Simple Sequence Repeat (SSR) characterisation of traditional apple cultivars (*Malus domestica* Borkh.) grown in the region of Madrid (Central Spain).

Alberto Arnal^{1, 2}, Yolanda Gogorcena³, Javier Tardío¹, Isabel Roldán-Ruiz⁴, Almudena Lázaro¹

¹ Departamento de Investigación Agroalimentaria. IMIDRA. Autovía A-2, km 38.2, Aptdo. 127, 28800, Alcalá de Henares, Madrid, Spain.

² Departamento de Biotecnología-Biología Vegetal. E.T.S.A.A.B. Universidad Politécnica de Madrid. Madrid, Spain.

³ Departamento de Pomología. Estación Experimental de Aula Dei – CSIC. Zaragoza, Spain.

⁴ Plant Sciences Unit. Institute for Agricultural and Fisheries Research (ILVO). Melle, Belgium

Corresponding Author:

Alberto Arnal¹

Departamento de Investigación Agroalimentaria. IMIDRA. Autovía A-2, km 38.2, Aptdo. 127, 28800, Alcalá de Henares, Madrid, Spain.

Email address: alberto.arnal@madrid.org

Key message

SSR analyses identified an intermediate but interesting genetic diversity in 23 traditional apple cultivars grown in mountainous rural areas from central Spain. Six of them have not been cited in previous research, indicating the existence of singular and valuable germplasm in the area.

Abstract

The genetic diversity of many Spanish apple collections has been evaluated, but some areas from central Spain have not been included. In the present work, a set of 41 accessions (33 diploids and 8 triploids) of 23 traditional apple cultivars collected in the region of Madrid and maintained at the IMIDRA collection were analysed with 13 Simple Sequence Repeats (SSRs) to evaluate its diversity. Three reference cultivars ('Fuji', 'Granny Smith' and 'Reineta Gris') were included as

reference to align the SSR profiles. The average number of alleles per locus was 7.66, which increased to 8.75 with triploids, and the number of rare alleles was 36. Moreover, the observed heterozygosity (H_o) was 0.74 and the expected heterozygosity (H_e), 0.66, indicating disequilibrium. The probability to find two identical individuals was very low ($P_{ID} = 3.60 \times 10^{-12}$). After the cluster analysis, the 29% of the accessions were found to have redundant profiles. Principal Coordinate Analysis, clustering and Bayesian model suggested the presence of four well-defined subgroups in the diploid traditional apple. In clustering, diploid reference cultivars were nested with triploids and independently from diploid traditional cultivars. Finally, after comparing profiles and names with other traditional Spanish cultivars, we conclude that the studied collection shares the ubiquitous accessions from cultivar groups such as «camuesas», «peros», and ‘Reineta’, but contain novel and uncharacterized cultivars with potential opportunities in breeding programmes and rural development: ‘Agridulce’, ‘Amarillo de El Paular’, ‘Hojancas’, ‘Mailla blanca’, ‘Pepita de melón’ and ‘Rabudas’. As these cultivars were collected in such a small area with environmental distinctiveness, they constitute a singular germplasm reservoir, which deserves to be preserved and more deeply studied.

Keywords: SSR, multivariate analysis, traditional cultivars, apple.

Introduction

Apple (*Malus domestica* Borkh.) is the most important temperate fruit tree crop (Pereira-Lorenzo et al. 2018a), with more than 125 million tons cropped in 2018 worldwide (FAO 2020). Apple production is also important in Spain, reaching almost 563,000 tons in 2018, which accounted for 60% of pome fruit production. Apple crops are located mainly in Northern Spain, in the regions of Galicia, Aragon and Catalonia (MAPA 2019).

Worldwide apple production is restricted to a few cultivars (e.g., ‘Red Delicious’, ‘Golden Delicious’ and ‘Jonathan’) whose pedigrees are linked (Hokanson et al. 2001; Larsen et al. 2018). These cultivars have been extensively used as progenitors in apple breeding programmes, reducing consequently the genetic base of the species (Noiton and Alspach 1996; Hokanson et al. 2001). In fact, the commercial cultivars ‘Golden Delicious’, ‘Gala’, ‘Idared’ and ‘Red Delicious’ accounted for 45% of the European Union production in 2016, similarly to 2008 (World Apple and Pear Association 2010; Marconi et al. 2018). In Spain, according to Iglesias and Carbó

(2018), almost a 95% of apple production is based on ‘Golden’, ‘Gala’, ‘Delicious’, ‘Fuji’, ‘Granny Smith’ and ‘Reineta’.

In terms of ploidy, two main cytotypes are present in the apple tree: the diploid ($2n=34$) and the triploid ($3n=51$) one. Diploid individuals have regular meiosis and fully fertile seeds, whereas triploid cultivars seems to be the fusion between a normal gamete and an unreduced one (Larsen et al. 2018). In consequence, triploid varieties have aberrant meiosis that causes low viable and sterile pollen, and they are usually self-infertile (Crane and Lawrence 1930; Orange Pippin Ltd. 2019). Nevertheless, due to the advantages in fruit size and vigour of triploids in comparison to diploid cultivars, triploids are relatively common in old apple collections (Larsen et al. 2018) and at a higher ratio than the one estimated for spontaneous poliploidy (Janick et al. 1996).

In 1993, the Plant Genetic Resources Programme (NPGRP) was established in Spain and the Apple Tree Collections Network as a dependence to preserve the apple genetic resources. For that propose the apple collection from ‘Estación Experimental de Aula-Dei’ (EEAD-CSIC) conserve a duplicate of each accession. The goal of this instrument is to accomplish several priorities from the FAO Global Plan of Action, inserted in the Box 7.5 of The State of the World’s Plant Genetic Resources for Food and Agriculture (FAO 1997). In the nine ex-situ apple collections that sustain our National Apple Network, there are curated 1,746 *Malus* accessions (INIA 2020). Thus, during this time and following the mandatory of the FAO, the apple diversity of many of these collections has already been registered in the NPGRP and evaluated (Itoiz and Royo 2003; Ramos-Cabrer et al. 2007; Pereira-Lorenzo et al. 2008, 2017; Urrestarazu et al. 2012; Pérez-Romero et al. 2015). However, apple germplasm grown in the Guadarrama Range has never been included in these studies (Supplemental Figure 1).

Germplasm can be characterized using morphological and molecular markers, but nowadays genotyping is the most widespread to overcome the problems associated with morphological characterisation (Donini et al. 2000; Ban et al. 2014). Different molecular markers have been developed for apple germplasm characterisation, such as Random Amplification of Polymorphic DNA (RAPDs), Amplified Fragment Length Polymorphism (AFLP), Inter Simple Sequence Repeats (ISSRs) and Simple Sequence Repeats (SSR) (Dunemann et al. 1994; Reddy et al. 2002; Royo and Itoiz 2004; Kenis and Keulemans 2005). SSRs are among the most widely used for cultivar identification because of their easy methodology, polymorphism, great discriminatory power, reproducibility and relatively low cost in comparison to other molecular techniques

(Larsen et al. 2018). In fact, SSRs are still the most common markers for diversity studies in woody species (Van Laere et al. 2018), and, specifically, in apple characterisation used in many national germplasm banks, such as those of Lithuania (Mažeikienė et al. 2019), Brazil (Mariano et al. 2019), Greece (Ganopoulos et al. 2018), Italy (Marconi et al. 2018), Denmark (Larsen et al. 2017), France (Lassois et al. 2016), Portugal (Ferreira et al. 2016), Turkey (Uzun et al. 2016), Hungary (Király et al. 2012), Iran (Gharghani et al. 2009) and even used for comparisons at a pan-European level (Urrestarazu et al. 2016).

The aim of this work is to assess the genetic diversity of the traditional apple cultivars grown in the region of Madrid (Central Spain) to hold future local conservation decisions, to promote the apple diversity in Spain by completing the National Apple Network with genotypes adapted to mountainous areas and to provide information to increase the knowledge of the Apple Genetic Resources.

Material & Methods

Plant material

A collection of 41 traditional apple accessions, established in 2009 in ‘La Isla’, an IMIDRA Experimental Station (Arganda del Rey, Madrid), was evaluated to assess its genetic diversity (Table 1). This collection is formed mainly by 23 traditional cultivars sampled in rural areas of Madrid (Spain), most of them from mountainous areas in the north of the region (Supplemental Figure 1) and documented in an ethnobotanical survey (Aceituno-Mata 2010). The mean altitude of these locations is over 1,200 m above sea level. At this altitude, the predominant bioclimate is the supramediterranean, with the oro- and crioromediterranean climates prevailing at the peaks that reach over 2,000 m (Rivas Martínez 1987).

Two diploid commercial cultivars (‘Fuji’ and ‘Granny Smith’) from the same Experimental Station were included as references to anchor subsequent analysis. In addition, DNA samples of ‘Reineta Gris’ from the National Reference Germplasm Bank, located in ‘Estación Experimental de Aula Dei (EEAD-CSIC)’ was included in amplification and sequencing for size alignment. This alignment is compulsory to compare the studied germplasm with other published fingerprinting (Ferreira et al. 2016; Royo et al. 2017).

