoid not only used for pungency, but also has multiple pharmaceutical functions for human. Capsaicinoid has anticancer activities for skin, lung, pancreatic, and prostatic cancer cells (Clark and Lee, 2016). It also has pain relief function for rheumatoid arthritis (Fraenkel et al., 2004), and can be used to treat neropathic pain (Peppin and Pappagallo, 2014).

Capsaicinoid is important metabolite for food marker and also field of pharmacy. Therefore there have been large efforts to increase capsaicinoid content

of pepper. In 2007, *C. chinense* 'Bhut Jolokia' was the hottest pepper over the world with one million Scoville heat units (SHU) and other *C. chinense* accessions were also showed higher than 500,000 SHU (Bosland and Baral, 2007; Canto-Flick et al., 2008). Furthermore, another *C. chinense* 'Trinidad Moruga Scorpion' was found to have two million SHU, which is twice hotter than 'Bhut Jolokia' (Bosland et al., 2012). These extremely hot peppers were identified to accumulate capsaicinoid not only at the placental tissue where most of pepper contain capsaicinoid, but also at the pericarp tissue (Bosland and Coon, 2015). Recently, pepper breeders make more pungent varieties by artificial crosses and selections, including Carolina Reaper, Dragon's Breath Chili Pepper, and Pepper X which exceed 2-3 million SHU.

However not much is studied about the genetic factors affecting pungency of pepper. Non-functional alleles of a few genes were detected, and used for genetic study of the gene. The *Pun1* gene is most well-known capsaicinoid biosynthetic gene, which encodes capsaicin synthase in the final step of capsaicinoid biosynthesis pathway. *Pun1* is responsible for presence of pungency in pepper (Stewart et al., 2005), and three non-functional alleles including *pun1*, *pun1*^{2}, and *pun1*^{3} have been identified (Stellari et al., 2010; Stewart et al., 2005; Stewart et al., 2007). More than seven non-functional alleles of *putative aminotransferase* (*pAMT*) were detected from *C. chinense* accessions (Tanaka et al., 2015). All of these accessions synthesized capsinoid, a low-pungent capsaicinoid analogue, instead of capsaicinoid. In *pamt* mutant plants, an intermediate of phenylpropanoid pathway

name vanilline is catalyzed to vanillylalcohol instead of vanillylamine (Lang et al., 2009). Except those two genes, exact functions of the genes in the capsaicinoid biosynthetic pathway were not confirmed, due to the lack of genetic variation of non-pungent accessions. The other locus controls presence of pungency was designated *Pun2*, and identified from wild pepper *C. chacoense* (Stellari et al., 2010). This locus was confirmed to be different with *Pun1*, by allelism test with *pun1*, *pun1*^{2}, and *pun1*^{3} non-pungent accessions. However specific locus was not mapped as marker density was too low.

Candidate genes controlling pungency was predicted based on the capsaicinoid biosynthesis pathway and correlated expression level with capsaicinoid content. From gene expression analyses of different development stage fruits, *phenylalanine ammonia lyase* (*Pal*), *3-keto-acyl-ACP synthase* (*Kas*), *caffeic acid O-methyltransferase* (*Comt*), and *thioesterase* (*Fat*) showed correlation with expression of *Pun1*, and also capsaicinoid content (Aluru et al., 2003; Curry et al., 1999; Kim et al., 2001). In developing fruits, capsaicinoid start to accumulate around 20 day post anthesis (dpa), accumulate most around 30-40 dpa when the fruits are breaker stage, and start to be degraded becoming mature (Han et al., 2013; Jang et al., 2015; Zhang et al., 2016). Capsaicinoid content of pericarp tissue also showed similar accumulation pattern with placental tissue, and capsaicinoid biosynthetic genes were expressed in the pericarp (Tanaka et al., 2017). This research shows that even though the gene determines the accumulation of capsaicinoid in the pericarp tissue in unclear, but capsaicinoid biosynthesis

pathway works same way in placental and pericarp tissues.

Few years ago, *C. annuum* reference genomes were published (Kim et al., 2014; Qin et al., 2014) and recently two more *Capsicum* genomes were also reported (Kim et al., 2017). With the development of next-generation sequencing (NGS) method and assembled pepper reference genome, candidate genes that might be related to capsaicinoid biosynthesis have been purposed using RNA-Seq (Zhang et al., 2016), genome-wide association study (GWAS) (Nimmakayala et al., 2016), or meta-analysis of previous quantitative-trait locus (QTL) mapping results (Hill et al., 2017). Zhang and colleagues performed RNA-Seq from the placental tissue of highly-pungent *C. frutescens* 'Guijiangwang' by dividing five developmental stages (Zhang et al., 2016). By digital gene expression analysis, 135 genes with predicted function were identified and most of them were related to capsaicinoid biosynthesis. Also 20 of them were newly identified as candidate genes for pungency.

In 94 *C. annuum* accessions, GWAS was performed to identify candidate genes controlling capsaicinoid content (Nimmakayala et al., 2016). A total of 7,331 SNPs were used and 72 were found to be associated with capsaicinoid content. One of the linked genes was Ankyrin-like protein which has acyltransferase function similar like the protein encoded by *Pun1*. Therefore this research group proposed the gene as possible candidate for capsaicinoid content. In addition to GWAS, QTL analysis is popular method to identify the loci controlling quantitative traits. QTL controlling capsaicinoid content have been identified from several interspecific

populations (Ben-Chaim et al., 2006; Blum et al., 2003; Lee et al., 2016; Yarnes et al., 2013). In 2003, a bulked segregant analysis was used to identify the QTL from low- and high-pungency plants selected from an F_2 population derived from a cross between non-pungent *C. annuum* 'Maor' and pungent *C. frutescens* 'BG 2816' (Blum et al., 2003). These plants were screened with 400 restriction fragment length polymorphism markers, which resulted in the detection of a single pungency QTL, cap, on chromosome 7. Another population, derived from a cross between the mildly pungent *C. annuum* 'NuMex RNaky' and highly pungent *C. frutescens* 'BG2814-6' cultivars, was used to identify five QTL for capsaicin, four for dihydrocapsaicin, one for nordihydrocapsaicin, and five for total capsaicinoid content (Ben-Chaim et al., 2006). Yarnes et al. (2013) assessed recombinant inbred lines (RILs), derived from the same parental lines as Ben-Chaim et al. (2006), using a Pepper Genechip array of 16,188 unigenes, resulting in the detection of QTL on chromosomes 3, 4, 5, 6, 7, 10, and 11. Recently, four capsaicinoid QTL were also detected from an F_2 population derived from a cross between *C. annuum* 'NB1' and *C. chinense* 'Bhut Jolokia' (Lee et al., 2016). Direct comparisons between the QTL from different studies are not possible due to the limited numbers of common markers between populations and the low density of genetic maps. Although capsaicinoid biosynthesis genes may be located at these QTL, no likely candidate genes underlying the QTL have been proposed in these previous studies.

This study focused on identification of genetic factors controlling quantitative and qualitative level of capsaicinoid content in *Capsicum*.

REFERENCES

Aluru, M.R., Mazourek, M., Landry, L.G., Curry, J., Jahn, M. and O'Connell, M.A. (2003) Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in *Capsicum* fruit. *J. Exp. Bot.* **54**, 1655-1664.

Ben-Chaim, A., Borovsky, Y., Falise, M., Mazourek, M., Kang, B.C., Paran, I. and Jahn, M. (2006) QTL analysis for capsaicinoid content in *Capsicum*. *Theor. Appl. Genet.* **113**, 1481-1490.

Blum, E., Mazourek, M., O'Connell, M., Curry, J., Thorup, T., Liu, K., Jahn, M. and Paran, I. (2003) Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. *Theor. Appl. Genet.* **108**, 79-86.

Bosland, P.W. and Baral, J.B. (2007) 'Bhut Jolokia'—the world's hottest known chile pepper is a putative naturally occurring interspecific hybrid. *Hortscience* **42**, 222-224.

Bosland, P.W. and Coon, D. (2015) Novel formation of ectopic (nonplacental) capsaicinoid secreting vesicles on fruit walls explains the morphological mechanism for super-hot chile peppers. *J. Am. Soc. Hortic. Sci.* **140**, 253-256.

Bosland, P.W., Coon, D. and Reeves, G. (2012) 'Trinidad Moruga Scorpion' pepper is the world's hottest measured chile pepper at more than two million scoville heat units. *HortTechnology* **22**, 534-538.

Canto-Flick, A., Balam-Uc, E., Bello-Bello, J.J., Lecona-Guzman, C., Solis-Marroquin, D., Aviles-Vinas, A., Gomez-Uc, E., Lopez-Puc, G. and Santana-Buzzy, N. (2008) Capsaicinoids content in Habanero pepper (*Capsicum chinense* Jacq.): hottest known cultivars. *Hortscience* **43**, 1344-1349.

Caterina, M.J., Schumacher, M.A., Tominaga, M., Rosen, T.A., Levine, J.D. and Julius, D. (1997) The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816-824.

Clark, R. and Lee, S.H. (2016) Anticancer properties of capsaicin against human cancer. *Anticancer Res.* **36**, 837-844.

Curry, J., Aluru, M., Mendoza, M., Nevarez, J., Melendrez, M. and O'Connell, M.A. (1999) Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp. *Plant Sci.* **148**, 47-57.

Fraenkel, L., Bogardus, S.T., Concato, J. and Wittink, D.R. (2004) Treatment options in knee osteoarthritis: the patient's perspective. *Archives of Internal Medicine* **164**, 1299-1304.

Haak, D.C., McGinnis, L.A., Levey, D.J. and Tewksbury, J.J. (2012) Why are not all chilies hot? A trade-off limits pungency. *Proc. Biol. Sci.* **279**, 2012-2017.

Han, K., Jeong, H.J., Sung, J., Keum, Y.S., Cho, M.C., Kim, J.H., Kwon, J.K., Kim, B.D. and Kang, B.C. (2013) Biosynthesis of capsinoid is controlled by the *Pun1* locus in pepper. *Mol. Breed.* **31**, 537-548.

Hill, T.A., Chunthawodtiporn, J., Ashrafi, H., Stoffel, K., Weir, A. and Van Deynze, A. (2017) Regions underlying population structure and the genomics of organ size determination in *Capsicum annuum*. *Plant Genome* 1**0**.

Jang, S., Han, K., Jo, Y.D., Jeong, H.J., Siddique, M.I. and Kang, B.C. (2015) Substitution of a dysfunctional *pAMT* allele results in low-pungency but high levels of capsinoid in *Capsicum chinense* 'Habanero'. *Plant Breed. Biotechnol.* **3**, 119-128.

Kim, M., Kim, S., Kim, S. and Kim, B.-D. (2001) Isolation of cDNA Clones Differentially Accumulated in the Placenta of Pungent Pepper by Suppression Subtractive Hybridization. *Mol. Cells* **11**, 213-219.

Kim, S., Park, J., Yeom, S.I., Kim, Y.M., Seo, E., Kim, K.T., Kim, M.S., Lee, J.M., Cheong, K., Shin, H.S., Kim, S.B., Han, K., Lee, J., Park, M., Lee, H.A., Lee, H.Y., Lee, Y., Oh, S., Lee, J.H., Choi, E., Choi, E., Lee, S.E., Jeon, J., Kim, H., Choi, G., Song, H., Lee, J., Lee, S.C., Kwon, J.K., Lee, H.Y., Koo, N., Hong, Y., Kim, R.W., Kang, W.H., Huh, J.H., Kang, B.C., Yang, T.J., Lee, Y.H.,

Bennetzen, J.L. and Choi, D. (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* **18**, 210.

Kim, S., Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A., Seo, E., Choi, J., Cheong, K., Kim, K.T., Jung, K., Lee, G.W., Oh, S.K., Bae, C., Kim, S.B., Lee, H.Y., Kim, S.Y., Kim, M.S., Kang, B.C., Jo, Y.D., Yang, H.B., Jeong, H.J., Kang, W.H., Kwon, J.K., Shin, C., Lim, J.Y., Park, J.H., Huh, J.H., Kim, J.S., Kim, B.D., Cohen, O., Paran, I., Suh, M.C., Lee, S.B., Kim, Y.K., Shin, Y., Noh, S.J., Park, J., Seo, Y.S., Kwon, S.Y., Kim, H.A., Park, J.M., Kim, H.J., Choi, S.B., Bosland, P.W., Reeves, G., Jo, S.H., Lee, B.W., Cho, H.T., Choi, H.S., Lee, M.S., Yu, Y., Do Choi, Y., Park, B.S., van Deynze, A., Ashrafi, H., Hill, T., Kim, W.T., Pai, H.S., Ahn, H.K., Yeam, I., Giovannoni, J.J., Rose, J.K., Sorensen, I., Lee, S.J., Kim, R.W., Choi, I.Y., Choi, B.S., Lim, J.S., Lee, Y.H. and Choi, D. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **46**, 270-278.

Lang, Y., Kisaka, H., Sugiyama, R., Nomura, K., Morita, A., Watanabe, T., Tanaka, Y., Yazawa, S. and Miwa, T. (2009) Functional loss of *pAMT* results in biosynthesis of capsinoids, capsaicinoid analogs, in *Capsicum annuum* cv. CH-19 Sweet. *Plant J.* **59**, 953-961.

Lee, J., Park, S.J., Hong, S.C., Han, J.H., Choi, D., Yoon, J.B. and Havey, M. (2016) QTL mapping for capsaicin and dihydrocapsaicin content in a population of *Capsicum annuum* 'NB1' × *Capsicum chinense* 'Bhut Jolokia'. *Plant Breed.* **135**, 376-383.

Lee, J.H., An, J.T., Siddique, M.I., Han, K., Choi, S., Kwon, J.-K. and Kang, B.C. (2017) Identification and molecular genetic mapping of Chili veinal mottle virus (ChiVMV) resistance genes in pepper (*Capsicum annuum*). *Mol. Breed.* **37**, 121.

Nimmakayala, P., Abburi, V.L., Saminathan, T., Alaparthi, S.B., Almeida, A., Davenport, B., Nadimi, M., Davidson, J., Tonapi, K., Yadav, L., Malkaram, S., Vajja, G., Hankins, G., Harris, R., Park, M., Choi, D., Stommel, J. and Reddy,

U.K. (2016) Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. *Sci. Rep.* **6**, 38081.

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

Peppin, J.F. and Pappagallo, M. (2014) Capsaicinoids in the treatment of neuropathic pain: a review. *Ther. Adv. Neurol. Disord.* **7**, 22-32.

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Ling-Hu, C., Zhou, H., Lin, H., Gonzalez-Morales, S., Trejo-Saavedra, D.L., Tian, H., Tang, X., Zhao, M., Huang, Z., Zhou, A., Yao, X., Cui, J., Li, W., Chen, Z., Feng, Y., Niu, Y., Bi, S., Yang, X., Li, W., Cai, H., Luo, X., Montes-Hernandez, S., Leyva-Gonzalez, M.A., Xiong, Z., He, X., Bai, L., Tan, S., Tang, X., Liu, D., Liu, J., Zhang, S., Chen, M., Zhang, L., Zhang, L., Zhang, Y., Liao, W., Zhang, Y., Wang, M., Lv, X., Wen, B., Liu, H., Luan, H., Zhang, Y., Yang, S., Wang, X., Xu, J., Li, X., Li, S., Wang, J., Palloix, A., Bosland, P.W., Li, Y., Krogh, A., Rivera-Bustamante, R.F., Herrera-Estrella, L., Yin, Y., Yu, J., Hu, K. and Zhang, Z. (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl Acad. Sci.* **111**, 5135-5140.

Stellari, G.M., Mazourek, M. and Jahn, M.M. (2010) Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. *Heredity* **104**, 460-471.

Stewart, C., Jr., Kang, B.C., Liu, K., Mazourek, M., Moore, S.L., Yoo, E.Y., Kim, B.D., Paran, I. and Jahn, M.M. (2005) The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant J* **42**, 675-688.

Stewart, C., Jr., Mazourek, M., Stellari, G.M., O'Connell, M. and Jahn, M. (2007) Genetic control of pungency in *C. chinense* via the *Pun1* locus. *J. Exp. Bot.* **58**, 979-991.

Tanaka, Y., Nakashima, F., Kirii, E., Goto, T., Yoshida, Y. and Yasuba, K.I. (2017)

Difference in capsaicinoid biosynthesis gene expression in the pericarp reveals elevation of capsaicinoid contents in chili peppers (Capsicum chinense). *Plant cell reports* **36**, 267-279.

Tanaka, Y., Sonoyama, T., Muraga, Y., Koeda, S., Goto, T., Yoshida, Y. and Yasuba, K. (2015) Multiple loss-of-function putative aminotransferase alleles contribute to low pungency and capsinoid biosynthesis in *Capsicum chinense*. *Mol. Breed.* **35**, 142.

Tewksbury, J.J., Levey, D.J., Huizinga, M., Haak, D.C. and Traveset, A. (2008a) Costs and benefits of capsaicin-mediated control of gut retention in dispersers of wild chilies. *Ecology* **89**, 107-117.

Tewksbury, J.J. and Nabhan, G.P. (2001) Directed deterrence by capsaicin in chillies. *Nature* **412**, 403-404.

Tewksbury, J.J., Reagan, K.M., Machnicki, N.J., Carlo, T.A., Haak, D.C., Penaloza, A.L. and Levey, D.J. (2008b) Evolutionary ecology of pungency in wild chilies. *Proc. Natl Acad. Sci.* **105**, 11808-11811.

Yarnes, S.C., Ashrafi, H., Reyes-Chin-Wo, S., Hill, T.A., Stoffel, K.M. and Van Deynze, A. (2013) Identification of QTLs for capsaicinoids, fruit quality, and plant architecture-related traits in an interspecific *Capsicum* RIL population. *Genome* **56**, 61-74.

Zhang, Z.X., Zhao, S.N., Liu, G.F., Huang, Z.M., Cao, Z.M., Cheng, S.H. and Lin, S.S. (2016) Discovery of putative capsaicin biosynthetic genes by RNA-Seq and digital gene expression analysis of pepper. *Sci. Rep.* **6**, 34121.

CHAPTER I

QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*

The research described in this chapter has been submitted in *Plant Biotechnology Journal*.

ABSTRACT

Capsaicinoids are unique compounds produced only in peppers (*Capsicum* spp.). Several studies using classical quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) have identified QTL controlling capsaicinoid content in peppers; however, neither the QTL common to each population nor the candidate genes underlying them have been identified due to the limitations of each approach used. Here, we performed QTL mapping and GWAS for capsaicinoid content in peppers using two recombinant inbred line (RIL) populations and one GWAS population. Whole-genome resequencing and genotyping-by-sequencing (GBS) were used to construct high-density single-nucleotide polymorphism (SNP) maps. Five QTL regions on chromosomes 1, 2, 3, 4, and 10 were commonly identified in both RIL populations over multiple locations and years. Furthermore, a total of 109,610 SNPs derived from two GBS libraries were used to analyze the GWAS population consisting of 208 *C. annuum*-clade accessions. A total of 69 QTL regions were identified from the GWAS, 10 of which were co-located with the QTL identified from the two biparental populations. Within these regions, we were able to identify five candidate genes known to be involved in capsaicinoid biosynthesis. Our

results demonstrate that QTL mapping and GBS-GWAS represent a powerful combined approach for the identification of loci controlling complex traits.

INTRODUCTION

Hot peppers (*Capsicum* spp.) contain capsaicinoids, unique compounds that produce a burning sensation called pungency. Capsaicinoids are believed to protect pepper fruits from diseases such as *Fusarium* (Tewksbury et al., 2008), and enable the dispersal of their seeds by birds, which, unlike mammals, cannot detect the pungency and do not harm the seeds (Tewksbury and Nabhan, 2001). Humans use pungent peppers as a vegetable, in sauces, and in food additives (Aza-Gonzalez et al., 2011), while capsaicinoids are also used in pharmaceuticals and other medicines (Aza-Gonzalez et al., 2011; Luo et al., 2011).

The presence of capsaicinoids is mainly controlled by *Pun1*, which encodes capsaicin synthase (CS) (Stewart et al., 2005). CS functions in the final step of capsaicinoid biosynthesis, and the expression of *Pun1* is detected only in the fruits (Stewart et al., 2005; Stewart et al., 2007). Most non-pungent pepper cultivars have non-functional *Pun1* alleles, containing a deletion (*pun1*), a frameshift mutation (*pun1*^{2}), or an early stop codon (*pun1*^{3}) (Stellari et al., 2010; Stewart et al., 2005; Stewart et al., 2007). Mutations in another gene, *Putative Aminotransferase* (*pAMT*), convert biosynthesis of capsaicinoids into that of capsinoids, which are about 1,000 times less pungent than capsaicinoids (Lang et al., 2009). Several non-functional *pamt* alleles have been identified and used to breed high-capsinoid pepper varieties (Jang et al., 2015; Jeong et al., 2015; Tanaka et al., 2015; Tanaka et

al., 2014). Gene expression analyses have revealed other genes that function in the capsaicinoid biosynthesis pathway, including those encoding phenylalanine ammonia lyase (*Pal*), 3-keto-acyl-ACP synthase (*Kas*), and thioesterase (*Fat*) (Curry, 1999; Aluru, 2003; Kim, 2001). The expression levels of these genes correlate with the capsaicinoid content, but allelic variations affecting capsaicinoid biosynthesis have been identified only for *Pun1* and *pAMT* (Koeda, 2015).

The capsaicinoid content of peppers is controlled by quantitative trait loci (QTL) (Collins et al., 1995; Sanatombi and Sharma, 2008), which have been identified from several interspecific populations (Ben-Chaim et al., 2006; Blum et al., 2003; Lee et al., 2016b; Yarnes et al., 2013). However, direct comparisons between the QTL from different studies are not possible due to the limited numbers of common markers between populations and the low density of genetic maps. Although capsaicinoid biosynthesis genes may be located at these QTL, no likely candidate genes underlying the QTL have been proposed in these previous studies.

Traditional QTL mapping is highly dependent on the genetic diversity of the two parents, and the effects of the detected QTL can vary between populations. QTL regions can also be quite large, incorporating too many genes to investigate as potential candidate genes. The limitations of QTL analysis can be overcome using genome-wide association studies (GWAS), which can narrow down the candidate regions using natural populations. GWAS does have the potential for false positive error however, and validation of the results is necessary (Korte and Farlow, 2013; Zhu et al., 2008). The number of markers used in the GWAS highly affects its

results. Genotyping-by-sequencing (GBS) is one of the genotyping methods used for GWAS, and GBS-GWAS approaches have been successfully applied to the identification of candidate genes controlling quantitative traits in plant species including soybean (*Glycine max*), diploid alfalfa (*Medicago sativa*), chickpea (*Cicer arietinum*), and maize (*Zea mays*) (Navarro et al., 2017; Sakiroglu and Brummer, 2017; Sonah et al., 2015; Upadhyaya et al., 2016). There have been only two reports on the use of GWAS for analysis of capsaicinoid content in *Capsicum.* Using 176 simple sequence repeats and 96 *C. annuum* accessions, Nimmakayala et al. (2014) identified one marker on chromosome 1 that was associated with the capsaicin and dihydrocapsaicin contents, while Nimmakayala et al. (2016) used 7,331 single nucleotide polymorphisms (SNPs), of which 72 were found to be associated with capsaicinoid content, including in a candidate gene encoding an ankyrin-like protein which has acyltransferase function similar to CS.

GWAS can have high rates of false positive errors due to the population structures (Zhu et al., 2008). The combination of GWAS and QTL analyses can compensate for the limitations of each approach, enabling the identification of loci controlling agronomically important quantitative traits. Such combined approaches have been successfully used to identify candidate genes controlling flowering time, panicle architecture, leaf architecture, frost resistance, and seed-related traits in *Arabidopsis thaliana*, rice (*Oryza sativa*), maize, winter faba bean (*Vicia faba*), and soybean, respectively (Brachi et al., 2010; Crowell et al., 2016; Sallam et al., 2016; Sonah et al., 2015; Tian et al., 2011).

