



## Genetic Status of the Swedish Central collection of heirloom apple cultivars

Jonas Skytte af Sättra<sup>a,\*</sup>, Michela Troggo<sup>b</sup>, Firuz Odilbekov<sup>a</sup>, Jasna Sehic<sup>c</sup>, Helena Mattisson<sup>a</sup>, Inger Hjalmarsson<sup>d</sup>, Pär K. Ingvarsson<sup>e</sup>, Larisa Garkava-Gustavsson<sup>a</sup>

<sup>a</sup> Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden

<sup>b</sup> Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy

<sup>c</sup> Department of Plant Breeding – Balsgård, Swedish University of Agricultural Sciences, Fjälkestadvägen 459, 29194, Kristianstad, Sweden

<sup>d</sup> Department of Landscape Architecture, Planning and Management, Swedish University of Agricultural Sciences, Alnarp, Sweden

<sup>e</sup> Department of Plant Biology, Swedish University of Agricultural Sciences, 750 07, Uppsala, Sweden

### ARTICLE INFO

#### Keywords:

*Malus domestica*  
SNP array  
Genotyping  
Pedigree  
Gene bank  
Local cultivars  
SSR markers

### ABSTRACT

Cultivated apple is one of the most widely grown fruit crops worldwide. With the introduction of modern apple cultivars, from foreign and national breeding programs, the use of local cultivars decreased during the 20th century. In order to minimize genetic erosion and avoid loss of special genotypes, a number of local clonal archives were established across Sweden, with the goal of retaining old and local cultivars. About 220 apple cultivars, appointed for preservation, obtained the status of mandate cultivars. Initially, they were identified based on pomological traits, but prior to the establishment of the Swedish Central Collection they were genotyped with simple sequence repeat (SSR) markers. SSR markers helped to evaluate the status of the preserved material, as well as to find the best possible true-to-type source for propagation, thus guiding the establishment of the Central Collection. Recently, 215 accessions from this collection were genotyped using the 20 K apple Infinium® single nucleotide polymorphism (SNP) array, in order to gain insight into its genetic structure. The initial SSR analysis confirmed the identity of multiple samples with the same cultivar name grown in different locations and identified several mislabeled samples. In the subsequent SNP analysis we identified 30 clonal relationships and a number of parent-offspring relationships, including 18 trios. We also identified five cultivar samples with inconsistent ploidy levels between the SNP and SSR data, in some cases indicating problematic samples preserved in either the Central Collection or some of the local clonal archives. These cultivars need further investigation to ensure their true-to-typeness. Furthermore, the Swedish Central Collection has continued to grow since the onset of this work and now contains additional cultivars, which should be included in future studies. The results indicate that a number of the preserved mandate cultivars holds high potential value for modern breeding programs.

### 1. Introduction

The cultivated apple (*Malus domestica* Borkh.) is one of the world's most widely grown fruit crops. It belongs to the family Rosaceae, is diploid ( $2n = 34x$ ), and has life-history traits characteristic of perennial fruit crops, i.e., it is outcrossing and to a large extent self-incompatible, has a long juvenile period and a long life span, and is often clonally propagated (Gaut et al., 2015). This generally leads to a high degree of heterozygosity and a large proportion of genetic diversity being retained following domestication (Miller and Gross, 2011). These life-history traits also facilitate close genetic relationships between very old cultivars and those emerging from modern breeding programs. In Sweden, a program studying pollination compatibility between

cultivars was established in Alnarp in the 1920s. Seeds from interesting crosses were sown out and resulted in a number of cultivars being released (Nilsson, 1987). The current Swedish apple breeding program was established in Balsgård in the 1940s, and since then has released a number of cultivars adapted to the Scandinavian climate, both for commercial production and home gardening (Nybom, 2019). Some of the cultivars released from these two programs have been appointed for preservation, despite their modern source.

As modern cultivars increased in popularity among commercial growers, a national inventory and collection of local cultivars was initiated in 1979 by Nordiska Genbanken (now NordGen), to prevent loss of genetic resources. This resulted in the establishment of the first local clonal archives spread across the country and a compilation of

\* Corresponding author.

E-mail address: [Jonas.skytte.af.sattra@slu.se](mailto:Jonas.skytte.af.sattra@slu.se) (J. Skytte af Sättra).

<https://doi.org/10.1016/j.scienta.2020.109599>

Received 28 February 2020; Received in revised form 9 June 2020; Accepted 9 July 2020

Available online 17 July 2020

0304-4238/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

pomological descriptions (Nilsson, 1987). Today the responsibility for conservation of heirloom apple cultivars lies with the National Gene Bank of Vegetatively Propagated Crops, similarly to other Scandinavian countries, e.g. Norway (Gasi et al., 2016) and Finland (Heinonen and Bitz, 2019). Starting in 2013, the local clonal archives were complemented by the establishment of the Swedish Central Collection of mandate cultivars at SLU in Alnarp, with the majority of the trees being planted in 2014 and 2015. Mandate cultivars are heirloom cultivars assigned for preservation by the National Gene Bank and are defined as either i) being of local origin according to lore, or ii) originating from Swedish breeding programs, or iii) foreign cultivars with a long documented history of cultivation (Hjalmarsson, 2019; Hjalmarsson, 2020). Thus, the primary criteria for the selection of mandate cultivars was their specific cultural importance and not the inherent genetic diversity.

The Swedish Central Collection of mandate cultivars is expected to include genetic variation relevant for local adaptation, as it contains cultivars with a wide range of climate hardiness (Nilsson, 1987). This collection has the potential to become a rich source of favorable alleles for the Swedish apple breeding program. Current knowledge concerning relationships among the Swedish mandate cultivars is limited, and mainly based on historical pomological records (Nilsson, 1987), pedigree records in the breeding programs, and previous studies using SSR markers (Garkava-Gustavsson et al., 2008; Urrestarazu et al., 2016). Recently, the pedigrees of some cultivars in the collection have been elucidated using a large set of SNP markers (Muranty et al., 2020). Clarifying unknown pedigrees greatly enhances the prospects for germplasm utilization, by increasing the accuracy of pedigree-based quantitative trait loci analyses (Howard et al., 2017) and reducing the risk of unwanted inbreeding in future crosses.

Genotyping of gene bank material is crucial in order to ensure that genetic variation is maintained, but also to make gene bank material a useful resource for breeding. In the past, SSR markers have been widely used in germplasm studies of apple at the national, regional, and European level (Garkava-Gustavsson et al., 2013, 2008; Gasi et al., 2016; Heinonen and Bitz, 2019; Larsen et al., 2017; Lassois et al., 2016; Marconi et al., 2018; Patocchi et al., 2009; Routson et al., 2009; Urrestarazu et al., 2016; van Treuren et al., 2010). More recently, genotyping-by-sequencing (GBS), SNP arrays, and diversity array technology (DArT) markers have been used to unravel apple pedigrees (Larsen et al., 2018; Muranty et al., 2020; Ordidge et al., 2018; Vanderzande et al., 2017). The IRSC 8 K SNP array (Chagné et al., 2012) available for apple has been useful in elucidating pedigrees (Howard et al., 2017), while the 20 K Infinium® SNP array (Bianco et al., 2014) that followed is currently being used in an ongoing apple pedigree project (Howard et al., 2018). The latter has also been used in the construction of a high-density genetic map (Di Pierro et al., 2016) and QTL-discovery for several traits (Allard et al., 2016; Di Guardo et al., 2017; van de Weg et al., 2018). Thus, genotyping a collection with the 20 K SNP array not only provides robust assessments of pedigree inference and population structure, but also improves the usefulness of the collection by allowing integration with larger germplasms and facilitates its implementation in a breeding-by-design concept (Peleman and Van Der Voort, 2003) that utilizes published information on the genetic control of relevant traits.

In the present study, we first demonstrate how SSR markers were used to assist the establishment of the Swedish Central Collection of mandate cultivars by identifying the best possible true-to-type source of graft wood for the trees intended for preservation in the Central Collection. Recently, we used the 20 K apple SNP array to characterize 215 apple accessions from the Swedish Central Collection. The genotypic data was used to describe the status of the collection by identifying clonal relationships, putative 1st degree relationships, and possible parent-offspring relationships, and by assessing population structure.

