

1-9-2003

Development of microsatellite markers in Cannabis Sativa for fingerprinting and genetic relatedness analyses

Hussain J. Al-Ghanim
Florida International University

DOI: 10.25148/etd.FI13101576

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Other Chemistry Commons](#)

Recommended Citation

Al-Ghanim, Hussain J., "Development of microsatellite markers in Cannabis Sativa for fingerprinting and genetic relatedness analyses" (2003). *FIU Electronic Theses and Dissertations*. 1091.
<https://digitalcommons.fiu.edu/etd/1091>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

DEVELOPMENT OF MICROSATELLITE MARKERS IN *CANNABIS*
SATIVA FOR FINGERPRINTING AND GENETIC RELATEDNESS
ANALYSES

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

FORENSIC SCIENCE

by

Hussain J. Al-Ghanim

2003

QD0 A444
2003

To: Dean Arthur Herriott
College of Arts and Sciences

This thesis, written by Hussain J. Al-Ghanim, and entitled Development of Microsatellite Markers in *Cannabis sativa* for Fingerprinting and Genetic Relatedness Analyses, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Alan Meerow



Raymond Schnell



Kenneth Furton

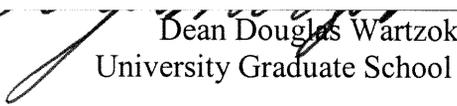


Jose Almirall, Major Professor

Date of Defense: January 9, 2003

The thesis of Hussain J. Al-Ghanim is approved.

Dean Arthur Herriott
College of Arts and Sciences



Dean Douglas Wartzok
University Graduate School

Florida International University, 2003

DEDICATION
TO MY PARENTS

إلى والدي

ACKNOWLEDGMENTS

In the name of Allah, the Beneficent, the Merciful

All praise is due to Allah

I would like to express my great appreciation to my major advisor Dr. Jose Almirall for his guidance and patience during my graduate study and for the opportunity to work on this interesting project. I am also grateful to my committee members, first and foremost Dr. Alan Meerow whose dedication is extremely admirable, Dr. Raymond Schnell for his valuable help, and Dr. Kenneth Furton for his recommendations.

A round of applause goes to Jason Clayton for helping me build good lab techniques. I also want to thank Wilber Quintanilla and all the members of the genetics lab at the United States Department of Agriculture-Agricultural Research Service-Subtropical Horticulture Research Station in Miami, Florida (USDA-ARS-SHRS) for their instructive training. All of you were very helpful and your assistance is so much appreciated and never forgotten.

The most heartfelt thanks and gratitude go to my parents. Without their patience, understanding, and love, the completion of this work would not have been possible. I thank my two brothers and sister for their encouragement and inspiration. To my grandmother who always prays for me, Allah bless you. To all the great people who also have me in their prayers, thank you forever.

Special thanks go to my sponsor, Dubai Police at the United Arab Emirates, for funding this research and for their infinite support throughout my undergraduate and graduate years.

I would like to thank Dr. Heather Coyle at the Connecticut Forensic Science Laboratory for her generous help to provide us with the DNA samples and Dr. Gary Shutler for his assistance in the project.

ABSTRACT OF THE THESIS

DEVELOPMENT OF MICROSATELLITE MARKERS IN *CANNABIS SATIVA* FOR FINGERPRINTING AND GENETIC RELATEDNESS ANALYSES

by

Hussain J. Al-Ghanim

Florida International University, 2003

Miami, Florida

Professor Jose Almirall, Major Professor

Microsatellite markers were developed for *Cannabis sativa* L. (marijuana) to estimate the level of polymorphism, usefulness for DNA typing (genotype identification), and to measure the genetic relationships between the different plants. Twelve different oligonucleotide probes were used to screen an enriched microsatellite library of *Cannabis sativa* in which 49% of the clones contained microsatellite sequences. Characterization of microsatellite loci in *Cannabis* revealed that GA/CT was the most abundant class of isolated microsatellites representing 50% overall. Eleven polymorphic SSR markers were developed, derived from dinucleotide motifs and eight from trinucleotide motifs. A total of 52 alleles were detected averaging 4.7 alleles/locus. The expected heterozygosity of the eleven loci ranged between 0.368 and 0.710 and the common probability of identical genotypes was 1.8×10^{-7} . The loci identified 27 unique profiles of the 41 *Cannabis* samples. The eleven microsatellite markers developed in this study were found to be useful for DNA fingerprinting and for assessing genetic relationships in *Cannabis*.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION	1
The history of <i>Cannabis</i>	1
<i>Cannabis</i> and the Law	2
<i>Cannabis</i> taxonomy	2
<i>Cannabis</i> botany	5
<i>Cannabis</i> as a drug of abuse	6
Forensic identification of marijuana	8
DNA era	9
Restriction fragment length polymorphism	10
PCR era	11
Random amplified polymorphic DNA.....	12
Amplified fragment length polymorphism	13
Microsatellite	14
II. MATERIALS AND METHODS.....	19
DNA isolation and preparation of genomic DNA	19
Preparation of the linkers	20
DNA/linker ligation	21
Pre-amplification for microsatellite enrichment	22
Preparation for enrichment of microsatellite	23
Enrichment of microsatellite sequence	24
Dynabeads/hybridization washes.....	25
Pre-cloning amplification.....	26
Fragment size fractionation.....	27
Cloning reaction.....	28
Plasmid recovery.....	29
M13 insert amplification.....	29
Exonuclease I treatment.....	30
Cycle sequencing	31
Primer design and fragment analysis	32
Determination of allele size	34
Statistical analyses	34
III. RESULTS	36
Characterization of isolated microsatellite sequences	36
Characterizations of the selected microsatellite sequences	37
Polymorphisms of microsatellite loci	42
Genetic identifications and relationships.....	55

IV.	DISCUSSION	66
	<i>Cannabis</i> genotyping	66
	Microsatellites in <i>Cannabis</i>	76
	Microsatellite polymorphism and applications	77
V.	CONCLUSION.....	82
	REFERENCES	84
	APPENDIX.....	92

LIST OF TABLES

TABLE	PAGE
1. Microsatellite enrichment success for <i>Cannabis sativa</i> and characterization of the microsatellite types	36
2. The 11 SSR markers and the primer sequences.....	40
3. List of information and association known between the 41 <i>C. sativa</i> plants.....	41
4. Calculation of different genetic parameters for <i>Cannabis</i> samples excluding the duplicates	43
5. Genotyping using P4 primer set, allele count, and allele frequencies	44
6. Genotyping using P6 primer set, allele count, and allele frequencies	45
7. Genotyping using P7 primer set, allele count, and allele frequencies	46
8. Genotyping using P9 primer set, allele count, and allele frequencies	47
9. Genotyping using P13 primer set, allele count, and allele frequencies	48
10. Genotyping using P14 primer set, allele count, and allele frequencies	49
11. Genotyping using P15 primer set, allele count, and allele frequencies	50
12. Genotyping using P17 primer set, allele count, and allele frequencies	51
13. Genotyping using P19 primer set, allele count, and allele frequencies	52
14. Genotyping using P24 primer set, allele count, and allele frequencies	53
15. Genotyping using P25 primer set, allele count, and allele frequencies	54

LIST OF FIGURES

FIGURE	PAGE
1. The chemical structure of the four major chemical substances found in <i>Cannabis</i>	7
2. Cycle sequencing of the locus B01-CANN1 encoding for the (GAA) ₁₃ (A)(GAA) ₃	38
3. Cycle sequencing of the locus E07-CANN1 encoding for the (CTA) ₉	39
4. Cycle sequencing of the locus B02-CANN2 encoding for the (AAG) ₁₀	39
5. PCA of the 27 unique <i>Cannabis</i> genotypes	64
6. PCA of 26 different <i>Cannabis</i> genotypes (removing outlier G30).....	65
7. Homozygous profiles using P17	66
8. Heterozygous profiles using P25	67
9. Genotyping of <i>C. sativa</i> samples using P13 and the stutter peaks.....	69
10. Mixed preferential amplification of plus A and true allele in P24 locus	71
11. Genotyping using P14 and scoring by the true peaks	72
12. Genotyping using P19 and scoring by the true peaks	73
13. Genotyping using P24 and scoring by the plus A peaks.....	74
14. Genotyping using P7 and the special case	75
15. P17 and reproducibility between using the M13 fluorescent tail primer and the direct fluorescent primer	78
16. P6 and reproducibility between the results	79

LIST OF TREES

TREE	PAGE
1. NJ tree based on Chord's genetic distance across the 41 <i>Cannabis</i> plants.....	56
2. UPGMA tree based on Chord's genetic distance across the 41 <i>Cannabis</i>	57
3. NJ tree based on Chord's genetic distance across 27 different genotypes.	58
4. UPGMA tree based on Chord's genetic distance across 27 different genotypes	59
5. UPGMA tree based on Nei-unbiased genetic distance across 27 different genotypes.	60
6. NJ tree based on modified Roger genetic distance across 27 different genotypes	61
7. NJ plus bootstrap values based on DNA shared Alleles (DSA) genetic distance	62
8. NJ plus bootstrap values based Chord's genetic distance.....	63

Introduction

The history of *Cannabis*:

Cannabis sativa L. (marijuana) is one of the oldest cultivated plants. Depending on the person's point of view, *Cannabis* is useful, a reagent euphoria, or evil [1]. "To the agriculturist, *Cannabis* is a fibre crop; to the physician of a century ago it was a valuable medicine; to the physician of today it is an enigma; to the user a euphoriant; to the police, a menace; to the traffickers, a source of profitable danger; to the convict or parolee and his family a source of sorrow" [2].

Throughout history, *Cannabis* has been used as a source of hempen fiber for rope and fabric from its stems, food and oil from its seeds, and a psychoactive drug from its flowers and leaves. The oldest documented history of *Cannabis* dates to 10,000 B.C. in China. It is believed to be indigenous to Asia and to parts of Iran, southern Siberia, the Kirghiz Desert, India, and the Himalayas [3]. *Cannabis* has been known in China since Neolithic times 6,000 BP. Because almost every part of the plant is capable of being used, people of China use the extracted hemp fibers that are long and durable to make variety of items including ropes, clothes, papers, and fishing nets. In addition, the seeds were used as an important "grain" whereas leaves and roots were recommended for various medical illnesses such as gout, rheumatic pains, and constipation [3, 4].

Cannabis cultivation eventually made its way to India, Europe, and Africa. The Aryans introduced *Cannabis* into the Indian culture 4,000 years ago, and Scythians brought *Cannabis* to Europe 2,800 years ago [5]. By the 12th century, *Cannabis* was being utilized in the Islamic paper mills, but not until the 15th century in the English ones.

In the 1500's, Spanish troops brought *Cannabis* to South America, whereas the English brought *Cannabis* to North America for rope production around 1600's.

***Cannabis* and the Law:**

By 1860, the number of people taking marijuana for recreational use had sharply increased such that it replaced alcohol as a popular form of intoxication [6]. In 1870, *Cannabis* was listed in the United States Pharmacopoeia as a source of medicine for a variety of illnesses. In 1915, California was the first state to outlaw *Cannabis*, and many other states followed soon after. After that, the price driven by the Marijuana Tax Act of 1937 made it unrealistic to use marijuana as a medicine. By 1941, *Cannabis* was removed from the United States Pharmacopoeia. Congress passed the Control Substance Act of 1970 to classify psychoactive drugs into five schedules based on their characteristics. In United States, marijuana is listed in schedule I which is the most restrictive. Schedule I drugs present a high potential for abuse and no medical usage [1]. Legally, marijuana is all parts of any plant of the genus *Cannabis* including seeds, resin (hashish), and any derivative compound. In U.S. and Canada, it is a criminal offense for someone to possess *Cannabis* plants (marijuana), *Cannabis* resin (hashish), or any cannabinoid constituents.

***Cannabis* taxonomy:**

Cannabis has been introduced to lands and climates that are far and different from its natural habitat which forces it to adapt to the new environment and consequently develop certain characteristics that vary per location. Human intervention has contributed to the rise of various races and strains of the plant by selecting the plant for different purposes. For example, people grow *Cannabis* for longer and stronger fiber,

high oil content, or greater narcotic effect [7]. *Cannabis*, marijuana, and hemp are terms that are used interchangeably. However, *Cannabis* is the botanical genus of the plant and marijuana describes *Cannabis* plants that contain high Δ^9 tetrahydrocannabinol (THC) content and are used for their psychoactive potency [1]. Hemp is used to describe *Cannabis* plants that have low THC content and are cultivated for fiber. Therefore, there are two distinctive strains; one is generally cultivated for fiber (hemp) and the other for drug use (marijuana) [8].

There had been immense debates over how many species *Cannabis* plants contain. Many different taxonomic names and descriptions for *Cannabis* have been reported according the International Code of Botanical Nomenclature. Some of these specific names reported by Schultes (1974) [9] are: *C. americana* Houghton et Hamilton; *C. erratica* Sievers; *C. foetens* Gilibert; *C. generalis* H.L Krause; *C. gigantea* Crevost; *C. indica* Lamarck; *C. lupulus* Scopoli; *C. pedemontana* Camp; *C. ruderalis* Janischewsky, and *C. sativa* Linnaeus.

Historically, botanists recognized generally three major species of *Cannabis*: *C. sativa*, *C. indica*, and *C. ruderalis*. These three species had been carefully described and often distinguished by different growth habitats, characters of the seeds, and the major difference in the structures of the wood and fruit [7, 10]. The name *Cannabis sativa* is initially reported by Bauhin in 1623, but it is first published by Carlous Linnaeus in the seminal text *Species Plantarum* of 1753 which is considered the internationally accepted starting point for botanical nomenclature [11]. Linnaeus originally gave India as the country of origin for *Cannabis sativa* L. (L stands for Linnaeus), but later he realized that *C. sativa* was indigenous of Asia. The old polytypic view of *Cannabis* evolved in 1783

when Lamarck described *C. indica* to be very distinct from *C. sativa* [12]. Lamarck gave Asia as a source of origin for *C. indica*. In 1924, a Russian botanist, Janischewsky described a new species called *C. ruderalis*. He mentioned that *C. ruderalis* is known to inbreed in Volga region, western Siberia, central Asia, and Russia.

On the other hand, Small and Cronquist strongly support the concept that the variations within the genus *Cannabis* are due to the cultivation and selection of man [13]. They surveyed plants that represented a diverse population of *Cannabis* and concluded that there is no difference between wild and cultivated strains. They also conducted numerous studies pertaining to the cytology and breeding behavior, chemical variations, and morphology of seeds. They concluded that there is inevitably one species of *Cannabis* [13]. Hemp specialist Dewey also observes that when the different strains of *Cannabis* are cultivated together under new conditions, they do not remain distinct, thus cannot regard them as individual species [14]. Today, most botanists strongly support the view that *Cannabis* is a genus with a single highly variable species (*C. sativa*), which has diversified into a wide variety of ecotypes and cultivated races [15]. In addition, most state and federal laws acknowledge *Cannabis* as a single species. Currently, *Cannabis* is listed as belonging to [16]:

Kingdom – Plantae – Plant

Subkingdom – Embryophyta – Seed bearing

Phylum – Tracheophyta – Vascular plant

Subphylum – Pterosida – Fern like

Class – Angiospermae – Flowering plants

Subclass – Dicotyledonae – Two seed leaves, mulberry, Elm Trees

Order – Urticales – Hops and Hemp

Family – Cannabinaceae – Hops and Hemp

Genus – *Cannabis*

Species – *sativa*

***Cannabis* botany:**

Cannabis is a tall erect, dioecious plant, which means that the male and female flowers develop on separate plants. *Cannabis* can grow to a height of 6 meters (20 feet) in about 5-6 month [17] with adequate sunlight and ample irrigation. *Cannabis* is a diploid organism with $2n = 20$ chromosomes, two of which are sex chromosomes. Female plants have two X chromosomes while male plants have one X chromosome and a Y chromosome [18]. The female plant tends to be smaller and fuller in structure when compared to the male plant. Females also have more leaves located at the top of the plant with many leaves surrounding the flowers. The calyx is covered with resin exuding glandular trichomes (hairs). The clusters of 10 to 100 individual flowers are considered the “bud” of commercial marijuana. Many growers prefer to produce seedless buds (sinsemilla) by removing male plants before they mature and release pollen. Sinsemilla plants have a reputation of producing high potency marijuana since metabolites are invested in the production of resin rather than seeds. As the calyces swell, the glandular trichomes on the surface develop and secrete aromatic tetrahydrocannabinol (THC) laden resin. THC serves as a barrier to water loss in *Cannabis*. This is because THC is viscous hydrophobic oil that eludes on the surface of the plant and supplies a waxy coating to the leaves [19]. The three major forms of abuse are: marijuana which is prepared from the

dried flowering tops and leaves, hashish which consists of dried resin and compressed flowers, and hash oil which is prepared by extraction from hashish using distillation [1].

***Cannabis* as a drug of abuse:**

Cannabis is the only genus known to produce chemical substances called cannabinoids. These substances are responsible for the psychoactive ingredients of marijuana. The cannabinoids belong to a class of chemicals called terpenoids (terpene) [15]. These compounds are essential oils within the plant. They cannot produce the psychoactive action in the acid form in which they are found naturally in the fresh plant material. In order to generate the psychoactive potential of marijuana, the plant material must be heated to remove the carboxyl group. For this reason, and because cannabinoids (non-polar molecules) have low solubility in water, they are normally administered by smoking. According to a study carried out by the British Medical Association [20], there are more than 60 different cannabinoids and over 400 active components identified in different *Cannabis* plants. Cannabinol (CBN), cannabidiol (CBD), Δ^8 tetrahydrocannabinol (Δ^8 THC), and Δ^9 tetrahydrocannabinol (Δ^9 THC) have been identified as the major constituents (figure 1). The psychoactive effect is primarily associated with the concentration of Δ^9 THC while the others have no known pharmacological activity [8].

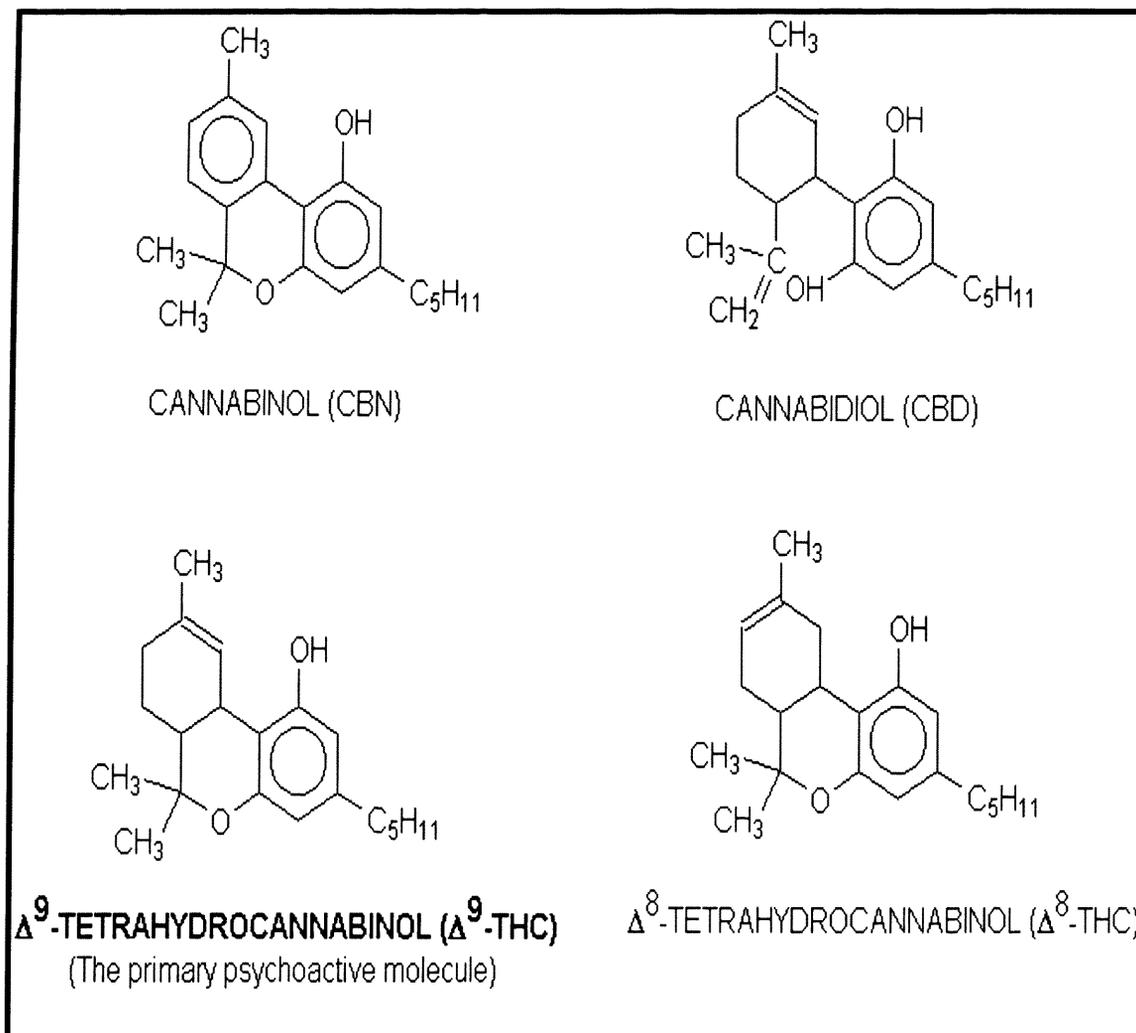


Figure (1): The chemical structure of the four major chemical substances found in *Cannabis*

Vollner reported that the fiber type of *Cannabis* is characterized by high CBD to THC ratio (>5) whereas the drug type has a much lower ratio (<0.2) [21]. The variation in the ratio of the cannabinoids has been also correlated with several factors such as environmental conditions, the stage of the plants' development, the time of the sample collection [22]), soil nutrients [23], and the intensity of light [24]. However, the results of most significant studies on *Cannabis* constituents agree that the ratios of cannabinoids

(THC, CBN, and CBN) and/or their absence in some sample are mainly under genetic control [25].

Forensic identification of marijuana:

In forensic investigations, the plant-like materials that are confiscated by the police are usually accompanied by the question: “Is this plant material *Cannabis*?” [15]. The common procedures utilized by forensic science laboratories for *Cannabis* identification are based on the study of morphological (botanical) features of the material and on the presence of certain chemical compounds. The forensic experts consider that the tests commonly employed can identify *Cannabis* beyond reasonable doubt. The botanical identification of *Cannabis* is determined microscopically. The analysis of the leafy material by the microscope is based on the observation of certain morphological features. The presence of characteristic glandular hairs, cystolithic hairs, and non-glandular hairs are strongly diagnostic for *Cannabis* [15]. The upper surface of the *Cannabis* leaves has short hairs called cystoliths that often contain calcium carbonate crystals. In addition, when the plant is about to flower, the upper surface of the leaves has glandular hairs that contain resin [26]. When all the above criteria are met, the microscopic analysis can be conclusive for the identification of *Cannabis*. However, sometimes the drying process, the age of plants, and the harvesting and preserving conditions of the material can significantly affect the botanical features of *Cannabis*. Moreover, Nakamura reported more than 80 different plant species contain cystolithic hairs resembling those found in *Cannabis* [27]. Therefore, forensic experts now rely on both microscopic and chemical techniques to conclusively distinguish *Cannabis*.

The chemical identification of *Cannabis* is carried out by using several different techniques such as the Duquenois test, thin layer chromatography (TLC), gas chromatography (GC), high performance liquid chromatography (HPLC), and/or mass spectroscopy (MS). The modified Duquenois test [28] is one of the best color tests used in drug identification of marijuana. This test can indicate that the unknown material is marijuana when a final purple color reaction is observed. TLC, GC, and HPLC are the commonly used chromatographic techniques to detect the major *Cannabis* constituents (THC, CBD, and CBN). MS is also an excellent and widely used method to identify marijuana [29].

