BMC Genomics

Molecular genetic diversity and population structure analyses of rutabaga accessions from Nordic countries as revealed by single nucleotide polymorphism markers

Zhiyu Yu[†], Rudolph Fredua-Agyeman[†], Sheau-Fang Hwang and Stephen E. Strelkov[*]

Abstract

Background: Rutabaga or swede (*Brassica napus ssp. napobrassica* (L.) Hanelt) varies in root and leaf shape and colour, flesh colour, foliage growth habits, maturity date, seed quality parameters, disease resistance and other traits. Despite these morphological differences, no in-depth molecular analyses of genetic diversity have been conducted in this crop. Understanding this diversity is important for conservation and broadening the use of this resource.

Results: This study investigated the genetic diversity within and among 124 rutabaga accessions from five Nordic countries (Norway, Sweden, Finland, Denmark and Iceland) using a 15 K single nucleotide polymorphism (SNP) *Brassica* array. After excluding markers that did not amplify genomic DNA, monomorphic and low coverage site markers, the accessions were analyzedwith 6861 SNP markers. Allelic frequency statistics, including polymorphism information content (PIC), minor allele frequency (MAF) and mean expected heterozygosity ($\overline{H}$e) and population differentiation statistics such as Wright's F-statistics (F_{ST}) and analysis of molecular variance (AMOVA) indicated that the rutabaga accessions from Norway, Sweden, Finland and Denmark were not genetically different from each other. In contrast, accessions from these countries were significantly different from the accessions from Iceland ($P <$ 0.05). Bayesian analysis with the software *STRUCTURE* placed 66.9% of the rutabaga accessions into three to four clusters, while the remaining 33.1% constituted admixtures. Three multivariate analyses: principal coordinate analysis (PCoA), the unweighted pair group method with arithmetic mean (UPGMA) and neighbour-joining (NJ) clustering methods grouped the 124 accessions into four to six subgroups.

Conclusion: Overall, the correlation of the accessions with their geographic origin was very low, except for the accessions from Iceland. Thus, Icelandic rutabaga accessions can offer valuable germplasm for crop improvement.

Keywords: *Brassica*, SNP, AMOVA, Population differentiation, PCoA, UPGMA and NJ

* Correspondence: strelkov@ualberta.ca
[†]Zhiyu Yu and Rudolph Fredua-Agyeman contributed equally to this work.
Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada

Yu *et al. BMC Genomics*   (2021) 22:442

Page 2 of 13

Background

Brassica napus ssp. napobrassica (L.) Hanelt, called 'rutabagge' in Sweden, 'rutabaga' in the USA and Canada, and 'swede' in the UK, New Zealand and Australia, is a cool-weather root crop thought to have been derived from the natural or spontaneous hybridization between *B. rapa* (turnip) and *B. oleracea* (cabbage or kale) [1]. Rutabaga is often assumed to have originated in Sweden, but may have come from Finland [2, 3]. Nevertheless, it was distributed from Sweden (where it grew in the wild before 1400) to England, Germany and other European countries around the end of the eighteenth century [4] and was introduced to North America by European immigrants in the early nineteenth century [5]. Therefore, the Nordic countries are considered the center of rutabaga domestication and diversity.

Rutabagas are grown for use as a table vegetable and as fodder for animals [3]. The roots are rich in vitamins A, C and fibre; are low in calories and have trace amounts of vitamin B_1, B_2, potassium, calcium, magnesium and iron [3, 6]. Like most cruciferous vegetables, they have antioxidant and anti-cancer properties [7]. The leaves have much higher levels of protein (17–18%) than the roots (0.6–2.0%) [8, 9]. However, most of the components are non-protein nitrogen (urea and ammonia), which can be converted into protein by microbes in the stomach of ruminants, but not in pigs [10]. Rutabagas vary considerably in morphology, disease resistance, seed yield and quality parameters such as erucic acid and glucosinolate content [3, 11]. Breeding efforts have targeted root appearance and flesh colour, earliness, drought tolerance, improvement in resistance to diseases, broadening genetic diversity and quality traits associated with the seeds [3, 6, 12–15]. Quantitative traits such as root length, diameter and fresh weight are also of interest for crop improvement [16].

Genetic variation in plants is a key pillar of biodiversity and provides the resources for the development of new and improved cultivars with desirable characteristics [17]. In addition, studying diversity in natural plant populations makes it possible to understand genetic exchange or gene flow within and between populations [18]. Many genetic diversity studies have utilized simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers due to their abundance and co-dominant nature. However, PCR amplification of genomic DNA using SSR markers can produce sequence artifacts because of errors in Taq DNA polymerase activity and the formation of chimeric and heteroduplex molecules [19–21]. The production of artifacts, particularly in the case of highly polymorphic SSR markers, can cause difficulties in allele size calling [22]. Alleles of the same sized products may have different sequences [23]. This can also affect the quality of genotyping data. Random amplified polymorphic DNA (RAPD) markers are dominant markers with low reproducibility and accuracy, while random fragment length polymorphism (RFLP) markers have a low discrimination power and can be costly [24].

In contrast, SNPs arise because of point mutations and hence most SNPs are biallelic, which leads to greater accuracy in genotyping; these markers also offer the advantage of co-dominance. In addition, SNP-based systems lend themselves to automation, and hence a larger number of markers (tens of thousands or higher) can be screened within a shorter time in comparison with the use of SSR markers [25]. The high heritability of SNPs makes them the marker of choice for studying genetic diversity and phylogeny in crop species with ancient genome duplications such as *B. napus* [26]. A major drawback is that SNP calling is difficult for polyploid species such as *B. napus* [25]. In addition, SNP markers used for genetic diversity studies should be neutral or be present in non-coding regions to eliminate bias introduced by selection when inferring population structure. Therefore, SNP arrays used for genotyping require extensive validation to confirm their usefulness for general application. Genome resequencing is an alternative to array-based methods and generally yields over a million SNP markers [27–30].

Previous molecular studies indicated that spring oilseed rape, winter oilseed rape, fodder and vegetable types, and rutabagas formed separate clusters of *B. napus* [31–33]. Bus et al. [31] used 89 SSR markers to estimate genetic diversity in 509 *B. napus* inbred lines, of which 73 were swedes or rutabagas. Similarly, Diers and Osborn [32] used 43 RFLP markers to group 83 *B. napus* lines including two rutabagas. Mailer et al. [33] reported that a set of 100 RAPD markers could identify four rutabaga accessions among 23 cultivars of *B. napus*. Zhou et al. [27] used 30,877 SNP markers to differentiate 300 *Brassica* accessions into spring, semi-winter and winter ecotypes. Gazave et al. [28] genotyped 782 *B. napus* accessions with 30,881 high quality SNP markers and reported three major subpopulations, of which the highest variance was found in the spring and winter samples. Whole genome sequencing has indicated that winter oilseeds, which include rutabagas, may be the original form of *B. napus* and that this crop may have multiple origins [29, 30].

One hundred seventy-one rutabaga accessions are available (assessed on January 11th, 2021) from the Nordic Genetic Resource Center, Alnarp, Sweden. Of these, 145 accessions are from the Nordic countries, 20 are from France, four are from Germany and one accession each is from Estonia and the United Kingdom. Many of these are landraces with great genetic variability that can be exploited in rutabaga and other *Brassica*

Yu *et al. BMC Genomics*   (2021) 22:442

Page 3 of 13

breeding programs around the world. The genetic diversity and variability that exist within and among rutabaga accessions and populations from the Nordic countries have not been examined. Understanding this diversity is important for conservation and broadening the use of this important resource. Therefore, the aim of the present study was to use high-throughput genotyping with *Brassica* SNP markers to estimate genetic diversity in rutabaga accessions from five Nordic countries (Norway, Sweden, Finland, Denmark and Iceland).

Results

SNP marker characteristics

Thirteen thousand seven hundred four SNP markers on the 15 K SNP *Brassica* chip were used to screen the 124 rutabaga accessions and three rutabaga cultivars. Among these, 31% (4213 SNPs) were monomorphic, 5% (701 SNPs) were low coverage site markers, and 14% (1929 SNPs) were missing data points for > 5% of the accessions. Thus, filtering removed ≈ 50% of the SNP markers, while the remaining ≈ 50% (6861 SNPs) were retained for the diversity analysis. This comprised 4390 A-genome and 2471 C-genome SNP markers.

