scientific reports

Check for updates

OPEN

Unlocking Spanish pear genetic diversity: strategies for construction of a national core collection

Patricia Irisarri[1,2,11], Jorge Urrestarazu[3,4,11], Ana Ramos-Cabrer[5,11], Santiago Pereira-Lorenzo[5], María Encarnación Velázquez-Barrera[6,7], María Belén Díaz-Hernández[5], Enrique Dapena[8], Valero Urbina[9], Josep Dalmases[9], Domingo Ríos-Mesa[6,7], Sara Crespo-Martínez[4], Maite Loidi[4], Luis Gonzaga Santesteban[3,4], Javier Ascasíbar-Errasti[10], Pilar Errea[1,2], Carlos Miranda[3,4]✉ & Ana Pina[1,2]✉

Spanish pear germplasm collections are crucial for preservation, research, and breeding efforts. However, genetic diversity and structure is unknown at national level. A coordinated national project analyzed 1251 accessions from 7 Spanish pear collections using an internationally recognized set of 14 SSRs to enhance the utilization of these collections. Key findings included the identification of 760 unique genotypes (490 diploids and 270 triploids). Notably, genotypes represented by a single accession accounted for 49% of the total, indicating high vulnerability of this material. Using a Bayesian clustering method revealed two main genetic groups, G1 containing most foreign cultivars and G2 retaining local Spanish cultivars, which were further divided into two other subgroups using a nested approach, revealing moderate but significant differentiation among them. The populations were renamed according to the origin of the reference samples assigned to each group as 'South' (G1.1), 'Western Europe-1' (G1.2), 'Western Europe-2' (G2.1) and 'No-Pyrus communis' (G2.2). The results led to the creation of a 'generalist' collection, aiming to maximize genetic diversity representativeness, starting with 68 genotypes but expanding to 111 to achieve better allele recovery. This core collection is a valuable resource for genetic studies and conservation, enhancing efforts to preserve pear biodiversity.

Keywords Biodiversity, Conservation, Core collection, Genebank management, Microsatellite (SSRs), Pyrus

European pear (*Pyrus communis* L.) is an important fruit crop grown in temperate climate regions. Despite its wide geographical distribution, global pear production is highly dependent on a limited number of cultivars. Recent years have witnessed significant shifts in agricultural practices, with traditional pear varieties being replaced by more productive alternatives that offer improved economic outcomes. This trend, coupled with the dominance of a small number of widely distributed cultivars in pear production and breeding, poses a potential threat to genetic diversity within the species. The annual global production of pears amounts to approximately 26.32 million tons, with Europe contributing 9.7% to this total. Spain ranks as the fourth largest producer of pears in Europe[1]. Spain emerges as a significant contributor, with Catalonia leading in pear production with

[1]Departamento de Ciencia Vegetal, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Zaragoza 50059, Spain. [2]Instituto Agroalimentario de Aragón-IA2 (CITA-Universidad de Zaragoza), Zaragoza 50013, Spain. [3]Instituto de Investigación Multidisciplinar en Biología Aplicada (IMAB), Universidad Pública de Navarra (UPNA), Pamplona 31006, Spain. [4]Departamento de Agronomía, Biotecnología y Alimentación, Universidad Pública de Navarra (UPNA), Pamplona 31006, Spain. [5]Departamento de Producción Vexetal e Proxectos de Enxeñaría, Universidade de Santiago de Compostela (USC), Lugo 27002, Spain. [6]Centro de Conservación de la Biodiversidad Agrícola de Tenerife, Cabildo Insular de Tenerife, Puerto de la Cruz 38400, Spain. [7]Departamento de Ingeniería Agraria y del Medio Natural, Universidad de La Laguna, San Cristóbal de La Laguna 38071, Spain. [8]Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA), Villaviciosa 33300, Spain. [9]Departament d'Hortofruticultura, Jardineria i Botànica Lleida, Universitat de Lleida (UDL-IRTA), Lleida 25178, Spain. [10]Centro de Investigaciones Agrarias de Mabegondo. Xunta de Galicia, A Coruña, Spain. [11]Patricia Irisarri, Jorge Urrestarazu and Ana Ramos-Cabrer contributed equally to this work. ✉email: carlos.miranda@unavarra.es; apina@cita-aragon.es

around 45% of the total, followed closely by La Rioja and Aragon. An example of the dominance of a small number of cultivars is found in the main pear varieties grown in Spain, which are 'Conference', 'Blanquilla', 'Ercolini', 'Dr Jules Guyot', and 'Williams'[2]. Other cultivars such as 'Doyenne du Comice', 'Mantecosa Bosc' and 'Grand Champion' are cultivated at a smaller rate.

The awareness of this situation led to the development of initiatives that focus on the recovery and preservation of the genetic resource. Pear germplasm collections serve as an essential source of variation for breeding programs and play a key role in conserving diversity in an accessible way for the research community and stakeholders such as plant breeders in public, non-profit, and private research sectors. In these living collections, the confirmation of germplasm identity and pedigrees is a critical aspect of managing germplasm collections and maintaining genetic diversity within the collection. It is important to highlight the importance of fingerprinting the accessions housed in national collections, since it impacts not only collection curators and breeders, but also it has widespread implications of germplasm mix-ups to the research community as well as to commercial growers. Several studies used the approach of genetic characterization to harness diversity in pear germplasm collections, their use promoted by national and international genebanks[3,4]. Large ex-situ pear germplasm collections are held in genebanks worldwide, including Oregon[5], UK National Pear Collection at Brogdale[6] and Western Europe[7,8]. The largest pear gene bank in the world can be found in Oregon, maintained by the US Department of Agriculture, Agricultural Research Service (USDA-ARS), where there are more than 2,500 unique clones and seedlings[5]. The Spanish national *Pyrus* (pear) collections are coordinated by the INIA (National Institute of Agricultural and Food Research and Technology), and Spanish pear gene banks can mainly be found in northwestern (Asturias and Galicia), northeastern Spain (Navarre, Aragon and Catalonia), and Canary Island. All the pear accession "passport" data can be found at INIA database, which provide information on the environmental conditions in which the accession was originally cultivated, as well as latitude, longitude, and altitude, which are fundamental initial information for each accession stored in the bank. Many of the national accessions have been extensively characterized and evaluated for important phenotypic, agronomic and genetic traits[9–11]. However, these studies were conducted using different sets of phenotypic parameters and SSRs, making difficult the comparison of accession identities across studies.

Consequently, more attention has been paid to harmonize methods for easy comparison between species in recent years. The Spanish apple collection was a good starting point to investigate functional genetic variation and efficient utilization of germplasm collections useful for applied breeding efforts[12]. It should be taken into account that it is almost never possible to phenotype a large portion of the available germplasm due to high costs, challenges with adaptation, restricted facility resources and time pressure. Therefore, efficient utilization of germplasm collections can be time intensive, laborious, and expensive in the context of harnessing genetic resources. A possible solution is the establishment of core collections, defined as the smallest group of accessions that is representative of the whole genetic diversity within the collection[4,13]. These collections provide a meaningful genetic representation of the diversity within a given fruit species, allowing researchers to access a wide range of genetic and phenotypic traits. This is crucial for breeding programs, as it provides a rich gene pool for the selection of desirable traits such as disease resistance, climate adaptability, and fruit quality.

Despite the age of high-throughput genotyping and sequencing platforms, the usefulness of SSRs is still evidenced in a large number of studies to evaluate *Pyrus* diversity (identity, pedigree, and genetic diversity[3,14,15]). Previous works in pear (Pyrus spp.) germplasm collections have been conducted in various countries to assess genetic diversity and verify plant pedigrees, such as Sweeden[16], Italy[3,14], Hungary[8], Spain[10,11,15] and German and Romanian national collections[7]. In an attempt to allow accession comparisons between studies and germplasm collections, the European Cooperative Program for Plant Genetic Resources (ECP/GR) identified a standard set of 17 microsatellites (SSRs) to be used in future studies[6,17]. In addition, a newly created USPGR (US *Pyrus* Genetic Resources) fingerprinting set has been recently published[18], with similar results to SNP-based genotyping platforms (70 K pear AxiomTM array).

The aims of this study were: (1) to assess the germplasm identity of pear material grown in Spain (2) to elucidate the genetic diversity and relationships between the cultivated genotypes and the reference cultivars, and (3) to determine the genetic structure inherent in the Spanish germplasm with a standardized set of microsatellite markers, as well as the identification of the origins of its diversity. Finally, we define the Spanish pear collection, optimize the conservation strategy of Spanish pear biodiversity, outline the importance of conserving these valuable resources, and highlight the need for conservation strategies that are appropriately designed for pear species.

