Molecular characterisation of Vesuvian apricot cultivars: implications for the certification and authentication of protected plant material

By R. RAO*, M. BENCIVENNI, M. LA MURA, T. ARAUJO-BURGOS and G. CORRADO
Dipartimento di Scienze del Suolo, della Pianta, dell’Ambiente e delle Produzioni Animali, via Università 100, Università degli Studi di Napoli Federico II, 80055 Portici (Naples), Italy (e-mail: rao@unina.it)
(Accepted 21 September 2009)

SUMMARY
In the EU, the production of local varieties with well-known quality characteristics may be defended by a Protected Designation of Origin (PDO) or a Protected Geographical Indication (PGI) label. Although the award of either status is based on meeting a number of conditions, the registration, distinction, and protection of such plant material is based on morphological traits. Here, using seven SSR and four AFLP primer pairs, we report on the molecular characterisation of a group of 36 apricot (Prunus armeniaca) varieties from southern Italy. This group included all the varieties approved for the “Albicocca Vesuviana” (Vesuvian Apricot) PGI label. Cluster analysis, based on genetic distances, clearly differentiated the 11 PGI cultivars from the other genotypes. Nonetheless, among the 11 PGI-cultivars, two pairs were found to have identical SSR and AFLP profiles. In addition, molecular analysis indicated the presence of mis-labelling and erroneous denominations of trees from the PGI area. The data revealed that DNA fingerprinting should be always deployed to complement the use of morphological traits in the description of plant material during the institution of a PDO or PGI label, and to evaluate the presence of genetic inconsistencies in orchards.

A pricot [Prunus armeniaca L. (2n = 16)] is a member of the family Rosaceae and was probably domesticated in western China and Central Asia. According to UN FAO estimates (http://www.fao.org/corp/statistics/en/), apricot cultivation is spread throughout the World. All major apricot-producing countries lie around the Mediterranean basin and in the Middle East, with the exception of Japan. Italy is the third largest producer in the World, after Turkey and the Islamic Republic of Iran. Approx. 40% of Italian apricot production comes from the Campania region, where the most important area for apricot cultivation lies around Mount Vesuvius. Despite yearly fluctuations in yield, this area accounts for approx. 60% of regional production, and just < 20% of Italian production (Mainolfi et al., 2006). A considerable amount of apricot fruit (≤ 80%) is processed locally and sold canned, in syrup, or used for nectars, juices, jams, and pastes. Vesuvian apricots have superior characteristics (e.g., high sugar content, uniform colour, proper texture, pulp yield, and a rich flavour) that make them well-suited for processing and as ingredients in traditional confectioneries and pastries (Forlani and Pugliano, 1997). The remaining apricot fruit production is distributed to local markets (Mainolfi et al., 2006), as Vesuvian apricots become marketable later than fruit imported from other countries. Unfortunately, imported apricots are becoming more widespread in this industrial sector because of their more regular availability. In addition, the area devoted to apricot cultivation around Mount Vesuvius is gradually decreasing, mainly because of the urbanisation of rural areas. The vast majority of orchards are small, family-businesses (≤ 5 ha), making it difficult to increase profits through technology-based improvements in productivity.

The region of cultivation, and cultivar, are important factors that influence the levels of carotenoids in apricot fruit, which are usually higher in cultivars grown in the Mediterranean region (Dragovic-Uzelac et al., 2007). In order to protect Vesuvian apricot production, an EU Protected Geographic Indication (PGI) is shortly to be registered (“Albicocca Vesuviana”). This trade name will be crucial for the promotion of a product with superior characteristics, the maintenance of biodiversity, improvements to the incomes of farmers (in return for a “genuine effort to improve quality”), the preservation of rural areas around Mt. Vesuvius, and to give accurate information about the origin of a product that is easily identifiable by local consumers.

