Impact of low and mid-density markers genotyping on CIMMYT's tropical maize breeding

Sudha Nair, Xuecai Zhang, Manje Gowda
CIMMYT Global Maize Program

01-02-2024

Outline

- CIMMYT maize molecular breeding
- Genomic resources and genotyping platforms
- Trait marker pipeline: MABC and Forward breeding
- Genomic selection
- Marker-based quality control
- Final takeaways

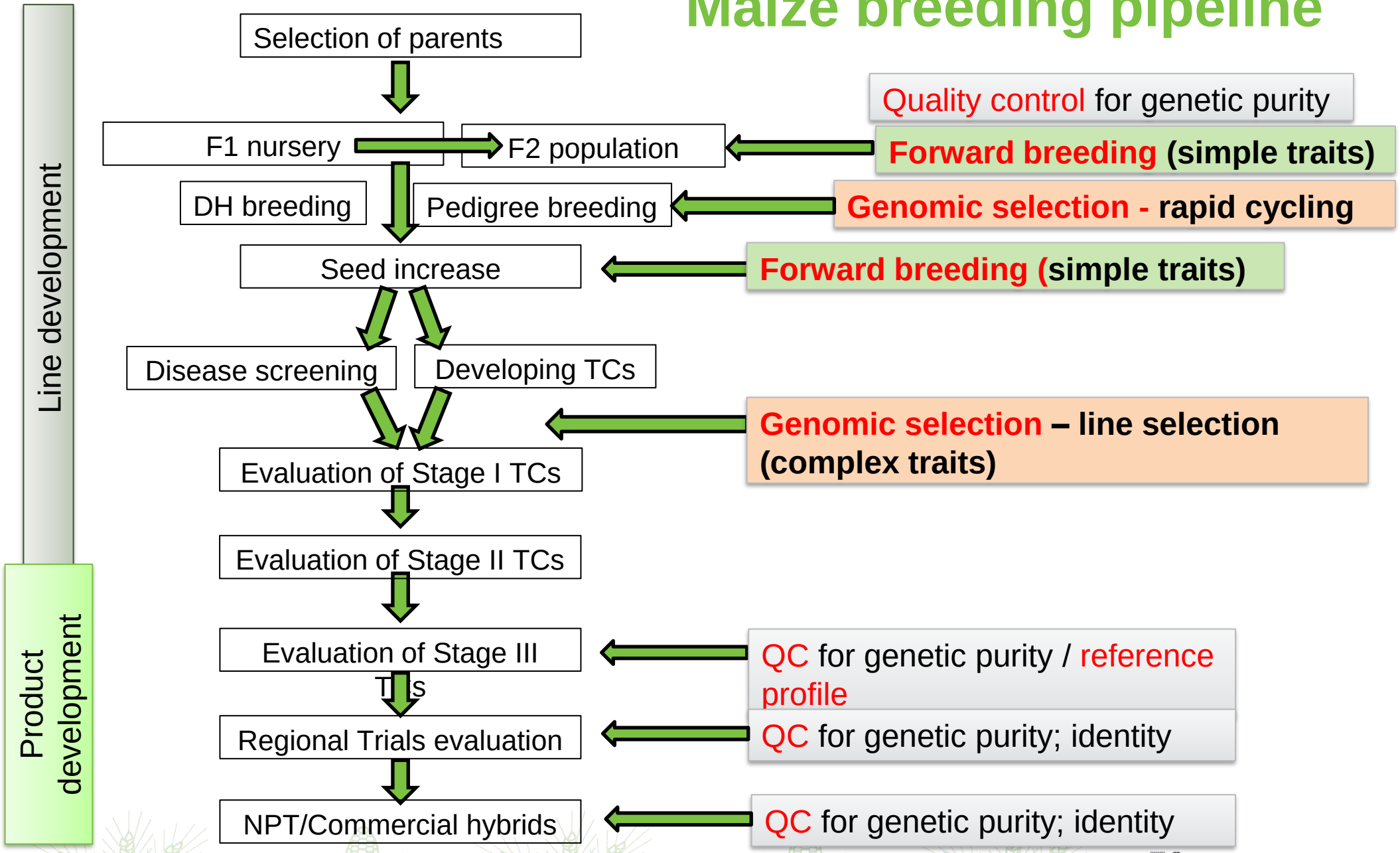
CIMMYT Maize Molecular Breeding

Latin America

E&S Africa

Asia

Maize breeding pipeline



CIMMYT

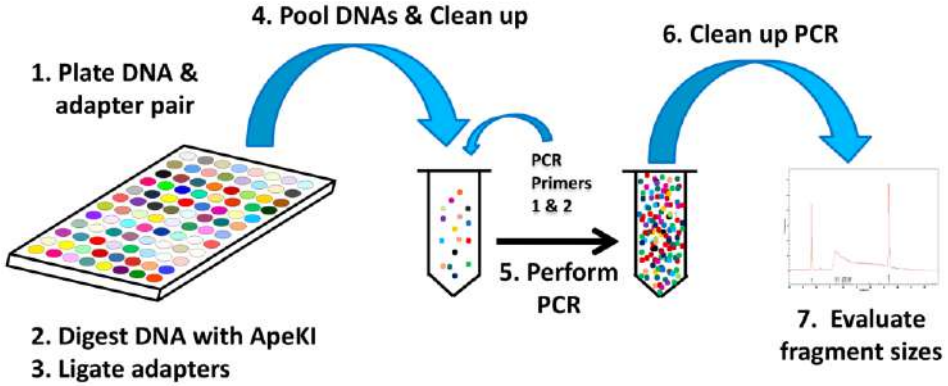
Genomic resources and genotyping platforms

Genotyping by sequencing

Genotyping-by-sequencing (GBS) is one of most widely used reduced representation sequencing methods

GBS has been extensively applied for the sequencing of more than 17,000 maize materials (https://www.panzea.org/) with ~1 Million imputed SNP calls mapped to reference genome

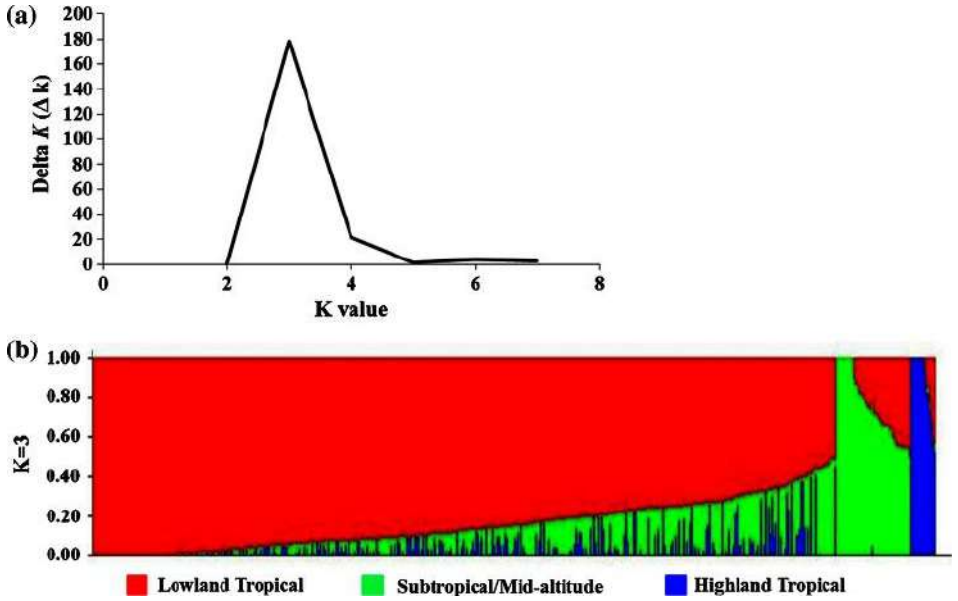
Three generations of maize haplotype maps are constructed using this data and have been applied to studies on many aspects of maize genetics.



(Elshire et al., 2011)

CIMMYT has assembled five association mapping panels which are genotyped with GBS, and used to map various traits of breeding significance

95% of released CIMMYT Maize Lines (CMLs) are genotyped using GBS



(Wu et al., 2011)

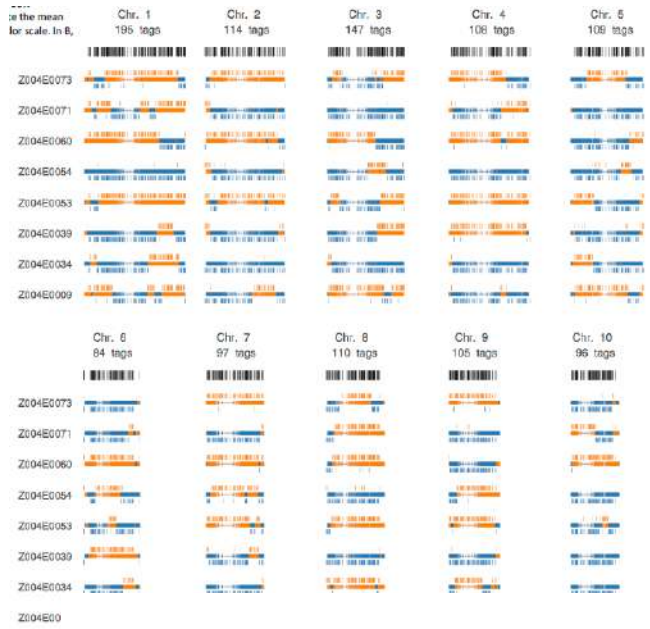
rAmpSeq genotyping

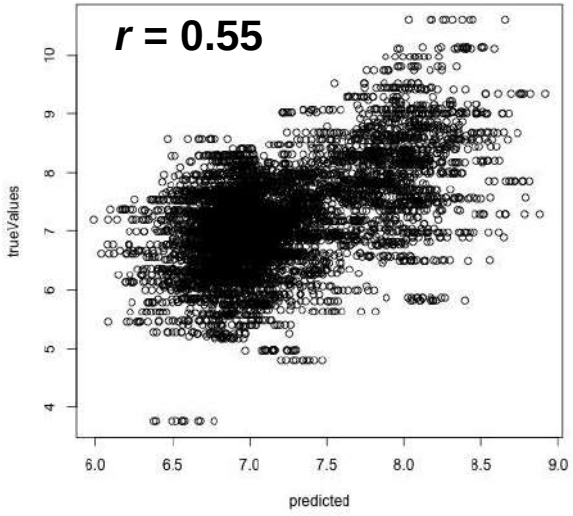
Methodology used conserved regions to design PCR primers for amplifying thousands of middle repetitive regions, followed by bioinformatic scans to identify DNA sequence polymorphisms in the low copy intervening sequences (Buckler et al., 2016).

Cost effective mid-density genotyping strategy for large-scale genomic selection projects.

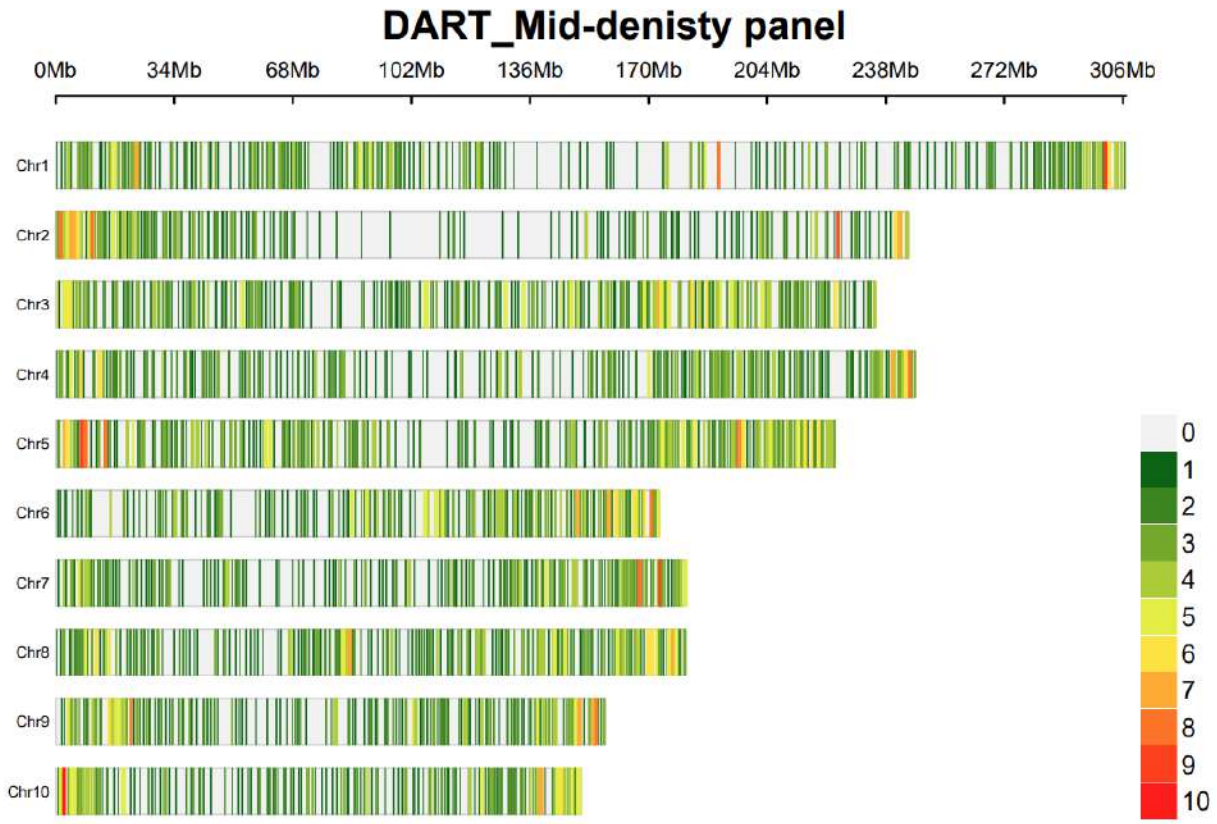
For maize samples genotyped for genomic selection, ~7000 tags were generated on an average

Routine use of genomic selection as a strategy in maize breeding pipelines initiated in 2017 employing this genotyping platform





DArTag Mid-density panel



Currently being used for genomic selection, genetic diversity studies, QTL mapping, reference profile

- ~7000 SNP sequences submitted to DArT for development of MD panel; Version 1 of the MD panel had 1898 genome-wide SNPs
- Version 2 (current) MD panel has 3305 genome wide SNP markers developed from sequencing data of >10,000 breeding lines and landraces belonging to different breeding programs from Latin America, Africa and Asia
- Mainly derived from the genomics resources from CIMMYT and IITA including whole-genome re-sequencing (WGS), genotyping by sequencing (GBS), DArTseq genotyping and maize HapMap3
- The average marker density of the panel is about 1 SNP per 0.72 Mbp, distributed across 10 chromosomes
- Comprises trait markers from several discovery studies in maize along with random markers spread across the genome, including the 70 QC KASP markers

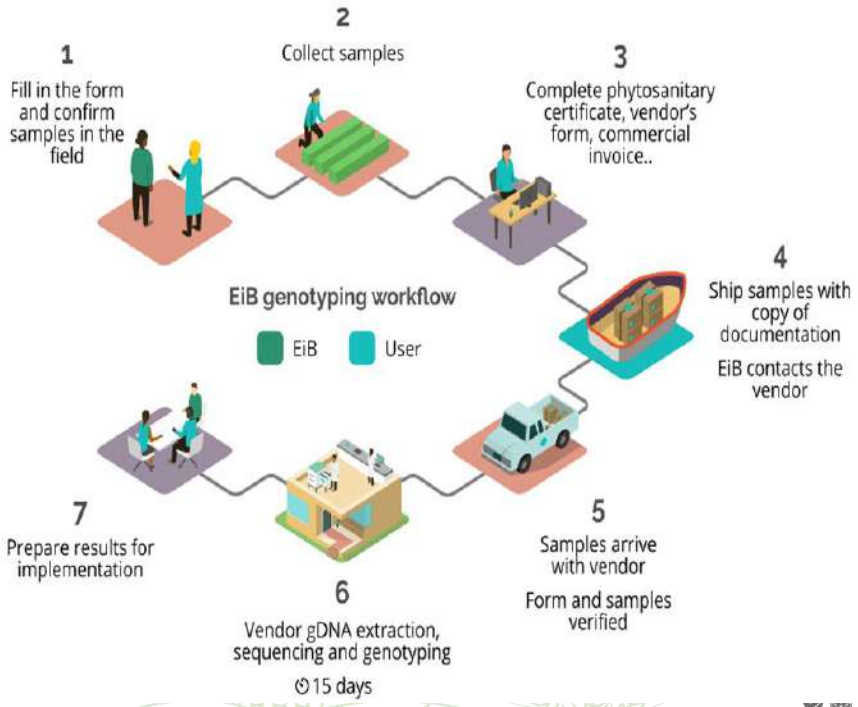
https://excellenceinbreeding.org/toolbox/services/maize-mid-density-genotyping-services

