

SUPPLEMENTARY INFORMATION

Population Structure of Genotypes and Genome-Wide Association Studies of Cannabinoids and Terpenes Synthesis in Hemp (*Cannabis sativa* L.)

Plants

Marjeta Eržen^{1,2,*}, Andreja Čerenak¹, Tjaša Cesar², Jernej Jakše²

¹Department for Plants, Soil and the Environment, Slovenian Institute of Hop Research and Brewing, Cesta Žalskega tabora 2, 3310 Žalec, Slovenia

²Department for Agronomy, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia

* Correspondence:

Corresponding Author: Marjeta Eržen, Biotechnical faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia

Table S1: Barcode sequences with adapters used in NGS library preparation. Adapters are in *Italics*.

Barcode	Barcode sequence with adapters 5'-3'
A_Barcode-1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGAT
A_Barcode-2	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGAT
A_Barcode-3	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGAT
A_Barcode-4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGAT
A_Barcode-5	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGAT
A_Barcode-6	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTTCGAT
A_Barcode-7	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGAT
A_Barcode-8	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGAT
A_Barcode-9	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGAT
A_Barcode-10	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGAT
A_Barcode-11	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGAT
A_Barcode-12	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTTCGAT
A_Barcode-13	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACGAT
A_Barcode-14	CCATCTCATCCCTGCGTGTCTCCGACTCAG TTGGAGTGTCGAT
A_Barcode-15	CCATCTCATCCCTGCGTGTCTCCGACTCAG TCTAGAGGTCGAT
A_Barcode-16	CCATCTCATCCCTGCGTGTCTCCGACTCAG TCTGGATGACGAT
A_Barcode-17	CCATCTCATCCCTGCGTGTCTCCGACTCAG TCTATTTCGTCGAT
A_Barcode-18	CCATCTCATCCCTGCGTGTCTCCGACTCAG AGGCAATTGCGAT
A_Barcode-19	CCATCTCATCCCTGCGTGTCTCCGACTCAG TTAGTCGGACGAT
A_Barcode-20	CCATCTCATCCCTGCGTGTCTCCGACTCAG CAGATCCATCGAT
A_Barcode-21	CCATCTCATCCCTGCGTGTCTCCGACTCAG TCGCAATTACGAT
A_Barcode-22	CCATCTCATCCCTGCGTGTCTCCGACTCAG TTCGAGACGCGAT
A_Barcode-23	CCATCTCATCCCTGCGTGTCTCCGACTCAG TGCCACGAACGAT
A_Barcode-24	CCATCTCATCCCTGCGTGTCTCCGACTCAG AACCTCATTTCGAT
A_Barcode-25	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATACGAT

A_Barcode-26	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTACAACCTCGAT
A_Barcode-27	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCGAT
A_Barcode-28	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCGGAATCGAT
A_Barcode-29	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGACCACTCGAT
A_Barcode-30	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGGTTATCGAT
A_Barcode-31	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCAAGCTGCGAT
A_Barcode-32	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTTACACACGAT
A_Barcode-33	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTCATTGAACGAT
A_Barcode-34	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCATCGTTCGAT
A_Barcode-35	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGCCATTGTCGAT
A_Barcode-36	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGGAATCGTCGAT
A_Barcode-37	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGAGAATGTCGAT
A_Barcode-38	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGGACGGACGAT
A_Barcode-39	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAACAATCGGCGAT
A_Barcode-40	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACATAATCGAT
A_Barcode-41	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCACTTCGCGAT
A_Barcode-42	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACGAATCGAT
A_Barcode-43	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGACACCGCGAT
A_Barcode-44	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGGCCAGCGAT
A_Barcode-45	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGCTTCCTCGAT
A_Barcode-46	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGTCCGAACGAT
A_Barcode-47	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGCAACCACGAT
A_Barcode-48	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTAAGAGACGAT
A_Barcode-49	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAACATAACGAT
A_Barcode-50	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAATGGCGAT
A_Barcode-51	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGAGCCTATTTCGAT
A_Barcode-52	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCATGGAACGAT
A_Barcode-53	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGGCAATCCTCGAT
A_Barcode-54	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGGAGAATCGCGAT
A_Barcode-55	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCACCTCCTCGAT
A_Barcode-56	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCATTAAATTCGAT
A_Barcode-57	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGCAACGGCGAT
A_Barcode-58	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAGAACACGAT
A_Barcode-59	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTTGATGTTTCGAT
A_Barcode-60	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGCTCTTCGAT
A_Barcode-61	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTCGGATCGAT
A_Barcode-62	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTGCTTACGAT
A_Barcode-63	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTTAGAGTTCGAT
A_Barcode-64	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGAGTTCCGACGAT
A_Barcode-65	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGGCACATCGAT
A_Barcode-66	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCAATCATCGAT
A_Barcode-67	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTACCAGTCGAT
A_Barcode-68	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAAGAAGTTCGAT
A_Barcode-69	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCAATTGGCGAT
A_Barcode-70	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTACTGGTCGAT
A_Barcode-71	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGGCTCCGACGAT

A_Barcode-72	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CGAAGGCCACACGAT</i>
A_Barcode-73	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TCTGCCTGTCGAT</i>
A_Barcode-74	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CGATCGGTTTCGAT</i>
A_Barcode-75	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TCAGGAATACGAT</i>
A_Barcode-76	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CGGAAGAACCCTCGAT</i>
A_Barcode-77	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CGAAGCGATTTCGAT</i>
A_Barcode-78	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CAGCCAATTCTCGAT</i>
A_Barcode-79	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CCTGGTTGTCGAT</i>
A_Barcode-80	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TCGAAGGCAGGCGAT</i>
A_Barcode-81	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CCTGCCATTTCGCGAT</i>
A_Barcode-82	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TTGGCATCTCGAT</i>
A_Barcode-83	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CTAGGACATTTCGAT</i>
A_Barcode-84	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CTTCCATAACGAT</i>
A_Barcode-85	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CCAGCCTCAACGAT</i>
A_Barcode-86	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CTTGGTTATTCGAT</i>
A_Barcode-87	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TTGGCTGGACGAT</i>
A_Barcode-88	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CCGAACACTTCGAT</i>
A_Barcode-89	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TCCTGAATCTCGAT</i>
A_Barcode-90	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CTAACCACGGCGAT</i>
A_Barcode-91	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CGGAAGGATGCGAT</i>
A_Barcode-92	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CTAGGAACCGCGAT</i>
A_Barcode-93	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CTTGTCCAATCGAT</i>
A_Barcode-94	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TCCGACAAGCGAT</i>
A_Barcode-95	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CGGACAGATCGAT</i>
A_Barcode-96	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TTAAGCGGTCGAT</i>

Table S2: Primers with sequences used in NGS library amplification.

Primer	Sequence 5'-3'
P1amp	CCACTACGCCTCCGCTTTCCTCTCTATG
T_PCR_A	CCATCTCATCCCTGCGTGTCT

Collection of Images S1: Photos of individual phenotypes within varitey 'Carmagnola selected', 'Tiborszallasi' and 'Finola selection' based on visual traits.

Phenotype CI ('Carmagnola selected')



Phenotype CII ('Carmagnola selected')



Phenotype TI ('Tiborszallasi')



Phenotype TII ('Tiborszallasi')



Phenotype TIII ('Tiborszallasi')



Phenotype TIV ('Tiborszallasi')



Phenotype TV ('Tiborszallasi')



Phenotype FI ('Finola selection')



Phenotype FII ('Finola selection')



Phenotype FIII ('Finola selection')



Phenotype FIV ('Finola selection')



File S1: SNP calling

SNP calling was performed using the Genome Analysis Toolkit (GATK) (<https://gatk.broadinstitute.org/hc/en-us>), following the GATK Best Practices Workflows (<https://gatk.broadinstitute.org/hc/en-us/sections/360007226651-Best-Practices-Workflows>).

The procedure included the following steps:

1 Creating a sequence dictionary file from a reference sequence using the 'CreateSequenceDictionary'.

Reference genome in FASTA format was in file `GCA_900626175_seqret.fasta`

command: `gatk CreateSequenceDictionary -R GCA_900626175_seqret.fasta`

2 Identification of duplicates using 'MarkDuplicates'.

The file `MarkDuplicates.bash` was written (below), which was run with a bash command `MarkDuplicates.bash`

```
for f in *.bam; do
```

```
base=${f%.bam}
```

```
gatk MarkDuplicates -I »$f« -O »$base«_marked_duplicates.bam -M
»$base«_marked_duplicates_metrics.txt 1>«$base«_markduplicates.1log
2>«$base«_markduplicates.log
```

done

3 Sorting of BAM files using 'SortSam'.

File SortSam.bash was written (below) which was run with a bash command SortSam.bash

```
for f in *_marked_duplicates.bam; do
base=${f%.bam}
gatk SortSam -I $f -O »$base«_sorted.bam -SORT_ORDER coordinate
1>«$base«_sortsam.1log 2>«$base«_sortsam.2log
```

done

4 Recalibration of qualitative values using the 'BaseRecalibrator' according to known SNP positions and tool 'ApplyBQSR.bash'.

For known 4537 SNP positions file in VCF format 4537_pozicij.vcf.was used.

