Journal of Agriculture and Food Research 24 (2025) 102498

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

# Genetic determinants of leaf protein content and extractability in sugar beet: Opportunities for leaf biomass valorisation

Hugo Rijken [a], Aamir Saleem [a], Juan V. Renedo [c], Alain Tossens [c], Marieke E. Bruins [b], Luisa M. Trindade [a,*]

[a] Plant Breeding. Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB, Wageningen, the Netherlands
[b] Wageningen Food & Biobased Research, Wageningen University & Research, Bornse Weilanden 9, 6708 WG, Wageningen, the Netherlands
[c] United Beet Seeds B.V., Industriepark 15, 3300, Tienen, Belgium

ARTICLE INFO

*Keywords:*
Side stream valorisation
Leaf nitrogen content
Whole crop use
Alternative proteins
Genome-wide association study

ABSTRACT

Sugar beet (Beta vulgaris L.) leaves represent an underutilized agricultural byproduct with potential for valorisation through leaf protein extraction. However, current extraction yields are too low to render this protein source economically viable. To date, variation in leaf protein yield among sugar beet germplasm remains largely unexplored. To assess this variation and unravel the genetic architecture underlying relevant traits, we conducted a genome-wide association study (GWAS). A panel of 180 sugar beet accessions was evaluated across three field trials in the Netherlands. Phenotyping included total and extractable leaf protein content, extractable RuBisCO, protein extractability, dry matter content, and leaf yield. Moderate heritability was observed for total protein content and dry matter content, while the extractable protein fractions, protein extractability and leaf yield exhibited low heritability. The GWAS identified 45 significant marker–trait associations and underlying quantitative trait loci (QTLs) related to key traits influencing leaf protein extraction. Candidate gene analysis revealed several genes with putative roles in RuBisCO assembly, chloroplast development, and chloroplast protein import. Our findings indicate that breeding for higher total leaf protein content is feasible, while gains through enhanced extractability may be limited. This study lays the foundation for the genetic improvement of sugar beet leaf protein traits. However, further functional validation of the identified QTLs and candidate genes is essential to confirm their practical utility.

## 1. Introduction

Sugar beet (*Beta vulgaris* L. subsp. *vulgaris*) is a widely grown crop for the extraction of sugar from its tap roots. It is estimated that around 20 % of the world's sugar is derived from sugar beet and about half of this is produced within the European Union [1]. While the sugar beet is grown for its tap root, about 20–34 % of the plant's biomass are leaves [2]. These leaves can be valorised by the utilisation of the leaf protein. It is estimated that 1 ha of sugar beet produces about 500 kg of leaf protein annually [3], but currently the leaves are not collected and are returned to the soil during beet harvest. Valorising green leaf biomass offers opportunity to enhance the efficiency and sustainability of food systems while providing farmers with additional income. Leaf protein extraction can yield high quality protein isolates that can be used as a food ingredient. One of the proteins of interest is the photosynthesis enzyme Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). RuBisCO is a soluble protein that constitutes approximately 25 % of the total protein in leaves and its protein isolate shows comparable or superior functional properties to that of whey and soy protein isolates [4].

Sugar beet leaves contain approximately 22.8 % protein on a dry matter basis [5], yet current extraction methods recover only around 6 % of the total protein [6]. Protein extraction typically involves mechanically disrupting the leaves to break open the cell walls and release intracellular contents, including proteins. The resulting juice is then further processed, fractionated and concentrated, to obtain a leaf protein isolate. At this stage, fractionation methods often rely on the solubility and re-solubilization potential of RuBisCO. Several studies have focused on understanding and improving the process of leaf protein extraction [7,8]. Endogenous factors such as the plant cell wall and phenolic compounds are recognized as key barriers to efficient extraction. The cell wall, in particular, poses a physical barrier to protein release. Research in tomato has shown that variation in cell wall composition

* Corresponding author.
*E-mail address:* luisa.trindade@wur.nl (L.M. Trindade).

https://doi.org/10.1016/j.jafr.2025.102498
Received 28 August 2025; Received in revised form 23 October 2025; Accepted 30 October 2025
Available online 31 October 2025
2666-1543/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

can influence protein extractability; specifically, higher levels of pectin-derived galacturonic acid were negatively correlated with extractability [9]. Furthermore, enzyme-assisted extraction methods that incorporate cell wall-degrading enzymes show improved protein yields from sugar beet leaves [10]. Phenolic compounds are also known to hinder protein extractability, due to their ability to form complexes with proteins [11]. These interactions can induce structural changes in the proteins, reducing their solubility and, consequently, their extractability [12–14]. Developmental factors have also been reported to influence extractability. Protein extractability decreased in aging tomato leaves, due to senescence-regulated protein degradation [15]. Plant age has also been suggested to affect extractability, with reduced extraction efficiency in older plants [16], although no consistent effect was observed for sugar beet leaves harvested at 3–8 months after sowing [17].

Genetic variation may also play a significant role in determining protein extraction yield. In various crops, studies have reported genotypic differences in total leaf protein content [18–21], as well as in the levels of soluble protein fractions specifically [22–24]. When it comes to protein extractability, evidence of genotypic variation has been reported in tomato, where [9] found that the wild species *Solanum pennellii* exhibited notably lower extractability than commercial tomato varieties of *Solanum lycopersicum*. In sugar beet, the extent of variation in leaf protein content and extractability across different germplasm has not yet been characterized. However, given the intraspecific variation observed in other crops, it is reasonable to expect similar diversity within sugar beet. The existence of genetic variability opens a promising avenue for improving leaf protein yield through the selective breeding of sugar beet varieties that are more suitable for leaf protein extraction.

To accelerate breeding for leaf protein yield, it is beneficial to identify and understand the genetic basis of the trait. This information can help to speed up the selection process and to understand the biological mechanism(s) that control the trait's variation. Genome wide association studies (GWAS) are a powerful approach to associate DNA markers to variation in complex traits. In sugar beet, GWAS studies have led to the detection of quantitative trait loci (QTL) for important agronomic traits like low bolting tendency [25], sugar content and root yield [26]. These QTL can be utilized to speed up the selection of the trait, saving valuable time and money. Furthermore, the identification and investigation of genes surrounding the QTL can help to give a better understanding of the mechanisms involved in leaf protein extraction. In sugar beet, research on the genetics basis of leaf protein content and extraction yield remains limited, with only few studies exploring these traits at the genomic level. GWAS offers a valuable approach to uncover the genetic mechanisms underlying variation in leaf protein traits and to identify candidate loci for future breeding.

In this study, a GWAS approach was performed to identify genetic markers associated with key traits relevant to leaf protein extraction: total leaf protein content, extractable protein content, extractable RuBisCO, protein extractability, dry matter content, and leaf yield. The aim was to develop molecular tools that can support the breeding of sugar beet varieties with improved protein extraction yield. A panel of 180 sugar beet accessions was evaluated across three field trials conducted in The Netherlands. Leaves were harvested at the end of the growing season to reflect a realistic harvest scenario, where leaves are collected as a byproduct while the beet root remains the primary yield component.

## 2. Methods

### 2.1. Plant material and field trials

The plant material used in this study consisted of 180 *Beta vulgaris* accessions that were provided by the sugar beet breeding company SESVanderHave (SvdH; Tienen, Belgium). The accessions are three-way hybrids, which resulted from crossings between a homozygous pollinator line and a hybrid component utilized as a tester. Two testers and 180 pollinators were used to create the panel, with 144 pollinators crossed with one tester, and 36 pollinators crossed with the second tester. The panel was evaluated in three field trials across two locations in 2022 and 2023 (Table 1). In trial WOU22, an alpha-lattice design with two replications per accession was used. In trials NOP23 and WOU23, a randomized complete block design (RCBD) was used, also with two replicates. The plants were sown in plots with three rows of 30 plants. The plots were located next to each other, without border spacing. To avoid cross contamination during sampling, we used only the middle row of every plot for phenotyping. According to the environmental data of the trial locations (Supplementary Fig. S1), daily temperature, total precipitation and daily solar radiations varied both within trials along the growing seasons, and across the three trials. Total precipitation ranged from 328 to 380 mm in WOU22 and WOU23 to 205 mm in NOP23. Average daily temperature was slightly lower in WOU22 (16.6 $\pm$ 4.0 °C) than WOU23 (17.5 $\pm$ 0.2 °C) and NOP23 (17.2 $\pm$ 3.0 MJ $m^{-2}$). Similarly, WOU22 received less daily solar radiations (18.3 $\pm$ 5.9 MJ $m^{-2}$) as compared to WOU23 (19.6 $\pm$ 5.8) and NOP23 (19.7 $\pm$ 6.8 MJ $m^{-2}$). These environmental differences may have contributed variation in leaf protein traits among trials.

### 2.2. Sample collection

To emulate the scenario where leaves are harvested as a byproduct to the beet harvest, we collected leaf samples at the end of the growing season (Table 1). Leaves were harvested without the petiole and collected in 50 $\times$ 80 cm monofilament net bags. For trial WOU22, all leaves from 10 plants in a plot's middle row were collected. In 2023, leaf sampling included all the plants in the middle row, with 3–4 young fully expanded leaves collected from each plant.

### 2.3. Sample processing

The bulk leaf sample was mixed and split into two subsamples (A and B). Subsample A was used to determine the dry matter and total protein content of the leaves. Subsample B was used to determine the amount of extractable protein. To determine extractable protein and protein extractability of the total sample, we used the parameters of both subsamples (A and B). In trial WOU22, the dry leaf yield per plant was determined by measuring the fresh weight of the bulk leaf sample before splitting. Leaf yield (LY) was calculated using the fresh weight of the bulk leaf sample, the dry matter content of subsample A, and the number of plants sampled (equation (1)).

$$Leaf\ yield = \frac{Fresh\ weight\ (bulk) * Dry\ matter\ proportion(A)}{Nplants} \quad (1)$$

After splitting of the bulk leaf sample, we placed subsample A back into the monofilament net bag and recorded the fresh weight of the leaves. The leaves were then dried at 50 °C for 48 h. The dried leaves were weighed, milled (1 mm) and stored in airtight containers. The fresh and dry weight were used to calculate the dry matter content of the leaves. Near-infrared spectral absorbance (400–2500 nm) of the dried and milled leaf samples were recorded with a FOSS NIRs analyzer (FOSS

**Table 1**
Field trials. Overview of field trials, including trial code, location, date of sowing and leaf harvest, and age at harvest.

| Trial | Location | Sowing | Leaf harvest | Age at harvest (days) |
|---|---|---|---|---|
| WOU22 | Wouwse Plantage (NL) | April 22, 2022 | October 3, 2022 | 164 |
| NOP23 | Emmeloord (NL) | May 16, 2023 | August 28, 2023 | 104 |
| WOU23 | Wouwse Plantage (NL) | May 2, 2023 | September 18, 2023 | 139 |

DS2500). NIRs based prediction models were previously developed in Ref. [27], and used in this study to predict total leaf protein content (LPC), and extractable RuBisCO content (ER) of the samples.

