Tree Genetics & Genomes (2021) 17:32
https://doi.org/10.1007/s11295-021-01513-y

ORIGINAL ARTICLE

Advanced genebank management of genetic resources of European wild apple, *Malus sylvestris*, using genome-wide SNP array data

Joukje Buiteveld[1] · Herma JJ Koehorst-van Putten[2] · Linda Kodde[2] · Ivo Laros · Giorgio Tumino[2] · Nicholas P. Howard[3] · Eric van de Weg[2] · Marinus J.M. Smulders[2]

Received: 8 December 2020 / Revised: 12 April 2021 / Accepted: 10 May 2021

Abstract

The Netherlands' field genebank collection of European wild apple (*Malus sylvestris*), consisting of 115 accessions, was studied in order to determine whether duplicates and mistakes had been introduced, and to develop a strategy to optimize the planting design of the collection as a seed orchard. We used the apple 20K Infinium single nucleotide polymorphism (SNP) array, developed in *M. domestica*, for the first time for genotyping in *M. sylvestris*. We could readily detect the clonal copies and unexpected duplicates. Thirty-two *M. sylvestris* accessions (29%) showed a close genetic relationship (parent-child, full-sib, or half-sib) to another accession, which reflects the small effective population size of the in situ populations. Traces of introgression from *M. domestica* were only found in 7 individuals. This indicates that pollination preferentially took place among the *M. sylvestris* trees. We conclude that the collection can be considered as mainly pure *M. sylvestris* accessions. The results imply that it should be managed as one unit when used for seed production. A bias in allele frequencies in the seeds may be prevented by not harvesting all accessions with a close genetic relationship to the others in the seed orchard. We discuss the value of using the SNP array to elaborate the *M. sylvestris* genetic resources more in depth, including for phasing the markers in a subset of the accessions, as a first step towards genetic resources management at the level of haplotypes.

Keywords European wild apple, · *Malus sylvestris*, · *Malus domestica*, · 20K Infinium SNP array, · Genotyping, · Ex situ collection, · In situ conservation, · Haplotype

Introduction

European wild apple or crabapple (*Malus sylvestris* (L.) Mill.) is a major contributor to the chloroplast and nuclear genomes of the cultivated apple, *Malus domestica* (Coart et al. 2006; Cornille et al. 2012, 2013, 2014; Nikiforova et al. 2013; Sun et al. 2020). At this moment, *Malus sylvestris* is an endangered tree species in Europe due to destruction of suitable habitats and conversion of open-type forest into mature forest types (Stephan et al. 2003; Coart et al. 2003; Jacques et al. 2009).

Communicated by C. Chen

✉ Joukje Buiteveld
joukje.buiteveld@wur.nl

✉ Marinus J.M. Smulders
rene.smulders@wur.nl

Herma JJ Koehorst-van Putten
herma.koehorst@wur.nl

Linda Kodde
linda.kodde@wur.nl

Ivo Laros
ivo.laros@wur.nl

Giorgio Tumino
giorgio.tumino@wur.nl

Nicholas P. Howard
nick.howard@fresh-forward.nl

Eric van de Weg
eric.vandeweg@wur.nl

[1] Centre for Genetic Resources, the Netherlands, Wageningen University & Research, Wageningen, the Netherlands

[2] Plant Breeding, Wageningen University & Research, Wageningen, the Netherlands

[3] Fresh Forward Breeding & Marketing, Eck en Wiel, the Netherlands

Published online: 03 June 2021

In the Netherlands, *M. sylvestris* is rare, occurring as single individuals or small groups of (6-34) trees scattered across the Eastern part of the Netherlands (Koopman et al. 2007). There are various threats to these wild apple populations. Firstly, the small forest patches in which the wild apple trees occur formerly consisted of coppice wood but as this practice has been discontinued, large trees of other species create low-light conditions, therefore the apple trees grow poorly and flower seldomly. Hence, there is very little regeneration of the population. Secondly, spontaneous apple trees resulting from thrown-away 'cultivated' apples appear in forest edges and may hybridize with the remaining wild apple trees. Because of these threats, ex situ conservation measures were taken in 2002 by establishing a field genebank for *M. sylvestris* in the Netherlands (in vivo ex situ). Therefore, the original trees at different localities within the Netherlands were vegetatively propagated by grafting. For safety and seed orchard purposes, each accession was represented by 2 to 5 copies in the collection. At the same time, these wild apple populations were genetically characterized using selected microsatellite markers (eight microsatellites across linkage groups plus seven additional markers on linkage group 10; Koopman et al. 2007). These markers revealed a weak signal of introgression from cultivated apple in the genotyped accessions, but no direct descendants of cultivated apple varieties were detected. Based on this information, some of the sampled trees were excluded from the collection.

Ex situ collections which are maintained as living trees in a field genebank, are expensive to establish and maintain and therefore they preferably serve multiple purposes. They have to be efficiently and properly managed. The aim of the ex situ field collection of wild apple in the Netherlands is twofold: 1) to assure the long-term conservation of the genetic diversity of wild apple as present in the Netherlands, and 2) to facilitate use of the collection by various users for research, and seed and plant production for re-introduction in nature. Therefore, the collection is also managed as a seed orchard.

Efficient germplasm management consists of preserving the highest diversity with the smallest number of accessions. Since the establishment of the collection, about 15 years ago, several issues have been raised and questions asked about the size and composition that is needed for optimal genetic management of such a small collection of European wild apple. These were addressed by the present study and are introduced below.

First, it is necessary to ensure the genetic integrity of the accessions (Zurn et al. 2020), which may be affected by mislabelling during genebank management (e.g. accessions with the same code but different DNA profiles or accessions with different codes but the same DNA profile). The genetic integrity of an accession can also be compromised by rootstocks overgrowing the scion, a potential threat as the grafting success of *M. sylvestris* was sometimes relatively poor due to variable quality of budwood and poor compatibility. Furthermore, detection and removal of accessions with introgression from *M. domestica* can help to maintain the integrity of the genebank collection.

In addition, there is little knowledge on the degree of genetic relatedness (sibship, family structures) among the accessions and how this affects the genetic diversity of the collection. The majority of the accessions in the collection are from small, relict populations, so family relations among accessions may be expected. Insight in the degree of relatedness among accessions can contribute to reducing genetic redundancy in the collection.

Second, regarding acquisition of new material as an alternative for adding clonal copies of the original trees to the collection could be through collecting seeds in situ and adding their seedlings as new accessions to the collection. However, incorporating seedlings in the collection as an alternative may increase the chance of including hybrid material with cultivated apple. In addition, compared to the original mature trees sampled before, present-day seedlings may be more related to each other due to shared parentage and genetic drift in a population that is decreasing in size and has low and skewed fruit set.

Third, insight into population structure and genetic relatedness among the accessions can be used to optimize the planting design of the collection as a seed orchard (e.g. should genotypes from the original populations be kept separate as conservation seed orchards or should they be mixed as a single seed orchard) or even to plan a mating design with specific crosses between genotypes that are genetically most dissimilar. It may also support the harvest strategy in the seed orchard for maximising the genetic diversity in the seed lot as well in in situ populations to be used for reintroduction.

Therefore, this study had three objectives: 1) to detect the extent of mislabelling, redundancy, and introgression with *M. domestica* in the *M. sylvestris* collection; 2) to assess the level of genetic structure and relatedness (parent-offspring or sibling relationships) in the collection, and 3) to define a genebank management strategy for the use of seed-derived material from the collection.

Compared to microsatellite markers (as used by e.g., Larsen et al. 2006; Koopman et al. 2007; Bitz et al. 2019; Reim et al. 2020), genome-wide SNP data enable a much finer overview of the genetic composition of the accessions, their genetic distance and genetic relationships, and the occurrence of introgression. We decided to address these questions using genomic tools that have been developed in the last 15 years for *M. domestica*. Notably, we used one of the arrays with single nucleotide polymorphism (SNP) markers with a known degree of polymorphisms in cultivated germplasm, namely the 20K SNP Infinium array (Bianco et al. 2014), the dense, genome-wide genetic reference map for most of the 20K SNP array markers (Di Pierro et al. 2016), and the reference

genome sequences for the cultivated apple (Velasco et al. 2010; Daccord et al. 2017; Zhang et al. 2019). Using these resources, duplicates may be identified (Pikunova et al. 2014), parent-offspring relations may be established (Evans et al. 2011; Howard et al. 2017, 2018; Van de Weg et al. 2018; Vanderzande et al. 2019; Skytte AF Sätraa et al. 2020), and possible introgression from cultivated germplasm may be identified by comparison to the genetic data available for a large part of the cultivated germplasm (Peace et al. 2019).

