Population structure in landrace barley (Hordeum vulgare L.) during the late 19th century crop failures in Fennoscandia

Nils Forsberg, Matti W. Leino and Jenny Hagenblad

Population structure in landrace barley (*Hordeum vulgare* L.) during the late 19th century crop failures in Fennoscandia

Nils E G Forsberg[1,2], Matti W Leino[2,3,4] and Jenny Hagenblad[2]

[1] Norwegian University of Science and Technology, Department of Biology, N-7491 Trondheim, Norway

[2] IFM-Biology, Linköping University, SE-581 83 Linköping, Sweden

[3] Nordiska museet, Swedish Museum of Cultural History, Box 27820, SE-115 93 Stockholm, Sweden

[4] The Archaeological Research Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden

Corresponding author: Jenny Hagenblad, IFM-Biology, Linköping University, SE-581 83 Linköping, Sweden. Phone: +46 13 286686. Email: Jenny.Hagenblad@liu.se

Word count: 6484

Abstract

Agricultural disasters and the subsequent need for supply of relief seed can be expected to influence the genetic composition of crop plant populations. The consequences of disasters and seed relief have, however, rarely been studied since specimens sampled before the events are seldomly available. A series of crop failures struck northern Fennoscandia (Norway, Sweden and Finland) during the second half of the 19th century. In order to assess population genetic dynamics of landrace barley (*Hordeum vulgare*), and consequences of crop failure and possible seed relief during this time period, we genotyped seeds from 16 historical accessions originating from two time periods spanning the period of repeated crop failure. Reliable identification of genetic structuring is highly dependent on sampling regimes and detecting fine-scale geographic or temporal differentiation requires large sample sizes. The robustness of the results under different sampling regimes was evaluated by analysing subsets of the data and an artificially pooled dataset. The results led to the conclusion that six individuals per accession were insufficient for reliable detection of the observed genetic structure. We found that population structure among the data was best explained by collection year of accessions, rather than geographic origin. The correlation with collection year indicated a change in genetic composition of landrace barley in the area after repeated crop

43 failures, likely a consequence of introgression of relief seed in local
44 populations. Identical genotypes were found to be shared among some
45 accessions, suggesting founder effects and local seed exchange along known
46 routes for trade and cultural exchange.
47

Introduction

Extreme climate events are a constant threat to low-yielding agricultural areas and understanding and predicting the long-term genetic effects on genetic composition after disasters and seed relief is important for food security (Ferguson et al. 2012). While the status and recovery of crops in the aftermath of recent disasters and conflicts have been studied (e.g. Sperling 2001; Jones et al. 2002; Ferguson et al. 2012; Fuentes et al. 2012) most studies have been from contemporary Africa.

In the northern parts of Fennoscandia (Norway, Sweden and Finland), above the 65th parallel, lies the northernmost limit of cereal cultivation (Bjørnstad 2012). Archaeobotanical studies have shown that cereal cultivation has a long history in the region (Bergman and Hörnberg 2015; Josefsson et al. 2017), with finds dating back to at least 500 BC along the coast and 1400 AD in the interior (Bergman and Hörnberg 2015). The region covers vast land areas but agricultural land is restricted to small and isolated locations. Due to the harsh climatic conditions in the region the only cereal species with sufficient hardiness for cultivation is barley (*Hordeum vulgare*) (Bjørnstad 2012).

Barley is a diploid species that is almost completely self-fertilizing across a range of environments (Abdel-Ghani et al 2004). The landraces grown in Northern Fennoscandia were well-known for their adaption to the short growing season through fast maturation but at the cost of smaller harvests. Agricultural literature from the late 19th and early 20th century tell of a original type of barley cultivated in this region known as “lappkorn” (Lapponian barley) or “finnkorn” (Finnish barley) (Grotenius, 1896; Hellström, 1917). Indeed, Forsberg et al. (2015) studying as few as six individuals from each of 31 historical accessions from all over Fennoscandia and Denmark, showed how six-row barley from northernmost Fennoscandia was, as a group, genetically differentiated from six-row barley elsewhere in the region. Similar results were obtained by Lempiäinen-Avci et al. (2018) focusing on Finnish barley.

In spite of the barley’s well-known hardiness, Fennoscandia has historically repeatedly suffered from crop failures (Dribe et al. 2015). Extreme weather in the region during the years 1866 - 1869 resulted in several consecutive years of crop failure (Häger et al. 1978; Nelson 1988). In 1867 the spring was so cold that anomalies of such a magnitude are only expected to occur a few times in a millennium (Jantunen and Ruosteenoja, 2000). In addition, autumn came early this year imposing harvest of yet unripe cereals. The following years were only marginally better (Nelson 1988). The poor

harvests during this period contributed to a culmination in emigration from Sweden to America (Grym 1959; Nelson 1988) and in northern Finland the human population shrank by 5% from 1865 to 1870 from the combined effects of emigration, starvation and starvation-related disease (Tilastokeskus 1875; Pitkänen 1992). Yield rates, expressed as the ratio of harvest volume compared to seed volume, from the Norrbotten region in northernmost Sweden for the years 1865 - 1900, reveal that crop failures also occurred in 1877, 1888 and 1892 (Statistiska centralbyrån 1856-1905). Yield rates from Finland and Norway show a similar general pattern, with consecutive years with low yield during the second half of the 1860s and additional sporadic years with poor yield during the period 1870-1900 (Tilastokeskus 1875; Nelson 1988). The drastic and frequent loss of seed from crop failure may have resulted in a loss of indigenous genetic barley diversity through population bottlenecks. During the crop failure in northern Sweden 1867 - 1869, both national and international efforts were made to alleviate famine (Häger et al. 1978; Nelson 1988). Emergency relief was mostly provided in the form of flour, not seed, and the seed import from outside the region was not particularly increased during the period (Nelson 1988). Whether farmers stayed true to their local landraces and saved what little harvest they had for seeding the next year's crop or whether what seed import there was led to the addition of novel genetic diversity to the local landraces is not known.

Few extant landraces are available from Northern Fennoscandia. In contrast, the area is unusually well endowed when it comes to accessions of historical seed samples (Leino et al. 2009; Leino 2010). During the late 19th century the northernmost Fennoscandia was the target of several seed collection missions with the purpose of obtaining material to display at fairs and exhibitions (Leino 2010). The specimens, mostly six-row barley, gathered during some of these missions remain at museums across Fennoscandia (Leino et al. 2009; Leino 2010). The age of the material ensures that it represents genuine landrace barley, as plant improvement for six-row barley in Fennoscandia did not begin until the early 20th century (Osvald 1959). Although the seeds are no longer viable, genetic analysis of DNA is possible (e.g. Leino et al. 2009; Forsberg et al. 2015). Historical accessions collected from northernmost Fennoscandia generally fall into two distinct temporal classes, 1867 - 1870 and 1893 - 1896, thus spanning the years of crop failure. This provides an opportunity to study the famine years' effect on the crop's genetic composition. The Fennoscandian crop failures of the late 19th century can thus serve as an excellent case study of how the genetic composition of landrace crops changes after a period of continuous poor harvests.

Studies of genetic structure and spatial distribution of crops has received considerable attention in recent years (e.g. Olsen and Schaal 1999; Londo et al. 2006; Jones et al. 2011; Oliveira et al. 2012, Yelome et al. 2018). In most cases such studies have relied on the genotyping of single seeds or pooled DNA from multiple seeds thereby increasing the number of accessions or populations that can be screened. Other studies have prioritized genotyping of multiple individuals of each population (e.g. Papa et al. 1998; Demissie et al. 1998; Leino and Hagenblad 2010; Forsberg et al. 2015, Hagenblad et al. 2017). Computer simulations and microsatellite data from *Arabidopsis thaliana* suggests that the number of sampled individuals per accession can affect the ability to detect genetic clusters (Fogelqvist et al. 2010). The power to detect genetic structuring over short periods of time or limited geographical ranges, where the genetic variation within populations is much greater than the diversity among populations, may thus be strongly affected by the sampling regime.

