Genet Resour Crop Evol
https://doi.org/10.1007/s10722-024-02015-1

RESEARCH ARTICLE

Exploring cannabidiol variations, investigation of genetic diversity, population structure and unveiling male-specific genetic marker in industrial hemp (*Cannabis sativa* L.)

Ahmad Alsaleh · Güngör Yılmaz

Received: 4 February 2024 / Accepted: 4 May 2024
© The Author(s) 2024

Abstract *Cannabis sativa* played a pivotal role across different industries. Recently, industrial hemp, particularly in the case of Cannabidiol (CBD), gained attention for its therapeutic potential. This study evaluated cannabinoid variability and genetic diversity within 43 industrial hemp individuals, primarily Turkish, across various plant parts and growth stages, and inflorescences of females showed significant CBD content. The highest contents were observed in Turkish landraces (0.55–8.05% with an average of 3.26%), making them valuable genetic resources for high CBD. Genetic structure revealed distinct populations based on gender and influenced by geographical origin. Analysis of Molecular Variance showed 92% of genetic variation observed within populations and indicated a promising source of novel allelic diversity in the Turkish gene pool. Turkish females showed significant genetic diversity [No. of Different Alleles (Na) 1.507, No. of Effective Alleles (Ne) 1.226, Shannon's Information Index (I) 0.258, and Percentage of Polymorphic Loc (%P) 74.91], exhibiting richer genetic variation than their international peers. Principal Coordinate Analysis unveiled gender-specific genetic differences, and admixture clusters shed light on genetic interactions and historical connections among diverse populations. Unweighted Pair-Group Method with Arithmetic Averaging highlighted unique genetic profiles and distinct genetic lineages. Genome-wide association study revealed a highly significant male-specific genetic marker explained 50% of the phenotypic variation. These findings inform future breeding strategies and conservation efforts and contribute to varietal identification methods, Marker-Assisted Selection, and efficient cultivar development in upcoming programs.

Keywords *Cannabis sativa* · Industrial hemp · Cannabidiol (CBD) · Genetic diversity · MTA · Male-genetic marker

Introduction

Cannabis sativa, an ancient cultivated plant, has played a pivotal role in human history, contributing to various industries such as food, textiles, construction, pharmaceuticals, and recreation (Burgel et al. 2020; Hammami et al. 2021). In recent years, the intricate tapestry of this botanical marvel of Industrial hemp has led to significant attention to its potential benefits,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10722-024-02015-1.

A. Alsaleh (✉)
Agriculture and Food, Hemp Research Institute, Yozgat Bozok University, 66200 Yozgat, Turkey
e-mail: aal_saleh@yahoo.com

G. Yılmaz
Field Crops, Faculty of Agriculture, Yozgat Bozok University, 66200 Yozgat, Turkey
e-mail: gungor.yilmaz@bozok.edu.tr

Published online: 15 May 2024

especially the therapeutic, as Cannabidiol (Chen and Pan 2021), a non-psychoactive compound, has emerged as a subject of extensive exploration due to its remarkable therapeutic potential, including anti-inflammatory, analgesic, and anxiolytic effects (Yang et al. 2020; Burgel et al. 2020). The pharmacological significance of CBD has spurred investigation aimed at maximizing its yield, specifically in high-CBD hemp cultivars cultivated under diverse environmental conditions (Yang et al. 2020; Abdollahi et al. 2021). To optimize cannabinoid production, breeders must understand genetic variability, environmental cues, and gender-specific traits (Moliterni et al. 2004; van Bakel et al. 2011). Notably, Toth et al. (2020) emphasized the pivotal role of female plants in CBD production, as Crispim Massuela et al. (2022) underscored. The significance of aligning growth phases with CBD content in industrial hemp is underscored by Burgel et al. (2020), highlighting the ongoing need for advancement in cultivation and harvesting and specifically addressing the unexplored dynamics of cannabinoid accumulation in industrial hemp genotypes during different growth stages. This involves discerning the impact of genotype and harvest time on biomass yield, separated into leaves, inflorescence, and cannabinoid yield. Assessing CBD content throughout various growth stages of industrial hemp is imperative for understanding variability and its diverse potential applications across industries. This knowledge is instrumental in developing new hemp strains with heightened CBD content and improved therapeutic benefits (Crispim Massuela et al. 2022; Mostafaei Dehnavi et al. 2022). Abdollahi et al. (2021) further emphasized the intricate relationship between cannabinoid content and genetic traits, growth stages, and environmental conditions. Yang et al. (2020) also recognized a gap in current replicated research on this topic and documented a noteworthy escalation in Tetrahydrocannabinol (THC), CBD, and Cannabigerol (CBG) concentrations during flower development of hemp. Hemp, characterized by its diploid nature and unique sex chromosomes, possesses a complex genetic framework being crucial for its agricultural and industrial traits (Moliterni et al. 2004; van Bakel et al. 2011).

Uncovering the genetic diversity and structure of industrial hemp is paramount for hemp breeding programs, aiming to develop cultivars with enhanced agronomic and industrial traits, underscoring the importance of comprehending the genetic construction of this invaluable plant (Sawler et al. 2015). Baloch et al. (2017) asserted that the analysis of plant genetic diversity is pivotal in plant genetics, breeding, conservation, and evolution. Therefore, understanding the genetic and cannabinoid variation of hemp is not only vital for enhancing its agricultural and industrial characteristics but also holds significance in exploring its therapeutic potential, particularly concerning CBD. Analyzing genetic diversity and structure within industrial hemp populations facilitates identifying and conserving unique genetic resources being crucial for future cultivar development (Shams et al. 2020). In recent years, deoxyribonucleic acid (DNA) markers have emerged as powerful tools for plant genetic analysis (Nadeem et al. 2018). Unlike traditional breeding methods, molecular markers enable swift and precise identification of plants with desired features, such as high CBD content and early detection of plant gender. Significant progress in plant breeding has been witnessed following the introduction of DNA molecular markers (Alsaleh et al. 2022), enabling genetic characterization at early developmental stages. Inter Simple Sequence Repeat (ISSR) markers have demonstrated their value in genetic diversity analysis across various investigations, including hemp research by Kayıs et al. (2010), Lata et al. (2010), Zhang et al. (2014), Khatak et al. (2016), and Shams et al. (2020). These markers have also played a pivotal role in investigating genetic diversity in different crops.

Notably, earlier research in other crops underscores the continued significance of ISSR markers in genetic diversity analyses. The appeal of ISSR markers lies in their high polymorphism, reproducibility, and cost-effectiveness, making them indispensable tools in unraveling genetic variations in diverse plant species. The early separation of male and female plants in cannabis cultivation holds agricultural significance, given the gender-specific influence on key traits (Chandra et al. 2020). Numerous researchers have undertaken various experiments to identify molecular markers for early-stage sex determination. Noteworthy investigations conducted by Törjek et al. (2002) and Sakamoto et al. (2005) have meticulously outlined the transformation of male-specific markers into Sequence Characterized Amplified Region (SCAR) markers. Furthermore, Rode et al. (2005) exemplified this phenomenon by

reporting the identification of alleles for SSR markers and revealed distinctions between different X chromosomes. In a related study, Borin et al. (2021) stated understanding the gender-specific traits in hemp is paramount, especially for agriculturally significant attributes.

The quest for a reliable molecular marker for early-stage sexual differentiation in hemp is crucial, necessitating further validation through additional features linked to gender traits. Sesiz et al. (2024) highlighted the utility of a genome-wide association study (GWAS) for discovering genetic markers associated with significant characteristics, providing a more precise and cost-effective alternative to traditional breeding methods. Utilizing molecular markers in GWAS can accurately pinpoint the location of genes controlling vital traits such as gender, offering valuable information for developing more effective breeding techniques, including Marker-Assisted Selection (MAS). Integrating reliable markers for sexual differentiation in the early stages of the crop facilitates efficient sex screening during seedling, enabling the removal of undesired plants and conserving vital resources like labor, water, fertilizers, and space. Beyond streamlining cultivation practices, this approach ensures the intentional cultivation of highly desirable female or male plants (Rode et al. 2005; Borin et al. 2021).

Despite the global economic significance of industrial hemp, a noticeable gap exists in investigations into the CBD content and systematic research on the genetic structure diversity of industrial hemp cultivars or landraces cultivated particularly in Türkiye. Therefore, this research aims to measure CBD content by exploring the development of cannabinoids in the inflorescences and leaves of industrial hemp under open-field conditions during two growth stages: flowering and harvest time. Additionally, the study seeks to assess the variability and genetic structure of industrial hemp populations. Furthermore, the research includes the detection of gender-specific markers in hemp plants. This information can potentially help the development of new hemp cultivars with enhanced agronomic and industrial properties. Moreover, the study may contribute to preserving genetic resources in industrial hemp, fostering the breeding of novel cultivars with improved traits specific to the Turkish context. The implications of this research extend beyond national borders, offering valuable insights into optimizing the cultivation and harvesting of industrial hemp for CBD production on a global scale.

