



GWAS-Assisted Genomic Prediction for Forage Yield Under Deficit Irrigation Management in Elite Alfalfa (*Medicago sativa* L.) Germplasm



Harpreet Kaur, Cesar A. Medina, Christopher Pierce,
Arshdeep Singh Gill, Jon Reich, Steve Damon, Long-Xi Yu, Ian
Ray



Introduction

- Alfalfa (*Medicago sativa* L.) is a perennial forage legume.
- Most cultivars are autotetraploid ($2n=4x=32$).
- Arid and semi-arid regions in the U.S. possess limited water resources for alfalfa irrigation.
- Genome-wide Association Study (GWAS) and Genomic Prediction

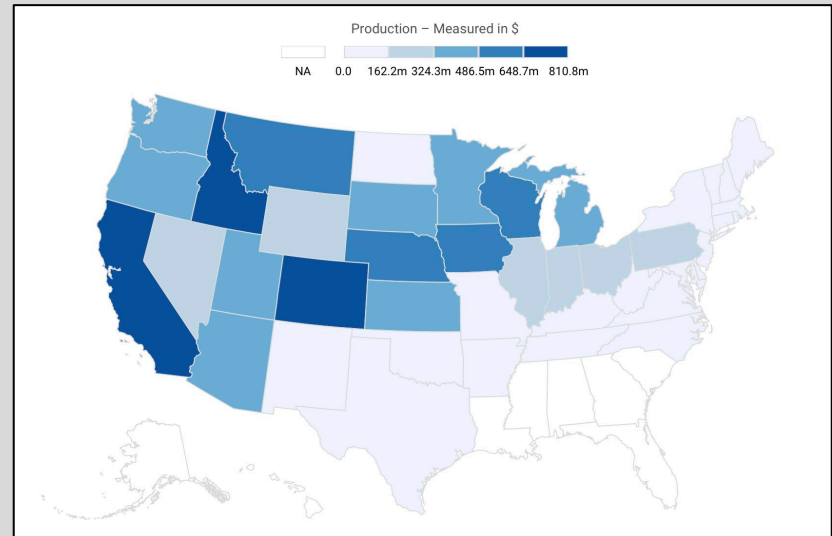


Image source: <https://nass.usda.gov>

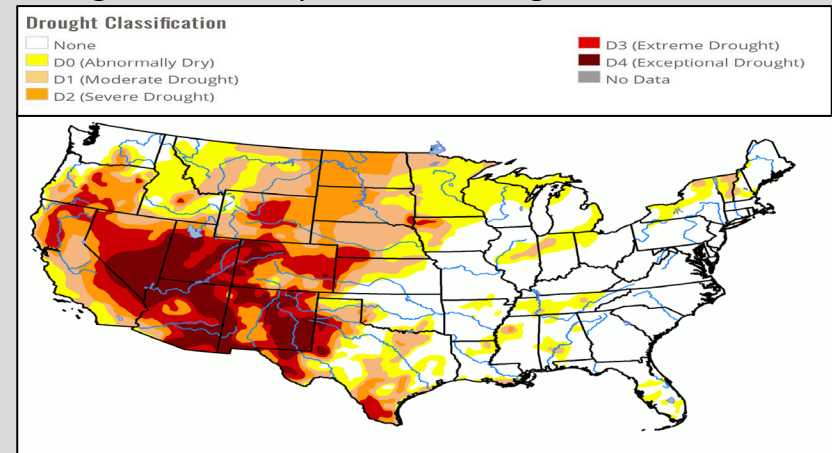


Image source: <https://droughtmonitor.unl.edu>

Step 1: Phenotyping and Genotyping

- **Phenotyping**

- **Elite germplasm** – 215 maternal half sib families
- Forage yield under deficit irrigation management
- Two locations – NM and CA
- Three years - 2018, 2019, 2020

- **Genotyping**

- 215 maternal parents
- Genotype-by-sequencing – 12,884 Single Nucleotide Polymorphism (SNP) marker

Statistical associations by integrating phenotypic and genotypic data using R based software GWASpoly (Rosyara et. al., 2016)

Step 2: Genome-wide association study

Location	Month Year	Marker	Score	Effect
NM	April 2018	chr4_06222	5.539	NA
	May 2018	chr1_00200	5.325	-0.407
	June 2018	chr1_00200	6.107	-0.148
			6.107	NA
			6.107	-0.298
			6.107	-0.352
	October 2018	chr2_01985	5.754	0.252
			5.754	0.503
			5.754	0.510
	CA	May 2020	chr3_03883	5.300