Material and methods

Plant material

The ex situ genebank collection of *M. sylvestris* in Roggebotzand (the Netherlands) is a field collection of trees managed by the State Forest Service, existing as grafted trees (with clonal replicates) from budwood that was collected at six locations in the Netherlands since 2001 (Table 1; Fig. 1; Koopman et al. 2007). Trees with phenotypic characteristics that could point to cultivated apples such as hairiness of the underside of the leaves, large leaves, or fruits with red blush (Maes 2013) were excluded during that sampling.

For the current study, young leaves were collected from 137 trees in the genebank, 14 trees in situ of which one tree was considered to be a putative hybrid (Supplementary Table S1) and, additionally, leaves of 29 one-year old seedlings in a nursery (Zundert). In total 180 samples were collected. The 137 trees sampled in the genebank comprise of 115 accessions and 22 clonal replicates of these accessions (10 accessions with 1 replicate and 6 accessions with 2 replicate trees). The original location of the accessions and trees sampled in situ is given in Table 1. The seedlings were grown from seeds collected from dropped apples picked up at two of the in situ locations (Nijmegen and Veluwe). The young leaves were stored at −80C, freeze dried, and genomic DNA was extracted according to Fulton et al. (1995).

Table 1 Origin of the *M. sylvestris* accessions and in situ sampled trees

Population	Origin*	Location	No. of trees
1	Winterswijk	Ex situ	10
2	Veluwe	Ex situ	10
3	Zeldersche Driessen	Ex situ	29
4	Vijlenerbos	In situ	7**
5	Slenaken	In situ	1
6	Meinweg	In situ	6
7	Nijmegen	Ex situ	23
8	Sint Jansberg	Ex situ	29
9	Drenthe	Ex situ	36

* Names of the populations follow Koopman et al. (2007). The locations are also depicted in Fig. 1. Details per sample in Supplementary Table S1

** One of these samples was putatively introgressed, as it had a morphology intermediate between wild and cultivated apples. It was included as a control

SNP genotyping

All samples were genotyped with the 20K Infinium SNP array (Bianco et al. 2014) at the Fondazione Edmund Mach according to the procedures described by Chagné et al. (2012) and Antanaviciute et al. (2012). The array contains 18019 SNPs (Bianco et al. 2014). Most of the SNPs are located on the genetic map of Di Pierro et al. (2016). This was the first time that the array was used in *M. sylvestris*.

SNP calling for the 180 *M. sylvestris* samples was done as part of a larger set of samples that also included *M. domestica* varieties and selections, using Genome Studio (Illumina Inc., San Diego, CA, USA; http://www.illumina.com) with manually adjusted cluster definitions obtained through an ongoing apple pedigree project (Howard et al. 2018). Within the set of *M. sylvestris* samples, markers with >10% missing values were removed, along with monomorphic markers and markers with a minor allele frequency (MAF) <5%. Genotypes with >5% missing values were also removed. Triploids were detected by examining B allele frequency (BAF) plots in Genome Studio (Chagné et al. 2015).

The first step in the analysis was the identification of duplicate genotypes. Screening for identical genotypes, among all pairs of *M. sylvestris* samples, was done using sets of SNPs for which data were available for both genotypes of the pair.

From the large reference dataset on *M. domestica* produced with the 20K array available to us through the EU-FruitBreedomics project, 33 popular varieties grown in commercial fruit orchards and private gardens in the first half of the 20th century in north-western Europe, or popular varieties for consumption in that period (Supplementary Table S1), were added to the filtered *M. sylvestris* dataset to be able to determine the degree of differentiation between the two species, the occurrence of hybridization and, in case of introgression from cultivated varieties, the regions of introgression and possibly direct parentage of some of the cultivated trees. These 33 varieties were also used as outgroup in the Structure analysis (Pritchard et al. 2000). We were also able to check against SNP data of a set of 65 rootstocks, which were provided by NIAB-EMR, UK (unpublished), in order to check if the genetic integrity of an accession was compromised by rootstocks overgrowing the scion.

Parent-offspring relations

Since apples were picked near in situ trees, which were also genotyped, we expected to find genetic relationships between

Fig. 1 Locations of the *M. sylvestris* populations in the Netherlands

the seedlings and some adult trees. Parentage reconstruction analysis to detect parent-offspring relationships was done for (1) the filtered unique *M. sylvestris genotypes* and (2) this set plus the set of 33 *M. domestica* varieties to enable the identification of feral trees (direct offspring of varieties) and of introgression events (plants with a relationship to *M. domestica* but less than direct offspring) (Cronin et al. 2020).

We used a custom R script to determine Mendelian errors between genotypes in any putative set of parent-child (P-C) relationship (duo) as well as complete parentage (trio or P-P-C). In case of trios, the script counts the number of Mendelian errors observed in the offspring, given the genotypes of two putative parents and under the assumption of parent-child relation. The analysis for duos counts the number of opposite homozygous calls, which are Mendelian errors when assuming a parent-child relationship. We used a low percentage of opposite homozygosity as indication for putative parent-offspring relations, and higher percentages as indicative of more distant relationships or, if a variety was involved, as a sign of introgression of *M. domestica* into *M. sylvestris* (see Vanderzande et al. 2019). In addition, the data were also compared to the dataset from the ongoing apple pedigree reconstruction project described in Howard et al. (2018), which includes thousands of varieties.

Population structure and genetic distance to *M. domestica*

For the comparison with cultivated apples, a dataset was added of SNPs of 33 *M. domestica* varieties that were popular varieties grown in commercial fruit orchards and private gardens in the first half of the twentieth century in north-western Europe, or that were popular varieties for consumption in that period (Supplementary Table 1). A principal component analysis (PCA) was performed by cmdscale (R core team, 2017) of all filtered unique *M. sylvestris* genebank accessions, plus the 33 *M. domestica* varieties.

Structure version 2.3.4 (Pritchard et al. 2000) was also run on the filtered unique genebank accessions together with the *M. domestica* varieties. Posterior probabilities between 0.5 and 0.85 for the largest group assignment in the *M. sylvestris* samples were taken as putative cases of introgression from *M. domestica* to *M. sylvestris*.

We also performed a Structure analysis per chromosome, using the genetic map of Di Pierro et al. (2016) as guidance for

chromosome-specific SNP sets, as introgression from *M. domestica* may vary among chromosomes (Supplementary Tables S3A and S3B). For this analysis, we selected around 120 markers per linkage group (every fifth marker on the genetic map), 2099 markers in total, and looked for q >0.50 on each linkage group separately as tentative *M. domestica* inferred ancestry.

F_{st} values between the mature *M. sylvestris* genotypes and *M. domestica* varieties and among the *M. sylvestris* populations from which the trees were originally sourced, were calculated using Fstat 2.94 (Goudet 1995), using 600 randomly chosen SNPs.

Results

Genotyping and quality filtering

An analysis of missing SNP values per sample produced only two samples with >5% missing values, which were subsequently excluded, leaving 178 *M. sylvestris* samples for further analyses. One was most likely a mixture of two genotypes and thus the results of a DNA sampling error, but the other sample, D180283 from population Drenthe, was triploid. A maximum likelihood shared haplotype test which is part of the apple pedigree reconstruction project (described by Howard et al. 2018) indicated that its parents are most likely also *M. sylvestris*.

Of the 18019 SNPs on the 20K SNP array, 3552 (20%) were excluded because of >10% missing values, and 5348 markers (30%) were excluded because of <5% MAF. An additional 1007 markers (6%) were completely monomorphic in the dataset of 178 *M. sylvestris* samples. This resulted in 8112 SNPs for further analyses (Supplementary Table S2).

