Journal of Agricultural Science; Vol. 11, No. 2; 2019
ISSN 1916-9752 E-ISSN 1916-9760
Published by Canadian Center of Science and Education

Identification of Early Tomato Fruit Ripening Loci by QTL-seq

E. Ruangrak[1,2,3], Yongchen Du[2], Nang Myint Phyu Sin Htwe[3], Pimpan Pimorat[4] & Jianchang Gao[2]

[1] Department of Technology and Industry, Prince of Songkla University, Pattani Campus, Pattani, Thailand

[2] Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China

[3] Urban Agriculture Technology Research Group, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani, Thailand

[4] Faculty of Agricultural Technology, Valaya Alongkorn Rajabhat University under the Royal Patronage, Sakaeo Campus, Sakaeo, Thailand

Correspondence: Jianchang Gao, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China. Tel: 81-0736-77-0345. E-mail: gaojianchang@caas.cn

Received: September 7, 2018 Accepted: November 1, 2018 Online Published: January 15, 2019

doi:10.5539/jas.v11n2p51 URL: https://doi.org/10.5539/jas.v11n2p51

The research is financed by the National Natural Science Foundation of China (31471874) and the National Key Research and Development Program of China (2016YFD0101703).

Abstract

QTL-seq has been successfully studied in identifying major QTLs, markers, and candidate genes associated with traits that are important for crop improvement. Tomato earliness is an economically important trait and is a major current research focus recently. This paper reports the identification of tomato early ripening fruit locus facilitated by QTL-seq using a novel next-generation sequencing technology. Two DNA pools of phenotypes of F2 offspring from crosses between the Bone MM (early ripening fruit, P1) and 071-440 (late ripening fruit, P2) cultivars of (*Solanum lycopersicum*) were bulked for sequencing and alignment analysis. Sequencing results revealed 434 SNP markers on chromosome 11, a candidate QTL at position 52,048,208 bp (named er-fruit) and a candidate gene, Solyc11g071510.1.1. The "er-fruit" as confirmed by the traditional QTL method was related to the early fruit ripening trait in tomato. Additionally, BLAST analysis to known homologies for Solyc11g071510.1.1 gene encodes glycoside hydrolases (GHs). GHs are functionally associated with cell wall degradation, fruit softening and ripening. Thus, GHs may be important in fruit softening, stimulating early fruit ripening in tomato. Our results confirmed that QTL-seq is effective method to identify candidate QTL loci, candidate genes and candidate markers.

Keywords: QTL-seq, tomato, early ripening fruit, glycoside hydrolases

1. Introduction

Earliness in tomato is one of the factors that needs much concern in recent years due to climatic changes and increased world's population. The ability to bring their products earlier to the market in the season can produce better income for growers (Kevany et al., 2008). Earliness in tomatoes consists of three stages; (1) flowering time, (2) fruit setting time, and (3) fruit ripening time (Powers, 1941). The environmental factors such as temperature and light intensity play a significant role in the expression of any components for early maturity (Kerr, 1955; Adams et al., 2001). It has been reported "Early Cherry' alleles caused reductions in both ripening time and fruit weight by using RAPD marker analysis in F_2 population derived from a cross between Lycopersicon esculentum'E6203' (normal ripening) and Lycopersicon esculentum'Early Cherry' (early ripening) (Doganlar et al., 2000).

Early fruit ripening is commercially important and effective trait for tomato (Gur et al., 2010). A QTL (dw1) of the tomato that linked to phenotypic traits, increased yield (quantitative) and earliness (qualitative) have been identified although it caused a decline in fruit firmness (Inai et al., 2006). The tomato is classified as a climacteric fruit that needs phytohormone ethylene to ripen and it also coordinates expression of thousands of genes regulating fruit softening and increasing color development, sugars, acids, and aroma production (Klee & Giovannoni, 2011). The important fruit ripening phenotypes have been distinguished by rin, nor, Nr and Cnr mutants that have been

provided novel insights into the control of ripening processes (Thompson et al., 1999). In addition, the cell wall modification for softening of the fruit tissues is affected by transcriptional factors nor, rin, and ethylene receptor Never-ripe (Nr) because the transcription level of cell wall degrading enzymes polygalacturonase and pectate-lyase were not observed in rin, nor, and Nr mutants during tomato fruit ripening (Osorio et al., 2011). Smith and Gross (2000) proposed that a member of glycoside hydrolase family 35, β-galactosidase II, may be involved in Gal metabolism during cell wall degradation for softening of tomato fruit, conversion of chloroplasts into chromoplasts, fruit growth, and senescence.

Next-generation sequencing (NGS) technology was proved as a quick accurate and successful method of genome analysis (Takagi et al., 2013a) which involves categorizing molecular markers associated to target genes or genotyping a pair of bulked DNA samples from two dissimilar extreme phenotypes and connecting the markers with QTLs related with chosen traits of research interest (Michelmore et al., 1991; Giovannoni et al., 1991; Mansur et al., 1993; Darvasi & Soller, 1994). The new approach has been proposed as a means of developing rapid QTL map through the MutMap (Abe et al., 2012), MutMap-Gap (Takagi et al., 2013b), Mutmap+ (Fekih et al., 2013), and QTL-seq (Takagi et al., 2013a) approaches. QTL-seq has developed (Fekih et al., 2013; Takagi et al., 2013a) to replace traditional QTL mapping which is labour-intensive, time-consuming and involves substantial costs associated with the development of DNA markers, genotyping and the generation of a large number of progenies during advanced segregating generations (Takagi et al., 2013a). Moreover, rapid identification of the QTL region (marker and candidate gene) associated with the traits of interest can be performed in the F2 population. The QTL-seq has been employed previously to identify QTLs underlying disease resistance traits in rice (Takagi et al. 2013a), the early flowering trait in cucumber (Lu et al., 2014), seed weight trait in the chickpea (Das et al., 2015), fruit weight, locule number (Illa-Berenguer et al., 2015) and early flowering traits in tomato (Ruangrak et al., 2018). In the present study, we used QTL-seq to identify the QTL for early ripening trait in tomato progenies of cross between naturally selected Bone MM (earliness) and 071-440 (lateness) cultivars.

2. Method

2.1 Plant Materials and Phenotypic Evaluation

S. lycopersicum cv. *Bone MM* (Earliness (E); P1 from Russia) and 071-440 (Lateness (L); P2 from China) were used as parents (Figure A1). The genetic backgrounds of Bone MM and 071-440 are extremely different for first fruit ripening characteristics. For the phenotypic evaluation, the first fruit ripening time was visually scored by counting the days from the first flower opening (anthesis) to the first fruit ripening of each plant, developing 90% red color on fruit surface. The data were used for frequency distribution analysis. F2 progeny showing two extremes (early and late) of first fruit ripening times were isolated and pooled into two bulks (each bulk comprising 30 individuals). The experiment was performed in the tunnel type green house (at day/night average temperatures of 28 °C/15 °C) at the Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China (39.96° N, 116.33° E).

2.2 QTL-seq Analysis

Two DNA bulks of extreme early (41-45 days after anthesis) and late (55-59 days after anthesis) ripening times categories were extracted as equal volumes of DNA samples from the F2 progeny by following previously described DNA isolation methods (Abe et al., 2012; Takagi et al., 2013a). The genomic DNA extraction was performed from fresh tomato leaves using the Cetyl Trimethyl Ammonium Bromide (CTAB) method. The whole-genome sequencing was performed using an Illumina Genome IIx sequencer. Pair-end sequencing libraries (read length 100 bp) with 500 bp insert sizes were prepared for sequencing. The short reads were aligned to the *S. lycopersicum* reference genome sequence (//ftp.ensemblgenomes.org/pub/plants/release-22/fasta/solanum_lycopersicum/dna/Solanum_lycopersicum.SL2.40.22.dna.toplevel.fa) with BWA software (Li & Durbin, 2009). SNP-calling was performed using SAM tools software (Li et al., 2009) and then converted into SNP-index as reported (Abe et al., 2012; Fekih et al., 2013; Takagi et al., 2013a). A given result was based on short reads harbouring the SNP being different from the reference sequence (Fekih et al., 2013; Lu et al., 2014). A SNP-index of E- and L-Ripening bulks was subtracted to obtain a Δ(SNP-index). Fisher's exact test (Fisher, 1922) was used to evaluate the statistical significance of the Δ(SNP-index) values. The detection of functionally annotated putative SNPs and the annotation of the candidate polymorphic marker locus were performed using ANNOVAR software (Wang et al., 2010).

2.3 Traditional QTL Analysis

To verify the results of the QTL-seq, conventional QTL analysis using InDel (insertion or deletion) markers was used. Two hundred and three InDel markers were identified from chromosome 11 (Table B1) by aligning E-Ripening Bulk Illumina reads to the reference genome (//ftp.ensemblgenomes.org/pub/plants/release-22/fasta/

solanum_lycopersicum/dna/Solanum_lycopersicum.SL2.40.22.dna.toplevel.fa) with BWA/SAMtools software (Li & Durbin, 2009). The primers of the InDel markers were designed using Primer 3 software (http://simgene.com/Primer3). The DNA of 190 F2 plants was diluted to approximately 50-100 ng/µl and prepared for InDel marker performance analysis as described in Ruangrak et al. (2018).

2.4 Protein BLAST Function

BLAST searches for homologies to known protein functions at the Sol Genomics Network (SGN; http://solgenomics.net/) website were used to infer the function of the candidate gene (Fernandez-Pozo et al., 2015).

3. Results

3.1 Phenotyping and Distribution of Early Fruit Ripening Time in F2 Population

The early fruit ripening time was scored from the days after anthesis (first flowering) to the ripening of fruit (first fruit) of 1200 individuals of F2 obtained from crossing Bone MM (P1) to 071-440 (P2) cultivars (Figure A1). The frequency distribution of the fruit ripening time is shown in Figure 1. The average number of days after anthesis to ripening of fruit on Bone MM (P1) was approximately 9 days earlier than that in 071-440 (P2), whereas F1 plants ripened 5 days later than P1 and 4 days earlier than P2.

