JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, 2012
doi: 10.1111/1556-4029.12005
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

CRIMINALISTICS

Nicholas Shirley,[1] *B.S.; Lindsay Allgeier,*[1] *M.S.; Tommy LaNier,*[2]*; and Heather Miller Coyle,*[1] *Ph.D.*

Analysis of the NMI01 Marker for a Population Database of *Cannabis* Seeds*,†,‡

ABSTRACT: We have analyzed the distribution of genotypes at a single hexanucleotide short tandem repeat (STR) locus in a *Cannabis sativa* seed database along with seed-packaging information. This STR locus is defined by the polymerase chain reaction amplification primers CS1F and CS1R and is referred to as NMI01 (for National Marijuana Initiative) in our study. The population database consists of seed seizures of two categories: seed samples from labeled and unlabeled packages regarding seed bank source. Of a population database of 93 processed seeds including 12 labeled *Cannabis* varieties, the observed genotypes generated from single seeds exhibited between one and three peaks (potentially six alleles if in homozygous state). The total number of observed genotypes was 54 making this marker highly specific and highly individualizing even among seeds of common lineage. Cluster analysis associated many but not all of the handwritten labeled seed varieties tested to date as well as the National Park seizure to our known reference database containing Mr. Nice Seedbank and Sensi Seeds commercially packaged reference samples.

KEYWORDS: forensic science, DNA, population genetics, short tandem repeat, NMI01, *Cannabis sativa*

Cannabis sativa, marijuana, is one of the largest illicit cash crops in the United States. In the United States, criminal organizations have increasingly used remote areas of park land for the illegal cultivation of high-potency marijuana. These criminal organizations are threatening the safety of visitors to the parks, damaging the environment, and funding violent drug cartels (1,2). Although the local, state, and federal law enforcement agents have taken appropriate action to remove these sites whenever possible through a coordinated effort by the National Marijuana Initiative (NMI), Office of National Drug Control Policy, and Department of Justice, it is difficult to eliminate the problem because the geographic source (originator) of the material is largely unknown.

To further complicate tracing the origin of these cultivation sites is the lack of genetic characterization of *Cannabis*. The genetics of *C. sativa* is not well defined due to the illegal nature of the substance and the difficulty in obtaining labeled samples. Whatever genetics is listed on the Internet or on seized seed packets are subject to interpretation without having an independent method to classify the samples. In a mutual effort to define the genetics and establish an effective DNA test to genotype marijuana samples, labeled and unlabeled seed samples were surveyed with the NMI01 short tandem repeat (STR) marker originally identified by Hsieh et al. (3). Unlabeled seeds were seized from an illegal marijuana cultivation site on National Park land in the United States and genotyped with NMI01 as well.

The concept of matching unknown evidentiary samples to a known reference database is prevalent in forensic science. There are databases for: bullet and cartridge casings (Integrated Ballistic Identification System), paint (Paint Data Query and National Automotive Paint File), glass (Glass Evidence Reference Database), shoe prints (TreadMark, SoleMate, TreadMate), documents (Forensic Information System for Handwriting and International Ink Library), arson (Ignitable Liquids Reference Collection), tablet and capsule drug identification (Ident-A-Drug, PharmInfoNet), fingerprints (Integrated Automated Fingerprint Identification System), and human DNA (Combined DNA Index System [CODIS]) (4,5).

Here, we describe a seed database and establish genotypes for our population. Our database is unique in that many of the seed samples have packaging information that identifies a genetic variety and a supplier which allows us to build a reference database for comparative purposes to unlabeled marijuana evidentiary seizures. Our labeled samples are also more likely to be genetically related (sibs) than random street seizure samples from unlinked cases based on the standard genetic crossing required to maintain *Cannabis* variety heritage. This research is presented as part of an ongoing effort to eliminate large-scale illegal cultivation by drug dealers in the State and Federal parks; and as part of a DNA mapping initiative by the Drug Enforcement Agency (DEA).

[1]Forensic Science Department, Henry C. Lee College of Criminal Justice & Forensic Sciences, University of New Haven, 300 Boston Post Road, West Haven, CT 06516.

[2]Director, National Marijuana Initiative, HIDTA-ONDCP Suite 1900, 1010 Second Avenue, San Diego, CA 92101.

*Funding provided by the Office of National Drug Control Policy (ONDCP), National Marijuana Initiative (NMI).

