Tree Genetics & Genomes (2016) 12:36
DOI 10.1007/s11295-016-0997-8

ORIGINAL ARTICLE

Genetic pool structure of local apple cultivars from Portugal assessed by microsatellites

V. Ferreira[1] · A. M. Ramos-Cabrer[2] · V. Carnide[1] · O. Pinto-Carnide[1] · A. Assunção[3] · A. Marreiros[4] · R. Rodrigues[5] · S. Pereira-Lorenzo[2] · I. Castro[1]

Received: 17 November 2015 / Revised: 17 February 2016 / Accepted: 22 March 2016

Abstract A set of 87 apple accessions, located in three Portuguese apple germplasm collections, plus eight reference cultivars, were analyzed using 16 SSRs with the aim of assessing their genetic diversity and structure and evaluating relationships among them. Among the accessions studied, 64 unique genotypes were identified, 51 diploids and 13 putative triploids, revealing 19 groups of synonyms and 4 of homonyms. The genetic analyses performed by Bayesian model-based clustering (Structure) revealed a clear differentiation of two major groups (RPP1 and RPP2), one of them (RPP1) corresponding to old local Portuguese accessions, some of them putatively derived from ancient hybridization with Reineta, and the second one (RPP2) including all reference accessions and local Portuguese accessions considered to be of more recent origin. These results were confirmed by factorial correspondence analysis (FCA) and molecular variance analysis (AMOVA) and are in accordance with the Jaccard coefficients. Results indicate that the Portuguese apple germplasm includes an important and distinct gene pool of cultivars with respect to reference international cultivars, being a relevant source of genetic diversity for present and future breeding programs, originated at least 200 years ago, when some of these cultivars were firstly named. Moreover, this study will be useful to improve the apple germplasm management and in the definition of a preservation strategy concerning the Portuguese apple landraces.

Keywords *Malus* × *domestica* Borkh · Local cultivars · Microsatellite (SSR) · Genetic diversity · Germplasm · Conservation

Communicated by E. Dirlewanger

Electronic supplementary material The online version of this article (doi:10.1007/s11295-016-0997-8) contains supplementary material, which is available to authorized users.

✉ I. Castro
icastro@utad.pt

[1] Centro de Investigação e de Tecnologias Agro-Ambientais e Biológicas (CITAB), Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal

[2] Departamento de Producción Vexetal, Escola Politécnica Superior, Universidade de Santiago de Compostela, Campus de Lugo, 27002 Lugo, Spain

[3] Divisão de Apoio ao Setor Agroalimentar, Direção Regional de Agricultura e Pescas do Norte, Quinta de S. José, S. Pedro de Merelim, 4700-859 Braga, Portugal

[4] Direção de Serviços de Desenvolvimento Agroalimentar e Rural, Direção Regional de Agricultura e Pescas do Algarve, Apartado 282, Patacão 8001-904, Faro, Portugal

[5] Centro de Investigação de Montanha (CIMO), Escola Superior Agrária de Ponte de Lima, Instituto Politécnico de Viana do Castelo (ESAPL/IPVC), 4990-706 Refóios, Portugal

Introduction

Apple belongs to the Rosaceae family and *Malus* genus. Despite *Malus* taxonomy be complex, unclear, and likely to be revised in the future, it is described as comprising about 30 species and several subspecies (Robinson et al. 2001).

The domesticated apple (*Malus* × *domestica* Borkh.) is one of the most relevant cultivated fruit crops worldwide, in particular on temperate zones, and the fourth most economically important, following citrus, grape, and banana (Hummer and Janick 2009) being cultivated for a long time, with references dating back to ancient Greece and Roman times (Zohary et al. 2012).

The cultivated apple was initially domesticated from the wild apple *Malus sieversii* (Ldb.) Roem in the Tian Shan

Published online: 07 April 2016

Mountains in Central Asia about 4000 to 10,000 years ago. The Silk Route brought the domesticated apples westwards along, where they could have hybridized with other wild apples, such as *Malus baccata* (L.) Borkh. in Siberia, *Malus orientalis* Uglitz. in the Caucasus, and *Malus sylvestris* Mill. in Europe (Cornille et al. 2014).

Although more than 7000 apple cultivars have been documented across the globe, nowadays only few of them dominate the world fruit production, namely cultivars with an ancient origin and broadly spread, such as 'McIntosh' (1800s), 'Jonathan' (1820s), 'Cox's Orange Pippin' (1830s), 'Granny Smith' (1860s), 'Delicious' (1870s), and 'Golden Delicious' (1890s), to few well-adapted genotypes, such as 'Red Delicious', 'Golden Delicious' and 'Jonathan,' and to polyclonal cultivars, 'Red Delicious,' 'Gala,' and 'Fuji' (Janick et al. 1996; Noiton and Alspach 1996). This trend led to the standardization and genetic uniformity of commercial apple cultivars. As a consequence, many traditional and locally well-adapted cultivars have been considered obsolete, replaced, and almost excluded due to the low productivity, lack of size uniformity of their fruits which, in most cases, do not meet the standard of modern cultivars, leading to a dramatic loss of genetic variability. However, to prevent the loss of genetic diversity, these old and local cultivars should be adequately conserved.

In Portugal, the genetic erosion has been recognized as a major problem, and since the 1990s, considerable effort has been exerted in the preservation of the genetic resources of fruit trees (Queiroz et al. 2015). As a result, several national reference collections for apple were established in regional centers of the Ministry of Agriculture, namely in the North (Felgueiras), Center (Viseu), and South (Tavira) mainland Portugal (Barata et al. 2008). Some accessions are putative local ecotypes, some of them represent old cultivars, still planted nowadays, such as 'Bravo de Esmolfe' and 'Riscadinha de Palmela,' both present throughout the country, and 'Porta da Loja,' mainly found in Northwest Portugal (Veloso et al. 2008). References to 'Bravo de Esmolfe' date back two centuries (Mota 1919) and together with 'Riscadinha de Palmela' are the only Portuguese regional cultivars currently cultivated under designation 'Protected Denomination of Origin' (DOP).

Within apple collections, microsatellite markers have been favored over other to establish unique genetic identities or fingerprints and to assess genetic diversity (Evans et al. 2009). A set of these markers has been recommended by the European Cooperative Programme for Plant Genetic Resources (ECPGR) for the study of apple and pear genetic resources, in order to harmonize information gathered by different groups (Evans et al. 2009). An adequate molecular characterization is essential for an efficient germplasm management, by identifying clonal relationships, synonyms, homonyms, and propagation or labeling errors (true-to-type correspondence of accessions).

Markers proposed by ECPGR have already been used to characterize apple collections namely in USA (Potts et al. 2012), Spain (Pereira-Lorenzo et al. 2008; Urrestarazu et al. 2012), and Italy (Liang et al. 2015). Despite this set of SSR markers has been used to characterize Portuguese pear landraces (Bassil et al. 2009; Queiroz et al. 2015), Portuguese apple landraces have not yet been genotyped by the standard SSR markers that span apple genome. Even though most of the old germplasm accessions are considered of low commercial interest, they are diversity repositories and a valuable source of allelic variability, especially on apple breeding programs in order to obtain new cultivars with economical value and adaptive traits, and thus, the interest in them is increasing.

On this way, the main goal of this study was to evaluate the genetic diversity and relationships of Portuguese local apple accessions by SSR markers proposed by the ECPGR for apple genotyping, in order (1) to determine the genetic identity of the apple material through identification of duplicates, synonyms, and homonyms that are difficult to distinguish using standard morphological descriptors; (2) to assess the genetic diversity and structure in the studied germplasm; (3) compare the obtained genotypes with other collections or genotypes for association studies in order to contribute for a better management of apple collections.

Material and methods

Plant material

A total of 87 accessions, composed by old and local autochthonous apple germplasm, were studied (Table 1). This material was maintained in three governmental germplasm collections: 'Escola Superior Agrária de Ponte de Lima' (ESAPL) (23 accessions), 'Quinta de Sergure' from the 'Direção Regional de Agricultura e Pescas do Norte' (DRAPNorte) (39 accessions), and 'Centro de Experimentação Agrária de Tavira' from the 'Direção Regional de Agricultura e Pescas do Algarve' (DRAPAlgarve) (25 accessions), located in Minho, Douro Litoral, and Algarve Portuguese provinces, respectively (Fig. 1). Multiple accessions with the same designation sampled in one or more germplam collections were given different codes. In addition, eight apple accessions ['Delicious,' 'Fiesta,' 'Prima,' 'Worcester Pearmain,' 'Michelin' (Cider), 'Malling 9' (rootstock), '*Malus floribunda* 821,' and '*Malus robusta* 5'], chosen as a reference set by ECPGR, were also analyzed (Table 1). The reference accessions were provided by the Research Institute of Horticulture and Seeds, INRA, Agrocampus Ouest, University of Angers, France.

