Genome-wide association mapping reveals novel genes associated with coleoptile length in a worldwide collection of barley

Hao Luo
Murdoch University https://orcid.org/0000-0003-0992-611X
Camilla Beate Hill
Murdoch University
Gaofeng Zhou
Murdoch University
Xiao-Qi Zhang
Murdoch University
Chengdao Li (C.Li@murdoch.edu.au)

Research article

Keywords: Coleoptile length, deep seeding, barley, GWAS

Posted Date: March 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-18401/v1

Version of Record: A version of this preprint was published on July 22nd, 2020. See the published version at https://doi.org/10.1186/s12870-020-02547-5.

Abstract

Background

Drought is projected to become more frequent and severe in a changing climate, which requires deep sowing of crop seeds to reach soil moisture. Coleoptile length is a key agronomic trait in cereal crops such as barley, as long coleoptiles are linked to drought tolerance and improved seedling establishment under early water-limited growing conditions.

Results

In this study, we detected large genetic variation in a panel of 328 diverse barley (Hordeum vulgare L.) accessions. To understand the overall genetic basis of barley coleoptile length, all accessions were geminated in the dark and phenotyped for coleoptile length after 2 weeks. The investigated barleys had significant variation for coleoptile length. We then conducted genome-wide association studies (GWASs) with more than 30,000 molecular markers and identified 8 genes and 12 intergenic loci significantly associated with coleoptile length in our barley panel. The Squamosa promoter-binding-like protein 3 gene (SPL3) on chromosome 6H was identified as a major candidate gene. The missense variant on the second exon changed serine to alanine in the conserved SBP domain, which likely impacted its DNA-binding activity.

Conclusion

This study provides genetic loci for seedling coleoptile length along with candidate genes for future potential incorporation in breeding programmes to enhance early vigour and yield potential in water-limited environments.

Background

Geminating seedlings of monocotyledons have a coleoptile, which is a sheath-like tissue covering the primary leaf to protect the emerging shoot as it breaks through the soil to the surface. The coleoptile is essential for early crop establishment and its length determines the maximum depth at which the seed can be sown [28, 33, 38, 39, 50]. If seeds are sown at a depth greater than their coleoptile length, it may result in lower emergence rate, reduced early growth, fewer tiller numbers, and decreased grain yield [11, 38, 43]. In agricultural growing areas prone to drought, the topsoil moisture can be insufficient for seed germination, and seeds need to be sown deeper to access enough moisture [23, 44] and lower temperatures [25]. Therefore, the varieties with longer coleoptiles are preferable in water-limited growing regions; for example, winter wheat grown in the low water supply areas of the Pacific Northwest of the United States are sown at depth of 10 to 20 cm [45]. Deep seeding may also reduce the threat of damage by mice or other animals [6], and protect the seedlings from pre-emergent herbicides [31].

Auxins are a class of plant hormones that can modify plant cell walls, and are essential for coleoptile cell elongation and expansion [7]. In cereal coleoptiles, the most significant cell wall modifications induced by auxin are the decline of noncellulosic glucan content [16, 27, 41] and the degradation of 1,3:1,4-β-glucan [42], by activating exo- and endo-β-glucanases associated with cell walls [18, 19, 22]. This partial degradation resulted in the cell wall loosening, therefore increasing the cells' extensibility. Auxin also enhances the synthesis of H^+-ATPase level in the plasma membrane to increase the H^+ extrusion into the apoplast to adjust to an optimum pH for cell wall enlargement [12]. Some studies suggest that the potassium channel gene Zea mays K^+ channel 1 (ZMK1) is upregulated by auxin and essential for coleoptile elongation by maintaining K^+ accumulation and turgor [34]. Some other cell wall-bound proteins were also reported to be key regulators for cell extension, such as α-expansin in rice [17] and β-expansin in wheat coleoptiles [9]. Nucleoside diphosphate (NDP) kinase genes were also reported to be involved in coleoptile elongation [32]. Although auxin is known to activate a group of genes responsible for cell expansion, the exact mechanisms underlying coleoptile elongation remain unclear.

Barley varieties show significant differences for coleoptile length. Paynter and Clarke (2010) [33] determined the coleoptile length for a total of 44 barley cultivars with different breeding origins, early growth habits (erect or prostrate) and pedigrees. In this collection the coleoptile length ranged from 38.7 mm (cultivar Morrell from Western Australia (WA)) to 92.9 mm (cultivar Doolup from WA), with an average of 70.2 mm. They concluded that coleoptile length was not associated with breeding origin and early growth habit in their barley collection. Takeda and Takahashi (1999) [51] scored 5,082 barley and 1,214 wheat varieties and found significant differences in deep-seeding tolerance, which related to the coleoptile length, first internode length and the seed size. Subsequent studies conducted QTL analyses for coleoptile length using several barley doubled haploid (DH) mapping populations: Takahashi et al. (2001) [50] used two different DH populations (Harrington × TR306 and Steptoe × Morex) and identified QTLs for deep-seeding tolerance, coleoptile length and first internode length on the long arm of chromosome 5H, corresponding with QTLs for abscisic acid and gibberellic acid response. Takahashi et al. (2008) [49] used another Harrington × TR306 population and mapped QTLs for coleoptile elongation on chromosomes 1H, 2H, 4H, 5H, 6H and 7H. However, only 127 markers for the Harrington × TR306 population, and 223 markers for the Steptoe × Morex population were available. As a result, detected QTLs spanned 2 to 5 cM intervals across the seven barley chromosomes, which was too low-resolution to pinpoint candidate genes. To date, no specific candidate genes for barley coleoptile length have been reported.

In this study, we performed genome-wide association mapping with more than 30,000 genetic markers to map the marker-trait associations (MTAs) for coleoptile length. We used a worldwide collection of mainly domesticated barley cultivars, including a large proportion of barley cultivars grown in the driest regions in the world such as Australia. The aims of this study were (i) to investigate the phenotypic variation of coleoptile length in a diverse worldwide collection of barley genotypes, (ii) to determine genomic regions associated with coleoptile length via GWAS, and (iii) to identify and characterise the most likely candidate genes underlying the MTAs.

Results

Coleoptile length in barley accessions

The coleoptile length recorded in two independent experiments correlated with each other significantly (R^2= 0.87). Therefore, the average data of the two experiments was used to represent the varieties' coleoptile length. The investigated barleys had significant variation for the coleoptile length and its heritability was estimated at 0.54. The average length of the 328 barley accessions was at 5.40 cm, with the longest variety Russia24 at 7.51 cm and the shortest variety CDC Unity at 3.27 cm. The vast majority of all varieties in the barley panel (292 varieties, 88.8%) had coleoptile length ranging from 4 to 6.51 cm, only 27 varieties (8.2%) ranging from 6.50 to 7.51 cm and 10 varieties (3.0%) ranging from 3.27 to 3.99 cm (Figure 1 (A)). The barleys from different origins, in different row types and with different growth habits were compared (Figure 1 (B), (C) and (D)). The coleoptile length separated the barley origins into two subgroups: group 1 including Australia (mean at 5.60 cm), Africa (mean at 5.74 cm) and Asia (mean at 5.86 cm); group 2 including Europe (mean at 5.26 cm) and America (mean at 5.23 cm). The Australia originated barleys had longer coleoptile than European ($p=3.79\times10^{-4}$) and North and South American varieties ($p=1.26\times10^{-3}$), African longer than European ($p=2.51\times10^{-2}$) and North and South American ($p=4.26\times10^{-2}$), Asian longer than European ($p=1.93\times10^{-4}$) and North and South American ($p=8.87\times10^{-4}$) (Figure 1 (B)). No significant difference was found among the members in either group. Barleys with different row types (two-row or six-row) had a similar coleoptile length (the two-row and six-row means were 5.40 cm and 5.37 cm, respectively) and no significant difference were detected. Similarly, no significant difference was found between the barleys with different growth habits (the spring barleys and winter barleys means were both 5.38 cm). In conclusion, the coleoptile length correlated to the breeding origins but not to the row types and growth habits. The varieties with relatively long coleoptile (>6.50 cm) and their origins, row types, growth habits and coleoptile length are listed in Table S1.

