Theor Appl Genet (2013) 126:1285–1303
DOI 10.1007/s00122-013-2053-5

ORIGINAL PAPER

# SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium

**Cristina Esteras · Gelsomina Formisano · Cristina Roig · Aurora Díaz · José Blanca · Jordi Garcia-Mas · María Luisa Gómez-Guillamón · Ana Isabel López-Sesé · Almudena Lázaro · Antonio J. Monforte · Belén Picó**

Received: 4 November 2012 / Accepted: 19 January 2013 / Published online: 5 February 2013


**Abstract** Novel sequencing technologies were recently used to generate sequences from multiple melon (*Cucumis melo* L.) genotypes, enabling the in silico identification of large single nucleotide polymorphism (SNP) collections. In order to optimize the use of these markers, SNP validation and large-scale genotyping are necessary. In this paper, we present the first validated design for a genotyping array with 768 SNPs that are evenly distributed throughout the melon genome. This customized Illumina GoldenGate assay was used to genotype a collection of 74 accessions, representing most of the botanical groups of the species. Of the assayed *loci,* 91 % were successfully genotyped. The array provided a large number of polymorphic SNPs within and across accessions. This set of SNPs detected high levels of variation in accessions from this crop's center of origin as well as from several other areas of melon diversification. Allele distribution throughout the genome revealed regions that distinguished between the two main groups of cultivated accessions (*inodorus* and *cantalupensis*). Population structure analysis showed a subdivision into five subpopulations, reflecting the history of the crop. A considerably low level of LD was detected, which decayed rapidly within a few kilobases. Our results show that the GoldenGate assay can be used successfully for high-throughput SNP genotyping in melon. Since many of the genotyped accessions are currently being used as the parents of breeding populations in various programs, this set of mapped markers could be used for future mapping and breeding efforts.

Communicated by H. Nybom.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-013-2053-5) contains supplementary material, which is available to authorized users.

C. Esteras · C. Roig · J. Blanca · B. Picó (✉)
COMAV, Institute for the Conservation and Breeding of Agricultural Biodiversity, Universitat Politècnica de València (UPV), Camino de Vera s/n, 46022 Valencia, Spain
e-mail: mpicosi@btc.upv.es

G. Formisano
Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples 'Federico II', Via Universita' 100, 80055 Portici, Italy

A. Díaz · A. J. Monforte (✉)
Instituto de Biología Molecular y Celular de Plantas (IBMCP), Universitat Politècnica de València (UPV)-Consejo Superior de Investigaciones Científicas (CSIC), Ciudad Politécnica de la Innovación (CPI), Ed. 8E, C/Ingeniero Fausto Elio s/n, 46022 Valencia, Spain
e-mail: amonforte@ibmcp.upv.es

J. Garcia-Mas
IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB, Carretera de Cabrils Km 2, 08348 Cabrils, Barcelona, Spain

M. L. Gómez-Guillamón · A. I. López-Sesé
Instituto de Hortofruticultura Subtropical y Mediterránea, Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), Est. Exp. La Mayora, Algarrobo-Costa, Málaga, Spain

A. Lázaro
IMIDRA (Madrilean Research Institute for Rural Development, Agriculture and Food), Autovía A-II, Km 38,200, Alcalá de Henares, 28800 Madrid, Spain

## Introduction

Melon (*Cucumis melo* L., $2n = 2x = 24$) is one of the most important fruit crops worldwide. Current world production of melon is over 22 million tons, and is of particular importance in Mediterranean and East Asian countries (FAO 2012). Modern melon seeds are hybrids and have a significant and growing economic impact.

This species is also of interest for its particular biological characteristics and high diversity. *C. melo* is considered to be divided into two subspecies, *melo* and *agrestis* (Jeffrey 1980), each of which includes several cultivar groups that have been merged or split in several different classification proposals (Naudin 1859; Munger and Robinson 1991; Pitrat et al. 2000). One of the latest is that proposed by Pitrat (2008), who divided the subspecies *melo* into ten groups (*cantalupensis*, *reticulatus*, *adana*, *chandalak*, *ameri*, *inodorus*, *chate*, *flexuosus*, *dudaim*, and *tibish*) and the subsp. *agrestis* into five (*momordica*, *conomon*, *chinensis*, *makuwa*, and *acidulus*). However, some of these groups are quite heterogeneous, and accessions displaying intermediate features are difficult to classify.

Both African and Asian origins have been proposed for the species (Sebastian et al. 2010). During its domestication, melon underwent an intense process of diversification, and primary and secondary centers of diversity now spread from Eastern Asia to the Mediterranean Sea (Pitrat 2008). Nowadays, the species comprises wild, feral, and cultivated genotypes, with those of commercial interest mostly belonging to the *cantalupesis* and *inodorus* groups (Fernández-Trujillo et al. 2011).

High-throughput transcriptome data of melon were recently generated using novel sequencing technologies (Roche 454 pyrosequencing and the SOLID platform) (Blanca et al. 2011, 2012). The availability of these large collections of expressed sequence tags (ESTs), derived from multiple melon genotypes, has enabled the in silico identification of single nucleotide polymorphisms (SNPs). These and other melon SNP resources are freely accessible at the Melogene database (http://melogene.net/). The recent publication of the sequence of the melon genome (Garcia-Mas et al. 2012) has provided an additional resource to map all these SNPs in the melon genome (http://melonomics.net/).

Single nucleotide polymorphisms are valuable markers for the construction of high-resolution genetic maps, for the study of population structure, and for the discovery of marker–trait relationships in association-mapping experiments. However, efforts to validate the in silico-identified SNPs have been limited to a few tens of markers in a subset of melon genotypes (Blanca et al. 2011, 2012). In other crops, high-throughput SNP discovery has been accompanied by the development of array-based genotyping platforms that permit the rapid scoring of several hundred or thousand markers in parallel. The combination of the GoldenGate genotyping assay (Illumina, San Diego, CA, http://www.illumina.com/) with VeraCode or BeadExpress technologies provides one of the most robust and flexible platforms for SNP genotyping. GoldenGate platforms have been developed and successfully used in barley, wheat, rice, maize, soybean, chickpea, apple, conifers, and other plants (Rostoks et al. 2006; Hyten et al. 2008; Akhunov et al. 2009; McMullen et al. 2009; Graham et al. 2010; Yamamoto et al. 2010; Yan et al. 2010; Chancerel et al. 2011; Khana et al. 2012; Choudhary et al. 2012). This high-throughput method for multiplexing SNP assays has remained untested in melons up to this point.

Given the great multiplexing potential of the GoldenGate assay, we felt it would be relevant to utilize this technology for large-scale SNP validation in melon using a diverse germplasm set. We designed a customized 768-SNP GoldenGate assay, using markers distributed throughout the genome, that had been discovered through the resequencing of a number of diverse accessions, including the parents of the reference melon genetic map (Díaz et al. 2011) and of the DHL line used for sequencing the entire melon genome (Garcia-Mas et al. 2012). In this study, we used this platform to genotype a diverse set of melon accessions (which included landraces, modern cultivars, exotic accessions, and wild relatives), generating valuable information about the diversity, genetic structure, and differentiation between melon populations. This is also the first study of the extent of linkage disequilibrium (LD) in melons. The platform presented herein is a new resource for accelerating germplasm management in melon breeding programs.

## Materials and methods

### Melon genotypes

The panel of melon genotypes used in this study consisted of 93 accessions belonging to both *C. melo* subsp. *melo* and *C. melo* subsp. *agrestis*. Their classification in botanical groups was according to Pitrat (2008). The accessions were studied to address three different objectives: to study the genetic diversity of melon using the largest SNP genotyping platform designed to date, to compare the variability measurements obtained using AFLP markers as a complementary technology, and to undertake the first analysis of the extent of LD in melon. A summary of the accessions used in each assay is included in Table 1 and a more detailed description of melon genotypes is available in Online Resource Table 1. In addition, photographs of all of the analyzed accessions can be found in Online Resource Fig. 1.

**Table 1** Summary of the melon accessions analyzed in each assay

| Group | Accessions | Total | SNPs | AFLPs | LD |
|---|---|---|---|---|---|
| *Cucumis melo subsp. melo* | | | | | |
| *Inodorus* | Commercial cultivars | 4 | 4 | 2 | 2 |
| | Spanish landraces | 32 | 30 | 17 | 14 |
| | Other European, Asian and African landraces | 6 | 4 | 4 | 4 |
| *Ameri* and *other* | European, Asian and African landraces | 9 | 3 | 2 | 6 |
| *Cantalupensis and reticulatus* | European, Asian and American commercial cultivars and landraces | 20 | 11 | 11 | 14 |
| *Flexuosus and chate* | European and Asian accessions | 4 | 4 | 4 | 0 |
| *Dudaim* | Asian accession | 1 | 1 | 1 | 0 |
| *Cucumis melo subsp. agrestis* | | | | | |
| *Momordica* | Indian accessions | 4 | 4 | 4 | 0 |
| *Conomon* | Far-Eastern accessions | 5 | 5 | 5 | 0 |
| *Acidulus, tibish, chito and agrestis* | African and Asian accessions, including wild genotypes | 8 | 8 | 7 | 0 |
| All groups | | 93 | 74 | 57 | 40 |

We selected a subset of 74 genotypes for the genotyping assay with the SNP GoldenGate platform. The two best-represented groups of the subsp. *melo* were *inodorus* and *cantalupensis*, which are both very commercially relevant (Table 1). Representatives of most market classes within each group were analyzed. The *inodorus* group included four commercial cultivars (three 'Piel de Sapo' and one 'Honeydew') and a set of 30 Spanish landraces, representing the full range of diversity of traditional Spanish melons ('Piel de Sapo', 'Amarillo', 'Blanco', 'Rochet', 'Tendral,' and other intermediate types) (Online Resources Table 1 and Fig. 1). Most are only cultivated for self-consumption or for sale in local markets. The 'Piel de Sapo' group, however, is widely cultivated for national and international markets. Several representatives of *inodorus* landraces from Eastern Europe, North Africa and Asia were also included. The *cantalupensis* group included 11 accessions that are widely or locally grown in Europe, such as the 'Charentais' types ('Vedrantais' and 'Nantais Oblong'), the cantaloupe cultivars from France ('Prescott', 'D'Alger' and 'Noir des Carmes') and Israel ('Noy Israel'), some types included in the *reticulatus* group from the USA and Italy ('Dulce', Topmark, PMR45 and 'Popone'), and one gynoecious breeding line, WI 998 (Online Resource Fig. 1). Three landraces from Western Asia of the *ameri* and *chandalak* groups, four snakemelons (*flexuosus* and *chate*) from Europe and the Middle East, and one representative of the ornamental aromatic group *dudaim* were also analyzed (Online Resource Table 1 and Fig. 1).

The best-represented groups of the subspecies *agrestis* were *momordica* and *conomon* (Table 1), as they are widely used as breeding materials, and included accessions from India and Far-East countries, respectively. Apart from these groups, several non-sweet and non-climacteric *agrestis* accessions were assayed: one *acidulus* accession from India, one *tibish* from Sudan (considered to be the most primitive form of melon), one *chito*, and five wild *agrestis* melons with small fruits from Africa and India (Online Resource Table 1 and Fig. 1).

A subset of 57 genotypes of those included in the SNP array was also analyzed using AFLPs (Table 1, Online Resource Table 1). A set of 40 accessions, mostly belonging to the *inodorus* and *cantalupensis* groups, was selected for the LD analysis (Table 1, Online Resource Table 1). Twenty-one were selected from those included in the SNP genotyping assay, and a new set of 19 accessions was added to better represent most of the genetic variability of the cultivated groups.