Allelic patterns and genetic diversity indices among and within populations

Figure 1 shows the origin and sample sizes of the rutabaga accessions used for this study. Allelic patterns and genetic diversity summary statistics at any given locus or averaged across the 6861 SNP loci for the rutabaga accessions separately for each country and for the whole collection are presented in Table S1 and Fig. 2A to D.

The proportion of polymorphic loci (%P) detected separately for the NOR-, SWE-, FIN- and DNK- subpopulations was significantly higher (range 88.5–99.6%) than for the ISL-subpopulation (67.9%) ($P < 0.05$) (Table S1). The mean number of alleles per locus (Na) was highest in the SWE-subpopulation (2.236 ± 0.005) and lowest in the ISL-subpopulation (1.707 ± 0.006) (Table S1). Similarly, the mean number of effective alleles per locus (Ne) and Shannon's information index (I) were significantly higher in the SWE-subpopulation (1.590 ± 0.004 and 0.535 ± 0.002, respectively) compared with the ISL-subpopulation (1.299 ± 0.004 and 0.305 ± 0.003, respectively) (Table S1). In addition, the mean number of alleles with a frequency ≥ 5% (Na Freq ≥ 5%) and mean number of common alleles found in ≤50% of the subpopulations (Na common ≤ 50%) were lowest for the ISL-subpopulation (Fig. 2A). Thus, most of the genetic diversity indices for the NOR-, SWE-, FIN- and DNK-subpopulations were not significantly different from each other. They were, however, all significantly different from the ISL-subpopulation ($P < 0.05$).

The diversity of the SNP markers expressed as the polymorphic information content (PIC) is presented in Fig. 2B. The number of markers with PIC > 0.2 was
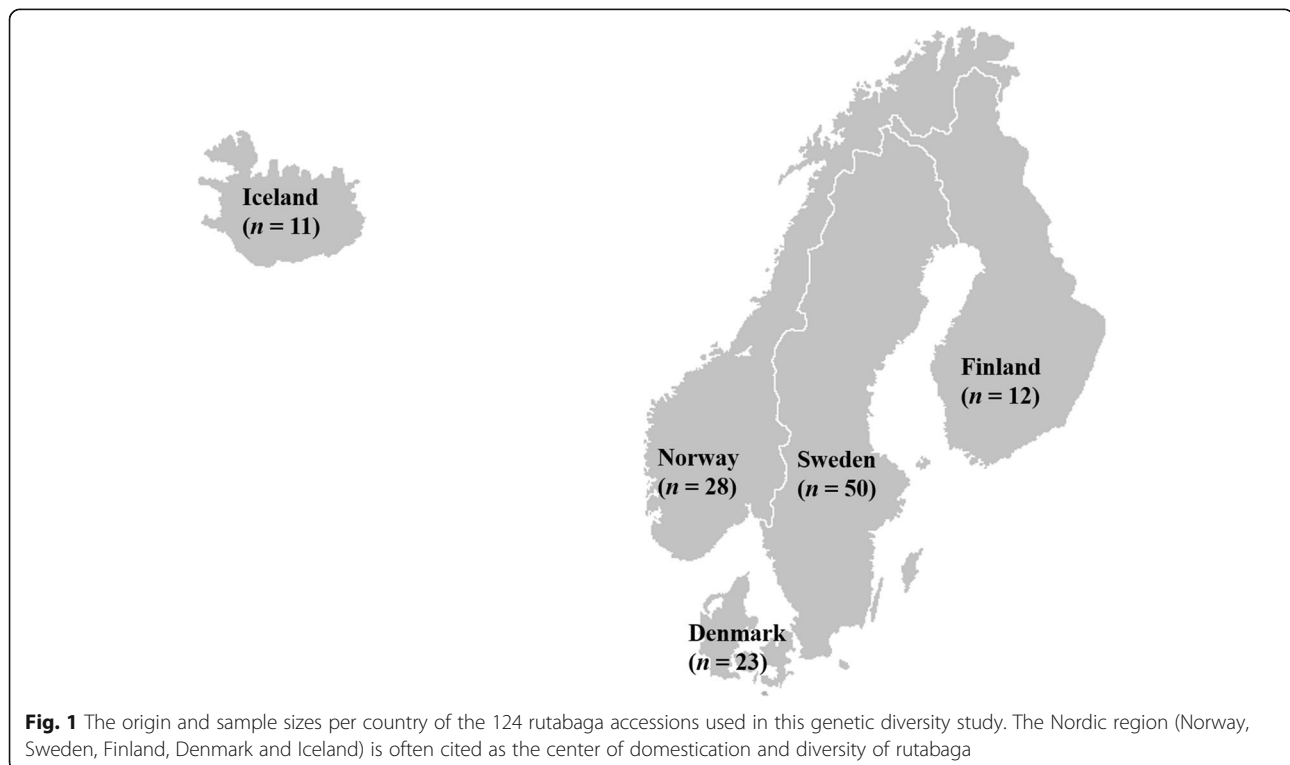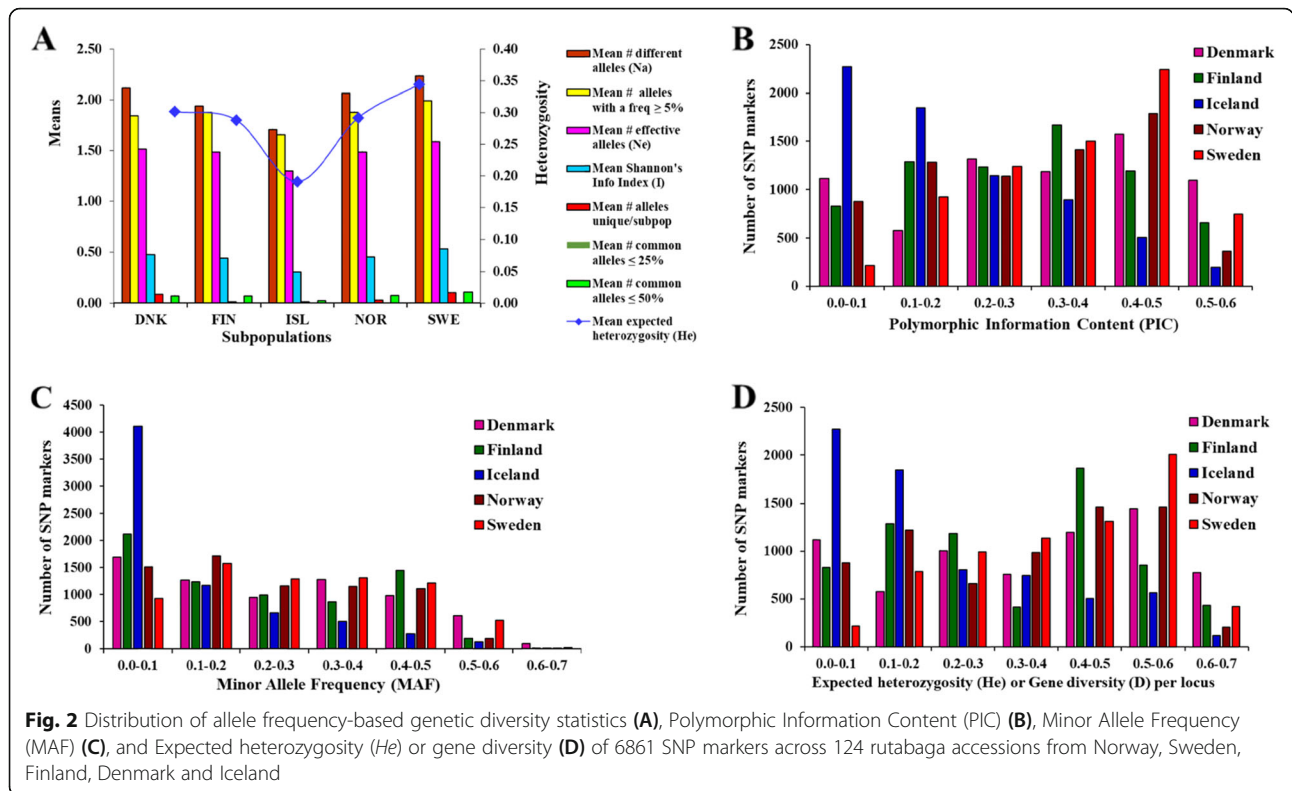

