

# Prospects for improving Alfalfa Yield Using Genomic- and Phenomic-Based Breeding

Francis, M. W.\*; Pap, D.\*; Krill-Brown, A.\*; Brummer, E. C.\*  
\*University of California, Davis

**Key words:** Alfalfa; genomic selection; remote sensing; plant breeding

## Abstract

Alfalfa (*Medicago sativa* L.) is a perennial outcrossing legume that is cultivated as an important forage crop in many parts of the world. Yield is the most important trait for profitable alfalfa production, yet over the last 30 years yield improvement in California has stagnated. Current breeding methods focus on recurrent phenotypic selection; however, alternatives incorporating genomic- and phenomic-based information may enhance genetic gain and help to address the lack of yield improvement. Here we attempt to increase the yield potential of alfalfa using genomic selection (GS) in combination with high throughput phenotyping (HTP). A total of 193 families from two closely related elite populations were sown in the greenhouse and transplanted into mini sward plots at two locations near Davis, CA in May 2020. The trial was managed as a high-input system under full irrigation. Families were genotyped and phenotyped for biomass yield by mechanical harvest and a combination of drone and tower-based remote sensors across 12 harvests, 3 in the establishment year (2020), 7 in the first full year of production (2021) and 2 in 2022. Alfalfa yields ranged from 13-27 tonnes DM/hectare/year with a number of half-sib families outperforming popular cultivars in the first 2 years of production. Biomass volume predicted from the drone-based cameras had a moderate prediction accuracy with an overall  $R^2$  of 0.55. Some individual harvests reached accuracies as high as 0.85. Genotyping resulted in a dataset with 6,838 SNPs. Allele frequencies were used to generate a relationship matrix for GS. Narrow-sense heritability for dry matter yield was 0.31 and the predictive ability of the GS model was 0.15.

## Introduction

Alfalfa (*Medicago sativa* L.) is a legume forage crop primarily grown for hay, silage and pasture. Cultivated alfalfa is an outcrossing autotetraploid ( $2n = 4x = 32$ ) with a genome size of 800-1000 Mb (Blondon et al. 1994). Yield is the most important trait for profitable alfalfa production, yet somewhat inexplicably, yield improvement in alfalfa has stalled over the last ~30 years (Brummer & Casler 2014). Often referred to as the 'Queen of forages', alfalfa is one of California's most important crops, driving production of the state's most significant agricultural enterprise: dairy. In 2021, 580,000 hectares of alfalfa were harvested for hay and haylage with a total value of \$836MM (USDA 2022). When combined with the \$7.5B value of milk and dairy production, alfalfa accounts for a significant portion of California's agricultural GDP. A variety of explanations have been proposed to explain the reduced yield improvements in alfalfa. Firstly, yield improvement in forage crops already lagged far behind that of annual grain crops due in part to its perennial nature requiring multiple years of evaluation before selection can be made; the additional demands of winter survival; and the inability to make gains in the harvest index as all above ground biomass is harvested. Additionally, yield is often not selected for explicitly in breeding programs, instead the focus has been on improvements in disease and insect resistance. Although such a breeding focus can lead to realisation of yield potential, it is not increasing yield per se. Lastly, the requirement in selecting for forage quality and long-term persistence diverts attention from direct yield improvement.

Our research seeks to address the lack of improvement in alfalfa yield by focusing on increasing its yield potential. This can be accomplished by implementing genomic selection in our breeding program accompanied by high-throughput phenotyping technology. Genomic selection could increase the rate of genetic gain in alfalfa yield by reducing the cycle time length for typical selection programs. In addition, the genomic data could be used to improve other important, complex, quantitatively inherited traits including nutritive value, seed yield and multi-year persistence. One of the major constraints on the efficacy of genomic selection is the size of the training population, where the accuracy of genomic estimated breeding values (GEBVs) increase as the size of the training population increases (Jannink, Lorenz, & Iwata, 2010). Traditionally, phenotypic data collection in alfalfa breeding programs requires harvesting and weighing individual plots up to 10 times a year and is the largest constraint of trial size. By utilising modern high-throughput phenotyping technologies on a drone-based platform, we can reduce the amount of harvesting required and can greatly increase trial size, thus improving prediction accuracy.