Subsample B was collected in airtight and lightproof aluminium zipbags (30 × 50 cm, 9 l) and stored at 4 °C to minimize degradation of the leaves. Within three days after harvest, a twin screw press (Angel Juicer 7500, Angel Juicer Co. LTD, South-Korea) was used to press the leaves. The juice was collected and weighed to determine the percentage of juice pressed from the leaves. A 2 ml subsample of the juice was taken and centrifuged at 20,000 g for 45 min to precipitate any solids. A subsample of the supernatant (500 μl) was transferred into a tin capsule (20 × 8 mm, Elemental microanalysis), and dried for 48 h at 60 °C. The nitrogen content of the supernatant was measured through combustion of the dried tin capsules in a rapid N-exceed nitrogen and protein analyzer (Elementar).

The extractable protein content and protein extractability of the bulk leaf sample was calculated by using parameters measured on both subsamples separately. To do this, the dry matter and total protein content of subsample A were combined with the percentage of juice and the supernatant protein of subsample B. Extractable protein (g/100 g dried leaves) was determined as the ratio of protein in the juice and the dry matter content of the leaves, as shown in equation (2).

$$Extractable\ protein = \frac{(Percentage\ of\ juice\ (B) * Protein\ in\ supernatant(B))}{Dry\ matter\ content\ (A)} \quad (2)$$

The extractable protein was then expressed as a percentage of total protein to obtain protein extractability (g/100 g protein) in equation (3):

$$Protein\ extractability = \frac{Extracted\ protein}{Total\ protein} * 100 \quad (3)$$

### 2.4. Statistical analysis of the phenotype data

Phenotype data was corrected for spatial heterogeneity in the field by using the SpATS mixed model approach, which utilizes 2-dimensional smoothing with P-splines [28]. In the mixed model, genotype was set as a fixed factor to obtain the best linear unbiased estimates (BLUEs) of genotypic means for each trial. An additional mixed model was fitted where genotype was set as a random factor to obtain the broad sense heritability of the phenotypes per trial. The BLUEs of genotypic means as well as the broad sense heritability estimates were also obtained across the three environments. To obtain BLUEs across trials, a linear mixed model (equation (4)) was fitted using the lme4 package in R [29].

$$Y = Intercept + Genotype + Environment + Block + GEI + Residual \quad (4)$$

Where, 'Y' is response variable, 'Intercept' represents the mean value of the response variable, 'Genotype' is a fixed effect of the accession, 'Environment' is a fixed effect of the trial, 'Block' is a random effect of the block nested within the trial, 'GEI' is a random effect of the genotype by environment interaction, and 'Residual' represents the error term. The significance of the GEI term was tested using the 'ranova' function implemented in the lmerTest R package [30]. To obtain broad-sense heritability estimates, the model in equation (4) was refitted, but with Genotype set as a random factor to obtain the variance components. These variance components were then used to calculate the broad sense heritability in the following formula:

$$H^2 = VG \Big/ \left( VG + \frac{VGEI}{nE} + \frac{V\varepsilon}{nE * nBlock} \right) \quad (5)$$

Where, 'VG' represents the variance component of Genotype, 'VGEI' represents the variance component of the genotype by environment interaction, '$V\varepsilon$' represents the variance component of the error term, 'nE' represents the number of environments, and 'nBlock' the number of blocks in the trial design.

### 2.5. Genotyping and imputation

Leaf samples from individual pollinator plants, which were used for hybrid production, were selected for DNA fingerprinting analysis. Leaf disks of 1 $cm^2$ were freeze dried at −80 °C and ground before DNA extraction. DNA extraction was performed using an in-house DNA extraction method, yielding an average concentration of approximately 10 ng/μl. For genotyping, a proprietary Axiom array containing 22,332 Single Nucleotide Polymorphisms (SNP) was used. These SNPs are distributed over the nine chromosomes of the sugar beet genome, and were designed to provide uniform genome cover. The majority of the SNPs are located in coding regions. Genotyping analysis with the Axiom array was performed by Eurofins Genomics Europe (AgriGenomics Products & Services A/S) in Denmark. The genotyping matrix was generated using the Axiom Suite Software.

The raw dataset included 12,956 single-nucleotide polymorphism (SNP) markers covering the whole sugar beet genome. The SNPs were filtered and imputed using the codeMarkers function in statgenGWAS R package [31]. Markers with more than 10 % missing values, duplicate markers, and markers with a minor allele frequency below 5 % were first removed from the dataset. In addition, accessions with more than 10 % missing values were also removed. For the remaining dataset, missing values were replaced with a random value calculated using allele frequencies per SNP. In the end, 2536 markers scored on 167 genotypes were used for further analysis.

### 2.6. Population structure

Population structure of the GWAS panel was assessed and visualized through Principal Component Analysis (PCA), followed by K-means clustering to group individuals into genetically similar clusters. PCA was conducted on the imputed SNP dataset using the prcomp function in R, and clustering was performed using the kmeans function from the base stats package R Core Team [32]. The optimal number of clusters (K) was determined using the elbow method. The number of significant PCs was determined using PCAtest in R [33], which evaluates whether the observed correlation structure between variables exceeds random expectation. It does so by comparing eigenvalue-based statistics (Psi and Phi) to null distributions generated through random permutations.

### 2.7. GWAS procedure

Single trait GWAS analysis was performed both within, and across environments depending on the BLUEs used as input. The analysis was performed using the statgenGWAS package [31] in R. Population structure was taken into account by fitting a Van Raden kinship matrix and adding it to the GWAS model. The effectiveness of population structure correction was evaluated by assessing quantile-quantile (QQ) plots of expected versus observed P values. The rate of false GWAS discoveries was controlled following the approach of Brzyski et al. [34] for multiple testing correction. Specifically, we partitioned the genome-wide SNPs into clusters of correlated markers (based on LD threshold = 0.5), retained only the top representative within each cluster (based on marginal P-value), and then applied a modified Benjamini-Hochberg procedure on these representatives to control FDR at the cluster level rather than the SNP level. This approach prevents the over-counting of correlated SNPs as independent discoveries and offers power while still maintaining rigorous error control.

### 2.8. Linkage disequilibrium decay

To define the genomic window size surrounding significant markers for candidate gene exploration, the decay of linkage disequilibrium (LD) was assessed for each chromosome. In the analysis of LD, a correction for

population structure was applied similarly to the method described by Adetunji and co-workers [35]. To obtain LD between pairs of markers, two separate linear regression models were fit. In the first model, linear regression was applied to predict the response of a marker using only a genetic relatedness (GR) component as a predictor. This GR component consisted of significant PCs identified above. In the second model, the paired marker was added to the model as an additional predictor variable, alongside the GR component. The difference in $r^2$ (second model – first model) between the two models was taken to represent LD (expressed as $r^2$) between the pair of markers. If the difference resulted in a negative value, the LD was set to 0. LD was estimated for all marker pairs within 20 Mb of each other. To estimate LD decay over distance, the $r^2$ values were binned in an ascending order with steps of 10 Kb between pair of SNPs for each chromosome, and the 90th percentile $r^2$ value of each bin was taken. These values were used to estimate LD decay over distance by fitting a non-linear model, using the equation from Ref. [36]. The distance where the fitted $r^2$ dropped to half of its maximum value was taken as the size of the genomic region to investigate for candidate gene analysis.

### 2.9. Candidate gene analysis

Genomic regions surrounding significant markers were investigated to gather information on the putative candidate genes involved in the variation of the investigated trait. All genes within the LD decay window, upstream and downstream of the significant marker, were selected from the genome assembly of the EL10.2 sugar beet genome [37]. Based on the gene's annotation, candidate genes were selected that may explain variation in the mapped trait. Special attention was given to candidate genes related to chloroplast development and division, protein synthesis and degradation, protein/peptide transporters, photosynthesis, and genes related to synthesis and break down of the cell wall.

## 3. Results

### 3.1. Phenotypic variation

In a panel of 180 sugar beet (*Beta vulgaris*) accessions, leaves were sampled and leaf protein content (LPC), extractable leaf protein content (ELPC), extractable RuBisCO (ER), protein extractability (PE) and dry matter content (DM) were determined. High phenotypic variation was observed for all traits (Fig. 1). LPC of sugar beet leaves ranged from 18.56 % to 30.93 %. A difference in LPC was also observed between trials: samples from trial WOU23 had a mean value of 21.82 %, trial WOU22 26.05 % and NOP23 27.19 %. Trial WOU23 also had lower ELPC and ER compared to that of the other two trials. PE showed a broad range of variation in all three trials, as highlighted by a coefficient of variation (CV%) of 8.07 % in WOU22, 10.15 % in WOU23, and 8.59 % in NOP23. WOU22 had the oldest plants at time of harvest, and a slight reduction in PE was observed. Leaf dry matter content varied from 9.48 % to 16.61 %, and in trial NOP23 dry matter content was approximately 2 % lower compared to that of WOU22 and WOU23 (Fig. 1). Leaf yield (LY) was only measured in the WOU22 trial and had a mean of 22.31 g dry biomass per plant, with a CV of 19.76 % (supplementary, Table S1).

