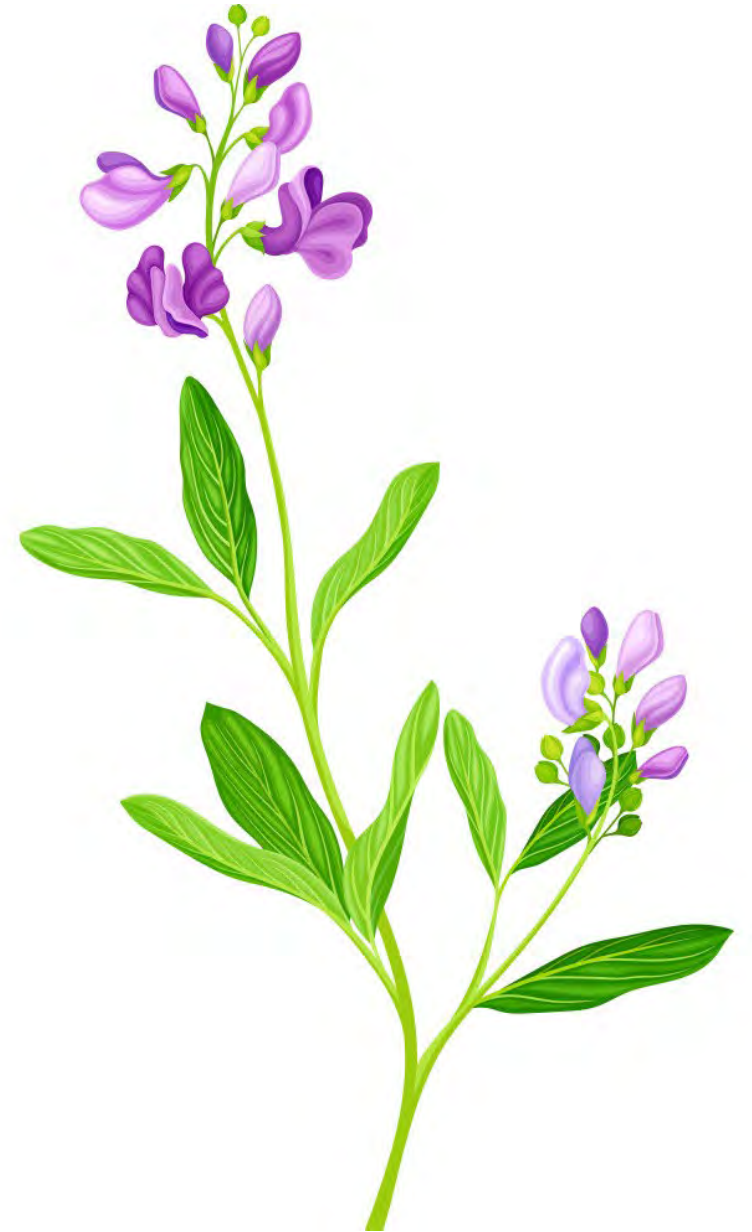


Development of a Universal Training Population System (TPS) to Accelerate Alfalfa Genomic Selection

Cesar Medina

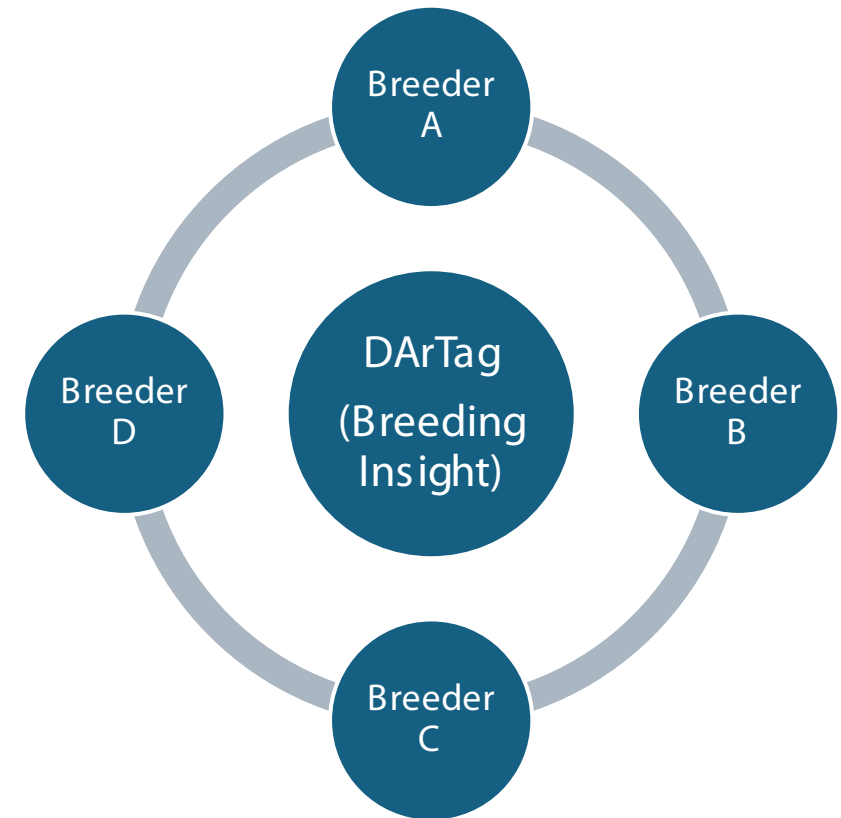
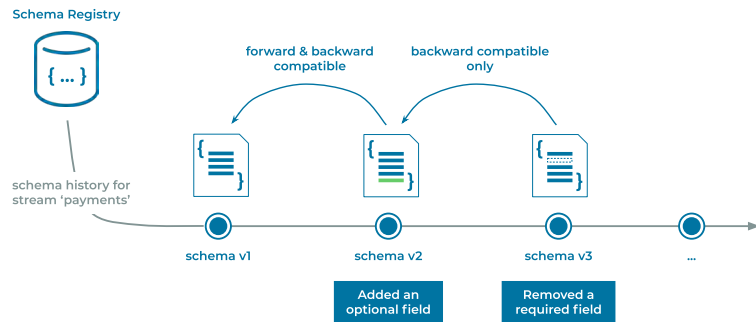
NAAIC

2026-07-13



Will AI replace the breeders?

- In the age of AI, soft skills are assets that are difficult to replace
- As scientists, we still have critical thinking, adaptability, and communication
- This project represents a significant collaborative achievement within the alfalfa community.



DArTag 3k array offers retro-compatibility across all projects

Objective

- To develop a universal Training Population System (TPS) for alfalfa breeding in North America
 - To characterize the genetic diversity and population structure of alfalfa populations to assess genotype redundancy and interconnections among projects genotyped using the DArTag markers.
 - To develop an integrated genomic prediction model that incorporates population structure, quantitative trait effects, and genotype-by-environment interactions.
 - Generate estimated breeding values for key agronomic traits.

