

Relationship Between SNPs Density & Prediction Accuracy in Alfalfa Genomic Selection

Bill Biligetu¹ and Hu Wang²

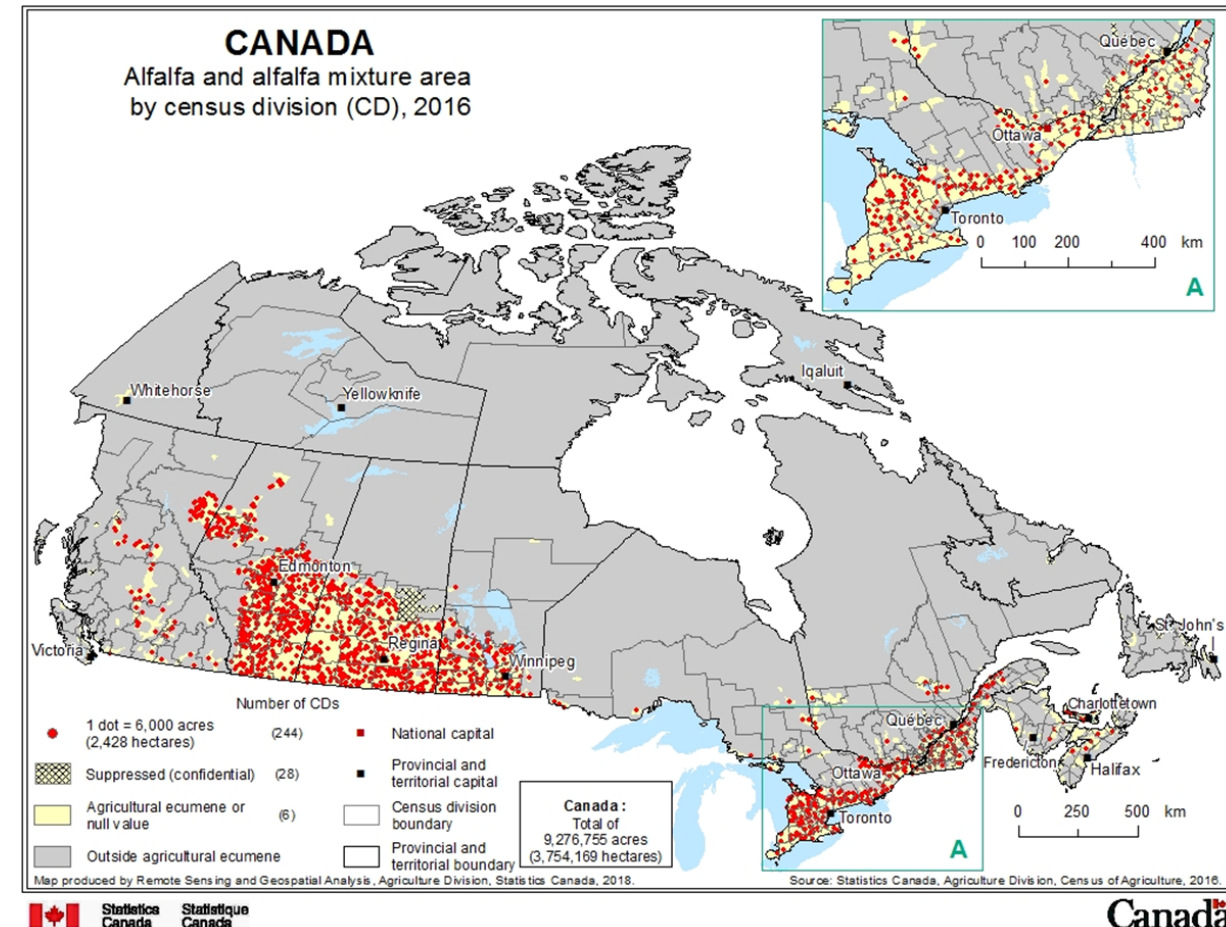
¹Crop Development Centre, University of Saskatchewan, Canada