In this study, we performed QTL mapping in one intraspecific and one interspecific RIL population of *Capsicum*. High-density genetic maps and phenotype data from multiple environments were used to ensure an accurate linkage analysis. In addition, a total of 208 *C. annuum*-clade accessions, including *C. annuum*, *C. chinense*, and *C. frutescens*, were genotyped by GBS and analyzed using GWAS. By comparing the physical locations of the QTL identified in this study and previous work, five candidate genes in the capsaicinoid biosynthesis pathway were proposed.

MATERIALS AND METHODS

Plant materials

Two RIL populations were used in this study. The intraspecific population of 120 RILs ($F_{7:10}$) was derived from *C. annuum* 'Perennial' × *C. annuum* 'Dempsey' (Han et al, 2016). Among them, 56 RILs were pungent and were used for the QTL analysis. An interspecific population of 85 RILs ($F_{9:11}$) derived from *C. annuum* 'TF68' × *C. chinense* 'Habanero', provided by Prof. Byung-Dong Kim of Seoul National University (Kim et al., 2010), was also used. These populations were referred to as 'PD' and 'TH' RILs, respectively, following their parental names. The 'PD' RIL population was grown in Hana Seed Co., Ltd., in Anseong (2011 and 2012a) and Seoul National University farm in Suwon, Republic of Korea (2012b), while 'TH' RILs were grown in Anseong (2013 and 2014). The plants in both Anseong and Suwon were grown in plastic greenhouses; however, the plants were grown in soil in Anseong and in pots in Suwon. Five plants were grown for each line.

A sub-population of *Capsicum* core collections was used for GWAS (Lee et al., 2016a). To reduce the effect of the major gene controlling pungency, *Pun1*, all accessions were genotyped using the MAP1 marker (Rodríguez-Maza et al., 2012; Stewart et al., 2005). A total of 208 accessions were selected, including 140 from the CC240 core collection (Lee et al., 2016a) and 68 additional accessions (Table I-

1). Five plants of each accession were grown in plastic greenhouses at the RDA-Genebank in Jeonju, Republic of Korea.

Table I- 1. Accessions used for GWAS

Species	Number of accessions		
	CC240*	Additional accessions	Total
Capsicum annuum	110	35	145
Capsicum chinense	16	26	42
Capsicum frutescens	14	7	21
Total	140	68	208

*Lee et al., 2016a

Evaluation of capsaicinoid content

The placental tissue of fruits from ‘PD’ RIL and ‘TH’ RIL was dissected for capsaicinoid extraction to reduce the effect of fruit size. Capsaicinoids were extracted following the method of Han et al. (2013). For the GWAS population, the modified protocol of Han et al. (2013) was used to extract the capsaicinoid. In short, three biological replicates were prepared by freeze-drying whole fruits, which were then ground using a hand blender (HR2860; Koninklijke Philips, Amsterdam, the Netherlands) and stored in sealed containers at -80℃. To extract multiple samples at one time, 0.1 g of pepper powder was placed in a 2-mL microcentrifuge tube and mixed with 1.5 mL of a 6:4 ethyl acetate:acetone solution. After incubation for one day at 37℃, 1 mL supernatant was transferred to a 1.5-mL microcentrifuge tube and dried using a centrifugal speed vacuum concentrator SVQ-70 (Operon, Gimpo, Republic of Korea). The pellet was dissolved in 1 mL methyl alcohol and filtered using a 0.2-μm syringe filter (PN4450; Pall Corporation, Port Washington, NY, USA).

The filtered extracts were transferred to a high-performance liquid chromatography (HPLC) vial (5182-0715; Agilent Technologies, Santa Clara, CA, USA). The contents of capsaicin and dihydrocapsaicin were quantified using HPLC in the National Instrumentation Center for Environmental Management (Seoul, Republic of Korea). Capsaicin and dihydrocapsaicin standards were purchased from Sigma-Aldrich (St. Louis, MO, USA; M2028 and M1022, respectively).

gDNA extraction and genotyping-by-sequencing

DNA was extracted from the 'TH' RIL and GWAS populations using the CTAB method (Lee et al., 2017) and diluted to 50 ng/μL with distilled water. GBS libraries of 'TH RIL' were generated by digestion with PstI/MseI using a SBG 100-Kit v2.0 (Keygene N.V., Wageningen, the Netherlands), while those of the GWAS population were constructed manually following digestion with PstI/MseI and EcoRI/MseI, according to a previously reported protocol (Jo et al., 2017; Truong et al., 2012). In either case, DNA was digested with the restriction enzymes and adapters were ligated to it. The libraries were amplified with selective primers, which used 'GA' for 'TH RIL' and 'TA' for the GWAS population. Amplified libraries generated from 85 'TH' RILs and two replicates of each of the population parents were pooled in a single tube. The libraries of the GWAS population were pooled in five tubes. All tubes were sequenced in separate lanes of a HiSeq 2000 (Illumina, San Diego, CA, USA) at Macrogen (Seoul, Republic of Korea).

Reference-based SNP calling

Raw 101-bp reads of the GWAS and 'TH' RIL libraries were trimmed to a minimum length of 80 bp and filtered for a minimum quality of Q20. The filtered reads were aligned to the *C. annuum* 'CM334' reference chromosomes v1.55 (Kim et al., 2014) using the Burrows-Wheeler Aligner program v0.7.12 (Li and Durbin, 2010). For SNP calling and filtering, the GATK Unified Genotyper v3.3-0 was

used (DePristo et al., 2011). SNPs from the 'TH' RIL population were filtered for a minimum genotype quality of 20 and a minimum read depth of 3. For the GWAS population, SNPs were filtered for a minor allele frequency > 0.03, a calling rate > 0.6, and an inbreeding coefficient (F) > 0.8.

Bin map construction for the RILs

Missing data for parents were imputed using the FSFHapImputation plugin for Tassel 5 (Swarts et al., 2014), and the recombination breakpoints of the RILs were detected using a sliding window approach (Han et al., 2016; Huang et al., 2009). The ratio of SNPs with maternal and paternal genotypes was calculated for each window, defined as 18 linked SNPs, and the overall genotype of each window was decided. Ratios of > 0.7, 0.3–0.7, and < 0.3 were scored as maternal, heterozygous, and paternal genotypes, respectively. With the exception of the threshold for the recombination breakpoints, the methods described by Huang et al. (2009) were used. The genetic locations of the bins were decided using the Carthagene program (De Givry et al., 2005).

The bin map of 'PD' RIL was constructed using SNPs from the re-sequencing data and the sliding window approach (Han et al., 2016). The bin maps of both 'PD RIL' and 'TH RIL' were constructed based on the *C. annuum* 'CM334' reference genome (Kim et al., 2014), and were compared using physical locations on the reference genome by the Marker Browser Phyzen Genomics Institute (Seongnam, Republic of Korea) and the MapChart v2.2 program (Voorrips, 2002).

QTL analysis for capsaicinoid content

QTL controlling the contents of capsaicin, dihydrocapsaicin, and total capsaicinoid (the combined capsaicin and dihydrocapsaicin contents) were independently detected for 'PD RIL' and 'TH RIL'. Composite interval mapping was performed using Windows QTL Cartographer v2.5 (Wang et al., 2012) and the LOD threshold was determined by 500 permutation tests ($P < 0.05$) for each trait. When the genetic locations of the QTL (at a 99% significance level) overlapped in the plants grown in the different environments, they were defined as a single QTL. The physical locations of the QTL from 'PD RIL' and 'TH RIL' were also compared with the genetic and physical location of bins linked to the QTL. Epistatic effects between the QTL were identified using a MIM analysis with a Bayesian information criterion (BIC-X) model using the default options.

Genome-wide association analyses for capsaicinoid content (this experiment was done by Hea-Young Lee, and all the data related to GWAS analysis was used under permission)

The 109,610 filtered SNPs detected from the 208 individuals of the GWAS population were used for association mapping. The population structure estimation (PCA and Kinship matrixes) and GWAS (based on the compressed mixed linear model) were conducted using the R package Genomic Association and Prediction Integrated Tool (Lipka et al., 2012) with default settings. The *P*-values of SNPs

from GWAS underwent an FDR analysis, and the FDR-adjusted *P*-value of 0.05 was used to set the significant threshold level.

Haplotype block estimation and candidate gene identification

The haplotype block of the GWAS population was estimated using PLINK v1.9 (Chang et al., 2015) with the following settings: '--no-parents --allow-no-sex --blocks-max-kb 2000 --blocks-inform-frac 0.9 --blocks-strong-highci 0.85 --blocks-recomb-highci 0.8'. Candidate genes located at the associated regions were identified, and their functions were annotated using Blast2Go (Gotz et al., 2008).

RESULTS

Measurement of capsaicinoid content in the biparental populations

'Perennial' is a pungent small pepper line, while 'Dempsey' is a non-pungent bell pepper cultivar. Due to the non-functional *pun1* allele of the paternal line 'Dempsey', the 'PD' RIL population created in this cross had a 1:1 segregation ratio of pungency, comprising 56 pungent and 64 non-pungent RILs. Capsaicinoid content was evaluated from plants grown in three different environments. The capsaicinoid contents in the placenta of fruits from the pungent parent 'Perennial' grown in Anseong were 38,013 and 31,518 μg/g dry weight (DW) in 2011 and 2012a, whereas the capsaicinoid content of placental tissues from plants grown in Suwon (2012b) was 81,257 μg/g DW (Table I- 21). The average capsaicinoid contents of the placental tissues from the pungent RILs were 16,555, 13,005, and 22,058 μg/g DW in 2011, 2012a, and 2012b, respectively (Table I- 2A; Figure I- 1). Transgressive segregation was observed in 2011 and 2012a.

The parents of the 'TH' RILs, 'Habanero' and 'TF68', are both pungent, with 'Habanero' found to be more pungent than 'TF68' (Table I- 3Table I- 2; Figure I- 1B). The capsaicinoid contents of the placental tissues in 'TF68' and 'Habanero' were 5,672 and 89,825 μg/g DW, respectively, in 2013, and 7,199 and 73,819 μg/g DW, respectively, in 2014. The average placental tissue capsaicinoid contents of

the 85 ‘TH’ RILs phenotyped for QTL mapping were 25,809 and 23,953 μg/g DW in 2013 and 2014, respectively. In both years, RILs more pungent than ‘Habanero’ were identified.

The distribution of total capsaicinoid content showed a positive skew in both the ‘PD’ RIL and ‘TH’ RIL populations. A large percentage of the ‘TH’ RILs had a lower or similar capsaicinoid content compared with ‘TF68’, and these skewed distributions were also detected for the individual capsaicin and dihydrocapsaicin contents (data not shown). The contents of capsaicin, dihydrocapsaicin, and the total capsaicinoids showed a high level of correlation between all environments, with Pearson correlation coefficients of between 0.64 and 0.99 (Figure I- 2).

Table I- 2. Capsaicinoid contents (μg/g DW) of Perennial, Dempsey, and 'PD' RIL plants grown in three different environments

	2011			2012a			2012b		
	P	D	RIL	P	D	RIL	P	D	RIL
CAP	25,043	ND	9,675 ± 9,589.9	16,970 ± 1,368.7	ND	6,970 ± 6,846.4	42,793 ± 8,729.5	ND	11,812 ± 10,993.2
DICAP	12,969	ND	6,880 ± 7,203.3	14,548 ± 1,439.3	ND	6,035 ± 6,381.5	38,463 ± 5,463.7	ND	10,449 ± 9,586.0
TOTAL	38,013	ND	16,555 ± 16,595.8	31,518 ± 2,808.0	ND	13,005 ± 13,159.0	81,257 ± 14,193.1	ND	22,058 ± 20,255.4

ND, not detected

CAP, capsaicin content; DICAP, dihydrocapsaicin content; TOTAL, total capsaicinoid content

2012a and 2012b means different grown places and detailed information is described at material and method

P, Perennial; D, Dempsey

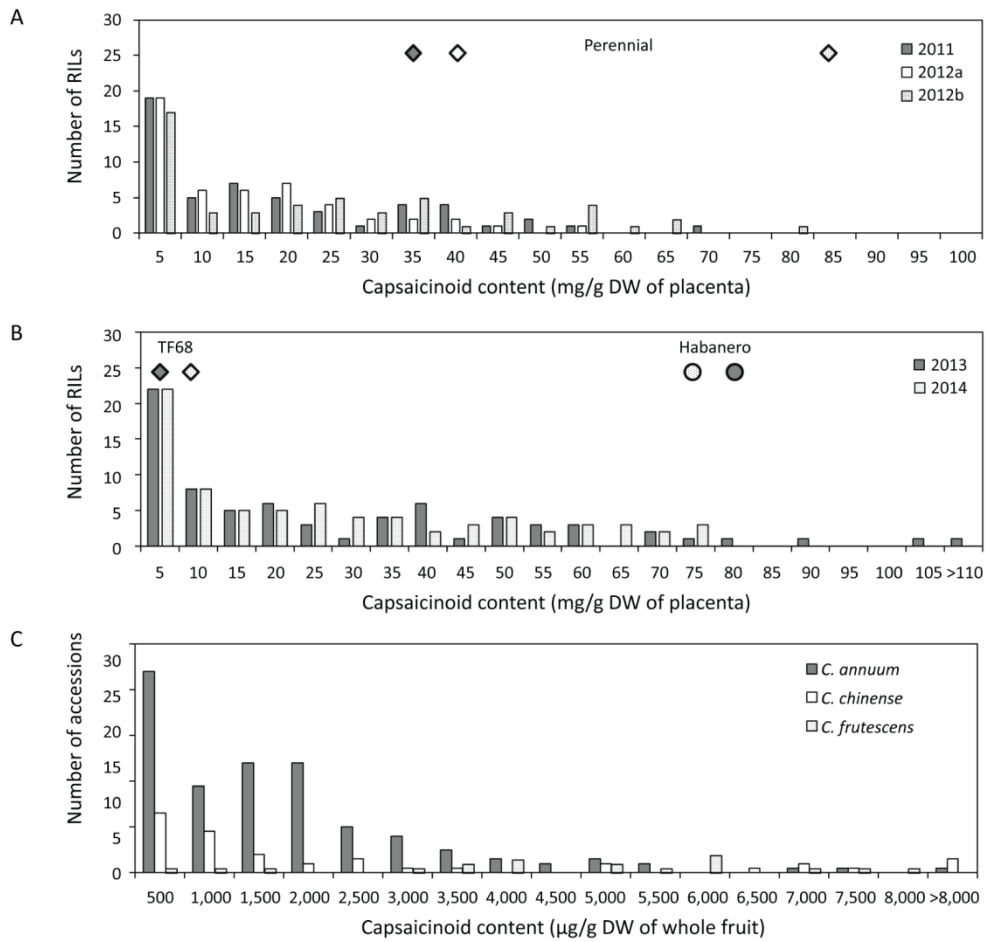


Figure I- 1. Capsaicinoid contents of the 'PD' RILs (A), 'TH' RILs (B), and GWAS population (C). Diamonds and circles show the average contents of the maternal parents (Perennial or TF68) and the paternal parent (Habanero), respectively. Dempsey, the paternal parent of 'PD' RILs, was non-pungent.

Table I- 3. Capsaicinoid contents (μg/g DW) of TF68, Habanero, and 'TH' RIL plants grown in two different environments

Trait	2013			2014		
	T	H	RIL	T	H	RIL
CAP	3,280 ± 1,585.8	50,484	12,646 ± 14,584.0	4,068 ± 4,067.8	38,102 ± 1,486.4	11,640 ± 11,317.0
DICAP	2,392 ± 1,054.2	39,341	13,163 ± 15,531.4	3,131 ± 3,130.8	35,717 ± 2,607.3	12,314 ± 12,132.9
TOTAL	5,672 ± 2,572.7	89,825	25,809 ± 28,508.7	7,199 ± 7,198.6	73,819 ± 3,644.1	23,953 ± 22,595.7

CAP, capsaicin content; DICAP, dihydrocapsaicin content; TOTAL, total capsaicinoid content

T, TF68; H, Habanero

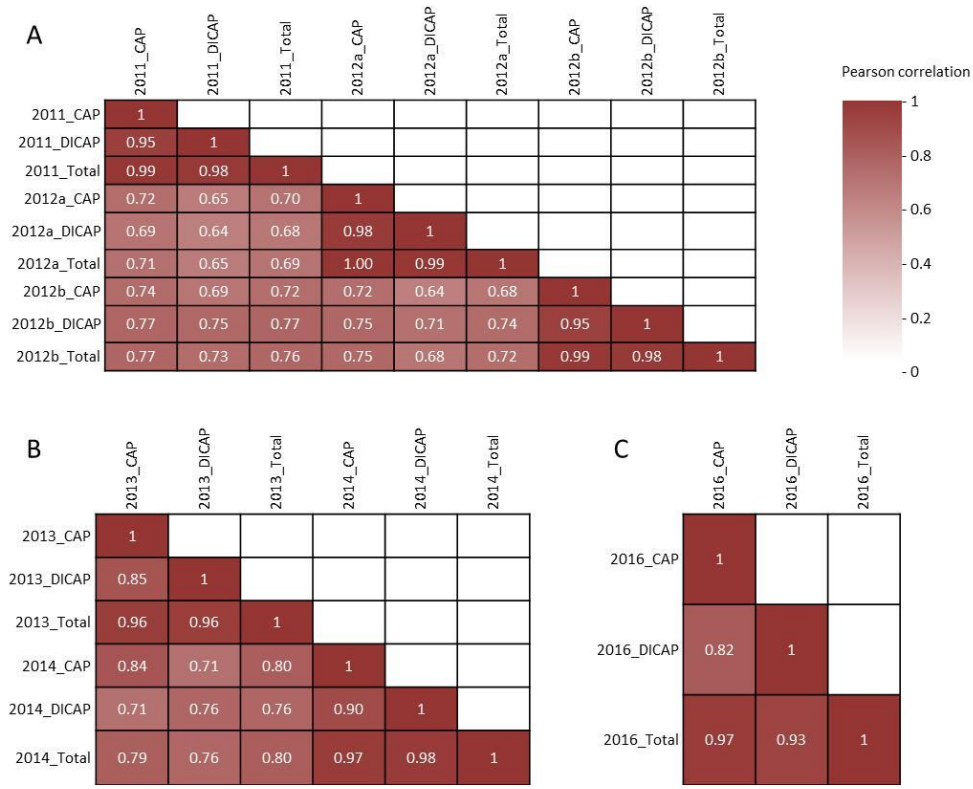


Figure I- 2. Correlation between the contents of capsaicin, dihydrocapsaicin, and total capsaicinoids in 'PD' RILs (A), 'TH' RILs (B), and the GWAS population (C). CAP, capsaicin; DICAP, dihydrocapsaicin; Total, total capsaicinoid.

Bin map of biparental populations

Genotypes of ‘TH’ RILs were analyzed using GBS after the preparation of libraries from PstI/MseI-digested DNA. The average number of reads per sample was around 4 million, and a total of 8,587 SNPs were detected by aligning the sequences obtained from GBS to the *C. annuum* ‘CM334’ reference genome (Table I- 4). The SNPs were more densely distributed at the ends of the chromosomes (Figure I- 3A). To correct for missing data and genotyping error, a sliding window approach was used (Han et al., 2016). Recombination breakpoints were determined using 18 consecutive SNPs as one sliding window, and a high-density bin map of the ‘TH’ RIL population was constructed. The map consisted of 1,089 bins with an average genetic distance of 1.0 cM (Table I- 5). Among the 12 linkage groups, the genetic distance of chromosome 1 was longest and chromosome 8 was shortest.

For the genotyping of the ‘PD’ RIL population, previously reported whole-genome resequencing data were used (Han et al., 2016). Due to the higher density of SNPs and larger number of RILs in ‘PD’ than ‘TH’, more recombination breakpoints were identified in ‘PD’ RILs (3,983) than in ‘TH’ RILs (2,386); therefore, the average distance between bins was shorter in ‘PD’ RILs (0.5 cM). The total genetic lengths of the ‘PD’ RIL and ‘TH’ RIL maps were estimated to be 1,372 cM and 1,127 cM, respectively. The two bin maps were compared with the *C. annuum* ‘CM334’ reference genome, and the physical position of each bin was determined (Figure I- 4). With the exception of chromosome 8, the overall genetic and physical positions of the bins were colinear. More bins were detected in the

middle of chromosomes for ‘PD’ than in ‘TH’, which reflected the distribution of the SNPs (Figure I- 3B).

Table I- 4. Number of sequencing reads and SNPs used for GWAS and QTL mapping

	TH RIL	PD RIL	GWAS population
# of accessions (lines)	85	56	208
Genotyping method	GBS (PstI/MseI)	WGS	GBS (PstI/MseI and EcoRI/MseI)
Avg. # of reads per sample	4,103,757	15,738,890*	3,326,422
Total # of SNPs	8,587	1,431,214	109,610
Avg. distance b/w SNPs (bp)	328,662	2,713	25,093

*Number of sites sequenced by paired-end reads

WGS, whole genome resequencing; GBS, genotyping-by-sequencing

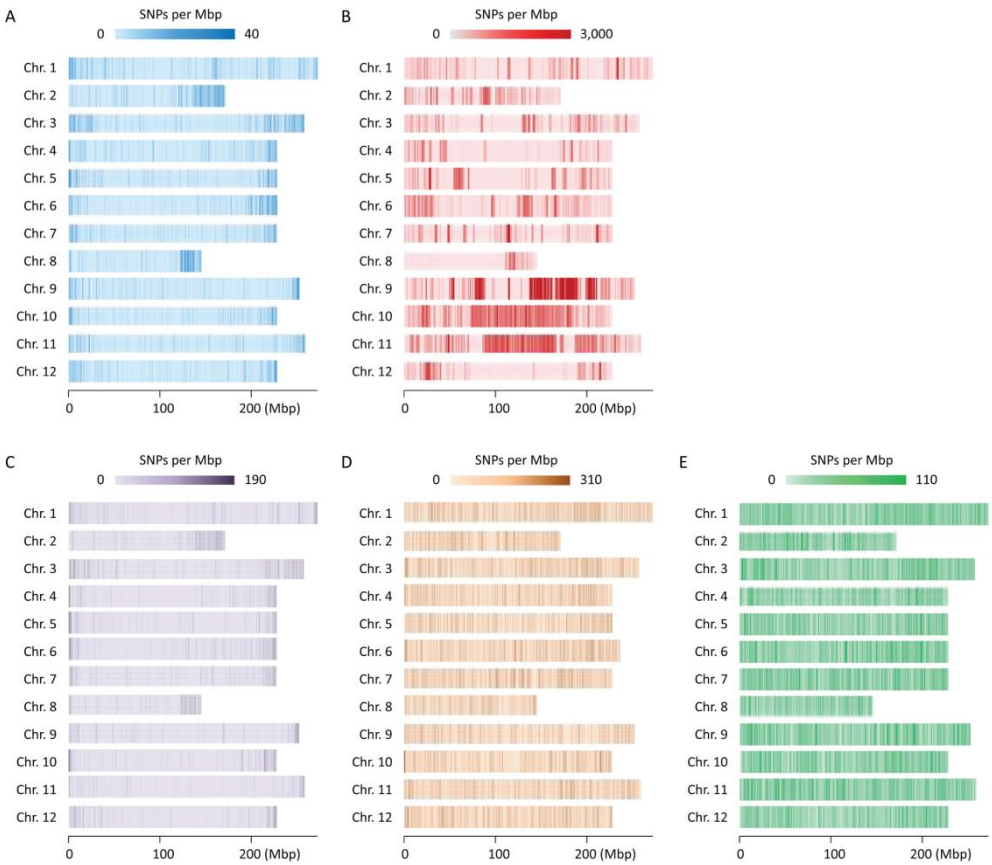


Figure I- 3. SNP density (number of SNPs per Mbp) of the 'TH' (A) and 'PD' (B) RIL populations, as well as the GWAS populations digested using PstI/MseI (C), EcoRI/MseI (D) and PstI/MseI and EcoRI/MseI (E).