For each accession, genomic DNA was extracted from 100 mg of woody shoot following the protocol supplied in the NucleoSpin Plant Kit (Macherey Nagel, Duren, Germany). The

Table 1 Set of 87 Portuguese apple accessions and eight references studied, indicating germplasm collection

Accession	Collection
Arouca	DRAPNorte
Baionesa Ceboleira	DRAPNorte
Belle de Boskoop	DRAPNorte
Bravo de Esmolfe 1	DRAPAlgarve
Bravo de Esmolfe 2	DRAPNorte
Camoesa D. Ester	DRAPNorte
Camoesa Fina 1	DRAPNorte
Camoesa Fina 2	ESAPL
Camoesa Rosa 1	DRAPNorte
Camoesa Rosa 2	DRAPNorte
Camoesa de Coura	ESAPL
Camoesa Pedra	ESAPL
Capela Sra. das Neves	DRAPNorte
Casanova	DRAPAlgarve
D. Emília	DRAPAlgarve
D. Manuel	DRAPNorte
Espelho	DRAPAlgarve
Espriega Clara	DRAPNorte
Espriega 1	DRAPNorte
Espriega 2	ESAPL
Espriega Parda	DRAPNorte
Espriega Verde	DRAPNorte
Focinho de Burro	DRAPNorte
Gigante D'Oiro	DRAPAlgarve
Gigante do Douro 1	DRAPNorte
Gigante do Douro 2	ESAPL
Gilbarbeda Amarela	ESAPL
Gilbarbeda Branca	ESAPL
Maçã Ácida	DRAPAlgarve
Maçã Cigana	DRAPAlgarve
Maçã da Pedralva	DRAPAlgarve
Maçã de Outubro	DRAPAlgarve
Maçã de S. João	DRAPNorte
Maçã de Vinho	DRAPNorte
Maçã do Carrascalinho	DRAPAlgarve
Malápio 1	DRAPAlgarve
Malápio 2	ESAPL
Malápio Bico de Pardal	DRAPAlgarve
Malápio de Pé Curto	DRAPAlgarve
Malápio do Norte	DRAPAlgarve
Malápio Pé de Porco	DRAPAlgarve
Maria Gomes 1	DRAPAlgarve
Maria Gomes 2	DRAPAlgarve
Marmela	DRAPNorte
Martingil	DRAPNorte
Moleirinha 1	DRAPNorte
Moleirinha 2	ESAPL
Pardo de Sovinha	DRAPNorte
Pardo Lindo 1	DRAPNorte
Pardo Lindo 2	ESAPL
Pé de Cera 1	DRAPNorte
Pé de Cera 2	ESAPL
Pé de Cera 3	ESAPL
Pedregal	DRAPAlgarve
Perna de Pisco	ESAPL
Pêro da Avó	DRAPAlgarve
Pêro da Lixa	DRAPNorte
Pêro de Coura	DRAPNorte
Pêro de Mesa	DRAPAlgarve
Pêro Limão	DRAPNorte
Pêro Tomate	DRAPAlgarve
Pêro Vermelho	DRAPAlgarve
Pink Lady Rosa	DRAPNorte
Pipo de Basto 1	DRAPNorte
Pipo de Basto 2	ESAPL
Porta da Loja 1	DRAPNorte
Porta da Loja 2	ESAPL
Porta da Loja 3	ESAPL
Porta da Loja 4	ESAPL
Porta da Loja 5	ESAPL
Rajada Vermelha	ESAPL
Refóios	DRAPNorte
Reineta Espanhola	DRAPNorte
Repinau	DRAPNorte
Riscadinha da Feira	DRAPNorte
Riscadinha das Malhas	DRAPNorte
S. João	ESAPL
S. Miguel	DRAPNorte
Saínho	DRAPAlgarve
Sangue de Boi	ESAPL
Setúbal	DRAPAlgarve
Três ao Prato	DRAPNorte
Verdeal 1	DRAPNorte
Verdeal 2	ESAPL
Verdinho de Vila Real	DRAPNorte
Vermelha de Refóios	ESAPL
Zé Luís	DRAPAlgarve
Delicious	IRHS—reference
Fiesta	IRHS—reference
Malling 9 (Rootstock)	IRHS—reference
Malus floribunda 821	IRHS—reference
Malus robusta 5	IRHS—reference
Michelin (Cider)	IRHS—reference
Prima	IRHS—reference
Worcester Pearmain	IRHS—reference

DRAPNorte Quinta de Sergude, Direção Regional de Agricultura e Pescas do Norte, *DRAPAlgarve* Centro de Experimentação Agrária de Tavira, Direção Regional de Agricultura e Pescas do Algarve, *ESAPL* Escola Superior Agrária de Ponte de Lima, *IRHS* Institut de Recherche en Horticulture et Semences

DNA concentration was determined by UV spectrometer (Nanodrop® ND-1000, ThermoFisher Scientific, USA)

followed by quality check using a 1.0 % (*w*/*v*) horizontal agarose gel electrophoresis using TBE buffer and ethidium bromide staining. DNA working dilutions were adjusted to 10 ng/μL.

Microsatellite amplification

A set of 16 microsatellites (Hokanson et al. 1998; Liebhard et al. 2002; Vinatzer et al. 2004; Silfverberg-Dilworth et al. 2006) was used (Table 2). These microsatellites include the common set of ECPGR SSR markers for *Malus* characterization and all are being used in the frame of the FruitBreedomics European project for genotyping numerous apple accessions allowing interlaboratory data comparison. Fifteen of the markers used belong each to 1 of the 17 linkage groups of the apple genome, ensuring independence among loci. The forward primers were labeled with 6-FAM, VIC, NED, or PET fluorescent dye, and three multiplex PCRs were designed (denoted as A, B, and C) (Table 2).

Polymerase chain reactions for the three multiplex were performed in a final volume of 10 μL using 10 ng of DNA template, 0.10 μM of each primer, except for CH02c11 and CH02c06 for which 0.15 and 0.40 μM were used, respectively, and 1× PCR Master mix of QIAGEN kit multiplex PCR (Qiagen, Hilden, Germany). The PCR reactions were carried out in a thermal cycler with the following temperature profile for multiplexes A and B: an initial denaturation step at 95 °C for 15 min, followed by five touchdown cycles at 95 °C for 30 s, 65–1 °C/cycle for 1 min, and 72 °C for 1 min, followed by 30 cycles at 95 °C for 30 s, 60 °C for 1 min, 72 °C for 1 min, and a final step at 72 °C for 30 min. Multiplex C was modified, as a touchdown cycle number was raised to seven, and annealing temperature for the last 30 cycles was 58 °C instead of 60 °C.

Eight reference cultivars [‘Delicious,’ ‘Fiesta,’ ‘Prima,’ ‘Worcester Pearmain,’ ‘Michelin’ (Cider), ‘Malling 9’ (rootstock), ‘*Malus floribunda* 821,’ and ‘*Malus robusta* 5’] were used as control in each run to ensure size accuracy and to minimize run-to-run variation. PCR products were separated by electrophoresis in a 3.0 % (*w*/*v*) horizontal agarose gel using TBE buffer and ethidium bromide staining. The PCR patterns were recorded using the Image Gel-Doc^{TM} XR+ (BIO RAD).

The fragment analyses were performed on a ABI PRISM 3730 sequencer (Applied Biosystems, Foster City, CA, USA) and PCR products analyzed and sized with Peak Scanner Software version 1.0, using the GeneScan 500-LIZ as internal standard (Applied Biosystems, Foster City, CA, USA).

★ Escola Superior Agrária de Ponte de Lima (ESAPL)
◆ Quinta de Sergude, Direção Regional de Agricultura e Pescas do Norte (DRAPNorte)
✚ Centro de Experimentação Agrária de Tavira, Direção Regional de Agricultura do Algarve (DRAPAlgarve)

Fig. 1 Regions where local apple accessions included in this study were collected

Table 2 Microsatellite code, linkage group, PCR details, and size range (bp) of 16 SSR loci analyzed in this study

Locus	Linkage group number	Multiplex	Dye	Size range (bp)	Forward primer sequence 5′ → 3′	Reverse primer sequence 5′ → 3′
CH01f02[a]	12	A	FAM	159–214	ACCACATTAGAGCAGTTGAGG	CTGGTTTGTTTTCCTCCAGC
CH01f03b[a]	9	A	NED	138–183	GAGAAGCAAATGCAAAACCC	CTCCCCGGCTCCTATTCTAC
CH01h01[a]	17	B	FAM	97–143	GAAAGACTTGCAGTGGGAGC	GGAGTGGGTTTGAGAAGGTT
CH01h10[a]	8	A	PET	89–135	TGCAAAGATAGGTAGATATATGCCA	AGGAGGGATTGTTTGTGCAC
CH02c06[a]	2	C	PET	203–266	TGACGAAATCCACTACTAATGCA	GATTGCGCGCTTTTTAACAT
CH02c09[a]	15	B	VIC	231–257	TTATGTACCAACTTTGCTAACCTC	AGAAGCAGCAGAGGAGGATG
CH02c11[a]	10	B	PET	205–241	TGAAGGCAATCACTCTGTGC	TTCCGAGAATCCTCTTCGAC
CH02d08[a]	11	A	VIC	205–258	TCCAAAATGGCGTACCTCTC	GCAGACACTCACTCACTATCTCTC
CH03d07[a]	6	C	VIC	181–227	CAAATCAATGCAAAACTGTCA	GGCTTCTGGCCATGATTTTA
CH04c07[a]	14	B	VIC	95–141	GGCCTTCCATGTCTCAGAAG	CCTCATGCCCTCCACTAACA
CH04e05[a]	7	A	PET	177–229	AGGCTAACAGAAATGTGGTTTG	ATGGCTCCTATTGCCATCAT
CH05f06[a]	5	B	NED	165–191	TTAGATCCGGTCACTCTCCACT	TGGAGGAAGACGAAGAAGAAAG
CHVf1[b]	1	C	FAM	129–174	ATCACCACCAGCAGCAAAG	CATACAAATCAAAGCACAACCC
GD12[c]	3	C	NED	140–191	TTGAGGTGTTTCTCCCATTGGA	CTAACGAAGCCGCCATTTCTTT
GD147[c]	13	C	PET	127–160	TCCCGCCATTTCTCTGC	GTTTAAACCGCTGCTGCTGAAC
Hi02c07[d]	1	C	VIC	103–151	AGAGCTACGGGGATCCAAAT	GTTTAAGCATCCCGATTGAAAGG

[a] Liebhard et al. (2002)

[b] Vinatzer et al. (2004)

[c] Hokanson et al. (1998)

[d] Silfverberg-Dilworth et al. (2006)

Molecular analyses

Allelic profile of each accession was determined for each SSR locus. Considering that apple accessions can be polyploid, the software SPAGeDI v1.2g (Hardy and Vekemans 2002) was used to compute genetic information statistics, as this software supports analyses of datasets containing individuals with different ploidy levels. Genetic information statistics included the number of total alleles per locus (A), the number of rare allele per locus (B), alleles with frequency <0.05, effective number of alleles [$Ae=(\sum pi^2)^{-1}$, where p_i is the frequency of the ith allele], observed heterozygosity (Ho), and expected heterozygosity (He). The discrimination power (D) was also calculated as $D=1-\sum pi^2$ (Tessier et al. 1999), where pi represents the frequency of the ith genotype.

