

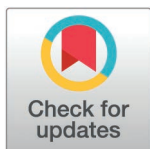
RESEARCH ARTICLE

Exploring the genetic architecture underlying dietary fiber content in Colombian Andean blueberry (*Vaccinium meridionale* Swartz)

Ginna Patricia Velasco Anacona¹, Angie Carolina Guevara Correa², Carlos-Eduardo Narváez-Cuenca², Teresa Mosquera Vásquez¹, Johana Carolina Soto Sedano¹ ^{3*}

1 Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias Agrarias, Genética de rasgos de interés agronómico, Bogotá, Colombia, **2** Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Departamento de Química, Food Chemistry Research Group, Bogotá, Colombia, **3** Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Departamento de Biología, Genética de rasgos de interés agronómico, Bogotá, Colombia

* jcsotos@unal.edu.co



OPEN ACCESS

Citation: Anacona GPV, Correa ACG, Narváez-Cuenca C-E, Vásquez TM, Soto Sedano JC (2026) Exploring the genetic architecture underlying dietary fiber content in Colombian Andean blueberry (*Vaccinium meridionale* Swartz). PLoS One 21(6): e0344321. <https://doi.org/10.1371/journal.pone.0344321>

Editor: Muhammad Qasim Shahid, South China Agricultural University, CHINA

Received: February 18, 2026

Accepted: May 15, 2026

Published: June 4, 2026

Copyright: © 2026 Anacona et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: This work was funded by the Ministry of Science, Technology, and Innovation of Colombia (MinCiencias) No.

Abstract

Dietary fiber composition is a major determinant of fruit nutritional quality, yet its genetic basis remains poorly characterized in wild *Vaccinium* species. Here, we combined extensive phenotyping with a genome-wide association study (GWAS) to dissect the genetic control of dietary fiber traits in Colombian agraz (*Vaccinium meridionale* Swartz). Total dietary fiber (TDF), insoluble dietary fiber (IDF), and soluble dietary fiber (SDF), the SDF/IDF ratio, and maturity index (MI) were quantified in fruits from 119 genotypes, representing the most comprehensive evaluation of dietary fiber fractions in fresh *Vaccinium* fruit to date. Using a GWAS approach based on 16,709 high-quality SNPs across 69 genotypes, we identified 24 QTLs associated with dietary fiber content across 15 chromosomes, revealing a polygenic architecture underlying this trait. A TDF QTL (Chr41:26883013) directly co-localized with *VaccDscf31-augustus-gene-268.33*, a 7-deoxyloganetin glucosyltransferase, embedded within a glycosyltransferase-rich LD block. IDF variation was associated with *VaccDscf33-processed-gene-116.2* (pectin methylesterase 15) while the SDF/IDF ratio co-localized with *VaccDscf55-augustus-gene-9.30*, encoding a xyloglucan endotransglucosylase/hydrolase. Together, the integration of high-resolution phenotyping with QTL mapping connects natural variation in dietary fiber content and composition to specific biosynthetic, remodeling, and regulatory pathways, providing actionable molecular targets for marker-assisted and genomic selection aimed at improving nutritional quality, texture, and processing traits in *Vaccinium* breeding programs.

BPIN 2020000100075-2021). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Blueberries are widely recognized as a health-protective fruit, primarily due to their bioactive compounds, especially flavonoids (such as anthocyanins), flavonols (such as quercetin [1] glycosides), phenolic acids (such as chlorogenic acid) [1], and dietary fiber [2]. The Colombian blueberry, known as agraz (*Vaccinium meridionale* Swartz), is a wild species belonging to the Ericaceae family, that grows spontaneously in the mountainous regions of Colombia, Ecuador, Peru, Venezuela, and Jamaica [3]. Its fruit is an edible berry with an acidic and astringent taste, valued by consumers who demand natural sources of beneficial compounds. It is consumed fresh, dehydrated or in preparations such as wines, nectars, and jams [4]. Previous studies indicate that agraz fruits contain a variety of polyphenolic compounds and dietary fiber, which are associated with significant bioactive properties, such as high antioxidant capacity [4], antiproliferative, and cytotoxic effects against cancer cells [5], antimicrobial activity, and protection against colorectal cancer [6,7].

Dietary fiber is an important component in foods since it is related to human health. A recent meta-analysis including 17,155,277 individuals reported that high dietary fiber intake is associated with a reduced risk of several chronic diseases, particularly cardiovascular disease mortality, pancreatic cancer, and diverticular disease [8]. Although we have recently reported information on the chemical and genetic architecture of polyphenolic composition in Colombian *V. meridionale* fruits [9], information regarding the variability of dietary fiber fractions in this species remains limited. Moreover, studies evaluating dietary fiber composition across large genotype collections are scarce not only for agraz but also for other *Vaccinium* species.

Dietary fiber in fruits is a complex and dynamic trait primarily determined by the composition and structural organization of plant cell wall polysaccharides [10]. It is mainly composed of cellulose, hemicellulosic polysaccharides, lignin, and pectins, which together define the mechanical properties, texture, and nutritional quality of fleshy fruits [10].

From a nutritional perspective, dietary fiber is conventionally classified into insoluble dietary fiber (IDF) and soluble dietary fiber (SDF). IDF is largely associated with cellulose, lignin, and hemicellulosic polysaccharides [11], while SDF, is dominated by pectins and hemicellulosic polysaccharides [12]. Hemicellulosic polysaccharides are present in both fractions; however, they differ markedly in structural complexity, branching patterns, and association with the plant cell wall matrix. In the IDF fraction, hemicelluloses are predominantly composed of structurally integrated polymers that are tightly bound to cellulose microfibrils and lignin [11]. These typically include less branched xylans and structural mannans or glucomannans, which contribute to cell wall rigidity and exhibit limited water solubility [10]. In contrast, the hemicellulosic polysaccharides found in the SDF fraction are generally more loosely associated with the cell wall or are readily extractable [11]. These polymers are often characterized by higher degrees of branching and structural heterogeneity, including arabinoxylans with extensive arabinose substitutions, xyloglucans, and mixed-linkage β -glucans, which enhance hydration properties and solubility [12].

Understanding the structural differences between fiber fractions requires consideration of the biosynthetic pathways that determine cell wall polymer assembly and remodeling. Cellulose is synthesized by *CesA* genes, which encode the subunits of the cellulose synthase enzyme complex, visible in the plasma membrane [13]. Hemicelluloses comprise a diverse group of heterogeneous polysaccharides that function as cross-linking components between cellulose microfibrils in the primary cell wall, including heteromannans, xyloglucans, heteroxylans, and mixed-linkage glucans [14]. Their biosynthesis is mediated predominantly by glycosyltransferases (GTs), including uridine 5'-diphospho-glucuronosyltransferases (UDP-glycosyltransferases). The synthesis of the xyloglucan backbone is catalyzed by members of the cellulose synthase-like (CSL) family, enzymes belonging to the GT2-CSL superfamily [14]. Within this group, distinct CSL clades specialize in the synthesis of different hemicellulosic polymers: CSL-C proteins are involved in xyloglucan chain elongation, CSL-A members participate in the biosynthesis of mannan and glucomannan backbones, and CSL-F enzymes contribute to mixed-linkage glucan formation [14]. In contrast, the assembly of xylan backbones is mediated by glycosyltransferases of the GT43 family, which exhibit β -1,4-xylosyltransferase activity [15]. Hemicellulose remodeling involves enzymes such as xyloglucan endotransglucosylase/hydrolases (XTHs), which modify xyloglucan chain length and connectivity [16], as well as β -xylosidases and α -L-arabinofuranosidases, which participate in the degradation and restructuring of hemicellulosic polysaccharides during cell wall remodeling [17,18].

In addition to polysaccharide components, the insoluble fiber fraction also includes lignin, a complex phenolic polymer that reinforces secondary cell walls and contributes to tissue rigidity. Lignin biosynthesis occurs through the phenylpropanoid pathway and is mediated by the coordinated action of enzymes such as cinnamoyl-CoA reductase (CCR), 4-cinnamoyl-CoA ligase (4CL), shikimate O-hydroxycinnamoyltransferase (HCT), and peroxidases (PODs), which contribute to secondary wall strengthening and tissue rigidity [19]. In *Vaccinium corymbosum*, genome-wide analyses identified 92 caffeic acid O-methyltransferase (COMT) genes, with specific members such as *VcCOMT40* and *VcCOMT92* showing a direct relationship with lignin accumulation during fruit development, underscoring their relevance for fruit firmness and shelf life [20].

Finally, pectin, a heterogeneous group of complex polymers composed mainly of homogalacturonan (HG) and rhamnogalacturonan I and II (RG-I and RG-II) [21,22]. Homogalacturonan is the most abundant pectic domain and consists of linear chains of (1 $\rightarrow$ 4)- α -D-galacturonic acid residues, whose backbone is synthesized in the Golgi apparatus by galacturonosyltransferases (GAUTs) [23]. Following biosynthesis, pectin undergoes extensive remodeling mediated by enzymes such as pectin methylesterases (PMEs) and polygalacturonases (PGs), which regulate the degree of methylesterification, depolymerization, and interactions with other cell wall polymers, thereby modulating pectin solubility and the mechanical properties of the cell wall [24].

Based on the above considerations, we hypothesize that there are genetic mechanisms underlying the natural variability in the dietary fiber in *V. meridionale*. Accordingly, this study aimed to quantify IDF, SDF, total dietary fiber (TDF), and the SDF/IDF $\times$ 100 ratio in a large collection of ripe agraz fruit genotypes collected in rural areas of the Colombian Departments of Santander, Boyacá, and Cundinamarca. In addition, a genome-wide association study (GWAS) was conducted in a representative agraz subpopulation to identify quantitative trait loci (QTL) potentially associated with dietary fiber traits. The results of this study provide deeper insight into the genetic mechanisms underlying dietary fiber content in agraz and establish a scientific foundation for future breeding strategies aimed at improving fiber-related nutritional quality. Moreover, these findings highlight the importance of advancing research in wild and underutilized crops with high nutritional and commercial potential, emphasizing their value as genetic reservoirs for crop improvement and sustainable agriculture.

Materials and methods

Fruit samples

A previously reported collection of 118 fruit genotypes sampled from rural areas of Santander, Boyacá, and Cundinamarca (Colombia) was established as a genetically diverse *V. meridionale* population [9,25]. Also, a control commercial *V. meridionale* sample was analyzed [9] to complete a collection of 119 fruit material (S1 Table).

