Analyses of phenotypic and genomic based genetic diversity highlight useful trait variation for the improvement of grain and protein yield of pea (*Pisum sativum* L.) grown in Norway

Stefano Zanotto [1]*, Chloé Grieu[3], Jon Arne Dieseth[2], Muath Alsheikh[2], Wendy Waalen[3], Susanne Windju[2] and Anne Kjersti Uhlen[1]

[1]Department of Plant Science, Faculty of Biosciences, Norwegian University of Life Sciences (NMBU), P.O. Box 5003, N-1432 Ås, Norway
[2], Graminor AS, 2322 Ridabu, Norway
[3]Division of Food Production and Society, Norwegian Institute for Bioeconomy Research (NIBIO), Ås, Norway

*Corresponding author: stefano.zanotto@nmbu.no

Abstract

Pea (*Pisum sativum* L.) plays a central role for achieving the goal of increased self-sufficiency of plant-based protein for food and feed in Norway. This study characterized 36 pea cultivars of mainly Nordic and eastern European origin grown in a multi-environment field trial in Norway for various agronomic traits including grain and protein yield. The same accessions were also genotyped with a single nucleotide polymorphism (SNP) chip array to investigate the genetic diversity both at the phenotypic and genomic levels. Variance component (VC) analyses found that genotype (G) and genotype by year by location (G × Y x L) interaction VC were the main sources of variation for all the considered traits, with G x Y x L being strongly affected by the different environmental conditions which characterized the two test years. Measurements of genetic distance based on phenotypic and genomic data were significantly correlated and distinguished the material in three main clusters which were partially associated with their geographic origin. This study identified pea cultivars which outperformed the reference Norwegian variety Ingrid both in term of grain and protein yields, which can be successfully grown in Norway and used as source of genetic diversity for future breeding efforts targeting the Norwegian environment.

Keywords

G X E, breeding, population structure, adaptation, multi-location field trials

Acknowledgments

We thank the staff at SKP Friland at Vollebekk and the staff at Bjørke and Apelsvoll for the excellent agronomic management of the field trials. We also acknowledge Guro Augusta Rogstad Sørheim, Chloé Grieu, and Vilde Gadderud for help with data collection at Vollebekk, Apelsvoll, and Bjørke, respectively.

Conflict of interest

The authors declare no conflict of interest.

Funding: The research was supported by the project “Future Protein Crops” funded by the Foundation for Research Levy on Agricultural Products (FFL), the Agricultural Agreement Research Fund (JA) in Norway, and industry partners (NFR project no. 326701). The research was also supported by the project “GreenPlantFood” funded by the Reseasch Council of Norway and industry partners (NFR project no. 319049).

Data Availability: The data presented in this study are available on request from the corresponding author.

Author contributions

Stefano Zanotto: conceived the idea for this study, formal analyses, writing original draft, review, and editing., **Jon Arne Dieseth**: investigation, review and editing, **Muath Alsheikh**: investigation, review and editing, **Susanne Windju**: investigation, review and editing, **Wendy Waalen**: investigation, review and editing, **Chloé Grieu**: investigation, review and editing **Anne Kjersti Uhlen**: project administration and funding, review, and editing,

ORCID

Anne Kjersti Uhlen https://orcid.org/0000-0002-7655-9754

Stefano Zanotto https://orcid.org/0000-0001-6588-7237

Chloé Grieu https://orcid.org/0000-0002-3291-5868

Wendy Waalen https://orcid.org/0000-0001-9307-417X

Introduction

Field pea (*Pisum sativum* L.) is a temperate grain legume well adapted to cool temperatures and can be grown also on higher latitudes under long days and short growing seasons. The seed has a high protein content and is well suited both for human foods and animal feed. The pea crop has been cultivated at high latitudes in the Nordic countries since historical times. The production volume has varied, but it has persistently occupied a rather small part of the arable land (about 1.3 % of the total arable area was cultivated with pea and faba bean in 2018, Abrahamsen et al., 2018) that are dominated by cereals. However, renewed interest has been seen during the last decade in the Nordic countries as well as in the EU, driven by demand for high protein crops to replace imported plant protein (EU Farm to Fork strategy as part of the European Green Deal). Increase in the production areas has also been driven by the agronomical and environmental benefits of including grain legumes in a cropping system (Stagnari et al., 2017; van Loon et al., 2023). In Norway pea is grown as a spring crop, due to the severe winters which hinder the use of winter types. The main pea growing region in Norway is located at approximately 60°N, which represents the highest latitude of large-scale cultivation for this crop worldwide.

High-yielding and adapted pea varieties are a prerequisite for successful cultivation and farmer adoption. Specifically, for the pea cultivation regions in southeast of Norway, there is a need for varieties adapted to the prevalent weather conditions characterized by cool spring and then rapidly increasing temperatures together with long days in the first part of the life cycle, and relatively cool and humid weather with decreasing temperatures during the maturation and harvest period. The low spring temperatures can delay emergence and seedling development and thus move maturation and harvest later in the autumn with risk of unfavorable conditions due to decreasing temperatures and increasing precipitation as autumn progresses. On the other hand, the long photoperiod during summer combined with the closer to optimal temperatures in this period can accelerate plant development (Carlson-Nilsson et al., 2021), giving a positive compensating effect. As pea is a facultative long-day plant (Lake et al., 2021), these effects can be strengthened at higher latitudes.

The pea ideotypes for such northern latitudes are not well defined, but the ability to mature early and at decreasing temperatures are likely among the most important. Vanhala et al. (2016) investigated flowering time adaptation in Swedish pea landraces and corresponding diversity in main flowering time genes. The results showed that flowering time adaptation had a complex genetic background. However, strong genetic control of physiological traits related to plant growth and development like days to flowering, days to maturity or flowering duration are reported in many studies, i.e, through high heritability estimates (Ali et al., 2021; P. Annicchiarico et al., 2017; Tolessa, 2017; Uhlarik et al., 2022, Sadras et al., 2019).

High and stable yields combined with improved ability to withstand lodging are among the main breeding goals. Today, semi-leafless varieties are dominating, having more tendrils but large stipules (*afaf StSt*) improving the resistance to lodging without sacrificing in leaf area to produce high yield (Tran et al. 2022). Several studies have focused on physiological traits important for yield potential in pea, such as exploring the interrelationships between phenological development, duration of the growth stages, formation of the yield components and yield stability. In many studies, seed yields were positively associated with yield components as thousand seed weight (TSW) and number of pods per plant, but also to plant height, days to flowering and days to maturity (Ali et al. 2021, Carlson-Nilsson et al. 2021, Uhlarik et al. 2022).

Protein content is an important quality trait in pea, both when used for food and feed. Daba and Morris (2022) recently reviewed the variation in protein content in pea and found a general variation from 13-38 % with the most common variation from 20-30%. Many investigations have shown negative correlations between protein and starch content as well as between protein content and seed yield. However, these relationships were not always negative, and positive correlations between protein content and yield were also reported (Červenski et al., 2017; Karjalainen and Kortet, 1987). In other studies, the protein content was reported to be independent of the grain yield (Carlson-Nilsson et al., 2021; Ferrari et al., 2016; Mohammed et al., 2018a). Heritability estimates for protein content have been shown to be moderately high (reviewed by Daba and Morris 2022). Jha et al. (2013) studied 169 pea accessions and reported that the eastern European accessions were generally higher in protein content compared with western European and Canadian accessions. A similar study by Tzitzikas et al. (2006) reported that brown pea genotypes had higher average protein content than green or yellow (averages of 26.7%, 21.5% and 21.8%, respectively).

Pea varieties which are high-yielding and adapted to the short growing season are needed to support the increase in grain legume production in Norway. However, there is no pea breeding program in Norway to date, and the available varieties on the market are developed by breeding companies which direct their activities towards regions at latitudes further south. In this study, a set of modern spring pea varieties from breeding companies in northern, central and eastern regions of Europe were collected and tested for agronomic performance in multi-location field trials at northern latitudes of approximately 60° N in Norway. The aims were to analyze the effect of genotype (G), location (L) and genotype by location (G X L) on agronomic traits important for the cultivation of peas in Norway, subjected to cool and wet conditions during pod development and maturation. This analysis was complemented by using molecular markers to measure genetic diversity at the genomic level. The aims of this study?) were to:

(i) identify genotypes that could be recommended for cultivation in Norway

(ii) assist breeding of pea cultivars adapted to the Norwegian conditions, by pointing at strategies for pre-breeding and selecting genotypes to be used as parents for crossings to initiate a future Norwegian breeding program.

Material and methods

Plant material and field trials

Varieties and advanced breeding lines of field peas were collected from European breeders and tested in field trials at three locations in Norway in 2022 and 2023. The selection of varieties was done through contact with breeding companies in northern and central Europe, which provided varieties of spring peas that were assumed to be adapted to Nordic climates. All were of the semi-leafless type (*afafStSt*). The number of varieties tested were 44 and 51 in the seasons 2022 and 2023, respectively. An overview of the varieties tested in the different seasons and of the breeding companies providing the material is given in the supplementary material (**S1**). A set of 36 genotypes (varieties and breeding lines) common across all test locations and years were considered for the analyses described below. The three locations were Vollebekk (research farm at Norwegian University of Life Science, Ås, 59^{o}39′ N, 10^{o}45′ E), Bjørke (research farm of Graminor AS, Hamar, 60^{o}80′ N, 11^{o}20′ E), and Apelsvoll (research farm at the Norwegian Institute of Bioeconomy Research, Toten, 60^{o}70′ N, 10^{o}87′ E). All locations are within the main grain legumes and cereals producing region in southeast of Norway. Among these, Apelsvoll is representative for growth conditions of slightly cooler temperatures and shorter duration of the growth season while Vollebekk is representative for growth conditions of slightly warmer temperatures, early sowing time and longer growth season. An overview of the daily mean temperature and precipitation at the three locations in the two seasons is provided in the supplementary material (**S2**). The field trials were laid out as a randomized complete block design with three replicates. Seed rate was set to 100 viable seeds/m2, and the trials were sown when the soil reached optimal moisture content in the spring. Plot size varied at the different locations, from 9,6m^2 at Bjørke to 12,8m^2 at Apelsvoll. Weeds were controlled by using the herbicide Basagran (Bentazon, 870 g/kg). The field trials were carried out on fertile soil of optimal pH. No fertilization nor *Rhizobium* inoculation was applied, but PK fertilized was applied to secure optimal plant growth. An overview of management practices at the different locations is given in the supplementary material (**S3**).

Phenotyping

Plant development stage was assessed weekly using the BBCH scale (Weber and Bleiholder, 1990) on single plots, and the data was used to calculate the number of days from sowing to BBCH60 (first flower open), BBCH70 (end flowering/first pods final length) and BBCH89 (fully ripe). Plant height was recorded in cm from soil surface to top of canopy, using the average of two measurements per plot. Plant height was recorded weekly from the BBCH60 stage to harvest. Maximum height (PLH_max) and height at harvest (PLH_harv) were used in statistical analyses. The field trials were harvested using a plot combine harvester. A seed lot per plot was collected and dried to moisture content below 15%. Grain ield in kg/ha at 15% moisture (YLD) was calculated. Seed samples taken from single plots was cleaned and the weight of thousand seeds (TSW) was recorded in g. Protein content (PRT) was analysed by Near infrared transmission (NIT) with InfratecTM NOVA instrument (Foss, DK) for the field trials at Apelsvoll and Bjørke. For the field trials conducted at Vollebekk, the seed samples were milled (using Stein Laboratory mill, Fred Stein Laboratories, Inc. Atchison, Kansas) and N content was analysed by the combustion method (Dumas, ISO 16634 (2008)) performed by LABTEK, NMBU, Norway. Protein content was calculated using the conversion factor 6,25 and given as percent of dry weight. Protein yield (PRT_YLD) was calculated from the grain yield and protein content and is given in kg/ha. For all traits phenotyping was done plot-wise for all three replicates per experiment. In some cases, only two full replicates were used. An overview of the phenotyping

methods used for each trait and the number of considered replicates per experiment is given in the supplementary material (**S4**).

Statistical analysis of phenotypic data

A mixed linear model (MLM) with the fixed effects of year, location and their interaction and the random effects of genotype and the various interactions (genotype X year, genotype X location, genotype X year X location and replicate nested within location and year) was fitted for each phenotypic trait by the restricted maximum likelihood (REML) with the breedR package (Muñoz and Sanchez, 2020) using the following formula:

$$Y_{sijk} = \mu + \beta_s + u_i + \gamma_{si} + v_j + \alpha_{sj} + w_{ij} + \delta_{isj} + z_{sik} + \epsilon_{sijk} \; (eq.1, \text{across years})$$

Where: Y_{sijk} is the raw phenotype, μ is the overall mean; β_s is the fixed effect of the *s*-th year; u_i represents the fixed effect of the *i*-th location; γ_{si} is the fixed effect of the interaction between the *s*-th year and the *i*-th location; $v_j \sim N(0, \sigma_v^2)$ represents the random effect of the *j*-th genotype; $\alpha_{sj} \sim N(0, \sigma_\alpha^2)$ represents the random effect of the interaction between the the *s*-th year and the *j*-th genotype; $w_{ij} \sim N(0, \sigma_w^2)$ represents the random effect of the interaction between the *i*-th location and the *j*-th genotype; $\delta_{isj} \sim N(0, \sigma_\delta^2)$ represents the random effect of the interaction between the *s*-th year, the *i*-th location and the *j*-th genotype, $z_{sik} \sim N(0, \sigma_z^2)$ represents the random effect of *k*-th replicate nested within the *s*-th year and the *i*-th location and $\epsilon_{sijk} \sim N(0, \sigma_\epsilon^2)$ is the residual error term.

Variance component (VC) analyses of each phenotypic trait were conducted with the restricted maximum likelihood (REML) procedure using the model described in *eq*. 1. The standard error of each variance component was calculated with the confint() function of the R package stats. Also, the significance of each variance component was assessed as the p-value of the Likelihood Ratio Test (LRT) by dropping the associated single term in the model. As an example, to assess the significance of the genotypic variance component, the genotype term was dropped from the original model and the ANOVA test was performed with the anova() function of the R package stats between the reduced and original models. The absolute values for each VC were then used to calculate the percentage of the total phenotypic variance explained by each VC for each phenotypic trait, as well as the broad sense heritability (h^2) using the following equation:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GxL}^2}{nL} + \frac{\sigma_{GxY}^2}{nY} + \frac{\sigma_{GxYxL}^2}{nYx\,nL} + \frac{\sigma_{Rep}^2}{nR} + \sigma_\epsilon^2/nLxnR} \; (eq.2)$$

where σ_G^2, σ_{GxL}^2, σ_{GxY}^2 , σ_{GxYxL}^2 σ_{Rep}^2 and σ_ϵ^2 are the genotype, genotype by location, genotype by year, genotype by year by location, replicate and residual error variance components and *nL, nY* and *nR* are the number of locations, years and replicates per location, respectively.

The environmental individual effects (at plot level) due to year, location and the various interactions were subtracted from the observed phenotype to obtain adjusted phenotypes which were then used (as genotypic means) to conduct the multivariate analyses described below.

The choice of assigning year and location as fixed effect in model 1 was justified by the facts that: 1) the three test locations represent three climatically distinguished micro environments within the major pea cultivation area in Norway; 2) the two test years 2022 and 2023 were characterized by two distinguished climatic trends that were representative of the climate which characterizes the current

(2022) and predicted (2023) growing seasons in Norway, the latter characterized by a dry spring and early summer and exceptionally wet and cool late summer.

A second mixed model considering each year separately was performed for the grain yield with the fixed effects genotype, location and their interaction and the random effect replicate within location to assess the differences between the evaluated genotypes across locations within the two test years using the following formula:

$$Y_{ijk} = \mu + u_i + v_j + w_{ij} + z_{ik} + \epsilon_{ijk}\,(mod.2, \text{within year})$$

Where: Y_{ijk} is the raw phenotype, μ is the overall mean; u_i represents the fixed effect of the *i*-th location; v_j represents the fixed effect of the *j*-th accession; w_{ij} represents the fixed effect of the interaction between the *i*-th location and the *j*-th accession; $z_{ik} \sim N(0, \sigma_z^2)$ represents the random effect of *k*-th replicate nested within the *i*-th location and $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$ is the residual error term.

All the data was checked for normality of the residuals' distribution and the homogeneity of the variance through the Shapiro Wilk and Levene's tests using the shapiro.test() and leveneTest() functions of the R packages stats and car (R Core Team, 2025), respectively.

Relationships among phenotypic traits

The relationship between the phenotypic traits was investigated as phenotypic and genotypic correlations. Phenotypic correlations were calculated with the Pearson method for all traits across locations and years. The genetic correlation was assessed using a multi-trait model as explained in (Gutierrez et al., 2024). Briefly, for each random effect a full covariance matrix is estimated. The cov2cor() function in the R package stats was used to compute the genetic correlation between traits from the additive genetic variance–covariance matrix. The genetic correlation among traits was extracted and compared with the phenotypic correlation. Descriptive analysis and correlations of the phenotypic data were conducted with the R 4.3.3 software (R Core Team, 2025).

Multivariate analyses of adjusted phenotypes

The matrix of genotypic means for the adjusted phenotypes was scaled and centered before performing multivariate analyses. Centering ensures that each variable has a mean of zero, and scaling standardizes the variance of each variable to 1. Principal component analysis (PCA) and hierarchical clustering analysis (HCA) were performed with PCA() and hclust() function of the factormineR (Husson et al., 2008) and the stats package function in R, respectively.

For HCA the distance matrix was calculated using Euclidean distance with the dist() function of the stats package, and the clustering algorithm used was ward.D2. Dendrograms were created with the fviz_dend() function from the factoextra package (Kassambara and Mundt, 2020) in R. The NbClust() function from the NbClust package (Charrad, et al. 2020) was used to identify the best number of clusters, employing various statistical indexes for this purpose. The results were complemented by the visual assessment of the dendrograms to choose the optimal number of clusters.

Genotyping and SNP filtering

A total of 43 genotypes, (phenotyped in the field during the 2022 test year **S1**) were genotyped with the GenoPea 13.2 K SNP Array chip (Tayeh et al., 2015) by the company SGS Trait Genetics, Gatersleben, Germany. DNA was extracted from pooled leaves of 10 young seedlings for each accession. A total of 11634 SNPs with known map position were identified for this plant material, SNPs with unknown genome position were initially discarded. Then, SNPs with more than 20%

missing values or overall minor allele frequency (MAF) < 0.01 were removed. This resulted in a set of 8859 SNPs of which 8558 (96.6%) were located on chromosomes and 301 on unanchored scaffolds. The number of missing values per marker and per individual, as well as the allele frequency distribution, after filtering, is shown in supplementary material (**S5** and **S6**). Missing data after filtering (0.65% of all sample and SNP genotype calls) were imputed by replacing each missing data point with the mean allele frequency across all accessions per SNP. The distribution of the filtered SNPs along the genome is displayed in the supplementary material (**S7**).

Linkage disequilibrium

The linkage disequilibrium (LD) for each single chromosome was calculated using only the SNP markers with known chromosomal position. LD was estimated as the squared allele frequency correlations (r^2) for each pairwise combination of markers distanced within 1 Mbp with the pcor.shrink () function of the package corpcor (Schafer, et al., 2021) in R studio. LD was then plotted against the genomic distance between marker pairs and the LD decay was visualized with the LOESS regression model. The LD decay was estimated as the point where the fitted curve reached half of its maximum value (Gutierrez et al., 2024).