CIMMYT

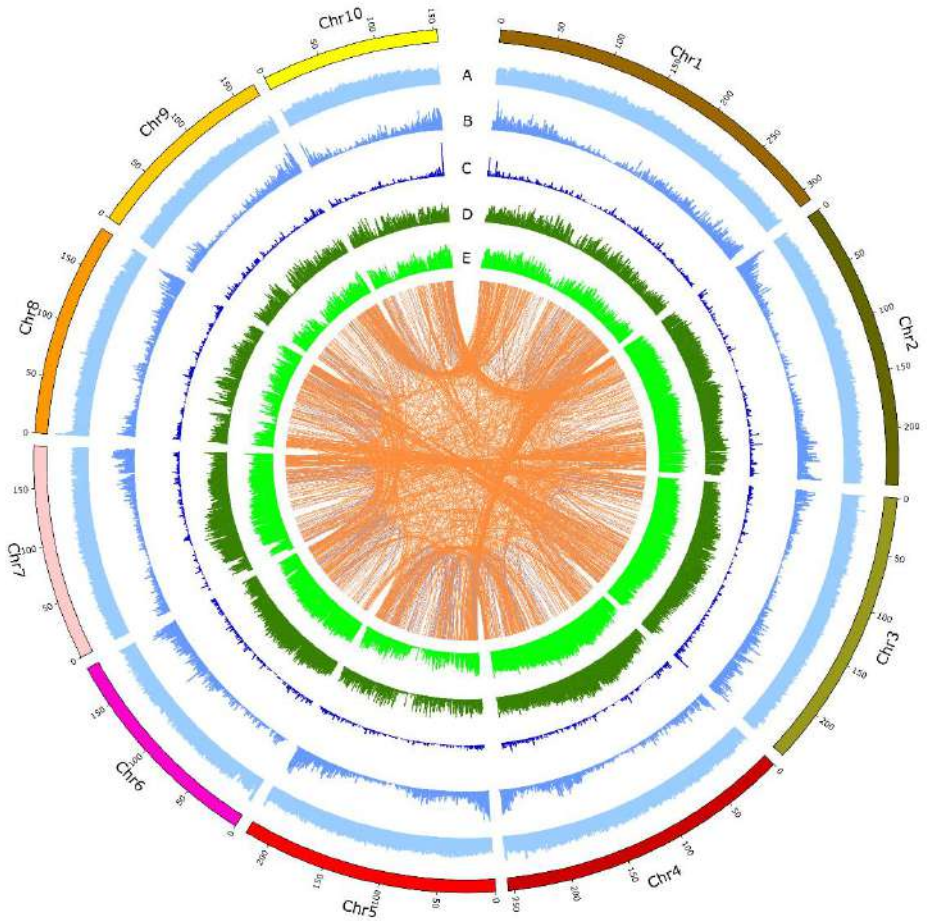
Maize low-density SNP sets

- The EiB low-density genotyping service is a shared KASPTM genotyping platform serving all CGIAR centers and their partner programs KASP markers for maize_v2.xlsx
- Includes trait markers for breeding-relevant traits for different tropical product profiles which are outcomes of trait mapping studies of breeding-relevant traits at CIMMYT and IITA
- Includes a common set of 50 SNPs used in quality control at various stages across all maize breeding pipelines

Applications:
Marker assisted back crossing
Forward breeding
Quality control



Genome assembly of CML495 and re-sequencing of 90 key CMLs



Overview of CML495 genome assembly.
A transposable element; **B** gene density; **C** presence and absence (PAV) relative to B73_RefGen_v4; **D** SNPs relative to B73_RefGen_v4; **E** SVs relative to B73_RefGen_v4; and synteny with lines representing paralogous gene pairs between and within pseudo-chromosomes in CML495 (purple and orange lines).

Genomic feature	CML495
Length of CML495 assembly (bp)	2,208,341,065
Length of 10 pseudochromosomes (bp)	2,162,116,483
Maximum scaffold length (bp)	310,254,704
N50 of scaffold (bp)	221,728,245
LTR assembly index, LAI	24.40
Complete BUSCOs (%)	96.50
Number of unanchored contigs	1,031
Number of genes	37,947
Number of genes in 10 pseudochromosomes	37,879
Number of transcripts	47,760
Genes with RNA-seq support	29,758
Total size of transposable elements (bp)	1,677,097,145

CIMMYT

Re-sequenced CMLs and elite DT lines

Subgroup1 Dought Tolerance 19 Lines	
CML343	LAPOSTASEQ-C3-FS17-1-2-3-2-1-B
CML405	LAPOSTASEQ-C0-B*3-12-1-1-B
CML477	LAPOSTASEQ-C3-B-B-1-2-B
CML488	DTPW-C8-F31-4-2-1-5-B
CML494	LAPOSTASEQ-C4-F7-1-2-2-2-2-B
CML564	DTPY-C9-F46-3-9-1-2-2-1-3-B*7
CML566	(LAPOSTASEQ-C7-F96-1-2-1-1-B*3/CML444//CML444)-DH16-B
CML567	(LAPOSTASEQ-C7-F71-1-2-1-2-B*3/CML539//CML539)-DH3-B
CML568	(LAPOSTASEQ-C7-F71-1-2-1-2-B*3/CML539//CML539)-DH20-B
CML576	(CLFAWW11/CML494)-B-24-2-2-B-B-1-B-8-B-B
CML592	(CML536/DTPW-C9-F109-2-6-1-1-B)-DH13-B-B
DPW9F104	DTPWC9-F104-5-4-1-1-B-B
DTPY9F46	DTPYC9-F46-1-2-1-2-B-B-B
DTPY9F74	DTPYC9-F74-1-1-1-1-B-B-B-B
LPSC7F64	La Posta Sequia C7 F64-2-6-2-2-B-B-B
CML444	P43-C9-1-1-1-1-1-B
CML544	((CML395/CML444)-B-4-1-3-1-B/CML444//(TUXPSEQ-C1-F2/P49SR)-F2-45-7-1-2-B)-2-1-2-2-B
CML522	CML444 IR P43-C9-1-1-1-1-1-B
CML373	P43SR-4-1-1-2-1-B-8-1-B

Over 600 candidate genes relevant to DT tolerance

Zm00001d050107	opr8	12-oxophytodienoate reductase 3
Zm00001d004358	abi28	ABI3VP1 transcription factor (Fragment)
Zm00001d051194	adc2	Arginine decarboxylase
Zm00001d030803	ZmARF2	Auxin response factor
Zm00001d030801	ZmARF1	Auxin response factor
Zm00001d004384	saur21	Auxin-induced protein 10A5
Zm00001d049659	saur37	auxin-responsive protein SAUR21-like
Zm00001d030955	ZmCBL2-1	Calcineurin B-like protein 3
Zm00001d015743	ZmHy5	cAMP response element binding (CREB) protein
Zm00001d049889	DREB2C	Dehydration-responsive element-binding protein 2C
Zm00001d014113	ERF096	Ethylene-responsive transcription factor ERF098
Zm00001d025409	RAP2-6/ereb21	Ethylene-responsive transcription factor ERF115
Zm00001d032223	ga20ox6	Fe2OG dioxygenase domain-containing protein
Zm00001d009646	scl1/ZmGRAS11	GA repressor DELLA
Zm00001d042187	grx8	Glutaredoxin domain-containing protein
Zm00001d019429	mybr30	Glutathione S-transferase T3
Zm00001d047491	jmj3	Growth-regulating factor 5
Zm00001d016381	hdac1	Histone deacetylase
Zm00001d050074	PEPR1	Leucine-rich repeat receptor-like protein kinase PEPR1
Zm00001d024268	nac110	NAC domain-containing protein
Zm00001d010743	ZmCIPK19	Non-specific serine/threonine protein kinase
Zm00001d005484	ZmPLDδ3/pld12	Phospholipase D
Zm00001d033552	mha13	Plasma membrane ATPase
Zm00001d025303	hak19	Potassium transporter
Zm00001d050069	ZmTPS7.2/trps8	probable alpha,alpha-trehalose-phosphate synthase [UDP-forming] 9
Zm00001d005003	cal4	Putative calmodulin-like protein 2
Zm00001d011669	myb124	Putative MYB DNA-binding domain superfamily protein
Zm00001d028759	pdc3	Pyruvate decarboxylase
Zm00001d042188	SAPK3/SnRK2-10	Serine/threonine-protein kinase SAPK3
Zm00001d023931	myb109	Transcription factor MYB83 (Fragment)
Zm00001d021191	ZmbZIP22/vip1	Transcription factor RF2a
Zm00001d004383	lug6	WD40 repeat-containing protein HOS15
Zm00001d039245	WRKY6	WRKY domain-containing protein
Zm00001d047309	wrky61	WRKY53 transcription factor

Trait markers in breeding pipelines

Trait marker pipeline

Stage-gate 1

Stage-gate 2

Stage-gate 3

Discovery project proposal

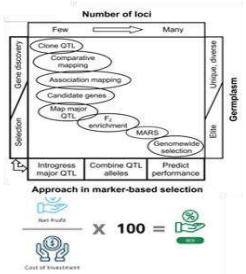
Discovery

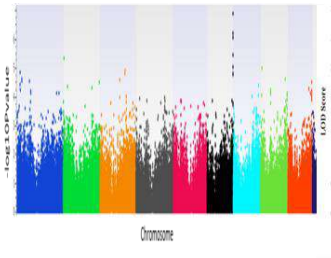
Validation in independent populations

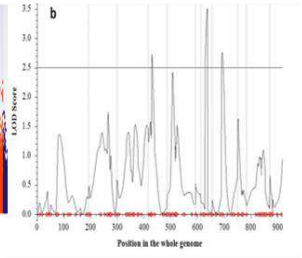
Validation in breeding populations

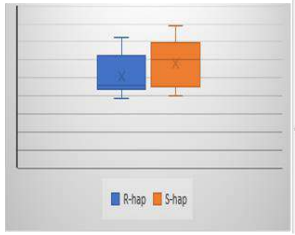
Technical validation of genotyping assays

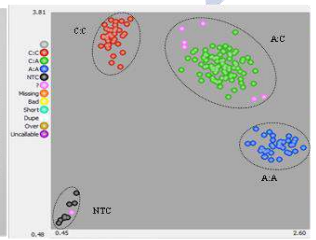
Deployment











Fusarium Ear Rot
Fall army worm

Fusarium Ear Rot
Common Rust
Fall Army Worm

Gray Leaf Spot
Kernel Zinc

Turcicum Leaf Blight
Tar Spot Complex

Maize Streak Virus
Maize Lethal Necrosis
Pro-Vitamin A
Tar spot complex

CGIAR

CIMMYT MR

Molecular marker deployment strategies

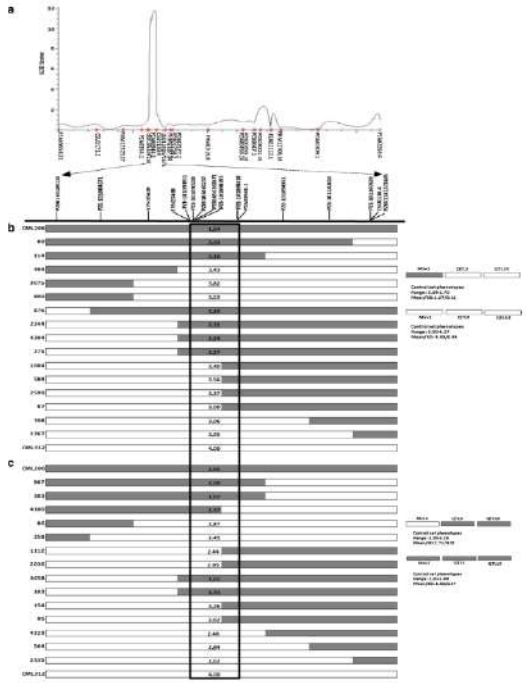
Marker assisted back crossing	MAS to transfer high value rare alleles in the breeding pool from trait donors Favorable allele frequency: null to low Effect size: large
Forward breeding	MAS in routine breeding crosses for increasing the frequency of specific high-value alleles in breeding populations Favorable allele frequency: medium to high Effect size: moderate to large
Genomic selection	Prediction of breeding value based on genome-wide markers Favorable allele frequency: medium to high Effect size: multiple small effect

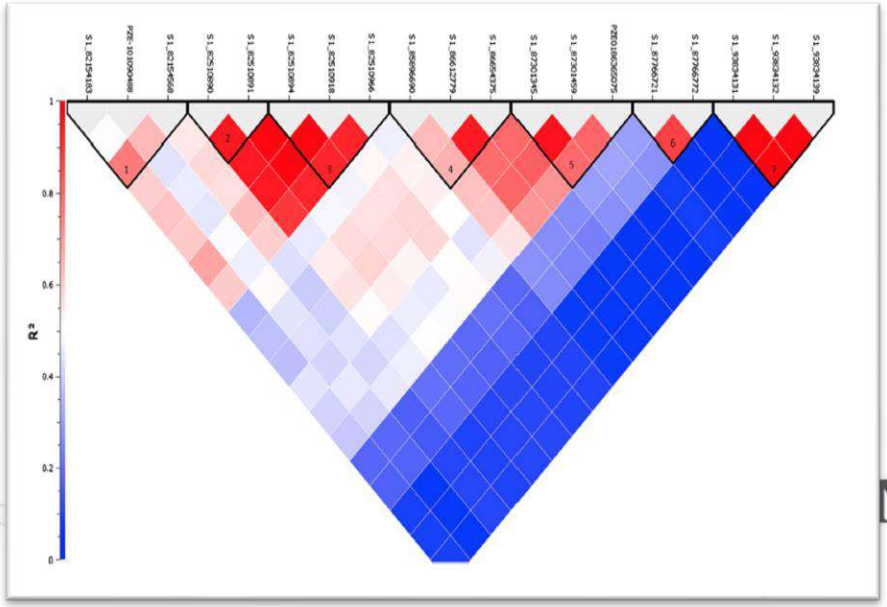
CGIAR

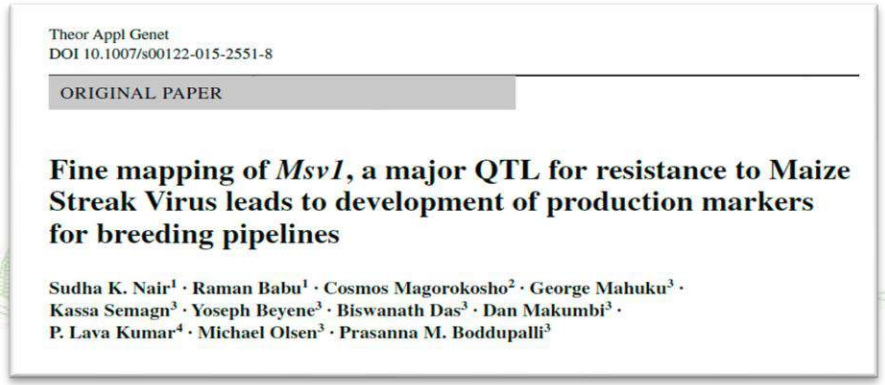
CIMMYT MR

Production markers for *Msv1* for Maize streak virus

- Followed a QTL isogenic recombinant (QIR) strategy to fine map to a genetic interval of 0.87 cM
- Validation confirmed 94% accuracy for selection with favourable haplotype
- Routine use in forward breeding







Theor Appl Genet
DOI 10.1007/s00122-015-2551-8

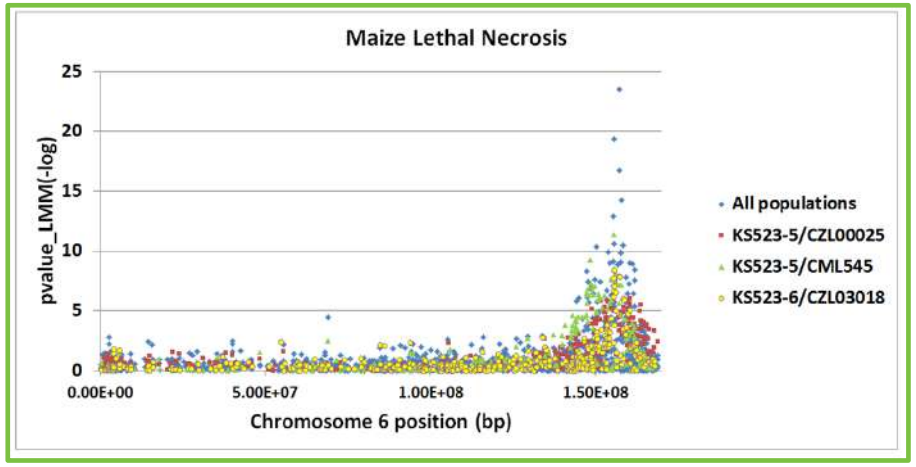
ORIGINAL PAPER

Fine mapping of *Msv1*, a major QTL for resistance to Maize Streak Virus leads to development of production markers for breeding pipelines

