

ASOCIACIJSKE ŠTUDIJE CELOTNEGA GENOMA IN POPULACIJSKA STRUKTURA RAZLIČNIH GENOTIPOV NAVADNE KONOPLJE (*Cannabis sativa L.*)

Marjeta ERŽEN¹, Andreja ČERENAK², Tjaša CESAR³ in Jernej JAKŠE⁴

Izvirni znanstveni članek / Original scientific article

Prispelo / Received: 16. 10. 2023

Sprejeto / Accepted: 6. 12. 2023

Izvleček

Zaradi tujeprašnosti vetrocvetnosti in dvodomnosti konoplje večkrat prihaja do navzkrižnega križanja dveh sort in ker so sorte konoplje populacijske, se znotraj sorte lahko pojavi neizenačenost med posameznimi rastlinami, kar privede do pojava različnih fenotipov. Namen naše raziskave je bil ugotoviti, ali se fenotipi, odbrani na podlagi vizualnih lastnosti, med sabo razlikujejo tudi na genetskem nivoju. Znotraj treh različnih sort Carmagnole selected, Tiborszallasi in selekcije Finole smo odbrali fenotipe glede na vizualne lastnosti rastlin. Iz listov smo izolirali DNA in izdelali knjižnice NGS. Sledila je hibridizacija knjižnic NGS z biotiniliziranimi vabami, obogatitev ulovljenih knjižnice s sondami ter sekvenciranje na napravi Ion Proton. Izvedli smo populacijsko strukturo genotipov in asociacijske študije celotnega genoma (GWAS). Z orodjem Admixure smo določili, da se preučevane sorte lahko razdelijo v največ pet skupin glede na populacijsko strukturo. Carmagnola selected se na genotipskem nivoju ni razlikovala, medtem ko smo lahko sorto Tiborszallasi, in selekcijo Finole razdelili v dve skupini. Na podlagi analize PCA smo na genetski ravni sorte ločili med sabo, medtem ko znotraj sort posameznih fenotipov ni bilo mogoče razlikovati. Na podlagi analize MLM smo analizirali 45 različnih fenotipskih lastnosti, v nadaljevanju pa izpostavljamo samo dve. Pri lastnosti CBD-A določili osem signifikantnih SNP, medtem ko smo pri lastnosti α-pinenu določili 37 signifikantnih SNP. Za vse signifikantne pozicije smo določili tudi gene.

Ključne besede: *Cannabis sativa L.*, genotipi, GWAS, konoplja, populacijska struktura

¹ Mag. inž. agr., Inštitut za hmeljarstvo in pivovarstvo Slovenije (IHPS), e-pošta: marjeta.erzen@ihps.si

² Izs. prof. dr., IHPS, e-pošta: andreja.cerenak@ihps.si

³ Mag. biotehnol., Univerza v Ljubljani, Biotehniška fakulteta (UL BF), e-pošta: tjasja.cesar@bf.uni-lj.si

⁴ Prof. dr., UL BF, e-pošta: jernej.jakse@bf.uni-lj.si

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

Abstract

In the context of hemp, due to cross-pollination and dioecy, crossbreeding of two varieties often occurs. Since hemp varieties are population-based, differences among individual plants can arise within a single variety, leading to the emergence of different phenotypes. The aim of our research was to determine whether phenotypes selected based on visual characteristics also differ at the genetic level. Within three different varieties, namely Carmagnola selected, Tiborszallasi, and Finole selection, we selected phenotypes based on visual plant traits. We isolated DNA from the leaves and constructed NGS libraries. Hybridization of NGS libraries with biotinylated probes, enrichment of captured libraries with probes, and sequencing on the Ion Proton platform were performed. Population structure of genotypes and whole-genome association studies (GWAS) were conducted. Using the Admixture tool, we determined that all three examined varieties could be divided into up to five groups based on population structure. Carmagnola selected did not show genotypic differentiation, whereas both Tiborszallasi and Finole selections could be divided into two groups. Genetic differentiation between varieties was observed through PCA analysis, but differentiation between individual phenotypes within varieties was not possible. Through MLM analysis 45 different phenotype traits were analysed. In this article two traits will be exposed. Eight significant SNPs were identified for the CBD-A trait, while 37 significant SNPs were identified for the α-pinene trait. Genes were also assigned to all significant positions.

Keywords: *Cannabis sativa L.*, genotypes, GWAS, hemp, population structure

1 UVOD

Konoplja (*Cannabis sativa L.*), ena izmed najstarejših gojenih rastlin na svetu (Andre in sod., 2016), je najbolj prepoznavna zaradi psihoaktivnega učinka kanabinoida THC. THC je eden od več kot sto kanabinoidov, ki imajo pomembno vlogo pri zdravljenju različnih bolezni (Namdar in sod., 2018). Drugi najpomembnejši sekundarni metaboliti so komponente eteričnega olja, odgovorne za vonj rastline in imajo sinergistični učinek s kanabinoidi (Aizpurua-Olaizola in sod., 2016).

Konoplja je dvodomna in tujeprašna rastlina, kar lahko povzroči navzkrižno križanje dveh rastlin (Clarke in Merlin, 2013). Zaradi populacijskih sort pa prav tako lahko prihaja do neizenačenosti med rastlinami in pojava različnih fenotipov znotraj sorte (Barcaccia in sod., 2020).

