ISSN 0006-3509, Biophysics, 2020, Vol. 65, No. 2, pp. 237–240. © Pleiades Publishing, Inc., 2020.
Russian Text © The Author(s), 2020, published in Biofizika, 2020, Vol. 65, No. 2, pp. 276–279.

The Signatures of Ecological Adaptation in the Genomes of Chickpea Landraces

A. B. Sokolkova[a], P. L. Chang[b], N. Carrasquila-Garcia[b], N. V. Noujdina[c], D. R. Cook[b], S. V. Nuzhdin[a, d], and M. G. Samsonova[a, *]

[a] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, 195251 Russia
[b] Faculty of Plant Pathology, University of California, Davis, CA 95616, USA
[c] Faculty of Geography, University of California, Los Angeles, CA 90095, USA
[d] University of Southern California, Los Angeles, CA 90089, USA
*e-mail: m.samsonova@spbstu.ru

Received December 4, 2019; revised December 4, 2019; accepted January 22, 2020

Abstract—The seed bank of the Vavilov All-Russia Institute of Plant Genetic Resources (VIR) contains a wide range of landraces that were collected almost 100 years ago, in which natural selection might leave signatures through crop diversification. In this study, we analyzed 407 landraces sampled at centers of the origin of the chickpea and at sites of secondary diversity. We hypothesize that a fraction of single nucleotide polymorphisms might have exhibited strong selection to a range of environmental conditions that chickpea experienced during domestication and subsequent geographic distribution. Using the BayPass package we identified 13 polymorphisms; these assort by environmental conditions and are strong candidates for local adaptation.

Keywords: chickpea, bioclimatic analysis, genotyping by sequencing (GBS), single nucleotide polymorphisms (SNPs), linkage disequilibrium

DOI: 10.1134/S0006350920020244

In the early 20th century (1911–1940), systematic efforts to collect and preserve the diversity of crops were made under the leadership of N.I. Vavilov. Currently, they are stored in the collection of the Vavilov All-Russia Institute of Plant Genetic Resources (VIR) in St. Petersburg [1]. Geographic variation and genetic diversity of the majority of crops collected at that time should apparently reflect the historical conditions of their cultivation that formed in the previous millennia. The possibility to identify the signatures of historical selection for adaptation to various environmental conditions is of particular interest [2].

In this study, we investigated this possibility using the data for chickpea (*Cicer arietinum*), whose cultivation of which in Russia is currently rapidly increasing. Initially, chickpea was domesticated approximately 10000 years ago in the Fertile Crescent region (Middle East) and then spread to India (~6000 years ago) as well as to Ethiopia and North Africa (~3000 years ago) [3]. During its secondary diversification, chickpea began to be cultivated in a number of new environments and climatic conditions with the use of new cultivation methods, which apparently was made possible due to selection of mutations that segregated in the population or the emergence of new mutations. This hypothesis was tested using the biogeographical, genomics, and computational biology methods, with which specific haplotypes that are potentially responsive to selection were identified. The results can be used in the future to improve crops.

MATERIALS AND METHODS

Genetic material and genomic data. Data were obtained from plants that grow in countries such as Ethiopia, Lebanon, Morocco, Turkey, and India as well as in wider regions of Central Asia and the Mediterranean. Genotyping by sequencing (GBS) data and single nucleotide polymorphisms (SNPs) were taken from [4]. All Illumina data are available in the database of the National Center for Biotechnology (code BioProject PRJNA388691). SNPs were identified using the GATK pipeline [5] and further filtered using VCFtools [6]. A total of 2579 SNPs passed all filters and 407 samples were selected for further analysis.

Genetic and bioclimatic analysis. BayPass package [7] was used to identify the genetic markers associated with bioclimatic variables specific to particular area. For each bioclimatic variable (see Table 1) the mean

Abbreviations: SNPs, single nucleotide polymorphisms; SMS, synthetic morphology score; BF, Bayes factor.