## 2. Material and methods

### 2.1. Plant material

For SSR analysis, young leaves from a total of 204 apple cultivars were collected in May – early June (2007–2009) and stored at  $-80^{\circ}\text{C}$  until use. Samples were collected from different clonal archives, the stock collection of the Swedish Elite Plant Station (EPS, holding true-to-type and virus-free genotypes in stock orchards in Fjälkestad), the research- and breeding-oriented germplasm collection at the Swedish University of Agricultural Sciences (SLU-Balsgård) and the Institute of Horticulture in Kiev (Ukraine). If several trees of the same cultivar were present in different collections, all those were sampled and analyzed resulting in 340 samples (Supplementary File 1).

For SNP genotyping, leaf samples were collected from 215 accessions of *Malus domestica* Borkh. in the Swedish Central Collection of mandate cultivars (Alnarp, Sweden), including two rootstocks: ‘A2’ on which the collection is grafted and ‘Bemali’ (Supplementary File 1). Young leaves from growing shoots of a single tree were collected at the end of May 2018, freeze-dried, and stored at  $-80^{\circ}\text{C}$ . Extraction of total genomic DNA was performed using the DNeasy Plant Mini Kit (Qiagen) following the standard protocol. The quality and concentrations of the DNA samples were measured by spectrometry (Nanodrop, Thermo Scientific). The sample names used in this work are the accepted names according to the Swedish Utility and Cultivated Plants Database (SKUD) in January 2020, whenever such were available.

### 2.2. DNA-amplification and SSR analysis

Genomic DNA was isolated from approximately 100 mg of frozen leaves, using the Qiagen DNeasy™ Plant Mini Kit (Qiagen) following the manufacturer’s protocol. Twelve primer pairs, developed by Liebhard et al. (2002) were used for SSR-analysis: CH01h02, CH02c06, CH02c09, CH02c11, CH02d08, CH04c06, CH04c06, CH04e05, COL, CH02b10, CH04e02 and CH02c02b. The forward primers were fluorescently labeled at the 5′-end using either FAM or HEX.

Amplifications were performed for each primer pair separately as described in (Garkava-Gustavsson et al., 2008). PCR-products were pseudo-multiplexed before being analyzed on a 3730 DNA analyzer (Applied Biosystems). Size of the amplified products was calculated based on an internal standard (500ROXTM Size Standard (Applied Biosystems)) and evaluated using GeneMapper® software v. 3.0 (Applied Biosystems).

### 2.3. Flow cytometry

The flow cytometry analyses were conducted on the samples showing more than two alleles at the Plant Cytometry Services (JG Schijndel) in the Netherlands. The fresh leaves were chopped in ice-cold buffer with DAPI, and flow cytometry was performed on a CyFlow ML (Partec GmbH, Münster, Germany) using *Lactuca sativa* as internal standard, and ‘Cox Orange’ as a ploidy level standard.

### 2.4. Analysis of SSR-data

The SSR fragments were scored in terms of loci and alleles, and thus the allelic composition of each sample was determined. Then Simple Matching coefficient (Sokal and Michener, 1958) was calculated for all the analyzed samples with NTSYS-pc statistical package v. 2.2 (Rohlf, 2005) to produce a similarity matrix and detect accessions with identical allelic profiles (a few mis-matches in one allele were allowed for the genotypes to be considered as identical). Allelic compositions were examined to verify the identity of cultivars with their known sports, to assure that accessions with the same name had identical profile, as well as to identify problematic cases (e.g., obvious identification mistakes and mislabeling).

Next, all triploid cultivars and a tetraploid cultivar were removed from the data set. The genetic diversity of the remaining 150 unique profiles of diploid cultivars was estimated using GenAlex 6.503 (Peakall and Smouse, 2006, 2012). Genetic diversity parameters analyzed included: number of different alleles (Na), number of effective alleles (Ne), Shannon’s information index (I), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity (uHe) and inbreeding coefficient (F).

### 2.5. SNP genotyping

The SNP genotyping was carried out using the 20 K Infinium® SNP array (Illumina Inc.) (Bianco et al., 2014). Samples comprising 200 ng of genomic DNA from each of the 215 accessions were analyzed, following the standard Illumina protocol detailed in Chagné et al. (2012). Intensity data were analyzed using the workflow described by Vanderzande et al. (2019), with one major deviation. In Vanderzande et al. (2019) the cluster definitions were obtained from ASSiST (Di Guardo et al., 2015), while in this project we used manually adjusted cluster definitions and a subset of SNPs obtained through an ongoing apple pedigree project (Howard et al., 2018) consisting of 10,368 SNPs. The cluster definitions were loaded to Genome Studio (GS), v 2.0 (Illumina Inc.) was implemented, and the SNP genotype calls obtained were used for further analyses. Sample quality was assessed using the overall call rate and the histogram distributions of the B-allele frequency and the R-parameter.

### 2.6. SNP-data analysis

Putative ploidy levels were determined using frequency plots according to Chagné et al. (2015), as described by Vanderzande et al. (2019). Duplicate individuals were identified using PLINK 1.9 (Chang et al., 2015), using the Pi\_Hat parameter as an approximation of the global pairwise identity-by-descent (IBD) of individuals. PLINK utilizes a method-of-moments approach to estimate proportion of alleles shared IBD (PI\_HAT,  $\hat{\pi}$ ):

$$\hat{\pi} = P(Z = 1)2 + P(Z = 2),$$

where  $P(Z = 1)$  and  $P(Z = 2)$  is the global estimate of the proportion of loci with 1 and 2 alleles in common, respectively (Purcell et al., 2007).

Duplicates as well as known sports were used as quality controls, to verify the accuracy of the SNP genotyping procedure. In the case of unexpected duplicates, these were further examined for morphology of individual trees, records from sample preparations, historical descriptions and SSR data from local archives. All duplicates were removed from the dataset, which was then analyzed with PLINK to identify putative 1st degree relationships, i.e., pairs of individuals with coefficients of relationship equal to 0.5, such as full siblings and parent-offspring relationships.

The available pedigree records were verified using the R-script for counting Mendelian inconsistent errors, as described by Vanderzande et al. (2019). The pedigree records were adjusted for any errors detected and the script was re-run to identify possible parent-offspring combinations, using thresholds of 30 and 60 Mendelian inconsistent errors for parent-offspring (P-O) and parent-parent-offspring (P-P-O) combinations, respectively. To illustrate the results, network plots were generated using the iGraph package (Csardi and Nepusz, 2006) in RStudio version 1.1.463.

Assessment of population structure was performed following removal of synonymous samples, but without pruning for linkage disequilibrium (LD) or minor allele frequencies (MAF). Principal component analysis (PCA) was carried out in PLINK and plotted in Excel 2016 (Microsoft). Statistical analysis was carried out using a model-based clustering as implemented in the software STRUCTURE v2.3.4 (Falush et al., 2007, 2003; Pritchard et al., 2000). STRUCTURE was run with Parallel GNU (Tange, 2011) while assuming admixture, correlated

allele frequencies between population, and implementing the LD model, using the genetic positions from Di Pierro et al. (2016). Initially, a single-replicate analysis was run assuming 1–10 subpopulations (K), in order to determine suitable settings. Five independent replicates were then performed, with a burn-in of 15,000 and a run length of 50,000 for each K from 1 to 5. Next, the most probable number of subpopulations was estimated using the method from Evanno et al. (2005), as implemented in Structure Harvester (Earl and von Holdt, 2012).

## 3. Results

### 3.1. Ploidy level and duplicates

All primer pairs produced clear and consistent DNA-profiles, which were used in the analysis. In all loci except two (CH04c06 L1 and CH02c02b) three alleles were found in some genotypes. Some of these were previously known triploid cultivars, e.g. ‘Gravensteiner’. However, some cultivars were not previously described as being triploid and their ploidy level was confirmed by flow cytometry analyses, e.g. ‘Sköldinge’. Analysis of the similarity matrix confirmed a number of samples with identical SSR profiles, as expected as multiple samples of the same cultivar had been collected from different locations. Others concerned samples with an SSR based indication of a need for further investigation, i.e. were 9 cases of samples with the same name having distinct SSR profiles when collected from different collections. In addition, there were 49 cases of multiple cultivars sharing the same SSR profile, indicating potentially synonymous cultivars or mislabeling (Supplemental File 1).

