BMC Plant Biology

RESEARCH

Open Access

Population structure, genetic diversity, and GWAS analyses with GBS-derived SNPs and silicodart markers unveil genetic potential for breeding and candidate genes for agronomic and root quality traits in an international sugar beet germplasm collection

Noor Maiwan Bahjat[1], Mehtap Yıldız[1*], Muhammad Azhar Nadeem[2,3], Andres Morales[4,5], Josefina Wohlfeiler[6], Faheem Shahzad Baloch[2,7], Murat Tunçtürk[8], Metin Koçak[1], Yong Suk Chung[7], Dariusz Grzebelus[9], Gökhan Sadik[1], Cansu Kuzğun[1] and Pablo Federico Cavagnaro[6,9*]

Abstract

Background Knowledge about the degree of genetic diversity and population structure is crucial as it facilitates novel variations that can be used in breeding programs. Similarly, genome-wide association studies (GWAS) can reveal candidate genes controlling traits of interest. Sugar beet is a major industrial crops worldwide, generating 20% of the world's total sugar production. In this work, using genotyping by sequencing (GBS)-derived SNP and silicoDArT markers, we present new insights into the genetic structure and level of genetic diversity in an international sugar beet germplasm (94 accessions from 16 countries). We also performed GWAS to identify candidate genes for agriculturally-relevant traits.

Results After applying various filtering criteria, a total of 4,609 high-quality non-redundant SNPs and 6,950 silicoDArT markers were used for genetic analyses. Calculation of various diversity indices using the SNP (e.g., mean gene diversity: 0.31, MAF: 0.22) and silicoDArT (mean gene diversity: 0.21, MAF: 0.12) data sets revealed the existence of a good level of conserved genetic diversity. Cluster analysis by UPGMA revealed three and two distinct clusters for SNP and DArT data, respectively, with accessions being grouped in general agreement with their geographical origins and their tap root color. Coincidently, structure analysis indicated three (K = 3) and two (K = 2) subpopulations for SNP and DArT data, respectively, with accessions in each subpopulation sharing similar geographic origins and root color; and comparable clustering patterns were also found by principal component analysis. GWAS on 13 root and

*Correspondence:
Mehtap Yıldız
mehtapyildiz@gmail.com
Pablo Federico Cavagnaro
cavagnaro.pablo@inta.gob.ar

Full list of author information is available at the end of the article

leaf phenotypic traits allowed the identification of 35 significant marker-trait associations for nine traits and, based on predicted functions of the genes in the genomic regions surrounding the significant markers, 25 candidate genes were identified for four root (fresh weight, width, length, and color) and three leaf traits (shape, blade color, and veins color).

Conclusions The present work unveiled conserved genetic diversity–evidenced both genetically (by SNP and silicoDArT markers) and phenotypically- exploitable in breeding programs and germplasm curation of sugar beet. Results from GWAS and candidate gene analyses provide a frame work for future studies aiming at deciphering the genetic basis underlying relevant traits for sugar beet and related crop types within *Beta vulgaris* subsp. *vulgaris*.

Keywords *Beta vul*garis, Genetic diversity, Genotyping by sequencing, Germplasm characterization, GWAS, Candidate genes

Introduction

Next-generation sequencing (NGS) methods have facilitated new research opportunities to study plants with and without a reference genome. NGS technologies differ from the traditional Sanger sequencing method in terms of its massively parallel analysis, high throughputness, and cheaper cost. Although NGS increases access to genome sequences, the downstream data analysis and biological interpretations continue to be a challenge to comprehending genomes. Nucleotide variance profiling and extensive genetic marker analysis are made possible by the rapid and affordable production of large amounts of sequence data using NGS [1]. From sequencing library generation to modern bioinformatics methods for post-sequencing procedures, the production, assembly, and analysis of these sequence reads require a variety of experimental approaches, depending on the type of NGS application used [2].

Diversity Array Technology (DArT) is a sequence-independent genotyping method that generates genome-wide fingerprints. DArT is a hybridization-based microarray technique that is capable of measuring thousands of DNA fragments to hybridization arrays with high levels of multiplexing [3]. This widely used technology has the advantage of yielding a high level of polymorphism and no prior sequence information about the genome is required. The main applications of DArT have been for the construction of genetic linkage maps, and QTL and genetic diversity analyses in numerous crops, including wheat [4], barley [5], common bean [6], and safflower [7], among others. Over the last decade, DArT has generated two types of markers; silicoDArT and single nucleotide polymorphism (SNP). SilicoDArT are microarray markers that are dominant and scored for the presence or absence of a single allele [8]. DArTseq-based SNPs are co-dominant markers. Both types of markers have been effectively used to study population structure, genetic diversity, genome wide associations, and genetic mapping in several crop species, including maize [9], common bean [10], cassava [11], and Barbara groundnut [12].

SNP is a one-base variation in a single DNA nucleotide that occurs at a specific position in the genome. SNPs represent useful markers in plant genetic research because they reflect both genetic diversity that occurs naturally and genetic drift that occurs as a result of plant breeding [13, 14]. Thus, due to their advantages concerning high-throughput detection and simple integration of genotypic data, SNP markers have been widely used to create DNA fingerprints of germplasm collections in several crops, including maize [15], wheat [16], melon [17], and pepper [18].

Genome-wide association studies (GWAS) coupled with NGS data analyze thousands of genetic variants–most commonly, SNPs- across many genomes to find those statistically associated with a specific trait. This methodology overcomes several limitations of the traditional gene and quantitative trait loci (QTL) mapping strategy used in plants, characterized by the used of biparental crosses and yielding limited allelic diversity and genomic resolution [19]. Thus, GWAS allows the utilization of phenotypically-characterized germplasm collections, without the need for structured populations (e.g., F2 or F3 progenies), providing both greater allelic diversity and mapping resolution, sometimes to the gene level [20]. This methodology has been successful in identifying QTL that explain large portions of the phenotypic variation for many plant traits. Once a phenotype-marker association is identified, searches for candidate genes in the nearby region, followed by their expression analysis in plants with contrasting phenotypes for a given trait, may provide valuable insight into a phenotype's underlying biology.

Beet, botanically known as *Beta vulgaris* L. (2n = 18), belongs to the *Amaranthaceae* family, sub-family Betoideae, and Caryophyllales order. *Beta vulgaris* subsp. *vulgaris* represents a species complex composed of different crop types, including sugar beet [21], a major industrial source of table sugar; the leafy vegetable known as chard, Swiss chard, or spinach beet [22]; the root vegetable commonly referred to as garden beet or beetroot [23]; and mangelwurzel or fodder beet, used as a fodder crop [24].

All of these crop types are sexually compatible and therefore inter-crossable. Sugar beet is a biennial plant that forms its tap root in the first year and produces flowers in the second. In Europe, sugar beet competes with sugarcane for the manufacture of sugar and ethanol. Additionally, the crop is rich in carbohydrates that are utilized as a source of feed for nursing animals like cattle, and the pulp may be used as an adhesive in both cosmetic products and printing ink [25].

The modern sugar beet was developed in the middle of the eighteenth century as a selection from mangelwurzel (fodder beets) in Central Europe, but its utilization for food and medicinal purposes has increased significantly in recent decades [26]. The cultivation of sugar beet is mostly concentrated within the latitudinal ranges of 30–60 degrees north and 25–35 degrees south. This geographical distribution is attributed to the physiological adaptability of the crop to temperate conditions. Today, approximately 20% of the world's sugar is produced from sugar beet, while the rest, 80%, is produced from sugarcane. Currently, the crop is commercially cultivated in more than 50 countries and Europe is considered the main sugar beet producer [27]. Northern Europe is the main region of sugar beet production [28], while the Russian Federation is the world's top producer, with 41.2 thousand metric tons per year.

Currently, the world is facing several issues, including climate change and a rapidly increasing population. Climate change represents a major threat to agriculture production, causing various biotic and abiotic stresses and ultimately lowering crop yield [29]. Besides climate change, the human population is increasing and demanding the availability of more food for the present and upcoming generations. The available commercial cultivars are prone to various biotic and abiotic stresses, evidencing the importance of new cultivar development. All of the issues discussed above can be addressed through harnessing crop genetic diversity [30]. Germplasm characterization provides detailed insight into genetic diversity and allows the identification of new variants that can be used in breeding programs [31, 32]. Also, the identification of loci that cosegregate with specific traits of interest–e.g., by means of GWAS or linkage mapping- is a pre-requisite for implementing marker-assisted breeding strategies [20].

Several genetic marker systems have been used to analyze the genetic structure in *B. vulgaris* subsp. *vulgaris* and some of its wild relatives [21, 33–36]. However, these studies used relatively few accessions and/or number of markers. Thus, the evaluation of a broad germplasm collection from diverse geographical origins using a large number of informative markers may provide a more robust assessment of the genetic diversity and population structure in this species. Also, marker-trait association analysis for agronomically important traits may reveal QTL and candidate genes for traits such as root biomass yield and sugar content, which are highly relevant for sugar beet production. Other traits of interest for beet breeders are root shape and pigment content. In the case of chard, leaf-related traits (e.g., leaf size, shape, and color) are of major importance. The inter-crossability nature of different crop types within *Beta vulgaris* subsp. *vulgaris* makes it possible to introgress superior variants of such phenotypic traits from one another.

In this study, a collection of 94 phenotypically different accessions of *Beta vulgaris* L. subsp. *vulgaris* from diverse geographical origins were used to characterize genetic diversity and population structure in this taxon composed mostly of sugar beet materials, using GBS-derived SNP and silicoDArT markers. In addition, the SNP markers data set was used for GWAS and candidate gene analyses on 13 root and leaf phenotypic traits, of importance for sugar beet and other crop types within *B. vulgaris* subsp. *vulgaris*. The resulting data provide novel information concerning the genetic diversity, population structure, and genetic relationships among the accessions in this taxonomic group, and reveal QTL and candidate genes for nine and seven phenotypic traits, respectively. Altogether, the reported information will serve as the foundation for future genetic studies and will be instrumental in breeding programs of this species.

Results

In total, 41,080 raw SNP and 45,063 silicoDArT markers were called. SNPs had a substantially greater amount of missing data (823,874; 21.3%) than the SilicoDArT markers (224,515; 5.3%). All the called markers were polymorphic, of which 12,497 SNPs (30.4%) and 25,394 silicoDArTs (56.4%) were present in at least one individual of every population, considering the geographical origins of the accessions as the populations. The latter subset of markers werefurther selected while the remaining ones (i.e., those with missing data in all the individuals of any population) were discarded. This criterion, which makes marker selection more stringent, resulted in fewer available markers for downstream analyses but allowed more robust estimates of genetic diversity and population structure parameters, as an excess of missing data are known to negatively influence such estimates. Within the latter set of markers, 8,676 SNPs (69.4%) and 24,629 silicoDArTs (97.0%) had repeatability of 100%, and, of those, 8,676 (100%) and 19,614 (79.6%) presented call rates > 95% and were further selected. From the latter set of markers, we kept those with MAF > 5%, resulting in a total of 4,609 SNP and 6,950 silicoDArT markers that were used for downstream genetic analyses in the sugar beet germplasm.

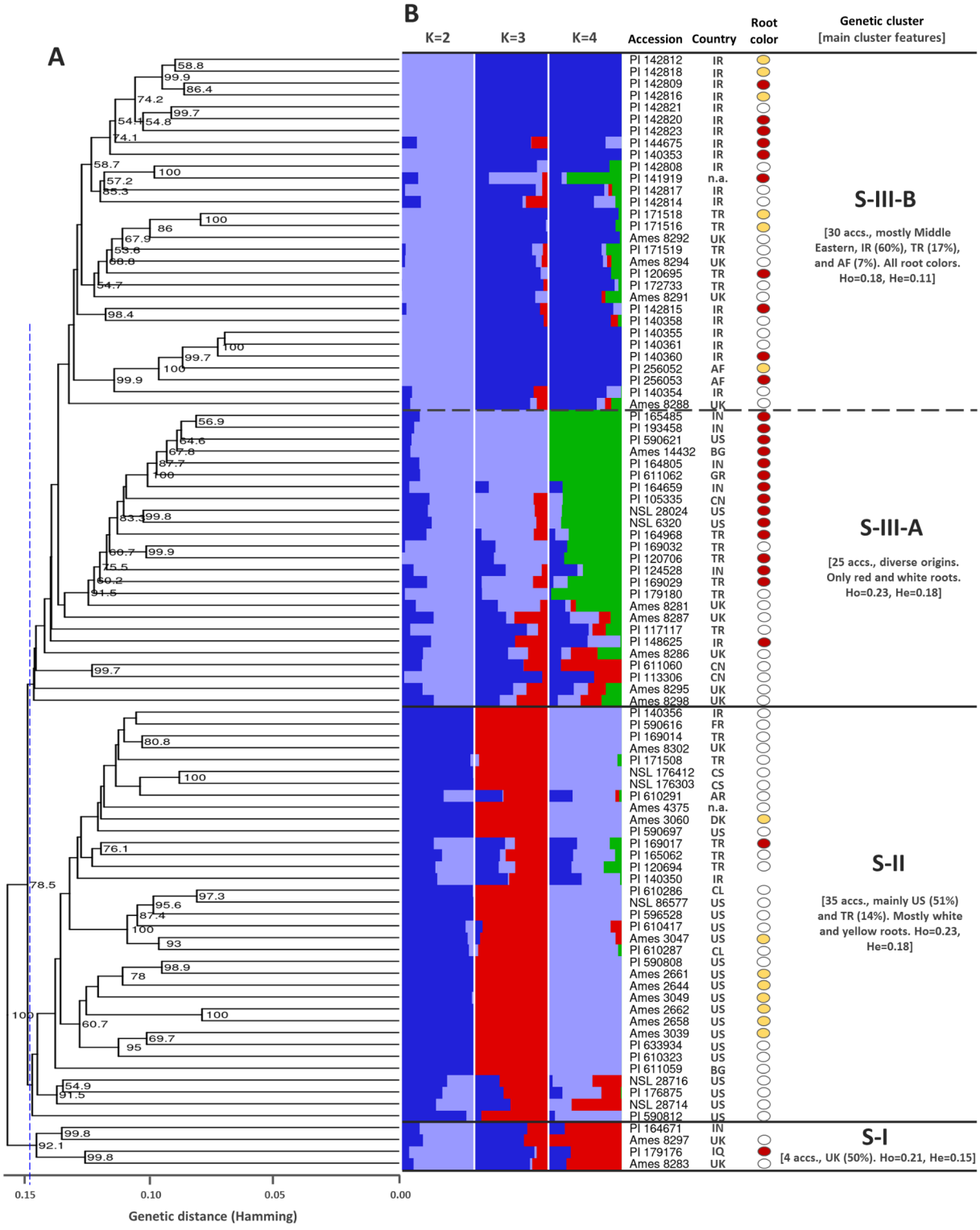

Fig. 1 (See legend on next page.)

Polymorphism of SNP and silicodart markers in the sugar beet germplasm

In the SNP markers dataset, a total of 9,218 alleles were identified, indicating that all loci were bi-allelic. The SNPs average genetic diversity, as estimated by Nei's genetic diversity index (NGD), was 0.31 (range = 0.09–0.50), their mean polymorphic information content (PIC) was 0.25 (range = 0.09–0.38), and their mean minor allele

(See figure on previous page.)

Fig. 1 Genetic relationships and population structure for 94 sugar beet accessions based on 4,609 SNP markers. (**A**) UPGMA dendrogram based Hamming genetic distance (GD). Clusters S-I, S-II, and S-III, indicate the groups revealed by clustering the accessions at a GD < 0.145 (indicated by the vertical dashed blue line). Cluster S-III was further classified in two sub-clusters, S-III-A and S-III-B. Horizontal bold and dashed lines separate genetic clusters and sub-clusters, respectively. Branch support (> 50%) is based on 1000 bootstrap replications and shown as a percentage. (**B**) Genetic structure of the sugar beet accessions considering different optimal number of populations (i.e., K = 2, K = 3, and K = 4). Each accession is represented by a horizontal bar partitioned into two (K = 2), three (K = 3), or four-colored segments (K = 4), indicating their relative membership to the considered clusters. Root external colors -white, yellow, and reddish purple- are indicated by colored circles on the right of the accession's country of origin. Results from *post hoc* analyses of the optimal K, testing K values from 1 to 12, are presented in Supplementary Fig. S1. ISO country codes are as follows: AF: Afghanistan; AR: Argentina; BG: Bulgaria; CL: Chile; CN: China; CS: Serbia and Montenegro; DK: Denmark; FR: France; UK: United Kingdom; GR: Greece; IN: India; IQ: Iraq; IR: Iran; ET: Ethiopia; TR: Turkey; US: United States of America. For two accessions, information on the country of origin was 'not available' (n.a.). Main features of each cluster are described. Ho: observed heterozygocity; He: expected heterozygocity

frequency (MAF) was 0.22 (range = 0.05–0.50). Also, their mean observed (H_o) and expected heterozygosity (H_e) was 0.21 and 0.15, respectively, whereas the mean inbreeding coefficient (F_i) was 0.30. For silicoDArT markers, analyzed globally in the sugar beet collection, the total number of different alleles was 13,900, also indicating a bi-allelic status for all the loci. We found lower mean NGD for these markers (NGD = 0.21) as compared to SNPs. SilicoDArTs also had lower mean PIC (0.18, range = 0.09–0.37), MAF (0.12, range = 0.05–0.48), and F_i values (-0.20, range = -0.73-0.16) than SNPs, but greater mean H_o (0.25) and H_e (0.30).

Genetic diversity, population structure, and principal coordinate analyses based on GBS-derived SNP markers

Pairwise genetic distances among the accessions, as estimated by Hamming distance (HGD), ranged from 0.145 to 0.363, with an average of 0.279. The fact that none of the pair-wise comparisons had a HGD value of zero indicates the absence of duplicate accessions (i.e., synonymies) in this collection. The two most genetically dissimilar accessions were Ames 2644 and Ames 8297 (HGD = 0.362), from the United States and United Kingdom, respectively; whereas the most genetically similar ones were PI 140,361, PI 140,355, and PI 140,360 (all with pairwise HGD values of 0.145), all of them from Iran. Based on the accessions pair-wise HGD values, a UPGMA dendrogram was constructed to visualize genetic relationships among the sugar beet accessions (Fig. 1A). By clustering the accessions with HGD values of 0.145 or less, three major clusters were revealed for these SNP data, namely S-I, S-II, and S-III. Cluster S-III was further classified into two subgroups, S-III-A and S-III-B, based on evident separation of the latter groups. A general association was found between the genetic clusters and the countries of origin of the accessions. Cluster S-I was the smallest group of the dendrogram, composed of four accessions, two of them from UK (50%) and the rest were from Iraq and India. Cluster S-II comprised 35 accessions, the majority of which were from the USA (18 accessions, representing ~ 51% of the total taxa included in the cluster), followed by five accessions from Turkey (14%), while the remaining accessions belonged to seven countries, with little contribution each to the overall geographic representation. Cluster S-III-A was composed of 25 accessions from diverse origins, including (in decreasing order) Turkey (6 accessions), India (5), the UK (5), the USA (3), and China (3), among others with fewer representatives. Cluster S-III-B comprised 30 accessions, with a vast majority of materials from the Middle East and surrounding regions, predominating from Iran (18 accessions representing 60% of the taxa), Turkey (5 accessions, 17%), and Afghanistan (2 accessions, ~ 7%).

