Predicting plant trait dynamics from genetic markers

Hobby, David; Tong, Hao; Heuermann, Marc; Mbebi, Alain J.; Laitinen, Roosa A. E. ...

2025-05
Nature Research
http://hdl.handle.net/10138/595717
Hobby, D, Tong, H, Heuermann, M, Mbebi, A J, Laitinen, R A E, Dell'Acqua, M, Altmann, T & Nikoloski, Z 2025, 'Predicting plant trait dynamics from genetic markers', Nature plants, vol. 11, no. 5. https://doi.org/10.1038/s41477-025-01986-y

nature plants

Article

https://doi.org/10.1038/s41477-025-01986-y

Predicting plant trait dynamics from genetic markers

Received: 22 August 2024

Accepted: 14 March 2025

Published online: 17 April 2025

Check for updates

David Hobby[1,2,6], Hao Tong[1,2,6], Marc Heuermann[3,6], Alain J. Mbebi[1,2], Roosa A. E. Laitinen[4], Matteo Dell'Acqua[5], Thomas Altmann[3] & Zoran Nikoloski[1,2]

Molecular and physiological changes across crop developmental stages shape the plant phenome and render its prediction from genetic markers challenging. Here we present dynamicGP, an efficient computational approach that combines genomic prediction with dynamic mode decomposition to characterize the temporal changes and to predict genotype-specific dynamics for multiple morphometric, geometric and colourimetric traits scored by high-throughput phenotyping. Using genetic markers and data from high-throughput phenotyping of a maize multiparent advanced generation inter-cross population and an *Arabidopsis thaliana* diversity panel, we show that dynamicGP outperforms a baseline genomic prediction approach for the multiple traits. We demonstrate that the developmental dynamics of traits whose heritability varies less over time can be predicted with higher accuracy. The approach paves the way for interrogating and integrating the dynamical interactions between genotype and environment over plant development to improve the prediction accuracy of agronomically relevant traits.

The phenome of a plant comprises the entirety of traits expressed at any given time and is the integrated outcome of the effects of genetic factors, environmental conditions and their complex interactions. Understanding how the crop phenome changes over time can help predict individual traits at specific timepoints in crop development. However, this problem is challenging not only because of the intricate dependence between individual traits but also because of differences in how the phenomes of specific genotypes change over the plant life cycle.

The classical approach of genomic prediction (GP) in crops trains machine learning models using data on traits measured in a population of genotypes at a specific timepoint based on genetic markers[1]. These models can be used to predict the traits in other genotypes for which genetic markers are available, thus foregoing additional measurements; the accuracy of predictions depends on representativeness of the training set and on the genetic architecture of the predicted trait[2]. As a result, GP accelerates genetic gain and reduces the need for labour-intensive field phenotyping in a test set of genotypes. While GP has been extended to allow and improve the simultaneous prediction of multiple correlated traits[3–5], existing GP approaches have not yet addressed the problem of predicting the dynamics of multiple traits, that is, the expression of multiple traits at different timepoints across the entire period of growth of the plant.

Advances in high-throughput phenotyping (HTP) technologies, including non-destructive approaches such as hyperspectral, multispectral, fluorescence and thermal imaging, have enabled the detailed capture of plant morphometric, geometric and colourimetric traits at various growth stages and have resulted in large-scale time-resolved heterogeneous datasets on crop phenotypes[6–10]. These developments offer the possibility of predicting time-resolved traits from genetic

[1]Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany. [2]Bioinformatics Department, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany. [3]Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research, Seeland OT Gatersleben, Gatersleben, Germany. [4]Organismal and Evolutionary Biology Research Programme, Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland. [5]Institute of Plant Sciences, Scuola Superiore Sant'Anna, Pisa, Italy. [6]These authors contributed equally: David Hobby, Hao Tong, Marc Heuermann. e-mail: zoran.nikoloski@uni-potsdam.de

markers, with the potential to revolutionize the GP. Recent efforts have cast the prediction of the dynamics of multiple morphometric traits as a spatio-temporal sequence prediction problem. The proposed solutions rely on deep learning approaches, such as spatio-temporal long short-term memory or memory-in-memory networks[11], generative adversarial networks[12] and a U-net encoder–decoder[13]. However, rather than directly predicting time-resolved crop traits, these approaches aim to predict a time series of images that are then used to infer the traits of interest. These studies do not incorporate genetic marker data, limiting their relevance in predicting multiple, time-resolved traits for unseen genotypes; furthermore, the underlying models are, presently, challenging to apply.

The time-resolved prediction of crop phenotypes can also be addressed by dynamic mode decomposition (DMD)[14]. DMD is an advanced data-driven method used in decomposing data from complex, time-dependent systems into spatio-temporal modes that provide low-dimensional description of the systems' dynamics. These modes render DMD particularly useful for understanding and predicting the behaviour of systems characterized by high-dimensional data. DMD has been applied across different fields, including fluid mechanics, engineering, epidemiology[14] and neuroscience[15]; however, no attempts have yet been made to apply DMD to data from HTP technologies or to combine this method with genetic markers.

Here, we combine DMD with GP to predict the genotype-specific dynamics of multiple morphometric, geometric and colourimetric crop traits using HTP data measured on a maize population of recombinant inbred lines (RILs) as well as an *Arabidopsis thaliana* diversity panel. Our approach, termed dynamicGP, showed consistent prediction accuracy for multiple traits and across multiple timepoints. When compared with a baseline GP approach, dynamicGP exhibited higher accuracies at the majority of timepoints and irrespective of traits, showing that it can effectively capture the modes and, therefore, the developmental dynamics of untested genotypes based on genetic marker data.

Results

DMD can predict dynamics of geometric traits in maize

DMD is a data-driven diagnostic and time-series prediction method that has found diverse applications from engineering to finance[14] and has only recently been applied in the biological sciences[16]. To illustrate the effectiveness of DMD, we applied it to predict the dynamics of growth-related traits derived from multimodal HTP imaging measured at 25 timepoints, 5 days per week for 5 weeks, beginning at day 15 after sowing, in a maize multiparent advanced generation inter-cross (MAGIC) population of 347 RILs of which genotyping data were available for 330 (Methods and Supplementary Figs. 1 and 2). A 1-day shift of the second of three HTP experiments resulted in a 2-day gap and, thus, five consecutive measurements per week, as these were adjusted between the three experiments based on the days after sowing (DAS). By using network clustering, we selected 50 traits as representatives for the compendium of 498 top, side and combined measurements of leaf-colour and leaf-geometric traits (Methods, Table 1 and Supplementary Table 1).

For a single maize line, the time-resolved phenotype was arranged into a $p \times T$ matrix X, where p is the number of traits and T is the number of timepoints (Fig. 1a). Two submatrices, X_1 and X_2, which are offset by a single timepoint are derived from X (Fig. 1a) and then used to calculate a best-fit linear $p \times p$ operator (matrix) A, linking the phenotype at one timepoint with the phenotype in the following timepoint (Fig. 1a). We note that the usage of linear operator does not assume that the traits truly change linearly with time. The operator A is usually non-symmetric and can be computed directly using the classical DMD approach ('Algorithm 1' section in the Methods).

We first tested how accurately the operator A, calculated using the data from a specific genotype, was able to recreate the underlying time-resolved phenotypes. The collection of predicted (P) traits, $x_{P,t+1}$, at timepoint $t+1$ can be obtained from the multiple traits, x_t, at timepoint t using the operator A. The essential decision one can then make is whether the multiple traits $x_{P,t+1}$ are obtained by using: (1) measured traits x_t at timepoint t, resulting in the iterative version of dynamicGP or (2) predicted traits $x_{P,t}$ at timepoint t, resulting in the recursive version of dynamicGP, since the prediction at any timepoint t is obtained by recursively unfolding the process forwards from an available measurement of x_t at a given initial timepoint $t=1$. Since the operator A captures the dynamics of multiple traits, it allows prediction of the multiple traits at any timepoint, irrespective of where it lies in the temporal expression of traits. We note that the recursive scenario is particularly relevant in applications, since it is expected to minimize the need for phenotyping.

Table 1 | The representative traits for the maize MAGIC population (includes five representative traits for the maize MAGIC population as depicted in Fig. 3a–c)

Minimum	Average blue value of plant pixel colours from the top images
Q1	Average saturation of the fraction of green coloured pixels in top-view images
Mean	Mean of the saturation value of the plant pixel colours in the top-view images
Q3	Standard deviation of the blue value of the plant pixel colours in the fluorescent side-view images
Maximum	Average blue value of the plant pixel colours from the top images

We found that the operator A, derived from the classical DMD approach, was able to nearly perfectly recreate the training data when used iteratively, with a mean prediction accuracy of ~1 across all traits over the timepoints within a 5-day block of daily measurements and a mean prediction accuracy of 0.70 (±0.20) for the timepoints immediately following the 2-day gap, for which no data were available (Extended Data Fig. 2). In the recursive scenario, the classical DMD approach resulted in a perfect recreation of the training data for the first 5-day block; however, following the first 2-day gap, the prediction accuracy decreased rapidly as a result of error propagation (Extended Data Fig. 2). These results indicated that the classical DMD approach yielded models that overfitted to the training data and were not robust to slight deviations in the input data.

We additionally tested another DMD algorithm, named Schur-based DMD, which has improved numerical stability[17]. This algorithm uses the singular vectors associated with the r largest singular values of X_1 alongside the Schur decomposition to reconstruct another operator A_r ('Algorithm 2' section in the Methods), which is used in place of the operator A in the prediction procedure outlined above. We note that like the operator A, A_r is not necessarily symmetric. In comparison with A from the classical DMD, we found that use of A_r in the iterative scenario resulted in decreased mean prediction accuracy across all traits and all timepoints of 0.84 (±0.18). However, in the recursive scenario, which is particularly relevant for practical applications, we observed robust performance, supported by the mean prediction accuracy of 0.78 (±0.16) and 0.79 (±0.13) across all traits and all timepoints and for the last timepoint, respectively (Extended Data Fig. 3). These findings demonstrate that the Schur-based DMD exhibited good predictive performance and prompted our investigation of the extent to which the elements of A_r and corresponding matrices used in its derivation can be predicted from genetic markers.