Table 1. The list of bioclimatic covariates and their abbreviations

Bioclimatic covariate	Abbreviation
Average annual temperature, °C * 10	BIO_1
Average difference between the maximum and minimum temperature/day, °C * 10	BIO_2
Isothermality, %	BIO_3
Temperature seasonality, standard deviation * 100	BIO_4
Maximum temperature of the warmest month, °C * 10	BIO_5
Minimum temperature of the coldest month, °C * 10	BIO_6
Average annual temperature range, °C * 10	BIO_7
Average temperature of the wettest quarter, °C * 10	BIO_8
Average temperature of the warmest quarter, °C * 10	BIO_{10}
Average temperature of the coldest wet quarter, °C * 10	BIO_{11}
Precipitation for the year, mm	BIO_{12}
Precipitation for the wettest month, mm	BIO_{13}
Precipitation for the driest month, mm	BIO_{14}
Seasonality of precipitation, mm	BIO_{15}
Precipitation of the wettest quarter, mm	BIO_{16}
Precipitation of the driest quarter, mm	BIO_{17}
Precipitation of the warmest quarter, mm	BIO_{18}
Precipitation of the coldest quarter, mm	BIO_{19}
Digital relief model, m	DEM

Table 2. The synthetic morphology scores calculated on the basis of bioclimatic covariates

Synthetic morphology scores	Bioclimatic variables
SMS1	BIO_1, BIO_3, BIO_6, BIO_8, BIO_{11}, BIO_{15}
SMS2	BIO_{12}, BIO_{13}, BIO_{16}, BIO_{18}
SMS3	BIO_5, BIO_{10}

values were calculated separately for each of the six geographically different groups. Since many bioclimatic variables are strongly correlated, the mean values of each of the five groups of strongly correlated variables were replaced with the synthetic morphology score (SMS), which is the first principal component of data (Table 2). Three bioclimatic variables were not strongly correlated with all other variables and their mean values were used in the analysis without transformation. Thus, we constructed eight variables for studying the association with genetic characters using BayPass. The analysis was performed on the assumption of independence of variables. Each SNP was estimated by the importance sampling algorithm [7], which computes the Bayes factor (BF) value, empirical *P*-value, and respective regression coefficients. Associations were considered significant when the empirical *P*-value was greater than 4. To quantify the strength of association of SNPs with the variables we used the factor assessment scale in accordance with Jeffreys rule [8]: very strong evidence at $15 < BF < 20$ and decisive evidence at $BF > 20$. Manhattan plots were obtained in R using the CMplot library [9].

RESULTS

A total of 407 samples were divided into six separate groups reflecting the geographical location: Ethiopia (ETHI), India (IND), Lebanon (LEB), Morocco (MOR), Turkey (TUR), and Central Asia (UZB). The Mann–Whitney–Wilcoxon test [10] was used to identify the differences between the groups for each bioclimatic covariate.

The analysis of associations of the previously found 2579 SNPs with environmental gradients was performed using the BayPass package [7]. This package

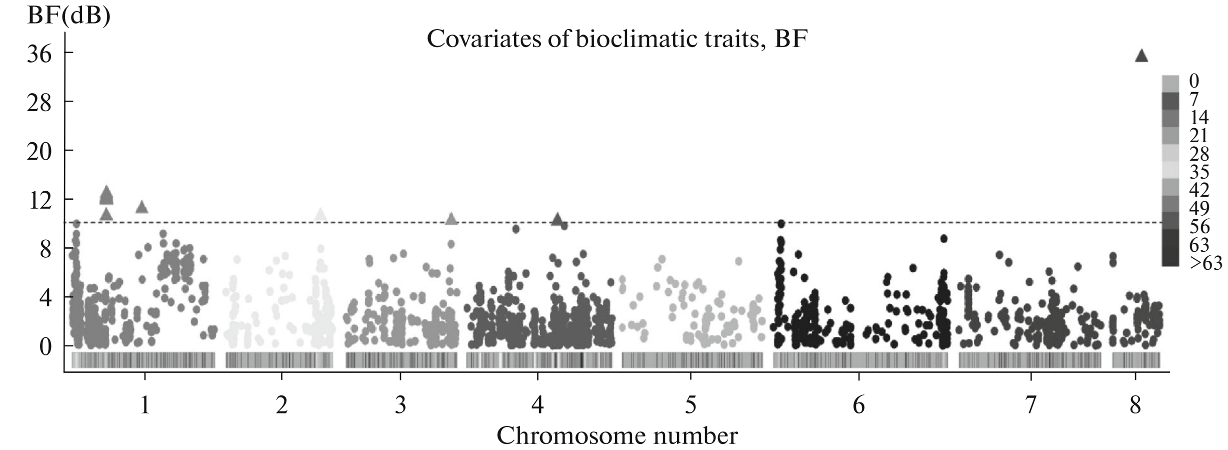

Fig. 1. An analysis of associations using BayPass software. SNPs with BF > 15 are shown with triangles. When one item is associated with a number of bioclimatic covariates with different BF values, only the most significant SNP it presented. The distribution density of SNPs along the chromosomes is shown at the bottom of the Manhattan plot.

