Genetic diversity and population structure of core watermelon (*Citrullus lanatus*) genotypes using DArTseq based SNPs

Xingping Yang[a,d,*], Runsheng Ren[a,b,d], Rumiana Ray[b], Jinhua Xu[a], Pingfang Li[a], Man Zhang[a], Guang Liu[a], Xiefeng Yao[a], Andrzej Kilian[c]

[a] Institute of Vegetable Crops/Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Jiangsu Academy of Agricultural Sciences, 50 Zhongling Street, Nanjing 210014, China

[b] Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, UK

[c] Diversity Arrays Technology, University of Canberra, Kirinari St., Bruce, ACT 2600, Australia

[d] These authors contributed equally to this work

[*]Corresponding author. Tel.: +86-25-84392262; Fax: +86-25-84390262. should be addressed; E-mail address: xingping@jaas.ac.cn (X.P. Yang)

Abstract

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. &Nakai var. *lanatus*] is an economically important vegetable belonging to the Cucurbitaceae family. Genotypes that exhibit agronomically important traits are selected for the development of elite cultivars. Understanding the genetic diversity and the genotype population structure based on molecular markers at genome level can speed up the utilization of diverse genetic resources for varietal improvement. In the present study, we carried out an analysis of genetic diversity based on 3882 SNP markers across 37core watermelon genotypes including the most widely used watermelon varieties and wild watermelon. Based upon the SNP genotyping data of the 37 watermelon genotypes screened, gene diversity and polymorphism information content values across chromosomes varied between 0.03-0.5 and 0.02-0.38, with averages of 0.14 and 0.13, respectively. The two wild watermelon genotypes were distinct from cultivated varieties and the remaining thirty-five cultivated genotypes were differentiated into three major clusters. Twenty genotypes were grouped in cluster I. Eleven genotypes were grouped in cluster II. Three advanced breeding lines of yellow fruit flesh and genotype SW043 were grouped in cluster III.The results from Neighbor-Joining (NJ) dendrogram, principal coordinate analysis and STRUCTURE analysis approaches were consistent and the grouping of genotypes was generally in agreement with their origins. Here we reveal the genetic relationships among the core watermelon genotypes maintained at Jiangsu Academy of Agricultural Sciences, China. The molecular and phenotypic characterization of the existing core watermelon genotypes, together with specific agronomic characteristics can be utilized by researchers and breeders for future watermelon improvement.

Keywords：Watermelon; Genetic diversity; Molecular marker; SNPs; Core genotype

Introduction

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. &Nakai var. *lanatus*, 2n = 2x = 22] is an economically important vegetable crop that is grown on approximately 7% of the world's cultivated land area with an annual production of about 109 million tons (FAO Statistics 2013, http://faostat.fao.org/). China is the leading watermelon producing country accounting for 71% of the world's production. More recently, in China, the economic and nutraceutical importance of watermelon has increased thus challenging breeders to develop new varieties combining high fruit quality with enhanced resistance to biotic and abiotic stresses.

A number of watermelon genotypes and cultivars possess good morphological and horticultural traits, and some of them have been selected as core breeding materials and used extensively as parental lines in watermelon breeding programs. A major challenge for breeders is to be able to accurately estimate the relationship between parents prior to initiating hybridization. This is particularly important since cultivated watermelon has been shown to exhibit narrow genetic diversity (Levi *et al*. 2001a; 2001b), resulting in progenies displaying undesirable agronomic characteristics obtained from the crosses. Knowledge of the genetic relationships and characterization of watermelon diversity including core genotypes that are being used as parents in the current watermelon breeding programs can facilitate efficient management and improved utilization of available genotype resources.

Generally, genetic variation in plants can be characterized using morphological and molecular methods. However, the use of molecular markers has the advantage of improved reliability and repeatability (Powell *et al*., 1996). Molecular markers previously used for characterization of genetic diversity in watermelon genotypes include random amplified polymorphic DNA (RAPD) (Mujaju *et al*., 2010), amplified fragment length polymorphism (AFLP) (Che *et al*., 2003),simple sequence repeat (SSR) (Kwon *et al*., 2010) and expressed sequence tags-SSRs (EST-SSR) (Mujaju *et al*., 2013). However, these marker systems utilize

limited molecular markers and are primarily gel-based, costly and time consuming. In contrast, single nucleotide polymorphism (SNP) markers, offer substantial advantages over PCR-based methods such as large-scale genotyping, the generation of high abundance sequence information and whole genome coverage to allow genetic studies (Gupta *et al*., 2008; Singh *et al*., 2013). The method has been widely applied to various crop species for genetic variability analysis, such as rice (Singh *et al*., 2013), maize (Van Inghelandt *et al*., 2010), wheat (Nielsen *et al*., 2014) and melon (Esteras *et al*., 2013).With the falling costs of DNA sequencing and availability of whole genome sequence of watermelon (Guo *et al*., 2013), a new SNP-based marker platform known as DArTseq was developed recently by combination Diversity Array Technology (DArT) marker system with Illumina short read sequencing method (Sansaloni *et al*., 2011).

Here we used DArTseq-based SNP markers on whole genome level to analyze the genetic diversity of core watermelon genotypes that have been widely and commonly used in watermelon breeding programs of China. Our results will facilitate further exploitation of these genotypes by researchers and breeders for watermelon improvement.

Materials and Methods

Plant materials and DNA extraction

In the present study, a set of 37 watermelon core genotypes including elite watermelon cultivars, inbred lines and wild watermelon genotypes were evaluated. Single seed of each watermelon genotype was first pre-germinated and then transferred into pots (15 cm x 15 cm) filled with compost mix of sand and peat1:1 (v:v) Levington M1 compost (Monro Group) for growing.At four-leaf stage, one or two leaf pieces for each plant were collected for DNA extraction. The leaf samples were first placed in plastic bags with tiny holes and then vacuum-dried in SuperModulyo freeze dryer (Thermo Savant, USA) for 7 days, and stored before DNA extraction. Seedlings were grown under controlled growth room conditions with

a 16/8 h light/dark cycle and temperature of 28°C day and 20°C night.The main characteristics obtained from trials carried out over the last two years (2011 and 2012) at the watermelon breeding Research Station, Jiangsu Academy of Agricultural Sciences (JAAS), are given in Table S1.

Freeze-dried leaf samples (0.1g) of were ground completely in a 2 mL Eppendorf tube with added garnet sand (0.15mm/0.7 mm) using FastPrep®-24 Instrument (MP Biomedicals Inc., UK). DNA extraction was performed using modified cetyltrimethylammonium bromide (CTAB) method (Ren *et al*. 2012). The quality and quantity of DNA samples were determined by agarose gel analysis and DNA concentration was adjusted to 50-100 ng μL^{-1}.

DArTseq based SNP analysis

Based on the preliminary tests for appropriate enzyme combinations in cucurbits, restriction enzymes combination PstI-MseI was chosen for the digestion of a mixture of DNA samples. After digestion, DNA samples are processed in digestion/ligation reactions principally as per Kilian *et al*. (2012) but replacing a single PstI-compatible adaptor with two different adaptors corresponding to two different Restriction Enzyme (RE) overhangs. The PstI-compatible adapter was designed to include Illumina flowcell attachment sequence, sequencing primer sequence and "staggered", varying length barcode region, similar to the sequence reported by Elshire *et al*, (2011). Reverse adapter contained flowcell attachment region and MseI-compatible overhang sequence. Only "mixed fragments" (PstI-HpaII) are effectively amplified in 30 rounds of PCR using the following reaction conditions: 1 min at 94℃for initial denaturation; 30 cycles each consisting of 20 sec at 94℃ for denaturation, 30 sec at 58 ℃ for annealing, 45 sec at 72 ℃ for extension; and finally a 7 min extension step at 72 ℃. After PCR equimolar amounts of amplification products from each sample of the 96-well microtiter plate are bulked and applied to c-Bot (Illumina) bridge PCR followed by sequencing on Illumina Hiseq2500. The sequencing (single read) was run for 77 cycles.

