

RESEARCH

Open Access



Non-linear regression models for time to flowering in wild chickpea combine genetic and climatic factors

Konstantin Kozlov¹, Anupam Singh², Jens Berger³, Eric Bishop-von Wettberg⁴, Abdullah Kahraman⁷, Abdulkadir Aydogan⁶, Douglas Cook⁵, Sergey Nuzhdin² and Maria Samsonova^{1*}

From 11th International Multiconference "Bioinformatics of Genome Regulation and Structure\Systems Biology" - BGRS\SB-2018
Novosibirsk, Russia. 20-25 August 2018

Abstract

Background: Accurate prediction of crop flowering time is required for reaching maximal farm efficiency. Several models developed to accomplish this goal are based on deep knowledge of plant phenology, requiring large investment for every individual crop or new variety. Mathematical modeling can be used to make better use of more shallow data and to extract information from it with higher efficiency. Cultivars of chickpea, *Cicer arietanum*, are currently being improved by introgressing wild *C. reticulatum* biodiversity with very different flowering time requirements. More understanding is required for how flowering time will depend on environmental conditions in these cultivars developed by introgression of wild alleles.

Results: We built a novel model for flowering time of wild chickpeas collected at 21 different sites in Turkey and grown in 4 distinct environmental conditions over several different years and seasons. We propose a general approach, in which the analytic forms of dependence of flowering time on climatic parameters, their regression coefficients, and a set of predictors are inferred automatically by stochastic minimization of the deviation of the model output from data. By using a combination of Grammatical Evolution and Differential Evolution Entirely Parallel method, we have identified a model that reflects the influence of effects of day length, temperature, humidity and precipitation and has a coefficient of determination of $R^2 = 0.97$.

Conclusions: We used our model to test two important hypotheses. We propose that chickpea phenology may be strongly predicted by accession geographic origin, as well as local environmental conditions at the site of growth. Indeed, the site of origin-by-growth environment interaction accounts for about 14.7% of variation in time period from sowing to flowering. Secondly, as the adaptation to specific environments is blueprinted in genomes, the effects of genes on flowering time may be conditioned on environmental factors. Genotype-by-environment interaction accounts for about 17.2% of overall variation in flowering time. We also identified several genomic markers associated with different reactions to climatic factor changes. Our methodology is general and can be further applied to extend existing crop models, especially when phenological information is limited.

Keywords: Wild chickpea, Model, Climatic factors, GWAS

*Correspondence: m.g.samsonova@gmail.com

¹Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya, 195251 St. Petersburg, Russia

Full list of author information is available at the end of the article



Background

Chickpea (*Cicer arietinum* L.), is the second most cultivated grain legume crop, grown in more than 50 countries of the world (ICARDA). Chickpea, which was originally domesticated in Southeastern Turkey, has been adapted to various environmental and climatic conditions across the globe from subtropical conditions in South Asia and East Africa to Northern regions of temperate North America. The time duration for chickpea to reach its reproductive phase is often limited by changing temperatures, rainfall pattern, daylength or competition for use of land by other crops in rotation [1, 2]. In Mediterranean and temperate regions, chickpeas are sown in spring where the day length and temperature increase towards the reproductive period, while in subtropical regions (Center and Southern India, Ethiopia, Queensland Australia) it is planted in the start of the dry season after the monsoonal rainy season when daylengths tend to be shorter and temperatures cooler. In the more temperate northern parts of India, the reproductive phase of spring-planted chickpea coincides with decreasing temperature and day length, whereas in the southern and the central parts of the country it falls within terminal drought (the end of the dry season) [3, 4]. Hence, chickpea breeding has focused on developing varieties differing in their growth duration to be able to adapt to different latitudes and sowing regimes [3, 5–7]. To achieve consistent yield, crop duration must closely match the available growing season [8]. Chickpea cultivars and landraces become increasingly temperature responsive as from the Mediterranean through northern, central and southern India, because these disparate origins have selected for contrasting phenological regulators [3]. This information is invaluable for modeling crop performance. For example, Vadez et al. (2012, 2013) [9–11] considered climatic factors like expected rainfall to predict performance of chickpeas in different geographical locations.

Several successful plant models like SSM [10, 12], DSSAT [13–17], APSIM [18] and others [19, 20] have been developed for legumes. These models use differential equations to describe biophysical and biochemical processes like photosynthesis, water uptake etc. and account for impact of genotype, soil, weather and economic factors. The influence of weather conditions is assessed using concepts like Heat Unit Index (HUI) [20], Crop Heat Units (CHI), Degree Days (DD), Biological Days (BD) [9] – all of them quantitatively characterizing the rate of progression to the next phenological phase on a daily basis. Both DD and BD could depend on temperature, water content and photoperiod. This formalism was applied to develop individual models for important crops. For example, DSSAT was used to simulate growth and yield in soybean [21] and chickpea [22] among several other crops [23–26]. For more than three decades these models have been applied

in research projects of different countries. The SSM model was successfully tested using independent data from a wide range of growth [10, 12] and environmental conditions including Iran [27] and water deficit in India [11]. Considerable manipulations are required to adapt the DSSAT model to new environments and cultivars [28–32], limiting the utilization of these models. As varieties are constantly changing because of new releases that can cope with emerging pathogens and pests, as well as shifting consumer demands, the need for flexible models that can adjust to new varieties is high.

In an era of rapidly advancing genomic technologies and approaches, updated modeling approaches that can be tailored to genotype-specific effects are essential. Next generation sequencing and high throughput genotyping lead to identification of thousands of molecular markers (SSR, SNP, STMS, ESTs, CISP, DArT) [33] making it possible to construct chickpea genetic maps [34, 35] and ultimately to dissect the effect of different loci on key traits like flowering time. A combination of Sanger, 454/FLX and Illumina reads have been used to generate in transcriptome and genome assemblies for chickpea [34, 36–38].

Due to these advances in sequencing technologies and data acquisition, the genome-wide association study (GWAS) has become an important approach to understand the genetics of natural variation and traits of agricultural importance. Recent examples of GWAS in agriculturally important plants include identification of photoperiodic flowering time genes in sorghum [33], frost tolerance genes in barley (*Hordeum vulgare* L.) [34]; leaf architecture [35] and resistance to southern leaf blight genes in maize [36] as well as several agricultural traits in rice [37], to name a few. To extend GWAS to the analysis of genotype-by-environment ($G \times E$) interactions bioclimatic variables can be used as a GWAS phenotype. Association between bioclimatic variables at a site of an accession's origin and SNPs can indicate climatic adaptation [39]. While GWAS is a good method to identify genomic regions associated with important traits, typical GWAS designs require controlled planting of replicated accessions. This can quickly become logistically daunting and expensive across many sites.

Crop models may complement GWAS approaches by accounting for the influence of environmental factors [16]. However the models developed in the pre-genomic era considered genotype influence at best as a set of given “genetic coefficients” that do not correspond to actual genes [40]. Therefore these models were unable to simulate gene-by-environment interactions, thereby limiting their utility in predicting phenological characteristics of cultivars across different geographical locations and genotypes [9]. Mathematical models and tools that combine genetic and climate data to predict agronomic traits will

greatly benefit breeders by simulating the performance of any given well-characterized genotype in any given well-characterized environment [41, 42].

While some of the current crop models consider the influence of local environmental conditions and others global climate changes for locally grown varieties, here, we built a new model using the flowering time of two species of wild chickpeas (*Cicer reticulatum* L. and *C. echinospermum*) collected at 21 different sites in Turkey and grown in 4 distinct environmental conditions. We further use our model to test two important hypotheses. Firstly, we propose that besides local environmental factors, chickpea phenology may be strongly predicted by accession geographic origin. Secondly, as the adaptation to specific environments is blueprinted in genomes, the effects of genes on flowering time may be conditioned on environmental factors. We check these hypotheses by statistical modeling of chickpea responses to climate change scenarios conditional on geographic site of origin and genotype.

Materials and methods

Dataset of wild chickpea accessions

The dataset consists of wild chickpea (*Cicer reticulatum* L. and *Cicer echinospermum*). Accessions were collected at 21 sites in five regions in Turkey (see Additional file 1: Table S1) by von Wettberg et al. [43]. These wild accessions were planted in climatically distinct sites in Turkey (Sanliurfa and Ankara, autumn and spring-sowing) and Australia (Floreat, near Perth, WA and Mt.Barker, WA). Being grown in contrasting environments the phenotype data on time to flowering is highly diverse. The distribution of time to flowering for the whole dataset is shown in Fig. 1. The time to flowering ranges from 64 to 221 day. Details on the phenotyping experiment and its subsequent analysis will be presented in the future manuscripts (Berger, J.: Analysis of phenotyping of wild chickpea in diverse environments, in preparation).

Climatic data was downloaded from NNDC Climate Data on-line [44]. The summary of agroclimatic factors as well as results of testing their correlation with flowering time are given in Additional file 1: Table S2 and S3, respectively.

