LETTERS
https://doi.org/10.1038/s41588-019-0414-y

nature genetics

Genomic prediction of maize yield across European environmental conditions

Emilie J. Millet[1,2,4], Willem Kruijer[1], Aude Coupel-Ledru[2,5], Santiago Alvarez Prado[2,6], Llorenç Cabrera-Bosquet[2], Sébastien Lacube[2], Alain Charcosset[3], Claude Welcker[2], Fred van Eeuwijk[1] and François Tardieu[2*]

The development of germplasm adapted to changing climate is required to ensure food security[1,2]. Genomic prediction is a powerful tool to evaluate many genotypes but performs poorly in contrasting environmental scenarios[3–7] (genotype × environment interaction), in spite of promising results for flowering time[8]. New avenues are opened by the development of sensor networks for environmental characterization in thousands of fields[9,10]. We present a new strategy for germplasm evaluation under genotype × environment interaction. Yield was dissected in grain weight and number and genotype × environment interaction in these components was modeled as genotypic sensitivity to environmental drivers. Environments were characterized using genotype-specific indices computed from sensor data in each field and the progression of phenology calibrated for each genotype on a phenotyping platform. A whole-genome regression approach for the genotypic sensitivities led to accurate prediction of yield under genotype × environment interaction in a wide range of environmental scenarios, outperforming a benchmark approach.

Grain yield is the outcome of many physiological processes responding to environmental conditions that rapidly fluctuate throughout the growing season. As a consequence, genes and quantitative trail loci underlying yield and its constituent processes vary between experiments (genotype × environment interaction, G × E)[11–13] because the corresponding alleles have positive, negative or null effects depending on environmental scenarios[14]. Evolution has constrained the interactions between physiological processes such that integrated component traits like grain number respond in a relatively simple way to changes in environmental conditions[14]. The problem of G × E for yield can therefore be addressed via the genetic factors underlying the response of its component traits to environmental conditions. The sensitivity of yield mainly varies with the timing of stressing periods in relation to phenological events[14,15]. It is now possible to predict flowering time of many accessions on the basis of marker profiles and environmental indices[8] and the dates of phenological events can be modeled for a range of maize genotypes throughout Europe in current and future climatic conditions[16]. Here, we approached the problem of genomic prediction for yield by first identifying the environmental drivers for G × E of yield components (grain number and weight) and then predicting, from marker profiles, the corresponding genotypic sensitivities to measured environmental conditions (Supplementary Fig. 1). Prediction of plant phenology using data from a phenotyping platform was essential in the construction of genotype-specific environmental characterizations. Our approach out-performed a state-of-the-art genomic prediction method using an environmental relatedness matrix alongside a genetic kinship matrix[3].

Grain yield and related traits were measured for a diversity panel of 246 dent maize hybrids in 25 field experiments under rainfed and irrigated conditions along a climatic transect from western to eastern Europe and an experiment in Chile (Fig. 1a). Grain yield varied strongly between experiments (for example 1.2–12.9 t ha^{-1} for the reference hybrid) and hybrids (for example 2.1–7.1 t ha^{-1} in one experiment). It was closely related to grain number across experiments (for example $r = 0.96$ for the reference hybrid[12]).

We first modeled the progression of leaf phenological stages of individual hybrids in each field (Fig. 1a) because a direct evaluation of the timing of reproductive events would be infeasible. This was on the basis of measurements in a robotized phenotyping platform[17]. The progression of the number of appeared leaves closely followed thermal time (corrected for the effect of meristem temperature) and, for the reference hybrid, was consistent with field observations (Fig. 1b). The rate of progression presented an appreciable genotypic variability (Fig. 1c,d). For each site, the dates of leaf stages of all hybrids were then calculated on the basis of thermal times and temperature records.

The progression of leaf stages allowed estimation, for all hybrids, of the dates of four phenological events that defined three phenological phases (Fig. 2; Supplementary Fig. 2). (1) A first phase, called 'vegetative' hereafter, elapsed from apical floral transition to silk initiation on ovaries[18]. These two events were not directly measured but strictly correspond to leaf stages[19] as observed in the phenotyping platform. (2) The flowering phase elapsed from silk initiation to the date at which grains cannot abort[20]. (3) The grain-filling phase elapsed from the end of abortion to physiological maturity[21]. Hybrids differed in their timing of development and therefore experienced different environmental conditions in a same field during each of the phenological phases (Fig. 2a–c). These differences in conditions experienced generated wide ranges of genotype-specific intercepted radiation (R_{int}), soil water potential (Ψ_{soil}) and meristem temperature.

We then examined the relationship between yield components (grain number and weight) and environmental conditions (Fig. 3). G × E was observed for grain number depending on the amount of radiation intercepted by plants during the vegetative phase (R_{int}),

[1]Biometris, WUR, Wageningen, the Netherlands. [2]LEPSE, INRA, Université Montpellier, SupAgro, Montpellier, France. [3]GQE-Le Moulon, INRA, Université Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France. [4]Present address: Biometris, WUR, Wageningen, the Netherlands. [5]Present address: University of Bristol, School of Biological Sciences, Bristol, UK. [6]Present address: IFEVA and CONICET, Buenos Aires, Argentina. *e-mail: francois.tardieu@inra.fr