Sequences generated from each lane are processed using proprietary DArT analytical pipelines and the corresponding fastq files are first processed to filter away poor quality sequences, applying more stringent selection criteria to the barcode region compared to the rest of the sequence. Accordingly, the assignments of the sequences to specific samples carried in the "barcode split" step are very reliable. Approximately 2,500,000 (+/- 7%) sequences per barcode/sample are used in marker calling. Finally, identical sequences are collapsed into "fastqcall files". These files are used in the secondary pipeline for DArT PL's proprietary SNP and Silico DArT (presence/absence of restriction fragments in representation) calling algorithms (DArTsoft14).

Data Analysis

Marker attributes for each marker locus were calculated using software PowerMarker Version 3.25 (Liu and Muse, 2005), including major allele frequency, gene diversity, polymorphic information content, gene diversity and heterozygosity. Based on genetic similarity, a dendrogram was constructed by application of the unweighted pair group method with Neighbor-Joining (NJ) cluster analysis using software Darwin Version 5.0 (Perrier *et al*., 2003). It was also used to perform principal coordinate analysis (PCoA) to visualize the genetic relationships among individual watermelon genotypes.

Software program STRUCTURE v2.3 (Pritchard *et al*., 2000) was used to infer population structure. Estimation of the best K value (the number of clusters) was performed by evaluating K=1 to 8 with the admixture and correlated allele frequency models. Five independent runs were done for each K. Each run consisted of a burn-in period of 10,000 iterations followed by 10,000 Monte Carlo Markov Chain (MCMC) iterations. Both LnP(D) in STRUCTURE output and its derived ΔK method (Evanno *et al*., 2005) were used to determine the K value.

Results

Characterization of SNP markers

A total of 4808 polymorphic SNP markers that have a scoring reproducibility of 99.7% and a call rate of 98.8% were identified by genotyping of 37 watermelon genotypes using the DArTseq platform. In order to make subsequent analysis more reliable, SNP markers with a missing proportion of > 5% were excluded and a total of 3882 SNP markers were used for further analysis. The physical position along the chromosome of SNP markers were determined based on results of the alignment to the reference watermelon genome of 97103 (minimum base identity > 90% and E-value < 10^{-5}, Guo *et al*., 2013). The number of SNP markers was not evenly distributed across the 11 chromosomes and it ranged from 204 for chromosome 04 to 478 for chromosome 05, with an average number of 352 (Fig.S1).

Based upon the SNP genotyping data of the 37 watermelon genotypes screened, gene diversity and polymorphism information content values across chromosomes ranged from 0.03 to 0.5 and from 0.02 to 0.38, respectively (Fig.S2).The mean gene diversity and PIC values were 0.14 and 0.13, respectively and their distributions showed that 91% (for gene diversity) and 93% (for PIC value) of the markers were both in the range from 0.05 to 0.2 (Fig.S2).

Genetic diversity analysis

A Neighbor-joining (N-J) dendrogram of 37 watermelon genotypes based on the Jaccard's similarity matrix data obtained with the 3882 SNP markers was shown in Fig.1. The N-J dendrogram broadly separated 2 wild genotypes from the 35 cultivated genotypes (Fig.1a). In order to visualize the relationships among the 35 cultivated genotypes more clearly, a new dendrogram was generated by excluding the 2 wild genotypes and the results showed that the SNP markers were able to detect a high variability among the 35 watermelon cultivars (Fig 1b). In this dendrogram, the 35 cultivars were classified into 3 clusters. Cluster I contained 11 genotypes in total and all of the 11 genotypes were either elite cultivars from US or breeding

lines with known US racial background. Cluster II consisted of 20genotypes, of which 2 were from US, 1 were from Korea, 2 were from Australia, 2 were from Japan, 1 was from Taiwan and 12 were from China (3 were from Xinjiang, 6 were from Jiangsu, 3 were from Beijing). The third main group was formed by three Jiangsu-derived breeding lines, and one breeding line from Taiwan was also included in this cluster.

Principal Coordinate analysis

The Jaccard's similarity matrix generated from marker scores was also used for Principal Coordinate Analysis (PCoA) to visualize the genetic relationships between the watermelon genotypes and the results showed that the 37 genotypes were clearly classified into two groups. Group I included only 2 wild genotypes and Group II only cultivated genotypes (Fig.S3a).The 35 cultivated genotypes assigned to Group II were further arranged in three separated clusters (Fig.S3b), which was generally consistent with the Neighbor-Joining analysis above. The first and the second principal axis explain 25.9% and 11.2% of the variation, respectively.

Population Structure Analysis

For all the 37 genotypes, the sharp division of LnP(D) score and the peak value of delta K score were both obtained at K = 2 (Fig.2a, 2b and Fig. 4a).This means that the 37 genotypes were partitioned into two groups, which corresponded to the two wild genotypes and the 35 cultivated genotypes. When the 35 genotypes were examined, the LnP(D) score increased continuously with the increase of K from 1 to 8, but the most apparent change appeared when K increased from 2 to 3 (Fig. 3a). Delta K also peaked at a K-value of three (Fig.3b).These results mean that the 35 genotypes should be divided into three populations. The results obtained from the STRUCTURE analysis are in good agreement with those obtained from Neighbor-Joining dendrogram and the principal coordinate analysis. Each genotype is represented by a vertical column and different colors represent different subpopulations. The

proportion of a given genotype's color bar represents the proportion that genotype belongs to the corresponding subpopulation (Fig. 4b).

Discussion

Polymorphism of SNP markers

A total of 3882 SNP markers were used in the present study to provide detailed molecular characterization of core watermelon genotypes. The number distribution of markers on the 11 chromosomes varied greatly and their distribution was also reflective of chromosome size in watermelon (Guo *et al*., 2013).

Polymorphic information content (PIC) refers to the usefulness of a marker for detecting polymorphism. Due to the bi-allelic nature of SNPs, PIC values range between 0 to 0.5, which are lower than PIC values for multi-allelic markers, such as SSR, AFLP and RAPD that can range between 0.5 to 1.0. Although with a low average PIC value (0.13), the SNP markers in the present study have greater abundance and co-dominant inheritance pattern increasing their effectiveness in discriminating the genotypes compared to RAPDs, SSRs and AFLPs markers used in previous watermelon genetic diversity studies.

Genetic diversity of wild and cultivated germplasm

Selection of genetically diverse parents with high genetic variability is a key step in hybridization programs. In the present study, the genetic relationship based on three different approaches (NJ dendrogram, principal coordinate analysis and population structure analysis) gave similar results. The 37 watermelon genotypes in the present study can be classified into two differentiated clusters: wild genotypes and cultivated genotypes. Similar distinct clustering pattern of wild and cultivated watermelon genotypes has also been reported by Hwang *et al*. (2011) who clustered the 32 watermelon genotypes into two major clusters based on AFLP and EST-SSR markers. Cluster I included all adapted watermelon cultivars and Cluster II included the four wild-type species (PI 189225, PI 386024, PI 494817 and PI

632755). These wild-type PIs are known to exhibit high levels of resistance against various diseases, for example, PI 189225, PI 632755 and PI 386024 have been reported to be resistant to powdery mildew (*Podsphaera xanthii* race 2W) (Tetteh et al., 2010), whilst PI 494817 is moderately resistant to bacterial fruit blotch (*Acidovorax citrulli*), a significant threat to watermelon around the world (Hopkins and Thompson, 2002). Using molecular markers-based introgression, breeders can develop new cultivated varieties with superior disease resistance from the wild-type germplasm described above. Gichimu *et al*. (2009) also reported high morphological diversity between unimproved accessions (wild accession and landrace) and commercial cultivars.