Konopljin genom pri ženskih rastlinah meri 818 Mb, pri moških pa 843 Mb. Sestavljen je iz devetih parov avtosomnih kromosomov in enega para spolnih kromosomov (van Bakel in sod., 2011).

Prvi osnutek konopljinega genoma so leta 2011 objavili (van Bakel in sod., 2011) za sorto Purple Kush. Leta 2018 pa so (Grassa in sod., 2018) objavili genom sorte cs10 na ravni kromosomov. Z razvojem naprednejših in zmogljivejših tehnik sekvenciranja je mogoče sekvencirati daljša zaporedja ter veliko število zaporedij v kratkem času, pri nižjih stroških, večji kakovosti in manjši možnosti napake (Ambardar in sod., 2016). Z novejšimi tehnikami sekvenciranja so se razvijala tudi nova bioinformacijska orodja, hkrati pa je bilo omogočeno *de novo* sekvenciranje genomov kateregakoli organizma (Kulski, 2016). Sodobne tehnologije večinoma temeljijo na skupnih korakih, kot so: fragmentacija DNA, ligacija adapterjev ter sekvenciranje fragmentov.

Kar nekaj raziskav je že bilo narejenih z namenom ugotavljanja genetske raznolikosti predvsem med različnimi kultivarji konoplje pa tudi med podvrstami (Kojoma et al., 2002; Laverty et al., 2019; Soler et al., 2017; Grassa et al., 2021). Namen naše raziskave je bil ugotoviti populacijsko strukturo različnih genotipov navadne konoplje in ugotoviti ali so razlike vidne na fenotipski ravni vidne tudi na genetski ravni. Z asociacijskimi študijami celotnega genoma (GWAS) pa smo želeli opredeliti gene, ki se nahajajo na signifikantnih mestih za fenotipski lastnosti (CBD-A in α-pinien) ter jih z njima povezati.

2 MATERIALI IN METODE

2.1 Odbira fenotipov glede na vizualne lastnosti

Znotraj treh različnih sort konoplje (Carmagnola selected, Tiborszallasi in selekcija Finole) smo glede na vizualne lastnosti posameznih rastlin znotraj sort, ki smo jih spremljali skozi celotno rastno dobo, konec rastne dobe odbrali različne fenotipe (Eržen in sod., 2023). Odbirali smo jih glede na velikost rastline, barvo rastline, razvejanost, dolžino in kompaktnost socvetij, velikost listov in antociansko obarvanje listnih pecljev. Znotraj sorte Carmagnola selected smo določili dva fenotipa, in sicer C1 in C2, znotraj sorte Tiborszallasi pet fenotipov: T1, TII, TIII, TIV, TV, ter znotraj selekcije Finola štiri fenotipe: F1, FII, FIII, FIV. Da smo ohranili rastline ne opršene, smo enkrat tedensko posevezek pregledali ter izločili rastline moškega spola. Za izolacijo DNA smo pri vsakem fenotipu vzorčili do 20 rastlin na fenotip. Liste smo po vzorčenju zamrznili na -20 °C.

2.2 Izdelava knjižnice NGS

DNA smo prvo izolirali s pomočjo CTAB po hišnem protokolu (Kump in Javornik, 1996). Izolirano DNA smo z ultrazvočno kopeljo razbili na fragmente povprečne dolžine od 250 bp do 350 bp. Soniciranje je trajalo 30 minut. Uspešnost soniciranja smo preverili na agaroznem gelu, nato pa vzorce izpareli do 40 µL z evaporatorjem. Sledilo je popravilo koncev fragmentov pri čemer smo uporabili µL 10X T4 DNA ligaznega pufra (New England Biolabs, ZDA), 2 µL dNTP (10mM), 1 µL T4 polinukleotidne kinaze (10 U/µL) (New England Biolabs, ZDA), 0,33 µL T4 polinukleotidne polimeraze (3U/ µL) (New England Biolabs, ZDA), 6,76 µL

deionizirane vode ter 40 µL vzorca. Reakcijo smo termostatirali v cikličnem termostatu pri 25 °C (25 min) in 12 °C (10 min) in ohladili na 4 °C. Reakcijo smo očistili z magnetnimi delci MagSi-NGSprep Plus (Magtivio, Nizozemska) (1,8-kratni volumen vzorca) in odstranili supernatant ter dodali 180 µL 70-odstotnega etanola. Fazo čiščenja smo ponovili še dvakrat. Po čiščenju smo izvedli ligacijo adapterjev. Uporabili smo naslednjo mešanico: 20 µL očiščenega vzorca, 3 µL 10X T4 DNA ligaznega pufra (New England Biolabs, ZDA), 0,42 µL P1 adapterja, 0,42 µL barkodnih oligonukleotidov, ki so bili za vsak vzorec drugi, 3 µL dNTP (10mM), 1 µL T4 DNA ligaze (New England Biolabs, ZDA), 0,66 µL deionizirane vode in 1,5 µL Bst 2.0 WarmStart DNA polimeraze (New England Biolabs, ZDA). Reakcijo smo termostatirali v cikličnem termostatu pri 22 °C (30 min) in pri 50 °C (20 min) ter ohladili na 4 °C. Reakcijo smo ponovno očistili z magnetnimi delci in 70-odstotnim etanolom. Sledilo je pomnoževanje knjižnice NGS. Uporabili smo mešanico PCR 10 µL očiščene knjižnice, 10 µL 5X KAPA HiFi Fidelity pufra (Roche, Švica), 1,5 µL dNTP (10 mM), 2,5 µL P1amp začetnega oligonukleotida, 2,5 µL T_PCR_A začetnega oligonukleotida, 22,5 µL deionizirane vode in 2 µL HotStart DNA polimeraze (Roche, Švica). Pomnoževanje knjižnice je potekalo v cikličnem termostatu z začetno denaturacijo pri 95 °C 3 minute, sledilo je 8 ciklov pri 98 °C (20 s), 60 °C (30 s) in 72 °C (30 s), zaključno podaljševanje je potekalo 1 minuto pri 72 °C, sledilo je ohlajanje na 4 °C. Produkte PCR smo ponovno očistili z magnetnimi delci in 85-odstotnim etanolom. DNA smo raztopili v 20 µL TE pufra. Sledila je kvantifikacija z napravo Agilent 2100 DNA Bioanalyzer in uporabo komercialnega kompleta DNA 1000 (Agilent Technologies Inc., ZDA).

