A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Construction of an SNP linkage map using genotyping-by-sequencing in *Capsicum chinense*

GBS를 이용한 고추(*Capsicum chinense*)의
SNP 연관 지도 작성

FEBRUARY, 2016

JIWOONG HAN

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

Construction of an SNP linkage map using genotyping-by-sequencing in *Capsicum chinense*

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

BY
JIWOONG HAN

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

FEBRUARY, 2016

APPROVED AS A QUALIFIED DISSERTATION OF JIWOONG HAN
FOR THE DEGREE OF MASTER OF SCEINCE
BY THE COMMITTEE MEMBERS

CHAIRMAN	**Eun Jin Lee, Ph.D.**
VICE-CHAIRMAN	**Byoung-Cheorl Kang, Ph.D.**
MEMBER	**Doil Choi, Ph.D.**

Construction of an SNP linkage map using genotyping-by-sequencing in *Capsicum chinense*

JIWOONG HAN

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSTRACTS

Construction of linkage map is important for genetic studies and marker-assisted breeding programs. Recent advances in next-generation sequencing technologies allow the rapid and efficient generation of high-density linkage maps. In this study, we used a modified genotyping-by-sequencing (GBS) protocol involving two restriction enzymes (*PstI-MseI*). A total of 87 F_2 individuals derived from a cross of *C. chinense* 'Habanero' and *C. chinense* 'Jolokia' were used to construct an SNP linkage map. To imputate missing data, a novel method combining sliding window and bin mapping was used to obtain reliable genotype data. A total of 1,000 bin markers selected from 217,152 SNPs were used for

construction of a linkage map. The total genetic distance was about 3654.1 cM, and the average distance between bin markers was 3.9 cM. These bin markers were evenly distributed over 14 linkage groups, corresponding to 12 pepper chromosomes. This research shows that GBS together with the sliding window imputation method enabled us to rapidly develop a genetic map in an F_2 generation. Using the linkage map, QTL mapping was performed for capsaicinoids content in the placenta and pericarp tissues. Also, we analyzed *Pun1* expression to test whether capsaicinoids are synthesized in the pericarp. The results showed that only Jolokia showed *Pun1* expression whereas Habanero and F_1 hybrid showed no expression in pericarp. Distribution of capsaicinoids content in the pericarp of F_2 plants also indicated that capsaicinoids might be regulated by a recessive gene(s). When QTL analysis was performed using the GBS linkage map and capsaicinoids content data, one QTL was identified for capsaicinoids in the placenta. Six QTLs were detected on chromosome 6 and 9 for capsaicinoids content in the pericarp. Bin markers linked to the two major QTLs will be used for selection plants containing high capsaicinoids content.

Keywords: Genotyping-by-sequencing (GBS), sliding window, bin-map, SNP linkage map, single nucleotide polymorphism (SNP)

Student number: 2014-20018

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

RILs	Recombinant inbred lines
GBS	Genotyping-by-sequencing
SNP	Single nucleotide polymorphism
NGS	Next generation sequencing
cM	centi Morgan (the unit of genetic distance)
LOD	Logarithm of the odds
MAS	Marker-assisted selection
MAF	Minor-allele frequency
RFLPs	Restriction fragment length polymorphisms
SSRs	Microsatellites
CTAB	Cetyl trimethylammonium bromide
BWA	Burrow wheelers alignment
SAM	Sequence alignment/map
GATK	GenomeAnalysisTK
AF	Allele frequency
LG	Linkage group
QTLs	Quantitative trait loci

INTRODUCTION

In the field of plant genetics and breeding, molecular markers have become essential. Amid of diverse molecular markers single nucleotide polymorphisms (SNPs) is eminently substantial and appropriate for large scale analysis (Rafalski et al., 2002; Zhu et al., 2003). During recent years, next generation sequencing (NGS) technologies have been utilized for detection of SNPs present in genome (Elshire et al., 2011). Although high-throughput genotyping platforms are available, these are still a time consuming and expensive process. Series of sequential advancements in sequencing technologies and base calling softwares are letting NGS technologies provide higher throughputs per run. Day to day up-grading sequencing technologies resulted in the development of inexpensive platform known as genotype-by-sequencing (GBS). Recently, GBS is emerging as a simple but highly powerful, convenient, and economically worthy replacement to many whole-genome resequencing technologies to deliver large number of SNPs in a range of plant species. It is also anticipated that the volume of sequencing data produced and its standard will keep on increasing per run, thus it allows higher plexing but lower cost per sample. This practice will enable plant breeders to obtain useful genomic information even in plant species with a big genome (Elshire et al., 2011).

Two distinct approaches of GBS have been developed (Poland et al., 2012a). First one is the restriction enzyme digestion which can be used for marker-assisted

selection (MAS) events. Genome complication can be minimized after digesting the DNA with selected restriction enzymes prior to the ligation of adapters. The second is multiplex enrichment PCR; in this strategy, numerous SNPs are outlined in a part of genome. In this methodology for the amplification of interested area a defined set of primers are designed (Wang et al., 2014).

Many researchers have used GBS as an effective approach in genetic diversity studies (Poland et al., 2012b; Fu et al., 2014; Jarquín et al., 2014; Pootakham et al., 2015). Poland et al., (2012a) used GBS approach to map over 34,000 SNPs and 240,000 tags onto Barley, and 20,000 SNPs and 367,000 tags on the wheat reference map. They also found 41,371 SNPs in a set of 254 advanced breeding lines from CIMMYT's semiarid wheat breeding program (Poland et al., 2012b). The application of GBS was also applied to economically important conifer species lacking reference sequences, lodgepole pine (*Pinus contorta*) and white spruce (*Picea glauca*) and about 60,000 SNPs per species were obtained. After stringent filtering, 17,765 and 17,845 high-coverage SNPs were obtained without missing data for lodgepole pine and white spruce, respectively (Chen et al., 2013). The GBS approach also has been found to be convenient for genetic analysis and marker development of rapeseed, lupin, lettuce, switchgrass, soybean, and maize (Bus et al., 2012; Truong et al., 2012; Yang et al., 2012; Lu et al., 2013; Sonah et al., 2013). Pootakham et al., (2015) detected 21,471 SNPs in oil palm out of 3,417 fully informative SNP markers. And then, they placed 1,085 markers on a linkage map which spanned 1429.6 cM and had an average distance of markers every 1.26

cM. Iquira et al., (2015) mapped 8,339 SNP markers in soybean using GBS technique.

In pepper, many species has been existing over the world. However, most of previous studies were mainly conducted in *C. annuum*. *C. chinense* is the hottest pepper across the globe, and the demand has been increasing constantly. And previous developed linkage maps in pepper were consisted of RFLP, AFLP, SNP-based markers using populations with wide-crosses. Marker densities are relatively low below 500 markers per map. Current GBS analysis will provide additional information of imputation methods and SNP linkage map that can be utilized for genetic analysis of various trait in *Capsicum*. In this study, GBS mapping and capsaicinoids content data were used to detect QTLs controlling capsaicinoids content in the pericarp. We confirmed that *C. chinense* 'Jolokia' contains high capsaicinoids content in the pericarp tissues leading to exceedingly high pungency (Bosland et al., 2015). Furthermore, six QTLs were detected on chromosome 6 and 9 for capsaicinoids content in the pericarp.

LITERATURE REVIEW

1. Linkage map

Genetic linkage maps permit the elucidation of genome structure and organization and enable the identification of molecular markers linked to traits, leading ultimately to the elucidation of the genetic basis of the trait of interest. A genetic map represents a key tool for genetic studies and should contain enough information to identify and position genes/QTLs that control traits of agronomic interest (Pedraza-Garcia et al., 2010). During last several decades, linkage maps have been developed for many diverse plant species (Ward et al., 2013). Traditionally, transferable linkage map development has been developed through the scoring of restriction fragment length polymorphisms (RFLPs) (Tanksley et al., 1992), microsatellites (SSRs) and gene specific markers (Sargent et al., 2006) in a segregating progeny. Using such markers, saturated reference linkage maps for many plant species have been developed. Reference maps inform the selection of markers for mapping in other progenies (Eshed, Yuval, and Dani Zamir, 1995; Sargent et al., 2012; Silfverberg et al., 2006) and have been used to anchor, order and orientate physical map BAC contigs, and genome sequencing scaffolds for the assignment of pseudo-chromosomes for whole genome sequence initiatives (Khurana and K. Gaikwad, 2005; Shulaev et al., 2011; Velasco et al., 2007; Zhebentyayeva et al., 2008; Kalyanaraman et al., 2010).

2. Genetic map in pepper

Previously, several researchers were constructed linkage maps in pepper. Developed linkage maps were mostly consisted of RFLP, AFLP, SNP-based markers. Marker densities were relatively low below the 500 markers per map. Kang et al., (2001) have constructed a molecular linkage map of pepper (*Capsicum* spp.) in an interspecific F_2 population of 107 plants with 150 RFLP and 430 AFLP markers. The resulting linkage map consists of 11 large (206–60.3 cM) and 5 small (32.6–10.3 cM) linkage groups covering 1,320 cM with an average map distance between framework markers of 7.5 cM. Another map was constructed using a population of 126 F_8 recombinant inbred lines derived from a cross between YCM334 and Tean. 281 AFLPs, 101 EST-SSRs, 37 consensus SSRs and 1 CAPS marker were mapped and distributed in 19 linkage groups (LGs). Based on distribution of the consensus markers, 14 linkage groups were assigned into 12 chromosomes of pepper. The map covered 2177.5 cM with an average of 5.2 cM (Truong et al., 2010). Park et al., (2014) utilized expressed sequence tags (ESTs) to develop single-nucleotide polymorphism (SNP) markers. Three different types of PCR-based markers derived from pepper ESTs were developed. A total of 512 markers, comprising 214 IBP, 143 COSII, 48 eSNP, and 107 previously reported markers, were mapped on 12 linkage groups (LGs) of the ‘‘AC99’’ F_2 population. This newly constructed interspecific map (AC2) covered 2,335.6 cM with an average marker interval distance of 4.5 cM.

3. Single nucleotide polymorphism

Single nucleotide polymorphisms (SNPs) are the most abundant mutations in genomes of related species. To develop high-density linkage maps, researchers have shifted from anonymous markers such as AFLPs and microsatellites to direct analysis of sequence variations, including single nucleotide polymorphisms (SNPs) (Zhang et al., 2013; Huang et al., 2014; Pootakham et al., 2014). SNPs are abundant in plant genomes, and their usefulness as genetic markers have been well established over the past decade. Recent studies of sequence diversities have shown that SNP frequencies in plants are one in every 100–300 bp (Edwards et al., 2007). Their ubiquity makes them valuable for genetic mapping, particularly for the generation of high-density linkage maps. Advances in high-throughput next generation sequencing technologies have enabled initial efforts in SNP discovery.

4. Next-generation sequencing

The advent of the next-generation sequencing technology holds the promise for a methodological leap forward in genotyping and genetic mapping. The new sequencing techniques not only increase sequencing throughput by several orders of magnitude but also allow simultaneously sequencing a large number of samples using a multiplexed sequencing strategy (Craig et al., 2008; Cronn et al., 2008). These recent technical advances have paved the way for the development of a sequencing-based highthroughput genotyping method that combines advantages of

time and cost effectiveness, dense marker coverage, high mapping accuracy and resolution, and more comparable genome and genetic maps among mapping populations and organisms (Huang et al., 2009).

5. Genotyping-by-sequencing

Recently, a high-throughput and low-cost genotyping method named genotyping-by-sequencing (GBS) has been developed and has proven its efficiency in other crops such as maize and barley (Elshire et al., 2011). The proposed approach is simple and suitable for rapidly generating high-density genetic maps. GBS can yield hundreds to thousands of SNPs (single nucleotide polymorphisms) without the need for any prior characterization of candidate loci. A few protocols have been described to date including the original protocol involving the enzyme ApeKI (Elshire et al., 2011), demonstrated in maize and barley, and an alternative protocol involving enzymes PstI and MspI, demonstrated in barley and wheat (Poland et al., 2012). The choice of enzymes is a key factor in determining the degree of complexity reduction that is achieved. More recently, Sonah et al., (2013) showed that it was possible to further optimize the degree of complexity reduction and thus increase both the number and quality of markers obtained, by using PCR primers including selective bases during the preparation of GBS libraries. In addition, appropriate sequencing and genetic analysis tools are available to provide its efficiency to GBS method (Moumouni, K. H. et al., 2015).

