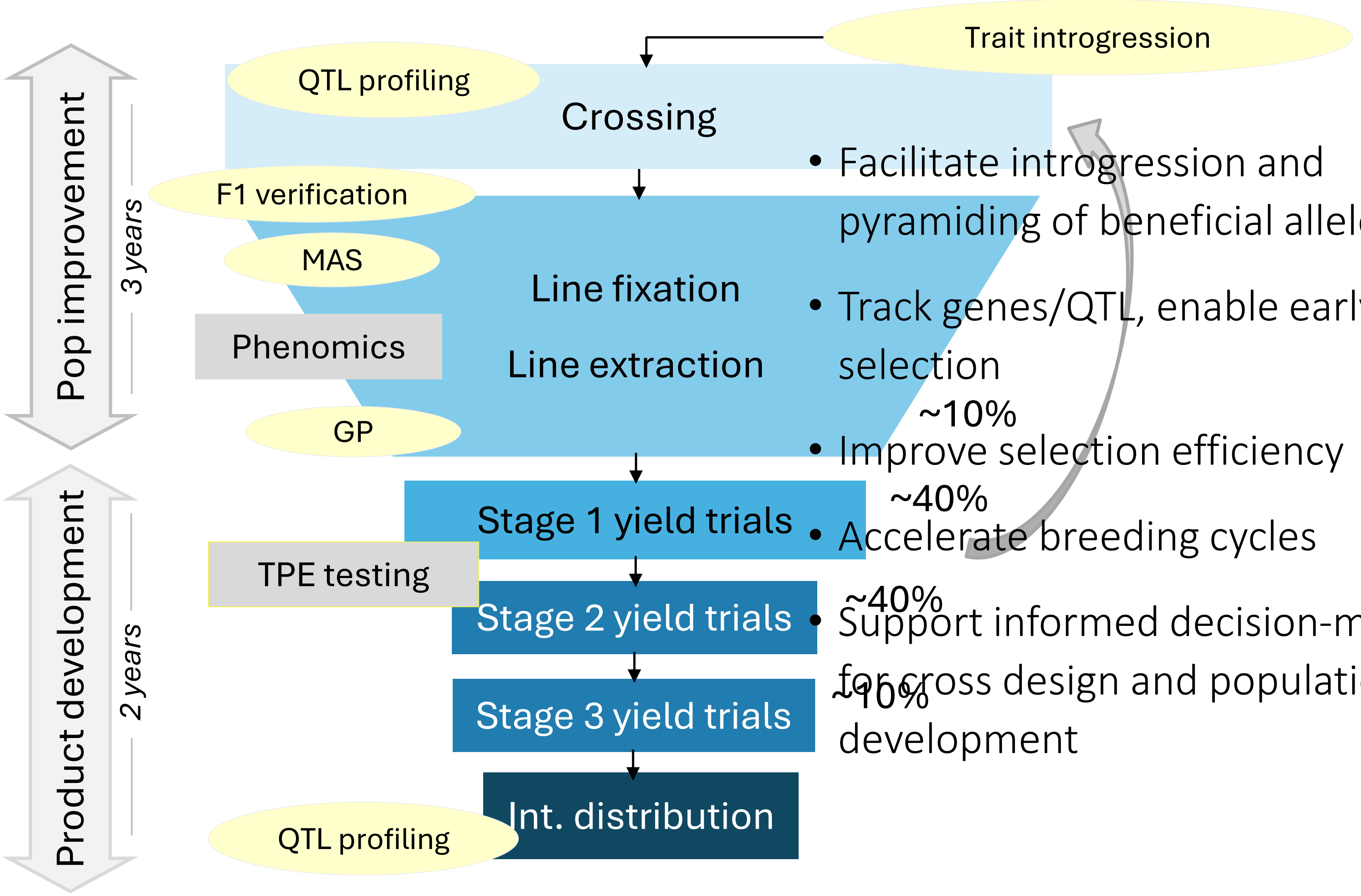
Harnessing sequenced-based haplotypes within IWYP

Susanne Dreisigacker, Matthew Reynolds, Shifeng Cheng, Cong Feng, Jiemeng Xu, Carolina Rivera, Simon Griffith, Jose Crossa

IWYP annual meeting • September-12, 2025

Molecular marker deployment in the GWP



Pop improvement

3 years

Product development

2 years

Trait introgression

QTL profiling

Crossing

F1 verification

MAS

Line fixation

Line extraction

Phenomics

GP

Stage 1 yield trials

TPE testing

Stage 2 yield trials

Stage 3 yield trials

Int. distribution

QTL profiling

~10%

~40%

~40%

~10%

- Facilitate introgression and pyramiding of beneficial alleles
- Track genes/QTL, enable early selection
- Improve selection efficiency
- Accelerate breeding cycles
- Support informed decision-making for cross design and population development

Deployment of IWYP derived molecular markers

Trait introgression

Gene	Trait	Source	Status
TaGW2-6A	Grain weight	IWYP76	DW elite
GL2	Grain length	IWYP76	AL, BC2F3
TaCol-5B	Spike architecture	-	AL
Qht-1B,2B,6D	Heat tolerance	IWYP64	AL, BC2F3
Qmst-3B,5B,7D	Heat/Drought tolerance	SEED	AL, BC2F3
TaGW2-6A,B,D	Grain weight	IWYP76	BC2F3
HB-D2	Spike architecture	IWYP-AP09	BC2F4
7T.D2 (Ae. muticum)	Grain number, biomass	IWYP48	F1Top, BC2F4