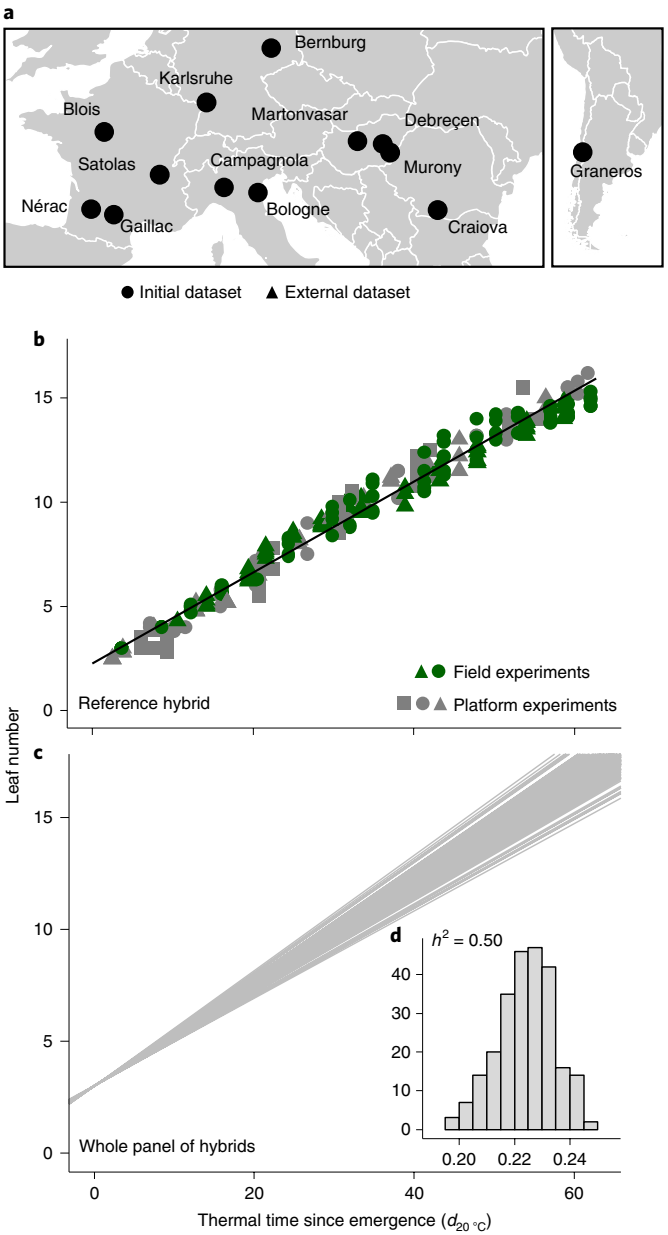

Fig. 1 | Progression of leaf phenological stages in a phenotyping platform and in two field experiments. a, Map of the experimental sites (drawn with the map_data() function of the R package ggplot2). **b**, Progression of the number of appeared leaves as a function of time corrected for temperature (thermal time) for the reference hybrid in three platform experiments and two field experiments. **c**, Progression of the number of appeared leaves for 246 hybrids in the phenotyping platform. Each line corresponds to one genotype. **d**, Histogram of leaf appearance rate (leaf $d_{20\,°C}^{-1}$) across the panel and narrow-sense heritability.

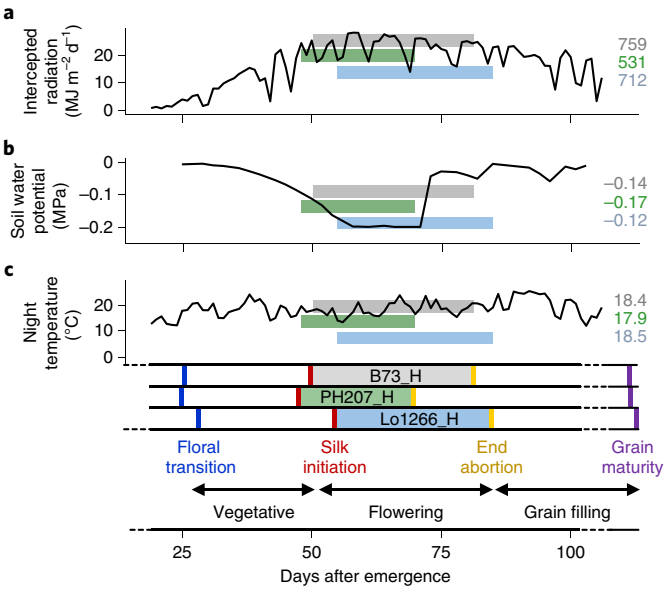

Fig. 2 | The calculation of environmental indices over phenological phases for three hybrids in one experiment. The plant cycle was split into three phases separated by floral transition (blue vertical bar), beginning of silk initiation (dark red bar), end of abortion (yellow bar) and grain maturity (purple bar). These phenological phases occurred at different thermal times for each hybrid in a given field as shown for the flowering phase of three hybrids: B73_H (gray shade), PH207_H (green shade) and Lo1266_H (blue shade). **a**–**c**, Time courses of environmental variables in experiment Gai12R and average or cumulative environmental indices calculated over the flowering phase for the three hybrids (same color codes), for example −0.14, −0.17 and −0.12 MPa for mean soil water potential.

average night temperature over the flowering phase (T_{night}) as in wheat and rice[22,23] and soil water potential during the flowering phase (Ψ_{soil}). For example, an increase in T_{night} by 3 °C decreased grain number by 15 and 50% for the least and most sensitive hybrids. The mean weight of individual grains was not dependent on the environmental conditions during the post-flowering phase, consistent with earlier studies reporting the robustness of grain filling, except in extreme conditions[24,25]. It was therefore considered to be only genotype-dependent and estimated using a model without G×E terms (Methods).

We modeled grain yield of the *i*th hybrid in the *j*th experiment, $\widehat{GY_{i,j}}$, as the product of estimated individual grain weight ($\widehat{GW_i}$) and grain number ($\widehat{GN_{i,j}}$):

$$\widehat{GY_{i,j}} = \widehat{GN_{i,j}} \times \widehat{GW_i} \quad (1)$$

$\widehat{GN_{ij}}$ was further dissected via a factorial regression:

$$GN_{i,j} = \mu + e_j + g_i + \beta_{1,i}\, T_{night_{i,j}} + \beta_{2,i}\, R_{int_{i,j}} + \beta_{3,i}\, \Psi_{soil_{i,j}} + \varepsilon_{i,j} \quad (2)$$

where, μ is an intercept, e_j and g_i are environmental and genotypic main effects and ε_{ij} the residual; $\beta_{1,i}$, $\beta_{2,i}$ and $\beta_{3,i}$ are the genotype-dependent sensitivities to environmental indices, $T_{night_{i,j}}$, $R_{int_{i,j}}$, and $\psi_{soil_{i,j}}$ sensed by the *i*th hybrid in the *j*th experiment (Fig. 2). Sensitivities showed large genotypic variability and high narrow-sense heritabilities, with $h^2 = 0.51 - 0.78$ (Fig. 3).

Grain yields predicted from equations (1) and (2) (Fig. 4) showed clear correlations with observed genotypic means from single site analyses: $r = 0.60 - 0.91$ across hybrids in each experiment (denoted 'per experiment' hereafter) and $r = 0.71 - 0.97$ for a given hybrid across experiments (denoted 'per hybrid' hereafter) (Fig. 4; Supplementary Tables 1 and 2).

