

# Implementation of Genomic Selection in CIMMYT –Africa Maize Breeding Program

Yoseph Beyene, M. Gowda, J. Crossa, M. Olsen; K. Robbins, P. Perez, J. Burgueno, and  
BM. Prasanna

Presentation to the EiB Breeding Optimization Community of Practice Monthly  
Meeting, 12 May 2022

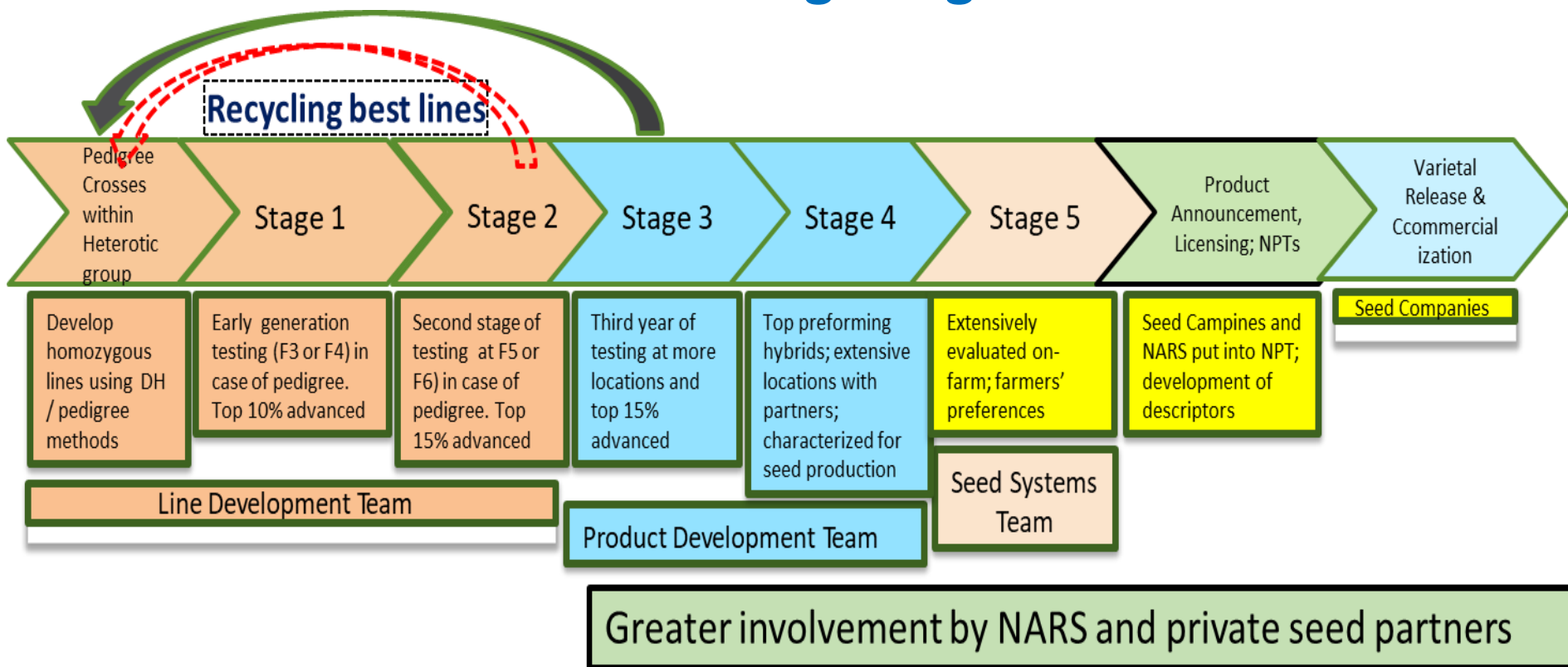


# CIMMYT Maize Breeding Pipelines in ESA

Code	Pipeline description	Target Countries	Est. Area (M ha)
EA-PP1	<b>Early/intermediate-maturing, white maize</b> varieties with <b>drought-tolerance, NUE</b> and resistance to GLS, TLB, Ear rots, MSV and <b>MLN</b> , and suitable for food use in Eastern African <b>tropical rainfed dry/wet mid-altitude areas</b> .	Ethiopia, Kenya, Uganda, Tanzania (Northern)	3.17
EA-PP2	<b>Late-maturing, white maize</b> varieties with <b>drought tolerance, NUE</b> , and resistance to GLS, TLB, MSV, common rust, and ear rots, adapted to <b>Eastern African tropical rainfed wet, upper mid-altitude areas</b> , and used mainly for food purposes.	Ethiopia, Kenya, Uganda, Tanzania (Southern)	3.38
EA-PP3	<b>Late-maturing, multiple stress-tolerant white maize</b> varieties for the <b>Eastern African tropical rainfed highlands</b> and used mainly for food purposes.	Ethiopia, Kenya, Uganda, Tanzania (Southern)	1.75
SA-PP1	<b>Intermediate/late maturing</b> , nitrogen-use efficient (NUE), <b>drought- and heat-tolerant white maize</b> varieties for the Southern African tropical rainfed mid-altitude/transition areas and used mainly for food purposes.	Mozambique, Zimbabwe, Malawi, Zambia, Tanzania (Southern); drought-prone smallholder farm areas in South Africa	3.74
SA-PP2	<b>Early-maturing, drought-, heat- and low soil pH stress-tolerant white maize</b> varieties for the <b>Southern African dry/wet lowland and mid-altitude areas</b> and used mainly for food purposes.	Zimbabwe, Malawi, Zambia, Tanzania (Southern), drought-prone smallholder farmers, about 2-3% of total maize area)	2.03

14 million hectares

# CIMMYT Maize Breeding: Stage-Gate Process



- **Stage 1** – First testcross evaluation; one tester; 2 reps, 3-5 sites
- **Stage 2** – Selected lines (10-15% S.I.) from Stage 1 trials; 3 testers; 2 reps, 8-10 sites
- **Stage 3** – Selected lines from Stage 2 trials (15% S.I.); Cross with 5 testers; 2 reps, 10-15 locations
- **Stage 4 (Regional On-station Trials)** – Best products from Stage 3; 2 rows, 3 reps, 25-35 locations
- **Stage 5 (Regional On-Farm Trials)** – 30-50 on-farm trials per Product Profile; Farmers' preferences
- **Final Product Advancement Meeting** to identify products/pre-commercial hybrids to be announced to the partners through CIMMYT Website

# Criteria for Product Advancement

## Stage 1 to Stage 2 [Line Development Team]

- Line entering Stage 1 TC are pre-selected for disease resistance (e.g., MSV, MLN) using markers.
- GEBV of the line analyzed
- Hybrid performance (yield under optimal, and abiotic and biotic stresses)
- Selection intensity of 10-15% applied
- Independent culling/selection index
- Number of lines used in Stage 1 varies from 1000 to 1500.

## Stage 2 to Stage 3 [Line Development Team]

- GCA of the line with testers
- Hybrid performance at more locations relative to Stage 1 ( optimal; abiotic and biotic stresses)
- Sites with low heritability ( $<0.10$ ) discarded
- Selection intensity of 10-15% applied.
- Lines selected for recycling

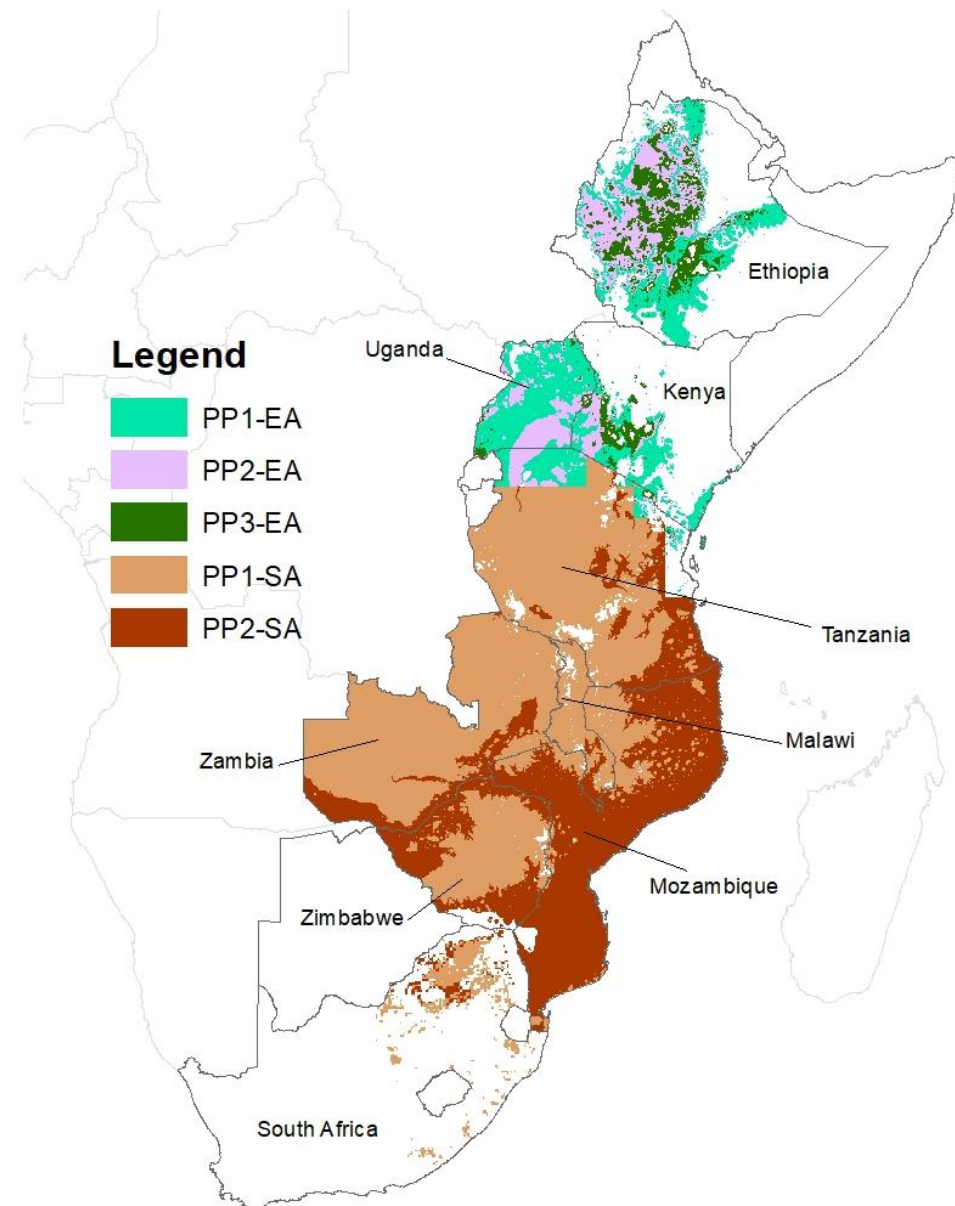
## Stage 3 to Stage 4 [Product Development Team; Advancement Committee]

- Hybrid performance at more locations (10-15) relative to Stage 2 (optimal; abiotic and biotic stresses)
- Sites with low heritability ( $<0.10$ ) discarded
- Selection intensity of 15% applied



