Shopping in the diversity supermarket; using big data to bring diversity to breeding

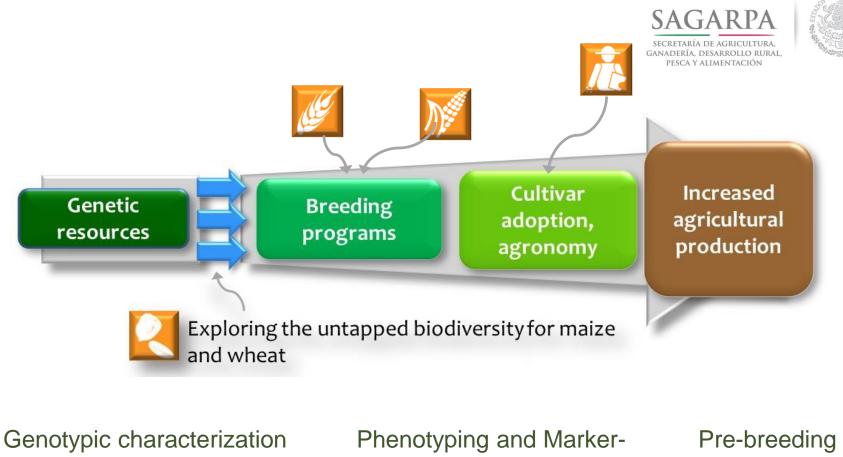
Sarah Hearne – CIMMYT International Maize and Wheat Improvement Centre Big Data in Agriculture 14th May 2018

s.hearne@cgiar.org; cimmyt-mab-seed@cgiar.org

Seeds of Discovery (SeeD) 2010 (MasAgro Biodiversidad)



CIMMYT



of germplasm banks and public elite germplasm

trait associations

Diversity, diversity everywhere but not a drop to drink



Inbred donors and elite lines- basis of breeding in public and private sector

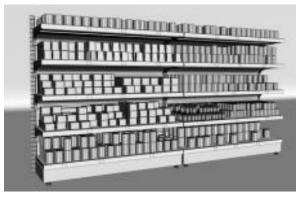
Barriers to use



Adaptation, linkage drag

Barriers to use













Where does Big Data fit?

Genotyping of germplasm



GBS

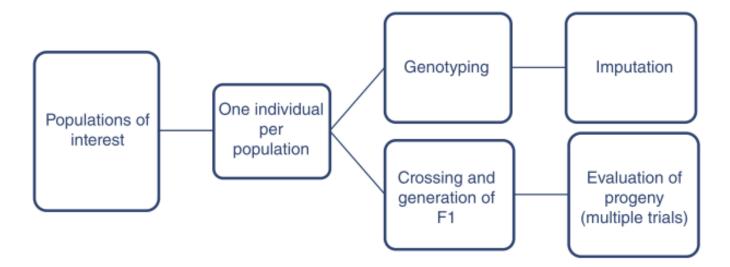
1M ref genome aligned SNP Imputation QC filtered to ~350k

DArT seq GWAS:- 110k SNP ~50%mapped to ref genome No imputation

Bank: 1M markers, 650k SNP filtered to 350-470k SNP (5x) Allele frequencies ~45% mapped to ref genome

Marker trait associations - GWAS

F1 association mapping (FOAM)



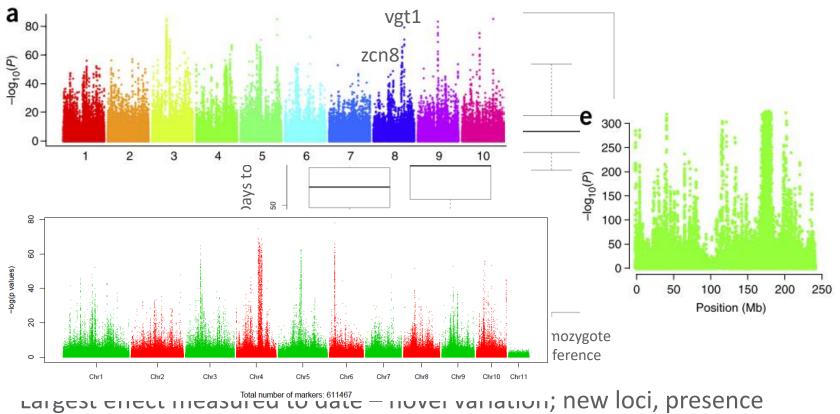
3500 accessions, flowering, plant & ear height.....

