

Supplementary Materials

Leveraging Data from the Genomes to Fields Initiative to Investigate G×E in Maize in North America

Marco Lopez-Cruz, Fernando M. Aguade, Jacob D. Washburn, Natalia de Leon, Shawn M. Kaeppler,
Dayane Lima, Ruijuan Tan, Addie Thompson, Laurence Willard De La Bretonne, and Gustavo de los Campos

This file contains:

- Supplementary Figures S1-S8
- Supplementary Tables S1-S4
- Supplementary Files S1-S3

1. Supplementary Figures

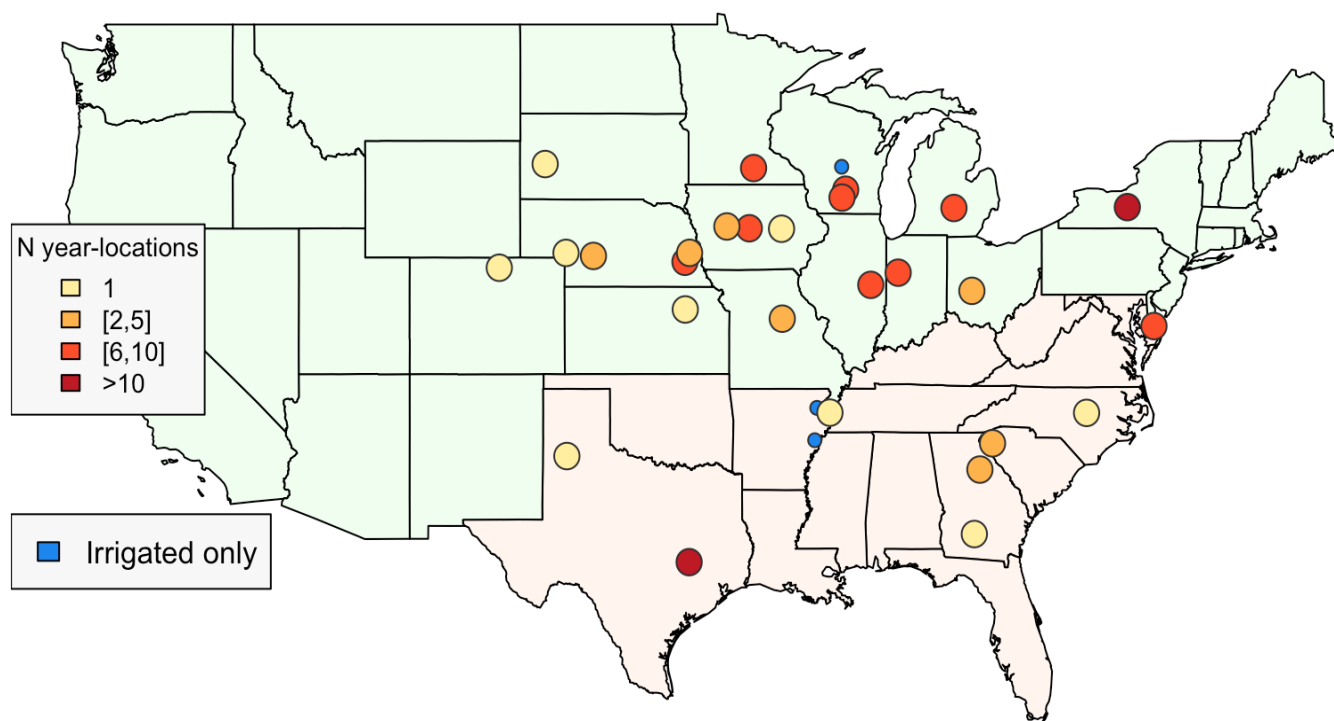


Figure S1. Testing locations in Genomes to Fields trials (only non-irrigated locations were used in this study)