Table I- 5. Bin map of the 'TH' RIL population

Chr.	Number of SNPs	Number of bins	Physical length of bin (Mb)		Genetic distance of bin (cM)	
			Mean	Total	Mean	Total
1	920	118	2.3	272.7	1.3	155.2
2	804	93	1.8	171.1	1.2	112.1
3	1,060	127	2.0	257.9	0.9	111.7
4	657	102	2.2	222.6	0.8	81.7
5	672	105	2.2	233.5	0.7	78.2
6	794	101	2.3	236.9	0.9	89.9
7	679	78	3.0	231.9	1.0	80.5
8	541	35	4.1	145.1	1.1	39.6
9	614	88	2.9	252.8	1.0	88.5
10	589	83	2.8	233.6	1.3	105.4
11	613	88	3.0	259.7	1.0	87.3
12	644	71	3.3	235.7	1.4	97.2
Total	8,587	1,089	2.5	2,753.5	1.0	1,127.3

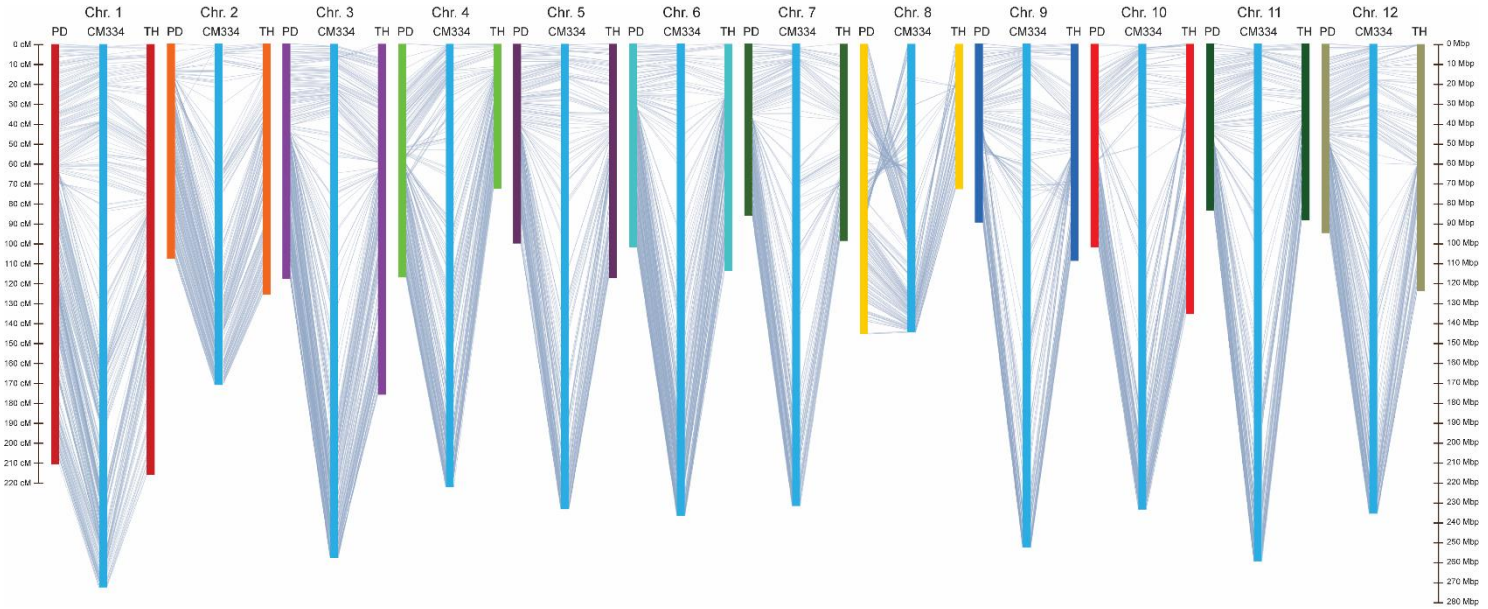


Figure I- 4. Comparison of the genetic maps of 'PD' and 'TH' RILs with the physical map. Bars on the left and right show the genetic map position (cM) and the physical map position (Mbp), respectively. PD, genetic map of 'PD' RILs; TH, genetic map of 'TH' RILs; CM334, physical map of the *C. annuum* 'CM334' reference genome.

QTL mapping for capsaicinoid content

QTL controlling the contents of capsaicin, dihydrocapsaicin, and total capsaicinoid were detected in 'PD' RILs and 'TH' RILs (Table I- 6; Table I- 7). For 'PD', the capsaicinoid contents of 56 RILs and an ultra-high-density bin map of 120 RILs were used to identify 5, 9, and 8 QTL for the capsaicin, dihydrocapsaicin, and total capsaicinoid contents, respectively. The QTL were located on chromosomes 1, 2, 3, 4, 6, 10, and 12 (Table I- 6). Five QTL, *PD-cap10* (capsaicin-related), *PD-dicap1.1* and *PD-dicap10.2* (dihydrocapsaicin-related), and *PD-total2* and *PD-total10.2* (total capsaicinoid-related) were independently identified in plants grown in two environments. Moreover, with the exception of five QTL, the majority of the QTL regions corresponded to more than two traits, such as the region containing *PD-cap10*, *PD-dicap10.2*, and *PD-total10*. RILs with the 'Perennial' SNP genotypes at the QTL showed an increased capsaicinoid level, except for those at *PD-dicap2.1*, *PD-dicap12*, and *PD-total12*. *PD-dicap1.1* showed the highest LOD value (8.7) and *PD-dicap10.2* showed the highest R^2 value (28.8%).

The capsaicinoid contents of the 'TH' RILs grown in two environments were evaluated for the QTL analysis of this population. A total of 8, 5, and 9 QTL for the capsaicin, dihydrocapsaicin, and total capsaicinoid contents were detected, located on chromosomes 1, 2, 3, 4, 6, and 10 (Table I- 7). Among these 22 QTL, 15 had negative additive effects, meaning plants with the 'Habanero' genotype showed an increased capsaicinoid content. *TH-cap2.2* showed the highest LOD score, and

explained 18.8% of total phenotypic variation of the capsaicin contents. As seen for 'PD' RILs, 15 of the 'TH' RIL QTL also controlled two traits; however, no QTL regions were detected to control all three traits.

The physical locations of the QTL detected in 'PD' and 'TH' RILs were compared using the *C. annuum* 'CM334' reference genome (Figure I- 5A). Among the 22 QTL from each population, 9 QTL from 'PD' RILs and 7 from 'TH' RILs were located at the same positions. On chromosome 1, *PD-cap1*, *PD-dicap1.1*, *PD-total1.1*, and *TH-cap1.4* were located at 39.6–60.6 Mbp. *PD-dicap2.1*, *TH-cap2.2*, and *TH-total2* were located 124.8–132.2 Mbp along chromosome 2, while *PD-cap3* and *TH-total3.2* were located 225.1–237.8 Mbp along chromosome 3. *PD-total4.2* and *TH-cap4* appeared to be located near the markers of a single region on chromosome 4, but their positions were relatively distant, located at 110.4 and 177.4 Mbp, respectively. The largest numbers of QTL were located at 9.6–23.9 Mbp along chromosome 10, including *PD-cap10*, *PD-dicap10.2*, *PD-total10.2*, *TH-dicap10*, and *TH-total10*. The QTL located on chromosome 10 could explain the contents of capsaicin, dihydrocapsaicin, and total capsaicinoids, and the R^2 value was higher than 15.0%. Five genetic regions on chromosomes 1 (39.6-60.6 Mbp), 2 (124.8-132.2 Mbp), 3 (225.1-237.8 Mbp), 4 (110.4-177.4 Mbp), and 10 (9.6-23.9 Mbp) were considered to be common QTL, and the QTL located on chromosome 10 were considered the most significant.

Table I- 6. QTL controlling capsaicin, dihydrocapsaicin, and total capsaicinoid contents detected in 'PD' RIL

Trait	QTL	Year	Chr.	Location (cM)	LOD	R^2 (%)	Direction*	Additive effect
CAP	*PD-cap1*	2011	1	50.4–52.2	7.0	25.3	+	5.0
	PD-cap2	2012a	2	83.5–87.8	5.0	18.4	+	3.5
	PD-cap3	2012b	3	82.7–90.2	5.0	14.4	+	4.9
	PD-cap6	2012b	6	47.3–50.5	7.0	22.3	+	5.3
	PD-cap10	2011, 2012a	10	28.6–35.4	4.6–6.6	15.2–25.2	+	3.6–3.9
DICAP	*PD-dicap1.1*	2011, 2012b	1	47.1–61.1	4.6–8.7	13.8–27.7	+	3.7–3.9
	PD-dicap1.2	2011	1	117.8–119.5	4.0	11.0	+	2.5
	PD-dicap1.3	2011	1	121.5–128.3	4.0	11.0	+	2.5
	PD-dicap2.1	2012a	2	50.0–56.4	3.0	11.2	-	2.3
	PD-dicap2.2	2012a	2	82.8–87.1	5.0	15.5	+	3.0
	PD-dicap4	2011	4	45.4–50.3	5.0	13.0	+	2.8
	PD-dicap10.1	2011	10	18.8–26.1	5.0	15.2	+	2.9
	PD-dicap10.2	2012a, 2012b	10	29.1–34.6	6.6–7.3	21.2–28.8	+	3.6–4.6
	PD-dicap12	2012a	12	28.4–32.6	3.0	11.1	-	2.4
TOTAL	*PD-total1.1*	2011	1	50.0–60.5	3.0	11.1	+	8.8
	PD-total1.2	2011	1	122.3–128.8	4.0	10.2	+	5.4
	PD-total2	2012a, 2012b	2	83.2–88.0	4.0–5.2	15.5–17.1	+	6.5–10.5
	PD-total4.1	2011	4	42.7–49.2	4.0	10.7	+	5.8
TOTAL	*PD-total4.2*	2012b	4	67.3–72.5	4.0	15.2	+	8.2
	PD-total10.1	2011	10	22.1–24.9	5.0	14.8	+	6.8
	PD-total10.2	2011, 2012a	10	28.6–32.8	4.9–7.1	15.7–27.2	+	6.9–7.3
	PD-total12	2012a	12	28.4–32.5	4.0	11.6	-	5.0

*Genotypes that increase the pungency level. + means the genotype resembles that of Perennial

Underlined QTL were common to both populations

CAP, capsaicin; DICAP, dihydrocapsaicin; TOTAL, total capsaicinoid

2012a and 2012b means different grown places and detailed information is described at material and method

Table I- 7. QTL controlling capsaicin, dihydrocapsaicin, and total capsaicinoid contents detected in 'TH' RIL

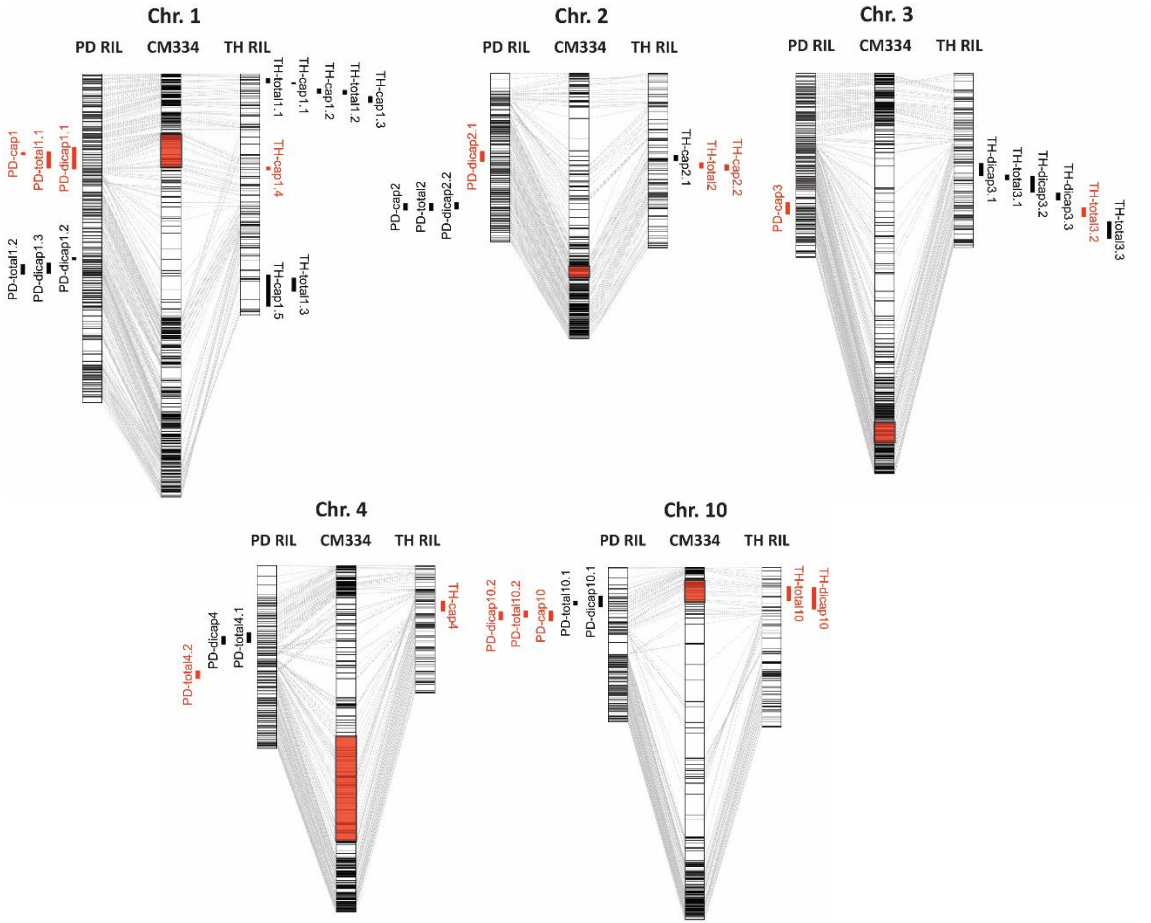
Trait	QTL	Year	Chr.	Location (cM)	LOD	R^2 (%)	Direction*	Additive effect
CAP	*TH-cap1.1*	2014	1	5.6–6.6	4.1	8.4	-	5.1
	TH-cap1.2	2014	1	9.6–12.6	4.5	9.1	-	5.0
	TH-cap1.3	2014	1	14.5–18.5	7.2	13.7	-	6.3
	TH-cap1.4	2013	1	59.8–61.7	5.4	21.1	-	10.7
	TH-cap1.5	2013,2014	1	129.2–149.3	4.2, 4.6	6.1–11.8	-	3.5–6.7
	TH-cap2.1	2014	2	52.3–56.0	9.8	18.5	-	6.2
	TH-cap2.2	2014	2	58.4–62.0	9.9	18.8	-	6.6
	TH-cap4	2014	4	22.6–29.3	4.7	10.3	+	4.4
DICAP	*TH-dicap3.1*	2014	3	57.6–65.7	4.6	9.5	+	4.1
	TH-dicap3.2	2013,2014	3	65.7–76.0	5.5–5.6	11.2–18.2	+	4.2–7.7
	TH-dicap3.3	2013	3	76.6–81.3	3.6	12.4	+	7.0
	TH-dicap6	2013	6	83.9–86.8	6.0	19.7	-	12.4
	TH-dicap10	2014	10	13.3–27.5	5.2	10.7	-	4.9
TOTAL	*TH-total1.1*	2014	1	3.1–6.0	6.2	12.9	-	11.4
	TH-total1.2	2014	1	10.6–13.3	7.9	15.8	-	12.8
	TH-total1.3	2013	1	131.0–139.7	3.8	12.2	-	14.3
	TH-total2	2014	2	57.3–60.8	8.7	16.1	-	12.6
	TH-total3.1	2013	3	65.0–68.1	4.8	14.4	+	11.9
	TH-total3.2	2014	3	86.0–91.7	3.5	6.8	+	8.4
	TH-total3.3	2014	3	94.8–106.1	5.4	10.0	+	10.8
	TH-total6	2013	6	83.9–86.6	3.8	12.9	-	20.7
	TH-total10	2014	10	12.7–22.0	7.8	15.5	-	13.0

*Genotypes that increase the pungency level. + means the genotype resembles that of TF68.

Underlined QTL were common to both populations

CAP, capsaicin; DICAP, dihydrocapsaicin; TOTAL, total capsaicinoid

A


Chr. 1
PD RIL
CM334
TH RIL
PD-total1.2
PD-dicap1.3
PD-dicap1.2
PD-total1.1
PD-dicap1.1
TH-total1.1
TH-cap1.4
TH-total1.3
TH-cap1.5
TH-cap1.1
TH-total1.2
TH-cap1.2
TH-cap1.3
PD-cap2
PD-total2
PD-dicap2.2
PD-dicap2.1
TH-cap2.1
TH-total2
TH-cap2.2
PD-cap3
TH-dicap3.1
TH-total3.1
TH-dicap3.2
TH-dicap3.3
TH-total3.2
TH-total3.3
Chr. 2
Chr. 3
Chr. 4
PD-total4.2
PD-dicap4
PD-total4.1
TH-cap4
Chr. 10
PD-dicap10.2
PD-total10.2
PD-cap10
PD-total10.1
PD-dicap10.1
TH-total10
TH-dicap10

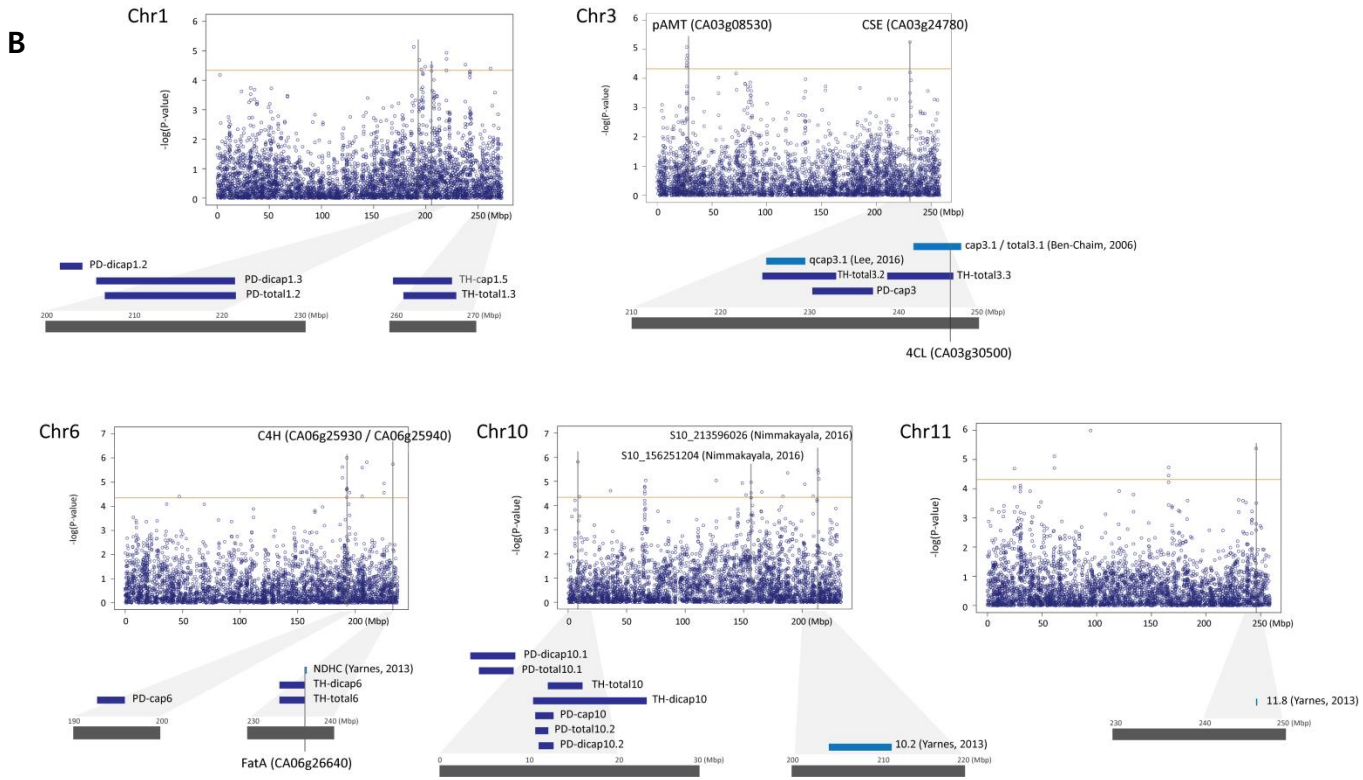


Figure I- 5. Comparison of QTL and SNPs associated with capsaicinoid contents. (A) Common QTL detected from both 'PD' and 'TH' RIL populations. The QTL marked in red are common to both populations. (B) Manhattan plot from GWAS and co-located QTL. The threshold of the -log(*P*) was 4.3. QTL and genes related to capsaicinoid biosynthesis are indicated.

Epistatic control of capsaicinoid content

In both ‘PD’ and ‘TH’ RILs, the distribution graphs of capsaicinoid content showed a positive skew (Figure I- 1A, B); there were more mildly pungent RILs than extremely pungent RILs. A skewed distribution indicates that there may be epistatic interactions between the QTL. Using multiple-interval mapping (MIM), epistatic effects between common QTL were detected (Table I- 8; Table I- 9). In ‘PD’ RILs, additive-by-additive epistases between *PD-cap1* and *PD-cap10* (capsaicin, 2011), *PD-dicap2.1* and *PD-dicap10.2* (dihydrocapsaicin, 2012a), *PD-dicap1.1* and *PD-dicap10.2* (dihydrocapsaicin, 2012b), and *PD-total1.1* and *PD-total10.2* (total capsaicinoid, 2011) were detected (Table I- 8). These individual QTL and the interactions between them could explain 17.5 to 45.4% of the variation in their respective capsaicinoid contents.

Additive-by-additive epistatic effects between the common QTL were detected in only one environment (2014) for ‘TH’ RILs, occurring between *TH-cap2.2* and *TH-cap4* for capsaicin content, and *TH-total2*, *TH-total3.2*, and *TH-total10* for total capsaicinoid content (Table I- 9). The effects of the individual QTL (*TH-total2*, *TH-total3.2*, and *TH-total10*) and their epistatic effects could explain 57.2% of the total capsaicinoid content variation in 2014.

Table I- 8. Epistatic effects of major QTL in 'PD' RILs

Trait	Year	QTL	R^2 (%)*	Total R^2 (%)
CAP	2011	*PD-cap1*	14.8	45.4
		PD-cap10	21.8	
		PD-cap1×PD-cap10	8.8	
DICAP	2012a	*PD-dicap2.1*	-0.4	17.5
		PD-dicap10.2	10	
		PD-dicap2.1×PD-dicap10.2	7.9	
	2012b	*PD-dicap1.1*	15.4	40.8
		PD-dicap10.2	22.4	
		PD-dicap1.1×PD-dicap10.2	3.0	
TOTAL	2011	*PD-total1.1*	11.3	41.9
		PD-total10.2	17.2	
		PD-total1.1×PD-total10.2	13.4	

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

CAP, capsaicin; DICAP, dihydrocapsaicin; TOTAL, total capsaicinoid

Table I- 9. Epistatic effects of major QTL in 'TH' RILs

Trait	Year	QTL	R^2 (%)*	Total R^2 (%)
CAP	2014	*TH-cap2.2*	13.7	20.4
		TH-cap4	5.3	
		TH-cap2.2×TH-cap4	1.4	
TOTAL	2014	*TH-total2*	37.3	57.2
		TH-total3.2	-0.4	
		TH-total10	18.5	
		TH-total2×TH-total3.2	-0.6	
		TH-total2×TH-total10	2.3	
		TH-total3.2×TH-total10	0.1	

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

CAP, capsaicin; TOTAL, total capsaicinoid

SNPs and haplotype blocks of GWAS population

To validate the QTL detected from the biparental populations, a GWAS study for capsaicinoid content was performed using 208 *C. annuum*-clade accessions, including 145 from *C. annuum*, 42 from *C. chinense*, and 21 from *C. frutescens* (Table I- 1). The accessions were genotyped using the GBS method; GBS libraries were constructed using two restriction enzyme sets, PstI/MseI and EcoRI/MseI, from which 14,461 and 119,710 significant SNPs were detected, respectively (data not shown). Even distribution of SNPs was detected using EcoRI/MseI than PstI/MseI (Figure I- 3C, D). After filtering the SNPs for minor allele frequencies and calling rate, a total of 109,610 SNPs were selected for further study (Table I- 4). In contrast to 'PD' and 'TH' RILs, the SNPs of the GWAS population were relatively evenly distributed, with an average distance between SNPs of 25,093 bp (Figure I- 3E). Using these SNPs, the *Capsicum* accessions were divided into three subgroups using a principal component analysis (PCA) (Figure I- 6A) and a phylogenetic analysis (Figure I- 6B). These analyses showed that the accessions of the GWAS population were grouped according to their expected species groups, *C. annuum*, *C. chinense*, and *C. frutescens*. Eight accessions were not included in any of the subgroups, and four accessions were in different subgroups from the species that was described in their passport data in Genebank where collecting *Capsaicum* germplasm. The population structure determined from the PCA was applied for the GWAS.

