

Whole-Genome Sequencing on 220 Alfalfa (*Medicago sativa* L.) Accessions Identified Loci Associated with Fall Dormancy

Fan Zhang

Chinese Academy of Agricultural Sciences

Junmei Kang

Chinese Academy of Agricultural Sciences

Ruicai Long

Chinese Academy of Agricultural Sciences

Mingna Li

Chinese Academy of Agricultural Sciences

Yan Sun

China Agricultural University

Fei He

Chinese Academy of Agricultural Sciences

Xueqian Jiang

Chinese Academy of Agricultural Sciences

Changfu Yang

Chinese Academy of Agricultural Sciences

Xijiang Yang

Chinese Academy of Agricultural Sciences

Jie Kong

Chinese Academy of Agricultural Sciences

Zhen Wang

Chinese Academy of Agricultural Sciences

Zhiwu Zhang

Washington State University

Qingchuan Yang (✉ qchyang66@163.com)

Chinese Academy of Agricultural Sciences <https://orcid.org/0000-0002-5926-9798>

Research Article

Keywords: GWAS, fall dormancy, *Medicago sativa*, DREB1C, freezing tolerance

Posted Date: November 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1006624/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Fall dormancy (FD) is one of the most important traits of alfalfa (*Medicago sativa*) for cultivar selection to overcome winter damage. Regrowth plant height following autumn clipping is an indirect way to evaluate FD. Although transcriptomics, proteomics analysis, and QTL mapping have revealed some important genes correlated with FD, the genetic architecture of this trait is still unclear. There are no applicable genes or markers for selection, which hinders progress in the genetic research and molecular breeding for the trait. We conducted whole-genome sequencing (WGS) on 220 alfalfa accessions at 10x depth. Among the 875,023 SNPs, seven of them were associated with FD using genome-wide association study (GWAS). One SNP located on chromosome 6 is in linkage disequilibrium with dehydration-responsive element-binding protein 1C (*DREB1C*). Furthermore, seven DREB genes are clustered in this region, one of which has previously been shown to enhance freezing tolerance in the model plant *Medicago truncatula*. The candidate genes uncovered by our research will benefit the transgenic and CRISPR-Cas9 research of FD in alfalfa. This gene will also be useful for marker development and assisted selection of FD for alfalfa.

Key Message

Seven SNP associated to fall dormancy in alfalfa and one candidate gene *DREB1C* were identified using GWAS method combined with whole genome sequencing.

Introduction

Fall dormancy (FD) is the adaptive growth characteristic of alfalfa in autumn and can be evaluated by the regrowth plant height at 25-30 days after final cutting (Teuber et al. 1998).

The taller alfalfa grows in the fall, the less dormant it is. The FD level of alfalfa varieties can be classified from 1 to 11, where 1 represents most dormant and 11 represents least dormant (Teuber et al. 1998). Dormancy is induced by falling temperature and shortening photoperiod. Dormancy can improve plant survival in winter and increase freezing tolerance. FD alfalfa cultivars have better winter survival and hardiness than non-dormant alfalfa cultivars (Barnes et al. 1979; Smith 1961; STOUT 1985; STOUT and HALL 1989). FD level can be considered one of the indices for alfalfa cultivar selection in specific regions. There is a trade-off between FD and yield. High FD could increase overwintering ability, but decrease the potential yield (Avci et al. 2018; STOUT and HALL 1989). Breaking down the linkage between FD and yield from genetic level could benefit for both yield and winter hardiness. Studying the genetic architecture of FD could be used to improve those high yield alfalfa cultivars with bad cold tolerance (Li et al. 2015; Munjal et al. 2018).

The genetic mechanism of FD in alfalfa has been studied using genomics, transcriptomics, and proteomics. QTL mapping between different FD level alfalfa germplasms has been conducted in several studies. Seventy-one QTL related to FD traits have been identified, and 15 of them can be detected in more than one environment (Li et al. 2015). Another study identified 45 significantly associated QTL for FD (Adhikari et al. 2018). Both of these groups found that FD and winter hardiness have independent

inheritance and could be improved independently. A gene expression-related experiment showed that a cold acclimation-responsive gene, RootCAR1, was positively associated with alfalfa winter survival (Cunningham et al. 2001). More FD-related genes have been found using transcriptomics technology. Forty-four genes were identified by comparison of the transcriptomes from non-FD versus FD cultivars' leaves. The transcription of the genes IAA-amino acid hydrolase ILR1-like 1, abscisic acid receptor PYL8, and monogalactosyldiacylglycerol synthase-3 were suggested to be involved in regulating FD in alfalfa (Du et al. 2017). Eight significantly differentially expressed transcription factors related to CBF and ABRE-BFs were identified (Liu et al. 2019). Furthermore, FD-related proteins were analyzed using proteomics and metabolomics. A total of 90 proteins were different between FD and non-FD alfalfa. One of these was thiazole biosynthetic enzyme (*MsThi*), which is important for alfalfa growth (Du et al. 2018). Raffinose family oligosaccharide (RFO) metabolism was shown to be involved in short photoperiod-induced freezing tolerance in dormant alfalfa cultivars (Bertrand et al. 2017).

Genome-wide association studies (GWAS) are widely used to locate precise SNPs associated with phenotypes using historical recombination information. With the development of the sequencing method and genetic analysis method of GWAS, the candidate SNP markers for important quantitative traits have been identified across different populations (Biazzi et al. 2017; Wang et al. 2016; Yu et al. 2016). Based on 198 accessions, Longxi-Yu reported a set of GWAS results. It includes 19 SNPs associated with drought resistance (Zhang et al. 2015). 36 SNPs associated with salt tolerance (Yu et al. 2016), 42 associated with plant growth and forage production (Liu and Yu 2017), and 131 markers associated with 26 forage quality traits (Lin et al. 2020). Using the 336 genotypes, Wangzan conducted GWAS for several key alfalfa traits, including fiber-related traits and digestibility (Wang et al. 2016), crude protein and mineral concentrations (Jia et al. 2017), and nine biomass-related traits (Wang et al. 2020). Another study reported several SNPs associated with forage quality using half-sib progeny developed from three cultivars (Biazzi et al. 2017). Furthermore, several genes were revealed to be related to some important traits of alfalfa. For example, several stress-responsive genes associated with yield under water deficit were identified, including leucine-rich repeat receptor-like kinase, B3 DNA-binding domain protein, translation initiation factor IF2, and phospholipase-like protein (Liu and Yu 2017). Two markers linked to *NIR-NBS-LRR* genes were significantly associated with *Verticillium* wilt resistance (Yu et al. 2017). A cell wall biosynthesis gene associated with several cell wall-related traits that affect alfalfa nutritive value has been identified (Sakiroglu and Brummer 2017). *MsACR11* has an effect on plant height, which could significantly increase the height of transgenic arabidopsis (Wang et al. 2020).

Although many GWAS have been performed in alfalfa, few of these have focused on FD. Furthermore, with the lack of reference genome in alfalfa, most previous studies have used the *Medicago truncatula* genome as a reference (Sakiroglu and Brummer 2017; Wang et al. 2020; Yu et al. 2017). With the technological development of the genome assembly, the genome of the cultivar "Zhongmu No.1" has been released. A flowering locus T homolog gene, *MsFTa2*, may be associated with fall dormancy and salt resistance and was identified using the GWAS method and the new reference genome for alfalfa (Shen et al. 2020).

In this study, we evaluated FD (we define FD as regrowth plant height at 30 days following autumn clipping) in 220 accessions and conducted WGS with the mean depth of 10x sequencing for every accession. A total of 875,023 SNPs were detected in the alfalfa reference genome and used for conducting GWAS. The objectives of this experiment were (1) evaluate the linkage disequilibrium (LD) and population structure of alfalfa, (2) identify SNPs associated with FD, and (3) infer the candidate genes controlling FD.

Materials And Methods

Plant materials and phenotyping

The plant materials used in this study consisted of 220 accessions collected from all over the world. These accessions include 55 cultivars, 26 cultivated materials, 95 landraces, 4 breeding materials, 16 wild materials, and 24 uncertain improvement status materials. Among them, 26 accessions come from the database of the Medium Term Library of National Grass Seed Resources of China. Another 194 accessions come from the database of the U.S. National Plant Germplasm System (USDA GRIN). The geographic sources of the accessions include China, the United States, Turkey, Afghanistan, Tashkent Uzbekistan, Spain, Russia, Morocco, France, Argentina, and others (Table S1).