Population structure and Linkage disequilibrium (LD) analysis

A Neighbor-joining (NJ) tree based on genetic distances for the barley population in this study (328 barley accessions) was constructed, incorporating their origins, row types and growth habits respectively (Figure 2 (A)). The barley collection in this study presented high genetic diversity and covered accessions from worldwide origins (35 countries), with different row types (two-row and six-row) and in different growth habits (winter, spring, and facultative) (Figure 2 (A) (B); Supplemental Figure S1). The result of PCA with separation based on row-type, growth habit, or geographic location are presented in Figure 2 (B).

The optimal K value (number of subpopulations) of the barley germplasm collection in this study was predicted using ADMIXTURE v.1.3.0 [1]. It showed that the optimal number of subpopulations was K=7 according to the Δ cross-validation error (Figure 2 (C) and Supplementary Figure S2). The population structures of 328 barley germplasm with K value from 2 to 12 were listed in Supplementary Figure S3. The pairwise LD decay (r2) analysis was performed on each chromosome and decreased with physical distance (Supplemental Figure S4).

Association analysis

The association study was performed using two subsets of markers: subset MAF01 (MAF>0.01) including 33,146 markers and subset MAF05 (MAF>0.05) including 23,193 markers. Significant MTAs (qFDR<0.05) were identified on all 7 chromosomes (Supplemental Figure S5 (A) and (B)). The quantile–quantile (QQ) plots indicated that the GLM model was suitable and efficient for this study (Supplemental Figure S5 (C) and (D)). All the significant MTAs identified using the two marker cut-offs were listed in Supplemental Table S2. Totally, there were 128 markers identified significantly associated with coleoptile length (qFDR<0.05) using both marker subsets, representing 53 genic loci (loci within genes) and 54 intergenic loci (loci between genes) (Supplemental Table S2), explaining 4.07–6.81% phenotypic variation (r^2). Apart from the common markers identified by both MAF01 and MAF05, MAF01 identified 49 additional markers, representing extra 19 genic loci and 9 intergenic loci (Supplemental Table S2), explaining 3.37-5.54 % phenotypic variation (r^2). Apart from the common markers identified by MAF01 and MAF05, MAF05 identified 21 additional markers, representing extra 7 genic loci and 9 intergenic loci (Supplementary Table S2), explaining 3.49-4.07 % phenotypic variation (r^2).

The highly significant associated loci (qFDR<0.01, $-\log_{10}(q)>2.0$) identified by either MAF01 or MAF05 are listed in Table 1. There were 5 genic loci identified by MAF01 and MAF05 in common, including HORVU1Hr1G073010 (unknown function), HORVU1Hr1G076430 (*FT*), HORVU1Hr1G077230 (*CSLC6*), HORVU5Hr1G058300 (*TPPB*) and HORVU5Hr1G007340 (*LRR-RLK*). One genic locus within gene HORVU6Hr1G019700 (*SPL3*) was only detected in MAF01. Two genic loci harbouring genes HORVU6Hr1G022770 (*VIN3*), and HORVU6Hr1G022500 (*BTBD*) were only detected in MAF05. The top significantly associated markers (with the lowest qFDR) for each candidate gene are summarized for their effect on the coleoptile length (Figure 3). For the *SPL3* gene, the marker C6H53910826 showed 1.42 cm longer coleoptile ($p=2.85\times10^{-6}$) when the alternative allele T is present, explaining 5.54% phenotypic variation. For the unknown function gene HORVU1Hr1G073010, the marker D1H500582726 showed 0.24 cm shorter coleoptile ($p=1.52\times10^{-2}$) when the alternative allele T is present, explaining 6.52% phenotypic variation. For the *FT* gene, the marker C1H514098702 showed 0.34 cm longer coleoptile ($p=4.81\times10^{-4}$) when the alternative allele C is present, explaining 6.68% phenotypic variation. For the *CSLC6* gene, the marker D1H516785422 showed 0.33 cm shorter coleoptile ($p=1.96\times10^{-3}$) when the alternative allele G is present, explaining 6.28% phenotypic variation. For the *TPPB* gene, the marker D5H456061421 showed 0.99 cm longer coleoptile ($p=4.74\times10^{-9}$) when the alternative allele C is present, explaining 6.45% phenotypic variation. For the *LRR-RLK* gene, the marker D5H014097066 showed 0.49 cm shorter coleoptile ($p=1.05\times10^{-3}$) when the alternative allele G is present, explaining 6.10% phenotypic variation. For the *VIN3* gene, the marker C6H72969182 showed 0.51 cm shorter coleoptile ($p=1.71\times10^{-9}$) when the alternative allele A is present, explaining 5.91% phenotypic variation. For the *BTBD* gene, the marker D6H071745828 showed 0.52 cm longer coleoptile ($p=1.73\times10^{-9}$) when the alternative allele G is present, explaining 5.87% phenotypic variation. In conclusion, all the associated markers within genes had strong effects on the coleoptile length. There were 12 intergenic loci identified by MAF01, MAF05 or both.

Possible genes responsible for coleoptile length were searched around the loci and listed in Supplementary Table S3.

Major candidate gene for coleoptile length

The markers within gene *SPL3* showed association with coleoptile length for both MAF01 and MAF05. In MAF01, there were 5 markers on *SPL3* identified to be significant (0.01<qFDR<0.05) and 11 markers on *SPL3* showed highest significance of all associations (qFDR=5.4×10^{-3}). In MAF05 there were 2 markers on *SPL3* identified to be significant (0.01<qFDR<0.05) (Supplementary Table S2). Although the significant markers within eight candidate genes all had strong effects on the coleoptile length, the marker C6H53910826 (and other 10 markers with qFDR=5.4×10^{-3} in MAF01) on *SPL3* showed the most significant effect: the mean length was at 5.37 cm when the allele C presented and 6.79 cm when the allele T presented (p=2.85×10^{-6}) (Figure 3 (A)). In conclusion, multiple markers on *SPL3* had been identified to be significantly associated by different methods and some of the markers had the highest association index (qFDR=5.4×10^{-3}) and had the strongest effect on the phenotype. Therefore, *SPL3* was considered the major candidate gene associated to the coleoptile length in this study.