34 trials

Romero et al 2017, Nature Genetics



GWAS - flowering



in elite tropical material (CMLs) varies– ~1000 genes

Romero et al 2017, Chen, Palacios, Willcox, Burgueno, Hearne unpublished

Extending and expanding germplasm bank data

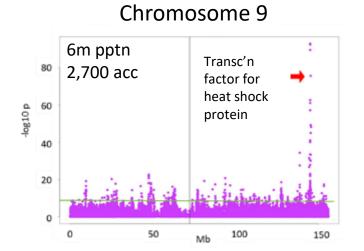
- ~17500 landrace accessions with collection site "data"
- 2y curation effort

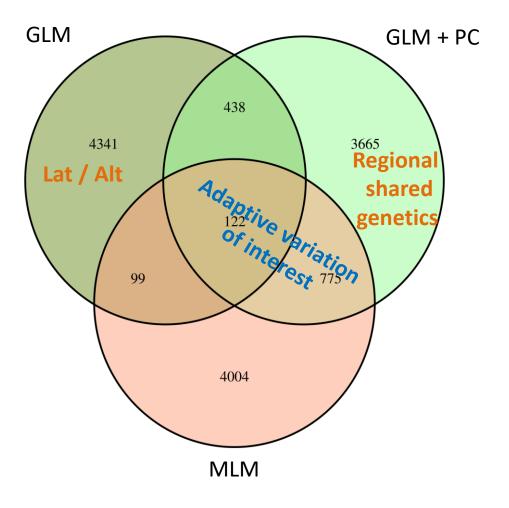
Kai Sonder

• Derive GIS climate variables from collection sites

a 5 · C · -					Maize Accession GIS extracted d	ata curated 2014v2 - Excel							
ile Home Insert Page Layout Fr	ormulas	Data Revi	ew View	ACRO	BAT 🛛 🖓 Tell me what you want to do						H	EARNE, Sarah Jane	(CIMMYT) & Sh
	- A * A	= = =	<i>≫</i> - ∰w	rap Text	Number -	Normal_Shee	Normal	Bad			×	AutoSum -	AT P
ste	A				enter - \$ - % , * % % Conditio	al Format as Good	Neutral	Calcu	lation		ete Format	🖶 Fill -	Sort & Find &
💞 Format Painter " 🛛 🖉 👘 👘	A. A.	2.8.2	THE THE MAN	erge oc u		g* Table*	Reduat					& Clear *	Filter * Select *
Clipboard G Font	15		Alignment		S Number S	Stj	yles			C	sts	Edit	ting
920 * 1 × ✓ <i>f</i> _x CIMM	IYTMA-0255	17											
A 8 C	D	E	F	G	4 1	JKLMN	OPOR	S S	TU	v		w	X
GID +1 accid = collectno	vearc v L	at New 👻 I	ong New - A	It (I - 1	* Location	* di * di * di * di * di * di			dd - dd - d	d6MonthGrow	An - ddFlon	werMonti - ddF	lowerMonti - de
262176 CIMMYTMA-025520 ARZM 04 023	1980	-29.35	-58.6	60	3 PERUGORRIA&CORRIENTES	421 422 388 448 480	569 447 413 46	50 471	412 410		459	448	480
262175 CIMMYTMA-025519 ARZM 04 021	1980	-30.05	-57.82	84	8 PARAJE ESTINGANA&CORRIENTES	431 446 411 496 492	581 497 452 48	482	425 422		487	496	492
262174 CIMMYTMA-025518 ARZM 04 020	1980	-30.05	-57.82	84	3 COLONIA LIBERTAD& CORRIENTES	431 446 411 496 492	581 497 452 48	87 482	425 422		487	496	492
262173 CIMMYTMA-025517 ARZM 04 016	1980	-30.25	-57.63	50	3 PARAJE TACUABA&CORRIENTES	431 446 411 496 492	581 497 452 48	87 482	425 422		487	496	492
262172 CIMMYTMA-025516 ARZM 04 003	1980	-28.55	-56.05	78	4 SANTO TOME&CORRIENTES	491 510 473 499 516	577 509 502 53	37 525	502 475		513	516	577
262171 CIMMYTMA-025515 ARZM 03 053	1980	-32.310879	-59.692253	57	3 COLONIA 21 CRUCECITA&ENTRE RIOS	448 412 401 423 471	595 489 464 48	482	422 423		465	423	471
262170 CIMMYTMA-025513 ARZM 03 051	1980	-31.395742	-59.210403	60	3 COLONIA VIRARO&ENTRE RIOS	432 397 373 435 462	594 449 451 46	50 457	396 397		452	435	462
262169 CIMMYTMA-025512 ARZM 03 050	1980	-31.187587	-59.411532	65	3 COLONIA AVIGDOR&ENTRE RIOS	432 397 373 435 462	594 449 451 40	50 457	396 397		452	435	462
262168 CIMMYTMA-025511 ARZM 03 047	1980	-31.187587	-59.411532	65	3 COLONIA AVIGDOR&ENTRE RIOS	432 397 373 435 462	594 449 451 46	50 457	396 397		452	435	462