# Trait Packages and Advancement Thresholds

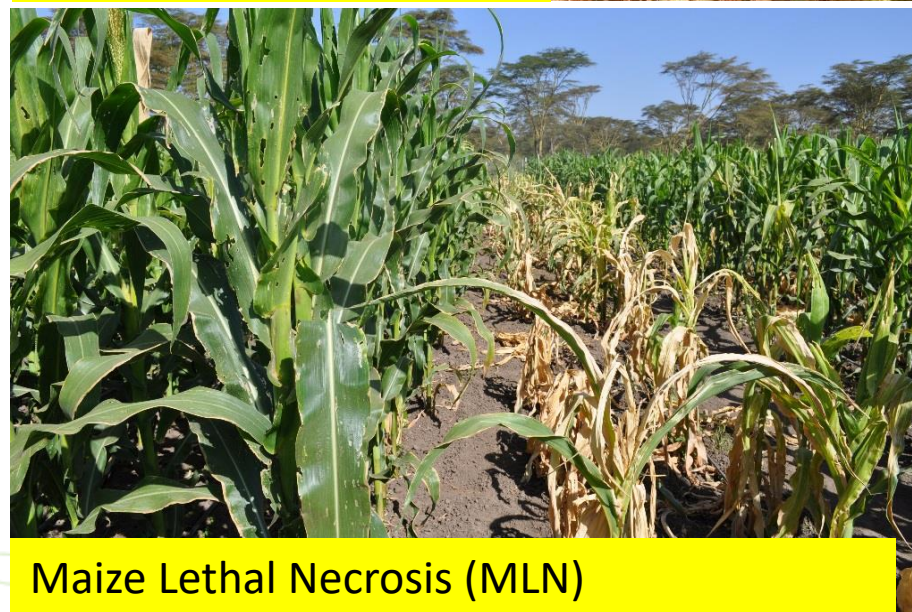
Basic Traits	Product target	PHB30G19, DK777, H517
Yield	>10% greater than commercial checks)	1
Grain color	White	1
Maturity	Equivalent to best check	1
Plant height	±10 cm of best check	2
Ear height	±10 cm of best check	2
Grain moisture	Equivalent to best check	2
Drought tolerance	>10% > mean of checks	1
Nitrogen use efficiency (NUE)	>10% > mean of checks	2
TLB resistance	<4.0 (1-9 scale)	1
GLS resistance	<4.0 (1-9 scale)	1
MSV resistance	<3.0 (1-9 scale)	1
Common rust resistance	<4.0 (1-9 scale)	2
Fusarium ear rot resistance	Less than 10% incidence	1
Bare tips (poor husk cover)	Less than 10% incidence	1
MLN resistance	10 % > of the mean of checks	1
Value-added traits	Product target	Market priority
<i>Striga</i> tolerance	>10% or equal to the best trait checks	2
FAW resistance	≤5 leaf damage; ≤3 ear damage (1-9 scale)	2



1 = top priority; 2 = lower priority.



# Sampling TPEs





# GS being applied at CIMMYT maize breeding to accelerate the rate of genetic gain in two ways

1. Rapid cycle genomic selection (RCGS)
2. Predict the performance of un-phenotyped genotypes at an early stage of testing



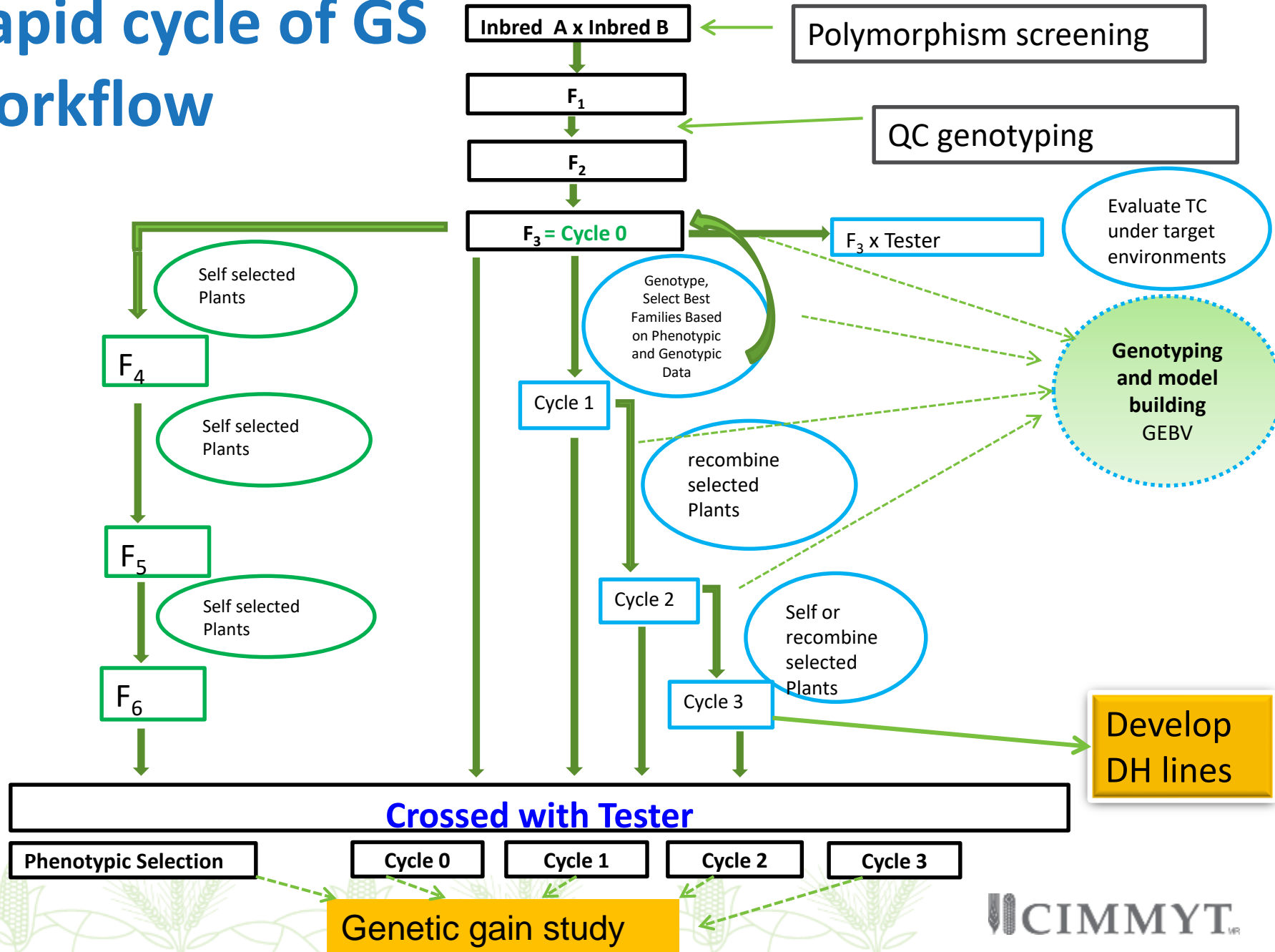
# Rapid generation advancement using GS

- 34 bi-parental populations (total =6252 F<sub>2:3</sub>, each with aprox. 184 progenies)
- Each pop. phenotyped in 2-4 managed water-stressed, 3-4 well-watered environments, and genotyped with low and high-density markers
- Genetic gain studies completed for 18 populations





