Genetics and Molecular Research

Analysis of genetic diversity and population structure in a tomato (*Solanum lycopersicum* L.) germplasm collection based on single nucleotide polymorphism markers

T. Wang[1,2], Q.D. Zou[2], S.Y. Qi[2], X.F. Wang[2], Y.Y. Wu[2], N. Liu[3], Y.M. Zhang[2], Z.J. Zhang[2] and H.T. Li[2]

[1]College of Horticulture, Shenyang Agricultural University, Shenyang, Liaoning, China
[2]Institute of Vegetables, Liaoning Academy of Agricultural Sciences, Shenyang Liaoning, China
[3]Institute of Soil and Water Conservation, Liaoning Academy of Agricultural Sciences, Chaoyang, Liaoning, China

Corresponding authors: H.T. Li / Z.J. Zhang
E-mail: haitao57@sohu.com / huixue224@163.com

Genet. Mol. Res. 15 (3): gmr.15038209
Received December 7, 2015
Accepted March 11, 2016
Published July 29, 2016
DOI http://dx.doi.org/10.4238/gmr.15038209

ABSTRACT. Knowledge of genetic diversity is important to assist breeders in the selection of parental materials and in the design of breeding programs. In this study, we genotyped 348 inbred tomato lines, representing vintage and contemporary fresh-market varieties, by using 52 single nucleotide polymorphisms (SNPs); 45 of these were found to be polymorphic. The average minor allele frequency and unbiased expected heterozygosity were 0.315 and 0.356, respectively. Population structure analysis revealed that contemporary germplasm

could be distinctly divided into six subpopulations representing three market classes and breeding programs (pink, green, and red). Vintage germplasm could be separated into at least two subpopulations, and more admixtures were found in vintage lines than in contemporary lines. These findings indicate that contemporary inbred lines are more diversified than vintage inbred lines. AMOVA of vintage and contemporary lines was performed. A significant difference was found ($P < 0.01$), which explained 17.4% of the total genetic variance. Subsequently, we constructed a core collection using 45 polymorphic SNP markers. The data showed that all alleles were captured by only 2% of lines, indicating that more alleles, as well as rare alleles, could enable more variation to be captured in the core collection. These data allow us to discard redundant inbred tomato lines and to select elite inbred lines, which will accelerate the breeding process.

Key words: Tomato; Genetic diversity; Population structure; SNP marker

INTRODUCTION

Tomato (*Solanum lycopersicum* L.) is one of the most economically important vegetable crops in the world. It is also a model system used to investigate fruit development (Carrari and Fernie, 2006) and carbohydrate metabolism (Obiadalla-Ali et al., 2004), and for quantitative trait locus (QTL) analyses (Zhang et al., 2012). The tomato is believed to have originated from the Andean region (Bai and Lindhout, 2007). Mesoamerica was probably the first center of domestication, and although the exact date of domestication is unknown, tomato was already being cultivated in southern Mexico, and probably in other areas, as early as 500 BC. By the 16th century, Spanish explorers started to cultivate tomatoes in Europe. Around the 17th century, the tomato was introduced to China by the Portuguese. Presently, China is the world's largest tomato producer, with an annual production of 50 million tons, which accounts for about one quarter of the global output (FAO, 2012).

During its long evolution and domestication, a large number of contemporary tomato cultivars with different shapes and colors have been developed from wild species. Compared to their wild progenitors, cultivated tomatoes exhibit far greater phenotypic variation, including increased fruit size, diverse colors and shapes, and improved quality and taste. However, the large range of phenotypic variation is not representative of greater genetic variation in domesticated tomato. Cultivated tomato exhibits lower genetic diversity than other autogamous crops (Miller and Tanksley, 1990), which can be explained by a series of population bottlenecks (Ranc et al., 2008). The long period of selection by humans narrowed the genetic base of the cultivated tomato. For this reason, the polymorphism level of molecular markers in the cultivated tomato gene pool is lower than that in other autogamous species (Archak et al., 2002). In recent years, researchers have started to develop molecular markers for cultivated tomato and to construct intraspecific genetic linkage maps (Shirasawa et al., 2010; Hamilton et al., 2012; Yang et al., 2014). Therefore, increased numbers of molecular markers with a higher level of polymorphism can be applied for QTL mapping and genetic analysis in the cultivated tomato.