6. Sliding window and bin map

Limitations of GBS include a relatively large proportion of missing data and a small, but rarely corrected, percentage of SNP genotyping sequencing errors (Chen et al., 2014). Recently, sliding-window approach was developed for imputing missing SNP data, where adjacent SNPs with the same genotype in an interval are combined into bins that demarcate recombination locations across the whole population (Yu et al., 2011; Huang et al., 2009). The bin-map method is demonstrated to be more powerful for detecting QTL than traditional methods and has also been employed for fine mapping of yield-associated loci in rice and sorghum and root-knot nematode resistance QTL in soybean (Zou et al., 2012; Gao et al., 2013; Xu et al., 2013).

MATERIALS AND METHODS

Plant materials and DNA extraction

Two hottest varieties of peppers, *C. chinense* 'Habanero', and *C. chinense* 'Jolokia' were used as parental lines in this study. These parental lines were used to create an interspecific F_2 mapping population. Total genomic DNA was extracted from 87 F_2 plants and their parental lines using modified cetytrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). For DNA extraction and genotyping, two lip disks of young pepper leaves were used.

Library preparation and sequencing

DNA concentrations were diluted to 80 ng/μL for library preparation. Preparation of *PstI-MseI* sequencing libraries was carried out according to the protocol described by SBG 100-Kit $_{(v2.0)}$ (Keygene, Netherland). Single-end sequencing was performed on two lanes of an Illumina HiSeq2000 at the Macrogen Inc. (Seoul, Korea).

Sequencing data processing and SNP calling

The CLC Genomics Workbench $_{(v6.5)}$ was used to check sequencing quality (QC) and trim the sequence reads. Quality trimming value was set to 0.01 (=Q20).

For trimming, ambiguous nucleotides were removed to allow maximal 2 nucleotides. Sequence length less than 70 was removed (Figure 1). A pipeline implemented in Python program was used for the processing of trimmed data. This pipeline uses several software tools as Burrow Wheelers Alignment (BWA), Sequence Alignment/Map (SAM) Tools, and Picard (Li and R. Durbin, 2009; Li et al., 2009; Truong et al., 2012). The raw reads were aligned to the reference genome using BWA and SAM Tools. Picard comprises of Java-based command-line utilities that manipulate SAM files, and duplicate marking and sorting were executed by Picard. The SNPs were further sorted and filtered using GenomeAnalysisTK (GATK) (DePristo et al., 2011). GATK criteria were based on the QUAL≥30, allele frequency (AF), and minimum sequencing depth 3 (Figure 1).

SNP dataset filtering and genotyping

Processed GBS reads were aligned against the reference genome of *C. annuum* 'CM 334' for SNP calling. The following criteria used for further SNP filtering: (i) SNP loci with no calling data in one or both parents was removed; (ii) The SNPs in the parents as ab x ab (both parents heterozygous), aa x ab (male parent heterozygous), ab x aa (female parent heterozygous), and aa, bb (different calling data in same parents) were removed. In addition, the SNPs in the parents as aa x aa or bb x bb (same calling data in both parents) were also removed; (iii).Threshold for removal was set to P ≤0.05. Lastly, we performed parent-based genotyping for this. The genotypes were labeled as A and B for homozygous

genotypes, and H for heterozygous genotypes. ‘Habanero’ and individuals following ‘Habanero’ genotypes were coded as “A”, whereas ‘Jolokia’ and individuals following ‘Jolokia’ genotypes were coded as “B” (Figure 1) (Rabbi et al., 2014; Moumouni et al., 2015).

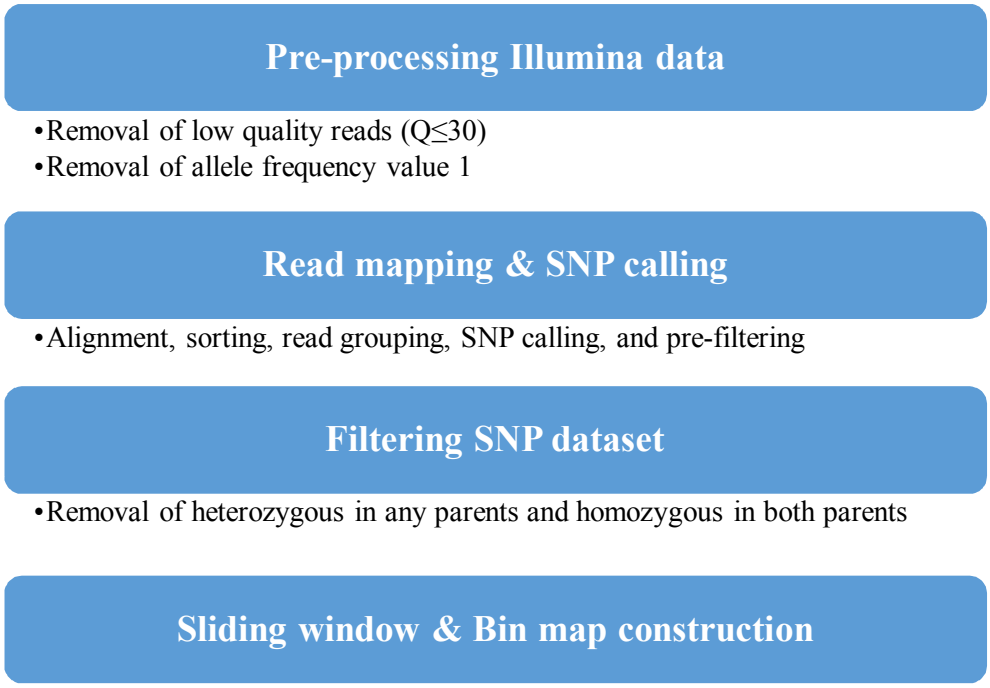


Fig.1. Workflow of SNP calling and bin map construction.
GBS data was pre-processed to remove low quality reads and allele frequency value of 1. The reads were aligned to the reference genome of *C. annuum* 'CM 334'. Using these data, SNP calling and pre-filtering were performed using GenomeAnalysisTK (GATK). After further SNP filtering process like the wrong genotype assignment, high-quality SNP dataset were generated. To construct SNP linkage map, sliding widow and bin map approach were used in this study.

Bin map and SNP linkage map construction

We used *C. annuum* chromosome information with the reference in this study. High-density genetic maps of populations with high linkage disequilibrium contain many redundant markers that provide no new information, but increase the computational requirements of mapping. Furthermore, a small percentage of genotypes are falsely called due to sequencing error. To address these issues, a modified version of the sliding window approach was applied (Huang et al., 2009; Chen et al., 2014). A slightly modified sliding window approach was used to investigate recombination breakpoints and construct a bin map. The genotype of each window was called with a window size of 18.8 Mb and moving window size of 50 kb. For each F_2 individuals, the ratio of SNP alleles from Habanero and Jolokia within the window was calculated. When more than 70% of SNPs had one parental genotype, the window was called as homozygous; otherwise, the window was called as heterozygous. Adjacent windows with same genotype were combined into a recombination bin. Recombination breakpoints were determined as the position where a different genotype emerged. Based on the recombination breakpoint position, a physical bin map was constructed (Huang et al., 2009).

To construct the linkage map, SNP representing each bin were used as genetic markers. A linkage map was constructed using CarthaGene software based on Linux operating system (De Givry et al., 2005). Recombination frequencies between markers were estimated using a likelihood-based approach. To construct a linkage group (LG), criteria value was set LOD score threshold of 3.0 and a

maximum distance of 30 cM. Among several LGs, the best map order was determined based on LOD scores and recombination frequencies. Distances between bin markers were determined in cM using the Kosambi mapping function. The bin marker loci have been named with the prefix "HJ" (for "Habanero x Jolokia"). The resulting linkage map was then drawn using the MapChart2.2 software (Voorrips. R, 2002).

Alignment of linkage map to physical map

After construction of the linkage map using bin markers, Linkage map and physical map were compared using Marker browser function in GBS browser developed by the PhyZen Genomics Institute (Seoul, Korea). The physical map information was based on the genome sequence of *C. annuum* (Kim et al., 2014).

Isolation of RNA and cDNA synthesis

Placenta and pericarp tissues were used for the analysis of the *Pun1* gene and capsaicinoids content. Pericarp was divided to three sections from the basal part to the end of fruits. Total RNA was extracted from placenta and middle part of pericarp at the break stage of fruits using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. cDNA synthesis was performed immediately after RNA isolation. Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) was used for cDNA synthesis.

Expression analysis of *Pun1* by semi-qPCR

To test *Pun1* expression, cDNA product was used to perform semi-qPCR. Pepper actin primer set was used for the internal standard. Image J program was performed to check the bands intensity by calculating the gross area. Semi-qPCR was performed by the combination of a primer set designed to amplification of the 1st exon sequence of *Pun1*. PCR was conducted in a 25 μL reaction volume containing 5 μL of cDNA, 10 pmol of each primer, 2 μL of dNTP, 2.5 μL of Hipi buffer and 0.2 μL of *Taq* polymerase. PCR condition was as follows: 95℃ for 4 minutes followed by 35 cycles of 95℃ for 30 seconds, 55.5℃ for 30 seconds and 72℃ for 1 minute and final extension of 10 minutes at 72℃.

Measurement of capsaicinoids content by HPLC analysis

For HPLC analysis, three pepper fruits from F_2 individuals were harvested when the fruits were fully colored. The placenta and pericarp tissue of each fruit were divided to measure each part of capsaicinoids content. Freeze-dried powder was extracted with 7.5 mL of ethyl acetate and acetone mixture (6:4) by shaking at 37℃ for 24 h, and 3 mL of supernatant was evaporated in an Automatic Environmental SpeedVac System AES1010 (Operon, Gimpo, Korea). The pellet was dissolved in 1 mL 99.9% methanol (Honeywell Burdick & Jackson, Muskegon,

MI, USA), and a 500 μL aliquot was used for HPLC analysis. HPLC analysis was performed in National Instrumentation Center for Environmental Management (NICEM, Seoul, Korea) according to the method described by Han et al., (2013).

QTL analysis

Capsaicinoids content from placenta and pericarp was used for QTL analysis. Composite interval mapping (CIM) was performed with Windows QTL Cartographer 2.5 (Wang et al., 2012). LOD threshold was set by 2.5. The closest bin marker to the QTL with the highest LOD was selected to represent each QTL in placenta and pericarp. The proportion of phenotypic variation explained by each QTL was estimated using R^2 (%) value. Box plots of major QTLs were calculated by quartile of capsaicinoids content in pericarp.

RESULTS

Genotyping by sequencing

A single end sequencing was executed to generate GBS dataset of the parents and 87 F_2 individuals obtained from the cross 'Habanero x Jolokia'. The GBS protocol achieved complexity reduction using both a two-enzyme restriction digest (*PstI-MseI*). After sequencing on two lanes, the data from read 2 was removed and the total number of reads was 244,534,479. The average number of reads was approximately 2.7 Mb per individual. The number of reads per sample ranged between 0.4 and 6.8 Mb. After trimming the data by CLC Genomics Workbench (v 6.5), the number of reads per sample was reduced to 0.3 and 5.9 Mb with an average of 2.3 Mb. Most of read length were about 100 bp. These data was enough to be used for SNP calling (Table 1). For mapping the GBS data to the reference genome, SNP and genotyped data were generated using several software tools as described in MATERIALS AND METHODS. The parent-based genotyping was performed using the sequence data. The reads of parents and 87 F_2 individuals were aligned to the reference genome sequences (*C. annuum*) to obtain the physical positions of each SNP. A total of 217,152 SNPs were identified in the GBS datasets.

Table 1. Number and average length of reads before and after trimming.

Name	Number of reads	Average length	Number of reads after trim	Percentage trimmed	Average length after trim
Total	244,534,479.0	-	213,488,294.0	-	-
Average	2,687,192.1	82.4	2,346,025.2	86.00	88.8
Minimum	432,372.0	50.0	282,301.0	37.33	88.5
Maximum	6,814,225.0	84.6	5,885,089.0	90.14	89.0

Bin map construction

Using a sliding window approach, the recombination maps were converted into a skeleton bin map for further genetic analysis (Figure 2; Van Os et al., 2006). Adjacent SNPs with the same genotype were considered as a single recombination bin. SNPs representing each bin were used for linkage map construction. Using 217,152 SNPs in 87 F_2 populations, we obtained 1,007 bin markers along the 12 chromosomes as described in method. The average physical length of bin markers was 2.75 Mb, ranging from 100 kb to 73.6 Mb. Most of bin marker length were more than 1 Mb (Figure 3 and Table 2).

