Euphytica (2026) 222:70
https://doi.org/10.1007/s10681-026-03721-w

RESEARCH

Genome-wide association study of sucrose content in a multi-environment sugar beet panel (*Beta vulgaris* L.)

Mehmet Arslan · Omar Gaoua · Özge Doğanay · Anıl Mehmet Baltacı · Özgür Özmen

Received: 18 November 2025 / Accepted: 27 March 2026 / Published online: 9 April 2026
© The Author(s) 2026

Abstract Sugar content is a significant factor shaping yield and processing quality in sugar beet (*Beta vulgaris* L.), but its genetic basis remains incompletely understood. To investigate this trait, we conducted a genome-wide association study (GWAS) using 177 genetically diverse accessions genotyped with 29,300 DArTseq-derived single-nucleotide polymorphism (SNP) markers. Sucrose concentration ranged from 8.2 to 17.8% (w/w), with a mean of approximately 14.5%. Broad-sense heritability was estimated at 0.76, indicating strong genetic control across environments. Association analyses were performed using GLM, MLM, and FarmCPU models based on genotype-specific BLUPs. No SNP surpassed the Bonferroni-adjusted genome-wide significance threshold. However, several chromosomal regions displayed consistent suggestive signals across models, particularly on chromosomes 3, 4, 7, and 9. These associations were modest in magnitude and varied in statistical support among the models. Given the moderate population size and average marker spacing (~19–20 kb) relative to the rapid linkage disequilibrium decay (~10 kb), some genomic regions may not have been fully captured, potentially reducing the power to detect small-effect loci. Overall, the findings support a quantitatively controlled genetic architecture for sucrose content characterized by multiple loci with small effects. These results highlight the potential value of genomic selection approaches over single-marker strategies for improving sucrose accumulation in sugar beet breeding programs.

Keywords *Beta vulgaris* · Genomic selection · GWAS · DArTseq markers · Sucrose content

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10681-026-03721-w.

M. Arslan (✉)
Faculty of Agriculture, Department of Agricultural Biotechnology, Erciyes University, 38280 Kayseri, Türkiye
e-mail: mehmetarslan@erciyes.edu.tr

O. Gaoua · Ö. Doğanay
Graduate School of Natural and Applied Sciences, Department of Agricultural Biotechnology, Erciyes University, 38280 Kayseri, Türkiye

A. M. Baltacı · Ö. Özmen
Kayseri Şeker A. Ş., Şeker Mahallesi, Osman Kavuncu Bulvarı No: 314., Kocasinan, Kayseri, Türkiye

Introduction

Sugar beet (*Beta vulgaris* L.) is the second most important source of table sugar worldwide after sugarcane (Dohm et al. 2014), contributing approximately 25–30% of global sugar production, with annual outputs exceeding 34 million tons (International Sugar Organization 2022). Since its domestication and systematic improvement began in the late eighteenth century, breeding efforts have increased the root sucrose concentration from approximately

4–6% to more than 18% in modern cultivars (De Lucchi et al. 2021; Sandell et al. 2022). Despite this substantial progress, further improvement of sugar content remains a primary objective in breeding programs, as even modest increases can significantly increase sugar yield per unit area.

The sucrose concentration in sugar beet is considered a quantitative trait influenced by both genetic and environmental factors. Numerous quantitative trait loci (QTLs) associated with sugar content have been reported across all nine chromosomes (Weber et al. 1999; Würschum et al. 2011). Early biparental mapping studies identified genomic regions on chromosomes 2, 3, and 4, whereas subsequent association mapping efforts detected additional loci on chromosomes 3, 4, and 5 (Würschum et al. 2011). However, reported QTL positions and effect magnitudes vary among populations and environments, and many detected loci account for only a limited fraction of the total phenotypic variance. These inconsistencies suggest that the genetic determinants of sucrose accumulation may depend strongly on germplasm composition, the environmental context, and marker resolution.

Although several studies have provided valuable insights, the genetic architecture of the sucrose content in cultivated sugar beet remains incompletely understood (Wang et al. 2023). Different germplasm panels frequently yield partially overlapping sets of associated loci, and some chromosomal regions have been less consistently represented in previous analyses (Weber et al. 1999; Wang et al. 2023). In addition, sucrose concentration is closely linked to other agronomic traits, particularly root yield, with negative correlations often observed between these components (Wang et al. 2023). This relationship complicates the simultaneous improvement of both traits and underscores the need for a clearer understanding of the genomic factors contributing to sugar accumulation.

The development of high-throughput genotyping platforms has substantially expanded the scope of genome-wide association studies (GWASs) in crops. Genotyping-by-sequencing (GBS) approaches enable the discovery of thousands of single-nucleotide polymorphisms (SNPs) across diverse germplasm collections (Liu et al. 2023), providing broader genome coverage than traditional linkage mapping populations do (Cuevas et al. 2017). DArTseq technology offers a cost-effective method for generating genome-wide SNP datasets in sugar beet and has been successfully applied in multiple crops for dissecting complex yield and quality traits, including maize (Xiao et al. 2017), rice (Wang et al. 2020), tomato (Lin et al. 2014), sorghum (Niu et al. 2020), cotton (Fang et al. 2017), and beans (Wu et al. 2020).

Large-scale sugar beet studies further illustrate both the potential and limitations of GWASs. For example, Liu et al. (2023) analyzed 977 accessions via high-density SNP data and detected numerous marker–trait associations. Wang et al. (2024) integrated GWAS with transcriptomic analyses to investigate genomic regions associated with biomass and sugar-related traits. These studies highlight that the number and strength of detected associations depend strongly on panel size, marker density, and statistical modeling, emphasizing the importance of adequate genomic coverage and robust phenotypic integration across environments.

Unlike certain crops in which a single major gene dramatically alters sugar accumulation (e.g., the *su1* mutation in sweet corn; Creech 1965; Tsai and Nelson 1966; Boyer and Shannon 1983; Rahman et al. 1998; Tracy 1996), sugar beet has not been shown to possess a universally segregating major-effect locus controlling sucrose concentration across cultivated materials. Instead, current evidence suggests that sucrose accumulation reflects the combined influence of multiple genomic regions, the detection of which depends on the underlying linkage disequilibrium structure and the extent of genome coverage achieved by the marker set. Accordingly, understanding LD patterns and marker distributions is critical for interpreting association results.

In the present study, we performed a genome-wide association analysis of sucrose content using a genetically diverse panel of cultivated sugar beet accessions, including modern cultivars, breeding lines, and USDA-GRIN materials evaluated across multiple field environments. Phenotypic data were modeled using mixed-effect approaches to account for environmental and genotype × environment variation prior to association testing. Our objectives were to (1) quantify phenotypic variation and heritability across environments, (2) evaluate genome-wide marker–trait associations under Bonferroni-adjusted significance thresholds, and (3) assess how marker density relative to linkage disequilibrium patterns influences the

detection of association signals in this panel. Clarifying these factors will contribute to improved experimental design and support breeding strategies based on genome-wide prediction or targeted marker deployment.

