Scientia Horticulturae 243 (2019) 567–574

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Identification of loci associated with fruit traits using genome-wide single nucleotide polymorphisms in a core collection of tomato (*Solanum lycopersicum* L.)

Ngan Thi Phan[a,1], Lan Thi Trinh[a,1], Mi-Young Rho[c], Tae-Sung Park[c], Ok-Rye Kim[c], Jun Zhao[d], Hyoung-Moon Kim[d], Sung-Chur Sim[a,b,*]

[a] Dept. of Bioresources Engineering, Sejong University, Seoul, Republic of Korea
[b] Plant Engineering Research Institute, Sejong University, Seoul, Republic of Korea
[c] Vegetable Research Division, National Institute of Horticultural & Herbal Science, Rural Development Administration, Jenju, Republic of Korea
[d] Dept. of Applied Statistics, Konkuk University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
GWAS
Allele mining
Genotyping by sequencing
Population structure
Germplasm collection

ABSTRACT

Advances in sequencing technology have facilitated allele mining with high resolution for quantitative traits in crop species. In this study, a genome-wide association study (GWAS) was conducted to detect quantitative trait loci (QTL) for fruit traits in tomato. We developed a core collection of 192 tomato accessions representing contemporary breeding lines, vintage varieties, and wild species. Phenotypic variations in this collection were evaluated for six traits including fruit shape, fruit color, pericarp thickness, fruit weight, fruit height, and fruit width over two years in field trials. Genotyping by sequencing (GBS) detected a total of 140,072 SNPs in the core collection and 8,550 SNPs were selected based on missing data rate ($<$ 20%) and minor allele frequency ($\geq$5%) for GWAS. The 192 accessions were divided into seven clusters and the resulting membership coefficient matrix was used to account for population structure. We identified two loci for fruit color, seven loci for fruit shape, 11 loci for pericarp thickness, 13 loci for fruit weight, seven loci for fruit height, and 10 loci for fruit width at $P < 0.0001$ using the compressed mixed linear model. These loci explained 7.11–37.64% of total phenotypic variance. We also found that nine loci on seven chromosomes showed significant associations with multiple traits, suggesting pleiotropic effects of QTL. A total of 41 loci associated with the six fruit traits in our study will be a useful resource for marker-assisted selection to improve fruit traits in tomato.

1. Introduction

Tomato (*Solanum lycopersicum*) is an economically important vegetable cultivated worldwide and a model plant for basic research in the Solanaceae family. This species has diverse genetic variations in fruit traits such as shape, weight, and color. The fruit traits are important to determine the market values of tomato cultivars. Therefore, tomato breeding programs have made efforts to improve these traits according to consumer's demands. Genetic dissection of the fruit traits has also been extensively studied and a number of QTL have been identified in tomato (Lippman and Tanksley, 2001; Muños et al., 2011; Phan and Sim, 2017; Ranc et al., 2012; Rodríguez et al., 2011; Xu et al., 2013). However, QTL analysis using biparental populations (linkage analysis) has low mapping resolution due to a few recombination events and thus associations between QTL and markers are often ineffective in complex breeding populations (Holland, 2007; Pérez-de-Castro et al., 2012). Association mapping is an effective approach to detect QTL with high mapping resolution as it is based on germplasm collections and multi-parent segregating populations representing greater recombination events relative to biparental populations (Yu and Buckler, 2006). This method can also investigate more alleles than linkage analysis (Gupta et al., 2005; Zhu et al., 2008). Despite these advantages, association mapping was not commonly applied to vegetable crops due to lack of genome-wide markers and suitable mapping populations until the mid-2000s.