Detection of duplicates

In the total dataset 23 sets of duplicates were found. No duplicates were found in the 29 seedlings. One duplication was observed among the 14 trees that were sampled in situ in forest patches, which was not expected. In the remaining 135 samples from the field genebank, 18 genotypes were found twice and 4 genotypes three times. Hence, 23 genotypes were represented by 48 samples. From these 48 samples, the members of 23 pairs of samples were identical for all the marker scores that could be compared (between 7855 and 8053 SNPs, depending on the number of missing values in one of the two samples), for 4 pairs there was 1 mismatch, and for 3 pairs there were 2-3 mismatches between the members of a pair. Non-duplicate pairs showed at least 18% differences in the genotype scores. Duplicate samples could thus be easily and accurately distinguished from non-duplicate samples.

During the sampling, we had included a number of clonal replicates of 16 accessions. All clonal replicates were correctly retrieved except two: the mixed sample (a DNA sampling error) and a different genotype than expected, which may indicate a case of mislabelling in the field genebank. Five sets of duplicates were found in the genebank, which were not expected forehand. Three sets of these accessions may have been originally sampled from an ancient coppice system that was not recognised in the field, and thus these trees will probably be genuine ramets. In the other two cases, the duplicated accessions originated from different localities and these are probably caused by mislabelling during the nursery phase or planting in the field collection.

Removing duplicates and mislabelled trees reduced the number of unique genebank genotypes to 110, plus 13 genotypes sampled in situ. The average heterozygosity of the 123 unique genotypes plus the 29 seedlings was 0.30 while the median was 0.29 (Fig. 2).

Parentage analysis—finding trios to detect parent-offspring and sibling relationships A distinct gap in Mendelian errors when assuming perfect trios, was observed between a small group of trios and the rest (Fig. 3A). Those with 2% or less Mendelian errors were considered "realistic trios" and these are listed in Table 2. The great majority of the <2% errors was due to null alleles (on average ~25 per genotype), indicating that the genotype calling made very few mistakes.

For one accession in the field genebank, both parents could be traced in the genebank itself. The tree and its parents originated from population Zeldersche Driessen, where the distance among them was less than 5 m.

We identified both parents for 25 of the 29 seedlings in the dataset (10 seedlings in the Nijmegen population and 15 in the

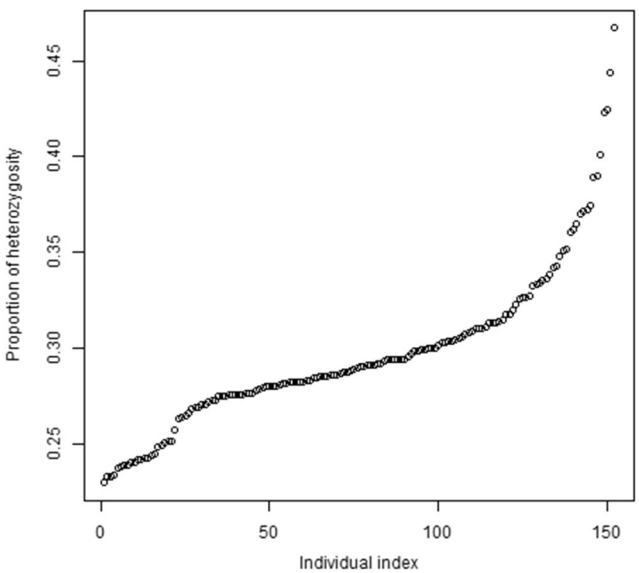

Fig. 2 Distribution of level of heterozygosity of the 152 genotypes (110 accessions, 13 in situ trees and 29 seedlings collected in situ)

Table 2 Parentage analysis of apple seedlings obtained from two in situ populations, based on trios. The two parents are in random order. The five full-sib families found are indicated by numbers

Population	Seedling	Parent 1	Parent 2	Full-sib family group	Number of compared SNP markers	Number of trio Mendelian errors	% trio Mendelian error
Veluwe	D180314	D180163	D180156	1	4585	32	0.70
	D180301	D180163	D180156	1	4586	32	0.70
	D180313	D180163	D180156	1	4588	33	0.72
	D180315	D180163	D180156	1	4574	33	0.72
	D180309	D180157	D180165	2	4101	33	0.80
	D180317	D180157	D180165	2	4097	32	0.78
	D180319	D180157	D180165	2	4105	46	1.12
	D180307	D180157	D180165	2	4104	45	1.10
	D180320	D180157	D180165	2	4100	41	1.00
	D180305	D180165	D180163	3	4293	28	0.65
	D180302	D180165	D180163	3	4293	36	0.84
	D180316	D180165	D180163	3	4293	28	0.65
	D180318	D180156	D180165		4504	32	0.71
	D180308	D180159	D180157		4355	40	0.92
	D180300	D180159	D180161		4768	21	0.44
Nijmegen	D180304	D180238	D180211	4	4297	20	0.47
	D180306	D180238	D180211	4	4301	21	0.49
	D180322	D180238	D180211	4	4299	20	0.47
	D180321	D180213	D180247	5	3854	41	1.06
	D180310	D180213	D180247	5	3852	39	1.01
	D180312	D180213	D180247	5	3851	38	0.99
	D180323	D180213	D180247	5	3849	35	0.91
	D180311	D180213	D180222		3708	34	0.92
	D180303	D180223	D180222		3972	42	1.06
	D180299	D180240	D180222		3941	40	1.01

Veluwe population). In total, 13 different trees were the parents of these 25 seedlings; 6 of the Nijmegen population and 7 of the Veluwe population. The number of offspring per parent ranged from 1 to 9. The trios we identified enabled us to reconstruct full-sibs as well as half-sibs among these offspring. Five full-sib families were present (Table 2), which

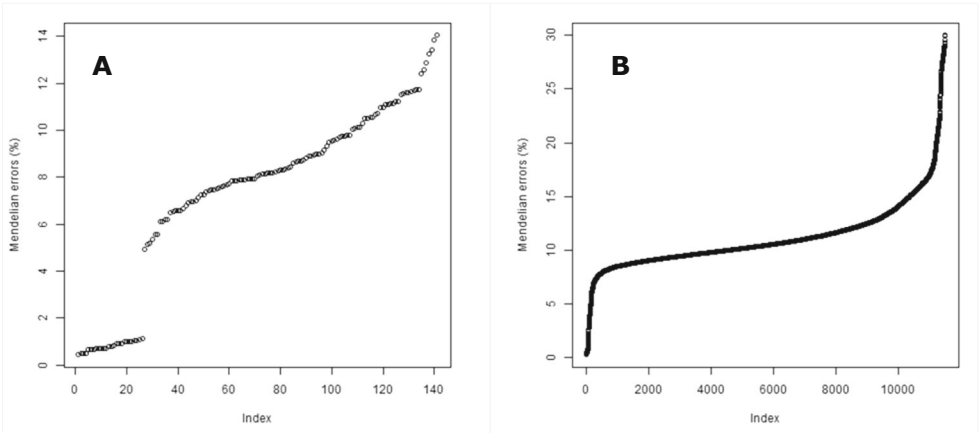

Fig. 3 Distinguishing Mendelian errors when assuming (a) complete parentage P-P-C (trio) and (b) parent-child P-C (duo) Mendelian relationships in possible combinations of the *M. sylvestris* genotypes. Trios and pairs are indexed according to the percentage Mendelian error. In Fig. 3A, 1.7% error is the highest value for true trios that were part of the set. In Fig. 3B, the inflexion point is around 7% error, and only 2.2% (247/11476) of the possible pairs of duos has less than 7% error

may include offspring from reciprocal crosses. The distance between parental trees of a seedling in the natural populations varied between 21 and 969 m in Nijmegen and between 14 and 85 m in Veluwe. The parents of the 19 full-sibs were all located at less than 55 m distance from each other.

Relatedness analysis using duos

There was no clear gap in Mendelian error rate (based on opposite homozygosity) among the tested potential P-C relationships (Fig. 3B), due to which the differentiation between real and random pairs was less evident than with trios. The inflexion point was at around 7% error. This is comparable to the error limit used by Muranty et al. (2020), although they had a much larger set of SNPs. In the absence of a gap, we used the Mendelian error rate in each parent-child (duo) relationship from the 25 trio relationships in the seedlings that we already confirmed (Table 2) as reference values. Among these 50 duo parent-child relationships the Mendelian error rate ranged between 0.27% and 0.90% whereas the Mendelian error rate among 28 full-sib relationships from this material ranged between 1.13% and 3.99% and the 53 relationships between true half-sibs ranged 3.04-6.81%. In contrast, all relationships between seedlings from different populations, which we assume to be unrelated, ranged from 7.32 to 23.63%. This is similar to the 6.26–23.58% error for relationships between the mature trees from different populations that also may be assumed not to be closely related. Therefore, a Mendelian error rate below 0.90% was used for confirming parent-child relationships, below 3.5% for full-sibs and below 7% for half-sibs or third-degree relationships.