†Presented in part at the 62nd Annual Meeting of the American Academy of Forensic Sciences (AAFS), February 22–27, 2010, in Seattle, WA; 2010 NMI-CAMP Annual Meeting, May 10–13, in San Diego, CA; Northeastern Association of Forensic Scientists (NEAFS), November 4–7, 2009, in Long Branch, NJ.

‡The research presented here in this article reflects the work and opinions of the authors and not necessarily that of the agencies listed herein.

Received 15 Aug. 2011; and in revised form 13 Dec. 2011; accepted 28 Dec. 2011.

Materials and Methods

Sample Collection

Seeds, both labeled and unlabeled, were collected for testing by Tommy LaNier at NMI. Tommy LaNier coordinated the seed collection effort with various supporting agencies specifically for this research program. The information given to the agencies for collection was to collect seeds with commercial labels and without commercial labels and send to the NMI for shipment to the University of New Haven (UNH; West Haven, CT) via a traceable shipping method. It is remarkably difficult to obtain these types of seized materials but with this coordinated effort, we were able to obtain the samples necessary for this project. Labeled sample varieties reported in this study include: Mr. Nice Seedbank commercial packages, Sensi Seeds commercial packages, and handwritten labeled packages (home-grown or repackaged samples).

Sample Processing and Preparation for Polymerase Chain Reaction

Seeds were visually identified as *C. sativa*, and DNA was extracted following manual grinding of single seeds in liquid nitrogen using a previously published procedure (6,7). For a positive control, a large sample of leaf *Cannabis* was provided courtesy of the DEA so that the same batch could be processed simultaneously with the seeds and provided a positive control for the seeds as well as a migration control for the capillary electrophoresis instrument.

Marker Specificity

The NMI01 polymerase chain reaction (PCR) primers were previously tested for species specificity to *Cannabis* by Hsieh et al. (3). However, their article did not contain the names of the plant species that were tested. The authors of that study kindly agreed to provide the information to us. The species that they tested were the following: *Humulus japonicas, Nicotiana tabacum, Dicliptera chinensis, Impatiens wallerana, Anredera cordifolia, Passiflora suberosa, Ambrosia elatior, Sedum formosanum, Polygala paniculata, Oxalis corymbosa, Plantago asiatica, Gonostegia hirta, Boehmeria densiflora, Pasania harlandii, Litsea cubeba, Randia cochinchinensis, Polygonum nepalense, Clinopodium gracile, Maesa tenera,* and *Melastoma candidum* (Dr. James Lee, personal communication). In addition, we tested *Hedera* (ivy) and *Nepeta* (catnip) as our negative controls. Any of these plant species that exhibited no PCR products could be used as negative controls.

PCR

Sample amplifications were performed using a total reaction volume of 25 μL, using 2.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems, Inc., Foster City, CA), 0.5 μL dNTP (New England BioLabs, Ipswich, MA), 2.5 μL Standard Taq reaction buffer (New England BioLabs), and 1.5 μL each of forward and reverse primer (3) (CS1F and CS1R, respectively; Applied Biosystems). The CS1F primer was labeled with a 5′-fluorescent tag, and the PCR was for 32 cycles of 94°C denaturation for 60 sec, 55°C annealing for 60 sec, 72°C extension for 60 sec, and one 72°C extension for 30 min (5).

TABLE 1—*Bin sets used for genotyping the seed population database with the NMI01 marker.*

Bin	Start	Location	End
1	121.5	122	122.5
1.1	122.5	123	123.5
1.2	123.5	124	124.5
1.3	124.5	125	125.5
1.4	125.5	126	126.5
1.5	126.5	127	127.5
2	127.5	128	128.5
2.1	128.5	129	129.5
2.2	129.5	130	130.5
2.3	130.5	131	131.5
2.4	131.5	132	132.5
2.5	132.5	133	133.5
3	133.5	134	134.5
3.1	134.5	135	135.5
3.2	135.5	136	136.5
3.3	136.5	137	137.5
3.4	137.5	138	138.5
3.5	138.5	139	139.5
4	139.5	140	140.5
4.1	140.5	141	141.5
4.2	141.5	142	142.5
4.3	142.5	143	143.5
4.4	143.5	144	144.5
4.5	144.5	145	145.5
5	145.5	146	146.5
5.1	146.5	147	147.5
5.2	147.5	148	148.5
5.3	148.5	149	149.5
5.4	149.5	150	150.5
5.5	150.5	151	151.5
6	151.5	152	152.5
6.1	152.5	153	153.5
6.2	153.5	154	154.5
6.3	154.5	155	155.5
6.4	155.5	156	156.5
6.5	156.5	157	157.5
7	157.5	158	158.5
7.1	158.5	159	159.5
7.2	159.5	160	160.5
7.3	160.5	161	161.5
7.4	161.5	162	162.5
7.5	162.5	163	163.5
8	163.5	164	164.5
8.1	164.5	165	165.5
8.2	165.5	166	166.5
8.3	166.5	167	167.5
8.4	167.5	168	168.5
8.5	168.5	169	169.5
9	169.5	170	170.5
9.1	170.5	171	171.5
9.2	171.5	172	172.5
9.3	172.5	173	173.5
9.4	173.5	174	174.5
9.5	174.5	175	175.5
10	175.5	176	176.5
10.1	176.5	177	177.5
10.2	177.5	178	178.5
10.3	178.5	179	179.5
10.4	179.5	180	180.5
10.5	180.5	181	181.5
11	181.5	182	182.5
11.1	182.5	183	183.5
11.2	183.5	184	184.5
11.3	184.5	185	185.5
11.4	185.5	186	186.5
11.5	186.5	187	187.5
12	187.5	188	188.5
12.1	188.5	189	189.5
12.2	189.5	190	190.5