SPL3 is a transcription factor gene located on chromosome 6H:53,909,817 to 53,916,886 (7,070 bp), consisting of 5'UTR, 3'UTR, four exons and three introns (Figure 4 (A)). The gene encodes a protein with 474 amino acids (aas). The conserved SBP domain is central functional region of this transcription factor, and contains a plant-specific DNA-binding domain. All the variants and amino acid substitutions are summarized in Figure 4 (B). There were five variants on exons, including four missense variants and one synonymous variant. The missense variant at position 53,913,050 replaced serine with alanine in the SBP domain, likely impacting its DNA-binding activity. Furthermore, this marker C6H53913050 was one of the markers showing highest significance in this study (qFDR=5.4×10^{-3}). Other three missense variants included glutamic acid replaced with lysine at position 53,913,549, alanine replaced with valine at position 53,913,335 and 53,910,588. Five variants were found in 5' or 3' UTR and eight variants were in the introns (Figure 4 (B)). Figure 4 (C) showed the position of all detected variants, including significant association and non-significant association with coleoptile length, and the LD plot surrounding these markers.

To further understand the role of *SPL3* gene in barley coleoptile growth, we measured its expression in coleoptile tissue of two varieties (CDC Unity and CI5791), representing two major haplotypes of *SPL3* (C/T at marker C6H53910826) in the population. The comparison of the *SPL3* expression between CDC Unity and CI5791, and in dark and under natural daylight is presented in Supplementary Figure S6. The data represented the actin normalized target gene expression relative to control (the *SPL3* gene expression in coleoptile of CDC Unity in dark) (considered as 1). There was no statistically significant difference between CDC Unity and CI5791, either in dark or under natural daylight. However, in CDC Unity, the expression under light decreased by 43% compared with dark conditions. Similarly, in CI5791, the expression under light decreased by 39% compared with dark conditions. In conclusion, the coleoptile

length variation between two *SPL3* alleles was not due to the gene expression, but for the both alleles there was a steady decline in gene expression when the coleoptile was exposed under daylight.

Discussion

More frequently occurring drought conditions as result of changing climate had become a major challenge in many growing regions of the world. Long coleoptile barleys have the potential to help mitigate limitations on emergence and improve the field establishment of barleys in water limited areas. In this study we performed GWAS on coleoptile length using 328 barley accessions with more than 30,000 molecular markers. As a result, the coleoptile length showed significant variation in this barley population and the GWAS identified 8 genes and 12 intergenic loci with high significance, providing a valuable source for selecting long coleoptile barleys in the future.

In this study, the seeds were sown in washed river sand, a commonly used method for coleoptile measurement [37, 49], and more closely related to the seeds' natural germination conditions compared with the wet filter paper method [4]. Previous studies indicated that, apart from the genetic factors, both the seed source and the germination methodology may affect the coleoptile length [35, 49], which explained the variations of our measurement compared to earlier studies [4, 33].

SPL3 is the major candidate gene for coleoptile length in this study. The SPL3 protein belongs to SBP family and contains the conserved SBP domain. It was an important plant-specific transcription factor regulating plant growth and development, and the two zinc finger structures are located in the core region of the SBP domain [24]. According to the gene variants, there are four amino acid substitutions, one of which is a serine to alanine change located in the second zinc-finger structure of SBP domain [8]. This amino acid change was likely to change the SPL3 DNA binding affinity. We showed that SPL3 is expressed in the coleoptile tissue and the expression increased in dark conditions, in accordance with the light sensitivity of coleoptile growth. Although the expression in dark only increased by 60–70% compared with daylight conditions, the effect can be significantly amplified since SPL3 is a transcription factor potentially activating other genes. However, the gene expression between two haplotypes in the population was identical, so the coleoptile length variation of the two genotypes was probably due to the protein difference rather than the expression. Further studies are needed to determine the function and the downstream target of this genes, as well as how it affects coleoptile growth in barley.

A number of candidate genes identified in this study were shown to regulate cell growth in other plants. CSLC6 on 1H, encoding an enzyme involved in cell wall biosynthesis, was determined a candidate gene due to highly significant associations with coleoptile length. The gene, a member of the cellulose-synthase-like family, was proposed to encode the catalytic subunit of enzymes that synthesize hemicellulose backbones [40]. A previous study showed that hemicelluloses were involved in regulation of wall modification, elongation and growth [53]. Therefore, this gene may influence the coleoptile length by regulating cell wall synthesis. GA2ox (HORVU1Hr1G076730) on 1H was identified as another candidate gene, and it is also a known semi-dwarf gene [26]. This gene is essential in the Gibberellin (GA)

catabolic pathway and regulates plant growth by inactivating endogenous bioactive GAs. Previous studies indicated that many semi-dwarf genes had a negative effect on coleoptile length [3, 10]. This gene may regulate the plant growth by GA pathway in multiple aspects, including the coleoptile elongation. EXPB2 (HORVU1Hr1G054230) on 1H was only identified by MAF01 with significant association. Members in this gene family had been showed to regulate coleoptile growth in other cereal crops [9, 17]. Previous studies showed that the wheat coleoptile length was controlled by multiple genes on different chromosomes, and the combination of QTLs with minor effect could improve coleoptile length obviously [29, 39]. Further work needs to be done to investigate how the genes and loci identified in this study had an impact on barley coleoptile and it would provide an opportunity to breed barley varieties with preferable coleoptile length by manipulating multiple genes/loci.

For cereal crops grown in drought prone regions, cultivars with longer coleoptile can emerge quicker and have better early vigour. Longer coleoptiles can not only protect the crops from early drought but also improve water use efficiency, and are more sustainable in low water supply areas. In the past decades, coleoptile elongation has been studied mainly in wheat, with a main focus on dwarf or semi-dwarf genes that had only little effect on the coleoptile length [10]. Only a few other studies announced coleoptile length-related loci other than dwarf genes [3]. In this study, by genome wide association analysis, 20 loci were identified at high significance and more than 100 loci were detected at lower significance, providing a wide range of genetic factors interacting with coleoptile elongation. This study provided a comprehensive overview of the genetic pool regulating coleoptile length. The genes detected in this study need further investigation regarding to their variation in the barley accessions and their mechanism regulating coleoptile cell growth. It opens the opportunity to understand the genetic network of crop early vigour and field establishment.

Conclusions

In this study, a worldwide collection of barleys (328 accessions) showed a large phenotypic variation for coleoptile length. The genome wide association analysis using more than 30,000 markers identified 8 genes and 12 intergenic loci with high significance (qFDR < 0.01), as well as 71 genes and 60 intergenic loci with lower significance (0.01 < qFDR < 0.05). SPL3 was determined as the novel major candidate gene for coleoptile length in this study. It had the strongest effect on the coleoptile length among the candidate genes and was identified by multiple markers and different methods. SPL3 was expressed in the coleoptile but no significant expression difference was found between haplotypes. The substitution of serine for alanine in the second zinc-finger structure of SBP domain in SPL3 likely impacted its DNA-binding activity. This work provides a valuable overview of genetic factors responsible for coleoptile length in barley, detecting genome-wide loci with improved resolution, making it easier to pinpoint the responsible genes in the future research. It provides opportunity to improve seedling early vigour and stand establishment for barley and other crops by better understanding the mechanism regulating coleoptile growth.

