

University of Vermont

ScholarWorks @ UVM

---

College of Agriculture and Life Sciences Faculty  
Publications

College of Agriculture and Life Sciences

---

10-1-2020

## Genome-wide association study in accessions of the mini-core collection of mungbean (*Vigna radiata*) from the World Vegetable Gene Bank (Taiwan)

Alena Sokolkova

*Peter the Great St. Petersburg Polytechnic University*

Marina Burlyaeva

*N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources*

Tatjana Valiannikova

*Kuban Branch of Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR)*

Margarita Vishnyakova

*N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources*

Roland Schafleitner

*AVRDC - The World Vegetable Center*

Follow this and additional works at: <https://scholarworks.uvm.edu/calsfac>



Part of the [Agriculture Commons](#), [Community Health Commons](#), [Human Ecology Commons](#), [Nature and Society Relations Commons](#), [Place and Environment Commons](#), and the [Sustainability Commons](#)

---

### Recommended Citation

Sokolkova A, Burlyaeva M, Valiannikova T, Vishnyakova M, Schafleitner R, Lee CR, Ting CT, Nair RM, Nuzhdin S, Samsonova M, von Wettberg E. Genome-wide association study in accessions of the mini-core collection of mungbean (*Vigna radiata*) from the World Vegetable Gene Bank (Taiwan). *BMC plant biology*. 2020 Oct;20(1):1-9.

This Article is brought to you for free and open access by the College of Agriculture and Life Sciences at ScholarWorks @ UVM. It has been accepted for inclusion in College of Agriculture and Life Sciences Faculty Publications by an authorized administrator of ScholarWorks @ UVM. For more information, please contact [donna.omalley@uvm.edu](mailto:donna.omalley@uvm.edu).

---

**Authors**

Alena Sokolkova, Marina Burlyaeva, Tatjana Valiannikova, Margarita Vishnyakova, Roland Schafleitner, Cheng Rwei Lee, Chau Ti Ting, Ramakrishnan Madhavan Nair, Sergey Nuzhdin, Maria Samsonova, and Eric von Wettberg

RESEARCH

Open Access



# Genome-wide association study in accessions of the mini-core collection of mungbean (*Vigna radiata*) from the World Vegetable Gene Bank (Taiwan)

Alena Sokolkova<sup>1</sup>, Marina Burlyeva<sup>2</sup>, Tatjana Valiannikova<sup>3</sup>, Margarita Vishnyakova<sup>2</sup>, Roland Schafleitner<sup>4</sup>, Cheng-Ruei Lee<sup>5</sup>, Chau-Ti Ting<sup>5</sup>, Ramakrishnan Madhavan Nair<sup>6</sup>, Sergey Nuzhdin<sup>1,7</sup>, Maria Samsonova<sup>1\*</sup> and Eric von Wettberg<sup>1,8\*</sup>

From Fifth International Scientific Conference "Plant Genetics, Genomics, Bioinformatics, and Biotechnology" (PlantGen2019) Novosibirsk, Russia. 24-29 June 2019

## Abstract

**Background:** Mungbean (*Vigna radiata* (L.) R. Wilczek, or green gram) is important tropical and sub-tropical legume and a rich source of dietary protein and micronutrients. In this study we employ GWAS to examine the genetic basis of variation in several important traits in mungbean, using the mini-core collection established by the World Vegetable Center, which includes 296 accessions that represent the major market classes. This collection has been grown in a common field plot in southern European part of Russia in 2018.

**Results:** We used 5041 SNPs in 293 accessions that passed strict filtering for genetic diversity, linkage disequilibrium, population structure and GWAS analysis. Polymorphisms were distributed among all chromosomes, but with variable density. Linkage disequilibrium decayed in approximately 105 kb. Four distinct subgroups were identified within 293 accessions with 70% of accessions attributed to one of the four populations. By performing GWAS on the mini-core collection we have found several loci significantly associated with two important agronomical traits. Four SNPs associated with possibility of maturation in Kuban territory of Southern Russia in 2018 were identified within a region of strong linkage which contains genes encoding zinc finger A20 and an AN1 domain stress-associated protein.

(Continued on next page)

\* Correspondence: [m.g.samsonova@gmail.com](mailto:m.g.samsonova@gmail.com); [Eric.Bishop-Von-Wettberg@uvm.edu](mailto:Eric.Bishop-Von-Wettberg@uvm.edu)

<sup>1</sup>Peter the Great St. Petersburg Polytechnic University, Department of Applied Mathematics, St. Petersburg, Russia

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



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** The core collection of mungbean established by the World Vegetable Center is a valuable resource for mungbean breeding. The collection has been grown in southern European part of Russia in 2018 under incidental stresses caused by abnormally hot weather and different photoperiod. We have found several loci significantly associated with color of hypocotyl and possibility of maturation under these stressful conditions. SNPs associated with possibility of maturation localize to a region on chromosome 2 with strong linkage, in which genes encoding zinc finger A20 and AN1 domain stress associated protein (SAP) are located. Phenotyping of WorldVeg collection for maturation traits in temperate climatic locations is important as phenology remains a critical breeding target for mungbean. As demand rises for mungbean, production in temperate regions with shorter growing seasons becomes crucial to keep up with needs. Uncovering SNPs for phenology traits will speed breeding efforts.