Cultivated germplasm resources provide the genetic basis for both breeding and

genetic research. With the introduction of contemporary commercial hybrids, an increasing number of vintage varieties have been abandoned or preserved in germplasm banks (Casals et al., 2011). The majority of vintage varieties are cultivated by home gardeners. This can be attributed to the high yield and disease-resistance offered by contemporary commercial varieties. However, vintage tomato varieties can provide some interest traits, such as biotic stress-resistance and high-quality fruits (Hawtin et al., 1996). Therefore, it is essential to analyze the genetic diversity of vintage tomato varieties.

Genetic diversity can be evaluated using morphological traits or DNA markers. Morphological traits are the simplest way to investigate genetic diversity, but are often influenced by the environment (Zhou et al., 2015a). DNA markers help us to understand genetic variation at the DNA level. Therefore, several molecular marker systems have been applied to the study of genetic diversity in crops, including restriction fragment length polymorphism (RFLP) (Garcia-Mas et al., 2000), random-amplified polymorphic DNA (Korkmaz and Dogan, 2015), amplified fragment length polymorphism (van Berloo et al., 2008), simple sequence repeats (SSR) (Zhou et al., 2015b), single nucleotide polymorphism (SNP), and insertion-deletion (Sim et al., 2011) markers.

SNPs are the most frequent type of sequence variation in the genome (Gupta et al., 2008). Recent advances in sequence analysis have triggered a shift toward the use of SNPs in many species, particularly in model organisms. In the present study, 348 inbred tomato lines representing both vintage and contemporary varieties were genotyped using 52 highly polymorphic SNPs. The main objectives of this research were to characterize these SNPs in a large number of tomato germplasm, and to compare the genetic variability of two tomato groups. Furthermore, these data may enable us to discard redundant tomato lines and to construct a core collection.

MATERIAL AND METHODS

Plant materials

A collection of 348 cultivated inbred tomato lines was selected for use in our study. One hundred of these lines were from the China National Vegetable Germplasm Bank located at the Institute of Vegetables and Flowers of Chinese Academy of Agricultural Sciences, 174 lines were from Liaoning Academy of Agricultural Sciences, and 74 lines were obtained from the Centre for Genetic Resources, the Netherlands (CGN, www.cgn.wur.nl). The germplasm collection consisted of 189 vintage inbred lines and 159 contemporary fresh-market inbred lines. The vintage germplasm collection included traditional varieties, landraces, and heirlooms and represents early tomato improvement efforts. Contemporary lines represent commercial fresh-market hybrids in China. Most of these contemporary lines were selected from the offspring of commercial hybrids, and each line underwent at least six generations of self-pollination. Detailed information on the materials is provided in **Table S1**.

Molecular markers

Fifty-two SNP markers (**Table S2**) were selected to genotype the 348 tomato inbred lines. Twenty-two SNP markers were designed from expressed sequence tags that had been mapped onto intraspecific linkage maps in tomato (Shirasawa et al., 2010). The remaining 30

SNPs were discovered in cultivated tomato via sequencing (Hamilton et al., 2012) and showed a higher degree of polymorphism among the tomato cultivars (Sim et al., 2011).

Genotyping

Genomic DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) method (Stewart and Via, 1993). SNPs were scored as cleaved amplified polymorphic sequences (CAPS), and restriction enzyme cut sites were detected using SNP-RFLPing V2.0 (Chang et al., 2010). SNPs without restriction enzyme sites were converted to dCAPS markers, and the PCR primers were designed using Primer 3 (primer3.ut.ee) based on the sequences between the two marker alleles. All primers were commercially synthesized by Sangon Biotech (Shanghai, China). Genotypes of SNP markers were identified through the size of restricted fragments.

PCR was performed in a 10-μL volume, and each reaction contained 3 μL (5-20 ng) genomic DNA template, 5 pM each primer, 0.25 mM each dNTP, 1X PCR buffer, and 0.5 U *Taq* DNA polymerase (Sangon Biotech). Amplifications were performed in an ABI thermal cycler programmed for 5 min at 95°C followed by 35 cycles of 30 s at 94°C, 45 s at an appropriate annealing temperature, and a 30-s extension at 72°C. Final reactions included an extended incubation at 72°C for 10 min. SNP marker amplicons were digested for 4 h in a 20-μL reaction volume with 1 U appropriate restriction enzyme (Thermo, Shanghai, China). After digestion, 5 μL from each reaction mixture was separated on a 3% agarose gel and visualized using a BIO-RAD gel imaging system.