Flowchart for generation of TPS

DArTag 3k array offers retro compatibility among the projects



Development of genotyping platform



DArTag file
3,000 target SNPs
~10,000 Microhaplotypes

Population genomics analysis

Unifying datasets by leveraging background knowledge

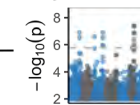
>10,000 alfalfa genotypes



Population structure



Major QTLs

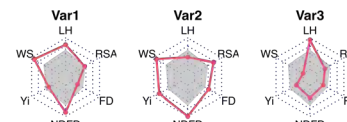


Enviromics (weather, soil)



Training Population System (TPS)

New material ranking



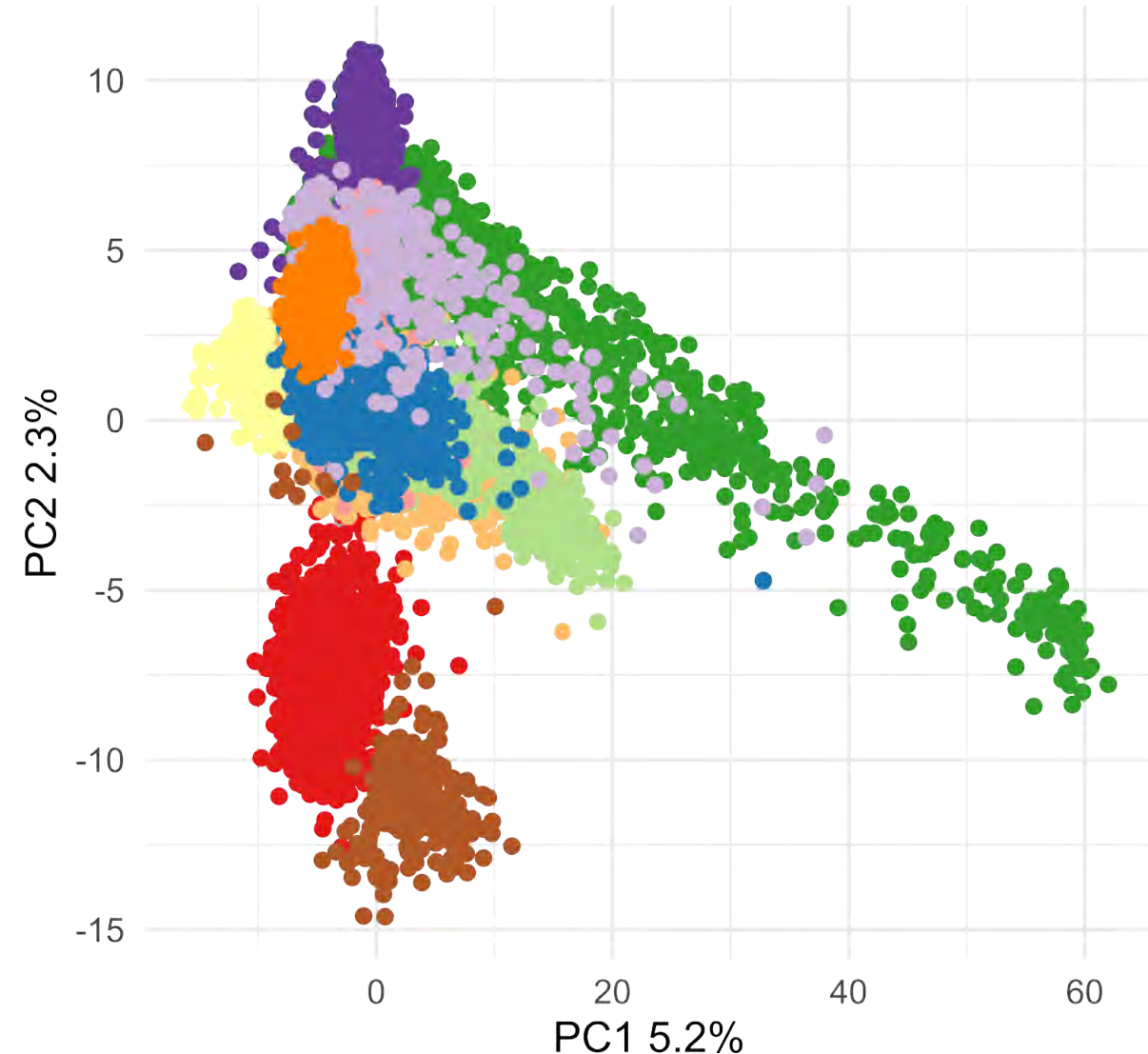
Training population system designed to predict the GEBV of new genotypes

Alfalfa breeding projects included in the TPS

Code	DArT_code	SY	Geno	Pheno	Loc
GSDIG	DAI21-6679	ZX	1,500	500	MN
PI2022	DAI22-7513	ZX	1,070	1,200	MN
RSA	DAI23-7788	ZX	2,200	1,190	MN
WARP21	DAI23-8031	ZX/HR	1,233	1,053	MN
UW drone	DAI23-8688	ZX	1,100	950	MN
Aphan00	DAI25-10215	DS	376	376	MN
BigApp	DAI22-7249	MP	648	NA	UT
Roza2019	DAI22-7535	MW	450	450	WA
CAGE20	DAI22-7011	MW	600	600	WA
pH_tolerance	DAI23-8202	ZX/Alf	345	345	GA
N2 fixation	DAI25-10222	DS/BG	376	376	ND
Aphan01	DAI23-8158	DS	357	357	MN
Total			10,255	7,397	5

- Twelve projects from five locations were selected to build the TPS
- The projects cover more than 10,200 genotypes (10K)
- However, phenotypic data are sparse due to differing original breeding objectives

