International Journal of Legal Medicine (2020) 134:1603–1612
https://doi.org/10.1007/s00414-020-02338-5

ORIGINAL ARTICLE

Evaluation of two 13-loci STR multiplex system regarding identification and origin discrimination of Brazilian *Cannabis sativa* samples

Lucas de Oliveira Pereira Ribeiro[1,2] · Eduardo Avila[1,2,3] · Roberta Fogliatto Mariot[2,4] · Mauro Sander Fett[2,3,5] · Flávio Anastácio de Oliveira Camargo[2,5] · Clarice Sampaio Alho[1,2]

Received: 16 December 2019 / Accepted: 9 June 2020 / Published online: 25 June 2020

Abstract

According to the Brazilian Federal Police (BFP), the Brazilian *Cannabis sativa* illicit market is mainly supplied by drugs originated from Paraguay and Northeastern Brazil (Marijuana Polygon region). These two known routes, the increasing indoor cultivations (supported by online market), and drugs from Uruguay are also in BFP's sight. Forensic tools to aid police intelligence were published in the past years. In genetics, microsatellites have gained attention due to their individualization capability. This study aims to evaluate the effectiveness and efficiency of two STR multiplex systems previously proposed in 94 *Cannabis sativa* samples seized in Brazil. Principal coordinate analyses (PCoA), forensic parameters, and genetic structure analysis were executed. Both panels were effective in individualizing and origin discriminating all samples, and the system proposed in 2015 demonstrated better results. For this marker set, the probability of identity for a random individual is approximately one in 65 billion; also, the PCoA shows a clear genetic distinction among samples according to its origin. Bayesian inference populational structure analysis indicated a significant genetic diversity among seizure groups, matching with its origin. Overall, the STR multiplex systems were able to achieve its purpose in individualizing and differentiating, according to geographic region, Brazilian *Cannabis* sp. samples.

Keywords *Cannabis sativa* · Forensic genetics · Genetic markers · Origin tracking · Individualization

Introduction

Consumption of *Cannabis sativa* (from this point on referred to as cannabis) as a hallucinogen dates back to 4000 BC, and although consumption levels have stabilized or declined in Europe, in the America Continent, overall use has increased [1, 2]. According to the last survey published by the United Nations Office on Drugs and Crime, America has the highest rate of marijuana production and traffic, with four out of the top five countries with the largest drug markets located in this continent. Behind the United States of America (USA), Mexico, Paraguay, and Egypt, respectively, Brazil is the fifth nation in that list. In 2016, 225 tons of cannabis was seized by the Brazilian Federal Police (BFP—Polícia Federal Brasileira), 5% of all seized marijuana worldwide [2].

It is known through police reports and investigations that Brazil's cannabis market is mainly supplied by drugs cultivated in three countries: Paraguay, Brazil, and Uruguay. South, Southeast, and Midwest geopolitical regions are supplied by Paraguay and Uruguay, while North and Northeast are

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00414-020-02338-5) contains supplementary material, which is available to authorized users.

✉ Lucas de Oliveira Pereira Ribeiro
oprlucas@gmail.com

[1] Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil

[2] Instituto Nacional de Ciência e Tecnologia – INCT Ciências Forenses, Rio de Janeiro, Brazil

[3] Departamento da Polícia Federal no Rio Grande do Sul, Porto Alegre, Brazil

[4] Department of Chemistry and Biochemistry, Florida International University, Miami, FL 33199, USA

[5] Departamento de Ciências do Solo, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

supplied by the Marijuana Polygon ("Polígono da Maconha")—a region around São Francisco river basin comprising two Brazilian states: Bahia and Pernambuco. Moreover, the exploration of the Solimões Route ("Rota do Solimões"—a pathway alongside Solimões river used to traffic cocaine into Brazil) and the online market of cannabis seeds also contribute to the supply chain [2–6].

The rise in cannabis recreational and medicinal consumption influences decriminalization and legalization movements all over the globe. These phenomena allied with police, academic, and industrial interests (e.g., genetic quality control, regulation, and traffic tracking) urge the development of tools and methods aimed to study, analyze, and control *Cannabis* sp. DNA analyses have been increasingly performed over the last two decades to identify cultivars, origin tracking, and species differentiation [7–15].

BFP, the main law enforcement agency responsible for the national narcotraffic eradication, still lacks scientific methods that could help to achieve this goal, albeit some advances in the area are emerging. For species identification, two Brazilian studies proposed the ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL) gene, present in the chloroplast, as a DNA barcode for cannabis [16, 17]. DNA barcode, although not suitable for individualization, is a technique that focuses on DNA regions that are conserved within a species and vary among them, which makes DNA barcode widely used for species differentiation.

In another approach, the analyses of short tandem repeats (STRs) have gained attention as tracking tools. Present in both vegetal and animal genomes, STRs are repetitive short sequences, up to six nucleotides, located at a defined gene, locus, or in a non-coding region [18]. STRs for cannabis were first described in 2003 [19, 20], and to date, several other studies have evaluated and reported new markers optimized for forensic purposes [12, 18, 21–23]. Aiming to validate and construct a reliable marker set, Houston et al. [12, 23] developed two multiplex panels following the International Society of Forensic Genetics (ISFG) and The Scientific Working Group on DNA Analysis Methods (SWGDAM) guidelines [24, 25]. In both studies, the authors demonstrated the viability and applicability of STR systems in individualizing and differentiating cannabis samples. To improve comprehension, the STR multiplex system based on Houston et al. study [12] will be referred to as Panel 2015, and the system based on Houston et al. [23] as Panel 2017.

In a previous study, Fett et al. [26] used Panel 2015 to evaluate 72 Brazilian cannabis samples seized by the BFP between 2014 and 2017. Results demonstrated an efficient and reliable panel, which was able to individualize and discriminate 100% of the samples based on the geographical group where they were seized. The present study aimed to compare Panel 2015's and Panel 2017's effectiveness and efficiency in individualizing and tracking the origin of 94 Brazilian cannabis samples. We also calculated forensic parameters (i.e., mean number of alleles, observed and expected heterozygosity, polymorphic content index, power of exclusion, power of discrimination, and *p* value for Hardy-Weinberg equilibrium) and analyzed genetic variance among the groups for all 19 markers collectively.