In recent years, DNA technology has been given close attention in the analysis of marijuana. The DNA sequencing analysis have been employed successfully in plant genetic studies and recently employed as new DNA method in forensic investigation for *Cannabis* identification. The sequence analysis of the nuclear ribosomal DNA of two internal transcribed spacer (ITS1 and ITS2) as well as the sequences of the *trnL* intron of the intergenic spacer between the *trnL* 3' exon and *trnF* gene of chloroplast DNA have been the most useful for the identification of *Cannabis* [15].

DNA era:

One of the central missions for many important braches of biological science is the analysis of genetic diversity and relatedness between and within different species, populations, and individuals. The common techniques of evaluating the genetic variability such as comparative anatomy, morphology, embryology, and physiology have been strongly complemented by the different molecular markers. The development of molecular markers are based on polymorphisms found in proteins or DNA, a source of

data which has been used by research laboratories in a variety of fields such as taxonomy, phylogeny, ecology, genetics, forensics, and plant breeding [30].

For many years, protein genetic markers (allozymes) have been the molecular markers of choice. Since the 1960s, allozyme electrophoresis has been employed in many organisms from bacteria to various animal and plant species. These markers were used in numerous fields such as biochemistry, genetics and breeding for different purposes such as population structure, mating system, and systematics [31]. However, in recent years, this trend has shifted dramatically towards the DNA molecule as a powerful source of polymorphisms. The unique portions of an individual's DNA sequence has made it possible to study the genetic diversity and relatedness between organisms. A wide variety of techniques to visualize DNA polymorphisms have been developed and molecular markers have been derived from those techniques.

Restriction fragment length polymorphism:

David Botstein and his coworkers were the first to exploit the small variations found between people at the genetic level to be used for human gene map construction [32]. This type of variation they used is called restriction fragment length polymorphism (RFLP). In 1985, while searching for disease markers in DNA, Alec Jeffreys discovered a unique application for the RFLP to the science of human identification. "DNA fingerprinting" is a term first described by Jeffreys for a method that allows for fast detection of many highly variable DNA loci based on variations in the length of the repetitive DNA sequence clusters [33, 34]. The DNA fingerprinting method detects polymorphisms by hybridization of specific multilocus "probes" to restriction fragments separated by electrophoresis. The result is a multi-banded pattern whose complexity

suggests a fingerprint unique to each individual. This leads to the first forensic use of DNA in England that made use of DNA fingerprinting. The police investigation together with the DNA fingerprinting was able to identify Colin Pitchfork as murderer of two young girls in the English Midlands. Consequently, an innocent suspect was freed for a crime that he did not commit based on DNA evidence [35].

The most useful forms of RFLP markers arise from variation in the number of tandemly repeated DNA sequence occurring between two restriction sites. These sequences are called variable number of tandem repeats (VNTR) or minisatellite [33, 34]. This type of DNA fragment consists usually of a 10 to 60 base motif that is found within and between genes. Clusters of such sequences are widely dispersed in the genomic DNA. The number of the repeats at a given locus is variable, each of these variations make a VNTR allele [35]. Many of these loci have dozens of alleles thus heterozygosity is common, resulting in a high degree of polymorphisms among these loci. The standard DNA testing employed in forensics uses four to ten different VNTR probes to discriminate between individuals. Despite its informative input in numerous fields and applications, the RFLP technique has limitations that restrict its wider acceptance. The RFLP pattern requires fairly high molecular weight DNA (50-200 micrograms of DNA per individual) to generate a DNA fingerprint pattern of a particular organism [35]. Moreover, RFLP profiles take a longer time to develop and require better quality DNA.

PCR era:

In 1985, a new technique called the polymerase chain reaction (PCR) [36, 37] invented by Kary Mullis has changed the face of molecular biology forever and started a new era of PCR-based fingerprinting techniques. PCR is able to amplify any DNA

sequence of interest to high copy numbers and hence the name molecular “Xeroxing” [35]. To amplify a particular DNA sequence, two single-stranded oligonucleotide primers are designed to complement a specific sequence on the template DNA. Upon the introduction of PCR, a wide variety of PCR-based techniques were developed to detect polymorphism. A thorough review of these techniques can be found in Innis et al. (1990) [38]. All the PCR-based techniques require only small amounts of the genomic DNA (1/10 to 1/100 the amount needed of RFLP) [35], and the results can be obtained in a short time. Applying such techniques dominate the field of molecular biology to generate different molecular markers for various applications.

Random amplified polymorphic DNA:

Based on PCR technology, Williams introduced a method that employs the use of one or two GC-rich PCR primers of arbitrary ten base oligonucleotides to generate some molecular markers [39]. This technique that is termed random amplified polymorphic DNA (RAPD) and the RAPD markers are easily developed based on PCR amplification of random loci in the genome of a species. Because the 10 base primer(s) sequence is arbitrary, no prior knowledge of the genomic DNA sequence composition is needed; therefore the primers can be used universally for the DNA analysis of eukaryotes and prokaryotes [30]. These primers have the possibility of annealing at a number of different loci in the genomic DNA. For amplification products to occur, the primer(s) have to bind at two different sites on opposite strands of the DNA template and within an amplifiable distance of each other. Since multiple fragments are produced by a single amplification, the polymorphisms are detected as presence or absence of each fragment in the pattern.

Primers prefer to bind to those target sites with a high degree of homology. Thus, the number of PCR products is directly related to the number and orientation of the sequences that are complementary to the primers in the genome. Polymorphisms in the RAPD markers are attributed to several types of events. The first type is the insertion of a large piece of DNA between the two annealing sites. The deletion of a DNA fragment holding primer-annealing sites can also result in fragment variations. In addition, some nucleotide substitution can affect the affinity of the primers to the annealing sites [30]. However, because these RAPD fragments are scored as dominant markers, the heterozygous state cannot be distinguished from the homozygous state. RAPD markers cannot be used to estimate the allele frequencies for population genetics studies nor it can be used for comparative mapping [39, 40]. In addition, the allele frequencies are also needed for DNA fingerprinting in order to estimate the individual's probability of match.

Amplified fragment length polymorphism:

Another method that employs the use of specific oligonucleotide primers is termed amplified fragment length polymorphism (AFLP) technique [41]. AFLP technology is based on the selective PCR amplification of restriction fragments from the total digest of DNA and generates a banding pattern that can be used as a unique identifier. AFLP is a highly selective method for fingerprinting genomic DNA of any origin and complexity. AFLP combines several techniques which results in its high power of discrimination. The technique is used to generate high density genetic marker maps for the selected desired traits in various crops such as barley, tomato, and potatoes. Similar to RFLP and RAPD analysis, AFLP derives from mutations in restriction sites or length variation of restriction fragments [42]. AFLP offers several advantages over the

other DNA fingerprinting systems. The most important are the capacity to examine an entire genome for polymorphism. AFLP has been shown to be more reproducible than the older RAPD technique because it utilizes two specific PCR primers in the DNA amplification reaction.

The AFLP method entails several steps [43]. The first step involves restriction digestion of the genomic DNA with two specific enzymes, a rare cutter and a frequent cutter. Then, adaptors are added to the ends of the DNA fragments to generate known sequence for PCR amplification. Primers are designed to recognize the known adaptor sequence and one, two, or three additional base pairs (any one of the four possible nucleotides: A, T, G, or C) of the restriction fragments. The additional base pairs are known as selective nucleotides. PCR amplification is specific for the fragments where the primers are able to anneal. Generally, two separate selective rounds of PCR are carried out. The first round is called “pre-selective PCR amplification” which only employs one selective nucleotide. The second round is known as “selective amplification” [43] which uses the same selective nucleotide of the first round plus one or two new ones. Normally, AFLP will generate 50 to 100 amplified fragments that can be separated on polyacrylamide gel using electrophoresis [44]. AFLP detects polymorphism as presence or absence of a band.

Microsatellite:

Microsatellite DNA has evolved as a robust molecular marker technique. A microsatellite [45] is also known as simple sequence repeat (SSR) [46] or short tandem repeat (STR) [47]. Microsatellites are DNA sequences of six or fewer bases that are repeated in tandem arrays (i.e. CTCTCTCTCTCT) [48]. The most abundant type of

the motifs is the dinucleotide repeat. Microsatellites are evenly distributed in human and other mammalian genomes as well as in plants [49]. They are 10-fold more frequent in the human genome than in plant and vary significantly among species of plants [50, 51]. Microsatellites can be found in any location on the genome, but they are less abundant in protein-coding regions (exons) than in non-coding regions (introns). In addition, the total microsatellite content in the genome is related to the overall genomic size of the organism [52].

Weber defines the microsatellites structure in terms of three categories [53]: perfect repeats (without interruption i.e. $(N_1N_2)_x$); imperfect repeats (interrupted by non-repeat bases i.e. $(N_1N_2)_x(N)y(N_1N_2)_z$), and compound or complex repeats (two or more repeats runs adjacent to each other i.e. $(N_1N_2)_x(N_3N_4)_y$). The variation in the number of repeats among different fragments provides the basis for the polymorphism. Microsatellite loci can be highly polymorphic, with a recent report of two loci in barley having 28 and 37 alleles [54]. Once these microsatellites have been found, the unique DNA sequences flanking the repeats are used to design pairs of oligonucleotide primers for PCR. Because the primers are complementary to unique DNA sequences, they can be designed to be about 23 bases long, which improve the specificity of the amplification and the reproducibility of results. The alleles differ in the number of repeats of the microsatellite sequence. For example, two alleles arise if they differ by two dinucleotide repeats, thus the two possible CT alleles generated could be CT_6 (CTCTCTCTCTCT) and CT_8 (CTCTCTCTCTCTCTCT). The different alleles can be amplified by PCR and separated by gel or capillary electrophoresis that detects the length difference in the

number of microsatellite repeats. Since the amplified DNA fragments from the microsatellite loci differ in size, heterozygosity can be directly assessed. If the individual is homozygote, then both parents contribute the same allele, whereas the heterozygote's individual receive different alleles from his/her parents [55]. Unlike with RAPDs and AFLPs where Hardy-Weinberg equilibrium must be assumed to calculate heterozygosity and allele frequencies, microsatellites allow a direct test of Hardy-Weinberg equilibrium and estimation of allele frequency and heterozygosity. This method repeatedly identifies a single locus and targets hypervariable regions of the genome. Polymorphism is generally detected as a length difference in the amplified product. This length difference may be as small as one base pair.

The inherent instability of the microsatellite loci makes them very useful for evolutionary and genetic studies. Mutation rates of microsatellite loci are usually higher than the mutation rates at other loci within the same genome. Microsatellite mutation rates *in vivo* are estimated to be from 10^{-2} events per locus per generation in *E. coli* [56] to around 6×10^{-6} events per locus per generation in *Drosophila* [57]. Two models have been proposed to explain the mechanism of microsatellite instability. The first model indicates that microsatellite instability is a result of an increased rate of unequal crossing-over within the repeats. This unequal crossing over is caused by recombination among homologous chromosomes that are imperfectly aligned. The second model argues that microsatellite instability is a result of an elevated rate of slip-strand mispairing (SSM) during DNA replication. This SSM process starts with the DNA polymerase "slippage" during replication, which leads to partial dissociation of the original template and the newly replicated strand. Consequently, this may cause misaligned reassociation of the

two strands and failure in mismatch repair. The results of a number of experiments argue in favor of SSM to be the major cause of the instability [56, 59].

Several important advantages make microsatellite markers the method of choice for DNA fingerprinting and analysis of genetic relationships. Microsatellite is usually a single locus with multiple alleles and this robust technique can be easily distributed between different laboratories as primer sequences [45]. Thus microsatellites act as universal genetic markers by serving as common tool for collaborative research [43, 60]. Microsatellites can also be used in multiplex PCR where several microsatellite loci can be assayed in the same amplification reaction. Microsatellite markers are codominant, highly informative, reproducible, have high discriminatory power, and easily genotyped [61-64]. In forensics, sometime is important to detect mixed samples profiles, which only can be detected by SSR. Furthermore, the use of fluorescence detection techniques improves the resolution and accuracy in calling the allelic profiles as well as expedites data collection and thus lowers the costs. The recent advancement in electrophoresis and analysis instrumentations that provides high-throughput and semi-automated analysis allow for effective utilization of plant genetic diversity. Because of these advantages, microsatellites have become well suited for wide range of applications in genetic mapping [65], fingerprint and genotype identification [66], seed purity evaluation and germplasm conservation [67], genetic relatedness and paternity studies [68], and marker-assisted selection [69].

In human forensics, microsatellite markers are employed in the development of unique allelic profiles to determine the individual's identity. Thus, the forensic DNA community employs the use of tetranucleotide microsatellite markers for DNA

fingerprinting analysis. Today, most forensic laboratories are using the 13 CODIS loci, which are capable of generating DNA profiles with high discriminating power [55]. The importance of plant DNA fingerprinting to forensics began in 1993 when the legal precedent was established for the use of plant DNA evidence in criminal casework. In *Bogan vs. State of Arizona*, a woman was killed and her body was dumped in the Arizona desert under a Palo Verde tree. Seedpods from a Palo Verde tree found in Bogan's car were identified as a DNA match to the same tree, where the murder victim was found. The plant DNA profiles used in this case were RAPDs-based [70].

Cannabis (marijuana) is one of the most widely used illicit drugs in United States and throughout the world [71]. Marijuana is encountered frequently in daily criminal cases. Molecular marker systems based on RAPD [72] and AFLP [73, 74] have been developed for *C. sativa* and have been used for fingerprinting analysis. However, RAPDs are dominant markers with serious reproducibility problems. AFLP analysis detects multiple loci with high reproducibility, however, AFLP are also dominant markers [64]. RFLP markers are highly polymorphic but they are very labor intensive. Microsatellite markers have been developed for some plants for general purposes [52, 65] but have not been applied forensically to *C. sativa*. To date, forensic labs and the general forensic science community lack a simple, standardized technique for fingerprinting analysis of marijuana plants that have been seized as evidence in a case. The objective of this work was to develop a number of microsatellite markers capable of individualizing *Cannabis* plants for fingerprinting and genetic relatedness analyses.

Materials and Methods

The construction of the microsatellite library was performed at the genetics lab at the United States Department of Agriculture-Agricultural Research Service-Subtropical Horticulture Research Station (USDA-ARS-SHRS), Miami, Florida. The enrichment protocols utilized were based on a modified version of the method described by Edwards et al. (1996) [75]. Microsatellite loci were developed by a marker enrichment technique, which consisted of: 1) hybridizing extracted genomic DNA of a known cultivar of *C. sativa* with specific repeat unit probes, 2) sequencing positive clones, 3) designing oligonucleotide primers on either side of the repeat region, and 4) testing loci for polymorphism by sampling different unrelated individuals.

Microsatellite Enrichment Protocols:

DNA isolation and preparation of genomic DNA:

Cannabis genomic DNA was provided by Dr. Heather Coyle (Connecticut State Forensic Science Laboratory, CSFSL, USA) and Dr. Gary Shutler (Royal Canadian Mounted Police, RCMP, Canada) which was extracted with a QIAGEN DNeasy kit (QIAGEN Operon, Alameda, CA) according to the manufacturer's recommendations [74]. Then, 360 ng of the genomic DNA (18 ng/ μ l) was digested with 5 U of the restriction endonuclease enzyme Sau3AI (10 U/ μ l), 2.5 μ l of 10X reaction buffer (Life Technologies, Inc., Gaithersburg, MD), and dH₂O to a final volume of 25 μ l. This mixture was incubated at 37° C for 90 minutes followed by inactivation of enzyme at 65° C for 10 minutes. To confirm that the genomic DNA was cut by Sau3AI, a 500 ng of the

digested DNA was electrophoresed on a 0.7% agarose gel using 50% TBE buffer and visualized with ethidium bromide.

Preparation of the linkers:

The Sau linkers were prepared by the following oligonucleotides:

•**Sau-L-A: 5'-GCGGTACCCGGGAAGCTTGG-3'**

•**Sau-L-B: 5'-GATCCCAAGCTTCCCGGGTACCGC-3'**

The Sau-L-B was phosphorylated by mixing together 5 µg of Sau-L-B oligomer in TB (5 µl), 5 µl of 10X Kinase buffer, 5 µl of 10 mM ATP, 3 µl of 10 U/µl of T4 polynucleotide kinase (New England Biolabs, Inc., Beverly, MA), and dH₂O for a final volume of 50 µl. The reaction mixture was incubated at 37° C for one hour and then at 65° C for 15 minutes. After that, 5 µg of the Sau-L-A oligomer was diluted in 50 µl of 1X kinase buffer containing 2 mM ATP and mixed with 5'-phosphorylated Sau-L-B mixture. This mixture made up the Sau linkers after running the following thermocycle program: 2 minutes at 85° C, 15 minutes at 65° C, 15 minutes at 37° C, 15 minutes at 22° C, and finally incubating for 15 minutes on ice.

To purify the linkers, the mixture was extracted with equal volume of phenol:chloroform:isoamyl alcohol (PCI) using a ratio of 25:24:1. The aqueous layer containing the linkers was transferred to a clean tube and Sau linkers were precipitated using 11 µl of 3 M NaOAc (pH 5.2) to a final concentration of 0.3 M and 333 µl of absolute ethanol. The resulting solution was incubated on ice for 30 minutes and centrifuged at 12,000 rpm for 15 minutes at 0° C. The supernatant was discarded and the pellet was rinsed once with 500 µl of ice cold 70% ethanol. The ethanol solution was

centrifuged at 12,000 rpm for 15 minutes at 0° C and the supernatant was poured off. Finally, the recovered Sau linkers pellet was air-dried at room temperature for 10 minutes and then resuspended in 10 µl of nuclease-free water.

DNA/linkers ligation:

In the ligation step, 200 ng of the digested DNA was mixed with 8 µl of 5X ligase buffer, 2 µl of T4 DNA ligase (Life Technologies, Inc., Gaithersburg, MD), about 1 µg of Sau linkers (1 µl), and dH₂O to a final volume of 40 µl. This reaction mix was incubated at 4° C for 72 hours. The ligation reaction was stopped by heating at 65° C for 10 minutes and then put on ice. The ligation reaction was purified to remove excess linkers with Performa DTR Gel Filtration Cartridges (Edge Biosystems, Inc., Beverly, MA) according to the manufacturer's recommendations. The success of the ligation reaction was verified using a PCR amplification in which the PCR mix had the following:

Reagent	Amount (µl)
Linkers-ligated DNA	5
10X buffer with 15 mM MgCl ₂	2.5
dNTP (2.5 mM each)	2
Sau-L-A primer (10 µM)	1
<i>Taq</i> polymerase (5 U/µl)	0.5
dH ₂ O	14
Total	25

The reaction mixture was amplified using the following PCR program: 94° C for 3 minutes, (94° C for 1 minute, 68° C for 1 minute, 72° C for 1 minute) for 30 cycles, 72° C for 10 minutes, and hold at 4° C. Finally, 5 µl of PCR product was electrophoresed on a 0.9% agarose gel and 50% TBE at 109 voltage for one hour. A successful ligation

reaction would show a visible “smear” of amplified PCR products when visualized with ethidium bromide.

Pre-amplification for microsatellite enrichment:

The following reaction mix was prepared in triplicates:

Reagent	Amount (μ l)
Linkers-ligated DNA	1
10X buffer with 15 mM MgCl ₂	5
Sau-L-A primer (10 μ M)	2
dNTP (2.5 mM each)	4
<i>Taq</i> polymerase (5 U/ μ l)	0.5
dH ₂ O	37.5
Total	50

The triplex mixtures were amplified using the following PCR program: 94° C for 3 minutes, (94° C for 1 minute, 68° C for 1 minute, 72° C for 1 minute) for 20 cycles, 72° C for 10 minutes, and hold at 4° C. The triplex reaction mixtures were pooled together and purified using the PCI extraction in which equal volume of PCI was added to the mix, vortexed thoroughly, centrifuged at 13,000 rpm for 10 minutes, and the top layer was transferred to a clean tube. Then the solution was ethanol precipitated by adding 5 M of NaCl to a final concentration of 100 mM (3 μ l) and 2.5X volumes of ice-cold absolute ethanol (383 μ l). The resulting solution was incubated on ice for 60 minutes and centrifuged at 13,000 rpm for 20 minutes on 0° C. The supernatant was carefully discarded and the DNA pellet was dried in a vacuum centrifuge and resuspended in 29 μ l of nuclease free water. 4 μ l of the result was electrophoresed in a 0.9% agarose gel and 50% TBE to check if enough DNA was recovered during the pre-amplification.

Preparation for enrichment of microsatellite:

Twelve different biotinylated oligonucleotide probes were employed all of which contain Biotin-ATAGAATAT at the 5' end [75]. Those probes consisted of two types of dinucleotide repeats: (CT)₁₅ and (GT)₁₅ and ten types of trinucleotide repeats: (CAA)₁₀, (ATT)₁₀, (GCC)₁₀, (ACC)₁₀, (AGG)₁₀, (CTT)₁₀, (AGC)₁₀, (ACG)₁₀, (ACT)₁₀, and (ATC)₁₀ (Invitrogen, Chicago, IL). Each type of the biotinylated oligonucleotide probes was resuspended in TE (pH 8) to make 1 µg/µl the stock solution. In a separate tube, each probe solution was diluted 1/10 with nuclease free water. A master mix containing the 12 probes was prepared 10 µl of the diluted solution (120 µl total volume).

To prepare the dynabeads for enrichment, 100 µl aliquot of the stock dynabeads solution was washed to remove the sodium azide preservative as described by the following manufacturer's steps:

1. The Dynabeads M-280 Streptavidin (Dyna, Oslo, Norway) were resuspended in sodium azide by gently shaking the vial.
2. 100 µl of dynabeads solution was transferred to a clean microcentrifuge tube.
3. The tube was placed in the Dynal MPC magnet for 2 minutes without disturbance.
4. The supernatant was removed carefully with a pipette and, as possible, avoiding to touch the inside wall of the tube where the magnetic beads are attracted.
5. The tube was removed from the magnet and 100 µl of the 2X B&W wash buffer was added to resuspend the dynabeads again. The 2X B&W buffer was made by mixing 10 mM of tris-HCl (pH 7.5), 1 mM of EDTA, and 2 M of NaCl.
6. The tube was placed in the Dynal MPC magnet for 2 minutes and the wash buffer was then removed.

7. The wash was repeated 4 more times using 100 μ l of 2X B&W buffer as described by step 5 and 6.
8. After the final wash, the dynabeads were resuspended in another 100 μ l of 2X B&W buffer.
9. In this step, 100 μ l of the 12 biotinylated probes mix made earlier was added to the dynabeads and mixed thoroughly (100 ng/ μ l total oligo concentration).
10. The dynabeads were allowed to hybridize with the biotinylated probes by placing the tube on a mixing table for 30 minutes at room temperature.
11. The tube was placed in a MPC magnet for 3 minutes and the supernatant was removed afterward.
12. The dynabeads were washed 3 times with 100 μ l of 1X B&W buffer as above.
13. Finally, the dynabeads were resuspended in 500 μ l of 2X B&W buffer.

Enrichment of microsatellite sequence:

The pre-amplified linkers-ligated DNA was denatured by incubating at 95° C for 7 minutes and immediately placed on ice for at least 5 minutes. To enrich for the microsatellite-containing fragments, the following mix was prepared:

Reagent	Amount (μ l)
Denatured DNA	25
Sau-L-A oligo (1 μ g)	1
Hybridization buffer*	474
Dynabeads in 2X buffer	500
Total	1000

*The hybridization buffer ingredients:

50% formamide, 3X SSC, 25 mM Na-phosphate (pH 7), and 0.5% SDS.

The reaction solution was mixed overnight at room temperature on a mixing table.