## Methods

In May 2020 we established a field trial in two locations: Yolo (38°32'09 N, 121°46'32 W) and Solano (38°31'39 N, 121°46'18 W) Counties in the Sacramento Valley. Trial entries consisted of CUF 101, UC-Impalo-WF, Highline, and ~100 half-sib families from each of two closely related but distinct, elite UC Davis breeding populations derived from UC-Impalo-WF. Plants were grown in trays in the greenhouse for 2.5 months before being transplanted into the field. Plots are 2' X 3' 4", comprised of 6 rows of 4 plants (24 plants total) spaced 8 inches apart to emulate competition faced in a commercial alfalfa field. Each location contains two reps consisting of 203 plots (812 plots total).

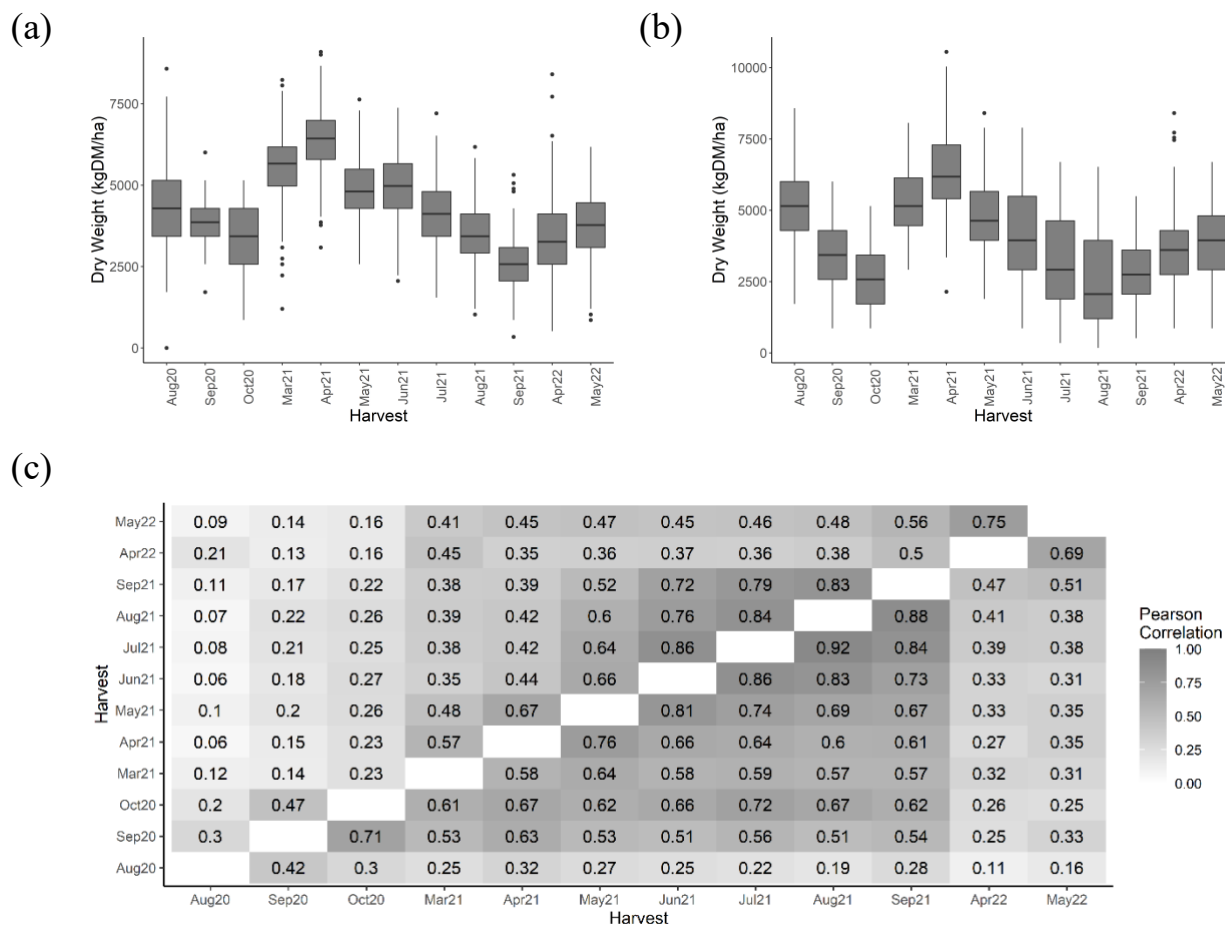
Yield was measured each harvest using a John Deere self-propelled forage harvester with the 36A research plot harvester attachment from RCI Engineering. Harvests occurred on a 28-day cutting schedule throughout the growing season. In addition, prior to each harvest, a drone-based Red-Edge-M™ multispectral camera was flown at a height of 15 m to capture images for estimation of forage biomass volume. Raw images were stitched using Pix4D and biomass estimates calculated from the resulting files using QGIS.