Fig. 1 The origin and sample sizes per country of the 124 rutabaga accessions used in this genetic diversity study. The Nordic region (Norway, Sweden, Finland, Denmark and Iceland) is often cited as the center of domestication and diversity of rutabaga

Fig. 2 Distribution of allele frequency-based genetic diversity statistics **(A)**, Polymorphic Information Content (PIC) **(B)**, Minor Allele Frequency (MAF) **(C)**, and Expected heterozygosity (*He*) or gene diversity **(D)** of 6861 SNP markers across 124 rutabaga accessions from Norway, Sweden, Finland, Denmark and Iceland

highest for the SWE-subpopulation (5725 ≈ 83%) and DNK-subpopulation (5170 ≈ 75%), intermediate for the FIN- and NOR-subpopulations (4701–4726 ≈ 69%), and lowest among for the ISL-subpopulation (2742 ≈ 40%). The PIC averaged across the 6861 SNPs separately for each population followed similar patterns as the allelic and genetic diversity, with the highest PIC occurring in the SWE-subpopulation (0.35) and the lowest in the ISL-subpopulation (0.18).

The number of SNP markers with minor allele frequency (MAF) ≤ 0.1 was of the order ISL- (4106 ≈ 60%) > FIN- (2115 ≈ 31%) > DNK- (1690 ≈ 25%) > NOR- (1518 ≈ 22%) > SWE-subpopulations (933 ≈ 14%). Thus, the frequency of minor alleles was highest for the ISL-subpopulation, intermediate for the FIN-, DEN- and NOR-subpopulations, and lowest for the SWE-subpopulation (Fig. 2C).

The expected heterozygosity per locus (H_e), also called gene diversity (*D*), followed similar patterns as the rest of the parameters measured with the exception of the MAF (Fig. 2D). Analyses of the gene pool structure ($\overline{H}_e$, expected heterozygosity averaged over all 6861 loci) of the rutabaga accessions from each country suggested that there was no significant difference in the genetic variability of the rutabaga accessions from Sweden (0.345 ± 0.002), Denmark (0.301 ± 0.002), Norway (0.292 ± 0.002), and Finland (0.288 ± 0.002). These accessions

were, however, genetically different from the accessions from Iceland (0.191 ± 0.002) (Table S1).

Genetic differentiation among regions, populations and within accessions

Pairwise comparisons of population differentiation using the fixation statistics index (F_{ST}) are presented in Table 1. The F_{ST} values for all 10 pairwise combinations of all five subpopulations ranged from 0.032 to 0.133. Pairwise F_{ST} values for NOR/SWE, NOR/FIN and SWE/FIN ranged from 0.032 to 0.067 (lowest); the values for NOR/DNK, SWE/DNK and FIN/DNK ranged from 0.050 to 0.88 (intermediate); whereas the F_{ST} values for the ISL/NOR, ISL/SWE, ISL/DNK and ISL/FIN ranged from 0.103 to

Table 1 Pairwise correlation of the fixation index or F_{ST} values between subpopulations of rutabaga accessions from Denmark, Finland, Iceland, Norway and Sweden

	DNK	FIN	ISL	NOR	SWE
DNK	0.000				
FIN	0.088	0.000			
ISL	0.133	0.124	0.000		
NOR	0.067	0.067	0.103	0.000	
SWE	0.050	0.032	0.106	0.042	0.000

F_{ST} values between subpopulations ; *DNK* Denmark, *FIN* Finland, *ISL* Iceland, *NOR* Norway, *SWE* Sweden

0.133 (highest). Overall, the lowest F_{ST} value was found between the SWE- and FIN-subpopulations and the highest between the ISL- and DNK-subpopulations (Table 1).

The analysis of molecular variance (AMOVA) of the distance matrices obtained with TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) and GenAlEx software for the rutabaga accessions were highly correlated (Tables S2a and S2b). The AMOVA among and within the five populations partitioned the overall genetic variance into three parts: ≈ 94% attributable to within population differences, whereas ≈ 5% and ≈ 1% of the variation occurred among populations and among regions, respectively ($P = 0.108$) (Fig. 3A). This suggested only minor differences in the entire rutabaga populations from the different countries.

Pairwise comparison of the AMOVA (Φ_{PT}) between the populations, however, revealed a higher genetic variance (18 to 27%) between the ISL-subpopulation and the NOR-, SWE-, FIN- and DNK-subpopulations (Table 2). Furthermore, the rutabaga accessions from Iceland and Denmark were the most genetically diverse ($\Phi_{PT} = 27\%$), followed by accessions from Iceland and Finland ($\Phi_{PT} =$

24%). In contrast, rutabaga accessions from Sweden and Finland were the most similar ($\Phi_{PT} = 2\%$) followed by accessions from Norway and Sweden ($\Phi_{PT} = 7\%$).

Thus, the vast majority of the genetic variability could be attributed to within population differences. Nevertheless, the pairwise comparison of the subpopulations suggested that considerable variation existed between the rutabagas from the different countries.

Cluster analyses

The principal coordinate analysis (PCoA) based on the 6861 SNP markers clustered the 124 rutabaga accessions into six heterogeneous subgroups (Fig. 3B) using the first (PCoA1 ≈ 14.7% of genetic variance) and second (PCoA2 ≈ 11.4% of genetic variance) principal coordinates. Clearly, the rutabaga accessions from Sweden, Norway and Finland were distributed across almost all of the subgroups (P1 to P6 in Fig. 3B). In contrast, the accessions from Iceland and Denmark were concentrated in subgroup P3 and subgroups P1 and P2, respectively (Fig. 3B).
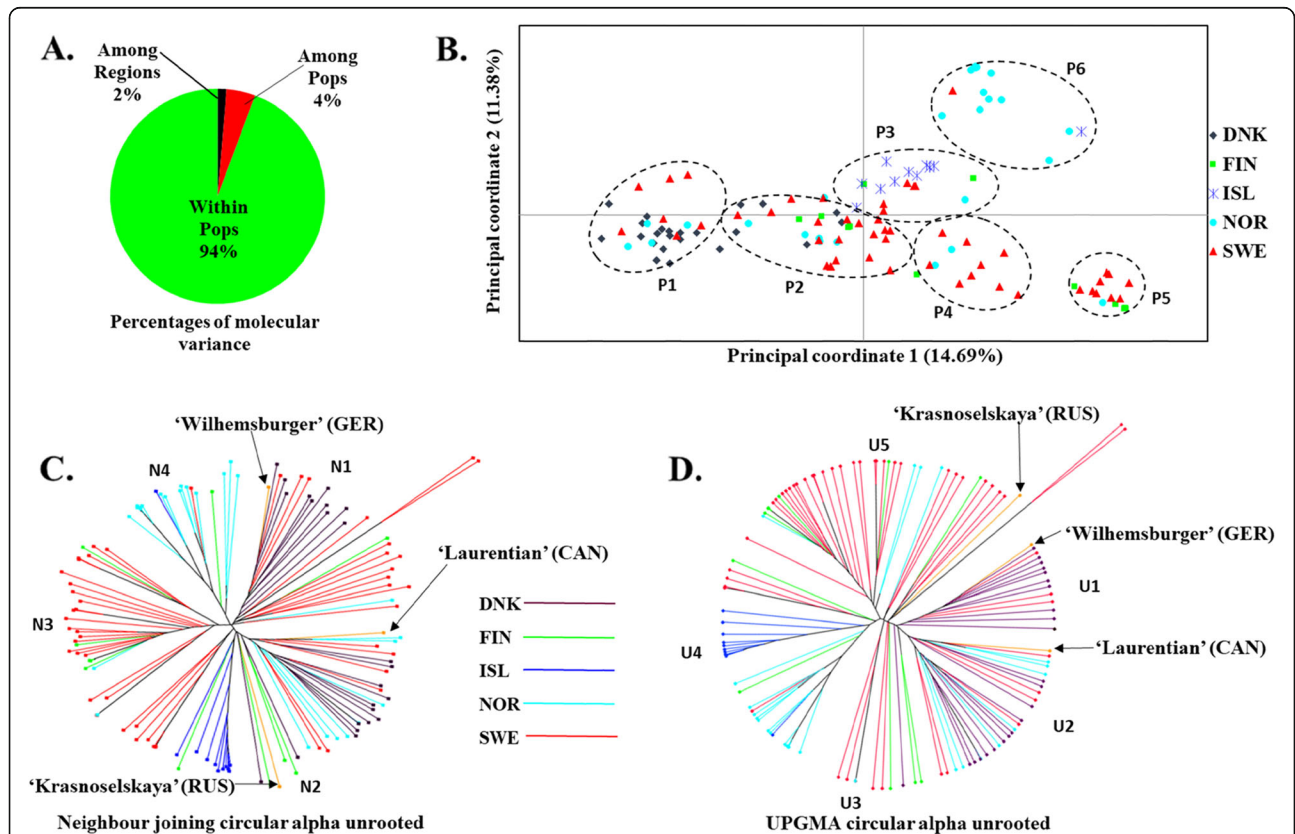