Population structure by PCA of 10k genotypes



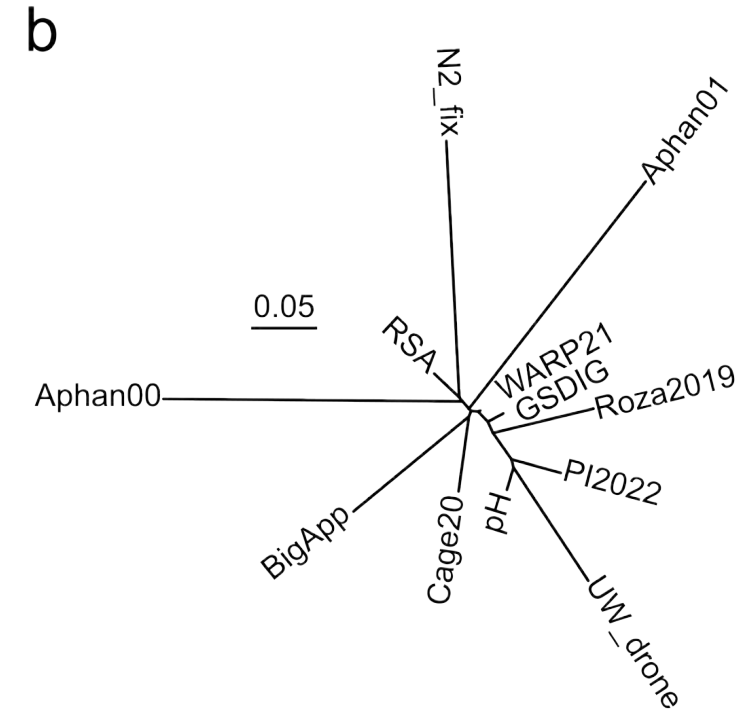
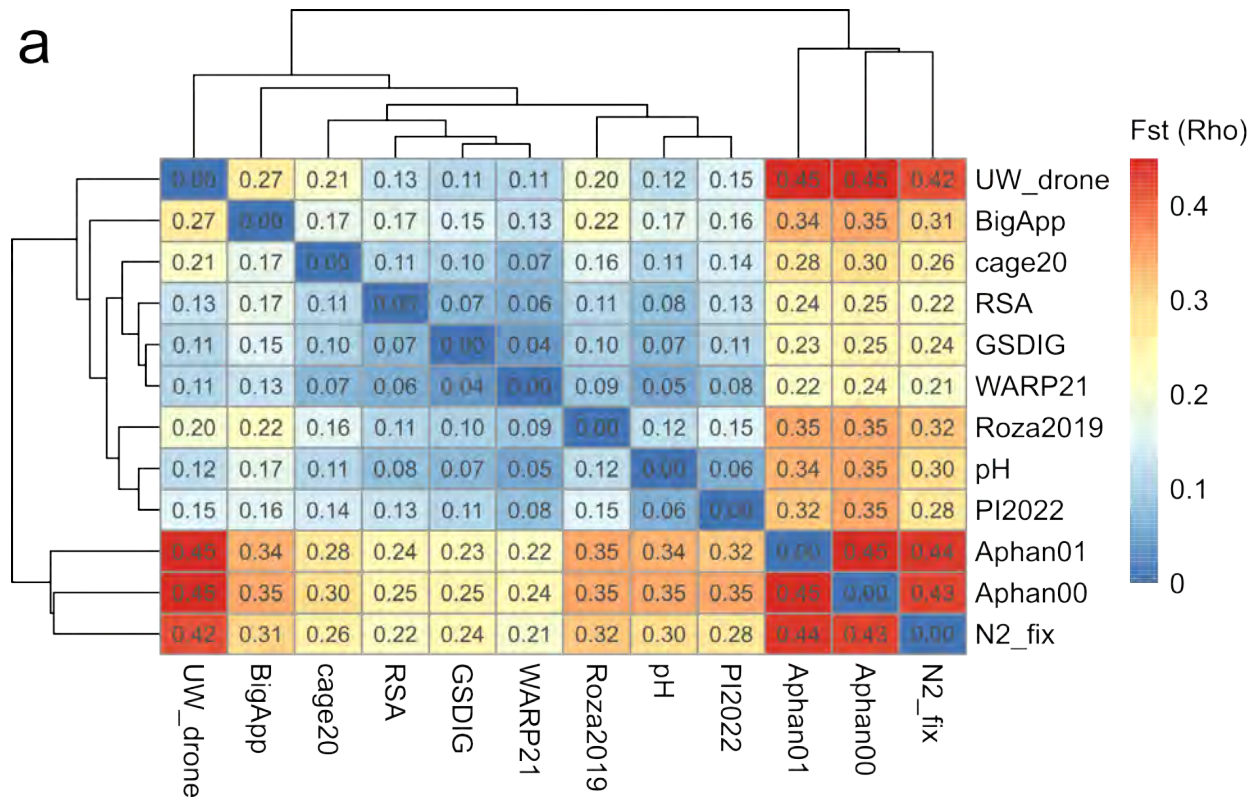
DArT

- DAI21_6679 (GSDIG)
- DAI22_7011 (Cage20)
- DAI22_7249 (BigApp)
- DAI22_7513 (PI2022)
- DAI22_7535 (Roza2019)
- DAI23_7788 (RSA)
- DAI23_8031 (WARP21)
- DAI23_8158 (Aphan00)
- DAI23_8202 (pH)
- DAI23_8688 (UW_drone)
- DAI25_10215 (Aphan01)
- DAI25_10222 (N2_fix)

Project	Ae	Ho	Hs
Aphan00	1.50	0.33	0.29
Aphan01	1.51	0.35	0.29
N2_fix	1.52	0.34	0.30
BigApp	1.54	0.34	0.31
Cage20	1.54	0.34	0.31
GSDIG	1.55	0.35	0.32
WARP21	1.56	0.34	0.32
RSA	1.56	0.35	0.33
Roza2019	1.57	0.39	0.33
PI2022	1.60	0.40	0.34
UW_drone	1.60	0.43	0.34
pH	1.62	0.43	0.35

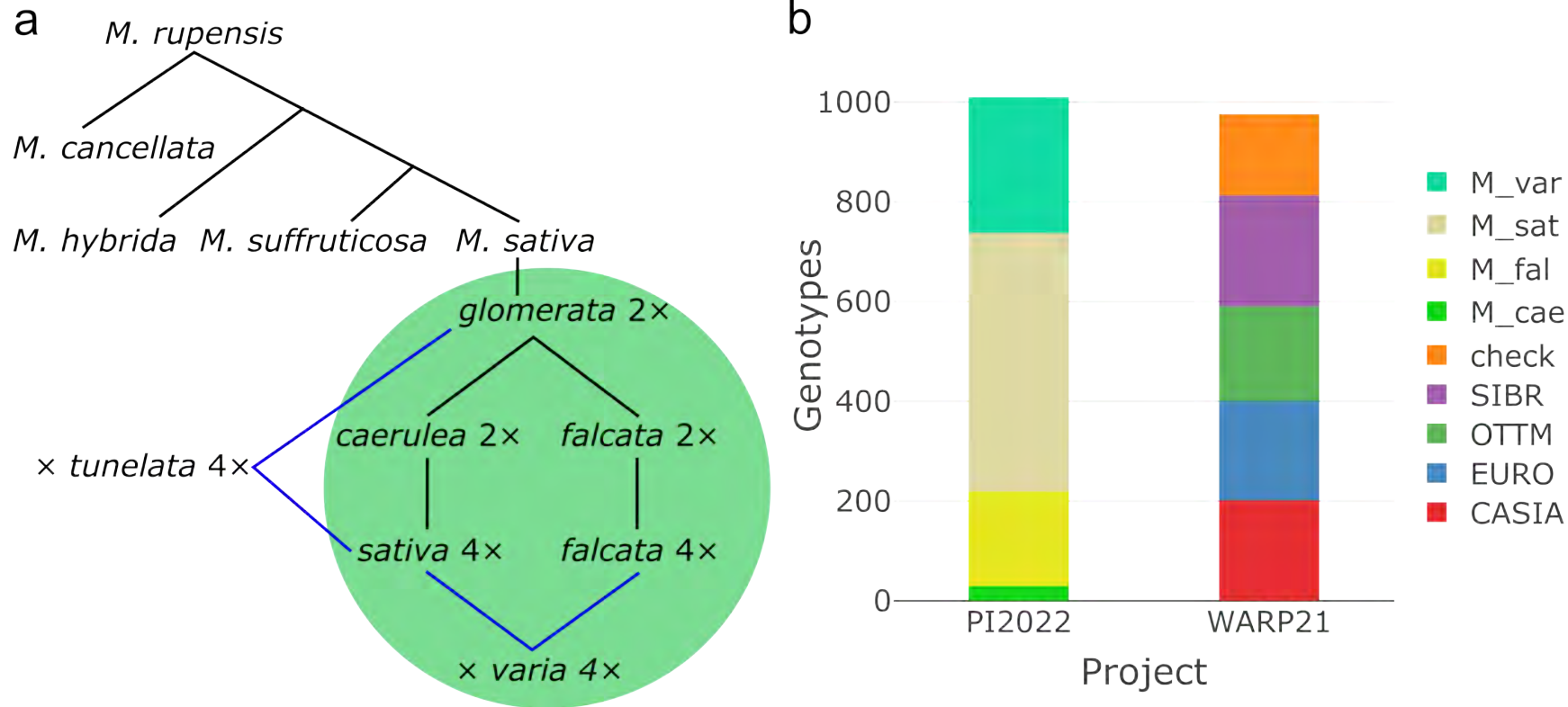
Ae = Effective number of alleles; Ho = Observed Heterozygosity;
Hs = Expected Heterozygosity