Materials and methods

Plant material

In this research, a total of 43 industrial hemp (*Cannabis sativa* L.) genotypes were examined. These genotypes were obtained from the Hemp Research Institute, Yozgat Bozok University, and comprised 19 female (♀) plants, 19 male (♂) plants, and five monoecious (⚥) plants. These plants were collected from twelve local Turkish landraces, one Turkish-released cultivar, six foreign cultivars from different international locations, and five unknown landraces of untraceable origin. The cultivation began in the spring of 2021 at the dedicated research and implementation area of the Hemp Research Institute in Yozgat, Türkiye. The site coordinates are 39°49′7.19″ N latitude and 34°48′9.59″ E longitude. The experimental period 2021 had precipitation of 509.6 mm, the mean temperature was 10.1 °C, and no crop was planted the previous year. The cultivation included 19 dioecious and 5 monoecious genotypes, and their identities can be found in Supplementary Table S1. The Federer's seminal work in 1956 inspired the experimental design, which was meticulously executed. The plants were planted with a precise 20 cm spacing between rows, and the area was kept weed-free and disease-free through the strategic application of herbicides and fungicides. Agronomic standards and meticulous plant protection practices were upheld throughout the cultivation period. The harvesting was done in August and September, timed uniquely for each plant to align with their maturation profiles. We categorized the collected samples as follows: (FF): Inflorescence harvested during the flowering stage; (FH): Inflorescence harvested during the harvesting stage; (LF): Leaves harvested during the flowering stage; and (LH): Leaves harvested during the harvesting stage.

Isolation of genomic DNA and ISSR analysis

Fresh leaves were explicitly acquired to isolate total genomic DNA from each, encompassing females, males, and monoecious plants. According to Doyle

and Doyle, the Cetyl trimethyl ammonium bromide (CTAB) protocol was employed. (1987) at the Laboratory of Science and Technology Application and Research Center (Bilim ve Teknoloji Uygulama ve Araştırma Merkez; BİLTEM), Yozgat Bozok University, Yozgat-Türkiye. The consequent DNA was evaluated for quantity and quality using 0.8% agarose gel electrophoresis. Following isolation, the DNA was diluted to a concentration of 25 ng μl^{-1} and used for ISSR analysis. For performing ISSR analysis, at the beginning, 50 ISSR primers were chosen and used for primary screening on four individuals to determine levels of polymorphism and assess primer functionality. The 20 primers that yielded the highest polymorphism were then used to screen all individuals (Table S2). The PCR amplification protocol was based on Alsaleh et al. (2020) and involved amplification of ISSRs from approximately 50 ng of genomic DNA, 0.4 mM Deoxyribonucleotide triphosphate (dNTP), 0.8 pmol ISSR primers, 0.2 U Taq DNA polymerase, and 1X PCR buffer (10 mM Tris HCl pH 8.3, 50 mM KCl, three mM MgCl2) in a 10 µl PCR reaction. The Mastercycler was programmed to perform an initial denaturation step at 94 °C for 2 min, followed by 40 cycles of 1 min at 94 °C at the specific annealing temperature for each primer (as listed in Table S2). The final extension step was carried out at 72 °C for 2 min, followed by a further extension at 72 °C for 10 min, and a final cooling step to 4 °C. The ISSR-PCR products were then run on 2.5% Agarose gels submerged in 1X of Tris, Borate, and Ethylenediaminetetraacetic acid (TBE) buffer using an electrophoresis chamber alongside a standard size marker (100 bp DNA Ladder) to estimate the molecular size. The gels were stained with ethidium bromide solution (at a final concentration of 0.5 µg/ml) for 20 min, and the PCR fragments were visualized on an ultraviolet (UV) transilluminator.

High-performance liquid chromatography (HPLC) quantification of CBD contents

CBD assessment was conducted on different parts (leaves and whole inflorescence) of the tested hemp individuals', as outlined in Table S3. Samples were collected at two distinct growth stages: at flowering and at harvest. The analysis involved manually collecting leaves and whole inflorescence from each individual at each sampling stage. To ensure optimal preservation until the commencement of the required investigations, the samples were securely stored in paper bags within a controlled environment characterized by darkness, low temperature, and low humidity. Then, the collected plant samples were kept in a dark oven at 40 °C until they became dry. They were then ground, weighed, and extracted in methanol at a ratio of 1:100. The extracted samples were re-diluted at 1:100 in methanol, passed through filters, and taken into HPLC vials for analysis. CBD detection was carried out using an HPLC instrument (Shimadzu-LC-20AT, Japan) with a photo-diode array (PDA) detector (Shimadzu-SPD-M20, Japan) at a wavelength of 210 nm. The HPLC system was equipped with a specialized column (Cannabis Analyzer for Potency, 2.7 µm, 150 mm × 4.6 mm, Shimadzu, USA) for cannabinoid detection. The column temperature was maintained at 35 °C, and the mobile phase system used was A (0.1% formic acid in acetonitrile): B (0.1% formic acid in water) in an isocratic mode with a solvent flow rate of 0.3 ml/min. The sample injection amount was 10 µL. The HPLC procedures were conducted at BİLTEM laboratories at Yozgat Bozok University, Türkiye. The cannabinoid content of the samples was determined from the absorbance values of each sample using the calibration chart created by analyzing the calibration solutions in the HPLC device in the first stage and the extracted cannabis plant samples in the second step.

Statistical analysis of phenotypic and molecular data

Forty-three individuals underwent systematic categorization into five discrete groups (A, B, C, D, and E), delineated by their gender, origin, and the intricate insights garnered from population structure analysis. Group A was made up of 13 females from Türkiye, and Group B was composed of 13 males from Türkiye. Group C included six female individuals, out of which five had an unknown lineage. These females were accompanied by the foreign cultivar "Finola." To avoid confusion, we have referred to these six females as Unknown females. Similarly, Group D comprised males sourced from five individuals of untraceable origin, and one male individual of the foreign "Finola" cultivar also accompanied these males. To avoid confusion, we will refer these six males as unknown males. Lastly, Group E comprises a unique set of five monoecious cultivars of wholly foreign

origin. We conducted initial analyses, beginning with Descriptive Statistics to explore group differences. Subsequently, Analysis of Variance (ANOVA) was applied to the data from CBD Content Analysis using Microsoft Excel Significant differences were highlighted using Tukey's HSD (honestly significant difference) test with the rstatistics package (Kassambara 2023). Visual representations were generated through Violin Plots to illustrate distribution and variability across groups, utilizing Jeffrey's Amazing Statistics Program (JASP) software version 0.11.1 (Love et al. 2019). Additionally, Factorial Analysis was carried out, and variance interpretation was achieved through Principal Component Analysis (PCA) using XLSTAT 2016.02.28451. The variables used in PCA and their corresponding eigenvalues were documented.

Lastly, Hierarchical Clustering Classification was performed using JASP software version 0.11.1, presenting a hierarchical tree that delineates clusters and relationships among data points on the factorial map. Analysis of Molecular Variance (AMOVA) was done using Genetic Analysis in Excel (GenAlEx) version 6.501 software, developed by Peakall and Smouse (2001). 574 ISSR loci were meticulously recorded from the data generated by 20 carefully selected primers. Each locus was scrutinized twice, adhering to the method delineated by Buntjer (1999), and this verification process was executed within the Cross Checker 2.91 software. The ISSR loci data were binary, with '1' signifying the presence of a marker and '0' denoting its absence within the genomic representation of each respective sample. These binary scores were systematically organized into a binary matrix. The determination of population structure was accomplished using the model-based clustering method, as implemented in STRUCTURE version 2.3.4 software, as described by Pritchard et al. (2000). This analysis was performed under the parameters number of clusters or populations assumed K = 10, with 5000 burning iterations and 50,000 repetitions over ten iterations. Subsequently, the results were subjected to further analysis using Structure Harvester version 0.6.94 software, developed by Earl and Von Holdt (2012), to estimate the optimal number of clusters (K) within the population, following the methodology proposed by Evanno et al. (2005). Principal Coordinate Analysis (PCoA) was conducted employing GenAlEx version 6.501 software. Genetic distance calculations were performed using Diversity Analysis and Representation for Windows (DARWin) version 6.0.21 software by the methodology described by Perrier et al. (2003). The proportion of phenotypic variation explained by gender for each marker was estimated using the coefficient of determination (R^2) value in Trait Analysis by aSSociation, Evolution and Linkage (TASSEL 5) version 5.2.86 software (Bradbury et al. 2007). The Bonferroni threshold for multiple testing and an adjusted corrective entry was applied to determine significant associations (Kaler and Purcell 2019). To further enhance our understanding of genetic relationships and individual similarities, we utilized Tassel. This tool enabled the creation of a tree diagram and facilitated the detection of significant similarity among the studied individuals using the Unweighted Pair-Group Method with Arithmetic Averaging (UPGMA).

Results

Cannabidiol variations assessment

HPLC analysis revealed varying levels of cannabinoid constituents. In Table S3, an overview of CBD levels is presented for different combinations of hemp plants. Female individuals such as Van (♀-FF: 8.501%), Maltepe1 (♀-FF: 5.309%), and Mısır Çarşısı (♀-FF: 4.965%) exhibit notably high CBD content in female inflorescence during the flowering stage. These individuals present valuable genetic reservoirs for high-CBD hemp landraces. Individuals such as Kavacik (♀-FF: 1.490%) from Turkish landraces and Finola (♀-FF: 0.554%) and USO31 (♀-FF: 0.822%) from foreign individuals, respectively, displayed relatively lower CBD content. Male inflorescence, like ♂-FF varied from 0.01 to 1.693%, exhibits stable CBD content across individuals. Monoecious individuals showed varied CBD content, and individuals such as Futura75 (⚥-FF: 1.898%) and Santhica27 (⚥-FF: 1.707%) present promising options for cannabinoid extraction. On the other hand, female inflorescence (♀-FF) displayed notable variations in CBD concentrations, demonstrating higher CBD contents compared to other parts or growth stages, ranging from 0.554% to 8.501%.