Materials and methods

Plant materials and experimental design

A panel of sugar beet genotypes was assembled to capture the genetic diversity of sugar content. The association panel consisted of 177 sugar beet accessions, including breeding lines, modern hybrids, and diverse open-pollinated cultivars (all belonging to *Beta vulgaris* subsp. *vulgaris*). These genotypes were obtained from a commercial breeding program and the USDA-ARS Western Regional Plant Introduction Station in Pullman, Washington. Detailed information on the accession names, origins, and improvement statuses is provided in Table S1.

The sugar beet genotypes were cultivated over three consecutive years (2020, 2021, and 2022) in Kayseri–Kocasinan (38°45′19.9″N, 35°24′53.0″E; 1055 m asl), Yozgat–Boğazlıyan (39°08′58.6″N, 35°12′32.6″E; 1098 m asl), and Sivas–Şarkışla (39°22′13.6″N, 36°24′14.9″E; 1180 m asl), Turkey. The seeds were sown in a randomized complete block design with three replicates. Each genotype was grown in two 5-m row plots, with a spacing of 20 cm within rows and 45 cm between rows.

The plants were cultivated under standard agronomic practices for sugar beet. The plots were thinned to the optimal stand, and routine fertilizer and pest control measures were applied as needed. Fifteen sugar beet roots per plot were randomly selected and sent to the Kayseri Sugar Company for determination of sugar content (%) via a Betalyzer (Automated System for Quality Analysis of Sugar Beet, Anton Paar OptoTec GmbH, Seelze, Germany).

Phenotypic data analysis and BLUP estimation

The experiment consisted of three locations evaluated over three years, resulting in nine environments. Sugar content (%) was analyzed using a linear mixed-effects model implemented in R (lme4 package). Phenotypic values used for GWAS were estimated as best linear unbiased predictions (BLUPs) across environments, with environmental effects accounted for in the mixed-model framework. The following model was fitted:

$$y_{ij} = \mu + E_j + G_i + (GE)_{ij} + \varepsilon_{ij}$$

where y_{ij} represents the observed sugar content of genotype i in environment j; μ represents the overall mean; E_j represents the fixed effect of the environment; G_i represents the random effect of the genotype; $(GE)_{ij}$ represents the random genotype × environment interaction; and ε_{ij} represents the residual error. Random effects and residuals were assumed to be independently and normally distributed.

Genotype-specific BLUPs extracted from this model ranged from negative to positive values and followed an approximately normal distribution (see Fig. 1 and Table S1). Negative BLUP values indicate genotypes performing below the overall mean, whereas positive values indicate superior genotypes (see Table S1).

Broad-sense heritability (H^2) across environments was estimated from variance components as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GE}^2}{n} + \frac{\sigma_e^2}{nr}}$$

where σ_G^2 denotes genotypic variance, σ_{GE}^2 represents genotype × environment interaction variance, σ_e^2 represents residual variance, n represents the number of environments, and r represents the number of replications per environment.

The model was fitted via the `lmer` function in the lme4 R package (Bates et al. 2015), and genotype-specific BLUPs were subsequently used as phenotypic inputs for GWAS.

DNA extraction and DArTseq genotyping

Genomic DNA was extracted from the young leaf tissue of each accession via a CTAB-based method (Doyle and Doyle 1987). DNA quality was evaluated on agarose gels and quantified via a fluorometric assay. Approximately 200 ng of high-quality DNA per sample was sent to Diversity Arrays Technology (DArT Pty Ltd., Australia) for DArTseq genotyping (Sansaloni et al. 2011; Gruber et al. 2019). DArTseq is a genotyping-by-sequencing approach that

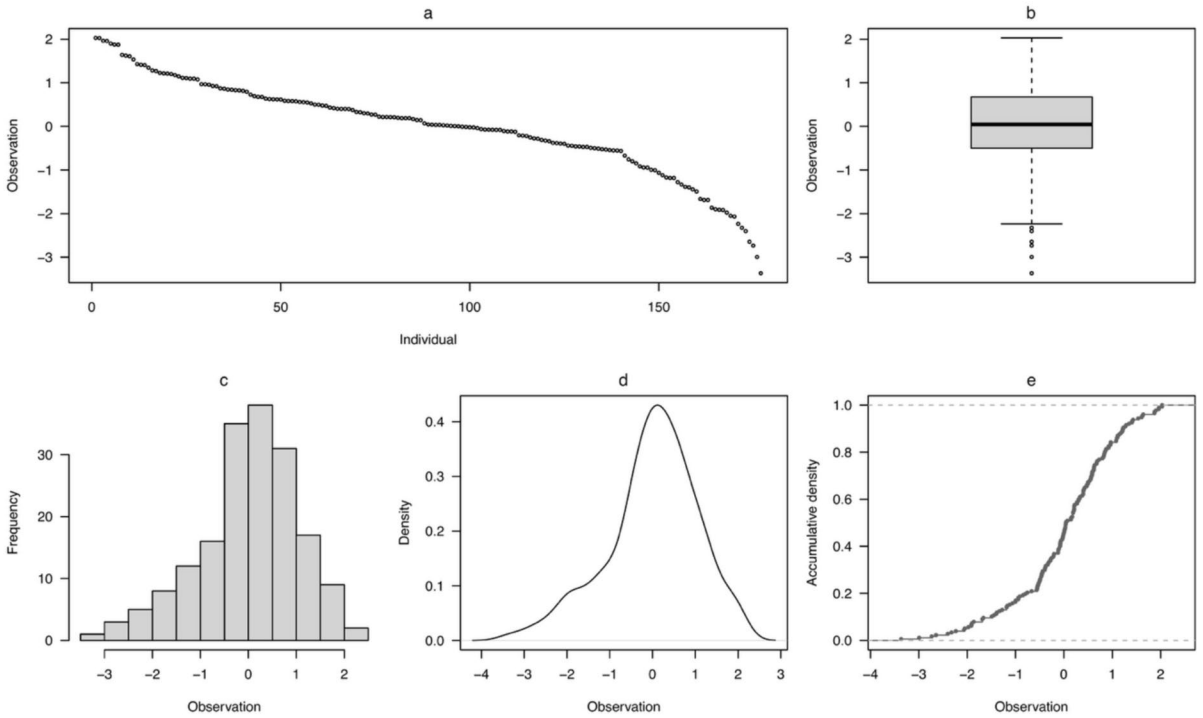


Fig. 1 **a**–**e** Distribution of genotype-specific BLUP values for sugar content across environments in 177 sugar beet genotypes used for GWAS. **a** Ranked BLUP values across genotypes; **b** boxplot summarizing variation and outliers; **c** histogram of BLUP frequency distribution; **d** density plot of BLUP values; and **e** cumulative distribution of BLUP values

combines genome complexity reduction with next-generation sequencing to discover and genotype SNP markers. Briefly, DNA samples were digested with a proprietary restriction enzyme combination optimized for sugar beet, barcoded adapters were ligated, and pooled libraries were sequenced on an Illumina platform. Sequence reads were processed via the DArT proprietary pipeline to call SNPs, and marker positions were assigned against the *Beta vulgaris* EL10.2 sugar beet reference genome (McGrath et al. 2023).