²Grasslands Research Centre, AgResearch, Palmerston North, New Zealand

June 25 2024

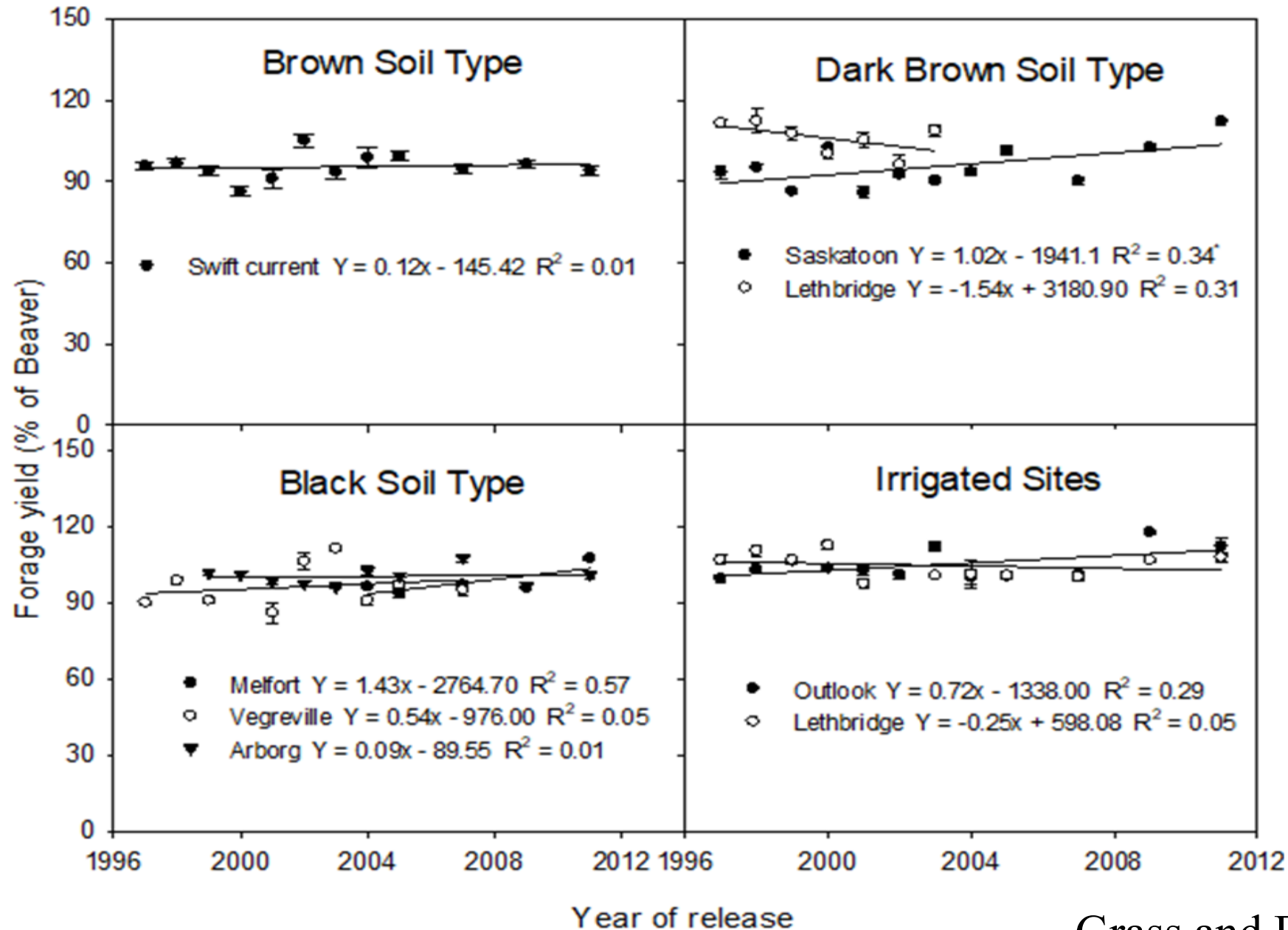
Alfalfa production in western Canada

- Western three provinces: 3 million ha in grass-alfalfa mixtures
- Winter survival is important
- Alfalfa fall dormancy score: <3-4
- 1-2 cut systems or animal grazing

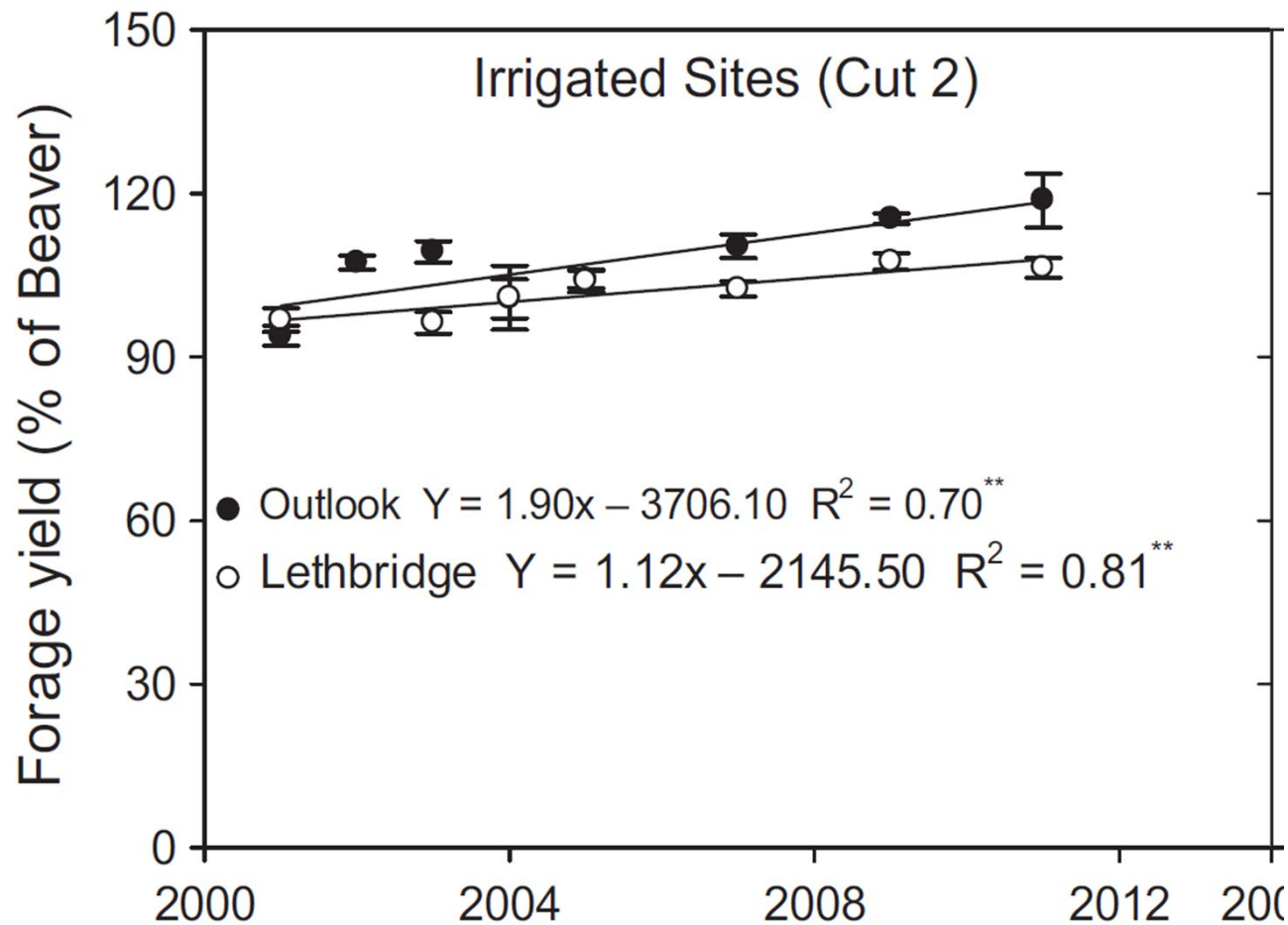


(Statistic Canada, 2016)

