

# Diversity And Population Structure of Nordic Potato Cultivars and Breeding Clones Grown Under Long Daylength in Europe

**Catja Selga**

Swedish University of Agricultural Sciences: Sveriges lantbruksuniversitet

**Pawel Chrominski**

NordGen: Nordic Genetic Resource Center

**Ulrika Carlson-Nilsson**

NordGen: Nordic Genetic Resource Center

**Mariette Andersson**

Swedish University of Agricultural Sciences: Sveriges lantbruksuniversitet

**Aakash Chawade**

Swedish University of Agricultural Sciences: Sveriges lantbruksuniversitet

**Rodomiرو Ortiz** (✉ [rodomiرو.ortiz@slu.se](mailto:rodomiرو.ortiz@slu.se))

Swedish University of Agricultural Sciences: Sveriges lantbruksuniversitet <https://orcid.org/0000-0002-1739-7206>

---

## Research Article

**Keywords:** potato, genetic diversity, population structure, potato breeding, genebank

**Posted Date:** December 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1116756/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The genetic diversity and population structure of breeding germplasm is central knowledge for crop improvement. To gain insight into the genetic potential of the germplasm used for potato breeding in a Nordic breeding program as well as the collections from the Nordic genebank (NordGen), 133 potato genotypes were genotyped using the Infinium Illumina 20K SNP array. After SNP filtering, 11 610 polymorphic SNPs were included in the analysis. In addition, data from three important breeding traits – percent dry matter and uniformity of tuber shape and eye – were scored to measure the variation between groups. The genetic diversity among the genotypes was estimated using principal coordinate analysis based on the genetic distance between individuals, as well as by using the software STRUCTURE. Both methods suggest that the collected breeding material and the germplasm from the genebank are closely related, with a low degree of population structure between the groups. The phenotypic distribution among the genotypes revealed significant differences, especially between farmer's cultivars and released cultivars and breeding clones. The percent heterozygosity was similar between the groups, with a mean average of 58–60%. Overall, the breeding germplasm and the collection of genotypes from the Nordic gene bank seems to be closely related with similar genetic background. This gains insight that the genetic potential of available Nordic potato breeding germplasm is low, and for genetic hybridization purposes, genotypes from outside the Nordic region should be employed.

## Introduction

The cultivated potato (*Solanum tuberosum* L.) originates from the Andes in South America, where it was domesticated 8 000–10 000 years ago (Spooner et al. 2005). The import of potato landraces, particularly from southern Chile to northern latitudes in Europe, resulted in potato cultivars already adapted to long-days. This adaptation was later found to be very favourable for cultivation in Europe and contributed to the genetic background of many popular cultivars grown today (Ghislain et al. 2009, Hardigan 2017; Gutaker et al. 2019). The potato was first introduced in Europe in the second half of the 16<sup>th</sup> century (Hawkes and Francisco-Ortega 1993) and from early on, several different genotypes were brought to Europe from South America (Ríos et al. 2007). Over the course of the 17<sup>th</sup> and 18<sup>th</sup> centuries, potato gained acceptance as a food crop in Europe (Ames and Spooner 2008), and became particularly popular in northern Europe (McNeill 1999).

Potato reached the Nordic region of Europe (Sweden, Denmark, Norway, Finland, and Iceland) later than the rest of the countries on in Europe. It was introduced to Sweden in 1658; i.e., about 50 years after other European sites (Burton 1989). The first potato plants cultivated in the Nordic region originated from several European sites (Bjor 2001; Erjefält 2001; Ólafsson 2001, Tolstrup 2001; Varis 2001). These first plants established the genetic background to the farmer's varieties (sometimes referred to as landraces (Maxted et al. 2020)), which were popular before professional potato breeding began in the early 20<sup>th</sup> century (Erjefält 2001) and often resulted in the sowing of potato true seeds rather than tubers in early day potato cultivation (Osvold 1965).

Sweden was the first country in the Nordic region to start a formal potato breeding program through work at the Swedish Seed Association (Svalöf) in 1903. This breeding program changed ownership throughout the 20<sup>th</sup> century and is since 2006 situated at the Department of Plant Breeding at the Swedish University of Agricultural Sciences (SLU).

Since 1979, the Nordic Genetic Resource Center (NordGen) has collected released potato cultivars, old farmer's cultivars, and breeding clones from the Nordic countries for future conservation (Veteläinen et al. 2005). The potato accessions at NordGen are kept *in vitro*, and metadata such as morphological characteristics, collecting site, release year, and pedigree information are kept in their database GENBIS. Their collection of farmer's cultivars was characterised by Veteläinen et al. in 2005 studying a wide range of morphological traits and by using 63 amplified fragment length polymorphism (AFLP) markers. In this study, no population structure based on country origin could be revealed either by studying the phenotypic or genotypic data. After the study was published, new tools and methods have made high-throughput genotypic data more accessible and affordable with the release of the potato genome reference sequence (Potato Genome Consortium 2001).

In our research that we report herein we studied the diversity of 133 potato genotypes from SLU and NordGen. Genotypic data was collected using the GGPv3.0 array (Vos *et al.* 2015), and the single nucleotide (SNP) data collected were used to evaluate the genetic structure and level of heterozygosity within each and among the three populations. Phenotypic data were also recorded for three important breeding traits as a measure of the variation between the groups.

## Materials And Methods

### *Plant material*

In total, the sample of our study included 133 genotypes from the germplasm collections at the potato-breeding program at SLU (n = 49), NordGen (n = 75), and a selection of nine released cultivars popularly grown in Sweden today and commonly used as parents for the breeding germplasm at SLU (Supplementary table S1). From SLU, breeding clones that had undergone at least two cycles of selection (T<sub>3</sub>) were selected, and clonal generation were defined for each genotype until five cycles of selection (T<sub>6+</sub>). The information about the nine released cultivars regarding both country of origin (breeder or breeding company) and release year were available through the Potato Pedigree Database (Hutten and Berloo, 2001; Berloo et al. 2007). The germplasms from NordGen contained genotypes of three different clonal types: breeding clones (n = 8), cultivars (n = 38) and farmer's cultivars (n = 29). The origin countries of the genotypes from NordGen were defined as the Nordic country in which the genebank did the collecting. Regarding cultivars, there are two, 'Early Rose' and 'Marius' that are non-Nordic (the former released in the USA and the other in Poland) but in the database they are assigned to Sweden and Norway because of their former extensive cultivation in both.

### *Phenotypic data*

Phenotypic information was available for 109 of the total 133 genotypes. Three breeding traits on potato tubers were investigated – percent dry matter content, tuber eye depth, and tuber shape uniformity. Phenotypic information was lacking for two of the released cultivars, and for seven of SLU breeding clones. From the NordGen collection, 60 out of 75 genotypes had phenotypic data available, recorded as described by Veteläinen *et al.* (2005). The phenotypic data for the genotypes belonging to SLU and cultivars grown today were taken from 20–plant plots grown at three field sites in Sweden (Helgegården 56°02'N 14°07'E, Mosslunda 55°98'N 14°10'E and Umeå 63°84'N 20°26'E) during 2018. Tubers were sown on 14–15<sup>th</sup> May and harvested on 11–12<sup>th</sup> October. Tuber shape uniformity was recorded using a scale ranging from 1 (non-uniform) to 9 (uniform). Tuber eye depth was recorded on a scale ranging from 1 (deep) to 9 (shallow). Specific gravity was estimated as tuber weight in air divided by subtraction of tuber weight in air and tuber weight in water. The estimates on specific gravity were converted to percent dry matter content as described by Mosley and Chase (1993) and adapted to a scale from 1 (low percent dry matter content) to 9 (high percent dry matter content).