We identified three parent-child relationships among the genebank accessions (0.34-0.53% error) (Table 3): two in Zeldersche Driessen and one in Nijmegen. In addition, we found four putative full-sibs (1.47–2.51% error) and 14 putative half-sibs (4.29–5.24% error) through unknown parents. In each case, the pairs were collected from the same population. We also found 12 relationships with 5.38–6.16% error, always from the same population, which may represent second- or third-degree relatedness. Overall, 32 of 110 unique accessions (29%) were closely related to each other in the field genebank at the level of parent-offspring, siblings and half-sibs.

In addition to the duo relationships among accessions in Table 3, accession D180244 was parent of seedling D180298 (0.84% error) in the Nijmegen population. Accession D180211, along with being a parent in three of the trios in Nijmegen, was also a parent of seedling D180324 (0.56% error). The other parents remain unknown.

Thus, in total, we have identified 90% of the parents of the seedlings in the dataset, namely both parents for 25 seedlings, one parent for two seedlings, and neither parent for two seedlings.

In addition, when comparing the data with the dataset from the ongoing apple pedigree reconstruction project described in Howard et al. (2018), which includes thousands of varieties, it was found that D180295 is an offspring of the variety ‘Vlaamsche Schijveling’.

Population structure and genetic distance to *M. domestica*

For a comparison with cultivated apples, the SNP data of 33 *M. domestica* varieties grown in commercial fruit orchards and private gardens in the first half of the twentieth century in north-western Europe, or used for consumption in that period, were added. When the SNP data were filtered for this extended set of samples to optimise the set of SNPs for population differentiation analysis, a set of 11551 SNP markers remained, or close to 3500 more markers (Supplementary Table S2). About 100 additional markers with too many missing values in the *M. sylvestris* accessions were now included, as well as almost 500 markers that were monomorphic in *M. sylvestris*, but the great majority (more than 2900 SNPs) were now included as MAF > 5%, due to the polymorphic scores contributed by the *M. domestica* genotypes. Thus, 29% of the SNPs that are polymorphic in *M. domestica* were monomorphic or had very low MAF in our limited set of *M. sylvestris* genotypes.

The mature *M. sylvestris* populations were not much differentiated. The F_{st} values varied between 0.011 and 0.062 (Table 4). In contrast, the F_{st} value between the groups of *M. sylvestris* and *M. domestica* samples was 0.307.

Within the 123 unique mature *M. sylvestris* genotypes, no obvious structure could be detected in the PCA (Fig. 4A). The genotypes form one group with a few outliers on the first axis. These are unlikely to be rootstock sprouts, as no relationship was detected with 65 rootstocks.

The *M. sylvestris* genotypes and the domestic varieties clearly clustered separately in a PCA, where the first principal component explained 23.2% of the observed variance (Fig. 4B). Seven individuals were in between the *M. sylvestris* and *M. domestica* clusters. One parent of D180202 was identified as ‘Sterappel’ (syn ‘Reinette Rouge Etoilee’ or ‘Early Red Calville’). D180202 was sampled in situ in Vijlenerbos, which is in the southern part of the country, and the tree is located at the edge of the forest and was included as it had an intermediate morphology.

Structure analyses

The Structure analysis of the 123 unique mature *M. sylvestris* and the 33 *M. domestica* samples at K=2 separated the two species well (Fig. 5). The seven intermediate *M. sylvestris* samples in the PCA are also the seven samples with the highest level of admixture (q=0.30–0.85). When we performed the analysis for the 17 chromosomes separately, 23 individuals had $q > 0.50$ for tentative *M. domestica* inferred

Table 3 Close genetic relationships between accessions in the *M. sylvestris* genebank, based on duo analysis. (P-C is parent-child, FS is fullsib, HS is halfsib)

Accession 1	Accession 2	Number of compared SNP markers for the duo comparison	Number of duo Mendelian errors	% duo Mendelian error	Population of both accessions	Inferred relationships
D180171	D180184	4641	16	0.34	Zeldersche Driessen	reliable P-C
D180211	D180259	4479	22	0.49	Nijmegen	reliable P-C
D180175	D180178	4739	25	0.53	Zeldersche Driessen	reliable P-C
D180161	D180163	5039	74	1.47	Veluwe	reliable FS
D180263	D180267	4474	89	1.99	Drenthe	reliable FS
D180173	D180192	4879	120	2.46	Zeldersche Driessen	reliable FS
D180179	D180193	4772	120	2.51	Zeldersche Driessen	reliable FS
D180217	D180230	4426	190	4.29	Nijmegen	reliable HS
D180169	D180194	4409	195	4.42	Zeldersche Driessen	reliable HS
D180171	D180174	4585	205	4.47	Zeldersche Driessen	reliable HS
D180267	D180286	4410	201	4.56	Drenthe	reliable HS
D180276	D180287	4773	228	4.78	Drenthe	reliable HS
D180224	D180259	4423	212	4.79	Nijmegen	reliable HS
D180276	D180284	4532	217	4.79	Drenthe	reliable HS
D180279	D180284	4120	199	4.83	Drenthe	reliable HS
D180183	D180190	4246	213	5.02	Zeldersche Driessen	reliable HS
D180179	D180182	4431	224	5.06	Zeldersche Driessen	reliable HS
D180177	D180183	4258	217	5.1	Zeldersche Driessen	reliable HS
D180171	D180191	4443	230	5.18	Zeldersche Driessen	reliable HS
D180169	D180190	4357	227	5.21	Zeldersche Driessen	reliable HS
D180238	D180251	4124	216	5.24	Nijmegen	reliable HS
D180263	D180286	4199	226	5.38	Drenthe	putative 2nd generation
D180293	D180296	3371	186	5.52	Drenthe	putative 2nd generation
D180167	D180179	4469	258	5.77	Zeldersche Driessen	putative 2nd generation
D180149	D180155	5008	303	6.05	Winterswijk	putative 2nd generation
D180268	D180274	4960	300	6.05	Drenthe	putative 2nd generation
D180147	D180152	4815	292	6.06	Winterswijk	putative 2nd generation
D180275	D180290	4756	290	6.1	Drenthe	putative 2nd generation
D180270	D180274	4890	299	6.11	Drenthe	putative 2nd generation
D180148	D180155	5007	307	6.13	Winterswijk	putative 2nd generation
D180222	D180248	4204	258	6.14	Nijmegen	putative 2nd generation
D180274	D180275	4769	294	6.16	Drenthe	putative 2nd generation
D180285	D180290	5019	309	6.16	Drenthe	putative 2nd generation

introgression, including the seven intermediate genotypes from the PCA and the overall Structure analysis (Supplementary Tables 3A and 3B). In 'Sterappel'-offspring D180202, we found the signal on 12 of the 17 chromosomes. The other six intermediate genotypes had signals on two to six linkage groups. Of the other sixteen individuals, five had a signal on two chromosomes and 11 had a signal on only one chromosome.

Across linkage groups, only five linkage groups showed introgression in five or more individuals, notably linkage groups 5 (in 12 individuals) and 2 (in eight individuals). Introgression on linkage group 17, on which the incompatibility locus is located (Maliepaard et al. 1998), occurred in five individuals. As the set of 33 varieties was small, we did not investigate the chromosomal segments at the haplotype level.