A rather clear association was also found between genetic clusters and tap root color, as evidenced by the fact that 25 of the 27 accessions (92.6%) with reddish purple root were included in cluster S-III, whereas only one accession (3.7%) with this root color was included in clusters I and II (Fig. 1B). Within cluster S-III, subcluster S-III-A was particularly rich in reddish-purple rooted materials, with 60% of the accessions exhibiting this color phenotype, while in cluster S-III-B red beets accounted for 37% of the accessions. Conversely, red beets represented only 2.8% and 25% of the accessions in clusters S-II and S-I, respectively. In contrast, the white and yellow root materials largely predominated in clusters S-II and S-III-A, representing these color phenotypes, combined, 95% and 60% of all the accessions in these clusters, respectively. Interestingly, yellow-rooted materials were only present in the latter two clusters.

Genetic structure in the sugar beet germplasm was estimated using the SNP data set. Analysis of the optimal number of populations indicated three clusters (K = 3) as most probable, considering both the mean likelihood and ΔK methods (Supplementary Fig. S1). Figure 1B depicts the clustering results considering two (K = 2), three (K = 3), and four (K = 4) subpopulations in this germplasm collection. The resulting clusters from the STRUCTURE analysis, considering either K = 3 or K = 4, fully coincided with the clustering results of UPGMA, as shown in Figs. 1A and B.

Principal coordinates analysis (PCoA) using the SNP markers captured 13.5% of the total variation in the first two principal coordinates and grouped the accessions in general agreement with the clusters obtained by UPGMA and STRUCTURE analyses. As shown in Fig. 2A, three

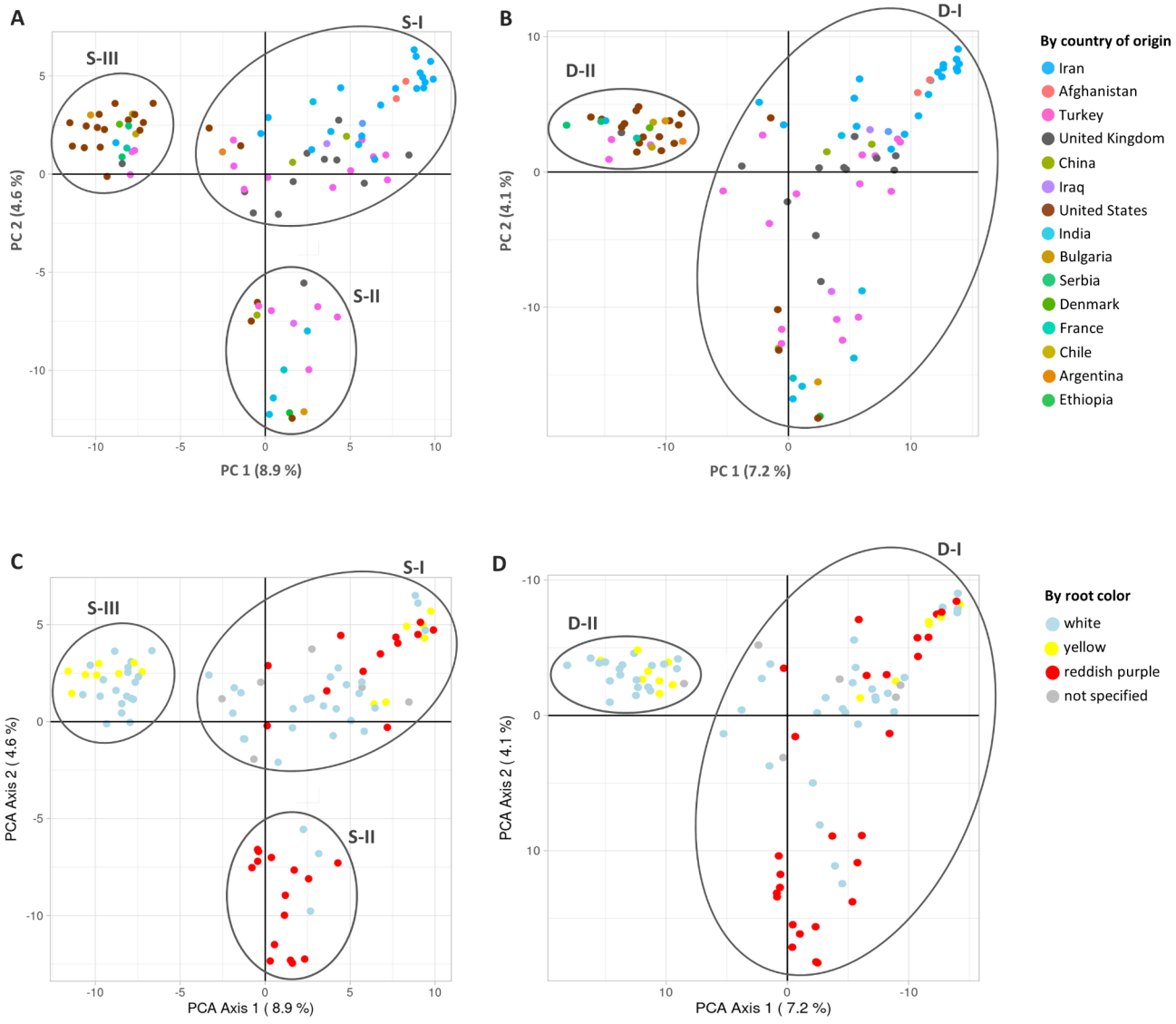

Fig. 2 Genetic clustering of 94 sugar beet accessions by country of origin (**A**, **B**) and root color (**C**, **D**) based on principal coordinate analysis of 4,609 SNPs (**A**, **C**) and 6,950 silicoDArT markers (**B**, **D**)

major groups could be identified. The first -and largest- group of accessions, denominated S-I and located mostly in the upper-right quadrant, comprised most of the Middle Eastern materials, including 19 accessions from Iran, 10 from Turkey, two from Afghanistan, and one from Iraq. In addition, this cluster also included most of the materials from the UK (10 accessions). Another group of materials, S-II, composed of 18 accessions of mixed origins, could be identified in the center and lower-right quadrant of the biplot, and included sugar beets from Turkey (4), India (4), the US (2), and various other geographical origins with less than two accession per country. Lastly, a distinct cluster, S-III, was revealed due to its clear separation from the other clusters, appearing in the far left side of the upper-left quadrant. This cluster comprised 26 accessions and included most of the accessions from the US (15), and a few materials from diverse European origins. Altogether, a strong correspondence was found between the SNP-based PCoA clusters S-I (predominantly Middle Eastern materials), S-II (materials of mixed origins), and S-III (predominantly from the US) with the UPGMA clusters S-III-B, S-III-A, and S-II, respectively (Figs. 1 and 2A). Also, the PCoA clusters showed a clear association with the root color phenotypes, as indicated by the fact that all the accessions with reddish purple root were included in clusters S-I and S-II (corresponding to the UPGMA clusters S-III-B and S-III-A, respectively) whereas cluster S-III contained only accessions with white and yellow root (in correspondence with the UPGMA cluster S-II) (Fig. 2C).

Results from AMOVA and general genetic diversity statistics for the SNP-based UPGMA clusters are

presented in Table 1, and they showed that most of the genetic variation found was among the accessions (i.e., within clusters), representing 71.3% of the total variation, rather than among clusters, which accounted for 28.7% of the total variation. Pair-wise genetic differentiation (F_{ST}) values among the UPGMA clusters revealed that S-II and S-III-A were genetically most similar (F_{ST} = 0.017), whereas S-I and S-III-B were the most different pair of clusters (F_{ST} = 0.037). Supplementary Table S1 presents a similar AMOVA analysis considering the geographical origins of the accessions (instead of UPGMA clusters), indicating that most of the variation found was among the accessions (i.e., within countries) (83.6%) rather than among countries (16.4%). Additionally, pair-wise F_{ST} values among countries indicated that accessions from Afghanistan and Chile were genetically most similar (F_{ST} = 0.070) whereas accessions from Iraq and Iran showed the greatest differentiation level (F_{ST} = 0.164).

Genetic diversity, population structure, and principal coordinate analyses based on silicodart markers

Analysis of silicoDArT markers in the sugar beet collection revealed pairwise genetic distance (Hamming) values among the accessions in the range of 0.139–0.439, with an average of 0.339. The two most genetically dissimilar accessions were NSL 176,303 and PI 140,355 (HGD = 0.439), from Serbia and Montenegro and Iran, respectively, whereas the two most similar ones were Ames 2662 and Ames 2658 (HGD = 0.139), both from United States. The fact that all the silicoDArT-based pairwise GD values were > 0 confirmed the absence of duplicate materials in this collection.

Figure 3A presents a UPGMA dendrogram depicting relationships among the accessions, which was constructed from a matrix of pairwise HGD values estimated from silicoDArT data. By clustering the accessions with HGD ≤ 0.175, two major clusters were revealed, denominated D-I and D-II. Cluster D-I was composed of 34 accessions, the majority of them (18 accessions) from United States, which represent 54% of the taxa included in this cluster. The materials from the USA included in cluster D-I account for 90% of the total accessions from this country. The other taxa included in D-I have diverse European, Asian, or South American origins. Cluster D-II conforms the largest group with 60 accessions. Four subclusters could be identified within D-II, designated as A, B, C, and D. Subcluster A represents the most external branch of D-II, and includes six accessions from UK and the Middle East. Subcluster B is composed of 17 accessions from diverse origins, including (in decreasing order) Turkey (5 accessions), India (5), US (3), UK (2), and China (2). Subcluster C represents a small branch, composed of only two accessions from UK. Subcluster D is the largest subgroup, with 35 accessions, and exhibits a strong prevalence of Middle Eastern accessions, with a total of 27 taxa from this region (accounting for 77% of the plant materials in this subcluster), predominating accessions from Iran (18), Turkey (7), and Afghanistan (2).

The silicoDArT-base dendrogram evidenced a clear association between the resulting clusters and the tap root color of the accessions, with all the reddish-purple beets included in cluster D-II, whereas D-I contained only accessions with white and yellow roots (Fig. 3B). Within D-II, subcluster B was particularly rich in red beets, comprising this color phenotype ~ 88% of the accessions included in this subcluster, followed by subcluster D where red beets represented 31% of the accessions. In contrast, cluster D-I comprised most of the white-colored beets, representing this phenotype 73.5% of the total accessions in the cluster. Interestingly, yellow-rooted accessions were only present in cluster D-I and in subcluster D of D-II.

Genetic structure was estimated with silicoDArT data. To this end, different models using K from 2 to 12 were tested (Supplementary Fig. S2). Based on their mean

Table 1 Genetic diversity statistics and AMOVA for SNP data, by UPGMA clusters

Cluster	Number of accessions (%)	mean NGD	mean PIC	mean MAF	mean Ho	mean He	mean F_i	F_{ST} among clusters[#]		
								S-II	S-III-A	S-III-B
S-I	4 (4.3%)	0.30	0.24	0.21	0.18	0.11	0.40	0.023	0.025	0.037
S-II	35 (37.2%)	0.30	0.24	0.22	0.23	0.18	0.23		0.017	0.029
S-III-A	25 (26.6%)	0.30	0.24	0.22	0.23	0.18	0.22			0.031
S-III-B	30 (31.9%)	0.29	0.24	0.21	0.21	0.15	0.28			
AMOVA										
Source of variation										
Among clusters		28.7%								
Among accessions/within clusters		71.3%								
Overall FST		0.027								
Overall FIS		0.286								
Overall FIT		0.306								

NGD. Nei's genetic diversity index; PIC. Polymorphic index content; MAF. Minor allele frequency; Ho. Observed heterozygosity; He. Expected heterozygosity; Fi. Inbreeding coefficient. [#] Pair-wise genetic differentiation (F_{ST}) among UPGMA clusters

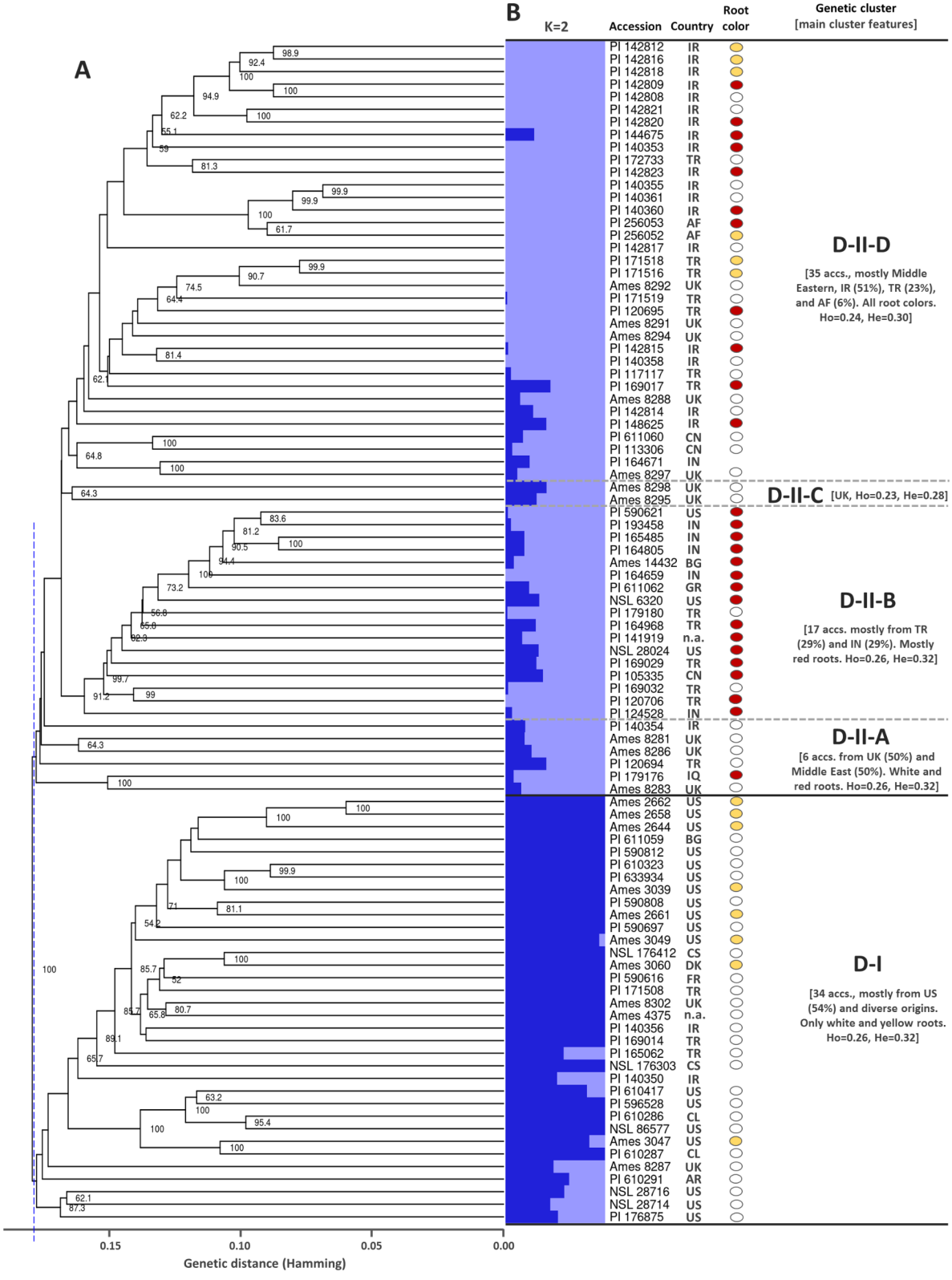

Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Genetic relationships and population structure for 94 sugar beet accessions based on 6,950 silicoDArT markers. (**A**) UPGMA dendrogram based Hamming genetic distance (GD). Clusters D-I and D-II grouped accessions with GD < 0.175 (indicated by the vertical dashed blue line). Horizontal black lines separate genetic clusters, and four subclusters within D-II are separated by gray dashed lines and indicated with gray letters A-D. Branch support (> 50%) is based on 1000 bootstrap replications and shown as a percentage. (**B**) Estimated genetic structure for the sugar beet collection. Each accession is represented by a horizontal bar partitioned into two colored segments (K = 2), indicating their relative membership to the two clusters. Root external colors -white, yellow, and reddish purple- are indicated by colored circles on the right of the accession's country of origin. Results from *post hoc* analyses of the optimal K, testing K values from 1 to 12, are presented in Supplementary Fig. S2. ISO country codes are as follows: AF: Afghanistan; AR: Argentina; BG: Bulgaria; CL: Chile; CN: China; CS: Serbia and Montenegro; DK: Denmark; FR: France; UK: United Kingdom; GR: Greece; IN: India; IQ: Iraq; IR: Iran; ET: Ethiopia; TR: Turkey; US: United States of America. For two accessions, information on the country of origin was 'not available' (n.a.). Main features of each cluster are described. Ho: observed heterozygocity; He: expected heterozygocity

likelihoods, indicating a maximum likelihood value for two populations in this collection (Supplementary Fig. S2 A), and the fact that running STRUCTURE using different models, considering two to five populations, consistently revealed only two clusters (Supplementary Fig. S2 C), we selected two populations (K = 2) as the most parsimonious model for this dataset, although ΔK analysis suggested K = 4 as optimal (Supplementary Fig. S2 B). Thus, the resulting clusters from the STRUCTURE analysis using K = 2 separated the accessions into two clusters, which coincided with clusters D-I and D-II of the UPGMA dendrogram (Fig. 3B).