Dietary fiber content

IDF and SDF contents were measured by the enzymatic-gravimetric method as described in the AOAC 985.29 and AOAC 991.43 methods [26] in the freeze-dried-ground samples. Both components (IDF and SDF) were experimentally found by performing three replications. Corrections because of ash and protein contents were performed for both, SDF and IDF contents. Additionally, TDF content was calculated as the sum of SDF and IDF. For sample digestion, 0.5 g of each freeze-dried sample was weighed and subjected to treatments with α -amylase, protease, and amyloglucosidase, adjusting the pH and temperature at each stage to facilitate the digestion of specific components. Dietary fiber kit was purchased from Sigma-Aldrich (Darmstadt, Germany). IDF content was determined by filtering the resulting digestate, followed by sequential washing of the retained residue with water, 78% aqueous ethanol, 95% aqueous ethanol, and acetone; the residue was then dried at 105 °C until constant weight. SDF was determined by mixing the filtrate with 95% (v/v) aqueous ethanol at a ratio of 1:4 (filtrate: aqueous ethanol, v/v). The mixture was allowed to stand overnight at 4 °C to precipitate soluble fiber components. The precipitate was recovered by filtration and sequentially washed with 78% (v/v) aqueous ethanol, 95% aqueous ethanol, and acetone. The recovered SDF was then dried at 105 °C to constant weight. Both types of fiber contents were corrected for protein and ash content. Method reproducibility was assessed by performing technical replicates for selected genotypes. Results are expressed as mean values $\pm$ standard deviation, reported on both a dry weight (DW) and fresh weight (FW) basis. To evaluate whether the measured dietary fiber variables were associated with fruit maturity stage or harvesting altitude, data for both the maturity index (MI) and altitude were obtained from previous research [9].

Statistical analysis

Dietary fiber data set of the fruit collection was analyzed by the Kolmogorov-Smirnov test for normality, one-way ANOVA, Box-plot representation, principal component analysis (PCA), and hierarchical clustering. Spearman's correlation coefficients were calculated to evaluate potential dependencies among dietary fiber parameters and to determine whether those variables were related to harvesting altitude and MI. The mentioned statistical analyses were done using R-Studio software (2025.09.2+418), with the packages: ggplot2 [27], nortest [28], tidyr [29], dplyr [30], ggrepel [31], tibble [32], and ragg [33].

Genotypic data, population structure, and genome-wide association analysis

The GWAS were performed using a previously generated single nucleotide polymorphism (SNP) dataset (S2 Table) for *V. meridionale*, originally described by [9]. Briefly, genotyping was conducted using a genotyping-by-sequencing (GBS) approach, and raw sequencing reads were processed to identify high-quality SNPs. Clean reads were mapped to the *V. corymbosum* cv. Draper v1.0 reference genome [34], using the set of the longest scaffolds (scaffolds 1–48) to represent the 48 chromosomes. SNP markers with more than 30% missing data, those with a minor allele frequency (MAF) lower than 0.01, as well as INDELs, multiallelic, and monomorphic *loci* were excluded. In addition, genotypes presenting more than 70% missing data were excluded. Following quality filtering, 69 of the initial 118 phenotyped genotypes were retained, yielding 16,709 high-quality SNPs across the filtered genotypes for downstream analyses.

Population genetic structure, genome linkage disequilibrium (LD), and genome-wide SNP distribution were previously characterized using this SNP dataset, as described by Narváez et al. [9]. Population structure was inferred using a sNMF v1.2 (Non-Negative Sparse Matrix Factorization) algorithm, implemented in the LEA v3.10.2 package in R [35]. The number of ancestral populations *K* were evaluated, using 100 independent runs for each value and a regularization parameter (α) set to 100. Individuals were then classified into *K* ancestral populations, when their ancestry proportion reached or exceeded 80%. The resulting ancestry *K* were subsequently included in the association model to account for potential confounding effects due to population stratification. Linkage disequilibrium (LD), expressed as the average

pairwise correlation coefficient (r^2) between SNPs, was calculated using the complete set of SNP markers identified in the *V. meridionale* population

To identify significant quantitative trait loci (QTL) associations, a GWAS were performed independently for each phenotypic trait that exhibited significant variability among the 69 genotypes, as determined by ANOVA, using seven different association mapping models, as follows: General Linear Model (GLM) [36], Mixed Linear Model (MLM) [37], Multiple Loci Mixed Model (MLMM) [38], Compressed Mixed Linear Model (CMLM) [39], Fixed and Random Model Circulating Probability Unification (FarmCPU) [40], Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) [41], and Settlement of MLM Under Progressively Exclusive Relationship (SUPER) [42]. Significance of marker–trait associations was initially assessed using nominal p -values derived from the GWAS models. Multiple testing was accounted for by applying a false discovery rate (FDR) correction, and associations with FDR-adjusted p -value < 0.05 were retained as significant. Quantile–quantile plots were examined to evaluate model performance and control of Type I error [43].

Candidate gene identification

Genomic positions of the identified QTLs were mapped and predicted functional annotation was performed to classify each variant as intergenic, intronic, coding, or regulatory. using the *V. corymbosum* cv. Draper v1.0 reference genome [34]. Candidate genes were identified within a LD window of ± 161.782 kb surrounding each significant QTL, according to the LD decay previously reported for the genotype-population [9]. Gene annotation information was retrieved from the VacciniumGDB database (<https://www.vaccinium.org/>).

Functional annotation of candidate genes was performed using protein homology and curated biological information from the UniProt (<https://www.uniprot.org/>) and NCBI (<https://www.ncbi.nlm.nih.gov/>) databases. Gene Ontology (GO) analyses were conducted using the OmicsBox cloud platform (<https://www.biobam.com/omicsbox/>).

Results

Dietary fiber fractions across the fruit collection

Information on the SDF, IDF, TDF contents (in DW) as well as the SDF/IDF $\times 100$ ratio is presented in S1 Table. While in the fruit collection the IDF content ranged from 19.9 ± 1.2 g/100 g DW (in G13) to 47.0 ± 2.0 g/100 g DW (in E02) (2.4-fold change), the SDF content ranged from 0.9 ± 0.2 g/100 g DW (in CP 23) and 0.9 ± 0.6 g/100 g DW (in CP40) to 6.1 ± 1.6 g/100 g DW (in G08) (6.5-fold change). Additionally, the TDF content ranged from 23.1 ± 2.0 g/100 g DW (in G03) to 49.6 ± 1.4 g/100 g DW (in E02). In all cases, the IDF fraction showed higher levels than the SDF fraction, with a SDF/IDF $\times 100$ ratio ranging from $3.0 \pm 1.8\%$ (in SC22) to $24.9 \pm 4.9\%$ (in G08), highlighting a compositional profile dominated by the IDF. The coefficients of variation (CV) observed among biological replicates were notably high, ranging from 0.6 to 25.7% for IDF, 3.5 to 97.2% for SDF, and 0.8 to 22.2% for TDF, reflecting substantial inherent biological variability among samples. In contrast, CV values obtained from technical replicates were consistently lower, ranging from 0.3 to 10.4% for IDF, 1.3 to 24.9% for SDF, and 0.1 to 9.3% for TDF, indicating that the analytical method itself exhibits good reproducibility. Furthermore, The IDF and TDF values recorded in the control fruit sample (29.7 ± 0.5 and 30.6 ± 0.2 g/100 g DW, respectively) fall within the mid-range of the values detected across the fruit collection. In contrast, the SDF content in such control sample (1.0 ± 0.4 g/100 g DW) is at the lowest end of the observed range for the fruit collection. When analyzing data of the fruit collection together with the control sample, the ANOVA indicated that IDF, SDF, TDF, and the SDF/IDF $\times 100$ ratio were statistically influenced by the genotype (F value < 0.0001). Considering a serving size of approximately 150 g (FW) and a recommended daily total dietary fiber intake of 30 g [8], the control ripe fruit sample, with a content of 4.8 ± 1.3 g TDF/100 g FW, would contribute only about 24% of the daily requirement. In contrast, several genotypes within the fruit collection showed a markedly higher nutritional contribution. Four genotypes supplied 50–57% of the recommended daily intake (V05, 50%; V04, 55%; S03, 55%; and S01, 57%).

Correlations among variables

No significant associations were found between dietary fiber fractions (IDF, SDF, TDF) or the SDF/IDF $\times$ 100 ratio and the altitude at which fruits were collected, suggesting that altitude did not substantially influence fiber composition under the conditions studied. In contrast, the MI was positively associated with IDF ($r=0.2655$, $p<0.01$), SDF ($r=0.2217$, $p<0.05$), and TDF ($r=0.3003$, $p<0.001$). A strong and significant positive correlation was found between IDF and TDF ($r=0.9622$, $p<0.001$), reflecting the predominance of the IDF over the SDF content across samples. Furthermore, SDF showed a positive, strong, and significant correlation with the SDF/IDF $\times$ 100 ratio ($r=0.9128$, $p<0.001$). A negative, strong, and significant correlation was found between IDF and the ratio SDF/IDF $\times$ 100 ($r=-0.5661$, $p<0.001$). All these results are presented in [S1 Fig](#).

Principal component analysis (PCA) and cluster analysis

PCA revealed that the first two components accounted for 99.5% of the total inertia ([Fig 1](#)). Proximity of vectors confirm the relationship described in the precedent subheading between IDF and TDF as well as between SDF and SDF/IDF $\times$ 100 ratio. Clustering analysis yielded three groups with biological interest. Group I (blue) comprises 14 genotypes and was characterized by the greatest IDF and TDF contents, along with intermediate SDF content and SDF/IDF $\times$ 100 ratio. The elevated TDF values observed in genotypes E02 (49.6 ± 1.3 g/100 g DW), O02 (47.5 ± 0.7 g/100 g DW), B07 (47.5 ± 1.1 g/100 g DW), B06 (47.4 ± 5.6 g/100 g DW), and T02 (45.1 ± 1.5 g/100 g DW), identified as outliers in the box-plot analysis ([S2 Fig](#)), suggest a superior capacity for dietary fiber accumulation, which may be associated with genotype-dependent differences in cell wall composition. Group II (green) includes most of the genotypes (82 out of the 119 fruit material collected) and was characterized by intermediate IDF and TDF contents and the lowest SDF content and SDF/IDF $\times$ 100 ratio. In contrast, Group III (orange), comprising 23 genotypes, showed the lowest IDF and TDF content but the greatest SDF content and SDF/IDF $\times$ 100 ratio. Within this group, genotypes G08 and GA24 exhibited the greatest SDF contents (6.1 ± 1.5 g/100 g DW and 6.1 ± 0.4 g/100 g DW) and the greatest SDF/IDF $\times$ 100 ratios ($24.9 \pm 4.9\%$ and $24.0 \pm 3.2\%$, respectively). These genotypes were identified as outliers in the Box-plot analysis ([S2 Fig](#)), indicating a distinct fiber composition profile dominated by soluble fractions. Taken together, Groups I and III represent contrasting dietary fiber profiles, with Group I characterized by a high TDF and IDF content, whereas Group III is distinguished by a higher proportion of soluble fiber.