We then used equations (1) and (2) to predict grain yield for non-observed environments and/or hybrids (Fig. 4). Environmental main effects (e_j) were predicted using the same environmental indices that described the G×E, calculated from sensor data

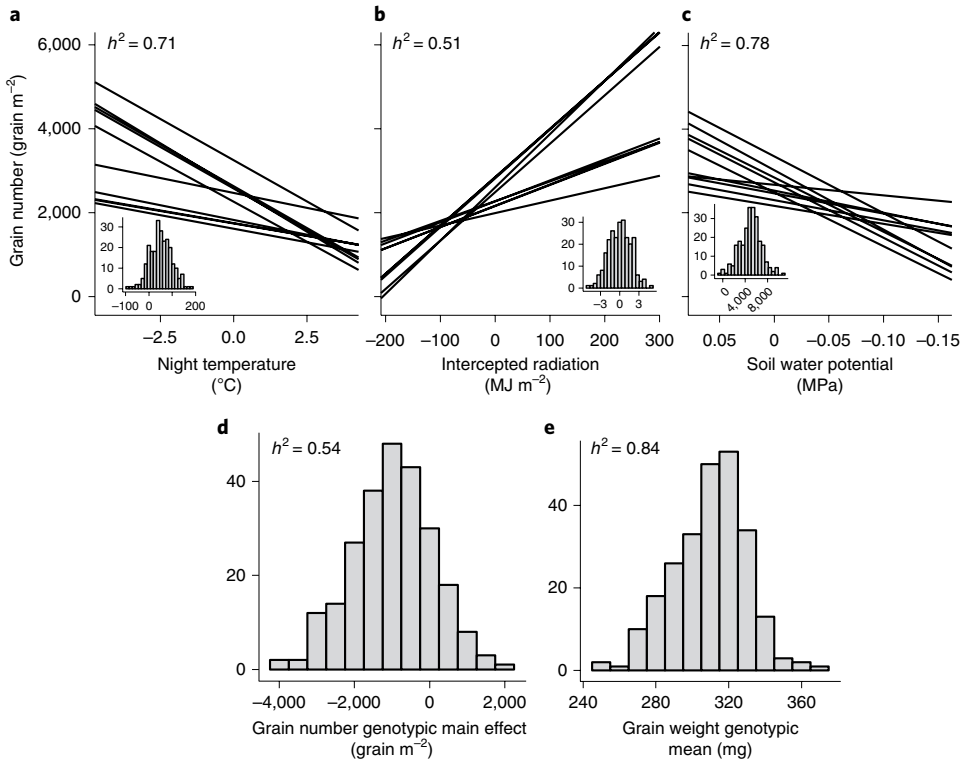

Fig. 3 | Variability of genotype-specific response curves of grain number to environmental indices and of genotypic means of grain number and individual grain weight. a, Grain number as a function of night temperature averaged during the flowering phase. **b**, Grain number as a function of cumulative intercepted light calculated over the vegetative phase. **c**, Grain number as a function of soil water potential averaged during the flowering phase. For better legibility in **a** to **c**, environmental indices are centered and only five hybrids with maximum and minimum responses are shown, one line per hybrid. The variability of sensitivities from the factorial regression model (equation (2)) across the whole panel is shown in histograms. **d**, Variability of the genotypic main effect for grain number. **e**, Variability of the genotypic mean for grain weight.

(Methods). Genotype-specific parameters (GW_i for grain weight and g_i, $\beta_{1,i}$, $\beta_{2,i}$ and $\beta_{3,i}$ for grain number) were predicted from marker profile by whole-genome regression (Supplementary Fig. 3a–e). Genotype-specific environmental indices were calculated as functions of sensor data and genomic predictions for the progression of leaf stages and for flowering time (Supplementary Fig. 3f–h).

Prediction accuracy was assessed in a cross-validation scheme under stratified random sampling of hybrids and experiments. We drew ten times a random training set of 200 hybrids (G) and 20 experiments (E) from the initial 246×25 dataset, keeping the remaining 46 hybrids (nG) and five experiments (nE) as test sets, with a stratification strategy for respecting the proportions of genetic groups and of environmental scenarios (Table 1 and Methods). Finally, we evaluated the prediction performance of our method with an external validation using independent datasets from external environments and/or external hybrids (Table 1; Supplementary Fig. 1; Methods). In addition, we benchmarked our method by comparing it with a genomic prediction method in which G×E is modeled using an environmental kinship[3] (Methods).

- For the prediction of already tested hybrids in new experiments $(G,nE)_{cross}$ of the cross-validation, correlations between observed and predicted grain yields were $r=0.43–0.85$ per experiment and $r=0.71–0.97$ per hybrid (Fig. 4; Supplementary Tables 1 and 2).
- For new hybrids in already tested experiments $(nG,E)_{cross}$, correlations were $r=0.21–0.71$ per experiment and $r=0.66–0.96$ per hybrid (Fig. 4; Supplementary Tables 1 and 2). In contrast, for this same dataset, predictions obtained from the benchmark method using the environmental kinship produced poor yield predictions, with $r=-0.12–0.44$ per experiment (Supplementary Table 3).
- For new hybrids in new experiments $(nG,nE)_{cross}$, our approach produced medium to good predictions: $r=0.20–0.74$ per experiment.
- For external experiments with already tested hybrids, $(G,nE)_{ext1}$, we obtained medium to high correlations, from $r=0.38–0.80$ per experiment and $r=0.55–0.83$ per hybrid (Fig. 4; Supplementary Tables 1 and 2). For external new hybrids in a second set of external experiments, $(nG,nE)_{ext2}$, correlations between predicted and observed yields ranged from $r=0.32–0.39$ per experiment, although with an appreciable bias (Fig. 4; Supplementary Tables 1 and 2).

Overall, our approach predicted yield for every combination of hybrid and experiment from marker profiles, high-resolution sensor data and platform-based predictions of phenology. For that, we dissected yield into its components, grain number and weight, then modeled G×E of grain number in a factorial regression model with genotype-specific sensitivities to genotype-specific environmental indices and finally predicted the genotypic main effect and sensitivities from marker profiles (Supplementary Fig. 1). This improved the prediction accuracy in comparison to a benchmark multi-environment genomic prediction in which G×E is modeled using an environmental kinship[3] (55% increase, Supplementary Table 3). Because our field experiments covered a wider range of environmental scenarios than related studies[3,5,7,26] and our approach produced reliable yield predictions on independent external datasets, our method presents a broader applicability than competing statistical genetic approaches[3,5]. It is noteworthy that our analysis was performed