DNA extraction

DNA was extracted from young leaves using a modified protocol of DNeasy 96 Plant Kit® (Qiagen, Hilden, Germany). Later, the product of these extractions was diluted to 5 µg/ml in ultrapure water for the subsequent amplification

SSR analysis

A set of 13 unlinked SSRs were applied to evaluate the apple collection diversity (Table 2). The SSR markers CH01f02, CH01h01, CH01h10, CH02c09, CH02c11, CH02d08, CH03d07, CH04c07, CH04e05 and CH05f06 were defined by Liebhard et al. (2002), GD12 and GD147 by Hokanson et al. (1998), and Hi02c07 by Silfverberg-Dilworth et al. (2006). Eleven of them (CH01h01, CH01f02, CH02c09, CH01h10, CH012c11, CH04c07, GD12, CH02d08, Hi021c07, GD147, CH04e05) were recommended by ECP/GR *Malus/Pyrus* Working Group, with a proposed standard methodology to compare among different laboratories (Lateur et al. 2012).

Polymerase chain reactions (PCR) were performed in a 2.700 thermal cycler (Applied Biosystems®), using three multiplex PCRs [i.e. A, B and C; (Urrestarazu et al. 2012)] in a final volume of 10 µl using 10 ng of DNA template, 0.10 µM of each primer, except for CH02c11 (0.15 µM), and 1× PCR Master mix of QIAGEN kit multiplex PCR (Qiagen, Hilden, Germany).

The PCR cycling conditions for the A and B multiplexes were the following: pre-incubation for 15 min at 95 °C, followed by five touchdown cycles at 95 °C for 30 s, 65 °C–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 last cycle ending with a final 30 min extension at 72 °C. Multiplex C was slightly different, as touchdown cycle number was raised to seven, and annealing temperature for the last 30 cycles was 58 °C instead of 60 °C.

To evaluate the amplification results, 1 µl of the PCR product was separated by electrophoresis on a 2% agarose gel. Amplification bands were detected using the GelDoc 2000 (Bio-Rad Laboratories, CA, USA) transilluminator and visualised by the Quantity One analysis Software (Bio-Rad Laboratories, CA, USA).

The PCR product were then mixed with formamide and the GeneScan 500 LIZ Size standard and run in an ABI PRISM 3130XL sequencer (Applied Biosystems, Foster City, CA, USA). In order to reduce the effects of artefacts in the results of the sequencing, 1 µl of the PCR product were diluted 4 or 36 times in ultrapure water prior sequencing. Then the fragment analysis and sizing were carried out using the Peak Scanner Software version 1.0 (Applied Biosystems).

Diversity analysis

The ploidy level of each accession was assessed by the maximum count of the number of alleles of each locus. Then, a diversity analysis was performed for diploid accessions, calculating manually in a spreadsheet the following variables: number of alleles per locus (A), number of rare alleles per locus (B , number of alleles whose frequency was <0.05), number of effective alleles (A_e), observed heterozygosity (H_o), expected heterozygosity (H_e) and probability of identity (P_{ID}). Values of A_e were obtained as $1/\sum p_i^2$, (Morgante et al. 1994); H_o as the ratio of heterozygous genotypes per locus (Nei 1978); H_e as $1 - \sum p_i^2$ (Nei 1978) and P_{ID} as $\sum p_i^4 + \sum \sum (2p_i p_j)^2$ (Waits et al. 2001). In all the formulas, p_i represents the allele frequencies of the i th loci. In the case of P_{ID} , p_j is the frequency of the j th allele, used only if $i \neq j$.

In addition, the actual or observed number of genotypes (G_o), the possible genotypes (G_p) and the percentage of represented genotypes in the collection (R) were also evaluated.

The possible genotypes (G_p) is given by all the combination of alleles by locus. This value was calculated as a combination with repetition of A_i alleles taken two by two, where A_i is the number of alleles for a given locus:

$$G_p = \binom{2 + A_i - 1}{2} = \frac{(A_i + 1)!}{2(A_i - 1)!}$$

Finally, the optimal number of loci selected to distinguish all unique genotypes was determined experimentally according to Tessier et al. (1999), after removing repeated genotypes.

Principal Coordinates Analysis

A Principal Coordinates Analysis (PCoA) with all SSR markers was calculated using the R package ‘adegetnet’ (Jombart and Ahmed 2008) after preparing data with the *loci2genind* function from ‘pegas’. Loading plots scoring contribution for each allele per locus and principal coordinates helped to identify discriminant alleles.

Bayesian analysis of population structure

A Bayesian model using the Structure 2.3.4 software (Pritchard et al. 2000) was applied to study population structure. The admixture model considered unlinked loci and correlated allele frequencies and was computed separately for diploids and triploids.

The number of tested populations (k) ranged from 2 to 10 in diploids and from $k = 2$ to 5 in triploids. These differences lie on the very low number of samples. Genotypes were assigned to the group where a highest membership was obtained, considering such affinity with the probability of membership (qI) ≥ 0.8 . This cut value was established in other Spanish and

European collection studies (Urrestarazu et al. 2012; Larsen et al. 2017; Pereira-Lorenzo et al. 2017; Miranda et al. 2018). Lower probabilities assigned the cultivar to an admixed origin. Twenty runs per k were evaluated, each consisting of a burning period length of 30,000 steps followed by 1,000,000 Monte Carlo Markov Chain (MCMC) replicates, as recommended by Porras-Hurtado et al. (2013). Additionally, the optimum value of k was computed with the Evanno's ΔK statistic from Structure Harvester (Earl and Von Holdt 2012).

Cluster analysis

Genotypes were clustered through the unweighted pair group method arithmetic average (UPGMA) (Sokal and Michener 1958) using the Nei's distance (Nei 1973) and the bootstrap method with 10,000 replications. This analysis was carried with traditional and reference accessions (N=44) using the function *aboot* of the 'poppr' R package (Kamvar et al. 2014).

Results

Ploidy level

Eight traditional accessions showed a range of three to five loci with three amplified alleles, thus indicating triploidy (Supplemental Table 1). Triploid accessions belong to four cultivars: 'Agridulce' (MAMO), 'del Ortel' (ORCA), 'Hojancas' (MHHO) and 'Reineta' (REHO; REMO, REPR1, REPR2 and REPU). Three alleles were identified at five loci in 'MAMO' (CH01f02, CH02c11, CH02d08, CH03d07 and CH05f06); at four loci in 'ORCA' (CH01f02, CH01h01, CH02c09, CH02c11 and CH04c07) and at three loci in 'MHHO' (CH01f02, CH02d08 and CH05f06). Concerning the cultivar 'Reineta', full triploid loci were found in CH01h01, CH02c09, CH02c11, CH02d08, CH03d07, CH04c07 and CH05f06 for all genotypes except for REPR1. As expected, the reference cultivars 'Fuji' and 'Granny Smith' were diploids and 'Reineta Gris', triploid.

Genetic diversity

All SSR produced polymorphic amplicons that were adjusted to their bins in most of the accessions (Supplemental Figure 2). However, GD147 was removed from the study because the quality of profiles was poor to allow alignments between samples and references. The parameters of diversity of the remaining 12 SSRs are summarised in Table 3.

The total number of alleles per locus (A) amplified in all loci was 92 in diploids accessions (105 with triploids), with an average number of alleles per locus of 7.66 (8.75 with triploids); ranging

from 4 alleles (locus GD12 and Hi02c07) to 12 alleles (locus CH01f02). The total number of effective alleles (A_e) scored in our study was 47.16, with an average value of 3.93, meaning around the 50% of the initial value of A . CH02c11 was the most diverse locus in A_e (7.14 alleles) and Hi02c07 the least one (1.57 alleles). In addition, 36 alleles (40% of A) were considered rare. Moreover, rare allele (B) distribution among loci was slightly correlated with A , being 6 in CH01f02 (the most diverse) and zero in Hi02c07 (who was the least diverse). Heterozygosity and genotype frequencies for each locus were evaluated for diploids (Figure 1). Mean values of observed heterozygosity (H_o) and expected heterozygosity (H_e) were 0.80 and 0.71, respectively. H_o varied along the loci amplified, ranging from 0.24 in Hi02c07 to 1 in CH01h10. Observed genotypes per locus (G_o) scored an average value of 10.08. The loci with more G_o were CH01f02 and CH01h01, which accounted for 14 genotypes; whereas in Hi02c07 only 4 were found. The average value of representativeness (R) was 33%, being GD12 (60%) the only loci with more than the 50% of the combinations. However, 6 out of 12 loci shown a low number of allele combinations (representativeness below the 30%). The probability of identity of each SSR in the overall studied accessions was $P_{ID} = 3.60 \times 10^{-12}$, with a mean per loci of 0.12. The optimum number of loci to distinguish all genotypes after removing identical accessions was three for the 21 unique traditional diploid genotypes (Table 4) and only one (CH01h01, CH02c09 or CH03d07) for the five unique triploid genotypes (MAMO, ORCA, MHHO, REPR1 and the other 'Reineta').