The building blocks of dynamicGP are heritable

For each genotype, the operators A and A_r, as well as the intermediate component matrices in the Schur-based DMD ('Algorithm 2' section in the Methods), can be determined from phenotypic data. Here, we treat each individual entry of these intermediate component matrices as a trait in a single-trait GP model (Fig. 1b). Our approach, termed

https://doi.org/10.1038/s41477-025-01986-y

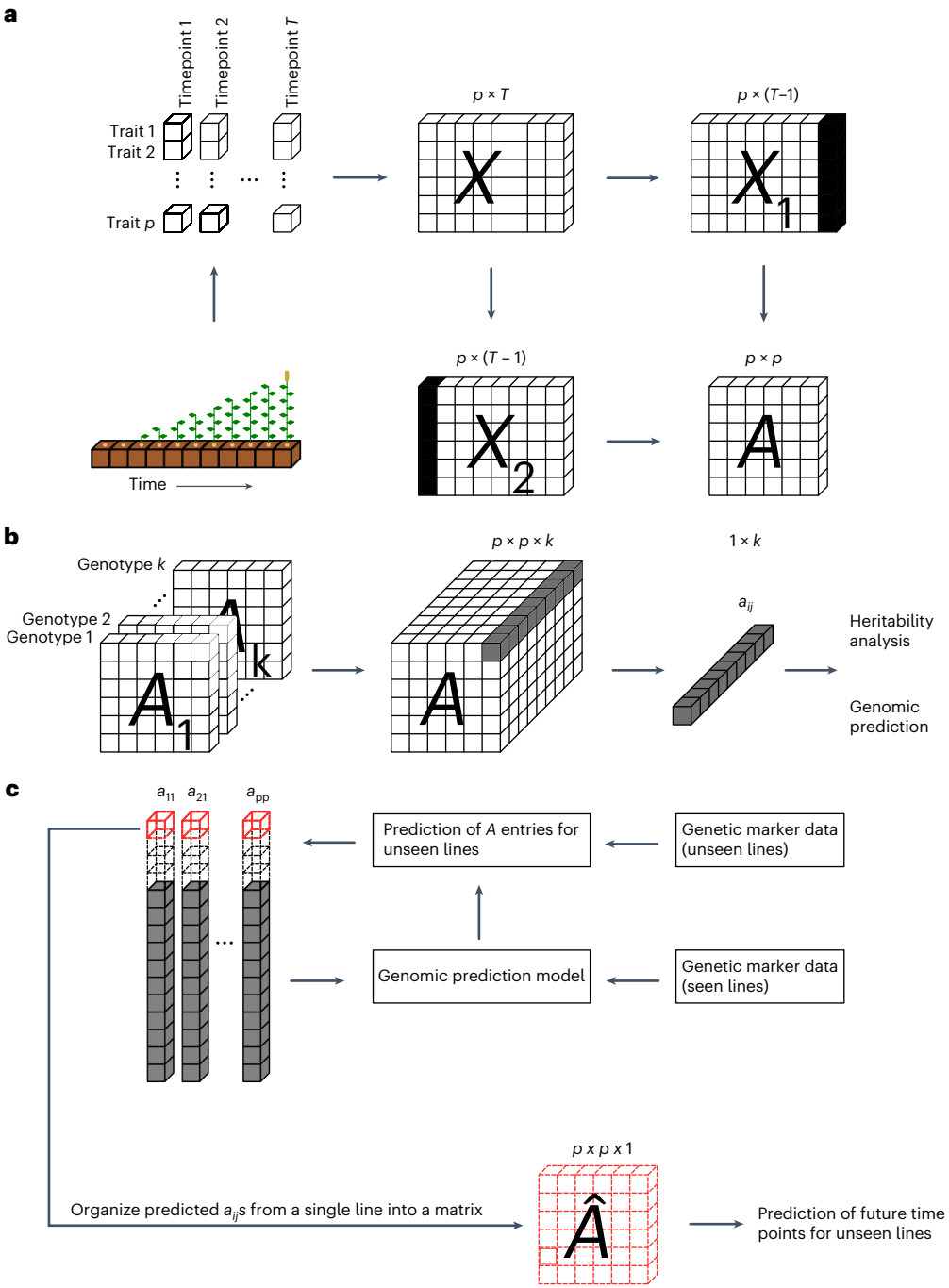


Fig. 1 | A schematic representation of dynamicGP. a, For a single genotype with measurements for p traits at T timepoints, we seek to find a time-invariant best-fit linear operator A, which transforms the phenome, given by measurements for p traits, at time t into the phenome at time $t+1$, according to $x_{t+1} = Ax_t$. When organized into matrix form for all timepoints, we obtain a $p \times T$ matrix X. Two submatrices, X_1 and X_2, are offset by a single timepoint and are derived from X by omitting the measurements at the last and the first timepoint, respectively, and are then used to calculate A using equation (3) of the classical DMD method in algorithm 1 (Methods). **b**, When the time-resolved data are available for k genotypes, we obtain a $p \times p \times k$ tensor whose elements can be treated as traits in heritability analyses and GP. **c**, The models for the operator entries are trained and then used to obtain predictions of entries, $\hat{a}_{ij}$ (red), for unseen lines, which are then gathered into a matrix that is then used to predict future timepoints for the line.

dynamicGP, uses genetic markers as predictors and the individual matrix entries as the response variables in ridge-regression best linear unbiased prediction (RR-BLUP) models, which are collected together and organized into their respective matrices (Fig. 1c). These are, in turn, combined with selected phenotypic measurements to make longitudinal predictions of plant traits for unseen genotypes.

Here, the use of the Schur-based DMD has the advantage that its intermediate components contain fewer entries than either A or A_r. This renders the components of Schur-based DMD a potential entry point for the computationally efficient prediction of A_r from genetic markers. We examined the marker-based heritability of the entries of all the intermediate matrices of the Schur-based DMD using 70,846 single nucleotide polymorphisms (SNPs) from the MAGIC maize population (Methods). These matrices include those of the singular value decomposition of $X_1 = U\Sigma V^{\mathrm{T}}$, the rank-reduced representation of the operator A, denoted by $\widetilde{A}$, the matrices Q and R resulting from its Schur

https://doi.org/10.1038/s41477-025-01986-y

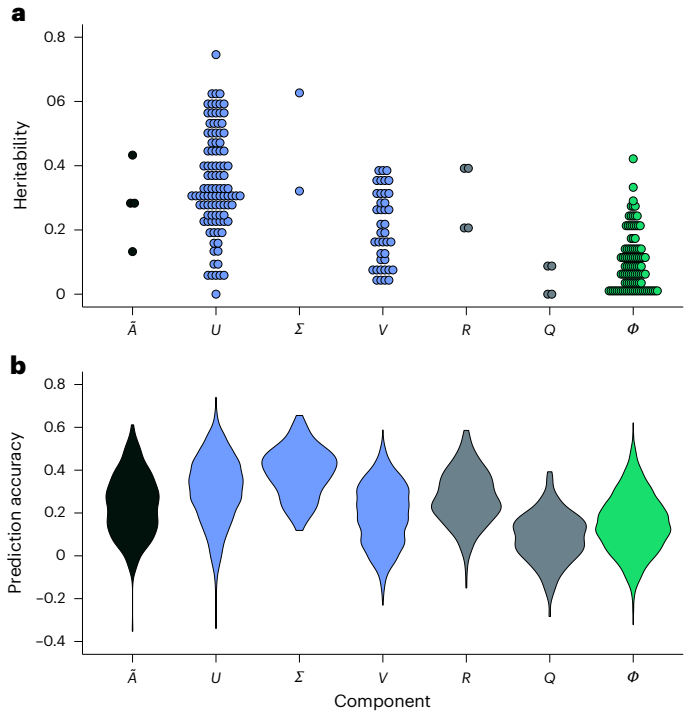


Fig. 2 | Heritability estimates and prediction accuracy of the building blocks of dynamicGP on the maize dataset. a, The heritability estimates for each entry of the matrices from algorithm 2 used as building blocks of dynamicGP, as obtained by genomic-relatedness-based restricted maximum likelihood (GREML). Here, $\tilde{A}$ denotes the matrix resulting from the rank-reduced proper orthogonal decomposition-projected representation in algorithm 2; U, Σ and V are obtained from the singular value decomposition of X_1; R and Q are matrices obtained from the Schur decomposition of $\tilde{A}$, while Φ collects the projected modes from DMD of the operator A. **b**, The prediction accuracy of the building blocks of dynamicGP using RR-BLUP in 20 iterations of a 5-fold cross-validation.

decomposition as well as the projected DMD modes, Φ ('Algorithm 2' section in the Methods).

We first observed that the non-zero heritability measures for the entries of these matrices were obtained by using the first two singular vectors ($r = 2$, Methods). Specifically, we found that the mean heritability of $\tilde{A}$ entries was 0.28, ranging from 0.13 to 0.43. Similarly, the entries of matrix U_r exhibited a mean heritability of 0.35, ranging from ~0 to 0.75. Moreover, the entries of Σ_r displayed a mean heritability of 0.47, while those of V_r showed a mean heritability of 0.20, ranging from 0.03 to 0.39 (Fig. 2a). Regarding the Schur decomposition of $\tilde{A}$, the mean heritability of R entries was 0.30, ranging from 0.20 to 0.39. In addition, the Q entries demonstrated a mean heritability of 0.04, ranging from ~0 to 0.09. Notably, the mean heritability of Φ entries was 0.10, ranging from ~0 to 0.42 (Fig. 2a). These findings indicated that substantial parts of the building blocks of dynamicGP exhibit moderate to high heritability, suggesting that they may be predictable from genetic markers, and we can therefore use them to predict operator A_r.

To test this hypothesis, we used 20 iterations of a 5-fold cross-validation to examine the genomic predictability of the components of the singular value decomposition of $X_1$and the entries of $\tilde{A}$ using SNPs. The entries of $\tilde{A}$ exhibited a mean prediction accuracy of 0.24 (±0.15), while the entries of matrix U_r demonstrated a slightly higher mean accuracy of 0.31 (±0.16). Moreover, the entries of Σ_r showed a mean prediction accuracy of 0.39 (±0.12), whereas those of $V_r$displayed a slightly lower mean accuracy of 0.19 (±0.14). Regarding the Schur decomposition of matrix $\tilde{A}$, the entries of R exhibited a mean accuracy of 0.27 (±0.13), while those of Q showed a lower mean accuracy of 0.08 (±0.12); in addition, the mean accuracy of the entries of Φ was 0.15 (±0.14) (Fig. 2b). These results suggested that the building blocks of dynamicGP can indeed be predicted from genomic data using standard GP models.

Two versions of dynamicGP differ in performance in maize

Having established that the elements of the building blocks of dynamicGP are heritable and predictable, we next used the predicted values for each entry in matrices Φ and R. Specifically, we used ten iterations of a 5-fold cross-validation with a validation step to recreate the predicted operator A_r for each unseen line (Fig. 1c). These predicted operators were used to make longitudinal predictions of all 50 traits over the entire time domain using the two versions of dynamicGP, iterative and recursive, defined similarly to the usage of DMD (Methods).