PCoA using the silicoDArT data captured 11.3% of the total variation in the first two principal coordinates, and grouped the accessions in general agreement with the clusters obtained by UPGMA and STRUCTURE analyses. As shown in Fig. 2B, only two major groups could be clearly identified. The first and largest group, denominated D-I, included 66 accessions from diverse origins, located in the middle and the right quadrants of the biplot. This group included all but one of the Iranian accessions (20), which were mostly located in the upper-right quadrant, as well as most of the Turkish (15), all but one of the British accessions (11), and all the Indian (5) and Afghan (2) materials. The other group, D-II, was smaller (28 accessions), and it was entirely located in the upper-left quadrant, and clearly separated from D-I. It comprised most of the accessions from the US (15 accessions, representing 75% of the total taxa from this country), along with a few accessions from diverse European (8) or South American (3) origins. Altogether, a strong correspondence was found between the silicoDArT-based PCoA clusters D-I (predominantly Middle Eastern, British, and Indian materials) and D-II (predominantly from US) with the silicoDArT-based UPGMA clusters D-I and D-II, respectively (Figs. 2 and 3). Similarly, PCoA clustering revealed an association with the accession's root color phenotypes, coincidently with results from UPGMA clustering, as evidenced by the inclusion of all the red beets in cluster D-I, while D-II only contained white and yellow-rooted accessions (Fig. 2D).

Results from AMOVA and general genetic diversity statistics for the silicoDArT-based UPGMA clusters and subclusters are presented in Table 2. To this end, two models were considered, one with only the two major clusters D-I and D-II (model 1), and the other one considering D-I and the four subclusters within D-II (i.e., subclusters A, B, C, and D) (model 2). These data revealed that most of the genetic variation was among the accessions (i.e., within clusters), representing 74.5–77.6% of the total variation (depending on the model), rather than among clusters, which accounted for 22.4–25.5% of the total variation. Pair-wise F_{ST} values among the UPGMA clusters indicated a generally low genetic differentiation among the clusters, with D-I, D-II-A, and D-II-B showing virtually no differentiation (F_{ST} = -0.018 to -0.028), whereas D-II-C and D-II-D were the most genetically different pair of clusters (F_{ST} = 0.072). Supplementary Table S2 presents a similar AMOVA analysis considering the geographical origins of the accessions, indicating that most of the variation found was among the accessions (i.e., within countries) (83.5%) rather than among countries (16.5%). Additionally, pair-wise F_{ST} values among countries indicated that accessions from Serbia and Montenegro and Chile were genetically most similar (F_{ST} = 0.070) whereas accessions from Iraq and Iran showed the greatest differentiation level (F_{ST} = 0.120), coinciding with results obtained with SNP markers (Supplementary Table S1).

GWAS and candidate gene analyses

A total of 13 root and leaf phenotypic traits of agronomic and industrial relevance were analyzed by GWAS in the beet germplasm collection. In total, 35 significantly associated genomic regions were revealed for nine traits [namely, root fresh weight (RFW), root width (RW), root length (RL), root external color (REC), root sugar content (RSC), leaf length (LL), leaf shape (LS), leaf blade color (LBC), and leaf veins red color intensity (LVRCI)], whereas no association was found for the remaining four traits [root shape (RS), root length-width ratio (RLWR), root tip shape (RTS), and leaf margin undulations (LMU)] (Fig. 4; Table 3; Supplementary Table S3). Detailed results from searches for candidate genes within a 60 kb region spanning the SNP marker with strongest association with a trait (i.e., ± 30 kb on each side of the peak marker), for each associated region, are presented in Supplementary Table S3, whereas Table 3 shows summarized data with

Table 2 Genetic diversity statistics and AMOVA for silicodart data, by UPGMA clusters and subclusters

Model	Cluster	Number of accessions (%)	mean NGD	mean PIC	mean MAF	mean Ho	mean He	mean F_i	F_{ST} among clusters[#]				
									D-II	D-II-A	D-II-B	D-II-C	D-II-D
1	D-I	34 (36.2%)	0.21	0.18	0.12	0.25	0.30	-0.20	0.004				
	D-II	60 (63.8%)	0.21	0.18	0.13	0.25	0.30	-0.20					
2	D-I	34 (36.2%)	0.21	0.18	0.13	0.26	0.32	-0.22		-0.018	-0.028	0.032	0.019
	D-II-A	6 (6.3%)	0.21	0.18	0.13	0.26	0.32	-0.22			-0.025	0.035	0.021
	D-II-B	17 (18.1%)	0.22	0.18	0.13	0.26	0.32	-0.23				0.026	0.012
	D-II-C	2 (2.1%)	0.19	0.16	0.11	0.23	0.28	-0.20					0.072
	D-II-D	35 (37.2%)	0.20	0.17	0.12	0.24	0.30	-0.22					
AMOVA													
Source of variation			Model 1		Model 2								
Among clusters			22.4%		25.5%								
Among accessions/within clusters			77.6%		74.5%								
Overall FST			0.004		0.015								
Overall FIS			-0.204		-0.216								
Overall FIT			-0.198		-0.198								

Two models were considered: (1) with only the two major UPGMA clusters, D-I and D-II; (2) considering D-I and the four subclusters of D-II (i.e., A, B, C, and D). NGD. Nei's genetic diversity index; PIC. Polymorphic index content; MAF. Minor allele frequency; Ho. Observed heterozygosity; He. Expected heterozygosity; Fi. Inbreeding coefficient. [#] Pair-wise genetic differentiation (F_{ST}) among UPGMA clusters

selected candidates based on their predicted gene function. The selection of the ± 30 kb window for searching for candidate genes was based on linkage disequilibrium (LD) decay curves, as presented in Supplementary Fig. S3). Among the tap root traits potentially related with root biomass yield, three, four, and four significant associations with RFW, RL, and RW, respectively, were identified across six chromosomes. For RFW, three candidate genes were identified in the respective QTL regions, namely a putative U-box type E3 ubiquitin ligase gene, a midasin gene, and a polycomb group protein gene. The proteins encoded by these genes are involved in protein ubiquitination, ribosome biogenesis, and the formation of protein complexes that regulate the expression of other genes via histone modifications, respectively. The potential roles of these genes related to RFW are described in the Discussion section. For RW, four candidate genes were identified in two QTL regions, and they encoded a nudix hydrolase, a small polypeptide protein of the DEVIL family, and two glutathione S-transferases. Their gene products are involved in regulating cellular energy metabolism and hormone signaling, act as regulatory molecules that coordinate various plant processes, and modulate antioxidant responses, respectively; and they have been shown to influence root development and growth by different mechanisms (discussed in the following section). For RL, five candidate genes in two associated regions were identified, and comprised two putative expansin-like genes, which encode cell wall proteins that participate in cell growth and stress responses; a nudix hydrolase, involved in cell energy metabolism and hormone signaling; a DNA-directed RNA polymerase II, involved in the synthesis of protein-coding mRNAs and microRNAs; and a histone acetyltransferase, known to epigenetically regulate the expression of other genes. These genes have been shown to influence RL in other species (discussed below).

For root color, which reflects the absence (white) or presence of different betalain pigments, such as betaxanthins (yellow) or betacyanins (red-violet), a single marker-trait association was found in Chr. 4, and this region included a putative ABC transporter gene, known to be involved in the vacuolar transport and accumulation of secondary metabolites, such as anthocyanins, conditioning red or purple pigmentation in organs of various species. Its possible role as a candidate for betalain pigmentation in the tap root of beets is discussed below. For sugar content, the most important industrial trait in sugar beet, a single association was identified in Chr. 7, which harbored five genes, but no clear candidates could be identified for this trait, presumably because most of these genes encoded uncharacterized proteins (Supplementary Table S3).

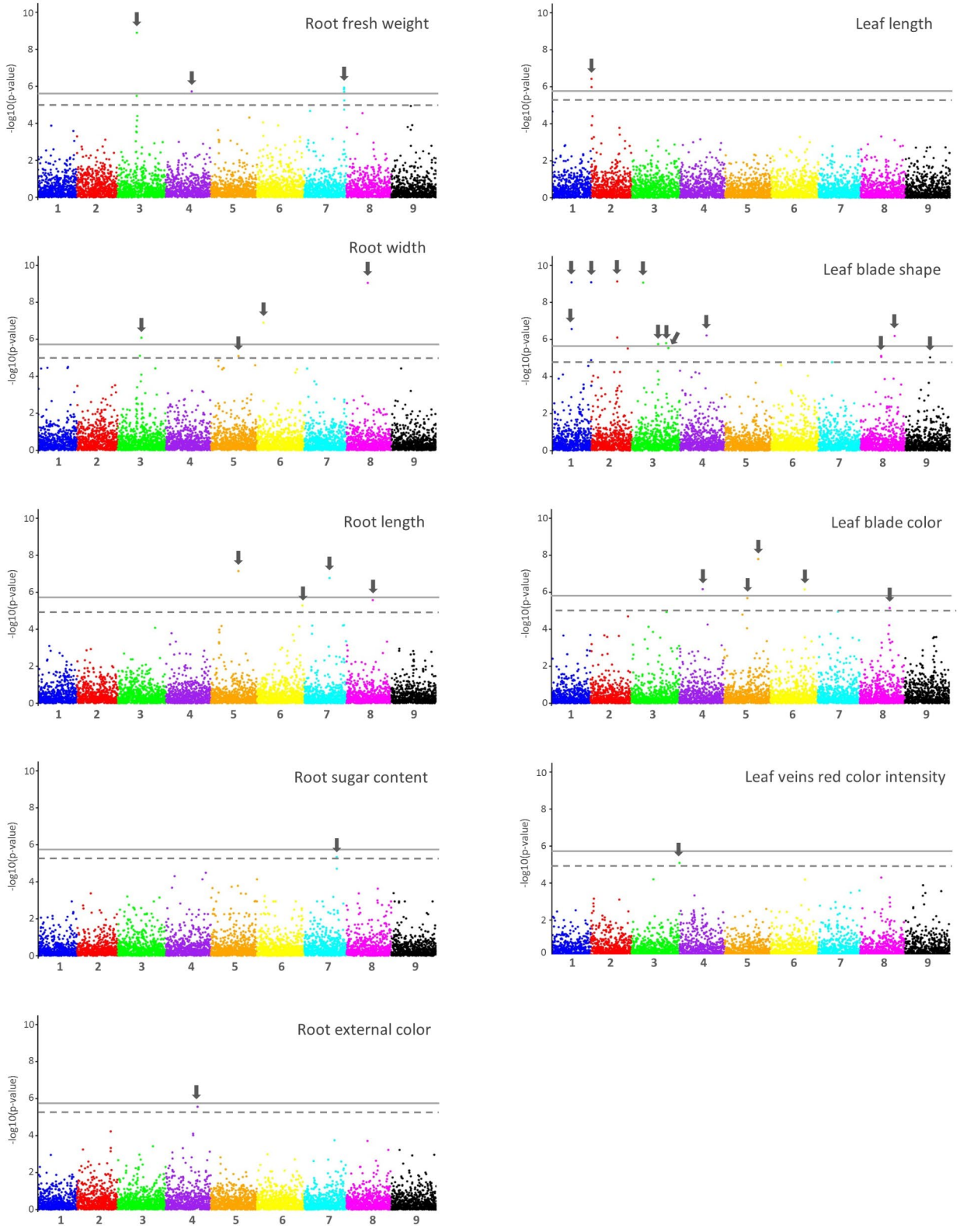

Fig. 4 Manhattan plots summarizing association mapping results for nine root and leaf phenotypic traits of sugar beet. The horizontal dashed and solid lines indicate significance thresholds based on the false discovery rate (FDR) and Bonferroni adjusted methods for $p = 0.05$, respectively. The peak SNP markers for genomic regions significantly associated with each trait, used for searching for candidate genes in a region spanning ± 30 kb from the peak marker, are indicated with gray arrows. Detailed information for these significant marker-trait associations is presented in Table 4 and Supplementary Table S3

Table 3 Summary of peak SNPs and candidate genes for GWAS analysis on nine phenotypic traits[a]

Trait	Peak signal						Region of search for candidate genes (peak SNP ± 30 kb)	Number of genes in region[a]	Candidate genes	Position of candidate gene (distance from peak SNP, in bp)	Annotation
	Chr.	Position	SNP genotype	F value	*p* value	-log10(*p*)					
Root fresh weight (RFW)	3	32,320,057	A/G	26.6	1.3E-09	8.9	32,290,057–32,350,057	3	LOC104883815	32,313,983–32,323,684 (0)	U-box domain-containing protein 14 (PUB14)
	7	7,356,347	G/A	16.2	1.2E-06	5.9	7,326,347–7,386,347	5	LOC104899776	7,309,211–7,360,747 (0)	midasin (MDN1)
	4	51,421,666	A/G	15.6	1.9E-06	5.7	51,391,666–51,451,666	7	LOC104890632	51,419,872–51,431,674 (0)	polycomb group protein EMBRYONIC FLOWER 2
Root width (RW)	8	3,861,150	T/A	27.2	8.9E-10	9.0	3,831,150–3,891,150	8	LOC104902151	3,860,143–3,871,244 (0)	nudix hydrolase 20
									LOC125492567	3,886,866–3,887,548 (+ 25716)	small polypeptide DEVIL 10
	6	22,238,837	C/A	19.4	1.3E-07	6.9	22,208,837–22,268,837	2	NA	-	-
	3	37,834,380	G/A	16.7	8.2E-07	6.1	37,804,380–37,864,380	5	LOC104906825	37,856,030–37,858,515 (+ 21650)	glutathione S-transferase U7
									LOC104906851	37,823,005–37,823,358 (-11375)	probable glutathione S-transferase
	5	4,716,543	C/A	13.6	8.0E-06	5.1	4,686,543–4,746,543	4	NA	-	-
Root length (RL)	5	47,076,179	G/A	20.2	7.2E-08	7.1	47,046,179–47,106,179	5	LOC104892825	47,053,192–47,056,677 (-19502)	expansin-like B1
									LOC104892824	47,062,172–47,071,170 (-5009)	expansin-like B1
									LOC125497584	47,064,547–47,067,347 (-8832)	nudix hydrolase 2-like
	7	44,799,149	T/C	19.0	1.7E-07	6.8	44,769,149–44,829,149	3	LOC104899323	44,794,810–44,795,503 (-3646)	DNA-directed RNA polymerase II subunit 4
									LOC104902093	44,797,232–44,807,922 (0)	histone acetyltransferase type B catalytic subunit
	8	4,798,785	G/A	15.1	2.7E-06	5.6	4,768,785–4,828,785	5	NA	-	-
	6	7,893,749	G/A	14.2	5.2E-06	5.3	7,863,749–7,923,749	4	NA	-	-
Root external color (REC)	4	55,080,253	A/C	15.0	2.8E-06	5.6	55,050,253–55,110,253	8	LOC104890386	55,063,756–55,068,952 (-11301)	ABC transporter F family member 5
Root sugar content (RSC)	7	54,256,064	A/T	20.9	4.6E-06	5.3	54,226,064–54,286,064	5	NA	-	-
Leaf length (LL)	2	10,036,510	A/T	17.8	3.8E-07	6.4	10,006,510–10,066,510	4	NA	-	-
	2	10,174,700	G/C	16.4	1.1E-06	6.0	10,144,700–10,204,700	1	NA	-	-
Leaf shape (LS)	1	9,908,448	T/A	27.3	8.2E-10	9.1	9,878,448–9,938,448	7	LOC104905284	9,945,774–9,948,227 (+ 27326)	homeobox-leucine zipper protein ATHB-4
	1	28,907,662	G/A	27.3	8.3E-10	9.1	28,877,662–28,937,662	2	NA	-	-
	2	48,345,608	G/A	27.5	7.4E-10	9.1	48,315,608–48,375,608	5	LOC104907732	48,299,440–48,310,628 (-34980)[b]	pentatricopeptide repeat-containing protein
	3	18,882,943	C/T	27.2	8.6E-10	9.1	18,852,943–18,912,943	4	NA	-	-
	1	29,039,062	T/A	18.3	2.7E-07	6.6	29,009,062–29,069,062	4	LOC104896300	29,046,791–29,048,788 (+ 7729)	3-ketoacyl-CoA synthase
	4	52,151,192	C/T	17.1	6.1E-07	6.2	52,121,192–52,181,192	5	LOC104890605	52,120,751–52,122,755 (-28437)	NAC domain-containing protein 82

Table 3 (continued)

Trait	Peak signal						Region of search for candidate genes (peak SNP ± 30 kb)	Number of genes in region[a]	Candidate genes	Position of candidate gene (distance from peak SNP, in bp)	Annotation
	Chr.	Position	SNP genotype	F value	*p* value	-log10(*p*)					
	8	54,895,227	C/A	17.0	6.5E-07	6.2	54,865,227–54,925,227	4	LOC104900640	54,916,471–54,917,525 (+21244)	ethylene-responsive transcription factor 12
									LOC104900641	54,885,509–54,899,925 (0)	receptor-like protein 4
	3	49,421,662	A/G	15.8	1.6E-06	5.8	49,391,662–49,451,662	6	NA	-	-
	3	41,661,861	C/A	15.6	1.8E-06	5.7	41,631,861–41,691,861	3	LOC104888714	41,667,607–41,670,067 (+5746)	receptor-like protein kinase 46
	2	54,744,919	A/G	14.9	3.1E-06	5.5	54,714,919–54,774,919	8	LOC104905962	54,719,887–54,727,449 (-17470)	DEAD-box ATP-dependent RNA helicase 18
	3	51,789,180	T/C	15.0	2.9E-06	5.5	51,759,180–51,819,180	5	LOC104889622	51,824,781–51,826,055 (+35601)[b]	calmodulin-binding protein 25
	8	36,635,339	T/C	13.6	7.8E-06	5.1	36,605,339–36,665,339	2	NA	-	-
	8	36,736,050	G/C	13.5	8.8E-06	5.1	36,706,050–36,766,050	2	NA	-	-
	9	31,605,120	C/T	13.4	9.3E-06	5.0	31,575,120–31,635,120	4	LOC104904740	31,592,859–31,598,730 (-6390)	protein SCARECROW
Leaf blade color (LBC)	5	54,764,697	T/C	22.5	1.6E-08	7.8	54,734,697–54,794,697	4	NA	-	-
	4	48,689,236	G/A	17.0	6.8E-07	6.2	48,659,236–48,719,236	5	NA	-	-
	6	58,608,595	A/G	16.9	7.0E-06	6.2	58,578,595–58,638,595	5	NA	-	-
	5	4,373,833	A/G	15.4	2.1E-06	5.7	4,343,833–4,403,833	4	LOC104894280	4,399,608–4,402,532 (+25775)	dof zinc finger protein DOF5.6
	8	51,664,047	T/G	13.7	7.1E-06	5.1	51,634,047–51,694,047	4	NA	-	-
Leaf veins red color intensity (LVRCI)	5	9,796,059	G/T	13.6	8.2E-06	5.1	9,766,059–9,826,059	2	LOC104894023	9,801,823–9,803,992 (-5764)	NAC domain-containing protein

[a] Only traits with significant associations are presented [i.e., root shape (RS), root length-width ratio (RLWR), root tip shape (RTS), and leaf margin undulations (LMU) were excluded from the table]. [b] Candidate gene identified outside the GWAS region (± 30 kb from peak marker). For each trait, the associated regions are presented in decreasing order based on their significance level. NA. No candidate gene found in GWAS region. Additional information on GWAS and candidate gene analyses are presented in Supplementary Table S4

Among the leaf traits, 14 significant associations across six chromosomes were identified for LS, representing the largest number of QTLs identified for a single trait in this study. A total of 10 candidate genes were identified for nine of these associations. These genes encoded a putative transcription factor (TF) of the homeobox-leucine zipper gene family, a 3-ketoacyl-CoA synthase, a pentatricopeptide repeat protein, a DEAD-box ATP-dependent RNA helicase, two receptor-like protein kinases, a calmodulin-binding protein, a the TFs belonging to the NAC, ethylene-responsive factor (ERF), and SCARECROW gene families. Their encoded proteins are known to regulate the expression of other genes–e.g., by means of regulating transcription, modulating RNA metabolism, or acting as signal molecules- in key developmental and physiological pathways, and their activity has been associated with LS, through different mechanisms, in various species (discussed below). For LBC, conditioned by the presence or absence of the red pigment betacyanin, five associations were identified in four chromosomes. In one of these associated regions, in Chr. 5, a putative candidate gene encoding a 'DNA-binding-with-one-finger' (DOF) TF, known to regulate anthocyanin pigmentation in some species (discussed below), was identified. Lastly, a single region in Chr. 5 was associated with the red pigmentation in leaf veins (LVRCI), also determined by the presence and content of betacyanins, and this region harbored a NAC TF. NAC TFs have been shown to control anthocyanin biosynthesis by regulating the expression of other TFs (e.g., MYBs and bHLHs) directly involved in the transcriptional regulation of key enzymes of this pathway. Its potential role in conditioning LVRCI in sugar beet is discussed below.