Materials and methods

Cannabis sativa samples

A total of 94 cannabis specimens were analyzed; all samples were seized within the Brazilian territory by Brazilian Federal Police. The samples were divided into seven different groups according to the year and region of apprehension. A detailed description of the groups is exhibited in Table 1.

As positive controls, 2 cannabis samples "1-D1" and "4-A2" (grouped as "EUA") with a known genetic profile for both panels were kindly provided by Dr. David Gangitano (Department of Forensic Science, College of Criminal Justice, Sam Houston State University, Huntsville, TX).

Extraction and quantification

Fett et al. [26] described two kits used for DNA isolation [27, 28]. According to their results, DNeasy Mericon Food Kit performed better and, hence, was used for DNA extraction in this study, following the same protocols. Samples' DNA concentrations were assessed by Qubit dsDNA HS Assay Kit (Invitrogen Carlsbad, CA, USA), following manufacturer instructions.

PCR multiplex conditions

Aiming to analyze and compare the two proposed panels [12, 23], two cannabis STR multiplex were evaluated. Table 2 lists all markers, their respective characteristics, final optimal concentrations, and dyes. The amplification of both panels was performed via polymerase chain reaction (PCR) using the Multiplex PCR Kit (Qiagen, Valencia, CA, USA) on a VERITI 96-well Thermo Cycler (Thermo Fisher Scientific, Waltham, MA, USA).

All 94 samples were amplified using Panel 2015 and Panel 2017. PCR volumes and conditions for Panel 2015 were performed as described in Fett et al. [26], and PCR volumes and conditions for Panel 2017 followed Houston et al. [23].

Single PCR reactions were required for samples that showed polyploidy content in one or more locus and when locus dropout occurred. All single PCR reactions were prepared at a 12.5 μL volume using 2 ng of the template (≈ 2 μL), 6.25 μL of Taq PCR Master Mix (Qiagen, Hilden, Germany), 0.625 μL of each primer (forward and reverse), and 4 μL of

Table 1 Sample characterization

	N	RA[a]	Year	Material	Origin
N15	24	Marijuana Polygon	2015	Leaves	Brazilian Northeastern
N17	12	Marijuana Polygon	2017	Leaves	Brazilian Northeastern
N18	16	Marijuana Polygon	2018	Leaves	Brazilian Northeastern
PF	31	Postal sorting and distribution facility[b]	2014–2015	Seeds[c]	Europe[d]
APR15	5	Rio Grande do Sul	2015	Pressed cannabis[e]	Paraguay[d]
APR17	5	Rio Grande do Sul	2017	Pressed cannabis[e]	Paraguay[d]
APR18	1	Rio Grande do Sul	2018	Pressed cannabis[e]	Unknown

[a] Region where the apprehension was done. [b] Facility where packages shipped to Brazil from abroad are received and processed for customs. [c] Seeds were seized by BFP and planted for confirmation tests; leaves and flowers were provided for genetic tests. [d] The region of origin could not be verified and followed the BFP reports. [e] A brick of cannabis consisted of dry leaves, flower, and stem pressed together, usually, more than one individual is present

distilled water. The single PCR reaction was also performed on the Thermo Cycler previously described and its conditions followed: activation for 5 min at 95 °C, followed by 30 cycles of 30 s at 95 °C, 90 s at 57 °C, and 30 s at 72 °C, and a final extension of 30 min at 60 °C. All PCR reactions, single and multiplex, included one negative and one positive control.

Fragment analysis

Capillary electrophoresis was carried out in a 3500 Genetic Analyzer (Thermo Fisher Scientific Inc., Massachusetts, USA), used for fragment separation and detection. For both single and multiplex PCR reactions, an aliquot of 1 μL of PCR multiplex product was added to 9.6 μL Hi-Di Formamide® and 0.4 μL LIZ® 600 Size Standard (Thermo Fisher Scientific Inc., MA, USA), totalizing 11 μL of reaction. Then, the samples were denatured at 95 °C for 3 min followed by 3 min in − 20 °C and loaded on the 3500 Genetic Analyzer. The conditions utilized for both Panels (2015 and 2017) followed as described on Houston et al. [23]: oven at 60 °C; pre-run 15 kV, 180 s; injection 1.5 kV for 8 s; run 19.5 kV, 1330 s; capillary length 36 cm; polymer POP-4™; and dye set G5.

Genotyping was performed using Gene Mapper ID-X software (Thermo Fisher Scientific Inc., MA, USA) v. 1.4. A bin set and an allelic ladder, for both panels, were kindly provided by Dr. Giantano and used in each run for genotyping and allele assignment. A threshold of 150 relative fluorescence units (RFUs) was set according to the literature [23]. Allele nomenclature used for allele calls was developed by Houston et al. [12, 23], following forensic and nomenclature guidelines [25, 29]. Budowle et al. [30] recommended for DNA mixture samples a within-laboratory standard protocol for correct assignment of peaks. In order to maintain the reliability of peak assignments, Fett et al. [26] used forensic guidelines to develop and implement a laboratory protocol for cannabis' mixture, which was used when necessary.

Locus 4910 new allele

Considering the marker 4910 present in Panel 2017, an off-bin and off-ladder peak between allele 4 and 10 was observed in 36 of all evaluated samples. To verify the existence of a new, previously unreported allele, a single PCR reaction for three homozygous samples was performed followed by a Sanger sequencing procedure. For PCR amplification, cycling sequencing, and sequencing, the same instruments described above were used. BigDye® Terminator Sequencing Kit (Thermo Fisher Scientific Inc., MA, USA) was used according to manufacturer protocol. Run conditions are as follows: oven temperature at 55 °C; pre-run 15 kV, 180 s; injection 1.2 kV for 12 s; run 15 kV, 1700 s; capillary length 36 cm; polymer POP -4[tm]; and dye set Z. The software SeqScape v 3.0 (Thermo Fisher Scientific Inc., MA, USA) was used to perform data analysis.

Statistical and concordance analyses

All analyses were performed for each Panel (2015 and 2017); as well for all 19 markers collectively, results were recorded and compared among each. To determine genetic parameters and evaluate STR marker efficiency for forensic purposes, GenAlEx v.6.503 software [31] and the STRAF online tool [32] were used. Total number of alleles (A), effective number of alleles (A_e), observed heterozygosity (H_O), expected heterozygosity (H_e), and polymorphic information content (PIC) were all calculated for each locus considering all samples as a single population and for each seizure group separately. Additionally, the power of exclusion (PE), and probability matching (PM) were calculated for each locus per Panel. The combined power of exclusion (CPE), combined probability matching (CPM), and power of discrimination (PD) were estimated for all locus on each panel [33, 34].