Haplotype blocks were calculated in each chromosome using less stringent

options than the default settings. About 90% of SNPs were grouped into 5,513 blocks, and each block contained 3–138 SNPs, with an average of 18 SNPs (Table I- 10. The block size varied between 3 bp and 2 Mbp, with average block sizes of 567, 438, 547, 395, 465, 526, 434, 225, 506, 505, 535, and 370 kbp for the twelve chromosomes, respectively. Genome-wide, the average haplotype block size was 409 kbp, which was larger than the average distance between the SNPs used for the GWAS.

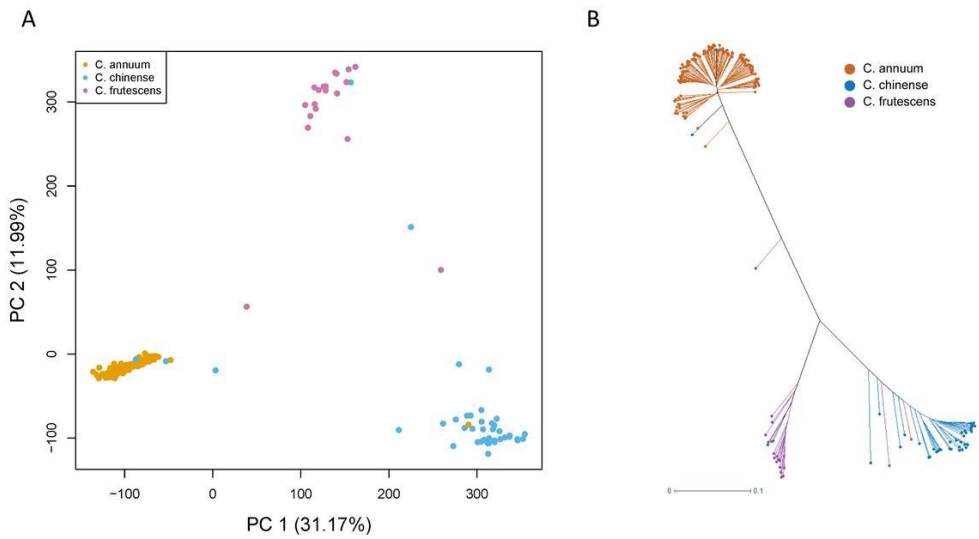


Figure I- 6. Population structure of the GWAS population, with a principal component analysis (A) and a phylogenetic tree (B) determined from 109,610 SNPs. Dark orange, blue, and purple colors indicate *C. annuum*, *C. chinense*, and *C. frutescens*, respectively.

Table I- 10. Haplotype block estimated by genotyping-by-sequencing of the GWAS population

Chr	Block size (Kbp)			Number of blocks
	Minimum	Maximum	Average	
1	0.006	1999.59	394.5	567
2	0.006	2000	293.5	438
3	0.006	1999.98	381.5	547
4	0.011	1999.92	470.4	395
5	0.013	1999.98	400.3	465
6	0.003	1999.51	358.8	526
7	0.007	1999.97	442.7	434
8	0.013	2000	537.8	225
9	0.005	1999.41	416.9	506
10	0.003	1999.91	384.0	505
11	0.006	1999.94	402.8	535
12	0.005	1999.99	540.0	370
Total	0.003	2000	409.0	5513

GWAS for capsaicinoid content

For the 208 accessions comprising the GWAS population, the capsaicinoid content was evaluated from freeze-dried whole fruits. Their total capsaicinoid contents varied from 2 to 16,082 μg/g DW in the whole fruit (Figure I- 1C), with ten accessions containing less than 10 μg/g DW. Of the 20 accessions with the highest capsaicinoid contents, eight were *C. frutescens*, seven were *C. chinense*, and five accessions were *C. annuum*. The three most pungent accessions, 'Habanero', '9007', and 'Spain 5', were all accessions of *C. chinense*, which is well known for its pungency (Bosland and Baral, 2007; Bosland et al., 2012; Canto-Flick et al., 2008).

We analyzed the association of SNPs with the capsaicin, dihydrocapsaicin, and total capsaicinoid contents using GWAS. A total of 99 SNPs were associated with capsaicin, 9 were linked to dihydrocapsaicin, and 42 SNPs were associated with the total capsaicinoid content; however, the SNPs associated with the dihydrocapsaicin and total capsaicinoid contents did not exceed the false discovery rate (FDR) threshold, so only the 99 capsaicin-associated SNPs were considered significant. These 99 SNPs were grouped into 69 genomic regions using a haplotype block estimation. Using gene annotation data and SNPs located on haplotype blocks, 213 genes located on 69 associated regions were found and their functions were predicted (data not shown).

Among 69 regions, 10 regions on chromosomes 1, 3, 6, and 10 were co-located with QTL detected in the present study (Figure I- 5B), while four regions

on chromosomes 10 and 11 were linked to previously detected QTL and SNPs (Nimmakayala et al., 2016; Yarnes et al., 2013). On chromosome 1, three regions incorporating six SNPs were co-located with *PD-dicap1.3* and *PD-total1.2*, while another SNP was co-located with *TH-cap1.5* and *TH-total1.3*. One capsaicin-associated SNP was detected between 230.53 and 231.21 Mbp on chromosome 3, which corresponded to the location of the *TH-total3.2* and *PD-cap3* QTL. On chromosome 6, three regions containing eight SNPs were detected together with *PD-cap6*. Two regions on chromosome 10, each containing a single SNP, were also validated by QTL; SNP 10_8241800 was co-located with *PD-dicap10.1*, *PD-total10.1*. And the other SNP 10_9608580 was co-located with *TH-total10*, *TH-dicap10*, *PD-cap10*, *PD-dicap10.2*, and *PD-total10.2*. A total of 55 new regions associated with capsaicin contents were detected, with an average –log(*P*) value of 4.84.

Candidate gene prediction for QTL controlling capsaicinoid content

From the QTL mapping and GWAS, we were able to identify candidate genes involved in the capsaicinoid biosynthesis pathway. Among the candidate genes from GWAS, two genes expected to function in the capsaicinoid biosynthesis pathway were identified (Table I- 11). *pAMT*, located on chromosome 3, was strongly linked to seven significantly associating SNPs. pAMT mediates the formation of vanillylamine, which is the final step of the phenylpropanoid pathway for the biosynthesis of capsaicinoids (Lang et al., 2009); therefore, *pAMT* is a plausible candidate gene for the control of capsaicinoid content. On chromosome 6, *cinnamate 4-hydroxylase* (*C4H*) was located around 400 kbp away from SNP 6_232803485. C4H is also involved in the phenylpropanoid pathway, and has catalytic activity in the biosynthesis of coumarate from cinnamate (Curry et al., 1999; Mazourek et al., 2009). The comparison of the QTL mapping and GWAS results led to the identification of *caffeoyl shikimate esterase* (*CSE*), located on chromosome 3, as a candidate gene. Two QTL from this study, one QTL from Lee (2016b), and one SNP from the GWAS were linked to *CSE* (Table I- 11; Figure I-5B). Even though the role of *CSE* in the capsaicinoid biosynthesis pathway is unknown, CSE is known to hydrolyze caffeoyl shikimate, which is an intermediate of phenylpropanoid pathway (Vanholme et al., 2013). From the QTL mapping results, *TH-total3.3*, located at 239.4–246.9 Mbp on chromosome 3, was associated

with the gene encoding 4-coumaroyl-CoA ligase (4CL), which was previously linked to other capsaicinoid QTL, *cap3.1* and *total3.1* (Ben-Chaim et al., 2006). Another gene, encoding acyl-ACP thioesterase (*FatA*), functions in the fatty acid biosynthesis pathway, and was associated with the QTL *TH-dicap6* and *TH-total6*, as well as QTL *6.8*, which was previously found to be linked to nordihydrocapsaicin content (Yarnes et al., 2013). In summary, we propose five candidate genes for the control of capsaicinoid content, *pAMT*, *C4H*, *CSE*, *4CL*, and *FatA*, each with known or potential functions in the capsaicinoid biosynthesis pathway.

To compare the capsaicinoid contents of plants to their genotypes at these candidate genes, individual plants from the RIL and GWAS populations were grouped by their genotypes at *pAMT*, *C4H*, *CSE*, *4CL*, and *FatA*. Bin markers located within 1 Mbp of the candidate genes were used, due to the lack of genotype information for the candidate genes themselves. For 'PD' RILs, PD3-bin56, PD6-bin174, PD3-bin200, PD3-bin216, PD6-bin179 were used, while TH3-bin1, TH6-bin91, TH3-bin98, TH3-bin112, and TH6-bin93 were used for 'TH' RILs. For the GWAS population, the genotypes of SNPs 3_26745367, 6_232803485, and 3_230603011, which showed association with capsaicin content, were used to draw box plots for *pAMT*, *C4H*, and *CSE*, respectively. The accessions were also separated by their genotypes at SNPs 3_246744919 and 6_234337365, which were not associated with capsaicin content, but were the closest SNPs to *4CL* and *FatA*. In 'PD' RILs, only *CSE* was associated with a significant difference in capsaicin

content, whereas *C4H*, *FatA*, and *4CL* were associated with significant differences in capsaicin content in 'TH' RILs (Figure I- 7). In the GWAS population, differences in the genotypes of all five candidate genes led to highly significant differences in capsaicinoid contents.

Table I- 11. Candidate genes related to capsaicinoid biosynthesis and their associated QTL or SNPs

Candidate gene (CDS)	PD RIL	TH RIL	GWAS population	Previous study
pAMT (CA03g08530)	-	-	3_26745322, 3_26745328, 3_26745367, 3_26770544, 3_26770554, 3_26770560, 3_27438287	-
C4H (CA06g25930 / CA06g25940)	-	-	6_232803485	-
CSE (CA03g24780)	*PD-cap3*	*TH-total3.2*	3_230603011	*qcap3.1* (Lee et al., 2016b)
4CL (CA03g30500)	-	*TH-total3.3*	-	*cap3.1*, *total3.1* (Ben-Chaim et al., 2006)
FatA (CA06g26640)	-	*TH-dicap6*, *TH-total6*	-	*6.8* (Yarnes et al., 2013)

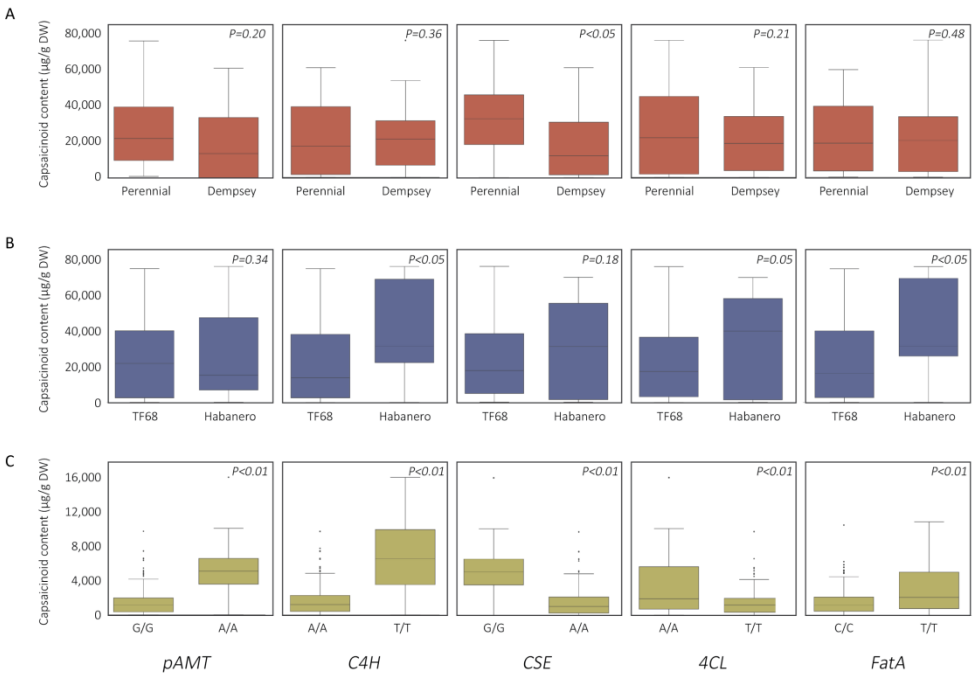


Figure I- 7. Box plots of capsaicinoid contents regulated by the five candidate genes. Total capsaicinoid contents resulting from the 'PD' RIL (A) and 'TH' RIL (B) parental genotypes related to each candidate gene in plants grown in 2012b and 2014, respectively. (C) Total capsaicinoid content resulting from reference ('CM334') and alternative genotypes related to each candidate gene in plants of the GWAS population. RILs and accessions were classified by the genotype of the most closely linked marker to the target genes. *P*-values were calculated from an equal variance t-test for 'PD' and 'TH' RIL', and from an unequal variance t-test for the GWAS population.

DISCUSSION

Global comparison of QTL for capsaicinoid content

Using the *C. annuum* 'CM334' reference genome, we compared the physical locations of capsaicinoid-related QTL detected in multiple analyses and capsaicinoid content-associating SNPs from GWAS. The QTL from our research were also compared with those detected in other studies (Ben-Chaim et al., 2006; Blum et al., 2003; Lee et al., 2016b; Nimmakayala et al., 2016; Yarnes et al., 2013). Before now, the comparison of QTL from different studies was not feasible, due to the lack of a reference genome or common markers used for the genetic maps. Here, we used the primer sequences linked to the QTL to BLAST the 'CM334' reference genome (v1.55), or if primer sequences were not publicly available, the closest marker with information was used. In total, eight QTL located on chromosomes 2, 3, and 6 were validated. A shared QTL on chromosome 3, which contains *PD-cap3* and *TH-total3.2*, was co-located with *qcap3.1*, which was previously identified in the F_2 population derived from a *C. annuum* 'NB1' × *C. chinense* 'Bhut Jolokia' cross (Lee et al., 2016b). *TH-total3.3* was also linked to *cap3.1* and *total3.1* detected in the *C. annuum* 'NuMex Rnaky' × *C. frutescens* 'BG2814-6' $F_{2:3}$ population (Ben-Chaim et al., 2006). *TH-dicap6* and *TH-total6* were co-located with a QTL controlling nordihydrocapsaicin detected in a *C. frutescens* '2814-6' × *C. annuum* 'NuMex Rnaky' RIL (Yarnes et al., 2013). The

Pun1 locus was detected from 'PD' RILs in the QTL *PD-cap2*, *PD-dicap2.1*, and *PD-total2*. We expected that *Pun1* would be detected due to the segregation of pungency in some RILs having heterozygous loci. Previously, *cap* and *cap7.2*, located on chromosome 7, were detected from two different populations and were thought to constitute a major QTL (Ben-Chaim et al., 2006; Blum et al., 2003). This locus was expected to 202-203 Mbp on chromosome 7 using linked marker information, but no QTL or associated SNPs were located in this locus from our QTL or GWAS analyses. Ben-Chaim et al. and Blum and colleagues used *C. frutescens* as a parent to develop interspecific populations. The genetic diversity of this species is unlikely to be represented in our intraspecific population derived from *C. annuum* ('PD' RILs) or our interspecific population generated by a cross between *C. annuum* and *C. chinense* ('TH' RILs). The 21 *C. frutescens* accessions used in the GWAS population may therefore not be sufficient to enable our detection of the *cap/cap7.2* locus in the present study. The other QTL detected only one population or environment might be the minor QTL and showed the large environmental effect on identification of QTL. To overcome difficulties for detecting minor QTL and figure out the significant QTL affect multiple pepper accessions, we compared the results with GWAS.

We could compare the QTL with the capsaicin-associated SNPs detected from GWAS, but not with the dihydrocapsaicin- and total capsaicinoid-associated SNPs, as their associations were not significant. From the raw GWAS results, nine and 42 SNPs were associated with the dihydrocapsaicin and total capsaicinoid contents,

respectively. Among them, five and 36 SNPs were also detected as being associated with capsaicin content, and only these reached the FDR threshold for this association. Several GWAS and QTL mapping studies have demonstrated that one locus can control multiple highly correlated traits (Bauchet et al., 2017; Ben-Chaim et al., 2006; Crowell et al., 2016; Han et al., 2016; Wang et al., 2011). We detected a high correlation between capsaicin, dihydrocapsaicn, and total capsaicinoid contents; therefore, it is possible that some SNPs can affect these traits simultaneously. If the capsaicinoid contents of the GWAS population were evaluated repeatedly, we would expect to identify more significantly associated SNPs, enabling the validation of more QTL using GWAS.

Candidate genes controlling capsaicinoid content

Using a genome-based approach, we found five candidate genes for controlling capsaicinoid contents in pepper: *pAMT*, *C4H*, *4CL* and *CSE* from the phenylpropanoid pathway, and *FatA*, from the fatty acid pathway (Lang et al., 2009; Mazourek et al., 2009; Stewart et al., 2005). In plants, the phenylpropanoid pathway is known to be related to the biosynthesis of amino acids and diverse secondary metabolites (Fraser and Chapple, 2011; Vogt, 2010), and its involvement in the production of capsaicinoids was predicted based on intermediates and genes identified in other plant species (Curry et al., 1999; Mazourek et al., 2009). The expression of genes in the phenylpropanoid pathway was also found to be related to the capsaicinoid content of *Capsicum* fruits (Arce-Rodriguez and Ochoa-Alejo,

2017; Curry et al., 1999; Kim et al., 2009; Kim et al., 2014; Liu et al., 2012; Stewart et al., 2005; Zhang et al., 2016).

pAMT is the final enzyme of the phenylpropanoid pathway, and mediates the formation of vanillylamine (Curry et al., 1999; Lang et al., 2009). The loss of *pAMT* function results in the biosynthesis of capsinoid, which has little to no pungency, rather than the pungent capsaicinoid (Jang et al., 2015; Lang et al., 2009; Tanaka et al., 2010a; Tanaka et al., 2010b; Tanaka et al., 2015). In pungent peppers, *pAMT* was expressed only in the fruits, and its expression level in the developmental stages was positively correlated with that of other capsaicinoid biosynthesis genes, c*affeoyl-CoA 3-O-methyltransferase*, *Kas*, *branched-chain amino acid aminotransferase*, and *Pun1*, as well as *C4H*, another candidate gene identified in our current study (Arce-Rodriguez and Ochoa-Alejo, 2017; Sarpras et al., 2016; Zhang et al., 2016). C4H functions at the endoplasmic reticulum to catalyze the reaction from cinnamate to coumarate (Mazourek et al., 2009), and another candidate gene, *4CL*, encodes the enzyme that acts immediately after C4H in this pathway (Mazourek et al., 2009). Another candidate gene, *CSE*, is located on chromosome 3 in a region that was detected in both the QTL mapping and the GWAS. CSE functions in the lignin biosynthetic pathway in the *Arabidopsis*, *Medicago*, *Populus*, and *Panicum* genera, while in maize it has only a slight esterase activity (Ha et al., 2016; Vanholme et al., 2013). With the exception of this activity, its other functions in the Solanaceae family are unknown; therefore, further genetic studies are needed to elucidate its activities in *Capsicum*.

Compared with the phenylpropanoid pathway, not much is known about the fatty acid pathway because branched-chain fatty acids are not common metabolites in most plant species outside of the Solanaceae family (Mazourek et al., 2009). FatA functions at the last stage of the fatty acid pathway and regulates the chain length of the fatty acids (Aluru et al., 2003). The expression of *FatA* during fruit maturation is correlated with capsaicinoid content (Aluru et al., 2003; Keyhaninejad et al., 2014; Zhang et al., 2016). Various transcription factors have been suggested to control the expression of the capsaicinoid biosynthesis genes, including *Erf*, *Jerf*, and *CaMYB31* (Arce-Rodriguez and Ochoa-Alejo, 2017; Keyhaninejad et al., 2014); however, none of them were detected in our QTL and GWAS analyses.

The association of *pAMT* with pungency was detected only from the GWAS population, in which four *C. chinense* and fourteen *C. frutescens* accessions had a minor allele at SNP 3_26745367 that was linked to *pAMT* (data not shown). All accessions with minor alleles at SNPs linked to *C4H* and *CSE* were also *C. chinense*. In previous studies, *C. chinense* has been reported to have diverse *pamt* non-functional alleles affecting the levels of various capsaicinoids and capsinoids (Jang et al., 2015; Koeda et al., 2014; Tanaka et al., 2010a; Tanaka et al., 2010b; Tanaka et al., 2015). Effects for *4CL* and *FatA* were detected only from the 'TH' RIL analysis in our study; even though no significant SNPs were identified in these regions from the GWAS, the capsaicinoid contents of plants with different alleles at these regions showed a varied distribution (Figure I- 7C). For *4CL*, the majority of

the minor alleles were detected from *C. chinense* and *C. frutescens* accessions in the GWAS population. Species-specific genetic variation can be used to breed highly pungent pepper cultivars by introgressing the candidate genes of *C. chinense* or *C. frutescens* into *C. annuum*.

SNP detection by GBS for QTL study

We used the double-digestion method to make GBS libraries for the 'TH' RIL and GWAS populations. PstI/MseI enzymes were used to digest both populations, with the additional use of EcoRI/MseI enzymes for the GWAS population. An *in silico* analysis revealed that approximately three times more effective cut sites (100–600 bp length fragments) were predicted when using the EcoRI/MseI enzymes to digest the 'CM334' reference genome than when PstI/MseI were used (data not shown), and that EcoRI/MseI made 30 times more cut sites in the regions where few SNPs were detected using PstI/MseI. The use of both sets of enzymes to construct GBS libraries therefore enabled the acquisition of sufficient SNPs for the GWAS.

The percentage of SNPs located in genic regions was highest in 'TH' RILs, and the SNP distribution graph revealed that they were concentrated in euchromatin regions (Table I- 12; Figure I- 3A). A similar biased distribution of SNPs was observed in soybean GBS results generated using ApeKI (Sonah et al., 2013). Sonah and colleagues reported the ratio of SNPs located in soybean genic regions was as high as 39.5%, which was very similar to the proportion in the 'TH'

RIL population used PstI/MseI. Like ApeKI, PstI has partial methylation sensitivity in plants (Elshire et al., 2011; Pootakham et al., 2016; Truong et al., 2012), which could result in the identification of high SNP densities in genic regions. In the 'PD' RIL and GWAS populations however, only 1.2% and 3.2% of SNPs, respectively, were located in genic regions. Of the two enzyme sets, libraries using EcoRI/MseI showed more number of SNPs than PstI/MseI in the GWAS population and might bring even distribution of SNPs. The large number of SNPs in our study demonstrates the effectiveness of using two enzyme sets for GBS. This approach could reduce the costs for genotyping and increase the number of effective SNPs in comparison with the use of one enzyme or one enzyme set.