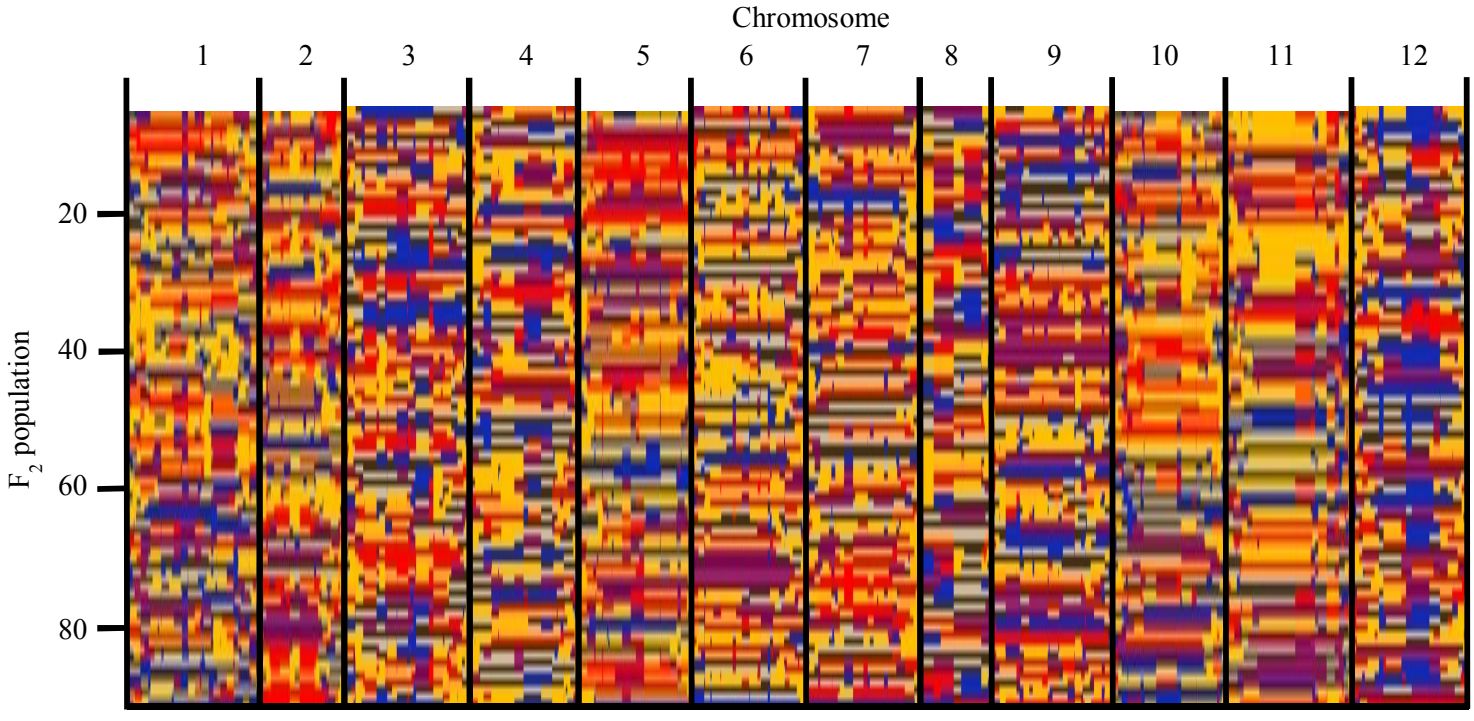


Fig.2. Recombination bin map of *C. chinense* 'Habanero' × 'Jolokia' F_2 population.

Bin map consists of 1,007 bin markers inferring from 217,152 SNPs in F_2 population. Red, 'Habanero' genotype; blue, 'Jolokia' genotype; and yellow, heterozygous, respectively.

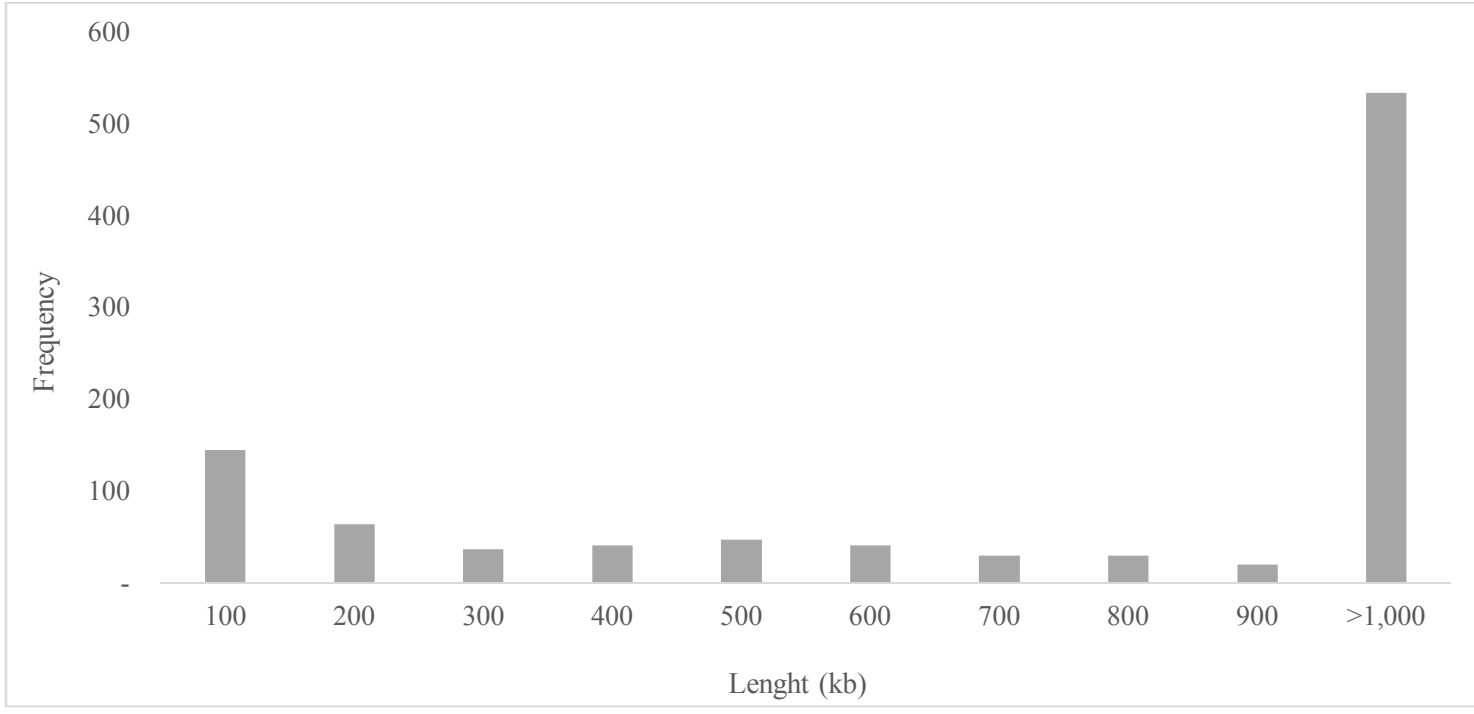


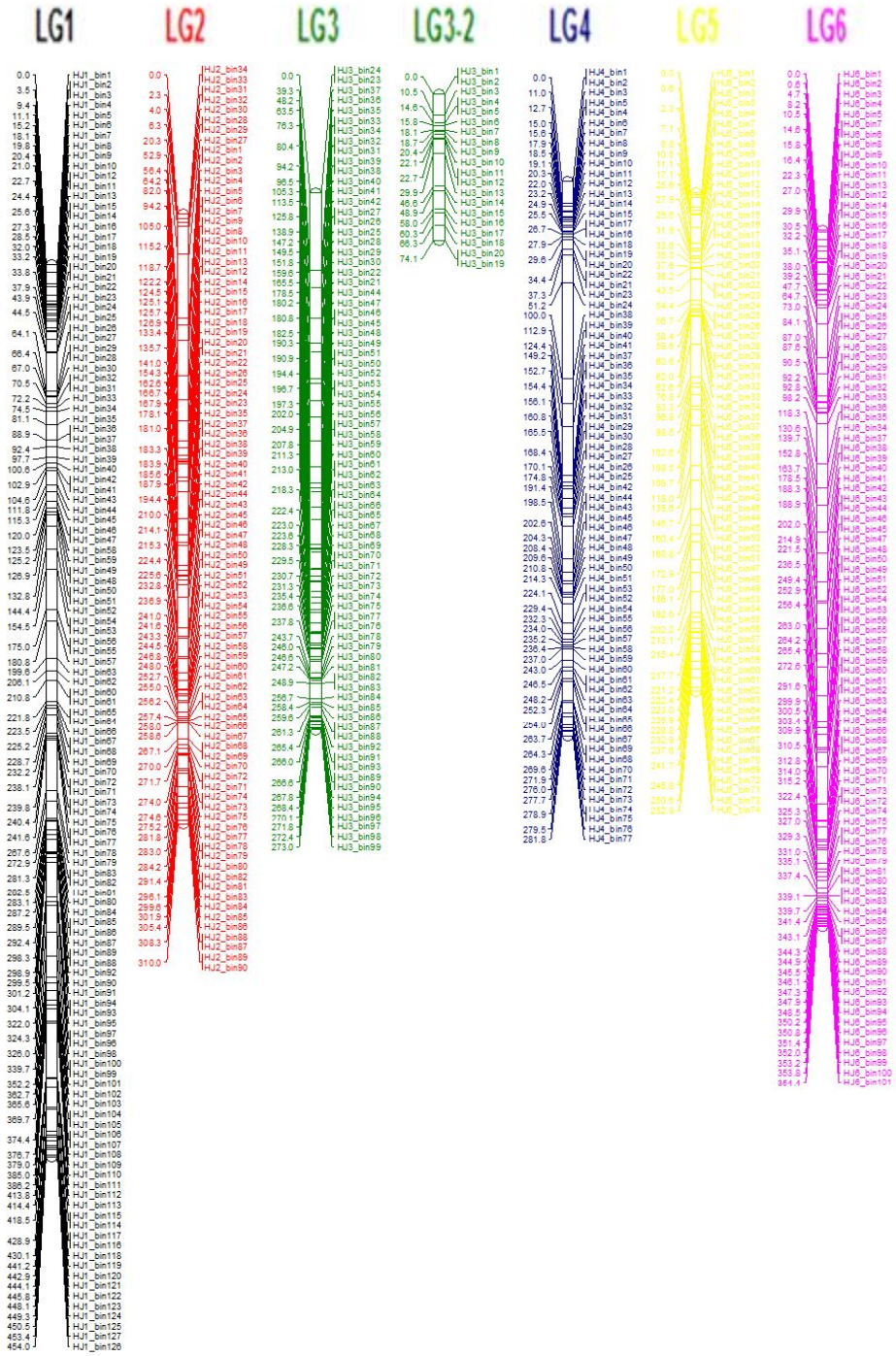
Fig.3. Distribution of bin marker length.
The mean length of bin marker was 2.7 Mb and ranged from 100 kb to 73.6 Mb with median of 1.2 Mb.

Table 2. The total number of SNPs and bin markers mapped to the 'Habanero x Jolokia' linkage map.