Interpopulation genetic distances



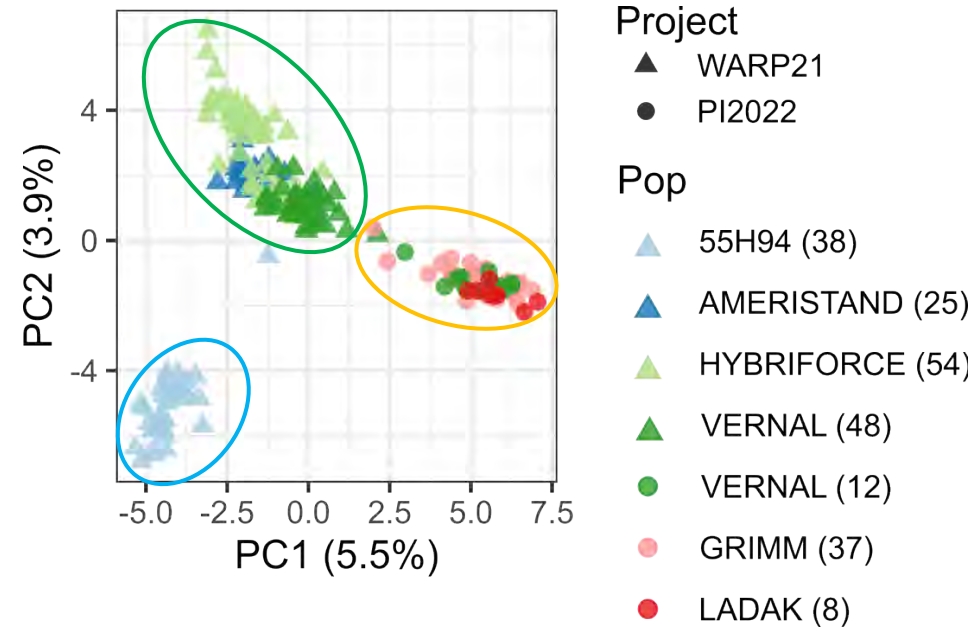
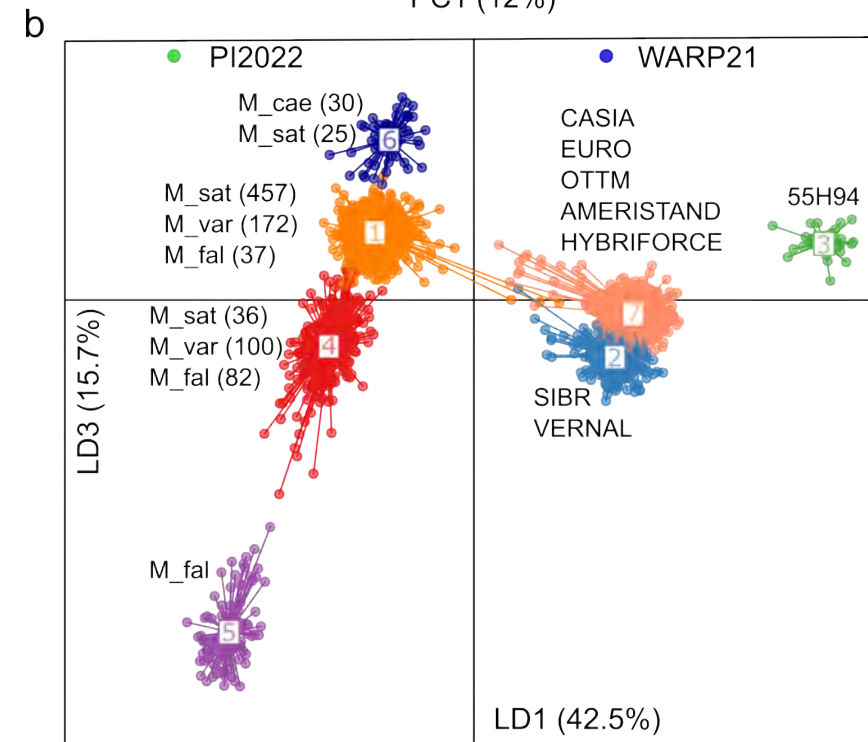
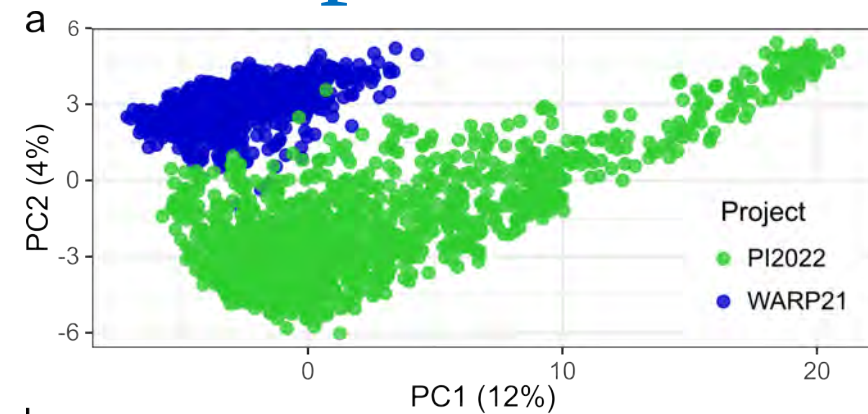
- The Fst matrix displays the genetic distances between projects, while the NJ tree represents the most distant projects.
- Mapping populations (Aphan00, Aphan01, N2_fix, UW_drone) showed the highest Fst values, compared to more diverse populations

