GWAS between flax accessions and agronomically important phenotypic traits

A. Sokolkova[1], T. Rozhmina[2], A. Kanapin[1], A. Samsonova[1], S. Surkova[1], M. Bankin[1], M. Samsonova[1]

1 Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
2 Federal Research Centre for Bast Fiber Crops, Torzhok, Russia

1. Introduction

Flax is one of the most important crops, both in Russia and in other countries. In modern conditions, fibre flax is considered as a strategic crop in Russia, since it is able to replace raw cotton, which has moved into the category of imported raw materials. Oil flax is used for the production of seeds, which are characterized by valuable nutritional and therapeutic properties. The flax collection created at the Federal Research Centre for Bast Fiber Crops is one of the largest in the world and covers almost all the genetic diversity of the crop. The collection was screened for resistance to Fusarium wilt and highly resistant samples were identified.
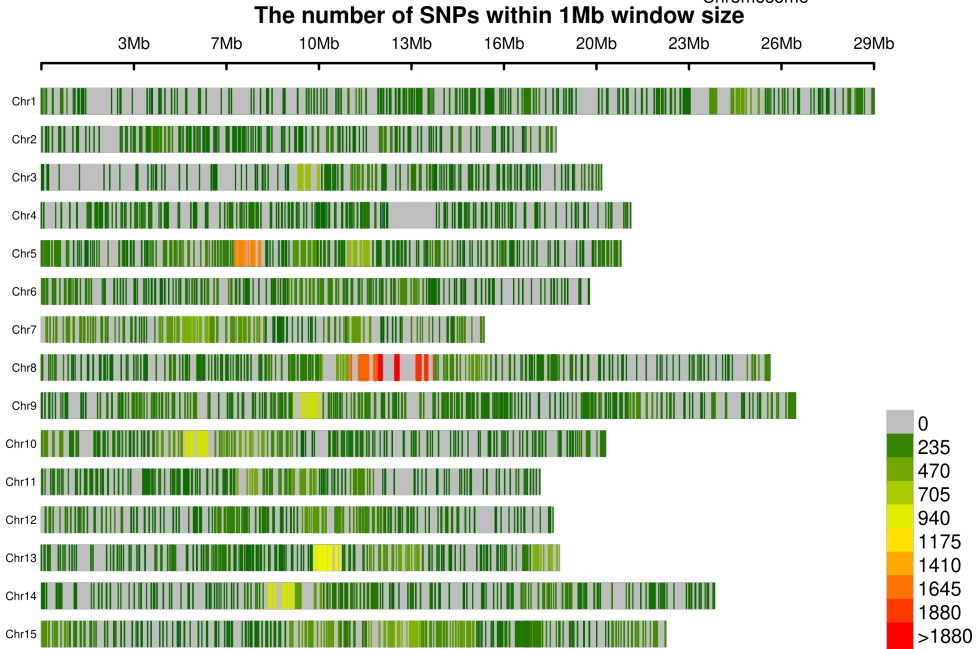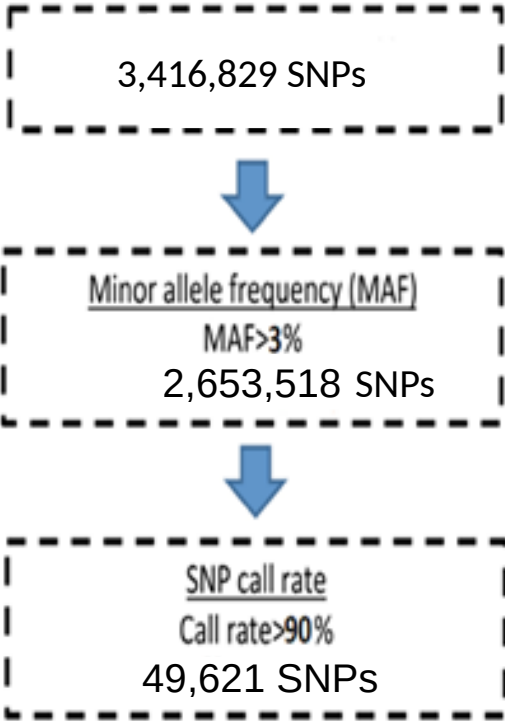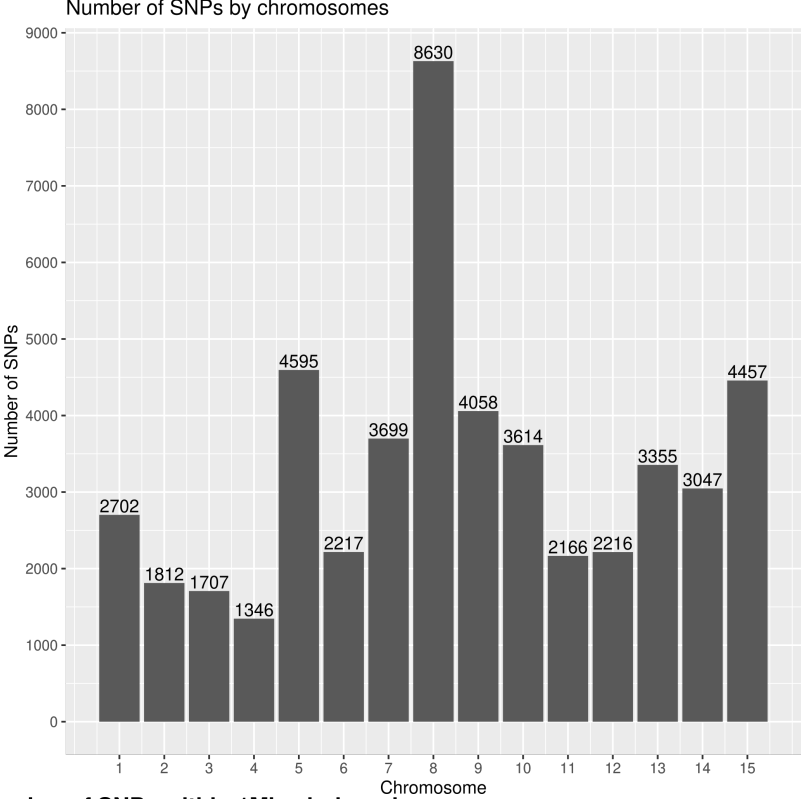Phenotyping of 306 accessions was performed three times at experimental station in Torzhok, Russia: in 2019 and two times in 2020.
Here we present the results of GWAS for 306 accessions from this collection.