In October 2017, the seeds of 220 accessions were planted in the greenhouse of the Chinese Academy of Agricultural Sciences (CAAS) in Langfang, Hebei Province, China (39.59°N, 116.59° E). The area has a continental monsoon climate with an average temperature of 11.9 °C/year. The coldest month is January with an average temperature of -4.7 °C and the hottest month is July with an average temperature of 26.2 °C. The annual rainfall is 554.9 mm. The soil type is loam soil with the PH value of 7.37 and soil organic matter content is 1.69%. Individuals were transplanted to the field of CAAS in April 2018. A randomized complete block design with three replications was used for the experimental design. Every accession included 5 individuals placed 30 cm apart in a single row for one replication. The accession spacing was 65 cm between rows and columns. To keep all individuals uniform, they were clipped to a height of 5 cm after transplant. To keep same evaluation standard with field management of alfalfa, all individuals were clipped at the early flowering stage (when 10% of plants begin flowering). Plants were clipped three times in 2018 and four times in 2019. Fall dormancy was collected using plant regrowth one month after final clipping (31 Oct 2018 and 29 Oct 2019). The tallest stem height of one individual was considered FD. The correlation information between FD of the mean value of different individuals (3 replication x 5 individual=15 individuals) and FD of sequencing individual was analyzed using ggplot2 (Wickham 2011).

The variance of fall dormancy was analyzed using a generalized linear model (GLM) as follows:

$$\text{height} = \text{accession} + \text{ind}(\text{accession}) + \text{year} + \text{accession} * \text{year} + \text{replicates} + \text{error}$$

All factors were random, and the GLM was performed using PROC GLM (SAS Institute, 2010). The broad-sense heritability (H^2) was computed as follows:

$$H2 = \sigma_g^2 / (\sigma_g^2 + \frac{\sigma_{gy}^2}{y} + \frac{\sigma^2}{yl})$$

where σ_g^2 is the variance of accession; σ_{gy}^2 is the interaction variance of accession and year; σ^2 is the residual variance. The items y and i in the equation refer to the number of years and individuals, respectively.

Sequencing and SNP calling

Phenotyping was conducted on 15 individuals per accession and one individual with the typical phenotype was selected for sequencing. Young leaves were selected at the early regrowth stage (9 July 2019). Total DNA was extracted using the CWBIO Plant Genomic DNA Kit (CoWin Biosciences, Beijing, China), according to the manufacturer's protocol. At least 6 ug of genomic DNA from each accession was used to construct a sequencing library following the manufacturer's instructions (Illumina Inc.). Paired-end sequencing libraries with an insert size of approximately 300 bp were sequenced on an Illumina NovaSeq 6000 sequencer at BerryGenomics company. The data size of every accession is 10 Gb and the average Q30 is 85%. The raw data has been uploaded to the National Genomics Data Center (NGDC, <https://bigd.big.ac.cn/>) under BioProject PRJCA004024 and NCBI Sequence Read Archive with BioProject ID: PRJNA739212. Sequencing data were first quality filtered using Trimmomatic software with default parameters (Bolger et al. 2014). Paired-end sequencing reads were mapped to the alfalfa reference genome (haploid genome with 8 chromosomes (Long et al. 2021)) with BWA-MEM using default parameters (Li 2013). SAMtools were used to translate SAM file to BAM file and sort BAM files using default parameters (Li et al. 2009).

Picard Tools was used to mark duplicate reads (<http://broadinstitute.github.io/picard/>), and Genome Analysis ToolKit was used to correct indels which can be mistaken for SNPs (Van der Auwera et al. 2013). SAMtools mpileup and VarScan were used to detect SNPs (Koboldt et al. 2012). Furthermore, SNP data were filtered using VCFtools (Danecek et al. 2011) with a missing rate of less than 10%, a minor allele frequency of more than 0.05, and mean read depth greater than 20. The SNP distribution on the chromosome was generated using the R code from a GWAS study in beef cattle (Zhou et al. 2019).

Population structure and linkage disequilibrium

Population structure was calculated using the TASSEL 5 software with principal component analyses (PCA) (Bradbury et al. 2007). The 3D PCA results combined with geography information were plotted using R package scatterplot3d (Ligges and Mächler 2002). The population structure subgroup was estimated using admixture and the figure was generated using R package pophelper (Francis et al, 2017). LD information was calculated using the software PopLDdecay with default parameters (Zhang et al. 2019). The VCF file containing information on all 875,023 SNP markers was imported to PopLDdecay. The LD results among all accessions were used to estimate the LD of alfalfa. The distance versus mean R^2 within 30 kb of LD was plotted using R package ggplot2.

Association mapping

The Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) method was used to carry out the GWAS (Huang et al. 2019) for two kinds of FD phenotype (sequencing individual and mean of 15 individuals). GWAS was conducted using the BLINK C version software. The Bonferroni multiple test correction was used to determine the significant SNP threshold ($P=0.01/875,023=1.14 \times 10^{-8}$). The BLINK method uses iterations to select a set of markers associated with a trait. These associated markers are fitted as a covariate for testing the remaining markers. This method better controls false positives than the kinship approach. The real data and simulated data results showed that BLINK has higher statistical power than other methods, such as MLM, FarmCPU, et al. (Huang et al. 2019). The Manhattan and Q-Q plots of GWAS results were created using the R package qqman (Turner et al. 2014).

Results

Phenotypic variance

To identify the genetic difference in FD for different individual among the same accession, the variance of the phenotype was analyzed. The effect of accession, year, individual, and accessions–year interaction were estimated (Table1). FD was recorded for two years: 2018 and 2019. The FD correlation between sequencing individual and mean of 15 individuals was used to estimate individual variance. The correlation R^2 value was 0.46 and 0.55 in 2018 and 2019. The correlation R^2 value across two years was 0.47 (Figure 1). The FD correlation R^2 between 2018 and 2019 among sequencing individuals and accession mean were 0.41 and 0.27 (Figure S1). The variance of FD in 220 accessions was analyzed using ANOVA. Significant differences in accession, year, and accession-year interaction were observed at the 0.001 probability level ($p < 0.001$). The variation of individuals was not significant at the 0.05 probability level ($p = 0.12$), indicating that the difference in FD between individuals is not big enough and the mean value can be used to represent the FD of one accession (Table 1). The broad-sense heritability was 69.2% for FD.

Genotype calling and allele frequencies

A total of 32.2 million SNPs were detected using the BWA-SAMTools-VarScan pipeline. After filtering the SNPs with a missing value $>10\%$, minimum mean read depth >20 , and minor allele frequency (MAF) >0.05 , there were 875,023 SNPs that passed the threshold and were confirmed for further analysis. Among these 875,023 SNPs, the average SNP density in the genome was one SNP per 26.95 bp (among the 814,517 SNP with SNP interval less than 300 bp) (Figure 2a, b). There were 60,506 SNP with an interval larger than 300bp, and the average SNP interval for these was 11128.66 bp (Figure 2b). The distribution of the MAF value showed that the value equal to 0.06 has the highest density (Figure 2c, detail data not shown). The SNP number range is similar for MAF values from 0.2 to 0.5 (Figure 2c). These results demonstrate that the markers can be assigned to unbiased chromosomes and distinct positions.

Population structure

The accessions used in our research originated from different countries, especially the USDA-GRIN collection, representing wide coverage of genetic diversity on alfalfa. To analyze population structure, the genome-wide SNP data generated by WGS was used for principal component analysis (PCA) and population structure analysis. The PCA scatter-plots showed a weak population structure for the 220 accessions consisting of different geographical origins with no clear subpopulation (Figure 3). In the scatterplot of PC1-3, most accessions are mixed. But some accessions coming from China can be separated from those with other origins (red color). Similarly, some accessions from South America can be separated from the others (cyan color). The accessions from Europe, North America, and other Asia countries have wide and mixed distribution. The geographic distribution of accessions interacted with PC1 showed similar results and the accessions coming from China and South America were different from others (Figure S2). The first three PCs explained 2.6%, 1.8%, and 1.2% genetic variance separately; the total genetic variance explained by the first three PCs is 5.6%. Population structure analyses were conducted by admixture. The best subgroup was estimated by cross-validation (CV) error with a K value range from 2 to 10. The CV results showed K=3 has the smallest error (Figure S3 and Figure S4). The relationship between subgroup were illustrated using population structure results from K=2 to K=4. The accessions coming from China can be separate for K=3. The accessions from South America and Africa can be separate for K=4 (Figure 4). These mixed clusters indicate that, although accessions come from different geographical origins, they still have a high degree of genetic relatedness. Furthermore, the variance of FD among different geographic origins has a similar difference pattern. The accessions coming from Africa have the highest FD, while the accessions coming from China have lower FD. The FD variance of South America was in the middle of Africa and China. The FD variance of accessions coming from other origins (Europe, North America, and other Asia) is either similar to FD in China or South America (Figure S5).