Results

Molecular characterization of Spanish pear resources using SSRs

The 14 SSRs data set was adjusted to allow comparison between the 7 Spanish pear collections (Fig. 1), using the product sizes of the reference cultivars as a guideline to determine size ranges and conversions. Genotypes were considered duplicates if they matched at all alleles in all 14 SSRs. The SSR analysis allowed the differentiation of 760 unique genotypes within the 1251 accessions evaluated (Table S1). All SSRs were highly polymorphic. The number of alleles per locus ranged between 17 (EMPc11) and 33 (CH03g07), with an average of 23.6 alleles per locus (Table 1). The observed heterozygosity (Ho) ranged from 0.37 (CH04e03) to 0.91 (CH01d09) with a mean of 0.77. The expected heterozygosity (He) varied from 0.39 (CH04e03) to 0.93 (CH01d09), with a mean of 0.82, and their discrimination power (PD) varied from 0.63 (CH04e03) to 0.99 (CH01d09). A significant number of rare alleles (frequency lower than 0.05) was identified with a mean of 18 alleles per locus, ranging from 10 (CH-Vf1 and EMPc11) and 27 (CH03g07). Moreover, a total of 44 unique alleles alleles (alleles present in only one accession) were found in 35 accessions. The Wright's fixation index (Fis) was also determined for the unique genotypes. The inbreeding coefficient ranged from 0.1349 (CH02b10) to -0.0496 (EMPc11) in the overall set of unique genotypes, with an average value of -0.0227 for all loci. This low value suggested no loss of heterozygosity

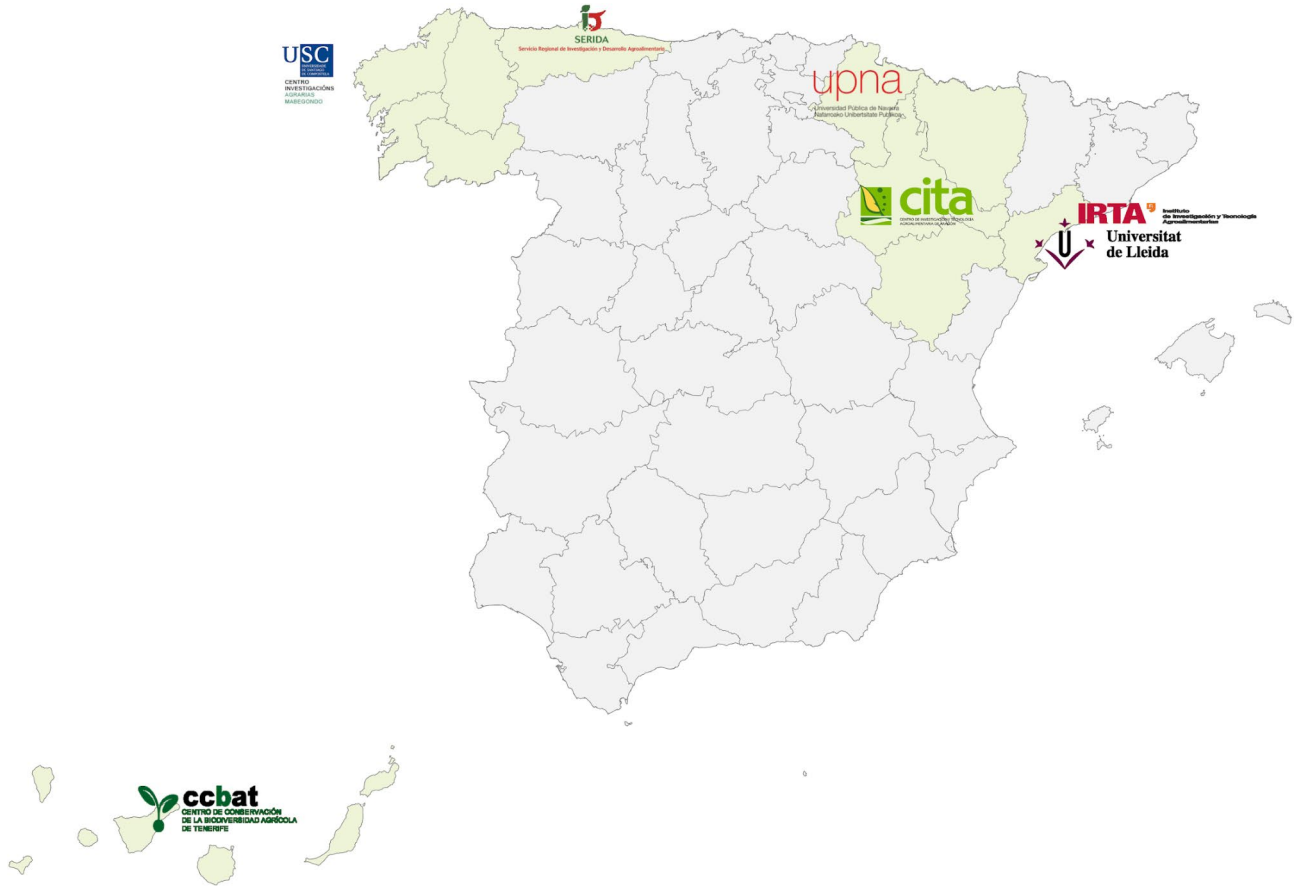

Fig. 1. Geographic location of the Spanish pear germplasm collections included in this study. UPNA (Universidad Pública de Navarra), CITA (Centro de Investigación y Tecnología Agroalimentaria de Aragón), USC (Universidade de Santiago de Compostela), Serida (Servicio Regional de Investigación y Desarrollo Agroalimentario), CCBAT (Centro de Conservación de la Biodiversidad Agrícola de Tenerife), and UdL (Universitat de Lleida-IRTA).

among the accessions analyzed. As many as 270 individuals out of 760 unique genotypes (35.5%) showed three alleles at two or more loci, and they were considered triploids. Considering all genotypes that showed three alleles in at least two locus, the highest number of genotypes with three alleles was found at EMPc11 and CH-Vf1 loci, with 100 individuals, respectively. The lower level of triploids was detected with CH04e03 with 22 accessions, among these, there were well known triploids, such as 'Roma', 'Beurré Alexander Lucas' and 'Don Guindo'.

Based on the microsatellite data, a 39% duplication degree was observed in the total 1251 accessions. Synonymies and homonymies were detected in the samples. The range of clonality varied between pear collections with the lowest values for CITA (17%), IRTA (16%) and Serida (15%) and the highest values for USC-CIAM (49%), CCBAT (45%), UPNA (36%) and UDL (34%). The comparison of SSR profiles revealed 152 groups of SSR duplicates involving 643 accessions, ranging from 2 to 33 accessions each one (Table S1). In total, 608 unique genotypes were represented by a single accession, indicating a situation of high vulnerability of these pear genetic resources held at the seven national collections. 'Williams' was the most represented genotype in these 152 groups of synonyms with 32 accessions within the group belonging to different geographical locations. The following was the Galician cv 'Urraca' with 26 accessions. Some duplicates were expected, as the accessions received very similar or identical denominations (e.g. the identity groups with 'Williams' and 'Williams Mollerusa'). However, most groups of duplicates comprised accessions that have different names (or whose name is unknown) and different geographical origin, suggesting that those plants have been spread through grafting. For instance, a group of 19 accessions with the same SSR profiles but different cultivar names: 'Manteca', 'Peral de agua', 'Bergamota' and 'Don Guindo' (genetic group 159) (Table S1). Additionally, different SSR profiles were obtained between the reference cultivar 'Williams' and 'Max Red Barlett' for three loci CH02b10, EMPc117 and GD147, despite the last one being a bud mutation derived from 'Williams'. Finally, some grafting propagation error was detected for the cultivars 'Malacara' and 'Magallon', given the same SSR profile at CITA collection compared to Navarra collection.

Locus	Allelic Range	A	B	C	He	Ho	PD	Fis	Alleles[a]
CH.Vf1	127–172	19	10	53	0.88	0.77	0.97	0.0793	**127**, 129, 131, 133, 135, **137**, 139, 141, **143**, **144**, **145**, 147, 149, 152, **154**, **156**, **158**, **162**, **172**
CH01d08	241–310	25	19	76	0.84	084	0.95	-0.0271	**241**, **243**, 245, **247**, **249**, **253**, **257**, **262**, **274**, **276**, **279**, 281, 283, **284**, **285**, 287, **289**, 291, **293**, **295**, **297**, 299, **301**, **305**, **310**
CH01d09	120–179	26	17	65	0.93	0.91	0.99	0.0015	**120**, **122**, **126**, **128**, **130**, 132, **134**, 136, **138**, 140, 142, 144, **146**, **148**, **152**, 154, 156, 158, 160, **162**, **164**, **168**, **172**, **174**, **176**, **179**
CH01f07	171–219	26	19	73	0.89	0.85	0.98	0.0141	**171**, **173**, **175**, 177, **179**, **181**, 182, **183**, 185, **187**, 189, 190, **191**, **193**, 195, **197**, **199**, **201**, **203**, **205**, **207**, 209, **212**, **214**, **216**, **219**
CH02b10	110–162	27	20	74	0.90	0.73	0.98	0.1349	**110**, **114**, **120**, 122, 124, **126**, 128, 130, 132, 134, **136**, 138, **139**, **141**, **143**, **145**, **146**, **147**, **148**, **149**, **151**, **153**, **154**, **155**, **157**, **159**, **162**
CH02c11	202–250	24	17	71	0.89	0.78	0.97	-0.0039	**202**, **206**, **208**, **210**, **212**, **214**, 216, 218, 220, **222**, 224, **226**, 228, **230**, **232**, **234**, **236**, 238, 240, **242**, **244**, **246**, **248**, **250**
CH03d12	85–161	27	22	81	084	0.73	0.95	0.0798	**85**, **90**, 92, **94**, **96**, **102**, **104**, **106**,108, 110, 112, **114**, **116**, **118**, **120**, **122**, 125, **127**, **129**, **132**, **134**, **139**, **142**, **149**, **157**, **159**,**161**
CH03g07	199–288	33	27	82	0.89	0.83	0.97	0.0121	**199**, **203**, **205**, **208**, **211**, **215**, **217**, **221**, **223**, **225**, 227, **229**, **231**, 233, **235**, **237**, **239**, 243, 245, **246**, **247**, 249, **251**, **253**, **255**, 257, **259**, **261**, **263**, **265**, **267**, **269**, **288**
CH04e03	176–212	14	12	86	0.39	0.37	0.63	0.0156	**176**, 179, **181**, **185**, **187**, **189**, **194**, **196**, **197**, **199**, **202**, 204, **206**, **212**
CH05c06	78–120	19	14	74	0.78	0.80	0.92	-0.0442	**78**, **82**, **84**, 88, **90**, 92, **94**, **96**, 98, **100**, **102**, **104**, **106**, 108, **110**, 112, **114**, **118**, **120**
EMPc11	125–171	17	10	59	0.83	0.85	0.96	-0.0496	**125**, **132**, **136**, 138, 140, **142**, 144, **145**, 146, **148**, 150, **152**, 154, 156, **158**, **160**, **171**
EMPc117	84–140	24	20	83	0.86	0.77	0.97	0.0825	**84**, **88**, **90**, **92**, **93**, **95**, **97**, **99**, **101**, **103**, **105**, **107**, **109**, **111**, 113, 115, 117, 119, **121**, **123**, **125**, **128**, **130**, **140**
GD142	126–204	31	25	81	0.90	0.87	0.98	0.0131	**126**, **128**, **134**, **138**, **140**, **143**, **146**, **148**, **150**, **152**, 154, **156**, 158, **160**, 162, **164**, 166, 168, **170**, **172**, **174**, 176, **178**, **180**, **182**, **184**, **186**, **188**, **194**, **198**, **204**
GD147	117–167	19	15	79	0.72	0.63	0.91	-0.0102	**117**, **121**, **123**, 125, 127, **129**, **131**, 133, 135, **137**, **139**, **141**, **143**, **145**, **150**, **154**, **156**, **162**, **167**
Mean		23.6	18	74	0.82	0.77	0.94	-0.0213	