In this study we have investigated the temporal consequences of crop failure and subsequent relief on the genetic composition of 19th century landrace barley in Northern Fennoscandia. To facilitate detection of relatively small effects on a regional scale we sampled up to 20 individuals from each accession. By creating subsets and artificially mimicking the output from single seed sampling and pooling of DNA extracts we also assessed the

effect of different sampling regimes on the ability to detect genetic clustering.

Materials and Methods

Sample selection

Twenty grains from each of 16 accessions of landrace six-row barley were chosen for the study (Table 1). Some of the specimens had previously been part of the Forsberg et al. (2015) study, but new accessions from northernmost Fennoscandia were added and the number of grains from each accession were more than tripled to increase the power to detect fine-scale genetic structure beyond that of Forsberg et al. (2015). The accessions were obtained from three different 19th century seed collections; Tromsø University Museum in Norway (TR, four accessions), Mustiala Agricultural College in Finland (MU, two accessions) and the Swedish Museum of Cultural History in Sweden (NM, ten accessions) (Leino 2010). Grain had been gathered from farmers during two distinct three-year periods in the 19th century that were classified into an "Early" (1867 – 1870, seven accessions) and a "Late" (1893 – 1896, nine accessions) class (Table 1). Maps for geographic representation of accession origin and geographic genetic

177 structure were generated using ArcGIS (ESRI, Redlands, CA, USA) with
178 geographic base data from the “ESRI data and maps v. 9.3” database (2008).
179
180 *DNA-Analysis*
181 DNA was extracted from individual seeds from each accession using
182 FastDNA Spin Kits and the FastPrep Instrument (MP Biochemicals, Solon,
183 OH, USA). Extractions were performed at a laboratory separate from that
184 where SNP genotyping was carried out to reduce the risk of contamination.
185 A negative control was included in each extraction series and a total of nine
186 negative controls were included in the genotyping. Genotyping was
187 performed using an Illumina Golden Gate assay (Illumina Inc., San Diego,
188 CA, USA) for the C-384 barley SNP set detailed by Moragues et al. (2010).
189 The robustness of the C-384 SNP set on historical barley landrace material
190 was shown in Forsberg et al. (2015).
191
192 The resulting data were processed and studied with the BeadStudio 3.1.3.0
193 software package (Illumina Inc., San Diego, CA, USA). Quality control
194 based on CG10 scores led to the exclusion of 26 low performance samples,
195 including all nine negative controls. Samples with more than 25 % missing
196 data (39 samples), markers with more than 20 % missing data (92 SNPs)
197 and monomorphic SNPs (140 SNPs) were also excluded, in that order. High

198 quality genotypes for 152 SNP variable markers were obtained from 275
199 individuals.
200
201 *Genetic diversity*
202 Within-accession genetic diversity was calculated as Nei's h (Nei 1973),
203 using a purpose-written script in the statistical software R (R development
204 core team 2013, version 3.0.2). The distribution of genetic diversity was
205 further studied through AMOVA (Excoffier et al. 1992) and F_{ST} statistics
206 (Weir and Cockerham 1984) between pairs of accessions. Pairwise F_{ST} was
207 also calculated between the Early and Late classes of accessions and
208 between groups defined by country of origin. F_{ST} significance was estimated
209 using permutation tests with 1000 permutations. AMOVA was performed
210 with country of origin and age class as discrete groups. The proportion of
211 total genotype sharing, i.e. individuals that were scored as identical, within
212 and between accessions was also calculated. AMOVA, pairwise F_{ST} and total
213 genotype sharing were calculated using the *Arlequin 3.5* software (Excoffier
214 and Lischer 2010). *Arlequin* was set to infer haplotype definitions from the
215 distance matrix and to allow for 25% missing data per loci.
216
217 *Population structure*
218 Population structure was assessed in R using principal component analysis
219 (PCA) and the SNP data was analysed both at an accession level and on an

220 individual level. For the individual level, each homozygous SNP was treated
221 as either 1 or 0 and missing data were replaced with the allele frequency in
222 the full dataset of the allele designated as ‘1'. For the accession level PCA,
223 allele frequencies of each accession for each of the SNPs were calculated
224 and treated as independent variables. PCoA was included as a comparison
225 with PCA and was assessed using the *ape* R package (Popescu et al 2012).
226 PC dispersion, the mean pairwise distance in PC-space between individuals
227 within accessions, was calculated as the average distance between
228 individuals belonging to the same accession in a multidimensional space
229 calculated from all principal components according to Forsberg et al.
230 (2015). Population clustering was explored using two different methods,
231 *structure* (Pritchard et al. 2000; Falush et al. 2007, version 2.3.3) and
232 Discriminant Analysis of Principal Components, DAPC (Jombart et al.
233 2010). Genotype data was analysed as haploid, as suggested for *structure*
234 clustering for predominantly self-fertilizing species by Nordborg et al.
235 (2005), treating heterozygous loci as missing data. The admixture model
236 was used and simulations were run with a burn-in period set to 25 000
237 iterations and estimates based on the following 50 000 iterations for one
238 through ten clusters (K = 1 to 10). Potential multimodality of the clustering
239 analyses was resolved by merging 20 runs for each value of K using the
240 CLUMPP software (Jakobsson et al. 2007). CLUMPP merging used the
241 Greedy Algorithm method and results were visualized with the Distruct 1.1

242 software (Rosenberg 2004). The optimal number of clusters was assessed
243 using the H' statistic from CLUMPP and the ΔK value calculated as
244 suggested by Evanno et al. (2005). In addition to analysis of the full data set,
245 accessions were divided into the Early and Late classes and analysed
246 separately in *structure*, to assess the geographic genetic structure within the
247 temporal classes. DAPC was performed using the *Adegenet* R package
248 (Jombart et al. 2011). All principal components were used for prior group
249 clustering and the 10 most principal components were used to prevent over-
250 fitting. The DAPC analysis was repeated 20 times and the results were
251 merged using CLUMPP to resolve multimodality. The merged results were
252 visualized with the Distruct 1.1 software.
253
254 *Analysis of covariation of genetic structure with geographic and temporal*
255 *information*
256 To pinpoint underlying causes for the observed population clustering, as
257 determined by *structure* and PCA, clustering was tested for correlation with
258 geographic and temporal variables. Cluster membership from *structure* and
259 the two most informative principal components of the PCA were tested
260 against the latitude, longitude, altitude, country of origin and age of the
261 accessions using a multiple linear regression. Geographic parameters
262 (altitude, latitude and longitude) were used as numerical variables, country
263 of origin was defined as categorical variables. The temporal variable,

264 defined as the collection year of the accessions, was tested both as a
265 numerical variable and as a categorial variable with the temporal classes
266 Early or Late (Table 1). Simultaneous testing of geographic parameters and
267 country of origin was performed using multiple linear regression models
268 with either cluster membership from the merged s*tructure* simulations with
269 the highest support or PC1 or PC2 score from the PCA as the regressand.
270 Both accession-level cluster membership and individual cluster membership
271 were used as two separate levels of testing. Accession level data was
272 analysed using fixed effect models while individual level data was analysed
273 both with fixed effect models and mixed effect models. Since genotyping
274 was performed on several different plates, plate identity of the samples was
275 included as a random effect for the mixed effect models. The comparison
276 between the two temporal classes was performed using a two-sample t-test,
277 under the assumption that the data was normally distributed (confirmed
278 through Kolmogorov-Smirnov tests). Correlations where covariations were
279 found between explanatory variable were, additionally, analysed using
280 partial correlation, to compensate for the detected covariation. All statistical
281 testing was performed using R.
282
283 *Effect of sampling regime on detection of population structure*
284 The effect of sampling regime on detection of population structure was
285 assessed by repeating principal component and *structure* analyses using

subsets of the data, created to simulate smaller sample sizes and DNA pooling. All subsets were compared with the full dataset under the assumption that the full dataset would have a more accurate fit to the underlying genetic distribution than the subsets. Ten replicates each of single-individual and six-individual sample schemes were randomly generated from the full dataset. An artificially pooled dataset was generated using data from all individuals in each accession and used the most frequent allele for each locus in a given accession as the pooled genotype.

