How microsatellite diversity helps to understand the domestication history of melon

L. Serres-Giardi and C. Dogimont*
INRA UR 1052 Research Unit for Genetic Improvement of Fruit and Vegetables Domaine St Maurice, BP 94, 84 143 Montfavet Cedex France
*Corresponding author e-mail: catherine.dogimont@avignon.inra.fr

Keywords: Cucumis melo, genetic structure, phylogeny, SSR, wild melons

Abstract
Melon (Cucumis melo L.) population structure remains incomplete because of the sampling weakness in some botanical groups and for wild melons. The purpose of this study was to assess the genetic subdivisions in melon germplasm using a more representative sampling of the worldwide diversity and to investigate the localization of melon domestication and diversification. To reach this objective, a set of 20 microsatellite markers was used to genotype 713 accessions including 635 cultivated and feral melons, 66 wild melons and 12 relative Cucumis sp. A two-level structure in melon was revealed using population genetics statistics, clustering methods with genetic distances and Bayesian assignment method. Accessions split into two groups, fitting very well the subspecies melo and agrestis. Agrestis group was clearly substructured according to geographical origins, with African, Far-Eastern and Indian subgroups. The substructure of melo group was less resolute and the four clusters obtained grouped together several botanical groups originating from Europe, America, Middle East or Central Asia. African and Asian wild Cucumis melo were assigned to two distinct clusters in agrestis group, each with cultivated melons from the same respective geographical origin, suggesting that melon was domesticated twice. The domestication of wild African melon would have had a weak impact on the diversification of cultivated melon, producing only Sudanese tibish and seinat types. The domestication of wild Asian melon, probably in India, would have first produced Indian cultivated melons and then be at the origin of Far-Eastern melons. Introduction of Eastern melons in the West and successive breeding activities may have then produced melo ssp. These findings are consistent with recent data on the Cucumis genus structure and origin, and with the high frequency of resistance genes found in melon Indian accessions.
INTRODUCTION

Melon (*Cucumis melo* L.) is an important crop cultivated worldwide. It belongs to the genus *Cucumis* (Kirkbride 1993). *C. melo* is a diploid (2n=24), allogamous and insect-pollinated species. Widely considered as originating from East Africa, recent phylogenetic data on the *Cucumis* genus demonstrated that *C. melo* originates from Asia, together with its closely related species *C. sativus* (Sebastian et al. 2010; Telford et al. 2011).

Wild melons are rampant, highly branched vines with small leaves and flowers. Fruits are small; they have thin, bitter and non sweet flesh enclosing many small seeds. Wild melons have a wide geographical distribution and are found in Africa, in the Indian subcontinent and have been also described in Australia (Kirkbride 1993; Telford et al. 2011). Domesticated melons typically have larger fruits and seeds and an exceptional diversity of shapes and colours, which corresponds to a diversity of use. Some cultigroups are grown for eating the immature non-sweet fruits and others for the mature sweet fruits. The seeds of some melons are consumed; some melons are cultivated for their aromatic or decorative characteristics (Pitrat et al. 2000).

*C. melo* was described to split into two subspecies *C. melo* subsp. *meloides* and *C. melo* subsp. *agrestis*. Various infraspecific classifications have been proposed, mainly based on fruit shape and colour and flesh type. According to Pitrat (Pitrat et al. 2000), sixteen botanical groups were defined: five botanical groups in *C. melo* subsp. *agrestis* (*acidulus*, *chinensis*, *conomon*, *makuwa* and *momordica*) and eleven in *C. melo* subsp. *meloides* (*adana*, *ameri*, *cantaloupensis*, *chandalak*, *chato*, *chito*, *dudaim*, *inodorus*, *flexuosus*, *reticulatus* and *tibish*). No genetic barriers preventing crosses between the two subspecies, between the botanical groups or with wild melons have been reported.

The genetic diversity of *C. melo* has been described using various molecular markers, including RAPDs, RFLPs and microsatellite markers (SSR). Until now, most studies have focused on either a set of accessions from a single geographical origin or a limited number of accessions (Tanaka et al. 2007; Phan et al. 2010; Soltani et al. 2010; Aierken et al. 2011; Fergany et al. 2011; Yildiz et al. 2011, for some recent ones).

