

IZBIRA SNP POZICIJ GENOMA ZA DOLOČITEV GENETSKE RAZNOLIKOSTI PRI NAVADNI KONOPLJI (*Cannabis sativa* L.)

Marjeta ERŽEN¹, Andreja ČERENAK², Tjaša CESAR³, Lucija LUSKAR⁴ in Jernej JAKŠE⁵

Izvirni znanstveni članek / Original scientific article

Prispelo / Received: 7. 11. 2022

Sprejeto / Accepted: 24. 11. 2022

Izvleček

Konoplja (*Cannabis sativa* L.) je starodavna rastlina, ki vsebuje številne sekundarne metabolite, med drugimi tudi kanabinoide, ki so značilni predvsem za to vrsto. Rastline so heterozigotne, sorte populacijske, česar posledica je pojav številnih fenotipov znotraj sorte, kar se kaže na neizenačenosti rastlin znotraj populacije. Da bi preverili raznolikost fenotipov tudi na genetskem nivoju, smo na referenčni genom konoplje CBDRx kartirali transkriptomske NGS (angl. new generation sequencing, nove generacije sekvenciranja) podatke sort Finola, Purple Kush in Santhica 27. Poiskali smo SNP pozicije skupne vsem trem sortam in po dvema sortama skupaj. Določili smo heterozigotne SNP (angl. single-nucleotide polymorphism, polimorfizem posameznega nukleotida). Izdelali smo pregled vseh polimorfizmov ter porazdelitev polimorfizmov po kromosomih. Največ polimorfizmov je bilo pri sorti Purple Kush (782.406). Sorta Finola je imela največ polimorfizmov na kromosomu 4 (31.587), sorta Santhica 27 na kromosomu 1 (18.337) in sorta Purple Kush prav tako na kromosomu 1 (124.049). Med vsemi variantami polimorfizma je bilo največ SNV (angl. single-nucleotide variants, variant posameznega nukleotida). Glede na presek vseh treh sort se največ polimorfizmov pojavlja na prvem kromosomu. Za izbrane SNP pozicije smo naročili izdelavo komercialnega kompleta oligonukleotidnih sond za lovljenje. Od skupno 4.631 izbranih SNP pozicij je bilo izdelanih 4.086 sond za lovljenje. Z uporabo SNP polimorfizmov in tehnologijo visoko zmogljivega sekvenciranja je možnosti za analize konoplje na genetskem nivoju še veliko.

Ključne besede: *Cannabis sativa* L., genom, konoplja, SNP, sonde za lovljenje

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

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

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

⁴ Mag. biotehnol., IHPS, e-pošta: lucija.luskar@ihps.si

⁵ Prof. dr., BF UL, e-pošta: jernej.jakse@bf.uni-lj.si

DETERMINATION OF SNP GENOME POSITIONS FOR CAPTURE PROBES DESIGN FOR IDENTIFICATION OF GENETIC DIVERSITY IN HEMP (*Cannabis sativa* L.)

Abstract

Hemp (*Cannabis sativa* L.) is an ancient plant which contains many different secondary metabolites, among others also cannabinoids known especially for *Cannabis* plants. Plants are heterozygous and varieties are population which usually leads to different phenotypes within varieties. In order to determine differentiation of phenotypes also on genetic level, transcriptome NGS (angl. new generation sequencing) data of varieties 'Finola', 'Purple Kush' in 'Santhica 27' were mapped on reference hemp genome of 'CBDRx'. Common SNPs (angl. single-nucleotide polymorphism) positions of all three varieties and two and two varieties together were determined and heterozygous SNPs were also determined. Overview of polymorphism and distribution of SNPs on individual chromosomes were determined. The highest number of SNPs were at variety 'Purple Kush' (782.4060). 'Finola' has the highest number of polymorphisms on chromosome 4 (31.587), 'Santhica 27' on chromosome 1 (18.337), and 'Purple Kush' also on chromosome 1 (124.049). Of all variants of polymorphisms SNVs (angl. single-nucleotide variants) have the highest number. According to all three varieties the highest number of polymorphisms were on chromosome 1. Selected SNPs were used for commercial set of oligonucleotides design for capture probes. Of the 4.631 SNP positions determined, 4.086 capture probes were designed. With SNP polymorphisms and high throughput sequencing there is still a lot of possibilities for hemp analysis on genetic level.