Discussion

In this study of European wild apple genetic resources in the Netherlands, we used the 20K SNP array developed for cultivated apple (*M. domestica*; Di Pierro et al. 2016) for the first time in *M. sylvestris*. A total of 8048 SNPs passed our filtering, with a median level of heterozygosity of 0.29 (55% of the samples between 0.28 and 0.30), which is only slightly lower than the value of 0.34 (range 0.30–0.38 for more than 90% of

Table 4 Pairwise F_{st} values as measure of differentiation among *M. sylvestris* populations and from *M. domestica*

Population	Number of individuals	Winterswijk	Veluwe	Zeldersche Driessen	Vijlenerbos	Slenaken	Meinweg	Nijmegen	Sint Jansberg	Drenthe	33 *M. domestica* varieties
		1	2	3	4	5	6	7	8	9	10
Winterswijk	7	0									
Veluwe	6	0.037	0								
Zeldersche Driessen	26	0.037	0.028	0							
Vijlenerbos	7	0.062	0.048	0.029	0						
Slenaken	1	NA	NA	NA	NA	0					
Meinweg	5	0.058	0.050	0.023	0.011	NA	0				
Nijmegen	20	0.040	0.033	0.011	0.026	NA	0.025	0			
Sint Jansberg	22	0.035	0.029	0.012	0.023	NA	0.023	0.017	0		
Drenthe	29	0.015	0.029	0.034	0.045	NA	0.043	0.035	0.036	0	
M. domestica varieties	33	0.284	0.268	0.290	0.176	NA	0.232	0.272	0.260	0.271	0

the samples) found by Muranty et al. (2020) using a much larger set of SNPs from the apple Axiom 480K SNP array (Bianco et al. 2016) in commercial apple varieties, which have been selected and therefore may have an elevated level of heterozygosity. Among the accessions we found the first triploid *M. sylvestris* accession.

The 20K array was developed for genome-wide coverage in *M. domestica* with some minor input from *M. floribunda* and *M. micromalus,* but not from *M. sylvestris.* This might cause ascertainment bias, even though *M. sylvestris* is a major contributor to the genome of *M. domestica*. The microsatellite data used earlier in *M. sylvestris* (Koopman et al. 2007) and selected based on map positions in *M. domestica* had not shown any ascertainment bias. When the set of SNP data from 123 *M. sylvestris* genotypes was filtered together with SNP data of 33 *M. domestica* varieties made with the same 20K array, this resulted in a set of 11551 SNP that passed the threshold, an increase of about 3500 SNP markers that now were sufficiently polymorphic while they were (almost) monomorphic in *M. sylvestris* (we used a MAF threshold of 5%). It is possible that some of these markers are located in regions in cultivated apple that are not derived from *M. sylvestris* (Sun et al. 2020). However, adding a diverse group (even cultivars) to accessions with a narrow background can result in a comparable increase in the number of polymorphic SNPs. For example, a combination of indica and japonica rice accessions had 8–25% more polymorphic SNPs from the RiceSNP50 array compared to each subspecies separately (Chen et al. 2014). For a thorough analysis of unique *M. sylvestris* regions, a much larger set of *M. sylvestris* accessions should be screened, and it should include populations across its distribution area. These may be compared with the available SNP data of over 4000 *M. domestica* old and modern varieties and breeding lines that have been genotyped and will be analysed as part of an ongoing pedigree reconstruction project (Howard et al. 2018).

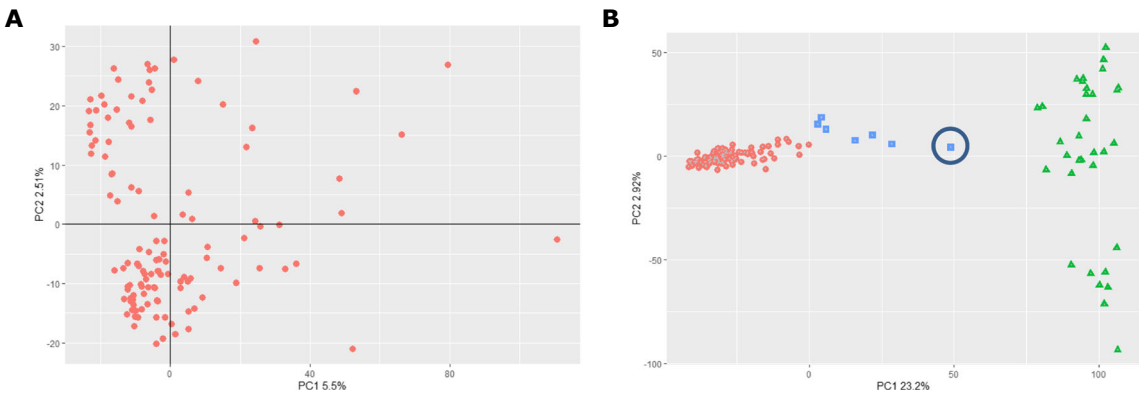

Fig. 4 Principal component analysis of 123 *M. sylvestris* genotypes without (A) and with (B) 33 additional *M. domestica* varieties. PC1 of Fig. 4B represents the difference between the two species; red dots are *M. sylvestris*, green triangles are *M. domestica*. The blue squares are *M. sylvestris* accessions with obvious introgression. D180202, the accession with one *M. domestica* parent, is the encircled blue square closest to *M. domestica*

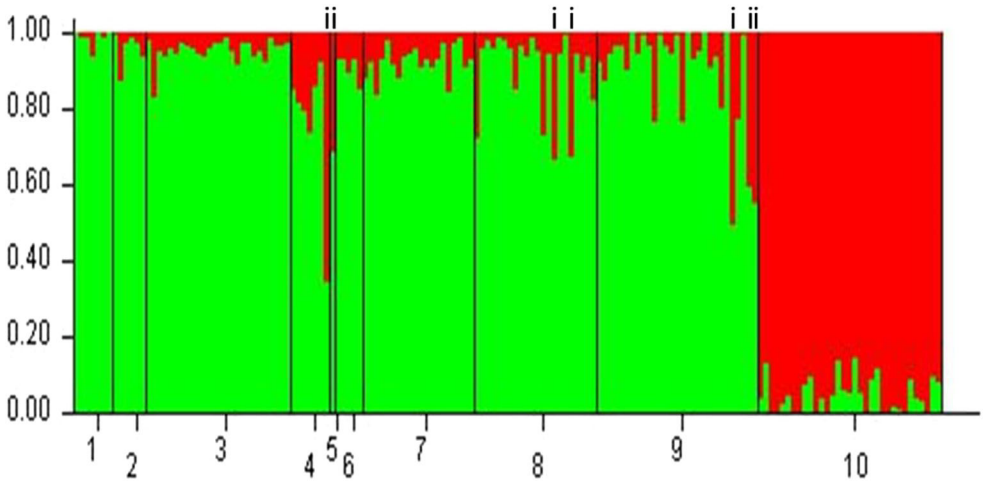

Fig. 5 Structure analysis of the *M. sylvestris* accessions in the field genebank along with the 33 *M. domestica* reference varieties, at K=2 (i.e., assuming two groups). Each individual is represented by a vertical line, the black lines separate the nine populations where the trees originate from (in the order as presented in Table 1 and Supplemental Table 1). Group 10 = the 33 *M. domestica* varieties. The individually estimated posterior probability to each of the two clusters are indicated by colors. The seven intermediate samples do have the highest level of admixture (q=0.30–0.85)

Evaluation of the ex situ management

In situ and ex situ conservation are complementary to each other and should both be applied to adequately conserve the genetic diversity of a species (Volis and Blecher 2010). Naturally, *M. sylvestris* trees are found in open forests and along forest edges (Schnitzler et al. 2014). Their current occurrence in small forest patches in the fragmented landscape in the Netherlands restricts the possibilities for implementing in situ conservation strategies. According to the Euforgen technical guidelines, the establishment of an ex situ collection, which also serves as a seed, is the most suitable and efficient conservation measure to undertake for European wild apple (Stephan et al. 2003). In the Netherlands, the ex situ collection was established as a field genebank in Roggebotzand in 2002 and does not meet these guidelines with respect to number of clones per seed orchard. This collection is used as a seed orchard to produce pure *M. sylvestris* seeds for forest and landscape purposes. Hereto, the collection has been used to establish three separate seed orchards, each containing 31 to 41 accessions from one (Drenthe), two (Nijmegen and Jansberg), or three populations (Winterswijk, Zeldersche Driessen, and Veluwe). The three seed orchards are officially approved as a 'seed source' under the category 'source identified' in the national register of approved basic material.

Errors may occur at many stages during collection of material, the propagation phase at the nursery or during planting in the field. The results of the current study demonstrate that the SNP analysis is an efficient tool to assist collection management in quality control. Overall, the clonal analysis identified the expected duplicate trees except in one case. Although only a small number of ramets were analysed, this confirms that erroneous labelling is rare, and we did not detect any sprouted rootstocks overgrowing the scions. The analysis also revealed some redundancy in the collection. Among the total of 115 accessions in the ex situ collection, 5 duplicates were found, indicating an immediate redundancy of 4.3%. Two of these unexpected duplicate groups were due to mislabeling, but three groups were sampled together, and they may be the result of coppice practices, which was common up to the early twentieth century (Koopman et al. 2007).

