Impact of fruit shape selection on genetic structure and diversity uncovered from genome-wide perfect SNPs genotyping in eggplant

Weilun Liu · Zongwei Qian · Jian Zhang · Jingjing Yang · Mingsheng Wu · Lorenzo Barchi · Haiyan Zhao · Honghe Sun · Yanling Cui · Changlong Wen (ID)

Abstract The recently constructed eggplant genome provides an opportunity for genome-wide marker exploration to be used for variety identification and genetic analysis. SNPs with high density across the genome and genetic stability are suitable for DNA fingerprint and genetic analysis. In this study, we used the eggplant reference genome of line 63/7 and resequencing data of 45 unique lines to select 219 genome-wide perfect SNPs in the eggplant. The high-throughput SNP genotyping technique target SNP-seq was used to successfully validate the 219 SNPs across 377 eggplant varieties to establish the unique DNA fingerprint for each variety. In addition, we chose 36 SNPs as a core SNP set with the ability to differentiate 95% of 377 eggplant varieties. Model-based structure and principal component analysis suggested three major population groups and indicated a significant impact of long-term selection for fruit shape in eggplant varieties. Further genetic diversity analysis within the three population groups showed a considerably narrow genetic background in round- and oval-fruited eggplants. This could be a result of limited choices of varieties used for breeding of round- and oval-fruited eggplants in Asia, indicated by its low observed heterozygosity (H_o) value (0.152) and high inbreeding coefficient (0.442). Finally, a genome-wide association study based on the 219 SNPs identified five associated SNPs located near the *SUN* and *OVATE* homologs, which had conserved function in controlling the fruit shape. This study signals a risk of genetic erosion in the round- and oval-fruited eggplants and provides valuable information for future variety management and breeding programs.

Weilun Liu, Zongwei Qian and Jian Zhang contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11032-019-1051-y) contains supplementary material, which is available to authorized users.

W. Liu · Z. Qian · J. Zhang · J. Yang · H. Sun · Y. Cui · C. Wen
Beijing Vegetable Research Center (BVRC), National Engineering Research Center for Vegetables, Beijing Academy of Agricultural and Forestry Sciences, Beijing 100097, China

W. Liu · Z. Qian · J. Zhang · J. Yang · H. Sun · Y. Cui · C. Wen
Beijing Key Laboratory of Vegetable Germplasms Improvement, Beijing 100097, China

W. Liu · Z. Qian · J. Zhang · J. Yang · H. Sun · Y. Cui · C. Wen (✉)
Key Laboratory of Biology and Genetics Improvement of Horticultural Crops (North China), Ministry of Agriculture and Rural affairs, Beijing 100097, China
e-mail: wenchanglong@nercv.org

M. Wu · H. Zhao
Beijing Seed Management Station, Beijing 100088, China

L. Barchi
DISAFA, Plant Genetics and Breeding, University of Turin, 10095 Grugliasco, Italy

Introduction

The eggplant (*Solanum melongena* L.), also known as aubergine or brinjal, belongs to the *Solanaceae* family. It is considered one of the most important cultivated *Solanaceous* crops with a global production of 52.3 million tons and cultivation of 1.8 million ha (FAO Statistics 2017, http://www.fao.org/). It is cultivated and consumed worldwide, of which China produces about half of global production, ranking it the largest eggplant producer (32.8 million tons, 785,000 ha), followed by India, Egypt, and Turkey (FAO Statistics 2017, http://www.fao.org/). Eggplant is also regarded as one of the healthiest vegetables due to its high amounts of vitamins, minerals, and bioactive compounds such as chlorogenic acid and anthocyanins with antioxidant properties (Cericola et al. 2013; Plazas et al. 2013).

Eggplant is an Old World species and it was hypothesized to be domesticated from the wild species *Solanum insanum* (Ranil et al. 2017; Taher et al. 2017). Eggplant cultivation was documented in China 2000 years ago, and continual selection and breeding resulted in eggplant varieties with different fruit shapes and colors (Hurtado et al. 2012). However, like other domesticated crops, the domestication outside of its wild ancestor natural habitat and anthropogenic selection resulted in a drastic reduction in genetic diversity (Acquadro et al. 2017; Flint-Garcia 2013). The focus on limited varieties as genetic materials in breeding would cause a danger of genetic erosion (Munoz-Falcon et al. 2009; Cericola et al. 2013). Despite several genetic research studies reporting eggplant genetic diversity (Cericola et al. 2013; Hurtado et al. 2012; Naegele et al. 2014), they focused mostly on eggplant germplasms rather than on cultivated varieties and hybrids, and a limited number of DNA markers were tested. Therefore, comprehensive studies assessing the genetic variations among the commercial eggplant hybrids are lacking, which could help to better understand their genetic basis and relationships.

In past decades, there has been a continuing increase in the cultivation of eggplants, and hundreds of commercial varieties have been released into the seed market. To protect the economic interest of eggplant breeders, seed producers, and eggplant producers, rapid variety identification and authentication methods are becoming necessary. The traditional variety identification method based on the morphological characteristics through field inspection was time consuming and easily being influenced by environmental conditions. Genotyping using molecular markers has been proven to be a reliable alternative method that could be more accurate and accessible (Munoz-Falcon et al. 2009; Gao et al. 2012; Jamali et al. 2019). Among the molecular markers, SNP as a third-generation marker system with high density across the genome and high genetic stability has been widely used for genetic background selection and marker-assisted selection breeding, as well as for map-based clone and QTL mapping (Cheng et al. 2016; Wu et al. 2018a). It has also been successfully applied and used to establish DNA fingerprint in several crops, such as rice and maize, but has not been much studied in eggplant cultivars (Shirasawa et al. 2004; Tian et al. 2015).

Although a large number of SNPs have been discovered in eggplant based on RAD sequencing (Barchi et al. 2011), RNA sequencing (Gramazio et al. 2016), genotyping-by-sequencing (Acquadro et al. 2017), and whole genome resequencing (Barchi et al. 2019a). Compared with the other major *Solanaceae* crops like tomato and potato (Kim et al. 2014; Hamilton et al. 2011; Vos et al. 2015), genome-wide SNPs for eggplant have been relatively unexplored. The draft genome of eggplant SME_r2.5.1 was published (Hirakawa et al. 2014), and a well-assembled genome of the eggplant line 63/7 anchored at chromosome level (Barchi et al. 2019b) was constructed. This provides an opportunity to perform resequencing in diverse eggplants to explore genetic variations and genome wide SNPs. For example, a recent study used resequencing data from eight *S. melongena* and one *S. incanum* accessions to identify SNPs for genotyping 422 eggplant accessions with Single Primer Enrichment Technology (SPET) (Barchi et al. 2019a). Studies have shown that exploring on a wider gene pool is necessary to avoid polymorphism bias and is more suitable for developing markers for use in a much wider range of germplasm and varieties (Wang et al. 2014; Vos et al. 2015). The use of variomes from large resequencing data has been demonstrated to be useful and efficient for selecting suitable SSR markers, named perfect SSRs for genotyping in cucumber (Yang et al. 2019). The perfect SSRs were selected based on their polymorphism, stable motif, and flanking sequence conservation in the cucumber variome. In this study, we utilized the eggplant variome based on the resequencing data of 45 eggplant lines and selected 219 genome-wide perfect SNPs that are informative and have conserved flanking sequences. A newly established high-throughput SNP genotyping technique,

target SNP-seq, was successfully applied to genotype 377 eggplant varieties. Further genetic analysis of the 377 eggplant varieties provided valuable insight into the effect of human selection on population structure and genetic diversity. A genome-wide association study identified several SNPs associated with fruit shape. This study provides valuable information for eggplant variety management and for marker-assisted breeding in the future.