Fig. 3 Analysis of molecular variance (AMOVA) partitioning of molecular variance among regions, populations and within accessions **(A)**. Principal coordinates analysis (PCoA) **(B)**, Neighbour joining (NJ) **(C)**, and Unweighted pair group method with arithmetic mean (UPGMA) **(D)** analyses with 6861 SNP markers grouped the 124 rutabaga accessions from Norway, Sweden, Finland, Denmark and Iceland into 6, 4 and 5 subgroups, respectively. The positions of the three out-groups, Laurentian (CAN), Wilhemsburger (GER) and Krasnoselskaya (RUS), are indicated on the NJ and the UPGMA trees

Yu *et al. BMC Genomics*   (2021) 22:442

Page 6 of 13

Table 2 Pairwise comparison between population genetic variance of 124 rutabaga accessions from Denmark, Finland, Iceland, Norway and Sweden

	DNK	FIN	ISL	NOR	SWE
DNK	–				
FIN	16%	–			
ISL	27%	24%	–		
NOR	14%	12%	21%	–	
SWE	9%	2%	18%	7%	–

Values indicate genetic variance between populations
DNK Denmark, *FIN* Finland, *ISL* Iceland, *NOR* Norway, *SWE* Sweden

The neighbour-joining (NJ) based on the 6861 SNP markers clustered the 124 rutabaga accessions into four major branches (Fig. 3C). The unrooted phylogenetic trees indicated that the accessions from Sweden were distributed into three of the branches (N1, N2 and N3), those from Norway, Finland and Denmark were segregated into two of the branches (N2 and N4, N2 and N3 and N1 and N2, respectively), whereas accessions from Iceland were concentrated in one branch (N2) (Fig. 3C).

The unweighted pair group method with arithmetic mean (UPGMA) based on the 6861 SNP markers indicated that the trees for the 124 rutabaga accessions were clustered into five major branches (Fig. 3D). The accessions from Sweden, Norway, and Finland were widely distributed across at least four of the major branches (Fig. 3D). Similar to the branching patterns in the NJ analysis, the rutabaga accessions from Denmark and Iceland clustered into two branches (U1 and U2) or one branch (U4), respectively.

Overall, the three multivariate analyses (PCoA + NJ + UPGMA) suggested the existence of four to six groups in the rutabaga accessions. However, correlations with their geographic origin were very low, except for the accessions from Iceland.

The unrooted trees used to depict the NJ and UPGMA do not imply a known ancestral root of the three outgroups (which are coloured orange in Fig. 3C and D). However, the results suggested that the rutabaga 'Wilhemsburger' was in the first branch (N1 of the NJ unrooted tree), while 'Laurentian' and 'Krasnoselskaya' both were grouped in the second branch (N2 of the NJ unrooted tree) (Fig. 3C). In the case of the unrooted tree used to depict UPGMA, 'Wilhemsburger', 'Laurentian' and 'Krasnoselskaya' were grouped in the first (U1), second (U2) and fifth branch (U5), respectively (Fig. 3D).

The NJ and UPGMA representation of the similarity matrices as a phylogram (Figs. S1a and S1b) and a circular rooted (Figs. S2a and S2b) diagram are included in the Supplementary Materials. These indicate even closer groupings of the accessions based on their geographic origins.

Bayesian population structure analysis

The *STRUCTURE* analysis was run 11 times with the accessions unassigned and 11 times with the accessions assigned to their respective countries of origin. Table 3 summarizes the *STRUCTURE* results used to infer the

Table 3 Determination of the number of cluster sets in 124 rutabaga accessions from Denmark, Finland, Iceland, Norway and Sweden using the Evanno et al. (2005) and Puechmaille et al. (2016) methods

Structure ran #	Burn-in lengths	MCMC* lengths	Number of clusters (K)	Number of Reps	Number of populations[a]		Number of Populations[β]			
					ΔK (Unassigned)[a]	ΔK (Assigned)[b]	MedMedK	MedMeaK	MaxMedK	MaxMeaK
1	5000	5000	10	10	9	8	3	4	4	4
2	5000	5000	10	20	8	8	3	4	4	4
3	10000	10000	10	10	9	8	3	4	4	4
4	10000	10000	10	20	2	9	3	4	4	4
5	20000	20000	10	10	8	2	3	4	4	4
6	20000	50000	10	10	9	2	3	4	4	4
7	50000	50000	10	10	3	3	3	4	4	4
8	10000	100000	10	10	8	6	3	4	4	4
9	20000	100000	10	10	2	9	3	4	4	4
10	50000	100000	10	10	9	2	3	4	4	4
11	100000	100000	10	10	3	9	3	4	4	4

*$*MCMC$ Markov Chain Monte Carlo
[a] The ad hoc ΔK method (Evanno et al. 2005); [a] Accessions unassigned to any population or country; [b] Accessions assigned to their countries of origin
[β] The median (MedMedK and MaxMedK) or mean (MedMeaK and MaxMeaK) estimators used to determine which subpopulations belonged to a cluster (K) (Puechmaille et al. 2016)

Yu *et al. BMC Genomics*   (2021) 22:442

Page 7 of 13

population genetic structure of the rutabaga accessions from the Nordic countries. The number of clusters (K) determined following the method of Evanno et al. [34] indicated ΔK statistic values of $K = 2$ to 9, while the four alternative statistics (MedMedK, MedMeaK, MaxMedK and MaxMeaK) determined following Puechmaille [35] and Li and Liu [36] indicated 3 to 4 clusters (Table 3). Increasing the number of replications from 10 to 20 produced cluster numbers similar to the above. These suggested that the Puechmaille [35] and Li and Liu [36] method was more consistent than the Evanno et al. [34] method for inferring the population genetic structure of the rutabaga accessions from the Nordic countries. Based on the ΔK statistic values, there was no significant difference in *STRUCTURE* run # 1, 2, 3 and 8 for analysis done with the accessions unassigned and for analysis with the accessions assigned to their respective countries of origin. In contrast, significant differences were found for *STRUCTURE* run # 4, 5, 6, 9, 10 and 11. The two methods produced approximately the same number of clusters ($K = 3$ to 4) at Burn-in and MCMC lengths each of 50,000 and at $K = 1–10$ and for 10 replicates (i.e. run #7) (Table 3).