Methods

Plant material

A barley diversity panel was assembled comprising a large collection of over 4,000 landrace, cultivated, and research barley varieties of diverse origin preserved at the Western Barley Genetics Alliance at Murdoch University (Perth, Australia). For this study, 328 barley accessions were selected representing domesticated, landrace, and breeding accessions from 35 countries and were selected based on diversity in coleoptile length, and geographic origins to represent European, Asian, North and South American, African and Australian breeding programs. The selection of domesticated barley originated from various breeding programs and represented the whole variety of cultivated barley lifeforms, including two- and six-row genotypes with winter, spring, and facultative growth habits. A list of all plant materials is provided in Supplemental Figure S1.

Phenotyping

Seeds with normal size and plumpness were selected from each accession of 328 varieties. The seeds were sown at 2 cm depth in washed river sand (12% v/w moisture) and stored in the dark at 20 °C. After 14 days the etiolated seedlings were pulled out from the river sand and the coleoptile length from the embryo end of grain to the coleoptile tip was measured manually using a ruler. Eight biological replicates were used for each accession for coleoptile length measurement, and it was repeated two times. Absolute measurements that deviated from mean values by more than 2 standard deviation units were considered as outliers and set as missing values. The average values of repeats were used to represent the varieties' coleoptile lengths.

DNA extraction and high-throughput genotyping

The collection of 328 barley varieties were grown in glasshouse and harvested at the three-leaf stage. Genomic DNA extraction was performed using a rapid cetyl-trimethyl- ammonium bromide (CTAB) method [46].

The SNP and InDel markers were obtained from targeted re-sequencing of phenology genes, low coverage sequencing (1x) of DNA libraries and a whole genome profiling for DArTseq markers as described before [14, 15, 30]. Filtering of the marker dataset using MAF01 (MAF>0.01) and MAF05 (MAF>0.05) retained 33,146 and 23,193 genetic markers respectively.

Population structure and genotypic data analysis

Genetic distance based on IBS (identity-by-state similarity) were calculated and a Neighbor-Joining (NJ) dendrogram was computed, all using TASSEL v.5.2.39 software [5].

The model-based clustering algorithm of ADMIXTURE v.1.3.0 was used to investigate subpopulation structure of the barley diversity panel. Prior to population structure analysis in ADMIXTURE, the genotype

dataset was LD pruned using Plink v1.939 with the window size set to 50 kb, step size to 5, and the pairwise r^2 threshold set to 0.5, yielding 19,014 genetic variants. As described previously [14], a preliminary analysis was performed in 100 replicate runs by inputting successive values of K from 1 to 20. A 10-fold cross validation (CV) procedure was performed with 100 different fixed initial seeds for each K-value. The most likely K-value was determined using ADMIXTURE's CV error values. The software CLUMPP [20] v.1.1.2 was used to obtain the optimal alignments of 100 replicates for each K-value. Individual genotype membership proportions were averaged across runs according to the permutation with the greatest symmetric similarity coefficient. The output from CLUMPP for the optimal K was used to make plots using the cluster visualization program Pophelper v.2.2.3 (http://royfrancis.github.io/pophelper/) implemented in R software (http://www.R-project.org/).

Principal component analysis (PCA) was also conducted based on all markers data using TASSEL v.5.2.39 to summarize the genetic structure and variation present in the barley germplasm. The first two principal components were plotted against each other using 'scatter plot' function in Microsoft Excel 2016. NJ trees were constructed using the Java application Archaeopteryx v.0.9909 [13] based on genetic distances calculated in TASSEL v.5.2.39.

LD analysis

Genome-wide LD analysis was performed by pair wise comparisons among the intra-chromosomal genetic markers using Plink v.1.93 [36]. LD was estimated by using squared allele frequency correlations (r^2) between the intra-chromosomal pairs of loci. To investigate the extent of LD decay, intra-chromosomal r^2 values were plotted against the physical distance (kb) between markers. Curves were fitted by second-degree LOESS using R software v.3.5.1 (http://www.R-project.org/).

Association analysis

Genome wide association studies were performed using 33,146 (MAF>0.01) and 23,193 (MAF>0.05) genetic markers, respectively in TASSEL [5]. Different statistical models were used to calculate P-values for putative MTAs which included population structure (Q) and the kinship matrix (K) to account for population structure to avoid spurious associations.

We used the General Linear Model (GLM) with population structure (Q) matrix (the first three PCs) and kinship (K) matrix (matrix of genetic similarity based on simple matching coefficients, which was also used for constructing the neighbour-joining tree) as a correction for population structure. Results from different Q and K matrices were compared, and according to the quantile-quantile (Q-Q) plot, the GLM model incorporating Q was suitable for this study.

For the GLM model and analyses, multiple testing was performed to assess the significance of marker trait associations using the qvalue [47] v.2.8.0 R package (R 3.4.2) employing the smoother method [48], an extension of the false discovery rate (FDR) method. Lambda was selected as 0 which estimates $\pi i(0)=1$, which produces a list of significant tests equivalent to the [2] procedure and is considered a

conservative case of the qvalue methodology. Only markers with qFDR < 0.05 were considered to be significant. The Manhattan plot was drawn with qqmanv.0.1.4 [52].

Broad-sense heritability (H^2) was calculated using the following equation by treating genotype and environment as random effects according to Equation 1:

$$H^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (1)$$

where σ_a^2 and σ_e^2 represent the variance derived from genotypic and environmental effects, respectively.

Candidate gene expression

Seeds with normal size and plumpness were selected from barley varieties CDC Unity and CI5791, representing the two major haplotypes of the primary candidate gene *Squamosa promoter-binding-like protein 3* (*SPL3*) in the population. Eight seeds per variety were sown at 2 cm depth in washed river sand (12% v/w moisture) and stored at 20 °C in the dark and under normal daylight, respectively, with three technical repeats. The coleoptile was collected at day 7 and ground to a fine powder with liquid nitrogen. The total RNA was extracted using TRIsure (Bioline, #38032) following the manufacturer instruction. The RNA extract was quantified using NanoDrop™ One Microvolume UV-Vis spectrophotometer (Thermo Scientific) and approximately equal amount of RNA (~1 µg) was used for reverse transcription. The reverse transcription was done using SensiFAST™ cDNA Synthesis Kit (Bioline, # 65053) as per manufacturer instruction. The real-time PCR was done using SensiFAST™ SYBR® Lo-ROX Kit (Bioline, #94005) and ViiA™ 7 Real-Time PCR System (Applied Biosystems), performed in triplicate for each cDNA sample. A pair of primers were designed to flank and straddle the intron to avoid the DNA contamination: SPL3_exp_F (ACAGTGCAGCCGGTTTCATG) and SPL3_exp_R (GAACATATGGAGCCTGACCGA). The target gene expression was normalized using Actin as a reference [21]. The thermocycling was set as follows: 96 °C for 2 min, 40 cycles at 96 °C for 15 s, 62 °C for 15 s and 72 °C for 15 s, and a final extension at 72 °C for 5 min. The relative expression levels of target genes were determined as $2^{-\triangle Ct}$. The size of the PCR product was estimated by 2% agarose gel.