Material and methods

Plant material and DNA isolation

A total of 45 eggplant lines from China, South Asia, Japan, and Europe were resequenced, and 377 commercial eggplant varieties with variations in fruit shapes and color provided by Beijing Seed Management Station were applied to establish the DNA fingerprint in this study (Table S1). Total DNA were extracted from fresh true leaves following a CTAB-based method (Fulton et al. 1995).

Genome-wide SNP discovery

Resequencing data for 45 eggplant lines were used for genome-wide SNP discovery (Genome Sequence Archive at http://bigd.big.ac.cn/, accession number CRA001645). All of high-quality reads were mapped to the reference genome (Barchi et al. 2019b) using BWA, and the mapped reads were filtered to remove PCR duplication. SNP discovered from the 45 eggplant lines was conducted with GATK with stringent filtering criteria. Detected SNP with the following criteria: (1) minor allele frequency (MAF) > 0.4; (2) miss rate < 0.2; (3) heterozygosity < 0.2; (4) no sequence variation in the flanking region of 100 bp were selected as candidate perfect SNPs (Yang et al. 2019). Then a minimal number of SNPs representing the total genetic diversity of the 45 eggplant lines across the genome was selected (Henning et al. 2015; Yang et al. 2019) for subsequent genotyping by target SNP-seq.

SNP genotyping by target-seq

We designed a multiplex PCR panel with primers targeting sequence flanking the 150-bp regions containing each of the selected perfect SNPs. The target SNP-seq library construction procedure was based on the target SSR-seq protocol (Yang et al. 2019). It included a multiplex PCR reaction followed by another PCR reaction to add a universal adaptor and a specific barcode to the amplified DNA fragments from each sample. The purified PCR product from each sample was then combined and the target SNP-seq library was ready to be sequenced on the Illumina Hiseq X ten platform (Molbreeding Biotechnology Company, Shijiazhuang, China).

The raw sequencing reads were demultiplexed and assigned back to sample identities based on the specific barcodes via the Illumina bcl2fastq pipeline (Illumina, San Diego, CA, USA). The reads for each sample were then mapped to eggplant reference genome of eggplant line 67/3 (Barchi et al. 2019b) using BWA with default parameters (Li and Durbin 2009). Specific SNP variant bases were identified and SNP genotype data were extracted using the GATK software (McKenna et al. 2010). For each SNP locus, the allele with the highest number of reads was considered as major allele and the allele with the second highest number of reads as minor allele. A SNP locus with a read frequency of the major allele greater than 0.8 was considered as homozygous and a SNP locus with a read frequency of major and minor alleles both larger than 0.3 was considered as a heterozygous genotype (Yang et al. 2019).

Population structure analysis of eggplant varieties

The SNP genotype data were analyzed by using the model-based program STRUCTURE (Pritchard et al. 2000). Ten independent runs for K value (number of population groups assigned) ranging from 1 to 10 were performed using an admixture model with a burn-in period of 10,000 steps followed by 100,000 Monte Carlo Markov Chain simulations. The most likely K value was determined by considering ΔK (Evanno et al. 2005), the second order rate change of (estimated log probability of data) LnP (D) with respect to K implemented in STRUCTURE HARVESTER (Earl and von Holdt 2012). All varieties were assigned into corresponding groups based on their proportional membership probability (Q). Principal component analysis (PCA) and principal co-ordinate analysis (PCoA), and an unrooted neighbor-joining tree with Nei's standard genetic distance were performed using the ape and poppr packages in R software (Nei 1978; Kamvar et al. 2014).

Marker polymorphism and population differentiation analysis in eggplant varieties

The minor allele frequency (MAF), gene diversity (GD), observed heterozygosity (H_o), and polymorphic information content (PIC) were calculated using a perl script (Yang et al. 2019). To understand the population differentiation, an analysis of molecular variance (AMOVA) between and within groups was conducted using the poppr R package (Kamvar et al. 2014). Also, the estimation of pairwise F statistics (F_{st}) among groups was performed using the hierfstat R package.

Core SNPs set development for variety identification

The Perl method developed in Yang's study was used to select a core set of SNPs representing the total genetic diversity and with high discerning power (Yang et al. 2019). The saturation curve of the discernibility was plotted by pairwise comparison of varieties genotypes.

Core eggplant varieties selection

A pairwise comparison matrix by calculating the numbers of differential SNPs between each variety was built within each population (Yang et al. 2019). Fewer differential SNPs indicated closer kinship with others. The top 20% varieties with close kinship were considered as core varieties in each group.

GWAS analysis

The GWAS analysis was performed with Tassel v5.2.25 (Bradbury et al. 2007) using the mixed linear model (MLM, PCA + K-model), taking into account both kinship and structure (Yu et al. 2006). Three types of eggplant fruit shape: round, oval, and long were scored from 1 to 3, respectively. Association with an adjusted p value less than 0.005 was declared significant. For markers that were significantly associated with a trait, a general linear model with all fixed-effect terms was used to estimate R^2, the amount of phenotypic variation explained by each marker.

Results

Genome-wide perfect SNPs discovery based on resequencing data of 45 eggplant lines

Based on the resequencing data of 45 eggplant lines, the total reads were mapped to the newly assembled reference genome of eggplant (Barchi et al. 2019b) to discover genome-wide polymorphic SNPs. A total of 26,029,890 SNP loci were obtained, of which 93,582 SNPs that had a MAF more than 0.4, and a missing rate and heterozygosity less than 0.2 were selected as candidate SNP loci for genetic analysis. Analyzing the 100-bp sequence flanking the 93,582 SNPs from the resequencing data, 1925 SNPs with no sequence variations in flanking sequences were chosen as perfect SNPs. A modified minimal marker protocol (Henning et al. 2015; Yang et al. 2019) was applied to select the minimal numbers of SNPs to represent the genetic diversity in the resequencing data of the 45 eggplant lines. Overall, 251 SNPs were selected, of which 219 successfully passed the multiplex PCR panel design for subsequent target SNP-seq genotyping. The 219 SNPs were distributed from chromosome one to 12 (Table S3) and the phylogenetic tree of the 45 eggplant lines using the 219 SNPs showed classification and relationships largely consistent with that determined using the total genome-wide SNPs (Fig. S1).