### 3.2. Heritability estimates

The broad-sense heritability was estimated for LPC, ELPC, ER, PE, and DM. These traits were analysed within-trials and across-trials. In the analysis across-trials, almost half of the phenotypic variation in LPC is due to genetic factors, indicated by a heritability estimate of 0.44

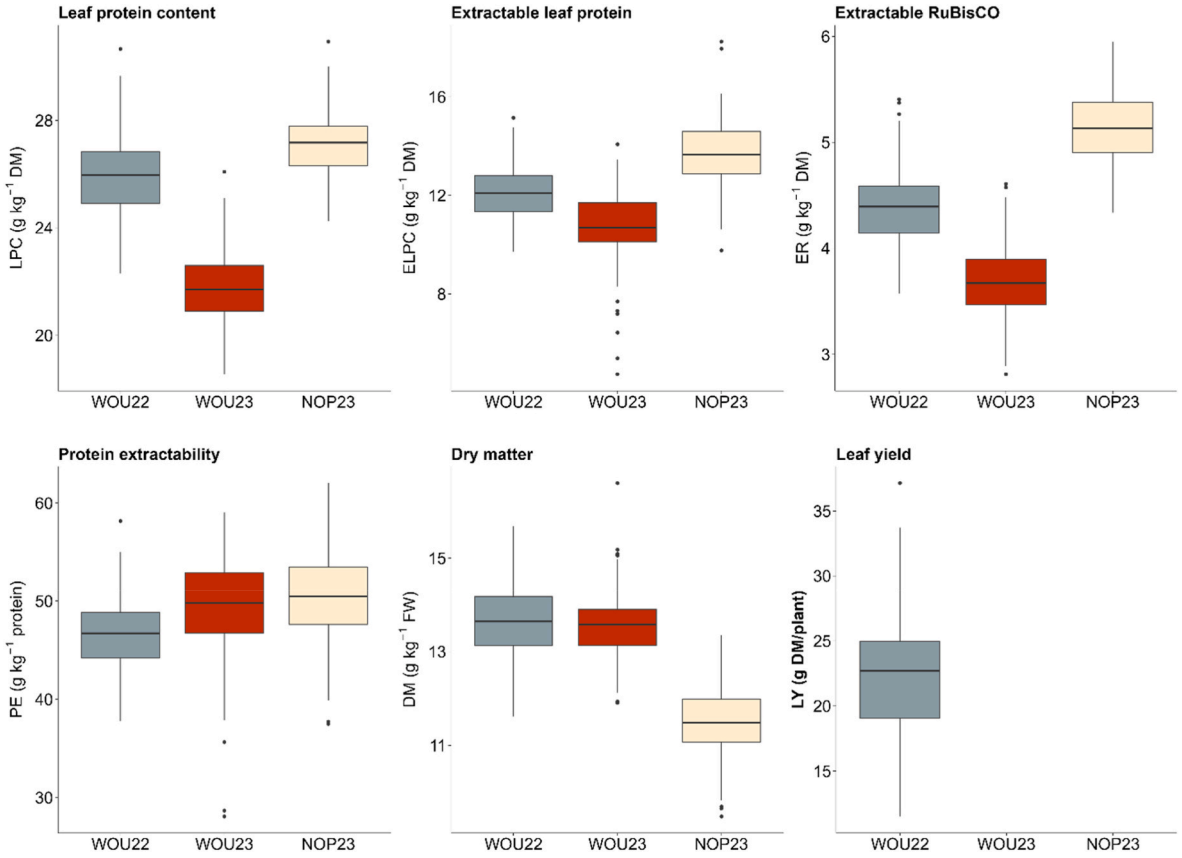



**Fig. 1.** Phenotypic variation. Boxplots showing statistical distribution of all traits (BLUEs data), across the three trials. WOU22 = grey, WOU23 = red, NOP23 = white. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 2**
Broad-sense heritability. Heritability estimates for leaf protein content, extractable leaf protein content, extractable RuBisCO, protein extractability, dry matter and leaf yield. Estimated within- and across-trials where data was available. G = Genotype, E = Environment, Block = Block, GEI = Genotype by Environment Interaction, ε = residual variance, H2 = Broad-sense heritability. Outliers were not considered during heritability calculations. Values in brackets are the confidence intervals of heritability estimates.

| Trait | Within-trial heritability | | | Across trial variance components and heritability | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | WOU22 | WOU23 | NOP23 | G | E | Block | GEI | ε | $H^2$ |
| Leaf protein (LPC) | 0.43 [0.38, 0.67] | 0.18 [0, 0.45] | 0.39 [0.30, 0.67] | 0.4 | 7.8 | 0.34 | 0.3 | 2.77 | 0.44 [0.26, 0.57] |
| Extractable protein (ELPC) | 0.13 [0, 0.33] | 0.22 [0.01, 0.52] | 0.12 [0, 0.40] | 0.1 | 1.9 | 0.49 | 0.1 | 2.36 | 0.22 [0, 0.39] |
| Extractable RuBisCO (ER) | 0.15 [0.0, 0.43] | 0.1 [0, 0.39] | 0.45 [0.38, 0.65] | 0 | 0.5 | 0.02 | 0 | 0.19 | 0.18 [0, 0.37] |
| Protein extractability (PE) | 0.04 [0, 0.24] | 0.23 [0.01, 0.52] | 0.17 [0, 0.46] | 0 | 3.3 | 2.37 | 1.2 | 31.1 | 0 [0,0] |
| Dry matter (DM) | 0.43 [0.43, 0.70] | 0.03 [0, 0.35] | 0.29 [0.11, 0.53] | 0.2 | 1.4 | 0.94 | 0 | 1.04 | 0.46 [0.29, 0.57] |
| Leaf yield (LY) | 0.25 [0.06, 0.51] | – | – | – | | | | | |

(Table 2). A large part of the sums of squares in the LPC model went to the environment (E) factor, indicating that the environment strongly affects this phenotype. DM had a similar heritability at 0.46, but ELPC and ER were lower at 0.22 and 0.20. For PE, Genotype (G) could not explain any of the variation, resulting in a heritability estimate of 0. Modelling PE resulted in a large residual variance, indicating that the other factors in the model, Environment (E), Block (B) and Genotype by environment interaction (GEI) could also not explain the variation in PE. For all traits, the effect of GEI was not found to be significant ($P > 0.05$). Estimating the broad-sense heritability within trials yielded varying results. The heritability of ER was notably higher in trial NOP23 (0.45) compared to other trials and the across-trial analysis. Some heritability was also observed for PE in trials WOU23 and NOP23 (0.23 and 0.17). Based on the heritability estimates, the across-trial BLUEs for LPC, ELPC, ER and DM were taken to the next step for GWAS analysis. Within-trial BLUEs were also used for GWAS, but only when the within-trial heritability exceeded that of the across-trial analysis or when across-trial data were unavailable. This applied to LY in WOU22 ($H^2 = 0.25$), ER in NOP23 ($H^2 = 0.45$), and PE in WOU23 and NOP23 ($H^2 = 0.23$ and 0.17, respectively).

### 3.3. Population structure and linkage disequilibrium

Population structure of the GWAS panel was investigated through Principal Component Analysis, followed by K-means clustering. Through this approach, four distinct clusters were detected within the panel (Fig. 2A). The first cluster (coloured red) corresponds to genotypes that were crossed with the second tester. The remaining clusters are linked to breeding efforts adressing different biotic stresses, with clusters 2 (green) and 3 (blue) corresponding to normal rhizomania and nematode resistant accessions, and cluster 4 (purple) to strong rhizomania and rhizoctonia resistant accessions. The first two principal components explain only 34.72 % of the variance. The first 24 principal components were statistically significant ($P < 0.05$), and these were used to correct for genetic relatedness in the analysis of linkage disequilibrium decay.

In case of no correction for genetic relatedness, LD was highly inflated even for marker pairs beyond a distance of 30 Mb (Fig. 2B). Correcting for genetic relatedness removed the high LD values at further distances, but retained a high LD between physically close markers (Fig. 2C). LD decay was determined for each individual chromosome. Accordingly, LD dropped to half of its maximum value at distances from 140 Kb to 500 Kb, depending on the chromosome (Table 3).

### 3.4. Association mapping

Association mapping analysis was performed to detect associations between markers and the investigated traits, across the whole genome. The significance of marker-trait associations was inspected through the Manhattan plots displaying the LOD cores ($-\log_{10}(P)$) (Figs. 4–5). Additionally, the Q-Q plots were considered to inspect the effectiveness of population structure correction (Supplementary Fig. S2). Significant marker-trait associations (FDR-adjusted $P < 0.05$) were detected on all chromosomes (Fig. 3). Using the across trial genotypic means for total protein (LPC), extractable protein (ELPC), extractable RuBisCO (ER) and dry matter (DM), 21 markers were identified explaining 4.42–9.24 % of the total variation in the phenotype (Fig. 4, Table 4). Another 24 markers were identified through further analysis of ER, protein extractability (PE) and leaf yield (LY) using the adjusted genotypic means from specific trials (Fig. 5, Table 4). Co-localisations of significant marker-trait associations for different traits were observed across several chromosomes. Notably, a hotspot was detected at the end of

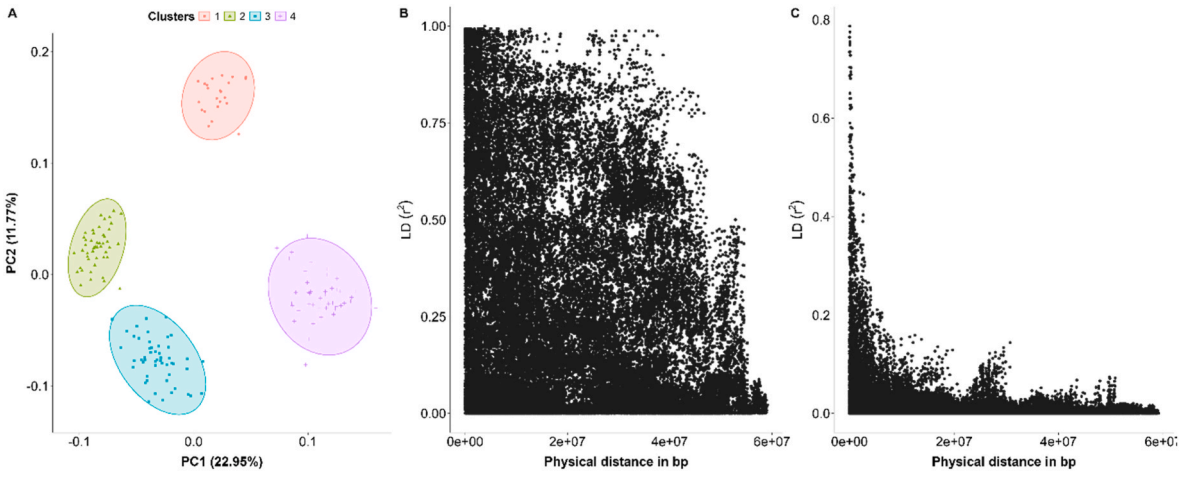



**Fig. 2.** Population structure and linkage disequilibrium. A) PCA plot of the marker data, K-means clustering resulted in four clusters. B) Pairwise LD between markers on chromosome 8 without correcting for genetic relatedness. C) Pairwise LD between markers on chromosome 8, with correction for genetic relatedness.