## SNP genotyping array

### *SNP selection*

A set of 768 putative SNPs were included in the GoldenGate platform (Online Resource Table 2). This set comprised SNPs selected from three different sources:

(a) Five hundred and forty-eight SNPs were selected from a collection of several tens of thousands of polymorphisms identified in silico by Blanca et al. (2011) using an EST collection produced by 454 and Sanger sequencing. In that study, the 454 ESTs came from the two genotypes used as the parents of the melon genetic map [*C. melo* subsp. *melo* var. *inodorus* cv. 'Piel de Sapo' T111 (In-Ps-T111) and *C. melo* subsp. *agrestis* var. *conomon* cv. 'Songwhan Charmi' (Con-SC) (Deleu et al. 2009)] and also as parents of the DHL line used in the genome sequencing project (Garcia-Mas et al. 2012). The Sanger ESTs were generated from these two accessions as well as from seven

additional genotypes [two *inodorus* cultivars, 'Piel de Sapo Piñonet' (In-Ps-Piñ) and 'TamDew', four *cantalupensis* cultivars ['Vedrantais' (Can-Ved), C35, 'Dulce' (Can-Dul) and 'Noy Israel' (Can-NY)] and one additional *conomon* type, Pat81 (Con-Pat81)]. Several criteria were used to select these SNPs from the whole collection in order to optimize their validation in a GoldenGate platform. SNPs with a flanking sequence of at least 120 bp and with no other SNPs or introns in their vicinity were selected preferentially using the filters described in Blanca et al. (2011). All the selected SNPs were polymorphic between T111 and SC in silico.

(b) The second set of 158 SNPs was selected from those already mapped in the previous version of the melon genetic map by Deleu et al. (2009). These SNPs were primarily derived from an in silico Sanger EST collection (González-Ibeas et al. 2007) and were selected for being polymorphic between T111 and SC. After validation by resequencing, these SNPs were mapped using CAPS, SNAPShot, and pyrosequencing (Deleu et al. 2009). It was necessary to adapt these SNPs to the GoldenGate genotyping assay by using a 120-bp flanking sequence.

(c) The third set of SNPs was specifically selected for the LD analysis from three different genome contigs. A total of 62 SNPs identified in these contigs were included in the platform. Details about the selection strategy of these SNPs are described below.

This SNP platform was previously used to increase the density of the melon genetic map and to anchor the genome sequence during the genome sequencing project. To do this, the available mapping population (Deleu et al. 2009) was included in the GoldenGate genotyping assay and marker sequences were subjected to BLAST against the available version of the melon genome (Garcia-Mas et al. 2012; http://melonomics.net/).

All the information about the SNPs used in the GoldenGate platform is included in Online Resource Table 2: the SNPs and flanking sequence, their map and genome position, the annotation details of the corresponding unigenes, and the number of reads of each allele in the sequenced genotypes (for the SNPs selected in silico by Blanca et al. (2011)).

### *GoldenGate assay*

The genotyping with the GoldenGate platform was carried out at the Spanish "Centro Nacional de Genotipado" [CEGEN-ISCIII, (http://www.cegen.org)]. The sequence of each selected *locus*, including the polymorphic nucleotide and a 60-bp flanking sequence, was submitted by CEGEN-ISCIII to the Illumina Assay Design Tool (ADT) (Illumina). Designability scores were then given to each *locus*. These scores ranged from 0 to 1.0. The GoldenGate genotyping assay was conducted as described elsewhere (Blanca et al. 2012). The custom oligos that were used for the Illumina assay, 3 at each of the 768 different SNP *loci*, as well as the Illumina scores are listed in Online Resource Table 2. The GoldenGate assay was deployed on the BeadXpress R platform using Veracode R technology (Illumina). The PCR products, labeled Cy3 or Cy5 depending on the allele, were hybridized to glass Veracode micro-beads, each bearing a locus-specific barcode via the corresponding Illumicode sequence. Then, each SNP was identified by its IllumiCode and alleles were discriminated by their fluorescent signals on a Veracode BeadXpress Reader. The automatic allele calling for each *locus* was accomplished using GenomeStudio (Illumina). The clusters were manually edited when necessary.

Total DNA of the whole collection was extracted from young leaves using the CTAB method with minor modifications (Esteras et al. 2012a). To improve the quality of the obtained DNA, 70 % ethanol containing 15 mM ammonium acetate was used in the last wash, and the DNA was treated with RNase A. DNA concentrations in TE buffer were adjusted to 50 ng/μl, with the PicoGreen fluorescence being measured on an ABI7900 apparatus (Applied Biosystems). DNA of the 74 selected genotypes was delivered to CEGEN for genotyping. In order to assess the quality of the SNP calls, we duplicated three accessions that were randomly selected using independent DNA preparations. The proportion of consistent calls across all accessions was 100 %.

### AFLP genotyping

A subset of melon accessions was also genotyped using AFLP markers (Table 1 and Online Resource Table 1). AFLP analysis was performed as reported in Fita et al. (2011) using six primer combinations: *Eco*RI-ACA/*Mse* I-CA, *Eco*RI-AGG/*Mse* I-CA, *Eco*RI-ACG/*Mse* I-CA, *Eco*RI-ACC/*Mse* I-CA, *Eco*RI-AGC/*Mse* I-CA, and *Eco*RI-AAC/*Mse* I-CA.

### Data analysis

The results of the SNP genotyping were used to calculate major allele frequency, gene diversity, and polymorphism information content (PIC) for each *locus*. In addition, Nei et al.'s (1983) genetic distances (Nei et al. 1983) across the genotypes and neighbor-joining (NJ) tree were calculated using PowerMarker 3.5 (Liu and Muse 2005). The NJ tree was plotted with TreeView 1.6.6. AFLP data were analyzed with Popgene v. 1.32 (Yeh et al. 2000). A principal component analysis (PCA) was done on the SNP genotyping data using smartPCA (Patterson et al. 2006). This method takes into account the codominant nature of these

markers and was specifically developed to work optimally with this kind of data. For the AFLP data, the multivariate analysis was a principal coordinate analysis carried out on a Jaccard distance matrix with a custom script, available upon request, which implemented the method described by Krzanowski (2000).

The population structure underlying the genotyped collection of accessions was analyzed using STRUCTURE v2.2 (Pritchard et al. 2000). STRUCTURE's quantitative clustering method uses a Bayesian approach to identify subpopulations and to assign individuals to these populations. Given a sample of individuals, $K$ populations are assumed and individuals are assigned to these populations. $K$ values from 2 to 10 were tested for 20 independent runs for each $K$ with 1,000,000 iterations and a run length of 1,000,000. The five log likelihood values for each $K$ were then plotted to find the $K$s around a plateau of the likelihood values. We also identified the best $K$ using the delta $K$ method (Evanno et al. 2005). The proportion of the genome of an individual that belongs to each inferred population (admixture) was estimated. We also studied genetic differentiation by calculating Wright's $F$ statistics (Fit, Fis, and Fst) values for each SNP between the *inodorus* and *cantalupensis* groups using Powermarker 3.5 and Genepop 4.0.10 (Raymond and Rousset 1995; Rousset 2008; http://genepop.curtin.edu.au/).

### LD analysis

The extent of LD in melon had not previously been studied, so in order to choose an appropriate set of SNPs for LD analysis, we followed a sequential approach: (1) define the minimum distance necessary to detect LD between SNPs based on LD analysis in a single genomic region and sample of melon accessions; (2) identify SNPs in additional genomic regions that separated according to the results obtained in (1); (3) genotype the selected SNPs in a larger germplasm sample; (4) calculate LD within the *inodorus* and *cantalupensis* accessions.

For step (1), eight amplicons located in positions 41, 42, 45, 49, 53, 113, and 204 Kb of the contig consisting of two previously sequenced BAC clones, 60K17 (van Leeuwen et al. 2003) and I3J4 (Deleu et al. 2007) (included in a region covering 163 Kb of the CM3.5_scaffold00003 in the current version of the melon genome, positions 5,604,802 to 5,768,125: http://melonomics.net) were sequenced in 40 accessions [mostly *inodorus* and *cantalupensis* (Table 1)]. This represented most of the genetic variability of cultivated melons in the COMAV genebank collection (Esteras et al. 2009; Online Resource Table 1). Pair-wise $r^2$ values between informative SNPs (Major Allele Frequency, MAF $< 0.95$) were calculated with GGT 2.0 (van Berloo 2008) and double-checked with TASSEL (Bradbury et al. 2007). The decay of LD over physical distance was studied by plotting $r^2$ against distances in bp, fitting the data using a second-degree locally weighted scatterplot smoothing (LOESS; Breseghello and Sorrells 2006) implemented in an Excel plug-in (Peltier 2009). The minimum distance for LD detection was defined at the distance where the LOESS fitted curve reached a plateau.

For step (2), a set of amplicons covering two regions of about 12 Kb in the melon genomic scaffolds scaffold00016 (position from 5,215,669 to 5,226,781) and scaffold00011 (from 413,312 to 424,397) were sequenced in 17 melon accessions selected from those used in step 1 (Online Resource Table 1). Informative SNPs (MAF $< 0.95$) were then defined. For step (3), 62 SNPs that were evenly distributed in the three genomic regions screened in (1) and (2), and located at the maximum separation according to (1), were included in the GoldenGate assay and, consequently, the 74 accessions used for the SNP diversity assay were genotyped. Additional SNPs from the SNP genotyping assay with known physical positions were also included in the LD analysis. LD was studied as described above. Baseline $r^2$ was calculated from the parametric 95th percentile of the distribution of the unlinked markers.

## Results

### SNP validation

The GoldenGate genotyping assay was carried out successfully, with 91 % of the SNPs (698) genotyped in most of the accessions, taking into account both monomorphic and polymorphic *loci* (Online Resource Table 3). The remaining 70 SNPs had different clustering or amplification problems (more than three clusters, absence of or low cluster separation, low intensity, amplification of only one allele, and/or a high percentage of failed samples) (in red in Online Resource Table 2). The average designability rank scores for failed SNPs (0.86) was remarkably lower than that of the successful SNPs (0.91), but most scores were $>0.6$, which is considered to be the optimal threshold for a GoldenGate assay. Various matching problems with primers, the existence of additional SNPs in the annealing regions, non-allele-specific matches, the existence of more than two alleles and/or the amplification of a duplicated genomic region might be the cause of these genotyping problems. We blasted the 120-bp sequences used for the SNP design against the current version of the whole melon genome (http://melonomics.net/). The percentage of markers that mapped in more than one genomic region and displayed low sequence similarity with the genomic sequence was higher in failed than in working SNPs (Online Resource Table 2). The percentage of failed

markers was lower (44/548, 8 %) in the group of SNPs selected in silico by Blanca et al. (2011) than in those previously mapped with different technologies by Deleu et al. (2009) (20/158, 13 %) (Table 2). This result suggests that the specific filters applied for the selection of the in silico-identified SNPs (avoiding introns, additional SNPs in the vicinity and highly variable regions) actually optimize their use in a GoldenGate platform.

Of the new in silico selected SNPs, 92 % (504/548) produced high-quality signals in the GoldenGate platform. Among these successful SNPs, 53 were monomorphic in all accessions, including the two parents of the mapping population (T111 and PI 161375) (in blue in Online Resource Tables 2 and 3), indicating that these were false SNPs. Thus, the final success rate after validation was about 90 %. All of these SNPs were in silico-selected for variability between T111 and PI 161375, but the in silico screening was conducted using ESTs from seven additional genotypes (Blanca et al. 2011; Online Resource Table 2). In addition, when possible, SNPs with several reads per allele were also selected. One hundred and forty out of these 504 SNPs had two or more reads per allele. The percentage of these markers that were validated as real SNPs (138/140, 99 %) was higher than that of those SNPs having only one read in one or both alleles (313/364, 86 %).

Most of the polymorphic SNPs were mapped (443 new, 126 previously mapped SNPs and 33 selected for LD) using the reference mapping population employed in previous melon maps (Fernandez-Silva et al. 2008; Deleu et al. 2009; Díaz et al. 2011) (Table 2). This mapping information was used to anchor the melon genome sequence (Garcia-Mas et al. 2012) and is available at http://melonomics.net/. One set of SNPs was heterozygous or failed in one of the parents and could not be mapped (in orange in Online Resource Tables 2 and 3).