**Keywords:** Mungbean, Core-collection, Phenotyping, GWAS, Population structure, Phenology traits

## Background

Mungbean (*Vigna radiata* (L.) R. Wilczek, or green gram) is an important but underutilized tropical and sub-tropical legume that is a rich source of dietary protein and micronutrients and plays a vital role in restoring soil fertility in a variety of crop rotational systems. Mungbean has been of the greatest agronomic and cultural significance in India, which accounts for most of the 6 million hectares of it that are produced globally, as well as other tropical countries such as China, Indonesia, Thailand, Myanmar, and the Philippines [1]. However, it has increasing production in temperate regions such as the arid regions of Southern Europe, Queensland in Australia, the Southern Great Plains of the US, and sporadically in the Southern Russian Federation, where it can complement winter annuals such as wheat and barley due to its short duration, allowing two crop cycles per year in places where fallow seasons have previously been common. Currently mungbean accounts for about 8.5% of the world production area for grain legume crops (without counting the oilseed legume soybean) [2].

Although historically considered an orphan legume [3], considerable and significant genetic resources are accumulated in 35 genebanks of the world [4], some of which are well phenotyped. The phenotypic data of the oldest collection in the N. I. Vavilov All-Russian Institute for Genetic Resources (VIR) has been used to trace the trends of crop breeding over the past 100 years [5]. Furthermore, genomic resources have recently been developed for mungbean that allow molecular dissection of important agronomic and domestication traits. The cultivated mungbean genome was completed in 2014, complemented by a genome of its wild relatives *V. radiata* var. *sublobata* and *V. reflexo-pilosa* [6]. High-density molecular markers have been developed, allowing characterization of diversity panels such as core collections. The most widely used core-collection and mini core-collection have been developed in the World Vegetable Center (WorldVeg) [7], which has an international mandate for mungbean breeding and improvement.

Given the broad geographic range over which mungbean is produced, and the importance of phenology to allowing it to complete its lifecycle before inhospitable conditions (drought, winter cold, or the production demands of more economically valuable rotational partners such as wheat or rice), assessing the genetic basis of flowering time and related agronomic traits is critical for finding useful genetic variation for ongoing breeding efforts. Genome-wide association studies (GWAS) have emerged as powerful approach for finding genetic variation in germplasm. This approach has been used in a range of legumes, from soybean [8] and the model legume *Medicago truncatula* [9–11], to other legumes previously considered orphans such as chickpea [12, 13], cowpea [e.g., [14] and pigeonpea [15].

Here we employ GWAS to examine the genetic basis of variation in several important traits in mungbean, using the mini-core collection of mungbean established by the WorldVeg, which includes 296 accessions that represent the major market classes of mungbean and all the micro-satellite alleles present in a collection of 1400 accessions [7]. This collection has been grown in common field plot in southern European part of Russia in 2018.