# Rapid cycle of GS workflow



# Training set: TC evaluation under Optimum, Drought and Diseases

Drought



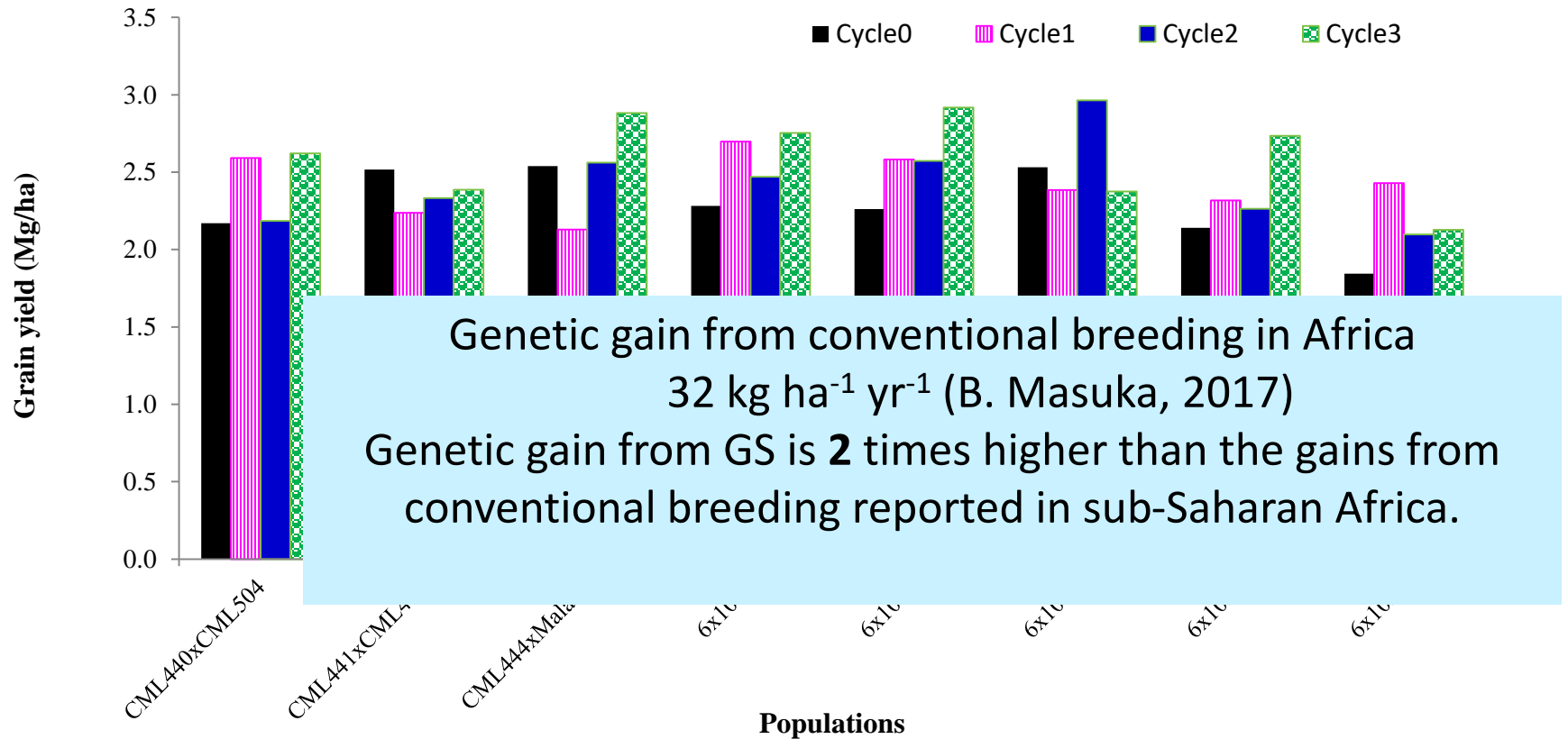
Optimum



Diseases



# RCGS results: Gain in grain yield under drought environments in SSA

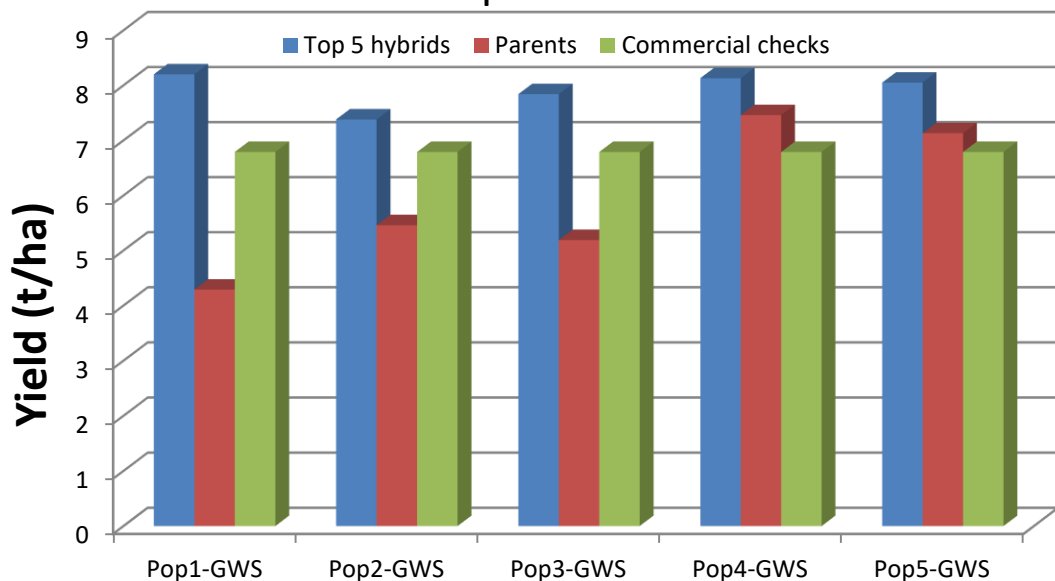


Overall gain in GY : 70.5 kg ha<sup>-1</sup> year<sup>-1</sup>

Beyene et al. (2015) Crop Sci. 55:154–163

# Performance of hybrids developed from lines derived via RCGS

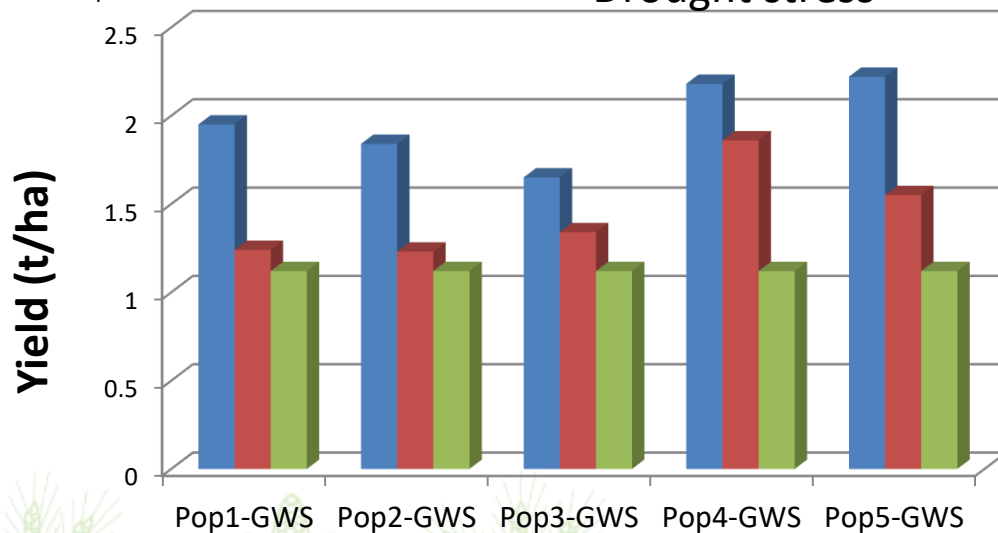
## Optimum



Gain over the commercial checks = 8.7 to 20.8%

Gain over the parents = 9.0 to 91.1%

## Drought stress



- Gain over the commercial checks = 47.3 to 98.2%

- Gain over the parents = 17.2 to 57.3%



# Lines developed through RCGS are being used as the parent of allocated hybrids to partners

Line	Parent in # allocated hybrids
CKLMARS1C3S50268	1
CKLMARS1C3S50080	2
CKLMARS1C3S50113	3
CKLMARS1C3S50140	2
CKLMARS1C3S50137	1





# DT tolerant and susceptible hybrids at Kiboko-2021





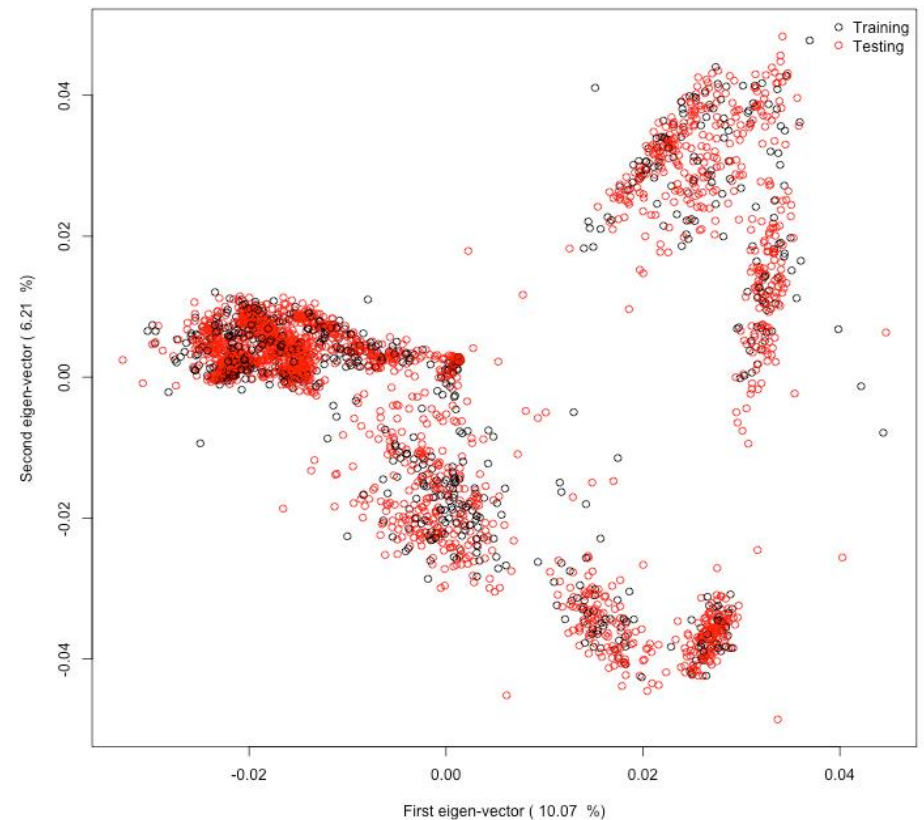
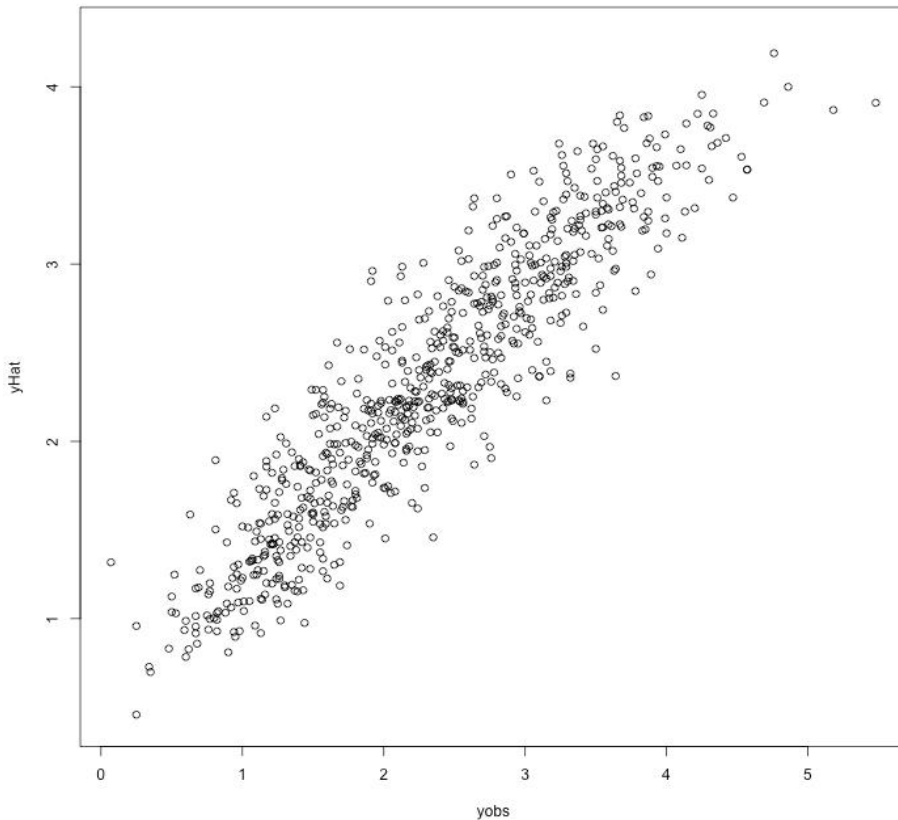
# Genomic selection using “test half-predict-half” strategy

# lines genotyped (Stage I)= 2511

# DH lines phenotyped (test crosses) =855

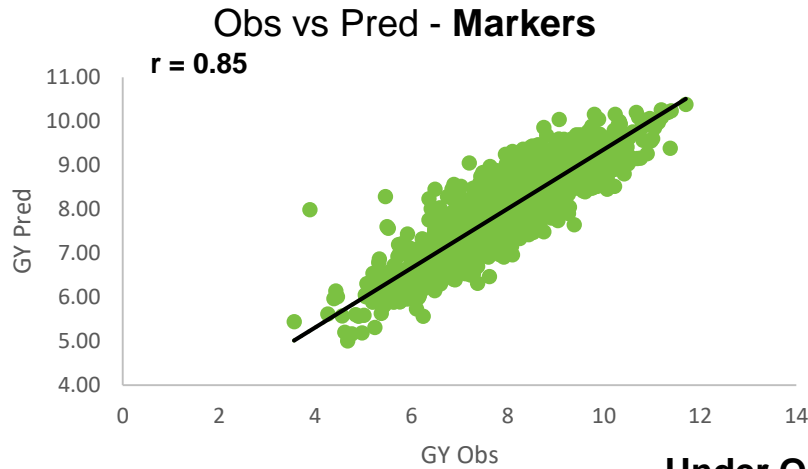
# sites=2 optimum and one managed drought