Dynabeads/hybridization washes:

The washes were carried on as follows:

1. The hybridization reaction was placed in the MPC magnet for 3 minutes at room temperature.
2. Leaving the tube in the MPC magnet, the hybridization solution was pipetted out.
3. 1.0 ml of wash buffer #1 (2X SSC and 0.01% SDS) was added.
4. The reaction tube was then removed from the MPC and the solution was mixed gently to resuspend the dynabeads. After that, the solution was incubated at 42° C for 3 minutes in waterbath without agitation.
5. Again, the tube placed in MPC for 2 minutes at 42° C (MPC was placed in the waterbath from now on).
6. The supernatant was removed and fresh 1 ml of wash buffer #1 was added.
7. The washes were repeated 4 times (steps 3-6) with wash buffer #1.
8. The washes were repeated 3 times using wash buffer #2 (0.5X SSC and 0.01% SDS) as described in steps 3-6.
9. After the final wash, the entire wash buffer was removed and the dynabeads were resuspended in 200 µl of nuclease free water then transferred to a clean 500 µl.
10. The DNA fragments were eluted and captured by denaturing at 95° C for 7 minutes and placing the tube immediately on ice for at least 10 minutes.
11. The tube was then place in MPC magnet (set on ice) and incubated for 2 minutes.
12. The supernatant (enriched DNA) was removed and transferred to a 1.5 µl tube. The dynabeads were resuspended in 500 µl of 2X B&W buffer and stored at 4° C.

13. The recovered enriched DNA was concentrated by adding 5 M NaCl to a final concentration of 100 mM NaCl (4 μ l) and 2.5X volumes of ice-cold absolute ethanol. The solution was incubated on ice for 60 minutes and centrifuged at 13,000 rpm for 20 minutes at 0° C.
14. The supernatant was poured off and the DNA pellet was centrifuged in vacuum for 30 minutes.
15. The DNA was resuspended in 25 μ l of nuclease free water.

To further increase the DNA fragments containing microsatellites, second PCR amplification and enrichment were performed.

Pre-cloning amplification:

After the second enrichment of the microsatellite containing fragments, the PCR reaction mix was prepared in triplicate containing the following:

Reagent	Amount (μ l)
Linkers-ligated DNA	1
10X buffer with 15 mM MgCl ₂	5
Sau-L-A primer (10 μ M)	2
dNTP (2.5 mM each)	4
<i>Taq</i> polymerase (5 U/ μ l)	0.5
dH ₂ O	37.5
Total	50

The pre-cloning amplification was carried on by denaturing at 94° C for 3 minutes, followed by 25 cycles of 1 minute at 94° C, 1 minute at 68° C, and 1 minute at 72° C. The final extension was 10 minute at 72° C and hold at 4° C. The triplex PCR reactions were pooled together and purified once by equal volume of PCI. 3 μ l of 5 M NaCl was added to the reaction tube and mixed thoroughly (100 mM NaCl final concentration).

Fragment size fractionation:

First, the SizeSep 400 Spun Column Sepharose (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) was equilibrated as described by the manufacturer. The equilibration wash buffer was prepared to contain 10 mM tris-HCl (pH 8), and 100 mM NaCl. Then the spun column was inverted several times to resuspend the Sepharose CL-4B gel. The top and bottom caps were removed and placed aside. The column was allowed to drain out for about 10 minutes but without drying the gel. The bottom cap was replaced and 2 ml of equilibration buffer was added. Next, the top cap was replaced and the column was inverted several times to resuspend the gel again. Both caps were removed again and the column was allowed to drain as before. The wash with the equilibration buffer was repeated twice more. At the final wash, the buffer was stopped just as the last of the buffer enters the top of the gel and then both caps were replaced.

Second, the size-fractionation of the enriched DNA fragments was performed following the manufacturer's recommendations:

1. A 1.5 ml microcentrifuge tube was placed in the bottom of a 30 ml cortex tube. The caps were removed from the spun column and placed inside the microcentrifuge tube so that the tip of the column would be inside the tube.
2. The entire cortex tube was centrifuge at 400 x g for 2 minutes.
3. The column was removed and placed upright in a rack so that the gel should be compact and dry.
4. The enriched DNA solution was slowly applied to the top of the gel by carefully avoiding the run of the DNA past the sides of the gel.
5. The loaded column was placed back into the 30 ml cortex tube with a clean tube.

6. The cortex tube was centrifuged at 400 x g for 2 minutes.
7. The column was discarded and the effluent collected in the microcentrifuge tube was saved for the next step.

Third, the size-fractionated DNA was concentrated by ethanol precipitation as described previously. The final DNA recovered was resuspended in 25 μ l of nuclease free water.

Cloning reaction:

The cloning reaction was performed using the TOPO TA Cloning Kit (Invitrogen Corp., Chicago, IL) as described by the manufacturer's recommendations:

1. The ligation reaction was made by mixing 4 μ l of the enriched DNA, 1 μ l of salt solution, and 1 μ l of TOPO vector.
2. The reaction was incubated at room temperature for 30 minutes.
3. 2 μ l of the ligation reaction mixture was transferred to one vial of competent *E. coli* cell. The result was mixed gently without a pipette and the remaining ligation reaction mixture was store at - 20° C.
4. The vial was incubated on ice for 30 minutes.
5. The plasmid vectors were transformed into *E. coli* cells by heat shocking at 42° C for 30 seconds without agitation.
6. The vial was placed immediately on ice.
7. 250 μ l of SOC broth was added to the *E. coli* cells followed by incubating at 37° C for 60 minutes with shaking at 200 rpm.

8. 30 μ l aliquot of the transformation mix was plated on separate pre-warmed Luria-Bertani agar plates with 100 μ g/ml Ampicillin using a sterilized spreader. The entire volume of the reaction mixture was plated out.
9. The plates were placed at 37° C overnight.

Plasmid recovery:

Sterile toothpicks were used to inoculate 96-well plates containing 100 μ l of SOC broth with 100 μ g/ml Ampicillin. The broth culture plates were covered with aluminum foil tape and placed at 37° C overnight. 25 μ l of the cells culture from each well was transferred to another 96-well plate. The plates were centrifuged to pellet cells and then were inverted and spun for a short time to remove the broth. The cells were resuspended with 50 μ l of 10 mM tris-HCl (pH 8). The plates were covered with foil tape and stored at -20° C until using in the next step. These cells were used as a template for insert PCR with M13 primers (Life Technologies, Inc., Gaithersburg, MD). The remaining 75 μ l of the cells culture were store at -20° C after adding 25 μ l of glycerol and spinning at 1700 x g for 20 minutes to pellet cells.

M13 insert amplification:

The following amplification mix was prepared:

Reagent	1 rxn (μ l)	105 rxn (one plate, μ l)
10X buffer with 15 mM MgCl ₂	1.5	157.5
dNTP (2.5 mM each)	0.6	63
M13 Forward	0.3	31.5
M13 Reverse	0.3	31.5
<i>Taq</i> polymerase (5 U/ μ l)	0.06	6.3
dH ₂ O	9.24	970.2
Total	12	1260

12 μ l of master mix was added to each well of 96-well PCR plate. To the appropriate well, 3 μ l of the cells culture solution was added as well. The reaction mix was amplified using the following PCR program: 94° C for 5 minutes, (94° C for 30 seconds, 60° C for seconds, and 72° C for 2 minutes) for 30 cycles, and final extension for 10 minutes at 72° C then holding at 4° C. 5 μ l of PCR products from two columns were electrophoresed on 2% agarose gel with 50% TBE. This was to check for the success of M13 amplification and size fractionation in which all plasmid inserts should be 400 bp or above.

Exonuclease I treatment:

In a half-skirted plate, the following reaction mix was prepared:

Reagent	1 rxn (μ l)	105 rxn (one plate, μ l)
Exonuclease I (20 U/ μ l)	0.025	2.63
Exonuclease 10X buffer	2	210
dH ₂ O	12.975	1,362
Total	15	1,575

15 μ l of reaction mix was added to each well of half-skirted plate. Then, 5 μ l of M13 PCR products were added to the appropriate wells. The PCR products were incubated using the following thermalcycle program: 37° C for 60 minutes, 65° C for 20 minutes, and hold on 4° C. 40 μ l of absolute ethanol was added to each well. The plate was covered with aluminum foil tape and mixed by inverting for several times. The plate was incubated for 15 minutes at room temperature. Then the plate was centrifuged at 1700 x g for 30 minutes. The foil tape was removed followed by inverting the plate and spinning at 10 x g for 30 seconds. The plate was air dried for at least 15 minutes, covered with aluminum foil tape, and saved at -20° C until used for cycle sequencing.

Cycle sequencing:

The PCR products amplified from plasmid inserts were sequenced using BigDye Terminator™ Cycle Sequencing Ready Reaction Kit (Version 2.0) (Applied Biosystems, Foster City, CA) on an MJ Research Thermocycler (MJ Research, Inc., Waltham, MA).

The cycle sequencing mix was prepared with the following:

Reagent	1 rxn (μl)	104 rxn (one plate, μl)
T7 primer (1.6 pmol/μl)	1	104
5X buffer	1.75	182
BigDye Terminator	0.5	52
dH ₂ O	6.75	702
Total	10	1040

10 μl of the sequencing mix was added to each well containing the previously dried insert PCR product. The cycle sequencing reaction was carried out using the following PCR program: 96° C for 10 seconds, (50° C for 5 seconds, 60° C for 4 minutes) for 26 cycles and hold at 4° C. After that, 40 μl of the 76% ethanol was added to each well plate to concentrate the sequencing products. The plate was then covered with aluminum foil tape and mixed thoroughly by inverting. The sequencing products were allowed to precipitate overnight at room temperature on dark location. The next day, the plate was taken out and centrifuged at 1810 x g for 45 minutes (4000 rpm). The foil tape was removed and plate was inverted and centrifuged for 10 x g for 30 seconds. The plate was then air dried for at least 15 minutes.

The sequencing products were denatured by adding 10 μl of Hi-Di formamide to each well and incubating at 95° C for 2 minutes then placed immediately on ice for at least 10 minutes. Next, the plate was centrifuged at 3500 rpm for 20 seconds and loaded

into the ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). The sequencing products were electrophoresed on the ABI 3100 and analyzed with DNA Sequencing Analysis Software 3.7 (Applied Biosystems, Foster City, CA). The sequences generated from the automatic sequencer were analyzed using the GCG Wisconsin Packages (Accelrys, Madison, WI) to identify those that contain microsatellite loci. Microsatellite loci were imported into Sequencher v.4.1 to sort, clean up, and generate consensus sequences for primer design.

Primer design and fragment analysis:

The sequence files for those that have enough flanking sequences at both ends were combined with the others that needed reverse sequencing using T3 universal primer to be used for primer design. Based on the flanking sequences, the PCR primers were designed using GCG Wisconsin Packages (Accelrys, Madison, WI). The GCG program applied specific criteria when selecting the optimal primers candidates. Some of these most important criteria were:

GCG Primer Design Selection Criteria	
Primer Size:	18-22 nucleotides
Primer GC Content:	40-55%
Primer Tm:	50-65° C
Product Length:	100-350 bp
Product GC Content:	40-55%
Product Tm:	70-95° C
Difference in Primer Tm:	< 2.0° C
Duplicate Primer Endpoints is Not Allowed	

One primer pair was selected and synthesized for each locus from QIAGEN Operon (Alameda, CA). All primer pairs were tested by PCR with six different annealing temperatures (50° C, 53.2° C, 55.5° C, 58.4° C, 61.8° C, and 64.6° C). The optimal

annealing temperature was determined to be 53° C at which the amplicons had the highest intensity as measured by relative fluorescent unit (rfu). With some modification, the M13 fluorescent tail primer method described by Roy et al. (1996) [76] was used as a screening technique to detect polymorphism among the different loci and subsequently for fragment analysis. In this method, 19 oligonucleotide bases was added at the 5' end of the forward PCR primers having the following sequence:

5'- CACGACGTTGTAACGAC -3'

This above sequence was identical to the fluorescently labeled (fam) universal M13 forward (-29) primer, which was also included in the amplification reaction.

For fragment analysis, the tailed forward primers gave a complementary sequence to the M13 primer, which was employed for priming in the PCR amplification to subsequently produce fluorescently-labeled PCR product. The PCR reaction was prepared (for one reaction) by adding 1 µl of 10X buffer with 15 mM MgCl₂, 0.2 µl of dNTP's (2.5 mM each), 0.25 µl of forward tail primer (10 µM), 0.25 µl of reverse primer (10 µM), 0.6 µl M13 tag primer (1 µM with fam), 0.1 µl of *taq* polymerase (5 U/µl), 1 µl of DNA template (10 ng/µl), and dH₂O to a final volume of 10 µl. The PCR reaction was carried on by denaturing at 94° C for 5 minutes, followed by 40 cycles of 45 seconds at 94° C, 45 seconds at 53° C, and 1 minute at 72° C. The final extension was 10 minute at 72° C and hold at 4° C. For those primer pairs that did not produce PCR product with M13-tail fluorescent label method, direct 5'-labeled fluorescent primers were used (Invitrogen Corp., Chicago, IL).

Determination of allele size:

The PCR products were electrophoresed in a capillary electrophoresis instrument (ABI Prism 3100) using Performance Optimized Polymer 4 (POP4) (Applied Biosystems, Foster City, CA). Each sample was prepared by mixing 1.5 µl of the PCR product with 12 µl of Hi-Di formamide and 0.1 µl of GeneScan 500 ROX fluorescently labeled size standard (Applied Biosystems, Foster City, CA). The PCR products were denatured by incubating at 95° C for 2 minutes and immediately cooling on ice for at least 10 minutes. Samples were injected electrokinetically at 3 kV for 10 seconds and were run at 60° C for 45 minutes at 15 kV. The data generated was imported into GeneScan 3.7 software (Applied Biosystems, Foster City, CA) for fragment size determination. The final allele size determination of the microsatellite data was performed using Genotyper 3.7 software (Applied Biosystems, Foster City, CA).

Statistical analyses:

All samples were scored for the allele designations based on repeats size which were employed in different statistical analyses. To investigate some genetic parameters of polymorphism, the following was calculated: allele frequencies, number of alleles per locus (n), effective number of alleles (n_e), observed heterozygosity (H_o), expected heterozygosity (H_e), and probability of identical genotypes (PI). The genetic parameters were determined using the eleven microsatellite over 33 diploid *Cannabis* plants (excluding duplicate samples). Observed heterozygosity (H_o) was obtained by dividing the number of heterozygous plants over the total number of plants tested for each locus. The degree of polymorphism was measured using expected heterozygosity (H_e) [77]:

$$H_e = 1 - \sum P_i^2.$$

where P_i is the frequency of the i th alleles for each locus in *Cannabis*. Probability of identical genotypes (PI) was estimated according to Paetkau et al. (1995) [78]:

$$PI = \sum P_i^4 - \sum \sum (2 P_i P_j)^2$$

where P_i is the frequency of i th allele and P_j is the frequency of $(i + 1)$ th allele detected.

Effective number of alleles (n_e) was calculated based on Morgante's formula [65]:

$$n_e = (\sum p_i^2)^{-1}$$

The genetic relationships among the unique *Cannabis* profiles were analyzed using the neighbor joining method (NJ) [79] and unweighted pair group method (UPGMA) [80]. These cluster analyses (NJ and UPGMA) represented collections of statistical methods that were used to divide a set of members into groups (clusters). Group members would share certain properties in common and those members that were genetically related would have close genetic distance whereas those that were less related would be further a part in the dendrogram. The analyses might reveal associations and structure in data which could provide some insight into the research topic. The NJ and UPGMA trees were generated based on the following genetic distance: Chord [81], unbiased Nei [82], and modified Roger [83]. In order to give a confidence limit to the relationships between *Cannabis* plants, 2000 replicas of bootstrapping [84] were performed with the NJ method to test support at the branch nodes. Those nodes with bootstrap values below 50% are considered unsupported. Principal component analysis (PCA) was also used to show correlation between the data sets in 3 dimensional space across the plants tested [80]. The calculations of genetic distances, cluster analyses, and PCA were performed using the NTSYSpc v.2.1 package [85]. The bootstrap was performed using the TreeMaker program [86].

Results

Characterization of isolated microsatellite sequences:

The cloning step of the enriched *Cannabis* DNA generated 685 clones, from which 192 clones were sequenced (two 96-well plates). Ninety-five (95) clones were considered useful as they contained either dinucleotide motifs with nine or greater repeat units, or they contained trinucleotide motifs with five or greater repeat units. The types of microsatellite motifs identified were consistent with the twelve types of oligonucleotide probes that were used for the enrichment. The isolated microsatellite sequences were as follows: 51% dinucleotide repeats, 49% trinucleotide repeats, 79% perfect repeats, 14% imperfect repeats, and 7% compound repeats (table 1).

Total clones of enriched DNA	685
Total sequencing reactions	192
Total microsatellite inserts (%)	49%
Frequency of microsatellites in <i>Cannabis sativa</i> (%)	
GAT/CTA	GA/CT
AAG/TTC	GTT/CAA
GT/CA	CAT/GTA
ACG/TGC	GGA/CCT
10.2%	50.0%
15.3%	16.3%
1.0%	5.1%
1.0%	1.0%
Type of repeat (%)	
By nucleotide string	
Di	51.0%
Tri	49.0%
By form	
Perfect	78.6%
Imperfect	14.3%
Compound	7.1%

Table (1): Microsatellite enrichment success for *Cannabis sativa* and characterization of the microsatellite types

The isolated microsatellites had two types of dinucleotide repeats and six types of trinucleotide repeats. The majority of microsatellite loci were composed of a GA/CT dinucleotide motif representing 50% overall. The most common isolated trinucleotide motifs were GTT/CAA, AAG/TTC, and GAT/CTA representing 16%, 15%, and 10% respectively, of all detected microsatellites (table 1). The maximum repeat units recorded for dinucleotide motifs were 49 repeats and 17 repeats for the trinucleotide motifs. The number of repeat units ranged between 15 bp (5 trinucleotides) to 98 bp (49 dinucleotides). A complete list of the frequency of microsatellite types recovered from *Cannabis sativa* is shown in table 1.

Characterizations of the selected microsatellite sequences:

Thirty-six clones had suitable flanking regions for GCG primer design. From these 36 loci, seven could not generate primer sets because of low annealing temperature particularly due to a high concentration of nucleotides A and T. Of 29 primer pairs designed by GCG, 25 sets were selected, synthesized and tested for polymorphism. Fourteen primer pairs were eliminated because they produced no PCR products, nonspecific products, or complex (uninterpretable) products. These products were mainly due to the high tendency of primers to produce palindromes or primer dimers between the two primer sequences. In addition, two of those fourteen primer pairs were monomorphic. The remaining eleven loci, examples shown in figure 2-4, were found to be polymorphic and reliable for scoring the different alleles across the 41 *Cannabis* samples (table 2). The original genomic DNA used in the library construction (G40, table 3) was also included to provide a positive size control. All of the amplified products were in the expected size range and the PCR products ranged from 105 bp to

339 bp. The eleven STR markers were derived from three dinucleotide repeats, five trinucleotide repeats, one compound trinucleotide repeat, and two imperfect trinucleotide repeats (table 2).

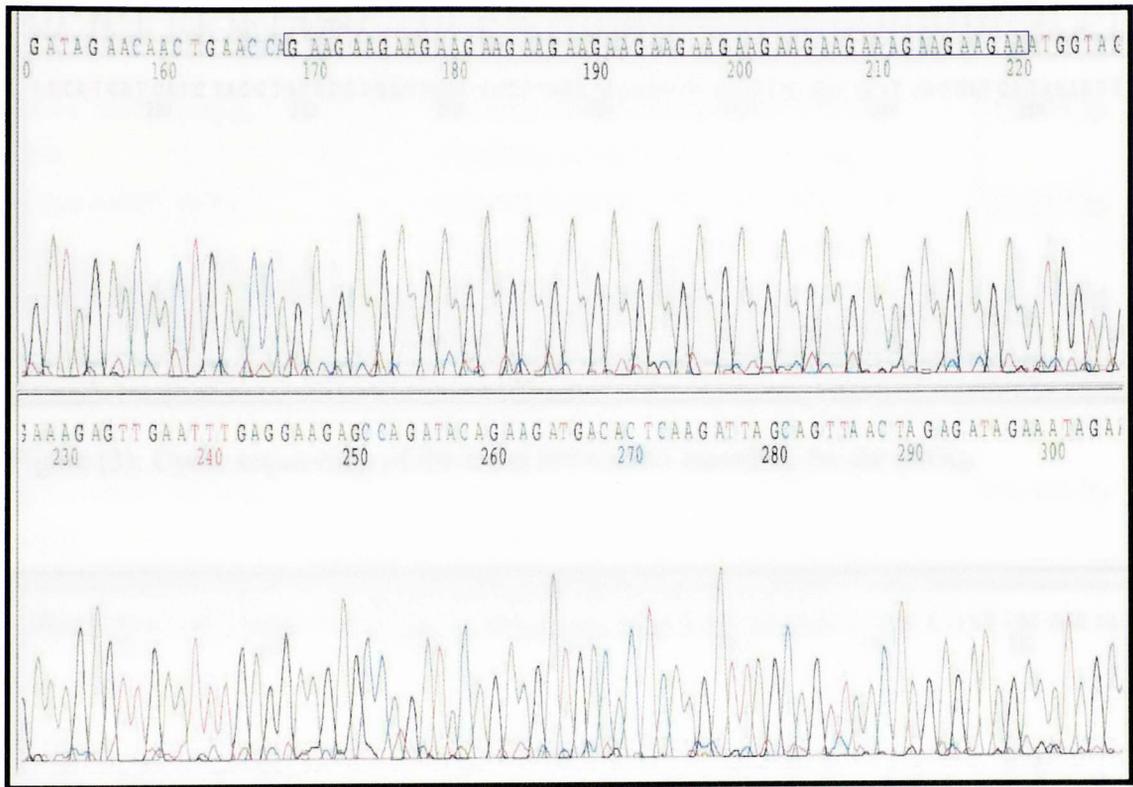


Figure (2): Cycle sequencing of the locus B01-CANN1 encoding for the (GAA)₁₃(A)(GAA)₃