Genotyping analysis of eggplant varieties using target SNP-seq

The selected 219 perfect SNPs were successfully genotyped in the 377 eggplant varieties using the target SNP-seq method. The average read depth per SNPs in the 377 varieties was 1041 and 58% of the samples were sequenced at a depth greater than 1000× (Fig. S2A). Three hundred and seventy-two out the 377 varieties (98.7%) exhibited more than 98% alignment rate (Fig. S2B). Of these aligned reads, all varieties aligned to the target SNP region at a rate greater than 95%, with the average target region alignment rate of 97.4% (Fig. S2C). Only three SNP markers showed missing data with the highest missing rate being just 0.53%. This demonstrated that the selected 219 SNPs were highly conserved and well genotyped. In addition, the target SNP-seq uniformity index (Fig. S2D) was analyzed to calculate the proportion of the coverage above 10% of mean depth value for each variety (Nishio et al. 2015) and to

infer the level of accuracy. Almost all the varieties (99.5%) had a uniform index higher than 98% indicating a high level of accuracy for the target SNP-seq.

Using these 219 SNPs, a unique genetic fingerprint was established for each of the 377 eggplant varieties. The PIC values of 219 SNPs genotype in eggplant varieties ranged from 0.078 to 0.375 with an average of 0.313, of which 68.64% of all SNPs exhibited PIC value higher than 0.3 (Fig. 1a). There were 119 SNPs displaying MAF value higher than 0.3 (54.3%) with mean value of 0.311 (Fig. 1b). The observed heterozygosity (H_o) displayed an average value of 0.224 with 57.9% of all SNPs above 0.2 (Fig. 1c). Moreover, the mean genetic diversity (GD) value of 219 SNPs was 0.398 ranging from 0.0813 to 0.5 for individual markers (Fig. 1d). This result indicated that the 219 perfect SNPs of eggplant are informative with good discriminating capacity, and suitable for variety identification and genetic diversity analysis.

Population structure analysis in eggplant varieties

The population structure of 377 eggplant varieties were analyzed based on the 219 perfect SNPs in the eggplant genome. The 377 varieties include major contemporary eggplant varieties originated in China, as well as imported from other countries like the Netherlands and France. The model-based structure analysis showed that the best K value was $K = 2$ (Fig. 2a). At $K = 2$, the population were differentiated based on their origins and geographic distribution; the 299 varieties in Pop1 mostly originated in China and East Asia, whereas the 78 varieties in Pop2 are mainly introduced from Europe or have consanguinity of European varieties, like cultivars 'Jin Qie 320' and '17z36' (Fig. 2b). Both Pop1 and Pop 2 contain eggplants with various fruit colors and shapes. To further analyze the genetic structure of eggplant varieties, the population structure at $K = 3$ was studied. Pop1 was divided into two sub-populations Pop1A and Pop1B, with clear separation of fruit shape in eggplants (Fig. 2b). PCA and PCoA were conducted to assess the population structure (Fig. 2c). The two-dimensional plots of PCA and PCoA clearly indicated three clusters of eggplant varieties, which was consistent with Pop1A, Pop1B, and Pop2 inferred by structure analysis at $K = 3$. In addition, a phylogenetic analysis using an unrooted neighbor-joining tree was calculated from pairwise genetic distances (Fig. 3), which was in
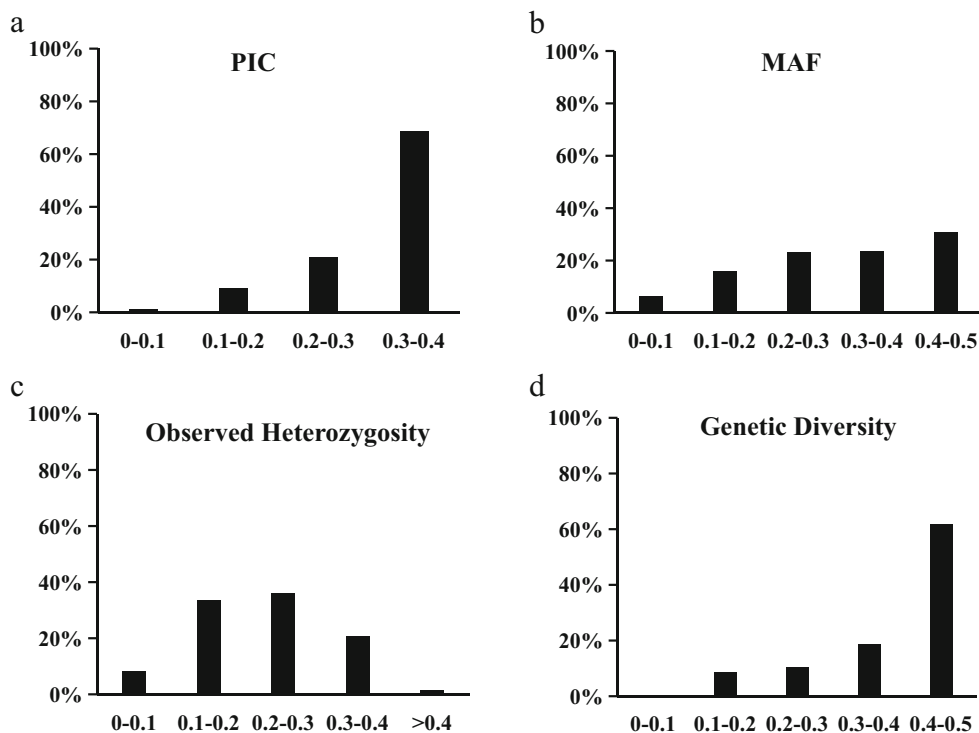