In conclusion, we demonstrated that analysis using genome-based QTL mapping and GWAS is a useful tool for the identification of candidate genes associated with capsaicinoid content, which was not easily achieved in previous studies using low-density genetic maps. The candidate genes and their associated SNPs detected here will be helpful to improve our understanding of capsaicinoid biosynthesis, and could be applied to the breeding of high-pungency peppers. We also confirmed the minor effects of each locus and the epistatic effects between QTL, revealing that multiple markers should be used together for marker-assisted selection.

Table I- 12. Distribution of SNPs in genic and intergenic regions

Population	Number of SNPs (%)		
	Genic	Intergenic	Total
PD RIL	16,581 (1.2)	1,414,633 (98.8)	1,431,214
TH RIL	3,307 (38.5)	5,280 (61.5)	8,587
GWAS population	3,453 (3.2)	106,157 (96.8)	109,610

REFERENCES

Aluru, M.R., Mazourek, M., Landry, L.G., Curry, J., Jahn, M. and O'Connell, M.A. (2003) Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in *Capsicum* fruit. *J. Exp. Bot.* **54**, 1655-1664.

Arce-Rodriguez, M.L. and Ochoa-Alejo, N. (2017) An R2R3-MYB transcription factor regulates capsaicinoid biosynthesis. *Plant Physiol.* **174**, 1359-1370.

Aza-Gonzalez, C., Nunez-Palenius, H.G. and Ochoa-Alejo, N. (2011) Molecular biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). *Plant Cell Rep.* **30**, 695-706.

Bauchet, G., Grenier, S., Samson, N., Segura, V., Kende, A., Beekwilder, J., Cankar, K., Gallois, J.L., Gricourt, J., Bonnet, J., Baxter, C., Grivet, L. and Causse, M. (2017) Identification of major loci and genomic regions controlling acid and volatile content in tomato fruit: implications for flavor improvement. *New Phytol.* **215**, 624-641.

Ben-Chaim, A., Borovsky, Y., Falise, M., Mazourek, M., Kang, B.C., Paran, I. and Jahn, M. (2006) QTL analysis for capsaicinoid content in *Capsicum*. *Theor. Appl. Genet.* **113**, 1481-1490.

Blum, E., Mazourek, M., O'Connell, M., Curry, J., Thorup, T., Liu, K., Jahn, M. and Paran, I. (2003) Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. *Theor. Appl. Genet.* **108**, 79-86.

Bosland, P.W. and Baral, J.B. (2007) 'Bhut Jolokia'—the world's hottest known chile pepper is a putative naturally occurring interspecific hybrid. *Hortscience* **42**, 222-224.

Bosland, P.W., Coon, D. and Reeves, G. (2012) 'Trinidad Moruga Scorpion' pepper is the world's hottest measured chile pepper at more than two million scoville

heat units. *HortTechnology* **22**, 534-538.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J. and Roux, F. (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* **6**, e1000940.

Canto-Flick, A., Balam-Uc, E., Bello-Bello, J.J., Lecona-Guzman, C., Solis-Marroquin, D., Aviles-Vinas, A., Gomez-Uc, E., Lopez-Puc, G. and Santana-Buzzy, N. (2008) Capsaicinoids content in Habanero pepper (*Capsicum chinense* Jacq.): hottest known cultivars. *Hortscience* **43**, 1344-1349.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7.

Collins, M.D., Wasmund, L.M. and Bosland, P.W. (1995) Improved method for quantifying capsaicinoids in *Capsicum* using highperformance liquid chromatography. *Hortscience* **30**, 137-139.

Crowell, S., Korniliev, P., Falcao, A., Ismail, A., Gregorio, G., Mezey, J. and McCouch, S. (2016) Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nat. Commun.* **7**, 10527.

Curry, J., Aluru, M., Mendoza, M., Nevarez, J., Melendrez, M. and O'Connell, M.A. (1999) Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp. *Plant Sci.* **148**, 47-57.

De Givry, S., Bouchez, M., Chabrier, P., Milan, D. and Schiex, T. (2005) CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* **21**, 1703-1704.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. and Daly, M.J. (2011) A framework for variation discovery and

genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498.

Doyle, J.J. and Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11-15.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379.

Fraser, C.M. and Chapple, C. (2011) The phenylpropanoid pathway in *Arabidopsis*. *Arabidopsis Book* **9**, e0152.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J. and Conesa, A. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420-3435.

Ha, C.M., Escamilla-Trevino, L., Yarce, J.C., Kim, H., Ralph, J., Chen, F. and Dixon, R.A. (2016) An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *Plant J.* **86**, 363-375.

Han, K., Jeong, H.-J., Sung, J., Keum, Y.S., Cho, M.C., Kim, J.H., Kwon, J.K., Kim, B.D. and Kang, B.C. (2013) Biosynthesis of capsinoid is controlled by the *Pun1* locus in pepper. *Mol. Breed.* **31**, 537-548.

Han, K., Jeong, H.J., Yang, H.B., Kang, S.M., Kwon, J.K., Kim, S., Choi, D. and Kang, B.C. (2016) An ultra-high-density bin map facilitates high-throughput QTL mapping of horticultural traits in pepper (*Capsicum annuum*). *DNA Res.* **23**, 81-91.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T. and Han, B. (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**, 1068-1076.

Jang, S., Han, K., Jo, Y.D., Jeong, H.J., Siddique, M.I. and Kang, B.C. (2015) Substitution of a dysfunctional *pAMT* allele results in low-pungency but high levels of capsinoid in *Capsicum chinense* 'Habanero'. *Plant Breed. Biotechnol.*

3, 119-128.

Jeong, H.-S., Jang, S., Han, K., Kwon, J.K. and Kang, B.C. (2015) Marker-assisted backcross breeding for development of pepper varieties (*Capsicum annuum*) containing capsinoids. *Mol. Breed.* **35**, 226.

Jo, J., Purushotham, P.M., Han, K., Lee, H.R., Nah, G. and Kang, B.C. (2017) Development of a Genetic Map for Onion (Allium cepa L.) Using reference-free genotyping-by-sequencing and SNP assays. *Front. Plant Sci*. **8**, 1606.

Keyhaninejad, N., Curry, J., Romero, J. and O'Connell, M.A. (2014) Fruit specific variability in capsaicinoid accumulation and transcription of structural and regulatory genes in *Capsicum* fruit. *Plant Sci.* **215-216**, 59-68.

Kim, J.S., Park, M., Lee, D.J. and Kim, B.D. (2009) Characterization of putative capsaicin synthase promoter activity. *Mol. Cells* **28**, 331-339.

Kim, O.R., Cho, M.C., Kim, B.D. and Huh, J.H. (2010) A splicing mutation in the gene encoding phytoene synthase causes orange coloration in Habanero pepper fruits. *Mol. Cells* **30**, 569-574.

Kim, S., Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A., Seo, E., Choi, J., Cheong, K., Kim, K.T., Jung, K., Lee, G.W., Oh, S.K., Bae, C., Kim, S.B., Lee, H.Y., Kim, S.Y., Kim, M.S., Kang, B.C., Jo, Y.D., Yang, H.B., Jeong, H.J., Kang, W.H., Kwon, J.K., Shin, C., Lim, J.Y., Park, J.H., Huh, J.H., Kim, J.S., Kim, B.D., Cohen, O., Paran, I., Suh, M.C., Lee, S.B., Kim, Y.K., Shin, Y., Noh, S.J., Park, J., Seo, Y.S., Kwon, S.Y., Kim, H.A., Park, J.M., Kim, H.J., Choi, S.B., Bosland, P.W., Reeves, G., Jo, S.H., Lee, B.W., Cho, H.T., Choi, H.S., Lee, M.S., Yu, Y., Do Choi, Y., Park, B.S., van Deynze, A., Ashrafi, H., Hill, T., Kim, W.T., Pai, H.S., Ahn, H.K., Yeam, I., Giovannoni, J.J., Rose, J.K., Sorensen, I., Lee, S.J., Kim, R.W., Choi, I.Y., Choi, B.S., Lim, J.S., Lee, Y.H. and Choi, D. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **46**, 270-278.

Koeda, S., Sato, K., Tomi, K., Tanaka, Y., Takisawa, R., Hosokawa, M., Doi, M., Nakazaki, T. and Kitajima, A. (2014) Analysis of Non-pungency, aroma, and

origin of a *Capsicum chinense* cultivar from a Caribbean island. *J. Jpn. Soc. Hortic. Sci.* **83**, 244-251.

Korte, A. and Farlow, A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29.

Lang, Y., Kisaka, H., Sugiyama, R., Nomura, K., Morita, A., Watanabe, T., Tanaka, Y., Yazawa, S. and Miwa, T. (2009) Functional loss of *pAMT* results in biosynthesis of capsinoids, capsaicinoid analogs, in *Capsicum annuum* cv. CH-19 Sweet. *Plant J.* **59**, 953-961.

Lee, H.Y., Ro, N.Y., Jeong, H.J., Kwon, J.K., Jo, J., Ha, Y., Jung, A., Han, J.W., Venkatesh, J. and Kang, B.C. (2016a) Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. *BMC Genet.* **17**, 142.

Lee, J., Park, S.J., Hong, S.C., Han, J.-H., Choi, D., Yoon, J.B. and Havey, M. (2016b) QTL mapping for capsaicin and dihydrocapsaicin content in a population of *Capsicum annuum* 'NB1' × *Capsicum chinense* 'Bhut Jolokia'. *Plant Breed.* **135**, 376-383.

Lee, J.-H., An, J.-T., Siddique, M.I., Han, K., Choi, S., Kwon, J.K. and Kang, B.C. (2017) Identification and molecular genetic mapping of Chili veinal mottle virus (ChiVMV) resistance genes in pepper (*Capsicum annuum*). *Mol. Breed.* **37**, 121.

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

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z. (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397-2399.

Liu, S., Chen, C., Chen, G., Cao, B., Chen, Q. and Lei, J. (2012) RNA-sequencing tag profiling of the placenta and pericarp of pungent pepper provides robust candidates contributing to capsaicinoid biosynthesis. *Plant Cell Tiss. Org. Cult.* **110**, 111-121.

Luo, X.J., Peng, J. and Li, Y.J. (2011) Recent advances in the study on

capsaicinoids and capsinoids. *Eur. J. Pharmacol.* **650**, 1-7.

Mazourek, M., Pujar, A., Borovsky, Y., Paran, I., Mueller, L. and Jahn, M.M. (2009) A dynamic interface for capsaicinoid systems biology. *Plant Physiol.* **150**, 1806-1821.

Navarro, J.A.R., Willcox, M., Burgueno, J., Romay, C., Swarts, K., Trachsel, S., Preciado, E., Terron, A., Delgado, H.V., Vidal, V., Ortega, A., Banda, A.E., Montiel, N.O., Ortiz-Monasterio, I., Vicente, F.S., Espinoza, A.G., Atlin, G., Wenzl, P., Hearne, S. and Buckler, E.S. (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* **49**, 476-480.

Nimmakayala, P., Abburi, V.L., Abburi, L., Alaparthi, S.B., Cantrell, R., Park, M., Choi, D., Hankins, G., Malkaram, S. and Reddy, U.K. (2014) Linkage disequilibrium and population-structure analysis among *Capsicum annuum* L. cultivars for use in association mapping. *Mol. Genet. Genomics* **289**, 513-521.

Nimmakayala, P., Abburi, V.L., Saminathan, T., Alaparthi, S.B., Almeida, A., Davenport, B., Nadimi, M., Davidson, J., Tonapi, K., Yadav, L., Malkaram, S., Vajja, G., Hankins, G., Harris, R., Park, M., Choi, D., Stommel, J. and Reddy, U.K. (2016) Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. *Sci. Rep.* **6**, 38081.

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

Rodríguez-Maza, M.J., Garcés-Claver, A., Park, S.W., Kang, B.C. and Arnedo-Andrés, M.S. (2012) A versatile PCR marker for pungency in *Capsicum* spp. *Mol. Breed.* **30**, 889-898.

Sakiroglu, M. and Brummer, E.C. (2017) Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. *Theor. Appl.*

Genet. **130**, 261-268.

Sallam, A., Arbaoui, M., El-Esawi, M., Abshire, N. and Martsch, R. (2016) Identification and verification of QTL associated with frost tolerance using linkage mapping and GWAS in winter faba bean. *Front. Plant Sci.* **7**, 1098.

Sanatombi, K. and Sharma, G.J. (2008) Capsaicin content and pungency of different *Capsicum* spp. cultivars. *Not. Bot. Horti. Agrobot. Cluj Napoca* **36**, 89-90.

Sarpras, M., Gaur, R., Sharma, V., Chhapekar, S.S., Das, J., Kumar, A., Yadava, S.K., Nitin, M., Brahma, V., Abraham, S.K. and Ramchiary, N. (2016) Comparative analysis of fruit metabolites and pungency candidate genes expression between Bhut Jolokia and other *Capsicum* species. *PLoS ONE* **11**, e0167791.

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

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

Stellari, G.M., Mazourek, M. and Jahn, M.M. (2010) Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. *Heredity* **104**, 460-471.

Stewart, C., Jr., Kang, B.C., Liu, K., Mazourek, M., Moore, S.L., Yoo, E.Y., Kim, B.D., Paran, I. and Jahn, M.M. (2005) The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant J* **42**, 675-688.

Stewart, C., Jr., Mazourek, M., Stellari, G.M., O'Connell, M. and Jahn, M. (2007) Genetic control of pungency in *C. chinense* via the *Pun1* locus. *J. Exp. Bot.* **58**, 979-991.

Swarts, K., Li, H., Romero Navarro, J.A., An, D., Romay, M.C., Hearne, S., Acharya, C., Glaubitz, J.C., Mitchell, S., Elshire, R.J., Buckler, E.S. and Bradbury, P.J. (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome* **7**, 1-12.

Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T. and Yazawa, S. (2010a) Newly mutated *putative-aminotransferase* in nonpungent pepper (*Capsicum annuum*) results in biosynthesis of capsinoids, capsaicinoid analogues. *J. Agric. Food Chem.* **58**, 1761-1767.

Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T. and Yazawa, S. (2010b) Novel loss-of-function *putative aminotransferase* alleles cause biosynthesis of capsinoids, nonpungent capsaicinoid analogues, in mildly pungent chili peppers (*Capsicum chinense*). *J. Agric. Food Chem.* **58**, 11762-11767.

Tanaka, Y., Sonoyama, T., Muraga, Y., Koeda, S., Goto, T., Yoshida, Y. and Yasuba, K. (2015) Multiple loss-of-function putative aminotransferase alleles contribute to low pungency and capsinoid biosynthesis in *Capsicum chinense*. *Mol. Breed.* **35**, 142.

Tanaka, Y., Yoneda, H., Hosokawa, M., Miwa, T. and Yazawa, S. (2014) Application of marker-assisted selection in breeding of a new fresh pepper cultivar (*Capsicum annuum*) containing capsinoids, low-pungent capsaicinoid analogs. *Sci. Hort.* **165**, 242-245.

Tewksbury, J.J., Levey, D.J., Huizinga, M., Haak, D.C. and Traveset, A. (2008) Costs and benefits of capsaicin-mediated control of gut retention in dispersers of wild chilies. *Ecology* **89**, 107-117.

Tewksbury, J.J. and Nabhan, G.P. (2001) Directed deterrence by capsaicin in chillies. *Nature* **412**, 403-404.

Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R., McMullen, M.D., Holland, J.B. and Buckler, E.S. (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* **43**, 159-162.

Truong, H.T., Ramos, A.M., Yalcin, F., Ruiter, M.d., Poel, H.J.A.v.d., Huvenaars, K.H.J., Hogers, R.C.J., Enckevort, L.J.G.v., Janssen, A., Orsouw, N.J.v. and Eijk, M.J.T.v. (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. *PLoS ONE* **7**, e37565.

Upadhyaya, H.D., Bajaj, D., Narnoliya, L., Das, S., Kumar, V., Gowda, C.L., Sharma, S., Tyagi, A.K. and Parida, S.K. (2016) Genome-wide scans for delineation of candidate genes regulating seed-protein content in chickpea. *Front. Plant Sci.* **7**, 302.

Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., Kim, H., Cross, J., Morreel, K., Araujo, P., Welsh, L., Haustraete, J., McClellan, C., Vanholme, B., Ralph, J., Simpson, G.G., Halpin, C. and Boerjan, W. (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* **341**, 1103-1106.

Vogt, T. (2010) Phenylpropanoid biosynthesis. *Mol. Plant* **3**, 2-20.

Voorrips, R.E. (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77-78.

Wang, L., Wang, A., Huang, X., Zhao, Q., Dong, G., Qian, Q., Sang, T. and Han, B. (2011) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. *Theor. Appl. Genet.* **122**, 327-340.

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

Yarnes, S.C., Ashrafi, H., Reyes-Chin-Wo, S., Hill, T.A., Stoffel, K.M. and Van Deynze, A. (2013) Identification of QTLs for capsaicinoids, fruit quality, and plant architecture-related traits in an interspecific *Capsicum* RIL population. *Genome* **56**, 61-74.

Zhang, Z.X., Zhao, S.N., Liu, G.F., Huang, Z.M., Cao, Z.M., Cheng, S.H. and Lin, S.S. (2016) Discovery of putative capsaicin biosynthetic genes by RNA-Seq and digital gene expression analysis of pepper. *Sci. Rep.* **6**, 34121.

Zhu, C., Gore, M., Buckler, E.S. and Yu, J. (2008) Status and Prospects of Association Mapping in Plants. *Plant Genome* **1**, 5-20.

CHAPTER II

Identification of a novel locus controlling pungency in *Capsicum*

ABSTRACT

Capsaicinoid is a unique compound of hot peppers (*Capsicum* spp.) and it gives burning sensation called pungency. The *Pun1* gene encoding acyltransferase has been known to control the presence of pungency in cultivated peppers including *C. annuum* and *C. chinense*. In this study, we identified a novel locus controlling pungency named *Pun3*. A total of 92 recombinant inbred lines (RILs) and 201 F2 plants derived from a cross between the non-pungent *C. annuum* 'YCM334' and pungent *C. annuum* 'Tean' were used. RILs and parental lines were sequenced by the genotyping-by-sequencing (GBS) method and 17,646 SNPs were detected. A bin map derived from the SNPs was constructed with a total map length of 1,800 cM and the average marker distance of 1.0 cM. Presence of pungency in RIL and F_2 population was evaluated using Gibb's reagent and HPLC. The segregation ratios of an 1:1 and a 3:1 for pungency and non-pungency were observed in RILs and F_2, respectively, demonstrating the presence of a single locus controlling pungency. Using the high-density genetic map and phenotype data, *Pun3* was mapped to chromosome 7. Additional markers were developed using genome sequence information of *C. annuum* 'CM334', 'Zunla-1' and *C. chinense* 'PI159236', and seven candidate genes were identified in the target region. Five of them were predicted to encode MYB transcription factors, and expression analysis of the candidate showed that the MYB4 expression was significantly reduced in

‘YCM334’. Therefore, the MYB transcription factor was predicted to regulate capsaicinoid biosynthesis. Analyses of promoter sequences of the genes between ‘YCM334’ and ‘Tean’ showed no difference. From the results of this study, we are proposing a capsaicinoid regulation model showing the regulation of capsaicinoid content in pepper.

INTRODUCTION

Pepper (*Capsicum* spp.) is the only genus which can synthesize capsaicinoid. There were limited numbers of studies to find the locus controlling presence of capsaicinoid. The major gene controlling pungency is *Pun1*, which encodes capsaicin synthase with acyltransferase function (Stewart et al., 2005). Capsaicin synthase synthesizes capsaicinoid using vanillylamine from phenylpropanoid pathway and branched-fatty acids from fatty acid pathway. As capsaicin synthase is the enzyme function in the final step for capsaicnoid biosynthesis, most of non-pungent peppers are known to have non-functional *pun1* alleles (Stellari et al., 2010; Stewart et al., 2005; Stewart et al., 2007). Three non-functional alleles including *pun1*, *pun1*^{2}, and *pun1*^{3} were identified in cultivated accessions, *C. annuum*, *C. chinense*, and *C. frutescens*. *pun1* has 2.5 kbp deletion in promoter region and first exon, *pun1*^{2} has early stop codon in first exon, and *pun1*^{3} has INDEL in second exon. Using allelic information of *Pun1*, multiple markers to detect pungent and non-pungent pepper have been developed and have been used for molecular breeding of pungent peppers (Lee et al., 2005; Rodríguez-Maza et al., 2012).

Also *Pun1* regulates the biosynthesis of capsinoid, an analogue of capsaicinoid but almost non-pungent (Han et al., 2013). Capsinoid biosynthesis is also regulated by *putative aminotransferase* (*pAMT*) (Lang et al., 2009). pAMT

catalyze vanillylamine from vanillin in phenylpropanoid pathway. When pepper has non-functional *pamt* allele, low-pungent capsainoid are synthesized instead of pungent capsaicin (Lang et al., 2009; Tanaka et al., 2010a; Tanaka et al., 2010b). More than seven non-functional *pamt* alleles have been identified, and most of them were detected from *C. chinense* (Jang et al., 2015; Tanaka et al., 2015; Tanaka et al., 2018). Non-functional *pamt* alleles have been introgressed *C. annuum* elite accessions to breed high-capsinoid pepper cultivars (Jeong et al., 2015; Tanaka et al., 2014).

Another locus regulates presence of pungency was identified from wild-type pepper *C. chacoense* (Stellari et al., 2010). From allelism test, it was shown that non-pungent *C. chacoense* has a different allele from *pun1* alleles in other *Capsicum* species. This locus was mapped on pepper chromosome 7, and named as *pun2*. It was proposed that *pun2* might be the ortholog of *cap*, a QTL detected to control capsaicin content (Blum et al., 2003). Except *Pun1*, *Pun2*, and *pAMT*, no other locus or gene was identified to regulate presence of pungency.

In previous studies, due to the low density genetic map and lack of gene expression data for the parental lines, it took lots of time and effort to find the candidate genes. However genome sequence of *C. annuum* have been sequenced in 2014 (Kim et al., 2014; Qin et al., 2014) and recently two more reference genomes of *C. baccatum* and *C. chinense* were reported (Kim et al., 2017a). Using reference genome, meta-analysis of multiple researches for the same trait and identification of candidate genes using annotation data will be possible. Also data analysis of

next-generation sequencing including whole-genome resequencing, genotyping-by-sequencing (GBS), and RNA-Seq will be easier than analysis without reference genome (Conesa et al., 2016; Poland and Rife, 2012). With the reference genome of *C. annuum*, candidate genes and loci in peppers were identified for disease resistance, morphological traits, pungency related traits, and so on (Arce-Rodriguez and Ochoa-Alejo, 2017; Cheng et al., 2016; Han et al., 2016; Kim et al., 2017b; Lee et al., 2017; Liu et al., 2016; Nimmakayala et al., 2016).

We found a non-pungent *C. annuum* accession 'YCM334' having functional *Pun1* allele. In this study, we constructed a high-density genetic map using GBS, and identified a novel locus controlling presence of pungency in 'YCM334'. And by RNA-Seq and fine-mapping using high-resolution melting (HRM) markers, a candidate gene encoding R2R3-MYB transcription factor was proposed. These results will be helpful to understand the biosynthetic pathway of capsaicinoid.

MATERIALS AND METHODS

Plant materials and allelism test

RIL and F_2 populations derived from a cross between non-pungent *C. annuum* 'YCM334' and pungent *C. annuum* 'Tean' were used (Truong et al., 2010). Each population was referred to 'YT RIL' and 'YT F_2', respectively. A total of 141 RILs and 219 F_2 plants were grown for three years (2012, 2013, and 2017) and one year (2017) at plastic house located in Anseong, Republic of Korea, respectively. More than three mature fruits per plant were harvested, and placental tissue was dissected. Placental tissues were stored at -20^{o}C until used. In 2017, mature green stage fruits were also harvested and tested for pungency directly.