Table 1. Measures of genetic diversity for the 14 SSR loci: allelic range (bp), number of alleles per locus (A), number of infrequent alleles (B) ($p < 0.05$), percentage of infrequent alleles with respect to the total number of alleles (C), expected heterozygosity (he), observed heterozygosity (Ho), discriminant power (PD), values of Wright's fixation index (Fis) and alleles identified in the overall set of pears cultivars. [a]The infrequent alleles ($p < 0.05$) are highlighted in bold

Structure population and differentiation

As a whole, 760 unique genotypes were analyzed by STRUCTURE since population structure can be used to optimize collection efforts and greatly reduce the number of individuals required to conserve allelic diversity. The Bayesian analysis of unique genotypes was processed with Structure Harvester and showed a most probable hierarchical division into two groups ($K = 2$, $\Delta K = 625$) (Fig. 2A). Furthermore, the results showed a less pronounced peak at $K = 4$ ($\Delta K = 25$), indicating that the germplasm could be divided into 4 groups. The plot of the average log-likelihood values for *K*s ranging from 1 to 10 and the distribution of *K*-values[19] according to *K*-values are shown in Fig. 2. The hierarchical genetic structure was investigated at $K = 2$, one group clustered 362 genotypes (G1) and the second 398 genotypes (G2). The proportion of genotypes assigned to each population was not symmetric, and many accessions were strongly assigned to one population or another, both considered strong indications that a real population structure exists[20]. Group 1 (G1) mostly included international cultivars such as 'Williams', 'Roma', 'Abbe Fetel' and 'Canal Red'; and group 2 (G2) is characterized by local varieties common in Spain ('Castell', 'Abugo', 'and 'Magallon') and the international cv. 'Beurre Hardy'.

A nested Bayesian clustering was applied to those two groups and two subgroups were found at each one, respectively. The first group (G1) analyzed showed a most probable division into 2 groups with $\Delta K \approx 1100$ and in the second group (G2) with $\Delta K = 75$ (Figs. 2B-C and 3).

Descriptive statistics for the four structure groups obtained are shown in Table 2. A threshold value of $qI \geq 0.80$ was used to strongly assign individuals to the groups. The affinity of most individuals with their assigned subgroups was strong, and the proportions of genotypes with high probability of availability for each subgroup were high. The classification of the accessions with ≥ 0.8 was 66% for the subgroup G1.1, 72.16% for the G1.2, 74.52% for the G2.1 and 72.04% for the G2.2. 216 genotypes (28.46%) could not be consistently assigned to any subgroup. Genetic diversity indexes were calculated for the four subgroups obtained by the Bayesian model-based clustering method (Table 2). The four subgroups were compared in terms of the total number of alleles, exclusive, mean expected heterozygosity (H_e), 1-D and Evenness index. All these diversity values were higher in G2.2 group compared to the other groups, followed by the G2.1 group. He varied from 0.74 (G1.2) to 0.84 (G2.2), revealing a high proportion of heterozygous individuals in the four subgroups. The percentage of alleles represented in each sub-group was 55. 0% (G1.1), 63. 1% (G1.2), 77. 6% (G2.1) and 88. 6% (G2.2). A total number of 2 alleles in 14 SSRs appeared as exclusive in G1.1 only, whereas 44 alleles were exclusive in G2.2. Interestingly, the mean number of alleles per locus in group G2.2 was higher (292) than in the other subgroups, despite the number of accessions forming the group (186) were lower than in G2.1 and G1.2 (212

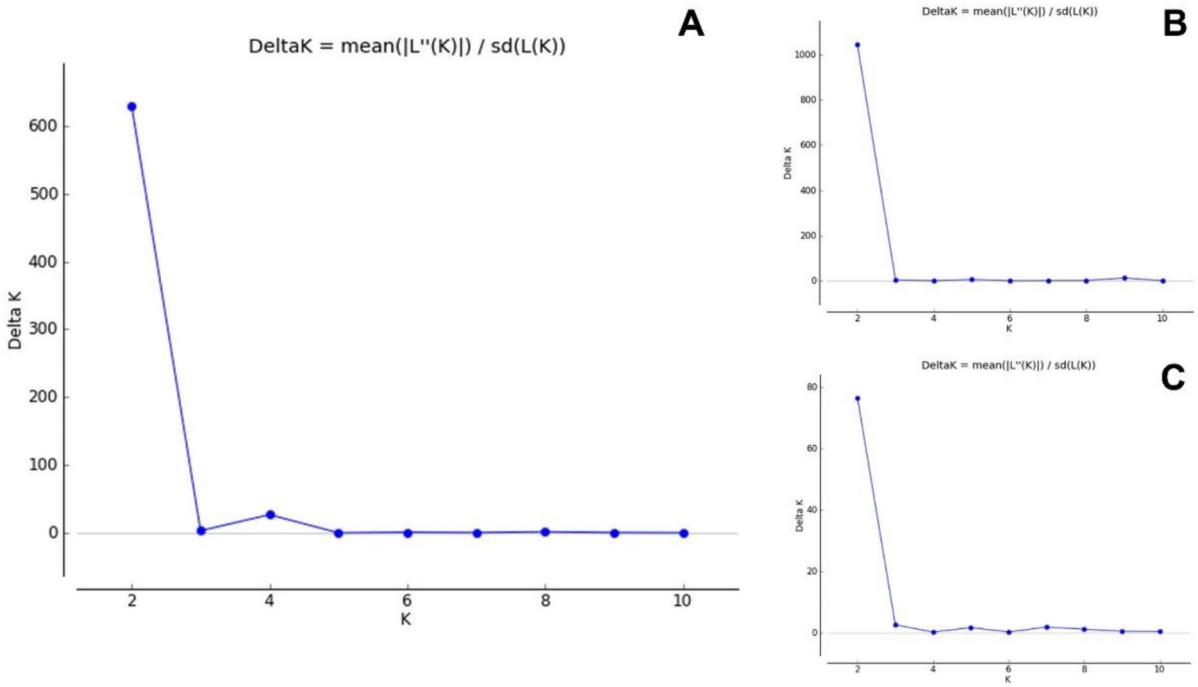

Fig. 2. Delta *K* values over 200 runs for increasing *K*-values, from 1 to 10. (**A**) Plot for the analysis on the 760 unique genotypes giving a robust ΔK maximum at K = 2. Plot for the sub-structure analysis for G1 (**B**) and G2 (**C**).

genotypes, respectively). Although there was no straightforward correspondence between the geographic origin of the accessions and their clustering, some trends can be noted. The four well-differentiated subgroups obtained were named according to the origin of the reference samples assigned to each. Hence, they were renamed 'South' (G1.1), 'Western Europe-1' (G1.2), 'Western Europe-2' (G2.1) and 'No-*Pyrus communis*' (G2.2).

The group called 'South' includes references such as 'Blanquilla' and some of Italian origin, 'Ercolini', 'Roma' and 'Coscia Precoce'. In the 'Western Europe-1' group (G1.2), some European reference cultivars were found, such as 'Beurre Guiffard', 'Monsallard', 'Williams', 'Jules Guyot', 'Flor de Invierno', 'Limonera', and 'Comice', among others. In the 'Western Europe-2' group, 3 reference cultivars are clustered, one of them ('Beurre Hardy') with strong assignation (Table 3). Finally, the 'No-*P.communis*' group contains Asian varieties such as 'Sinseiki', 'Nijiseiiki', 'Hosui' and also some local native cultivars such as 'Magallon' and 'Castell'. The genetic discrimination in subgroups found using Structure was also confirmed with minimum spanning network base using the SSR data. The results clearly showed that the four subgroups 'South, 'No-*P.communis*', 'Western Europe-1' and 'Western Europe-2' clustered separately (Fig. 4).