Next generation sequencing (NGS) technologies have accelerated to uncover whole genome sequences and to explore genome-wide single nucleotide polymorphisms (SNPs) in plant (Varshney et al., 2014). In

* Corresponding author at: Dept. of Bioresources Engineering, Sejong University, Seoul, Republic of Korea.
E-mail address: sungchur@sejong.ac.kr (S.-C. Sim).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scienta.2018.09.003
Received 4 May 2018; Received in revised form 19 August 2018; Accepted 3 September 2018
0304-4238/ © 2018 Elsevier B.V. All rights reserved.

tomato, an international consortium completed a reference genome assembly (The Tomato Genome Consortium, 2012) and resequencing of germplasm panels identified a large number of SNPs across 12 chromosomes (Liu et al., 2014; The 100 Tomato Genome Sequencing Consortium et al., 2014). Although resequencing is effective in detecting genome-wide SNPs, it is not always cost-effective. Genotyping by sequencing (GBS) is also a high-throughput approach based on genome complexity reduction method and is cost-effective to genotype a large number of samples such as mapping populations (Elshire et al., 2011; Kumar et al., 2013). Thus, GBS has emerged as a powerful tool for several applications in crop species including high-resolution genetic mapping (Poland et al., 2012; Spindel et al., 2013), genetic diversity analysis (Labate et al., 2014; Romay et al., 2013), QTL mapping (Bastien et al., 2014; Czerednik et al., 2015) and genomic selection (Crossa et al., 2013; Poland et al., 2012). In tomato, 14,043 SNPs were identified in a collection of 120 accessions representing three wild species *S. peruvianum*, *S. neorickii*, and *S. chmielewskii* using the GBS method for population genetics study (Labate et al., 2014). QTL mapping for fruit quality traits were also conducted with 3125 SNPs detected by GBS in the inbred backcross line population from an interspecies cross between *S. pimpinellifolium* x *S. lycopersicum* (Celik et al., 2017).

NGS-based SNP discovery and high-throughput genotyping platforms have facilitated association mapping, especially genome-wide association study (GWAS) in vegetable crops. In addition, several statistical methods have been developed for improving accuracy and efficiency of GWAS (Yu and Buckler, 2006; Zhang et al., 2010). With these powerful tools, GWAS have identified favorable alleles for disease resistance, plant architecture, fruit morphology, and fruit quality in tomato (Ranc et al., 2012; Ruggieri et al., 2014; Sacco et al., 2015; Sauvage et al., 2014; Shirasawa et al., 2013; Sim et al., 2015; Zhang et al., 2015). In the present study, we conducted GWAS to explore novel alleles for six fruit traits (fruit color, fruit shape, pericarp thickness, fruit weight, fruit height, and fruit width) in a core collection of 192 tomato accessions representing diverse genetic backgrounds. The GBS method was applied to genotype the core collection using genome-wide SNPs. Phenotypic evaluations for the six traits were performed over two years in field trials. The SNP markers significantly associated with these traits can be a useful resource for improving fruit traits through marker-assisted selection in tomato breeding programs.

2. Materials and methods

2.1. Plant materials

The 98 contemporary tomato breeding lines were collected from the National Institute of Horticultural and Herbal Science (NIHHS) in Rural Development Administration (RDA), Republic of Korea (ROK). In addition, the 94 germplasm accessions were assembled from the National Agrobiodiversity Center (NAC) in RDA, the Germplasm Resources Information Network (GRIN) in the U.S. Department of Agriculture, the C. M. Rick Tomato Genetics Resource Center (TGRC), and Sejong University. This tomato collection consisted of 191 cultivated accessions (*S. lycopersicum*) and a wild species (*S. pimpinellifolium*) from ≥ 18 countries including ROK, Russia, the United States, Uzbekistan, and China (Supplementary Table S1).