We found that the iterative approach yielded more consistent accuracies across the investigated timepoints compared with the recursive approach (Fig. 3a). Across all traits and timepoints, the iterative approach resulted in a mean prediction accuracy of 0.44 (±0.32). The average blue value of plant pixel colours from top images, which quantifies the blue–yellow colour space of visible light (see Supplementary Table 1 for trait descriptions), emerged as the best-performing trait, with a mean prediction accuracy of 0.85 (±0.07). In contrast, the mean of the *a*value of plant pixel colours in side-view images showed the lowest prediction performance of −0.20 (±0.14). Conversely, the recursive approach had a mean prediction accuracy of 0.22 (±0.25) across all traits and timepoints. Here, mean of the *a*value of plant pixel colours in fluorescent top-view images could be predicted with mean prediction accuracy of 0.52 (±0.14), while the normalized fraction of pixels with mean 12% brightness in top images, as quantified in the hue, saturation and value colour space, again displayed the lowest mean prediction accuracy of −0.18 (±0.13). The recursive predictions, as expected, tended towards zero at later timepoints as prediction errors compounded over time.

We observed that traits exhibiting consistent heritability over time, indicated by smaller values for the coefficient of variation, tended to demonstrate higher prediction accuracies across all timepoints. This relationship was supported by a moderate negative Pearson correlation coefficient of −0.46 between the mean prediction accuracy for the traits from iterative dynamicGP across all timepoints and the coefficient of variation of the heritability estimates across the entire timepoints (Fig. 3d). The corresponding correlation coefficient for recursive dynamicGP was −0.41 (Supplementary Fig. 2). Therefore, in line with expectations, traits whose heritability does not vary during development were found to be better predicted by dynamicGP approach.

DynamicGP outperforms baseline models in maize

To assess the predictive performance of dynamicGP, we compared it with a GP baseline approach using RR-BLUP models, which is the standard in breeding programmes. The RR-BLUP models were trained with data on each trait from the first timepoint in a training set and used to predict the trait in all subsequent timepoints for the testing set. Inspecting the difference of performance between dynamicGP and the baseline for traits with different predictability, we found that both the iterative and recursive versions of dynamicGP outperformed the baseline RR-BLUP models consistently across all investigated timepoints (Extended Data Fig. 4). Specifically, the mean prediction accuracy across all traits for the first timepoint, on which the models were trained, was 0.26 (±0.15). While this accuracy was maintained when the baseline RR-BLUP models were applied to predict the traits over the first three subsequent timepoints, already at the second timepoint, both versions of dynamicGP outperformed the baseline (Extended Data Fig. 4). Moreover, the iterative version of dynamicGP outperformed both the baseline and the recursive version of dynamicGP for every timepoint.

We examined the differences in mean prediction accuracy across all timepoints between dynamicGP and the baseline models for each trait. The best-performing traits using the recursive method, including

https://doi.org/10.1038/s41477-025-01986-y

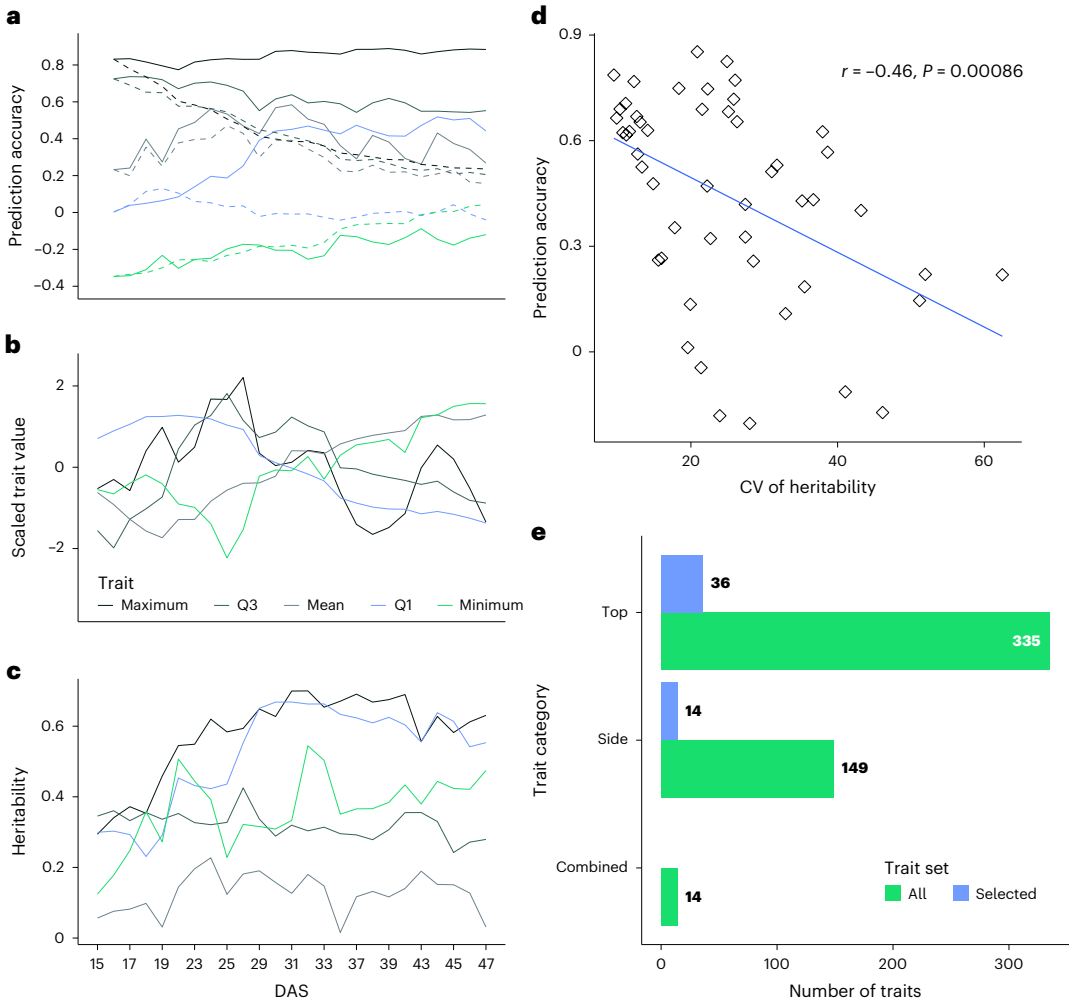


Fig. 3 | Consistency of trait heritability across time influences the prediction accuracy of dynamicGP on the maize dataset. a, The longitudinal prediction accuracy of representative traits corresponding to the maximum, minimum and mean, as well as the first and third interquartiles of all 50 traits across the 24 predicted timepoints (Table 1). The operator A_t was obtained from dynamicGP using 5-fold cross-validation (CV). The solid and dashed lines represent predictions from the iterative and recursive versions of dynamicGP, respectively. **b**, The time-resolved, scaled mean trait values across all accessions, depicting the different dynamics of traits. **c**, The heritability of the five representative traits across time (see Extended Data Fig. 1 for depiction of the dynamics of these traits). The colours for the particular traits are maintained in **a**–**c**. **d**, A Pearson correlation between the mean prediction accuracy of traits from iterative dynamicGP and the coefficient of variation of heritability estimates of traits across time (The *P* value is derived from two-sided test from a sample of 50 points). **e**, A category membership of traits in the full dataset (498 image-drived traits) compared with the traits that were selected by our clustering method for inclusion in the analysis (50 image-derived traits).

the mean of the *a* value of plant pixel colours in fluorescent top-view images and the mean of the *x* values in the fluorescent *xyz* colour space from top images, showed mean prediction accuracies above 0.5 of all timepoints, which was 15% higher (0.52 versus 0.45) than the best-performing trait in the baseline models (Fig. 4a). In the iterative method the best-performing traits, namely the average blue value of plant pixel colours from top images, as well as the mean of the blue value of plant pixel colours in side-view images (Supplementary Table 1), showed mean accuracies above 0.8 at all timepoints, which was 89% higher (0.85 versus 0.45) than the best trait from the baseline model predictions (two-sided *t*-test, $P < 0.001$ for all timepoints, Fig. 4a). Furthermore, both versions of dynamicGP yielded mean squared errors much lower than the RR-BLUP baselines (Extended Data Fig. 5).

We performed an additional test to determine the ability of our method to predict unseen lines in unseen timepoints (Supplementary Fig. 3, scenario 2). To this end, we trained a dynamicGP model on only the first 20 timepoints and tested its performance on the final 5 timepoints. We approached this in two ways: (1) beginning with the 1st timepoint (day 15) and predicting through to the end of the time series (day 47) and (2) beginning with the 20th timepoint (day 40) and predicting the 21st through the 25th (from day 43 to day 47). We compared these predictions with those from the RR-BLUP baseline models trained on day 15 and day 40, respectively, and used in predicting the remainder of the time series through to day 47. We observed that in both the tested configurations, iterative and recursive dynamicGP outperformed the respective baseline models (Supplementary Fig. 4). Furthermore, the two dynamicGP versions both outperformed the baseline with regards to the mean squared error (Supplementary Fig. 5) and in terms of the number of better predicted traits (Fig. 4b).