Based on the AMOVA analysis, significant variance differences were also observed between the four subgroups previously identified by the nested-Bayesian clustering. The overall F_{st} value of 0.059 suggested moderate but highly significant ($P < 0.001$) differentiation between subgroups. This value was higher than that obtained when the differentiation between major groups (G1 and G2) was considered (F_{st} 0.037, $P < 0.001$). The results obtained by spanning network and AMOVA were coherent with those obtained by Structure and clustering analysis (Fig. 1S).

Design of the Spanish pear core collection

For the definition of the core collection, 453 genotypes corresponding only to accessions were considered, after discarding those corresponding to reference cultivars (85 genotypes), prospections (79 genotypes corresponding to trees geotagged but not yet preserved in a collection) and, finally, 143 genotypes strongly assigned to groups pooling the international references in the STRUCTURE analysis performed using population information (Table S1). The diversity and distance values obtained for the optimization strategies are summarized in Table 4. As expected, each strategy best optimized the distance measures and criteria that were expressly included in the optimized core collection (OC) selection criteria. The pure strategies obtained better levels of optimization of A-NE (average distance between each genotype of the collection and the nearest entry) and E-NE (average distance between each entry and the nearest entry) mean distances, but in return their performance was worse when it came to optimizing SH (Shannon-Weaver diversity index) distance and, above all, allele recovery. Regarding the preservation of genetic structure and ploidy, the E-EC (Entry-based enhanced core) strategies (pure or mixed) produced the most unbalanced results, as they were strongly biased towards triploid genotypes

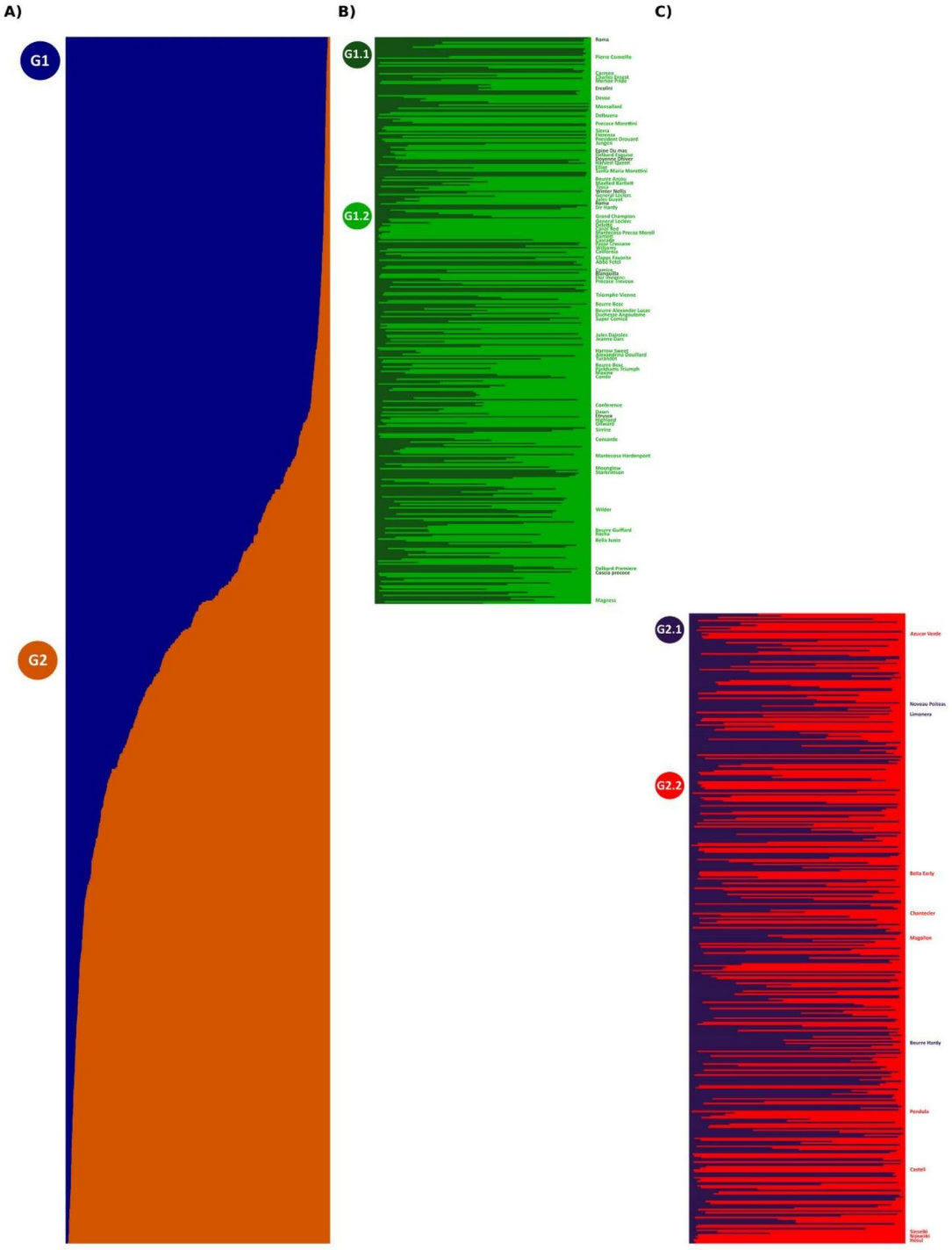

Fig. 3. Graphical display of the results of the STRUCTURE analyses (K = 2), nested-structure groups and placement of reference cultivars using 14 SSRs. (**A**) Structure analysis for the complete set of genotypes, (**B**) nested Structure analysis for the first subgroup (G1), (**C**) nested Structure analysis for the second subgroup (G2). Blue= 'G1', orange= 'G2', dark green = 'G1.1', light green = 'G1.2', dark blue = 'G2.1' and red = 'G2.2'.

and those from foreign population groups ('Western Europe'). The A-NE_P strategy selected genotypes from each reconstituted population in a proportion more similar to the original, especially for the larger OC, as well as maintaining a proportion of diploid genotypes more similar to the original. The A-NE_M strategy, although very efficient while maintaining distance optimization and allele recovery, showed a bias toward accessions assigned

	Group	Number of genotypes in the subgroups				% Genotypes	Number of alleles		Average		
		Genotypes	Local accessions	References[a]	Strongly assigned (qI > 0.8.)		Total	Exclusive	He	1-D	Evenness
'South'	G1.1	150	141	9	99	66.00	182	2	0.773	0.771	0.717
'Western Europe-1'	G1.2	212	146	66	153	72.17	209	7	0.740	0.738	0.664
'Western Europe-2'	G2.1	212	209	3	158	74.53	257	18	0.807	0.805	0.666
'No_Pyrus *communis*'	G2.2	186	177	9	134	72.04	292	44	0.836	0.834	0.695
Total		760	673	87	544						

Table 2. Summary descriptive information for each of the four genetic subgroups identified by bayesian clustering analyses using 14 SSR markers in the 760 unique genotypes. Summary statistics include the partitioning of number of individuals (local and reference cultivars) in each subgroup, accession percentage with a robust assignation to the group (q > 0.8), number of alleles (total and exclusive) and average expected heterozygosity (H_e), 1-D and evenness index.

in admixis to 'Western Europe' groups or associated with other *Pyrus* species, and to over-represent triploid genotypes.

The choice was made to create a 'generalist' or 'multipurpose' type collection, in which the aim was to maximize the representativeness of the genetic diversity of the collection. With this in mind, the most appropriate distance measure was A-NE, and A-NE_P, with 68 genotypes retained (15%), the most efficient strategy, as it optimized the distance criteria better than the others, and the representation of genetic structure and diversity were also the closest to the set of candidate genotypes. However, the allele recovery was moderate (77%), so an additional set of genotypes was selected to optimize this parameter as well. Table 5 shows the results of completing the 15% A-NE_P optimized core with the genotypes providing missing alleles (A-EN_P + full allele recovery). When multiple genotypes shared the same missing alleles, priority was given to those that optimized genetic distances, had historical significance, or offered unique phenotypic traits. In the final step, genotypes carrying previously unrecovered alleles were incorporated. If multiple genotypes had the same allele, the same expert knowledge selection process was applied and the final core collection involved 111 genotypes (25% of the candidates). The A-NE and SV distances were still better optimized than for other strategies at 15% size. Although the unique alleles recovered were mostly from genotypes assigned in admixis to the 'Western Europe' or 'wild pear' groups, the proportion of genotypes belonging to the Spanish pear still better than for any other strategy with almost full allele recovery.