## Diversity in melon accessions

Six hundred and twenty-eight SNPs were polymorphic in the set of 74 accessions analyzed using the GoldenGate platform (Online Resource Table 3). Despite most SNPs having been selected for poymorphism between T111 and SC (Deleu et al. 2009; Blanca et al. 2011), many of them also detected polymorphism between accessions belonging to the same or to different groups. The number of SNPs that detected polymorphism within the different groups and their average MAF is detailed in Table 3. The polymorphism detected with AFLPs in the subset of accessions is also indicated.

The lowest percentage of polymorphism was found, as expected, among the commercial *inodorus* 'Piel de Sapo' accessions, which are also quite phenotypically uniform (Online Resource Fig. 1). The group of Spanish *inodorus* landraces displayed a substantial degree of polymorphism, only slightly lower than that of the European, Asian and African landraces (including accessions of the *inodorus* and *ameri* groups and those from various other countries) (Table 3). The *cantalupensis* group was more variable than the *inodorus* group. Also, large sets of SNPs detected polymorphism within *flexuosus*, *momordica* and *agrestis*. A remarkably low variability was found within the *conomon* group, which included highly phenotypically variable accessions (variable in sex type, fruit traits, mechanisms of fruit ripening, sugar content, etc.) (Online Resource Fig. 1).

The relationships between accessions were studied using SNPs and AFLPs. Even though there were fewer AFLP markers than SNPs (190 vs. 698, respectively), and the set of accessions used in each analysis was not identical, both multivariate analyses had similarities (Fig. 1a, b). The three main coordinates explained 57.7 and 59.2 % of the observed variation (in the SNPs and AFLPs, respectively), separating accessions from the two subspecies *melo* and

**Table 2** Summary of the SNP genotyping results obtained using the GoldenGate platform

| SNP source | SNPs included in the GG platform/average designability score | Failed (%)/ average designability score | Monomorphic (%)/ average designability score | Polymorphic (%)/ average designability score | Mapped (%)/ average designability score |
|---|---|---|---|---|---|
| In silico detected Blanca et al. (2011) | 548/0.92 | 44 (8 %)/0.89 | 53 (9.7 %)/0.94 | 451 (82.3 %)/0.92 | 443 (80.8 %)/0.92 |
| Previously mapped Deleu et al. (2009) | 158/0.87 | 20 (12.7 %)/0.84 | 12 (7.6 %)/0.88 | 126 (79.7 %)/0.87 | 126 (79.7 %)/0.87 |
| Identified by sequencing for LD | 62/0.83 | 6 (9.7 %)/0.81 | 5 (8 %)/0.87 | 51 (82.3 %)/0.83 | 33 (53.2 %)/0.85 |
| All SNPs | 768 | 70 (9.1 %) | 70 (9.1 %) | 628(81.8 %) | 602 (78.4 %) |

**Table 3** Number (percentage) of polymorphic SNPs in the different groups and average MAF

| Group | *Inodorus* | *Inodorus* | *Inodorus* and *ameri* | *Cantalupensis* and *reticulatus* | *Flexuosus* and *chate* | *Momordica* | *Conomon* | *Acidulus tibish*, *chito* and *agrestis* |
|---|---|---|---|---|---|---|---|---|
| Accessions<br>SNP source | Commercial Piel de sapo | Spanish landraces | European, Asian and African landraces | European, Asian and American commercial cultivars and landraces | European and Asian accessions | Indian accessions | Far-Eastern accessions | African and Asian accessions and wild genotypes |
| In silico detected Blanca et al. (2011) | 41 (8.1 %) 0.66 | 185 (36.7 %) 0.86 | 203 (40.3 %) 0.75 | 267 (53.0 %) 0.75 | 305 (60.5 %) 0.69 | 282 (56.0 %) 0.69 | 33 (6.6 %) 0.69 | 222 (44.0 %) 0.76 |
| Previously mapped Deleu et al. (2009) | 12 (8.7 %) 0.67 | 57 (41.3 %) 0.86 | 66 (47.8 %) 0.74 | 78 (56.5 %) 0.74 | 86 (62.3 %) 0.68 | 65 (50 %) 0.67 | 4 (2.9 %) 0.75 | 58 (42.0 %) 0.8 |
| Identified by sequencing for LD | 4 (7.1 %) 0.67 | 42 (75 %) 0.88 | 27 (48.2 %) 0.74 | 42 (75 %) 0.62 | 41 (73.2 %) 0.60 | 31 (55.4 %) 0.73 | 9 (16.1 %) 0.73 | 20 (35.7 %) 0.76 |
| All working SNPs | 57 (8.2 %) | 284 (40.7 %) | 296 (42.4 %) | 387 (55.4 %) | 432 (61.9 %) | 378 (54.2 %) | 46 (6.6 %) | 300 (43 %) |
| AFLPs | – | 64 (33.7 %) | 75 (39.5 %) | 62 (32.7 %) | 70 (36.8 %) | 60 (31.6 %) | 66 (34.6 %) | 147 (77.4 %) |

Degree of polymorphism obtained with AFLPs (percentage)

Percentages are indicated over the total working SNPs. Groups of accessions used for SNP and AFLP assays are detailed in Online Resource Table 1

*agrestis*, placing *momordica*, *flexuosus,* and *dudaim* accessions in an intermediate position. According to the PCA constructed with the SNPs, the most distant groups were the *inodorus* and the *conomon*, whereas in the AFLPs, both *conomon* and *agrestis* from Africa and Asia were the most divergent groups from the *inodorus* cultivars.

## Population structure

All SNP and AFLP data were used for the model-based clustering method as implemented in STRUCTURE. The delta *K* method suggested $K = 4$ as the best for the data set consisting of SNP markers, but it failed to identify the best *K* for the AFLP data set. The best replicate that gave maximum log likelihood values was obtained when *K* was set at four. The non-parametric analysis suggested $K = 5$ for both data sets, so we decided to accept this value as the best *K*. The accessions included in each of these groups are shown in Fig. 2a, b. This subpopulation division can also be observed in the neighbor-joining tree based on pair-wise genetic distances (Fig. 3).

Two distinguished populations that corresponded to the two predefined *inodorus* and *cantalupensis* groups were found within the subspecies *melo*. The *inodorus* group showed low levels of admixture (identified in black in Fig. 2a). The second population (orange in Fig. 2a), including all the *cantalupensis* accessions, also showed low admixture. The AFLP data also supported this differentiation, although they showed a higher level of admixture in the *inodorus* group (Fig. 2b). An even higher level of admixture was observed in the rest of the accessions belonging to the subspecies *melo*. The landraces from Eastern Europe, North Africa and Asia belonging to the varieties *inodorus* and *ameri* (morphologically close to the *cantalupensis* group) (Online Resource Fig. 1) shared genetic background with the *inodorus*, *cantalupensis*, and *momordica* populations. Similar admixture levels were found in the *flexuosus* landraces from Asia, including non-sweet melons consumed traditionally as cucumbers.

Furthermore, STRUCTURE analysis suggested the existence of at least three populations within the subspecies *agrestis* (red, blue, and green in Fig. 2a). These populations corresponded to Far Eastern *conomon* accessions, African *agrestis* and the *momordica* from India. The AFLPs supported the *conomon* and African *agrestis* populations, and showed significantly higher levels of admixture in the *momordica* group (Fig. 2b). Several subsp. *agrestis* accessions showed relatively high levels of admixture with both marker systems. The non-sweet genotypes belonging to *acidulus* and small-fruited *agrestis* (mostly from India) shared genetic background with *conomon*, *momordica*, and African *agrestis*. This combination also occurs in the Arya genotype. This is a very unique accession as it is morphologically similar to the *flexuosus* types (usually

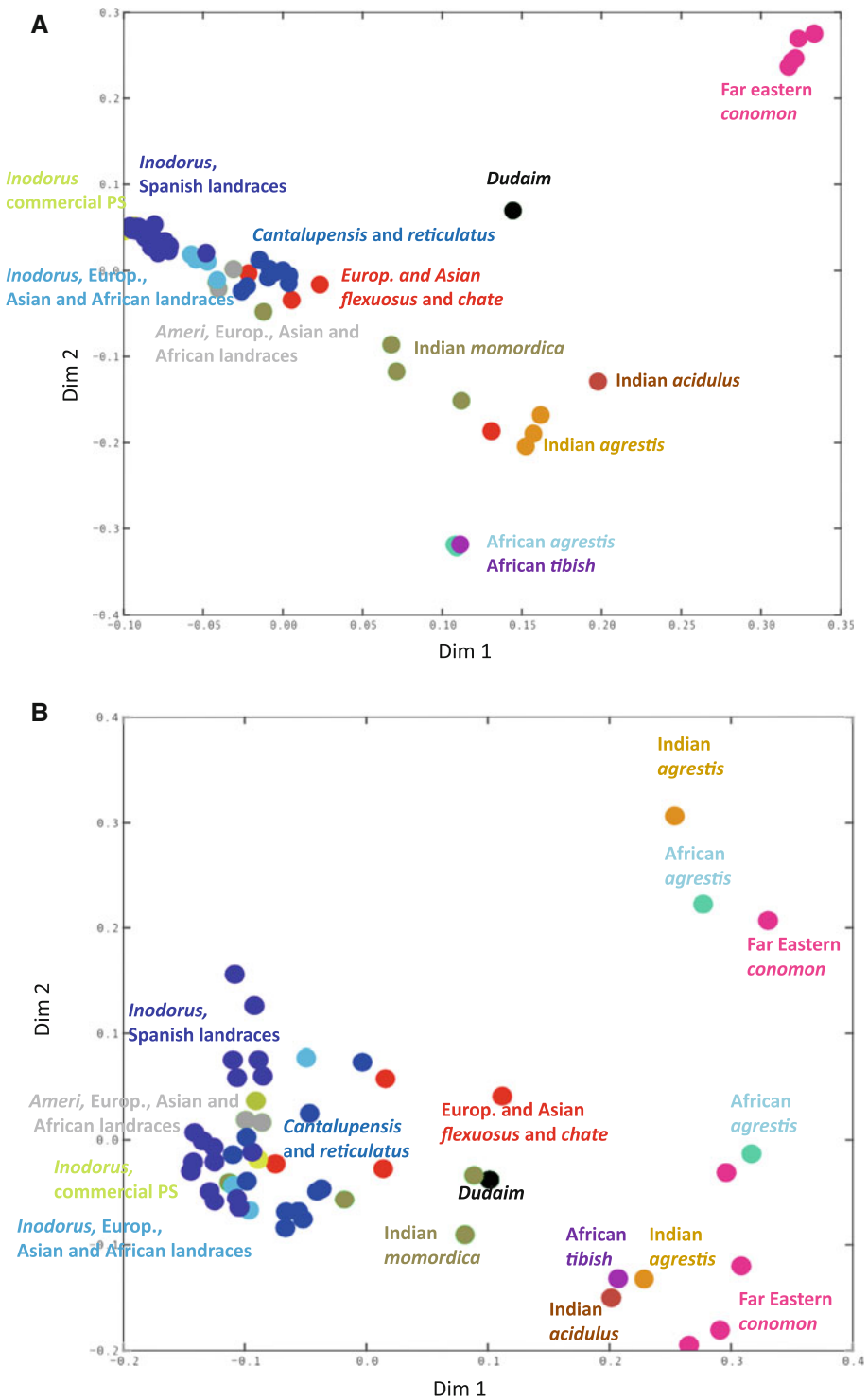



**Fig. 1** Principal Component Analysis (PCA) based on 628 SNP markers distributed throughout the melon genome (**a**) and Principal Coordinate Analysis (PCoA) based on 190 AFLP polymorphic *loci* (**b**). Cultivar groups are the same as those described in Table 3 (color figure online)

classified within the subspecies *melo*) (Online Resource Fig. 1) but is molecularly similar to the *agrestis* subspecies. The only analyzed *dudaim* accession, which is quite aromatic and used as an ornamental, shared genetic background with four of the five differentiated populations: *inodorus*, *conomon*, *momordica*, and *cantalupensis*.

Spanish *inodorus* landraces

A set of 284 SNPs (Table 3), distributed throughout the genome, were also polymorphic within the analyzed group of Spanish landraces. 'Blanco' and 'Tendral' melons were more variable than 'Amarillo' and 'Piel de Sapo' (Table 4).