Alfalfa yield in western Canada



■ Alfalfa Regrowth (1997 -2011) in Western Canada



- Breeders need to improve many traits (cold tolerance, disease resistance, grazing tolerance...)
- Out-crossing, polyploid form
- Longer breeding cycles

Alfalfa breeding in Western Canada



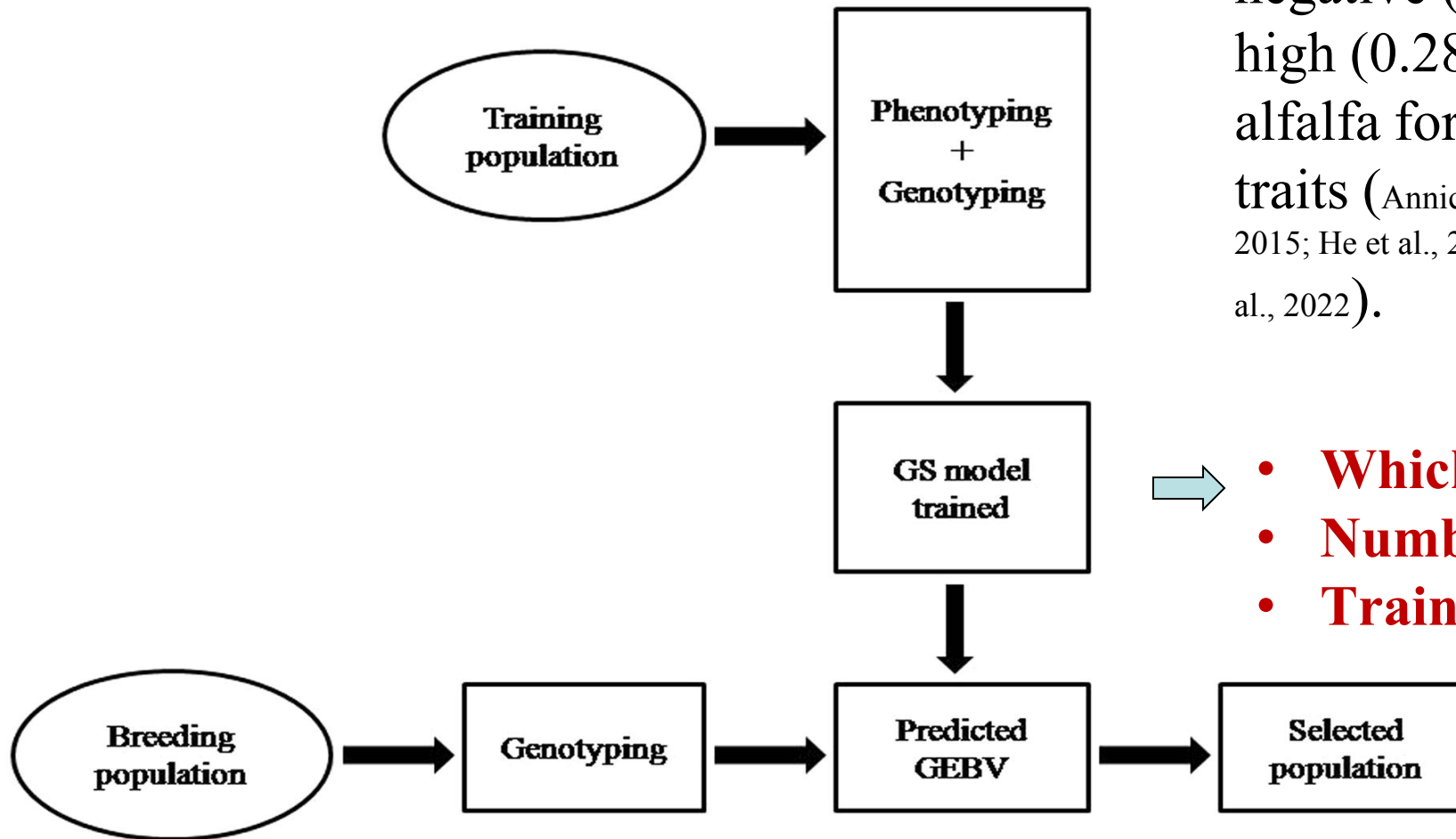
First: survive winter
($FD = <4$)



Other traits

Genomic selection

- Net genetic merit of a genotype using molecular markers distributed across the genome



- negative (-0.03), low (0.05-0.10) or high (0.28-0.51) GS accuracies for alfalfa forage yield and nutritive traits (Annicchiarico et al., 2022; Annicchiarico et al., 2015; He et al., 2022; Jia et al., 2018; Murad Leite Andrade et al., 2022).