Data analysis

For all 348 inbred lines, as well as the subpopulations, minor allele frequency (MAF), number of different alleles (N_A), observed heterozygosity (H_O), unbiased expected heterozygosity (H_E), and Shannon's information index (I) were calculated with the GenAlEx 6.5 software (Peakall and Smouse, 2012). Cluster analysis was performed with PowerMarker v3.25 (Liu and Muse, 2005), a neighbor-joining tree was constructed based on shared alleles, and the MEGA6.0 software (Tamura et al., 2013) was used to view the phylogenetic tree.

Population structure was investigated using the STRUCTURE software (Pritchard et al., 2000). In order to identify the optimal number of populations (K), the value of K was set from 1 to 10, and three individual runs for each K with a burn-in of 10,000 iterations and a run length of 100,000 iterations were applied in addition to an admixture model and correlated allele frequencies. In the model-based method, individuals with inferred ancestry ≥0.65 were assigned to the corresponding population, and individuals with inferred ancestry <0.65 were assigned to a mixed population. Subsequently, the delta K method described by Evanno et al. (2005) was used to test the optimal K value.

Core collection

The core collection of 348 tomato cultivars was constructed by the MSTRAT software version 4.1 (Gouesnard et al., 2001). Two algorithms were used for analysis, namely maximization (M) strategy and random sampling. We compared the results obtained using both algorithms and selected the minimum size that captured the genetic variation present in

all tomato lines. To identify the minimal size of the core collection, MSTRAT was run with 20 replicates and 20 maximum iterations.

RESULTS

Characterization of SNP

We used 52 SNPs to genotype 348 inbred tomato lines (**Table S1**). Weak or unrecognized bands were scored as missing values. Seven markers were discarded due to a lack of polymorphism or because >20% missing values were found for all lines. Ninety alleles were detected for the remaining 45 polymorphic SNP markers. Base changes involved A/C (9), A/G (13), A/T (1), C/G (4), T/G (7), and C/T (11). A list of these informative SNPs including chromosomes, base changes, MAF, and $H_{E,}$ is provided in **Table S2**. Of the 45 polymorphic SNP markers, only two showed a value for MAF of less than 0.05. Approximately 82.2% (37/45) of SNPs had an MAF of more than 0.2. The average MAF was 0.315, ranging from 0.015 to 0.493 (Figure 1).

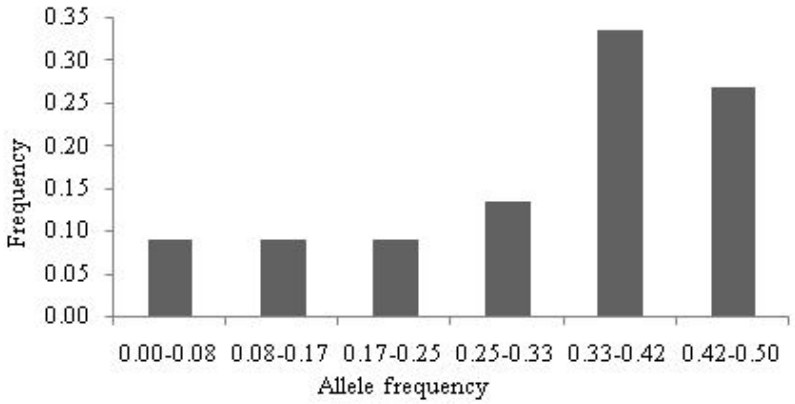

Figure 1. Distribution of minor allele frequencies for 45 polymorphic single nucleotide polymorphism markers.

Structure analysis

Population structure of the 348 tomato inbred lines was investigated using the STRUCTURE 2.3.4 program (Pritchard et al., 2000). The results showed that the maximum delta K was detected at K = 2 and the second delta K peak was detected at K = 9 (Figure 2). Given that the delta K peak at K = 2 may be an artifact (Vigouroux et al., 2008), the 348 tomato cultivars were further divided into nine populations (Figure 3). A total of 252 (72.41%) inbred lines were clearly assigned to a single population. The vintage inbred lines were divided into three populations consisting of 34, 56, and 29 lines, respectively. There were 67 lines (35.45%) with ancestry <0.65 that were classified as admixtures, and 3 lines were misclassified as contemporary tomato lines. The contemporary tomato inbred lines were divided into six populations consisting of 23, 25, 25, 36, 13, and 8 lines, respectively. The remaining 29 lines (18.24%) were classified as admixtures.