Raw SNPs were filtered for quality in R. Markers with > 20% missing data and minor allele frequency (MAF) < 0.05 were removed. After filtering, 29,300 high-quality SNPs remained across the nine sugar beet chromosomes. SNPs deviating from Hardy–Weinberg equilibrium were not excluded given the breeding history and structure of the panel. Missing genotypes were imputed via the mean allele frequency method implemented in GAPIT where needed.

Population structure and linkage disequilibrium

The population structure within the association panel was evaluated via principal component analysis (PCA) on the basis of genome-wide SNP data. PCA was performed in R via the *prcomp* function on the filtered marker matrix. The first three principal components were extracted and visualized via two- and three-dimensional plots. A scree plot was generated to assess the proportion of genetic variance explained by each component.

Pairwise relatedness among accessions was estimated by calculating an identity-by-state (IBS) kinship matrix via the EMMA algorithm implemented in GAPIT. The resulting kinship coefficients were visualized as a heatmap to examine patterns of genetic relatedness across the panel. Genome-wide linkage disequilibrium (LD) was assessed by calculating the squared allele−frequency correlation coefficient (r^2) between pairwise SNP markers as a function of physical distance.

GWAS models and statistical analysis

Genome-wide association analyses were conducted to identify marker–trait associations for sucrose content (%), represented by genotype-specific best linear unbiased predictions (BLUPs) estimated across environments. The use of BLUPs allowed adjustment for environmental variation and genotype-by-environment effects prior to association testing.

Three statistical models were implemented within GAPIT: a general linear model (GLM), a mixed linear model (MLM), and the multilocus approach Fixed and Random Model Circulating Probability Unification (FarmCPU). In the GLM analysis, population structure was accounted for by including the first two principal components derived from genome-wide SNP data as fixed covariates. The MLM incorporates both principal components and a kinship (K) matrix to correct for relatedness among individuals. FarmCPU was executed via an iterative framework that alternates between fixed and random effect modeling, with principal components and kinship integrated as implemented in GAPIT.

Marker–trait associations were tested via BLUP values as continuous phenotypic responses. Genome-wide significance was evaluated via a Bonferroni-adjusted threshold ($\alpha = 0.05$) on the basis of the number of markers analyzed. In addition, a suggestive threshold ($P < 1 \times 10^{-4}$) was used to identify loci showing moderate statistical support.

Quantile–quantile (QQ) plots were generated to examine the distribution of observed versus expected *P*–values, and Manhattan plots were constructed to visualize marker associations across chromosomes.

Software and statistical analysis

All the statistical analyses were conducted in R (version 4.4.1; R Core Team, 2024). Genome-wide association analyses were performed using the GAPIT package (v3.5.0). Data processing and visualization were carried out using standard R packages, including *dplyr*, *ggplot2*, and *qqman*.

Results

Phenotypic variation in sugar content

Considerable variation in sugar content was observed among the 177 sugar beet genotypes evaluated across three locations over three years (2020, 2021 and 2022). Mean sucrose concentration ranged from 8.2 to 17.8%, with an overall mean of approximately 14.5%.

Analysis of variance (Table S3) revealed that year and year × location interaction had highly significant effects on sugar content ($P < 0.001$), indicating strong environmental influence. Replication effects were not significant. Therefore, phenotypic data were not averaged across environments. Instead, genotype-specific best linear unbiased predictions (BLUPs) were estimated using a mixed model approach to account for environmental variation and used for subsequent GWAS analysis.

The genotype-specific BLUP values showed an approximately normal distribution (Fig. 1). The ranked plot (Fig. 1a) illustrates continuous variation among genotypes, while the boxplot (Fig. 1b) summarizes the central tendency, dispersion, and outliers. The histogram and density plot (Fig. 1c, d) further support a near bell-shaped distribution with slight deviation in the lower tail. The cumulative distribution plot (Fig. 1e) shows a smooth and continuous increase across the panel, indicating quantitative variation rather than distinct phenotypic classes.

Most genotypes (approximately 70%) exhibited sugar content between 12 and 16%, whereas a smaller subset of accessions represented the upper (~ 17–18%) and lower (~ 8–11%) extremes of the distribution. No clear clustering of phenotypic classes was observed, supporting the continuous nature of variation for this trait.

Broad-sense heritability (H^2) across environments was estimated at 0.76, indicating that a large proportion of the observed phenotypic variance was attributable to genetic differences among genotypes. This relatively high heritability suggests strong genetic control of sugar content under the evaluated field conditions, although environmental variation and genotype × environment interaction effects still contributed to total variance.

Overall, the extent of phenotypic variation and the heritability estimate confirm that the panel provides

adequate genetic diversity and statistical power for genome-wide association analysis.

Genomic diversity and population structure

Genome-wide population structure was evaluated via principal component analysis (PCA) on the basis of the SNP dataset. The first three principal components explained 4.47%, 2.06%, and 1.56% of the total genetic variance, respectively (Fig. 2). The 3D PCA scatter showed a diffuse cloud of points without discrete, well-separated clusters, indicating that the panel did not split into strong genetic subgroups. The scree plot displayed a steep drop in explained variance after PC1, with subsequent components each contributing only a small fraction of the variance, which is consistent with limited stratification.

Pairwise relatedness was assessed via the kinship matrix estimated in GAPIT. The kinship heatmap (Fig. 3) was dominated by low off-diagonal values, suggesting that most accessions are only weakly related, whereas a small number of localized blocks with higher kinship likely reflect a shared breeding background among subsets of materials. Overall, the combination of PCA and kinship patterns indicates that major confounding due to strong population subdivision is unlikely, but mild relatedness exists and should be accounted for in association testing. Therefore, PC covariates were used in GLM-based analyses, and kinship correction was applied in mixed-model approaches.

SNP quality metrics and genome-wide LD patterns

After quality filtering, 29,300 SNP markers were retained for downstream analyses. The marker-level summary metrics produced by GAPIT are shown in Fig. 4. The minor allele frequency (MAF) values range from approximately 0.05 to 0.50, indicating that the retained markers include both low- and moderate-frequency variants suitable for association testing. Marker heterozygosity values varied broadly (approximately 0.10 to 0.80), with a noticeable concentration at intermediate levels, which is consistent with the outcrossing biology of sugar beet and the heterogeneous breeding background of the panel. The marker-wise LD metric (r^2) was generally low across the genome, with occasional higher values, as expected when strong LD occurred only between a subset of closely linked markers.