DynamicGP can be applied to data from *A. thaliana*

To demonstrate the applicability of dynamicGP, we tested our approach with the BLUEs for 132 traits in time series consisting of 13 timepoints measured in a diversity panel of *A. thaliana* composed of 382 genotypes (experiment number 3 (EXP3) from ref. 18). After clustering of the traits following the same protocol as for maize, 45 traits were used in dynamicGP. We observed that although algorithms 1 and 2 performed as expected (Supplementary Fig. 6 and

https://doi.org/10.1038/s41477-025-01986-y

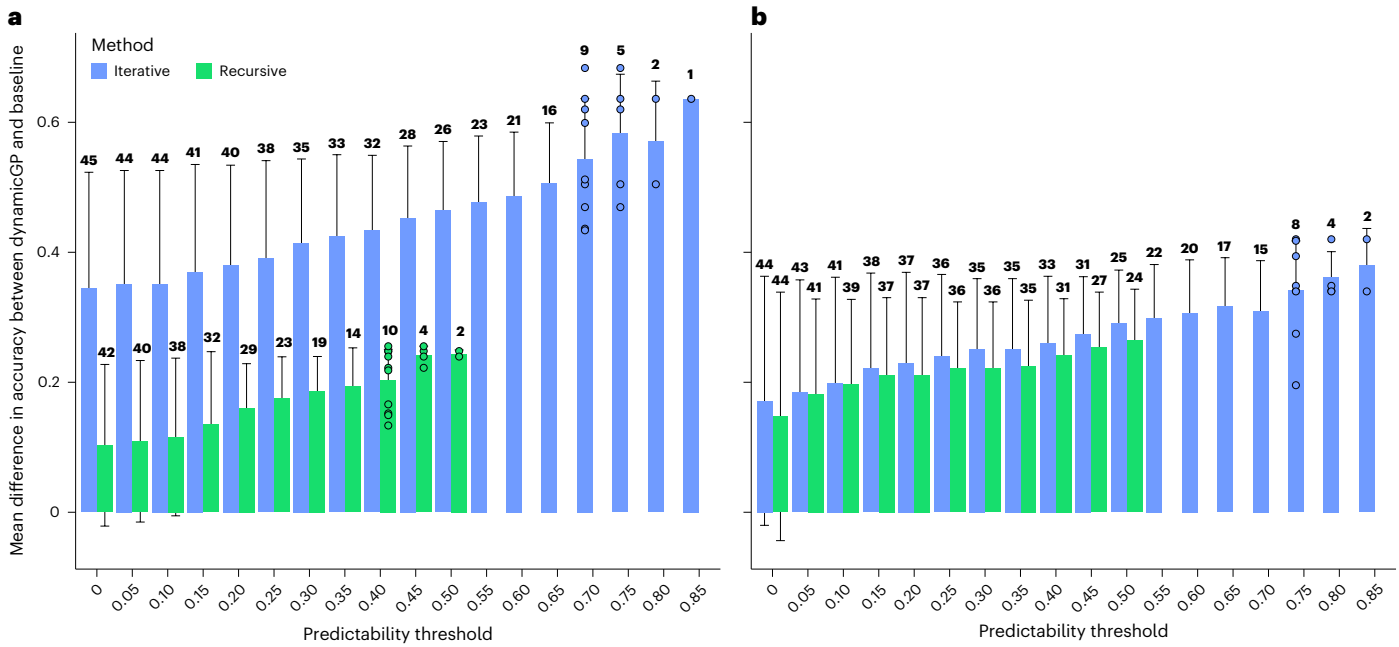


Fig. 4 | DynamicGP outperforms the baseline models on the maize dataset. **a**, The prediction of the complete time-series in iterative and recursive configurations in unseen lines. **b**, The predictions of only the last five timepoints in unseen lines in models trained with the data from the first 20 timepoints in the maize dataset. We determined the difference in performance of the iterative and recursive versions of dynamicGP and baseline RR-BLUP models, for traits whose predictability is above a threshold value, specified on the *x* axis. The number of traits with predictability larger than a specified threshold value is indicated above the respective bar. The best predicted traits with dynamicGP exhibited the greatest difference from the baseline predictions. The bars indicate the mean (±standard deviation) difference in trait prediction accuracies of traits that have mean prediction accuracies above the threshold. The number of samples (*n*) is specified on the top of each bar.

Extended Data Fig. 6), the matrix entries of the intermediate component matrices exhibited much lower heritabilities than the traits in the maize dataset, above, across all values for the number of components, *r* (Supplementary Figs. 7 and 8). As a result, we observed low prediction accuracies for the components of R and Φ (Extended Data Fig. 7). Despite this finding, we obtained positive prediction accuracies for some of the included traits (Extended Data Figs. 8a and 9). Note that the time-resolved heritabilities for the traits in the *A. thaliana* diversity panel were in general lower than those observed in the maize MAGIC dataset (Extended Data Fig. 8c). Interestingly, both iterative and recursive versions of our method again outperformed RR-BLUP baseline predictions in terms of the mean prediction accuracy across all traits and timepoints (Extended Data Fig. 9), mean squared error after the third timepoint (Supplementary Fig. 9) and in the number of traits with higher mean prediction accuracy across the full time series (Extended Data Fig. 10). These results further demonstrate the predictive power of dynamicGP over the baseline GP models.

Discussion

Developing computational approaches that can predict the dynamics of crop growth and developmental plasticity for unseen genotypes in new environments holds the promise to revolutionize breeding practices. Such approaches can provide insights into how the genotype affects the time-resolved phenome and also can highlight the temporal dependence between traits comprising the phenome. While there have been important developments in GP of multiple traits[3–5], they essentially operate with snapshot data and do not capture the dynamics of crop traits.

Here, we introduce dynamicGP, a computational approach that resolves the problem of predicting the growth dynamics across development in crops for which time-series measurements of morphometric and geometric growth-related traits for multiple genotypes are available from HTP platforms[6–10]. DynamicGP essentially applies GP to the building blocks of DMD, thus facilitating the prediction of multiple traits over time of unseen genotypes. Using a time-resolved HTP data collected for 5 days per week for 5 weeks beginning at day 15 after sowing from RILs of a maize MAGIC population, we showed that both the iterative and the recursive versions of dynamicGP outperformed the baseline GP approach. This finding was confirmed by applying dynamicGP to an independent HTP dataset from an *A. thaliana* diversity panel obtained from a different phenotyping platform.

In addition, we show that at the predictability threshold of 0.5 for two of the traits, recursive dynamicGP exhibited a better performance compared with the baseline GP approach. These traits were both top-down images measured in the fluorescence colour space. For the iterative dynamicGP, we found nine traits above a threshold of 0.7 where dynamicGP outperformed the baseline, including two traits above a threshold of 0.8, both measured in the visible colour space, namely, the average blue value of plant pixel colours from top images and the mean of blue value of plant pixel colours in side-view images (Supplementary Table 1). Further, we observed that the mean squared error of dynamicGP was lower for both the iterative and recursive versions than the baseline, with log-transformed mean squared errors of −4.30 and −3.89 for iterative and recursive, respectively, and −3.00 for the baseline (Extended Data Fig. 5). Therefore, dynamicGP yields predictions that are closer to the true values than the RR-BLUP baseline models.

The performance improvement of dynamicGP over the classical GP approach is due to the usage of time-resolved phenotypic data. While the availability of such data is rapidly increasing due to deployment of HTP in controlled environments and on the field, dynamicGP is more resource intensive than the classical GP. However, dynamicGP does not rely on model fitting (for example, of growth curves) and subsequent usage of the model parameters in GP, thus avoiding propagation of error and often unfounded assumptions about models that best describe the dynamics of single traits. Instead, dynamicGP builds on

the advantages of DMD that allows the simultaneous investigation of multiple traits over time.

Although dynamicGP is applicable with any time-resolved phenotypic data on multiple traits, we tested its performance with traits retained after correlation analysis. The selection of traits was conducted to avoid the consideration of highly correlated traits that may affect predictive performance. Similar to any data-driven approach, we hypothesize that the performance of dynamicGP depends on the number of traits included and timepoints considered, which may vary between datasets. Both the recursive and iterative versions of dynamicGP offer the possibility to make predictions about plant phenotypes of unseen genotypes given measurements of specified timepoints in time, thus reducing the need for extensive temporal measurements. In this regard, while dynamicGP provides the means to overcome the challenge of predicting temporal phenomes for unseen genotypes, it inherits the limitations of GP, such as the transferability of the models to different populations and environments[2].

Future developments of dynamicGP can rely on extensions of DMD to consider effects of environmental factors[19]. These will facilitate further refinements of the proposed approach that are expected to have very substantial impact on breeding crop varieties adapted to particular regions as well as precision agriculture. While the application of dynamicGP considered morphogeometry and physiology-related traits, future extensions to dynamicGP will consider their relationship to agronomic phenotypes, assessed longitudinally or in a single timepoint.

Methods

Data generation and reuse

In three independent HTP experiments, 347 RILs and nine founder lines of a MAGIC maize population[20] were screened for natural variation in the automated HTP facility for large plants at the Leibniz Institute of Plant Genetics and Crop Plant Research in Gatersleben, Germany[21]. The panel was divided into three equal-sized subgroups and analysed in three consecutive experiments: 2113MH (116 RILs and nine founders), 2121MH (116 RILs and nine founders), and 2137MH (115 RILs and eight founders). To ensure compatibility of the data, each HTP experiment was supplemented by an additional randomly drawn overlap of 21, 21 or 27 RILs, respectively, half of which came from each of the other two experiments. The experimental design followed a complete randomized block design with three replicates per genotype (three carriers with two plants each) overlapping with three blocks corresponding to 4 of the 12 lanes of the HTP system. The image acquisition began at 15 DAS for 31 timepoints per experiment and 6 days per week. An image acquisition event was a carrier with two plants phenotyped by one image in the top view and four images in the side view at 22°, 45°, 112° and 135°. Altogether 347,740 images were taken (2113MH: 114,710 images, 91,768 side view, 22,942 top view; 2121MH: 116,840 images, 93,472 side view, 23,368 top view; 2137MH: 116,190 images, 92,952 side view, 23,238 top view). The traits were derived from images by IAP version 2.3.0 (ref. 22) and were analysed by alignment by DAS. A shift of the sowing date by 1 day in 2121MH allowed alignment to DAS for only 5 days per week, so that 2 days were missing each week. Consequently, the timepoints were not consecutive but were organized in 5-day blocks, with one measurement taking place each day and no measurements for 2 days over 5 weeks.

The analysed data comprised measurements for 498 traits from side, top and combined categories derived from multicolour-space image analysis[18] (Fig. 3e and Supplementary Table 1). A linear mixed model with random effect of genotype, experiment, lane, genotype-by-experiment interaction and replicate nested in experiment was built for each timepoint. The trait heritability was calculated by the ratio between the genotype variance component and total variance component. A summation of the best linear unbiased prediction value of the genotype and fixed effect of intercept were used as the final trait values.

The genotyping data were available for 330 maize RILs consisting of 79,557 high-quality SNPs derived from SPET genotyping[23]. The missing genotyping data were imputed using Beagle 5.2 (ref. 24). We then removed the SNPs with a minor allele frequency <0.05, resulting in a final set of 70,846 SNPs for GP models.

The *A. thaliana* diversity panel consisted of 384 accessions phenotyped for 132 traits at 15 timepoints in the same facility as the maize and corresponds to the constant light treatment in EXP3 from an existing study[18]. The genomic data were available for 382 accessions, comprising 207,257 SNPs after filtering for a minor allele frequency smaller than 0.05. The first timepoint was removed due to the large number of missing values, leaving 14 timepoints for the analysis. The missing trait values were imputed using the R package 'missForest'[25].