Principal Coordinates Analysis

As shown in Figure 2, the percentage of cumulative variance explained by the first two principal coordinates (PCo 1 and PCo 2) accounted for 52.10%, being necessary six PCo to reach the 80% of accumulated variability. Four clusters were grouped, which were well distributed thorough the 2-D plot drawn with the two first PCo. Group 1, with negative value of PCo 1, included ten accessions: 'Camueso tardío' (CTAP), four «peros» (PAPR, PPHI, PPMO, PRBU), three 'Rojillo' (RJHO, RTAV, RTEV), and two 'Rojo' (RJPR, RJPU). Group 2, located in negative values in both axis, grouped nine accessions: two 'Camuesa' (CABU, CAVA), two 'de Chapa' (CHCA, CHVA), 'Pepita de melón' (PECA, PEHI, PEVA), 'Rabudas' (RAHI) and one 'Temprano' (TEPI). Group 3, was characterised by extremely positive values of PCo1 and negative values of PCo 2, and casted only three accessions, being all of them «peros» (PAHO, PAPU, PGHI). Finally, located in the first quadrant with positive values of both PCo, Group 4

gathers thirteen accessions: the two reference cultivars ('Fuji' and 'Granny Smith'), 'Amarillo de El Paular' (APRA), 'Camueso temprano' (CTEP), 'Esperiega' (ESPU), 'Mailla blanca' (MBHI), 'Hojancas' (MHPR), 'de Rosa' (RORA), 'San Felipe' (SFCA), one 'Temprano' (TECH) and the three traditional 'Verde Doncella' (VDCA, VDHO, VDTI).

A threshold loading value of 0.07 was used to select the most discriminating alleles per locus, leading to a satisfactory separation among the four groups. The PCo 1 discriminated along alleles Hi02c07.112 and Hi02c07.114 and PCo 2 did along CH02d08.254, CH05f06.180. As a result, there were no compromising confusions between groups.

In the case of triploids the first three principal components accounted more than the 95%, but no differentiating alleles per locus were found as the variance is shared by many alleles from different SSRs (data not shown).

Bayesian Analysis of populations

The optimum k value obtained in the Bayesian analysis for diploids and triploids accessions was both k= 4 (Figure 3). In the case of diploids, Evanno's ΔK was 6.05. The reference population (in red) was composed by the two apple references ('Fuji' and 'Granny Smith') and shared with 'Verde Doncella' (VDCA, VDHO, VDTI), 'Camueso temprano' (CTEP), 'Amarillo de El Paular' (APRA), 'Temprano' (TECH), 'Esperiega' (ESPU), 'Pero pardo' (PPMO) and 'Mailla blanca' (MBHI). In the pale green population were gathered all 'Camuesa' (CAVA, CABU), 'de Chapa' (CHVA, CHCA) and 'Pepita de melón' (PEVA, PEHI, PECA).

The dark green population, composed by some «peros» (PAHO, PAPU and PGHI), 'Hojancas' (MHPR), 'San Felipe' (SFCA), 'Temprano' (TECH) and 'de Rosa' (RORA), showed the lowest admixture levels, as none of them shown accessions with $qI < 0.8$.

On the other hand, only six (17% of diploids) accessions have low probability of membership ($qI < 0.8$): 'Camueso tardío' (CTAP), 'Mailla blanca' (MBHI), 'Rabudas' (RAHI), 'de Rosa' (RORA), 'Pero real' (PRBU) and 'Rojillo' (RTAV).

In triploid accessions, Evanno's ΔK was 0.858. In this ploidy level, four 'Reineta' accessions (REHO, REMO, REPR2 and REPU) and 'Reineta Gris' scored a $qI \geq 0.8$ and belong to the same. The other triploids 'Agridulce' (MAMO), 'Hojancas' (MHHO), 'del Ortel' (ORCA) and 'Reineta' (REPR1) were highly admixed.

Cluster analysis

Figure 4 shows the UPGMA hierarchical clustering carried out with the local and reference accessions (N=44). A 29% of duplicates were found. The biggest duplicate set was the one formed by some «peros», ‘Rojo’ and ‘Rojillo’ (PAPR, PPHI, RJHO, RJPU, RJPR and RTEV) and the ‘Reineta’ (REHO, REMO, REPR2, REPU and ‘Reineta Gris’, cluster I.1.a). On the other hand, some accessions with the same name were grouped in distant clusters. This is the case of ‘Hojancas’, ‘Pero de Aragón’, ‘Reineta’ and ‘Temprano’. The dendrogram is clearly structured in two main sets. The first cluster (Cluster I) separates triploids and the diploid reference cultivars (‘Granny Smith’ and ‘Fuji’) from the rest of traditional cultivars (Cluster II). Cluster II retains exclusively traditional cultivars in several subgroups and branches grown alone. The most distinct accessions were ‘Verde Doncella’ (VDTI, VDCA and VDHO) and ‘Camueso tardío’ (CTEP), as they were separated from the rest of the cultivars. The remaining accessions were structured in several groups: Cluster II.1.a contains mostly «peros» and its duplicates (PAPR, PPHI, RJHO, RJPU, RJPR and RTEV). Cluster II.1.b includes exclusively the set of ‘Camuesa’ (CABU and CAVA) ‘de Chapa’ (CHCA and CHVA) and ‘Pepita de melón’ (PECA, PEHI and PEVA). The remaining clusters grouped mainly accessions of group 4, as nested ‘Hojancas’ (MHPR) and its duplicates (SFCA, TECH), de Rosa’ (RORA), ‘Pero de Aragón’ (PAPU) and its duplicates (PAHO PGHI), ‘Mailla blanca’ (MBHI) and ‘Esperiega’ (ESPU).

Discussion

Ploidy level

Eight traditional accessions from four cultivars (19.51%) shown at least three fully heterozygous loci, meaning that they are likely to be triploids (Urrestarazu et al. 2012; Marconi et al. 2018). The percentage of triploids was similar to those described in other studies, such as the 18.5% of triploids found in Lassois et al. (2016), 19.5% in Liang et al. (2015) and 20% in Marconi et al. (2018). But lower than that found in other Iberian apple collections, where the level of triploids was 24% (Urrestarazu et al. 2012), 27.86% (Pereira-Lorenzo et al. 2007), 29% (Ramos-Cabrera et al. 2007) and even 32% (Pina et al. 2014).

The detection of triploids might be linked to the number of loci analysed as there are more chances to uncover a third allele of full heterozygous loci. In fact, the studies mentioned above used a number of loci generally higher than our, as oscillated from 15 (Liang et al. 2015) to 24

(Lassois et al. 2016). Nevertheless, in Ramos-Cabrer et al. (2007), almost a 30% of triploids were detected with only 10 primers, so it is also possible to detect triploids with a low number of SSR. Five triploids accessions are labelled as ‘Reineta’. The notable presence of ‘Reineta’ accessions in several municipalities is explained by two facts: first, its ethnobotanical importance in the past (Guerra et al. 2010) and second, its medium to late blooming (Pina 2005). As a result, it is possible to speculate that ‘Reineta’ might be suitable in mountainous and cold environments as it was suggested (Pina et al. 2014).

Genetic diversity

The quantitative genetic diversity of the IMIDRA apple collection based on 12 SSR can be considered intermediate, because direct values of genetic diversity are below or similar to other studies (see A, A_e and B in Table 3) as unfolded below.

The mean number of alleles per locus (A) amplified in all loci, including triploids, was 8.75, which is lower than values reached in other Spanish collections. The most complete study of traditional cultivars in Spain scored a value of 18.61 with 1,453 accessions (Pereira-Lorenzo et al. 2017), others scored 12.3 with 77 accessions (Ramos-Cabrer et al. 2007), 12.8 with 183 accessions (Pina et al. 2014) and 16.69 with 493 accessions (Urrestarazu et al. 2012).

This apparent reduced diversity can be explained by the quite low number of accessions (41) from this local collection. In fact, other collections with a similar size (Galli et al. 2005; Garkava-Gustavsson et al. 2008; Gaši et al. 2010; Patzak et al. 2012) observed similar or even lower, with only 6.5 alleles per locus in the case of Goulão and Oliveira (2001) leading us to consider that in our region there is an intermediate apple diversity.

The number of effective alleles (A_e) for all SSR loci was 47.16, which accounts around the half of the total number of alleles per locus (A). Our results are in agreement with other authors (Pina et al. 2014; Liang et al. 2015; Ferreira et al. 2016; Uzun et al. 2016; Lassois et al. 2016; Marconi et al. 2018). Rare alleles (B) might also be as influenced by the collection size, as Garkava-Gustavsson et al. (2008) scored a 35% of rare alleles with 68 cultivars, Pina et al. (2014) reached 57% with 183 accessions, and in Liang et al. (2015) and Larsen et al. (2017) that proportion rose to almost a 70% with 383 and 448 accessions, respectively.