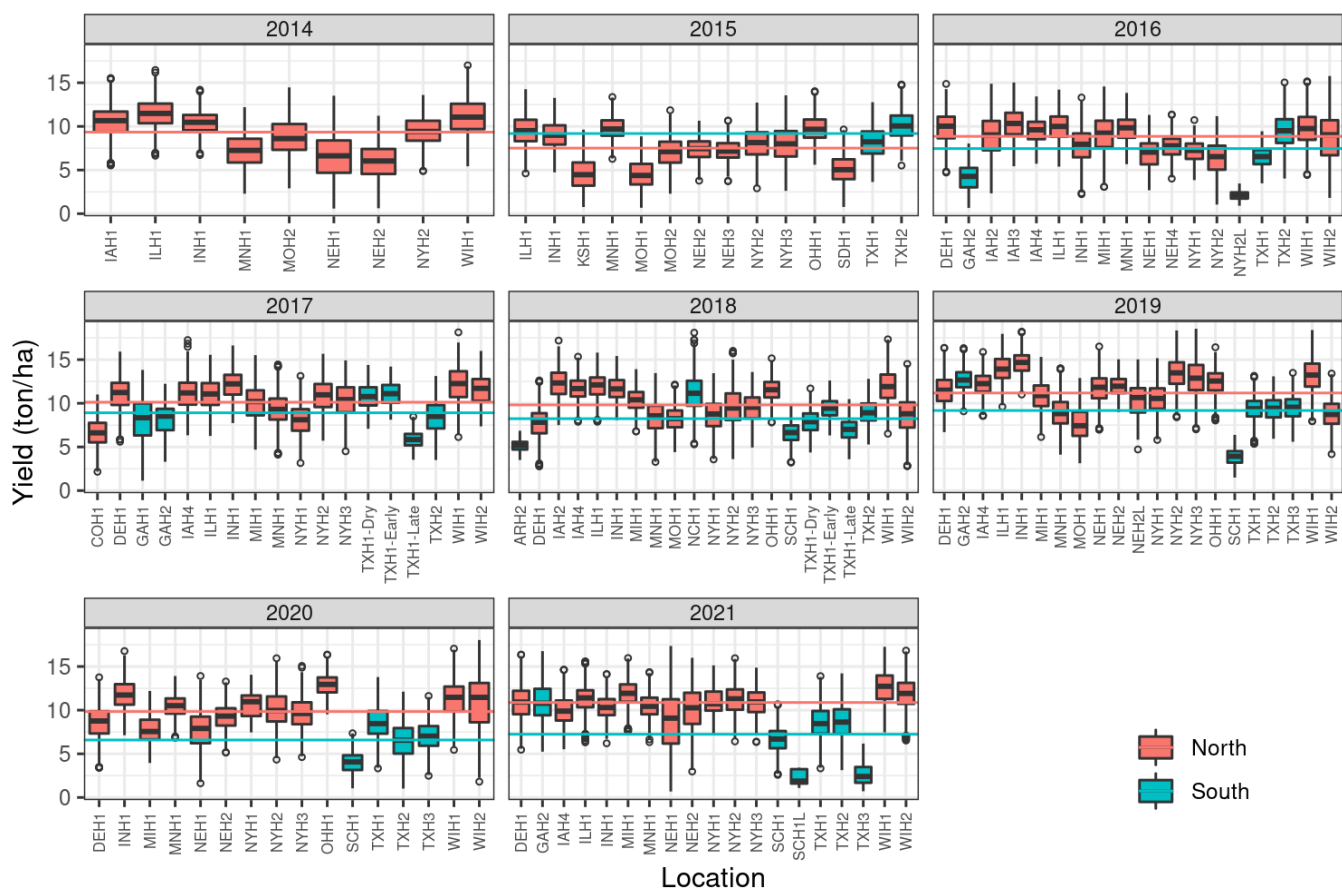


Figure S2. Boxplot of grain yield by year, region, and location within region.

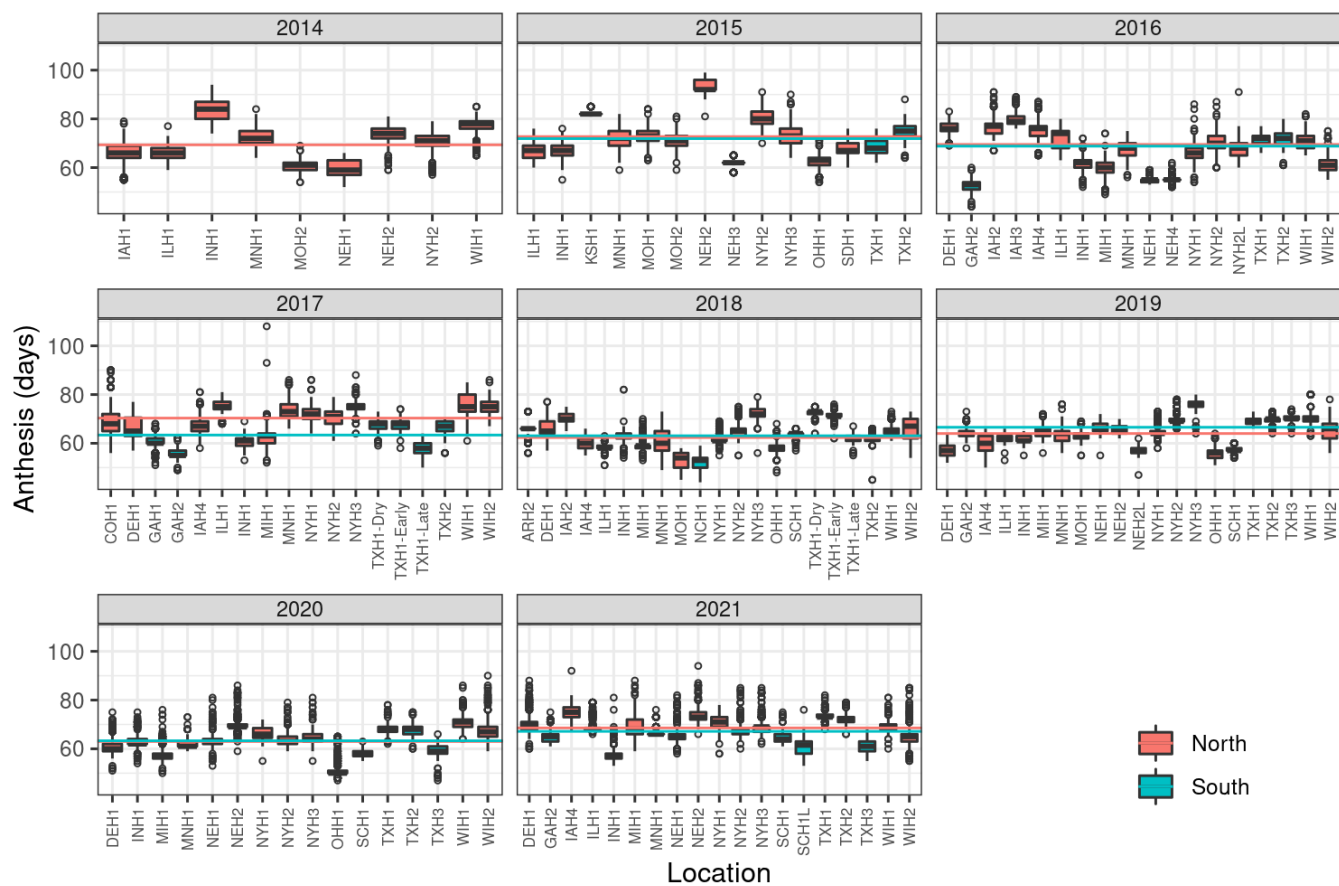


Figure S3. Boxplot of days-to-anthesis by year, region, and location within region.