Cluster analysis conducted using only the *inodorus* varieties, both Spanish and Eurasian references, showed a certain degree of cultivar grouping (Fig. 4). Commercial 'Piel de Sapo' grouped with 'Piel de Sapo' landraces. 'Amarillo' and 'Blanco' types were similar, forming two groups with accessions of the two classes, and the two 'Rochet' types were separated from the other groups. A group of 'Tendrals' was the most similar to the Eurasian landraces, mainly those from Madrid that are also called Negros (Black Tendral). Some unclassified accessions were similar to others included in well-defined groups. For example, the 'Loperano' landrace was quite similar to the Amarillo type 'Caña dulce', both of which are from Jaén, and the 'Loperos' melon from Córdoba was similar to the 'Blanco Escrito' collected in Cádiz. Also, the landrace 'Coca' is similar to 'Blanco Tempranillo', both of which are from Jaén. We can conclude that geographical origin was also important and that Andalusian melons are more similar due to their origin area than to their morphological class. All landraces were distinguished with this set of markers (Online Resource Table 3). In Online Resource Fig 2, a unique genotype graphic for each of these landraces is represented.

## Candidate *loci* under selection

The position of each of the 628 polymorphic SNPs in the melon genome and location of the 602 mapped SNPs in the genetic map is detailed in Online Resource Tables 2 and 3 (60, 30, 56, 69, 34, 57, 49, 52, 57, 53, 48, and 37 SNPs were located in linkage groups (LG) from I to XII, respectively—this information is also available at http://melonomics.net/). There were no differences in the average PIC among the SNPs located in the different LG in the whole germplasm collection (Fig. 5a, b). When analyzed by group, the *cantalupensis* and *momordica* shared a significant increase in PIC in LG VII, which is not observed in the other groups. There are several regions in LG VII that were fixed in the *inodorus* cultivars and were highly variable in *cantalupensis* and *momordica* (e.g., from 19.3 to 27.3 cM, and from 44.9 to 59.9 cM—marked in yellow in Online Resource Tables 2 and 3). The variability of the SNPs along each LG was also graphically represented for each genotype (Online Resource Fig. 2). In this analysis, we can also see several regions that seem to have been fixed for contrasting alleles in the two main cultivated groups, *inodorus* and *cantalupensis*.

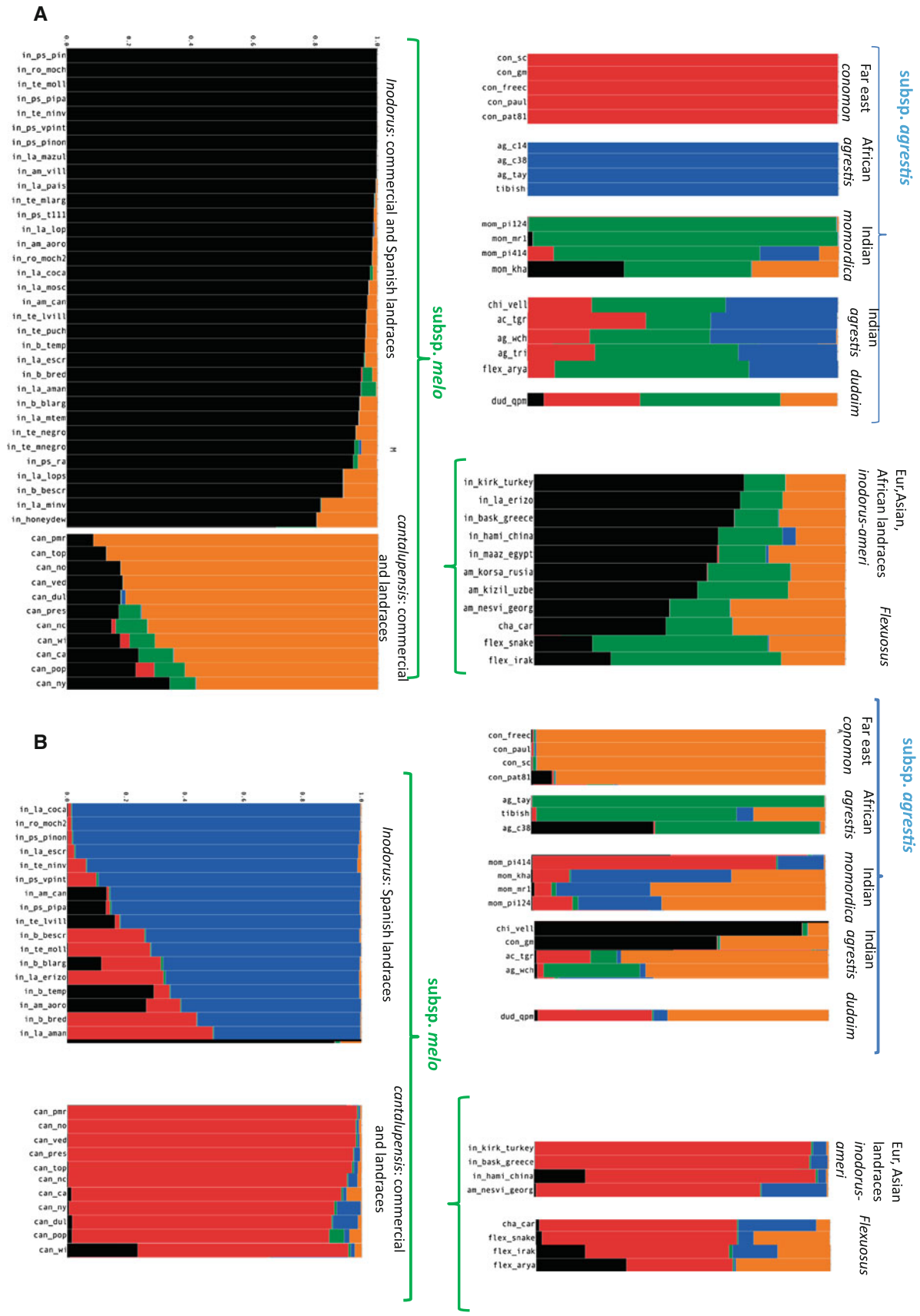
We conducted further analyses to identify candidates for *loci* that are under selection between these two major groups (*inodorus* and *cantalupensis*). The genetic structure of the populations was analyzed using the Wright *F* parameters, Fis, Fit, and Fst. These parameters were calculated for each of the 365 markers that were polymorphic in these two sets of accessions (Online Resource Table 3). There was a remarkably high degree of fixation (high Fis values) in each population. We observed considerable variation for Fst within this set of *loci*. A number of *loci* were identified with Fst $\sim$ 0 (104 *loci* with Fst $< 0.1$), indicating similar allelic frequencies in both populations. However, most of the Fst values revealed a large divergence between these two groups (ranging from 0.1 to 0.95, with an average Fst of 0.52). We found, for example, a number of *loci* with high levels of Fst (>0.75) (marked in purple in Online Resource Tables 2 and 3). For these *loci*, the annotation of the corresponding unigenes and fixed or more frequent alleles in each population are listed in Online Resource Table 4.

## LD analysis within *inodurus* and *cantalupensis* accessions

We followed a sequential approach in order to choose an appropriate set of SNPs for the LD analysis (see "Materials and methods" for a detailed description). In the first step, a total of 47 informative SNPs (MAF $< 0.95$) were identified in eight selected amplicons from the contig 60K17-I3J4 (van Leeuwen et al. 2003; Deleu et al. 2007) in a set of 40 accessions (Table 1, Online Resource Table 1). These SNPs map at scaffold00003 of the genomic sequence of melon (http://melonomics.net; Garcia-Mas et al. 2012). Sixty-six percent of these SNPs separated within <12 Kb, and the rest from 60 to 163 Kb. In general, LD decayed rapidly within 1 Kb and dropped to very low levels after 4 Kb (Online Resource Fig. 3a), although some differences were observed between amplicons.

Based on these results, 10 Kb from two additional genome sequences (scaffold00011 and scaffold00016) were sequenced in a subset of 17 accessions in order to define more SNPs and to do a preliminary check of the extent of LD in different genomic regions (Online Resource Table 1). A total of 71 new informative SNPs were identified. The second preliminary LD analysis showed that LD also decayed rapidly within 1 Kb (Online Resource Fig. 3b), whereas a very low proportion of SNPs showed strong LD up to 8 Kb away.

To perform the definitive LD analysis in a larger germplasm sample, 62 SNPs that were evenly distributed within a 10-Kb region from those three genomic regions were included in the previously defined GoldenGate array, and the 74 accessions used for the genetic diversity analysis were genotyped. Given that the genomic position of the rest of the SNPs of the GoldenGate array was also known, the SNPs that were separated by <20 Kb were also included in the LD analysis. SNPs located in different chromosomes (Online Resource Table 3) were used to calculate the background LD between unlinked SNPs.


A
in_ps_pin
in_ro_moch
in_te_moll
in_ps_pipa
in_te_ninv
in_ps_vpint
in_ps_pinon
in_la_mazul
in_am_vill
in_la_pais
in_te_mlarg
in_ps_tlll
in_la_lop
in_am_aoro
in_ro_moch2
in_la_coca
in_la_mosc
in_am_can
in_te_lvill
in_te_puch
in_b_temp
in_la_escr
in_b_bred
in_la_aman
in_b_blarg
in_la_mtem
in_te_negro
in_te_mnegro
in_ps_ra
in_la_lops
in_b_bescr
in_la_minv
in_honeydew
Inodorus: commercial and Spanish landraces
can_pmr
can_top
can_nc
can_ved
can_dul
can_pres
can_nc
can_wi
can_ca
can_pop
can_ny
cantalupensis: commercial and landraces
subsp. melo
con_sc
con_gm
con_freec
con_paul
con_pat81
ag_c14
ag_c38
ag_tay
tibish
mom_pi124
mom_mr1
mom_pi414
mon_kha
chi_vell
ac_tgr
ag_wch
ag_tri
flex_arya
dud_qpm
Far east conomon
African agrestis
Indian momordica
Indian agrestis dudaim
subsp. agrestis
in_kirk_turkey
in_la_erizo
in_bask_greece
in_hami_china
in_maaz_egypt
am_korsa_rusia
am_kizil_uzbe
am_nesvi_georg
cha_car
flex_snake
flex_irak
Eur,Asian, African landraces inodorus-ameri
Flexuosus
B
in_la_coca
in_ro_moch2
in_ps_pinon
in_la_escr
in_te_ninv
in_ps_vpint
in_am_can
in_ps_pipa
in_te_lvill
in_b_bescr
in_te_moll
in_b_blarg
in_la_erizo
in_b_temp
in_am_aoro
in_b_bred
in_la_aman
Inodorus: Spanish landraces
can_pmr
can_no
can_ved
can_pres
can_top
can_nc
can_ca
can_ny
can_dul
can_pop
can_wi
cantalupensis: commercial and landraces
subsp. melo
con_freec
con_paul
con_sc
con_pat81
ag_tay
tibish
ag_c38
mom_pi414
mom_kha
mom_mr1
mom_pi124
chi_vell
con_gm
ac_tgr
ag_wch
dud_qpm
Far east conomon
African agrestis
Indian momordica
Indian agrestis dudaim
subsp. agrestis
in_kirk_turkey
in_bask_greece
in_hami_china
am_nesvi_georg
cha_car
flex_snake
flex_irak
flex_arya
Eur, Asian landraces inodorus-ameri
Flexuosus

◀ **Fig. 2** Populations identified with STRUCTURE in the germplasm collection analyzed with SNPs (**a**) and with AFLPs (**b**). Inferred population structure in a collection of melon accessions (commercial cultivars, landraces, exotic types, and wild genotypes) using the model-based program STRUCTURE (Pritchard et al. 2000). Results shown are for $K = 5$. Each accession's genome is represented by a single *horizontal line*, which is partitioned into colored segments in proportion to the estimated membership in the five subpopulations (color figure online)

Due to the strong population structure in melon germplasm (Fig. 2a, b), the *inodorus* and *cantalupensis* accessions were analyzed separately. The SNPs located in scaffold00016 were not polymorphic within *inodorus*, and only two haplotypes were observed within *cantalupensis*. As a result, these SNPs were discarded for the *inodorus* LD analysis and two analyses were performed for *cantalupensis*, one with and one without these SNPs. The decay of LD over genetic distances for the three analyses is depicted in Fig. 6. LD within *inodorus* decayed up to 6 Kb away, being on average high ($r^2 > 0.5$) and highly significant ($p < 0.01$) at less than 3 Kb, but low ($r^2 \approx 0.3$) and residually significant ($p < 0.05$) at distances higher than 6 Kb. This result contrasted with the preliminary LD analysis, probably due to the broader genetic diversity of the germplasm used before (which included not only *i-nodorus* but also *cantalupensis* and related accessions). LD decay within *cantalupensis* depended strongly on the inclusion of the SNPs located in scaffold00016, meaning that LD was above the background $r^2$ ($p < 0.05$) when scaffold00016 SNPs were included for distances larger than 10 Kb, but was below the background $r^2$ even for distances below 1 Kb when those SNPs were discarded. A specific genetic phenomenon, out of the scope of the current research, is most likely underlying the unexpected low genetic variability and high LD at scaffold00016. Thus, discarding the scaffold00016 region, LD within *cantalupensis* may be considered to be extremely low.

