



Genome-Wide Association Studies Identifying Multiple Loci Associated With Alfalfa Forage Quality

Sen Lin¹, Cesar Augusto Medina¹, O. Steven Norberg², David Combs³, Guojie Wang⁴, Glenn Shewmaker⁵, Steve Fransen⁶, Don Llewellyn⁷ and Long-Xi Yu^{1*}

¹ Plant Germplasm Introduction Testing and Research, Agricultural Research Service, United States Department of Agriculture, Prosser, WA, United States, ² Franklin County Extension Office, Washington State University, Pasco, WA, United States, ³ Department of Dairy Science, University of Wisconsin, Madison, WI, United States, ⁴ Eastern Oregon Agricultural and Natural Resource Program, Oregon State University, La Grande, OR, United States, ⁵ Kimberly R&E Center, University of Idaho, Kimberly, ID, United States, ⁶ Irrigated Agriculture Research and Extension Center, Washington State University, Prosser, WA, United States, ⁷ Department of Animal Sciences, Washington State University, Pullman, WA, United States

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Bernadette Julier,
INRAE, France
Quanzhen Wang,
Northwest A&F University, China

*Correspondence:

Long-Xi Yu
longxi.yu@usda.gov

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 31 December 2020

Accepted: 30 April 2021

Published: 18 June 2021

Citation:

Lin S, Medina CA, Norberg OS, Combs D, Wang G, Shewmaker G, Fransen S, Llewellyn D and Yu L-X (2021) Genome-Wide Association Studies Identifying Multiple Loci Associated With Alfalfa Forage Quality. *Front. Plant Sci.* 12:648192. doi: 10.3389/fpls.2021.648192

Autotetraploid alfalfa is a major hay crop planted all over the world due to its adaptation in different environments and high quality for animal feed. However, the genetic basis of alfalfa quality is not fully understood. In this study, a diverse panel of 200 alfalfa accessions were planted in field trials using augmented experimental design at three locations in 2018 and 2019. Thirty-four quality traits were evaluated by Near Infrared Reflectance Spectroscopy (NIRS). The plants were genotyped using a genotyping by sequencing (GBS) approach and over 46,000 single nucleotide polymorphisms (SNPs) were obtained after variant calling and filtering. Genome-wide association studies (GWAS) identified 28 SNP markers associated with 16 quality traits. Among them, most of the markers were associated with fiber digestibility and protein content. Phenotypic variations were analyzed from three locations and different sets of markers were identified by GWAS when using phenotypic data from different locations, indicating that alfalfa quality traits were also affected by environmental factors. Among different sets of markers identified by location, two markers were associated with nine traits of fiber digestibility. One marker associated with lignin content was identified consistently in multiple environments. Putative candidate genes underlying fiber-related loci were identified and they are involved in the lignin and cell wall biosynthesis. The DNA markers and associated genes identified in this study will be useful for the genetic improvement of forage quality in alfalfa after the validation of the markers.

Keywords: alfalfa, GWAS, markers, forage quality, GBS

Abbreviations: ADF, acid detergent fiber; aNDF, neutral detergent fiber analyzed with amylase; aNDFom, amylase neutral detergent fiber organic matter; ADICP, acid detergent insoluble crude protein; ADICP_pct, ADICP as a percent of crude protein fraction; CP, crude protein; M6Std_Mperton_48, milk per ton estimated from the “Milk 2006” model using 48-h NDFD (standard method); M6Trad_Mperton_48, milk per ton estimated from the “Milk 2006” model using 48-h NDFD (traditional method); NDFdStd_24, NDF digestibility standard in 24 h; NDFdStd_30, NDF digestibility standard in 30 h; NDFdTrad_30, NDF digestibility traditional in 30 h; NDFdTrad_48, NDF digestibility traditional in 48 h; NDFdTrad_120, NDF digestibility traditional in 120 h; NDFdTrad_240, NDF digestibility traditional in 240 h; NDFind_DKC, indigestible NDF (David K. Combs’ protocol); NDFkd_DKC, NDF digestion rate (David K. Combs’ protocol); NDFStd_48, NDF digestibility standard in 48 h; NDICP, neutral detergent insoluble crude protein; NFC, non-fibrous carbohydrates; RFQ, relative forage quality index; RFV, relative feed value index; NRC01_Lignin_NEL, Dairy National Research Council 2001 equation for net energy for lactation; ProtoSol, soluble protein; Sugar_ESC, sugar of ethanol soluble carbohydrates; Sugar_WSC, sugar of water soluble carbohydrates; TTNDFD, total tract NDF digestibility; TTNDFD_DKC, total tract NDF digestibility (David K. Combs’ protocol); uNDF30, 30-h undigested neutral detergent fiber; uNDF120, 120-h undigested neutral detergent fiber; uNDF240, 240-h undigested neutral detergent fiber.

INTRODUCTION

Alfalfa is so called “The Queen of Forages” due to its high value of nutrients beneficial to livestock performance (Hrbáčková et al., 2020). The quality of alfalfa can directly affect daily animal response and gains (Popp et al., 2000). However, alfalfa quality varied remarkably by genotype, and the genetic basis that influences the quality traits is still poorly understood. The autotetraploid alfalfa and its outcrossing characteristics make it more challenging for genetic gain in the economical traits. Fiber content and digestibility are both major factors that affect alfalfa quality. Acid detergent fiber (ADF) and neutral detergent fiber (NDF) are two important parameters on the estimation of fiber quality in forage. The difference between ADF and NDF is that ADF is part of NDF without the hemicellulose, which can be removed by boiling the hay in a detergent solution in acid conditions (Bonsi et al., 1995). Thus, ADF only contains cellulose, lignin and cutin. Lower ADF value is associated with higher quality of alfalfa (Riday et al., 2002). Since NDF takes hemicellulose into account, which is partially digestible by livestock, it presents the forage amount that can be utilized by livestock, and predicts the energy value of alfalfa. The amount of NDF digested before passing from rumen can be estimated by NDF digestibility (NDFD) in 24–48 h. NDFD in longer times (120–240 h) is used to measure the slow and indigestible components. Recently, an *in vitro* assay was developed for the prediction of total tract NDF digestion (TTNDFD), which is a direct predictor of the performances of cattle when fed with forages that differ in fiber digestibility (Lopes et al., 2015). Other broadly used indexes, such as relative feed value (RFV) and relative forage quality (RFQ) are also related to the fiber quantity and digestibility in alfalfa (Undersander et al., 2002). Both RFV and RFQ have been used for ranking forage quality based on the intake potential and digestible matter.

Lignin is an anti-quality component and its content directly affects forage quality. Genetic improvement has been made in alfalfa to reduce the lignin content. Manipulation of the genes involved in lignin biosynthesis reduced lignin content and improved the digestibility of alfalfa for ruminant animals (Jung et al., 2012). Syringyl (S)-lignin, p-hydroxyphenyl (H)-lignin and guaiacyl (G)-lignin are three basic units of lignin in plants. Of those, G-lignin is the primary component in alfalfa. It was reported that knockdown of caffeic acid 3-O-methyltransferase (COMT) reduced G-lignin and total lignin contents (Guo et al., 2001). Down-regulation of P450 enzymes cinnamate 4-hydroxylase (C4H), coumarate 3-hydroxylase (C3H) or ferulate 5-hydroxylase (F5H) lowered lignin content in the transgenic lines of alfalfa. Repression of C3H provided a better digestibility of alfalfa compared to C4H or F5H (Reddy et al., 2005). Cellulose and hemicellulose are two major components in the plant cell wall. Unlike lignin, cellulose and hemicellulose are partially digestible in rumen. Increasing the amount of either cellulose or hemicellulose is favorable for improvement of alfalfa digestibility (Jung et al., 2012; Adesogan et al., 2019). Several genes, such as glycosyltransferase and beta-glucosidase, have been reported to play important roles on cellulose and hemicellulose biosynthesis

and metabolism (Perrin et al., 1999; Scheller and Ulvskov, 2010; Singhania et al., 2013).

Protein content is one of important characteristics affecting the quality of alfalfa. Crude protein (CP) is used for measuring protein in the forage and its content ranges from 18 to 25%. Increasing the intake of protein can significantly improve the cattle growth rate, immunity, and milk production. Acid detergent insoluble crude protein (ADICP) and neutral detergent insoluble crude protein (NDICP) are used to estimate rumen protein availability.

Genetic improvement of forage quality requires understanding the genetic base at the whole genome level. An efficient technique for high-throughput identification of markers/genes responsive to quality traits would be beneficial for genetic improvement. Recently, Next Generation Sequencing (NGS) technology has been used for developing high density SNPs at the genome-wide scale (Rocher et al., 2015). Genome-wide association studies (GWAS) provide advanced tools to identify genetic loci associated with traits of interest using the high-density markers throughout the whole genome (Scheben et al., 2017). It has been successfully applied in the identification of DNA markers associated with agronomic traits in alfalfa and its close relative *M. truncatula* (Alqudah et al., 2020). High-density linkage maps were constructed using GBS markers in both *M. sativa* and *M. truncatula* (Li et al., 2014; Yu et al., 2020). GWAS have been used for identification of significant markers associated with biomass and cell wall biosynthesis in *M. truncatula* (Sakiroglu and Brummer, 2017), and with yield, quality, biotic, and abiotic stress resistance in alfalfa (Biazzi et al., 2017; Hawkins and Yu, 2018; Lin et al., 2020; Medina et al., 2020). However, how genetic factors influence alfalfa quality is still unknown because of the complexity of the alfalfa genome and the quantitative trait controlled by multiple genes.