Abbreviations

GWAS
Genome-wide association studies
SPL3
Squamosa promoter-binding-like protein 3
ZMK1
Zea mays K^+ channel 1
NDP

Nucleoside diphosphate
DH
Doubled haploid
QTL
Quantitative trait locus
MTA
Marker-trait association
CTAB
Cetyl-trimethyl- ammonium bromide
MAF
Minor allele frequency
IBS
Identity-by-state similarity
NJ
Neighbor-Joining
LD
Linkage disequilibrium
PCA
Principal component analysis
CV
Cross validation
GLM
General Linear Model
Q-Q
Quantile-quantile
FDR
False discovery rate
aa
Amino acid
CSLC6
Cellulose-synthase-like C6
GA2ox
Gibberellin 2-oxidase
GA
Gibberellic acid
EXPB2
Expansin-B2
FT
Flowering Locus T
TPPB

Trehalose-6-phosphate phosphatase
LRR-RLK
Leucine-rich repeat receptor-like protein kinase family
VIN3
Vernalization insensitive 3
BTBD
BTB/POZ domain-containing protein
WA
Western Australia

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The dataset supporting the conclusions of this article is included within the Additional file 10.

Funding

This work was supported by funding from Department of Primary Industries and Regional Development Western Australia (DPIRD); and Western Australian State Agricultural Biotechnology Centre (SABC).

Acknowledgments

Not applicable.

Authors' contributions

HL designed the experiment, phenotyped coleoptile length of barley collection, analysed the data, drafted and revised the manuscript. CBH analysed the data, performed the GWAS and population structure analyses, provided the marker data, and revised the manuscript. GZ analysed the data. XQZ provided the marker data. CL supervised the experiment and provided critical feedback. All authors discussed the results and contributed to the final manuscript.

References

1. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. Bmc Bioinformatics. 2011;12.
2. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological. 1995;57(1):289-300.
3. Bovill WD, Hyles J, Zwart AB, Ford BA, Perera G, Phongkham T, et al. Increase in coleoptile length and establishment by Lcol-A1, a genetic locus with major effect in wheat. Bmc Plant Biol. 2019;19.
4. Box AJ, Jefferies SP, Barr AR, editors. Emergence and establishment problems of hulless barley – a possible solution. 9th Australian Barley Technical Symposium; 1999.
5. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633-5.
6. Brown PR, Singleton GR, Tann CR, Mock I. Increasing sowing depth to reduce mouse damage to winter crops. Crop Protection. 2003;22(4):653-60.
7. Catala C, Rose JKC, Bennett AB. Auxin-regulated genes encoding cell wall-modifying proteins are expressed during early tomato fruit growth. Plant Physiol. 2000;122(2):527-34.
8. Chen XB, Zhang ZL, Liu DM, Zhang K, Li AL, Mao L. SQUAMOSA Promoter-Binding Protein-Like Transcription Factors: Star Players for Plant Growth and Development. J Integr Plant Biol. 2010;52(11):946-51.
9. Gao Q, Zhao MR, Li F, Guo QF, Xing SC, Wang W. Expansins and coleoptile elongation in wheat. Protoplasma. 2008;233(1-2):73-81.
10. Grover G, Sharma A, Gill HS, Srivastava P, Bains NS. Rht8 gene as an alternate dwarfing gene in elite Indian spring wheat cultivars. Plos One. 2018;13(6).
11. Hadjichristodoulou A, Della A, Photiades J. Effect of Sowing Depth on Plant Establishment, Tillering Capacity and Other Agronomic Characters of Cereals. Journal of Agricultural Science. 1977;89(Aug):161-7.
12. Hager A, Debus G, Edel HG, Stransky H, Serrano R. Auxin-Induced Exocytosis and the Rapid Synthesis of a High-Turnover Pool of Plasma-Membrane H+-Atpase. Planta. 1991;185(4):527-37.
13. Han MV, Zmasek CM. phyloXML: XML for evolutionary biology and comparative genomics. Bmc Bioinformatics. 2009;10.
14. Hill CB, Angessa TT, McFawn LA, Wong D, Tibbits J, Zhang XQ, et al. Hybridisation-based target enrichment of phenology genes to dissect the genetic basis of yield and adaptation in barley. Plant Biotechnol J. 2019;17(5):932-44.
15. Hill CB, Wong D, Tibbits J, Forrest K, Hayden M, Zhang XQ, et al. Targeted enrichment by solution-based hybrid capture to identify genetic sequence variants in barley. Sci Data. 2019;6.
16. Hoson T, Nevins DJ. Beta-D-Glucan Antibodies Inhibit Auxin-Induced Cell Elongation and Changes in the Cell-Wall of Zea Coleoptile Segments. Plant Physiol. 1989;90(4):1353-8.

17. Huang J, Takano T, Akita S. Expression of alpha-expansin genes in young seedlings of rice (Oryza sativa L.). Planta. 2000;211(4):467-73.
18. Huber DJ, Nevins DJ. Beta-D-Glucan Hydrolase Activity in Zea Coleoptile Cell-Walls. Plant Physiol. 1980;65(5):768-73.
19. Inouhe M, Nevins DJ. Auxin-Enhanced Glucan Autohydrolysis in Maize Coleoptile Cell-Walls. Plant Physiol. 1991;96(1):285-90.
20. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23(14):1801-6.
21. Jia QJ, Tan C, Wang JM, Zhang XQ, Zhu JH, Luo H, et al. Marker development using SLAF-seq and whole-genome shotgun strategy to fine-map the semi-dwarf gene ari-e in barley. Bmc Genomics. 2016;17.
22. Kotake T, Nakagawa N, Takeda K, Sakurai N. Auxin-induced elongation growth and expressions of cell wall-bound exo- and endo-beta-glucanases in barley coleoptiles. Plant and Cell Physiology. 2000;41(11):1272-8.
23. Kuhling I, Redozubov D, Broll G, Trautz D. Impact of tillage, seeding rate and seeding depth on soil moisture and dryland spring wheat yield in Western Siberia. Soil & Tillage Research. 2017;170:43-52.
24. Li XY, Lin EP, Huang HH, Niu MY, Tong ZK, Zhang JH. Molecular Characterization of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family in Betula luminifera. Front Plant Sci. 2018;9.
25. Lindstrom MJ, Papendick RI, Koehler FE. Model to Predict Winter-Wheat Emergence as Affected by Soil Temperature, Water Potential, and Depth of Planting. Agronomy Journal. 1976;68(1):137-41.
26. Lo SF, Yang SY, Chen KT, Hsing YL, Zeevaart JAD, Chen LJ, et al. A Novel Class of Gibberellin 2-Oxidases Control Semidwarfism, Tillering, and Root Development in Rice. Plant Cell. 2008;20(10):2603-18.
27. Loescher W, Nevins DJ. Auxin-Induced Changes in Avena-Coleoptile Cell-Wall Composition. Plant Physiol. 1972;50(5):556-&.
28. Mohan A, Schillinger WF, Gill KS. Wheat Seedling Emergence from Deep Planting Depths and Its Relationship with Coleoptile Length. Plos One. 2013;8(9).
29. Murphy K, Balow K, Lyon SR, Jones SS. Response to selection, combining ability and heritability of coleoptile length in winter wheat. Euphytica. 2008;164(3):709-18.
30. Mwando E, Han Y, Angessa TT, Zhou G, Hill CB, Zhang XQ, et al. Genome-Wide Association Study of Salinity Tolerance During Germination in Barley (Hordeum vulgare L.). Front Plant Sci. 2020;11:118.
31. Osullivan PA, Weiss GM, Friesen D. Tolerance of Spring Wheat (Triticum-Aestivum L) to Trifluralin Deep-Incorporated in the Autumn or Spring. Weed Research. 1985;25(4):275-80.
32. Pan L, Kawai M, Yano A, Uchimiya H. Nucleoside diphosphate kinase required for coleoptile elongation in rice. Plant Physiol. 2000;122(2):447-52.