Allelism test was performed by crossing 'YCM334' with *C. annuum* 'ECW30R' and *C. chacoense* 'PI260433-np'. F_1 plants from each cross together with F_1 from 'YCM334' and 'Tean' were grown in the field of Seoul National University farm in Suwon, Republic Korea. Mature fruits were harvested and freeze-dried for HPLC analysis.

Evaluation of pungency

Presence of pungency was tested using Gibb's reagent method (Jeong et al., 2012). Briefly, when capsaicinoid and Gibb's reagent reacted in the filter paper,

blue color was shown. For this test, samples were prepared in two ways. Fresh fruits in mature green stage were cut horizontally and stamped in the paper, capsaicinoid extracts were used for mature fruits. Capsaicinoid from placental tissues was extracted following the protocol of Han et al. (2013). Content of capsaicin and dihydrocapsaicin was measure by HPLC in National Instrumentation Center for Environmental Management (Seoul, Republic of Korea).

Sequence and gene expression analysis of *Pun1*

Genomic DNA of 'YCM334', 'Tean', RIL, and F_2 plants were extracted using hexadecyl trimethyl ammonium bromide (CTAB) method (Lee et al., 2017). Concentration of DNA was measure using Epoch^{TM} Microplate Spectrophotometer (BioTek, Winooski, VT, USA). DNA was diluted to final concentration of 50 ng/μl.

For sequencing and expression analysis of *Pun1,* we followed the protocol described from Han et al. (2013). Primers were designed based on the *Pun1* sequence of *C. annuum* 'Thai Hot' (NCBI accession number: AY819029.1). cDNA of 'YCM334', 'Tean', pungent wild-type 'Yuwolcho', 'Perennial', and non-pungent accession with non-functional *pun1* allele 'Dempsey' was used for *Pun1* expression analysis.

gDNA extraction and genotyping-by-sequencing (GBS)

GBS libraries were constructed using two different restriction enzyme sets,

PstI/MseI and EcoRI/MseI (described in chapter I). gDNA of 92 RILs and two replicated of each parents were digested by restriction enzyme, and 96 different barcode adapter were aligned to each digested DNA. Separately amplified 96 libraries were pooled in one tube and sequenced in one lane for each restriction enzyme set by HiSeq 2500 (Illumina, San Diego, CA, USA) at Macrogen (Seoul, Republic of Korea).

Raw sequencing reads were filtered with minimum length 80 bp, Phred quality score Q20, and maximum ambiguous bases 2. After demultiplexing reads to each sample using barcode sequences, reads were aligned to *C. annuum* 'CM334' reference genome v.1.55 (Kim et al,, 2014; http://peppergenome.snu.ac.kr/) using the Burrows-Wheeler Aligner program v0.7.12 (Li and Durbin, 2010). Bam files from PstI/MseI and EcoRI/MseI were merged to one file for each parental line and RIL, and used for SNP calling. SNPs were identified using the GATK Unified Genotyper v3.3-0 (DePristo et al., 2011) and filtered with minimum depth 3, and calling rate 0.9. Also Chi-square test was performed to get the SNPs showed expected 1:1 ratio of maternal ('YCM334') and paternal ('Tean') genotyped RILs. Filtered SNPs were separated to each chromosome following the information from 'CM334' reference genome, and used for bin map construction.

Bin map of 92 lines was constructed based on sliding window approach (Han et al., 2016). Ratio of SNPs with maternal and paternal genotype in consecutive 21 SNPs was calculated in 92 RILs separately. When the ratio was below 0.3, between 0.3-0.7, and over 0.7, window was called as paternal, heterozygous, and maternal

genotype, respectively. And the points where the genotype changes were called as recombination breakpoints. By combining all recombination breakpoints of 92 RILs, physical position of bin markers were determined following the protocol of Huang et al. (2009).

Genetic mapping of *Pun3*

High-density genetic map of bin markers was constructed using the Carthagene program (De Givry et al., 2005) and *Pun3* was mapped in the high-density genetic map. For fine-mapping of target region, high-resolution melting (HRM) marker was designed. Based on the physical position of linked bin markers, SNPs located on the region were selected. Also SNPs detected from RNA-Seq data and resequencing of 'YCM334' and 'Tean' (Kang et al., 2016) were used to design linked primers. PCR and HRM were performed as reported previously (Han et al., 2013).

Physical location of primers linked to *Pun3* was identified by BLAST to *C. annuum* 'CM334' (Kim et al., 2014; Hulse-Kemp et al., 2018) *C. annuum* 'Zunla-1' (Qin et al., 2014; http://peppersequence.genomics.cn/), and C. chinense 'PI159236' (Kim et al., 2017) reference genomes. Candidate genes located on the target region was determined by the coding sequence (CDS) information of both reference genomes.

RNA-Seq of ‘YCM334’ and ‘Tean’

Fruits of ‘YCM334’ and ‘Tean’ were harvested 16, 26, and 36 day post anthesis (dpa) and placental tissues were dissected to extract RNA. Total RNA was extracted from placental tissues using Total RNA Extraction kit (MGmed, Seoul, Republic of Korea) following manufacturer’s protocol. Library construction and sequencing were performed at Macrogen (Seoul, Republic of Korea). TruSeq RNA Sample Prep Kit v2 was used to make RNA-Seq library, and six libraries including three replicates for each parental lines were sequenced in one lane of HiSeq 4000 (Illumina, San Diego, CA, USA).

Sequence reads were filtered and analyzed using CLC Genomics Workbench 8.0 (https://www.qiagenbioinformatics.com/). Good quality reads with the length over 80 bp and Phred quality score Q20 were selected. Reference genomes and annotation data of *C. annuum* ‘Zunla-1’ (Qin et al., 2014) and *C. chinense* ‘PI159236’ v.1.2 (Kim et al., 2014) were used for analysis. Expression values in total read number aligned to the genes were calculated and genes showed the false discovery rate (FDR) corrected *P*-value higher than 0.05 were considered as candidate genes.

RT-PCR and sequence analysis

Gene-specific primers and primers to amplify the possible promoter regions were designed based on the CDS sequence of Kim et al. (2017) and Qin et al.

(2014). Total RNA was reverse transcribed to cDNA using M-MLV reverse transcriptase (Thermo Fisher Scientific) and cDNA was diluted five times. RT-PCR was performed with the gene-specific primers and the pepper actin primer. Reaction was composed of 3 μl of diluted cDNA, 10× PCR buffer, 2.5mM dNTPs, 10 pmoles/μl of each primer, and 1 unit of Ex Taq polymerase (Takara Korea Biomedical Inc., Seoul, Republic of Korea). PCR condition was initial 5 min pre-denaturation at 94℃, 30 cycles of 30 s at 94℃, 30 s at 58℃, and 1 min at 72℃, and 10 min final extension at 72℃. RT-PCR PCR products were electrophoresed on 1% agarose gel.

To determine the sequence of candidate gene and promoter regions, same primers used for RT-PCR and additional primers amplify the ~3kb 5' upstream region of first exon were used for PCR. PCR mixtures composition and PCR condition were same as RT-PCR, while the elongation time was changed to 80 s. The PCR products were purified using Zimoclean PCR Purification Kit (Invitrogen Korea, Seoul, Korea) and nucleotide sequencing was performed with 3730xl DNA analyzer (Thermo Fisher Scientific) at Macrogen (Seoul, Republic of Korea).

RESULTS

Presence of pungency

For accurate evaluation, presence of pungency was determined from the placental tissue where capsaicinoid synthesized and accumulated. Non-pungent parental line 'YCM334' did not show blue color from Gibb's reagent test, and capsaicinoid was not detected by HPLC analysis (Table II- 1). Pungent parent 'Tean' showed pungency from both Gibb's reagent method and HPLC analysis. Mean capsaicinoid content of 'Tean' was 6,967 μg/g dry weight of placental tissue.

Pungency of RIL and F_2 populations were evaluated at least two times, from breaker stage fruits and mature stage fruits. If one of the stages showed pungent phenotype, that line was determined as pungent line. A total of 139 RILs were evaluated for pungency, and showed 1:1 ratio of pungent and non-pungent lines (Table II- 1). Segregation of pungent and non-pungent plants in F_2 populations showed 3:1 ratio. Among 214 F_2 plants, 93 plants showed unclear results from Gibb's reagent method, therefore, capsaicinoid content was measured by HPLC analysis for confirm the phenotype. A total of 24 plants were pungent, and their mean capsaicinoid content was 973 μg/g of freeze-dried placental tissue, which is much lower than pungent parental line 'Tean'.

Table II- 1. Segregation of pungency in RIL and F_2 populations

Line	Number of line		Pungent : non-pungent	Capsaicinoid content (μg/g dry weight of placenta)		
	Pungent	Non-pungent		Capsaicin	Dihydrocapsaicin	Total capsaicinoid
YCM334	0	2	0:2	ND	ND	ND
Tean	3	0	3:0	2,656 ± 1,639.4	4,310 ± 2,410.8	6,967 ± 4,029.0
RIL	81	58	1:1 (P>0.05)	-	-	-
F_2	157	57	3:1 (P>0.05)	-	-	-

ND, not detected; -, not evaluated

Allelism test with *pun1* and *pun2* mutant accessions

Segregation ratio of pungent and non-pungent lines showed 1:1 in 'YT RIL' and 3:1 in 'YT F_2'. Therefore, non-pungency of 'YCM334' was expected to be controlled by a single gene. As *Pun1* was known to be the major gene controlling pungency of pepper (Stewart et al., 2005), sequence of *Pun1* was analyzed from 'YCM334' and 'Tean'. There was no significant mutation on promoter and genic regions in 'YCM334' (Figure II- 1). However, from RT-PCR *Pun1* was not expressed in placental tissue of 'YCM334'.

To confirm that 'YCM334' has functional *Pun1*, and uncharacterized gene *Pun2* (Stellari et al., 2010), allelism was tested. F_1 plants from a cross between 'YCM334' and 'ECW30R' with non-functional *pun1* allele had pungent fruits with 2,213 μg of capsaicinoid of g dried placental tissue (Table II- 2). Also same result was shown from F_1 plants from a cross between 'YCM334' and 'PI260433-np' with non-functional *pun2* allele. And from a cross between 'YCM334' and 'Tean', non-pungent phenotype of 'YCM334' was determined as recessive. Therefore another gene controlling pungency in 'YCM334' was named as *Pun3*.

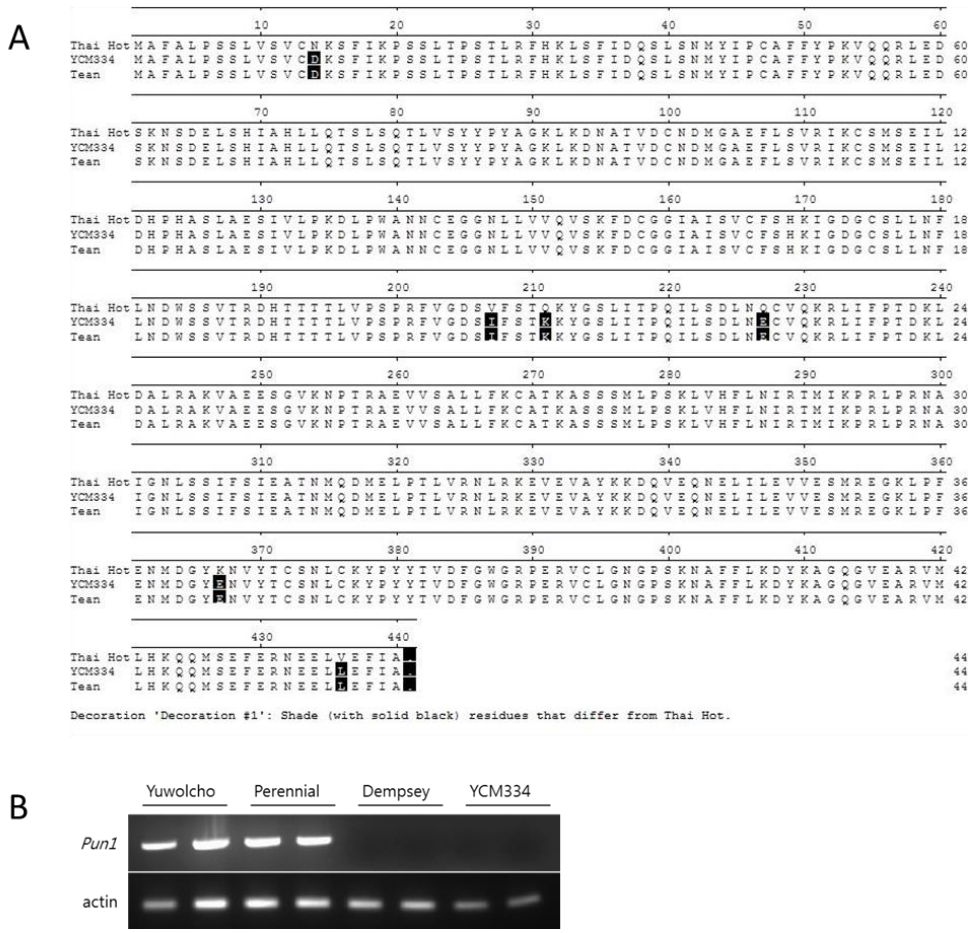


Figure II- 1. Translated amino acid sequences (A) and gene expression (B) of *Pun1*. 'Thai Hot', 'Yuwolcho', and 'Perennial' are pungent accessions and 'Dempsey' is a non-pungent accession with *pun1* allele.

Table II- 2. Allelism test between non-pungent accessions with *pun1* and *pun2*

	C. annuum 'ECW30R' (*pun1*/*pun1*)	*C. chacoense* 'PI260433-np' (*pun2*/*pun2*)	*C. annuum* 'Tean' (wild-type)
C. annuum 'YCM334'	2,213[a]	2,787	2,708

[a]Total capsaicinoid content of mature stage placenta (μg/g dry weight)

Genetic mapping of *Pun3*

To find the genetic position of *Pun3*, high-density genetic map of 'YT RIL' was constructed using GBS (Table II- 3). From the PstI/MseI and EcoRI/MseI restriction enzyme sets, 2,335 and 19,454 SNPs were detected, respectively (data not shown). After combining results from two GBS results, SNPs following 1:1 segregation ratio of maternal and paternal genotype RILs were selected for further study. Filtered 17,646 SNPs were used for sliding window approach (Han et al., 2016) and 1,850 bin markers were detected from twelve chromosomes. Average physical length of bin markers was 1.5 Mbp and mean genetic distance of bin markers was 1.0 cM with the range 0.7-1.2 cM which shows the high-density of the genetic map.

From the phenotype and genotype data of 92 RILs, *Pun3* was mapped on 70.3-92.4 cM of chromosome 7. Using the physical position of the linked bin markers, *Pun3* was expected to be located 187.8-205.3 Mbp on physical position (Figure II- 2). Closest bin marker was 7.6 cM far from *Pun3*, and there were 286 genes were located based on annotation data of 'CM334' reference (v.1.55).

Table II- 3. Bin map constructed by GBS in 92 RIL

Chr.	# of SNPs	# of bins	Physical length (Mbp)		Genetic distance of bin (cM)	
			Mean	Total	Mean	Total
1	1517	201	1.4	272.7	1.2	246.8
2	1005	160	1.1	171.1	1.1	170.8
3	1749	215	1.2	257.9	1.2	250.4
4	963	138	1.6	222.6	0.8	116.6
5	1819	154	1.5	233.5	0.8	124.3
6	1585	174	1.4	236.9	1.0	167.3
7	1233	138	1.7	231.9	0.8	116.2
8	398	69	2.1	145.1	1.5	106.2
9	2845	166	1.5	252.8	0.7	121.7
10	1242	136	1.7	233.6	0.9	124.3
11	2032	146	1.8	259.7	0.8	118.9
12	1258	153	1.5	235.7	0.9	136.4
Total	17,646	1,850	1.5	2753.5	1.0	1,800.0

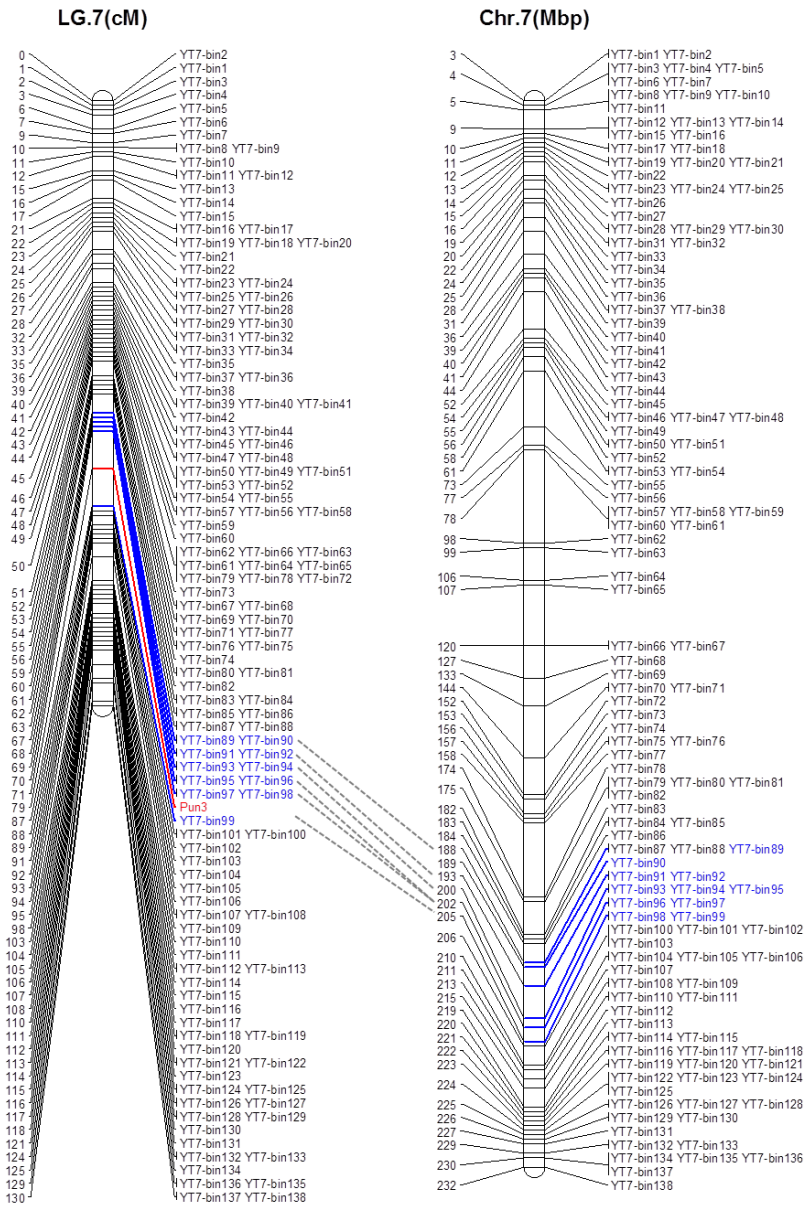


Figure II- 2. Genetic mapping of *Pun3* using a bin map constructed by GBS. LG.7 and Chr.7 shows the genetic and physical position of bins, respectively. Bins marked with blue color shows linkage with the *Pun3* locus. Genetic and physical position of *Pun3* linked bin markers were compared with dotted lines.

Fine-mapping of *Pun3*

HRM markers were developed using the SNPs detected from GBS and RNA-Seq in this study and previously reported resequencing data (Kang et al., 2016). A total of 139 'YT RILs' were genotyped with developed markers on the target region (Table II- 4; Figure II- 3), and three markers including 199496, 939_end2, and 12980_pro were found to be linked to *Pun3*. Additionally, when 214 'YT F_2' was used for fine-mapping of *Pun3*, 199496, 939_end2, and 12980_pro showed three, one and three recombinants in this population, respectively. Therefore, *Pun3* was mapped between those three markers. By comparing the physical positions of the markers in 'CM334' (Hulse-Kemp et al., 2018), 'Zunla-1' (Qin et al., 2014), and 'PI159236' (Kim et al., 2017) pseudomolecules, the target region could be reduced to a physical length of less than 1 Mbp.

Based on 'Zunla-1' and 'PI159236' reference genomes, *Pun3* was mapped to 192.2-193.1 Mbp and 199.3-200.1 Mbp on chromosome 7, respectively. There were five predicted CDS in the target region of 'Zunla-1' including Capana07g001603, Capana07g001604, Capana07g001606, Capana07g001609, and Capana07g001614 (Figure II- 3). Furthermore, CC.CCv1.2.scaffold1257.1, CC.CCv1.2.scaffold1257.2, CC.CCv1.2.scaffold1257.3, CC.CCv1.2.scaffold939.7, CC.CCv1.2.scaffold939.5, CC.CCv1.2.scaffold939.4, CC.CCv1.2.scaffold939.2, and CC.CCv1.2.scaffold939.1 were located on the target region spanning 0.8 Mbp in the 'PI159236' genome (Table II- 5; Figure II- 3).

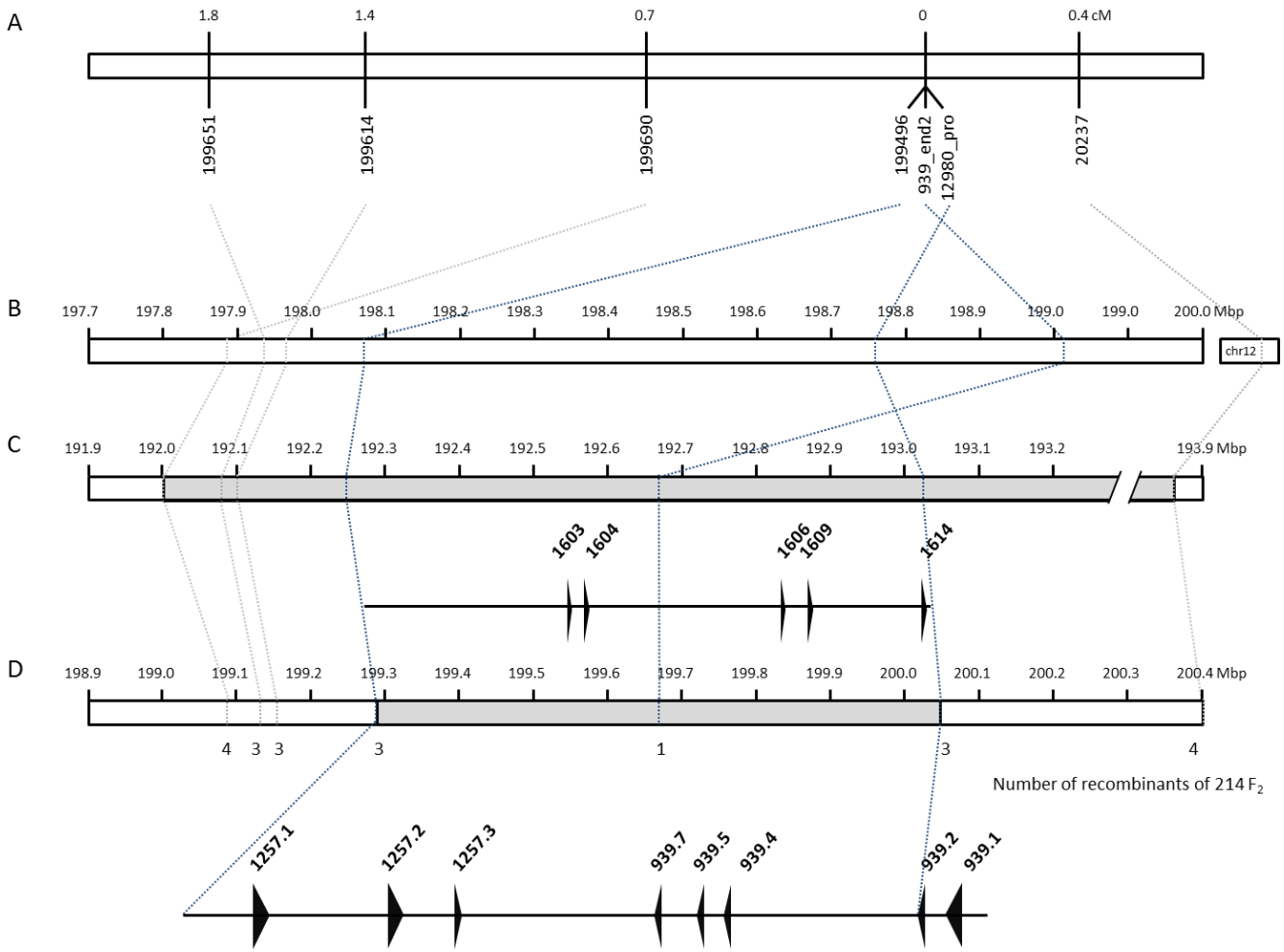

A
1.8
1.4
0.7
0
0.4 cM
199651
199614
199690
199496
939_end2
12980_pro
20237
B
197.7
197.8
197.9
198.0
198.1
198.2
198.3
198.4
198.5
198.6
198.7
198.8
198.9
199.0
199.0
200.0 Mbp
chr12
C
191.9
192.0
192.1
192.2
192.3
192.4
192.5
192.6
192.7
192.8
192.9
193.0
193.1
193.2
193.9 Mbp
1603
1604
1606
1609
1614
D
198.9
199.0
199.1
199.2
199.3
199.4
199.5
199.6
199.7
199.8
199.9
200.0
200.1
200.2
200.3
200.4 Mbp
4
3
3
3
1
3
4
Number of recombinants of 214 F2
1257.1
1257.2
1257.3
939.7
939.5
939.4
939.2
939.1

Figure II- 3. Fine-mapping of the *Pun3* locus based on the multiple *Capsicum* reference genomes. Genetic location of *Pun3* in YT RIL (A), and physical locations of the markers in 'CM334' (B), 'Zunla-1' (C), and 'PI159236' (D) pseudomolecules were compared. Numbers marked in 'PI159236' shows the number of recombinants in YT F_2 population, and candidate genes in the target region are shown (D).