(Continued)

TABLE 1—*Continued.*

Bin	Start	Location	End
12.3	190.5	191	191.5
12.4	191.5	192	192.5
12.5	192.5	193	193.5
13	193.5	194	194.5
13.1	194.5	195	195.5
13.2	195.5	196	196.5
13.3	196.5	197	197.5
13.4	197.5	198	198.5
13.5	198.5	199	199.5
14	199.5	200	200.5
14.1	200.5	201	201.5
14.2	201.5	202	202.5
14.3	202.5	203	203.5
14.4	203.5	204	204.5
14.5	204.5	205	205.5
15	205.5	206	206.5
15.1	206.5	207	207.5
15.2	207.5	208	208.5
15.3	208.5	209	209.5
15.4	209.5	210	210.5
15.5	210.5	211	211.5
16	211.5	212	212.5
16.1	212.5	213	213.5
16.2	213.5	214	214.5
16.3	214.5	215	215.5
16.4	215.5	216	216.5
16.5	216.5	217	217.5
17	217.5	218	218.5
17.1	218.5	219	219.5
17.2	219.5	220	220.5
17.3	220.5	221	221.5
17.4	221.5	222	222.5
17.5	222.5	223	223.5
18	223.5	224	224.5
18.1	224.5	225	225.5
18.2	225.5	226	226.5
18.3	226.5	227	227.5
18.4	227.5	228	228.5
18.5	228.5	229	229.5
19	229.5	230	230.5
19.1	230.5	231	231.5
19.2	231.5	232	232.5
19.3	232.5	233	233.5
19.4	233.5	234	234.5
19.5	234.5	235	235.5
20	235.5	236	236.5
20.1	236.5	237	237.5
20.2	237.5	238	238.5
20.3	238.5	239	239.5
20.4	239.5	240	240.5
20.5	240.5	241	241.5
21	241.5	242	242.5
21.1	242.5	243	243.5
21.2	243.5	244	244.5
21.3	244.5	245	245.5
21.4	245.5	246	246.5
21.5	246.5	247	247.5
22	247.5	248	248.5
22.1	248.5	249	249.5
22.2	249.5	250	250.5
22.3	250.5	251	251.5
22.4	251.5	252	252.5
22.5	252.5	253	253.5
23	253.5	254	254.5
23.1	254.5	255	255.5
23.2	255.5	256	256.5
23.3	256.5	257	257.5
23.4	257.5	258	258.5
23.5	258.5	259	259.5
24	259.5	260	260.5

(Continued)