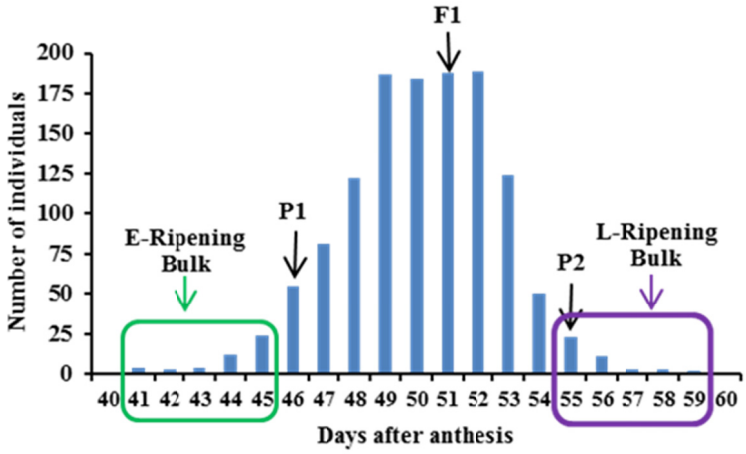

Figure 1. Frequency distribution of ripening time as determined in P1, P2, F1 and F2 populations. The DNAs of early and late ripening plants were bulked to make an early ripening bulk (E-Ripening bulk) and a late ripening bulk (L-Ripening bulk), respectively

3.2 Sequencing of the Early Fruit Trait

QTL-seq is based on the cross between two parents that have extreme phenotypic differences followed by selfing of F1 individuals to breed F2 offspring. The DNA samples from 30 F2 individuals showing the phenotypic extremes were bulked in an equal ratio (Figure A1) and subjected to whole genome sequencing with 100 bp paired-end reads for a depth of more than 10X coverage. To summarize the genotyping and sequencing data assessment, all sample sequence data were considered to be of sufficiently high quality and showed a normal GC distribution. The short reads were aligned to the tomato reference genome sequence using BWA software and the alignment was performed, using SAMTOOLS software for SNP-calling. The results of the alignment suggested that the reference genome size is 781666411 bp, and the mapping rate of the aligned samples varied from 93.14% to 96.3%, for the reference genome (Table B2). The average coverage depth ranged between 16.31X to 21.09X and the 1X coverage (at least one base coverage) was above 99.56% (Table B2). Thus, a normal alignment resulted could be considered valuable for the detection of nucleotide change and subsequent analyses.

Furthermore, the ANNOVAR software tool (Wang et al., 2010) was used to locate putative SNPs with respect to predicted genes and to classify them based on function. In the early ripening population, a total of 1,870,498 bp SNPs were detected. The largest number of SNPs (1,622,942 bp, or 86.77%) were classified in the intergenic region group, while the second largest number of SNPs fell into the intronic group (92,638 bp, or 4.95%) and the third largest group were located upstream of a gene (63,263 bp, or 3.38%) followed by downstream of a gene (52,837 bp, or 2.82%). Furthermore, in terms of SNPs (< 2%) in coding regions, both non-synonymous (18,955

bp, or 1.01%) and synonymous (13,230 bp, or 0.71%) SNPs were identified. The missense type non-synonymous SNPs again subdivided into SNPs belonging to a stop gain (468 bp, or 0.03%), a stop loss (168 bp, or 0.01%), and splicing group (180 bp, 0.01%) (Table B3). After screening of each type of SNPs, 1,870,498 polymorphic marker loci were filtered and plotted separately into a part of the SNP-index site and SNP-index graphs from the E- and L-Ripening bulks (Figures S3 and S4, respectively). To identify QTL positions corresponding to the early ripening trait, Δ(SNP-index) values were compared to identified candidate SNPs (Figure A5). The Δ(SNP-index) was calculated based on the difference from two SNP-index graphs of E- and L-Ripening bulks, for which putative SNP positions in two different extreme traits should exhibit unequal contribution from two parental genomes (Fekih et al., 2013; Takagi et al., 2013a). Based on the analysis, 434 candidate SNPs were identified and were located on chromosome 11 (Table B4). Annotations of the 434 polymorphic marker loci were generated using ANNOVAR software (Wang et al., 2010).

3.3 Genomic Region Corresponding to the Early Fruit Ripening Trait

In this experiment, a genomic region located from 51.0 to 53.0 Mb on chromosome 11 was detected which was associated with the early ripening time phenotype (Figures 2(a)-2(c)). An average Δ(SNP-index) was computed in a 1 Mb interval using a 1 Kb sliding window (Figure A2). The Δ(SNP-index) graphs of the candidate genes were plotted with confidence intervals under the null hypothesis of no QTL ($P < 0.05$). The positions of early fruit ripening time candidate QTL gene were located on chromosome 11. A QTL position was located on the genomic region of 52,048,208 bp in the Solyc11g071350.1.1 gene (Figures 3(a) and 3(c)).

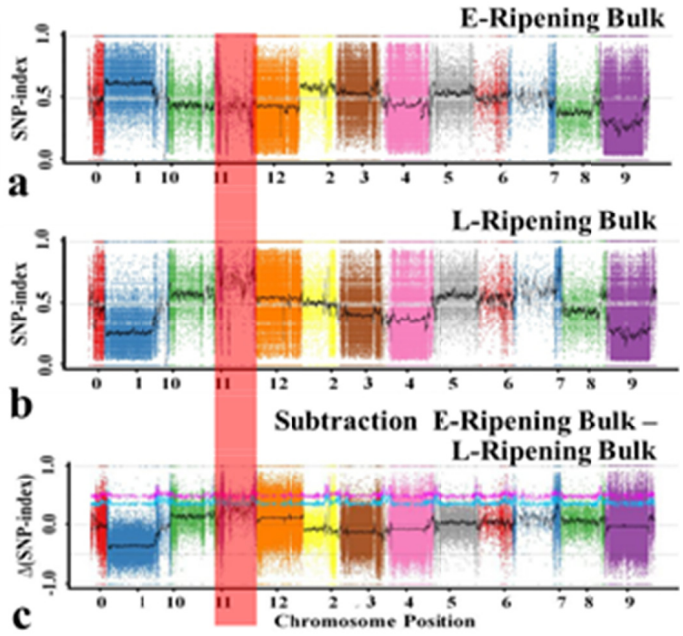

Figure 2. SNP-index graphs of (a) E-Ripening bulk, (b) L-Ripening bulk and (c) Subtraction of E-Ripening bulk to L-Ripening bulk or Δ(SNP-index) graph from QTL-seq analysis. The X-axis shows the position of the twelve chromosomes and the Y-axis shows the SNP-index (a and b) and Δ(SNP-index) as given in (c). The Δ(SNP-index) graph (c) was plotted with statistical confidence intervals under the null hypothesis of no QTL ($P < 0.05$). The calculation of Δ(SNP-index), the subtraction of the two SNP-indexes: Δ(SNP-index) = SNP-index (E-Ripening bulk) – SNP-index (L-Ripening bulk). A 1,000 times permutation test selected 95% (blue line), 99% (purple line) confidence level as the screening threshold using Fisher's exact test. The black lines are average values of the SNP-index or Δ(SNP-index) drawn from the sliding window analysis

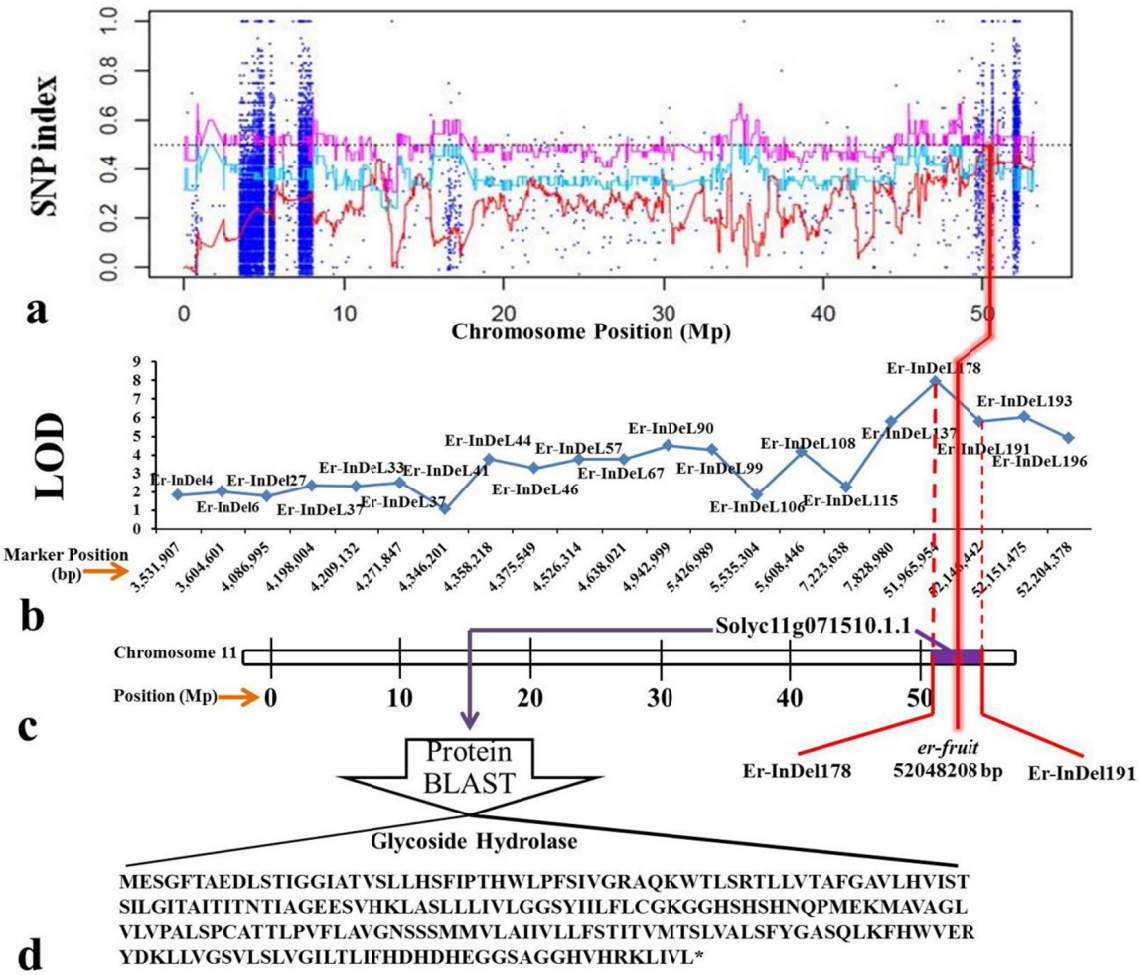

Figure 3. Illustration of the QTL-seq result and confirmation. (a) The SNP-index Δ(SNP-index) plot of chromosome 11 shows the subtraction of E-Ripening bulk and L-Ripening bulk at ripening time and the red lines are average values of the SNP-index. Additionally, 1,000 times permutation test selected 95% (blue line) and 99% (purple line) confidence levels as the screening threshold. (b) The LOD graph of Linkage analysis with 21 Indel markers shows the identification of traditional QTLs on chromosome 11. The Y-axis shows LOD values and the X-axis shows marker positions on chromosome 11. (c) The location of the candidate QTL position and candidate genes on chromosome 11. (d) BLAST protein analysis shows the candidate gene. The red lines are average values of the SNP-index gene

In the detection of QTLs corresponding to early fruit ripening, a total of 1,870,498 SNPs were identified between the Bone MM and 071-440 genomes, and the SNP-index was calculated for each SNP. Graphs showing the relationships between the SNP-index and genomic positions are given in Figures 2(a)-2(c). Highly contrasting patterns of SNP-index graphs were found for E-bulk and L-bulk in the region, suggesting that the early fruit ripening was mainly associated with Bone MM-type genomic segments in the 51.0 to 53.0 Mb region of chromosome 11 (Figures 2(a)-2(c)). On the other hand, late fruit ripening was associated with the 071-440-type genome in the same region, indicating that values of the Δ(SNP-index) which possibly correspond to QTLs governing the difference between the E- and L-progeny. Statistical confidence intervals of the Δ(SNP-index) were calculated for all the SNP positions with given read depths under the null hypothesis of no QTLs, and were plotted along with the Δ(SNP-index) (Figure 2(c)).