Koncentracijo DNA smo izmerili s qPCR, pri čemer smo uporabili komplet KAPA Library Quantification Kit for Ion Torrent Platforms (Kapa Biosystems, ZDA). Knjižnice smo nato ekvimolarno združili v štiri reakcije do skupno 2000 ng DNA na reakcijo ter nato volumen skupne knjižnice izpareli z evaporatorjem do končnega volumna 7 µL.

2.3 Hibridizacija knjižnice NGS

Pripravili smo hibridizacijsko mešanico z biotiniliziranimi vabami (MyBites, Arbor Bioscience, ZDA). To smo inkubirali v cikličnem termostatu 10 minut pri 60 °C ter na sobni temperaturi 5 minut. Hibridizacijsko mešanico smo alikvitirali v štiri tubice v mirkocentrifugirki po 18,5 µL. Sledila je priprava mešanice začetnih oligonukleotidov, s katerimi hibridiziramo adapterje ter preprečimo nastanek konkaternerov. Tudi te smo po 5 µL alikvitirali v štiri tubice v mikrocentrifugirki in dodali 7 µL knjižnice v vsako tubico. Sledila je inkubacija v cikličnem termostatu pri 95 °C 5 minut, nato smo temperaturo spustili do 65 °C in v ciklični termostat dodali stripe s hibridizacijsko mešanico. Inkubacija je potekala 5 minut, nato pa smo 18 µL hibridizacijske mešanice prenesli v PCR stripe z začetnimi oligonukleotidi. Sledila je ponovna inkubacija pri 65 °C, 16 ur. Po inkubaciji smo pripravili streptavidinske magnetne kroglice (MyBites, Arbor Bioscience, ZDA), ki smo jih po 120 µL alikvitirali v štiri tubice v mirkocentrifugirki, dali na magnet in odstranili supernatant. Dodali smo 800 µL vezavnega pufra, ponovno dali na magnet, odstranili supernatant in postopek

ponovili še dvakrat. Sprane streptavidinske magnetne kroglice smo raztopili v 280 µL vezanega pufra ter jih alikovtirali v štiri tubice v mikrocentrifugirko po 70 µL. Kroglice smo segrevali na 65 °C 2 minuti ter nato v vsako tubico v mikrocentrifugirki k kroglicam dodali 30 µL ulovljene hibridizirane reakcije. Sledila je 5 minutna inkubacija pri 65 °C ter štirikratno spiranje z 180 µL Wash X pufrom. Reakcijo smo raztopili v 30 µL pufra E in pomnožili v cikličnem termostatu, pri čemer smo uporabili komercialni komplet KAPA HiFi HotStart ReadyMix (Kapa Biosystems, ZDA). Denaturacija je potekala 2 minuti pri 98 °C, sledilo je 10 ciklov pri 98 °C (20 s), 60 °C (30 s), 72 °C (30 s). Zaključno podaljševanje je potekalo 5 minut pri 72 °C. Reakcijo smo ohladili na 8 °C. Sledil je prenos supernatanta v nov strip ter čiščenje z magnetnimi kroglicami MagSi-NGSprep Plus. Reakcijo smo nato raztopili v 20 µL TE pufra. Koncentracijo reakcije smo določili s pomočjo qPCR, po postopku, ki je opisan zgoraj.

2.4 Sekvenciranje

Hibridizirane oz. ulovljene knjižnice s sondami smo združili v eno reakcijo s skupno koncentracijo 100 pM. Reakcijo smo nato pomnožili s sistemom Ion Onetouch 2 System (Thermo Fisher Scientific, ZDA), pri čemer je pomnoževanje potekalo 4 ure in 50 minut. Z uporabo kroglic Dynabeads My One Streptavidin C1 (Thermo Fisher Scientific, ZDA) smo obogatili vezane fragmente na kroglicah. Za obogatitev smo uporabili napravo Ion OneTouch ES (Thermo Fisher Scientific, ZDA). Obogatitev je potekala 37 minut. Reakciji smo dodali delce Ion PI Controle Ion Sphere Particles (Thermo Fischer Scientific, ZDA) in sekvenčne oligonukleotide. Postopek obogatitve je bil izveden po protokolu (Ion PI™ ..., 2017a). Sledila je priprava sekvenatorja, ki je potekal po protokolu (Ion PI™ ..., 2017b), ter nalaganje vzorca na sekvenčni čip Ion PI Chip v3 (Thermo Fisher Scientific, ZDA), ki smo mu še prej dodali 6 µL Ion PI™ Hi-Q Sequencing Polymerase (Thermo Fischer Scientific, ZDA) in 60 µL 50 % pufra za prileganje, ter zagnali sekvenciranje.