Extent of present and past hybridization

The study indicated a high frequency of pure *M. sylvestris* trees in the data set, as only 7 *M. sylvestris* individuals were detected with admixed ancestry in the Structure analysis. These individuals originally came from four different populations: Drenthe (3), Nijmegen (2), Slenaken (1), and Vijlenerbos (1). The individual tree in the forest edge of Vijlenerbos with mixed ancestry was already noticed during sampling as having intermediate morphological characters and was therefore included as a control. For this tree, we determined that one of its parents was the variety Sterappel, syn. Reinette Rouge, first described in 1830 and widely cultivated in Europe in the first half of the twentieth century (Smith 1971).

In the Netherlands, conditions for hybridization with domesticated apple exist. Old maps from the beginning of the nineteenth century show that practically every village in large parts of the Netherlands was surrounded by a belt of orchards. The largest expansion of apple growing took place between 1920 and 1930 and reached its peak around 1950 as the agricultural crisis around 1900 caused many farmers to switch from grain to fruit growing. Nowadays, cultivation of commercial apples is still an important activity in several regions

of the Netherlands, notably in the Betuwe (Gelderland), South Limburg, and Zeeland. This means that the sampling locations Slenaken, Vijlen, Meinweg, as well as Nijmegen and Sint Jansberg are situated in apple-growing regions. On the other hand, we also found introgressed apple trees in Drenthe, which never had extensive commercial apple cultivation, although farms would have had apple trees for home use.

The low frequency of introgression from *M. domestica* to *M. sylvestris* in our study indicates that gene flow between the two species occurred only at low frequency. There are several studies on hybridization between *M. sylvestris* and *M. domestica* [see for a review Cornille et al. 2014 and references therein], which show rates of hybridization varying from 6.8% (Coart et al. 2003) to 36.7% (Cornille et al. 2013). In a comprehensive sampling of *M. sylvestris* across Europe (1889 individuals) Cornille et al. (2015) detected that 23.1% of the genotypes showed signs of introgression from *M. domestica*. A recent study investigating the frequency of hybrids in *M. sylvestris* populations in Northern Britain (Ruhsam et al. 2019) also found a significant amount of hybridization, as an average of 27% of the sampled *M. sylvestris* trees were the product of hybridization. However, they found clear differences in the extent of hybridization among the sampled regions, which differ in geographical features and intensity of land use. The lowest percentage of samples with hybrid origin were found in areas characterized by rugged terrain, high proportion of woodland of natural origin and, likely, a low number of cultivated apple trees in orchards. The highest percentage of hybrid trees collected in the field were found in areas with a long tradition of cultivating apple trees. Cornille et al. (2013) also suggested that environmental factors such as the distance of *M. sylvestris* populations to *M. domestica* orchards, as well as ecological or geographic factors, may affect hybridization rates. Larsen and Kjær (2009) found in their study that pollination distances above 50 m are quite common in natural populations, while pollinations above 300 m also occur, at a low frequency. Feurtey et al. (2017) found pollination distances between *M. domestica* and *M. sylvestris* up to a few km, but only 5% was above 1 km. In our study, the sampling of the trees was primarily focussed on old forest areas, where relict populations of *M. sylvestris* still exist in the Netherlands, in combination with stringent selection criteria for the typical 'wild' phenotypic characteristics. Trees with phenotypic characteristics that could point to cultivated apples such as hairiness of the underside of the leaves, large leaves, or fruits with red blush were excluded from sampling (Maes 2013). Apple trees occurring in the forest edges and along roads were also excluded as these may be derived from thrown-away cores (feral apples). Therefore, our results on percentage of hybrids found in the genebank material may be a realistic estimate of real hybridization rates in the natural populations.

Expanding the genebank collection with seedlings

To expand the ex situ field genebank in order to capture more of the genetic diversity that is still present in *M. sylvestris* populations, but not represented in the field genebank yet, the question was raised if seedling material collected in situ could be used instead of vegetative propagation of the original trees through budwood. Although none of the seedlings showed to be the result of hybridization with cultivated apple, nearly all seedlings were derived from parents that were already present in the genebank. In addition, a lot of effort would be required to keep the contribution of the parents balanced. Therefore, this is not recommended. Stocking in situ populations with seedlings from a seed orchard, is discussed below.

Threats to in situ populations

Our results suggest that the main threat for *M. sylvestris* populations in the Netherlands is not introgression from cultivated apple. In addition to the 6 'intermediate' genotypes, 16 other genotypes showed introgression on 1 or 2 chromosomes. These might be random signals, or the remnants of an old introgression event, for which the age cannot be determined easily. At least, they do not point at recent introgression events. The set of reference varieties we used was small, but it was tuned towards the region of north-west Europe and varieties of the beginning of the twentieth century.

A more pressing threat is the fact that the populations are small and isolated. Close genetic relationships at the level of parent-offspring, siblings and half-sibs were found in the populations, as already indicated by the shared haplotype analysis of Koopman et al. (2007). Overall, 32 (29%) of the accessions had a close relative among the unrelated accessions in the population from which they came. In a self–incompatible species such as *M. sylvestris*, small population size and genetic relatedness may, due to limiting mate availability and genetic drift, lead to further population decline. These results imply that maintaining genetic diversity within these populations may become a problem in further generations, even if the level of heterozygosity of the accessions does not indicate this yet. There is no doubt that the loss of genetic diversity may compromise the ability of the populations to evolve and to cope with environmental changes and eventually reduces their chances of long-term existence (Booy et al. 2000).

Therefore, to maintain the genetic diversity in these populations, the most effective conservation strategy would be to increase the size of these populations by restocking with seed from the ex situ collection. With such small numbers of trees and occurrence of closely related individuals, sourcing from the same location is not effective, nor feasible, so it should be done through seed from other populations.

An essential requirement for in situ conservation of the relict *M. sylvestris* populations is that restocking is accompanied by improvements of the habitat conditions and proper management actions. During Medieval times and all through the nineteenth century, coppicing was an important form of forest management in the Netherlands and other parts of Europe. Oaks, beech, and other tree species were periodically coppiced. This brought light into the forest and a light-demanding species such as *M. sylvestris* benefited from this. However, current forest practice has resulted in relatively dense forest canopies. These conditions make *M. sylvestris* a weak competitor and makes it difficult for it to rejuvenate. Therefore, restocking should be accompanied by management actions such as thinning of surrounding trees, to allow sufficient light to be transmitted through the canopy. This will enable the trees to flower, to form fruit and to regenerate. Besides unfavorable light conditions due to the change in forest practice, current difficulties encountered in regeneration are also the result of browsing; therefore, young saplings also need to be protected from wild game.

Seed orchard design and in situ restocking

The SNP analysis confirmed that little population structure exists in the ex situ collection, but it did identify the existence of close relationships. The results strongly support the notion to combine all accessions in one seed orchard instead of their current partitioning in three different seed orchards according to their geographic origin. In order to facilitate random crosses between the genetically most distant accessions and thereby to maximize the potential genetic variability in the seed lots for restocking in forest areas, it is important to neutralise the effect of closely related genotypes, so that allele frequencies remain constant. The strategy that we propose, based on the genome-wide SNP data, is to focus for the seed orchard on the 71% of the 110 accessions that have no close genetic relationships with each other. Without the need to remove accessions in the collection, we recommend to harvest all accessions but suggest that within the seven pairs with parent-child or full-sib relationships, sharing half their genome, only one accession is used for seed harvest.

For a seed orchard, different scenarios are possible for harvesting in order to restock natural populations, including 1) harvest in bulk, 2) harvest per mother tree, mix equally and distribute half-sibs equally over the populations, and 3) harvest per mother tree and distribute the half-sibs over populations while taking into account the genotypes of the mother trees. We recommend the harvest strategy to follow the size of the population that is desired. Harvesting in bulk, which is less costly, could be appropriate for commercial plantings and restocking of large populations, while for restocking of very small populations the background of the mother trees should be taken into account and the frequency of related individuals (half-sibs) in the seed harvest should be controlled, even though this takes more effort from the seed orchard manager.