Fig. 1 Genetic diversity analysis with 219 perfect SNPs in the 377 eggplant varieties. Polymorphic information content (PIC; **a**), minor allele frequency (MAF; **b**), observed heterozygosity (H_o; **c**), genetic diversity (GD; **d**)
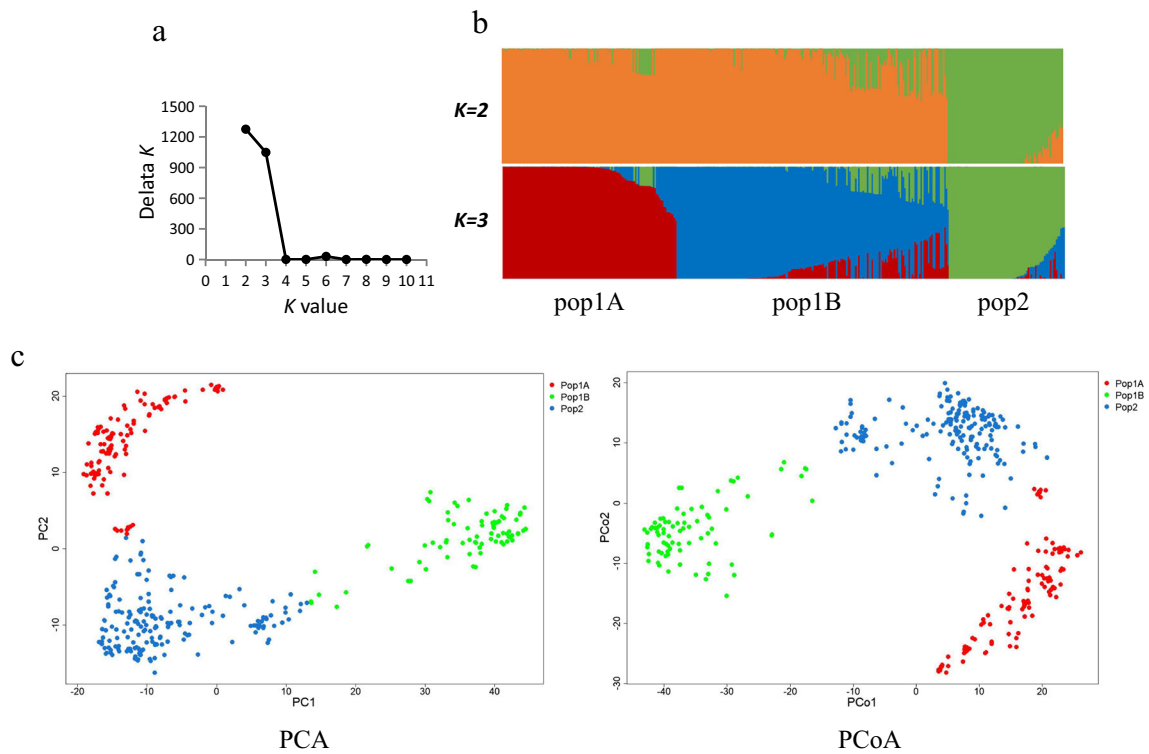
Fig. 2 Population structure, PCA, and PCoA analysis of the 377 eggplant varieties. (**a**) Delta K plots derived from 219 SNPs genotyping result. (**b**) Population classification at $K = 2$ (Pop1 colored in orange and Pop2 colored in green) and at $K = 3$ (Pop1A colored in red, Pop1B colored in blue, and Pop2 colored in green). (**c**) Principal component analysis and principal coordinate analysis from 219 SNPs genotyping result. The varieties belonging to Pop1A, Pop1B, and Pop2 were colored the same as that in (**b**)

accordance with the population structure inferred by the structure analysis, PCA and PCoA. All these suggested the 377 eggplant varieties could be classified into three populations. The presence of mixture was observed within the three populations (membership coefficient of its own population less than 0.8). Pop1A and Pop2 contained a small number of mixtures, whereas Pop1B contained a relatively higher amount of mixture with membership coefficient from both Pop1A and Pop1B. Interestingly, Pop1A was mainly composed of round- and oval-fruited eggplants, and most of the eggplant varieties in Pop1B had elongated fruits. This highlighted that fruit shape was strongly correlated to population structure for eggplant varieties in Asia, and that fruit shape is a major trait for selection, which had an impact on the genetic structure of eggplant varieties.

In this study, the AMOVA was conducted to assess the population structure of 377 eggplant varieties based on the 219 perfect SNPs. The AMOVA showed that the differences between three suggested populations (Pop1A, Pop1B, and Pop2) contributed 36.9% of the

variation, and the minimum variation (14.7%) occurred within populations (Table 1). Meanwhile, the difference between varieties contributed 48.4% of the total variations. At the meantime, a pairwise F_{st} estimation between the three populations was also performed to test for significant variations between the populations. Within Pop1, the round- and oval-fruited eggplants in Pop1A represented a distinct population from the long-fruited eggplants in Pop1B ($F_{st} = 0.2614$). Pop2 was genetically differentiated from both Pop1A ($F_{st} = 0.5023$) and Pop1B ($F_{st} = 0.3966$), indicating a closer relationship with Pop1B (Table S4). The relatively closer relationship between Pop1B and Pop2 may be partly due to their higher level of gene exchange suggested by the mixture between them.

Core SNPs set for genetic diversity analysis and variety identification

Commonly, a small number of highly informative SNP markers are selected as a core marker set for easy and
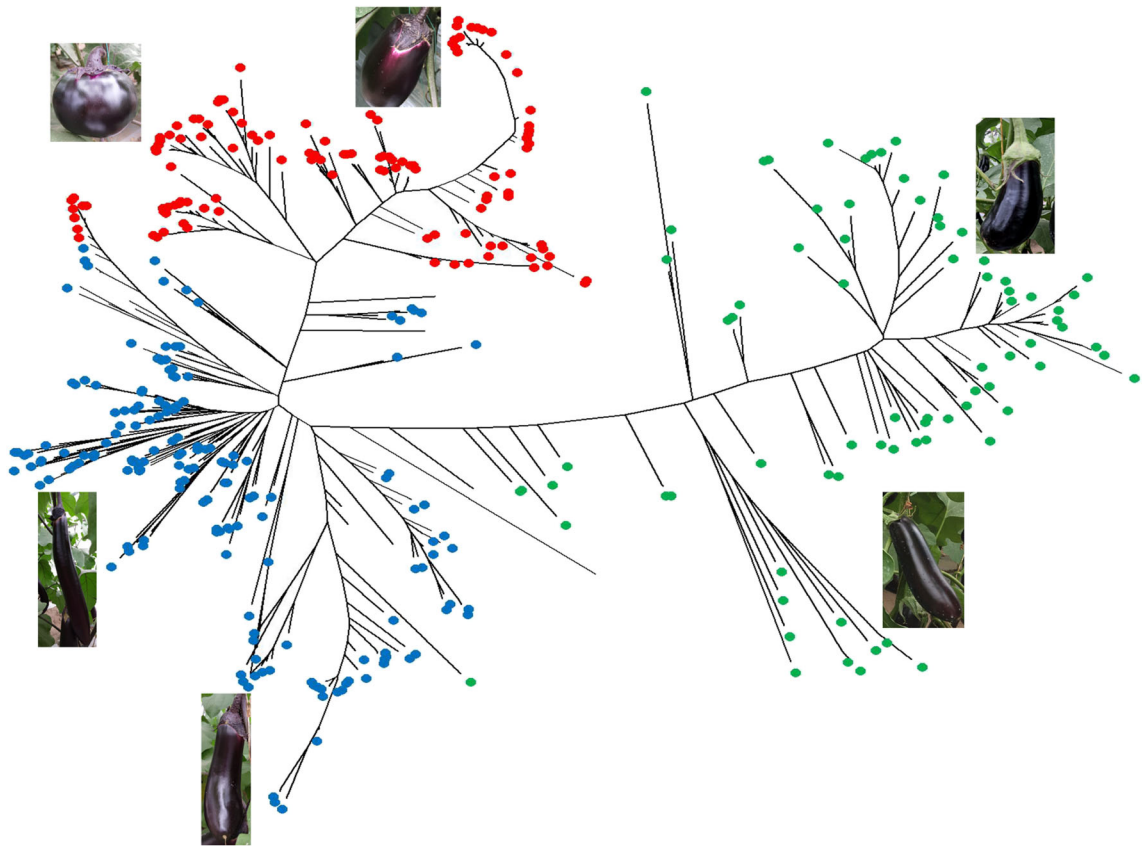
Fig. 3 Unrooted neighbor-joining tree of 377 eggplant varieties based on the 219 perfect SNPs. The varieties in Pop1A, Pop1B, and Pop2 are colored as in Fig. 3