2.5 Analiza sekvenčnih podatkov

Po sekvenciranju smo surove sekvenčne podatke uvozili v program CLC Genomics Workbench 22. Podatke smo očistili slabih regij, izločili prekratka zaporedja ter izvedli analizo kvalitete podatkov. Nato smo na referenčni genom konoplje cs10 (GCA_900626175.2) kartirali sekvencirane odčitke. Sledil je postopek določanja SNP, za kar smo uporabili orodje GATK (angl. Genome Analysis Toolkit), pri čemer smo dobili končno datoteko VCF. Nato smo vzorce genotipizirali na podlagi 4537 analiziranih pozicij SNP.

2.6 Populacijska struktura genotipov

Za določanje populacijske strukture smo uporabili orodje Admixture 1.3.0. Genotipske podatke v obliki datoteke VCF smo pretvorili v obliko Plink in z orodjem Admixture pognali ukaz, za katerega smo predvidevali razporeditev genotipov v najmanj eno skupino in največ 20 skupin. Za vsako skupino je bila izračunana napaka

navzkrižnega preverjanja (CV) (angl. cross-validation error). Skupina z najnižjo CV vrednostjo je nato določila število populacij. Za vizualen prikaz razporeditve genotipov smo uporabili orodje Structurly 0.1.0.

2.7 Asociacijske študije celotnega genoma (GWAS)

Za izvedbo asociacijskih študij celotnega genoma smo uporabili orodje Tassel (različica 5.2.88). V program Tassel smo uvozili datoteko VCF z genotipskimi podatki. Manjkajoče genotipske podatke iz datoteke smo nadomestili z ukazom 'LD KNNi Imputation'. Ta temelji na kombinaciji algoritma k-najbližjih sosedov in neuravnovezenosti vezave. Sledilo je filtriranje genotipskih podatkov, kjer smo pogostost redkega alela določili pri 0,05, najvišjo porazdelitev heterozigotnosti pri 1,0, odstranili smo insercije in delecije ter SNP z majhno statistično verjetnostjo.

Sledila je izvedba metode glavnih komponent (PCA), pri čemer smo število komponent nastavili na 5. Na podlagi prvih dveh komponent smo s pomočjo programa R Studio izrisali grafe PCA.

Sledila je analiza linearnega mešanega modela (MLM), s katero smo opravili končno asociacijsko študijo celotnega genoma (GWAS), pri čemer smo 171 genotipskih podatkov združili z dvema fenotipskima podatkoma, in sicer CBD-A in α-pineno. Na podlagi združenih podatkov in analize Kinship smo izvedli analizo linearnega mešanega modela (MLM). Podatke smo vizualno prikazali v programu R studio z grafi tipa Manhattan, pri čemer smo z namenom zmanjšanja možnosti lažno pozitivnih vrednosti določili mejo FDR (angl. False Discovery Rate). Meja FDR je bila določena z ukazom 'qvalue'. Na podlagi p-vrednosti pridobljenimi z analizo MLM smo s paketom 'qqman' vizualno prikazali podatke z uporabo grafov tipa Manhattan.

Na podlagi pridobljenih signifikantnih pozicij SNP smo poiskali gene, ki se za posamezne lastnosti nahajajo na teh pozicijah.

3 REZULTATI Z RAZPRAVO

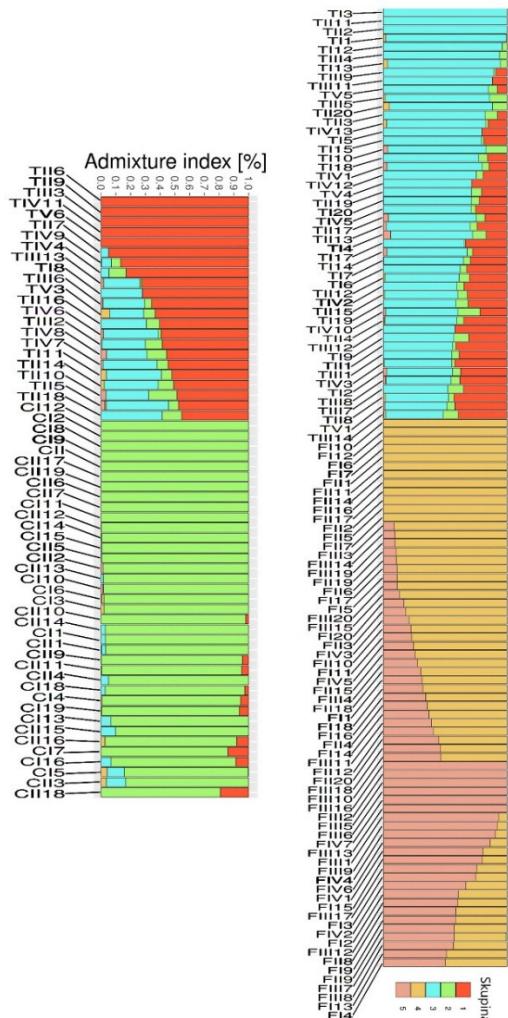
Za sekvenciranje 171 vzorcev smo uporabili dva čipa. Pri prvem čipu smo bili pri nalaganju na čip 90 % uspešni, povprečna dolžina odčitkov je bila 193 bp, skupno pa smo pridobili 87.961.831 uporabnih odčitkov, kar predstavlja 66 %. Pri drugem čipu smo bili 87 % uspešni. Povprečna dolžina odčitkov je bila 132 bp, pridobili pa smo 85.279.421 uporabnih odčitkov, kar, tako kot pri prvem čipu, predstavlja 66 %.