The aim of the present work was to further elucidate the genetic structure and evolutionary history in *C. melo* by using SSR markers on a wide-based collection of melon consisting of more than 700 accessions. In the present paper, we use a set of 20 microsatellite markers to survey the genetic diversity occurring in a collection spanning most of the diversity collected around the world (Table 1). The collection comprises 66 wild accessions and 635 cultivated and feral accessions, from all the 16 botanical groups defined by Pitrat (Pitrat et al. 2000). Our main objective was (1) to quantify the genetic diversity available in this collection using ‘neutral’
molecular markers (2) to test for the presence of an underlying genetic structure (3) to localise the processes of domestication and diversification of cultivated melons.

MATERIALS AND METHODS

Plant material

A collection comprising 701 accessions was chosen from the 2,400 *C. melo* accessions maintained at INRA Avignon. Six hundred and thirty-five cultivated and feral accessions were chosen to be representative of all geographical origins and botanical groups (Table 1). Their assignments to the botanical groups were made according to Pitrat et al. (2000). Sixty-six wild accessions were included originating from Africa, India and Central Asia. Accessions from 12 *Cucumis* species were added as outgroups. Most of the accessions have been collected long before this study and maintained by self-pollination at INRA Avignon.

DNA extraction and detection of SSR loci

Genomic DNA was isolated from fresh young leaves using the DNA extraction kit (Qiagen). Eight seedlings per accession were sampled in bulk. Twenty melon SSR markers, previously developed and used in genetic diversity studies were chosen (Díaz et al. 2011). The SSR markers have been mapped and were distributed in all 12 linkage groups, one or two per linkage group.

Data analysis

Genetic distance matrix between pairs of accessions was estimated from an index of dissimilarity based on a simple matching implemented in DARwin 5.0.156. It was used to perform a principal coordinate analysis (PCoA) and to construct a neighbour-joining tree using the DARwin software. The model-based program Structure (Pritchard et al. 2000) was used to infer population structure using a model allowing for admixture and correlated allele frequencies, with a burn-in length of 100,000 iterations and a run length of 100,000 iterations. The tested values of the number of groups K ranged from one to 30, with 10 replicate runs for each K. The most likely number of groups was determined using the magnitude of ΔK (Evanno et al. 2005). Inter-population genetic differentiation was measured by pairwise *Fst* calculated using GENETIX 4.05 (Belkhir et al. 2004). The intra-population level of gene diversity was measured by the expected heterozygosity using GENETIX. A home-made script computed the number of alleles, non-rare alleles (frequency higher than 5%) and private alleles (specific to one population) using R.

RESULTS

Microsatellite diversity

The 20 SSR loci were polymorphic and yielded a total of 272 alleles in *C.*
The average number of alleles per locus was 13.6 and 3.5 when removing rare alleles (i.e. with a frequency lower than 0.05). Combining data from the 20 microsatellite loci, we found 660 different multilocus genotypes, including 637 unique genotypes.

**Genetic structure of the collection**

We tested the existence of genetic structure in the melon collection. We determined the number (K) of diverged groups or clusters using the software Structure. The first peak of ΔK-value was obtained when testing the K=2 hypothesis, indicating the presence of two main genetic groups. The majority of the accessions were clearly assigned to one of the two groups, but some accessions appeared with high admixture level between the two groups. Principal coordinates analysis (PCoA) also allowed to distinguish two groups with a transitional zone. One group (I) essentially included accessions of the botanical groups *adana*, *ami*, *cantalupensis*, *chandalak*, *chate*, *flexuosus*, *inodorus* and *reticulatus*, corresponding to the *C. melo* subsp. *melo*. The other group (II) essentially included accessions of the botanical groups *acidulus*, *chinensis*, *conomon*, *makuwa* and *momordica*, which belong to the *C. melo* subsp. *agrestis*. *Chito* and *tibish* and *seinat* accessions, which have been reported to belong to *C. melo* subsp. *melo* (Pitrat et al. 2000) indubitably clustered in the group II (Table 2). 58 out of the 66 wild accessions were assigned to the group II. The eight wild melon accessions included in the group I all had wild global look but presented some typical characters of the cultivated melons, orange or sweet flesh for example; they likely resulted from gene flow with cultivated melons.