Discussion

An in-depth analysis of genetic characterization has been performed for seven pear collections preserved at different locations in Spain, where pears have been grown for a long time. Most of the collections were established before molecular identification became feasible. At that time, the criteria used for selecting the germplasm to be maintained in collections were mainly based on morphology (pomology), eco-geography, and/or passport information. As a consequence, unintended redundancies are likely to occur within and between collections. Therefore, standardization of the molecular analyses and harmonization of the SSR sets are essential to facilitate the comparisons of genetic characterizations between different labs, and the findings found in this study have significant implications for the management and utilization of pear germplasm collections. In this work, 1251 pear accessions were analyzed with 14 SSRs (13 of the ECPGR-approved list and CH02c11), and high levels of polymorphism were encountered among the 760 unique accessions. The SSRs used displayed a high degree of polymorphism and discriminating power, highlighting the adequacy of selected markers and ensured the accuracy of the analysis to fingerprint germplasm collections[3,11,15]. Sixty-one per cent of the genotypes found in the Spanish collections were unique, most of them (49%) represented only by one accession, which reflects the vulnerability of this material. As expected, the observed heterozygosity at the locus CH04e03 was significantly lower than the values at the other 13 loci (0.37 vs. 0.91), meaning that a high number of individuals at this locus are homozygous, with the less discrimination power as reported by several authors[8,11,15,16]. The high level of diversity in pear germplasm found at the Spanish national level agreed with results obtained in other European countries, such as Spain[11,15], Italy[3], Sweden[16], Swiss[4]. The inclusion of *Pyrus pyrifolia* 'Hosui', 'Shinseiki' and 'Nijisseiki', *Pyrus sacilifocia* 'Pendula' and *Pyrus calleryana* 'Chantecler' in the reference set of cultivars, resulted in a wide range of allele sizes as reported by Evans et al.[6].

Regarding ploidy, the estimated percentage of triploids based on the number of genotypes with two or more loci with three alleles was 35.5%, slightly higher than in earlier studies in Italy[3,21], Spain mountain areas[11] and cultivars from Northwestern Spain[10] (20, 23.2 and 27%, respectively). Despite triploid cultivars are often selected and propagated by farmers due to their desirable phenotypic traits, such as larger fruit compared to diploid, incorporating these triploid genotypes into breeding programs remains challenging. This is mainly due to their irregular gamete formation and potential sterility[21].

Although the use of high-density SNP-based genotyping for pear germplasm characterization has recently increased[5,22], SSRs are still the markers of choice for routine fingerprinting analyses, especially in conservation studies where the priority is not represented by a high marker density per sample, rather by the inclusion of a large number of samples. In this study, we demonstrated that SSRs are still useful to avoid redundancy in the collections, reduce their management costs, and as a tool to provide true-to-type cultivars to nurseries and the

GR_K2	GG	Name_	*P*(Q)	GR_K2_1	*P*(Q)
2	97	Beurre Hardy	***0.962***	2.1	**0.980**
2	76	Limonera	0.728	2.1	**0.927**
2	132	Noveau Poiteau	0.712	2.1	**0.972**
2	298	Hosui (P. pyrifolia)	***0.991***	2.2	**0.970**
2	290	Nijiseiiki (P. pyrifolia)	***0.990***	2.2	**0.978**
2	287	Shinseiki (P. pyrifolia)	***0.990***	2.2	**0.978**
2	263	Pendula (P. salicifolia)	***0.972***	2.2	**0.988**
2	121	Castell	***0.979***	2.2	**0.981**
2	11	Magallon	***0.937***	2.2	0.567
2	310	Chanticleer (P. calleryana)	***0.931***	2.2	**0.983**
2	500	Bella Early	***0.904***	2.2	**0.955**
2	611	Azucar Verde	0.558	2.2	**0.943**
1	36	Blanquilla	***0.967***	1.1	**0.983**
1	625	Doyenne Dhiver	***0.980***	1.1	**0.976**
1	28	Roma	***0.991***	1.1	**0.967**
1	177	Roma	***0.976***	1.1	**0.936**
1	617	Coscia Precoce	0.639	1.1	**0.827**
1	143	Winter Nellis	***0.977***	1.1	0.554
1	66	Ercolini	***0.985***	1.1	0.540
1	127	Etrusca	***0.913***	1.1	0.528
1	126	Epine Du Mas	***0.981***	1.1	0.510
1	115	Abbe Fetel	***0.969***	1.2	**0.876**
1	591	Alexandrina Douillard	***0.944***	1.2	**0.808**
1	196	Bartlett	***0.974***	1.2	**0.973**
1	608	Bella Junio	0.710	1.2	**0.735**
1	117	Beurre Alexander Lucas	***0.959***	1.2	**0.772**
1	119	Beurre Anjou	***0.978***	1.2	0.562
1	118	Beurre Bosc	***0.940***	1.2	**0.968**
1	587	Beurre Bosc	***0.961***	1.2	**0.942**
1	1	Beurre Guiffard	0.727	1.2	0.756
1	614	California	***0.973***	1.2	**0.975**
1	618	Canal Red	***0.974***	1.2	**0.967**
1	623	Carmen	***0.986***	1.2	**0.825**
1	120	Cascade	***0.974***	1.2	**0.967**
1	122	Charles Ernest	***0.986***	1.2	0.692
1	597	Clapps Favorite	***0.970***	1.2	**0.978**
1	77	Comice	***0.968***	1.2	**0.948**
1	123	Concorde	***0.881***	1.2	**0.973**
1	615	Condo	***0.937***	1.2	**0.973**
1	124	Conference	***0.926***	1.2	**0.963**
1	592	Dawn	***0.915***	1.2	**0.966**
1	613	Delbard Exquise	***0.980***	1.2	**0.860**
1	593	Delbard Premiere	0.652	1.2	**0.800**
1	612	Delbuena	***0.982***	1.2	**0.981**
1	594	Delette	***0.975***	1.2	**0.928**
1	125	Devoe	***0.984***	1.2	**0.867**
1	595	Dir Hardy	***0.976***	1.2	**0.927**
1	155	Duchesse Angouleme	***0.958***	1.2	**0.959**
1	626	Elliot	***0.979***	1.2	0.542
1	620	Fiorenza	***0.981***	1.2	**0.982**
1	40	Flor Invierno	***0.967***	1.2	0.666
1	128	GeneralLeclerc	***0.976***	1.2	0.793
1	211	General Leclerc	***0.975***	1.2	**0.937**
1	129	Grand Champion	***0.976***	1.2	**0.858**
1	616	Harrow Sweet	***0.945***	1.2	**0.988**
1	621	Harvest Queen	***0.980***	1.2	**0.907**

Continued

GR_K2	GG	Name_	*P*(Q)	GR_K2_1	*P*(Q)
1	609	Highland	***0.898***	1.2	**0.926**
1	130	Jeanne Darc	***0.952***	1.2	**0.947**
1	598	Jules Dairoles	***0.953***	1.2	0.681
1	13	Jules Guyot	***0.976***	1.2	**0.974**
1	619	Jungen	***0.981***	1.2	**0.955**
1	599	Magness	0.516	1.2	**0.964**
1	600	Mantecosa Hardenpont	***0.866***	1.2	**0.913**
1	610	Mantecosa Precoz_Morell	***0.974***	1.2	**0.927**
1	131	Maxine	***0.937***	1.2	**0.838**
1	197	MaxRed Bartlett	***0.978***	1.2	**0.972**
1	601	Merton Pride	***0.986***	1.2	**0.968**
1	2	Monsallard	***0.984***	1.2	0.705
1	602	Moonglow	***0.838***	1.2	**0.954**
1	603	Onward	***0.897***	1.2	**0.972**
1	133	Packhams Triumph	***0.939***	1.2	**0.828**
1	134	Passe Crassane	***0.974***	1.2	**0.930**
1	135	Pierre Corneille	***0.988***	1.2	0.656
1	182	Precoce Morettini	***0.982***	1.2	**0.824**
1	136	Precoce Trevoux	***0.966***	1.2	**0.973**
1	137	President Drouard	***0.981***	1.2	**0.873**
1	138	Rocha	0.721	1.2	**0.952**
1	604	Santa Maria Morettini	***0.979***	1.2	0.708
1	605	Sierra	***0.981***	1.2	**0.948**
1	606	Sirrine	***0.888***	1.2	**0.986**
1	607	Starkrimson	***0.829***	1.2	**0.976**
1	139	Super Comice	***0.957***	1.2	**0.973**
1	140	Tosca	***0.977***	1.2	**0.814**
1	141	Triomphe Vienne	***0.962***	1.2	**0.956**
1	624	Turandot	***0.943***	1.2	0.788
1	142	Wilder	0.754	1.2	0.882
1	10	Williams	***0.973***	1.2	0.983

Table 3. Reference cultivars used in the study indicating group assignment by STRUCTURE analysis. GG (genetic group), name, and the probability of assignment for each genotype to the genetic groups is provided. Strongly assigned individuals ($qI > 0.8$.) Are indicated in bold.

research community. One of the key outcomes of this study was the identification of a high clonality (39%), suggesting that a high number of redundant accessions are being conserved. This has a direct utility for pear germplasm conservation since the elimination of unnecessary duplicates could lead to reduce the costs and efforts required for collection management. The results also highlight the importance of a coordinated analysis, integrating both phenotypic and genetic data, to optimize the effort for the knowledge of the real diversity available and to facilitate its subsequent use in breeding. Usually, pear trees are clonally propagated by grafting. Although grafting enhances the genetic conservation, also complicates germplasm curation, facilitating the mislabeling of individuals as homonyms or synonyms and hindering both conservation and utilization of these important genetic resources[23]. In the Spanish collections, we found some synonyms and homonyms, and also identified putative labeling errors. The majority of detected synonymies and homonymies agree with other studies[11,15,18]. In some instances, synonyms may be due to small changes in the names due to differences in local localization, 'Williams' vs. 'Williams Mollerusa' and, in other cases, a cultivar may have been renamed to associate its ripening with some seasonal events 'Pera de invierno'. The percentage of duplicates revealed in pear germplasm collections from other countries was 17% in Bosnia and Herzegovina[24], 29% in Sweden[16], 48% in Tunisia[25], whereas only 1.6% duplicates were found within the German cultivar collection[7]. Through more efficient curation practices, the diversity and potential of pear germplasm can be better conserved and utilized.