Analysis of population genetic structure

The population structure was analyzed with the snmf() and Q() functions of the R package LEA (Frichot and François, 2015) which output the results in a similar way as the STRUCTURE software although it employs different algorithms and was indicated as being more accurate for self-pollinating species (Crosta et al., 2023). Filtered SNP data, as described in the section above, were used as input in the analyses. The optimal number of genetic clusters was selected after visual assessment of the cross-entropy parameter plot estimated through cross-validation. In addition, a discriminant analysis of principal component (DAPC) was performed with the R package adegenet (Jombart and Ahmed, 2011) on the same dataset (matrix of filtered and imputed SNPs) to visualize pattern of structural variation in the genomic data in the bidimensional space. DAPC analysis first identifies the best number of clusters in the data and then maximize the differences between clusters. The find.clusters() function first performs a PCA on the data, then performs a clustering algorithm (k-means) that considers a range of potential K values to group the data. The goodness of fit of each K value is indicated by a BIC score, where the lowest BIC indicates the best fit. The dapc() function performs a discriminant analysis with the chosen K value, finding the axis of variation that best discriminates among the groups. The results are then visualized similarly to a PCA plot, however, DAPC analysis typically results in a stronger grouping than that shown by PCA.

Relationship between phenotypic and genetic diversity

A reduced SNP matrix containing only the data for the 36 genotypes phenotyped in both the 2022 and 2023 field trials (**S1**) was used for further analyses of genetic structure through both ordination (PCA) and classification (HCA). The method used for HCA was the same as that described for the phenotypic dataset. Finally, the consistency between estimates of dissimilarity provided by the morphophysiological traits measured in the field and molecular markers for the 1296 possible pairs of distances (36 distances based on genotypes by 36 distances based on phenotypes) was assessed by simple correlation analysis as well as through visual comparison of the dendrograms and biplot conducted on (adjusted) phenotypic, and genotypic data respectively.

Results

Trait variation and interrelationship

The results of the analysis of variance across years (*eq.1*) found that year had a significant effect for all the phenotypic traits (**table 1**). The first testing year (2022) had more favorable growth and harvesting conditions which resulted in higher grain yield, protein content and protein yield across all testing locations. The maximal plant height was also higher in 2022 compared to 2023. Plant height at harvest was higher in 2022 at Vollebekk and Apelsvoll than at Bjørke, while in 2023 plants were affected by heavy lodging at Vollebekk and Apelsvoll. Finally in 2023 there was quicker plant development in terms of the measured stages of the BBCH scale than in 2022, likely because of the early drought during the spring of 2023. The variation of all the phenotypic traits for the two years and locations separately, is shown in the supplementary material (**S8**). The fixed effects of location and of the interaction of year and location were significant for all traits but maximal plant height (**table 1**). In 2022 grain yield and protein yield were considerably lower at Bjørke compared to Apelsvoll and Vollebekk. TSW was significantly lower at Vollebekk in 2022 compared to Apelsvoll and Bjørke, while the opposite was observed in 2023. Similarly to TSW, protein content was lower at Vollebekk in 2022 than at Apelsvoll and Bjørke, while the lowest protein yield was measured at Bjørke in 2022. For all the other traits (excluding maximal plant height) the effect of location, although significant, did not follow a clear trend and the year by location interaction likely played a stronger role as proved by the different patterns observed within single years (**S8**).

Table 1. F values for the fixed effects variance components (±SE) for the random effects estimated with the model in *eq.* 1 and line mean broad-sense heritability (±SE) for the phenotypic traits registered for 36 pea genotypes across two years and three locations.

	YLD	TSW	PRT	PRT_YLD	PLH_max	PLH_har	BBCH60	BBCH70	BBCH79
Fixed Effect									
L‡	45.31***	6.78***	16.402**	34.62***	ns†	21.58***	23.59***	2.60*	11.80***
Y	182.88***	146.64***	4.155*	143.55***	102.31***	197.12***	638.48***	616.77***	51.85***
L X Y	31.41***	42.27**	4.547*	8.65**	ns	91.40***	7.85**	61.88***	21.94***
Variance Component									
G	620.05 (474.50)	536.89 (136.90)***	0.96 (0.30)***	30.40 (20.61)	38.38 (13.83)***	36.037 (14.66)***	0.73 (0.35)**	3.87 (1.29)***	5.79 (1.45)***
L X G	ns	ns	0.45 (0.16)***	ns	ns	18.26 (11.09)*	ns	3.06 (0.70)***	ns
Y X G	746.40 (536.39)*	27.73 (24.78)	ns	ns	24.47 (8.67)***	ns	ns	0.38 (0.27)*	ns
Y X L X G	3511.48 (509.05)***	126.93 (25.74)***	0.25 (0.11)***	169.01 (26.52)***	13.03 (3.87)***	38.61 (12.19)***	1.52 (0.34)***	0.63 (0.36)**	6.97 (1.34)***
Rep	265.09 (109.07)***	20.12 (8.97)***	0.48 (0.16)***	30.54 (11.03)***	11.74 (4.66)***	9.47 (3.91)***	0.43 (0.17)***	1.27 (0.42)***	4.04(1.31)***
Res	2330.62 (161.37)	226.32 (17.07)	0.50 (0.04)	68.27 (5.44)	26.37 (2.34)	86.77 (6.472)	3.21 (0.25)	3.25 (0.26)	9.05 (0.72)
h^2	0.37(0.21)	0.91 (0.02)	0.71 (0.07)	0.38 (0.21)	0.65 (0.35)	0.58 (0.17)	0.49 (0.13)	0.66 (0.27)	0.62 (0.16)

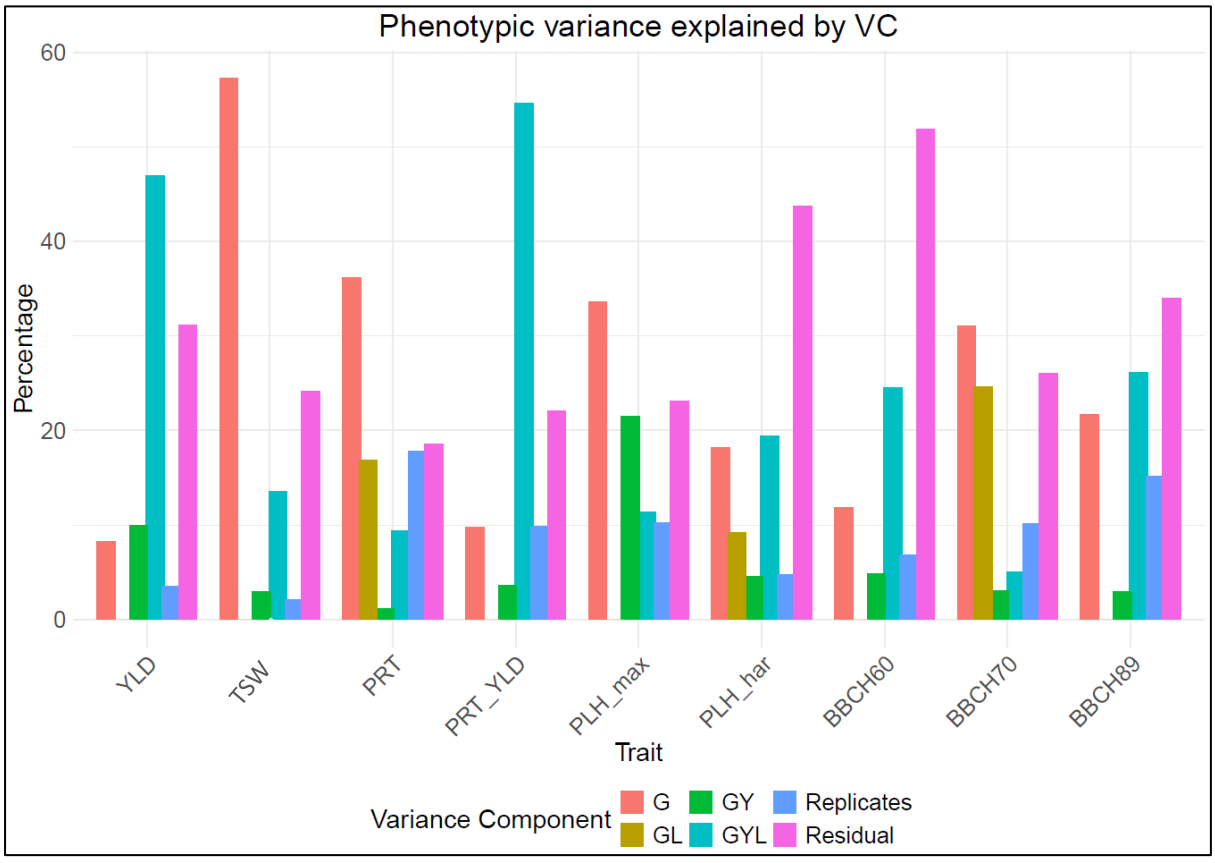
*, **, *** significance at the 0.1,0.05 and 0.01 levels, respectively.

†ns, not significant.

‡ L = Location; Y = Year; L X Y = Location X Year interaction; G = Genotype; L X G = Location X Genotype interaction; Y X G = Year X Genotype interaction; Y X L X G = Year X Location X Genotype interaction; Rep = Replicate; Res = Residuals.

The variance component analysis (**table 1**) reported a significant effect of the genotype variance components (VC) for all traits but grain yield and protein yield which, nonetheless, had a p-value just above 0.1. The genotype by location VC was significant only for protein content, plant height at harvest and BBCH70, while the genotype by year VC was significant only for grain yield, maximal plant height and BBCH70. For all traits the genotype by location by year VC was strongly significant, proving the strong effect of the within year genotype by location interaction. The replicate VC was also significant for all traits although its magnitude was low relatively to the other VC (**Figure 1**).

Figure 1. Percentage of phenotypic variance explained by each variance component. YLD = Grain Yield; TSW = Thousand Seed Weight; PRT = Protein Content; PRT_YLD = Protein Yield; PLH_max = Maximal Plant Height; PLH_har = Plant Height at Harvest; BBCH60 = Full Flowering; BBCH70 = end flowering/first pods final length and BBCH89 = fully ripe.



The genotypic VC had the strongest contribution to the total phenotypic variance for TSW, protein content, maximal plant height and BBCH70 (**Figure 1**) which also had the highest broad sense heritability (> 0.60, **table 1**). Grain yield and protein yield were affected by a strong within year genotype by location interaction as shown in **Figure 2**, proving that different genotypes performed differently at different locations within the 2022 and 2023 growing seasons, which were strongly affected by the different weather conditions described above (**S2**). Plant height at harvest, BBCH60, BBCH89 and to a minor extent BBCH70 were characterized by a high residual error VC because of the strong within trial (field) variation in lodging which affected mainly the recording of plant height at harvest, and the plant development after flowering.

The analysis of variance for grain yield within years and across locations (*eq. 2*) found a significant effect of genotype, location and their interaction in both years (results not shown). In 2022, as mentioned above, the grain yield was significantly lower at Bjørke than at Apelsvoll and Vollebekk while in 2023 the grain yield was significantly lowest at Vollebekk. Genotype ranking was strongly different between the two testing years, with only one genotype (Colin) appearing in the top 10 genotypes in both years, being 10th and first in 2022 and 2023 respectively. A specific comparison was performed with the Tukey method between all genotypes and the reference Norwegian variety Ingrid, which accounts for most of the pea production in Norway. In 2022 Ingrid was ranked 16th, however, the difference with all of the 15 genotypes having higher grain yield across the three locations was not statistically significant ($p>0.1$). Only two genotypes, Leva and Cysterski, had a significantly lower grain yield than Ingrid. In 2023 Ingrid was ranked 27th and the top eight genotypes in the rank (Colin, Nemo, Mefisto, Egle, Turnia, Kazek, Mandaryn and Sisu) had all a statistically significant ($p>0.05$) higher grain yield than Ingrid.

The strongest association among traits (after phenotypic adjustment across years and locations) were found between grain yield and protein yield, protein yield and BBCH89, plant height at harvest and maximal plant height, maximal plant height and BBCH70 and all combinations of the BBCH measurements (all >0.50, **table 2**). The importance of these associations was confirmed by magnitude of the genetic correlations (**S9**) which also showed a strong association (>0.50) of maximal plant height with all BBCH measurements.

Table 2. Phenotypic correlations calculated as Pearson correlations on traits genotypic mean across the whole dataset.

	TSW	PRT	PRT_YLD	PLH_max	PLH_har	BBCH60	BBCH70	BBCH89
YLD	0.23	-0.07	0.86***	0.47***	0.34**	0.35**	0.48***	0.49***
TSW		-0.19	0.12	0.13	0.27	-0.15	-0.04	-0.07
PRT			0.39**	-0.16	0.11	0.28*	0.11	0.26
PRT_YLD				0.37**	0.32*	0.49***	0.46***	0.55***
PLH_max					0.56***	0.29*	0.52***	0.49***
PLH_har						0.06	0.31*	0.42**
BBCH60							0.68***	0.73***
BBCH70								0.90***

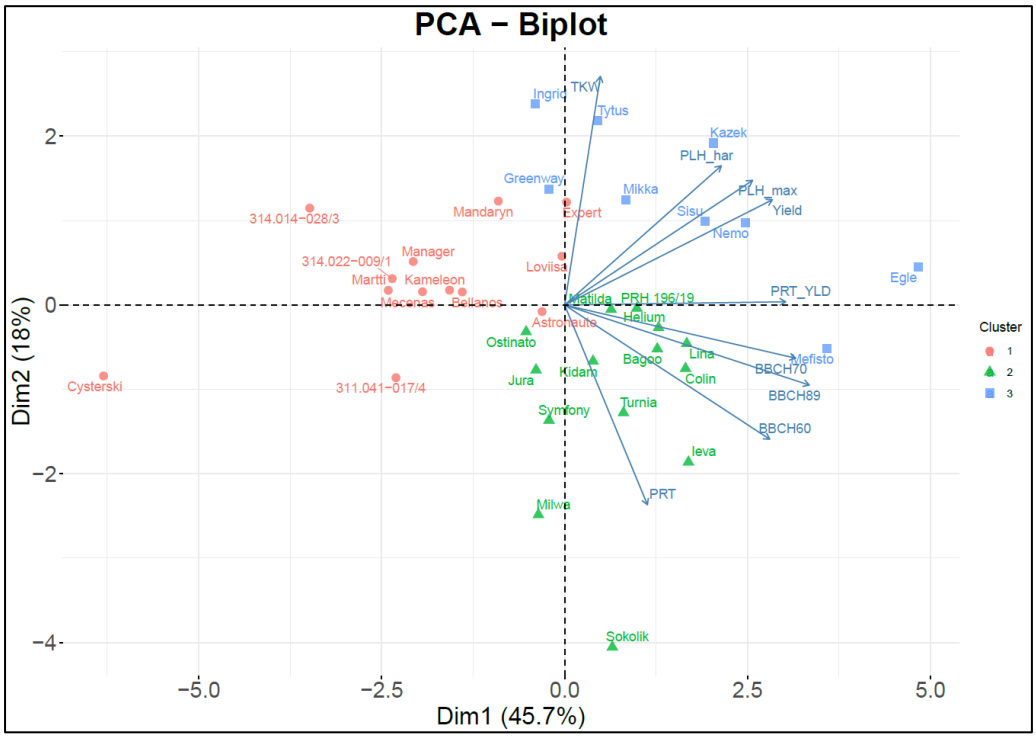
*, **, *** significance at the 0.1,0.05 and 0.01 levels, respectively.

The genetic association between the traits was further investigated through a PCA performed on the adjusted phenotypes, which retains only the genetic component of the variance (**Figure 2A**). The genotypes were clustered in three different groups by the HCA performed on the same matrix of adjusted phenotypes (**Figure 2B**). One group (blue in **Figure 2A** and **B**) was quite diverse and comprehended genotypes with high values for maximal plant height and plant height at harvest as well as high grain yield (Kazek, Sisu, Nemo top right of **figure 2A**). Belonging to the same group, the genotype Tytus and Ingrid were characterized by high TSW, while Eagle and Mefisto were more similar to the previous subgroup but also characterized by higher protein yield and later development (right side of **Figure 2A**). The second group (green in **Figure 2A** and **B**) comprehended genotypes which were characterized by average performance in terms of grain and protein yields (placed near the center of **figure 2A**) although some of them, Colin, Lina, Bagoo and Leva had a later plant development (higher values of BBCH measurements) while Leva, Milwa and Sokolik also had a higher protein content than all the other genotypes (lower part of **figure 2A**). Finally, the third group (red in **figure 2A** and **B**) were characterized by lower-than-average performance across years and

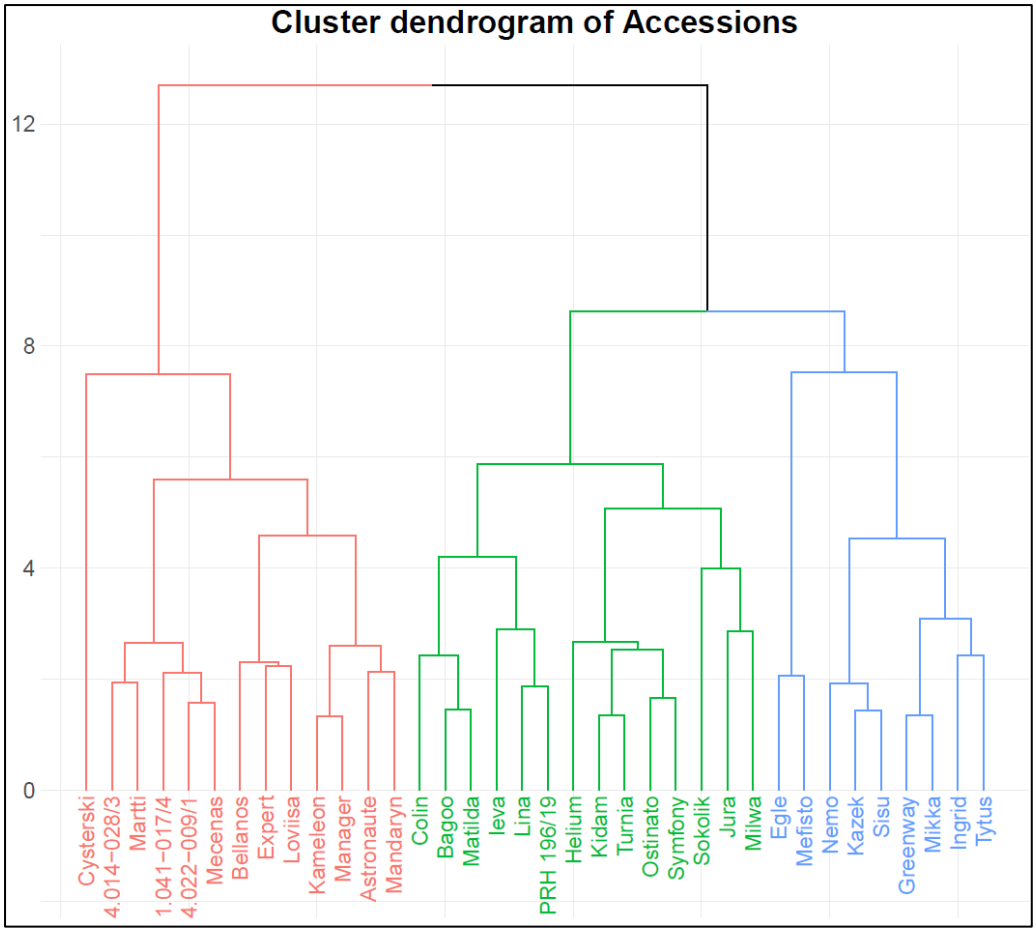
locations and were positioned in the center and left side of **figure 2A**. It is worth noting that since the multivariate analyses presented in **Figure 2** were performed on adjusted phenotypes, the associations among traits and grouping of genotypes are stable across environments, since they are based only on the genetic part of the phenotypic variance.