Further results of the Structure analysis showed that each group could be divided into subgroups: four subgroups (A to D) in group I and three subgroups in group II (E to G). The three subgroups E to G were clearly distinguishable in the PCoA results while the four subgroups in group I overlapped. In the same way very few accessions of the subgroups E to G were determined as admixed while admixture was frequent between accessions of the four subgroups A to D.

The three subgroups E to G clearly corresponded with the geographical origin of accessions. The group E consisted of accessions from India, including accessions of the botanical groups *acidulus* and *momordica*, unknown-type Indian cultivated melons and 27 Asian wild melons. It also included the eight accessions of the botanical group *chito* of this study, which were collected in South America and West Indies. The group F consisted of cultivated (16 *tibish* accessions) and wild melons (30 accessions) from Africa. The group G consisted of accessions from Far East, belonging to the botanical groups, *chinensis*, *conomon* and *makuwa*.

The sense of the four subgroups A to D was more horticultural than geographical. Subgroup A mainly consisted of *inodorus* melons from Europe,
Middle East, America and Africa; it also included the *ameri* melons recognized as cultigroup *ananas*. The group B mainly consisted of European, American and African accessions of the *reticulatus* and *cantalupensis* botanical groups. The group C grouped accessions of the botanical groups *chate*, *chandalak*, *adana* and *ameri*-except *ananas* cultigroup- and 19 *flexuosus* of which 13 were Sudanese melons. The group D included melons which showed high admixture level between groups I and II. These melons are mainly unknown types from India and *flexuosus* from India, Maghreb and Middle East.

**Phylogenetic analysis**

Phylogenetic analysis supported the genetic structuration of melon in the seven genetic groups defined by the Structure analysis. The three subgroups E, F, G of the group II (‘*agrestis*’) were well isolated from each other whereas the four subgroups of the group I (‘*melo*’) were less distinct. According to the tree root defined with other *Cucumis* species, the African subgroup F represented a distinctive population. The Far East accessions of the subgroup G and the American feral *chito* descended from the subgroup E grouping Indian melons. The subgroup D was intermediate between the three other ‘*melo*’ subgroups and the subgroup E, proving that melons of the subsp. *melo* came from Indian ‘*agrestis*’ melons. Relationships between the three ‘*melo*’ subgroups, A, B and C were less obvious.

Wild melons were split into two groups, an African one and an Asian one. The distribution of cultivated melons on the phylogeny in relation to wild melons suggested that cultivated melons had two distinct origins: African *tibish* and *seinat* came from the African wild pool whereas all others cultivated melons came from the Asian wild pool.

**Genetic diversity and differentiation among groups**

Genetic diversity was higher within the group II ‘*agrestis*’ (He=0.61) than within the group I ‘*melo*’ (He=0.45) in spite of lower sample size (Fig. 1). The subgroups E and D which comprise a large number of Indian accessions had a much higher genetic diversity than other subgroups (He$_E$=0.68; He$_D$=0.60). Genetic diversity of the subgroup F, which comprises African accessions, was also high (He=0.49). On the other hand, subgroup G, which grouped Far East accessions, was genetically poor (He=0.18) and shared all the alleles with the group E.

Pairwise *Fst* confirmed the proximity between the subgroups A, B and C of the group I and the intermediate place of the subgroup D between the subgroups A, B, C, and the subgroups E, F, G of the group II. Asian and African wild melons were respectively extremely close to Indian (subgroup E) and African (subgroup F) cultivated melons.
DISCUSSION

Infraspecific genetic differentiation of C. melo

A worldwide collection of melon, currently maintained and evaluated at INRA Avignon (France) is publicly available. Using 20 highly polymorphic SSR markers, we confirmed the uniqueness of most accessions of the collection and revealed a large amount of genetic diversity and allelic variation within the melon species. The average number of alleles per locus was 13.6, whereas it was only 3.0 in a broadly based collection of Cucurbita comprising 104 accessions and using a very large set of SSRs (134) (Gong et al. 2012).