Plots of MedMedK, MedMeaK, MaxMedK and MaxMeaK as well as log-likelihood (*lnK*) against the number of clusters suggested the presence of subpopulations in the accessions (Fig. 4 and S3). Based on a threshold for similarity score of 70%, 66.1% of the accessions were placed into one of the three clusters while 33.9% were classified as admixtures (Table 4). Excluding the admixture, 91.3% of the accessions from Denmark and 72.7% of the accessions from Iceland were present in only one cluster (1 and 2, respectively). In contrast, 58.3% of the accessions from Finland and 42.0% of the accessions from Sweden were present in clusters 1 and 3, while 75.0% of the accessions from Norway were present in clusters 1 and 2 (Table 4). The German rutabaga 'Wilhemsburger' was placed in cluster 1 along with some of the accessions from Denmark, Finland, Norway and Sweden. The Canadian rutabaga 'Laurentian' and the Russian rutabaga 'Krasnoselskaya' were admixtures. Overall, the number of clusters (3 to 4) obtained in the *STRUCTURE* analysis with the Puechmaille [35] and Li and Liu [36] method was consistent and comparable with the 4–6 subgroups obtained in the multivariate analysis. In contrast, the number of clusters determined following Evanno et al. [34] were not consistent and varied widely.

Clustering of genotypes with similar names
The NJ, UPGMA and *STRUCTURE* analyses placed the majority of the accessions with similar names but with different accession numbers into the same cluster, irrespective of their countries of origin. For example, the three analyses placed all six 'Wilhemsburger' accessions (FGRA112D, FGRA107D, FGRA108D, FGRA110D, FGRA106D and FGRA109D) in the same cluster as 'Wilhemsburger' from Germany, which was used as an outgroup (Fig. 4D and S2). Similarly, the NJ and UPGMA analyses placed all six (FGRA120S, FGRA118S, FGRA121S, FGRA119S, and FGRA117S) 'Östgota' accessions into one group (Fig. S2), while the *STRUCTURE* analysis placed five of the six into one group (except FGRA116S) (Fig. 4D). In the case of 'Bangholm' accessions, both NJ and UPGMA captured 13 of the 16 accessions into one group, while the remaining three accessions (FGRA 003, FGRA011 and FGRA008) were placed into two groups (Fig. S2). The *STRUCTURE* analysis placed 15 of the 16 'Bangholm' accessions (except FGRA008) in the same cluster (Fig. 4D). Therefore, the clustering of the rutabaga accessions using NJ, UPGMA and *STRUCTURE* analyses was very consistent.

Discussion
A comprehensive body of literature exists on rutabagas in the main Nordic languages (Personal communication, Prof. Ann-Charlotte Wallenhammar, Swedish University of Agricultural Sciences). This probably reflects the transmission of seeds and information on agronomic practices for rutabaga cultivation in the Nordic region since medieval times [4]. Turesson [40–42] observed that when the same species of plants were grown in different habitats over many years, they differed from each other in stature, colour, morphology and texture of leaves, stem, flowers and seed. Consequently, rutabagas that are adapted to different climatic and geographic environments will develop different morphological traits.

In this study, SNP markers and combinations of allele- and distance-based population genetics statistics, multivariate clustering and Bayesian methods were used to examine genetic diversity and differentiation in rutabaga accessions from Norway, Sweden, Finland and Denmark and Iceland. Diers and Osborn [32] used rutabaga accessions as an out-group in genetic diversity studies of *B. napus*, whereas Mailer et al. [33] and Bus et al. [31] compared rutabagas with spring oilseed rape, winter oilseed rape, fodder and vegetable types. Fewer than 100 SSR, RFLP and RAPD markers, however, were used in those studies compared with the 6861 SNP markers in the current study. In contrast, Gazave et al. [28] and Zhou et al. [27] identified 1,081,925 and 1,197,282 SNP markers using an Illumina Hiseq single-end sequencing and Specific-Locus Amplified Fragment sequencing (SLAF-Seq), respectively. Similarly, An et al. [29] and Lu et al. [30] obtained 372,546 and 675,457 high-quality SNPs by RNA-sequencing, respectively. The four studies used over 30,000 SNP markers for genetic structure analysis, which is $\approx 4 \times$ the 6861 markers used in our study.
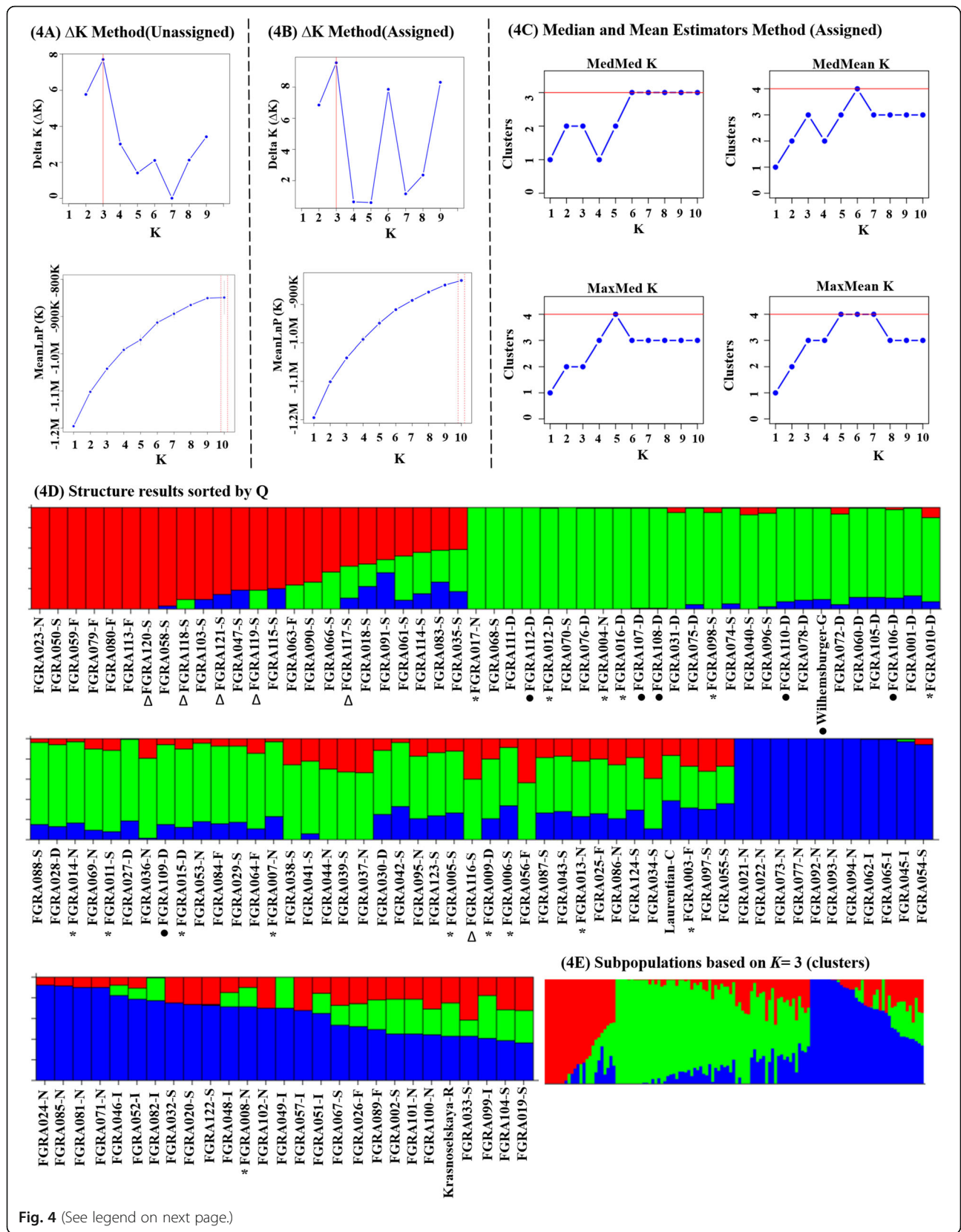
Fig. 4 (See legend on next page.)

Yu *et al. BMC Genomics*   (2021) 22:442

Page 9 of 13

(See figure on previous page.)