Keywords: *Cannabis sativa* L., capture probes, genome, hemp, SNP

1 UVOD

Konoplja (*Cannabis sativa* L.) je starodavna kmetijska rastlina, ki spada v družino Cannabaceae in izvira iz centralne Azije (Hillig, 2005). Uporablja se za pridelavo vlaken, pridelavo semen, za prehrano ljudi in živali, v medicini za zdravljenje različnih bolezni. Je dvodomna rastlina, nekatere sorte pa so lahko tudi enodomne (Barcaccia in sod., 2020).

Konoplja vsebuje številne sekundarne metabolite (več kot 500 različnih) (Namdar in sod., 2018), kamor spadajo kanabinoidi, terpeni, flavonoidi, steroidi, alkaloidi in lignani (Janatová in sod., 2018). Med najpomembnejše spadajo kanabinoidi in terpeni, ki imajo številne terapevtske učinke na zdravje ljudi in živali. Do danes je bilo določenih več kot 100 različnih kanabinoidov. Med najpomembnejše spadajo kanabidiol (CBD), Δ -9-tetrahidrokanabinol (Δ -9-THC), kanabigerol (CBG), kanabikromen (CBC) in njihove karboksilne oblike (Namdar in sod., 2018). Največ

kanabinoidov se nahaja v trihomih na ženskih socvetjih, v koreninah in semenih kanabinoidov ni (Russo in Marcu, 2017).

Konoplja je diploidna rastlina ($2n = 20$), sestavljena iz devetih avtosomnih kromosomov in enega para spolnih kromosomov. Ženski haploidni genom je ocenjen na velikost 818 mega baznih parov (Mbp), moški haploidni genom pa 843 Mbp (Kovalchuk in sod., 2020). Z namenom ugotavljanja genetske variabilnosti in razlik na genetskem nivoju med različnimi tipi konoplje kot so *Cannabis sativa* spp. *sativa*, *Cannabis sativa* spp. *indica*, *Cannabis sativa* spp. *ruderalis* je bilo narejenih že kar nekaj študij (Gao in sod., 2014; Sawler in sod., 2015; Shirley in sod., 2013). Soler in sod., (2017) so z gSSR markerji ugotavljali razlike znotraj sorte in med sortami. Ugotovili so visoko informativnost markerjev pri preučevanju variabilnosti izražanja sekvenc THC-A in CBD-A sintaz. Za razlikovanje indijske in navadne konoplje na genomskem nivoju so Sawler in sod., (2015) z uporabo SNP markerjev dokazali značilno razliko in ugotovili, da je raven heterozigotnosti pri navadni konoplji višja kot pri indijski konoplji, kar kaže na to, da so sorte navadne konoplje pridobljene iz širše genetske osnove. S sekvenciranjem dolgih in kratkih odčitkov je bilo posekvenciranih 100 različnih genomov konoplje z namenom izdelave genetske karte in identifikacije lokusov kvantitativnih lastnosti (Grassa in sod., 2018). Za sorti Purple Kush in Finola je bila izdelana kombinirana genetska in fizična karta (Lavery in sod., 2019). Van Bakel in sod., (2011) so izdelali osnutek genoma in transkriptoma z namenom preučevanja ekspresije genov za kanabinoide in prekursorske poti pri sortah Finola in Purple Kush, medtem ko so Grassa in sod., (2021) sestavili genom na nivoju kromosomov.

Ker je konoplja tujeprašna vrsta in sorte populacijske, prihaja velikokrat do genetsko neizenačenih sort, znotraj sort pa se pojavijo številni fenotipi, kot posledica heterozigotnosti (Barcaccia in sod., 2020). V preteklosti je bilo veliko genetskih študij narejenih na podlagi izboljšanja pridelka, prehranske kvalitete in zmanjšanja dovzetnosti rastlin za bolezni in škodljivce (Aardema in DeSalle, 2021). Vedno bolj pa so pri konoplji zanimive genetske raziskave na področju sinteze kanabinoidov, predvsem CBD in THC, razlike med tipi konoplje (Gilmore in sod., 2003; Soler in sod., 2017) ter izbira markerjev za določanje spola in kemotipa pri konoplji (Toth in sod., 2020).