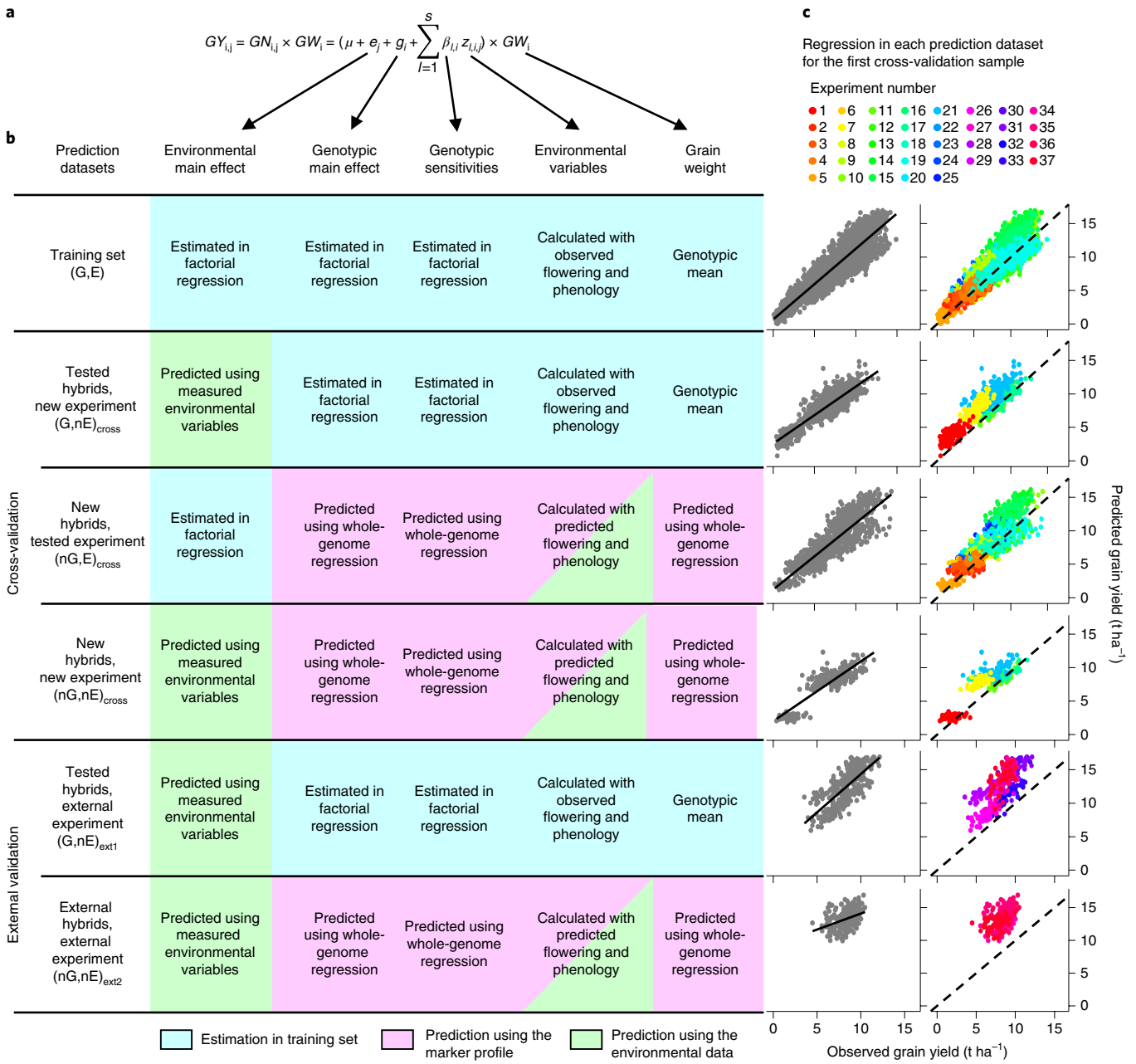

Fig. 4 | Yield prediction: method and results for each dataset. a, Equation for the dissection of grain yield ($GY_{i,j}$). $GN_{i,j}$ grain number of hybrid *i* in experiment *j*, GW_i, mean grain weight of hybrid *i*. For $GN_{i,j}$, μ, intercept; e_j, environmental main effect in field *j*; g_i, genotypic main effect for hybrid *i* ;$\beta_{l,i}$, sensitivity of hybrid *i* to the genotype-specific environmental index $z_{l,i,j}$ (intercepted light, night temperature or soil water potential). **b**, Model parameters, either estimated or predicted depending on whether hybrids and experiments were used as training (G and E, respectively) or test set (nG and nE, respectively). For already tested G and E (light blue cells), parameters were estimated from measured data via the factorial regression for grain number (equation (2)) or averaged across experiments for grain weight (Methods). For nE, the environmental main effects were predicted using measured environmental data (green cells). For nG, genotypic parameters were predicted using a genomic prediction model (magenta cells). **c**, Relationships between predicted and observed grain yield in each dataset for the first cross-validation sample, either cross-validation or external validation sets. In the first column, the overall correlation is presented with a linear regression across hybrids and experiments (black line). The second column shows the same correlations, in which each experiment is represented with a color. Dashed black lines indicate the first bisector. Experimental fields numbered as in Supplementary Table 1.

in a panel with a reduced flowering window, so predictions were not driven by the large effect on yield of flowering time through cumulative photosynthesis and stress avoidance[16,27]. Similarly, the genetic structure of the panel had a limited impact on the prediction (Supplementary Table 4, Supplementary Fig. 4 and Supplementary Note 1). Our approach offers good prospects for breeding programmes aiming at the development of new hybrids adapted to the diversity of current and future climatic conditions[10,12].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0414-y.

Received: 30 October 2018; Accepted: 8 April 2019;
Published online: 20 May 2019

Table 1 | Summary table of the different datasets used for training, test and validation of the prediction method

	Dataset		Number of hybrids					Number of experiments					Range of environmental indices		
	Type	Name	Total	Iodent	Lancaster	Stiff Stalk	Diverse dents	Total	WW-Cool	WW-Hot(day)	WW-Hot	WD	T_{night}	R_{int}	Ψ_{soil}
Initial		-	246	39	45	55	107	25	6	4	9	6	15.0;23.2	39;499	−0.228;−0.009
Training		(G,E)	200	32	36	44	88	20	5	3	7	5	15.2;23.1	39;482	−0.225;−0.010
Cross-validation	Test	$(G,nE)_{cross}$	200	32	36	44	88	5	1	1	2	1	16.1;22.0	61;445	−0.225;−0.010
	Test	$(nG,E)_{cross}$	46	7	9	11	19	20	5	3	7	5	15.2;23.1	44;445	−0.225;−0.010
	Test	$(nG,nE)_{cross}$	46	7	9	11	19	5	1	1	2	1	16.3;21.9	65;414	−0.225;−0.010
External validation		$(G,nE)_{ext1}$	32	4	4	5	19	12	8	2	2	0	14.6;20.9	150;493	−0.125;−0.015
		$(nG,nE)_{ext2}$	56	-	-	-	-	4	4	0	0	0	16.4;18.2	267;460	−0.044;−0.021

The four genetic groups were identified in another study and are described in Methods. The environmental scenarios are the same as in previous work[12]. They combine soil water status (WW or WD) and temperature (Cool, Hot(day), Hot) measured during the flowering phase. Environmental indices are those used to model the G×E in grain number (equation (2)).