2. Materials and Methods

The dataset:

- 306 flax accessions
- 8X depth genome sequencing >> 3,416,829 SNPs
- phenotypic traits: plant height, length of technical part of the stem, technical part weight, inflorescence length, number of bolls, seeds per plant, 1000 seed weight, the diameter of the stem, the number of internodes, distance between internodes, stem slenderness, stem taperingness, ultimate fiber length and resistance to Fusarium wilt

Number of SNPs by chromosomes

3,416,829 SNPs

↓

Minor allele frequency (MAF)
MAF>3%
2,653,518 SNPs

↓

SNP call rate
Call rate>90%
49,621 SNPs

The number of SNPs within 1Mb window size

3. Results

Structure-like Population Genetic Analysis in R package LEA (Falush et al., 2007), based on 3,416,829 SNPs, was used to analyze the structure of the population. A range of sub-populations (K=1:10) were tested and a K-value of 6 was determined to best capture the structure of the population based on minimal cross-entropy. An individual accession with more than 55% identity from a single sub-population was classified as representative of that sub-population.
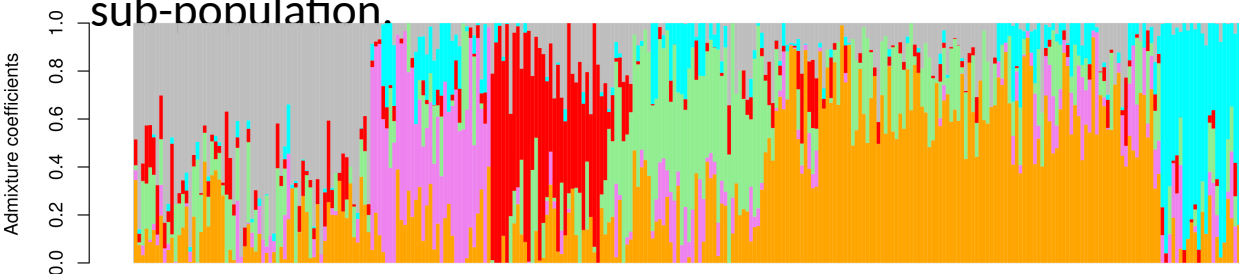
Principle component analysis (PCA), based on 49,621 SNPs, revealed the first 5 significant PCs explained 22,6% of the variance of all markers.