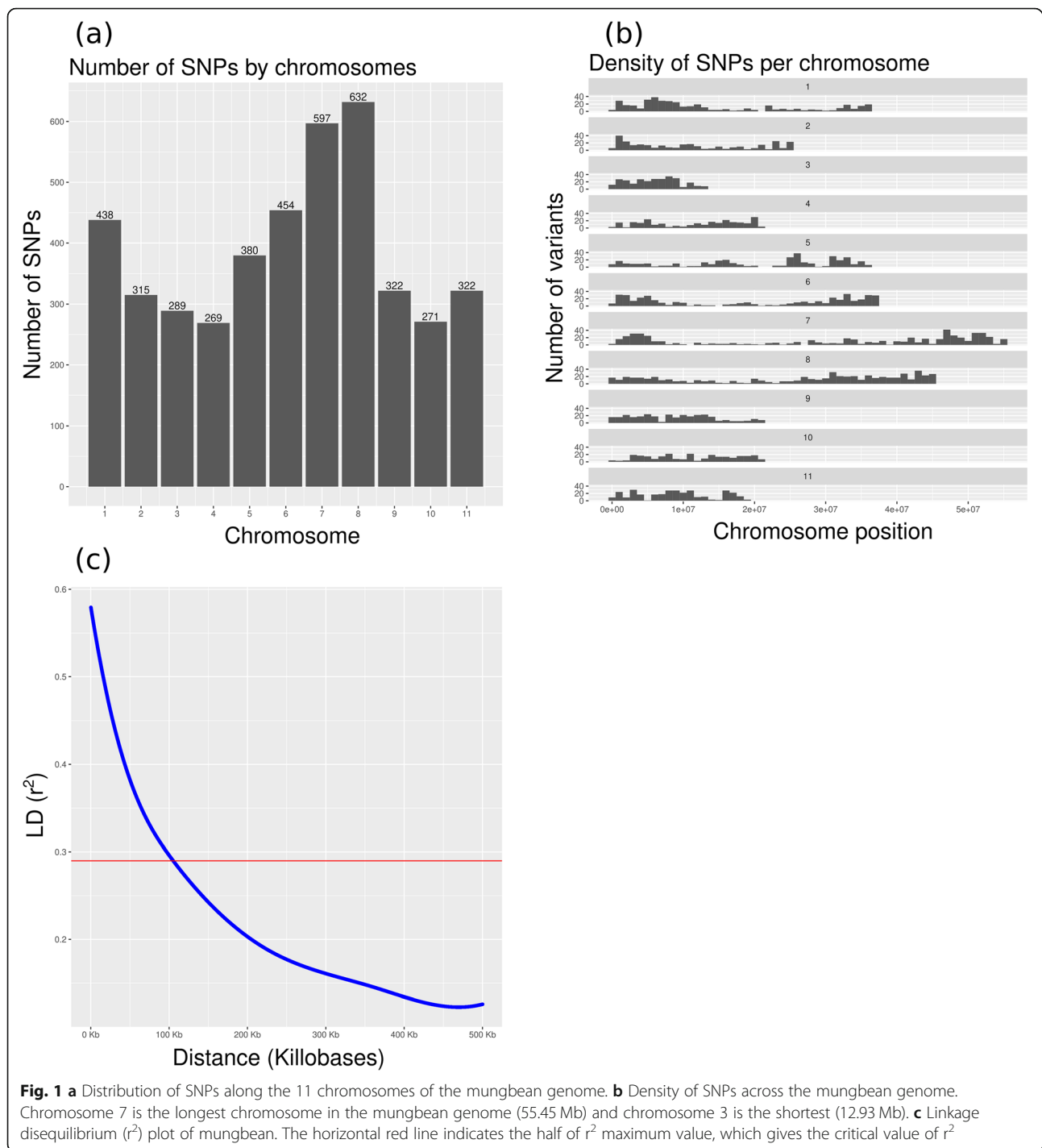
## Results

### Marker polymorphism analysis

Genotyping by Sequencing (GBS) was used to survey polymorphism within the genomes of 293 accessions. SNPs were filtered to retain polymorphisms present in at least 90% of genotypes with a minor allele frequency of at least 3%. Four thousand two hundred eighty-nine polymorphisms are distributed among all chromosomes, but with variable density (from 269 SNPs on chromosome 4 to 632 SNPs on chromosome 8) and 752 SNPs are on scaffolds (Fig. 1a, b).

### Estimation of linkage disequilibrium

Linkage disequilibrium was estimated between 5041 SNPs and the half of  $r^2$  maximum value ( $r^2 = 0.29$ ) was taken as a critical value. The smothering spline of intra-

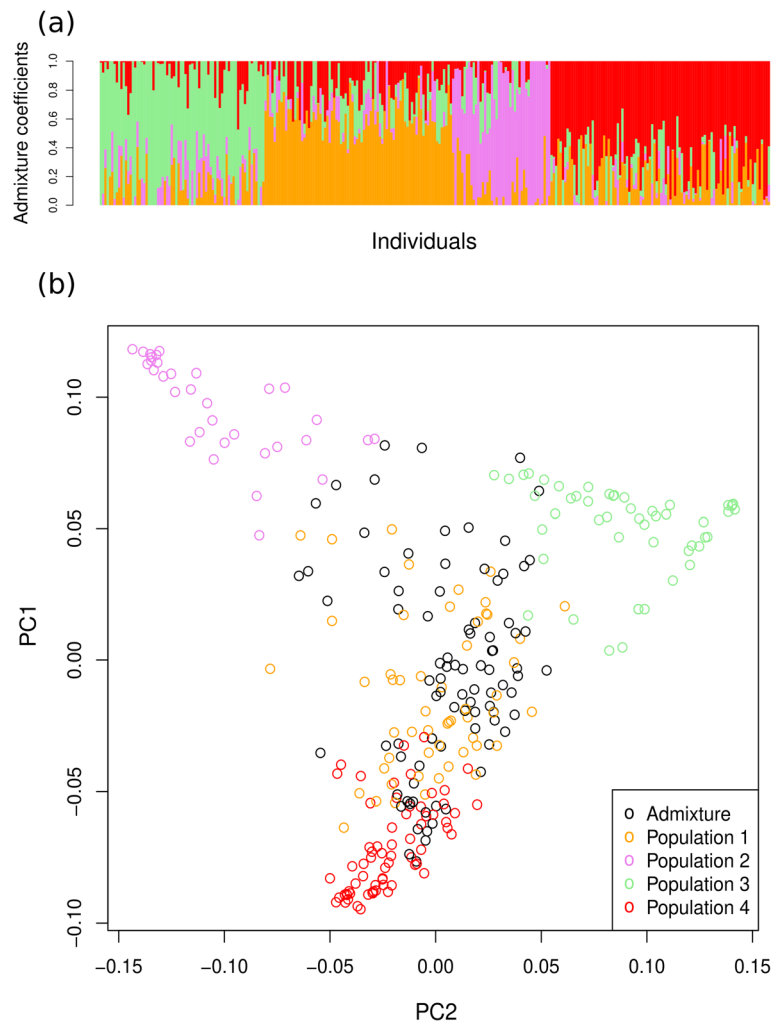


chromosomal  $r^2$  values intersects this threshold at approximately 105 kb physical distance (Fig. 1c).