Cilj naše raziskave je bil poiskati večje število heterozigotnih SNP (angl. Single-nucleotide polymorphisms, polimorfizem posameznega nukleotida) pozicij, ki bi služile za nadaljnjo izdelavo sond za lovljene, na podlagi katerih bi lahko ugotovili populacijsko strukturo in genetsko diverziteto različnih fenotipov konoplje.

2 MATERIALI IN METODE

2.1 Izbira referenčnega genoma in transkriptomskih podatkov konoplje

V podatkovni zbirki NCBI (National Center for Biotechnology Information) smo na podlagi iskalnih parametrov izbrali referenčni genom za konopljo CBDRx (BioProject PRJEB29284), ki je sestavljen na nivoju kromosomov. Nato smo v podatkovni zbirki SRA (Sequence Read Archive) izbrali transkriptomске NGS podatke sort Finola (SRR351933, SRR7630403, SRR351932), Purple Kush (SRR352210, SRR352208, SRR352205) in Santhica 27 (SRR5210008, SRR5209988, SRR5209953), jih uvozili v program CLC Genomic Workbench Version 20.0.4. Iskali smo predvsem transkriptomске podatke navadne konoplje pri čemer so bili v trenutku raziskav na voljo samo podatki sort Finola in Santhica 27. Nato smo dodali še sorto Purple Kush, ki pa je sicer opredeljena kot indijska konoplja. Naše predvidevanje je bilo, če bodo polimorfizmi skupni trem izbranim sortam iz SRA podatkovne baze, bomo povečali verjetnost prisotnosti polimorfizma tudi v naših vzorcih. Vsako sorto posebej smo kartirali na izbrani genom konoplje z orodjem Map Reads to Reference. Poiskali smo polimorfna mesta in tarčne regije pri kartiranih odčitkih, kjer smo uporabili orodje Basic Variant Detection. Izbirali smo samo heterozigotne SNP pozicije. Z orodjem Venny 2.0 (<https://bioinfogp.cnb.csic.es/tools/venny/>) smo na podlagi Vennovega diagrama poiskali skupne SNP pozicije vsem trem izbranim sortam ter pozicije SNP za dve in dve sorti skupaj. Za preverjanje delovanja SNP na naših vzorcih smo določili skupno 60 SNP pozicij, 30 skupnih vsem tem sortam in 30 skupnih sortama Finola in Santhica 27 naključno po kromosomih. Določili smo sekvence SNP v skupni dolžini 501 bp (levo in desno od SNP pozicije 250 bp). Z orodjem BLAST smo nato poiskali homologna zaporedja in kratka ujemanja med dvema zaporedjema iz katerih to zaporedje naredi lokalne poravnave. Pozicije, ki so vsebovale odčitke, ki so imeli več mest prileganja v genomu, smo izločili iz analize. V programu Primer 3 smo izdelali končnih 14 parov začetnih oligonukleotidov (preglednica 1) s katerimi smo preverili ustreznost izbire SNP pozicij na naših vzorcih.

Preglednica 1: Začetni oligonukleotidi, njihove pozicije na kromosomih in sekvence parov začetnih oligonukleotidov.

Začetni oligonukleotid	Kromosom	Pozicija na kromosomu	Sekvence parov začetnih oligonukleotidov	
SNP 1	kromosom 1	443869	L: CCGTAGAAGGTGGCAAATGT R: TGCTTGTTTTCTTGGTTTTAGG	Pozicije skupne sortam Finola, Santhica 27, Purple Kush
SNP 2	kromosom 2	7186803	L: AGCCATTCCAAAAGCATTCC R: TTACAGCTTGTCAGCAAT	
SNP 3	kromosom 3	1187767	L: TAGGAATTGAACCGGATTGC R: TCAGCCTGCAATAATCGAAA	
SNP 4	kromosom 4	26919313	L: TTTTAGCGGGAACAACAACC R: AAGGAGGGAATTGGAAGAGC	
SNP 5	kromosom 4	66059933	L: AACTTGCAGCTCAAGGGAAA R: AAATCCACCATGGAAGGACA	