A companion paper studying the genetic association of flowering time in one of the wild chickpeas (*Cicer reticulatum* L.) has identified six suggestive polymorphic sites associated with flowering time (Singh, A.: Genome-wide association studies in wild chickpea, in preparation). These SNPs were identified as the best SNPs after running a mixed linear model (MLM) in TASSEL, which associated flowering time (phenotype) with the genotypes using site/year/season as a factor to account for their effect on phenotype. Additional file 1: Table S6 presents number of times the reference allele for a SNP

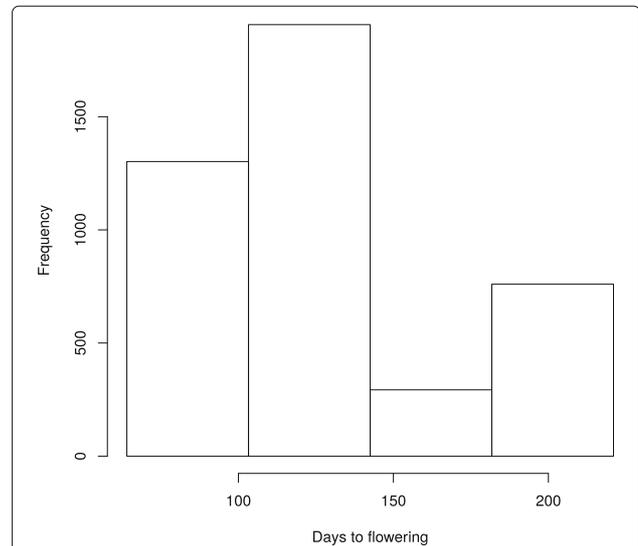


Fig. 1 Distribution of time to flowering for the whole dataset. The range for time to flowering is from 64 to 221 day

associated with flowering time is present in plant genotypes. To access genotype-environment interactions we group plants into 18 groups – one for each alternative (ALT) and reference (REF) allele combination (ALT/ALT, REF/ALT and REF/REF) – for each SNP and built a model (1) for each group.

Regression model for time to flowering

We model a time period from sowing to flowering as a linear combination of N control functions F_n , $n = 0, \dots, N-1$ of agroclimatic factors. Thus, the model takes the form (1)

$$y_i = \beta_0 + \sum_{n=0}^{N-1} \beta_{n+1} \cdot F_n(\mathbf{X}_i) + \varepsilon_i \quad i = 0, \dots, I-1 \quad (1)$$

where y_i is modeled phenotype (time from sowing to flowering) for each plant i from a group of the size I , β_n are coefficients, $n = 0, \dots, N$, that are to be found to minimize the discrepancy between data and model, \mathbf{X}_i is a vector of agroclimatic factors and ε_i is a standard error. The number of coefficients is $N+1$ because β_0 is an intercept.

In comparison with previous models in our approach control functions F_n are automatically composed in analytic form from the expressions of climatic factors. Thus, a wider range of non-linear dependencies between the phenotype and factors is explored (see “Analytic form of control function” on page 4).

To study the adaptation to environment of origin we represent collection sites as $L = 21$ binary variables, where $l = 1, \dots, L$ enumerates locations: Baristepe1, Baristepe2, Baristepe3, Beslever, Cermik, Cudi, Cudi2, Dereici, Destek, Egil, Gunasan, Kalkan, Karabahece, Kayatepe, Kesentas, Ortanca, Oyali, Sarikaya, Savur1, Sirnak1, Siv-Diyar (see column 1 in Additional file 1:

Table S1). For each plant enumerated with $i = 0, \dots, I - 1$ one of the L variables d_i^l takes the value '1' to indicate collection site and others are '0'. The interaction between control function and location is modeled by an additional term in the regression function that has the form of a weighted sum of $N \cdot L$ pairwise products of control functions F_n and each binary site variable d_i^l .

Consequently, a model with information about a collection site takes the form (2).

$$y_i = \beta_0 + \sum_{n=0}^{N-1} \beta_{n+1} \cdot F_n(\mathbf{X}_i) + \sum_{n=0}^{N-1} \sum_{l=1}^L \zeta_{l \cdot N+n} \cdot F_n(\mathbf{X}_i) \cdot d_i^l + \varepsilon_i \tag{2}$$

where in addition to notations used in (1) new regression coefficients $\zeta_{l \cdot N+n}$ define the influence of function F_n of climatic factors on phenotype of plants collected at site l so that condition $\zeta_{l \cdot N+n} \neq 0$ points on plant adaptation to the site. As a result, this model makes it possible to regress a range of climatic variables describing the phenotyping site (e.g. day length, temperature, precipitation etc.) independently for each of our 21 collection sites.

We denote K number of SNP and $J = 3$ combinations of alternative (ALT) and reference (REF) alleles ALT/ALT, ALT/REF and REF/REF by 0, 1, and 2, respectively. Then to include GWAS results into the model we define $J \cdot K$ groups of plants so that members of the same group have the same combination of alleles in one of the SNP positions. Thus we define a matrix D with the number of rows equal to the number of plants I and $J \cdot K$ columns. Then, the elements of matrix D are defined by (3). Thus, the form of the regression function adapts to the allele combination of a plant by changing the weights of control functions.

$$d_i^{3k+j} = \begin{cases} 1 & \text{if in plant } i \text{ the combination for SNP } k \text{ is } j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Consequently, a model with genetic information takes the form (4).

$$y_i = \beta_0 + \sum_{n=0}^{N-1} \beta_{n+1} \cdot F_n(\mathbf{X}_i) + \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} \rho_{(3k+j)N+n} \cdot F_n(\mathbf{X}_i) \cdot d_i^{3k+j} + \varepsilon_i \tag{4}$$

where in addition to notations used in (1) new regression coefficients $\rho_{(3k+j)N+n}$, define the effect of genotype-by-climatic factor interaction.

Analytic form of control function

In previous studies different forms of dependencies between phenotype and climatic factors have been considered [45–50]. For example, “segmented”, “beta”, “quadratic” and “dent-like” functions were considered in [10]. A product of quadratic functions of day length

and mean temperature was used in iterative regression analysis (IRA) [51] to characterize a developmental speed per day. An interphase speed was calculated as a product of the effects of day length, water deficit and temperature in [52].

We propose a more general approach, in which the analytic form of a control function together with regression coefficients and a set of predictors are inferred automatically by stochastic minimization of the deviation of the model output from data. We use a combination of Grammatical Evolution (GE) [53, 54], LASSO [55] and Differential Evolution Entirely Parallel (DEEP) [56, 57] method to recover analytic form of F_n , find regression coefficients and determine the set of climatic factors, respectively [58]. Differential Evolution was proposed by Storn and Price in 1995 [59] as a heuristic stochastic optimization method. DEEP was developed by us for application in the field of bioinformatics [56]. It includes several recently proposed enhancements [57, 60]. More details can be found in Additional file 1: Section S5.

In GE, the analytic function form is built by decoding the sequence called “word” of L integers called codons. Decoding is performed according to simple rules of substitution that establish a correspondence between codons and either an elementary arithmetic operation: '+', '-', '*', '/', or expression: X , $(X - \text{Const})$, $1/(X - \text{Const})$, where X is a name of a predictor and Const is some constant number (see Additional file 1: Section S2). To make estimation of regression coefficients with LASSO method reliable we performed 4-fold cross-validation at this stage so that the model was build using 75% of samples for training and the rest 25% was used for evaluation of the model.

Statistical tests

We used standard statistical techniques for hypothesis testing implemented in R system for statistical computing [61]. We used multiple-way analysis of variance (MANOVA) with the Pillai test statistic [62] and ANOVA with the Fisher test statistic to check for significance in the difference of effects of climatic factors on phenotype between locations and genotypes. For pairwise comparison of the influence of climatic factors on phenotype between genotypes and locations we applied the Wilcoxon-Mann-Whitney test.

The Spearman’s rank correlation was used to estimate correlation between allele frequencies and climatic factors at primary collection sites (geographic sites of origin).

Software tools

Although a few Grammatical Evolution (GE) implementations are freely available (see e.g. [54, 63]) they either lack a specific set of expressions or show low performance in experimental runs due to interpreted language (data

not shown). Consequently a decision was made to implement GE in C++ using Armadillo [64], mlpack [65], HDF5 [66], HighFive [67] and Qt [68] as these packages provide efficient matrix operations, the LASSO method, data input-output and utility functions, respectively. The code is open-source on GitLab [69].

TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) [70] was developed in Java, and is compatible with multiple operating systems (Windows, Linux and Mac OS). TASSEL can implement several different GWAS models like general linear model (GLM) and MLM using a GUI or command line version of the software.

Results

We first performed ANOVA test to check for differences in mean time to flowering between accessions collected at different sites in Turkey to demonstrate that flowering time is an adaptive trait in chickpea. The difference in means was significant with criterion value $F = 2.003$ and $p = 0.005 < 0.05$.