Discussion

Sugar has become an integral component of human diets and is a valuable raw resource in the food, beverage, and pharmaceutical sectors. It is a common name for sucrose, which is mainly obtained from two major crops, sugar cane and sugar beet. Like other crops, sugar beet production is negatively affected by the changing climatic conditions [27, 37]. Under these conditions, germplasm characterization is considered an effective means to explore the available diversity and identify superior materials that can be used for the development of improved cultivars. Regarding sugar beet germplasm, the United States Department of Agriculture's (USDA) National Germplasm System contains ~ 2700 accessions belonging to various species of the genus *Beta*, whereas the Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben, Germany, conserves more than 2000 accessions of sugar beet. Besides these large collections, countries having large-scale sugar beet production usually maintain locally-adapted germplasm. Researchers have employed several molecular markers to investigate genetic diversity and population structure in sugar beet and closely related species [21, 36, 38–42]. To this end, most of the available studies used a few PCR-based markers [38–42], whereas only two reports used large-scale whole genome covering markers [21, 36]. In addition, given the large number of sugar beet germplasm available worldwide, different studies may yield substantially different results, based on the particular selection of the accessions used across studies. This will likely affect general estimates of genetic diversity and population structure indices, genetic relationships among the accessions, and the possibility of identifying significant QTLs–and thereby candidate genes for traits of interest- by GWAS (i.e., the latter analysis relies on the degree of genotypic and phenotypic variation associated with particular phenotypic traits in the germplasm collection studied). Thus, to provide an accurate and informative assessment of the genetic diversity and population structure, and to be able to identify QTLs and candidate genes for traits of interest in sugar beet and other closely related–and genetically inter-crossable- crop types (e.g., beet root, chard, and mangelwurzel), this study evaluated 94 accessions phenotypically and geographically diverse (from 16 countries and four continents) of *Beta vulgaris* subsp. *vulgaris*, along with 4609 SNPs and 6950 silicoDArT markers.

Estimates of various genetic diversity indices (NGD, PIC, MAF, H_o, H_e, and F_i) revealed the existence of a good level of genetic diversity which is conserved in the USDA sugar beet germplasm. It was observed that the SNP markers data set yielded greater values for these diversity indices as compared to silicoDArTs. These findings are in line with those from an earlier study [42] reporting greater PIC value and polymorphism frequency with SNPs as compared to DArT markers. Their study characterized genetic diversity in 54 sugar beet hybrid varieties using DArT, SNP, and SSR markers, and reported greater mean PIC values for all three marker systems than the PIC values found in the present study with SNPs and silicoDArTs. Similarly, another study using 'restriction enzyme associated DNA' (RAD)-derived SNP markers for genotyping sugar beet accessions [36], reported slightly greater PIC and H_e values than those found in the present study. The slightly greater genetic diversity found in these previous reports–as estimated by PIC and H_e indices- as compared to our study may be due to differences in the germplasm(i.e., presumably, a more diverse germplasm collection was used in those studies) and the type and amount of molecular markers used.

SNP markers-based genetic relationships

Estimates of pair-wise genetic distance (HGD) among the accessions revealed Ames 2644 and Ames 8297 (HGD = 0.362) as the most genetically distinct genotypes.

Table 4 Geographical origin and main morphological characteristics for the tap root and leaves of 94 sugar beet accessions

Accession[a]	Geographical origin Country (State)	Root traits							Leaf traits				
		Fresh weight (g)	Length (cm)	Width (cm)	Length-width ratio	Shape (UPOV code)[b]	Shape of tip (UPOV code)[b]	External color	Length (cm)	Blade color	Blade shape	Veins red color intensity	Margin undulations
Ames 14,432	Bulgaria (Plovdiv)	393.0	21.0	12.0	1.8	circular (3)	rounded (2)	reddish purple	17.0	red	medium	strong	medium
Ames 2644	USA (Utah)	118.0	16.0	5.0	3.2	obovate (4)	pointed (1)	yellow	19.0	green	narrow	weak	weak
Ames 2658	USA (Utah)	307.4	37.0	8.5	4.4	circular (3)	rounded (2)	yellow	13.1	green	narrow	weak	medium
Ames 2661	USA (Utah)	606.3	23.0	8.0	2.9	very narrow obovate (6)	pointed (1)	yellow	22.2	green	broad	weak	strong
Ames 2662	USA (Utah)	184.3	27.0	7.5	3.6	narrow oblong (5)	pointed (1)	yellow	20.0	green	medium	weak	medium
Ames 3039	USA (California)	369.3	28.0	6.5	4.3	obovate (4)	pointed (1)	yellow	18.1	green	broad	weak	weak
Ames 3047	USA (California)	465.5	44.0	6.5	6.8	obovate (4)	pointed (1)	yellow	22.1	green	broad	weak	medium
Ames 3049	USA (California)	120.7	29.0	11.0	2.6	obovate (4)	pointed (1)	yellow	27.0	green	medium	weak	weak
Ames 3060	Denmark	856.3	18.0	9.0	2.0	circular (3)	depressed (4)	yellow	26.9	green	medium	weak	strong
Ames 4375	No info	1300.0	30.0	13.0	2.3	very narrow obovate (6)	pointed (1)	white	12.1	green	medium	medium	medium
Ames 8281	England, UK	20.0	12.0	2.0	6.0	obovate (4)	pointed (1)	white	5.2	green	medium	absent	absent
Ames 8283	England, UK	74.0	28.0	6.0	4.7	very narrow obovate (6)	pointed (1)	white	11.1	green	narrow	absent	medium
Ames 8286	England, UK	10.0	16.0	2.0	8.0	very narrow obovate (6)	rounded (2)	white	3.2	green	narrow	absent	weak
Ames 8287	England, UK	104.0	20.0	5.0	4.0	obovate (4)	depressed (4)	white	5.0	green	broad	absent	weak
Ames 8288	England, UK	322.0	20.0	5.0	4.0	obovate (4)	flat (3)	white	14.8	green	broad	medium	weak
Ames 8291	England, UK	10.0	20.0	3.0	6.7	transverse medium elliptic (2)	depressed (4)	white	17.2	green	broad	absent	weak
Ames 8292	England, UK	na	na	na	na	na	na	white	na	na	na	na	na
Ames 8294	England, UK	256.0	19.0	7.0	2.7	narrow oblong (5)	depressed (4)	white	18.1	green	broad	absent	medium
Ames 8295	England, UK	508.0	26.0	13.0	2.0	obovate (4)	flat (3)	white	22.0	green	broad	absent	medium
Ames 8297	England, UK	10.0	17.0	3.0	5.7	obovate (4)	depressed (4)	white	3.1	green	narrow	absent	weak
Ames 8298	England, UK	na	na	na	na	na	na	white	na	na	na	na	na
Ames 8302	England, UK	570.0	26.0	11.0	2.4	obovate (4)	depressed (4)	white	24.9	green	broad	absent	medium
NSL 176,303	Serbia and Montenegro	555.0	28.0	14.0	2.0	circular (3)	rounded (2)	white	15.1	green	broad	absent	weak
NSL 176,412	Serbia and Montenegro	374.0	7.0	28.0	0.3	obovate (4)	depressed (4)	white	16.4	green	broad	absent	medium
NSL 28,024	USA (Wyoming)	106.0	14.0	7.0	2.0	circular (3)	rounded (2)	reddish purple	14.2	green/red	broad	strong	weak
NSL 28,714	USA (Wyoming)	412.0	36.0	9.0	4.0	narrow oblong (5)	flat (3)	white	20.1	green	broad	absent	medium
NSL 28,716	USA (Wyoming)	572.0	24.0	11.0	2.2	transverse narrow elliptic (1)	depressed (4)	white	18.0	green	broad	absent	medium
NSL 6320	USA (Illinois)	308.0	28.0	10.0	2.8	circular (3)	flat (3)	reddish purple	14.1	green	broad	weak	weak
NSL 86,577	USA (Colorado)	546.0	22.0	11.0	2.0	narrow oblong (5)	pointed (1)	white	16.3	green	medium	absent	absent
PI 105,335	China	66.0	20.0	4.0	5.0	very narrow obovate (6)	pointed (1)	reddish purple	15.2	red	broad	strong	medium

Table 4 (continued)

Accession[a]	Geographical origin Country (State)	Root traits							Leaf traits				
		Fresh weight (g)	Length (cm)	Width (cm)	Length-width ratio	Shape (UPOV code)[b]	Shape of tip (UPOV code)[b]	External color	Length (cm)	Blade color	Blade shape	Veins red color intensity	Margin undulations
PI 113,306	China	376.0	31.0	7.0	4.4	circular (3)	pointed (1)	white	21.1	green	broad	absent	weak
PI 117,117	Turkey	272.0	24.0	8.0	3.0	narrow oblong (5)	pointed (1)	white	17.2	green	broad	absent	absent
PI 120,694	Turkey	1150.0	20.0	17.0	1.2	circular (3)	flat (3)	white	20.3	green	broad	absent	absent
PI 120,695	Turkey	830.0	28.0	14.0	2.0	transverse narrow elliptic (1)	flat (3)	reddish purple	17.9	green	broad	weak	weak
PI 120,706	Turkey	508.0	21.0	13.0	1.6	transverse medium elliptic (2)	rounded (2)	reddish purple	27.1	green	broad	absent	weak
PI 124,528	India	952.0	21.0	14.0	1.5	transverse medium elliptic (2)	rounded (2)	reddish purple	24.0	green	broad	medium	weak
PI 140,350	Iran	na	na	na	na	na	na	na	na	na	na	na	na
PI 140,353	Iran	314.0	17.0	11.0	1.5	transverse medium elliptic (2)	rounded (2)	reddish purple	14.3	green	broad	medium	weak
PI 140,354	Iran	584.0	16.0	13.0	1.2	transverse narrow elliptic (1)	flat (3)	white	21.4	green	broad	weak	weak
PI 140,355	Iran	692.0	26.0	18.0	1.4	transverse medium elliptic (2)	flat (3)	white	14.1	green	broad	absent	weak
PI 140,356	Iran	310.0	21.0	9.0	2.3	transverse narrow elliptic (1)	5	white	24.2	green	broad	absent	weak
PI 140,358	Iran	785.0	19.0	14.0	1.4	circular (3)	rounded (2)	white	13.3	green	broad	medium	weak
PI 140,360	Iran	965.0	24.0	21.0	1.1	transverse medium elliptic (2)	flat (3)	reddish purple	18.8	green	broad	absent	weak
PI 140,361	Iran	278.0	18.0	12.0	1.5	circular (3)	pointed (1)	white	25.2	green	broad	absent	weak
PI 141,919	No info	222.0	18.0	10.0	1.8	circular (3)	rounded (2)	reddish purple	11.2	green	broad	medium	weak
PI 142,808	Iran	820.0	24.0	18.0	1.3	transverse medium elliptic (2)	flat (3)	white	17.0	green	broad	absent	weak
PI 142,809	Iran	1328.0	30.0	25.0	1.2	transverse narrow elliptic (1)	flat (3)	reddish purple	31.4	green	broad	medium	weak
PI 142,812	Iran	1158.0	42.0	32.0	1.3	transverse narrow elliptic (1)	flat (3)	yellow	30.2	green	broad	absent	weak
PI 142,814	Iran	344.0	21.0	10.0	2.1	obovate (4)	depressed (4)	white	18.1	green	broad	absent	weak
PI 142,815	Iran	2430.0	34.0	28.0	1.2	transverse medium elliptic (2)	depressed (4)	reddish purple	28.7	green	broad	absent	weak
PI 142,816	Iran	980.0	31.0	21.0	1.5	transverse narrow elliptic (1)	depressed (4)	yellow	27.1	green	broad	absent	absent
PI 142,817	Iran	192.0	23.0	8.0	2.9	circular (3)	depressed (4)	white	18.0	green	broad	absent	weak
PI 142,818	Iran	812.0	20.0	23.0	0.9	transverse medium elliptic (2)	rounded (2)	yellow	26.2	green	broad	absent	weak
PI 142,820	Iran	452.0	27.0	11.0	2.5	transverse narrow elliptic (1)	depressed (4)	reddish purple	22.3	green	broad	absent	weak
PI 142,821	Iran	1514.0	27.0	24.0	1.1	transverse narrow elliptic (1)	flat (3)	white	36.0	green	broad	absent	weak

Table 4 (continued)

Accession[a]	Geographical origin Country (State)	Root traits							Leaf traits				
		Fresh weight (g)	Length (cm)	Width (cm)	Length-width ratio	Shape (UPOV code)[b]	Shape of tip (UPOV code)[b]	External color	Length (cm)	Blade color	Blade shape	Veins red color intensity	Margin undulations
PI 142,823	Iran	76.0	30.0	7.0	4.3	circular (3)	depressed (4)	reddish purple	20.2	green	broad	weak	weak
PI 144,675	Iran	2676.0	40.0	36.0	1.1	transverse narrow elliptic (1)	depressed (4)	reddish purple	32.1	green	broad	weak	weak
PI 148,625	Iran	1476.0	24.0	22.0	1.1	transverse medium elliptic (2)	depressed (4)	reddish purple	23.2	green	broad	absent	weak
PI 164,659	India	482.0	17.0	11.0	1.5	circular (3)	pointed (1)	reddish purple	9.4	green/red	broad	strong	medium
PI 164,671	India	na	na	na	na	na	na	na	na	na	na	na	na
PI 164,805	India	580.0	20.0	14.0	1.4	transverse medium elliptic (2)	rounded (2)	reddish purple	19.0	green/red	broad	strong	medium
PI 164,968	Turkey	1078.0	37.0	18.0	2.1	transverse medium elliptic (2)	rounded (2)	reddish purple	20.2	green/red	broad	very strong	medium
PI 165,062	Turkey	569.0	37.0	14.0	2.6	narrow oblong (5)	pointed (1)	white	19.9	green	broad	absent	medium
PI 165,485	India	188.0	15.0	8.0	1.9	circular (3)	pointed (1)	reddish purple	13.1	green/red	broad	very strong	medium
PI 169,014	Turkey	718.0	28.0	16.0	1.8	transverse medium elliptic (2)	depressed (4)	white	18.1	green	broad	absent	strong
PI 169,017	Turkey	1138.0	33.0	16.0	2.1	transverse medium elliptic (2)	depressed (4)	reddish purple	28.5	green	broad	medium	medium
PI 169,029	Turkey	873.0	35.0	13.0	2.7	circular (3)	depressed (4)	reddish purple	28.1	green/red	broad	very strong	medium
PI 169,032	Turkey	740.0	17.0	7.0	2.4	circular (3)	depressed (4)	white	8.2	green	medium	absent	weak
PI 171,508	Turkey	310.0	31.0	16.0	1.9	circular (3)	depressed (4)	white	16.1	green	broad	absent	medium
PI 171,516	Turkey	158.0	20.0	8.0	2.5	circular (3)	pointed (1)	yellow	9.0	green	broad	absent	medium
PI 171,518	Turkey	1132.0	30.0	16.0	1.9	transverse medium elliptic (2)	flat (3)	yellow	24.3	green	broad	absent	weak
PI 171,519	Turkey	1088.0	31.0	17.0	1.8	transverse medium elliptic (2)	depressed (4)	white	20.0	green	broad	absent	weak
PI 172,733	Turkey	na	na	na	na	na	na	white	na	na	na	na	na
PI 176,875	Turkey	708.0	70.0	10.0	7.0	obovate (4)	depressed (4)	white	22.3	green	broad	absent	medium
PI 179,176	Iraq	26.0	13.0	4.0	3.3	narrow oblong (5)	pointed (1)	reddish purple	8.4	green	broad	medium	medium
PI 179,180	Turkey	127.0	33.0	5.0	6.6	narrow oblong (5)	depressed (4)	white	17.1	green	broad	absent	medium
PI 193,458	Ethiopia	308.0	20.0	13.0	1.5	transverse medium elliptic (2)	depressed (4)	reddish purple	16.3	green/red	broad	strong	medium
PI 256,052	Afghanistan	120.0	26.0	8.0	3.3	circular (3)	depressed (4)	yellow	12.2	green	narrow	absent	weak

Table 4 (continued)

