

Supplementary data

Genomic prediction for yield improvement and safeguarding genetic diversity in CIMMYT spring wheat (*Triticum aestivum* L.)

Ephrem Habyarimana

Table S1. Description of 19 traits measured in up to six trials in two CIMMYT populations, the 29th (C29SAWSN) and 30th (C30SAWSN) semi-arid bread wheat screening nurseries.

Trait abbreviation	Trait description	Population	Location, year and treatment
<i>Disease resistance</i>			
FHBIND	Fusarium head blight (FHB) index: the product of FHB severity and FHB incidence multiplied by 100	C29SAWSN	Batan 2010
LR	Leaf rust: score severity 1-100	C29SAWSN	Batan 2009 and 2010, Obregon 2009: small observation plots
		C30SAWSN	Batan 2010 and 2011, Obregon 2010: small observation plots
SB	Spot blotch (AUDPC, area under disease progress curve)	C29SAWSN	Agua Fria 2010 to 2012
SR	Stem rust: score severity 1-100	C29SAWSN	Njoro 2009 and 2010
YR	Yellow rust: score severity 1-100	C29SAWSN	Toluca 2009; Santa Catalina (Ecuador) 2009
		C30SAWSN	Boximo 2010; Toluca 2010; Santa Catalina (Ecuador) 2011
<i>Agronomic traits</i>			
DTH	Days to heading: number of days from planting to heading	C29SAWSN	Batan 2009; Toluca 2009; Obregon 2009: full irrigation bed, drought bed, and small observation plots; Obregon 2010: drought bed, drought flat, full irrigation bed, and small observation plots; Agua Fria 2010 and 2011
		C30SAWSN	Obregon 2010: drought bed, full irrigation bed; Obregon 2011: full irrigation bed, full irrigation flat, heat bed, and drought flat
PH	Plant height (cm)	C29SAWSN	Toluca 2009; Obregon 2009: drought bed, full irrigation bed; Obregon 2010: drought bed, drought flat, full irrigation bed, small observation plots
GY	Grain yield (t ha ⁻¹)	C29SAWSN	Obregon 2009: drought bed, full irrigation bed; Obregon 2010: heat bed, drought bed, drought flat, full irrigation bed, full irrigation flat

C30SAWSN Obregon2010: drought bed, full irrigation bed;
Obregon 2011: heat bed, drought flat, full
irrigation bed, full irrigation flat

Table S1. Continued.

Trait abbreviation	Trait description	Population	Location, year and treatment
<i>Bread-making quality</i>			
ALVPL	Alveograph ratio between dough elasticity (P) and dough extensibility (L)	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
FLPRO	Flour protein content at 14% moisture basis (%)	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
FLRSDS	Flour sodium dodecyl sulfate sedimentation (mL g ⁻¹)	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
GRNHRD	Grain hardness index (%): based on particle size index [Approved Method 55-30, AACC International (2000)]	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
GRPRO	Grain protein content at 12.5% moisture (%)	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
LOFVOL	Bread loaf volume (mL x 100g flour ⁻¹)	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
MIXTIM	Mixograph mixing time (min) for optimum dough development	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
MIXTYP	Mixographic type: 1 (weak) to 8 (strong) dough mixing tolerance scores	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
SDSEDIND	Sodium dodecyl sulfate (SDS) sedimentation index: SDS sedimentation divided by protein concentration in flour	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
TESTWT	Test weight (kg hL ⁻¹)	C29SAWSN	Obregon 2009: full irrigation bed; Obregon 2010: drought bed
TKW	Thousand kernel weight (g)	C29SAWSN	Obregon 2010: heat bed, full irrigation bed
		C30SAWSN	Obregon 2011: heat bed, full irrigation bed

[†] Bed: each entry is planted on 2 beds, 80cm wide, 2.8 m long, with 3 rows per bed. Flat: flat planting on 1.3 x 4.5 m plots, Drought: one pre-seeding irrigation and one irrigation after germination applied to create very severe drought stress conditions. Full irrigation: 5 irrigations applied. Heat: late sowing (90 days delay) resulting in considerable heat stress at flowering and grain filling.

Table S2. Mean cross-validation accuracy for plant height, days to heading and grain yield measured in the C29SAWSN (29th CIMMYT semi-arid wheat screening nursery), using six models, and different numbers of genotyping-by-sequencing (GBS) markers.