Speed Breeding

Advancement 3 to 4 generations/year

CIMMYT Toluca, Mexico

Deployment of IWYP derived molecular markers

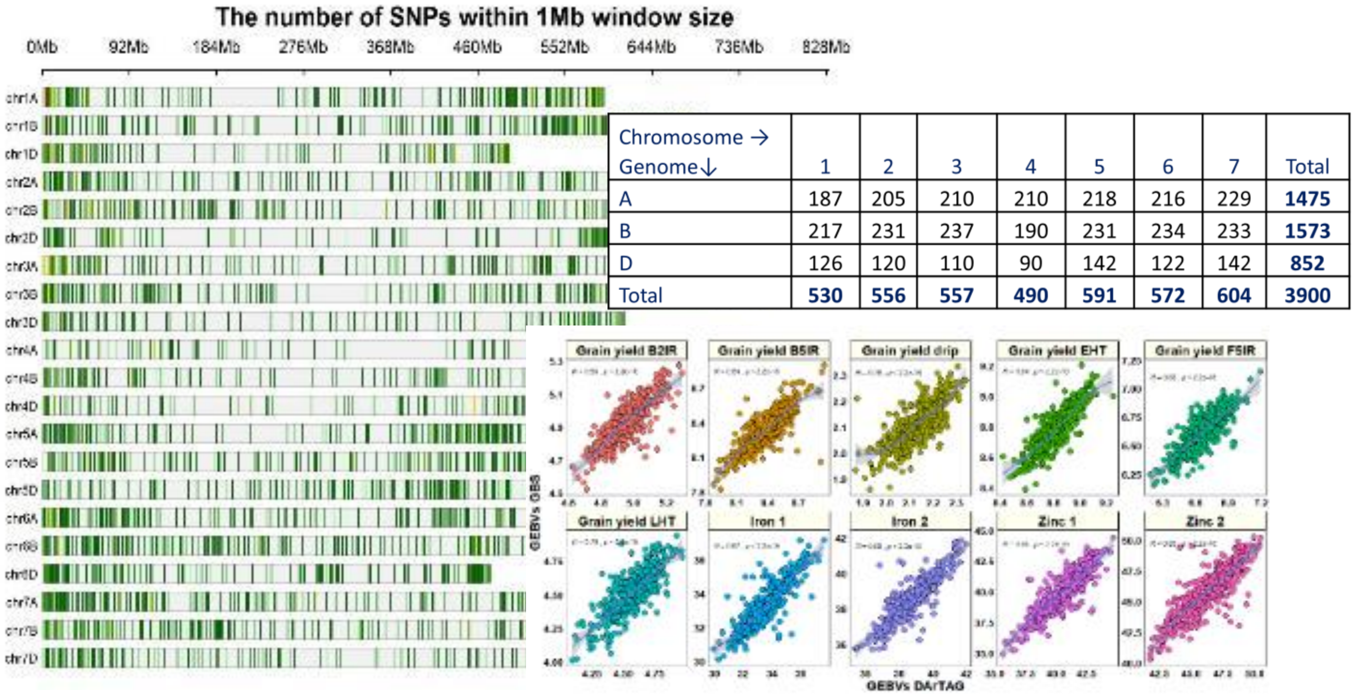
QTL profiling ↔ GP

Gene	Trait	Source
TaToe1-B1	Flowering, HI	IWYP25
TaFT3-B1	Flowering, HI	IWYP25
Qtkw.cim-6A	Grain weight	IWYP76
Qtkw.jic-6A	Grain weight	IWYP76
Wapo-A1	Spikelets/Spike	IWYP708
FT-A2	Spikelets/Spike	IWYP708
GNI1	Grain number	-
GNI2	Grain number	-

TaDArTAG panel 2.0

DArT diversity arrays technology

https://excellenceinbreeding.org/toolbox/services/wheat-39k-mid-density-genotyping-services-0



Chromosome → Genome↓	1	2	3	4	5	6	7	Total
A	187	205	210	210	218	216	229	**1475**
B	217	231	237	190	231	234	233	**1573**
D	126	120	110	90	142	122	142	**852**
Total	**530**	**556**	**557**	**490**	**591**	**572**	**604**	**3900**