Accession[a]	Geographical origin Country (State)	Root traits							Leaf traits				
		Fresh weight (g)	Length (cm)	Width (cm)	Length-width ratio	Shape (UPOV code)[b]	Shape of tip (UPOV code)[b]	External color	Length (cm)	Blade color	Blade shape	Veins red color intensity	Margin undulations
PI 256,053	Afghanistan	532.0	28.0	14.0	2.0	transverse medium elliptic (2)	depressed (4)	reddish purple	12.0	green	broad	absent	weak
PI 590,616	France (Hauts-de-France)	74.0	21.0	6.0	3.5	circular (3)	depressed (4)	white	4.0	green	narrow	absent	weak
PI 590,621	USA	114.0	21.0	7.0	3.0	circular (3)	rounded (2)	reddish purple	11.2	red	broad	very strong	medium
PI 590,697	USA	468.0	26.0	11.0	2.4	circular (3)	depressed (4)	white	18.4	green	broad	absent	medium
PI 590,808	USA	622.0	22.0	9.0	2.4	circular (3)	depressed (4)	white	20.3	green	broad	absent	medium
PI 590,812	USA	28.0	19.0	4.0	4.8	obovate (4)	depressed (4)	white	10.1	green	broad	absent	medium
PI 596,528	USA	538.0	32.0	10.0	3.2	obovate (4)	depressed (4)	white	20.9	green	broad	Absent	weak
PI 610,286	Chile	718.0	27.0	10.0	2.7	circular (3)	depressed (4)	white	22.1	green	broad	absent	medium
PI 610,287	Chile	340.0	33.0	9.0	3.7	circular (3)	depressed (4)	white	12.8	green	broad	absent	medium
PI 610,291	Argentina	na	na	na	na	na	na	white	na	na	na	na	na
PI 610,323	USA	374.0	31.0	8.0	3.9	circular (3)	depressed (4)	white	20.0	green	broad	absent	medium
PI 610,417	USA	100.0	23.0	6.0	3.8	circular (3)	depressed (4)	white	11.1	green	broad	absent	medium
PI 611,059	Bulgaria (Plovdiv)	898.0	13.0	39.0	0.3	transverse medium elliptic (2)	depressed (4)	white	27.3	green	broad	absent	medium
PI 611,060	China	10.0	17.0	2.0	8.5	narrow oblong (5)	depressed (4)	white	5.0	green	broad	absent	weak
PI 611,062	Greece	72.0	21.0	5.0	4.2	transverse narrow elliptic (1)	flat (3)	reddish purple	18.3	red	broad	very strong	medium
PI 633,934	USA	786.0	35.0	12.0	2.9	obovate (4)	depressed (4)	white	20.2	green	broad	absent	medium
Mean		**546.1**	**25.2**	**11.9**	**2.8**				**18.0**				
SD		**485.9**	**8.7**	**7.3**	**1.7**				**6.9**				

[a] All the accession correspond to *Beta vulgaris* L. subsp. *vulgaris*

[b] Numbers in parenthesis correspond to the official phenotypic codes of the International Union for the Protection of New Varieties of Plants (UPOV Code: BETAA_VUL_GVC). The phenotypes of UPOV codes for root shape and root tip shape are depicted in Supplementary Fig. S4

In addition, these accessions present highly-contrasting phenotypes for several morphological traits (Table 4). Genetically and phenotypically contrasting accessions are of great interest for breeding, as these plant materials can be instrumental for future research studies aiming at deciphering the genetic basis underlying traits of interest (e.g., by means of developing mapping populations, followed by linkage mapping and candidate gene analyses), or introgressing new traits into valuable genetic backgrounds. Keeping this in view, pairs of accessions with high HGD and contrasting phenotypes, as identified in this study, could be prioritized and included in sugar beet breeding programs.

The UPGMA-based dendrogram grouped the sugar beet germplasm into three main clusters, namely S-I, S-II, and S-III. The geographical origin and root color of the accessions play an integral part in their clustering into a respective group, and similar findings have been reported previously in other species [43]. Cluster S-I was the smallest group (four accessions), whereas a total of 35 accessions were included in Cluster S-II, with ~51% of the clustered accessions being from USA, 17% from Europe, and 14% from Turkey. It is generally accepted that sugar beet originated in Germany (German Silesia) back in the 18th century and later it was dispersed to other parts of the world [44]. It is highly possible that Turkey received sugar beet from European countries, because of its vicinity to Europe and its particular geographic location, connecting Asia and Europe and being an important path in the Silk Route. Later in the 19th century, this crop was introduced in North America, most likely from Europe, and those materials were used for initiating germplasm collections, conservation, characterization, and breeding activities at the USDA [45]. This could be one of the main reasons behind the genetic similarity found between USA materials and accessions of Turkish and European origins. Cluster S-III was the largest group, harboring a total of 55 accessions and, on a broad scale, this cluster grouped accessions from the Middle East, Southeast, and East Asia, accounting these geographical origins for ~73% of the clustered plant materials, whereas the remaining taxa were mostly from Europe (accounting for 20% of the accessions). The main cluster S-III was further subdivided into subclusters S-III-A and S-III-B. The accessions in S-III-A have diverse origins, mainly from Asia, Europe, and USA, whereas S-III-B showed a clear predominance of materials from Middle East countries, representing this region more than 80% of the total clustered accessions. This suggests that countries from this region share similar preferences for sugar beet cultivars and desired traits, probably favoring locally adapted materials, and -most likely- they have shared genetic material of sugar beet throughout history. This is conceivable, given that the three countries from which most of these accessions derived–i.e., Turkey, Iran, and Afghanistan- are neighbors, with Iran sharing borders with the other two countries. Also, it is believed that sugar beet was first introduced in Iran and Afghanistan through the route of the Silk Road by, where Turkey served as the nexus between continents for trader travelers. The genetic relationships observed in the sugar beet germplasm by UPGMA were fully supported by results from population STRUCTURE analysis, revealing three (K = 3) or four (K = 4) possible subpopulations in this collection, as well as by PCoA, showing three genetically distinct groups of accessions; showing a rather clear association between the genetic clusters and root color of the accessions, and partially followed geographic origin (Figs. 1 and 2A and C).

The fact the root color was associated with the genetic clusters (e.g., red beets predominated in cluster S-III, particularly in subcluster S-III-A, whereas S-II contained, almost exclusively, white and yellow-rooted beets; Fig. 1) and geographical origins of the accessions (e.g., ~60% of the red-rooted beets were from the Middle East, mainly from the neighbor countries Iran and Turkey, and an additional 20% were from India) suggests that red-rooted materials have been selected from a few and/or closely related genetic stocks grown in specific regions, whereas white and yellow beets seem to have diverse origins.

The fact that our analysis of molecular variance based on SNP markers revealed that most of the genetic variation can be attributed to differences among the accessions (i.e., within clusters) rather than among subpopulations or clusters coincides with previous findings, reporting greater SNP variation within groups than among groups in sugar beet germplasm [33]. Together, these data suggest that in sugar beet germplasm a small proportion of the total variation is unique to any given accession, and that within-group analyses are required to evidence genetic variants that may be of value for breeding purposes.

SilicoDArT markers-based genetic relationships

Estimates of pair-wise genetic distance (HGD) among the accessions revealed NSL 176,303 (from Serbia) and PI 140,355 (from Iran) (HGD = 0.439) as the most genetically distinct genotypes. Noticeably, this pair of most-contrasting accessions differed from that obtained with SNP markers, and the former exhibited greater HGD value. This is likely due to differences in the type of markers and their polymorphisms across the accessions. Nonetheless, these genetically-contrasting materials may be of interest for sugar beet breeding programs, as discussed above.

The UPGMA-based dendrogram separated the sugar beet germplasm into two major clusters, D-I and D-II, which partially followed geographic origins and showed

a more robust association with root color phenotypes, in general agreement with results obtained with SNP data. For example, cluster D-I included 90% of the accessions from the USA, as well as some materials from other origins, and contained exclusively white- and yellow-rooted materials, strongly resembling cluster S-II of the SNP-based dendrogram (Figs. 1 and 3). Similarly, cluster D-II, which was further subdivided into subclusters A, B, C, and D, exhibited a clear concordance with the SNP-based cluster S-III and its subclusters. For instance, subcluster D-II-B (silicoDArTs) contained most of the taxa included in subcluster S-III-A (SNPs), comprising only red and white-rooted beets, with a strong predominance of red beets. As discussed earlier, the joined clustering of genetically-similar red-rooted accessions suggests that red beet materials were originally selected from a few and/or closely related genetic stocks. On the other hand, the largest subcluster D-II-D (silicoDArTs) comprised accessions of different color phenotypes predominantly from the Middle East, mainly from Iran and Turkey, accounting this region for 77% of the clustered materials, showing a strong resemblance–and sharing most of the clustered taxa- with subcluster S-III-B (SNPs) comprising 83% of its accessions from the Middle East. As discussed above, this suggests that neighbor countries from this region have similar preferences and accessibility for sugar beet cultivars and that they may have shared genetic material throughout history. The partial associations among genetic clusters, geographical origins, and root color phenotypes observed with silicoDArTs were further confirmed by STRUCTURE and PCoA analyses, revealing two distinct genetic groups that shared most of the clustered accessions with the those in clusters D-I and D-II from the UPGMA dendrogram (Figs. 2B and D and 3).

Results from AMOVA using silicoDArTs data confirmed that most of genetic variation in this germplasm was due to differences among the accessions, accounting for 74.5–77.6% of the total variation, rather than among clusters (22.4–25.5% of the variation); in full agreement with data from SNPs analysis, showing 71.3% and 28.7% of the total variation attributed to differences among accessions and among clusters, respectively (Tables 1 and 2).

GWAS enabled the identification of candidate genes conditioning tap root and leaf phenotypic traits

In the present study, GWAS revealed 35 significant marker-trait associations for nine phenotypic traits, and further analysis of the predicted function of the genes in the associated regions identified 25 candidate genes for four tap root (RFW, RW, RL, REC) and three leaf traits (LS, LBC, LVRCI) (Fig. 4; Table 4, Supplementary Table S3). Among the tap root traits, RFW is directly related with root biomass yield, which is of great importance for the sugar beet crop as a major industrial source of sugar. For this trait, a region in Chr. 3 with the strongest association harbored a gene (LOC104883815) annotated as U-box protein 14 (PUB14), which fully co-localized with the peak GWAS signal. U-box type E3 ubiquitin ligases (PUBs) determine substrate specific recognition during posttranslational ubiquitination of proteins, thereby modulating numerous plant signaling pathways associated with plant growth, development, biotic and abiotic stresses, and hormone signaling [46]. With regards to root growth, some members of this family have been shown to regulate asymmetric cell division and cell proliferation in the root meristem, conditioning final root architecture [47]. In *Arabidopsis*, deleterious mutants of *AtPUB9* displayed severely reduced root growth and lateral root formation, suggesting that this PUB regulates root development [48]. In corn, 10 PUB genes were found to be highly upregulated in the primary root during different stages of root development, suggesting that these genes participate in root growth and morphogenesis [49]. In potato, *PUB14* co-localized with the largest-effect QTL for early maturity, a trait that results in fast tuber growth, suggesting that *PUB14* was involved in maturity and regulation of tuber formation [50]. Results from previous studies suggest that PUBs exert many of their effects in plants, including root growth, by regulating hormone signaling, providing evidence of how targeted degradation of proteins by ubiquitination affects downstream transcriptional regulation of hormone-responsive genes in the auxin, gibberellin (GA), abscisic acid (ABA), ethylene (ETH), and jasmonate (JA) pathways [50]. In addition, ubiquitin-mediated proteolysis may also act upstream of the hormonal signaling cascades by regulating hormone biosynthesis, transport and perception [50]. A search for cis-acting regulatory elements in the 2000 bp region upstream from the transcription start site and in introns of LOC104883815, revealed 12 hormone-response elements, comprised of JA (1), ETH (4), ABA (3), SA (3), and GA (1); and three metabolic and growth related elements associated with meristem expression (Supplementary Fig. S4). Coincidently, some of these same responsive elements, namely GA and meristem-expression related elements, were also present in the promoter regions of a set of PUB genes that were highly expressed during root development in maize [49]. It has been shown in *Arabidopsis* that apical root meristem-specific genes play a crucial role in regulating root development [52], whereas GA determines cell growth polarity in the root cortex of maize [53] and promotes primary root growth in rice [54]. Altogether, these data suggest LOC104883815 as a strong candidate gene for RFW, but additional studies aiming at functionally characterizing this gene and its relationship with root growth in

sugar beet are necessary to test this hypothesis. Another region in Chr. 7, associated with RFW, harbored a midasin (MDN1) gene (LOC104899776) that colocalized with the peak marker. MDN1 is an essential AAA (ATPase associated with various cellular activities) protein that removes assembly factors from distinct precursors of the ribosomal 60 S subunit, thereby strongly influencing ribosome biogenesis [55]. In *Arabidopsis*, MDN1 is essential for early development of the embryo and for root development and growth, being the latter effect attributed to the upregulation, by MDN1, of cell proliferation in the root apical meristem [55, 56]. Mutants of this gene exhibited retarded root growth, severely smaller root size (~threefold shorter), and a dwarf plant phenotype as compared to their wild type counterparts in *Arabidopsis* [56] and maize [57]. According to Birne [58], these phenotypes are commonly observed for ribosomal protein mutations in several plant species, and may be caused by a defective protein synthesis in root cells as a result of inadequate ribosome biogenesis. Lastly, in Chr. 4, another region associated with RFW contained seven genes, including LOC104890632, which colocalized with the peak marker and was annotated as an embryogenic flowering 2 (*EMF2*) gene. *EMF2* encodes a Polycomb group (PcG) protein involved in the formation of protein complexes that maintain gene silencing via histone modification, and this gene -along with Polycomb repressive complex (PRC) proteins- has been shown to epigenetically inhibit primary root growth and lateral root development, whereas the loss-of-function of *EMF2* resulted in long primary roots and increased production of lateral roots [59]. Thus, it is possible that mutations in this gene may influence the growth and architecture–and thereby RFW- in the tap root of sugar beet.

For RW, the region with strongest association, found in Chr. 8, harbored 7 genes, including a putative nudix hydrolase 20 (LOC104902151) that fully colocalized with the peak GWAS signail. Nudix hydrolases are hydrolytic enzymes capable of cleaving nucleoside diphosphates linked to other moieties [60]. In a recent GWAS study in sugar beet, another member of the nudix hydrolase gene family, namely nudix hydrolase 15 (BVRB_8g182070), was strongly associated with root shape and root groove depth [61]. The enzyme encoded by this gene can hydrolyze NADPH, an essential cofactor required for cell growth and proliferation in roots and leaves [62]. Coincidently, in another GWAS report in sweet potato, a nudix hydrolase homologue was found to be strongly associated with continuous storage root formation and bulking, and this effect might be associated with hormones promoting lateral root initiation [63]. In that same region, a gene encoding a small polypeptide DEVIL 10 (LOC125492567) protein was also identified. Small polypeptides, typically containing less than 150 amino acids, can act as important regulatory molecules that coordinate cellular responses required for differentiation, growth, and development [64]. A subfamily of such small peptides, called DEVIL-like (DVL), have been shown to regulate root growth under abiotic stress conditions by affecting ABA-related gene expression in *Arabidopsis* [65]. Conversely, DVLs from fox millet were shown to suppress cell division and elongation in root by inhibiting auxin signaling and transport, which lead to reduced root growth and a short root phenotype [66], and similar phenotypes were observed in *Arabidopsis* plants overexpressing two DVL proteins [67], suggesting that DVLs are involved in the regulation of primary root growth. These data point at LOC125492567 as a possible candidate gene for root growth and RW in sugar beet. Another significant association with RW was found in Chr. 3, which harbored two putative glutathione-S-transferase (GST) genes (LOC104906825 and LOC104906851). GSTs play a major role in the plant's antioxidant responses, by limiting oxidative damage caused by different types of stresses, but they also can influence root growth and development. For example, *GSTU7*, a homolog of the sugar beet GSTs found in this region, has been shown to promote primary root growth and root length in *Arabidopsis* [68]. *GSTU7* acts as a glutathione peroxidase. Although the exact mechanism by which *GSTU7* affects root growth is still unclear, it was hypothesized that it may be related to the formation of H_2O_2, as this reactive oxygen species (ROS) has been associated with root development and the activity of peroxidases that control the transition from proliferation to differentiation in the root by balancing ROS between root zones [69–71]. Very recently, overexpression of a GST from sweet potato, namely *IbGSTL2*, resulted in increased starch content and increased amylopectin/amylose ratio, suggesting that GSTs may influence starch biosynthesis and promote growth in storage roots [72].

For RL, the strongest association was found in Chr. 5, and this region contained three candidate genes, two of them encoding putative expansin-like B1 (EXLB1) proteins (LOC104892825 and LOC104892824) and the other one encoding a nudix hydrolase 2-like protein (LOC125497584). Expansins are structural cell wall proteins that participate in cell growth and stress responses by regulating cell wall expansion through weakening the hydrogen bond between cell wall polysaccharides, and they interact with plant hormones to coordinate many physiological and cellular processes of plant growth [73]. A Brassica homologue of the two expansin genes found in sugar beet, namely *BrEXLB1*, was found to promote primary root growth by increasing the size of the root elongation zone, thereby increasing RL [74]. Similarly, the soybean expansin *GmEXP1* was responsible for primary root elongation, and its overexpression

resulted in longer primary roots in soybean and tobacco [75]. Altogether, these data suggest LOC104892825 and LOC104892824 as strong candidates for RL in sugar beet. As described above, homologues of the other candidate gene found in this region, a putative nudix hydrolase, were previously associated–by GWAS analysis- with storage root shape and root groove depth in sugar beet [61] and with continuous storage root formation and bulking in sweet potato [63]. Another region in Chr. 7, also associated with RL, contained a putative RNA polymerase II (LOC104899323). RNA polymerase II (Pol II) synthesizes protein-coding mRNAs and microRNAs (miRNAs), and its transcription activity is regulated by multiple proteins that either control transcriptional initiation, usually interacting with the promoter region of Pol II, or they modulate–by other mechanisms- the efficiency of mRNA synthesis after the initiation stage. Mutations in some of the components of these 'protein-Pol II' complexes have been repeatedly shown to influence many aspects of primary root growth, including RL [76, 77]. Additional studies are needed to test whether this gene influences RL in sugar beet. Close to the latter gene, we identified a putative histone acetyltransferase (HAT, LOC104899324) that colocalized with the peak GWAS signal. Histone acetylation is a HAT-mediated epigenetic modification that regulates the expression of numerous genes involved in development and other plant processes. A number of HAT homologues have been shown to regulate root growth and morphogenesis, thereby influencing RL, across several species [59, 78].