**Table 3**
Distances (Kb) at which LD ($r^2$) decayed to half of its maximum value, determined for each chromosome and corrected for genetic relatedness.

| Chromosome | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| LD decay distance (Kb) | 140 | 280 | 180 | 320 | 500 | 310 | 230 | 450 | 230 |

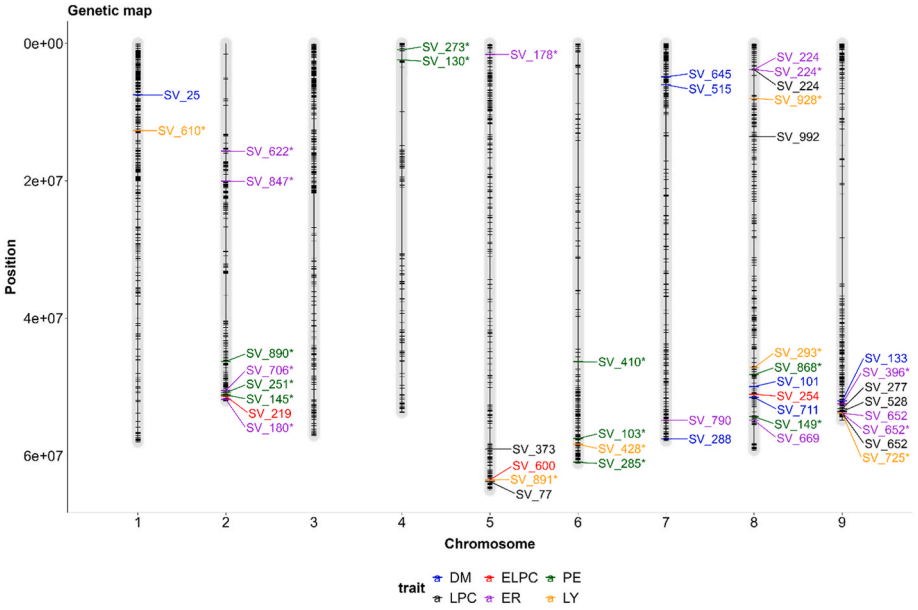



**Fig. 3.** Genetic map plot of significant marker-trait associations. The x-axis represents the chromosome number, and the y-axis shows the positions of markers. Marker positions reflecting marker density are highlighted as black markings within the depicted chromosome bars. Markers with significant associations are highlighted with annotations of marker IDs for leaf dry matter (blue), leaf protein content (black), extractable leaf protein content (red), extractable RuBisCO (purple), protein extractability (green) and leaf yield (orange). Annotations with * represent the markers that were only found using within-trial GWAS. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

chromosome 5, where markers associated with LPC, ELPC and LY were identified. Similarly, the end of chromosome 9 emerged as another hotspot, with markers associated with LPC, DM, ER and LY. A third hotspot was identified at the end of chromosome 2, where traits related to the extractable protein proportion were prevalent, as markers for ELPC, ER and PE were identified (Fig. 3).

### 3.5. Across-trial GWAS

The across-trial GWAS analysis identified 21 markers associated with leaf protein content (LPC), extractable leaf protein content (ELPC), extractable RuBisCO (ER) and dry matter content (DM) (Fig. 4). For LPC, significant marker-trait associations were identified on chromosomes 5, 8 and 9, with the strongest association peak observed at marker SV_528 on chromosome 9, explaining 8.99 % of the phenotypic variance in LPC (Table 4). For ELPC, significant associations were identified on chromosomes 2, 5 and 8, where marker *SV_600* on chromosome 5 showed the highest significance and explained 8.97 % of the phenotypic variance (8.97 %). Notably, marker *SV_224* was significantly associated with both LPC and ER, explaining 6.9–7.2 % of the phenotypic variation. For DM, significant marker-trait associations were detected on chromosomes 1, 7, 8 and 9, with marker SV_288 and SV_515 on chromosome 7 explaining 7.1–9.2 % of phenotypic variance in DM (Table 4).

### 3.6. Within-trial GWAS

Further investigation into extractable RuBisCO (ER), protein extractability (PE) and leaf yield (LY) using within-trial data resulted in an additional 24 significant marker-trait associations (Table 4; Fig. 5). Notably, two markers SV_224 and SV_652 on chromosomes 8 and 9, respectively—previously detected in across-trial GWAS for ER—showed the strongest signals for ER in trial NOP23 (LOD = 3.23–3.39), explaining 8.4–8.9 % of phenotypic variation. For PE, significant associations were also detected in NOP23, with markers SV_149 and SV_868 explaining 12.3–13.05 of phenotypic variation, suggesting their potential role in extraction efficiency. For LY, six trial-specific associations were detected in WOU22, among which SV_293 and SV_928 on chromosome 8 explained the highest phenotypic variance 12.0–13.5 %.

### 3.7. Candidate genes

All genes surrounding the significant markers were extracted from the EL10.2 genome assembly [37]. The number of genes extracted varied from 8 to 117 (Table 4), and depended on the size of the LD window per chromosome (Table 3), as well as the density of genes within the specific region. A considerable proportion of the significant markers were located within, or in close proximity to, annotated genes, consistent with the genome-wide distribution of SNPs in coding regions. These loci may therefor reflect direct or regulatory effects of candidate genes underlying the observed trait variation. For each trait, a selection of candidate genes was made that may explain the observed variation in the trait (Table 5, the full list of genes in Supplementary Table S2). Putative candidate genes for LPC included genes related to nitrate/nitrogen transporters (NRT1/PTR family), RuBisCO assembly (cpn60beta) and chlorophyll content (FRS7). FRS7 was also a candidate gene for ELPC, along with genes related to cell wall modification and protein ubiquitination. For ER, the candidate genes related to protein transport (SEC31), chloroplast protein import (TOC90 and ClpC), and cell wall modification (CGR2). Marker *SV_149*, which was strongly associated with PE is positioned inside the PAT24 gene. This gene regulates cellulose biosynthesis through its interaction with CESA3. Other putative candidate genes for PE included genes related to the cell wall (GAUT9),

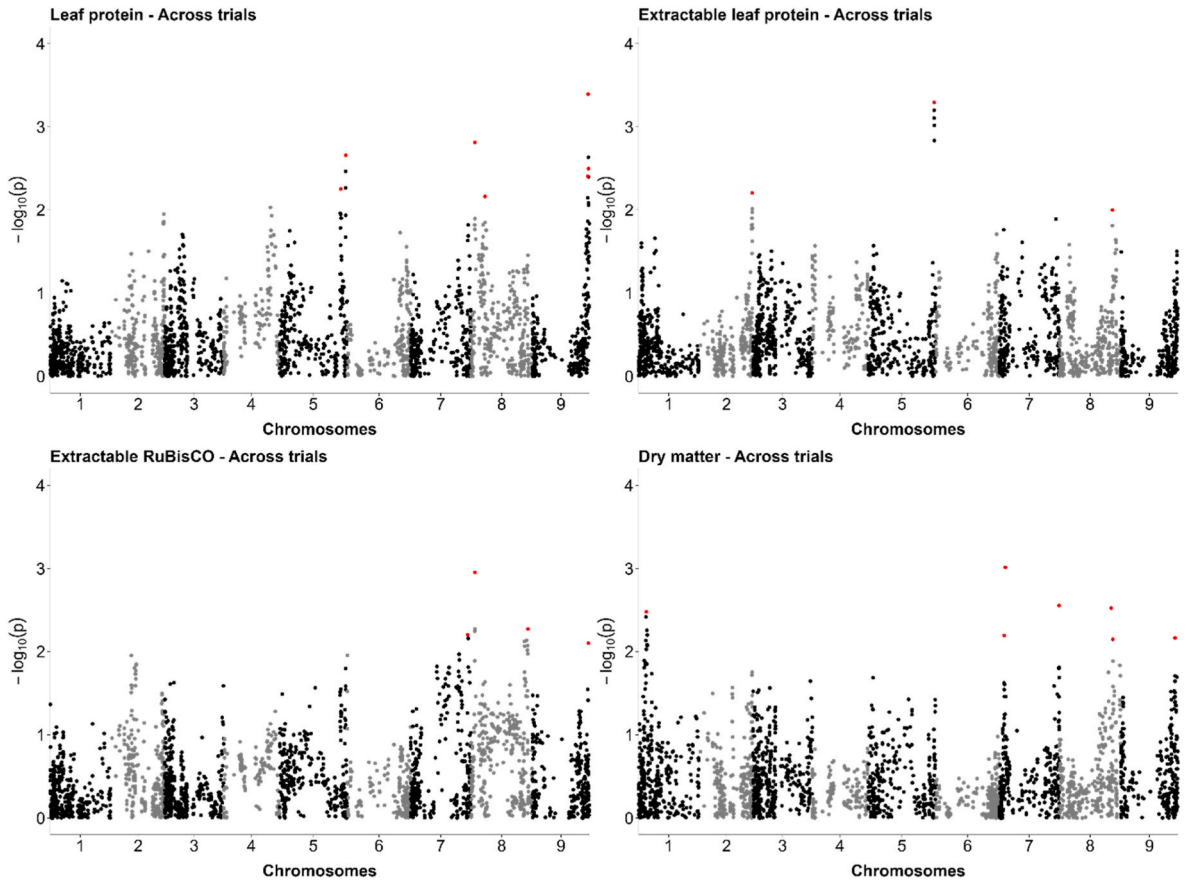



**Fig. 4.** Manhattan plots of across-trial GWAS analysis for leaf protein and related traits. The x-axis represents the chromosome position, and the y-axis shows the LOD scores (-$\log_{10}$(p)). Each dot represents a genetic marker. Significant (FDR, $P < 0.05$) marker-trait associations are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

chlorophyll degradation (PPH), and chloroplast protein import (TIC20-v, ClpC). Candidate genes for DM included FBA3 and enolase, both are known to be involved in glycolysis, an energy producing metabolic pathway. For LY, candidate genes included genes related to leaf development (AGAL) and cell wall modification (CGR2). For some of the marker trait associations listed in Table 4, no candidate genes could be proposed. A full list of all genes within the LD window of significant markers is given in the supplementary.

## 4. Discussion

Sugar beet leaves are an agricultural byproduct with potential for valorisation through leaf protein extraction. In this study we investigated the variation present in sugar beet germplasm for key traits related to leaf protein extraction. Furthermore, QTL and putative candidate genes associated with these traits were identified via GWAS.

### *4.1. Extensive phenotypic variation for leaf protein traits in sugar beet*

LPC, ELPC, ER and PE showed a wide range of variation within the set of 180 sugar beet accessions used in this study. This variation was comparable across all trials. LPC ranged from 18.56 % to 30.93 %, which is comparable to earlier reported values for leaf protein content in sugar beet leaves [5,17,38]. Protein extractability (PE) ranged from 28.08 % to 62.05 % in this study, which can be compared to previously reported values for nitrogen recovery by Kiskini and colleagues [17], where values ranging from 27.7 % to 55.6 % are reported. PE has been reported to decrease with plant age [16]. While individual trials were harvested on the same day, a similar trend can still be observed when comparing the trials for PE. WOU22 had the oldest plants at harvest and also lower values for PE, although this difference was not large and accessions with high values of PE were still observed in WOU22. Remarkably, the observed values for LPC were much lower in trial WOU23 compared to that of NOP23 and WOU22 trials. This difference may be attributed to variations in soil fertility, as WOU23 also received a lower nitrogen fertilizer dosage (data not shown). A similar pattern was observed for both ELPC and ER, suggesting that environmental conditions and management practices have a substantial impact on leaf protein yield in sugar beet.