262167 CIMMYTMA-025510 ARZM 03 045	1980	-30.38	-58.75	67	3 BASUALDO&ENTRE RIOS	418 422 373 447 466	568 447 427 46	51 457	398 397		454	447	466
262166 CIMMYTMA-025509 ARZM 03 040	1980	-30.355498	-58.740503	69	3 FELICIANO&ENTRE RIOS	418 422 373 447 466	568 447 427 48	61 457	398 397		454	447	466
262165 CIMMYTMA-025508 ARZM 03 037	1980	-30.5833333	-58.5	77	3 COLONIA 9&ENTRE RIOS	431 421 384 459 464	568 473 439 43	4 457	411 409		461	459	464
262164 CIMMYTMA-025507 ARZM 03 034	1980	-30.5833333	-58.5	77	3 COLONIA TUNA&ENTRE RIOS	431 421 384 459 464	568 473 439 47	4 457	411 409		461	459	464
262163 CIMMYTMA-025506 ARZM 03 032	1980	-30.5833333	-58.5	77	3 COLONIA TUNA&ENTRE RIOS	431 421 384 459 464	568 473 439 41	4 457	411 409		461	459	464
262162 CIMMYTMA-025505 ARZM 03 030	1980	-30.5833333	-58.5	77	3 COLONIA TUNA&ENTRE RIOS	431 421 384 459 464	568 473 439 47	4 457	411 409		461	459	464
262161 CIMMYTMA-025504 ARZM 03 029	1980	-30.5833333	-58.5	77	3 COLONIA TUNA&ENTRE RIOS	431 421 384 459 464	568 473 439 47	4 457	411 409		461	459	464
262160 CIMMYTMA-025503 ARZM 03 022	1980	-30.65	-58.02	48	3 COLONIA 16&ENTRE RIOS	431 421 384 459 464	568 473 439 47	4 457	411 409		461	459	464
262159 CIMMYTMA-025501 ARZM 03 020	1980	-30.83	-57.9	59	3 SANTA ELOISA&ENTRE RIOS	442 445 409 482 490	592 497 477 49	98 483	436 421		486	482	490
262158 CIMMYTMA-025500 ARZM 03 019	1980	-30.83	-57.9	59	8 SANTA ELOISA&ENTRE RIOS	442 445 409 482 490	592 497 477 49	483	436 421		486	482	490
262157 CIMMYTMA-025499 ARZM 03 016	1980	-30.8	-57.92	61	3 VILLA DEL ROSARIO&ENTRE RIOS	442 445 409 482 490	592 497 477 49	98 483	436 421		486	482	490
262156 CIMMYTMA-025498 ARZM 03 013	1980	-30.82	-57.98		3 SANCHEZ SAUCE&ENTRE RIOS	442 445 409 482 490			436 421		486	482	490
262155 CIMMYTMA-025497 ARZM 03 012	1980	-30.82	-57.98		3 SANCHEZ SAUCE&ENTRE RIOS	442 445 409 482 490					486	482	490
262154 CIMMYTMA-025496 ARZM 03 011	1980	-30.82	-57.98		3 SANCHEZ SAUCE&ENTRE RIOS	442 445 409 482 490	592 497 477 49	98 483	436 421		486	482	490
262153 CIMMYTMA-025495 ARZM 03 010	1980	-30.82	-57.98		3 SANCHEZ SAUCE&ENTRE RIOS	442 445 409 482 490			436 421		486	482	490
262152 CIMMYTMA-025494 ARZM 03 009	1980	-30.82	-57.98		3 SANCHEZ SAUCE&ENTRE RIOS	442 445 409 482 490			436 421		486	482	490
262151 CIMMYTMA-025493 ARZM 03 008	1980	-30.8	-57.92		3 VILLA DEL ROSARIO&ENTRE RIOS	442 445 409 482 490			436 421		486	482	490
262150 CIMMYTMA-025492 ARZM 03 006	1980	-30.9	-57.93		3 SANTA ANA&ENTRE RIOS	442 445 409 482 490					486	482	490
262149 CIMMYTMA-025491 ARZM 03 005	1980	-30.9	-57.93		3 SANTA ANA&ENTRE RIOS	442 445 409 482 490					486	482	490
262148 CIMMMUTMAN,025400 A07M 03.004	1080	.22.17	.58 77	51	RIGANTA ANITAR ENTRE DIGS	AA6 A11 308 AAA A61	501 186 178 15	93.1.59	423 434		.455		.461