Developmental stages also played a crucial role; female inflorescence at the flowering stage (♀-FF) exhibited a mean CBD content of 3.258%, which

decreased to 1.746% at the harvesting stage (♀-FH) (Table 1). Similarly, within monoecious plants, significant intra-plant variability was observed. Inflorescence from the flowering stage (⚥-FF) had an average CBD content of 1.409%, while inflorescence from the harvesting stage (⚥-FH) averaged 0.873%. Despite comparatively lower CBD content, monoecious individuals remained valuable for CBD production, emphasizing the pivotal importance of precise harvest timing to optimize CBD yields. Additionally, ANOVA analysis underscored significant variations in CBD levels observed across various plant parts and developmental stages, as elucidated in Table S4a. The calculated F-statistic of 18.11, surpassing the critical F-value (1.96), indicated highly significant differences in CBD content among these groups. Female inflorescence during the flowering stage (♀-FF) exhibited the highest average CBD content at 3.258%, emphasizing the pinnacle of cannabinoid synthesis during this phase. Similarly, Monoecious inflorescence at the flowering stage (⚥-FF) displayed an average CBD content of 1.409%, comparable to specific sets of female inflorescence, signifying their potential as valuable sources of cannabinoids. In contrast, leaves from various stages across females, males, and monoecious plants displayed significantly lower CBD contents, ranging from 0.16 to 0.57% (Table S4b). Tukey's Honestly Significant Difference (HSD) test identified a group with noteworthy differences in CBD content across plant types and growth stages, as detailed in Table S5. Gender-Based Differences in CBD Content: Significant variations in CBD content among females were observed across different plant parts. The Kruskal–Wallis test results, detailed in Table S6, reinforce our earlier findings, confirming substantial disparities in CBD content among traits ($H(9) = 80.609$, $p < 0.001$). This non-parametric analysis aligns with the outcomes of the ANOVA, underscoring the robustness of the identified differences.

Furthermore, the chi-squared test conducted, characterized by a test statistic of 162.995 and 26 degrees of freedom, revealed highly significant differences (< 0.001), as detailed in Table S6. Finally, Violin plots were utilized to visually represent CBD content distributions across various plant types and growth stages. This graphical approach showcased the data spread, deepening our understanding of the dataset's intricacies (refer to Fig. 1).

Principal Component Analysis (PCA) confirmed the outcomes observed in other statistical analyses, revealing distinct eigenvalues. The primary components (F1 and F2) were pivotal in elucidating the most significant variance in CBD contents (Fig. S1). Specifically, F1 contributed 54.8% to the total variance, while F2 accounted for an additional 31.5%, collectively explaining 86.389% of the variability (Fig. 2). The Factor analysis elucidated patterns within the dataset. RC1 exhibited substantial loadings for all traits, indicating a shared underlying factor influencing these traits. Specific characteristics such as female-FF, female-FH, female-LH, and female-LF demonstrated high loadings on RC1, suggesting a common genetic or environmental basis. RC2 indicated distinct associations, primarily involving monoecious groups, pointing towards unique factors governing monoecious hemp genotypes (Fig. S2).

Conversely, monoecious samples exclusively from foreign individuals exhibited distinct CBD patterns (Fig. 2). Comparative analysis of sexual types unveiled notable differences in CBD content. Dioecious plants such as Medical and Van displayed

Table 1 Descriptive Statistics of Cannabidiol Contents of Leaves and Inflorescences at Flowering and Harvesting Stages in an Industrial Hemp Panel

	♀-LF	♀-LH	♀-FF	♀-FH	♂-LF	♂-FF	⚥-LF	⚥-LH	⚥- FF	⚥-FH
Mean	0.572	0.432	3.258	1.746	0.248	0.757	0.318	0.163	1.409	0.873
Std. Deviation	0.318	0.273	1.956	1.119	0.206	0.476	0.096	0.087	0.254	0.278
Coefficient of variation	0.556	0.632	0.6	0.641	0.832	0.629	0.301	0.533	0.18	0.319
Minimum	0.102	0.092	0.554	0.409	0.009	0.01	0.084	0.0005	0.579	0.008
Maximum	1.417	0.885	8.501	3.684	0.556	1.693	0.573	0.453	1.898	1.567

♀: Female, ♂: male, ⚥: monoecious, FF: inflorescence from the flowering stage, FH: inflorescence from harvesting stage, LF: leaves from the flowering stage, and LH: leaves from harvesting stage; CV%: Coefficient of Variation

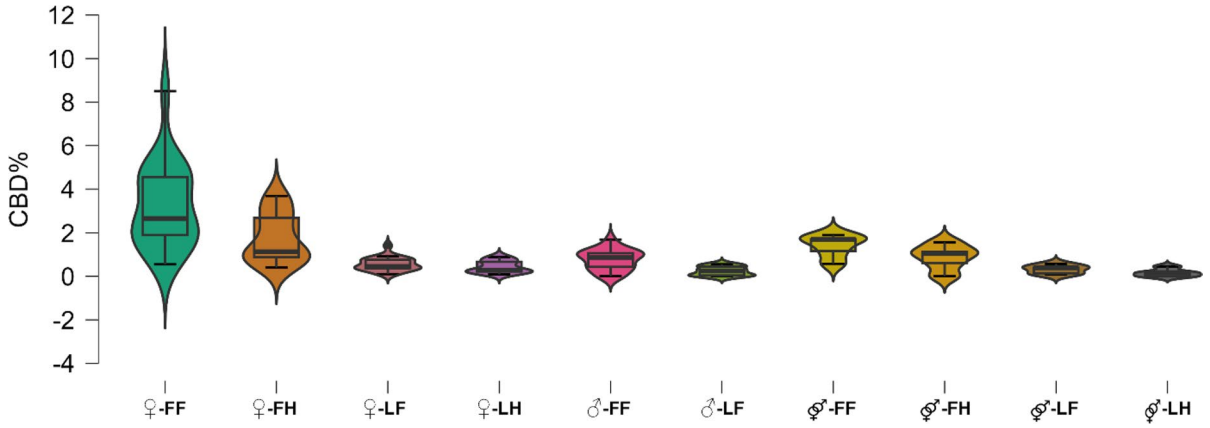


Fig. 1 The distribution of Cannabidiol (CBD) content across different plant parts, types, and growth stages of female (♀), male (♂), and monoecious (⚥) plants, with samples categorized as follows: FF (inflorescence from the flowering stage), FH (inflorescence from harvesting stage), LF (leaves from the flowering stage), and LH (leaves from harvesting stage)

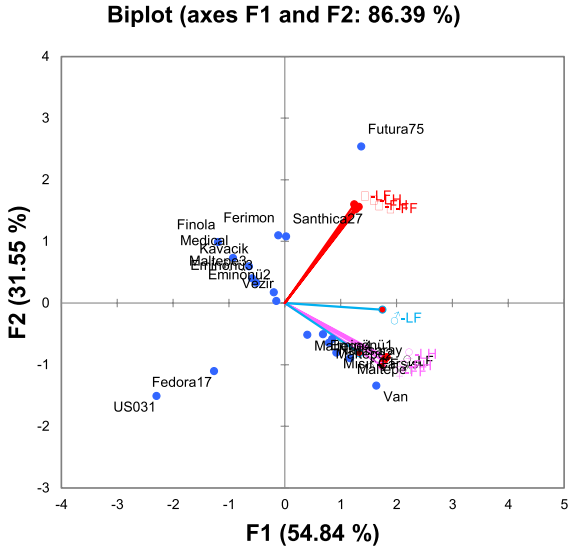


Fig. 2 Bioplot diagram of cannabinoid profiles across hemp genotypes: Influence on CBD synthesis in individuals and distinct pattern

diverse CBD profiles, reflecting genetic heterogeneity within this category. Conversely, monoecious individuals from different international origins exhibited specific cannabinoid patterns.