In addition, the average probability of identity (P_{ID}) was relatively high (0.12) comparing to other Spanish collections (Urrestarazu et al. 2012) and some Italian germplasm (Liang et al. 2015; Marconi et al. 2018), but not as high as in other studies (Ganopoulos et al. 2018).

As several studies with similar size reported greater diversity (Galli et al. 2005; Garkava-Gustavsson et al. 2008; Gaši et al. 2010; Patzak et al. 2012) it can be assumed that the low genetic diversity is due to an environmental or cultural pressure in the cultivar selection that produced a bias in the diversity found, probably accompanied by certain genetic erosion.

The intermediate diversity found are also reflected in the low number of observed genotypes (G_o) and representativeness (R), suggesting that only a scarce set of combinations might have been selected and preserved over time in high mountainous rural areas of Madrid, characterised by such specific conditions: hard winters and quite dry summers, characteristic of the supramediterranean climate (Rivas Martínez 1987). For example, the locus CH01h10, with 5 different alleles, only scored 7 combinations from 15 possible ($R=47\%$), being very frequent the genotype 94/102 (27%) and not existing, for example, accessions with 112/120 (Figure 1).

This germplasm collection contains interesting findings as those resulted from the expected and observed heterozygosity (H_o and H_e), Our collection is not in equilibrium as found in other collections from the Iberian Peninsula (Pina et al. 2014; Ferreira et al. 2016).

The number of optimum loci to differentiate the collection is in accordance with other studied collections. Only three SSR markers were needed due to the combination of a moderate representativeness (R), low overlapping between genotypes, and secondary due to the high observed heterozygosity (H_o). In the apple collection studied by Galli et al. (2005) 4 SSR were needed to discriminate 66 commercial cultivars. In the study of Wichmann et al. (2018) only three loci discriminated 40 accessions. Larger collections may require more loci, as seven loci were needed to discriminate the 1,127 accessions from Poles et al. (2018) study.

PCoA, Bayesian analysis of populations and Cluster analyses

A clear separation of the reference varieties from traditional cultivars have been shown in all statistical analyses. This consistency is common in Iberian studies (Ferreira et al. 2016; Pereira-Lorenzo et al. 2017), highlighting the ancient hybridisation and singularity of the Spanish apple traditional cultivars.

In cluster analysis (Figure 4), all triploids but ORCA were classified separately from the diploid accessions, as reported other authors (Pereira-Lorenzo et al. 2007; Garkava-Gustavsson et al. 2008; Wichmann et al. 2018). This separation can also be explained by the fact that five of the triploids accessions belong to ‘Reineta’ cultivar or are presumably related to this genotype.

The maximum Nei's distance in the cluster analysis (0.16) was lower than in other collections characterised, such as 0.21 in Uzun et al. (2016), 0.32 in Guarino et al. (2006), 0.6 in Goulão and Oliveira (2001) and 0.78 in Pina et al. (2014), however the dendrogram topology (Figure 4) is similar to the mentioned studies. The fact that we didn't find distant accessions as in other studies is compatible with intermediate levels of diversity.

Mislabelling in old collections, typographical mistakes, or name replacement due to farmer preferences often occurs (Foroni et al. 2012), sometimes due to language preferences (Heinonen and Bitz 2019). This is probably the case of «camuesa» and «pero» accessions that share the same name but grouped in different clusters as found in other collections (Pina et al. 2014).

On the other hand, duplicates of molecular profiles are often obtained because phenotypic variants maintained in germplasm collections are indistinguishable by using molecular markers. Duplicate profiles were found between CABU and CHVA (from Group 1 in PCoA and Cluster II.1.b) and MHPR, SFCA and TECH (from Group 4 in PCoA and Cluster II.2.a.). It is demonstrated that SSR can hide false clones (Larsen et al. 2017) masking punctual variations (Garkava-Gustavsson et al. 2008; Cipriani et al. 2010; Patzak et al. 2012) detectable with morphological traits and sensorial descriptions. Therefore, it is necessary to use other molecular markers (such as retrotransposons) to elucidate the presumed spontaneous mutations (Venturi et al. 2006; Zhao et al. 2010). In addition, a morphological approach is required to characterize clones supporting the conservation of those putative clones.

Relationships between cultivars can be revealed up to some point using different approaches. Results from a close relationship between 'Camuesa' (CAVA, CABU), 'de Chapa' (CHCA, CHVA) and 'Pepita de melón' (PECA, PEHI, PEVA) was found as indicated in the dendrogram (Cluster II.1.b), the PCoA (Group 1) and the Bayesian model (pale green population). These results were consistent with those reported by Pereira-Lorenzo and coworkers (2018b) referring 'Camuesa' as an admixed cultivar. In this case, we suspect that 'de Chapa' and 'Pepita de melón' are probably a type of 'Camuesa', either from one or several punctual mutations or by sexual reproduction.

The presence of synonyms and homonyms of «pero» in this collection highlights the importance of this cultivar in our region. In fact, the cultivar 'Pero de la Hiruela' was reported in the past and probably cultivated in the study area (Aceituno-Mata 2010; Ayuntamiento de La Hiruela 2020). We did not conserve any accession with this name in our collection but some «pero» accessions

were collected ('Pero gordo' and 'Pero pardo') in our study area. On the other hand, we also sampled in La Hiruela locality several accessions named 'Pero de Aragón', which is alleged to be the synonym of that cultivar (Aceituno-Mata 2010).

Traditional cultivar names are shared among different rural regions of Spain. Therefore, 'del Ortel', 'Reineta' and 'Verde Doncella' accessions from the IMIDRA collection might be the same cultivars as those collected in other Spanish regions, as shared very similar profiles (Royo et al. (2017). On the contrary, the profiles found in 'Esperiega' (ESPU), «peros» and 'Verde Doncella' differed in accessions with similar names in other Iberian collections (Ferreira et al. 2016; Royo et al. 2017). In the case of «peros», the comparison of profiles with 'Pero de Cehegín' and 'Peruco Lardero', resulted to be different (Ferreira et al. 2016; Royo et al. 2017; Martínez et al. 2020).

Further discussion on «camueas» and «peros» can be developed with morphological and ethnobotanical data. In the case of ethnobotanical research, «camueas» and «peros» are mentioned as two distinct types of apple due to differences in morphological and utilisation aspects (Arnal et al. 2018). Nevertheless, we speculate that the genetic background of these two apple types is so heterogeneous that it is very difficult to find a short number of discriminant SSR to differentiate them in consistent groups. Such heterogeneity can be possible either by punctual mutations due to the grafting stress or cross-pollination between highly heterozygous genotypes (Garkava-Gustavsson et al. 2008), which widen diversity and, consequently, the origin of new and distinct cultivars.

The loss of the apple genetic resources in the study area

There is no evidence that inhabitants from rural areas of Madrid intensively cultivated ubiquitous, commercial genotypes along with the traditional ones, except for the 'Reineta' varietal group, which is supposed to be old enough to have been cultivated and introgressed into the apple genetic resources of the region (Aceituno-Mata 2010).

The absence of cultivation of international varieties supports our idea that the threat to apple genetic resources conservation in this region is the abandonment of rural areas rather than variety reconversion. The migration from rural areas to larger cities is a serious problem not only in our region and other local areas (FEMP 2017; Marconi et al. 2018) but also in many regions of the European Union, as evidenced by the abundant legislation in rural development that tries to prevent it, such as the Regulation (EU) N° 1305/2013 in support of rural development by the

European Agricultural Fund for Rural Development (EAFRD). Nevertheless, the infusion of commercial cultivars into the traditional germplasm population is possible and happened specially in ‘Reineta’ (Pereira-Lorenzo et al. 2003).

Finally, to our knowledge, the studied collection contains six exclusive traditional cultivars, not recorded previously in other Spanish collections (Royo et al. 2017): ‘Amarillo de El Paular’, ‘Agridulce’, ‘Hojancas’, ‘Mailla blanca’, ‘Pepita de melón’ and ‘Rabudas’.

Conclusions

The present study is the first aiming to describe the diversity and variability of the apple trees growing in the centre of the Iberian Peninsula. This study increase the Apple Genetic Resources in Spain, thus proposing new cultivars for rational and sustainable utilisation. Many cultivars grown in the study area are probably the same as those cultivated in other regions of the Iberian Peninsula. Nevertheless, six cultivars reported herein have not been previously listed or characterized (‘Agridulce’, ‘Amarillo de El Paular’, ‘Hojancas’, ‘Mailla blanca’, ‘Pepita de melón’ and ‘Rabudas’). Taking into account that they were collected in a mountainous small area, they constitute a unique germplasm reservoir which deserves to be urgently preserved and more deeply studied. In conclusion, our research has shed some light in the Spanish apple diversity and should be taken into consideration for subsequent actions, such as registry in the National Apple Network and rational decision making in cultivar protection according to the requirements and opportunities given by the EAFRD and under the inspiration of the new Sustainable Development Objectives from the United Nations (2015-2030).