fast study in genetic diversity analysis and variety identification. In this study, 36 SNPs were chosen as the core SNPs set with the ability to differentiate 95% of the 377 eggplant varieties (Fig. 4a). Moreover, the PCA and PCoA analysis using the 36 markers showed three well-separated clusters, which was consistent with the analysis using all 219 perfect SNPs (Fig. 4b). The neighbor-joining tree built using the 36 core SNPs also showed a clear classification in three groups (Fig. 4c). Therefore, these 36 core SNPs were sufficient for representing the genetic diversity of the 377 varieties with high efficiency in variety identification.

SNP polymorphism and genetic diversity within populations

The polymorphism of genome-wide perfect 219 SNP markers in eggplant genome was further evaluated within the three populations. Pop1A showed the lowest amount of polymorphic markers (180 SNPs), while both Pop1B and Pop2 had 215 polymorphic SNPs (Table 2). The five SNP markers with no polymorphism in Pop1B were behaving polymorphic in Pop2. Also, the five non-polymorphic SNPs in Pop2 displayed polymorphism in Pop1A and Pop1B. This

Table 1 Analysis of molecular variance (AMOVA) among populations and within populations in eggplant varieties

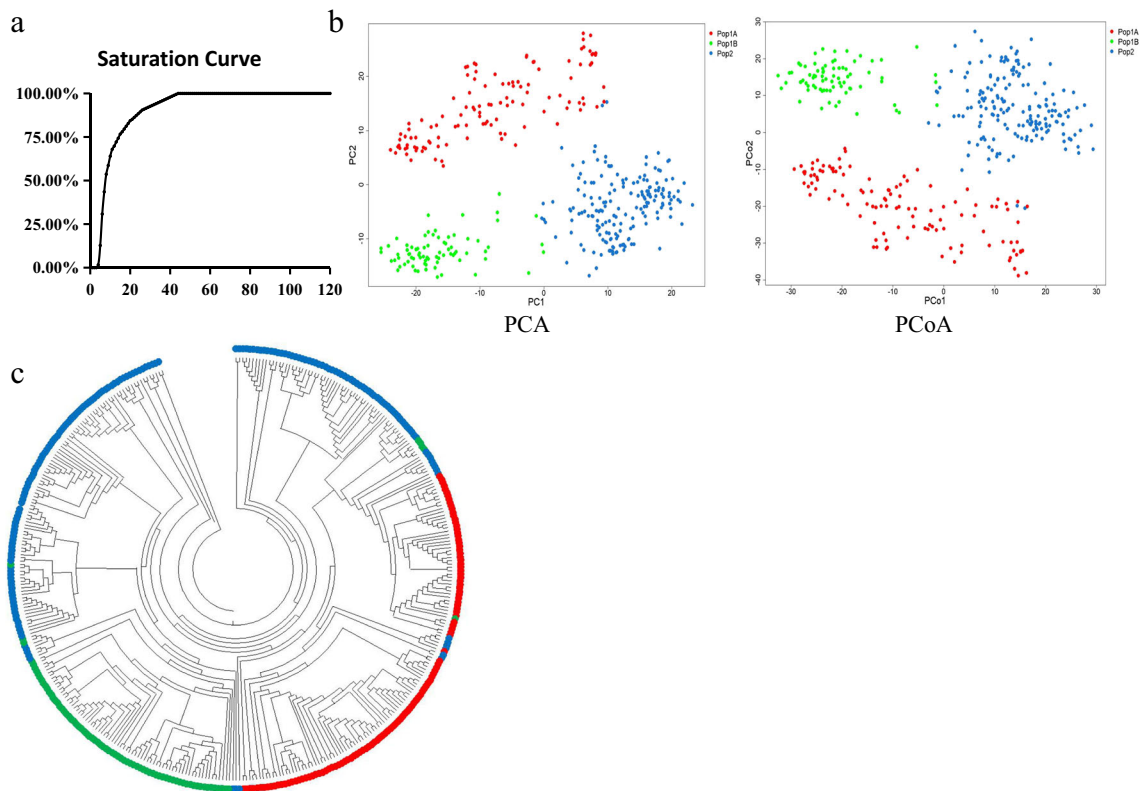| Source of variation	df	Sum of squares	Mean of squares	Variance components	Percentage of variation
Between populations	2	17,896.6	8948.3	37.5	36.9
Between samples within populations	374	29,629.4	79.2	14.1	14.7
Within samples	377	18,571.4	49.3	49.3	48.4
Total	753	66,097.4	87.8	101.8	100

Fig. 4 Genotyping and genetic analysis of eggplant varieties based on 36 core SNPs. (**a**) The saturation curve of 219 perfect SNPs identifying in 377 eggplant varieties. (**b**) PCA and PCoA analysis using 36 core SNPs. (**c**) Unrooted neighbor-joining tree of 377 eggplant varieties based on the 36 core SNPs. The varieties in Pop1A, Pop1B, and Pop2 are colored as in Fig. 3

further indicated the differentiation between the three populations. The lower polymorphic rate of SNPs in Pop1A in turn caused lower average GD and PIC value (0.248 and 0.199), compared to that in Pop1B (0.311 and 0.248) and in Pop2 (0.307 and 0.248) (Table 2). This demonstrated that a narrower genetic background is present in the round- and oval-fruited eggplant varieties (Pop1A). We accessed the inbreeding coefficient in the three populations. Pop2 had lowest inbreeding coefficient on average at 0.065, and Pop1B had a moderate inbreeding coefficient of 0.154. Meanwhile, Pop1A showed significantly higher inbreeding coefficient at 0.442 (Table 2). The high inbreeding coefficient in Pop1A corresponded to its low average H_o using either 219 SNPs (0.124) or its 180 polymorphic SNPs (0.152) (Table S7). This indicated a high frequency of inbreeding events among the round- and oval-fruited eggplants in Pop1A, which may explain their low genetic diversity detected by the 219 SNPs.