Clustering and trait selection

The traits were first clustered using the Mantel correlations over all genotypes and timepoints, represented as a matrix. To this end, a network was created with nodes representing traits and edges denoting intertrait Mantel correlations of a value above 0.96. Modularity clustering, as implemented in the R package 'igraph'[26], was used to create clusters of traits, from which the trait with the highest mean SNP-based heritability over the considered timepoints was selected as a cluster representative in the following analyses. This resulted in 50 representative traits from three categories for the maize dataset (Fig. 3e). The mean Mantel correlation between all traits was 0.63. Using this clustering method, the mean within-cluster intertrait Mantel correlation was 0.97, while the mean intercluster correlation was 0.51. The investigations of the hierarchical clustering based on the distance matrix resulting from the Mantel correlation resulted in clusters with mean within-cluster intertrait correlations approximately equal to the intercluster correlations and was therefore deemed unsatisfactory (see Supplementary Table 2 for a comparison of modularity and hierarchical clustering). The data for each trait were then minimum–maximum normalized across all the lines and timepoints and were used in modelling. The above outlined network-based clustering method was repeated on the 132 traits in the *A. thaliana* diversity panel[18], yielding 45 unique clusters from which we selected the representative traits that were used in the subsequent analysis.

DMD

The DMD determines a time-invariant best-fit linear operator A that transforms the measurements of data at one timepoint into the measurements at the following timepoint, that is,

$$x_{t+1} = Ax_t, \tag{1}$$

where x is a column vector of traits, and A is a $p \times p$ matrix. When the x vectors of all timepoints are concatenated into matrices, equation (1) becomes

$$X_2 = AX_1, \tag{2}$$

where X_1 and X_2 are $p \times (T-1)$ matrices that are offset by a single timepoint (Fig. 1b). From $X_1$and X_2, the matrix A can be directly calculated using

$$A = X_2 X_1^{\dagger}, \tag{3}$$

where $X_1^{\dagger}$ indicates the Moore–Penrose pseudoinverse of X_1. The operator A allows us to predict the columns of X beginning with x_1. In fluid mechanics p is often very large (that is millions), and T is in tens of thousands, and there are many algorithms that rely on lower-dimensional representations. We cannot use these representations since they rely

on eigenvalue decomposition that renders the lower-dimensional approximation symmetric. In our case, with $p = 50$ and $T = 25$, we can compute A directly from the data. However, our tests demonstrated that this approach is sensitive to deviations in input data. To address these issues and to improve numerical stability, we used algorithm 2, known as the Schur-based DMD[17], to find the modes and to obtain an approximation of A, denoted as A_r, detailed below.

Algorithm 1. Classical DMD:

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t, \tag{4}$$

$$\boldsymbol{X}_2 = \boldsymbol{A}\boldsymbol{X}_1, \tag{5}$$

$$\boldsymbol{A} = \boldsymbol{X}_2\boldsymbol{X}_1^{\dagger}, \tag{6}$$

Algorithm 2. Schur-based DMD:

$\boldsymbol{X}_1 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$	(7)	where equation (7) shows a singular value decomposition of $\boldsymbol{X}_1$.
$\widetilde{\boldsymbol{A}} = \boldsymbol{U}_r^T\boldsymbol{X}_2\boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1}$	(8)	where equation (8) shows a rank-reduced representation of $\boldsymbol{A}$ projected onto the r POD modes of $\boldsymbol{U}$.
$\widetilde{\boldsymbol{A}} = \boldsymbol{Q}^T\boldsymbol{R}\boldsymbol{Q}$	(9)	where equation (9) identifies two matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ through a Schur decomposition of $\widetilde{A}$.
$\boldsymbol{\Phi} = \boldsymbol{X}_2\boldsymbol{V}_r\sum_r^{-1}\boldsymbol{Q}$	(10)	where equation (10) shows projected DMD modes.
$\boldsymbol{A}_r = \boldsymbol{\Phi}\boldsymbol{R}\boldsymbol{\Phi}^{\dagger}$.	(11)	where equation (11) shows a reconstruction of truncated $\boldsymbol{A}_r$.

In a Schur-based DMD, U_r, V_r and Σ_r^{-1} denote the matrices from the singular value decomposition of X_1 restricted to the first r columns. To account for the 2-day gaps in our data, we modified our X_1 and X_2 such that X_1 omitted the timepoint immediately before the gap, and X_2 omitted the timepoint immediately following the gap so that the difference between the columns of the two matrices was always a single time step. An R implementation of algorithms 1 and 2 is available at https://github.com/dobby978/dynamicGP.

Heritability and GP analyses

To quantify the extent to which the selected traits and all the elements of all matrices defined in algorithm 2 can be predicted by genetic markers, we calculated the SNP-based heritability using genome-wide complex trait analysis (GCTA)[27]. The genomic relatedness matrix was generated using TASSEL 5 (ref. 28). We then investigated the accuracy of the GP using a RR-BLUP model, as implemented in the R package 'rrBLUP'[29], to predict individual elements of these matrices. The prediction accuracy for the elements was quantified as the mean Pearson correlation coefficient between the true and predicted values in 20 iterations of 5-fold cross-validations.

Selection of the r value in Schur-based DMD

The number of singular values r was selected by analysing the heritability and predictability of components in a Schur-based DMD using $r \in [2, 3, 4, 5, 6]$. The heritabilities of the singular values and entries of the first two singular vectors are relatively high (0.47 and 0.35 on average, respectively; Supplementary Fig. 10). Most entries of the singular vectors associated with the smaller singular values exhibited near-zero heritability (Supplementary Fig. 10). For the Schur decomposition of $\widetilde{A}$ with $r = 2$ ($\widetilde{A}_2$), the entries of Q and R exhibited a non-zero heritability (0.04 and 0.29 on average, respectively), while the entries of the projected DMD modes, Φ, had a mean heritability of 0.10 (Supplementary Fig. 11). We observed similar findings regarding to predictability that the mean prediction accuracies for the first two singular values and the entries of the singular vectors as well as of $\widetilde{A}$ were at least 0.1, while for other values of r, the prediction accuracy for the entries of $\widetilde{A}_r$ were near zero (Supplementary Fig. 12). Similarly, the entries of the matrices Q and R obtained from $\widetilde{A}_2$ showed a non-zero prediction accuracy, while Q and R of $\widetilde{A}_r$, with r larger than 2, exhibited a near-zero prediction accuracy. The matrix Φ had a non-zero mean prediction accuracy for an r value of, at most, 3 (Supplementary Fig. 13).

As r is a hyperparameter that must be selected before training any machine learning models, we additionally tested the heritability of the matrix elements in only the training–testing set of a training–testing–validation configuration (Supplementary Fig. 3, scenario 1). In these tests, we again observed that the matrix elements corresponding to the first two singular vectors, suggesting that this method of selection of r holds when the lines in the validation set are removed (Supplementary Figs. 14 and 15). Therefore, we used the first two singular vectors ($r = 2$) in the subsequent analysis, as the matrix entries corresponding to larger values of r were not heritable nor predictable with sufficient accuracy to contribute any useful information. This also had the benefit of acting as a noise filter, which reduced overfitting. The performance of DMD algorithm 2 across all timepoints with a different number of singular vectors retained is shown in Supplementary Fig. 16.

Implementation of dynamicGP

The predicted (P) Q and R matrices, denoted Φ_P and R_P, for each line were created by collecting the entries obtained from the RR-BLUP models and placing them in their corresponding location in the matrices; Φ_P and R_P were then used in

$$A_P = \Phi_P R_P \Phi_P^{\dagger}, \tag{12}$$

to obtain a prediction of A_r, defined in algorithm 2 (equation (8)). For each iteration of the 5-fold cross-validation, we predicted the A_P operators in the testing fold using the model from data of the training folds. The A_P operators were then used to predict the trait values in testing fold. The accuracies of prediction were assessed for each trait using the Pearson correlation between the predicted trait values and the measured trait values. This resulted in 100 predictions of each trait at each timepoint, and the mean value was reported as the final predication accuracy.

Iterative and recursive versions of dynamicGP

Trait values in the following timepoints were predicted in two scenarios, namely iterative and recursive. In the iterative version of dynamicGP, we used the measured trait values at each timepoint to predict the values at the following timepoint, that is,

$$x_{P,t+1} = A_P x_t. \tag{13}$$

In the recursive version of dynamicGP, we used the measured trait values at $t = 1$ and then used the predicted values to predict the next timepoint over the time interval of interest, namely

$$\begin{aligned} x_{P,t+1} &= A_P x_t \ \text{ for } t = 1, \\ x_{P,t+1} &= A_P x_{P,t} \ \text{ for } t > 1. \end{aligned} \tag{14}$$

Reporting summary

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

Data availability

A panel of 347 MAGIC maize lines and their 9 founder lines were phenotyped at high throughput in the tier 1 phase of the CAPITALISE project[30]. The *A. thaliana* phenotyping dataset corresponds to EXP3 obtained from a published study[31]. The genotyping data for maize and *A. thaliana*

are available via Zenodo at https://doi.org/10.5281/zenodo.14959484 (ref. 32). Source data are provided with this paper.

Code availability

All code that was used to generate the results of this study is available via GitHub at https://github.com/dobby978/dynamicGP and via Zenodo at https://doi.org/10.5281/zenodo.14959484 (ref. 32).