Although several apricot cultivars have been described in the Vesuvian area (Forlani and Pugliano, 1997; Forte, 1987), only 11 will be accepted for the production of PGI-certified fruit. These traditional cultivars result from selections by local peasants, are particularly adapted to the Vesuvian environment (Massai and Pennone, 2007), and probably represent a distinct genetic pool. This would be consistent with the evidence that genetic differentiation of apricot is linked mainly to its geographical origin (De Vicente et al., 1998; Hagen et al., 2002; Hormaza, 2002; Zhebentyayeva et al., 2003).

The present study describes the molecular characterisation of 36 apricot cultivars grown in the Vesuvian area, including the 11 to be granted the “Albicocca Vesuviana” (Vesuvian Apricot) PGI trade-name. As genetic inconsistencies can be specially relevant in traditional plant material, our aims were: i) to study genetic relationships; and ii) to identify allelic combinations that could distinguish the PGI cultivars from other locally-grown cultivars.

*Author for correspondence.
MATERIALS AND METHODS

Plant material


Young leaves were collected at the Dipartimento di Arboricoltura, Botanica e Patologia Vegetale, Università di Napoli Federico II, or at the Azienda Sperimentale Improsta (Eboli), Consorzio per la Ricerca Applicata in Agricoltura, Regione Campania.

To test the genetic uniformity of trees, leaves from the cultivars ‘Ceccona’, ‘Pelleccoliella’, and ‘Portici’ were harvested on different farms in the area described in the “Albicoceca Vesuviana” PGI regulation.

DNA isolation, SSR analysis, and data collection

Genomic DNA isolation from young leaves and polymerase chain reaction (PCR) amplifications were carried out as described (Corrado et al., 2009; Rao et al., 2006). To genotype the plants, we used one primer pair (UDP97-402) developed in peach (Testolin et al., 2006), and six primer pairs developed in apricot, UDAp407, UDAp410, UDAp411, and UDAp420 (Messina et al., 2004), and UDAp480, and UDAp446 (R. Testolin, unpublished; Table I).

PCR reactions (20 µl) were performed using the AmpliTaq Gold polymerase mixture (Applied Biosystems, Milan, Italy) with the following conditions: 5 min denaturation at 94°C, followed by 35 cycles of 30 s denaturation at 94°C, 30 s following conditions: 5 min denaturation at 94°C, and six primer pairs developed in apricot, UDAp407, UDAp410, UDAp411, and UDAp420 (Messina et al., 2004), and UDAp480, and UDAp446 (R. Testolin, unpublished; Table I). PCR reactions (20 µl) were performed using the following conditions: 5 min denaturation at 94°C, followed by 35 cycles of 30 s denaturation at 94°C, 30 s at the annealing temperature listed in Table I, and 1 min of elongation at 72°C. The last cycle was followed by a final incubation for 5 min at 72°C. The amplification products were resolved using an ABI PRISM 3100 Avant (Applied Biosystems, Milan, Italy) and allele sizes were calculated using GeneScan 3.7 software (Applied Biosystems), as described (Caramante et al., 2009).

AFLP analysis and band scoring

AFLP assay was performed using the AFLP analysis system (Invitrogen Life Technologies, Milan, Italy) with some modifications. Genomic DNA (125 ng) was double-digested with Eco RI and Mse I for 2 h at 37°C and the restriction enzymes were inactivated at 72°C for 15 min. The DNA fragments were ligated for 20 h to Eco RI and Mse I adapters in adapter ligation solution using T4 DNA ligase (Invitrogen Life Technologies) according to the manufacturer’s instructions. For pre-selective amplification, 2.5 µl of a 10-fold diluted ligation mixture was amplified by 20 cycles of 94°C for 30 s, 56°C for 60 s, 72°C for 60 s using the Eco RI and Mse I pre-amplification primers.