Defining genetic population structure

STRUCTURE software was employed, coupled with the Structure Harvester program, to investigate intraspecific differentiation and population structure to gain detailed insights. The model-based clustering approach identified the optimal number of populations (K) as 3, a determination supported by a significant Delta K value of 7 (Fig. S3). The results exhibited a high correspondence based on sex nature, emphasizing the reliability of the clustering. Considering a Delta K value of 7.08386 at K = 3, the individuals were distinctly clustered mainly based on sex, forming three clusters (Fig. 3a). The first cluster comprised Unknown females, monoecious plants, and two Turkish females (Kavcik and Van). The second cluster exclusively included males, whether Unknown, foreign or domestic. The third cluster was composed solely of Turkish females, highlighting clear gender-based genetic distinctions. The mean F-statistics (Fst) values (0.3168, 0.1479, and 0.1115) suggest varying degrees of genetic differentiation among the three clusters. Further exploration using K5 and K8 revealed additional layers of genetic structuring (Fig. 3a). At K5, individuals clustered not only based on their sex nature but also relatively on their geographical origins, forming six groups. The first group consisted of nine Turkish females. The second group included Unknown females, monoecious samples, and two Turkish females (Kavcik and Van), indicating potential genetic similarity despite geographical disparities. The third group comprised nine Turkish males, underscoring their unique genetic structure. The fourth group consisted of four males, three Turkish, and the male unk-5. The fifth group comprised

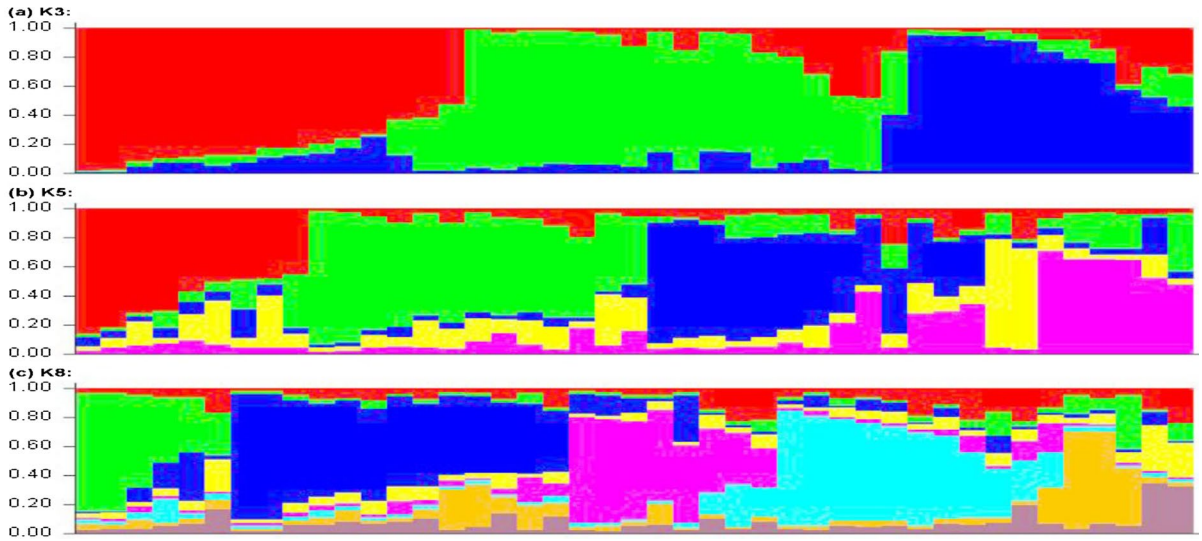


Fig. 3 The survey of genetic relationship among 43 cannabis individuals using STRUCTURE at K = 3, 5, and 8 (a**, b, and c**), respectively. Distinct groups are represented by different colors, and each bar symbolizes a sample. The predominance of a color over others in a bar signifies that the sample belongs to the group represented by that color

two Turkish females (Maltepe2 and Narlisaray), forming a distinctive cluster. The last group included six males, with the majority being Unknown, alongside the presence of the Turkish landraces Narlisaray (Fig. 3b). In the K8 analysis, the individuals were partitioned into seven clusters, describing relative similarities based on gender and geographical origin. The first cluster included six Turkish females, and the second cluster comprised Unknown females, monoecious plants, and Turkish females (Kavcik and Van). The third cluster included four Unknown males, while the fourth cluster consisted of four males, three Turkish, alongside the Unknown-1. The fifth cluster included exclusively male individuals, predominantly Turkish, with the presence of the Unknown-5. The sixth cluster comprised three Turkish females, indicating their unique genetic composition. The final group included Turkish females of Narlisaray and Maltepe2, underscoring their distinct genetic profile within the broader population (Fig. 3c).

Exploring genetic diversity in hemp populations and principal coordinate analysis (PCoA)

Genetic distance analysis revealed intraspecific variation in *cannabis sativa* individuals. Within Turkish females, distinct genetic distances were apparent, averaging 0.223. For example, Maltepe4 and Eminönü3 displayed a genetic distance of 0.151, indicating a notable genetic lowest differentiation between Turkish female individuals. In contrast, Maltepe2 and Kavacik exhibited a higher genetic distance of 0.313, suggesting a substantial genetic distinction between these Turkish female individuals. Among Turkish male individuals, a range of genetic distances was observed, with an average higher than Turkish females 0.226. Female individuals also exhibited varying genetic distances, reflecting their genetic diversity. For instance, a relatively low genetic distance of 0.133 was observed between Finola and Unk-1, indicating a close genetic relationship among these females. Conversely, a higher genetic distance of 0.215 was noted between Unk-2 and Unk-3, suggesting a more distant genetic relationship between these individuals. Similar patterns of genetic differentiation were observed among Unknown male individuals. The genetic distance matrix highlighted significant diversity, with values ranging from 0.127 to 0.215.

In contrast, Unkown females also displayed a more downward average genetic distance (0.171) than their male counterparts (0.183). Unknown males also exhibited a close genetic relationship. The monoecious individuals displayed intriguing patterns of genetic differentiation. The higher genetic distance

was observed between Ferimon and US03 (0.249), while the lowest one was observed between Ferimon and Santhica27 (0.142).

Interestingly, monoecious individuals exhibited an average genetic distance of 0.194. This intermediate value suggests a moderate genetic diversity within this subset.

The pairwise population matrix of Nei genetic distances provides, as evidenced in Table S7, an additional, comprehensive overview of the genetic differentiations among distinct hemp populations. Analyzing the genetic distances between Turkish females and Turkish males reveals a relatively low genetic distance of 0.031, indicating a moderate level of genetic divergence within the Turkish Cannabis population. Comparing Turkish females with Unknown females or males reveals higher genetic distances (0.044 and 0.037, respectively), signifying more significant genetic divergence between these groups. The lowest genetic distance between Turkish and Unknown males (0.027) suggests a closer genetic relationship. At the same time, the highest distance was noted between Unknown females and Turkish males (0.057). Notably, the genetic distances between the monoecious population and other groups are moderate, ranging from 0.039 to 0.055.

Furthermore, the genetic distances among monoecious individuals and other populations are notably intermediate, emphasizing a relatively homogenous genetic composition. In parallel, the Nei genetic identity analyses, as illustrated in Table S8, accentuated significant genetic similarities among specific population pairs. Turkish and Unknown males exhibit the highest important genetic identity of 0.973. Turkish females and Turkish males also show a remarkably high genetic identity of 0.969, indicating a shared genetic background between these gender groups. Comparing Turkish females with Unknown females reveals a genetic identity value of 0.957. These values suggest substantial genetic overlap between Turkish and Unknown populations, emphasizing the potential historical gene flow and shared genetic heritage. The genetic identity among monoecious populations and their divergence from other groups (ranging from 0.947 to 0.962) is noteworthy. The relatively genetic identity values within the monoecious populations underscore the genetic coherence among monoecious individuals, indicating a shared genetic background despite geographical distinctions.

PCoA represented in Fig. 4 analysis delineated four distinct clusters within the population, with notable nuances in their composition. The Turkish female and male populations, as well as Unknown female and male populations, exhibited clear and separate genetic profiles. A particularly intriguing revelation was identifying an admixture cluster encompassing Unknown females, monoecious populations, and one Turkish female individual.

The PCoA focusing on different gender populations (females, males, and monoecious) revealed substantial distinctions. Specifically, males exhibited notable differences from the female or monoecious population and maintained a significant separation from each other (Fig. S4). In contrast, the monoecious population was intricately entwined and closely associated with the female population, indicating a higher degree of genetic similarity between monoecious individuals and females. PCoA conducted based on the genders and origins of the studied groups (Fig. 5) revealed precise divergent hereditary positions. Notably, the data displayed logical arrangements for the individuals or populations, with the

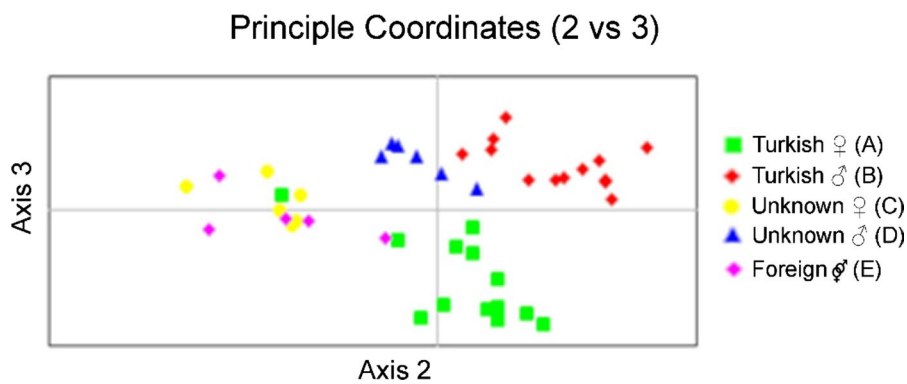


Fig. 4 Relationship between individuals visualized by Principal Coordinate Analysis (PCoA) using Inter Simple Sequence Repeat (ISSR) data

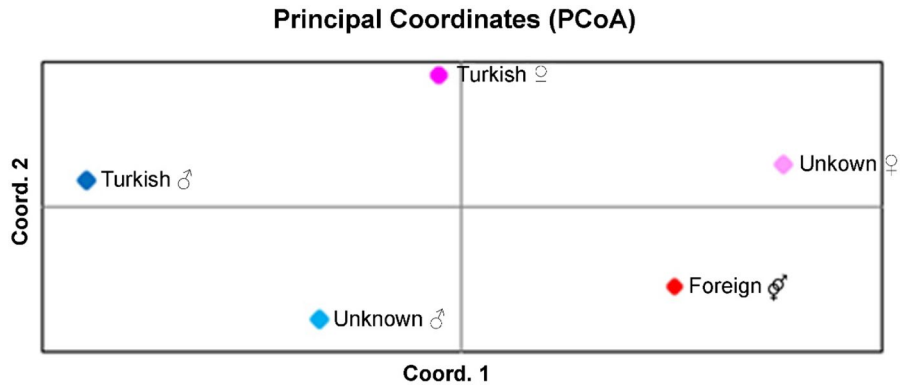


Fig. 5 Principal Coordinates Analysis based on the genders and origin of studied groups

closest genetic distance observed between monoecious and Unknown female groups. Conversely, the most significant genetic disparities were evident in the placements between Unknown females or monoecious and Turkish males.