References

1. Tester, M. & Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **327**, 818–822 (2010).
2. IPCC. *Climate Change 2014: Synthesis Report* (eds Core Writing Team, Pachauri R. K. & Meyer L. A.) (IPCC, 2014).
3. Jarquín, D. et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* **127**, 595–607 (2014).
4. Cooper, M., Technow, F., Messina, C., Gho, C. & Totir, L. R. Use of crop growth models with whole-genome prediction: application to a maize multienvironment trial. *Crop Sci.* **56**, 2141–2156 (2016).
5. Burgueño, J., de los Campos, G., Weigel, K. & Crossa, J. Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. *Crop Sci.* **52**, 707–719 (2012).
6. Ly, D. et al. Whole-genome prediction of reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.) by genomic random regression. *Field Crops Res.* **216**, 32–41 (2018).
7. Roorkiwal, M. et al. Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype x environment interaction on prediction accuracy in chickpea. *Sci. Rep.* **8**, 11701 (2018).
8. Li, X., Guo, T., Mu, Q., Li, X. & Yu, J. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. *Proc. Natl Acad. Sci. USA* **115**, 6679–6684 (2018).
9. Chenu, K., Deihimfard, R. & Chapman, S. C. Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt spatial and temporal trends. *New Phytol.* **198**, 801–820 (2013).
10. Harrison, M. T., Tardieu, F., Dong, Z., Messina, C. D. & Hammer, G. L. Characterizing drought stress and trait influence on maize yield under current and future conditions. *Glob. Change Biol.* **20**, 867–878 (2014).
11. Ribaut, J.-M., Hoisington, D. A., Deutsch, J. A., Jiang, C. & Gonzalez-de-Leon, D. Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. *Theor. Appl. Genet.* **92**, 905–914 (1996).
12. Millet, E. J. et al. Genome-wide analysis of yield in europe: allelic effects vary with drought and heat scenarios. *Plant Physiol.* **172**, 749–764 (2016).
13. Bonneau, J. et al. Multi-environment analysis and improved mapping of a yield-related QTL on chromosome 3B of wheat. *Theor. Appl. Genet.* **126**, 747–761 (2012).
14. Tardieu, F., Simonneau, T. & Muller, B. The physiological basis of drought tolerance in crop plants: a scenario-dependent probabilistic approach. *Annu. Rev. Plant Biol.* **69**, 733–759 (2018).
15. Parent, B. et al. Quantifying wheat sensitivities to environmental constraints to dissect genotype x environment interactions in the field. *Plant Physiol.* **174**, 1669–1682 (2017).
16. Parent B. et al. Maize yields over Europe may increase in spite of climate change, with an appropriate use of the genetic variability of flowering time. *Proc. Natl Acad. Sci. USA* **115**, 10642–10647 (2018).
17. Cabrera-Bosquet, L. et al. High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytol.* **212**, 269–281 (2016).
18. Fuad-Hassan, A., Tardieu, F. & Turc, O. Drought-induced changes in anthesis-silking interval are related to silk expansion: a spatio-temporal growth analysis in maize plants subjected to soil water deficit. *Plant Cell Environ.* **31**, 1349–1360 (2008).
19. Girardin, P. *Ecophysiologie du Maïs: Fonctionnement de la Plante et de la Culture* (Association Générale des Producteurs de Maïs, 2000).
20. Andrade, F. H. et al. Kernel number determination in maize. *Crop Sci.* **39**, 453–459 (1999).
21. Borrás, L. & Westgate, M. E. Predicting maize kernel sink capacity early in development. *Field Crops Res.* **95**, 223–233 (2006).
22. Hatfield, J. L. et al. Climate impacts on agriculture: implications for crop production. *Agron. J.* **103**, 351–370 (2011).
23. Welch, J. R. et al. Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures. *Proc. Natl Acad. Sci. USA* **107**, 14562–14567 (2010).
24. Fisher, R. A. Yield potential of dwarf spring wheat and the effect of shading. *Crop Sci.* **15**, 607–613 (1975).
25. Slafer, G. A. & Savin, R. Source–sink relationships and grain mass at different positions within the spike in wheat. *Field Crops Res.* **27**, 85–89 (1994).
26. Messina, C. D. et al. Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). *Eur. J. Agron.* **100**, 151–162 (2018).
27. Rio, S., Mary-Huard, T., Moreau, L. & Charcosset, A. Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. *Theor. Appl. Genet.* **132**, 81–96 (2018).

Acknowledgements

We are grateful to A. Grau, B. Suard, C. Check, P. Sidawy and T. Laisné for technical assistance in the Phenoarch experiments, as well as to key persons from the 2014 field experiments at Arvalis, Euralis and KWS. We are also grateful to S. Nicolas and S. Negro for the genotyping (imputation and quality check) and to R. Rincent for advice on the use of the BGLR package. This work was supported by the EU project FP7-244374 (DROPS), the Agence Nationale de la Recherche projects ANR-10-BTBR-01 (Amaizing) and ANR-11-INBS-0012 (Phenome), the Netherlands Scientific Organisation for Research NWO-STW project 11145 Learning from Nature and the EU project H2020 731013 (EPPN^{2020}).

Authors contributions

E.J.M., F.T., C.W. and F.v.E. designed the research and analyzed the field experiments. E.J.M., F.T and F.v.E. wrote the paper with the contributions of W.K., S.A.P., L.C.B., A.C.L., C.W., S.L., and A.C. F.v.E., E.J.M. and W.K. performed the genomic prediction. L.C.B. and S.A.P. performed and analyzed the platform experiments.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0414-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to F.T.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Genetic material. Data of field experiments and panel of hybrids were previously described[12]. The panel of hybrids results from the cross of a common flint parent (UH007) with 246 dent lines that maximized the diversity in the dent group while keeping a restricted flowering window (7 d; ref. [28]). It was genotyped with a 50 K Infinium HD Illumina array[29]. Four genetic groups were identified using the ADMIXTURE software[28], namely Non Stiff Stalk (Iodent, 39 lines), Lancaster (45 lines), Stiff Stalk (55 lines) and diverse-dent lines (107 lines) that do not belong to the former three heterotic groups (Table 1). This structure appeared consistent with the history of dent maize breeding programs and the differentiation of heterotic groups. The IBS (identity-by-state) varied from 0.58 to 0.95 (average 0.66)[28]. A principal coordinate analyses (PCoA) was performed on the IBD (identity-by-descent) and confirmed the grouping with a clear separation between the three main groups (Iodent, Lancaster and Stiff Stalk) whereas the diverse-dent lines did not present a clear structure.

The hybrid B73 × UH007 was considered as the reference hybrid because of its median precocity and performance and because the B73 genome is one of the reference genomes in maize[30].

An external panel of 56 maize hybrids was used for external validation of our methods. Briefly, they were generated by crossing 56 additional dent lines from European breeding programs and recent US Corn Belt with the tester UH007. These hybrids are a subset of a larger panel presented in detail in another study[27].

Field experiments. The panel of hybrids was evaluated in 25 experiments (combinations of location × year × water treatment) carried out in 2012 or 2013, at ten sites, either rainfed or irrigated. Sites were distributed on a west–east transect across Europe in France, southern Germany, northern Italy, Hungary and Romania at latitudes from 44° to 49° N compatible with the adaptation region of the panel. One experiment was performed in Chile (34° S). The map of experimental sites was drawn with the map_data() function of the R package ggplot2 (ref. [31]; Fig. 1). Experiments were designed as alpha-lattice designs with two and three replicates for watered or rainfed treatments, respectively. An environmental classification was performed on the basis of soil water potential and temperature scenarios[10,12].