In this study, a total of 200 alfalfa accessions were planted in three locations for 2 years. Thirty-four quality traits were measured using NIRS. The best linear unbiased estimate (BLUE) was used for analyzing phenotypic variations of each trait. The genotyping was done in the same panel of germplasm by GBS and over 46,000 SNPs were obtained. Marker and trait data were used to identify marker loci associated with quality traits using GWASPoly with multiple models. Our goal is to determine genetic factors that influence forage quality in alfalfa. The genetic information gained from this study would help in genetic improvement of alfalfa with higher quality.

MATERIALS AND METHODS

Plant Materials

A diverse panel of germplasms composed of 148 accessions selected from USDA-ARS National Plant Germplasm System, and 52 cultivars from S&W Seed Co., Alforex SeedsTM, Legacy Seeds and Blue River Hybrids were used in this experiment. Hi-Gest360 and 22338_VernalFD2 were used as higher or lower quality checks, respectively. Most of the germplasms were originally collected from Northwest and central regions in US and Canada, including Idaho, Montana, Nebraska, Washington, and

North Dakota, South Dakota, British Columbia, Saskatchewan, and Manitoba. Other germplasms were collected from different countries in Asia, Europe, and Africa, including Afghanistan, China, India, Lebanon, Oman, Yemen, Bulgaria, Germany, Russia, Spain, Turkey, and Algeria. All the accessions were listed in **Supplementary Table 1**.

Field Experiment

The panel of accessions were planted in three locations in Prosser, Washington (WA), Kimberly, Idaho (ID), and Union, Oregon (OR). Each location contained 11 blocks and each block contains 20 plots with 4.6 m long and 1.2 m wide. An augmented randomized complete block design (ARCB) was applied in field trial (Federer, 1961; Federer and Raghavarao, 1975). The field at each location was separated to 11 blocks. Twenty accessions including two checks, Hi-Gest® 360 (high quality check) and Vernal (low quality check) were planted in each block. We have one replicate for all accessions but eleven replicates for checks in each location. The phenotypic data from the two checks were used as covariates to correct the phenotypic variation across three locations for 2 years. Field plots were irrigated regularly until harvesting. Samples for quality measurement were collected at mid-bud stage from the first cut in each location. The seeding date in OR, ID, and WA were May 4, May 10, and April 20, respectively, in 2018. The first cut date in OR, ID, and WA were Jul 11, Jul 10, and Jun 29 in 2018, and Jun 4, Jun 5, May 14 in 2019, respectively. The average temperature between seedling date and first cut in 2018 were 15.6, 17.3, and 16.8°C at locations in OR, ID, and WA, respectively. Soil pH varied from 6.3 to 8.2, cation exchange capacity varied from 18.7 to 40.7 across the three locations. Fertilizer nutrients were applied in each location based on the soil parameters.

Phenotyping and Data Analysis

Shoot samples were collected from first cuts in each location and used for analysis of the 34 traits related to forage quality. The samples collected for quality test were from the same plants for DNA extraction. All samples were dried at 57°C, and then transferred to Hammer mill (Meadow Mills, Wilkesboro, NC 28659) for first grinding and then to Wiley Mill (Thomas Scientific, United States) for second grinding to 2 mm, followed by a final grinding in the sampling mill (Udy Cyclone mill, Fort Collins, CO 80524) through a 1 mm screen. Sample powders were analyzed for 34 quality factors including CP, indigestible CP (ADICP) soluble protein (Protsol), fiber (ADF and NDF), fiber digestibility (NDFD) undigested NDF (uNDF), non-fiber carbohydrates (sugar, NFC) and for several indexes of forage fiber quality (TTNDFD, RFV, and RFQ) and other factors by NIRS using a scanning monochromator (FOSSNIR Systems 5000, Silver Spring, MD, United States) by Rock River Laboratory (Watertown, WI 53094). The reference calibrations for the above components used proprietary equations developed from more than 100,000 alfalfa samples. Global H statistics of each sample were submitted and the average Global H of the sample set was monitored to assure that the spectra of samples submitted were consistent with the calibration samples. Global H is a leverage value scaled so that the average value in the calibration set is 1.0.

It measures the squared distance of a spectrum to the average spectrum (Marten et al., 1984; Karn et al., 1991; Shenk and Westerhaus, 1991).

Best linear unbiased estimates (BLUEs) were calculated with the lme4 package in R for analyzing phenotypic data. Genotype was assumed as a fixed effect, and environmental factors such as location, year, and block were treated as a random effect. When analyzing phenotypic data by location, environmental factors included year and block. Analysis of correlation amongst 34 quality traits was performed by Pearson coefficients using BLUEs.

DNA Samples Preparation and Sequencing

Leaves of 24 representative plants from each accession were collected as a pool for extraction of DNA using Qiagen DNeasy 96 Plant Kit (Qiagen, CA) according to the manufacturer's protocol. DNA samples were tested by measuring the concentration and absorbance ratio of 260/280 and 260/230 with Nanodrop 1000 spectrometer (Thermo Fisher Scientific). DNA samples were sent to the sequencing facility at Oregon State University for library preparation following digestion by restriction enzyme *ApeKI*. Illumina HiSeq2000 system was used for sequencing DNA libraries. To increase coverage, each library was sequenced using two channels with double strands. On average, the sequencing dataset contained over 6 million reads of 136 bp length per sample. The *M. truncatula* genome sequence (Mt A17 r5.0) was used as a reference genome for alignment of the reads using Bowtie2 with highly sensitive parameters (Langmead and Salzberg, 2012). Variant calling was performed using NGSEP v4.0 with the following parameters: maximum base quality score 30, minimum genotyping quality 40, ploidy = 4 (Duitama et al., 2014). The SNPs with minor allele frequency (MAF) lower than 0.05 or missing value higher than 20% were removed. A total of 46,792 (8.1% missing calls) SNPs were obtained after filtering. Missing values in the variants were imputed via hidden Markov model in NGSEP V4.0.

Population Structure Analysis

The genotypic data was used for principal component analysis (PCA). First, the vcf file containing 46,792 SNPs was transformed to bed format using the plink software. Then eigenvec file containing the data of PC1, PC2, and PC3 was generated using GCTA (Yang et al., 2011). A phylogenetic tree was generated using the neighbor joining method and plotted by iTOL (Ciccarelli et al., 2006).

Genome-Wide Association Analysis and Identification of Candidate Genes

The R package of GWASpoly, which takes the allele dosage into account, was used for GWAS (Rosyara et al., 2016). Six different marker-effect models including additive, general, diplo-general, diplo-additive, 1-dom and 2-dom were applied for analysis of marker-trait association. Both Q and K matrix were included in the analysis. Q matrix was obtained from PCA. K matrix was determined by default in GWASpoly after inputting the phenotype and SNP data. False discovery rate (FDR) of 0.05 was

used as threshold for significance selection. Candidate genes were localized by retrieving the positions of significant markers in *M. truncatula* genome database¹.

Gene Annotation

Gene annotation was based on the SNP position in *M. truncatula* genome dataset V5.0. National Center for Biotechnology Information² and UniProt database³ were also used for annotation by extracting a 2-kb flanking region of each significant marker and searching the homologs by BLAST.

RESULTS

Development of SNP Markers Using Genotyping by Sequencing

GBS generated more than 1,146 million reads with the average length of 136 bases. More than 53.9% reads were successfully aligned to the *M. truncatula* reference genome. Minor allele frequency (MAF) was distributed between 0.05 and 0.5 (Figure 1A). The proportion of SNPs with MAF between 0.05–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, and 0.4–0.5 were 24.7, 20.3, 21.0, 18.0, and 16.1%, respectively. After variant calling and filtering, 46,792 meaningful SNPs were obtained. Among those, 4,796 were intergenic variants and up to 63.8% SNPs were present in the coding regions with 13,040 synonymous variants and 9,796 missense variants (Figure 1B). Six hundred and twenty-three and 1,955 variants were identified at the 5' and 3' untranslated regions (UTR), respectively. In addition, there were 923 variants located at the upstream of the 5' UTR, while 419 variants were in the downstream region. The distribution and density of the SNPs on 8 chromosomes were illustrated in Figure 1C. Chromosome 3 had the most with 7,258 SNPs. The highest density of SNPs was observed on chromosome 1 with 122.98 SNPs in every mega-base pair (Table 1).