Authors’ contribution statement

Alberto Arnal, Almudena Lázaro and Javier Tardío designed the research; Alberto Arnal and Almudena Lázaro sampled the accessions; Yolanda Gogorcena provided materials and supervised amplification profiles; Alberto Arnal, Yolanda Gogorcena and Isabel Roldán-Ruiz performed the research and analysed the data. Alberto Arnal wrote the draft of the manuscript and Yolanda Gogorcena, Javier Tardío and Almudena Lázaro, revised and commented the manuscript.

Acknowledgements

We thank gratefully to Sabine Van Glabeke and Hylde Mulle for their dedication at the sequencing and fragment analysing in the ILVO Research Centre (Belgium) during summer 2018.

This work was supported by the IMIDRA (Madrid Institute for Research and Rural Development in Food and Agriculture), under the local project ‘Prospection, characterization, ethnobotanical documentation and conservation of traditional horticultural cultivars of the Community of Madrid’ (ref. FP16-ETNOBOT)’ and a predoctoral contract funded by the INIA (National Institute for Agricultural and Food Research and Technology). We thank to Alvaro Blanco for his work as responsible for the Apple Germplasm Bank of EEAD-CSIC whose maintenance is partially supported by the INIA, under the project ‘RF2011-00017-C05-05’ and the ‘Estación Experimental de Aula Dei’ (EEAD-CSIC) allowing us to use laboratory facilities. Partial support was obtained from the Spanish State Research Agency (AEI) co-financed with the European Regional Development Fund (FEDER) (project AGL2017-83358-R to Y.G; AEI/FEDER, UE), and the Aragón Government (Group A09-20R).

Conflict of interest

The authors declare that they have no conflict of interest.

Data archiving statement

Accession information is described in the Table 1 and their genotype profiles for all 12 SSR included in the Supplemental Table 1.

References

- Aceituno-Mata L (2010) Estudio Etnobotánico y Agroecológico de la Sierra Norte de Madrid. Autonomous University of Madrid. Science Faculty
- Arnal A, Lázaro A, Aceituno-Mata L, et al (2018) *Malus domestica* Borkh. In: Tardío J, Pardo de Santallana M, Morales R, et al. (eds) *Inventario Español de los Conocimientos Tradicionales Relativos a la Biodiversidad Agrícola*. Madrid, pp 227–240
- Ayuntamiento de La Hiruela (2020) Fiesta de la recolección del Pero.

488 <https://turismolahiruela.es/que-hacer/fiestas-tradicionales/recoleccion-del-pero/>. Accessed
489 29 May 2020

490 Ban SH, Yun WH, Kim GH, et al (2014) Genetic identification of apple cultivars bred in Korea
491 using simple sequence repeat markers. *Hortic Environ Biotechnol* 55:531–539. doi:
492 10.1007/s13580-014-0017-6

493 Cipriani G, Spadotto A, Jurman I, et al (2010) The SSR-based molecular profile of 1005
494 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and
495 reveals a large admixture amongst varieties of different geographic origin. *Theor Appl*
496 *Genet* 121:1569–1585. doi: 10.1007/s00122-010-1411-9

497 Crane MB, Lawrence WJC (1930) Fertility and vigour of apples in relation to chromosome
498 number. *J Genet* 22:153–163. doi: 10.1007/BF02983844

499 Donini P, Cooke RJ, Reeves JC (2000) Molecular markers in variety and seed testing. *Dev Plant*
500 *Genet Breed* 5:27–34. doi: 10.1016/S0168-7972(00)80005-5

501 Dunemann F, Kahnau R, Schmidt H (1994) Genetic relationships in *Malus* evaluated by RAPD
502 “Fingerprinting” of cultivars and wild species. *Plant Breed* 113:150–159. doi:
503 10.1111/j.1439-0523.1994.tb00717.x

504 Earl DA, Von Holdt BM (2012) Structure Harvester: a website and program for visualizing
505 Structure output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361.
506 doi: 10.1007/s12686-011-9548-7

507 FAO (2020) FAOSTAT. In: FAO Stat. database World Wide Web.
508 <http://www.fao.org/faostat/en/#data/QC>. Accessed 8 Apr 2020

509 FAO (1997) The State of the World’s Plant Genetic Resources for Food and Agriculture - Part 2.
510 Rome, Italy

511 FEMP (2017) Listado de medidas para luchar contra la despoblación en España

512 Ferreira V, Ramos-Cabrera AM, Carnide V, et al (2016) Genetic pool structure of local apple
513 cultivars from Portugal assessed by microsatellites. *Tree Genet Genomes* 12:. doi:
514 10.1007/s11295-016-0997-8

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

518 Galli ZS, Halász G, Kiss E, et al (2005) Molecular identification of commercial apple cultivars

519 with microsatellite markers. HortScience 40:1974–1977

520 Ganopoulos I, Tourvas N, Xanthopoulou A, et al (2018) Phenotypic and molecular
 521 characterization of apple (*Malus × domestica* Borkh.) genetic resources in Greece. Sci
 522 Agric 75:509–518. doi: 10.1590/1678-992x-2016-0499

523 Garkava-Gustavsson L, Kolodinska Brantestam A, Sehic J, Nybom H (2008) Molecular
 524 characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis.
 525 Hereditas 145:99–112. doi: 10.1111/j.0018-0661.2008.02042.x

526 Gaši F, Simon S, Pojskić N, et al (2010) Genetic assessment of apple germplasm in Bosnia and
 527 Herzegovina using microsatellite and morphologic markers. Sci Hortic (Amsterdam)
 528 126:164–171. doi: 10.1016/j.scienta.2010.07.002

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

533 Goulão L, Oliveira CM (2001) Molecular characterisation of cultivars of apple (*Malus ×*
 534 *domestica* Borkh.) using microsatellite (SSR and ISSR) markers. Euphytica 122:81–89. doi:
 535 10.1023/A:1012691814643

536 Guarino C, Santoro S, De Simone L, et al (2006) Genetic diversity in a collection of ancient
 537 cultivars of apple (*Malus × domestica* Borkh.) as revealed by SSR-based fingerprinting. J
 538 Hortic Sci Biotechnol 81:39–44

539 Guerra M, Valenciano JB, Marcelo V, Casquero PA (2010) Storage behaviour of “Reinette du
 540 Canada” apple cultivars. Spanish J Agric Res 8:440–447

541 Heinonen M, Bitz L (2019) How to discover traditional varieties and shape in a National
 542 Germplasm Collection: the case of Finnish seed born apples (*Malus × domestica* Borkh.).
 543 Sustainability 11:7000. doi: 10.3390/su11247000

544 Hokanson SC, Lamboy WF, Szewc-McFadden AK, et al (2001) Microsatellite (SSR) variation in
 545 a collection of *Malus* (apple) species and hybrids. Euphytica 118:281–294. doi:
 546 10.1023/A:1017591202215

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

550 Iglesias I, Carbó J (2018) El cultivo del manzano en España: situación actual e innovación
 551 varietal. [https://www.interempresas.net/Horticola/Articulos/208474-El-cultivo-del-](https://www.interempresas.net/Horticola/Articulos/208474-El-cultivo-del-manzano-en-Espana-situacion-actual-e-innovacion-varietal.html)
 552 [manzano-en-Espana-situacion-actual-e-innovacion-varietal.html](https://www.interempresas.net/Horticola/Articulos/208474-El-cultivo-del-manzano-en-Espana-situacion-actual-e-innovacion-varietal.html). Accessed 5 Jul 2020
 553 INIA (2020) National Inventory of Plant Genetic Resources.
 554 webx.inia.es/web_inventario_nacional/. Accessed 5 Jul 2020
 555 Itoiz R, Royo JB (2003) Isoenzymatic variability in an apple germplasm bank. *Genet Resour*
 556 *Crop Evol* 50:391–400. doi: 10.1023/A:1023970926516
 557 Janick J, Cummins JN, Brown SK, Hemmat M (1996) Apples. In: Janick J, Moore JN (eds) *Fruit*
 558 *Breeding, Tree and Tropical Fruits*. pp 1–78
 559 Jombart T, Ahmed I (2008) adegenet: a R package for the multivariate analysis of genetic
 560 markers. *Bioinformatics* 24:1403–1405
 561 Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of
 562 populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281
 563 Kenis K, Keulemans J (2005) Genetic linkage maps of two apple cultivars (*Malus x domestica*
 564 Borkh.) based on AFLP and microsatellite markers. *Mol Breed* 15:205–219. doi:
 565 10.1007/s11032-004-5592-2
 566 Király I, Redeczki R, Erdélyi É, Tóth M (2012) Morphological and molecular (SSR) analysis of
 567 old apple cultivars. *Not Bot Horti Agrobot Cluj-Napoca* 40:269–275
 568 Larsen B, Gardner K, Pedersen C, et al (2018) Population structure, relatedness and ploidy levels
 569 in an apple gene bank revealed through genotyping-by-sequencing. *PLoS One* 13:e0201889.
 570 doi: 10.1371/journal.pone.0201889
 571 Larsen B, Toldam-Andersen TB, Pedersen C, Ørgaard M (2017) Unravelling genetic diversity
 572 and cultivar parentage in the Danish apple gene bank collection. *Tree Genet Genomes*
 573 13:14. doi: 10.1007/s11295-016-1087-7
 574 Lassois L, Denancé C, Ravon E, et al (2016) Genetic Diversity, Population Structure, Parentage
 575 Analysis, and Construction of Core Collections in the French Apple Germplasm Based on
 576 SSR Markers. *Plant Mol Biol Report* 34:827–844. doi: 10.1007/s11105-015-0966-7
 577 Lateur M, Ordidge M, Engels J, Lipman E (2012) Report of a Working Group on *Malus / Pyrus*.
 578 Bioversity International, Rome, Italy
 579 Liang W, Dondini L, De Franceschi P, et al (2015) Genetic Diversity, Population Structure and
 580 Construction of a Core Collection of Apple Cultivars from Italian Germplasm. *Plant Mol*