3.4 Traditional QTL Analysis

The highest LOD (Log of Odds) score (LOD = 7.94) was observed in the interval of InDel markers located at 46.5 and 53.0 Mb. This interval was corresponded to the genomic region identified by the QTL-seq method (Figures 2(a)-2(c)). To verify the early ripening QTL which was identified by the QTL-seq method, 190 F2

plants were used in a traditional QTL analysis. Among 203 InDel markers from chromosome 11, 21 InDel markers were polymorphic between the E- and L-Ripening bulks as shown in Table 1.

Table 1. The information of InDel markers used in the traditional QTL analysis

Marker name	Primer sequences (5′→3′)	Position (bp)	LOD Value[a]	PVE[b]	Kruskal-Wallis test (*P*)[c]
Er-InDel4	F: CAGAATTGAGCAACATTCAA R: AGCTTGGATTCCCTTCTATC	3 531 907	1.83	11.73	**
Er-InDel6	F: CTAGGGATAGGCATTTTCTG R: GTACTTAATCAAAGCTCAGCC	3 604 601	2.03	11.67	**
Er-InDel27	F: TGTTTCTATTCGTGAACCAT R: TTGTCAAATTCATGATTAAAAG	4 086 995	1.78	11.75	**
Er-InDel31	F: GGGAAAACCTTAGAATCTTGA R: TTAGGTAGCGTTTTATGGGA	4 198 004	2.34	11.58	**
Er-InDel33	F: TTTGCATAGTTTTTGCTCCT R: ACCACACCAAATTGACTTTC	4 209 132	2.32	11.59	***
Er-InDel37	F: ATCCCACGATTAAATCAGC R: TCAATGCTCCCTCACTTATT	4 271 847	2.51	11.53	****
Er-InDel41	F: AGGAATTATGGGGGATTACA R: CAAACATCGAATGAACAACA	4 346 201	1.06	10.97	**
Er-InDel44	F: TTTAGAAGGATGGCCAGATA R: TCGAACGTGACCAATAAAAT	4 358 218	3.76	11.17	******
Er-InDel46	F: CTTCTGGGGTACTCTCTCCT R: CGAAATTGATATACTATCGGTG	4 375 549	3.28	11.31	*****
Er-InDel57	F: CAAATATACCCGAATCTCCA R: CTATGAGCGAAACTCCAAGT	4 526 314	3.77	11.17	*****
Er-InDel67	F: AGTCACGAGCTTGAAATTCT R: AACGAGCCATTATTGTCCTA	4 638 021	3.77	11.17	****
Er-InDel90	F: CATTTTCGGTAAGTTTTTGG R: TGTCGAAAAAGAATTAAACGA	4 942 999	4.52	10.96	*******
Er-InDel99	F: ACCCTCCAAAAATACATGC R: GGATGAAATGGAAAAGACAG	5 426 989	4.28	11.03	*******
Er-InDel106	F: GCATTCATCTAAAGGCAAAC R: GAACAGATCTCACTTCGGTC	5 535 304	1.85	11.72	**
Er-InDel108	F: GTAGTGCAACCAAAGACCAC R: TAGCCTAATTGGTCGAGTGT	5 608 446	4.17	11.06	*******
Er-InDel115	F: GTTAGGTTTCAGTTGCCGT R: CGAACTTAGTCCATCACCAT	7 223 638	2.28	11.60	***
Er-InDel137	F: TCAAGTTTCCTTTTGCTTTC R: AGTCCCTATCCACAGATCCT	7 828 980	5.77	10.62	*******
Er-InDel178	F: TGTCGTCACTGACTATTTGG R: CTCCTTGAGGAAAGGACTCT	51 965 954	7.94	10.06	*******
Er-InDel191	F: TCATCTTTCGAGTCGAGATT R: TATCCATTTTGTATAGGGGC	52 146 442	5.80	10.61	*******
Er-InDel193	F: TGAAGGAAACAATGTCACAA R: TACCTGAAAAGAAATCGGAA	52 151 475	6.05	10.55	*******
Er-InDel196	F: TGTTTGTCACAAGTATCTGTTG R: AAATTAGTCGCGTTCCATAC	52 204 378	4.90	10.86	*******

Note. [a]LOD value was performed by MQM mapping analysis. [b]PVE was the percentage of phenotypic variance calculated by MQM mapping in MapQTL 6.0. Significance: *P = 0.05; **P = 0.01; ***P = 0.005; ****P = 0.001; *****P = 0.0005; ******P = 0.0001; *******P = 0.00005 (P values from Kruskal-Wallis test are indicated by an asterisk).

These 21 markers were applied to the segregating population for QTL analysis. MQM mapping analysis identified a major QTL for early fruit ripening time delimited by two InDel markers Er-InDel178 and

Er-InDel191, which were physically located in the region range of 51,965,954 to 52,146,442 Mb on chromosome 11 (Figure 3(b)). A LOD threshold value of 3.8 was used for declaration of a QTL. The LOD values in this region ranged from 1.06 to 7.94 with the highest peak at marker locus Fl-InDel178 (7.94). This interval was corresponded to the genomic region identified by the QTL-seq method (Figures 2(a)-2(c)). In addition, the candidate gene was aligned to the Solyc11g071350.1.1, encodes glycoside hydrolases (GHs), which was analyzed by BLAST through the Sol Genomics Network (SGN; http://solgenomics.net/) website (Fernandez-Pozo et al., 2015).

4. Discussion

QTL-seq is a powerful tool for identifying candidate QTL loci and candidate genes using NGS technology as previously reported (Takagi et al., 2013a; Lu et al., 2014; Das et al., 2015; Illa-Berenguer et al., 2015). This study is aimed to rapidly identify the candidate QTL locus and gene related to the early fruit ripening of the tomato using QTL-seq. Our aim was achieved successfully using naturally selected varieties from Russia (earliness) and China (lateness). The results of phenotyping and the distribution of early fruit ripening time demonstrated that multiple genes control fruit ripening time because the frequency distribution is close to a normal (Gaussian) distribution (Takagi et al., 2013a) (Figure 1). Thus, results suggest that the F2 population can further benefit from the use of QTL-seq analysis. The QTL mapping results confirmed the QTL-seq analysis, supporting the proposition of the QTL located on 52,048,208 bp was a major QTL associated with the early ripening fruit phenotype.

The normal distribution of the F2 population clearly allowed the performance of QTL-seq, which is based on the crossing of two parents that have extreme phenotypic differences followed by selfing of F1 individuals to generate F2 progeny. Takagi et al. (2013a) suggested that an F2 population is much easier to generate than RILs of complex generations. DNA samples of F2 individuals showing extreme phenotypes, *i.e.* those exhibiting the earliest and latest extreme values of fruit ripening phenotype were bulked in an equal ratio and subjected to whole genome sequencing. In this study, the high base accuracy of Q30 varied from 90.08% (for L-Ripening bulk) to 92.29% (P1), with an average of 91.30% (Table B5) suggested that the sequencing data of all the samples corresponded to low error probabilities and sufficiently high quality. Alignment analysis of the sequencing data showed a candidate QTL located on 52,048,208 bp on the Solyc11g071510.1.1 gene on chromosome 11 and this was confirmed by the result of the traditional QTL method which was consistent with the QTL-seq analysis.

Furthermore, the result of BLAST protein function analysis suggested that this candidate gene encodes glycoside hydrolases (GHs). GHs function as common degradation enzymes with a bond between a carbohydrate, a protein, lipid or another moiety, and are found in many kinds of organisms such as archaea, bacteria, animals and plants (Tyler et al., 2010). Consequently, genes encoding GHs are comparatively abundant in plants where they are involved in processes of starch metabolism, defense, and cell-wall remodeling (Tyler et al., 2010). GH genes play important roles in synthesizing carbohydrate-active enzymes in photosynthesis and in constructing carbohydrate- rich cell walls (Coutinho et al., 2003). Other functions of GHs in plants include pathogen defense, the degradation of starch, and hormone signalling (Minic, 2008). GH genes express to regulate functions in plant cell wall synthesis, renovation, and degradation (Minic & Jouanin, 2006; Lopez-Casado et al., 2008). In this context, GHs which participate in the degradation of cell wall polysaccharides are also implicated in the governance of plant cell wall loosening, the regulation of growth and development, germination, abscission, cell adhesion and fruit ripening (Fischer & Bennett, 1991; Minic, 2008). GH genes also play an important role during fruit ripening, with multiple enzymes promoting the disassembly of cell wall polysaccharides or polysaccharide domains and contribute to modifications in cell wall construction. The most characterized and studied cell wall degrading proteins in fruits were reviewed by Owino, Ambuko, and Mathooko (2005). These include GH enzymes such as polygalacturonases (PGs), ß-D-galactosidases, endo-ß-1,4-D-glucanases, and to a lesser extent endo-ß-mannanases, ß-D-xylosidases, α-D-galactosidase, and XET (Minic, 2008). ß-galactosidase II plays an important role in degrading galactan and the rise in its activity through tomato ripening suggests a possible role for this enzyme in tomato softening (Smith & Gross, 2000). During fruit ripening, pectin and some hemicellulosic polysaccharides gradually develop solubility and depolymerize by the release of neutral sugar residues from side chains of matrix polysaccharides (Huber & O'Donoghue, 1993; Brummell & Labavitch, 1997).

5. Concludsion

In summary of this study, as confirmed by traditional QTL and BLAST protein function analysis, QTL-seq detection found that a GH gene is related to the early fruit ripening trait in the tomato as GH genes are

functionally associated with cell wall degradation, fruit softening and ripening fruit. Thus, GHs may be important in fruit softening that stimulate early fruit ripening of tomato. These results established that QTL-seq is rapid and effective method to identify candidate QTL loci, candidate genes and candidate markers. In addition, our results are important for plant breeding and crop improvement because early ripening is not only one of the major earliness traits in tomato but also one of the important agronomical traits in crop plants.