Prospects for future research on genetic relationships in *M. sylvestris*

Our current analyses on genetic relationships is based on single SNP data. However, the network of discovered genetic relationships does allow phasing of the SNP markers, allowing us to reconstruct multi-SNP haplotypes. The genotypes that were parent to several offspring, in varying combinations, as shown in the trio and duo analyses in Table 3 combined, reveal an entire network of genetic relationships. From a conservation point of view, identifying separate haplotypes would, in the future, allow management of diversity at the haplotype level. Such an advanced approach would not be meaningful for the small Dutch genebank, but it may be explored at a European level.

Analyses of the length of shared haplotypes are powerful for the identification and reconstruction of distant genetic relationships, in resolving generation order, and demographic history (Koopman et al. 2007; Gusev et al. 2012; Harris and Nielsen 2013; Leitwein et al. 2020). In the Rosaceae, through various projects, approaches, software and workflows are being developed to perform such analyses in an adequate and efficient manner (Howard et al. 2018; Laurens et al. 2018; Peace et al. 2019). These infrastructures are expected to become publicly available soon (Howard and Peace, personal communication). They have been mostly developed in *M. domestica* and sweet cherry (*Prunus avium* L.). Our dataset may serve as an initiation point for such analyses in *M. sylvestris* and lead to a joint international initiative on a much wider germplasm collection.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11295-021-01513-y.

Acknowledgements We thank René van Loon and Marina Fijten for assisting in collecting the samples, and Felicidad Fernández of NIAB-EMR for making available unpublished SNP data on rootstock varieties. The reference SNP data in *M. domestica* came from the FruitBreedomics project no 265582: "Integrated approach for increasing breeding efficiency in fruit tree crops", which was co-funded by the EU seventh Framework Programme. The study was part of the Research Program 'Circular and Climate Neutral' (KB-34-013-003) funded by the Dutch Ministry of Agriculture, Nature and Food Quality.

Code availability Not applicable

Author contribution Conception: JB, MJMS; Sampling: JB; DNA extraction: LK, IL; Data analysis: JB, HK, GT, NH, EW; Analysis and writing: JB, HK, NH, EW, MJMS; Final editing: JB, HK, NH, MJMS. All authors have read and approved the final version.

Declarations

Ethics approval Not applicable

Consent to participate See author contribution

Consent for publication See author contributions

Competing interests The authors declare no competing interests.

Data archiving statement Information and data of genotypes is provided in the Supplementary tables.

References

Antanaviciute L, Fernández-Fernández F, Jansen J, Banchi E, Evans KM, Viola R, Velasco R, Dunwell JM, Troggio M, Sargent DJ (2012) Development of a dense SNP-based linkage map of an apple rootstock progeny using the Malus Infinium whole genome genotyping array. BMC Genomics 13:203. https://doi.org/10.1186/1471-2164-13-203

Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo N, Salvi S, Viola R, Gut I, Laurens F, Chagné D, Velasco R, van de Weg E, Troggio M (2014) Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (*Malus* × *domestica* Borkh). PLoS One 9:e110377. https://doi.org/10.1371/journal.pone.0110377

Bianco L, Cestaro A, Linsmith G, Muranty H, Denance C, Théron A, Poncet C, Micheletti D, Kerschbamer E, Di Pierro EA, Larger S, Pindo M, van de Weg E, Davassi A, Laurens F, Velasco R, Durel C-E, Troggio M (2016) Development and validation of the Axiom® Apple 480K SNP genotyping array. Plant J 86:62–74. https://doi.org/10.1111/tpj.13145

Bitz L, Heinonen M, Moisander S, Tanhuanpää P, Sarvarinne J (2019) Differential allelic richness between *Malus sylvestris* l. and *Malus* × *domestica* borkh. from Finland as a measure of genetic loss. Sustainability 11:6949. https://doi.org/10.3390/su11246949

Booy G, Hendriks RJJ, Smulders MJM, van Groenendael JMBVB (2000) Genetic diversity and the survival of populations. Plant Biol 2:379–395. https://doi.org/10.1055/s-2000-5958

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C (2012) Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. PLoS One 7:e31745. https://doi.org/10.1371/journal.pone.0031745

Chagné D, Kirk C, Whitworth C, Erasmuson S, Bicknell R, Sargent DJ, Kumar S, Troggio M (2015) Polyploid and aneuploid detection in apple using a single nucleotide polymorphism array. Tree Genet Genomes 11:94. https://doi.org/10.1007/s11295-015-0920-8

Chen H, Xie W, He H et al (2014) A high-density SNP genotyping array for rice biology and molecular breeding. Mol Plant 7:541–553. https://doi.org/10.1093/mp/sst135

Coart E, Vekemans X, Smulders MJM, Wagner I, van Huylenbroeck J, van Bockstaele E, Roldán-Ruiz I (2003) Genetic variation in the endangered Wild apple (*Malus sylvestris* (L.) Mill.) in Belgium as revealed by AFLP and microsatellite markers. Consequences for conservation. Mol Ecol 12:845–857. https://doi.org/10.1046/j.1365-294X.2003.01778.x

Coart E, Van Glabeke S, De Loose M, Larsen AS, Roldán-Ruiz I (2006) Chloroplast diversity in the genus Malus: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). Mol Ecol 15:2171–2182. https://doi.org/10.1111/j.1365-294X.2006.02924.x

Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang XG, Tenaillon MI, Giraud T (2012) New Insight into the History of Domesticated Apple: Secondary Contribution of the European Wild Apple to the Genome of Cultivated Varieties. PLoS Genet 8:e1002703. https://doi.org/10.1371/journal.pgen.1002703

Cornille A, Giraud T, Bellard C, Tellier A, Le Cam B, Smulders MJM, Kleinschmit J, Roldan-Ruiz I, Gladieux P (2013) Post-glacial recolonization history of the European crabapple (Malus sylvestris Mill.), a wild contributor to the domesticated apple. Mol Ecol 22:2249–2263. https://doi.org/10.1111/mec.12231

Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P (2014) The domestication and evolutionary ecology of apples. Trends Genet 30:57–65. https://doi.org/10.1016/j.tig.2013.10.002

Cornille A, Feurtey A, Gélin U, Ropars J, Misvanderbrugge K, Gladieux P, Giraud T (2015) Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. Evol Appl 8:373–384. https://doi.org/10.1111/eva.12250

Cronin D, Kron P, Husband BC (2020) The origins and evolutionary history of feral apples in southern Canada. Mol Ecol 29:1776–1790. https://doi.org/10.1111/mec.15277

Daccord N, Celton JM, Linsmith G, Becker C, Choisne N, Schijlen E, van de Geest H, Bianco L, Muranty H, Micheletti D, Di Pierro EA, Gouzy J, Gaillard S, Velasco R, Durel CE, Rees DJG, Laurens F, Lespinasse Y, Quesneville H, Aubourg S, Weigel D, van de Weg E, Troggio M, Bucher E (2017) The apple genome: evolution and methylome dynamics of early fruit development. Nat Genet 49:1099–1106. https://doi.org/10.1038/ng.3886

Di Pierro EA, Gianfranceschi L, Di Guardo M, Koehorst-van Putten HJJ, Kruisselbrink JW, Longhi S, Troggio M, Bianco L, Muranty H, Pagliarani G, Tartarini S, Letschka T, Lozano Luis L, Garkava-Gustavsson L, Micheletti D, Bink MCAM, Voorrips RE, Aziz E, Velasco R, Laurens F, van de Weg WE (2016) A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Horticulture Research 3:16057. https://doi.org/10.1038/hortres.2016.57

Evans KM, Patocchi A, Rezzonico F, Mathis F, Durel CE, Fernández-Fernández F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Gianfranceschi L, Komjanc M, Lateur M, Madduri M, Noordijk Y, van de Weg WE (2011) Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: trueness-to-type of cultivars and their parentages. Mol Breed 28:535–547. https://doi.org/10.1007/s11032-010-9502-5

Feurtey A, Cornille A, Shykoff JA, Snirc A, Giraud T (2017) Crop-to-wild gene flow and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation strategy of the wild apple in Europe. Evol Appl 10:180–188. https://doi.org/10.1111/eva.12441

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Report 13:207–209. https://doi.org/10.1007/BF02670897

Goudet J (1995) FSTAT version 1.2: a computer program to calculate F-statistics. J Hered 86:485–486