New MTA form recent gene discovery panels

Panels

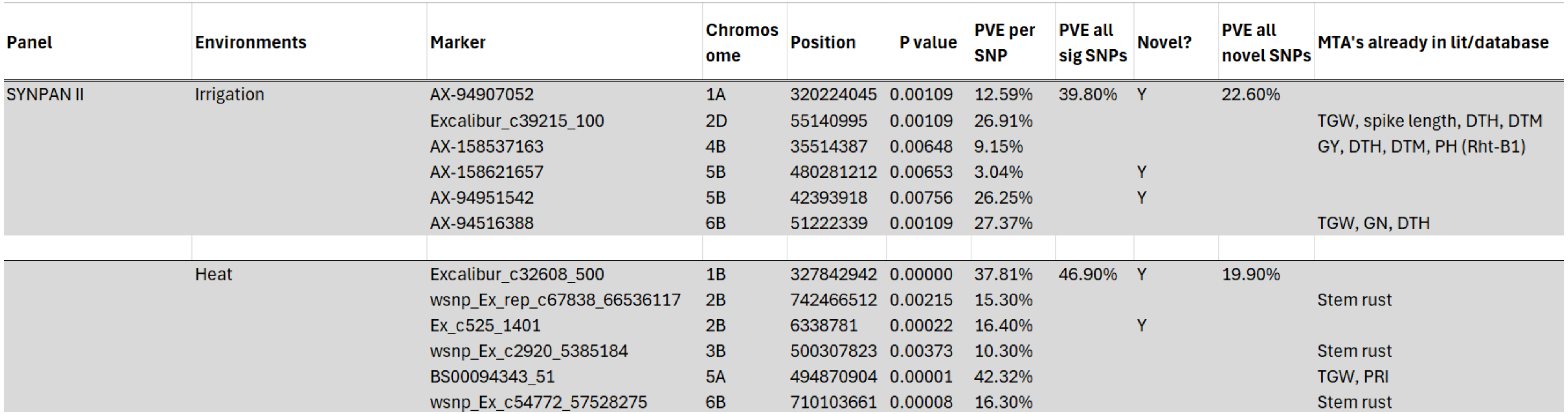
MOLPAN

NAMPAN

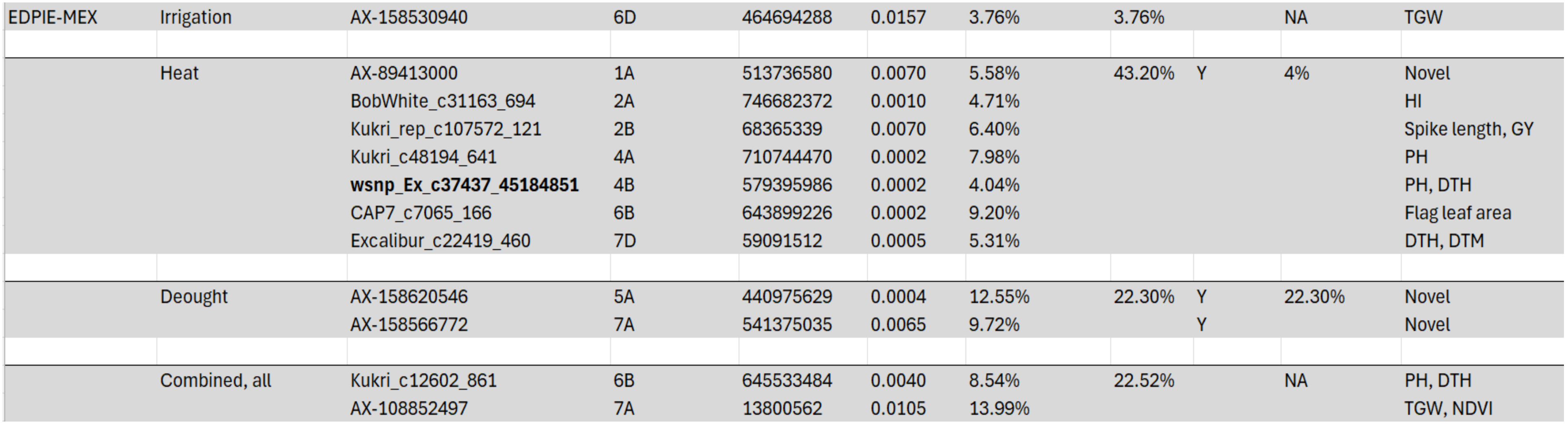
EDPIE

SYNPANII

ISo-D lines



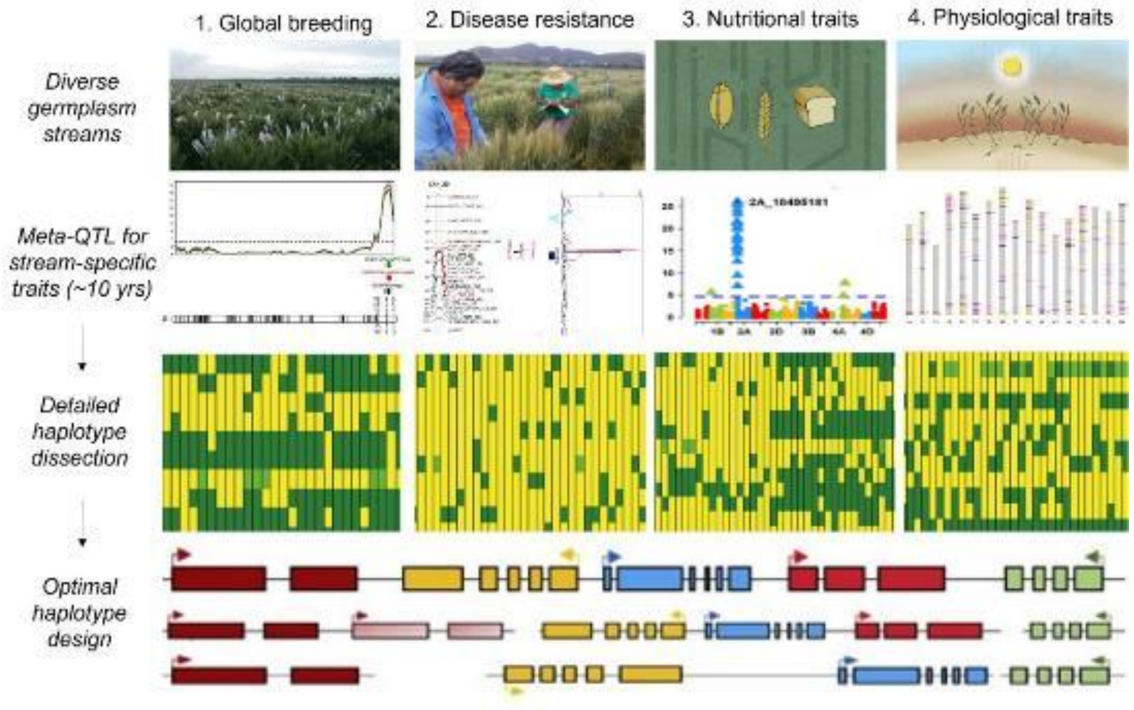
Panel	Environments	Marker	Chromosome	Position	P value	PVE per SNP	PVE all sig SNPs	Novel?	PVE all novel SNPs	MTA's already in lit/database
SYNPAN II	Irrigation	AX-94907052	1A	320224045	0.00109	12.59%	39.80%	Y	22.60%	
		Excalibur_c39215_100	2D	55140995	0.00109	26.91%				TGW, spike length, DTH, DTM
		AX-158537163	4B	35514387	0.00648	9.15%				GY, DTH, DTM, PH (Rht-B1)
		AX-158621657	5B	480281212	0.00653	3.04%		Y		
		AX-94951542	5B	42393918	0.00756	26.25%		Y		
		AX-94516388	6B	51222339	0.00109	27.37%				TGW, GN, DTH
	Heat	Excalibur_c32608_500	1B	327842942	0.00000	37.81%	46.90%	Y	19.90%	
		wsnp_Ex_rep_c67838_66536117	2B	742466512	0.00215	15.30%				Stem rust
		Ex_c525_1401	2B	6338781	0.00022	16.40%		Y		
		wsnp_Ex_c2920_5385184	3B	500307823	0.00373	10.30%				Stem rust
		BS00094343_51	5A	494870904	0.00001	42.32%				TGW, PRI
		wsnp_Ex_c54772_57528275	6B	710103661	0.00008	16.30%				Stem rust



EDPIE-MEX	Irrigation	AX-158530940	6D	464694288	0.0157	3.76%	3.76%		NA	TGW
	Heat	AX-89413000	1A	513736580	0.0070	5.58%	43.20%	Y	4%	Novel
		BobWhite_c31163_694	2A	746682372	0.0010	4.71%				HI
		Kukri_rep_c107572_121	2B	68365339	0.0070	6.40%				Spike length, GY
		Kukri_c48194_641	4A	710744470	0.0002	7.98%				PH
		wsnp_Ex_c37437_45184851	4B	579395986	0.0002	4.04%				PH, DTH
		CAP7_c7065_166	6B	643899226	0.0002	9.20%				Flag leaf area
		Excalibur_c22419_460	7D	59091512	0.0005	5.31%				DTH, DTM
	Deought	AX-158620546	5A	440975629	0.0004	12.55%	22.30%	Y	22.30%	Novel
		AX-158566772	7A	541375035	0.0065	9.72%		Y		Novel
	Combined, all	Kukri_c12602_861	6B	645533484	0.0040	8.54%	22.52%		NA	PH, DTH
		AX-108852497	7A	13800562	0.0105	13.99%				TGW, NDVI

GlobalWheatG2B: Harnessing sequence-based haplotypes

Collaboration with AGIS-CAAS and JIC



WP1. Foundational and structural analyses
- Population dynamics
- Linkage and haplotype structures
- Introgressions