- Very few significant marker-trait associations identified based on 0.05 FDR threshold (Score > 5)

Threshold Score more than 1

Example: Marker chr1_00056

In CA

**Regrowth
cycle**

**Effect
(Score)**

Threshold Score more than 1

Example: Marker chr1_00056

In CA

Regrowth cycle	May 2018	June 2019	July 2019	July 2018	Aug 2019	Sept 2019	Total 2018	Total 2019
Effect (Score)	-0.442 (1.068)	-0.365 (1.055)	-0.580 (2.775)	-0.420 (2.345)	-0.454 (2.454)	-0.222 (2.118)	-1.420 (1.488)	-1.292 (1.161)

Threshold Score more than 1

Example: Marker chr1_00056

In CA

Regrowth cycle	May 2018	June 2019	July 2019	July 2018	Aug 2019	Sept 2019	Total 2018	Total 2019
Effect (Score)	-0.442 (1.068)	-0.365 (1.055)	-0.580 (2.775)	-0.420 (2.345)	-0.454 (2.454)	-0.222 (2.118)	-1.420 (1.488)	-1.292 (1.161)

In NM

July 2020	June 2018	June 2019	June 2020	May 2018	Oct 2019	Oct 2020	Total 2018	Total 2019	Total 2020
-0.501 (2.267)	-0.144 (1.351)	-0.205 (1.32)	-0.452 (1.568)	-0.211 (1.757)	-0.327 (1.96)	-0.421 (2.807)	-0.434 (1.195)	-0.833 (1.254)	-1.369 (2.462)

Effect of DNA Markers on Forage Yield and Quality

Yield DNA Marker	NM NDF '18	NM NDF '19	NM IVD'18	NM IVD '19	NM LIG '18	NM LIG '19	NM CP '18	NM CP '19	CA Yld Tot18	CA Yld Tot19	NM Yld Tot181920
1500			-0.39				-0.33		1.26	2.49	1.03
2500									-1.97	-2.84	
4500	-0.52	-0.79	0.49	0.70		-0.16	0.40	0.44		-1.99	-0.83
6500									1.82	2.14	1.34
8000		-0.59		0.42		-0.11			-2.25	-2.37	-1.74
9800									-1.78	-2.23	-1.31
11000									1.66	1.90	1.43
12884	0.55		-0.44						1.84	2.25	1.33

NM, New Mexico; CA, California; Yld, Yield; NDF, neutral detergent fiber; IVD, in vitro digestibility; LIG, lignin; CP, crude protein; Tot, total. Direction of DNA marker effects for yield & quality are consistent over time and make sense biologically!!

Effect of DNA Markers on Forage Quality and Yield

Forage Quality DNA Marker	NM NDF '18	NM NDF '19	NM IVD '18	NM IVD '19	NM LIG '18	NM LIG '19	NM CP '18	NM CP '19	CA Yld Tot18	CA Yld Tot19	NM Yld Tot181920
5000	-0.42	-0.32	0.34	0.27	-0.08	-0.08	0.26	0.18		-0.71	
6329	-0.78	-0.41	0.70	0.35	-0.18	-0.12	0.48	NA	1.06		
6330	0.59	0.45	-0.63	-0.36	0.18	0.12	-0.37	-0.19			

NM, New Mexico; CA, California; Yld, Yield; NDF, neutral detergent fiber; IVD, in vitro digestibility; LIG, lignin; CP, crude protein; Tot, total.