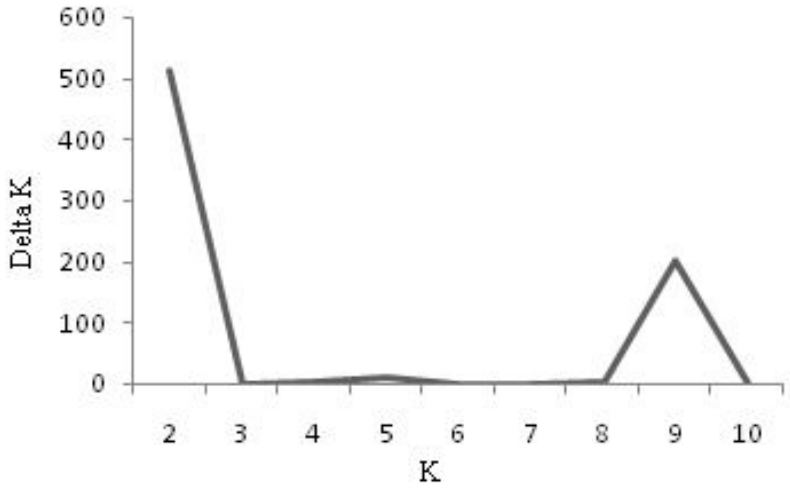

Figure 2. Delta K values for different numbers of populations assumed in the STRUCURE analysis.

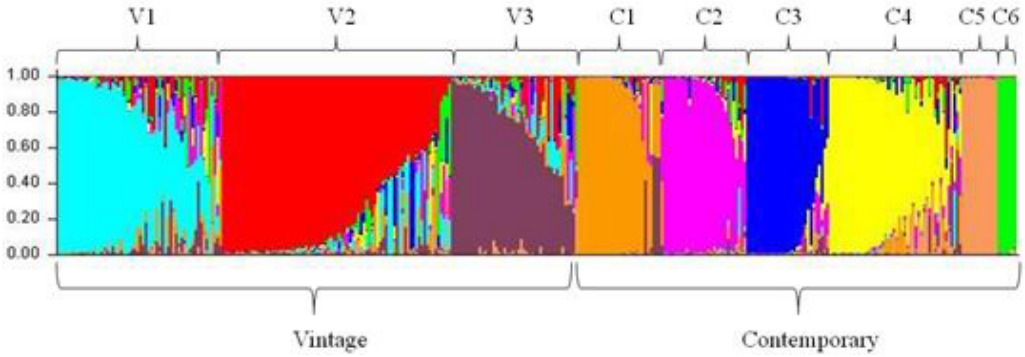

Figure 3. Classification of 348 inbred tomato lines into nine populations using the STRUCTURE program based on SNP molecular markers.

We also constructed a neighbor-joining tree based on shared alleles. The 348 tomato inbred lines were clustered into three major clusters (Figure 4). Two major clusters were consistent with the first and second vintage populations. The third major cluster contained almost all of the contemporary lines, and the lines of six populations were completely separated. However, some vintage varieties were clustered into the contemporary cluster. Contemporary inbred lines were divided into six clusters, which were supported by neighbor-joining clustering analysis. All of the inbred lines in the first cluster (C1) were pink tomato lines and the second cluster (C2) included all the green tomato lines, which possess the *rin* gene and have been used to improve tomato fruit firmness. The remaining four clusters (C3-C6) comprise 82 red tomato lines. The vintage inbred lines were separated into at least two clusters. The first cluster (V1) included 18 cultivated cherry tomatoes and 16 medium-fruited tomatoes. The second cluster (V2) contained 56 lines, with most of these tomatoes (82.1%) having originated from the Netherlands. The third cluster (V3) was probably an admixed

cluster; the neighbor-joining tree and the phenotypic characters of these lines also supported this hypothesis. There were three cherry tomatoes and 10 pink tomatoes within this cluster. The ancestry of the pink tomato could explain why 17 lines of this cluster were clustered closed to C1.