Phenotypic variation between and within the groups for each trait was visualized using the R package 'vioplot' (Adler and Kelly, 2020). To detect variation of the means of each of the groups – either by source of genotypes (i.e. NordGen, SLU or grown cultivar) or type of clone (i.e. breeding clone, released cultivar or farmer's cultivar), an analysis of variance was conducted for each of the three breeding traits. If the analysis of variance was significant ( $P < 0.05$ ), Tukey's range test (Tukey 1949) was conducted to determine which of the three groups differed from one another.

### *SNP genotyping*

From each of the 133 genotypes, approximately 0.25 g of leaf tissue was sampled for SNP genotyping. The leaf material was collected on ice and kept at -80 °C until shipping to SGS – TraitGenetics GmbH (Gatersleben, Germany) for DNA extraction, genotyping and allele calling. The GGPv3.0 array (Vos *et al.* 2015) was used for genotyping. For genotype calling, the software 'Illumina GenomeStudio' (Illumina, San Diego, CA) was utilized, scoring four alleles per locus. To ensure the quality of the SNPs to our data set, SNPs with over 10% missing rate were discarded, leaving 14370 SNPs. Genotype callings were translated from base format (ATCG) to numeric format and missing values were imputed using the function *read.GWASpoly* from the R package 'GWASpoly' (Rosyara *et al.* 2016). The imputation method deployed was population mode, where the most frequent allelic state is used to impute missing marker genotypes. Monomorphic markers and markers with a minor allele frequency below 0.5 were discarded, leaving 11610 SNPs to be used in further analysis. The R statistical program was used for analysis (R Core Team, 2021).

### *Genotypic data analysis*

Population structure was investigated using the set of 11 610 SNPs. The allele format was converted from numeric format (0–4) to an AB format, where A represents the major allele and B represents the minor allele at each given locus. The function *stampConvert* from the R package 'StAMPP' (Pembleton

et al. 2013) was used to calculate the allele frequency at each genotype. The genetic distance between individuals was calculated based on Nei's distance (Nei 1972) using the function *stamppNeisD* from the same R package. To investigate population structure, the genetic distance matrix was displayed in two ways. A heatmap was used to visualise the genetic kinship among the genotypes based on Nei's genetic distance. In addition to the heatmap, a principal coordinate analysis (PCoA) (Gower 1966) was carried out to determine the singular value decomposition. The PCoA was carried out using the function *pcoa* from the R package 'ape' (Paradise and Schliep 2019) without correction for negative Eigenvalues. The PCoA was visualized as a scatter plot using the R package 'ggplot2' (Wickham 2016).

The genetic relationships among the genotypes from the NordGen population were further investigated with a hierarchical clustering analysis. The Euclidian distance was calculated among the 75 genotypes from NordGen using 11 610 SNPs in numeric format (0–4). Subsequently, a neighbour joining dendrogram was generated using the function *set* from the R package 'dendextend' (Galili 2015). The country of origin and clonal type for each genotypes was visualized using the function *colored\_bars* from the same package.

The population structure for all genotypes (n=133) was also determined using the 'STRUCTURE' software (Pritchard et al. 2000) with an admixture model. As 'STRUCTURE' is not developed for tetraploid species, the genotypic data was "diploidized" where the three heterozygote states (AAAB, AABB, and ABBB) were converted to a single heterozygote (AB). Three replicates were performed for each suggested number of populations (*K*) ranging from 1 to 10 with 50 000 replicates of the Markov Chain Monte Carlo after 15 000 burn-ins. The optimal *K* value was determined as the log-likelihood of *K*, and visualized using the 'STRUCTURE HARVESTER' web application (Earl and VonHodt 2012). The independent outputs from the 3 replicates of the optimal *K* were permuted using the software 'CLUMPP' (Jakobsson and Rosenberg 2007) based on the algorithm *FullSearch*. The output was subsequently visualised using the software 'DISTRUCT' (Rosenberg 2004).

The level of heterozygosity was measured as the percent of heterozygote SNPs in each genotypes, using a diploid model where the three tetraploid heterozygote states were scored as one. The level of heterozygosity over time was estimated using cultivar release year for the 51 cultivars where these data were available. Pearson's correlation coefficient was calculated to establish any association between release year and level of heterozygosity.

## Results

### *Phenotypic variation*

The sample of genotypes included in this study is of the table potato type grown in the Nordic region of Europe – particularly Sweden. The genotypes were grouped based on two criteria 1) source of the germplasm (i.e. cultivars grown today, NordGen and, SLU), or 2) clonal type (i.e. breeding clones, released cultivars and farmer's cultivars). Phenotypic characteristics of importance to potato marketing – percent dry matter in the tuber flesh, tuber eye depth and tuber shape uniformity are shown in Fig. 1. Variation

among the three sources of data (NordGen, SLU and cultivar grown today) was observed and a significant ( $P < 0.001$ ) difference was found for percent dry matter and tuber shape uniformity. According to Tukey's range test the accessions from NordGen displayed, on average, significantly different percent dry matter and tuber shape uniformity than the two other groups (SLU and cultivar grown today). An alternative division of genotypes was by grouping them by clonal type – breeding clones, released cultivars or farmer's cultivar. There were significant differences for all measured phenotypic traits – percent dry matter, tuber eye depth, and tuber shape uniformity among these three groups ( $P < 0.001$ ) (Fig. 2). For percent dry matter and tuber eye depth, Tukey's range test presented differences between farmer's cultivars and the other two groups. For tuber shape uniformity, Tukey's range test presented differences between breeding clones and the two other groups.

### *Genetic structure*

After marker reduction, 11610 SNP markers remained for all 133 genotypes. Nei's genetic distance was used to do the PCoA among all genotypes (Fig. 3). A weak population structure was revealed by the PCoA, with a smaller cluster of SLU breeding clones from early clonal generations ( $T_3$  and  $T_4$ ) separating from the remnant accessions (Fig. 3B). The PCoA revealed no clear population structure based on country of origin for the obsolete cultivars held at NordGen (Fig. 3A), or breeding clone generation (Fig. 3C). A heatmap based on Nei's genetic distance between genotypes also revealed a similar pattern as the PCoA, with a grouping of early selection breeding clones from SLU having a smaller group than remaining accessions (Fig. 4B). The optimal  $K$  value related to the population structure was two, using a diploid genotyping model for the software 'STRUCTURE' (Pritchard *et al.* 2000). Most genotypes were an admixture of the two populations, suggesting a weak population structure (Fig. 4A and Fig. 4C). However, two clusters in the heatmap did correspond to the two subpopulations proposed by the results from 'STRUCTURE', a subset containing predominantly farmer's cultivars, and a subset containing breeding material from SLU.

### *Genetic diversity of potatoes from NordGen*

Using the Euclidian distance from the SNP marker data, a dendrogram was built for the 75 accessions from NordGen (Fig. 5). The 75 accessions were grouped according to both country of origin (Sweden, Denmark, Finland, Norway or Iceland), and type of clone (farmer's cultivar, released cultivar or breeding clone). A pattern emerged after grouping for type of clone (Fig. 5B): the majority of the released cultivars and breeding clones formed a distinct group from the farmer's cultivars.

A subset of 25 accessions from NordGen was further investigated using another dendrogram, which facilitating comparing it to a previous study using AFLP markers (Veteläinen *et al.* 2005). This dendrogram showed a completely different branching among the subset of accessions (Supplementary Fig. S1) compared to the previously available dendrograms based on either genotypic or phenotypic data. The dendrogram containing the smaller subset of individuals did not display any grouping according to country of origin (Supplementary Fig. S1).

## *Level of heterozygosity*

The level of heterozygosity was defined as the number of heterozygote allele callings in 14730 SNP markers and was determined for all genotypes. Percent heterozygosity did not differ much among accessions, with a few exceptions. A trend of heterozygosity among accessions from the three groups based on source of the germplasm (NordGen, SLU and grown cultivar) was not observed, neither when studying the genotypes separately (Fig. 6A) or by group (Fig. 6B). The percent heterozygosity was overall lower in SLU breeding clones than in NordGen accessions. However, the mean level of heterozygosity was around 48 to 50 % for all three groups based on source of the germplasm.