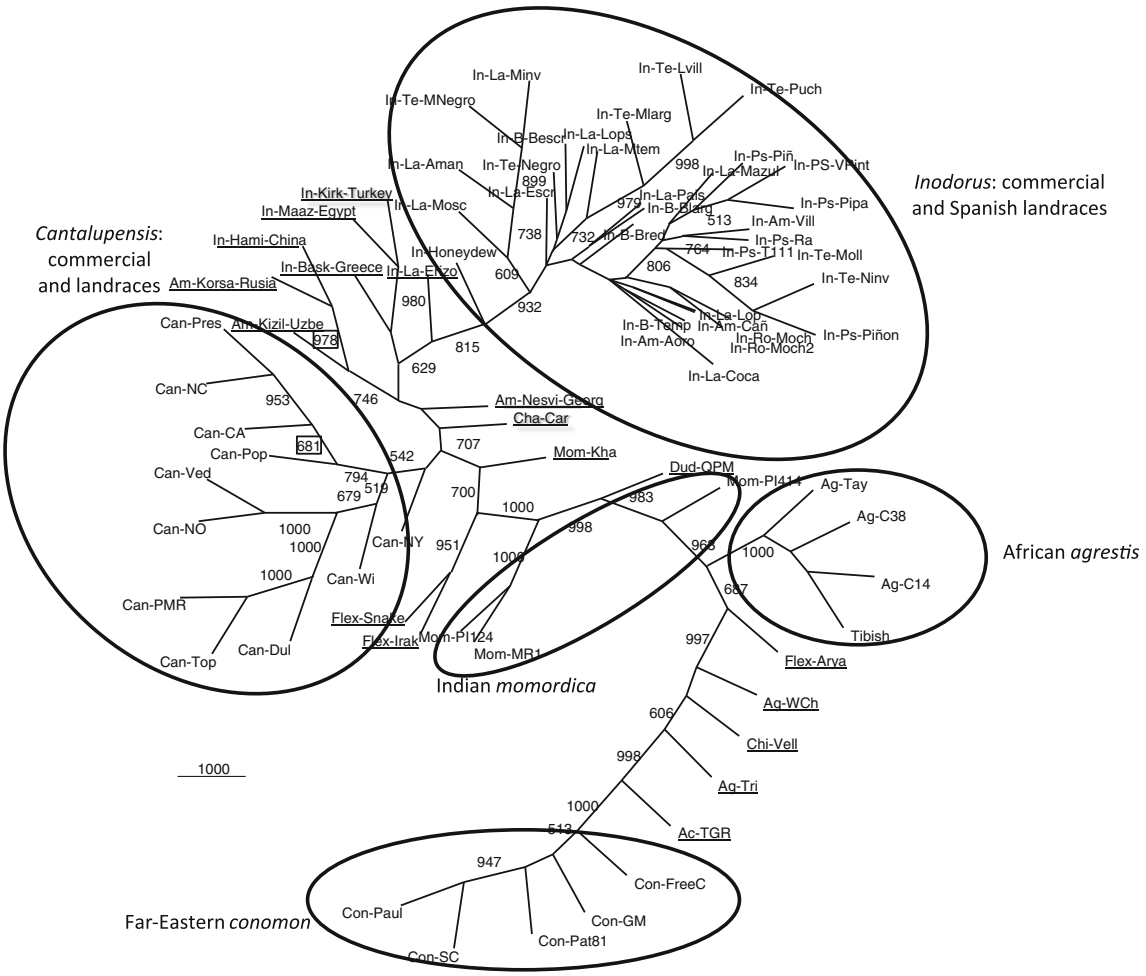



**Fig. 3** NJ tree constructed with SNP results. Accessions included in each of the five populations defined with STRUCTURE are indicated. *Underlined* are the accessions with high levels of admixture

**Table 4** Number (percentage) of polymorphic SNPs in the different groups of Spanish landraces analyzed in this study and average MAF

| Markers | Piel de sapo | Amarillo | Blanco | Tendral | Other |
|---|---|---|---|---|---|
| SNPs | 46 (16.2 %) 0.69 | 114 (40.1 %) 0.7 | 128 (45.1 %) 0.7 | 113 (39.8 %) 0.71 | 258 (90.8 %) 0.82 |
| AFLPs | 25 (13.2 %) | 21 (11.1 %) | 40 (21.05 %) | 25 (13.2 %) | 32 (16.8 %) |

Degree of polymorphism obtained with AFLPs (percentage)

Percentages are indicated over the polymorphic SNPs in the *Inodorus* Spanish landraces

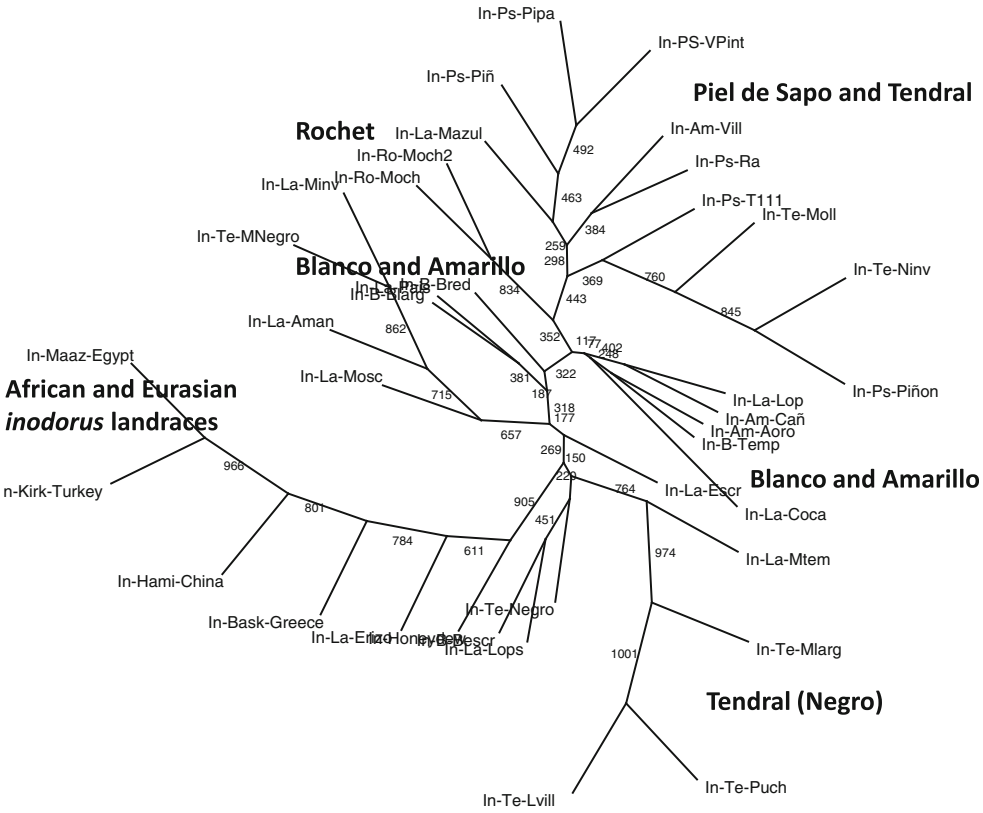



**Fig. 4** NJ tree constructed with SNP results in Spanish landraces. The main traditional Spanish types are indicated and compared to other *inodorus*, African and Eurasian landraces

## Discussion

High-throughput SNP genotyping using a GoldenGate platform has provided a means to study the genetic diversity, the population structure, and the LD in melons for the first time. The high validation rate found for the selected SNPs confirms the quality of the in silico identification and of the filtering process performed previously. Similar validation values have been reported in other GoldenGate platforms generated in soybean, barley, maize, pea, and squash (Rostoks et al. 2006; Hyten et al. 2008; Yan et al. 2010; Blanca et al. 2012). The results also confirm the importance of the in silico selection of SNPs with several reads per allele in order to increase the validation percentage. Other sequencing technologies that provide higher coverage than Sanger or 454, such as Illumina or SOLID, result in a larger number of reads per allele. These are quite useful for true SNP detection (Blanca et al. 2012). However, they require a transcriptome and/or genome of reference to facilitate the mapping process of their shorter reads.

This is the largest set to date of SNPs validated for genetic diversity studies in melon (Deleu et al. 2009; Blanca et al. 2011, 2012; Clepet et al. 2011), and the first to be implemented in a high-throughput genotyping array. Knowing the polymorphism of the selected SNPs in this diverse germplasm collection will help to determine their usefulness in future genetic diversity studies or mapping efforts, especially when some of the analyzed accessions are parents of different mapping populations (Harel-Beja et al. 2010; Yuste-Lisbona et al. 2010).

Different types of molecular markers had been used previously to characterize elite melon germplasm (commercial cultivars, hybrids, and breeding lines, mainly from the USA and European markets), mostly belonging to the *inodorus* and *cantalupensis* groups. These studies usually

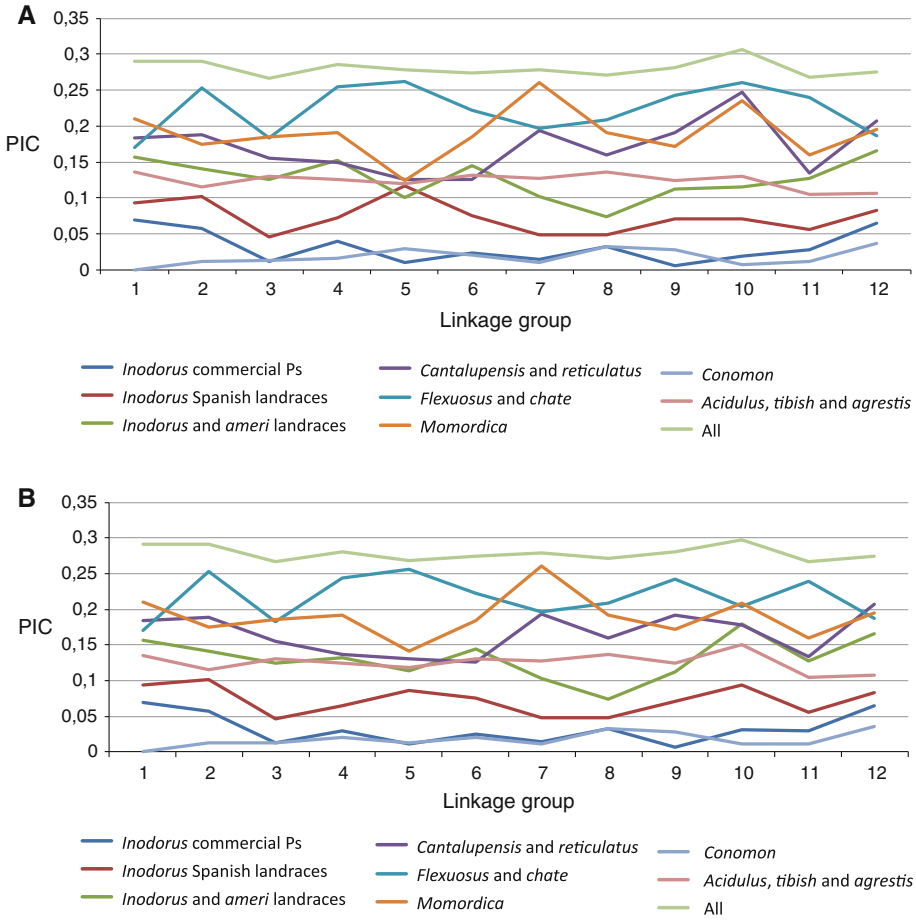



**Fig. 5** Average PIC of the SNPs located in each LG. Including all SNPs (**a**). Discarding the LD SNPs (**b**) (color figure online)