These findings have practical implications for both breeding and industrial applications. The observed genetic diversity suggests potential to select accessions with superior protein content and extractability for developing dual-purpose sugar beet varieties. Moreover, the influence of environmental and management factors, particularly nitrogen fertilization and harvesting time, highlights the need to optimize cultivation practices to enhance protein yield and extraction efficiency without causing significant trade-offs with other economically important traits in sugar beet.

### *4.2. Heritability provided an estimate of the genetic determination of traits*

For LPC, we obtained moderate heritability (0.44) across trials. This is comparable to previous studies analysing total leaf protein and/or nitrogen content in different crops [22,39]. ELPC and ER were also found to be heritable across trials, albeit at lower levels (0.22 and 0.20, respectively). Notably, a higher heritability estimate for ER (0.45) was observed within the NOP23 trial, suggesting reduced environmental

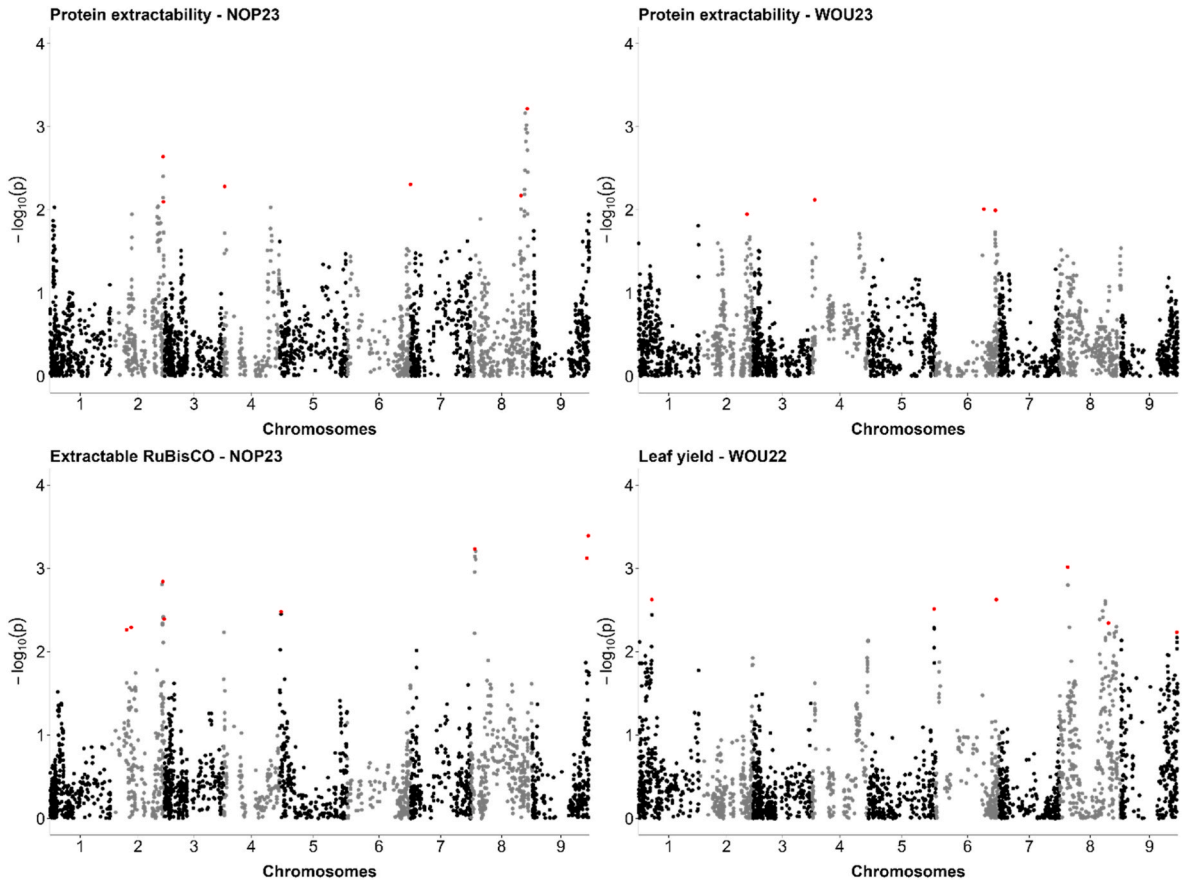



**Fig. 5.** Manhattan plots of within-trial GWAS analysis for leaf protein and related traits. Association mapping using within-trial data was performed for PE in trial NOP23 and WOU23, ER in trial NOP23, and LY in trial WOU22. The x-axis represents the chromosome position, and the y-axis shows the LOD scores ($-\log_{10}(p)$), respectively. Each dot represents a genetic marker. Significant (FDR, $P < 0.05$) marker-trait associations are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

noise or a stronger genotypic effect in this specific environment. These heritability estimates are consistent with previously reported values for soluble protein content in alfalfa, where estimates from 0.23 to 0.46 were reported [23,40–42]. The heritability estimates for PE were inconsistent, as some heritability was observed within individual trials, but this signal disappeared when data were analysed across trials. This could be attributed to varying environmental conditions, or differences in plant age at the time of leaf sampling. Another contributing factor may be the measurement of PE itself, as it is a derived trait calculated from multiple parameters, where small measurement errors can accumulate, leading to greater overall variability in PE. Enhancing the reliability of this phenotype, for example through a more robust protein extraction protocol and by standardizing field conditions, could improve the heritability of PE.

For traits with low heritability, selection is inherently less efficient, as the phenotype is more heavily influenced by environmental noise and less reflective of underlying genetic differences. In sugar beet, improving leaf protein yield may therefore be more effectively achieved by selecting for increased total leaf protein content (LPC), which shows higher heritability. In association mapping, low-heritability traits require more careful interpretation, as the underlying genetic signal may be weak. However, they can still reveal meaningful marker-trait associations provided that the genetic effects are correctly estimated with efficient modelling approaches such as used by SpATS and lme4, precisely accounting for field and environmental heterogeneity. Additionally, a large sample size of the GWAS helps to detect the subtle associations. Such associations merit further validation and may have the potential to inform future breeding strategies.

### *4.3. Association mapping identified genetic determinants of leaf protein traits*

Association mapping identified 45 markers, each associated with one or more of the following traits: total protein (LPC), extractable protein (ELPC), extractable RuBisCO (ER), protein extractability (PE), dry matter (DM), and leaf yield (LY). Based on the significant marker-trait associations, putative candidate genes, underlying the genetic markers, were identified. In the following paragraphs, we highlight and explain some of the candidate genes for specific regions where multiple marker-trait associations for different traits co-localized, as well as for markers that showed a stronger significance.

### *4.4. Co-localising marker-trait associations on chromosomes 5 and 9*

The ends of chromosomes 5 and 9 were identified as hotspots for several marker-trait associations for different traits. On chromosome 5, marker-trait associations for LPC, ELPC and LY were identified within a distance of ~250 Kb. FAR1-related sequence 7 (FRS7) was identified as a potential candidate gene in this region. FRS7 is part of the FRS-FRF family of proteins involved in several cellular processes such as chloroplast division and biogenesis, chlorophyll biosynthesis and the regulation of flowering time and growth [43–45]. FRS7 was shown to interact with FRS12 to form a complex that negatively regulates flowering time and diurnal growth in *Arabidopsis* [45]. In sugar beet, the exact function of FRS7 has not been investigated. However, in spinach, which also belongs to the *Amaranthaceae* family and is therefore related to sugar beet, FRS7 has been reported as a putative candidate gene for

**Table 4**
List of significant marker-trait associations. Significant (FDR, P < 0.05) marker-trait associations resulting from the GWAS analysis. N. of genes = the number of genes found within the LD decay window of the significant marker. Potential QTLs are highlighted as bold and italic.