Table II- 4. Primer sequences used in this study

Primer name	Sequence (5' → 3')	Product size (bp)
199496_F	CCACCAAAACGTCAAAATTAG	121
199496_R	ATGTCGAGGTGAGTTTGAGC	
199614_F	CAACGAGTCACTCTCTCCCTT	99
199614_R	TAGTCGGAGTGAGAGGGAAG	
199651_F	CCTGAGAACATTGAGATTCAAACA	121
199651_R	GGCGTCTTTCAAAATTTGCT	
199690_F	CACGGACTTGGCTTATCCTT	121
199690_R	GAAGTTGTAGCAGTTTCTTGCTC	
939_end2_F	TGCAGTAAACTTAATTCGTGCAGT	84
939_end2_R	ACAGAGATTGTTGGAAGCATGC	
12980_pro_F	TTGCGTAAGTGTTGGATGGA	225
12980_pro_R	AACCGCAAGCATGAAAGAAT	
20237_F	TCATGCCCTATTTTTGCCAAGG	84
20237_R	GCGGTGACTTTGTGCGAAAT	
CC.1257.2_F	ACACCTTGCTGTGATGAAAATG	178
CC.1257.2_R	GCCTTAAGTAATTCAGCCATCG	
CC.1257.3_F	AACGTGGACACCTGAAGAAGAT	217
CC.1257.3_R	ATTTCCAAGTTGAGCATGAAGG	
CC.939.7_F	TGTGGGAAGAGCTGCAGATTGA	193
CC.939.7_R	TCTTATGATGAGCATCCCAGTG	
CC.939.4_F	AACTTACCAGGTAGATCGGACA	231
CC.939.4_R	ACTTGATGAAACCTCTGGGAA	
CC.939.2_F	TAACTAGCGGAAGCTTTTGGAC	197
CC.939.2_R	TCCATCTTCAGCTTCTTCTTCA	
CC.939.1_F	GACGAAAATGGAATGAAGAAGG	228
CC.939.1_R	AGCATGGAGGTTCAAGATGATT	
Actin_F	CTTGTCTGTGATAATGGAACAG	189
Actin_R	GGGATACTTCAAGGTGAGAATA	

Table II- 5. Candidate genes of *Pun3* on 'PI159236' reference genome

CDS	Physical location in 'PI159236' (bp)		Homologous gene in 'Zunla-1'	Definition
	Start	End		
CC.CCv1.2.scaffold1257.1	199,369,060	199,381,531	Capana01g003923	Hypothetical protein BC332_19184
CC.CCv1.2.scaffold1257.2	199,522,271	199,524,162	Capana07g001603	Transcription factor MYB34
CC.CCv1.2.scaffold1257.3	199,596,439	199,597,438	Capana07g001604	Myb-related protein Myb4
CC.CCv1.2.scaffold939.7	199,799,307	199,800,001	Capana07g001606	Transcription factor MYB32
CC.CCv1.2.scaffold939.5	199,830,057	199,834,620	Capana09g000758	Hypothetical protein BC332_19188
CC.CCv1.2.scaffold939.4	199,838,633	199,840,007	Capana07g001609	Myb-related protein Myb4
CC.CCv1.2.scaffold939.2	200,046,103	200,052,084	Capana07g001614	Myb-related protein Myb4
CC.CCv1.2.scaffold939.1	200,084,109	200,090,827	Capana07g001604	Hypothetical protein BC332_19185

RNA-Seq for candidate gene identification

To select the candidate genes located on the target region and showing different expression between 'YCM334' and 'Tean', we performed RNA-Seq for parental lines. RNA was extracted from placental tissues at three developmental stages. Filtered reads were aligned to *C. annuum* 'Zunla-1' reference genome and annotation data which is from the same species as 'Tean' and 'YCM334'. Numbers of raw reads were 37 million on average, and 56-60% reads were aligned in 'Zunla-1' (Table II- 6). A total of 729, 3961, and 2091 genes showed significantly different expression levels at 16, 26, and 36 dpa, respectively (Figure II- 4). Down-regulated and up-regulated genes in 'YCM334' were most identified at 26 dpa.

Expression levels of the genes involved in capsaicinoid biosynthetic pathway and candidate genes of *Pun3* were analyzed by counting the number of reads aligned to each gene (Figure II- 5). The gene controlling the presence of pungency, *Pun1*, was expressed only in 'Tean' at 26 and 36 dpa (Figure II- 5A). *BCAT*, *KasI*, and *Fat* which regulate fatty acid biosynthesis showed relatively high levels of expression in 'Tean' than those in 'YCM334', especially at 36 dpa. By contrast, *pAMT* and *Comt* in the phenylpropanoid pathway did not show any significant difference between 'Tean' and 'YCM334'. Eight CDS were predicted in 'PI159236', and two of them were homologous with one CDS of 'Zunla-1' (Table II- 5). Therefore, expression levels of seven CDS were analyzed (Figure II- 5B). Capana07g001606 and Capana07g001609 were not expressed in both parental lines where only one read was mapped to Capana07g001609. Capana01g003923

and Capana09g000758 were expressed in both parental lines. Capana07g001603 and Capana07g001614 were highly expressed in ‘Tean’ at 36 dpa and they were homologous to CC.CCv1.2.scaffold1257.2 and CC.CCv1.2.scaffold939.2, respectively. In addition, Capana07g001604 which is homologous to CC.CCv1.2.scaffold1257.3 and CC.CCv1.2.scaffold939.1 was expressed highly in ‘Tean’. From RNA-Seq, Capana07g001603, Capana07g001604, and Capana07g001614 were selected as candidate genes for *Pun3*.

Table II- 6. RNA-Seq reads used for analysis

Sample	Stage (dpa)	Total read	Filtered read*	Mapped read in pairs
Tean	16	35,969,676	31,794,440	18,367,990
		42,860,678	38,350,122	22,622,454
		42,894,256	38,146,764	22,711,378
	26	33,519,414	30,244,324	17,746,500
		37,073,552	33,496,656	19,691,086
		30,985,268	28,196,810	16,532,608
	36	30,652,786	27,859,100	16,207,696
		44,556,878	40,606,610	23,635,874
		39,695,056	35,094,044	20,676,056
YCM334	16	42,764,246	38,929,348	22,921,094
		35,450,204	32,349,124	19,128,156
		34,564,880	31,495,626	18,441,196
	26	36,695,372	33,389,000	18,783,704
		39,484,656	36,112,366	21,058,078
		38,623,066	35,418,582	20,643,428
	36	38,179,308	34,241,638	19,708,970
		40,505,980	36,419,388	20,713,158
		35,045,874	31,379,862	18,233,538

*Reads were filtered with the minimum length 80 bp and Phred quality score 20

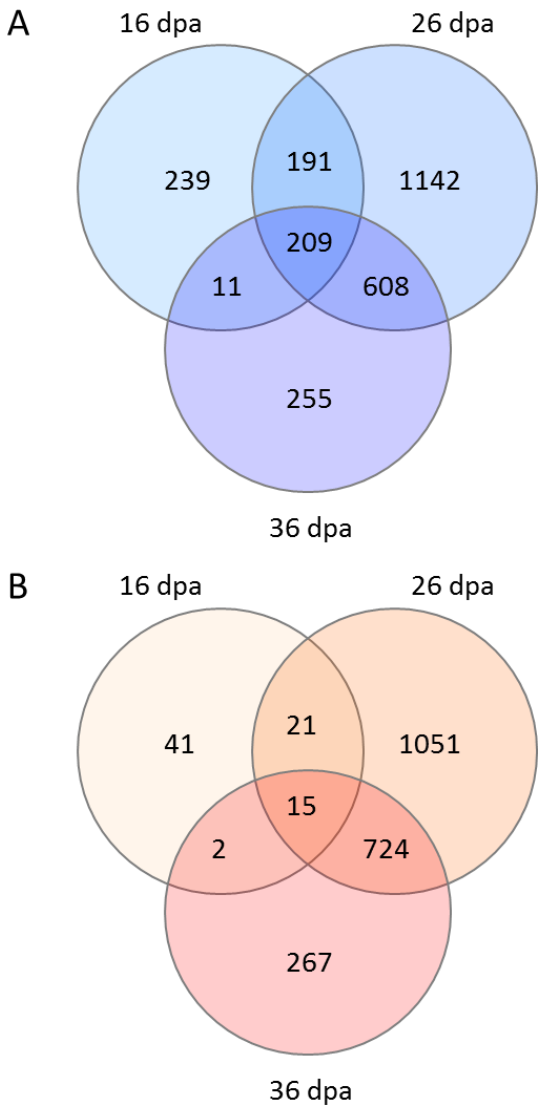


Figure II- 4. Number of down-regulated (A) and up-regulated (B) genes in 'YCM334' compared to 'Tean'. FDR-controlled *P*-value < 0.05 was considered as statistically significant.

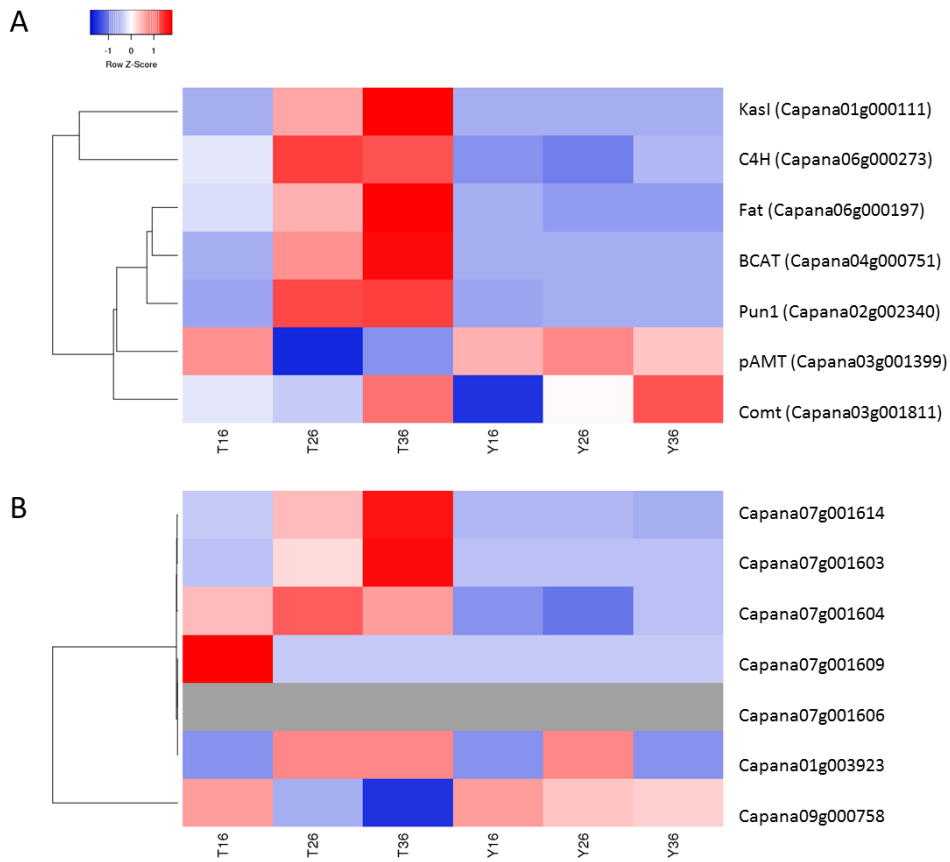


Figure II- 5. Heat map of genes involved in capsaicinoid biosynthetic pathway (A) and the *Pun3* locus (B) between 'Tean' and 'YCM334'. Total number of reads aligned to each gene was used. Gray color means no read was aligned to the gene.

Candidate gene prediction and sequence variation

In the ‘PI159236’ reference genome, eight CDS were predicted on the target region, and six of them were homologous to the CDS identified from ‘Zunla-1’. From RNA-Seq data , three CDS from ‘Zunla-1’ and two CDS from ‘PI159236’ showed differential expression between ‘Tean’ and ‘YCM334’. Capana07g001603, Capana07g001604, and Capana07g001614 all encodes MYB-transcription factors and have more than 75% similarity each other (Figure II- 6). Therefore, to test specific expression of this gene only in ‘Tean’, RT-PCR analysis was performed (Figure II- 7). CC.CCv1.2.scaffold1257.3 and CC.CCv1.2.scaffold939.1 were amplified from ‘Tean’ and they are homologous to Capana07g001604 of ‘Zunla-1’. Additionally, CC.CCv1.2.scaffold939.1 sequence was similar to Capana07g001614 (Figure II- 6B).

To find reveal the sequence difference in CC.CCv1.2.scaffold1257.3 (Capana07g001604) between ‘YCM334’ and ‘Tean’, gene coding regions and 3.7 kbp upstream regions containing the promoter sequence were sequenced. In the predicted promoter region, three SNPs were detected between ‘YCM334’ and ‘Tean’ (Figure II- 8A). Furthermore, one synonymous mutation at the first exon and one mutation on second intron were identified in YCM334’ (Figure II- 8B).

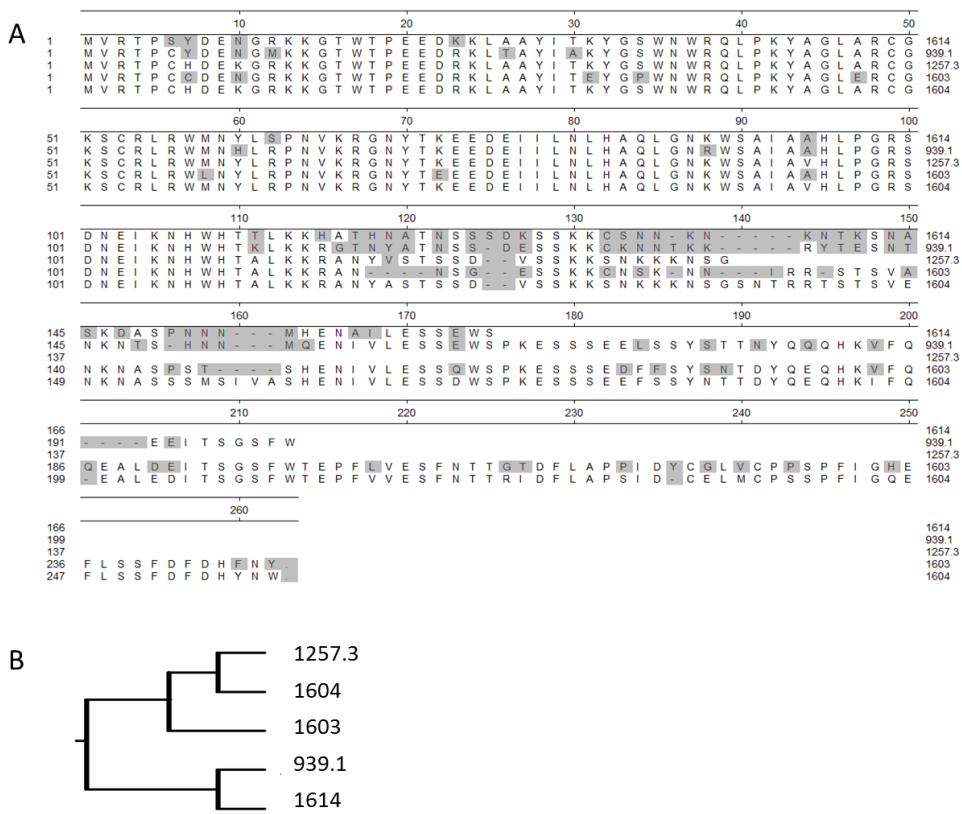


Figure II- 6. Amino acid sequence alignment of candidate genes from 'Zunla-1' and 'PI159236' (A) and UPGMA phylogenetic tree (B). 1603, 1604, 1614, 939.1, and 1257.3 are abbreviations of Capana07g001603, Capana07g001604, Capana07g001614, CC.CCv1.2.scaffold939.1, and CC.CCv1.2.scaffold1257.3, respectively. Shaded sequences are polymorphic from 1604.

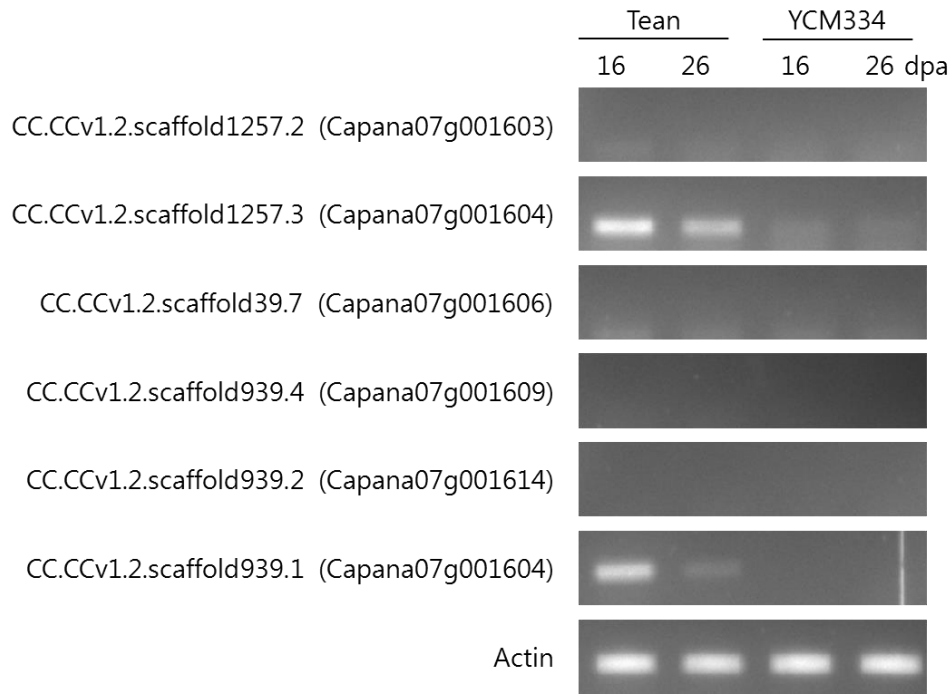


Figure II- 7. Gene expression of candidate genes located on the *Pun3* locus

A

Figure II- 8. Promoter (A) and genic region (B) sequences of Capana07g001604. Shaded regions with purple and black color are variants and exon, respectively. First intron and second exon region sequence were not analyzed and predicted based on ‘Zunla-1’ reference genome.

DISCUSSION

Novel gene controlling pungency

C. annuum 'YCM334' is derived from *C. annuum* 'CM334' and *C. annuum* 'Yolo Wonder' (Truong et al., 2010) and known for its resistance to *Phytophthora capsici* and *Ralstonia solanacearum* (Kang et al., 2016; Liu et al., 2014). Previously, a RIL population derived from a cross between 'YCM334' and 'Tean' have been used for construction of genetic map (Truong et al., 2010) and finding the major QTL for resistance to *P. capsici* (Liu et al., 2014). Using the previous map information and the RIL population, we attempted to map a novel locus controlling pungency. As 'YCM334' is derived from a cross between *C. annuum* accessions, we expected that 'YCM334' would have non-functional allele of the *Pun1* gene. However, 'YCM334' had the same functional *Pun1* allele as other pungent pepper varieties (Stewart et al., 2005; Stewart et al., 2007), but the *Pun1* gene was not expressed in the placental tissue (Figure II- 1). As the promoter region of *Pun1* in 'YCM334' did not show any significant difference from'Tean', we concluded that *Pun3* may control the expression of *Pun1*.

The other locus reported to control pungency was *Pun2*, which was found in the wild accession *C. chacoense* 'PI260433' (Stellari et al., 2010). Although the identity of *Pun2* was not found, but the gene was mapped on chromosome 7 where *Pun3* is also mapped. Furthermore, the closest marker to *Pun2* was located near

201.4 Mbp on chromosome 7 which is very close to the *Pun3* region. However, when we performed an allelism test between 'YCM334' and non-pungent 'PI260433', all the F1 plants had pungent fruits indicating that *Pun2* and *Pun3* are the different loci. Another non-pungent accession, *C. chinense* 'No,3341', was identified (Koeda et al., 2014). This research group showed that non-pungent accession 'No.3341' has functional *Pun1*, *pAMT,* and *O-methyltransferase* alleles which involved in capsaicinoid biosynthetic pathway (Koeda et al., 2015; Koeda et al., 2014). Due to the difficulty of interspecific cross, we could not perform an allelism test between 'YCM334' and 'No.3341'.

Next-generation sequencing (NGS) based genetic mapping

For genetic mapping, SNPs were identified by GBS of 92 'YT RILs'. We used two different restriction enzyme (RE) sets, PstI/MseI and EcoRI/MseI. From our previous research, distribution of SNPs was relatively even when GBS libray was prepared using EcoRI/MseI whereas SNPs derived from PstI/MseI were mostly distributed at the end of each chromosome (described in Chapter I). By combining two RE sets, we could detect SNPs located in genic regions (PstI/MseI) as well as intergenic regions (EcoRI/MseI). Compared to PstI/MseI, EcoRI/MseI RE set generated about eight times more number of SNPs between 'YCM334' and 'Tean'. However, as EcoRI/MseI RE set generates GBS library derived mostly from intergenic regions, we did not use SNPs directly to construct a linkage map. Instead, a sliding window approach was used to construct a bin map and to

minimize problems caused by genotyping errors and missing data. A bin map was successfully used to map QTL in rice, sorghum, maize, and pepper (Chen et al., 2014; Han et al., 2016; Wang et al., 2011; Zou et al., 2012). For single gene mapping, the bin map may not be appropriate to find the accurate and narrow region, because bin marker expands 160-1,070 kbp depends on the crop and population. Therefore, we approximately map the *Pun3* locus using bin map, and fine-map the target region using SNPs. We expect this approach is more appropriate for the crops having a large and complicated genome with multiple repeated sequences.