Fig. 4 Bayesian cluster analysis of 124 rutabaga accessions from Norway, Sweden, Finland, Denmark and Iceland estimated using the software *STRUCTURE* based on 6861 SNP markers. The optimal value of *K*, determined by the method of Evanno et al. (48) with populations unassigned **(A)** or assigned **(B)** to their respective countries, as well as by the method of Puechmaille [37] and Li and Liu [38] **(C)**, suggested that the 124 rutabaga accessions could be placed into 3 or 4 clusters (*K* = 3 most likely). Detailed Bayesian clustering of 124 rutabaga accessions by the CLUMPAK program [39] **(D-E)**. Each column or rectangular bar represents the individual rutabaga accessions used in this study, while each colour represents one gene pool and the stacked bars with different colours represent admixtures with their shared ancestry components **(D)**. Simplified view suggests three ancestral populations **(E)**

In general, the use of more markers improves the accuracy of subpopulation clustering. Moreover, DNA sequencing can provide information on gene-rich regions and is particularly useful for species with limited or no genome information. Nonetheless, DNA sequencing is more costly than the use of array-based SNP markers, although extensive validation of the SNP markers is needed before they can be used for genetic diversity studies.

The average of 2.012 ± 0.003 alleles per SNP locus obtained in this study was less than the 4.78 alleles per SSR locus found by Bus et al. [31]. This was expected, since SSR markers are multi-allelic codominant markers while SNP markers are often bi-allelic. The mean expected heterozygosity ($\overline{H}$ e) or gene diversity (D) of 0.283 ± 0.001 obtained in this study was less than the 0.43 reported by Bus et al. [31], likely because the 73 accessions examined in the latter represented a more diverse collection from 19 countries with a much wider geographical distribution (Europe, North America, Asia, New Zealand and the North Africa). The lower allelic diversity summary statistics with SNP markers compared with SSR markers also has been reported in rice [43, 44], barley [45], mushrooms [46] and other species. An et al. [29] reported SNP density of 0.286–1.080 per kb. The SNP density in this study ranged from 0.009 to 0.184 kb per chromosome [12]. This is a reflection of the higher numbers of SNPs detected by genome resequencing methods.

The pairwise fixation index (F_{ST}) obtained in the current study ranged from 0.032 to 0.133, which was within the 0.054 reported by Bus et al. [31]. Therefore, the use of SNP and SSR markers confirmed that genetic differentiation in the rutabaga accessions is low and there is a high degree of genetic exchange within these accessions. The observation in earlier studies that rutabagas clustered separately from spring, winter, fodder and vegetable *Brassica* species [31–33] could be due to many rutabagas being landraces with different morphological adaptions to various geographic and climatic regions. The significantly higher pairwise F_{ST} values for ISL-DNK (0.133), ISL-FIN (0.124), ISL-SWE (0.106) and ISL-NOR (0.103) compared with the pairwise F_{ST} values for DNK, FIN, NOR and SWE (range 0.032 to 0.088) may reflect enrichment caused by lack of mating between the Icelandic sub-population and the rest of the subpopulations. As an island in the North Atlantic, Iceland is geographically isolated from the other Nordic countries. This isolation, combined with possible differences in microclimatic and soil conditions, may have resulted in less exchange of germplasm between Iceland and Norway, Denmark, Finland and Sweden.

The hierarchical clustering of the genotype data by the use of NJ, UPGMA and PCoA analyses yielded 4 to 6 subgroups in the rutabaga accessions. Previous studies [35, 47] reported that the *STRUCTURE* program did not reliably identify the main clusters within a subpopulation. Therefore, in this study, many runs of the *STRUCT*

Table 4 Inferred ancestry of 124 rutabaga accessions from Denmark, Finland, Iceland, Norway and Sweden based on membership coefficients

Cluster (*K*)	Origin of rutabaga[α]					Total Number	Ave genetic Distance within subgroup[β]
	DNK (23)	FIN (12)	ISL (11)	NOR (28)	SWE (50)		
1[*]	21	2	0	8	11	42	0.1805
2	0	0	8	13	4	25	0.4218
3	0	5	0	1	10	16	0.2245
Admixture[†♣]	2	5	3	6	25	41	
In-Cluster	91.3%	58.3%	72.7%	78.6%	50.0%	66.9%	
Admixture	8.7%	41.7%	27.3%	21.4%	50.0%	33.1%	

[α] Sample sizes of the Accessions were assigned to a specific cluster (*K*) if *P* ≥ 0.70 and those that did not meet this threshold were considered as admixture
Placement of out-groups:[*]'Wilhemsburger' (Germany) = *K*1 and [†]'Laurentian' (Canada) and [♣] 'Krasnoselskaya' (Russia) = Admixture
[β]The average genetic distance within each cluster or subgroup was obtained from the *STRUCTURE* results

URE program were carried out to obtain convergence before the 'best' number of clusters was determined. In addition, we compared the number of inferred clusters from the alternative statistics, MedMedK, MedMeaK, MaxMedK and MaxMeaK [35] to the ΔK method [34] to decide the 'best' number of clusters. The possible number of clusters was 2, 3, 4, 6, 8 or 9, with the 'best' or most consistent of the two methods being 3 to 4. The variable numbers of clusters in the PCoA, UPGMA, NJ and *STRUCTURE* analyses could be due to the low levels of genetic variability and the high genetic exchange among rutabaga accessions from the Nordic countries. This was confirmed further by the high degree (33.1%) of admixtures detected by the Bayesian population structure analyses. Inspection of clusters obtained by the use of the four genetic structure analysis methods showed that the majority (90–100%) of accessions with similar names were in the same group irrespective of their countries of origin. This suggests that the 6861 SNP markers were able to reliably determine the number of clusters in the 124 rutabaga accessions from the Nordic countries. The population structure analyses were in agreement with the allele diversity summary statistics obtained in this study.

Conclusion
Our results showed that the majority of the genetic differences in rutabaga accessions from the Nordic countries were present within the Icelandic subpopulations, while accessions from Norway, Sweden, Finland and Denmark were genetically very similar. Given these findings, based on molecular genetics analyses, there may be value in additional and more detailed study of the morphological traits of accessions originating from these different countries. Lastly, these rutabaga accessions should be of great interest to breeders for increasing genetic diversity in *B. napus*.

Materials and methods
Plant material
One-hundred twenty-four rutabaga accessions collected from the Nordic Genetic Resource Center and used for genome wide association studies by Fredua-Agyeman et al. (12) were included in this study. These consisted of 23 accessions from Denmark, 12 from Finland, 11 from Iceland, 28 from Norway and 50 accessions from Sweden (Fig. 1). Hereafter, the rutabaga accessions from the five countries will be referred to as the DNK-, FIN-, ISL-, NOR- and SWE-subpopulations, respectively. In addition, seeds of three commercial rutabaga cultivars, 'Laurentian' from Canada, 'Wilhemsburger' from Germany and 'Krasnoselskaya' from Russia, were included as the out-group. Details on the accessions are presented in Table S3. Two to four seeds of each accession were grown in 13 × 13 × 15 cm pots filled with Sunshine Mix #4 Aggregate Plus Growing Mix (Sungro Horticulture Canada Ltd) and kept in a growth chamber with a 16 h/8 h (22 °C) day/night cycle for 4 weeks. Leaf tissue (~ 0.25 g) was collected from two plants of each accession in 1.5 mL microcentrifuge tubes on ice. The samples were stored at – 20 °C and shipped on dry ice for SNP genotyping.

SNP genotyping and filtering
SNP genotyping of the 124 rutabaga accessions and three commercial cultivars was performed with a 15 K SNP *Brassica* array at TraitGenetics GmbH, Gatersleben, Germany. These markers were part of 34,248 A and C genome-specific and polymorphic SNP markers identified during an analysis of 432 *B. napus* and ancestral diploid genotypes by Clarke et al. [25]. The majority of the SNP assays targeted single loci within the genome to limit the impact of genome duplication and to simplify SNP calling [25]. The SNP marker positions were mapped to the A genome of *B. rapa* [48], the C genome of *B. oleracea* [49] and the A and C genomes of *B. napus* [50]. The majority of the SNP loci were found in non-coding regions [27].