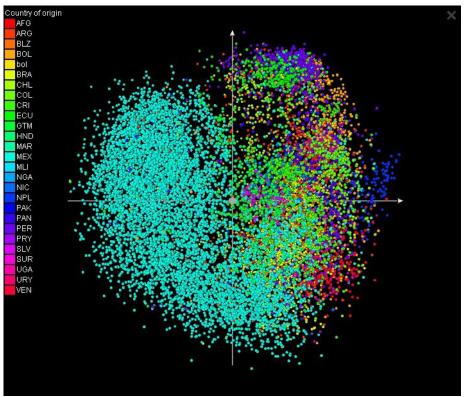




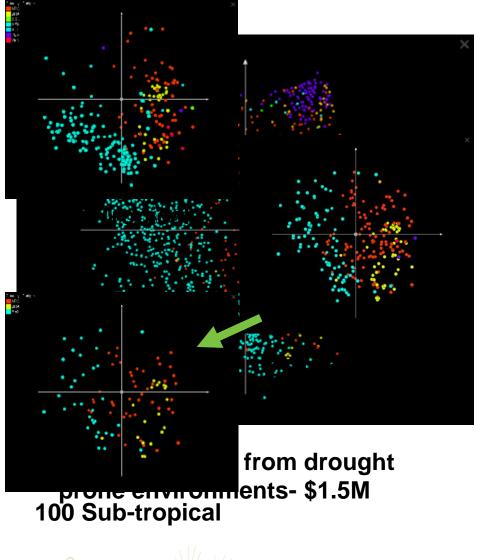
Alberto Romero, Arcadio Valdes- unpublished, HuiHui Li, Jorge Franco, Jose Crossa

Molecular and environmental diversity – shift from cores to breeder panels

MDS euclidian genetic distances

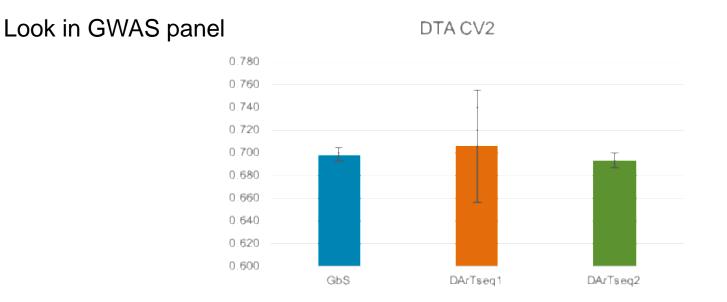


15,384 landraces



Smarter selection?

GS with landraces



MarkerBefore QCAfter QCGBS(1M) 352048335319DArTseq111069840880DArTseq2946445027

Gorjanc *et al* 2016 BMC Genomics Crossa and Hearne unpublished



MORE?