Using a Bayesian clustering method allowed the identification of two genetic groups: G1 containing most foreign cultivars and G2 retaining local Spanish cultivars, which were further divided in two other subgroups (G1.1, G1.2, G2.1, G2.2) by a nested approach revealing moderate but significant differentiation among them. These results revealed complexity of genetic diversity and important variations within both local and foreign pear cultivars. A closer examination of the individuals within each subgroup revealed that reference cultivars from Southern Europe ('Roma', 'Etrusca', 'Blanquilla', 'Ercolini') were allocated in subgroup G1.1. G1.2 ('Western Europe-1') included cultivars from Western Europe ('Barlett', 'Concorde', 'Beurre Bosc', 'Conference', 'General Leclerc'). Similarly, G2.1 ('Western Europe-2') included some reference cultivars from Western Europe but to a

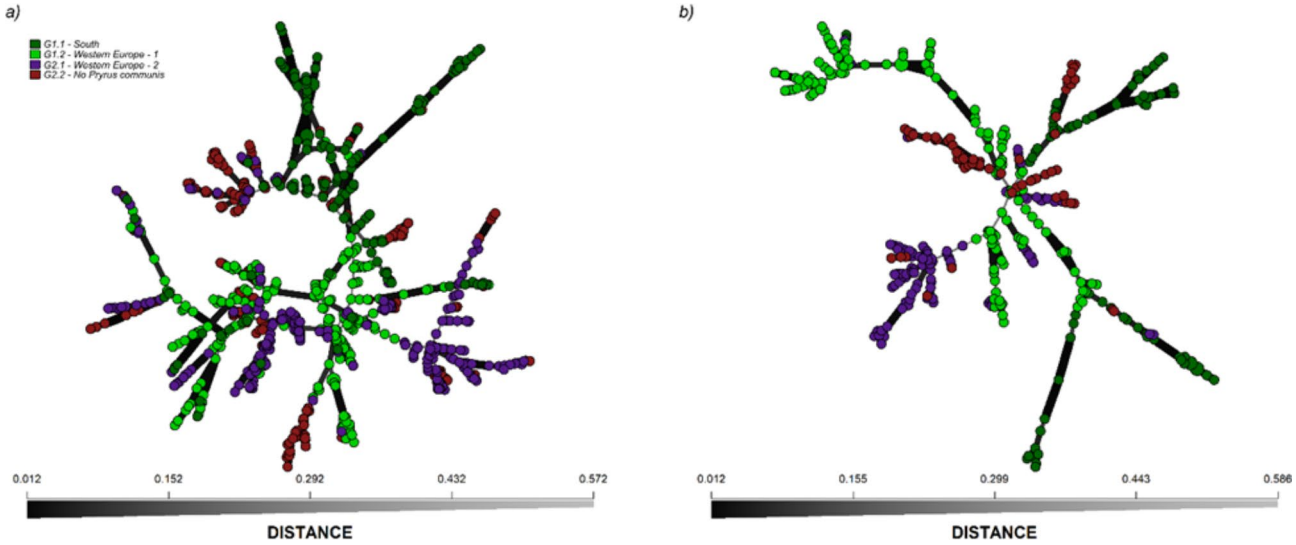

Fig. 4. Minimum spanning network of 14 simple sequence repeat loci (SSR s) from the 760 unique pear genotypes (**A**) and strongly assigned (**B**) belonging to the seven pears Spanish collections. Color codes for each genetic subgroups identified after nested Structure analysis: dark green= 'South', green light= 'Western Europe 1', dark blue = 'Western Europe 2' and red brown= 'No *P.communis*'.

Optimization strategy	Core size (%)	Average dist				Proportion of genotypes (%)		
		A-NE	E-NE	S_H	C_V (%)	Gr. 'South'	Gr. 'No Pyrus Com'	Diploids
A-NE_{P*}	5	0.36	0.43	4.75	0.51	30.4	43.5	56.5
	10	0.32	0.41	4.79	0.59	35.6	37.8	57.8
	15	0.29	0.44	4.98	0.77	27.9	45.6	48.5
A-NE_{M*}	5	0.38	0.50	5.18	0.76	21.7	60.9	34.8
	10	0.34	0.50	5.22	0.89	13.3	62.2	35.6
	15	0.30	0.48	5.19	0.96	14.7	52.9	36.8
E-NE_{P*}	5	0.53	0.70	5.06	0.66	13.0	56.5	4.3
	10	0.47	0.64	5.09	0.78	17.8	46.7	2.2
	15	0.41	0.60	5.10	0.86	20.6	42.6	2.9
E-NE_{M*}	5	0.50	0.64	5.24	0.79	8.7	60.9	8.7
	10	0.42	0.60	5.25	0.90	11.1	60.0	13.3
	15	0.36	0.56	5.23	0 0.95	11.8	52.9	20.6
A-NE_P and full recovery	25	0.25	0.41	5.06	1.00	19.8	46.8	55.9
Genotypes sampled from the full collection (n=453)						***27.8***	***38.0***	***62.0***

Table 4. Genetic parameters of the optimized core subsets selected by purely analytical procedures and by mixed procedures. A-NE: average distance between each genotype of the collection and the nearest entry, E-NE: average distance between each entry and the nearest entry, S_H: Shannon-Weaver diversity index, C_V: allelic coverage in percentage. $*_P$ and $_M$ means pure and mixed strategy, respectively.

lesser extension ('Beurre Hardy', 'Limonera', 'Noveau Poiteau' and 'Mantecosa Hardy'). The last group (G2.2) was more heterogeneous including *Pyrus pyrifolia* 'Hosui', 'Shinseiki' and 'Nijisseiki', *Pyrus salicifolia* 'Pendula', *Pyrus calleryana* 'Chantecler' as well as some *Pyrus communis* accessions and was renamed 'No-*Pyrus communis*'. Notably, the higher mean number of alleles per locus (292) and more exclusive alleles (44) in the subgroup G2.2 compared to others, indicated a high degree of genetic diversity within this group, which could be a key reservoir of genetic variation for breeding and conservation purposes. These populations were similar as in other studies that characterize the genetic diversity of national or local germplasm collections using genetic markers[7,10,11,26]. Remarkable differences between the inferred groups and subgroups were revealed at the allelic level, pointing out further evidence of a division of the germplasm into different partitioning levels. The results obtained by spanning network and AMOVA analysis were coherent with those obtained by Structure, validating the existence of genetic structure in the analyzed Spanish pear germplasm.

To the best of our knowledge, despite its relevance, few studies have been performed to design core collections in pear, which is crucial for the long-term conservation and use of pear diversity. In this study, the core collection

was established using a mixed approach that combine formal analytical procedures and 'expert knowledge', as it was reported in other studies with a good efficiency of CC definition[4,41]. Core collections (CCs) were developed with the objective of obtaining a conservation collection and, therefore, it was advisable to define them using the criteria for generalist CCs[13]. Ideally, each accession in the whole collection should be represented in the core by an entry that is most similar to itself. Thus, CCs have been validated following the indications by Odong et al.[13] that is, using preferably distance-based indices and criteria not used in the selection phase, supplemented by other classic indices suited to the evaluation of generalist collections such as Shannon Index (SH) and allele coverage (CV).

The CCs sampled by mixed strategies highlighted a trade-off in the effectiveness of optimizing genetic distances, so that the decision must be determined in terms of fitness-for-use. If our CCs had been selected with the aim to represent the extreme values in the collection, optimizing E-NE should have been the objective[13] and, therefore, E-NE_p would be the best-suited strategy. The differences in the final CCs sampled were small, revealing a certain "buffer" effect of the OC subset. Such effect could be expected, as the OC subset accounts for ca. 60% of the final CCs, and the genotypes in it were rarely included in the optimized subsets sampled by pure optimization of the distances. Similar results were obtained in the definition of the Swiss pear core collections[4], and confirm that mixed selection methods can generate CCs that are similarly efficient to those obtained by purely analytical methods. The final core collection included 111 genotypes, with Spanish pear cultivars which have played an important role in breeding history, are popular, prestigious or emblematic among local growers and/or consumers, or exhibit some phenotypic features of interest, such as ´Don Guindo', 'Pera de San Juan', 'Ceremeño', 'Blanquilla', 'Abugo', 'Tendral' and 'Pera de Canuel', among others.