Genome-wide LD decay was examined by plotting r^2 against physical distance (Fig. 5). LD decreased rapidly with increasing distance, and the smoothed trend approached background levels at ~ 10 kb. This rapid decay supports the expectation of short haplotype blocks in sugar beet. However, given a genome

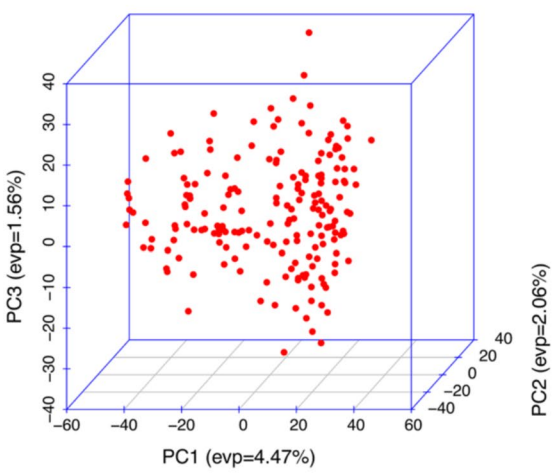


Fig. 2 Principal component analysis (PCA) of sugar beet genotypes on the basis of SNP data. (Left) Three-dimensional PCA plot showing the distribution of genotypes along the first three principal components (PC1 = 4.47%, PC2 = 2.06%, PC3 = 1.56%). (Right) Scree plot illustrating the variance explained by the top ten principal components, with the curve flattening after the first few PCs

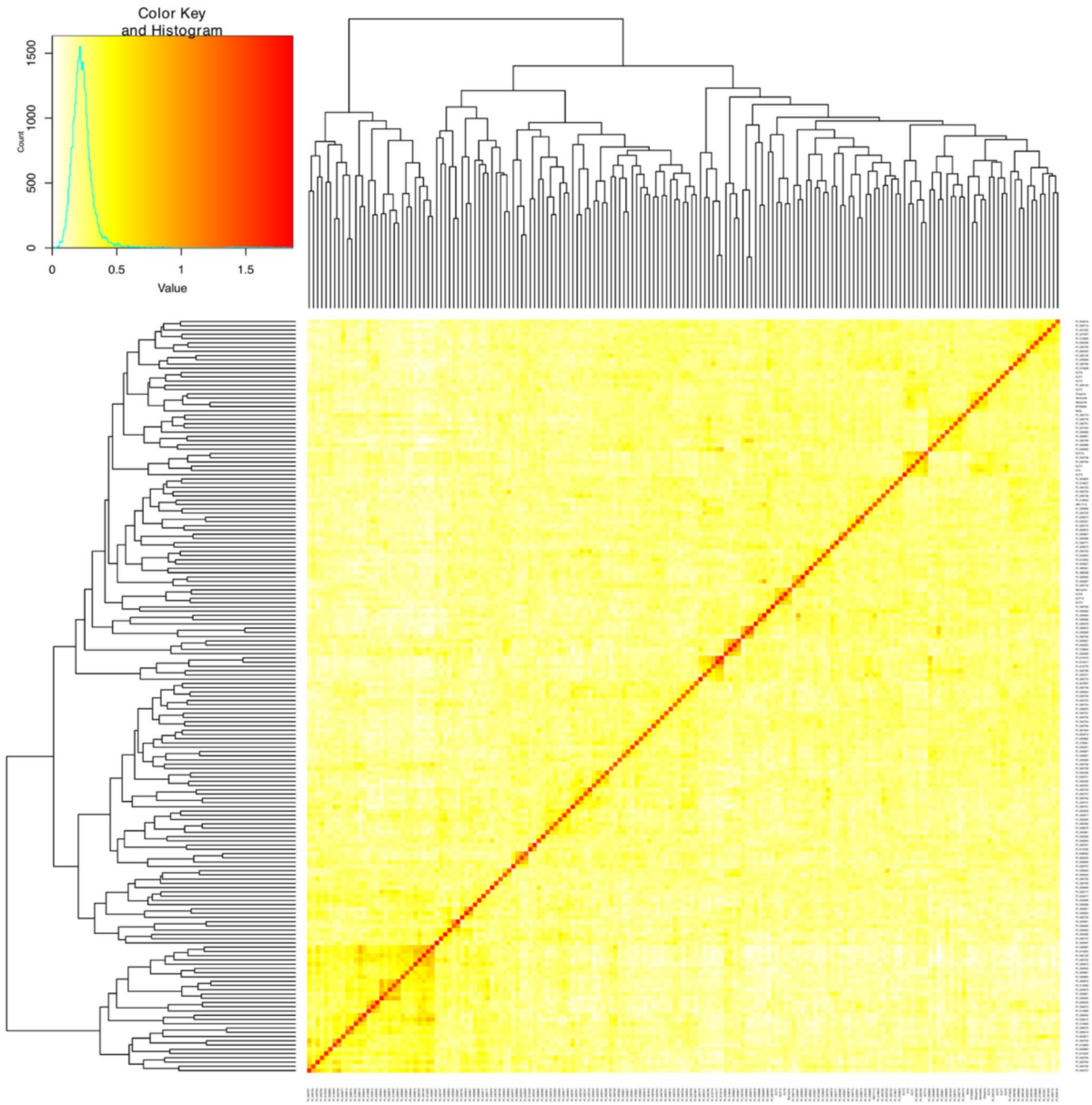


Fig. 3 Kinship heatmap of the sugar beet panel. The matrix illustrates pairwise genetic relatedness among genotypes, with hierarchical clustering highlighting groups of individuals with higher similarity

size of ~ 560 Mb and ~ 29 k SNPs, the average marker spacing is on the order of ~ 19–20 kb, which is greater than the LD decay distance. Consequently, although the SNP set provides broad genome-wide information, some genomic intervals may be insufficiently tagged, which can reduce the power to detect loci with small effects and may contribute to the absence of Bonferroni-significant associations.

Association mapping for sugar content

Genome-wide association analyses were performed via GLM, MLM, and FarmCPU models with genotype-specific BLUP values as phenotypic inputs. Across all the models, no SNP exceeded the Bonferroni-adjusted genome-wide significance threshold (Fig. 6). The quantile–quantile plots showed adequate