3.1 Populacijska struktura genotipov

Na podlagi genotipskih podatkov smo z orodjem Admixture razdelili v pet skupin, saj je imela skupina 5 najnižjo CV vrednost, in sicer 0,44940.

Vseh pet skupin smo vizualizirali s programom StructuRly. Skupine so prikazane na Sliki 1. Glede na genotipske podatke lahko rečemo, da je sorta Carmagnola selected na genetskem nivoju zelo homogena, saj celotna spada v eno skupino, kljub temu, da smo jo na fenotipskem nivoju lahko razdelili v dve skupini. Sorta Tiborszallasi in

selekcija Finole pa sta manj homogeni, saj lahko vsako sorte razdelimo v dve skupini, pri čemer so fenotipi znotraj posameznih sort naključno razdeljeni po skupinah. Kljub temu, da je konoplja populacijska sorta, so fenotipi znotraj sort dovolj izenačeni. Glede na barve, ki označujejo posamezne skupine, lahko vidimo, da imata sorte Carmagnola selected in Tiboszallasi nekaj podobnosti, medtem ko se selekcija Finole popolnoma loči od teh dveh sort, saj ne najdemo nobene sorodnosti s sortama.



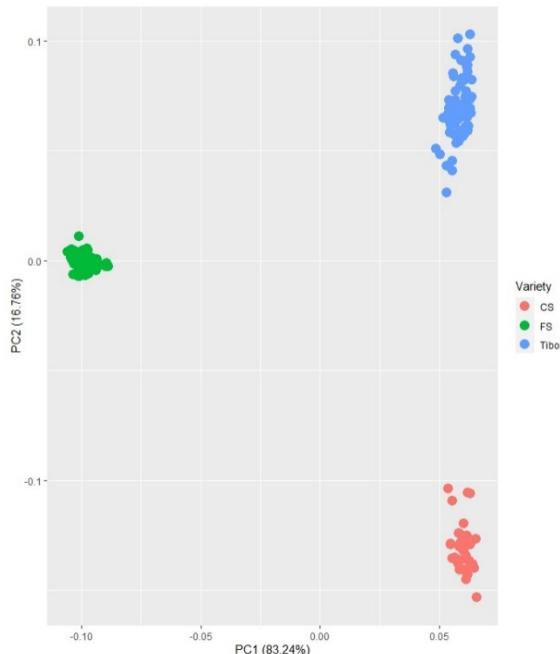
Slika 4: Populacijska struktura 171 različnih genotipov narejena na podlagi treh različnih sort konoplje, ki smo ji lahko na podlagi vizualnih lastnosti razdelili na različne fenotipe znotraj posamezne sorte. Rdeča in modra barva predstavlja sorte Tiborszallasi, zelena barva predstavlja sorto Carmagnola selected, rumena in bež barva pa predstavlja selekcijo Finole.

Homogenost sorte Carmagnola selected lahko pojasnimo s tem, da izvira iz stare italijanske sorte Carmagnola, katero so kot izvorno rastlino uporabljali pri žlahtnjenju številnih novih sort (Clarke in Merlin, 2013). Večja homogenost sorte Carmagnola selected se vidi tudi v tem, da smo pri tej sorti lahko določili je dva fenotipa, medtem ko smo pri ostalih dveh sortah določili štiri oz. pet fenotipov. V svoji raziskavi so sicer Sawler in sod. (2015) ugotovili, da je industrijska konoplja bolj heterogena od indijske konoplje ter da se podvrsti na genetski ravni signifikantno ločita med sabo. Heterogenost industrijske konoplje je odraz širšega genetskega bazena, prav tako pa je pri indijski konoplji bolj pogosto žlahtnjenje med bližnjimi sorodniki.

3.2 Asociacijske študije celotnega genoma (GWAS)

Po filtrirjanju podatkov v programu Tassel smo od skupno 4537 pozicij SNP dobili 3670 pozicij SNP.