Chromosome	Number of SNPs	Linkage group (LG)	Number of bins	Average distance between bin markers (cM)	Average length of bins (Mb)	LG length (cM)	Total physical length (Mb)
1	23,316	1	127	3.6	2.18	454.3	272.3
2	17,400	2	90	3.5	1.88	310	165.7
3	25,056	3	78	3.5	2.59	273.1	253.4
		3-2	20	3.9		74.3	
4	17,748	4	77	3.7	2.96	282.1	222.1
5	17,226	5	74	3.5	3.18	253.3	232.5
6	23,055	6	101	3.5	2.36	354.4	236.2
7	19,488	7	88	3	2.63	262.2	229
8	8,178	8	28	3.3	3.97	89.6	143
		8-2	7	4.7		28.1	
9	16,443	9	89	4.2	2.84	368.3	250.1
10	18,183	10	76	5.5	3.09	340.4	231.6
11	13,572	11	65	4.2	3.95	266.7	256.5
12	17,487	12	80	3.8	2.93	297.3	231.8
Average	18,096	-	71.4	3.9	2.75	261	227
Total	217,152	14	1,000	-	-	3654.1	2,724.2

SNP linkage map construction

We observed 217,152 SNPs and 1,007 bins. A linkage map was constructed with these bins as markers. CarthaGene software was used to construct a genetic linkage map. We obtained 1,000 bin markers into a genetic linkage map of 14 linkage groups, corresponding to the 12 pepper chromosomes. Chromosome 3 and 8 were divided to two linkage groups. The total distance of the linkage map was 3654.1 cM, approximately 3.9 cM per bin. This resulted in a stable map configuration with 1,000 mapped bin markers, an average chromosome length of approximately 261 cM, and the average number of bin markers of 71.4 per linkage group (LG) (Table 2). Every groups was somewhat larger than those in the previous mapping studies in *Capsicum* spp. This results may be due to sequencing errors and missing data.


LG1
LG2
LG3
LG3-2
LG4
LG5
LG6

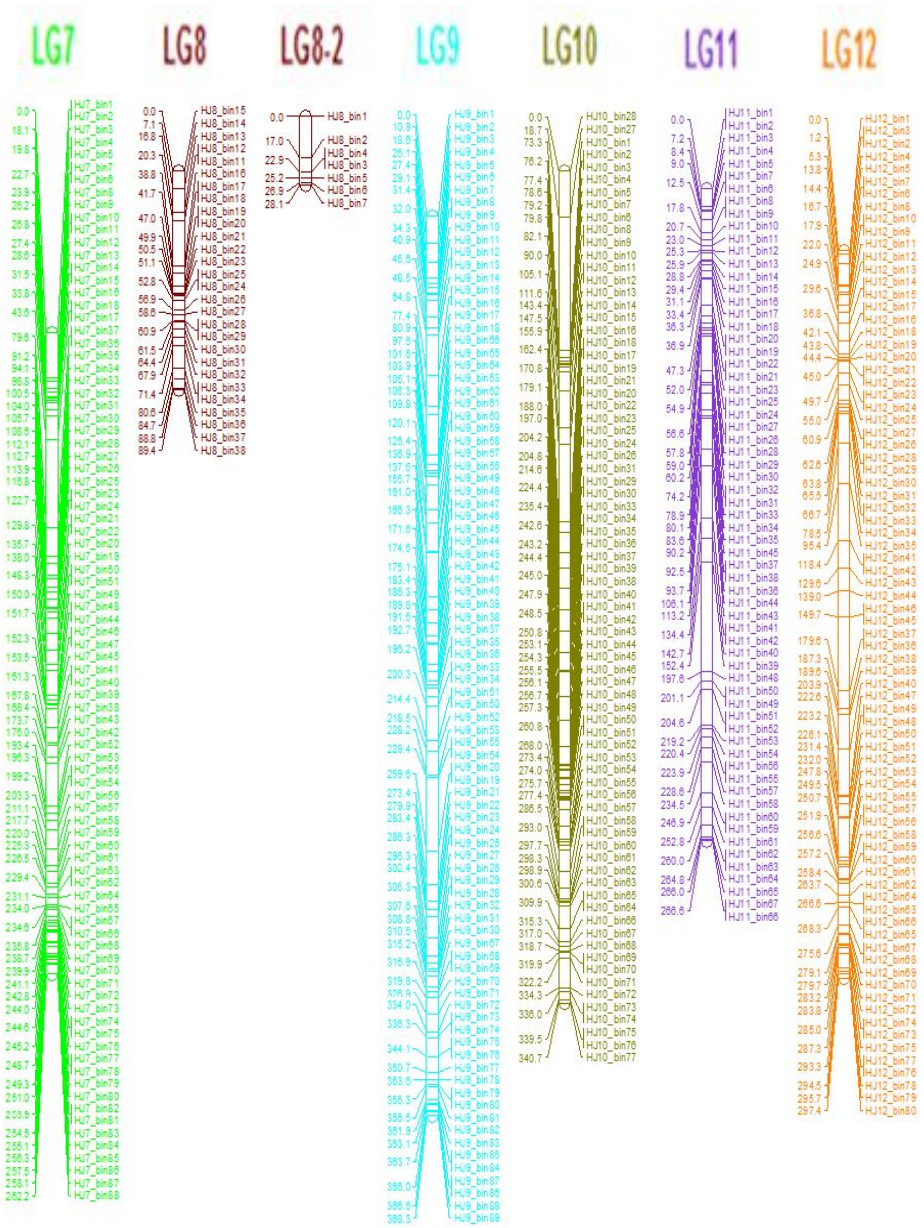
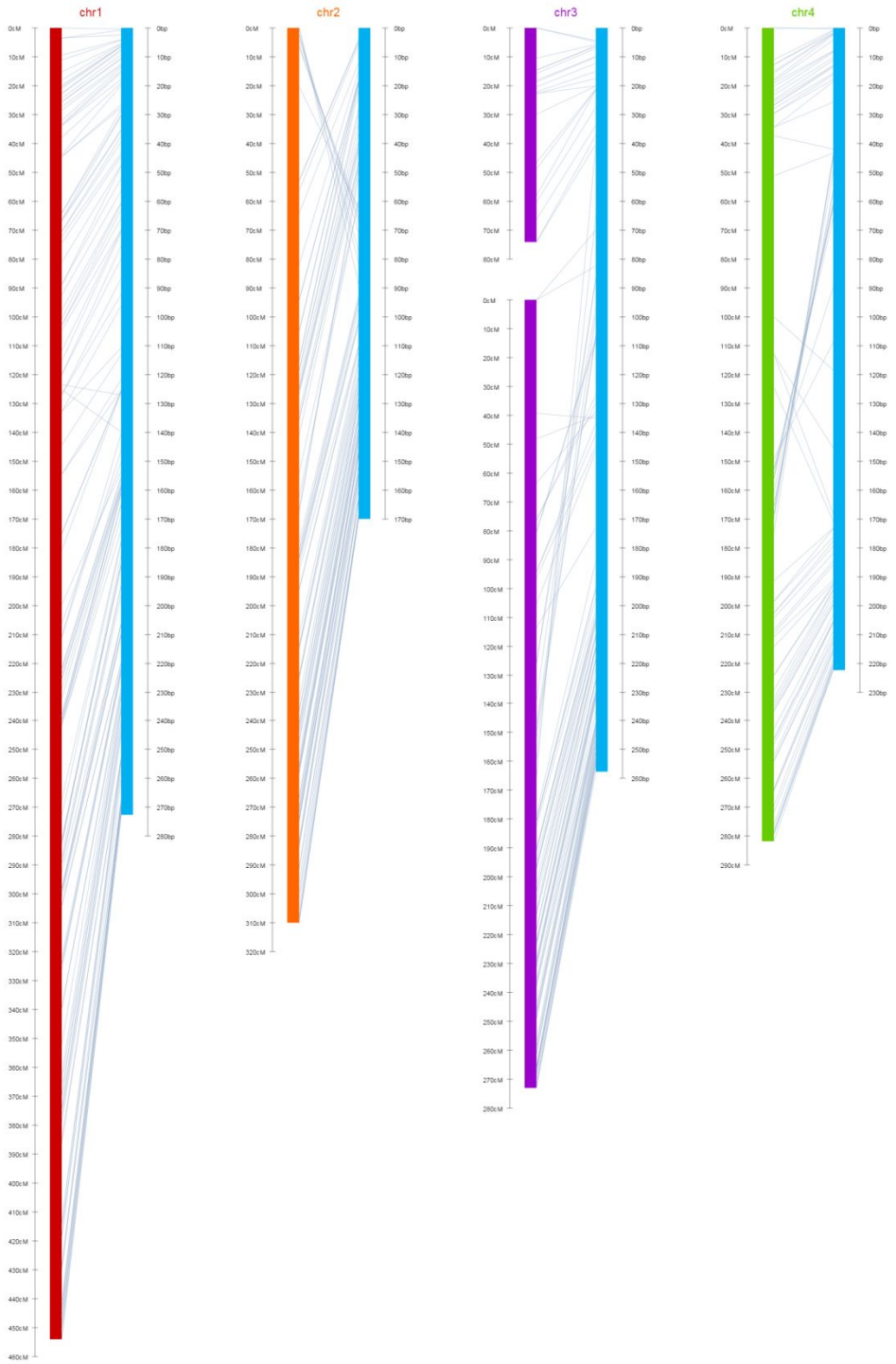
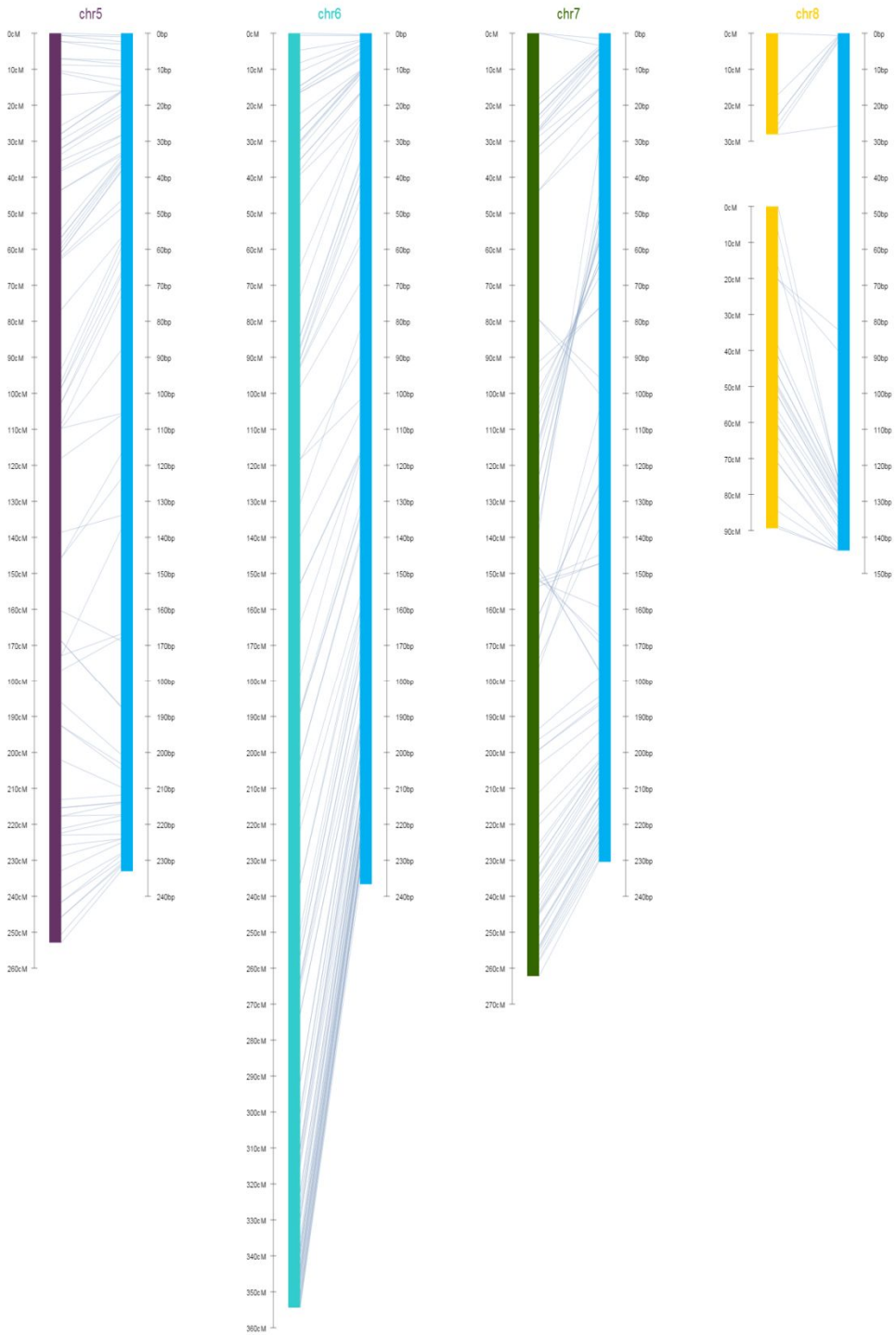


Fig.4. Construction of SNP linkage map.
'Habanero x Jolokia' LGs are denoted LG1-LG12 corresponding to Chr1-Chr12 in pepper. Additional linkage groups are LG3-2 and LG8-2. Each bin marker loci which corresponds to the 12 pepper chromosome are painted by different color. Cumulative distances are indicated to the left in (cM). Bin marker loci are labeled HJ00_bin 000 to the right on each linkage group.