- ➡
- **Which models**
 - **Number of molecular markers**
 - **Training to test set ratio**

Materials and Methodology

Plant materials:

14 alfalfa populations collected from long-term grazing sites (140 genotypes)

Experimental design:

A randomized complete block design (RCBD) with two blocks

Location:

Saskatoon Research Farm

Phenotypic data collection time:

2018, 2019 and 2020 growing seasons

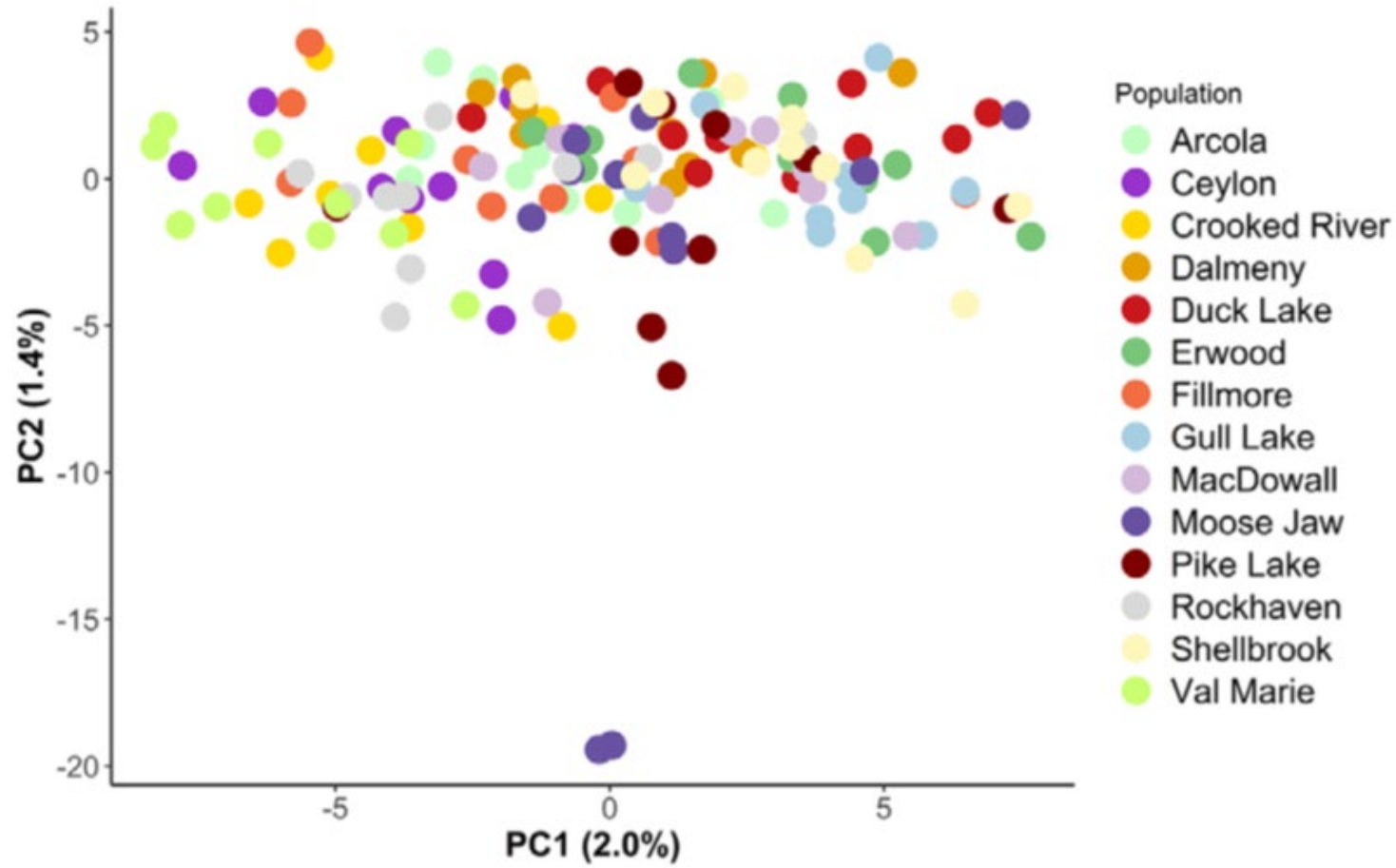
Genomic data: GBS 125bp to generate 4,932 SNPs