Phenotypes

Genotypes



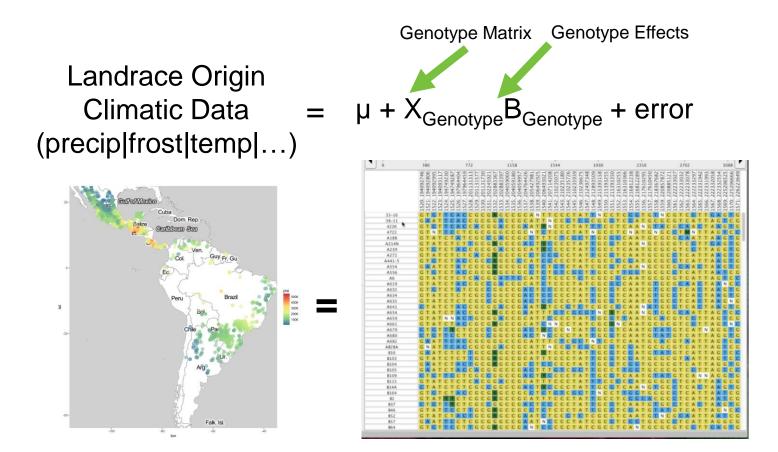
Environments







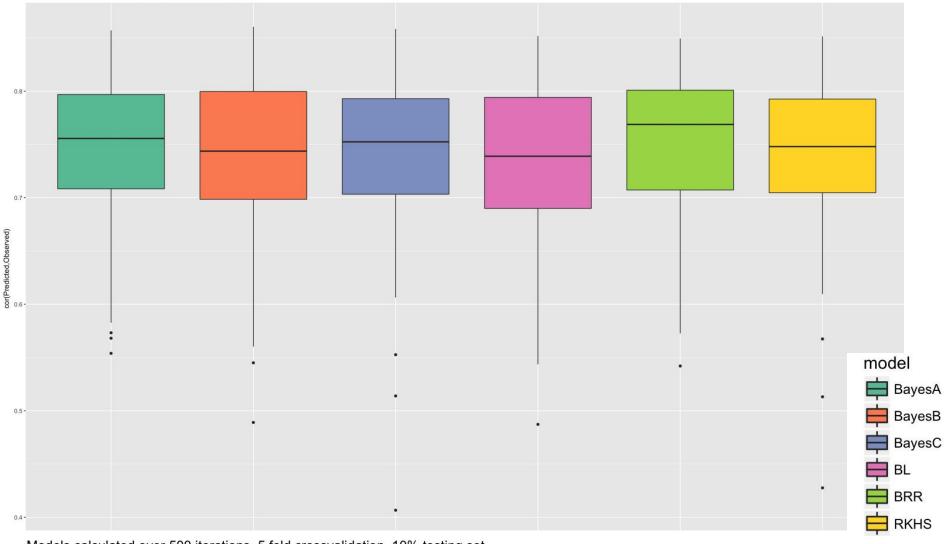
Predict landrace collection site environments?



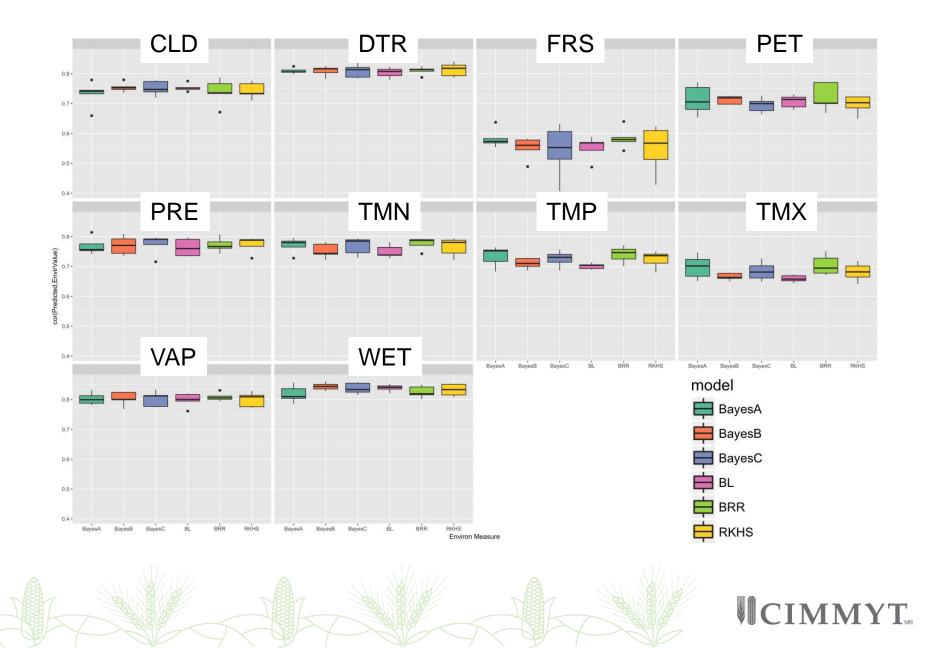
Arcadio Valdes, 2700 acc GWAS panel, 120k markers (single SNP 50bp window) imputed GBS