Relationship between linkage map and physical map

The linkage map developed using the 1,000 bin markers was aligned to the reference genome and detected physical positions of bins (Figure 5). The most of bin markers per each chromosome were well correlated to the physical position of the sequence. Although high marker density map was constructed by GBS, some regions has no bin markers on the *C. annuum* physical map. The gaps were mainly observed at near the centromere regions. Since that region is composed of hetero-chromatin and tandemly repeating regions, there may no polymorphic sequences and/or restriction enzyme sites used in this study. However, most of linkage groups (LGs) in "Habanero x Jolokia" showed collinearity except for some differences like an inversion. Linkage groups (LGs) Chr 1, Chr 2, Chr 5, Chr 6, Chr 8, Chr 10, Chr 11, and Chr 12 showed collinearity except for few inversions. On the other hand, linkage groups (LGs) Chr 3, Chr 4, Chr 7, and Chr 9 showed more inversions than others relatively (Figure 5).


chr1
chr2
chr3
chr4


chr5
chr6
chr7
chr8

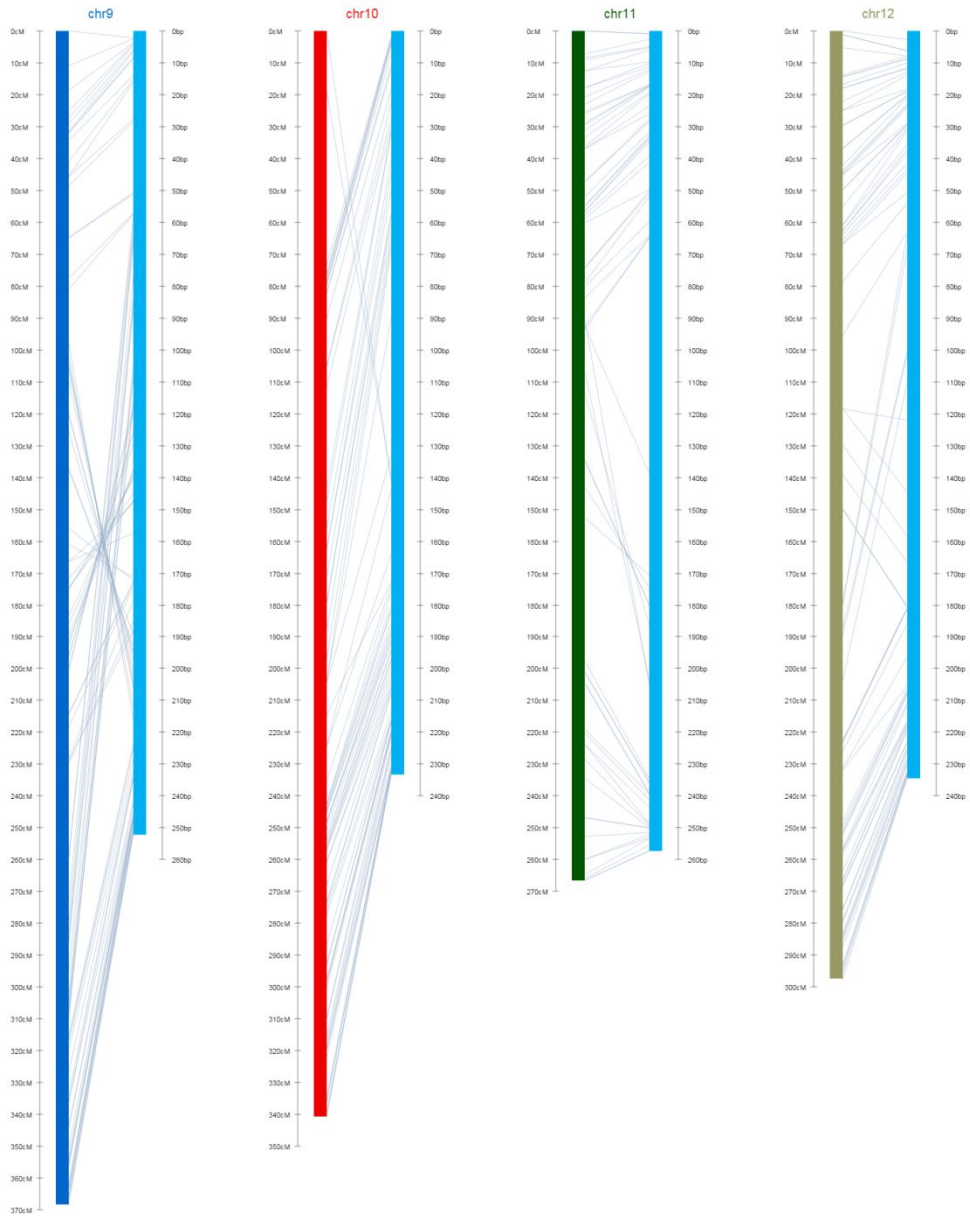


Fig.5. Comparison of the bin map with CM334 physical map.

In each pepper chromosome, 'Habanero x Jolokia' genetic map and distances are indicated to the left in cM. On the right, '*C. annuum*' physical map and distances are given in bps. Lines between genetic map and physical map represent the positions of each bin marker corresponding to the same location. Collinearity and inversions are observed in this study.

Measurement of capsaicinoids content in placenta and pericarp

Capsaicinoids content in placenta and pericarp were measured by HPLC analysis (Figure 6). Pericarp tissues without containing placental tissues were prepared. Two or three fruits were harvested when they are at fully colored and breaker stages. Capsaicinoids content in Habanero placenta and in pericarp at the breaker stage was 125,560 μg/gDW (μg per g dry weight) and 513 μg/gDW, respectively. Capsaicinoids content in Jolokia placenta and pericarp tissues were 104,522 μg/gDW and 34,154 μg/gDW, respectively. Capsaicinoids content in F_1 hybrid placenta was 98,601 μg/gDW whereas 244 μg/gDW in pericarp. Capsaicinoids content of Jolokia shows 66~139 times higher than that of Habanero and F_1 hybrid in the pericarp. Capsaicinoids content in placenta and pericarp tissues of a sweet pepper was also measured (Table 3).

To compare expression of the *Pun1* gene, semi-qPCR was carried out (Figure 7). Semi-qPCR analysis using RNA extracted from placenta tissues of all samples showed expected size of amplicons (571 bp) except for that of sweet pepper. By contrast, *Pun1* amplicon was obtained only from RT-PCR using RNA extracted from pericarp tissues of Jolokia.

Capsaicinoids content in placenta and pericarp tissues of F_2 plants were also measured by HPLC analysis using placenta of 93 plants and pericarp of 89 plants. In the pericarp, most of F_2 plants showed below 15,000 μg/gDW capsaicinoids. In

the placenta, capsaicinoids content in F_2 population showed a normal distribution (Figure 10 and 11).

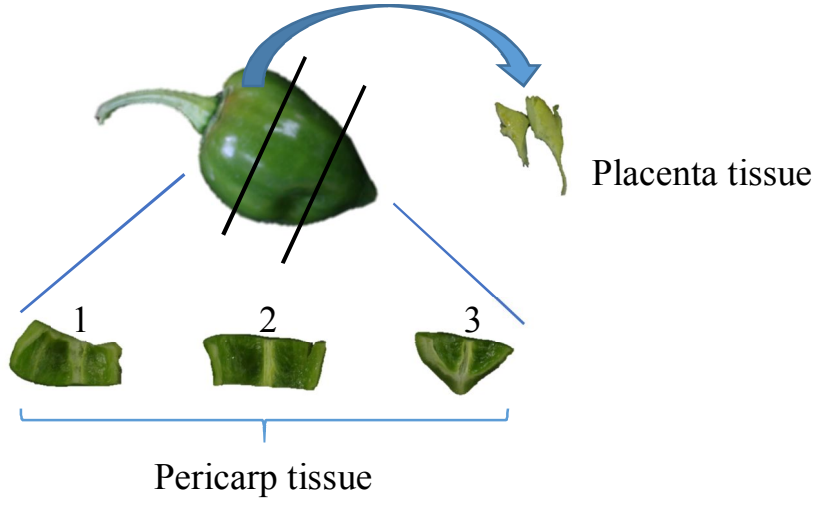


Fig.6. Sampling method of pericarp tissues for capsaicinoids analysis.
Habanero, Jolokia, F_1 hybrid, and Yolo Wonder were used in this study. Placenta and pericarp tissues were used for the analysis of the *Pun1* gene and capsaicinoids content. Pericarp was divided into three parts. RNA was extracted from the middle part (# 2) of pericarp at the break stage of fruits.

Table 3. Capsaicinoids content in placenta and pericarp

Species	Stage	Tissue	Contents (μg/gDW)*		
			Capsaicin	Dihyhdro capsaicin	Total
Habanero	Breaker	Placenta	81,979 ± 1,674	43,580 ± 2,095	125,560 ± 3,770
		Pericarp	389 ± 170	246 ± 0	513 ± 344
	Fully colored	Placenta	37,611 ± 16,140	21,621 ± 9,496	59,232 ± 25,141
		Pericarp	239 ± 151	144 ± 0	287 ± 165
Jolokia	Breaker	Placenta	60,634 ± 15,457	43,887 ± 9,047	104,522 ± 24,039
		Pericarp	23,946 ± 12,291	10,208 ± 5,662	34,154 ± 17,918
Habanero x Jolokia F_1	Breaker	Placenta	61,887 ± 14,381	36,714 ± 16,828	98,601 ± 30,782
		Pericarp	235 ± 276	27 ±0	244 ± 268
Paprika type	Fully colored	Placenta	224 ±44	165 ± 21	389 ± 22
		Pericarp	14 ± 4	12 ± 5	26 ± 10

*Capsaicinoids content were measured by HPLC analysis in placenta and pericarp. Pericarp between placental tissues was prepared. Fruits were harvested when they changed to be fully colored and breaker stages

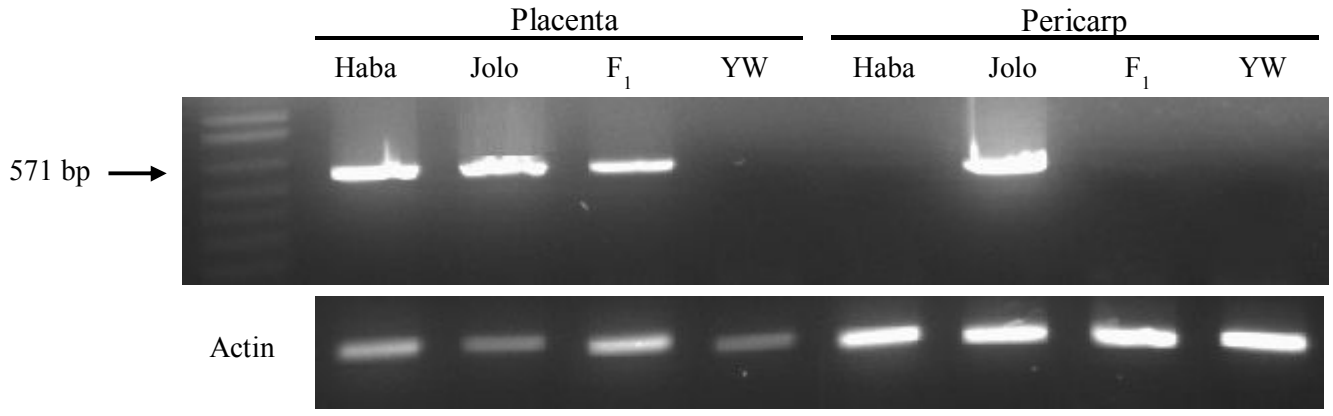


Fig.7. Expression analysis of *Pun1* by RT-PCR.
The primer designed in 1st exon sequence of *Pun1* was used in this study. Semi-qPCR was performed to compare *Pun1* expression difference in placenta and pericarp tissues. Pericarp of Jolokia showed a band with571 bp in size while Habanero, F_1 hybrid and Yolo wonder showed no band in pericarp. Yolo wonder was used for negative control which is non-pungent plant. RNA extraction was performed at the breaker stage. *Haba* Habanero, *Jolo* Jolokia, *F_1* F_1 hybrid, *YW* Yolo wonder