The percent heterozygosity among genotypes when grouped by clonal type followed the same trend and did not show any difference between the groups (Fig.7A). Farmer's cultivars had the greatest variation for heterozygosity, and the breeding clones did show a lower percent heterozygosity than released cultivars (Fig. 7B). Nonetheless, the mean level of heterozygosity was very similar between the three clonal types (48–49 %).

The level of heterozygosity was checked over time using available information of year of release for 51 cultivars. Pearson's product-moment correlation showed a slight but not significant ( $P > 0.05$ ) negative trend (-0.27) towards lower levels of heterozygosity over time (Fig. 8).

## **Discussion**

Understanding the genetic diversity of breeding germplasm is one of the fundamental tools for a plant breeder, to make informed crosses for the development of new and improved cultivars. By understanding the genetic diversity of *in vitro* collections of potato accessions in genebanks, plant breeders may utilize it wisely in their crossing schemes. In this study, we used SNPs from an array containing 22K SNP markers to assess the genetic diversity of accessions used for breeding table potatoes for the Swedish market. The accessions included breeding clones from the SLU breeding program; germplasm from NordGen made up of old breeding clones, obsolete cultivars and farmer's cultivars; and a few cultivars grown today often used as breeding parents in the SLU breeding program.

The number of genotypes (population size) are unevenly distributed among the groups being compared throughout all analysis in this study. The largest difference in population size was found when comparing the three groups based on source of germplasm (NordGen, SLU and cultivars grown today), the group cultivars grown today contain only nine accessions (five with phenotypic data available), and NordGen contain 75 accessions. There are about 25 cultivars commonly grown for the potato market today in Sweden (personal communication: Anders Andersson, Potatisodlarna). Among these 25 cultivars are 'King Edward', 'Bintje', 'Connect' and 'Carolus', which we have included in this study. The most commonly grown early maturing cultivar in Sweden by far is 'Solist'. Both the phenotypic and genotypic diversity could be limited to the number of individuals, which must be considered when discussing the results from this study. The grouping based on clonal type has a more even distribution of population size, spanning from 29 (farmer's cultivars) to 57 (breeding clones) genotypes.

The phenotypic data from the NordGen accessions has been extensively described in a previous publication (Veteläinen et al. 2005). We combined these old phenotypic records with new data recorded for clones from the SLU breeding program to compare the phenotypic diversity among both germplasm sources. Phenotypic traits such as percent dry matter are affected by environmental factors, for example post-harvest storage conditions and soil type (Wilson and Lindsay 1969; Schrippers 1976). There were no cultivars used as checks in common for the study by Veteläinen et al. and the field trial described here, hence it was not possible to determine any environmental effects on the three phenotypic traits. The results show that on the group level, all measured phenotypic traits differ among the three groups based on source of the germplasm (Fig. 1). Tuber shape uniformity significantly differs among the groups, and the highest variation for this trait was noted for SLU accessions, and more specifically, among the breeding clones. An explanation to why the genotypes from SLU exhibits a large variation in tuber shape uniformity may be that this trait is not considered as important compared to other breeding traits during the early selection cycles of the breeding program ( $T_1$ - $T_4$ ). A larger proportion of the genotypes from NordGen has a higher percent dry matter content compared to the other two groups, a character that is correlated to mealy potato tubers (Burton 1989). When studying the phenotypic variation among the different groups based on clonal type (Fig. 2), it becomes clear that the higher scores for percent dry matter from NordGen originated from the farmer's cultivars within this population. The farmer's cultivars have, on average, higher percent dry matter than the breeding clones and cultivars grown today in Sweden. The largest spread of percent dry matter scores was noticed among the breeding clones. This result suggests that this trait was not the target when making selections in the potato breeding program. The farmer's cultivars also differ from the two other clonal types regarding tuber eye depth. The farmer's cultivars seem to have much deeper tuber eyes compared to the two other clonal types. Potato breeders are often selecting for shallow eyes as this makes tubers easier to peel for the consumers of table potato. These farmer's cultivars did not result from breeding efforts, but instead they reflect what characters potato growers and consumers have favoured. Hence, the tuber characters favoured by end users differ geographically and over time. In Sweden today, 'King Edward' remains the favourite table potato mainly due to its appreciated flavour and good properties for the production of mashed potatoes as well for other uses in gastronomy. Such a finding shows that flavour and possibility to use a cultivar for many purposes should be considered when breeding new cultivars.

In general, potato exhibits a low degree of population structure (D'hoop et al. 2010; Ortiz 2015), with European cultivars in particular stemming from a very narrow genetic base (Glendinning 1983; Srivastava et al. 2016). The results from our study suggest that there is a very limited population structure among potato cultivars and breeding clones in the Nordic region. In line with what previously was reported by Veteläinen et al. (2005), no population structure was observed based on country of origin (Fig. 3A). The NordGen definition of country of origin is, however, limited to and defined as the country in which the accession was collected, and it is known that several of the farmer's cultivars have been grown in more than a single Nordic country. This as well as the limited genetic base of European potato cultivars in general may have affected the lack of structure based on country of origin (Fig. 3A).

Population structure was examined using a PCoA based on Nei's pairwise distance among genotypes (Fig. 3A-C) and STRUCTURE (Pritchard et al. 2000) with  $K$  ranging from 0 to 10 (Fig. 4A and Fig. 4C). The  $K$  with the maximum likelihood was 2, thus suggesting two subpopulations in the data. This theoretical number of subpopulations is to a limited extent explained in the results from the PCoA and heatmap. A group of breeding clones from SLU does cluster separately from the other material. When studying this group more closely, it appears that clones from the early cycles of selection in SLU breeding program are represented in this outlying group (Fig. 3B, Fig. 4C). Breeding clones from SLU which have undergone a larger number of cycles of selection are in the same cluster with the rest of the material, thereby suggesting that the potato breeders are actively making selections towards clones that are similar to cultivars grown in Sweden today and farmer's cultivars previously grown in the Nordic region.

The bar plot from the STRUCTURE output suggests a possible subpopulation division based on farmer's cultivars versus the remaining materials (Fig. 4A). This division of subpopulations is unclear when studying the PCoA (Fig. 3C). However, there is a cluster represented by farmer's cultivars appearing in the heatmap (Fig. 4B), which is based on the same genetic data as the PCoA. A similar population structure, where farmer's cultivars grouped in a separate subpopulation was found in Chinese potato collections (Wang et al. 2019). It is still important to keep in mind that almost every genotype was an admixture of both theoretical subpopulations, which suggests a weak population structure.

No structure based on country of origin was observed, which could be explained by the narrow genetic base of the sampled accessions. The pedigree information is limited for potato in general, but it is not inconceivable that the farmer's cultivars included in this study would appear in the pedigree of the breeding clones either from SLU or among the NordGen accessions. Several of the cultivars grown in Sweden today are actively used as breeding parents at the SLU potato breeding program. It would have been interesting to include genotypes with a wider geographical background, especially outside Europe to get an estimation if the Nordic region has been uniquely differentiated compared to potato grown elsewhere in the world.

Previous research found that market class was a good biological explanation to subpopulations in cultivated tetraploid potato (Hirsch et al. 2013; Igarashi et al. 2019; Pandey et al. 2021). In this study, all potato accessions are classified as table potato with one exception, the cultivar 'Kuras' from the population of cultivars grown in Sweden today. 'Kuras' is grown for starch production (Bauw et al. 2006). We did not, however, observe this as an outlier in our genetic structure analysis, but instead it grouped close to the other Dutch cultivars in the study (data not shown).