Two independent external datasets were considered to test the prediction model for grain yield. The first one, $(G,nE)_{ext1}$ involved a subset of 32 already tested hybrids in 12 external experiments distributed across Europe, in France, Germany, northern Italy and Hungary, at latitudes from 44° to 52° N (Supplementary Table 5). The second independent dataset, $(nG,nE)_{ext2}$ involved the 56 external maize hybrids, in four external experiments carried out in 2014 in France and Germany at latitudes from 46° to 52° N (Supplementary Table 5).

Grain yield, yield components and dates of anthesis or silking were measured in each field for the 246 hybrids. Grain yield was recorded simultaneously with grain water content and adjusted to 0.15 g g^{-1} dry matter. Individual grain weight was measured in two samples of 200 grains and grain number was calculated by dividing yield by individual grain weight. The number of appeared leaves for the reference hybrid was scored twice a week in two field locations, Ner12 and Ner13 (ref. [12]) and three times during the vegetative phase in the other experiments (Supplementary Fig. 2).

Analyses were performed using the R software[32], v.3.4.3. For each experiment, genotypic means were calculated using a mixed model on the basis of fixed hybrid and replicate effects, random spatial effects and spatially correlated errors (see previous work[12] for details). The best linear unbiased estimators (BLUEs) were used in subsequent analyses.

Meristem temperature was calculated every hour for each field experiment, on the basis of a simplified energy balance in which, in comparison with T_{air}, $T_{meristem}$ increases with light (import of energy) and decreases with VPD_{air} (loss of energy due to latent heat associated with transpiration)[12,33]. The soil water potential (Ψ_{soil}) used here was the mean value measured with tensiometers in two or three plots sown with the reference hybrid[12].

The amount of light intercepted by the reference hybrid during a phenological phase (R_{int}) was obtained in each experiment by simulations using the crop model APSIM[34] with climatic data for the experiment (Supplementary Fig. 5). Briefly, APSIM was run at a 3-h time step in well-watered conditions (irrigation whenever soil water reserve decreased below 100% of its maximum value). The vector of APSIM model parameters for simulation was calibrated for the reference hybrid using measured leaf area in one experiment (Mauguio, France, 2011, six dates through the growing season). Notably, intercepted light as calculated for a given site is that of the reference hybrid (not for each individual hybrid) so it can be considered as an environmental variable. Hence, genotypic differences in coefficients of the regression between R_{int} and a yield component are not only due to genetic differences in photosynthesis rate but also, potentially, to differences in light interception between hybrids.

Platform experiments. The panel was evaluated in three experiments at the Phenoarch platform[17,35] (https://www6.montpellier.inra.fr/lepse/M3P) in Spring (2012 and 2013) and Winter (2013), thereby generating a large range of incident light (3–7 MJ m^{-2} d^{-1}), evaporative demand (0.5–3.5 kPa) and temperature (18–30 °C). Meristem temperature was recorded every 15 min with thermocouples placed at immediate proximity of the meristems of eight plants of the reference hybrid, together with other environmental conditions that are not referred to here. Three plants per hybrid were grown under well-watered conditions in every experiment. The numbers of visible and ligulated leaves of every plant were scored twice a week and the total number of leaves recorded at the end of the vegetative phase.

Defining phenological phases. Maize is a diecious plant in which male and female organs are located at different positions. The tassel is located at the plant apex and the ear at the axil of an intermediate leaf, typically at the 11th leaf for a hybrid bearing 17 leaves in total. The dates of four phenological events were estimated for all hybrids, namely floral transition on the apex, silk initiation on ovaries, end of grain abortion and grain maturity (Fig. 2). Direct identification of these event requires the dissection of the ear, an impossible task at high throughput for hundreds of hybrids in the field. Hence, they were not directly observed but they strictly correspond to leaf stages[19] so corresponding dates were estimated on the basis of the progression of leaf number. This resulted in genotype-specific thermal times for floral transition and silk initiation, corresponding to leaf numbers (n) of $n=\frac{n_f}{2}-1$ for floral transition and $n=n_f-2$ for silk initiation[19], where n_f is the final leaf number of the studied hybrid (Fig. 2). The phenological phase between these first two events is referred to as 'vegetative', following the current usage in modeling. Its beginning and end were calculated in thermal time units for each hybrid on the basis of leaf stage measurements in the platform. Corresponding dates in calendar time were calculated in each field on the basis of hourly temperature records. The flowering phase elapsed from the date at which silks were initiated on ovaries to the date at which abortion does not occur anymore in most environmental scenarios[20,36]. The latter was estimated as 14 equivalent days at 20 °C after female flowering (silking), observed in each field[36]. The calendar dates for end of the flowering phase were calculated in all fields on the basis of this thermal time and on temperature records. The grain-filling phase elapsed from end of abortion to physiological maturity defined as the date at which grain water content decreases to 0.36 g g^{-1} dry matter[21]. To estimate this date, we simulated the time course of grain water content before physiological maturity, following a published method (dry-down phase)[37]. Briefly, the rate of decrease in grain water content is inversely proportional to the water to be removed, calculated as the difference between the current water content and the equilibrium water content. Grain water content was measured at harvest in every hybrid of every field and equilibrium water content was calculated every day on the basis of air relative humidity and air temperature[37]. We estimated grain water content for each day backward from harvest, until it reached 0.36. Grain number is determined during the first two phenological phases and grain weight during the second and the third phase[15,20,38].

Selection of environmental variables affecting grain yield components. Environmental conditions as observed during the first two phenological phases were used as predictors in genotype-specific regression models for grain number. To avoid confounding effects, R_{int}, T_{night} and T_{max} were analyzed in well-watered experiments only (WW scenario), in 20 experiments with soil water potential above −0.1 MPa (ref. [12]). Indeed, the interactions between the effects of T_{max}, T_{night} and R_{int} on grain number were not significant in WW experiments (Supplementary Tables 6 and 7).

During the first phase, grain number in WW experiments was related to R_{int}, T_{night} *and* T_{max} ($r=0.62$, −0.59 and −0.66 respectively, for the correlation with grain number of the reference hybrid). Because T_{night} and T_{max} are correlated with R_{int} and because of the well-known effect of intercepted light on biomass[39], we only included R_{int} in this phase in the final regression models. This effect during an early phase confirms the link between vegetative growth (for example leaf elongation rate on the platform) and growth of reproductive organs in controlled conditions and fields (for example silk growth)[40,41].