- The phenotypic data were used to predict the remaining untested lines

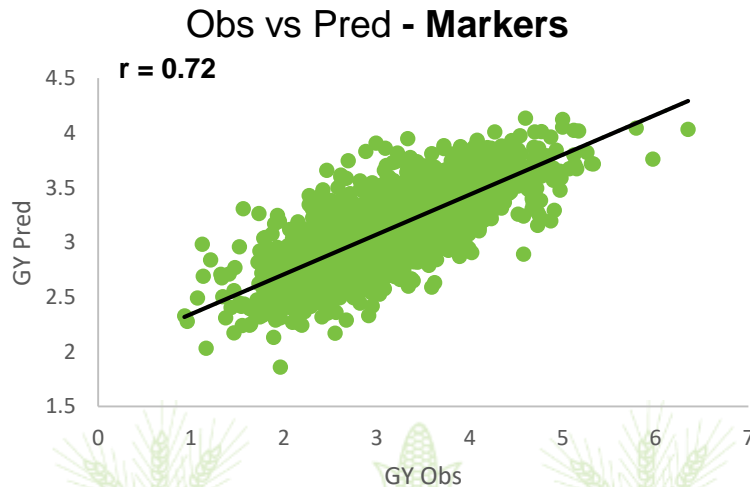
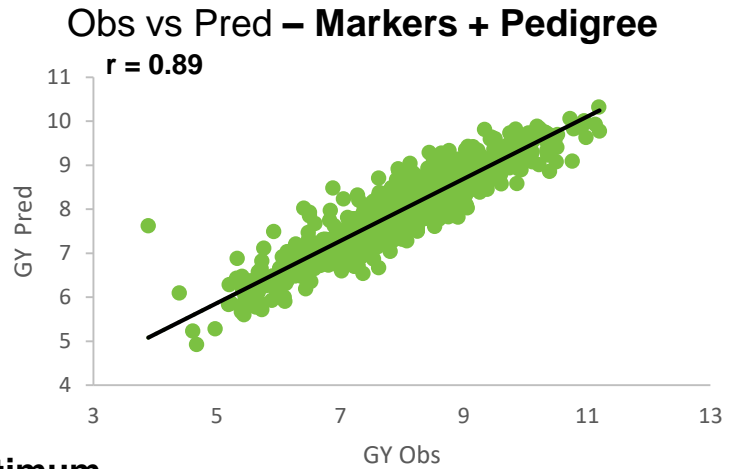


Correlation (observed and predicted): Optimum: 0.84 and Drought: 0.92

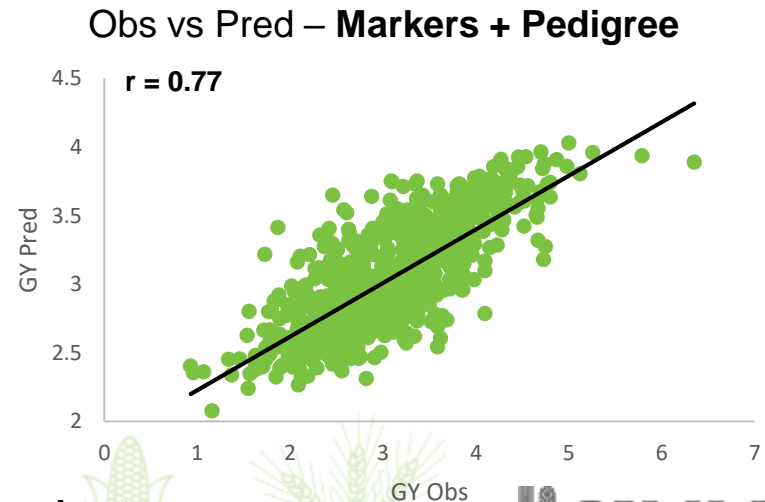
# Genomic selection using test half-predict-half strategy –incorporating pedigree



**Under Optimum**



**Under drought stress**





# Cross-validated prediction accuracy within and across testers

Management/trait	Tester 1	Tester 2	Across testers
	111	742	853
GY-WW	0.16±0.12	0.60±0.03	0.67±0.05
GY-WS	0.22±0.18	0.64±0.07	0.65±0.05
MOI-WW	0.16±0.14	0.58±0.01	0.65±0.04
MOI- WS	0.44±0.16	0.61±0.06	0.57±0.05
AD-WW	0.41 ±0.13	0.70 ±0.07	0.75 ±0.04
AD- WS	0.49±0.20	0.63±0.04	0.67±0.05
PH-WW	0.14±0.12	0.65±0.03	0.70±0.03
PH-WS	0.17±0.12	0.72±0.04	0.72±0.04



# Advancement of lines based on GEBV and PS:

Population	CML536/LPS-F64
# DH genotyped	166
# DH lines phenotyped	88
# lines selected based on Phenotype	21
# lines selected based on GEBV	19

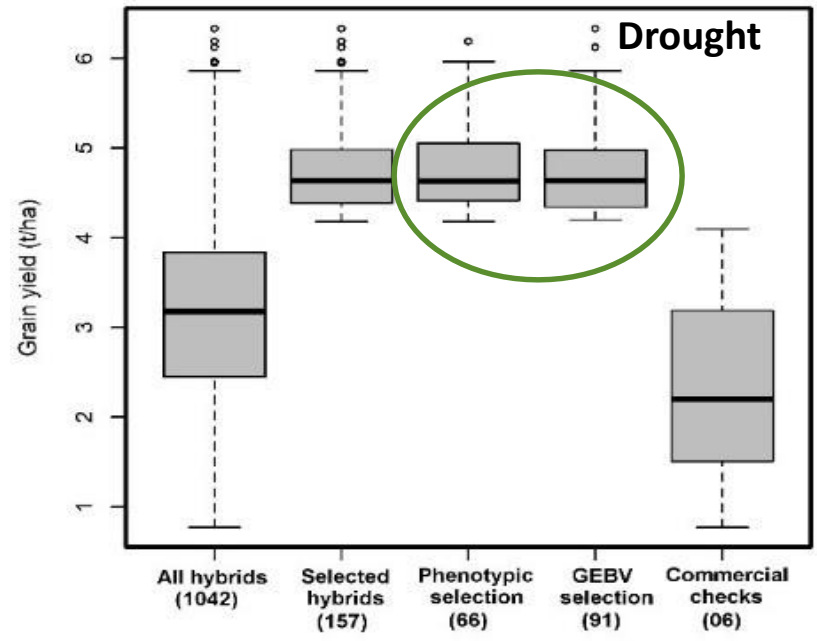
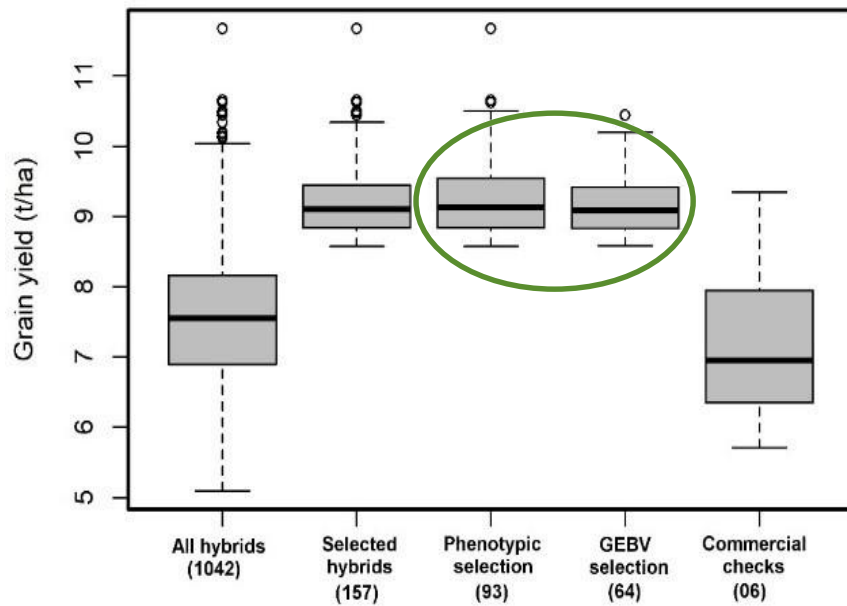
Check	GY(t/ha)_Opt	GY(t/ha)_MD	MOI-Opt	PH_Opt
H517	6.5	1.8	16.4	267.4
Pioneer 30G19	6.2	2.9	17.9	254.5
WH505	7.6	3.1	17.9	257.3
Heritability	0.64	0.52	0.3	0.8
Genotype Variance	0.70	0.21	0.3	79.8
GenxLoc Variance	0.13		0.3	16.2
Residual Variance	2.06	0.39	4.0	81.5
Grand Mean	6.81	3.10	18.2	248.8
LSD	1.43	0.93	1.4	11.6
CV	21.07	20.27	10.9	3.6
n Replicates	2	2	2	2
n Locations	3	1	3	3

SN	name	Observed GY_BLUE_optimal	Predicted GY_BLUE_optimal	Observed GY_BLUE_drought	Predicted GY_BLUE_drought
1	CKDHL152921	9.1	7.6	3.1	3.1
2	CKDHL152857	8.8	7.6	4	4
3	CKDHL152610	8.4	7.7	4.7	3.8
4	CKDHL152563	8.4	7.7	4.2	3.6
5	CKDHL152554	8.3	7.7	3.6	3.3
6	CKDHL152653	8.3	7.8	3.3	3.3
7	CKDHL152616	8.2	7.3	4.3	4.1
8	CKDHL152617	8.2	7.2	3.8	3.4
9	CKDHL152821	8.2	7.3	3	3.1
10	CKDHL152733	8.1	7.5	4.2	3.4
11	CKDHL152658	8.1	7.2	3.4	3.2
12	CKDHL152638	8.1	7.5	3.6	3.1
13	CKDHL152976	8	7.5	3.3	3.4
14	CKDHL152591	8	7.6	3.7	3.6
15	CKDHL152906	7.6	7.2	3.6	3.2
16	CKDHL152751	7.5	7.3	3.6	3.8
17	CKDHL152769	7.4	7.2	4.7	3.8
18	CKDHL153005	7.4	7.1	4	3.7
19	CKDHL152929	7.3	7	3.9	3.8
20	CKDHL152866	7.1	7.2	4.3	3.7
21	CKDHL152962	7	7	4.3	3.5
1	CKDHL152820	NA	7.6	NA	3.5
2	CKDHL152994	NA	7.5	NA	3.7
3	CKDHL152529	NA	7.5	NA	3.5
4	CKDHL152590	NA	7.5	NA	3.3
5	CKDHL152811	NA	7.4	NA	3.8
6	CKDHL152682	NA	7.4	NA	3.4
7	CKDHL152579	NA	7.4	NA	3.5
8	CKDHL152927	NA	7.4	NA	3.5
9	CKDHL152759	NA	7.4	NA	3.9
10	CKDHL152890	NA	7.4	NA	3.3
11	CKDHL152689	NA	7.4	NA	3.1
12	CKDHL152632	NA	7.3	NA	3.3
13	CKDHL152773	NA	7.3	NA	3.9
14	CKDHL152862	NA	7	NA	3.9
15	CKDHL152813	NA	7	NA	3.8
16	CKDHL152879	NA	7.2	NA	3.8
17	CKDHL152777	NA	7	NA	3.7
18	CKDHL152849	NA	7.2	NA	3.7
19	CKDHL152778	NA	7.2	NA	3.6



# Comparison of GS vs PS

Category	# lines	# testers	# of hybrids
All stage II hybrids	348	3	1042
Hybrids advance through phenotype	176	3	526
Hybrids advance through GEBV	172	3	516



## Stage1 TC:

**50% Phenotyping + 50% genotyping**

Beyene et al (2019). *Front. Plant Sci.*  
10:1502. doi: 10.3389/fpls.2019.01502

<b>GS:PS cost ratio</b>	<b>0.68</b>
-------------------------	-------------



# Performance of the top hybrids (advanced via phenotypic and GEBV) for grain yield and other agronomic traits compared to the best check under optimum and drought conditions