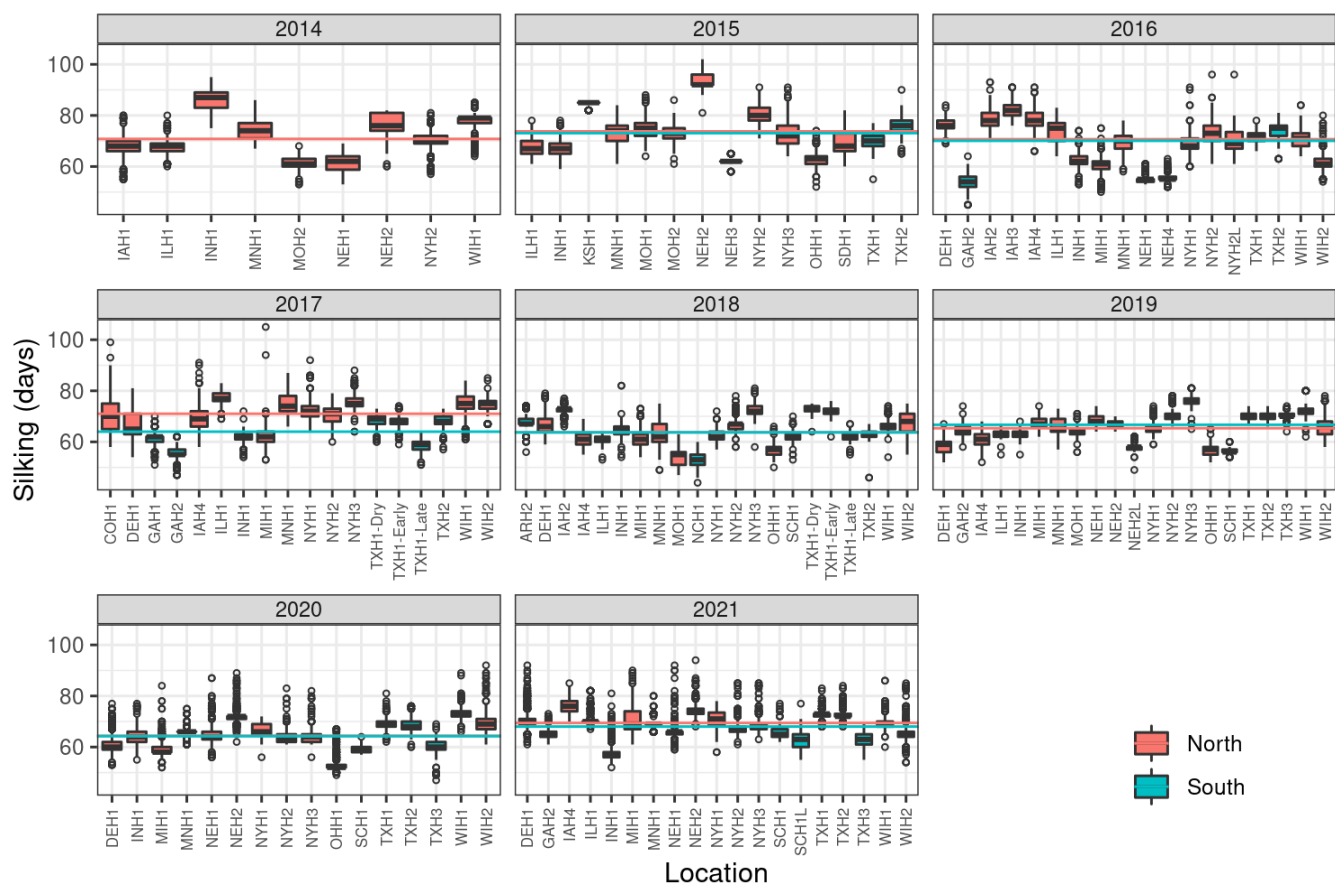


Figure S4. Boxplot of days-to-anthesis by year, region, and location within region.