### 3.2. Genetic diversity

The number of alleles varied from 2 (CH04c06 L1) to 20 (CH02c06) (Table 1). Loci with a high number of alleles also produced high values for  $H_o$  and  $H_e$ , and vice versa. For each locus,  $H_o$  values (from 0.30 to 0.89) were quite similar to  $H_e$  values (from 0.34 to 0.88). The highest  $H_o$  was found at locus CH02b10, while the lowest at locus CH02c02b. Locus CH02b10 yielded the highest  $H_e$  value, while locus CH04c06 L1 had the lowest. All loci had F coefficients below 0.1, except CH02c02b which deviated with a value of 0.49.

### 3.3. SNP quality check and ploidy

Of the 10,368 SNPs considered here, 406 showed significant deviations from Hardy-Weinberg equilibrium (‘ChiTest100’ column in the

**Table 1**  
Genetic diversity in terms of number of different alleles (Na), number of effective alleles (Ne), Shannon’s information index (I), observed heterozygosity (Ho) expected heterozygosity (He), unbiased expected heterozygosity (uHe) and inbreeding coefficient (F) from the 150 unique diploid genotypes analyzed.

Locus	Na	Ne	I	Ho	He	UHe	F
CH01h02	12	3.06	1.44	0.67	0.67	0.68	0.00
CH02c06	20	7.77	2.34	0.86	0.87	0.87	0.01
CH02c09	11	7.58	2.13	0.85	0.87	0.87	0.02
CH02c11	14	6.97	2.15	0.83	0.86	0.86	0.03
CH02d08	15	4.41	1.85	0.79	0.77	0.78	-0.02
CH04c06 L1	2	1.51	0.52	0.37	0.34	0.34	-0.11
CH04c06 L2	12	6.45	2.07	0.86	0.84	0.85	-0.02
CH04e05	13	3.84	1.76	0.75	0.74	0.74	-0.01
COL	11	5.38	1.85	0.75	0.81	0.82	0.08
CH02b10	17	8.57	2.37	0.89	0.88	0.89	-0.01
CH04e02	11	4.29	1.68	0.76	0.77	0.77	0.01
CH02c02b	8	2.43	1.16	0.30	0.59	0.59	0.49
Mean	12	5.00	1.78	0.72	0.75	0.75	0.04
Total	146	62					

**Table 2**

Sample ploidy level inconsistencies.  $C_{SSR}$  indicates ploidy level expected from simple sequence repeat (SSR) data, validated by flow cytometry and  $C_{SNP}$  indicates ploidy level based on single nucleotide polymorphism (SNP) data.

Sample	$C_{SSR}$	$C_{SNP}$
'Bosebo Sötäpple'	3n	2n
'Cellini'	2n	3n
'Holländare'	3n	2n
'Skälbyäpple'	3n	2n
Pomme de Cannelle'	3n	2n

GS SNP table) in this collection. Six of these SNPs were not called, 35 displayed an excess of homozygosity and the remaining 365 SNPs displayed an excess of heterozygosity. Since the fraction of markers deviating from Hardy-Weinberg equilibrium was low ( $< 5\%$ ), these markers were included in the downstream analyses.

Using cluster definitions and a robust subset of SNPs from an ongoing apple pedigree project (Howard et al., 2018), all samples had an overall call rate above 0.88 (except for the tetraploid 'Alfa68', which had a call rate of 0.80) and p50 GC values between 0.75 and 0.81. All samples had B-allele frequency (distinct heterozygous and homozygous peaks) and R-parameter histograms (low values on the x-axis) in GS, indicating good sample quality without contamination (distinct heterozygous and homozygous peaks), as described by Vanderzande et al. (2019). Next, we inspected the B-allele frequency plots generated using the method described in Chagné et al. (2015) which, together with the B-allele frequency histograms in GS, allowed us to identify sample ploidy. A few samples had ploidy levels differing from those revealed based on SSR markers (Table 2). Interestingly, 'Tegnér'säpple' showed a B-allele frequency plot indicating a translocation/aneuploidy at Linkage Group 15 (Supplementary File 2)

### 3.4. Duplicates in the Swedish Central Collection

Duplicate samples were defined as having global pairwise estimates of IBD  $> 0.999$ , as there were no sample pairs with IBD values between 0.99 and 0.7. Analysis for the presence of duplicates resulted in 185 unique samples out of the total 215 apple accessions tested. In the case of triploid synonymous samples, their B-allele frequencies were investigated to ensure that they had the same heterozygous state (i.e., either 'AAB' or 'ABB'). All synonymous triploid samples had B-allele frequencies with correlations above 0.99, indicating that they were identical. All expected duplicates were identified, both duplicate samples and color sports. However, we identified a number of unexpected duplicates and these were compared against the SSR data obtained from the local clonal archives. Two cases, 'Rödluvan'/'Borsdorfer' and 'A2'/'Björnegårdsäpple', were clear examples of mislabeling or propagation mistakes, and the true-to-type individuals were revealed based on examination of morphological characters. In the case of 'Cox's Orange Pippin'/'Skälbyäpple', the IBD coefficients to other cultivars were investigated and were found to be consistent with the pedigree of 'Cox's Orange Pippin'. Duplicate samples were excluded from further analysis (Table 3).

### 3.5. 1st degree relationships, P-O relationships, and possible crosses

Having removed duplicates, the remaining 185 samples were re-analyzed for pair-wise IBD. A threshold for putative 1st degree relationships was set at IBD = 0.484 (0.5–1/64), guided by the IBD values of samples with known relationships. A draft pedigree based on previous SSR studies (Urrestarazu et al., 2016) and historical records (Svensson, 2005) concerning the cultivars was then constructed and analyzed for pedigree errors using an R-script as described by Vanderzande et al. (2019). Erroneous parents (3 cases; Supplementary File 1) were removed and the pedigree was re-analyzed to search for

possible parent-offspring combinations. Nineteen possible crosses were identified, all of which appeared plausible according to the historical record of cultivar ages (Nilsson, 1987; Svensson, 2005) (Table 4). Most of the offspring were local Swedish cultivars, with two exceptions, 'Transparente Blanche' and 'Gestreifter Wintercalvill'. These are thought to be of Baltic/Russian and German origin, respectively, and both have possible parents that are either very old or of unknown origin (Nilsson, 1987).

### 3.6. Possible parents

A number of possible parent-offspring relationships within the collection were also identified, heavily biased towards a few cultivars. The cultivars with the highest number of possible parent-offspring relationships were 'Gimmersta' (29), 'Grågylling' (19), 'Vitygyling' (16), 'Rosenhäger' (7), and 'Klockhammarsäpple' (6). For most cultivars, the number of possible parent-offspring relationships was in line with the number of 1st degree relationships estimated from IBD values (Supplementary File 1).

### 3.7. Network

Based on the possible parent-parent-offspring combinations identified, parent-offspring relations, and estimated 1st degree relationships, a single network was identified for the collection, comprising 115 cultivars interconnected through 1st degree relationships (Fig. 1). Forty-nine of the 185 unique cultivars were found to have no 1st degree relationships within the collection. A number of putative 1st degree relationships that are not supported by the test for Mendelian inconsistent errors were also highlighted. Notably, 'Sködinge' and 'Alnarps Favorit' had a low number of possible parent-offspring relationships, while having a relatively high number of putative 1st degree relationships.

### 3.8. Genetic structure

In PCA, the first three principal components explained 9.0, 6.1, and 4.8 % of the genetic variation in the SNP set in the collection, with no evident population structure (Fig. 2). Accordingly, the STRUCTURE analysis yielded the highest  $L(K)$  value for  $K = 1$  and  $\Delta K$  dropped to 5 already at  $K = 2$ , indicating that the Swedish Central Collection most likely consists of one sub-population (Supplementary File 1).