The narrow genetic base of cultivated watermelon (Levi et al., 2001a; Levi et al., 2001b) poses a challenge for watermelon breeding programs combining specific quality characteristics, such as high fruit and sugar yield and pest and disease resistance. Thus, the exploitation of the wild watermelon genotypes as genetic source to improve resistance and tolerance to biotic and abiotic stress is essential for the development of new varieties (Thies and Levi, 2007). Indeed, the wild watermelon genotype PI 189225 that is included in our studies is a known source of resistance to various diseases such as powdery mildew (Tetteh *et al*., 2010, 2013), anthracnose (Boyhan *et al*., 1994), gummy stem blight (Gusmini *et al*., 2005). 'G10' is another wild watermelon genotype that is characterized by its good resistance to Fusarium wilt (unpublished data). The wide differences observed between the wild and cultivated genotypes can provide valuable information for the utilization of wild watermelon in improving disease resistance by interspecific backcross between the wild germplasm and cultivated breeding lines (Levi et al., 2010). The identification and the use of the molecular markers linked to resistance genes in the wild-type germplasm will speed up the introgression of desirable traits into new varieties.

Cluster I of cultivated germplasm

Cluster analysis revealed the presence of genetic diversity according to origin and evidence of relationships between genotypes from different origins. For example, Cluster I was constituted mainly by cultivars from US and Cluster II was constituted mainly from East-Asia or Austria.

Cultivars with moderate fruit characteristics but good resistance to Fusarium wilt, such as ‘All Sweet’, ‘Charleston Gray’, ‘Smokylee’, ‘Sugarlee’ and ‘Crimson Sweet’ were differentiated in Cluster I. In addition to the good fruit quality, they can all be used to derive lines with resistance to Fusarium wilt. ‘LW022’and ‘LW023’were two advanced breeding lines and were bred and selected from US cultivars. ‘9 Jiu’ has good resistance to Fusarium wilt and should make it an excellent parent in hybrid crosses, especially for producing unprotected cultivars. ‘SSD’ was a large icebox variety with an attractive striped rind pattern, good internal red pigmentation, and higher soluble sugar content than other icebox varieties. SSD could be crossed with a small parent to produce an icebox-sized fruit or with a large-fruited breeding line to produce a standard-size fruit (Cralll et al. 1994). ‘AU-GSC’ and ‘AU-RS’, both have multiple-disease resistance (resistance to anthracnose, gummy stem blight, and Fusarium wilt) and they can be used as resistant materials for developing or enhancing disease resistance in watermelon cultivars.

Cluster II of cultivated germplasm

Among the cultivars in cluster II, ‘P1-3’and ‘P3-1’ are two inbred lines with dark skin, good quality and early maturity, and they can be used to synthesize hybrids with small fruits under the protected green house condition. The two inbred lines ‘SW055-1’ and ‘SW057’, which had high soluble sugar content, can be used for development of hybrids with good fruit quality. The inbred line ‘P4’ had considerable yield advantage over other watermelon plants; thus, it should be useful in breeding programs aiming to develop good fruit quality, resistance and high fruit yield cultivars. ‘Red flesh 8424’ has a solid dark green rind and red flesh and

was used as the maternal parent to produce the most popular commercial cultivar ‘8424’ in East China. ‘K3’ is elongate, with sweet, flavorful red flesh and thick rind. Many breeders have made use of ‘K3’ as sources of good fruit quality for the development of new hybrids. In previous studies, we selected some breeding lines, including ‘MW022’, ‘J2’, ‘P4’, ‘Hong 5-2’, ‘SW055-1’, ‘MW097’, ‘MW099’, ‘MW095’, ‘MW096’, ‘Furong F8’, ‘MW026’, ‘MW089’, with excellent fruit quality. However, they were much less resistant to diseases. Thus, these breeding lines with different fruit and plant characteristics were selected in order to synthesize hybrids with a range of fruit and plant patterns to attend to the farmers´ and consumers´ preferences. Particularly, cultivar ‘Sugarbaby’ and ‘Calhoun Gray’ were apart from other US cultivars and were grouped into Cluster II. This could be ascribed to the more frequent use of these two elite cultivars for the watermelon breeding programs resulting in close genetic relationships with the cultivars from East Asia.

Cluster III of cultivated germplasm

Interestingly, three advanced breeding lines (‘R-1-3’, ‘R-1-2’and ‘R-2-1-2’) from Jiangsu province and one advanced breeding line (‘SW043’) from Taiwan were located in an independent cluster (Cluster III), suggesting a unique genetic background to other cultivated genotypes. ‘R-1-3’, ‘R-1-2’, ‘R-2-1-2’and ‘SW043’were the four advanced breeding lines with the same yellow color of fruit flesh. The color of fruit flesh has been shown to be an important indicator of genetic relationships among watermelon types and most wild watermelon types have white, light green or yellow flesh while most cultivated have red flesh (Wang *et al*., 2011).Thus, this genetic clustering reflected that the color of fruit flesh may be one of the main causes for this clustering, but other influencing causes for this clustering need to be investigated further.

In conclusion, this study using DArTseq based SNP markers revealed the relationships and genetic diversity among 37 core watermelon genotypes. The new

information will be useful to breeders to maximize the parental diversity for new crosses within breeding programs and development of varieties with improved resistance to abiotic and biotic stress.

Conclusions

In this study, we used 3882 SNP markers to evaluate the genetic diversity and population structure of watermelon genotypes and our results showed that apart from distinct grouping of wild genotypes from cultivated watermelon genotypes, there are also three main groups in the 35 cultivated genotypes. The grouping of genotypes based on the large number of SNP markers will also be useful in providing a theoretical foundation for effectively utilizing these genotypes in future watermelon breeding programs.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

This research was supported by grants from the Ministry of Science and Technology (2012BAD02B03-14), the Jiangsu Provincial Major Support Program for Agricultural (BE2012323), the China Postdoctoral Science Foundation (2013M541624) and the Jiangsu Province Postdoctoral Science Foundation (1301068B). The authors would also like to acknowledge the Jiangsu Academy of Agriculture Sciences, China and The University of Nottingham, UK for funding of this research.

References

Boyhan G, Norton J, Abrahams B and Wen H (1994) A new source of resistance to anthracnose (Race 2) in watermelon. *HortScience* 29: 111-112.

Che KP, Liang CY, Wang YG, Jin DM, Wang B, Xu Y, Kang GB and Zhang HY (2003) Genetic assessment of watermelon genotype using the AFLP technique. *HortScience* 38: 81-84.

Cralll JM, Elmstrom GW, and McCuistion FT (1994)SSDL: A High-quality IceboxWatermelon Breeding Line Resistantto Fusarium Wilt and Anthracnose. HortScience 29(6):707-708.

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

Esteras C, Formisano G, Roig C, Díaz A, Blanca J, Garcia-Mas J, Gómez-Guillamón ML, López-Sesé AI, Lázaro A and Monforte AJ (2013) SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium. *Theoretical and Applied Genetics* 126: 1285-1303.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology* 14: 2611-2620.