Genome-wide association analysis

Analysis of population structure across the 69 genotypes indicated that $K=3$ was the most likely number of genetic clusters ([S3 Fig](#)). This structure was incorporated as a covariate in the GWAS model to control stratification effects. A total of 24 QTLs were identified across the evaluated traits, showing levels of significance ranging from 3.39×10^{-10} to 2.71×10^{-7} . Among these, the QTL Chr10:31193910 was detected in association with both TDF and IDF, exhibiting the highest statistical significance, whereas the QTL Chr41:26883013, associated with TDF, showed the lowest significant value ([Table 1](#)).

Significant associations were detected using the MLMM, BLINK, and FarmCPU models after correction for population structure and multiple testing. The Q-Q plots indicated an adequate fit between observed and expected p -values for each analyzed trait, supporting the robustness of the applied association models ([Figs 2 and 3](#)). The Manhattan plots illustrate the genomic distribution of the significant QTLs across the genome ([Figs 2 and 3](#)).

Regarding their genomic distribution, the QTLs were located in different functional regions, including nine QTLs in introns, nine within coding sequences (CDS), one in the 5' UTR, three in the 3' UTR, and two in intergenic regions. Among the QTLs located in coding regions, five nucleotide substitutions resulted in non-synonymous amino acid changes, while four corresponded to synonymous substitutions ([Table 2](#)).

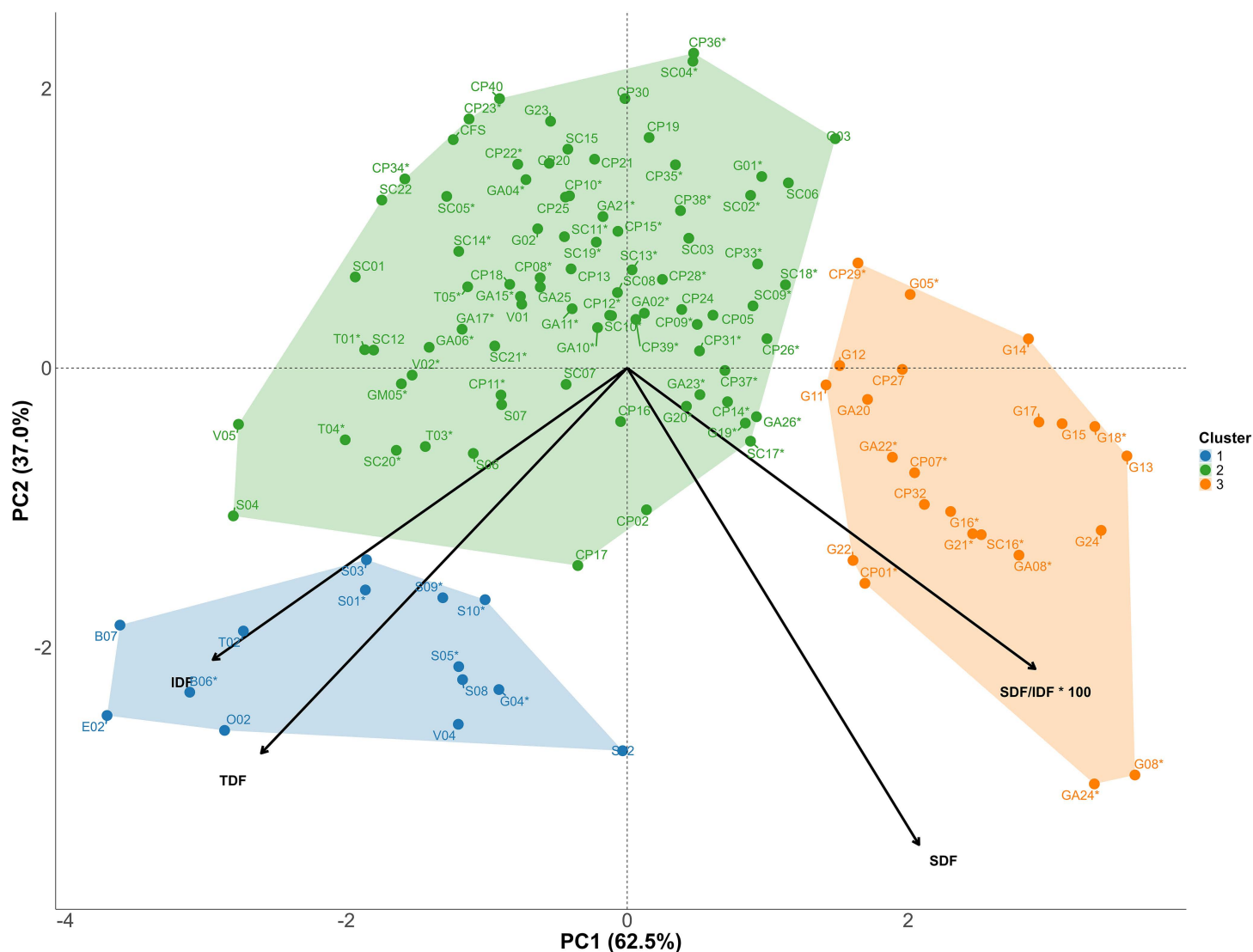


Fig 1. Principal component analysis (PCA) biplot with clustering based on insoluble dietary fiber (IDF), soluble dietary fiber (SDF), total dietary fiber (TDF), and the SDF/IDF $\times$ 100 ratio. Vectors representing each dietary fiber variable are shown. Three distinct clusters were identified. Group I (blue) comprises 14 genotypes and is characterized by the greatest IDF and TDF contents, along with intermediate SDF content and SDF/IDF $\times$ 100 ratio. Group II (green) includes 82 genotypes and is characterized by intermediate IDF and TDF contents and the lowest SDF content and SDF/IDF $\times$ 100 ratio. Group III (orange), comprising 23 genotypes, is characterized by the lowest IDF and TDF contents but the greatest SDF content and SDF/IDF $\times$ 100 ratio. Genotypes marked with an asterisk are those included in the genome-wide association study (GWAS).

<https://doi.org/10.1371/journal.pone.0344321.g001>

Based on a linkage disequilibrium window of ± 161.782 kb (S4 Fig), a total of 501 candidate genes were identified across all traits. These genes were defined exclusively by their physical proximity to significant QTLs (S3 Table).

Functional characterization of candidate genes based on Gene Ontology

Among the 501 analyzed genes, 380 were successfully annotated with Gene Ontology terms. Within the Cellular Component category, the most represented terms were intracellular anatomical structure (119 genes), membrane (108),

Table 1. Statistical summary of QTLs significantly associated with dietary fiber-related traits identified by GWAS in *Vaccinium meridionale* populations.

Trait	QTL ^e	Allele	<i>p</i> -value	H&B.P.Value ^f	Allelic Effect	MAF ^g	Model
TDF ^a	Chr41: 26883013	C/G	2.71×10^{-7}	5.0×10^{-3}	4.916	0.06	FarmCPU
	Chr36: 6980209	A/G	1.47×10^{-7}	2.0×10^{-3}	2.630	0.37	BLINK
	Chr10: 31193910	G/T	3.39×10^{-10}	0.0×10^0	-5.171	0.09	MLMM
IDF ^b	Chr10: 31193910	G/T	1.42×10^{-10}	2.37×10^{-6}	-0.077	0.09	BLINK
	Chr47: 11475242	C/T	3.57×10^{-7}	2.98×10^{-3}	-0.085	0.05	BLINK
	Chr28: 2741502	G/T	1.70×10^{-6}	2.84×10^{-2}	0.075	0.07	MLMM
SDF ^c	Chr7: 5871188	C/T	1.13×10^{-6}	6.31×10^{-3}	0.289	0.41	FarmCPU
	Chr7: 5871210	A/G	1.13×10^{-6}	6.31×10^{-3}	0.289	0.41	FarmCPU
	Chr7: 5871211	A/T	1.13×10^{-6}	6.31×10^{-3}	0.289	0.41	FarmCPU
SDF/IDF ratio	Chr55: 950149	G/T	1.02×10^{-9}	1.71×10^{-5}	-7.505	0.07	MLMM
	Chr44: 2352242	A/C	2.20×10^{-6}	1.84×10^{-2}	-4.350	0.23	MLMM
MI ^d	Chr2: 38594279	C/T	1.40×10^{-5}	2.93×10^{-2}	-1.004	0.09	BLINK
	Chr6: 38156513	G/C	9.28×10^{-24}	1.55×10^{-19}	-13.507	0.01	FarmCPU
	Chr6: 38156521	C/A	1.80×10^{-20}	1.51×10^{-16}	3.500	0.01	FarmCPU
	Chr13: 32668778	C/T	2.70×10^{-11}	9.02×10^{-8}	-2.056	0.04	BLINK
	Chr15: 10224969	C/T	1.25×10^{-6}	3.47×10^{-3}	1.839	0.03	FarmCPU
	Chr18: 3682850	C/T	5.00×10^{-7}	1.67×10^{-3}	-1.251	0.48	MLMM
	Chr21: 21155	G/A	8.05×10^{-17}	1.29×10^{-12}	-4.506	0.04	BLINK
	Chr25: 5250255	G/A	3.64×10^{-15}	2.03×10^{-11}	-3.014	0.11	MLMM
	Chr30: 12983665	A/T	1.69×10^{-8}	7.07×10^{-5}	1.438	0.49	MLMM
	Chr32: 27271453	A/C	7.68×10^{-9}	2.14×10^{-5}	-1.921	0.04	BLINK
	Chr42: 3842884	A/G	9.97×10^{-12}	4.16×10^{-8}	-3.475	0.04	BLINK
	Chr43: 2102167	C/A	8.86×10^{-16}	4.93×10^{-12}	-4.448	0.04	BLINK
	Chr46: 5969978	C/T	3.32×10^{-7}	7.93×10^{-4}	-0.848	0.45	BLINK
	Chr303: 34520	G/A	6.79×10^{-6}	1.62×10^{-2}	-1.895	0.03	FarmCPU

a TDF: Total dietary fiber

b IDF: Insoluble dietary fiber

c SDF: Soluble dietary fiber

d MI: Maturity index

e QTL: The QTLs are reported as Chr: position, where Chr indicates the chromosome and Position corresponds to the physical position (bp) in the reference genome

f H&B.P.Value: Benjamini-Hochberg and Benjamini-Yekutieli (false discovery rate) test.

g MAF: Minor allele frequency

<https://doi.org/10.1371/journal.pone.0344321.t001>

organelle (100), and cytoplasm (76), while a smaller proportion of genes was associated with the endomembrane system (22). In the Molecular Function category, genes were predominantly involved in binding and enzymatic activities, with organic cyclic compound binding (83 sequences) and small molecule binding (79) being the most frequent, followed by transferase activity (54), hydrolase activity (36), catalytic activity acting on a protein (33), carbohydrate derivative binding (30), and oxidoreductase activity (20). Regarding Biological Process, the dominant term was metabolic process (118 sequences), accompanied by a substantial representation of regulation of cellular process (50), establishment of localization (35), cellular response to stimulus (33), response to stress (30), and cellular component organization or biogenesis (26), whereas transmembrane transport, signaling, signal transduction, and cell communication were each represented by 22 sequences (Fig 4).

