Scientia Horticulturae 203 (2016) 38–46

Contents lists available at ScienceDirect

Scientia Horticulturae

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

Construction of a high-density genetic map for watermelon (*Citrullus lanatus* L.) based on large-scale SNP discovery by specific length amplified fragment sequencing (SLAF-seq)

Jianli Shang[1], Na Li[1], Nannan Li, Yongyang Xu, Shuangwu Ma*, Jiming Wang*

Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, The Laboratory of Melon Crops, Zhengzhou, Henan Province 450009, China

ARTICLE INFO

Article history:
Received 8 December 2015
Received in revised form 16 February 2016
Accepted 4 March 2016
Available online 19 March 2016

Keywords:
Watermelon
High-density
Genetic map
SLAF-seq

ABSTRACT

Watermelon (*Citrullus lanatus* L.) is an important vegetable crop worldwide, but a high-density genetic map of elite watermelon cultivars has not yet been reported. The present study reported the large-scale discovery of single nucleotide polymorphisms (SNPs) by specific length amplified fragment sequencing (SLAF-seq) for high-density genetic map construction using two elite watermelon cultivars. The map contained 2634 SNPs distributed on 11 linkage groups (LGs) corresponding to the number of chromosome pairs in watermelon. The total length of the linkage map was 1,906.31 cM, with an average distance of 0.72 cM between adjacent markers. Notably, 35 unmapped scaffolds, which contained 91 SNPs and covered 4.56 Mb, were dispersedly anchored to LG1, LG2, LG4, LG7, LG8, LG9 and LG11. Although disagreements were observed on LG1 and LG11, the local collinearity between the genetic and physical distances, haplotype map and heat map indicated that the genetic map herein is of high quality. The present study reports the first high-density linkage map of two watermelon elite parents.

1. Introduction

Watermelon (*Citrullus lanatus* L.) is an important vegetable crop in the warmer regions of the world (Maynard, 2001; Paris et al., 2013). China is the largest producer of watermelons, with an agricultural area of 1.82 million Ha and the production of 70 million tons of watermelon in 2013 (FAO, http://faostat.fao.org).

Genetic maps are a key resource for understanding genome organization and evolutionary relationships (Wang et al., 2011) and assist in assigning and orientating sequence assemblies to correct chromosome locations (Ren et al., 2012). Moreover, high-density linkage maps provide the basis for quantitative trait loci (QTL), the map-based cloning of major genes underlying important traits, and marker-assisted selection. Several unsaturated genetic maps of watermelon have been constructed based on isozymes (Navot and Zamir, 1987; Zamir et al., 1984), randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), and sequence related amplified polymorphism (SRAP) markers, but only a limited number of simple sequence repeat (SSR) markers had been used (Levi et al., 2006, 2009) prior to 2012. The development of next-generation sequencing (NGS) technologies and the completion of the whole-genome sequencing of watermelon (Guo et al., 2013) have facilitated the generation of several linkage maps for watermelon. The first high-density genetic map harboring 698 SSRs, 219 insertion/deletions (InDels) and 36 structure variations (SVs) was reported using a population derived from a cross of an elite cultivar (97103) and the United States Plant Introduction (PI) 296341-FR (*C. lanatus* var. *citroides*) (Ren et al., 2012). Three single nucleotide polymorphism (SNP) maps for watermelon containing 338, 378 and 358 SNP markers were constructed and compared using three populations developed from crosses between two elite cultivars, an elite cultivar and a wild egusi accession, and an elite cultivar and a wild citron accession (Sandlin et al., 2012). An integrated genetic map containing 698 SSR, 219 InDel, 36 SV and 386 SNP markers (Ren et al., 2014) was constructed by merging data from four mapping experiments (Ren et al., 2012; Sandlin et al., 2012) using different parental lines, including three subspecies of watermelon. A genetic map containing 232 SNPs and 50 SSRs was constructed using two plant introduction accessions as parents (*C. lanatus* var. *citroides* × *C. lanatus* var. *lanatus*) (Nimmakayala et al., 2014). To identify QTL and SNPs associated with resistance, a genetic map harboring 266 SNPs was constructed using a population derived from two closely related watermelon cultivars (Lambel