Gichimu BM, Owuor BO, Mwai GN and Dida MM (2009) Morphological characterization of some wild and cultivated watermelon (*citrullus* sp.) accessions in Kenya. *Journal of Agricultural and Biological Science* 4: 10-18.

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, Min J, Guo X, Murat F, Ham BK, Zhang Z, Gao S, Huang M, Xu Y, Zhong S, Bombarely A, Mueller LA, Zhao H, He H, Zhang Y, Zhang Z, Huang S, Tan T, Pang E, Lin K, Hu

Q, Kuang H, Ni P, Wang B, Liu J, Kou Q, Hou W, Zou X, Jiang J, Gong G, Klee K, Schoof H, Huang Y, Hu X, Dong S, Liang D, Wang J, Wu K, Xia Y, Zhao X, Zheng Z, Xing M, Liang X, Huang B, Lv T, Wang J, Yin Y, Yi H, Li R,Wu M, Levi A, Zhang X, Giovannoni JJ, Wang J, Li Y, Fei Z, Xu Y (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature genetics* 45: 51-58.

Gupta P, Rustgi S and Mir R (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101: 5-18.

Gusmini G, Song R and Wehner TC (2005) New sources of resistance to gummy stem blight in watermelon. *Crop science* 45: 582-588.

Hopkins DL, Thompson CM (2002). Evaluation of *Citrullus* sp. germplasm for resistance to *Acidovorax avenae* subsp. *citrulli*. Plant Disease 86: 61-64.

Hwang J, Jumsoon K, Byeonggu S, Kwanghwan K and Younghoon P (2011) Genetic diversity in watermelon cultivars and related species based on AFLPs and EST-SSRs. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 39: 285-292.

Jones E, Sullivan H, Bhattramakki D and Smith J (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). *Theoretical and Applied Genetics* 115: 361-371.

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P and Uszynski G (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. *Data Production and Analysis in Population Genomics* 888: 67-89.

Kwon YS, Oh YH, Yi SI, Kim HY, An JM, Yang SG, Ok SH and Shin JS (2010) Informative SSR markers for commercial variety discrimination in watermelon (*Citrullus lanatus*). *Genes & Genomics* 32: 115-122.

Levi A, Thomas CE, Keinath AP and Wehner TC (2001a) Genetic diversity among watermelon (*Citrullus lanatus* and *Citrullus colocynthis*) accessions. *Genetic Resources and Crop Evolution* 48: 559-566.

Levi A, Thomas CE, Wehner TC and Zhang X (2001b) Low genetic diversity indicates the need to broaden the genetic base of cultivated watermelon. *HortScience* 36: 1096-1101.

Levi A, Wechter WP, Harris KR, Davis AR and Fei Z (2010) High-frequency oligonucleotides in watermelon expressed sequenced tag-unigenes are useful in producing polymorphic polymerase chain reaction markers among watermelon genotypes. *Journal of the American Society for Horticultural Science* 135: 369-378.

Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129.

Mujaju C, Sehic J and Nybom H (2013) Assessment of EST-SSR Markers for Evaluating Genetic Diversity in Watermelon Accessions from Zimbabwe. *American Journal of Plant Sciences* 4: 1448.

Mujaju C, Werlemark G, Garkava-Gustavsson L and Nybom H (2010) High Levels of RAPD and SSR Marker Diversity in Landraces of Watermelon (*Citrullus lanatus*) in Southern Africa. *Acta Horticulturae* 918: 291-296.

Nielsen NH, Backes G, Stougaard J, Andersen SU and Jahoor A (2014) Genetic Diversity and Population Structure Analysis of European Hexaploid Bread Wheat (*Triticum aestivum* L.) Varieties. *PloS one* 9: e94000.

Perrier X, Flori A and Bonnot F (2003) Methods of data analysis. *Genetic diversity of cultivated tropical plants.* 33-63.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for genotype analysis. *Molecular breeding* 2: 225-238.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D and Kilian A (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*, *BMC Proceedings*. 5: P54.

Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi R, Singh N and Singh R (2013) Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. *PloS one* 8: e84136.

Tetteh AY, Wehner TC and Davis AR (2010) Identifying resistance to powdery mildew race 2W in the USDA-ARS watermelon genotype collection. *Crop science* 50: 933-939.

Tetteh AY, Wehner TC and Davis AR (2013) Inheritance of Resistance to Powdery Mildew Race 2 in *Citrullus lanatus* var. *lanatus*. *HortScience* 48: 1227-1230.

Thies JA and Levi A (2007) Characterization of watermelon (*Citrullus lanatus* var. *citroides*) genotype for resistance to root-knot nematodes. *HortScience* 42: 1530-1533.

Van Inghelandt D, Melchinger AE, Lebreton C and Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics* 120: 1289-1299.

Wang YH, Behera TK and Kole C (2011) Genetics, genomics and breeding of cucurbits. CRC Press.