Gusev A, Palamara PF, Aponte G, Zhuang Z, Darvasi A, Gregersen P, IPe'er (2012) The Architecture of Long-Range Haplotypes Shared within and across Populations. Mol Biol Evol 29:473–486. https://doi.org/10.1093/molbev/msr133

Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genet 9:e1003521. https://doi.org/10.1371/journal.pgen.1003521

Howard NP, van de Weg E, Bedford DS, Peace CP, Vanderzande S, Clark MD, Li Teh S, Cai L, Luby JJ (2017) Elucidation of the 'Honeycrisp' pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (*Malus*× *domestica*) pedigree-connected SNP data set. Horticulture Research 4:17003. https://doi.org/10.1038/hortres.2017.3

Howard NP, DC Albach, JJ Luby (2018) The identification of apple pedigree information on a large diverse set of apple germplasm and its application in apple breeding using new genetic tools. In: 18th international conference on organic fruit-growing: proceedings of the conference, 19-21 February 2018. Foerdergemeinschaft Oekologischer Obstbau e. V.(FOEKO), Hohenheim, Germany 88–91

Jacques D, Vandermijnsbrugge K, Lemaire S, Antofie A, Lateur M (2009) Natural distribution and variability of the wild apple (*Malus sylvestris*) in Belgium. Belg J Bot 142:39–49

Koopman WJM, Li Y, Coart E, van de Weg WE, Vosman B, Roldán-Ruiz I, Smulders MJM (2007) Linked versus unlinked markers: multilocus microsatellite haplotype sharing as a tool to estimate gene flow and introgression. Mol Ecol 16:243–256. https://doi.org/10.1111/j.1365-294X.2006.03137.x

Larsen AS, Kjær ED (2009) Pollen mediated gene flow in a native population of *Malus sylvestris* and its implications for contemporary gene conservation management. Conserv Genet 10:1637. https://doi.org/10.1007/s10592-008-9713-z

Larsen AS, Asmussen CB, Coart E, Olrik DC, Kjaer ED (2006) Hybridization and genetic variation in Danish populations of European crab apple (*Malus sylvestris*). Tree Genet Genomes 2:86–97. https://doi.org/10.1007/s11295-005-0030-0

Laurens F, Aranzana MJ, Arus P, Bassi D, Bink M, Bonany J, Caprera A, Corelli-Grappadelli L, Costes E, Durel CE, Mauroux JB, Muranty H, Nazzicari N, Pascal T, Patocchi A, Peil A, Quilot-Turion B, Rossini L, Stella A, Troggio M, Velasco R, van de Weg E (2018) An integrated approach for increasing breeding efficiency in apple and peach in Europe. Horticulture Research 5:11. https://doi.org/10.1038/s41438-018-0016-3

Leitwein M, Duranton M, Rougemont Q, Gagnaire P-A, Bernatchez L (2020) Using Haplotype Information for Conservation Genomics. Trends Ecol Evol 35:245–258. https://doi.org/10.1016/j.tree.2019.10.012

Maes B (2013) Inheemse bomen en struiken in Nederland en Vlaanderen. herkenning, verspreiding, geschiedenis en gebruik. Third edition Uitgeverij Boom (in Dutch)

Maliepaard C, Alston FH, Van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, Van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, Den Nijs APM, Periam PN, Rikkerink E, Roche P, C Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Ginkel MV-V, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73. https://doi.org/10.1007/s001220050867

Muranty H, Denancé C, Feugey L, Crépin JL, Barbier Y, Tartarini S, Ordidge M, Troggio M, Lateur M, Nybom H, Paprstein F, Laurens F, Durel CE (2020) Using whole-genome SNP data to reconstruct multi-generation pedigrees involving numerous old apple cultivars. BMC Plant Biol 20:2. https://doi.org/10.1186/s12870-019-2171-6

Nikiforova SV, Cavalieri D, Velasco R, Goremykin V (2013) Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. Mol Biol Evol 30:1751–1760. https://doi.org/10.1093/molbev/mst092

Peace CP, Bianco L, Troggio M, van de Weg E, Howard NP, Cornille A, Durel C-E, Myles S, Migicovsky Z, Schaffer RJ, Costes E, Fazio G, Yamane H, van Nocker S, Gottschalk C, Costa F, Chagné D, Zhang X, Patocchi A, Gardiner SE, Hardner C, Kumar S, Laurens F, Bucher E, Main D, Jung S, Vanderzande S (2019) Apple whole genome sequences: recent advances and new prospects. Horticulture Research 6:59. https://doi.org/10.1038/s41438-019-0141-7

Pikunova A, Madduri M, Sedov E, Noordijk Y, Peil A, Troggio M, Bus VGM, Visser RGF, van de Weg E (2014) 'Schmidt's Antonovka' is identical to 'Common Antonovka', an apple cultivar widely used in Russia in breeding for biotic and abiotic stresses. Tree Genet Genomes 10:261–271. https://doi.org/10.1007/s11295-013-0679-8

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

Reim S, Lochschmidt F, Proft A, Höfer M (2020) Genetic integrity is still maintained in natural populations of the indigenous wild apple species *Malus sylvestris* (Mill.) in Saxony as demonstrated with nuclear SSR and chloroplast DNA markers. Ecology and Evolution. https://doi.org/10.1002/ece3.6818

Ruhsam M, Jessop W, Cornille A, Renny J, Worre R (2019) Crop-to-wild introgression in the European wild apple *Malus sylvestris* in Northern Britain. Forestry 92:85–96. https://doi.org/10.1093/forestry/cpy033

Schnitzler A, Arnold C, Cornille A, Bachmann O, Schnitzler C (2014) Wild European Apple (*Malus sylvestris* (L.) Mill.) Population Dynamics: Insight from Genetics and Ecology in the Rhine Valley. Priorities for a Future Conservation Programme. PLoS One 9:e96596. https://doi.org/10.1371/journal.pone.0096596

Skytte AF Sätraa J, Troggio M, Odilbekov F, Sehic J, Mattisson H, Hjalmarsson I, Ingvarsson PK, Garkava-Gustavsson L (2020) Genetic Status of the Swedish Central collection of heirloom apple cultivars. Sci Hortic 272:109599. https://doi.org/10.1016/j.scienta.2020.109599

Smith MWG (1971) National apple register of the United Kingdom. Pinner Tolcarne Drive, Pinner, Middx. HA5 2DT, and Ministry of Agriculture, Fisheries and Food, Great Britain

Stephan BR, Wagner I, Kleinschmit J (2003) EUFORGEN Technical guidelines for genetic conservation and use for wild apple and pear (*Malus sylvestris* and *Pyrus pyraster*). International Plant Genetic Resources Institute. Rome, Italy http://www.euforgen.org/fileadmin//templates/euforgen.org/upload/Publications/Technical_guidelines/Technical_guidelines_Malus-Pyrus.pdf

Sun X, Jiao C, Schwaninger H et al (2020) Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. Nat Genet. https://doi.org/10.1038/s41588-020-00723-9

Van de Weg E, Di Guardo M, Jänsch N, Socquet-Juglard D, Costa F, Baumgartner I, Broggini GAL, Kellerhals M, Troggio M, Laurens F, Durel CE, Patocchi A (2018) Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. Mol Breed 38:5. https://doi.org/10.1007/s11032-017-0755-0

Vanderzande S, Howard NP, Cai L, Da Silva LC, Antanaviciute L, Bink MCAM, Kruisselbrink JW, Bassil N, Gasic K, Iezzoni A, Van de Weg E, Peace C (2019) High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. PLoS One 14:e0210928. https://doi.org/10.1371/journal.pone.0210928

Velasco R et al (2010) The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat Genet 42:833–839. https://doi.org/10.1038/ng.654

Volis S, Blecher M (2010) Quasi in situ: a bridge between ex situ and in situ conservation of plants. Biodivers Conserv 19:2441–2454. https://doi.org/10.1007/s10531-010-9849-2

Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, Wang D, Tian Y, Yang C, Meng M, Yuan G, Kang G, Wu Y, Wang K, Zhang H, Wang D, Cong P (2019) A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat Commun 10:1494. https://doi.org/10.1038/s41467-019-09518-x

Zurn JD, Nyberg A, Montanari S et al (2020) A new SSR fingerprinting set and its comparison to existing SSR- and SNP-based genotyping platforms to manage Pyrus germplasm resources. Tree Genet Genomes 16:72. https://doi.org/10.1007/s11295-020-01467-7