Principal coordinate analysis (PCoA) based on individual pairwise codominant genotypic distance and a PCoA based on

Table 2 Characteristics of all 19 STR markers and their disposal in Panel 2015 and Panel 2017

	Marker	Dye	Motif	Type of repeat	Observed alleles	Final primer concentration (μM)
Panel 2015	D02[a]	FAM	(GTT)	Simple	6, 7, 8	0.04
	C11[a]	FAM	$(TGA)_x(TGG)_y$	Compound	13, 14, 15, 18, 21	0.05
	H09	FAM	(GA)	Simple	11, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24	0.08
	B01	FAM	$(GAA)_x(A)(GAA)_y$	Compound	10, 12, 13, 14, 15	0.09
	E07	VIC	(ACT)	Simple	7, 8, 9	0.30
	305[a]	VIC	(TGG)	Simple	4, 8, 9	0.08
	308	VIC	(TA)	Simple	3, 5, 6, 7, 8, 9, 10, 11, 12, 13	0.13
	B05[a]	VIC	(TTG)	Simple	7, 8, 9, 10	0.03
	H06[a]	VIC	(ACG)	Simple	7, 8, 9	0.07
	302	PET	$(ACA)_x(ACA)_y(ACA)_z$	Compound	4, 22, 29, 31, 33, 35, 36, 37	0.08
	301	PET	(TTA)	Simple	15, 16, 17, 19, 20, 21, 22, 23, 24, 25	0.30
	501[a]	NED	(TTGTG)	Simple	4, 5, 6	0.10
	CS1[a]	NED	(CACCAT)	Simple	10, 11, 12, 14, 17, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31	0.14
Panel 2017	501[a]	FAM	(TTGTG)	Simple	4, 5, 6	0.10
	9269	FAM	(ATAA)	Simple	5.3, 6	0.10
	4910	FAM	$(AAGA)_x(TAGA)_y(AAAA)_z$	Compound	4, 5, 10, 15	0.20
	5159	FAM	(AGAT)	Simple	3, 4, 4.2, 6, 7, 10	0.30
	305[a]	VIC	(TGG)	Simple	4, 8, 9	0.10
	9043	VIC	$(TCTT)_x(CCTT)_y(TCTT)_z$	Compound	3, 5, 6, 7	0.15
	B05[a]	VIC	(TTG)	Simple	7, 8, 9, 10	0.15
	1528	VIC	(ATTA)	Simple	6, 7	0.30
	D02[a]	PET	(GTT)	Simple	6, 7, 8	0.15
	C11[a]	PET	$(TGA)_x(TGG)_y$	Compound	11, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24	0.15
	H06[a]	PET	(ACG)	Simple	7, 8, 9	0.10
	3735	NED	(TATG)	Simple	3, 4, 5, 6, 7	0.25
	CS1[a]	NED	(CACCAT)	Simple	11, 12, 14, 17, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31	0.15

[a] Markers overlapped in both panels

pairwise population F_{ST} were performed. Furthermore, analysis of molecular variance (AMOVA) with 10,000 permutations was performed to access the amount of genetic variability observed among and within groups. A Region category was created, willing to assess differences between all major seizure regions (Region Northeastern covered N15+N17+N18; Foreign stands for PF; Southern comprised APR15+APR17). The APR18 seizure group was removed from AMOVA analysis due to being comprised of a single sample. Also, Hardy-Weinberg equilibrium was estimated for all alleles ($\alpha = 0.05$).

To estimate the concordance among the panels, more than two-thirds of samples (72) were processed using the Panel 2015 [26]. The genotyping results for the seven microsatellites (D02 CANN1, C11 CANN1, ANUCS 305, B05 CANN1, H06 CANN2, ANUCS 501, and CS1) that overlapped with the Panel 2017 were recorded and compared.

Genetic structure

Broadly, the genetic structure can be defined as any pattern(s) that explains genetic diversity and distribution observed within and among a set of populations [35]. Structure v2.3.4 software was used to assess the occurrence of genetic structure among all seven seizure groups with five independent runs for each *K* value, ranging from 2 to 5. Using an admixture model and allele frequencies correlated, a 100,000-length burn-in period followed by 100,000 steps Monte Carlo Morkov Chain (MCMC) procedure was applied. To estimate the best *K* value fitting the data, the obtained results in the analysis of the genetic structure of populations were processed through CLUMPAK SERVER online tool [36]. To obtain a more accurate appraisal of genetic diversity and structure occurrence, genotypes of all 19 loci collectively were used.

Results and discussion

STR multiplex amplification

Regarding Panel 2017, all parameters and conditions, despite the change of polymer POP-7™ to POP-4™, described by Houston et al [23] were replicated in this study. All 94 samples and positive controls analyzed showed a full DNA profile with no locus dropout (Supplementary 1). However, in Panel 2015, 24 samples suffered locus dropout at ANUCS 301, ANUCS 308, B01 CANN1, H09 CANN2, ANUCS 302, and E07 CANN1 markers; a similar occurrence was also observed by Fett et al. [26] and Houston et al. [12]. Primer-primer interactions and weak primer binding are known to cause dropouts; Houston and collaborators reported primer-primer interactions of 302/D02, 302/C11, 302/308, B05/308, B02/H11, and B02/301 loci. Notwithstanding, the annealing temperature of locus 301, 302, 308, and D01 was experimentally calculated and reported as seven to five degrees less than the original PCR protocol [12]. Following Fett and collaborators, we performed a touchdown PCR (55 °C to 60 °C) aiming to minimize these dropouts, although unspecific peaks start to appear. To avoid misgenotyping of these loci, single PCR reactions, as described in "Materials and methods," were performed for all samples.

For all samples, the STR profile of the seven markers overlapped in both panels (D02 CANN1, C11 CANN1, ANUCS 305, B05 CANN1, H06 CANN2, ANUCS 501, and CS1) were obtained. Profiles were compared to evaluate the concordance between multiplex systems, and a full concordance of STR genotypes was observed for all 94 samples and positive controls.