Sudha K. Nair[1] · Raman Babu[1] · Cosmos Magorokosho[2] · George Mahuku[3] · Kassa Semagn[3] · Yoseph Beyene[3] · Biswanath Das[3] · Dan Makumbi[3] · P. Lava Kumar[4] · Michael Olsen[3] · Prasanna M. Boddupalli[3]

Production markers for MLN: *qMLN_06.157* discovery

- **Modified QTLseq experiment**
- **Three populations involving KS23 sources**
- **Tissue sampled thousands of F2 plants**
- **Inoculated plants**
- **Genotyped most resistant and susceptible individuals**



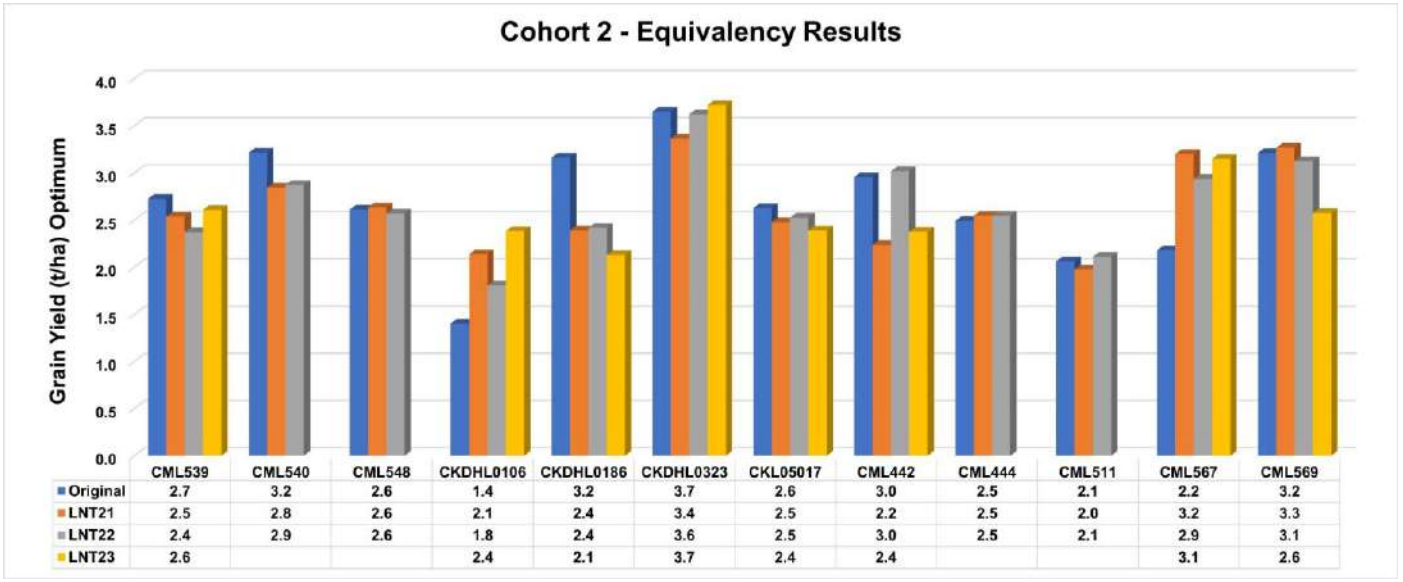
Special thanks to Diversity Arrays Technology and IGSS for genotyping support

CGIAR

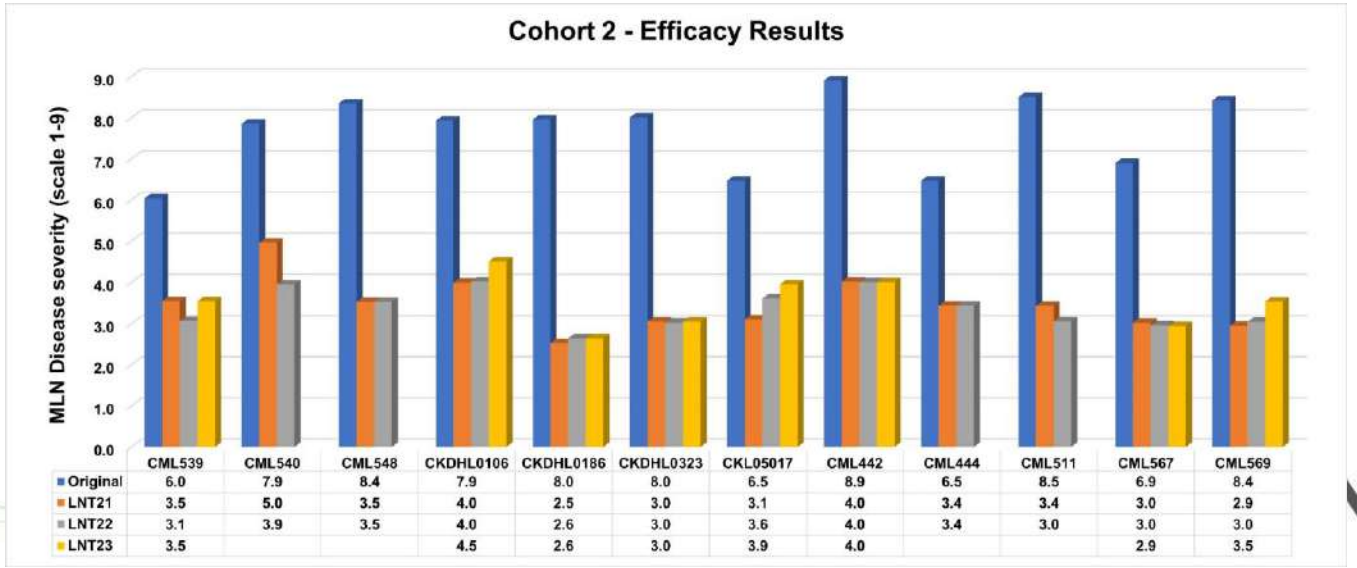
CIMMYT MR

MABC for *qMLN_06.157*: TI pipeline

>110 elite lines have been introgressed with *qMLN_06.157*

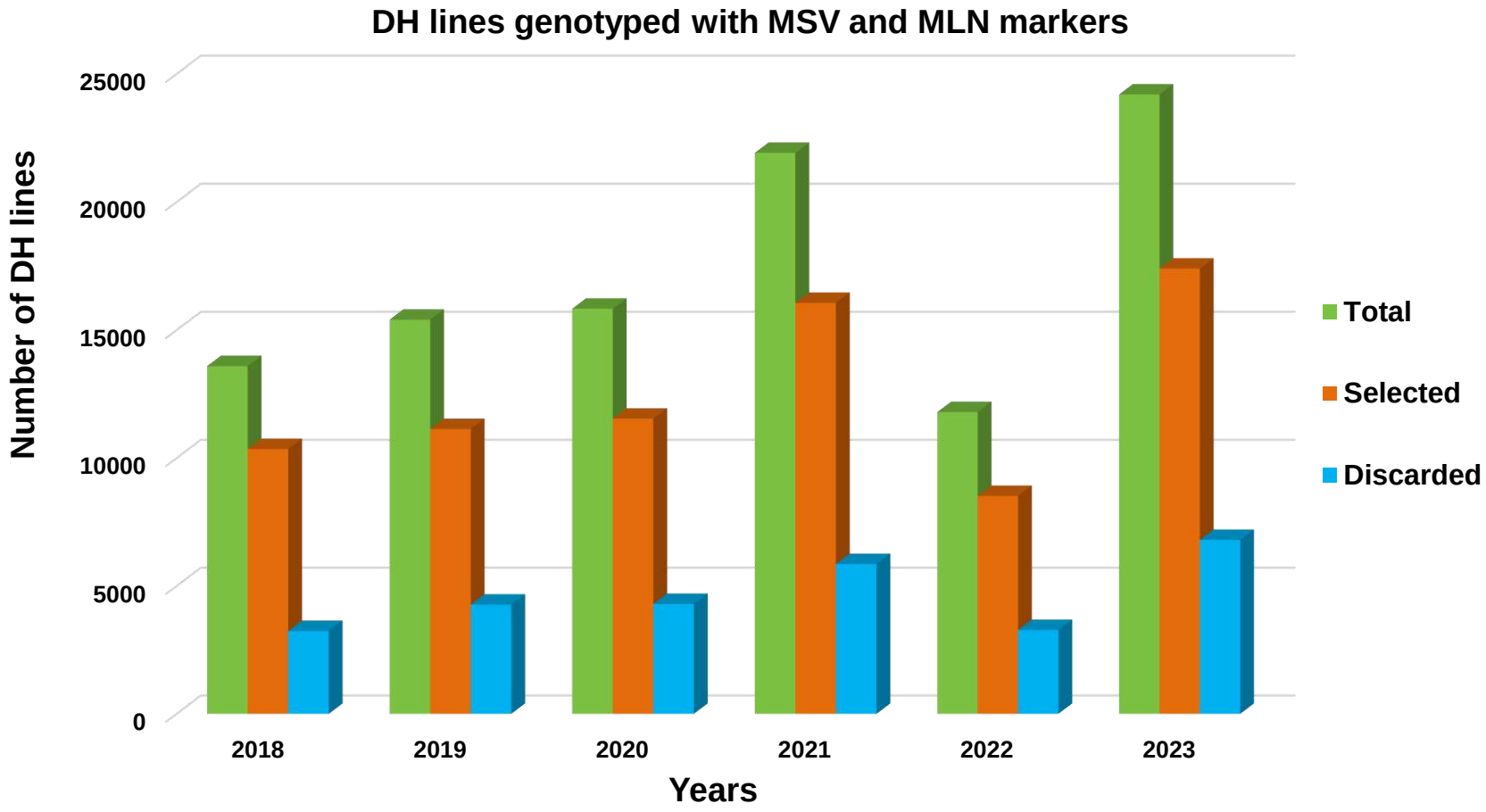


	CML539	CML540	CML548	CKDHL0106	CKDHL0186	CKDHL0323	CKL05017	CML442	CML444	CML511	CML567	CML569
Original	2.7	3.2	2.6	1.4	3.2	3.7	2.6	3.0	2.5	2.1	2.2	3.2
LNT21	2.5	2.8	2.6	2.1	2.4	3.4	2.5	2.2	2.5	2.0	3.2	3.3
LNT22	2.4	2.9	2.6	1.8	2.4	3.6	2.5	3.0	2.5	2.1	2.9	3.1
LNT23	2.6			2.4	2.1	3.7	2.4	2.4			3.1	2.6



	CML539	CML540	CML548	CKDHL0106	CKDHL0186	CKDHL0323	CKL05017	CML442	CML444	CML511	CML567	CML569
Original	6.0	7.9	8.4	7.9	8.0	8.0	6.5	8.9	6.5	8.5	6.9	8.4
LNT21	3.5	5.0	3.5	4.0	2.5	3.0	3.1	4.0	3.4	3.4	3.0	2.9
LNT22	3.1	3.9	3.5	4.0	2.6	3.0	3.6	4.0	3.4	3.0	3.0	3.0
LNT23	3.5			4.5	2.6	3.0	3.9	4.0			2.9	3.5

Forward Breeding for MLN and MSV

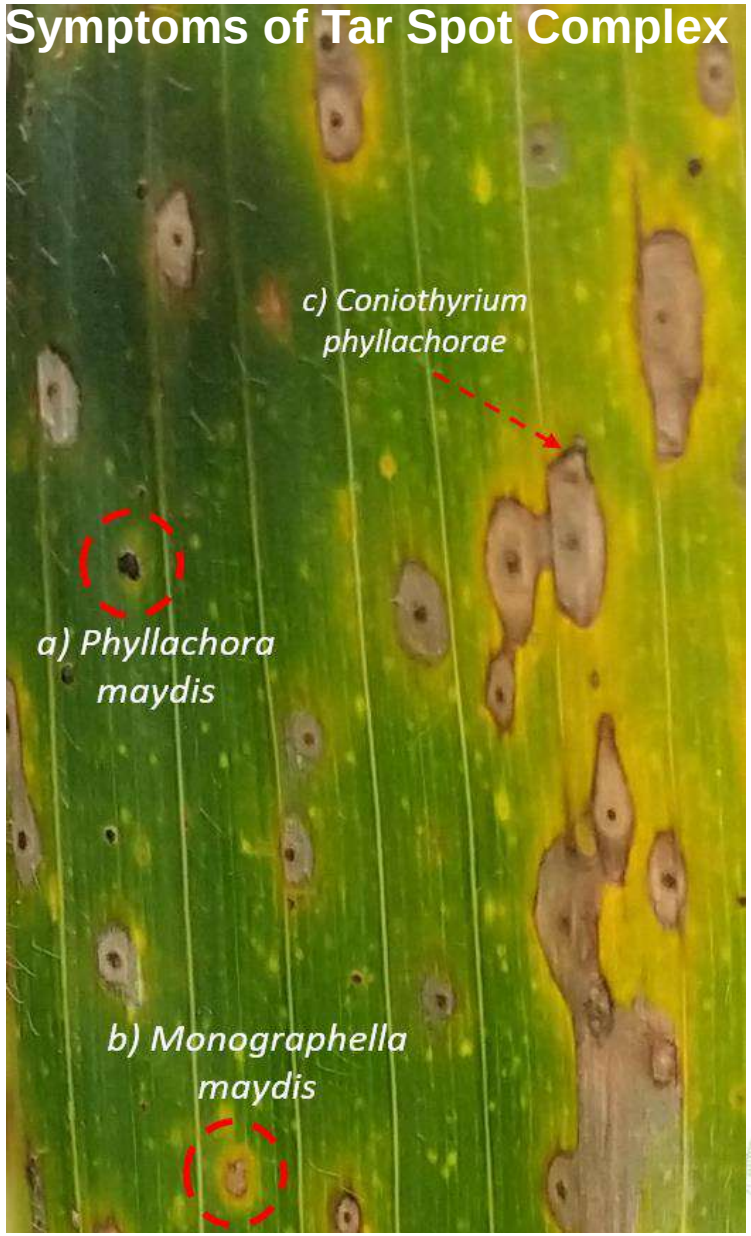


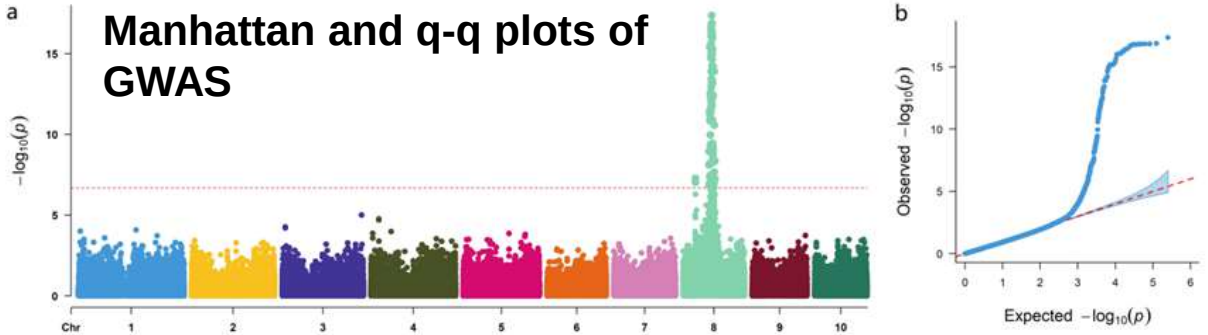
Forward breeding -Total number of lines genotyped and selected for MLN and MSV resistance haplotypes from last six years in ESA breeding pipelines