Conclusion

The genetic diversity available in Spanish pear germplasm collections could be assessed efficiently using a set of 14 SSRs markers, allowing the identification of the accessions that can be prioritized for applied breeding efforts. The initial genetic pool of 1251 pear accessions were reduced to a total of 760 unique pear genotypes, with 49% of the genotypes represented by only one accession. These results reflected a situation of high vulnerability of these pear genetic resources held at the seven national collections. Furthermore, the differential grouping of international pear reference cultivars (mainly in G1) and most of the local Spanish accessions (in G2) suggested that the germplasm collection analyzed in this work represents a very peculiar and largely unexplored source of pear biodiversity. These groups were further divided in two other subgroups by a nested approach, and the four populations were renamed based on origin, resulting in groups: 'South' (G1.1), 'Western Europe-1' (G1.2), 'Western Europe-2' (G2.1), and 'No-*Pyrus communis*' (G2.2). Finally, for the first time, a Spanish pear core collection was defined using different strategies. 111 genotypes were proposed for establishing of the core collection (CC), in such a way that this CC could be useful when available resources do not allow to assay a larger number of plants and as a first step in genetic association studies. This core collection thus guarantees the preservation of rare alleles, which is critical for maintaining the genetic diversity of a population and likely involved in plant adaptation to environmental challenges. Nowadays, it would be interesting a collaborative effort between international *Pyrus* collections to compare data at international level since it would help the impact curation and management decisions as worldwide level, ensuring that labeling errors between repositories are not occurring and verifying representation of key accessions.

Methods

Plant material and DNA extraction

A total of 1251 accesions representing 7 Spanish pear collections were evaluated, 86 of them being local and foreign reference cultivars (Table 3, Table S1). Of the 1251 accessions, the majority are *Pyrus communis* (1246), *Pyrus pyrifolia* (3), one *Pyrus calleryana* ('Chanticleer') and one *Pyrus salicifolia* ('Pendula'). The accessions were maintained at seven locations from Northern Spain and the Canary Islands: 254 from Centro de Investigación y Tecnología Agraria de Aragón (CITA) in Zaragoza, Aragon; 103 from the Universidad Pública de Navarra in Pamplona, Navarra; 369 from the Centro de Investigaciones Agrarias de Mabegondo (CIAM) in Galicia, and 193 from the Cabildos (CCBAT) from Tenerife, La Palma and Gran Canaria, all evaluated at the Universidad de Santiago de Compostela (USC); 100 from the Servicio Regional de Investigación y Desarrollo de Asturias (SERIDA) in Villaviciosa, Asturias; 91 from the Instituto de Investigación y Tecnología Agroalimentaria (IRTA) in Lleida, Catalonia; and 141 from the Universidad de Lleida (UdL) in Gimenells, Catalonia. Serida accesions were analysed at CITA of Aragon and UdL accessions by the Universidad Pública de Navarra.

Newly expanded leaves of each accession were collected in spring (March to April), immediately frozen in liquid nitrogen and stored at -20 ºC until use. Genomic DNA was collected from young fresh leaves or vegetative buds following the procedure used by Hormaza[27]. Quantification of each DNA sample was performed using a NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, DE, USA) and all samples were diluted to a final concentration of 10 ng μL^{-1}.

Microsatellite amplification

14 pear SSR primer combinations were used (Table S2), including 13 SSRs recommended by the European Cooperative Program for Plant Genetic Resources (ECPGR)[6] and the remaining one (CH02c11) successfully used in other pear or apple diversity studies[11,15,28]. Two SSRs were in common with the USPGR set (CH01d08 and CH04e03).

Three multiplex PCRs were conducted, named MMA, MMB and MMC, in which forward primers were labelled with 6-FAM, VIC, NED or PET fluorescent at 5´ end (see for details Urrestarazu et al.[15] and Bielsa et al.[11]), with a final volume of 10 µL, using 10 ng of DNA template, 0.2, 0.4, or 0.6 µM of each primer and 1X

QIAGEN Multiplex PCR Master Mix (Qiagen, Holden, Germany). The temperature profile for all three PCR reactions encompassed an initial 15 min denaturation step at 95 °C, followed by 10 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, 55 °C for 1 min, 72 °C for 1 min and a final step of 30 min at 72 °C. DNA amplification products were checked using agarose gel electrophoresis. Fluorescently labelled DNA fragments were separated on an ABI Prism 3730 (Applied Biosystems, Carlsbad, CA, USA) by capillary electrophoresis and analyzed and sized with Peak Scanner Software 2.0 (Applied Biosystems, Foster City, CA, USA).

The SSR profiles were aligned with regards to their allelic sizes by combining data from all the collections, thanks to the genotyping of numerous reference cultivars shared between collections (Table 3).

Data analysis. Genetic diversity, differentiation and structure analysis

The genetic diversity between cultivars was evaluated with SPAGeDI software[29] analyzing the number of polymorphic markers, total number of alleles, average number of alleles per marker, the observed heterozygosity (H_o), the expected heterozygosity (H_e)[30], Wright's F statistics (F_{is}, F_{st}), allelic frequencies, unique (exclusive) alleles and the number and percentage of rare alleles (frequency lower than 0.05). SSR markers were also evaluated using discrimination power ($PD = 1 - \sum p_i^2$)[31], where p_i represents the frequency of the i^{th} genotype. The analysis of molecular variance AMOVA that estimates the fraction of the genetic variation among and within populations was performed by Genodive version 3.05[32]. Both software packages are able to analyze data files containing diploid and triploid accessions together. In order to determine genetic relationships between the accessions studied, an Neighbor-joining (NJ) cluster analysis of the similarity matrix was performed using genotypes using poppr 2.9.6[33] and ape 5.8[34] packages in RStudio 2023.06.0[35] environment of R 4.4.0[36].

The SSR profiles of the 760 unique genotypes were used to investigate the population structure through the Bayesian model-based clustering procedure of STRUCTURE ver. 2.3.4[20]. As the dataset included diploid and triploid genotypes, the software was run using the recessive alleles approach, encoding the genotypes as detailed by Urrestarazu et al.[37]. The analysis was run for K values ranging from 1 to 10 inferred clusters with 10 independent runs each, and under admixture model with correlated allele frequencies. The tests were done applying a burn-in period of 200,000 and 500,000 iterations for data collection. The best K value was determined through the ΔK method[19] by using Structure Harvester ver. 0.6.1 application[38]. The results suggested that the $K = 2$ groups could be further structured in subgroups, and a second level (nested) structure analysis was performed individually for each $K = 2$ group (G1 and G2). The 760 unique genotypes were assigned to the groups according to their highest membership coefficient, considering a strong affinity when the assigning probability (qI) was ≥ 0.80[3,11,12].

The genetic discrimination in groups and subgroups found using Structure was confirmed by a minimum spanning network (MSN) generated for all genotypes using poppr 2.9.6[33] and ape 5.8[34] packages in RStudio 2023.06.0[35] environment of R 4.4.0[36]. Finally, to validate the genetic variation between the groups and subgroups defined by Structure, AMOVA analyses were performed by Genodive[32]. Descriptive statistics, including variation between groups or subgroups (F_{st}) and diversity within subgroups including H_e, number of polymorphic alleles, and number of private alleles (those present only in one sub-group) were estimated using Genodive[32].

Definition of the Spanish traditional pear core collection (STPCC)

The STPCC should maximize the representativeness of the genetic diversity contained in the Spanish national pear collections, prioritizing the inclusion of genotypes native to Spain, for budgetary and management reasons. For that, in the definition process it was performed first a final STRUCTURE analysis using the population information obtained previously for the genotypes with a membership $qI \geq 0.8$ (PopFlag = 1) whereas no information (PopFlag = 0) was applied to those ones with $qI < 0.8$[15]. After that, there were considered those genotypes that, in that final analysis of the genetic structure, had been assigned to the groups pooling the Spanish references, as well as those genotypes clustered in admixis ($qI < 0.8$) in the groups pooling the international references. However, genotypes corresponding to reference cultivars or only to prospections (i.e., potential accessions, trees geotagged but not yet preserved in a collection) were excluded regardless of their clustering level with a genetic group.

The core collection was defined using a mixed approach[4], in which first an optimized core collection (OC) was selected by maximizing the allele representativeness of the genotypes, and in a second step, the OC was supplemented with genotypes selected by other criteria, such as historical relevance, exceptional pomological features, or allowing full allele recovery.

In total, four sampling strategies were tested to define the OC, using the advanced stochastic local search method (ASLS) implemented in Core Hunter 3.2.1[39]:

- Pure accession-to-nearest-entry (A-NE_p): Optimization (minimization) of the average Bruvo (DB) distance[40] between each genotype of the collection and the nearest entry in the core (A-NE), as defined by Odong et al.[13].
- Mixed accession-to-nearest-entry (A-NE_M): simultaneous optimization of A-NE distance, the Shannon allele diversity index (S_H) and the proportion of alleles recovered (Allele coverage, C_V). In the optimization process, A-NE was assigned a relative weight of 50%, whereas S_H and C_V were assigned a 25% weight each.
- Pure entry-to-nearest-entry (E-NE_p): Optimization (maximization) of the average DB between each entry and the nearest neighboring entry in the core (E-NE), as defined by Odong et al.[13].
- Mixed entry-to-nearest-entry (E-NE_M): similar to A-NE_M but optimizing E-NE instead.

Cores were sampled for 5% (23 genotypes), 10% (45 genotypes) and 15% (68 genotypes) sizes of the original collection. One hundred independent runs per strategy and core size were sampled, and the final OCs were defined with the genotypes most frequently selected in the runs. The adequacy of the OCs was finally evaluated

using as criteria the values of A-NE, E-NE, S_H, C_V and their fidelity in representing the composition and structure of the original set of genotypes, i.e., the distribution of the genotypes in the reconstructed populations and the proportion of diploid and triploid genotypes.