Model with interactions between climatic factors and locations

Next, to estimate the effect of interaction between climatic factors at phenotyping sites and sampling locations in Turkey on flowering time we built a model (2). A series of numerical experiments were performed, as several runs are needed to obtain a reliable solution with stochastic optimization. By several trial-and-error attempts (data not shown) it was established that the number of control functions $N = 12$ and the length of the “word” $L = 5$ were the best parameters for the model. The population size for DEEP was set to 500.

We obtained several solutions with coefficient of determination > 0.85 and different analytic forms of the control functions (data not shown). We selected the model (5) as it reflects the influence of effects of day length, temperature, humidity and precipitation in the phenotyping environment and has coefficient of determination $R^2 = 0.97$.

$$\begin{aligned} \text{TTF} = & 59.49 + 74.95D_{x10}^{\min} + 19.83/(T_{x5}^{\min} - 0.03) - 1.98D_{x10}^{\text{mean}} \\ & - 53.18(D_{x50}^{\text{mean}} + 1/(U_{x10-15}^{\text{mean}} - 23.31)) - 13.04D_{x10-15}^{\text{mean}} \\ & - (0.05 \cdot \text{Baristepe1} + 0.12 \cdot \text{Baristepe3} + 0.29 \cdot \text{Beslever} \\ & + 0.31 \cdot \text{Dereici} + 0.45 \cdot \text{Kayatepe} + 0.03 \cdot \text{Kesentas} \\ & + 0.74 \cdot \text{Siv - Diyar} + 0.01 \cdot \text{Sarıkaya} + 0.10 \cdot \text{Sirnak1} \\ & + 0.20 \cdot \text{Oyali}) \cdot (D_{x50}^{\text{mean}} + 1/(U_{x10-15}^{\text{mean}} - 23.31)) \\ & + (0.03 \cdot \text{Cudi2} + 0.16 \cdot \text{Destek} + 0.09 \cdot \text{Gunasan} \\ & + 0.46 \cdot \text{Kesentas} + 0.43 \cdot \text{Oyali} \\ & + 0.28 \cdot \text{Sirnak1}) \cdot D_{x10}^{\min} \\ & + (0.41 \cdot \text{Cudi} + 0.05 \cdot \text{Karabahce} - 0.003 \cdot \text{Kesentas} \\ & + 0.02 \cdot \text{Oyali} - 0.12 \cdot \text{Sirnak1}) \cdot D_{x10-15}^{\text{mean}}, \end{aligned} \quad (5)$$

where D_{x10}^{\min} , D_{x50}^{mean} , D_{x10-15}^{mean} denote minimum day length over 10 days after sowing, mean day length over a period of 50 days and from 10 to 15 days after sowing, respectively, T_{x5}^{\min} denotes minimum temperature over 5 days after sowing, U_{x10-15}^{mean} denotes mean relative humidity over an interval from 10 to 15 days after sowing and D_{x10}^{mean} denotes average precipitation over 10 days after sowing.

The analysis of relative difference in the sum of squares for a model with and without a term describing an interaction between climatic factor at the phenotyping site and the accession geographic site of origin allows us to conclude that sampling collection site-by-phenotyping environment interaction accounts for about 14.7% of variation in time period from sowing to flowering.

We found that day length-by-collection site interaction is important for locations Baristepe3, Cudi, Cudi2, Destek, Gunasan, Karabahce, and both day length and humidity-by-collection site interaction are important for Baristepe1, Beslever, Dereici, Kayatepe, Kesentas, Oyali, Siv-Diyar, Sarıkaya, and Sirnak1 sampling sites. There were no interactions between climatic factors and collection sites in Baristepe2, Cermik, Egil, Kalkan, Ortanca and Savur1.

Basic flowering time models for locations

To analyze how climatic factors at phenotyping sites affect flowering time of plants collected at different locations we built basic models (1) for groups of plants sampled at each location separately. We present selected models with the highest coefficients of determination (R^2) between simulated and observed flowering time for each group in Additional file 1: Section S3. The distributions of time to flowering for these groups are presented in Additional file 1: Figure S2. Due to the stochastic nature of the procedure ten runs were performed with the same algorithmic parameters using different seeds for the random number generator to obtain an ensemble of models. Various factors and their combinations were selected as predictors by stochastic optimization.

Consequently, the effect of phenotyping environment day length, temperature, precipitation, humidity and their pairwise combinations on flowering time for plant groups was estimated with a coefficient of determination averaged over the ensemble of models, taking into account only terms dependent on the factor in question. The resulting coefficient values for each factor and factor combination are presented for all collection sites in Additional file 1: Figure S3–S7.

We compared the mean effect of climatic factors or factor combinations between accessions from different locations in Turkey with multiple-way ANOVA (MANOVA) using the values of coefficients of determination obtained in model runs as dependent variables and location as an independent variable. A Pillai statistics of 1.697 with $p = 0.037 < 0.05$ confirmed the statistical significance

of differences in mean influences of factors and their combinations on phenotype between locations.

Further, we applied one-way ANOVA to test the difference in effects on flowering time between accessions from different locations for each climatic factor or factor combination individually. Temperature, precipitation and their combination showed significant differences in the means of coefficient of determination values with $F = 3.617, p = 1.806e - 06$ and $F = 2.233, p = 0.003$ and $F = 2.038, p = 0.008$ respectively.

Analysis of the climatic factor effect on phenotype

We continue our analysis of effects of climatic factors on flowering time for accessions from different locations with a pair-wise comparison method. Firstly, the direction and extent of each factor influence on phenotype was estimated as a finite difference approximation of the partial derivative of a regression function (1) in respect to the factor. Figure 2 presents the box plots of factor influence estimators calculated for model ensembles and for each location.

It is evident that both effects of day length and temperature on flowering time are location-dependent. For accessions collected at some locations increasing day length (e.g. Egil) or temperature (e.g. Ortanca) speeds up the rate of flowering, while at other locations the response to

these factors is reversed (e.g. Kesentas). Surprisingly there was a consistent effect of precipitation across all accessions whereby higher precipitation reduced the time to flowering. This result is consistent with negative correlation between precipitation measures and time to flowering (see Additional file 1: Table S3). In comparison with precipitation the influence of humidity is opposite for most locations: the flowering time increases with rise of humidity. The influences of factor combinations are comparatively negligible.

Next, we compared the means of estimators of a factor’s influence on phenotype for location pairs with a Wilcoxon-Mann-Whitney test. Statistically significant differences in means between locations pairs are presented in Figs. 3, 4, 5, and 6 for day length, temperature, precipitation and humidity, respectively.

Flowering time model with climatic factor-by-genotype interaction

Different genotypes may react differently to climatic factors. Here we check this hypothesis using the flowering time model (4) with the interaction term between climatic factors and genotype. We identified six SNPs associated with flowering time (see Additional file 1: Table S5). Here we subdivided all the plants into 18 groups, each containing similar allele combination at one of six polymorphic

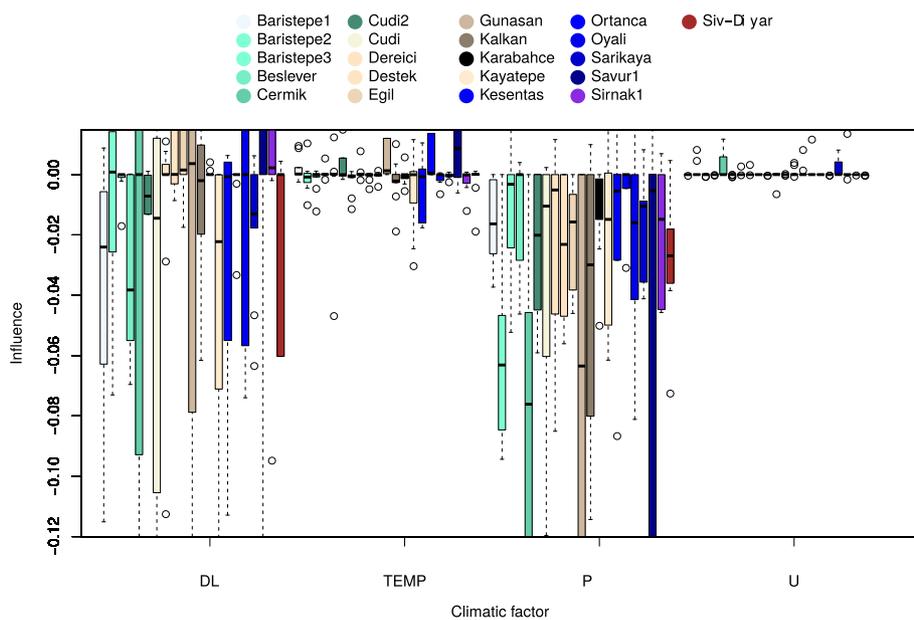


Fig. 2 Analysis of climatic factor effects on phenotype. Box plots of climatic factor influence estimators calculated for model ensembles and for each location as a finite difference approximation of the partial derivative of a regression function (1) in respect to the factor. Each box covers two quantiles from 25 to 75% of influence’s variation with a horizontal line at median value of the estimated influence. Empty circles represent outliers. Boxes located higher than zero mark on vertical axis represent a positive influence of a factor on flowering time. In this case increasing the factor speeds up flowering. Other boxes represent an opposite case. “DL”, “TEMP”, “P” and “U” correspond to factors related to day length, temperature, precipitation and humidity, respectively