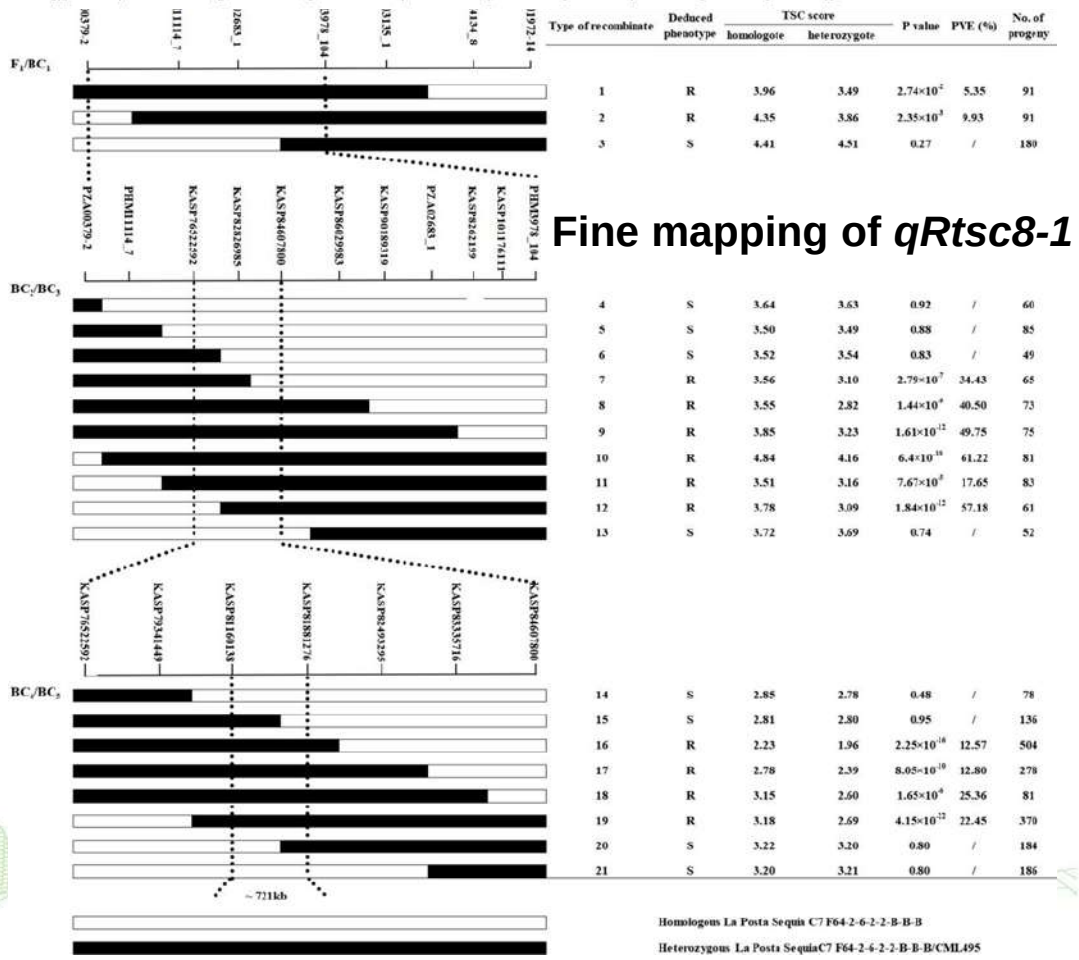
CGIAR

CIMMYT

Production markers for *qRtsc8-1* for maize tar spot complex







- ❑ Donor line of CML495 - HGA
- ❑ *qRtsc8-1* in an ~ 721 kb interval between production markers of KASP81160138 and KASP81881276
- ❑ Routine use in forward breeding
- ❑ Candidate genes *GRMZM2G063511* and *GRMZM2G073884* were identified

CIMMYT

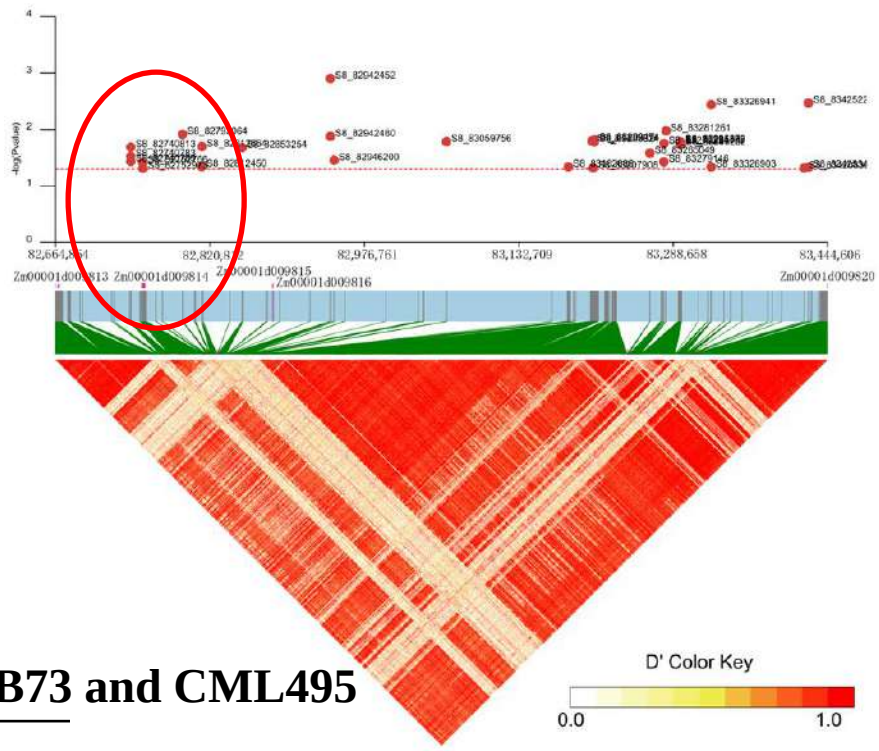
Fine mapping of *qRtsc8-1* for resistance to Tar spot complex

Pos.	CML495	B73	Pos.	CML495	B73
3	H	Q	532	S	N
11	S	G	542	V	I
18	A	P	561	N	H
141	S	N	589	S	G
151	L	V	744	N	T
191	Y	H	748	-	T
239	H	R	798	S	L
282	-	P	923	M	I
287	L	I	981	S	T
329	A	V	985	-	S
362	I	M	1053	V	I
444	P	S	1076	T	I
453	K	T	1093	H	D
486	G	S	1118	Q	L

Mutation site of *Zm00001d009814* between B73 and CML495

SAAS	Predicted_score	Predicted_class
R239H	0.209	neutral
T453K	0.251	neutral
S486G	0.441	neutral
I923M	0.724	functional
T981S	0.107	neutral
I1053V	0.642	functional
I1076T	0.798	functional
D1093H	0.752	functional

PPVED prediction of single amino acid mutations

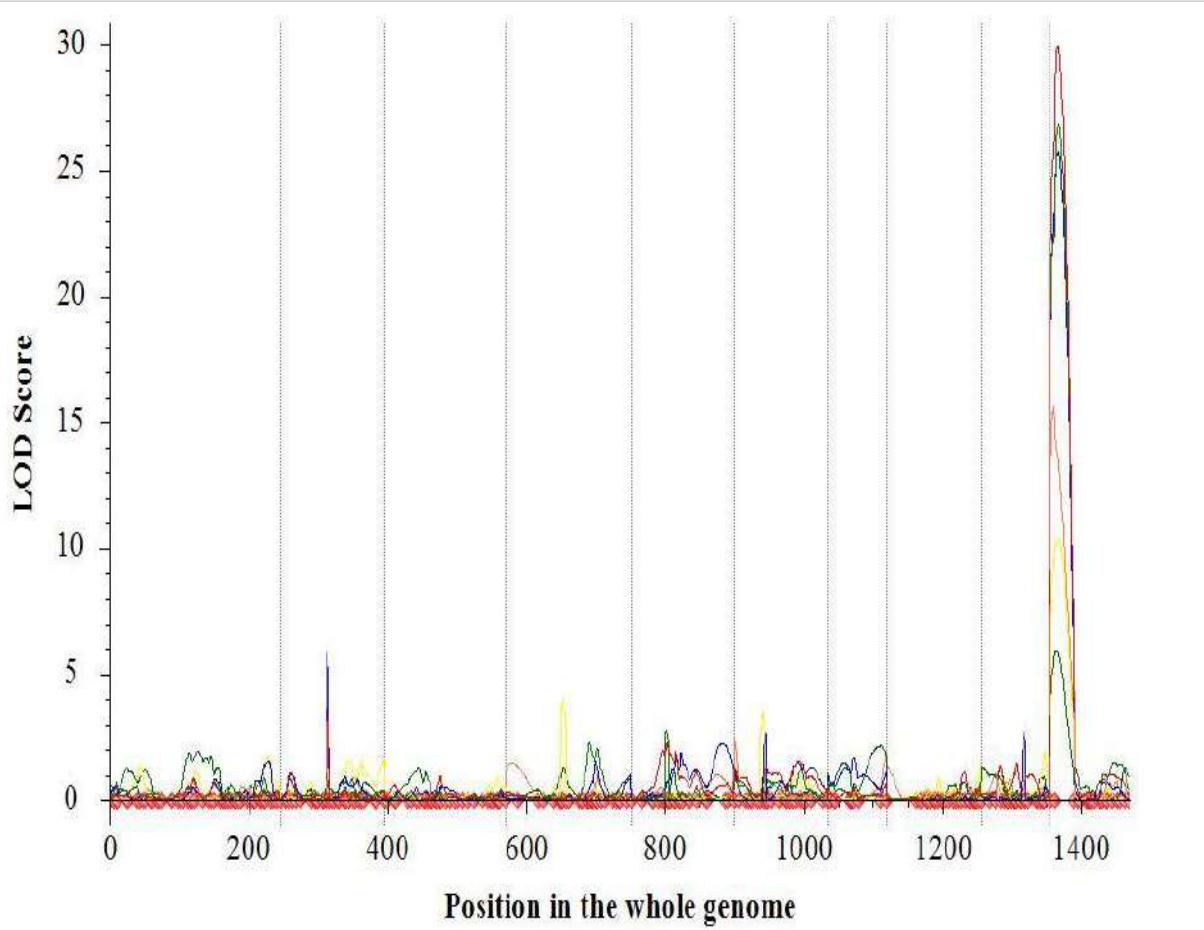


Candidate gene-based GWAS in the qRtsc8-1 fine-mapping region of 721 kb including *Zm00001d009814*, a leucine-rich repeat receptor-like protein kinase

TSC: Discovery stage of *qRtsc10-1*

qRtsc10-1 from CML576 detected by QTL mapping in (CML576/CLWN244) DH population with 225 DH lines evaluated in 7 reps

Locs	Reps	Chr.	LeftMarker (position)	RightMarker (position)	LOD	PVE (%)	Add
Chiapas20B-1	2	10	303(2090851)	304 (5051173)	27	38	-0.46
Chiapas20B-2	2	10	303 (2090851)	304 (5051173)	26	35	-0.32
Chiapas17B-1	1	10	303(2090851)	304 (5051173)	10	17	-0.69
Colombia19B-1	1	10	303(2090851)	304 (5051173)	6	12	-0.39
Colombia19B-2	1	10	302 (1206271)	303 (2090851)	16	24	-0.36
Chiapas20B-1 & 2	4	10	303 (2090851)	304 (5051173)	30	40	-0.35
Across locations	7	10	303(2090851)	304 (5051173)	30	40	-0.35



TSC score:

CML576 ranged from 1 to 4 across locations, with an average of 2.23;

CLWN244 ranged from 5 to 9 across locations, with an average of 7.13;

DH lines ranged from 3.18 to 7.62 across locations, with a heritability of 0.85, and LSD of 0.68.

Populations	No. of samples	snpZM00263	snpZM00266	snpZM00267	snpZM00286	snpZM00288	snpZM00269	snpZM00270	snpZM00289
((CML576/CLWN244)@40-B/CLWN244)-13	92	4.28e-08 ***	7.66e-13 ***	3.64e-12 ***	2.4e-11 ***	2.4e-11 ***	1.48e-10 ***	2.39e-07 ***	4.24e-05 ***
((CML576/CLWN244)@233-B/(CML576/CLWN244)@15-B)-68	105	0.523	5.03e-12 ***	8.12e-12 ***	5.2e-09 ***	5.2e-09 ***	1.53e-10 ***	3.29e-09 ***	5.09e-05 ***
((CML576/CLWN244)@10-B/(CML576/CLWN244)@15-B)-35	143	0.179	3.8e-08 ***	4.04e-09 ***	5.76e-08 ***	3.86e-08 ***	2.07e-06 ***	2.18e-05 ***	0.688
((CML576/CLWN244)@40-B/(CML576/CLWN244)@15-B)-27	156	0.00464 **	0.00051 ***	0.000179 ***	0.00483 **	0.00483 **	0.00133 **	0.000715 ***	NA
((CML576/CLWN244)@127-B/(CML576/CLWN244)@211-B-48	165	5.47e-08 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	2.58e-12 ***	2.03e-09 ***	1.14e-05 ***
((CML576/CLWN244)@158-B/(CML576/CLWN244)@15-B)-26	157	1.99e-06 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	0.0157 *
((CML576/CLWN244)@158-B/(CML576/CLWN244)@167-B)-6	167	4.47e-06 ***	1.81e-12 ***	1.01e-13 ***	NA	NA	NA	NA	NA
((CML576/CLWN244)@233-B/(CML576/CLWN244)@167-B)-13	136	0.813	2.83e-07 ***	1.34e-12 ***	0.571	0.571	0.569	0.571	
((CML576/CLWN244)@76-B/(CML576/CLWN244)@167-B)-3	104	NA	NA	NA	0.541	0.456	0.119	0.411	0.627
((CML576/CLWN244)@76-B/(CML576/CLWN244)@15-B)-35	80	NA	NA	NA	NA	NA	NA	NA	0.0882
((CML576/CLWN244)@92-B/(CML576/CLWN244)@211-B)-81	132	0.323	NA	NA	NA	NA	NA	NA	NA

Intertek SNP ID *)	Customer SNP ID	Comments
snpZM00262	S10-219993	1, bad quality
snpZM00263	S10-826897	2
snpZM00264	S10-1475501	
snpZM00265	S10-1501085	
snpZM00266	S10-2246125	3
snpZM00267	S10-3599696	4
snpZM00268	S10-3855495	5, bad quality
snpZM00285	S10-4127370	
snpZM00286	S10-4163326	6
snpZM00287	S10-4183669	
snpZM00288	S10-4184535	7
snpZM00269	S10-4262659	8
snpZM00270	S10-4336876	9
snpZM00289	S10-4478412	10
snpZM00271	S10-5178159	

TSC: Fine mapping of *qRtsc10-1*

Marker-trait association in fine-mapping populations

snpZM266 (2.24 Mb)- snpZM267 (3.60 Mb), 1.36 Mb interval

snpZM263 (0.83 Mb)- snpZM267 (4.16 Mb), 3.33 Mb interval

CGIAR

CIMMYT

TSC: Fine mapping of qRtsc10-1

Marker validation in 499 breeding lines, including 279 CMLs and 220 coded breeding lines

KASP markers	Favorable allele[b]				Unfavorable allele				P-value[e]	
	Allele	Frequency	TSC	AUDPC	Allele	Frequency	TSC	AUDPC	TSC	AUDPC
K10-219993	A	0.56	5.13	57.16	C	0.44	5.29	60.06	0.0563	0.0029**
K10-826897	T	0.52	5.15	57.69	A	0.48	5.29	59.70	0.1083	0.0401*
K10-1475501	A	0.38	5.16	58.00	T	0.62	5.23	58.79	0.4347	0.4297
K10-1501085	G	0.77	5.19	58.09	C	0.23	5.25	59.79	0.5496	0.1442
K10-2028072	T	0.43	5.14	57.16	C	0.57	5.27	59.66	0.1406	0.0120*
K10-2246125	G	0.60	5.17	57.77	T	0.40	5.27	59.70	0.2262	0.0518
K10-3599696	T	0.45	5.11	57.95	C	0.55	5.30	59.17	0.0259*	0.2147
K10-3855495	A	0.53	5.15	57.41	T	0.47	5.26	59.59	0.1811	0.0251*
K10-4262659	A	0.54	5.19	58.80	C	0.46	5.22	58.15	0.6755	0.5015
K10-4336876	C	0.64	5.16	57.49	G	0.36	5.28	60.30	0.1772	0.0058**

