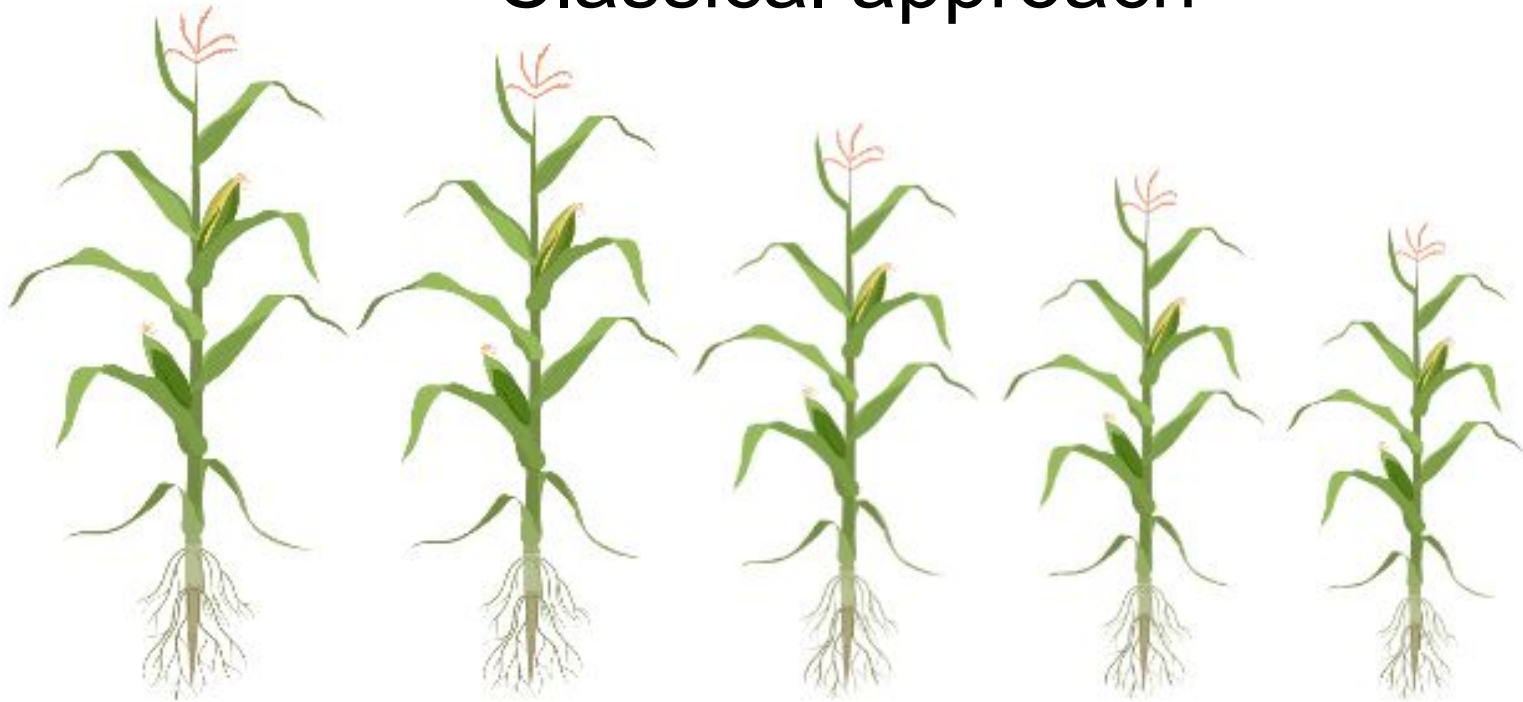# Breeding Program for Maize Improvement

Team members:
Vasiliy Zubarev
Vladislav Mityukov

Classical approach

Heterozygous source population

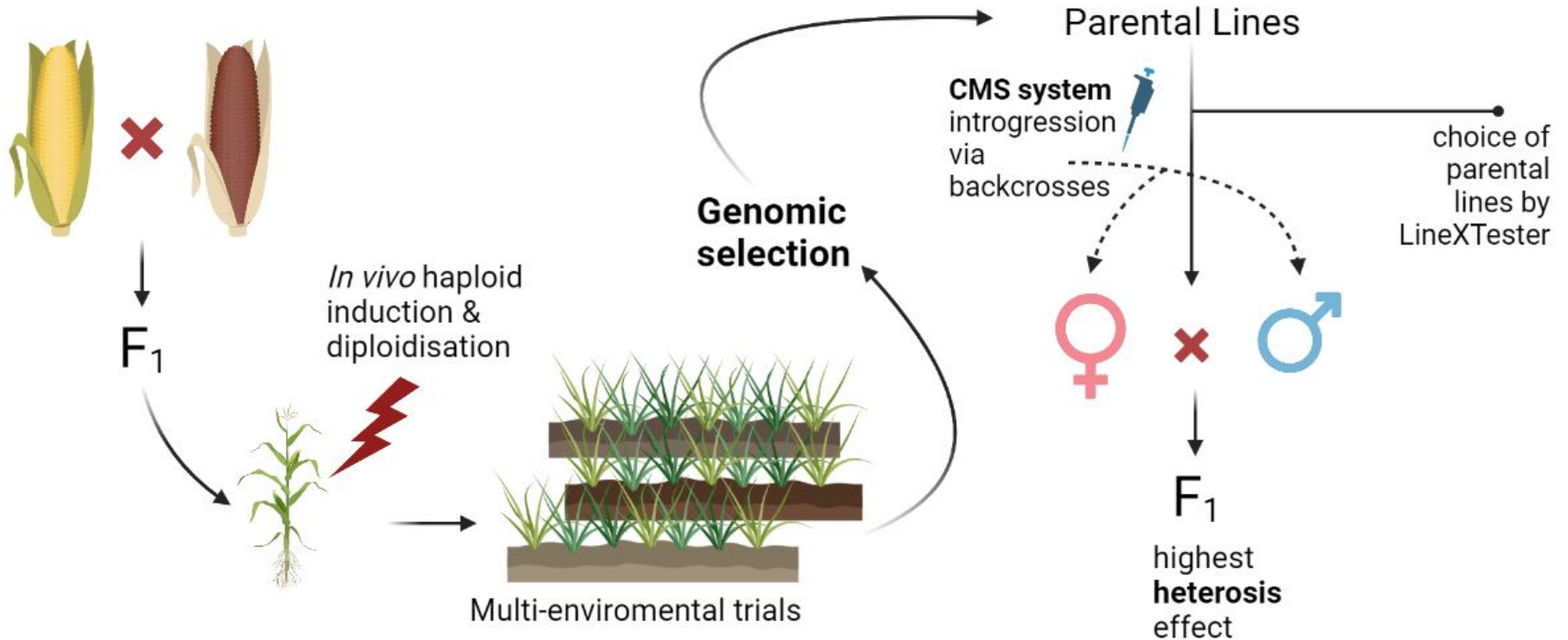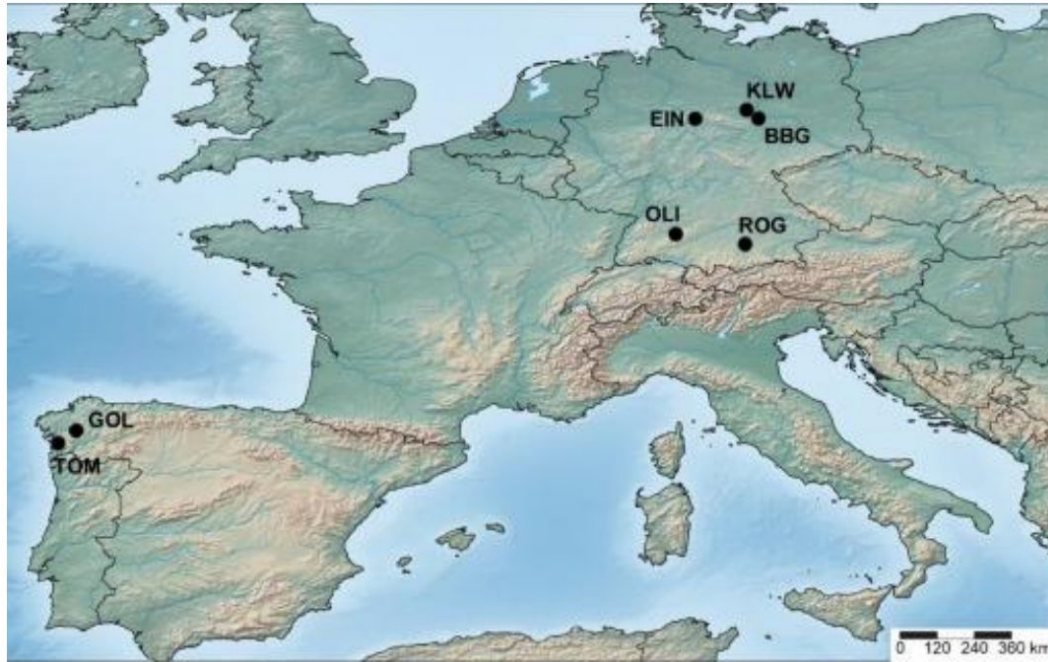N rounds of selfing

~90 % Homo inbred line

Our solution

For speeding up selection process

Breeding scheme

Multi-Environmental Trial

Locations of trials: Bernburg (BBG), Einbeck (EIN), Golada (GOL), Klein Wanzleben (KLW), Oberer Lindenhof (OLI), Roggenstein (ROG), Tomeza (TOM).

Phenotypic varibles measured:

EV_V4, EV_V6 - "ealy vigor" scores at V4 and V6 stages

PH_V4, PH_V6 - plant height at the same stages

PH_final - final plant height

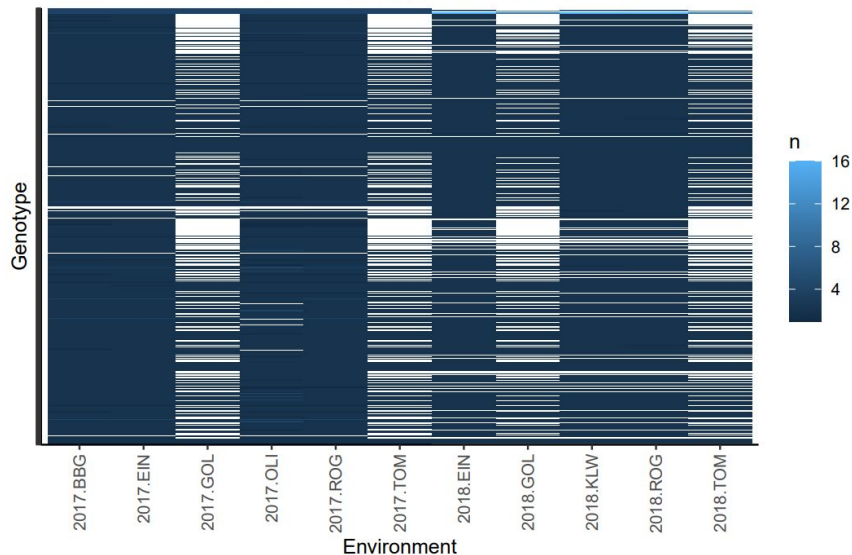**MF, FF** - days to male and female flowering respectively

LO - lodging score

TILL - tillering score

5

(In)completeness of the data

Number of observations per genotype and environment

There is a limited set of lines tested in some locations (especially GOL and TOM)

NA fraction per place, year and parameter
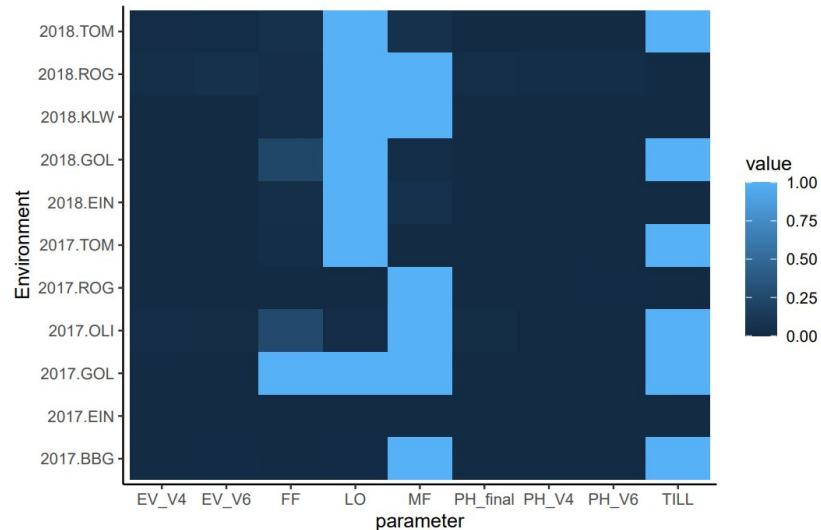
Some variables are totally not measured in some locations (in light blue)

Inter-relationships in the variables set
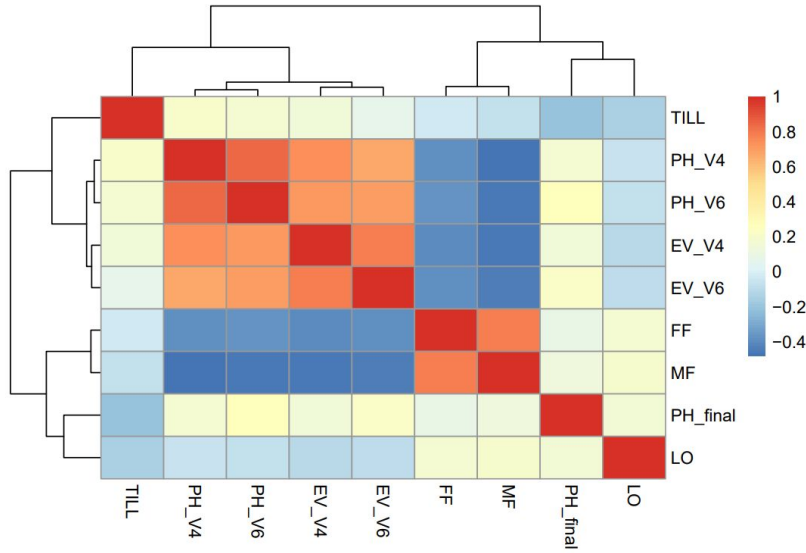
Two groups are highly correlated:

A) Variables representing early growth rate (EV_V4/6, PH_V4/6)
B) Time to flowering (MF, FF)
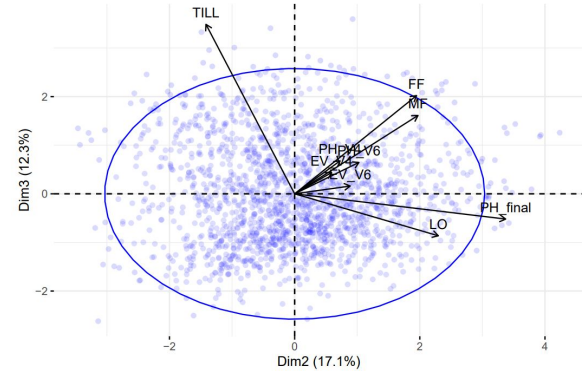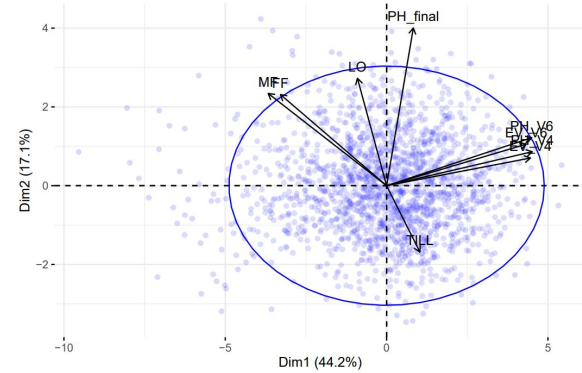
Best lines selection

We applied Finlay-Wilkinson model:

$Y=Env+Genotype+Genotype:Env\_mean+Block+error,$

where Y can be one of: Early vigor, Flowering time, Lodging score.

Genotype - environment interaction in the traits of interest
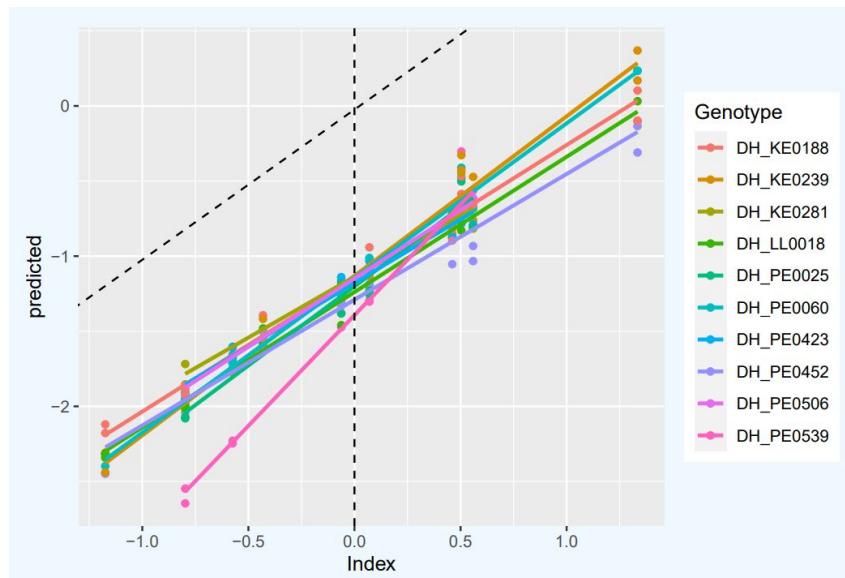
Finlay-Wilkinson regression lines for 10 lines with earliest flowering
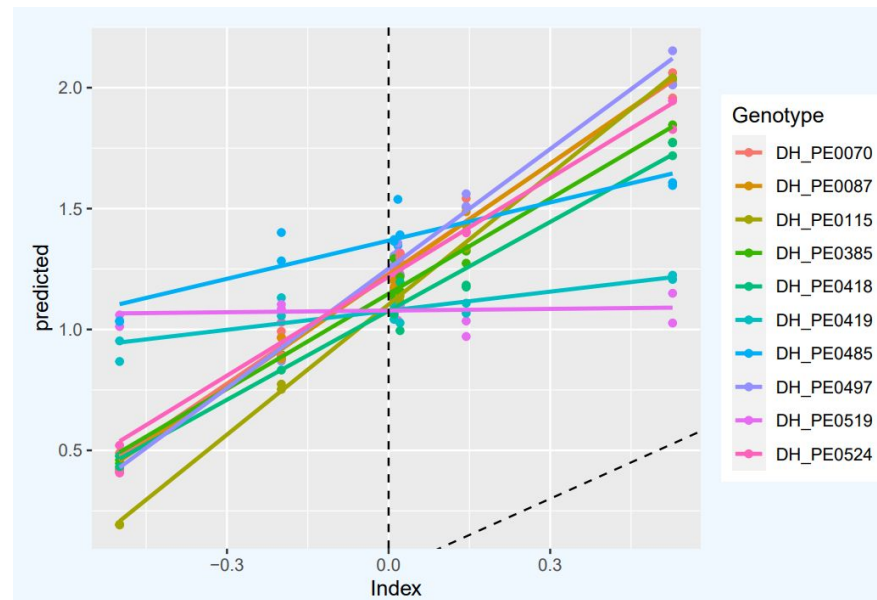
Finlay-Wilkinson regression lines for 10 lines with highest early vigor.

LineXtester

LineXtester

ANALYSIS LINE x TESTER: yield

ANOVA with parents and crosses
================================

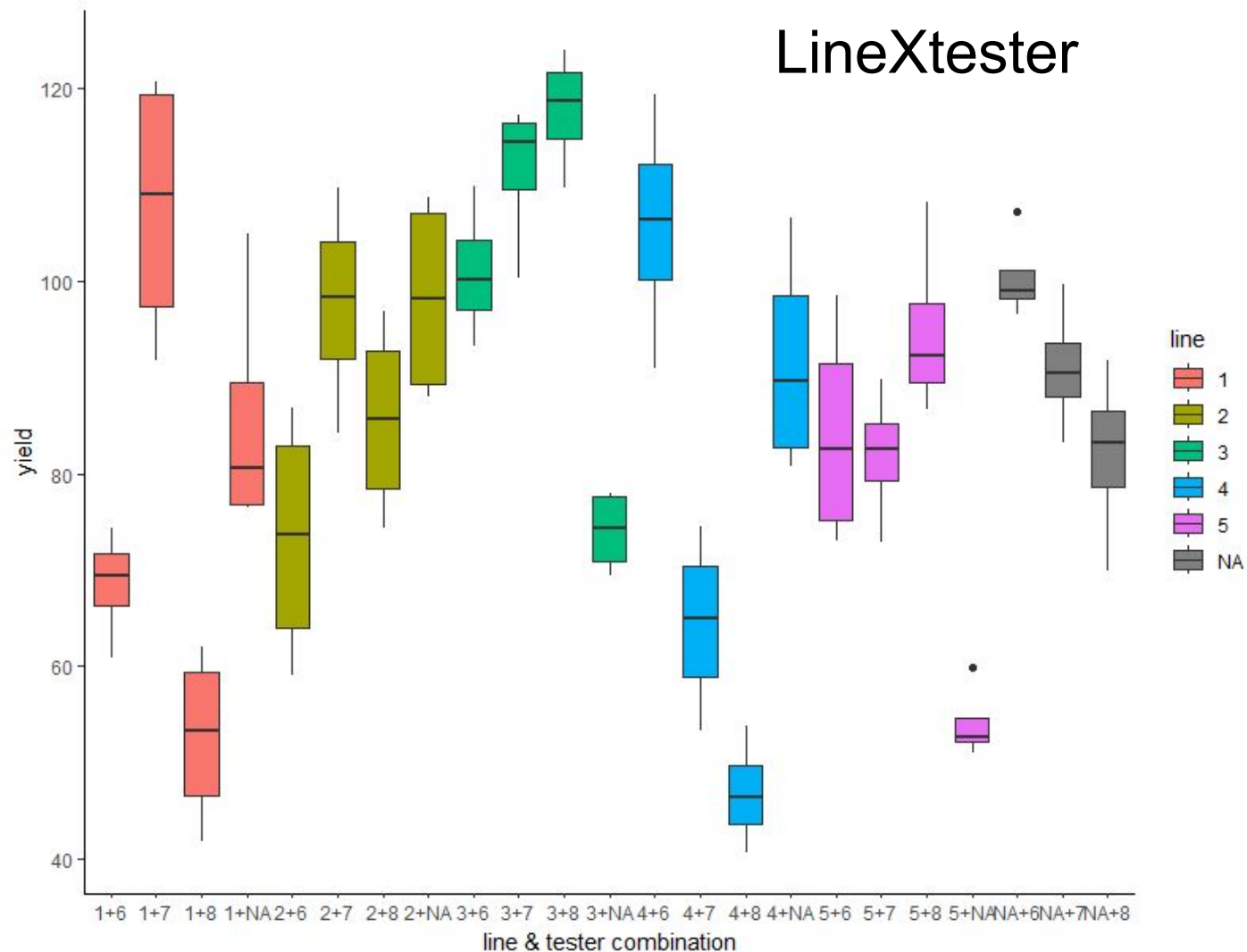|	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Replications	3	83.00012	27.66671	0.304	0.8224
Treatments	22	32553.20239	1479.69102	16.249	0.0000
Parents	7	6299.88519	899.98360	9.883	0.0000
Parents vs. Crosses	1	53.66287	53.66287	0.589	0.4455
Crosses	14	26199.65433	1871.40388	20.551	0.0000
Error	66	6010.03298	91.06111		
Total	91	38646.23549			

ANOVA for line X tester analysis
================================

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lines	4	10318.361	2579.59035	1.457	0.3009
Testers	2	1718.926	859.46289	0.485	0.6327
Lines X Testers	8	14162.367	1770.29590	19.441	0.0000
Error	66	6010.033	91.06111		

GCA Effects:
===========
Lines Effects:

1	2	3	4	5
-9.960	-0.718	23.817	-13.870	0.732

Testers Effects:

6	7	8
0.292	6.404	-6.697

SCA Effects:
===========

Lines	Testers 6	7	8
1	-8.019	24.959	-16.940
2	-12.546	5.717	6.828
3	-9.461	-4.918	14.378
4	33.136	-14.321	-18.815
5	-3.111	-11.438	14.548

ANOVA for line X tester analysis including parents
==

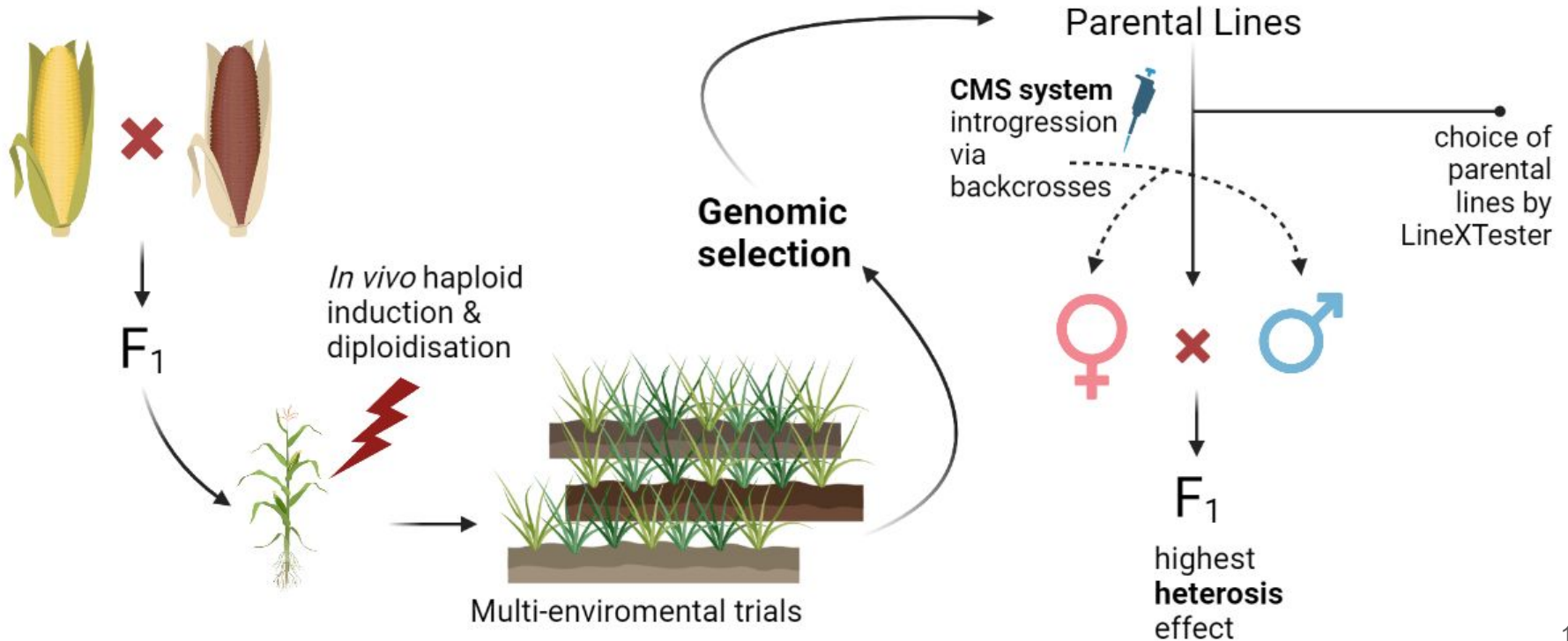|	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Replications	3	83.00012	27.66671	0.304	0.822
Treatments	22	32553.20239	1479.69102	16.249	0.000
Parents	7	6299.88519	899.98360	9.883	0.000
Parents vs. Crosses	1	53.66287	53.66287	0.589	0.445
Crosses	14	26199.65433	1871.40388	20.551	0.000
Lines	4	10318.36140	2579.59035	1.457	0.300
Testers	2	1718.92577	859.46289	0.485	0.632
Lines X Testers	8	14162.36716	1770.29590	19.441	0.000
Error	66	6010.03298	91.06111		
Total	91	38646.23549			

Certification of the best varieties

Steps:

1. Application
2. Verification of application
3. Verification of new variety (small-scale trials)
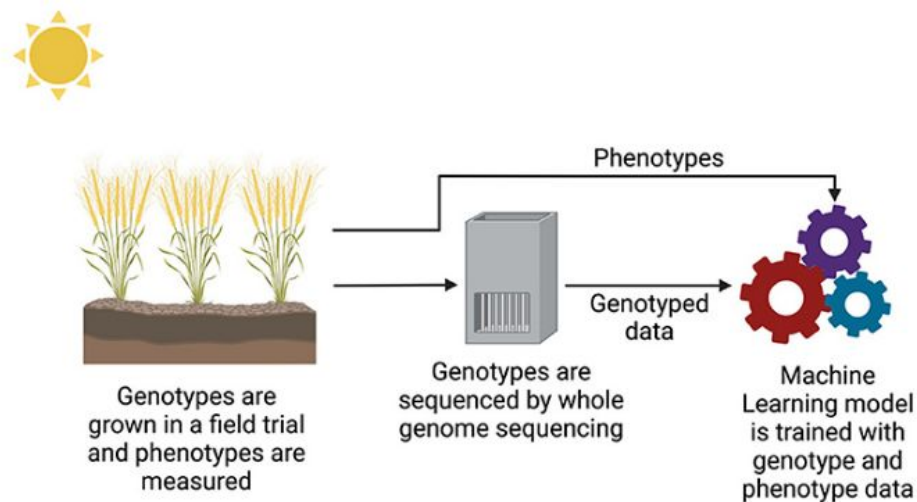4. State register certification of new plant variety

Properties of new variety/hybrid:

- novel (the seeds not sold or transferred to others for use less than 1 year in Russia before the date of application for a patent)
- clear breeding achievement in comparison with any other well-known variety existing at the time
- uniform (homogeneous)
- stable in various environmental conditions
- stable in main traits (they should stay unchanged and not disappear)
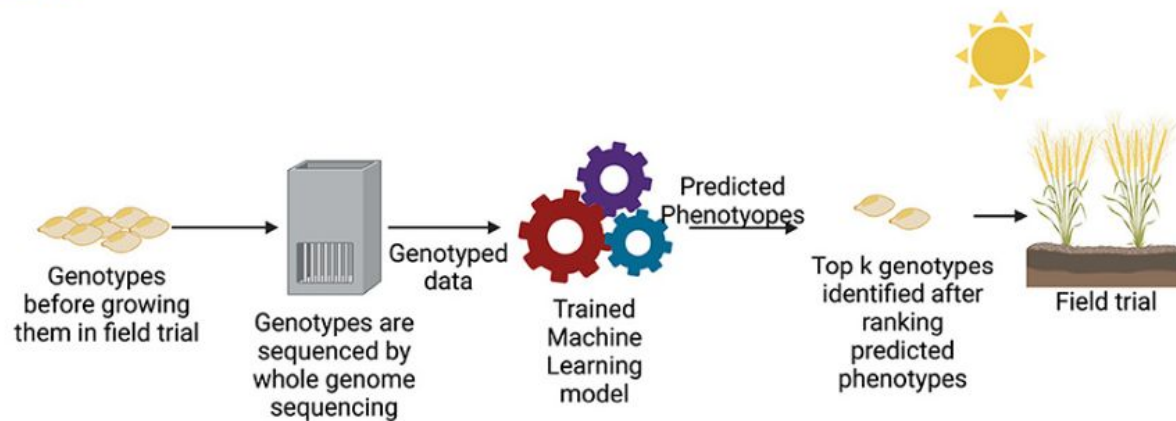- unique name for commercialization
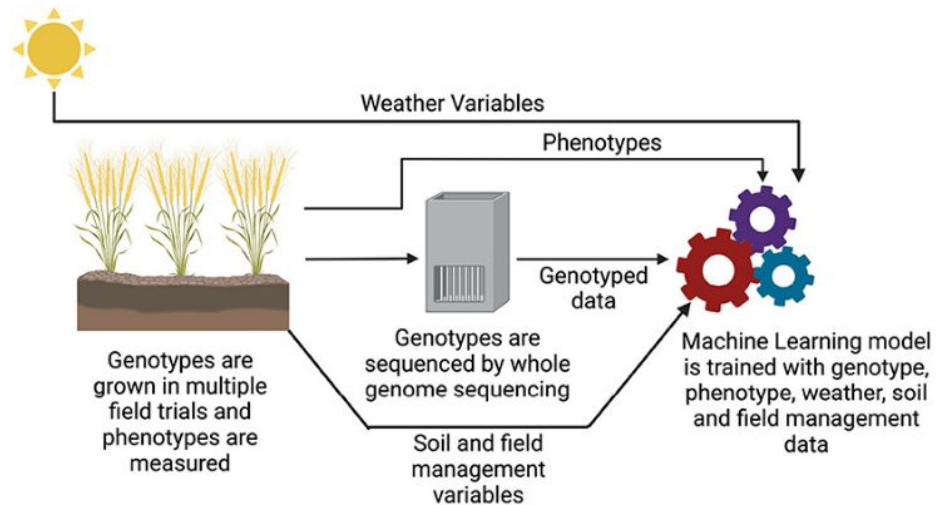
Breeding scheme MVP

Training Phase

Phenotypes

Genotypes are grown in a field trial and phenotypes are measured

Genotypes are sequenced by whole genome sequencing

Genotyped data

Machine Learning model is trained with genotype and phenotype data

Test Phase

Genotypes before growing them in field trial

Genotypes are sequenced by whole genome sequencing

Genotyped data

Trained Machine Learning model

Predicted Phenotyopes

Top k genotypes identified after ranking predicted phenotypes

Field trial

Training Phase

Weather Variables

Phenotypes

Genotyped data

Genotypes are grown in multiple field trials and phenotypes are measured

Genotypes are sequenced by whole genome sequencing

Soil and field management variables

Machine Learning model is trained with genotype, phenotype, weather, soil and field management data

Test Phase

Weather variables

Genotypes before growing them in field trial

Genotypes are sequenced by whole genome sequencing

Genotyped data

Trained Machine Learning model

Predicted Phenotyopes

Top k genotypes identified after ranking predicted phenotypes in a new environment

Field trial in a new environment

Soil and field management variables

Population improvement | Product development

Parents
32-128

Candidates
F1 640 / F2 10240

DH lines
N_2

Year — Parents 50

1/2 — DH lines N_1

3 — Headrows N_1

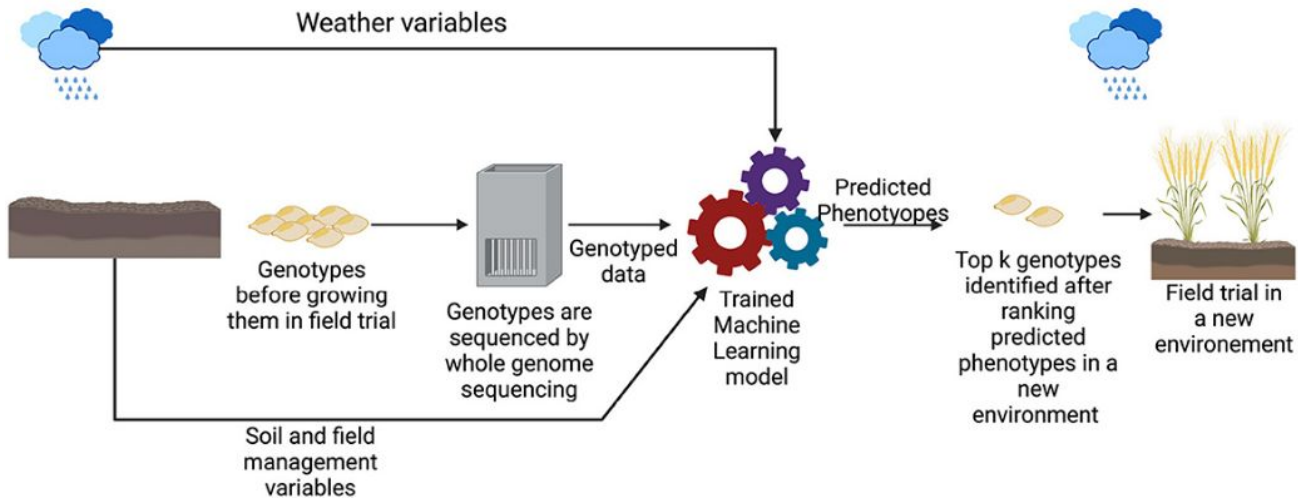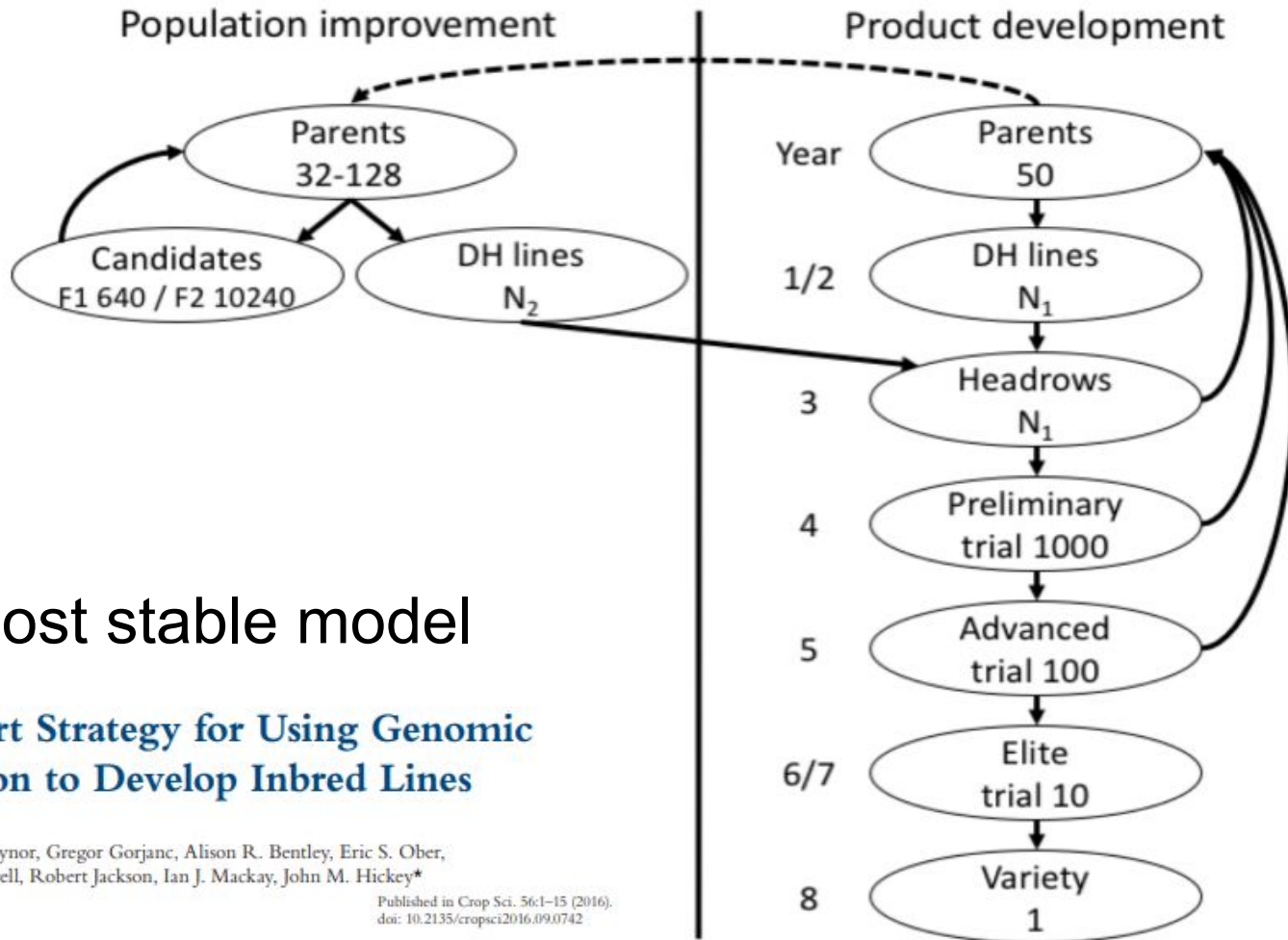4 — Preliminary trial 1000

5 — Advanced trial 100

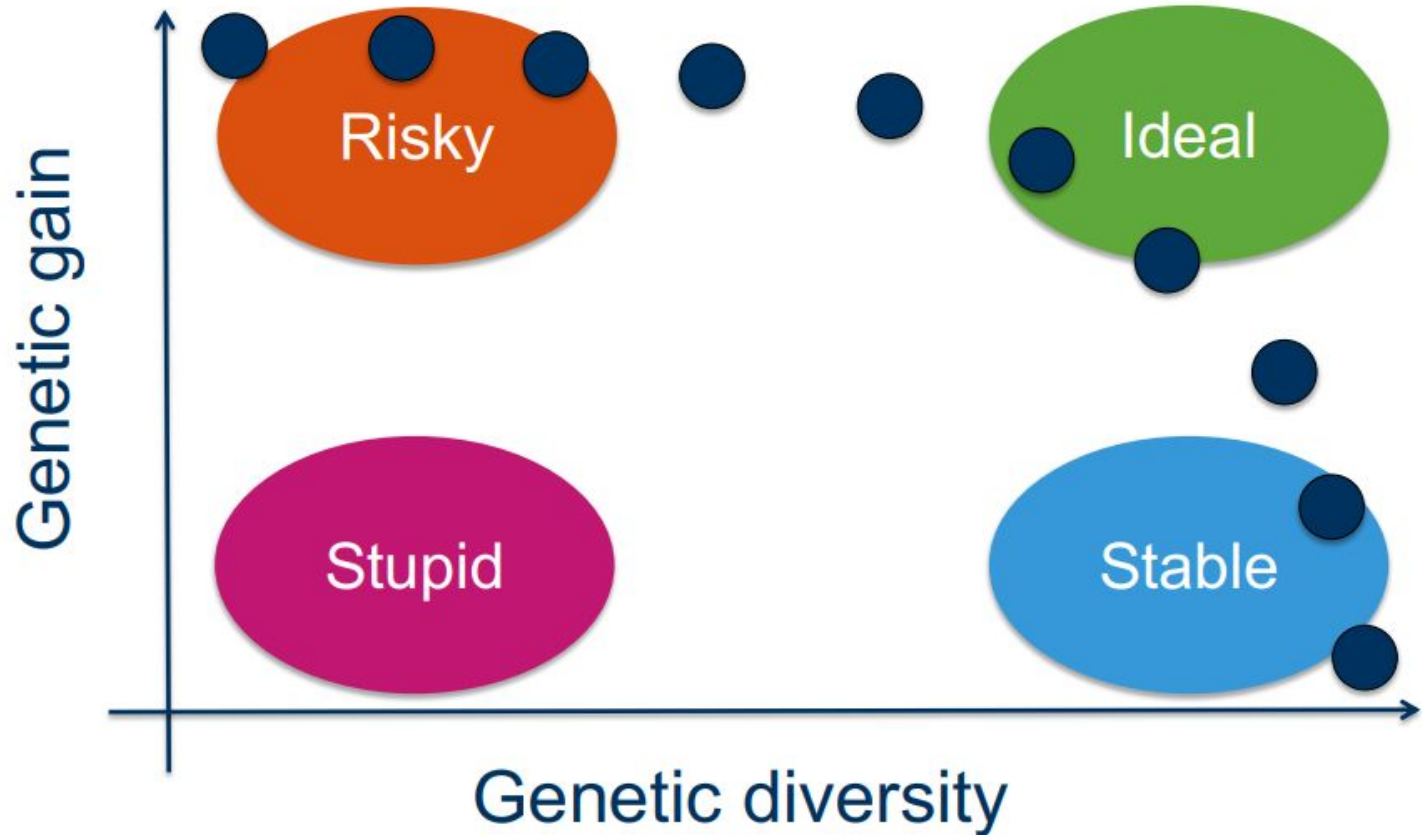6/7 — Elite trial 10

8 — Variety 1

The most stable model

A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines

R. Chris Gaynor, Gregor Gorjanc, Alison R. Bentley, Eric S. Ober, Phil Howell, Robert Jackson, Ian J. Mackay, John M. Hickey*

Breeder's dilemma

Further improvement

lauzingaretti/
DeepGP

A tool to implement Genomic Prediction
Experiments using Deep Learning

2 Contributors 1 Issue 13 Stars 8 Forks

frahik/**BMTME**

Bayesian Multi-Trait Multi-Environment for genomic
selection[R package] [Dev version]

3 Contributors 2 Issues 12 Stars 9 Forks

YinLiLin/**KAML**

Kinship Adjusted Multi-Loci Best Linear
Unbiased Prediction

4 Contributors 11 Issues 36 Stars 11 Forks

PyBrOpS: a Python package for breeding program simulation and optimization for multi-objective breeding

Robert Z. Shrote, Addie M. Thompson

doi: https://doi.org/10.1101/2023.02.10.528043

This article is a preprint and has not been certified by peer review [what does this mean?].

Posted February 13, 2023.

Download PDF
Print/Save Options

Email
Share
Citation Tools

Crop Yield Prediction Using Deep Neural Networks

Saeed Khaki[*] and Lizhi Wang

Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, United States

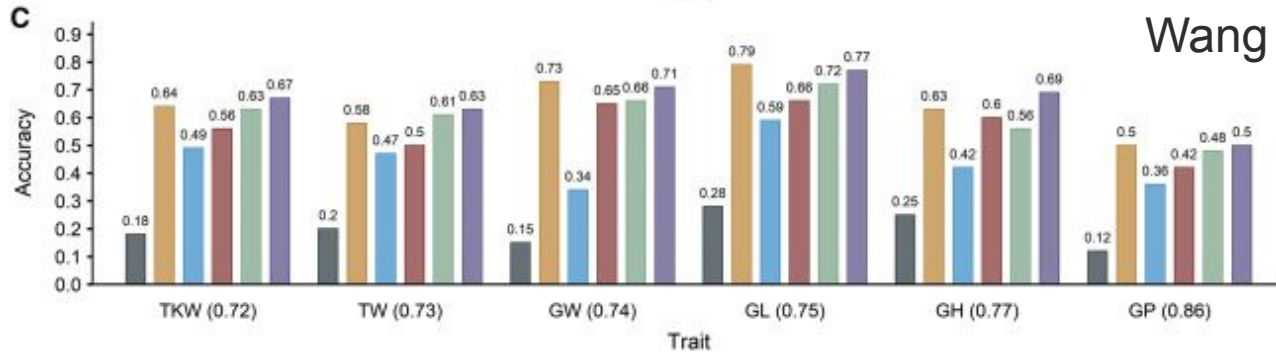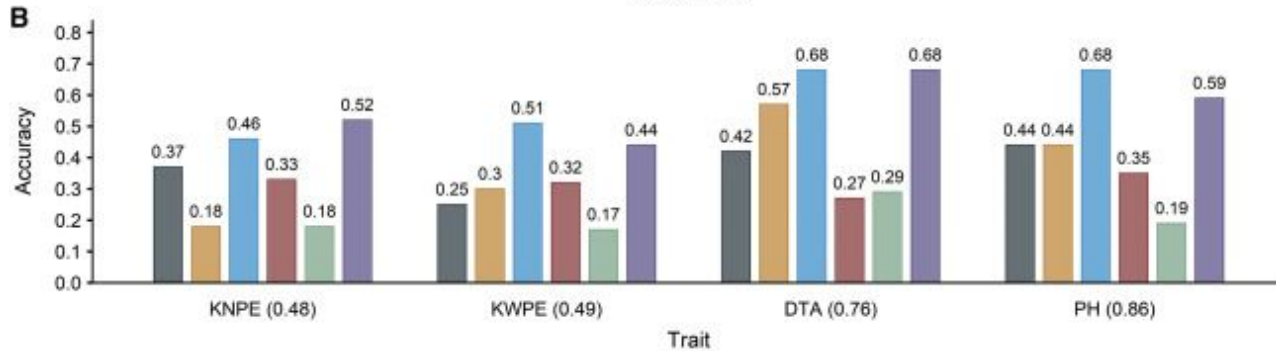| Model	Training RMSE	Training correlation coefficient (%)	Validation RMSE	Validation correlation coefficient (%)
DNN	12.01	84.01	12.81	81.44

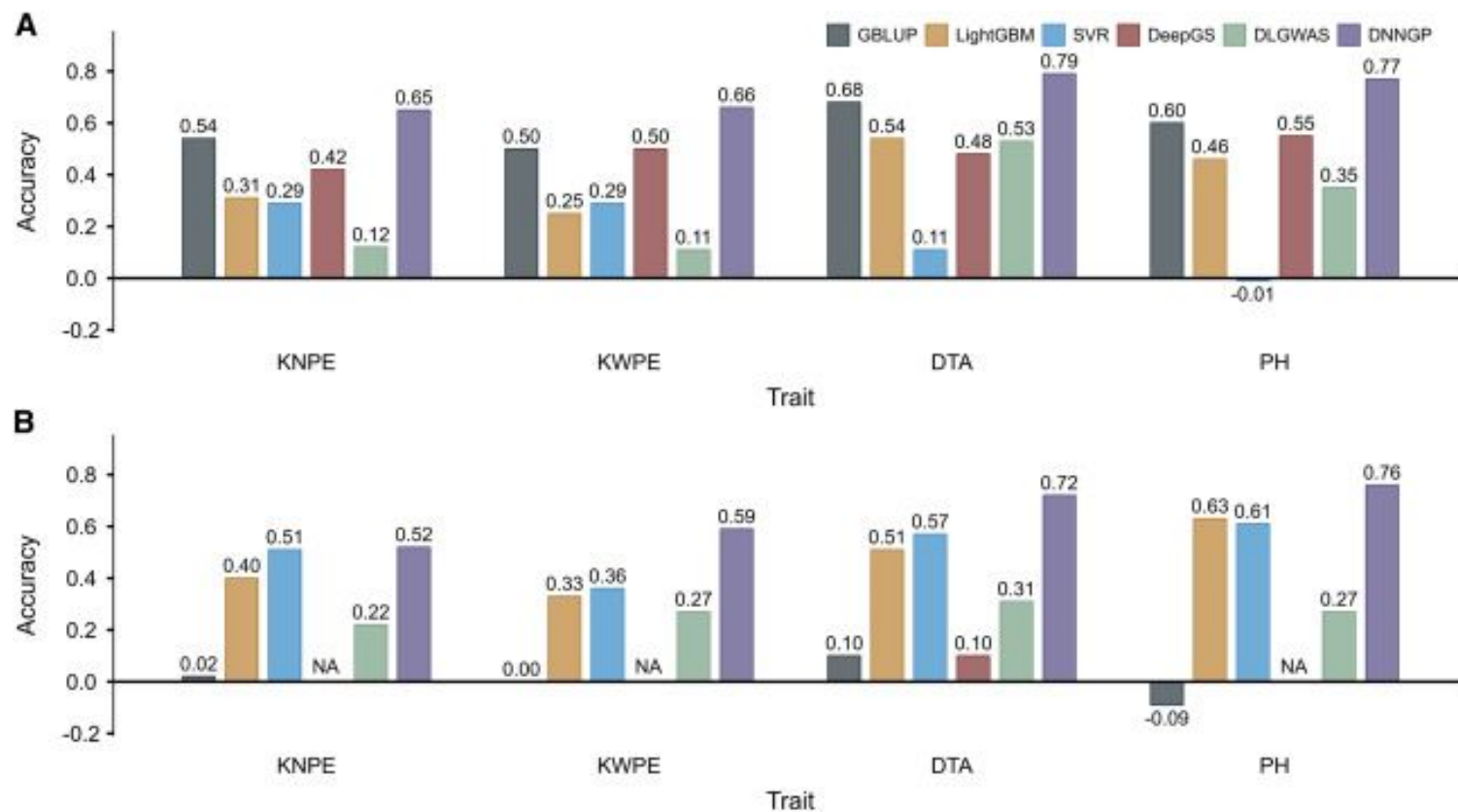The DNN model used 50 genetic markers and 20 environmental components selected by feature selection method.
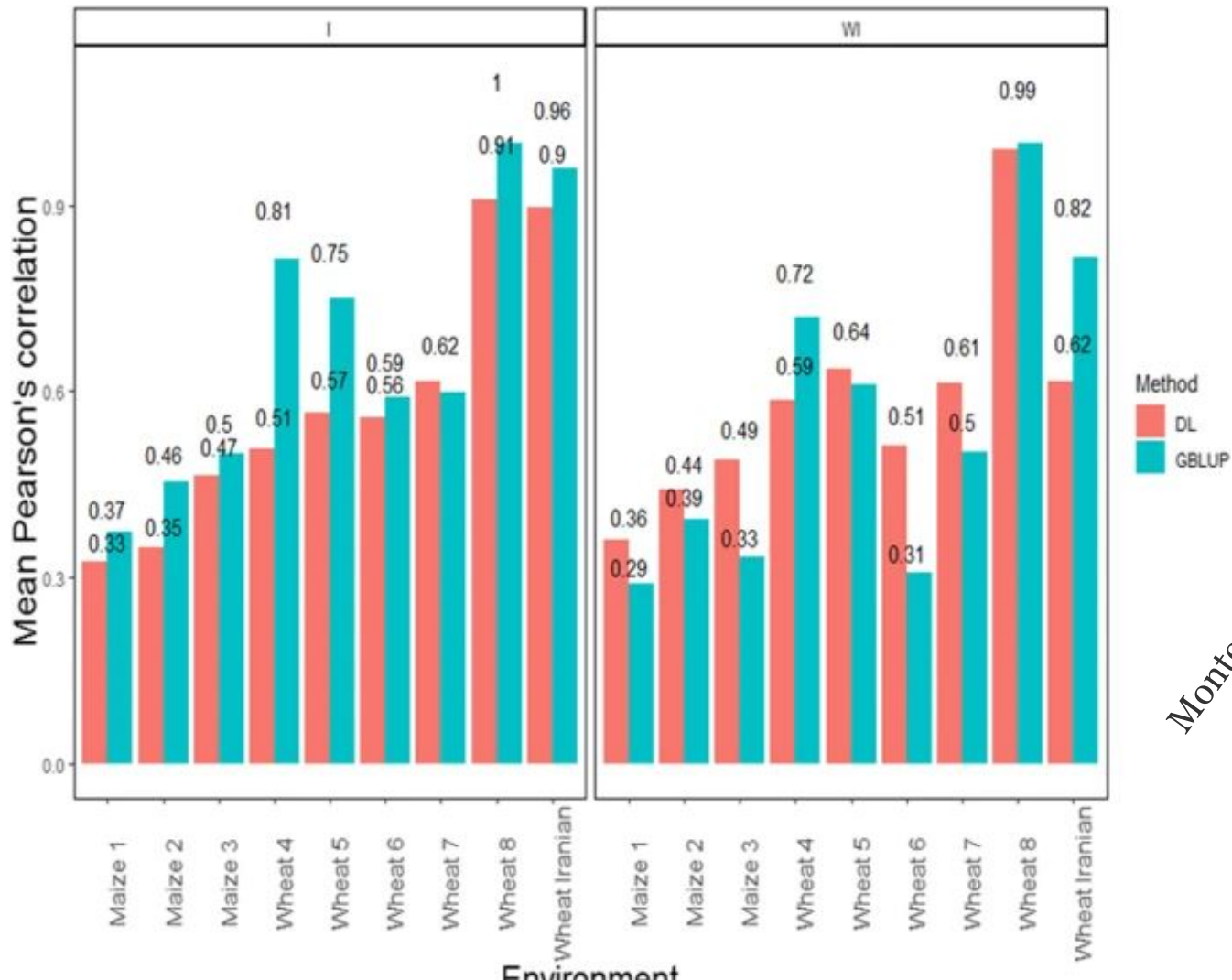
Montesinos-López et al., 2018
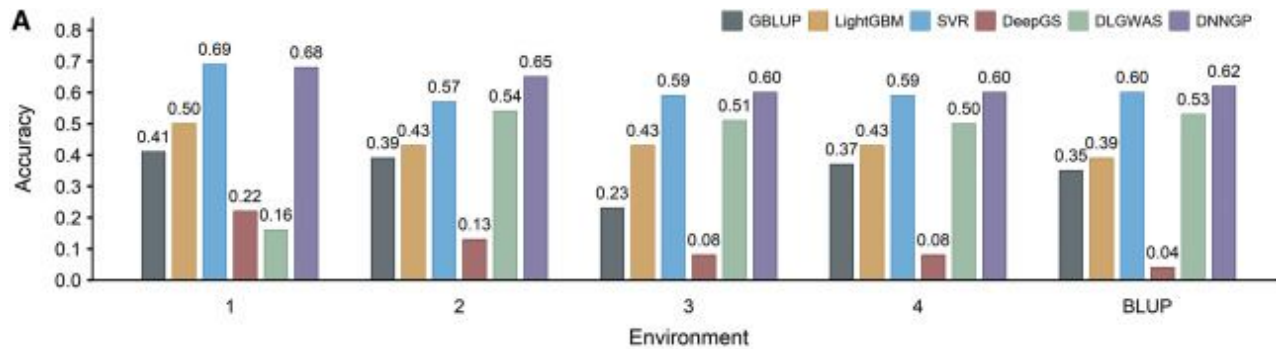
A - wheat599
B - maize1404
C - wheat2000

Wang et al. 2023

A - SNP

B - RNA-seq

Maize1404

Accounting for epistasis improves genomic prediction of phenotypes with univariate and bivariate models across environments

Elaheh Vojgani ✉, Torsten Pook, Johannes W. R. Martini, Armin C. Hölker, Manfred Mayer, Chris-Carolin Schön & Henner Simianer

Conclusion

We proposed breeding scheme for maize that includes DH lines obtaining, MET, genomic selection and CMS introduction.

We identified 5 highly promising DH lines for further corn hybrid production. Lines have one of the following traits: early vigor, early flowering, lodging resistance.

Obtained lines was evaluated in LineXtester analysis and show good performance.

Thank you for your attention!

Github repo