



Evaluation of genetic markers for the analysis of THC levels of Cannabis sativa samples using principal component analysis – A preliminary study

S. Cisana^a, M. Omedei^a, M. Di Nunzio^b, F. Seganti^a, V. Brenzini^c, A. Coppi^c, A. Berti^d, C. Di Nunzio^e, P. Garofano^a, E. Alladio^{a,f,*}

^a Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Orbassano (TO), Italy

^b Forensic Genetics Laboratory - Legal Medicine Unit Department of Medicine, University of Barcelona, Spain

^c Department of Biology, University of Florence, Italy

^d Reparto CC Investigazioni Scientifiche di Roma, Sezione di Biologia, Roma, Italy

^e Forensic Genetics Laboratory, Ceinge-Federico II University of Naples, Italy

^f Department of Chemistry, University of Torino, Italy

ARTICLE INFO

Keywords:

Cannabis sativa

Tetrahydrocannabinol (THC)

Short tandem repeats (STRs)

Principal Component Analysis (PCA)

ABSTRACT

Cannabis sativa is a worldwide commercial plant used for medicinal purposes, food and fiber production, and also as a recreational drug. Therefore, the identification and differentiation between legal and illegal *C. sativa* is of great importance for forensic investigations. In this study, principal component analysis (PCA), an exploratory data analysis technique, was tested to correlate the specific genotype with the concentration of tetrahydrocannabinol (THC) in the samples. *C. sativa* samples were obtained from legal growers in Piedmont, Italy, and from illegal drug seizures in the Turin region. DNA was extracted, quantified, amplified with a 13-loci multiplex STR and finally analyzed with an automated sequencer. The results showed a trend in the analyzed samples as they differed by their THC content and allele profiles. PCA yielded two clusters of samples that differed by specific allele profiles and THC concentrations. Further validation studies are needed, but this study could provide a new approach to forensic investigation and be a valuable aid to law enforcement in significant marijuana seizures or in tracing illicit drug trafficking routes.

1. Introduction

Cannabis sativa is widely used worldwide as a multipurpose plant as a source of fiber, food, and oil, but is also commonly used as a medicine and recreational drug due to its Δ^9 -tetrahydrocannabinol (Δ^9 -THC) content [1,2]. Despite the wide range of legal uses for cannabis, cultivation, possession, and sale are still prohibited by law in several countries. In Italy, the legal limit for Δ^9 -THC content is 0.2 %, based on D.P. R. 309/1990 Testo Unico on psychoactive drugs. Recently, a tolerance threshold of 0.6 % was introduced for industrial purposes (Legge n. 242 del 2 dicembre 2016) [3], by taking into account the variability that can occur during the growing process in cultivation, so that the plants with Δ^9 -THC content within 0.6 % are not criminal (and therefore seized and destroyed). However, the interest in the 0.6 % threshold stems from the fact that products with a Δ^9 -THC content greater than 0.2 % are sold illegally and are found in seizures by investigative agencies aimed at assessing the effective psychoactive potency of plants, which is

commonly attributed to products with a Δ^9 -THC content greater than 0.5–0.6 %. In this study, a multidisciplinary procedure with principal component analysis (PCA) was applied to *C. sativa* samples to correlate STR profiles and Δ^9 -THC concentration, considering the threshold of 0.6 % Δ^9 -THC.

2. Materials and methods

C. sativa samples were obtained from legal growers in Piedmont (Italy) and from illegal drug seizures in the Turin region and analyzed at the Centro Regionale Antidoping e di Tossicologia "A. Bertinaria" in Orbassano (Turin, Italy). Gas chromatography coupled with mass spectrometry (GC-MS) [4] was used to quantify the Δ^9 -THC content of the samples, while a 13-loci multiplex system STR [5–8] was tested to obtain their STR profile. Both methods were fully optimized and validated. Subsequently, a sparse version of PCA [9] was applied to analyze and examine the collected STRs data and evaluate the behavior of the

Abbreviations: THC, Tetrahydrocannabinol; STRs, Short tandem repeats; PCA, Principal Component Analysis.