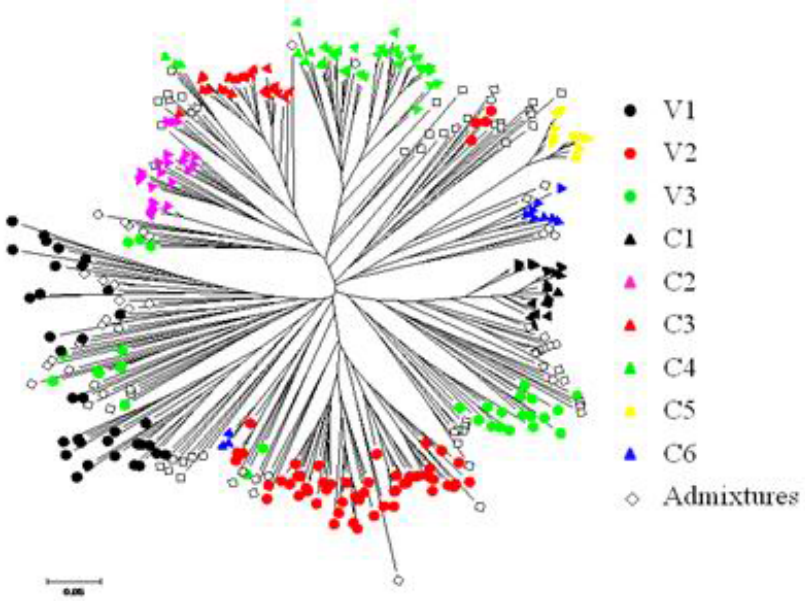

Figure 4. Neighbor-joining tree calculated for the 348 inbred tomato lines.

Assessment of genetic diversity

In order to differentiate between vintage and contemporary germplasm, AMOVA was performed between the two groups. The results showed that the difference among the two groups was significant ($P < 0.01$) and explained 17.4% of the total genetic variance. Pairwise F_{ST} values were calculated among the nine populations, Each population significantly differed from the others, and the differentiation between contemporary populations (0.424-0.867) was larger than that between vintage populations (0.120-0.313) (Table 1).

Table 1. Pairwise estimates of F_{ST} values for nine model-based populations.

Subpopulation	V1	V2	V3	C1	C2	C3	C4	C5
V2	0.313							
V3	0.208	0.120						
C1	0.472	0.603	0.382					
C2	0.475	0.460	0.321	0.625				
C3	0.508	0.608	0.405	0.678	0.498			
C4	0.466	0.488	0.329	0.548	0.424	0.483		
C5	0.546	0.563	0.415	0.867	0.687	0.798	0.590	
C6	0.337	0.417	0.234	0.646	0.459	0.610	0.428	0.638

All AMOVA-based F_{ST} estimates were significant ($P < 0.01$).

For all inbred tomato lines, the average H_E was 0.356 (Table 2), ranging from 0.028 (2325_361) to 0.497 (solcap_snp_sl_3355). Among the nine model-based subpopulations, V1 possessed the highest estimates of N_A, I, and H_E, because this subpopulation mainly contained cherry tomato. The six contemporary tomato subpopulations showed lower diversity than vintage subpopulations, which could be explained by the higher variation among populations. AMOVA for the six populations showed that 59.3% of the variation was found among populations.

Table 2. Summary of SNP diversity parameters in subpopulations.

Sample	N	N_A	I	H_O	H_E
Vintage	189	90	0.563	0.014	0.388
V1	34	85	0.484	0.011	0.328
V2	56	83	0.358	0.010	0.233
V3	29	84	0.450	0.012	0.302
Contemporary	159	87	0.475	0.014	0.324
C1	23	58	0.121	0.004	0.078
C2	25	72	0.263	0.031	0.171
C3	25	65	0.182	0.027	0.116
C4	36	73	0.308	0.001	0.205
C5	13	51	0.075	0.000	0.050
C6	11	69	0.291	0.024	0.197
Total	348	90	0.519	0.014	0.356

N: sample size; N_A: number of different alleles; H_O: observed heterozygosity; H_E: unbiased expected heterozygosity; I: Shannon's information index.

Core collection

Random and M strategy algorithms were used to generate a core tomato collection. The results showed that the M strategy was more efficient than random sampling (Figure 5). The redundancy curve of the M strategy peaked when only seven inbred lines were sampled. This indicated that 2% of lines were sufficient to represent all SNP variation among 348 inbred tomatoes.