## 4. Discussion

### 4.1. SSR-analysis of local clonal archives

The SSR analysis of the local clonal archives elucidated several accession that were either duplicates or mislabelings. Some of the interesting cases involved the different accessions traditionally referred to as 'Antonovka'. 'Antonovka' from Bergianska trädgården has the same SSR profile as 'Antonovka' from Julita and 'Antonovka Pamtorutka' from Balsgård. The latter was found to be identical to 'Antonovka Polotora Funtovaja' by Urrestarazu et al. (2016), which has been the preferred name in subsequent studies (Muranty et al., 2020). In contrast, 'Antonovka' from Ekebyhov is identical to 'Antonovka Kamenichka' at Balsgård and a sample of 'Antonovka' imported from the Institute of Horticulture, Kiev (Ukraine). According to the same studies as above, the preferred name for this cultivar has been 'Antonovka Obyknovennaja' (Muranty et al., 2020; Urrestarazu et al., 2016). 'Antonovka Polotora Funtovaja' and 'Antonovka Obyknovennaja' was recently found to have a parent-offspring relationship, which was directed based on historical data with the latter suggested as being one of the possible parents (Muranty et al., 2020). For the establishment of the Central Collection 'Antonovka' from Bergianska trädgården was used as graft wood source, i.e. 'Antonovka Polotora Funtovaja'. Thus, the collection

**Table 3**

Synonymous sample names based on identity-by-decent (IBD) analysis. Samples 1-4: Synonymous sample names. SSR: Simple sequence repeat data from the samples collected from the local archives confirming that the samples are synonymous (Same) or indicating that the sampled tree was not true to type (Diff). TTT: Phenotypic characters of the sampled trees in situ indicated that one of the sampled trees was not true-to-type (Diff).

Sample 1 (used name)	Sample 2	Sample 3	Sample 4	SSR	TTT
'Grågylling'	'Grågylling, Skokloster'	'Rödgylling'	'Hedemoraäpple'	n/a	–
'Ullströmsäpple'	'Gislaved'	'Hagbyberga'		Same	–
'Åkerö'	'Åkerö från Gripsholm'	'Åkerö' (mother tree)		Same	–
'Cox's Orange Pippin'	'Skälbyäpple'			Diff	–
'A2'	'Björnegårdsäpple'			Same	Diff
'Aroma'	'Amorosa'			n/a	–
'Alexander'	'Annero'			Same	–
'Aspa'	'Eneroths Klaräpple'			n/a	–
'Brita Horn'	'Ringstad'			n/a	–
'Gravensteiner'	'Roter Gravensteiner'			Same	–
'Gyllenkroks Astrakan'	'Gyllenkroks Astrakan' (mother tree)			n/a	–
'Suislepp'	'Gul Höstkalvill'			Same	–
'Kristinaäpple'	'Nådig Frun'			Same	–
'Melonenäppel'	'Rött Melonäpple'			Same	–
'Mio'	'Röd Mio'			Same	–
'Sävstaholm'	'Sävstaholm' (mother tree)	'P.J. Bergius'		n/a	–
'Vitgylling'	'Ramsta'			Same	–
'Borsdorffer'	'Rödluvan'			Diff	Diff
'Särsö'	'Spässerud'			Same	–
'Sjöholmsäpple'	'Ås Baron'	'Stor Klar Astrakan'		Same	–
'Calville Blanche d'Été'	'Calville Rouge d'Été'			n/a	–
'Stenkyrke'	'Stenkyrke' (mother tree)			n/a	–
'Transparente Blanche'	'Tistads Vaxgylling'			Same	–
'Kavlås'	'Wickstrands Favorit'			Same	–

**Table 4**

Possible crosses identified using the R-script described in Vanderzande et al. (2019). Offspring, Parent 1 & Parent 2: Possible parent-parent-offspring trios identified, maternal/paternal parent cannot be identified. Age Offspring, Age Parent 1 & Age Parent 2: Recorded approximate first appearance of offspring and parent cultivars, from Nilsson (1987). n/a = unknown origin, old = cultivars of unclear origin, but older than ~1850, \*arrival in Sweden, \*\* from Svensson (2005).

Offspring	Parent 1	Parent 2	Age Offspring	Age Parent 1	Age Parent2
'Adelhill'	'Alexander'	'Oranie'	n/a	1800s*	n/a
'Åkerö'	'Grågylling'	'Vitgylling'	1700s	old	old
'Bläsingeäpple'	'Stenkyrke'	'Summer Pearmain'	1900s	1800s	old
'Edsele'	'Rosenhäger'	'Vitgylling'	1900s	n/a	old
'Fredriksdalsäpple'	'Ullströmsäpple'	'Klockhammarsäpple'	1900s	1800s	1800s
'Frösvidal'	'Hampus'	'Grågylling'	1800s	1700s	old
'Grönsö'	'Grågylling'	'Vitgylling'	1800s	n/a	old
'Hanaskogsäpple'	'Oranie'	'Vitgylling'	1800s	n/a	old
'Hornsberg'	'Hampus'	'Gimmersta'	1800s	1700s	1800s
'Mälardalens Vitgylling'	'Klockhammarsäpple'	'Vitgylling'	n/a	1800s	old
'Mölnbacka Skogsäpple'	'Holmserudsäpple'	'Rosenhäger'	1800s	n/a	n/a
'Råby Rubin'	'Sävstaholm'	'Maglemer'	1900s**	1800s	1700s
'Sandbergs Röda'	'Grågylling'	'Vitgylling'	1900s	old	old
'Sjöholmsäpple'	'Gimmersta'	'Vitgylling'	1800s	1800s	old
'Sparreholm'	'Gimmersta'	'Rosenhäger'	1800s	1800s	n/a
'Stäringe Karin'	'Gimmersta'	'Grågylling'	1800s	1800s	old
'Gestreifter Wintercalvill'	'Holmserudsäpple'	'Rosenhäger'	n/a	n/a	n/a
'Transparente Blanche'	'Aspa'	'Vitgylling'	1800s *	1800s	old

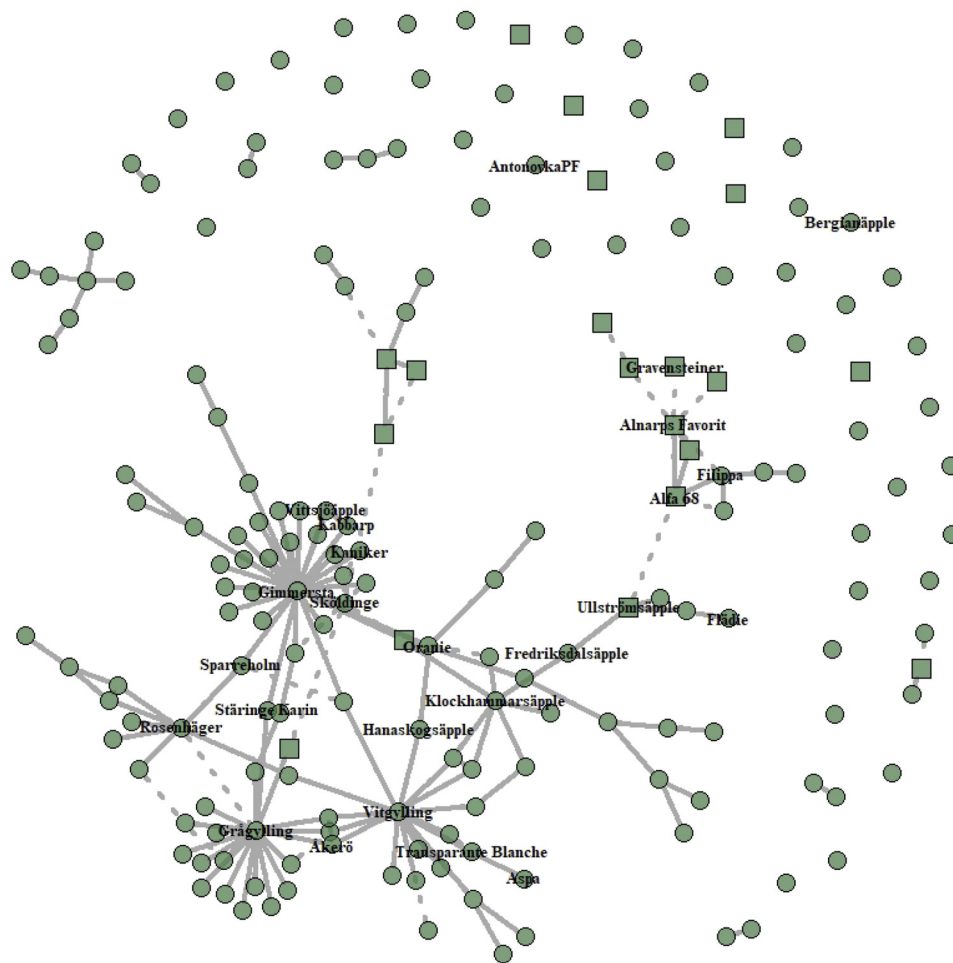
might benefit from including 'Antonovka Obyknovennaja' as well as it is likely to have a long history of cultivation in Sweden, even though there has been some confusion in past as both of these cultivars have commonly been designated as simply 'Antonovka'.