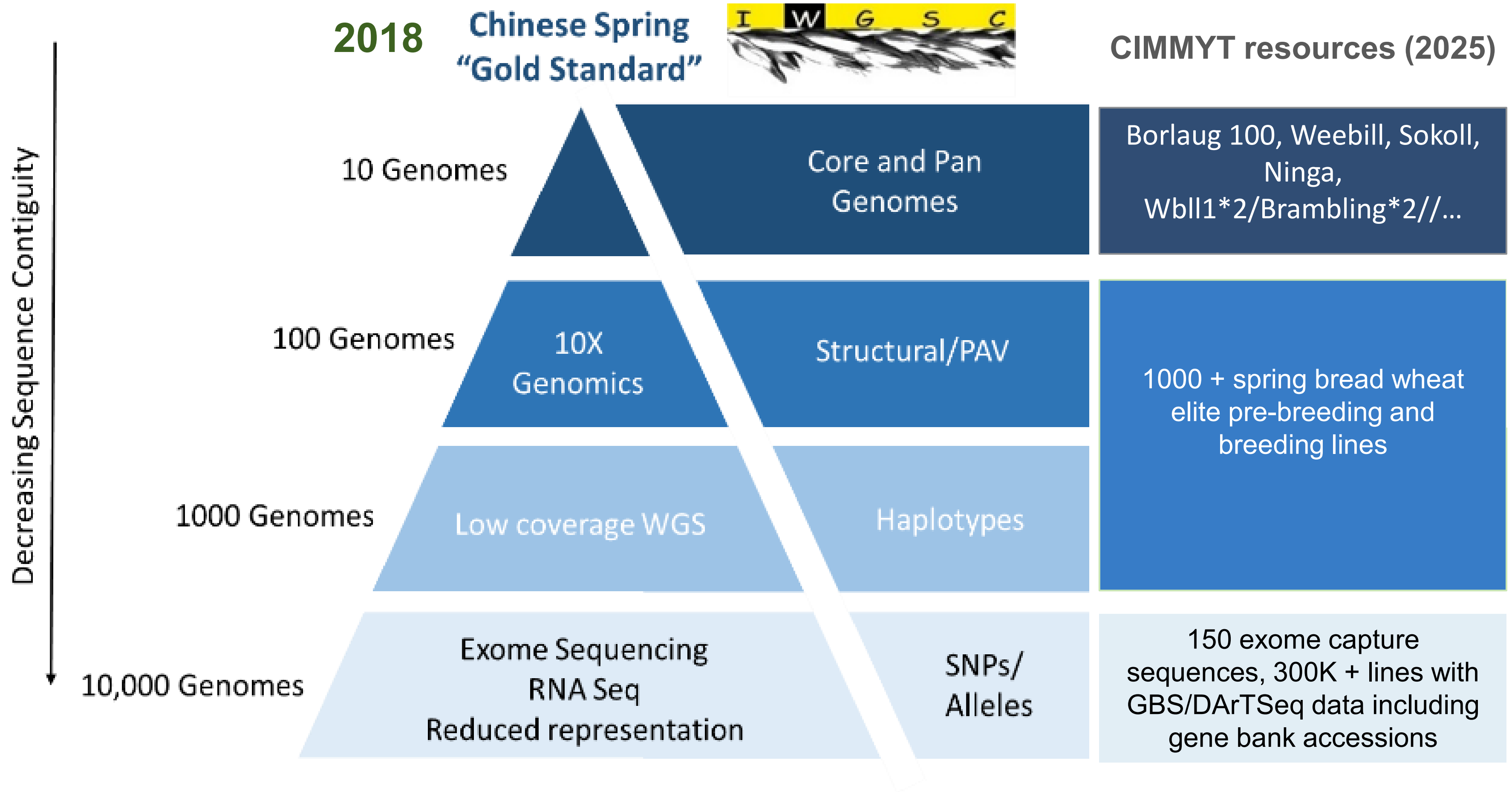
WP2. Gene discovery (focus on physiological traits)

WP3. In silico gene and haplotype mining

WP4. Genomic selection using whole genome sequence data

Adoption of Sequence-based Genomic Tools

Current CIMMYT sequence resources



Decreasing Sequence Contiguity

2018 **Chinese Spring "Gold Standard"** IWGSC

CIMMYT resources (2025)

10 Genomes — Core and Pan Genomes — Borlaug 100, Weebill, Sokoll, Ninga, Wbll1*2/Brambling*2//…

100 Genomes — 10X Genomics — Structural/PAV

1000 Genomes — Low coverage WGS — Haplotypes

1000 + spring bread wheat elite pre-breeding and breeding lines

10,000 Genomes — Exome Sequencing RNA Seq Reduced representation — SNPs/ Alleles — 150 exome capture sequences, 300K + lines with GBS/DArTSeq data including gene bank accessions

IWYP International Wheat Yield Partnership

Research to Deliver Wheat for the Future

Watkins & Worldwide Wheat G2B

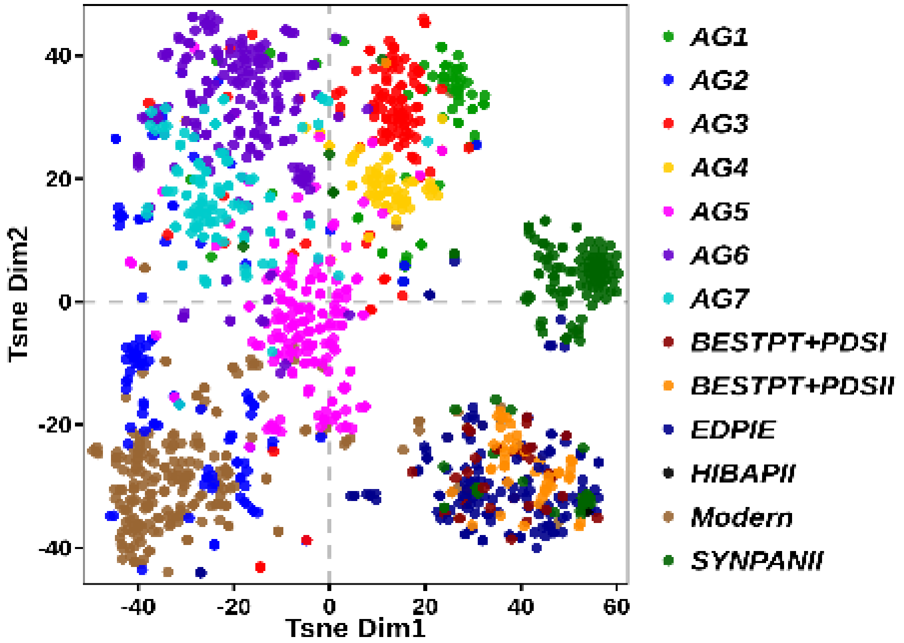
Decode, Dissect, Discover, Design and Deliver

Foundational and structural sequence analyses

Cong Feng,
AGIS-CAAS

Mayor steps

- Sequencing of approx. 900 CIMMYT Wheat lines (several hundred Gb)
- SNP Calling and filtering
- Mapping WatkinsG2B SNPs to CS2.1
- Merging CIMMYT and WatkinsG2B data
- Comparative population structure analyses
- Building haplotype maps
- Evaluation of diverse structural elements



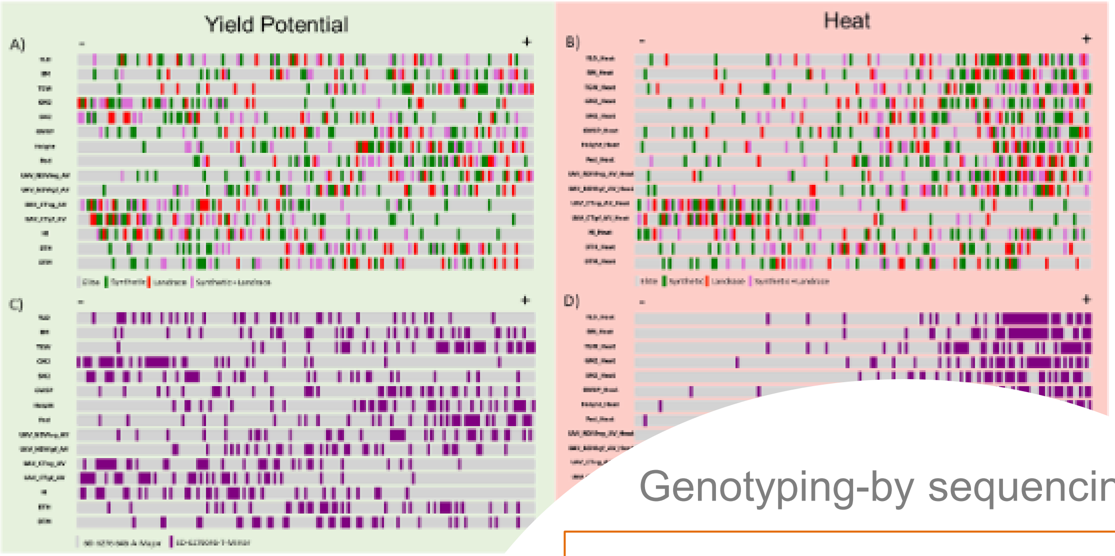
First observations of the initial pop structure:

1. CIMMYT wheat and a few modern varieties in WatkinsG2B have similar genetic backgrounds, but the majority is different
2. Synthetic hexaploidy wheat for a different group

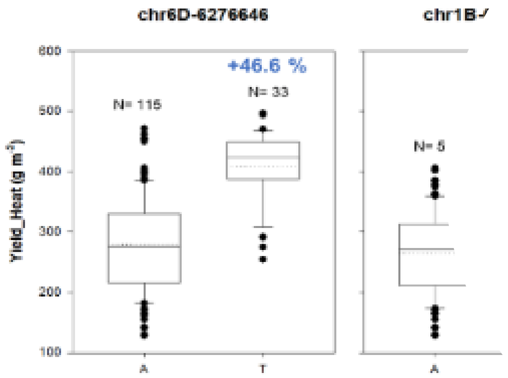
Gene discovery in HIBAP, EDPIE and SYNPAN

HIBAP1

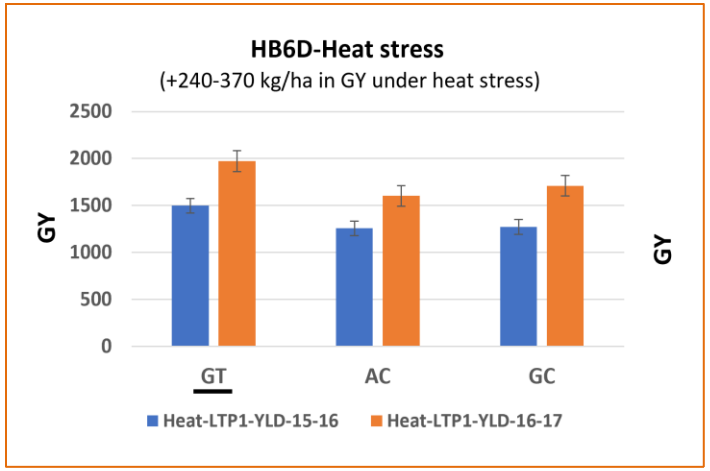
Exome Capture



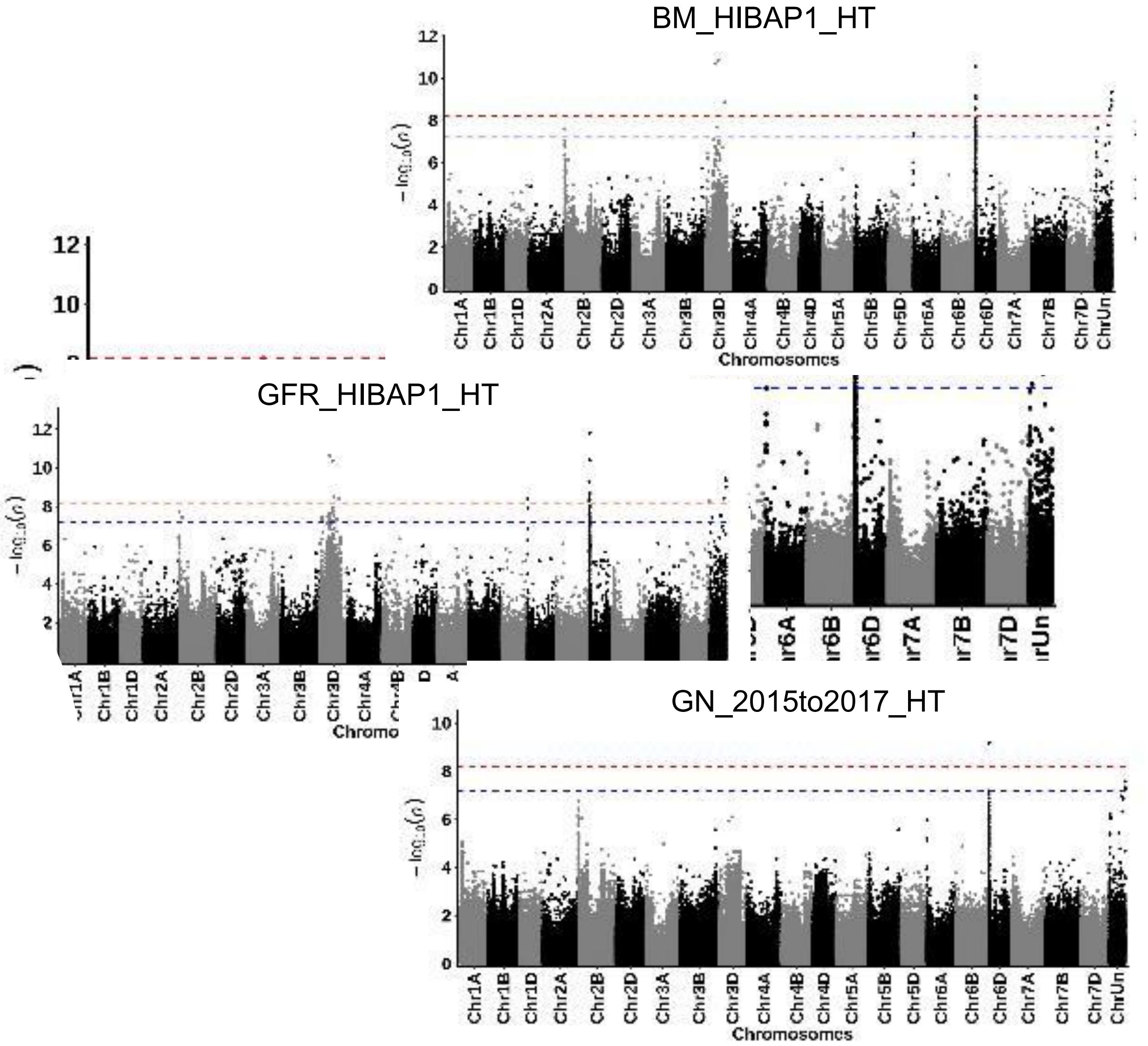
Genotyping-by sequencing



Molero et al., 2023

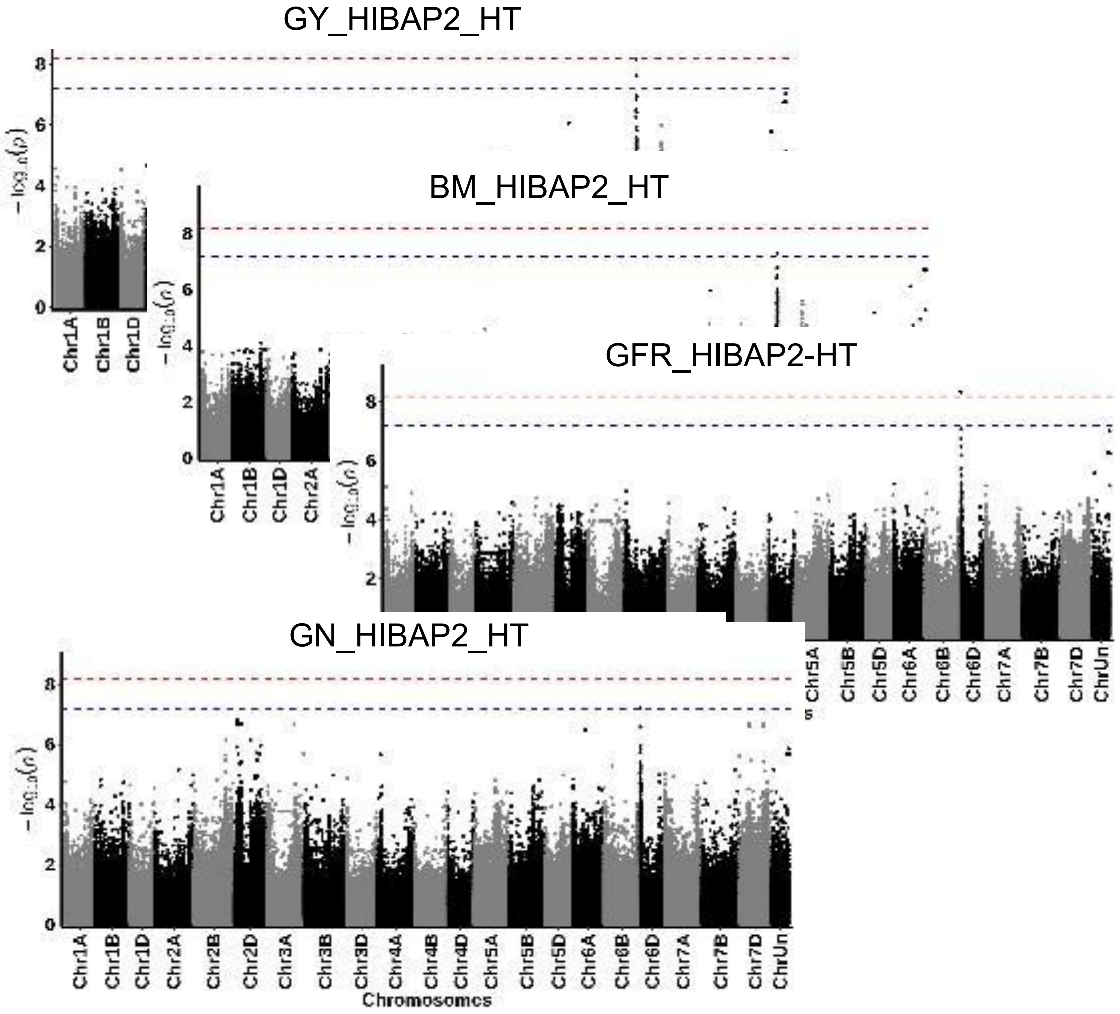


Singh et al. 2021



Gene discovery in HIBAP, EDPIE and SYNPAN

HIBAP2



HIBAP3

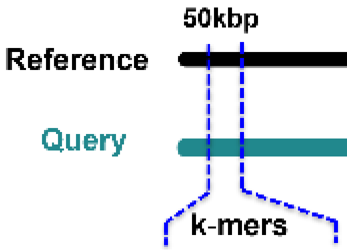