581 Biol Report 33:458–473. doi: 10.1007/s11105-014-0754-9
 582 Liebhard R, Gianfranceschi L, Koller B, et al (2002) Development and characterisation of 140
 583 new microsatellites in apple (*Malus × domestica* Borkh.). Mol Breed 10:217–241. doi:
 584 10.1023/A:1020525906332
 585 MAPA (2019) Avance del Anuario de Estadística del año 2018. Ministerio de Agricultura, Pesca
 586 y Alimentación, Madrid
 587 Marconi G, Ferradini N, Russi L, et al (2018) Genetic characterization of the apple germplasm
 588 collection in central Italy: the value of local varieties. Front Plant Sci 9:1460. doi:
 589 10.3389/fpls.2018.01460
 590 Mariano LC, Zchonski FL, Da Silva CM, Da-Silva PR (2019) Genetic variability in a Brazilian
 591 apple germplasm collection with low chilling requirements. PeerJ 6:e6265. doi:
 592 10.7717/peerj.6265
 593 Martínez R, Legua P, Hernández F, et al (2020) Molecular, physico-chemical and sensory
 594 characterization of the traditional Spanish apple variety “Pero de Cehegín.” Agronomy 10,
 595 1093. doi:10.3390/agronomy10081093
 596 Mažeikienė I, Bronė Šikšnianienė J, Baniulis D, et al (2019) SSR analysis based on molecular
 597 characterisation of apple germplasm in Lithuania. Zemdirbyste-Agriculture 106:159–166.
 598 doi: 10.13080/z-a.2019.106.021
 599 Miranda C, Errea P, Urrestarazu J, et al (2018) Definición del núcleo optimizado de la colección
 600 de conservación del manzano español. Actas Hortic 80:67–71
 601 Morgante M, Rafalski A, Biddle P, et al (1994) Genetic mapping and variability of seven
 602 soybean simple sequence repeat loci. Genome 37:763–9
 603 Nei M (1978) Estimation of average heterozygosity and genetic distane from a small number of
 604 individuals. Genetics 89:
 605 Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci
 606 70:3321–3323
 607 Noiton DAM, Alspach PA (1996) Founding clones, inbreeding, coancestry, and status number of
 608 modern apple cultivars. J Am Soc Hortic Sci 121:773–782
 609 Orange Pippin Ltd. (2019) Triploid apple varieties.
 610 <https://www.orangepippintrees.co.uk/articles/about-orange-pippin-fruit-trees>
 611 Patzak J, Paprštejn F, Henychová A, Sedlák J (2012) Genetic diversity of Czech apple cultivars

612 inferred from microsatellite markers analysis. *Hortic Sci* 39:149–157. doi:
 613 10.17221/218/2011-HORTSCI
 614 Pereira-Lorenzo S, Fischer M, Ramos-Cabrer AM, Castro I (2018a) Apple (*Malus* spp.)
 615 Breeding: Present and Future. In: Al-Khayri J, Jain SM, Johnson D V. (eds) *Advances in*
 616 *Plant Breeding Strategies: Fruits*. Springer International Publishing, pp 3–29
 617 Pereira-Lorenzo S, Ramos-Cabrer AM, Ascasibar-Errasti J, Piñeiro-Andión J (2003) Analysis of
 618 apple germplasm in northwestern Spain. *J Am Soc Hortic Sci* 128:67–84. doi:
 619 10.21273/jashs.128.1.0067
 620 Pereira-Lorenzo S, Ramos-Cabrer AM, Díaz-Hernández MB (2007) Evaluation of genetic
 621 identity and variation of local apple cultivars (*Malus × domestica* Borkh.) from Spain using
 622 microsatellite markers. *Genet Resour Crop Evol* 54:405–420. doi: 10.1007/s10722-006-
 623 0003-7
 624 Pereira-Lorenzo S, Ramos-Cabrer AM, Ferreira V, et al (2018b) Genetic diversity and core
 625 collection of *Malus × domestica* in northwestern Spain, Portugal and the Canary Islands by
 626 SSRs. *Sci Hortic (Amsterdam)* 240:49–56. doi: 10.1016/j.scienta.2018.05.053
 627 Pereira-Lorenzo S, Ramos-Cabrer AM, González-Díaz AJ, Díaz-Hernández MB (2008) Genetic
 628 assessment of local apple cultivars from La Palma, Spain, using simple sequence repeats
 629 (SSRs). *Sci Hortic (Amsterdam)* 117:160–166. doi: 10.1016/j.scienta.2008.03.033
 630 Pereira-Lorenzo S, Urrestarazu J, Ramos-Cabrer AM, et al (2017) Analysis of the genetic
 631 diversity and structure of the Spanish apple genetic resources suggests the existence of an
 632 Iberian genepool. *Ann Appl Biol* 171:424–440. doi: 10.1111/aab.12385
 633 Pérez-Romero LF, Suárez MP, Dapena E, Rallo P (2015) Molecular and morphological
 634 characterization of local apple cultivars in southern Spain. *Genet Mol Res* 14:1487–1501.
 635 doi: 10.4238/2015.February.20.4
 636 Pina A (2005) Descripción de variedades tradicionales del banco de germoplasma nacional de
 637 manzano. Universidad de Zaragoza
 638 Pina A, Urrestarazu J, Errea P (2014) Analysis of the genetic diversity of local apple cultivars
 639 from mountainous areas from Aragon (Northeastern Spain). *Sci Hortic (Amsterdam)* 174:1–
 640 9. doi: 10.1016/j.scienta.2014.04.037
 641 Poles L, Micheletti D, Banchi E, et al (2018) Genetic diversity investigation of the apple
 642 germplasm available at the Fondazione Edmund Mach. In: *Proceedings IV International*

643 Symposium on Molecular Markers in Horticulture. Acta Horticulturae, pp 155–163

644 Porras-Hurtado L, Ruiz Y, Santos C, et al (2013) An overview of Structure: applications,
 645 parameter settings, and supporting software. Front Plant Sci 4:1–13. doi:
 646 10.3389/fgene.2013.00098

647 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
 648 genotype data. Genetics 155:945–59

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

651 Reddy MP, Sarla N, Siddiq EA (2002) Inter simple sequence repeat (ISSR) polymorphism and
 652 its application in plant breeding. Euphytica 128:9–17. doi: 10.1023/A:1020691618797

653 Rivas Martínez S (1987) Memoria del mapa de series de vegetación de España, 1ª Edition.
 654 I.C.O.N.A. Ministerio de Agricultura, Pesca y Alimentación, Madrid

655 Royo JB, Itoiz R (2004) Evaluation of the discriminance capacity of RAPD, isoenzymes and
 656 morphologic markers in apple (*Malus × domestica* Borkh.) and the congruence among
 657 classification. Genet Resour Crop Evol 51:153–160

658 Royo JB, Miranda C, Urrestarazu J, et al (2017) Lleida University apple and pear germplasm
 659 collection. Molecular identification

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

663 Sokal RR, Michener CD (1958) A Statistical Methods for Evaluating Relationships. Univ Kansas
 664 Sci Bull 38:1409–1448

665 Tessier C, David J, This P, et al (1999) Optimization of the choice of molecular markers for
 666 varietal identification in *Vitis vinifera* L. Theor Appl Genet 98:171–177. doi:
 667 10.1007/s001220051054

668 Urrestarazu J, Denancé C, Ravon E, et al (2016) Analysis of the genetic diversity and structure
 669 across a wide range of germplasm reveals prominent gene flow in apple at the European
 670 level. BMC Plant Biol 16:1–20. doi: 10.1186/s12870-016-0818-0