#### Population structure of the Mungbean mini-core collection

Structure-like Population Genetic Analysis in R package LEA [16] was used to analyze the structure of the population of the 293 accessions from the mini-core collection of

mungbean and a K-value of 4 was determined to best capture the structure of the population based on minimal cross-entropy criterion (Additional file 1: Fig. S1). Using 0.55 as the likelihood to cluster for each accession in the four populations, 204 (70%) accessions were attributed to one of the four populations (Fig. 2a). The first population consists of 54 accessions, the second consists of 33 accessions, the third consists of 50 accessions and the fourth



**Fig. 2** **a** Population structure of 293 mungbean accessions at  $K=4$ . Each vertical bar represents a single accession, the length of each bar represents the proportion contributed by each population. Population 1 is color-coded orange, population 2 is color-coded violet, population 3 is color-coded green and population 4 is color-coded red. **b** Scatter plots of the first two principal components of PCA analysis based on 5041 SNPs. Each dot represents an accession color-coded according to membership (based on > 55% of identity) to populations identified from structure analysis

consists of 67 accessions. The remaining 89 out of 293 accessions were admixed (Additional file 2: Table S1; Additional file 3: Table S2).

Patterns of population differentiation were also analyzed using principle components (PCA). Figure 2b depicts the PCA plot for genetic data of the first versus second components and Additional file 4: Fig. S2 depicts a summary of variation and covariation attributed to the first five principle components that explained 32% of the variance of all markers.

Fixation index ( $F_{ST}$ ) among the four sub-populations was calculated using VCFtools [17]. Populations 1 and 4 are the most closely related ( $F_{ST} = 0.08$ ) and populations 2, 3 and 4 show considerable degree of differentiation ( $F_{ST} = 0.42, 0.43, 0.37$  respectively) (Additional file 5: Table S3).

### Single trait associations

Genetic and phenotypic data (list of phenotypes and units of measurement are in Additional file 6: Table S4) were weakly concordant, as described in Additional file 7: Table S5, which shows proportion of phenotypic variance explained by genotype for traits measured in Kuban. Neither strong positive nor negative correlation was observed between phenotypic traits (Additional file 8: Table S6).

GWAS analysis was implemented with and without the first five PCA axes scores used as covariates for all phenotypic data. The best type of analysis was chosen for each trait separately based on genomic control parameter ( $\lambda_{GC}$ ). The analysis revealed significant associations with two phenotypic traits (possibility of maturation and hypocotyl colour), measured in Kuban

(Fig. 3; Additional file 9: Fig. S3; Additional file 10: Fig. S4; Additional file 11: Table S7).

To annotate the regions in genome where significant SNPs are located, the genome was divided into haplotype blocks (haploblocks) based on linkage disequilibrium (Additional file 12: Table S8).

Four SNPs associated with possibility of maturation in Kuban territory of Southern Russia, localize to a region of 250 kb on chromosome 2 with strong linkage. Identified within this region gene *Vradi02g04380* encodes zinc finger A20 and AN1 domain stress-associated protein. Stress-associated proteins (SAPs), such as A20/AN1 zinc-finger domain-containing proteins play important roles in abiotic stress signaling in plants. SAPs are important candidates for preventing the loss of yield caused by exposure to environmental stresses [18].

## Discussion

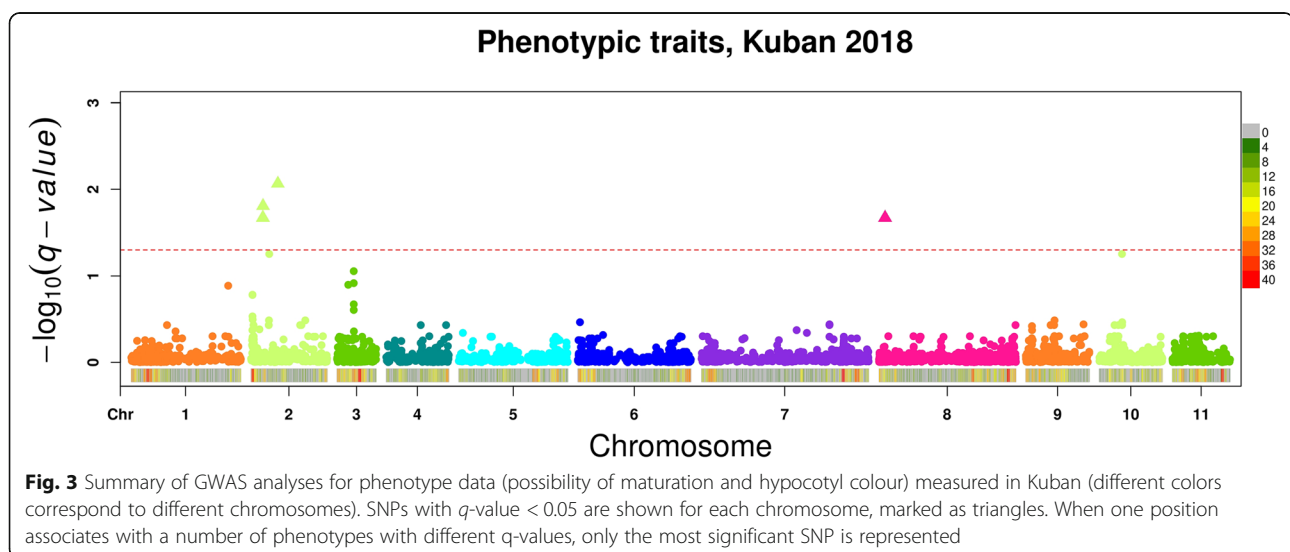
Mungbean, long an orphan legume crop of the tropical world, is emerging as a multi-faceted crop suited to a range of agroecologies. As a rich source of protein and micronutrients that can be consumed in many ways (dahl, soups, sprouts, flour, etc), and with seeds that can be rapidly cooked, it has a place in a number of traditional cuisines across Asia, and in increasingly diverse global palettes. It is a key ingredient in the new vegetarian egg (Just eggs), showing its adaptability as a protein source. Furthermore, with a short duration lifecycle, it fits into a range of crop rotations. Adapting varieties to a range of conditions from tropical to temperate regions is necessary to meet rising global demand.