Statistical Analysis of Quality Traits

Statistical analyses were performed among all 34 quality traits. Frequency distributions of 4 quality traits (NDFdTrad_120, TTNDFD, ProtoSol, ADICP_pct) related to fiber digestibility and protein content were shown in Figure 2. The phenotypic data of all the traits were normally distributed. NDFdTrad_120 varied from 57.97 to 69% (Figure 2A). The minimum and maximum values of TTNDFD were 40.23 and 50.97%, respectively (Figure 2B). ProtoSol (soluble protein) was ranging from 32.43 to 38.71%, with the mean value of 35.78% (Figure 2C). ADICP_pct varied from 2.12 to 2.89%, with the mean value of 2.51% (Figure 2D). Other traits were also well distributed as shown in histograms (Supplementary Figure 2). Broad sense heritability (H^2) was calculated for all 34 traits (Figure 2E). The H^2 of 26 traits were greater than 0.5. Among them, the H^2 values of 6 traits (NDFStd_48, TTNDFD, uNDF240, uNDF120, TTNDFD_DKC, and NDFdStd_30) were higher than

0.7, indicating most of phenotypic variations of these traits were genetically controlled. Lower heritability was found in Sugar_WSC and NDICP with H^2 less than 0.3, with the lowest heritability of 0.1 in NDICP. The correlations between 34 traits were analyzed based on Pearson coefficients and 3 groups were classified (Figure 2F). Among the highly positive correlation groups, TTNDFD is highly correlated to NDFStd_48, with the coefficient value of 0.99. The coefficient between aNDF and aNDFom was 0.98, followed by TTNDFD_DKC and TTNDFD, NDFdTrad_48, and NDFdTrad_120, with the coefficients greater than 0.97. High correlations were also detected between RFQ and M6Std_Mperton_48, NDFdStd_24 and NDFdStd_30, M6Trad_Mperton_48 and M6Std_Mperton_48, NDFStd_48 and TTNDFD_DKC, uNDF30 and uNDF120, and uNDF240 and uNDF30, with the coefficients higher than 0.95. In addition, a highly negative correlation (−0.99) was observed between NDFind_DKC and NDFdTrad_240. RFV was negatively correlated with ADF, aNDF, and aNDFom, with coefficients lower than −0.96. RFQ was also negatively correlated with uNDF120 and uNDF30, with the coefficients of −0.96 for both. Starch, NFC, Sugar_ESC, Sugar_WSC, ADICP, NDICP, Ash, and NDFkd_DKC were classified into the intermediate group according to their correlations. Moderately positive correlation was observed between ADICP and NDFkd_DKC with the coefficient 0.25, whereas Sugar_ESC and NDFkd_DKC were in moderately negative correlation with the coefficient of −0.35.

Population Structure Analysis

Genotypic data were used for cluster analysis using neighbor joining method. The alfalfa accessions were grouped into 3 clusters (Figure 3). Among those, 63 accessions were classified to cluster 1, 64 accessions were classified to cluster 2 and 65 accessions were in cluster 3. Cluster 1 contained a mixture of accessions from multiple countries and regions, including United States, Canada, China, Mediterranean, and central and eastern Europe. Cluster 2 contained the accessions from Turkey and central Asia. Cluster 3 contained accessions from United States except one from Canada. The result was consistent with the PCA where similar groups were observed (Supplementary Figure 1).

Genome-Wide Association Analysis

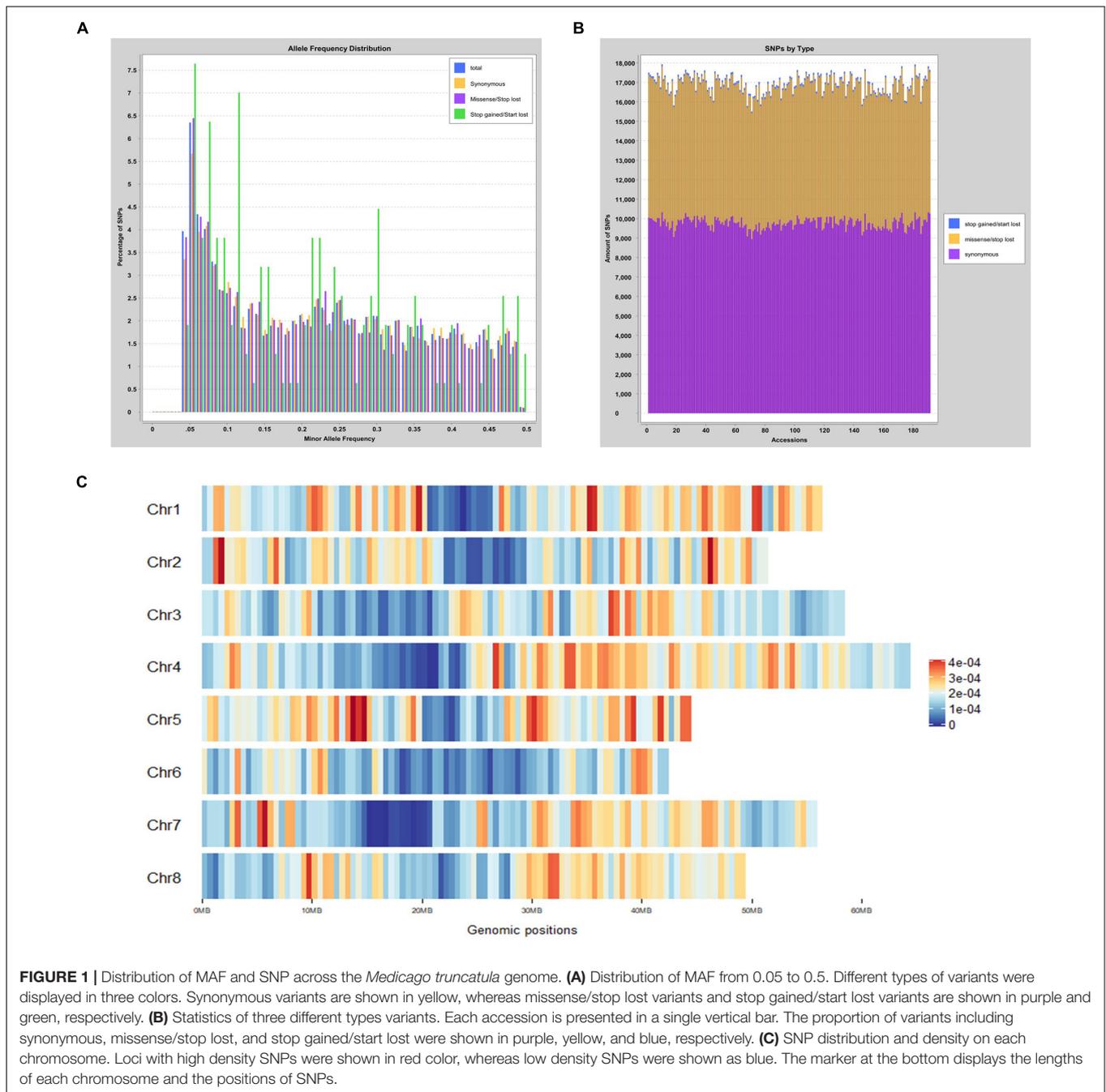
GWASPoly was used for GWAS with allelic dosage using different models, and the FDR of 0.05 was used as the threshold for the selection of significant associations. In total, 28 significant markers were obtained. Several markers were identified in more than one model. Generally, the traits associated with these 28 markers can be categorized into two groups.

The first group contained markers associated with traits related to fiber digestibility, including Lignin, uNDF30, uNDF120, TTNDFD, TTNDFD_DKC, NDFdTrad_30, NDFdTrad_48, NDFdTrad_120, NDFdStd_30, NDFStd_48, and RFQ. Two markers, MtChr4_33424222, and MtChr4_33424238 (at the same locus) on chromosome 4 were associated with 9 traits related to fiber digestibility, including NDFdTrad_120, NDFdTrad_30, NDFdTrad_48, TTNDFD, TTNDFD_DKC, NDFStd_48, RFQ, and uNDF30 as well as uNDF120

¹<https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/>

²<https://www.ncbi.nlm.nih.gov/>

³<https://ebi12.uniprot.org/>



(Figures 4A–I). Additionally, three markers including two on chromosome 2 (MtChr2_45854840, MtChr2_46050634) and one on chromosome 8 (MtChr8_47585557) were associated with NDFdTrad_30 (Supplementary Figure 3). Markers associated with Lignin content were identified on chromosome 1 (MtChr1_43382331) and chromosome 5 (MtChr5_13533783) (Figure 4J). Moreover, the marker MtChr1_43382331 was also associated with uNDF30 and uNDF120 (Supplementary Figure 3). Marker MtChr4_30582636 identified on chromosome 4 was associated with TTNDFD, TTNDFD_DKC, and RFQ (Figures 4D,E,G). Another TTNDFD-associated marker

MtChr4_32514187 was identified on chromosome 4 (Figure 4D). MtChr1_48744767 and MtChr6_522150 were associated with TTNDFD_DKC (Supplementary Figure 3). Two markers MtChr5_1952458 and MtChr8_30029709, associated with NDFkd_DKC, were identified on chromosomes 5 and 8, respectively (Figure 4K). Marker MtChr6_1927418 associated with NDFdStd_30 with $-\log p$ -value 6.11, was detected on chromosome 6 (Figure 4L and Supplementary Figure 3).