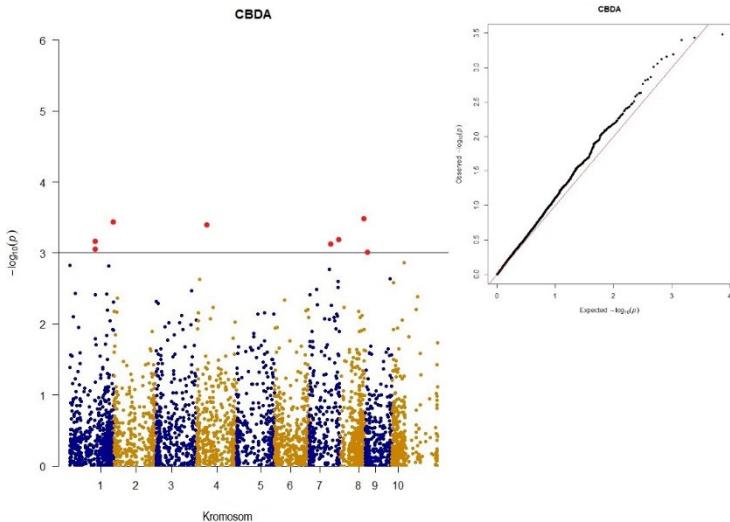
Z izvedbo analize PCA in vizualizacijo podatkov smo ugotovili, da se vse tri sorte popolnoma ločijo med sabo. Na Sliki 2 je prikazan graf analize PCA glede na vse tri sorte (Carmagnola selected, Tiborszallasi in selekcija Finole). Komponenta PC1 pojasni 83,41 % variance, komponenta PC2 pa 16,76 %.



Slika 5: Graf analize PCA na podlagi sort Carmagnola selected, Tiborszallasi in selekcije Finole.

Fenotipov znotraj posameznih sort ni bilo mogoče razlikovati.

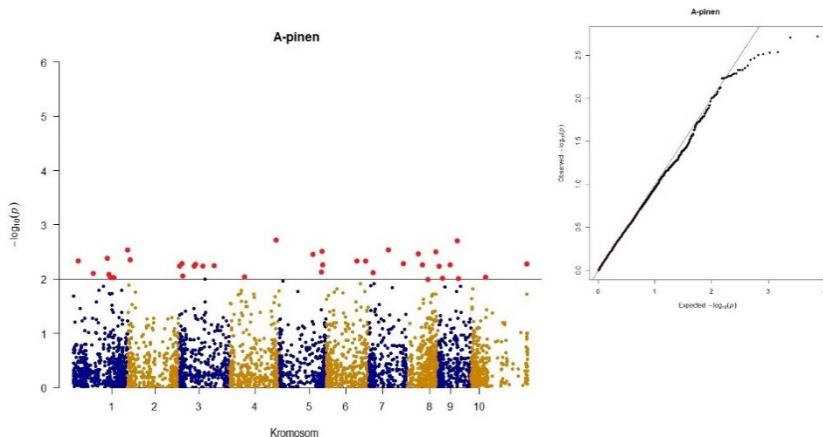
Asociacijske študije celotnega genoma smo izvedli na podlagi dveh fenotipskih podatkov vseh treh sort konoplje, in sicer na podlagi vsebnosti CBD-A in vsebnosti α -pinena. Na podlagi MLM analize in p-vrednosti posameznih signifikantnih pozicij SNP smo izrisali grafe tipa Manhattan za vsako lastnost.



Slika 6: Graf tipa Manhattan za lastnost CBD-A. Različne barve ločijo kromosome med seboj.

Lastnost CBD-A je imela nad statistično značilno FDR mejo 8 signifikantnih SNP: tri na kromosому 1, en na kromosomu 4, dva na kromosomu 7, en na kromosomu 8 in en na kromosomu 9. Na signifikantnih pozicijah SNP kromosoma 1 se nahaja protein TSS, ki vsebuje domeno Clu (angl. protein TSS (Clu domain-containing protein), ta ima vlogo pri organizaciji organelov, dvokomponentni odzivni regulator- verjetni PRP 95 (angl. two-component response regulator-like PRR95), ki ima vlogo pri fosforilaciji in oligomerna kompleksna Golgijeva podenota 4 (angl. conserved oligomeric Golgi complex subunit 4-like), ki zagotavlja normalno delovanje Golgijevega aparata. Na signifikantnih pozicijah SNP na kromosomu 4 se nahaja verjetna glukan 1,3-alfa glikozidaza iz družine glikozidne hidrolaze, ki vsebuje domeno proteina N-terminal (angl. probable glucan 1,3-alpha-glucosidase- Glycoside hydrolase family 31 N-terminal domain-containing protein), ta sodeluje pri metabolnih procesih ogljikovih hidratov. Na kromosomu 7 najdemo encimu podoben barberinski most 15 (angl. berberine bridge enzyme-like 15), ki ima oksidoreduktazno aktivnost, v enako družino kot barberinski most pa spadajo tudi kanabinoidi CBD-A, THC-A in CBC-A (Onofri in sod., 2015). Na kromosomu 8 se nahajata dva gena, in sicer verjeten transkripcijski faktor WRKY 72 (angl. probable WRKY transcription factor 72), ta je vključen v rast in razvoj rastlin ter je vključen v stresni odziv. Verjetno pa deluje tudi kot represor ali aktivator stresnega odziva. Matoušek in sod. (2016) so pri hmelju

našli transkripcijske faktorje iz družine WRKY, ki imajo precej verjeten učinek na biosintezo lupulina. Na kromosomu 8 se nahajata tudi protein IQ-DOAIN 30 (angl. protein IQ-DOMAIN 31) ter beta-glukozidazi podoben SFR2, kloroplastičen gen (angl. beta-glucosidase-like SFR2, chloroplastic).