There were also some clear cases of mislabelling, e.g. 'Eva-Lotta' in one of the local archives appears to be identical to 'Alice' and 'Hornsberg' in one of the local archives was identical to 'Hanaskogsäpple'. Revealing of such cases provided valuable support for the choice of graft wood source for the establishment of the central Swedish Central Collection. Notably, some accessions from the breeding collection in Balsgård appears to be not true to type, including 'Domö Favorit', 'Menigasker' and 'Silva' which all have different SSR profiles than accessions with the same names at the local clonal archives. Though beyond the scope of this study, this might be a relevant note, as several international studies including old Swedish cultivars have been based on samples obtained from the Balsgård collection

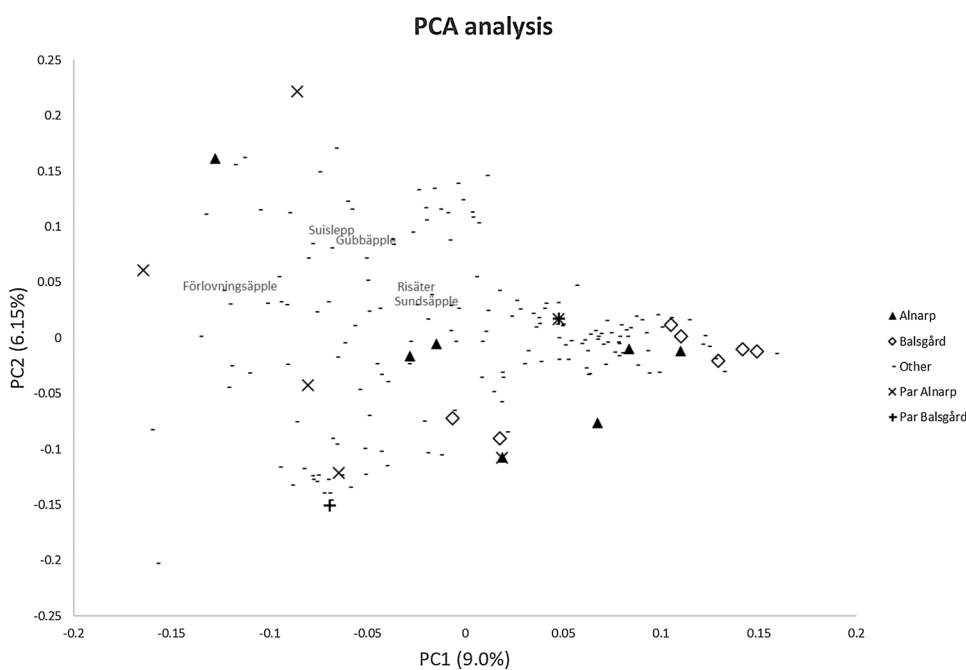
(Muranty et al., 2020; Urrestarazu et al., 2017, 2016).

Of the 12 SSR-markers utilized here, only 5 are common with those used by Urrestarazu et al. (2016) and their usefulness for comparison against a larger germplasm is therefore limited. Nevertheless, SSR profiles have been made available to the MUNQ-database and MUNQ genotype codes have been assigned as previously described, whenever possible (Muranty et al., 2020; Urrestarazu et al., 2016).

In the analysis for genetic diversity only diploid accessions were considered, as triploids are considerably less efficient as parents. Thus, diversity conserved in polyploids has limited impact as an actual resource. Considering the diversity parameters analyzed, the local archives seem to retain a high degree of genetic diversity. The mean observed heterozygosity (Table 1) is close to the expected heterozygosity, and quite high, which is expected considering the life history traits of the species and being in line with comparable studies (Gaut et al., 2015; Larsen et al., 2017; Lassois et al., 2016; Marconi et al.,



**Fig. 1.** Network plot of the Swedish Central Collection of heirloom apple cultivars. Diploid accessions are indicated as circles, triploid and tetraploid accessions as squares. Solid lines indicate possible parent-offspring and parent-parent-offspring relationships, including confirmed pedigree records; dotted lines indicate putative 1st degree relationship. Some of the cultivars mentioned in the discussion are designated with names.



**Fig. 2.** A two-dimensional plot of principal component analysis (PCA) of the Swedish Central Collection of heirloom apple cultivars. Cultivars originating from the current Swedish breeding program in Balsgård (diamond), from the program in Alnarp (triangle) and their respective parents (star and cross) are indicated. Cultivars which can be considered as suitable parents for future breeding efforts, for northern Sweden, are labeled with names.

2018).

It should be noted, however, that the main criterion for the establishment of the Swedish Central Collection was not genetic diversity *per se*, but rather cultural diversity. Consequently, several duplicates with different names that were revealed by analysis with SSR markers were nonetheless considered as distinct mandate cultivars. While being genetically identical, duplicate accessions might still be sports, which can be difficult to distinguish without detailed investigations of morphology in common gardens. Another aspect is the richness in folk lore accompanying many of the old cultivars, which cases can be just as important as the genetic properties of an old cultivar.

#### 4.2. Central collection - ploidy and synonyms

From the SNP markers analysis, we found the Swedish Central Collection of mandate cultivars to be comprised of 190 diploid, 24 triploid, and 1 tetraploid accessions, corresponding to 89, 10, and 1 % of all unique samples. Slightly higher frequencies of triploids have recently been reported for Belgian and Danish collections (15 % and 19 %, respectively). Similarly, we found the Swedish central collection to contain 86 % unique accessions, which is in line with the 79 % and 85 % reported for the Belgian and Danish collections, respectively (Larsen et al., 2018; Vanderzande et al., 2017).

When comparing ploidy levels with those obtained during SSR analysis of the same cultivars, but from the local clonal archives, we found five inconsistent cases. For one such case, 'Skälbyäpple', the sample analyzed from the local clonal archive was triploid and genetically identical with 'Gravensteiner', and accordingly a different source for budwood was used, which had not been analyzed prior to propagation. In the Central Collection, 'Skälbyäpple' was found to be morphologically similar to 'Cox's Orange Pippin' (Nilsson, 1987), while also having estimated global pairwise-IBD to other accessions that are in line with the pedigree of 'Cox's Orange Pippin'. According to pomological description 'Skälbyäpple' appears to be quite similar to 'Cox's Orange Pippin', thus it appears likely that 'Skälbyäpple' is synonymous with 'Cox's Orange Pippin'. The other accessions with inconsistent ploidy levels identified here were 'Bosebo Sötäpple', 'Cellini', 'Holländare' and 'lPomme de Cannelle' (Table 2). Similar to 'Skälbyäpple', 'lPomme de Cannelle' was propagated from source not analyzed by SSR markers, thus explaining the inconsistency. However, none of the other three samples appears to be synonymous with any other accession in the Central Collection, and this inconsistency thus requires further investigation.

The consistency between the use of SNP array data and flow cytometry to determine ploidy levels has previously been established (Chagné et al., 2015), making incorrect calling of ploidy levels unlikely. Consequently, the accessions from both the local clonal archives and the central collection should be analyzed using the same genetic markers, to clarify these inconsistencies. For example, 'Cellini' has previously been described as diploid (Urrestarazu et al., 2016), making it likely that the 'Cellini' accession present in the central collection is not true-to-type. Interestingly, although the 'Cellini' sample might not be true-to-type, it could still be a valuable part of the collection, as it does not appear to be synonymous with anything else in the collection. Similarly, 'Bergianäpple' has historically been sold as 'Antonovka' (Hjalmarsson, 2019), but it is now recognized as a separate cultivar with, according to our data, no close relationship to 'Antonovka' (IBD = 0.14). Regarding the synonymous samples, for nine out of the 24 pairs, SSR data were lacking for at least one of the accessions. This illustrates the usefulness of the current thorough examination of the collection, and complements the SSR-based information acquired during the establishment of the gene bank. Some of the duplicate pairs/trios reported here are either duplicates from different sources, e.g., 'Gyllenkroks Astrakan' (EPS) and 'Gyllenkroks Astrakan' (mother tree), or well-known sports such as 'Sävstaholm' and 'P.J. Bergius'. The remaining duplicate pairs would benefit from further genetic

investigation of the material in the local clonal archives and from morphological studies, in order to establish whether they are synonymous accessions, mixed-up trees, or sports. For example, 'Hedemoraäpple' appears to be synonymous to 'Grågylling' in the Swedish Central Collection and these two cultivars are described as having rather similar morphology, but 'Hedemoraäpple' is not described as having the twisted stem that is characteristic of 'Grågylling' (Nilsson, 1987). The accessions of 'Rödgylling' from the local clonal archives also had the same SSR profiles as 'Hedemoraäpple' and some accessions labeled as 'Grågylling'. According to lore, 'Hedemoraäpple' is said to originate from a seedling, and thus it would be interesting to investigate whether it is a mix-up or a stem morphology sport, or whether the 'Hedemora' tree investigated previously was not old enough to express the twisted stem trait clearly. Another example is 'Björnegårdsäpple', which is described as "particularly valuable", with a taste and aroma reminiscent of both pineapple and wild strawberries (Nilsson, 1987). In both the Swedish Central Collection and the local clonal archive analyzed, it seems to have been lost and overtaken by shoots from the rootstock. Thus, a new source has to be found for this cultivar through inventory of sources closer to the supposed origin, preferably the mother tree.