TABLE 1—*Continued.*

Bin	Start	Location	End
24.1	260.5	261	261.5
24.2	261.5	262	262.5
24.3	262.5	263	263.5
24.4	263.5	264	264.5
24.5	264.5	265	265.5
25	265.5	266	266.5
25.1	266.5	267	267.5
25.2	267.5	268	268.5
25.3	268.5	269	269.5
25.4	269.5	270	270.5
25.5	270.5	271	271.5
26	271.5	272	272.5
26.1	272.5	273	273.5
26.2	273.5	274	274.5
26.3	274.5	275	275.5
26.4	275.5	276	276.5
26.5	276.5	277	277.5
27	277.5	278	278.5
27.1	278.5	279	279.5
27.2	279.5	280	280.5
27.3	280.5	281	281.5
27.4	281.5	282	282.5
27.5	282.5	283	283.5
28	283.5	284	284.5
28.1	284.5	285	285.5
28.2	285.5	286	286.5
28.3	286.5	287	287.5
28.4	287.5	288	288.5
28.5	288.5	289	289.5
29	289.5	290	290.5
29.1	290.5	291	291.5
29.2	291.5	292	292.5
29.3	292.5	293	293.5
29.4	293.5	294	294.5
29.5	294.5	295	295.5
30	295.5	296	296.5
30.1	296.5	297	297.5
30.2	297.5	298	298.5
30.3	298.5	299	299.5
30.4	299.5	300	300.5
30.5	300.5	301	301.5
31	301.5	302	302.5
31.1	302.5	303	303.5
31.2	303.5	304	304.5
31.3	304.5	305	305.5
31.4	305.5	306	306.5
31.5	306.5	307	307.5
32	307.5	308	308.5
32.1	308.5	309	309.5
32.2	309.5	310	310.5
32.3	310.5	311	311.5
32.4	311.5	312	312.5
32.5	312.5	313	313.5
33	313.5	314	314.5
33.1	314.5	315	315.5
33.2	315.5	316	316.5
33.3	316.5	317	317.5
33.4	317.5	318	318.5
33.5	318.5	319	319.5
34	319.5	320	320.5
34.1	320.5	321	321.5
34.2	321.5	322	322.5
34.3	322.5	323	323.5
34.4	323.5	324	324.5
34.5	324.5	325	325.5
35	325.5	326	326.5
35.1	326.5	327	327.5
35.2	327.5	328	328.5
35.3	328.5	329	329.5
35.4	329.5	330	330.5

(Continued)

TABLE 1—*Continued.*

Bin	Start	Location	End
35.5	330.5	331	331.5
36	331.5	332	332.5
36.1	332.5	333	333.5
36.2	333.5	334	334.5
36.3	334.5	335	335.5
36.4	335.5	336	336.5
36.5	336.5	337	337.5
37	337.5	338	338.5
37.1	338.5	339	339.5
37.2	339.5	340	340.5
37.3	340.5	341	341.5
37.4	341.5	342	342.5
37.5	342.5	343	343.5
38	343.5	344	344.5

Fragment Analysis

One microliter of PCR product was mixed with ROX 500 size standard and Hi-Di formamide (both, Applied Biosystems) and heat denatured for 3 min at 95°C. Fragment separation and detection were performed on an ABI 3130 Genetic Analyzer using Performance Optimized Polymer (POP7), multi-capillary DS-32 (Dye Set F) (both, Applied Biosystems). Data were analyzed using GeneMapper 4.0 software (Applied Biosystems), and alleles were assigned using a customized panel and bin set (Table 1) (UNH). The bin set was calculated by taking the size of the fragment and subtracting 90 bases flanking the 5′ side and 26 bases flanking the 3′ side (including the primer sequence) and dividing the repeat length by 6. The relative fluorescent units (RFU) threshold for peak height determination was set at a minimum of 50 RFU. For comparisons of the seed relationships, Network 4.6.0.0 software (www.fluxus-engineering.com) was used with whole number allele values for data entry. GENEMARKER v.1.97 software by SoftGenetics (www.softgenetics.com) was also used for cluster analysis.

Sizing Precision

The 3130 ABI DNA sequencer used for *Cannabis* genotyping was tested for sizing precision by preparing ABI AmpF*l*STR Profiler Plus Ladder (positive control sample) or nuclease free water (negative control sample) using the following formula. Per sample well: 24 μL Hi-Di formamide + 1 μL GeneScan 500 ROX (Applied Biosystems) size standard + either 1.5 μL of ABI AmpF*l*STR Profiler Plus Ladder or water. The ABI MicroAmp optical 96-well reaction plate was covered with MicroAmp optical adhesive film and denatured at 95°C for 3 min using the GeneAmp PCR System 9700 (Applied Biosystems). Samples were then cooled on ice for 3 min prior to performing capillary electrophoresis on a 36 cm capillary array. The data collection settings were:

Run Voltage	15,000
Injection Voltage	1200
Injection Duration	16
Temperature	60
Laser Power	15

Precision of the instrument is being included in this publication as it is critical to the sizing of the DNA fragments due to lack of a commercially available *Cannabis* allelic ladder and was used to create the custom UNH bin sets. The positive