The H' value from the software CLUMPP after grouping the 20 replicate *structure* simulations for each K was used to compare the robustness of the clustering and to determine whether the same number of clusters were detected for the subsets. The sum of squares of the difference in cluster assignment after CLUMPP for each subset and the full dataset for each accession was calculated and compared in R. Principal Component data were compared with Procrustes analysis using the procOPA function included in the *shapes* package of R, with mirroring of axes allowed. Only the two principal components that explained the most variation were used in the analysis.

To determine whether the clustering output from the single-sample, six-sample and pooled subsets resulted in different conclusions than that from

the full data set, clustering information from *structure* and PCA from the subsets was subjected to the same additional analysis as the full dataset. Clustering information was tested with multiple linear regression with geographical parameters using multiple linear regression. Non-significant variables were excluded by order of decreasing p values. A two-sample t-test was used for detecting co-dependence of clustering with temporal class.

Results

Diversity within and between accessions

Within-accession genetic diversity (Nei's h) ranged from 0.043 to 0.160, with an average of 0.113 (Table 1). No significant difference was found between the within-accession genetic diversity of the different temporal classes "Early" and "Late" (two sample t-test, $M_{Early} = 0.107$, $SD_{Early} = 0.033$, $M_{Late} = 0.118$, $SD_{Late} = 0.038$, $p = 0.559$). No significant geographic trend in within-accession diversity was observed and diversity was not correlated with either altitude, latitude, longitude or country of origin (all $p > 0.05$). Highly diverse accessions could be found both from both northern ($h_{TR7} = 0.147$) and southern parts ($h_{NM668} = 0.152$ and $h_{NM669} = 0.160$) of the region. Large differences in within-accession genetic diversity could also be seen when comparing nearby accessions. For example, the genetic diversity of NM633 ($h_{NM633} = 0.043$) differed markedly from that of its nearest neighbours

NM751 (h_{NM751} = 0.125, distance ≈ 79 km) and NM599 (h_{NM599} = 0.122, distance ≈ 92 km), all Late accessions. On the other hand, MU69 (Early) and NM751 (Late), the accessions with the shortest geographic distance, had similar levels of genetic diversity (h_{MU69} = 0.121 vs. h_{NM751} = 0.125, distance ≈ 6 km).

Pairwise F_{ST} values between accessions across loci ranged from being slightly negative to a value of 0.362, when comparing NM1597 to MU1 (Supplementary table 1). Plotting F_{ST} values against geographic distance indicated no pattern of isolation by distance neither in the full dataset nor in the early or late groups considered separately (Supplementary figure 1) and geographic distance and pairwise F_{ST} values were not significantly correlated in either dataset (all accession pairs, c = -0.042, p = 0.645; early accession pairs, c = 0.024, p = 0.919; late accession pairs, c = -0.174, p = 0.310). Indeed, low F_{ST} values were not necessarily linked to short geographic distances, in particular when comparing between temporal classes. For example, Swedish NM1587 shared most similarity with the Norwegian accession TR8 with an origin 437 km away but from the same age class (F_{ST} = 0.03). NM1587 was in contrast quite different from its geographically nearest accession NM669, with an origin only 42 km away but belonging to a different age class (F_{ST} = 0.15) (Table 1, Supplementary table 1). Likewise, NM1597 was more similar to NM789, cultivated some 300 km away (F_{ST} = 0.05), than the nearest accession NM668 with an origin only 100 km away

(F_{ST} = 0.32) (Table 1, Supplementary table 1). F_{ST} comparisons between different countries of origin and different temporal classes, respectively, gave low, albeit significant, values. On a country level, F_{ST} indicated isolation by distance, with the largest difference between the most distantly located countries: Norway and Finland (F_{ST} = 0.0684***) followed by the Sweden - Norway (F_{ST} = 0.0421***) and Sweden - Finland (F_{ST} = 0.0416***) comparisons, both with similar F_{ST} values. The difference between temporal classes (F_{ST} = 0.0526***) was lower than the Norway – Finland comparison but higher than the F_{ST} values of the Sweden – Norway and Sweden – Finland comparisons.

Genetic structuring in northern Fennoscandian barley

The results of the DAPC clustering were largely similar to those of the *structure* clustering, although with a lower proportion of admixture (Supplementary table 2). Similarly, results from PCA and PCoA were highly correlated (accession level PC1 vs PCo1 and PC2 vs PCo2: c = -1; individual level PC1 vs PCo1: c = - 0.997; individual level PC2 vs PCo2: c = - 0.987). Hence, only *structure* and PCA results are presented below. Both H' values and ΔK suggested that a two-cluster model best described the distribution of the genetic diversity (Supplementary table 3) and membership to these clusters were used downstream as the response variable in a regression analysis. Five of the Early accessions (the Finnish

MU69, the Swedish NM1587 and NM1597 and the Norwegian TR1 and TR5) and three of the Late accessions (the Swedish NM633 and NM789) clustered together (light grey in Figure 1 and Figure 2), five of the Late accessions (the Finnish MU1 and the Swedish NM668, NM669, NM727 and NM751) clustered in a second group (dark grey in Figure 1 and Figure 2) while the remaining accessions (NM599, TR7 and TR8) were highly admixed. *Structure* results from analysis of the temporal classes separately yielded similar distributions as the full dataset, without apparent geographic structure (Supplementary table 4, 5).

PCA was performed both on an accession level and on an individual level. The first and second principal components explained a very high proportion of the total genetic diversity in the accession level analysis (Figure 3A; PC1 = 47.48 %, PC2 = 14.02 %) and a smaller proportion in the individual level PCA (Figure 3B; PC1 = 17.31 %, PC2 = 8.90 %). As expected, given the high explanatory power of PC1, the distribution of accessions along PC1 (Figure 3A) was highly similar to the *structure* clustering. The individual level PCA showed a shift in the genetic composition between the Early and Late samples along both PC1 and PC2 (Figure 3B). Despite low mean PC dispersion in the individual level PCA, NM1597 and NM633 had the highest PC dispersion variance of the accessions studied (Table 1), indicative of within-accession substructure.

Temporal class is an explanatory parameter for genetic structuring

In the accession level model (Table 2) no significant correlation with genetic clustering was found for either of the geographic parameters (latitude, longitude, altitude or country of origin) when the variables were tested as single regressions (all $p > 0.05$). Population clustering was, however, significantly correlated with temporal class, both from *structure* clustering ($p = 0.035$ and $r^2 = 0.23$), PC1 ($p = 0.039$ and $r^2 = 0.12$) and PC2 ($p = 0.028$ and $r^2 = 0.25$). The Early and Late temporal classes resulted in similarly high correlations with both cluster membership from *structure* (two sample t-test, $p = 0.0259$) and principal component score for PC1 (two sample t-test, $p = 0.029$). When using multiple linear regression with temporal class, latitude, longitude, altitude and country of origin as regressors the temporal link was obscured. Temporal class remained the most significant variable in the full model ($p = 0.131$ for PC1 and $p = 0.106$ for *structure* clustering, Supplementary table 6), and the effect of temporal class became significant after consecutively removing the least significant variables. Using *structure* clustering, temporal class became significant when longitude and latitude were excluded ($p = 0.047$), for PC1 when longitude was excluded ($p = 0.040$) and for PC2 when altitude, longitude and country of origin were excluded ($p = 0.040$). To assess whether this was an effect of uneven spatial sampling, correlations between harvest year (i.e.