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

674 Uzun A, Ozongun S, Gulsen O, et al (2016) Determination of genetic relatedness among Turkish
 675 apple germplasm based on ISSR markers. J Appl Bot Food Qual 89:. doi:
 676 10.5073/jabfq.2016.089.010
 677 Van Laere K, Hokanson SC, Contreras R, Van Huylenbroeck J (2018) Woody ornamentals of
 678 the temperate zone. In: Developments in Crop Science. pp 803–887
 679 Venturi S, Dondini L, Donini P, Sansavini S (2006) Retrotransposon characterisation and
 680 fingerprinting of apple clones by S-SAP markers. Theor Appl Genet 112:440–444. doi:
 681 10.1007/s00122-005-0143-8
 682 Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes
 683 in natural populations: cautions and guidelines. Mol Ecol 10:249–256. doi: 10.1046/j.1365-
 684 294X.2001.01185.x
 685 Wichmann B, Galli ZS, Molnár S, et al (2018) Molecular identification of old Hungarian apple
 686 varieties. Int J Hortic Sci 13:1–9. doi: 10.31421/ijhs/13/3/743
 687 World Apple and Pear Association (2010) European apple and pear crop forecast. Brussels,
 688 Belgium
 689 Zhao G, Dai H, Chang L, et al (2010) Isolation of two novel complete Ty1-copia
 690 retrotransposons from apple and demonstration of use of derived S-SAP markers for
 691 distinguishing bud sports of *Malus domestica* cv. “Fuji”. Tree Genet Genomes 6:149–159.
 692 doi: 10.1007/s11295-009-0236-7
 693

694 Table 1: Set of 41 studied accessions of 23 traditional apple cultivars from the IMIDRA
695 collection and three commercial cultivars used as reference. Ploidy level was estimated
696 according to the maximum number of alleles found in all loci.
697

Traditional name	Accession abbreviation	Ploidy level	Municipality/Collection
Agridulce	MAMO	3	Montejo de la Sierra
Amarillo de El Paular	APRA	2	Rascafría
Camuesa	CABU	2	Bustarviejo
	CAVA	2	Valdemanco
Camueso tardío	CTAP	2	Puebla de la Sierra
Camueso temprano	CTEP	2	Puebla de la Sierra
de Chapa	CHCA	2	Canencia
	CHVA	2	Valdemanco
del Ortel	ORCA	3	Canencia
de Rosa	RORA	2	Rascafría
Esperiega	ESPU	2	Puebla de la Sierra
Hojancas	MHHO	3	Horcajuelo de la Sierra
	MHPR	2	Prádena del Rincón
Maílla blanca	MBHI	2	La Hiruela
Pepita de melón	PECA	2	Canencia
	PEHI	2	La Hiruela
	PEVA	2	Valdemanco
Pero de Aragón	PAHO	2	Horcajuelo de la Sierra
	PAPR	2	Prádena del Rincón
	PAPU	2	Puebla de la Sierra
Pero gordo	PGHI	2	La Hiruela
Pero pardo	PPHI	2	La Hiruela
	PPMO	2	Montejo de la Sierra
Pero real	PRBU	2	Bustarviejo
Rabudas	RAHI	2	La Hiruela
Reineta	REHO	3	Horcajuelo de la Sierra
	REMO	3	Montejo de la Sierra
	REPR1	3	Prádena del Rincón
	REPR2	3	Prádena del Rincón
	REPU	3	Puebla de la Sierra
Rojillo	RJHO	2	Horcajuelo de la Sierra
	RTAV	2	Valdemanco
	RTEV	2	Valdemanco
Rojo	RJPR	2	Prádena del Rincón
	RJPU	2	Puebla de la Sierra
San Felipe	SFCA	2	Carabaña
Temprano	TECH	2	Chinchón
	TEPI	2	Pinilla del Valle
Verde doncella	VDCA	2	Canencia
	VDHO	2	Horcajuelo de la Sierra
	VDTI	2	Tielmes
Reference cultivars			
Fuji	Fuji	2	IMIDRA
Granny Smith	Granny Smith	2	IMIDRA
Reineta Gris	Reineta Gris	3	EEAD-CSIC

698 Table 2. Microsatellite code, sequence, linkage group, PCR details and size range (bp) of 13 SSR loci used in this study.

SSR name	Direction	Primer sequence (5'-3')	Primer Length (bp)	Melting Temp. °C	Locus	Linkage Grouping	Fluorophore	Product Length (bp)	Reference
CH01f02	F	ACCACATTAGAGCAGTTGAGG	21	59.4	single	12	FAM	174-206	Liebhard et al., 2002
	R	CTGGTTTGTTCCTCCAGC	20	58.4					
CH01h01	F	GAAAGACTTGCA GTGGGAGC	20	60.5	single	17	FAM	104-134	Liebhard et al., 2002
	R	GGAGTGGGTTTGAGAA GGTT	20	58.4					
CH01h10	F	TGCAAAGATAGGTAGATATATGCCA	25	60.9	single	08	PET	94-114	Liebhard et al., 2002
	R	AGGAGGGATTGTTTGTGCAC	20	58.4					
CH02c09	F	TTATGTACCAACTTTGCTAACCTC	24	60.1	single	15	VIC	233-257	Liebhard et al., 2002
	R	AGAAGCAGCAGAGGAGGATG	20	60.5					
CH02c11	F	TGAAGGCAATCACTCTGTGC	20	58.4	single	10	PET	219-239	Liebhard et al., 2002
	R	TTCCGAGAATCCTCTTCGAC	20	58.4					
CH02d08	F	TCCAAAATGGCGTACCTCTC	20	58.4	single	11	VIC	210-254	Liebhard et al., 2002
	R	GCAGACACTCACTCACTATCTCTC	24	65.2					
CH03d07	F	CAAATCAATGCAAACTGTCA	21	53.5	single	06	VIC	186-226	Liebhard et al., 2002
	R	GGCTTCTGGCCATGATTTTA	20	56.4					
CH04c07	F	GGCCTTCCATGTCTCAGAA G	20	60.5	single	14	VIC	98-135	Liebhard et al., 2002
	R	CCTCATGCCCTCCACTAACA	20	60.5					
CH04e05	F	AGGCTAACAGAAATGTGGTTTG	22	58.4	presumed multi-locus	07	PET	174-227	Liebhard et al., 2002
	R	ATGGCTCCTATTGCCATCAT	20	56.4					
CH05f06	F	TTAGATCCGGTCACTCTCCA CT	22	62.1	single	05	NED	166-184	Liebhard et al., 2002
	R	TGGAGGAA GACGAA GAA GAAAG	22	60.3					
GD12	F	TTGAGGTGTTTCTCCCATTTGGA	22	60.3	unknown	03	NED	141-191	Hokanson et al., 1998
	R	CTAACGAAGCCGCCATTTCTTT	22	60.3					
GD147	F	TCCCGCCATTTCTCTGC	17	54.8	single	13	PET	135-155	Hokanson et al., 1998
	R	GTTTAAACCGCTGCTGCTGAAC	22	62.1					
Hi02c07	F	AGAGCTACGGGGATCCAAAT	20	58.4	presumed multi-locus	01	VIC	108-149	Silfverberg-Dilworth et al., 2006
	R	GTTTAAGCATCCCGATTGAAAGG	23	61.1					

699

Table 3. Genetic diversity of the 41 traditional apple accessions studied with 12 SSR: number of alleles per locus (A), number of effective alleles (A_e), number of rare alleles per locus (B), observed (H_o), expected heterozygosity (H_e), probability of identity (P_{ID}), observed genotypes (G_o), possible genotypes (G_p) and representativeness, or percentage of observed genotypes over possible (R). In brackets, values including both diploid and triploid accessions. The overall P_{ID} was = 3.60×10^{-12} .

Locus	A	A_e	B	H_o	H_e	P_{ID}	G_o	G_p	R
CH01f02	12 (13)	5.51	6	0.94	0.84	0.06	14	78	17.95
CH01h01	8	4.28	3	0.85	0.77	0.13	14	36	38.89
CH01h10	5	3.56	0	1.00	0.72	0.07	7	15	46.67
CH02c09	7 (9)	3.33	3	0.79	0.70	0.19	8	28	28.57
CH02c11	9 (11)	7.14	2	0.94	0.86	0.05	11	45	24.44
CH02d08	8 (10)	3.29	4	0.76	0.70	0.18	10	36	27.78
CH03d07	8 (9)	4.36	3	0.88	0.77	0.13	13	36	36.11
CH04c07	9 (10)	4.60	4	0.97	0.78	0.11	11	45	24.44
CH04e05	9	2.63	5	0.70	0.62	0.09	11	45	24.44
CH05f06	9 (11)	3.84	5	0.67	0.74	0.15	12	45	26.67
GD12	4 (5)	3.05	1	0.82	0.67	0.25	6	10	60.00
Hi02c07	4 (5)	1.57	0	0.24	0.36	0.08	4	10	40.00
Total	92 (105)	47.16	36	-	-	-	-	-	-
Average	7.66 (8.75)	3.93	3.00	0.80	0.71	0.12	10.08	35.75	33.00

708 Table 4: Accumulation of SSR loci required to discriminate all 21 diploid unique genotypes. The remaining 12 diploid accessions are
709 separated by a slash (/) as scored identical SSR profile.