#### 4.3. 1st degree relationships

##### 4.3.1. Parent-offspring test

Using a parent-offspring test for Mendelian inconsistent errors, we were able to reject some stipulated relationships. For example, 'Kaniker' was speculated by the early 1900s' pomologist Carl Dahl to be the parent of 'Oranie' (Dahl, 1943), but according to our current SNP marker analysis these cultivars clearly did not share a parent-offspring relationship. On the other hand, the data suggest that they might both be the offspring of 'Gimmersta', and their pairwise IBD estimates indicate either a half-sibling or a grandparent-offspring relationship (0.28). According to tradition 'Flädie' is supposed to originate from a seed of a 'Gravensteiner' apple (Dahl, 1943), which is not supported by our data. However, these cultivars appear to share a cryptic relationship (IBD = 0.37). Interestingly, Larsen et al. (2017) also reported a complete lack of parent-offspring relationships involving 'Gravensteiner', despite some suggested by pomologies. Yet another example that has been the cause of speculation is 'Sparreholm'. It was discovered as a pippin by the gardener L. G. Hedlund, who believed it to be the offspring of 'Rosenhäger' and an astrakhan-type cultivar (Dahl, 1943). According to our results, this speculation was correct, as it was identified as the possible offspring of 'Rosenhäger' and 'Gimmersta', the latter being of astrakhan type. More recently, Garkava-Gustavsson et al. (2008) speculated that 'Oranie' could be one of the parents of 'Hanskogsäpple', based on SSR profiles, which our data seem to support. Similarly, 'Kabbarp' and 'Vittsjöäpple' were found to be genetically similar according to our SSR-analysis. Accordingly, our SNP data suggest that they are likely full siblings (IBD = 0.6), with 'Gimmersta' as one of the parents, as supported by the Mendelian inconsistent errors test.

##### 4.3.2. Possible parent-parent-offspring combinations

From the analysis of the Swedish Central Collection we identified several possible parent-parent-offspring combinations, all of which seem plausible based on historical records of the cultivar ages. All accessions involved were diploid, except for the triploid 'Ullströmsäpple' which is a possible parent of 'Fredriksdalsäpple' (Table 4, Supplementary File 1). Triploid apples perform much better as pollen donors than as mothers (Sato et al., 2007), so it is reasonable to assume that 'Ullströmsäpple' is the father. However, the procedure used here calls triploids only as heterozygous, regardless of their heterozygous state, and some proportion of true heterozygous SNPs can be expected not to be called, as they will have signals outside the heterozygous clusters. Thus, further clarification of putative parent-offspring relationships would clearly benefit from automated SNP calling of triploid samples, allowing more robust integration of triploids with diploid datasets.

#### 4.3.3. Possible parent-offspring combinations

The most common possible parents were all old cultivars of unknown origin, e.g., ‘Gimmersta’ (29 P-Os), ‘Grågylling’ (19 P-Os), ‘Vitgylling’ (16 P-Os), ‘Rosenhäger’ (7 P-Os) and ‘Klockhammarsäpple’ (6 P-Os). Thus ‘Gimmersta’, a sparsely spread local cultivar considered to be related to an astrakhan-type cultivar (Nilsson, 1987), surprisingly emerged as having the highest number of possible parent-offspring relationships in the collection. At the same time, a very old cultivar which was previously among the most common grown in central Sweden, ‘White Astrachan’ (Nilsson, 1987), is currently missing from the Central Collection. In our SSR analysis of local archives, ‘Gimmersta’ was found to be identical to samples of ‘White Astrachan’, as well as the not true-to-type accession of ‘Arvidsäpple’ preserved in the breeding collection in Balsgård. The accession of ‘Arvidsäpple’ preserved at Balsgård has previously been found to be synonymous to ‘Astrakan Bilyi’ (Urrestarazu et al., 2016). In our analysis of SNP data from the Swedish central collection, ‘Stäringe Karin’ was found to be the offspring of ‘Gimmersta’ x ‘Grågylling’, whereas Muranty et al. (2020) found the same cultivar to be the offspring of ‘White Astrachan’ x ‘Grågylling’. While ‘Gimmersta’ is described as having red fruit over color (Nilsson, 1987) a plausible explanation is that ‘Gimmersta’ is in fact a sport of ‘White Astrachan’, which would then explain the rather large number of possible parent-offspring relationships.

#### 4.3.4. Residual putative 1st degree relationships

Two cultivars, ‘Sköldinge’ and ‘Alnarps Favorit’, had far more putative 1st degree relationships than possible parent-offspring relationships. Many of the 1st degree relationships suggested by IBD values, but not passing the test threshold for Mendelian inconsistent errors, involved triploid cultivars or occurred within the network of 1st degree relationships. Due to their higher heterozygosities, triploids are expected to have higher IBD using the method-of-moments technique. Despite this, a few recorded heteroploid relationships were confirmed by the test for Mendelian inconsistent errors (‘Alfa 68’ - ‘Filippa’ and ‘Alnarps Favorit’ - ‘Alfa 68’).

Remaining couples (those not involving ‘Sköldinge’ and ‘Alnarps Favorit’) with putative 1st degree relationships, but failing the test threshold for Mendelian inconsistent errors, generally had quite high IBD values (up to 0.6). However, they also had Z0 parameters above 0.1 (Supplementary File 1), illustrating that the method-of-moments technique implemented in PLINK generally inflates IBD values (Morrison, 2013). The structure of the Swedish Central Collection is similar to the Danish collection, in that an equal proportion of the collection belongs to a single network of parent-offspring relationships and has a similar distribution of 1st degree relationships (Larsen et al., 2017).

#### 4.4. Genetic structure

The three first principal components explained a greater proportion of the genetic variation (9.0 and 6.15 % for the first two dimensions) than in two recent studies with similar numbers of cultivars that used GBS (2.87 and 2.47 %) or the 8 K apple SNP array (5.9 and 4.5 %) for genotyping (Larsen et al., 2018; Vanderzande et al., 2017). No population structure was evident from the PCA analysis, even though the dataset was not pruned for markers in LD. As reviewed by Miller and Gross (2011), perennial fruit crops generally exhibit limited population structure due to their life history traits such as clonal propagation, outcrossing, and long juvenile periods. Thus, as the Swedish Central Collection was established with the aim of preserving mainly local cultivars, a lack of population structure is to be expected. Comparable studies have found other European collections to consist of different subpopulations. However, they also contain modern cultivars (Lassois et al., 2016; Marconi et al., 2018; Pereira-Lorenzo et al., 2017; Vanderzande et al., 2017) and wild material (Larsen et al., 2018), in addition to local cultivars. An investigation of Norwegian *ex situ* collections found separate clustering between cultivars from southern