SNP 6	kromosom 6	78288161	L: GAAAACAGGTGTGGGAAGGA R: CCCGTTTGCAACATTTCTCT	Pozicije skupne sortama 'Finola' in 'Santhica 27'
SNP 7	kromosom 9	3445493	L: TTCGTTTTGATGTATGCACTCC R: TGCATGCTTAGACCCATCTG	
SNP 8	kromosom 10	55429963	L: TTCAGCACACCACGACATAA R: AGGGTTGGGTGAATGAATGA	
SNP 9	kromosom 1	5266236	L: TCTCCTTGATCAGCAACCAA R: TGCTCTCCTCCCTCAACAGT	
SNP 10	kromosom 1	1037237	L: AAGCTTCACCTTCTGCGAAA R: CAAATGCCGGAGTTTACTT	
SNP 11	kromosom 3	7382936	L: TTTCCCCGATCTTAGGGTTT R: TGGGAAAGTGAGGAGACTGG	
SNP 12	kromosom 4	69568690	L: TTCGCTGAAAACGACAAATG R: CCCGTCTAATCGGAAATTGA	
SNP 13	kromosom 6	70711381	L: ACTGCCTTCGTTTTCACCAG R: ATTCAGGGCCATGTCAAAAAG	
SNP 14	kromosom 8	42763723	L: GCACAAGAACTAATGGGCTGT R: ATATGGTGTTGGTGCGGTTT	

Pozicije SNP smo izbrali tako, da so bile enakomerno porazdeljene po kromosomih. Med njimi je bilo približno 100.000 bp razlike. Izbrane pozicije smo v obliki BED datoteke posredovali podjetju Arbour Bioscience za izdelavo sond za lovljenje.

2.2 Statističen pregled podatkov in odbira SNP pozicij

Glede na dobljene podatke izbranih SNP smo v programu Excel 2013 naredili statistiko števila vseh polimorfizmov in posameznih tipov polimorfizma ter porazdelitev SNP po posameznih kromosomih.

2.3 Testiranje delovanja SNP na vzorcih

Z namenom ugotoviti primernost postopka izbire SNP pozicij in strategijo izbire heterozigotnih SNP smo 14 naključno izbranih SNP pozicij (preglednica 1) uporabili za preverjanje njihove polimorfnost na osmih izbranih vzorcih konoplje in sicer na dveh različnih genotipih sorte Carmagnola CS, treh različnih genotipih sorte Tiborszallasi in treh različnih genotipih sorte selekcije Finola.

Izbrane pare začetnih oligonukleotidov smo pomnožili z uporabo verižne reakcije s polimerazo (PCR). PCR produkte smo nato očistili z ExoSAP-IT™ in inkubirali v cikličnem termostatu z naslednjim temperaturnim profilom; začetna denaturacija je potekala 3 minute pri 95°C, nato je sledilo 99 ciklov pri 96°C (10 s), 50°C (10 s) in 60°C (4 min). Namnoževanje je potekalo 7 minut pri 72°C. Za sekvenčno reakcijo smo uporabili kit BigDye Terminator 3.1 z naslednjo PCR mešanico: 2 µL 5x pufra, 0,5 µL pripravljene mešanice BigDye Terminator v3.1, 0,2 µL začetnih oligonukleotidov, 3,8 µL vode. PCR mešanici smo dodali 3,5 µL vzorca. Končno čiščenje sekvenčne reakcije smo opravili z etanolom in EDTA. Pripravljene vzorce smo poslali na sekvenčno analizo na Oddelek za zootehniko Biotehniške fakultete.

Končne rezultate smo obdelali v programu CodoneCode Aligner 9.0.1, kjer so bili elektroforegrami sekvenciranja tudi ročno pregledani.