Acknowledgements

The authors would like to thank the members of the Fresh-Tomato research group at the Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences and the Key Laboratory of Horticultural Crops Genetic Improvement, Ministry of Agriculture of the People's Republic of China for providing general help and support of this work, the Research and Development Office (RDO), Prince of Songkla University for providing general help in manuscript preparation. This study was funded by the National Natural Science Foundation of China (31471874).

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., … Terauchi, R. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology, 30*, 174-178. https://doi.org/10.1038/nbt.2095

Adams, S. R., Cockshull, K. E., & Cave, C. R. J. (2001). Effect of temperature on the growth and development of tomato fruits. *Annals of Botany, 88*(5), 869-877. https://doi.org/10.1006/anbo.2001.1524

Brummell, D. A., & Labavitch, J. M. (1997). Effect of antisense suppression of endopolygalacturonase activity on polyuronide molecular weight in ripening tomato fruit and in fruit homogenates. *Plant Physiology, 115*, 717-725. https://doi.org/10.1104/pp.115.2.717

Coutinho, P. M., Stam, M., Blanc, E., & Henrissat, B. (2003). Why are there so many carbohydrate-active enzyme-related genes in plants? *Trends in Plant Science, 8*, 563-565. https://doi.org/10.1016/j.tplants.2003.10.002

Darvasi, A., & Soller, M. (1994). Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics, 138*, 1365-1373.

Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Laxmi, K. V., … Parida, S. K. (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Research, 22*, 193-203. https://doi.org/10.1093/dnares/dsv004

Doganlar, S., Tanksley, S. D., & Mutschler, M. A. (2000). Identification and molecular mapping of loci controlling fruit ripening time in tomato. *Theoretical and Applied Genetics, 100*(2), 249-255. https://doi.org/10.1007/s001220050033

Fekih, R., Takagi, H., Tamiru, M., Abe, A., Natsume, S., Yaegashi, H., … Terauchi, R. (2013). MutMap+: Genetic mapping and mutant identification without crossing in rice. *PLoS ONE, 8*, e68529. https://doi.org/10. 1371/journal.pone.0068529

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., … Mueller, L. A. (2015). The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Research, 43*, D1036-D1041. https://doi.org/10.1093/nar/gku1195

Fischer, R. L., & Bennett, A. B. (1991). Role of Cell Wall Hydrolases in Fruit Ripening. *Annual Review of Plant Physiology and Plant Molecular Biology, 42*, 675-703. https://doi.org/10.1146/annurev.pp.42.060191.003331

Fisher, R. A. (1922). On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society, 85*, 87-94. https://doi.org/10.2307/2340521

Giovannoni, J. J., Wing, R. A., Ganal, M. W., & Tanksley, S. D. (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Research, 19*, 6553-6558. https://doi.org/10.1093/nar/19.23.6553

Gur, A., Osorio, S., Fridman, E., Fernie, A. R., & Zamir, D. (2010). hi2-1, a QTL which improves harvest index, earliness and alters metabolite accumulation of processing tomatoes. *Theoretical and Applied Genetics, 121*, 1587-1599. https://doi.org/10.1007/s00122-010-1412-8

Huber, D. J., & O'Donoghue, E. M. (1993). Polyuronides in avocado (*Persea americana*) and tomato (*Lycopersicon esculentum*) fruits exhibit markedly different patterns of molecular weight downshifts during ripening. *Plant Physiology, 102*, 473-480. https://doi.org/10.1104/pp.102.2.473

Illa-Berenguer, E., Van Houten, J., Huang, Z., & van der Knaap, E. (2015). Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theoretical and Applied Genetics, 128*, 1329-1342. https://doi: 10.1007/s00122-015-2509-x

Inai, S., Ichihashi, E., Sayama, H., & Ishimura, E. (2006). Practical use of QTL associated with yield and fruit quality in tomato. *Acta Horticulturae, 724*, 45-50. https://doi.org/10.17660/ActaHortic.2006.724.4

Kerr, E. A. (1955). Some factors affecting earliness in the tomato. *Canadian Journal of Agricultural Science, 35*(3), 300-308.

Kevany, B. M., Taylor, M. G., & Klee, H. J. (2008). Fruit-specific suppression of the ethylene receptor LeETR4 results in early-ripening tomato fruit. *Plant Biotechnology Journal, 6*, 295-300. https://doi.org/10.1111/j.1467-7652.2007.00319.x

Klee, H. J., & Giovannoni, J. J. (2011). Genetics and control of tomato fruit ripening and quality attributes. *Annual Review of Genetics, 45*, 41-59. https://doi.org/10.1146/annurev-genet-110410-132507

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., … Homer, N. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*, 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Lopez-Casado, G., Urbanowicz, B. R., Damasceno, C. M., & Rose, J. K. (2008). Plant glycosyl hydrolases and biofuels: A natural marriage. *Current Opinion in Plant Biology, 11*, 329-37. https://doi.org/10.1016/j.pbi.2008.02.010

Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., … Huang, S. (2014). QTL-seq identifies an early flowering QTL located near Flowering Locus T in cucumber. *Theoretical and Applied Genetics, 127*, 1491-1499. https://doi.org/10.1007/s00122-014-2313-z

Mansur, L. M., Orf, J., & Lark, K. G. (1993). Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics, 86*, 914-918. https://doi.org/10.1007/BF00211041

Michelmore, R. W., Paran, I., & Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America, 88*, 9828-9832. https://doi.org/10.1073/pnas.88.21.9828

Minic, Z. (2008). Physiological roles of plant glycoside hydrolases. *Planta, 227*, 723-40. https://doi.org/10.1007/s00425-007-0668-y

Minic, Z., & Jouanin, L. (2006). Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiology and Biochemistry, 44*, 435-49. https://doi.org/10.1016/j.plaphy.2006.08.001

Osorio, S., Alba, R., Damasceno, C. M. B., Lopez-Casado, G., Lohse, M., Zanor, M. I., … Fernie, A. R. (2011). Systems biology of tomato fruit development: combined transcript, protein, and metabolite analysis of tomato transcription factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. *Plant Physiology, 157*, 405-425. https://doi.org/10.1104/pp.111.175463

Owino, W. O., Ambuko, J. L., & Mathooko, F. M. (2005). Molecular basis of cell wall degradation during fruit ripening and senescence. *Stewart Postharvest Review, 1*, 1-10. https://doi: 10.2212/spr.2005.3.3

Powers, L. (1941). Inheritance of quantitative characters in crosses involving two species of Lycopersicon. *Journal of Agricultural Research, 63*, 149-174. https://doi.org/10.3923/pjbs.2006.2770.2776

Ruangrak, E., Su, X., Huang, Z., Wang, X., Guo, Y., Du, Y., & Gao, J. (2018). Fine mapping of a major QTL controlling early flowering in tomato using QTL-seq. *Canadian Journal of Plant Science, 98*, 672-682. https://doi.org/10.1139/cjps-2016-0398

Smith, D. L., & Gross, K. C. (2000). A Family of at least seven β-galactosidase genes is expressed during tomato fruit development. *Plant Physiology, 123*, 1173-1184. https://doi.org/10.1104/pp.123.3.1173

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., … Terauchi, R. (2013a). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal: For Cell and Molecular Biology, 74*, 174-183. https://doi.org/10.1111/tpj.12105

Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., … Terauchi, R. (2013b). MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. *The New Phytologist, 200*, 276-283. https://doi.org/10.1111/nph.12369

Thompson, A. J., Tor, M., Barry, C. S., Vrebalov, J., Orfila, C., Jarvis, M. C., … Seymour, G. B. (1999). Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant. *Plant Physiology, 120*, 383-390. https://doi.org/10.1104/pp.120.2.383

Tyler, L., Bragg, J. N., Wu, J., Yang, X., Tuskan, G. A., & Vogel, J. P. (2010). Annotation and comparative analysis of the glycoside hydrolase genes in Brachypodium distachyon. *BMC Genimics, 11*, 600. https://doi.org/10.1186/1471-2164-11-600

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research, 38*, e164. https://doi.org/10.1093/nar/gkq603