| Trial | Trait | Marker | Chr | Position (bp) | Allele frequency | Effect | LOD | % Var | N. of genes |
|---|---|---|---|---|---|---|---|---|---|
| Across trials | Total protein | SV_373 | 5 | 59,032,251 | 0.31 | 0.30 | 2.25 | 5.40 | 57 |
| | | SV_77 | 5 | 63,751,827 | 0.21 | −0.34 | 2.66 | 6.68 | 117 |
| | | ***SV_224*** | ***8*** | ***3,824,790*** | ***0.14*** | ***−0.42*** | ***2.81*** | ***7.17*** | ***56*** |
| | | SV_992 | 8 | 13,585,740 | 0.12 | −0.40 | 2.16 | 6.62 | 31 |
| | | SV_277 | 9 | 53,156,660 | 0.43 | 0.29 | 2.41 | 6.75 | 36 |
| | | ***SV_528*** | ***9*** | ***53,573,872*** | ***0.45*** | ***−0.32*** | ***3.39*** | ***8.99*** | ***40*** |
| | | ***SV_652*** | ***9*** | ***53,814,634*** | ***0.07*** | ***−0.50*** | ***2.50*** | ***5.65*** | ***39*** |
| | Extractable protein | SV_219 | 2 | 51,319,788 | 0.41 | −0.23 | 2.21 | 5.46 | 44 |
| | | ***SV_600*** | ***5*** | ***63,516,249*** | ***0.20*** | ***−0.35*** | ***3.29*** | ***8.97*** | ***106*** |
| | | SV_254 | 8 | 51,073,786 | 0.06 | −0.39 | 2.00 | 5.05 | 64 |
| | Dry matter | SV_25 | 1 | 7,504,500 | 0.11 | −0.29 | 2.48 | 6.35 | 8 |
| | | SV_645 | 7 | 4,881,054 | 0.47 | −0.16 | 2.19 | 5.51 | 25 |
| | | ***SV_515*** | ***7*** | ***6,039,625*** | ***0.11*** | ***0.30*** | ***3.01*** | ***7.12*** | ***31*** |
| | | ***SV_288*** | ***7*** | ***57,554,692*** | ***0.44*** | ***−0.21*** | ***2.55*** | ***9.24*** | ***54*** |
| | | SV_101 | 8 | 49,938,584 | 0.05 | −0.43 | 2.52 | 5.64 | 48 |
| | | SV_711 | 8 | 51,507,463 | 0.11 | −0.26 | 2.15 | 5.51 | 66 |
| | | SV_133 | 9 | 52,042,099 | 0.19 | −0.23 | 2.17 | 5.90 | 49 |
| | Extractable RuBisCO | SV_790 | 7 | 54,832,015 | 0.10 | −0.10 | 2.20 | 5.02 | 29 |
| | | ***SV_224*** | ***8*** | ***3,824,790*** | ***0.14*** | ***−0.10*** | ***2.95*** | ***6.93*** | ***56*** |
| | | SV_669 | 8 | 54,848,505 | 0.08 | −0.09 | 2.27 | 5.23 | 57 |
| | | ***SV_652*** | ***9*** | ***53,814,634*** | ***0.07*** | ***−0.11*** | ***2.10*** | ***4.42*** | ***39*** |
| Within-trial NOP23 | Protein extractability | SV_251 | 2 | 50,827,199 | 0.20 | 1.77 | 2.64 | 7.20 | 47 |
| | | SV_145 | 2 | 51,147,540 | 0.48 | −1.30 | 2.10 | 5.45 | 49 |
| | | SV_273 | 4 | 946,781 | 0.10 | −1.66 | 2.28 | 5.54 | 63 |
| | | SV_285 | 6 | 60,965,074 | 0.05 | −3.33 | 2.31 | 7.77 | 31 |
| | | ***SV_868*** | ***8*** | ***48,208,794*** | ***0.29*** | ***1.84*** | ***2.17*** | ***13.00*** | ***51*** |
| | | ***SV_149*** | ***8*** | ***54,306,342*** | ***0.07*** | ***−3.04*** | ***3.22*** | ***12.32*** | ***63*** |
| | Extractable RuBisCO | SV_622 | 2 | 15,692,586 | 0.36 | 0.10 | 2.26 | 5.63 | 24 |
| | | SV_847 | 2 | 20,069,779 | 0.36 | −0.11 | 2.29 | 9.60 | 22 |
| | | SV_706 | 2 | 50,528,251 | 0.24 | 0.13 | 2.84 | 7.96 | 54 |
| | | SV_180 | 2 | 51,820,310 | 0.33 | −0.10 | 2.39 | 6.88 | 45 |
| | | SV_178 | 5 | 1,617,451 | 0.11 | 0.15 | 2.48 | 5.68 | 80 |
| | | ***SV_224*** | ***8*** | ***3,824,790*** | ***0.14*** | ***−0.15*** | ***3.23*** | ***8.91*** | ***56*** |
| | | SV_396 | 9 | 52,440,896 | 0.18 | −0.14 | 3.12 | 7.48 | 44 |
| | | ***SV_652*** | ***9*** | ***53,814,634*** | ***0.08*** | ***−0.20*** | ***3.39*** | ***8.44*** | ***39*** |
| Within-trial WOU23 | Protein extractability | SV_890 | 2 | 46,280,319 | 0.14 | −1.64 | 1.95 | 3.85 | 46 |
| | | SV_130 | 4 | 2,402,011 | 0.49 | 7.89 | 2.12 | 4.27 | 53 |
| | | SV_410 | 6 | 46,329,343 | 0.25 | −2.04 | 2.01 | 4.00 | 21 |
| | | SV_103 | 6 | 57,489,224 | 0.40 | −1.30 | 1.99 | 3.96 | 20 |
| Within trial WOU22 | Leaf yield | SV_610 | 1 | 12,695,445 | 0.23 | 1.67 | 2.62 | 7.06 | 17 |
| | | SV_891 | 5 | 63,563,850 | 0.20 | 1.66 | 2.51 | 6.84 | 108 |
| | | SV_428 | 6 | 58,345,870 | 0.09 | 2.32 | 2.62 | 7.06 | 20 |
| | | ***SV_928*** | ***8*** | ***8,019,496*** | ***0.19*** | ***2.07*** | ***3.01*** | ***12.07*** | ***55*** |
| | | ***SV_293*** | ***8*** | ***47,189,427*** | ***0.23*** | ***−2.05*** | ***2.34*** | ***13.56*** | ***52*** |
| | | SV_725 | 9 | 53,841,912 | 0.39 | −1.35 | 2.23 | 7.25 | 40 |

chlorophyll A content [46,47]. Chlorophyll content as measured by fluorescence parameters were shown to be strongly correlated with leaf nitrogen content in sugar beet [48]. This correlation increased up to 0.86 in the sugar accumulation stage, which corresponds to the leaf harvesting time in this study. We therefore speculate that FRS7 might play a role in regulating the abundance of chloroplasts in the cell, thereby affecting the total leaf nitrogen content as well as chlorophyll fluorescence. In turn, this may increase photosynthesis rates, creating an influx of carbon to be used for growth, leading to an effect on LY as well. Further functional validation of FRS7 could provide insight into its role in determining leaf protein extraction yield in sugar beet.

On chromosome 9, marker-trait associations were identified for LPC, DM, ER and LY, within a genomic distance of ~1.8 Mb. Three different candidate genes were identified within this region, namely NAT1, cpn60beta and CGR2. We propose NAT1 as a promising candidate because the significant marker *SV_528* (linked to LPC) is located within NAT1, a gene that encodes a nucleobase-ascorbate transporter protein. Another member of the same NAT gene family, NAT4, shares a high sequence similarity with *Leaf permease 1* (LPE1) from maize, where it is required for proper chloroplast development [49]. In the sugar beet genome, NAT4 is positioned approximately 644 Kb from NAT1, placing it within this QTL region but outside the linkage disequilibrium (LD) decay window defined for chromosome 9 in this study. However, as LD can vary substantially within chromosomes [35], it is possible that LD extends over a longer distance at this locus, and that NAT4 is in LD with marker *SV_528* and other markers in this QTL region. We hypothesize that, NAT1 in sugar beet may fulfil a similar role to NAT4/LPE1 in maize, influencing variation in leaf protein content likely by regulating the development of protein-rich chloroplasts. Further investigation into the local LD structure and functional characterization of NTA1 will be essential to confirm its role in regulating leaf protein content in sugar beet.

Another candidate gene cpn60beta co-localized with *SV_277* (linked to LPC) within the same region on chromosome 9. Cpn60beta encodes the Chaperonin 60 beta protein, which is involved in the assembly of the photosynthesis enzyme RuBisCO [50]. In *Arabidopsis*, cpn60beta has been reported to play a role in acclimating photosynthesis to heat stress by protecting RuBisCO from thermal denaturation [51]. Furthermore, a reduction in cpn60beta protein level has been shown to impair plastid division [52]. Therefore, cpn60beta gene may represent a key candidate influencing variation in leaf protein content, likely through its role in RuBisCO assembly and stability for photosynthesis. The third identified candidate gene within the hotspot on chromosome 9 is CGR2, a pectin methylesterase. Pectin methylesterification through CGR2 and CGR3 has been reported to promote rosette growth, influence carbon partitioning, photosynthesis and respiration efficiency [53,54]. In fact,

**Table 5**
Candidate genes. Proposed candidate genes for each trait. Markers that were located inside their proposed candidate gene are highlighted in bold.

| Trait | Marker | Chr | Candidate gene(s) | Distance from marker (bp) | Annotation |
|---|---|---|---|---|---|
| Total protein | *SV_373* | 5 | LOC104894118; LOC104894119; LOC104894112 | 344,862; 306,327; 492,180 | protein NRT1/PTR FAMILY 2.9; protein NRT1/PTR FAMILY 2.11; protein NRT1/PTR FAMILY 4.6 |
| | *SV_77* | 5 | LOC104894455 | 16,197 | protein FAR1-RELATED SEQUENCE 7 (FRS7) |
| | *SV_277* | 9 | LOC104904761 | 57,211 | RuBisCO large subunit-binding protein subunit beta, chloroplastic (cpn60beta) |
| | *SV_528* | 9 | LOC104904798 | 0 | nucleobase-ascorbate transporter 1 (NAT1) |
| Extractable protein | *SV_219* | 2 | LOC104907005 | 6039 | E3 ubiquitin-protein ligase At3g02290 |
| | *SV_600* | 5 | LOC104894455 | 251,776 | protein FAR1-RELATED SEQUENCE 7 (FRS7) |
| | *SV_254* | 8 | LOC104901863; LOC104901677 | 9733; 51,783 | Pectin esterase-like; E3 ubiquitin-protein ligase listerin |
| Extractable RuBisCO | *SV_706* | 2 | LOC104907477 | 37,110 | protein transport protein SEC31 homologue B |
| | *SV_180* | 2 | LOC104906759 | 2719 | probable aquaporin NIP-type |
| | *SV_224* | 8 | LOC104900707 | 0 | E3 ubiquitin-protein ligase (RMA1H1) |
| | *SV_669* | 8 | LOC104901941; LOC104901918 | 0; 414,967 | translocase of chloroplast 90, chloroplastic (TOC90); chaperone protein ClpC, chloroplastic |
| | *SV_652* | 9 | LOC104904816 | 7473 | probable pectin methylesterase (CGR2) |
| Protein extractability | *SV_251* | 2 | LOC104907343; LOC104907441 | 13,185; 192,482 | probable galacturonosyltransferase 9 (GAUT9); pheophytinase (PPH) |
| | *SV_273* | 4 | LOC104891860 | 81,981 | protein TIC 20-v, chloroplastic |
| | *SV_285* | 6 | LOC104897279 | 2590 | ARM REPEAT PROTEIN INTERACTING WITH ABF2 (ARIA) |
| | *SV_149* | 8 | LOC104901906; LOC104901914 | 0; 85,755 | protein S-acyltransferase 24 (PAT24); chaperone protein (ClpC), chloroplastic |
| Dry matter | *SV_645* | 7 | LOC104898399 | 51,469 | fructose-bisphosphate aldolase 3, chloroplastic |
| | *SV_515* | 7 | LOC104898495 | 8908 | Enolase |
| Leaf yield | *SV_428* | 6 | LOC104897154 | 24,121 | probable alpha-galactosidase B (AGAL) |
| | *SV_891* | 5 | LOC104894455 | 204,175 | protein FAR1-RELATED SEQUENCE 7 |
| | *SV_928* | 8 | LOC104900922; LOC104900924 | 26,702; 0 | mitochondrial pyruvate carrier 1 (MPC1); ubiquinol oxidase 4, chloroplastic/chromoplastic (AOX4) |
| | *SV_725* | 9 | LOC104904816 | 19,804 | probable pectin methylesterase (CGR2) |

overexpression of CGR2 in *Arabidopsis* resulted in highly expanded leaves, coupled with a reduced number of chloroplasts per leaf area [54]. Due to the large size of the QTL region on chromosome 9 (~1.8 Mb), it remains unclear whether the observed variation in leaf protein content is driven by a single gene or by multiple linked candidate genes. Additional research is necessary to pinpoint the causal gene(s) underlying this region and to clarify their contribution co-localising traits.

Taken together, our results suggest that larger morphological and physiological changes in the leaves, thereby affecting variation in LPC, ELPC, ER, LY and DM, may underlie the observed QTLs on chromosome 5 and 9. Genetic pathways controlling chloroplast development, RuBisCO assembly, and growth inducing cell wall modifications represent promising targets for breeding efforts aimed at increasing the protein content of the leaves. Moreover, these insights can guide the selection of genotypes and the development of molecular tools for improving protein yield and extraction efficiency in sugar beet, supporting its valorisation as a dual-purpose crop for both sugar and protein production.