Figure 2. Results of multivariate analyses of the phenotypic traits measured on 36 genotypes across two years and 3 locations. **A** Principal Component Analysis and **B** Hierarchical clustering analysis performed on the matrix on adjusted phenotypes. In both panels the genotypes were colored following the cluster assignment indicated by the HCA performed on adjusted phenotypes.

A



B



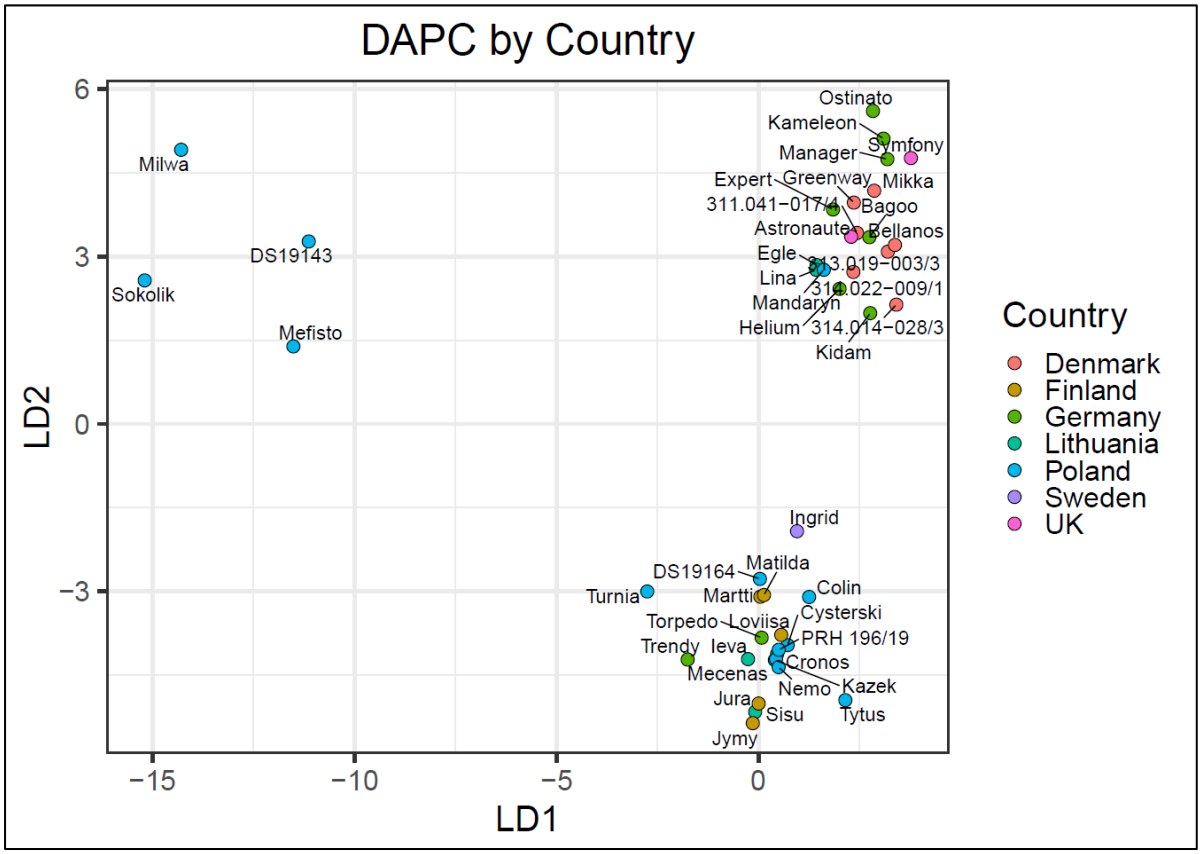
LD and Population structure

The LD decay was very rapid for each chromosome (LD plots are shown in the supplementary material **S10**) perhaps not surprisingly, since genotyping tools such as SNP arrays markers are already selected for being in low LD to avoid redundancy. However, the SNP distribution was even across the chromosomes, providing good coverage of all the genome (**S7**).

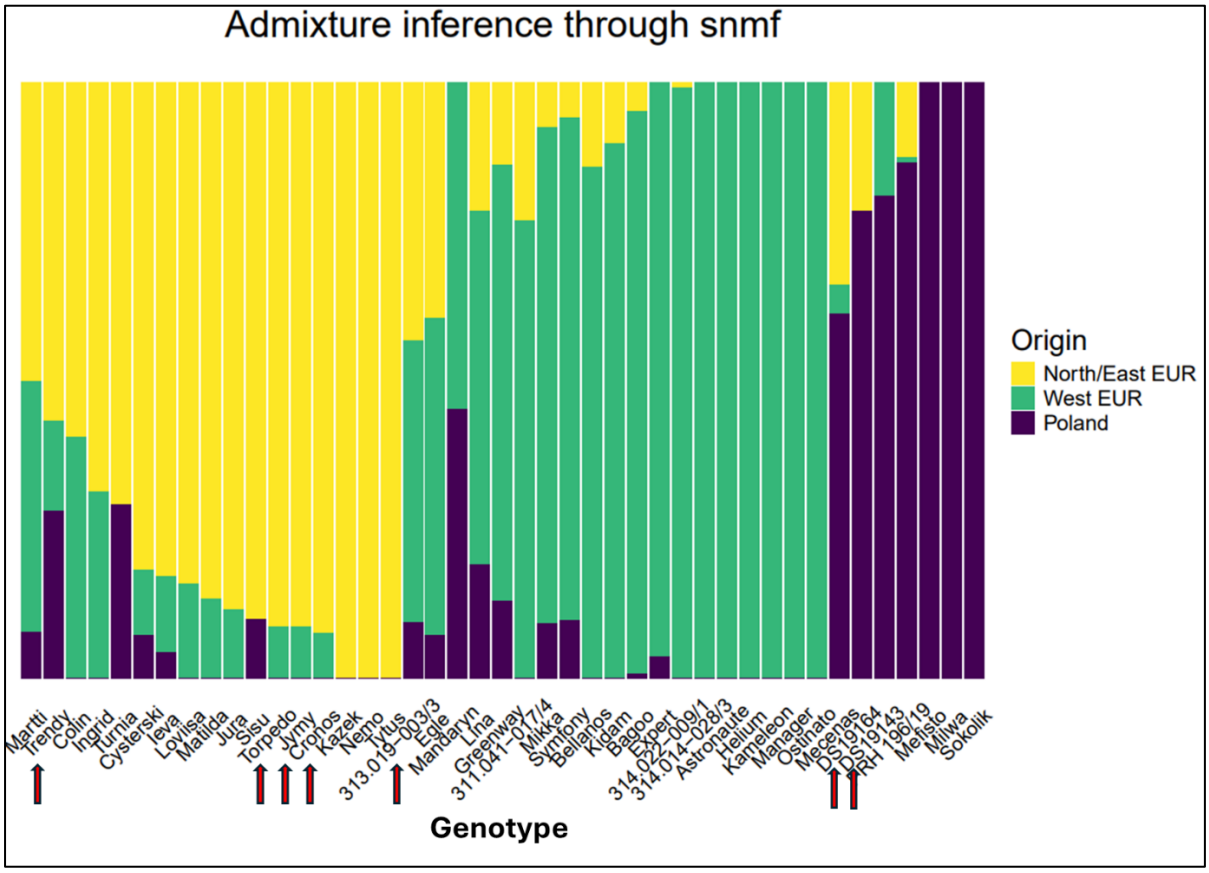
The optimal number of genetic clusters as suggested by the K-means statistic was equal to three (**S11**). The structurization of the genotypes (complete dataset of 43 genotypes) in three groups partly reflected the geographic origin of the plant material, as shown in **figure 3 A** and **B**. The genomic data clearly distinguish the material in three genetic backgrounds, with two major groups comprehending: 1) genotypes with western European origin (up right corner in **figure 3A** and green bars in **3B**) and 2) a mixture of genotypes with northern and eastern European origin (low right corner in **figure 3A** and yellow bars in **3B**). The third group was less numerous and was composed exclusively of genotypes bred in Poland, which were clearly distinct from the rest (up left corner in **figure 3A** and purple bars in **3B**).

Figure 3. Results of population genetic structure performed on the genotypic matrix of filtered and imputed SNPs for all the 43 genotypes for which genotypic data were available. **A** Discriminant analysis of principal components (DAPC) and **B** sNMF analysis for admixture inference. The genotypes in both figures were color coded depending on their country of origin (breeder provenance) in figure **A** and for macro-area of origin in figure **B**. The genotypes highlighted by the red arrows in figure **B** are those absent from the phenotypic dataset.

A

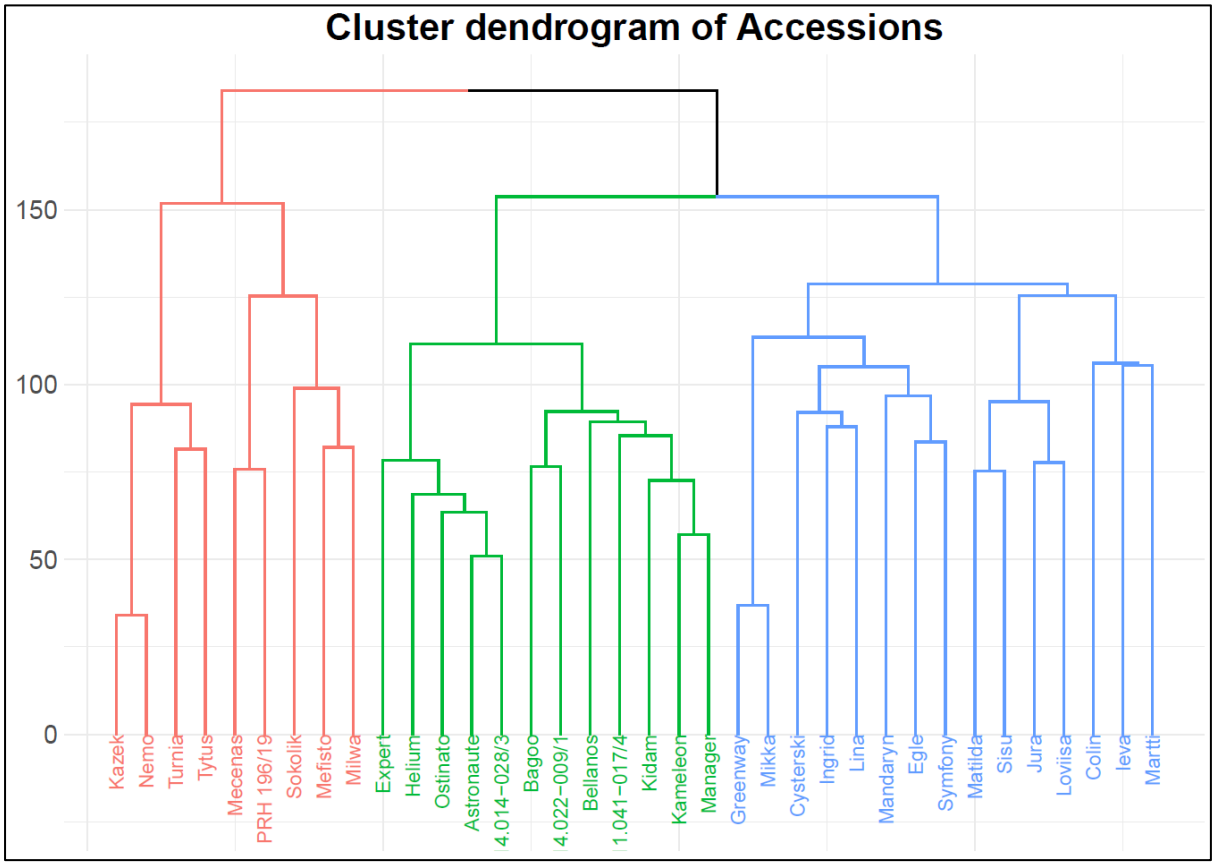


B



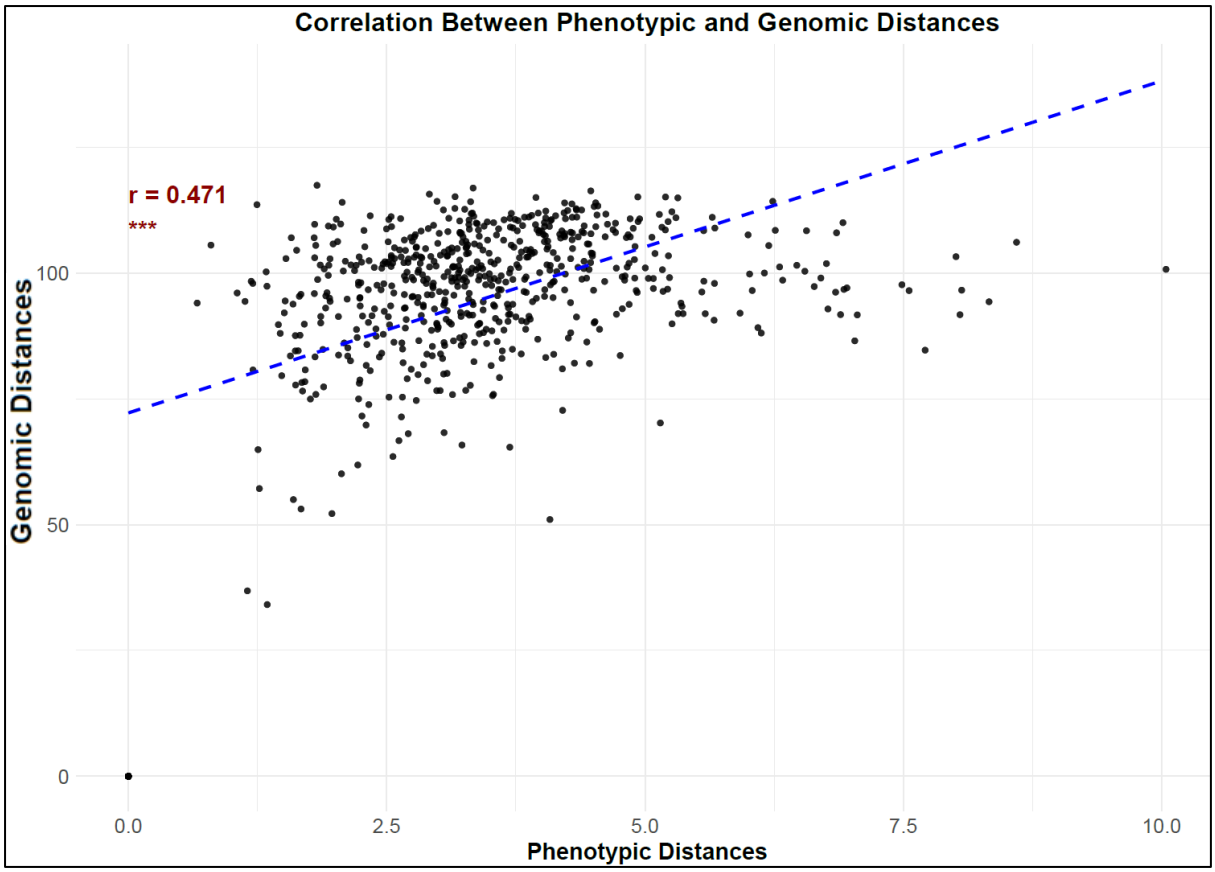
When considering only the 36 genotypes present in both the 2022 and 2023 field trials at all locations, the HCA analysis of the genomic data highlights once again three major groups as indicated in **figure 4** and **S12**. The groups highlighted in **figure 4** were close to those identified by the DAPC on the whole dataset, with the accessions most closely related in terms of geographic origin having also the shorter genetic distance.

Figure 4. Hierarchical clustering analysis (HCA) performed on the matrix of filtered and imputed SNPs for the 36 accessions included in the field trials in 2022 and 2023 at all locations



To test the consistency between the classification of the 36 accessions based on the (adjusted) phenotypic and genomic data a simple correlation was performed between the genotypes distance matrixes based on the phenotypic and genomic data, respectively. The moderately high correlation coefficient of r=0.47 (P<0.01), shows that there is consistency in the differentiation of the considered genotypes in terms of their phenotypic and genomic based genetic diversity (**figure 5**).

Figure 5. Results of correlation analysis between measures distance matrixes based on the (adjusted) phenotypic and genomic data, respectively, for the 36 genotypes present in the 2022 and 2023 field trials at all locations.



Discussion

Trait variation and interrelationship

This study identified considerable phenotypic variation for all the evaluated traits, with both location and year playing an important role in determining genotype performance. This confirms the importance of multi-location field trials of diverse material in the preliminary steps of a breeding program, to identify the correct breeding strategy and preselect high performing and stable genotypes (Annicchiarico et al., 2009; Gauch Jr., 2013). In this case, the two testing years were strongly different in terms of the weather conditions under the growing season and thus resulted in clear differences in genotype performance. Year had the strongest contribution to the phenotypic variation for all traits, in particular grain and protein yields were considerably lower at all locations in 2023. Location had a significant effect on all traits, excluding plant height, which was stable across testing locations but was reduced by the suboptimal growing conditions in 2023 as for grain and protein yields.

The weather conditions in 2023 included both an early drought (mid-May to mid-June) and thereafter frequent rain that lasted for the rest of the growth period with several events of unusual intense rainfalls in the period prior to harvest (August). Trials were irrigated during the early drought period to ensure satisfactory canopies, but a faster phenological development was nonetheless recorded. It is likely that the frequent precipitation lasting for the whole period from pod and seed development until maturation was the main reason for the lower grain yield in 2023. This could be due to both the cloudy weather reducing irradiance, and to the severe lodging that could reduce both assimilation and translocation into seeds and increase losses at harvest. Even though the weather conditions in 2023 were considered strongly unusual, the increased precipitation in the later part of the growth season with events of very intense rainfalls is in line with climate change predictions for these areas (Sjulgård et al., 2023). The weather conditions observed in 2023 will be more common in Norway; this on one hand highlighted the challenges of grain legumes cultivation in the future, but on the other it resulted in an opportunity for this research project to investigate how different genotypes performed in significantly different testing years. All traits but protein content and BBCH70 had a low, if not absent G x L interaction effect. In particular, for grain and protein yields, the traits with highest interest for plant breeders, the biggest source of variation was the within year G x L interaction. This suggests that the performance of different genotypes at different locations was more strongly affected by the weather conditions that differentiated the two growing seasons than by the different locations. For example, genotypes with earlier development and maturation were more negatively affected by the late harvest induced by the long period with rains and cold/cool temperatures characterizing the second part of the summer in 2023. Some of these genotypes showed also strong lodging, that in combination with the late harvest resulted in heavy yield losses. The strong difference in genotype ranking for grain yield between 2022 and 2023 suggests that breeding efforts for Norway should target the development of pea varieties with broad adaptation for the whole region rather than different varieties having specific adaptation to different subregions. This approach is justified by the fact that: 1) the major pea/grain growing region in Norway is restricted to a relatively delimited area with similar climatic conditions; 2) as proven by our results, the strongest source of genotype by environment interaction is temporal (different years with different weather conditions) rather than geographical (different locations with different weather conditions); 3) pea is still considered a minor crop in Norway which does not allow for major investments in terms of breeding. Therefore, the identification of genotypes with stable, even if average, yield performance will likely result in higher grain and protein yields over time, and an

increased adoption of this crop which is still considered by farmers (together with faba bean) less reliable than cereals because of heavy yield losses in unfavorable years (de Visser et al., 2014).