The second group contained markers associated with traits related to protein content, including ProtoSol, ADICP, ADICP_pct, and NDICP. Both ADICP and NDICP are

measurements of insolubility of crude protein. A significant marker on chromosomes 8 (MtChr8_41724151) was associated with ProtoSol (Figure 4M). MtChr3_20251386 and MtChr3_20251391 on chromosome 3 with MtChr5_41544991 on chromosome 5 were associated with NDICP (Figure 4N). Seven markers significantly associated with ADICP_pct were on chromosome 5 and detected in different models (Figure 4O and Supplementary Figure 3). Besides the 7 clustered SNPs, another two markers MtChr1_16144476 and MtChr1_36174006 on chromosome 1 were also associated with ADICP_pct. An ADICP associated marker MtChr2_6808038 was detected on chromosome 2 (Figure 4P).

Annotation of Functional Genes Associated With Quality Traits

The flanking sequences of significant markers were used for searching putative candidate genes using the BLAST search as described above. Of the 28 significant markers identified, 27 were encompassed by the annotated genes with different functions (Table 2). ProtoSol-associated marker MtChr8_41724151 was in a gene region of methyltransferase. Markers (MtChr1_16144476 and MtChr1_36174006) associated with ADICP_pct on chromosome 1 were adjacent to genes encoding Casein kinase 1 (CK1) family protein kinase and broad complex, tram track, bric-a-brac/POX virus and zinc finger (BTB/POZ) protein, respectively. Marker MtChr2_6808038 associated with ADICP was closely located to the gene encoding a tubby-like protein. Two of the NDFdTrad_30-associated markers on chromosomes 2 and 8 were closely located at the genes encoding thioredoxin-like protein and E3 ubiquitin-protein ligase, respectively. The TTNDFD-associated marker MtChr4_32514187 was located at the coding region of sulfurtransferase gene. Two markers, MtChr3_20251386 and MtChr3_20251391, were located at the same locus and they were associated with NDICP. They were encompassed by the gene encoding 1,3-beta-glucan synthase. Another NDICP associated gene surrounding MtChr5_41544991 was annotated as U3 small nucleolar RNA-associated protein. MtChr4_30582636 was in the

coding region of a gene homologous to 2',3'-cyclic-nucleotide 3'-phosphodiesterase. Three markers MtChr5_1952458, MtChr8_30029709, and MtChr4_43382331, associated with fiber related traits, were annotated as cycloartenol synthase, 1-phosphatidylinositol 4-kinase, and sodium/calcium exchanger membrane protein, respectively.

Additionally, MtChr4_33424222 and MtChr4_33424238, associated with multiple fiber related traits, were located near a hypothetical protein with unknown function. Similarly, markers MtChr5_13533783 and MtChr2_45854840 associated with Lignin and NDFdTrad_30, respectively, were also located at the gene regions of hypothetical proteins, as well as a cluster of markers associated with ADICP_pct on chromosome 5.

Comparison of Markers Identified in Different Environments

To investigate markers associated with alfalfa quality traits under different environments, we used the means of 2-year phenotypic data collected from different locations and analyzed separately by location with the same genotypic data for GWAS. The average value of CP content in WA was lower than those in ID and OR, while ProtoSol was the highest in OR among the 3 locations. For digestibility-related traits such as NDFdTrad_120 and TTNDFD, average values were higher in ID than those in OR and WA. However, WA had the highest values of Lignin and uNDF traits. The phenotypic variations of accessions of 34 quality traits by location are presented in Supplementary Figure 4. One hundred and fifty-six markers were associated with 24 quality traits (Figure 5). 47, 64, and 46 markers were identified in WA, OR, and ID, respectively (Figure 5 and Supplementary Tables 2–4). Compared with those identified using adjusted means from three locations, two, two, and four markers from WA, ID, and OR, respectively, overlapped with those identified using adjusted means. One marker (MtChr1_43382331) associated with Lignin was identified in WA, OR, and the adjusted means. This marker was associated with TTNDFD and NDFStd_48 that negatively correlated with Lignin. The rest of markers were not overlapping with those identified using adjusted means, indicating alfalfa quality was also strongly affected by environments. Markers consistently expressed in different locations will be useful in MAS after validation.

TABLE 1 | SNP distribution and frequency across the genome of *M. truncatula*.

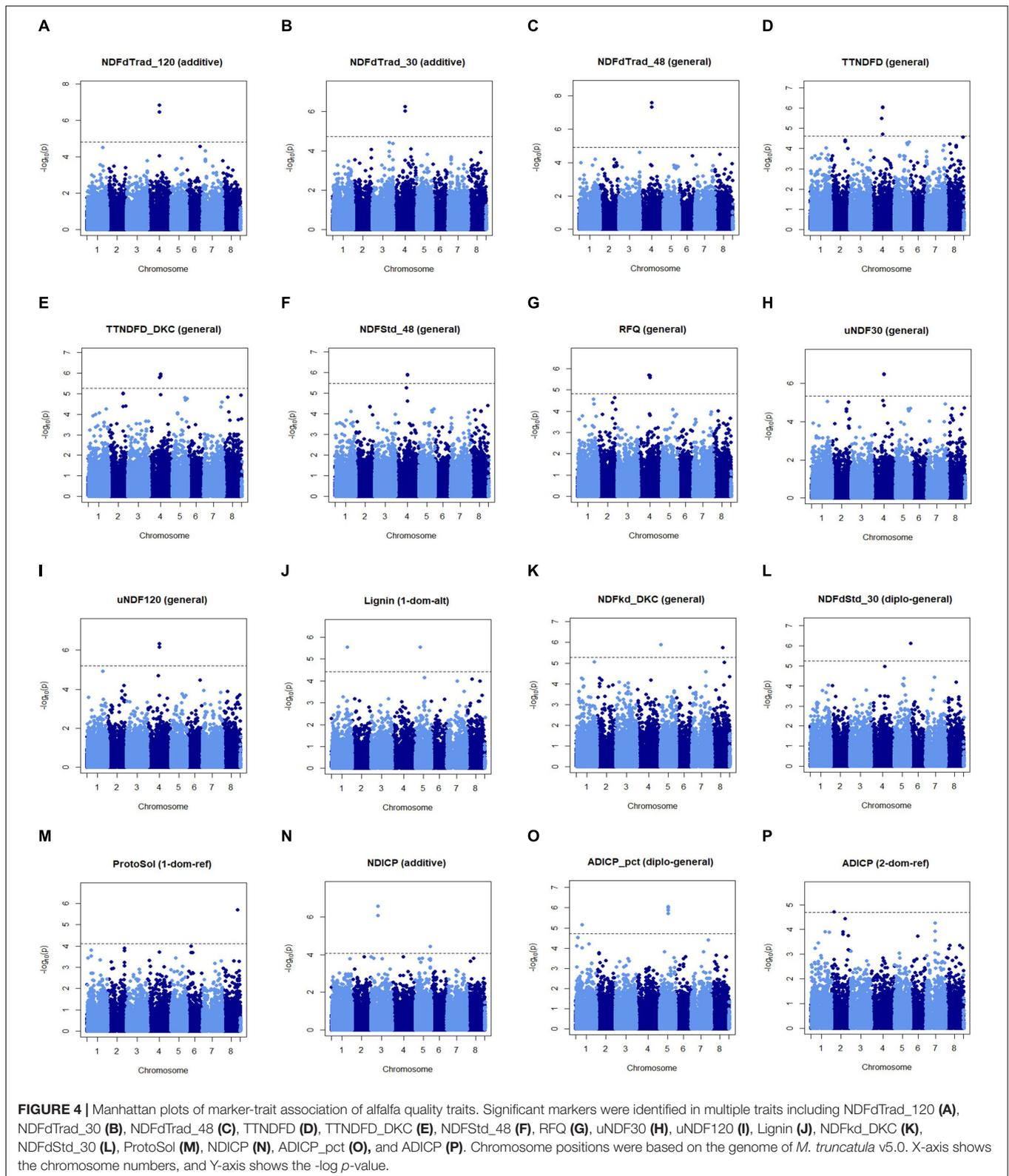
Chromosome	Number of SNPs	Lengths (bp)	SNPs/Mb
1	6,974	56,706,830	122.98
2	5,741	51,972,579	110.46
3	6,037	58,931,556	102.44
4	7,258	64,763,011	112.07
5	5,760	44,819,618	128.52
6	3,549	42,866,092	82.79
7	5,745	56,236,587	102.16
8	5,627	49,719,271	113.18
Oc01	18	1,205,179	14.94
Oc05	17	150,389	113.04
Oc09	5	63,286	79.01
MT	61	271,618	224.58

Oc01, Oc05, and Oc09 are scaffolds; MT, mitochondria.