### 4.5. Genomic regions associated to extractable protein-related traits

On chromosome 2, a region of approximately ~1.3 Mb hosted significant markers for variation in ELPC, ER and PE. All traits which are primarily aimed at quantifying the extractable portion of leaf protein. Here, two candidate genes can be proposed; GAUT9 and PPH. GAUT9 is a galacturonosyltransferase, which is required for the synthesis of pectin, an important constituent of the cell wall [55]. The cell wall has been reported to act as a barrier during the extraction process [7,9,56]. The variation highlighted on this locus may therefore reflect changes in the composition of the cell wall through GAUT9, influencing the cell wall's effect on protein extractability. The other candidate gene, PPH, is an enzyme involved in the breakdown of chlorophyll during leaf senescence [57]. In tomato, senescence-regulated protein degradation has been reported to reduce protein extractability. Given the high variability in the germplasm we used, different accessions may also show varying rates of senescence. PPH can be a potential candidate influencing variation in protein extraction efficiency through the regulation of leaf senescence. Additional research can provide further insight on the function of PPH and its possible role in leaf protein extraction yield in sugar beet.

At the end of chromosome 8, a relatively strong marker-trait association (*SV_149*) was identified for PE in NOP23. At approximately ~500 Kb distance, marker *SV_669* was associated with ER in the across-trials analysis. Putative candidate genes for this region include PAT24, ClpC and TOC90. Marker *SV_149* is intragenic within PAT24, which is involved in regulating cellulose biosynthesis through its interaction with CESA3 [58]. PAT24 may therefore affect the composition or characteristics of the cell wall, more specifically cellulose, and therefore protein extractability. The other two candidate genes in this region, ClpC and TOC90, are both associated with chloroplast protein import [59]. Specifically TOC90, was shown to support the accumulation of photosynthetic proteins into chloroplasts [60]. Variation tied to these genes may therefore specifically influence the import of the RuBisCO subunits for assembly in the chloroplast.

These results suggest that multiple genetic mechanisms may influence variation in extractable protein, and protein extractability in sugar beet. Genes involved in pathways related to cell wall modification, senescence regulation and the import of proteins into the chloroplast emerge as promising candidates influencing leaf protein content and extraction efficiency in sugar beet. Further functional validation is required to confirm their roles. Once validated, the identified QLs could serve as valuable targets for molecular breeding aimed at increasing protein extractability.

### 4.6. Strength and limitations of our GWAS analysis

To our knowledge, this is the first study aimed at quantifying and identifying the genetic architecture of traits related to leaf protein and protein extractability, in sugar beet as well as in any other crop. The broad sense heritability of many traits investigated in this study was

relatively low to moderate (0.2–0.3), reflecting the substantial environmental sensitivity of NIR-derived and enzymatic parameters such as RuBisCO content. A low heritability can reduce the power of GWAS. However, we minimized the influence of field heterogeneity and improved the accuracy of genetic estimates by fitting the spatial model (SpATS) and deriving genotype BLUEs. Additionally, we chose to run the GWAS analysis across trials when traits showed a reliable heritability across trials. This decision was based on the fact that a significant genotype-by-environment interaction was not present in our data. Importantly, the traits investigated were novel and biologically relevant to sugar beet, representing physiological processes tightly linked to nitrogen assimilation and photosynthetic performance. Thus, even associations explaining small proportions of phenotypic variance may provide valuable insights in the genetic control of these traits and guide future breeding efforts. GWAS analysis using the within trial BLUEs was only performed in cases where the within-trial heritability was higher than across trials. In these cases, potentially relevant marker-trait associations could still be identified that may have been overshadowed by differing environmental factors in other trials. Overall, adding the vanRaden kinship matrix to the GWAS model proved to be adequate to control false positive associations, as shown by the Q-Q plots for each analysis (Supplementary Figs. S2 and S3). For certain analyses (PE; WOU23, ELPC; Across trials), the Q-Q plots showed deflated significance values. Q-Q plot deflation can indicate that the model is overcorrecting for population structure. However, removing this component from the model and fitting a zero kinship matrix instead, did not, or minimally changed the observed deflation. Also in the analyses without deflation, we did not observe strong inflation of the significance values. This likely reflects the genetic complexity of the assessed traits, as they are likely to be built up from many different genes with small effects, which are more difficult to model by the GWAS. A larger sample size would have a better chance of picking up these small effect QTL compared to the 180 accessions used in this study. In spite of that, the GWAS analysis in our population identified a broad range of QTLs from diverse genomic locations by employing the correlation based clustering approach implemented in the FDR method for assigning significant markers.

To obtain the window size in which to search for candidate genes near significant markers, we determined the LD decay following the approach of Adetunji et al. [35]. This method corrects for genetic relatedness in calculations of LD, which was necessary as LD decay based on uncorrected data showed a very slow decay, and LD was spanning to a very long distance. This approach also provided the LD decay estimates per individual chromosome, which is advantageous as genome wide LD decay estimates are likely to be flawed due to intense selection patterns within our panel. Furthermore, obtaining the LD decay windows in this way provided shorter window sizes for certain chromosomes (down to 140 Kb), compared to a genome wide LD decay of 501 Kb provided by Wang et al. [61] in sugar beet. This is advantageous, as it reduces the amount of genes in LD with a significant marker and therefore simplifies candidate gene analysis. However, markers for different, but correlated traits were sometimes found close together, but outside the LD decay window of each other. It is possible that these markers were identified as significant because they are both in LD with the same locus affecting both traits. Which may suggest that the estimated LD decay was too short for these regions, and may have excluded interesting candidate genes in the neighbouring regions. Still, for most significant markers in this study, we proposed candidate genes for the traits based on the gene's (limited) functional annotation, and stated our hypotheses about the factors which possibly played roles in the trait's variation, e.g. chloroplast abundance, protein transport, and cell wall modification. While this is a good first approach, we emphasize the need for functional studies into the proposed candidate genes in order to validate their association with leaf protein.

## 5. Conclusion

This is the first study aimed at quantifying and dissecting the genetic component involved in leaf protein extraction in sugar beet. The total protein content of sugar beet leaves was found to be moderately heritable, providing a new avenue for increasing leaf protein yield through breeding of higher protein content in the leaves. Lower heritability was found for extractable protein, as well as protein extractability, indicating that a genetics approach targeting these specific traits may have limited success. Association mapping identified 45 marker-trait associations related to leaf protein (extractability), leaf yield or dry matter content. Based on these marker-trait associations, several candidate genes are proposed to play a role in regulating leaf protein content and extractability. The candidate genes include those involved in RuBisCO assembly (cpn60beta), cell wall modification (CGR2, PAT24, GAUT9), chloroplast development (NAT1, FRS7), and chloroplast protein import (TOC90, ClpC). However, further research is needed to validate these genes and clarify their potential role in regulating leaf protein content and extractability. As leaf protein extraction is intended to increase the overall use of the sugar beet, it is crucial that breeding for higher leaf protein content does not have a negative impact on the sugar yield extracted from the beet root. Therefore, it is important to investigate the extent of competition over resource allocation between the leaves and the beet root.

## CRediT authorship contribution statement

**Hugo Rijken:** Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Aamir Saleem:** Writing – review & editing, Supervision, Methodology, Formal analysis, Data curation. **Juan V. Renedo:** Writing – review & editing, Resources, Formal analysis. **Alain Tossens:** Resources, Formal analysis. **Marieke E. Bruins:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Luisa M. Trindade:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (OpenAI, free version) to improve the readability and clarity of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the manuscript.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hugo Rijken reports financial support was provided by Dutch Ministry of Economic affairs - TKI. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jafr.2025.102498.

## Data availability

Data will be made available on request.