File BaseRecalibrator.bash was written (below), which was run with a bash command BaseRecalibrator.bash

```
for f in *_sorted.bam; do
base=${f%.bam}
gatk BaseRecalibrator -I »$f« -R GCA_900626175_seqret.fasta -known-
sites 4537_pozicij.vcf -O »$base«_recal_data.table
```

done

File ApplyBQSR.bash was written (below), which was run with a bash command ApplyBQSR.bash

```
for f1 in *_marked_duplicates_sorted.bam; do
base1=${f1%.bam}
base2=«${base1}_marked_duplicates_sorted»«
for f2 in *marked_duplicates_sorted_recal_data.table; do
base3=${f2%.table}
if [ »$base1« == »${base3}_recal_data»« ]; then
gatk ApplyBQSR -R GCA_900626175_seqret.fasta -I »$f1« -bqsr-recal-
file »$f2« -O »$base2«_processed.bam 1>«$base2«_ApplyBQSR1.log
2>«$base2«_ApplyBQSR2.log
```

fi

done

done

5 Determination of SNP positions using 'HaplotypeCaller'.

File HaplotypeCaller.bash was written (below), which was run with a bash command HaplotypeCaller.bash

```

for f in *_processed.bam; do
base1=${f%.bam}
base2=${base1%_processed}

gatk HaplotypeCaller -R GCA_900626175_seqret.fasta -I »$f« -O
»$base2«.vcf -ERC GVCF -native-pair-hmm-threads 150
1>«$base2«_HaplotypeCaller.log1 2>«$base2«_HaplotypeCaller.log2

done

```

6 Combining individual VCF files to one VCF file using the 'CombineGVCFs'.

File CombineGVCFs.bash was written (below), which was run with a bash command
CombineGVCFs.bash

gatk CombineGVCFs \

```

-R ../GCA_900626175_seqret.fasta -variant ../CI1.vcf -variant
../CI2.vcf -variant ../CI3.vcf -variant ../CI4.vcf -variant
../CI5.vcf -variant ../CI6.vcf \ --variant ../CI7.vcf \ --variant
../CI8.vcf \ --variant ../CI9.vcf \ --variant ../CI10.vcf \ --
variant ../CI11.vcf \ --variant ../CI12.vcf \ --variant ../CI13.vcf
\ --variant ../CI14.vcf \ --variant ../CI15.vcf \ --variant
../CI16.vcf \ --variant ../CI18.vcf \ --variant ../CI19.vcf \ --
variant ../CII1.vcf \ --variant ../CII2.vcf \ --variant ../CII3.vcf
\ --variant ../CII4.vcf \ --variant ../CII5.vcf \ --variant
../CII6.vcf \ --variant ../CII7.vcf \ --variant ../CII8.vcf \ --
variant ../CII9.vcf \ --variant ../CII10.vcf \ --variant
../CII11.vcf \ --variant ../CII12.vcf \ --variant ../CII13.vcf \ --
variant ../CII14.vcf \ --variant ../CII15.vcf \ --variant
../CII16.vcf \ --variant ../CII17.vcf \ --variant ../CII18.vcf \ --
variant ../CII19.vcf \ --variant ../FI1.vcf \ --variant ../FI2.vcf \
--variant ../FI3.vcf \ --variant ../FI4.vcf \ --variant ../FI5.vcf \
--variant ../FI6.vcf \ --variant ../FI7.vcf \ --variant ../FI9.vcf \
--variant ../FI10.vcf \ --variant ../FI11.vcf \ --variant
../FI12.vcf \ --variant ../FI13.vcf \ --variant ../FI14.vcf \ --
variant ../FI15.vcf \ --variant ../FI16.vcf \ --variant ../FI17.vcf
\ --variant ../FI18.vcf \ --variant ../FI20.vcf \ --variant
../FII1.vcf \ --variant ../FII2.vcf \ --variant ../FII3.vcf \ --
variant ../FII4.vcf \ --variant ../FII5.vcf \ --variant ../FII6.vcf
\ --variant ../FII7.vcf \ --variant ../FII8.vcf \ --variant
../FII9.vcf \ --variant ../FII10.vcf \ --variant ../FII11.vcf \ --
variant ../FII12.vcf \ --variant ../FII14.vcf \ --variant
../FII15.vcf \ --variant ../FII16.vcf \ --variant ../FII17.vcf \ --
variant ../FII18.vcf \ --variant ../FII19.vcf \ --variant
../FII20.vcf \ --variant ../FIII1.vcf \ --variant ../FIII2.vcf \ --
variant ../FIII3.vcf \ --variant ../FIII4.vcf \ --variant
../FIII5.vcf \ --variant ../FIII6.vcf \ --variant ../FIII7.vcf \ --
variant ../FIII8.vcf \ --variant ../FIII9.vcf \ --variant
../FIII10.vcf \ --variant ../FIII11.vcf \ --variant ../FIII12.vcf \
--variant ../FIII13.vcf \ --variant ../FIII14.vcf \ --variant
../FIII15.vcf \ --variant ../FIII16.vcf \ --variant ../FIII17.vcf \