For tap root color (REC) a single association was found in Chr. 4, and this region harbored a putative ATP-binding cassette (ABC) transporter (LOC104890386). ABC transporters have been demonstrated to be involved in vacuolar transport of plant secondary metabolites [79], including that of anthocyanins, and–by facilitating their accumulation in the vacuole- ABC transporters may condition anthocyanin pigmentation in several crop species, including grapes [80], apple [81], and purple carrots [82]. Given the close relationship between anthocyanins and betalains, showing comparable vegetative and reproductive functions in the plant [83], it is conceivable that betalain transport may be conducted by a transporter of the same class. Indeed, based on comparative transcriptome and metabolome analyses in the betalain-containing flowers of *Mirabilis jalapa*, an ABC transporter (*Mj7*) was reported as a major candidate gene for betalain pigmentation, presumably influencing the transport and accumulation of these pigments in the vacuole [84]. Although, together, these data suggest LOC104890386 as a candidate for REC, functional analyses of this gene will be required to evaluate its possible role in betalain transport in sugar beet.

For leaf shape (LS), 10 candidate genes were identified for nine significantly associated regions in six chromosomes. One of the strongest associations with LS was in Chr. 1, and this region contained a gene encoded as homeobox-leucine zipper protein ATHB4 (LOC104905284), a transcription factor belonging to the class II homeodomain leucine-zipper gene family, known to be involved in the regulation of adaptive responses to the environment and in plant development [85]. ATHB4, in particular, was shown to strongly influence leaf morphology, by regulating the patterning of the adaxial domain in the leaf primordium early in development, with the loss-of-function of this gene resulting in strongly abaxialized leaves [86], whereas ATHB4-overexpressed transgenic lines yielded leaves with reduced blade expansion in *Arabidopsis* [87]. Together, these data point at LOC104905284 as a major candidate for LS in sugar beet. Another association with LS, located in a different region of Chr. 1, contained a putative '3-ketoacyl-CoA synthase' (KAS) gene (LOC104896300). KAS enzymes catalyze the first step in the biosynthesis of very long chain fatty acids (VLCFAs), with chain lengths between C20 and C34, which are essential for the production of cuticular waxes, membrane lipids and sphingolipids [88]. In addition, VLCFAs function as signaling molecules that affect hormone transport and regulate organ development [88]. In *Medicago truncatula*, *WFL*, an homologous gene of the sugar beet KAS found in Chr. 1, was shown to strongly affect the morphology of leaves and floral organs, and it is intensively expressed in leaf primordia and flowers, with mutants of this gene exhibiting extremely wrinkled and fused leaves, as well as floral organ fusion defects [89]. Similar alterations in leaf morphology have been repeatedly observed in plants with disrupted VLCFA biosynthesis in *Arabidopsis* [90–92]. Based on these data, we hypothesize that this KAS-encoding gene conditions LS in sugar beet, as a result of variations in VLCFA production among the accessions. In Chr. 2, we identified another strong association with LS, and this region harbored a gene encoding a putative pentatricopeptide repeat (PPR) protein (LOC104907732). PPR proteins are RNA-binding proteins involved in RNA metabolism–including RNA editing, cleavage, splicing, stabilization and translation- in chloroplasts or mitochondria. Previous studies have shown that PPR genes are associated with leaf development, and some PPRs can strongly influence LS in species like *Arabidopsis*, where the activity of this gene can lead to crinkled vs. normal leaves [93]; grape, associated with the development of a leaf rolling phenotype [94]; and melon, conditioning deeply-lobed vs. round leaves [95]. Based on these data, it is conceivable that LOC104907732 may play a similar role conditioning LS in sugar beet. Another association with LS was identified

in Chr. 5, and this region included a putative DEAD-box ATP-dependent RNA helicase gene (LOC104905962). This type of helicase plays a crucial role in gene regulation through RNA metabolism, thereby affecting many plant processes. It was previously shown that an *Arabidopsis* DEAD-box RNA helicase gene, namely *RH10*, had a profound effect on LS, possibly via regulating the expression of abaxial-determining genes to promote the adaxial development of leaves, thereby affecting the polarity establishment of leaves [96]. A region in Chr. 3 was also associated with LS, and contained a putative 'receptor-like protein kinase' gene (LOC104888714). Some receptor-like kinases have been shown to regulate epidermal cell differentiation in leaves and other organs, influencing anticlinal cell division and cell morphology, thereby altering the surface phenotype of these organs. Thus, several epidermally expressed receptor-like kinases have been identified and functionally characterized, with their mutants consistently exhibiting alterations in LS associated with abnormal differentiation of epidermal tissue [97]. Also in Chr. 3, another region associated with LS included a putative calmodulin-binding protein (CaMBP, LOC104889622). The family of CaMBPs mediates calcium (Ca^{2+}) signaling to microtubules, membrane subdomains, and the nucleus during cell growth, and several of its gene members were shown to regulate the shape of leaves and other organs in *Arabidopsis* [98–100] and tomato [101], being these phenotypes associated with alterations in microtubule patterning and orientation. An association with LS was also found in Chr. 4, and this region contained a putative NAC transcription factor (LOC104890605). NAC genes are involved in different developmental and physiological processes such as organ formation, root development, and response to biotic and abiotic stresses. Several NAC genes were shown to regulate LS in various model and crop species, including *Arabidopsis* [102], tomato [103], the Brassicaceae *Cardamine hirsuta* [96], orchid [104], petunia [105], and *Antirrhinum* [106], possibly acting via transcriptional regulation of genes involved in leaf developmental patterning. Altogether, these data suggest LOC104890605 as a strong candidate for LS in sugar beet. Another association with LS in Chr. 8 revealed two candidate genes, a receptor-like protein (LOC104900641), which role was described above, and a putative ethylene-responsive factor (ERF) (LOC104900640). ERFs are TFs involved in various plant processes, including hormone-induced organ development and growth, and response to ethylene and stress. Based on the fact that some ERFs have be shown to regulate LS in species like poplar [107] and maize [108], we hypothesize that LOC104900640 may play a similar role conditioning LS in sugar beet. Lastly, a marginally-statistically association with LS was identified in Chr. 9, and this region contained a putative SCARECROW (SCR) TF, known to be involved in the regulation of root and leaf development. This gene has been associated with changes in LS in various species, possibly by regulating cell division planes and proliferation in this organ [109, 110], yet its role in determining LS in sugar beet needs to be further investigated by means of functional analyses.

For leaf blade color (LBC), five associations were revealed in four chromosomes. One of these associations, in Chr. 5, harbored a putative 'DNA-binding-with-one-finger' (DOF) transcription factor (TF) (LOC104894280), known to regulate anthocyanin pigmentation in some species [111, 112]. In sugar beet, LBC varies from red-violet to green, these phenotypes being determined by the accumulation, or not, of red betalains called betacyanins. Anthocyanin and betalains, although mutually exclusive in their occurrence across species (i.e., in the plant kingdom, different taxa may have either–but not both- of these pigments, being betalains only present in the Caryophyllales), share several structural [e.g., glucosyltransferases, acyltransferases, *Anthocyanidin synthase* (*ANS*), *Dihydroflavonol4-reductase* (*DFR*)] and regulatory genes involved in their biosynthetic pathways [84, 113]. Among the latter, TF families including MYB, bHLH, DOF, AP2, NAC, TCP, WD40 and WRKY play crucial roles in anthocyanins biosynthesis by regulating the expression of key enzyme genes. To date, only some of the TFs regulating anthocyanin pigmentation have been directly associated with betalain production in the Caryophyllales, namely an R2R3 MYB in beet [114], two bHLHs in spinach [115], and several MYBs and one WRKY in pitaya [116–119]. Given the fact that betalains and anthocyanins assume similar physiological roles *in planta*, it is conceivable that additional betalain-regulatory elements belonging to the other TF families involved in anthocyanin biosynthesis, such as DOF, may be discovered in the future. Supporting this hypothesis is the fact that binding sites for DOF and other TFs (namely, MYB, bHLH, bZIP, and WRKY) were identified in the promoter region of two DOPA dioxygenase genes involved in the production of betalamic acid, one of the first steps in betalains biosynthetic pathway, in *Phytolacca americana* [120].

The red color in leaf veins (LVRCI) in sugar beet, which can vary from very intense pigmentation to absence of red color, is also due to the production of betacyanins. A single association with LVRCI was identified in Chr. 5, and this region harbored a putative NAC transcription factor (LOC104894023). NAC TFs have been shown to control anthocyanin biosynthesis by regulating the expression of other TFs–e.g., MYBs and bHLHs- that directly regulate the transcription of enzyme genes of this pathway in many plant species, including chrysanthemum [121], apple [122], peach [123], pear [124], and *Arabidopsis* [125]. Although to date NAC TFs have not

been reported to condition betalain production, considering that anthocyanins and betalains share a number of enzymes and TFs in their biosynthesis, it is conceivable that the NAC gene positionally associated with LVRCI may be involved in the red coloration of leaf veins in sugar beet.

It is worth noticing that no candidate genes were identified for several of the significant associations found, and this may be related, at least in part, with the fact that such regions contained a large proportion of genes with uncharacterized proteins (Supplementary Table S3). Thus, future improvements in annotation of the sugar beet genome will likely benefit the identification of candidate genes conditioning relevant traits in this species. In addition to genome annotation quality, linkage disequilibrium (LD) and SNP proximity to the causal gene variant can also influence candidate gene prediction (i.e., if LD extends over large genomic regions, pinpointing the causal gene may be challenging). In these cases, fine-mapping approaches can help narrow down candidate genes. Also, the absence of significant associations for four of the 13 phenotypic traits analyzed -namely root shape, root length-width ratio, root tip shape, and leaf margin undulations- may be due to insufficient variation for these traits within our germplasm collection and/or relatively low sample size. Also, gene-environment interactions can influence genetic effects, thereby potentially masking marker-trait associations. Thus, future GWAS analyses using a large number of accessions with clear and broad variation for these traits, ideally tested under different environmental conditions, may reveal significant associations and candidate genes for these traits. Future GWAS studies in sugar beet should also account for differences in allele frequencies across subpopulations (e.g., using PCA or Structure analysis) and pre-select markers with low minor allele frequency (MAF), as was done in the present work. Other factors that may influence the detection of marker-trait associations, but are virtually impossible to control in GWAS experiments, are gene-gene interactions (epistasis), as non-additive interactions can affect the expression of some traits, potentially confounding simple SNP-trait associations; and the genomic position of the SNP markers (i.e., SNPs located in regulatory sequences are more likely to have functional effects, making it easier to detect associations if functional annotation is considered). Improving genome annotation and expanding the extent of phenotypic variation may be particularly relevant for root sugar content, for which a single marginally significant QTL–but no candidate gene- was identified, given the importance of this trait in sugar beet. Lastly, although this study identified candidate genes based on their positional colocalization with significant GWAS signals and their predicted functions based on genome annotations, their candidacy for genetically-controlling each of these traits needs to be confirmed in future studies, e.g., by gene expression analysis in phenotypically contrasting plant materials, and/or using gene editing or genetic transformation approaches.

Conclusions

In this study, a large number of polymorphic GBS-derived co-dominant SNP and dominant silicoDArT markers were used to explore genetic diversity and population structure in an international sugar beet germplasm. In addition, the SNPs dataset was used for GWAS and candidate gene analyses on 13 root and leaf phenotypic traits of relevance for sugar beet and related crop types within *B. vulgaris* subsp. *vulgaris.* Both marker systems revealed substantial genetic diversity in this collection, as estimated by various diversity indices, identifying genetically and phenotypically contrasting accessions, without the presence of duplicate materials. Clustering analysis by both marker systems grouped the accessions in partial agreement with their geographical origins and their tap root color, suggesting a common history of phenotypic selection and utilization of closely related genetic stocks in some regions (e.g., in Middle East countries). GWAS led to the identification of 35 significant associations for nine traits and, based on predicted functions of the genes in the associated regions, 25 candidate genes were identified for four root (fresh weight, width, length, and color) and three leaf traits (shape, blade color, and veins color). Altogether, the gathered data is expected to be useful for sugar beet germplasm curation, marker-assisted breeding, and genomic selection; and provide a framework for future research aiming at investigating the genetic basis of some of the phenotypic traits studied herein.

Materials and methods

Plant material, phenotyping, and DNA extraction

A total of 94 sugar beet accessions were used as plant material in this study. These germplasm belong to 16 countries from four continents, and they present clear phenotypic variation for agronomic and root quality traits (Table 4). The leaf and root phenotypes of the accessions are presented in Supplementary Fig. S5. Seeds of each accession were provided by the United States Department of Agriculture (USDA). The seeds were sawn in germination trays and cultivated in a growth chamber at the Agricultural Biotechnology Department, Van Yuzuncu Yil University, Faculty of Agriculture, Turkey. After three weeks of culture, seedlings of the sugar beet accessions were transplanted to the experimental field of the Field Crops Department of the same institution, where they were grown for six months using conventional agricultural practices for the crop, after which the beet plants were harvested and immediately prepared for phenotyping by separating the foliage and root plant parts. Three

plants per accession (i.e., three biological replicates) were used for estimating the phenotypic traits. For phenotyping of root and leaf morphometric traits, the protocols and criteria for distinctness, uniformity, and stability tests proposed by the International Union for the Protection of New Varieties of Plants (UPOV) were used (UPOV Code: BETAA_VUL_GVC). Based on these criteria, a total of 13 traits were phenotyped. For root traits, root shape (measured along the longitudinal axis) was categorized as transverse narrow elliptic (numerically coded as 1), transverse medium elliptic (2), circular (3), obovate (4), narrow oblong (5), or very narrow obovate (6); 'root tip shape' was scored as pointed (1), rounded (2), flat (3), or depressed (4); 'root external color' was classified as yellow, white, or reddish purple; 'root fresh weight' was determined in a scale and expressed in grams; root length and width were expressed in centimeters (cm), and 'root length-width ratio' was also calculated. For root shape, the mean length/width ratios for the different categories were as follows: 'transverse narrow elliptic' (1) = 0.40; transverse medium elliptic (2) = 0.60; circular (3) = 1.00; obovate (4) = 1.80; narrow oblong (5) = 2.60; very narrow oblong (6) = 3.50. Root sugar concentration was estimated using the polarimetric method proposed for sugar beet by the International Commission for Uniform Methods of Sugar Analysis (ICUMSA) [126] and values were expressed as percentage of fresh weight. For leaf traits, 'leaf blade color' was visually scored as green, green and red (green/red), or red; 'leaf shape' was scored as narrow elliptic, medium elliptic, or broad elliptic; 'leaf veins red color intensity' was scored as weak, medium, strong, or very strong; 'leaf margin undulations' categories were weak, medium, strong, or very strong; and 'leaf length' was measured and expressed in cm.

Genomic DNA isolation was performed from young and visually healthy leaves following the CTAB protocol [127] with modifications proposed by the Diversity Arrays Technology (DArT) (https://ordering.diversityarrays.com/files/DArT_DNA_isolation.pdf). The DNA integrity was checked by 0.8% agarose gel electrophoresis and its quantitation and quality assessment (260/280 and 260/230 ratios) were performed using a Nanodrop spectrophotometer (DeNovix DS-11 FX, USA). DNA samples were diluted to a final concentration of 50 ng μl^{-1} and sent to DArT for genotyping by sequencing (GBS) analysis.

Genotyping by sequencing

A high-throughput genotyping method called DArTSeq™ available at the Diversity Arrays Technology Pty Ltd. (Canberra, Australia) was used to perform GBS analysis for the sugar beet germplasm. Detailed information about the GBS protocol used can be found in our previous studies [6, 128]. This technique is based on reducing the complexity of the genomic DNA through the use of *Pst*I restriction enzyme to enrich genomic representations with single-copy sequences. Thus, DNA samples were *Pst*I-digested and then ligated to adapters following the protocol of Kilian et al. [129]. Diversity Arrays Technology provides both SNP and presence-absence-based variants known as silico-DArT markers. Scoring of the resulting markers was obtained through the DArTsoft14 software plugin in KDCompute application (http://www.kddart.org/ kdcom pute.html).

Genetic diversity analyses

Data from SNP and SilicoDArT markers were analyzed independently. Raw data were loaded and filtered in R software [130] version 4.2 using dartR package v2 [131] using the following criteria. All SNPs and silicoDArT markers that had > 5% missing data were removed, as well as markers missing in all individuals of at least one population, considering as populations the countries of origin of the accessions. Markers with a reproducibility score (RepAvg) < 100% were also removed, as well as those that originated from the same DNA fragment, as they were considered redundant (not informative). SNPs with a minor allele frequency (MAF) lower than 5% were also discarded. The resulting SNPs and silicoDArT data were used for genetic analyses in the sugar beet germplasm collection.

Simple agglomerative hierarchical clustering was performed using poppr R package [132, 133]. Pair-wise genetic dissimilarity (GD) values using Hamming distance were calculated among the accessions with the 'bitwise.dist' function. Following the calculation of GD values, a distance matrix was generated and used to construct dendrograms using the Unweighted Paired Group Method with Arithmetic means (UPGMA) with 'aboot' and visualized using the package 'ggtree' [134]. Principal coordinate analysis (PCoA) was performed using 'gl.pcoa', a wrapper function implemented in dartR v2, and the first two principal coordinates were plotted. The genetic structure of the populations was assessed with Bayesian clustering algorithms of the fastSTRUCTURE software [135], an implementation of STRUCTURE [136] specifically made to handle genomic SNP matrix data. Distruct barplots were constructed in R using the package 'pophelper' [137]. Selection of the optimum number of populations (K) was done using the *post hoc* methods proposed previously [138], by running fastSTRUCTURE using 50,000 burn-in steps and 50,000 MCMC steps with 100 replicates of K ranging from 1 to 12, and the most parsimonious model was selected based on their mean likelihood and their delta K values. Analysis of molecular variance were performed using pegas AMOVA as implemented in dartR [131] using (i) countries and (ii) clusters inferred from UPGMA tree as subpopulations.

General genetic statistics were independently calculated for populations (countries) and UPGMA genetic clusters using the 'popgen' function implemented in the snpReady package [139, 140].