References

1. Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
2. Alemu, A. et al. Genomic selection in plant breeding: key factors shaping two decades of progress. *Mol. Plant* **17**, 552–578 (2024).
3. Gill, H. S. et al. Multi-trait multi-environment genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Front. Plant Sci.* **12**, 709545 (2021).
4. Karaman, E., Lund, M. S. & Su, G. Multi-trait single-step genomic prediction accounting for heterogeneous (co)variances over the genome. *Heredity* **124**, 274–287 (2020).
5. Montesinos-López, O. A. et al. A singular value decomposition Bayesian multiple-trait and multiple-environment genomic model. *Heredity* **122**, 381–401 (2019).
6. Li, D. et al. High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Front. Bioeng. Biotechnol.* **8**, 623705 (2021).
7. Song, P., Wang, J., Guo, X., Yang, W. & Zhao, C. High-throughput phenotyping: breaking through the bottleneck in future crop breeding. *Crop J.* **9**, 633–645 (2021).
8. Sun, D. et al. Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality. *Plant Methods* **15**, 54 (2019).
9. Zhang, H., Wang, L., Jin, X., Bian, L. & Ge, Y. High-throughput phenotyping of plant leaf morphological, physiological, and biochemical traits on multiple scales using optical sensing. *Crop J.* **11**, 1303–1318 (2023).
10. Poorter, H. et al. Pitfalls and potential of high-throughput plant phenotyping platforms. *Front. Plant Sci.* **14**, 1233794 (2023).
11. Wang, C. et al. Predicting plant growth and development using time-series images. *Agronomy* **12**, 2213 (2022).
12. Yasrab, R., Zhang, J., Smyth, P. & Pound, M. P. Predicting plant growth from time-series data using deep learning. *Remote Sens.* **13**, 331 (2021).
13. Chang, S., Lee, U., Hong, M. J., Jo, Y. D. & Kim, J.-B. Time-series growth prediction model based on U-Net and machine learning in *Arabidopsis*. *Front. Plant Sci.* **12**, 721512 (2021).
14. Kutz, J. N., Brunton, S. L., Brunton, B. W. & Proctor, J. L. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (SIAM, 2016).
15. Brunton, B. W., Johnson, L. A., Ojemann, J. G. & Kutz, J. N. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci. Meth.* **258**, 1–15 (2016).
16. Hasnain, A. et al. Learning perturbation-inducible cell states from observability analysis of transcriptome dynamics. *Nat. Commun.* **14**, 3148 (2023).
17. Thitsa, M., Clouatre, M., Verriest, E., Coogan, S. & Martin, C. A numerically stable dynamic mode decomposition algorithm for nearly defective systems. *IEEE Control Syst. Lett.* **5**, 67–72 (2021).
18. Heuermann, M. C., Meyer, R. C., Knoch, D., Tschiersch, H. & Altmann, T. Strong prevalence of light regime-specific QTL in *Arabidopsis* detected using automated high-throughput phenotyping in fluctuating or constant light. *Physiol. Plantarum.* **176**, e14255 (2024).
19. Proctor, J. L., Brunton, S. L. & Kutz, J. N. Dynamic mode decomposition with control. *SIAM J. Appl. Dyn. Syst.* **15**, 142–161 (2016).
20. Dell'Acqua, M. et al. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol.* **16**, 167 (2015).
21. Junker, A. et al. Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Front. Plant Sci.* **5**, 770 (2015).
22. Klukas, C., Chen, D. & Pape, J.-M. Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. *Plant Physiol.* **165**, 506–518 (2014).
23. Ferguson, J. N. et al. The genetic basis of dynamic non-photochemical quenching and photosystem II efficiency in fluctuating light reveals novel molecular targets for maize (*Zea mays*) improvement. Preprint at *bioRxiv* https://doi.org/10.1101/2023.11.01.565118 (2023).
24. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
25. Stekhoven, D. J. & Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).
26. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJ. Compl. Syst.* **1695**, 1–9 (2006).
27. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
28. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
29. Endelman, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**, 250–255 (2011).
30. Heuermann, M. DataCite Commons: a phenotypic dataset of natural variation in image-related traits by high-throughput phenotyping of 347 MAGIC maize lines and their 9 founder lines in the CAPITALISE project. *e!DAL* https://doi.org/10.5447/ipk/2025/0 (2025).
31. Heuermann, M. C. et al. Strong prevalence of light regime-specific QTL in *Arabidopsis* detected using automated high-throughput phenotyping in fluctuating or constant light. *Physiol. Plant*. https://doi.org/10.1111/ppl.14255 (2024).
32. Hobby, D. et al. Code and data to reproduce the findings related to dynamicGP. *Zenodo* https://doi.org/10.5281/zenodo.14959484 (2025).

Acknowledgements

Part of this research was funded by the Horizon Europe research and innovation programme, project BOLERO (breeding for coffee and cocoa root resilience in low-input farming systems based on improved rootstock), HORIZON-CL6-2021-BIODIV-01-13, under grant agreement no. 101060393 (to Z.N.). The project was also supported by the European Union's Horizon 2020 research and innovation programme (grant no. 862201 to M.D., T.A. and Z.N.).

Author contributions

D.H. and H.T. analysed the data, implemented and tested the approach and performed research. M.H. performed experiments and analysed the data. A.J.M. analysed the data and contributed to the development of the approach. R.A.E.L. contributed to the development of the approach. M.D. and T.A. provided materials, supervised research and acquired funding. Z.N. designed research, supervised research, acquired funding and contributed to the development of the approach. D.H., H.T. and Z.N. wrote the first version of the manuscript. All authors contributed to revising and finalizing the paper.

Funding
Open access funding provided by Max Planck Society.

Competing interests
The authors declare no competing interests.

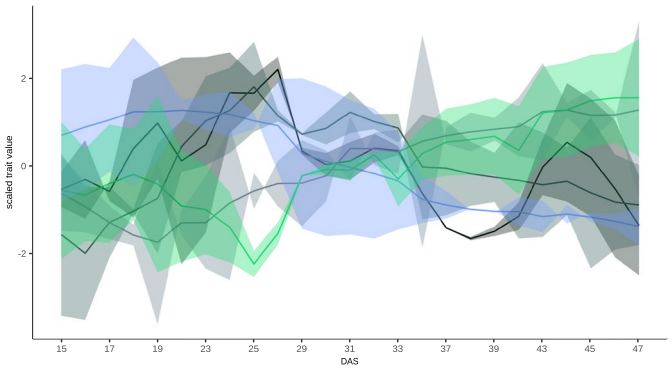
Additional information
Extended data are available for this paper at https://doi.org/10.1038/s41477-025-01986-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-025-01986-y.

Correspondence and requests for materials should be addressed to Zoran Nikoloski.

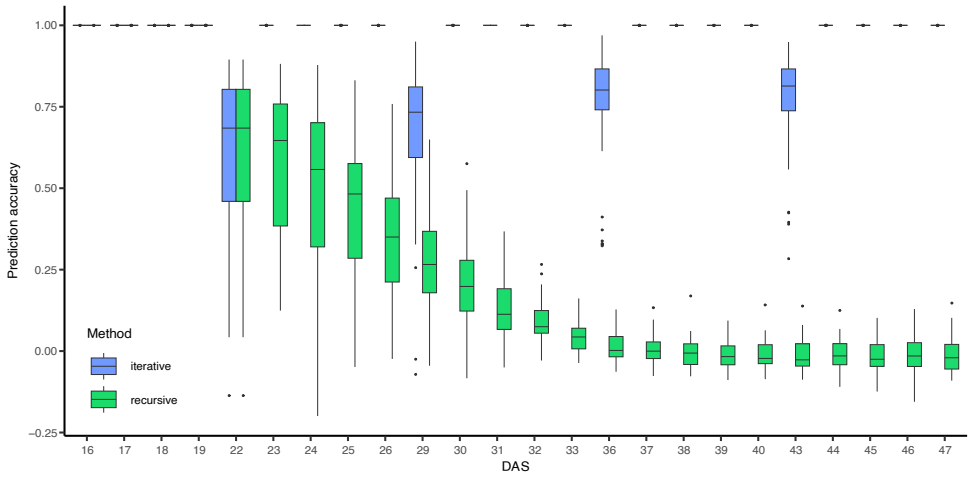
Peer review information *Nature Plants* thanks Osval A. Montesinos-López and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41477-025-01986-y



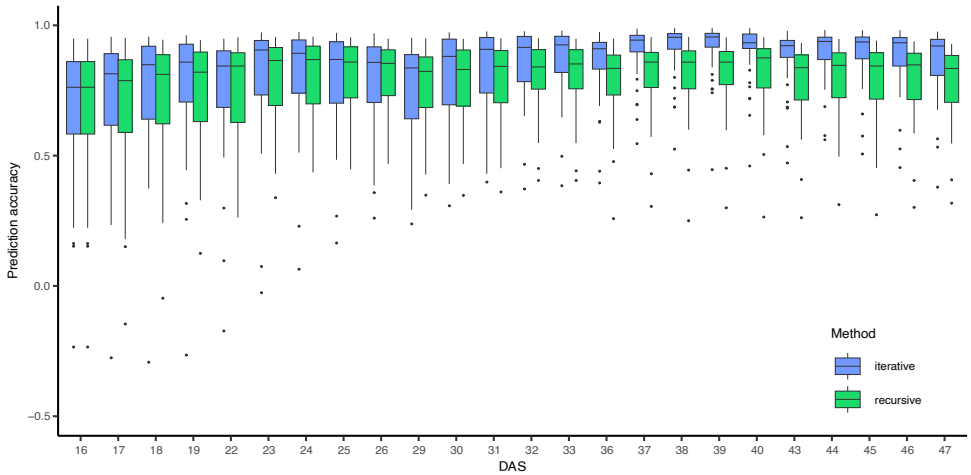
Extended Data Fig. 1 | Time-resolved dynamics of five selected traits from the maize HTP data set. Expanded version of Fig. 3b, showing time-resolved, scaled mean trait values across all lines, depicting the different dynamics of traits with standard deviation depicted as shaded area.

https://doi.org/10.1038/s41477-025-01986-y



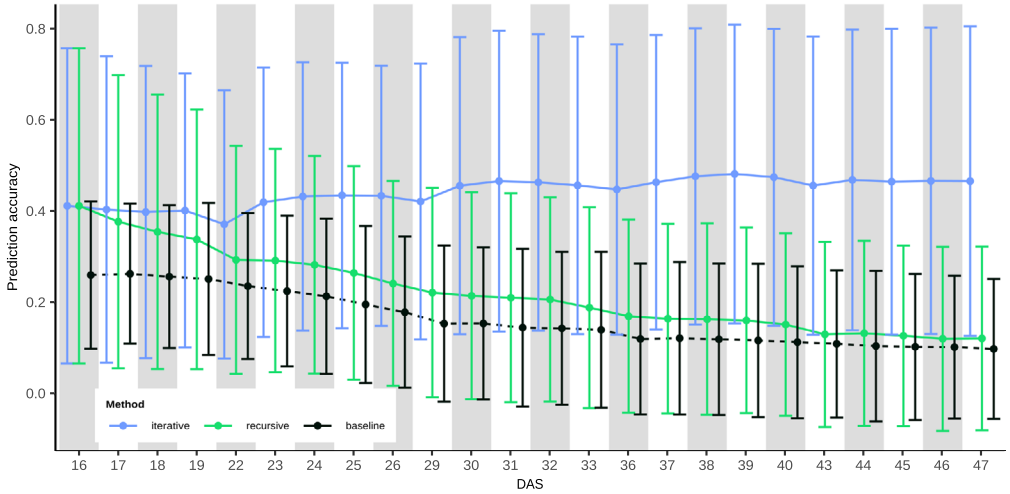
Extended Data Fig. 2 | Performance of DMD Algorithm 1 on the maize MAGIC data set. The operator **A** recreates training data near perfectly for the first week. However, after the first two-day gap, past time point 5, the recursive model (green) began to accrue errors and tended towards 0 over the subsequent time points. The performance of the iterative model was unaffected aside from the days immediately following the two-day gap (after $t = [5, 10, 15, 20]$). Prediction accuracy represents the mean accuracy across 50 traits using the operator **A** calculated based on Algorithm 1, similar to that presented in Fig. 1. Horizontal lines denote the median, boxes indicate the interquartile range (IQR), whiskers indicate the extended range of 1.5× IQR.