The relatively high TSW, protein content and maximal plant height h^2 values confirms previous findings in Pisum sativum by (Singh and Singh, 2011; Tolessa 2017, Ali et al. 2021, Uhlarik et al 2022) which also found high h^2 for the same traits. This work found that phenological development and maturation time, measured as days to end of flowering (BBCH_70) and days to full ripe (BBCH_89) had also quite high $h^2$and genotypic VC, with the traits being stable across environments both in terms of different years and locations. The stability of these traits is particularly relevant for the selection of the right material for cultivation in Norway, which favors early maturing genotypes because of the challenges of wet and cool autumns which often hinder harvesting of late maturing material (Carlson-Nilsson et al., 2021).

Trait association was investigated through correlation and PCA on adjusted phenotypes, to identify associations that are based on the genetic component of the phenotypic variance and thus, that are stable across different environments. Grain yield had the highest correlation with protein yield, but none of these traits were significantly correlated with grain protein content. This was also found in other studies (Carlson-Nilsson et al., 2021; Ferrari et al., 2016; Mohammed et al., 2018). One of the main goals of Norwegian agriculture is that of increasing grain legume production to replace imported protein feed ingredients. Thus, the increase of protein yield (as protein production per unit area) over time is a more relevant goal than the increase of grain protein content by itself. Therefore, the high correlation between grain yield and protein yield is positive news in the current context, where the farmers are simply paid for the tonnage of grain produced with no premium paid for improved quality for example in terms of protein content. Yield (both grain and protein yield) was positively correlated with BBCH measurements as well as maximal plant height and to a lesser extent plant height at harvest, suggesting that taller genotypes with longer vegetative cycle results in higher yields. These findings are in line with previous literature (Ali et al., 2021; Carlson-Nilsson et al., 2021; Uhlarik et al., 2022) and indicate that despite the cool and wet harvesting season that often characterize cultivation areas in Norway (as we experienced in 2023), genotypes with longer growing cycles outperformed early genotypes in terms of grain yield. However, it is worth nothing that the early maturing genotypes might have been penalized in terms of grain and protein yields because they were harvested too late, resulting in pod shattering. For logistical reasons all plots at all field trials considered in this work were harvested at once.

Several genotypes had higher grain yield than the reference Norwegian variety Ingrid in both the 2022 and 2023 growing seasons, suggesting that new genotypes might be considered for cultivation in Norway. In particular, when considering the results of the model conducted across years and locations Colin, Kazek, Nemo, Helium and Egle were the five highest ranked genotypes, thus having the most stable grain yield across years and locations.

The results of multivariate analyses on the matrix of adjusted phenotypes indicated the classification of the tested genotypes in three different groups, characterized by distinct adaptation pattern. Of these three groups, one included genotypes with lower-than-average performance in terms of grain yield (red in **figure 2A** and **B**), and thus might not be suggested for cultivation in Norway or as parents of new crosses within a future Norwegian breeding program. However, such low yielding genotypes might still be useful if they are carrying traits that are considered useful by plant breeders. In this study this might be the case for some very early genotypes which could be used as a source of earliness in a breeding effort aiming at creating very early types which could extend the cultivation areas or be cultivated as a pre-crop to winter cereals. Most of the high yielding and stable genotypes belonged to one of the remaining two groups, blue or green in **figure 2A** and **B**). Plant height seems

to be important for achieving a high grain yield, however, considering the susceptibility of pea to lodging especially in wet conditions, plant height at harvest is probably a better predictor of a genotype that is less prone to lodging. Finally, among the third group (green in **figure 2A** and **B**), some genotypes might be considered as a source of genetic diversity for higher protein content. Considering that the trait was not significantly correlated with grain yield and that a considerable genetic diversity was identified (high G VC and h^2), an increase of protein content of the high yielding genotypes might be achieved through crossing with those having higher protein content without compromising their productivity. Most of the genotypes with high protein content had eastern European origin, something which was observed in a previous study considering a wider panel of genotypes (Jha et al., 2013). The results of the structure analyses performed on genomic data confirmed that this material has indeed a different genetic background and therefore, might carry genes associated with higher protein content that can improve this trait in higher yielding genotypes of western European origin such as those in blue in **figure 2A** and **B**.

Population structure and marker based genetic diversity

To complement the evaluation of genetic diversity in terms of phenotypic traits measured in the field, this study used molecular markers to measure the diversity at the genomic level. The most popular method for finding associations between a phenotypic trait with molecular markers is perhaps genome wide association study (GWAS). GWAS requires a sufficient number of evaluated genotypes to identify reliable trait-marker associations, especially for quantitative traits which are strongly affected by L and G x L variations. The number of evaluated genotypes in this work was not sufficient to perform a GWAS. However, the analysis of population structure clearly identified diverse genotypes in terms of their genomic makeup that can be used as source of useful diversity for several traits.

The different analyses performed to evaluate population structure identified three main clusters which partially coincided with the different geographic origin of the genotypes. Genotypes with different provenances are often characterized by distinct phenotypes and geographic origin is often associated with structurization at the genomic level (Annicchiarico et al., 2017; Crosta et al., 2022; Pavan et al., 2022). The climatic conditions at the place of origin of a genotype often induce new mutations that will be gradually fixated within the population if it is useful for adaptation to that specific environment (Zanotto et al., 2021). In the process of plant breeding, new genotypes that are source of useful traits are useful only if they indeed carry new genes that are associated with the trait of interest. Therefore, the selection of parents for new crosses with the aim of trait improvement should be done from genotypes belonging to different genetic pools. The results of this study were promising to this extent, since: 1) high yielding and stable genotypes were identified among different clusters and 2) genotypes with high protein content belonged to a different cluster than those with high and stable grain yield. From a breeding perspective this suggests that there is the potential for improving grain yield by crossing different high yielding genotypes belonging to different genomic clusters. Also, the protein content of high yielding genotypes can be improved through crossing with genotypes with high protein content belonging to a different genomic cluster. These results are of high relevance for the long-term strategy of improving plant protein production by unit area, or protein yield.

Furthermore, the relatively high and strongly significant correlation between the measure of genetic distance between genotypes at the phenotypic and genomic levels, indicates that the observed phenotypic diversity is indeed associated with different alleles carried by different genotypes. In future breeding efforts a wider panel of plant material should be evaluated in field trials and genotyped with the same or similar tools. SNP chip technology proved reliable in describing genetic

diversity (Tran et al., 2023, 2022), is repeatable over time and allows to join different datasets since the tested markers are always the same (and often associated with relevant phenotypic traits). This is particularly useful for the development of genomic selection models, which requires the availability of a lot of phenotypic and genotypic data from many different genotypes across different environments (Annicchiarico et al., 2019; Annicchiarico et al., 2017; Crosta et al., 2023), something which is often not realistic within a single research project.

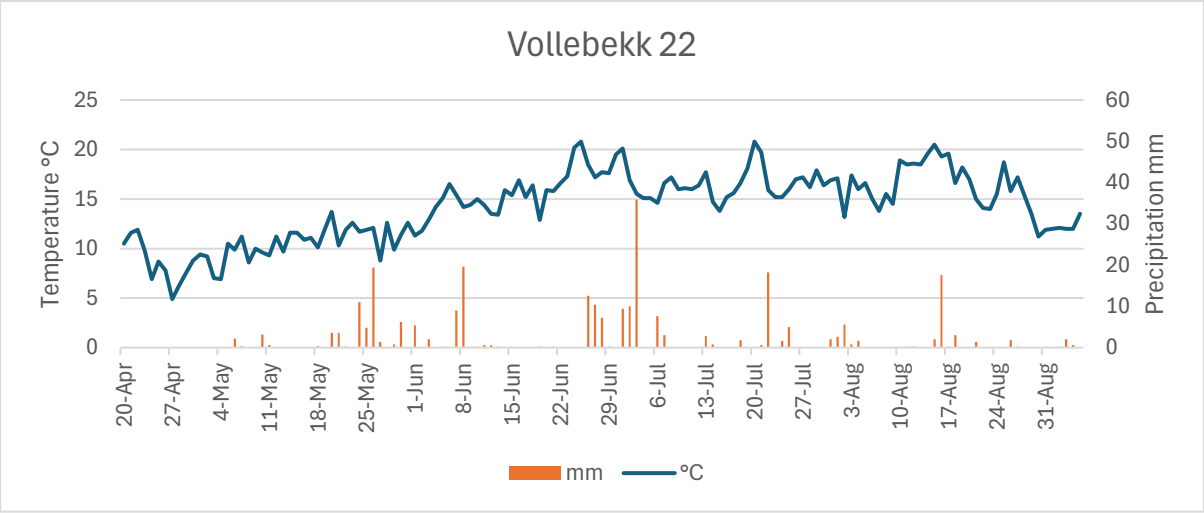
In this perspective, this study was successful in exploring available genomic tools for describing genetic diversity in a considerably small collection of pea genotypes, and to relate it to the observed phenotypic diversity for several important agronomic traits. A future step should be the evaluation of a larger number of genotypes and building a dataset of genotypic and phenotypic observations robust enough for developing a genomic prediction model. This process works iteratively, progressively including new genotypes with the goal of improving the traits, and it was demonstrated as being more time and resource effective than traditional breeding schemes (Crosta et al., 2022). At the same time the evaluation of new genotypes done in such research projects (particularly varieties and advanced breeding lines) has an important value for cultivar recommendations for farmers as well as data sharing to the providers (collaborating breeding institutions) of such genotypes. Therefore, we foresee a continuation of the current project as pre-breeding activity for the Norwegian environment which might result in the longer terms into a Norwegian based breeding program for grain legumes.

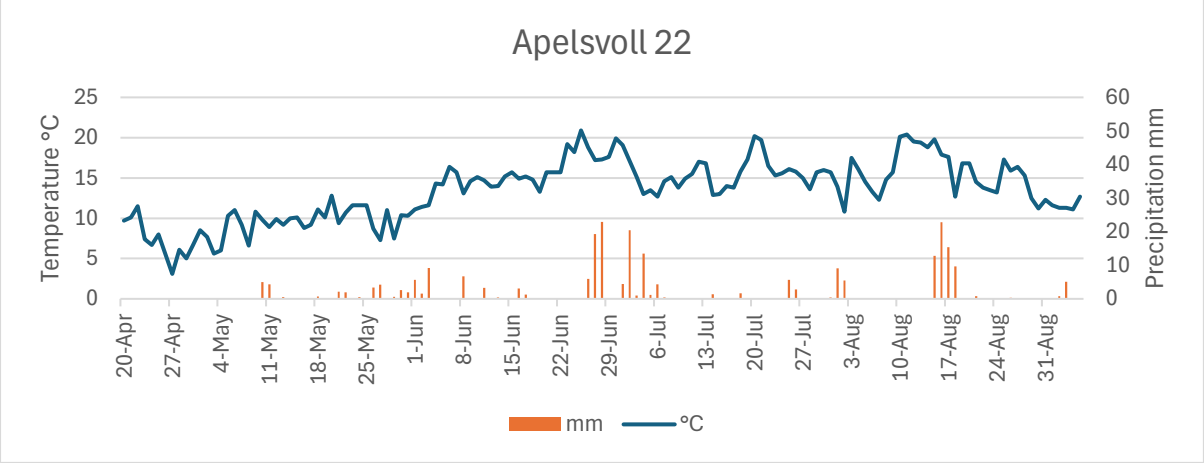
Supplementary material

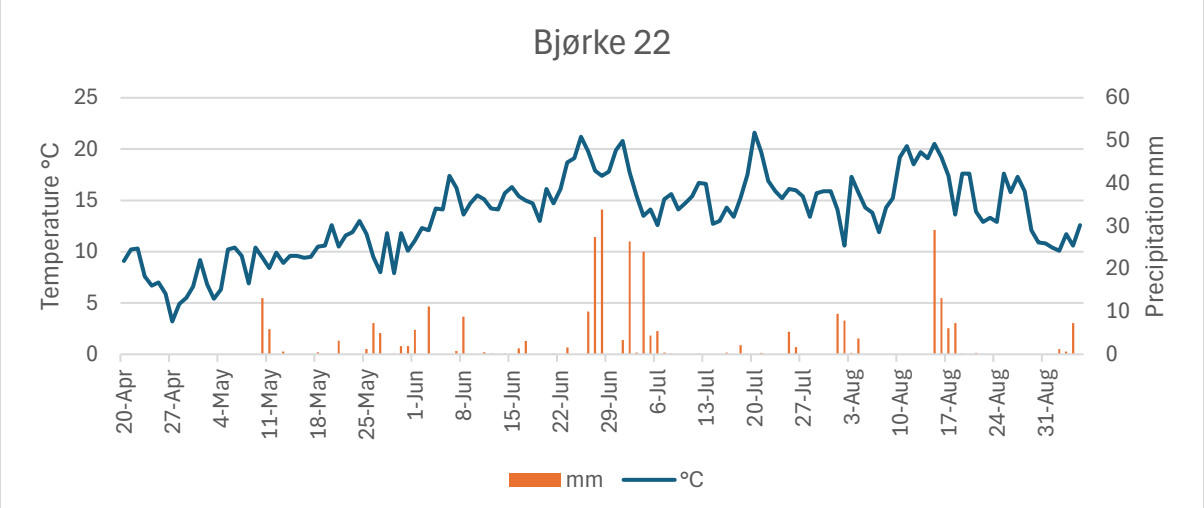
S1. Overview of the pea varieties tested in field trials in 2022 and 2023, breeding companies and countries of origin. The varieties written in bold letters are genotyped and phenotyped in two seasons at three locations.

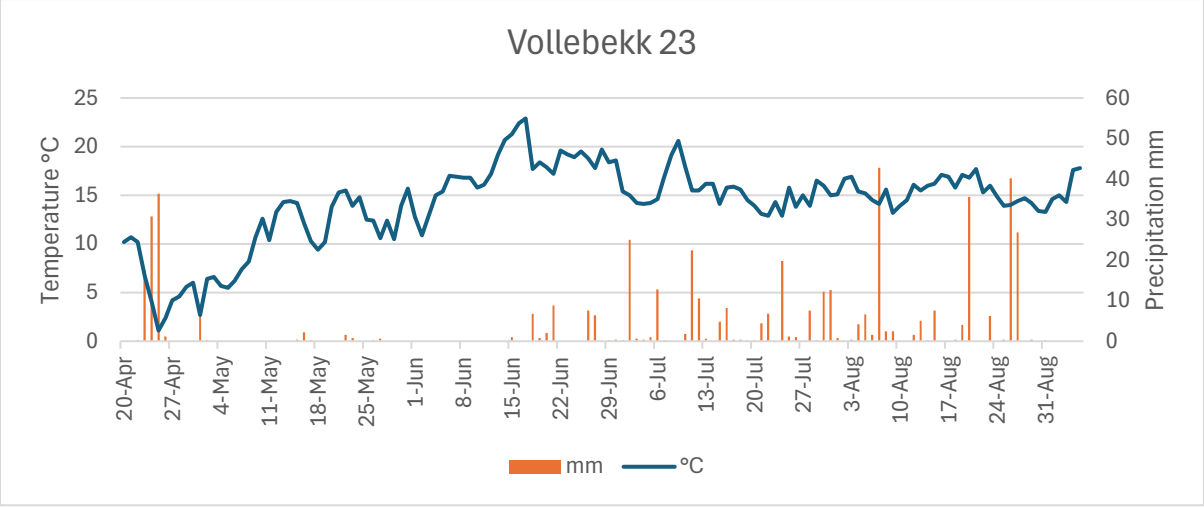
Variety/line	Breeder	Country	2022	2023
311.041-017/4	Nordic seed	Denmark	x	x
313.019-003/3	Nordic seed	Denmark	x	x
314.014-028/3	Nordic seed	Denmark	x	x
314.022-009/1	Nordic seed	Denmark	x	x
Akord	Phr	Poland		x
Asgard	P.H.Petersen	Germany		x
Astronaute	RAGT	France	x	x
Bagoo	KWS	Germany	x	x
Bellanos	NordicSeed	Denmark	x	x
Colin	Danko	Poland	x	x
Cronos	P.H. Petersen	Germany	x	
Cysterski	Phr	Poland	x	x
DS19143	Danko	Poland	x	x
DS19164	Danko	Poland	x	
DS19179	Danko	Poland		x
DS21247	Danko	Poland		x
DS21269	Danko	Poland		x
Egle DS	Lammc	Lithuania	x	x
Expert	P.H. Petersen	Germany	x	x

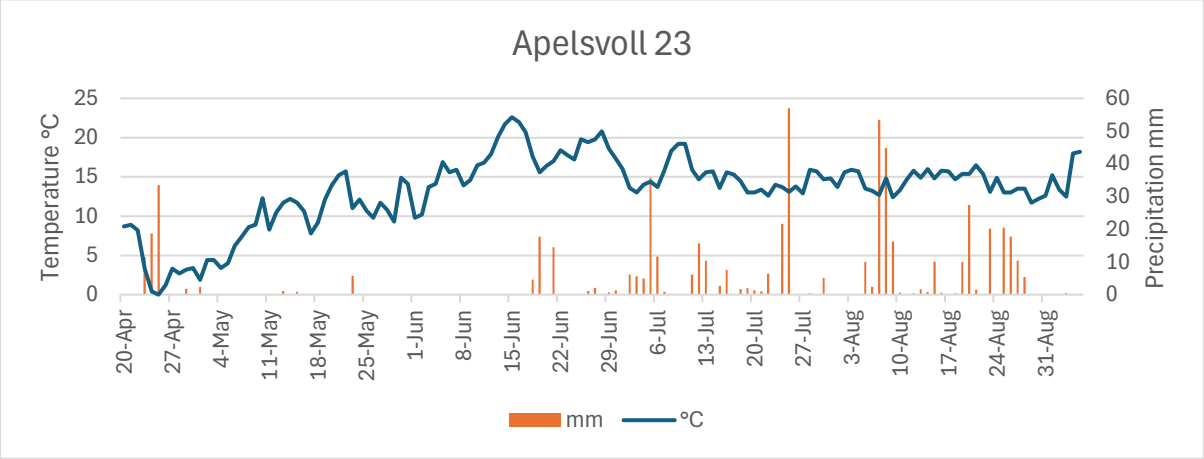
Greenway	NordicSeed	Denmark	x	x
Heikki	Boreal	Finland		x
Helium	P.H. Petersen	Germany	x	x
Iconic	RAGT			x
Ieva DS	Lammc	Lithuania	x	x
Impact	Nordic Seed	Denmark		x
Ingrid	Lantmännen	Sweden	x	x
Jura DS	Lammc	Lithuania	x	x
Jymi	Boreal	Finland	x	
Kameleon	KWS	Germany	x	x
Kazek	Danko	Polen	x	x
Kidam	KWS	Germany	x	x
Lempi	Boreal	Finland		x
Lina DS	Lammc	Lithuania	x	x
Loviisa	Boreal	Finland	x	x
Manager	KWS	Germany	x	x
Mandaryn	Hr smolice	Poland	x	x
Martti	Boreal	Finland	x	x
Matilda	Boreal	Finland	x	x
Mecenas	Hr smolice	Poland	x	x
Mefisto	Hr smolice	Poland	x	x
Mikka	NordicSeed	Denmark	x	x
Milwa	Hr smolice	Poland	x	x
Nemo	Danko	Polen	x	x
Ostinato	NPZ	Germany	x	x
PRH196/19	Hr smolice	Poland	x	x
Roch	Phr	Poland		x
Sisu	Boreal	Finland	x	x
Sokolik	Phr	Poland	x	x
Starski	Phr	Poland	x	x
Symfony	RAGT	France	x	x
Texas	P.H. Petersen	Germany	x	x
Turnia	Phr	Poland	x	x
Tytan	Phr	Poland	x	x
Tytus	Danko	Polen	x	x