TABLE 2—*Observed genotypes for the NMI01 marker in a* Cannabis *seed database.*

No.	Sample	Allele 1	Allele 2	Allele 3	Source
1	Royal Hawaiian	16.2	26		Unknown
2	Royal Hawaiian	23.1	26		Unknown
3	Royal Hawaiian	16.2	26		Unknown
4	Royal Hawaiian	11.1	16.2	26	Unknown
5	Nat. Park seizure	17.3	27.1		National Park Service
6	Nat. Park seizure	2.2	25.2	30.2	National Park Service
7	Nat. Park seizure	2.1	17.3	26.1	National Park Service
8	Nat. Park seizure	26.2	30.3		National Park Service
9	Nat. Park seizure	2.1	17.3	28.1	National Park Service
10	Nat. Park seizure	25.2	27.1		National Park Service
11	Nat. Park seizure	2.1	26.1	27.2	National Park Service
12	Nat. Park seizure	23.2	27.1		National Park Service
13	Dreamtime	23.2			Mr. Nice Seedbank
14	Dreamtime	16.4	28.1		Mr. Nice Seedbank
15	Dreamtime	13.4			Mr. Nice Seedbank
16	Dreamtime	2.2	11.5	13.4	Mr. Nice Seedbank
17	Dreamtime	28.1			Mr. Nice Seedbank
18	Dreamtime	2.2	23.2	28.1	Mr. Nice Seedbank
19	Dreamtime	2.2	11.5	13.4	Mr. Nice Seedbank
20	Dreamtime	16.4	29		Mr. Nice Seedbank
21	Dreamtime	2.2	11.5	13.4	Mr. Nice Seedbank
22	Dreamtime	28.1			Mr. Nice Seedbank
23	Dreamtime	16.3	23.1		Mr. Nice Seedbank
24	Orient Express	2.2	13.4	26.2	Unknown
25	Orient Express	2.2	22.3	28.1	Unknown
26	Orient Express	2.2	13.4	26.2	Unknown
27	Orient Express	13.4	26.2		Unknown
28	Orient Express	13.4	26.2		Unknown
29	Orient Express	2.2	13.4	26.2	Unknown
30	Orient Express	22.3	26.2		Unknown
31	Orient Express	13.4	28.1		Unknown
32	Orient Express	2.2	22.3	26.2	Unknown
33	Orient Express	2.2	13.4	26.2	Unknown
34	Shark Shock	16.4	28.1		Mr. Nice Seedbank

(Continued)

TABLE 2—*Continued.*

No.	Sample	Allele 1	Allele 2	Allele 3	Source
35	Shark Shock	13.4	28.1		Mr. Nice Seedbank
36	Shark Shock	16.4	28.1		Mr. Nice Seedbank
37	Shark Shock	13.4	16.4		Mr. Nice Seedbank
38	Shark Shock	16.4	28.1		Mr. Nice Seedbank
39	Shark Shock	13.4	16.4		Mr. Nice Seedbank
40	Shark Shock	13.4	28.1		Mr. Nice Seedbank
41	Shark Shock	13.4	16.4		Mr. Nice Seedbank
42	MOD	17.3	28.1		Unknown
43	MOD	17.3	28.1		Unknown
44	MOD	26.2	28.1		Unknown
45	MOD	17.3	28.1		Unknown
46	MOD	26.2	28.1		Unknown
47	MOD	26.2	28.1		Unknown
48	MOD	17.3	28.1		Unknown
49	MOD	26.2	28.1		Unknown
50	MOD	17.3	28.1		Unknown
51	MOD	17.3	28.1		Unknown
52	MOD	26.2	28.1		Unknown
53	MOD	17.3	28.1		Unknown
54	MOD	17.3	28.1		Unknown
55	Black Widow	2.2	11.5	28.1	Mr. Nice Seedbank
56	Black Widow	2.2	11.5	13.4	Mr. Nice Seedbank
57	Black Widow	2.2	13.4	20.3	Mr. Nice Seedbank
58	Black Widow	13.4			Mr. Nice Seedbank
59	Black Widow	2.2	13.4	20.3	Mr. Nice Seedbank
60	Black Widow	2.2	20.3	28.1	Mr. Nice Seedbank
61	Black Widow	2.2	20.3	26.2	Mr. Nice Seedbank
62	Black Widow	2.2	20.3	28	Mr. Nice Seedbank
63	WM	13.4	16.3		Unknown
64	WM	13.4	20.3		Unknown
65	WM	16.3			Unknown
66	WM	16.3	20.3		Unknown
67	WM	2.2	16.4	20.3	Unknown
68	WM	2.2	13.4	20.3	Unknown
69	WM	2.2	16.3		Unknown
70	WM	13.4	20.3		Unknown
71	WM	13.4	16.3		Unknown
72	La Nina	2.2	13.4	16.4	Mr. Nice Seedbank
73	La Nina	2.2	13.4	20.3	Mr. Nice Seedbank
74	La Nina	2.2	16.3	20.3	Mr. Nice Seedbank
75	La Nina	2.2	13.4	20.3	Mr. Nice Seedbank
76	La Nina	16.3	20.3		Mr. Nice Seedbank
77	La Nina	13.4	20.3		Mr. Nice Seedbank
78	Bazooka	22.2	25.2		Unknown
79	Bazooka	22.2	24.2		Unknown
80	American Dream	16.3	25.2		Sensi Seeds
81	American Dream	16.3	28		Sensi Seeds
82	American Dream	16.2	25.2		Sensi Seeds
83	American Dream	16.3	28		Sensi Seeds
84	Skunk no. 1	22.2			Gypsy Nirvana's Seed Boutique
85	Skunk no. 1	22.2	27.2		Gypsy Nirvana's Seed Boutique
86	Skunk no. 1	16.3	22.2		Gypsy Nirvana's Seed Boutique
87	Skunk no. 1	16.3	22.2		Gypsy Nirvana's Seed Boutique
88	Skunk no. 1	16.3	22.2		Gypsy Nirvana's Seed Boutique
89	Sugar Babe	26	28		Unknown
90	Sugar Babe	28			Unknown
91	Sugar Babe	28			Unknown
92	Sugar Babe	26	28		Unknown
93	Sugar Babe	13.3	28		Unknown