Entry	Pedigree	Advancement	Optimum (GY/ha)	% increase	Drought (GY/t)	AD	%MOI	TLB	% ER
20	(CML543/CML566)//CKDHL150421	Phenotype	9.5	20.4	3.9	72.3	19.2	2.1	3.1
97	(CML395/CML444)//CKDHL1500213	GEBV	9.4	19.1	3.8	75.2	20.1	2.2	1.3
86	(CML543/CML566)//CKDHL1500261	GEBV	9.1	14.7	5.1	73.0	20.0	2.1	1.3
58	(CML395/CML444)//CKDHL150399	Phenotype	9.0	13.6	3.8	72.9	19.7	2.2	2.7
55	(CML395/CML444)//CKDHL1500041	Phenotype	8.9	12.8	3.8	72.7	19.5	2.5	1.8
76	(CML395/CML444)//CKDHL150431	GEBV	8.9	12.7	3.3	73.0	20.6	2.0	0.5
77	(CML543/CML566)//CKDHL150431	GEBV	8.8	11.0	4.3	74.3	21.6	2.3	5.5
126	(CML322/CML543)//CKDHL1500215	GEBV	8.7	10.0	4.4	70.7	19.5	2.0	3.9
56	(CML543/CML566)//CKDHL1500041	Phenotype	8.7	9.8	3.6	74.2	18.2	2.0	4.6
65	(CML543/CML566)//CKDHL150342	Phenotype	8.7	9.6	3.5	73.1	19.1	2.5	1.6
98	(CML543/CML566)//CKDHL1500213	GEBV	8.7	9.1	3.7	76.0	21.3	2.5	0.9
21	(CML322/CML543)//CKDHL150421	Phenotype	8.5	7.6	3.2	70.1	19.5	2.0	0.8
11	(CML543/CML566)//CKDHL150339	Phenotype	8.5	7.5	4.8	73.6	20.4	2.4	1.9
137	WH505		7.9		3.3	73.9	18.8	2.3	1.0
136	PH 30G19		7.9		3.6	67.8	20.7	2.0	4.1
139	DK 777		7.1		3.9	71.9	18.4	2.6	5.5
138	H516		6.7		2.2	69.9	18.2	2.0	6.6
140	DK 8031		5.7		2.1	66.9	16.9	2.2	8.7
	Heritability		0.6		0.5	0.9	0.6	0.6	0.4
	Grand Mean		7.5		3.6	72.6	19.7	2.5	3.3
	LSD		1.3		1.4	1.4	1.8	0.7	4.2
	CV		16.3		18.6	1.9	9.1	16.1	113.6
	n Locations		5		1	5	5	2	4

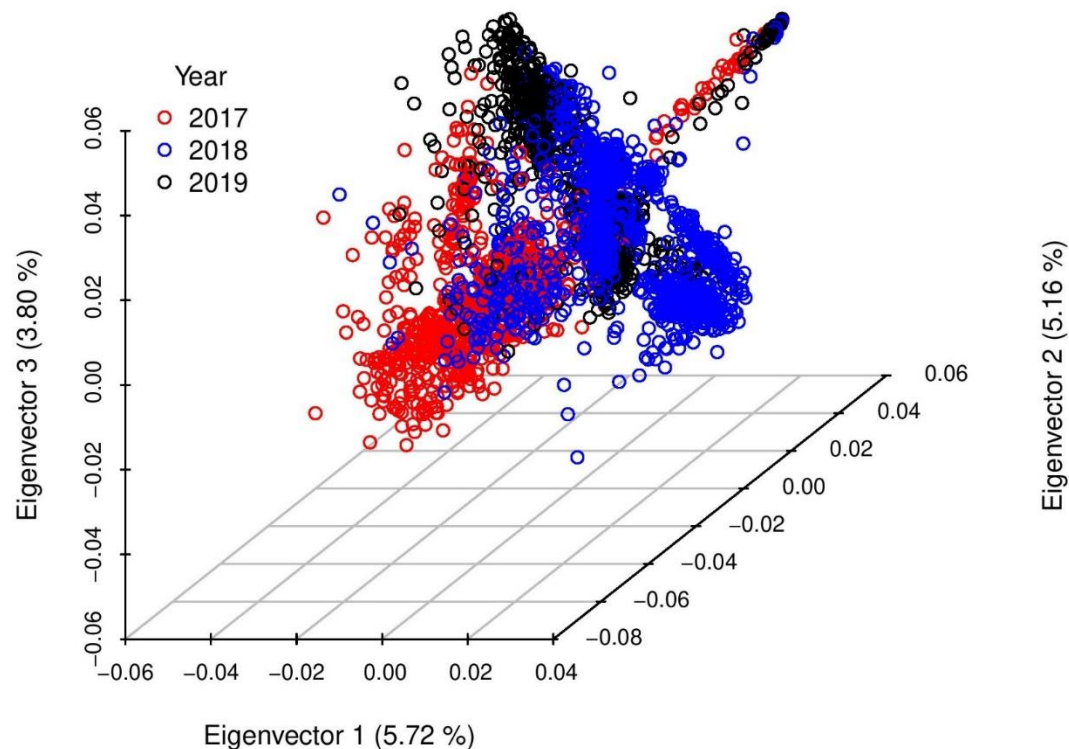


# GS using historical Data – Prediction across Years

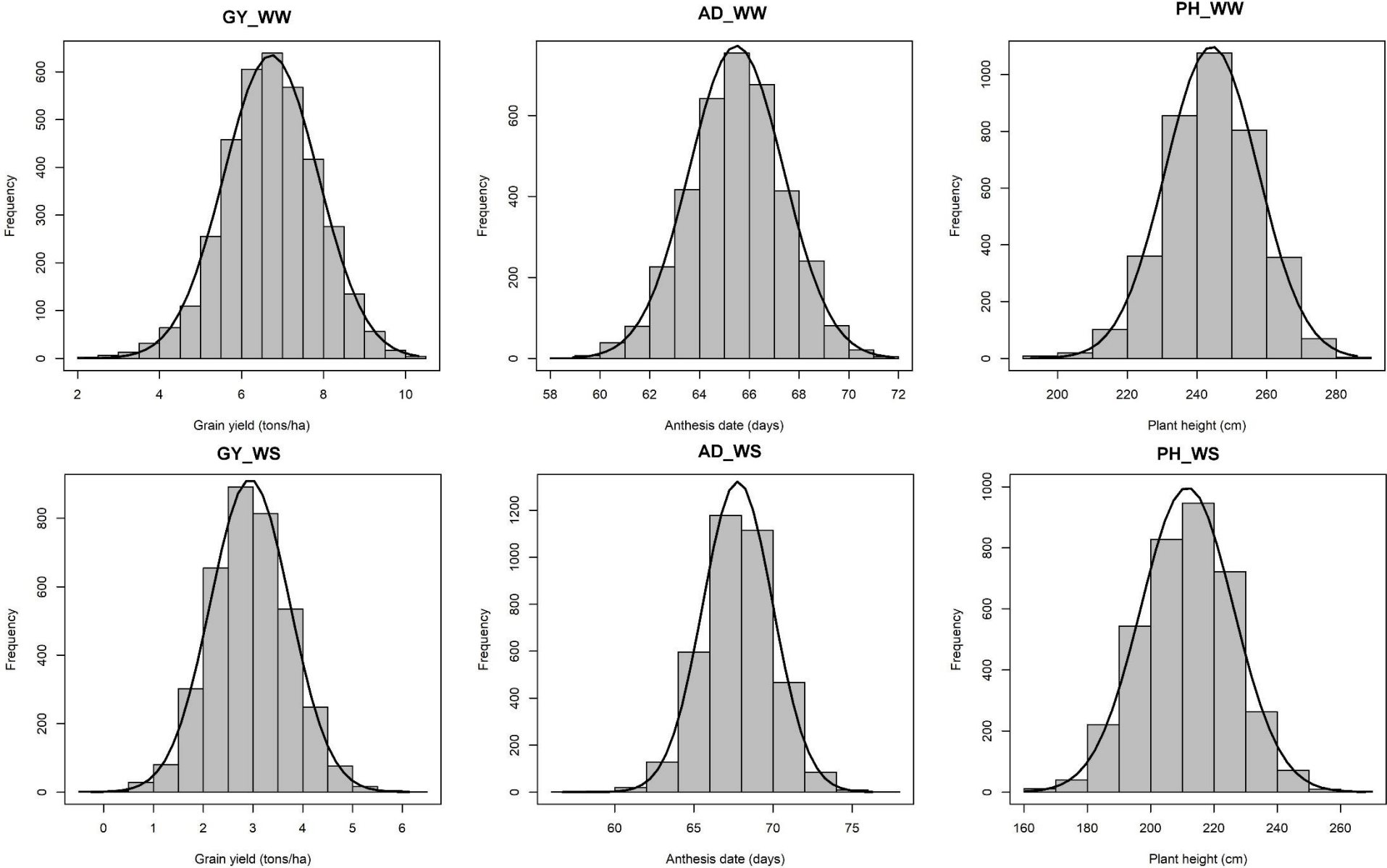
Scenario 1	2017 data (N)	Training set (TRN): 2017 data + x % of 2018)	Testing set (TST): (100-x) % 2018
	923		923 (0%)
		1065 (10%)	1281
		1035 (30%)	996
		1635 (50%)	712
		1919 (70%)	427
		2204 (90%)	142
Scenario 2	2017+ 2018 pooled data(N)	Training set: 2017 + 2018 + x % of 2019	Testing set: (100-x) % 2019
	2346		2346 (0%)
		2418 (10%)	650
		2563 (30%)	505
		2707 (50%)	361
		2851 (70%)	217
		2996 (90%)	117

Year	Number of lines phenotyped and genotyped records	Management
2017	923	Optimum and managed drought
2018	1423	Optimum and managed drought
2019	722	Optimum and managed drought
Total	3068	

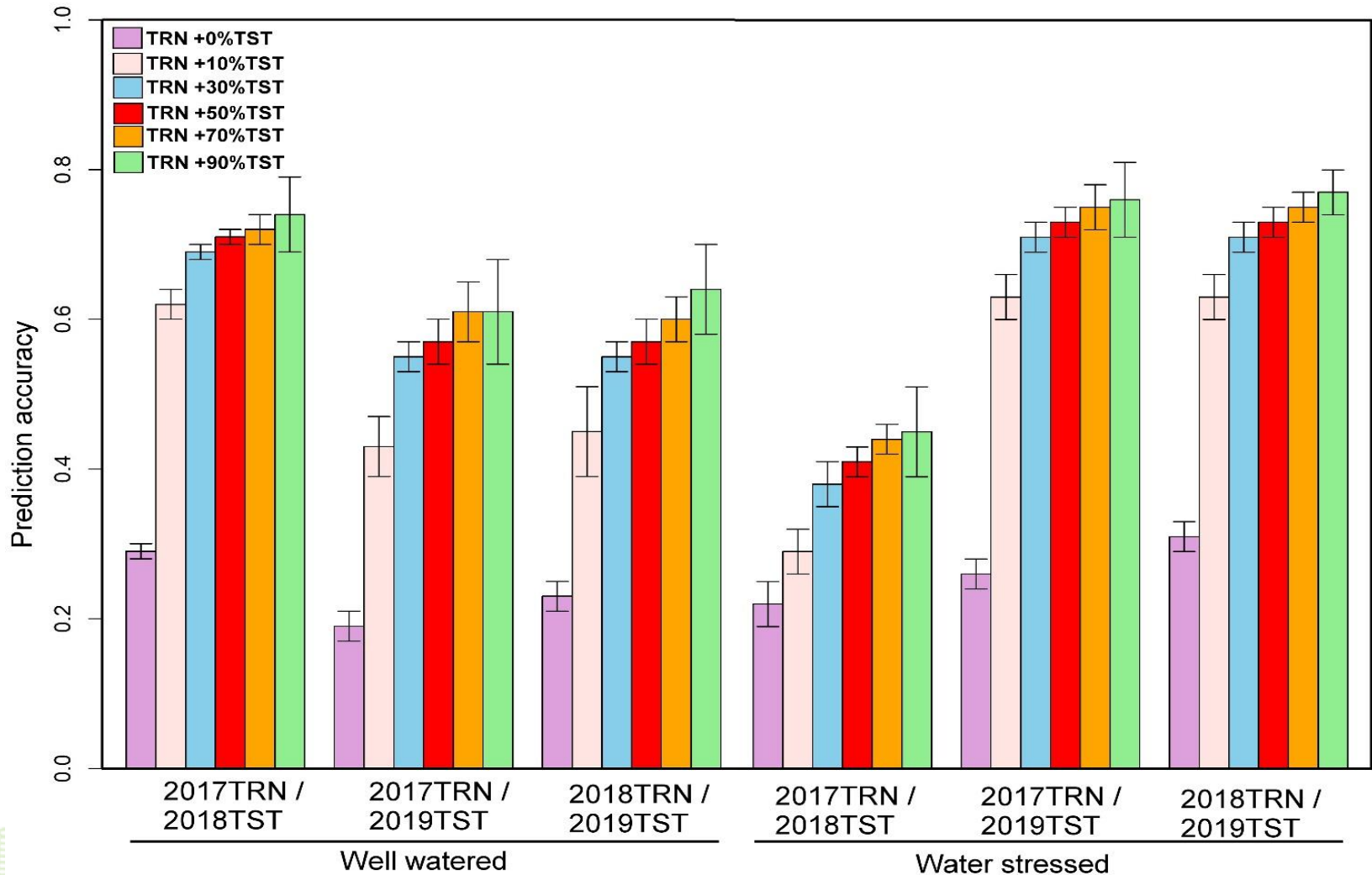
5173 markers (after QC)