The genetic diversity of the potato germplasm from NordGen has been examined previously by Veteläinen et al. (2005). Their study, focusing on finding morphological characters to identify duplicates among accession, included a fraction of the accessions used in our study. All of them were farmer's cultivars. Our results using a 22K SNP array supports what Veteläinen et al. (2005) noticed; i.e., duplicated accessions of potato did not exist in the NordGen genebank. The study also included a genetic diversity study assessment using 63 AFLPs. The dendrogram from their study based on AFLPs was

different than ours based on SNPs (Supplementary Fig. S1). The dendrogram by Veteläinen et al. (2005) based on their 57 morphological characters did not match with the dendrogram from our study (Supplementary Fig. S1). A separate dendrogram containing all accessions from the NordGen population was also drawn using the SNP data (Fig. 5). As in the PCoA (Fig. 3A), no population structure was revealed based on country of origin from the dendrogram. When the accessions were assigned as type of clone, the structure could be explained by the clustering in the dendrogram to a certain degree. Two subgroups were revealed, where most of the released cultivars and breeding clones branched separately from the farmer's cultivars.

Crop yield potential has successfully been improved through heterosis for other crops (Breiger 1950; Duvick 2001; Lippman and Zamir 2007). It has been theorized that over time, potato breeding will increase the level of heterozygosity due to heterosis related to increased tuber yield (Mendoza and Haynes 1974; Hirsch et al. 2013). In line with what was found by Hirsh et al. (2003) studying US potato cultivars, the level of heterozygosity did not change significantly over time in our study. However, the number of cultivars were significantly larger in the later years compared to the early years, making the estimations skewed.

The average percent heterozygosity ranged from 48 to 50 % for each population included in this study. This is a bit lower than what was recorded previously in other germplasms. The average percent heterozygosity was 56 % or 57% in germplasm from the USA or Japan respectively (Hirsh et al. 2013, Igarashi et al. 2019). Another US germplasm set was investigated by Pandey et al. (2021) with an average percent heterozygosity of 60 %. In both studies in the USA as well as in the study in Japan, a very similar or identical SNP array was utilized for genotyping as in our study. Hence, the genetic variability of the Nordic germplasm is slightly lower. The percent heterozygosity did not vary between populations or types of clones, thus contradicting the theory that breeding would increase the germplasm's level of heterozygosity, which may otherwise suggest limited genetic gains for tuber yield in potato breeding as already noted in the USA (Douches et al. 1996).

Employing SNP arrays for genotyping may lead to ascertainment bias in population research such as ours (Nielsen 2004; Albrechtsen et al. 2010; Heslot et al. 2013; Geibel et al. 2021). SNP arrays often show an underrepresentation of SNPs with extreme allele frequency and are limited to the heterozygosity of the loci found using a limited panel of genotypes included in the development of the array. The 22K SNP array used in this study was developed using 569 unique accessions selected from all over the world, but with an emphasis on European cultivars (Vos et al. 2015). Several of the cultivars grown in Sweden today were included in the development of the SNP array ('Bintje', 'Kuras', 'Sarpö Mira', 'Desirée' and 'Bionica'), and some of these cultivars are used as parents in the SLU breeding program. However, only one non-Nordic accession kept at NordGen was included when the array was developed (the US cultivar 'Early Rose'). Hence, a lot of the genetic diversity present in this group may be unavailable when using this genotyping method.

In conclusion, the germplasm from NordGen and the potato breeding program at SLU seems to be closely related. There is a slightly larger variation spread among the breeding clones from SLU according to our genetic analyses. While the spread of phenotypes might be larger among the farmer's cultivars kept at NordGen. The results generated will be of interest to potato breeders in Sweden and other countries of the Nordic region as they consider introducing accessions from NordGen to expand the genetic diversity in their breeding programs.

## Declarations

### *Funding*

The authors thank grant and other funding provided by the Swedish University of Agricultural Sciences (SLU), the Swedish Research Council Formas for the project (2019– 2022). “Genomisk prediktion i kombination med högkapacitetsfenotypning för att öka potatisens knölskörd i ett föränderligt klimat”, and Stiftelsen för miljöstrategisk forskning (Mistra) for the “Mistra Biotech” project (2012–2020).

### *Conflicts of interest*

The authors declare no conflicts of interest.

### *Availability of data*

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

### *Code availability*

Not applicable.

### *Authors contribution*

Catja Selga designed the experiment, collected germplasm from SLU, performed all analyses and generated all figures and tables and wrote the original draft. Pawel Chrominski collected germplasm from NordGen and contributed to the original draft. Ulrika Carlsson-Nilsson contributed to original draft. Mariette Andersson reviewed and edited the paper. Aakash Chawade designed the experiment, reviewed and edited the paper. Rodomiro Ortiz designed the experiment and contributed to the original draft.

### *Ethical approval*

Not applicable.

### *Consent to participate*

Not applicable.

Not applicable.

## Acknowledgements

The authors would like to thank Fredrik Reslow for his assistance in collecting the breeding trait data for the SLU germplasm.

## References

1. Adler D, Kelly ST (2020) *vioplot: violoin plot*. R package version 0.3.6, URL: <https://github.com/TomKellyGenetics/vioplot>
2. Albrechtsen A, Cilius Nielsen F, Nielsen R (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Mol Biol Evol* 27:2534–2547. <https://doi.org/10.1093/molbev/msq148>
3. Ames M, Spooner DM (2008) DNA from herbarium specimens settles a controversy about origins of the European potato. *Am J Bot* 95:252–257. <https://doi.org/10.3732/ajb.95.2.252>
4. Bauw G, Nielsen HV, Emmersen J, Nielsen KL, Jørgensen M, Welinder KG (2006) Patatins, Kunitz protease inhibitors and other major proteins in tuber of potato cv. Kuras. *FEBS J* 273:3569–3584. <https://doi.org/10.1111/j.1742-4658.2006.05364.x>
5. van Berloo R, Hutten RCB, van Eck HJ, Visser RGF (2007) An online potato pedigree database resource. *Potato Res* 50:45–57. <https://doi.org/10.1007/s11540-007-9028-3>
6. Bjor T (2001) The history of the potato in Norway.. In: In: Veteläinen M, Bennvid I-B (eds) *Kartofler i Norden. En beskrivelse af gamle kartoffelsorter i Nordisk Genbank (The Nordic potato book)*. CAL-förlaget AB, Halmstad, p 55
7. Breiger FG (1950) The genetic basis of heterosis in maize. *Genetics* 35:420–445
8. Burton WG (1989) *The Potato*, vol 21, 3rd edn. Longman Scientific & Technical, New York, pp 401–402
9. Duvick DN (2001) Biotechnology in the 1930s: the development of hybrid maize. *Nat Rev Genet* 2:69–74. <https://doi.org/10.1038/35047587>
10. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
11. Eriksson D, Carlson-Nilsson U, Ortíz R, Andreasson E (2016) Overview and breeding strategies of table potato production in Sweden and the Fennoscandian region. *Potato Res* 59:279–294. <https://doi.org/10.1007/s11540-016-9328-6>
12. Erjefält L (2001) The history of the potato in Sweden.. In: In: Veteläinen M, Bennvid I-B (eds) *Kartofler i Norden. En beskrivelse af gamle kartoffelsorter i Nordisk Genbank (The Nordic potato book)*. CAL-förlaget AB, Halmstad, pp 65–66