GWAS analysis

A total of 5,645 GBS-derived SNP markers were used for GWAS analysis. Only SNP markers were used for GWAS, as these are codominant and more informative markers than silicoDArTs, which are dominant. Association analysis between SNP genotypes and 13 phenotypic traits was conducted using TASSEL (Trait Analysis by Association, Evolution, and Linkage, version 5.2.82) with a General Linear Model (GLM) approach, using as reference genome sequence the EL10_1.0 assembly of the sugar beet inbred line EL10 [141]. The selected traits are of agronomic and industrial importance, among which eight correspond to the tap root (i.e., RFW, RW, RL, RLWR, RS, RTS, RSC, and REC) and the remaining five are leaf traits (i.e., LL, LS, LBC, LVRCI, and LMU). Population structure was accounted for by including the first three principal components derived from a principal component analysis (PCA) as covariates in the model. To enhance result stability and control for false positives, 100 permutations were applied in the association analysis. Manhattan plots were generated to visualize the association signals for each trait. False discovery rate (FDR) and the Bonferroni correction were calculated, resulting in -log10(p) values of 5.05 and 7.00, respectively, and both thresholds were indicated in Fig. 4 to visualize which markers surpass each threashold, indicating different stringency levels. For estimation of the FDR, the Benjamini-Hochberg procedure was applied, and an FDR adjusted p-value (q-value) of 0.05 was considered as the threshold for declaring significant marker-trait associations. Qualitative phenotypic traits were encoded numerically to facilitate the association analysis. Leaf blade color was categorized as follows: green (1), green/red (2), and red (3). Leaf blade shape was represented by narrow (1), medium (2), and broad (3). For leaf vein red color intensity, categories included absence of red coloration (0), weak coloration (1), medium (2), strong (3), and very strong (4). Leaf margin undulations was scored as absence of undulations (0), weak undulations (1), medium (2), and strong (3). Finally, root external color was encoded as white (0), yellow (1), and reddish purple (2). These numerical codes allowed for the standardized input of qualitative traits within the association model, enabling consistent analysis across phenotypes.

Other bioinformatic analyses

To elucidate transcriptional modulatory pathways associated with LOC104883815, a candidate gene for RFW, a cis-acting regulatory element analysis was carried out in the 2000 bp region upstream of the transcription start site and in introns of this gene, using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; accessed on 17 November 2024) [142].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06525-7.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

The accessions used in this study were provided by the United States Department of Agriculture (USDA).

Author contributions

M.Y., F.S.B., M.A.N., and P.F.C. conceived and directed the study. M.Y., N.M.B.A., and M.A.N. conducted most of the experiments, including growing plants in the field, plant phenotyping, tissue sampling, and DNA isolation and preparation for sequencing. A.M., J.W., and P.F.C. conducted bioinformatics analyses and and prepared Figs. 1, 2, 3 and 4. M.Y., M.A.N., N.M.B.A., and P.F.C. wrote the original draft. M.Y., M.A.N., P.F.C., D.G., F.S.B., M.T., M.K., and Y.S.C. reviewed and edited the manuscript. MY supervised the project. All the authors have read approved the final version of the manuscript.

Funding

Scientific Research Projects Unit (BAP) of Van Yüzüncü Yıl Üniversitesi, project FBA-2022-10150. This research was also supported by the Brain Pool Plus program, funded by the Ministry of Science and ICT, National Research Foundation of Korea (RS-2024-00403759) to Faheem Shehzad Baloch.

Data availability

The datasets generated and/or analysed during the current study are available in the GitHub repository, with direct web links to SNP and silicoDArT data at https://github.com/JosefinaWohlfeiler/gl_snp_unfiltered.EL10_v1.0.vcf.git and https://github.com/JosefinaWohlfeiler/Report_DSb22-7501_SilicoDArT_1.csv, respectively. Additional information is presented as supplementary data in the Additional files 1-3.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Agricultural Biotechnology, Faculty of Agriculture, Van Yuzuncu Yil University, Van 65080, Turkey
[2]Department of Biotechnology, Faculty of Science, Mersin University, Yenişehir, Mersin 33343, Turkey
[3]Department of Field Crops, Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, Sivas 58140, Turkey
[4]Instituto Nacional de Tecnología Agropecuaria (INTA) Estación Experimental Agropecuaria La Consulta, La Consulta M5567, Argentina
[5]Facultad de Ciencias Agrarias, Laboratorio de Biología Molecular, Universidad Nacional de Cuyo (UNCuyo), Instituto de Biología Agrícola de Mendoza (IBAM CONICET, Luján de Cuyo M5534, Argentina

[6]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto Nacional de Tecnología Agropecuaria (INTA) Estación Experimental Agropecuaria Mendoza, Luján de Cuyo M5534, Argentina
[7]Department of Plant Resources and Environment, Jeju National University, Jeju 63243, Republic of Korea
[8]Department of Field Crops, Faculty of Agriculture, Van Yuzuncu Yil University, Van 65090, Turkey
[9]Department of Plant Biology and Biotechnology, Faculty of Biotechnology and Horticulture, University of Agriculture in Krakow, Krakow, Poland

Received: 10 December 2024 / Accepted: 8 April 2025
Published online: 24 April 2025

References

1. Singh VK, Singh AK, Singh S, Singh BD. Next-Generation sequencing (NGS) tools and impact in plant breeding. In: Al-Khayri J, Jain S, Johnson D, editors. Advances in plant breeding strategies: breeding, biotechnology and molecular tools. Cham: Springer; 2015. pp. 563–612.
2. Bräutigam A, Gowik U. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biol. 2010;12:831–41.
3. Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian G. Diversity Arrays Technology (DArT): A generic high density genotyping platform. In: 7th International Safflower Conference, Berlin, Germany. 2008.
4. Zhang L, Liu D, Guo X, Yang W, Sun J, Wang D, Sourdille P, Zhang A. Investigation of genetic diversity and population structure of common wheat cultivars in Northern China using dart markers. BMC Genet. 2011;12:42.
5. Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A. Diversity arrays technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci U S A. 2004;101:9915–20.
6. Nadeem MA, Habyarimana E, Çiftçi V, Nawaz MA, Karaköy T, Comertpay G, Shahid MQ, Hatipoğlu R, Yeken MZ, Ali F, Ercişli S, Chung G, Baloch FS. Characterization of genetic diversity in Turkish common bean gene pool using phenotypic and whole-genome DArTseq-generated silicodart marker information. PLoS ONE. 2018;13:e0205363.
7. Ali F, Nadeem MA, Barut M, Habyarimana E, Chaudhary HJ, Khalil IH, Alsaleh A, Hatipoğlu R, Karaköy T, Kurt C, Aasim M. Genetic diversity, population structure and marker-trait association for 100-seed weight in international safflower panel using silicodart marker information. Plants. 2020;9:652.
8. Alam M, Neal J, O'Connor K, Kilian A, Topp B. Ultra-high-throughput DArTseq-based silicodart and SNP markers for genomic studies in macadamia. PLoS ONE. 2018;13:e0203465.
9. Tomkowiak A, Bocianowski J, Spychała J, Grynia J, Sobiech A, Kowalczewski PŁ. DArTseq-based high-throughput silicodart and SNP markers applied for association mapping of genes related to maize morphology. Int J Mol Sci. 2021;22:5840.
10. Baloch FS, Nadeem MA. Unlocking the genomic regions associated with seed protein contents in Turkish common bean germplasm through genome-wide association study. Turk J Agric for. 2022;46:113–28.
11. Adu BG, Akromah R, Amoah S, Nyadanu D, Yeboah A, Aboagye LM, Amoah RA, Owusu EG. High-density DArT-based silicodart and SNP markers for genetic diversity and population structure studies in cassava (Manihot esculenta Crantz). PLoS ONE. 2021;16:e0255290.
12. Odesola KA, Olawuyi OJ, Paliwal R, Oyatomi OA, Abberton MT. Genome-Wide association analysis of phenotypic traits in Bambara groundnut under drought-stressed and non-stressed conditions based on DArTseq SNP. Front Plant Sci. 2023;14:1104417.
13. Thomson MJ. High-throughput Snp genotyping to accelerate crop improvement. Plant Breed Biotech. 2014;2:195–212.
14. Shavrukov Y. Comparison of SNP and CAPS markers application in genetic research in wheat and barley. BMC Plant Biol. 2016;16(Suppl 1):11.
15. Zhang X, Zhang H, Li L, Lan H, Ren Z, Liu D, Wu L, Liu H, Jaqueth J, Li B, Pan G, Gao S. Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. BMC Genomics. 2016;17:697.
16. Kumar D, Chhokar V, Sheoran S, Singh R, Sharma P, Jaiswal S, Iquebal MA, Jaiswar A, Jaisri J, Angadi UB, Rai A, Singh GP, Kumar D, Tiwari R. Characterization of genetic diversity and population structure in wheat using array based SNP markers. Mol Biol Rep. 2020;47:293–306.
17. Abu Zaitoun SY, Jamous RM, Shtaya MJ, Mallah OB, Eid IS, Ali-Shtayeh MS. Characterizing Palestinian snake Melon (Cucumis Melo Var. flexuosus) germplasm diversity and structure using SNP and DArTseq markers. BMC Plant Biol. 2018;18:246.
18. Taranto F, D'Agostino N, Greco B, Cardi T, Tripodi P. Genome-wide SNP discovery and population structure analysis in pepper (Capsicum annuum) using genotyping by sequencing. BMC Genomics. 2016;17:943.
19. Borevitz JO, Nordborg M. The impact of genomics on the study of natural variation in Arabidopsis. Plant Physiol. 2003;132:718–25.
20. Tibbs Cortes L, Zhang Z, Yu J. Status and prospects of genome-wide association studies in plants. Plant Genome. 2021;14:e20077.
21. Galewski P, McGrath JM. Genetic diversity among cultivated beets (Beta vulgaris) assessed via population-based whole genome sequences. BMC Genomics. 2020;21:189.
22. Lehner R, Blazek L, Minoche AE, Dohm JC, Himmelbauer H. Assembly and characterization of the genome of Chard (Beta Vulgaris Ssp. Vulgaris Var. cicla). J Biotech. 2021;333:67–76.
23. Bangar SP, Singh A, Chaudhary V, Sharma N, Lorenzo JM. Beetroot as a novel ingredient for its versatile food applications. Crit Rev Food Sci Nutr. 2022;63:8403–27.
24. Maravić N, Teslić N, Nikolić D, Dimić I, Šereš Z, Pavlić B. From agricultural waste to antioxidant-rich extracts: green techniques in extraction of polyphenols from sugar beet leaves. Sustain Chem Pharm. 2022;28:100728.
25. Mall AK, Misra V, Santeshwari, Pathak AD, Srivastava S. Sugar beet cultivation in India: prospects for bio-ethanol production and value-added co-products. Sugar Tech. 2021;23:1218–34.
26. Ebmeyer H, Fiedler-Wiechers K, Hoffmann CM. Drought tolerance of sugar beet - Evaluation of genotypic differences in yield potential and yield stability under varying environmental conditions. Eur J Agron. 2021;125:126262.
27. Qi A. Sugar beet production under changing climate: opportunities and challenges. In: Misra V, Srivastava S, Mall AK, editors. Sugar beet cultivation, management and processing. Singapore: Springer; 2022. pp. 407–27.
28. Vladu M, Tudor VC, Mărcuță L, Mihai D, Tudor AD. Study on the production and valorization of sugar beet in the European union. Rom Agric Res. 2021;38.
29. Wiebe K, Robinson S, Cattaneo A. Chapter 4 - Climate change, agriculture and food security: impacts and the potential for adaptation and mitigation. In: Campanhola C, Pandey S, editors. Sustainable food and agriculture. Rome: The Food and Agriculture Organization of the United Nations and Elsevier Inc; 2019. pp. 55–74.
30. Nadeem MA. Deciphering the genetic diversity and population structure of Turkish bread wheat germplasm using iPBS-retrotransposons markers. Mol Biol Rep. 2021;48:6739–48.
31. Yildiz M, Koçak M, Nadeem MA, Cavagnaro P, Barboza K, Baloch FS, Keleş D. Genetic diversity analysis in the Turkish pepper germplasm using iPBS retrotransposon based markers. Turk J Agric for. 2020;44:1–14.
32. Zia MAB, Demirel U, Nadeem MA, Çaliskan ME. Genome-wide association study identifies various loci underlying agronomic and morphological traits in diversified potato panel. Physiol Mol Biol Plants. 2020;26:1003–20.
33. Adetunji I, Willems G, Tschoep H, Bürkholz A, Barnes S, Boer M, Malosetti M, Horemans S, van Eeuwijk F. Genetic diversity and linkage disequilibrium analysis in elite sugar beet breeding lines and wild beet accessions. Theor Appl Genet. 2014;127:559–71.
34. Andrello M, Henry K, Devaux P, Desprez B, Manel S. Taxonomic, Spatial and adaptive genetic variation of Beta section Beta. Theor Appl Genet. 2016;129:257–71.
35. Liang XM, Pi Z, Wu ZD, Li SN. Constructing DNA fingerprinting and evaluating genetic diversity among sugar beet (Beta vulgaris L.) varieties based on four molecular markers. Sugar Tech. 2023;25:1361–73.
36. Stevanato P, Broccanello C, Biscarini F, Del Corvo M, Sablok G, Panella L, Stella A, Concheri G. High-throughput RAD-SNP genotyping for characterization of sugar beet genotypes. Plant Mol Biol Rep. 2014;32:691–6.
37. Hoffmann CM, Huijbregts T, van Swaaij N, Jansen R. Impact of different environments in Europe on yield and quality of sugar beet genotypes. Eur J Agron. 2009;30:17–26.
38. Srivastava S, Pathak AD, Kumar R, Joshi BB. Genetic diversity of sugar beet genotypes evaluated by microsatellite DNA markers. J Environ Biol. 2017;38:777–83.