After genotyping, filtering was done to remove monomorphic and low coverage site SNP markers, those with minor allele frequency (MAF) < 0.05, and SNPs with missing data for > 5% of the accessions. Six thousand eight hundred sixty-one SNP markers were retained for the calculation of the genetic diversity indices and the population structure analyses. These comprised 4390 A-genome and 2471 C-genome SNP markers distributed across all 19 chromosomes of *B. napus* [12].

Allele frequency-based population structure analyses
The proportion of polymorphic loci (%P), the mean number of alleles per locus (Na), the mean number of effective alleles per locus (Ne), the mean expected heterozygosity ($\overline{H}_e$), the mean unbiased expected heterozygosity ($U\overline{H}_e$), the mean number of alleles with a frequency ≥ 5% (Na Freq ≥ 5%), mean number of common alleles found in ≤25% and ≤ 50% of the subpopulations (Na comm ≤ 25% and Na comm ≤ 50%; respectively) and Shannon's information index (I), within and among subpopulations, as well as Wright's [51] genetic differentiation F-statistics (F_{ST}) between the populations were determined with GenAlEx 6.5 [37, 52]. The F_{ST} values were assessed at 1000 random permutations across the 6861 loci.

In addition, the polymorphism information content (PIC), minor allele frequency (MAF) and the expected heterozygosity at any given locus (He) also called gene diversity (D) [38] were evaluated for the DNK-, FIN-, ISL-, NOR- and SWE-subpopulations and the entire population using POWERMAKER v3.25 [39].

Yu *et al. BMC Genomics*   (2021) 22:442

Page 11 of 13

Distance-based population structure analyses

The genetic and similarity distance matrices within and among the subpopulations were calculated for the 6861 SNP markers and the 124 accessions using both GenAlEx 6.5 [37, 52] and TASSEL v5.2.2.5 [53].

The matrices were used to test the hierarchical partitioning of the analysis of molecular variance (AMOVA) among regions, populations and within accessions and their level of statistical significance was assessed based on 10,000 permutations [54]. In addition, patterns in the population were inferred or visualized by principal coordinates analysis (PCoA) [55]. The AMOVA and PCoA were conducted with GenAlEx 6.5. The unweighted pair group method with arithmetic mean (UPGMA) [56] and neighbour-joining (NJ) [57] clustering methods implemented in TASSEL v5.2.2.5 [53] were used to generate phylogenetic trees.

Bayesian population structure analyses

A Bayesian clustering approach, applying a Markov Chain Monte Carlo (MCMC) algorithm implemented in the population-genetic software *STRUCTURE* v2.3.4 [58], was used to assign the 124 rutabaga accessions from the various countries into a number of genetically homogeneous clusters (K) based on the 6861 SNP markers. In addition, *STRUCTURE* was used to assign the rutabagas 'Laurentian' (Canada), 'Wilhemsburger' (Germany) and 'Krasnoselskaya' (Russia) into the Nordic subpopulations showing similar variation patterns.

STRUCTURE was run using the admixture model with correlated allele frequencies at a series of burn-in lengths from 5000 to 100,000 iterations and MCMC run lengths from 5000 to 100,000 permutations. In addition, *STRUCTURE* was run with the accessions unassigned to any population or country and with the accessions assigned to their countries of origin. Runs for each $K = 1$–10 were replicated 10 or 20 times. These were done to determine the parameters needed to reach convergence. The most likely number of clusters (the ad hoc ΔK test) and average log-likelihood plots were determined following Evanno et al. [34] with STRUCTURE HARVESTER v0.6.94 [59]. Secondly, the median (MedMedK and MaxMedK) or mean (MedMeaK and MaxMeaK) estimators of the "best" K were used to group subpopulations into clusters with STRUCTURESELECTOR [35, 36]. Accessions were assigned to a specific cluster if the probability of membership was ≥0.70, with those that did not meet this threshold considered as an admixture.

Statistical analysis

Statistical significance between means of the parameters (pairwise and overall) was established by Fisher's protected least significant difference (LSD) test ($P \leq 0.05$) using SAS v. 9.4 (SAS Institute, Inc., Cary, NC, USA).

Abbreviations

PIC: Polymorphism information content; %P: The proportion of polymorphic loci; *Na*: The mean number of alleles per locus; *Ne*: The mean number of effective alleles per locus; $\overline{H}_e$: The mean expected heterozygosity; $U \overline{H}_e$: The mean unbiased expected heterozygosity; *Na Freq* ≥ 5%: The mean number of alleles with a frequency ≥ 5%; *Na comm* ≤ 25% and *Na comm* ≤ 50%: Mean number of common alleles found in ≤25% and ≤ 50% of the subpopulations, respectively; *I*: Shannon's information index; F_{ST}: Wright's genetic differentiation F-statistics; MAF: Minor allele frequency; D: Gene Diversity; TASSEL: Trait analysis by association evolution and linkage; MCMC: Markov Chain Monte Carlo; AMOVA: Analysis of molecular variance; PCoA: Principal coordinate analysis; UPGMA: The unweighted pair group method with arithmetic mean; NJ: Neighbour-joining

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07762-4.

<div style="border:1px solid">

Additional file 1.

</div>

Availability of data and materials

The accession numbers of the rutabaga genotypes used in the current study are provided in Table S3 of the supplementary materials. In addition, the datasets generated during the current study are available in the manuscript or the supplementary materials. The SNPs used for genotyping are available from Clark et al. [51] or can be downloaded at https://static-content.springer.com/esm/art%3A10.1007%2Fs00122-016-2746-7/MediaObjects/122_2016_2746_MOESM3_ESM.pdf. Any other information can be obtained from the corresponding author.

Declarations

Ethics approval and consent to participate

The collected seeds were for research and their use complied with the guidelines and regulations of the University of Alberta, the Canada Food Inspection Agency and the International Treaty for Plant Genetic Resources.

Consent for publication

Not applicable.

Competing interests

On behalf of all authors, the corresponding author states that they have no competing interests.

