J Forensic Sci, May 2004, Vol. 49, No. 3
Paper ID JFS2003150
Available online at: www.astm.org

Hsing-Mei Hsieh,[1] Ph.D.; Rur-Jyun Hou,[1] M.S.; Ku-Feng Chen,[1] B.S.; Li-Chin Tsai,[1] M.S.; Su-Wen Liu,[1] M.S.; Kuo-Lan Liu,[2] M.S.; Adrian Linacre,[3] Ph.D.; and James Chun-I Lee,[1] Ph.D.

Establishing the rDNA IGS Structure of *Cannabis sativa**

ABSTRACT: The rDNA intergenic spacer (IGS) structure of *Cannabis sativa* was established and can be used for classification and identification of this species. In this study, DNA fragments of rDNA IGS were amplified by PCR from *Cannabis sativa* plant extracts and a 1387 bp fragment was obtained. DNA sequence analysis revealed six different repeat motifs. In the middle of the IGS sequence, there were three sequence motifs, and the same three sections of DNA were then repeated with minor variation in sequence. The terminal region of the IGS was composed of another three different repeat units; multiple copies of these terminal repeat motifs were present in no discernible order. Within six repeat motifs, point variations were observed in five. The DNA sequence of the locus was compared with all the plant sequences registered in GenBank by the Fasta program of GCG software with the result that this DNA fragment was significantly different from any other DNA sequence recorded to date. The most similar sequence was that of Hops (*Humulus lupulus*), but with a similarity of only 88.9% over 579 bp. These specific and complex variations of IGS may be related to the species and geographic distributions.

KEYWORDS: forensic science, *Cannabis sativa*, rDNA, IGS (intergenic spacer)

The analysis of DNA polymorphisms is used widely in botanical studies for the analysis of phylogeny, classification, genetic relationship, species identification, and even plant individualization. The techniques in molecular analysis of plants usually involve whole genome studies such as RFLP (restriction fragment length polymorphism) (1), RAPD (random amplified polymorphic DNA) (2), or AFLP (amplified fragment length polymorphism) (3). Additionally there are allele-specific markers, either coding or non-coding regions, such as rbcL (large subunit of ribulose-1,5-bisphosphate carboxylase) (4), ITS (internal transcribed spacer) (5), IGS (intergenic spacer) (6) and microsatellites (7,8). The rDNA IGS (ribosomal DNA intergenic spacer), separating the 26S and 18S rRNA coding regions, is typically composed of reiterated subrepeats (9) and differences in the number and sequence of these subrepeats are related to most of the length variation between rDNA repeat units within closely related species, among populations, and even between individuals (10).

Species identification and origin determination of the plant *Cannabis sativa* (*C. sativa*) are a major issue in forensic science. Identification of *C. sativa* has long been based on the observation of cystolithic hairs on the leaves (11). The legal requirement in most countries is to confirm the presence of cannabinoids (12–16). DNA polymorphism tests have been used to determine not only the presence of cannabis but also their possible geographical origin (17–23). Among these tests, RAPD can provide an indication of whole genome variation at both species and population level (17), but it has been superseded by tests of specific sequence due to problems with reproducibility. Among the specific markers, the internal transcribed spacer (ITS) I and II of nuclear ribosomal DNA and intergenic spacer region of chloroplast DNA (trnL-trnF IGS) provides species-specific identification and some polymorphic information (18–23). The highly polymorphic STR loci predict the genetic relationship (24,25). Due to the high degree of variability in the structure of rDNA IGS, this locus provides a possible source of variation among related species, within species of the same populations, and even within individual plants (26–30). In order to apply the use of the polymorphic rDNA IGS to the classification of *C. sativa*, the structure of rDNA IGS was established in this study.