DISCUSSION

Genotyping Call and GWAS With Allelic Dosage

In this study, we used GBS for genotyping the association panel of 200 alfalfa varieties and obtained over 46,000 SNPs. These high-density markers were subsequently used for association mapping of 34 alfalfa quality traits in the panel of association. Although several gaps were observed on chromosomes, most of these regions were potentially close to centromere that are considered to contain redundant repeats and methylated sequences with lower gene density (Wong et al., 2006; Mizuno et al., 2011).



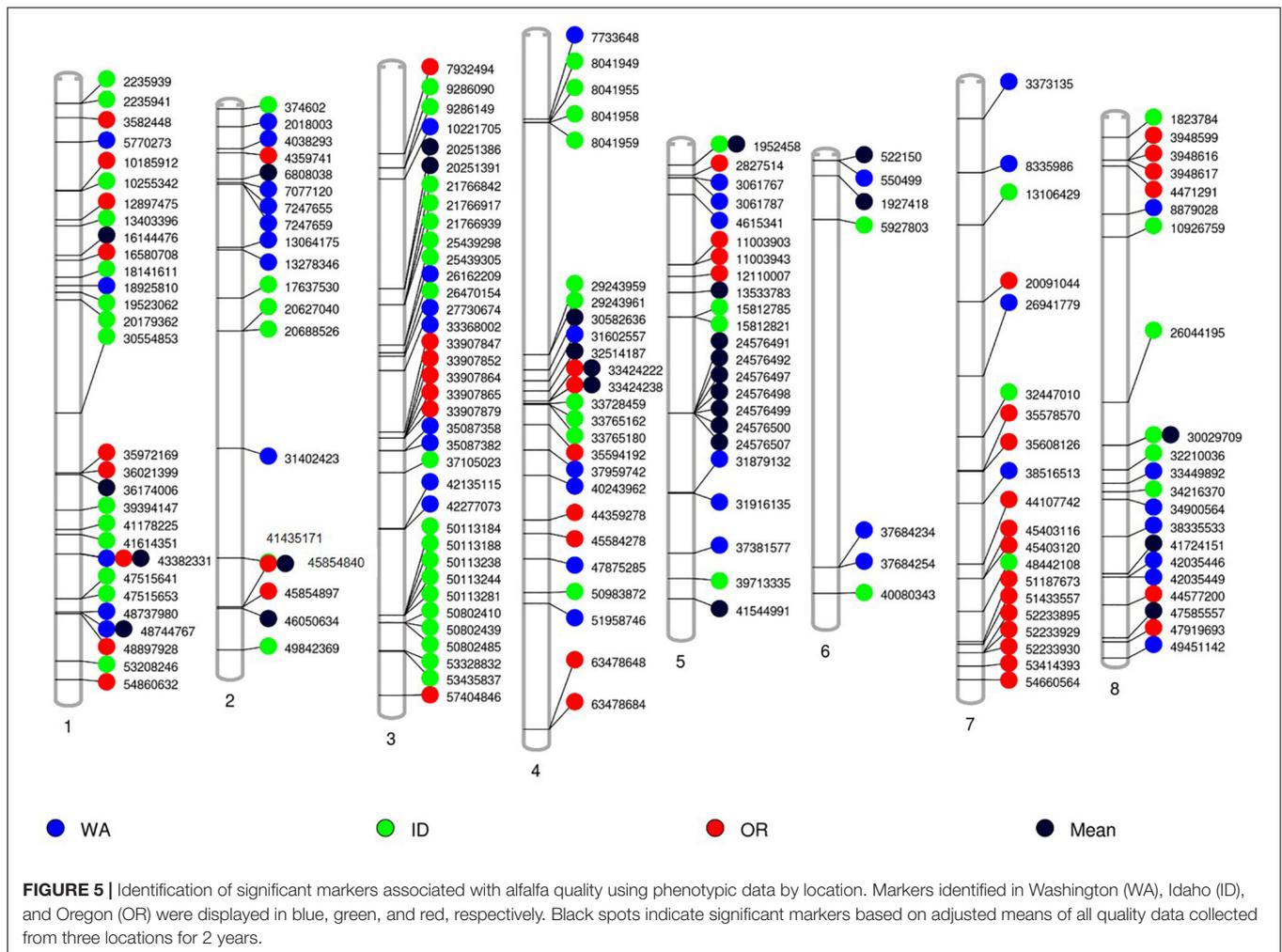
this study almost covered all the essential quality-related characteristics. Generally, high quality forage is rich in protein and can supply high amount of nutrients to the

livestock (Hoveland and Monson, 1980). However, crude protein content and the digestion rate decline during the maturity of alfalfa. Analysis of the correlation between

protein content and digestibility would provide a better clue on the quality changes during the growth period of alfalfa. In the present study, the result of the correlation analysis found a total of 97 pairs of traits that were strongly positively correlated with the coefficients greater than 0.7, whereas other pairs of traits were strongly negatively correlated with the coefficients lower than -0.7 . The positive correlations were mainly observed in two groups

TABLE 2 | Significant markers associated with quality traits identified using *M. truncatula* genome as references.

Marker	Ref	Alt	$-\log p$	R^2	Trait	Annotation
MtChr8_41724151	A	G	5.7	0.2	ProtoSol	Putative tRNA [cytosine(34)-C(5)]-methyltransferase
MtChr2_6808038	A	G	4.71	0.06	ADICP	Putative transcription factor TUBBY family
MtChr1_16144476	A	G	5.15	0.19	ADICP_pct	Putative protein kinase CK1-CK1 family
MtChr1_36174006	A	G	5.38	0.27		Chromatin remodeling and transcription regulator BTB-POZ-MATH family
MtChr5_24576491	C	G	5.72	0.09		Hypothetical protein
MtChr5_24576492	A	T	6.04	0.09		
MtChr5_24576497	C	T	6.01	0.1		
MtChr5_24576498	G	T	6.04	0.09		
MtChr5_24576499	A	T	5.9	0.09		
MtChr5_24576500	A	G	6.04	0.09		
MtChr5_24576507	A	T	6.07	0.1		
MtChr3_20251386	C	T	6.56	0.12	NDICP	1,3-Beta-glucan synthase
MtChr3_20251391	A	G	6.06	0.11	NDICP	
MtChr5_41544991	A	G	4.43	0.08	NDICP	U3 small nucleolar RNA-associated protein
MtChr1_43382331	A	T	5.55	0.16	Lignin	Sodium/calcium exchanger membrane region
	A	T	5.73	0.18	uNDF120	
	A	T	5.82	0.18	uNDF30	
MtChr2_45854840	A	C	5.29	0.01	NDFdTrad_30	Hypothetical protein
MtChr2_46050634	A	G	5.34	0.01	NDFdTrad_30	Putative thioredoxin-like protein
MtChr8_47585557	C	T	5.5	0.01	NDFdTrad_30	Aminoacyltransferase, E1 ubiquitin-activating enzyme
MtChr4_30582636	C	T	5.7	0.09	RFQ	2',3'-cyclic-nucleotide 3'-phosphodiesterase
	C	T	5.49	0.04	TTNDFD	
	C	T	5.79	0.04	TTNDFD_DKC	
MtChr4_32514187	A	C	4.71	0.03	TTNDFD	Sulfurtransferase
MtChr4_33424222	C	T	5.62	0.21	NDFdTrad_120	Hypothetical protein
	C	T	6.25	0.19	NDFdTrad_30	
	C	T	7.59	0.21	NDFdTrad_48	
	C	T	5.92	0.15	NDFStd_48	
	C	T	5.68	0.16	RFQ	
	C	T	6.06	0.15	TTNDFD	
	C	T	5.86	0.14	TTNDFD_DKC	
	C	T	6.33	0.19	uNDF120	
	C	T	6.48	0.17	uNDF30	
MtChr4_33424238	C	T	5.26	0.2	NDFdTrad_120	
	C	T	6.02	0.19	NDFdTrad_30	
	C	T	7.32	0.21	NDFdTrad_48	
	C	T	5.88	0.15	NDFStd_48	
	C	T	5.59	0.16	RFQ	
	C	T	6.03	0.16	TTNDFD	
	C	T	5.94	0.15	TTNDFD_DKC	
	C	T	6.16	0.18	uNDF120	
	C	T	6.47	0.17	uNDF30	
MtChr5_13533783	A	G	5.55	0.11	Lignin	Hypothetical protein
MtChr5_1952458	G	T	5.9	0.01	NDFkd_DKC	Cycloartenol synthase
MtChr8_30029709	A	G	5.75	0.07	NDFkd_DKC	1-Phosphatidylinositol 4-kinase
MtChr6_1927418	A	G	6.11	0.04	NDFdStd_30	Preprotein translocase subunit SecE
MtChr1_48744767	G	T	4.6	0.21	TTNDFD_DKC	Hypothetical protein
MtChr6_522150	C	T	4.56	0.11	TTNDFD_DKC	840 bp away from a hypothetical gene



of traits. The first group included ADF, aNDF, aNDFom, ADICP_pct, NDFind_DKC, Lignin, uNDF30, uNDF120, and uNDF240. The second group included M6Std_Mperton_48, M6Trad_Mperton_48, NRC01_Lignin_NEL, Fat, RFV, RFQ, CP, ProtoSol, NDFdTrad_30, NDFdTrad_48, NDFdTrad_120, NDFdTrad_240, NDFdStd_24, NDFdStd_30, NDFdStd_48, TTNDFFD, and TTNDFFD_DKC. It was not surprising that the traits within the groups were positively correlated as they are in the same categories. Likewise, negative correlations were found between the two groups as they contained opposite traits. Although some traits are highly correlated, the genetic basis behind each trait may be different. Moreover, each of the traits has different characteristics when evaluating alfalfa quality. For example, Lignin, ADF, and NDF are highly correlated as both lignin and cellulose are counted in ADF and hemicellulose is counted in NDF besides lignin and cellulose. NDFD240 represents the indigestible fiber and it is related to gut fill in ruminants (Cotanch, 2015). NDFD30 and NDFD48 are usually used to index how fast fiber degrades (Hoffman and Combs, 2003). On the other hand, the result of correlations among those traits are also helpful to simplify the evaluation process on

alfalfa quality in the future since it tells us which traits are highly correlated.