3 REZULTATI Z RAZPRAVO

SRA je prosto dostopna baza podatkov surovih sekvenčnih odčitkov pridobljenih z NGS (angl. Next generation sequencing; tehnologija visoko zmogljivostnega sekvenciranja) tehnologijo (Kodama, Shumway in Leinonen, 2012). V podatkovni bazi smo našli podatke masovnega paralelnega sekvenciranja transkriptov za sorte Finola, Purple Kush in Santhica 27, ki smo jih kartirali na genom konoplje CBDRx (BioProject PRJEB29284). Na podlagi analize SNP regij je bila narejena statistika na kartiranih odčitkih. Iz preglednice 2 je razvidno, da je bilo največ polimorfizmov pri sorti Purple Kush (782.406), najmanj pa pri sorti Santhica 27 (122.033). Pri sorti Finola je bilo največ SNP na kromosomu 4 (31.587), pri sorti Santhica 27 na kromosomu 1 (18.337) in pri sorti Purple Kush prav tako na kromosomu 1 (124.049). Pri sortah Santhica 27 in Purple Kush je bilo največ heterozigotnih SNP, medtem ko je imela sorta Finola največ homozigotnih SNP. Med vsemi variantami SNP je bilo največ SNV (angl. single nucleotide variants; variant posameznega nukleotida), sledili so MNV (multi-nucleotide variants; več-nukleotidne variante), delecije, insercije, najmanj pa je bilo zamenjav. Daleč največ SNV je bilo pri sorti Purple Kush.

Preglednica 2: Porazdelitev SNP po posameznih kromosomih pri treh različnih sortah konoplje.

Sorta	Variante	Porazdelitev SNP po kromosomih										Homozigotnost	Heterozigotnost
		krom1	krom2	krom3	krom4	krom5	krom6	krom7	krom8	krom9	krom10		
Finola	295.651	45.969	29.713	26.448	31.587	24.201	28.505	23.099	30.789	23.845	31.495	149.967	145.684
Santhica 27	122.033	18.337	11.758	10.717	14.012	9.515	12.037	8.159	13.401	10.686	13.411	59.590	62.443
Purple Kush	782.406	124.049	83.625	77.684	78.340	66.134	74.202	55.965	81.805	70.486	70.116	320.849	461.557

Legenda: krom- kromosom.

V podatkovni zbirki NCBI lahko najdemo 13 različnih genomskih podatkov za konopljo, od tega je pet takšnih, ki so sestavljeni na nivoju kromosomov, ostali so sestavljeni le na nivoju sosesk (angl. Contig) ali ogrodij (angl. Scaffold). Referenčni genom, uporabljen v naši raziskavi, je do sedaj najboljši sestavljen genom (Grassa in sod., 2021).

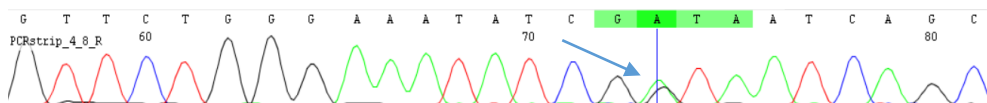
S funkcijo »SNP calling« (iskanje pozicij SNP) smo določili na kateri poziciji se polimorfizem nahaja ali na kateri poziciji se vsaj ena od baz razlikuje od referenčne sekvence (Nielsen in sod., 2011).

Preglednica 3: Porazdelitev heterozigotnih SNP po kromosomih, glede na presek dveh ali treh sort konoplje.

Št. Kromosoma in njegova dolžina/ skupina sort	Kromosom 1 (101.209 kb)	Kromosom 2 (96.347 kb)	Kromosom 3 (94.671 kb)	Kromosom 4 (91.941 kb)	Kromosom 5 (88.182 kb)	Kromosom 6 (79.355 kb)	Kromosom 7 (71.238 kb)	Kromosom 8 (64.622 kb)	Kromosom 9 (61.561 kb)	Kromosom X (104.666 kb)
F, S27, PK	176	75	83	84	72	113	67	116	85	89
F, S27	327	175	216	262	174	195	155	253	213	251
F, PK	635	394	333	357	307	345	292	416	293	286
S27, PK	361	184	221	218	171	253	144	275	217	171
Skupaj po kromosomih	1499	828	853	921	724	906	658	1060	808	797
Porazdelitev	67,5 kb/SNP	116,4 kb/SNP	111,9 kb/SNP	99,8 kb/SNP	121,8 kb/SNP	87,6 kb/SNP	108,4 kb/SNP	61,0 kb/SNP	76,2 kb/SNP	131,3 kb/SNP

F-'Finola', S27- 'Santhica 27', PK- 'Purple Kush'.