Both GBS and RNA-Seq data were analyzed based on 'CM334' reference genome v.1.55. Using GBS data, a high-density genetic map was constructed and *Pun3* was mapped. Furthermore, we could detect multiple SNPs within the target region from RNA-Seq data, and used for fine-mapping. By genotyping with HRM markers and comparison of multiple *Capsicum* reference genomes, eight and five candidate genes were selected from *C. chinense* 'PI159236' and *C. annuum* 'Zunla-1'. Furthermore from RNA-Seq, expressions of candidate genes in the placenta tissues between parental lines were compared. A total of three genes showed differential expression, while only one of them was amplified from RT-PCR. In the *Pun3* target region, five putative MYB transcription factors having highly similar sequences were identified (Figure II- 6). Due to high sequence similarity and duplication of MYB transcription factors in the regions, the sequences of our tartget region could be misassembled in the reference genomes.

HRM marker ‘20237’ was mapped to chromosome 12 in ‘CM334’ (Figure II- 3), and Capana07g001603 and Capana07g001604 were also mapped to chromosome 12 in ‘CM334’. Therefore, construction of genetic map and development of markers using NGS can be applied not only genetic mapping of specific trait, but also for improve reference genomes by re-assembling the regions with repeated sequences.

MYB transcription factor regulates capsaicinoid

MYB transcription factor is one of the largest groups of transcription factors in plants (Ambawat et al., 2013). Among the MYB transcription factors, R2R3-type MYB domain is most common in plants and regulates many biological processes including plant tissue development, response to abiotic stresses, and metabolism. MYB transcription factor also controls the phenylpropanoid pathway and related secondary metabolites like anthocyanin (Dubos et al., 2008; Gonzalez et al., 2008; Quattrocchio et al., 2006).

In pepper, 91 MYB transcription factors and 83 MYB-related transcription factors were predicted (Kim et al., 2014), several MYB was identified to control anthocyanin biosynthesis of pepper fruits (Borovsky et al., 2004; Stommel et al., 2009), to respond to abiotic stress (Seong et al., 2008), and to affect capsaicinoid content (Arce-Rodriguez and Ochoa-Alejo, 2017). Arce-Rodriguez and colleagues found that CaMYB31, one of the MYB transcription factors identified in *C. annuum* regulates multiple genes in capsaicinoid biosynthesis pathway, and the

expression is affected by hormone or abiotic stresses. CaMYB31 was predicted as Capana07g001604 which is one of our candidate genes. Even though we could not find any significant variants in this gene, a downstream gene affected by CaMYB31 might regulate pungency in 'YCM334'. Another candidate gene was CC.CCv1.2.scaffold939.2 encoding Myb-related protein Myb4 (Table II- 5). To confirm that the MYB transcription factor is the gene encoding *Pun3*, further sequence analysis of upstream regions need to be done.

RNA-Seq results of 'YCM334' and 'Tean' showed that expressions of the *Pun1* gene and three fatty-acid pathway related genes, *Kas*, *BCAT*, and *FatA*, were reduced in non-pungent 'YCM334'. When the *Pun1* gene was silenced, the expressions of *Kas*, *BCAT*, *FatA*, and *pAMT* were significantly reduced (Arce-Rodriguez and Ochoa-Alejo, 2015). Furthermore, silencing of CaMYB31 resulted in significant reduction of expressions of capsaicinoid biosynthesis genes and eventually leads to reduction of capsaicinoid content. Therefore, from the results, two possible regulation models can be proposed. First, precursors or capsaicinoid itself might be the negative feedback regulators and down-regulate the capsaicinoid biosynthetic genes (Arce-Rodriguez and Ochoa-Alejo, 2015; Kim et al., 2009). Additionaly, MYB transcription factors, candidate genes for *Pun3*, regulates *Pun1* and other capsaicinoid biosynthetic genes especially fatty-acid related genes. In Arabidopsis and tomato, MYB complex was considered to regulate accumulation of anthoycanin, another secondary metabolite, and expression of biosynthetic genes (Cao et al., 2017; Gonzalez et al., 2008). Further studies are necessary to find other

transcription factors or investigate the interaction of MYB and stuructural genes.

Here, we find a novel locus controlling presence of pungency in *C. annuum* pepper. *Pun3* was predicted to encode MYB transcription factor which regulates multiple genes involved in capsaicinoid biosynthesis. We figured out that non-pungent phenotype can be caused by mutation in one MYB transcription factor, and revealing the function of *Pun3* will be helpful to understand the capsaicinoid biosynthesis in pepper.

REFERENCES

Ambawat, S., Sharma, P., Yadav, N.R. and Yadav, R.C. (2013) MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* **19**, 307-321.

Arce-Rodriguez, M.L. and Ochoa-Alejo, N. (2015) Silencing *AT3* gene reduces the expression of *pAmt*, *BCAT*, *Kas*, and *Acl* genes involved in capsaicinoid biosynthesis in chili pepper fruits. *Biol. Plantarum* **59**, 477-484.

Arce-Rodriguez, M.L. and Ochoa-Alejo, N. (2017) An R2R3-MYB transcription factor regulates capsaicinoid biosynthesis. *Plant Physiol.* **174**, 1359-1370.

Blum, E., Mazourek, M., O'Connell, M., Curry, J., Thorup, T., Liu, K., Jahn, M. and Paran, I. (2003) Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. *Theor. Appl. Genet.* **108**, 79-86.

Borovsky, Y., Oren-Shamir, M., Ovadia, R., De Jong, W. and Paran, I. (2004) The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to *Anthocyanin2* of Petunia. *Theor. Appl. Genet.* **109**, 23-29.

Cao, X., Qiu, Z., Wang, X., Van Giang, T., Liu, X., Wang, J., Wang, X., Gao, J., Guo, Y., Du, Y., Wang, G. and Huang, Z. (2017) A putative R3 MYB repressor is the candidate gene underlying atroviolacium, a locus for anthocyanin pigmentation in tomato fruit. *J. Exp. Bot.* **68**, 5745-5758.

Chen, Z., Wang, B., Dong, X., Liu, H., Ren, L., Chen, J., Hauck, A., Song, W. and Lai, J. (2014) An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F_2 maize population. *BMC Genomics* **15**, 433.

Cheng, J., Qin, C., Tang, X., Zhou, H., Hu, Y., Zhao, Z., Cui, J., Li, B., Wu, Z., Yu, J. and Hu, K. (2016) Development of a SNP array and its application to genetic

mapping and diversity assessment in pepper (*Capsicum* spp.). *Sci. Rep.* **6**, 33293.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80-92.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X. and Mortazavi, A. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13.

De Givry, S., Bouchez, M., Chabrier, P., Milan, D. and Schiex, T. (2005) CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* **21**, 1703-1704.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. and Daly, M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498.

Dubos, C., Le Gourrierec, J., Baudry, A., Huep, G., Lanet, E., Debeaujon, I., Routaboul, J.M., Alboresi, A., Weisshaar, B. and Lepiniec, L. (2008) MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *Plant J.* **55**, 940-953.

Gonzalez, A., Zhao, M., Leavitt, J.M. and Lloyd, A.M. (2008) Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* **53**, 814-827.

Han, K., Jeong, H.J., Sung, J., Keum, Y.S., Cho, M.C., Kim, J.H., Kwon, J.K., Kim, B.-D. and Kang, B.-C. (2013) Biosynthesis of capsinoid is controlled by the *Pun1* locus in pepper. *Mol. Breed.* **31**, 537-548.

Han, K., Jeong, H.J., Yang, H.B., Kang, S.M., Kwon, J.K., Kim, S., Choi, D. and Kang, B.C. (2016) An ultra-high-density bin map facilitates high-throughput QTL mapping of horticultural traits in pepper (*Capsicum annuum*). *DNA Res.* **23**, 81-91.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T. and Han, B. (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**, 1068-1076.

Hulse-Kemp, A.M., Maheshwari, S., Stoffel, K., Hill, T.A., Jaffe, D., Williams, S.R., Weisenfeld, N., Ramakrishnan, S., Kumar, V., Shah, P., Schatz, M.C., Church, D.M. and Van Deynze, A. (2018) Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic. Res.* **5**.

Jang, S., Han, K., Jo, Y.D., Jeong, H.J., Siddique, M.I. and Kang, B.C. (2015) Substitution of a dysfunctional *pAMT* allele results in low-pungency but high levels of capsinoid in *Capsicum chinense* 'Habanero'. *Plant Breed. Biotechnol.* **3**, 119-128.

Jeong, H.J., Hwang, D.Y., Ahn, J.T., Chun, J.Y., Han, K.E., Lee, W.-M., Kwon, J.K., Lee, Y.-J. and Kang, B.C. (2012) Development of a simple method for detecting capsaicinoids using Gibb's reagent in pepper. *Hortic. Sci. Technol.* **30**, 294-300.

Jeong, H.-S., Jang, S., Han, K., Kwon, J.K. and Kang, B.C. (2015) Marker-assisted backcross breeding for development of pepper varieties (*Capsicum annuum*) containing capsinoids. *Mol. Breed.* **35**, 226.

Kang, Y.J., Ahn, Y.K., Kim, K.T. and Jun, T.H. (2016) Resequencing of *Capsicum annuum* parental lines (YCM334 and Taean) for the genetic analysis of bacterial wilt resistance. *BMC Plant Biol.* **16**, 235.

Kim, J.S., Park, M., Lee, D.J. and Kim, B.D. (2009) Characterization of putative capsaicin synthase promoter activity. *Mol. Cells* **28**, 331-339.

Kim, S., Park, J., Yeom, S.I., Kim, Y.M., Seo, E., Kim, K.T., Kim, M.S., Lee, J.M., Cheong, K., Shin, H.S., Kim, S.B., Han, K., Lee, J., Park, M., Lee, H.A., Lee,

H.Y., Lee, Y., Oh, S., Lee, J.H., Choi, E., Choi, E., Lee, S.E., Jeon, J., Kim, H., Choi, G., Song, H., Lee, J., Lee, S.C., Kwon, J.K., Lee, H.Y., Koo, N., Hong, Y., Kim, R.W., Kang, W.H., Huh, J.H., Kang, B.C., Yang, T.J., Lee, Y.H., Bennetzen, J.L. and Choi, D. (2017a) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* **18**, 210.

Kim, S., Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A., Seo, E., Choi, J., Cheong, K., Kim, K.T., Jung, K., Lee, G.W., Oh, S.K., Bae, C., Kim, S.B., Lee, H.Y., Kim, S.Y., Kim, M.S., Kang, B.C., Jo, Y.D., Yang, H.B., Jeong, H.J., Kang, W.H., Kwon, J.K., Shin, C., Lim, J.Y., Park, J.H., Huh, J.H., Kim, J.S., Kim, B.D., Cohen, O., Paran, I., Suh, M.C., Lee, S.B., Kim, Y.K., Shin, Y., Noh, S.J., Park, J., Seo, Y.S., Kwon, S.Y., Kim, H.A., Park, J.M., Kim, H.J., Choi, S.B., Bosland, P.W., Reeves, G., Jo, S.H., Lee, B.W., Cho, H.T., Choi, H.S., Lee, M.S., Yu, Y., Do Choi, Y., Park, B.S., van Deynze, A., Ashrafi, H., Hill, T., Kim, W.T., Pai, H.S., Ahn, H.K., Yeam, I., Giovannoni, J.J., Rose, J.K., Sorensen, I., Lee, S.J., Kim, R.W., Choi, I.Y., Choi, B.S., Lim, J.S., Lee, Y.H. and Choi, D. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **46**, 270-278.

Kim, S.B., Kang, W.H., Huy, H.N., Yeom, S.I., An, J.T., Kim, S., Kang, M.Y., Kim, H.J., Jo, Y.D., Ha, Y., Choi, D. and Kang, B.C. (2017b) Divergent evolution of multiple virus-resistance genes from a progenitor in *Capsicum* spp. *New Phytol.* **213**, 886-899.

Koeda, S., Sato, K., Tanaka, Y., Takisawa, R. and Kitajima, A. (2015) A *Comt1* loss of function mutation is insufficient for loss of pungency in *Capsicum*. *Am. J. Plant Sci.* **06**, 1243-1255.

Koeda, S., Sato, K., Tomi, K., Tanaka, Y., Takisawa, R., Hosokawa, M., Doi, M., Nakazaki, T. and Kitajima, A. (2014) Analysis of non-pungency, aroma, and origin of a *Capsicum chinense* cultivar from a Caribbean island. *J. Jpn. Soc. Hortic. Sci.* **83**, 244-251.

Lang, Y., Kisaka, H., Sugiyama, R., Nomura, K., Morita, A., Watanabe, T., Tanaka, Y., Yazawa, S. and Miwa, T. (2009) Functional loss of *pAMT* results in biosynthesis of capsinoids, capsaicinoid analogs, in *Capsicum annuum* cv. CH-19 Sweet. *Plant J.* **59**, 953-961.

Lee, C.-J., Yoo, E., Shin, J.H., Lee, J., Hwang, H.S. and Kim, B.D. (2005) Non-pungent *Capsicum* contains a deletion in the capsaicinoid synthetase gene, which allows early detection of pungency with SCAR markers. *Mol.Cells.* **19**, 262-267.

Lee, J.-H., An, J.-T., Siddique, M.I., Han, K., Choi, S., Kwon, J.-K. and Kang, B.-C. (2017) Identification and molecular genetic mapping of Chili veinal mottle virus (ChiVMV) resistance genes in pepper (*Capsicum annuum*). *Mol. Breed.* **37**, 121.

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

Liu, L., Venkatesh, J., Jo, Y.D., Koeda, S., Hosokawa, M., Kang, J.H., Goritschnig, S. and Kang, B.C. (2016) Fine mapping and identification of candidate genes for the sy-2 locus in a temperature-sensitive chili pepper (*Capsicum chinense*). *Theor. Appl. Genet.***129**, 1541-1556.

Liu, W.Y., Kang, J.H., Jeong, H.S., Choi, H.J., Yang, H.B., Kim, K.T., Choi, D., Choi, G.J., Jahn, M. and Kang, B.C. (2014) Combined use of bulked segregant analysis and microarrays reveals SNP markers pinpointing a major QTL for resistance to *Phytophthora capsici* in pepper. *Theor. Appl. Genet.***127**, 2503-2513.

Nimmakayala, P., Abburi, V.L., Saminathan, T., Alaparthi, S.B., Almeida, A., Davenport, B., Nadimi, M., Davidson, J., Tonapi, K., Yadav, L., Malkaram, S., Vajja, G., Hankins, G., Harris, R., Park, M., Choi, D., Stommel, J. and Reddy, U.K. (2016) Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. *Sci. Rep.* **6**, 38081.

Poland, J.A. and Rife, T.W. (2012) Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome* **5**, 92.

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Ling-Hu, C., Zhou, H., Lin, H., Gonzalez-Morales, S., Trejo-Saavedra, D.L., Tian, H., Tang, X., Zhao, M., Huang, Z., Zhou, A., Yao, X., Cui, J., Li, W., Chen, Z., Feng, Y., Niu, Y., Bi, S., Yang, X., Li, W., Cai, H., Luo, X., Montes-Hernandez, S., Leyva-Gonzalez, M.A., Xiong, Z., He, X., Bai, L., Tan, S., Tang, X., Liu, D., Liu, J., Zhang, S., Chen, M., Zhang, L., Zhang, L., Zhang, Y., Liao, W., Zhang, Y., Wang, M., Lv, X., Wen, B., Liu, H., Luan, H., Zhang, Y., Yang, S., Wang, X., Xu, J., Li, X., Li, S., Wang, J., Palloix, A., Bosland, P.W., Li, Y., Krogh, A., Rivera-Bustamante, R.F., Herrera-Estrella, L., Yin, Y., Yu, J., Hu, K. and Zhang, Z. (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl Acad. Sci. USA* **111**, 5135-5140.

Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J. and Koes, R. (2006) PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* **18**, 1274-1291.

Rodríguez-Maza, M.J., Garcés-Claver, A., Park, S.W., Kang, B.C. and Arnedo-Andrés, M.S. (2012) A versatile PCR marker for pungency in *Capsicum* spp. *Mol. Breed.* **30**, 889-898.

Seong, E.S., Guo, J. and Wang, M.H. (2008) The chilli pepper (*Capsicum annuum*) MYB transcription factor (CaMYB) is induced by abiotic stresses. *J. Plant Biochem. Biotechnol.* **17**, 193-196.

Stellari, G.M., Mazourek, M. and Jahn, M.M. (2010) Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. *Heredity* **104**, 460-471.

Stewart, C., Jr., Kang, B.C., Liu, K., Mazourek, M., Moore, S.L., Yoo, E.Y., Kim, B.D., Paran, I. and Jahn, M.M. (2005) The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant J* **42**, 675-688.

Stewart, C., Jr., Mazourek, M., Stellari, G.M., O'Connell, M. and Jahn, M. (2007) Genetic control of pungency in *C. chinense* via the *Pun1* locus. *J. Exp. Bot.* **58**, 979-991.

Stommel, J.R., Lightbourn, G.J., Winkel, B.S. and Griesbach, R.J. (2009) Transcription factor families regulate the anthocyanin biosynthetic pathway in *Capsicum annuum*. *J. Amer. Soc. Hort. Sci.* **134**, 244-251.

Tanaka, Y., Fukuta, S., Koeda, S., Goto, T., Yoshida, Y. and Yasuba, K.i. (2018) Identification of a novel mutant *pAMT* allele responsible for low-pungency and capsinoid production in chili pepper: accession 'No. 4034' (*Capsicum chinense*). *Hortic. J.* doi: 10.2503/hortj.OKD-115

Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T. and Yazawa, S. (2010a) Newly mutated *putative-aminotransferase* in nonpungent pepper (*Capsicum annuum*) results in biosynthesis of capsinoids, capsaicinoid analogues. *J. Agric. Food Chem.* **58**, 1761-1767.

Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T. and Yazawa, S. (2010b) Novel loss-of-function *putative aminotransferase* alleles cause biosynthesis of capsinoids, nonpungent capsaicinoid analogues, in mildly pungent chili peppers (*Capsicum chinense*). *J. Agric. Food Chem.* **58**, 11762-11767.

Tanaka, Y., Sonoyama, T., Muraga, Y., Koeda, S., Goto, T., Yoshida, Y. and Yasuba, K. (2015) Multiple loss-of-function putative aminotransferase alleles contribute to low pungency and capsinoid biosynthesis in *Capsicum chinense*. *Mol. Breed.* **35**, 142.

Tanaka, Y., Yoneda, H., Hosokawa, M., Miwa, T. and Yazawa, S. (2014) Application of marker-assisted selection in breeding of a new fresh pepper cultivar (*Capsicum annuum*) containing capsinoids, low-pungent capsaicinoid analogs. *Sci. Hort.* **165**, 242-245.

Truong, H.T.H., Kim, K.T., Kim, S., Chae, Y., Park, J.H., Oh, D.G. and Cho, M.C. (2010) Comparative mapping of consensus SSR markers in an intraspecific F8 recombinant inbred line population in *Capsicum*. *Hortic. Environ. Biotechnol.*

51, 193-206.

Wang, L., Wang, A., Huang, X., Zhao, Q., Dong, G., Qian, Q., Sang, T. and Han, B. (2011) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. *Theor. Appl. Genet.* **122**, 327-340.

Zhang, Z.X., Zhao, S.N., Liu, G.F., Huang, Z.M., Cao, Z.M., Cheng, S.H. and Lin, S.S. (2016) Discovery of putative capsaicin biosynthetic genes by RNA-Seq and digital gene expression analysis of pepper. *Sci. Rep.* **6**, 34121.

Zou, G., Zhai, G., Feng, Q., Yan, S., Wang, A., Zhao, Q., Shao, J., Zhang, Z., Zou, J., Han, B. and Tao, Y. (2012) Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. *J. Exp. Bot.* **63**, 5451-5462.

ABSTRACT IN KOREAN

고추는 매운맛을 내는 물질인 캡사이시노이들 합성하는 유일한 식물로, 국내외에서 채소뿐만 아니라 각종 양념, 첨가물, 의약품으로서 소비되고 있다. 캡사이시노이드 합성 경로는 애기장대와 같이 잘 밝혀진 합성경로를 가지고 있는 식물을 바탕으로 예측되었다. 특히 이 경로상에 위치하는 유전자에 대한 발현 실험을 통해 실제 캡사이시노이드 함량과 연관이 있는 후보 유전자들이 선발되었다. 하지만 여전히 유전자들이 실제로 캡사이시노이드 함량을 직접적으로 조절하는지, 조절한다면 어떤 역할을 하는지에 대한 연구가 매우 부족한 실정이다.

따라서 본 연구에서는 두 가지 연구를 통해 고추의 캡사이시노이드 합성에 관여하는 유전인자를 탐색하고자 하였다. 첫 번째 연구에서는 매운맛 정도, 즉 캡사이시노이드 함량을 조절하는 양적조절유전자좌(QTL)를 탐색하였다. 이를 위하여 두 개의 근동질유전자계통 집단을 활용하였다. 보다 정확한 캡사이시노이드 함량 분석을 위하여, 캡사이시노이드가 주로 합성되고 축적되는 조직인 과실의 태좌로부터 함량을 측정하였다.

또한 차세대유전체분석 기술을 기반으로 하여 다수의 단일염기다형성(SNP) 마커를 탐색하였으며, 이를 통해 고밀도 유전자 지도를 작성하였다. 작성한 유전자 지도와 조사한 표현형 정보를 바탕으로 두 집단에서 각각 QTL 을 찾았고, SNP 를 탐색하는데 활용하였던 'CM334' 고추의 표준유전체 서열을 바탕으로 두 집단의 공통적인 QTL 을 염색체 1, 2, 3, 4, 10 번에서 확인하였다. 또한 보다 정밀한 유전자 탐색을 위해서 본 연구실에서 고추의 핵심집단을 이용해 수행한 전장유전체상관성분석(GWAS) 연구 결과와 QTL 분석 결과를 비교하였다. 그 결과 총 10 개의 지역이 QTL 분석과 GWAS 분석에서 공통적으로 캡사이시노이드 함량을 조절하는 것으로 확인하였다. 또한 캡사이시노이드 합성 경로에 위치한다고 밝혀져 있거나, 연결되어 있을 것으로 예측되는 다섯 개의 후보 유전자를 선발할 수 있었다. 두 번째 연구에서는 새롭게 발견된 맵지 않은 고추 계통을 활용하여 캡사이시노이드 합성 유무를 결정하는 새로운 유전자좌인 *Pun3* 를 탐색하였다. 맵지 않은 계통인 *C. annuum* 'YCM334' 와 매운 계통인 *C. annuum* 'Tean' 을 교배하여 분리집단을 육성하였으며, 이들의 캡사이시노이드 합성

유무에 대한 표현형 조사를 수행하였다. 또한 genotyping-by-sequencing (GBS) 방법을 활용하여 고밀도 유전자 지도를 작성하여 염색체 7 번에 *Pun3* 를 매핑하였다. SNP 정보를 이용하여 *Pun3* 지역을 정밀하게 좁혀나갔으며, 'CM334', 'Zunla-1', *C. chinense* 'PI159236' 표준유전체 서열을 기반으로 여덟 개의 후보 유전자를 찾았다. 이 중 다섯 개의 유전자는 MYB transcription factor 를 암호화하는 것으로 예측되었으며, 기존 연구로부터 캡사이시노이드 합성 경로와 연관이 있을 것으로 보고되었다. 부모 계통의 RNA-Seq 과 RT-PCR 을 통한 유전자 발현 분석 결과 'YCM334'에서만 유전자가 발현하지 않는 MYB transcription factor 두 개를 최종 선발하였다.

두 연구 결과로부터 고추의 캡사이시노이드 함량과 합성 유무를 조절하는 후보 유전자들을 선발할 수 있었다. 추가적인 유전자 검증 연구가 필요하지만, 이 결과는 캡사이시노이드 합성에 대한 이해를 돕는데 유용할 뿐만 아니라 고신미 품종을 육성하기 위한 분자육종법으로의 활용도도 높을 것으로 판단된다.