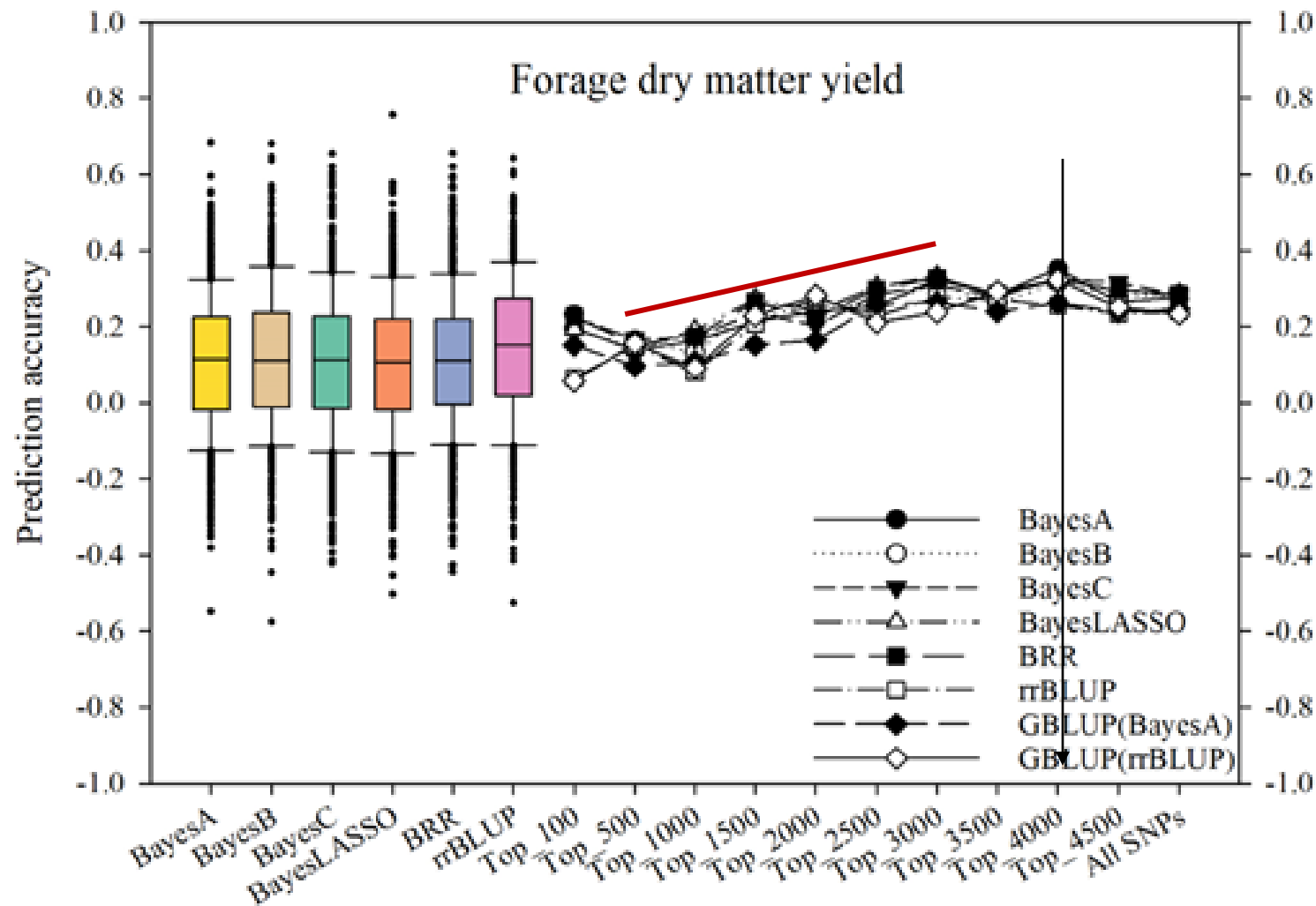
Marker effect of each SNP

- To determine the effect of SNP markers, the mean-squared estimated- marker effects (MSEMEs) for each SNP for each trait were obtained by 1000 repeats. (***Pérez & de los Campos, 2014***)
 - ❑ The SNP markers were grouped into the most informative 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, and 4500 based on the ranks
- The proportion of SNPs in the transcript regions was calculated using the alfalfa reference genome by Chen et al. (2020).
 - ❑ the SNPs located in the region for each trait increased
 - ❑ Candidate gene annotation of the top 100 SNPS showed 80%–90% of the 100 SNPs are situated within protein coding regions associated with candidate genes