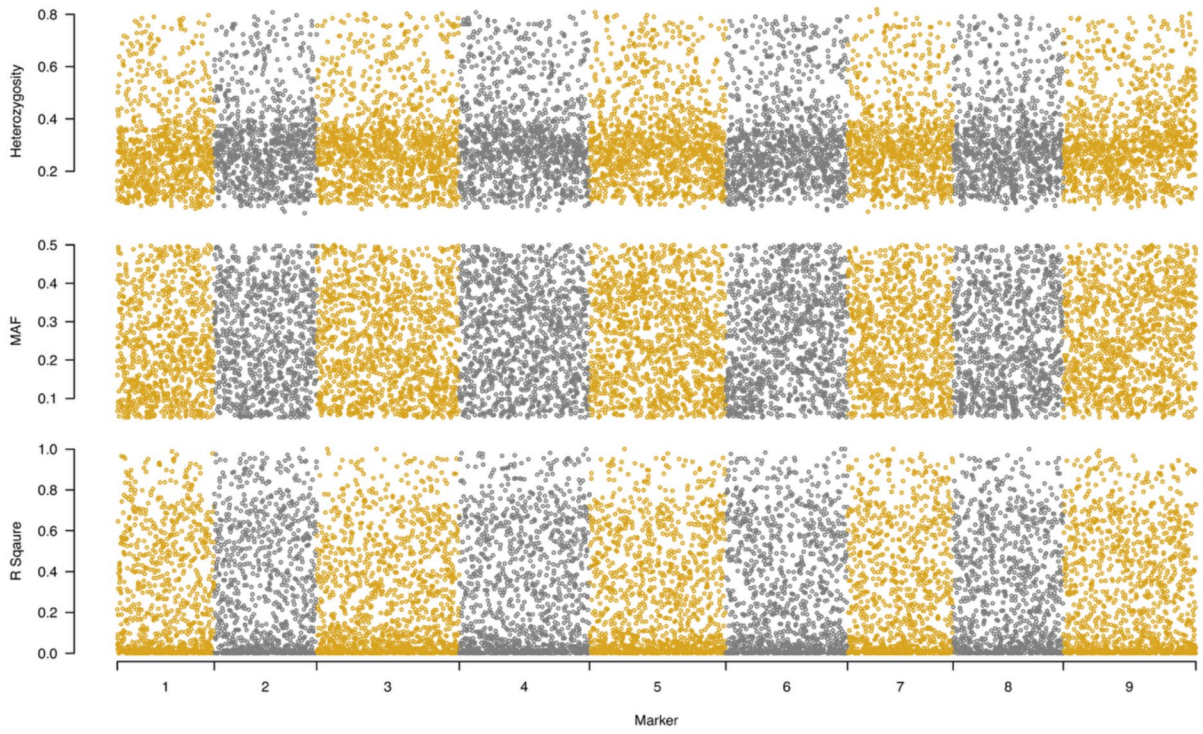


Fig. 4 Genome-wide distribution of marker properties across the nine sugar beet chromosomes. From top to bottom: Marker heterozygosity, minor allele frequency (MAF), and linkage disequilibrium (R^2). Markers are plotted according to chromosomal position, with alternating colors indicating different chromosomes

control of population structure and relatedness in the MLM and FarmCPU models, whereas the GLM exhibited mild inflation.

Although no marker reached genome-wide significance, several loci presented elevated $-\log10(P)$ values across the models (see Table 1 and Table S2). The most consistent signals were located on chromosomes 4, 9, 3, and 7. The strongest association was detected on chromosome 4 at 55.3 Mb, with *P*–values of 8.37×10^{-5} (GLM), 1.09×10^{-4} (MLM), and 1.17×10^{-4} (FarmCPU). This SNP had a minor allele frequency (MAF) of 0.35 and an estimated additive effect of -0.67.

A second locus on chromosome 9 at 19.9 Mb showed comparable statistical support across the models, although with a lower MAF (0.15). Additional loci on chromosome 3 (6.5–7.1 Mb) and chromosome 7 (29.8 Mb) presented moderate association signals but did not meet the Bonferroni threshold. Across loci, estimated allelic effects ranged from -0.42 to -0.67, indicating modest contributions to phenotypic variation. The allelic dosage effects for representative SNPs are illustrated in Supplementary Fig S1, which shows the distribution of BLUP-adjusted sucrose content values across allele classes.

Taken together, these results indicate that no major-effect locus was detected in this panel. The observed associations are consistent with a complex genetic architecture involving multiple loci with relatively small effects.

Detailed statistics for all the suggestive loci are presented in Supplementary Table S2.

Discussion

Phenotypic variation and genetic signals in a multi-environment context

The sugar beet panel exhibited broad and continuous variation in sugar content (8.2–17.8%), with an approximately normal distribution and an overall mean near 14.5%. Such a pattern is typical of complex quality traits shaped by numerous underlying

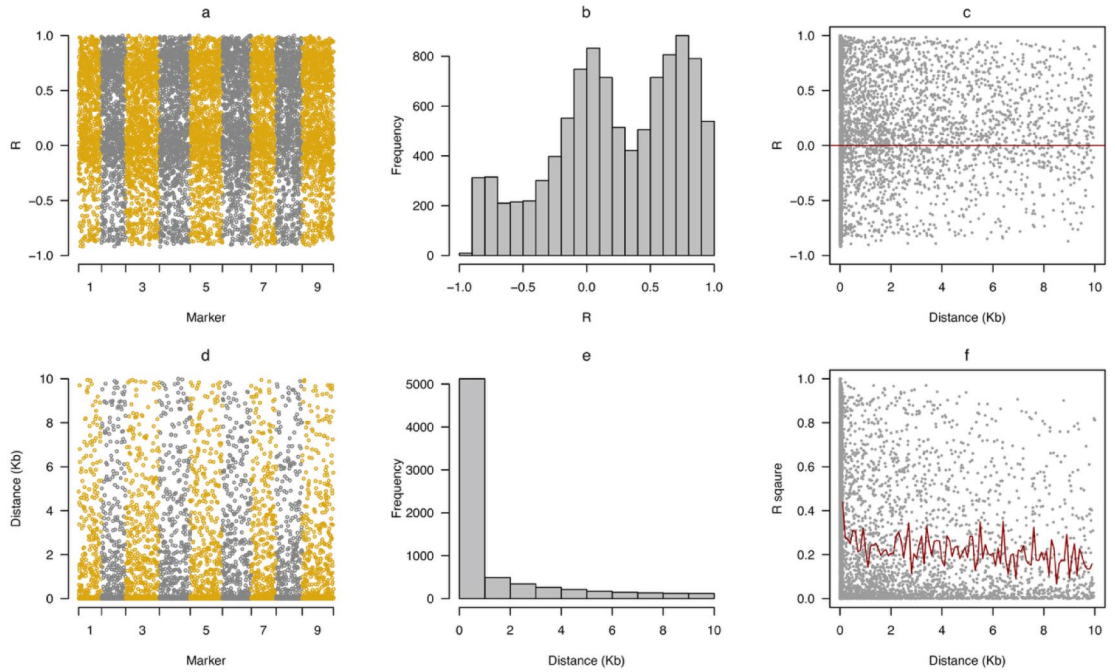


Fig. 5 Genome-wide linkage disequilibrium (LD) in the sugar beet association panel. **a** Distribution of pairwise correlation coefficients (r) across chromosomes; **b** frequency distribution of r; **c** relationship between r and physical distance (kb); **d** inter-marker distances across chromosomes; **e** distribution of inter-marker distances; and (f) LD decay (r^2)

genetic factors plus environmental modulation, rather than by discrete Mendelian classes. Multiple location studies in sugar beet have long reported that sugar concentrations and related quality traits vary continuously and are influenced by the environment while still possessing measurable genetic variance, supporting the suitability of diverse breeding panels for quantitative genetic analysis (Schneider et al. 2002).