Locus 4910 allele 5

The sequencing results confirmed an occurrence of five tandem repeats of the motif, corroborating with the hypothesis of a new allele. From the Brazilian Northeast, 28 samples (≈ 54%), 13 of them in homozygous, showed the presence of this new allele. From Southern Brazil seizures, 8 samples (≈ 72%) demonstrated the same peak. In total, 36 samples exhibited the presence of this undescribed allele 5, indicating a possible private allele for Southern America samples. This new allele was reported and can be accessed under the GenBank Accession code: MK084769.1.

Forensic parameters

Considering Panel 2015 and Panel 2017 individually, allele frequencies were calculated for all data set and used to compute the main forensic parameters (Table 3). To reach the most informative locus, we analyzed the mean number of alleles, the effective number of alleles, polymorphic content index,

Table 3 Parameters of forensic interest of Panel 2015 and Panel 2017 in all *Cannabis sativa* ($n = 94$) seizure in Brazil

	Markers	301	302	308	B01	E07	H09	305	501	B05	C11	CS1	D02	H06	1528	3735	4910	5159	9043	9269
Panel 2015	*A*	4	3	3	3	2	4	2	2	3	2	6	2	2						
	A_e	1.66	1.81	2.25	1.66	1.53	2.47	1.57	1.23	2.08	1.43	4.11	1.80	1.38						
	H_o	0.36	0.59	0.19	0.29	0.40	0.60	0.45	0.15	0.68	0.22	0.85	0.53	0.26						
	H_e	0.34	0.43	0.48	0.36	0.29	0.57	0.32	0.14	0.50	0.24	0.72	0.44	0.21						
	PIC	0.66	0.72	0.77	0.59	0.38	0.72	0.33	0.22	0.48	0.40	0.88	0.44	0.39						
	PE	0.15	0.22	0.05	0.06	0.06	0.20	0.11	0.02	0.25	0.06	0.53	0.12	0.06						
	PD	0.84	0.85	0.86	0.74	0.60	0.88	0.56	0.37	0.73	0.60	0.96	0.70	0.61						
	pHW	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.11	0.00*	0.26	0.00*	0.00*	0.00*	0.00*						
Panel 2017	*A*							2.00	1.63	2.75	2.25	5.63	2.00	2.00	1.38	2.88	2.50	2.75	2.50	1.75
	A_e							1.57	1.23	2.08	1.43	4.11	1.80	1.38	1.13	2.21	1.91	1.56	1.99	1.27
	H_o							0.45	0.15	0.68	0.22	0.85	0.53	0.26	0.08	0.44	0.58	0.34	0.51	0.09
	H_e							0.32	0.14	0.50	0.24	0.72	0.44	0.21	0.09	0.45	0.47	0.26	0.45	0.18
	PIC							0.33	0.22	0.48	0.40	0.88	0.44	0.39	0.12	0.69	0.54	0.29	0.57	0.37
	PE							0.11	0.02	0.25	0.06	0.53	0.12	0.06	0.00	0.21	0.14	0.06	0.21	0.00
	PD							0.56	0.37*	0.73	0.60*	0.96*	0.70*	0.61*	0.17*	0.86*	0.78*	0.50	0.80	0.56*
	pHW							0.11	0.00*	0.26	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.06	0.01*	0.00*

A, mean number of alleles; A_e, mean effective number of alleles; H_o, mean observed heterozygosity; H_e, mean expected heterozygosity; *PIC*, polymorphic content index; *PE*, power of exclusion; *PD*, power of discrimination; *pHW*, *p* value for Hardy-Weinberg equilibrium

* Significant *p* value

Table 4 AMOVA of Panel 2015 and Panel 2017 for all dataset

	Source	Df	SS	MS	Est. Var.	%
Panel 2015	Among Regions	3	148.694	49.565	1.194	28%*
	Among Pops	3	17.012	5.671	0.091	2%*
	Among Indiv	88	294.552	3.347	0.371	9%*
	Within Indiv	95	247.500	2.605	2.605	61%*
	Total	189	707.758		4.261	100%
Panel 2017	Among Regions	3	109.532	36.511	0.794	22%*
	Among Pops	3	22.543	7.514	0.184	5%*
	Among Indiv	88	248.061	2.819	0.217	6%**
	Within Indiv	95	226.500	2.384	2.384	67%*
	Total	189	606.637		3.580	100%

Df, degrees of freedom; *SS*, sum-of-squares; *MS*, mean of squares; %, percentage of variance explained

*P(rand ≥ data) = 0.001

**P(rand ≥ data) = 0.003

power of exclusion, and power of discrimination. The three most informative loci were CS1, 308, and H09 in Panel 2015 and CS1, 3735, and 9043 in Panel 2017. Similar results were observed by Hsieh et al. [37], Valverde et al. [38], Houston et al. [12, 23], and Fett et al. [26]. It is important to mention that for Houston et al. [23] the 5159 marker was the second most informative locus, which was not observed in our sample, likely due to a high prevalence of allele 6 in the Brazilian seizures (Supplementary 2).

In five samples seized in Northeastern Brazil, one to four alleles of equal peak height were observed for locus CS1. DNA of both samples was extracted directly from leaf material, excluding mixture contamination. This locus was previously described as polyallelic and could be explicated by region duplication or polyploidy [18, 39, 40]. These events are common in vegetal species and may cause none or minimal phenotype/function interference [41]. For those samples, an internal laboratory protocol was followed to determine the genotype [26].

The *p* value for Hardy-Weinberg equilibrium was calculated for all samples as one population (Table 3) and all groups of seizures (Supplementary 3). The APR18 seizure group was not included in the calculation since it comprises only one sample. Regarding the seizure groups, the number of loci in disequilibrium ranges from 1 (in APR17 group) to 7 (in the N15 group). Considering all samples, the *p* value is significant to eleven loci in Panel 2015 and ten loci in Panel 2017. It is expected that an increasing number of loci being tested would also increase the total number of loci that are in disequilibrium and, hence, may need correction, although a rational explanation of why the observed deviations occurred is more important [42]. The observed Hardy-Weinberg disequilibrium in many loci can be explained by three main factors: (i) asexual reproduction of individuals, common and wide spread in the cannabis illicit market; (ii) the observed substructure caused by population stratification; (iii) and the selection of markers (loci) employed in this study, which are designed to provide origin differentiation among groups of samples. All mentioned factors are the main contributors to the observed disequilibrium.