(Continued)

control samples were run with each data set and provided reproducible results. Given the large number of alleles detected in this study, it would be possible to create an allelic ladder in future studies for genotypic comparisons.

Results and Discussion

Instrument Precision

The 3130 ABI DNA Sequencer used for *Cannabis* genotyping was tested with replicate samples for sizing precision by preparing ABI AmpF*l*STR Profiler Plus Ladder, a commercially available standard used routinely in human identification. As anticipated from previously reported studies on STR marker validation and consistent with the instrument manufacturer, the precision of the instrument was no >0.5 base pairs, which is the generally accepted standard deviation in peak sizing used for human identification. This information was used to set the allele designations for the NMI01 marker such that our customized bin panel could consistently represent DNA fragments that migrate to the same location on our detection system. These fragments were not sequenced for nucleotide content, so our alleles are based on our bin sets for fragment size. The bin sets used for the seed database are listed in Table 1.

Marker Specificity

The NMI01 PCR primers were previously tested for *Cannabis* species specificity by Hsieh et al. (3) and by the authors using

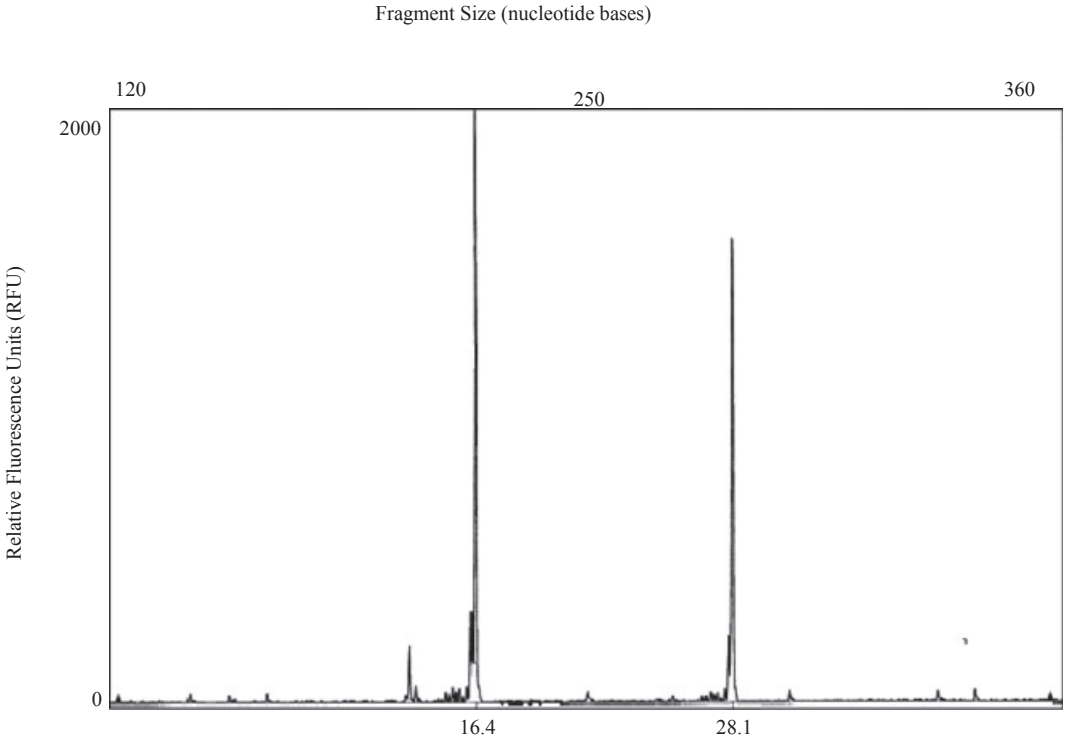

FIG. 1—*DNA electropherogram for* Cannabis sativa *"Shark Shock" seed. The genotype is a 16.4, 28.1 with* x*-axis in nucleotide bases (120–360);* y*-axis in relative fluorescence units (0–3500).*