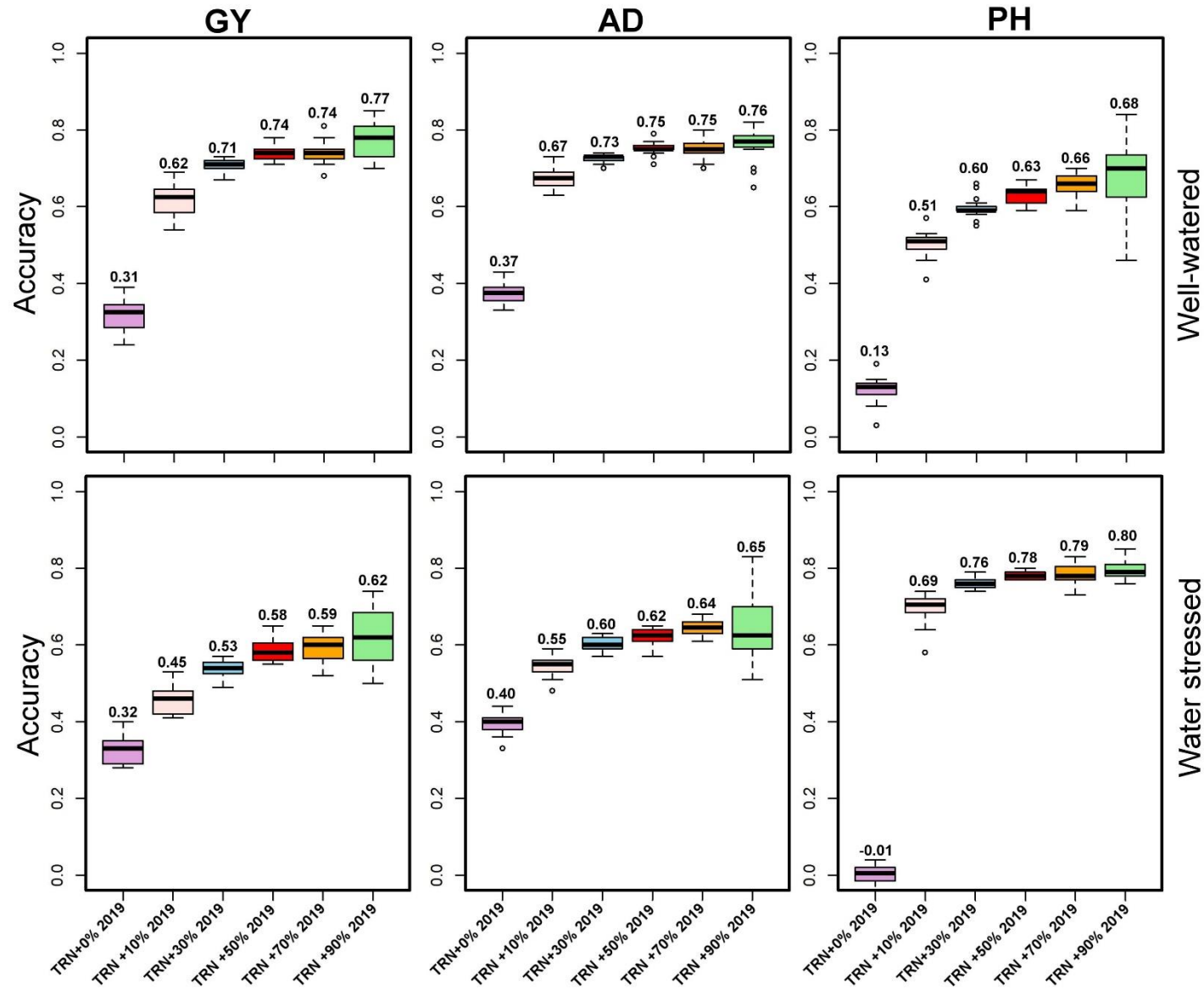
# Phenotypic distribution of GY, AD and PH under optimum (top) and managed drought (bottom) conditions



# Prediction accuracies for grain yield using one-year data to predict another year's data and converting 10%, 30%, 50%, 70% and 90% of the data from the TST to TRN



# Results of Predication Across Years when two years data used to predict third-year data



2017+2018 data use as training set to predict 2019

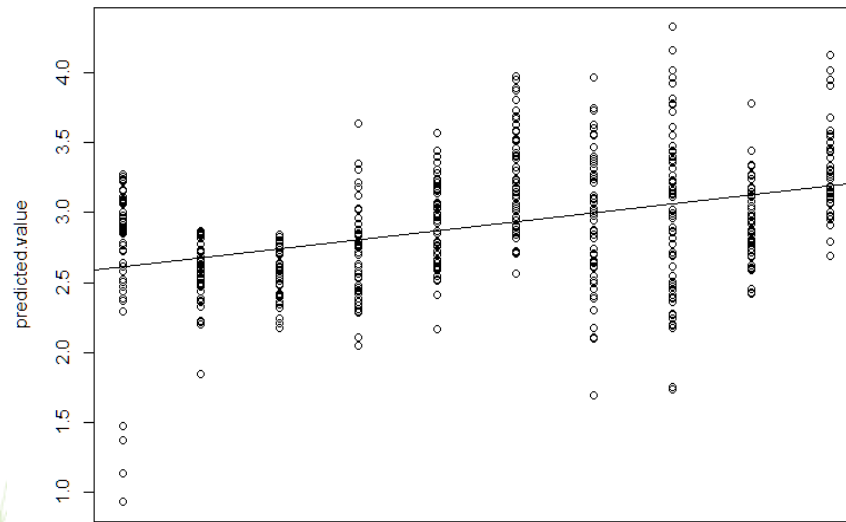
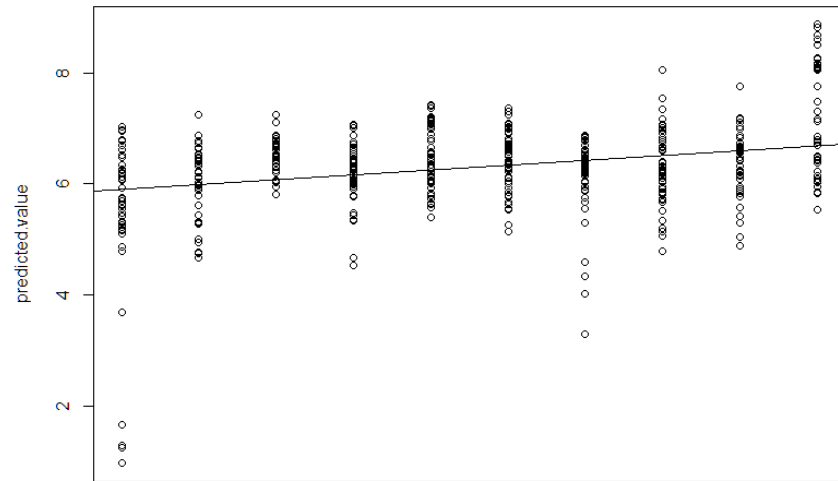
Beyene et .al, 2021



# Genetic gain estimation Grain yield (2011-2020)

Year	Optimum	Drought
2011	45	50
2012	44	44
2013	39	39
2014	43	43
2015	51	51
2016	53	53
2017	54	54
2018	65	65
2019	44	44
2020	38	38
<b>Total</b>	<b>476</b>	<b>481</b>
<b>Genetic gain (%)</b>	<b>1.61</b>	<b>2.4</b>