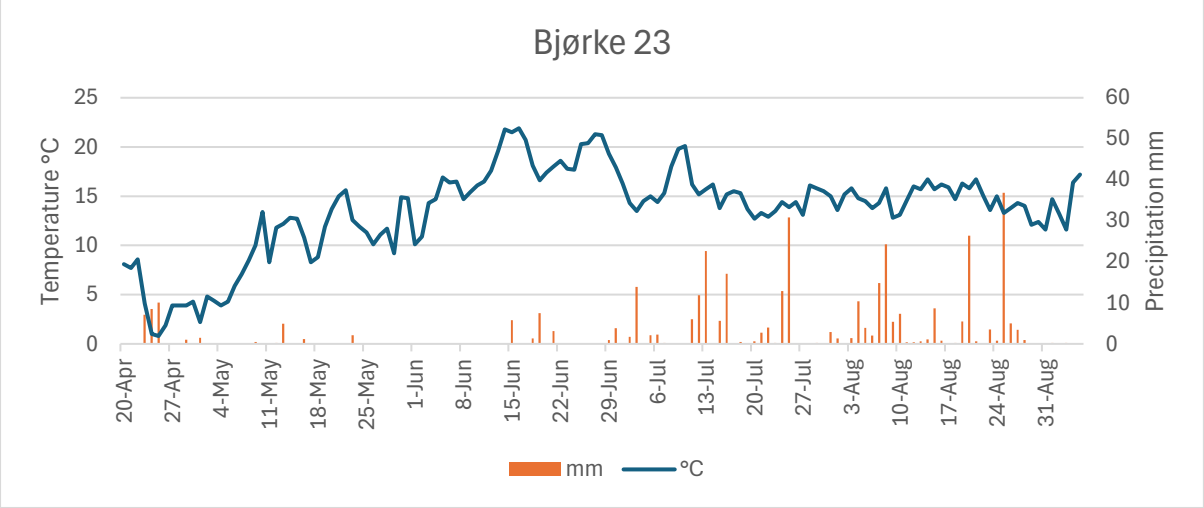

Vollebekk 22
Temperature °C
Precipitation mm
25
20
15
10
5
0
60
50
40
30
20
10
0
20-Apr
27-Apr
4-May
11-May
18-May
25-May
1-Jun
8-Jun
15-Jun
22-Jun
29-Jun
6-Jul
13-Jul
20-Jul
27-Jul
3-Aug
10-Aug
17-Aug
24-Aug
31-Aug
mm
°C


Apelsvoll 22
Temperature °C
Precipitation mm
25
20
15
10
5
0
60
50
40
30
20
10
0
20-Apr
27-Apr
4-May
11-May
18-May
25-May
1-Jun
8-Jun
15-Jun
22-Jun
29-Jun
6-Jul
13-Jul
20-Jul
27-Jul
3-Aug
10-Aug
17-Aug
24-Aug
31-Aug
mm
°C


Bjørke 22
Temperature °C
Precipitation mm
25
20
15
10
5
0
60
50
40
30
20
10
0
20-Apr
27-Apr
4-May
11-May
18-May
25-May
1-Jun
8-Jun
15-Jun
22-Jun
29-Jun
6-Jul
13-Jul
20-Jul
27-Jul
3-Aug
10-Aug
17-Aug
24-Aug
31-Aug
mm
°C







S2. Mean daily temperature (°C) and precipitation (mm) at the three locations Vollebekk, Apelsvoll and Bjørke in 2022 and 2023.

S3. Details of management practices for each field experiment in 2022 and 2023.

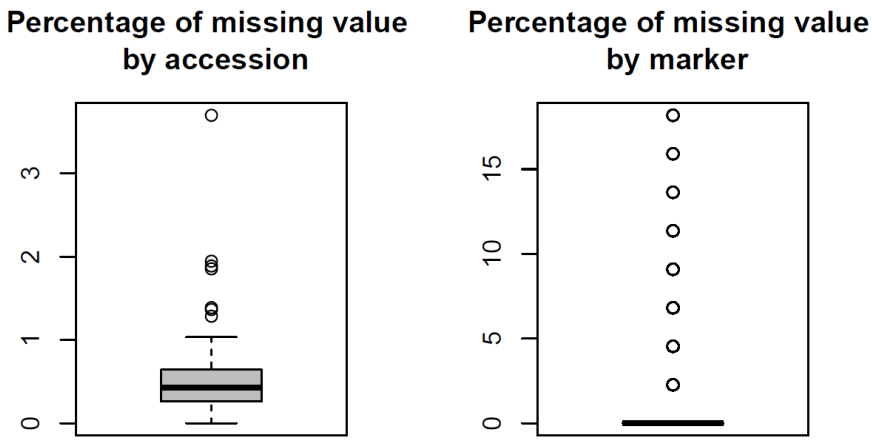
	Vollebekk 2022	Bjørke 2022	Apelsvoll 2022	Vollebekk 2023	Bjørke 2023	Apelsvoll 2023
Sowing date	22.04	06.05	26.4	08.05	10.5	11.5
Fertilizer	30kg OPTI-PK 0-11-21 (Yara)	-	-	30kg OPTI-PK 0-11-21 (Yara)	-	-
Plot size	1,5m*8m	1,5m*6m	1,5m*8m	1,6m*6m	1,5m*6m	1,6m*8m
Herbicide treatment	Basagran[1] 05.06 14.06	Basagran[1] 10.06	Basagran[1] 18.05 25.05 01.06	Basagran[1] 05.06 12.06	Basagran[1] 14.06 20.06	No treatment against weeds
Harvest date	01.08[2] 09.08[3]	29.08	25.08	28.08	22.08[2] 31.08[3]	05.09
Irrigation	20-25 mm in May_June	20-25 mm I May-June	2*20-25 mm in May-June	3*20-25 mm in May-June		One time in week 24 (20-25 mm)
Soil type	Silty loam	Clay loam	Loam	Silty loam	Clay loam	Loam
pH	6,4	7,3	5,6	6,3		5,7

[1] Basagran (containing 870 g/kg bentazon) were applied at full dose (100-140g/ha), or at two or three dosages using 15-20g/ha, 20-25 g/ha and 30-35g/ha, respectively

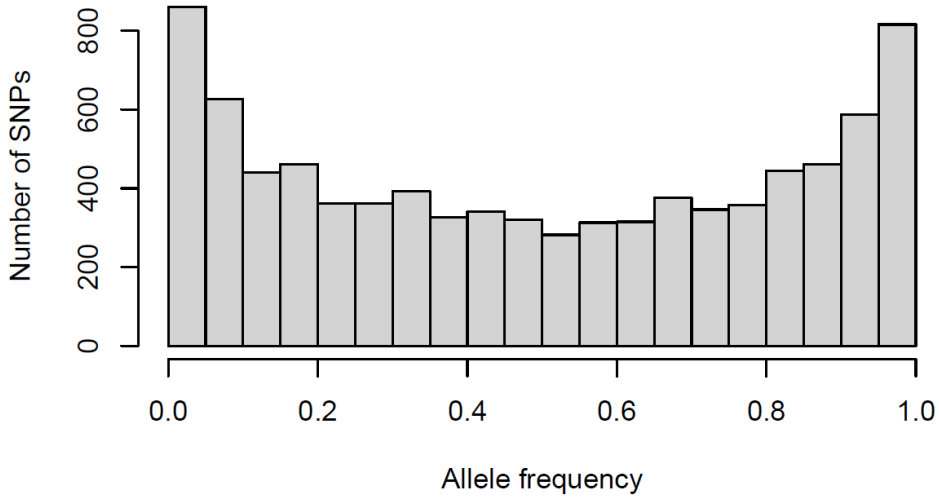
[2] Harvest date for the earliest varieties, [3] Harvest date for the medium to late varieties.

S4. Number of replications for registration of the phenotypic traits in the field trials at Vollebekk, Bjørke and Apelsvoll in 2022 and 2023

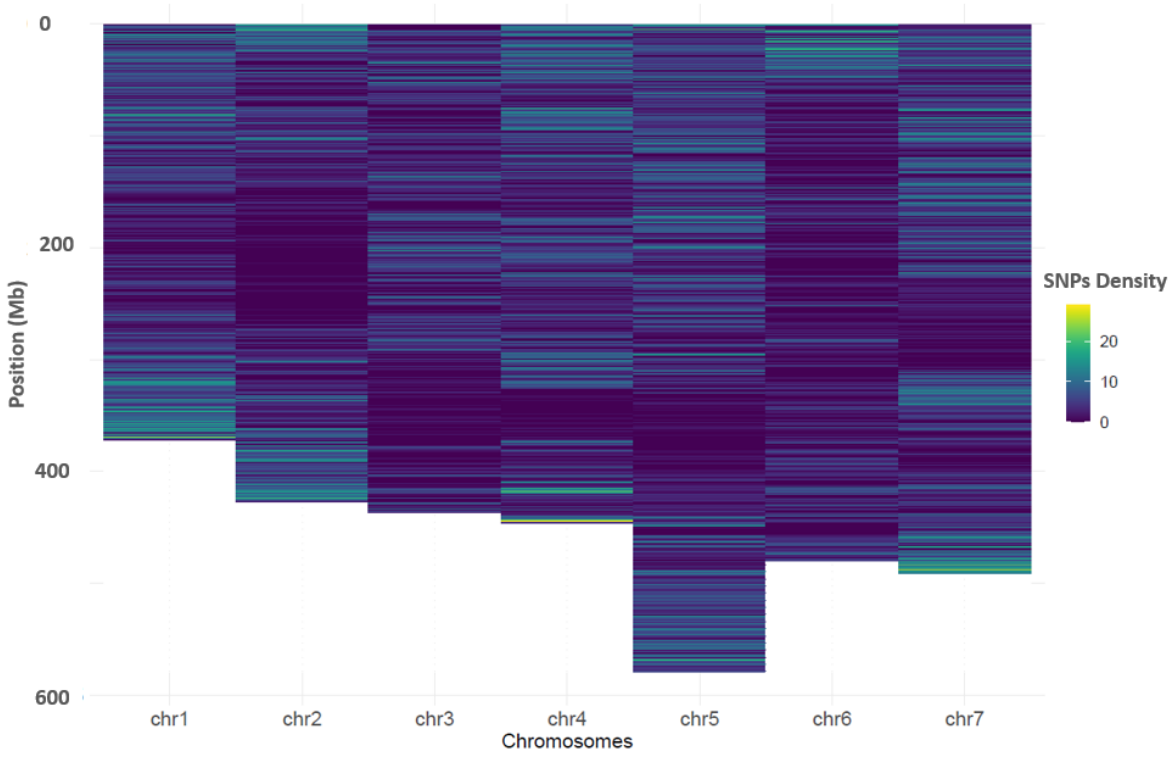
	Vollebekk 2022	Bjørke 2022	Apelsvoll 2022	Vollebekk 2023	Bjørke 2023	Apelsvoll 2023
BBCH 60	2	3 reps	2 reps	3 reps	3 reps	2 reps
BBCH70	2 reps	3 reps	2 reps	3 reps	3 reps	2 reps
BBCH89	2 reps	3 reps	2 reps	3 reps	3 reps	2 reps
PLH_max	2 reps	3 reps	2 reps	3 reps	3 reps	-
PLH_har	2 reps	3 reps	-	3 reps	3 reps	3 reps
YLD	3 reps	3 reps	3 reps	3 reps	3 reps	3 reps
TSW	2 reps	3 reps	3 reps	2 reps	3 reps	3 reps
PRT	2 reps	3 reps	3 reps	-	3 reps	3 reps
PRT_YLD	2 reps	3 reps	3 reps	-	3 reps	3 reps



S5. Boxplots of the fraction of missing values by accession and SNP, respectively, for the 44 pea accessions after removing SNPs with more than 20% missing values and MAF<0.01.

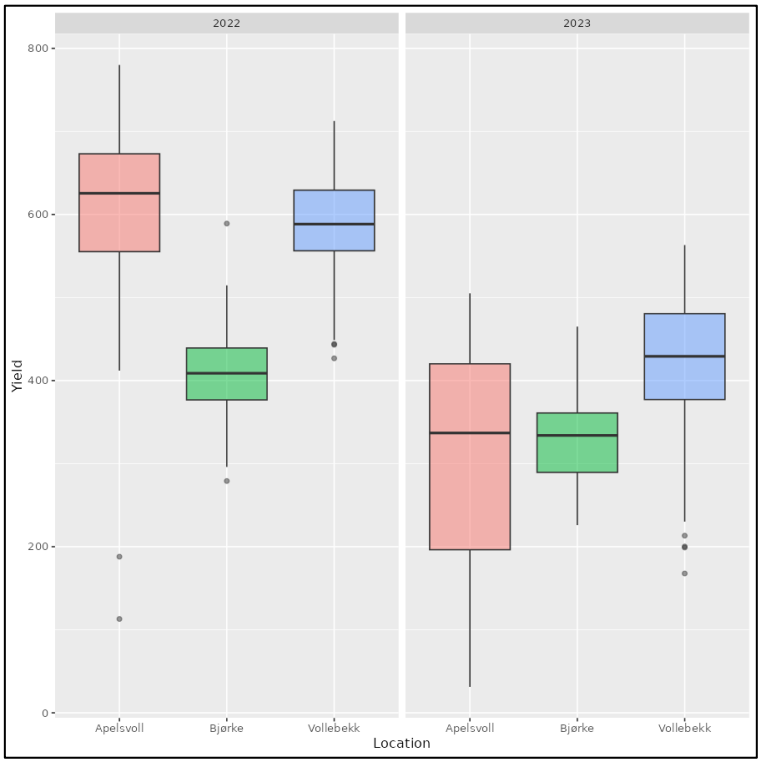


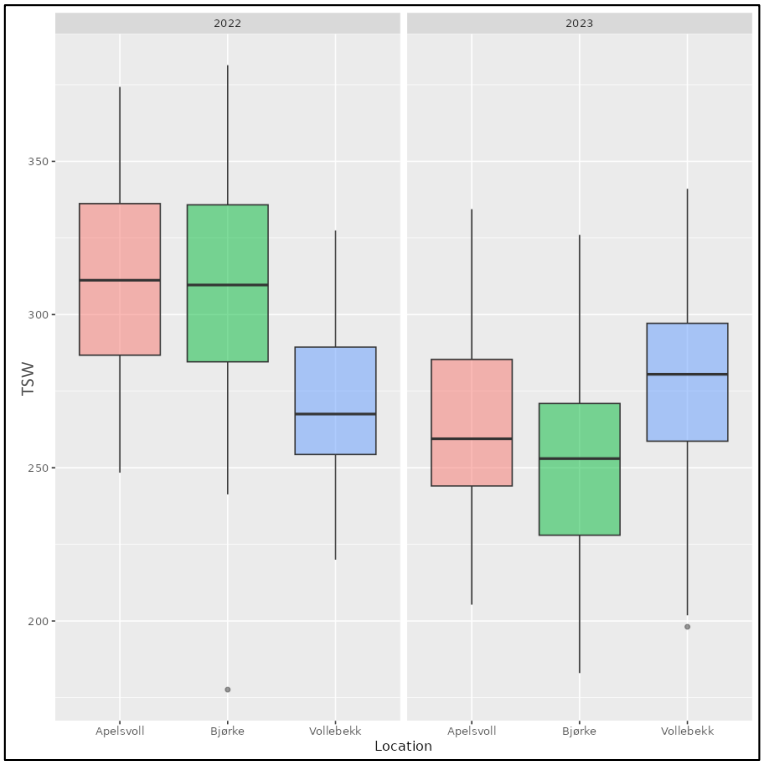
S6. Histogram of the distribution of allele frequencies (across all accessions) after removing SNPs with more than 20% missing values and MAF<0.01.

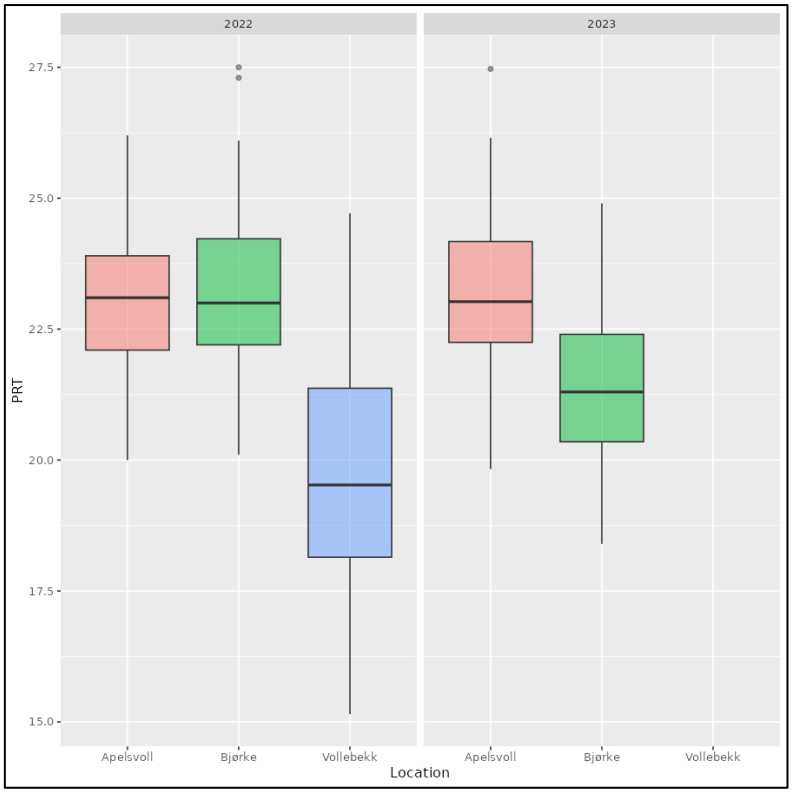


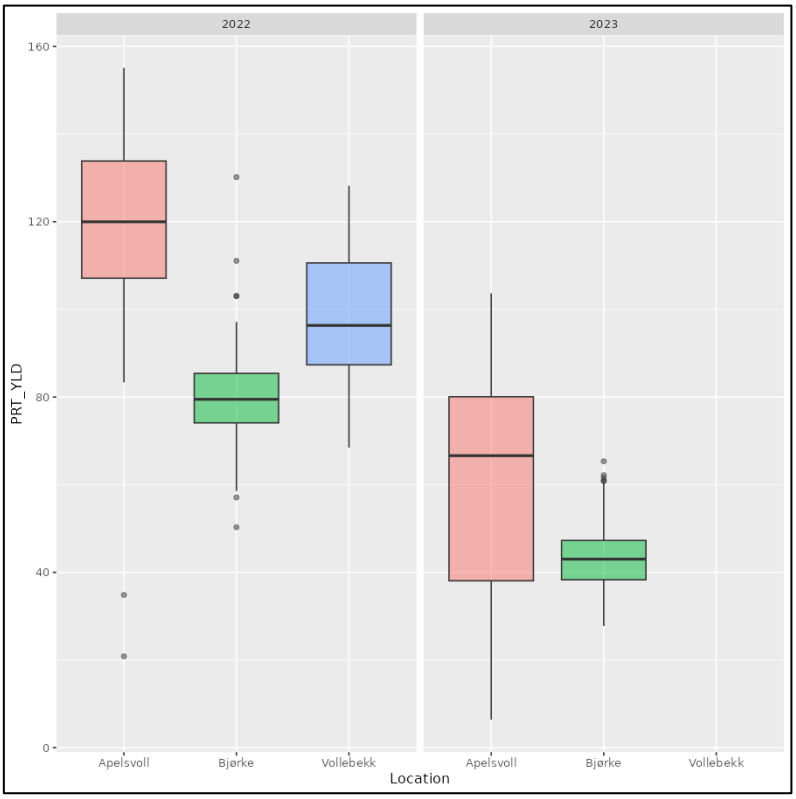
S7. Heat maps of each chromosome representing SNPs density within windows of 1 Mb. The y-axis reports the distance in Mbp.