Materials and Methods

Sample Preparation and DNA Extraction

A leaf sample was provided by a prosecutor from a drug trafficking case in Taiwan. Approximately 10 mg of leaf tissue was pulverized under liquid nitrogen in a mortar and pestle. The powder was transferred to a 1.5 mL microcentrifuge tube and DNA was extracted by a commercial kit (Plant Genomic DNA Miniprep System, Viogene, Shijr, Taipei, Taiwan). The isolated DNA was quantified with 0.7% agarose gel electrophoresis and also by ultraviolet detection using a spectrophotometer. The sample was stored at −20°C until used.

rDNA IGS Amplification

A forward primer on the 3′ end of 26S and a reverse primer on the 5′ end of 18S of rDNA sequences (22) were designed to amplify the IGS DNA of the plant extracts. The sequences of the

[1] Department of Forensic Science, Central Police University, 56 Shu-Jen Road, Kwei-San, Taoyuan 33334, Taiwan, ROC.
[2] Criminal Investigation Bureau, National Police Administration, Taiwan, ROC.
[3] Forensic Science Unit, Department of Pure & Applied Chemistry, University of Strathclyde, Glasgow G1 1XW, United Kingdom.
* This work was financially supported by NSC90-2420-H-015-001-QC from the National Science Council, ROC, and was presented in poster form at the 2003 American Academy of Forensic Science meeting.
Received 29 April 2003; and in revised form 4 Dec. 2003; accepted 31 Dec. 2003; published 7 April 2004.

forward and reverse primer were 5′-AGGTTAGTTTTACCCTACT-3′ and 5′-GCAGGATCAACCAGGTAGCA-3′, respectively. PCR amplifications were performed in a 50 μL volume containing 0.3 μM each of primer, reaction buffer (10 mM Tris-HCl, pH 8.3, 2.5 mM $MgCl_2$, 50 mM KCl, 0.1% (w/v) gelatin), 250 μM dNTP, 2.5 unit of VioTaq DNA polymerase (Viogene) and 10 ng of genomic DNA. The PCR amplifications were conducted in a 2400 Perkin-Elmer thermal cycler with the following conditions: denaturation at 94°C for 1 min, annealing at 50°C for 1 min and extension at 72°C for 6 min for 40 cycles, and followed by a 30-min extension at 72°C.

PCR Products Cloning and Sequencing

PCR products were separated by agarose gel electrophoresis. A DNA fragment of about 1.5 kb was found and eluted by agarose gel purification kit (Gel Extraction System, Viogene). The purified DNA fragments were ligated to a pSTBlue-1 vector (pSTBlue-1 AccepTor Vector Kit, Novagen, San Diego, CA). The colonies were screened by PCR with T1 (5′-GCAGACGCGTTACGTATCGG-3′) and T2 (5′-CTAGCCTAGGCTCGAGAAGC-3′) primers. T1 and T2 primers were designed according to the sequences of the pSTBlue-1 vector. Sequencing of the PCR products was performed using the T1 primer or T2 primer and the BigDye™ Terminator Kit (ABI PRISM™ BigDye™ Terminator Cycle Sequencing Ready Reaction Kit, Applied Biosystems, Foster, CA). The cycle sequencing products were separated with 5% denatured Long Ranger™ gel (FMC BioProducts, Rockland, ME) and detected by an ABI 373A DNA sequencer.

Sequence Confirmation

The sequencing results were confirmed by re-performing the PCR and sequencing of two partially overlapping fragments. The primers were designed according to the above sequencing results. The sequences of two primer sets were 5′-AGGTTAGTTTTACCCTACT-3′/5′-AAACTGGGTACATGGACACC-3′ and 5′-TCCATCAACA-CAAGTTGCGG-3′/5′-GTTATCCGAGTAGCAGATACC-3′. The conditions of PCR amplifications were similar to the previous descriptions. A single primer of each primer set was used for sequencing.

Sequence Analysis

The DNA sequence was compared with all the plant sequences registered in the GenBank by Fasta program of GCG software (Wisconsin Package Version 10.3, Accelrys Inc., San Diego, CA). The repeat sequence was searched manually.

Results

One DNA fragment (1387 bp, Fig. 1) containing the PCR products of rDNA IGS of *C. sativa* was cloned and sequenced.