Pri vseh kombinacijah se največ heterozigotnih SNP nahaja na prvem kromosomu, kjer se SNP povprečno nahaja na vsake 67,5 kb (preglednica 3), najmanj heterozigotnih SNP pa se nahaja na kromosomu X, kjer se SNP pojavi povprečno šele na vsake 131,3 kb. Največ heterozigotnih SNP je bilo v kombinaciji med sortama Finola in Purple Kush na prvem kromosomu (635). Najmanj SNP pa v kombinaciji vseh treh sort na kromosomu 7 (67). Skupni polimorfizmi več sort hkrati nam dajo večjo možnost, da se bodo pojavili tudi v naših vzorcih.



Slika 1: Elektroforegram in prikaz potrditve heterozigotnosti SNP. Nukleotid na poziciji G (puščica) ima prisoten tudi vrh za A, kar pomeni heterozigotnost rastline na tem lokusu.

Z izdelanimi 14 pari začetnih oligonukleotidov smo potrdili pomnoževanje izbranih lokusov najprej s PCR reakcijo in gelsko elektroforezo na vseh osmih izbranih vzorcih. Pomnožke smo nato še sekvencirali po Sangerju in preverili heterozigotnost oz. homozigotnost analiziranih rastlin za izbrane lokuse. Od skupno 14 parov začetnih oligonukleotidov je bilo 8 heterozigotnih (57 %), kar je potrdilo primernost postopka izbiranja SNP pozicij. Na sliki 1 je prikazana sekvence dela pomnožka s heterozigotnim SNP. S tem se potrdi variabilnost znotraj določenega vzorca/sorte.

V izdelavo smo poslali 4.631 anotiranih SNP pozicij na podlagi katerih so izdelali 4.086 sond za lovljenje dolžine 80 bp, ki imajo 100 % kvaliteto pokritosti. Vsak SNP je pokrit s tremi sondami. Petit in sod. (2020) so za pojasnjevanje genetske arhitekture kvalitete konopljinih vlaken določili 612.452 SNP, kar je precej več kot v naši raziskavi, namreč uporabili so 6 sort na podlagi katerih so pridobili NGS podatke, medtem ko smo mi uporabili samo 3 sorte.

Uporaba genomike pri analizah konoplje se je začela približno pred 25 leti z uporabo RAPD, AFLP in RFLP markerjev, nato so jih nadomestili bolj informativni SSR markerji (Barcaccia in sod., 2020). Leta 2011 so van Bakel in sod. objavili prva osnutka genoma dveh sort in s tem odprli pot sodobnim tehnikam molekularnega žlahtnjenja tudi z uporabo SNP markerjev. Soorni in sod. (2017) so v analizi različnih iranskih vzorcev konoplje našli 24 tisoč visoko informativnih SNP markerjev. Markerji so zelo uporabni pri analizah genetske raznolikosti različnih sort konoplje, pri odkrivanju polimorfizmov v povezavi z geni (npr. kanabinoidna sintaza) in pri žlahtniteljskih programih.

4 ZAKLJUČEK

Z uporabo SNP markerjev se bo povečalo znanje o genetski raznolikosti konoplje in s tem boljše razumevanje za izboljšanje sort. S SNP lahko hitro in učinkovito analiziramo variabilnost genetskih regij, razmerja med genotipi rastlin in njihove željene lastnosti. Kar nekaj genetskih analiz je že bilo narejenih na konoplji v povezavi z genetsko variabilnostjo med rastlinami konoplje (med podvrstama *C. sativa* in *C. indica*), z geni za kanabinoide in njihovo izražanje. Še vedno pa je veliko neznanega na področju razlikovanja in povezave med fenotipi in genotipi, pri čemer bo uporaba SNP polimorfizmov in tehnologije visoko zmogljivega sekvenciranja lahko pomembno doprinesla k takšnim raziskavam.