Appendix A

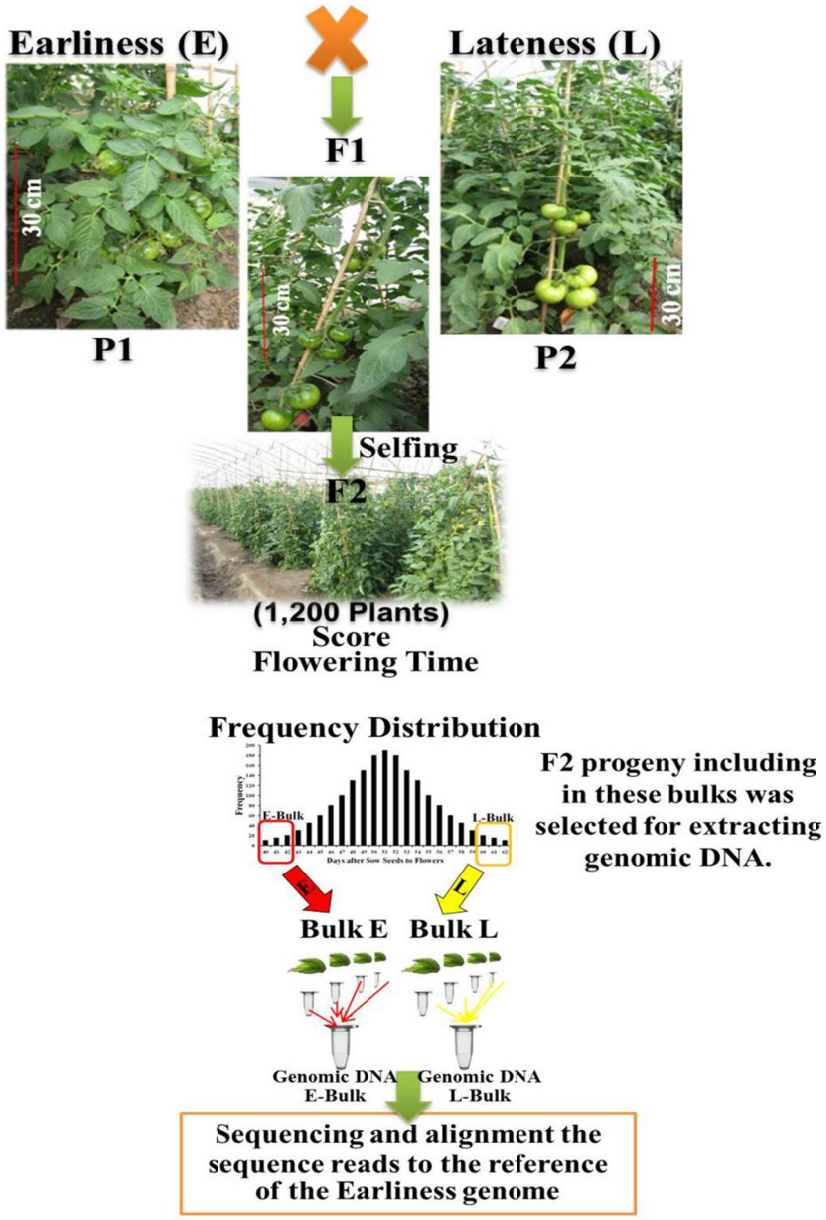

Figure A1. Illustration of the QTL-seq method used in this experiment

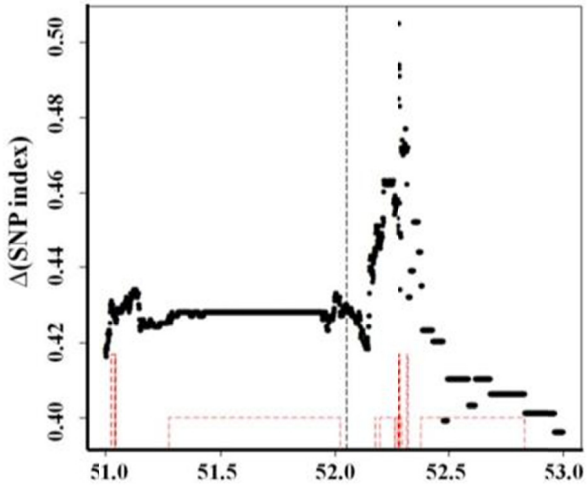

Figure A2. An average of Δ(SNP-index) was computed in a 1 Mb interval using a 1 Kb sliding window. The ΔSNP-index graphs of the candidate genes were plotted with confidence intervals under the null hypothesis of no QTL ($P < 0.05$). The positions of first fruit ripening time candidate gene (Solyc11g071510.1.1: 51M-53M) is located on chromosome 11

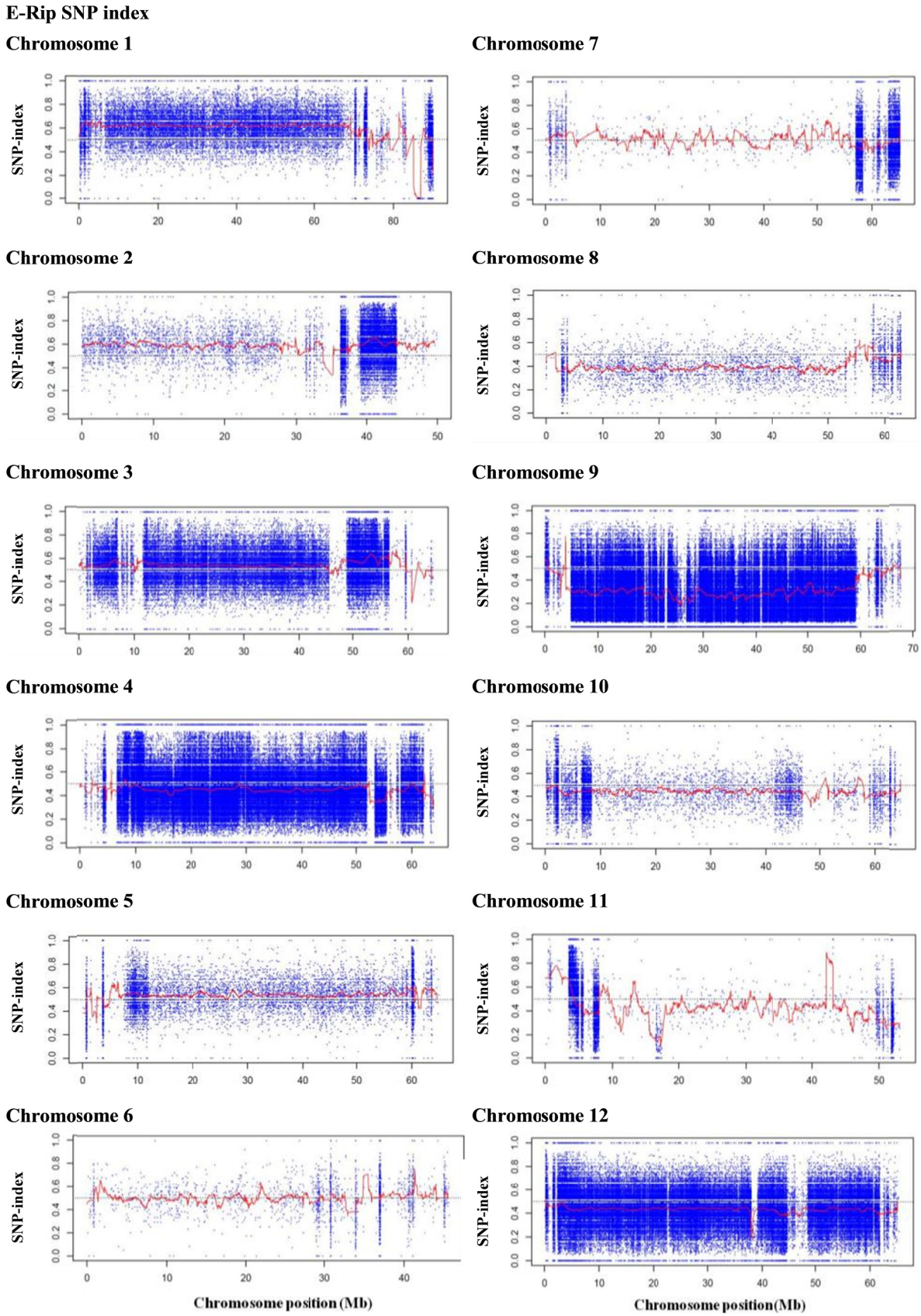

Figure A3. The SNP-index plots of twelve chromosomes which showed bulked DNA of the early fruit ripening (E-Ripening Bulk) at first fruit ripening time. The red lines are average values of the SNP-index or Δ(SNP-index) drawn by sliding window analysis

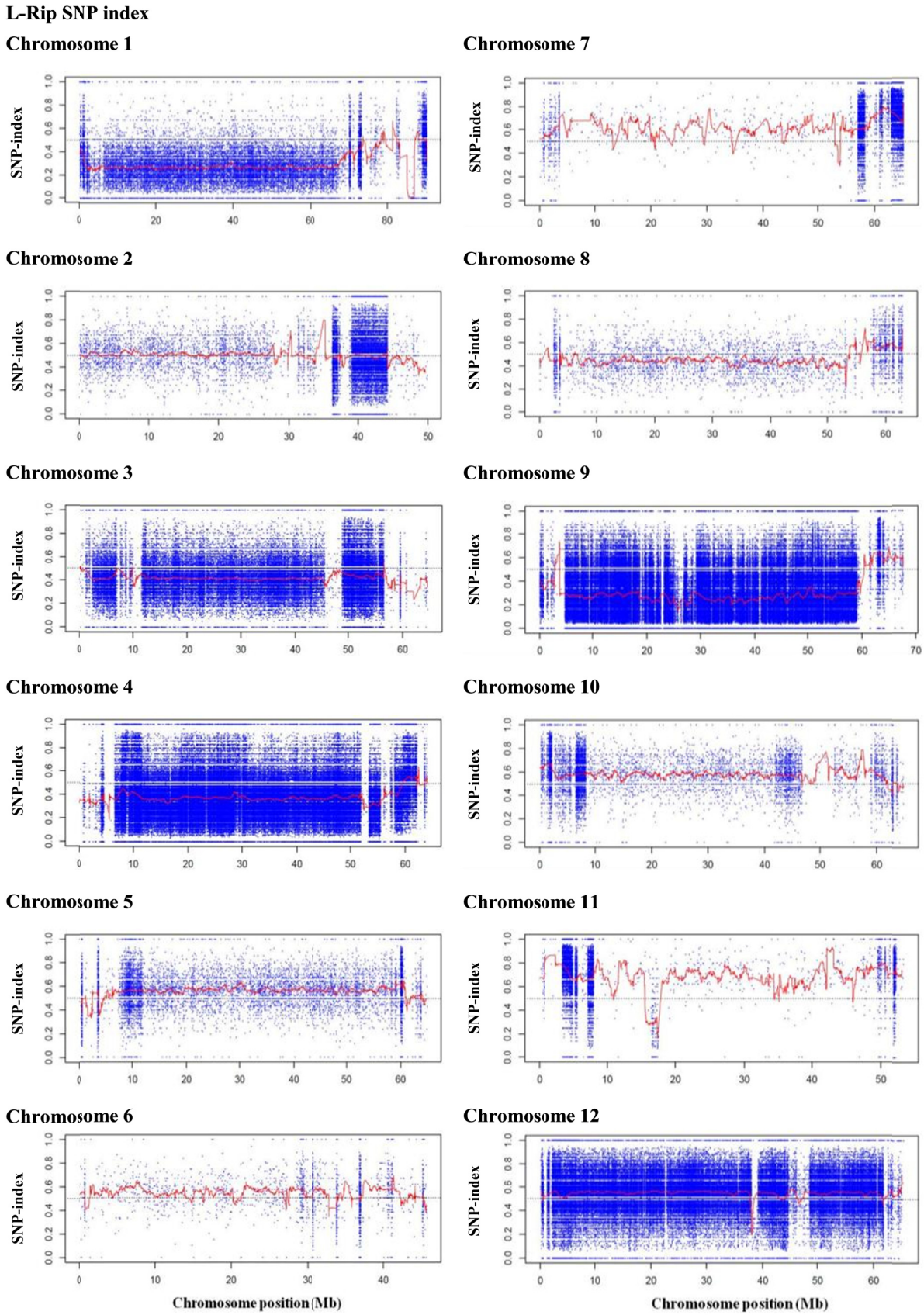

Figure A4. The SNP-index plots of twelve chromosomes which showed bulked DNA of the late fruit ripening (L-Ripening Bulk) at first fruit ripening time. The red lines are average values of SNP-index or Δ(SNP-index) drawn by sliding window analysis