Broad-sense heritability across the environments was high ($H^2 = 0.76$), indicating that a substantial proportion of the observed phenotypic variance was attributable to genetic differences among genotypes under the tested field conditions. In practical breeding terms, this magnitude of heritability is consistent with strong selection potential, provided that phenotyping is replicated and the environment is properly modeled. Best linear unbiased prediction (BLUP) and related mixed-model estimators are widely recommended for multi-environment trials because they reduce environmental noise and provide stable estimates of genotypic value when trials are unbalanced or subject to heterogeneous error variance (Piepho et al. 2008).

Although sugar content had high heritability, the data also indicate that environment and genotype × environment components contributed meaningfully to total variance. This is consistent with broader evidence from beet quality phenotypes showing that genotype is often a major driver but interaction with environment can be non-trivial, reinforcing the value of mixed-model phenotypic integration across years and locations rather than relying solely on arithmetic means (Hanson and Goldman 2019).

Population structure and relatedness in breeding germplasm

PCA and kinship analyses suggest that the panel behaves as a largely unstructured breeding population. The first three PCs explained only 4.47%, 2.06%, and 1.56% of the total marker variance and did not reveal discrete clusters, while the kinship matrix was dominated by low off-diagonal values with only small blocks of moderate relatedness. Such patterns are compatible with breeding germplasm shaped by historical recombination and germplasm exchange,

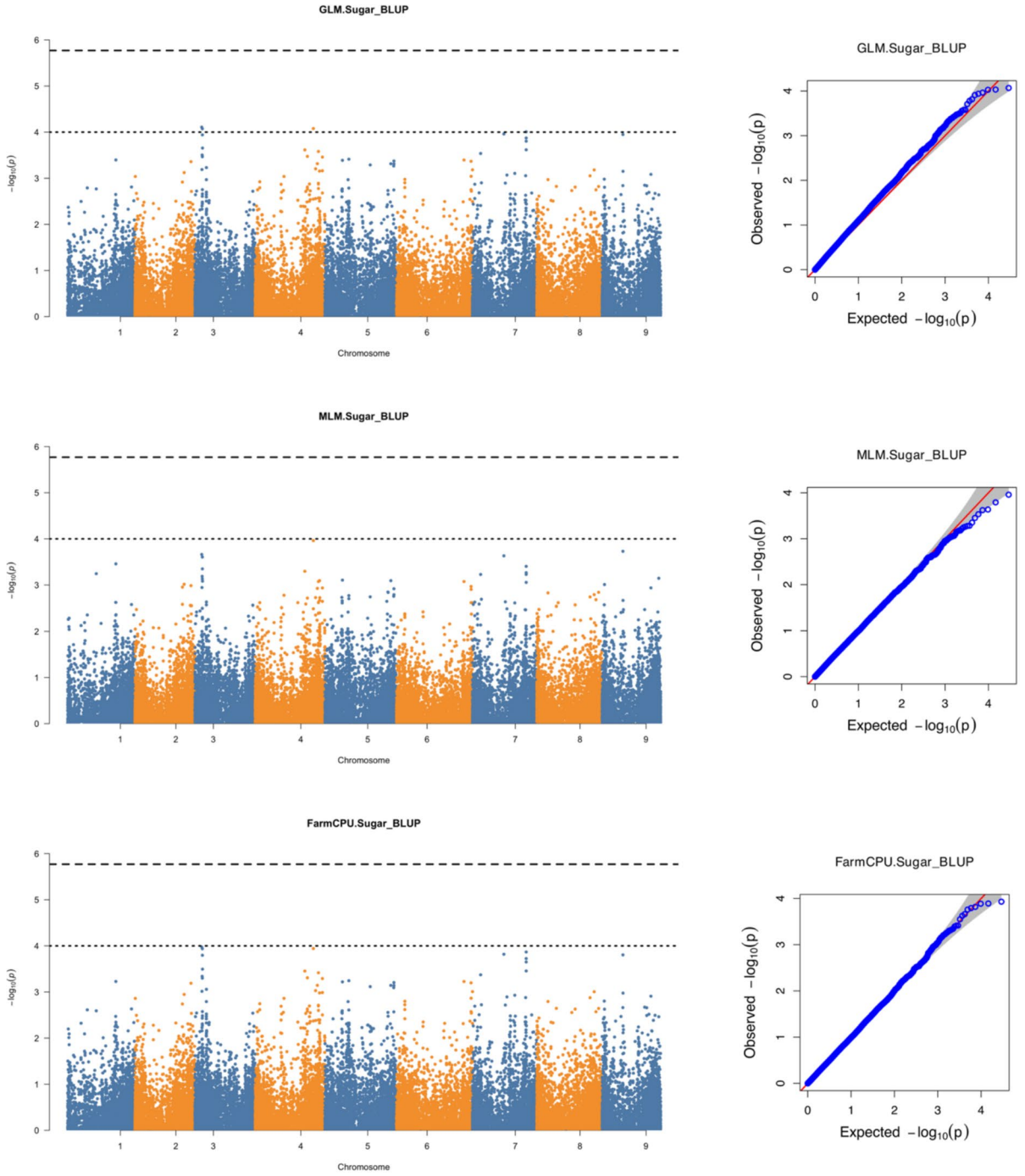


Fig. 6 Genome-wide association results for sucrose content (Sugar_BLUP) in the 177-accession sugar beet panel. Manhattan plots (left) and corresponding quantile–quantile (QQ) plots (right) are shown for the GLM, MLM, and FarmCPU models. The dashed horizontal line indicates the Bonferroni-adjusted significance threshold. QQ plots display the observed versus expected $-\log_{10}(P)$ values to assess model fit and control for population structure and relatedness

Table 1 Top-ranked SNPs associated with sucrose content identified across GWAS models. Allelic effects (β) were obtained from the MLM implemented in GAPIT

Trait	Best SNP markers	Chr	Position (bp)	*P* (GLM)	*P* (MLM)	*P* (FarmCPU)	MAF	Effect
Sucrose content	55,940,646|F|0–65:C > T-65:C > T	4	55,279,415	8.37×10^{-5}	1.09×10^{-4}	1.17×10^{-4}	0.35	−0.67
	3,656,679|F|0–67:C > G-67:C > G	9	19,975,240	1.13×10^{-4}	1.86×10^{-4}	1.28×10^{-4}	0.15	−0.65
	80,939,740|F|0–17:G > A-17:G > A	3	6,505,545	7.80×10^{-5}	2.18×10^{-4}	1.30×10^{-4}	0.20	−0.51
	27,022,523|F|0–23:C > A-23:C > A	7	29,794,291	1.09×10^{-4}	2.33×10^{-4}	1.52×10^{-4}	0.32	−0.60
	3,656,457|F|0–46:G > A-46:G > A	3	7,083,731	8.47×10^{-5}	2.48×10^{-4}	1.61×10^{-4}	0.33	−0.42

which tends to blur clear population subdivision. Whole-genome studies of cultivated beet types similarly report substantial shared variation and complex relationships among cultivated forms, consistent with the idea that breeding histories can produce weakly clustered structure rather than sharply separated subpopulations (Galewski and McGrath 2020).