For selective amplifications, 5 µl of a 30-fold diluted pre-amplified DNA reaction was used as template in a 20 µl reaction with the Mse I selective primers in Table II and the Eco- ACT primer, 5’-end labelled with hexa-chloro-6-carboxyfluorescein (HEX). PCR reactions included one cycle of 94°C for 30 s, 65°C for 30 s, 72°C for 60 s, 12 touch-down PCR cycles lowering the annealing temperature by 0.7 °C per cycle, and 23 cycles of 94°C for 30 s, 56°C for 60 s, and 72°C for 60 s. An appropriate dilution of the selective amplification reaction was added to 0.5 µl GeneScan-400 ROX size standard (Applied Biosystems) and Hi-Di Formamide (Applied Biosystems) to a final volume of 10 µl. Prior to capillary electrophoresis in an ABI PRISM 3730 (Applied Biosystems), samples were heated for 5 min at 94°C and chilled on ice. Fragment sizes, interpolated to the internal standard according to the Local Southern algorithm, and peak intensities, were calculated using Peak Scanner Version 1.0 software (Applied Biosystems). Only those fragments between 50 – 450 bp were used for scoring. We considered those fragments that fulfilled the default quality requirements of the Peak Scanner software for AFLPs, in duplicate PCR experiments, to be reproducible.

<table>
<thead>
<tr>
<th>SSR locus</th>
<th>Core</th>
<th>Primer sequence (5’ → 3’)</th>
<th>Ta (°C)</th>
<th>Allelic size range (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(TT)</td>
<td>AGACGACACAGCTTCTTCTGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDAp410</td>
<td>(AG)13</td>
<td>TGGTGAGAAGAGAGACG</td>
<td>56</td>
<td>117–177</td>
<td>Messina et al. (2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAACGGGTTGTGTACAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDAp411</td>
<td>(GA)13</td>
<td>TGGTGGAGAAGAGAGACG</td>
<td>56</td>
<td>78–104</td>
<td>Messina et al. (2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTCCCCCACCTTACATG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDAp420</td>
<td>(CT)20</td>
<td>TTTTCTCTCTTCCCTATG</td>
<td>56</td>
<td>158–178</td>
<td>Messina et al. (2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCGAAGACTTATGTTTCTGAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDAp480</td>
<td>(GA)8</td>
<td>GGTCAACAGACCCG</td>
<td>56</td>
<td>121–181</td>
<td>unpublished sequence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGTTCTGGAATTGATGTACATG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDAp446</td>
<td>(GA)8</td>
<td>CTTCCCTCTAGATTTCATG</td>
<td>54</td>
<td>121–179</td>
<td>unpublished sequence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CGTGGTTTGGGACATAGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDP97-402</td>
<td>(AG)7</td>
<td>TCCCATACCAACAAAAAACCCTG</td>
<td>54</td>
<td>118–140</td>
<td>Testolin et al. (2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGAGAAGGCTTGTCCTT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ta, annealing temperature.

*5’-end-labelled with 6-carboxyfluorescein (FAM) dye.

†5’-end-labelled with hexa-chloro-6-carboxyfluorescein (HEX) dye.
Data analysis
Cultivars that showed a single amplified SSR allele were considered, by convention, to be homozygous at that locus. For each SSR, we calculated the number of effective alleles, the observed heterozygosity, the polymorphic information content, the unbiased expected heterozygosity, the Fixation Index, the probability of identity, and the estimated frequency of null alleles as reported by Corrado et al. (2009).

Genetic similarity (GS) between genotype $i$ and $j$ was calculated using the Jaccard coefficient:

$$GS_{ij} = a / (n – d)$$

where $a$ was the number of matching amplified DNA fragments, $n$ was the total number of fragments, and $d$ was the number of matching absent fragments.

A dendrogram was constructed using the UPGMA algorithm, supporting nodes with a bootstrap re-sampling of 1,000 cycles.

For each AFLP primer pair, we calculated the average percentage of reproducible fragments per sample. Only reproducible fragments were used to determine: i) the percentage of polymorphic loci found in a reaction; ii) the number of effective alleles ($Ne$), where $Ne = 1 / \left( p_i^2 + q_i^2 \right)$; and iii) Nei’s average gene diversity per locus ($H$), which, for an AFLP marker, is equivalent to the average expected heterozygosity $He$ (Nei, 1973); and (iv) the polymorphic index content ($PIC$):

$$PIC = \frac{\sum (1 – p_i^2 – q_i^2)}{n}$$

where $p_i$ and $q_i$ were the frequencies of the presence and absence of an allele at locus $i$, respectively; and $n$ was the number of loci examined.