Linkage disequilibrium

To estimate the confidence interval of GWAS results, LD was analyzed using a total of 875,023 markers. The r^2 of LD across all chromosomes was used to estimate the LD of alfalfa. A rapid drop in r^2 was observed with the increase in physical distance. Over all the accessions, the value of r^2 decreased quickly within 2 kb physical distance, then decreased slowly afterward (Figure 2d). The value of r^2 decreased to 0.1 after 18 kb across all chromosomes. The results showed that extensive recombination took place during alfalfa evolution and distribution.

GWAS and candidate genes for FD

To identify SNP markers associated with FD and compare the effect of different kind of phenotype, genotypic and phenotypic data were analyzed using the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) method in the BLINKC version software. SNP effect was calculate using the BLINK R version Prediction function. Both kind of phenotype (mean of 15 individuals and sequencing individual) were used to conduct GWAS analysis.

The quantile-quantile (Q-Q) plot results of marker-trait associations for FD were illustrated using observed versus expected p-values (Figure 5). Any deviation from the expected red line implies SNP association with

FD. In the present study, a total of four SNPs passed the 1% threshold after a Bonferroni correction ($p < 1.14 \times 10^{-8}$) and were associated with FD of mean of 15 individuals and three SNPs associated with FD of sequencing individual (Figure 5, Table 2). Among them, two markers, chr2__23380918 and chr2__80663885, were located on chromosome 2. In addition, 2 markers (chr6__740339 and SNP chr6__69737395) were located on chromosome 6. Another three markers associated with FD of sequencing individual were located on chromosome 5 (chr5__36702273), 7 (chr7__49431088) and 2 (chr2__58174960). The FD difference between different genotypes can be observed in the seven SNP boxplots. There is no homozygous genotype CC for SNP chr6__740339 and there is no homozygous genotype TT for SNP chr5__36702273 and no AA for SNP chr2__58174960 (Figure 6). These results may suggest that CC, TT or AA genotype is the deleterious variance for alfalfa.

To identify potential candidate genes linked to SNP associated with FD, candidate genes were screened based on the up- and downstream 18 kb LD interval information. Of the seven significantly associated SNP, four of them are linked to known genes in the alfalfa genome (Table 2). There is one candidate gene (*Msa0867660*) with the annotated function of signal anchor protein connected to chr6__740339. Only one gene (*Msa0892030*), dehydration responsive element binding protein 1C (DREB1C), was connected with the SNP chr6__69737395 in the 36 kb genomic interval based on LD information (Figure S6). We further checked the genes near SNP chr6__69737395 with a larger scope (500 kb interval up- and downstream). Another six DREB genes were found upstream of this SNP (chr6:69234712-69716598). One of the six genes (*Msa0891950*) has been validated to be associated with freezing tolerance in *Medicago truncatula* (Figure S7) (Chen et al. 2010). The transgenic *Medicago truncatula* has lower plant height than normal plants and has better tolerance for freezing. We further checked all DREB genes on chromosome 6, and found that there were nine in total (Table S2). Another two DREB genes were *Msa0893350* and *Msa0893360* (chr6:76949157-76957350). There is only one cluster region with seven DREB genes on chromosome 6, and this region corresponds to our significant GWAS signal. Another two SNP have more than one candidate gene. The SNP chr5__36702273 is connected with one hypothetical protein gene (*Msa0729370*) and transcription factor FAMA isoform X2 gene (*Msa0729380*). The SNP chr7__49431088 is connected with two transcription factor bHLH155 genes (*Msa1022700* and *Msa1022710*).

Discussion

The FD variance of alfalfa

Because of the outcrossing and self-incompatibility feature of alfalfa, the genotype between different individuals is different. Most studies collect phenotypes using cloned plants (Li et al. 2015; Liu and Yu 2017; Sakiroglu and Brummer 2017; Wang et al. 2016). For one study, seed plants of the same accession were used for phenotype collection (Yu et al. 2016). Compared with cloned plants, the advantages of using seed plants are that they are easier to establish and have better accession representation. Furthermore, some experiments and assessment of phenotypes cause damage to the plants, such as winter survival, disease resistance, and root-related phenotypes. It is easier to rebuild the population using seed than by cloning plants. We checked the FD variance of individuals based on our experiment. To

check the influence of individuals' genetic differences for FD among the same accessions, we nested individuals in accession. Our results showed that the variance of individuals is not significant, while the variance of accessions is significant (Table 1). That means we could use the mean value to serve as a representative phenotype of one accession. Even though there is genetic variance among different individuals, their mean value is the whole picture of one accession. Furthermore, the correlation between sequencing individual plants and the mean value of different plants is high enough. The correlation R^2 value is 0.46 and 0.55 in 2018 and 2019 (Figure 1). While the correlation R^2 between 2018 and 2019 is 0.41 and 0.27 for mean FD and sequencing individual FD (Figure S1). It showed that environmental effect has bigger influence for a single individual. The mean value of multiple individuals is more stable among the different environments. In order to test the genetic variance of these two kinds of FD phenotype (sequencing individual and mean of 15 individuals). We tested the GWAS signal for both kinds of phenotypes and found associated SNP (Figure 5, Table S3). However, there is no overlap SNP for them. This phenomenon has been observed in previous studies (Heller and Yekutieli 2014; Lin 2019; Rietveld et al. 2014). We think FD maybe controlled by several different genes. Our two kinds of phenotypes corresponding to two different kinds of genes. Combine both associated SNP could uncover a better understanding of FD (Rietveld et al. 2014). Combining ANOVA and correlation results of FD and GWAS results, we think the mean of FD could be used for genetic analysis of alfalfa.

Our experiment also showed that FD evaluation in one location for three years is very hard. Because the accessions were collected from all over the world. The fall dormancy level is very different. For the accessions with low dormancy ability, their winter hardiness is weak and winter survival is a problem. It is hard to control the balance between sample diversity and local adaptation. If we want to collect more trails data, these accessions not adapted local environment will die and our phenotype will be biased. That's why we only collected two years' data. Our purpose is to gain more genetic variance of alfalfa. However, more test trails with different climatic environments may be a possible way to solve this problem.

The genetic diversity of alfalfa

The improvement of alfalfa is mainly based on outcrossing, and the genetic exchange of alfalfa always exists in alfalfa-growing regions (Li and Brummer 2012). To this end, nine distinct germplasms with global backgrounds were introduced in North America to cultivate new varieties (Barnes 1977). It is more difficult to distinguish alfalfa accessions based on geographic origin than for self-pollinating plants, such as soybean (Zhou et al. 2015) and rice (Huang et al. 2012). Also, diploid alfalfa (*M. sativa subsp. caerulea* and *M. sativa subsp. falcata*) can be clustered based on subspecies and geography (Sakiroglu and Brummer 2017; Şakiroğlu et al. 2010); the genetic diversity of tetraploid alfalfa (cultivated alfalfa) is more complicated. Based on our PCA results and population structure results, the genetic differentiation among different origins is not very high, except the ones from China and South America.

The introduction of alfalfa to South America took place around the sixteenth century, while it was introduced to China 2000 years ago (Annicchiarico et al. 2015). Differentiation has occurred along with cultivation in the local region. This phenomenon has also been identified in previous studies (Qiang et al.