Genetic similarity and core varieties analysis

To further understand the gene exchange and genetic background of the 377 eggplant varieties, a genetic similarity matrix was built based on the number of differential SNP genotypes between varieties within each population. Fewer differential SNP genotypes between varieties indicates a closer relationship. Varieties in both Pop1B and Pop2 showed higher numbers of differential SNPs, at 95.72 and 96.81, respectively. Pop1A displayed the lowest average numbers of differential SNPs at 73.09, which further indicated its narrower genetic background (Fig. S3). Furthermore, the top 20% of varieties with minimum differential SNPs in each subgroup were selected as the core eggplant varieties (Table S6). We chose 24 varieties to represent Pop1 including 'Niu Xin Qie' and 'Jiu Ye,' 36 in Pop1B including 'Hei Jao Zi', 'Jin Qie 218,' and 'Ya Shu16–2,' and 16 in Pop2 including 'Brigitte' and 'Sharapova 10-203.'

Table 2 Marker polymorphism within the defined three population groups

Item	Total	Pop1A	Pop1B	Pop2
Number of accessions	377	116	183	78
Number of markers used	219	219	219	219
Number of polymorphic markers	220	180	215	215
Observed heterozygosity	0.224	0.124	0.260	0.289
Genetic diversity	0.398	0.248	0.311	0.307
MAF	0.311	0.185	0.235	0.226
PIC	0.313	0.199	0.248	0.248
Inbreeding coefficient	0.435	0.442	0.154	0.065

Genome-wide association analysis of fruit shape in 377 eggplant varieties

The genome-wide 219 SNPs were used to conduct a genome-wide association study to identify SNPs associated with fruit shape (round, oval, and elongated) in 377 eggplant varieties. There were five SNPs identified showing strong association with the fruit shape ($p <$ 0.005) on chromosomes 3, 6, and 7 (Table 3; Fig. S4). With the availability of reference eggplant genome, we were able to map previously identified fruit trait QTLs to the genome and compare their genomic locations with the identified SNPs in this study. It was found that three of five newly identified SNPs were co-localized with fruit trait QTLs in previous QTL mapping and association studies (Doganlar et al. 2002; Frary et al. 2014; Portis et al. 2014, 2015). The SmSNP152 on chromosome 3 was matched to the fs7.1 associated with fruit shape (Doganlar et al. 2002; Frary et al. 2014), as well as the fdmaxE03.ML (Portis et al. 2014), and the E03.2 (Portis et al. 2015) associated with fruit diameter in eggplant. Analyzing the tomato syntenic region of SmSNP152 (Hirakawa et al. 2014), we found that it contained one *SUN* homolog gene and two *OVATE* homolog genes (Table S8). The *SUN* and *OVATE* family

genes have been well studied as they control the fruit elongation in tomato (Liu et al. 2002; Xiao et al. 2008; Wu et al. 2018b). Moreover, the associated SmSNP025 and SmSNP034 both on chromosome 7 were analyzed to be matched to the known QTL regions underlying fruit length and shape in eggplant and are syntenic with regions of tomato genome containing *OVATE*-like genes (Portis et al. 2014). The SmSNP133 and SmSNP191 were not matched to the known QTLs published in research on eggplant, but their syntenic tomato regions were also observed to be harboring the *SUN* homologs (Table S8). Therefore, this study suggested that the *SUN* and *OVATE* homolog genes may play key roles in controlling the fruit shape in eggplant.

Discussion

Genome-wide perfect SNP discovery and its efficient utility in variety identification

In this study, we showed that the selection method for perfect SSRs (Yang et al. 2019) could be modified to select perfect SNPs from the variome of 45 eggplant material lines. Validation using the novel target SNP-

Table 3 SNPs associated with eggplant shape uncovered by association analysis

Associated marker	Genomic position	p value	PVE (%)	Previous QTL co-localized	Reference
SmSNP025	Chr07:136591058	1.37E−05	6.09	fsE07.ML, flE07.ML	Portis et al. 2014
SmSNP191	Ch00.12014:850453	9.35E−04	3.75	–	
SmSNP133	Chr06:97194057	9.61E−04	3.73	–	
SmSNP152	Chr03:93376781	3.60E−03	3.01	fs7.1, E03.2, fdmaxE03.ML	Doganlar et al. 2002; Portis et al. 2014, 2015
SmSNP034	Chr07:32028796	4.46E−03	2.89	E10.4	Portis et al. 2015

seq in a larger collection of 377 eggplant varieties demonstrated good performance and efficiency with an extremely low missing rate for the chosen 219 perfect SNPs. This showed that the target SNP-seq provides a reliable and efficient method for high-throughput genotyping. The simultaneous amplification of hundreds of target sites using multiplex PCR and combining of multiple samples in one single sequencing run greatly reduced the time and cost for large-scale genotyping. Also, the high PIC and MAF values of the 219 perfect SNPs indicated that they are informative, suggesting that they could be used for variety identification and authentication in a wider range of eggplant cultivars. Therefore, the use of variomes proved to be useful for marker selection and evaluation, which could help to reduce the number of markers that need to be validated and ensure marker quality. Combining the use of variomes with a high-throughput genotyping method, target SNP-seq could help to speed up the molecular marker development and genotyping for variety identification, as well as molecular breeding for many other crops.

Fruit shape as a major selection trait in eggplants

Model-based population structure analysis, PCA, and PCoA detected three differentiated populations, Pop1A, Pop1B, and Pop2, within the 377 eggplant varieties. This variety collection was firstly divided into two main groups, Pop1 including eggplant originating from China and East Asia, and Pop2 containing eggplants with consanguinity from Western Europe. This was consistent with the fact that Western Europe and China are two secondary eggplant diversity centers and previous genetic structure analysis (Cericola et al. 2013; Taher et al. 2017). Their molecular differentiation could be contributed by human selection, mutations and recombination as well as environmental adaptation in two different geographic regions. Within Pop1, two sub-populations, Pop1A and Pop1B, could be detected and were found to be correlated with fruit shape. The majority of the eggplants in Pop1A have round or oval fruit, whereas Pop1B is composed of long-fruited eggplant. This provided evidence at genetic level showing that fruit shape is a major selection trait in eggplants. This was also indicated in a previous study showing a correlation between phylogenetic classification and fruit shape within clades (Barchi et al. 2011). Therefore, fruit morphology, as a major trait for human selection and

breeding, had left a significant impact on the genetic structure of eggplants. This is also observed in other vegetable crops like tomato and pepper with selection for fruit shape as an important factor responsible for genetic structure, besides market specialization and environmental adaptation (Sim et al. 2009; Gonzalez-Perez et al. 2014).