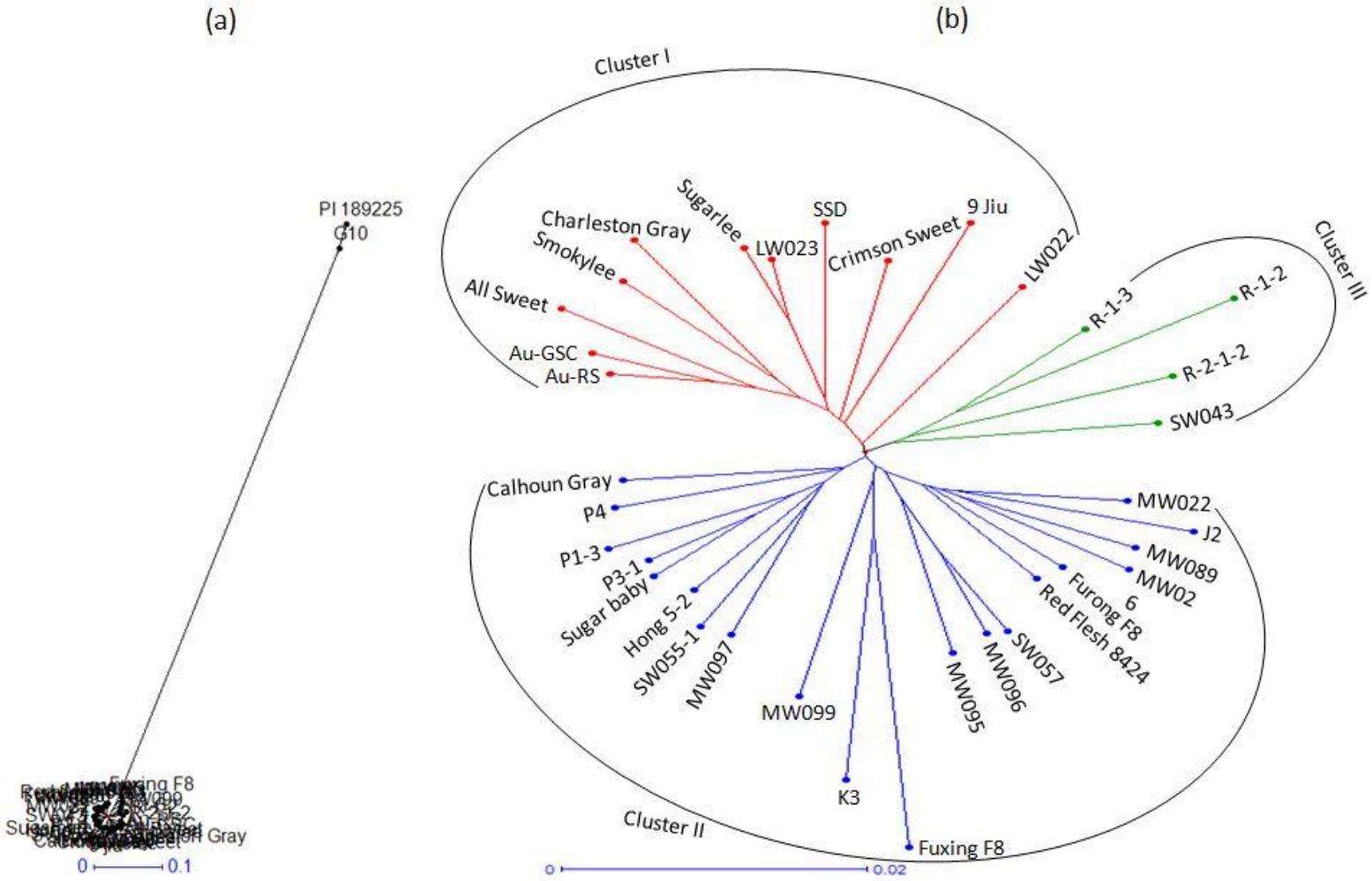

Fig. 1. Neighbor-Joining dendrogram showing the genetic relationships among 37 (a) and 35 (b) watermelon cultivars based on 3882 single nucleotide polymorphism (SNP) markers.

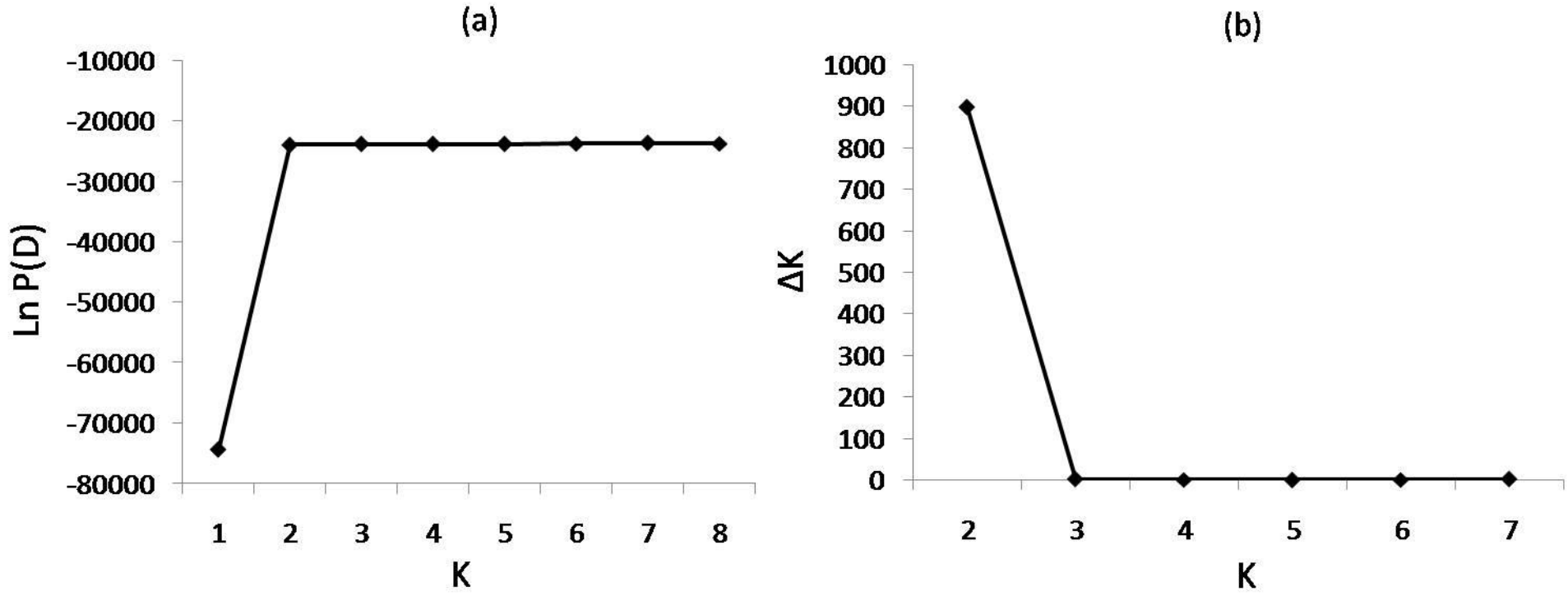

Fig. 2. Determination of the optimal value of K, based on five independent runs and K ranging from 1 to 8 based on 37 watermelon genotypes. (a) Evolution of the natural logarithm probability (LnP(D)); and (b) its derived statistics ΔK for each K value.

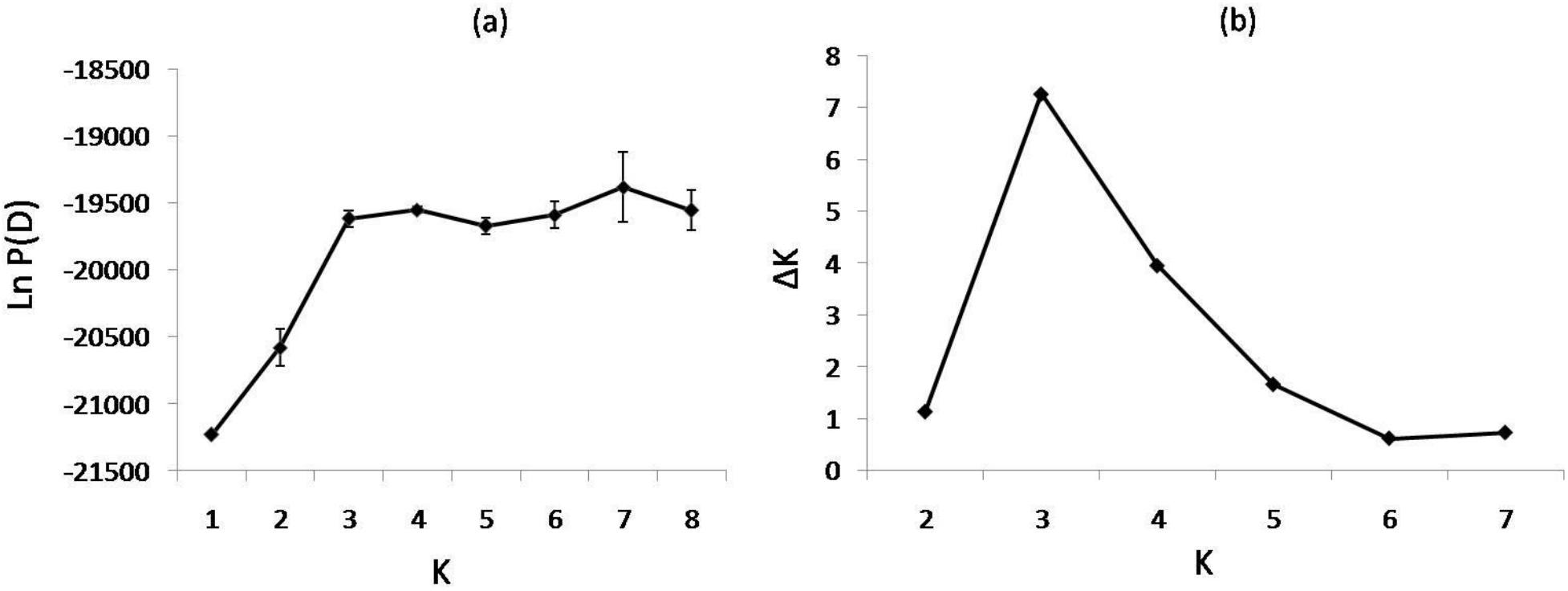

Fig. 3. Determination of the optimal value of K, based on five independent runs and K ranging from 1 to 8 based on 35 watermelon genotypes. (a) Evolution of the natural logarithm probability (LnP(D)); and (b) its derived statistics ΔK for each K value.