33. Paynter BH, Clarke GPY. Coleoptile length of barley (Hordeum vulgare L.) cultivars. Genetic Resources and Crop Evolution. 2010;57(3):395-403.
34. Philippar K, Fuchs I, Luthen H, Hoth S, Bauer CS, Haga K, et al. Auxin-induced K+ channel expression represents an essential step in coleoptile growth and gravitropism. P Natl Acad Sci USA. 1999;96(21):12186-91.
35. Pumpa J, Martin P, McRae F, Coombes N. Coleoptile length of wheat varieties. 2013.
36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
37. Radford BJ. Effect of Constant and Fluctuating Temperature Regimes and Seed Source on the Coleoptile Length of Tall and Semidwarf Wheats. Australian Journal of Experimental Agriculture. 1987;27(1):113-7.
38. Rebetzke GJ, Richards RA, Fettell NA, Long M, Condon AG, Forrester RI, et al. Genotypic increases in coleoptile length improves stand establishment, vigour and grain yield of deep-sown wheat. Field Crops Research. 2007;100(1):10-23.
39. Rebetzke GJ, Verbyla AP, Verbyla KL, Morell MK, Cavanagh CR. Use of a large multiparent wheat mapping population in genomic dissection of coleoptile and seedling growth. Plant Biotechnol J. 2014;12(2):219-30.
40. Richmond TA, Somerville CR. Integrative approaches to determining Csl function. Plant Mol Biol. 2001;47(1-2):131-43.
41. Sakurai N, Masuda Y. Auxin-Induced Changes in Barley Coleoptile Cell-Wall Composition. Plant and Cell Physiology. 1978;19(7):1217-23.
42. Sakurai N, Nevins DJ, Masuda Y. Auxin-Induced and Hydrogen Ion-Induced Cell-Wall Loosening and Cell Extension in Avena-Coleoptile Segments. Plant and Cell Physiology. 1977;18(2):371-80.
43. Schillinger WF, Donaldson E, Allan RE, Jones SS. Winter wheat seedling emergence from deep sowing depths. Agronomy Journal. 1998;90(5):582-6.
44. Schillinger WF, Papendick RI. Tillage mulch depth effects during fallow on wheat production and wind erosion control factors. Soil Science Society of America Journal. 1997;61(3):871-6.
45. Schillinger WF, Papendick RI. Then and now: 125 years of dryland wheat farming in the Inland Pacific Northwest. Agronomy Journal. 2008;100(3):S166-S82.
46. Stewart CN, Via LE. A Rapid Ctab DNA Isolation Technique Useful for Rapid Fingerprinting and Other Pcr Applications. Biotechniques. 1993;14(5):748-&.
47. Storey JD. A direct approach to false discovery rates. J Roy Stat Soc B. 2002;64:479-98.
48. Storey JD, Tibshirani R. Statistical significance for genomewide studies. P Natl Acad Sci USA. 2003;100(16):9440-5.
49. Takahashi H, Noda M, Sakurai K, Watanabe A, Akagi H, Sato K, et al. QTLs in barley controlling seedling elongation of deep-sown seeds. Euphytica. 2008;164(3):761-8.

50. Takahashi H, Sato K, Takeda K. Mapping genes for deep-seeding tolerance in barley. Euphytica. 2001;122(1):37-43.
51. Takeda K, Takahashi H. Varietal variation for the deep seeding tolerance in barley and wheat. Breeding Research. 1999;1:1-8.
52. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014:005165.
53. Yang JL, Zhu XF, Peng YX, Zheng C, Li GX, Liu Y, et al. Cell Wall Hemicellulose Contributes Significantly to Aluminum Adsorption and Root Growth in Arabidopsis. Plant Physiol. 2011;155(4):1885-92.

Supplementary Files Legend

Additional files

Additional file 1: Figure S1. Diversity of barley varieties. a) Geographic origins of barley varieties. Numbers of countries in each geographical region are indicated on top of the bars. b) Row type and growth habit of barley varieties. Numbers of barley varieties are indicated on top of the bars.

Additional file 2: Figure S2. Exploration of the optimal number of genetic subpopulations (K) using Δ cross-validation error and standard error values in the barley germplasm collection. A solid red line denotes the choice of K=7 which represents the most likely number of subpopulations within the barley germplasm collection.

Additional file 3: Figure S3. Plot of ancestry estimates inferred by ADMIXTURE for 329 worldwide barley accessions for 19,014 SNPs. Each colour represents a population, and the colour of individual haplotypes represents their proportional membership in the different populations. Membership coefficients for each population were merged across 100 replicate runs using the CLUMPP programme. The number of clusters (K) present in the entire population of 329 accessions was judged to be K = 7 based on the CV error. Shown are clusters 2 to 12.

Additional file 4: Figure S4: The extent of LD in the worldwide collection of domesticated barley varieties. Values are mean intra-chromosomal LD r^2 values for all intra-chromosomal pairs of SNPs binned by distance. Curves were fit by second-degree loess.

Additional file 5: Table S1 Barley accessions with long coleoptile.

Additional file 6: Table S2 Loci significantly associated with coleoptile length using two MAF ($q<0.05$)

Additional file 7: Table S3 Intergenic Loci with highly significant association ($qFDR<0.01$) and candidate genes

Additional file 8: Figure S5. Manhattan plots and quantile–quantile (QQ) plot of coleoptile length. (A) The Manhattan plot of coleoptile length by GLM model using minor allele frequency (MAF)<0.05. (B) The Manhattan plot of coleoptile length by GLM model using minor allele frequency (MAF)<0.01. The results were presented using -log10 of FDR adjusted p values (q value) against position on each of the seven chromosomes. Horizontal dashed lines showed the genome wide significant threshold set at qFDR<0.05 (blue) and qFDR<0.05 (red). (C) QQ plot for coleoptile length based on -log10 of FDR adjusted p-values (q-value) by GLM model using MAF<0.05. (D) QQ plot for coleoptile length based on -log10 of FDR adjusted p-values (q-value) by GLM model using MAF<0.01.