Markers Associated With Multiple Highly Correlated Traits

In present study, we identified two markers, MtChr4_33424222 and MtChr4_33424238, at the same locus associated with nine traits. All of them were related to fiber digestibility and they were highly correlated with each other. The highest correlation was found between NDFStd_48 and TTNDFFD with the coefficient value of 0.99, whereas the lowest coefficient 0.69 was observed between RFQ and TTNDFFD_DKC among the nine traits. The most negatively correlated pair was uNDF120 and RFQ with coefficient -0.96 . In general, uNDF120 and uNDF30 were negatively correlated with the other seven traits including NDFdTrad_30, NDFdTrad_48, NDFdTrad_120, NDFdStd_48 RFQ, TTNDFFD, and TTNDFFD_DKC, as uNDF120 and uNDF30 reflected the amount of undigested fiber which are negative factors for alfalfa quality. The associations between these two markers and the nine fiber-related traits indicate that the associated loci are possibly involved in the regulation of

TABLE 3 | Potential candidate genes within 50-kb of significant SNPs.

Trait	SNP	Adjacent gene	Annotation	Range
Lignin	MtChr1_43382331	MtrunA17Chr1g0195561	Pectinesterase	43,434,578–43,435,483
TTNDFD_DKC	MtChr1_48744767	MtrunA17Chr1g0202641	Galacturan 1,4-alpha-galacturonidase	48,714,585–48,719,814
NDFdTrad_120	MtChr4_33424222	MtrunA17Chr4g0033401	Beta-glucosidase	33,394,821–33,397,755
TTNDFD	MtChr4_32514187	MtrunA17Chr4g0032331	Hexosyltransferase	32,456,246–32,458,704
Lignin	MtChr5_13533783	MtrunA17Chr5g0412101	7-Deoxyloganic acid glucosyltransferase	13,557,474–13,559,854
NDFdStd_30	MtChr6_1927418	MtrunA17Chr6g0451381	Glycosyltransferase	1,948,594–1,953,305
TTNDFD_DKC	MtChr6_522150	MtrunA17Chr6g0449561; MtrunA17Chr6g0449521	Galactose oxidase; galactinol–sucrose galactosyltransferase	537,765–539,858; 493,534–498,577

multiple pathways including lignin biosynthesis and other cell wall components such as cellulose or hemicellulose. Further investigation of the candidate genes underlying these loci would be helpful to understand the genetic basis of fiber digestibility in alfalfa.

Putative Candidate Genes With Functions Involved in Forage Quality

Among significant markers identified in the present study, fourteen were associated with fiber-related traits; The same significant markers, MtChr4_33424222 and MtChr4_33424238 were detected in nine fiber-related quality traits and they were highly correlated each other. In our previous study, over 100 markers were associated with alfalfa quality under drought conditions (Lin et al., 2020). Our present results showed multiple candidate genes associated with forage quality which had similar functions to genes identified in the previous study. Sakiroglu and Brummer (2017) found that a gene locus associated with ADF and NDF on chromosome 8 was involved in cell wall biosynthesis, arabinose, and xylose content. Biazzi et al. (2017) identified three markers associated with CP, NDF digestibility (NDFD), and acid detergent lignin (ADL) of leaf, together with multiple significant markers in association with CP and ADL of stem. We identified 28 significant markers associated with protein content and fiber-related traits. Alfalfa with low content of lignin is highly desirable as rumen digestibility is negatively correlated with lignin content of alfalfa (Reddy et al., 2005). In the present study, marker MtChr1_43382331 was annotated as a sodium/calcium exchanger. This locus was associated with both uNDF120 and lignin, indicating that the exchange of sodium and calcium is likely to affect the composition of cell wall. The calcium level in plants is in a strong relationship with the amount of pectin in then cell wall (Reid and Smith, 2000). Furthermore, calcium homeostasis and binding on membranes of plant cells is affected by sodium (Rengel, 1992). Marker MtChr5_1952458 is associated with NDFkd_DKC, located in the gene coding region of cycloartenol synthase. It has been reported that cellulose biosynthesis is strikingly influenced by the sterol composition, and the high cycloartenol content often causes a disrupted cell wall structure and reduced cellulose (Schrack et al., 2004). Marker MtChr2_46050634 associated with NDFdTrad_30 was

located at the region of gene encoding thioredoxin-like protein. Thioredoxin is a class of protein involving redox signaling and relief of oxidative stress. It exists in a variety of organisms and is involved in many biological processes (Gelhay et al., 2005). Recent studies showed that thioredoxin-like protein affects lignin deposition, polymerization and cell wall synthesis (Alkhalifoui et al., 2007; Shin et al., 2020). Another marker (MtChr8_30029709) associated with NDFkd_DKC was located at the gene region of 1-phosphatidylinositol 4-kinase (PIK4), a transferase responsible for transferring phosphorus-containing groups. A recent study showed that PIK4 plays an important role in the trafficking of cellulose synthase complexes (Fujimoto et al., 2015). Marker MtChr4_30582636 was in the gene encoding region of 2',3'-cyclic-nucleotide 3'-phosphodiesterase (CNPase), a member of the 2H phosphoesterase superfamily. This type of enzyme has RNA ligase activity and may also be involved in tRNA splicing in plants. Slabas et al. (2004) detected t-RNA ligase in cell wall of Arabidopsis, implying the potential function of the enzyme on cell wall synthesis. Marker MtChr8_47585557 in candidate gene encoding E3 ubiquitin ligase was also associated with NDFdTrad_30 in the present study. E3 ubiquitin ligase is a protein that recruits ubiquitin and catalyzes its transfer from E2 subunit to the substrate directly or indirectly (Buetow and Huang, 2016). Borah and Khurana (2018) found that E3 ligase subunit F-box Kelch 1 (FBK1) was involved in root secondary cell wall thickening and lignin biosynthesis in rice. Another ubiquitin ligase protein, Arabidopsis Tóxicosen Levadura54 (ATL54) identified in Arabidopsis, participated in programmed cell death and secondary cell wall formation, although no significant variance on lignin content and composition in whole mature stem was observed by altering the expression level of *ATL54* (Noda et al., 2013). TTNDFD associated marker MtChr4_32514187 was encompassed by a gene encoding sulfurtransferase (Str). The interaction between Str and thioredoxin was found in different organisms, suggesting the involvement of Str in redox homeostasis, in turn affecting lignin biosynthesis (Papenbrock et al., 2011).

ADICP associated marker MtChr2_6808038 was in a *Tubby* gene. Tubby-like protein (TLP) is a large family and widespread in many plant species. However, only a small proportion of TLPs referring to stress response and

pollen development have been identified in plants (Wang et al., 2018). In the present study, marker MtChr2_6808038 identified in the gene region of TLP was associated with ADICP, which represents the portion of indigestible protein in rumen. The same gene was also identified previously in association with dry matter yield and early seedling growth (Sakiroglu and Brummer, 2017). Two annotated genes closely located with three markers were associated with NDICP, which is fiber-bound and considered to be an indigestible protein in rumen and the small intestine. One of the genes encoding 1,3-beta-glucan synthase is responsible for callose formation in plants. Callose is a polysaccharide in plant cell walls and important for a variety of processes in plant development and response to biotic and abiotic stresses (Chen and Kim, 2009). In addition, Cell wall rigidity was changed by regulating expression of this type of gene in Arabidopsis (Oliveira-Garcia and Deising, 2013).