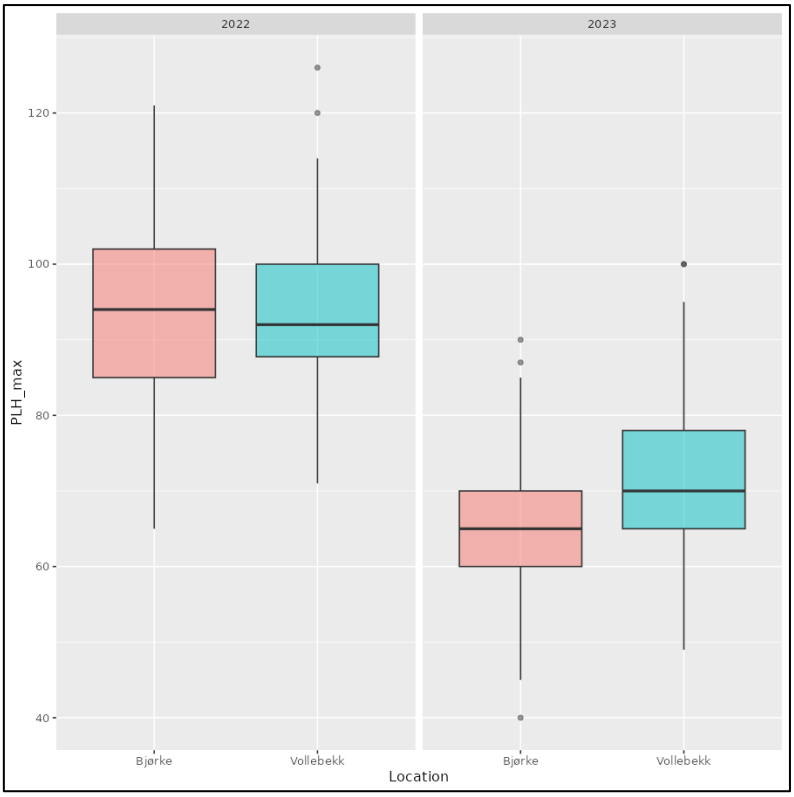
S8. Distribution of phenotypic traits for 36 pea genotypes across 2 years (2022 and 2023) and 3 locations (Apelsvoll, Bjørke and Vollebekk). Yield = Grain Yield; TSW = Thousand seed Kernel Weight; PRT = Protein Content; PRT_YLD = Protein Yield; PLH_max = Maximal Plant Height; PLH_har = Plant Height at Harvest; BBCH60 = first flower open; BBCH70 = end of flowering; BBCH89 = full ripe.

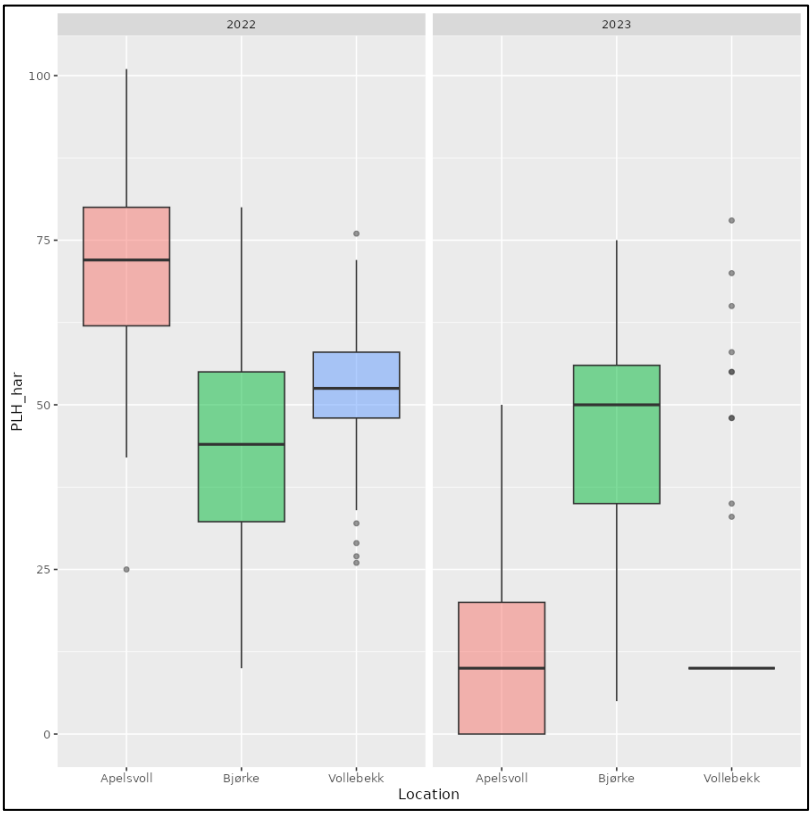


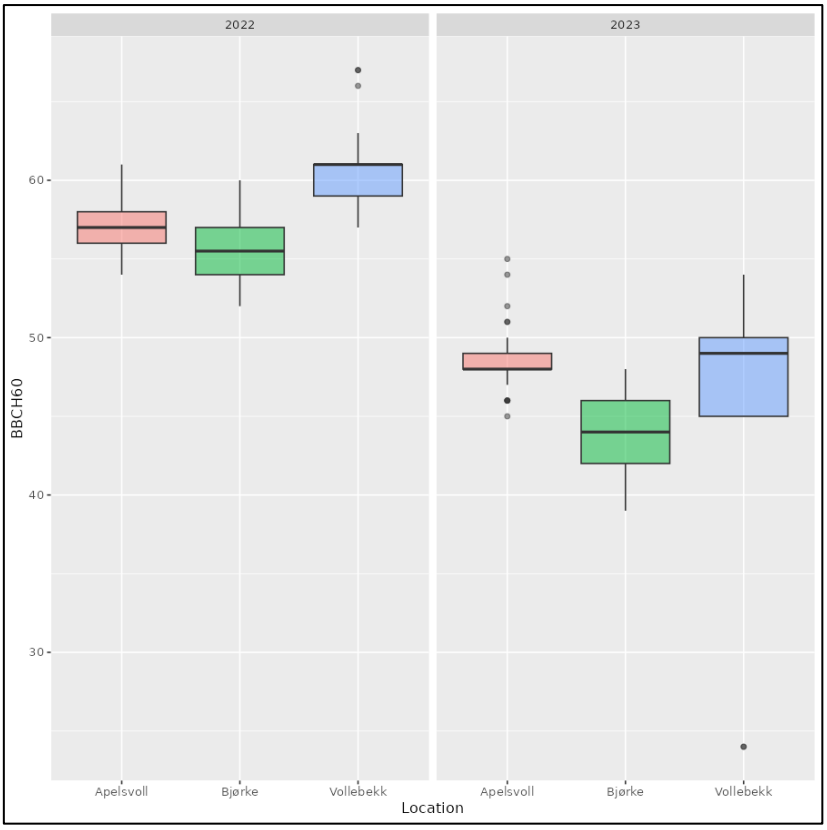


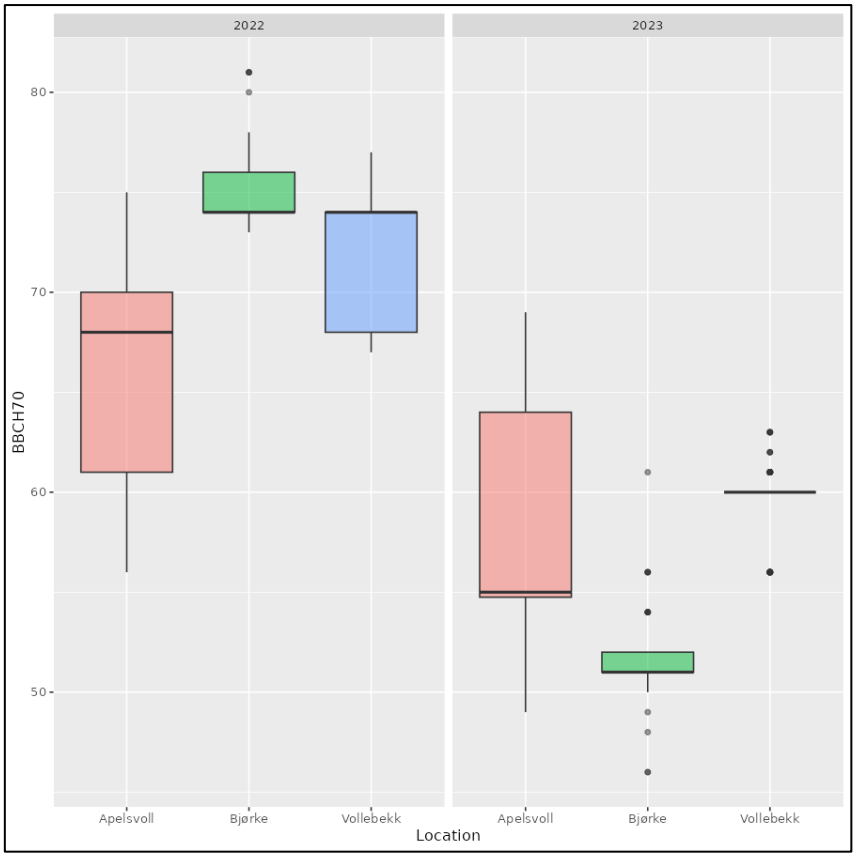

2022
2023
27.5
25.0
22.5
20.0
17.5
15.0
PRT
Apelsvoll
Bjørke
Vollebekk
Location

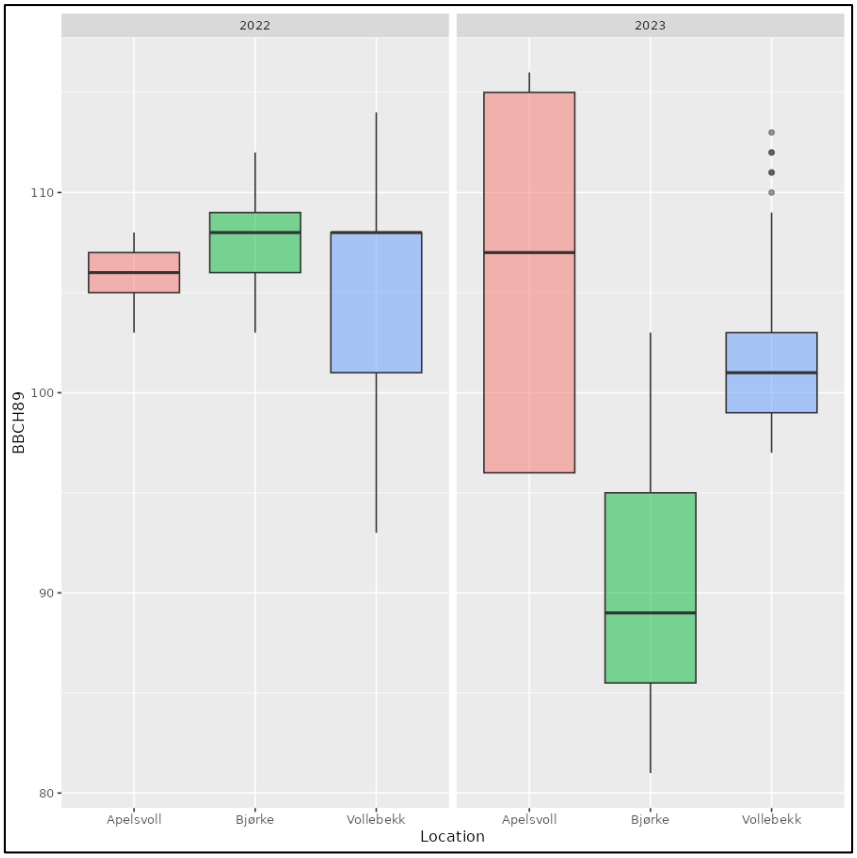

2022
2023
160
120
80
40
0
PRT_YLD
Apelsvoll
Bjørke
Vollebekk
Location


2022
2023
120
100
80
60
40
PLH_max
Bjørke
Vollebekk
Bjørke
Vollebekk
Location


2022
2023
100
75
50
25
0
PLH_har
Apelsvoll
Bjørke
Vollebekk
Apelsvoll
Bjørke
Vollebekk
Location


