

# ANALYSIS OF A REGRESSION MODEL FOR TIME TO FLOWERING IN WILD CHICKPEA WITH GENOTYPE-BY-ENVIRONMENT INTERACTIONS

**K.N. Kozlov<sup>1</sup>, Anupam Singh<sup>2</sup>, Eric Bishop-von Wettberg<sup>3</sup>, Sergey Nuzhdin<sup>2</sup>, and M.G. Samsonova<sup>1</sup>**

<sup>1</sup>Peter the Great St. Petersburg Polytechnic University, St. Petersburg 195251, Russia, kozlov\_kn@spbstu.ru

<sup>2</sup>Program Molecular and Computation Biology, University of California, Los-Angeles, USA

<sup>3</sup>Department of Plant and Soil Science, University of Vermont, 05405 Burlington, VT, USA

Chickpea is one of the most popular legumes, cultivated in 50 countries around the world. New varieties are being developed by introgressing wild landraces. Accurate prediction of crop flowering time is required for reaching maximal farm efficiency. Crucial to this effort are predictive models that connect agricultural traits to climatic factors.

A regression model for time to flowering is constructed by the combination of Grammatical Evolution (GE), LASSO and Differential Evolution Entirely Parallel (DEEP) method to recover analytic form of a dependence, find regression coefficients and determine the set of climatic factors, respectively. In a proposed approach, the unknown parameters of the model are inferred automatically by stochastic minimization of the deviation of the model output from data [1]. The method was applied to predict the flowering time in a dataset of wild chickpea collected at in five regions in Turkey by von Wettberg et al [2]. These wild accessions were planted in climatically distinct sites in Turkey and Australia. Being grown in contrasting environments the phenotype data on time to flowering is highly diverse. The dataset is subdivided into 18 groups according to allele combination at six SNPs associated with flowering time that were identified earlier.

In this work we analyze a regression model for time to flowering in a dataset of wild chickpea using bootstrap approach. We performed 1999 model adaptations using datasets resampled from original one with repetitions. The analysis revealed that the 95% confidence intervals for five out of six regression coefficients didn't contain zero and thus represent well established influence of climatic factor on time to flowering. Coefficients of genotype-by-environment interactions had 98% of confidence intervals that didn't include zero for each SNP.

The study was supported by the RFBR grant 18-29-13033.

## References

1. Kozlov, K.N., Novikova, L.Y. et al. A Mathematical Model of the Impact of Climatic Factors on Soybean Development // *Biophysics*, Vol. 63, issue 1, 2018, pp. 175–176.
2. von Wettberg, E.J., Chang, P.L. et al. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation // *Nature Communications* 9, 2018, 649