Selecting two projects: PI2022 and WARP21



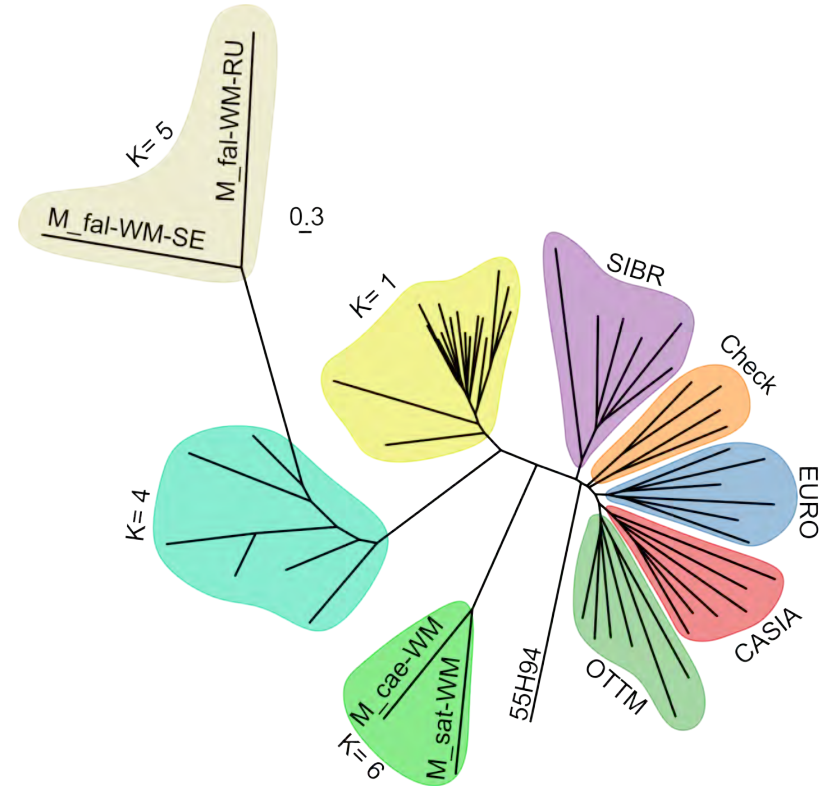
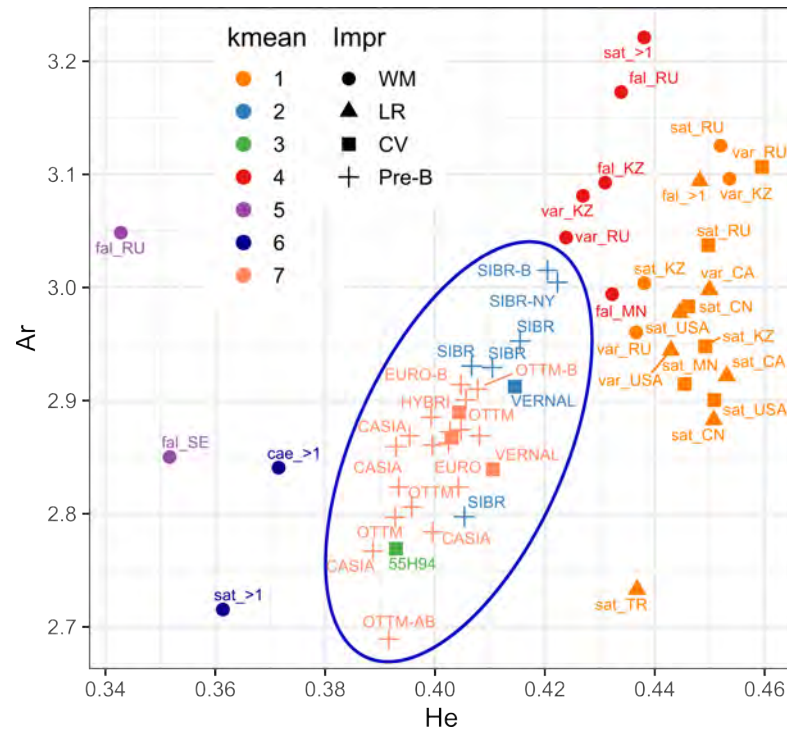
- PI2022 includes 1,010 genotypes from plant introduction (PI) germplasm accessions from the four main taxonomic groups
- WARP21 includes 976 genotypes composed of 24 pre-breeding populations grouped in four genetic pools and four commercial cultivars

Population structure of PI2022 and WARP21



- DAPC separates the two projects along the x-axis and groups the genotypes into seven clusters.
- PCA at cultivar levels indicates that genotypes are grouped into three clusters: 55H94, cultivars from WARP21, and cultivars from PI2022

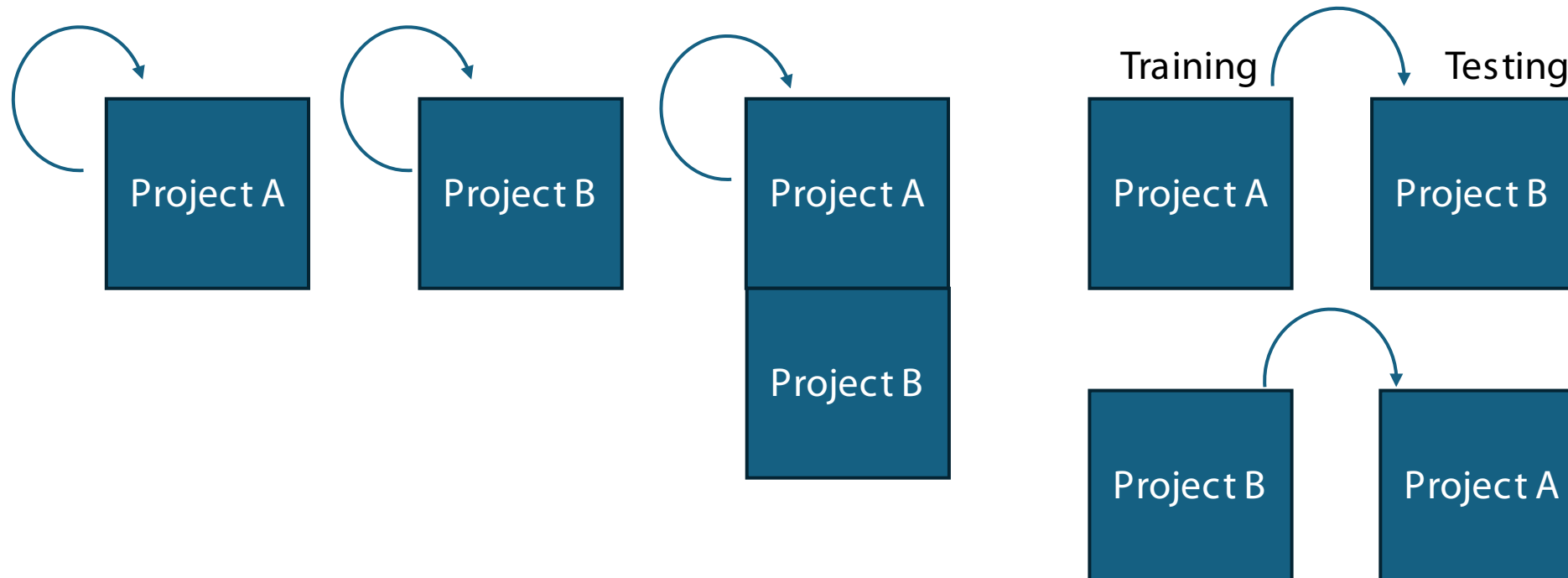
Intra and inter-genetic diversity of PI2022 and WARP21



- Patterns of genetic diversity across 56 alfalfa populations from the PI2022 and WARP21 projects.
- Populations were defined by K-means clustering, improvement level, taxonomy, and country of origin.
- The NJ tree resolves nine distinct clusters that align with K-means clustering and geographical genetic pools.