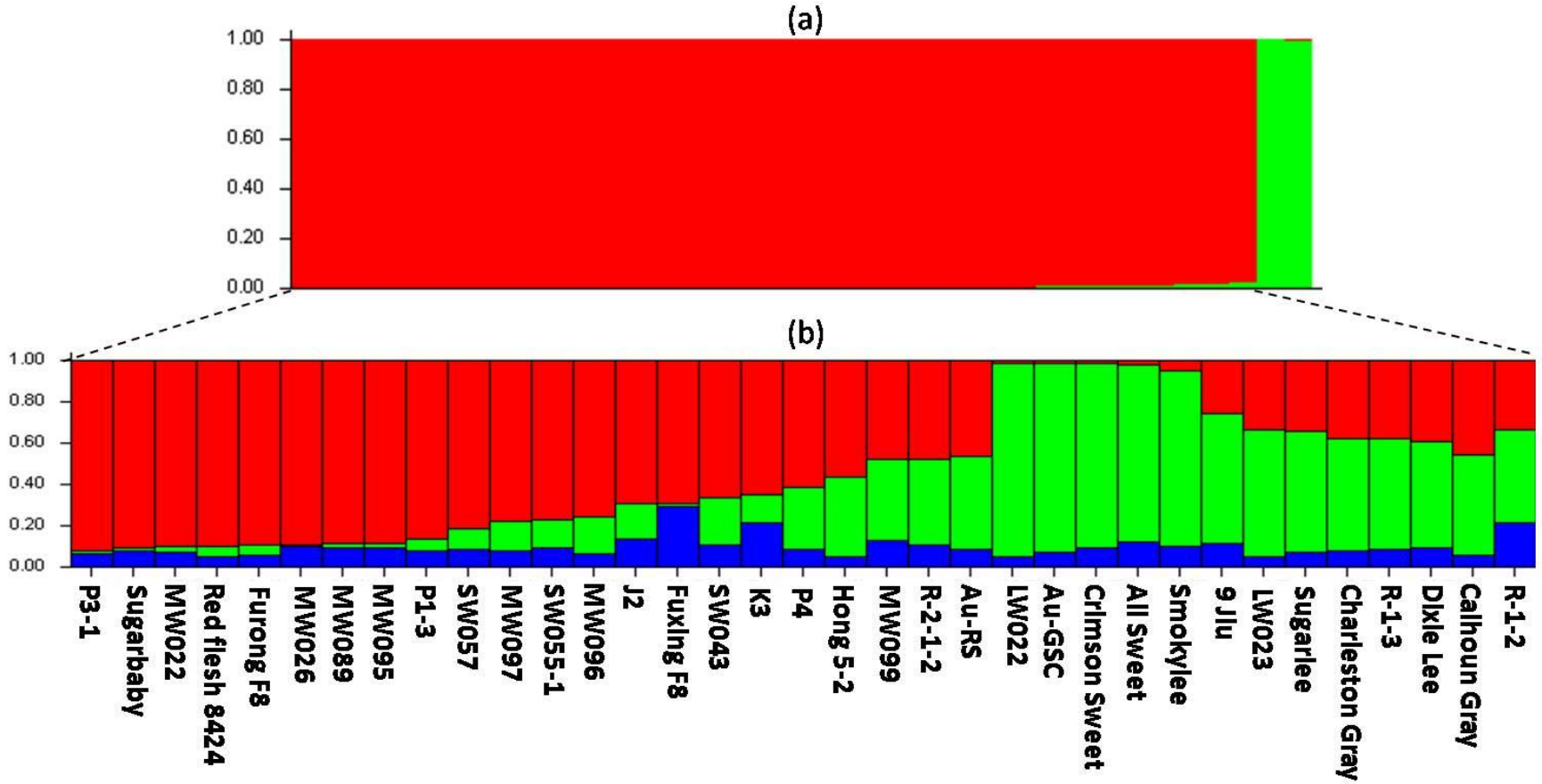

Fig. 4. Structure structure of the genotypes based on 3882 single nucleotide polymorphism (SNP) markers. (a) Optimal population structure (K=2) for the 37 watermelon genotypes and (b) Optimal population structure (K=3) for the 35 watermelon genotypes. Each individual is represented by a narrow vertical bar , which is partitioned into coloured segments in proportion to the estimated membership to the 3 populations.