Genodive software package (Meirmans and Van Tienderen 2004) was used to carry out analysis of molecular variance (AMOVA) (Excoffier et al. 1992; Michalakis and Excoffier 1996) and to estimate the F-statistics (Weir and Cockerham 1984) inbreeding coefficient (F_{IS}) and goodness-of-fit (F_{IT}).

The set of 16 co-dominant SSRs amplicons were organized in a square matrix in which the codes "0" and "1" were used for allele absence and presence, respectively. The Jaccard coefficient (JC) was then computed based on the binary data using the XLSTAT-Pro software package, which does not consider the shared absence of a character as a similarity (Lo et al. 2009). Genotypes were clustered by the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Sokal and Michener 1958), and a dendrogram was constructed using the XLSTAT.

A factorial correspondence analysis (FCA) was performed using Genetix4 (Belkhir et al. 2004).

Software Identity 1.0 (Wagner and Sefc 1999) was used to determine the cumulative likehood ratios for possible parentages between cultivars.

In order to study population structure and assign individuals to populations based on the diploid SSR genotypes, a model-based Bayesian procedure was applied, implemented using the Structure software (Pritchard et al. 2000), by using the admixture model with unlinked loci and correlated allele frequencies. Computation of K (unknown) RPPs (reconstructed panmictic populations) of individuals testing K = 1 to 15, assuming that the sampled cultivars were from anonymous trees of unknown origin (using the options usepopinfo = 0, popflag = 0), was done. This clustering approach assigns individuals probabilistically to reconstructed panmitic populations based on genotype. Assignment of a cultivar to a RPP was based on a probability of membership qI of 80 %, while a lower probability meant that this accession could have several parental RPPs. Five replicate runs per K value were carried out, each consisting of a burning period length of 30,000 steps followed by 1,000,000 MCMC (Monte Carlo Markov Chain) replicates. The Structure software estimates the most likely number of clusters (K) by calculating the log probability of

data for each value of K. We used Structure Harvester (Earl and vonHoldt 2012) to assess the best K value supported by the data.

In addition to the studied accessions, available profiles for 13 SSR markers of 46 reference cultivars with different origins (22 modern and old international cultivars and 24 Spanish cultivars) representing a wide range of genetic diversity were also added to the study and included in the cluster analysis (Jaccard coefficient) described above (Fig. 1 supplementary data).

Results

Characterization of SSRs loci and genetic diversity

All SSRs were polymorphic and amplified in a single locus. Among the 95 accessions studied, 64 unique genotypes were found: 51 diploids and 13 putative triploids (Table 1 supplementary data).

The 16 SSR loci amplified a total of 210 alleles in the 64 unique genotypes (56 local accessions and eight references), varying from a minimum of 9 (CH02c09 and CH01f03b) to a maximum of 21 different alleles (CH03d07) per locus, with a mean number of total alleles per locus of 11.5 (Table 3). The mean of the effective number of alleles was 4.9, ranging from 2.8 (Hi02c07) to 9.9 (CH02c06) (Table 3). A total of 120 rare alleles were identified and, as indicated by the difference between the average value of total alleles per locus (11.5) and effective number of alleles (4.9), most alleles had frequencies lower than 0.05.

The set of local material showed a total of 188 alleles, with a mean number of total alleles per locus of 10.0, and an average of effective number of alleles of 4.8. A total of 96 rare alleles were identified in the set of local material, which indicates a substantial level of diversity.

For the overall set, the observed heterozygosity ranged from 0.537 (Hi02c07) to 0.990 (CH02c06), with an average of 0.668 across loci, while the expected heterozygosity ranged from 0.647 to 0.906, for the same two loci, respectively, with a mean value of 0.756 (Table 3). Regarding to the set of local material, the observed heterozygosity ranged from 0.529 (Hi02c07) to 0.989 (CHVf1 and CH01f02), with a mean value of 0.661. On the other hand, the expected heterozygosity ranged from 0.617 (CH04c05) to 0.906 (CH01f02), with an average of 0.749.

For the overall set, discrimination power varied from 0.642 (Hi02c07) to 0.899 (Ch02c06), with a mean value of 0.750. For the local material, the parameter ranged from 0.389 (CH04e05) to 0.898 (CH01f02), with a mean value of 0.742 (Table 3).

Table 3 Measures of genetic diversity at two different levels: overall set and set of local material. Number of alleles per locus (A), number of rare alleles (B), effective number of alleles (Ae), observed (Ho) and expected (He) heterozygosity, discrimination power (*D*), inbreeding coefficient (F_{IS}), and goodness-of-fit (F_{IT})

Locus	Overall set ($n = 64$)								Set of local material ($n = 56$)						
	A	B	Ae	Ho	He	*D*	F_{IS}	F_{IT}	A	B	Ae	Ho	He	*D*	F_{IS}
CH01f02	17	11	9.829	0.979	0.905	0.898	−0.089**	−0.105**	16	8	9.803	0.989	0.906	0.898	−0.105**
CH01f03b	9	6	3.860	0.863	0.747	0.741	−0.156*	−0.183*	8	4	3.822	0.885	0.745	0.738	−0.183**
CH01h01	13	6	7.057	0.800	0.865	0.858	0.051	0.050	11	4	6.819	0.793	0.861	0.853	0.050
CH01h10	14	10	3.786	0.642	0.741	0.736	0.108	0.102*	11	8	3.730	0.644	0.738	0.732	0.102
CH02c06	20	12	9.918	0.990	0.906	0.899	−0.094**	−0.098**	18	10	9.210	0.989	0.899	0.891	−0.098**
CH02c09	9	6	6.606	0.958	0.855	0.847	−0.094*	−0.124*	10	4	6.397	0.977	0.851	0.844	−0.124**
CH02c11	11	2	9.413	0.863	0.901	0.894	−0.001	0.008	11	2	8.910	0.862	0.895	0.888	0.008
CH02d08	13	8	4.981	0.853	0.805	0.799	−0.097*	−0.107*	11	6	4.842	0.851	0.800	0.794	−0.107*
CH03d07	21	15	8.276	0.937	0.885	0.879	−0.041	−0.076	19	14	8.140	0.954	0.885	0.877	−0.076*
CH04c07	13	6	7.301	0.853	0.870	0.863	0.097*	0.071*	12	7	7.058	0.874	0.866	0.858	0.071
CH04e05	14	10	2.992	0.632	0.671	0.666	0.102	0.031*	11	7	2.574	0.644	0.617	0.389	0.031
CH05f06	10	4	6.640	0.863	0.856	0.849	0.015	0.023	9	2	6.797	0.862	0.860	0.853	0.023
CHVf1	13	9	4.161	0.768	0.765	0.760	0.024	0.021	10	6	3.800	0.759	0.737	0.730	0.021
GD12	11	5	5.467	0.884	0.824	0.817	−0.059	−0.118	11	6	5.745	0.931	0.833	0.826	−0.118*
GD147	12	5	6.467	0.895	0.852	0.845	−0.022	−0.046	11	4	6.175	0.920	0.845	0.838	−0.046
Hi02c07	10	5	2.796	0.537	0.647	0.642	0.140*	0.126*	9	4	2.711	0.529	0.637	0.631	0.126*
Mean	11.5	5.5	4.927	0.668	0.756	0.750	−0.012	−0.031	10	4	4.765	0.661	0.749	0.742	−0.031
Total	210	120							188	96					

*$P < 0.05$; **$P < 0.01$

A broad variation depending on the locus was observed for inbreeding coefficient values in the whole set and local set, with a mean value of −0.012 and −0.031, respectively (Table 3). This situation was enhanced for the goodness-of-fit, with a mean value of −0.031.

Cultivar identification and cluster analysis

Comparison of SSR profiles revealed 19 groups of genotypes with the same SSR profile, involving 50 accessions (Table 1 supplementary data). Duplicates were identified in accessions within the same collection and among different collections. Some duplicates were expected, once correspond to accessions with the same or very similar denomination, namely within 'Bravo de Esmolfe,' 'Camoesa Fina,' 'Gigante do Douro,' 'Maçã S. João,' 'Maria Gomes,' 'Moleirinha,' 'Pardo Lindo,' 'Pé de Cera,' 'Pipo de Basto,' 'Porta da Loja, ' and 'Verdeal' accessions (Table 4 and Table 1 supplementary data). The remaining groups of duplicates comprised accessions with different names and, in some cases, also different geographical origin, suggesting that they were spread through grafting, resulting in several groups of synonyms (Table 4).

Comparison of SSR data with those published for material from a Galician germplasm collection (CIAM—'Centro de Investigaciones Agrarias de Magebondo') (Ramos-Cabrer et al. 2007) allowed the detection for the first time of several Portuguese/Spanish synonyms (Table 4, Fig. 1 supplementary data). Also, two unknown genotypes present in CIAM (unpublished data), 'Torres Agrelo-14' and 'Torres Agrelo-15,' correspond in the 16 SSR loci amplified with the samples 'Pêro de Mesa' and 'Belle de Boskoop' (Table 4), denoting in the last case the misnaming of the supposed international cultivar present in DRAPN collection.