allows identification of the genetic markers associated with population variables such as bioclimatic variables at collection sites. In accordance with the recommendations of the authors of the BayPass package, eight covariates were built for each of the six geographically distinct groups: five synthetic morphology estimates (Tables 1 and 2) and three averages non-correlated covariates (BIO2_av, BIO19_av, and DEM_av). Each SMS is a linear combination of averaged values of correlated bioclimatic covariates, which has the highest possible variance. SMS1, SMS3, and SMS5 correspond to the temperature bioclimatic covariates, whereas SMS2 and SMS4 correspond to the precipitation bioclimatic variables. The following criteria were used to assess the significance of association: an empirical Bayesian *P*-value > 4 and (according to the Jeffreys rule [8]) BF > 15.

The analysis of associations using BayPass software revealed 13 SNPs that were significantly associated with covariates (Fig. 1). Eight SNPs on chromosome 1 and one SNP on chromosome 2 were associated with the variable BIO2_av, representing an average difference between the maximum and minimum temperatures in the same day. These eight SNPs are located on chromosome 1 in a 115-kb region with a strong linkage disequilibrium. Only one SNP was associated with SMS5 (Ca4: 30608189). SMS4 was associated with two SNPs on chromosome 1 (Ca1: 23621481) and chromosome 3 (Ca3: 36690022). Variables SMS1 and SMS3, which include temperature characteristics, were jointly associated with one SNP on chromosome 8 (Ca8: 10314452).

CONCLUSIONS

For thousands of years, breeders and farmers have been engaged in breeding crops with desired phenotypes [11]. By combining ecological and genomic data, it is now possible to identify the haplotypes that had been selected by local farmers. In this study, we have identified 13 single nucleotide polymorphisms associated with bioclimatic variables at sample collection sites that mark the genome regions that apparently experienced the effect of selection aimed at adapting to regional growing conditions during their secondary diversification.

FUNDING

Bioclimatic and biogeographical analyses, as well the BayPass analysis, were supported by the Russian Science Foundation (project no. 16-16-00007 (ABS, SVN, MGS)). The study was also supported by the cooperation agreement from the US Agency for International Development in the framework of the Feed the Future program (no. AID-OAA-A-14-00008 (DRC, PLCh, SVN)), the Zumberge Research and Innovation Fund (SVN), the Plant Genome Program of the US National Science Foundation (grant no. IOS-1339346 (DRC)).

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest. This article does not contain any studies involving animals or human participants performed by any of the authors.

REFERENCES

1. M. A. Vishnyakova, M. O. Burlyaeva, S. V. Bulyntsev, et al., S.-kh. Biol. **52** (5), 976 (2017).
2. E. Plekhanova, M. A. Vishnyakova, S. Bulyntsev, et al., Sci. Rep. **7**, 4816 (2017).
3. R. J. Redden and J. D. Berger, in *Chickpea Breeding and Management*, Ed. by S. S. Yadav, R. Redden, W. Chen, and B. Sharma (CABI, Wallingford, UK, 2007), pp. 1–13.

4. E. J. von Wettberg, P. L. Chang, A. Greenspan, et al., Nature Commun. **9**, Art. No. 649 (2018). https://doi.org/10.1038/s41467-018-02867-z

5. A. McKenna, M. Hanna, E. Banks, et al., Genome Res. **20**, 1297 (2010).

6. P. Danecek, A. Auton, G. Abecasis, et al., Bioinformatics **27**, 2156 (2011).

7. M. Gautier, Genetics **201**, 1555 (2015).

8. H. Jeffreys, *Theory of Probability,* 3rd ed. (Oxford Univ. Press, Oxford, 1961).

9. CMplot: Circle Manhattan Plot. URL: https://github.com/YinLiLin/R-CMplot. Cited June 20, 2018.

10. H. B. Mann and D. R. Whitney, Ann. Math. Stat. **18** (1), 50 (1947).

11. N. Maxted, M. E. Dulloo, and B. V. Ford-Lloyd, *Enhancing Crop Gene Pool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement* (CABI, Oxfordshire, UK, 2016).

Translated by M. Batrukova