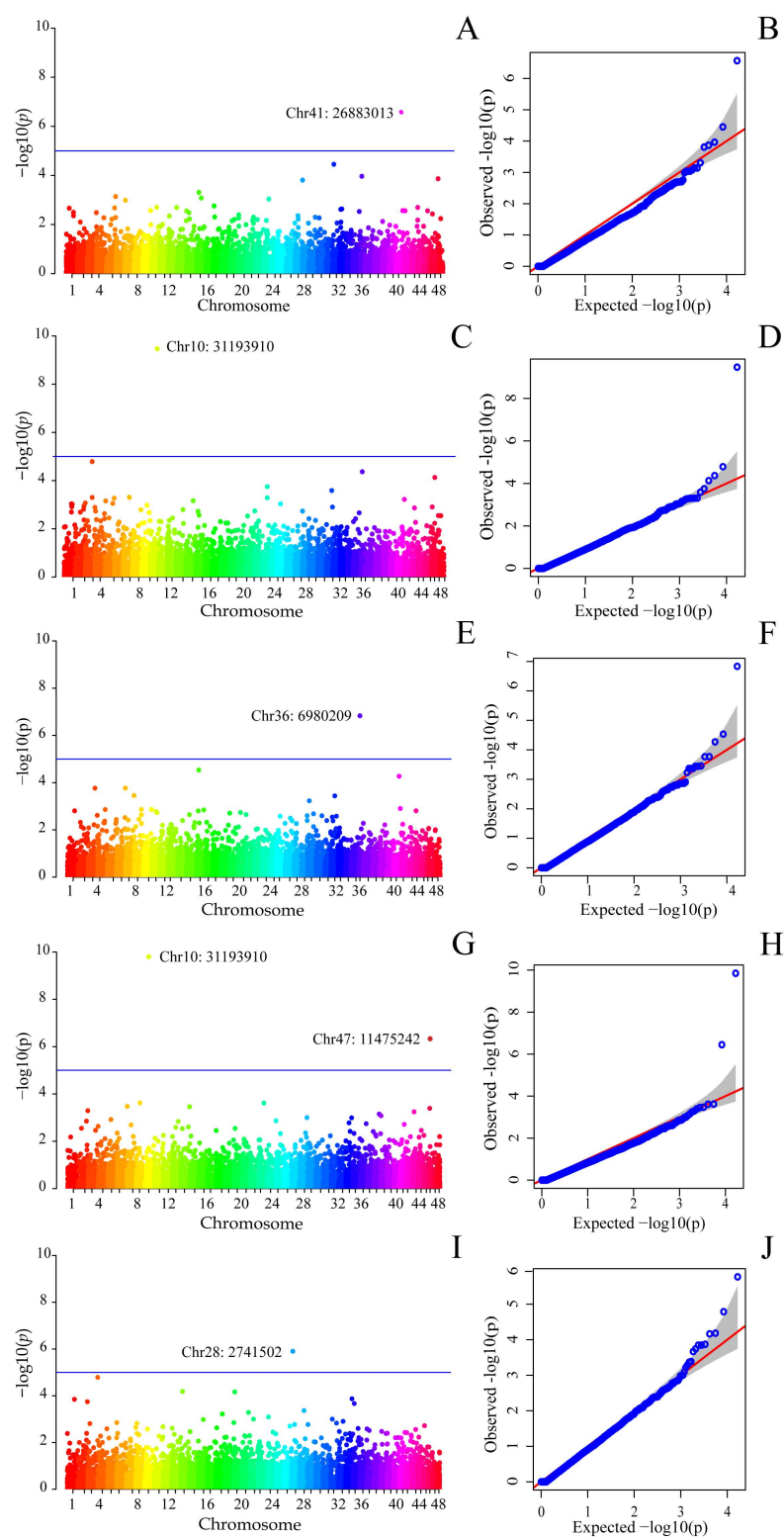


Fig 2. Manhattan and Q-Q plots of estimated $-\log_{10}(p\text{-value})$ for SNP associations with dietary fiber traits identified by GWAS. (A, B) Total dietary fiber (TDF) by the FarmCPU model. (C, D) TDF by the MLM model. (E, F) TDF by the BLINK model. (G, H) Insoluble dietary fiber (IDF) by the BLINK model. (I, J) IDF by the MLM model. The horizontal dashed line indicates the genome-wide significance threshold for phenotype-genotype associations.

<https://doi.org/10.1371/journal.pone.0344321.g002>

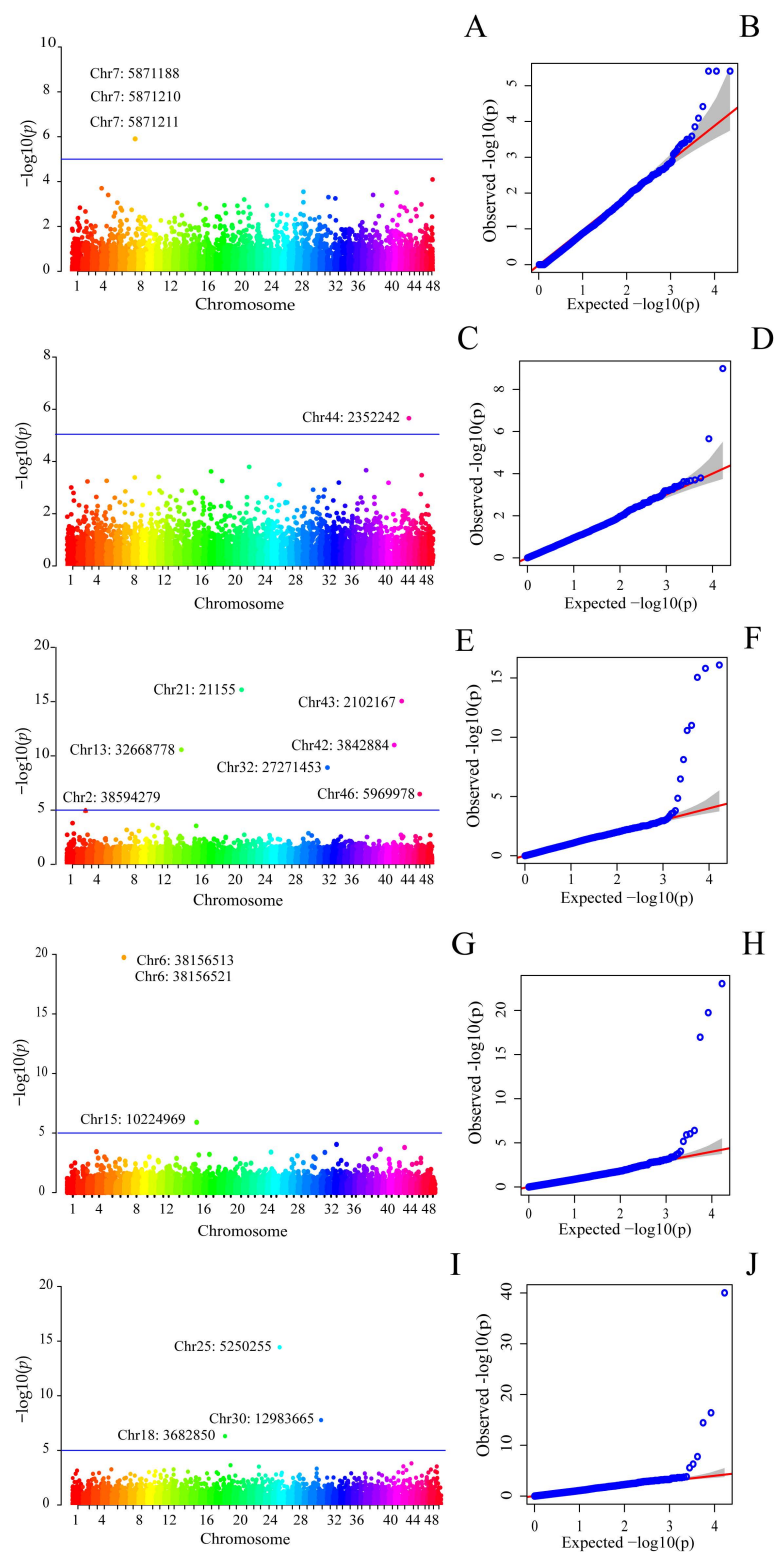


Fig 3. Manhattan and Q-Q plots of estimated $-\log_{10}(p\text{-value})$ for SNP associations with dietary fiber traits identified by GWAS. (A, B) Soluble dietary fiber (SDF) by the FarmCPU model. (C, D) SDF/IDF $\times 100$ ratio by the MLM model. (E, F) maturity index (MI) by the BLINK model. (G, H) MI by the FarmCPU model. (I, J) MI by the MLM model. The horizontal dashed line indicates the genome-wide significance threshold for phenotype-genotype associations.

<https://doi.org/10.1371/journal.pone.0344321.g003>

Table 2. Functional annotation and predicted coding effects of QTL-associated candidate genes identified in *Vaccinium meridionale*.