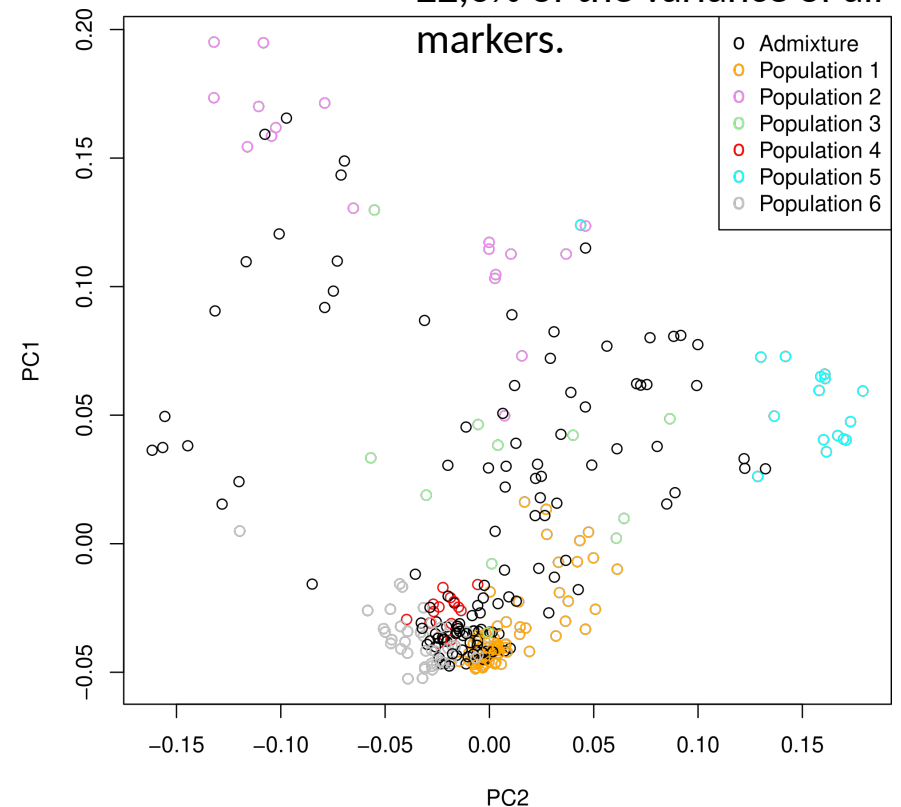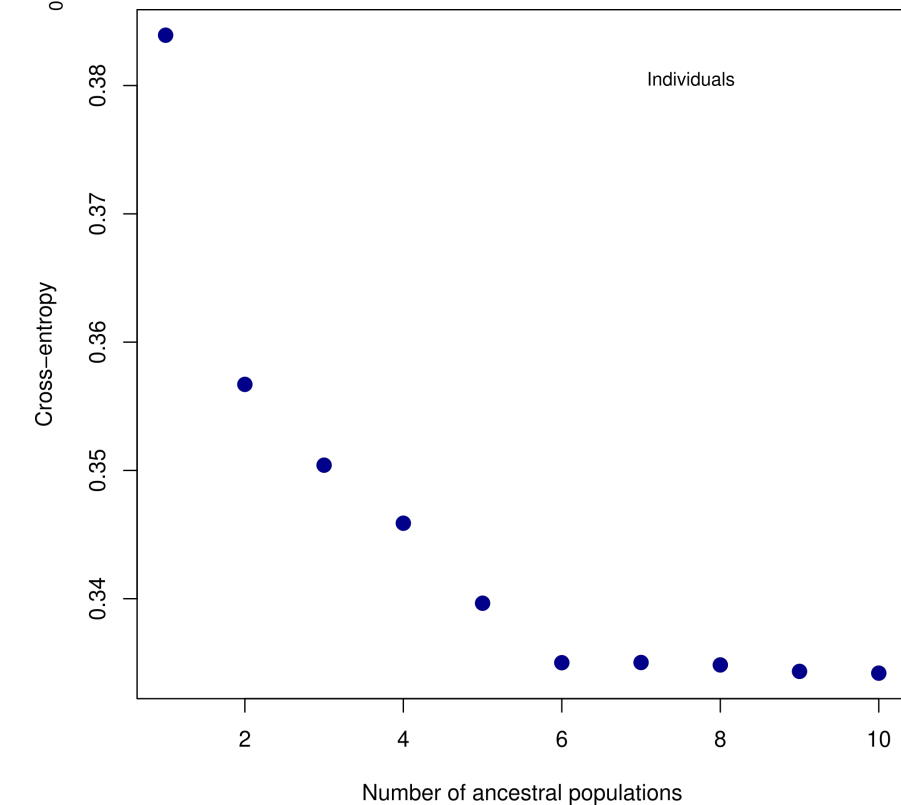
GWAS analysis was performed using the mixed linear model (MLM) with 5 Principal component analysis (PCA) axes scores and Q matrix from analysis of population structure used as covariates for all phenotypic data using TASSEL software (Bradbury et al., 2007).
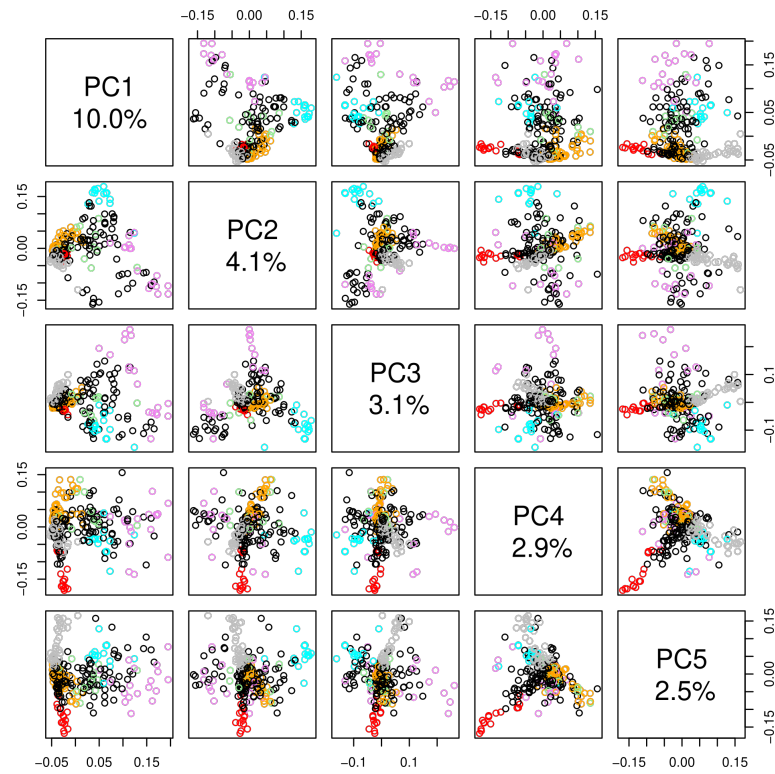
SNP Position	Chromosome	Phenotype	Is in the region of strong linkage	Is in genic region
1288166	1	resistance to Fusarium wilt (2019 and 2020_1)	+	Lus10025717
1288653	1	resistance to Fusarium wilt (2019 and 2020_1)	+	Lus10025717
3541833	4	OilContent (2020_1)	+	-
11649683	6	OilContent (2020_1)	+	-
15247725	8	OilContent (2020_1)	+	-
6717907	11	stem taperingness (2019)	-	-
11470097	12	OilContent (2020_1)	+	-
7329810	13	stem taperingness (2019)	-	-
11074736	15	OilContent (2020_1)	+	-
13965162	15	stem taperingness (2019)	+	-

4. References:

1. Falush D., Stephens M. and Pritchard J.K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7, 574-578.

2. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software forassociation mapping of complex traits in diverse samples. Bioinformatics 2007, 23, 2633–2635.