The core collection of mungbean established by the WorldVeg is a valuable resource for mungbean breeding. Two hundred ninety-three accessions that passed all filters and remained for our analysis were collected from

17 countries all over the world. 70% of accessions were attributed to one of the four populations identified in structure analysis. Populations 1 and 4 consist of accessions from Indian subcontinent (India and Pakistan), they are clustered together as seen in PCA plot (Fig. 2b). Population 2 represents accessions from Middle East and population 3 represents accessions from Southeast Asia and Oceania. In spite of the large degree of admixture, the separation of accessions into four groups on the PCA plot is quite clear (Fig. 2b). According to calculations of  $F_{ST}$  among the four sub-populations, two populations (populations 1 and 4 representing Indian subcontinent) were closely related, whereas others showed considerable degree of differentiation.

LD was estimated between 5041 SNPs and fell to half of the maximal  $r^2$  values at approximately 105 kb physical distance. The extent of LD reflect the effective size of population and the extent of population admixture. It is typically smaller in natural populations, strongly extended in bread cultivars, and intermediate among landraces [19].

Phenology will remain a critical breeding target for mungbean. As demand rises for mungbean, production in temperate regions with shorter growing seasons can help demand. Furthermore, with climate change, seasons in semi-arid tropical regions, such as much of South Asia and East Africa, may also become shorter. In the Kuban region in 2018 abnormally hot weather persisted through the flowering and beginning of maturation phases of some accessions and the development of the accessions was extremely non-uniform. This non-uniformity was elicited by plant exposure to different photoperiod in conjunction with drought, and represents plant response to these incidental stresses.





By performing GWAS on mungbean mini-core collection planted in Southern Russia, we have found several loci significantly associated with maturation and colour of hypocotyl. We found that SNPs associated with possibility of maturation in Kuban localize to a region on chromosome 2 with strong linkage in which genes encoding zinc finger A20 and AN1 domain SAP are located. SAPs are known to prevent the loss of yield caused by environmental stresses [18]. Uncovering SNPs for these traits will speed breeding efforts for production of mungbean suitable for cultivation in temperate regions. Furthermore, understanding correlations phenology with other traits such as seed size or leaf shape, will be necessary to adapt new varieties to changing agroecological conditions.

## Conclusions

Here we perform GWAS in accessions of mungbean, using the mini-core collection established by the World-Veg. The collection has been grown in southern European part of Russia in 2018 under incidental stresses caused by abnormally hot weather and different photoperiod. We have found several loci significantly associated with maturation and colour of hypocotyl. We found that SNPs associated with possibility of maturation in Kuban localize to a region on chromosome 2 with strong linkage, in which genes encoding zinc finger A20 and AN1 domain SAP are located. SAPs are known to prevent the loss of yield caused by environmental stresses. Phenotyping of this collection for maturation traits in temperate climatic locations is important for future breeding applications.

## Methods

### Plant materials

The mungbean (*Vigna radiata*) minicore collection was assembled from accessions held in trust under the multilateral system by the World Vegetable Center genebank as described in [7]. Seed of this collection was received from the World Vegetable Center with a Standard Material Transfer Agreement. Seed of the collection is available to the public (<http://seed.worldveg.org/>) (see Additional file 2: Table S1).

### Phenotyping

Phenotyping of 293 accessions was performed at Kuban experimental station of VIR in 2018.

Kuban experimental station of VIR is located at 45°22' N, 40°37' E in the steppe zone on the Prikubanskaya plain, 80 km from the beginning of the spurs of the Caucasian ridge. The soil is a weakly leached chernozem. The humus content in the surface horizons of the soil is 3.6–4.6%.

The climate near Kuban is temperate continental, with hot summers characterized by insufficient moisture and

by extreme instability of all climatic elements. The average monthly temperature of the coldest month, January, is 2.6 °C, and the warmest, July, 23 °C. The average annual rainfall is 565 mm.

Sowing was carried out on May the 5th. The summer had insufficient rainfall. The average air temperature was 24.4 °C, which is 2.2 °C higher than normal. The absolute maximum of summer air temperature of 39.9 °C was observed in June - during the period of juvenile plant growth. In June, high air temperatures and lack of precipitation led to soil and atmospheric drought. Abnormally hot weather persisted throughout July - during flowering and the beginning of maturation of some accessions. Twenty-eight days with dry winds were noted, of which 18 days in August. Precipitation during the summer was 73 mm, which is 40% of the norm.

The development of the accessions was extremely non-uniform. The first selective harvest of beans to avoid their cracking began in the second decade of July. The last accessions have been harvested at the first decade of October in 116 days after sowing, when night temperatures dropped to 0.3 °C, which is below the optimum for the development of mungbean plants. One hundred twenty-eight accessions did not enter the ripening stage.