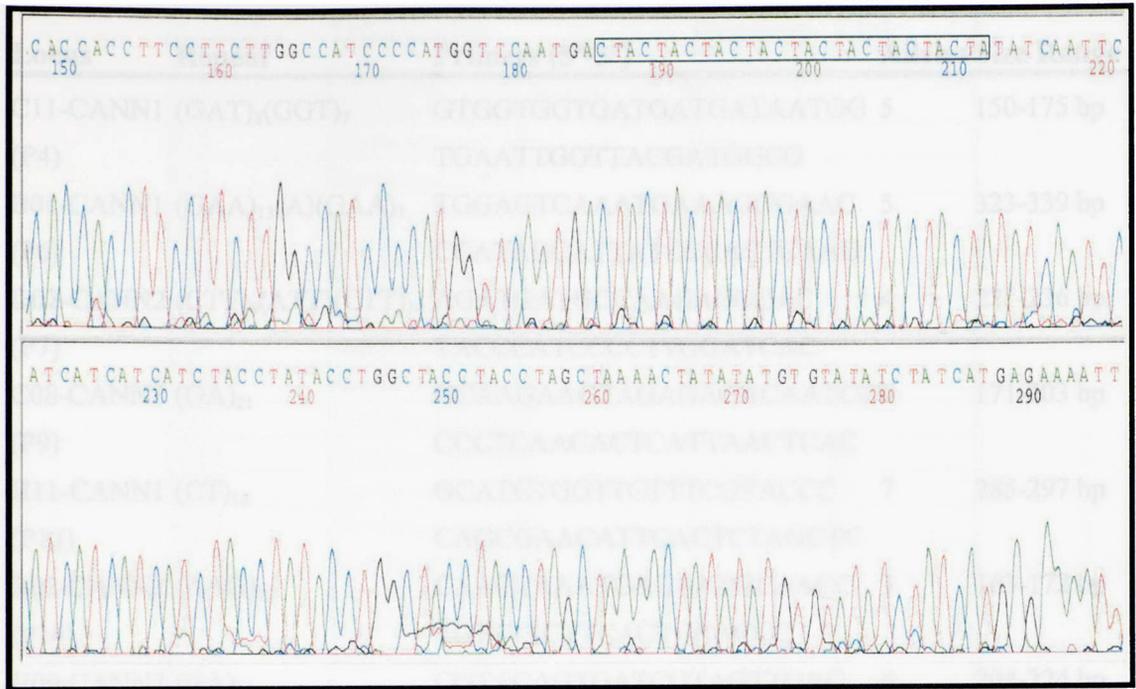


Figure (3): Cycle sequencing of the locus E07-CANN1 encoding for the (CTA)₉

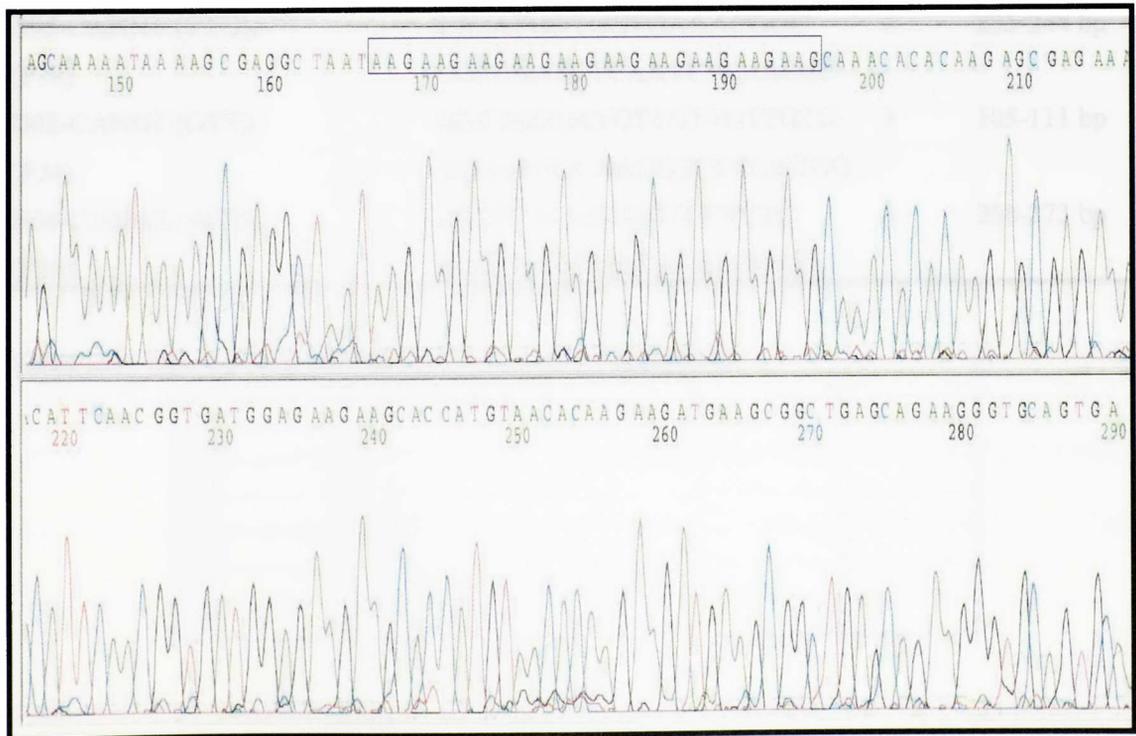


Figure (4): Cycle sequencing of the locus B02-CANN2 encoding for the (AAG)₁₀

Locus	Repeat	Primers (5'-3')	Alleles Size Range
C11-CANN1 (P4)	(GAT) ₈ (GGT) ₇	GTGGTGGTGATGATGATAATGG TGAATTGGTTACGATGGCG	5 150-175 bp
B01-CANN1 (P6)	(GAA) ₁₃ (A)(GAA) ₃	TGGAGTCAAATGAAAGGGAAC CCATAGCATTATCCCACTCAAG	5 323-339 bp
D02-CANN2 (P7)	(CTT) ₆ (ATT)(CTT) ₁₀	AGATGATGCCAAGAGGCAC TACCCATCCCCTTGGATCAC	4 221-236 bp
C08-CANN2 (P9)	(GA) ₂₁	GCAAGAAGTAGAGAGACAATCC CCCTCAACACTCATTA ACTCAC	9 171-203 bp
H11-CANN1 (P13)	(CT) ₁₈	GCATGTGGTTGTTTCGTACCC CAGCGAACATTC ACTCTAGCTC	7 285-297 bp
B02-CANN2 (P14)	(AAG) ₁₀	CAACCAAATGAGAATGCAACC TGTTTTCTTCACTGCACCC	3 163-172 bp
H09-CANN2 (P15)	(GA) ₁₅	CGTACAGTGATCGTAGTTGAG ACACATACAGAGAGAGCCC	6 204-224 bp
E07-CANN1 (P17)	(CTA) ₉	CAAATGCCACACCACCTTC GTAGGTAGCCAGGTATAGGTAG	3 105-111 bp
B05-CANN1 (P19)	(TTG) ₉	TTGATGGTGGTCAAACGGC CCCCAATCTCAATCTCAACCC	4 235-244 bp
D02-CANN1 (P24)	(GTT) ₇	GGTTGGGATGTTGTTGTTGTG AGAAATCCAAGGTCCTGATGG	3 105-111 bp
H06-CANN2 (P25)	(ACG) ₇	TGGTTTCAGTGGTCCTCTC ACGTGAGTGATGACACGAG	3 266-273 bp

Table (2): The 11 SSR markers and the primer sequences

Samples	Information	Duplicate	Link
G1	Northeast-street#1	1	
G2	Northeast-street#1	1	
G3	Northeast-street#1	2	
G4	Northeast-street#1		
G5	Northeast-street#1	3	
G6	Northeast-street#1	3	
G7	Northeast-street#1	4	
G8	Northeast-street#1		
G9	Northeast-street#1		
G10	Northeast-street#1	5	
G11	Northeast-street#2		
G12	Northeast-street#2		A
G13	Northeast-street#2		A
G14	Northeast-street#2	6	
G15	Northeast-street#2	7	
G16	Northeast-street#1	2	
G17	Northeast-street#1	4	
G18	Northeast-street#1	5	
G19	Northeast-street#2	6	
G20	Northeast-street#2	7	
G21	Canada		
G22	Canada		9
G23	Canada		E
G24	Canada		
G25	Canada		C
G26	Canada		10-C
G27	Canada		E
G28	Canada		10-C
G29	Canada		
G30	Canada		
G31	Canada		11
G32	Canada		D
G33	Canada		D
G34	Canada		11
G35	Canada		D
G36	Canada		9
G37	Canada		
G38	Canada		B
G39	Canada		B
G40	Used in library construction	8	9
G41	Canada	8	

Table (3): List of information and association known between the 41 *C. sativa* plants (personal communication with Dr. Coyle at CSFSL and Dr. Shutler at RCMP)

Continue Table (3):

- * Identical number (1-8) code for duplicate samples
- * Identical letter code for related samples
- A: closely related by AFLP
- B: originated from the same grower
- C: related by AFLP
- D: originated from the same grower
- E: related by AFLP
- 9: identical based on AFLP
- 10: identical based on AFLP
- 11: identical based on AFLP

Polymorphisms of microsatellite loci:

A total of 52 alleles were detected across the eleven loci. The number of alleles per locus ranged from three at loci P14, P17, P24, and P25 to nine at locus P9 (table 4). On average, 4.7 alleles and 2.4 effective alleles per locus were detected. The difference between the average number of alleles per locus and the effective number of alleles per locus infers the presence of rare alleles that occurred in few genotypes. Allele frequencies for 33 *Cannabis* samples excluding the duplicates were generally low (ranging from 0.015 to 0.773) especially at loci with a large number of alleles. The level of polymorphism detected at each locus was evaluated by expected heterozygosity (H_e). Expected heterozygosity ranged between 0.368 and 0.710 with a mean value of 0.568. The observed heterozygosity (H_o) ranged between 0.152 and 0.727 with an average of 0.529. The total probability that two unrelated individuals would have the same genotype across all eleven loci by chance (PI) was estimated to be low (1.8×10^{-7}) (table 4), thus resulting in high discrimination power between unrelated individuals.

Locus	<i>Ho</i>	<i>He</i>	<i>N</i>	<i>ne</i>	PI
P4	0.6061	0.5429	5	2.2089	0.258
P6	0.5758	0.5612	5	2.2782	0.2718
P7	0.1515	0.4926	4	1.971	0.2978
P9	0.4545	0.6901	9	3.2267	0.1281
P13	0.6364	0.5882	7	2.4281	0.2384
P14	0.6061	0.595	3	2.4694	0.2484
P15	0.7273	0.7103	6	3.4517	0.1293
P17	0.6667	0.5092	3	2.0374	0.3226
P19	0.6061	0.6396	4	2.7745	0.2014
P24	0.3658	0.3678	3	1.5812	0.4418
P25	0.4242	0.5478	3	2.2112	0.302
Average	0.5291	0.5677	4.7273	2.4217	1.8×10^{-7a}

Table (4): Calculation of different genetic parameters for *Cannabis* samples excluding the duplicates

^a Product of values across all loci analysis of *Cannabis* plant

Sample Info	150	153	156	165	175
P4-g1		0.5	0.5		
P4-g2		0.5	0.5		
P4-g3		0.5	0.5		
P4-g4		1			
P4-g5	0.5		0.5		
P4-g6	0.5		0.5		
P4-g7		0.5	0.5		
P4-g8			1		
P4-g9		0.5	0.5		
P4-g10		0.5		0.5	
P4-g11		1			
P4-g12		0.5		0.5	
P4-g13				1	
P4-g14		1			
P4-g15		0.5		0.5	
P4-g16		0.5	0.5		
P4-g17		0.5	0.5		
P4-g18		0.5		0.5	
P4-g19		1			
P4-g20		0.5		0.5	
P4-g21		0.5		0.5	
P4-g22		1			
P4-g23		0.5	0.5		
P4-g24		0.5	0.5		
P4-g25		1			
P4-g26		0.5	0.5		
P4-g27		1			
P4-g28		0.5	0.5		
P4-g29		0.5			0.5
P4-g30		1			
P4-g31		0.5	0.5		
P4-g32		0.5	0.5		
P4-g33		0.5	0.5		
P4-g34		0.5	0.5		
P4-g35		1			
P4-g36		1			
P4-g37		0.5		0.5	
P4-g38		0.5		0.5	
P4-g39		1			
P4-g40		1			
P4-g41		1			

Allele Count	2	50	19	10	1
Total Alleles	82				
Allele Freq	0.0244	0.609756	0.231707	0.121951	0.012195

Table (5): Genotyping using P4 primer set, allele count, and allele frequencies

Sample Info	323	326	329	333	339
P6-g1	0.5	0.5			
P6-g2	0.5	0.5			
P6-g3	0.5	0.5			
P6-g4	0.5	0.5			
P6-g5	0.5	0.5			
P6-g6	0.5	0.5			
P6-g7	0.5	0.5			
P6-g8	1				
P6-g9	1				
P6-g10	0.5				0.5
P6-g11	1				
P6-g12		1			
P6-g13		1			
P6-g14		1			
P6-g15	0.5				0.5
P6-g16	0.5	0.5			
P6-g17	0.5	0.5			
P6-g18	0.5				0.5
P6-g19		1			
P6-g20	0.5				0.5
P6-g21		1			
P6-g22	0.5	0.5			
P6-g23	0.5	0.5			
P6-g24		1			
P6-g25	0.5	0.5			
P6-g26	0.5		0.5		
P6-g27	0.5	0.5			
P6-g28	0.5		0.5		
P6-g29		1			
P6-g30		0.5		0.5	
P6-g31		1			
P6-g32		1			
P6-g33		1			
P6-g34		1			
P6-g35		1			
P6-g36	0.5	0.5			
P6-g37		0.5			0.5
P6-g38	0.5	0.5			
P6-g39	0.5	0.5			
P6-g40	0.5	0.5			
P6-g41	0.5	0.5			

Allele count	30	44	2	1	5
Total alleles	82				
Allele freq	0.365854	0.536585	0.02439	0.012195	0.060976

Table (6): Genotyping using P6 primer set, allele count, and allele frequencies

Sample Info	221	222	230	236
P7-g1	1			
P7-g2	1			
P7-g3			1	
P7-g4	1			
P7-g5				1
P7-g6				1
P7-g7			1	
P7-g8	0.5		0.5	
P7-g9	1			
P7-g10		1		
P7-g11	0.5	0.5		
P7-g12		1		
P7-g13		1		
P7-g14	1			
P7-g15		1		
P7-g16			1	
P7-g17			1	
P7-g18		1		
P7-g19	1			
P7-g20		1		
P7-g21	1			
P7-g22	1			
P7-g23			1	
P7-g24	0.5		0.5	
P7-g25	1			
P7-g26	1			
P7-g27	1			
P7-g28	1			
P7-g29	1			
P7-g30	0.5		0.5	
P7-g31	1			
P7-g32	0.5		0.5	
P7-g33	1			
P7-g34	1			
P7-g35	1			
P7-g36	1			
P7-g37	1			
P7-g38	1			
P7-g39	1			
P7-g40	1			
P7-g41	1			

Allele count	51	13	14	4
Total allele	82			
Allele freq	0.621951	0.158537	0.170732	0.04878

Table (7): Genotyping using P7 primer set, allele count, and allele frequencies

Sample Info	171	173	179	191	193	195	199	201	203
P9-g1					1				
P9-g2					1				
P9-g3			0.5		0.5				
P9-g4		0.5							0.5
P9-g5		1							
P9-g6		1							
P9-g7			0.5		0.5				
P9-g8					1				
P9-g9		0.5			0.5				
P9-g10					0.5	0.5			
P9-g11					1				
P9-g12					1				
P9-g13					1				
P9-g14						1			
P9-g15					0.5	0.5			
P9-g16			0.5		0.5				
P9-g17			0.5		0.5				
P9-g18					0.5	0.5			
P9-g19						1			
P9-g20					0.5	0.5			
P9-g21	1								
P9-g22		0.5				0.5			
P9-g23		1							
P9-g24				0.5	0.5				
P9-g25					1				
P9-g26					1				
P9-g27		0.5			0.5				
P9-g28					1				
P9-g29		0.5			0.5				
P9-g30							0.5	0.5	
P9-g31				1					
P9-g32					1				
P9-g33					0.5	0.5			
P9-g34				1					
P9-g35					1				
P9-g36		0.5				0.5			
P9-g37		0.5				0.5			
P9-g38					1				
P9-g39					1				
P9-g40		0.5				0.5			
P9-g41		0.5				0.5			

Allele count	2	15	4	5	39	14	1	1	1
Total alleles	82								
Allele freq	0.02439	0.18293	0.04878	0.060976	0.475610	0.170732	0.012195	0.0122	0.012195

Table (8): Genotyping using P9 primer set, allele count, and allele frequencies

Sample Info	285	287	289	291	293	295	297
P13-g1		0.5				0.5	
P13-g2		0.5				0.5	
P13-g3		1					
P13-g4		1					
P13-g5						1	
P13-g6						1	
P13-g7		1					
P13-g8	0.5					0.5	
P13-g9		0.5				0.5	
P13-g10		0.5				0.5	
P13-g11		1					
P13-g12		0.5				0.5	
P13-g13		0.5				0.5	
P13-g14		0.5		0.5			
P13-g15		0.5				0.5	
P13-g16		1					
P13-g17		1					
P13-g18		0.5				0.5	
P13-g19		0.5		0.5			
P13-g20		0.5				0.5	
P13-g21		1					
P13-g22		0.5				0.5	
P13-g23		1					
P13-g24						0.5	0.5
P13-g25		1					
P13-g26		0.5				0.5	
P13-g27		0.5			0.5		
P13-g28		0.5				0.5	
P13-g29						1	
P13-g30						1	
P13-g31		0.5				0.5	
P13-g32		1					
P13-g33		0.5		0.5			
P13-g34		0.5				0.5	
P13-g35		1					
P13-g36		0.5				0.5	
P13-g37					0.5	0.5	
P13-g38		0.5	0.5				
P13-g39		0.5	0.5				
P13-g40		0.5				0.5	
P13-g41		0.5				0.5	

Allele count	1	45	2	3	2	28	1
Total alleles	82						
Allele freq	0.012195	0.54878	0.02439	0.036585	0.02439	0.341463	0.012195

Table (9): Genotyping using P13 primer set, allele count, and allele frequencies

Sample Info	163	166	172
P14-g1		1	
P14-g2		1	
P14-g3	1		
P14-g4	1		
P14-g5	1		
P14-g6	1		
P14-g7	1		
P14-g8	0.5	0.5	
P14-g9	0.5	0.5	
P14-g10	0.5	0.5	
P14-g11	0.5	0.5	
P14-g12	0.5	0.5	
P14-g13	0.5	0.5	
P14-g14	1		
P14-g15	0.5	0.5	
P14-g16	1		
P14-g17	1		
P14-g18	0.5	0.5	
P14-g19	1		
P14-g20	0.5	0.5	
P14-g21		1	
P14-g22		0.5	0.5
P14-g23	0.5	0.5	
P14-g24	0.5	0.5	
P14-g25	0.5		0.5
P14-g26		0.5	0.5
P14-g27	1		
P14-g28		0.5	0.5
P14-g29		0.5	0.5
P14-g30	1		
P14-g31	0.5	0.5	
P14-g32	1		
P14-g33	1		
P14-g34	0.5	0.5	
P14-g35	0.5	0.5	
P14-g36		0.5	0.5
P14-g37	0.5		0.5
P14-g38		1	
P14-g39		1	
P14-g40		0.5	0.5
P14-g41		0.5	0.5

Allele Count	42	31	9
Total alleles	82		
Allele Freq.	0.512195	0.378049	0.109756

Table (10): Genotyping using P14 primer set, allele count, and allele frequencies

Sample Info	204	214	218	220	222	224
P15-g1			1			
P15-g2			1			
P15-g3			1			
P15-g4	0.5			0.5		
P15-g5			1			
P15-g6			1			
P15-g7			1			
P15-g8			1			
P15-g9		0.5	0.5			
P15-g10	0.5		0.5			
P15-g11		0.5	0.5			
P15-g12	0.5			0.5		
P15-g13	0.5			0.5		
P15-g14				1		
P15-g15	0.5		0.5			
P15-g16			1			
P15-g17			1			
P15-g18	0.5		0.5			
P15-g19				1		
P15-g20	0.5		0.5			
P15-g21			1			
P15-g22	0.5		0.5			
P15-g23		0.5	0.5			
P15-g24	0.5			0.5		
P15-g25	0.5		0.5			
P15-g26	0.5		0.5			
P15-g27	0.5		0.5			
P15-g28	0.5		0.5			
P15-g29	1					
P15-g30		0.5				0.5
P15-g31	0.5			0.5		
P15-g32	0.5	0.5				
P15-g33		0.5	0.5			
P15-g34	0.5			0.5		
P15-g35		0.5	0.5			
P15-g36	0.5		0.5			
P15-g37				1		
P15-g38			0.5		0.5	
P15-g39			0.5		0.5	
P15-g40	0.5		0.5			
P15-g41	0.5		0.5			

Allele count	21	7	39	12	2	1
Total allele	82					
Allele freq	0.256098	0.085366	0.47561	0.146341	0.02439	0.012195

Table (11): Genotyping using P15 primer set, allele count, and allele frequencies

Sample Info	105	108	111
P17-g1			1
P17-g2			1
P17-g3		0.5	0.5
P17-g4			1
P17-g5	0.5		0.5
P17-g6	0.5		0.5
P17-g7		0.5	0.5
P17-g8		0.5	0.5
P17-g9	0.5		0.5
P17-g10		0.5	0.5
P17-g11		0.5	0.5
P17-g12		0.5	0.5
P17-g13		0.5	0.5
P17-g14		0.5	0.5
P17-g15		0.5	0.5
P17-g16		0.5	0.5
P17-g17		0.5	0.5
P17-g18		0.5	0.5
P17-g19		0.5	0.5
P17-g20		0.5	0.5
P17-g21			1
P17-g22		0.5	0.5
P17-g23			1
P17-g24	0.5	0.5	
P17-g25			1
P17-g26			1
P17-g27		0.5	0.5
P17-g28			1
P17-g29		0.5	0.5
P17-g30		1	
P17-g31		0.5	0.5
P17-g32			1
P17-g33			1
P17-g34		0.5	0.5
P17-g35		0.5	0.5
P17-g36		0.5	0.5
P17-g37		0.5	0.5
P17-g38	0.5		0.5
P17-g39			1
P17-g40		0.5	0.5
P17-g41		0.5	0.5

Allele count	5	27	50
Total allele	82		
Allele freq	0.060976	0.329268	0.609756

Table (12): Genotyping using P17 primer set, allele count, and allele frequencies

Sample Info	235	238	241	244
P19-g1			1	
P19-g2			1	
P19-g3		0.5	0.5	
P19-g4			0.5	0.5
P19-g5		1		
P19-g6		1		
P19-g7		0.5	0.5	
P19-g8		1		
P19-g9		0.5		0.5
P19-g10		0.5	0.5	
P19-g11	0.5		0.5	
P19-g12		1		
P19-g13		1		
P19-g14			0.5	0.5
P19-g15		0.5	0.5	
P19-g16		0.5	0.5	
P19-g17		0.5	0.5	
P19-g18		0.5	0.5	
P19-g19			0.5	0.5
P19-g20		0.5	0.5	
P19-g21				1
P19-g22		0.5	0.5	
P19-g23			1	
P19-g24	0.5			0.5
P19-g25		0.5	0.5	
P19-g26		0.5	0.5	
P19-g27			0.5	0.5
P19-g28		0.5	0.5	
P19-g29			1	
P19-g30		1		
P19-g31			1	
P19-g32		0.5		0.5
P19-g33		0.5		0.5
P19-g34			1	
P19-g35		0.5		0.5
P19-g36		0.5	0.5	
P19-g37			1	
P19-g38		0.5	0.5	
P19-g39			1	
P19-g40		0.5	0.5	
P19-g41		0.5	0.5	

Allele count	2	32	37	11
Total allele	82			
Allele freq	0.02439	0.390244	0.45122	0.134146

Table (13): Genotyping using P19 primer set, allele count, and allele frequencies

Sample Info	105	108	111
P24-g1		0.5	0.5
P24-g2		0.5	0.5
P24-g3		1	
P24-g4		1	
P24-g5		1	
P24-g6		1	
P24-g7		1	
P24-g8		1	
P24-g9		1	
P24-g10		0.5	0.5
P24-g11	0.5	0.5	
P24-g12		1	
P24-g13		1	
P24-g14			1
P24-g15		0.5	0.5
P24-g16		1	
P24-g17		1	
P24-g18		0.5	0.5
P24-g19			1
P24-g20		0.5	0.5
P24-g21		0.5	0.5
P24-g22		0.5	0.5
P24-g23		1	
P24-g24		0.5	0.5
P24-g25		1	
P24-g26		1	
P24-g27		1	
P24-g28		1	
P24-g29		1	
P24-g30	1		
P24-g31		0.5	0.5
P24-g32		1	
P24-g33		1	
P24-g34		0.5	0.5
P24-g35		1	
P24-g36		0.5	0.5
P24-g37		1	
P24-g38		1	
P24-g39		1	
P24-g40		0.5	0.5
P24-g41		0.5	0.5
Allele count	3	61	18
Total alleles	82		
Allele freq	0.036585	0.743902	0.219512

Table (14): Genotyping using P24 primer set, allele count, and allele frequencies