2022
2023
BBCH60
60
50
40
30
Apelsvoll
Bjørke
Vollebekk
Location

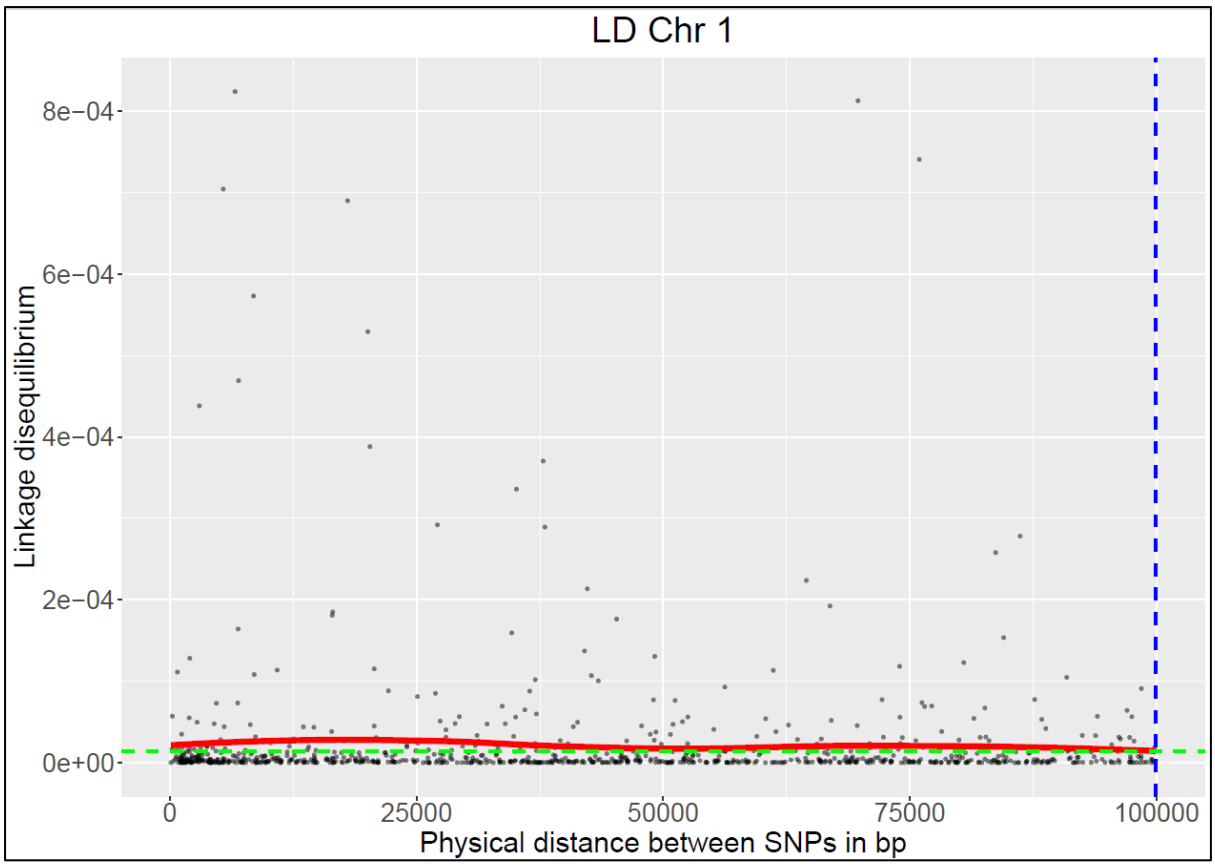

2022
2023
BBCH70
80
70
60
50
Apelsvoll
Bjørke
Vollebekk
Location

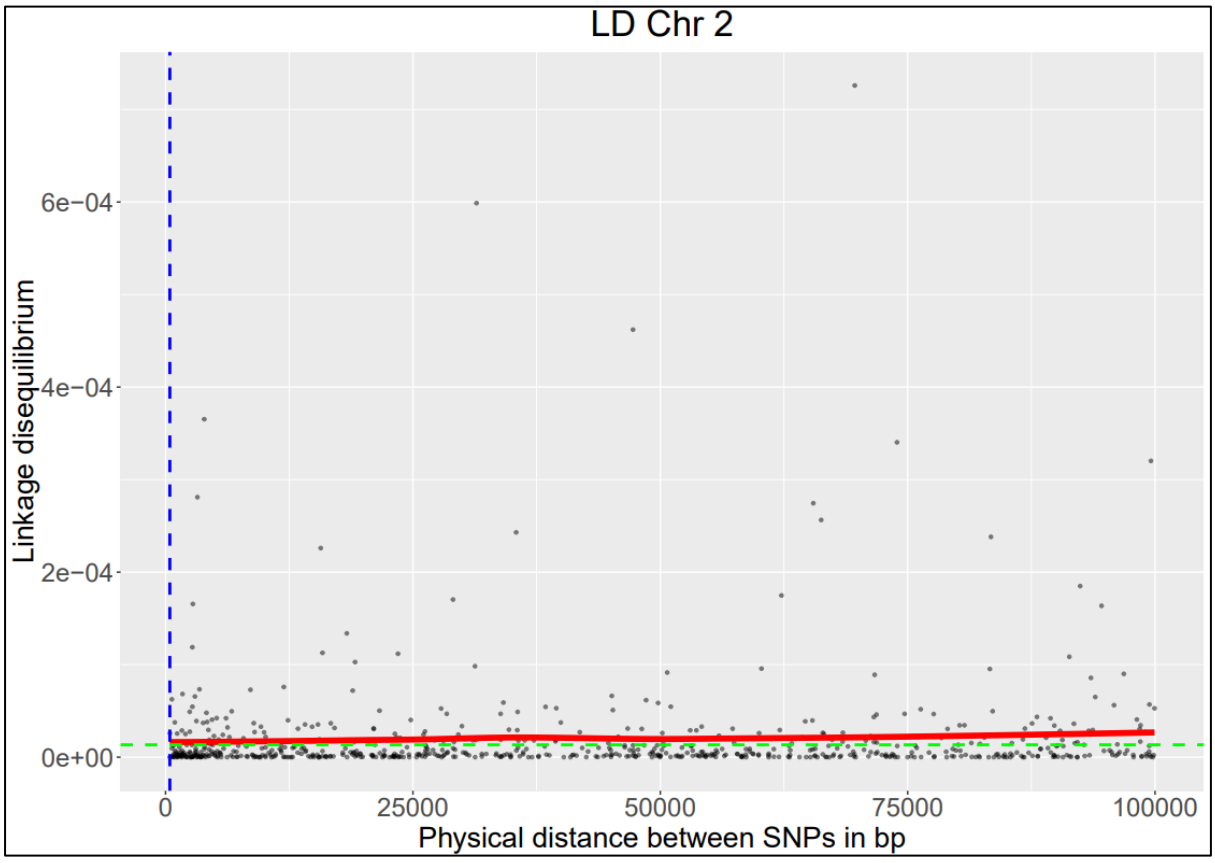


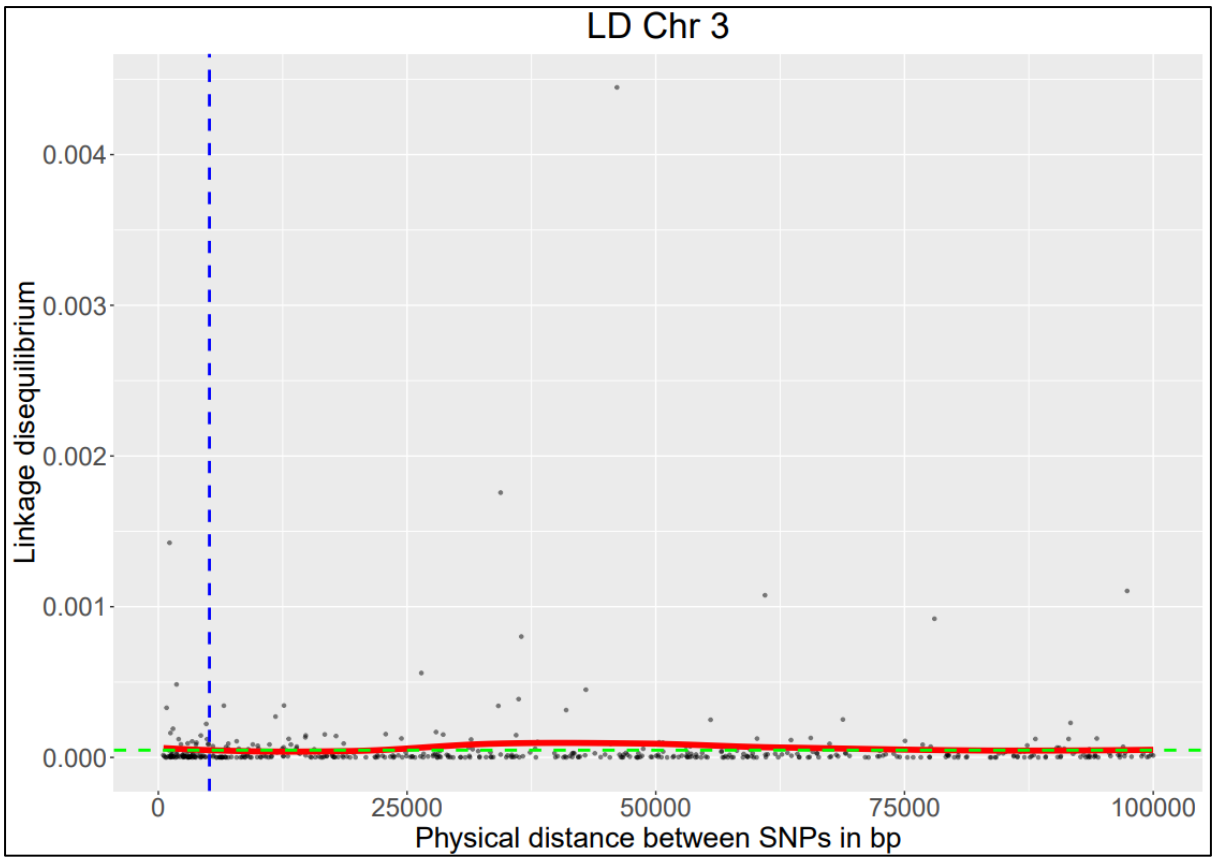
S9. Genotypic correlations for the traits measured across two years and three locations.

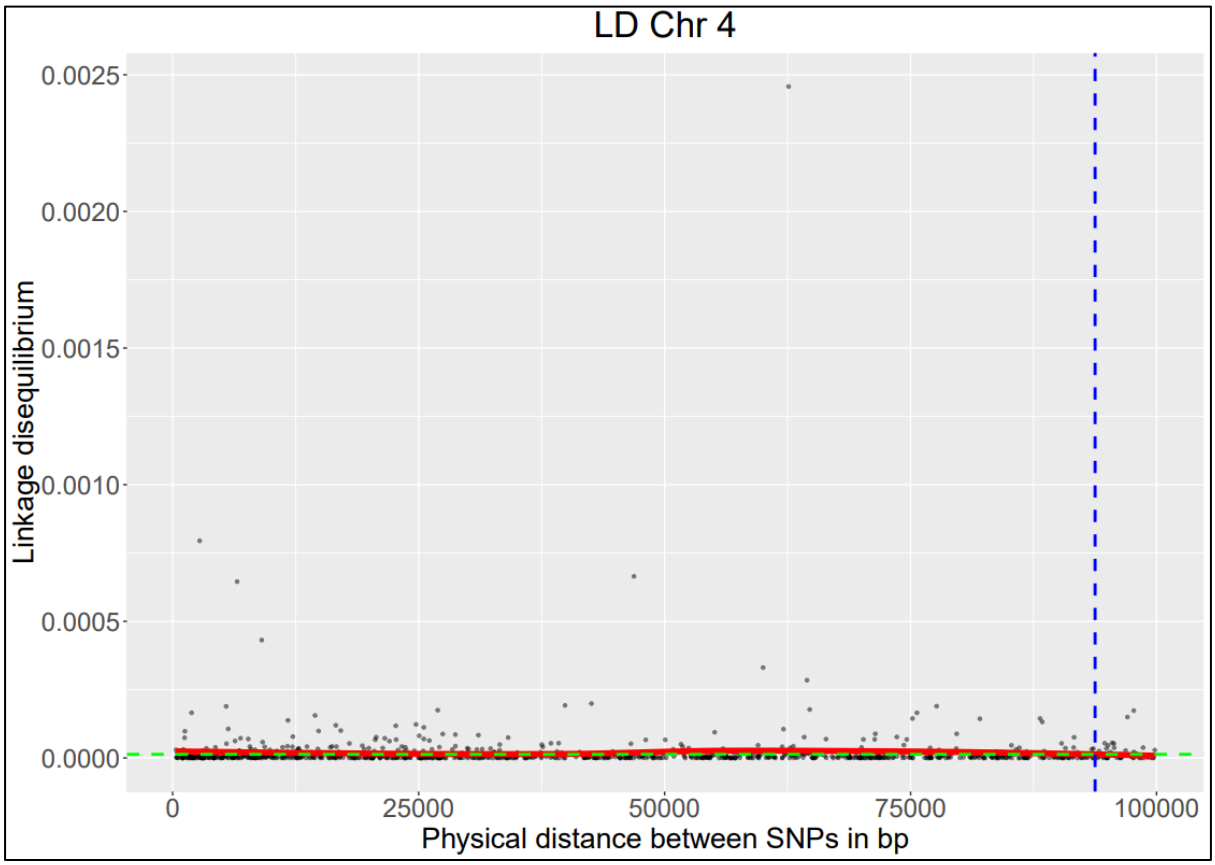
	TSW	PRT	PRT_YLD	PLH_max	PLH_har	BBCH60	BBCH70	BBCH89
YLD	0.24	-0.22	0.86	0.46	0.18	0.53	0.39	0.33
TSW		0.16	0.16	0.35	0.15	0.32	0.41	0.34
PRT			0.41	0.21	-0.05	0.14	0.30	0.36
PRT_YLD				0.38	0.34	0.52	0.46	0.54
PLH_max					0.12	0.59	0.70	0.63
PLH_har						0.03	-0.01	0.04
BBCH60							0.89	0.83
BBCH70								0.89

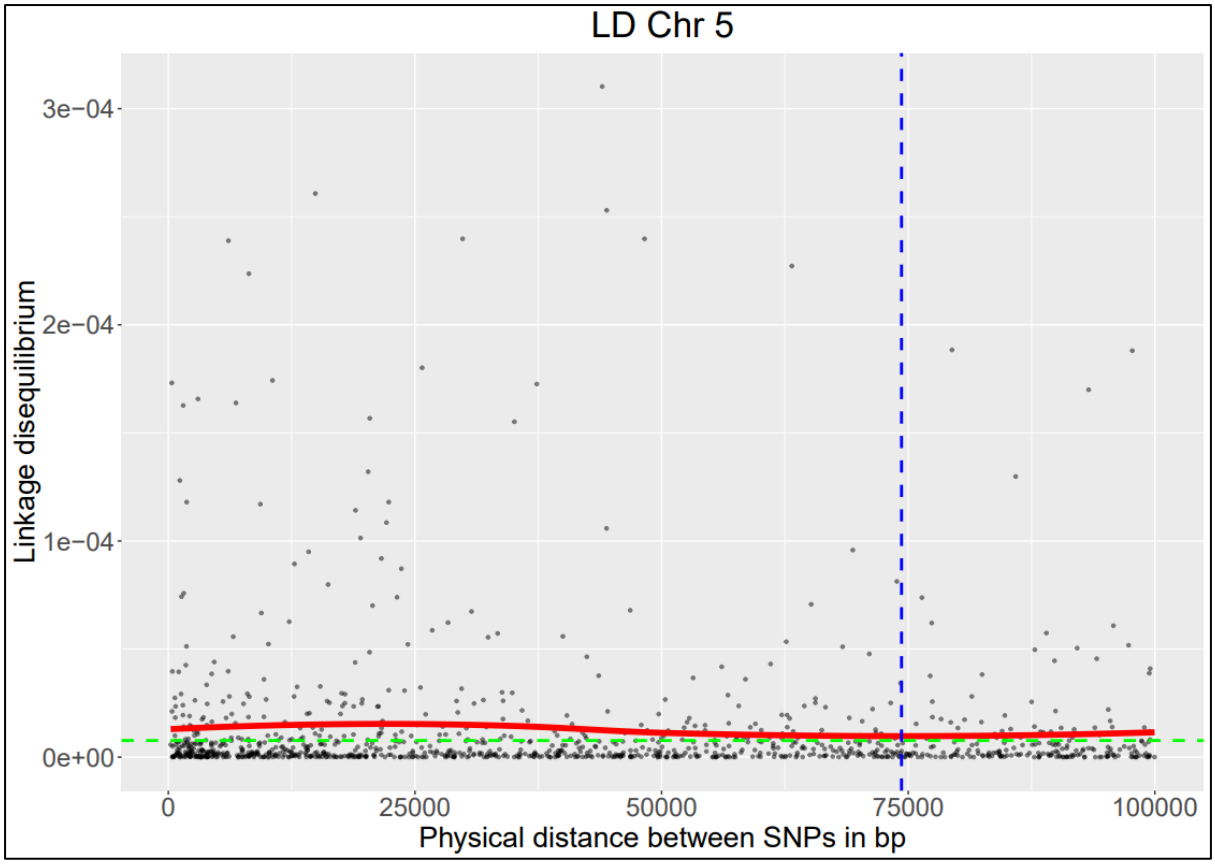
S10. Linkage disequilibrium (LD) decay for each chromosome as squared correlations of allele frequencies (r^2) between markers with a maximum distance of 1 Mbp. The X-axis shows the genomic distance in bp. The blue dotted line indicates the intersection between the LOESS curve (red) and the half of the average value at the minimal distance (dashed green line), highlighting the value of LD decay in base pairs (bp).

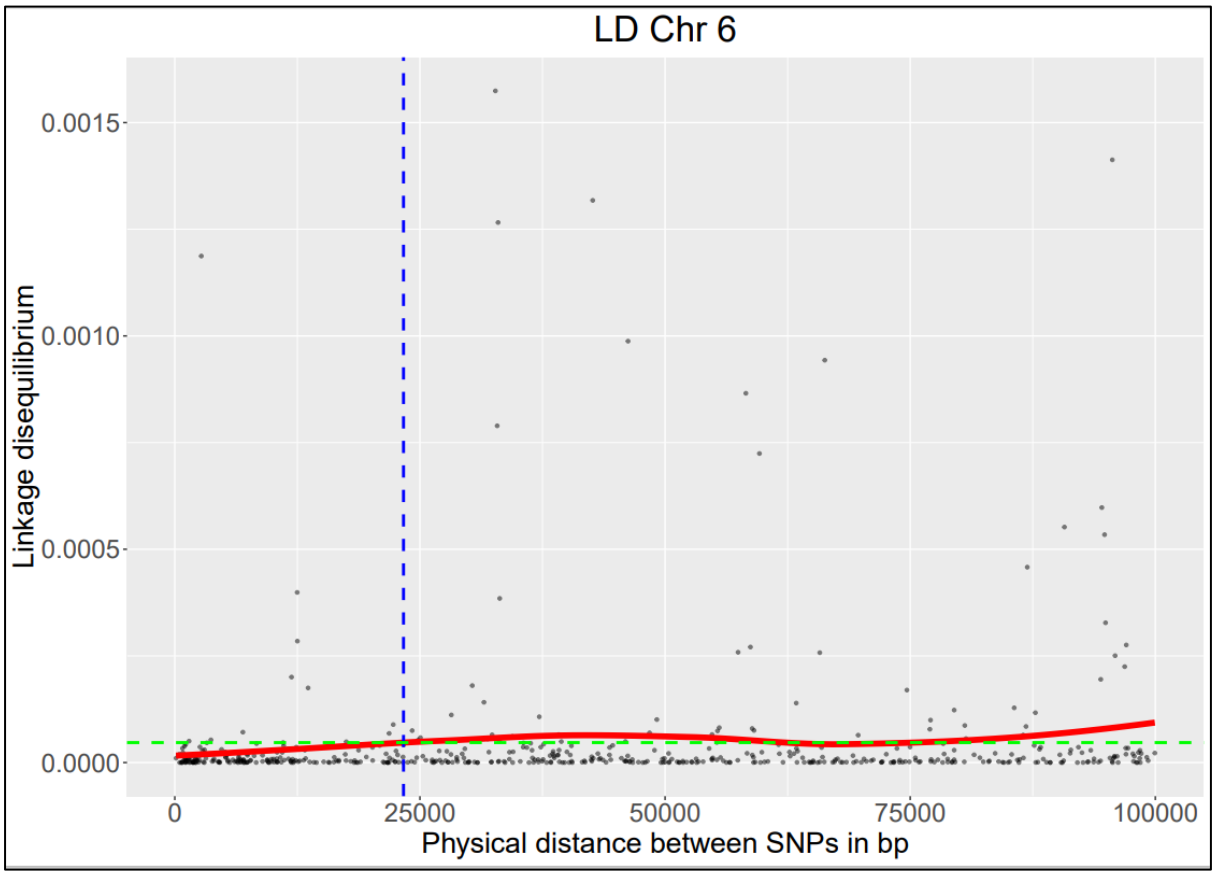

LD Chr 1
Linkage disequilibrium
8e−04
6e−04
4e−04
2e−04
0e+00
0
25000
50000
75000
100000
Physical distance between SNPs in bp

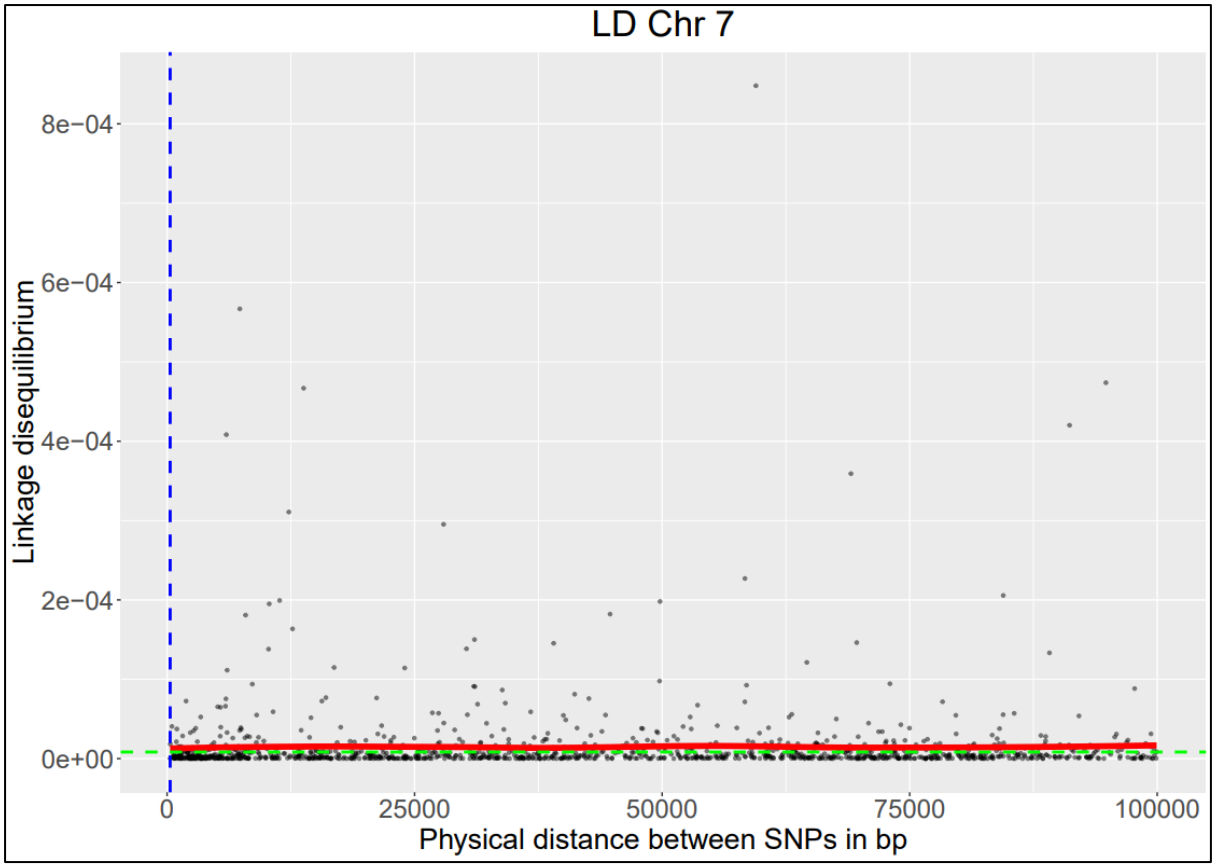

LD Chr 2
Linkage disequilibrium
6e−04
4e−04
2e−04
0e+00
0
25000
50000
75000
100000
Physical distance between SNPs in bp