Population Structure

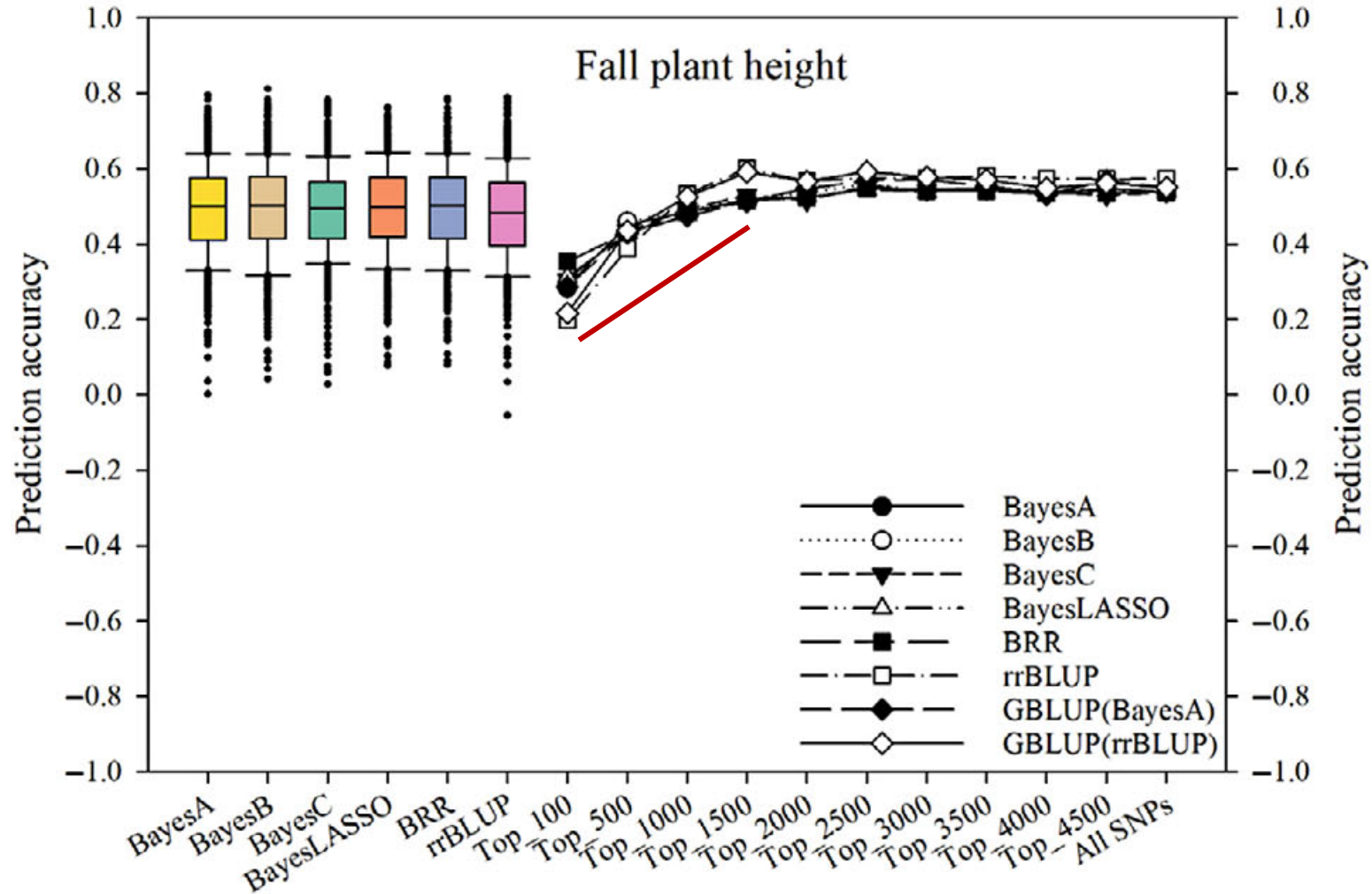


Which models are better for forage yield?



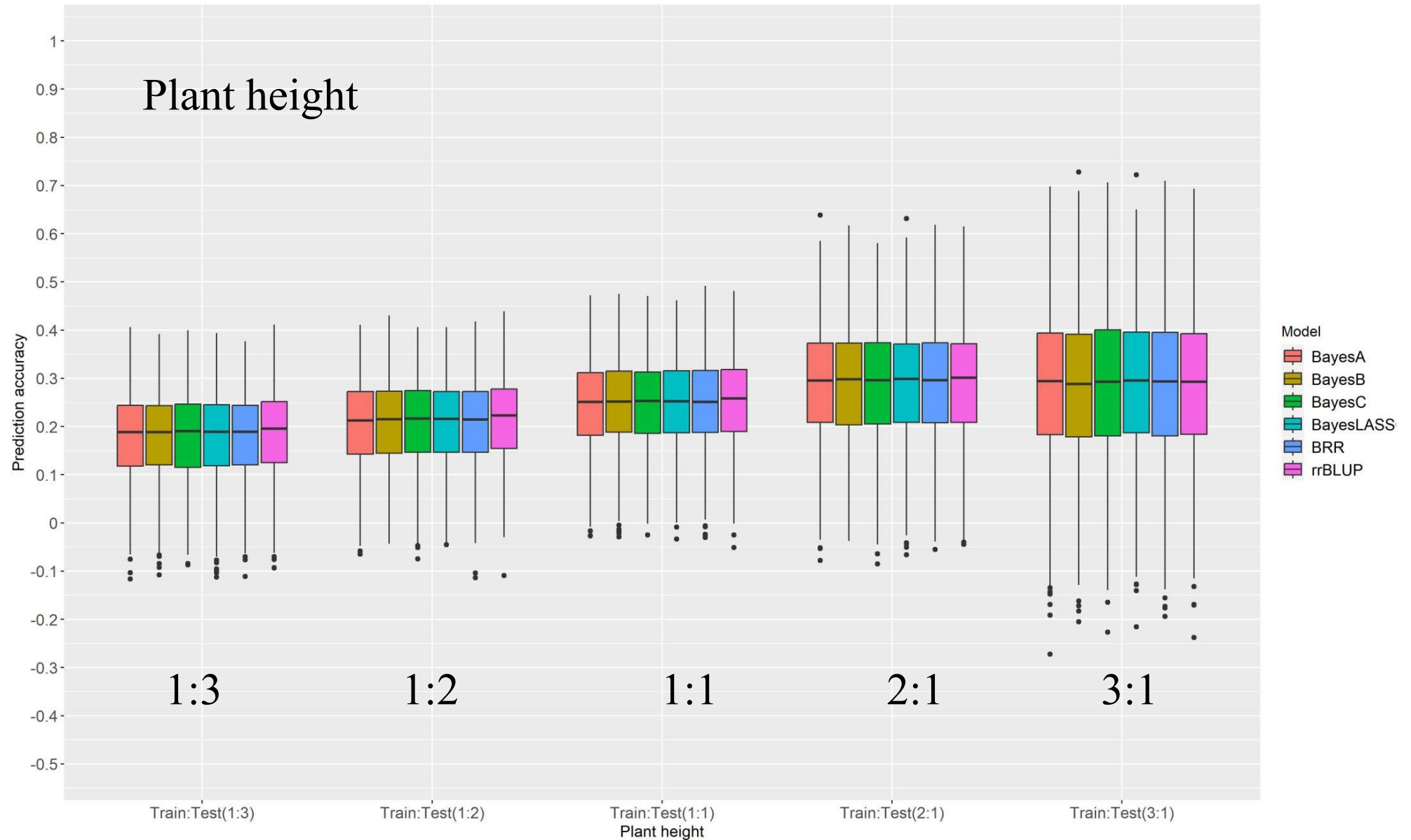
- Models are somewhat similar
- Prediction accuracy for DMV : 0.06 to 0.35 on SNPs density
- Highest accuracy of 0.35 for the top 4000 SNPs for BayesA method.