RESULTS
Polymorphism and cultivar variability

Molecular fingerprinting was carried out using primers designed to amplify six apricot SSRs and one transferable SSR of peach (Messina et al., 2004; Testolin et al., 2000), that detected polymorphisms at all loci. A total of 56 different alleles were scored, with an average of 8.00 ± 0.96 (mean ± SE) alleles per locus. The effective number of alleles (approx. 60% of the total), which enabled comparisons of allelic diversity to be made across loci independently of the allele frequency distribution, indicated that all loci were sufficiently diverse in the population analysed. This was also supported by the high PIC values, on average 0.78 ± 0.06 (mean ± SD). The average allele frequency per locus was $13.6 \pm 4.2\%$ (mean ± SD), but varied greatly, ranging from 43.8% for the 130 bp allele at the UDP97-402 locus to a minimum of 1.4% for five alleles (two each at the UDP410 and UDP420 loci and one at the UDP411 locus). While, as expected, the number of alleles was negatively correlated with the average allele frequency ($r = -0.95; P < 0.01$), the effective number of alleles was only weakly correlated with the average allele frequency ($r = 0.67; P = 0.21$). Overall, the data indicated the presence of a good level of polymorphism in the apricot population analysed, essentially due to the high number of alleles and to the frequency of heterozygous loci. The genetic indices also suggested the presence of locus-specific differences. The Fixation Index was substantially positive only for the UDP411 and UDP480 loci, pointing to the possible presence of undetected null alleles. Estimations of null allele frequency, from the deficiency of heterozygotes, indicated a significant probability, highest for the UDP411 locus. This was consistent with the very low value of $Ho$, which was essentially due to the lower number of alleles detected at this locus (Table III).

DNA samples from each pair of trees of the same cultivar showed the same SSR profile, with two exceptions. The first was ‘Bocuccia Spinosa’. At the UDP480 locus, one tree had a single peak (143 bp), while the second tree was heterozygous (143 bp and 181 bp). Genetic analysis favoured the possibility that such a difference was due to the presence of a null allele, rather than a homozygous locus. The second discrepancy was found for one tree labelled ‘Bocuccia Liscia’. At all seven SSR loci, this tree had an allelic profile identical to ‘Fracasso’. Therefore, it could be considered to be a case of mis-labelling or incorrect denomination.

The genetic relationships of the apricot biotypes were analysed using a distance-based hierarchical classification. The UPGMA cluster is shown in Figure 1. The dendrogram indicated the presence of a group with highly similar genotypes. This group, which originated at the node with a bootstrap value of 99 (Figure 1), was

---

**Table II**

<table>
<thead>
<tr>
<th>Selective primer</th>
<th>AR $^1$</th>
<th>Analysed bands</th>
<th>MR</th>
<th>PIC</th>
<th>Ne</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mse I-GAG</td>
<td>66.1%</td>
<td>35</td>
<td>34.2</td>
<td>0.14</td>
<td>1.24</td>
</tr>
<tr>
<td>Mse I-CAT</td>
<td>80.3%</td>
<td>71</td>
<td>64.5</td>
<td>0.19</td>
<td>1.32</td>
</tr>
<tr>
<td>Mse I-CTA</td>
<td>86.7%</td>
<td>113</td>
<td>39.2</td>
<td>0.17</td>
<td>1.28</td>
</tr>
<tr>
<td>Mse I-CTT</td>
<td>60.2%</td>
<td>64</td>
<td>40.6</td>
<td>0.17</td>
<td>1.27</td>
</tr>
</tbody>
</table>

$^1$ AR, average percentage of reproducible bands. MR, multiplex ratio. PIC, polymorphic index content. Ne, number of effective alleles.