Some presumed local accessions were identified as commercial cultivars, namely 'Vermelha de Refóios' identified as 'Delicious' (used as reference); the accessions 'Espriega 1,' 'Espriega Clara,' 'Espriega Parda,' and 'Espriega Verde' (with the same genotype) as 'Reineta' and 'Maçã de Outubro' and its synonym 'Setúbal' as 'Peas Good' (Table 4).

Four groups of homonyms were also observed, revealing accessions with similar designations but different SSR profiles. 'Camoesa Rosa 1 and 2,' 'Espriega 1 and 2,' 'Porta da Loja 1 and 3,' and 'Malápio 1 and 2' were found to be homonyms (Table 1 supplementary data).

Despite 'Marmela' being commonly designated as 'Três ao Prato de Covilhã,' the analyzed accessions 'Marmela' and 'Três ao Prato' revealed to be different genotypes.

Some similar designations like 'Riscadinha das Malhas' and 'Riscadinha da Feira' or 'Pêro Vermelho' and 'Pêro Tomate' could suggest duplicates. However, these accessions revealed different SSR profiles, corresponding to different genotypes (Table 1 supplementary data). Moreover, the designations 'Gilbarbeda Amarela' and 'Gilbarbeda Branca' suggest accessions differing in fruit color (yellow/'amarela' and white/'branca,' as indicated by accession names), mutants of the same variety. Given the inability of SSR markers to differentiate color mutant cultivars, as observed in other fruit species (Ferreira et al. 2016), 'Gilbarbeda' accessions were expected to have the same genotype. However, these two accessions revealed different SSR profiles, proving to be different genotypes (Table 1 supplementary data).

Within the Portuguese accessions, a total of 12 different genotypes were identified as putatively triploid ('Baionesa Ceboleira'; misnamed 'Belle de Boskoop'; 'D. Manuel'; ['Espelho/ Três ao Prato']; ['Espriega 1'/'Espriega Clara'/'Espriega Parda'/'Espriega Verde']; 'Espriega 2'; ['Gigante do Douro 1'/'Gigante do Douro 2'/Gigante D'oiro']; 'Gilbabeda Amarela'; 'Maçã da Pedralva'; ['Maria Gomes 1'/'Maria Gomes 2']; 'Martingil' and 'Porta da Loja 3') showing three alleles in four or more loci (Table 1 and Table 1 supplementary data). The reference 'Malling 9' (rootstock) showed a third allele only in the locus CH04c07.

The accessions 'Porta da Loja 1,' 2, 4, and 5 (diploid genotype) shared a complete set of alleles with the putative triploid genotypes 'Maria Gomes 1'/'Maria Gomes 2', 'Porta da Loja 3,' and the four genotypes identified as Reineta ('Espriega 1,' 'Espriega Clara,' 'Espriega Parda,' and 'Espriega Verde'), and are likely candidates to be the haploid donor of these cultivars. Moreover, 'Espriega 2' shared alleles in every loci with the mentioned Espriegas (synonymy of Reineta) indicating possible relation by hybridization. Also, 'Pêro da Lixa' and 'Camoesa Fina' (both diploid genotypes) shared at least one allele per locus, in 15 of the 16 loci analyzed, with the putative triploid accessions 'Gigante do Douro' and the misnamed 'Belle de Boskoop,' respectively (Table 1 supplementary data).

Eight main clusters (I–VIII) were obtained at a lower 0.20 JC, comprising the cluster II the highest number of local accessions (Fig. 2).

The cluster I included accessions from 'Baionesa Ceboleira' through 'Camoesa Rosa 2.' Inside this cluster, two putative triploid accessions, 'Baionesa Ceboleira' and 'D. Manuel,' shared at least 50 % of the alleles, with a JC over 0.5, grouping with the genotype 'Pipo de Basto,' an important local Portuguese cultivar. 'Maçã Ácida' group was also in this first cluster and includes four accessions comprising two different genotypes: 'Maçã Ácida' and its synonyms and 'Maçã da Pedralva,' which shared 50 % of the alleles, suggesting possible parental relationships (Fig. 2).

Cluster II included accessions from 'Reineta Espanhola' to 'Riscadinha da Feira.' This cluster revealed five small groups that were named 'Espriega,' 'Porta da Loja,' 'Verdeal,' 'Camoesa,' and 'Malápio.'

Cluster IIa included the reference accession 'Delicious' (genotype of 'Vermelha de Refóios'), which clustered together with 'Maçã do Carrascalinho,' 'D. Emília,' and 'Rajada

Table 4 List of local, Spanish, and international synonyms

Synonyms		
Local accessions	Spanish accessions	International cultivars
'Arouca'	'Camba'[a]	
'Baionesa Ceboleira'	'De Invierno'[b]	
'Belle de Boskoop' misnamed	'Torres Agrelo-15'[b]	
'Camoesa de Coura'; 'Camoesa Pedra'; 'Pêro de Coura'		
'D. Manuel'	'Tres En Cuna Y Otra'[a]	
'Espelho'; 'Três ao Prato'	'Do Apostol'[a]	
'Espriega 1'; 'Espriega Clara'; 'Espriega Parda'; 'Espriega Verde'		'Reineta'[a]
'Focinho de Burro'; 'Malápio 2'; 'Malápio do Norte'		
'Maçã Ácida'; 'Maçã Cigana'; 'Pedregal'		
'Maçã de Outubro'; 'Setúbal'		'Peas Good'[a]
'Maçã de S. João'; 'S. João'	'Roja de Verano'[a]	
'Malápio 1'; 'Malápio de Pé Curto'; 'Malápio Pé de Porco'		
'Marmela'; 'Pêro Limão'; 'Verdeal 1'; 'Verdeal 2'		
'Pêro de Mesa'	'Torres Agrelo-14'[b]	
'Pêro Vermelho'	'Sangre De Toro'[a]	
'Pipo de Basto 1'; 'Pipo de Basto 2'	'Peras Camba'[b]	
'Refóios'	'Blanca Plana'[a]	
'S. Miguel'	'Santiaguesa'[b]	
'Vermelha de Refóios'		'Delicious'

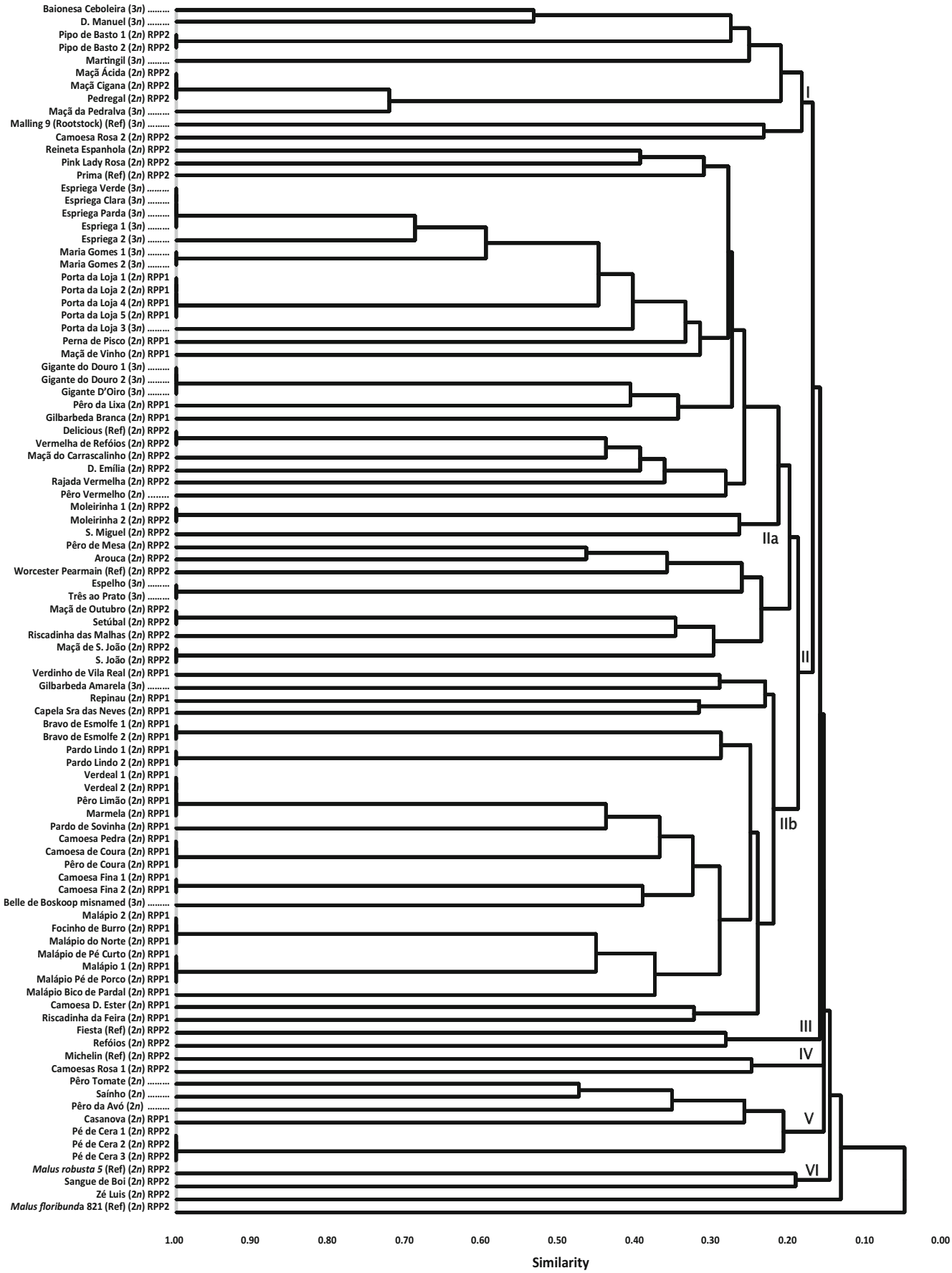
[a] Synonym profiles in Ramos-Cabrer et al. (2007) present at CIAM-'Centro de Investigaciones Agrarias de Mabegondo'