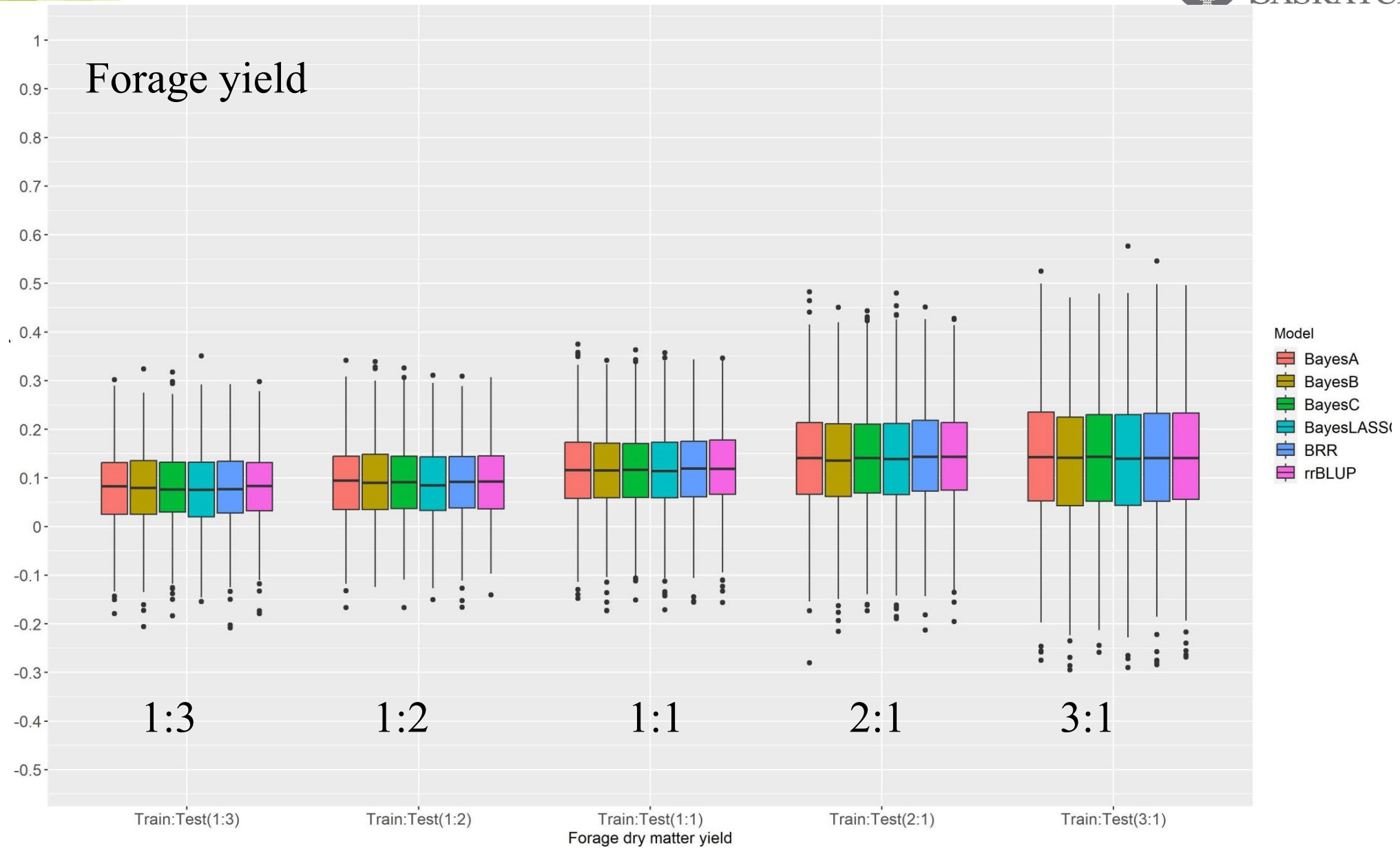
Which models are better for fall plant height?



- Prediction accuracy for SV ranged from 0.20 to 0.60 for SNPs density
- Highest prediction: top 1500 SNPs on rrBLUP method

Effect of training population size on prediction accuracy

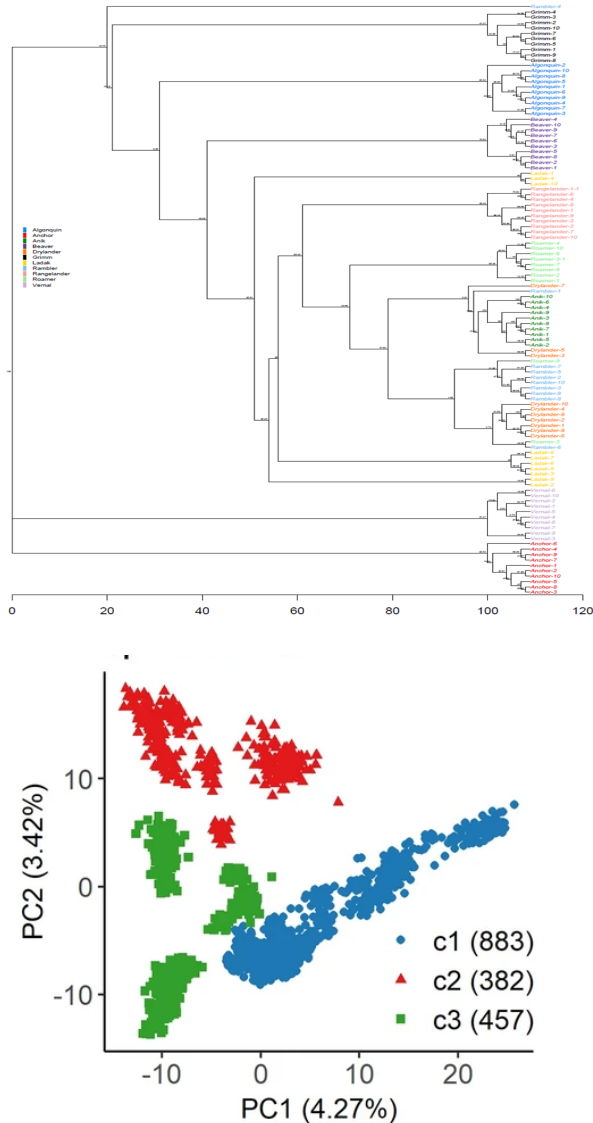




SOME TAKE AWAYS

- rrBLUP seems better for most tested traits, but the difference among GS is small
- Informative markers are more important than density of markers in most cases
- Some traits (i.e. yield) may require greater number of SNPs than other traits (i.e. fall height)