https://doi.org/10.1038/s41477-025-01986-y



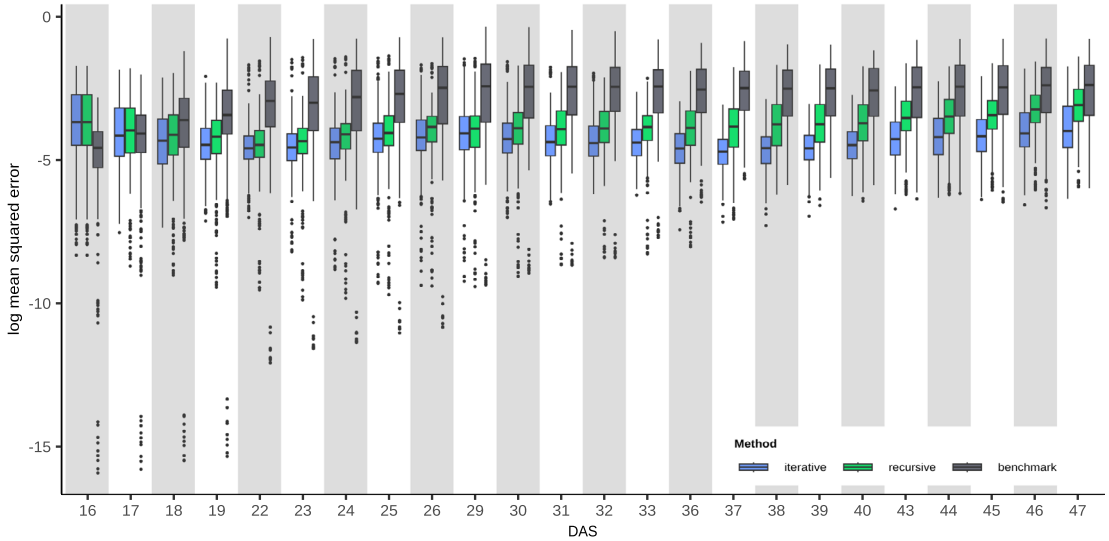
Extended Data Fig. 3 | Performance of DMD Algorithm 2 on the maize MAGIC data set. Algorithm 2 filters noise in the data and increases the prediction accuracy for time points following the two-day gaps (after $t = [5, 10, 15, 20]$) using the recursive approach (green). Shown is the mean accuracy across 50 traits using the operator $\mathbf{A}_r$ calculated based on Algorithm 2 with the number of singular vectors, r, included in the truncated model fixed to a value of two. Horizontal lines denote the median, boxes indicate the interquartile range (IQR), whiskers indicate the extended range of 1.5× IQR.

https://doi.org/10.1038/s41477-025-01986-y



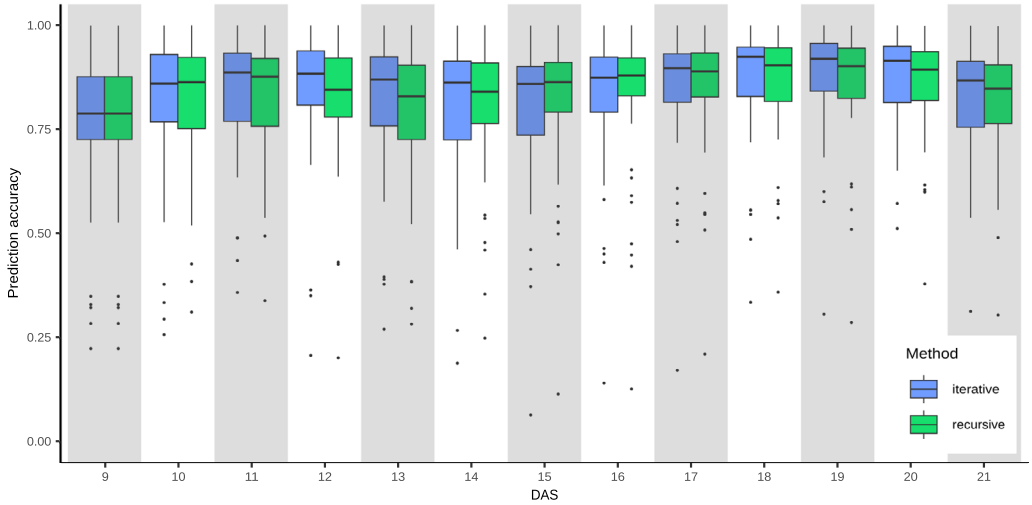
Extended Data Fig. 4 | Qualitative comparison of dynamicGP and baseline over all time points in a validation configuration for the maize MAGIC data set. The iterative version of dynamicGP (blue) yields mean prediction accuracies over all traits that are greater than the equivalent prediction accuracies from recursive dynamicGP (green) in 10 iterations of nested 5-fold cross validation with a validation step. Both versions of dynamicGP outperformed RR-BLUP baselines (black) at all time points. Lines represent mean model performance across all traits and error bars denote standard deviations.

https://doi.org/10.1038/s41477-025-01986-y



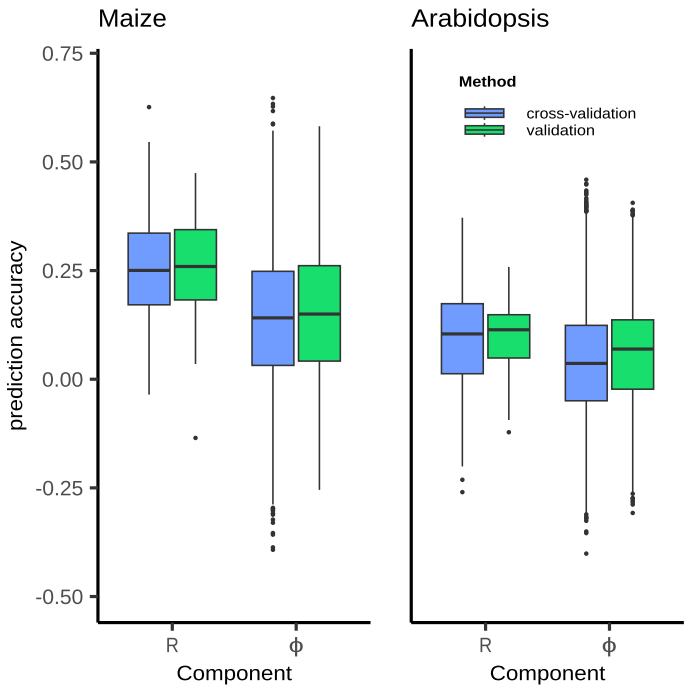
Extended Data Fig. 5 | Quantitative comparison of dynamicGP and baseline over all time points in a validation configuration for the MAGIC maize data set. The iterative version of dynamicGP (blue) yields mean mean squared error (MSE) over all traits that are lower than the equivalent MSEs in recursive dynamicGP (green) in 10 iterations of nested 5-fold cross validation with a validation step. The baseline models yield lower MSEs across all traits for the first two time points, however both versions of dynamicGP outperformed the baselines (gray) at all subsequent time points. Horizontal lines denote the median, boxes indicate the interquartile range (IQR), whiskers indicate the extended range of 1.5× IQR.

https://doi.org/10.1038/s41477-025-01986-y



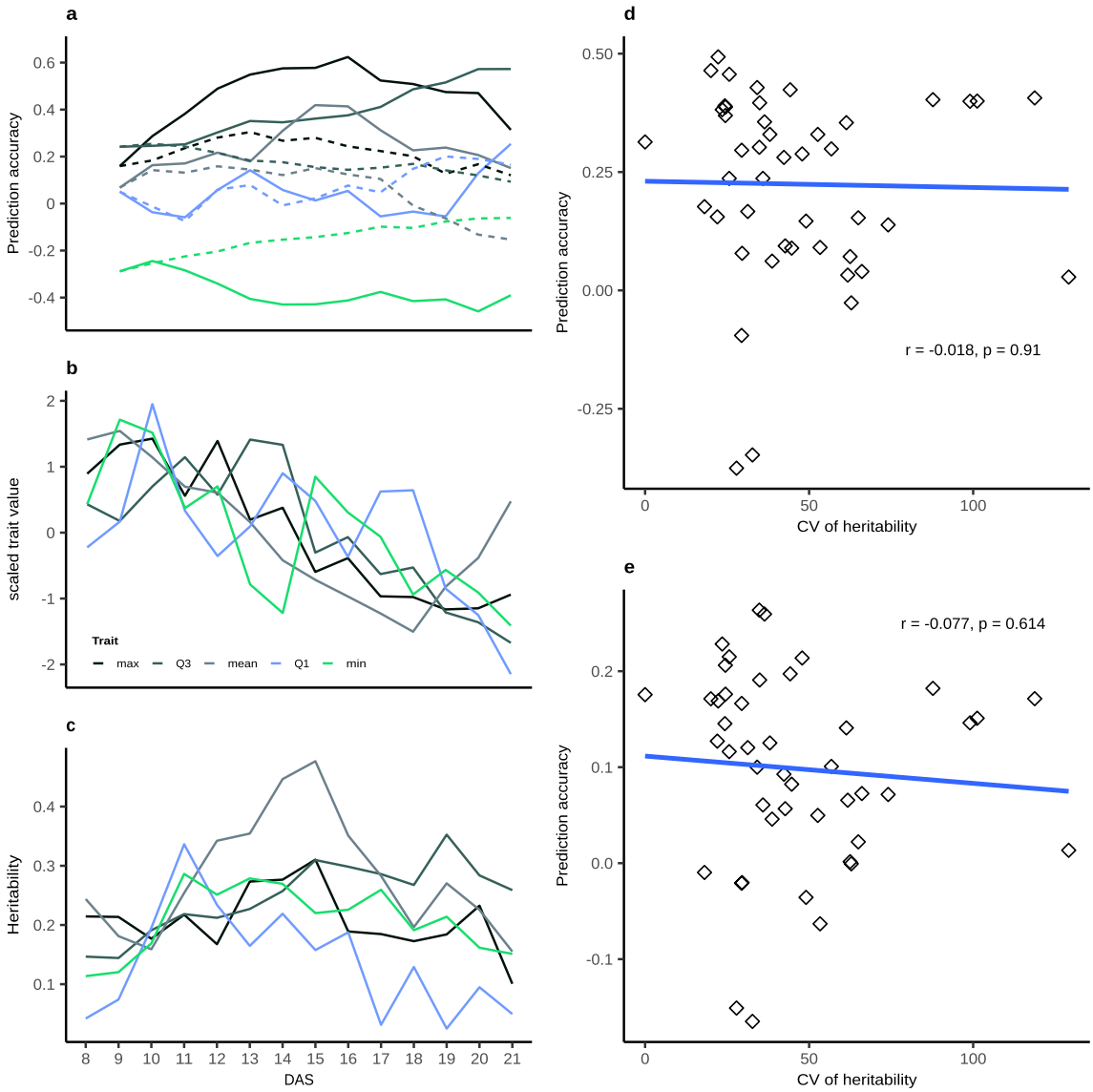
Extended Data Fig. 6 | Performance of DMD Algorithm 2 on data from an *A. thaliana* diversity panel. Shown is the mean accuracy across 45 traits using the operator $\mathbf{A}_r$ calculated based on Algorithm 2 with the number of singular vectors, r, included in the truncated model fixed to a value of two. Horizontal lines denote the median, boxes indicate the interquartile range (IQR), whiskers indicate the extended range of 1.5× IQR.