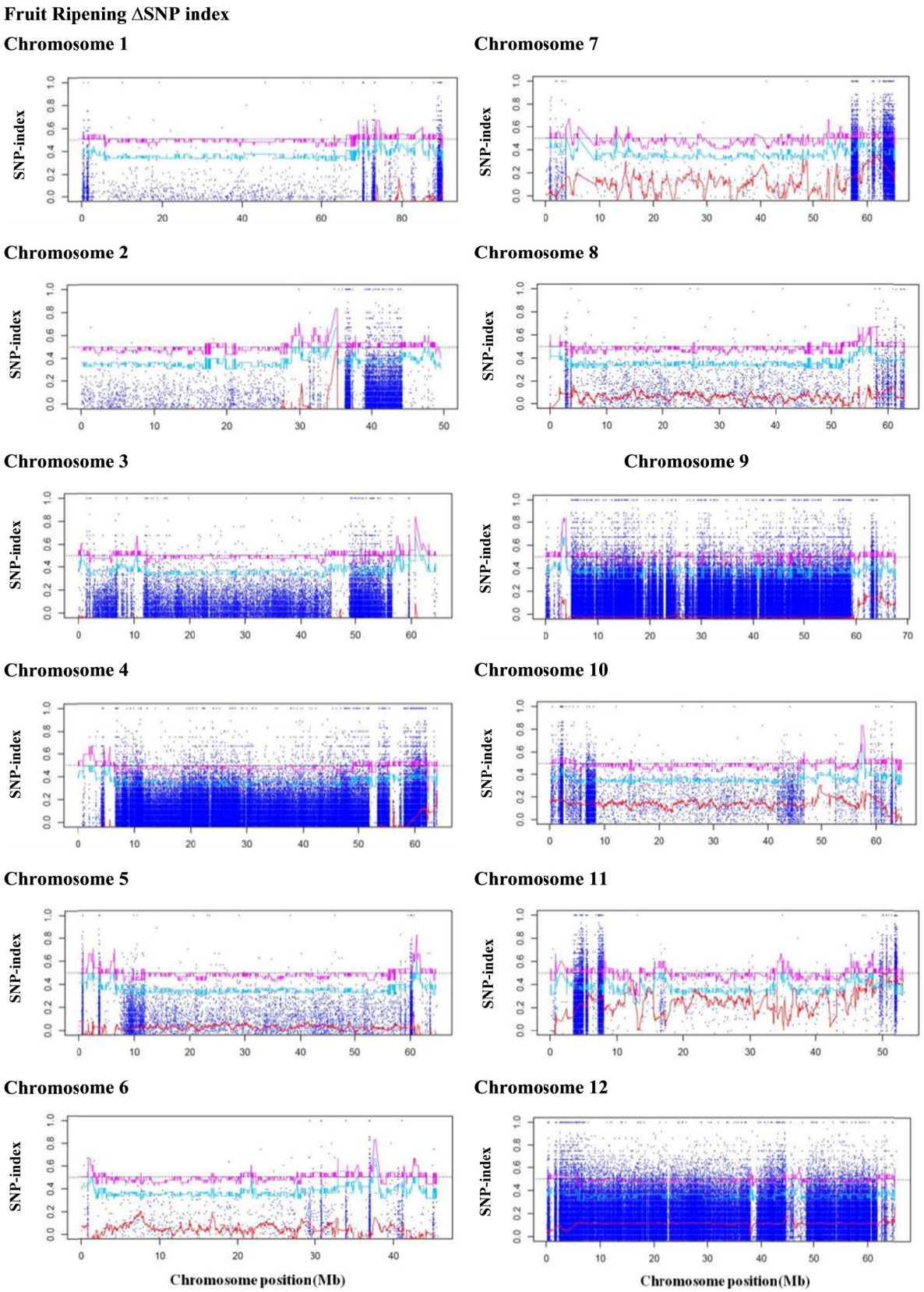

Figure A5. The Δ(SNP index) plots of twelve chromosomes which showed subtraction of E-Ripening Bulk and L-Ripening Bulk at first fruit ripening time. For the 1000 time permutation test selected 95% (blue line), 99% (purple line) confidence level as the screening threshold. The red lines are average values of SNP-index or Δ(SNP-index) drawn by sliding window analysis

Appendix B

Table B1. InDel markers were identified from chromosome 11

InDel	Position	Forward primer sequences	Reverse primer sequences (5′→3′)	size
1	2 838 463	AATATTCTGGCTTGTCGCTA	CAACCACAGGAGTAACCTTT	131
2	3 125 646	TACTCTTGTATAGTCCATTTCG	ATTTTGCTTCTTTGTTTACG	114
3	3 389 559	GCTTCTATTTAACAATTCCAAA	TGGAATTTTTCTCTTTTACCA	86
4	3 531 907	CAGAATTGAGCAACATTCAA	AGCTTGGATTCCCTTCTATC	156
5	3 597 324	TATTCCCCTCATTCCTTTTT	GAATGAAATGTGCAATGGTA	82
6	3 604 601	CTAGGGATAGGCATTTTCTG	GTACTTAATCAAAGCTCAGCC	158
7	3 675 820	TCTTCACTTGCAATCCTCTT	TGGCTTCAGAAATTTGTTTT	98
8	3 686 098	TTGGATCAAAATTTAGTTGG	TGATCGTAATTATTCAAAGAAA	109
9	3 707 839	TCTATTCGTTTGGGACAAAT	AATGATGAACTGAAAGGCAC	140
10	3 712 520	TGATCAAGATTTCAACACAAA	ATGCATTCAATGATCAACAA	99
11	3 715 610	TGAGTGGATAAAAATTCGGT	CTTCTTCTTCACACTCCACC	92
12	3 731 737	ATCAGTCGATGGTCTATTGG	CTAAATTTCTGTGGACACCC	152
13	3 756 267	TCAAAAATTCTCTCTCTCACAA	TCTTTCGGAAATAGAACAAAA	139
14	3 763 421	AAGTTGGACGTGTTGAGATT	GTTTTCATCAAGCGTCAAGT	95
15	3 776 827	ATTTCTTCTTTCCCTCCATT	CGTGACTAATTCAACTCATTTT	144
16	3 792 290	TGATTTAAGCTCTCATTCTTTT	GGAAATAGAAGAATCATCACAA	136
17	3 797 724	TTTATGTTTGTATAAAGCGAGC	TTGAATCGAAATAAAATGTTTG	81
18	3 799 535	TCATCAATAATAATGGGTCAAA	AATAGCAGCACTCACAAACA	136
19	3 962 344	GATTTTCAGTTCTTCATGGG	ACCGAAGCAACCATTAATAC	135
20	3 969 654	TATGTCAAACACTTTGCCTG	AACAATCTTCCAAACTCGAT	143
21	3 988 040	ATCTTCTTGTTGTTTCGACG	TTCATCATCAATCCCTCTTC	115
22	4 005 222	TGTTCAATCAAAAGTCATCG	TGATCCTGATCAGTTACAGGT	141
23	4 011 449	TCGGTCTGCAGAAATAATCT	TTGCATCCTTTTAAATCTTTG	155
24	4 033 022	CTGAATCGATGATGTAGGAA	TGCCACAACTTTTATACGAA	160
25	4 036 540	TGTCTCGACGATGTAAAGAA	TTTCCCAAAAACATATCACC	120
26	4 042 055	CATCTCTTTTGGAATAGACCC	ACAGGTGATTATCTGGTCAAA	100
27	4 086 995	TGTTTCTATTCGTGAACCAT	TTGTCAAATTCATGATTAAAAG	116
28	4 121 829	ATAAAGCCGCATATATTGGA	TATTTTAATGCGCTTCCTTT	148
29	4 125 698	TCCTCAAATTGTGTGAGATTT	CGAGCATTCATACTCGTTTT	157
30	4 187 615	AAAAGAAGGCTTTGATAAGTTC	TGCAATAACAAAGGAAGAAA	130
31	4 198 004	GGGAAAACCTTAGAATCTTGA	TTAGGTAGCGTTTTATGGGA	130
32	4 208 813	GAATGACAACCCTCTTTGAA	TTCTTCAACCATCACAAAAA	88
33	4 209 132	TTTGCATAGTTTTTGCTCCT	ACCACACCAAATTGACTTTC	134
34	4 236 711	CTTTCTTCCAAAATCAACAT	TTGAAATTGAGAGAGTACAAAG	84
35	4 237 861	TTGTTGTAGCATTGAACTGAG	TCTTTGGATCATTGATATTTTT	141
36	4 268 305	ATGCATTTTGTAAGTAAATCAA	TTGAGTGTAAAGGGCATAAC	103
37	4 271 847	ATCCCACGATTAAATCAGC	TCAATGCTCCCTCACTTATT	84
38	4 305 848	AAAACAGGCCAGCAATAATA	ACAAATTGAAACGAAGATGG	102
39	4 306 589	TCATTTAATATGAATCGCAAA	ATTGGTTAACAGGGTCACTT	114
40	4 316 701	TTCAATTCATCACCTACAACA	AGTTGTGACTAGCTCCTTGC	136
41	4 346 201	AGGAATTATGGGGGATTACA	CAAACATCGAATGAACAACA	154
42	4 347 369	CCCATCAAGGCATAGTATGTA	GTGACATGGTTATTCGACAA	148
43	4 352 220	AACGTCCCTTAAAACGTGTA	GAAATCGAGTCTCTGACGAA	113
44	4 358 218	TTTAGAAGGATGGCCAGATA	TCGAACGTGACCAATAAAAT	127
45	4 363 073	TTAAGGACAAGTAAGCTCCC	CTTGAGGCAAAACAAGAAGT	121
46	4 375 549	CTTCTGGGGTACTCTCTCCT	CGAAATTGATATACTATCGGTG	144
47	4 389 174	GTAAATCCTGAGTTGTTGGG	TTTGTTCTTTCTTAAACTCCG	86
48	4 389 947	ATTGGTGATCATCTTATCCG	TTGAACACAGAACCAGACAA	113