Application of genomic selection in alfalfa genomic selection



50 cultivars poly-cross (2015)



Seeded small field
isolated crossing
block (2016)

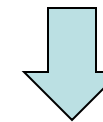


Selected 500
random seeds –
cloned using stem
cuttings (2017-28)

3 replicated, two identical
spaced nurseries at
Saskatoon and Swift Current
(2018-2022)



All plants were sequenced
(1,947 SNPs)



Developed populations based on
genomic model, and conventional
method (Forage yield, plant height,
fiber content)

What we observed:

- Overall genomic prediction was low (i.e. forage yield 0.06-0.13)
 - Multiple parents training population might have reduced GS accuracy
 - Sometimes identical clones showed different phenotypic values at two experimental sites
 - Target environment selection for G x E effect is important (Swift current vs. Saskatoon)
 - Genetic variation reduced because of identical clones
- Prediction was higher when selecting plants using single site data and year as 'environment in G x E

SL935AF Tag name: DM1			SL936AF Tag name: DM2			
Forage yield Saskatoon (Blue)			Forage yield Saskatoon (GEBV)			
Top 30 from two years data			Top 30 from Genomic Selection			
Genotype	BLUE		Genotype	Frequency	GEBV	Seed Wt
Gen_181	360.4		Gen_158	489	219.9	
Gen_8	345.9		Gen_98	489	217.9	
Gen_60	333.8833		Gen_67	464	215.4	
Gen_80	332.0833		Gen_40	445	214.8	
Gen_235	329.2167		Gen_188	435	214.8	
Gen_152	323.5		Gen_374	435	218.4	
Gen_114	323.4667		Gen_143	430	214.1	
Gen_286	318.95		Gen_114	426	214.4	
Gen_300	318.73		Gen_216	424	213.9	
Gen_340	316.4527		Gen_303	415	215.5	
Gen_141	310.6518		Gen_170	408	214.8	
Gen_145	307.9333		Gen_77	397	214.9	
Gen_40	304.3667	20% overlapping	Gen_230	387	215.6	
Gen_126	304.1333		Gen_330	382	214.9	
Gen_364	302.5167		Gen_343	379	214	
Gen_50	302.345		Gen_272	378	214.8	
Gen_1	302		Gen_241	370	213.7	
Gen_29	299.55		Gen_133	365	214.6	
Gen_203	298.5833		Gen_112	363	213.7	
Gen_381	297.4343		Gen_172	362	213.4	
Gen_338	297.2167		Gen_50	357	214	
Gen_65	296.0333		Gen_304	344	214.4	
Gen_62	292.5033		Gen_205	342	213.4	
Gen_11	289.9667		Gen_13	340	213.6	
Gen_247	289.9333		Gen_300	330	213	
Gen_148	288.9167		Gen_373	328	212.8	
Gen_245	288.3		Gen_8	326	212.2	
Gen_135	288.0243		Gen_364	319	212.9	
Gen_242	286.65		Gen_315	311	213	
Gen_390	280.0593		Gen_76	305	214	
187g			132g			

SL937AF			SL938AF Swift Current		
Forage yield Swift Current			Forage yield Swift Current		
Top 30 BLUE over two years data			Top 30 from Genomic selection		
Genotype ID	BLUE	Seed Wt	Genotype	Frequency	GEBV
Gen_382	477.0		Gen_124	493	270.2
Gen_137	463.8		Gen_385	475	271.1
Gen_35	459.3		Gen_157	467	268.4
Gen_105	429.4		Gen_256	464	269.1
Gen_34	419.4		Gen_91	461	268.8
Gen_250	411.4		Gen_179	459	268.6
Gen_210	408.0		Gen_314	446	269.6
Gen_309	408.0		Gen_304	444	269.7
Gen_228	407.1		Gen_263	430	268.5
Gen_226	406.1		Gen_218	423	268.9
Gen_121	403.7		Gen_21	414	269
Gen_302	393.9		Gen_360	414	269.6
Gen_271	393.0		Gen_286	410	269.1
Gen_60	392.9		Gen_186	392	269.4
Gen_91	390.0		Gen_192	392	269.5
Gen_372	389.7		Gen_199	367	268.1
Gen_367	389.6		Gen_152	366	269.4
Gen_154	388.1		Gen_386	366	268.1
Gen_260	382.7		Gen_175	360	269.3
Gen_65	382.3		Gen_72	357	268.4
Gen_373	378.9		Gen_102	356	269.1
Gen_312	372.4		Gen_259	353	269.5
Gen_148	366.0		Gen_136	352	268.2
Gen_23	365.9		Gen_331	348	268.6
Gen_199	365.6		Gen_118	335	268.6
Gen_48	364.6		Gen_154	329	267.4
Gen_109	362.0		Gen_140	322	269.2
Gen_115	361.8		Gen_302	321	268
Gen_89	360.5		Gen_30	310	268.9
Gen_143	358.6		Gen_132	302	268.4

105g

110g

Acknowledgements