LD Chr 3
Linkage disequilibrium
0.004
0.003
0.002
0.001
0.000
0
25000
50000
75000
100000
Physical distance between SNPs in bp

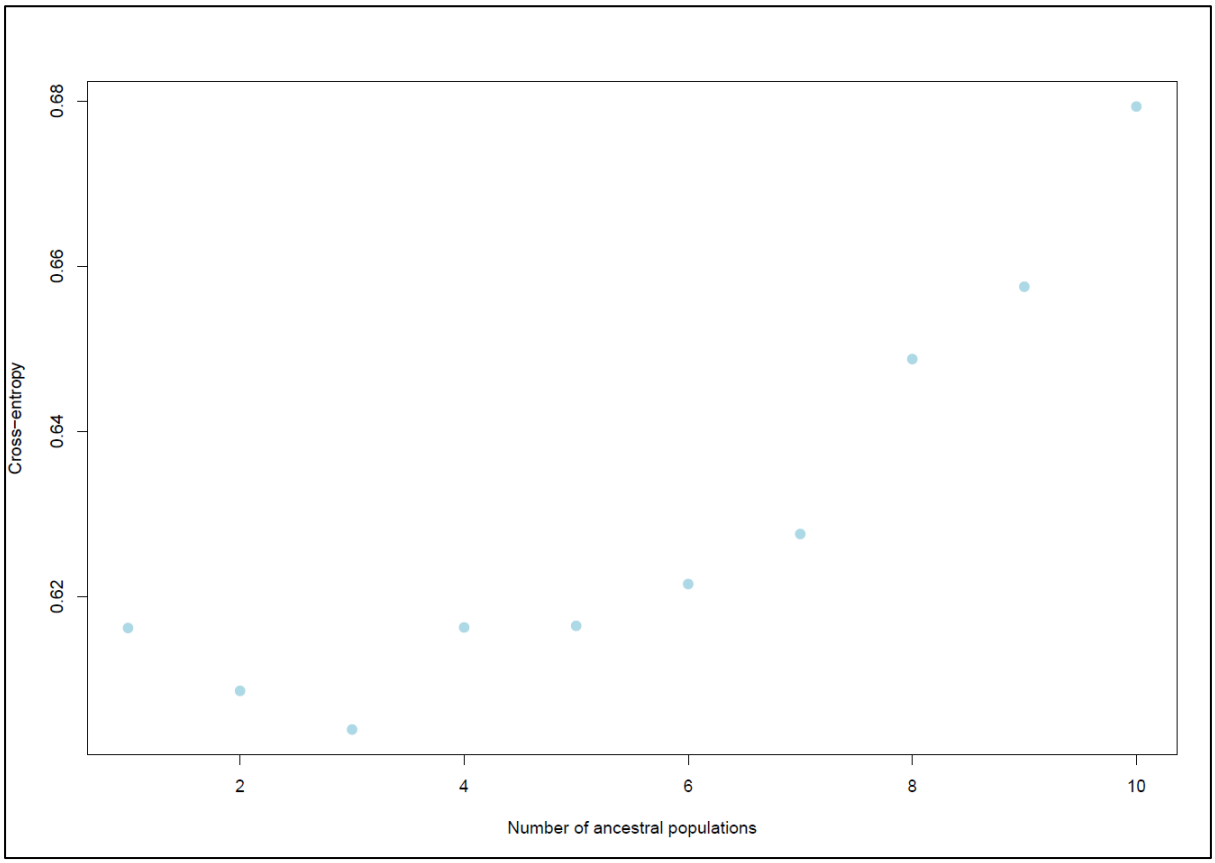

LD Chr 4
Linkage disequilibrium
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
25000
50000
75000
100000
Physical distance between SNPs in bp


LD Chr 5
3e−04
2e−04
1e−04
0e+00
Linkage disequilibrium
0
25000
50000
75000
100000
Physical distance between SNPs in bp

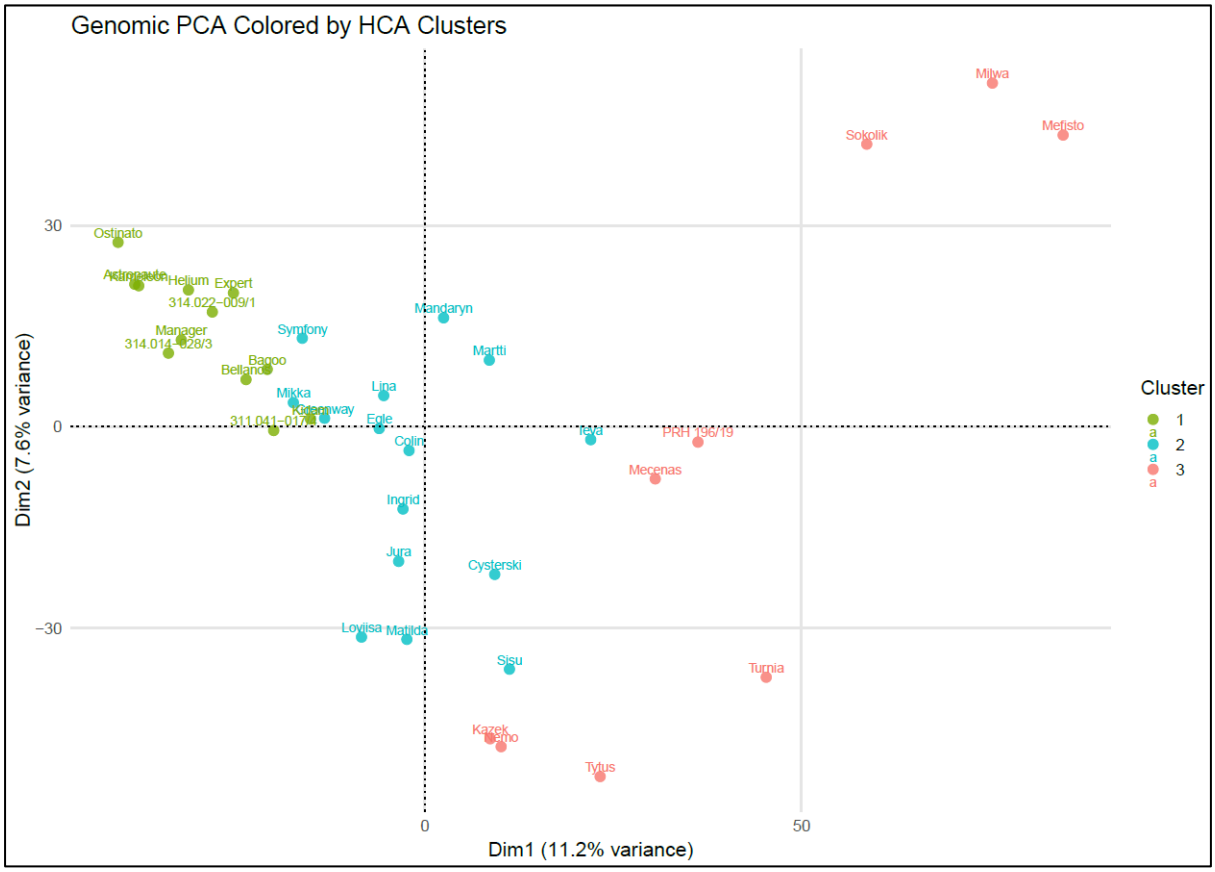

LD Chr 6
0.0015
0.0010
0.0005
0.0000
Linkage disequilibrium
0
25000
50000
75000
100000
Physical distance between SNPs in bp



S11. Optimal number of genetic clusters following a clustering algorithm (k-means), the number of clusters with the lowest cross entropy (3) was chosen to perform the DAPC.



S12. Population structure as PCA performed on the matrix of filtered and imputed SNPs for the 36 accessions included in the field trials in 2022 and 2023 at all locations. Accessions were colored depending on the cluster they were assigned by the HCA.



References

Ali, B., Kumar, S., Ahmed, W., 2021. Genetic Variability, Heritability and Correlation Coefficient in Production Traits of Pea (<i>Pisum sativum</i> L.) Genotypes. Int. J. Genet. Genomics 9, 78. https://doi.org/10.11648/j.ijgg.20210904.13

Annicchiarico, Paolo, Nazzicari, N., Pecetti, L., Romani, M., Ferrari, B., Wei, Y., Brummer, E.C., 2017. GBS-Based Genomic Selection for Pea Grain Yield under Severe Terminal Drought. Plant Genome 10, plantgenome2016.07.0072. https://doi.org/10.3835/plantgenome2016.07.0072

Annicchiarico, P., Nazzicari, N., Pecetti, L., Romani, M., Russi, L., 2019. Pea genomic selection for Italian environments. BMC Genomics 20, 603. https://doi.org/10.1186/s12864-019-5920-x

Annicchiarico, P., Romani, M., Cabassi, G., Ferrari, B., 2017. Diversity in a pea (Pisum sativum) world collection for key agronomic traits in a rain-fed environment of Southern Europe. Euphytica 213, 245. https://doi.org/10.1007/s10681-017-2033-y

Annicchiarico, P., Royo, C., Bellah, F., Moragues, M., 2009. Relationships among adaptation patterns, morphophysiological traits and molecular markers in durum wheat. Plant Breed. 128, 164–171. https://doi.org/10.1111/j.1439-0523.2008.01557.x

Carlson-Nilsson, U., Aloisi, K., Vågen, I.M., Rajala, A., Mølmann, J.B., Rasmussen, S.K., Niemi, M., Wojciechowska, E., Pärssinen, P., Poulsen, G., Leino, M.W., 2021. Trait Expression and Environmental Responses of Pea (Pisum sativum L.) Genetic Resources Targeting Cultivation in the Arctic. Front. Plant Sci. 12. https://doi.org/10.3389/fpls.2021.688067

Červenski, J.F., Danojević, D., Savić, A., 2017. Chemical composition of selected winter green pea (Pisum sativum L.) genotypes. J. Serbian Chem. Soc. 82, 1237–1246. https://doi.org/10.2298/JSC170323094C

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, 61(6), 1-36. URL http://www.jstatsoft.org/v61/i06/

Crosta, M., Nazzicari, N., Ferrari, B., Pecetti, L., Russi, L., Romani, M., Cabassi, G., Cavalli, D., Marocco, A., Annicchiarico, P., 2022. Pea Grain Protein Content Across Italian Environments: Genetic Relationship With Grain Yield, and Opportunities for Genome-Enabled Selection for Protein Yield. Front. Plant Sci. 12.

Crosta, M., Romani, M., Nazzicari, N., Ferrari, B., Annicchiarico, P., 2023. Genomic prediction and allele mining of agronomic and morphological traits in pea (Pisum sativum) germplasm collections. Front. Plant Sci. 14, 1320506. https://doi.org/10.3389/fpls.2023.1320506

Daba, S.D., Morris, C.F., 2022. Pea proteins: Variation, composition, genetics, and functional properties. Cereal Chem. 99, 8–20. https://doi.org/10.1002/cche.10439

de Visser, C.L.M., Schreuder, R., Stoddard, F., 2014. The EU's dependency on soya bean import for the animal feed industry and potential for EU produced alternatives. OCL 21, D407. https://doi.org/10.1051/ocl/2014021

Ferrari, B., Romani, M., Aubert, G., Boucherot, K., Burstin, J., Pecetti, L., Huart, M., Klein, A., Annicchiarico, P., 2016. Association of SNP markers with agronomic and quality traits of field pea in Italy. Czech J. Genet. Plant Breed. 52, 83–93. https://doi.org/10.17221/22/2016-CJGPB

Frichot, E., Francois, O. (2015). "LEA: an R package for Landscape and Ecological Association studies."_Methods in Ecology and Evolution_. <http://membres-timc.imag.fr/Olivier.Francois/lea.html>.

Fox J, Weisberg S (2019). _An R Companion to Applied Regression_, Third edition. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.

Gauch Jr., H.G., 2013. A Simple Protocol for AMMI Analysis of Yield Trials. Crop Sci. 53, 1860–1869. https://doi.org/10.2135/cropsci2013.04.0241

Gutierrez, N., Pégard, M., Solis, I., Sokolovic, D., Lloyd, D., Howarth, C., Torres, A.M., 2024. Genome-wide association study for yield-related traits in faba bean (Vicia faba L.). Front. Plant Sci. 15. https://doi.org/10.3389/fpls.2024.1328690

Husson, F., Le, S., and Josse, J. (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 25(1), 1-18. 10.18637/jss.v025.i01

Jha, A.B., Arganosa, G., Tar'an, B., Diederichsen, A., Warkentin, T.D., 2013. Characterization of 169 diverse pea germplasm accessions for agronomic performance, Mycosphaerella blight resistance and nutritional profile. Genet. Resour. Crop Evol. 60, 747–761. https://doi.org/10.1007/s10722-012-9871-1

Jombart, T. and Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. doi: 10.1093/bioinformatics/btr521

Karjalainen, R., Kortet, S., 1987. Environmental and genetic variation in protein content of peas under northern growing conditions and breeding implications. Agric. Food Sci. 59, 1–9. https://doi.org/10.23986/afsci.72238

Kassambara A, Mundt F. (2020). _factoextra: Extract and Visualize the Results of Multivariate Data Analyses_. R package version 1.0.7, <https://CRAN.R-project.org/package=factoextra>.

Sadras, V.O., Lake, L., Kutchartt, D.G., Calderini, D.F., 2021. Critical developmental period for grain yield and grain protein concentration in lentil. Field Crops Res. 270, 108203. https://doi.org/10.1016/j.fcr.2021.108203

Mohammed, Y.A., Chen, C., Walia, M.K., Torrion, J.A., McVay, K., Lamb, P., Miller, P., Eckhoff, J., Miller, J., Khan, Q., 2018. Dry pea (Pisum sativum L.) protein, starch, and ash concentrations as affected by cultivar and environment. Can. J. Plant Sci. 98, 1188–1198. https://doi.org/10.1139/cjps-2017-0338

Muñoz F, Sanchez L (2020). _breedR: Statistical Methods for Forest Genetic Resources Analysts_. R package version 0.12-5, <https://github.com/famuvie/breedR>.

Pavan, S., Delvento, C., Nazzicari, N., Ferrari, B., D'Agostino, N., Taranto, F., Lotti, C., Ricciardi, L., Annicchiarico, P., 2022. Merging genotyping-by-sequencing data from two ex situ collections provides insights on the pea evolutionary history. Hortic. Res. 9, uhab062. https://doi.org/10.1093/hr/uhab062

R Core Team (2025). R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Schafer J, Opgen-Rhein R, Zuber V, Ahdesmaki M, Silva APD, Strimmer. K (2021). _corpcor: EfficientEstimation of Covariance and (Partial) Correlation_. R package version 1.6.10, <https://CRAN.R-project.org/package=corpcor>.

Singh, R., Singh, P.M., 2011. Effect of sowing dates and varieties on yield and quality of garden pea seed. Veg. Sci. 38, 184–187.

Sjulgård, H., Keller, T., Garland, G., Colombi, T., 2023. Relationships between weather and yield anomalies vary with crop type and latitude in Sweden. Agric. Syst. 211, 103757. https://doi.org/10.1016/j.agsy.2023.103757

Stagnari, F., Maggio, A., Galieni, A., Pisante, M., 2017. Multiple benefits of legumes for agriculture sustainability: an overview. Chem. Biol. Technol. Agric. 4, 2. https://doi.org/10.1186/s40538-016-0085-1

Tolessa, T., 2017. Genetic Variation, Heritability, and Advances from Selection in Elite Breeding Materials of Field Pea (Pisum sativum L.) Genotypes. Agric. Res. Technol. Open Access J. 8. https://doi.org/10.19080/ARTOAJ.2017.08.555744

Tayeh, N., Aluome, C., Falque, M., Jacquin, F., Klein, A., Chauveau, A., Bérard, A., Houtin, H., Rond, C., Kreplak, J., Boucherot, K., Martin, C., Baranger, A., Pilet-Nayel, M.-L., Warkentin, T.D., Brunel, D., Marget, P., Le Paslier, M.-C., Aubert, G., Burstin, J., 2015. Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high-density, high-resolution consensus genetic map. Plant J. 84, 1257–1273. https://doi.org/10.1111/tpj.13070

Tran, C.T., Becker, H.C., Horneburg, B., 2022. Agronomic performance of normal-leafed and semi-leafless pea (Pisum sativum L.) genotypes. Crop Sci. 62, 1430–1442. https://doi.org/10.1002/csc2.20746

Tran, C.T., Beissinger, T.M., Becker, H.C., Horneburg, B., 2023. Genetic diversity of pea (Pisum sativum L.) genotypes differing in leaf type using SNP markers. Genet. Resour. Crop Evol. 70, 1085–1095. https://doi.org/10.1007/s10722-022-01487-3

Tzitzikas, E.N., Vincken, J.-P., de Groot, J., Gruppen, H., Visser, R.G.F., 2006. Genetic Variation in Pea Seed Globulin Composition. J. Agric. Food Chem. 54, 425–433. https://doi.org/10.1021/jf0519008

Uhlarik, A., Ćeran, M., Živanov, D., Grumeza, R., Skøt, L., Sizer-Coverdale, E., Lloyd, D., 2022. Phenotypic and Genotypic Characterization and Correlation Analysis of Pea (Pisum sativum L.) Diversity Panel. Plants 11, 1321. https://doi.org/10.3390/plants11101321

van Loon, M.P., Alimagham, S., Pronk, A., Fodor, N., Ion, V., Kryvoshein, O., Kryvobok, O., Marrou, H., Mihail, R., Mínguez, M.I., Pulina, A., Reckling, M., Rittler, L., Roggero, P.P., Stoddard, F.L., Topp, C.F.E., van der Wel, J., Watson, C., van Ittersum, M.K., 2023. Grain legume production in Europe for food, feed and meat-substitution. Glob. Food Secur. 39, 100723. https://doi.org/10.1016/j.gfs.2023.100723

Vanhala, T., Normann, K.R., Lundström, M., Weller, J.L., Leino, M.W., Hagenblad, J., 2016. Flowering time adaption in Swedish landrace pea (Pisum sativum L.). BMC Genet. 17, 117. https://doi.org/10.1186/s12863-016-0424-z

Weber, E., Bleiholder, H., 1990. Erläuterungen zu den BBCH-Dezimal-Codes für die Entwicklungsstadien von Mais, Raps, Faba-Bohne, Sonnenblume und Erbse-mit Abbildungen.

Erläut. Zu Den BBCH-Dezimal-Codes Für Entwicklungsstadien Von Mais Raps Faba-Bohne Sonnenblume Erbse-Mit Abbild. 42, 308–321.

Zanotto, S., Palmé, A., Helgadóttir, Á., Daugstad, K., Isolahti, M., Öhlund, L., Marum, P., Moen, M.A., Veteläinen, M., Rognli, O.A., Ergon, Å., 2021. Trait characterization of genetic resources reveals useful variation for the improvement of cultivated Nordic red clover. J. Agron. Crop Sci. https://doi.org/10.1111/jac.12487