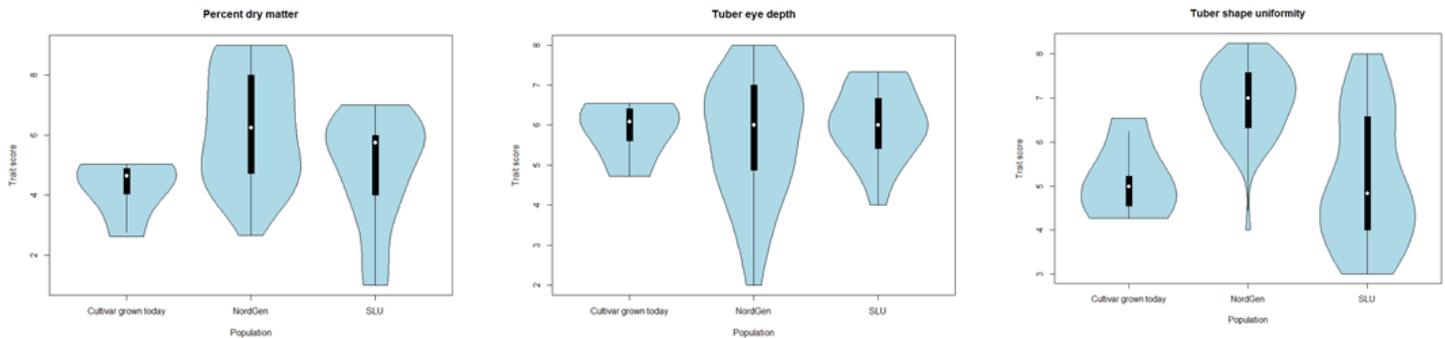
13. D'hoop BB, Paulo MJ, Kowitzanich K et al (2010) Population structure and linkage disequilibrium unravelled in tetraploid potato. *Theor Appl Genet* 121:1151–1170. <https://doi.org/10.1007/s00122-010-1379-5>
14. Douches DS, Maas D, Jastrzebski K, Chase RW (1996) Assessment of potato breeding progress in the USA over the last century. *Crop Sci* 36:1544–1552. <https://doi.org/10.2135/cropsci1996.0011183X003600060024x>
15. Galili T (2015) dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics* 31:3718–3720. <https://doi.org/10.1093/bioinformatics/btv428>
16. Geibel J, Reimer C, Weigend S, Weigend A, Pook T, Simianer H (2021) How array design creates SNP ascertainment bias. *PLoS ONE* 16:e0245178. <https://doi.org/10.1371/journal.pone.0245178>
17. Ghislain M, Núñez J, Herrera MR, Spooner DM (2009) The single *Andigenum* origin of neo-tuberosum potato materials is not supported by microsatellite and plastid marker analysis. *Theor Appl Genet* 118:963–969. <https://doi.org/10.1007/s00122-008-0953-6>
18. Glendinning DR (1983) Potato introductions and breeding up to the early 20th century. *New Phytology* 94:479–505
19. Gower JC (1966) Some distance properties of latent root and vector methods in multivariate analysis. *Biometrika* 53:325–338. <https://doi.org/10.1093/biomet/53.3-4.325>
20. Gutaker RM, Weiß CL, Ellis D et al (2019) The origin and adaptation of European potatoes reconstructed from historical genomes. *Nat Ecol Evol* 3:1093–1101. <https://doi.org/10.1038/s41559-019-0921-3>
21. Hardigan MA, Laimbeer PE, Newton L et al (2017) Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *PNAS* 114:E9999–E10008. <https://doi.org/10.1073/pnas.1714380114>
22. Hawkes JG, Francisco-Ortega J (1993) The early history of the potato in Europe. *Euphytica* 70:1–7. <https://doi.org/10.1007/BF00029633>
23. Heslot N, Rutkoski J, Poland J, Jannik J-L, Sorrells ME (2013) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS ONE* 8:e74612. <https://doi.org/10.1371/journal.pone.0074612>
24. Hirsch CN, Hirsch CD, Felcher K et al (2013) Retrospective view of North American potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. *G3* 3:1003–1013. <https://doi.org/10.1534/g3.113.005595>
25. Hutten RCB, van Berloo R (2001) An online potato pedigree database. URL: <http://www.plantbreeding.wur.nl/PotatoPedigree/>
26. Igarashi T, Tsuyama M, Ogawa K, Koizumi E, Sanetomo R, Hosaka K (2019) Evaluation of Japanese potatoes using single nucleotide polymorphisms (SNPs). *Mol Breeding* 39:9. <https://doi.org/10.1007/s11032-018-0917-8>

27. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
28. Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. *Trends Genet* 23:60–66. <https://doi.org/10.1016/j.tig.2006.12.006>
29. Maxted N, Hunter D, Ortiz R (2020) *Plant genetic conservation*. Cambridge University Press, Cambridge, p 546. <https://doi.org/10.1017/9781139024297>
30. McNeill WH (1999) How the potato changed the world's history. *Soc Res* 66:67–83
31. Mendoza HA, Haynes FL (1974) Genetic basis of heterosis for yield in autotetraploid potato. *Theor Appl Genet* 45:21–25
32. Mosley AR, Chase RW (1993) Selecting cultivars and obtaining healthy seed lots. *Potato Health Management*. APS Press, St Paul, MN, pp 19–27
33. Nei M (1972) Genetic distance between populations. *Am Nat* 106:283–292
34. Nielsen R (2004) Population genetic analysis of ascertained SNP data. *Hum Genomics* 1:218–224. <https://doi.org/10.1186/1479-7364-1-3-218>
35. Ólafsson S (2001) The history of the potato in Iceland.. In: In: Veteläinen M, Bennvid I-B (eds) *Kartofler i Norden. En beskrivelse af gamle kartoffelsorter i Nordisk Genbank (The Nordic potato book)*. CAL-förlaget AB, Halmstad, p 49
36. Ortiz R (2015) *Plant Breeding in the Omics Era*. Springer International Publishing AG, New York, pp 205–223. <https://doi.org/10.1007/978-3-319-20532-8>
37. Ortiz R, Selga C, Reslow F, Carlson-Nilsson U (2020) Svensk potatisförädling: Breeding the new table and crisp potatoes. *Sveriges Utsädesförenings Tidskrift* 1:16–26
38. Osvald H (1965) *Potatisen: odlingshistoria och användning*. Almqvist och Wiksells boktryckeri aktiebolag, Uppsala, Sweden, pp 129-130, 134
39. Pandey J, Scheuring DC, Koym JW et al (2021) Genetic diversity and population structure of advanced clones selected over forty years by a potato breeding program in the USA. *Sci Rep* 11:8344. <https://doi.org/10.1038/s41598-021-87284-x>
40. Paradise E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–582
41. Pembleton L, Cogan N, Forster J (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour* 13:946–952. <https://doi.org/10.1111/1755-0998.12129>
42. Potato Genome Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195. <https://doi.org/10.1038/nature10158>
43. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959