Table 6 Power of discrimination combined (PD_{comb}), power of exclusion combined (PE_{comb}), probability of identity (PI) of Panel 2015, Panel 2017, and 19 markers combined for all Brazilian *Cannabis sativa* samples

	PD_{comb}	PE_{comb}	PI[a]
Panel 2015	9.999999993E-01	0.892	6.50E-10
Panel 2017	9.9999996E-01	0.878	4.30E-08
19 markers	9.999999999995E-01	0.968	4.60E-13

[a] Considering all frequencies in all dataset

The results obtained on the AMOVA (Table 4) demonstrate that, for Panel 2015, 28% of the total variance in the dataset is due to Among-Region differences and only 61% due to within-individual differences. For Panel 2017, only 22% of total differentiation is due to Among-Region differences and 67% due to within-individual differences.

Considering only three major groups, Northeastern = N15+N17+N18; PF, and Southern = APR15+APR17+APR18, the Panel 2015 was able to correctly classify 99% of the samples' origins, while Panel 2017 accuracy was only 96% (Table 5).

Table 5 Classification of samples per major group and percentage of correct assignment for Panel 2015 and Panel 2017

	Panel 2015			Panel 2017		
	Northeastern	PF	Southern	Northeastern	PF	Southern
Northeastern	51	1		49	2	1
PF		31			31	
Southern			11	1		10
Percent accuracy	99%			96%		

Northeastern = groups N15+N17+N18; PF = Cannabis samples cultivated from foreign seeds seized by Brazilian postal services; Southern = groups APR15+APR17+APR18

Further, for all parameters (i.e., power of discrimination combined, power of exclusion combined, and the probability of Identity), Panel 2015 performed slightly better than Panel 2017 (Table 6).

The principal coordinate analysis (PCoA) of Panel 2015 visually distinguishes the seizure groups based on their geographical origin (Fig. 1). The principal component values Coord.1 and Coord.2 expressed ~ 36% and ~ 11%, respectively, of the explained genetic variance (PCoA of Panel 2017 can be consulted in Supplementary 4).

Northeastern Brazil samples even though seized in three different years (N15, N17, and N18), clearly formed a group together (exception of sample 2.3 from N17). Cannabis apprehended by law enforcement in the Rio Grande do Sul, state of Southern Brazil, in three different years (APR15, APR17, and APR18), also formed a separate group. Police investigation and academic studies suggest that the illicit drug markets of Rio Grande do Sul and Southern Brazil region are supplied by Paraguay [43, 44], probably due to the region proximity and extensive borders across both countries. The apprehensions done by BFP in postal facilities (PF) formed a long and scattered group along the *y*-axis. BFB reports suggest that the seeds came from Europe, as European websites are commonly used for seed acquisition and all brands identified in the labels could be found in those sites. Precise origin location could not be determined, which could explain the high dispersion of the group. Fett et al. [26] found similar results when analyzing 72 samples of the same dataset.

Overall, Panel 2015 proposed by Houston et al. [12] is more efficient and effective in identifying and region discriminating Brazilian cannabis samples. This result is somehow unexpected since Panel 2017 was supposed to present enhancements when compared with Panel 2015 [23]. Considering our data set, the random match probability (chance of two unrelated samples to have the same genotype) is one in 65 billion for Panel 2015, and one in 430 million for Panel 2017 (Table 6). The reduction of effectiveness and efficiency not expected for Panel 2017 could be explained by low heterozygosity presented by Brazilian samples in the new seven proposed loci (Table 3). It is important to mention that,