[b] Unpublished data

Vermelha.' Another accession included as reference, 'Worcester Pearmain,' was also included in this subcluster, clustering together with local accessions 'Pêro de Mesa' and 'Arouca.' Cluster IIa also contained three sets of accessions: the accessions with the same genotype, 'Espriega 1,' 'Espriega Clara,' 'Espriega Verde' and 'Espriega Parda,' 'Maria Gomes 1 and 2,' and the accession 'Espriega 2,' which revealed a similarity higher than 0.5, suggesting possible parental relationships (Fig. 2) and the 'Porta da Loja' group which includes the four accessions with the same genotype 'Porta da Loja 1, 2, 4, and 5' and the accession 'Porta da Loja 3.' 'Porta da Loja' is a regional cultivar, mainly confined to Northwest Portugal, where it is much appreciated due to its different taste, flavor, consistency, and large conserving capacity. On the other hand, the subgroup designated as cluster IIb (Fig. 2) included 'Verdeal' group, which contained 'Verdeal 1 and 2' and its synonymy accessions 'Marmela' and 'Pêro Limão,' and also 'Pardo de Sovinha' with a similarity close to 0.5 JC (Fig. 2). On the other hand, 'Camoesa' group included the three accessions with the same genotype 'Camoesa de Coura,' 'Camoesa Pedra,' and 'Pêro de Coura,' the accessions 'Camoesa Fina 1 and 2' and the misnamed 'Belle de Boskoop,' this last sub-grouped with 'Camoesa Fina 1 and 2.' Furthermore, 'Malápio' group included three different genotypes: the accessions with the same genotype 'Malápio de Pé Curto,' 'Malápio Pé de Porco,' and 'Malápio 1'; a second genotype found in the accessions 'Focinho de Burro,' 'Malápio do Norte,' and 'Malápio 2'; and a third accession with a different genotype, 'Malápio Bico de Pardal.' Cluster IIb also included a small group of four genotypes, 'Verdinho de Vila Real,' 'Gilbarbeda Amarela,' 'Repinau,' and 'Capela Sr[a] das Neves,' and the group of accessions corresponding to 'Bravo de Esmolfe' and 'Pardo Lindo,' which represent two regional cultivars with a high commercial value in Portugal, being 'Bravo de Esmolfe' cultivated under the 'Protected Denomination of Origin' (DOP) (Fig. 2).

Clusters III and IV included four accessions (Fig. 2), two in cluster III (reference 'Fiesta' and 'Refóios') and another two in cluster IV (reference 'Michelin' and 'Camoesa Rosa 1'). Furthermore, cluster V comprised five different genotypes, two of them sharing more than 50 % of the alleles, 'Pêro

Fig. 2 Dendrogram of genetic similarity generated using the Jaccard coefficient and the UPGMA method for the 95 accessions analyzed with 16 SSR loci. The reconstructed populations, RPP1 and RPP2, defined using Structure (Pritchard et al. 2000) for each diploid genotype clustered with $qI > 80$ are indicated. The ploidy level is indicated in *brackets*


Baionesa Ceboleira (3n) .....
D. Manuel (3n) .....
Pipo de Basto 1 (2n) RPP2
Pipo de Basto 2 (2n) RPP2
Martingil (3n) .....
Maçã Ácida (2n) RPP2
Maçã Cigana (2n) RPP2
Pedregal (2n) RPP2
Maçã da Pedralva (3n) .....
Malling 9 (Rootstock) (Ref) (3n) .....
Camoesa Rosa 2 (2n) RPP2
Reineta Espanhola (2n) RPP2
Pink Lady Rosa (2n) RPP2
Prima (Ref) (2n) RPP2
Espriega Verde (3n) .....
Espriega Clara (3n) .....
Espriega Parda (3n) .....
Espriega 1 (3n) .....
Espriega 2 (3n) .....
Maria Gomes 1 (3n) .....
Maria Gomes 2 (3n) .....
Porta da Loja 1 (2n) RPP1
Porta da Loja 2 (2n) RPP1
Porta da Loja 4 (2n) RPP1
Porta da Loja 5 (2n) RPP1
Porta da Loja 3 (3n) .....
Perna de Pisco (2n) RPP1
Maçã de Vinho (2n) RPP1
Gigante do Douro 1 (3n) .....
Gigante do Douro 2 (3n) .....
Gigante D'Oiro (3n) .....
Pêro da Lixa (2n) RPP1
Gilbarbeda Branca (2n) RPP1
Delicious (Ref) (2n) RPP2
Vermelha de Refóios (2n) RPP2
Maçã do Carrascalinho (2n) RPP2
D. Emília (2n) RPP2
Rajada Vermelha (2n) RPP2
Pêro Vermelho (2n) .....
Moleirinha 1 (2n) RPP2
Moleirinha 2 (2n) RPP2
S. Miguel (2n) RPP2
Pêro de Mesa (2n) RPP2
Arouca (2n) RPP2
Worcester Pearmain (Ref) (2n) RPP2
Espelho (3n) .....
Três ao Prato (3n) .....
Maçã de Outubro (2n) RPP2
Setúbal (2n) RPP2
Riscadinha das Malhas (2n) RPP2
Maçã de S. João (2n) RPP2
S. João (2n) RPP2
Verdinho de Vila Real (2n) RPP1
Gilbarbeda Amarela (3n) .....
Repinau (2n) RPP1
Capela Sra das Neves (2n) RPP1
Bravo de Esmolfe 1 (2n) RPP1
Bravo de Esmolfe 2 (2n) RPP1
Pardo Lindo 1 (2n) RPP1
Pardo Lindo 2 (2n) RPP1
Verdeal 1 (2n) RPP1
Verdeal 2 (2n) RPP1
Pêro Limão (2n) RPP1
Marmela (2n) RPP1
Pardo de Sovinha (2n) RPP1
Camoesa Pedra (2n) RPP1
Camoesa de Coura (2n) RPP1
Pêro de Coura (2n) RPP1
Camoesa Fina 1 (2n) RPP1
Camoesa Fina 2 (2n) RPP1
Belle de Boskoop misnamed (3n) .....
Malápio 2 (2n) RPP1
Focinho de Burro (2n) RPP1
Malápio do Norte (2n) RPP1
Malápio de Pé Curto (2n) RPP1
Malápio 1 (2n) RPP1
Malápio Pé de Porco (2n) RPP1
Malápio Bico de Pardal (2n) RPP1
Camoesa D. Ester (2n) RPP1
Riscadinha da Feira (2n) RPP1
Fiesta (Ref) (2n) RPP2
Refóios (2n) RPP2
Michelin (Ref) (2n) RPP2
Camoesas Rosa 1 (2n) RPP2
Pêro Tomate (2n) .....
Saínho (2n) .....
Pêro da Avó (2n) .....
Casanova (2n) RPP1
Pé de Cera 1 (2n) RPP2
Pé de Cera 2 (2n) RPP2
Pé de Cera 3 (2n) RPP2
Malus robusta 5 (Ref) (2n) RPP2
Sangue de Boi (2n) RPP2
Zé Luis (2n) RPP2
Malus floribunda 821 (Ref) (2n) RPP2
I
IIa
II
IIb
III
IV
V
VI
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Similarity