Analysis of molecular variance (AMOVA)

Employing AMOVA revealed that 8% of the total genetic variance (SS = 585.191, df = 4) was attributed to differences among these populations. The predominant portion of genetic variation (92%) was observed within the populations (SS = 3320.251, df = 38) (Fig. 4 and Table S9).

Genetic Diversity Within-Population: Turkish female and male populations exhibited higher genetic diversity (Na = 1.507 ± 0.036, Ne = 1.226 ± 0.011) than their Unknown counterparts, indicating a richer allelic variation among these Turkish groups (Table 2). Shannon's Information Index (I = 0.258 ± 0.009) also reflected this heightened diversity, emphasizing the complexity of genetic information within these populations. Expected Heterozygosity (He) and Unbiased Expected Heterozygosity (uHe) were also comparatively higher in Turkish populations.

Genetic Differences Between-Populations: Comparing populations, Turkish females and males displayed significantly higher genetic diversity (percentage of polymorphic loci = 74.91% and 70.73%, respectively) compared to Unknown females, Unknown males, and monoecious individuals.

Overall Genetic Diversity: The grand mean across loci and populations revealed an average genetic diversity (Na = 1.097 ± 0.018, Ne = 1.199 ± 0.005) and polymorphism (percentage of polymorphic loci = 53.52% ± 8.00%). Table S10 presents a detailed summary of band patterns among five distinct Cannabis populations, shedding light on the unique genetic signatures within each group. Turkish Female and

Table 2 Band Frequencies, Estimated Allele Frequencies, and Estimated Heterozygosity by Population for Binary (Diploid) Data Mean and standard error (SE) over Loci of the Populations

Pop	Group	Na	Ne	I	He	uHe	%P
Turkish ♀	A	1.507 ± 0.036	1.226 ± 0.011	0.258 ± 0.009	0.154 ± 0.006	0.161 ± 0.006	74.91%
Turkish ♂	B	1.432 ± 0.037	1.224 ± 0.011	0.253 ± 0.009	0.152 ± 0.006	0.159 ± 0.006	70.73%
Unknown ♀	C	0.810 ± 0.040	1.175 ± 0.012	0.175 ± 0.010	0.111 ± 0.007	0.122 ± 0.007	38.50%
Unknown ♂	D	0.927 ± 0.041	1.187 ± 0.011	0.196 ± 0.010	0.122 ± 0.007	0.136 ± 0.007	45.12%
Foreign ⚥	E	0.807 ± 0.040	1.185 ± 0.012	0.183 ± 0.010	0.117 ± 0.007	0.132 ± 0.008	38.33%
Grand Mean and SE over Loci and Pops							
Total		1.097 ± 0.018	1.199 ± 0.005	0.213 ± 0.004	0.131 ± 0.003	0.142 ± 0.003	53.52% ± 8.00%

Na = No. of Different Alleles; *I* = Shannon's Information Index = − 1* (p * Ln (p) + q * Ln(q)); *Ne* = No. of Effective Alleles = 1/(p^2 + q^2); *I* = Shannon's Information Index = − 1* (p * Ln (p) + q * Ln(q)); *He* = Expected Heterozygosity = 2 * p * q; *uHe* = Unbiased Expected Heterozygosity = (2N/(2N-1)) * He; and *P* = Percentage of Polymorphic Loci

Male populations exhibit 435 and 416 different bands, respectively. Notably, 55 private bands in Turkish Females and 43 in Turkish Males indicate distinctive genetic markers exclusive to each gender, underlining gender-specific genetic variability. Comparatively, Unknown female, Unknown male, and monoecious populations display 244, 273, and 243 bands, respectively. Among these, eight private bands in Unknown females, 6 in Unknown males, and 18 in monoecious individuals signify rare, population-specific genetic elements within the Unknown groups. Intriguingly, the absence of locally common bands (<=25% and <=50%) across all populations.

Genetic relationships among individuals revealed by UPGMA

As depicted in Fig. 6, the dendrogram revealed a complex phylogenetic structure characterized by the emergence of four primary branches; two were

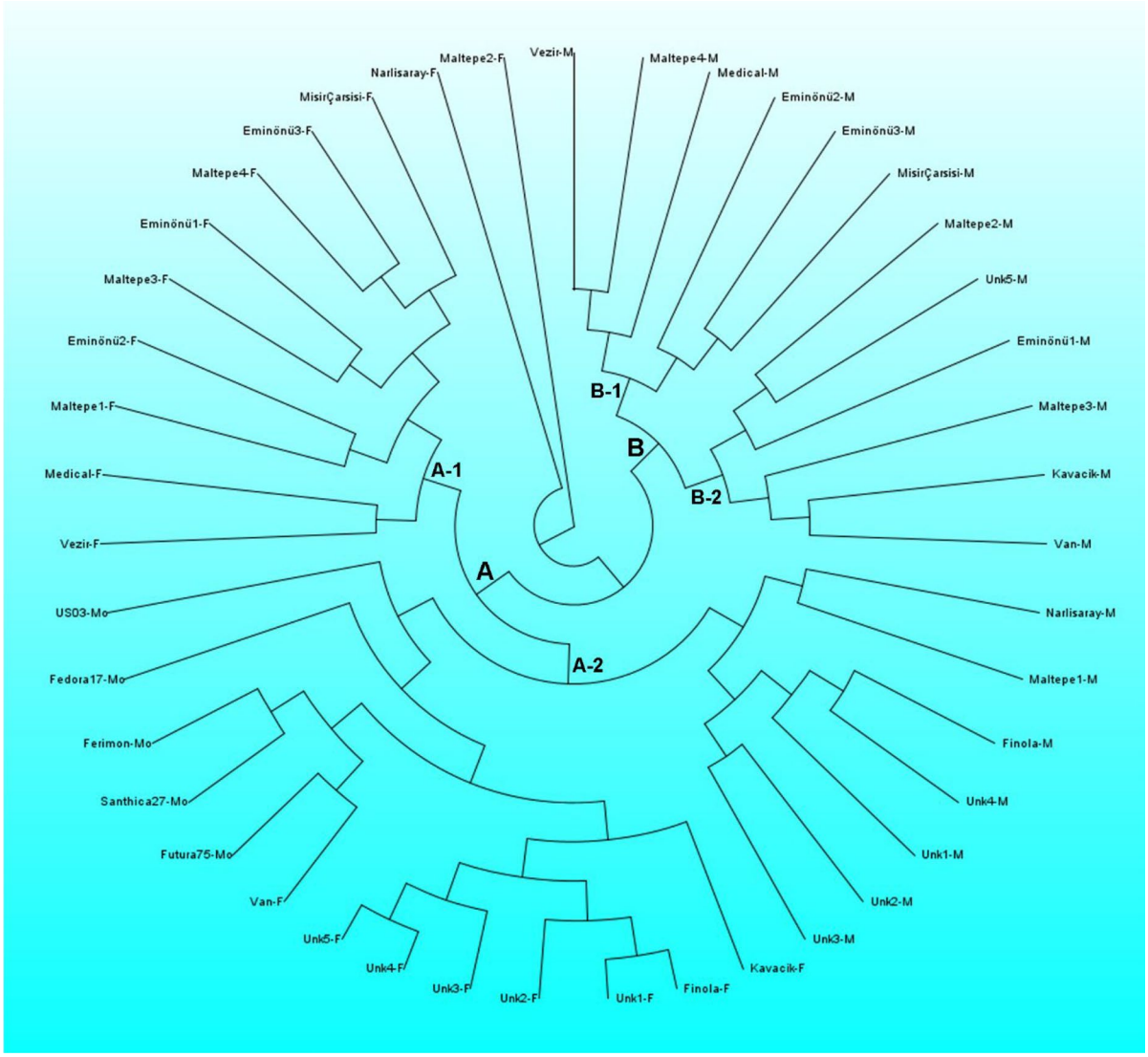


Fig. 6 Unweighted Pair-Group Method with Arithmetic Averaging (UPGMA) tree showing relationships between 43 industrial hemp individuals revealed by Inter Simple Sequence Repeat (ISSR) markers. Individual names are followed by F, M, and Mo letters, which correspond to females, males, and mon8oecious

separate, one for females of Narlisaray and one for Maltepe2 unique genetic profiles. The central cluster (A1), encompassing a diverse array of female and male plants, mainly Turkish, exhibited further complexity upon closer examination. It bifurcated into two sub-branches, one subbranch displayed striking homogeneity among nine genetically Turkish females (A-1), and the other was a more extensive cluster (A-2), encompassing a diverse array of female, male, and monoecious plants, exhibited further complexity upon closer examination, and revealing intriguing patterns of genetic relatedness. For instance, Unknown females formed a distinct sub-group, closely located to the Turkish Kavacik individual, and samples from the monoecious category clustered in the other sub-branch, including the female landraces of Van. The second sub-branch of the cluster (A2) primarily consisted of male individuals, incorporating Unknown group males and two Turkish males, Narlisaray and Maltepe1. A fascinating pattern emerged among males within the second main branch of dendrograms (B). Turkish males constituted a predominant cluster within this branch. Upon closer examination, this main branch revealed further complexity. It is divided into two distinct sub-branches, each with its unique composition. One sub-branch exclusively comprised males from Turkish individuals (B-1). In contrast, the other sub-branch (B-2) included a mixture of Turkish males and one male of unknown identity, specifically identified as "Unk-5."