During the second phase (encompassing flowering time), grain number in WW experiments was affected by T_{night}, that captured the largest part of the variability (for example $r=0.62$ for the correlation with grain number of reference hybrid), whereas the effects for T_{max} and R_{int} were not significant. Soil water potential, Ψ_{soil}, measured during the same second phase, affected grain number of water-deficit (WD scenario) plants. No direct correlation was observed between water potential and grain number but soil water potential was added as a significant predictor for grain number after its correction for R_{int} and T_{night}, a consequence of a partial correlation between soil water potential and grain number and a confirmation of the importance of the inter-relationships between light, temperature and soil water status[14].

On the basis of this stage-wise environmental variable selection procedure, we identified a final set of three environmental indices affecting grain number: T_{night} and Ψ_{soil} averaged over the second phase (encompassing flowering time) and R_{int} cumulated over the vegetative phase with no significant interaction between them. These indices were combined in a factorial regression model to dissect the G × E interaction of grain number (see main text and Fig. 4 for detail):

$$GN_{i,j}=\mu+e_j+g_i+\beta_{1,i}\,T_{night_{i,j}}+\beta_{2,i}\,R_{int_{i,j}}+\beta_{3,i}\,\Psi_{soil_{i,j}}+\varepsilon_{i,j} \quad (3)$$

The same procedure was applied to grain weight but no environmental index had a significant effect. We therefore fitted the additive model

$$GW_{i,j} = \mu + e_j + g_i + \varepsilon_{i,j} \quad (4)$$

and defined genotype-specific grain weight estimates $\widehat{GW_i} = \widehat{\mu + g_i}$. Because all datasets were balanced, $\widehat{\mu + g_i}$ was always equal to grain weight averaged across experiments.

Environmental scenarios were defined for stratified sampling (Table 1). They follow previously defined rules applying to the flowering period[12]. Three scenarios of temperature were identified, namely 'Cool' if mean night and maximum temperatures did not exceeded 20 and 33 °C, respectively during the flowering period, hot during days ('Hot(day)') if mean maximum temperatures exceeded 33 °C but mean night temperatures remained below 20 °C and 'Hot' in remaining cases. Well-watered and water-deficit scenarios involved a mean soil water potential above and below −100 kPa, respectively.

Narrow-sense heritabilities of genotype-specific parameters in the models for yield, grain number and grain weight were estimated at the genotypic level with a model assuming additive SNP effects using the R package heritability[42]. Overall, parameters showed a large genotypic variability with medium to large heritabilities: from −3,594 to 2,555 grains m^{-2} for g ($h^2 = 0.54$), from −86 to 182 grains m^{-2} °C^{-1} for β_1 ($h^2 = 0.71$), from −5.0 to 5.0 grains MJ^{-1} for β_2, ($h^2 = 0.51$), from −386 to 10,905 grains m^{-2} MPa^{-1} for β_3, ($h^2 = 0.78$) and from 246 to 365 mg for GW ($h^2 = 0.84$) (Fig. 3).

Selection of the genotyping data for genomic prediction. Genomic prediction of genotype-specific parameters in prediction models for yield, grain number and grain weight was performed using 41,722 SNPs originating from the 50 K Infinium HD Illumina array[29], combined with 1,312 SNPs identified at 120 published quantitative trait loci identified in the same panel for yield and related traits[12,35].

Training, test and validation of the genomic prediction. A random stratified sampling was repeated ten times by selecting 80% and discarding 20% of the 246 hybrids in each genetic group (Table 1) and/or selecting 80% and discarding 20% of the 25 experiments in each environmental scenario defined in an earlier study[12] (Table 1). Due to the limited number of experiments available, the two water-deficit scenarios (WD–Hot(day) and WD–Hot) were grouped while the three well-watered scenarios (WW–Cool, WW– Hot(day) and WW–Hot) were kept separate. This sampling method was chosen to maintain a good coverage of the total genetic space[43] and of the environmental space covered by the training set[44]. Four prediction datasets were created by combining hybrids and experiments depending on whether hybrids were used for parameter estimation (G) or not (nG) and whether the considered experiments were used for fitting the factorial regression with estimation of genotypic sensitivities (E) or not (nE) (Table 1). We drew ten times a random training set of 200 hybrids (G) and 20 experiments (E) from the initial 246 × 25 dataset, keeping the remaining 46 hybrids (nG) and five experiments (nE) as test sets. In the first dataset (G,E), 200 hybrids and 20 experiments were used as training set to estimate the traits values. The second prediction dataset $(G,nE)_{cross}$ involved the five experiments that were not used, together with the 200 training hybrids. The third prediction dataset $(nG,E)_{cross}$, involved the 46 hybrids that were not used in the 20 training experiments. The fourth prediction dataset $(nG,nE)_{cross}$, combined the 46 hybrids and five experiments that were not used. Two independent external datasets were also considered for an external validation of the prediction model for grain yield. The first one, $(G,nE)_{ext1}$ involved 12 external experiments that were not used and a subset of 32 already tested hybrids. The second one, $(nG,nE)_{ext2}$ involved 56 external maize hybrids in four external experiments that were not used (Table 1).

Genomic prediction model. A set of eight genotypic traits was considered. Five came from the grain yield dissection (equations (1) and (2)), namely the genotypic main effect for grain number, the three genotypic sensitivities and the genotypic mean grain weight (Supplementary Fig. 3a–e). The last three traits came from measurements in the platform or in the field and are used to define the phenological phases per hybrid, namely leaf appearance rate, silking date and final leaf number (Supplementary Fig. 3f–h). Genomic prediction was performed with a whole-genome regression model with the genotyping data previously described, using BayesR[45]. The above-mentioned traits were related to marker data following a linear regression model $\mathbf{y} = \mathbf{1}_n\mu + \mathbf{X\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ is the variable to predict, $\mathbf{X}$ is the n × m matrix of SNP genotypes, whose entries encode 0, 1 or 2 copies of the reference alleles and $\boldsymbol{\beta}$ is the m × 1 vector of SNP effects. For each SNP effect, BayesR assumes a prior distribution, which is a mixture of four zero mean normal distributions:

$$\pi_1 \times N(0, 0 \times \sigma_g^2) + \pi_2 \times N(0, 10^{-4} \times \sigma_g^2)$$
$$+\pi_3 \times N(0, 10^{-3} \times \sigma_g^2) + \pi_4 \times N(0, 10^{-2} \times \sigma_g^2)$$

The numbers π_1, π_2, π_3 and π_4 represent the probability of a SNP having no, small, intermediate or major effect, respectively. We ran the MCMC (Markov Chain Monte Carlo) chain for 60,000 iterations, including a burn-in period of 10,000 and used a thinning rate of 10.

Prediction of main environmental effect. The environmental main effects (e_j in equation (2)) for new experiments (nE) were predicted on the basis of main effects estimated in the training set (E). Estimated environmental main effects for the training set ($\widehat{e}_j$) were regressed on averaged environmental indices using the model:

$$\widehat{e}_j = a_2 T_{\text{night}_j} + a_1 R_{\text{int}_j} + a_3 \Psi_{\text{soil}_j} + \varepsilon_j$$

Estimates of a_1, a_2 and a_3 were used to predict environmental main effects for new experiments, which were then inserted in the factorial regression for prediction of grain yield (equation (2) and Fig. 4).