Yu *et al. BMC Genomics* (2021) 22:442

Page 12 of 13

References

1. Iñiguez Luy FL, Federico ML. The genetics of *Brassica napus* L. In: Bancroft I, Schmidt R, editors. Genetics and genomics of the Brassicaceae. New York Dordrecht, Heidelberg, London: Springer; 2011. p. 291–322.
2. Ahokas H. On the evolution, spread and names of rutabaga. In: MTT-agrifood research Finland 2004. Helsinki: Kave; 2004. p. 32.
3. Gowers S. Swedes and turnips. In: Bradshaw JE, editor. Root and tuber crops. Handbook of plant breeding, vol. 7. New York: Springer; 2010. p. 245–89. https://doi.org/10.1007/978-0-387-92765-7_8.
4. Harvey N. The coming of the swede to Great Britain: An obscure chapter in farming history. Agric Hist. 1949;23(4):286–8.
5. Sturtevant EL. Sturtevant's notes on edible plants. Geneva: New York Agr. Exper. Sta; 1919. p. 304–5. https://doi.org/10.5962/bhl.title.24577.
6. Bradshaw JE, Griffiths DW. Sugar content of swedes for stock feeding. J Sci Food Agric. 1990;50(2):167–72. https://doi.org/10.1002/jsfa.2740500204.
7. Pasko P, Bukowska-Strakova K, Gdula-Argasinska J, Tyszka-Czochara M. Rutabaga (*Brassica napus* L. var. *napobrassica*) seeds, roots, and sprouts: a novel kind of food with antioxidant properties and proapoptotic potential in Hep G2 hepatoma cell line. J Med Food. 2013;16(8):749–59. https://doi.org/10.1089/jmf.2012.0250.
8. Pivovarova NS. Breeding material of culinary turnip and swede and methods of evaluating it. Byulleten' Vsesoyuznogo Instituta Rastenievodstva. 1979;90:32–8.
9. Jung GA, Byers RA, Panciera MT, Shaffer JA. Forage dry-matter accumulation and quality of turnip, swede, rape, Chinese-cabbage hybrids and kale in the eastern USA. Agron J. 1986;78(2):245–53. https://doi.org/10.2134/agronj1986.00021962007800020006x.
10. Livingstone M, Jones AS, Mennie I. Swedes (*Brassica napus*) for growing pigs: chemical composition and use as a replacement for barley in the diet. Anim Feed Sci Technol. 1977;2(1):31–40. https://doi.org/10.1016/0377-8401(77)90038-4.
11. Gemmell DJ, Griffiths DW, Bradshaw JE. Effect of cultivar and harvest date on dry-matter content, hardness and sugar content of swedes for stockfeeding. J Sci Food Agric. 1990;53(3):333–42. https://doi.org/10.1002/jsfa.2740530306.
12. Fredua-Agyeman R, Yu Z, Hwang SF, Strelkov S. Genome-wide mapping of loci associated with resistance to clubroot in *Brassica napus* ssp. *napobrassica* (rutabaga) accessions from Nordic countries. Front Plant Sci. 2020;11:742. https://doi.org/10.3389/fpls.2020.00742.
13. Ayers GW, Lelacheur KE. Genetics of resistance in rutabaga to two races of *Plasmodiophora brassicae*. Can J Plant Pathol. 1972;52(6):897–900.
14. Hasan MJ, Rahman H. Genetics and molecular mapping of resistance to *Plasmodiophora brassicae* pathotypes 2, 3, 5, 6, and 8 in rutabaga (*Brassica napus* var. *napobrassica*). Genome. 2016;59(10):805–15. https://doi.org/10.1139/gen-2016-0034.
15. Hasan JM, Strelkov SE, Howard RJ, Rahman H. Screening of Brassica germplasm for resistance to *Plasmodiophora brassicae* pathotypes prevalent in Canada for broadening diversity in clubroot resistance. Can J Plant Pathol. 2012;92(3):501–15.
16. Grant I, Harney PM, Christie BR. Inheritance of yield and other quantitative characters in *Brassica napus* var. *napobrassica*. Can J Genet Cytol. 1982;24(4):459–65. https://doi.org/10.1139/g82-048.
17. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genet Res Int. 2015;2015:431487. https://doi.org/10.1155/2015/431487.
18. Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA. Phylogeographic studies in plants: problems and prospects. Mol Ecol. 1998;7(4):465–74. https://doi.org/10.1046/j.1365-294x.1998.00318.x.
19. Brakenhoff RH, Schoenmakers JG, Lubsen NH. Chimeric cDNA clones: a novel PCR artifact. Nucleic Acids Res. 1991;19(8):1949. https://doi.org/10.1093/nar/19.8.1949.
20. Cline JJ, Braman C, Hogrefe HH. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. 1996;24(18):3546–51. https://doi.org/10.1093/nar/24.18.3546.
21. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol. 2005;71(12):8966–9. https://doi.org/10.1128/AEM.71.12.8966-8969.2005.
22. Kulibaba RA, Liashenko YV. Influence of the PCR artifacts on the genotyping efficiency by the microsatellite loci using native polyacrylamide gel electrophoresis. Cytol Genet. 2016;50(3):16–23.
23. Estoup A, Tailliez C, Cornuet JM, Solignac M. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol Biol Evol. 1995;12(6):1074–84. https://doi.org/10.1093/oxfordjournals.molbev.a040282.
24. Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol. 2002;34(3):275–305. https://doi.org/10.1186/1297-9686-34-3-275.
25. Clarke WE, Higgins EE, Plieske J, Wieseke R, Sidebottom C, Khedikar Y, et al. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. Theor Appl Genet. 2016;129(10):1887–99. https://doi.org/10.1007/s00122-016-2746-7.
26. Hayward A, Mason AS, Dalton-Morgan J, Zander M, Edwards D, Batley J. SNP discovery and applications in *Brassica napus*. J Plant Bio. 2012;39(1):49–61. https://doi.org/10.5010/JPB.2012.39.1.049.
27. Zhou QH, Zhou C, Zheng W, Mason AS, Fan SY, Wu CJ, et al. Genome-wide SNP markers based on SLAF-Seq uncover breeding traces in rapeseed (*Brassica napus* L.). Front Plant Sci. 2017;8:648.
28. Gazave E, Tassone EE, Ilut DC, Wingerson M, Datema E, Witsenboer HMA, et al. Population genomic analysis reveals differential evolutionary histories and patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. Front Plant Sci. 2016;7:525.
29. An H, Qi X, Gaynor ML, Hao Y, Gebken SC, et al. Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. Nat Commun. 2019;10(1):2878. https://doi.org/10.1038/s41467-019-10757-1.
30. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, et al. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nat Commun. 2019;10(1):1154. https://doi.org/10.1038/s41467-019-09134-9.
31. Bus A, Körber N, Snowdon RJ, Stich B. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theor Appl Genet. 2011;123(8):1413–23. https://doi.org/10.1007/s00122-011-1676-7.
32. Diers BW, Osborn TC. Genetic diversity of oilseed *Brassica napus* germplasm based on restriction fragment length polymorphisms. Theor Appl Genet. 1994;88(6-7):662–8. https://doi.org/10.1007/BF01253968.
33. Mailer RJ, Scarth R, Fristensky B. Discrimination among cultivars of rapeseed (*Brassica napus* L.) using DNA polymorphisms amplified from arbitrary primers. Theor Appl Genet. 1994;87(6):697–704. https://doi.org/10.1007/BF00222895.
34. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x.
35. Puechmaille SJ. The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol Ecol Resour. 2016;16(3):608–27. https://doi.org/10.1111/1755-0998.12512.
36. Li YL, Liu JX. StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour. 2018;18(1):176–7. https://doi.org/10.1111/1755-0998.12719.
37. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012;28(19):2537–9. https://doi.org/10.1093/bioinformatics/bts460.
38. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(6):1358–70. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.
39. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005;21(9):2128–9. https://doi.org/10.1093/bioinformatics/bti282.
40. Turesson G. The genotypic response of the plant species to habitat. Hereditas. 1922;3:211–350.
41. Turesson G. The species and the variety as ecological units. Hereditas. 1922;3:100–13.
42. Turesson G. The plant species in relation to habitat and climate: contributions to the knowledge of genecological units. Hereditas. 1925;6(2):147–236.
43. Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian Rice varieties. PLoS One. 2013;8(12):e84136. https://doi.org/10.1371/journal.pone.0084136.

Yu *et al. BMC Genomics*   (2021) 22:442

Page 13 of 13

44. Gonzaga ZJ, Aslam K, Septiningsih EM, Collard BCY. Evaluation of SSR markers for molecular breeding in rice. Plant Breed Biotech. 2015;3(2):139–52. https://doi.org/10.9787/PBB.2015.3.2.139.

45. Varshney RK, Thiel T, Sretenovic-Rajicic T, Baum M, Valkoun J, Guo P, et al. Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. Mol Breed. 2008;22(1):1–13. https://doi.org/10.1007/s11032-007-9151-5.

46. Tsykun T, Rellstab C, Dutech C, Sipos G, Prospero S. Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. Heredity. 2017;119(5):371–80. https://doi.org/10.1038/hdy.2017.48.

47. Kalinowski ST. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity. 2011;106(4):625–32. https://doi.org/10.1038/hdy.2010.95.

48. The Brassica rapa Genome Sequencing Project Consortium, Wang X, Wang H, et al. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet. 2011;43:1035–9. https://doi.org/10.1038/ng.919.

49. Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, et al. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biol. 2014; 15(6):R77. https://doi.org/10.1186/gb-2014-15-6-r77.

50. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. 2014;345(6199):950–3. https://doi.org/10.1126/science.1253435.

51. Wright S. The interpretation of population structure by F-statistics with special regard to system of mating. Evolution. 1965;19(3):395–420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x.

52. Peakall R, Smouse PE. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6(1):288–95. https://doi.org/10.1111/j.1471-8286.2005.01155.x.

53. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5. https://doi.org/10.1093/bioinformatics/btm308.

54. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131(2):479–91. https://doi.org/10.1093/genetics/131.2.479.

55. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2:2074–93.

56. Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. Univ Kans Sci Bull. 1958;28:1409–38.

57. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.

58. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet. 2000;67(1):170–81. https://doi.org/10.1086/302959.

59. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61. https://doi.org/10.1007/s12686-011-9548-7.

Publisher's Note