Trait	QTL ^e	Region	Gene ID	Codon change ^f	Amino acid change ^g	Gene function	#CG ^h
TDF ^a	Chr41: 26883013	Intron	<i>VaccDscaff31-augustus-gene-268.33</i>	NT	NT	7-deoxyloganetin glucosyltransferase	23
	Chr36: 6980209	Intron	<i>VaccDscaff46-snap-gene-69.23</i>	NT	NT	DEAD-box ATP-dependent RNA helicase 5	14
	Chr10: 31193910	Intergenic	NT	NT	NT	–	21
IDF ^b	Chr10: 31193910	Intergenic	NT	NT	NT	–	21
	Chr47: 11475242	UTR 5'	<i>VaccDscaff33-augustus-gene-114.25-mRNA-1</i>	NT	NT	Cactin	28
	Chr28: 2741502	CDS	<i>VaccDscaff38-augustus-gene-27.36-mRNA-1</i>	AA[C]/AA[A]	N169K	Ankyrin repeat-containing protein BDA1	25
SDF ^c	Chr7: 5871188	Intron	<i>VaccDscaff12-snap-gene-58.40-mRNA-1</i>	NT	NT	Multidrug resistance efflux transporter EmrE	28
	Chr7: 5871210	CDS		[T]GT/[C]GT	R53R		
	Chr7: 5871211	CDS		AT[T]/AT[A]	I52I		
SDF/IDF ratio	Chr55: 950149	Intergenic	NT	NT	NT	–	29
	Chr44: 2352242	Intron	<i>VaccDscaff42-augustus-gene-23.32</i>	NT	NT	F-box/LRR-repeat protein 7	30
MI ^d	Chr2: 38594279	UTR3	<i>VaccDscaff2-augustus-gene-386.31</i>	NT	NT	Leucine-rich repeat-containing protein	27
	Chr6: 38156513	Intron	<i>VaccDscaff11-augustus-gene-381.25</i>	NT	NT	Family not named	33
	Chr6: 38156521	Intron		NT	NT		
	Chr13: 32668778	CDS	<i>VaccDscaff5-snap-gene-326.15</i>	[G]CT/[A]CT	S380T	Syntaxin SynN_4	23
	Chr15: 10224969	Intron	<i>VaccDscaff9-snap-gene-102.35</i>	NT	NT	DNA_polymerase_A	24
	Chr18: 3682850	CDS	<i>VaccDscaff15-processed-gene-37.1</i>	[G]TT/[A]TT	F224I	Catalase_3	8
	Chr21: 21155	Intron	<i>VaccDscaff27-snap-gene-0.31</i>	NT	NT	Family not named	12
	Chr25: 5250255	CDS	<i>VaccDscaff8-snap-gene-52.38</i>	[C]CC/[T]CC	S179S	AA_permease_2	30
	Chr30: 12983665	UTR3	<i>VaccDscaff19-augustus-gene-129.36</i>	NT	NT	Ethylene-responsive transcription factor ERF118	20
	Chr32: 27271453	CDS	<i>VaccDscaff32-snap-gene-272.30</i>	TT[A]/TT[C]	L57F	ATP-Binding Cassette Transporter	18
	Chr42: 3842884	CDS	<i>VaccDscaff24-augustus-gene-38.30</i>	[A]AG/[G]AG	K181E	Family not named	39
	Chr43: 2102167	UTR3	<i>VaccDscaff41-augustus-gene-21.19</i>	NT	NT	Family not named	14
	Chr46: 5969978	Intron	<i>VaccDscaff48-augustus-gene-59.47</i>	NT	NT	Family not named	40
	Chr303: 34520	CDS	<i>VaccDscaff303-snap-gene-0.58</i>	G[C]T/G[T]T	G5823V	ER lumen protein-retaining receptor	15

a TDF: Total dietary fiber

b IDF: Insoluble dietary fiber

c SDF: Soluble dietary fiber

d MI: Maturity index

e QTL: The QTLs are reported as Chr: position, where Chr indicates the chromosome and Position corresponds to the physical position (bp) in the reference genome

f Codon change: Codon variation at the SNP site is presented as Reference/Alternative. For example, AA[C]/AA[A] indicates a change from C (reference) to A (alternative).

g Amino acid change: Amino acid substitutions are indicated using the one-letter code (Original Position Alternative). The first letter represents the reference amino acid, the number of its position, and the second letter the alternative amino acid. K (Lysine), R (Arginine), I (Isoleucine), S (Serine), T (Threonine), F (Phenylalanine), L (Leucine), E (Glutamic acid), G (Glycine), V (Valine), and NT (Not transcribed)

h #CG: Number of candidate genes in LD with each QTL

<https://doi.org/10.1371/journal.pone.0344321.t002>

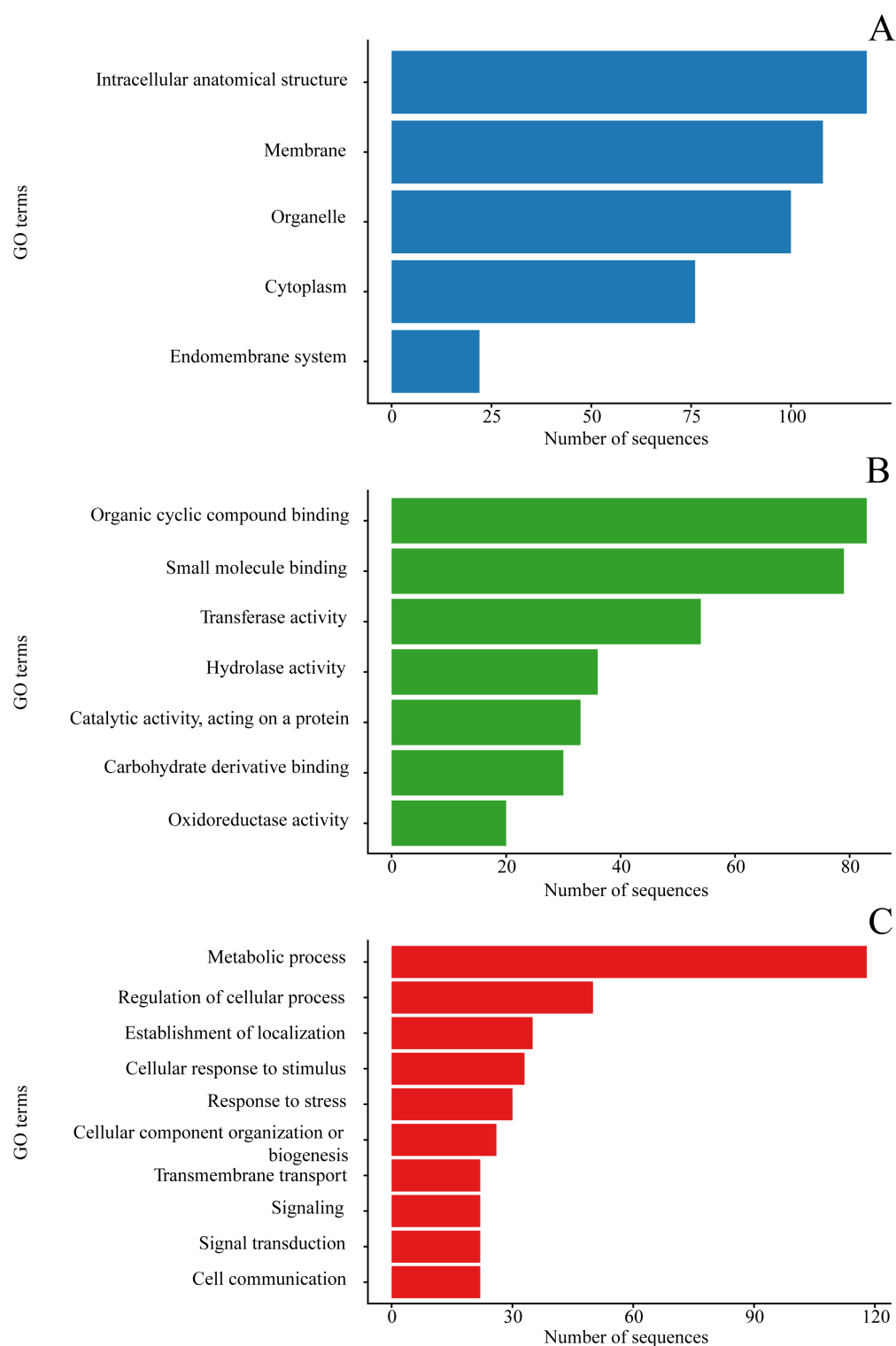


Fig 4. Gene Ontology (GO) classification of the annotated genes. Bar plots show the distribution of GO terms assigned to the 380 annotated genes across the three main GO categories: **(A)** Cellular Component, **(B)** Molecular Function, and **(C)** Biological Process. While the x-axis indicates the number of sequences associated with each term, the y-axis represents the GO terms within each category.

<https://doi.org/10.1371/journal.pone.0344321.g004>

Discussion

Genotype and physiological maturity drive dietary fiber variability in *V. meridionale* fruits

The contents of IDF found in the fruit collection of *V. meridionale* are greater than the values found in six cultivars of *V. corymbosum* (blueberry, 10.6–15.5 g/100 g DW) [44], and also greater than the one found in *Vaccinium myrtillus* (bilberry, 17.8 g/100 g DW) [45]. In contrast, the values reported of SDF not only in *V. myrtillus* (6.2 g/100 g DW) [45] but also in *V. corymbosum* (six cultivars, 2.4–4.3 g/100 g DW) [44] are comparable with the values for this dietary component found in the *V. meridionale* fruit collection. Predominance of IDF towards SDF coincides with such findings in *V. corymbosum* [44] and *V. myrtillus* (Reference) and it seems to be a common feature in different fruits [46]. IDF is associated with a positive effect on bulking fecal material, improves gastrointestinal motility, can be fermented at a certain extent by gut microbiota, and is more protective than SDFs against some metabolic diseases such as Type 2 diabetes mellitus. In contrast, SDF reduces the diffusion and absorption rates of glucose and lipids and is closely related to improving gut microbiota due to its high fermentability which is negatively correlated with the incidence rate of many chronic diseases, including inflammatory bowel disease, colorectal cancer, obesity, and type 2 diabetes mellitus [47,48]. This divergence highlights complementary nutritional targets for genotype selection. Moreover, the identified genotypes meeting daily dietary fiber requirements suggest that these materials would provide good candidates for breeding programs.

Although fruits were harvested at a similar external color, the MI exhibited considerable variability, suggesting heterogeneity in physiological maturity. A previous study analyzed ripe berries from a large set of *Vaccinium stamineum* genotypes harvested when pedicels changed from green to red, which was used as a visual indicator of maturation. However, despite the application of this criterion, considerable variability in physiological maturity was detected, as reflected by differences in pH, total soluble solids, and titratable acidity [49]. The authors concluded that both genotype and physiological maturity significantly affected these quality parameters. These findings are consistent with the variability observed in the present study and support the notion that visual harvest indices alone may not adequately represent physiological maturity in *Vaccinium* berries. Additionally, the MI may contribute to differences in the accumulation or structural modification of dietary fiber components, potentially through ripening associated cell wall polysaccharide remodeling [50].

Multiple QTLs suggest polygenic and regulatory control of dietary fiber composition in *V. meridionale*

Genome-wide association analyses suggest a complex and polygenic genetic architecture underlying dietary fiber-related traits in *V. meridionale*. Across TDF, IDF, SDF, the SDF/IDF ratio, and MI, multiple QTLs were identified across at least 15 chromosomes using complementary association models (BLINK, FarmCPU, and MLM). These QTLs mapped to both candidate coding and non-coding genomic regions, suggesting a genetic model in which dietary fiber traits are shaped by a combination of structural variation in key enzymes and regulatory variation affecting gene expression [10]. This pattern is consistent with the quantitative nature of dietary fiber composition, which depends on coordinated cell wall biosynthesis, remodeling, and developmental regulation rather than on single major-effect loci [51]. In this context, the identified genes are proposed as candidates based on genomic proximity and functional annotation, and their role in dietary fiber traits requires further validation.