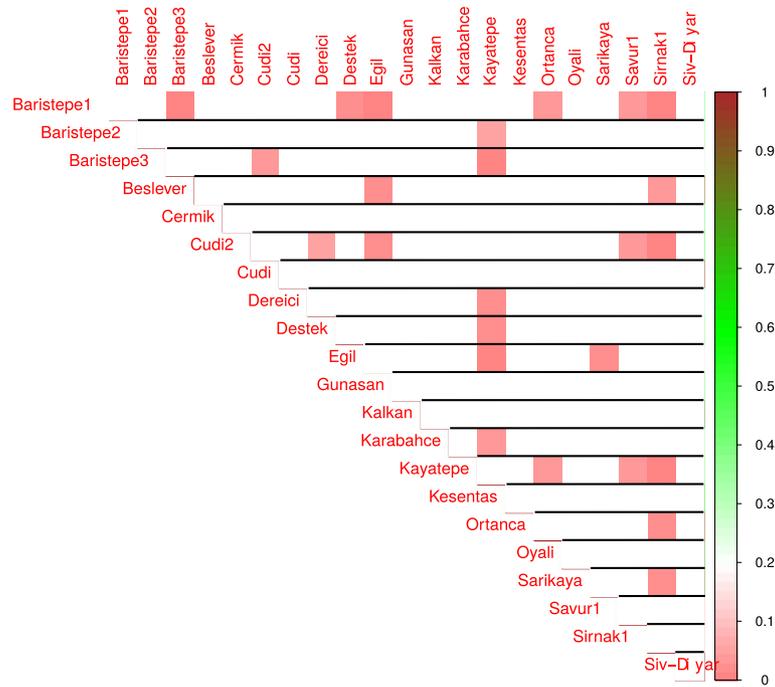


Fig. 3 Results of pair-wise comparisons of day length influence on flowering time. Mann-Whitney-Wilcoxon test was applied to compare the means of day length influence estimators for locations. Statistically significant differences in means are shown as red color gradation, cells with statistically non-significant comparisons are left blank

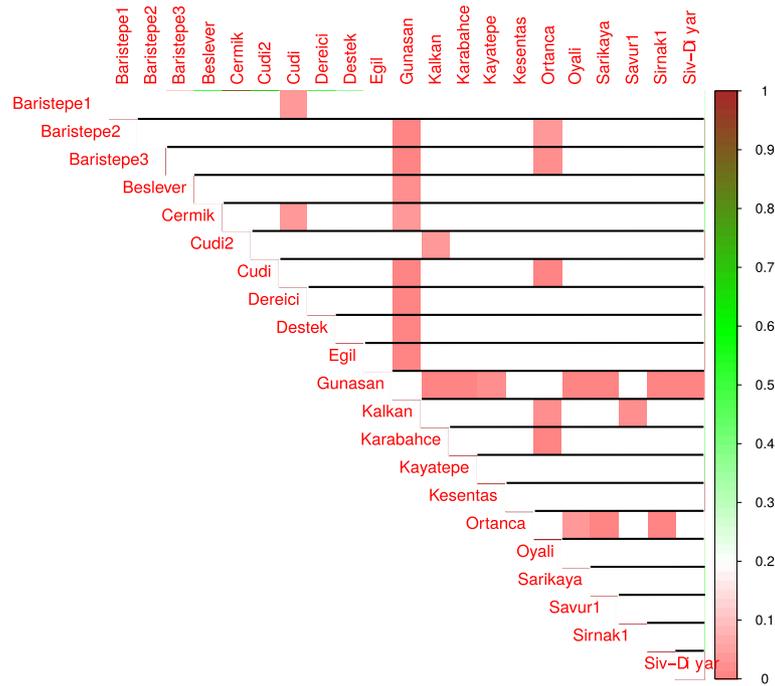


Fig. 4 Results of pair-wise comparisons of temperature influence on flowering time. Mann-Whitney-Wilcoxon test was applied to compare the means of temperature influence estimators for locations. Statistically significant differences in means are shown as red color gradation, cells with statistically non-significant comparisons are left blank

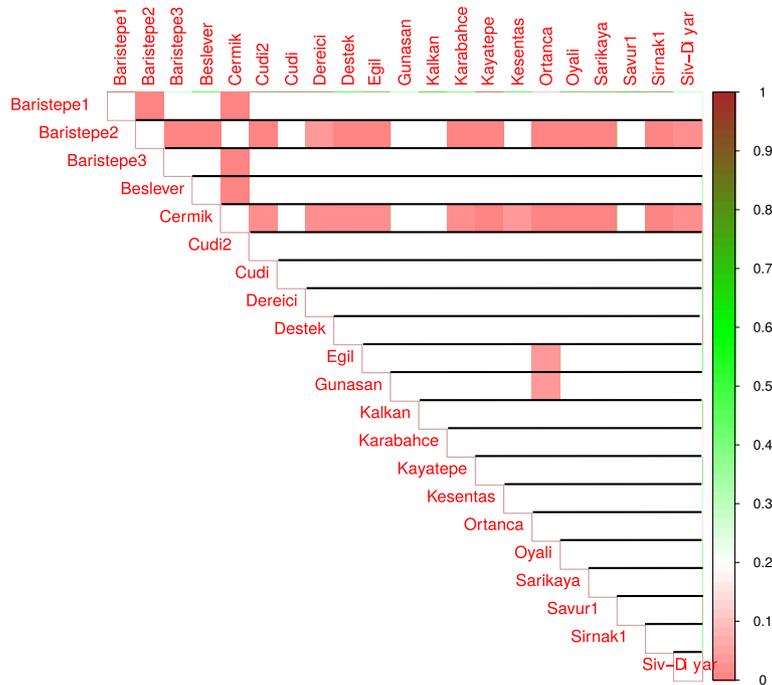


Fig. 5 Results of pair-wise comparisons of precipitation influence on flowering time. Mann-Whitney-Wilcoxon test was applied to compare the means of precipitation influence estimators for locations. Statistically significant differences in means are shown as red color gradation, cells with statistically non-significant comparisons are left blank

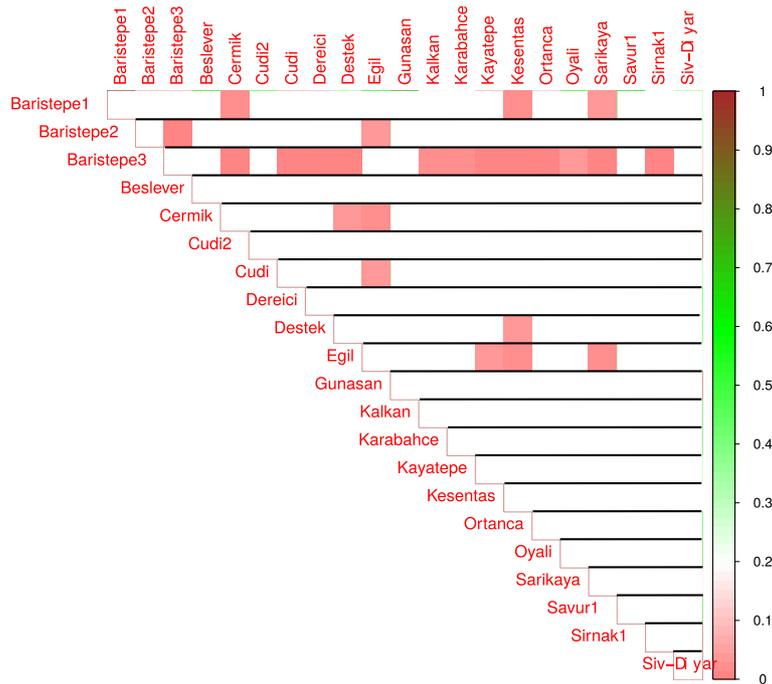


Fig. 6 Results of pair-wise comparisons of humidity influence on flowering time. Mann-Whitney-Wilcoxon test was applied to compare the means of humidity influence estimators for locations. Statistically significant differences in means are shown as red color gradation, cells with statistically non-significant comparisons are left blank

sites (see the “Regression model for time to flowering” section for more details). We further refer to these groups as SNP groups.

Ten runs were performed with the same algorithmic parameters but different seeds for random number generator. The model (6) with the best coefficient of determination $R^2 = 0.97$ was selected for further analysis.

$$\begin{aligned}
 \text{TTF} = & -5.71 \cdot T_{x5-10}^{\max} - 3.87 \cdot T_{x15-20}^{\max} - 0.39 \cdot (1/(D_{x60}^{\min} - 293.08) + T_{x10-15}^{\max}) \\
 & + 5.42 \cdot D_{x10-15}^{\text{sum}} / (D_{x15-20}^{\min} \cdot P_{x50}^{\text{mean}} + 1) \\
 & + 20.08 \cdot (T_{x5-10}^{\text{mean}} + (U_{x10-15}^{\text{mean}} - 0.004) / (T_{x5-10}^{\min} - 210.12)) \\
 & + (0.06 \cdot \text{snp5AA} + 0.49 \cdot \text{snp3RR}) \cdot T_{x5-10}^{\max} \\
 & + 0.05 \cdot \text{snp3AA} \cdot T_{x15-20}^{\max} \\
 & - (0.02 \cdot \text{snp1RR} + 0.002 \cdot \text{snp2AA} + 0.16 \cdot \text{snp2RR} \\
 & + 0.09 \cdot \text{snp3AA} + 0.50 \cdot \text{snp3RR} + 0.007 \cdot \text{snp4RR} \\
 & + 0.14 \cdot \text{snp5RR} + 0.04 \cdot \text{snp6RR}) \cdot (1/(D_{x60}^{\min} - 293.08) + T_{x10-15}^{\max}) \\
 & + 0.11 \cdot \text{snp4RR} \cdot D_{x10-15}^{\text{sum}} / (D_{x15-20}^{\min} \cdot P_{x50}^{\text{mean}} + 1) \\
 & + 0.54 \cdot \text{snp3AA} \cdot T_{x5-10}^{\text{mean}},
 \end{aligned} \tag{6}$$

where D_{x60}^{\min} , D_{x15-20}^{\min} denote minimum day length over 60 days after sowing and over a period from 15 to 20 day after sowing, respectively; D_{x10-15}^{sum} denotes sum of day lengths over a period from 10 to 15; T_{x15-20}^{\max} , T_{x10-15}^{\max} and T_{x5-10}^{\max} denote maximum temperatures over a periods from 15 to 20, from 10 to 15 and from 5 to 10 days after sowing, respectively; T_{x5-10}^{mean} and T_{x5-10}^{\min} denote mean and minimum temperatures over a period from 5 to 10 days after sowing, respectively; U_{x10-15}^{mean} denotes mean humidity of an interval from 10 to 15 days after sowing and P_{x50}^{mean} denotes mean precipitation of a period over 50 days after sowing.

While SNPs were identified only in *Cicer reticulatum* samples from 15 collection sites we are able to fit the model to the whole dataset giving appropriate values to the indicator variables – the elements of matrix D (see formulae 3 and 4).

The analysis of relative difference in the sum of squares for a model with and without the interaction terms between climatic factors and each SNP group allows us to conclude that genotype-by-environment interaction accounts for about 17.2% of variation in time period from sowing to flowering. All SNPs interact with temperature and day length. Additionally, SNP3 interacts with relative humidity and SNP4 interacts with precipitation.

To analyze the difference in response of SNP groups to climatic factors we built regression models (1) for each group separately. The distributions of time to flowering for these groups are presented in Additional file 1: Figure S8–S13. Selected models are presented in Additional file 1: Section S4.

Due to the stochastic nature of the procedure ten runs were performed with the same algorithmic parameters using different seeds for the random number generator to obtain an ensemble of models. Various agroclimatic factors and their combinations were selected as predictors by stochastic optimization.

We calculated the coefficients of determination for ensemble of models from which the terms that do not contain a predictor of a climatic factor or a combination of factors analyzed were excluded. The box plots of coefficient values for day length, temperature, precipitation, humidity and their combinations are presented in Additional file 1: Figure S14–S19 for all SNPs.

The multiple-way ANOVA (MANOVA) applied to the coefficient of determination as dependent variable and SNP group membership as an independent variable showed that the difference in mean effects of climatic factors on SNP groups is statistically significant (Pillai statistic value 1.287, $p = 1.333e - 09 < 0.05$).

Next we applied one-way ANOVA to test the influence of each climatic factor individually. The significant differences in the means of the coefficient determination values were observed for day length, humidity and the combination of precipitation and day length ($F = 2.102$, $p = 0.009497$ and $F = 6.642$, $p = 5.159e - 12$ and $F = 1.904$, $p = 0.0218$ respectively).

The next step in our analysis was the pair-wise comparison of climatic factor influences on flowering time between SNP groups. The direction and extent of each factor influence on phenotype was estimated as a finite difference approximation of the partial derivative of a regression function in respect to the factor. Additional file 1: Figure S20 presents the box plots of factor influence estimators calculated for model ensembles and for each SNP group.

The means of estimators of a factor influence on phenotype averaged over SNP groups were compared with a Mann-Whitney-Wilcoxon test. As is evident from an analysis of Table 1, climatic factors had divergent effects on genotypes with different reference alleles at five out of six polymorphic position analyzed. As an example, for SNP1 (T→G) day length has different effects on plants with ALT/ALT and REF/ALT, as well as REF/REF and ALT/ALT allele combinations. Precipitation influences plants with ALT/ALT and REF/REF combinations differently. In case of SNP2 (A→G) we found clear differences between genotypes with ALT/ALT and REF/REF for combination of day length with either temperature or precipitation. For SNP3 (C→T) humidity affects genotypes with ALT/ALT and REF/REF differently, day length – temperature combination exerts different influence on ALT/ALT and ALT/REF genotypes, as well as ALT/REF and REF/REF genotypes, day length – precipitation combination shows different effects on ALT/ALT

Table 1 Statistically significant differences in effects of climatic factors and their combinations on plant genotype

SNP	Factor	Genotype pairs	P value
1	DL	REF/ALT vs. REF/REF	0.035
1	P	ALT/ALT vs. REF/REF	0.032
1	DL	ALT/ALT vs. REF/REF	0.044
2	DL*TEMP	ALT/ALT vs. REF/REF	0.021
2	DL*P	ALT/ALT vs. REF/REF	0.040
3	DL*TEMP	ALT/ALT vs. REF/ALT	0.017
3	U	ALT/ALT vs. REF/REF	0.019
3	DL*TEMP	REF/ALT vs. REF/REF	0.045
3	DL*P	ALT/ALT vs. REF/ALT	0.049
5	DL	ALT/ALT vs. REF/ALT	0.042
5	DL	REF/ALT vs. REF/REF	0.043
5	P	REF/ALT vs. REF/REF	0.021
6	DL	ALT/ALT vs. REF/ALT	0.021

The pair-wise comparisons were performed with Mann-Whitney-Wilcoxon test. DL – day length, TEMP – temperature, P – precipitation; REF – reference allele, ALT – alternative allele for polymorphic site

and REF/ALT genotypes. For SNP5 (C→A) there is difference in influence of day length on REF/REF and REF/ALT, as well as REF/ALT and ALT/ALT genotypes. In addition, precipitation also affects differently ALT/REF and REF/REF genotypes. Different effects of day length on

ALT/ALT and REF/ALT genotypes is evident for SNP6 (A→G).

To further understand the relationship between precipitation and the allele frequency of the SNPs, we correlated the allele frequency of 15 populations (see Additional file 1: Table S4) at the putative GWAS SNPs with the mean annual precipitation at the primary collection sites of the genotypes. Allele frequency of the SNPs 1, 4, 5 and 6 have a linear relationship and are correlated with mean annual precipitation. This is indicative of the alleles being fixed in the populations which are found in the areas with high mean precipitation (see Fig. 7). Spearman's rank correlation between mean annual temperature and the allele frequency of the SNPs resulted in no significant relationship. This is indicative of 2 possible scenarios, a) the mean annual temperature value might not be indicative of critical time window affecting the time of flowering in the genotypes, b) the SNP alleles are not in genes involved in the pathways of temperature response (see Additional file 1: Figure S1).

Discussion

The lifecycle of chickpea is strongly determined by environmental factors. Consequently, its phenology is likely strongly predicted by geographic origin and local phenotyping environment, as demonstrated in domestic

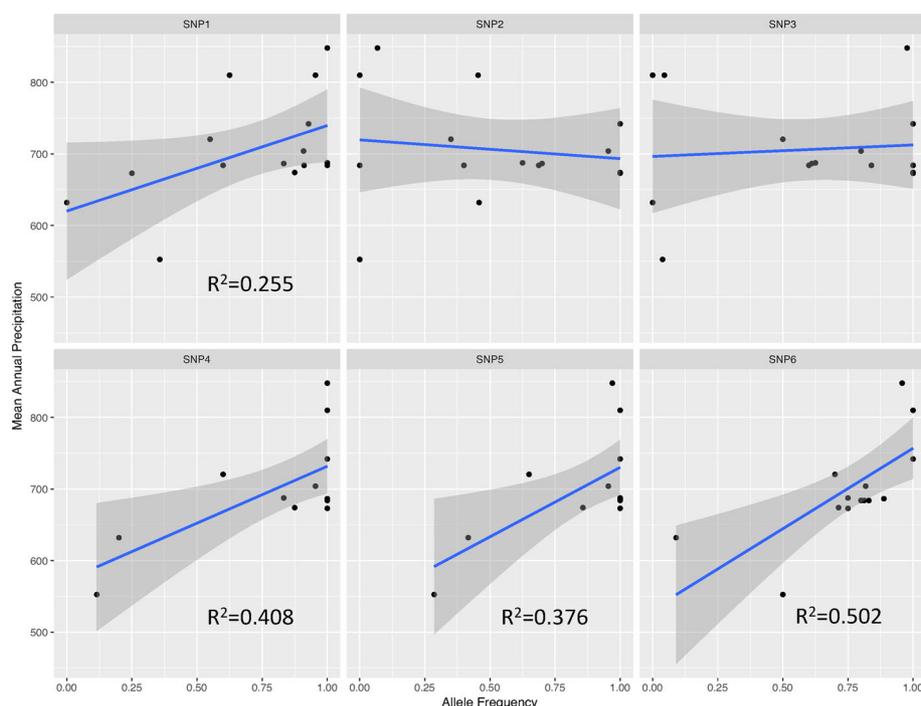


Fig. 7 Correlations of mean annual precipitation (mean_annual_prec) with allele frequency of the 6 GWAS SNPs calculated for 15 populations of the wild chickpeas (shown for completeness). Allele frequency of SNPs 1, 4, 5 and 6 are correlated with mean annual precipitation. The allele frequencies have a linear relationship at each of these significant SNPs, showing that these alleles are nearly fixed in the population in regions with high mean annual precipitation

chickpea cultivars and landraces originating from the Mediterranean to southern India [3]. Here we investigate this hypothesis in the wild progenitors of chickpea by statistical modeling of chickpea responses to environment conditional on geographic site of origin and genotype. Usually the extent of $G \times E$ interaction due to sampling site and environmental factors is modeled by state-of-the-art techniques such as AMMI and factorial regression or by using bioclimatic variables as a GWAS phenotype. Here we implemented a more general solution in which the analytic form of dependencies between predictors (climatic factors, collection sites and genotypes) and phenotype (flowering time) is automatically inferred by a stochastic optimization technique. Apart from automation the advantage of our approach resides in its ability to quickly examine different fits to the data and select the optimal one. We performed model parameterization on a wild chickpea dataset collected at 21 different locations in Turkey [43] grown in 4 different environments. GWAS analysis of the data identified six polymorphic sites responsible for flowering time variation independent of environmental conditions (Singh, A.: Genome-wide association studies in wild chickpea, in preparation).

We built two types of flowering time models – for the whole dataset and for groups of plants, that either originated from one sampling site or have similar allele combination at one of the 6 SNP positions.

Using the models for the whole dataset we found that 14.7% and 17.2% of variation in time to flowering is accounted for by interactions of climatic factors with geographic origin of the plant and its genotype, respectively. Contrary to previous approaches that measure the combined sensitivity of the phenotype to all environmental factors, our approach makes it possible to identify responses to specific environmental conditions and sampling locations in individual accessions, collection sites or SNP groups. In this case we have treated collection site as a model parameter which describes the composite influence of geography (latitude, altitude etc.) climate (day length, temperature) and biological interactions on phenotype. We found that in total 15 out of 21 sampling sites interact with different climatic factors at the phenotyping site, day length and humidity in particular. We also showed that all of six polymorphic sites identified in GWAS interact with temperature and day length, and that SNP3 and SNP4 additionally interact with relative humidity and precipitation respectively.

The influence of the geographic site of origin on plant phenology was further confirmed by applying a group-oriented approach. We found that wild chickpea accessions originating from different collection sites react differently to different environments. For example, plants collected at Baristepe1 react differently to day length

change in comparison to plants from locations Baristepe3, Destek, Egil, Ortanca, Savur1 and Sirnak1 (see Fig. 3).

Observing the relation between climatic factors at the site of genotype collection, we hypothesized that there should be an association between the allele frequency of the GWAS SNPs and climatic factors at genotype collection site. This was confirmed by strong correlations of allele frequency with collection site mean annual precipitation in 4 of the 6 SNP groups (Fig. 7). Three of these four SNPs, have fixed alleles (allele frequency 1) within populations with highest mean precipitation. This makes sense, given strong selection for climate-appropriate flowering time in Mediterranean annuals, which typically flower early to avoid terminal drought in low rainfall regions, but flower later to maximize their reproductive potential in longer season, high rainfall environments [71]. In this context we were interested to discover that there was no correlation of SNP allele frequencies with collection site annual mean temperature. This may be explained by strong site and SNP interaction for phenotyping temperature whereby genotypes collected at different locations responded differently to temperature (and precipitation). Thus, increasing temperature led to earlier flowering in some locations (e.g. Ortanca, Savur) and later flowering in others (e.g. Kesentas) (see Fig. 2), that makes it impossible to reveal dependencies between SNP frequencies and temperature with standard correlation analysis.

We were also able to demonstrate that certain environmental variables differently affect flowering time of genotypes with different allele combinations at five out of six polymorphic position analyzed. For example, different allele combinations at SNP1 differently react on day length change (see Table 1). This is an important characteristic of the SNP that might be used in practice.

We believe that the models we have developed here can be plugged into existing process-based models, such as SSM, to build a new generation of crop models that predicts aspects of crop performance based on genetic, geographic, environmental and management data. In an era of growing genomic information, these new models are essential. Specific subroutines modeling selected biological processes could be modified to incorporate effects on these variables without altering other processes within the model. With Grammatical Evolution and DEEP this can be achieved in automatic way, easing the adaptation of crop models in breeding programs around the world.

Conclusions

Analyzing patterns of adaptation is a key for defining strategies to cope with GxE interactions in breeding for either wide or specific adaptation. The phenology of adaptive traits, like flowering time, may be strongly predicted by plant geographic origin and local environmental factors. Here we tested this hypothesis by statistical modeling

of wild chickpea flowering time responses to different environmental conditions. Our results showed that 1) geographic origin of a plant is indeed a good predictor of flowering time in chickpea and 2) allele combinations at GWAS hits associated with flowering time are “environmentally responsive”, i.e. react differently to changes in climatic factors.

Our methodology is generic and can be further applied and extended to existing crop models.

Additional file

Additional file 1: Additional file 1 contains information on SNP based groups, climatic data for these groups, details on Grammatical evolution method. (PDF 634 kb)

Abbreviations

AMMI: Additive main effect and multiplicative interaction; APSIM: Agricultural Production Systems simulator; BD: Biological Days; CHU: Crop Heat Units; CISP: Conserved-intron scanning primers; DART: Diversity Arrays Technology; DD: Degree Days; DEEP: Differential Evolution Entirely Parallel; DSSAT: Decision Support System for Agrotechnology Transfer; EST: Expressed Sequence Tags; GE: Grammatical Evolution; GLM: General linear model; GUI: Graphical user interface; GWAS: Genome-wide association studies; HUI: Heat Unit Index; ICARDA: International Center for Agricultural Research in the Dry Areas; IRA: Iterative regression analysis; LASSO: Least absolute shrinkage and selection operator; MANOVA: Multiple-way analysis of variance; MLM: Mixed linear model; SNP: Single nucleotide polymorphism; SSM: Simple Simulation Modeling; SSR: Single sequence repeat; STMS: Sequenced tagged microsatellite site; TASSEL: Trait Analysis by aSSociation, Evolution and Linkage; TTF: Time to flowering

Acknowledgements

We thank Lyubov Novikova, Svetlana Surkova, Alena Sokolkova and Peter Chang for helpful discussions. This work was performed using computational resources of the Supercomputer Center of Peter the Great St. Petersburg Polytechnic University (<http://www.scc.spbstu.ru>).

Funding

Data collection and preparation, GWAS, as well as Spearman's correlation analysis of allele frequencies in populations (Fig. 7 and Additional file 1: Figure S1) was supported by a cooperative agreement from the United States Agency for International Development under the Feed the Future Program AID-OAA-A-14-00008 to D.R.C., E.J.B.v.W., S.V.N., A.F.W., A.K., V.V. and R.V.P. and by support from University of Southern California, Dornsife Chemical Biology Training Program. All other analyses were supported by RScF grant #16-16-00007. Publication of this article was supported by RScF grant #16-16-00007.

Availability of data and materials

The datasets and programs used and/or analyzed during the current study available from the corresponding author on request.

About this supplement

This article has been published as part of *BMC Plant Biology Volume 19 Supplement 2, 2018: Selected articles from BGRS\SB-2018: plant biology (part 2)*. The full contents of the supplement are available online at <https://bmcpantbiol.biomedcentral.com/articles/supplements/volume-19-supplement-2>.

Authors' contributions

Sergey Nuzhdin (SN), Maria Samsonova (MS), Eric Bishop-von Wettberg (EW) and Konstantin Kozlov (KK) conceived and designed the study. Jens Berger (JB), Douglas Cook (DC), Abdulkadir Aydogan (AA) and Abdullah Kahraman (AK) prepared the dataset. EW performed the sampling for wild races and prepared geo data. Anupam Singh (AS) performed GWAS analysis and Spearman's correlation analysis of allele frequencies in populations. KK, SN and

MS formulated the model. KK performed the calculations. KK, SN, and MS analyzed the results. AS, KK, SN, EW and MS wrote the text. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya, 195251 St. Petersburg, Russia. ²Program Molecular and Computation Biology, University of California, University Park, 24105 Los-Angeles, CA, USA. ³Commonwealth Scientific and Industrial Research Organization (CSIRO), Agriculture and Food, Underwood Ave, 6014 Perth, WA, Australia. ⁴Department of Plant and Soil Science, University of Vermont, 63 Carrigan Drive, 05405 Burlington, VT, USA. ⁵Department of Plant Pathology, University of California, One Shields Ave, 95616-8680 Davis, CA, USA. ⁶Central Research Institute for Field Crops (CRIFC), P.O. Box 226, 06042 Ankara, Turkey. ⁷Department of Field Crops, Faculty of Agriculture, Harran University, Osmanbey Campus, 63100 Sanliurfa, Turkey.

Published: 19 March 2019

References

1. Roberts EH, Hadley P, Summerfield RJ. Effects of temperature and photoperiod on flowering in chickpeas (*Cicer arietinum* L). *Ann Bot*. 1985;55(6):881–92.
2. Smithson JB, Thompson JA, Summerfield RJ. Chickpea (*Cicer arietinum* L). In: Summerfield RJ, Roberts RE, editors. *Grain Legume Crops*. London: Collins; 1985. p. 312–90.
3. Berger J, Milroy S, Turner N, Siddique K, Imtiaz M, Malhotra R. Chickpea evolution has selected for contrasting phenological mechanisms among different habitats. *Euphytica*. 2011;180:1–15.
4. Singh P, Virmani SM. Modelling growth and yield of chickpea (*Cicer arietinum* L). *Field Crop Res*. 1996;46:41–59.
5. Upadhyaya HD, Bajaj D, Das S, Saxena MS, Badoni S, Kumar V, Tripathi S, Gowda CLL, Sharma S, Tyagi AK, Parida SK. A genome-scale integrated approach aids in genetic dissection of complex flowering time trait in chickpea. *Plant Mol Biol*. 2015;89(4):403–20. <https://doi.org/10.1007/s11103-015-0377-z>.
6. Kumar V, Singh A, Mithra SVA, Krishnamurthy SL, Parida SK, Jain S, Tiwari KK, Kumar P, Rao AR, Sharma SK, Khurana JP, Singh NK, Mohapatra T. Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res*. 2015;22(2):133–45. <https://doi.org/10.1093/dnares/dsu046>.
7. Abbo S, Berger J, Turner N. Evolution of cultivated chickpea: Four bottlenecks limit diversity and constrain adaptation. *Funct Plant Biol*. 2003;30:1081–1087.
8. Ellis RH, Lawn RJ, Summerfield RJ, Qi A, Roberts EH, Chay PM, Brouwer JB, Rose J, Yeates SJ, Sandover S, et al. Towards the reliable prediction of time to flowering in six annual crops. v. chickpea (*Cicer arietinum*). *Exp Agric*. 1994;30(3):271–82. <https://doi.org/10.1017/S0014479700024376>.
9. Vadez V, Soltani A, Sinclair TR. Crop simulation analysis of phenological adaptation of chickpea to different latitudes of India. *Field Crops Res*. 2013;146:1–9.
10. Soltani A, Hammer G, Torabi B, Robertson M, Zeinali E. Modeling chickpea growth and development: Phenological development. *Field Crops Res*. 2006;99:1–13.
11. Vadez V, Soltani A, Sinclair TR. Modelling possible benefits of root related traits to enhance terminal drought adaptation of chickpea. *Field Crops Res*. 2012;137:108–15. <https://doi.org/10.1016/j.fcr.2012.07.022>.
12. Soltani A, Robertson M, Mohammad-Nejad Y, Rahemi-Karizaki A. Modeling chickpea growth and development: Leaf production and senescence. *Field Crops Res*. 2006;99:14–23.

13. Jones J, Antle J, Basso B, Boote KJ, Conant RT, Foster I, Godfray HCJ, Herrero M, Howitt RE, Janssen S, Keating B, Munoz-Carpena R, Porter C, Rosenzweig C, Wheeler TR. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agric Syst.* 2017;155:269–88.
14. Jones J, Antle J, Basso B, J Boote K, T Conant R, Foster I, Charles J Godfray H, Herrero M, E Howitt R, Janssen S, Keating B, Muñoz-Carpena R, Porter C, Rosenzweig C, R. Wheeler T. Brief history of agricultural systems modeling. 2016;155:240–254.
15. Jones J, Hoogenboom G, Porter C, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijssman AJ, Ritchie JT. The DSSAT cropping system model. *Eur J Agron.* 2003;18:1161.
16. J. Boote K, Jones J, Pickering N. Potential uses and limitations of crop models. *Agron J.* 1996;88:704–16.
17. Boote KJ, Jones J, White JW, Asseng S, Lizaso JJ. Putting mechanisms into crop production models. *Plant Cell Environ.* 2013;36(9):1658–1672.
18. Keating B, Carberry PS, Hammer G, Probert ME, Robertson MJ, Holzworth D, Huth NI, Hargreaves J, Meinke H, Hochman Z, Mclean G, Verburg K, Snow V, Dimes J, Silburn D, Wang E, Brown S, Bristow K, Asseng S, Smith C. An overview of APSIM, a model designed for farming systems simulation. *Eur J Agron.* 2003;18:267–88.
19. Battisti R, Sentelhas PC, Boote KJ. Sensitivity and requirement of improvements of four soybean crop simulation models for climate change studies in Southern Brazil. *Int J Biometeorol.* 2017. <https://doi.org/10.1007/s00484-017-1483-1>.
20. Williams JR, Jones CA, Kiniry JR, Spanel DA. The EPIC crop growth model. *Trans ASAE.* 1989;32(2):497–511.
21. Wilkerson GG, Jones J, Boote KJ, Ingram KT, Mishoe JW. Modeling soybean growth for crop management. *Trans Am Soc Agric Eng.* 1983;26(1):63–73.
22. Roorkiwal M, Rathore A, Das RR, Singh MK, Jain A, Srinivasan S, Gaur P, Chellapilla B, Tripathi S, Li Y, Hickey JM, Lorenz A, Sutton T, Crossa J, Jannink J-L, Varshney RK. Genome-enabled prediction models for yield related traits in chickpea. *Front Plant Sci.* 2016;7:1666.
23. Hoogenboom G, White JW, Jones J, Boote KJ. Beangro: A process-oriented dry bean model with a versatile user interface. *Agon J.* 1994;86(1):186–90.
24. Jones J, Keating B, Porter C. Approaches to modular model development. *Agric Syst.* 2001;70(2):421–43. [https://doi.org/10.1016/S0308-521X\(01\)00054-3](https://doi.org/10.1016/S0308-521X(01)00054-3).
25. Wajid A, Rahman MHU, Ahmad A, Khaliq T, Mahmood N, Rasul F, Bashir MU, Awais M, Hussain J, Hoogeboom G. Simulating the interactive impact of nitrogen and promising cultivars on yield of lentil (*Lens culinaris*) using CROPGRO-legume model. *Int J Agric Biol.* 2013;15(6):1331–6.
26. Ilkaee MN, Paknejad F, Golzardi F, Tookaloo MR, Habibi D, Tohidloo G, Pazoki A, Agayari F, Rezaee M, Rika ZF. Simulation of some of important traits in chickpea cultivars under different sowing date using cropgro-pea model. *Int J Biosci.* 2014;4(12):84–92.
27. Soltani A, Sinclair TR. A simple model for chickpea development, growth and yield. *Field Crops Res.* 2011;124:252–60.
28. Lal M, Singh KK, Srinivasan G, Rathore LS, Naidu D, Tripathi CN. Growth and yield responses of soybean in Madhya Pradesh, India to climate variability and change. *Agric For Meteorol.* 1999;93:53–70.
29. Chung U, Kim Y, Seo B, Seo M. Evaluation of variation and uncertainty in the potential yield of soybeans in South Korea using multi-model ensemble climate change scenarios. *Agrotechnology.* 2017;6(2):1000158.
30. Mohammed A, Tana T, Singh P, Molla A, Seid A. Identifying best crop management practices for chickpea (*Cicer arietinum* L.) in northeastern Ethiopia under climate change condition. *Agric Water Manag.* 2017;194:68–77.
31. Patil DD, Patel HR. Calibration and validation of CROPGRO (DSSAT 4.6) model for chickpea under middle GUJARAT agroclimatic region. *Int J Agric Sci.* 2017;9:4342–4.
32. Urgaya M. Modeling the impacts of climate change on chickpea production in Adaa Woreda (East Showa zone) in the semi-arid central rift valley of Ethiopia. *J Pet Environ Biotechnol.* 2016;7:288.
33. Bhosale SU, Stich B, Rattunde HFW, Weltzien E, Haussmann BI, Hash CT, Ramu P, Cuevas HE, Paterson AH, Melchinger AE, Parzies HK. Association analysis of photoperiodic flowering time genes in west and central African sorghum [*Sorghum bicolor* (L.) Moench]. *BMC Plant Biol.* 2012;12(1):32. <https://doi.org/10.1186/1471-2229-12-32>.
34. Visioni A, Tondelli A, Francia E, Psarwayi A, Malosetti M, Russell J, Thomas W, Waugh R, Pecchioni N, Romagosa I, Comadran J. Genome-wide association mapping of frost tolerance in barley (*Hordeum vulgare* L.). *BMC Genomics.* 2013;14(1):424. <https://doi.org/10.1186/1471-2164-14-424>.
35. Tian F, Bradbury P, Brown P, Hung H, Sun Q, Flint-Garcia S, Rocheford T, McMullen M, Holland J, Buckler E. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet.* 2011;43(2):159–62. <https://doi.org/10.1038/ng.746>.
36. Kump KL, Bradbury PJ, Visser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat Genet.* 2011;43(2):163–8. <https://doi.org/10.1038/ng.747>.
37. Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W, Feng H, Xie W, Lian X, Wang G, Luo Q, Zhang Q, Liu Q, Xiong L. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat Commun.* 2014;5:5087. <https://doi.org/10.1038/ncomms6087>.
38. Suwarno WB, Pixley KV, Palacios-Rojas N, Kaeppler SM, Babu R. Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theor Appl Genet.* 2015;128(5):851–64. <https://doi.org/10.1007/s00122-015-2475-3>.
39. Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE, Buckler ES, Brenton Z, Kresovich S, Morris GP. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv.* 2015;1(6). <https://doi.org/10.1126/sciadv.1400218>. <http://advances.sciencemag.org/content/1/6/e1400218.full.pdf>.
40. Hwang C, Correll MJ, Gezan SA, Zhang L, Bhakta MS, Vallejos CE, Boote KJ, Clavijo-Michelangeli JA, Jones J. Next generation crop models: A modular approach to model early vegetative and reproductive development of the common bean (*Phaseolus vulgaris* L.). *Agric Syst.* 2017;155:225–39. <https://doi.org/10.1016/j.agry.2016.10.010>.
41. Hatfield J, Walthall C. Meeting global food needs: Realizing the potential via genetics x environment x management interactions. *Agron J.* 2015;107:1215–26.
42. Tardieu F, Tuberosa R. Dissection and modelling of abiotic stress tolerance in plants. *Curr Opin Plant Biol.* 2010;13:206–12.
43. von Wettberg EJ, Chang PL, Başdemir F., Carrasquilla-Garcia N, Korbu LB, Moenga SM, Bedada G, Greenlon A, Moriuchi KS, Singh V, Cordeiro MA, Noujdina NV, Dinegde KN, Shah Sani SGA, Getahun T, Vance L, Bergmann E, Lindsay D, Mamo BE, Warschefsky EJ, Dacosta-Calheiros E, Marques E, Yilmaz MA, Cakmak A, Rose J, Migneault A, Krieg CP, Saylak S, Temel H, Friesen ML, Siler E, Akhmetov Z, Ozcelik H, Kholova J, Can C, Gaur P, Yildirim M, Sharma H, Vadez V, Tesfaye K, Woldemedhin AF, Tar'an B, Aydogan A, Bukun B, Penmetsa RV, Berger J, Kahraman A, Nuzhdin SV, Cook DR. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat Commun.* 2018;9:649. NNDc. Climate Data On-line. <https://www7.ncdc.noaa.gov/CDO/cdselect.cmd?datasetabbv=GSOD>. Accessed 30 Dec 2017.
44. Hammer GL, Vaderlip RL, Gibson G, Wade LJ, Henzell RG, Younger DR, Warren J, Dale AB. Genotype-by-environment interaction in grain sorghum. II. Effects of temperature and photoperiod on ontogeny. *Crop Sci.* 1989;29:376–84.
45. Horie T. Crop ontogeny and development. In: Boote KJ, Bennett JM, Sinclair TR, Paulsen GM, editors. *Physiology and Determination of Crop Yield*. Madison, USA: ASA, CSSA, and SSSA; 1994. p. 153–180.
46. Piper EL, Boote KJ, Jones J, Grimm SS. Comparison of two phenology models for predicting flowering and maturity date of soybean. *Crop Sci.* 1996;36:1606–14.
47. Yin X, Kropff MJ, Horie T, Nakagawa H, Centeno HGS, Zhu D, Goudriaan J. A model for photothermal responses of flowering in rice. i. model description and parameterization. *Field Crops Res.* 1997;51:189–200.
48. Robertson MJ, Carberry PS, Huth NI, Turpin JE, Probert ME, Poulton PL, Bell M, Wright GC, Yeates SJ, Brinsmead RB. Simulation of growth and development of diverse legume species in APSIM. *Aust J Agric Res.* 2002;53:429–46.
49. Robertson MJ, Watkinson AR, Kirkegaard JA, Holland JF, Potter TD, Burton W, Walton GH, Moot DJ, Wratten N, Farre I, Asseng S. Environmental and genotypic control of time to flowering in canola and Indian mustard. *Aust J Agric Res.* 2002;53:793–809.

51. Major DJ, Johnson DR, Tanner JW, Anderson IC. Effects of daylength and temperature on soybean development. *Crop Sci.* 1975;15:174–9.
52. Hodges T, French V. Soybean: soybean stages modeled from temperature, daylength and water availability. *Agron J.* 1985;77:500–5.
53. O'Neill M, Ryan C. Grammatical evolution. *EE Trans Evol Comput.* 2001;5(4):349–58. <https://doi.org/10.1109/4235.942529>.
54. Noorian F, de Silva A, Leong P. gramEvol: Grammatical Evolution in R. *J Stat Softw Artic.* 2016;71(1):1–26. <https://doi.org/10.18637/jss.v071.i01>.
55. Tibshirani R. Regression shrinkage and selection via the LASSO. *J R Stat Soc Ser B.* 1996;58(1):267–88.
56. Kozlov K, Samsonov A. DEEP – Differential Evolution Entirely Parallel Method for Gene Regulatory Networks. *J Supercomput.* 2011;57:172–8. <https://doi.org/10.1007/s11227-010-0390-6>.
57. Kozlov K, Samsonov AM, Samsonova M. A software for parameter optimization with differential evolution entirely parallel method. *PeerJ Comput Sci.* 2016;2:74. <https://doi.org/10.7717/peerj-cs.74>.
58. Kozlov K, Novikova LY, Seferova IV, Samsonova MG. Mathematical model of soybean development dependence on climatic factors. *Biofizika.* 2018;63:175–6.
59. Storn R, Price K. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Technical Report Technical Report TR-95-012, ICSI. 1995.
60. Zaharie D. Parameter adaptation in differential evolution by controlling the population diversity. In: Petcu D, editor. Proc. of 4th International Workshop on Symbolic and Numeric Algorithms for Scientific Computing. Timisoara, Romania: Analele Universitatii Timisoara; 2002. p. 385–97.
61. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. R Foundation for Statistical Computing. <https://www.R-project.org/>.
62. Pillai KCS. Regression shrinkage and selection via the LASSO. *Ann Math Stat.* 1955;26:117–21.
63. Peter Harrington. Genetic Programming C++ Code. <https://github.com/pbharrin/Genetic-Prog>. Accessed 30 Dec 2017.
64. Sanderson C, Curtin R. Armadillo: a template-based C++ library for linear algebra. *J Open Source Soft.* 2016;1:26.
65. Curtin RR, Cline JR, Slagle NP, March WB, Ram P, Mehta NA, Gray AG. mlpack: A scalable C++ machine learning library. *J Mach Learn Res.* 2013;14:801–5.
66. The HDF Group. Hierarchical Data Format, Version 5. <http://www.hdfgroup.org/HDF5/>. Accessed 30 Dec 2017.
67. The Blue Brain Project. HighFive - Header-only C++ HDF5 Interface. <https://github.com/BlueBrain/HighFive>.
68. The Qt Company. Qt Library. <https://www.qt.io/>. Accessed 30 Dec 2017.
69. Kozlov K. NLREG. <https://gitlab.com/mackoel/nlreg>. Accessed 30 Dec 2017.
70. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;23(19):2633–5. <https://doi.org/10.1093/bioinformatics/btm308>.
71. Berger J, Shrestha D, Ludwig C. Reproductive Strategies in Mediterranean Legumes: Trade-Offs between Phenology, Seed Size and Vigor within and between Wild and Domesticated Lupinus Species Collected along Aridity Gradients. *Front Plant Sci.* 2017;8:548.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