* Corresponding authors.
E-mail addresses: lina01@caas.cn, lina2341@163.com (N. Li), mashuangwu@caas.cn (S. Ma), pomology@163.com (J. Wang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.scienta.2016.03.007
0304-4238/

et al., 2014). Based on genotyping by sequencing (GBS), a high-resolution genetic map (547 skeletal SNP markers and 9933 add-on SNPs) was constructed using a mapping population that contained 113 F_2 progenies obtained from a cross of egusi and sweet watermelon (Reddy et al., 2014). More recently, a high-density DArTseq SNP-based genetic map of watermelon was constructed using an F_2 population derived from a cross between an elite cultivar and a wild germplasm (Ren et al., 2015). These genetic maps (which had been generated since 2012) exhibited several common patterns. First, several markers, including SSRs, InDels, SVs, and SNPs, were generated by NGS based on reference genome of watermelon and therefore facilitated a comparative analysis of collinearity, segregation distortion and QTL locations across several populations possible. Second, these maps represent all three subspecies of watermelon, with a particularly focus on inter-specific or inter-subspecific crosses. However, few maps of two elite watermelon cultivars have been published, and they include a limited number of markers (Lambel et al., 2014; Sandlin et al., 2012); a high-density genetic map of a population derived from two elite parents has not yet been published.

Recently, a new high-throughput strategy for *de novo* SNP discovery known as specific length amplified fragment sequencing (SLAF-seq) was developed utilizing the large-scale SNP development combined with NGS platforms (Sun et al., 2013). SLAF-seq has been successfully applied to construct high-density maps of sesame (Zhang et al., 2013), soybean (Qi et al., 2014), tea (Ma et al., 2015), cucumber (Wei et al., 2014; Xu et al., 2015), mei (a type of Japanese apricot) (Zhang et al., 2015a), rice (Mao et al., 2015), and grape (Guo et al., 2015). However, SLAF-seq has not been successfully applied to watermelon.

To increase marker density and exploit marker resources for elite watermelon cultivars, we constructed a high-density linkage map derived from a cross of two elite watermelon cultivars using SLAF-seq.

2. Materials and methods

2.1. Plant material and DNA extraction

An F_2 population of 93 plants derived from a cross between "ZXG01478" (the female parent) and "14CB11" (the male parent) was used to generate a linkage map. "ZXG01478" and "14CB11" are elite inbred lines that exhibit significant differences in several traits (Fig. 1), such as external stripe patterns on the fruits, fruit shapes, flesh colors, seed sizes, seed colors and resistance to *Fusarium oxysporum* f. sp. *niveum* Race 1. Both parents and their F_2 progeny were planted at the Fruits Research Institute of the Chinese Academy of Agricultural Sciences in Zhengzhou, China. Genomic DNA was extracted from fresh leaves of the parents and their F_2 progeny using a modified CTAB method (Doyle, 1987). DNA concentration was measured with a NanoDrop-1000 spectrophotometer (NanoDrop, Wilmington, DE, USA), and DNA quality was determined by electrophoresis in 0.8% agarose gels using λ DNA as a standard.