Glycosyltransferase (GT) gene families underpin fiber-related variation

A striking convergence between QTL location and functional annotation was observed for genes encoding glycosyltransferases across all fiber-related traits, supporting their central role in dietary fiber variation in *V. meridionale*. For TDF, the QTL Chr41:26883013 directly overlapped the glycosyltransferase gene *VaccDscaff31-augustus-gene-268.33*, annotated as a 7-deoxyloganetin glucosyltransferase and showing 56.31% sequence identity with its *Gardenia jasminoides* homolog, strongly suggesting this gene as a primary causal candidate underlying the QTL effect. This signal was further reinforced by the presence of five additional glycosyltransferase genes within the same LD block, spanning distances from 11.4 to

52.7 kb, including *VaccDscf31-processed-gene-268.6* (11,444 bp), *VaccDscf31-snap-gene-268.40* (16,617 bp), and *VaccDscf31-augustus-gene-268.31* (40,394 bp), all annotated as 7-deoxyloganetin glucosyltransferases, as well as *VaccDscf31-augustus-gene-268.38* (52,534 bp), encoding a UDP-glycosyltransferase 85A3, and *VaccDscf31-processed-gene-268.14* (52,689 bp), encoding UDP-glycosyltransferase 85A5. The tight physical clustering and functional coherence of these genes indicate a glycosylation hotspot likely contributing to variation in fiber biosynthesis or cell wall-associated secondary metabolites.

For IDF, the QTL Chr47:11475242 mapped within a LD block containing *VaccDscf33-snap-gene-113.46*, located 99.5 kb from the peak SNP and annotated as a probable glycosyltransferase homologous to At5g03795 from *Arabidopsis thaliana*. Although this gene does not overlap the QTL peak, its proximity and annotation support a role in the biosynthesis and modification of hemicellulosic polysaccharides, tightly bound to the cell wall matrix, and cellulosic polysaccharides, which constitute the main structural components of IDF [52,53]. In addition, glycosyltransferase mediated glycosylation has been shown to regulate lignin deposition and secondary wall development in fruit tissues, as demonstrated for PbUGT72AJ2 in pear (*Pyrus bretschneideri*), where altered uridine diphosphate glycosyltransferase (UGT) activity modulates lignin accumulation [54].

Similarly, for the SDF/IDF ratio, the QTL Chr55:950149 was associated with *VaccDscf55-augustus-gene-9.30* at 29.3 kb, encoding a xyloglucan endotransglucosylase/hydrolase. Xyloglucan-modifying enzymes regulate hemicellulose restructuring and wall extensibility, processes known to influence fiber solubility [55,56]. The physical proximity of this gene to the QTL therefore supports its candidacy as a contributor to variation in the relative proportions of soluble and insoluble dietary fiber.

Glycosyltransferase involvement extended to MI, as MI-associated QTLs colocalized with genes encoding enzymes involved in cell wall polysaccharide modification rather than pathways directly related to sugar or organic acid biosynthesis. Specifically, the QTL Chr18: 3682850 was associated with *VaccDscf15-snap-gene-38.27* (14.5 kb), encoding UDP-glycosyltransferase 92A1; Chr32: 27271453 with *VaccDscf32-augustus-gene-273.17* (85.2 kb), encoding a dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit 4; and Chr43: 2102167 with *VaccDscf41-augustus-gene-22.14* (124.9 kb), encoding a probable xyloglucan glycosyltransferase 12. These associations reinforce the role of glycosylation-driven cell wall remodeling during fruit maturation, a process in which xyloglucan-modifying enzymes are central to the assembly and restructuring of hemicellulose networks within the primary cell wall. Consistent with this interpretation, Yuan et al. [57] identified xyloglucan glycosyltransferases as core components of the cellulose–hemicellulose matrix, directly contributing to wall extensibility and polysaccharide architecture. Moreover, the significant positive correlations observed between MI and IDF, SDF, and TDF indicate that MI in *V. meridionale* integrates coordinated changes in dietary fiber composition, reflecting structural modifications of the cell wall that are characteristic of fruit ripening rather than solely the balance between soluble sugars and organic acids.

The enrichment of glycosyltransferase genes within fiber-associated QTLs is supported by functional studies demonstrating that GT involved in the assembly and remodeling of wall polysaccharides strongly influence fiber solubility and physicochemical properties [44]. Moreover, genome-wide characterization of the UDP-glycosyltransferase family (UGT subfamily) in blueberry (*V. corymbosum*) identified 361 VcUGTs, with marked expansion of groups E and G driven by genome duplication events [58]. The co-localization of UGTs within LD windows of multiple fiber-related QTLs in *V. meridionale* is therefore consistent with the evolutionary and functional prominence of this family in blueberry.

Functional evidence further supports this link: silencing of VcUGT160 (UGT85 subfamily) altered cytokinin homeostasis during blueberry fruit development [58]. Given the role of cytokinins in cell expansion and wall deposition, allelic variation in UGT genes near fiber-associated QTLs may influence dietary fiber traits through hormone-mediated developmental pathways. Collectively, the repeated association of specific glycosyltransferase families; UDP-glycosyltransferases, and xyloglucan endotransglucosylase/hydrolases with QTLs across multiple traits provides strong evidence that allelic variation in glycosylation pathways represents a major genetic determinant of dietary fiber composition in *V. meridionale*.

Cell wall remodeling enzymes and polysaccharide turnover

The QTLs identified for dietary fiber traits in *V. meridionale* revealed a consistent enrichment of candidate genes involved in cell wall remodeling and polysaccharide turnover, highlighting the importance of post-synthetic regulation of wall architecture in shaping fiber composition. Dietary fiber composition in fruits is increasingly recognized as a dynamic trait governed not only by primary biosynthetic pathways but also by extensive remodeling of pectic [59] and hemicellulosic polymers during development and ripening, which directly affects polysaccharide interactions, extractability, and solubility [10]. Consequently, the fiber-associated loci identified here point out to coordinated enzymatic and regulatory mechanisms controlling wall restructuring, rather than isolated variation in biosynthetic capacity.

A key component of this regulatory framework involves genes controlling pectin modification, given the central role of pectin structure in cell wall mechanics and fruit texture [60]. The IDF-associated QTL Chr47: 11475242 was linked to *VaccDscf33-processed-gene-116.2*, located 143.3 kb from the peak SNP and annotated as a probable pectin methyl-esterase. Although pectins are a major component of SDF, their physicochemical state is highly dynamic and can be altered by enzymatic modification, particularly through changes in the degree of methyl-esterification that influence solubility and interactions within the cell wall [61,62]. Pectin methyl-esterases regulate the demethylesterification of homogalacturonan, a process that reduces pectin solubility by promoting calcium-mediated cross-linking and enhanced interactions with cellulose and hemicelluloses [60,63]. Consequently, increased demethylesterification favors the formation of rigid pectate networks with reduced extractability, effectively shifting pectins toward the insoluble dietary fiber fraction [64,24]. Consistent with previous studies linking cell wall pectin to tissue mechanical properties, and intercellular adhesion and firmness in fruits [65,66]. The proximity of this pectin methyl-esterase gene to the IDF-associated QTL supports its role in modulating insoluble fiber accumulation through coordinated regulation of cell wall architecture rather than direct polymer identity.

In addition to enzymatic candidates, several QTL regions harbored regulatory genes implicated in signaling pathways that coordinate cell wall remodeling with developmental cues. Notably, receptor-like kinases (RLKs) were associated with the IDF QTL Chr28:2741502, including *VaccDscf38-snap-gene-27.46* at 29.7 kb, and with the MI QTL Chr25:5250255 through *VaccDscf8-snap-gene-54.38* at 152.3 kb. RLKs form a large family of plasma membrane localized receptors that perceive extracellular and developmental signals and transduce them to intracellular signaling networks, thereby influencing cell expansion, wall integrity maintenance, and polysaccharide remodeling [67,68]. Functional studies have shown that specific RLK subfamilies, such as wall-associated kinases (WAKs) and malectin-like receptor kinases (CrRLK1Ls), directly interact with cell wall components, including pectin, and trigger downstream pathways that regulate cell wall dynamics and development [69,70]. For example, WAKs can bind pectin fragments resulting from wall modification and initiate signaling that alters expression of cell wall remodeling genes, while CrRLK1Ls such as *FERONIA* mediate growth regulation in response to cell wall status [69]. The co-localization of these receptor kinases with fiber-related QTLs in *V. meridionale* indicates that allelic variation in RLK-mediated signaling could underlie differential regulation of wall remodeling during fruit development and maturation, integrating environmental and developmental signals with changes in wall architecture.

Taken together, our results provide the first comprehensive insight into the polygenic basis of dietary fiber variation in *V. meridionale*, revealing a complex and promising genetic framework that opens new avenues for molecular and precision breeding strategies in agraz. Nonetheless, expanding population size and incorporating functional validation will be essential to confirm causality and establish biological relevance. Thus, the associations identified here should be considered exploratory, and the detected QTLs and candidate genes require validation in larger, independent populations, including future studies involving gene expression analyses and functional approaches to confirm their role in the evaluated traits.

Conclusion

This study presents an extensive and underexplored diversity in dietary fiber composition within *Vaccinium meridionale*, identifying this species as a promising source of dietary fiber within the *Vaccinium* genus. The identification of genotypes capable of providing up to 57% of the recommended daily fiber intake highlights the outstanding functional potential of this

species. Moreover, the genetic architecture elucidated in this study indicates that cell wall glycosylation, remodeling processes, and putative regulatory pathways are major determinants of dietary fiber composition in *V. meridionale*. The consistent co-localization of fiber-related QTLs with genes encoding glycosyltransferases, pectin methylesterase enzymes, and xyloglucan-modifying enzymes provides promising targets for breeding strategies aimed at modulating both total fiber content and the balance between soluble and insoluble fractions. Given that dietary fiber strongly influences fruit texture, firmness, and extractability, allelic variation at these loci is expected to affect not only nutritional quality but also key fruit quality attributes relevant for fresh consumption and processing. In particular, genes involved in pectin and hemicellulose remodeling represent leverage points for fine-tuning fiber solubility, which are directly linked to recognized health benefits. Together, these results open future possibilities for developing a molecular framework to guide the breeding of *Vaccinium* varieties with improved dietary fiber profiles, with potential to integrate nutritional value and desirable fruit quality traits through marker-assisted and genomics-informed selection.