Repeat motifs were observed within the sequence of the IGS locus. Six different repeat motifs were identified within the DNA sequence of IGS clone. These repeat units were assigned with a letter from A to F, with the sequence and repeat structures shown in Table 1. Within these repeat motifs, point variations were observed in five of them. The repeat units were named with serial numbers. The positions and types of the repeat motifs are presented in Fig. 2. In the middle of the IGS sequence there were three sequence motifs, one being 25 bases in length (A repeat unit), another 33 bases in length (B repeat unit), and a third 29 bases in length (C repeat unit). The same three sections of DNA were then repeated once with minor variation in the DNA sequence. The terminal region of the IGS is composed of another three different repeat units, 17 bases in length (D repeat unit), 22 bases in length (E repeat unit), and 27 bases in length (F repeat unit). Multiple copies, which were 7, 5 and 3 copies of D, E and F repeat unit, respectively, were present but in no discernible order. The pattern of repeat units in the middle part of IGS was AB1C1 and then AB2C2. The repeat units in the terminal part of IGS showed no discernible pattern of distribution, rather repeat motifs D, E and F were present in a random order (D1, E1D2, F1E2D1, E2D1F2E3D3, and then D1F3E1D4).

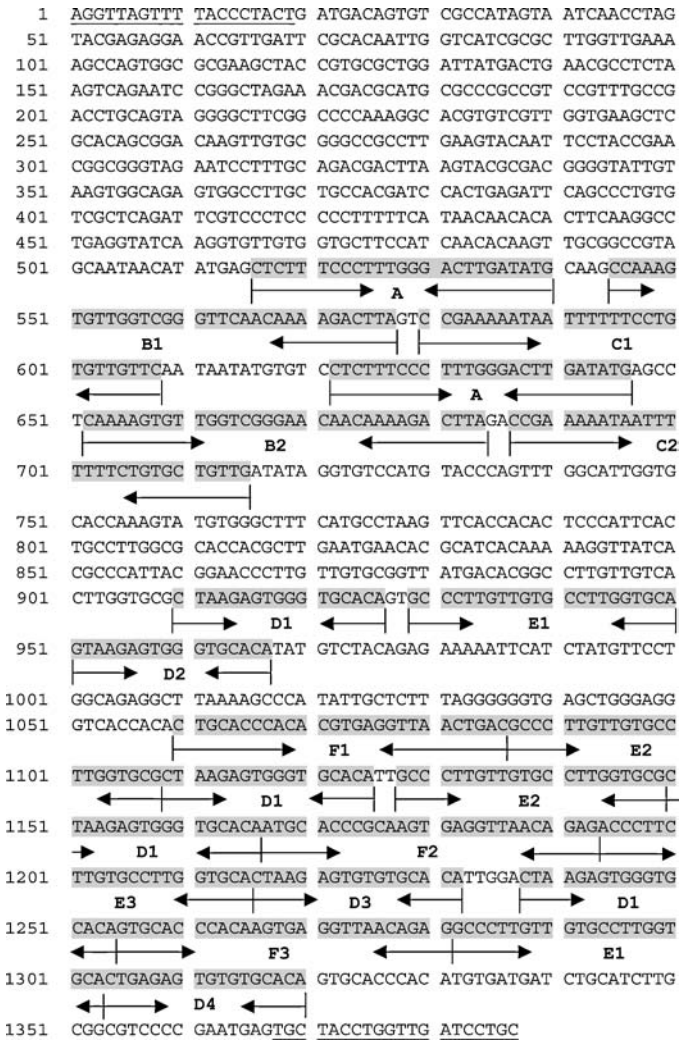

FIG. 1—*The rDNA IGS of Cannabis sativa. The primer sequences are underlined. The possible repeat sequences are shadowed and labeled under the sequences.*

The DNA sequence of 1387 bp was compared with all the plant sequences registered in GenBank by Fasta program of GCG software. The results showed that the most similar sequence was *Humulus lupulus*, but the similarity was only 88.9% in 579 bp. This similarity existed mostly in the 5′-flanking region (3′ end of 26S rDNA). By comparing with the sequence of *Humulus lupulus* (accession number AF223066), the result showed that the total length of Cannabis IGS was 984 bp, from position 387 to 1370 in Fig. 1. Position 1 to 386 and 1371 to 1387 represented the sequences of 26S rDNA and 18S rDNA, respectively. Until now, the lengths