To explore more candidate genes, we extended the search to the 50-kb region surrounding the significant SNPs (Table 3). Several genes were found to be potential candidates based on the gene functions. For example, a gene encoding pectinesterase was close to MtChr1_43382331, which catalyses the de-esterification to release pectin and methanol (Micheli, 2001). MtChr4_33424222 and MtChr4_33424238 were associated with nine quality traits, all of which were NDF-related traits with strong correlations each other. These two markers were adjacent to a beta-glucosidase gene less than 30 kb. Beta-glucosidase is an essential enzyme responsible for the hydrolysis of lignocellulose and plays important roles in cell wall formation (Singhania et al., 2013; Tsuyama and Takabe, 2015; Sampedro et al., 2017). NDFD associated marker MtChr6_1927418 was around 20 kb away from a gene encoding glycosyltransferase, which is implicated in the synthesis of polysaccharide backbone and side-chain linkages in the cell wall (Amos and Mohnen, 2019). Two genes, MtA17Chr6g0449561 and MtA17Chr6g0449521, encoding galactose oxidase and galactinol-sucrose galactosyltransferase were downstream and upstream of marker MtChr6_522150, respectively. Galactose oxidase is related to pectin synthesis and influences pectin properties (Šola et al., 2019), and involved in the modification and degradation of lignocellulose (Marjamaa and Kruus, 2018). Galactinol-sucrose galactosyltransferase belongs to glycosyltransferase, which is also involved in cell wall formation.

In summary, using GBS and GWAS, we developed 46,792 high quality SNPs and identified 28 significant markers associated with 16 quality-traits. Notably, some of the markers, such as MtChr4_33424222 and MtChr4_33424238 were associated with nine traits. Another lignin associated marker MtChr1_43382331 was consistently identified in multiple environments. Putative candidate genes underlying associated loci were identified. They may play roles in regulating forage quality. These markers and genes are going to be validated in our future studies. Once validated, they can be used for developing alfalfa with improved quality. Additionally, allelic dosage helped in finding high correlations between phenotypic traits and alleles at the associated loci and deepened our understanding of how genetic factors influence forage quality.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: Sequence Read Archive (SRA) under accession PRJNA666630.

AUTHOR CONTRIBUTIONS

L-XY and ON conceived the study. SL and CM carried out the bioinformatic analysis. SL and L-XY wrote the manuscript. ON, DC, GW, GS, SF, and DL conducted field trials and collected samples for quality evaluation. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Alfalfa and Forage Research Program (AFRP) (2017-70005-27095) from USDA National Institute of Food and Agriculture.

ACKNOWLEDGMENTS

We thank Edzard van Santen for his guidance on statistical analysis of phenotypic data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.648192/full#supplementary-material>

Supplementary Figure 1 | Population structure with principal component analysis. The top 2 components were used to represent the structure. Clusters 1, 2, and 3 were displayed in green, red and blue respectively.

Supplementary Figure 2 | Frequency distribution of different quality traits in alfalfa. (A) ADF, (B) ADICP, (C) aNDF, (D) aNDFom, (E) Ash, (F) CP, (G) Fat, (H) Lignin, (I) M6Std_Mperton_48, (J) M6Trad_Mperton_48, (K) NDFdStd_24, (L) NDFdStd_30, (M) NDFdTrad_30, (N) NDFdTrad_48, (O) NDFdTrad_240, (P) NDFind_DKC, (Q) NDFkd_DKC, (R) NDFStd_48, (S) NDICP, (T) NFC, (U) NRC01_Lignin_NEL, (V) Sugar_ESC, (W) RFQ, (X) RFV, (Y) Starch, (Z) Sugar_WSC, (AA) TTNDfD_DKC, (AB) uNDF30, (AC) uNDF120, (AD) uNDF240.

Supplementary Figure 3 | Manhattan plots of significant markers associated with alfalfa quality traits by additional models.

Supplementary Figure 4 | Phenotype distribution of 34 traits in 3 locations.

Supplementary Table 1 | Alfalfa accessions used in phenotype analysis and GWAS.

Supplementary Table 2 | Markers identified using phenotype data collected in WA.

Supplementary Table 3 | Markers identified using phenotype data collected in ID.

Supplementary Table 4 | Markers identified using phenotype data collected in OR.

Supplementary Table 5 | Genotypes of 3 significant markers associated with ProtoSol, ADICP_pct and NDFdTrad_48 in top and bottom 20 entries.

REFERENCES

- Adesogan, A. T., Arriola, K. G., Jiang, Y., Oyebade, A., Paula, E. M., Pech-Cervantes, A. A., et al. (2019). Symposium review: technologies for improving fiber utilization. *J. Dairy Sci.* 102, 5726–5755. doi: 10.3168/jds.2018-15334
- Alkhalifou, F., Renard, M., Vensel, W. H., Wong, J., Tanaka, C. K., Hurkman, W. J., et al. (2007). Thioredoxin-linked proteins are reduced during germination of *Medicago truncatula* seeds. *Plant Physiol.* 144, 1559–1579. doi: 10.1104/pp.107.098103
- Alqudah, A. M., Sallam, A., Stephen Baenziger, P., and Börner, A. (2020). GWAS: fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley – A review. *J. Adv. Res.* 22, 119–135. doi: 10.1016/j.jare.2019.10.013
- Amos, R. A., and Mohnen, D. (2019). Critical review of plant cell wall matrix polysaccharide glycosyltransferase activities verified by heterologous protein expression. *Front. Plant Sci.* 10:915. doi: 10.3389/fpls.2019.00915
- Biazzi, E., Nazzicari, N., Pecetti, L., Brummer, E. C., Palmonari, A., Tava, A., et al. (2017). Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. *PLoS One* 12:e0169234. doi: 10.1371/journal.pone.0169234
- Bonsi, M. L. K., Osuji, P. O., Tuah, A. K., and Umunna, N. N. (1995). *Vernonia amygdalina* as a supplement to teff straw (*Eragrostis tef*) fed to Ethiopian Menz sheep. *Agrofor. Syst.* 31, 229–241. doi: 10.1007/bf00712076
- Borah, P., and Khurana, J. P. (2018). The OsFBK1 E3 ligase subunit affects anther and root secondary cell wall thickenings by mediating turnover of a cinnamoyl-CoA reductase. *Plant Physiol.* 176, 2148–2165. doi: 10.1104/pp.17.01733
- Bourke, P. M., Voorrips, R. E., Visser, R. G., and Maliepaard, C. (2018). Tools for genetic studies in experimental populations of polyploids. *Front. Plant Sci.* 9:513. doi: 10.3389/fpls.2018.00513
- Buetow, L., and Huang, D. T. (2016). Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* 17, 626–642. doi: 10.1038/nrm.2016.91
- Chen, X.-Y., and Kim, J.-Y. (2009). Callose synthesis in higher plants. *Plant Signal. Behav.* 4, 489–492. doi: 10.4161/psb.4.6.8359
- Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287. doi: 10.1126/science.1123061
- Cotanch, K. (2015). “Using 240 hour uNDF in the field,” in *Cornell Nutrition Conference*, (NY, USA: William H. Miner Agricultural Research Institute East Syracuse).
- Duitama, J., Quintero, J. C., Cruz, D. F., Quintero, C., Hubmann, G., Foulquie-Moreno, M. R., et al. (2014). An integrated framework for discovery and genotyping of genomic variants from high-throughput sequencing experiments. *Nucleic Acids Res.* 42:e44. doi: 10.1093/nar/gkt1381
- Federer, W. T. (1961). Augmented designs with one-way elimination of heterogeneity. *Biometrics* 17, 447–473. doi: 10.2307/2527837
- Federer, W. T., and Raghavarao, D. (1975). On augmented designs. *Biometrics* 31, 29–35. doi: 10.2307/2529707
- Fujimoto, M., Suda, Y., Vernhettes, S., Nakano, A., and Ueda, T. (2015). Phosphatidylinositol 3-kinase and 4-kinase have distinct roles in intracellular trafficking of cellulose synthase complexes in *Arabidopsis thaliana*. *Plant Cell Physiol.* 56, 287–298. doi: 10.1093/pcp/pcu195
- Gelhay, E., Rouhier, N., Navrot, N., and Jacquot, J. P. (2005). The plant thioredoxin system. *Cell Mol. Life Sci.* 62, 24–35. doi: 10.1007/s00018-004-4296-4
- Guo, D., Chen, F., Inoue, K., Blount, J. W., and Dixon, R. A. (2001). Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin. *Plant Cell* 13, 73–88. doi: 10.1105/tpc.13.1.73
- Hawkins, C., and Yu, L.-X. (2018). Recent progress in alfalfa (*Medicago sativa* L.) genomics and genomic selection. *Crop J.* 6, 565–575. doi: 10.1016/j.cj.2018.01.006
- Hoffman, P., and Combs, D. (2003). Using NDF digestibility in ration formulation. *Focus forage* 6, 1–4.
- Hoveland, C. S., and Monson, W. G. (1980). “Genetic and environmental effects on forage quality,” in *Crop Quality, Storage, and Utilization*, ed. C. S. Hoveland (New York: John Wiley & Sons, Ltd), 139–168. doi: 10.2135/1980.cropquality.c6
- Hrbáčková, M., Dvřák, P., Takáč, T., Tichá, M., Luptovčíak, I., Šamajová, O., et al. (2020). Biotechnological perspectives of omics and genetic engineering methods in alfalfa. *Front. Plant Sci.* 11:592. doi: 10.3389/fpls.2020.00592
- Jung, H.-J. G., Samac, D. A., and Sarath, G. (2012). Modifying crops to increase cell wall digestibility. *Plant Sci.* 185, 65–77. doi: 10.1016/j.plantsci.2011.10.014
- Karn, J. F., Berdahl, J. D., and Dara, S. T. (1991). Forage quality analysis using near infrared reflectance spectroscopy (NIRS) technology in North Dakota. *Farm Res.* 48, 6–9.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, X., Wei, Y., Acharya, A., Jiang, Q., Kang, J., and Brummer, E. C. (2014). A saturated genetic linkage map of autotetraploid alfalfa (*Medicago sativa* L.) developed using genotyping-by-sequencing is highly syntenous with the *Medicago truncatula* genome. *G3* 4, 1971–1979. doi: 10.1534/g3.114.012245
- Lin, S., Medina, C. A., Boge, B., Hu, J., Fransen, S., Norberg, S., et al. (2020). Identification of genetic loci associated with forage quality in response to water deficit in autotetraploid alfalfa (*Medicago sativa* L.). *BMC Plant Biol.* 20:303. doi: 10.1186/s12870-020-02520-2
- Lopes, F., Ruh, K., and Combs, D. K. (2015). Validation of an approach to predict total-tract fiber digestibility using a standardized in vitro technique for different diets fed to high-producing dairy cows. *J. Dairy Sci.* 98, 2596–2602. doi: 10.3168/jds.2014-8665
- Marjamaa, K., and Kruus, K. (2018). Enzyme biotechnology in degradation and modification of plant cell wall polymers. *Physiol. Plant.* 164, 106–118. doi: 10.1111/pp1.12800
- Marten, G. C., Brink, G. E., Buxton, D. R., Halgerson, J. L., and Hornstein, J. S. (1984). Near infrared reflectance spectroscopy analysis of forage quality in four legume species. *Crop Sci.* 24, 1179–1182. doi: 10.2135/cropsci1984.0011183x002400060040x
- Medina, C. A., Hawkins, C., Liu, X.-P., Peel, M., and Yu, L.-X. (2020). Genome-wide association and prediction of traits related to salt tolerance in autotetraploid alfalfa (*Medicago sativa* L.). *Int. J. Mol. Sci.* 21:3361. doi: 10.3390/ijms21093361
- Micheli, F. (2001). Pectin methylsterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci.* 6, 414–419. doi: 10.1016/S1360-1385(01)02045-3
- Mizuno, H., Kawahara, Y., Wu, J., Katayose, Y., Kanamori, H., Ikawa, H., et al. (2011). Asymmetric distribution of gene expression in the centromeric region of rice chromosome 5. *Front. Plant Sci.* 2:16. doi: 10.3389/fpls.2011.00016
- Noda, S., Yamaguchi, M., Tsurumaki, Y., Takahashi, Y., Nishikubo, N., Hattori, T., et al. (2013). *ATL54*, a ubiquitin ligase gene related to secondary cell wall formation, is transcriptionally regulated by MYB46. *Plant Biotechnol.* 30, 503–509. doi: 10.5511/plantbiotechnology.13.0905b
- Oliveira-Garcia, E., and Deising, H. B. (2013). Infection structure-specific expression of β -1,3-glucan synthase is essential for pathogenicity of *Colletotrichum graminicola* and evasion of β -glucan-triggered immunity in Maize. *Plant Cell* 25, 2356–2378. doi: 10.1105/tpc.112.103499
- Papenbrock, J., Guretzki, S., and Henne, M. (2011). Latest news about the sulfurtransferase protein family of higher plants. *Amino Acids* 41, 43–57. doi: 10.1007/s00726-010-0478-6
- Perrin, R. M., DeRocher, A. E., Bar-Peled, M., Zeng, W., Norambuena, L., Orellana, A., et al. (1999). Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall biosynthesis. *Science* 284, 1976–1979. doi: 10.1126/science.284.5422.1976
- Popp, J. D., McCaughey, W. P., Cohen, R. D. H., McAllister, T. A., and Majak, W. (2000). Enhancing pasture productivity with alfalfa: a review. *Can. J. Plant Sci.* 80, 513–519. doi: 10.4141/P99-049
- Reddy, M. S. S., Chen, F., Shadle, G., Jackson, L., Aljoe, H., and Dixon, R. A. (2005). Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). *PNAS* 102, 16573–16578. doi: 10.1073/pnas.0505749102
- Reid, R. J., and Smith, F. A. (2000). The limits of sodium/calcium interactions in plant growth. *Funct. Plant Biol.* 27, 709–715. doi: 10.1071/pp00030
- Rengel, Z. (1992). The role of calcium in salt toxicity. *Plant Cell Environ.* 15, 625–632. doi: 10.1111/j.1365-3040.1992.tb01004.x
- Riday, H., Brummer, E. C., and Moore, K. J. (2002). Heterosis of forage quality in alfalfa. *Crop Sci.* 42, 1088–1093. doi: 10.2135/cropsci2002.1088