The following traits were recorded in the field: hypocotyl color, plant habit, days to 50% flowering, days to first mature pods, days to maturity, plant height, number of pods per plant, pod length, number of seeds per pod, 100 seed weight, plant weight and yield. Only 35 samples matured, thus we took only six traits for further analysis and transformed days to maturity to categorical trait possibility of maturation (whether the sample matured or not). The list of phenotypes and units of measurement are in Additional file 6: Table S4.

### Phenotyping data analyses

Shapiro-Wilk test for normality [20] was implemented to quantitative phenotypic traits. Spearman correlation coefficients of traits were calculated using the “rcorr” function from the “Hmisc” R library [21].

### Genotyping by sequencing (GBS) and SNP calling

Total genomic DNA was extracted from bulked young leaves of three single plants. GBS was performed at Diversity Array Technology P/L (DArT). The sequence data generated were then aligned to the mungbean reference genome sequence, *Vradi\_ver6* [6] to identify single nucleotide polymorphisms (SNPs).

SNPs were further filtered using VCFtools [17] to require minor allele frequency (MAF) > 3% and genotype call-rate > 90%, hence 5041 SNPs in 293 accessions passed all filters remained for further analysis.



### Genetic data analyses

Principal component analysis (PCA) was conducted using the “SNPRelate” R library [22]. Custom scripts in R [23] were used to plot depth and distribution of SNPs on chromosomes.

### Estimation of linkage disequilibrium (LD)

LD was estimated using the squared correlation coefficient ( $r^2$ ) between genotypes. VCFtools [17] was used to calculate intra-chromosomal  $r^2$  values. Only  $r^2$  values for SNPs with the physical distance between markers less than 500 kb were used to plot LD decay graph. LD decay was assessed by fitting a smooth spline of intra-chromosomal  $r^2$  values against the physical distance (bp) between markers in R [23]. The half of  $r^2$  maximum value was taken as a critical value. The threshold beyond which the LD was accepted as real physical linkage was estimated to be  $r^2 = 0.29$ . The intersection of the smothering spline of intra-chromosomal  $r^2$  values with this threshold was considered to be an estimate of the range of LD.

### Estimation of the proportion of phenotypic variance explained by all genome-wide SNPs

The GCTA program [24] was used to estimate the proportion of variance in phenotypes explained by all genome-wide SNPs. First, phenotypic data were normalized. Then, the genetic relationships among individuals from genome-wide SNPs were calculated using GCTA-GRM analysis. And, finally, GCTA-GREML analysis was performed to estimate the proportion of variance in a phenotype explained by all GWAS SNPs (i.e. the SNP-based heritability).

### Analysis of population structure

Structure-like Population Genetic Analysis in R package LEA [16] was used to analyze the structure of the population of the 293 accessions from the mini-core collection of mungbean. Choosing the number of subpopulations is based on a cross-entropy criterion. This criterion is based on the prediction of a fraction of masked genotypes (matrix completion), and on the cross-validation approach. Smaller values of the cross-entropy criterion usually mean better runs. Ten independent runs were performed for each simulated value of  $K$ , ranging from 1 to 10.  $K$ -value for which the cross-entropy curve exhibits a plateau was chosen ( $K = 4$ ). An individual accession with more than 55% identity from a single sub-population was classified as representative of that sub-population. Fixation index ( $F_{ST}$ ) among the four subpopulations was calculated using VCFtools [17].

### Mapping approaches

GWAS analysis was performed using a single-locus linear mixed model, implemented in FaST-LMM toolset

(Factored Spectrally Transformed Linear Mixed Models) [25]. Principal component analysis (PCA) of 5041 SNPs revealed that the first five significant principal components (PCs) explained 32% of the variance of all markers. The LMM model was implemented with and without the first five PCA axes scores used as covariates for all phenotypic data. The best type of analysis was chosen for each trait separately based on genomic control parameter ( $\lambda_{GC}$ ). We used a false discovery rate (FDR) [26] of 0.05 to determine significant trait associated loci separately for each trait. Manhattan plots were constructed using “CMplot” library [27] in R. Annotation of significant associated markers was performed using the Legume information system (LIS) [28] database.

### Estimation of haploblocks

To divide the genome into haplotype blocks (haploblocks) based on linkage disequilibrium, Haploview tools [29] were applied to the set of 4289 SNPs, located on chromosomes. Chromosomal regions with strong linkage were identified using default Haploview parameters (confidence interval for LD [0.7, 0.98]). Each haploblock was considered as the set of SNPs located within a given haploblock and used for annotation of significant associated markers.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12870-020-02579-x>.

**Additional file 1: Figure S1.** Cross-entropy plot for 293 mungbean accessions. X-axis indicates the number of ancestral populations, Y-axis represents the minimal cross-entropy. Ten independent runs were performed for each simulated value of  $K$ , ranging from 1 to 10.  $K$ -value for which the cross-entropy curve exhibits a plateau was chosen ( $K = 4$ ).

**Additional file 2: Table S1.** List of accessions, collection sites and accession's groups. Four populations were identified using Structure-like Population Genetic Analysis in R package LEA. An individual accession with more than 55% identity from a single sub-population was classified as representative of that sub-population. The remaining accessions were considered as admixture.