countries and traditional Scandinavian or very winter-hardy cultivars (Gasi et al., 2016). Similar clustering patterns have been proposed elsewhere (Garkava-Gustavsson et al., 2013; Urrestarazu et al., 2016), but as the Swedish Central Collection is well curated, containing only mandate cultivars, little clustering is expected. From a genetic resource point of view, a lack of population structure indicates that allelic variants underlying e.g., adaptation to a northern climate are present in an otherwise diverse genetic background, making it a promising germplasm to identify those underlying regions. Based on our PCA analysis, the cultivars released from the Swedish National breeding program at Balsgård seem to encompass only a small fraction of the genetic variation available in the mandate cultivar collection along the two first principal components. Furthermore, a majority of the parents of those cultivars are not represented in the collection (Fig. 2). The Swedish central collection can therefore be expected to contain considerable variation that can invigorate future breeding efforts. Based on the clustering of cultivars released from the Balsgård breeding program, diploid cultivars with coordinates below zero along principal component one and above zero along principal component two might serve to broaden the gene pool (upper-left part of Fig. 2). Considering cultivation in northern Sweden as a potential breeding goal for the Swedish breeding program, the cultivars ‘Gubbäpple’, ‘Risäter’ and ‘Suislepp’ as well as ‘Förlovningsäpple’ and ‘Sundsäpple’ might be interesting for crosses. ‘Gubbäpple’ is described as having big fruits similar to ‘Gravensteiner’, with a balanced sweetness and acidity as well as crispy flesh and an aroma reminiscent of almonds. ‘Risäter’ is described primarily suited as a cooking apple but is hardy, resistant to diseases and having broad branch angles. ‘Suislepp’ is a cultivar of Estonian origin, which is considered to combine excellent hardiness and a good, aromatic, taste. ‘Förlovningsäpple’ and ‘Sundsäpple’ are very briefly described in the literature, but originate from the northern part of Sweden and can thus be expected to be adapted to the climate (Hjalmarsson, 2019; Nilsson, 1987; Svensson, 2005). To our knowledge, these are not the recorded parents of any of the cultivars released from the Swedish breeding program in Balsgård, which has mainly been directed at commercial apple production in Sweden’s most southern province Scania, but this does not exclude the possibility that these cultivars have been used in unsuccessful crosses.

## 5. Conclusions

SSR markers were useful for first screening of the material preserved in local clonal archives and the collection at SLU-Balsgård and the Swedish Elite Plant Station in order to pinpoint putative identification mistakes as well as providing guidance in choosing the best possible source of graft wood for the Swedish Central Collection. The later has subsequently been analyzed using a subset of 10,368 SNPs from the 20 K SNP array. Thus, we obtained robust information from characterization of the Swedish central collection of mandate cultivars. A few, previously unknown, clonal relationships were identified and these might merit revision of cultivar names. It was also possible to identify putative ploidy levels and possible parent-offspring relationships. Notably, a few cases were identified where the SNP data were in conflict with data obtained from the presumed graft-wood source using SSR markers. In order to guarantee the true-to-typeness and genetic integrity of the Swedish Central Collection and its congruence with local clonal archives, these conflicting cases should be investigated further to identify the sources of the errors. Furthermore, the collection has continued to grow since the onset of this study, and thus molecular-based efforts should continue. Despite considerable phenotypic variation in traits related to hardiness and diverse geographic origins, the collection exhibits limited genetic structure. This information is expected to help improving the future development of the Swedish Central Collection and to facilitate use of the material in breeding programs.



## CRedit authorship contribution statement

**Jonas Skytte af Sättra:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Michela Troggio:** Methodology, Validation, Investigation, Resources, Writing - review & editing. **Firuz Odilbekov:** Methodology, Validation, Investigation, Resources, Data curation, Writing - review & editing. **Jasna Sehic:** Methodology, Validation, Investigation. **Helena Mattisson:** Methodology, Validation, Investigation. **Inger Hjalmarsson:** Conceptualization, Validation, Resources, Writing - original draft, Writing - review & editing. **Pär K. Ingvarsson:** Methodology, Validation, Writing - review & editing, Supervision. **Larisa Garkava-Gustavsson:** Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We are very grateful to the curators of the Swedish local clonal archives and Head of the Swedish Elite Plant Station for providing plant material as well as MSc Anna Zborowska for assistance with compiling the SSR data and accompanying cultivar information. We would also like to thank Caroline Denancé for providing MUNQ genotype codes. Financial support was obtained from the Swedish Board of Agriculture and the GroGrund “Accelerated and Cost efficient cultivar development through genome informed apple breeding”. The data analysis benefitted from the Udda server, Department of Plant Breeding, at the LTV Faculty, SLU.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.scienta.2020.109599>.

## References

- Allard, A., Bink, M.C., Martinez, S., Kelner, J.J., Legave, J.M., di Guardo, M., Di Pierro, E.A., Laurens, F., van de Weg, E.W., Costes, E., 2016. Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. *J. Exp. Bot.* 67, 2875–2888. <https://doi.org/10.1093/jxb/erw130>.
- Bianco, L., Cestaro, A., Sargent, D.J., Banchi, E., Derdak, S., Di Guardo, M., Salvi, S., Jansen, J., Viola, R., Gut, I., Laurens, F., Chagné, D., Velasco, R., van de Weg, E., Troggio, M., Chagne, 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>.
- Chagné, D., Crowhurst, R.N., Troggio, M., Davey, M.W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R.P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J.D., Iezzoni, A., Mockler, T., Wilhelm, L., Van de Weg, E., Gardiner, S.E., 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., Erasmussen, S., Bicknell, R., Sargent, D.J., Kumar, S., Troggio, M., 2015. Polyploid and aneuploid detection in apple using a single nucleotide polymorphism array. *Tree Genet. Genomes* 11. <https://doi.org/10.1007/s11295-015-0920-8>.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Csardi, G., Nepusz, T., 2006. The Igraph Software Package for Complex Network Research.
- Dahl, C.G., 1943. *Pomologi : Beskrivningar Över De Viktigaste I Sverige Odlade Fruktarterna. D. 1 Äpplen, 2nd ed. Bonnier, Stockholm.*
- Di Guardo, M., Micheletti, D., Bianco, L., Koehorst-van Putten, H.J., Longhi, S., Costa, F., Aranzana, M.J., Velasco, R., Arus, P., Troggio, M., van de Weg, E.W., 2015. ASSIST: an automatic SNP scoring tool for in- and outbreeding species. *Bioinformatics* 31, 3873–3874. <https://doi.org/10.1093/bioinformatics/btv446>.
- Di Guardo, M., Bink, M.C.A.M., Guerra, W., Letschka, T., Lozano, L., Busatto, N., Poles, L., Tadiello, A., Bianco, L., Visser, R.G.F., van de Weg, E., Costa, F., 2017. Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. *J. Exp. Bot.* 68, 1451–1466. <https://doi.org/10.1093/jxb/erx017>.
- Di Pierro, E.A., Gianfranceschi, L., Di Guardo, M., Koehorst-van Putten, H.J., Krusselbrink, J.W., Longhi, S., Troggio, M., Bianco, L., Muranty, H., Pagliarani, G., Tartarini, S., Letschka, T., Lozano Luis, L., Garkava-Gustavsson, L., Micheletti, D., Bink, M.C., Voorrips, R.E., Aziz, E., Velasco, R., Laurens, F., van de Weg, W.E., 2016. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. *Hortic. Res.* 3, 16057. <https://doi.org/10.1038/hortres.2016.57>.
- Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. <https://doi.org/10.1080/13235818.1996.10673675>.
- Falush, D., Stephens, M., Pritchard, J.K., 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7, 574–578. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>.
- Garkava-Gustavsson, L., Kolodinska Brantestam, A., Sehic, J., Nybom, H., 2008. Molecular characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis. *Hereditas* 145, 99–112. <https://doi.org/10.1111/j.0018-0661.2008.02042.x>.
- Garkava-Gustavsson, L., Mujaju, C., Sehic, J., Zborowska, A., Backes, G.M., Hietaranta, T., Antonius, K., 2013. Genetic diversity in Swedish and Finnish heirloom apple cultivars revealed with SSR markers. *Sci. Hortic. (Amsterdam)*. 162, 43–48. <https://doi.org/10.1016/j.scienta.2013.07.040>.
- Gasi, F., Kanlić, K., Stroil, B.K., Pojskić, N., Asdal, Å., Rasmussen, M., Kaiser, C., Meland, M., 2016. Redundancies and genetic structure among ex situ apple collections in Norway examined with microsatellite markers. *HortScience* 51, 1458–1462. <https://doi.org/10.21273/HORTSCI11212-16>.
- Gaut, B.S., Díez, C.M., Morrell, P.L., 2015. Genomics and the contrasting dynamics of annual and perennial domestication. *Trends Genet.* <https://doi.org/10.1016/j.tig.2015.10.002>.
- Heinonen, M., Bitz, L., 2019. How to discover traditional varieties and shape in a national germplasm collection: the case of Finnish seed born apples (*Malus × domestica* Borkh.). *Sustainability* 11, 7000. <https://doi.org/10.3390/su11247000>.
- Hjalmarsson, I. (Ed.), 2019. *Äldre äppelsorter för dagens trädgårdar, 1st ed. SLU, Alnarp.*
- Hjalmarsson, I., n.d. Conservation of fruit genetic resources in Sweden. *Acta Horti EUCARPIA symposium*, in press.
- Howard, N.P., van de Weg, E., Bedford, D.S., Peace, C.P., Vanderzande, S., Clark, M.D., Teh, S.L., Cai, L., Luby, J.J., 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. *Hortic. Res.* 4, 17003. <https://doi.org/10.1038/hortres.2017.3>.
- Howard, N.P., Albach, D.C., Luby, J.J., 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. *Ecofruit*. 88–91.
- Larsen, B., Toldam-Andersen, T.B., Pedersen, C., Ørsgaard, M., 2017. Unravelling genetic diversity and cultivar parentage in the Danish apple gene bank collection. *Tree Genet. Genomes* 13. <https://doi.org/10.1007/s11295-016-1087-7>.
- Larsen, B., Gardner, K., Pedersen, C., Ørsgaard, M., Migicovsky, Z., Myles, S., Toldam-Andersen, T.B., 2018. Population structure, relatedness and ploidy levels in an apple gene bank revealed through genotyping-by-sequencing. *PLoS One* 13. <https://doi.org/10.1371/journal.pone.0201889>.
- Lassois, L., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Hibrand-Saint-Oyant, L., Poncet, C., Lasserre-Zuber, P., Feugey, L., Durel, C.E., 2016. Genetic diversity, population structure, parentage analysis, and construction of core collections in the french apple germplasm based on SSR markers. *Plant Mol. Biol. Report.* 34, 827–844. <https://doi.org/10.1007/s11105-015-0966-7>.
- Marconi, G., Ferradini, N., Russi, L., Concezzi, L., Veronesi, F., Albertini, E., 2018. Genetic characterization of the apple germplasm collection in central Italy: the value of local varieties. *Front. Plant Sci.* 9, 1460. <https://doi.org/10.3389/fpls.2018.01460>.
- Miller, A.J., Gross, B.L., 2011. From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414. <https://doi.org/10.3732/ajb.1000522>.
- Morrison, J., 2013. Characterization and correction of error in genome-wide ibd estimation for samples with population structure. *Genet. Epidemiol.* 37, 635–641. <https://doi.org/10.1002/gepi.21737>.
- Muranty, H., Denancé, C., Feugey, L., Crépin, J.L., Barbier, Y., Tartarini, S., Ordidge, M., Troggio, M., Lateur, M., Nybom, H., Papstein, F., Laurens, F., Durel, C.E., 2020. Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biol.* 20, 2. <https://doi.org/10.1186/s12870-019-2171-6>.
- Nilsson, A., 1987. *Våra äppelsorter, 2:1. ed. Allmänna Förlaget, Stockholm.*
- Nybom, H., 2019. Apple production and breeding in Sweden. *Chron. Horticult.* 59, 21–25.
- Ordidge, M., Kirdwachai, P., Fazil Baksh, M., Venison, E.P., George Gibbins, J., Dunwell, J.M., 2018. Genetic analysis of a major international collection of cultivated apple varieties reveals previously unknown historic heteroploid and inbred relationships.