Tissue samples were taken from each plant of a single rep for genotyping. The median trifoliolate of the first fully expanded leaf counted from the growing tip of the plant was selected. DNA was extracted and genotyping-by-sequencing (GBS) libraries were constructed using a process modified from Elshire et al. (2011). Each library was sequenced in two lanes on Illumina HiSeq 4000 at the UC Davis genome centre. Raw reads were processed using the Tassel 5 GBS pipeline developed by Glaubitz et al. (2014) with the following parameters: biallelic single-nucleotide polymorphisms (SNPs) only, minimum mean read depth of 64, maximum mean read depth of 500, minor allele frequency of 0.03, maximum missing data of 10% and a minimum mapping quality of 30. After filtering 6,838 SNPs were considered for downstream analysis.

Alternative and reference read counts were extracted from the variant calling file to calculate allele frequencies which were then used to generate a genomic relationship matrix that was used for genomic selection. 10-fold cross validation was then used to determine the model's predictive ability.

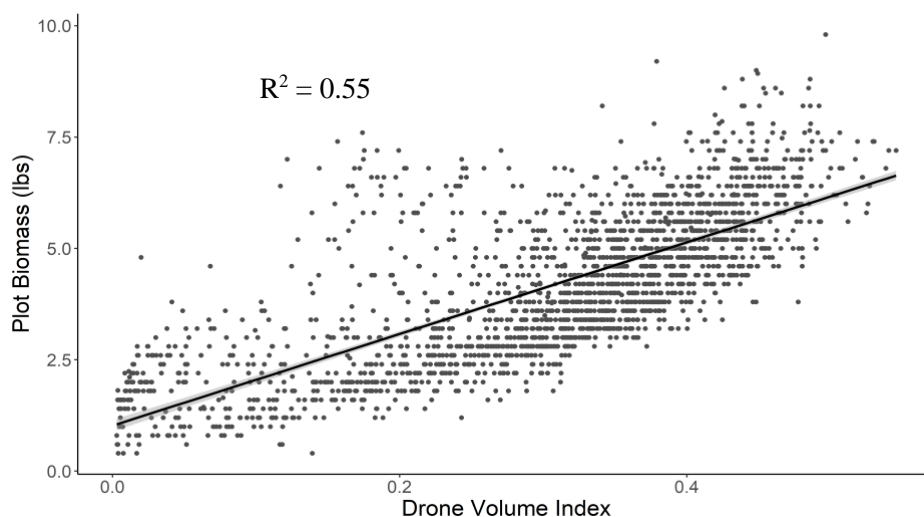
## Results and Discussion

### Phenotypic Data



**Figure 1. Dry matter yield (kgDM/ha) among alfalfa families across 12 harvests at (a) Solano County and (b) Yolo County trial locations. (c) Pearson correlation between phenotypic means for dry matter yield across 12 harvests in the upper diagonal for Solano County and lower diagonal for Yolo County trial locations.**

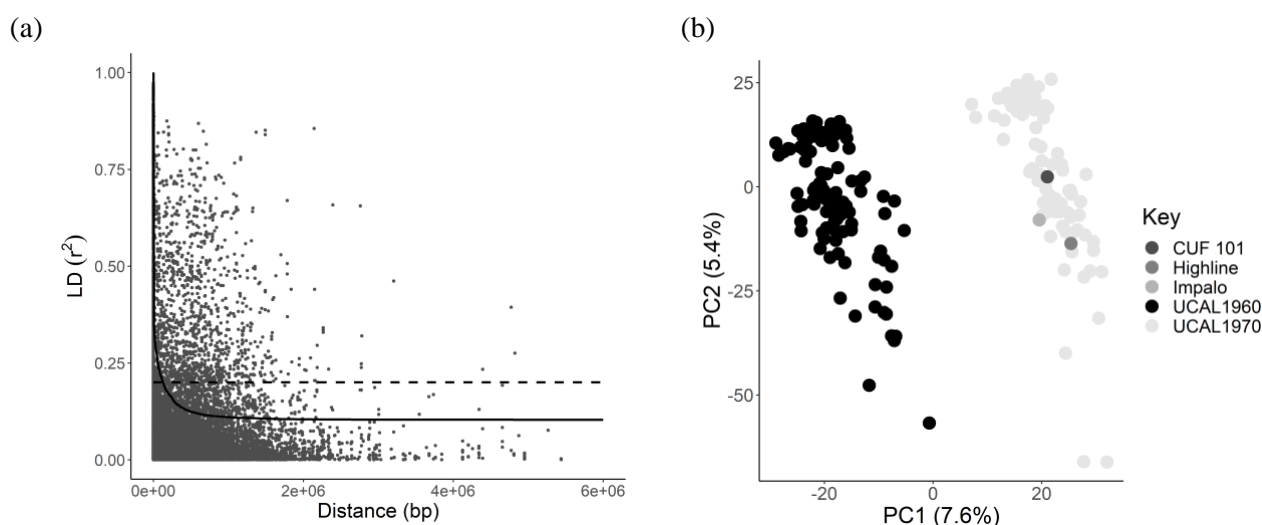
There was significant variation among families for dry matter yield (DMY) at each harvest and throughout the season. DMY at both locations was greatest for the spring harvests and lowest in late summer (Figure 1a and 1b). DMY across all trial entries ranged from 13-27 tonnes DM/ha/year. The correlation between harvests for DMY ranged from 0.06 to 0.92, with successive harvests having closer correlations than harvests further apart as would be expected (Figure 1c). Both populations in the trial had half-sib families that outperformed CUF 101, UC-Impalo-WF and Highline for total DMY. This, coupled with the wide range of phenotypes suggests potential for developing populations with increased yield potential.