Sample Info	266	269	273
P25-g1	0.5	0.5	
P25-g2	0.5	0.5	
P25-g3	0.5	0.5	
P25-g4	0.5	0.5	
P25-g5	0.5	0.5	
P25-g6	0.5	0.5	
P25-g7	0.5	0.5	
P25-g8		1	
P25-g9		1	
P25-g10		1	
P25-g11		0.5	0.5
P25-g12	1		
P25-g13	1		
P25-g14	1		
P25-g15		1	
P25-g16	0.5	0.5	
P25-g17	0.5	0.5	
P25-g18		1	
P25-g19	1		
P25-g20		1	
P25-g21	0.5	0.5	
P25-g22		1	
P25-g23	0.5		0.5
P25-g24	0.5	0.5	
P25-g25		1	
P25-g26		1	
P25-g27	0.5	0.5	
P25-g28		1	
P25-g29		1	
P25-g30	1		
P25-g31	1		
P25-g32	0.5	0.5	
P25-g33	1		
P25-g34	1		
P25-g35	0.5	0.5	
P25-g36		1	
P25-g37		1	
P25-g38	0.5		0.5
P25-g39	0.5		0.5
P25-g40		1	
P25-g41		1	

Allele count	33	45	4
Total alleles	82		
Allele freq	0.402439	0.54878	0.04878

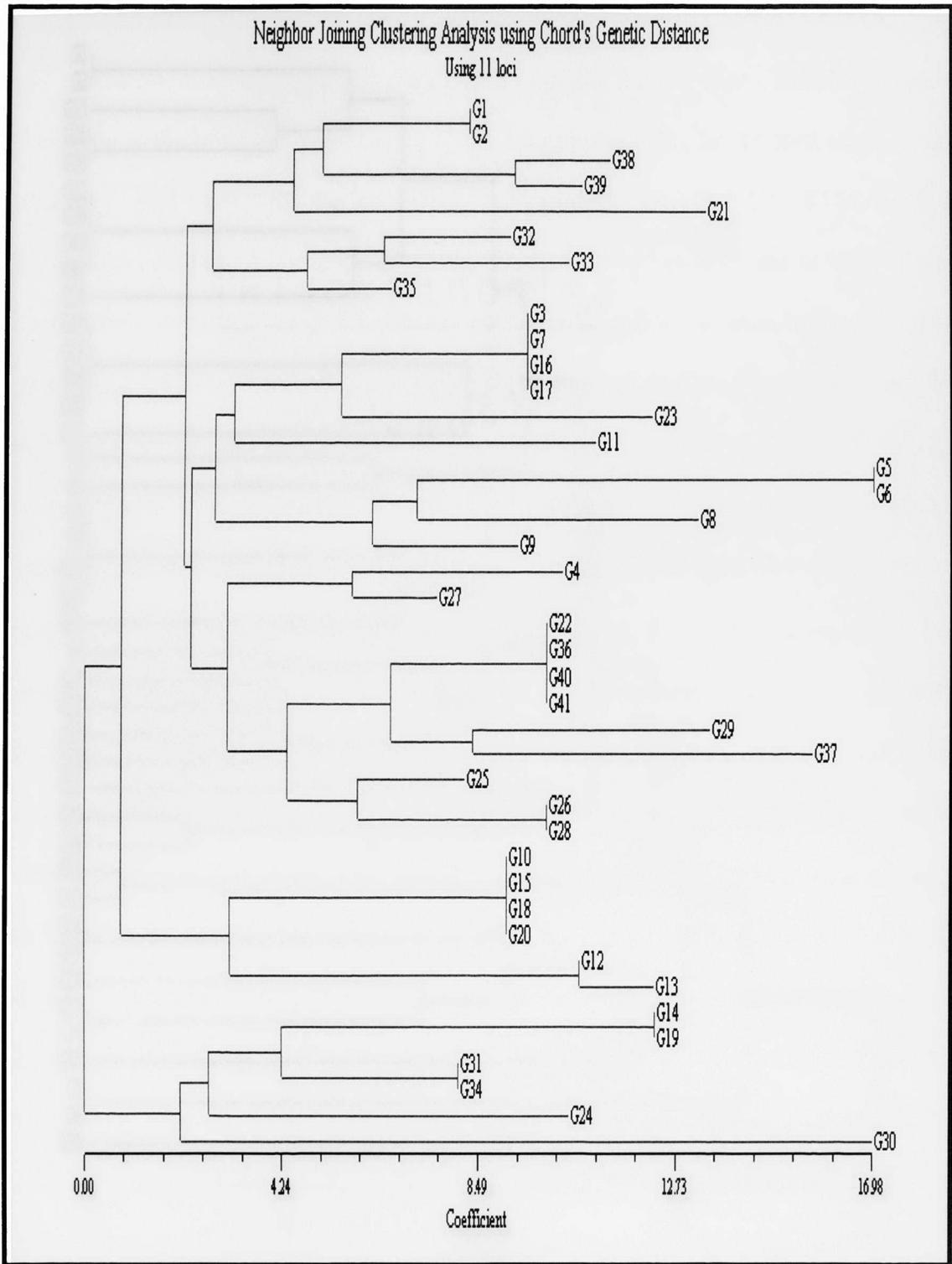
Table (15): Genotyping using P25 primer set, allele count, and allele frequencies

Genetic identifications and relationships:

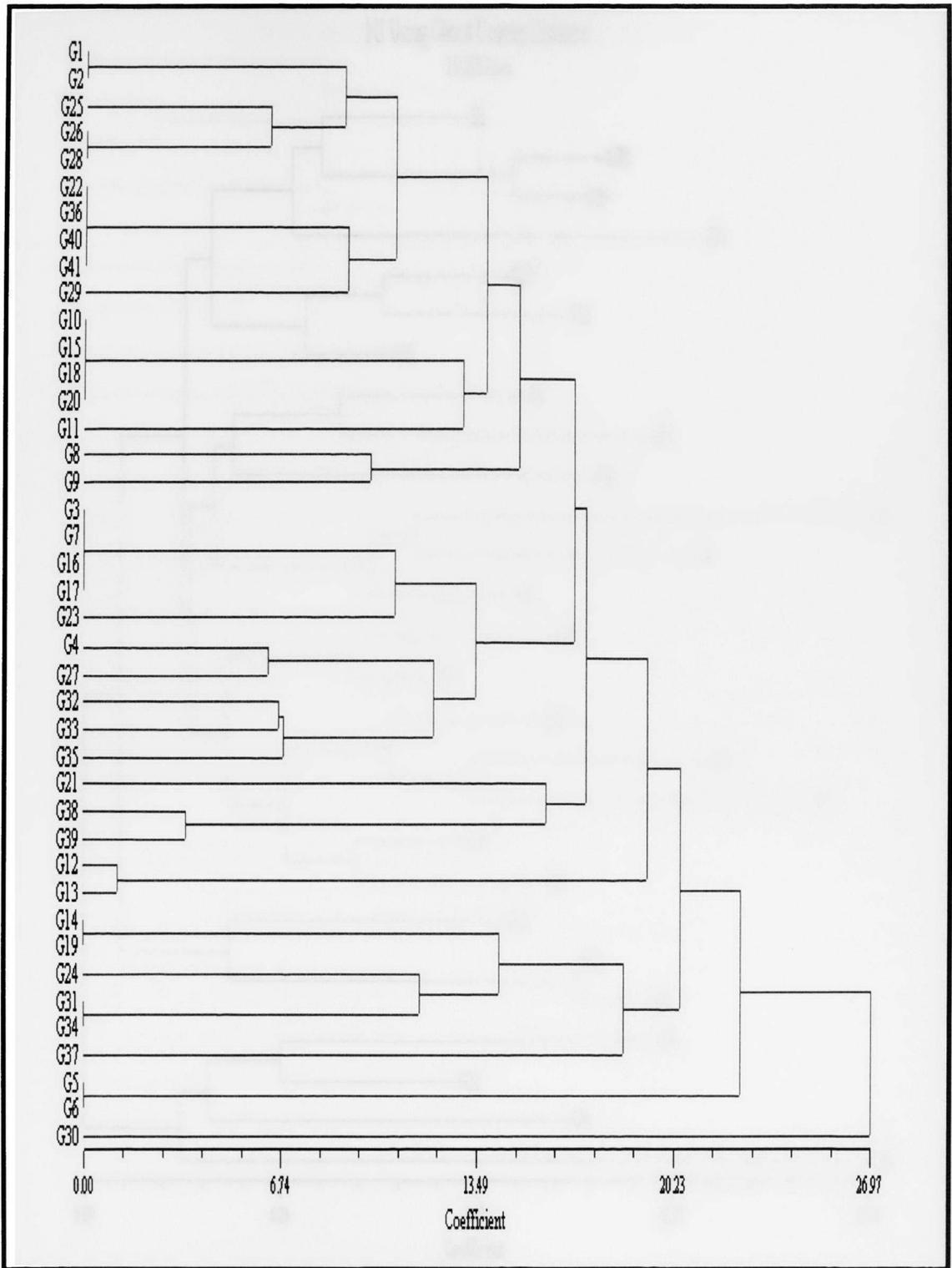
The microsatellite data was very useful to detect the identical *Cannabis* samples. According to the blind testing of the 41 *Cannabis* samples using the 11 SSR markers, the following sets had the same DNA fingerprint: G1/G2, G10/G15/G18/G20, G22/G36/G40/G41, G26/G28, G3/G7/G16/G17, G14/G19, G31/G34, and G5/G6. These identical sets were easily identified based on Chord's genetic distance by both NJ tree (tree 1) and UPGMA tree (tree 2) in which there were no genetic distance between the sets.

NJ and UPGMA methods were used to generate dendrograms using different distance coefficients to give different views on the genetic relationships among the 27 unique genotypes of *Cannabis* (tree 3-6). The NJ and UPGMA dendrograms were constructed using Chord's genetic distance (tree 3; 4), unbiased-Nei (tree 5), and with modified Roger (tree 6). Bootstrapping with 2000 replications was performed using the NJ method based on Chord's and DNA shared alleles genetic distance (tree 7 and 8 respectively). The only three clusters that could be considered supported were between the following: G12 and G13 with bootstrap values of 100%, G38 and G39 with bootstrap value of 85%, and G4 and G27 with bootstrap value of 55% (tree 8).

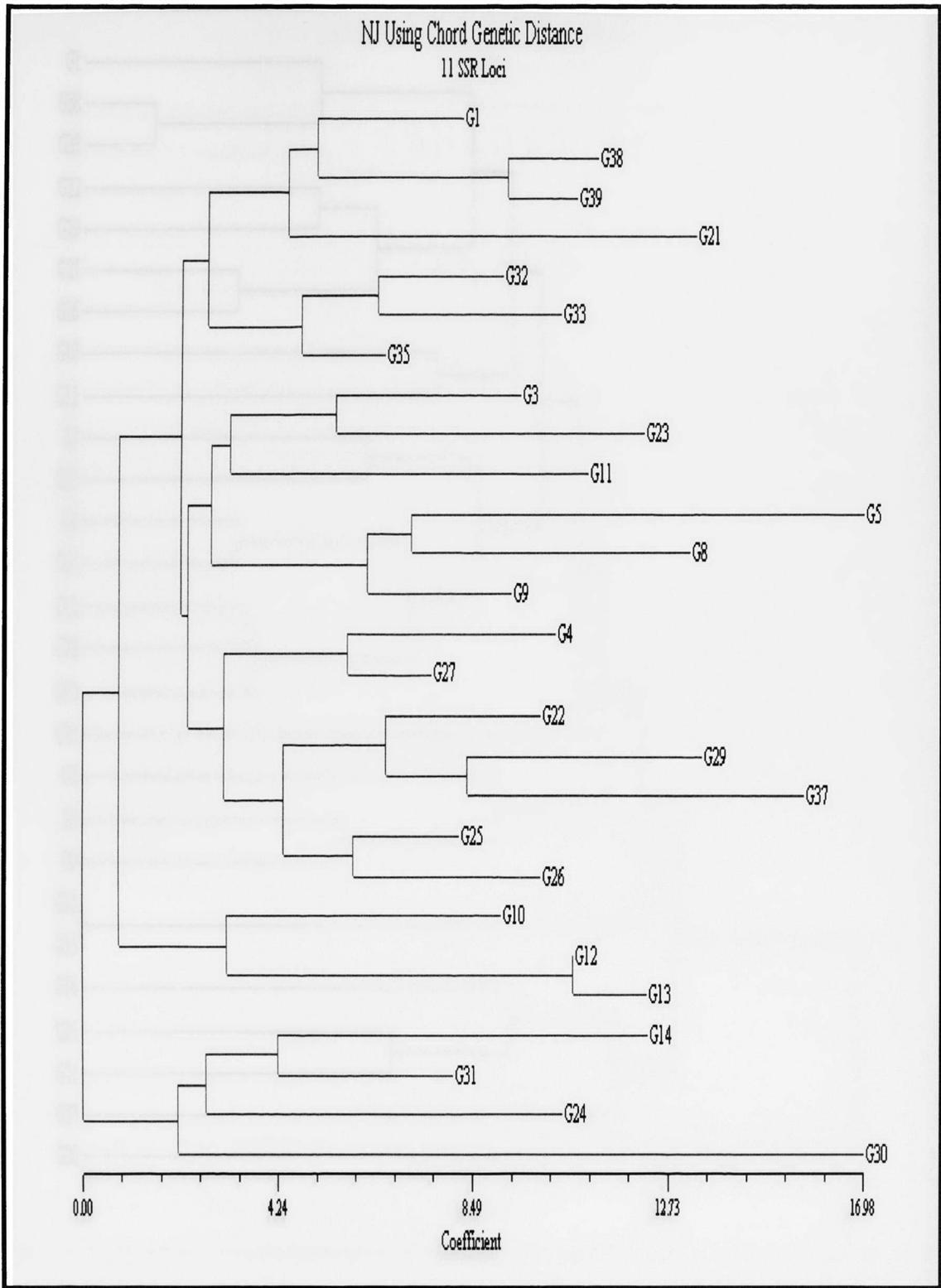
Principal Component Analysis (PCA) was performed on all 27 unique *Cannabis* genotypes in which 34.6% total variation was captured using three coordinates (figure 5). G30 had very distinct profile and could be easily recognized as outlier sample in PCA. After G30 was disregarded from the PCA testing, the remaining 26 genotypes became more widely dispersed on the PCA scatter plot in which 33.8% total variation was captured in the three coordinates (figure 6).



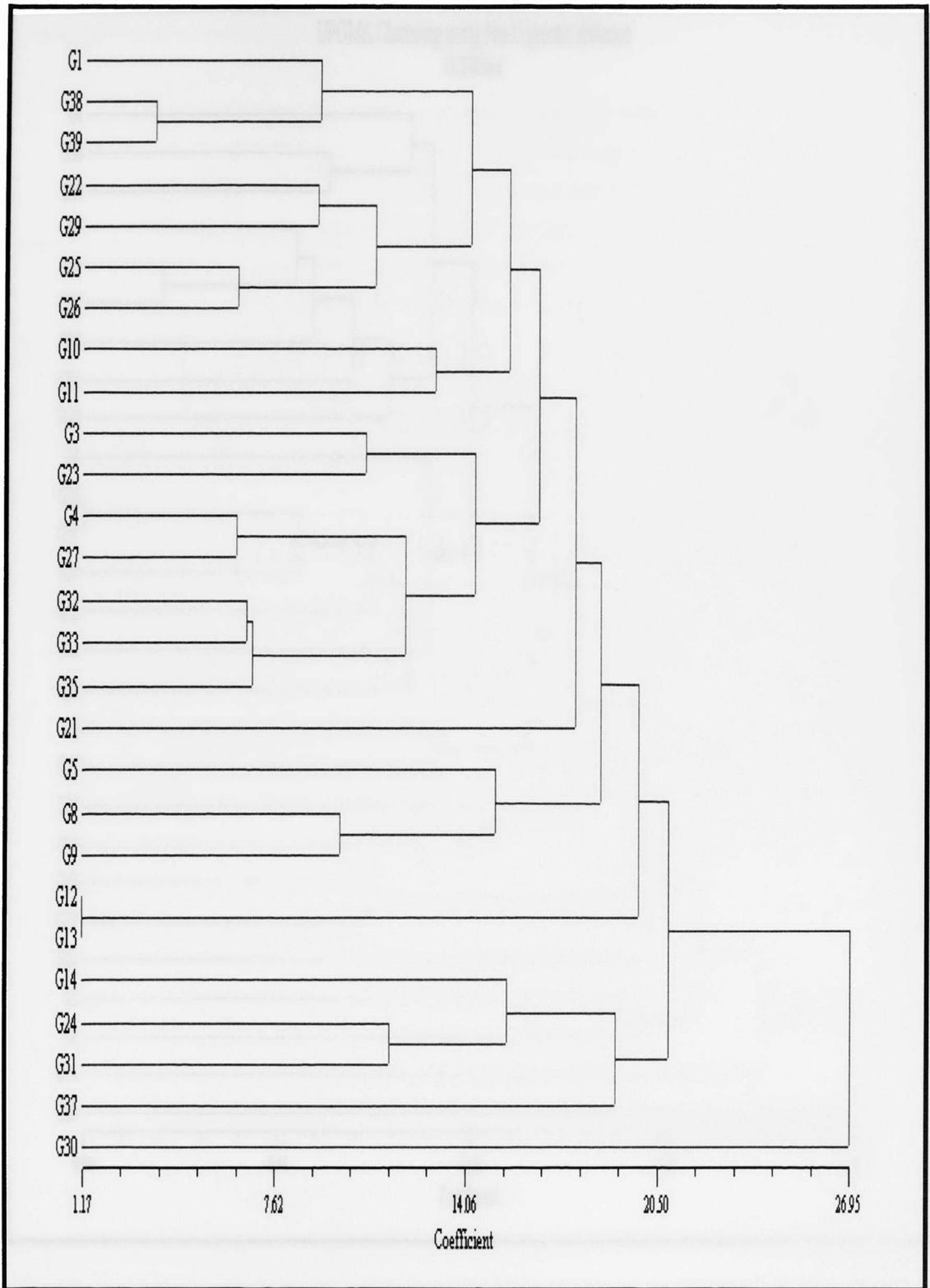
Tree (1): NJ tree based on Chord's genetic distance across the 41 *Cannabis* plants



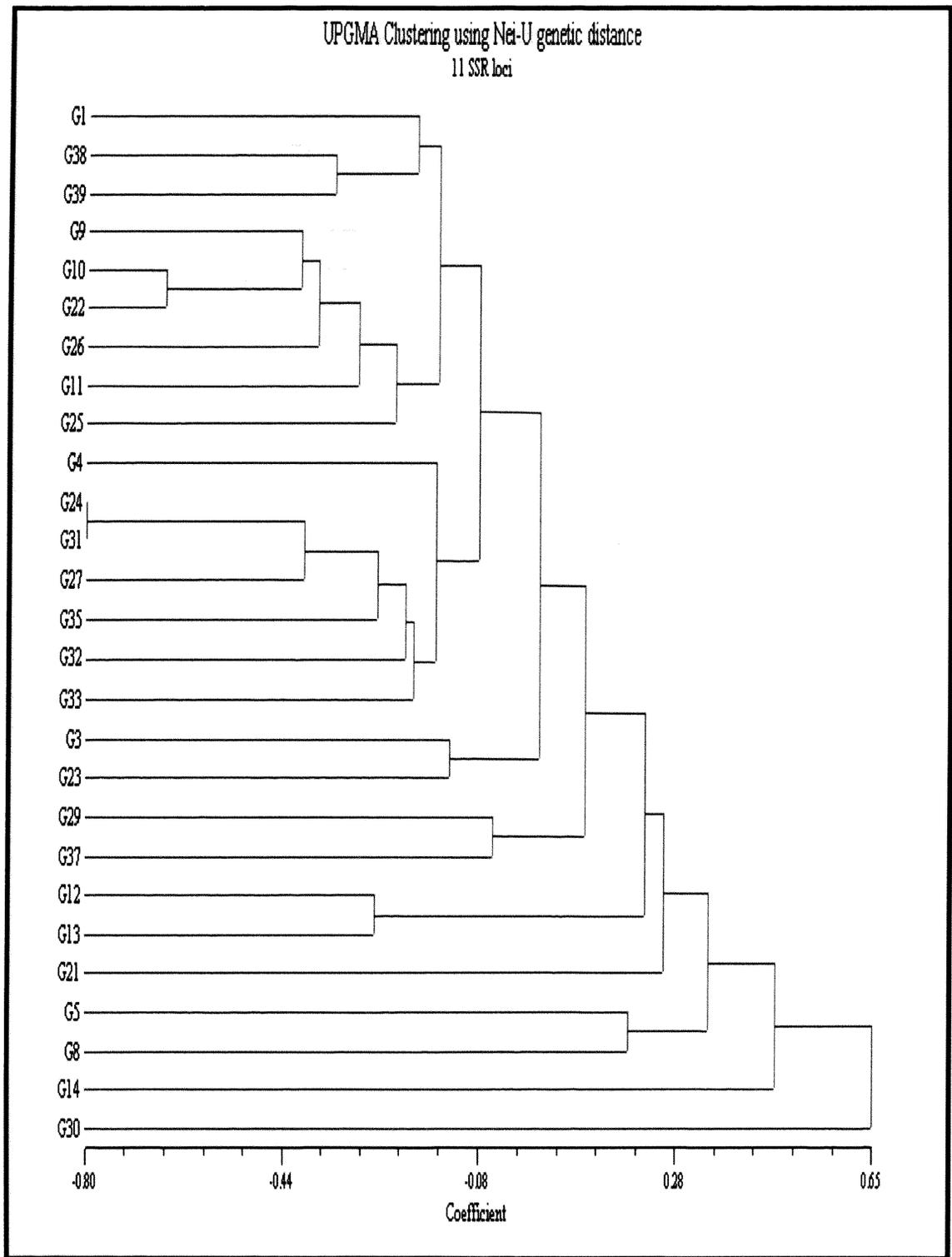
Tree (2): UPGMA tree based on Chord's genetic distance across the 41 *Cannabis* plants