2.2. SLAF library construction and high–throughput sequencing

SLAF-seq was performed as described in Sun et al. (2013) with minor modifications. Genomic DNA was first incubated at 37 °C with 0.6 U Mse I (NEB, Hitchin, UK), T4 DNA ligase (NEB), ATP (NEB), and *Mse*I adapter for 15 h; restriction-ligation reactions were heat-inactivated at 65 °C for 1 h and then digested with the restriction enzymes Hae III and SspI for 3 h. PCR was performed using diluted restriction-ligation samples, dNTP, Taq DNA polymerase (NEB) and an *Mse*I-primer containing barcode 1. The PCR products were purified using the E.Z.N.A.H Cycle Pure Kit (Omega, USA) and pooled. The pooled sample was incubated at 37 °C with MseI, T4 DNA ligase, ATP and a Solexa adapter. The sample was purified using a Quick Spin column (Qiagen, Germany) and then resolved on a 2% agarose gel. Fragments that were 450–500 bp (with indexes and adaptors) in size were isolated using a Gel Extraction Kit (Qiagen). These product fragments were then amplified by PCR with Phusion Master Mix (NEB) and Solexa Amplification primer mix to add barcode 2. The Phusion PCR settings are specified in the Illumina sample preparation guide. The samples were gel purified, and the 450–500 bp DNA was excised and diluted for sequencing. The selected SLAFs were then subjected to paired-end sequencing using an IlluminaHiSeqTM2500 (Illumina, Inc., San Diego, CA, U.S.) at Biomarker Technologies Corporation in Beijing (http://www.biomarker.com.cn/english/). To prevent false positive reads, the sequence error rate was estimated using the data from rice as a control. The ratio of high quality reads with quality scores greater than Q30 (indicating a 0.1% chance of an error, i.e., 99.9% confidence) in the raw reads and guanine-cytosine (GC) content were calculated for quality control.

2.3. SNP calling and genotype definition

Low-quality reads (quality scores <30) were filtered out and a pipeline combining Burrows–Wheeler Aligner (BWA) (Van Ooijen, 2004), Genome Analysis Toolkit (GATK) (DePristo et al., 2011) and Sequence Alignment/Map tools (SAMtools) (Li et al., 2009) was used to call the SNPs. The F_2 individual sequences were aligned against the 97103 reference genome (Guo et al., 2013) using BWA with parameters defined as Score (missed match) = 3, Score (opening gap) = 11 and Score (Gap extension) = 4. False alignments were always detected near InDels; therefore, local realignment was performed. The "Unified Genotyper" function of GATK was used for variant calling. SAMtools and GATK were used to identify SNPs, and their intersection was merged as the candidate SNP dataset. Only biallelic SNPs were retained as the final SNP dataset. Polymorphic markers were classified into four segregation patterns (hk × hk, lm × ll, nn × np and aa × bb). An F_2 population was obtained by selfing the F_1 of a cross between two fully homozygous parents with genotype aa or bb. Thus, only the SNPs whose segregation patterns were aa × bb were used to construct the genetic map. The average sequence depths of the SNPs were more than 20-fold in the parents and greater than 4-fold in the progeny. Any a progeny contained more than 70% of the SNPs in the parents, i.e., 70% integrity of SNPs in individuals. A $\chi 2$ test was conducted for each SNP with a null hypothesis that the two alleles at a locus segregated with a ratio of 1:1 in the F_2 population. All SNPs that significantly deviated from this ratio ($P < 0.001$) were excluded from the SNP dataset.

2.4. Genetic map construction and evaluation

Linkage mapping, marker ordering, error genotyping correction and map evaluation were performed using the Highmap software as described previously (Liu et al., 2014) with minor modifications. Firstly, markers were clustered into linkage groups (LGs) using a single-linkage clustering algorithm based on a pair-wise modified independence LOD score as the distance metric. Then, enhanced Gibbs sampling, spatial sampling and simulated annealing algorithms were combined to conduct an iterative process of marker ordering (Jansen et al., 2001; Van Ooijen, 2011). In brief, in the first stage of the ordering procedure, markers were selected using spatial sampling. One marker was randomly selected in a priority order of the test cross, and markers with a recombination frequency smaller than a given sampling radius were excluded from the marker set. Subsequently, simulated annealing was applied to search for the best map order. Summation of adjacent recombination fractions was calculated. The annealing system continued until, in several successive steps, the newly generated map order was