**Figure 2. Drone estimated biomass volume plotted against actual measured fresh plot biomass across all locations and harvests.**

Biomass volume predicted from the drone-based cameras had a moderate prediction accuracy with an overall  $R^2$  of 0.55 (Figure 2). The accuracy for some individual harvests was as high as 0.85. Variation in accuracy can be attributed to a variety of factors including weeds, non-uniform soil, lodging, gopher damage, wind and plant maturity. Anything that interferes with the natural height of the plants or soil may lead to spurious biomass volume estimates. For remote sensing to be successful for breeding, trials must be established in a level field with bare soil surrounding plots. They must also be free of pests and weeds and the timing of flights and harvests should be adjusted to minimise the risk of lodging and overgrown plots.

### ***Molecular markers, population genetics analyses and genomic selection***



**Figure 3. (a) Linkage disequilibrium (LD) decay between SNP pairs. The LD decay over physical distance was determined as the mean distance at the LD threshold of  $r^2 = 0.2$ . (b) Principal component analysis (PCA) of the genotype matrix showing the first two principal components that together explain 13% of the variation.**

A relationship matrix was used to perform a principal component analysis (Figure 3b). The first two principal components explained 13% of the variation and demonstrate clear population structure with the two populations present in the trial clustering independently. LD decay was measured based on the correlation between pairs of SNPs present on each chromosome (Figure 3a). Based on the threshold of  $r^2 = 0.20$ , LD decays at 130 Kb. After fitting the model, the narrow sense heritability ( $h^2$ ) for dry matter yield (DMY) was 0.31. 10-fold cross validation resulted in a predictive ability of 0.15 across all harvests.

### Conclusions and/or Implications

We have a wide range of phenotypes with many entries performing better than popular existing varieties for DMY in our yield evaluation trial. As is typical for yield, the predictive ability of our GS model is low, however there are more combinations of prediction scenarios to investigate as well as predicting forage quality. We have a significant amount of phenotypic data from both mechanical harvests and remote sensing which when coupled with the genomic information should allow us to make informed selections and develop new populations with improved yield potential

### Acknowledgements

This work was supported by grants to E. Charles Brummer from the U.S. Alfalfa Farmer Research Initiative (USAFRI) and the California Crop Improvement Association (CCIA). Matt Francis was supported by a fellowship from PGG Wrightson seeds.

### References

- Blondon, F., Marie, D., Brown, S., & Kondorosi, A. 1994. Genome size and base composition in *Medicago sativa* and *M. truncatula* species. *Genome*, 37(2), 264-270.
- Brummer, E. C., & Casler, M. D. 2014. Cool-Season Forages. In: *Yield Gains in Major U.S. Field Crops*. S. Smith, B. Diers, J. Specht, & B. Carver (eds). CSSA Special publ., pp. 33-51.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE*, 6(5). doi:<https://doi.org/10.1371/journal.pone.0019379>
- Galubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLOS ONE*, 9(2). doi: <https://doi.org/10.1371/journal.pone.0090346>
- Jannink, J.-L., Lorenz, A. J., & Iwata, H. 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, 9(2), 166-177. doi:<https://doi.org/10.1093/bfpg/elq001>
- USDA. 2022. 2021 State Agriculture Overview - California. Retrieved from: [https://www.nass.usda.gov/Quick\\_Stats/Ag\\_Overview/stateOverview.php?state=CALIFORNIA](https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=CALIFORNIA)