Next steps

- Run sequenced-based GWAS for all relevant HIBAP traits
- Explore haplotype diversity around major MTA
- Mine different haplotypes across CIMMYT elite lines and the Watkins collection
- Develop more presice, candidate gene-based molecular markers for breeding

Determination of recombination blocks via *kmer* comparisons

IBSpy
GitHub

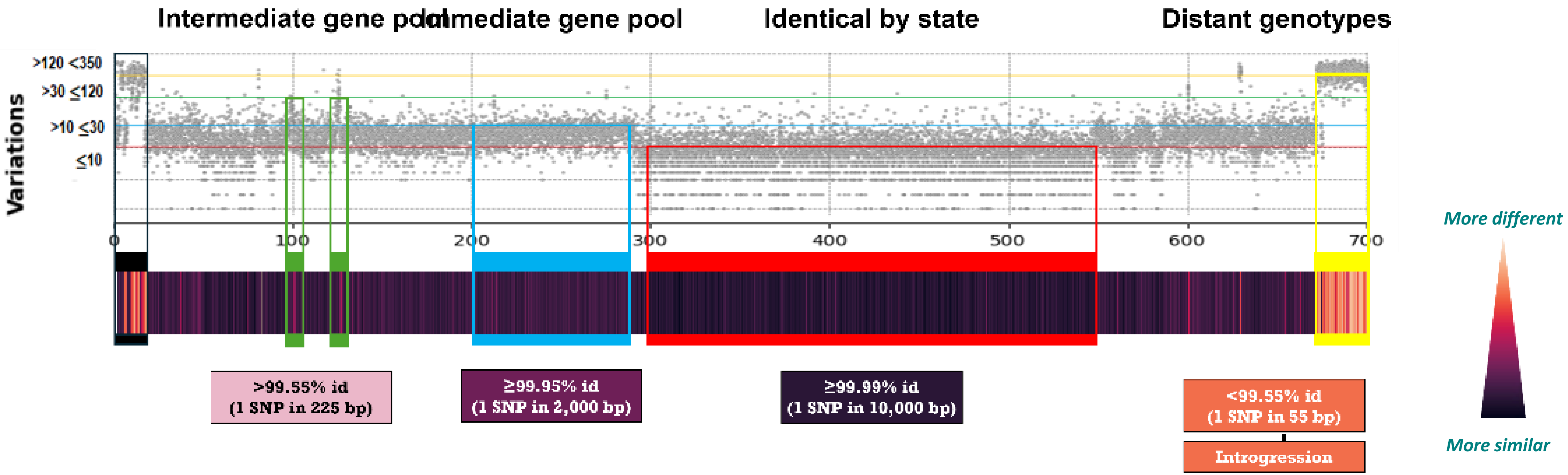
Identity-By-State python



Pipeline to determine recombination points blocks

Ricardo Ramírez González

Jesús Quiroz Chávez



>99.55% id
(1 SNP in 225 bp)