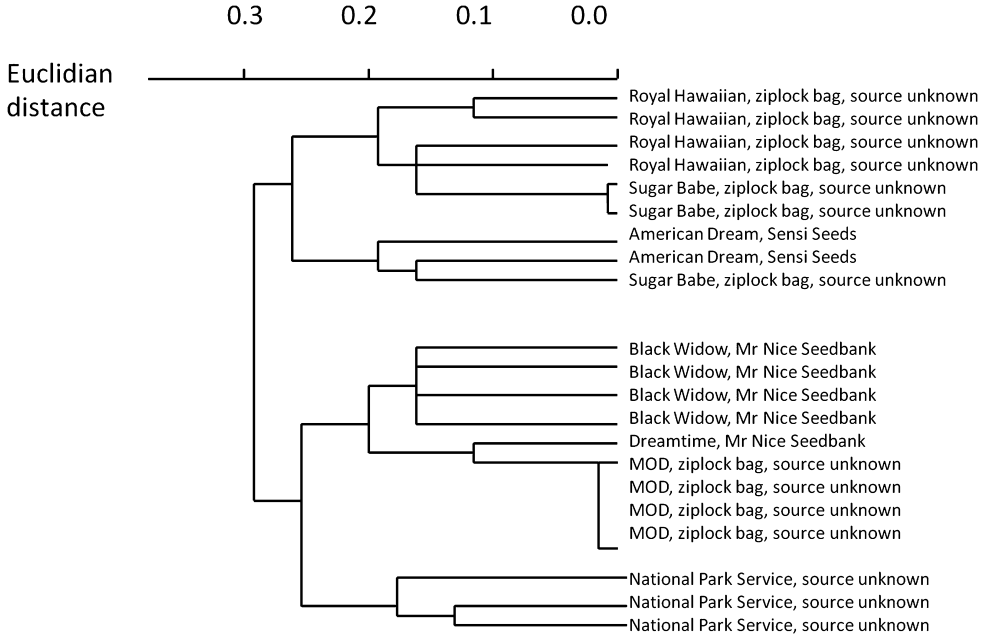

FIG. 2—*Cluster analysis of NMI01* Cannabis *seed samples.*

positive and negative control plant species. Of the species tested, no peaks were detected in plant species other than *Cannabis*. The baseline levels were low, between 0 and 30 RFU so fluorescent peaks detected in *Cannabis* were authentic DNA fragments and were scored if they reached >50 RFU set as the threshold for detection and genotype scoring.

Genotype Recovery

Of seeds processed, representing 12 labeled varieties and an unlabeled sample, a total of 54 genotypes (Table 2) was recovered. Each genotype recovered could contain up to three peaks; no more than three peaks were observed for any given sample. The multiple peaks observed could result from a number of pos-

sibilities including a duplicated locus on the same chromosome or from amplification from multiple sets of chromosomes, a genetic condition called polyploidy which is tolerated in plants (8). Another possibility is that the multiple peaks result from PCR primer homology to numerous sites within the same genome or from the combined genomes of the maternal seed coat and seed embryo, respectively (8,9). Genotypes with more than two DNA fragments were observed in bud, leaf, and seed samples, so the multiple peak pattern is not exclusively associated with marijuana seeds and hence not only a result of combined seed coat and embryo genetics. Regardless, this STR marker gives us a manner in which to genetically classify our samples for analysis and is highly reproducible in our samples.