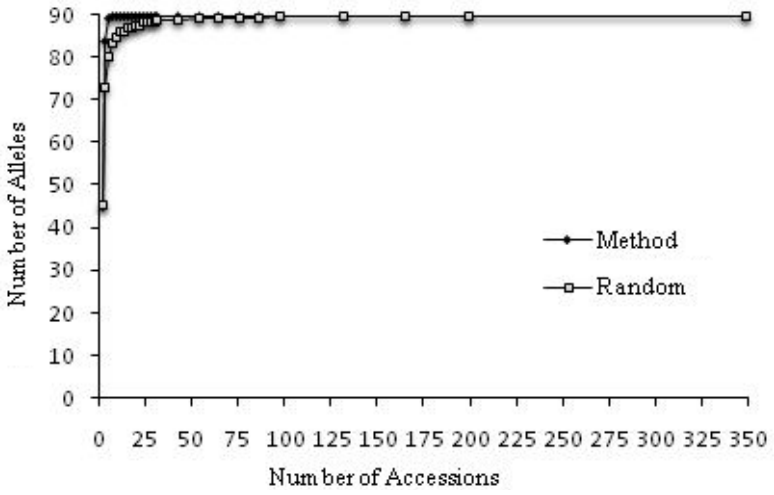

Figure 5. Redundancy curves obtained using the MSTRAT software.

DISCUSSION

SNPs are a predominant marker system used to investigate genetic variation, and a high density of SNPs is distributed across the whole tomato genome. The Tomato Genome Consortium (2012) estimated that 5.4 million SNPs exist between wild and domesticated genomes. Lin et al. (2014) discovered 11.6 million SNPs based on 360 accession sequences. Hamilton et al. (2012) identified 62,576 SNPs in cultivated tomato. Although different strategies were used, the number of SNPs identified in intra-species was much lower than that in inter-species. Compared to SSRs, SNP markers may be related to function. The functional annotation of SNPs can be obtained from public databases, and many SNP alleles are related to complex disorders (Wang et al., 2006). In the present study, 45 polymorphic SNP loci were used to successfully fingerprint tomato cultivars. Sequence information allowed us to visualize the genetic variability among vintage and contemporary tomato germplasm. In addition, these data can supplement other data obtained through genetic diversity analyses and association studies if the same materials were used.

Because SSRs are multi-allelic markers, they can provide a virtually unlimited number of alleles. Due to the high mutation rate, SSRs have a high number of alleles, most of which are found at low frequencies in the population (Lv et al., 2012). In contrast, SNPs are bi-allelic markers and the MAF was mostly in the 0.2-0.5 range (Hamblin et al., 2007). Of the 90 SNP alleles detected in the present study, 91.1% (82/90) were present in more than 20% of the collection sampled, with only two rare alleles identified (allele frequency <0.05). This finding is inherent to the characteristics of SNPs. Despite this, the average MAF (0.315) in this study was higher than that observed in other SNP-based studies (Corrado et al., 2014). Such bias could be attributable to the sampling of SNPs. All SNPs used here were selected according to the standard of higher polymorphism among cultivated tomato groups.

Cluster and structure analyses are used widely to study the genetic relationships between germplasm, and using a combination of both methods can effectively reveal genetic differentiation among complex populations (Li et al., 2010). In our study, vintage varieties were significantly different from contemporary varieties. Corrado et al. (2013, 2014) found the same results based on SNPs. Furthermore, the present findings confirmed the presence of subgroups within the vintage tomatoes, and the subpopulation structure was consistent with that observed in previous studies (Sim et al., 2011, 2012). Our findings also suggest that contemporary inbred lines can be divided into subpopulations that reflect specific market niches, and the subpopulations were found to be associated with tomato color (pink, green, and red). This is similar to the findings of past studies (Corrado et al., 2013), which also separated contemporary cultivars into three types (fresh, processing, and cherry). A key difference was that all contemporary varieties used in the present study were fresh-market tomatoes. The different colored subpopulations represent different market classes and breeding programs.

Genetic diversity for two predefined subpopulations and nine model-based subpopulations was measured using a number of different alleles, expected heterozygosity, and Shannon's information index. Significant differences were found among these cultivated tomatoes. Vintage varieties showed higher diversity than contemporary lines, which was not consistent with previously reports (Sim et al., 2012). This is mainly because cherry tomato is believed to be the ancestor of the cultivated tomato and possesses higher genetic variation than large-fruited types (Ranc et al., 2008). Our data also showed that most cherry tomatoes (51.7%) are classed as admixtures. This result was consistent with previous studies (Corrado et al., 2013).

The purpose of building a core collection is to provide a minimum number of resources and to represent the maximum genetic diversity of the species. At the DNA level, the number of alleles is an important indicator used to evaluate genetic diversity. Our data showed that all of the alleles were captured by only 2% of lines, which is lower than that reported in previous studies (4-25%) (Le Cunff et al., 2008; Ranc et al., 2008; Corrado et al., 2014; Song et al., 2014). This could be due to the lower number of rare alleles, as well as the higher expected heterozygosity, enabling all SNP alleles to be captured with a small number of individuals. Increasing the number of SNPs and constructing haplotypes may overcome this problem.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Science and Technology Supporting Project (#2014BAD01B0803) and the Natural Science Foundation of Liaoning Province of China (#2014027016).

REFERENCES

Archak S, Karihaloo JL and Jain A (2002). RAPD markers reveal narrowing genetic base of Indian tomato cultivars. *Curr. Sci.* 82: 1139-1143.

Bai Y and Lindhout P (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann. Bot. (Lond.)* 100: 1085-1094. http://dx.doi.org/10.1093/aob/mcm150

Carrari F and Fernie AR (2006). Metabolic regulation underlying tomato fruit development. *J. Exp. Bot.* 57: 1883-1897. http://dx.doi.org/10.1093/jxb/erj020

Casals J, Pascual L, Canizares J, Cebolla-Cornejo J, et al. (2011). The risks of success in quality vegetable markets: possible genetic erosion in Marmande tomatoes (*Solanum lycopersicum* L.) and consumer dissatisfaction. *Sci. Hortic. (Amsterdam)* 130: 78-84. http://dx.doi.org/10.1016/j.scienta.2011.06.013

Chang HW, Cheng YH, Chuang LY and Yang CH (2010). SNP-RFLPing 2: an updated and integrated PCR-RFLP tool for SNP genotyping. *BMC Bioinformatics* 11: 173. http://dx.doi.org/10.1186/1471-2105-11-173

Corrado G, Piffanelli P, Caramante M, Coppola M, et al. (2013). SNP genotyping reveals genetic diversity between cultivated landraces and contemporary varieties of tomato. *BMC Genomics* 14: 835. http://dx.doi.org/10.1186/1471-2164-14-835

Corrado G, Caramante M, Piffanelli P and Rao R (2014). Genetic diversity in Italian tomato landraces: Implications for the development of a core collection. *Sci. Hortic. (Amsterdam)* 168: 138-144. http://dx.doi.org/10.1016/j.scienta.2014.01.027

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x

FAO (2012). FAO Statistical Yearbook 2012. Food and Agriculture Organization (FAO) of the United Nations. http://faostat.fao.org/site/339/default.aspx. Accessed August 2013.

Garcia-Mas J, Oliver M, Gomez-Paniagua H and De Vicente M (2000). Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. *Theor. Appl. Genet.* 101: 860-864. http://dx.doi.org/10.1007/s001220051553

Gouesnard B, Bataillon TM, Decoux G, Rozale C, et al. (2001). MSTRAT: an algorithm for building germ plasm core collections by maximizing allelic or phenotypic richness. *J. Hered.* 92: 93-94. http://dx.doi.org/10.1093/jhered/92.1.93

Gupta PK, Rustgi S and Mir RR (2008). Array-based high-throughput DNA markers for crop improvement. *Heredity (Edinb)* 101: 5-18. http://dx.doi.org/10.1038/hdy.2008.35

Hamblin MT, Warburton ML and Buckler ES (2007). Empirical comparison of Simple Sequence Repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS One* 2: e1367. http://dx.doi.