- Rocher, S., Jean, M., Castonguay, Y., and Belzile, F. (2015). Validation of genotyping-by-sequencing analysis in populations of tetraploid alfalfa by 454 sequencing. *PLoS One* 10:e0131918. doi: 10.1371/journal.pone.0131918
- Rosyara, U. R., Jong, W. S. D., Douches, D. S., and Endelman, J. B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome* 9, 1–10. doi: 10.3835/plantgenome2015.08.0073
- Sakiroglu, M., and Brummer, E. C. (2017). Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. *Theor. Appl. Genet.* 130, 261–268. doi: 10.1007/s00122-016-2782-3
- Sampedro, J., Valdivia, E. R., Fraga, P., Iglesias, N., Revilla, G., and Zarra, I. (2017). Soluble and membrane-bound β -glucosidases are involved in trimming the xyloglucan backbone. *Plant Physiol.* 173, 1017–1030. doi: 10.1104/pp.16.01713
- Scheben, A., Batley, J., and Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol. J.* 15, 149–161. doi: 10.1111/pbi.12645
- Scheller, H. V., and Ulvskov, P. (2010). Hemicelluloses. *Annu. Rev. Plant Biol.* 61, 263–289. doi: 10.1146/annurev-arplant-042809-112315
- Schrack, K., Fujioka, S., Takatsuto, S., Stierhof, Y.-D., Stransky, H., Yoshida, S., et al. (2004). A link between sterol biosynthesis, the cell wall, and cellulose in Arabidopsis. *Plant J.* 38, 227–243. doi: 10.1111/j.1365-313X.2004.02039.x
- Shenk, J. S., and Westerhaus, M. O. (1991). Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci.* 31, 469–474. doi: 10.2135/cropsci1991.0011183x003100020049x
- Shin, J. S., Kim, S. Y., So, W. M., Noh, M., Yoo, K. S., and Shin, J. S. (2020). Lon domain-containing protein 1 represses thioredoxin γ 2 and regulates ROS levels in Arabidopsis chloroplasts. *FEBS Lett.* 594, 986–994. doi: 10.1002/1873-3468.13664
- Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C., and Pandey, A. (2013). Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour. Technol.* 127, 500–507. doi: 10.1016/j.biortech.2012.09.012
- Slabas, A. R., Ndimba, B., Simon, W. J., and Chivasa, S. (2004). Proteomic analysis of the Arabidopsis cell wall reveals unexpected proteins with new cellular locations. *Biochem. Soc. Trans.* 32, 524–528. doi: 10.1042/bst0320524
- Šola, K., Gilchrist, E. J., Ropartz, D., Wang, L., Feussner, I., Mansfield, S. D., et al. (2019). RUBY, a putative galactose oxidase, influences pectin properties and promotes cell-to-cell adhesion in the seed coat epidermis of Arabidopsis. *Plant Cell* 31, 809–831. doi: 10.1105/tpc.18.00954
- Tsuyama, T., and Takabe, K. (2015). Coniferin β -glucosidase is ionically bound to cell wall in differentiating xylem of poplar. *J. Wood Sci.* 61, 438–444. doi: 10.1007/s10086-015-1486-7
- Undersander, D., Moore, J. E., and Schneider, N. (2002). Relative forage quality. *Focus forage* 4, 1–2. doi: 10.1094/fg-2010-0125-01-rs
- Wang, M., Xu, Z., and Kong, Y. (2018). The tubby-like proteins kingdom in animals and plants. *Gene* 642, 16–25. doi: 10.1016/j.gene.2017.10.077
- Wong, N. C., Wong, L. H., Quach, J. M., Canham, P., Craig, J. M., Song, J. Z., et al. (2006). Permissive transcriptional activity at the centromere through pockets of DNA hypomethylation. *PLoS Genet.* 2:e17. doi: 10.1371/journal.pgen.0020017
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yu, L.-X., Zhang, F., Culma, C. M., Lin, S., Niu, Y., Zhang, T., et al. (2020). Construction of high-density linkage maps and identification of quantitative trait loci associated with *Verticillium* wilt resistance in autotetraploid alfalfa (*Medicago sativa* L.). *Plant Dis.* 104, 1439–1444. doi: 10.1094/PDIS-08-19-1718-RE

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Lin, Medina, Norberg, Combs, Wang, Shewmaker, Fransen, Llewellyn and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.