2015; Shen et al. 2020; Wang et al. 2020). The accessions coming from Europe and other Asia regions (regions except for China) are mixed and have a wide range. These accessions have the highest genetic diversity among all accessions with different geography. This result may be the reason that alfalfa originated from the Caucasus region: northeastern Turkey, Turkmenistan, and northwestern Iran (Michaud et al. 1988). This theory can also be strengthened by the knowledge that accession genetic diversity from North America and Africa is relatively low compared to European accessions. The cultivated alfalfa was exported to North America and Africa only around 200~300 years ago (Prosperi et al. 2014). This relatively short cultivation period also partially explains the low genetic diversity.

The linkage disequilibrium of alfalfa

Our results also showed that the LD distance is very short (R^2 reduced to 0.1 after 18 kb). This is different from previous studies (Sakiroglu and Brummer 2017) and (Li et al. 2014). One major reason may be because we used a large number of SNPs (around 58 times more than the previous report). With increased marker density, it is easier to estimate historical recombination events. Another study showed that LD can be decreased rapidly to less than 1 kb in alfalfa using candidate gene-related SNP (Herrmann et al. 2010). The rapid LD decrease is likely due to the outcrossing nature of alfalfa and the diverse germplasm we collected.

The candidate genes of FD

A total of seven SNPs were determined to be associated with FD based on the 1% Bonferroni test (Table 2 and Figure 5). Four out of seven SNPs can be mapped to candidate genes based on LD information, and one SNP is connected with a *DREB1C* gene. *DREB1C* has been shown to be associated with freezing tolerance. When the gene *CBF2/DREB1C* was disrupted, mutant *Arabidopsis* had a higher capacity to tolerate freezing than WT plants and were more tolerant to dehydration and salt stress (Novillo et al. 2004). Further validation has been done in model plant *Medicago truncatula*. Transgenic *Medicago truncatula* exhibited suppressed shoot growth and enhanced freezing tolerance (Chen et al. 2010). The transcriptomic analyses of different dormancy alfalfa cultivars showed that *DREB1C* was one of the most significantly differentially expressed genes and was upregulated in a fall dormant alfalfa cultivar (Liu et al. 2019). Another transcriptome sequencing analysis revealed that nine *DREB* unigenes responded to cold and/or freezing stresses (Shu et al. 2017). Even though FD and winter survival may be controlled by different QTL (Adhikari et al. 2018; Brummer et al. 2000; Li et al. 2015), FD serves as an indicator to evaluate freezing tolerance in the broad sense is possible. Regrowth plant height following autumn clipping can be considered a signal of fall dormancy and freezing tolerance. If an alfalfa cultivar has a higher regrowth plant height, it will also have less freezing tolerance (Sheaffer et al. 1992). We further confirmed that *DREB1C* contributes to FD using the GWAS method. And our gene expression data showed that the *DREB1C* gene can be induced by cold treatment. But the difference between high and low fall dormancy level alfalfa is not significant (Figure S8). Maybe our sample size (10 samples) used for expression analysis is not enough. However, combining several previous results from transgenic studies, transcriptomics, and GWAS showed that the *DREB1C* gene is very important for alfalfa freezing tolerance.

Validation of the *DREB1C* gene effects in alfalfa using gene-modification tools such as CRISPR/CAS9 (Chen et al. 2020) or a transgenic method (Chao et al. 2014) will give us a better understanding of the function of *DREB1C*.

In addition to the *DREB1C* gene, there should be other genes responsible for FD, because FD is a quantitative trait with continuous variance (Figure S9). We identified six genes associate with FD and one of which has a clear relationship with FD. Other genes identified by sequencing individual FD such as transcription factor FAMA isoform X2 and transcription factor bHLH155 may be also important for FD. These results tell us maybe there are different genes responsible for FD.

Moreover, there are some candidate genes responsible for FD that have been identified using omics technologies, such as *ABF4* from RNA-seq results (Liu et al. 2019) and *MsThi* from the iTRAQ-based method (Du et al. 2018). Cross-validation from multiple omics technologies could help us develop a more accurate understanding of FD. Our results showed that GWAS results could agree with RNA-seq data. Combining GWAS methods with other omics technologies on a bigger scale could be a possible way to uncover more details regarding FD for alfalfa (Scossa et al. 2021).

In conclusion, we conducted a GWAS analysis of FD in alfalfa using whole-genome sequencing on seeding plants. Weak population structure was found among different accessions belonging to different geographical origins. Seven SNPs were shown to be associated with FD. One candidate gene, *DREB1C*, was responsible for FD. This research will help us achieve a better understanding of the relationship between genotype and the FD phenotype. The results could serve as basic information for QTL mapping or candidate gene cloning to understand the mechanism of FD in alfalfa. Furthermore, analyzing the genetic basis of FD will help to develop cultivars with different FD.

Declarations

Funding

This work was supported by the National Natural Science Foundation of China (No. 31971758), and the breeding forage and grain legumes to increase China's and EU's protein self-sufficiency, collaborative research key project between China and EU (2017YFE0111000/EUCLEG 727312), and Key Projects in Science and Technology of Inner Mongolia (2021ZD0031) and the China Scholarship Council. The funding body played no role in the design of the study, the collection, analysis, and interpretation of data and the writing of the manuscript.

Competing interests

The authors declare that they have no conflicts of interest.

Availability of data and materials

All RAD raw sequence data were upload to the National Genomics Data Center (NGDC, <https://bigd.big.ac.cn/>) under BioProject PRJCA004024 (<https://ngdc.cncb.ac.cn/search/?dbId=biosample&q=PRJCA004024&page=1>) and NCBI Sequence Read Archive with Bioproject ID: PRJNA739212 (<https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA739212>). The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Code availability

The software used in the current study were supplied in M&M. Any questions related to software usage, please contract with Fan Zhang (15600085453@163.com).

Authors' contributions

QCY conceived and designed the experiments. FZ, JMK, RCL, MNL, YS, FH, XQJ, CFY, XJY, ZW, and ZWZ performed the experiments. JMK, RCL, MNL and YS plant the alfalfa seed in the greenhouse. FZ, JMK, RCL FH, XQJ, CFY and XJY collect the alfalfa phenotype. JK and ZW conduct gene expression experiment. FZ, ZW and ZWZ analyzed the data. FZ, ZWZ, and QCY wrote the paper. All authors read and approved the final manuscript.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Acknowledgements

We thank the database of the Medium Term Library of National Grass Seed Resources of China and the database of the U.S. National Plant Germplasm System (USDA GRIN) for providing alfalfa accessions. We thank China Scholarship Council (CSC) for their support the international study of Fan Zhang.

References

Adhikari L, Mohseni-Moghadam M, Missaoui A (2018) Allelopathic effects of cereal rye on weed suppression and forage yield in Alfalfa. *American Journal of Plant Sciences* 9:685

Annicchiarico P, Nazzicari N, Li X, Wei Y, Pecetti L, Brummer EC (2015) Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC genomics* 16:1-13

- Avci M, Hatipoglu R, Çınar S, Kiliçalp N (2018) Assessment of yield and quality characteristics of alfalfa (*Medicago sativa* L.) cultivars with different fall dormancy rating. *Legume Research* 41:369-373
- Barnes D (1977) Alfalfa germplasm in the United States: Genetic vulnerability, use, improvement, and maintenance. Department of Agriculture, Agricultural Research Service
- Barnes D, RE S, LJ E (1979) Fall dormancy in alfalfa: A valuable predictive tool
- Bertrand A, Bipfubusa M, Claessens A, Rocher S, Castonguay Y (2017) Effect of photoperiod prior to cold acclimation on freezing tolerance and carbohydrate metabolism in alfalfa (*Medicago sativa* L.). *Plant Science* 264:122-128
- Biazzi E, Nazzicari N, Pecetti L, Brummer EC, Palmonari A, Tava A, Annicchiarico P (2017) Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. *PLoS One* 12:e0169234
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635
- Chao Y, Zhang T, Yang Q, Kang J, Sun Y, Gruber MY, Qin Z (2014) Expression of the alfalfa CCCH-type zinc finger protein gene *MsZFN* delays flowering time in transgenic *Arabidopsis thaliana*. *Plant Science* 215:92-99
- Chen H, Zeng Y, Yang Y, Huang L, Tang B, Zhang H, Hao F, Liu W, Li Y, Liu Y (2020) Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nature communications* 11:1-11
- Chen J-R, Lü J-J, Liu R, Xiong X-Y, Wang T-x, Chen S-Y, Guo L-B, Wang H-F (2010) *DREB1C* from *Medicago truncatula* enhances freezing tolerance in transgenic *M. truncatula* and China Rose (*Rosa chinensis* Jacq.). *Plant Growth Regulation* 60:199-211
- Cunningham S, Gana J, Volenec J, Teuber L (2001) Winter hardiness, root physiology, and gene expression in successive fall dormancy selections from 'Mesilla' and 'CUF 101' alfalfa. *Crop science* 41:1091-1098
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156-2158
- Du H, Shi Y, Li D, Fan W, Wang G, Wang C (2017) Screening and identification of key genes regulating fall dormancy in alfalfa leaves. *PLoS one* 12:e0188964
- Du H, Shi Y, Li D, Fan W, Wang Y, Wang G, Wang C (2018) Proteomics reveals key proteins participating in growth difference between fall dormant and non-dormant alfalfa in terminal buds. *Journal of proteomics*