Figure 1 depicts a sample genotype which was generated from a commercially labeled seed sample called "Shark Shock" from Mr. Nice Seedbank (website: http://www.mrnice.nl). When seeds were compared, the genotypes showed genetic variation from one seed to another, while still showing some overall genetic similarities based on number of shared alleles. When seed samples were compared, those from a common genetic heritage grouped together while others were not associated with the reference database (data not shown) and likely have a distinctly different source. Our unknown seed samples from the National Park land were found to map to our reference database based on the genetics of the NMI01 marker. The added value of having reference genetic databases for *Cannabis* allows for the sourcing or verification of unknown seed samples to a commercially packaged seed source. This is particularly useful as an investigative lead and works on the same principle as that used for familial searching in the "Grim Sleeper" case where percent shared alleles in a human database led to the identification of a serial rapist (website: http://abcnews.go.com).

The benefits to using the NMI01 marker for seed analysis are that it can identify the seed as *Cannabis* due to PCR primer specificity, and it simultaneously genotypes the sample for data basing and mapping purposes. While a single, highly specific DNA marker may not be able to uniquely individualize a seed sample in the same way that the 13 core CODIS human identification markers are utilized (4), it can be used for cluster analysis based on commonality of shared alleles similar to other familial search methods for human DNA databases (4,5). Our cluster analysis for the unknown sample from the National Park shows that it groups with labeled varieties from a common genetic heritage, which contained a foundation line from Mr. Nice Seedbank called "Black Widow." Many of the handwritten labeled seed samples also were associated with this reference database as shown in Fig. 2 indicating a shared source.

Acknowledgments

The authors extensively thank the following agencies for their participation in this research: U.S. Department of Justice, Drug Enforcement Agency, Southwest Laboratory, National Park Service, U.S. Department of the Interior, Bureau of Land Management and the San Diego County Integrated Narcotic Task Force, National Marijuana Initiative, Office of National Drug Control Policy, and University of New Haven. Many thanks to University of New Haven students Jennifer Nabozny, Amanda ZeRuth, Casey Grenier, Amanda DeBiase, Samantha Dobshinsky, Caity Boland, John Hemenway, Melissa Hemler, Carley Lobacz, Matthew Kluko, Christopher Bailey, and Michael Johnson for their genotyping efforts for undergraduate research credit. Thank you to Valerie Taylor, Mark Carlson, and Dr. James Lee for their generous assistance with this study.

References

1. Office of National Drug Control Policy. Marijuana on public lands, http://www.WhiteHouseDrugPolicy.gov (accessed January 7, 2011).
2. Miller Coyle H, Palmbach T, Juliano N, Ladd C, Lee HC. An overview of DNA methods for the identification and individualization of marijuana. Croat Med J 2003;44(3):315–21.
3. Hsieh HM, Hou RJ, Tsai LC, Wei CS, Liu SW, Huang LH, et al. A highly polymorphic STR locus in *Cannabis sativa*. Forensic Sci Int 2003;131(1):53–8.
4. Niezgoda S. CODIS program overview. Profiles in DNA 1998;1(3):12–3.
5. Bowen R, Schneider J. Forensic databases: paint, shoe prints and beyond. NIJ Journal 2007;258:34–8.
6. Allgeier L, Hemenway J, Shirley N, LaNier T, Miller Coyle H. Field testing of DNA collection cards for *Cannabis sativa* with a single hexanucleotide marker. J Forensic Sci 2011;56(5):1245–9.
7. Miller CH, Shutler G, Abrams S, Hanniman J, Neylon S, Ladd C, et al. A simple DNA extraction method for marijuana samples used in amplified fragment length polymorphism (AFLP) analysis. J Forensic Sci 2003;48(2):343–7.
8. Miller CH. Forensic botany: principles and applications to criminal casework, 1st edn. Boca Raton, FL: CRC Press, 2005.
9. Vinkenoog R, Bushell C, Spielman M, Adams S, Dickinson HG, Scott RJ. Genomic imprinting and endosperm development in flowering plants. Mol Biotechnol 2003;25(2):149–84.

Additional information and reprint requests:
Heather Miller Coyle, Ph.D.
Forensic Science Department
Henry C. Lee College of Criminal Justice & Forensic Sciences
University of New Haven
300 Boston Post Road
West Haven, CT 06516
E-mail: Hcoyle@newhaven.edu