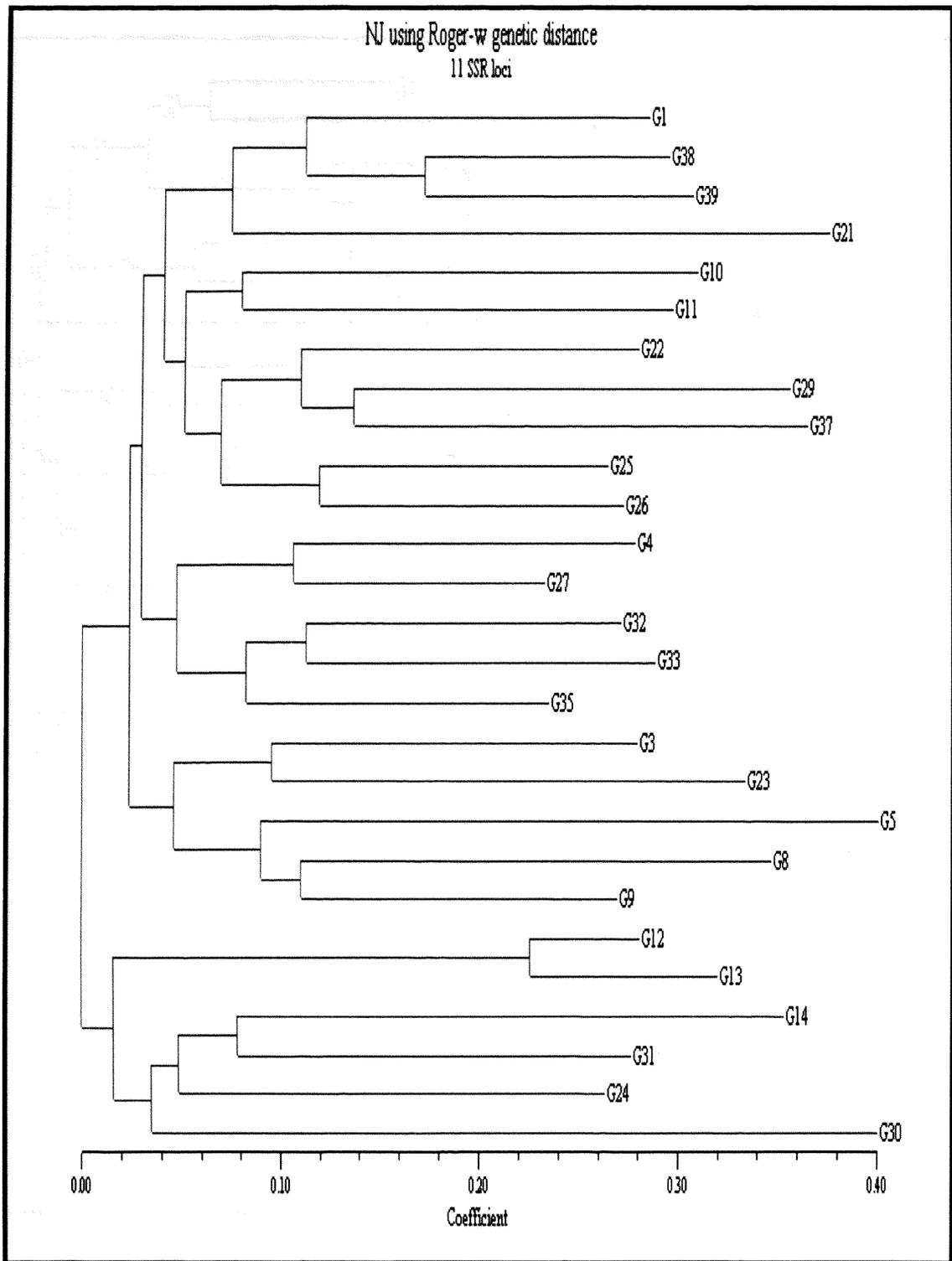
Tree (3): NJ tree based on Chord's genetic distance across 27 different genotypes



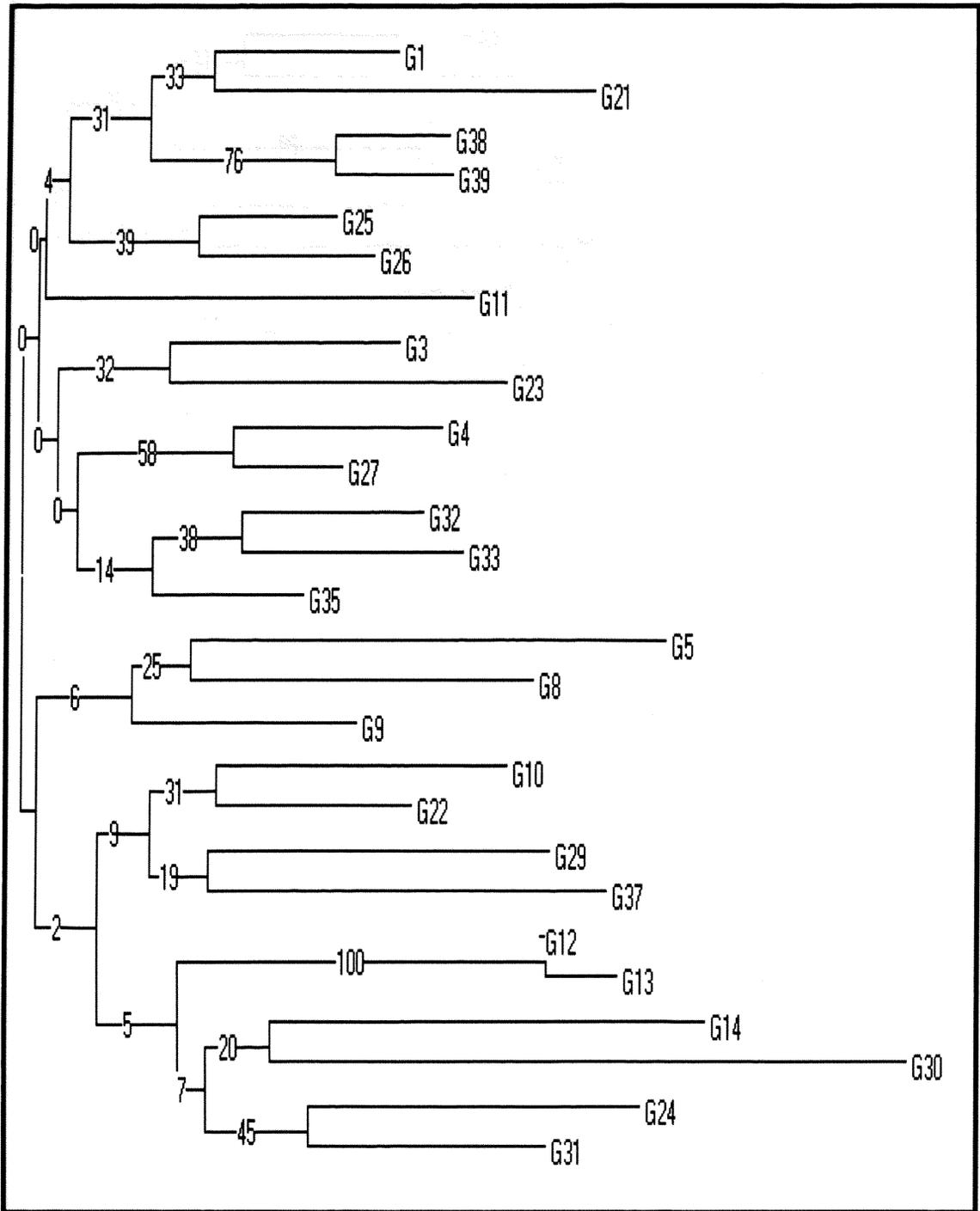
Tree (4): UPGMA tree based on Chord's genetic distance across 27 different genotypes



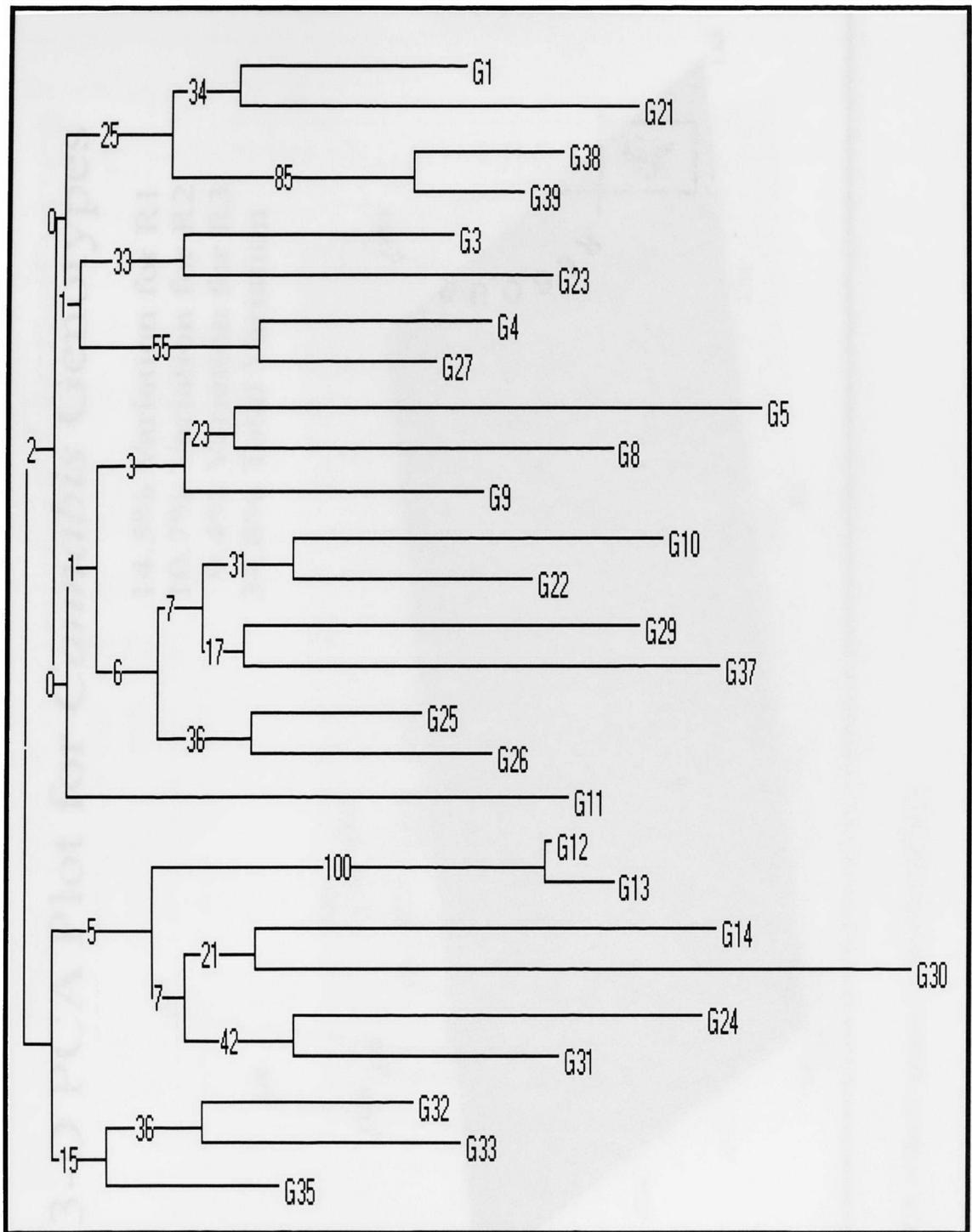
Tree (5): UPGMA tree based on Nei-unbiased genetic distance across 27 different genotypes



Tree (6): NJ tree based on modified Roger genetic distance across 27 different genotypes



Tree (7): NJ plus bootstrap values based on DNA shared Alleles (DSA) genetic distance



Tree (8): NJ plus bootstrap values based Chord's genetic distance

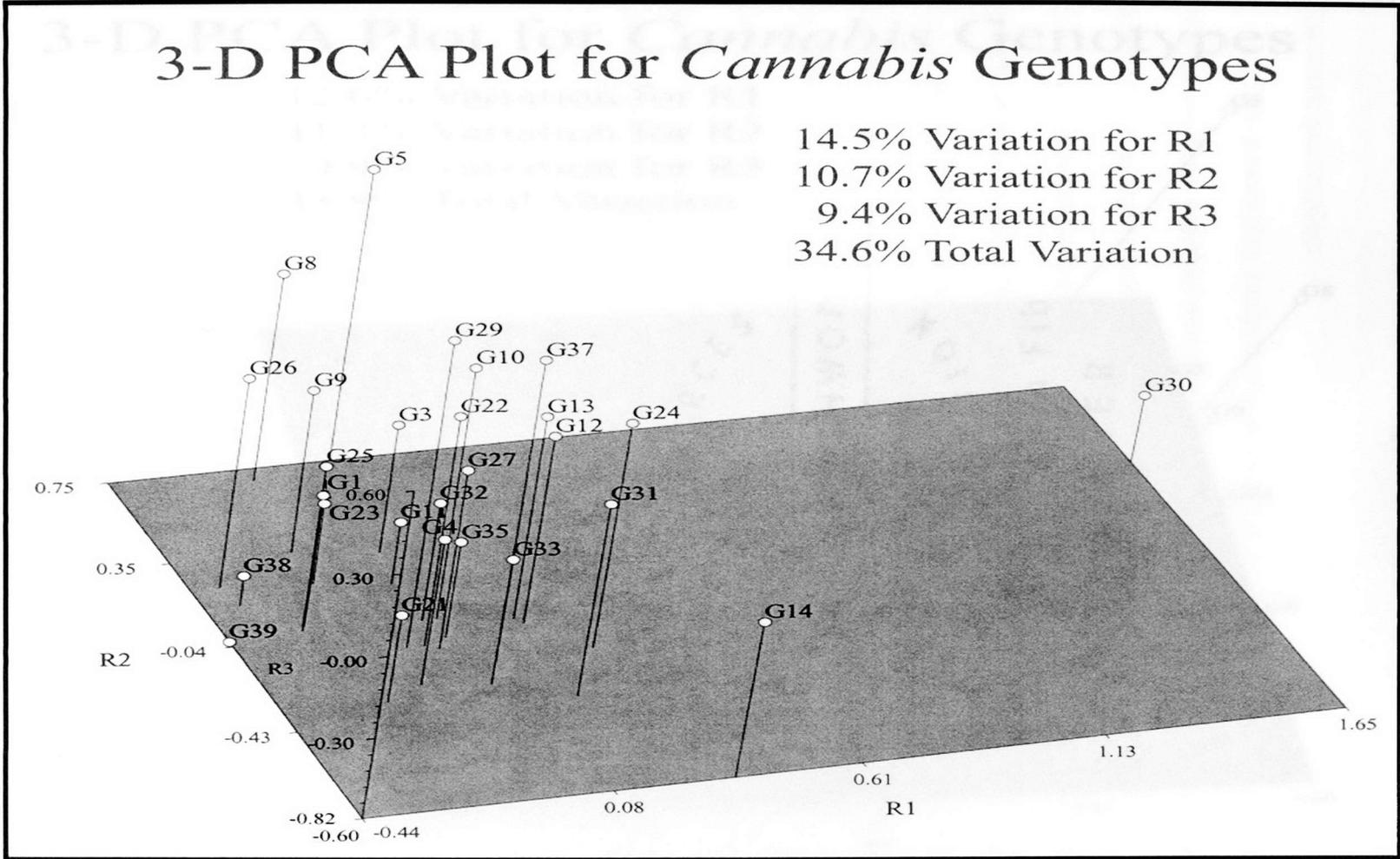


Figure (5): PCA of the 27 unique *Cannabis* genotypes

3-D PCA Plot for *Cannabis* Genotypes

12.6% Variation for R1
11.3% Variation for R2
9.9% Variation for R3
33.8% Total Variation

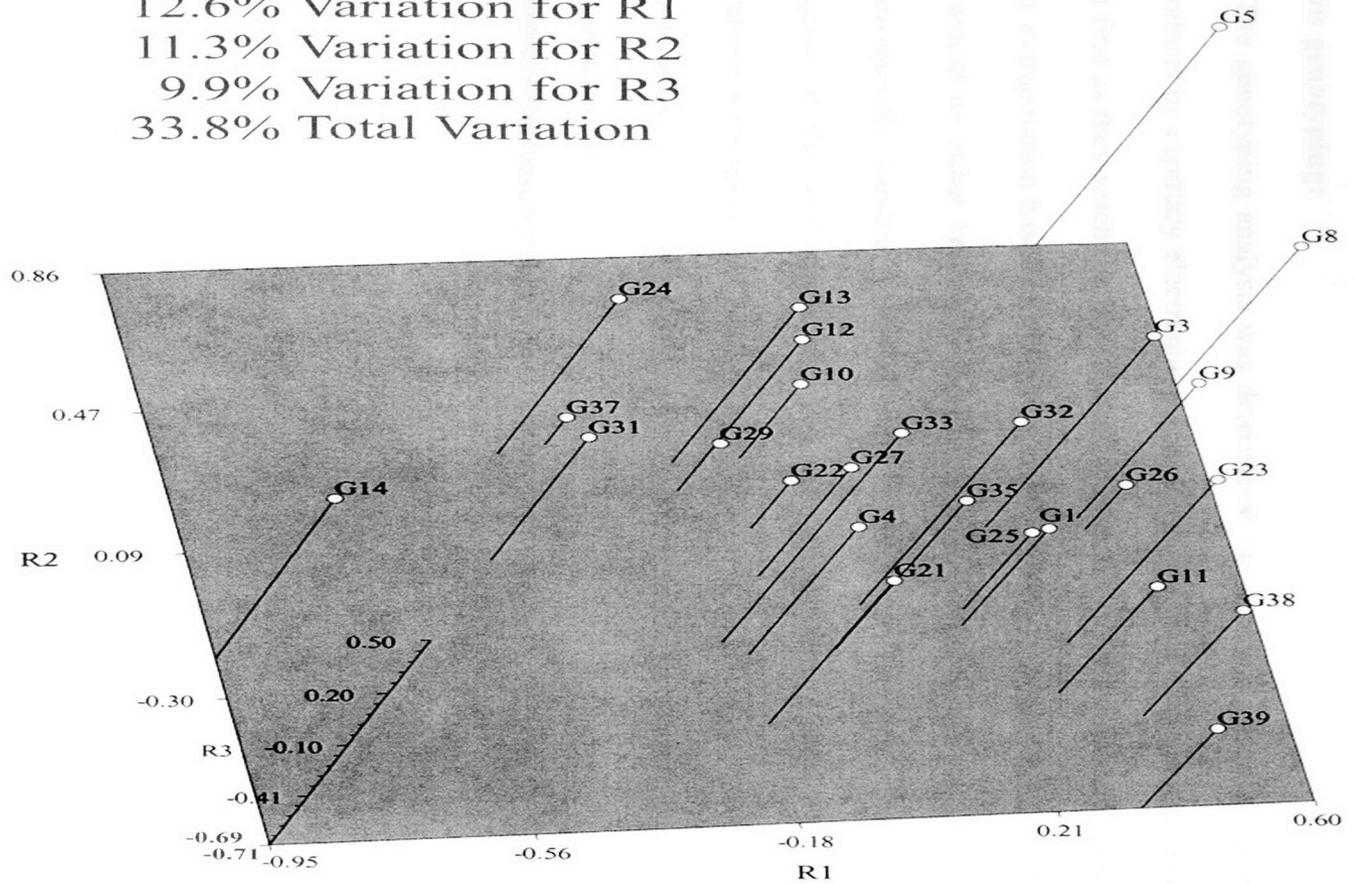


Figure (6): PCA of 26 different *Cannabis* genotypes (removing outlier G30)

Discussion

***Cannabis* genotyping:**

The genotyping analysis was determined by separating the fluorescently labeled PCR products by capillary electrophoresis using ABI 3100. The smaller fragments were detected first as they reach the laser-scanning detector. Then, the allele sizes were called by linear extrapolation based on the internal size standard. At each locus, the individual sample would be either homozygous (figure 7) or heterozygous (figure 8). However, some non-specific banding patterns were occasionally detected that did not affect interpretation of the SSR results as they fell outside the expected size range. Some considerations were taken into account when determining the allele sizes or in comparing allele sizes called with different labeled primer methods (M13-tail fluorescent label and direct fluorescent label).

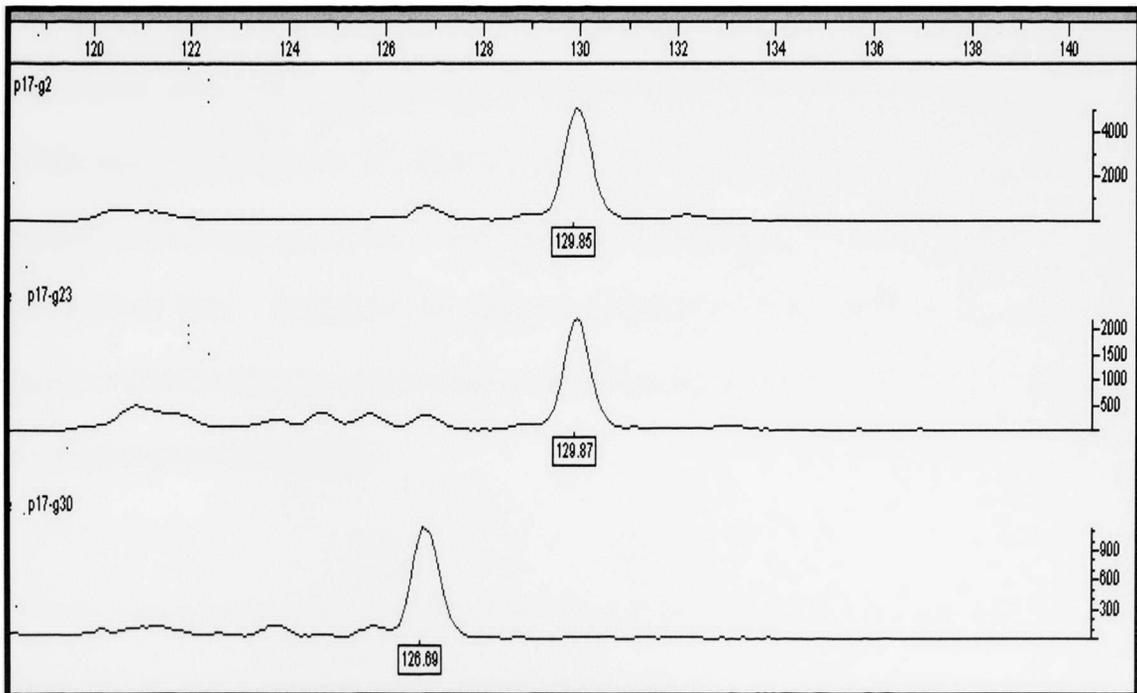


Figure (7): Homozygous profiles using P17

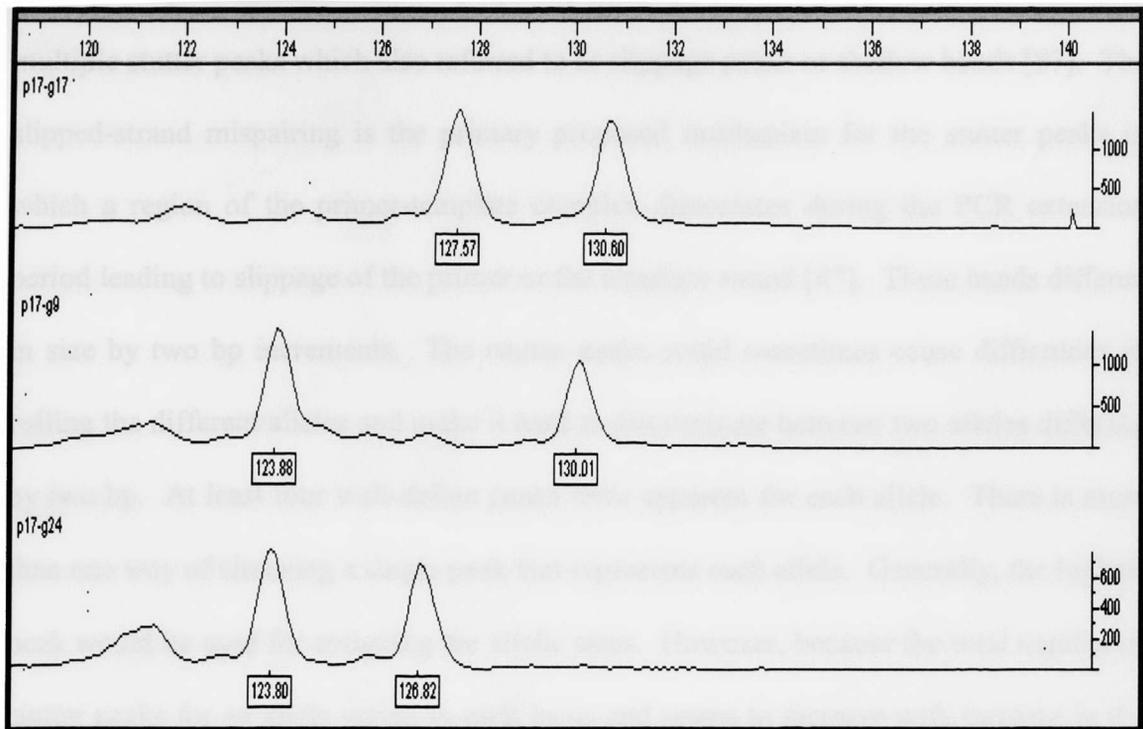


Figure (8): Heterozygous profiles using P25

First, the dinucleotide loci (P9, P13, and P15) composed of alleles that exhibited multiple stutter peaks which also referred to as slippage peaks or shadow bands [87]. The slipped-strand mispairing is the primary proposed mechanism for the stutter peaks in which a region of the primer-template complex dissociates during the PCR extension period leading to slippage of the primer or the template strand [87]. These bands differed in size by two bp increments. The stutter peaks could sometimes cause difficulties in calling the different alleles and make it hard to discriminate between two alleles differing by two bp. At least four well-define peaks were apparent for each allele. There is more than one way of choosing a single peak that represents each allele. Generally, the highest peak would be used for assigning the allelic sizes. However, because the total number of stutter peaks for an allele varied in each locus and seems to increase with increase in the number of repeat motifs, the highest peak at each locus may varied as well. In locus P13, for example, the second right-most peaks were observed to have the highest relative fluorescent units (rfu) amongst (figure 9) and therefore were chosen to designate the allelic sizes at such locus. The highest peak may or may not code for the true size of the template but in this case it did. These multiple stutter peaks were observed only in the dinucleotide loci. To reduce the amount of stutter products, someone could use SSR markers that have longer repeat units, imperfect repeat markers, and DNA polymerases with faster processivity [55].