Genomic prediction approaches for TPS

- PI2022 and WARP21 projects were phenotyped for Arch and PLH
- Approaches: 1) within-project 10-fold cv; 2) Join-project 10-fold cv; 3) across-projects 10-fold cv (Reciprocal Validation Strategy)

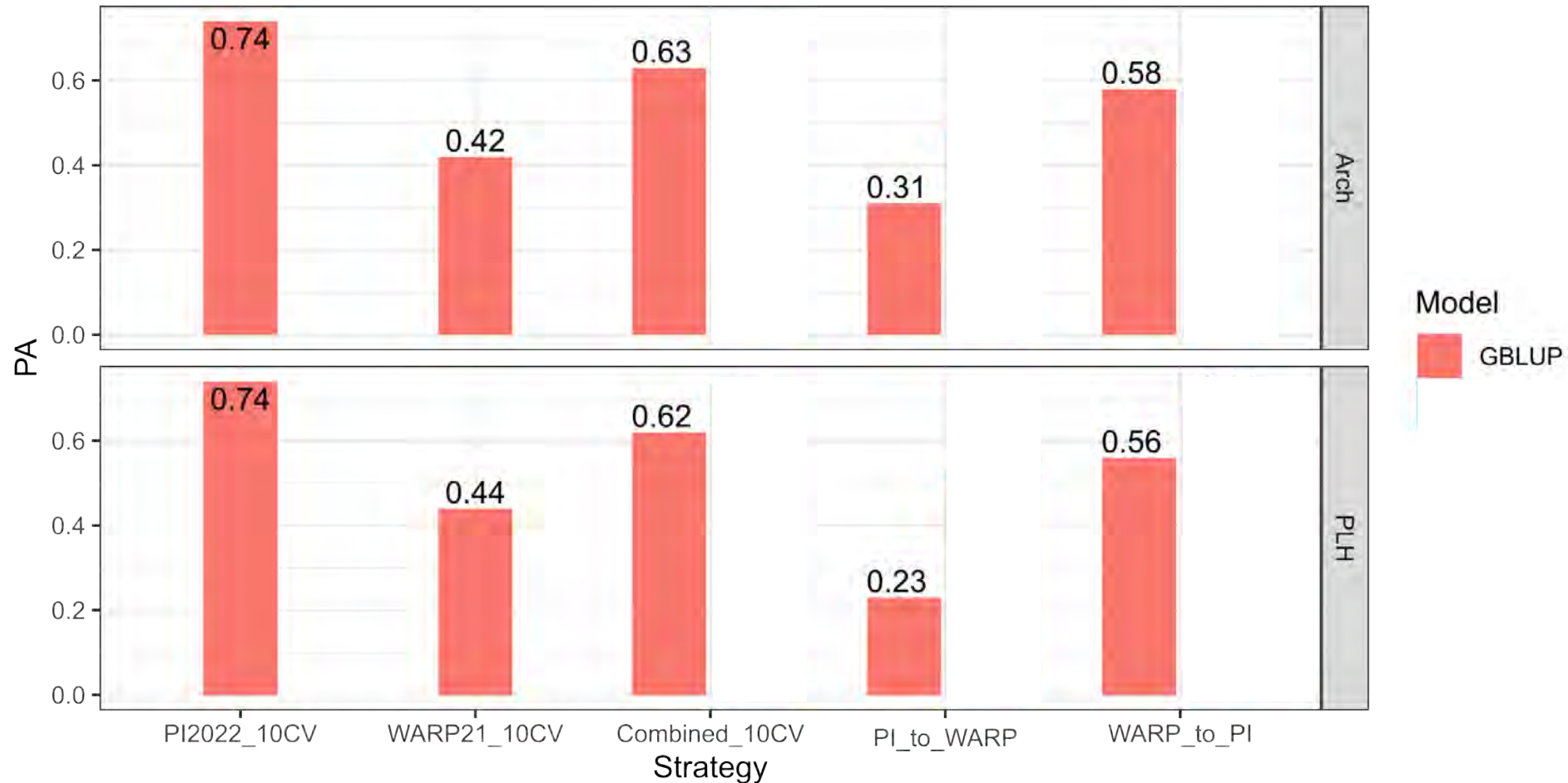


Approaches to increase the PA



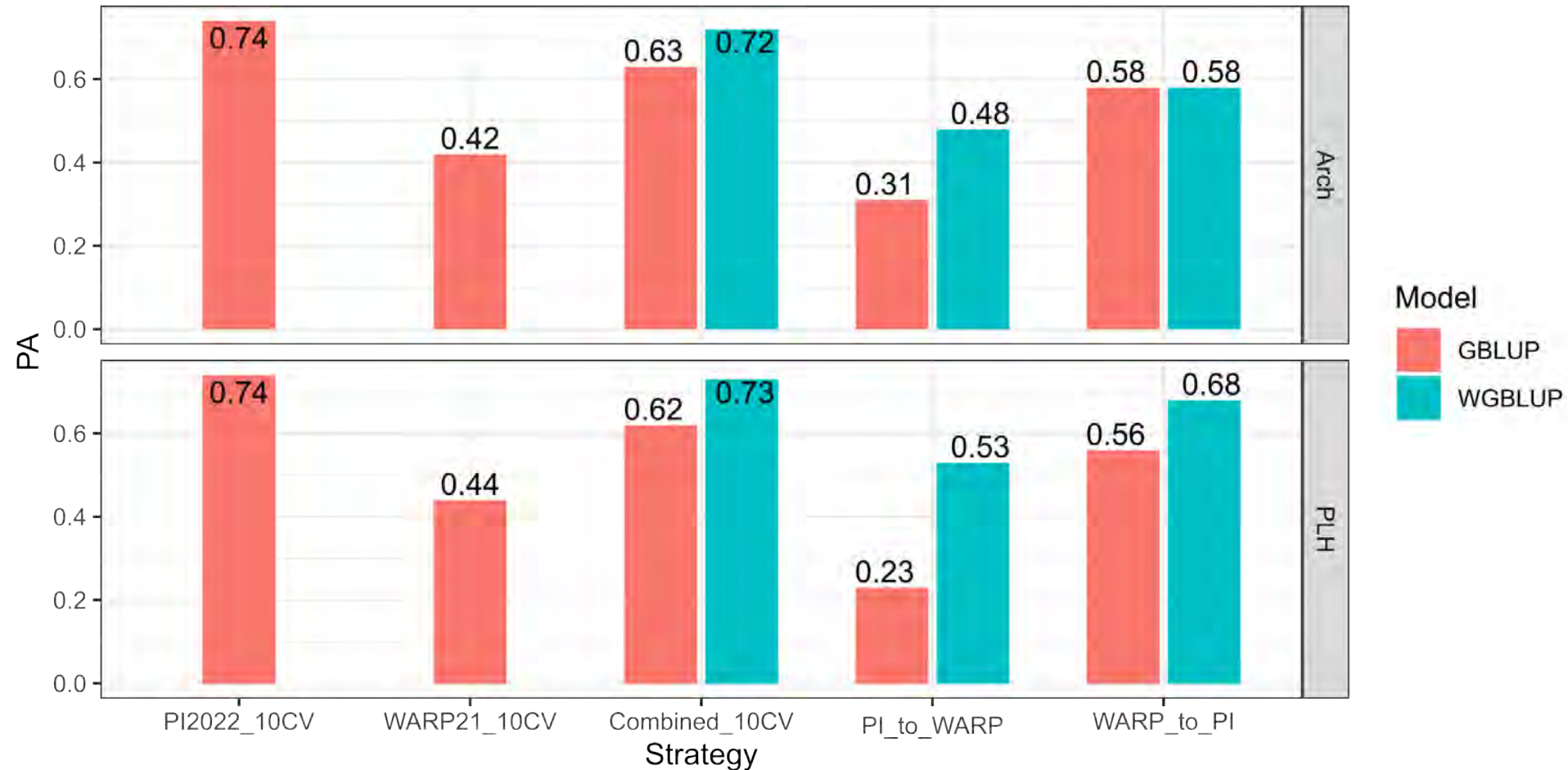
- within-project 10-fold cv: 30% lower PA in WARP (pre-breeding project) vs PI2022
- Join-project 10-fold cv: the PA is higher than the arithmetic mean of the two projects