SSR combination	Unique genotypes discriminated	Accessions discriminated	Unique genotypes remaining (%)
CH03d07	9	CTAP, ESPU, MBHI, PAHO/PAPU/PGHI, PAPR/PPII/RJHO/RJPR/RJPU/RTEV, PPMO, TEPI, VDCA, VDTI APRA, CTEP, CABU/CHVA, MHPR/SFCA/TECH, PECA/PEHI, PEVA,	49
CH03d07+CH01f02	19	RAHI, RORA, PRBU/RTAV, VDHO	90
CH03d07+CH01f02+CH02d08	21	CAVA, CHCA	0

710

711

Figure 1. Observed genotype distribution in the 33 traditional diploid apple accessions. Homozygosis was assumed when only one peak was detected in the sequencer.

Figure 2. Principal Coordinates Analysis (PCoA) based on 12 SSR loci from 35 diploid genotypes (33 traditional accessions and a 'Fuji' and a 'Granny Smith'). The small graph in the bottom left corner represents the sedimentation of the Principal Coordinates, with a cumulative value of 52.10% for PCo1 and PCo2.

Figure 3. Bayesian population structure implemented in the STRUCTURE software and based on 12 SSR loci. a) Population structure for $k=4$ of 33 diploid local accessions and 'Fuji' and 'Granny Smith'; colours in this figure are the same as those used for the groups obtained in PCoA (Figure 2). b). Population structure for $k=4$ for 8 triploid local accessions and 'Reineta Gris'.

745 *Supplemental Table 1: SSR profile of the 41 studied accessions and three reference cultivars through 12 microsatellites. In bold, homozygous loci; in*
746 *italics, not fully heterozygous loci in triploid accessions.*

Accession	CH01f02	CH01h01	CH01h10	CH02c09	CH02c11	CH02d08	CH03d07	CH04c07	CH04e05	CH05f06	GD12	Hi02c07
MAMO	170/180/188	<i>116/118</i>	91/97	243/255	208/216/230	211/229/254	193/207/227	<i>108/112</i>	175	174/178/182	<i>150/152</i>	114
APRA	170/194	116	91/101	251/253	228/236	211/248	197/227	108/112	175	186/190	150/184	114
CABU	174/186	112/118	91/97	249	228/232	254	193/197	98/116	175/218	180	154/184	114
CAVA	170/186	112	97/115	249/251	206/230	254	193/197	98/116	175/224	180/182	154/184	114
CTAP	170/184	118	97/107	249/251	214/234	211/254	207/209	98/108	175/200	182/186	148/184	114
CTEP	174/186	116/128	97/101	249	214/230	211/248	187/197	102/112	200/202	178	150/154	104/114
CHCA	170/186	112	97/115	249/251	206/230	213/254	193/197	98/116	175/224	180/182	154/184	114
CHVA	174/186	112/118	91/97	249	228/232	254	193/197	98/116	175/218	180	154/184	114
ORCA	170/182/186	112/116/118	<i>97/107</i>	245/247/251	214/234/238	<i>217/248</i>	<i>187/193</i>	98/108/120	175	<i>176/188</i>	<i>150/154</i>	114
RORA	184/206	118/120	91/97	249/251	214/222	213/248	193/225	108/112	175/200	188/190	150/154	114
ESPU	170/174	116/120	91/97	249	228/236	213/254	227	96/98	175/200	178/180	154	104/114
MHHO	170/182/186	<i>116/130</i>	<i>97/101</i>	245/255	208/230	225/229/254	<i>207/227</i>	<i>108/112</i>	<i>175/202</i>	174/178/182	150	114
MHPR	182/206	118/120	91/97	249/251	214/222	213/248	193/225	96/120	175/200	180	150/154	114
MBHI	170	112/114	91/97	249/253	228	211/248	187/193	96/98	200/224	180	150	104/114
PECA	170/194	112/116	101/107	251/253	228/232	254	193/197	98/108	175/224	178/180	154	108/114
PEHI	170/194	112/116	101/107	251/253	228/232	254	193/197	98/108	175/224	178/180	154	108/114
PEVA	170/188	112/118	97/115	249/251	228/232	254	193/197	98/116	175/224	180	154/184	108/114
PAHO	160/170	114/118	91/97	243/247	214/230	248/254	193	96/110	175/206	180	150/154	112
PAPR	170/184	112/118	97/107	249/251	234/238	248/254	197/209	98/108	175	178/186	154/184	114
PAPU	160/170	114/118	91/97	243/247	214/230	248/254	193	96/110	175/206	180	150/154	112
PGHI	160/170	114/118	91/97	243/247	214/230	248/254	193	96/110	175/206	180	150/154	112
PPHI	170/184	112/118	97/107	249/251	234/238	248/254	197/209	98/108	175	178/186	154/184	114
PPMO	172/192	118/120	91/97	249/253	234/238	211/256	193/203	98	175	178/186	154/184	104/114
PRBU	170/174	116/118	91/115	251	234/238	211/254	193/197	98/108	175	178/186	154/184	114
RAHI	170	112/114	101/107	249	238	254	187/197	96/98	175/226	180/186	148/184	114
REHO	<i>178/180</i>	112/116/130	<i>91/97</i>	231/243/253	208/216/230	211/229/254	207/215/227	106/110/112	<i>175/202</i>	174/178/182	<i>150/152</i>	<i>114/116</i>
REMO	<i>178/180</i>	112/116/130	<i>91/97</i>	231/243/253	208/216/230	211/229/254	207/215/227	106/110/112	<i>175/202</i>	174/178/182	<i>150/152</i>	<i>114/116</i>
REPR1	<i>160/170</i>	112/116/130	<i>97/115</i>	<i>240/251</i>	208/232	211/225/254	207/215/227	106/110/112	<i>175/202</i>	<i>178/186</i>	<i>150/154</i>	<i>104/108</i>
REPR2	<i>178/180</i>	112/116/130	<i>91/97</i>	231/243/253	208/216/230	211/229/254	207/215/227	106/110/112	175/202	174/178/182	<i>150/152</i>	<i>114/116</i>
REPU	<i>178/180</i>	112/116/130	<i>91/97</i>	231/243/253	208/216/230	211/229/254	207/215/227	106/110/112	175/202	174/178/182	<i>150/152</i>	<i>114/116</i>
RJHO	170/184	112/118	97/107	249/251	234/238	248/254	197/209	98/108	175	178/186	154/184	114
RTAV	170/174	116/118	91/115	251	234/238	211/254	193/197	98/108	175	178/186	154/184	114
RTEV	170/184	112/118	97/107	249/251	234/238	248/254	197/209	98/108	175	178/186	154/184	114
RJPR	170/184	112/118	97/107	249/251	234/238	248/254	197/209	98/108	175	178/186	154/184	114
RJPU	170/184	112/118	97/107	249/251	234/238	248/254	197/209	98/108	175	178/186	154/184	114
SFCA	182/206	118/120	91/97	249/251	214/222	213/248	193/225	96/120	175/200	180	150/154	114
TECH	182/206	118/120	91/97	249/251	214/222	213/248	193/225	96/120	175/200	180	150/154	114
TEPI	170/186	114/120	91/97	249	222/238	213/254	197/225	98/104	204/228	180/186	148/184	114
VDCA	182/206	118/130	91/97	247/253	234/238	229/254	203/227	108/110	175/202	170/182	150/154	114
VDHO	178/186	118	97/107	247/253	222/228	248	197/227	96/98	202/206	182/184	150	108/114
VDTI	160/186	128/140	91/97	231/255	230/232	205/224	193/207	108	175/218	164/182	150	114
Reference cultivars												
Fuji	182	116	91/97	231/243	230/234	211	227	106/118	175/202	164/174	148/154	116
Granny Smith	180/206	112/130	97	231/241	228/232	211/250	193/227	106/112	175	170/174	150/154	114/118
Reineta Gris	<i>178/180</i>	112/116/130	<i>91/97</i>	231/243/253	208/216/230	211/229/254	207/215/227	106/110/112	<i>175/202</i>	174/178/182	<i>150/152</i>	<i>116/118</i>

747

748 *Supplemental Figure 1. Areas of origin of the studied accessions (areas outlined in black) and location of IMIDRA collection (yellow square). MDT25 2015*
 749 *y BDLJE 2018. CC-BY 4.0 ign.es.*

Supplemental Figure 2. Example of unadjusted product amplification in the locus CH02d08 for four samples: CHCA, 'Granny Smith', TECH and SFCA.

753
 754 *Supplemental Figure 3. Loadingplots indicating the most discriminating alleles in PCo 1 and PCo 2 for the PCoA based on 12 SSR loci from 35 diploid*
 755 *genotypes (33 traditional accessions and a 'Fuji' and a 'GrannySmith'). The orange line indicates the threshold used in this study (0.07).*

756