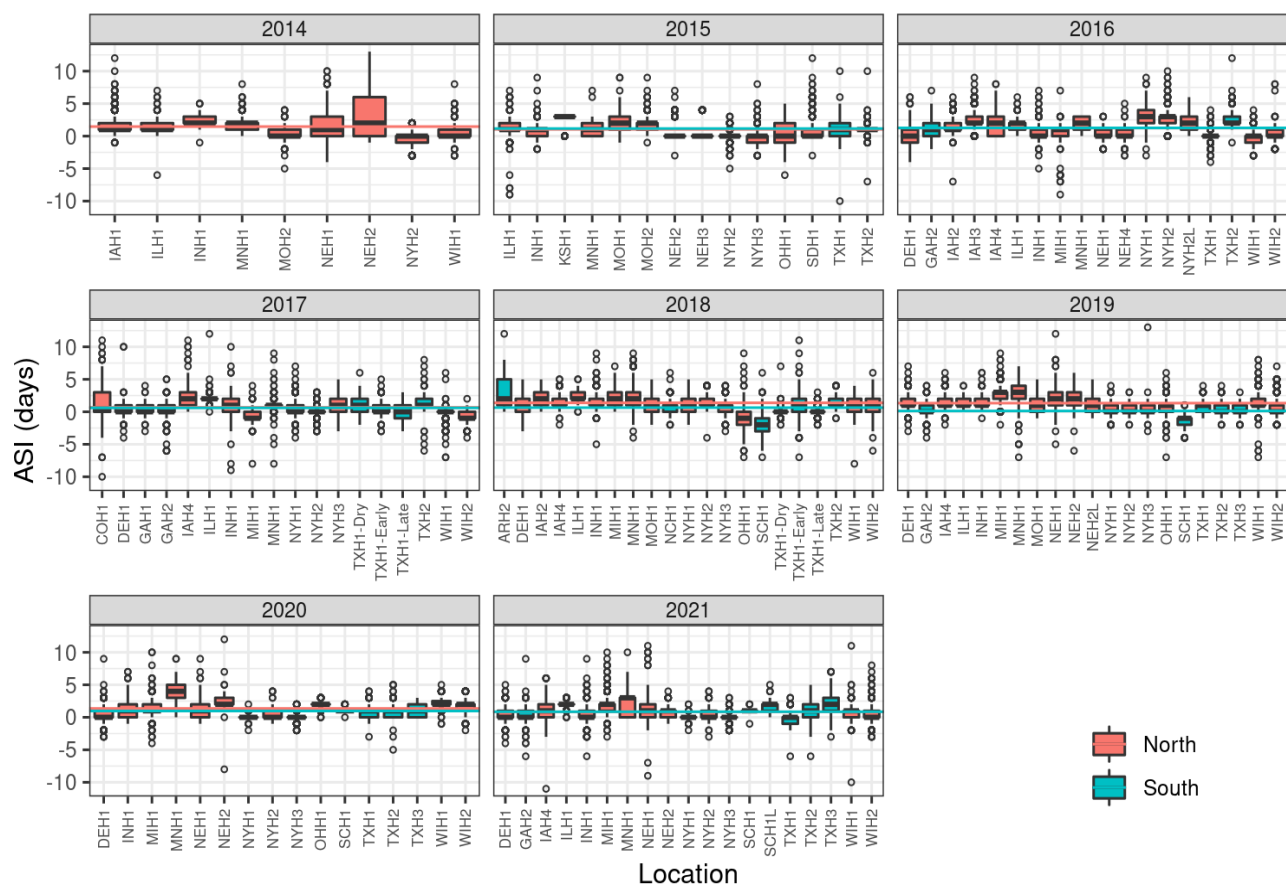


Figure S5. Boxplot of days-to-anthesis by year, region, and location within region.