418 not temporal class but the actual year of harvest) and geographic origin was
419 analysed. Harvest year was highly correlated with both sample latitudinal
420 origin ($r = -0.587$, $p = 0.019$) and longitudinal origin ($r = 0.530$, $p = 0.033$).
421 When using partial correlation to assess the effect of harvest year while
422 correcting for the spurious correlation with longitude or latitude, harvest
423 year tended to be associated with genetic clustering, although only
424 significantly so for PC2 (*structure* clustering latitude $p = 0.061$, longitude p
425 $= 0.098$; PC1 latitude $p = 0.075$ longitude $p = 0.098$; PC2 latitude $p = 0.035$,
426 longitude $p = 0.045$).
427
428 In the individual level model (Table 2) the effect of sample plate during
429 genotyping, if treated as a fixed effect, was found to be non-significant ($p =$
430 0.154). Latitude, longitude and country of origin, but not altitude, were
431 significantly correlated with *structure* clustering at the individual level if
432 tested as single correlations (all $p < 0.001$ except for altitude $p > 0.05$),
433 although each explained a very small portion of the variation ($r^2 = 0.0391$,
434 0.0470 and 0.0510 for latitude, longitude and country of origin,
435 respectively). Similar results were found for PC1 with significant
436 correlations but low explanatory power for longitude, latitude and country
437 of origin (all $p < 0.001$ except latitude $p < 0.01$; $r^2 = 0.0420$, 0.0345 and
438 0.0481 for longitude, latitude and country of origin, respectively). The
439 highest correlation and explanatory power for structure clustering and PC1

were found when testing the regression between individual cluster membership and harvest year (Table 2), which was highly significant both for *structure* clustering ($p < 0.001$, $r^2 = 0.134$) and PC1 ($p < 0.001$, $r^2 = 0.119$). Comparing the two temporal classes on the individual level revealed a significant difference in cluster membership from *structure* (two sample t-test, $p < 0.001$) and principal component score for PC1 (two sample t-test, $p < 0.001$). PC2 differed slightly showing correlation with altitude ($p = 0.029$, $r^2 = 0.014$), country of origin ($p < 0.001$, $r^2 = 0.094$) and harvest age ($p = 0.035$ and $r^2 = 0.013$).

Using multiple linear regression with either *structure* clustering or PC1 as regressand and altitude, longitude, latitude, country of origin and temporal class as regressors, resulted in temporal class and country of origin as significant (both $p < 0.001$ for both *structure* clustering and PC1, Supplementary table 6). In the multiple linear regression with PC2, however, only country of origin was significant. Mixed effect models including sample plate as random effect yielded similar results for structure clustering and PC1 (Supplementary table 6). Conversely, in mixed effect models for PC2 only country of origin and temporal class were significant (Supplementary table 6).

Analysing the genetic structure of each temporal classes separately yielded no significant geographic effects for the accession level (Supplementary table 7). Analysed on the individual level the longitudinal origin had the highest covariance with the genetic structure of both the Early and the Late class.

AMOVA provided additional support for the separation by age class. The bulk of the variation, some 85 %, was found within the accessions, with 11.13 % and 12.57 % of the variation present within temporal classes and countries respectively (Table 3). Although a minor part of the variation was found between temporal classes and among countries we note that the age class parameter explained 3.85 % of the variation, whereas the country of origin parameter explained less than half the amount, 1.64 %, of the variation in their respective models.

Effects of sampling procedures

Subsampling the dataset to sample sizes of one and six individuals per accession, respectively, reduced the number of informative SNPs to on average 75.6 % and 97.8 %, respectively (Table 4). An even higher loss of information was seen in the pooled sample, where only 21.7% of the SNPs were still informative, compared to the 152 SNPs in the full dataset. Sum of Squares of difference from the *structure* cluster designations from the full

dataset to those of the subsets showed that the six-individual sample size subsets aligned closer to the full dataset ($AvgSSQ_{6Ind.vs.Full} = 0.201$, $sdSSQ_{6Ind.vs.Full} = 0.077$) than the artificially pooled subset ($SSQ_{Pool.vs.Full} = 0.548$), and that the single individual subsets differed the most from the full dataset ($AvgSSQ_{1Ind.vs.Full} = 1.371$, $sdSSQ_{1Ind.vs.Full} = 0.409$).

Procrustes analysis of the two major PCs revealed that all six-individual subsets but one were more similar to the PCA of the full dataset than the pooled dataset was (Table 4). The average OSS (Ordinary Procrustes Sum of Squares) for subsets was significantly smaller for the six-individual subsets compared to the pooled sample (one sample t-test, $p < 0.001$), indicating that the principal components were more similar when comparing the full dataset with the six-individual subsets than with the pooled dataset. The PCA of the subsets using single individuals differed by far the most from the PCA of the full dataset (one sample t-test, $p < 0.001$).

The correlations between population structure and temporal and geographic parameters were also analysed for the subsets and compared with those of the full dataset (Supplementary table 8). Co-dependence with temporal class could only be detected in one out of the ten single-sample subsets. Significant correlations with altitude ($p < 0.05$) were detected in two single-sample subsets. In the six-sample subsets significant correlations with age

505 class was detected in four out of ten subsets for PC1 and *structure* clustering
506 and five of ten subsets for PC2. The artificially pooled dataset found the
507 same co-dependence with temporal class as the full dataset and an additional
508 correlation between latitudinal origin of the accessions and *structure*
509 clustering.
510
511 *Genotype sharing suggest long distance seed exchange*
512 Individuals sharing the same total genotypes, where every scored SNP was
513 identical, were found both within and among accessions. Six groups of
514 shared total genotypes that included individuals from several accessions, an
515 indication of seed exchange, were found. Three of these included more than
516 three individuals (Supplementary table 9). The majority of the three most
517 common shared total genotypes (genotype 1 – 3 in Supplementary table 9)
518 were found in accessions from the Torne Valley (MU69, NM599, NM633,
519 NM798 and NM789) along the Swedish-Finnish border (Figure 4). The
520 most common shared total genotype (genotype 1 in Supplementary table 9),
521 which occurred in 16 copies, was primarily shared between the least diverse
522 accessions, with six copies occurring in NM1597 and four copies in
523 NM633. In contrast with the Torne Valley accessions, these two accessions
524 were from geographically distant localities.
525
526

Discussion

Using a large number of individuals from each studied accession increased our power to detect fine-scale genetic structure in a geographic region that had previously seemed genetically relatively homogeneous (Forsberg et al. 2015). Although geographic origin was associated with genetic structuring parameters, the sampling time point better explained the genetic distribution of the data. The 30-year span separating the Early and Late accessions is infamous for the repeated crop failures occurring in the region.

Disastrous events have throughout history led to failure of food production and subsequent risk of starvation. In many cases relief efforts, either in the shape of food, or through supplies aiming to restore agricultural production, have alleviated the consequences. Modern examples are the restoration of agriculture after the hurricane Mitch disaster in Honduras in 1998 and the civil war in Rwanda 1994-1996. In both cases replacement seed from the CGIAR institutes played an important role (Varma et al. 2004). However, seed relief risks narrowing the local crop gene pool and introduce less adapted genotypes (Ferguson et al. 2012). Seed relief may thus affect the long-term efficiency of local agriculture.

During the worst years of crop failure in northern Sweden in the 1860s, many farmers had no seed for the spring sowing. After the most devastating

549 year, 1867, seed shortage was described as a general and severe problem in
550 the yearly agricultural reports collected by the regional Rural Economy and
551 Agricultural Society (Rydstedt et al. 1868). The following year, the same
552 source reports that many farmers had planted seed imported from more
553 southerly locations (Finell et al. 1869). Our findings of temporal genetic
554 structure during this time period corroborate these reports and suggest that
555 the composition of plant material changed as a result of seed aid and import
556 and that replacement seed was not only acquired locally. Although both
557 major clusters detected here were present in both temporal classes, there
558 was a considerable shift in the distribution of cluster membership (Figure 2)
559 when comparing accessions collected during the early famine years, 1867 –
560 1870, (Early accessions) to accessions collected 1893 – 1896, after the
561 famine (Late accessions).
562
563 Population crashes are expected to lead to a reduction in the genetic
564 diversity through increased genetic drift during the population bottleneck
565 (Nei et al. 1975). We were, however, unable to detect any general reduction
566 in within-accession genetic diversity among the late accessions. Ferguson et
567 al. (2012) showed that the genetic composition of cowpea changed
568 significantly after a severe flood in Mozambique in 2000, while maintaining
569 a similar level of diversity. Similarly, the varietal composition, but not
570 overall diversity, of beans in Rwanda was affected by the civil war in 1994-

1996 (Sperling 2001). The same pattern seems to have followed the 19th century crop failure in Northern Fennoscandia. The dependency of agrarian societies on crop plants for their sustenance means that crop failure calls for supplementary seed to be brought in from other regions. An input of new genetic diversity is thus expected to follow the reduction in population size, which could manifest itself in a shift in the structuring of the genetic diversity such as the one detected here. An input of new seed would also counteract the loss of genetic variation following a population crash and could explain why not such loss was detected here.