Slika 7: Graf tipa Manhattan za lastnost α -pinen. Različne barve ločijo kromosome med seboj.

Za lastnost α -pinen je bilo nad statistično značilno mejo FDR določenih 37 signifikantnih SNP. Pri določanju genov smo določili šest neopredeljenih genov. Na kromosому 1 smo določili F-box ponavljajoči protein (angl. F-box/kelch-repeat protein At4g39560), ribosomalni biogenezni protein NOP53 (angl. ribosome biogenesis protein NOP53), aneksin D3, ki se veže na kalcij in se odziva na stres, DEAD box ATP odvisno RNA helikazo 31 (angl. DEAD-box ATP-dependent RNA helicase 31), ki katalizira odvijanje DNA ali RNA, protein RBL, ki sodeluje pri prenosu signala. Na kromosому 2 se nahaja nifU podoben protein 2 (angl. nifU-like protein 2, chloroplastic), ki regulira sestavo železovih in žveplovih skupin. Na kromosому 3 smo določili rodanezi podobno domeno proteina 11 (angl. rhodanese-like domain-containing protein 11, chloroplastic), ki ima ob prisotnosti aktivne strani cisteina verjetno sulfotransferazno aktivnost, receptorju S podobna serin/threoninska protienska kinaza At4g27290 (angl. G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290), ki je vključena v abiotski stres in rastlini omogoča prilagajanje na spreminjačoče okoljske faktorje, protein IQ-domena 0, T-kompleksni protein 1 (angl. T-complex protein 1 subunit delta), šaperon, ki pomaga pri zvijanju proteinov pri hidrolizi ATP in verjetno neaktivna vijolična kislinska fosfataza 27 (angl. probable inactive purple acid phosphatase 27), ki katalizira hidrolizo širokega spektra aktivnih mono in diestrov fosforjeve kisline in anhidridov. Na kromosому 4 se nahajata gen beta galaktozidaza 3, ki sodeluje pri metabolnem procesu ogljikovih

hidratov in ribosomski biogenezni protein TSR3 homolog (angl. ribosome biogenesis protein TSR3 homolog). Na kromosому 5 se nahajajo protein 1 toleranten na nekovine (angl. metal tolerance protein 1), ki ima transportno aktivnost kovinskih kationov, taumatinu podoben protein 1 (angl. thaumatin-like protein 1), ki ima obrambno funkcijo ter odziv na bakterije, 2-oksidoglutarat-Fe(II) tip oksidoreduktaze hxnY (angl. 2-oxoglutarate-Fe(II) type oxidoreductase hxnY), ki veže kovinske ione in ima oksidoreduktazno aktivnost. Na kromosому 6 najdemo gen beta-heksosaminidaza 2, ki sodeluje pri metabolnih procesih ogljikovih hidratov in gen amidofosforibozil transferaraza (angl. amidophosphoribosyltransferase, chloroplastic-like). Na kromosumu 7 najdemo NR1/HIN1 podoben protein (angl. NDR1/HIN1-like protein 12), ki ima obrambno funkcijo, protein ATHB-14 leucinske zadrge-homeobox (angl. homeobox-leucine zipper protein ATHB-14), ki se veže na DNA in ima aktivnost transkripcijskega faktorja in vezavo lipidov in PHD prstni protein LFIN-LIKE 2 (angl. PHD finger protein ALFIN-LIKE 2). Na kromosumu 8 se nahaja transkripcijski terminacijski faktor MTEF18 (angl. transcription termination factor MTEF18, mitochondrial), ki se veže na promotor DNA in regulira mitohondrijsko podvajanje in transkripcijo, protein 4, ki se veže na plazmodezmatski kalus (angl. Plasmodesmata callose-binding protein 4) ter prolikopenska izomeraza (angl. prolycopeno isomerase, chloroplastic), ki sodeluje pri biosintetskem procesu karotenoidov. Na kromosumu 9 imamo fosfoprotein 1 (angl. stress-induced-phosphoprotein 1), protein ABA deficient 4 (angl. protein ABA DEFICIENT 4, chloroplastic), ki sodeluje pri biosintezi neoksantina in abscizinske kisline, protein RDM15, transkripcijski faktor SRM1, ki prav tako koordinira biosintezo abscizinske kisline in evkariontski translacijski iniciacijski faktor 2, podenota gama (angl. eukaryotic translation initiation factor 2 subunit gamma), ki se veže na tRNA. Na kromosomu 10 pa se nahajata gen za vijolično kislinsko fosfatazo 2 (angl. purple acid phosphatase 2) in protein GIGANTEA, ki regulira fotoperiodo cvetenja.