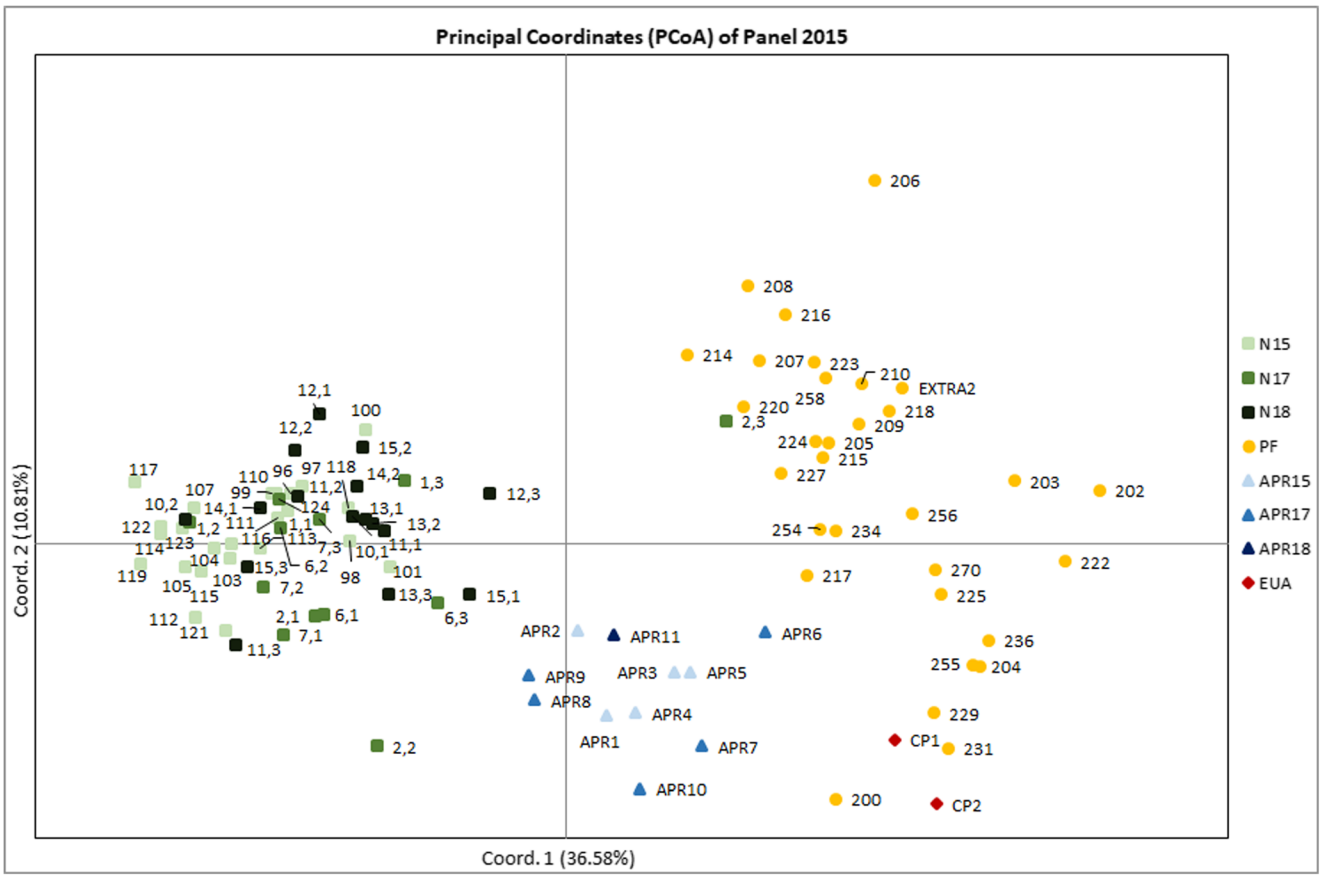

Fig. 1 Principal coordinate analysis (PCoA) of Panel 2015 based on genetic distance of all data set ($n = 94 + 2$ positive controls). Percentage of variation explained by the axes expressed inside the parenthesis. N15, *Cannabis* sp. samples seized at Northeastern Brazil in 2015; N17, *Cannabis* sp. samples seized at Northeastern Brazil in 2017; N18, *Cannabis* sp. samples seized at Northeastern Brazil in 2018; PF, *Cannabis* sp. seeds apprehended in postal office and cultivated by law enforcements; APR15, official drug seizures from Rio Grande do Sul in 2015; APR17, official drug seizures from Rio Grande do Sul in 2017; APR18, official drug seizures from Rio Grande do Sul in 2018; EUA, *Cannabis sativa* samples seized in USA used as positive control

Table 7 List of private alleles observed per group of samples for all markers

Seizure group	Locus	Allele	Freq
N15	308	6	0.021
N15	H09	22	0.042
N15	H09	23	0.021
N17	302	35	0.083
N18	B01	12	0.156
N18	C11	18	0.031
N15 + N17	5159	7	0.031
N17 + N18	301	23	0.067
Northeastern Brazil	301	19	0.031
Northeastern Brazil	CS1	25	0.101
Northeastern Brazil	CS1	26	0.313
Northeastern Brazil	H09	12	0.031
PF	308	3	0.048
PF	308	11	0.032
PF	4910	15	0.032
PF	5159	10	0.032
PF	9043	7	0.016
PF	301	24	0.065
PF	301	25	0.016
PF	302	22	0.016
PF	302	29	0.048
PF	305	9	0.032
PF	CS1	21	0.113
PF	CS1	30	0.016
PF	CS1	31	0.016
PF	H09	16	0.387
PF	H09	24	0.032
APR15 + APR17	CS1	11	0.300
APR17 +APR18	H09	17	0.350
EUA	302	4	0.500
EUA	302	37	0.250
EUA	C11	21	0.250
EUA	CS1	10	0.250

despite the better performance of Panel 2015, Panel 2017 is efficient in individualizing and discriminating Brazilian samples as well (Supplementary 5). To date, a few STR markers for cannabis were published and no national or international database was constructed. Thus, it is necessary to evaluate and select a group of markers that best suits the analyzed dataset; due to population heterozygosity, different markers can be useful for one scenario and not for others.

In order to obtain the genotypes for all 19 markers, the profiles of Panel 2015 and Panel 2017 were analyzed together. Considering all 19 loci, private alleles were observed for 11 loci (Table 7). The new allele 5 in locus 4910 was present in Northeastern (N15, N17, N18) and Southern apprehension (APR15, APR17, APR18) samples. As mentioned, the absence of this allele in PF and positive control samples suggests that it could be a private allele for South America samples. PF group presented the largest number of private alleles, possibly due to its distinct European origin and possible cannabis varieties or cultivar selection.

When analyzing results for all 19 markers collectively, an increase in power of exclusion combined and in probability of identity can be observed; however, no significant increment was observed in the power of discrimination combined (Table 6). The low gain in the information of all STR loci collectively could, also, be explained by the low heterozygosity of some markers in the Brazilian samples (Table 3).

The results of the analysis of the genetic structure of populations for all groups seem to suggest a significant genetic variance according to the major geographic region of the seizure (Northeastern, Southern, and Foreign samples) (Fig. 2). A diverse number of layers comprising general genetic structure can be visualized through alternative K value selection. Utilizing Evanno's method [45] the ideal number of clusters for our dataset was estimated as $K = 3$. The genetic difference between Brazilian and non-Brazilian samples could be observed with two clusters ($K = 2$), although at optimum K ($K = 3$) genetic structure seems to be associated with the major region origin of the seizures. At $K = 4$, some of the genetic

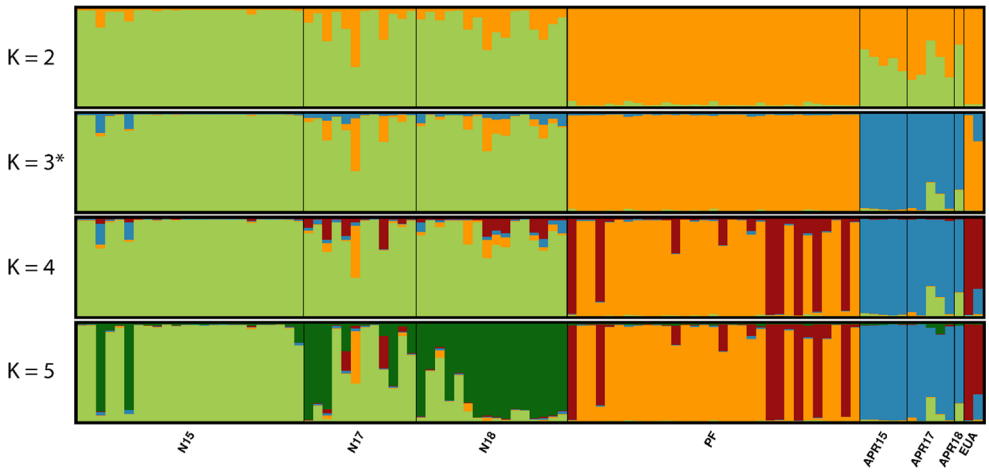

Fig. 2 Genetic structure of all samples analyzed based on 19 STR loci. Each row represents the increasing number of clusters (K). Vertical lines correspond to a different individual, with each color representing a relative proportion of association with each inferred cluster, assigned by distinct colors. Group abbreviations follow as described in "Materials and methods." *Ideal number of clusters according to Evanno's method [52]

diversity present in the PF group might be associated with North America's genetic inheritance. This association needs to be treated carefully since the North America group comprises only two samples. In addition, some level of genetic diversity was observed among Northeastern (N15, N17, N18) samples ($K = 5$).