Genomic selection in breeding pipelines

Three Potential GS Applications in Maize Breeding

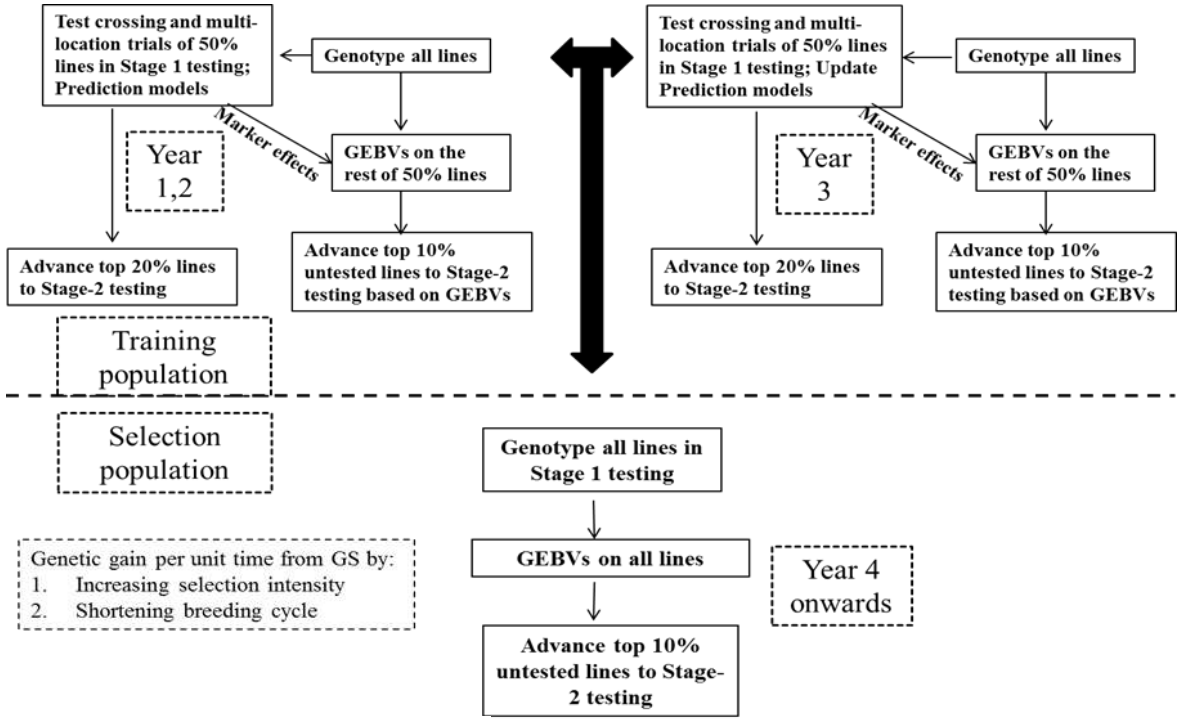
1. Prediction and selection of untested inbred lines
2. Prediction and selection of untested hybrids
3. Population improvement through rapid cycle recurrent selection

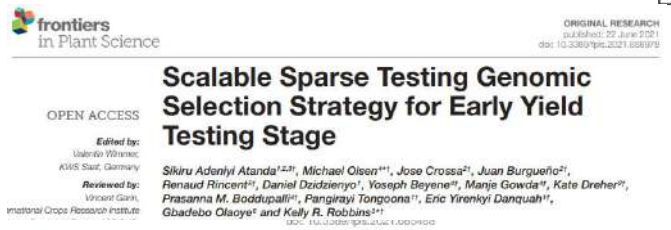
CGIAR

CIMMYT

Initial strategy for GS in maize breeding pipelines

- Final products from maize breeding pipelines are hybrids: three-way cross hybrids in ESA and single cross hybrids in Asia
- Initial strategy followed was test-half predict-half strategy
- 40-50% of the lines were test crossed and evaluated
- All lines were genotyped at mid-density and GEBVs estimated for the remaining 50% of the lines
- Lines with best BVs (10-20%) were selected for stage II evaluation
- This saves 30-45% resources with same selection efficiency

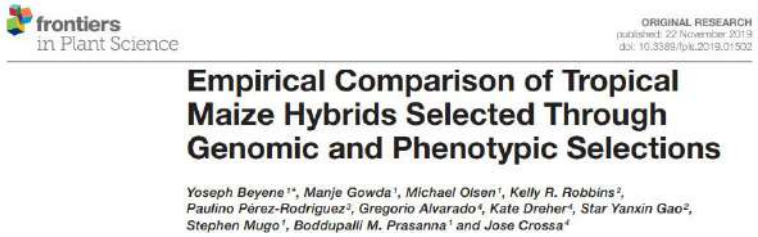



frontiers in Plant Science

ORIGINAL RESEARCH

OPEN ACCESS

Scalable Sparse Testing Genomic Selection Strategy for Early Yield Testing Stage


frontiers in Plant Science

ORIGINAL RESEARCH

Empirical Comparison of Tropical Maize Hybrids Selected Through Genomic and Phenotypic Selections


frontiers in Plant Science

Application of Genomic Selection at the Early Stage of Breeding Pipeline in Tropical Maize

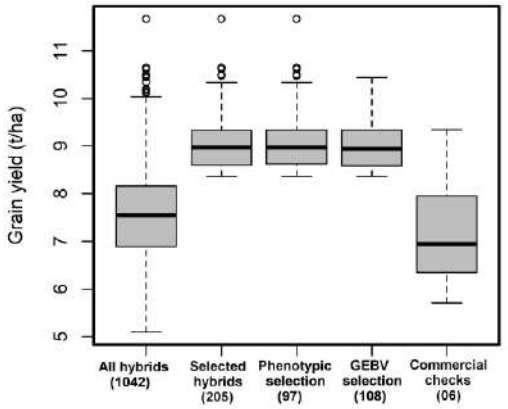

Theoretical and Applied Genetics (2021) 134:279–294

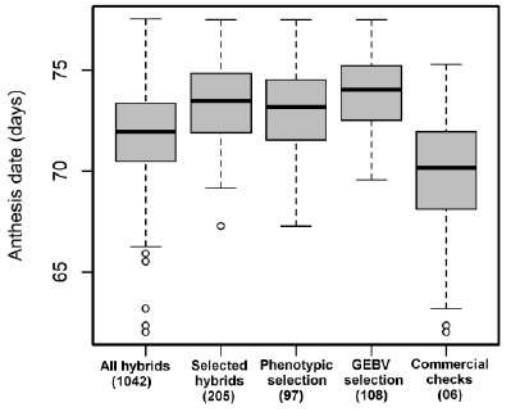
ORIGINAL ARTICLE

Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program

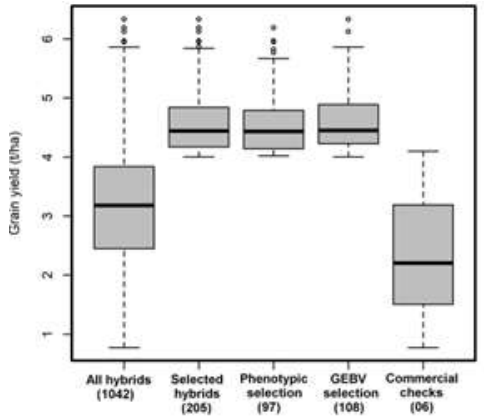
Comparison of PS vs GS at Stage II hybrids in three regions

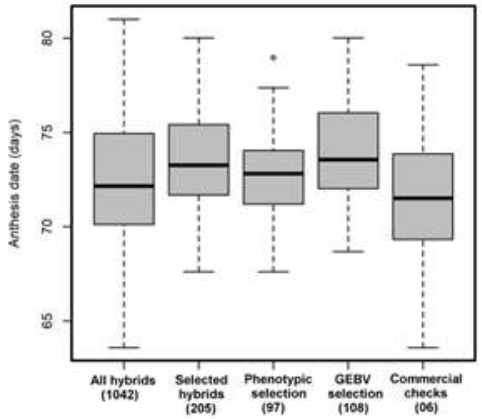
Optimum - ESA



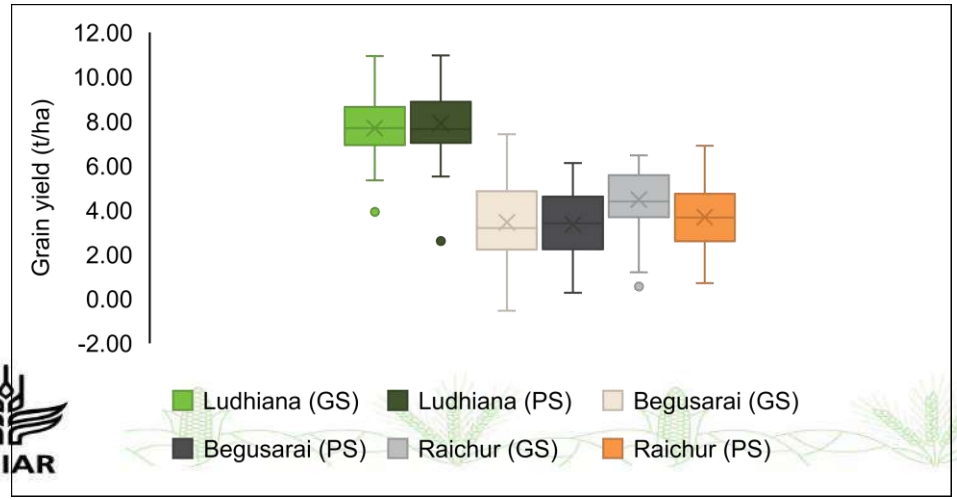


Managed Drought - ESA

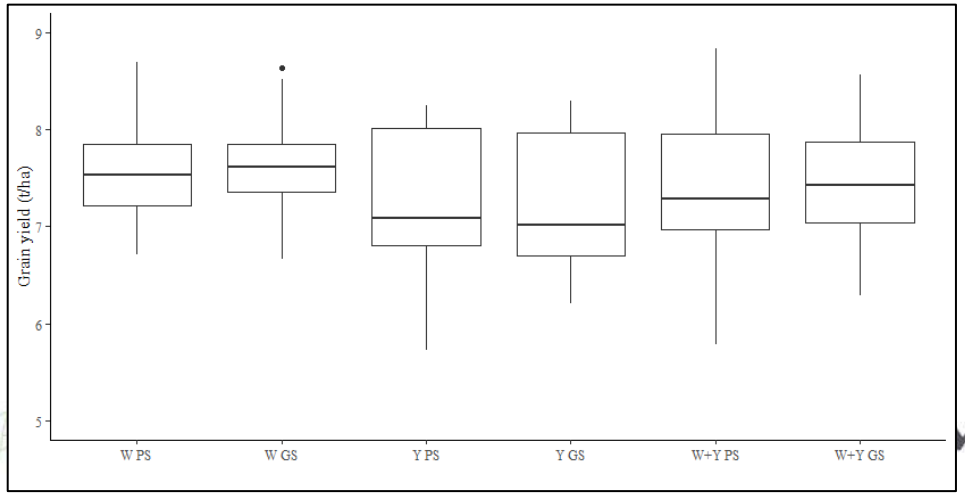




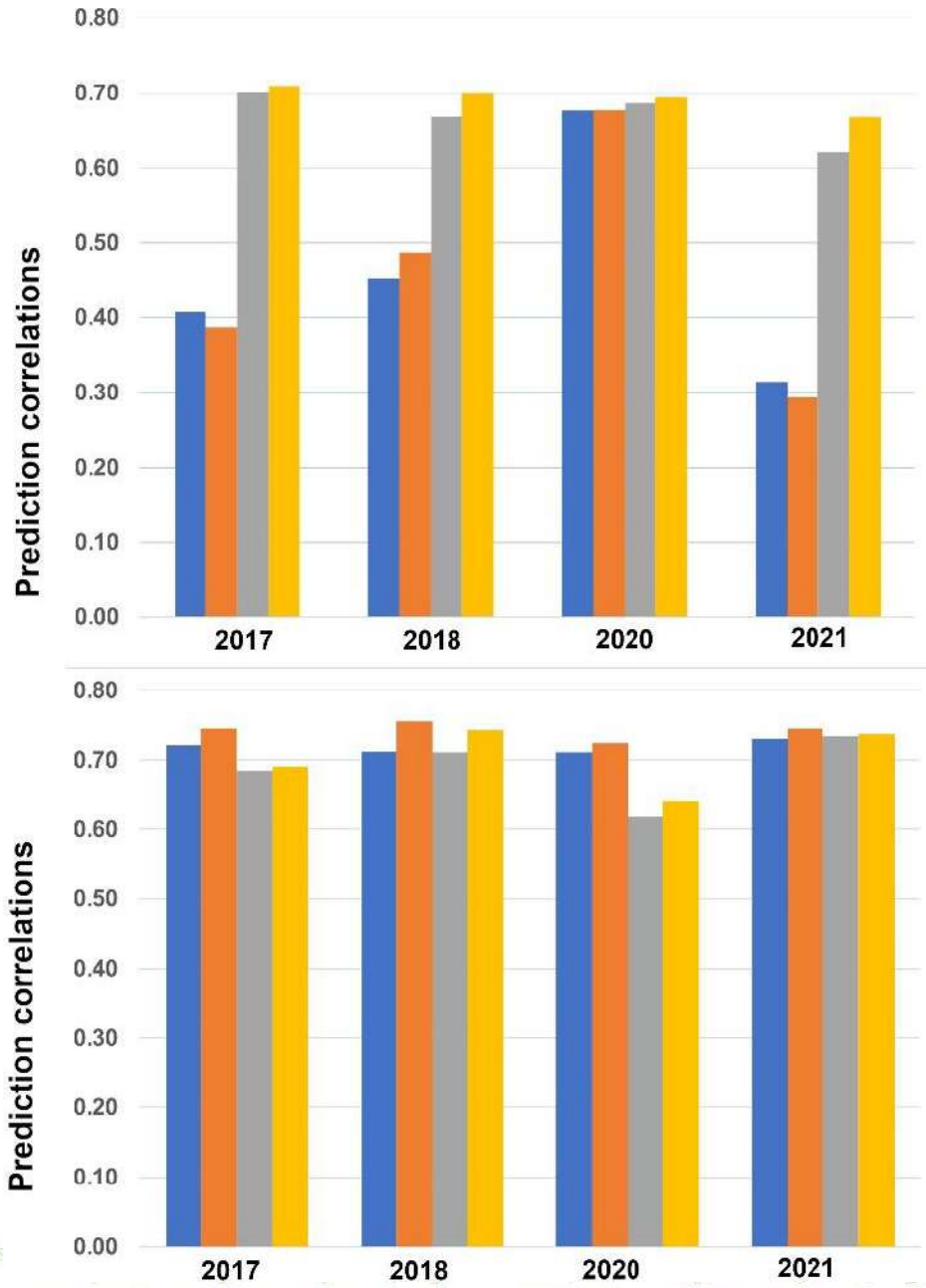
Heat stress – South Asia



Optimum and Drought - Latin America



GS – With different genotyping platforms



Year/Management	Managed drought	Optimum	Low Nitrogen
2017	423	423	-
2018	620	620	-
2019	246	244	-
2020	302	302	302
2021	499	499	499

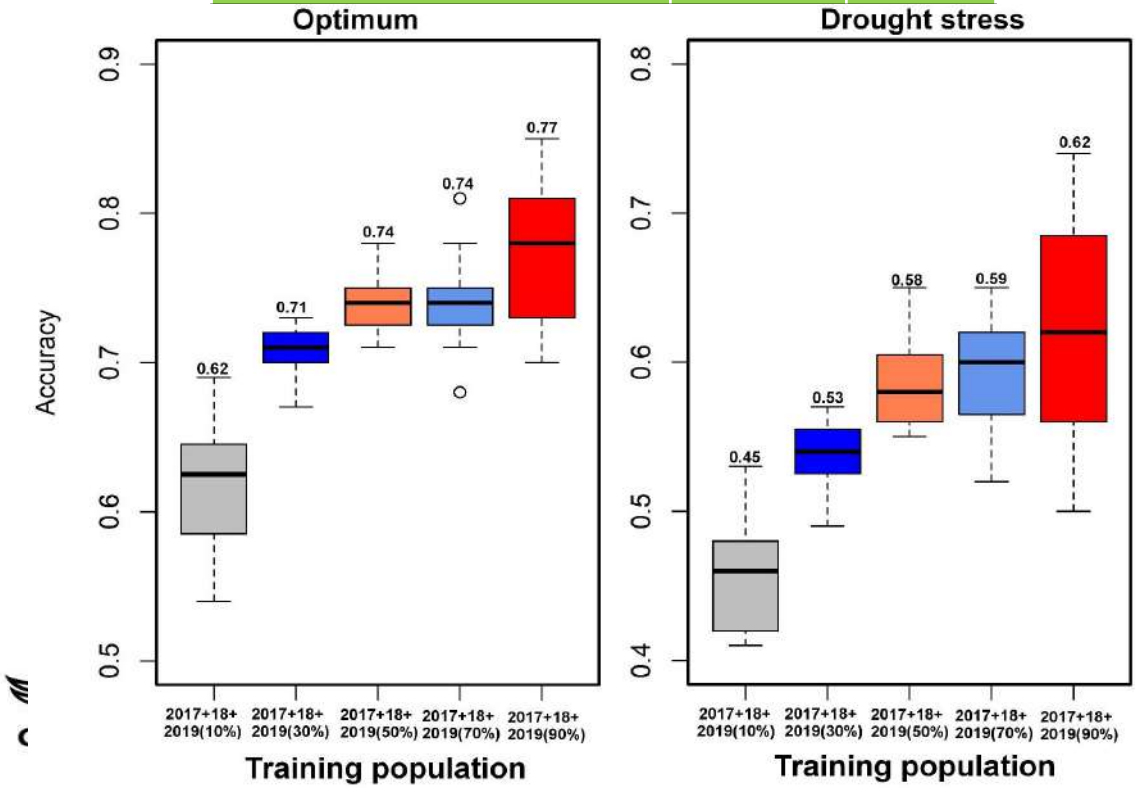
Table. Number of individuals with phenotypes and genotypes with both rAmpSeq and Dart markers by year