## References


[1] Crop productions and plant based products: sugar. https://agriculture.ec.europa.eu/farming/crop-productions-and-plant-based-products/sugar_en, 2024.
[2] M.E. Bruins, P. Speranza, A.C.K. Sato, Green leaves, in: S. Pyett, W. Jenkins, Mierlo Bv, L.M. Trindade, D. Welch, Zanten Hv (Eds.), Our Future Proteins - a Diversity of Perspectives, VU University Press, Amsterdam, 2023, pp. 231–242.
[3] Tenorio A. Tamayo, Sugar beet leaves for functional ingredients [Doctoral thesis], Wageningen University, 2017.
[4] A.H. Martin, O. Castellani, G.A. de Jong, L. Bovetto, C. Schmitt, Comparison of the functional properties of RuBisCO protein isolate extracted from sugar beet leaves with commercial whey protein and soy protein isolates, J. Sci. Food Agric. 99 (2019) 1568–1576.
[5] T.M. Lammens, M.C.R. Franssen, E.L. Scott, J.P.M. Sanders, Availability of protein-derived amino acids as feedstock for the production of bio-based chemicals, Biomass Bioenergy 44 (2012) 168–181.
[6] Tenorio A. Tamayo, J. Gieteling, G.A.H. de Jong, R.M. Boom, A.J. van der Goot, Recovery of protein from green leaves: overview of crucial steps for utilisation, Food Chem. 203 (2016) 402–408.
[7] Y.W. Sari, U. Syafitri, J.P.M. Sanders, M.E. Bruins, How biomass composition determines protein extractability, Ind. Crops Prod. 70 (2015) 125–133.
[8] E. Goktayoglu, M.H. Oztop, S. Ozcan, Proteomics approach to differentiate protein extraction methods in sugar beet leaves, J. Agric. Food Chem. 71 (2023) 9157–9163.
[9] M. Kleuter, Y. Yu, F. Pancaldi, M. Nagtzaam, A.J. van der Goot, L.M. Trindade, Cell wall as a barrier for protein extraction from tomato leaves: a biochemical study, Plant Physiol. Biochem. 208 (2024) 108495.
[10] A. Akyuz, S. Ersus, Optimization of enzyme assisted extraction of protein from the sugar beet (beta vulgaris L.) leaves for alternative plant protein concentrate production, Food Chem. 335 (2021) 127673.
[11] F. Shahidi, C.S. Dissanayaka, Phenolic-protein interactions: insight from in-silico analyses – a review, Food Production, Processing and Nutrition 5 (2023).
[12] A.A. Anoop, P.K.S. Pillai, M. Nickerson, K.V. Ragavan, Plant leaf proteins for food applications: opportunities and challenges, Compr. Rev. Food Sci. Food Saf. 22 (2023) 473–501.
[13] J. Kroll, H.M. Rawel, Reactions of plant phenols with myoglobin: influence of chemical structure of the Phenolic Compounds, J. Food Sci. 66 (2006) 48–58.
[14] T. Ozdal, E. Capanoglu, F. Altay, A review on protein–phenolic interactions and associated changes, Food Res. Int. 51 (2013) 954–970.
[15] M. Kleuter, Y. Yu, F. Pancaldi, A.J. van der Goot, L.M. Trindade, Prone to loss: Senescence-regulated protein degradation leads to lower protein extractability in aging tomato leaves, Plant Sci. 350 (2025) 112284.
[16] J.B. Butler, An investigation into some causes of the differences of protein expressibility from leaf pulps, J. Sci. Food Agric. 33 (1982) 528–536.
[17] A. Kiskini, A. Vissers, J.P. Vincken, H. Gruppen, P.A. Wierenga, Effect of plant Age on the quantity and quality of proteins extracted from sugar beet (Beta vulgaris L.) leaves, J. Agric. Food Chem. 64 (2016) 8305–8314.
[18] H. Hirano, Varietal differences of leaf protein profiles in Mulberry, Phytochemistry 21 (1982) 1513–1518.
[19] S. Terzić, J. Atlagić, I. Maksimović, T. Zeremski, M. Zorić, V. Miklič, et al., Genetic variability for concentrations of essential elements in tubers and leaves of Jerusalem artichoke (Helianthus tuberosus L.), Sci. Hortic. 136 (2012) 135–144.
[20] M.E. Olson, R.P. Sankaran, J.W. Fahey, M.A. Grusak, D. Odee, W. Nouman, Leaf protein and mineral concentrations across the "Miracle Tree" genus moringa, PLoS One 11 (2016).
[21] E. Biazzi, N. Nazzicari, L. Pecetti, E.C. Brummer, A. Palmonari, A. Tava, et al., Genome-Wide Association mapping and genomic selection for alfalfa (Medicago sativa) forage quality traits, PLoS One 12 (2017) e0169234.
[22] J.L. Hansen, D.R. Viands, J.C. Steffens, C.J. Sniffen, Heritability and improvement of protein and nitrogen concentrations in wilted alfalfa forage, Crop Sci. 32 (1992) 879–883.
[23] Genetic results from a diallel of eight high and eight low plants for fraction I (18S) protein content, in: D.H. Heinrichs (Ed.), Rept 23rd Alfalfa Improvement Conf, 1972.
[24] K. Ishimaru, N. Kobayashi, K. Ono, M. Yano, R. Ohsugi, Are contents of Rubisco, soluble protein and nitrogen in flag leaves of rice controlled by the same genetics? J. Exp. Bot. 52 (2001) 1827–1833.
[25] S. Ravi, G. Campagna, M.C. Della Lucia, C. Broccanello, G. Bertoldo, C. Chiodi, et al., SNP alleles associated with low bolting tendency in sugar beet, Front. Plant Sci. 12 (2021) 693285.
[26] T. Wurschum, H.P. Maurer, T. Kraft, G. Janssen, C. Nilsson, J.C. Reif, Genome-wide association mapping of agronomic traits in sugar beet, Theor. Appl. Genet. 123 (2011) 1121–1131.
[27] H. Rijken, H. Oertel F., C. Hirschmann B., A. Tossens, D. Hackenberg, J. Vegas, et al., Leaf protein extractability in sugar beet: Genetic control and breeding potential for dual-purpose cultivars,, Field Crops Research 336 (2025) 110215, https://doi.org/10.1016/j.fcr.2025.110215.
[28] M.X. Rodríguez-Álvarez, M.P. Boer, A. Fred, P.H.C. Eilers, Correcting for spatial heterogeneity in plant breeding experiments with P-splines, Spatial Statistics 23 (2018) 52–71.
[29] D. Bates, M. Machler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, J. Stat. Software 67 (2015) 1–48.
[30] A. Kuznetsova, P.B. Brockhoff, R. Christensen, B. H, lmerTest package: tests in Linear mixed effects models, J. Stat. Software 82 (2017) 1–26.
[31] B.J. van Rossum, W. Kruijer, Statgengwas: Genome Wide Association Studies, 2022.
[32] R Core Team, R: a Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2025.
[33] A. Camargo (Ed.), PCAtest: statistical significance of PCA, R package version 0.02, ed, 2024. https://github.com/arleyc/PCAtest.
[34] D. Brzyski, C.B. Peterson, P. Sobczyk, E.J. Candes, M. Bogdan, C. Sabatti, Controlling the rate of GWAS false discoveries, Genetics 205 (2017) 61–75.
[35] I. Adetunji, G. Willems, H. Tschoep, A. Burkholz, S. Barnes, M. Boer, et al., Genetic diversity and linkage disequilibrium analysis in elite sugar beet breeding lines and wild beet accessions, Theor. Appl. Genet. 127 (2014) 559–571.
[36] W.G. Hill, B.S. Weir, Variances and covariances of squared linkage disequilibria in finite populations, Theor. Popul. Biol. 33 (1988) 54–78.
[37] J.M. McGrath, A. Funk, P. Galewski, S. Ou, B. Townsend, K. Davenport, et al., A contiguous de novo genome assembly of sugar beet EL10 (Beta vulgaris L.), DNA Res. 30 (2023).
[38] P. Biondo, J. Boeing, E. Barizao, N. Souza, M. Matsushita, C. Oliveira, et al., Evaluation of beetroot (*Beta vulgaris* L.) leaves during its developmental stages: a chemical composition study, Food Sci. Technol. 34 (2014) 94–101.
[39] S. Mickelson, D. See, F.D. Meyer, J.P. Garner, C.R. Foster, T.K. Blake, et al., Mapping of QTL associated with nitrogen storage and remobilization in barley (Hordeum vulgare L.) leaves, J. Exp. Bot. 54 (2003) 801–812.
[40] B.G. Gegenbach, Genetic and Related Studies of Protein Content in *Medicago sativa* L, Univ. Illinois, 1971.
[41] H. Heinrichs D, J.E. Miltimore, Variation of fraction 1 protein content in alfalfa, Can. J. Plant Sci. 50 (1970) 537–539.
[42] J.E. Miltimore, J.M. McArthur, B.P. Goplen, W. Majak, R.E. Howarth, Variability of fraction 1 protein and heritability estimates for fraction 1 protein and total phenolic constituents in Alfalfa1, Agron. J. 66 (1974) 384–386.
[43] X. Ouyang, J. Li, G. Li, B. Li, B. Chen, H. Shen, et al., Genome-wide binding site analysis of FAR-RED ELONGATED HYPOCOTYL3 reveals its novel function in Arabidopsis development, Plant Cell 23 (2011) 2514–2535.
[44] W. Tang, W. Wang, D. Chen, Q. Ji, Y. Jing, H. Wang, et al., Transposase-derived proteins FHY3/FAR1 interact with PHYTOCHROME-INTERACTING FACTOR1 to regulate chlorophyll biosynthesis by modulating HEMB1 during deetiolation in Arabidopsis, Plant Cell 24 (2012) 1984–2000.
[45] A. Ritter, S. Inigo, P. Fernandez-Calvo, K.S. Heyndrickx, S. Dhondt, H. Shi, et al., The transcriptional repressor complex FRS7-FRS12 regulates flowering time and growth in Arabidopsis, Nat. Commun. 8 (2017) 15235.
[46] J.C. Dohm, A.E. Minoche, D. Holtgrawe, S. Capella-Gutierrez, F. Zakrzewski, H. Tafer, et al., The genome of the recently domesticated crop plant sugar beet (beta vulgaris), Nature 505 (2014) 546–549.
[47] N. Ji, Z. Liu, H. She, Z. Xu, H. Zhang, Z. Fang, et al., A genome-wide Association Study reveals the genetic mechanisms of nutrient accumulation in spinach, Genes 15 (2024).
[48] C. Fei, H. Fan, X.X. Fan, G. Xu, Estimation of total nitrogen content in sugar beet leaves based on chlorophyll fluorescence parameters, Photosynthetica 58 (2020) 869–872.
[49] N.P. Schultes, T.P. Brutnell, A. Allen, S.L. Dellaporta, T. Nelson, J. Chen, *Leaf permease1* gene of maize is required for chloroplast development, Plant Cell 8 (1996) 463–475.
[50] P.V. Viitanen, M. Schmidt, J. Buchner, T. Suzuki, E. Vierling, R. Dickson, et al., Functional characterization of the higher plant chloroplast chaperonins, J. Biol. Chem. 270 (1995) 18158–18164.
[51] M.E. Salvucci, Association of Rubisco activase with chaperonin-60beta: a possible mechanism for protecting photosynthesis during heat stress, J. Exp. Bot. 59 (2008) 1923–1933.
[52] K. Suzuki, H. Nakanishi, J. Bower, D.W. Yoder, K.W. Osteryoung, S. Y. Miyagishima, Plastid chaperonin proteins Cpn60 alpha and Cpn60 beta are required for plastid division in Arabidopsis thaliana, BMC Plant Biol. 9 (2009) 38.
[53] S.J. Kim, M.A. Held, S. Zemelis, C. Wilkerson, F. Brandizzi, CGR2 and CGR3 have critical overlapping roles in pectin methylesterification and plant growth in Arabidopsis thaliana, Plant J. 82 (2015) 208–220.
[54] S.M. Weraduwage, S.J. Kim, L. Renna, C.A. F, D.S. T, F. Brandizzi, Pectin methylesterification impacts the relationship between photosynthesis and plant growth, Plant Physiol. 171 (2016) 833–848.
[55] J.D. Sterling, M.A. Atmodjo, E. Inwood S, V.S. Kumar Kolli, H.F. Quigley, M. G. Hahn, et al., Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase, Proc. Natl. Acad. Sci. USA 103 (2005) 5236–5241.
[56] K. Rommi, T.K. Hakala, U. Holopainen, E. Nordlund, K. Poutanen, R. Lantto, Effect of enzyme-aided cell wall disintegration on protein extractability from intact and dehulled rapeseed (Brassica rapa L. and Brassica napus L.) press cakes, J. Agric. Food Chem. 62 (2014) 7989–7997.
[57] S. Schelbert, S. Aubry, B. Burla, B. Agne, F. Kessler, K. Krupinska, et al., Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during leaf senescence in Arabidopsis, Plant Cell 21 (2009) 767–785.
[58] L. Zhang, M.S. Thapa Magar, Y. Wang, Y. Cheng, Tip growth defective1 interacts with cellulose synthase A3 to regulate cellulose biosynthesis in Arabidopsis, Plant Mol. Biol. 110 (2022) 1–12.

[59] S. Kovacheva, J. Bedard, R. Patel, P. Dudley, D. Twell, G. Rios, et al., In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import, Plant J. 41 (2005) 412–428.
[60] A. Hiltbrunner, K. Grunig, M. Alvarez-Huerta, S. Infanger, J. Bauer, F. Kessler, AtToc90, a new GTP-binding component of the Arabidopsis chloroplast protein import machinery, Plant Mol. Biol. 54 (2004) 427–440.
[61] S. Wang, Z. Yue, C. Yu, R. Wang, Y. Sui, Y. Hou, et al., Genome-wide association study identifies the genetic basis of key agronomic traits in 207 sugar beet accessions, Hortic. Res. 11 (2024) uhae230.