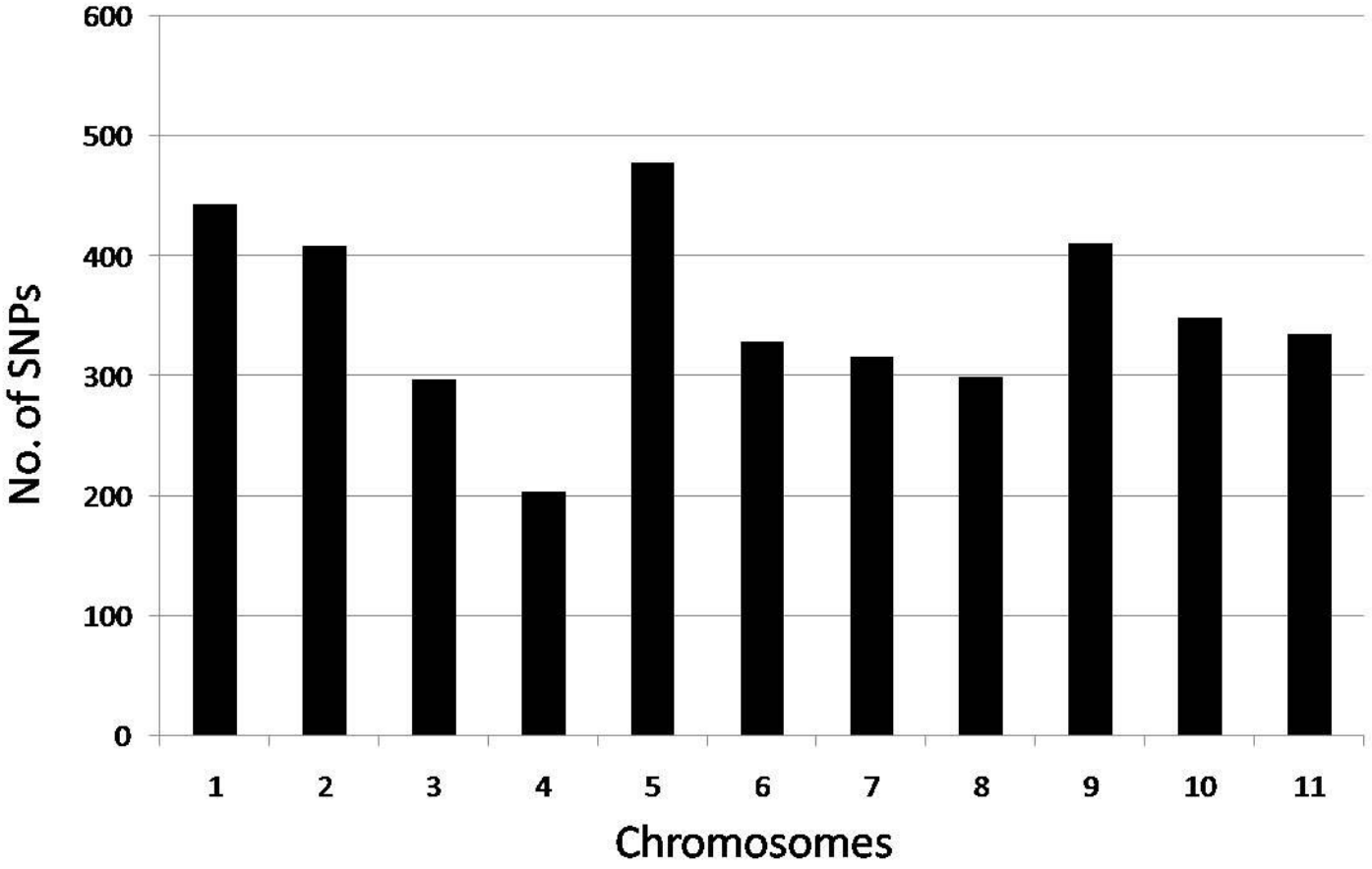

Fig. S1. Distribution of DArTseq based SNP markers on different watermelon chromosomes. The x axis represents the number of each watermelon chromosome. The y axis is the number of SNP markers and the number of SNPs on each chromosome was shown by the height of the bars.

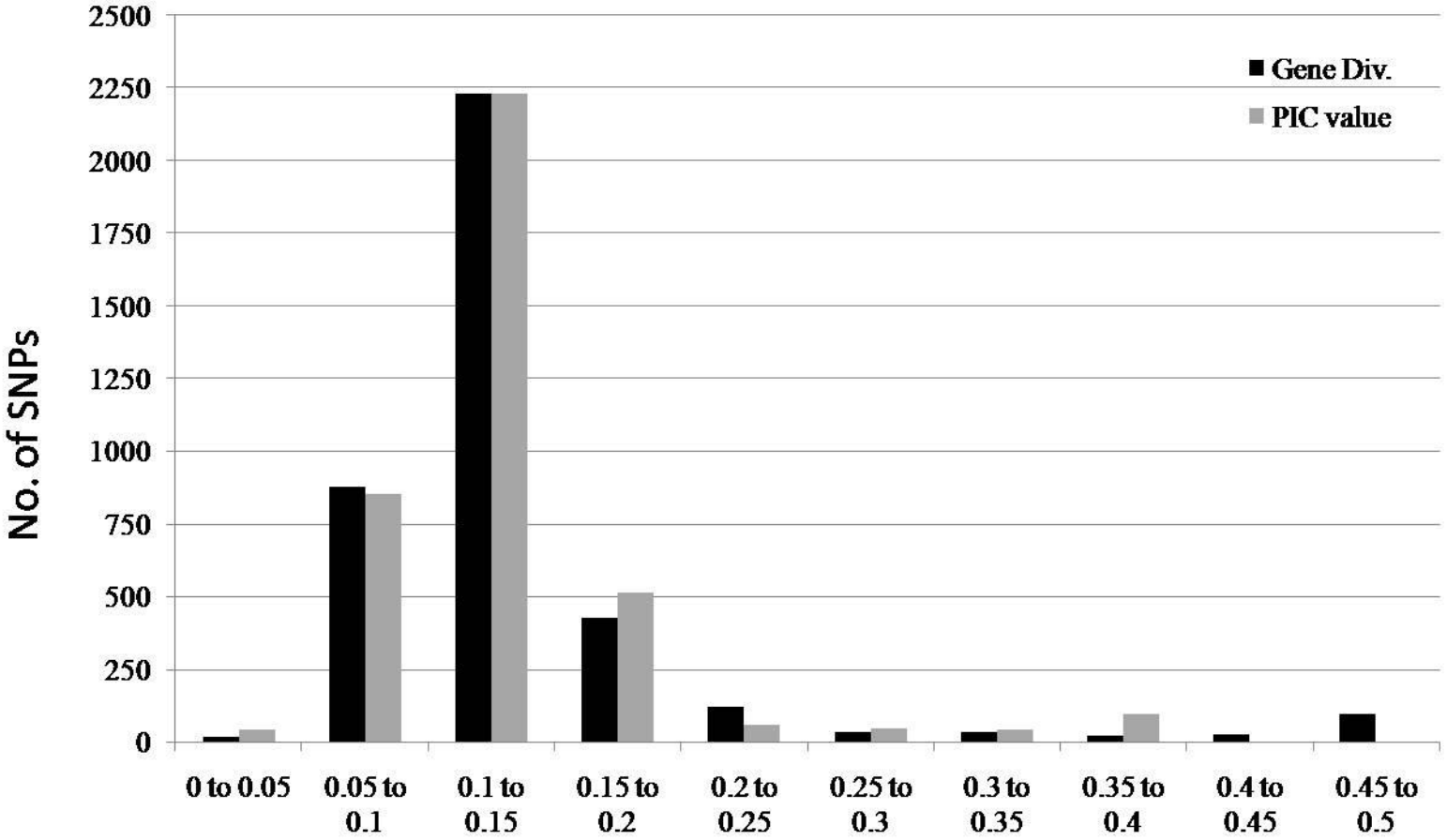

Fig. S2. Distribution of gene diversity and polymorphism information content values for 3882 single nucleotide polymorphism (SNP) markers used in the study.