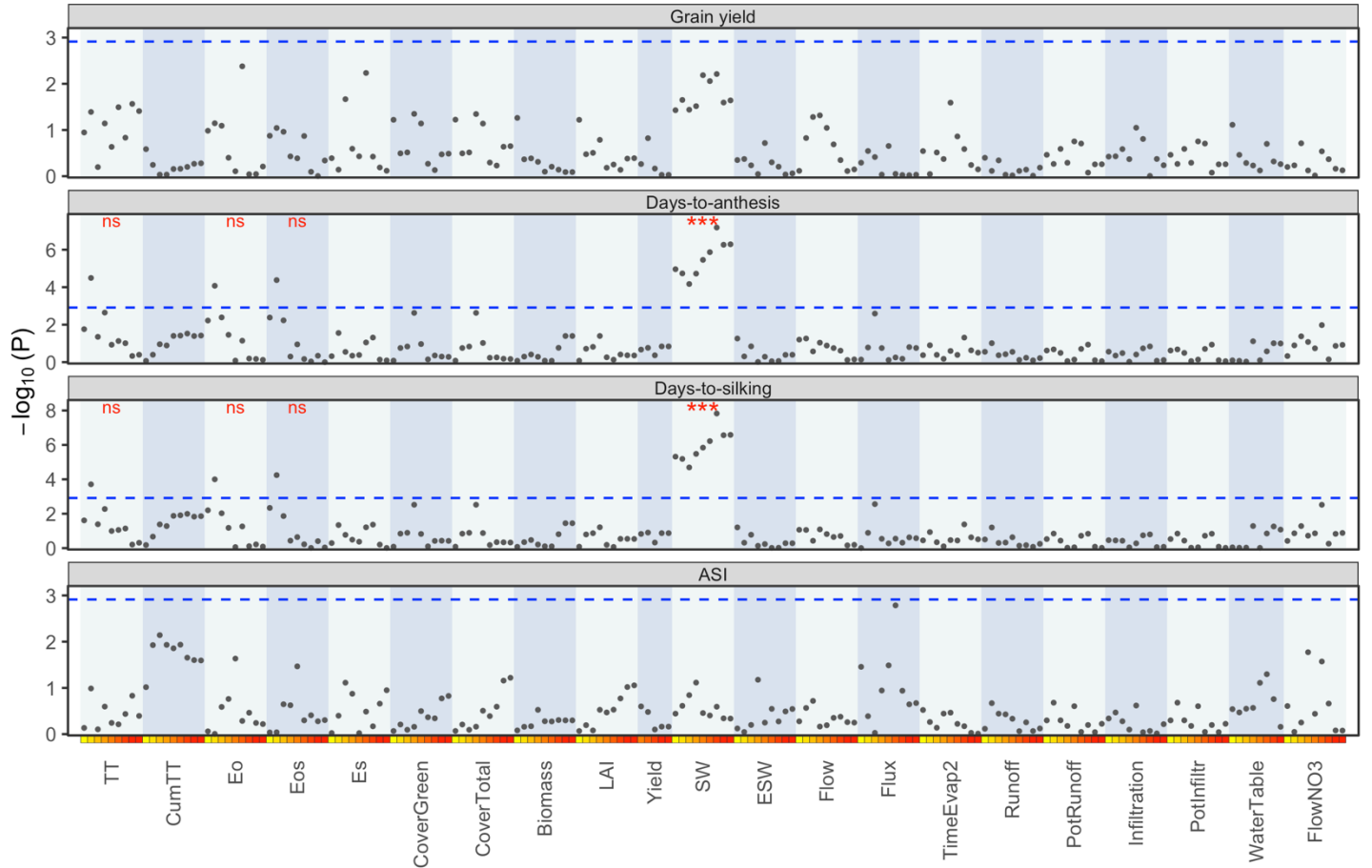


Figure S6: Association between individual environmental (co)variates with yield and flowering traits in the southern trials. Each point corresponds to an environmental covariate. The horizontal line gives the threshold for significance after Bonferroni adjustment. The red marks ('ns'= non-significant, '.', '*', '**', and '***', corresponding to P-values >0.1, and smaller than 0.1, 0.05, 0.01, and 0.001, respectively) indicate whether the EC group was enriched for significant associations in a Hypergeometric test. TT=Thermal time, CumTT=Cumulated thermal time, Eo=Potential evapotranspiration, Eos=Potential evaporation, Es=Realized evaporation, LAI=Leaf Area Index, SW=Soil Water, ESW=Extractable soil water, Flow=Unsaturated water movement between layers, Flux=Saturated water flux from each layer to layer below, TimeEvap2=Time since the start of second stage evaporation, PotRunoff=Potential runoff, PotInfiltr=Potential infiltration, and FlowNo3=Amount of Nitrogen leaching as NO₃ from each layer.

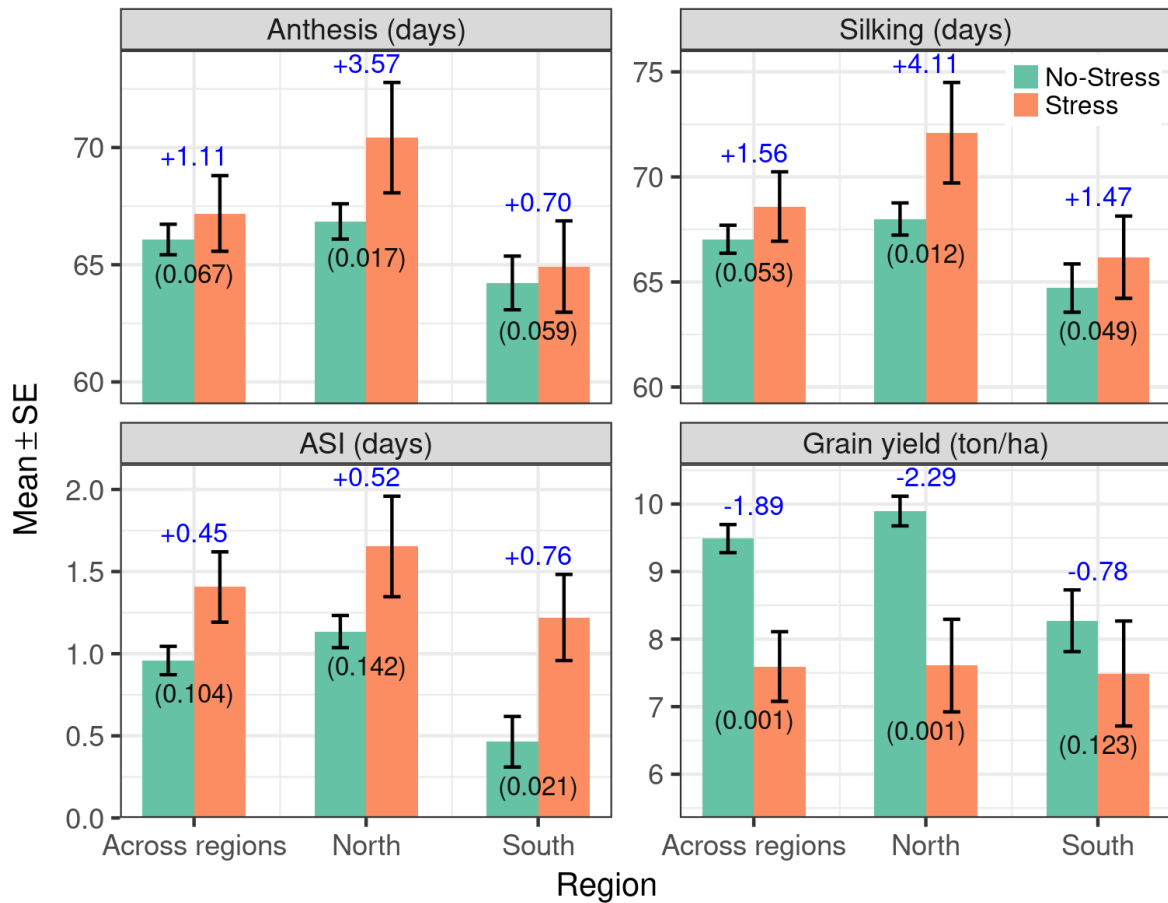


Figure S7: Average phenotype difference between year-locations with both drought (supply-demand-ratio, $SDR \leq 0.54$) and heat (number of days with temperature over 30°C , $HI30 > 36$) stress, and the rest of year-locations no meeting both stresses ($SDR \leq 0.54$ and $HI30 > 36$). See Figure 7 for separated analysis for SDR and $HI30$