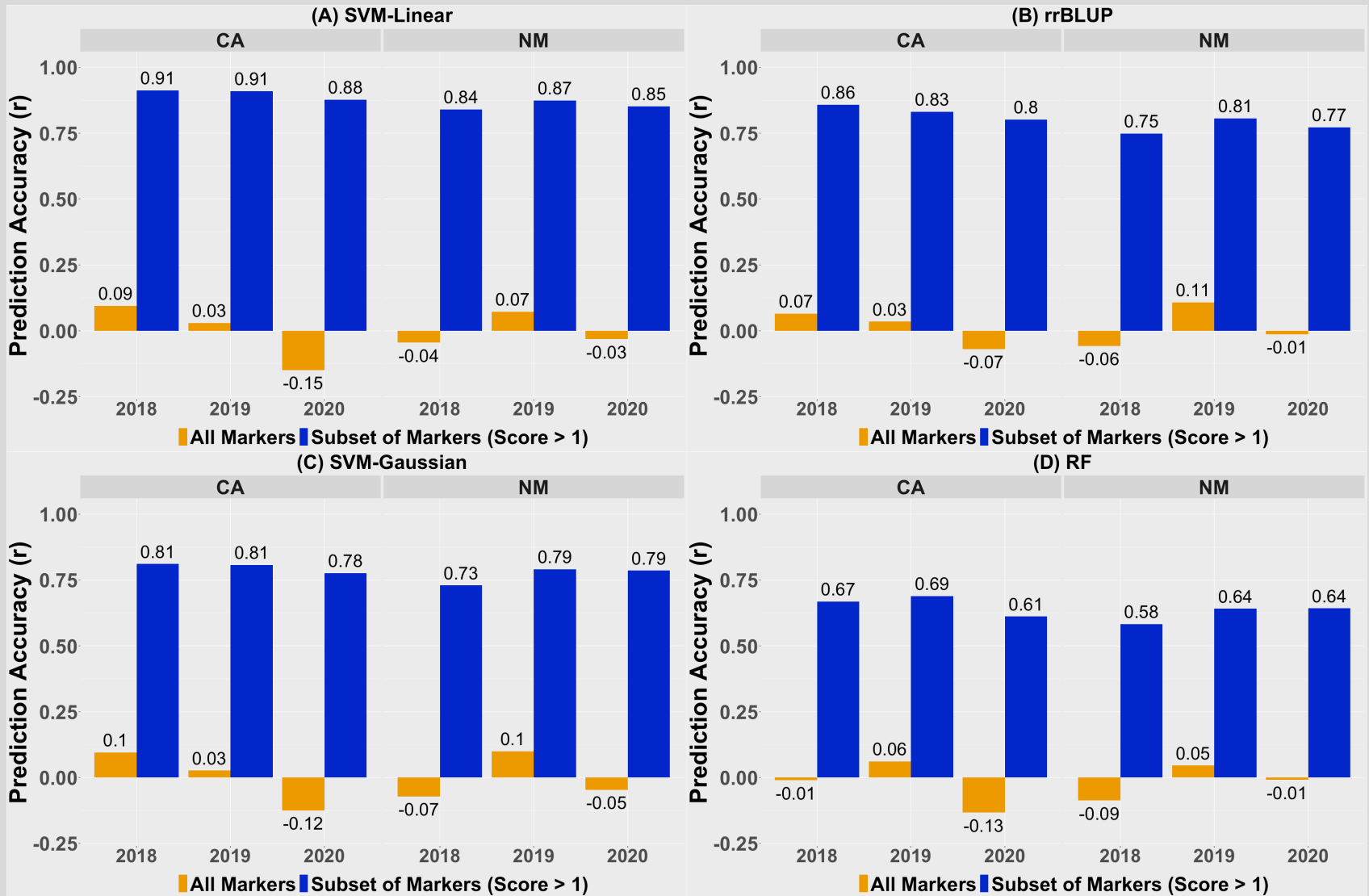
Step 3: Genomic Prediction

- Nine GP models - **Ridge Regression Best Linear Unbiased Prediction (rrBLUP), Genomic BLUP (GBLUP), Support Vector Machine (SVM) -Linear, SVM-Gaussian, Random Forest (RF), BayesA, BayesB, BayesC, and Bayesian LASSO (BL)**