**Additional file 3: Table S2.** Number of mungbean accessions in populations among the 17 countries of origin.

**Additional file 4: Figure S2.** Scatter plots of the first five principal components of PCA analysis based on 5041 SNPs. Each dot represents an accession. Color-coded is according to membership (based on > 55% of identity) to populations identified from structure analysis.

**Additional file 5: Table S3.** Fixation index ( $F_{ST}$ ) among the four subpopulations.

**Additional file 6: Table S4.** List of phenotypes measured at Kuban in 2018.

**Additional file 7: Table S5.** The proportion of variance in a phenotype explained by all GWAS SNPs (i.e. the SNP-based heritability) for traits measured in Kuban.

**Additional file 8: Table S6.** The correlation coefficients between phenotypic traits.

**Additional file 9: Figure S3.** Summary of GWAS analyses for phenotypic trait Possibility of maturation measured in Kuban. (a) SNP QQ-plot. (b) SNP Manhattan plot (different colors correspond to different

chromosomes). SNPs with  $q$ -value  $< 0.05$  are shown for each chromosome, marked as triangles.

**Additional file 10: Figure S4.** Summary of GWAS analyses for phenotypic trait Hypocotyl color measured in Kuban. (a) SNP QQ-plot. (b) SNP Manhattan plot (different colors correspond to different chromosomes). SNPs with  $q$ -value  $< 0.05$  are shown for each chromosome, marked as triangles.

**Additional file 11: Table S7.** Significant SNPs for phenotypic traits measured in Kuban. Results of GWAS analysis for phenotypic traits measured in Kuban, 2018. Annotation of significant associated markers was performed using Legume information system (LIS) database.

**Additional file 12: Table S8.** Haploblocks inferred by Haploview tools.

## Abbreviations

GWAS: Genome-wide association studies; SAP: Stress associated protein; WorldVeg: World Vegetable Center; SNP: Single nucleotide polymorphism; VIR: N. I. Vavilov All-Russian Institute for Genetic Resources; GBS: Genotyping by Sequencing; PCA: Principal component analysis;  $F_{ST}$ : Fixation index; LD: Linkage Disequilibrium; DaT: Diversity Array Technology P/L; MAF: Minor allele frequency; FaST-LMM: Factored Spectrally Transformed Linear Mixed Models; FDR: False discovery rate

## Acknowledgements

We thank Joyce Yen for description how GBS of mungbean mini-core collection was performed. EVW's crop domestication course provided helpful comments on the manuscript.

## About this supplement

This article has been published as part of BMC Plant Biology Volume 20 Supplement 1, 2020: Selected articles from the 5th International Scientific Conference "Plant genetics, genomics, bioinformatics, and biotechnology" (PlantGen2019). The full contents of the supplement are available online at <https://bmcpplantbiol.biomedcentral.com/articles/supplements/volume-20-supplement-1>.

## Authors' contributions

AS performed GWAS analysis. MB, TV and MV performed dataset phenotyping. RMN established and genotyped the mungbean mini-core collection. RS amplified and provided the mini-core collection. MG, MV, SN, CRL, CTT, RS, and EVW developed the project, oversaw the analysis and project staff, and develop the manuscript outline and analytical framework. AS wrote the manuscript with help from all the authors. All authors contributed to revisions and editing of the manuscript.

## Funding

Design of the study, analysis and interpretation of data (plant growing and phenotyping, genetic diversity analysis and GWAS), as well as writing and publication of this article were supported by Russian Science Foundation grant 18-46-08001. EVW is also supported by USDA Hatch funding through the Vermont State Experimental Station. CRL and CTT are supported by the Ministry of Science and Technology of Taiwan 107-2923-B-002-004-MY3. The collection of the genotypic data was funded by the Australian Centre for International Agricultural Research (ACIAR)-funded International Mungbean Improvement Network (CIM-2014-079) and other core donors to the World Vegetable Center: Republic of China (Taiwan), United States Agency for International Development (USAID), Australian Centre for International Agricultural Research (ACIAR), Thailand, Philippines, Korea, and Japan.

## Availability of data and materials

The sequence data are available from the National Center for Biotechnology database under BioProject PRJNA645721.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Peter the Great St. Petersburg Polytechnic University, Department of Applied Mathematics, St. Petersburg, Russia. <sup>2</sup>Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia. <sup>3</sup>Kuban Branch of Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR), Krasnodar region, Russia. <sup>4</sup>World Vegetable Center, Shanhua, Tainan 74199, Taiwan. <sup>5</sup>National Taiwan University, Taipei 106, Taiwan. <sup>6</sup>World Vegetable Center, South and Central Asia, ICRISAT Campus, Patancheru, Hyderabad, Telangana 502324, India. <sup>7</sup>University of Southern California, Los Angeles, CA 90089, USA. <sup>8</sup>University of Vermont, Burlington, VT 05405, USA.