173:126-138

- Heller R, Yekutieli D (2014) Replicability analysis for genome-wide association studies. *The Annals of Applied Statistics* 8:481-498
- Herrmann D, Barre P, Santoni S, Julier B (2010) Association of a CONSTANS-LIKE gene to flowering and height in autotetraploid alfalfa. *Theoretical and applied genetics* 121:865-876
- Huang M, Liu X, Zhou Y, Summers RM, Zhang Z (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* 8:giy154
- Huang X, Kurata N, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:497-501
- Jia C, Wu X, Chen M, Wang Y, Liu X, Gong P, Xu Q, Wang X, Gao H, Wang Z (2017) Identification of genetic loci associated with crude protein and mineral concentrations in alfalfa (*Medicago sativa*) using association mapping. *BMC plant biology* 17:1-7
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* 22:568-576
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079
- Li X, Alarcón-Zúñiga B, Kang J, Nadeem Tahir MH, Jiang Q, Wei Y, Reyno R, Robins JG, Brummer EC (2015) Mapping fall dormancy and winter injury in tetraploid alfalfa. *Crop Science* 55:1995-2011
- Li X, Brummer EC (2012) Applied genetics and genomics in alfalfa breeding. *Agronomy* 2:40-61
- Li X, Han Y, Wei Y, Acharya A, Farmer AD, Ho J, Monteros MJ, Brummer EC (2014) Development of an alfalfa SNP array and its use to evaluate patterns of population structure and linkage disequilibrium. *PLoS One* 9:e84329
- Ligges U, Mächler M (2002) Scatterplot3d-an r package for visualizing multivariate data. Technical Report
- Lin S, Medina CA, Boge B, Hu J, Fransen S, Norberg S, Yu L-X (2020) Identification of genetic loci associated with forage quality in response to water deficit in autotetraploid alfalfa (*Medicago sativa* L.). *BMC Plant Biology* 20:1-18
- Lin X (2019) Reproducibility and replicability in large scale genetic studies. National Academies Press

- Liu X-P, Yu L-X (2017) Genome-wide association mapping of loci associated with plant growth and forage production under salt stress in alfalfa (*Medicago sativa* L.). *Frontiers in Plant Science* 8:853
- Liu Z-Y, Baoyin T, Li X-L, Wang Z-L (2019) How fall dormancy benefits alfalfa winter-survival? Physiologic and transcriptomic analyses of dormancy process. *BMC plant biology* 19:1-13
- Long R, Zhang F, Zhang Z, Li M, Chen L, Wang X, Liu W, Zhang T, Yu L-X, He F (2021) Assembly of chromosome-scale and allele-aware autotetraploid genome of the Chinese alfalfa cultivar Zhongmu-4 and identification of SNP loci associated with 27 agronomic traits. *bioRxiv*
- Michaud R, Lehman W, Rumbaugh M (1988) World distribution and historical development. *Alfalfa and alfalfa improvement* 29:25-91
- Munjal G, Hao J, Teuber LR, Brummer EC (2018) Selection mapping identifies loci underpinning autumn dormancy in alfalfa (*Medicago sativa*). *G3: Genes, Genomes, Genetics* 8:461-468
- Novillo F, Alonso JM, Ecker JR, Salinas J (2004) CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in *Arabidopsis*. *Proceedings of the National Academy of Sciences* 101:3985-3990
- Peterson BG, Carl P, Boudt K, Bennett R, Ulrich J, Zivot E, Cornilly D, Hung E, Lestel M, Balkissoon K (2018) Package 'performanceanalytics'. R Team Cooperation
- Prosperi J-M, Jenczewski E, Muller M-H, Fournier S, Sampoux J-P, Ronfort J (2014) Alfalfa domestication history, genetic diversity and genetic resources. *Legume Perspectives* 4:13-14
- Qiang H, Chen Z, Zhang Z, Wang X, Gao H, Wang Z (2015) Molecular diversity and population structure of a worldwide collection of cultivated tetraploid alfalfa (*Medicago sativa* subsp. *sativa* L.) germplasm as revealed by microsatellite markers. *PLoS One* 10:e0124592
- Rietveld CA, Conley D, Eriksson N, Esko T, Medland SE, Vinkhuyzen AA, Yang J, Boardman JD, Chabris CF, Dawes CT (2014) Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychological science* 25:1975-1986
- Sakiroglu M, Brummer EC (2017) Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. *Theoretical and applied genetics* 130:261-268
- Şakiroğlu M, Doyle JJ, Brummer EC (2010) Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. *Theoretical and applied genetics* 121:403-415
- Scossa F, Alseekh S, Fernie AR (2021) Integrating multi-omics data for crop improvement. *Journal of Plant Physiology* 257:153352

- Sheaffer C, Barnes D, Warnes D, Lueschen W, Ford H, Swanson D (1992) Seeding-year cutting affects winter survival and its association with fall growth score in alfalfa. *Crop Science* 32:225-231
- Shen C, Du H, Chen Z, Lu H, Zhu F, Chen H, Meng X, Liu Q, Liu P, Zheng L (2020) The Chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasm provide genomic resources for alfalfa research. *Molecular Plant* 13:1250-1261
- Shu Y, Li W, Zhao J, Zhang S, Xu H, Liu Y, Guo C (2017) Transcriptome sequencing analysis of alfalfa reveals CBF genes potentially playing important roles in response to freezing stress. *Genetics and molecular biology* 40:824-833
- Smith D (1961) Association of fall growth habit and winter survival in alfalfa. *Canadian Journal of Plant Science* 41:244-251
- STOUT DG (1985) Fall growth and stand persistence of alfalfa in interior British Columbia. *Canadian journal of plant science* 65:935-941
- STOUT DG, HALL JW (1989) Fall growth and winter survival of alfalfa in interior British Columbia. *Canadian Journal of Plant Science* 69:491-499
- Teuber L, Taggard K, Gibbs L, McCaslin M, Peterson M, Barnes D (1998) Fall dormancy. Standard tests to characterize alfalfa cultivars:A-1
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics* 43:11.10. 11-11.10. 33
- Wang J, Zhang Z (2020) GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *bioRxiv*
- Wang Z, Qiang H, Zhao H, Xu R, Zhang Z, Gao H, Wang X, Liu G, Zhang Y (2016) Association mapping for fiber-related traits and digestibility in alfalfa (*Medicago sativa*). *Frontiers in plant science* 7:331
- Wang Z, Wang X, Zhang H, Ma L, Zhao H, Jones CS, Chen J, Liu G (2020) A genome-wide association study approach to the identification of candidate genes underlying agronomic traits in alfalfa (*Medicago sativa* L.). *Plant biotechnology journal* 18:611
- Wickham H (2011) *ggplot2*. *Wiley Interdisciplinary Reviews: Computational Statistics* 3:180-185
- Yu L-X, Liu X, Boge W, Liu X-P (2016) Genome-wide association study identifies loci for salt tolerance during germination in autotetraploid alfalfa (*Medicago sativa* L.) using genotyping-by-sequencing. *Frontiers in plant science* 7:956
- Yu LX, Zheng P, Zhang T, Rodriguez J, Main D (2017) Genotyping-by-sequencing-based genome-wide association studies on *Verticillium* wilt resistance in autotetraploid alfalfa (*Medicago sativa* L.). *Molecular*

Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35:1786-1788

Zhang T, Yu L-X, Zheng P, Li Y, Rivera M, Main D, Greene SL (2015) Identification of loci associated with drought resistance traits in heterozygous autotetraploid alfalfa (*Medicago sativa* L.) using genome-wide association studies with genotyping by sequencing. *PLoS One* 10:e0138931

Zhou J, Liu L, Chen CJ, Zhang M, Lu X, Zhang Z, Huang X, Shi Y (2019) Genome-wide association study of milk and reproductive traits in dual-purpose Xinjiang Brown cattle. *BMC genomics* 20:1-11

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature biotechnology* 33:408-414

Tables

Table 1. Analysis of variance for FD in GWAS population over two years.