The subdivision of the 701 melon accessions in two main clusters corroborated the taxonomical subdivision of C. melo into the two subspecies C. melo subsp. melo and C. melo subsp. agrestis. Western melons notably compound the main part of the ‘melo’ group while Eastern melons notably compound the majority of the ‘agrestis’ group, revealing a West-East geographical differentiation of cultivated melons. The botanical group chito, attributed to the subsp. melo (Pitrat et al. 2000), clearly clustered in the ‘agrestis’ genetic group, as previously found (Stepansky et al. 1999). Sudanese accessions of the recently described botanical group tibish and seinat were also classified into the subsp. melo and clustered in the ‘agrestis’ genetic group. These conflicts between botanical classification and genetic structure stress the limits of basing the botanical classification on characters which are difficult to evaluate and may have a continuous variability, as the pilosity of the female hypanthium (Pitrat et al. 2000).

In addition to this melo / agrestis structuration, a substructure of seven genetic subgroups was revealed by analyses at subspecies level. Within the ‘agrestis’ group, the subdivision in three subgroups is highly consistent with geographical distribution: the group E in India, the group F in Africa and the group G in Far East. We can infer from the genetic proximity of chito feral accessions from America with Indian melons of the group E that chito melons originated from India, as previously suggested (Decker-Walters et al. 2002). The substructure of ‘melo’ group in four clusters was less resolute, likely due to important gene flow from modern breeding programs.

The seven genetic groups did not exactly coincide with the botanical groups based on horticultural traits. Distinct botanical groups clustered in the same genetic subgroup and could not be differentiated: reticulatus and cantalupensis; chinensis, conomon and makuwa; acidulus, momordica and chito; adana, ameri, chate and chandalak. Inversely, flexuosus, ameri, dudaim and cantalupensis were included in several genetic subgroups. Flexuosus accessions (or snakemelons) from Sudan and Saudi Arabia, together with elongated melons of chate and adana, clustered in the genetic group C, whereas flexuosus from India and North Africa clustered in the intermediate genetic group D. This suggests, as reported recently for long-fruitied
cultivars in *C. pepo* subsp. *pepo* (Gong et al. 2012), that independent selection events for long fruits, occurred in *C. melo* subsp. *melo*. A great diversity among *flexuosus* accessions was previously reported (Yildiz et al. 2011; Soltani et al. 2010). Most *cantalupensis* were assigned to the subgroup B of ‘*melo*’. Brazilian *cantalupensis* clustered in the subgroup A with most *inodorus* melons. The six ornamental and aromatic *dudaim* accessions included in this study fall in three distinct genetic subgroups (D, E and G) from both ‘*melo*’ and ‘*agrestis*’ groups. Larger sampling would be required to understand the origin and spreading of the *dudaim* botanical group.

**Melon may have been domesticated twice, in Asia and in Africa**

The occurrence of a single domestication event or several independent domestication events can be settled by comparing the phylogenetic relationships between a representative sample of cultivated species and their wild forms or relatives. We showed that wild melons did not constitute a group distinct from the cultivated melons. Wild African melons clustered with the African subgroup of *C. melo* subsp. *agrestis*, while wild Asian melons clustered within the Indian subgroup of subsp. *agrestis*, suggesting two independent domestication events, in Africa and in Asia, from the *agrestis* subsp. wild gene pool. Two domestication events were also reported for *Cucurbita pepo*, in Mexico 10,000 years ago and in Eastern North Africa 5,000 years ago (Sanjur et al. 2002).

Our results are consistent with the fact that all the *melo* subsp. cultivated diversity may derive from the domestication of melons in Asia, by human selection and spreading. In contrast, domestication of melons in Africa may have had a marginal impact, only contributing to the *tibish* and *seinat* cultivated group, essentially grown in East Africa. However, the diversity of *tibish* and *seinat* is large (He=0.49) and may be especially relevant for the study of horticultural traits related to the domestication process and for pest and disease resistances.