* Corresponding author at: Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Orbassano (TO), Italy.

<https://doi.org/10.1016/j.fsigss.2022.10.004>

Received 17 September 2022; Accepted 5 October 2022

Available online 6 October 2022

1875-1768/© 2022 Elsevier B.V. All rights reserved.

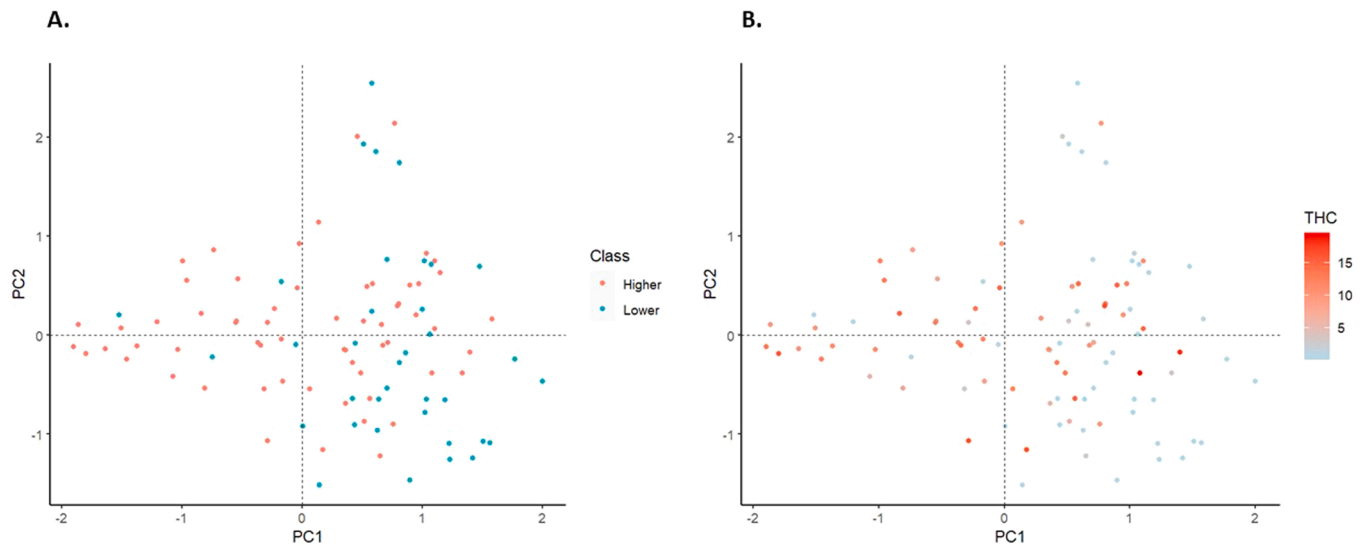


Fig. 1. PCA scores plots with samples are colored according to $\Delta 9$ -THC % values defined by quantitative analysis GC-MS. In Figure A, samples with $\Delta 9$ -THC % levels greater than 0.6 % w/w are labeled as "higher" (red circles), while samples with $\Delta 9$ -THC % levels less than 0.6 % w/w are labeled as "lower" (blue circles). In Figure B, a color scale is used to represent the *C. sativa* samples according to their measured $\Delta 9$ -THC % content.

samples in terms of their collection category and $\Delta 9$ -THC content. *C. sativa* samples were classified into two categories (i.e., 'lower' and 'higher') based on their $\Delta 9$ -THC content, using a threshold of 0.6 % w/w. PCA was calculated using the R environment (version 4.1.3) [10]. For this purpose, the following R packages were used: dplyr [11], mixOmics [12] and plotly [13].