Additional file 9: Figure S6. The relative expression (fold change) of SPL3 gene in coleoptile of CDC Unity and CI5791. The 'dark' indicated the gene expression in etiolated coleoptile tissue and the 'daylight' indicated the gene expression in normal coleoptile tissue under natural light. The expression was normalized using reference gene (actin) and relative to the etiolated coleoptile tissue of CDC Unity (converted to 1).

Additional file 10: The dataset supporting the conclusions of this article.

Table

Due to technical limitations, Table 1 is only available for download from the Supplementary Files section.

Figures

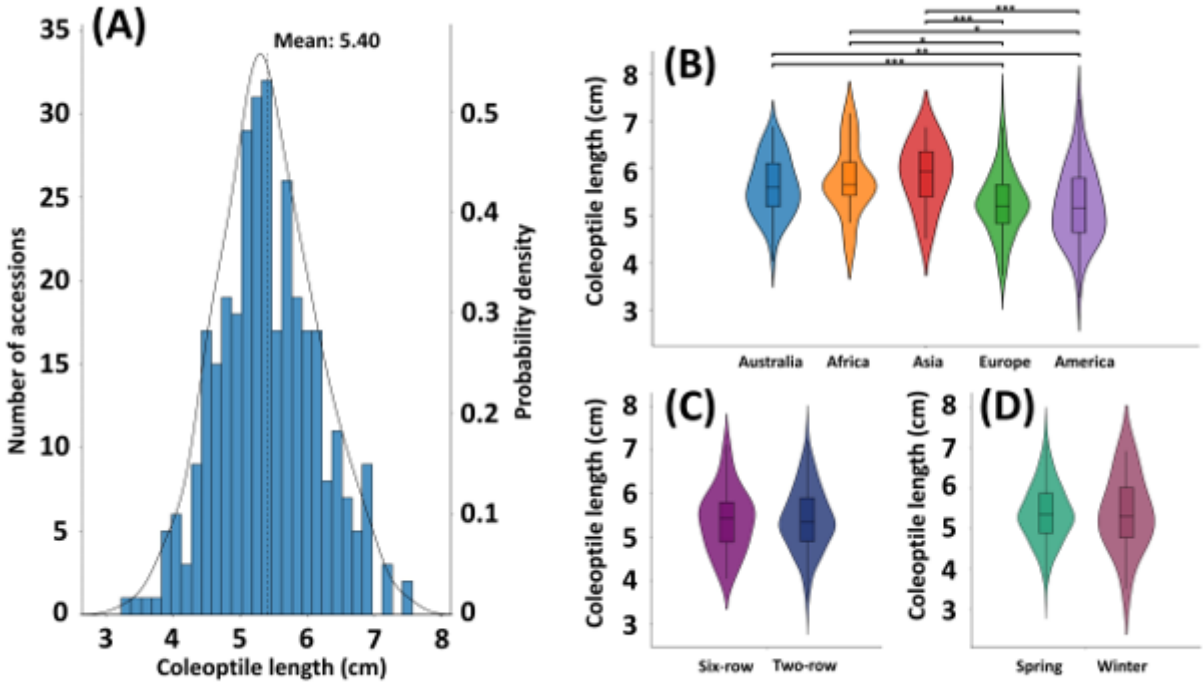

Figure 1

The coleoptile length in 328 barley accessions. (A)The distribution frequency of coleoptile length in 328 barley accessions. Comparison of coleoptile length from different origins (B), in different row types (C) and with different growth habits (D). The coleoptile length was averaged for two years experiments. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

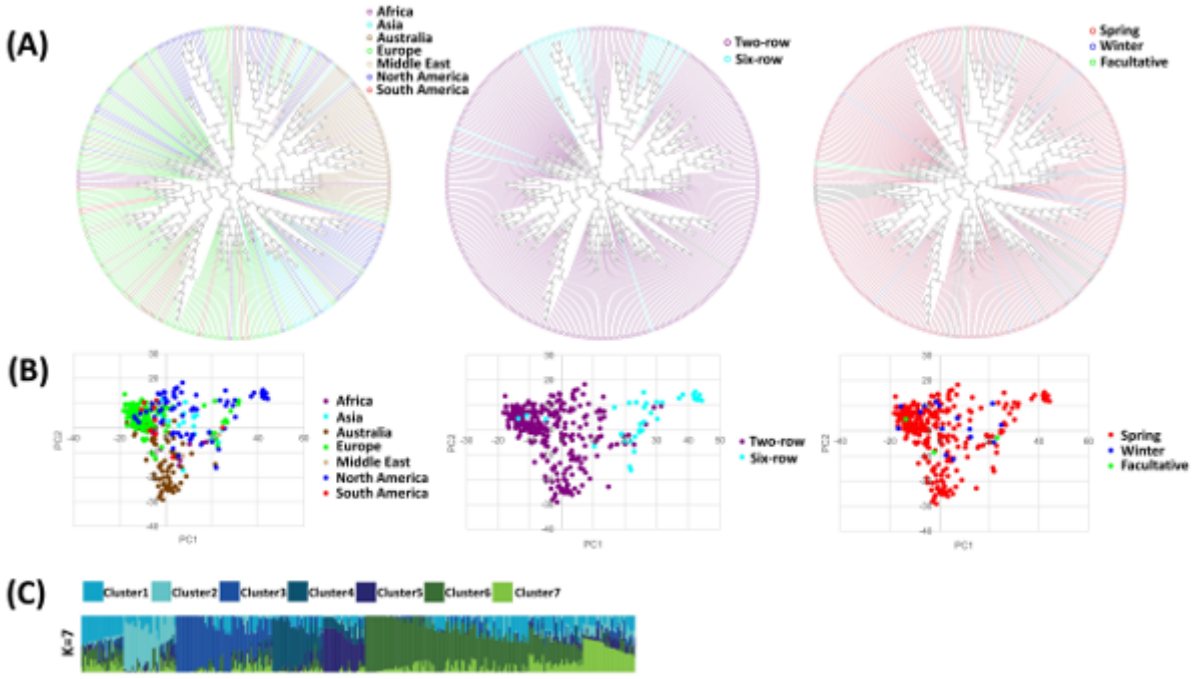

Figure 2

Population structure analysis of 328 barley accessions. (A) Phylogenetic neighbour-joining (NJ) tree of 328 barley accessions constructed based on genetic distances, their origins, row types and growth habits. (B) Principal component analysis (PCA) of the first two components of 328 barley accessions based on origins, row types and growth habits. (C) Plot of ancestry estimates inferred by ADMIXTURE for 328 worldwide barley accessions for 19,014 SNPs. Each colour represents a population, and the colour of individual haplotypes represents their proportional membership in the different populations. Membership coefficients for each population were merged across 100 replicate runs using the CLUMPP programme. The number of clusters (K) present in the entire population of 328 accessions was judged to be K = 7 based on the CV error.