The second step in the construction of the STPCC was carried out in the OC considered to be the most suitable. First, genotypes carrying more than one allele not recovered in the OC were added (optimized recovery). When more than one genotype carried the same missing alleles, priority was given to the genotypes that optimized genetic distances, had the greatest historical relevance, or provided phenotypic characters of interest that were not yet conserved. In the final step, those genotypes carrying an allele that had not been recovered yet were incorporated (full recovery). In the case of more than one genotype carrying the same allele, the procedure was the same as in the previous case.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding authors on reasonable request.

Received: 11 July 2024; Accepted: 23 October 2024
Published online: 04 November 2024

References

1. FAO. Food and Agriculture Organization of the United Nations. *FAOSTAT* (2023). https://www.fao.org/faostat/es/#home
2. MAPA. *Estadísticas Agrarias. Ministerio de Agricultura, Pesca y Alimentación* (2023). https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/encuestas-plantaciones-arboles-frutales/
3. Ferradini, N. et al. Characterization and phylogenetic analysis of ancient Italian landraces of Pear. *Front. Plant. Sci.* **8** (2017).
4. Urrestarazu, J. et al. Integration of expert knowledge in the definition of Swiss pear core collection. *Sci. Rep.* **9** (2019).
5. Montanari, S., Postman, J., Bassil, N. V. & Neale, D. B. Reconstruction of the largest pedigree network for pear cultivars and evaluation of the genetic diversity of the USDA-ARS national pyrus collection. *G3 Genes|Genomes|Genetics* **10**, 3285–3297 (2020).
6. Evans, K., Fernández-Fernández, F. & Govan, K. Harmonising fingerprinting protocols to allow camparisons between germplasm collections—Pyrus. *Acta Hortic.* **814**, 103–106 (2009).
7. Puskás, M. et al. Molecular and flow cytometric evaluation of pear (*Pyrus* L.) genetic resources of the German and Romanian national fruit collections. *Genet. Resour. Crop Evol.* **63**, 1023–1033 (2016).
8. Kocsisné, G. M. et al. Genetic diversity and similarity of pear (*Pyrus communis* L.) cultivars in Central Europe revealed by SSR markers. *Genet. Resour. Crop Evol.* **67**, (2020).
9. Miranda, C., Urrestarazu, J., Santesteban, L. G., Royo, J. B. & Urbina, V. Genetic diversity and structure in a collection of ancient Spanish pear cultivars assessed by microsatellite markers. *J. Am. Soc. Hortic. Sci.* **135**, 428–437 (2010).
10. Ferreira dos Santos, A. R., Ramos-Cabrer, A. M., Díaz-Hernández, M. B. & Pereira-Lorenzo, S. Genetic variability and diversification process in local pear cultivars from northwestern Spain using microsatellites. *Tree Genet. Genomes.* **7**, 1041–1056 (2011).
11. Bielsa, F. J., Irisarri, P., Errea, P. & Pina, A. Genetic diversity and structure of local pear cultivars from mountainous areas from Aragon (Northeastern Spain). *Agronomy.* **11**, 1778 (2021).
12. Pereira-Lorenzo, S. et al. Analysis of the genetic diversity and structure of the Spanish apple genetic resources suggests the existence of an Iberian genepool. *Ann. Appl. Biol.* **171**, 424–440 (2017).
13. Odong, T. L., Jansen, J., van Eeuwijk, F. A. & van Hintum, T. J. L. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theoretical and Applied Genetics* **126**, 289–305. https://doi.org/10.1007/s00122-012-1971-y (2013).
14. Baccichet, I. et al. Genetic and ploidy diversity of pear (*Pyrus* spp.) germplasm of Friuli Venezia Giulia, Italy. *Genet. Resour. Crop Evol.* **67**, 83–96 (2020).
15. Urrestarazu, J., Royo, J. B., Santesteban, L. G. & Miranda, C. Evaluating the influence of the microsatellite marker set on the genetic structure inferred in *Pyrus communis* L. *PLoS One* **10**, (2015).
16. Sehic, J., Garkava-Gustavsson, L., Fernández-Fernández, F. & Nybom, H. Genetic diversity in a collection of European pear (Pyrus communis) cultivars determined with SSR markers chosen by ECPGR. *Sci. Hortic.* **145** (2012).
17. Evans, K. M., Fernández-Fernández, F., Bassil, N., Nyberg, A. & Postman, J. Comparison of accessions from the UK and US National Pear Germplasm collections with s standardized set of microsatellite markers. *Acta Hortic.* 41–46. https://doi.org/10.17660/ActaHortic.2015.1094.2 (2015).
18. Zurn, J. D. et al. A new SSR fingerprinting set and its comparison to existing SSR- and SNP-based genotyping platforms to manage Pyrus germplasm resources. *Tree Genet. Genomes* **16** (2020).
19. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **14** (2005).
20. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
21. Draga, S., Palumbo, F., Miracolo Barbagiovanni, I., Pati, F. & Barcaccia, G. Management of genetic erosion: The (successful) case study of the pear (*Pyrus communis* L.) germplasm of the Lazio region (Italy). *Front. Plant. Sci.* **13** (2023).
22. Nybom, H. & Lācis, G. Recent large-scale genotyping and phenotyping of plant genetic resources of vegetatively propagated crops. *Plants* **10**, 1–30. https://doi.org/10.3390/plants10020415 (2021).
23. Guzzon, F. & Ardenghi, N. M. G. Could taxonomic misnaming threaten the ex situ conservation and the usage of plant genetic resources? *Biodivers. Conserv.* **27**, 1157–1172 (2018).
24. Gasi, F. et al. Assessment of European pear (*Pyrus communis* L.) genetic resources in Bosnia and Herzegovina using microsatellite markers. *Sci. Hortic.* **157** (2013).
25. Brini, W., Mars, M. & Hormaza, J. I. Genetic diversity in local Tunisian pears (*Pyrus communis* L.) studied with SSR markers. *Sci. Hortic.* **115**, 337–341 (2008).
26. Velázquez-Barrera, M. E., Ramos-Cabrer, A. M., Pereira-Lorenzo, S. & Ríos-Mesa, D. J. Genetic Pool of the Cultivated Pear Tree (Pyrus spp.) in the Canary Islands (Spain), Studied Using SSR Molecular Markers. *Agronomy* **12** (2022).
27. Hormaza, J. I. Molecular characterization and similarity relationships among apricot (*Prunus armeniaca* L.) genotypes using simple sequence repeats. *Theor. Appl. Genet.* **104**, 321–328 (2002).
28. Fernández-Fernández, F., Harvey, N. G. & James, C. M. Isolation and characterization of polymorphic microsatellite markers from European pear (*Pyrus communis* L). *Mol. Ecol. Notes* **6** (2006).
29. Hardy, O. J. & Vekemans, X. Spagedi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2** (2002).
30. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**, 5269–5273 (1979).

31. Tessier, C., David, J., This, P., Boursiquot, J. M. & Charrier, A. Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theor. Appl. Genet.* **98**, 171–177 (1999).
32. Meirmans, P. G. & Van Tienderen, P. H. Genotype and genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* **4** (2004).
33. Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. *Poppr*: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281. https://doi.org/10.7717/peerj.281 (2014).
34. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528. https://doi.org/10.1093/bioinformatics/bty633 (2019).
35. RStudio Team & RStudio *RStudio: Integrated Development for R* (PBC, 2020). http://www.rstudio.com/
36. R Core Team. *R: A Language and Enviroment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2024). http://www.r-project.org
37. Urrestarazu, J., Miranda, C., Santesteban, L. G. & Royo, J. B. Genetic diversity and structure of local apple cultivars from northeastern Spain assessed by microsatellite markers. *Tree Genet. Genomes*. **8**, 1163–1180 (2012).
38. Earl, D. A. & VonHoldt, B. M. Structure harvester: A website and program for visualizing structure output and implementing the evanno method. *Conserv. Genet. Resour.* **4** (2012).
39. De Beukelaer, H., Davenport, G. F. & Fack, V. Core Hunter 3: Flexible core subset selection. *BMC Bioinform.* **19**, 203 (2018).
40. Bruvo, R., Michiels, N. K., D'Souza, T. G. & Schulenburg, H. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol. Ecol.* **13**, 2101–2106 (2004).
41. Liu, Q. et al. Genetic diversity and population structure of pear (*Pyrus* spp.) collections revealed by a set of core genome-wide SSR markers. *Tree. Genet. Genomes*. **11**, 128. https://doi.org/10.1007/s11295-015-0953-z (2015).

Author contributions

P.E., A.P., J.U, L.G.S, C.M., S.P., A.R.C, M.B.D., V.U., J.D., E.D., and J.A. contributed to conceptualization, resources, and funding acquisition. A.P., P.I., J.U. and C.M conceived the experiments. J.U., A.R.C., M.E.V, S.C.M., M.L., P.I. and A.P. data collection. P.I., J.U., C.M. and A.P. analyzed the data. P.I., J.U., C.M. and A.P. wrote the original draft. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-77532-1.

Correspondence and requests for materials should be addressed to C.M. or A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Acknowledgements.
This research was funded by Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), grant numbers RTA2015-00052-C02-02 and RFP2015-00015-00-00, and the Gobierno de Aragon-European Social Fund, European Union (Grupo Consolidado A12).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.