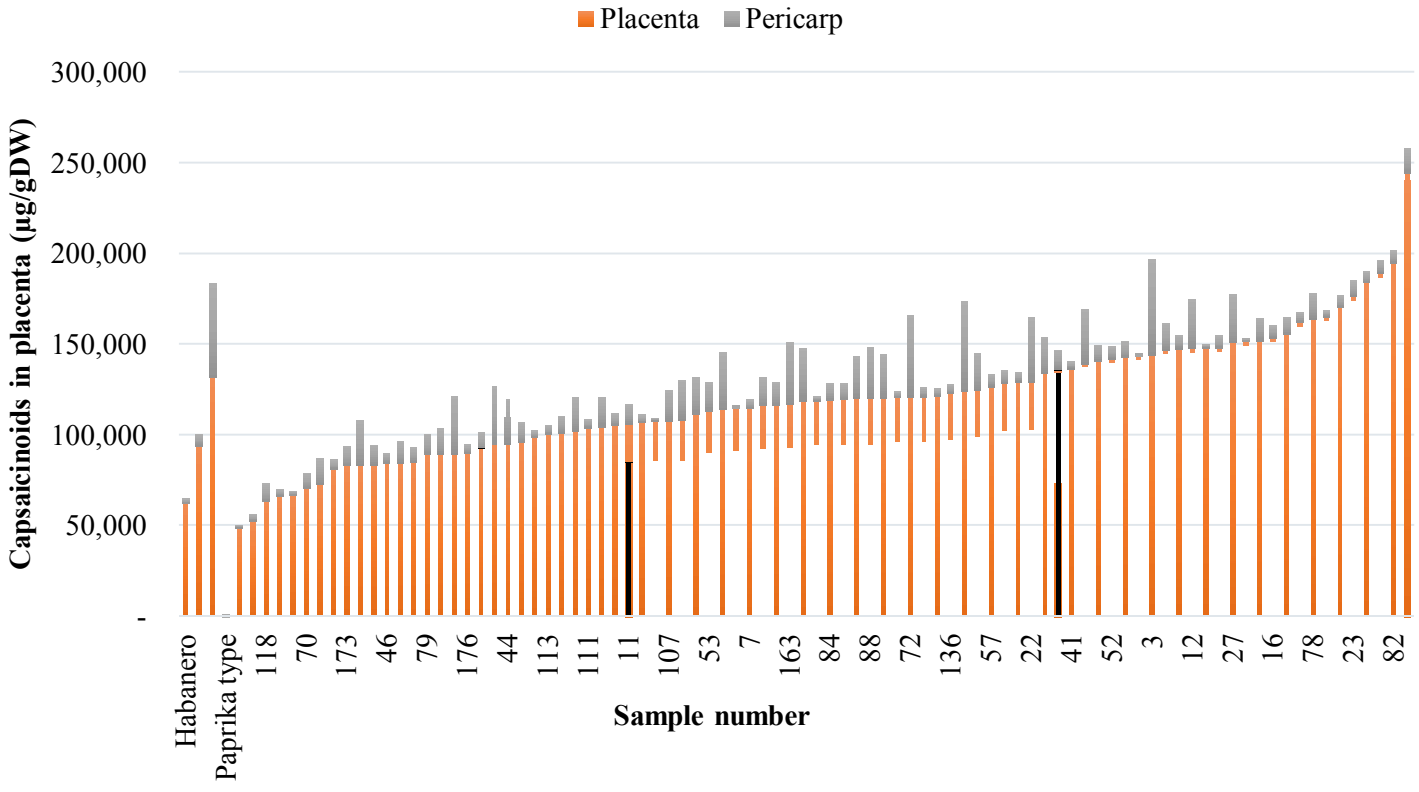


Fig.8. Capsaicinoids content distribution in F_2 population.
Capsaicinoids were measured in placenta and pericarp tissues of 88 F_2 plants. Fruits were harvested when they were fully colored. Three fruits of each individual were pooled for reducing sample variations. Orange bar means capsaicinoids content in placenta. Gray means capsaicinoids of pericarp.

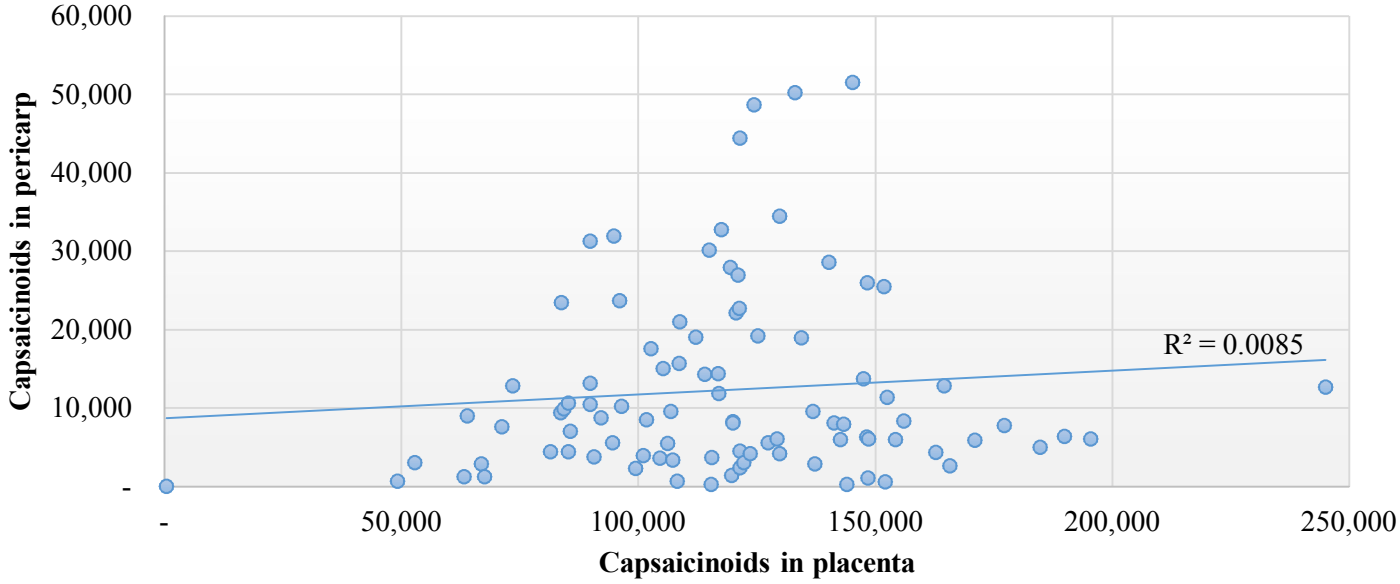


Fig.9. Correlation coefficient between capsaicinoids content in pericarp and placenta.
Correlation value (r) was 0.092 demonstrating no correlation between capsaicinoids content in placenta and pericarp tissues. The x-axis shows capsaicinoids in placenta and y-axis is capsaicinoids in pericarp.

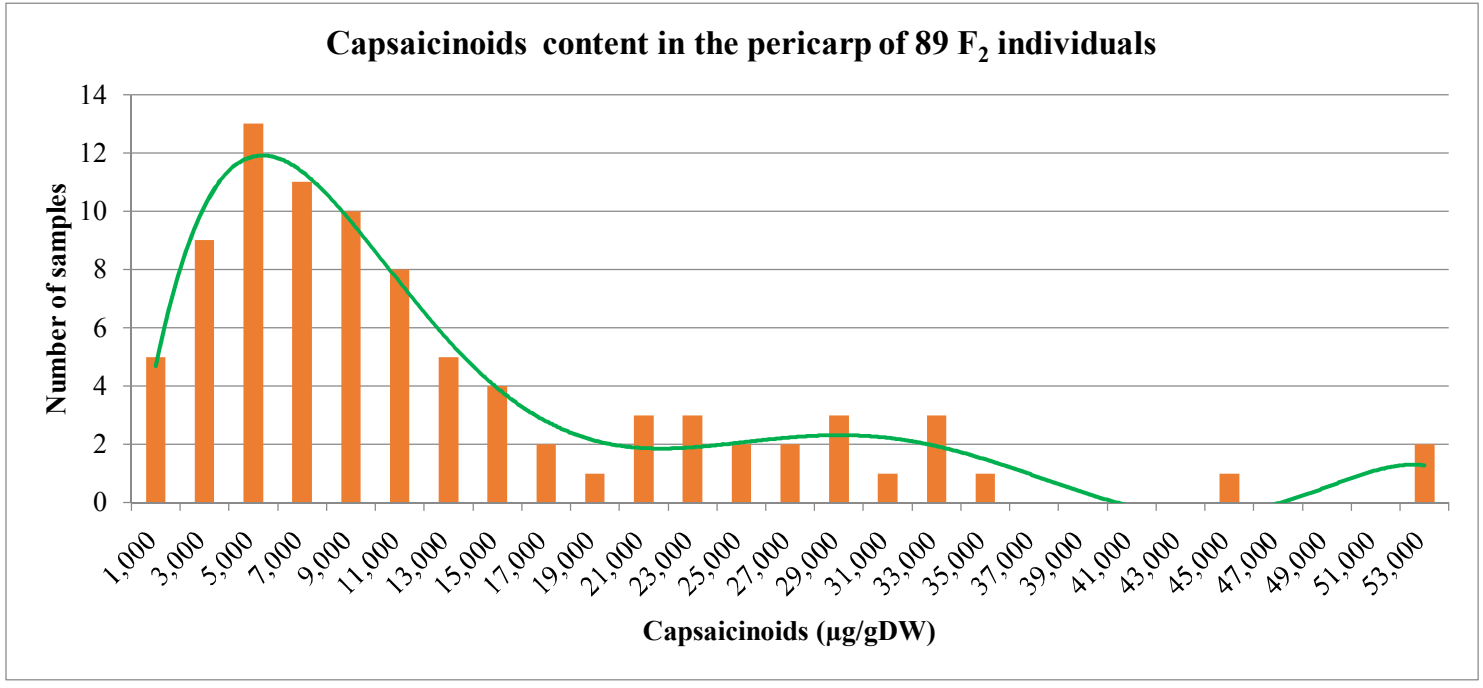


Fig.10. Distribution of capsaicinoids content in F_2 population.
Capsaicinoids content of 89 F_2 individuals in pericarp were measured by HPLC analysis. Fruits were harvested when they changed to be fully colored. The three pericarp per individual were pooled for reducing sample variations. All content was measured as µg/gDW (µg of compound per g dry weight of placenta and pericarp). The x-axis interval is 2,000 µg/gDW.