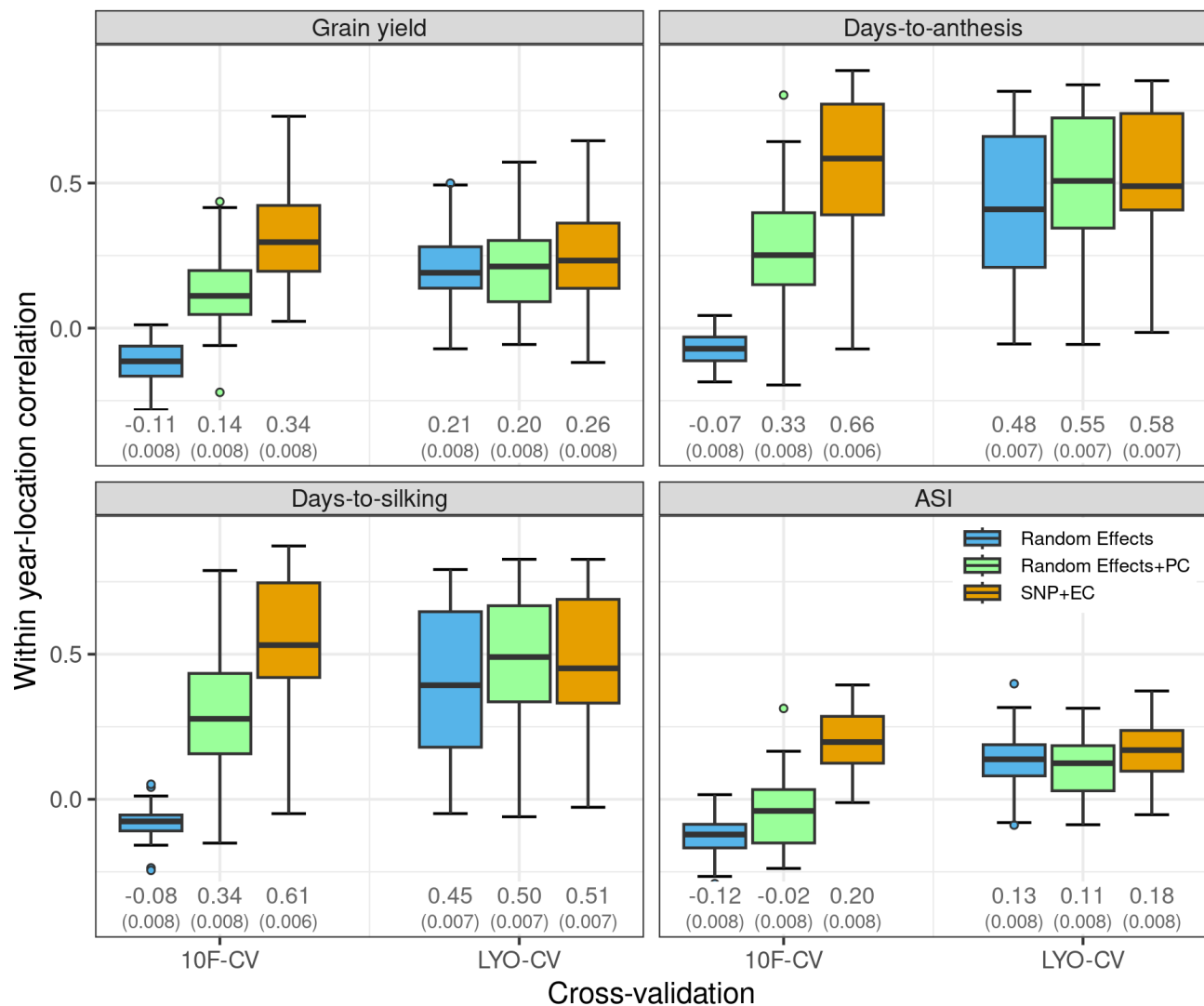


Figure S8. Within year-location prediction correlation for each model in the 10F-CV (hybrids assigned to one of 10 folds) and LYO-CV (leave-one-year-out) in southern locations (see Figure 8 for results for the northern locations). The 'Random Effects+PC' model (4) includes the top 5 SNPs-derived PCs plus the top 5 ECs-derived PCs. Numbers at the bottom are the pooled mean and standard error (in parenthesis).