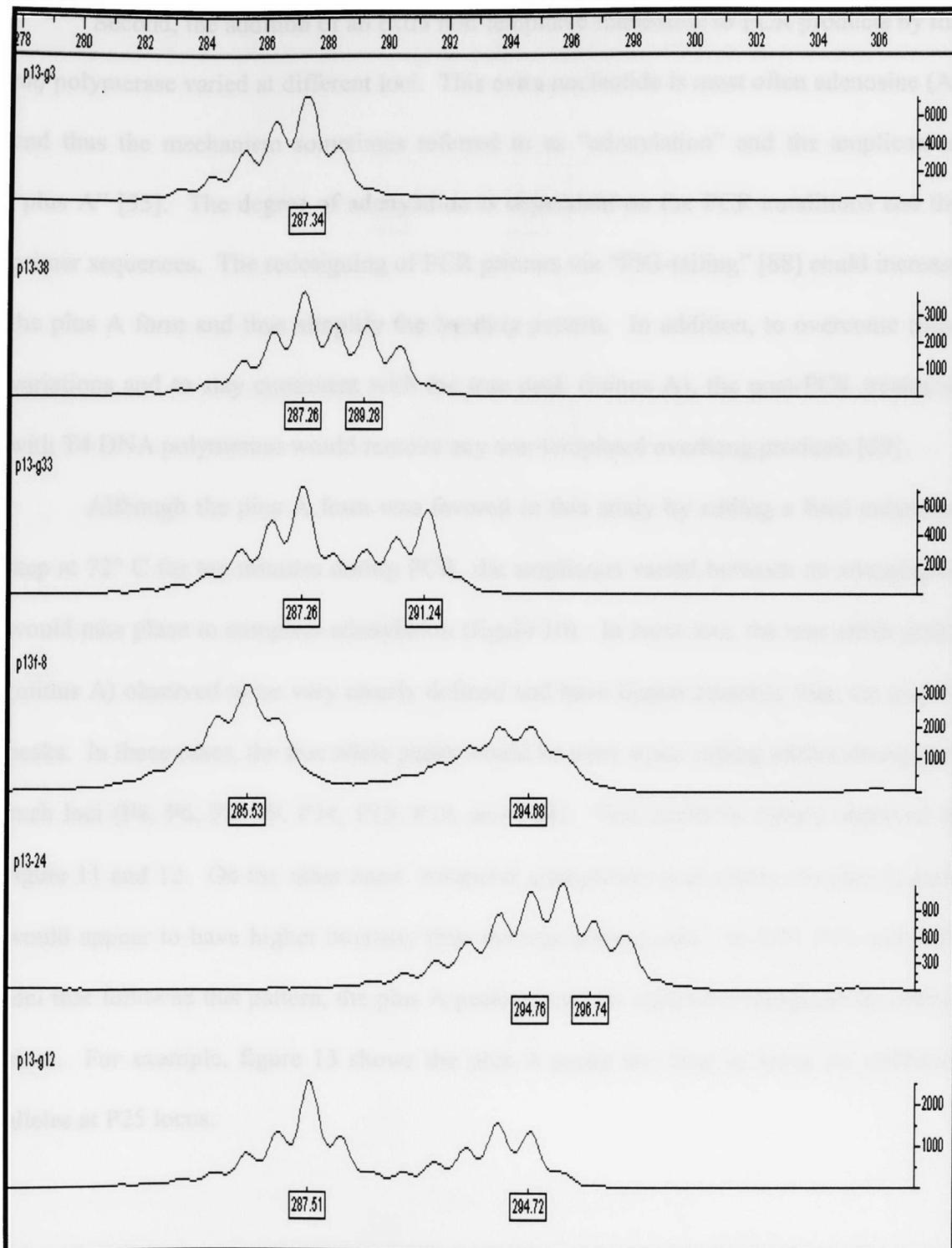


Figure (9): Genotyping of *C. sativa* samples using P13 and the stutter peaks

Second, the addition of an extra non-templated nucleotide to PCR products by the *taq* polymerase varied at different loci. This extra nucleotide is most often adenosine (A) and thus the mechanism sometimes referred to as “adenylation” and the amplicon as “plus A” [55]. The degree of adenylation is dependent on the PCR conditions and the primer sequences. The redesigning of PCR primers via “PIG-tailing” [88] could increase the plus A form and thus simplify the banding pattern. In addition, to overcome these variations and to stay consistent with the true peak (minus A), the post-PCR treatment with T4 DNA polymerase would remove any non-templated overhang products [89].

Although the plus A form was favored in this study by adding a final extension step at 72° C for ten minutes during PCR, the amplicons varied between no adenylation would take place to complete adenylation (figure 10). In most loci, the true allele peaks (minus A) observed to be very clearly defined and have higher intensity than the plus A peaks. In these cases, the true allele peaks would be used when calling alleles throughout such loci (P4, P6, P7, P9, P14, P13, P19, and P24). This could be clearly observed in figure 11 and 12. On the other hand, whenever adenylation was active, the plus A peak would appear to have higher intensity than the true allele peaks. In P15, P17, and P25 loci that followed this pattern, the plus A peaks would be utilized to designate the allelic sizes. For example, figure 13 shows the plus A peaks are used to score the different alleles at P25 locus.

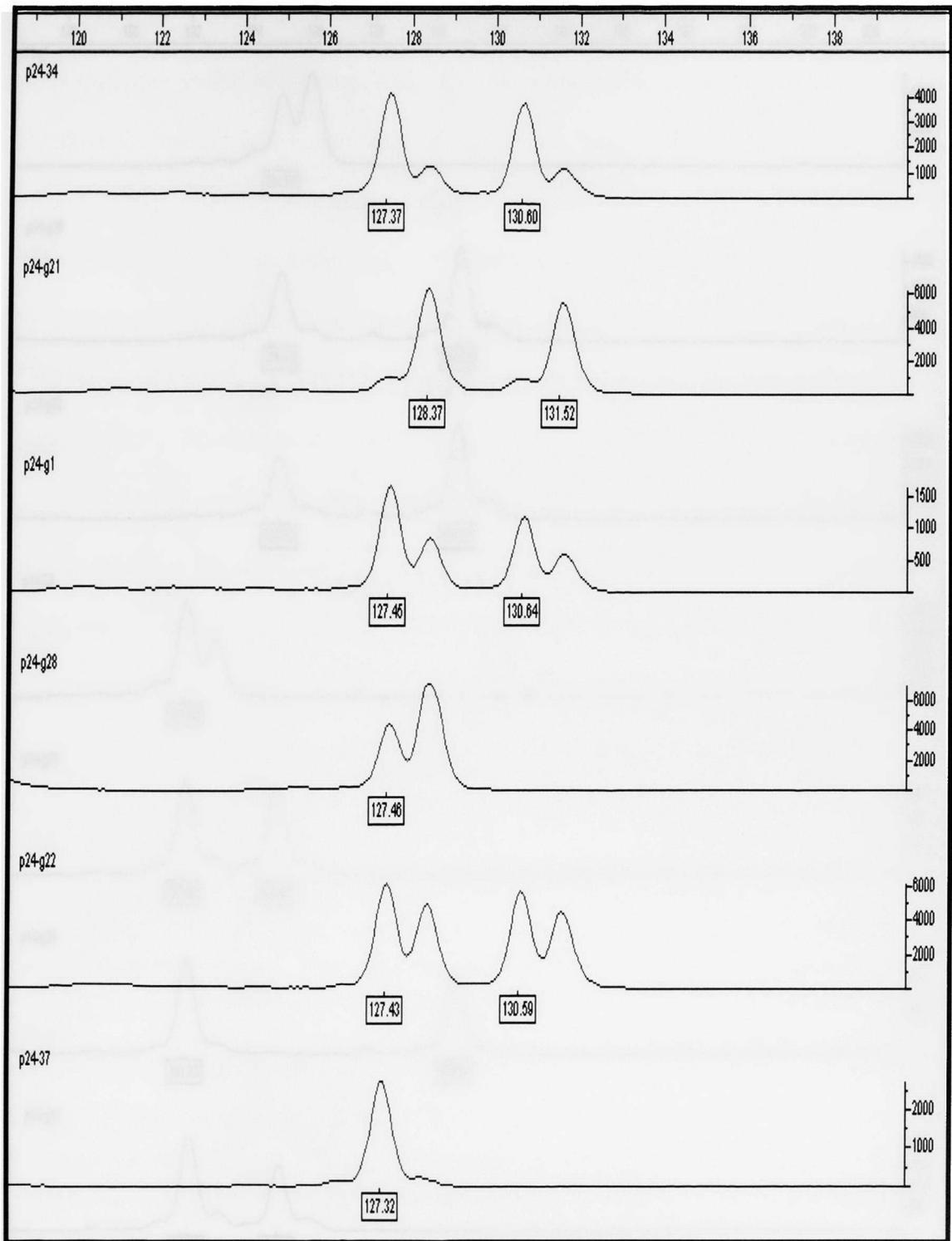


Figure (10): Mixed preferential amplification of plus A and true allele in P24 locus

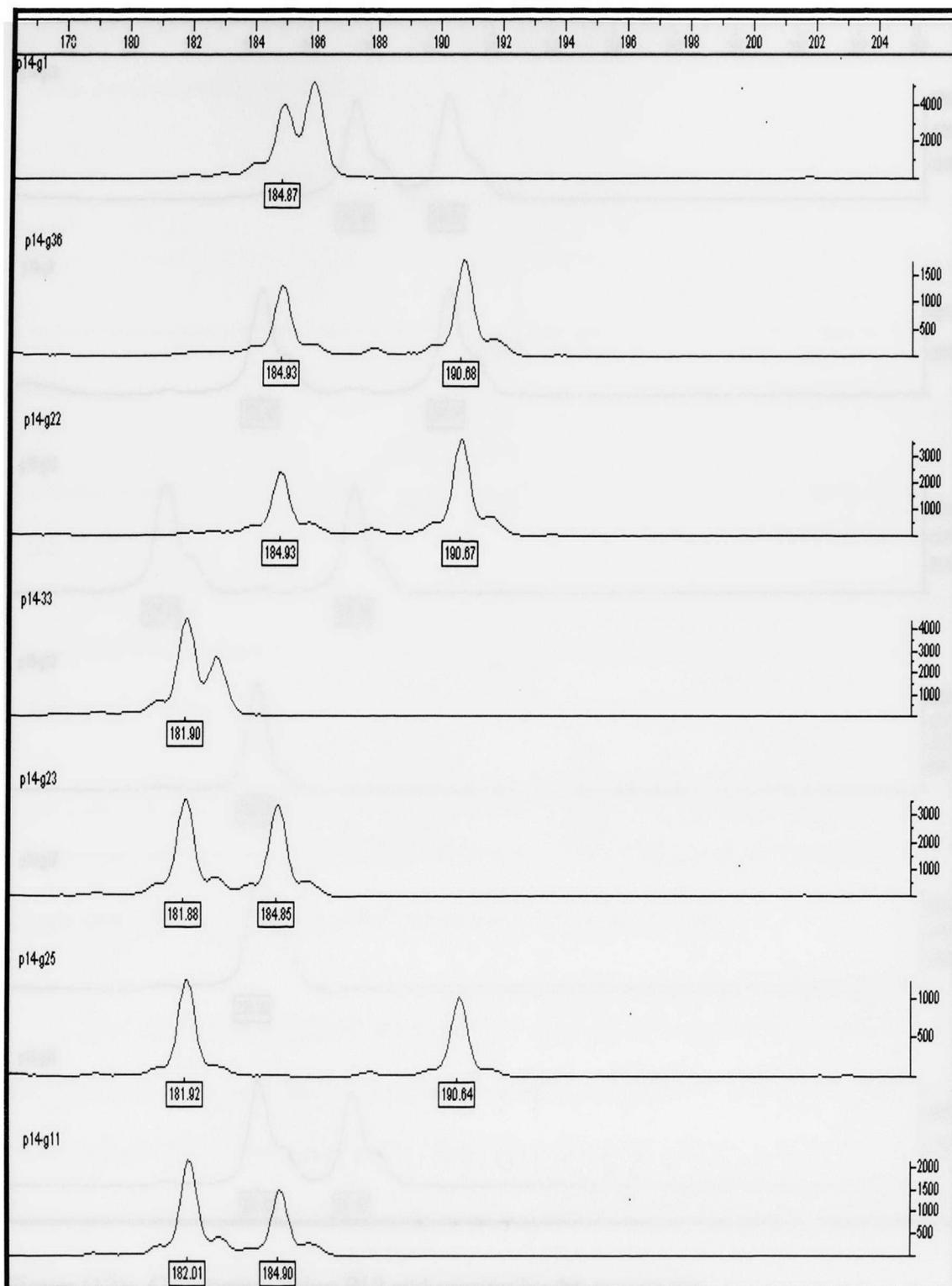


Figure (11): Genotyping using P14 and scoring by the true peaks

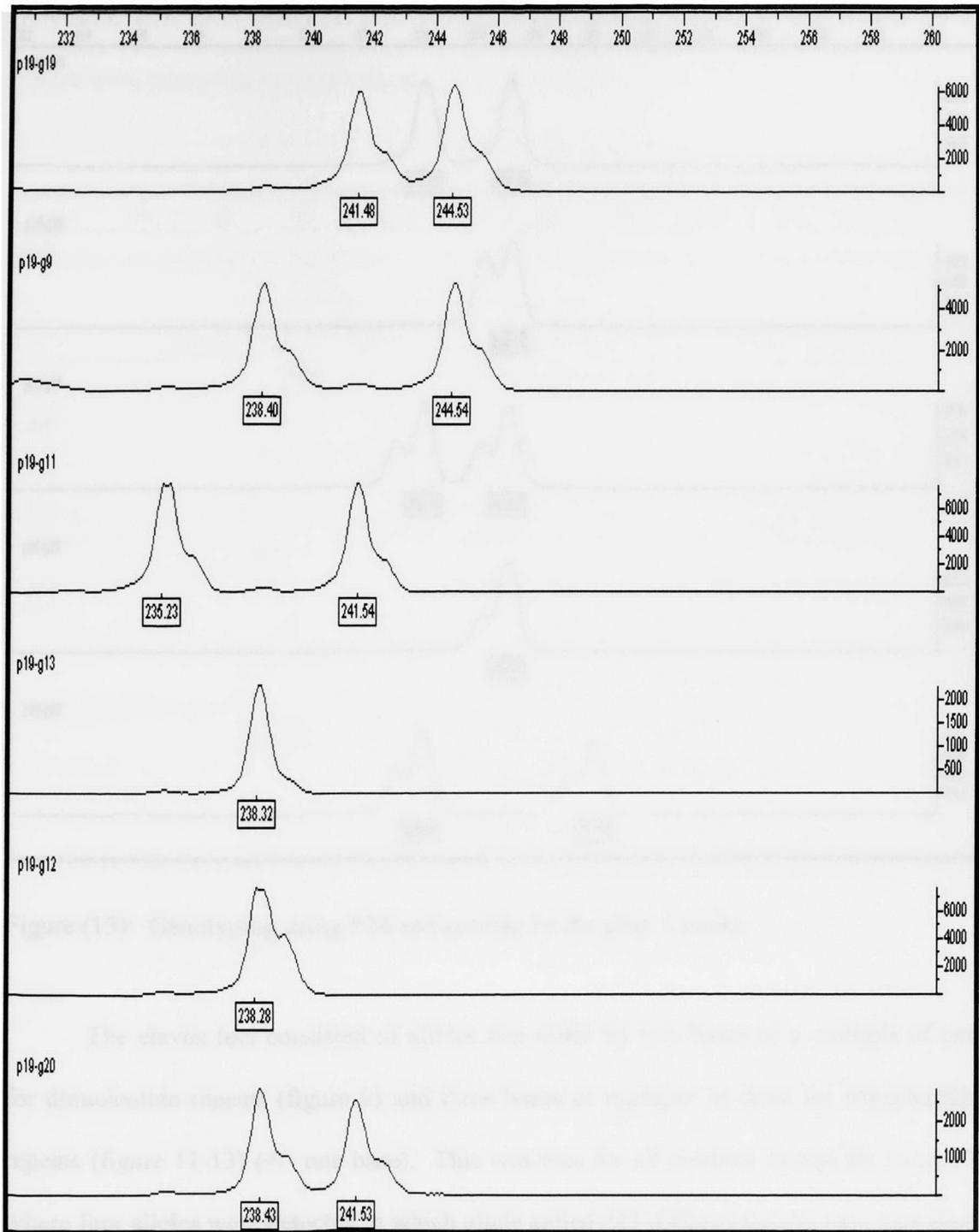


Figure (12): Genotyping using P19 and scoring by the true peaks

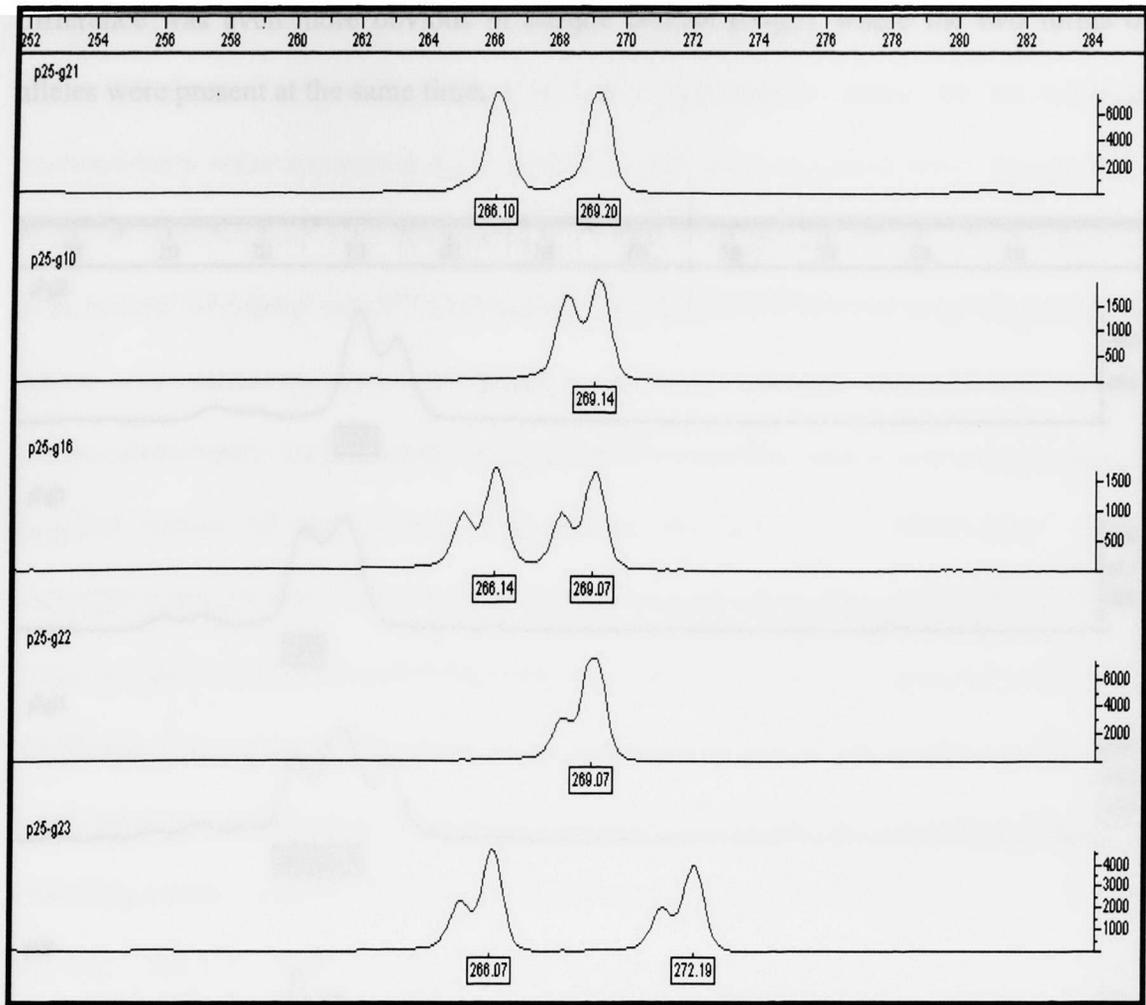


Figure (13): Genotyping using P24 and scoring by the plus A peaks

The eleven loci consisted of alleles that differ by two bases or a multiple of two for dinucleotide repeats (figure 9) and three bases or multiple of three for trinucleotide repeats (figure 11-13) (+/- one base). This was true for all markers except for locus P7 where four alleles were detected in which allele called 221 differed by one base pair from allele called 222. Allele 221 was a replica of allele 222 with one base insertion that was repeatedly observed and clearly distinguishable as it illustrated in figure 14. This

difference was even more obvious in sample labeled p7-g11 where the two forms of alleles were present at the same time.