Although the study area is vast and covers several different bio-climatic zones (Karlsen et al., 2006) previous studies have shown no significant geographic structure in barley within Northern Fennoscandia (Leino and Hagenblad 2010; Forsberg et al. 2015). Genetic structuring is associated with some of the geographic parameters investigated in this study. However, the associations are weaker than the temporal associations and may be an effect of uneven sampling with regards to sample age. Hellström (1917) describes barley from northern Sweden as being phenotypically relatively variable but suggests that differences were primarily evident in comparisons between landraces from different altitudes rather than latitudes or different municipalities. In this study we did not detect any relationship between altitude and genetic clustering. Unfortunately, contemporary metrological

593 data are not detailed enough to allow for further genetic-climatic
594 correlations and use of modern-day climatic data is problematic, as climate
595 has changed quite dramatically in this region during the past 150 years. Bio-
596 climatic zones and important agricultural parameters such as length of
597 growth season do, however, depend primarily on latitude and altitude in this
598 area (Karlsen et al., 2006), parameters with only minor correlation, and
599 lesser than that of sample age, with genetic structuring among the samples
600 studied here.
601
602 While the use of historical seed allows us to study past temporal and
603 geographic distribution of genetic diversity, access to samples limits the
604 quality of sampling. It was not possible to obtain a geographically even
605 distribution of Early and Late accessions from the area studied. For
606 example, all the Norwegian accessions are Early accessions while the
607 majority of the Swedish accessions are Late. It is therefore possible that the
608 detected difference between Early and Late accessions also has a
609 geographical component that cannot be discerned from the available
610 historical material. Lacking any possibility of improving the sampling, we
611 tentatively note that in the two cases with Early and Late accessions from
612 the same area (the Early MU69 vs the Late NM751 and the Early NM1587
613 vs the Late NM669) we do see a larger than average shift in the genetic

clustering ($\Delta Clustering_{MU69\text{-}NM751} = 0.47$ and $\Delta Clustering_{NM1587\text{-}NM699} = 0.69$, $\Delta Clustering_{Early\text{-}Late} = 0.31$).

The detection of temporal genetic structure was made possible by the large number of individuals analysed from each accession. Neither the single-individual nor the six-individual subset samples were able to reliably detect the temporal shift in genetic structuring identified in the full data set. The sampling of within-accession diversity has been shown to aid in the correct identification of genetic structure (Fogelqvist et al. 2010; Hagenblad et al. 2017) and our results corroborate these findings. In this study we find that the artificial pooling scheme, despite vastly reducing the number of informative SNPs, found the same significant correlations with age class as the full dataset, but performed worse than the six-individual sampling in terms of detecting the same genetic structure in *structure* and PC analyses. Pooling a large number of individual DNA extracts may be as useful as studying a small number of seeds on an individual basis, and is preferable to sampling single individuals in cases where the cost of genotyping is a limiting factor. It should, however, be noted that the required number of individuals per accession also depends on the diversity among accessions and the research questions being asked. In this study we found a pronounced need for a large sample size for assessing genetic structure since the genetic change over time was relatively small. In studies were the genetic variation

636 between accessions is small relative to the genetic variation within
637 accessions we advise the use of several tens of individuals per accession.
638
639 Individuals sharing a total genotype were detected among the accessions
640 along the long-established trade route of the Torne valley (Groth 1984), and
641 the total genotype sharing, and in several cases low F_{ST} values between
642 Torne Valley accessions, is likely the result of seed exchange in this area.
643 This is an example of local networks of seed exchange in areas with
644 common infrastructure and agricultural conditions. Such systems are
645 regularly formed in agrarian societies depending on landrace cultivation
646 (Thomas et al. 2011) and recent day examples show the particular
647 importance of such systems after disastrous events (Sperling 2001). Seed
648 trade was, however, probably not ubiquitous in the area. The low genetic
649 diversity of NM1597 (Kvikkjokk) and high F_{ST} values between NM1597 and
650 the neighbouring NM1587, NM668 and NM669 instead suggests isolated
651 farming in Kvikkjokk, tentatively also with bottleneck effects from the
652 agriculturally very demanding conditions described in the area by
653 contemporary sources (Laestadius, 1824).
654
655 The high degree of total genotype sharing between NM1597 from
656 Kvikkjokk and several accessions in the Torne valley, almost 300 km apart,
657 is puzzling, but has a possible historical explanation. The seeds from

Kvikkjokk (NM1597), characterized by a high presence of Genotype 1 (six out of 17 genotyped individuals), were donated by Johan Laestadius, vicar of Kvikkjokk 1860-1870. Johan's uncle Lars-Levi Laestadius provides a historical link between most sites with Genotype 1. L.L. Laestadius was an early 19th century vicar and botanist from Kvikkjokk with a considerable interest in agronomy. In 1826 L.L. Laestadius took up a position as vicar in Karesuando and in 1849 in Pajala. These two localities are the origin of the two accessions NM798 and NM633, respectively, which contain the largest proportion of Genotype 1 outside of NM1597. Whether the botanist and vicar or his family, upon moving, brought seeds with them and thereby influenced the genetic composition of barley in the Torne region, can probably never be established beyond speculation. Nevertheless, the possible effect of influential individuals on the distribution of genetic diversity of cultivated crops cannot be disregarded and remains a tantalizing thought.

Conclusions

By genetic analysis of a large number of samples per accession we have shown how the genetic composition of landrace barley in northern Fennoscandia changed during the latter part of the 19th century. This change occurred during a period characterized by repeated crop failures in the area, and the need for replacement seed after severe crop failure is most likely the

cause of the observed genetic change. This adds to the results of studies of more recent crop failures suggesting that genetic composition, but not genetic diversity, is primarily affected by severe crop failure.

Acknowledgement

This work was funded by the Norwegian institute of Science and Technology (NTNU), the Helge Ax:son Johnsons Foundation, the Hem i Sverige-fonden Foundation, the CF Lundström Foundation and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) grant number 2018-02845. Plant material was kindly provided from Mustiala Agricultural College with the help of Annika Michelson and Hannu Ahokas and from Tromsø University Museum with the help of Torbjørn Alm.

Competing Interests

The authors declare no competing interests.

Data archiving

Genotype data available from the Dryad Digital Repository (doi:10.5061/dryad.qv9s4mw9b)

706 **References**

707

708 Abdel-Ghani AH, Parzies HK, Omary A, Geiger HH (2004) Estimating the
709 outcrossing rate of barley landraces and wild barley populations collected
710 from ecologically different regions of Jordan. Theor Appl Genet 109: 588–
711 595.
712 Bergman I, Hörnberg G (2015) Early Cereal Cultivation at Sámi
713 Settlements: Challenging the Hunter–Herder Paradigm? Arctic Anthropol
714 52: 57-66.
715 Bjørnstad A, Abay F (2010) Multivariate patterns of diversity in Ethiopian
716 barley. Crop sci 50: 1579-1586.
717 Dribe M, Olsson M, Svensson P (2015) Famines in the Nordic countries,
718 AD 536 - 1875. No. 138. Lund University, Department of Economic
719 History.
720 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of
721 individuals using the software STRUCTURE: a simulation study. Mol Ecol
722 14:2611-2620.
723 Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of
724 programs to perform population genetics analyses under Linux and
725 Windows. Mol Ecol Resour 10: 564-567.