**Table III**

<table>
<thead>
<tr>
<th>SSR</th>
<th>Na</th>
<th>AAL</th>
<th>Ne</th>
<th>Ho</th>
<th>PIC</th>
<th>UHe</th>
<th>FI</th>
<th>EFNA</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>UD Ap407</td>
<td>9</td>
<td>0.11</td>
<td>5.75</td>
<td>0.94</td>
<td>0.82</td>
<td>0.84</td>
<td>-0.14</td>
<td>-0.66</td>
<td>0.09</td>
</tr>
<tr>
<td>UD Ap410</td>
<td>10</td>
<td>0.10</td>
<td>6.56</td>
<td>0.86</td>
<td>0.84</td>
<td>0.86</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>UD Ap411</td>
<td>6</td>
<td>0.20</td>
<td>3.43</td>
<td>0.19</td>
<td>0.71</td>
<td>0.72</td>
<td>0.72</td>
<td>0.31</td>
<td>0.22</td>
</tr>
<tr>
<td>UD Ap420</td>
<td>8</td>
<td>0.12</td>
<td>5.54</td>
<td>0.78</td>
<td>0.82</td>
<td>0.83</td>
<td>0.05</td>
<td>0.02</td>
<td>0.11</td>
</tr>
<tr>
<td>UD Ap480</td>
<td>6</td>
<td>0.14</td>
<td>5.12</td>
<td>0.47</td>
<td>0.80</td>
<td>0.82</td>
<td>0.41</td>
<td>0.17</td>
<td>0.12</td>
</tr>
<tr>
<td>UD Ap446</td>
<td>12</td>
<td>0.07</td>
<td>4.98</td>
<td>0.77</td>
<td>0.80</td>
<td>0.81</td>
<td>0.03</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>UDP97-402</td>
<td>5</td>
<td>0.20</td>
<td>3.16</td>
<td>0.62</td>
<td>0.68</td>
<td>0.70</td>
<td>0.08</td>
<td>0.03</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Na, number of alleles. AAL, average allele frequency. Ne, number of effective allele. Ho, observed heterozygosity. PIC, polymorphic index content. UHe, unbiased He. FI, fixation index. EFNA, estimated frequency of null alleles. PI, probability of identity.
characterised by having identical SSR profiles at four loci (UDAp407, UDAp410, UDAp411, and UDAp420). None of these trees were included in the “Albicocca Vesuviana” PGI regulation. In this group, ‘Sel. 2 (Portici)’ and ‘Stradona’ were the only trees that displayed a single amplification product at one locus, (UDAp480 and UDAp499, respectively), while the remaining biotypes had the same heterozygous allelic profile (121 bp and 153 bp at the UDAp480, and 129 bp and 165 bp at the UDAp446 loci). These plants were divided in two clusters, since some of them were heterozygotes (e.g., ‘Ananassa’) and others had a single amplification product at the UDP97-402 locus (e.g., ‘Grangicano’). Among the 11 PGI cultivars, two pairs of cultivars were found to have identical SSR profiles at all loci (‘San Castrese’ and ‘Palumella’; and ‘Pellechiella’ and ‘Portici’).

For these reasons, we extended the genetic analysis of these four PGI genotypes using AFLPs (Table II). Although related to a reduced numbers of genotypes, the best high-quality profiles (i.e., well-separated peaks, a high signal-to-noise ratio, a lack of shoulder or stutter peaks, and fragments distributed throughout the available size range) were found for the Mse I-CTA and Eco RI-ACT primer combination. Each of the four primer pairs was polymorphic, but they could not distinguish between ‘San Castrese’ and ‘Palumella’, or between ‘Pellechiella’ and ‘Portici’. To investigate the high degree of similarity between these genotypes further, we compared the reference trees at the Istituto Sperimentale per la Frutticoltura to trees growing on farms of the “Albicocca Vesuviana” PGI area. This assay also provided a clue concerning intra-cultivar variability. This analysis included three plants of ‘Pellechiella’ and three of ‘Portici’, and also trees of a PGI cultivar (‘Ceccona’) with a unique SSR profile. Figure 2 shows the UPGMA dendrogram based on genetic distances calculated from the SSR data. The data confirmed that ‘Pellechiella’ and ‘Portici’ had an identical SSR profile. Furthermore, ‘Portici’ sample 4 had a different profile at the UDAp410 locus compared to the reference tree and the other two samples. Finally, SSR fingerprints also indicated the presence of incorrect denominations on farms (i.e., ‘Ceccona’ 4 and ‘Portici’ 3).