Source	df	Type III SS	F-value	P-value	Variance
accession	219	111676.00	10.73	<.0001	15.41
year	1	116825.50	2458.49	<.0001	35.39
ind(accession)	866	43915.99	1.07	0.12	0.53
repeat	2	382.01	4.02	0.02	0.07
accession*year	204	22669.18	2.34	<.0001	4.24

Table 2. Significant SNP markers associated with FD.

FD type	SNP	Allele	Chr	P-value	SNP Effects	Candidate Gene	Gene Annotation
Mean of 15 individuals	chr2_23380918	C/T	2	2.49E-12	21.84	–	–
	chr2_80663885	C/A	2	5.82E-09	-22.17	–	–
	chr6_740339	A/C	6	1.21E-09	-17.86	Msa0867660	signal anchor
	chr6_69737395	T/C	6	8.59E-09	19.62	Msa0892030	DREB1C
Sequencing individual	chr5_36702273	A/T	5	2.43E-20	21.8371	Msa0729370, Msa0729380,	hypothetical protein, transcription factor FAMA isoform X2
	chr7_49431088	T/C	7	8.24E-10	-22.1709	Msa1022700, Msa1022710,	transcription factor bHLH155 (two)
	chr2_58174960	G/A	2	6.91E-09	-17.8609	–	–

Figures

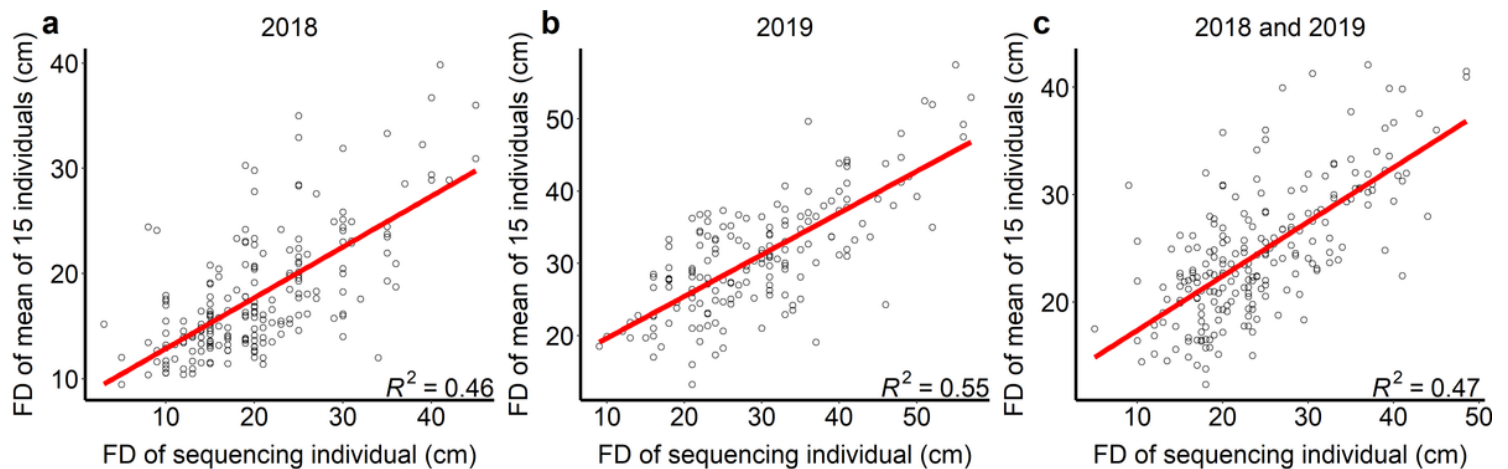


Figure 1

Correlation of FD between sequencing individual and mean of 15 individuals. Every cycle dot represented one accession. Regression line were drawn with red line and R squared value of linear regression were

present at the bottom right of the figure. a, the correlation information in 2018. b, the correlation information in 2019. c, the correlation information of 2018 and 2019.

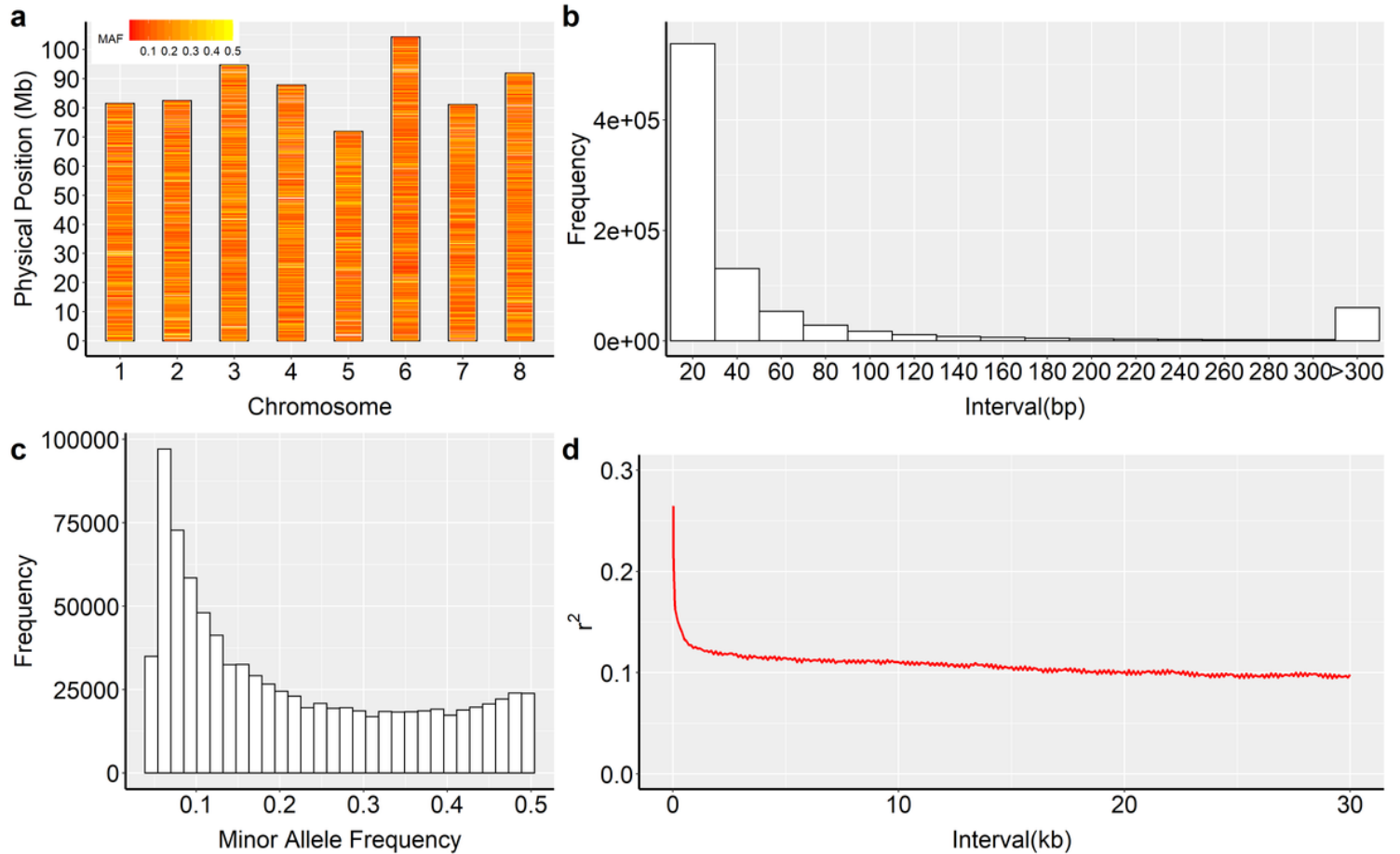


Figure 2

Properties of single nucleotide polymorphisms (SNPs). In total, 220 alfalfa accessions were genotyped by WGS; 875,023 SNPs passed filters and quality control. Marker distributions are displayed as a heatmap on 8 chromosomes by minor allele frequency (MAF) (a). Marker density is displayed by histogram according to the interval of adjacent SNPs (b). MAF distribution is shown in a histogram (c). LD decay is shown according to the red trend line (d).