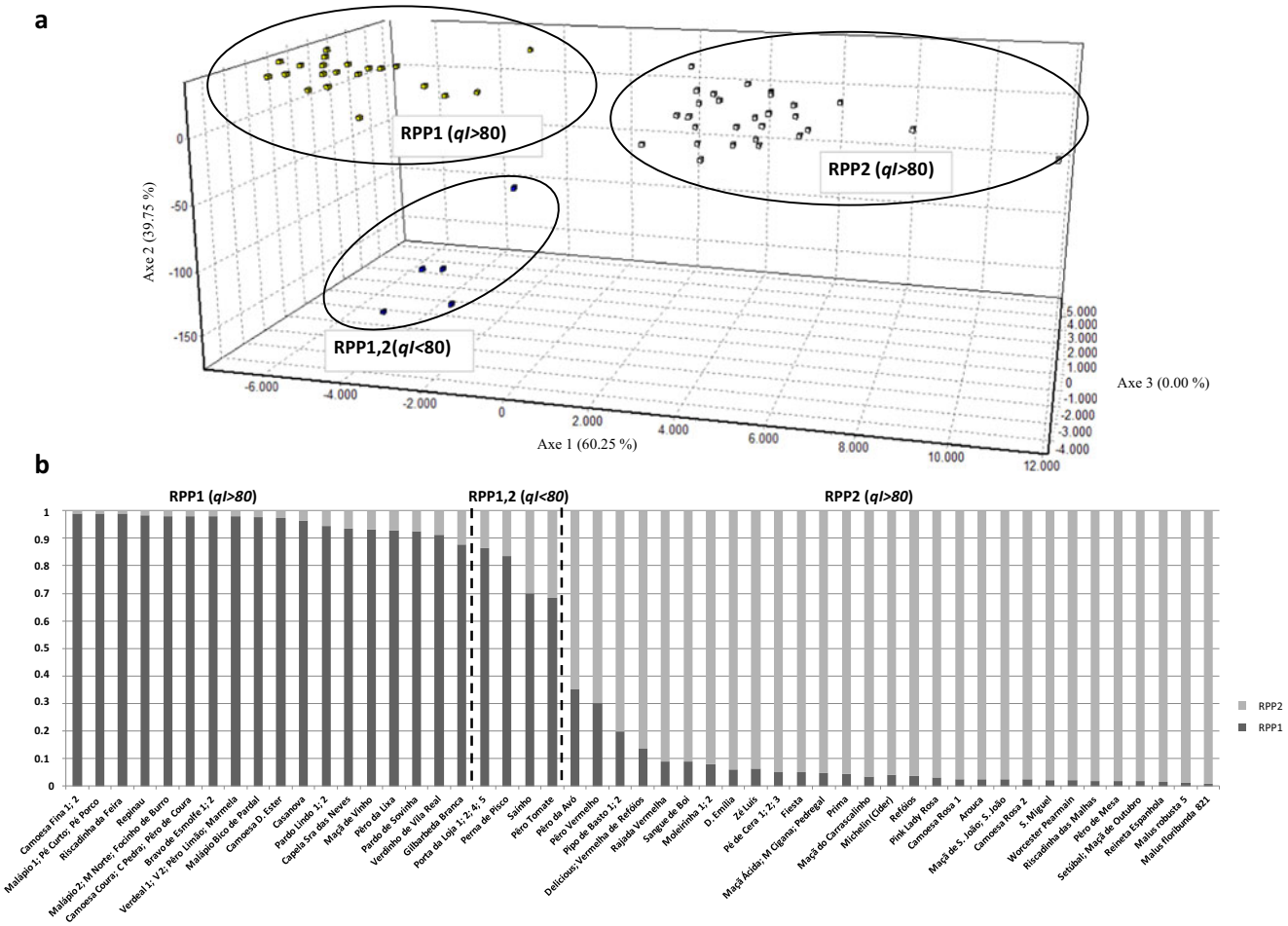


Fig. 3 **a** Multivariate analysis (factorial correspondence analysis) and **b** hierarchal structure analysis obtained using Structure (Pritchard et al. 2000), with K = 2, based on data for 16 SSRs in 51 unique diploid genotypes *Malus* spp.

Tomate' and 'Saínho,' with a similarity higher than 0.45 JC. The last, cluster VI comprised two accessions, the reference '*Malus robusta* 5' and 'Sangue de Boi.'

The accession 'Zé Luís' and the reference '*Malus floribunda* 821' correspond to quite distant genotypes, being completely different from the remaining material studied.

The results obtained with the comparative analysis performed with 46 reference cultivars revealed a clear cluster with most of the international cultivars, completely differentiated from the majority of the Portuguese accessions (Fig. 1 supplementary data), namely several accessions included in cluster IIb in Fig. 2. The Spanish reference cultivars grouped thoroughly mixed with the Portuguese accessions (Fig. 1 supplementary data).

Genetic structure and differentiation

A model-based Bayesian approach by using Structure (Pritchard et al. 2000) was conducted with the 16 SSRs on the 51 unique diploid genotypes, including local and reference cultivars. The most likely number of clusters (k) by calculating the log probability of data, ln[Pr(X/K)], for K = 2, estimated by Structure Harvester (Earl and vonHoldt 2012), which corresponded to strong differentiation of two main groups, one clustering 24 genotypes (RPP1) and a second one with 27 (RPP2), all of them with qI > 80 % (Fig. 3a, b).

RPP1 grouped only Portuguese diploid genotypes, 14 out 24 in cluster IIb, differentiated at a JC < 0.2. Between RPP1 and 2 (qI < 80) were found four intermediate genotypes from Algarve Province (Fig. 3b). RPP2 included all the diploid reference genotypes and 20 local Portuguese genotypes (Fig. 3a, b). A smaller peak at K = 4 was found when the posterior K statistics of Evanno et al. (2005) were applied, revealing substructure in RPP2 (Fig. 4). RPP2a, RPP2b, and RPP2c included 21 genotypes out of 27 from RPP2 when K = 2 with a qI > 80 plus one from RPP1 when K = 2 with a qI < 80. RPP2a, included five Portuguese accessions, *Malus floribunda* 821 and *Malus robusta* 5, as so the accession Michelin (Cider). RRP2b grouped six Portuguese accessions with the reference 'Worcester

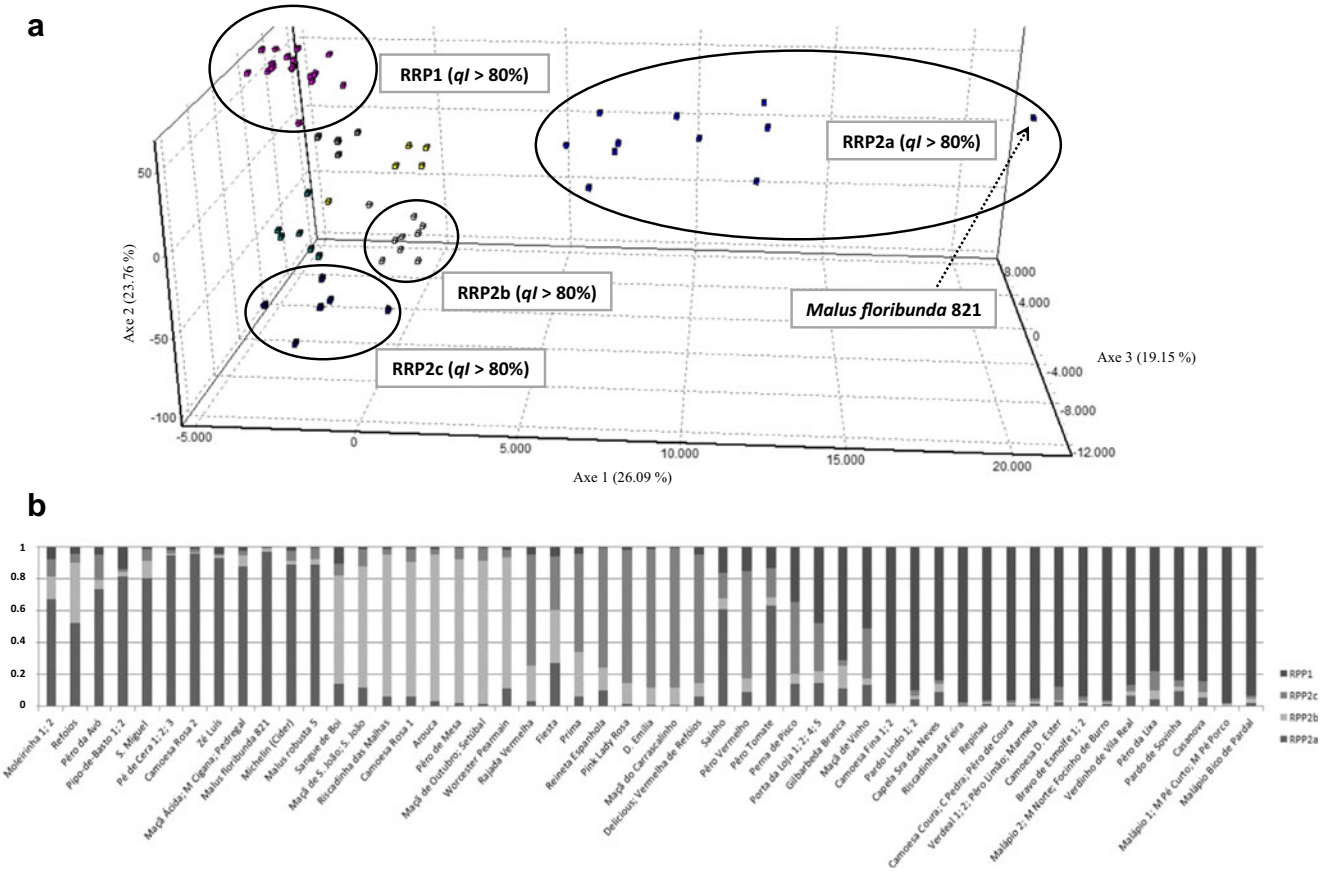


Fig. 4 **a** Multivariate analysis (factorial correspondence analysis) and **b** hierarchal structure analysis obtained using Structure (Pritchard et al. 2000), with K = 4, based on data for 16 SSRs in 51 unique diploid genotypes *Malus* spp.

Pearmain.' Finally, RPP2c had five Portuguese accessions related with 'Delicious' (Fig. 4).

FCA computed on all the 51 diploid accessions displayed two clear distinguish groups, RPP1 and RRP2 (Fig. 3a), the last group subdivided into RRP2a, RPP2b, and RPP2c (Fig. 4a), confirming the genetic discrimination results determined by the Structure analyses with K = 2 and K = 4 (Figs. 3 and 4).

The result of AMOVA analysis for K = 2, when carried out on accessions with $qI > 80$, indicated that the proportion of genetic differentiation within the two main groups accounted for most of the molecular variance (92.6 %) and only 7.4 % accounted for variation among the groups. When we performed the same analyses for K = 4 with a $qI > 80$, variation among the RPPs increased up to 13.2 %, with the highest pairwise F_{st} obtained between the Portuguese accessions (RPP1) and the genotypes related to 'Golden Delicious' in RPP2c (11.2 %, $P < 0.001$), and slighter lower with RPP2b related to 'Worcester Pearmain' (10.3 %, $P < 0.001$). The lower pairwise F_{st} was observed between RPP1 and RPP2a (5.9 %, $P < 0.001$), the group that include pure species and the accession used for cider production 'Michelin.'

Discussion

SSR markers and genetic diversity

The correct identification of accessions emphasize the importance of the germplasm collections studies using powerful tools, such as molecular markers, in order to avoid redundancy in collections, reduce the management costs, and distribute true-to-type cultivars to nurseries for propagation. The knowledge of the genetic diversity among accessions is a key point for the efficient use and conservation of endangered regional material.