report a limited genetic diversity in some groups (García et al. 1998; Staub et al. 2000). In our study, we found a low variability between the commercial *inodorus* 'Piel de Sapo' accessions, whereas the *cantalupensis* group is highly variable. Despite this reduced variability, a set of 57 SNPs that were evenly distributed in the genome were able to distinguish the three commercial 'Piel de Sapo' and could therefore be useful for fingerprinting cultivars of this valuable market class. The group of *inodorus* Spanish landraces still retains a substantial degree of variability. Most of the SNPs found to be polymorphic in the commercial 'Piel de Sapo' group are also polymorphic in the group of Spanish 'Piel de Sapo' landraces and/or in the entire *inodorus* landraces group, except for a group of nine SNPs (marked in light green in Online Resource Tables 2 and 3). These may represent introgressions from other groups, possibly including *agrestis* types, during the breeding process to which the landraces have not been subjected. One example is the marker *CmeIF4EB*, located in the gene that codifies for the eukaryotic translation initiation factor 4E-1, which confers resistance to viruses (Nieto et al. 2007) and has been introgressed into some commercial 'Piel de Sapo' cultivars from the SC *conomon* accession. Similar results have been reported in a recent study employing an Infinium SNP array in tomato (Sim et al. 2012) in which a differential allelic variation was found within different cultivar groups for several disease-resistance genes (conferring resistance to bacteria, fungi, and viruses), reflecting distinct introgression histories.

The snake and snap melons (*flexuosus* and *momordica*, non-sweet and low-sugar-content accessions) from the Near and Middle East and India were even more variable than the *acidulus, tibish* and wild small-fruited *agrestis* from Africa and India. These results, with the *cantalupensis* and *momordica* groups being the most variable of subsp. *melo* and *agrestis,* respectively, are consistent with the results presented by Blanca et al. (2012) after resequencing pools of genotypes belonging to the different groups. The low variability found within the *conomon* group also agrees with the resequencing results of Blanca et al. (2012). Other previous studies with different types of molecular markers (Tanaka et al. 2007) showed less variability within this group than in other wild and exotic groups. However, in our data, the *conomon* group is even less variable than the *inodorus* group of commercial cultivars. This unexpectedly low variability may be due to the fact that most of the SNPs were selected in silico for being monomorphic within a *conomon* variety. According to the

AFLPs, the *conomon* group was also less variable than the African and Indian *agrestis*, but its variability was similar to that found in *inodorus* or *cantalupensis*. Therefore, in this set of SNPs, the variability confined to the *conomon* group was underestimated.

Multivariate analyses with SNPs and AFLPs supported the subspecific separation and the intermediate position of *momordica*, *flexuosus*, and *dudaim* demonstrated by many previous studies (Silberstein et al. 1999; Monforte et al. 2003; Nakata et al. 2005; Deleu et al. 2009; Esteras et al. 2009; Blanca et al. 2011). However, although according to the SNP genotyping the most distant groups were the *inodorus* and the *conomon*, the AFLPs indicated that both *conomon* and *agrestis* from Africa and India were the more divergent groups from *inodorus*, which is more consistent with previous studies (Blanca et al. 2012). The SNP sampling process, in which those SNPs that were polymorphic between T111 and SC were selected, likely overestimates the genetic distance between these two groups. Despite this bias, which may have affected the study of the *inodorus*–*conomon* relationships, this set of markers is the largest one employed to date in melon and has been used successfully to analyze the population structure of this germplasm collection, which is representative of the entire species variability.

The population structure analysis suggested the existence of five subpopulations, two in subspecies *melo* and three in subspecies *agrestis*. In subspecies *melo*, the clear differentiation between the *inodorus* and *cantalupensis* groups differs from most previous studies, which show an ambiguous differentiation of both groups (Stepansky et al. 1999; Esteras et al. 2009). The resolution of the current results is most likely due to the large number of markers used. These two populations show distinct vine traits, flowering traits, and fruit morphologies. *Inodorus* yields fruits that are typically non-aromatic, non-climacteric, and that have a high sugar content, whereas *cantalupensis* fruits are typically aromatic, climacteric, and medium-sugar.

The high levels of admixture found in subsp. *melo* landraces from Eastern Europe, Asia and North Africa (sweet and non-sweet melons, *inodorus*, *ameri*, and *flexuosus*) is coherent with the highest levels of variability reported previously in this area. Central Asia and the Near and Middle East have been reported as the origin area for

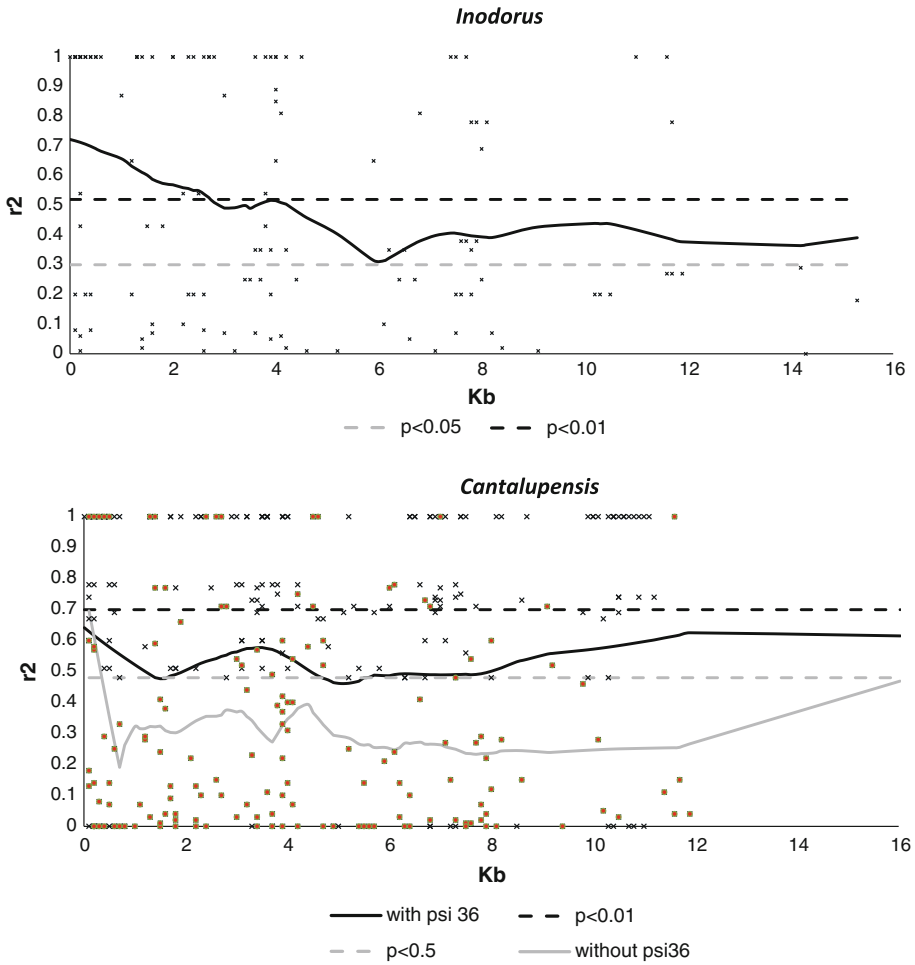



**Fig. 6** Linkage disequilibrium ($r^2$) versus physical distance (Kb) between *inodorus* and *cantalupensis* accessions and cultivars. The false discovery rates $p < 0.05$ and $p < 0.01$ are indicated with *black* and *grey dashed lines*, respectively. Curves were fitted by second-degree LOESS. For the *cantalupensis* group, the *grey curve* was calculated removing the SNPs included in scaffold00016 (color figure online)

the *inodorus* group, which was important for the diffusion of these types to Europe. Also, it is thought that *cantalupensis* melons spread to Europe from the eastern part of Turkey (Zhukovsky 1951). A recent extensive search of literature from the Roman and medieval periods suggests that sweet melons were present in Central Asia in the mid-ninth century and in Persia by the mid-tenth century, whereas the first reports of melons in Andalusia date from the second half of the eleventh century (Paris et al. 2012). These melons probably arrived from Central Asia as a consequence of Islamic conquest, trade and agricultural development. Our results are coherent with this idea of Eastern European melons and *cantalupensis* having been derived from these Eurasian varieties.

In the subspecies *agrestis*, all the *conomon* accessions, which are cultivated in Far Eastern countries, formed a subpopulation with low levels of admixture even though they were variable in fruit characteristics, with sweet and non-sweet, climacteric and non-climacteric fruits, and a wide range of shapes and colors (Online Resource Fig. 1). The *momordica* group also formed a subpopulation, but was the most heterogeneous. The results are consistent with the high variability and intermediate position of the *momordica* group reported in previous studies (Staub et al. 2004; Luan et al. 2008; Dhillon et al. 2007, 2009; Nhi et al. 2010). These studies suggested that snap melon landraces from Northern India might represent a central melon origin area from which oriental and occidental melon germplasm developed. This idea is also coherent with results obtained by resequencing melon transcriptomes (Blanca et al. 2012), where the *momordica* group was the one with the highest percentage of SNPs in common with other botanical groups of both subspecies, grown in both extremes of the geographical distribution of melon cultivation.

Additionally, all of the African accessions formed a separate group, which included small-fruited wild African melons and one accession of the *tibish* group (Online Resource Fig. 1). This is a type with medium-size fruits that is cultivated in Sudan, consumed as a cucumber, and is considered to be one of the most primitive forms of melon. The fact that it is not differentiated from wild African melons suggests that this variety was domesticated from African and not from wild Indian types. Africa has long been considered the center of origin of melons, but recent studies suggest that *C. melo* may have originated somewhere in Asia and then reached Africa from there (Renner et al. 2007). In fact, wild melons are very frequent in East and West Africa, and also from Central Asia to India (Whitaker and Bemis 1976; McCreight and Staub 1993; Rubatzky and Yamaguchi 1997). It is also possible that different domestication events took place in Africa and Asia (Bates and Robinson 1995), the most extensive of which may have occurred in Asia, generating more edible types. The few studies focused on African germplasm (Mliki et al. 2001; Akashi et al. 2002) suggested independent introductions from African germplasm into India and the Middle East. According to our results, African accessions are differentiated from Indian *momordica*, although they share genetic background with other genotypes from India (small-sized *agrestis* and *flexuosus*), which prevented us from distinguishing between the two hypotheses: that African melons come from Indian melons and/or that African melons reached India and crossed with Indian melons increasing the genetic diversity in this region.

Taking everything together, the presented results agree with a high diversity of non-sweet and low-sugar melons from India (*agrestis, acidulus*, and *momordica*). The variability existing in this region, with a likely contribution of African germplasm, might have given rise to the exotic *conomon*. Also, the high diversity of sweet and non-sweet melons from Central Asia and the Near and Middle East may have originated in this region, leading to the high diversity of *cantalupensis* and *inodorus* landraces still found in secondary centers of diversification, such as Spain.