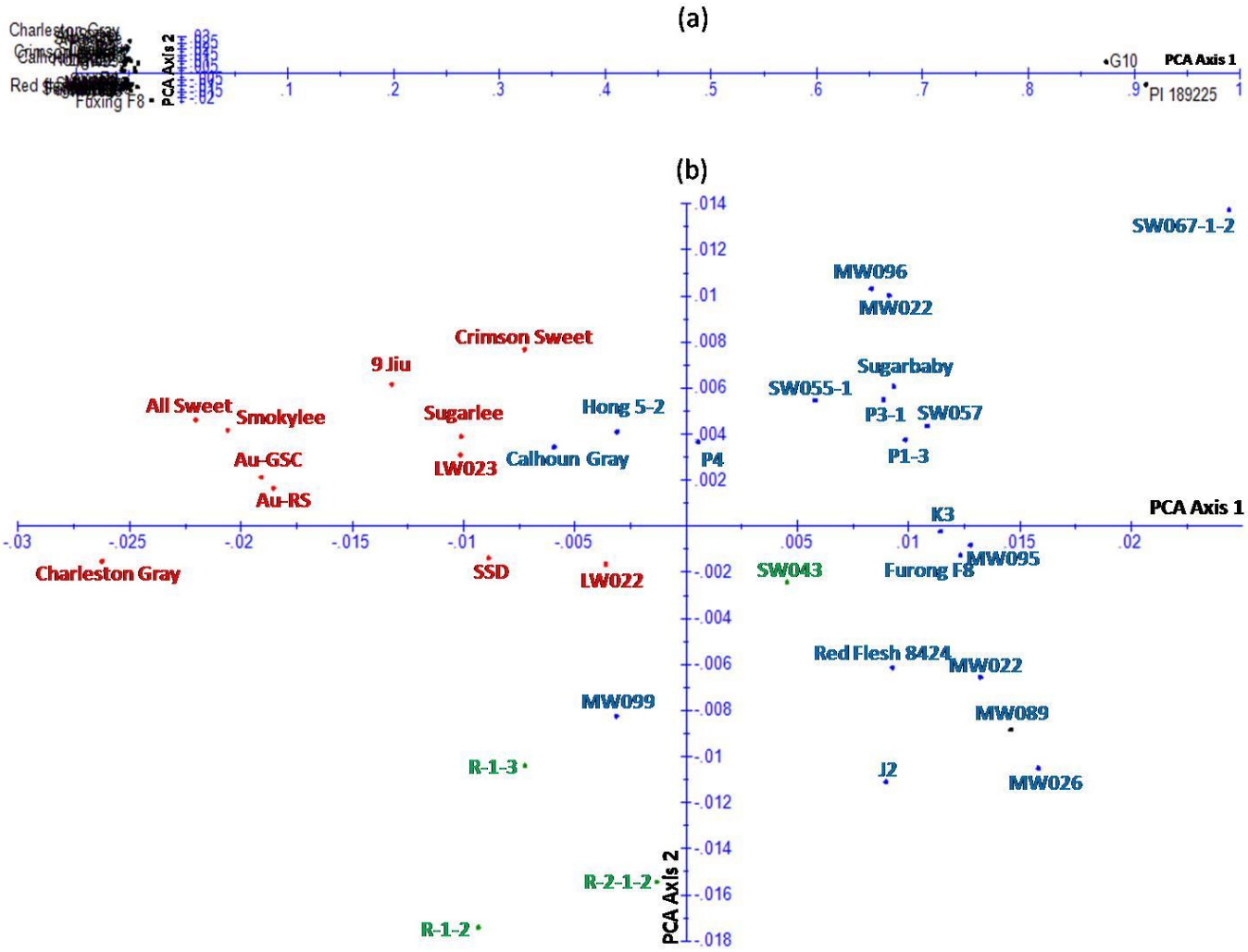

Fig. S3. Principal coordinate analysis of 37 (a) and 35 (b) watermelon genotypes based on Dice's distance calculated from 3882 single nucleotide polymorphism (SNP) markers.

Table S1. Details of breeding history, major characteristic traits and origin of the genotypes evaluated in the diversity study.

No.	Genotype name	Character	Fruit characteristics										Source
			weight (kg)	Rind color[a]	Stripe type[b]	Shape	Fruit Colour	Width(cm)	Length(cm)	Rind thickness(cm)	SSCC[c]	SSCE[d]	
1	Hong 5-2	Inbred line	2.60	GR	NS	Round	Red	16.9	16.4	0.5	10.5	8.7	Taiwan
2	SW043	Inbred line	1.72	GR	NS	Round	Yellow	15.0	14.5	0.5	11.9	9.6	Taiwan
3	Red flesh 8424	Inbred line	3.64	LR	No stripe	Round	Red	18.7	17.1	1.0	10.8	8.7	Xinjiang
4	MW095	Inbred line	2.80	GR	SS	Round	Red	18.2	16.8	0.9	9.6	8.4	Xinjiang
5	MW099	Inbred line	3.80	GR	WS	Round	Red	19.8	18.5	1.2	10.4	9.5	Xinjiang
6	G10	Wild type	2.15	GR	No stripe	Round	White	16.7	15.5	1.5	2.2	2.0	Xinjiang
7	P4	Inbred line	4.82	DG	No stripe	Round	Red	21.0	20.7	1.3	10.9	7.3	Jiangsu
8	K3	Variety	2.08	GR	NS	Round	Pink	16.3	15.9	0.6	10.5	8.4	Jiangsu
9	J2	Inbred line	2.90	GR	NS	Round	Pink	18.9	18.1	0.4	10.0	8.8	Jiangsu
10	P1-3	Inbred line	3.04	DG	No stripe	Elongated	Red	22.8	15.6	0.9	11.7	10.2	Jiangsu
11	R-1-2	Inbred line	1.40	GR	No stripe	Elongated	Yellow	22.1	10.9	0.5	8.8	7.6	Jiangsu
12	R-1-3	Inbred line	1.60	GR	No stripe	Round	Yellow	15.0	14.0	0.6	9.5	8.8	Jiangsu
13	R-2-1-2	Inbred line	0.89	GR	NS	Round	Yellow	12.4	11.5	0.5	9.0	8.7	Jiangsu
14	MW026	Inbred line	5.00	GR	NS	Round	White	21.9	20.4	1.0	8.4	7.1	Jiangsu
15	P3-1	Inbred line	4.32	DG	No stripe	Round	Red	22.5	20.4	1.1	11.1	8.7	Jiangsu
16	MW022	Inbred line	3.50	GR	NS	Round	Red	19.3	18.6	0.9	11.0	10.0	Beijing
17	MW096	Inbred line	2.30	DG	No stripe	Round	Red	16.7	15.9	0.8	10.3	9.0	Beijing
18	MW097	Inbred line	3.70	GR	NS	Round	Red	19.0	18.6	1.0	11.0	8.6	Beijing
19	Crimson sweet	Variety	3.00	YG	NS	Round	Pink	17.3	17.1	1.0	10.6	9.0	USA
20	LW023	Inbred line	5.20	GR	NS	Round	Red	22.5	20.9	1.1	11.1	9.0	USA
21	9 jiu	Variety	3.10	LR	No stripe	Elongated	Red	24.8	14.5	0.8	9.0	6.7	USA
22	LW022	Inbred line	5.46	LR	No stripe	Elongated	Red	33.7	16.2	1.3	10.8	9.5	USA
23	Sugarbaby	Variety	2.10	Black	No stripe	Round	Red	15.6	16.0	1.1	9.6	7.9	USA
24	SSD	variety	2.50	LR	NS	Round	Red	17.7	16.6	0.6	10.0	7.6	USA
25	Charleston	Variety	3.10	LR	No stripe	Elongated	Pink	23.8	16.1	1.2	9.0	6.9	USA

	Gray												
26	Calhoun Gray	Variety	5.00	LR	No stripe	Elongated	Pink	30.7	19.0	1.1	9.0	8.0	USA
27	Smokylee	Variety	8.10	DG	No stripe	Elongated	Red	34.6	20.4	1.7	11.1	8.6	USA
28	Au-GSC	Variety	2.34	GR	NS	Round	Red	15.0	15.6	0.7	9.6	6.2	USA
29	All Sweet	Variety	2.34	GR	NS	Elongated	Red	21.8	14.7	0.7	8.2	5.9	USA
30	Au-RS	Variety	2.60	GR	NS	Round	Red	16.4	16.7	0.6	7.2	5.7	USA
31	Sugarlee	Variety	5.2	GR	NS	Round	Red	22.5	20.9	1.1	11.1	9.0	USA
32	PI 189225	Wild type	0.58	GR	No stripe	Round	White	9.2	10.2	1.1	4.0	2.0	Zaire
33	MW089	Inbred line	3.50	GR	NS	Round	Red	19.1	18.3	0.9	10.5	9.0	Japan
34	FurongF8	Inbred line	2.70	GR	NS	Round	Red	17.4	16.9	0.6	10.3	9.0	Japan
35	Fuxing F8	Inbred line	1.80	GR	NS	Elongated	Red	22.3	12.3	0.5	11.2	9.6	Korea
36	SW055-1	Inbred line	2.55	LR	No stripe	Elongated	Red	23.0	14.2	0.8	10.2	9.6	Australia
37	SW057	Inbred line	1.90	GR	NS	Elongated	Red	21.4	13.0	0.6	11.0	9.0	Australia

[a] GR, Green; LR, Light green; YG, Yellow green; DG, Dark green

[b] NF, Narrow stripes; SS, Straight stripe; WS, Wide stripes

[c] SSCC, soluble sugars content in center position (%)

[d] SSCE, soluble sugar content in edge position (%)