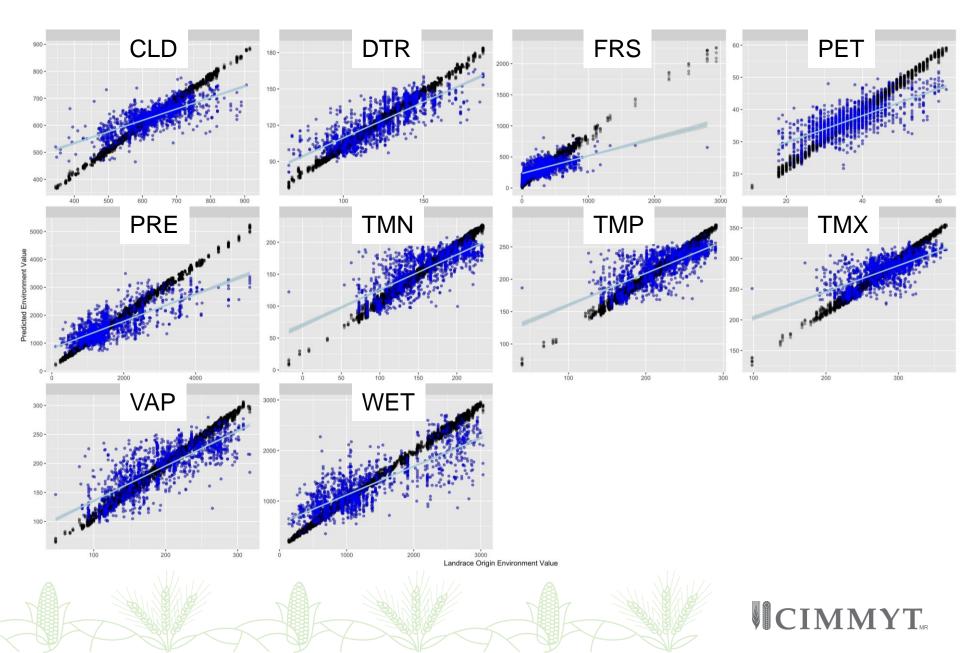
Landrace origin environment

Models calculated over 500 iterations. 5 fold crossvalidation. 10% testing set.



Models calculated over 500 iterations. 5 fold crossvalidation. 10% testing set.



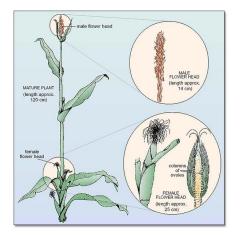


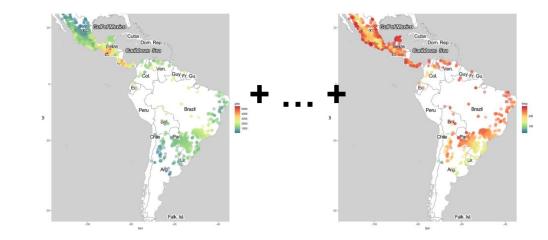
Predicting Flowering Time from Landrace collection site environment?

Field Trial Phenotype (DTA)