726 Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance
727 inferred from metric distances among DNA haplotypes: Application to
728 human mitochondrial DNA restriction data. Genetics 131: 479-491.
729 Falush D, Stephens M, Pritchard JK (2007) Inference of population
730 structure using multilocus genotype data: dominant markers and null alleles.
731 Mol Ecol Notes 7: 574-578.
732 Ferguson ME, Jones RB, Bramel PJ, Domínguez C, Torre do Vale C, Han J
733 (2012) Post-flooding disaster crop diversity recovery: a case study of
734 Cowpea in Mozambique. Disasters 36: 83–100
735 Finell JM, Burman GD, Rehausen W von, Fogelmarck SU, Sjöstedt U,
736 Hummel D et al. (1869) Hushållsgillenas årsberättelser Norrbottens läns
737 hushållningssällskaps handlingar 1869: 43-66.
738 Fogelqvist J, Nittyvuopio A, Ågren J, Savolainen O (2010) Cryptic
739 population genetic structure: the number of inferred clusters depends on
740 sample size Mol Ecol Resour 10: 314-323.
741 Forsberg N, Russell J, Macaulay M, Leino M, Hagenblad J (2015) Farmers
742 without borders—genetic structuring in century old barley (Hordeum
743 vulgare). Heredity 114: 195-206.
744 Flygare I (2011) The structure of agriculture. In: Jansson U, Wastenson L,
745 Aspenberg P (eds) National atlas of Sweden. Agriculture and forestry in
746 Sweden since 1900 - a cartographic description. Norstedt: Stockholm. pp
747 58-70.

748 Fuentes F, Bazile D, Bhargava A, Martínez E (2012) Implications of
749 farmers' seed exchanges for on-farm conservation of quinoa, as revealed by
750 its genetic diversity in Chile. J Agric Biol Sci 150: 702-716.
751 Grotenfelt G (1896) Landtbruket i Finland: en öfversikt. Hagelstam:
752 Helsingfors.
753 Groth Ö (1984) Norrbotten 1 In: Norrbottens historia.
754 Skrivarförlaget/Norrbottens bildningsförbund: Luleå
755 Grym E (1959) Från Tornedalen till Nordnorge. Luleå Bokförlag: Luleå
756 Hagenblad J, Zie J, Leino MW (2012) Exploring the population genetics of
757 genebank and historical landrace varieties. Genet Resour Crop Evol. 59:
758 1185-1199.
759 Hagenblad J, Morales J, Leino MW, Rodríguez-Rodríguez AC (2017)
760 Farmer fidelity in the Canary Islands revealed by ancient DNA from
761 prehistoric seeds. J Archaeol Sci 78:78-87.
762 Hellström P (1917) Norrlands jordbruk. Almqvist & Wiksell: Uppsala
763 Häger O, Torell C, Villius H (1978) Ett satans år: Norrland 1867. Sveriges
764 radio: Stockholm
765 Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and
766 permutation program for dealing with label switching and multimodality in
767 analysis of population structure. Bioinformatics 23:1801-1806.

768 Jantunen J, Ruosteenoja K (2000) Weather conditions in northern Europe in
769 the exceptionally cold spring season of the famine year 1867. Geophysica
770 36: 69-84.
771 Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal
772 components: a new method for the analysis of genetically structured
773 populations. BMC Genet. 11: 1-15.
774 Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of
775 genome-wide SNP data. Bioinformatics 27: 3070-3071.
776 Jones RB, Bramel P, Longley C, Remington T (2002) The need to look
777 beyond the production and provision of relief seed: Experiences from
778 Southern Sudan. Disasters 26: 302–315.
779 Jones H, Civan P, Cockram J, Leigh FJ, Smith LMJ, Jones MK et al. (2011)
780 Evolutionary history of barley cultivation in Europe revealed by genetic
781 analysis of extant landraces. BMC Evol Biol 11.1: 320
782 Josefsson T, Hörnberg G, Liedgren L, Bergman I (2017) Cereal cultivation
783 from the Iron Age to historical times: evidence from inland and coastal
784 settlements in northernmost Fennoscandia. Veg Hist Archaeobot 26: 259-
785 276.
786 Karlsen SR, Elvebakk A, Høgda KA, Johansen B (2006) Satellite-based
787 mapping of the growing season and bioclimatic zones in Fennoscandia.
788 Global Ecology and Biogeography 15: 416-430.

789 Laestadius LL (1824). Om möjligheten och fördelen af allmänna
790 uppodlingar i Lappmarken. Zacharias Haeggström: Stockhholm.
791 Leino MW (2010) Frösamlingar på museum. Nordisk museologi 1: 96-108.
792 Leino MW, Hagenblad J (2010) Nineteenth century seeds reveal the
793 population genetics of landrace barley (*Hordeum vulgare*). Mol Biol Evol
794 27: 964-973.
795 Leino MW, Hagenblad J, Edqvist J and Strese EMK (2009) DNA
796 preservation and utility of a historic seed collection. Seed Sci Res 19: 125-
797 135.
798 Lempiäinen-Avci M, Lundström M, Huttunen S, Leino MW, Hagenblad J
799 (2018) Archaeological and historical materials as a means to explore
800 Finnish crop history. Environmental Archaeology 1-16.
801 Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y, Schaal BA (2006)
802 Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple
803 independent domestications of cultivated rice, *Oryza sativa*. Proc Natl Acad
804 Sci U S A 103: 9578-9583.
805 Moragues M, Comadran J, Waugh R, Milne I, Flavell AJ, Russell JR (2010)
806 Effects of ascertainment bias and marker number on estimations of barley
807 diversity from high-throughput SNP genotype data. Theor Appl Genet 120:
808 1525-1534.
809 Nei M (1973) Analysis of gene diversity in subdivided populations. Proc
810 Natl Acad Sci U S A 70: 3321-3323.

811 Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and
812 genetic variability in populations. Evolution 29: 1–10.
813 Nelson MC (1988). Bitter Bread: the Famine in Norrbotten 1867-1868. PhD
814 thesis Uppsala University.
815 Oliveira HR, Campana M, Jones H, Hunt H, Leigh F, Lister DL et al (2012)
816 Tetraploid wheat landraces in the Mediterranean basin: taxonomy, evolution
817 and genetic diversity. PLoS One 7:e37063.
818 Olsen KM, Schaal BA (1999) Evidence on the origin of cassava:
819 phylogeography of *Manihot esculenta*. Proc Natl Acad Sci U S A 96: 5586-
820 5591.
821 Osvald H (1959). Åkerns nyttoväxter. Sv. litteratur: Stockholm.
822 Papa R, Attene G, Barcaccia G, Ohgata A, Konishi T (1998) Genetic
823 diversity in landrace populations of *Hordeum vulgare* L. from Sardinia,
824 Italy, as revealed by RAPDs, isozymes and morphophenological traits. Plant
825 Breeding 117: 523-530.
826 Pitkänen K (1992) The patterns of mortality during the Great Finnish
827 Famine in the 1860s. Acta Demographica 1992: 81-102.
828 Popescu AA, Huber KT, Paradis E (2012) ape 3.0: new tools for distance
829 based phylogenetics and evolutionary analysis in R. Bioinformatics, 28,
830 1536–1537.
831 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population
832 structure using multilocus genotype data. Genetics 155:945-959.

833 Rosenberg NA (2004) DISTRUCT: a program for the graphical display of
834 population structure. Mol Ecol Notes 4: 137-138.
835 Rydstedt G, Burman GD, Jacobsson JD, Schönfelt RF, Rehausen W von,
836 Berghmark D et al. (1868). Hushållsgillenas årsberättelser. Norrbottens läns
837 hushållningssällskaps handlingar 1868: 46-75.
838 Statistiska centralbyrån Landshövdingeämbetet i Norrbottens län (1856-
839 1905). Femårsberättelser Norrbottens län, 1856-1905. Bidrag till Sveriges
840 officiella statistik. H Kungl Maj:ts befallningshafvandes femårsberättelser.
841 Stockholm.
842 Sperling L (2001) The effect of the civil war on Rwandas bean seed systems
843 and unusual bean diversity. Biodivers Conserv 10: 989-1010.
844 Tilastokeskus (1875). Finlands Officiella Statistik II: Öfversigt af Finlands
845 ekonomiska tillstånd åren 1866-1870. Helsinki.
846 Yelome IO, Audenaert K, Landschoot S, Dansi A, Vanhove W, Silue D et
847 al. (2018) Analysis of population structure and genetic diversity reveals
848 gene flow and geographic patterns in cultivated rice (*O. sativa* and *O.*
849 *glaberrima*) in West Africa. Euphytica, 214:215
850 Varma S, Winslow M (2004) Healing wounds: How the international
851 centers of the CGIAR help rebuild agriculture in countries affected by
852 conflicts and natural disasters. Consultative Group on International
853 Agricultural Research (CGIAR), Washington, DC.