2.2. Genotyping by sequencing (GBS)

Genomic DNA was extracted from fresh, young leaves of three-week old seedlings using DNeasy Plant mini kit (QIAGEN, Valencia, CA, USA). GBS libraries were prepared using a protocol modified from (Elshire et al., 2011). First, the DNA samples were digested using the *Ape*KI enzyme (New England Biolabs, Ipswich, MA, USA), following the manufacturer's protocol. After digestion, the DNA fragments were ligated to the adapters with different barcodes that were assigned to each plant sample. The resulting DNA samples were then pooled and amplified by PCR to generate GBS libraries. The library quality was checked using the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The GBS libraries were sequenced using the HiSeq 2000 (Illumina Inc., San Diego, CA, USA) and the pair-end method. The raw sequencing reads were demultiplexed based on the individual barcode. The adapter and barcode sequences were then removed and the low quality sequences were trimmed using the SolxaQA v1.13 package (Cox et al., 2010). The Burrows-Wheeler Aligner (BWA) v 0.6.1-r104 program (Li and Durbin, 2009) was used to map the cleaned reads to the tomato reference genome v SL2.50. The mapped reads were extracted using SAMtools v 0.1.16 (Li et al., 2009) and a SNP matrix between the 192 tomato accessions was generated using the script of (Kim et al., 2014). The resulting SNPs were further filtered with > 5% of minor allele frequency and < 20% of missing data for GWAS.

2.3. Phenotypic evaluation and data analysis

The 192 tomato accessions were evaluated for six fruit traits including shape, color, pericarp thickness, weight, height, and width in field trials. Plants were first grown in greenhouse and six to seven-week old seedling were transplanted in the plastic covered field (high-tunnel) with 60 cm spacing between plants. The field trials were conducted using completely randomized design with four replicates per genotype in 2016 and 2017. For phenotypic evaluation, fully ripe fruits were harvested from the 2nd-4th flowering clusters. Fruit shape was determined using the 1–9 scales: 1 = flattened, 2 = slightly flattened, 3 = circular, 4 = rectangular, 5 = cylindrical, 6 = heart shaped, 7 = obovate, 8 = ovate, and 9 = pear-shaped. For fruit color (external), 1–6 scales were used: 1 = cream, 2 = yellow, 3 = orange, 4 = pink, 5 = red, and 6 = brownish. These scales were adapted from the International Union for the Protection of New Varieties of Plants (UPOV, 2001). For the other four traits (fruit weight, height, width, and pericarp thickness), actual values were measured using 10 fruits per genotype. The values of pericarp thickness were converted into 1–5 scales: 1 (< 3.59 mm), 2 (3.6–5.19 mm), 3(5.2–6.79 mm), 4(6.8–8.39 mm), and 5(> 8.4 mm).

Phenotypic variances of six traits were analyzed using three statistical methods. We used the Chi-squared test of independence (Pearson, 1900) and cumulative logit (McCullagh, 1980) methods for the nominal data (fruit shape and color) and the ordinal data (pericarp thickness), respectively. For the continuous data (fruit weight, height, and width), analysis of variance (ANOVA) was conducted. Correlation between fruit shape and color was measured using the Cramer's V coefficient (Cramér, 1946). This method was also used to calculate correlation coefficients between these two traits (nominal data) and pericarp thickness (ordinal data). Correlations among three traits with the continuous data were measured using the Pearson correlation coefficient (Pearson, 1895). The Spearman's rank correlation coefficient (Spearman, 1904) was used to analyze correlations between three continuous traits and pericarp thickness, while the eta method (Olejnik and Algina, 2003) was used between three continuous traits and two nominal traits. All of these analyses were conducted in R program (R Development Core Team, 2015).

2.4. Population structure and genome-wide association study (GWAS)

Population structure in the core collection was inferred using the STRUCTURE v2.3.4 program (Pritchard et al., 2000). To find the best K (number of clusters), we used the STRUCTURE model that allows for admixture and correlated allele frequencies. A series of K (1–10) was first tested in 10 independent simulations for each K with a burn-in of 10,000 iterations and a run length of 75,000 iterations. The selected K as candidates were further tested with a burn-in of 20,000 iterations and a run length of 100,000 iterations. The best K was then determined using the delta K method (Evanno et al., 2005) and a population

Download English Version:

https://daneshyari.com/en/article/9489625

Download Persian Version:

https://daneshyari.com/article/9489625

Daneshyari.com