All SSR loci analyzed in this study displayed a high degree of polymorphism with nine to 21 alleles per locus. The overall allelic diversity showed by the set of 16 SSR markers used reveals a high genetic variation in the apple germplasm evaluated, evidence of richness and singularity that can be found yet within germplasm collections representative of ancient cultivars and highlighting their importance as repositories of germplasm diversity.

When compared to other wide scale apple genetic diversity studies, the average number of alleles per locus (11.5) was similar to that reported by Pereira-Lorenzo et al. (2008) and Gharghani et al. (2009), in local and commercial genotypes,

but slightly lower than those reporter by van Treuren et al. (2010). Furthermore, regarding to the heterozygosity in local genotypes (0.75), this level of diversity is in agreement with the level of polymorphism, 0.73, reported for a similar number of accessions from the regions Basque Country, Asturias, and Galicia in Northern and Northwestern Spain (Pereira-Lorenzo et al. 2008), but slightly lower than that obtained in sets of local cultivars from Northeastern Spain, 0.82 (Urrestarazu et al. 2012) and 0.80 (Pina et al. 2014), in Italy, 0.80 (Liang et al. 2015), and Portuguese Azores Islands, 0.81, (Foroni et al. 2012).

The broad variation, depending on the locus observed, for F_{IS} and F_{IT} values in the whole set of cultivars, suggests a lack of close groups of related individuals within the germplasm analyzed, which is in accordance with the forced allogamy due to the self-incompatibility system of *Malus* × *domestica*, as also observed by Liang et al. (2015) for Italian apple germplasm.

Collection management and genetic relationships

Grafting is the main method of propagating selected clones in fruit trees. The average clonality in this study was high (30 %), denoting the difficulty in identifying the different accesssions morphologically. This value was similar to the clonality found in Italian apple cultivars (34 %) (Liang et al. 2015) and for a Dutch apple collection (32 %) (van Treuren et al. 2010), but lower than obtained in apple collections from Northeastern Spain (43.5 %) (Urrestarazu et al. 2012).

Regarding to the 64 unique genotypes identified, 19 groups of genotypes with the same SSR profile, involving 50 accessions, were reported as duplicates and appear within and among the three collections analyzed (Table 1 supplementary data). Some duplicates were expected due to the identical denominations. However, several duplicated genotypes were identified as synonyms (Table 4). Besides synonyms within and among collections, other synonyms between Portuguese and Spanish cultivars were also reported for the first time (Table 4), indicating the close genetic relationship for this crop in both countries as it happens in other species such as grapevine (Castro et al. 2011; Ferreira et al. 2015).

Conversely, four cases of homonyms have been as well identified. Misidentifications and mislabeling are considered an important cause of homonyms in germplasm collections. These kinds of misidentifications were reported in the Italian and Spanish apple germplasm (Urrestarazu et al. 2012; Liang et al. 2015), being these occurrences the result of modifications in the cultivar's name after the introduction in different collections and/or countries.

Among the 64 unique genotypes, 51 were diploids and 13 (≈20 %) putative triploids. Higher values, 24 and 29 %, were detected in Spanish germplasm by Urrestarazu et al. (2012) and Ramos-Cabrer et al. (2007), respectively. The amplification of three fragments in a single locus is not a proof of the triploid status, once they can be produced by duplication events, somatic mutations generating chimerical or mosaic states, giving rise to unique, not real alleles. However, these hypotheses are not being considered, once the third allele observed on the putative triploid accessions also exists in other diploid and putative triploid accessions, with the exception of the unique third allele observe in the locus CH04c07 in the reference 'Malling 9' (rootstock). Despite multiple alleles have been observed at several loci, indicating putative triploids, these findings need confirmation by chromosome count or flow cytometry.

Several diploid and putative triploid genotypes were found to share alleles in a high number of loci, namely within 'Porta da Loja,' 'Maria Gomes,' and the 'Espriega' accessions identified as 'Reineta,' suggesting several hybridization events. This kind of hybridization has been previously described in apple, namely for two cultivars from Netherlands, explaining how 'Boskoop' (triploid) was originated from 'Reinette de Caux' (Ramos-Cabrer et al. 2007).

Although the high level of variation observed in this study, several duplications within collections were observed and must be avoided due to the high maintenance cost of field collections; however, a certain level of duplication provides a safety backup system. On that way, the results obtained can be used in the future in the re-organization and composition of the collections towards improvement of their management efficiency.

The comparative analysis performed with 46 reference cultivars revealed a clear differential clustering of several international reference cultivars from most of Portuguese accessions (Fig. 1 supplementary data). However, some Portuguese accessions grouped together with reference cultivars which may indicate the introgression of international cultivars with few of these local accessions. Moreover, a common ancestral origin of Portuguese and Spanish cultivars is suggested by the mixed clustering of the reference Spanish cultivars and the Portuguese accessions (Fig. 1 supplementary data), denoting the uniqueness of this genetic pool as a possible origin of the autochthonous cultivars 'Camoesa' and 'Pero.'

Genetic structure and differentiation

Bayesian clustering analyses have proven to be powerful tools to analyze the genetic structure in tree species, namely apple. However, to our knowledge, this method was mainly used in wild populations of apple species and only more recently has been used to resolve the genetic structure in collections of domesticated apple (Pereira-Lorenzo et al. 2008; Urrestarazu et al. 2012; Pina et al. 2014).

This study revealed the existence of a clear structure, not only among the references but also within the local accessions analyzed. The first partition of genetic variation showed two strongly differentiated groups, RPP1 and RPP2. The major group RPP1 only comprises local accessions corresponding to a separate gene pool, independent from RPP2 that contains the reference accessions. This partition suggests the uniqueness of this local material, some of it with references dating back to the beginning of the nineteenth century (Mota 1919; Natividade 1922; Lima 1932). Genetic relationships found in this study between accessions 'Porta da Loja 1,' 2, 4, and 5 (diploid genotype) with the four genotypes identified as the commercial cultivar Reineta ('Espriega1,' 'Espriega Clara,' 'Espriega Parda,' and 'Espriega Verde') could explain the origin of this RPP1 by an early hybridization occurred in Portugal. The presence of some local accessions in RPP2 could be partially explained by possible incursions of foreign into local material, resulting in new cultivars by hybridization, that could have produced lineages that were subsequently propagated (Pereira-Lorenzo et al. 2010, 2011). A closer examination of accessions revealed that in RPP2 four local accessions, 'Camoesa Rosa 2,' 'Arouca,' 'Refóios,' and 'Camoesa Rosa 1' were allocated with 'Malling 9' (rootstock), 'Worcester Pearmain,' 'Fiesta,' and 'Michelin' (cider), respectively, suggesting that they can share more genetic information with these reference accessions rather than with the Portuguese germplasm analyzed. The group RPP1,2 ($qI < 80$ %) represents an intermediate group between the reconstructed populations RPP1 and RPP2, and includes four accessions ('Saínho,' 'Pêro Tomate,' 'Pêro da Avó,' and 'Pêro Vermelho'), all of them from Algarve region and with similar morphological traits (red skin and pear shape) (Fig. 3).

Moreover, one supposed local accession ('Vermelha de Refóios') revealed to be the reference commercial cultivar 'Delicious' highlighting the wide spreading of commercial cultivars. Also, the UPGMA clustering showed the admixture of two local accessions 'Sangue de Boi' and 'Zé Luís,' with the non-*domestica* reference material, *Malus floribunda* 821 and *Malus robusta* 5, revealing that those accessions may not be domesticated apple *Malus* × *domestica* Borkh.

The substructure revealed in RPP2 when K = 4 showed three subgroups, RPP2a, RPP2b, and RPP2c. RPP2a included five Portuguese accessions, plus '*Malus floribunda*,' '*Malus robusta*,' as well as the reference accession 'Michelin' (cider) (Fig. 4). The production of cider in the limits of DRAPN, Northwest Portugal, was of great importance until eleventh century, even higher than wine (Pereira 1962). Moreover, RPP2b grouped six Portuguese accessions with the reference 'Worcester Pearmain,' which is a seedling of Devonshire Quarrenden, obtained in 1874, suggesting the antiquity of the group. The last group, RPP2c, included five Portuguese accessions and the reference 'Delicious,' suggesting some kind of relationship (Fig. 4).

The overall F_{ST} value of 0.074 with K = 2 suggests a moderated differentiation between groups, which is in agreement with F_{ST} values obtained in previous studies of genetic diversity and structure of local apple cultivars, mainly from Northeastern Spain, 0.076 and 0.070 (Urrestarazu et al. 2012; Pina et al. 2014). Liang et al. (2015) observed an overall F_{ST} value of 0.056 among the four subgroups obtained with a core collection of apple cultivars from Italy, suggesting the existence of a weaker population differentiation among the subgroups than the differentiation obtained in the present study. However, a stronger differentiation was found between the Portuguese group (RPP1) and the genotypes related to 'Golden Delicious' accession in RPP2c (11.2 %, $P < 0.001$) indicating an older origin of this group from 'Reineta' and as a possible origin of cultivars 'Camoesa' and 'Pero.'

Conclusions

As far as we are aware, this is the first molecular characterization reported for the mainland Portuguese apple germplasm. This study comprised 87 accessions, plus eight accessions included as references. Additionally, a large-scale comparison of the local Portuguese material with a high number of reference international cultivars (including modern and old cultivars) with different origins helped to better understand the genetic variation of Portuguese apple ancient cultivars, allowing a fine delineation of this unique genetic pool. The initial pool of 87 Portuguese accessions was reduced to a total of 56 unique apple genotypes, including 12 putative triploids, by using a set of 16 polymorphic SSR markers.