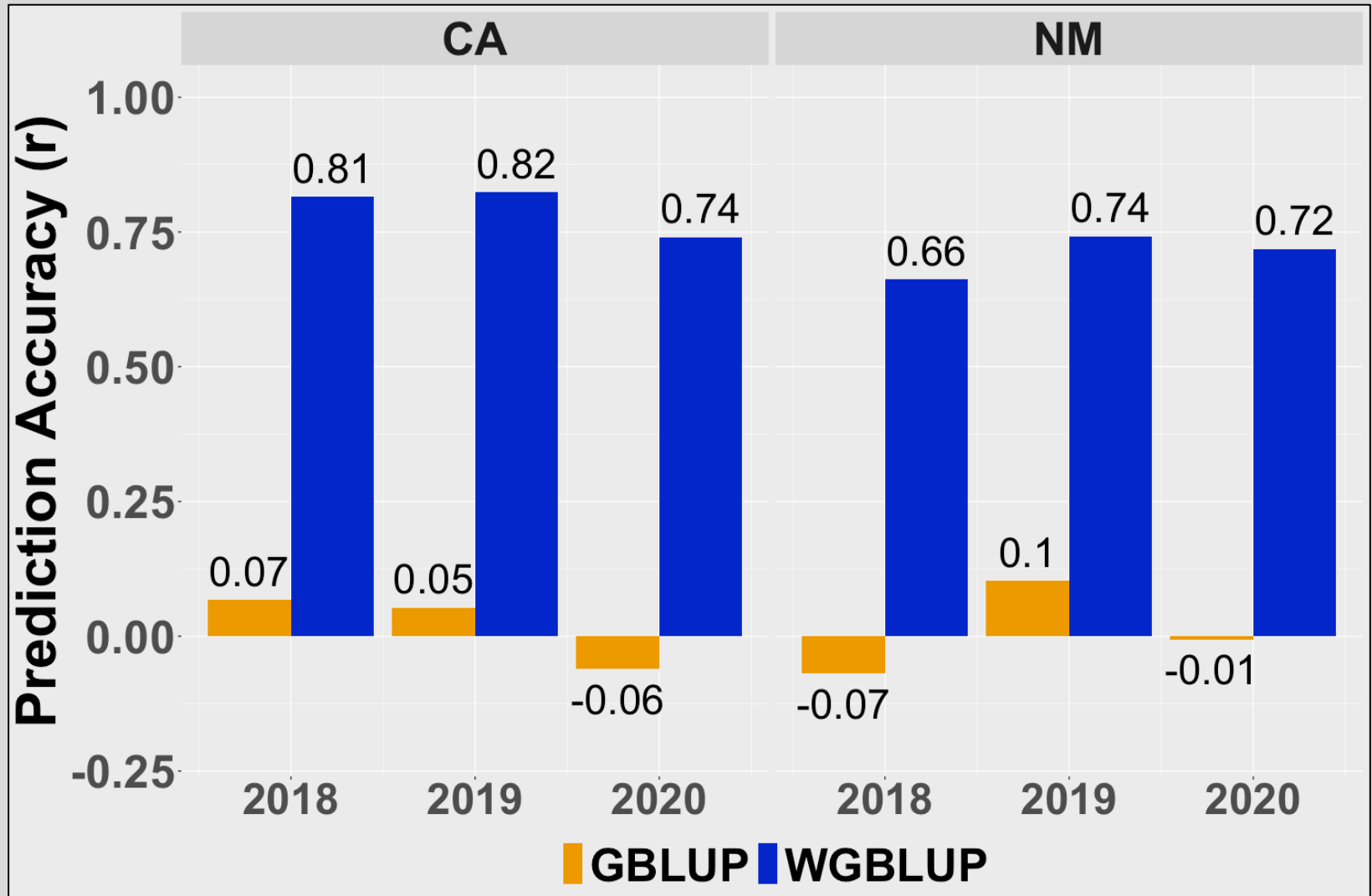
Three Approaches

- **Approach 1:** Genomic prediction modelling with all 12,884 markers
- **Approach 2:** 1,915 - 2,132 SNP markers identified with **Score** ($-\log_{10}$ p-value) more than 1 from GWAS results to run the GWAS-assisted genomic prediction
- **Approach 3:** Weighted GBLUP with all 12,884 markers using **Scores as weights** (Medina et. al., 2021) from GWASpoly 6 gene action models (Rosyara et. al., 2016)

GWAS-Assisted Genomic Prediction results



GWAS-Assisted Genomic Prediction results



Final Remarks

- Relaxation of GWAS analysis score thresholds may facilitate identification of a subset of biologically relevant SNP markers which can be included in genomic selection models to improve prediction accuracies.
- WGBLUP models also enhance prediction accuracies, and with a large number of markers, they are computationally less intensive compared to machine-learning and Bayesian models.
- Independent validation of these genomic prediction models is needed.
- Contingent upon validation, selected SNP markers may be useful for developing elite alfalfa germplasm that can be productive under deficit irrigation management.

Acknowledgements



Dr. Ian Ray
Christopher Pierce
Arshdeep Singh Gill



Dr. Jon Reich
Dr. Steve Damon



Dr. Long-Xi Yu
Dr. Cesar M. Culma



Funding support - Alforex Seeds &
New Mexico Agricultural Experiment Station



Genotyping & phenotyping – Alforex Seeds,
New Mexico State University, USDA-ARS
(Prosser, Washington)



Forage Quality Poster: P10

Thank you