- PLoS One 13. <https://doi.org/10.1371/journal.pone.0202405>.
- Patocchi, A., Fernández-Fernández, F., Evans, K., Gobbin, D., Rezzonico, F., Boudichevskaia, A., Dunemann, F., Stankiewicz-Kosyl, M., Mathis-Jeanneteau, F., Durel, C.E., Gianfranceschi, L., Costa, F., Toller, C., Cova, V., Mott, D., Komjanc, M., Barbaro, E., Kodde, L., Rikkerink, E., Gessler, C., Van De Weg, W.E., 2009. Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. *Tree Genet. Genomes* 5, 211–223. <https://doi.org/10.1007/s11295-008-0176-7>.
- Peakall, R., Smouse, P.E., 2006. GenALEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>.
- Peakall, R., Smouse, R.E., 2012. GenALEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>.
- Peleman, J.D., Van Der Voort, J.R., 2003. Breeding by design. *Trends Plant Sci.* 8, 330–334. [https://doi.org/10.1016/S1360-1385\(03\)00134-1](https://doi.org/10.1016/S1360-1385(03)00134-1).
- Pereira-Lorenzo, S., Urrestarazu, J., Ramos-Cabrera, A.M., Miranda, C., Pina, A., Dapena, E., Moreno, M.A., Errea, P., Llamero, N., Díaz-Hernández, M.B., Santesteban, L.G., Laquidain, M.J., Gogorcena, Y., Urbina, V., Dalmases, J., Ascasibar-Errasti, J., Royo, J.B., 2017. Analysis of the genetic diversity and structure of the Spanish apple genetic resources suggests the existence of an Iberian gene pool. *Ann. Appl. Biol.* 171, 424–440. <https://doi.org/10.1111/aab.12385>.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>.
- Rohlf, 2005. NTSYSp Numerical Taxonomy and Multivariate Analysis System.
- Routson, K.J., Reilley, A.A., Henk, A.D., Volk, G.M., 2009. Identification of historic apple trees in the Southwestern United States and implications for conservation. *HortScience* 44, 589–594. <https://doi.org/10.21273/HORTSCI.44.3.589>.
- Sato, M., Nyui, T., Takahashi, H., Kanda, H., 2007. Comparison of flowering and fruiting of seedlings from reciprocal crosses between diploid and triploid apple cultivars. *J. Japan. Soc. Hort. Sci.*
- Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38, 1409–1438.
- Svensson, H., 2005. *Äpplen i Sverige*, 2nd ed. Bokförlaget Prisma, Stockholm.
- Tange, O., 2011. GNU Parallel - the Command-line Power Tool. <https://doi.org/10.5281/zenodo.16303>.
- Urrestarazu, J., Denance, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Poncet, C., Lateur, M., Houben, P., Ordidge, M., Fernandez-Fernandez, F., Evans, K.M., Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Miranda, C., Gassmann, J., Kellerhals, M., Suprun, I., Pikunova, A.V., Krasova, N.G., Torutaeva, E., Dondini, L., Tartarini, S., Laurens, F., Durel, C.E., 2016. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biol.* 16, 130. <https://doi.org/10.1186/s12870-016-0818-0>.
- Urrestarazu, J., Muranty, H., Denancé, C., Leforestier, D., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Aubourg, S., Celton, J.M., Daccord, N., Dondini, L., Gregori, R., Lateur, M., Houben, P., Ordidge, M., Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Troggio, M., Bianco, L., Velasco, R., Poncet, C., Théron, A., Moriya, S., Bink, M.C.A.M., Laurens, F., Tartarini, S., Durel, C.E., 2017. Genome-wide association mapping of flowering and ripening periods in apple. *Front. Plant Sci.* 8, 1923. <https://doi.org/10.3389/fpls.2017.01923>.
- van de Weg, E., Di Guardo, M., Jansch, M., Socquet-Juglard, D., Costa, F., Baumgartner, I., Brogini, G.A.L., Kellerhals, M., Troggio, M., Laurens, F., Durel, C.-E., 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. <https://doi.org/10.1007/s11032-017-0755-0>.
- van Treuren, R., Kemp, H., Ernsting, G., Jongejans, B., Houtman, H., Visser, L., 2010. Microsatellite genotyping of apple (*Malus × domestica* Borkh.) genetic resources in the Netherlands: application in collection management and variety identification. *Genet. Resour. Crop Evol.* 57, 853–865. <https://doi.org/10.1007/s10722-009-9525-0>.
- Vanderzande, S., Micheletti, D., Troggio, M., Davey, M.W., Keulemans, J., 2017. Genetic diversity, population structure, and linkage disequilibrium of elite and local apple accessions from Belgium using the IRSC array. *Tree Genet. Genomes* 13. <https://doi.org/10.1007/s11295-017-1206-0>.
- Vanderzande, S., Howard, N.P., Cai, L., Da Silva Linge, C., Antanaviciute, L., Bink, M.C.A.M., Kruisselbrink, J.W., 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>.