Prediction of grain yield using environmental relatedness matrices. To benchmark our method, we compared it with a method in which G × E is modeled using an environmental kinship matrix[3]

$$\Omega = \frac{1}{L}\sum_{l=1}^{L} \mathbf{w_l}\mathbf{w_l^t}$$

where the $np \times 1$ vectors $\mathbf{w_l}$ are standardized genotype-specific environmental indices, for $L = 12$ environmental indices (R_{int}, Ψ_{soil}, T_{night} and $T_{\max}$, in three phenological phases) observed for n genotypes in p environments. Using the R package BGLR (v.1.0.5; ref. [46]), we fitted the mixed model

$$Y_{i,j} = \mu_j + g_i + \sum_{l=1}^{L} w_l(i,j)\gamma_l + ge_{i,j} + \varepsilon_{i,j}$$

where $Y_{i,j}$ is the grain yield of the ith hybrid in the jth experiment, μ_j are environment specific intercepts and $\mathbf{g} \sim N(0, \sigma_g^2 K)$ is a $n \times 1$ vector of genotypic main effects, K being a genetic relatedness matrix on the basis of the same 41,722 SNPs used in BayesR. γ_l is the main effect of the lth environmental variable and $w_l(i,j)$ is the element in w_l corresponding to hybrid i and experiment j. The $np \times 1$ vector $\mathbf{ge}$ contains G × E effects, distributed as

$$\mathbf{ge} \sim N(0, \Omega \circ (ZKZ^t))$$

where Z is the binary incidence matrix assigning observations to genotypes and ∘denotes the entrywise (Hadamard) product[3]. The latter distribution can be derived from the sum of products of random effects of environmental index with equal variance and infinitesimal SNP effects[3]. We ran the MCMC chain for 60,000 iterations, including a burn-in period of 10,000 and used a thinning rate of 5.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The field data, the accessions list and the genotypic information associated with this study are stored in GnpIS (ref. [47]) and can be downloaded at https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/IASSTN (ref. [48]). Phenological data in the phenotyping platform can be downloaded at http://www.phis.inra.fr/openphis/web/index.php?r=document%2Fview&id=371 after logging as guest into the PHIS (ref. [49]) information system (www.phis.inra.fr).

References

28. Negro, S. S. et al. Genotyping-by-sequencing and microarrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. Preprint at https://www.biorxiv.org/content/10.1101/476598v1 (2018).
29. Ganal, M. W. et al. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PloS ONE* **6**, e28334 (2011).
30. Schnable, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
31. Wickham, H *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
32. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
33. Guilioni, L., Cellier, P., Ruget, F., Nicoullaud, B. & Bonhomme, R. A model to estimate the temperature of a maize apex from meteorological data. *Agric. For. Meteorol.* **100**, 213–230 (2000).
34. Hammer, G. L. et al. Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. *J. Exp. Bot.* **61**, 2185–2202 (2010).

35. Alvarez Prado, S. et al. Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. *Plant Cell Environ.* **41**, 314–326 (2018).
36. Borrás, L. Control of kernel weight and kernel water relations by post-flowering source-sink ratio in maize. *Ann. Bot.* **91**, 857–867 (2003).
37. Maiorano, A., Fanchini, D. & Donatelli, M. MIMYCS.Moisture, a process-based model of moisture content in developing maize kernels. *Eur. J. Agron.* **59**, 86–95 (2014).
38. Oury, V., Tardieu, F. & Turc, O. Ovary apical abortion under water deficit is caused by changes in sequential development of ovaries and in silk growth rate in maize. *Plant Physiol.* **171**, 986–996 (2016).
39. Monteith, J. L. Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. Lond. B* **281**, 277–294 (1977).
40. Welcker, C. et al. A common genetic determinism for sensitivities to soil water deficit and evaporative demand: meta-analysis of quantitative trait loci and introgression lines of maize. *Plant Physiol.* **157**, 718–729 (2011).
41. Chapuis, R., Delluc, C., Debeuf, R., Tardieu, F. & Welcker, C. Resiliences to water deficit in a phenotyping platform and in the field: how related are they in maize? *Eur. J. Agron.* **42**, 59–67 (2012).
42. Kruijer, W. et al. Marker-based estimation of heritability in immortal populations. *Genetics* **199**, 379–398 (2015).
43. Bustos-Korts, D., Malosetti, M., Chapman, S., Biddulph, B. & van Eeuwijk, F. Improvement of predictive ability by uniform coverage of the target genetic space. *G3* **6**, 3733–3747 (2016).
44. Malosetti, M., Bustos-Korts, D., Boer, M. P. & van Eeuwijk, F. A. Predicting responses in multiple environments: issues in relation to genotype × environment interactions. *Crop Sci.* **56**, 2210–2222 (2016).
45. Moser, G. et al. Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet.* **11**, e1004969 (2015).
46. Perez, P. & de los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483–495 (2014).
47. Steinbach, D. et al. GnpIS: an information system to integrate genetic and genomic data from plants and fungi. *Database* https://doi.org/10.1093/database/bat058 (2013).
48. Millet, E. J. et al. *A Multi-site Experiment in a Network of European Fields for Assessing the Maize Yield Response to Environmental Scenarios* https://doi.org/10.15454/IASSTN (2019).
49. Neveu, P. et al. Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. *New Phytol.* **221**, 588–601 (2018).

nature research

Corresponding author(s): Francois Tardieu

Last updated by author(s): Mar 27, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	Data analysis was performed using R version 3.4.3 with the functions «lm» for the factorial regression and «cor» for the accuracies. Genomic prediction was performed using BayesR (https://github.com/syntheke/bayesR). We also used the BGLR R package version 1.0.5. and the R package ggmap version 2.6.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The field data, the accessions list and the genotypic information associated with this study are stored in GnpIS and can be downloaded at https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/IASSTN.
Phenological data in the phenotyping platform can be downloaded at http://www.phis.inra.fr/openphis/web/index.php?r=document%2Fview&id=371 after logging as guest into the PHIS information system (www.phis.inra.fr).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Genotypes : The primary set involves 246 maize dent lines chosen for maximising the diversity in the dent group while keeping a restricted flowering window. An external test set involved 56 maize hybrids. Experiments involved 25 combinations site x year x watering treatment for the initial set, and two external sets consisting of 12 experiments (Ext1) and four experiments (Ext2), respectively.
Data exclusions	Four experimental sites among the initial 29 were discarded in this study due to lack of sufficiently detailed environmental information.
Replication	Each experiment was designed as an alpha-lattice design with two and three replicates per hybrid for watered and rain-fed treatments, respectively.
Randomization	Experiments were designed as alpha-lattice designs.
Blinding	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging