

Introduction

- Alfalfa is an important forage crop for livestock production.
- Improving biomass yield and forage quality is essential for breeding superior cultivars. These traits are complex and controlled by many genes and environmental factors.
- GWAS identifies markers in linkage disequilibrium with QTLs associated with yield and quality traits.
- Multivariate models and machine learning may improve the detection and prioritization of important genomic regions.

Objective

- Identify genomic loci associated with alfalfa biomass yield and forage quality traits using single trait GWAS, multivariate GEMMA, and machine-learning approaches.

Material and Methods

- Phenotyping: Fifteen forage-quality traits were measured by NIRS from the from the 3rd regrowth cycle in April 2018, 2019 and 2020; trait BLUEs were used for analysis.
- Genotyping: Genotyping-by-sequencing generated 6,267 SNPs from 267 individuals. ZhongmuNo.1¹ was used as reference genome to identify SNPs.
- Single-trait GWAS: GWASpoly (Genome-Wide Association Studies for polyploids) Version 2.14 tested multiple polyploid gene-action models; significance was declared at FDR ≤ 0.05.
- Genome-wide association analysis of forage biomass was conducted in GWASpoly with FDR and suggestive threshold.
- Multivariate GWAS: GEMMA (mvLMM)² (Genome-wide Efficient Mixed Model Association using a multivariate Linear Mixed Model) evaluated biologically correlated bivariate and trivariate trait combinations using pLRT & FDR correction.
- Machine learning: Random Forest, LASSO, and XGBoost³ were used to rank informative SNPs & compare signals across biomass yield and forage-quality traits.

Results



Figure 1. Pearson correlation heatmap of BLUE values for 15 alfalfa forage-quality traits. Red indicates positive correlations, blue indicates negative correlations, and color intensity represents correlation strength. Strong relationships among fiber and digestibility traits were used to guide the selection of biologically relevant, less redundant trait combinations for multivariate GWA.

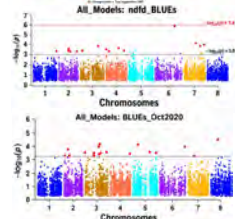


Figure 2. Manhattan plots from single-trait GWAS using GWASpoly. Marker-trait associations were evaluated across polyploid gene action models. Red points indicate significant SNPs for each trait, and black threshold lines indicate suggestive and red FDR-supported association levels.

Figure 3. Manhattan plot for biomass yield BLUEs using all single-trait GWAS models. Several suggestive SNP signals were observed across chromosomes, but no markers passed the FDR ≤ 0.05 threshold, indicating limited evidence for major-effect biomass-associated loci in this population.

Table 1. Summary of 11 unique SNPs significantly associated (FDR < 0.05) with forage quality traits in alfalfa identified using GWASpoly additive and dominant models. For each marker, the associated trait(s), significant genetic model(s), highest GWAS score (-log₁₀P), and estimated marker effects are reported. Several SNPs were associated with multiple correlated traits, indicating potential pleiotropic loci influencing forage quality.

Traits*	Marker	Models**	Best Score	Effects
IVTDM48, LIGNIN, UNDF240	Chr1_100221159	1-dom-alt, 1-dom-alt, 1-dom-alt	4.621	4.29, -1.22, -4.5
P	Chr2_10433805	2-dom-ref	5.428	-0.01
LIGNIN	Chr2_76453529	2-dom-alt	4.621	-1.22
IVTDM48, LIGNIN	Chr3_10510979	1-dom-alt, 1-dom-alt	4.621	4.29, -1.22
LIGNIN	Chr4_17631605	2-dom-alt	4.621	-1.22
ASH	Chr4_75750793	1-dom-alt	4.7	-0.27
ASH	Chr4_75750808	1-dom-alt	4.7	-0.27
IVTDM48, LIGNIN	Chr4_81820886	1-dom-alt, 1-dom-alt	4.621	4.29, -1.22
IVTDM48, LIGNIN	Chr4_81820908	1-dom-alt, 1-dom-alt	4.621	4.29, -1.22
IVTDM48, NDFD, NDFD	Chr6_81996729	1-dom-alt, 1-dom-alt, additive	5.835	-0.6, -0.59, -0.52
IVTDM48, LIGNIN	Chr7_72836002	1-dom-alt, 1-dom-alt	4.621	4.29, -1.22

*Trait abbreviations: CP, crude protein; ADF, acid detergent fiber; NDF, neutral detergent fiber; NDFD, neutral detergent fiber digestibility; ash, ash content; Lignin, acid detergent lignin; IVTDM48, in vitro true dry matter digestibility (48 h); aNDF48, digestible neutral detergent fiber (48 h); aNDF40m, ash-free adjusted neutral detergent fiber; uNDF240, undigested neutral detergent fiber after 240 h; uNDF240m, ash-free undigested neutral detergent fiber after 240 h; Ca, calcium; K, potassium; Mg, magnesium; P, phosphorus.

**Models: simplex-dominant alternative-allele (1-dom-alt), simplex-dominant reference-allele (1-dom-ref), duplex-dominant alternative-allele (2-dom-alt), and duplex-dominant reference-allele (2-dom-ref) models.