49	4 416 887	TAGGAGCCTAATCTGGAACA	TATCCAAGTATCTCATGCCC	132
50	4 450 426	GTTAAAAGTACCCCTAAGTGA	AAGCTTATCTTTATTAACTCCA	140
51	4 451 523	TACACCGGCTGTAAAGGTAG	ACCGAAAAGGGATAGAAAAT	159
52	4 456 449	TGGGGTTGTTCTGTTTTTAC	AAACACTGGGAAAATGTCAC	154
53	4 488 417	AAGGAACATGCTTCTTCAAA	TCATTATTGCTCAAAGAGGC	96
54	4 490 387	TTCAATGCCATGTATACCTTT	GTAGACTGTCTTGCAGAGGG	157
55	4 502 272	ACCTAAATGGACCATGAATCT	GCCAATATTGTAATCCCAAA	103
56	4 514 878	GGAAAAATTTGCCTTACATT	TTTTTAAACAAATGTCTCGAAT	130
57	4 526 314	CAAATATACCCGAATCTCCA	CTATGAGCGAAACTCCAAGT	112
58	4 542 717	TTTAATGTCATTCATGTGCAA	TAAATCAACGGAATCTGAGG	126
59	4 549 052	ACACGTGTATCACATTTTCCT	TGGCTAGCTATTTATACGCA	123
60	4 556 242	CAAAAGTGTTGACGTGCTTA	AAGATCACCCTCAACAATGA	152
61	4 556 864	CGTCAAACTATTGGGTCAGT	TGAAATCGTTTCTGGAAAAT	132
62	4 577 831	TTACAACAACTTCGTGCAAC	TTGAGTTTCGGGTATAAGTGA	155
63	4 605 203	AGCGAGACGTACCAAAAATA	AGACTTACGCCTCAATTTCA	88
64	4 618 882	ATAATTGCGCTCAATAGCA	GCCCAAGTCTCTTTCTCTTT	108
65	4 626 988	AAAAGCCCTTAAACAATCAA	TGATTTGTCAAAATGAACAAG	110
66	4 631 659	CGAATGCATCCTTTCTTAAA	CCCATTACCTAGCTTGACAG	145
67	4 638 021	AGTCACGAGCTTGAAATTCT	AACGAGCCATTATTGTCCTA	123
68	4 671 136	AACACTTGATCTACCAAACTCA	TTGGATTATGATTTTACCCTTC	144
69	4 703 160	CACCAAATTCTCTTCACCAT	AGAAATACACAGCAGTTGGG	111
70	4 704 795	TCCTAGCCATTTTGATGAGT	CTTTCATGAGCCTTTTCATC	144
71	4 719 917	TTTATTAATTTAAACGCCGC	TCTAAAAATGAGTGGACCAAA	133
72	4 722 366	AATGTTACATGACACCCCTT	GCTTGTTTTGTTTGTTCTCC	147
73	4 744 825	TATTTTAAACCCTTTTCCGT	GGTCAAACATTAGTCCTTTACA	132
74	4 747 534	CAAAATTCATAGCTAATTTCAA	ATTTAATTTTGTGGACGTGT	136
75	4 748 835	AAAAGGGTTAAAAGGTCAAC	TTTGGAGTAACTCTTCAAGG	114
76	4 749 358	ATGTATGCACCTTGTTTCTT	TTCTTGTATTTTGATTTAAACG	126
77	4 758 481	ATTACAAGCATCATAAGGGG	TTTATGGCATGAGAAGTCAA	139
78	4 765 664	GGAATTGGTAAGAGTCCACA	ATATGTCTCCCACTACCCCT	145
79	4 765 856	TGAGGTTGAAAGAATACCACA	GTTTTTCGACGTGAAATTGT	112
80	4 772 480	TTTTTCTGCTACAATCGAGTC	TGTTCGTGTCAACAAACATT	132
81	4 837 346	TTAATCACGTGCTTGCAATA	CAAACCCTAGCTATTCCATTT	156
82	4 838 504	CCTCTCTCCTCCCTCTCC	ATTTGTAAATATGGCAAGCG	152
83	4 866 926	TTTTCACCCTTACCATCG	TTTTTATAGTGGTGGGATAGG	157
84	4 867 489	ACACCTAAGGTTGACAATGG	CTTTGGGTCGAACTTTATTG	109
85	4 904 804	AAATTACGGGAAGTAGAAAAAC	TTTGTTTTGTTTAATGGAACC	108
86	4 920 067	AGAAAACTTTGGGAAAAAGG	TACAAGGCTCGTTTTTATGC	82
87	4 932 781	GTGAACGTTATTAAGTGGTTT	TTCATCACTATTTTAACCTTTT	155
88	4 937 081	ATAACCCCCATTCGTCTTT	TCTCTTTTTCTCGCATCAAT	139
89	4 941 284	TGTCAAACACCTTTTCATCA	CGGAATTCGGTAATAGAGTG	146
90	4 942 999	CATTTTCGGTAAGTTTTTGG	TGTCGAAAAAGAATTAAACGA	136
91	4 948 461	AAGTGAAAGTGTGCAACCAT	AGATTCCCTGAAACGTTCTT	153
92	4 983 330	TCCTTCAAGCCCATAAAATA	GATCATAACGACCCTATTGGT	134
93	4 985 718	TTCGTACATTGTGGTTGGTA	AATAATATCCTGGCGATCAA	157
94	5 001 460	AAAGATGTGAAGTCTAGCAAAT	TCGCGATATATATGTGTGTG	108
95	5 309 254	ACCAACCATCCCTTTATTTT	TCCATCACAATACTCAACGA	130
96	5 350 273	AAAGAAAGAAAAGAATCGGG	CTTTTGTTGAGCTTTGAAGG	158
97	5 361 713	AAAGCAAAATAGCGAGAGAA	CAACCCAAAAGCTTAATCAG	146
98	5 424 037	TGTTTGATATGTTATGTTCCCT	CCGAAACGAGACATTTAACTA	130
99	5 426 989	ACCCTCCAAAAATACATGC	GGATGAAATGGAAAAGACAG	157
100	5 428 746	AGGACAAATGAAAATCGAAA	CCAAATAAAATCAGCCCATA	138

101	5 463 728	CTCAATTTCAATCTTTATGCC	TAGCTAGAGGACCACAAACC	80
102	5 466 894	TGAATGGAGAAAAAGAAAAGA	AAATAGCCTAATTGGCAAAG	127
103	5 480 048	TGATACGTGAATAGGGATTG	AAAAAGAAAGCAAGATAAACG	83
104	5 515 271	GCCCTTGTCTTCTTCACTC	TAGTGAAAACGGGTTGAAGT	154
105	5 516 690	GCTTACGTTTGGACATTGTT	TTTTCAACCTAAGACCTCCA	139
106	5 535 304	GCATTCATCTAAAGGCAAAC	GAACAGATCTCACTTCGGTC	89
107	5 604 411	GGTCACCAAATATTCAAGGA	TTGCTGAAGTTTATCATGGA	92
108	5 608 446	GTAGTGCAACCAAAGACCAC	TAGCCTAATTGGTCGAGTGT	86
109	6 277 288	TCTATCCTTCCTTTTTGTGG	GGGGTGAATGTGACACTAAT	121
110	6 513 613	ACGGAGAAACCAAAAATGTA	AACACGGACGCATATATCTAA	121
111	6 555 783	TTCTTTTTAAAACGGAGAAAGT	TCACCTTCGATTCAATCTTT	108
112	7 179 708	TTTTTCCCTTATCCATTCAA	AGAAACCACCTACGAGATCA	129
113	7 207 977	CACACCCGCTAAAAGTTATT	GCCCAATCAACTATTTTTGT	107
114	7 208 496	GGAGATGAGCGAGATCTGTA	TCTCGCATGTATCCCTATCT	121
115	7 223 638	GTTAGGTTTCAGTTGCCGT	CGAACTTAGTCCATCACCAT	89
116	7 239 191	GGGAAAAATTTGTCTTTCTTG	CTATTTCTGTCCTTTGGCAC	149
117	7 281 558	TATCATGACCGATGTATTCG	TCTTTTGGAATTGACTTCTGA	140
118	7 291 112	TCAAAAGAAAAGAGAAGGTGA	CCTTTTTCCCTTTATCCAAT	155
119	7 292 091	ATTTGTGAATTTTTGGCATT	CACTATCCGACCCATTTTTA	122
120	7 313 245	ACCCTTTTCTTCTAACGGAA	AAAAATATGTTTGACGAGGATT	116
121	7 322 100	TTCCAAAGAACCGTTACACT	TTAACAACCTTCCACATTCC	141
122	7 346 925	ATTTTAGGTAGGTTTTCACGA	GACTTTTAACAGTGAGCGTG	80
123	7 439 153	GCTCGCATGATCTTAATTCT	TCGATAGCTCACTTTGAACA	108
124	7 541 872	CAAGAAAAATGCACAACAAA	CTCCTTTCCATTAGCATCAG	104
125	7 568 277	CCAAAGAACTGTTACACTTTCA	TTAACAACCTTCCACATTCC	127
126	7 574 712	TGGTCTAATGAAGCCTTTGT	ACCGGATAGGGATATTGACT	136
127	7 615 512	ATTTTTGAGAAGTTCGACCA	ATTGAGTCGCTCACGTAAAT	99
128	7 651 161	CATAAGCTCTTCAAATTGCC	TTGGAGAAAACACAACATCA	142
129	7 659 778	CCGTTTTAAAAAGATTGACA	ATGTGGCATGTTTAAGATCA	143
130	7 660 678	TGTCCCTTGCTTTACATTTT	AAGCTGTTTTGAGTATTGCC	112
131	7 679 022	TAATTGAAATGGATGAACCC	CTTCATCTTCATCTTCCTCG	153
132	7 693 896	TTTTTGTTATGTTTGGCAGA	TTCTCTCTCTCACGCTTTTT	92
133	7 716 571	ATGACTTCCAGCCAAATCTA	TCAAGCAATACAGAGTCGAA	147
134	7 717 308	AAGTTCTTTCCATTCTTCCC	GGCTTTCCAAATTCCTATCT	84
135	7 735 695	TGTGGGAGAGAGGTAAATTG	AATTACGTGGACTGACTTCG	133
136	7 741 318	GAATTGCATCGTTTTTCTTC	AAAATTAAAACCAACGGACA	113
137	7 828 980	TCAAGTTTCCTTTTGCTTTC	AGTCCCTATCCACAGATCCT	127
138	7 847 378	GAGAGACGTAGGAGAGAGG	TTCGTCAGAATATACAATTACA	145
139	7 903 813	TCTACTCGAGGGTTTAAGGA	CCGAACCAGATTACAAATAAA	142
140	7 928 572	TATGGCTGTCACAAACAGAG	AATTTCGATTCGATTTTGAG	91
141	7 946 864	ACCACATTTAAGCACCAAAC	ACGTGAAGTGTGAGTTTTATTG	133
142	7 949 256	GAATTTGGGTTTTGATTTTG	TAAAGTGAAGGGGTGAATTG	141
143	7 995 950	CACTTCATTAACGGGGTAGA	AGTGAGGGAACAATTTCTGA	109
144	45 230 496	CGGTCATCTAAAACACCATT	GAATCTTCGAACAGAGGCTA	155
145	45 258 180	CCCTGAAAAGTTAAGATAGCA	AGCATTATTTTACGGTTTAATG	97
146	45 370 701	ATGGATATATGGATGCCAAC	GATCCTTTTCCAAAGATGTG	147
147	45 709 190	TTTAGGGAAAATGCACAAGT	CGAAAAGGGGTAGAAAATTAC	141
148	46 125 680	GACGTGAACAACGTGACTAA	TACGCCTGACTCCTTATCAT	131
149	46 216 965	AAACTCAAATCCCAAATTCA	CTGTTGTTGATGCTGCTAGA	100
150	46 725 937	TACTTCAATTAAAGCTCGCC	CCCTTTGATCTATGTCGTGT	105
151	46 764 362	TTTTCTCTCTCTCTTTTCACG	GAATTGGGGTTGATGATAAG	107
152	46 858 505	AAATGCAAGTTTACAAGGGA	AAGACTTGGACTTGGACTTG	118