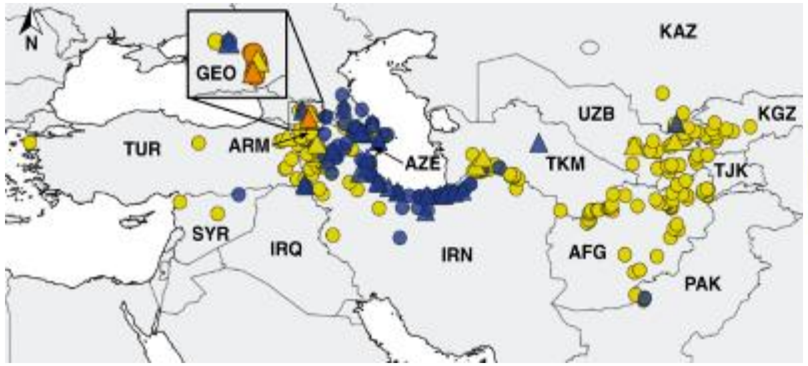
≥99.95% id
(1 SNP in 2,000 bp)

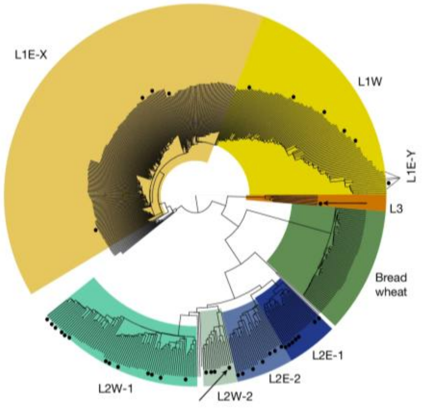
≥99.99% id
(1 SNP in 10,000 bp)

<99.55% id
(1 SNP in 55 bp)

Introgression

Determining the D-genome donors in the CIMMYT elite bread wheat line Borlaug 100





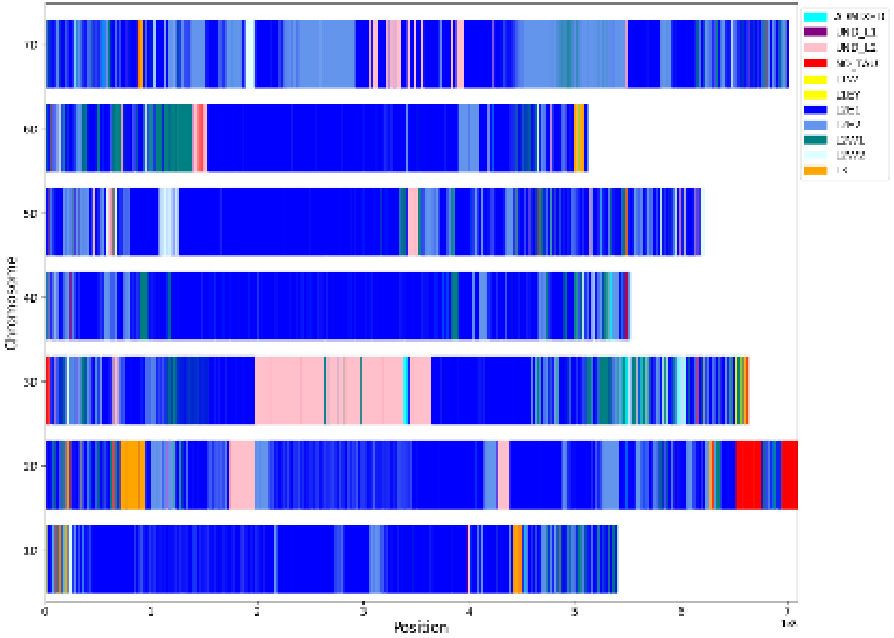
Gaurav et al., 2022

- *Ae. tauschii* can be characterized into three linages
- Linage 2 as the closest donor of the wheat D- genome

Fernando Galvan Delgadillo

The D-genome of Borlaug 100 is defined by approximately 95% by *Ae. tauschii* Lineage 2

CIMMYT reference genome: Borlaug 100

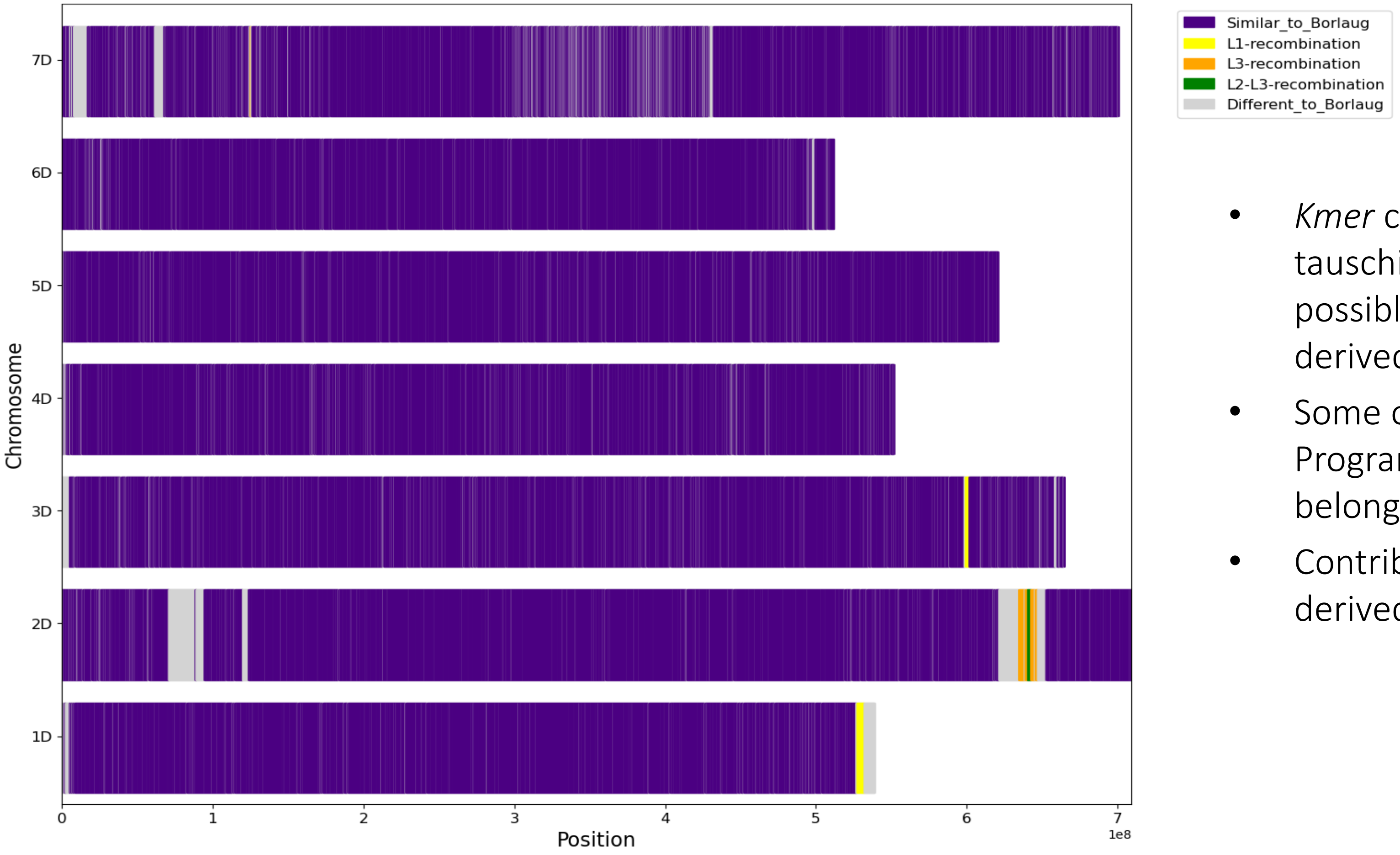


Lineages	Subpopulation	Contribution
Lineage 2	L2E1 L2E2 L2W1 L2W2 UND_L2	55.5% 17.6% 5.9% 1.8% 14.2% **95.04%**
Lineage 1	L1W **+** L1EY **+** UND_L1	**0.69 < 1%**
Lineage 3	–	**2.4%**
ADMIXED	–	**0.6%**
NO_TAU	–	**1.2%**