Landrace Origin Climatic Data

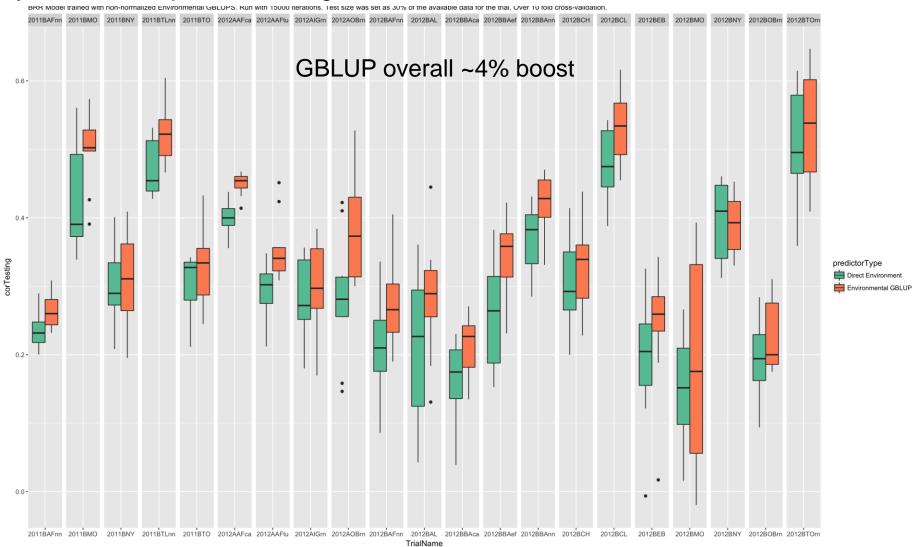
$$\mu + X_{precip}B_{temp} + X_{temp}B_{temp} + ... + X_{env}B_{env} + error$$





Landrace Origin Climatic Data GBLUP

Days to Anthesis predicted using landrace collection site environment- cross trials



Extend from GWAS panel to MGB



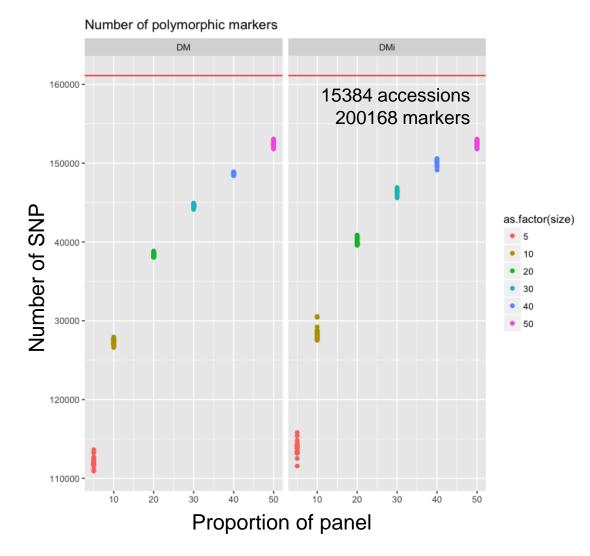
Where now?

Genotypes 24,500 LR 28,500 MGB

Environments 17,500

Phenotypes ?





EnvGS: Best bets - highest potential value (soc-econ links)

EnvGWAS: What is potentially novel



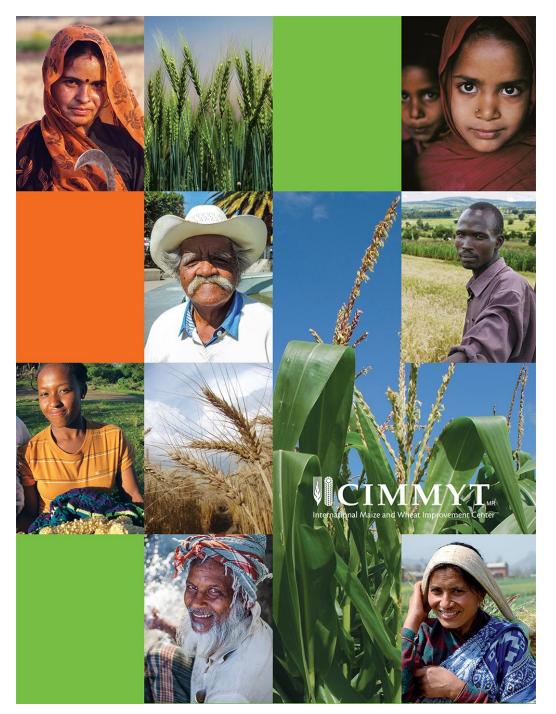


Thank you to fantastic,

colleagues, collaborators and

students







Modernización Sustentable de la Agricultura Tradicional

SAGARPA SECRETARÍA DE AGRICULTURA, GANADERÍA, DESARROLLO RURAL, PESCA Y ALIMENTACIÓN



Thank you for your interest!



research program on Maize



www.seedsofdiscovery.org