Results

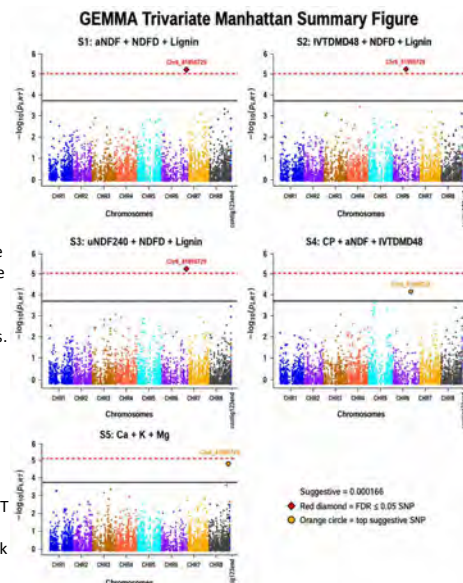


Figure 4. Five trivariate models were tested; Chr6_81996729 passed FDR ≤ 0.05 in three models

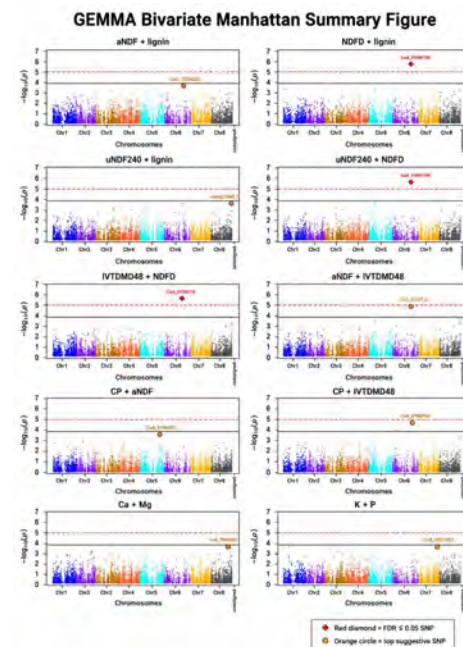


Figure 5. Ten bivariate models were tested; Chr6_81996729 was significant in five, supporting it as a robust locus for forage digestibility.

Key Findings

- Trait selection: Correlation analysis of 15 forage-quality traits revealed substantial redundancy. Representative traits were selected based on correlation structure and biological relevance.
- Principal Component Analysis (PCA) revealed two partially differentiated genetic clusters, confirming modest population structure; PC1 and PC2 explained 1.11% and 1.05% of the total SNP variation, respectively.
- Biomass yield: No SNPs reached FDR ≤ 0.05, indicating a highly polygenic trait and/or strong environmental influence while using suggestive threshold > 3 gives 433 associations and 205 SNPs.
- Single-trait GWAS: Identified 28 significant associations, representing 11 unique SNPs across seven forage-quality traits using the additive model alone identified only 1 unique SNP, while the combination of additive and 1-dom-alt models identified 11 unique SNPs. This indicates that including non-additive models increased SNP discovery.
- In contrast, using suggestive threshold 546 significant associations and 194 suggestive SNPs are detected.
- Lead locus: Chr6_81996729 was the strongest signal for NDFD under the additive model ($p = 1.46 \times 10^{-6}$; FDR = 0.0088) and was also detected across additional traits and gene-action models.
- Multivariate validation: Chr6_81996729 passed FDR correction in five trivariate and bivariate GEMMA models.
- Integrated evidence: GWASpoly, identified GEMMA identified Chr6_81996729 as the leading candidate region associated with forage digestibility and cell-wall quality.
- ML-prioritized SNPs were compared with GEMMA multivariate GWAS. The strongest overlap was found at Chr6_81996729 for ndfd_BLUEs and ivtdmd48_BLUEd and GEMMA FDR significant. Additional ML-GEMMA suggestive loci were detected at Chr3_80476692 for for ndfd_BLUEs and ivtdmd48_BLUEd and near Chr5_87862232/Chr5_87862239 for cp_BLUEs and ash_BLUEs.

Conclusion

- Biomass yield did not show robust FDR-significant SNP associations, suggesting a highly polygenic and environmentally influenced architecture.
- Forage-quality traits produced stronger signals. Single-trait analysis identified 28 FDR-significant associations representing 11 unique SNPs across seven traits, with more SNPs detected when multiple gene action models were considered.
- All the analysis prioritized consistently prioritized Chr6_81996729 as the strongest NDFD-centered locus.

Research Direction

- Future work will focus on validating Chr6_81996729 and other suggestive loci in independent populations, performing LD-based region mapping and gene annotation.
- Causal gene discovery and molecular marker development for informative breeding
- Additional multi-environment biomass analysis is needed because biomass yield did not show robust FDR-significant associations in the current study.
- Refining Machine learning models for more robust outcome.

Reference

- Shen C, Du H, Chen Z ...The Chromosome-Level Genome Sequence of the Autotetraploid Alfalfa and Resequencing of Core Germplasm Provide Genomic Resources for Alfalfa Research Molecular Plant, 2020; 13, 1250-1261
- Xiang Zhou and Matthew Stephens (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods* 11, 407–409.
- GitHub - Morteza-M-Saber/BacGWASim_ML_benchmark: The pipeline to evaluate ML and GWAS feature selection tools - GitHub

Acknowledgement

- Funding support - Alflorex Seeds & New Mexico Agricultural Experiment Station
- Genotyping & phenotyping – Alflorex Seeds, New Mexico State University, USDA-ARS (Prosser, Washington)