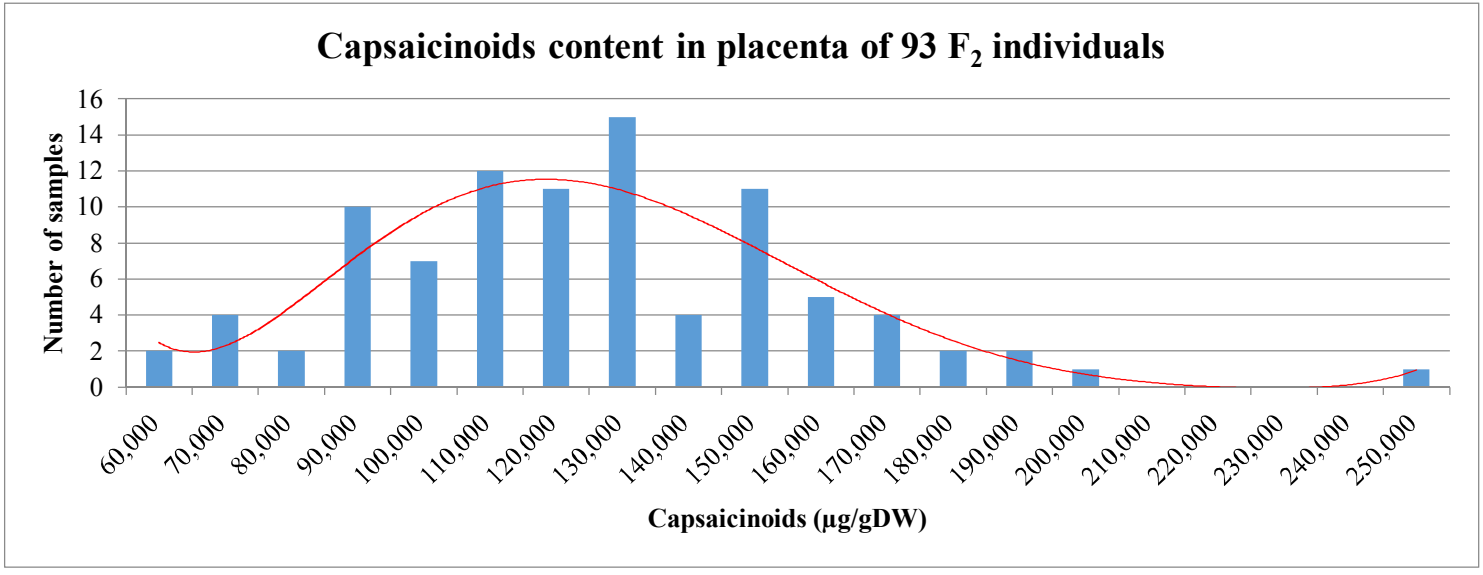


Fig.11. Distribution of capsaicinoids content in placenta tissues of F_2 population.
Capsaicinoids content of 93 F_2 individuals in placenta were measured by HPLC analysis. Fruits were harvested when they changed to be fully colored. The three placenta per individual were pooled for reducing sample variations. All content was measured as µg/gDW (µg of compound per g dry weight of placenta and pericarp). The x-axis interval is 10,000 µg/gDW.

QTL analysis for capsaicinoids content in placenta and pericarp

Capsaicinoids content in placenta and pericarp were measured by HPLC. QTL analysis was performed for each tissue. One and six QTLs controlling capsaicinoids content in placenta and pericarp tissues were found, respectively. A total of seven QTLs were mapped to the linkage map which was developed in this study. One QTL on the chromosome 12 was detected for capsaicinoids content in placenta. This QTL was located nearby HJ 12_bin 35 marker. LOD score was 3.41. By contrast, six QTLs were detected for capsaicinoids content in pericarp, and most of them were located on the chromosome 9. One QTL was located on the chromosome 6. *qCPE3* on the chromosome 9 was located at 109.31 cM corresponding to 200.8 Mb. This QTL showed the highest LOD score. Therefore, *qCPE3* appeared to be a major QTL of capsaicinoids in pericarp (Table 4, Figure 12).

Two of five QTLs for capsaicinoids content in pericarp showed high LOD score which is over the 4.3. These QTLs appeared to be major QTLs controlling capsaicinoids in pericarp. The capsaicinoids content of Jolokia genotype showed higher than those of Habanero and hetero genotype (Figure 13).

Table 4. Summary of QTLs for capsaicinoids content.

Trait	QTL	Chr	Bin marker	Position (cM)	Position (Mb)	Interval (cM)	LOD score	R^2 (%)
Capsaicinoids in placenta	*qCPL1*	12	HJ12_bin35	95.41	62.8	79.8-107.4	3.41	28.0
Capsaicinoids in pericarp	*qCPE1*	6	HJ6_bin84	339.11	224.3	331-339.7	2.53	20.8
	qCPE2	9	HJ9_bin66	99.51	216	84.9-101.6	4.35	48.7
	qCPE3	9	HJ9_bin61	109.31	200.8	106.3-119.9	4.48	50.5
	qCPE4	9	HJ9_bin45	171.61	147.3	160.6-183.4	3.43	33.5
	qCPE5	9	HJ9_bin39	191.51	137.7	183.4-200.2	3.37	44.2
	qCPE6	9	HJ9_bin26	302.41	84.6	286.3-307.6	3.56	56.5

CPL: Capsaicinoid in placenta
CPE: Capsaicinoid in pericarp

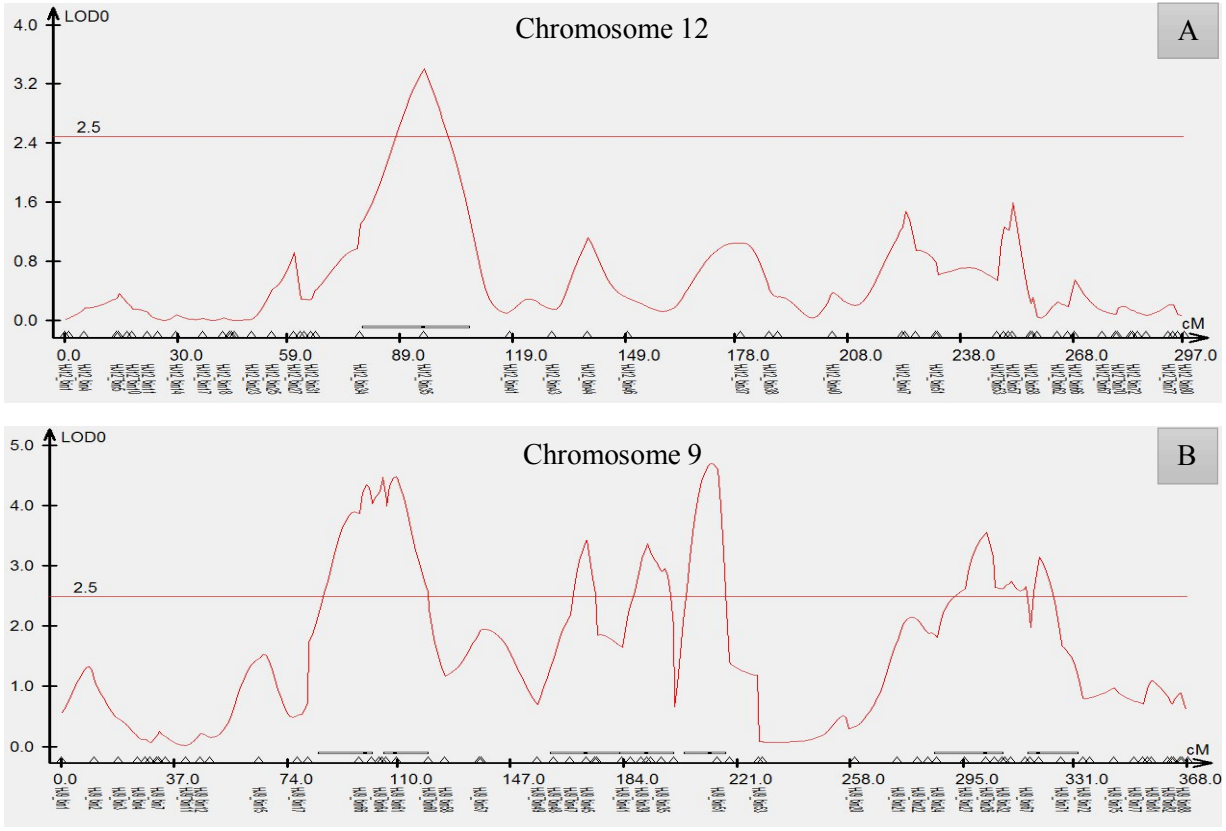


Fig.12. Positions of QTLs for capsaicinoids content in placenta and pericarp tissues.
One QTL was identified for capsaicinoids in the placenta on chromosome 12. This QTL is located nearby HJ12_bin35 marker. In the case of pericarp, five QTLs were detected and located on chromosome 9. Major QTL which has highest LOD score was located in 109.31 cM corresponding to 200.8 Mb. A: QTL in placenta, B: QTLs in pericarp

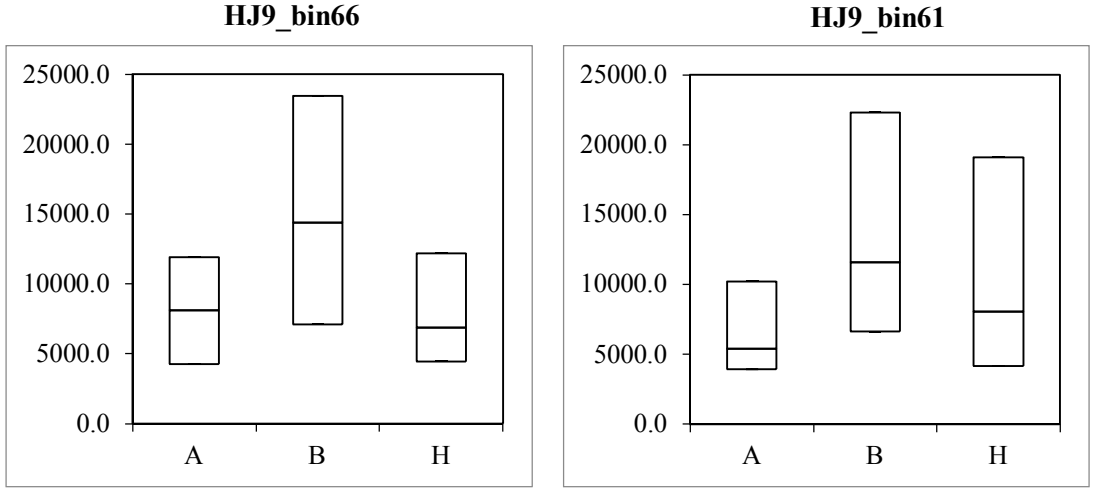


Fig.13. Box plots of major QTLs for capsaicinoids content in pericarp.
Two major QTLs show high LOD score over the 4.3. For both of the QTLs, the Jolokia genotype showed higher capsainoids content than those of Habanero and hetero genotype. A: Habanero, B: Jolokia, H: Hetero

DISCUSSION

Genotyping-by-sequencing (GBS) offers several competitive advantages over other genotyping methods. GBS makes us simultaneously discover and map a significant number of novel SNPs in any given crop species. There are two factors to consider before performing genotyping-by-sequencing: firstly the choice of restriction enzymes to reduce genome complexity and secondly the availability of a reference genome sequence (reviewed by Davey et al., 2011; Poland and Rife, 2012). Compared to RAD-seq, GBS method is a time and cost effective method because GBS offers small amount of library preparation steps. And GBS can be performed with small amounts of DNA (100~200 ng) and is amenable to a high level of multiplexing. In this study, we used two restriction enzymes, one rare cutting and the other frequent cutting to reduce genome complexity. This strategy not only allows us to effectively reduce genome complexity, but also create an even distribution of genomic fragments covering the length of a genome. This feature of GBS is especially important in the case of complex genomes such as pepper which has more than 3.5 Gb in genome size. The user can gain information about the sequence context of each SNP marker because GBS is based on the sequence technologies. Lastly, GBS shows a high correlation between densities of genes and SNPs. Therefore, this method could be used for mapping any known gene of interest.

In the previous studies on pepper, RFLP, AFLP, SNP-based markers were mostly used for linkage map construction. Especially, previous studies had several limitations: (i) the number of markers was relatively small usually less than 500 markers per linkage map; (ii) marker densities were relatively low; (iii) wide crosses were used for constructing a linkage map. High density maps are required for the precise mapping of important agronomic traits to be targeted in marker-assisted breeding. In this study, GBS was successfully used to produce a high density, saturated linkage map for a mapping population derived from a cross between two *C. chinense* lines. We have shown that GBS is an effective approach for the generation of marker-dense genetic map in pepper. The high-throughput GBS technology enabled us to place 217,152 SNPs loci. SNPs was expected to be more evenly distributed along the physical map. However, gaps near the centromere region were still observed in the high-density map. Some gaps may be resulted from lack of polymorphism in the mapping populations, which can be further improved by integrating other mapping populations. Gaps could also be filled in by using GBS libraries using different restriction enzymes, as shown in wheat (Saintenac et al., 2013). If gaps don't contain a suitable marker density, it may be useful to test additional enzyme combinations for GBS when a large investment has been made in phenotyping or trait analysis. Over 217,000 SNPs were identified and the genetic map contains 1,000 bin markers distributed on fourteen linkage groups spanning a total of 3654.1 cM. These bin makers met all of our filtering criteria and our markers have sufficient quality to be used for mapping.