In our study, Spanish landraces differentiated from the reference set of Eurasian *inodorus.* Spain is the fifth melon producer worldwide (the first in the EU) and the first world exporter (FAO 2012). There is a long tradition of melon cultivation in Spain. The most traditional area is that of central Spain, in which melons are cultivated extensively. This cultivation was based on the use of *inodorus* landraces selected over centuries by farmers (Esteras et al. 2012b). The main commercial *inodorus* type in the Spanish and international markets is the 'Piel de Sapo' class. However, other major classes, such as 'Amarillo', 'Blanco', 'Rochet', and 'Tendral', as well as several other less known types, still survive as landraces that appeal to local markets. Currently, most of the production of this landrace is being replaced by that of commercial hybrids, which are highly productive with short growing cycles and uniform fruits, belonging to the 'Piel de Sapo' market class. Morphological variation surveys of Spanish landraces have been carried out by several authors, confirming their distinctive morphological and sensorial characteristics (Gómez-Guillamón et al. 1998; Nuez et al. 1988; López-Sesé et al. 2003; Escribano et al. 2012). Our molecular results agree with those obtained in the only two previous studies in which the molecular diversity of Spanish melon germplasm was tested. In these two studies, conducted with RAPDs and SSRs (López-Sesé et al. 2003; Escribano et al. 2012), Spanish landraces were also different from those included as references, which suggests a lack of gene introgressions from other germplasm of distinct origin. One of these studies focused on traditional melons from Madrid, 'Black Tendral' landraces, such as 'Largo', 'Largo Negro' and

'Puchero', which were markedly different from all the other accessions examined (Escribano et al. 2012). Our results also confirm the singularity of this group of 'Tendrals', which were the more similar to Eurasian landraces.

The set of markers studied in the current report will be putatively useful for wider genetic diversity studies and for further mapping purposes, especially due to their even distribution in the melon genome. The study of the PIC variation in the SNPs throughout the different chromosomes revealed the existence of several regions with different variation patterns between subpopulations. Similar results have recently been reported in tomato, where a genome-scale SNP analysis has revealed genomic regions with high genetic variation between populations, suggesting that historical breeding practices have led to different patterns of genetic variation in cultivated tomato germplasm (Sim et al. 2012). In melons, one region in LGVII detected in this study, which is highly variable in *cantalupensis* and *momordica* and almost fixed in *inodorus,* contains several QTLs associated with fruit shape, size, flesh firmness, color, and soluble-solids content (Díaz et al. 2011). Some of the differential SNPs found in this region are located in melon unigenes with functions related to some of these traits, as is the case of SNP Cmpsnp978, located in unigene MELO3C017772, which codifies for a Phytoene Desaturase, which is an early enzyme of the carotenoid biosynthetic pathway. Most *inodorus* landraces have green or white flesh, whereas this trait is highly variable in *cantalupensis* and *momordica*, which range from green to different shades of orange. Melon flesh color is supposed to be controlled by two major genes, *gf* and *wf*, which interact epistatically and have been mapped at LG8 and LG9, respectively (Díaz et al. 2011). Different combinations of these two genes produce orange, green, and white fruit flesh. However, several QTLs involved in flesh color intensity variation have been reportedly mapped in LG3, 6, and 7 (Fernández-Trujillo et al. 2011).

By analyzing the genetic structure of populations using the Wright *F* parameters, specific regions of the genome that appear to be under selection were also highlighted, thereby distinguishing the two main groups of melons, *inodorus* and *cantalupensis* (Online Resource Table 4). The degree of gene differentiation found with Fst between *inodorus* and *cantalupensis* was wider than that reported for different groups of cultivated tomatoes, including processing, fresh market, and vintage accessions (Sim et al. 2012). Some of the SNPs with the highest Fst values were in LG2, in a region in which a QTL for powdery mildew resistance maps, but mostly were located in two regions of LG6 (32–38 and 70–85 cM). These are the same regions in which the *Cm-ETR2*, a subfamily II ethylene receptor homolog from melon, which is involved in fruit development and ripening, and the *Cm-ERF3*, an ethylene-responsive element-binding factor 3, have been mapped (Díaz et al. 2011). The genes of the first region were fixed in *inodorus* and variable in *cantalupensis*, whereas those located in the second region were fixed in *cantalupensis* and variable in *inodorus*. There were also several fixed markers dispersed throughout LG8 (in two regions in which *Cm-ACO3*, 1-aminocyclopropane-1-carboxylate oxidase 3, which is involved in ethylene synthesis and expressed in flowers, and a gene controlling flesh pH, have been located), in LG9 (in the region in which *Cm-ERF1* and *Cm-PEM3*, a pectin methyl esterase involved in fruit softening, map), in LG11 (in the region in which *Cm-ACS1* maps, the ACC synthase 1, which is wound-responsive and most likely plays a key role in the increased production of ethylene at the ripening stage) and in LG12 (in the regions where QTLs controlling powdery mildew resistance, net cover, and striped fruits have been located). The *loci* in these regions may represent candidates for functional analysis in order to identify the *loci* underlying phenotypic differences between these varietal groups. The fact that the selected regions include genes involved in the ethylene pathway and are related to flowering and fruit ripening is quite consistent with the plant and fruit characteristics of both groups. *Cantalupensis* are early flowering, with more female tendency and are sometimes monoecious and fully climacteric, whereas *inodorus* are considered to be late-flowering, mostly andromonoecious and non-climacteric, although some *inodorus* landraces show an intermediate behavior, with slight aroma and flesh softening, such as certain 'Blanco' and 'Rochet' types, or the 'Erizo' and 'Escrito' melons. Also, the differentiation in regions with resistance genes is consistent with the breeding history of these cultivars, as many *inodorus* landraces have not been subjected to breeding processes, and have therefore been less introgressed with resistance genes coming from *momordica*, *conomon*, or *agrestis* sources than the *cantalupensis* cultivars.

Even though a higher density SNP array should be analyzed to draw conclusions about genome-wide LD, the current results suggest that LD is expected to be low (in the range of a few Kb) when analyzing a broad sample of melon accessions. Some variability may be observed in specific genomic regions or depending on the germplasm. For a more thorough analysis, a density of one SNP every Kb or lower would be necessary to ensure the detection of LD in a genome-wide experiment. The extent of LD in melon is, therefore, much lower than in other species, such as tomato (van Berloo et al. 2008; Robbins et al. 2011), wheat (Breseghello and Sorrells 2006), peach (Aranzana et al. 2010), and barley (Comadran et al. 2011), where the extent of LD expands through higher genetic distances, even lower than in rice [where it is in the range of several hundreds of Kb (McNally et al. 2009; Zhao et al. 2010)]. On the other hand, the current results are within the range of LD found in maize, where LD generally decays rapidly within 1.5 Kb,

but can be larger in specific genomic regions or germplasm collections (Remington et al. 2001; Ching et al. 2002). A large proportion of melon varieties are monoecious. The andromonoecious cultivars also need external pollinators to fertilize the female flower, as the pollen produced by the hermaphrodite flower is not enough the ensure a high fruit set, and, concomitantly high seed production. Therefore, melon may be considered an outcrossing species. Also, melon is an Old World species, and was probably one of the most ancient crops cultivated in the Mediterranean region (Janick et al. 2007; Paris et al. 2012). Furthermore, as was discussed above, molecular diversity studies have shown that a large genetic variability in the primary centers of diversity (Mliki et al. 2001; Monforte et al. 2003; Dhillon et al. 2009; Roy et al. 2012; Fergany et al. 2011) and an intensive subsequent diversification have resulted in a large number of different cultivated botanical groups (Pitrat 2008). Due to the outcrossing reproduction system of this species, important intercrossing could have occurred during cultivar development by traditional farmers, which would have consequently reduced the degree of LD. Therefore, the low extent of LD in melon observed in the current report is in line with the reproductive system of the species and the history of melon cultivation.

**Acknowledgments** This project was carried out in the frame of the MELONOMICS project (2009–2012) of the Fundación Genoma España and with the contributions of the PLAT KKBE project PIM2010PKB-00691.

## References

Akashi Y, Fukunda N, Wako T, Masuda M, Kato K (2002) Genetic variation and phylogenetic relationships in East and South Asian melons, *Cucumis melo* L., based on the analysis of five isozymes. Euphytica 125:385–396

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119(3):507–517

Aranzana MJ, Abbassi el-K, Howad W, Arús P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genet 11:69

Bates DM, Robinson RW (1995) Cucumbers, melons and watermelons. In: Smartt J, Simmonds NW (eds) Evolution of Crop Plants, 2nd edn. Longman Scientific, Harlow, pp 89–96

Blanca J, Cañizares J, Ziarsolo P, Esteras C, Mir G, Nuez F, Garcia-Mas J, Picó B (2011) Melon transcriptome characterization. SSRs and SNPs discovery for high throughput genotyping across the species. Plant Genome 4(2):118–131

Blanca J, Esteras C, Ziarsolo P, Perez D, Fernández V, Collado C, Rodriguez R, Ballester A, Roig C, Cañizares J, Pico B (2012) Transcriptome sequencing for SNP discovery across *Cucumis melo*. BMC Genomics 13:280. doi:10.1186/1471-2164-13-280

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM (2007) Yogesh Ramdoss and Edward S. BucklerTASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Chancerel E, Lepoittevin C, Le Provost G et al (2011) Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. BMC Genomics 12:368

Ching A, Caldwell KS, Jung M, Dolan M, Smith OSH, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19. doi:10.1186/1471-2156-3-19

Choudhary S, Gaur R, Gupta S, Bhatia S (2012) Derived genic molecular markers: development and utilization for generating an advanced transcript map of chickpea. Theor Appl Genet 124(8):1449–1462

Clepet C, Joobeur T, Zheng Y, Jublot D, Huang M, Truniger V, Boualem A, Hernandez-Gonzalez ME, Dolcet-Sanjuan R, Portnoy V, Mascarell-Creus A, Caño-Delgado A, Katzir N, Bendahmane A, Giovannoni JJ, Aranda MA, Garcia-Mas J, Fei Z (2011) Analysis of expressed sequence tags generated from full-length enriched cDNA libraries of melon. BMC Genomics 12:252

Comadran J, Ramsay L, MacKenzie K, Hayes P, Close TJ, Muehlbauer G, Stein N, Waugh R (2011) Patterns of polymorphism and linkage disequilibrium in cultivated barley. Theor Appl Genet 122(3):523–531

Deleu W, González V, Monfort A, Bendahmane A, Puigdomènech P, Arús P, Garcia-Mas J (2007) Structure of two melon regions reveals high microsynteny with sequenced plant species. Mol Genet Genomics 278(6):611–622

Deleu W, Esteras C, Roig C, Gonzalez-To M, Fernandez-Silva I, Gonzalez-Ibeas D, Blanca J, Aranda MA, Arus P, Nuez F, Monforte AJ, Picó MB, Garcia-Mas J (2009) A set of EST-SNPs for map saturation and cultivar identification in melon. BMC Plant Biol 9:90

Dhillon NPS, Ranjana R, Singh K, Eduardo I, Monforte AJ, Pitrat M, Dhillon NK, Singh PP (2007) Diversity among landraces of Indian snapmelon (*Cucumis melo* var. *momordica*). Genet Resour Crop Evol 54:1267–1283

Dhillon NPS, Singh J, Fergany M, Monforte AJ, Sureja AK (2009) Phenotypic and molecular diversity among landraces of snapmelon (*Cucumis melo* var. *momordica*) adapted to the hot and humid tropics of eastern India. Plant Genet Resour: Charact Util 7(3):291–300. doi:10.1017/S1479262109990050

Díaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, Staub JE, Zalapa JE, Cuevas HE, Dace G, Oliver M, Boissot N, Dogimont C, Pitrat M, Hofstede R, Koert P, Harel-Beja R, Tzuri G, Portnoy V, Cohen S, Schaffer A, Katzir N, Xu Y, Zhang H, Fukino N, Matsumoto S, Garcia-Mas J, Monforte AJ (2011) A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (*Cucumis melo* L.). BMC Plant Biol 11:111. doi:10.1186/1471-2229-11-111

Escribano S, Lázaro A, Cuevas HE, López-Sesé AI, Staub JE (2012) Spanish melons (*Cucumis melo* L.) of the Madrid provenance: a unique germplasm reservoir. Genet Resour Crop Ev. doi:10.1007/s10722-011-9687-4