Supporting information

S1 Table. Phenotypic dataset of dietary fiber measurements in the *Vaccinium meridionale* population.
(XLSX)

S2 Table. SNPs data set for GWAS.
(TXT)

S3 Table. List of candidate genes in linkage disequilibrium (LD) surrounding significantly associated QTLs.
(XLSX)

S1 Fig. Spearman's correlation coefficients including IDF: Insoluble dietary fiber; SDF: Soluble dietary fiber; TDF: Total dietary fiber; SDF/IDF x 100 ratio, the altitude where each genotype was sampled and the maturity index. * Denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$.
(DOCX)

S2 Fig. Box-plot displayed for each of the measured response variables in the fruit collection, including the control commercial fruit. IDF: Insoluble dietary fiber; SDF: Soluble dietary fiber; TDF: Total dietary fiber; and SDF/IDF x 100 ratio.
(DOCX)

S3 Fig. Population structure analysis of the *V. meridionale* genotypes. LEA admixture coefficient analysis, K1 (green), K2 (purple) and K3 (orange).
(DOCX)

S4 Fig. Linkage disequilibrium (LD) decay at the genomic level calculated for the *V. meridionale* panel.
(DOCX)

Acknowledgments

The authors would like to thank the peasants in the various municipalities where the collection was carried out for their guidance and for serving as guides for the scientific team. We also thank the farmers who provided us with commercial materials.

Author contributions

Conceptualization: Carlos-Eduardo Narváez-Cuenca, Teresa Mosquera Vásquez, Johana Carolina Soto Sedano.

Data curation: Ginna Patricia Velasco Anacona, Angie Carolina Guevara Correa.

Formal analysis: Ginna Patricia Velasco Anacona, Angie Carolina Guevara Correa.

Funding acquisition: Teresa Mosquera Vásquez.

Investigation: Ginna Patricia Velasco Anacona, Carlos-Eduardo Narváez-Cuenca, Teresa Mosquera Vásquez, Johana Carolina Soto Sedano.

Methodology: Ginna Patricia Velasco Anacona, Angie Carolina Guevara Correa.

Project administration: Carlos-Eduardo Narváez-Cuenca, Teresa Mosquera Vásquez.

Resources: Teresa Mosquera Vásquez.

Software: Angie Carolina Guevara Correa.

Supervision: Carlos-Eduardo Narváez-Cuenca, Johana Carolina Soto Sedano.

Validation: Carlos-Eduardo Narváez-Cuenca, Johana Carolina Soto Sedano.

Writing – original draft: Ginna Patricia Velasco Anacona.

Writing – review & editing: Angie Carolina Guevara Correa, Carlos-Eduardo Narváez-Cuenca, Teresa Mosquera Vásquez, Johana Carolina Soto Sedano.

References

1. Montanari S, Thomson S, Cordiner S, Günther CS, Miller P, Deng CH, et al. High-density linkage map construction in an autotetraploid blueberry population and detection of quantitative trait loci for anthocyanin content. *Front Plant Sci.* 2022;13:965397. <https://doi.org/10.3389/fpls.2022.965397> PMID: 36247546
2. Miller K, Feucht W, Schmid M. Bioactive Compounds of Strawberry and Blueberry and Their Potential Health Effects Based on Human Intervention Studies: A Brief Overview. *Nutrients.* 2019;11(7):1510. <https://doi.org/10.3390/nu11071510> PMID: 31269727
3. Ligarreto-Moreno GA, Torres-Aponte WS, Ariza-Castillo A. Propagation of the neotropical fruit *Vaccinium meridionale* Swartz by air layering. *Propagación del frutal neotropical Vaccinium meridionale Swartz por acodos aéreos.* 2013.
4. Garzón GA, Soto CY, López-R M, Riedl KM, Browmiller CR, Howard L. Phenolic profile, in vitro antimicrobial activity and antioxidant capacity of *Vaccinium meridionale swartz* pomace. *Heliyon.* 2020;6(5):e03845. <https://doi.org/10.1016/j.heliyon.2020.e03845> PMID: 32490223
5. González M, Samudio I, Sequeda-Castañeda LG, Celis C, Iglesias J, Morales L. Cytotoxic and antioxidant capacity of extracts from *Vaccinium meridionale Swartz* (Ericaceae) in transformed leukemic cell lines. *J Appl Pharm Sci.* 2017;7:24–30. <https://doi.org/10.7324/JAPS.2017.70305>
6. Arango-Varela SS, Luzardo-Ocampo I, Maldonado-Celis ME, Campos-Vega R. Andean berry (*Vaccinium meridionale Swartz*) juice in combination with Aspirin modulated anti-inflammatory markers on LPS-stimulated RAW 264.7 macrophages. *Food Res Int.* 2020;137:109541. <https://doi.org/10.1016/j.foodres.2020.109541> PMID: 33233169
7. Osorio M, Posada L, Martínez E, Estrada V, Quintana G, Maldonado ME, et al. Bacterial nanocellulose spheres coated with meta acrylic copolymer: *Vaccinium meridionale swartz* extract delivery for colorectal cancer chemoprevention. *Food Hydrocolloids.* 2024;147:109310. <https://doi.org/10.1016/j.foodhyd.2023.109310>
8. Veronese N, Gianfredi V, Solmi M, Barbagallo M, Dominguez LJ, Mandalà C, et al. The impact of dietary fiber consumption on human health: An umbrella review of evidence from 17,155,277 individuals. *Clin Nutr.* 2025;51:325–33. <https://doi.org/10.1016/j.clnu.2025.06.021> PMID: 40651334
9. Narváez-Cuenca CE, Guevara-Correa AC, Velasco Anacona GP, Sepúlveda J, Soto Sedano JC, Mosquera-Vásquez T. New anthocyanin and non-anthocyanin phenolic compounds in wild bilberry (*Vaccinium meridionale Swartz*) and their genetic architecture. *ACS Agricultural Science & Technology.* 2023. <https://doi.org/10.1021/acsagscitech.5c00458>
10. Zhang X, Lu M, Ludlow RA, Ma W, An H. Transcriptome analysis reveals candidate genes for dietary fiber metabolism in *Rosa roxburghii* fruit grown under different light intensities. *Hortic Environ Biotechnol.* 2021;62:751–64. <https://doi.org/10.1007/s13580-021-00359-6>
11. Du X, Pérez-Boada M, Fernández C, Rencoret J, del Río JC, Jiménez-Barbero J, et al. Analysis of lignin-carbohydrate and lignin-lignin linkages after hydrolase treatment of xylan-lignin, glucomannan-lignin and glucan-lignin complexes from spruce wood. *Planta.* 2014;239(5):1079–90. <https://doi.org/10.1007/s00425-014-2037-y> PMID: 24531838
12. Mudgil D. The interaction between insoluble and soluble fiber. *Dietary fiber for the prevention of cardiovascular disease.* Elsevier. 2017:35–59. <https://doi.org/10.1016/B978-0-12-805130-6.00003-3>
13. Schneider R, Hanak T, Persson S, Voigt CA. Cellulose and callose synthesis and organization in focus, what's new? *Current Opinion in Plant Biology.* 2016;9–16. <https://doi.org/10.1016/j.pbi.2016.07.007>
14. Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, et al. Hemicellulose biosynthesis. *Planta.* 2013;238(4):627–42. <https://doi.org/10.1007/s00425-013-1921-1> PMID: 23801299

15. Anders N, Wilson LFL, Sorieul M, Nikolovski N, Dupree P. β -1,4-xylan backbone synthesis in higher plants: how complex can it be? *Front Plant Sci*. 2023;13. <https://doi.org/10.3389/fpls.2022.1076298>
16. Han Y, Zhu Q, Zhang Z, Meng K, Hou Y, Ban Q, et al. Analysis of xyloglucan endotransglycosylase/hydrolase (XTH) genes and diverse roles of isoenzymes during persimmon fruit development and postharvest softening. *PLoS One*. 2015;10(4):e0123668. <https://doi.org/10.1371/journal.pone.0123668> PMID: 25849978
17. Rohman A, Dijkstra BW, Puspaningsih NNT. β -Xylosidases: Structural Diversity, Catalytic Mechanism, and Inhibition by Monosaccharides. *Int J Mol Sci*. 2019;20(22):5524. <https://doi.org/10.3390/ijms20225524> PMID: 31698702
18. Lagaert S, Pollet A, Courtin CM, Volckaert G. β -xylosidases and α -L-arabinofuranosidases: accessory enzymes for arabinoxylan degradation. *Biotechnol Adv*. 2014;32(2):316–32. <https://doi.org/10.1016/j.biotechadv.2013.11.005> PMID: 24239877
19. Lu M, Ma WT, Liu YQ, An HM, Ludlow RA. Transcriptome analysis reveals candidate lignin-related genes and transcription factors in *Rosa roxburghii* during fruit ripening. *Plant Mol Biol Report*. 2020;38:331–42. <https://doi.org/10.1007/s11105-020-01193-3>
20. Liu Y, Wang Y, Pei J, Li Y, Sun H. Genome-wide identification and characterization of COMT gene family during the development of blueberry fruit. *BMC Plant Biol*. 2021;21(1):5. <https://doi.org/10.1186/s12870-020-02767-9> PMID: 33407129
21. Ropartz D, Ralet MC. Pectin Structure. Pectin: Technological and Physiological Properties. Cham: Springer International Publishing. 2020:17–36. https://doi.org/10.1007/978-3-030-53421-9_2
22. Lerouge P, Carrier M, Mollet J-C, Lehner A. The cell wall pectic rhamnogalacturonan II, an enigma in plant glycobiology. 2021:553–71. <https://doi.org/10.1039/9781839164538-00553>
23. Sterling JD, Atmodjo MA, Inwood SE, Kumar Kolli VS, Quigley HF, Hahn MG, et al. Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proc Natl Acad Sci U S A*. 2006;103(13):5236–41. <https://doi.org/10.1073/pnas.0600120103> PMID: 16540543
24. Wu H-C, Bulgakov VP, Jinn T-L. Pectin Methylsterases: Cell Wall Remodeling Proteins Are Required for Plant Response to Heat Stress. *Front Plant Sci*. 2018;9:1612. <https://doi.org/10.3389/fpls.2018.01612> PMID: 30459794
25. Sepúlveda J, Rondón González F, Soto Sedano JC, Velasco GP, Mosquera T, Delgado MC, et al. SNP Analysis Reveals Novel Insights into the Genetic Diversity of Colombian *Vaccinium meridionale*. *Genes (Basel)*. 2025;16(6):675. <https://doi.org/10.3390/genes16060675> PMID: 40565567
26. AOAC. Official methods of analysis. 16th ed. Gaithersburg: Gaithersburg. 1995.
27. Wickham H. ggplot2. New York, NY: Springer New York. 2009. <https://doi.org/10.1007/978-0-387-98141-3>
28. Gross J, LU. Nortest: Tests for normality. 2015.
29. Wickham H, VD, G M. Tidy: Tidy messy data. 2024.
30. Wickham H, F R, H L, M K, V D. dplyr: A grammar of data manipulation. 2023.
31. Slowikowski K. Ggrepel: Automatically position non-overlapping text labels with “ggplot2”. 2024.
32. Müller K, W H. Tibble: Simple data frames. 2023.
33. Pedersen TL, S M. Ragg: Graphic devices based on AGG. 2024.
34. Colle M, Leisner CP, Wai CM, Ou S, Bird KA, Wang J, et al. Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *Gigascience*. 2019;8(3):giz012. <https://doi.org/10.1093/gigascience/giz012> PMID: 30715294
35. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*. 2015;6(8):925–9. <https://doi.org/10.1111/2041-210x.12382>
36. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9. <https://doi.org/10.1038/ng1847> PMID: 16862161
37. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006;38(2):203–8. <https://doi.org/10.1038/ng1702> PMID: 16380716
38. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet*. 2012;44(7):825–30. <https://doi.org/10.1038/ng.2314> PMID: 22706313
39. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*. 2010;42(4):355–60. <https://doi.org/10.1038/ng.546> PMID: 20208535
40. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genet*. 2016;12(2):e1005767. <https://doi.org/10.1371/journal.pgen.1005767> PMID: 26828793
41. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*. 2019;8(2):giy154. <https://doi.org/10.1093/gigascience/giy154> PMID: 30535326
42. Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z. A super powerful method for genome wide association study. *PLoS One*. 2014;9:e107684. <https://doi.org/10.1371/journal.pone.0107684>
43. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. 2014. <https://doi.org/10.1101/005165>
44. Chevalier LM, Rioux L-E, Angers P, Turgeon SL. Low-Temperature Blanching as a Tool to Modulate the Structure of Pectin in Blueberry Purees. *J Food Sci*. 2017;82(9):2070–7. <https://doi.org/10.1111/1750-3841.13826> PMID: 28796321