2. Supplementary Tables

Table S1. Description of the environmental covariates simulated by APSIM

Num.	Variable	Units	Description
1	HI30	days	Number of days with maximum temperature over 30 Celsius degrees (°C)
2	CumHI30	days	Accumulated number of days with maximum temperature over 30°C since germination
3	TT	Celsius degrees	Thermal Time
4	CumTT	Celsius degrees	Accumulated thermal time since one day before germination
5	Eo	mm	Potential evapotranspiration of the whole soil-plant system
6	Eos	mm	Potential evaporation from soil surface
7	Es	mm	Actual (realized) soil water evaporation
8	ESW	mm	Extractable soil water from each layer relative to lower limit at 15Bar (LL15)
9	Flow	mm	Unsaturated water movement between layers
10	FlowNO ₃	kg/ha	Amount of Nitrogen leaching as NO ₃ from each soil layer
11	Flux	mm	Saturated water flux from each layer to the layer below
12	Infiltration	mm	Daily infiltration across the soil surface
13	PotInf	mm	Potential infiltration, this is set by microclimate and is rainfall less that intercepted by the canopy and residue component
14	PotRunoff	mm	Gets potential runoff
15	Runoff	mm	Daily runoff
16	SW	mm/mm	Amount of water in the soil in each layer
17	T	days	Time since start of second state evaporation
18	WaterTable	mm	Depth of the water table (10000 if no water table present)
19	CoverGreen	0-1	Total plant green cover from all organs
20	CoverTotal	0-1	Total plant cover from all organs
21	LAI	m ² /m ²	Leaf area index, live plant green
22	Biomass	kg/ha	Above ground biomass
23	Yield	kg/ha	Grain yield dry weight

Water variables ESW, Flow, FlowNO₃, Flux, and SW are provided by APSIM by each layer. They are aggregated across all layers to get a total value.

Table S2. Description of the environmental covariates simulated by APSIM

Period	Start	End	Length (days)*		
			Mean	Min	Max
GerEme	Germination	Emergence	7.72	5	14
EmeEnJ	Emergence	EndJuvenile	18.6	11	38
EnJFlo	EndJuvenile	FloralInitiation	1.18	1	2
FloFla	FloralInitiation	FlagLeaf	38.91	32	52
FlaFlw	FlagLeaf	Flowering	4.67	4	6
FlwStG	Flowering	StartGrainFill	8.63	6	12
StGEnG	StartGrainFill	EndGrainFill	34.31	24	73
EnGMat	EndGrainFill	Maturity	1.76	1	5
MatHar	Maturity	HarvestRipe	1.68	1	4

*Mean, minimum and maximum length of the periods were calculated across all 136 year-locations (2014-2021)

Table S3: Tester used in making the G2F hybrids

Tester group	N testers	Testers	N hybrids
North			
CG	11	C103,CG60,CG102,CG110,CG111,CGR01,CGR03,CG44,CG108,CG123,CG120	176
DK3IIH6	1	DK3IIH6	142
LH	5	LH162,LH145,LH38,LH51,LH210	50
LH82	1	LH82	199
LH198	1	LH198	207
LH185	1	LH185	223
LH195	1	LH195	757
PH	23	PHN82,PB80,PHG35,PHM49,PHR25,PHW03,PHHV4,PHRE1,PHG29,PHG47,PHG83,PHR55,PHW30,PHJ65,PHN11,PHN47,PHR03,PHW53,PHM57,PHR63,PHAJ0,PH207,PHW65	261
PHB	1	PHB47	255
PHK	3	PHK76,PHK05,PHK56	398
PHP	2	PHP02,PHP60	385
PHT	2	PHT69,PHTD5	572
PHZ	1	PHZ51	659
TX	5	TX205,TX6252,TX775,TX777,TX779	17
Other	7	H95,MO17,OH43,OH7B,S8324,Q381,NK787	41
South			
CG	5	C103,CG102,CGR03,CG108,CGR01	7
DK3IIH6	1	DK3IIH6	23
LH	4	LH162,LH51,LH210,LH38	25
LH82	1	LH82	76
LH198	1	LH198	10
LH185	1	LH185	32
LH195	1	LH195	695
PH	18	PHN82,PHG35,PHM49,PHJ65,PHM57,PHN11,PHN47,PHR03,PHR55,PHR63,PHW30,PHW53,PHHV4,PB80,PHR25,PHG47,PHAJ0,PH207	86
PHB	1	PHB47	56
PHK	1	PHK76	4
PHP	1	PHP60	5
PHT	2	PHT69,PHTD5	322
PHZ	1	PHZ51	522
TX	6	TX205,TX6252,TX775,TX777,TX779,TX714	19
Other	5	H95,MO17,OH43,NK787,Q381	25

Table S4: Analysis of Variance for Yield and Flowering traits in southern trials using two models.

	Yield		Anthesis		Silking		ASI	
Source	Random Effects	SNP+ EC	Random Effects	SNP+ EC	Random Effects	SNP+ EC	Random Effects	SNP+ EC
YEAR (Y)	0.197 (0.095)	--	0.199 (0.080)	--	0.215 (0.095)	--	0.114 (0.063)	--
LOC (L)	0.167 (0.096)	--	0.565 (0.111)	--	0.405 (0.113)	--	0.108 (0.057)	--
YL	0.508 (0.098)	--	0.330 (0.058)	--	0.338 (0.066)	--	0.319 (0.060)	--
(Total YL)	0.663 (0.008)	--	0.874 (0.005)	--	0.876 (0.005)	--	0.406 (0.008)	--
EC	--	0.712 (0.022)	--	0.878 (0.020)	--	0.895 (0.019)	--	0.412 (0.020)
Hybrid (G)	0.044 (0.003)	--	0.057 (0.001)	--	0.049 (0.001)	--	0.032 (0.003)	--
SNP	--	0.033 (0.006)	--	0.057 (0.009)	--	0.061 (0.008)	--	0.035 (0.005)
GxL	0.022 (0.002)	--	0.008 (0.001)	--	0.008 (0.001)	--	0.037 (0.006)	--
SNPxEC	--	0.065 (0.008)	--	0.035 (0.004)	--	0.032 (0.004)	--	0.057 (0.009)
Error	0.270 (0.004)	0.263 (0.004)	0.053 (0.001)	0.049 (0.001)	0.063 (0.001)	0.057 (0.001)	0.527 (0.008)	0.535 (0.007)

Y: Year, L: Location, YL: Year-location, G: Hybrid, EC: Environmental Covariates. All traits were standardized to a unit variance.