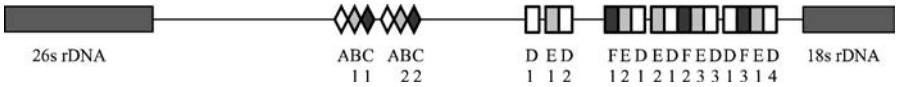

FIG. 2—*Possible repeat sequence structure of rDNA IGS of Cannabis sativa.*

TABLE 1—*Sequences of the repeat unit identified in rDNA IGS of Cannabis sativa. (Note: The highlights point out positions of sequence variations.)*

Name of repeat unit	Length	Sequences
A	25bp	CTCTTTCCCTTTGGGACTTGATATG
B1	33bp	CCAAAGTGTTGGTCGGGTTCAACAAAAGACTTA
B2	33bp	CAAAAGTGTTGGTCGGGAACAACAAAAGACTTA
C1	29bp	CCGAAAAATAATTTTTTCCTGTGTTGTTC
C2	29bp	CCGAAAAATAATTTTTTTCTGTGCTGTTG
D1	17bp	CTAAGAGTGCGTGCACA
D2	17bp	GTAAGAGTGCGTGCACA
D3	17bp	CTAAGAGTGTGTGCACA
D4	17bp	CTCAGAGTGTGTGCACA
E1	22bp	GCCCTTGTTGTGCCTTGGTGCA
E2	22bp	GCCCTTGTTGTGCCTTGGTGCG
E3	22bp	ACCCTTCTTGTGCCTTGGTGCA
F1	27bp	CTGCACCCACACGTGAGGTTAACTGAC
F2	27bp	ATGCACCCGCAAGTGAGGTTAACAGAG
F3	27bp	GTGCACCCACAAGTGAGGTTAACAGAG

of reported botanic rDNA IGS are from 1 to 12 kb; therefore, the *C. sativa* IGS is relatively short in length.

Discussion

The repeat sequence structure, shown in Fig. 2, was similar to the reported structure of other plant rDNA IGS (26–30). While the structure of the IGS in *C. sativa* showed some similarities to other plants, the sequence of this locus was greatly different from that of any other plant.

The typical transcription initiation site (TIS), TATA(G)TA(N) GGGG, could not be determined by comparison with other reported sequences (29). Therefore, it will be necessary to determine the exact TIS by S1 nuclease mapping. Another possibility is that this is just a non-functional pseudogene, so the TIS was not observed.

These specific and complex variations of IGS may be related to the species and geographic distributions. The sequence structure of rDNA IGS of *C. sativa* provides a potential means of classification and identification among different cultivars and even within individual plants, making it a valuable locus in the establishment of the population data of different accessions or origins of cannabis.

