GENOME-WIDE ASSOCIATION STUDY AND POPULATION STRUCTURE OF DIFFERENT HEMP (*CANNABIS SATIVA* L.) GENOTYPES

Marjeta Eržen[1], Andreja Čerenak[1], Jernej Jakše[2]

[1]*Slovenian Institute of Hop Research and Brewing, Žalec, Slovenia*
[2]*University of Ljubljana, Biotechnical Faculty, Department of Agronomy, Ljubljana, Slovenia*

E-mail: marjeta.erzen@ihps.si

Hemp (*Cannabis sativa* L.) is one of the oldest cultivated plants in the world, with a multipurpose range of uses. It is a dioecious annual plant [1]. Due to wind pollination, heterozygous plants frequently occur within varieties, leading to appearance of different phenotypes, which can manifest at both the chemical and genetic levels.

Based on visual traits, different phenotypes were identified within three different hemp varieties (Carmagnola selected, Tiborszallasi, and Finola selection). DNA of selected phenotypes was isolated and shotgun NGS libraries were designed using an in-house protocol. Libraries were hybridized with 4537 capture probes for target loci enrichment. After enrichment, samples were loaded on the Ion v3 chip and sequenced using the Ion Proton Sequencer. Two chips were used for 171 samples. Raw UBAM reads were imported into CLC Genomic Workbench 22. Reads were trimmed for low quality regions and mapped to the reference genome cs10. SNP calling was performed using the Genome Analysis Toolkit (GATK). Population structure analysis was conducted using ADMIXTURE 1.3.0 [2]. Cross-validation (CV) error was calculated for each group. Genome-wide association studies (GWAS) were conducted using Tassel 5 software [3]. Genotyping data were filtered, resulting in 3670 SNP positions out of 4537. These data were combined with three recorded phenotype traits (CBD, THC, and anthocyanin coloration of leaf petiole). Kinship analysis was performed to minimize false positives. GWAS analysis was done using MLM model, incorporating filtered and combined genotype and phenotype data, along with kinship matrix and PCA results. GWAS results were visualized using Manhattan plots with p-values in RStudio and the 'qqman' package [4]. False discovery rate (FDR) was determined using the 'qvalue' package at thresholds of 0.01, 0.001, or 0.0001. According to ADMIXTURE, genotype data were divided into 5 groups with a minimum CV value of 0.44940. The Carmagnola selected variety was homogeneous with no distinct groups, while Tiborszallasi and Finola selections were divided into two groups. GWAS revealed 14 significant SNPs above the FDR threshold associated with THC. Among them, two were on chromosome 1, one on chromosome 2, two on chromosome 4, one on chromosome 6, two on chromosome 7, four on chromosome 8, and one on chromosome 10, which was also the most significant SNP with a p-value of 0.000135 and an R^2 of 0.3474. For CBD, one significant SNP was associated with

chromosome 10. The p-value was 0.0000795, the marker's effect on the trait was 0.1479, and R^2 was 0.04245. Anthocyanin coloration of leaf petiole was associated with three statistically significant SNPs, all on chromosome 8. The most significant SNP had a p-value of 0.000504, a marker effect of 0.05, and an R^2 of 0.35167. Based on significant SNPs, genes located at defined positions were determined for each phenotypic trait. The GWAS analysis results serves as a valuable tool for understanding the mechanisms and functions of genes correlated with specific hemp traits.

References:

1. Hillig KW. Genetic evidence for speciation in Cannabis (Cannabaceae). Genet Resour Crop Evol 2005: 52:161–80.
2. ADMIXTURE 1.3.0. 2023. https://dalexander.github.io/admixture (accessed May 10, 2023).
3. 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: 2633–35.
4. Turner S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots 2018; https://doi.org/10.1101/005165.