From a GWAS standpoint, weak structure reduces the risk that association signals are artefacts of strong stratification, but relatedness must still be controlled. The modelling strategy used here (PC covariates in GLM and kinship in MLM; plus multi-locus approaches) reflects standard best practice for association mapping in structured or partially related samples and is supported by foundational mixed-model GWAS methods that explicitly address confounding due to both population structure and cryptic relatedness (Zhang et al. 2010).

Linkage disequilibrium, marker density, and implications for detection power

Genome-wide LD decayed rapidly, reaching background levels within ~ 10 kb. This rapid decay is consistent with the obligate outcrossing nature of sugar beet and aligns with previous reports describing short-range and heterogeneous LD patterns in elite germplasms (Li et al. 2011).

Rapid LD has important dual implications. It can enable relatively high mapping resolution when causal variants are tagged, but it also increases the requirement for dense marker coverage. In this study, the average intermarker spacing (~ 19–20 kb across an ~ 560 Mb genome) exceeded the distance at which LD approached the background. Consequently, some causal variants may not be sufficiently tagged by the available SNPs to generate strong association peaks, particularly where local marker density is low. The sugar beet reference genome provides the genomic scale for this spacing argument and supports the use of chromosome-level physical positions in interpreting association signals (Dohm et al. 2014; McGrath et al. 2023).

This LD–marker density mismatch is a plausible contributor to the absence of Bonferroni–significant loci, especially for a trait in which many loci are expected to have modest effects. More generally, GWAS theory and empirical evidence show that detection power decreases sharply when effect sizes are small and when genotyped markers are imperfect proxies for causal variants; stringent multiple-testing correction further increases the likelihood of observing "no genome-wide significant hits" even when trait heritability is high (Visscher et al. 2017).

GWAS results and interpretation relative to prior beet studies

Using BLUP-adjusted phenotypes and multiple GWAS models (GLM, MLM, FarmCPU), no marker surpassed the Bonferroni-adjusted threshold, but several loci showed consistent elevations of $-\log_{10}(P)$ across the models. The strongest and most reproducible signals were located on chromosomes 4, 9, 3, and 7 (e.g., chr4 at 55.3 Mb and chr9 at 19.9 Mb), with modest estimated additive effects (approximately −0.42 to −0.67) and MAF values ranging from 0.15 to 0.35. In this setting, the most conservative interpretation is that no single major-effect locus segregates at moderate frequency in the panel and that detectable effects are distributed across multiple regions.

This outcome is consistent with earlier sugar beet mapping studies showing the dispersed genetic control of sugar-related traits. Schneider et al. (2002) identified multiple QTLs for sugar content, yield, and

quality across environments, reinforcing the concept that sugar traits are polygenic and environmentally sensitive. Similarly, joint linkage–association mapping in sugar beet revealed that agronomically important traits are influenced by multiple loci rather than a single dominant determinant (Reif et al. 2010).

A breeding-panel GWAS by Würschum et al. (2011), conducted in 924 elite lines, also emphasized short-range LD and used genome-wide markers to detect QTLs relevant to breeding, illustrating that larger panels can recover associations for complex traits under similar biological constraints. Comparisons with more recent large-scale beet studies further contextualize the present results. For example, Liu et al. (2023) analyzed 977 Beta vulgaris germplasms and performed a GWAS for 13 traits, demonstrating that much larger panels can recover substantial numbers of marker–trait associations for complex agronomic traits under appropriate marker densities and modeling frameworks.

Similarly, Wang et al. (2024) combined resequencing-scale variant discovery with GWAS and RNA-seq across hundreds of accessions, identifying candidate regulators of sugar accumulation and biomass, including carbohydrate interconversion enzymes. These studies illustrate that increasing the sample size and variant density and integrating functional datasets can materially improve the discovery of complex traits.

Methodologically, the use of multilocus GWAS models (FarmCPU and related approaches) is appropriate for improving power and reducing confounding in complex traits by iteratively incorporating associated markers as covariates. FarmCPU and BLINK were explicitly developed to better control false positives and improve power relative to single-locus scans in settings with many markers and potentially polygenic architecture (Eu-Ahsunthornwattana et al. 2014; Liu et al. 2016).

However, even with these approaches, limited tagging resolution and modest sample size can still constrain the detection of loci with small effects, particularly under conservative Bonferroni correction (Visscher et al. 2017).

Implications for breeding and future work

The genetic architecture observed here is highly heritable, but no major-effect locus detected has direct breeding implications. For traits influenced by multiple small-effect loci, marker-assisted selection targeting a small number of individual SNPs is often inefficient; in contrast, genomic selection is designed to capture the aggregate contribution of genome-wide markers to breeding value. Sugar beet provides strong empirical support for genomic selection: Würschum et al. (2013) demonstrated genomic selection performance in sugar beet breeding populations evaluated in multilocation trials, and Hofheinz et al. (2012) reported genome-based prediction of testcross performance across breeding cycles, both of which support the practical value of genome-wide prediction when single-marker discovery is limited (Biscarini et al. 2014). Genomic prediction methods have also been extended to additional sugar beet traits and phenotypic types, supporting the wider applicability of genome-enabled selection in this crop.

From a research perspective, the present study identifies several concrete avenues to strengthen locus discovery and validation. Increasing marker density (e.g., via higher-density arrays, imputation, or resequencing) would better match the short LD observed in this panel and improve tagging of causal variants. This direction is directly supported by the genome-wide LD landscape reported for elite sugar beet germplasm (Adetunji et al 2014). Expanding the panel size would improve the statistical power for small-effect loci, as emphasized in broad GWAS syntheses.

Finally, integrating functional data (expression profiling in relevant tissues and developmental stages) and replication in independent panels would help discriminate stable, reproducible loci from sampling noise, an approach exemplified in recent beet studies combining GWAS and transcriptomics.

Taken together, the results of this study provide a carefully controlled multienvironment GWAS framework for sugar content in sugar beet breeding material, reveal highly heritable signals, document rapid LD decay, and identify a small set of reproducible candidate regions whose effects are modest.

Conclusion

The sucrose content in the evaluated panel of 177 sugar beet accessions exhibited substantial genetic control across environments, with broad-sense heritability estimated at 0.76. This finding indicates that

a considerable proportion of phenotypic variance was attributable to genotypic differences, supporting the feasibility of genetic improvement under field conditions.

Despite this high heritability, no marker surpassed the Bonferroni-adjusted genome-wide significance threshold. Several loci, particularly chromosomes 3, 4, 7, and 9, presented moderate and model-consistent association signals across the GLM, MLM, and FarmCPU analyses. The absence of major-effect loci segregating at moderate frequency suggests that the sucrose content in this germplasm panel is governed by multiple loci with individually small to moderate effects.

The rapid LD decay (~ 10 kb), together with an average marker spacing of ~ 19–20 kb, indicates that parts of the genome may not have been fully tagged by the SNP panel. In combination with the moderate sample size, this likely reduced the statistical power to detect small-effect quantitative trait loci. Therefore, the absence of genome-wide significant associations should be interpreted with caution.