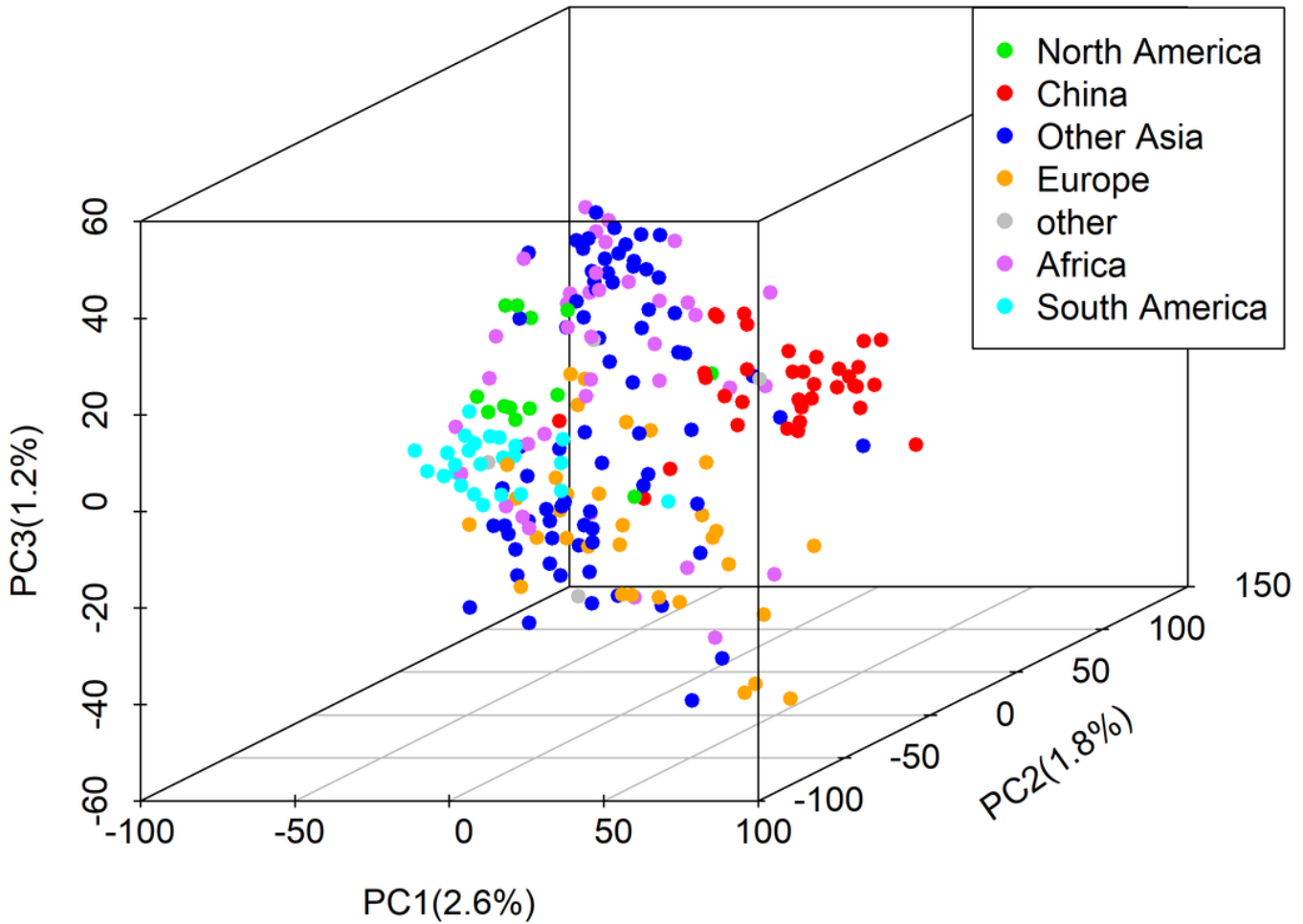


Figure 3

Population structure from the principal component analysis. A total of 875,023 SNPs and 220 alfalfa accessions were used to perform the principal component analysis. Population structure is shown as a 3D plot (D) of the first three principal components (PC) with colored points that define the seven groups. There are 13, 35, 73, 32, 4, 38, and 25 accessions in the groups corresponding to Africa, China, Europe, North America, other (not belonging to any other group), other Asia (except China), and South America, respectively.

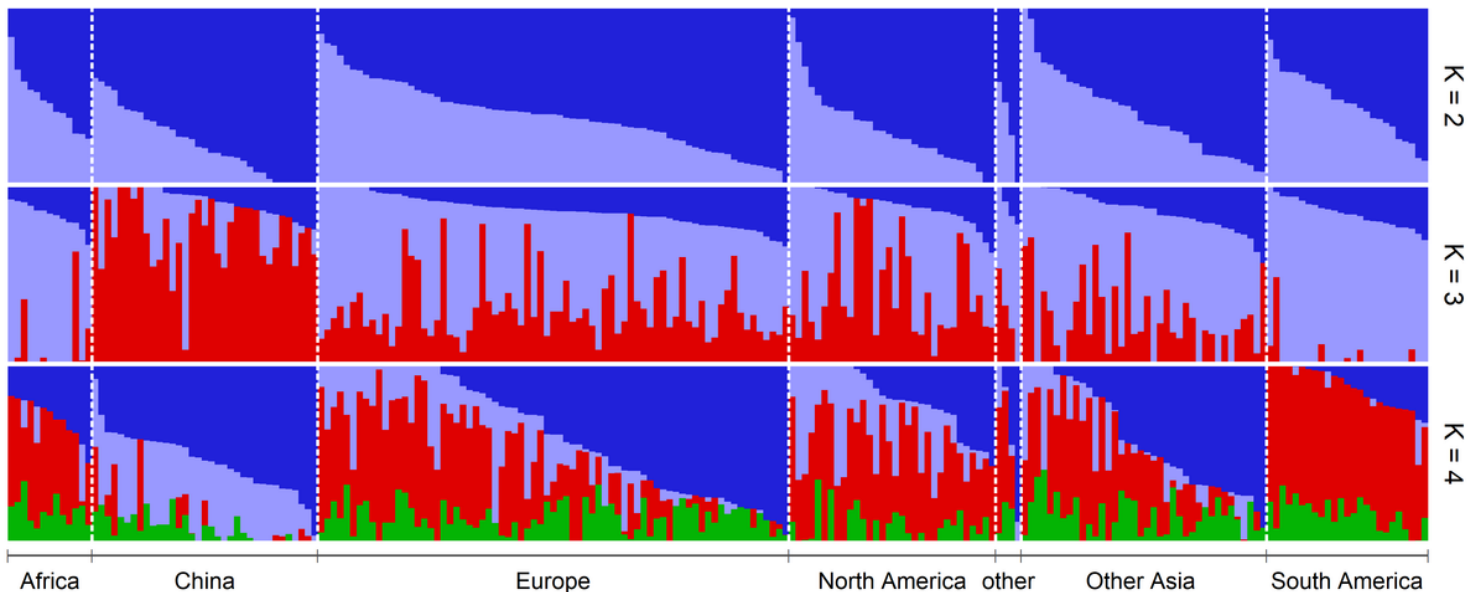


Figure 4

Population structure results from K=2 to K=4. K values were labeled on the right side of the figure. The individual order of subgroups was ordered by K=2. The geographical origins were shown at the bottom of the figure. Different origins were separated by a white dish line.