References

1. Atchison BA, Whitefield PR, Bottomley W. Comparison of chloroplast DNAs by specific fragmentation with EcoRI endonuclease. Mol Gen Genet 1976;148:263–9.
2. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Aci ds Res 1990;18:6531–5. [PubMed]
3. Vos P, Hogers R, Bleeker M, Reijans M, Lee TVD, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 1995;23:4407–14. [PubMed]
4. Clegg MT. Chloroplast gene sequences and the study of plant evolution. Proc Natl Acad Sci USA 1993;90:363–7. [PubMed]
5. Baldwin BG. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Mol Phylogenet Evol 1992;1:3–16. [PubMed]
6. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 1991;17:1105–9. [PubMed]
7. Gupta PK, Balyan HS, Sharma PC, Ramesh B. Microsatellites in plants: A new class of molecular markers. Curr Sci 1996;70:45–54.
8. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 1989;17:6463– 71. [PubMed]
9. Hemleben V, Ganal M, Gerstner J, Scheibd K, Torres R. Organization and length heterogeneity of plant ribosomal RNA genes. In: Kahl G, editor. Architecture of eukaryotic genes. VHC: Weinheim, 1987;371–83.
10. Rogers SO, Bendich AJ. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Mol Biol 1987;9: 509–20.
11. Mitosinka GT, Thornton JI, Hayes TL. The examination of cystolithic hairs of Cannabis and other plants by means of the scanning electron microscope. J Forensic Sci Soc 1972;12:521–9. [PubMed]
12. Mikes F, Hofmann A, Waser PG. Identification of (-)-delta 9-6a, 10a-trans-tetrahydrocannabinol and two of its metabolites in rats by use of combination gas chromatography-mass spectrometry and mass fragmentography. Biochem Pharmacol 1971;20:2469–76. [PubMed]
13. Vree TB, Breimer DD, van Ginneken CA, van Rossum JM, de Zeeuw RA, Witte AH. Identification of cannabivarins in hashish by a new method of combined gas chromatography-mass spectrometry. Clin Chim Acta 1971;34:365–72. [PubMed]
14. Vree TB, Breimer DD, van Ginneken CA, van Rossum JM. Identification in hashish of tetrahydrocannabinol, cannabidiol and cannabinol analogues with a methyl side-chain. J Pharm Pharmacol 1972;24:7– 12. [PubMed]
15. Thornton JI, Nakamura GR. The identification of marijuana. J Forensic Sci Soc 1972;12:461–519. [PubMed]
16. Smith RN. High pressure liquid chromatography on cannabis, identification of separated constituents. J Chromatogr 1975;115:101–6. [PubMed]
17. Gillan R, Cole MD, Linacre A, Thorpe JW, Watson ND. Comparison of *Cannabis sativa* by random amplification of polymorphic DNA (RAPD) and HPLC of cannabinoids: a preliminary study. Sci Justice 1995;35:169–77. [PubMed]
18. Siniscalco Gigliano G, Caputo P, Cozzolino S. Ribosomal DNA analysis as a tool for the identification of *Cannabis sativa* L. specimens of forensic interest. Sci Justice 1997;37:171–4. [PubMed]
19. Linacre A, Thorpe J. Detection and identification of cannabis by DNA. Forensic Sci Int 1998;91:71–6. [PubMed]
20. Siniscalco Gigliano G. Preliminary data on the usefulness of internal transcribed spacer I (ITS1) sequence in *Cannabis sativa* L. identification. J Forensic Sci 1999;44:475–7. [PubMed]
21. Siniscalco Gigliano G. Identification of *Cannabis sativa* L. (Cannabaceae) using restriction profiles of the Internal Transcribed Spacer II (ITS2). Sci Justice 1998;38: 225–30. [PubMed]
22. Wilkinson M, Linacre AM. The detection and persistence of *Cannabis sativa* DNA on skin. Sci Justice 2000;40:11–4. [PubMed]

23. Kohjyouma M, Lee IJ, Iida O, Kurihara K, Yamada K, Makino Y, et al. Intraspecific variation in *Cannabis sativa* L. based on intergenic spacer region of chloroplast DNA. Biol Pharm Bull 2000;23:727–30. [PubMed]
24. 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;131:53–8. [PubMed]
25. Gilmore S, Peakall R, Robertson J. Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations. Forensic Sci Int 2003;131:65–74. [PubMed]
26. Baldwin BG, Markos S. Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of Calycadenia (Compositae). Mol Phylogenet Evol 1998;10:449–63. [PubMed]
27. Nickrent DL, Patrick JA. The nuclear ribosomal DNA intergenic spacers of wild and cultivated soybean have low variation and cryptic subrepeats. Genome 1998;41:183–92. [PubMed]
28. Chou CH, Chiang YC, Chiang TY. Within- and between-individual length heterogeneity of the rDNA-IGS in *Miscanthus sinensis* var. glaber (Poaceae): phylogenetic analyses. Genome 1999;42:1088–93. [PubMed]
29. Fernandez M, Polanco C, Ruiz ML, Perez de la Vega M. A comparative study of the structure of the rDNA intergenic spacer of *Lens culinaris* Medik., and other legume species. Genome 2000;43:597–603. [PubMed]
30. Gupta PK, Sharma PK, Balyan HS, Roy JK, Sharma S, Beharav A, et al. Polymorphism at rDNA loci in barley and its relation with climatic variables. Theor Appl Genet 2002;104:473–81. [PubMed]

Additional information and reprint requests:
James Chun-I Lee, Ph.D.
Department of Forensic Science
Central Police University
56 Shu-Jen Road, Kwei-San
Taoyuan 33334
Taiwan ROC
Fax:+886-3-3275907
E-mail: jimlee@sun4.cpu.edu.tw