Identification of a sex-associated 785-bp fragment

The Mixed Linear Model (MLM), incorporating Population structure (Q) and Kinship matrix (K), revealed one marker-trait association (MTA) at a significance level of $p < 0.05$, particularly associated with the male gender (Table 3). A distinct 785 bp fragment was consistently observed in all male plants tested. At the same time, it was conspicuously absent in female and monoecious plants. This specific fragment 785 bp, generated by the *ISSR880* primer, exhibited a strong association with the male sex phenotype in hemp. The presence of this band exclusively in male individuals and its absence in females and monoecious plants suggest its potential as a sex-linked marker in hemp (Fig. S5).

Table 3 The marker that shows a significant association with male-gender using Mixed Linear Model (MLM), incorporating Population structure (Q) and Kinship matrix (K) models in industrial hemp

Trait	Marker	*p*	MarkerR^2	MAF
Male Gender	*ISSR880bp785*	7.15E−06	0.50	0.44

p The values of the association effect and significance, R^2 phenotypic variance imparted by each marker, *MAF* Minor Allele Frequency

Discussion

CBD variability in industrial hemp individuals

While industrial hemp is primarily cultivated for fiber and seed production, the presence of cannabinoids adds significant value to these crops. To optimize harvest outcomes, it is essential to understand the changes in chemical composition at various stages of hemp growth, which is pivotal in maximizing the value of this industrial crop. As a foundational step toward establishing a cannabis breeding program, our study aimed to characterize the CBD content in 43 hemp individuals, mainly from Türkiye. Our findings highlight the potential of Turkish hemp as a new region for diverse cannabis landraces with varying CBD content. Delving into various plant parts and growth stages, we uncovered substantial variations in CBD levels. Female plants, particularly during the flowering stage, exhibited significantly higher CBD content than other plant parts, highlighting the importance of precise developmental timing in maximizing cannabinoid production. This finding aligns seamlessly with previous research, including works by Burgel et al. (2020), Yang et al. (2020), and Abdollahi et al. (2021) all of which consistently demonstrated that the flowering stage represents the peak of CBD synthesis. This consensus not only strengthens the significance of our observations but also emphasizes the pivotal role played by the flowering stage in cannabinoid production across diverse hemp genotypes.

Furthermore, our comprehensive analysis indicated a general decrease in CBD contents across all studied individuals at the harvesting stage, highlighting the need for precision in timing the harvest to optimize CBD yields. Specific individuals, specifically Mısır Çarsısı, Maltepe1, and Van, demonstrated notably elevated CBD content (4.965%, 5.309%, and

8.501%) in female inflorescences. These landraces, identified as valuable genetic resources, hold promise for developing high-CBD hemp strains. These landraces exhibit substantial CBD content, underscoring their potential for medicinal and industrial uses.

In the Yozgat region, our investigation uncovered significant variations in CBD content among the same hemp cultivars, namely "Finola, USO 31, Santhica 27, Futura 75, Ferimon, and Fedora 17," with respective values of 0.554%, 0.822%, 2.040%, 3.086%, 1.794%, and 2.280%. Notably, these values were markedly higher than the CBD levels reported by Burgel et al. in 2020, which were 0.152%, 0.111%, 0.114%, 0.147%, 0.233%, and 0.331% for the corresponding cultivars.

Furthermore, Glivar et al. (2020) also documented lower CBD concentrations within two seasons for USO 31, Santhica 27, Ferimon, and Fedora 17, with values (except for Finola) of 0.757% and 0.479%. It is noteworthy that Finola exhibited relatively similar CBD concentrations in both our study and the findings of Glivar et al. (2020). Notably, Turkish individuals consistently displayed significantly higher CBD levels in comparison. These disparities in CBD content among the same hemp cultivars in different investigations highlight the influence of environmental factors, cultivation practices, and possibly genetic variations on cannabinoid production. The observed disparity in CBD content between Turkish, Unkown, and Foreign cultivars underscores the intricate nature of cannabinoid synthesis in industrial hemp.

Further genomic and molecular investigations are essential to unravel the underlying mechanisms to identify the specific genetic factors contributing to the elevated CBD levels observed in Turkish individuals. Environmental factors, encompassing climate, soil composition, and sunlight exposure, profoundly influence plant metabolism, including cannabinoid biosynthesis. Variations in these parameters between Turkish cultivation regions and foreign hemp-growing areas could underpin the observed differences in CBD content. Microclimatic conditions, in particular, might modulate the expression of genes related to cannabinoid production. Exploring the specific environmental nuances of each region is essential for gaining insights into the ecological regulation of CBD synthesis. The disparity in sample size between Turkish, Unknown, and foreign individual's raises intriguing questions about the potential influence of sample size on CBD content differences. It is worth noting that the notion that the smaller number of Unknown or foreign cultivars caused the lower CBD synthesis in their types should be reconsidered, especially since published reports by many researchers for the same sorts showed lower CBD contents for the same individuals; it aligns relatively with our conclusion. This consistency suggests that factors beyond sample size might be at play, urging further comprehensive investigations into the intricate dynamics of cannabinoid synthesis in different hemp genotypes.

Genetic population structure in *Cannabis Sativa*

Indeed, recent years have witnessed a surge in research exploring the genetic diversity within Cannabis species, including industrial hemp, employing molecular tools like ISSR markers. These investigations have made significant contributions to our understanding of the genetic structure of Cannabis populations. There are many reasons for utilizing ISSR primers to discern the genetic variatıon of industrial hemp: versatility and polymorphism of ISSR markers, cost-effectiveness and ease of use, and rapid results and applicability in population genetics. Moreover, several investigations have successfully utilized ISSR markers in various aspects of Cannabis genetics research. These investigations have explored genetic diversity, hermaphroditism, population structures, and the characterization of seized Cannabis, highlighting the versatility and effectiveness of ISSR markers in addressing diverse research questions in Cannabis genetics. Structure analyses unraveled intricate patterns of intraspecific differentiation. The model-based clustering approach identified three distinct populations (K = 3) within the studied individuals of *Cannabis sativa*, supported by a substantial Delta K value of 7, signifying robust clustering. These clusters correlated with the sex of the plants, revealing evident genetic distinctions among males, females, and monoecious individuals. Mean Fst values (0.3168, 0.1479, and 0.1115) indicated varying degrees of genetic differentiation among the three clusters, supporting the existence of distinct genetic subpopulations. Additional peaks in the Delta K plot suggested finer-scale genetic structures and further exploration with K5 and K8 uncovered other layers of gene structuring (Fig. S3). Exploring higher K-values (K = 5 and K = 8) revealed additional layers of genetic

structuring in the study. At K = 5, individuals not only clustered by sex but also showed relative grouping based on geographical origins, highlighting the influence of both genetic and geographical factors on population diversity. At K = 8, individuals were partitioned into seven clusters, showing similarities based on both gender and geographical origin. The observed genetic lineages within Turkish females, Unknown females, and various male groups emphasized the intricate interplay of genetic factors and geographic origins in shaping genetic diversity among *Cannabis sativa* individuals, suggesting and highlighting shared genetic elements between these individuals. This aligns with Liu et al. (2023) exploration of the genomic basis of geographical differentiation and breeding selection for plant architecture traits.

Genetic analysis of individuals and hemp population relationships

In the Turkish subset, both males and females exhibited diverse genetic configurations. Females showed a slightly lower average genetic distance. Analyzing the genetic distances between Turkish females and Turkish males reveals a relatively common genetic distance, indicating a moderate genetic divergence within the Turkish Cannabis population. While notable, this low divergence and the close genetic proximity indicate shared genetic traits influenced by historical and geographical factors, potentially indicating gender-specific gene flow. Unknown females and males displayed varying genetic distances, highlighting overall diversity. Despite geographical disparities, Unknown males exhibited close genetic relationships. In monoecious individuals, genetic distances indicate moderate diversity. The genetic identity among monoecious populations and their divergence from other groups is noteworthy. The relatively genetic identity values within the monoecious populations underscore the genetic coherence among monoecious individuals, indicating a shared genetic background despite geographical distinctions.

Gender-based variations were apparent, with lower genetic distances in females and an intermediate value in monoecious groups, suggesting nuanced genetic differentiation. It has been observed that males of Turkish origin and those of Unknown origin share a high genetic identity value of 0.973. Similarly, Turkish females were compared with Unknown origin and foreign monoecious females. Results indicate a genetic identity value of 0.957 and 0.962, respectively. These values indicate a substantial genetic overlap between Turkish, Unknown, and foreign populations, emphasizing the potential historical gene flow and shared genetic heritage. Genetic identity analysis can guide breeders in selecting parent plants with optimal genetic distances to enhance desired traits. The observed genetic distance also highlights the potential for harnessing diverse genetic traits from these individuals, contributing to developing hybrid cultivars with desired characteristics.