org/10.1371/journal.pone.0001367
Hamilton JP, Sim SC, Stoffel K, Deynze AV, et al. (2012). Single nucleotide polymorphism discovery in cultivated tomato via sequencing by synthesis. *Plant Genome* 5: 17-29. http://dx.doi.org/10.3835/plantgenome2011.12.0033
Hawtin G, Iwanaga M and Hodgkin T (1996). Genetic resources in breeding for adaptation. *Euphytica* 92: 255-266. http://dx.doi.org/10.1007/BF00022853
Korkmaz M and Dogan NY (2015). Biogeographic pattern of genetic diversity detected by RAPD and ISSR analysis in *Gypsophila* (Caryophyllaceae) species from Turkey. *Genet. Mol. Res.* 14: 8829-8838. http://dx.doi.org/10.4238/2015.August.3.6
Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, et al. (2008). Construction of nested genetic core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. *sativa*. *BMC Plant Biol.* 8: 31. http://dx.doi.org/10.1186/1471-2229-8-31
Li X, Yan W, Agrama H, Hu B, et al. (2010). Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. *Genetica* 138: 1221-1230. http://dx.doi.org/10.1007/s10709-010-9521-5
Lin T, Zhu G, Zhang J, Xu X, et al. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46: 1220-1226. http://dx.doi.org/10.1038/ng.3117
Liu K and Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129. http://dx.doi.org/10.1093/bioinformatics/bti282
Lv J, Qi J, Shi Q, Shen D, et al. (2012). Genetic diversity and population structure of cucumber (*Cucumis sativus* L.). *PLoS One* 7: e46919. http://dx.doi.org/10.1371/journal.pone.0046919
Miller JC and Tanksley SD (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theor. Appl. Genet.* 80: 437-448. http://dx.doi.org/10.1007/BF00226743
Obiadalla-Ali H, Fernie AR, Kossmann J and Lloyd JR (2004). Developmental analysis of carbohydrate metabolism in tomato (*Lycopersicon esculentum* cv. Micro-Tom) fruits. *Physiol. Plant.* 120: 196-204. http://dx.doi.org/10.1111/j.0031-9317.2004.0167.x
Peakall R and Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* 28: 2537-2539. http://dx.doi.org/10.1093/bioinformatics/bts460
Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
Ranc N, Muños S, Santoni S and Causse M (2008). A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (Solanaceae). *BMC Plant Biol.* 8: 130. http://dx.doi.org/10.1186/1471-2229-8-130
Shirasawa K, Isobe S, Hirakawa H, Asamizu E, et al. (2010). SNP discovery and linkage map construction in cultivated tomato. *DNA Res.* 17: 381-391. http://dx.doi.org/10.1093/dnares/dsq024
Sim SC, Robbins MD, Van Deynze A, Michel AP, et al. (2011). Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.). *Heredity (Edinb)* 106: 927-935. http://dx.doi.org/10.1038/hdy.2010.139
Sim SC, Van Deynze A, Stoffel K, Douches DS, et al. (2012). High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS One* 7: e45520. http://dx.doi.org/10.1371/journal.pone.0045520
Song Y, Fan L, Chen H, Zhang M, et al. (2014). Identifying genetic diversity and a preliminary core collection of Pyrus pyrifolia cultivars by a genome-wide set of SSR markers. *Sci. Hortic. (Amsterdam)* 167: 5-16. http://dx.doi.org/10.1016/j.scienta.2013.12.005
Stewart Jr CN and Via LE (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* 14: 748-750.
Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729. http://dx.doi.org/10.1093/molbev/mst197
Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635-641.
van Berloo R, Zhu A, Ursem R, Verbakel H, et al. (2008). Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. *Theor. Appl. Genet.* 117: 89-101. http://dx.doi.org/10.1007/s00122-008-0755-x
Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, et al. (2008). Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am. J. Bot.* 95: 1240-1253. http://dx.doi.org/10.3732/ajb.0800097
Wang P, Dai M, Xuan W, McEachin RC, et al. (2006). SNP Function Portal: a web database for exploring the function implication of SNP alleles. *Bioinformatics* 22: e523-e529. http://dx.doi.org/10.1093/bioinformatics/btl241

Yang J, Wang Y, Shen H and Yang W (2014). In silico identification and experimental validation of insertion-deletion polymorphisms in tomato genome. *DNA Res.* 21: 429-438. http://dx.doi.org/10.1093/dnares/dsu008

Zhang N, Brewer MT and van der Knaap E (2012). Fine mapping of fw3.2 controlling fruit weight in tomato. *Theor. Appl. Genet.* 125: 273-284. http://dx.doi.org/10.1007/s00122-012-1832-8

Zhou R, Wu Z, Cao X and Jiang FL (2015a). Genetic diversity of cultivated and wild tomatoes revealed by morphological traits and SSR markers. *Genet. Mol. Res.* 14: 13868-13879. http://dx.doi.org/10.4238/2015.October.29.7

Zhou R, Wu Z, Jiang FL and Liang M (2015b). Comparison of gSSR and EST-SSR markers for analyzing genetic variability among tomato cultivars (*Solanum lycopersicum* L.). *Genet. Mol. Res.* 14: 13184-13194. http://dx.doi.org/10.4238/2015.October.26.14

Supplementary material

Table S1. Description of 348 tomato accessions used in this study.

Table S2. Description of 52 molecular markers used in this study.