Identifying what were the first cultivated melon types in Asia represents a hard challenge. The large genetic diversity observed in Indian melon germplasm and the occurrence of wild melons in the genetic Indian subgroup (E) suggest that India could be a domestication centre for melon. Traces of cultivation of melons in India are attested from 2,000 BC (Bates and Robinson 1995) but further archaeological data would be required. Far East subgroup, including *conomon*, *makuwa* and *chinensis* melons have a large phenotypic diversity but a low genetic diversity and a large differentiation with all the other subgroups. This suggests Far-eastern melons would have experienced a severe bottleneck and subsequent genetic drift. They would originate from the same Indian gene pool, probably small-seed Indian melons as this trait was fixed in all Far-eastern melons. In parallel to the spread of melon cultivation towards Far-East, Indian melons would have been introduced towards the West, leading to the development of *melo* subsp.
ACKNOWLEDGMENTS

We thank N. Giovinazzo and V. Rittener for the DNA collection and the INRA Genotyping platform of Clermont-Ferrand for providing the SSR data. We thank D. Besombes for the maintenance of the melon collection and M. Pitrat for the assignment of botanical groups to the accessions. This research was supported in part by an INRA grant.

Literature cited


Fergany M, Kaur B, Monforte AJ, Pitrat M, Rys C, Lecoq H, Dhillon NPS, Dhaliwal SS (2011) Variation in melon (Cucumis melo) landraces adapted to the humid tropics of southern India. Genetic Resources and Crop Evolution 58: 225-243


Table 1. Number and geographical origin of the studied accessions.

<table>
<thead>
<tr>
<th>Type</th>
<th>Africa</th>
<th>America</th>
<th>Central Asia</th>
<th>Europe</th>
<th>Far East</th>
<th>India-Pakistan</th>
<th>Middle East</th>
<th>Unknown</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>feral and cultivated</td>
<td>88</td>
<td>70</td>
<td>46</td>
<td>171</td>
<td>108</td>
<td>65</td>
<td>75</td>
<td>12</td>
<td>635</td>
</tr>
<tr>
<td>wild melon</td>
<td>36</td>
<td>3</td>
<td>17</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>12</td>
<td>66</td>
</tr>
<tr>
<td>other <em>Cucumis sp.</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>70</td>
<td>49</td>
<td>171</td>
<td>108</td>
<td>82</td>
<td>80</td>
<td>29</td>
<td>713</td>
</tr>
</tbody>
</table>
Table 2. Table of contingency between botanical groups to which the accessions belong and genetic group assignment to one of the two groups determined by Structure.

<table>
<thead>
<tr>
<th>Botanical groups</th>
<th>I 'melo'</th>
<th>II 'agrestis'</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>acidulus</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>adana</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ameri</td>
<td>37</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>cantalupensis</td>
<td>76</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>chandalak</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>chate</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>chinensis</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>chito</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>conomon</td>
<td>9</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>dudaim</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>flexuosus</td>
<td>54</td>
<td>3</td>
<td>57</td>
</tr>
<tr>
<td>inodorus</td>
<td>100</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>makuwa</td>
<td>6</td>
<td>44</td>
<td>50</td>
</tr>
<tr>
<td>momordica</td>
<td>7</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>reticulatus</td>
<td>59</td>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>tibish and seinat</td>
<td>1</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>unknown</td>
<td>128</td>
<td>22</td>
<td>150</td>
</tr>
<tr>
<td>wild melon</td>
<td>8</td>
<td>58</td>
<td>66</td>
</tr>
<tr>
<td>total</td>
<td>502</td>
<td>190</td>
<td>692</td>
</tr>
</tbody>
</table>

Fig. 1. Principal coordinates analysis (PCoA) for seven cultivated melon groups and two wild melon groups based on pairwise Fst generated from microsatellite data. The circle diameter indicates genetic diversity (He). Filled arcs reflect the proportion of private alleles in each group. Cultivated groups are differentiated by K2 color code from Structure: light grey for group I and dark grey for group II. Wild melon groups are black.