Notably, the data PCoA displayed logical arrangements for the individuals or populations, with the closest genetic distance observed between monoecious and Unknown female groups. Conversely, the most significant genetic disparities were evident in the placements between Unknown females or monoecious and Turkish males.

Clear and separate genetic profiles by PCoA for females and males of Turkish, Unknown, and monoecious underscored significant gender distinctions. Despite the proximity between Turkish and Unknown males, their distinct clustering suggested similar genetic sources of localized adaptation or historical isolation, highlighting regional factors' impact on genetic differentiation. An intriguing revelation was identifying an admixture cluster comprising Unknown females, monoecious populations, and a Turkish female individual, indicating complex genetic intermingling and potential shared genetic heritage or historical connections. This admixture suggests interactions and gene flow among diverse groups, providing valuable insights into their evolutionary history. Focusing on different gender populations, the PCoA highlighted significant distinctions between males, females, and monoecious individuals, with males maintaining a considerable separation from females. The close association between the monoecious and female populations indicated higher genetic similarity. Furthermore, the PCoA based on both gender and geographic origin revealed precise divergent hereditary positions, emphasizing shared genetic traits between monoecious and Unknown female groups and significant genetic disparities between Unknown females and Turkish male groups. These findings underscore the influence of both gender and

geographic origin on the genetic structure of the *Cannabis sativa* populations studied.

Analysis of molecular variance

AMOVA analysis elucidated the intricate genetic dynamics in *Cannabis sativa* populations, highlighting the substantial impact of gender and geographical origin on overall genetic diversity. Eight percent of the total genetic variance was ascribed to population differences, emphasizing the vital role of localized environmental factors and individual genetic variation within designated groups. This finding resonates with prior research, aligning with research by Soorni et al. (2017) and Soler et al. (2017), highlighting substantial within-population variation compared to between-population estimates. The high variation within Iranian Cannabis populations (93.09% to 95.74%) observed by Soorni et al. underscores the importance of understanding the complexity of genetic diversity within specific regions. Additionally, Soler et al.'s findings, indicating significant variation among individuals within cultivars, emphasize the role of within-individual variation, especially concerning heterozygosity, in shaping the genetic landscape of Cannabis populations. The study by Shams et al. (2020) further supports our findings, emphasizing the intricate genetic diversity among studied populations. Their reported high genetic differentiation (Gst = 0.36) and commendable genetic diversity (Hs = 0.141) highlight the complexity of genetic interactions within Cannabis populations. Our study unveiled substantial differences between Turkish, Unknown, and foreign monoecious populations. Turkish female and male groups exhibited higher genetic diversity, as indicated by various parameters such as the number of alleles, Shannon's Information Index, Expected Heterozygosity, and Unbiased Expected Heterozygosity, indicating a broader genetic pool within these groups. Private congenital bands in Turkish females and males underscored gender-specific genetic variability unique to Turkish Cannabis individuals and emphasized the substantial genetic diversity inherent in Turkish Cannabis. In contrast, Unknown populations showed fewer private bands. The absence of commonly observed bands (<=25% and <=50%) across all populations is intriguing, suggesting a scarcity of widely shared genetic traits and highlighting the distinctive genetic composition within each population. These findings underscore the discernible genetic structure differentiating Turkish populations from foreign counterparts or Unknown populations. Regional adaptations, historical factors, or selective pressures potentially influence this structure.

Phylogenetic relationships

The resulting dendrograms revealed a complex phylogenetic structure characterized by distinct clusters, shedding light on the genetic profile of these *Cannabis sativa* individuals. Turkish female individuals displayed unique genetic profiles, with separate branches for Narlisaray and Maltepe2. This diversity emphasizes the significant genetic variation even among individuals of the exact geographical origin. Such different genetic signatures within Turkish female individuals underscore the importance of targeted breeding efforts and conservation strategies for preserving these unique genetic resources. The giant main branch (A) encompassed a diverse array of female, male, and monoecious plants, revealing intricate patterns of genetic relatedness. One sub-branch highlighted the homogeneity among nine genetically similar Turkish females (A-1), indicating shared genetic ancestry and a potential common breeding history. In contrast, another sub-branch exhibited a mixed composition, comprising Finola, Unknown females, monoecious plants, and Turkish females. Despite geographical disparities, this diverse group showed substantial genetic affinity, suggesting shared evolutionary trajectories and historical genetic exchanges among these populations. The proximity of Finola, Unknown females and monoecious plants within these sub-branches confirmed intriguing possibilities of gene relatedness, transcending geographical distances. Turkish males formed a predominant cluster within the second main branch (B), indicating a strong genetic affinity among these individuals. This clustering suggests the presence of shared genetic elements, reflecting a common genetic heritage among Turkish male hemp plants. Further analysis revealed complexity within this main branch, with two distinct sub-branches. One sub-branch exclusively comprised males from Turkish individuals, highlighting a specific genetic lineage among these domestic cultivars. In contrast, the other sub-branch (B-2) included a mix of Turkish

and Unknown males, precisely the cultivar "Unk-5." The coexistence of Turkish, Finola, and Unknown males in the same sub-branch (A-2) also indicates potential genetic kinship between these plants, suggesting genetic relationships that transcend geographical boundaries.

Significance of the identified sex-associated marker in hemp breeding

The detected MTA indicated a statistically significant association between the molecular marker *ISSR880bp785* and the male gender trait, with an explanation of 50% of phenotypic variation suggesting extremely substantial explanatory power for the observed variability in gender. These findings offer valuable insights into the potential role of the *ISSR880bp785* marker in influencing the gender trait in the studied plan. The discovery of the sex-associated 785-bp fragment, *ISSR880*, marks a significant breakthrough for the industrial hemp industry, carrying crucial implications for cultivation practices. Accurate gender identification is essential for optimizing hemp breeding efforts, with male plants playing a pivotal role in controlled cross-pollination to develop new hemp cultivars. This gender-specific marker expedites the identification and selection of male plants, streamlining the breeding process and facilitating the creation of hemp cultivars tailored to specific purposes. While *ISSR880* shows promise, acknowledging potential challenges is essential.

Further validation research must confirm its specificity across hemp cultivars, assess stability under diverse environmental conditions, and ensure applicability to different genetic backgrounds. Collaborative efforts among geneticists, breeders, and biotechnologists are crucial to addressing these challenges and fully leveraging *ISSR880*'s potential in industrial hemp breeding. The development of user-friendly assays, including sequence-characterized amplified region (SCAR) markers, should be a focus to facilitate widespread adoption in hemp breeding programs guided by the gender-specific marker *ISSR880*. Marker-assisted selection becomes a powerful tool, providing hemp breeders with a precise method for selecting male plants, ensuring a more efficient and targeted breeding process in industrial hemp cultivation.

Conclusion

In this investigation, our scrutiny delved into the intricate realms of cannabinoid variability and genetic diversity within *Cannabis sativa*, with a particular focus for the first time on Turkish hemp populations, a novel exploration in the scientific domain. Discerning noteworthy differences in CBD content across diverse plant parts and growth stages underscores the necessity for meticulous harvesting practices. The identification of high-CBD Turkish individuals unveils promising prospects for the augmentation of breeding programs aimed at optimizing cannabinoid yields. This regional specificity in CBD production has significant implications for scientific research and industrial applications. Our genetic analyses unraveled pronounced distinctions between male, female, and monoecious plants alongside subtle yet discernible genetic variations segregating Turkish, Unknown, and foreign monoecious populations. The richness of Turkish hemp landraces may provide novel allelic diversity. Of paramount significance is the unearthing of a sex-associated 785-bp fragment emanating from the *ISSR880* primer, which will serve as a distinctive gender-specific marker exclusive to male individuals within the industrial hemp milieu, a revelation poised to exert transformative influence within the hemp industry. The amalgamation of these findings regarding CBD content disparities and genetic intricacies is pivotal for shaping the trajectory of future research endeavors and conservation initiatives, providing invaluable guidance for the sustainable cultivation and reasonable utilization of industrial hemp.

Acknowledgements The authors thank Yozgat Bozok University, Scientific Research Projects Unit (Project No. 6608-BİLTEM/20-408), Yozgat, Türkiye, for the financial support granted to Ahmad Alsaleh.

Author contributions AA and GY conceptualized and established the methodology. AA performed molecular characterization, CBD assessment, and statistical analysis. AA and GY wrote the manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This work was supported by Yozgat Bozok University, Scientific Research Projects Unit (Project No. 6608-BİLTEM/20-408), Yozgat, Türkiye.

Data availability The manuscript provides all the data needed to conduct this study, except the binary scores matrix

analyzed during the current study, which is available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors confirm that this article's content has no conflict of interest.