3. Supplementary Files

File S1: Curated data, including SNP-genotypes, phenotypes, and environmental covariates ([download link](#)). **Folder structure:**

```
data
├── ECOV.csv
├── ECOV_KL.csv
├── ECOV_layered.csv
├── ECOV_period.csv
├── GENO.csv
├── MAP.csv
└── PHENO.csv
```

File S2: Workflow used to curate the data and to generate the environmental covariates. The compressed file contains folders each corresponding to a module in the workflow. The number preceding the name of the folder indicates the relative order of the module in the workflow ([download link](#)). **Folder content:**

```
├── 1_phenotypes
│   ├── code
│   │   └── 1_get_phenotype.R
│   └── output
├── 2_genotypes
│   ├── code
│   │   ├── 1_SNP_filter_codify.sub
│   │   └── 2_LD_prune_genotypes.R
│   └── output
├── 3_year_loc_summary
│   ├── code
│   │   └── 1_year_location_summary.R
│   └── output
├── 4_weather
│   ├── code
│   │   └── 1_get_met_apsim.R
│   └── output
├── 5_APSIM
│   ├── code
│   │   ├── 1_apsim_sim.R
│   │   └── 1_apsim_sim.sub
│   └── output
├── 6_ecov
│   ├── code
│   │   └── 1_get_env_cov.R
│   └── output
├── 7_GDD
│   └── code
│       └── 1_add_GDD_to_pheno.R
├── README
├── source
│   ├── Agronomic_information
│   └── g2f_2014_field_characteristics.csv
```

```

├── g2f_2015_agronomic_information.csv
├── g2f_2016_agronomic_information.csv
├── g2f_2017_agronomic\ information.csv
├── g2f_2018_agronomic\ information.csv
├── g2f_2019_agronomic_information.csv
├── g2f_2020_agronomic_information.csv
├── g2f_2021_agronomic_information.csv
├── Genotype
│   └── Hybrids_G2F_Genotype_Data_All_Years.vcf.zip
├── Metadata
│   ├── g2f_2014_field_characteristics.csv
│   ├── g2f_2015_field_metadata.csv
│   ├── g2f_2016_field_metadata.csv
│   ├── g2f_2017_field_metadata.csv
│   ├── g2f_2018_field_metadata.csv
│   ├── g2f_2019_field_metadata.csv
│   ├── g2f_2020_field_metadata.csv
│   ├── g2f_2021_field_metadata.csv
│   └── g2f_2022_field_metadata.csv
├── Phenotype
│   ├── g2f_2014_hybrid_data_clean.csv
│   ├── g2f_2015_hybrid_data_clean.csv
│   ├── g2f_2016_hybrid_data_clean.csv
│   ├── g2f_2017_hybrid_data_clean.csv
│   ├── g2f_2018_hybrid_data_clean.csv
│   ├── g2f_2019_phenotypic_clean_data.csv
│   ├── g2f_2020_phenotypic_clean_data.csv
│   └── g2f_2021_phenotypic_clean_data.csv
├── tools
│   ├── APSIM_functions.R
│   ├── ecov_utils.R
│   ├── Functions.R
│   ├── LD_prune.R
│   ├── read_metadata.R
│   └── read_phenotype.R

```

File S3: Scripts used for data analysis ([download link](#)). **Folder content:**

```

pipeline_analysis
├── 1_PC_genotypes
│   ├── code
│   │   ├── 1_get_G_matrix_PC.R
│   │   └── 2_make_Figure_1.R
│   └── output
├── 2_PC_ecov
│   ├── code
│   │   ├── 1_get_WWt_matrix_PC.R
│   │   └── 2_make_Figure_2.R
│   └── output
├── 3_ecov_varcomps
│   ├── code
│   │   ├── 1_ecov_variance_components.R
│   │   └── 2_make_Figure_3.R
│   └── output
├── 4_ecov_WAS
│   ├── code
│   │   ├── 1_ecov_WAS_SMR.R
│   │   └── 2_make_Figures_4_&_S6.R
│   └── output

```

```
|— 5_drought_heat_stress
|   |— code
|   |   |— 1_make_Figure_5.R
|   |   |— 2_make_Figure_6.R
|   |   |— 3_HI30_SDR_mixed_model.R
|   |   |— 4_make_Figures_7_&_S7.R
|   |— output
|— 6_genomic_prediction
|   |— code
|   |   |— 1_get_model_components.R
|   |   |— 2_ANOVA_mixed_model.R
|   |   |— 3_make_Tables_3_&_S4.R
|   |   |— 4_10F_CV_model.R
|   |   |— 5_LYO_CV_model.R
|   |   |— 6_make_Figures_8_&_S8.R
|   |— output
|— README
|— tools
|   |— c_utils.c
|   |— c_utils.o
|   |— c_utils.so
|   |— ecov_utils.R
|   |— Functions.R
|   |— get_folds.R
|   |— get_variance.R
|   |— TENSOR_EVD.R
```