39. Peng F, Pi Z, Li S, Wu Z. Genetic diversity and population structure analysis of excellent sugar beet (Beta vulgaris L.) germplasm resources. Horticulturae. 2024;10:120.
40. Sun B, Li S, Pi Z, Wu Z, Wang R. Assessment of genetic diversity and population structure of exotic sugar beet (Beta vulgaris L.) varieties using three molecular markers. Plants. 2024;13:2954.
41. Abbasi Z, Arzani A, Majidi MM. Evaluation of genetic diversity of sugar beet (Beta vulgaris L.) crossing parents using agro-morphological traits and molecular markers. J Agric Sci Technol. 2014;16:1397–411.
42. Simko I, Eujayl I, van Hintum TJL. Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. Plant Sci. 2012;184:54–62.
43. Rani J, Gulia V, Dhanda N, Kapoor M. Phenotypic characterization of periwinkle (*Catharanthus roseus* L. G. Don.) for the selection of promising materials for breeding programs. Discov Plants. 2024;1:34.
44. Kaffka SR, Grantz DA. Sugar crops. (Editor). Encyclopedia of agriculture and food systems. Volume 5. San Diego: Elsevier;. Academic; 2014. pp. 240–60. Neal Van Alfen.
45. Wang G, Chen Q, Yang Y, Duan Y, Yang Y. Exchanges of economic plants along the land silk road. BMC Plant Biol. 2022;22:619.
46. Trujillo M. News from the PUB: plant U-box type E3 ubiquitin ligases. J Exp Bot. 2018;69:371–84.
47. Kinoshita A, ten Hove CA, Tabata R, Yamada M, Shimizu N, Ishida T, Yamaguchi K, Shigenobu S, Takebayashi Y, Iuchi S, Kobayashi M, Kurata T, Wada T, Seo M, Hasebe M, Blilou I, Fukuda H, Scheres B, Heidstra R, Kamiya Y, Sawa S. A plant U-box protein, PUB4, regulates asymmetric cell division and cell proliferation in the root meristem. Development. 2015;142:444–53.
48. Deb S, Sankaranarayanan S, Wewala G, Widdup E, Samuel MA. The S-domain receptor kinase arabidopsis receptor Kinase2 and the U box/armadillo repeat-containing E3 ubiquitin Ligase9 module mediates lateral root development under phosphate starvation in Arabidopsis. Plant Physiol. 2014;165:1647–56.
49. Li X, Zhu L, Wu Z, Chen J, Wang T, Zhang X, Mei G, Wang J, Lv G. Classification and expression profile of the U-box E3 Bbiquitin ligase enzyme gene family in maize (Zea mays L). Plants. 2022;11:2459.
50. Li X, Xu J, Duan S, Zhang J, Bian C, Hu J, Li G, Jin L. Mapping and QTL analysis of early-maturity traits in tetraploid potato (Solanum tuberosum L). Int J Mol Sci. 2018;19:3065.
51. Dreher K, Callis J. Ubiquitin, hormones and biotic stress in plants. Ann Bot. 2007;99:787–822.
52. Liu YB. Cytokinin and auxin-regulated root meristem establishment during *Arabidopsis* somatic embryogenesis. Ph.D. Thesis, Shandong Agricultural University, Taian, China, 2011.
53. Baluška F, Parker J, Barlow PA. Role for gibberellic acid in orienting microtubules and regulating cell growth Polarity in the maize root cortex. Planta. 1993;191:149–57.
54. Liu DT, Jing YP, Shi HX, Zhong TT, Wang Z. Impact of IAA, GA3, and ABA on negative root phototropism and root growth of rice. Acta Agron Sin. 2014;40:1658–66.
55. Li PC, Li K, Wang J, Zhao CZ, Zhao SZ, Hou L, Xia H, Ma CL, Wang XJ. The AAA-ATPase MIDASIN 1 functions in ribosome biogenesis and is essential for embryo and root development. Plant Physiol. 2019;180:289–304.
56. Li PC, Yu SW, Li K, Huang JG, Wang XJ, Zheng CC. The mutation of Glu at amino acid 3838 of AtMDN1 provokes pleiotropic developmental phenotypes in Arabidopsis. Sci Rep. 2016;6:36446.
57. Qi W, Zhu J, Wu Q, Wang Q, Li X, Yao D, Jin Y, Wang G, Wang G, Song R. Maize reas1 mutant stimulates ribosome use efficiency and triggers distinct transcriptional and translational responses. Plant Physiol. 2016;170:971–88.
58. Byrne ME. A role for the ribosome in development. Trends Plant Sci. 2009;14:512–9.
59. Chen D-H, Huang Y, Jiang C, Si J-P. Chromatin-based regulation of plant root development. Front Plant Sci. 2018;9:1509.
60. Bessman MJ, Frick DN, O'Handley SF. The mutt proteins or nudix hydrolases, a family of versatile, widely distributed, housecleaning enzymes. J Biol Chem. 1996;271:25059–62.
61. Liu D, Tan W, Wang H, Li W, Fu J, Li J, Zhou Y, Lin M, Xing W. Genetic diversity and genome-wide association study of 13 agronomic traits in Beta vulgaris L. germplasms. BMC Genomics. 2023;24:413.
62. Corpas FJ, Aguayo-Trinidad S, Ogawa T, Yoshimura K, Shigeoka S. Activation of NADPH-recycling systems in leaves and roots of Arabidopsis thaliana under arsenic-induced stress conditions is accelerated by knock-out of nudix hydrolase 19 (AtNUDX19) gene. J Plant Physiol. 2016;192:81–9.
63. Bararyenya A, Olukolu BA, Tukamuhabwa P, Grüneberg WJ, Ekaya W, Low J, Ochwo-Ssemakula M, Odong TL, Talwana H, Badji A, Kyalo M, Nasser Y, Gemenet D, Kitavi M, Mwanga ROM. Genome-wide association study identified candidate genes controlling continuous storage root formation and bulking in hexaploid Sweetpotato. BMC Plant Biol. 2020;20:3.
64. Feng YZ, Zhu QF, Xue J, Chen P, Yu Y. Shining in the dark: the big world of small peptides in plants. aBIOTECH. 2023;4:238–56.
65. Yang L, Hu X, Ren M, Ma F, Fu J, Cui H. Stem-cell-expressed DEVIL-like small peptides maintain root growth under abiotic stress via abscisic acid signaling. Plant Physiol. 2024;194:2372–86.
66. Wang C, Wang T, Liu M, Zhang S, Wuet S. Small peptide SiDVL/RTFLs from Foxtail millet inhibit root growth through repressing auxin signaling in Transgenic Arabidopsis. Plant Cell Rep. 2024:43268.
67. Guo P, Yoshimura A, Ishikawa N, Yamaguchi T, Guo Y, Tsukaya H. Comparative analysis of the RTFL peptide family on the control of plant organogenesis. J Plant Res. 2015;128:497–510.
68. Ugalde JM, Lamig L, Herrera-Vásquez A, Fuchs P, Homagk M, Kopriva S, Müller-Schüssele SJ, Holuigue L, Meyer AJ. A dual role for glutathione transferase U7 in plant growth and protection from Methyl viologen-induced oxidative stress. Plant Physiol. 2021;187:2451–68.
69. Tsukagoshi H, Busch W, Benfey PN. Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. Cell. 2010;143:606–16.
70. Mase K, Tsukagoshi H. Reactive oxygen species link gene regulatory networks during Arabidopsis root development. Front Plant Sci. 2021;12:660274.
71. Eljebbawi A, Guerrero YDCR, Dunand C, Estevez JM. Highlighting reactive oxygen species as multitaskers in root development. iScience. 2021;24:101978.
72. Fan Y, Chen T, Xue L, Zhang H, Gao S, Zhao N, He S, Zhai H, Liu Q. A glutathione S-transferase IbGSTL2 interacts with IbcPGM to increase starch content and improve starch quality in sweet potato. Crop J. 2024. https://doi.org/10.1016/j.cj.2024.08.007.
73. Majda M, Robert S. The role of auxin in cell wall expansion. Int J Mol Sci. 2018;19:951.
74. Muthusamy M, Kim JY, Yoon EK, Kim JA, Lee SI. BrEXLB1, a Brassica rapa expansin-like b1 gene is associated with root development, drought stress response, and seed germination. Genes. 2020;11:404.
75. Lee DK, Ahn JH, Song SK, Choi YD, Lee JS. Expression of an expansin gene is correlated with root elongation in soybean. Plant Physiol. 2003;131:985–97.
76. Buendía-Monreal M, Gillmor CS, Mediator. A key regulator of plant development. Dev Biol. 2016;419:7–18.
77. Li Q, Wang Y, Sun Z, Li H, Liu H. The biosynthesis process of small RNA and its pivotal roles in plant development. Int J Mol Sci. 2024;25:7680.
78. Kumar V, Thakur JK, Prasad M. Histone acetylation dynamics regulating plant development and stress responses. Cell Mol Life Sci. 2021;78:4467–86.
79. Theodoulou FL. Plant ABC transporters. Biochim. Biophys Acta. 2000;1465:79–103.
80. Francisco RM, Regalado A, Ageorges A, Burla BJ, Bassin B, Eisenach C, Zarrouk O, Vialet S, Marlin T, Chaves MM, Martinoia E, Nagy R. ABCC1, an ATP binding cassette protein from grape Berry, transports Anthocyanidin 3-O-glucosides. Plant Cell. 2013;25:1840–54.
81. Xiang Y, Huang H-Y, Zhao Y-W, Wang C-K, Sun Q, Hu D-G. Role of an ATP binding cassette (ABC) transporter MdABCI17 in the anthocyanin accumulation of Apple. Scientia Hort. 2024;323:112502.
82. Bannoud F, Carvajal S, Ellison S, Senalik D, Gomez Talquenca S, Iorizzo M, Simon PW, Cavagnaro PF. Genetic and transcription profile analysis of tissue-specific anthocyanin pigmentation in Carrot root phloem. Genes. 2021;12:1464.
83. Stafford HA. Anthocyanins and betalains: evolution of the mutually exclusive pathways. Plant Sci. 1994;101:91–8.
84. Polturak G, Heinig U, Grossman N, Battat M, Leshkowitz D, Malitsky S, Rogachev I, Aharoni A. Transcriptome and metabolic profiling provides insights into betalain biosynthesis and evolution in Mirabilis jalapa. Mol Plant. 2018;11:189–204.
85. Carabelli M, Turchi L, Ruzza V, Morelli G, Ruberti I. Homeodomain-leucine zipper II family of transcription factors to the limelight: central regulators of plant development. Plant Signal Behav. 2013;8:e25447.
86. Bou-Torrent J, Salla-Martret M, Brandt R, Musielak T, Palauqui JC, Martínez-García JF, Wenkel S. ATHB4 and HAT3, two class II HD-ZIP transcription factors, control leaf development in Arabidopsis. Plant Signal Behav. 2012;7:1382–7.

87. Sorin C, Salla-Martret M, Bou-Torrent J, Roig-Villanova I, Martínez-García JF. ATHB4, a regulator of shade avoidance, modulates hormone response in Arabidopsis seedlings. Plant J. 2009;59:266–77.
88. Qin YM, Hu CY, Pang Y, Kastaniotis AJ, Hiltunen JK, Zhu YX. Saturated very-long-chain fatty acids promote cotton fiber and Arabidopsis cell elongation by activating ethylene biosynthesis. Plant Cell. 2007;19:3692–704.
89. Yang T, Li Y, Liu Y, He L, Liu A, Wen J, Mysore KS, Tadege M, Chen J. The 3-keto-acyl-CoA synthase WFL is involved in lateral organ development and cuticular wax synthesis in Medicago truncatula. Plant Mol Biol. 2021;105:193–204.
90. Bach L, Michaelson LV, Haslam R, Bellec Y, Gissot L, Marion J, Da Costa M, Boutin J-P, Miquel M, Tellier M, Domergue F, Markham JE, Beaudoin F, Napier JA, Faure J-D. The very-long-chain hydroxy fatty acyl-CoA dehydratase PASTICCINO2 is essential and limiting for plant development. Proc Natl Acad Sci USA. 2008;105:14727–31.
91. Lolle SJ, Cheung AI, Sussex IM. Fiddlehead: an Arabidopsis mutant constitutively expressing an organ fusion program that involves interactions between epidermal cells. Dev Biol. 1992;152:383–92.
92. Nobusawa T, Okushima Y, Nagata N, Kojima M, Sakakibara H, Umeda M. Synthesis of very-long-chain fatty acids in the epidermis controls plant organ growth by restricting cell proliferation. PLoS Biol. 2013;11:e1001531.
93. Hsieh WY, Liao JC, Hsieh MH. Dysfunctional mitochondria regulate the size of root apical meristem and leaf development in Arabidopsis. Plant Signal Behav. 2015;10:e1071002.
94. Yu YH, Li XF, Yang SD, Li SQ, Meng XX, Liu HN, Pei MS, Wei TL, Zhang YJ, Guo DL. Overexpression of VvPPR1, a DYW-type PPR protein in grape, affects the phenotype of Arabidopsis thaliana leaves. Plant Physiol Biochem. 2021;164:195–204.
95. Fang X, Zhu Z, Li J, Wang X, Wei C, Zhang X, Dai Z, Liu S, Luan F. Identification of chromosomal regions and candidate genes for round leaf locus in Cucumis melo L. Plants. 2024;13:1134.
96. Matsumura Y, Ohbayashi I, Takahashi H, Kojima S, Ishibashi N, Keta S, Nakagawa A, Hayashi R, Saéz-Vásquez J, Echeverria M, Sugiyama M, Nakamura K, Machida C, Machida Y. A genetic link between epigenetic repressor AS1-AS2 and a putative small subunit processome in leaf Polarity establishment of Arabidopsis. Biol Open. 2016;5:942–54.
97. De Smet I, Voß U, Jürgens G, Beeckman T. Receptor-like kinases shape the plant. Nat Cell Biol. 2009;11:1166–73.
98. Han HJ, Park HC, Byun HJ, Lee SM, Kim HS, Yun DJ, Cho MJ, Chung WS. The transcriptional repressor activity of ASYMMETRIC LEAVES1 is inhibited by direct interaction with calmodulin in Arabidopsis. Plant Cell Environ. 2012;35:1969–82.
99. Bürstenbinder K, Möller B, Plötner R, Stamm G, Hause G, Mitra D, Abel S. The IQD family of calmodulin-binding proteins links calcium signaling to microtubules, membrane subdomains, and the nucleus. Plant Physiol. 2017;173:1692–708.
100. Mitra D, Klemm S, Kumari P, Quegwer J, Möller B, Poeschl Y, Pflug P, Stamm G, Abel S, Bürstenbinder K. Microtubule-associated protein IQ67 DOMAIN5 regulates morphogenesis of leaf pavement cells in Arabidopsis thaliana. J Exp Bot. 2019;70:529–43.
101. Wu S, Xiao H, Cabrera A, van der Meulia T. SUN regulates vegetative and reproductive organ shape by changing cell division patterns. Plant Physiol. 2011;157:1175–86.
102. Larue CT, Wen J, Walker JC. A microRNA-transcription factor module regulates lateral organ size and patterning in Arabidopsis. Plant J. 2009;58:450–63.
103. Berger Y, Harpaz-Saad S, Brand A, Melnik H, Sirding N, Alvarez JP, Zinder M, Samach A, Eshed Y, Ori N. The NAC-domain transcription factor GOBLET specifies leaflet boundaries in compound tomato leaves. Development. 2009;136:823–32.
104. Valoroso MC, Lucibelli F, Aceto S. Orchid NAC transcription factors: a focused analysis of CUPULIFORMIS genes. Genes. 2022;13:2293.
105. Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell. 1996;85:159–70.
106. Weir I, Lu J, Cook H, Causier B, Schwarz-Sommer Z, Davies B. CUPULIFORMIS establishes lateral organ boundaries in Antirrhinum. Development. 2004;131:915–22.
107. Cheng Z, Zhu Y, He X, Fan G, Jiang J, Jiang T, Zhang X. Transcription factor PagERF110 inhibits leaf development by direct regulating PagHB16 in Poplar. Plant Sci. 2025;350:112309.
108. Jiang F, Guo M, Yang F, Duncan K, Jackson D, Rafalski A, Wang S, Li B. Mutations in an AP2 transcription factor-like gene affect internode length and leaf shape in maize. PLoS ONE. 2012;7:e37040.
109. Hughes TE, Langdale JA. SCARECROW is deployed in distinct contexts during rice and maize leaf development. Development. 2022;149:dev200410.
110. Mohanasundaram B, Palit S, Bhide AJ, Pala M, Rajoria K, Girigosavi P, Banerjee AK. PpSCARECROW1 (PpSCR1) regulates leaf blade and mid-vein development in Physcomitrium patens. Plant Mol Biol. 2024;114:12.
111. Ji N, Wang Q, Li S, Wen J, Wang L, Ding X, Zhao S, Feng H. Metabolic profile and transcriptome reveal the mystery of petal blotch formation in Rose. BMC Plant Biol. 2023;23:46.
112. Skirycz A, Jozefczuk S, Stobiecki M, Muth D, Zanor MI, Witt I, Mueller-Roeber B. Transcription factor AtDOF4;2 affects phenylpropanoid metabolism in Arabidopsis thaliana. New Phytol. 2007;175:425–38.
113. Shimada S, Otsuki H, Sakuta M. Transcriptional control of anthocyanin biosynthetic genes in the Caryophyllales. J Exp Bot. 2007;58:957–67.
114. Hatlestad GJ, Akhavan NA, Sunnadeniya RM, Elam L, Cargile S, Hembd A, Gonzalez A, McGrath JM, Lloyd AM. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. Nat Genet. 2014;47:92–6.
115. Cai X, Xu C, Wang X, Wang S, Zhang Z, Fei Z, Wang Q. Construction of genetic linkage map using genotyping-by-sequencing and identification of QTLs associated with leaf color in spinach. Euphytica. 2018;214:217.
116. Chen J, Xie F, Shah K, Chen C, Zeng J, Chen J, Zhang Z, Zhao J, Hu G, Qin Y. Identification of HubHLH family and key role of HubHLH159 in betalain biosynthesis by activating the transcription of HuADH1, HuCYP76AD1-1, and HuDODA1 in Pitaya. Plant Sci. 2023;328:111595.
117. Cheng M-N, Huang Z-J, Hua Q-Z, Shan W, Kuang J-F, Lu W-J, Qin Y-H, Chen J-Y. The WRKY transcription factor HpWRKY44 regulates CytP450-like1 expression in red Pitaya fruit (Hylocereus polyrhizus). Hort Res. 2017;4:17039.
118. Xie F, Chen C, Chen J, Chen J, Hua Q, Shah K, Zhang Z, Zhao J, Hu G, Chen J, Qin Y. Betalain biosynthesis in red pulp Pitaya is regulated via HuMYB132: a R-R type MYB transcription factor. BMC Plant Biol. 2023;23:28.
119. Xie F, Hua Q, Chen C, Zhang Z, Zhang R, Zhao J, Hu G, Chen J, Qin Y. Genome-Wide characterization of R2R3-MYB transcription factors in Pitaya reveals a R2R3-MYB repressor HuMYB1 involved in fruit ripening through regulation of betalain biosynthesis by repressing betalain biosynthesis-related genes. Cells. 2021;10:1949.
120. Takahashi K, Takamura E, Sakuta M. Isolation and expression analysis of two DOPA dioxygenases in Phytolacca americana. Z Naturforsch C J Biosci. 2009;64:564–73.
121. Wang Y, Wang Y, Zhou LJ, Peng J, Chen C, Liu S, Song A, Jiang J, Chen S, Chen F. CmNAC25 targets CmMYB6 to positively regulate anthocyanin biosynthesis during the post-flowering stage in chrysanthemum. BMC Biol. 2023;21:211.
122. Zhang S, Chen Y, Zhao L, Li C, Yu J, Li T, Yang W, Zhang S, Su H, Wang L. A novel NAC transcription factor, MdNAC42, regulates anthocyanin accumulation in red-fleshed Apple by interacting with MdMYB10. Tree Physiol. 2020;40:413–23.
123. Meng J, Sun S, Li A, Pan L, Duan W, Cui G, Xu J, Niu L, Wang Z, Zeng W. A NAC transcription factor, PpNAC1, regulates the expression of PpMYB10.1 to promote anthocyanin biosynthesis in the leaves of Peach trees in autumn. Hortic Adv. 2023;1:8.
124. Zhang J, Song B, Chen G, Yang G, Ming M, Zhang S, Xue Z, Han C, Li J, Wu J. Transcriptome analysis identified PyNAC42 as a positive regulator of anthocyanin biosynthesis induced by nitrogen deficiency in Pear (Pyrus spp). Horticulturae. 2024;10:980.
125. Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. Plant Cell Physiol. 2009;50:2210–22.
126. ICUMSA Method GS6-3. In: Bartens A, editor. Polarimetric sucrose content in sugar beet after clarification using aluminium sulphate– Official. Berlin:; 2024.
127. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13–5.
128. Baloch FS, Alsaleh A, Shahid MQ, Çiftçi V, Sáenz de Miera E, Aasim L, Nadeem M, Aktaş MA, Özkan H, Hatipoğlu H. A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. PLoS ONE. 2017;12:e0167821.
129. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Aschenbrenner-Kilian HC, Evers M, Peng M, Cayla K, Hok C, Uszynski P. Diversity arrays technology: a generic genome profiling technology on open platforms. In: Pompanon F, Bonin A, editors. Data production and analysis in population genomics. Methods in molecular biology. Volume 888. Totowa: Humana; 2012. pp. 67–89.
130. R Core Team. R: A language and environment for statistical computing. 2021. https://www.R-project.org. Accessed 15 March 2024.

131. Mijangos JL, Gruber B, Berry O, Pacioni C, Georges A. DartR v2: an accessible genetic analysis platform for conservation. Ecol Agric. 2022;13:2150–8.
132. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014;2:e281.
133. Kamvar ZN, Tabima JF, Brooks JC, Folarin D, Poppr. Genetic analysis of populations with mixed reproduction. R-packages. 2024. Available from: https://CRAN.R-project.org/package=poppr
134. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2016;8:28–36.
135. Raj A, Stephens M, Pritchard JK, fastSTRUCTURE. Variational inference of population structure in large SNP data sets. Genetics. 2014;197:573–89.
136. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
137. Francis RM. Pophelper: Tabulate, analyse and visualise admixture proportions. GitHub. 2024. Available from: https://github.com/royfrancis/pophelper
138. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: A simulation study. Mol Ecol. 2005;14:2611–20.
139. Granato ISC, Fritsche-Neto R, SnpReady. Preparing genotypic datasets in order to run genomic analysis. 2018.
140. Granato ISC, Galli G, de Oliveira Couto EG, e Souza MB, Mendonca LF, Fritsche-Neto R. SnpReady: A tool to assist breeders in genomic analysis. Mol Breed. 2018;38:102.
141. McGrath JM, Funk A, Galewski P, Ou S, Townsend B, Davenport K, Daligault H, Johnson S, Lee J, Hastie A, Darracq A, Willems G, Barnes S, Liachko I, Sullivan S, Koren S, Phillippy A, Wang J, Liu T, Pulman J, Childs K, Shu S, Yocum A, Fermin D, Mutasa-Göttgens E, Stevanato P, Taguchi K, Naegele R, Dorn KM. A contiguous de Novo genome assembly of sugar beet EL10 (Beta vulgaris L). DNA Res. 2023;30:dsac033.
142. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, de Peer YV, Rouzé P, Rombauts S. Plant CARE, a database of plant cis-acting regulatory elements and a portal to tools for in Silico analysis of promoter sequences. Nucleic Acids Res. 2002;30:325–7.

Publisher's note