44. R Core Team (2021) R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing). URL: <https://www.R-project.org/>
45. Ríos D, Ghislain M, Rodríguez F, Spooner DM (2007) What is the origin of the European potato? Evidence from Canary Island landraces. *Crop Sci* 47:1271–1280. <https://doi.org/10.2135/cropsci2006.05.0336>
46. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
47. Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome* 9:2. <https://doi.org/10.3835/plantgenome2015.08.0073>
48. Schrippers PA (1976) The relationship between specific gravity and percent dry matter in potato tubers. *Am Potato J* 53:111–122. <https://doi.org/10.1007/BF02854115>
49. Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ (2005) A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *PNAS* 102:14694–14699. <https://doi.org/10.1073/pnas.0507400102>
50. Srivastava A, Bhardwaj V, Singh BP, Khurana SMP (2016) Potato diversity and its genetic enhancement.. In: In: Rajpal VR, Rao SR, Raina SN (eds) *Gene Pool Diversity and Crop Improvement, Sustainable Development and Biodiversity*. Springer International Publishing, Switzerland, pp 187–226. <https://doi.org/10.1007/978-3-319-27096-8>
51. Tolstrup K (2001) The potato in Denmark.. In: In: Veteläinen M, Bennvid I-B (eds) *Kartofler i Norden. En beskrivelse af gamle kartoffelsorter i Nordisk Genbank (The Nordic potato book)*. CAL-förlaget AB, Halmstad, pp 32–35
52. Tukey JM (1949) Comparing individual means in the analysis of variance. *Biometrics* 5:99–114
53. Varis E (2001) The potato in Finland from the past to the present.. In: In: Veteläinen M, Bennvid I-B (eds) *Kartofler i Norden. En beskrivelse af gamle kartoffelsorter i Nordisk Genbank (The Nordic potato book)*. CAL-förlaget AB, Halmstad, pp 44–45
54. Veteläinen M, Gammelgård E, Valkonen JPT (2005) Diversity of Nordic landrace potatoes (*Solanum tuberosum* L.) revealed by AFLPs and morphological characters. *Genet Resour Crop Evol* 52:999–1010
55. Vos PG, Utidewilligen JGAML, Voorrips RE, Visser RGF, van Eck HJ (2015) Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. *Theor Appl Genet* 128:2387–2401. <https://doi.org/10.1007/s00122-015-2593-y>
56. Wang Y, Rashid MAR, Li X et al (2019) Collection and evaluation of genetic diversity and population structure of potato landraces and varieties in China. *Front Plant Sci* 10:139. <https://doi.org/10.3389/fpls.2019.00139>
57. Wickham H (2016) *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4. Available at <https://ggplot2.tidyverse.org/>

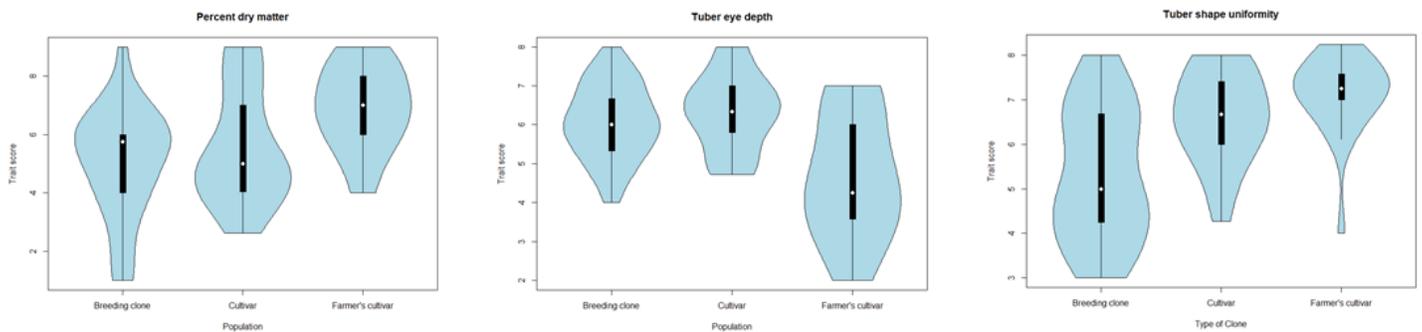
58. Wilson JH, Lindsay AM (1969) The relation between specific gravity and dry matter content of potato tubers. *Am Potato J* 46:323–328. <https://doi.org/10.1007/BF02862002>

## Figures



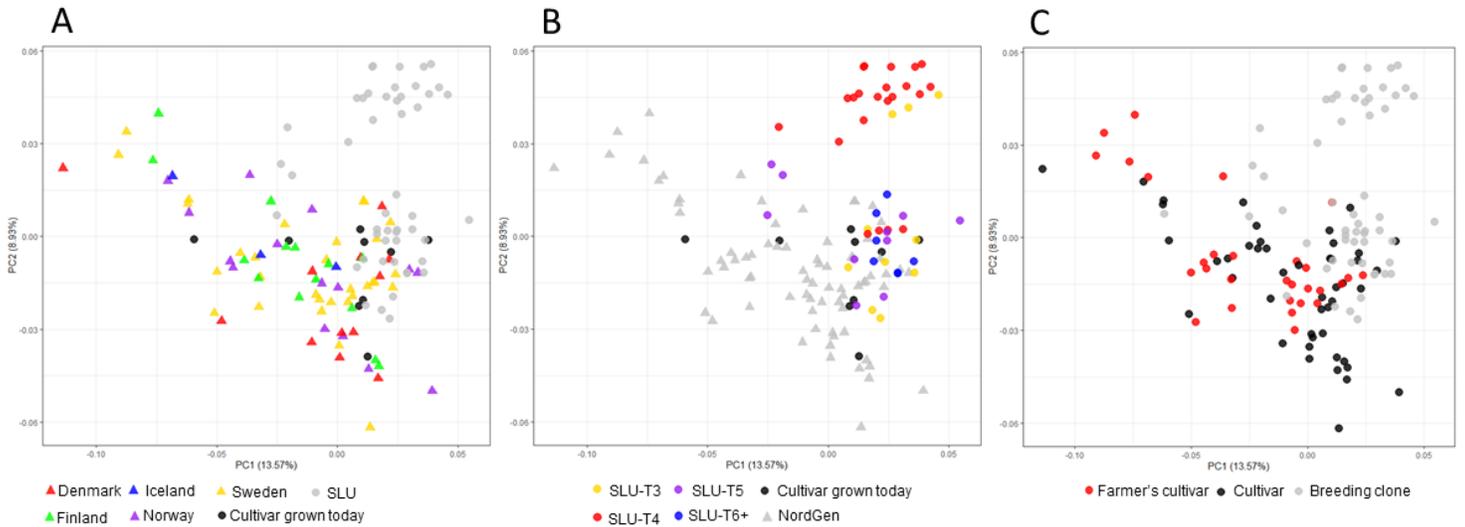
**Figure 1**

Phenotypic distribution for three breeding traits (percent dry matter, tuber eye depth and uniformity of tuber shape) across the three populations included in this study. Trait score can be found on y-axis, population are shown on the x-axis. The white point shows population median value.