Approaches to increase the PA



- The Reciprocal Validation Strategy (PI2022 $\rightarrow$ WARP or WARP $\rightarrow$ PI2022) reduces the PA for both traits
- PA in WARP $\rightarrow$ PI2022 were ~30% higher

Approaches to increase the PA



- The use of weights during the construction of the GRM increases the PA
- The Reciprocal Validation Strategy with WGBLUP increases the PA by ~30% for PLH (PI2022 → WARP).

Conclusion and future remarks

- The generation of standardized genetic diversity data will allow us to identify bottleneck effects, sources of genetic diversity, and potential relatedness among breeding projects.
- More than 10,000 genotypes provide a broad landscape of the population structure and genetic diversity within North American alfalfa projects.
- The use of p-values as weights during GRM generation increases the PA by ~30% in a Reciprocal Validation Strategy
- Future TPS efforts will focus on generating GEBVs for other traits (Forage Quality) for validation in new materials.

Thank you!



Ali Missaoui



UNIVERSITY OF MINNESOTA

Tarah Young



Barney Geddes



**Agricultural
Research
Service**

Zhanyou Xu
Deborah Samac
Heathcliffe Riday
Marilyn Warburton
Joseph Robins
Brian Irish



Dongyan Zhao, Meng Lin, Craig T. Beil,
Moirra J. Sheehan



LEGACY
Seeds

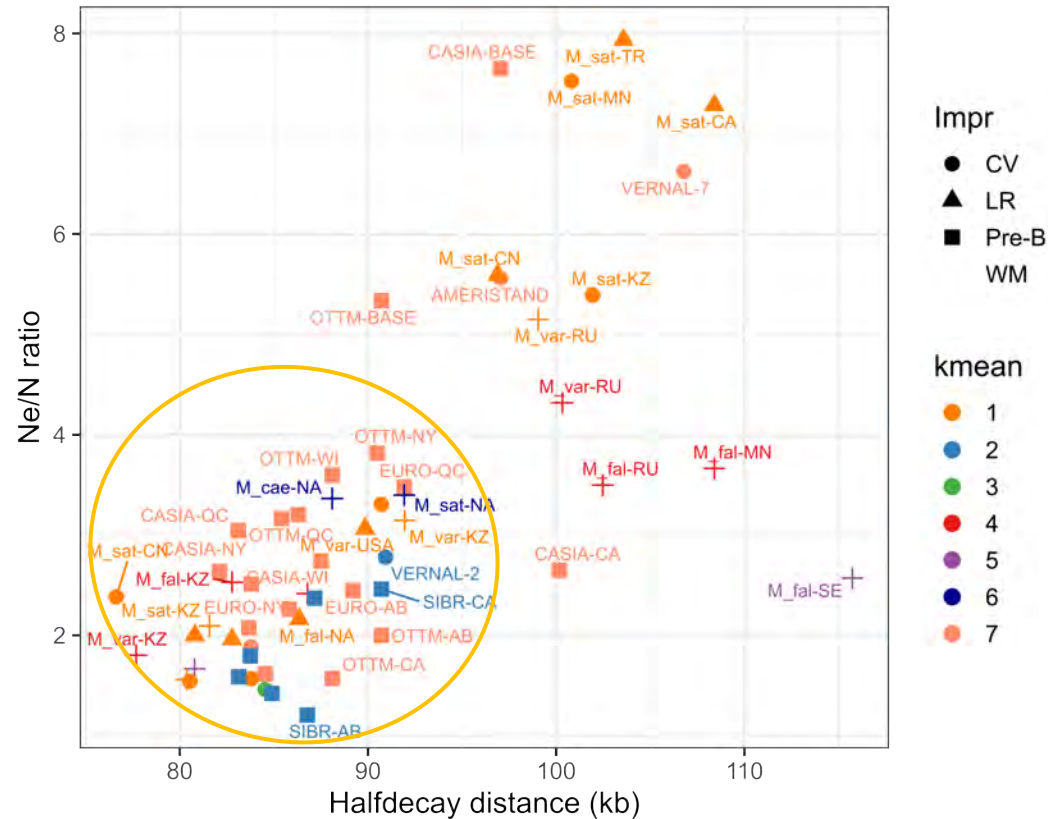
Olivia Steinmetz



Xuefeng Ma

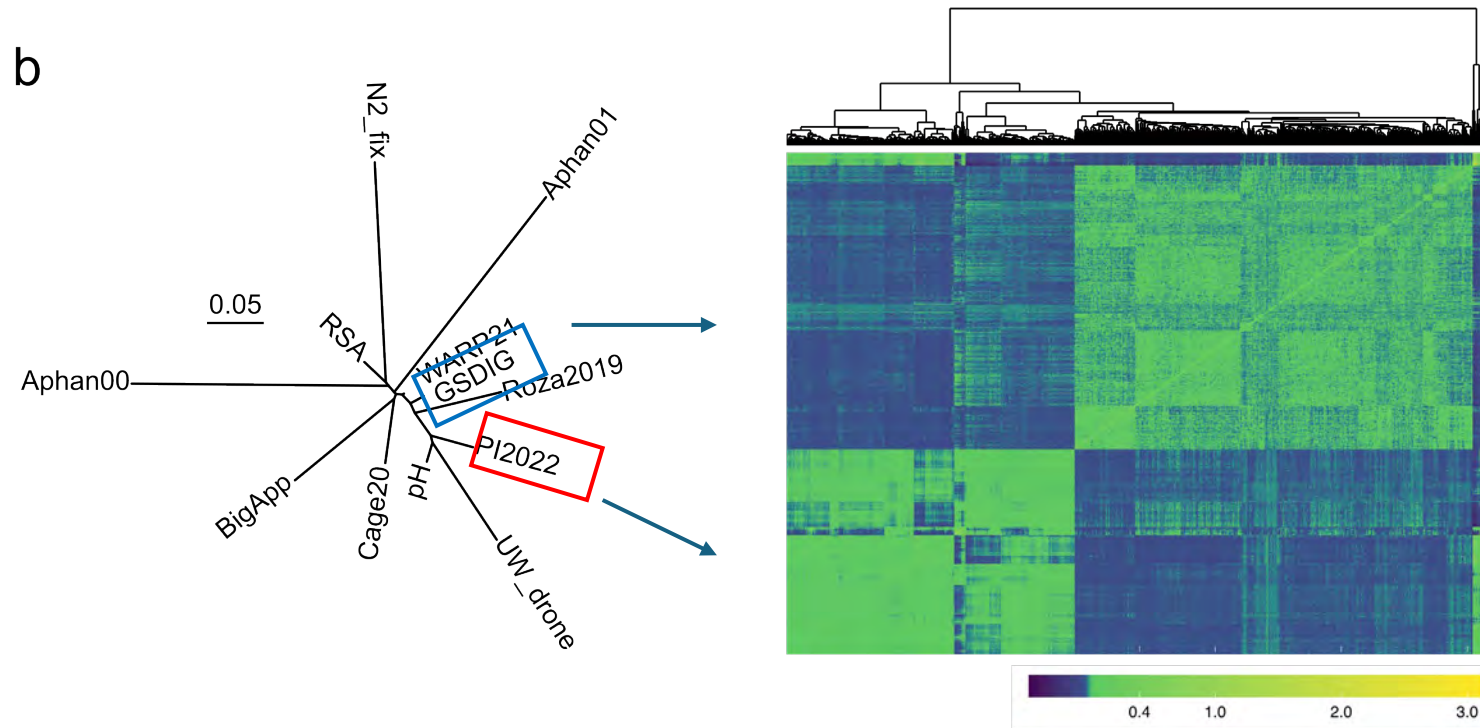
Extra

Intra genetic diversity of PI2022 and WARP21: Ne vs LD decay



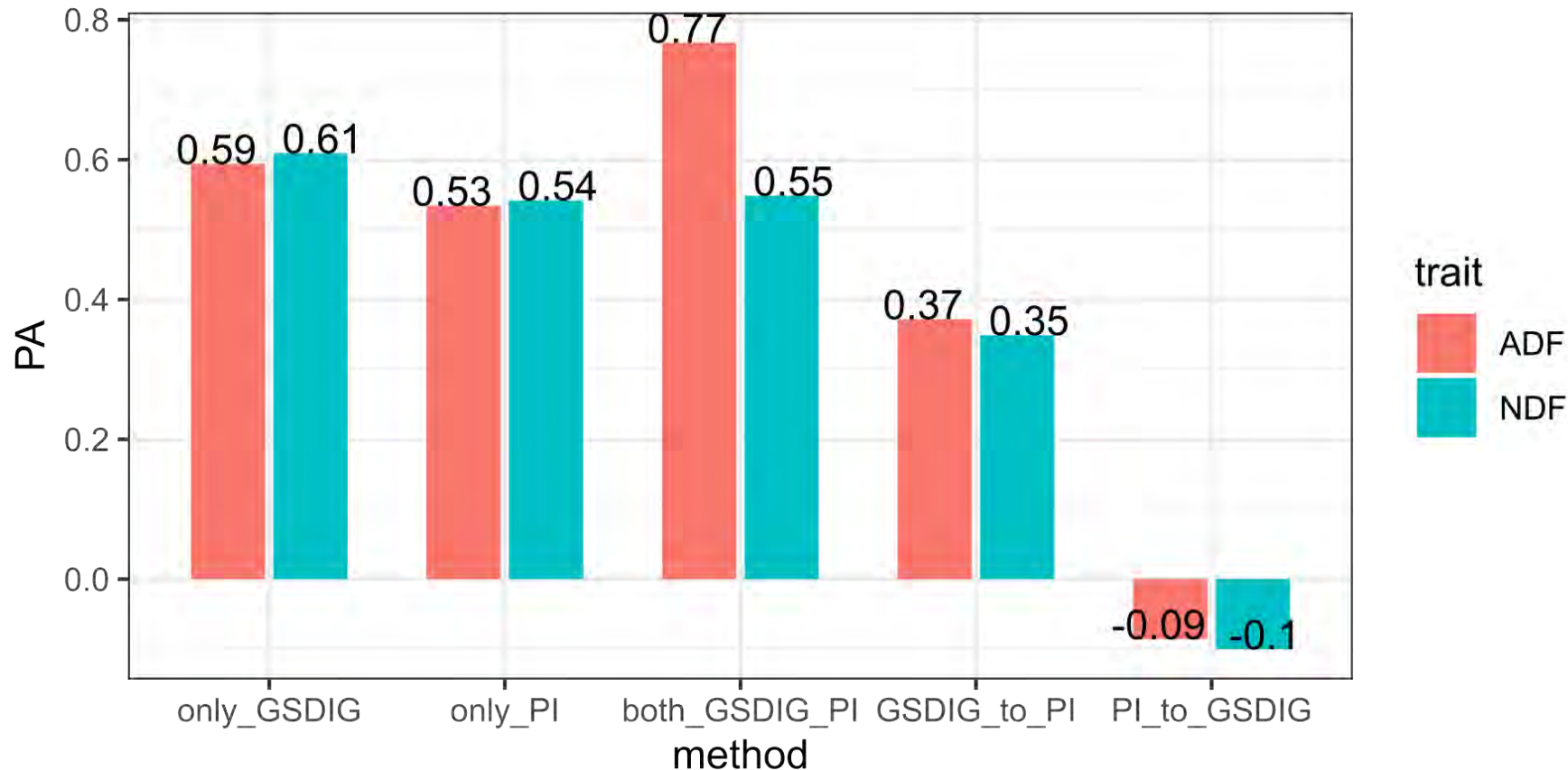
Comparing Ne vs LD half decay distance groups, most of the populations with low Ne and faster LDs, but also identify 14 populations with slow LD decay and/or larger Ne sizes

Is it possible to increase the PA in other traits?



- The F_{st} between GSDIG and PI2022 was 0.11; the comparison was done between a subset of 376 PIs and 489 genotypes of GSDIG
- A clear population structure is evident between the 200 genotypes of SAPA and the 500 genotypes from GSDIG.

Is it possible to increase the PA in other traits?



- All traits exhibited the highest PA under the joint strategy (PI + GSDIG), whereas performance was poor using the Reciprocal Validation Strategy (PI2022 $\rightarrow$ GSDIG).
- These results differ from those observed for Arch and PLH when comparing WARP and PI2022.