Esteras C, Lunn J, Sulpice R, Blanca J, Garcia-Mas J, Pitrat M, Nuez F, Picó B (2009) Phenotyping a highly diverse core collection of melons to be screened using EcoTILLING. Plant Gem Lisbon Plant Genomics European Meeting. p213

Esteras C, Gómez P, Monforte AJ, Blanca J, Vicente-Dólera N, Roig C, Nuez F, Picó B (2012a) High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. BMC Genomics 13:80. doi:10.1186/1471-2164-13-80

Esteras C, Nuez F, Picó B (2012b) Genetic diversity studies in Cucurbits using molecular tools. In: Wang Y, Behera TK, Kole

C (eds) Cucurbits: genetics, genomics and breeding of cucurbits. New Hampshire: Science Publishers Inc, Enfield

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

FAO (2012) FAOSTAT: http://faostat.fao.org/site/291/default.aspx. Accessed 4 January 2012

Fergany M, Kaur B, Monforte AJ, Pitrat M, Rys C et al (2011) Variation in melon (*Cucumis melo*) landraces adapted to the humid tropics of southern India. Genet Resour Crop Ev 58(2):225–243

Fernandez-Silva I, Eduardo I, Blanca J, Esteras C, Pico B, Nuez F, Arus P, Garcia-Mas J, Monforte AJ (2008) Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.). Theor Appl Genet 118:139–150

Fernández-Trujillo JP, Picó B, Garcia-Mas J, Álvarez JM, Monforte AJ (2011) Breeding for fruit quality in melon. In: Jenks MA, Bebeli P (eds) Breeding for fruit quality. Wiley-Blackwell, Ames Chapter 12

Fita A, Nuez F, Picó B (2011) Diversity in root architecture and response to P deficiency in seedlings of *Cucumis melo* L. Euphytica 181(3):323–339. doi:10.1007/s10681-011-0432-z

García E, Jamilena M, Álvarez JI, Arnedo T, Oliver JL, Lozano R (1998) Genetic relationships among melon breeding lines revealed by DNA markers and agronomic traits. Theor Appl Genet 96:878–885

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Cámara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutiérrez S, Blanca J, Cañizares J, Ziarsolo P, Gonzalez-Ibeas D, Rodríguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Melé M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez, Picó B, Gabaldón B, Roma G, Guigó R, Casacuberta JM, Arús P, Puigdomènech P (2012) The genome of melon (*Cucumis melo* L.). Genome amplification in the absence of recent duplication in an old widely cultivated species. PNAS, doi/10.1073/pnas.1205415109

Gómez-Guillamón ML, Sánchez F, Fernández-Muñoz R (1998) Caracterización de cultivares de melón. Actas de las Jornadas de Selección y Mejora de Plantas Hortícolas. Córdoba, Spain (in Spanish)

González-Ibeas D, Blanca J, Roig C, González-To M, Picó B, Truniger V, Gómez P, Deleu W, Caño-Delgado A, Arús P, Nuez F, Garcia-Mas J, Puigdomènech P, Aranda MA (2007) MELOGEN: an EST database for melon functional genomics. BMC Genomics 8:306

Graham IA, Besser K, Blumer S et al (2010) The genetic map of Artemisia annua L. identifies loci affecting yield of the antimalarial drug artemisinin. Science 327(5963):328–331. doi:10.1126/science.1182612

Harel-Beja R, Tzuri G, Portnoy V, Lotan-Pompan M, Lev S, Cohen S, Dai N, Yeselson L, Meir A, Libhaber SE, Avisar E, Melame T, van Koert P, Verbakel H, Hofstede R, Volpin H, Oliver M, Fougedoire A, Stalh C, Fauve J, Copes B, Fei Z, Giovannoni J, Ori N, Lewinsohn E, Sherman A, Burger J, Tadmor Y, Schaffer AA, Katzir N (2010) A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes. Theor Appl Genet 121:511–533

Hyten DL, Song Q, Choi I-Y et al (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952. doi:10.1007/s00122-008-0726-2

Janick J, Paris HS, Parrish DC (2007) The cucurbits of Mediterranean antiquity: identification of taxa from ancient images and descriptions. Ann Bot 100:1441–1457

Jeffrey C (1980) A review of the Cucurbitaceae. Bot J Linn Soc 81:233–247

Khana MA, Han Y, Zhao YF, Korban SS (2012) A high-throughput apple SNP genotyping platform using the GoldenGate™ assay. Gene 494(2):196–201. doi:10.1016/j.gene.2011.12.001

Krzanowski WJ (2000) Principles of Multivariate analysis: a User's Perspective. Oxford University Press, Oxford

Liu K, Muse SV (2005) Powermarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

López-Sesé AI, Staub JE, Gómez-Guillamón ML (2003) Genetic analysis of Spanish melon (*Cucumis melo* L.) germplasm using a standardized molecular-marker array and geographically diverse reference accessions. Theor Appl Genet 108(1):41–52

Luan F, Delannay I, Staub JE (2008) Chinese melon (*Cucumis melo* L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns. Euphytica 164:445–461

McCreight JD, Staub JE (1993) Indo–US *Cucumis* germplasm expedition. HortScience 28:467

McMullen MD, Kresovich S, Sanchez H et al (2009) Genetic properties of the maize nested association mapping population. Science 325(5941):737–740. doi:10.1126/science.1174320

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. PNAS 106(30):12273–12278

Mliki A, Staub JE, Zhangyong S, Ghorbel A (2001) Genetic diversity in melon (*Cucumis melo* L.): an evaluation of African germplasm. Genet Resour Crop Evol 48:587–597

Monforte AJ, Garcia-Mas J, Arús P (2003) Genetic variability in melon based on microsatellite variation. Plant Breed 122:153–157

Munger HM, Robinson RW (1991) Nomenclature of *Cucumis melo* L. Cucurbit Genet Coop Rpt 14:43–44

Nakata E, Staub JE, López-Sesé AI, Katzir N (2005) Genetic diversity of Japanese melon cultivars (*Cucumis melo* L.) as assessed by random amplified polymorphic DNA and simple sequence repeat markers. Genet Resour Crop Evol 52:405–419

Naudin C (1859) Review des cucurbitace'es cultive Museum. Ann Sci Natl Ser 4 Bot 12:79–164

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II Gene frequency data. J Mol Evol 19:153–170

Nhi PTP, Akashi Y, Hang TTM, Tanaka K, Aierken Y, Yamamoto T, Nishida H, Long C, Kato K (2010) Genetic diversity in Vietnamese melon landraces revealed by the analyses of morphological traits and nuclear and cytoplasmic molecular markers. Breeding Sci 60:255–266

Nieto C, Piron F, Dalmais M, Marco CF, Moriones E, Gómez-Guillamón ML, Truniger V, Gómez P, Garcia-Mas J, Aranda MA, Bendahmane A (2007) EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. BMC Plant Biol 7:34. doi:10.1186/1471-2229-7-34

Nuez F, Ferrando C, Díez MJ, Costa J, Catalá MS, Cuartero J, Gómez-Guillamón ML (1988) Collecting *Cucumis melo* L. in Spain. Cucurbit Genet Coop Rep 11:54–56

Paris HS, Amar Z, Lev E (2012) Medieval emergence of sweet melons, *Cucumis melo* (Cucurbitaceae). Ann Bot 110(1):23–33

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12):e190. doi:10.1371/journal.pgen.0020190

Peltier J (2009) LOESS Smoothing in Excel. http://peltiertech.com/WordPress/loess-smoothing-in-excel/

Pitrat M (2008) Melon (*Cucumis melo* L.). In: Prohens J, Nuez F (eds) Handbook of crop breeding Vol I: vegetables. Springer, New York, pp 283–315

Pitrat M, Hanelt P, Hammer K (2000) Some comments on infraspecic classification of cultivars of melon. Acta Hortic 510:29–36

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248–249

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. PNAS 98(20):11479–11484

Renner SS, Schaefer H, Kocyan A (2007) Phylogenetics of *Cucumis* (Cucurbitaceae): cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). BMC Evol Biol 7:58. doi:10.1186/1471-2148-7-58

Robbins MD, Sim SC, Yang W, Van Deynze A, van der Knaap E, Joobeur T, Francis DM (2011) Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. J Exp Bot 62(6):1831–1845

Rostoks N, Ramsay L, MacKenzie K et al (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103(49):18656–18661

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resources 8:103–106

Roy A, Bal SS, Fergany M, Kaur S, Singh H, Malik AA, Singh J, Monforte AJ, Dhillon NPS (2012) Wild melon diversity in India (Punjab State). Genet Resour Crop Ev 59(5):755–767

Rubatzky VE, Yamaguchi M (1997) World vegetable principles, production and nutritive values, 2nd edn. Chapman and Hall, International Thompson Publ, New York 853 p

Sebastian P, Schaefer H, Telford IRH, Renner SS (2010) Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species from melon is from Australia. Proc Nat Acad Sci USA 107:14269–14273

Silberstein L, Kovalski I, Huang R, Anagnostou K, Jahn MK, Perl-Treves R (1999) Molecular variation in melon (*Cucumis melo* L.) as revealed by RFLP and RAPD markers. Sci Hort 79:101–111

Sim S-C, Van Deynze A, Stoffel K et al (2012) High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. PLoS ONE 7(9):e45520. doi:10.1371/journal.pone.0045520

Staub JE, Danin-Poleg Y, Fazio G, Horejsi T, Reis N, Katzir N (2000) Comparative analysis of cultivated melon groups (*Cucumis melo* L.) using random amplified polymorphic DNA and simple sequence repeat markers. Euphytica 115:225–241

Staub JE, López-Sesé I, Fanourakis N (2004) Diversity among melón landraces (*Cucumis melo* L.) from Greece and their genetic relationships with other melon germplasm of diverse origins. Euphytica 136:151–166

Stepansky A, Kovalski I, Perl-Treves R (1999) Intraspecific classification of melons (*Cucumis melo* L.) in view of their phenotypic and molecular variation. Plant Syst Evol 217:313–332

Tanaka K, Nishitani A, Akashi Y, Sakata Y, Nishida H, Yoshino H, Kato K (2007) Molecular characterization of South and East Asian melon, *Cucumis melo* L., and the origin of group conomon var. makuwa and var. conomon revealed by RAPD analysis. Euphytica 153:233–247

van Berloo R (2008) GGT 2.0: Versatile software for visualization and analysis of genetic data. J Hered 99(2):232–236. (Available online at: http://dx.doi.org/10.1093/jhered/esm109)

van Berloo R, Zhu A, Ursem R, Verbakel H, Gort G, van Eeuwijk FA (2008) Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. Theor Appl Genet 117(1):89–101

van Leeuwen H, Monfort A, Zhang H-B, Puigdomenech P (2003) Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library, microcolinearity between *Cucumis melo* and *Arabidopsis thaliana*. Plant Mol Biol 51:703–718

Whitaker TW, Bemis WP (1976) Cucurbits, *Cucumis*, *Citrullus*, *Cucurbita*, *Lagenaria* (Cucurbitaceae). In: Simmonds NW (ed) Evolution of Crop Plants. Longrams, New York, USA, pp 64–69

Yamamoto T, Nagasaki H, Yonemaru J-I, Ebana K, Nakajima M, Shibaya T, Yano M (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267. doi:10.1186/1471-2164-11-267

Yan J, Yang X, Shah T et al (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451. doi:10.1007/s11032-009-9343-2

Yeh FC, Yang R, Boyle TJ, Ye Z, Xiyan JM (2000) POPGENE 32, microsoft windows-based freeware for population genetic analysis. Molecular biology and biotechnology centre. University of Alberta, Edmonton

Yuste-Lisbona FJ, López-Sesé AI, Gómez-Guillamón ML (2010) Inheritance of resistance to races 1, 2 and 5 of powdery mildew in the melon TGR-1551. Plant Breeding 129(1):72–75

Zhao K, Wright M, Kimball J, Eizenga G, McClung A et al (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS ONE 5(5):e10780. doi:10.1371/journal.pone.0010780

Zhukovsky P (1951) Agricultural Structure of Turkey (Anatolia). Türkiye Seker Fab. AS. Yay. 20. (in Turkish)