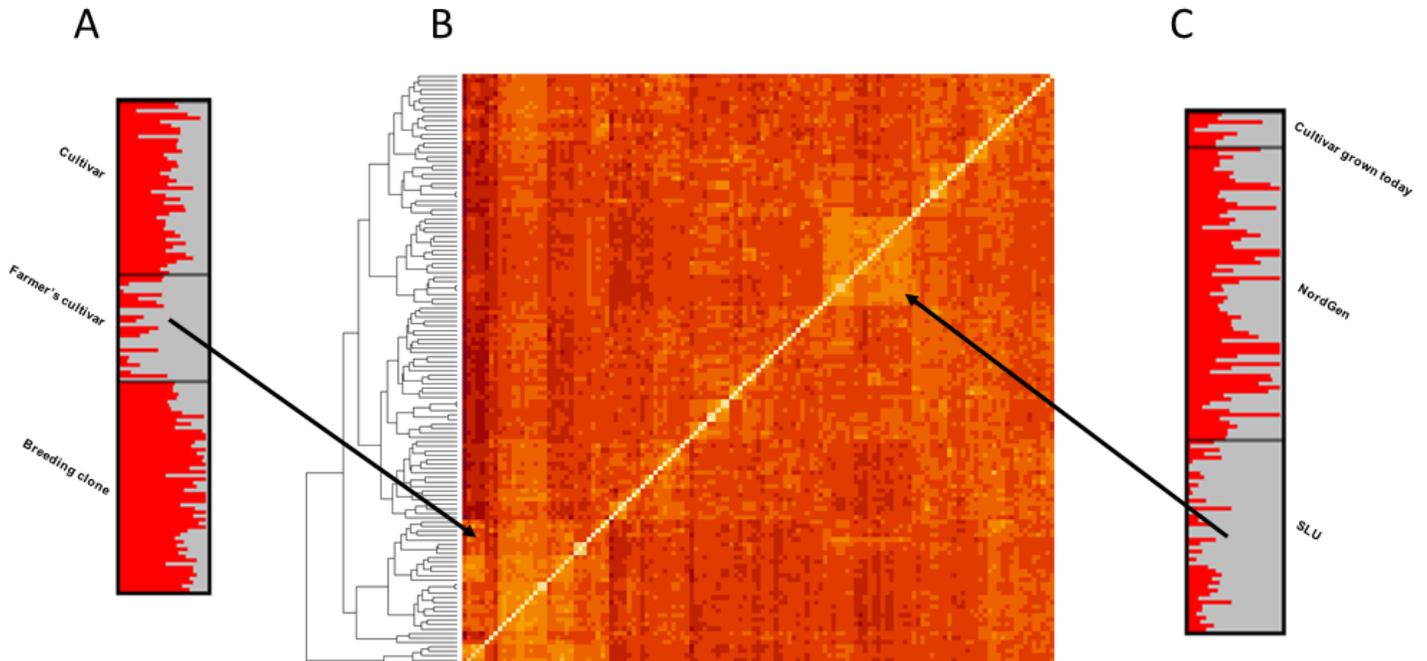
**Figure 2**

Phenotypic distribution for three breeding traits (percent dry matter, tuber eye depth and uniformity of tuber shape) for the three distinct types of clones included in this study. Trait score can be found on y-axis, population are shown on the x-axis. The white point shows population median value.



**Figure 3**

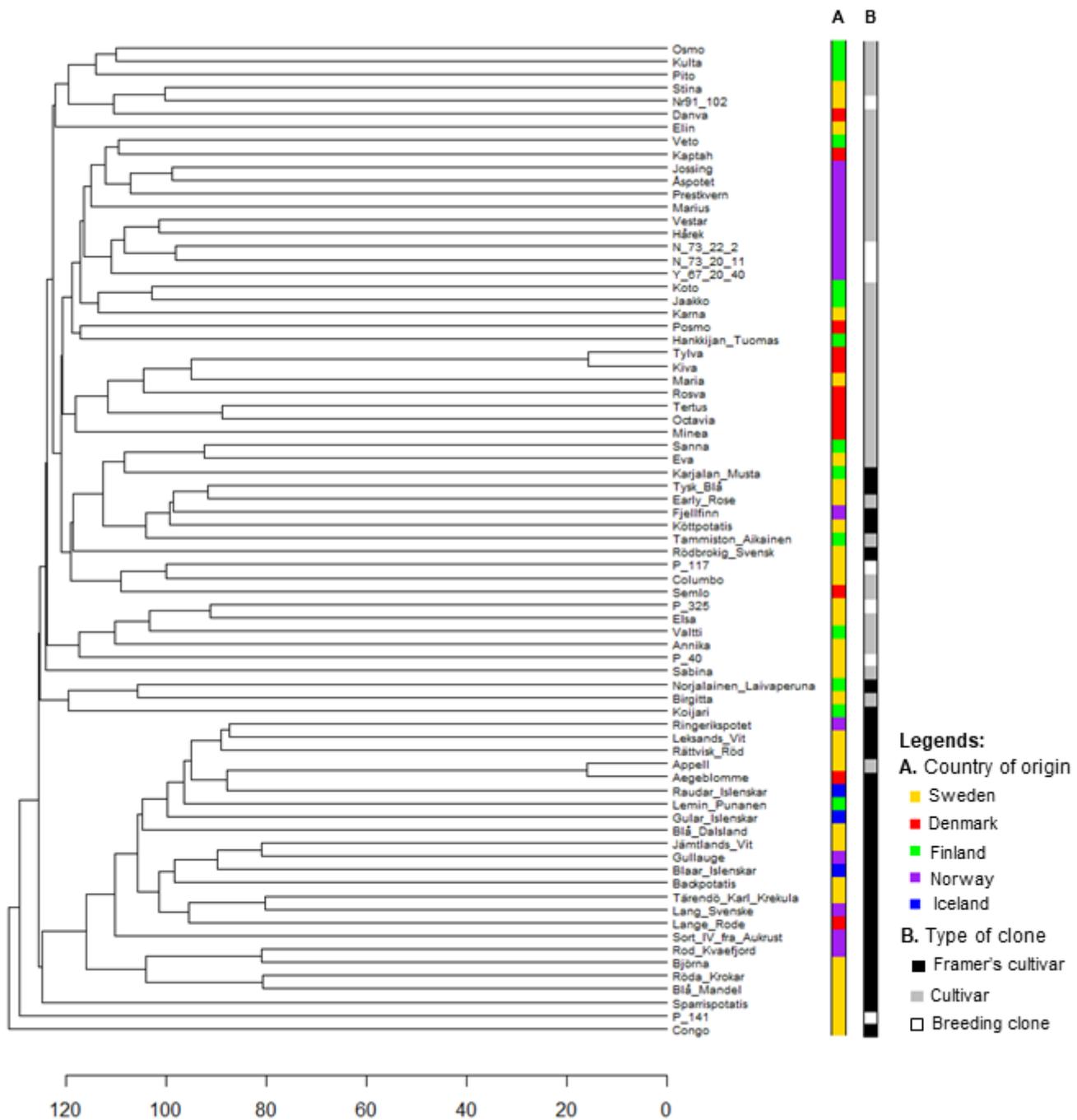
Genetic diversity revealed by a principal component analysis (PCoA) based on Nei's genetic distance of single nucleotide polymorphism (SNP) markers among the 133 accessions. Groupings are based on (A) population and country of origin for NordGen genotypes, (B) population and cycle of selection (T3 – T6+) for SLU breeding clones, and (C) clonal type.



**Figure 4**

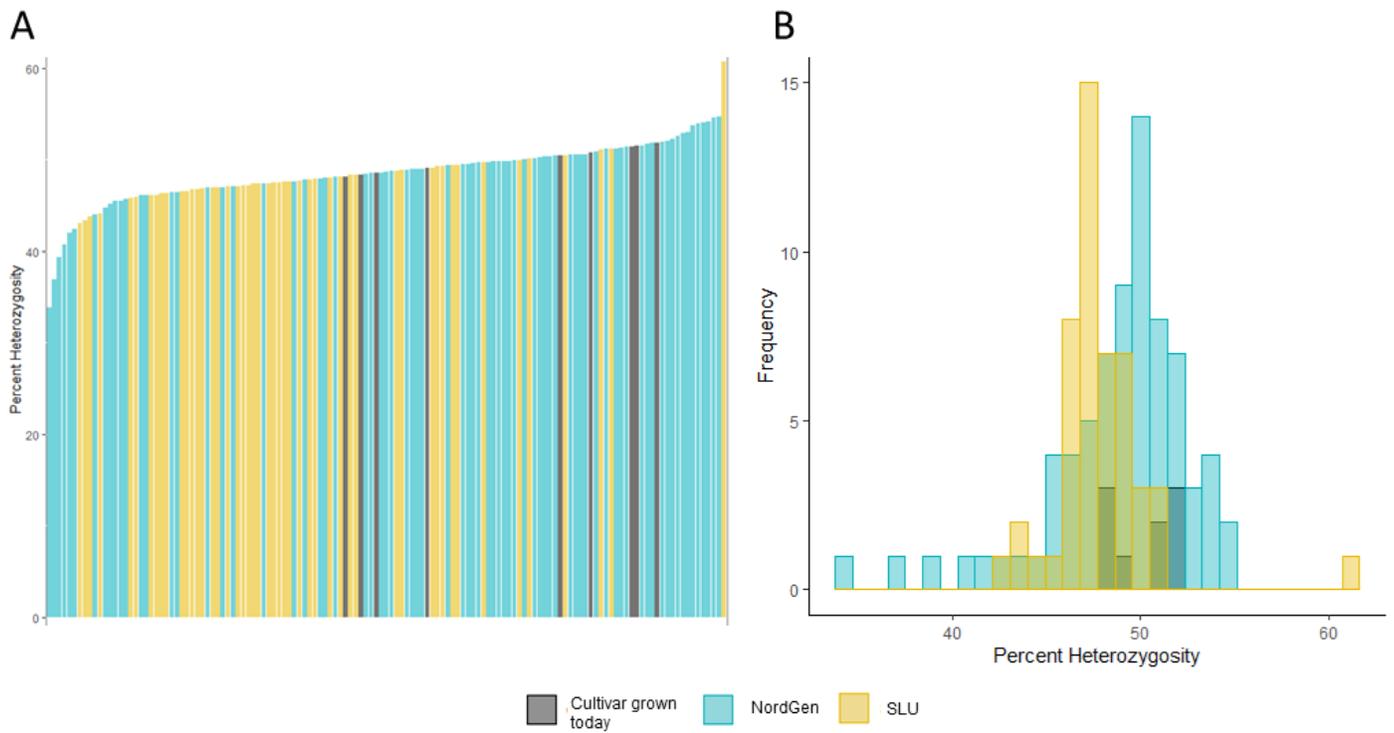
Population structure based on single nucleotide polymorphism (SNP) markers revealed by STRUCTURE (A and C) and heatmap with dendrogram showing the genetic diversity based on Nei's genetic (B) distance among the 133 accessions. The proposed number of subpopulations (K) was determined