Conclusion

The STR multiplex systems (Panel 2015 and Panel 2017) accomplished the aim of individualization and origin discrimination of all 94 Brazilian seized *Cannabis sativa* samples evaluated in this study. Despite the enhancements in Panel 2017 [12], for our dataset, the Panel 2015 demonstrated high efficiency and effectiveness regarding identification and origin discrimination of Brazilian cannabis samples. Corroborating with Fett et al. [26], Panel 2015 exhibits confident results and may be used as an intelligence tool for the Brazilian police to track drug pathways and establish a link between different crime cases. The analysis of all 19 markers collectively showed better results than Panel 2015. This increase in the informative power is expected due to the addition of six informative loci. In this study, we analyzed two 13-loci STR multiplex to obtain all 19 loci, which is not advantageous considering the effort and costs to do so. However, forensic panels with more informative loci should be considered.

To date, little is known about cannabis STR characteristics, such as global frequency, populational heterozygosity, and genome location. This may cause the choice of a non-informative, polyallelic, or genome duplicated loci when developing or analyzing forensic panels. To minimize these issues, a new microsatellite multiplex system comprising the most informative loci for Brazilian *Cannabis sativa* samples would be ideal for forensic and law enforcement purposes. Also, regarding international drug traffic and commercialization of legal cannabis, the creation of a collaborative, international database with reliable and efficient markers for all cannabis individuals would be ideal.

Acknowledgments We would like to thank Pietra Graebin and Catieli Lindholz for the technical support.

Funding information This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 (grant PRO-FORENSE 25/2014) and National Institute of Forensic Science and Technology (N°16/2014).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Iversen LL (2009) The science of marijuana, 2nd edn. Br J Clin Pharmacol 67:268. https://doi.org/10.1111/j.1365-2125.2008.03355.x
2. UNODC (2017) GLOBAL OVERVIEW OF DRUG DEMAND AND SUPPLY
3. De Backer B, Maebe K, Verstraete AG, Charlier C (2012) Evolution of the content of THC and other major cannabinoids in drug-type cannabis cuttings and seedlings during growth of plants*. J Forensic Sci 57:918–922. https://doi.org/10.1111/j.1556-4029.2012.02068.x
4. Cuypers E, Vanhove W, Gotink J, Bonneure A, van Damme P, Tytgat J (2017) The use of pesticides in Belgian illicit indoor cannabis plantations. Forensic Sci Int 277:59–65. https://doi.org/10.1016/J.FORSCIINT.2017.05.016
5. Alvarez A, Gamella JF, Parra I (2016) Cannabis cultivation in Spain: a profile of plantations, growers and production systems. Int J Drug Policy 37:70–81. https://doi.org/10.1016/j.drugpo.2016.08.003
6. Leandro Prazeres (2018) CV e Família do Norte exploram nova rota de tráfico de maconha na Amazônia - 26/02/2018 - UOL Notícias
7. Kohjyouma M, Lee IJ, Iida O et al (2000) Intraspecific variation in Cannabis sativa L. based on intergenic spacer region of chloroplast DNA. Biol Pharm Bull 23:727–730
8. Kojoma M, Iida O, Makino Y, Sekita S, Satake M (2002) DNA fingerprinting of Cannabis sativa using inter-simple sequence repeat (ISSR) amplification. Planta Med 68:60–63. https://doi.org/10.1055/s-2002-19875
9. Gao C, Xin P, Cheng C, Tang Q, Chen P, Wang C, Zang G, Zhao L (2014) Diversity analysis in Cannabis sativa based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. PLoS One 9:e110638. https://doi.org/10.1371/journal.pone.0110638
10. Valverde L, Lischka C, Erlemann S, de Meijer E, de Pancorbo MM, Pfeiffer H, Köhnemann S (2014) Nomenclature proposal and SNPSTR haplotypes for 7 new Cannabis sativa L. STR loci Forensic Sci Int Genet 13:185–186
11. Weiblen GD, Wenger JP, Craft KJ, ElSohly MA, Mehmedic Z, Treiber EL, Marks MD (2015) Gene duplication and divergence affecting drug content in Cannabis sativa. New Phytol 208:1241–1250. https://doi.org/10.1111/nph.13562
12. Houston R, Birck M, Hughes-Stamm S, Gangitano D (2015) Evaluation of a 13-loci STR multiplex system for Cannabis sativa genetic identification. Int J Legal Med 130:635–647. https://doi.org/10.1007/s00414-015-1296-x
13. Kitamura M, Aragane M, Nakamura K, Watanabe K, Sasaki Y (2016) Development of loop-mediated isothermal amplification (LAMP) assay for rapid detection of Cannabis sativa. Biol Pharm Bull 39:1144–1149. https://doi.org/10.1248/bpb.b16-00090
14. Houston R, Mayes C, King JL, Hughes-Stamm S, Gangitano D (2018) Massively parallel sequencing of 12 autosomal STRs in Cannabis sativa. Electrophoresis 39:2906–2911. https://doi.org/10.1002/elps.201800152
15. Houston R, Birck M, LaRue B, Hughes-Stamm S, Gangitano D (2018) Nuclear, chloroplast, and mitochondrial data of a US cannabis DNA database. Int J Legal Med 132:713–725. https://doi.org/10.1007/s00414-018-1798-4
16. Ribeiro A, Dias V, Mello I, Silva R, Sabino BD, Garrido RG, Seldin L, Moura-Neto RS (2013) O gene rbcL como barcode para identificação forense de Cannabis sativa. Saúde, Ética & Justiça 18:67. https://doi.org/10.11606/issn.2317-2770.v18ispep67-71
17. Mello ICT, Ribeiro ASD, Dias VHG, Silva R, Sabino BD, Garrido RG, Seldin L, de Moura Neto RS (2016) A segment of rbcL gene as