Consent for publication The authors read the manuscript and showed their willingness to publish this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abdollahi M, Sefidkon F, Peirovi A, Calagari M, Mousavi A (2021) Assessment of the cannabinoid content from different varieties of *Cannabis sativa* L. during the growth stages in three regions. Chem Biodiv 18:e2100247

Alsaleh A (2022) SSR-based genome-wide association study in Turkish durum wheat germplasms revealed novel QTL of accumulated platinum. Mol Biol Rep 49:11289–11300. https://doi.org/10.1007/s11033-022-07720-7

Alsaleh A, Doğrusoz MC, Basaran U, Tamkoc A, Avci MA (2020) Genetic diversity and molecular taxonomy study of genus *Festuca*. J Anim Plant Sci 30(4):931–943. https://doi.org/10.36899/JAPS.2020.4.0109

Alsaleh A, Baloch FS, Sesiz U, Nadeem MA, Hatipoğlu R, Erbakan M, Özkan H (2022) Marker-assisted selection and validation of DNA markers associated with cadmium content in durum wheat germplasm. Crop past Sci 73(7–8):943–956. https://doi.org/10.1071/CP21484

Baloch FS, Alsaleh A, Shahid MQ, Ciftci V, Saenz de Miera LE, Aasim M, Nadeem MA, Aktas H, Özkan H, Hatipoglu R (2017) A Whole Genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. PLoS ONE 12(1):e0167821. https://doi.org/10.1371/journal.pone.0167821

Borin M, Palumbo F, Vannozzi A, Scariolo F, Sacilotto GB, Gazzola M, Barcaccia G (2021) Developing and testing molecular markers in *Cannabis sativa* (Hemp) for their use in variety and dioecy assessments. Plants. https://doi.org/10.3390/plants10102174

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) Tassel: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635

Buntjer J (1999) Cross checker fingerprint analysis software v2.9. Wageningen University and Research Centre, The Netherlands

Burgel L, Hartung J, Pflugfelder A, Graeff-Hönninger S (2020) Impact of growth stage and biomass fractions on cannabinoid content and yield of different hemp (*Cannabis sativa* L.) genotypes. Agronomy 10:372. https://doi.org/10.3390/agronomy10030372

Chandra S, Lata H, ElSohly MA (2020) Propagation of cannabis for clinical research: an approach towards a modern herbal medicinal products development. Front Plant Sci 26(11):958. https://doi.org/10.3389/fpls.2020.00958

Chen C, Pan Z (2021) Cannabidiol and terpenes from hemp ingredients for future foods and processing technologies. J Fut Foods 1(2):113–127. https://doi.org/10.1016/j.jfutfo.2022.01.001

Crispim Massuela D, Hartung J, Munz S, Erpenbach F, Graeff-Hönninger S (2022) Impact of harvest time and pruning technique on total CBD concentration and yield of medicinal cannabis. Plants 11:140. https://doi.org/10.3390/plants11010140

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Federer WT (1956) Augmented designs. Hawaiian Planters 55:191–208

Glivar T, Eržen J, Kreft S, Zagožen M, Čerenak A, Čeh B, Benković ET (2020) Cannabinoid content in industrial hemp (*Cannabis sativa* L.) varieties grown in Slovenia. Ind Crops Prod 145:112082. https://doi.org/10.1016/j.indcrop.2019.112082

Hammami N, Prive JP, Joly DL, Moreau G (2021) Associations between cannabinoids and growth stages of twelve industrial hemp cultivars grown outdoors in Atlantic Canada. Ind Crops Prod 172:113997. https://doi.org/10.1016/j.indcrop.2021.113997

Kaler AS, Purcell LC (2019) Estimation of a significance threshold for genome-wide association studies. BMC Genom 20:618. https://doi.org/10.1186/s12864-019-5992-7

Kassambara A (2023) rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2, https://rpkgs.datanovia.com/rstatix/

Kayis SA, Hakki EE, Pinarkara E (2010) Comparison of effectiveness of ISSR and RAPD markers in genetic characterization of seized marijuana (*Cannabis sativa* L.) in Turkey. Afr J Agric Res 5(21):2925–2933

Khatak S, Mansi Ghai M, Dahiya S (2016) ISSR Marker based Inter and Intra- Specific Diversity Analysis in Different

Genotypes of Cannabis sativa. International Journal of Engineering Technology Science and Research. IJETSR. 3(4)

Lata H, Chandra S, Techen N, Khan IA, ElSohly MA (2010) Assessment of the genetic stability of micropropagated plants of *Cannabis sativa* by ISSR markers. Planta Med 76:97–100

Liu Y, Shen K, Yin C, Xu X, Yu X, Ye B, Sun Z, Dong J, Bi A, Zhao X, Xu D, He Z, Zhang X, Hao C, Wu J, Wang Z, Wu H, Liu D, Zhang L, Shen L, Hao Y, Lu F, Guo Z (2023) Genetic basis of geographical differentiation and breeding selection for wheat plant architecture traits. Genome Biol 24:114. https://doi.org/10.1186/s13059-023-02932-x

Love J, Selker R, Marsman M, Jamil T, Dropmann D, LyA VJ, Gronau QF, Šmíra M, Epskamp S, Matzke D, Wild A, Knight P, Rouder JN, Morey RD, Wagenmakers EJ (2019) JASP: graphical statistical software for common statistical designs. J Stat Softw. https://doi.org/10.18637/jss.v088.i02

Moliterni VMC, Cattivelli L, Ranalli P, Mandolino G (2004) The sexual differentiation of *Cannabis sativa* L.: a morphological and molecular study. Euphytica 140:95–106. https://doi.org/10.1007/s10681-004-4758-7

Mostafaei Dehnavi M, Ebadi A, Peirovi A, Taylor G, Salami SA (2022) THC and CBD fingerprinting of an elite cannabis collection from Iran: quantifying diversity to underpin future cannabis breeding. Plants 11:129. https://doi.org/10.3390/plants11010129

Nadeem MA, Nawaz MA, Shahid MQ, Dogan Y, Comertpay G, YıldızM HR, Ahmad F, Alsaleh A, Labhane N, Özkan H, Chung G, Baloch FS (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip 32(2):261–285. https://doi.org/10.1080/13102818.2017.1400401

Peakall R, Smouse PE (2001) GenAlEx V5: Genetic Analysis in Excel. Australian National University, Canberra, Australia, Population genetic software for teaching and research

Perrier X, Flori A, Bonnot F (2003) Methods for data analysis. In: Hamon P, Seguin M, Perrier X, Glazmann JC (eds) Genetic diversity of cultivated tropical plants. Science Publishers, Inc and Cirad, Montpellier, pp 31–63

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rode J, In-Chol K, Saal B, Flachowsky H, Kriese U, Weber WE (2005) Sex-linked SSR markers in hemp. Plant Breeding 124:167–170

Sakamoto K, Abe T, Matsuyama T, Yoshida S, Ohmido N, Fukui K, Satoh S (2005) RAPD markers encoding retrotransposable elements are linked to the male sex in *Cannabis sativa* L. Genome. https://doi.org/10.1139/g05-056

Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, Myles S (2015) The genetic structure of marijuana and hemp. PLoS ONE 10(8):e0133292. https://doi.org/10.1371/journal.pone.0133292

Sesiz U, Alsaleh A, Bektas H, Topu M, Özkan H (2024) Genome-wide association analysis of coleoptile length and interaction with plant height in durum wheat. Agron J 00:1–17. https://doi.org/10.1002/agj2.21488

Shams R, Azizi A, Hamzei J, Noroozisharaf A, Moghadam S, Kordrostami M (2020) Genetic structure and diversity of Iranian Cannabis populations based on phytochemical, agro-morphological and molecular markers. Ind Crops Prod 158:112950. https://doi.org/10.1016/j.indcrop.2020.112950

Soler S, Gramazio P, Figàs MR, Vilanova S, Rosa E, Llosa ER, Borràs D, Plazas M, Prohens J (2017) Genetic structure of *Cannabis sativa* var. indica cultivars based on genomic SSR (gSSR) markers: implications for breeding and germplasm management. Ind Crops Prod 104:171–178. https://doi.org/10.1016/j.indcrop.2017.04.043

Soorni A, Fatahi R, Haak DC, Salami SA, Bombarely A (2017) Assessment of genetic diversity and population structure in Iranian cannabis germplasm. Sci Rep 7:15668. https://doi.org/10.1038/s41598-017-15816-5

Törjek O, Bucherna N, Kiss E, Homoki H, Finta-Korpelova Z, Bocsa I, Nagy I, Heszky LE (2002) Novel male-specific molecular markers (MADC5, MADC6) in hemp. Euphytica 127:209–218

Toth JA, Stack GM, Cala AR, Carlson CH, Wilk RL, Crawford JL, Viands DR, Philippe G, Smart CD, Rose JKC, Smart LB (2020) Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. GCB Bioenergy 12:213–222. https://doi.org/10.1111/gcbb.12667

van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE (2011) The draft genome and transcriptome of *Cannabis sativa*. Genome Biol 12:R102. https://doi.org/10.1186/gb-2011-12-10-r102

Yang R, Berthold EC, McCurdy CR, da Silva BS, Brym ZT, Freeman JH (2020) Development of cannabinoids in flowers of industrial hemp (*Cannabis sativa* L.): a pilot study. J Agric Food Chem 68(22):6058–6064. https://doi.org/10.1021/acs.jafc.0c01211

Zhang LG, Chang Y, Zhang XF, Guan FZ, Yuan HM, Yu Y, Zhao LJ (2014) Analysis of the genetic diversity of Chinese native *Cannabis sativa* cultivars by using ISSR and chromosome markers. Genet Mol Res 13(4):10490–10500. https://doi.org/10.4238/2014.December.12.10

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.