In this study, we observed a significant lower level of genetic diversity in the round- and oval-fruited eggplants in Pop1A, indicated by their lower average GD and PIC values (Table 2). Furthermore, the low H_o and high inbreeding coefficient of Pop1A suggest a high level of inbreeding which could be the main reason behind the narrow background (Table 2). Also, a closer genetic similarity with the lowest average number of differential SNPs was observed for Pop1A (Fig. S3). This suggests that limited genetic material were used in the breeding program for round- and oval-fruited eggplants. Therefore, there is a high risk of genetic erosion in the round- and oval-fruited eggplants in Pop1A, which requires urgent attention in the breeding system. In Pop1A, a few varieties having membership coefficient from Pop1B based on structure analysis are all elongated eggplants. This suggests that introducing genetic material from elongated eggplants for round- or oval-fruited eggplants may be difficult to preserve the fruit shape trait. Genes promoting the elongated fruit may be dominant and there may be a higher chance of resulting long-fruited eggplants when crossing the round- or oval-fruited eggplants with the long-fruited eggplants. Marker-assisted selection (Zhou et al. 2003) could be used to quickly identify potential eggplants with the desired fruit traits which may help to introduce distant genetic materials for round- or oval-fruited eggplants at the same time to conserve the fruit shape. This would require an understanding of the genetic basis for fruit shape control in eggplant and to identify markers and genes associated with fruit shape, which can then be used for marker-assisted selection and molecular breeding.

Conservation of *SUN* and *OVATE*-like genes function in fruit shape control

The genotype of the 377 varieties using 219 genome-wide SNPs presented an opportunity for fruit shape association analysis in an attempt to discover SNPs that can be used for marker-assisted selection. This resulted in five associated SNPs to be detected. With location

information provided by the new reference genome, SmSNP025, SmSNP034, and SmSNP152 were found to be co-localized with QTL regions in previous studies (Table 3) (Doganlar et al. 2002; Frary et al. 2014; Portis et al. 2014, 2015). Particularly, the region of SmSNP152 at chromosome 3 has been found to be associated with fruit shape in QTL mapping studies using two independent F2 population in different locations (Doganlar et al. 2002; Frary et al. 2014; Portis et al. 2014) and one genome-wide association study (Portis et al. 2015). Repeated detection of this region highlights its significance in fruit shape control and its identification in a population derived from a cross between the cultivated eggplant, *S. melongena*, and its wild relative, *Solanum linnaeanum* (Doganlar et al. 2002), suggests its possible selection during domestication. This showed that these 219 genome-wide SNPs are useful for association study as we confirmed three QTLs and identified two novel regions associated with fruit shape in eggplant. The five associated SNPs would be valuable for marker-assisted selection programs and further QTL mapping and GWAS with higher marker density in these regions could help to identify the exact responsible genes.

The *SUN* and *OVATE*-like genes are two important gene families controlling fruit elongation, and they have been well mapped in several crops, such as tomato, pepper, cucumber, and melon (Liu et al. 2002; Xiao et al. 2008; Zygier et al. 2005; Pan et al. 2017; Che and Zhang 2019; Monforte et al. 2014). *SUN* encodes a protein promoting fruit elongation (Xiao et al. 2008), whereas *OVATE* encodes a protein playing a negative role in growth as its null mutation results in elongated fruit (Liu et al. 2002; Wu et al. 2018b). Both *SUN* and *OVATE* family genes have been shown to influence microtubule organization and dynamics with opposite effect on cell shape, and organ morphology (Lazzaro et al. 2018). In this study, examination of the syntenic region of the five associated SNPs in tomato genome found that all of them carry either *SUN* or *OVATE*-like genes. Particularly, the syntenic region in tomato genome of the SmSNP034 contains the *SlOFP20*, which has been fine-mapped and cloned showing to contribute to tomato fruit shape (Wu et al. 2018b). Also, the repeatably mapped region of SmSNP152 contains two *OVATE*-like genes and one *SUN*-like gene. The association found between regions containing *SUN* and *OVATE*-like genes and eggplant fruit shape suggests conservation of their function in fruit shape control in eggplant.

Conclusion

We showed that the eggplant reference genome and the variome of diverse genetic materials can be used to select 219 genome-wide perfect SNPs for variety identification and genetic analysis. Correlation between the eggplant fruit shape and genetic structure in Asia demonstrated that fruit shape is a major selection trait. The low genetic background detected using the 219 perfect SNPs in the round- and oval-fruited eggplants in Asia indicated a risk of genetic erosion and the urgent need to widen the choice of genetic materials in breeding programs. In our association study, we identified both previously detected and novel genomic regions associated with fruit shape in eggplant. Therefore, this study showed that the genome-wide perfect SNPs can be a valuable tool for variety identification and genetic analysis.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest.

Ethical approval   The authors declare that the experiments comply with the current laws of the country in which they were performed.

References

Acquadro A, Barchi L, Gramazio P, Portis E, Vilanova S, Comino C, Plazas M, Prohens J, Lanteri S (2017) Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. PLoS One 12:e0180774

Barchi L, Lanteri S, Portis E, Acquadro A, Vale G, Toppino L, Rotino GL (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. BMC Genomics 12:304

Barchi L, Acquadro A, Alonso D, Aprea G, Bassolino L, Demurtas O, Ferrante P, Gramazio P, Mini P, Portis E, Scaglione D, Toppino L, Vilanova S, Diez MJ, Rotino GL, Lanteri S, Prohens J, Giuliano G (2019a) Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. Front Plant Sci 10:1005

Barchi L, Pietrella M, Venturini L, Minio A, Toppino L, Acquadro A, Andolfo G, Aprea G, Avanzato C, Bassolino L, Comino C, Molin AD, Ferrarini A, Maor LC, Portis E, Reyes-Chin-Wo S, Rinaldi R, Sala T, Scaglione D, Sonawane P, Tononi P, Almekias-Siegl E, Zago E, Ercolano MR, Aharoni A, Delledonne M, Giuliano G, Lanteri S, Rotino GL (2019b) A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Sci Rep 9:11769

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635

Cericola F, Portis E, Toppino L, Barchi L, Acciarri N, Ciriaci T, Sala T, Rotino GL, Lanteri S (2013) The population structure and diversity of eggplant from Asia and the Mediterranean Basin. PLoS One 8:e73702

Che G, Zhang X (2019) Molecular basis of cucumber fruit domestication. Curr Opin Plant Biol 47:38–46

Cheng J, Qin C, Tang X, Zhou H, Hu Y, Zhao Z, Cui J, Li B, Wu Z, Yu J, Hu K (2016) Development of a SNP array and its application to genetic mapping and diversity assessment in pepper (Capsicum spp.). Sci Rep 6:33293

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) Conservation of gene function in the solanaceae as revealed by comparative mapping of domestication traits in eggplant. Genetics 161:1713–1726

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361. https://doi.org/10.1007/s12686-011-9548-7

Flint-Garcia SA (2013) Genetics and consequences of crop domestication. J Agric Food Chem 61:8267–8276

Frary A, Frary A, Daunay M-C, Huvenaars K, Mank R, Doğanlar S (2014) QTL hotspots in eggplant (Solanum melongena) detected with a high resolution map and CIM analysis. Euphytica 197:211–228

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Report 13:207–209

Gao P, Ma H, Luan F, Song H (2012) DNA fingerprinting of Chinese melon provides evidentiary support of seed quality appraisal. PLoS One 7:e52431