Overall, these results are consistent with a quantitatively controlled and distributed genetic basis of sucrose content in elite sugar beet germplasm. From a breeding standpoint, this genetic architecture supports the application of genome-wide prediction approaches rather than reliance on single-marker selection. Future studies using larger populations, increased marker density, and multienvironment genomic models will be necessary to further elucidate the genetic determinants of sugar accumulation.

Author Contributions Mehmet Arslan, Omar Gaoua, Özge Doğanay: data collection, data arrangement, formal analysis, and writing of the original draft. Omar Gaoua: performed genomic data processing and analysis. Anıl Mehmet Baltacı, Özgür Özmen: arrangement of experimental materials and data collection. All the authors read and approved the final manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK Project no. 119O327 and 222O328) and Kayseri Şeker A.Ş. Turkey.

Data availability The datasets generated and analyzed during the current study are available in the Supplementary Information files. Additional datasets and materials are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or nonfinancial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Adetunji I, Willems G, Tschoep H, Bürkholz A, Barnes S, Boer M, Malosetti M, Horemans S, van Eeuwijk F (2014) Genetic diversity and linkage disequilibrium analysis in elite sugar beet breeding lines and wild beet accessions. Theor Appl Genet 127(3):559–571

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48

Biscarini F, Stevanato P, Broccanello C, Stella A, Saccomani M (2014) Genome-enabled predictions for binomial traits in sugar beet populations. BMC Genet 15(1):87

Boyer CD, Shannon JC (1983) The use of endosperm genes for sweet corn improvement. Plant Breed Rev 1:139–161

Creech RG (1965) Genetic control of carbohydrate synthesis in maize endosperm. Genetics 52(6):1175–1186

Cuevas HE, Rosa-Valentin G, Hayes CM, Rooney WL, Hoffmann L (2017) Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. BMC Genomics 18(1):108

De Lucchi C, Biancardi E, Skaracis G, De Biaggi M, Pavli O, Ravi S, Chiodi C, Broccanello C, Stevanato P (2021) Sugar beet (Beta vulgaris ssp. vulgaris L.) improvement with next-generation breeding technology. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Advances in plant breeding strategies: vegetable crops, vol 8. Springer, Cham, pp 305–343

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O, Sörensen TR, Stracke R, Reinhardt R (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505(7484):546–549

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Eu-Ahsunthornwattana J, Miller EN, Fakiola M et al (2014) Comparison of methods to account for relatedness in

genome-wide association studies with family-based data. PLoS Genet 10(7):e1004445
Fang L, Wang Q, Hu Y et al (2017) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. Nat Genet 49(7):1089–1098
Galewski P, McGrath JM (2020) Genetic diversity among cultivated beets (*Beta vulgaris*) assessed via population-based whole-genome sequences. BMC Genomics 21:1–14
Gruber B, Unmack P, Berry O, Georges A (2019) dartR: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour 19(3):691–699
Hanson SJ, Goldman IL (2019) Genotype is primarily responsible for variance in table beet geosmin concentration, but complex genotype × environment interactions influence variance in total dissolved solids. J Am Soc Hortic Sci 144(6):429–438
Hofheinz N, Borchardt D, Weissleder K, Frisch M (2012) Genome-based prediction of test cross performance in two subsequent breeding cycles. Theor Appl Genet 125(8):1639–1645
International Sugar Organization (2022) ISO Sugar Yearbook 2022. International Sugar Organization, London. https://www.isosugar.org/publication/296/iso-sugar-yearbook-2022
Li J, Lühmann AK, Weißleder K, Stich B (2011) Genome-wide distribution of genetic diversity and linkage disequilibrium in elite sugar beet germplasm. BMC Genomics 12(1):484
Lin T, Zhu G, Zhang J et al (2014) Genomic analyses provide insights into the history of tomato breeding. Nat Genet 46(11):1220–1226
Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12(2):e1005767
Liu D, Tan W, Wang H et al (2023) Genetic diversity and genome-wide association study of 13 agronomic traits in 977 *Beta vulgaris* L germplasms. BMC Genomics 24(1):413
McGrath JM, Funk A, Galewski P et al (2023) A contiguous de novo genome assembly of sugar beet EL10 (*Beta vulgaris* L.). DNA Res 30(1):dsac033
Niu H, Ping J, Wang Y et al (2020) Population genomic and genome-wide association analysis of lignin content in a global collection of 206 forage sorghum accessions. Mol Breed 40(8):73
Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161(1):209–228
R Core Team (2024) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
Rahman A, Wong K, Jane J, Myers AM, James MG (1998) Characterization of SU1 isoamylase, a determinant of storage starch structure in maize. Plant Physiol 117(2):425–435
Reif JC, Liu W, Gowda M et al (2010) Genetic basis of agronomically important traits in sugar beet (*Beta vulgaris* L.) investigated with joint linkage association mapping. Theor Appl Genet 121:1489–1499
Sandell FL, StralisPavese N, McGrath JM et al (2022) Genomic distances reveal relationships of wild and cultivated beets. Nat Commun 13(1):2021
Sansaloni C, Petroli C, Jaccoud D et al (2011) Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. BMC Proc 5(Suppl 7):P54
Schneider K, Schäfer-Pregl R, Borchardt D, Salamini F (2002) Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor Appl Genet 104(6), 1107–1113
Tracy WF (1996) History, genetics, and breeding of supersweet (shrunken2) sweet corn. In: Janick J (ed) Plant Breeding Reviews, vol 14. Wiley, New York, pp 189–236
Tsai CY, Nelson OE (1966) Starch-deficient maize mutant lacking adenosine diphosphate glucose pyrophosphorylase activity. Science 151(3708):341–343
Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA et al (2017) 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101(1):5–22
Wang Q, Tang J, Han B, Huang X (2020) Advances in genome-wide association studies of complex traits in rice. Theor Appl Genet 133(5):1415–1425
Wang L, Zhang Z, Han P et al (2023) Association analysis of agronomic traits and construction of genetic networks by resequencing of 306 sugar beet (*Beta vulgaris* L.) lines. Sci Rep 13(1):15422
Wang S, Yue Z, Yu C et al (2024) Genome-wide association study identifies the genetic basis of key agronomic traits in 207 sugar beet accessions. Hortic Res 11(10):uhae230
Weber WE, Borchardt DC, Koch G (1999) Combined linkage maps and QTLs in sugar beet (*Beta vulgaris* L.) from different populations. Plant Breed 118(3):193–204
Wu J, Wang L, Fu J et al (2020) Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. Nat Genet 52(1):118–125
Würschum T, Maurer HP, Kraft T, Janssen G, Nilsson C, Reif JC (2011) Genome-wide association mapping of agronomic traits in sugar beet. Theor Appl Genet 123(7):1121–1131
Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. BMC Genet 14:1–8
Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017) Genome-wide association studies in maize: praise and stargaze. Mol Plant 10(3):359–374
Zhang Z, Ersoz E, Lai CQ et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42(4):355–360

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.