according to STRUCTURE as 2 (grey or red), while the assumed number of subpopulations was 3 – cultivar, farmer’s cultivar and breeding clone (A); or cultivar grown today in Sweden, NordGen and SLU (C). The arrows to the heatmap (B) denotes the clusters corresponding to the subpopulations (majority of grey individuals in each biological grouping) proposed by STRUCUTRE.



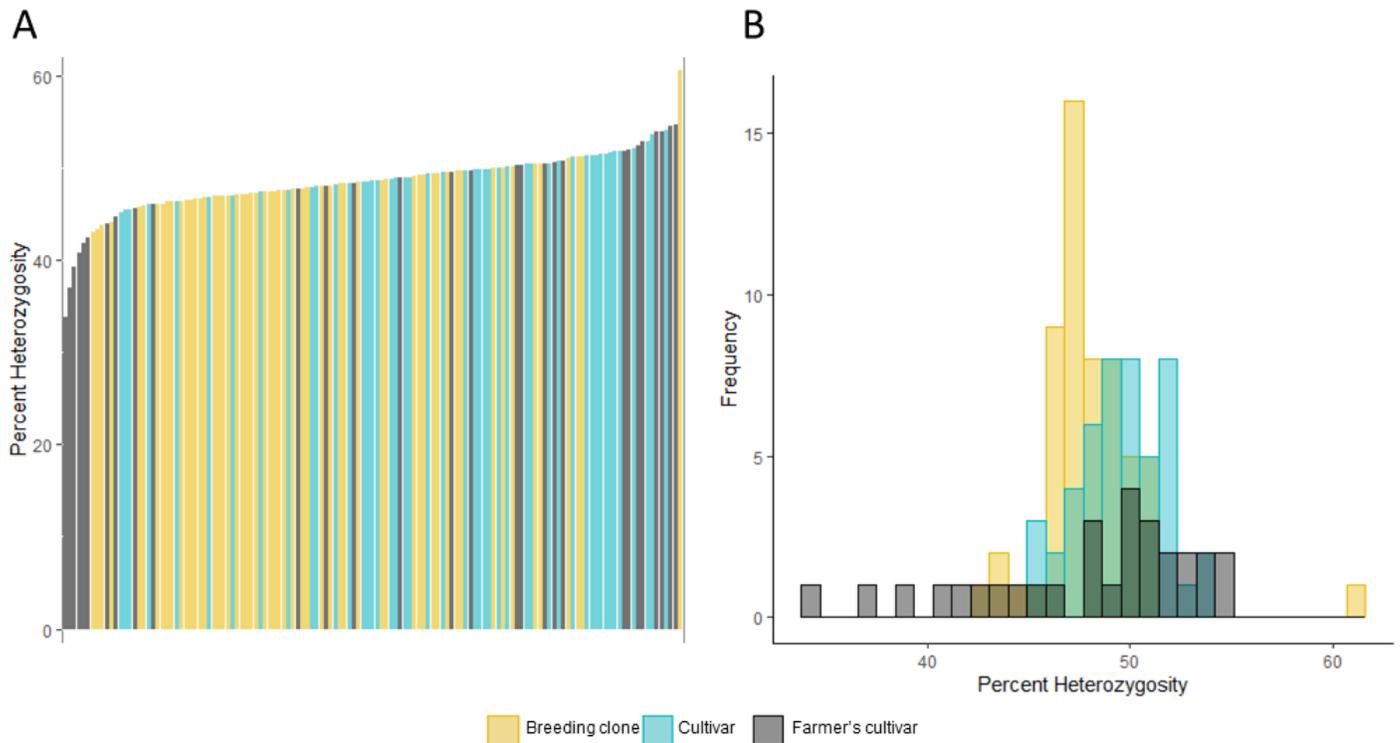
**Figure 5**

Dendrogram based on the Euclidian distance of single nucleotide polymorphism (SNP) markers among the 75 accessions from NordGen. Groupings made as per country of origin (A) or type of clone (B).



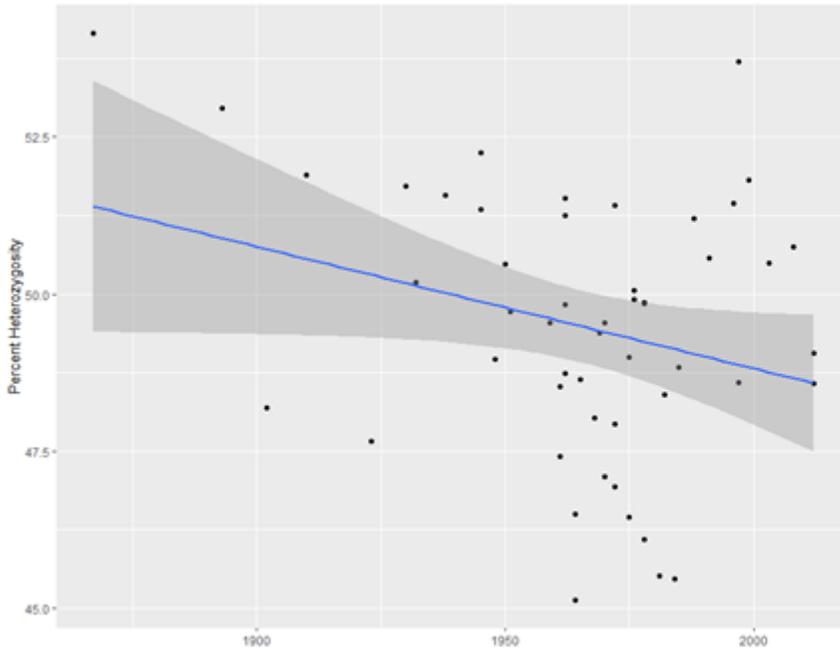
**Figure 6**

(A) Frequency distribution of percent heterozygosity among all 133 accessions and (B) histogram of percent heterozygosity among the three populations.



**Figure 7**

(A) Distribution of percent heterozygosity among all 133 accessions and (B) histogram of percent heterozygosity among the three types of clones.



**Figure 8**

Percent heterozygosity over release year for each of the 51 cultivars with release year information. The negative trend (illustrated by the regression line) is non-significant.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [figureS1catjaselga.tif](#)
- [supplementarytableS1.xlsx](#)