Gonzalez-Perez S, Garces-Claver A, Mallor C, Saenz de Miera LE, Fayos O, Pomar F, Merino F, Silvar C (2014) New insights into Capsicum spp relatedness and the diversification process of Capsicum annuum in Spain. PLoS One 9:e116276

Gramazio P, Blanca J, Ziarsolo P, Herraiz FJ, Plazas M, Prohens J, Vilanova S (2016) Transcriptome analysis and molecular marker discovery in Solanum incanum and S. aethiopicum, two close relatives of the common eggplant (Solanum melongena) with interest for breeding. BMC Genomics 17:300

Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, De Jong WS, Douches DS, Buell CR (2011) Single nucleotide polymorphism discovery in elite north American potato germplasm. BMC Genomics 12:302

Henning JA, Coggins J, Peterson M (2015) Simple SNP-based minimal marker genotyping for Humulus lupulus L. identification and variety validation. BMC Res Notes 8:542

Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H, Sato S, Isobe S, Tabata S, Fukuoka H (2014) Draft genome sequence of eggplant (Solanum melongena L.): the representative Solanum species indigenous to the Old World. DNA Res 21:649–660

Hurtado M, Vilanova S, Plazas M, Gramazio P, Fonseka HH, Fonseka R, Prohens J (2012) Diversity and relationships of eggplants from three geographically distant secondary centers of diversity. PLoS One 7:e41748

Jamali SH, Cockram J, Hickey LT (2019) Insights into deployment of DNA markers in plant variety protection and registration. Theor Appl Genet

Kamvar ZN, Tabima JF, Grunwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281

Kim JE, Oh SK, Lee JH, Lee BM, Jo SH (2014) Genome-wide SNP calling using next generation sequencing data in tomato. Mol Cells 37:36–42

Lazzaro MD, Wu S, Snouffer A, Wang Y, van der Knaap E (2018) Plant organ shapes are regulated by protein interactions and associations with microtubules. Front Plant Sci 9:1766

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760

Liu J, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Natl Acad Sci U S A 99:13302–13306

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Monforte AJ, Diaz A, Cano-Delgado A, van der Knaap E (2014) The genetic basis of fruit morphology in horticultural crops: lessons from tomato and melon. J Exp Bot 65:4625–4637

Munoz-Falcon JE, Prohens J, Vilanova S, Nuez F (2009) Diversity in commercial varieties and landraces of black eggplants and implications for broadening the breeders' gene pool. Ann Appl Biol 154:453–465

Naegele RP, Boyle S, Quesada-Ocampo LM, Hausbeck MK (2014) Genetic diversity, population structure, and resistance to Phytophthora capsici of a worldwide collection of eggplant germplasm. PLoS One 9:e95930

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590

Nishio SY, Hayashi Y, Watanabe M, Usami S (2015) Clinical application of a custom AmpliSeq library and ion torrent PGM sequencing to comprehensive mutation screening for deafness genes. Genet Test Mol Biomark 19:209–217

Pan Y, Liang X, Gao M, Liu H, Meng H, Weng Y, Cheng Z (2017) Round fruit shape in WI7239 cucumber is controlled by two

interacting quantitative trait loci with one putatively encoding a tomato SUN homolog. Theor Appl Genet 130:573–586

Plazas M, Lopez-Gresa MP, Vilanova S, Torres C, Hurtado M, Gramazio P, Andujar I, Herraiz FJ, Belles JM, Prohens J (2013) Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning. J Agric Food Chem 61:8871–8879

Portis E, Barchi L, Toppino L, Lanteri S, Acciarri N, Felicioni N, Fusari F, Barbierato V, Cericola F, Vale G, Rotino GL (2014) QTL mapping in eggplant reveals clusters of yield-related loci and orthology with the tomato genome. PLoS One 9: e89499

Portis E, Cericola F, Barchi L, Toppino L, Acciarri N, Pulcini L, Sala T, Lanteri S, Rotino GL (2015) Association mapping for fruit, plant and leaf morphology traits in eggplant. PLoS One 10:e0135200

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

Ranil RHG, Prohens J, Aubriot X, Niran HML, Plazas M, Fonseka RM, Vilanova S, Fonseka HH, Gramazio P, Knapp S (2017) Solanum insanum L. (subgenus Leptostemonum bitter, Solanaceae), the neglected wild progenitor of eggplant (S-melongena L.): a review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. Genet Resour Crop Ev 64:1707–1722

Shirasawa K, Monna L, Kishitani S, Nishio T (2004) Single nucleotide polymorphisms in randomly selected genes among japonica rice (Oryza sativa L.) varieties identified by PCR-RF-SSCP. DNA Res 11:275–283

Sim SC, Robbins MD, Chilcott C, Zhu T, Francis DM (2009) Oligonucleotide array discovery of polymorphisms in cultivated tomato (Solanum lycopersicum L.) reveals patterns of SNP variation associated with breeding. BMC Genomics 10:466

Taher D, Solberg SO, Prohens J, Chou YY, Rakha M, Wu TH (2017) World vegetable center eggplant collection: origin, composition, seed dissemination and utilization in breeding. Front Plant Sci 8:1484

Tian HL, Wang FG, Zhao JR, Yi HM, Wang L, Wang R, Yang Y, Song W (2015) Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. Mol Breed 35:136

Vos PG, Uitdewilligen JG, Voorrips RE, Visser RG, van Eck HJ (2015) Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. Theor Appl Genet 128:2387–2401

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, International Wheat Genome Sequencing C, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796

Wu J, Liu S, Wang Q, Zeng Q, Mu J, Huang S, Yu S, Han D, Kang Z (2018a) Rapid identification of an adult plant stripe rust resistance gene in hexaploid wheat by high-throughput SNP array genotyping of pooled extremes. Theor Appl Genet 131:43–58

Wu S, Zhang B, Keyhaninejad N, Rodriguez GR, Kim HJ, Chakrabarti M, Illa-Berenguer E, Taitano NK, Gonzalo MJ, Diaz A, Pan Y, Leisner CP, Halterman D, Buell CR, Weng Y, Jansky SH, van Eck H, Willemsen J, Monforte AJ, Meulia T, van der Knaap E (2018b) A common genetic mechanism underlies morphological diversity in fruits and other plant organs. Nat Commun 9:4734

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319: 1527–1530

Yang J, Zhang J, Han R, Zhang F, Mao A, Luo J, Dong B, Liu H, Tang H, Zhang J, Wen C (2019) Target SSR-Seq: a novel SSR genotyping technology associate with perfect SSRs in genetic analysis of cucumber varieties. Front Plant Sci 10:531

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhou PH, Tan YF, He YQ, Xu CG, Zhang Q (2003) Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. Theor Appl Genet 106:326–331

Zygier S, Chaim AB, Efrati A, Kaluzky G, Borovsky Y, Paran I (2005) QTLs mapping for fruit size and shape in chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato. Theor Appl Genet 111:437–445