5 LITERATURA

- Aardema M. L., DeSalle R. Can public online databases serve as a source of phenotypic information for Cannabis genetic association studies? *PLoS One*. 2021; 16.
- Barcaccia G., Palumbo F., Scariolo F., Vannozzi A., Borin M., Bona S. Potentials and Challenges of Genomics for Breeding Cannabis Cultivars. Vol. 11, *Frontiers in Plant Science*. 2020.
- Gao C., Xin P., Cheng C., Tang Q., Chen P., Wang C., Zang G., Zhao L. Diversity analysis in Cannabis sativa based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. *PLoS One*. 2014; 9(10): 1-7.
- Gilmore S., Peakall R., Robertson J. Short tandem repeat (STR) DNA markers are hypervariable and informative in Cannabis sativa: Implications for forensic investigations. *Forensic Science International*. 2003; 131(1): 65-74.
- Grassa C. J., Weiblen G. D., Wenger J. P., Dabney C., Poplawski S. G., Timothy Motley S., Michael T. P., Schwartz C. J. A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist*. 2021; 230(4): 1665-79.

- Grassa C., Wenger J., Dabney C., Poplawski S., Motley S. T., Michael T., Schwartz C. J. A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *New Phytologist*. 2018; 230: 1665-1679.
- Hillig K.W. Genetic evidence for speciation in Cannabis (Cannabaceae). *Genetic Resources and Crop Evolution*. 2005; 52(2): 161–80.
- Janatová A., Fraňková A., Tlustoš P., Hamouz K., Božik M., Klouček P. Yield and cannabinoids contents in different cannabis (*Cannabis sativa* L.) genotypes for medical use. *Industrial Crops and Products*. 2018; 112: 363–7.
- Kodama Y., Shumway M., Leinonen R.. The sequence read archive: Explosive growth of sequencing data. *Nucleic Acids Research*. 2012; 40: 2011–2013.
- Kovalchuk I., Pellino M., Rigault P., Van Velzen R., Ebersbach J., Ashnest J.R., MAu M., Schranz M. E., Alcorn J., Laprairie R. B., McKay J. K., Burbridge C., Schneider D., Vergara D., Kane N. C., Sharbel T. F. The Genomics of Cannabis and Its Close Relatives. *Annual Review of Plant Biology*. 2020; 71: 713–39.
- Laverty K. U., Stout J. M., Sullivan M. J., Shah H., Gill N., Deikus G., Sebra R., Hughes T. R., Page J. E., van Bakel H. A physical and genetic map of Cannabis sativa identifies extensive rearrangement at the THC/CBD acid synthase loci. *Genome Research*. 2019; 29: 146–56.
- 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 Products*. 2018; 113: 376–82.
- Nielsen R., Joshua S. P., Albrechtsen A., Song Y. S. Genotype and SNP calling from next-generation sequencing data. *Nature Review Genetic*. 2011; 12(6): 443–51.
- Petit J., Salentijn E. M. J., Paulo M. J., Denneboom C., van Loo E. N., Trindade L. M. Elucidating the Genetic Architecture of Fiber Quality in Hemp (*Cannabis sativa* L.) Using a Genome-Wide Association Study. *Frontiers in Genetic*. 2020; 11: 1–17.
- Russo E.B., Marcu J. Cannabis Pharmacology: The Usual Suspects and a Few Promising Leads. 1st ed. Vol. 80, *Advances in Pharmacology*. Elsevier Inc.; 2017. 67–134.
- 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.
- Shirley N., Allgeier L., Lanier T., Coyle H.M.. Analysis of the NMI01 Marker for a Population Database of Cannabis Seeds. *Journal of Forensic Science*. 2013; 58: 176–82.
- Soler S., Gramazio P., Figàs M. R., Vilanova S., Rosa E., Llosa E. R., Borrás D., Plazas M., Prohens J. 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.
- Soorni A., Fatahi R., Haak D. C., Salami S. A., Bombarely A. Assessment of Genetic Diversity and Population Structure in Iranian Cannabis Germplasm. *Scientific Report*. 2017; 7: 1–10.
- Toth J. A., Stack G. M., Cala A. R., Carlson C. H., Wilk R. L., Crawford J. L., Vainds D. R., Philippe G., Smart C. D., ROse J. K. C., Smart L. B. Development and validation of genetic markers for sex and cannabinoid chemotype in Cannabis sativa L. *Bioenergy*. 2020; 12(3): 213–22.
- 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): 1-17.