854 Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of
855 population structure. Evolution 38: 1358-1370.
856

857 **Table 1: Accession list with geographical information and genetic diversity for the accessions used in the study**

Accession	Origin	Country	Harvest Year	Age class	Lat	Long	Altitude (m.a.s.)	N_s	Nei's h	Mean PC-dispersion	Var PC-dispersion
TR1	Storjord	Norway	1869	Early	68.2	16.1	10	18	0.104	3.864	0.459
TR5	Ibestad	Norway	1869	Early	68.8	17.2	10	16	0.084	3.547	0.379
TR7	Balsfjord	Norway	1869	Early	69.3	19.3	50	18	0.117	4.109	0.536
TR8	Komagfjord	Norway	1869	Early	70.3	23.4	10	16	0.147	4.641	0.236
MU1	Rovaniemi	Finland	1893	Late	66.5	25.7	80	20	0.132	4.531	0.419
MU69	Muonio	Finland	1870	Early	68.0	23.7	240	15	0.121	4.268	1.106
NM1587	Jokkmokk	Sweden	1867	Early	66.6	19.8	250	7	0.126	3.889	0.156
NM1597	Kvikkjokk	Sweden	1867	Early	67.0	17.7	310	17	0.046	2.640	1.141
NM599	Matarengi	Sweden	1896	Late	66.4	23.7	50	19	0.122	4.218	0.919
NM633	Pajala	Sweden	1896	Late	67.2	23.4	170	18	0.043	2.302	1.191
NM668	Kurrokveik	Sweden	1896	Late	66.1	17.9	420	18	0.152	4.754	0.286
NM669	Vuollerim	Sweden	1896	Late	66.4	20.6	110	20	0.160	4.922	0.235

NM727	Sandön	Sweden	1896	Late	65.5	22.4	5	18	0.135	4.439	0.362
NM751	Kirtijokki	Sweden	1896	Late	67.9	23.5	200	15	0.125	4.223	0.433
NM789	Wouno	Sweden	1896	Late	65.8	24.1	10	20	0.082	3.421	0.649
NM798	Kuttainen	Sweden	1896	Late	68.4	22.8	300	20	0.105	3.869	1.034
Total								275	N/A	4.364	0.763

858 [a] Number of individuals remaining from each accession after quality control

Table 2: p and r^2 values for regression analysis of cluster membership. Negative adjusted r^2 values are given as 0 in the table.

Variable	*Structure* cluster membership (K = 2)				Principal component analysis (PC1)			
	Accession-level		Individual-level		Accession-level		Individual-level	
	p	r^2	p	r^2	p	r^2	p	r^2
Altitude	0.978	0.000	0.611	0.000	0.996	0.000	0.844	0.000
Latitude	0.366	0.000	< 0.001	0.039	0.338	0.000	0.001	0.035
Longitude	0.208	0.047	< 0.001	0.048	0.247	0.030	< 0.001	0.046
Country	0.589	0.000	< 0.001	0.051	0.566	0.000	< 0.001	0.048
Harvest year	0.035	0.23	< 0.001	0.134	0.039	0.219	< 0.001	0.119
Sample plate	N/A	N/A	0.193	0.005	N/A	N/A	0.037	0.017

Table 3: AMOVA of the genotypes of the studied accessions

Group	Source of variation	d.f.	Sum of Squares	Variance components	% of variation
Temporal classes	Among temporal classes	1	73.914	0.3632	3.85
	Among accessions within temporal classes	14	370.03	1.07453	11.38
	Within populations	259	2072.663	8.00256	84.77
	Total	274	2516.607	9.44028	
Countries	Among countries	2	79.539	0.15261	1.64
	Among accessions within countries	13	364.405	1.1724	12.57
	Within populations	259	2072.663	8.00256	85.79
	Total	274	2516.607	9.32757	

867 **Table 4: Number of informative SNPs and ordinary Procrustes sum of squares (OSS)**

868 **in the subsets**

869

	Informative SNPs		Procrustes analysis
Set	Average (s.d.)	Proportion	Average OSS (s.d.)
Full	152 (NA)	1	NA (NA)
1 individual/population	114.9 (8.279)	0.756	13.640 (2.370)
6 individuals/population	148.6 (1.174)	0.978	2.723 (0.872)
Artificial pooling	33 (NA)	0.217	4.141 (NA)

870

871

872 **Figure legends**

873

874 **Figure 1:** Genetic structure from 20 individual *structure* simulations

875 merged with the CLUMPP software for K = 2.

876

877 **Figure 2:** Map of spatial and temporal clustering from *structure* simulations

878 for K = 2.

879

880 **Figure 3:** PCA of SNP genotypes. Black circles signify accessions collected

881 1893-1896 and grey triangles signify accessions collected 1867-1870.

882 Results from analysis at A) accession level, and B) individual level.

883

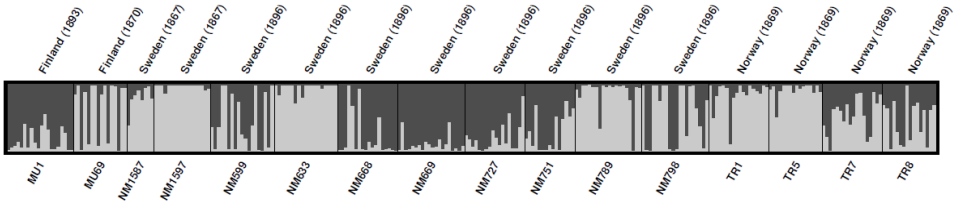
884 **Figure 4:** Geographic distribution of three most common shared genotypes

885 shown as bar plots at the geographic location of origin. The height of the

886 bars shows prevalence of the different shared genotypes.