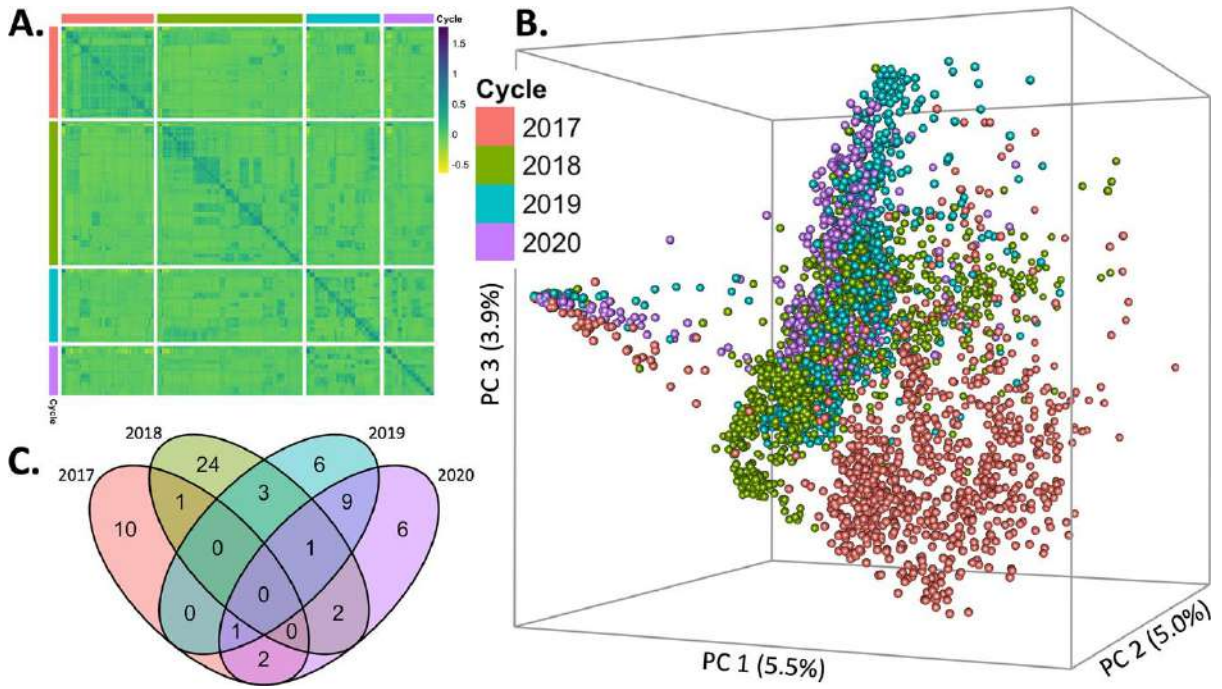
Stage 1 lines Genotyped (Phenotyped) in Five Product Profiles

GS Strategy	Year	EA-PP1	EA-PP2	EA-PP3	SAPP1	SAPP2
T50:P50 (TP - FR)	2017	1500(850)	1400(600)	-	-	-
T50:P50 (TP - FR)	2018	2600(1400)	-	-	-	-
T50:P50 (TP - FR)	2019	2500(890)	-	-	-	-
T50:P50 (TP-SpD)	2020	1550(720)	1300(510)	-	-	-
T50:P50 (TP - SpD)	2021	2050(860)	1200(460)	500	-	-
T50:P50 (TP -SpD)	2022	1880(635)	1460(520)	590	1450(420)	-
Test all (Sparse D)	2023	2500 (2500)	1200(425)	500	1500 (660)	570
	Total No of lines	**14180 (7355)**	**6500 (2515)**	**1590**	**2950 (1080)**	**570**

Prediction using historical data and GRM under drought stress and optimal conditions (EAPP1)

Training set	EST set	GY Opt
2017	2018	**0.34**
2018	2019	**0.28**
2019	2017	**0.21**
2017+2018	2019	**0.37**
2017+2018+10% of 2019	90%2019	**0.62**
2017+2018+30% of 2019	70%2019	**0.71**
2017+2018+50% of 2019	50%2019	**0.74**
2017+2018+70% of 2019	30%2019	**0.74**
2017+2018+90% of 2019	10%2019	**0.77**



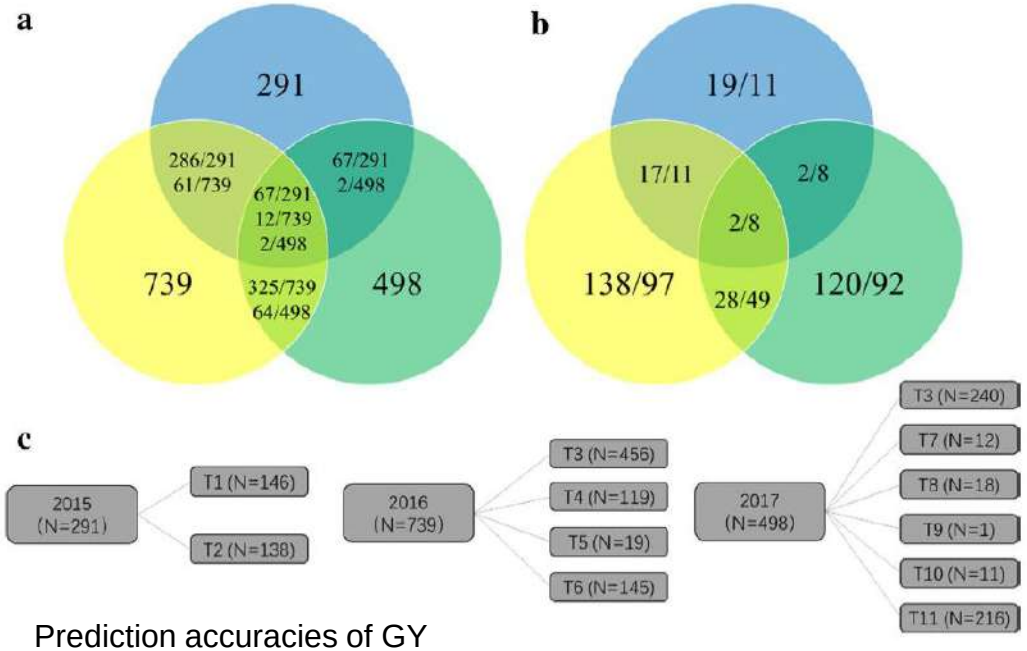


A Heatmap of the genomic relationship matrix. **B** First three PCs of the additive genomic relationship matrix, G. Dots represent individuals that are separated by colors for each year (2017, 2018, 2019 or 2020). **C** Venn diagram representing the number of common parents used to generate the DH lines at each year.

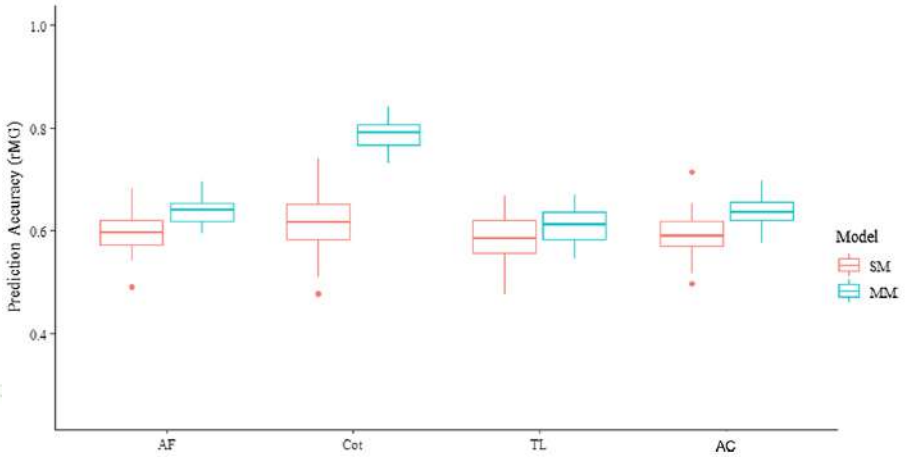
Prediction using historical data and multi-environmental model under optimal conditions in LA breeding pipelines

Training Population	Testing Population	Prediction accuracy
2015+2016	2017	0.38
2015+2017	2016	0.36
2016+2017	2015	0.50
2015+2016+ 50% 2017	50% 2017	0.50 (0.03)
2015+2017+ 50% 2016	50% 2016	0.56 (0.02)

Prediction accuracy improvement by increasing TRN size and strengthening the relationship between TRN and TST



Prediction accuracies of GY
estimated from cross-validation schemes, within and across location analyses, single-environment model (SM) and multiple-environment model (MM)



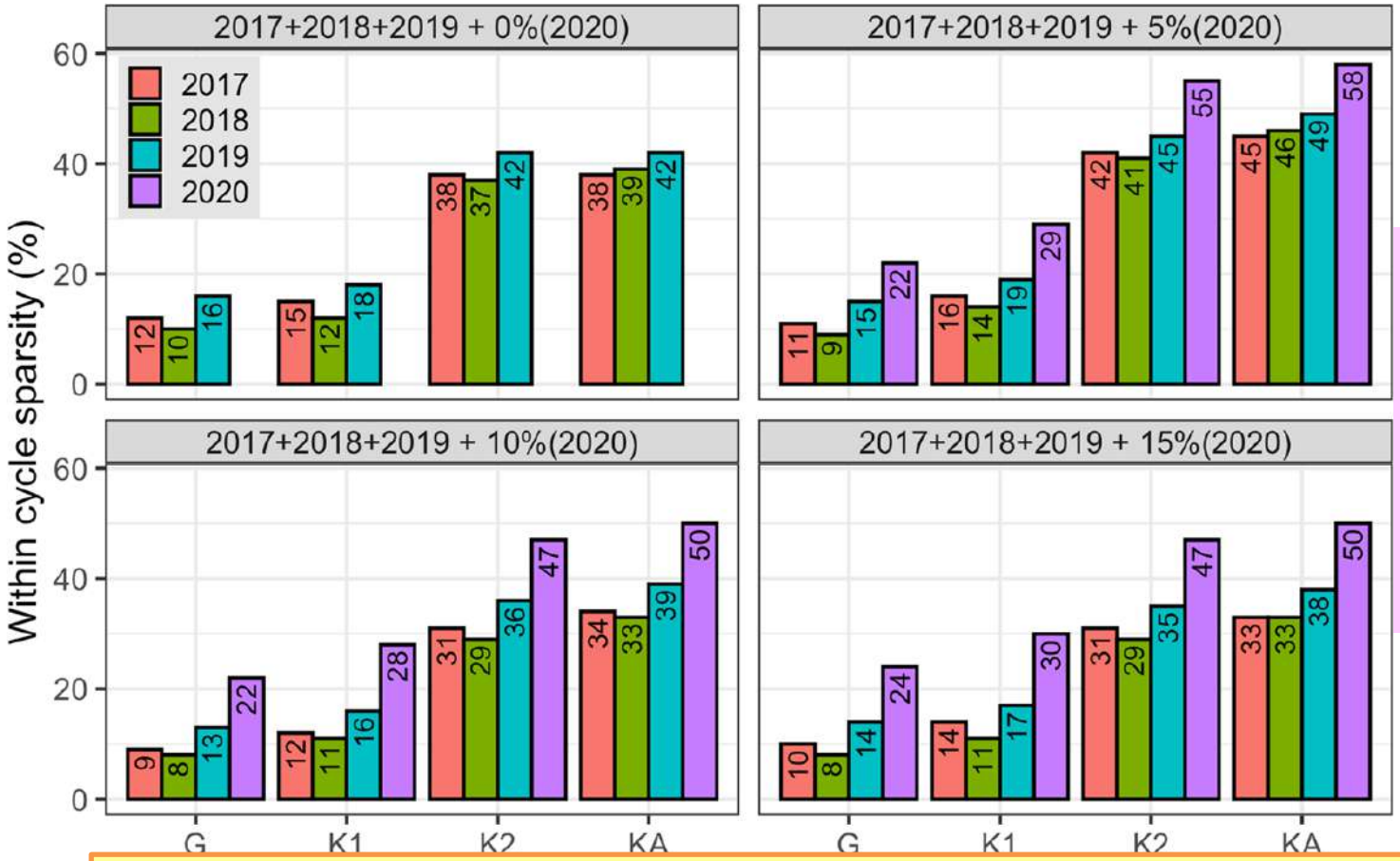
Theoretical and Applied Genetics (2020) 133:2869–2879
https://doi.org/10.1007/s00122-020-03638-5

ORIGINAL ARTICLE

Genomic prediction across years in a maize doubled haploid breeding program to accelerate early-stage testcross testing

Nan Wang[1,2] · Hui Wang[2,3,4] · Ao Zhang[5] · Yubo Liu[5] · Diansi Yu[2,3,4] · Zhuanfang Hao[1] · Dan Ilut[6] · Jeffrey C. Glaubitz[7] · Yanxin Gao[7] · Elizabeth Jones[7] · Michael Olsen[8] · Xinhai Li[1] · Felix San Vicente[2] · Boddupalli M. Prasanna[8] · Jose Crossa[2] · Paulino Pérez-Rodríguez[9] · Xuecai Zhang[2]

GS with Sparse Selection Index



Historical data – All lines are not necessary to be part of TP!