[§] N.GBS	[†] BL	PBL	BRR	PBRR	RKHS	PRKHS	Overall
2013	0.68	0.71	0.68	0.70	0.68	0.69	0.69
10000	0.67	0.69	0.67	0.69	0.67	0.68	0.68
20000	0.69	0.70	0.69	0.70	0.67	0.69	0.69
33762	0.70	0.70	0.70	0.70	0.67	0.68	0.69

[†]BL: Bayesian LASSO; BRR: Bayesian ridge regression; PBL: Bayesian LASSO with pedigree; PBRR: Bayesian ridge regression with pedigree; PRKHS: reproducing kernel Hilbert spaces with pedigree; RKHS: reproducing kernel Hilbert spaces; [§]N.GBS: number of GBS markers tested.

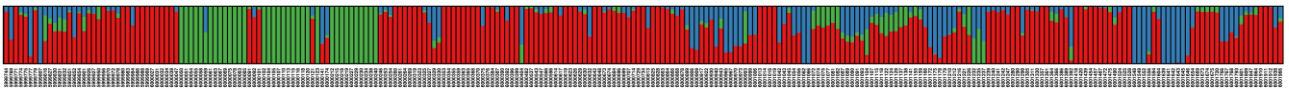


Fig S1. GBS-based population structure (K=3) of the 29th CIMMYT semi-arid wheat screening nursery (C29SAWSN) estimated using a model-based Bayesian algorithm implemented in the STRUCTURE software, and using an admixture model. Labels below identify the genotypes. Each individual is represented by a vertical box that is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters. Refer to the text for further details.

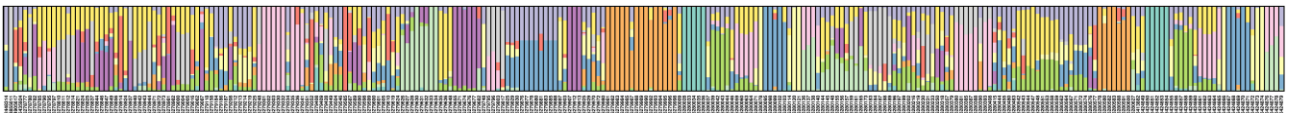


Fig S2. GBS-based population structure (K=12) of the 30th CIMMYT semi-arid wheat screening nursery (C30SAWSN) estimated using a model-based Bayesian algorithm implemented in the STRUCTURE software, and using an admixture model. Labels below identify the genotypes. Each individual is represented by a vertical box that is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters. Refer to the text for further details.

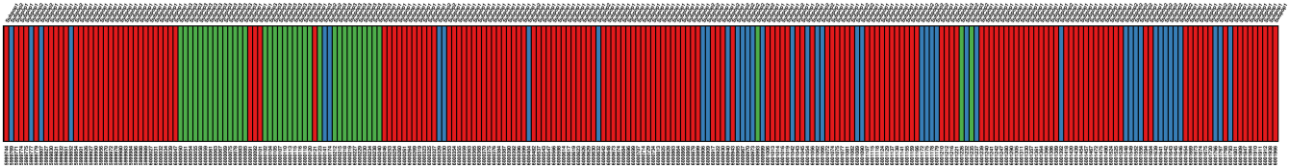


Fig S3. GBS-based population structure ($K=3$) of the 29th CIMMYT semi-arid wheat screening nursery (C29SAWSN) estimated using a model-based Bayesian algorithm implemented in the STRUCTURE software, and setting full membership threshold at 0.50. Each individual is represented by a vertical box whose color uniquely identifies the cluster the individual belongs to. Labels below and above identify genotypes and subpopulations, respectively. Refer to the text for further details.

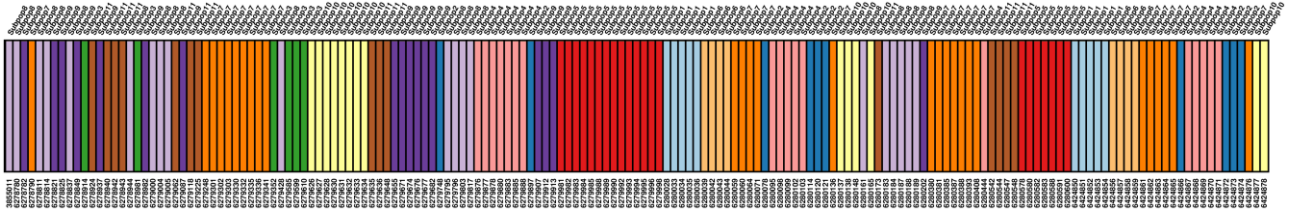


Fig S4. GBS-based population structure ($K=11$) of the 30th CIMMYT semi-arid wheat screening nursery (C30SAWSN) estimated using a model-based Bayesian algorithm implemented in the STRUCTURE software, and setting full membership threshold at 0.50. Each individual is represented by a vertical box whose color uniquely identifies the cluster the individual belongs to. Labels below and above identify genotypes and subpopulations, respectively. Refer to the text for further details.