https://doi.org/10.1038/s41477-025-01986-y



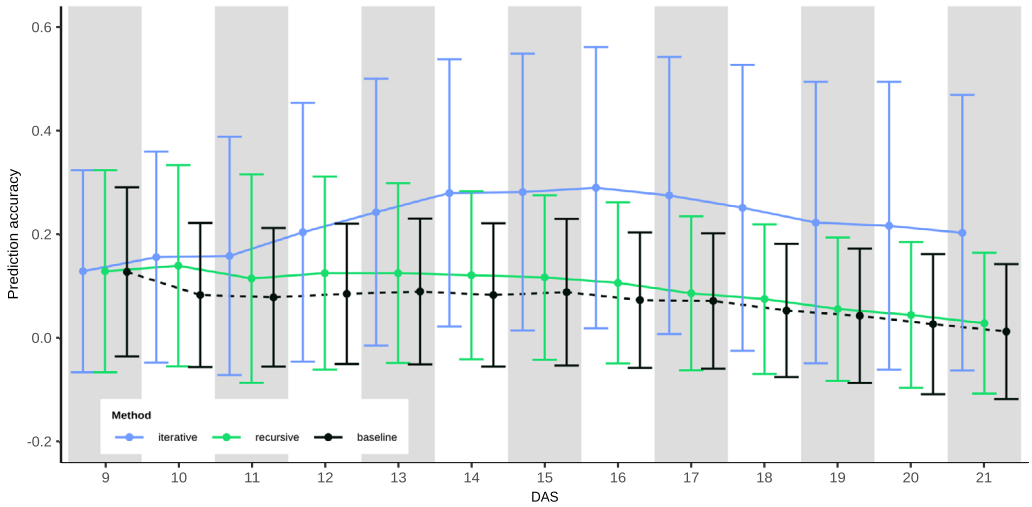
Extended Data Fig. 7 | Prediction accuracy of the elements of R and Φ in data sets of an *A. thaliana* diversity panel and a maize MAGIC population. Prediction accuracy of the elements of **R** and **Φ** after 10 iterations of 5-fold cross-validation with a validation step for Maize and Arabidopsis data sets. Horizontal lines denote the median, boxes indicate the interquartile range (IQR), whiskers indicate the extended range of 1.5× IQR.

https://doi.org/10.1038/s41477-025-01986-y



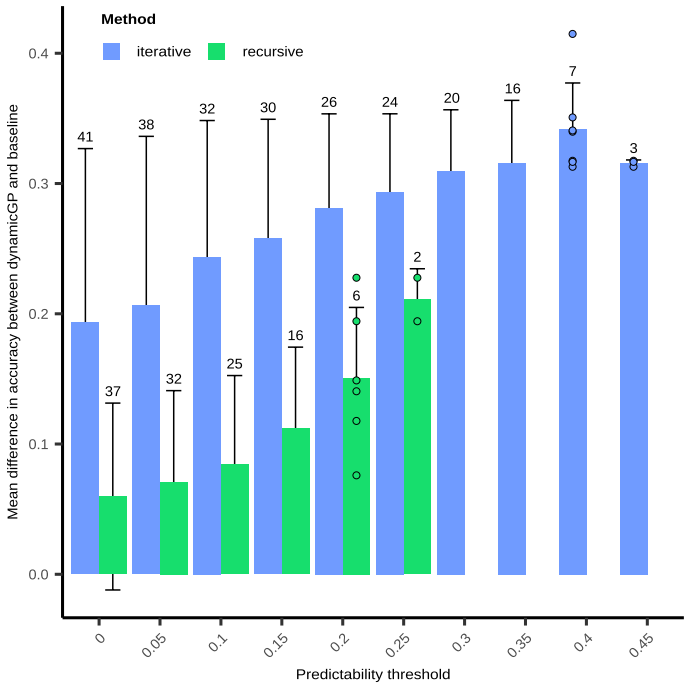
Extended Data Fig. 8 | Consistency of trait heritability over time less predictive of prediction accuracy on a data set of an *A. thaliana* diversity panel. **a**. Longitudinal prediction accuracy of representative traits corresponding to the maximum, minimum, mean, as well as the first and third interquartiles of all 45 traits across the 24 predicted time points (Supplementary Tables 3 and 4). The operator $\mathbf{A_r}$ was obtained from dynamicGP using 5-fold cross validation. The solid and dashed lines represent predictions from the iterative and recursive versions of dynamicGP, respectively. Colors for particular traits are maintained in panels **a** - **c**. **b**. Time-resolved, scaled mean trait values across all accessions, depicting the different dynamics of traits. c. Heritability of the five representative traits across time (p-value is derived from two-sided test from a sample of 50 points). **d** - **e**. Pearson correlation between the mean prediction accuracy of traits from iterative (**d**) and recursive (**e**) dynamicGP and the coefficient of variation of heritability estimates of traits across time.

https://doi.org/10.1038/s41477-025-01986-y



Extended Data Fig. 9 | Iterative dynamicGP outperformed recursive and baseline models over all time points on a data set of an *A. thaliana* diversity panel in a validation configuration. Despite lower prediction accuracy of **R** and **Φ**, non-zero prediction accuracy was achieved for both iterative and recursive dynamicGP in a validation configuration. The iterative version of dynamicGP (blue) yields mean prediction accuracies over all traits that are greater than the equivalent prediction accuracies recursive dynamicGP (green) in 10 iterations of nested 5-fold cross validation with a validation step. Both versions of dynamicGP outperformed RR-BLUP baselines (black) at all time points. Lines represent mean model performance across all traits and error bars denote standard deviations.

https://doi.org/10.1038/s41477-025-01986-y



Extended Data Fig. 10 | DynamicGP outperforms the baseline models on the *A. thaliana* data set. We determined the difference in performance of the iterative and recursive versions of dynamicGP and baseline RR-BLUP models, for traits whose predictability is above a threshold value, specified on the x-axis. The number of traits with predictability larger than a specified threshold value is indicated above the respective bar. The best predicted traits with dynamicGP exhibited the greatest difference from the baseline predictions. Bars indicate the mean (± SD) difference in trait prediction accuracies of traits which have mean prediction accuracies above the threshold.

nature portfolio

Corresponding author(s): Zoran Nikoloski

Last updated by author(s): Feb 27, 2025

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	Data was collected using the automated high-throughput phenotyping (HTP) facility for large plants at the Leibniz Institute of Plant Genetics and Crop Plant Research in Gatersleben, Germany, using a proprietary software.
Data analysis	Code is available in the form of an R package at https://github.com/dobby978/dynamicGP, available also at Zenodo XX.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data availability (DOIs) are described in the Data Availability Statement.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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	A common standard in genomic prediction studies is 100 iterations of 5-fold cross validation. In our preliminary investigations we determined that there was little difference between the results of 100 iterations and 20 iterations. We proceeded with 20 iterations for the sake of computational efficiency.
Data exclusions	We used modularity clustering to select representative traits from a larger data set, thereby excluding ~90% of the data, however this data was highly ($r > 0.96$) correlated with the included data.
Replication	The results from the analysis pipeline are fully reporducible.
Randomization	Randomization of cross validations folds was performed in R using sample() with replacement to create a vector with length corresponding to the number of lines with entries indicating the fold membership
Blinding	Blinding was not performed given that the analysis involved relatively simple comparisons.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*

Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

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

Antibodies

Antibodies used — *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation — *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s) — *State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*

Authentication — *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination — *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See ICLAC register) — *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance — *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*

Specimen deposition — *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods — *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals — *For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*

Wild animals — *Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Reporting on sex — *Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall*

	numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks — *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes — None.

Authentication — *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication. — *For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*

Files in database submission — *Provide a list of all files available in the database submission.*

Genome browser session
(e.g. UCSC) — *Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates — *Describe the experimental replicates, specifying number, type and replicate agreement.*

Sequencing depth — *Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*

Antibodies — *Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Peak calling parameters — *Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality — *Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*

Software — *Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation — *Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*

Instrument — *Identify the instrument used for data collection, specifying make and model number.*

Software — *Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See Eklund et al. 2016)

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☒ Functional and/or effective connectivity
☐	☒ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	We used Pearson correlation between the true and predicted trait values as an assessment of prediction accuracy, and as a method to create a correlation network between the traits, with edges corresponding to correlations of $r > 0.96$.
Graph analysis	We used graph analysis to cluster and select representative traits from a large set of phenomic traits. Nodes corresponded to traits and edges corresponded to inter-trait correlations above 0.96. Graphs were unweighted. Leading eigenvalue clustering, as implemented in R package "igraph" was used to determine clusters.
Multivariate modeling and predictive analysis	Ridge Regression Best Linear Unbiased Prediction (RR-BLUP) was used to predict 1) phenotypic traits and 2) elements of the matrices used in our method, from genomic data. No feature extraction or dimensionality reduction was performed. Models were trained and tested in 20 iterations of 5-fold cross validation. Dynamic Mode Decomposition was used to determine a best fit linear operator matrix which was then used to make predictions about future time points based on prior time points. Predictive performance was assessed with Pearson correlation between true and predicted values.