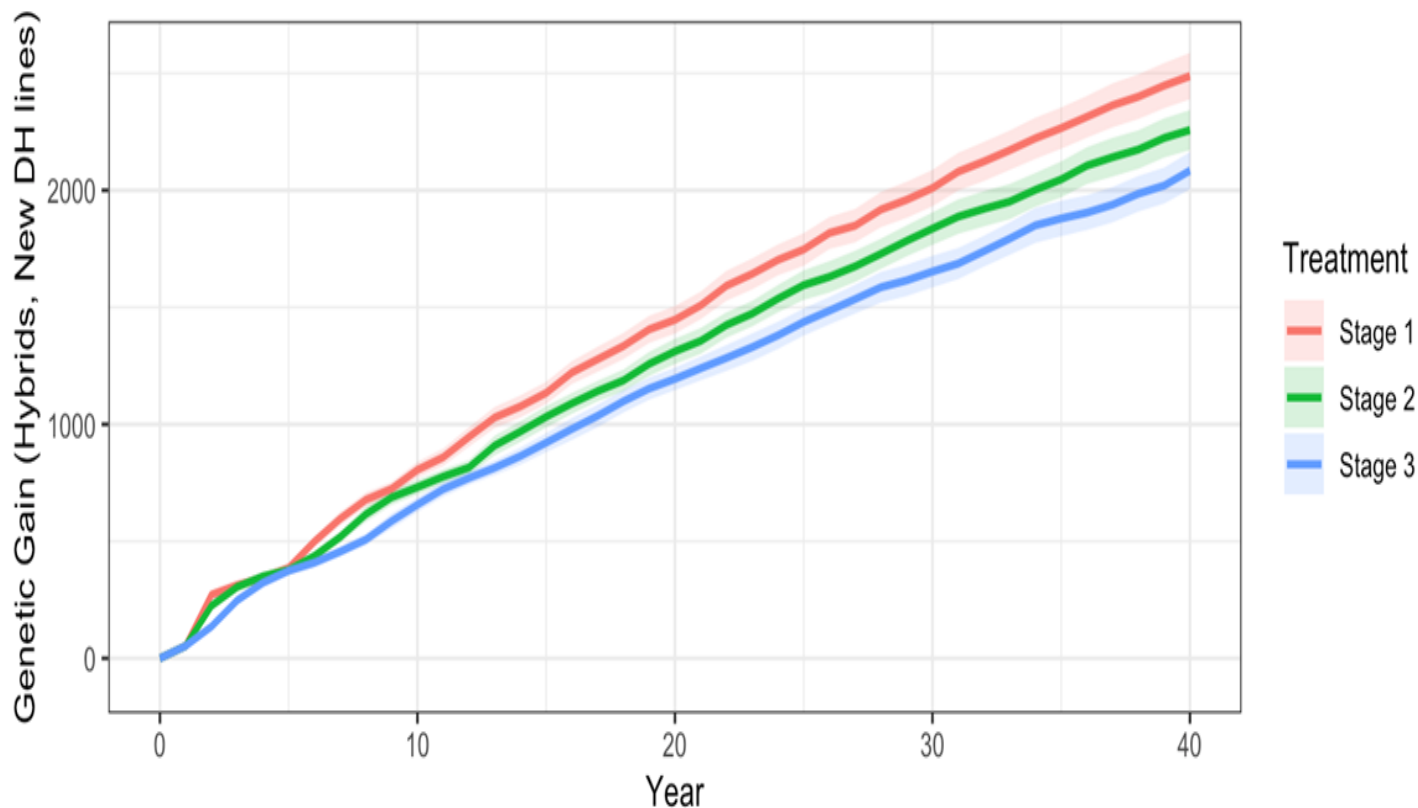
Optimum



Drought



# Simulation for Recycling Lines at Early Stage of Testing

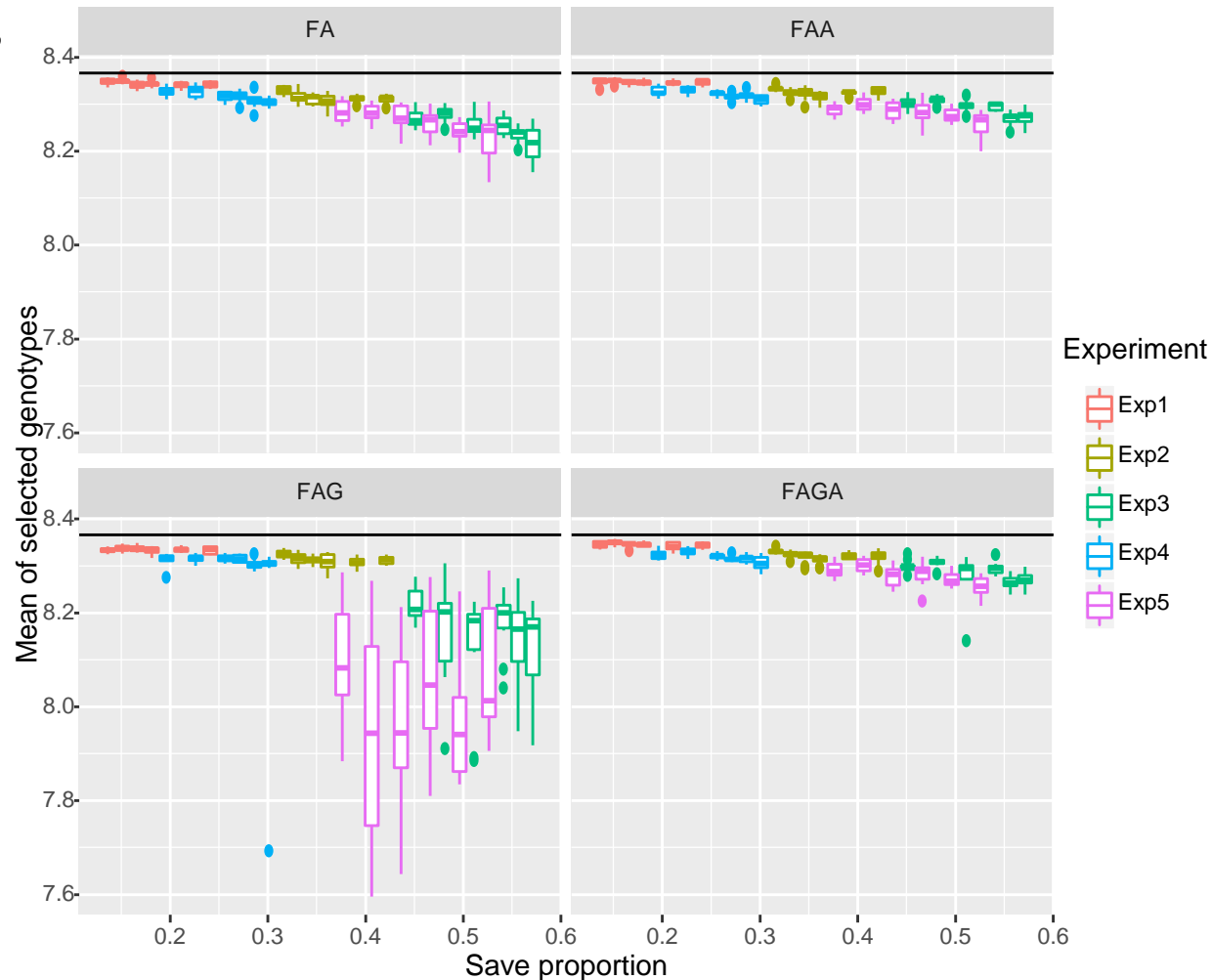


Simulation results comparing the current recycling at Stage 3 vs recycling after Stage 1 and 2 testing using data from EA-PP1. Recycling after Stage 1 or Stage 2 could deliver increase genetic gain by **17% and 9%** compared to recycling at Stage 3, respectively



# Sparse phenotyping to sample TPE

- Phenotypic data for 2018 Stage II trials (900 hybrids)
- Evaluated at 5 locations in Kenya
- Genotypic data for lines used in stage II trials
- Different experimental layout for sparse testing
- 4 types of analysis:
- Factorial analysis (FA),
- FA + CoP
- FA+ Marker data
- FA+ Marker + CoP



*The horizontal black line is the mean of the hybrids selected under complete phenotyping (8.36 t/ha)*

**Conclusion:** By saving 30% the phenotypic cost, 90% of the best hybrids were common between complete phenotyping and sparse phenotyping

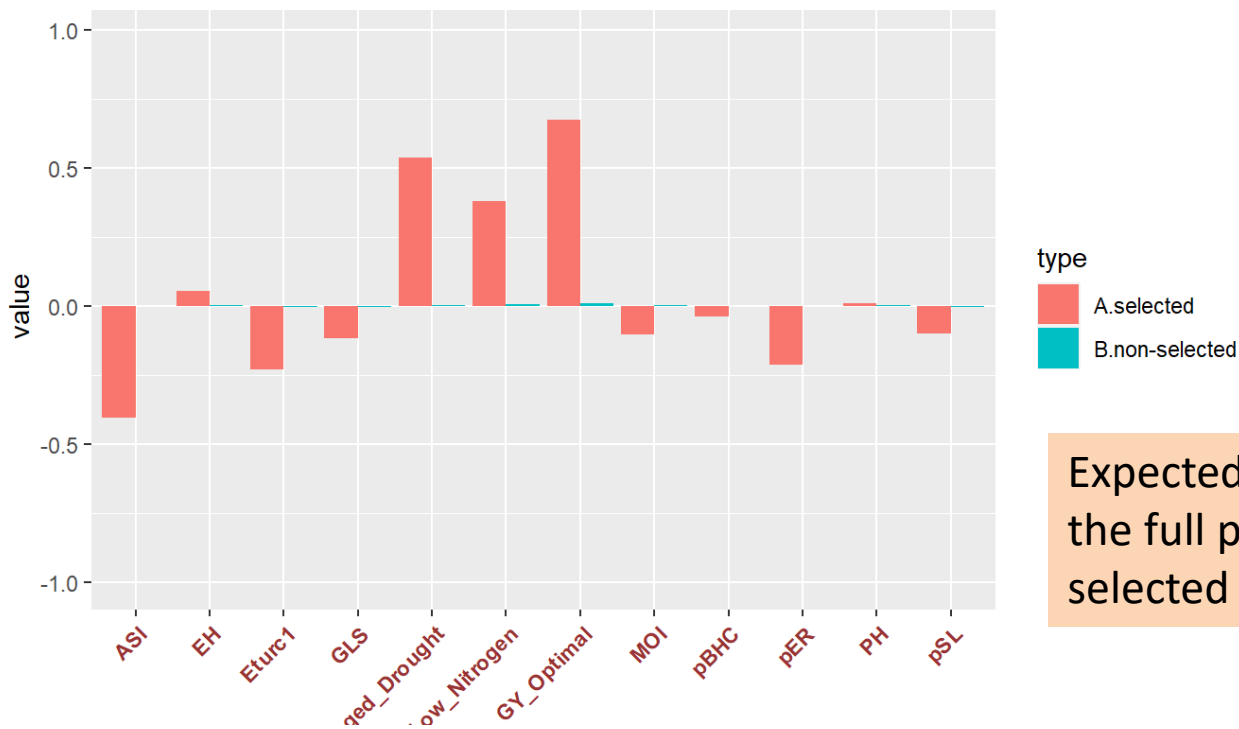
# Incorporating a selection index for selecting parents for recycling ( DESIRE software )

<b>No. of Entries</b>	<b>600</b>
<b>No. of Test Hybrids</b>	588
<b>No. of Checks</b>	12
<b>No. of Lines</b>	198
<b>No. of Testers</b>	3

Traits used for SI	H2
GY_Managed_Drought	0.83
GY_Managed_Low_Nitrogen	0.46
GY_Optimal	0.86
pER	0.90
pSL	0.87
pBHC	0.85
Eturc1	0.81
MOI	0.88
PH	0.95
EH	0.97
GLS	0.73
ASI	0.92

## Example of genetic merit

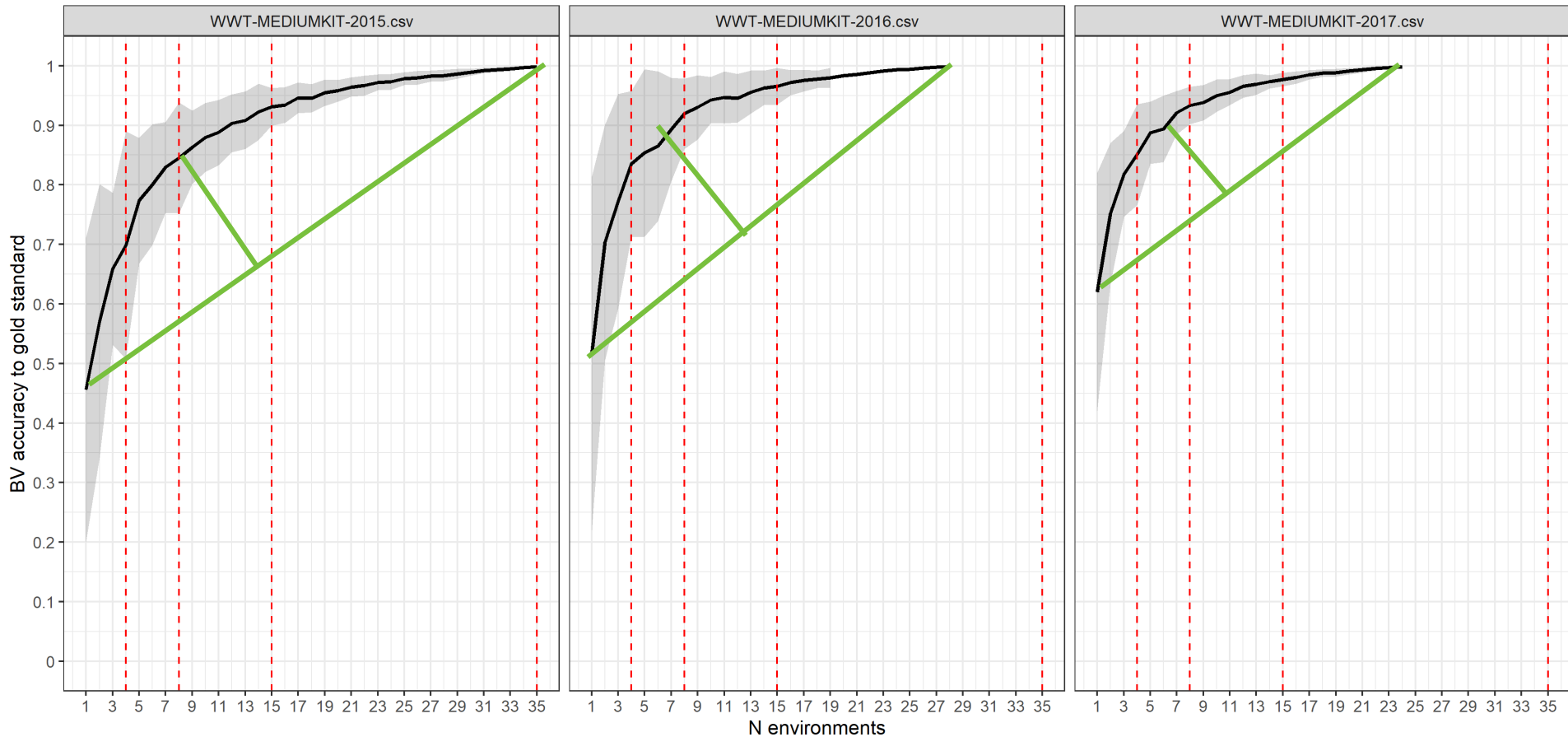
Line	Merit	Rank
CKDHL1715901	1.96	1
CKDHL1715896	1.77	2
CKDHL1715915	1.34	3
CKDHL1715480	-1.33	196
CKDHL1715260	-1.49	197
CKDHL1720872	-1.34	198



Expected changes in mean between the full population of lines versus a selected top 20% the lines using index.



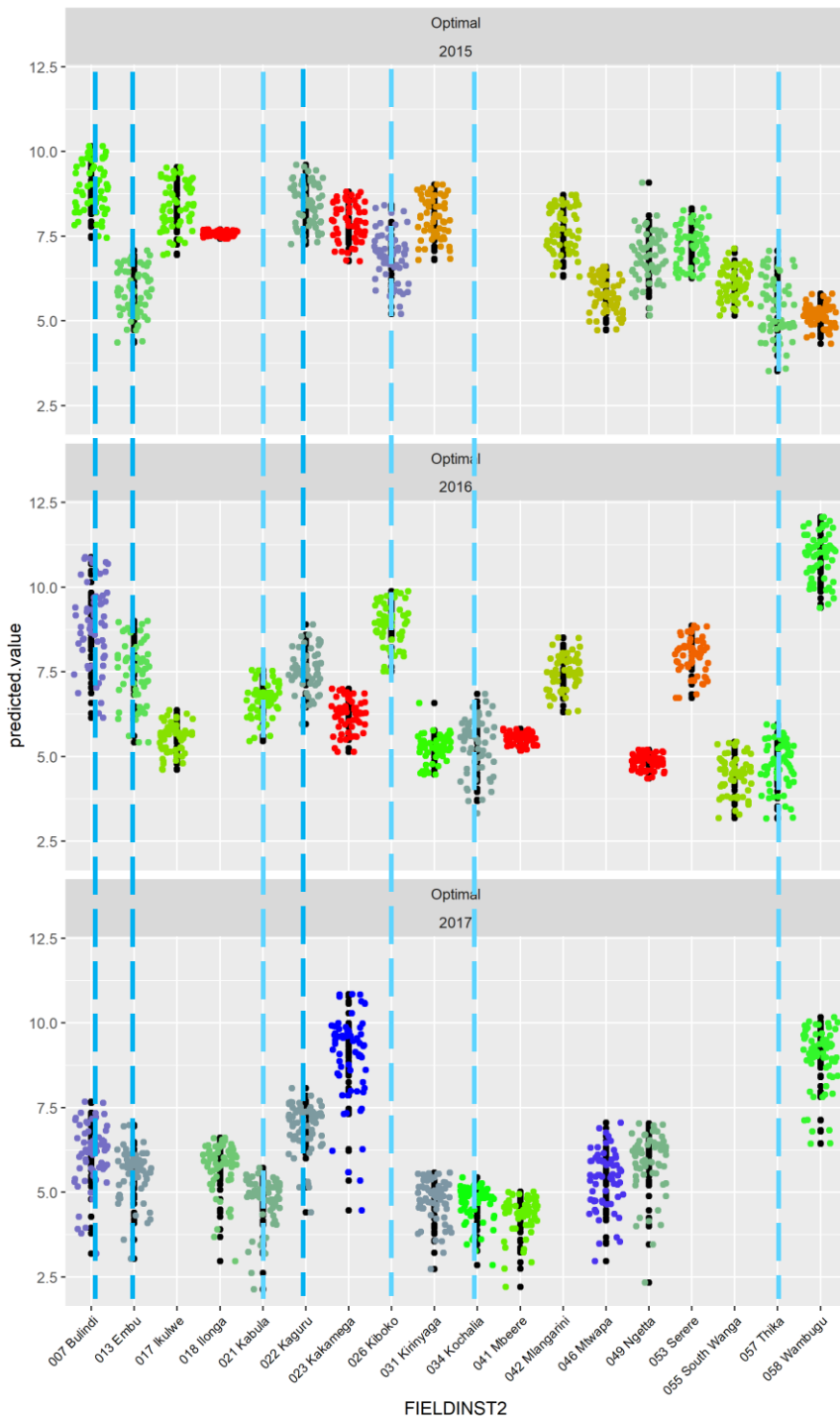
# Retrospective analysis for identifying optimal number of environment to sample TPE



Results from cross validation to know the accuracy between real (across the entire TPE) and estimated BV when selecting a given number of environments.



# Optimizing recycling through retrospective analysis to identify locations with high heritability and high genetic variance

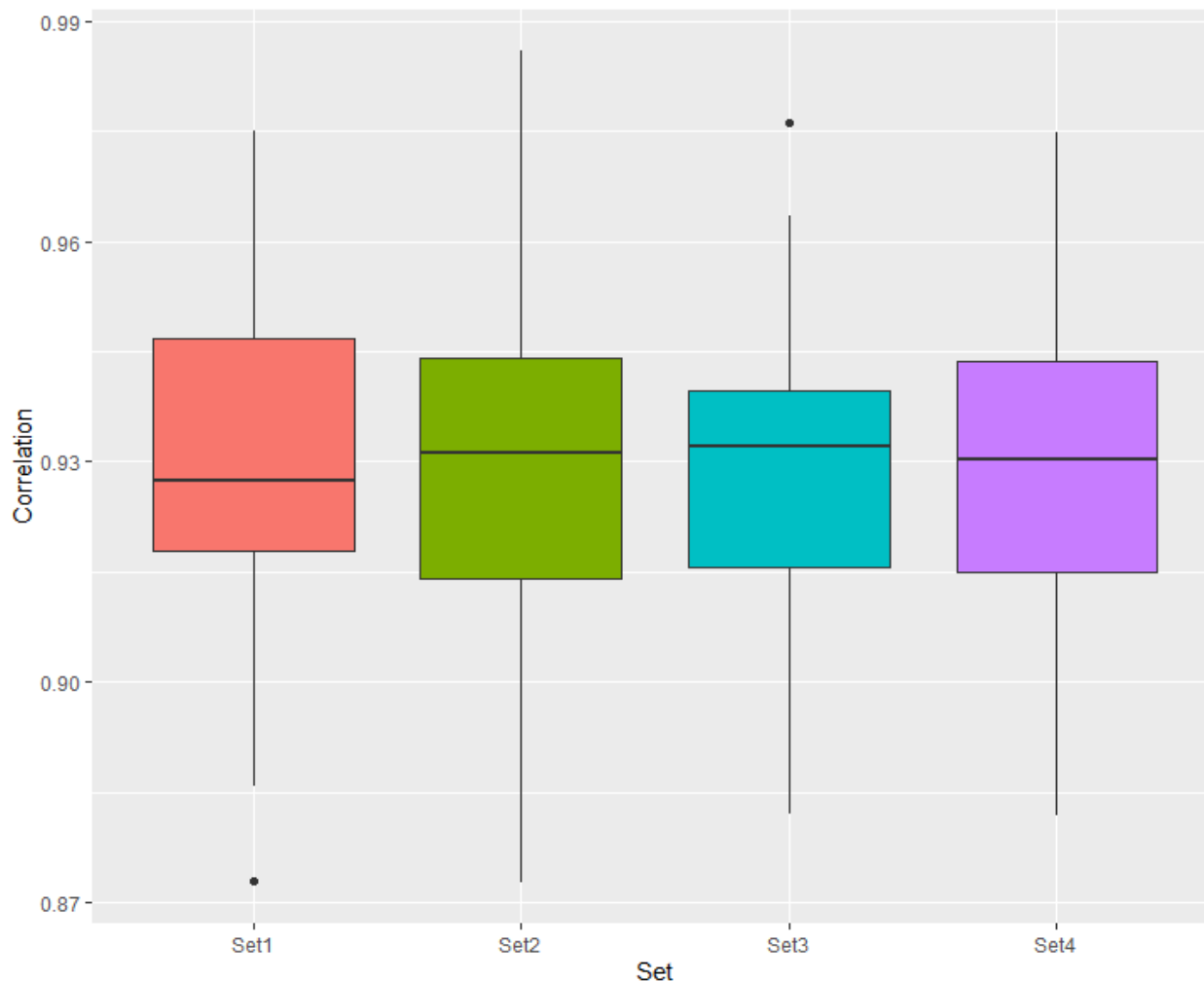


Results from single environment heritability across 3 years of data. We assume that environments with highest H2 and genetic variance represent the best locations for applying selection



# Results of sparse testcrossing

Option 3			
To phenotype			
Set	Lines	Testers	TC
1	48	1-2-3	144
2	85	1	85
3	85	2	85
4	85	3	85
Total			399
To predict			
Set	Lines	Testers	TC
1	85	2-3	170
2	85	1-3	170
3	85	1-2	170
Total			510



Spearman correlation between observed and predicted line-breeding value (n=50).  
 Set 1 line crossed with all testers, set 2-4 lines crossed with one tester, predicted with the other two testers.

# Lessons learned

1. Proper planning and coordination among
  - Conventional/ molecular breeder/Biometrician/ quantitative geneticist
2. Reliable service provider for efficient data turnaround time from leaf sampling to genotyping and data analysis.
3. Reasonable database for phenotypic and genotypic data storage/Links phenotypes with genotypes
4. Use of common genotyping platform to facilitate data sharing, reducing duplication and increase efficiency





**Thank you  
for your  
interest!**