4 ZAKLJUČEK

Zaradi populacijskih sort dvodomnosti ter tujeprašnosti vrste, prihaja pri konoplji pogosto do neizenačenosti in heterogenosti znotraj sorte. Zato je bil naš namen ugotoviti, ali se fenotipi odbrani na podlagi vizualnih lastnosti med sabo razlikujejo tudi na genetski ravni. Ugotovili smo, da se fenotipi na genetski ravni ne razlikujejo med sabo, medtem ko sorte se. Z analizo populacijske strukture sta se sorte Tiborszallasi in selekcija Finole lahko razdelili v dve skupini znotraj posamezne sorte, kar nakazuje na manjšo uniformnost kot pri sorti Carmagnola selected, ki je bila opredeljena kot ena skupina. Razlog za veliko raznolikost na fenotipskem nivoju bi lahko bil bolj okoljsko pogojen (vreme, lokacija, tla), kar pomeni, da bi lahko v več zaporednih letih poskusa ugotovili vpliv na število fenotipov znotraj sorte. Ker so bile sorte na genetski ravni bolj uniformne pomeni, da dejavniki, ki so vplivali na fenotipsko raznolikost niso genetsko pogojeni. Za boljše razumevanje konoplje na genetski ravni so potrebne nadaljnje analize delovanja različnih genov ter njihove povezave in vpliv na posamezne komponente, kar nam bi koristilo pri nadaljnji odbiri in žlahtnjenu novih sort z zaželenimi lastnostmi.

5 LITERATURA

- Aizpurua-Olaizola O., Soydaner U., Öztürk E., Schibano D., Simsir Y., Navarr, P., Etxebarria N., Usobiaga A. Evolution of the Cannabinoid and Terpene Content during the Growth of Cannabis sativa Plants from Different Chemotypes. *Journal of Natural Products.* 2016; 79(2): 324–331.
- Ambardar S., Gupta R., Trakroo D., Lal R., Vakhlu J. . High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology.* 2016; 56(4).
- Andre C. M., Hausman J. F., Guerriero G. *Cannabis sativa*: The plant of the thousand and one molecules. *Frontiers in Plant Science.* 2016; 7: 1–17.
- Barcaccia G., Palumbo F., Scariolo F., Vannozi A., Borin M., Bona S.. Potentials and Challenges of Genomics for Breeding Cannabis Cultivars. *Frontiers in Plant Science.* 2020; 11: 573299.
- Clarke R. C., Merlin M. D. *Cannabis: Evoluton and Ethnobotany* (1st ed.). University of California Press. 2013.
- Eržen M. Določanje kemijske sestave terpenov in kanabinoidov ter genetske raznolikosti odbranih fenotipov navadne konoplje (*Cannabis sativa* L.). Doktorska disertacija, Ljubljana. 118 str.
- Grassa C., Wenger J., Dabney C., Poplawski S., Motley S. T., Michael T., Schwartz C. J. Weible G. A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *BioRxiv.* 2018; 458083.
- Grassa C. J., Weible G. D., Wenger J. P., Dabney C., Poplawski S. G., Timothy Motley S., et al. A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist.* 2021 230:1665–79.
- Kojoma M., Iida O., Makino Y., Sekita S., Satake M.. DNA fingerprinting of *Cannabis sativa* using inter-simple sequence repeat (ISSR) amplification. *Planta Medica.* 2002;68:60–3. <https://doi.org/10.1055/s-2002-19875>.
- Kulski J. K. Next-Generation Sequencing – An Overview of the History, Tools, and “Omic” Applications. In *Next Generation Sequencing - Advances, Applications and Challenges.* 2016.
- Kump B., Javornik B. Evaluation of genetic variability among common buckwheat (*Fagopyrum esculentum* Moench) populations by RAPD markers. *Plant Science.* 1996; 114(2): 149–158.
- Laverty K. U., Stout J. M., Sullivan M. J., Shah H, Gill N., Deikus G., et al. A physical and genetic map of *Cannabis sativa* identifies extensive rearrangement at the 1 A physical and genetic map of *Cannabis sativa* identifies extensive rearrangement at the 2 T. *Genome Research.* 2019;29:146–56.
- Matoušek J., Kocábek T., Patzak J., Bříza J., Siglová K., Mishra A. K., Duraisamy G. S., Týcová A., Ono E., Krofta K. The “putative” role of transcription factors from HIWRKY family in the regulation of the final steps of prenylflavonid and bitter acids biosynthesis in hop (*Humulus lupulus* L.). *Plant Molecular Biology.* 2016; 92(3): 263–277.
- Namdar D., Mazuz M., Ion A., Koltai H. Variation in the compositions of cannabinoid and terpenoids in *Cannabis sativa* derived from inflorescence position along the stem and extraction methods. *Industrial Crops and Products.* 2018; 113(1): 376–382.
- Onofri C., De Meijer E. P. M., Mandolino G. Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry.* 2015; 116(1): 57–68.
- Sawler J., Stout J. M., Gardner K. M., Hudson D., Vidmar J., Butler L., Page J. E., Myles S. The genetic structure of marijuana and hemp. *PLoS ONE.* 2015; 10(8): 1–9.

- Soler S., Gramazio P., Figàs M. R., Vilanova S., Rosa E., Llosa E. R., et al. Genetic structure of *Cannabis sativa* var. indica cultivars based on genomic SSR (gSSR) markers: Implications for breeding and germplasm management. *Industrial Crops and Products*. 2017;104:171–8. <https://doi.org/10.1016/j.indcrop.2017.04.043>.
- van Bakel H., Stout J. M., Cote A. G., Tallon C. M., Sharpe A. G., Hughes T. R., Page J. E. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology*. 2011; 12(10).