Leveraging genetic innovations for accelerated breeding of climate resilient and nutritious crops
https://www.cabi.org/uk-cgiar-centre/crop-genetic-improvement-for-future-climate-resilience/

Estimating the contribution of *Ae. tauschii* in synthetic-derived lines using *k-mer* comparisons

MUCUY *Ae. tauschii* (224)



- *Kmer* comparison with sequence assemblies from the mayor tauschii lineages can help to determine recombination or possible contribution of Ae. tauschii accessions in Synthetic derived lines
- Some of the most used SHW in the CIMMYT Global Wheat Program include *Ae. tauschii* (205), *Ae. tauschii* (224) which belong to Lineage 1
- Contributions of *Ae. tauschii* accessions differ in synthetic derived lines and vary across the genome

Genomic selection using whole genome sequences

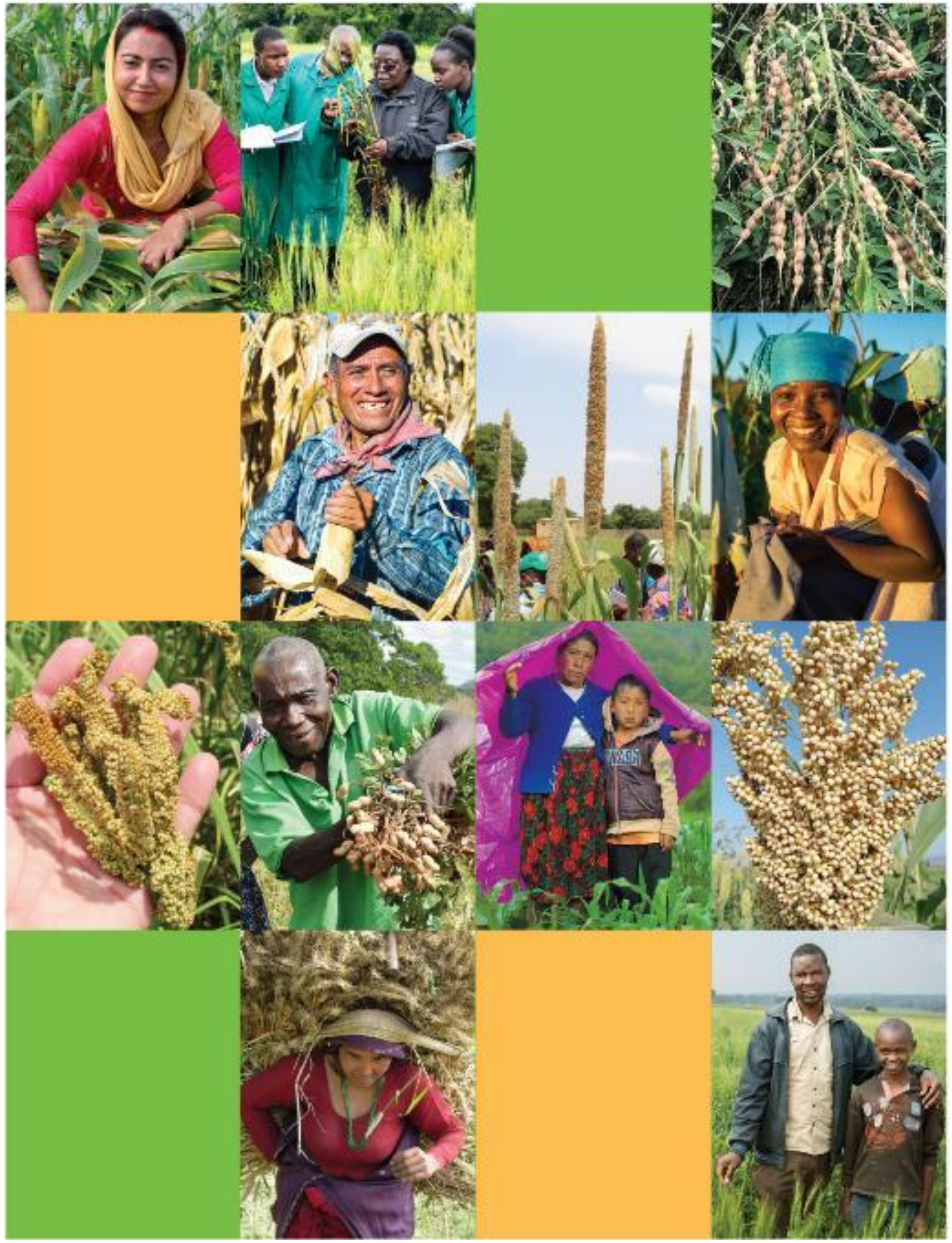

Osval Montesinos

1. Cross validation in EDPIE: 150 lines with phenotypic data (DTH, DTM, PH, GY) across 22 international environments
2. Whole genome sequence data

   Filtering criteria:

   - **Filtering by Read Depth (DP):** the number of sequencing reads. Two ranges:10-50, 51-100
   - **Filtering by Genotype Quality (GQ)**: the confidence that the assigned genotype (e.g., 0/0, 0/1, 1/1) is correct. GQ >= 30 and GQ >= 60.
   - **Filtering by Allelic Depth (AD):** the number of reads supporting each allele: AD>=5 and AD>=20.
   - **Filtering by Genotype Likelihood (PL):** PL gives the relative probability of each genotype PL>=30 and PL>=70
3. Prediction questions:
   - Increase of prediction accuracy in a multi-environmental context
   - Effect of different filtering treatments
   - Can prediction models inform on variants to be included in routinely used SNP panels

Treatments	min_DP	max_DP	GQ	AD	PL
T1	10	50	60	5	30
T2	10	50	60	5	70
T3	10	50	60	20	30
T4	10	50	60	20	70
T5	10	50	30	5	30
T6	10	50	30	5	70
T7	10	50	30	20	30
T8	10	50	30	20	70
T9	51	100	60	5	30
T10	51	100	60	5	70
T11	51	100	60	20	30
T12	51	100	60	20	70
T13	51	100	30	5	30
T14	51	100	30	5	70
T15	51	100	30	20	30
T16	51	100	30	20	70



Thank you for your interest!

CIMMYT | CGIAR