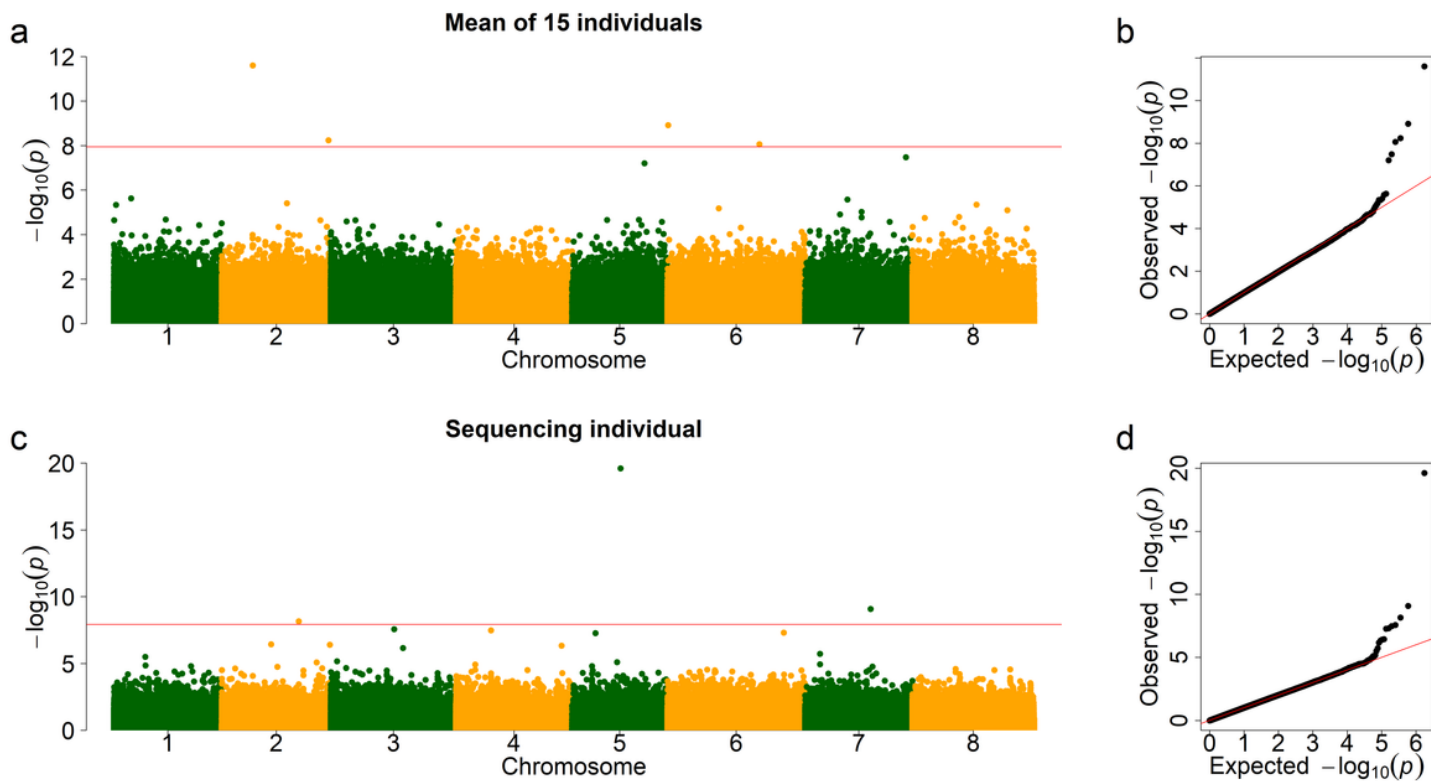


Figure 5

Manhattan and Quantile-Quantile (Q-Q) plot of FD. The genome-wide association study was performed by BLINK C version software, with a significant p-value threshold set at $P=0.01/875,023= 1.14 \times 10^{-8}$ (red line).

We identified four significantly associated SNPs for FD of mean of 15 individuals, which are shown in the Manhattan plot (a). Q-Q plots are displayed on the right panel (b). There are three associated SNP for FD of sequencing individual (c and d).

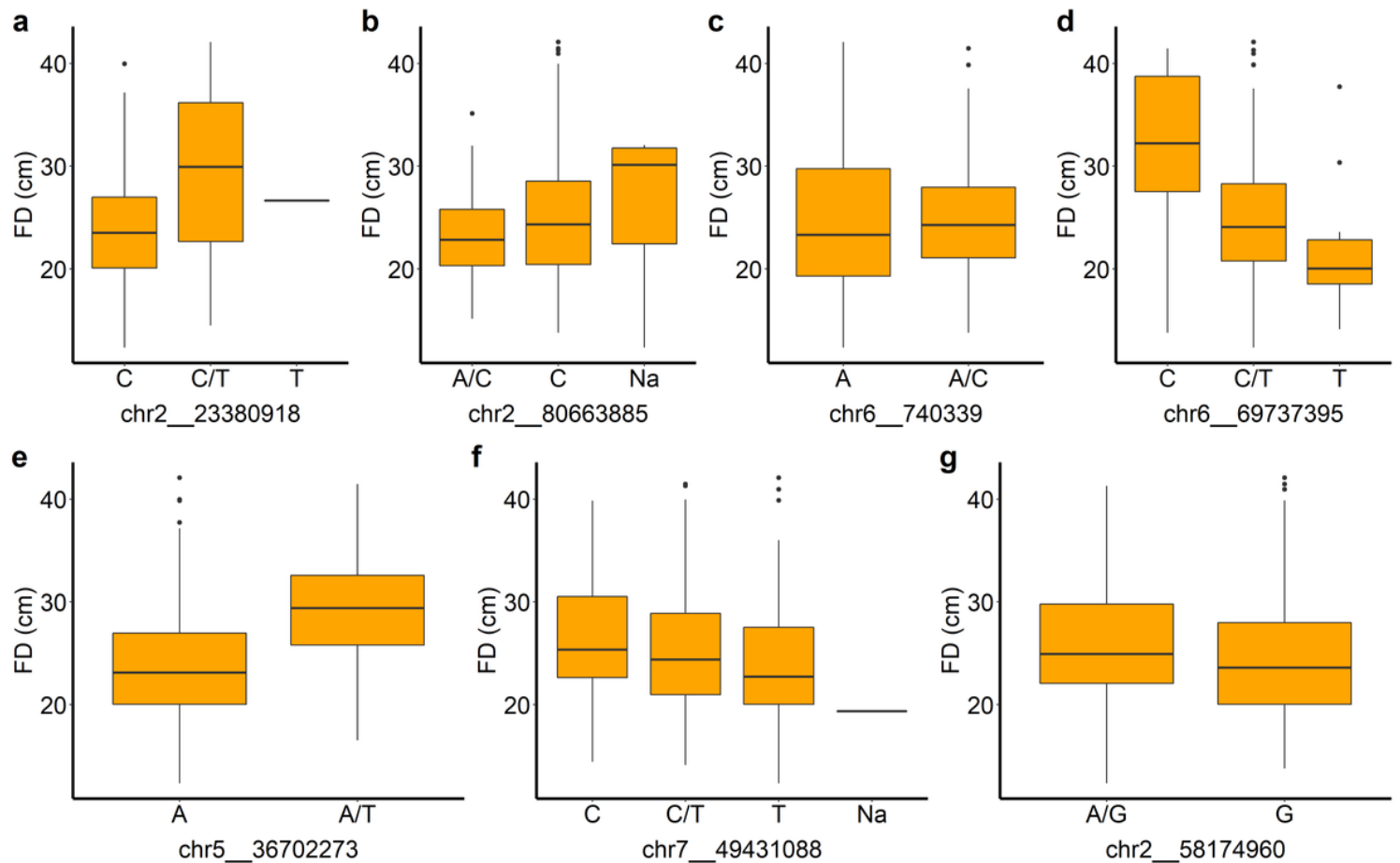


Figure 6

Distribution of FD among the genotypes of the associated SNPs. The SNP associated with FD of mean of 15 individuals were shown in upper half figure (a, b, c and d). The SNP associated with FD of sequencing individual were shown in lower half figure (e, f and g). There is only one individual with TT genotype for SNP chr2_23380918 and no individual with C genotype of CC on SNP chr6_740339. There is no homozygous genotype AA for SNP chr5_36702273 and no GG for SNP chr2_58174960.

Supplementary Files

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

- [supplementaryfileupload.pdf](#)
- [TableS1.xlsx](#)
- [TableS3.xlsx](#)