3. Results and discussion

Sparse PCA plotted the collected data and assessed the presence of subgroups or clusters within the observations, noting the separation into two main groups: legal and illegal samples. Along the PC1 axis (x-axis), a trend can be observed where most illegal samples are located on the left side of the graph with the most negative PC1 values, while legal samples have higher PC1 values on average (Fig. 1A). The correlation of PCA scores with $\Delta 9$ -THC % values is shown in Fig. 1B. Again, a trend along PC1 can be observed, with most of the legal samples on the right side of the graph, while the illegal samples have on average the lowest PC1 values.

4. Conclusions

The objective of this proof-of-concept study was to develop a multidisciplinary method to analyze the correlation of STR profiles and $\Delta 9$ -THC concentration in *C. sativa* samples. The combination of genetic profiles and measurement of $\Delta 9$ -THC concentration was used using a sparse PCA approach. Interesting trends and results were observed in separating samples into legal and illegal samples, according to the threshold of 0.6 % w/w. In the future, more machine learning approaches and samples of *C. sativa* will be tested to build more robust models. This approach could be a useful tool to help police trace the trafficking routes of cannabis samples and associate a cannabis plant with a specific geographic area.

Conflict of interest statement

None.

References

- [1] E. Small, A. Cronquist, A practical and natural taxonomy for caNNABIS, *Taxon* 25 (1976) 405–435, <https://doi.org/10.2307/1220524>.
- [2] S.A. Bonini, M. Premoli, et al., Cannabis sativa: a comprehensive ethnopharmacological review of a medicinal plant with a long history, *J. Ethnopharmacol.* 227 (2018) 300–315, <https://doi.org/10.1016/j.jep.2018.09.004>.
- [3] 2020 (https://www.gazzettaufficiale.it/atto/stampa/serie_generale/originario) (accessed April 30, Legge 2 dicembre 2016, n. 242. Disposizioni per la promozione della coltivazione e della filiera agroindustriale della canapa).
- [4] M. Pourseyed Lazarjani, S. Torres, et al., Methods for quantification of cannabinoid: a narrative review, *J. Cannabis Res* 2 (2020) 35, <https://doi.org/10.1186/s42238-020-00040-2>.
- [5] H.J. Alghanim, J.R. Almirall, Development of microsatellite markers in Cannabis sativa for DNA typing and genetic relatedness analyses, *Anal. Bioanal. Chem.* 376 (2003) 1225–1233, <https://doi.org/10.1007/s00216-003-1984-0>.
- [6] R. Houston, M. Birck, et al., Evaluation of a 13-loci STR multiplex system for Cannabis sativa genetic identification, *Int. J. Leg. Med.* 130 (2016) 635–647, <https://doi.org/10.1007/s00414-015-1296-x>.
- [7] R. Houston, M. Birck, et al., Developmental and internal validation of a novel 13 loci STR multiplex method for Cannabis sativa DNA profiling, *Leg. Med.* 26 (2017) 33–40, <https://doi.org/10.1016/j.legalmed.2017.03.001>.
- [8] M. Di Nunzio, V. Agostini, et al., A Ge.F.I. – ISFG European collaborative study on DNA identification of Cannabis sativa samples using a 13-locus multiplex STR method, *Forensic Sci. Int.* 329 (2021), 111053, <https://doi.org/10.1016/j.forsciint.2021.111053>.
- [9] R. Bro, A.K. Smilde, Principal component analysis, *Anal. Methods* 6 (2014) 2812–2831, <https://doi.org/10.1039/c3ay41907j>.
- [10] R Core Team, R: A language and environment for statistical computing, (2014). (<http://www.r-project.org/>).
- [11] H. Wickham, R. François, L. Henry, K. Müller, dplyr: A Grammar of Data Manipulation, (2020). (<https://cran.r-project.org/package=dplyr>).
- [12] F. Rohart, B. Gautier, A. Singh, K.A. Lê Cao, mixOmics: an R package for 'omics feature selection and multiple data integration, *PLoS Comput. Biol.* 13 (2017), <https://doi.org/10.1371/journal.pcbi.1005752>.
- [13] C. Sievert, plotly for R, (2018). (<https://plotly-r.com>).