```

```
--variant ../FIII18.vcf \ --variant ../FIII19.vcf \ --variant
../FIII20.vcf \ --variant ../FIV1.vcf \ --variant ../FIV2.vcf \ --
variant ../FIV3.vcf \ --variant ../FIV4.vcf \ --variant ../FIV5.vcf
\ --variant ../FIV6.vcf \ --variant ../FIV7.vcf \ --variant
../TI1.vcf \ --variant ../TI2.vcf \ --variant ../TI3.vcf \ --variant
../TI4.vcf \ --variant ../TI5.vcf \ --variant ../TI6.vcf \ --variant
../TI7.vcf \ --variant ../TI8.vcf \ --variant ../TI9.vcf \ --variant
../TI10.vcf \ --variant ../TI11.vcf \ --variant ../TI12.vcf \ --
variant ../TI13.vcf \ --variant ../TI14.vcf \ --variant ../TI15.vcf
\ --variant ../TI17.vcf \ --variant ../TI18.vcf \ --variant
../TI19.vcf \ --variant ../TI20.vcf \ --variant ../TII1.vcf \ --
variant ../TII2.vcf \ --variant ../TII3.vcf \ --variant ../TII4.vcf
\ --variant ../TII5.vcf \ --variant ../TII6.vcf \ --variant
../TII7.vcf \ --variant ../TII8.vcf \ --variant ../TII9.vcf \ --
variant ../TII10.vcf \ --variant ../TII11.vcf \ --variant
../TII12.vcf \ --variant ../TII13.vcf \ --variant ../TII14.vcf \ --
variant ../TII15.vcf \ --variant ../TII16.vcf \ --variant
../TII17.vcf \ --variant ../TII18.vcf \ --variant ../TII19.vcf \ --
variant ../TII20.vcf \ --variant ../TIII1.vcf \ --variant
../TIII2.vcf \ --variant ../TIII3.vcf \ --variant ../TIII4.vcf \ --
variant ../TIII5.vcf \ --variant ../TIII6.vcf \ --variant
../TIII7.vcf \ --variant ../TIII8.vcf \ --variant ../TIII9.vcf \ --
variant ../TIII11.vcf \ --variant ../TIII12.vcf \ --variant
../TIII13.vcf \ --variant ../TIII14.vcf \ --variant ../TIV1.vcf \ --
variant ../TIV2.vcf \ --variant ../TIV3.vcf \ --variant ../TIV4.vcf
\ --variant ../TIV5.vcf \ --variant ../TIV6.vcf \ --variant
../TIV7.vcf \ --variant ../TIV8.vcf \ --variant ../TIV9.vcf \ --
variant ../TIV10.vcf \ --variant ../TIV11.vcf \ --variant
../TIV12.vcf \ --variant ../TIV13.vcf \ --variant ../TV1.vcf \ --
variant ../TV3.vcf \ --variant ../TV4.vcf \ --variant ../TV5.vcf \ -
-variant ../TV6.vcf -O combined_all.vcf 1>1.log 2>2.log &
```

7 Samples genotyping using 'GenotypeGVCFs'.

genotyping of samples occurred according to 4,537 analyzed SNP positions (file 4537_pozicij.vcf). Data for monomorphic positions were kept using the `-include-non-variant-sites`

Run bash command:

```
gatk GenotypeGVCFs -R GCA_900626175_seqret.fasta -V combined_all.vcf
-O combined_all_genotyped.vcf -L 4537_pozicij.vcf -include-non-
variant-sites
```

8 Data were filtered according to the quality and depth of reads using 'VariantFiltration'

```
Run bash command: gatk VariantFiltration -R GCA_900626175_seqret.fasta -
V
combined_all_genotyped.vcf --filter-expression "DP < 5" --
filter-name "LowQual" --filter-expression "QUAL < 20" --
filter-name "LowQual20" -O test_filtered.vcf
```

Supplementary Data S1: [10.6084/m9.figshare.30762896](https://doi.org/10.6084/m9.figshare.30762896)

Supplementary Data S2: <https://doi.org/10.6084/m9.figshare.23761965>

Supplementary Data S3: <https://doi.org/10.6084/m9.figshare.23763873>

Supplementary Data S4: <https://doi.org/10.6084/m9.figshare.23764314>

Supplementary Data S5: <https://doi.org/10.6084/m9.figshare.23764482>

Supplementary Data S6: <https://doi.org/10.6084/m9.figshare.23765337>