**DISCUSSION**

A molecular analysis of Vesuvian PGI apricots using SSRs indicated slightly lower values for alleles detected and observed heterozygosity than those reported by Messina et al. (2004) and by Testolin et al. (2000), yet similar to the values reported in other studies (Hormaza, 2002; Romero et al., 2003; Zhebentyayeva et al., 2003). The average Fixation Index of our samples (a measure of population differentiation based on genetic polymorphism data) showed a lower value when compared to that reported by Romero et al. (2003), but similar to that obtained by analysing Tunisian apricots (Khadari et al., 2006). Taking into account that the trees analysed in the present study were cultivated in a limited geographic area, the data indicated a significant level of polymorphism and the presence of a potentially interesting level of heterozygosity among the Vesuvian apricot varieties. This can be explained by different factors: a bias towards polymorphic SSR loci, the low selection pressure applied by breeders to Vesuvian apricots over the past decades, and because the introduction of (and replacement with) new genotypes is rare (Forlani and Pugliano, 1997). Furthermore, cultivation of apricot in the Vesuvian area is broadly...
based on traditional farming, causing an appreciable elevated level of varietal promiscuity (Forlani and Pugliano, 1997).

In order to estimate the genetic relationships among samples, we performed an UPGMA hierarchical classification based upon the genetic distance matrix, which allotted the cultivars into well-separated clusters. One of these included genetically similar genotypes that were not included in the PGI regulation due to their limited importance and limited distribution. We cannot exclude the possibility that these biotypes were derived from a single entity, and that the very low level of genetic polymorphism observed was due to intra-cultivar variability. Members of this group may have different denotations because, along with the traditional varieties, the cultivation of apricot is also based on shared material that is usually named after local customs (Forlani and Pugliano, 1997).

SSR fingerprints indicated that ‘Portici’ and ‘Pellecchiella’, and ‘San Casterese’ and ‘Palumella’ were identical at the seven loci analysed. ‘San Casterese’ and ‘Pellecchiella’ and, as expected, also revealed cases of incorrect denomination (Caramante et al., 2009; Rao et al., 2006). In the future, it will be interesting to complement such molecular analysis with pomological and phenotypic descriptions, to assess possible differences between these PGI cultivars.

In conclusion, the results of the present study indicated the presence of good genetic differentiation in the Vesuvian apricot population, and confirmed that DNA molecular markers are a powerful tool to reveal the presence of genetic inconsistencies. Furthermore, we could identify “Albicocca Vesuviana”-specific profiles that can be used to improve the genetic traceability of PGI-labelled products (Melchiade et al., 2007).

Autochthonous animals and plants, specifically adapted to local environments, are the basic resource to make a product “typical”. This important level of biodiversity can be preserved by promoting the economic sustainability of traditional production systems by adequate legal protection which, in turn, will also encourage the economic stability of populations in rural areas. Thus, our work not only confirmed that DNA molecular markers could identify “Albicocca Vesuviana”-specific products by regional institutions should include on-farm molecular investigations before and after the institution of new labels.

We are grateful to Professor R. Testolin for the selection and gift of the primer pairs used, to Drs. R. Aversano and A. Nunziata for the protocols and assistance in the AFLP experiments, and to Professor M. Forlani for support and advice during the collection of plant material.

REFERENCES