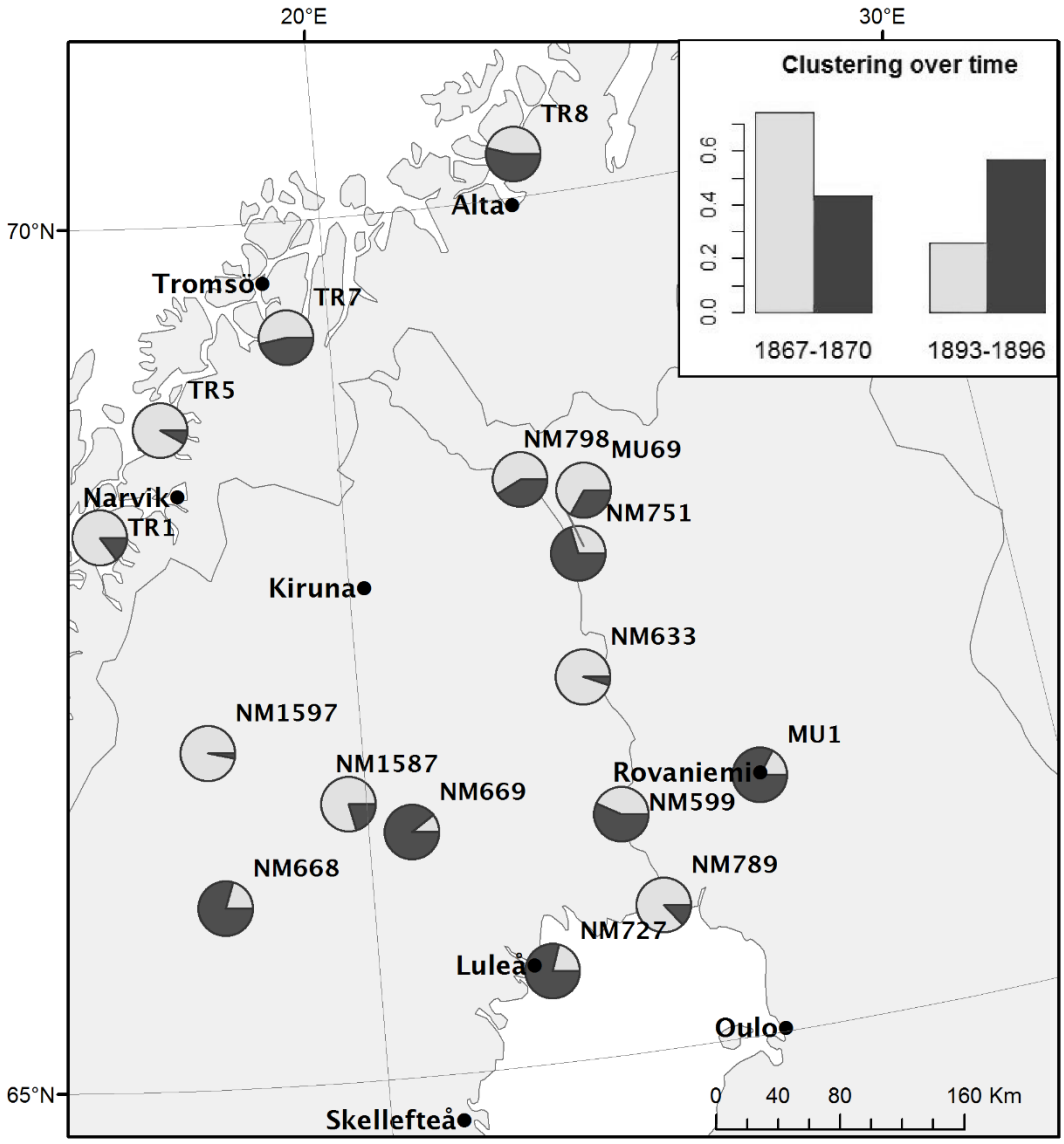
887

888 Figure 1
889



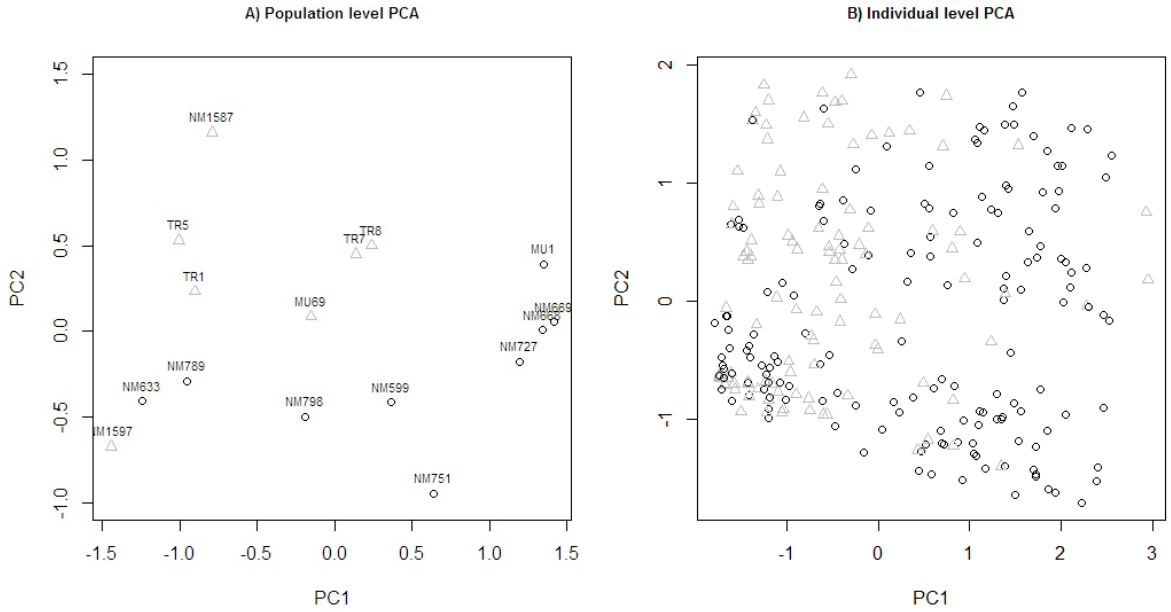
890
891

892 Figure 2



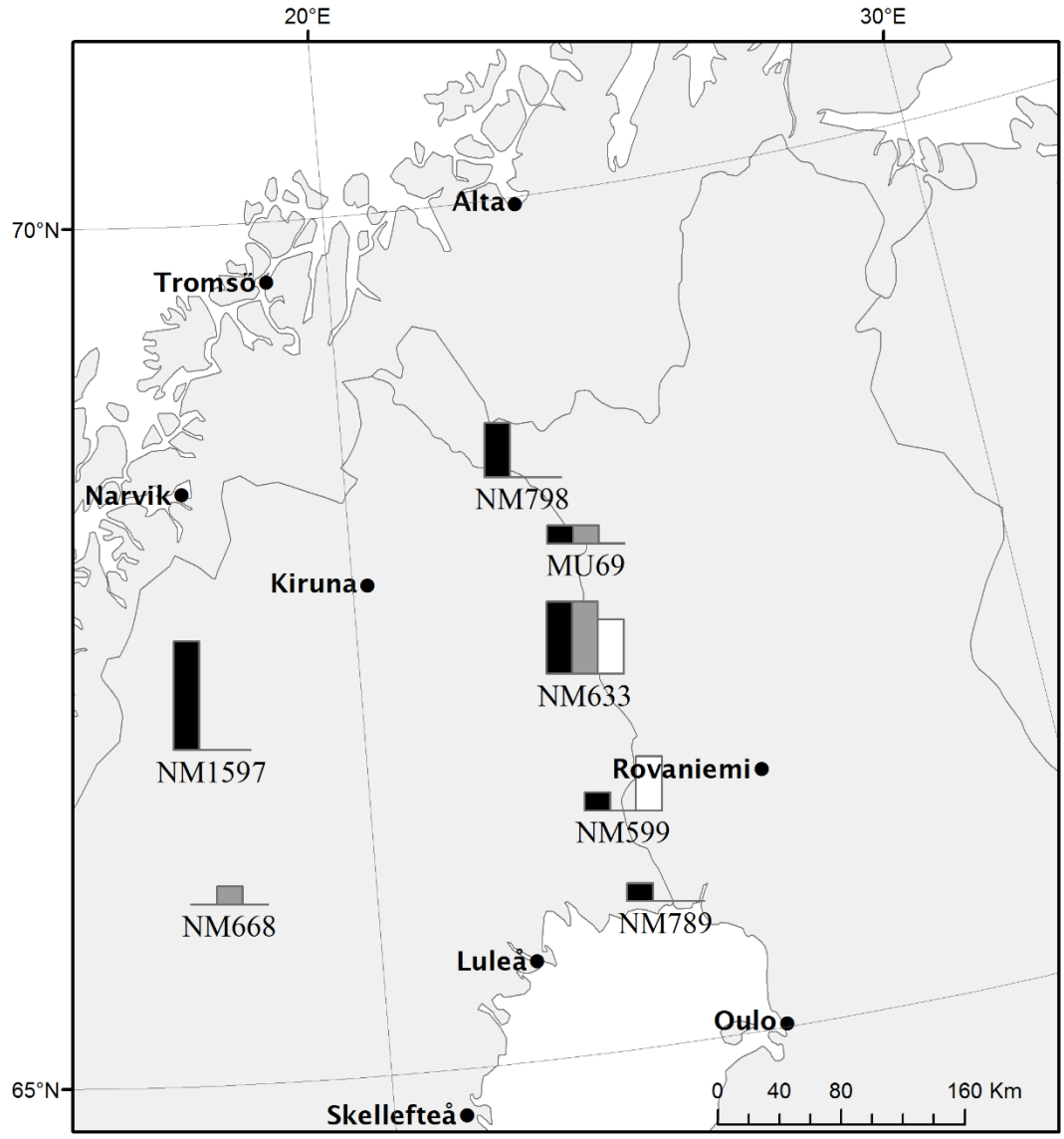
893
894

895 Figure 3



896

897 Figure 4



898