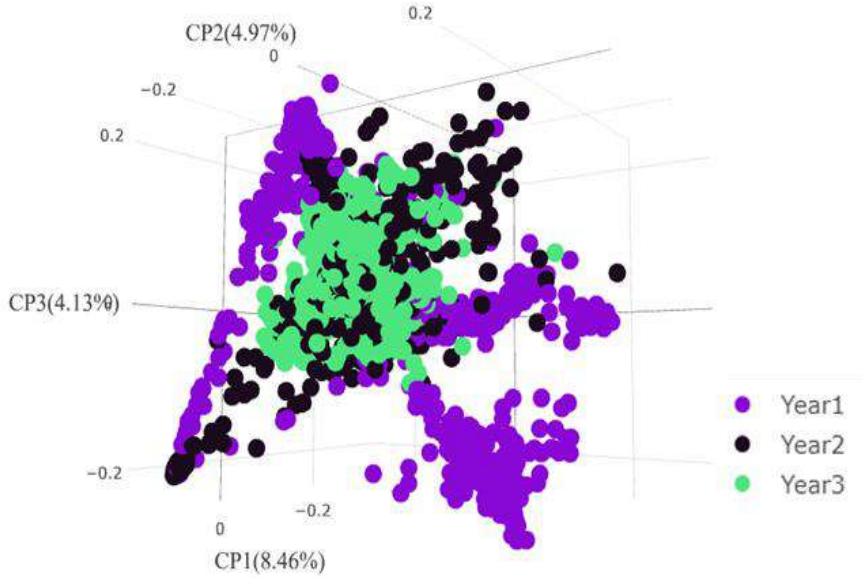
Only subset of lines are contributing for prediction accuracy

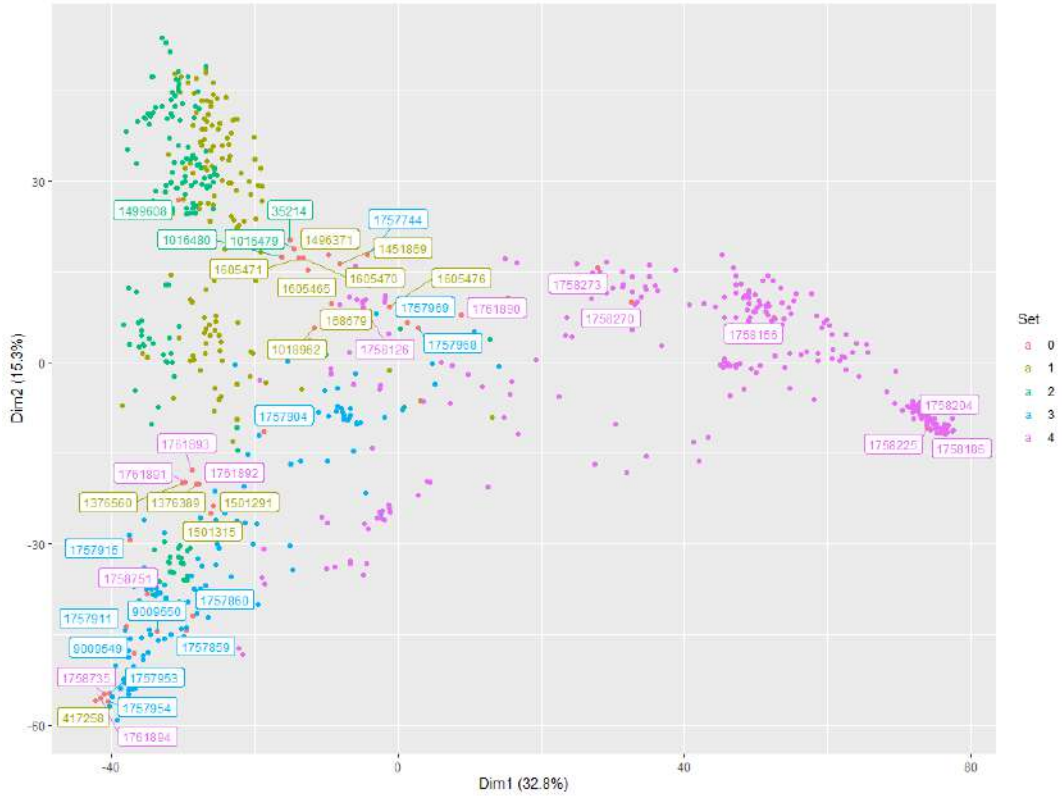
Proportion of the training individuals from each year that contributed to the prediction of genotypes from 2020, using SSIs with different relationship matrices

CGIAR

CIMMYT

Exploring genomic relationship with historical datasets for predictions under heat stress (SAHDT)





- In 2023, a sub-set of past data considering average GRM with respect to each individual in the testing set was used in predictions of new cohort, rather than using all available historical datasets, and the work is ongoing to test different algorithms for improving prediction accuracy using GRM.
- Historical training set optimization using GRM is being carefully considered

CGIAR

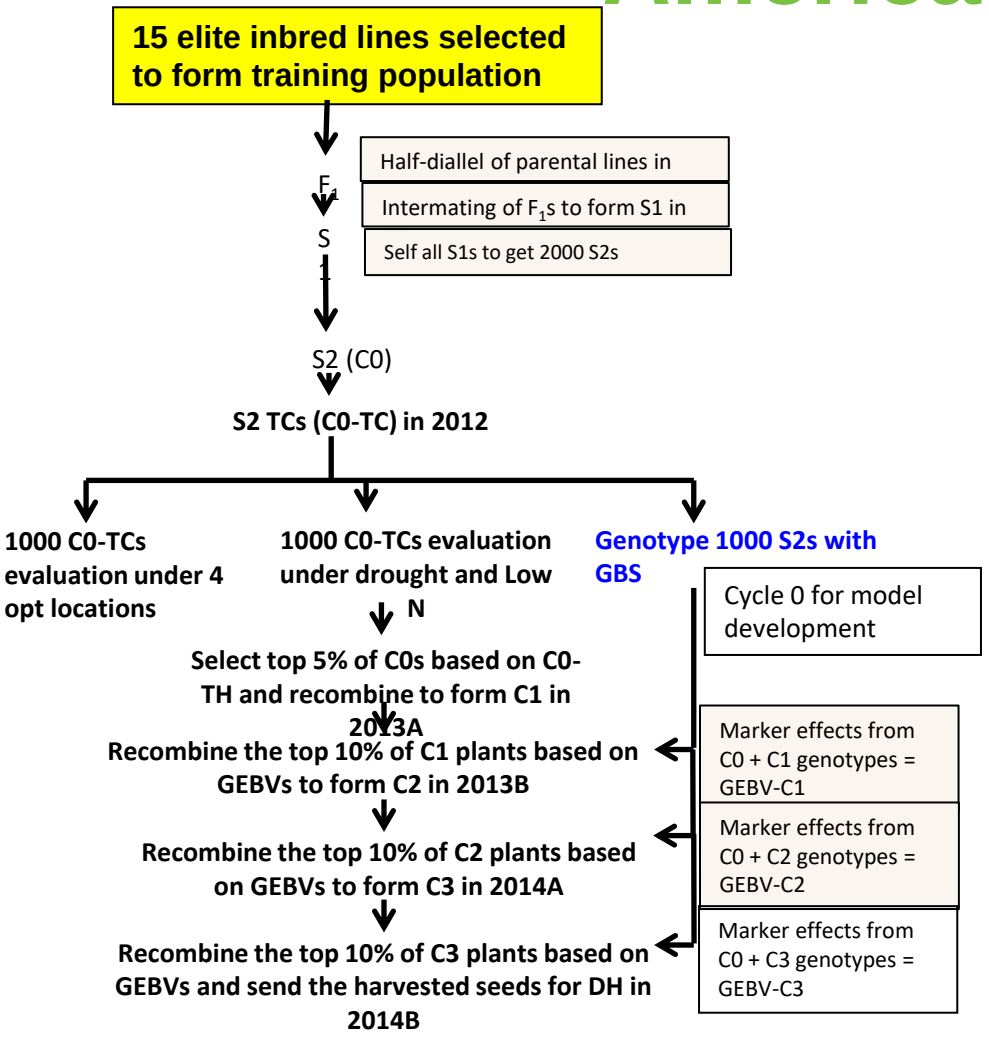
CIMMYT

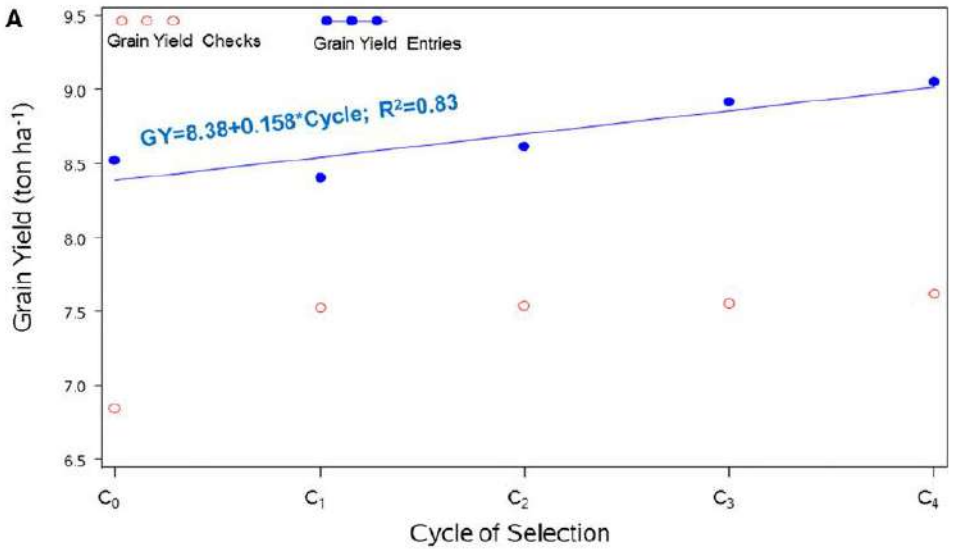
Lines developed through biparental based RCGS are used as parent in allocated hybrids to partners

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

Multi-parent based Rapid cycle GS : Latin America

Rapid-cycle Genomic Selection
1 cycle PS and 3 cycle GS accomplished in 4 years





Zhang et al. 2017

Overall gain in GY : 100.5 kg ha^{-1} $year^{-1}$

Rapid Cycling Genomic Selection in a Multiparental Tropical Maize Population

Xuecai Zhang, Paulino Pérez-Rodríguez, Juan Burgueño, Michael Olsen, Edward Buckler, Gary Atlin, Boddupalli M Prasanna, Mateo Vargas, Félix San Vicente ✉, José Crossa ✉

G3 Genes|Genomes|Genetics, Volume 7, Issue 7, 1 July 2017, Pages 2315–2326, https://doi.org/10.1534/g3.117.043141

CGIAR

CIMMYT

Multi-parent based Rapid cycle GS : Asia (Drought stress)

Selection of Inbred lines (8-12)

P1 x P2; P1 x P3,…., P9 x P10

F1 x F1

Selfing

$S_{ijA/B} \times Tester_{B/A}$

(Genotyping of $S_{2:3}$ lines)

Train prediction model

Phenotyping of TCs for Heat stress

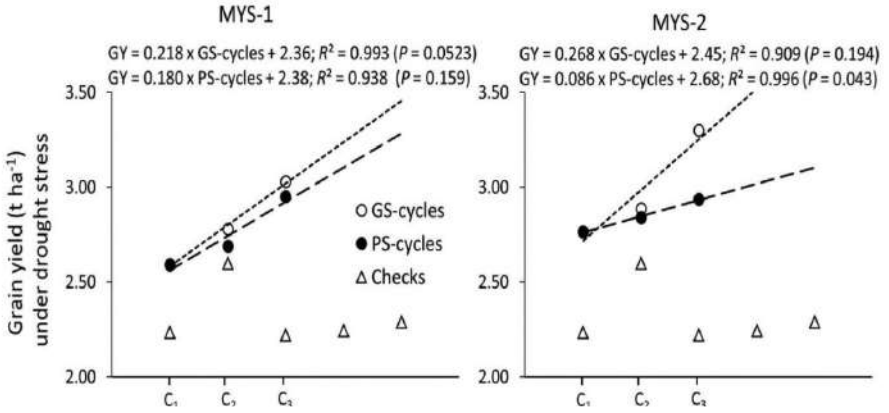
Intermate top 5-10% $S_{2:3}$ lines based on phenotype

Genotype progeny of Intermated lines

5-10% intermated based on GEBVs

C2 -C3

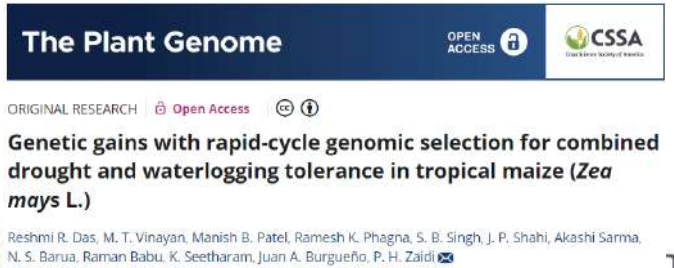
Doubled Haploid lines



Drought Stress:

MYS1: 109 kg ha^{-1} year^{-1}

MYS 2: 138 kg ha^{-1} year^{-1}


The Plant Genome

OPEN ACCESS

CSSA

ORIGINAL RESEARCH Open Access

Genetic gains with rapid-cycle genomic selection for combined drought and waterlogging tolerance in tropical maize (*Zea mays* L.)

Reshmi R. Das, M. T. Vinayan, Manish B. Patel, Ramesh K. Phagna, S. B. Singh, J. P. Shahi, Akashi Sarma, N. S. Barua, Raman Babu, K. Seetharam, Juan A. Burgueño, P. H. Zaidi

CIMMYT

Modified Rapid Cycle GS with historical data

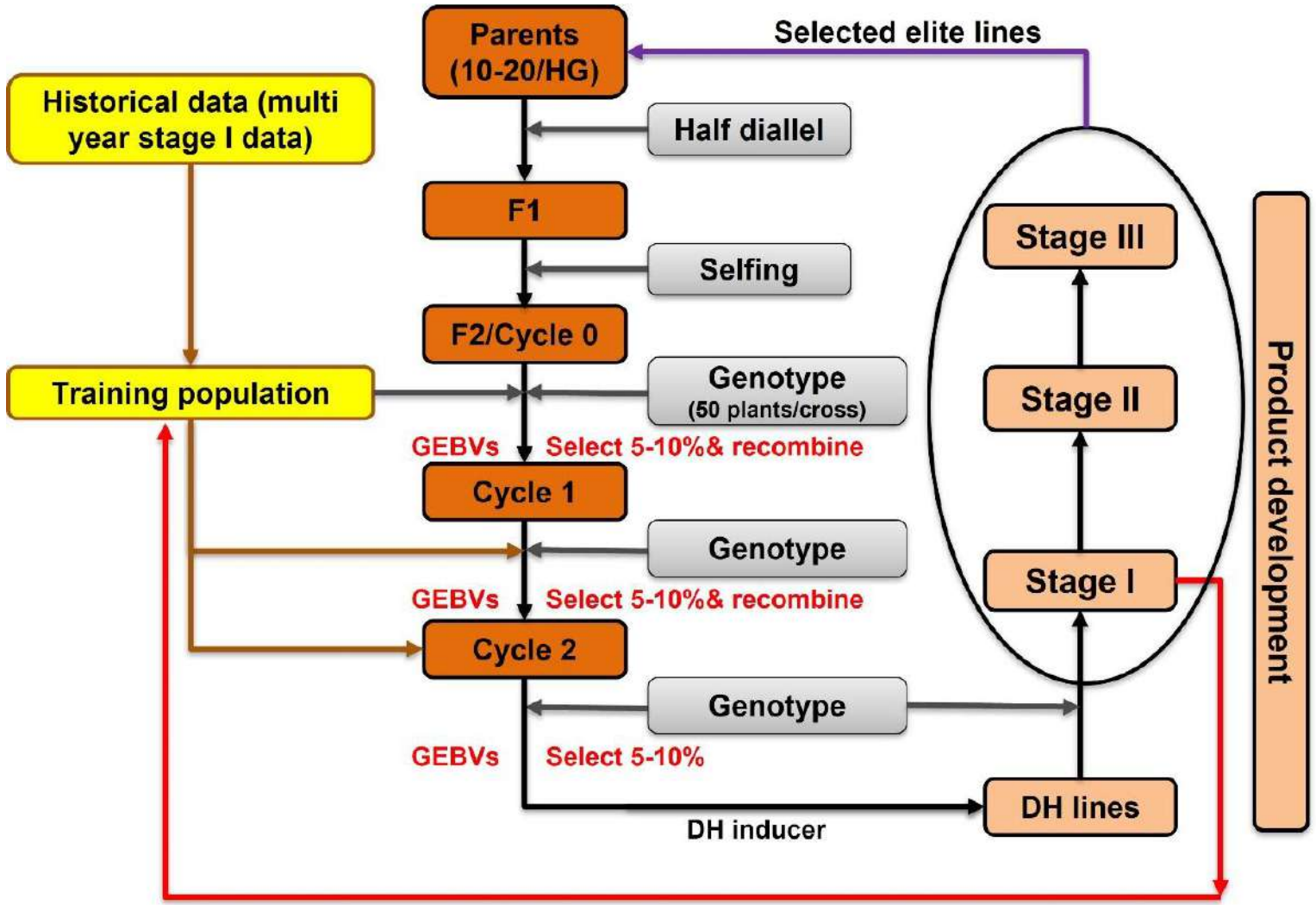
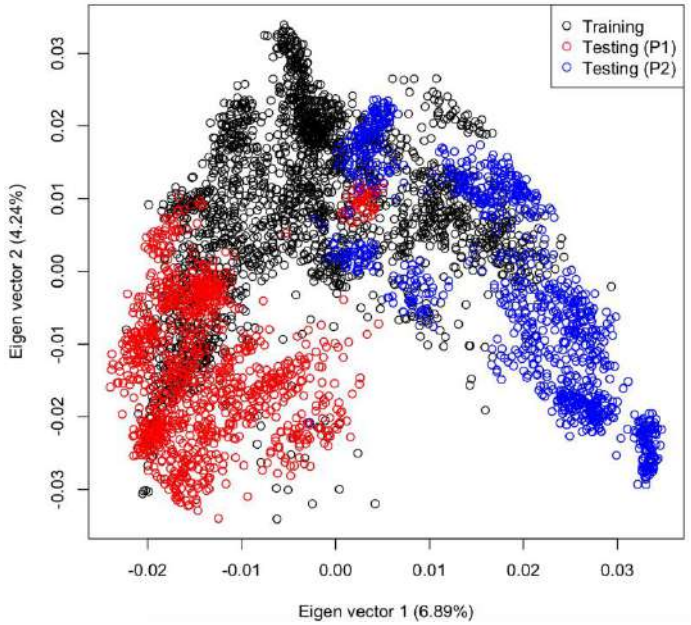