By using a model-based Bayesian clustering method, the accessions were structured in two main groups (RPP1 and RPP2). The clustering in two main groups resulting from the structure analysis was confirmed by factorial correspondence analysis and AMOVA approaches. The differential grouping of the reference cultivars (in RPP2) and most of Portuguese accessions (in RPP1) suggests that the Portuguese germplasm analyzed in this work represents a differentiated genetic pool of apple cultivars derived from ancient hybridization with 'Reineta' before the beginning of the nineteenth century, when there's the first written references in Portugal, and probably the origin of ancient cultivars 'Camoesa' and 'Pero.'

Present results highlight the relevance of autochthonous apple germplasm as a reservoir of genetic diversity, being the material analyzed a good example of genetic distinctness. A further integration of the data from other collections with different geographic regions in Portugal, Europe, or worldwide will make possible to preserve and exploit the whole apple gene pool.

Acknowledgments This work was supported by European Investment Funds by FEDER/COMPETE/POCI–Competitiveness and Internationalization Operational Programme, under the Project POCI-01-0145-FEDER-006958, National Funds by FCT–Portuguese Foundation for Science and Technology, under the project UID/AGR/04033/2013 and the scholarship SFRH/BD/96400/2013 and Galicia Norte Portugal European Grouping of Territorial Cooperation (GNP-EGTC) funds supporting an internship at University of Santiago de Compostela, under the IACOBUS program. The authors acknowledge the "Research Institute of Horticulture and Seeds" INRA, Agrocampus Ouest, University of Angers, for providing the eight reference DNA samples.

Compliance with ethical standards

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

Data Archiving Statement The authors declare that all the work described in this manuscript followed the standard Tree Genetics and Genomes policy. Our results are based on an original database of genetic profiles not submitted to a public database. Moreover, genetic profiles of our manuscript have been included as supplemental file.

References

Barata AM, Bettencourt E, Santos A et al (2008) The state of ex situ management. Instituto Nacional de Recursos Biológicos, Oeiras

Bassil N, Hummer K, Postman J et al (2009) Nomenclature and genetic relationships of apples and pears from Terceira Island. Genet Resour Crop Evol 56:339–352. doi:10.1007/s10722-008-9369-z

Belkhir K, Borsa P, Chikhi L et al (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations

Castro I, Martín JP, Ortiz JM, Pinto-Carnide O (2011) Varietal discrimination and genetic relationships of *Vitis vinifera* L. cultivars from two major Controlled Appellation (DOC) regions in Portugal. Sci Hortic 127:507–514. doi:10.1016/j.scienta.2010.11.018

Cornille A, Giraud T, Smulders MJM et al (2014) The domestication and evolutionary ecology of apples. Trends Genet 30:57–65. doi:10.1016/j.tig.2013.10.002

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

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. doi:10.1111/j.1365-294X.2005.02553.x

Evans KM, Fernández-Fernández F, Govan C (2009) Harmonising fingerprinting protocols to allow comparisons between germplasm collections—*Pyrus*. Acta Hortic 103–106. doi:10.17660/ActaHortic.2009.814.10

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491. doi:10.1007/s00424-009-0730-7

Ferreira V, Pinto-Carnide O, Mota T et al (2015) Identification of minority grapevine cultivars from Vinhos Verdes Portuguese DOC Region. Vitis 54:53–58

Ferreira V, Fernandes F, Pinto-Carnide O et al (2016) Identification of *Vitis vinifera* L. grape berry skin color mutants and polyphenolic profile. Food Chem 194:117–127. doi:10.1016/j.foodchem.2015.07.142

Foroni I, Baptista C, Monteiro L et al (2012) The use of microsatellites to analyze relationships and to decipher homonyms and synonyms in Azorean apples (*Malus* × *domestica* Borkh.). Plant Syst Evol 298:1297–1313. doi:10.1007/s00606-012-0637-1

Gharghani A, Zamani Z, Talaie A et al (2009) Genetic identity and relationships of Iranian apple (*Malus* × *domestica* Borkh.) cultivars and landraces, wild *Malus* species and representative old apple cultivars based on simple sequence repeat (SSR) marker analysis. Genet Resour Crop Evol 56:829–842. doi:10.1007/s10722-008-9404-0

Hardy OJ, Vekemans X (2002) Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620. doi:10.1046/j.1471-8286.2002.00305.x

Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* borkh. core subset collection. Theor Appl Genet 97:671–683. doi:10.1007/s001220050943

Hummer KE, Janick J (2009) Rosaceae: taxonomy, economic importance, genomics. In: Folta KM, Gardiner SE (eds) Genetics and Genomics of Rosaceae. Springer, New York, pp 1–17

Janick J, Cummins JN, Brown SK, Hemmat M (1996) Apples. In: Janick J, Moore JN (eds) Fruit breeding. Wiley, New York, pp 1–76

Liang W, Dondini L, De Franceschi P et al (2015) Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. Plant Mol Biol Report 33:458–473. doi:10.1007/s11105-014-0754-9

Liebhard R, Gianfranceschi L, Koller B, et al (2002) Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). Mol Breed 10:217–241. doi:10.1023/A:1020525906332

Lima JJ (1932) Método de caracterização das variedades de maçãs portuguesas ou tidas como tais. In: 2º congresso nacional de pomologia. Sociedade Pomológica Portuguesa, Lisboa, pp 38–335

Lo EYY, Stefanović S, Dickinson TA (2009) Population genetic structure of diploid sexual and polyploid apomictic hawthorns (Crataegus; Rosaceae) in the Pacific Northwest. Mol Ecol 18:1145–1160. doi:10.1111/j.1365-294X.2009.04091.x

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4:792–794. doi:10.1111/j.1471-8286.2004.00770.x

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064

Mota MC (1919) Catálogo geral de plantas, sementes e outros artigos. A Intermediária Limitada, Porto

Natividade JV (1922) A região de Alcobaça—Algumas notas para o estudo da sua agricultura, população e vida rural

Noiton DAM, Alspach PA (1996) Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J Am Soc Hortic Sci 121:773–782

Pereira BE (1962) A sidra no Norte de Portugal <Tratados de Antropologia e Etnologia>, Fase 3-4, Vol. XVIII, Porto, Soc, Portuguesa de Antropologia e Etnologia 1962, p. 362 e 366

Pereira-Lorenzo S, Ramos-Cabrer AM, González-Díaz AJ, Díaz-Hernández MB (2008) Genetic assessment of local apple cultivars from La Palma, Spain, using simple sequence repeats (SSRs). Sci Hortic 117:160–166. doi:10.1016/j.scienta.2008.03.033

Pereira-Lorenzo S, Costa RML, Ramos-Cabrer AM et al (2010) Variation in grafted European chestnut and hybrids by microsatellites reveals

two main origins in the Iberian Peninsula. Tree Genet Genomes 6: 701–715. doi:10.1007/s11295-010-0285-y

Pereira-Lorenzo S, Costa RML, Ramos-Cabrer AM et al (2011) Chestnut cultivar diversification process in the Iberian Peninsula, Canary Islands, and Azores. Genome 54:301–315. doi:10.1139/g10-122

Pina A, Urrestarazu J, Errea P (2014) Analysis of the genetic diversity of local apple cultivars from mountainous areas from Aragon (Northeastern Spain). Sci Hortic 174:1–9. doi:10.1016/j.scienta.2014.04.037

Potts SM, Han Y, Khan MA et al (2012) Genetic diversity and characterization of a core collection of *Malus* germplasm using simple sequence repeats (SSRs). Plant Mol Biol Report 30:827–837. doi:10.1007/s11105-011-0399-x

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

Queiroz A, Assunção A, Ramadas I et al (2015) Molecular characterization of Portuguese pear landraces (*Pyrus communis* L.) using SSR markers. Sci Hortic 183:72–76. doi:10.1016/j.scienta.2014.11.016

Ramos-Cabrer AM, Díaz-Hernández MB, Pereira-Lorenzo S (2007) Morphology and microsatellites in Spanish apple collections. J Hortic Sci Biotechnol 82:257–265

Robinson JP, Harris SA, Juniper BE (2001) Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant Syst Evol 226:35–58. doi:10.1007/s006060170072

Silfverberg-Dilworth E, Matasci CL, Van De Weg WE, et al (2006) Microsatellite markers spanning the apple (*Malus x domestica* Borkh.) genome. Tree Genet Genomes 2:202–224. doi:10.1007/s11295-006-0045-1

Sokal R, Michener C (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 28:1409–1438

Tessier C, David J, This P et al (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor Appl Genet 98:171–177. doi:10.1007/s001220051054

Urrestarazu J, Miranda C, Santesteban LG, Royo JB (2012) Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers. Tree Genet Genomes 8: 1163–1180. doi:10.1007/s11295-012-0502-y

van Treuren R, Kemp H, Ernsting G et al (2010) Microsatellite genotyping of apple (*Malus* × *domestica* Borkh.) genetic resources in the Netherlands: application in collection management and variety identification. Genet Resour Crop Evol 57:853–865. doi:10.1007/s10722-009-9525-0

Veloso MM, Duarte MC, Moreira P (2008) The state of in situ management. Instituto Nacional de Recursos Biológicos, Oeiras

Vinatzer BA, Patocchi A, Tartarini S, et al (2004) Isolation of two microsatellite markers from BAC clones of the *Vf* scab resistance region and molecular characterization of scab-resistant accessions in *Malus* germplasm. Plant Breed 123:321–326. doi:10.1111/j.1439-0523.2004.00973.x

Wagner HW, Sefc KM (1999) IDENTITY., p 4.0

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution (N Y) 38:1358. doi:10.2307/2408641

Zohary D, Hopf M, Weiss E (2012) Domestication of plants in the Old World: the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin, 4th edn. Oxford University Press, New York