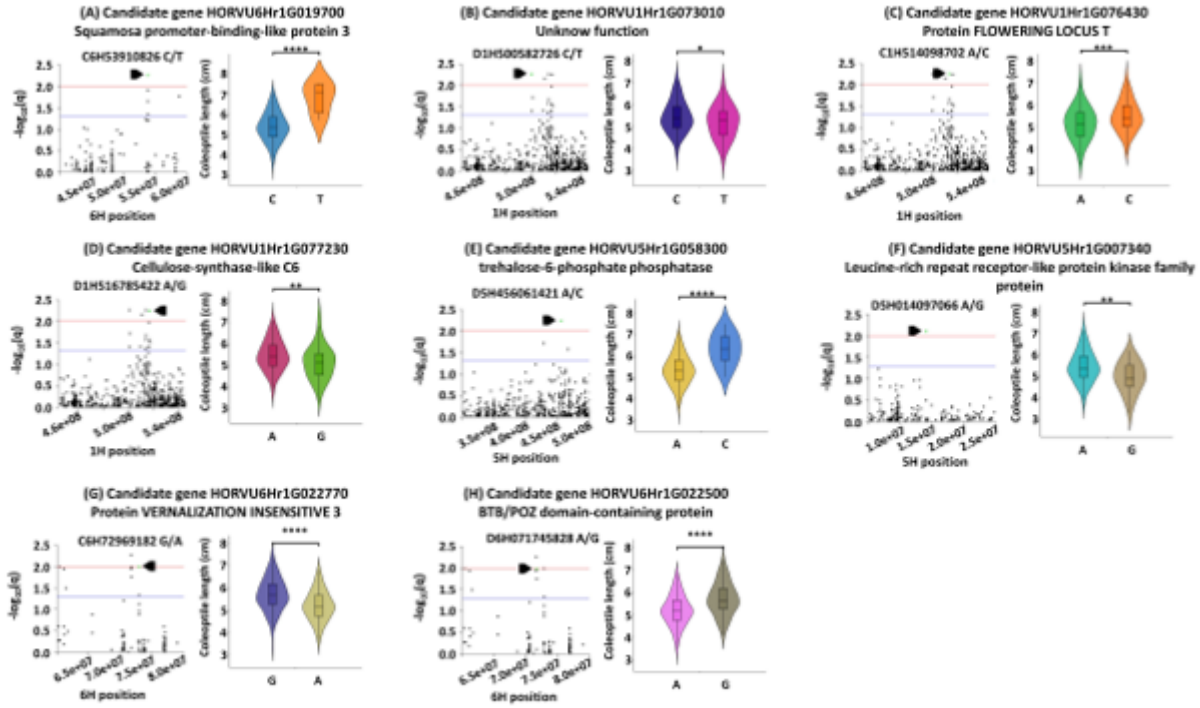

Figure 3

The genomic regions of eight candidate genes (A to H) had signals strongly associated with coleoptile length. Each included the Manhattan plot showing the physical position on the chromosome (left) and the coleoptile length variation between different alleles (right). For the Manhattan plot, the association significance was presented by -log10 of FDR adjusted P-values (q-values), and the signal with highest association (-log10(q)>2) for each candidate gene was highlighted in green. Horizontal lines indicated the genome-wide significant threshold selected by q-value cut-off at 0.05 (blue) and 0.01 (red). For the coleoptile length variation plot, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p<0.0001$.

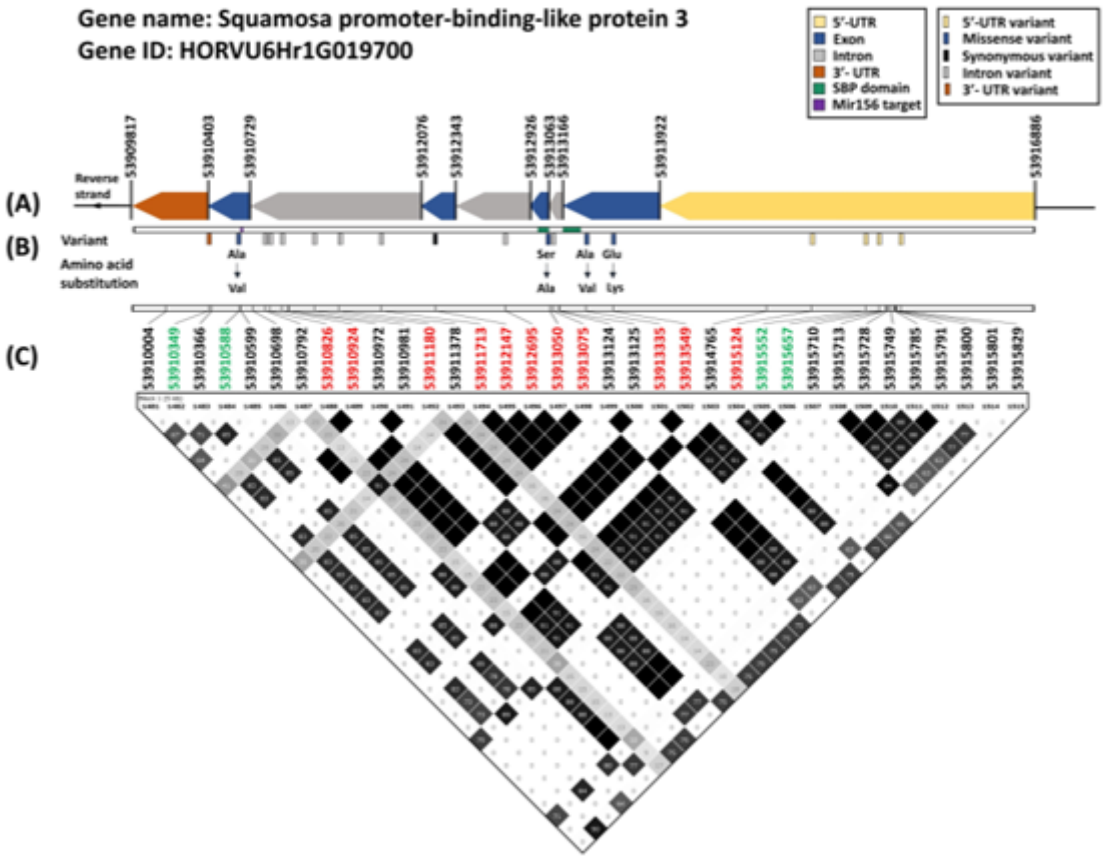

Figure 4

Summary of the SPL3 gene structure (A), amino acid substitution (B) and the local LD and haplotype blocks for genomic region with all 35 detected SNPs (C). (A) The gene structure was annotated in barley genome assembly IBSC v2 (Mascher et al., 2017). (B) the amino acid substitutions were predicted by detected GWAS markers. (C) The LD plot was generated in Haploview and indicated r2 values between pairs of SNPs multiplied by 100; white, r2 = 0; shades of grey, 0 < r2 < 1; black, r2 = 1. The haplotype block in the SPL3 genomic region were defined with the four-gamete rule method. The SNPs showing significant association (0.01<q value<0.05) in the GLM model were highlighted in green font. The SNPs showing highly significant association (q value<0.01) in the GLM model were highlighted in red font.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile3.docx
- Additionalfile1.docx
- Additionalfile10.xlsx
- Additionalfile2.docx

- Additionalfile6.docx
- Additionalfile4.docx
- Additionalfile5.docx
- Table1.docx
- Additionalfile8.docx
- Additionalfile9.docx
- Additionalfile7.docx