a potential tool for forensic discrimination of Cannabis sativa seized at Rio de Janeiro, Brazil. Int J Legal Med 130:353–356. https://doi.org/10.1007/s00414-015-1170-x
18. Köhnemann S, Nedele J, Schwotzer D, Morzfeld J, Pfeiffer H (2012) The validation of a 15 STR multiplex PCR for Cannabis species. Int J Legal Med 126:601–606. https://doi.org/10.1007/s00414-012-0706-6
19. Gilmore S, Peakall R, Robertson J (2003) Short tandem repeat (STR) DNA markers are hypervariable and informative in Cannabis sativa: implications for forensic investigations. Forensic Sci Int 131:65–74. https://doi.org/10.1016/S0379-0738**(02)00397-3**
20. Alghanim HJ, Almirall JR (2003) Development of microsatellite markers in Cannabis sativa for DNA typing and genetic relatedness analyses. Anal Bioanal Chem 376:1225–1233. https://doi.org/10.1007/s00216-003-1984-0
21. Howard C, Gilmore S, Robertson J, Peakall R (2008) Developmental validation of a *Cannabis sativa* STR multiplex system for forensic analysis. J Forensic Sci 53:???-??? https://doi.org/10.1111/j.1556-4029.2008.00792.x
22. Howard C, Gilmore S, Robertson J, Peakall R (2009) A *Cannabis sativa* STR genotype database for Australian seizures: forensic applications and limitations. J Forensic Sci 54:556–563. https://doi.org/10.1111/j.1556-4029.2009.01014.x
23. Houston R, Birck M, Hughes-Stamm S, Gangitano D (2017) Developmental and internal validation of a novel 13 loci STR multiplex method for Cannabis sativa DNA profiling. Legal Med 26:33–40. https://doi.org/10.1016/j.legalmed.2017.03.001
24. Linacre A, Gusmão L, Hecht W, Hellmann AP, Mayr WR, Parson W, Prinz M, Schneider PM, Morling N (2011) ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. Forensic Sci Int Genet 5:501–505. https://doi.org/10.1016/j.fsigen.2010.10.017
25. SWGDAM (2012) SWGDAM validation guidelines for DNA analysis methods. Forensic Sci Commun:1–13
26. Fett MS, Mariot RF, Avila E, Alho CS, Stefenon VM, de Oliveira Camargo FA (2019) 13-loci STR multiplex system for Brazilian seized samples of marijuana: individualization and origin differentiation. Int J Legal Med 133:373–384. https://doi.org/10.1007/s00414-018-1940-3
27. QIAGEN (2013) DNeasy® plant handbook. Qiagen, Valencia, CA
28. QIAGEN (2010) DNeasy® Mericon® food handbook. Qiagen, Valencia, CA
29. Olaisen B, Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr WR, Rand S (1998) DNA recommendations 1997 of the International Society for Forensic Genetics. Vox Sang 74:61–63
30. Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerrieri RA, Luttman JC, McClure DL (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 54:810–821. https://doi.org/10.1111/j.1556-4029.2009.01046.x
31. Smouse PE, Peakall R (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460
32. Gouy A, Zieger M (2017) STRAF-A convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet 30:148–151. https://doi.org/10.1016/j.fsigen.2017.07.007
33. BRENNER C (1990) Paternity index calculations in single locus hypervariable DNA probes : validation and other studies. Proc Int Symp Hum Identif 1989:21–53
34. Akram NA (2012) DNASF: a statistical package to analyze the distribution and polymorphism of CODIS STR loci in a heterogeneous population. J Forensic Res 3:1–7
35. Chakraborty R (1993) Analysis of genetic structure of populations: meaning, methods, and implications BT - human population genetics: a centennial tribute to J. B. S. Haldane. In: Majumder PP (ed) Springer US. MA, Boston, pp 189–206
36. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. https://doi.org/10.1111/1755-0998.12387
37. Hsieh H-M, Hou R-J, Tsai L-C, Wei CS, Liu SW, Huang LH, Kuo YC, Linacre A, Lee JCI (2003) A highly polymorphic STR locus in Cannabis sativa. Forensic Sci Int 131:53–58. https://doi.org/10.1016/S0379-0738(02)00395-X
38. Valverde L, Lischka C, Scheiper S, Nedele J, Challis R, de Pancorbo MM, Pfeiffer H, Köhnemann S (2014) Characterization of 15 STR cannabis loci: nomenclature proposal and SNPSTR haplotypes. Forensic Sci Int Genet 9:61–65. https://doi.org/10.1016/j.fsigen.2013.11.001
39. Shirley N, Allgeier L, LaNier T, Coyle HM (2013) Analysis of the NMI01 marker for a population database of Cannabis seeds. J Forensic Sci 58:S176–S182. https://doi.org/10.1111/1556-4029.12005
40. Allgeier L, Hemenway J, Shirley N, LaNier T, Coyle HM (2011) Field testing of collection cards for Cannabis sativa samples with a single Hexanucleotide DNA marker*,†. J Forensic Sci 56:1245–1249. https://doi.org/10.1111/j.1556-4029.2011.01818.x
41. Soltis PS, Soltis DE (2012) Polyploidy and genome evolution. Springer, Berlin Heidelberg
42. Ye Z, Wang Z, Hou Y (2020) Does Bonferroni correction "rescue" the deviation from Hardy-Weinberg equilibrium? Forensic Sci Int Genet 46:102254. https://doi.org/10.1016/j.fsigen.2020.102254
43. Moreira EM (2005) A despenalização do trabalhador rural preso no cultivo da Cannabis Sativa. **Confluência - Rev Interdiscip Sociol e Direito**
44. Shibuya EK, Sarkis JES, Negrini-Neto O, Martinelli LA (2007) Carbon and nitrogen stable isotopes as indicative of geographical origin of marijuana samples seized in the city of São Paulo (Brazil). Forensic Sci Int 167:8–15. https://doi.org/10.1016/j.forsciint.2006.06.002
45. EVANNO G, REGNAUT S, GOUDET J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x