45. Aura A-M, Holopainen-Mantila U, Sibakov J, Kössö T, Morkkila M, Kaisa P. Bilberry and bilberry press cake as sources of dietary fibre. *Food Nutr Res*. 2015;59:28367. <https://doi.org/10.3402/fnr.v59.28367> PMID: [26652738](#)
46. Li BW, Andrews KW, Pehrsson PR. Individual sugars, soluble, and insoluble dietary fiber contents of 70 high consumption foods. *Journal of Food Composition and Analysis*. 2002;15:715–23. <https://doi.org/10.1006/jfca.2002.1096>
47. Guan Z-W, Yu E-Z, Feng Q. Soluble Dietary Fiber, One of the Most Important Nutrients for the Gut Microbiota. *Molecules*. 2021;26(22):6802. <https://doi.org/10.3390/molecules26226802> PMID: [34833893](#)
48. Li M, Ma S. A review of healthy role of dietary fiber in modulating chronic diseases. *Food Res Int*. 2024;191:114682. <https://doi.org/10.1016/j.foodres.2024.114682> PMID: [39059940](#)
49. Matsumoto GO, Garcia A, Benevenuto J, Munoz PR. High-density linkage map and QTL analyses for fruit quality traits in the wild blueberry relative *Vaccinium stamineum*. G3 (Bethesda). 2026;16(1):jkaf263. <https://doi.org/10.1093/g3journal/jkaf263> PMID: [41223021](#)
50. Deng H, Wang X, Wang Y, Xiang Y, Chen M, Zhang H, et al. Dynamic Changes in Cell Wall Polysaccharides during Fruit Development and Ripening of Two Contrasting Loquat Cultivars and Associated Molecular Mechanisms. *Foods*. 2023;12(2):309. <https://doi.org/10.3390/foods12020309> PMID: [36673402](#)
51. Persson S, Wei H, Milne J, Page GP, Somerville CR. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc Natl Acad Sci U S A*. 2005;102(24):8633–8. <https://doi.org/10.1073/pnas.0503392102> PMID: [15932943](#)
52. Fangel JU, Petersen BL, Jensen NB, Willats WGT, Bacic A, Egelund J. A putative Arabidopsis thaliana glycosyltransferase, At4g01220, which is closely related to three plant cell wall-specific xylosyltransferases, is differentially expressed spatially and temporally. *Plant Sci*. 2011;180(3):470–9. <https://doi.org/10.1016/j.plantsci.2010.11.002> PMID: [21421394](#)
53. Feng Y, Jin Q, Liu X, Lin T, Johnson A, Huang H. Advances in understanding dietary fiber: Classification, structural characterization, modification, and gut microbiome interactions. *Compr Rev Food Sci Food Saf*. 2025;24(1):e70092. <https://doi.org/10.1111/1541-4337.70092> PMID: [39840651](#)
54. Wang H, Feng X, Zhang Y, Wei D, Zhang Y, Jin Q, et al. PbUGT72AJ2-Mediated Glycosylation Plays an Important Role in Lignin Formation and Stone Cell Development in Pears (*Pyrus bretschneideri*). *Int J Mol Sci*. 2022;23(14):7893. <https://doi.org/10.3390/ijms23147893> PMID: [35887241](#)
55. Stratilová B, Kozmon S, Stratilová E, Hrmova M. Plant Xyloglucan Xyloglucosyl Transferases and the Cell Wall Structure: Subtle but Significant. *Molecules*. 2020;25(23):5619. <https://doi.org/10.3390/molecules25235619> PMID: [33260399](#)
56. Vain T, Crowell EF, Timpano H, Biot E, Desprez T, Mansoori N, et al. The Cellulase KORRIGAN Is Part of the Cellulose Synthase Complex. *Plant Physiol*. 2014;165(4):1521–32. <https://doi.org/10.1104/pp.114.241216> PMID: [24948829](#)
57. Yuan W, Yao F, Liu Y, Xiao H, Sun S, Jiang C, et al. Identification of the xyloglucan endotransglycosylase/hydrolase genes and the role of PagXTH12 in drought resistance in poplar. *For Res (Fayettev)*. 2024;4:e039. <https://doi.org/10.48130/forres-0024-0036> PMID: [40027451](#)
58. Wang Y, Liu X, Zhao T, Li J, Chen L, Li Y, et al. Genome-wide characterization of the UDP-glycosyltransferases (UGT) family and functional analysis of VcUGT160 involved in dihydrozeatin glycosylation during blueberry fruits development. *BMC Genomics*. 2025;26(1):1044. <https://doi.org/10.1186/s12864-025-12267-5> PMID: [41239213](#)
59. Di Matteo A, Giovane A, Raiola A, Camardella L, Bonivento D, De Lorenzo G, et al. Structural basis for the interaction between pectin methyl-esterase and a specific inhibitor protein. *Plant Cell*. 2005;17(3):849–58. <https://doi.org/10.1105/tpc.104.028886> PMID: [15722470](#)
60. Huang W, Shi Y, Yan H, Wang H, Wu D, Grierson D, et al. The calcium-mediated homogalacturonan pectin complexation in cell walls contributes the firmness increase in loquat fruit during postharvest storage. *J Adv Res*. 2023;49:47–62. <https://doi.org/10.1016/j.jare.2022.09.009> PMID: [36198382](#)
61. Yüksel E, Kort R, Voragen AGJ. Structure and degradation dynamics of dietary pectin. *Crit Rev Food Sci Nutr*. 2025;65(29):6249–68. <https://doi.org/10.1080/10408398.2024.2437573> PMID: [39681562](#)
62. Jolie RP, Duvetter T, Van Loey AM, Hendrickx ME. Pectin methyl-esterase and its proteinaceous inhibitor: a review. *Carbohydr Res*. 2010;345(18):2583–95. <https://doi.org/10.1016/j.carres.2010.10.002> PMID: [21047623](#)
63. Anderson CT, Pelloux J. The Dynamics, Degradation, and Afterlives of Pectins: Influences on Cell Wall Assembly and Structure, Plant Development and Physiology, Agronomy, and Biotechnology. *Annu Rev Plant Biol*. 2025;76(1):85–113. <https://doi.org/10.1146/annurev-arplant-083023-034055> PMID: [39841930](#)
64. Chandel V, Biswas D, Roy S, Vaidya D, Verma A, Gupta A. Current advancements in pectin: extraction, properties and multifunctional applications. *Foods*. 2022;11:2683. <https://doi.org/10.3390/foods11172683>
65. Hyodo H, Terao A, Furukawa J, Sakamoto N, Yurimoto H, Satoh S, et al. Tissue specific localization of pectin-Ca²⁺ cross-linkages and pectin methyl-esterification during fruit ripening in tomato (*Solanum lycopersicum*). *PLoS One*. 2013;8(11):e78949. <https://doi.org/10.1371/journal.pone.0078949> PMID: [24236073](#)
66. Lara-Espinoza C, Carvajal-Millán E, Balandrán-Quintana R, López-Franco Y, Rascón-Chu A. Pectin and Pectin-Based Composite Materials: Beyond Food Texture. *Molecules*. 2018;23(4):942. <https://doi.org/10.3390/molecules23040942> PMID: [29670040](#)
67. Jia K, Wang W, Zhang Q, Jia W. Cell Wall Integrity Signaling in Fruit Ripening. *Int J Mol Sci*. 2023;24(4):4054. <https://doi.org/10.3390/ijms24044054> PMID: [36835462](#)

68. Steinwand BJ, Kieber JJ. The role of receptor-like kinases in regulating cell wall function. *Plant Physiol.* 2010;153(2):479–84. <https://doi.org/10.1104/pp.110.155887> PMID: [20410434](https://pubmed.ncbi.nlm.nih.gov/20410434/)
69. Liu C, Yu H, Voxeur A, Rao X, Dixon RA. FERONIA and wall-associated kinases coordinate defense induced by lignin modification in plant cell walls. *Sci Adv.* 2023;9(10):eadf7714. <https://doi.org/10.1126/sciadv.adf7714> PMID: [36897948](https://pubmed.ncbi.nlm.nih.gov/36897948/)
70. Yao X, Humphries J, Johnson KL, Chen J, Ma Y. Function of WAKs in Regulating Cell Wall Development and Responses to Abiotic Stress. *Plants (Basel).* 2025;14(3):343. <https://doi.org/10.3390/plants14030343> PMID: [39942905](https://pubmed.ncbi.nlm.nih.gov/39942905/)