Received: 29 January 2020 Accepted: 6 March 2020

Published: 14 October 2020

## References

- FAO. The State of Food and Agriculture 2018. Rome: Migration, agriculture and rural development; 2018.
- Establishing the International Mungbean Improvement Network. 2016. <https://avrdc.org/intl-mungbean-network>. Accessed 29 June 2019.
- Chivenge P, Mabhaudhi T, Modi AT, Mafongoya P. The potential role of neglected and underutilised crop species as future crops under water scarce conditions in sub-Saharan Africa. *Int J Environ Res Public Health*. 2015;12(6):5685–711. <https://doi.org/10.3390/ijerph120605685>.
- Bisht IS, Singh M. Asian vigna In: Genetic and Genomic Resources of Grain Legume Improvement. Singh M, Bisht IS, Eds. London: Elsevier Insights; 2013. p. 237–67. <https://www.sciencedirect.com/science/article/pii/B9780123979530001047/via%3Dihub>.
- Burlyayeva M, Vishnyakova M, Gurkina M, Kozlov K, Lee CR, Ting CT, Schafleitner R, Nuzhdin S, Samsonova M, von Wettberg E. Collections of Mungbean [*Vigna radiata* (L.) R. Wilczek] and urdbean [*V. mungo* (L.) Hepper] in Vavilov institute (VIR): traits diversity and trends in the breeding process over the last 100 years. *Genet Resour Crop Evol*. 2019;66(4):767–81.
- Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha B-K, Jun TH, Hwang WJ, Lee T, Lee J. Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat Commun*. 2014;5:5443.
- Schafleitner R, Nair R, Rathore A, Wang Y, Lin C, Chu S, et al. The AVRDC – the world vegetable center mungbean (*Vigna radiata*) core and mini core collections. *BMC Genomics*. 2015;16(1):344. <https://doi.org/10.1186/s12864-015-1556-7>.
- Hwang E-U, Song Q, Jia G, Specht JE, Hyten DL, Costa GL, Cregan PB. A genome wide association study of seed protein and oil content in soybean. *BMC Genomics*. 2014;15:1.
- Bonhomme M, André O, Badis Y, Ronfort J, Burgarella C, Chantret N, Prosperi JM, Briskine R, Mudge J, Debellé F, Navier H. High-density genome-wide association mapping implicates an F-box encoding gene in *Medicago truncatula* resistance to a phanomyces euteiches. *New Phytol*. 2014;201(4):1328–42.
- Michno J-M, Burghardt LT, Liu J, Jeffers JR, Tiffin P, Stupar R, Myers CL. Identification of candidate genes underlying nodulation-specific phenotypes in *Medicago truncatula* through integration of genome-wide association studies and co-expression networks. *BioRxiv*. 2018. <https://doi.org/10.1101/392779>.
- Kang Y, Torres-Jerez I, An Z, Greve V, Huhman D, Krom N, Cui Y, Udvardi M. Genome-wide association analysis of salinity responsive traits in *Medicago truncatula*. *Plant Cell Environ*. 2019 May;42(5):1513–31. <https://doi.org/10.1111/pce.13508>.
- Plekhanova E, Vishnyakova MA, Bulynytsev S, Chang PL, Carrasquilla-Garcia N, Negash K, et al. Genomic and phenotypic analysis of Vavilov's historic landraces reveals the impact of environment and genomic islands of agronomic traits. *Sci Rep*. 2017;7:4816.
- Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, et al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat Genet*. 2019;51:857–64. <https://doi.org/10.1038/s41588-019-0401-3>.
- Xu P, Wu X, Muñoz-Amatriain M, Wang B, Wu X, Hu Y, Huynh BL, Close TJ, Roberts PA, Zhou W, Lu Z. Genomic regions, cellular components and gene

- regulatory basis underlying pod length variations in cowpea (*V. unguiculata* L. Walp). *Plant Biotechnol J*. 2017;15(5):547–57.
15. Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, et al. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat Genet*. 2017;49:1082–8. <https://doi.org/10.1038/ng.3872>.
  16. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes*. 2007;7:574–8.
  17. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
  18. Ben Saad R, Safi H, Ben Hsoua A, Brini F, Ben RW. Functional domain analysis of LmSAP protein reveals the crucial role of the zinc-finger A20 domain in abiotic stress tolerance. *Protoplasma*. 2019;256(5):1333–44. <https://doi.org/10.1007/s00709-019-01390-2>.
  19. von Wettberg EJB, Chang PL, Başdemir F, Carrasquilla-Garcia N, Korbu LB, Moenga SM, et al. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat Commun*. 2018;9(1):649. <https://doi.org/10.1038/s41467-018-02867-z>.
  20. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52(3–4):591–611. <https://doi.org/10.1093/biomet/52.3-4.591>.
  21. Harrell Jr FE. Hmisc: Harrell Miscellaneous. R package version 4.1–1. 2018. <https://CRAN.R-project.org/package=Hmisc>. Accessed 20 Apr 2019.
  22. Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 2012;28(24):3326–8.
  23. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. <https://www.R-project.org/>. Accessed 15 Mar 2019.
  24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82.
  25. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. 2011;8:833–5.
  26. Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann Stat*. 2003;31:2013–35.
  27. CMplot: Circle Manhattan Plot. <https://github.com/YinLiLin/R-CMplot>. Accessed 10 June 2019.
  28. Dash S, Campbell JD, Cannon EK, Cleary AM, Huang W, Kalberer SR, et al. Legume information system (LegumeInfo. Org): a key component of a set of federated data resources for the legume family. *Nucl. Acids Res*. 2016;44:D1181–8.
  29. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–5.

## Publisher's Note

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

**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](https://biomedcentral.com/submissions)