Finally, our high-density map will contribute to a fundamental knowledge of the genome structure and have numerous applications in genomic research. High-density maps are valuable in applied breeding programs to enable genomic selection and precise mapping of agronomically important genes for marker-assisted selection.

Pepper is an important spice crop widely used in the world. Until recently, it has known that pungency controlling genes are specifically expressed in placenta tissues. However, according to a recent study, pericarp tissues also contain capsaicinoids in extremely pungent peppers such as *C. chinense* 'Jolokia' and 'Scorpion'. Therefore, we confirmed whether capsaicinoids are synthesized in the pericarp tissue of 'Jolokia' by *Pun1* expression analysis. As a result, *Pun1* was expressed in only the pericarp tissue of Jolokia and not in those of Habanero and F_1 hybrid plants. Distribution of capsaicinoids content in the pericarp of F_2 individuals indicated that capsaicinoids might be regulated recessive gene(s). Based on the correlation coefficient value (r=0.092), there was no correlation between capsaicinoids content in placenta and pericarp tissues.

To analyze QTLs controlling capsaicinoids content in placenta and pericarp, we used the linkage map developed by GBS. In the previous study, major QTLs for capsaicinoids content were detected on chromosome 3, 4, and 7 (Ben-Chaim et al., 2006; Yarnes et al., 2013). In these studies, capsaicinoids measurement was measured using whole fruits. Also, mapping populations were derived from one highly pungent and the other less pungent (or non-pungent) lines but there was no

population crossed between hot varieties peppers. In this study, placenta and pericarp tissues were separately analyzed QTLs for capsaicinoids content. In the placenta tissue, one QTL was identified for capsaicinoids on chromosome 12. Six QTLs in pericarp tissues were detected and located on chromosome 6 and 9. These results demonstrates that QTLs controlling capsaicinoids content in placenta and pericarp tissues could be different. Bin markers linked to the major QTLs can be used for selection plants containing high capsaicinoids content.

REFERENCES

Ben Chaim A, Borovsky Y, Falise M, Mazourek M, Kang BC, Paran I, Jahn M. 2006. QTL analysis for capsaicinoid content in Capsicum. Theoretical and Applied Genetics 113:1481-1490

Bosland PW, Coon D, Cooke PH. 2015. Novel Formation of Ectopic (Nonplacental) Capsaicinoid Secreting Vesicles on Fruit Walls Explains the Morphological Mechanism for Super-hot Chile Peppers. Journal of the American Society for Horticultural Science 140:253-256.

Bus A, Hecht J, Huettel B, Reinhardt R, Stich B. 2012. High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. BioMed Central Genomics 13:281.

Chen C, Mitchell SE, Elshire RJ, Buckler ES, El-Kassaby YA. 2013. Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. Tree Genetics & Genomes 9:1537-1544.

Chen Z, Wang B, Dong X, Liu H, Ren L, Chen J, Hauck A, Song W, 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. BioMed Central Genomics 15:433.

Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. Nature Methods 5:887-893.

Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Research 36:e122-e122.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499-510.

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

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491-498.

Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19:11-15.

Edwards D, Forster JW, Cogan NO, Batley J, Chagné D. 2007. Single nucleotide polymorphism discovery. Association mapping in plants. Springer, pp 53-76.

Elmer I, Humira S, François B. 2015. Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. BMC Plant Biology 15:5.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

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

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature Biotechnology 32:279-284.

Gao Z-Y, Zhao S-C, He W-M, Guo L-B, Peng Y-L, Wang J-J, Guo X-S, Zhang X-M, Rao Y-C, Zhang C. 2013. Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. Proceedings of the National Academy of

Sciences USA 110:14492-14497.

Han K, Jeong H-J, Sung J, Keum YS, Cho M-C, Kim J-H, Kwon J-K, Kim B-D, Kang B-C. 2013. Biosynthesis of capsinoid is controlled by the *Pun1* locus in pepper. Molecular Breeding 31:537-548.

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T. 2009. High-throughput genotyping by whole-genome resequencing. Genome Research 19:1068-1076.

Huang Y-F, Poland JA, Wight CP, Jackson EW, Tinker NA. 2014. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PLoS ONE.

Jarquín D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, Lorenz A. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. BioMed Central Genomics 15:740.

Kalyanaraman A, Fontana P, Bhatnagar S, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M. 2010. The genome of the domesticated apple. Malus 9 domestica Borkh. Nature Genetics 42:833839Verardo.

Kang B, Nahm S, Huh J, Yoo H, Yu J, Lee M, Kim B-D. 2001. An interspecific (*C. annuum*× *C. chinese*) F_2 linkage map in pepper using RFLP and AFLP markers. Theoretical and Applied Genetics 102:531-539.

Khurana P, Gaikwad K. 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, Seo E, Choi J, Cheong K, Kim K-T. 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nature Genetics 46:270-278.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genetics 9:e1003215.

Moumouni K, Kountche B, Jean M, Hash C, Vigouroux Y, Haussmann B, Belzile F. 2015. Construction of a genetic map for pearl millet, Pennisetum glaucum. L.. R. Br., using a genotyping-by-sequencing (GBS) approach. Molecular Breeding 35:1-10.

Park, S.W., Jung, J.K., Choi, E.A., Kwon, J.K., Kang, J.H., Jahn, M., & Kang, B.C. 2014. An EST-based linkage map reveals chromosomal translocation in Capsicum. Molecular Breeding, 34(3), 963-975.

Pedraza-Garcia F, Specht J, Dweikat I. 2010. A new PCR-based linkage map in *pearl millet*. Crop Science 50:1754-1760.

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome 5:103-113.

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

Poland JA, Rife TW. 2012. Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome 5:92-102.

Pootakham W, Jomchai N, Ruang-areerate P, Shearman JR, Sonthirod C, Sangsrakru D, Tragoonrung S, Tangphatsornruang S. 2015. Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). Genomics 105:288-295.

Pootakham W, Shearman JR, Ruang-areerate P, Sonthirod C, Sangsrakru D, Jomchai N, Yoocha T, Triwitayakorn K, Tragoonrung S, Tangphatsornruang S. 2014. Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava. Manihot esculenta Crantz.. PloS ONE 9:e116028.

Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, Jannink J-L, Kulakow PA. 2014. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. Virus Research 186:87-96.

Rafalski A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94-100.

Saintenac C, Jiang D, Wang S, Akhunov E. 2013. Sequence-based mapping of the polyploid wheat genome. G3: Genes| Genomes| Genetics 3:1105-1114.

Sargent D, Clarke J, Simpson D, Tobutt K, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta K. 2006. An enhanced microsatellite map of diploid Fragaria. Theoretical and Applied Genetics 112:1349-1359.

Sargent DJ, Passey T, Šurbanovski N, Girona EL, Kuchta P, Davik J, Harrison R, Passey A, Whitehouse A, Simpson D. 2012. A microsatellite linkage map

for the cultivated strawberry. *Fragaria×ananassa*. suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection. Theoretical and Applied Genetics 124:1229-1240.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP. 2011. The genome of woodland strawberry. Fragaria vesca. Nature Genetics 43:109-116.

Silfverberg-Dilworth E, Matasci C, Van de Weg W, Van Kaauwen M, Walser M, Kodde L, Soglio V, Gianfranceschi L, Durel C, Costa F. 2006. Microsatellite markers spanning the apple (*Malus* x *domestica Borkh*) genome. Tree Genetics & Genomes 2:202-224.

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

Tanksley S, Ganal M, Prince J, De Vicente M, Bonierbale M, Broun P, Fulton T, Giovannoni J, Grandillo S, Martin G. 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141-1160.

Truong HT, Ramos AM, Yalcin F, de Ruiter M, van der Poel HJ, Huvenaars KH, Hogers RC, van Enckevort LJ, Janssen A, van Orsouw NJ. 2012.

Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLoS ONE 7:e37565.

Truong HTH, Kim K-T, Kim S, Chae Y, Park J-H, Oh D-G, Cho M-C. 2010. Comparative Mapping of Consensus SSR Markers in an Intraspecific F_8 Recombinant Inbred Line Population in *Capsicum*. Horticulture Environment and Biotechnology 51:193-206.

Van Os H, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, de Jong W, Van Koert P. 2006. Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173:1075-1087.

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, FitzGerald LM, Vezzulli S, Reid J. 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2:e1326.

Voorrips R. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93:77-78.

Wang S, Basten C, Zeng Z, Cartographer WQ. 2012. Department of Statistics, North Carolina State University, Raleigh, NC.

Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C. 2014. The locust genome provides insight into swarm formation and long-distance flight. Nature Communications 5.

Ward JA, Bhangoo J, Fernández-Fernández F, Moore P, Swanson J, Viola R, Velasco R, Bassil N, Weber CA, Sargent DJ. 2013. Saturated linkage map construction in Rubus idaeus using genotyping by sequencing and genome-independent imputation. BioMed Central Genomics 14:2.

Xu X, Zeng L, Tao Y, Vuong T, Wan J, Boerma R, Noe J, Li Z, Finnerty S, Pathan SM. 2013. Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. Proceedings of the National Academy of Sciences USA 110:13469-13474.

Yang H, Tao Y, Zheng Z, Li C, Sweetingham MW, Howieson JG. 2012. Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. BioMed Central Genomics 13:318.

Yarnes SC, Ashrafi H, Reyes-Chin-Wo S, Hill TA, Stoffel KM, 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

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q. 2011. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PLoS ONE 6:e17595.

Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, Hong W, Zhang X. 2013. Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biology 13:141.

Zhebentyayeva T, Swire-Clark G, Georgi L, Garay L, Jung S, Forrest S, Blenda A, Blackmon B, Mook J, Horn R. 2008. A framework physical map for peach, a model Rosaceae species. Tree Genetics & Genomes 4:745-756.

Zhu Y, Song Q, Hyten D, Van Tassell C, Matukumalli L, Grimm D, Hyatt S, Fickus E, Young N, Cregan P. 2003. Single-nucleotide polymorphisms in soybean. Genetics 163:1123-1134.

Zou G, Zhai G, Feng Q, Yan S, Wang A, Zhao Q, Shao J, Zhang Z, Zou J, Han B. 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. Journal of Experimental Botany.

초 록

본 연구는 GBS 방법을 이용하여 고추의 연관지도를 작성하고 이를 이용하여 고추 과피의 capsaicinoid함량을 조절하는 유전자좌를 탐색하기 위하여 수행되었다. *Capsicum chinense* 종인 Habanero와 Jolokia를 교배한 F_2집단 87개체를 유전자 지도 작성에 이용하였다. 부모 계통과 F_2집단 87개체에서 DNA를 추출하고 두 가지 제한효소(*PstI*-*MseI*)로 절단한 후 GBS library를 제작하였다. 제작된 library를 이용하여 Illumina High Seq 2000 1 lane에서 염기서열 데이터를 얻고 SNP를 calling한 후 슬라이딩 윈도우 방법을 이용하여 빈맵을 작성하였다. 이후 1,000개의 bin을 대표하는 SNP 마커들을 이용하여 연관지도를 작성하였다. 그 결과 총 3,654.1 cM의 연관지도를 작성하였으며 bin 마커들 간의 평균적인 거리는 3.9 cM이었다. 14개의 연관 그룹을 기존의 표준유전체에 align한 결과 고추의 12개 염색체에 고르게 정렬되었다. 이렇게 작성된 유전자 지도를 태좌와 과피의 capsaicinoid함량을 조절하는 QTL 탐색에 이용하였다. 먼저, Habanero, Jolokia 이의 F_1 식물 과피에서 capsaicinoid를 측정한 결과 Jolokia의 과피에서만 검출되었다. 아울러 *Pun1* 유전자는 Jolokia의 과피에서만 발현이 되고 Habanero와 F_1에서는 발현을 확인할 수 없었다. 이를

통해 과피의 capsaicinoid함량은 열성 유전을 하는 것으로 판단하였다. F_2의 과피와 태좌에서 capsaicinoid를 측정한 결과와 이를 QTL 분석에 이용한 결과 총 일곱 개의 QTL을 찾을 수 있었으며 과피의 capsaicinoid함량을 조절하는 QTL은 염색체 9번에 주로 분포하였다.

주요어: 염기서열 기반 유전형 분석 (GBS), 슬라이딩 윈도우 (sliding window), 빈맵 (bin-map), 연관지도 (Linkage map), 단일 염기 다형성 (SNP)

학번: 2014-20018