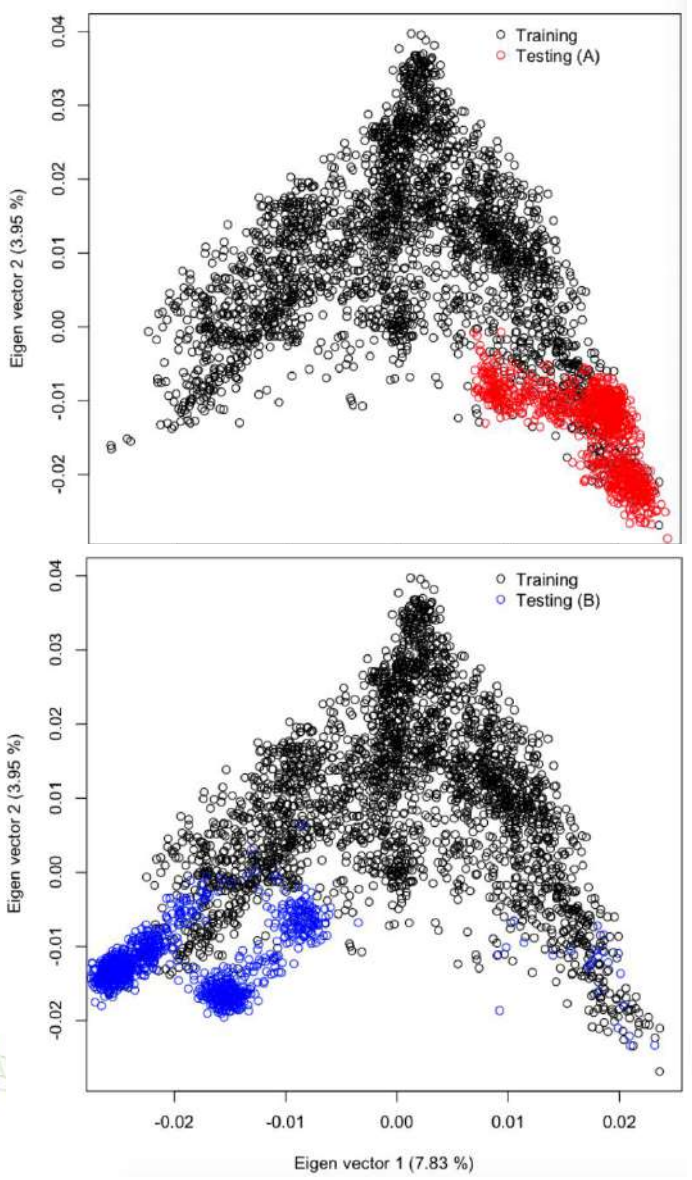
Figure. Flow chart showing the steps involved in RCGS cycle scheme

Modified RCGS with historical data

Cycle 0



Cycle 1



Best 10% of the lines were selected at cycle 1 and intermated, and the harvested seeds are submitted for DH induction

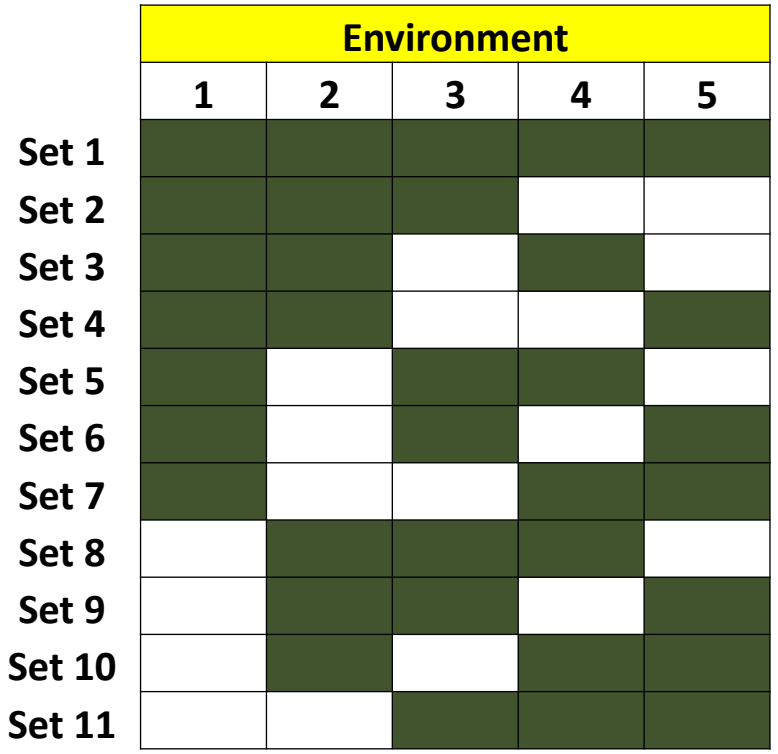
CGIAR

CIMMYT

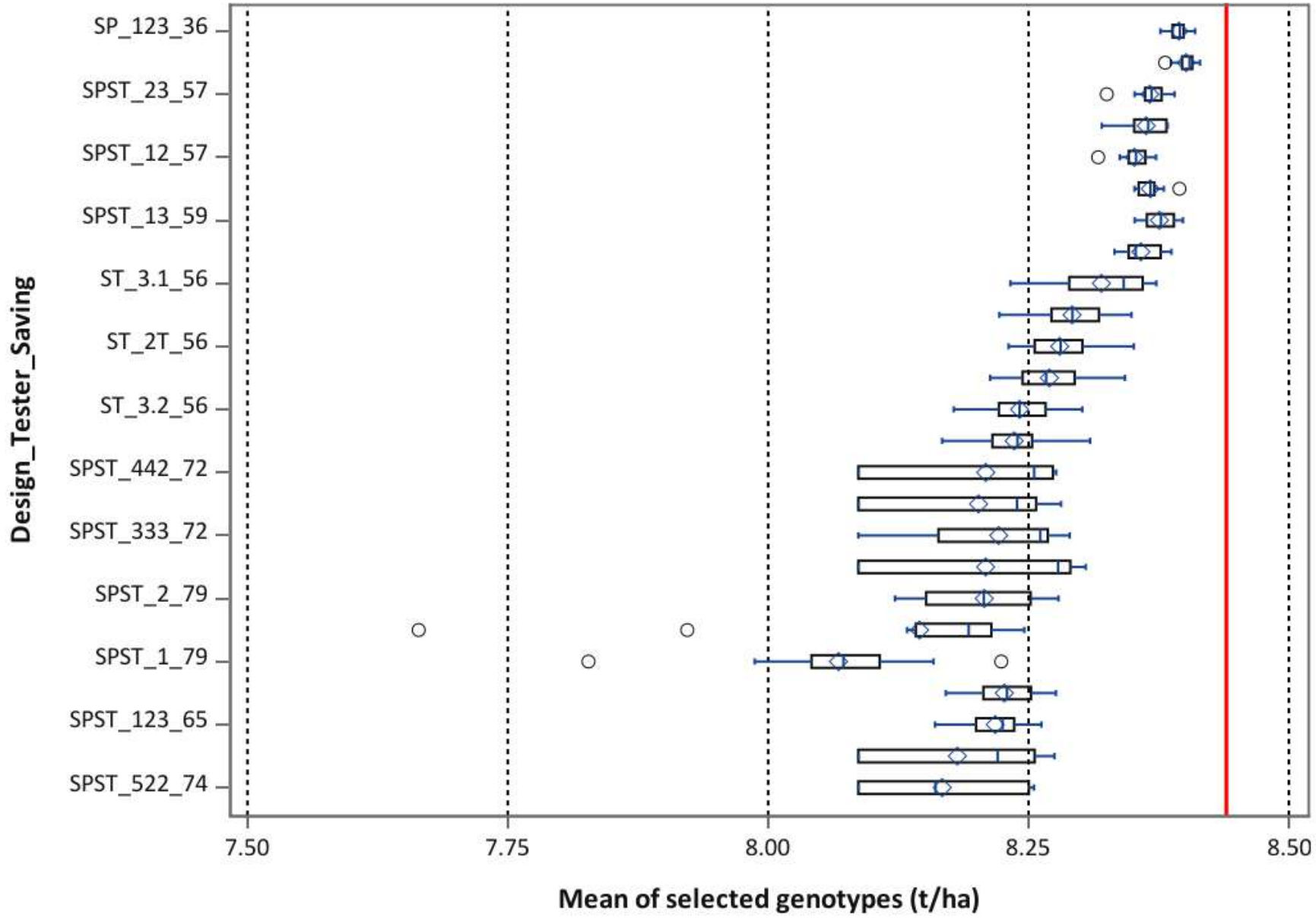
Objective of GS – Reduce both cost and cycle length

Future Strategy – Integration of sparse design with increased number of testers from two to six.

- Sparse design is planned in such a way that all lines get the opportunity to evaluated in the field with at least one tester and a set of lines are evaluated with all testers from each heterotic group.
- This approach mimics the combination of stage 1 and 2 evaluations, so helps to skip stage I evaluation from breeding pipeline ultimately reduce the breeding cycle time.
- Reducing the phenotyping plots up to 40% and reduces the cost significantly.
- Genotyping all lines, and phenotyping in chosen locations where all genotypes will be evaluated in more than two to three locations in the sparse design.



Sparse phenotyping & testcrossing (SPST) with historical data



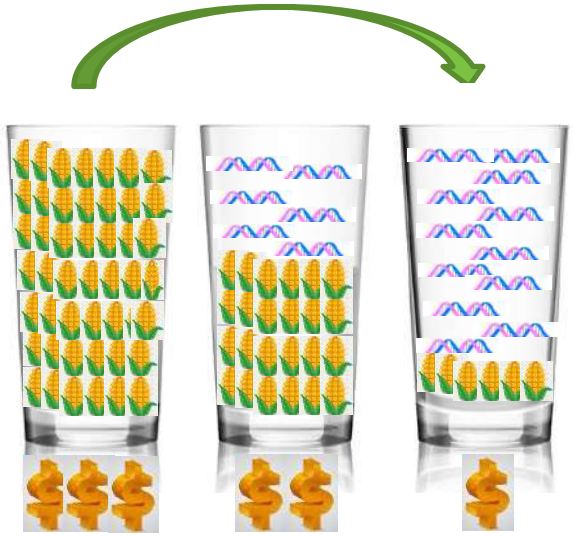
CGI

CIMMYT

Continuous Improvement in GS Strategy

- **Strategy I – completely replicated TP** – use **test-half and predict-half** strategy in PPs where GS is newly started – We use this approach in EAPP-2 in 2017-2019

- **Strategy II – Sparse design-based TP** – with **test-half and predict-half** strategy – Here test crosses are evaluated in sparse design – helps to increase one additional location with fixed number of plots as of test-half and predict-half strategy. This helps to capture better GxE interaction effect in the model. From 2020, we are using this approach for EAPP-I, and II.



- **Strategy III** – **Test all in Sparse design** – genotype all and test all in the product profile – Here we use sparse testing design where all the stage I lines are phenotyped, but sparse design is planned in such a way that total number of plots is reduced but all the genotypes are evaluated in field. GEBVs are estimated for all the lines and used for selection

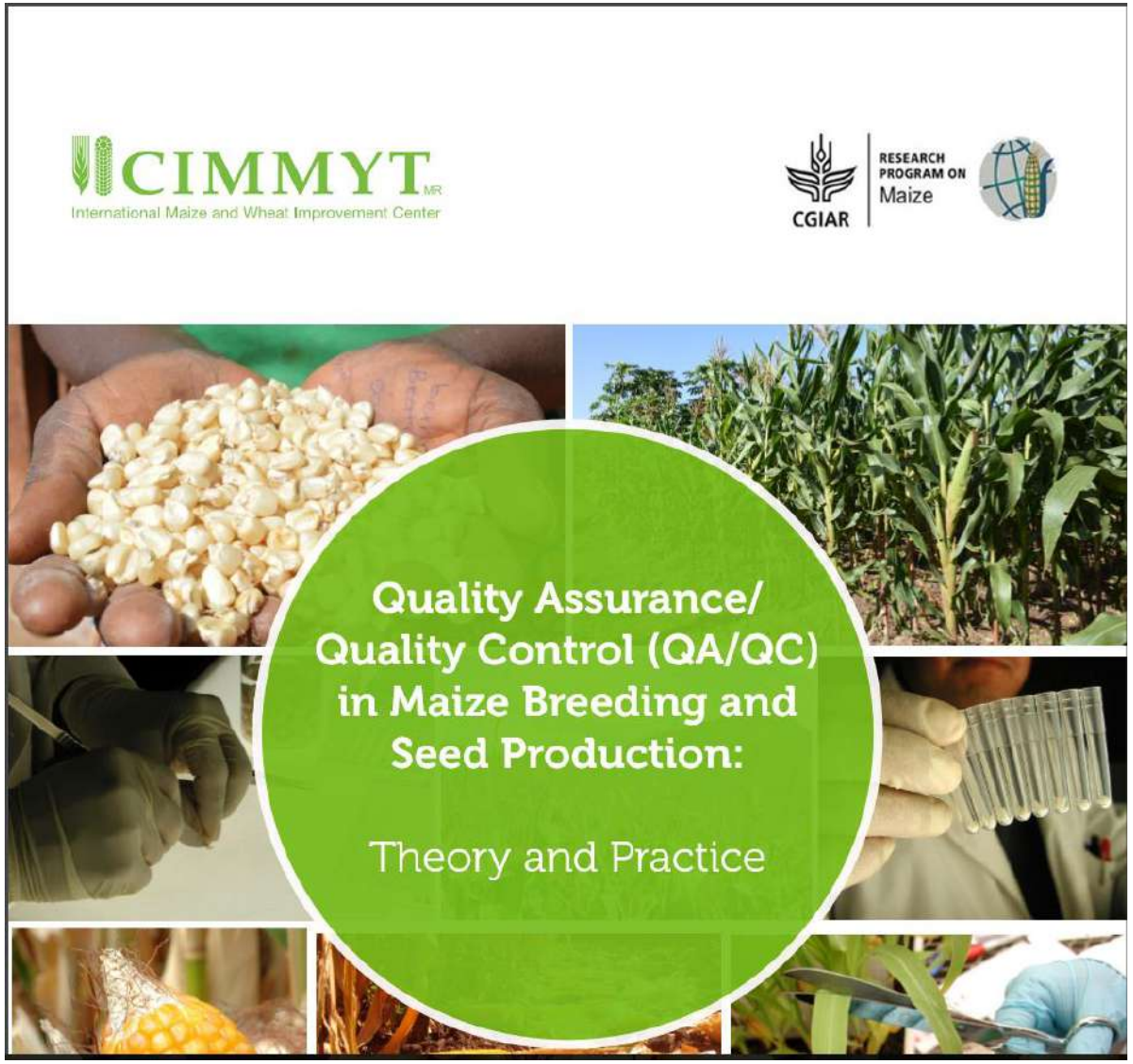
Modified Strategy – Use Sparse phenotyping with Sparse testcrossing Strategy III in combination with historical data which improves selection efficiency and reduce cycle length.

CGIAR

CIMMYT MR

QA & QC in maize breeding and seed production

- QC is a mainstay in many stages in all maize breeding pipelines starting from parental and F1 QC to line finishing
- Set of >350 lines being used in regional breeding programs and all released hybrids are genotyped with QC markers, and final Reference profile is available for both internal and partners' use
- All CMLs are genotyped with QC set and their reference profile is also available for internal and partners' use



A set of 100 informative markers were identified for routine application of QA&QC

CGIAR

CIMMYT

In Summary

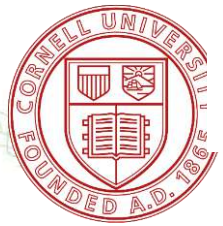
- Cost-effective and time and error sensitive genotyping platforms are key to use novel tools and strategies in breeding pipelines

- Application of molecular markers are decided by trait architecture, improvement of pipeline efficiency and cost optimization

- Maize has a trait marker pipeline that follows stage gate advancement process before trait markers are adopted for deployment in breeding pipelines

- Genomic selection is a routine strategy followed in all breeding pipelines and a continuous improvement of the existing strategies are explored for cost and time efficiency with support from breeders, quantitative geneticists and biometricians

Acknowledgements



Thank you for your interest!