153	47 566 184	AAATAATTTTCCGAACTCCC	GGAGAAAACGATTTCACAAA	96
154	47 973 834	TGCAGGAGTAGACTGAAACC	TTTTTGAGTGACTATATGCGAA	114
155	48 262 849	ATACTCGCTTTGCTTCAAAC	GCAATCCAAGCTACAGAAGT	133
156	48 404 570	CGCAACTCCTCTATCCTAAA	AGAGAATTAAATGGCGTTTG	109
157	48 771 672	TATCAACGTTTGGGATTTTC	CTCAATTTCATTTTGAACCTTT	126
158	48 874 074	GTGAACCAAACTTTAAACGG	GTTTGGTATTTCGTTGTGGT	150
159	49 204 628	CTCTTAACACCACTTTTGCC	ATGAAGAATTTTGGGGTTTT	90
160	49 256 128	ACCTATGGGGTCCTACTGTT	TGTATGCCTGTGATTTGTGT	120
161	49 808 261	TTTTTCGACGTAATTTCAGA	ACTAGACATAACATTGGGTCC	100
162	49 808 266	TTTTTCGACGTAATTTCAGA	ATAACATTGGGTCCAAGAAT	94
163	49 976 324	GCGCATACACCACAAAATTA	CACGTAGGCGGTCCTAATA	144
164	50 000 357	AATGAGTTCAAGGGGGTAAT	GCACATTTTGGGACATACTT	121
165	50 005 083	ATACATTTGTCCAAACCTCC	AAAAAGTATTGGTGTGGTTTT	112
166	50 159 967	TTGATATTGTTGACACCCAA	AAGGATGAAATACCGATGTG	151
167	50 593 726	CCCTGATCACTGCTAAACAT	CCACGTACTTACTTGTGGGT	137
168	50 623 429	TCAAATAATTTTGTTCTTGTGA	TCCCAATTTAAATAAAAGCA	154
169	50 806 100	ATTATATGTTTGGTCTCGGG	TCGACTAGTGGTTTGGTTTT	116
170	51 139 343	GCTCTTGGCAAAAGTAAAGA	TCGAAATTTCTGAAAGACGTA	160
171	51 197 947	CCCCATTAATCCTTTTTCA	CTTTAGTTGGTTTTGTTTGTGA	128
172	51 262 762	CGTGAAAAACTTCGAAAAAT	CGAGTTCGAGATTCTGTTTT	87
173	51 272 827	TTGAATTCATCGACAAATAAGA	CATGGAATAGGGATAGTGTCA	105
174	51 370 255	ATGGTTACCAGTCTCTGTCG	AGAAACAGGGGATACTTTGG	152
175	51 922 055	GTAGCCGTAGCACTTGAGAC	AACACGAAAAGAACATCACC	108
176	51 953 001	TTGAAGAAAAAGTTCACAAAAA	CCCTTTTTCCAAGAATGTAA	98
177	51 956 466	TGTTTGATTAAATTGGTGAAAT	ATAGCCTCTTTTCAATCACC	118
178	51 965 954	TGTCGTCACTGACTATTTGG	CTCCTTGAGGAAAGGACTCT	117
179	51 990 155	TCACAAAGGTTTGCTGAATA	TTAACATCATTTTGAGTGCG	118
180	52 006 616	CTCTCACCTCCCTCCTATCT	TAAGAGATAAATTGGGCTGG	159
181	52 034 258	TTGGGTTATTTCAGATTGTTC	GGGTACGGACCCTCTACTAT	127
182	52 045 558	TTGGTTACTGTCCTTTGGTT	AAGCAGTTTGCCATTTATGT	94
183	52 084 043	TTACCCCTTCTCTATTGTGC	TTTGGTTAAGCCAACTTTTC	141
184	52 086 183	TAGCATCCATGACTTTGTGA	ACAAGGATTGGTCAATATGC	134
185	52 107 486	TAAAGTCGTCGCTAAAGGTT	GCGACAACTTCATACCTCTT	123
186	52 119 815	TGACACTAAACTAATGCCGA	AAAGGCAAGAGGATTTGATA	121
187	52 120 967	AGGTTGATGGACAATATCTGA	CCATCCACTATCCCTCTTTA	155
188	52 125 009	AACTTTGCTTCACTTCTTGC	CATACAAACAGGCTTGAACA	155
189	52 139 230	CCCAAATGGACAAATCATTA	TTTGGAAATTTTATCCGTTG	103
190	52 143 232	GGATTACTGACCCTCCTCAT	ATCGAAAGATCTGTGTACGAA	110
191	52 146 442	TCATCTTTCGAGTCGAGATT	TATCCATTTTGTATAGGGGC	80
192	52 150 991	TGAAAGTCTGATGCACAAAG	TTCTCCATAACACAACACGA	121
193	52 151 475	TGAAGGAAACAATGTCACAA	TACCTGAAAAGAAATCGGAA	122
194	52 153 089	TCCTACAGCTATGTTGCTCA	CATTGTTCTGTTTCATGTCG	100
195	52 180 545	GGTCTTCAGGAATCTCAACA	CGAAGTTGTGGTTGTTACAG	122
196	52 204 378	TGTTTGTCACAAGTATCTGTTG	AAATTAGTCGCGTTCCATAC	141
197	52 268 761	TTTTACGTATCACGTTTATTTT	CATTTCACATGCGATAGTTA	135
198	52 270 046	TTGTTTTGTAAAATGATATGGA	TAAAATACACGTGATTCCG	96
199	52 321 549	AATGTCTGGTTCTGATGAGG	GGTTAGAAGCAAGTCCATTG	134
200	52 384 121	ATTTCTCAAATCTGCAAGGA	TTGTTTGCAAATTACTCTCAAC	122
201	52 388 226	TAGGAGGACCCTCTTAAAGC	CACCATCAGGTCTAGTCACC	94
202	52 747 446	TGTACCGTCTTTCTTTTGCT	GAAAAATGTCAATGAAGAGGA	142
203	52 855 027	ATGTTCCCTTCAATCATTTG	TTGTCAACAACCCAAACTAA	133

Table B2. Sequencing depth and coverage

Sample	Total reads[1]	Mapped reads[2]	Mapping rate (%)[3]	Average depth (X)[4]	Coverageat least 1X[5] (%)[5]	Coverageat least 4X[6] (%)[6]
P_1 (Bone MM)	144 593 794	139 246 400	96.30	17.68	99.19	98.27
P_2 (440-071)	136 509 578	127 146 260	93.14	16.31	96.04	94.17
E-Ripening Bulk	175 329 770	166 584 109	95.01	21.09	99.56	97.92
L-Ripening Bulk	152 560 340	144 286 570	94.58	18.19	99.23	95.82
Average	152 248 371	144 315 835	94.76	18.32	98.51	96.55
Total	608 993 482	577 263 339				

Note. [1]Effective sequencing data; [2]Comparison to the reference genome of the read numbers (including single end alignment and double end alignment); [3]Reference genome reads divided by the effective sequencing data; [4]The average sequencing depth; [5]A reference genome has at least 1 base covered per site accounting the genome; [6]A reference genome has at least 4 bases covered per site accounting the genome.

Table B3. SNP detection and annotation

Category		Number of SNPs
Exonic[1]	**Stop gain[2]**	468
	Stop loss[3]	168
	Synonymous[4]	13 230
	Non-synonymous[5]	18 955
Intronic[6]		92 638
Splicing[7]		180
Upstream[8]		63 263
Downstream[9]		52 837
upstream/downstream[10]		4 874
Intergenic[11]		1 622 942
Ts[12]		991 440
Tv[13]		879 058
Ts/Tv[14]		1 127
Total		1 870 498

Note. [1]The number of SNPs in an exon region, including stop gain, stop loss, non-synonymous and synonymous; [2]Introduction of a stop codon; [3]Loss of a stop codon; [4]Missense non-synonymous regions; [5]The number of SNPs presumed to be silent; [6]The number of SNPs in introns; [7]Splicing regions are located in the splice site (near the exon/intron boundaries of the 2 bp intron); [8]1 Kb downstream region; [9]1 Kb upstream region; [10]1 Kb upstream/downstream gene; [11]The number of SNPs in regions between genes; [12]Transitions (ts) are interchanges between a purine base and another purine (A↔G) or replacement of a pyrimidine with another pyrimidine (C↔T); [13]Transversions (tv) are interchanges between the purine and pyrimidine bases (T↔A, T↔G, C↔A, C↔G); [14]ts/tv; Transition/transversion ratio.

Table B4. The annotations of the candidate polymorphic marker loci

Category		Number of SNPs
Exonic[1]	**Stop gain**[2]	0
	Stop loss[3]	0
	Synonymous[4]	7
	Non-synonymous[5]	8
Intronic[6]		41
Splicing[7]		0
Upstream[8]		46
Downstream[9]		15
upstream/downstream[10]		2
Intergenic[11]		315
Ts[12]		267
Tv[13]		167
Ts/Tv[14]		1.598
Total		434

Note. [1]The number of SNPs in an exon region, including stop gain, stop loss, non-synonymous and synonymous; [2]Introduction of a stop codon; [3]Loss of a stop codon; [4]Missense non-synonymous regions; [5]The number of SNPs presumed to be silent; [6]The number of SNPs in introns; [7]Splicing regions are located in the splice site (near the exon/intron boundaries of the 2 bp intron); [8]1 Kb downstream region; [9]1Kb upstream region; [10]1 Kb upstream/downstream gene; [11]The number of SNPs in regions between genes; [12]Transitions (ts) are interchanges between a purine base and another purine (A↔G) or replacement of a pyrimidine with another pyrimidine (C↔T); [13]Transversions (tv) are interchanges between the purine and pyrimidine bases (T↔A, T↔G, C↔A, C↔G); [14]ts/tv; Transition/transversion ratio.

Table B5. Sequencing data quality

Sample	Raw Base (bp)[1]	Clean Base (bp)[2]	Effective Rate (%)[3]	Error Rate (%)[4]	Q20 (%)[5]	Q30 (%)[6]	GC Content (%)[7]
P_1 (Bone MM)	15 048 951 400	14 459 379 400	96.08	0.03	97.33	92.29	36.17
P_2 (071-440)	14 217 646 400	13 650 957 800	96.01	0.03	97.33	92.28	36.67
E-Ripening Bulk	18 162 823 000	17 532 977 000	96.52	0.03	96.61	90.56	36.5
L-Ripening Bulk	15 876 913 600	15 256 034 000	96.08	0.04	96.38	90.08	37.8
Average	15 826 583 600	15 224 837 050	96.17	0.03	96.91	91.3	36.79
Total	63 306 334 400	60 899 348 200					

[1]The original data yield of DNA sequences; [2]The effective data after filtering; [3]The ratio of raw data after filtering to obtain clean data, error rate and base error rate; [4]Base error rate; [5, 6]Phred quality scores ≥ Q20 and Q30; [7]The percentage of GC content.