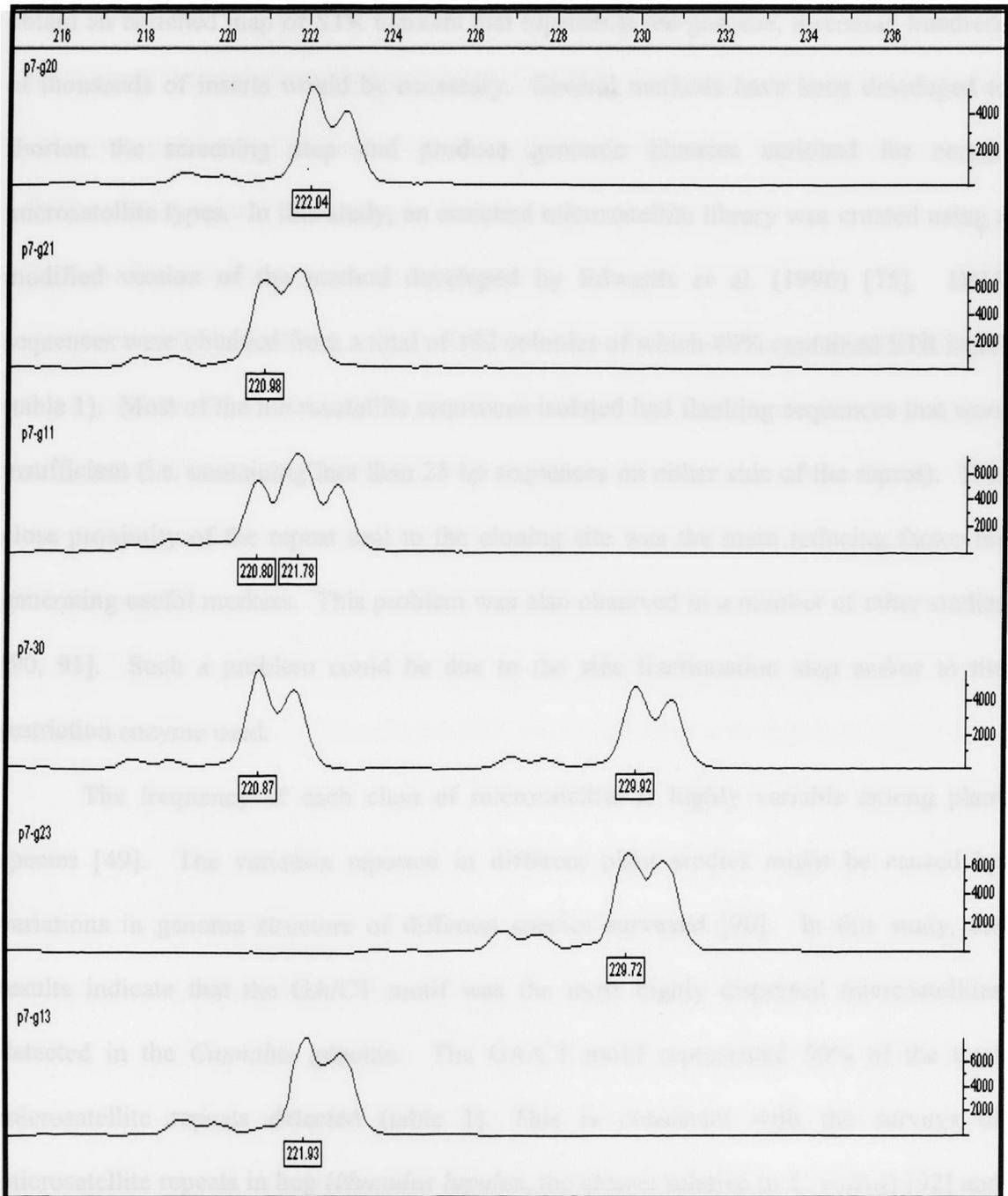


Figure (14): Genotyping using P7 and the special case

Microsatellites in *Cannabis*:

In plants, on average, there is one microsatellite every 33 kb whereas microsatellites occur approximately every 6 kb in the human genome [49]. Therefore, to obtain an enriched map of STR markers that represents the genome, screening hundreds of thousands of inserts would be necessary. Several methods have been developed to shorten the screening step and produce genomic libraries enriched for certain microsatellite types. In this study, an enriched microsatellite library was created using a modified version of the method developed by Edwards et al. (1996) [75]. DNA sequences were obtained from a total of 192 colonies of which 49% contained STR insert (table 1). Most of the microsatellite sequences isolated had flanking sequences that were insufficient (i.e. containing less than 25 bp sequences on either side of the repeat). This close proximity of the repeat unit to the cloning site was the main reducing factor for generating useful markers. This problem was also observed in a number of other studies [90, 91]. Such a problem could be due to the size fractionation step and/or to the restriction enzyme used.

The frequency of each class of microsatellite is highly variable among plant species [49]. The variation reported in different plant studies might be caused by variations in genome structure of different species surveyed [90]. In this study, the results indicate that the GA/CT motif was the most highly dispersed microsatellites detected in the *Cannabis* genome. The GA/CT motif represented 50% of the total microsatellite repeats detected (table 1). This is consistent with the surveys of microsatellite repeats in hop (*Humulus lupulus*, the closest relative to *C. sativa*) [92] and other studies [52, 63] in which GA/CT motif was the most abundant. However, other

reports of microsatellite repeats in mangrove [60] and potato [66] revealed a greater abundance of AC/TG motif over GA/CT motif. The number of trinucleotide repeats detected was similar to that of the dinucleotide repeats (table 1). Trinucleotide profiles are easier to score than the dinucleotide profiles because the variations in the number of core units of trinucleotide motifs are larger in length. In addition, the trinucleotide loci show more distinct allelic profiles by avoiding the stutter pattern that is often associated with the amplification of dinucleotide loci motifs [47, 55, 93].

Microsatellite polymorphism and applications:

Different RAPD, AFLP, RFLP, and SSR studies revealed various estimates for the genetic polymorphism, most of them were below or near the level of polymorphism obtained by SSR markers [62, 94, 95]. Microsatellite markers have shown high levels of polymorphism in many plant species including rice [96], tropical trees [97], Hop [95], oat [98], maize [99], wheat [100], sunflower [101], and many more. To discriminate between different cultivars of *C. sativa* and to potentially associate the samples of clonal origin using microsatellites, the selected STR loci would need to be highly polymorphic.

The discrimination power, ease of genotyping, and high reproducibility (figure 15, 16) emphasize that these eleven microsatellite markers are very useful for DNA fingerprinting in *Cannabis*. The blind test of microsatellite typing for the 41 samples matched the identities of the duplicates and the unique samples reported afterward by AFLP typing (table 3). G26/G28, G31/G34, and G22/G36/G40 were found to be identical by AFLP (table 3) and by the SSR markers (tree 1; 2), which strongly suggest that they could be clonally propagated. The remaining identical sets consisted of known duplicate samples (table 3), showing reproducibility of this typing technique. In addition,

the 11 SSR loci were very effective in uniquely identifying 27 profiles of the *Cannabis* samples tested (tree 1; 2). Microsatellite DNA also had high resolution and sensitivity for calculating the various genetic parameters presented in table 4.

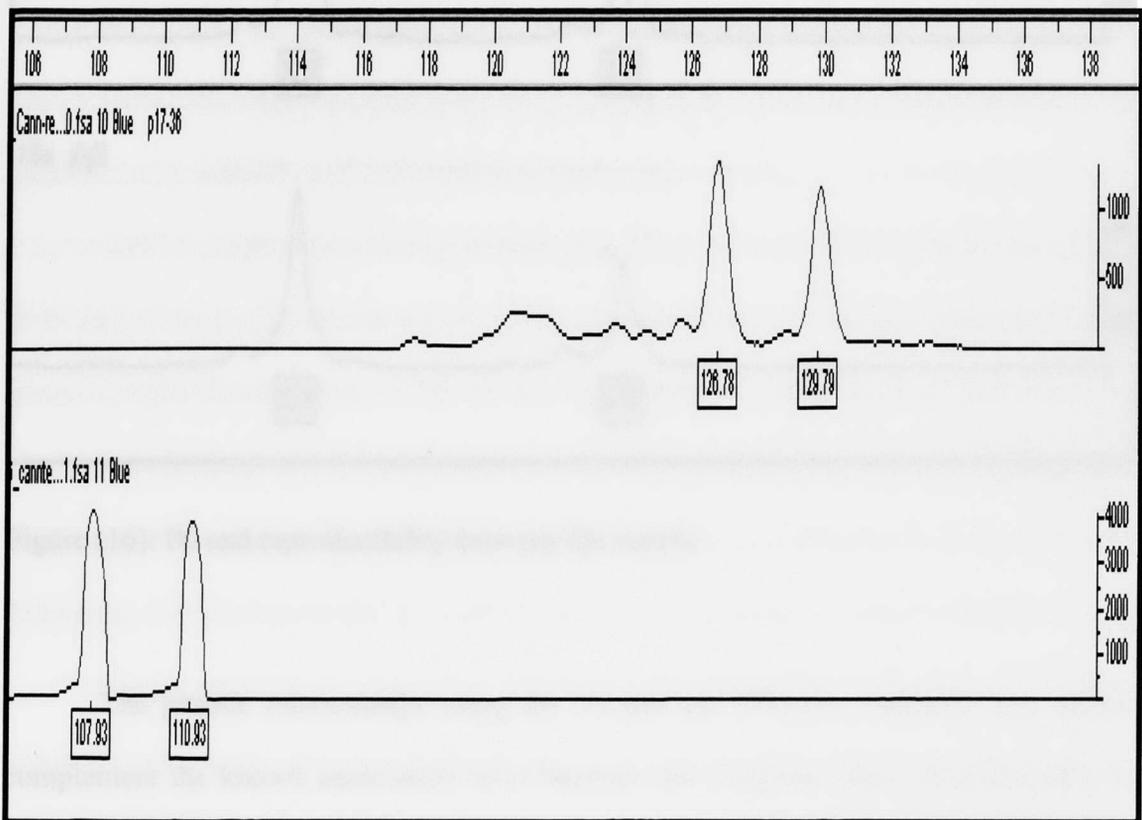


Figure (15): P17 and reproducibility between using the M13 fluorescent tail primer and the direct fluorescent primer

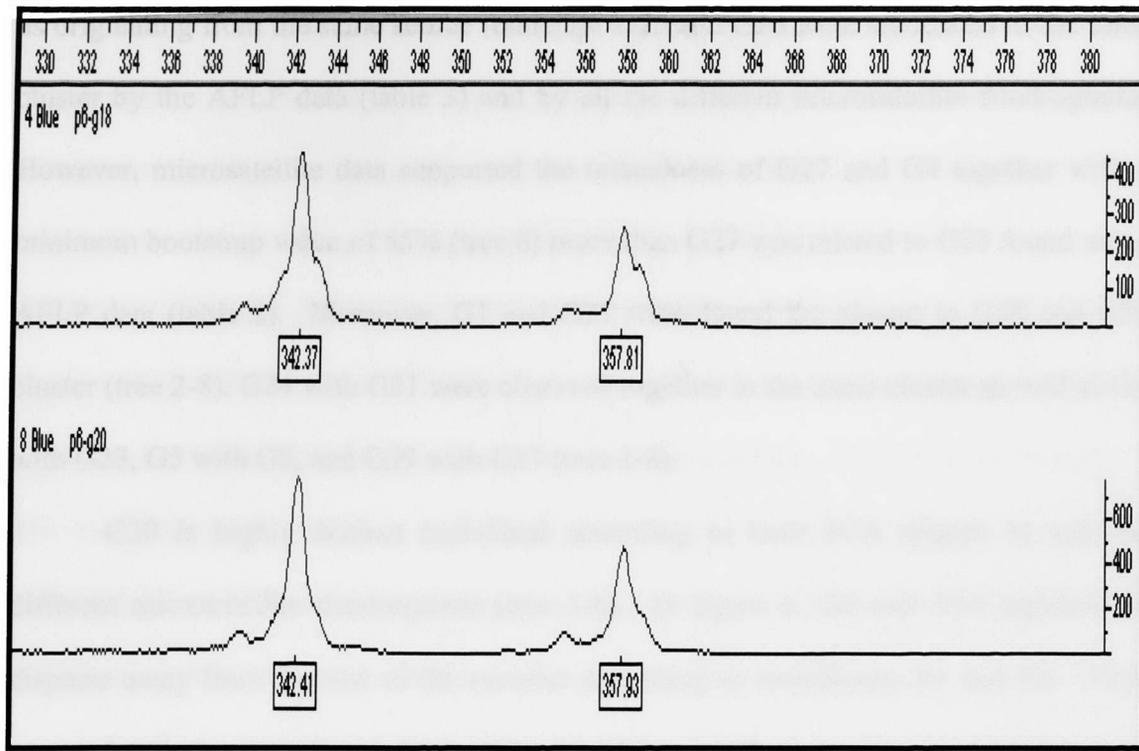


Figure (16): P6 and reproducibility between the results

The genetic relationships using the NJ and the UPGMA methods were able to complement the known association level between the *Cannabis* plant samples (table 3). First, samples G12 and G13 were found to have the strongest ties across all dendrograms (tree 1-8) supported by a bootstrap value of 100% (tree 7; 8). This relationship was in agreement with the information that both samples were closely related by AFLP analysis (table 3). In addition, all dendrograms were able to group G38 and G39 together with a minimum bootstrap value of 76% (tree 7). These results were consistent with the information that both samples came from the same grower (table 3). This information could link these samples to a specific source. Moreover, the seized samples G32, G33, and G35 clustered together in the microsatellite tree (tree 2-8) all of which were reported

as originating from the same source (table 3). G25 and G26 were associated in the same cluster by the AFLP data (table 3) and by all the different microsatellite dendrograms. However, microsatellite data supported the relatedness of G27 and G4 together with a minimum bootstrap value of 55% (tree 8) more than G27 was related to G23 found using AFLP data (table 3). Moreover, G1 and G21 were found the closest to G38 and G39 cluster (tree 2-8). G24 with G31 were observed together in the same cluster as well as G3 with G23, G5 with G8, and G29 with G37 (tree 2-8).

G30 is highly distinct individual according to both PCA (figure 5) and the different microsatellite dendrograms (tree 1-8). In figure 6, G5 and G14 appeared to disperse away from the rest of the samples according to coordinates R1 and R2. PCA scatterplot clustered G12 with G13, G33 with G32 and G35, and G39 with G38 (figure 6) which agreed with NJ and UPGMA analysis in most of grouping made (tree 2-8). Moreover, G29 was observed to be the closest to G37 by R3 coordinate (figure 6).

Many growers of *Cannabis* choose selective breeding to adapt to particular growing conditions and to increase the THC content, which contributes to the intoxicating effect of *Cannabis* [74]. Selective breeding of high THC content plants can be obtained using two methods. The first technique is clonal propagation using stem cuttings from clones containing high levels of THC. In this case, both the mother plant and subsequent cuttings (daughter clones) will have identical DNA profiles. Thus, a DNA analyst can easily detect plant materials that originated from the same source. Second, two plants can be cross-pollinated to generate seeds. After that, each seed can be grown into a plant that has its own unique DNA profile.

The eleven SSR markers developed in this study proved useful for DNA fingerprinting and genetic relatedness analyses. Since many organized marijuana growers use clonal propagation and hydroponic methods to produce plants that yield high drug content, this technique could be applied to provide linkage between a major source of marijuana and the smaller growers to assess distribution patterns by tracking clonal material. This information also could help to link cases and/or growers together with an aim to aid law enforcement agencies in drug eradication efforts. Other growers propagate their marijuana plants from seeds. The forensic application of this technique with seed grown plants could be used to associate a leaf found on a suspect to plant material found in a crime scene or to link an anonymous growing operation (i.e. marijuana cultivated in a particular field) to material found in a suspect's possession.

Conclusion

Microsatellite markers are used in a variety of applications and they are especially preferable when analyzing large numbers of samples. These markers provide excellent assessments of genetic diversity due to their high information content, ease of genotyping, codominant and multi-allelic nature, high discriminatory power, and reproducibility of results. Therefore, microsatellites are well suited for DNA fingerprinting and genetic relatedness between *Cannabis* plants.

The microsatellite enrichment library of *Cannabis* was composed of 49% positive repeat inserts. The microsatellite sequences isolated were dinucleotide repeats (51%) and trinucleotide repeats (49%). Of these, 79% were perfect repeats, 14% were imperfect repeats, and 7% were compound repeats. The most common microsatellite motif was the dinucleotide GA/CT which represented 50% of all microsatellites. The most common trinucleotide motifs were GTT/CAA, AAG/TTC, and GAT/CTA representing 16%, 15%, and 10% respectively of the microsatellites isolated (table 1). From the twenty-five primer pairs that were selected, synthesized, and tested for polymorphism, eleven loci were polymorphic and reliable markers. The eleven microsatellite markers were derived from three dinucleotide repeats, five trinucleotide repeats, one compound trinucleotide repeat, and two imperfect trinucleotide repeats.

A total of 52 alleles were detected across 41 *Cannabis* plants tested. The number of alleles per locus ranged from three to nine, averaging 4.7 alleles. The polymorphism level at each locus was evaluated by the expected heterozygosity (*He*), which varied between 0.368 and 0.710 with a mean value of 0.568. The probability of identical genotypes across all eleven loci was calculated to be 1.8×10^{-7} . This allows for high

discrimination capability between unrelated accessions. Thus, the eleven SSR markers were used to identify all the *Cannabis* plants that were duplicates and all samples that could be clonally propagated (tree 1, 2). The major genetic associations between the 27 unique genotypes were consistent when investigated by NJ, UPGMA, and PCA. These analyses could give useful investigative leads as to distribution patterns and source of origin of marijuana samples. The microsatellite DNA typing of *Cannabis* could aid the forensic community in criminal cases as well as assist the law enforcement agencies in drug eradication efforts.

References

- 1) Drug Identification Bible (DIB) (2001). Amera-Chem, Inc., Grand Junction, CO.
- 2) Mikuriya TH (1973). Marijuana: medical papers. 1839-1972. Medi-Comp Press, Oakland, CA.
- 3) Li Hui-Lin (1974). An Archeological and historical account of cannabis in China. *Economic Botany* 28 (4): 437-448.
- 4) Keng H (1974). Economic plants of ancient N. China as mentioned in Shih Ching (Book Of Poetry). *Economic Botany* 28 (4): 391-410.
- 5) Cannabis and Culture (CC) (1975). Mouton Publishers Company Aldine, Chicago, IL.
- 6) Grinspoon L, Bakalar JB (1993). "The history of cannabis" in marijuana: the forbidden medicine. New Haven Yale University Press.
- 7) Sakamoto K, Akiyama Y, Fukui K, Kamada H, Satoh S (1998). Characterization: genome sizes and morphology of sex chromosome in hemp (cannabis sativa L.). *Cytologia* 6:459-464.
- 8) Augrill S, Nillson JLG (1972). The chemistry and biological activity of cannabis. *Bull. Narc.* 24 (4):2935-2937
- 9) Schultes RE, Klein WM, Plowmna T, Lockwood, TE (1974). Cannabis and example of taxonomic neglect; *Bot Mus Leafl Harv Univ* 23:337.
- 10) Stearn WT (1974). Typification of cannabis sativa L.: *Botanical Museu Leaflets, Harvard Univ*, 23(9):325-336
- 11) Schultes, RE (1970). Random thoughts and queries on the botany of *Cannabis*; In Joyle CRB, Curry SH (Eds): *The botany and chemistry of cannabis*; Churcill: London, UK.
- 12) Joyce CR, Curry SH (1970). *The botany and chemistry of cannabis*. J&A Churchill, London.
- 13) Small E, Croquist A (1976). A practical and natural taxonomy for cannabis. *Taxon* 25 (4):405-435.
- 14) Strauss C, Mead M, Hamamsy L, Srinivas M (1975). *Cannabis and culture*. Mouton Publishers.

- 15) Siniscalco Gigliano G (2001). *Cannabis sativa* L.- botanical problems and molecular approaches in forensic investigations; *Forensic Sci Rev.* 13:1.
- 16) USDA, NRCS (2001). The PLANTS Database, Version 3.1 <http://plants.usda.gov>, National Plant Data Center, Baton Rouge, LA.
- 17) Clarke R (1981). *Marijuana Botany: The propagation and breeding of distinctive cannabis*. Ronin Publishing, Berkeley, CA.
- 18) Warmke HE, Davidson H (1943-44). Polyploidy investigation. *Year Book of the Carnegie Institution of Washington.* 43:135-139.
- 19) Pate D (1994). Chemical ecology of cannabis. *Journal of the International Hemp Association*, 2: 29, 32-37.
- 20) British Medical Association (BMA) (1997). *The misuse of drugs*. Harwood Academic, London.
- 21) Vollner L, Bieniek D, Korte, F (1986). Review of analytical methods for identification and quantification of Cannabis produces. *Regul Toxicol Pharm*, 6:348.
- 22) Latta R, Eaton B (1975). Seasonal fluctuations in cannabinoids content of Kansas marijuana. *Econ Bot*, 29:153.
- 23) Coffman C, Gentner W (1975). Cannabinoid profile and elemental uptake of *Cannabis sativa* L. as influenced by soil characteristics; *Agron J.*, 67:491.
- 24) Mahlberg PG, Hemphill JK (1983). Effect of light quality on cannabinoids content of *cannabis sativa* L. (Cannabaceae). *Chem Rev.*, 4:43.
- 25) Baker PB, Gough TA, Taylor BJ (1980). Illicitly imported Cannabis products: some physical and chemical features indicative on their origin. *Bull Narc* 32(2):31.
- 26) Lowry WT, Garriott JC (1975). On the legality of cannabis: The responsibility of the expert witness. *J. Forensic Science*, 20:624.
- 27) Nakamura GR (1969). Forensic aspects of cystolith hairs of cannabis and other plants; *J. Assoc Off Anal Chem* 52:5..
- 28) Butler WP (1962). Duquenois-Levine test for marihuana. *J Assoc Off Anal Chem*, 94:597.
- 29) De Forest P.D, Gaensslen RE, Lee HC (1983). *Forensic science: an introduction to criminalistics*. McGraw-Hill, Inc. NY.

- 30) Weising K, Nybom H, Wolff K, Meyer W (1995). DNA fingerprinting in plants and fungi. CRC Press, Inc., Boca Raton, FL.
- 31) Lewontin R, Hubby JL (1966). A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54: 595-609.
- 32) Botstein D, White RL, Skolnick M, David RW. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314–331.
- 33) Jeffreys AJ, Wilson V, Thein SL (1985). Hypervariable "minisatellite" regions in human DNA. *Nature* 314,67-73.
- 34) Jefferys AJ, Wilson V, Thein, SL (1985). Individuals specific "Fingerprinting" of human DNA. *Nature*, 326, 76-79.
- 35) Inman K, Rudin N (1997). An introduction to forensic DNA analysis. CRC Press Inc., Boca Raton, FL
- 36) Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, *Science*, 230: 1350-1354.
- 37) Saiki RK, Gelfand DH, Stoffel S, Scharf S, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Nature*, 230:487-497.
- 38) Innis MA, Gelfand DH, Sninsky JJ, White TJ (1990). PCR protocols, A Guide to Methods and Applications. Academic Press. San Deigo, CA.
- 39) Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18,6531-6535.
- 40) Fritsch P, Hanson MA, Spore CD, Pack PE, Rieseberg LH (1993). Constancy of RAPD primer amplification strength among distantly related taxa of flowering plants, *Plant Mol. Biol. Rep.*, 11:10-20.
- 41) Janssen P, Coopman R, Huys G, Swings J, Bleeker M, Vos P, Zabeau M, Kersters K (1996). Evaluation of the DNA fingerprinting method AFLP as an new tool in bacterial taxonomy. *Microbiology* 142, 1881-93.

- 42) Savelkoul PM, Aarts HM, Haas J de, Dijkshoorn L., Duim B., Otsen M., Rademaker JM, Schouls L, Lenstra JA (1999). Amplified-fragment length polymorphism analysis: the state of an art. *J. Clinical Microbiology*. 37: 3083-3091.
- 43) Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T (1997). Molecular tools in plant genetics resources conservation: a guide to the technologies. Rome, Italy; IRGRI.
- 44) Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995). Combined mapping of AFLP and RFLP markers in barley. *Mol Gen Genet* 249:65-73.
- 45) Litt M, Luty JA (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44,397-401.
- 46) Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK, Mao Yi-Pei, Ganten D, Dzau VJ, Lander ES (1991). Genetic mapping of a gene causing hypertensive rat. *Cell* 67,213-224.
- 47) Edwards A, Civitello A, Hammond HA, Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* 49,746-756.
- 48) Weber, JL, May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44,388-396.
- 49) Wang Z, Weber JL, Zhong G, Tanksley SD (1994). Survey of plant short tandem DNA repeats. *Theor Appl Genet*, 88;1-6.
- 50) Gupta PK, Balyan PC, Sharma PC, Ramesh B (1996). Microsatellites in plants: a new class of molecular markers. *Curr. Sci.*, 70:45-54.
- 51) Gupta PK, Varshney RK (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica*, 113:163-185.
- 52) Toth G, Gaspari Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genomes Research*, 10:967-981.
- 53) Weber JL (1990). Informativeness of human class of human (cC-dA)_n (dG-dT)_n polymorphisms. *Genomics*, 7:524-530.
- 54) Maroof MAS, Zhang Q, Biyashev RM (1994). Molecular marker analysis of powdery mildew resistance in barley. *Theor. Appl. Genet.* 88: 733-740.
- 55) Butler JM (2001). Forensic DNA Typing. Academic Press, CA, USA

- 56) Levinson G, Gutman GA (1987). High frequencies of short frame shifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res*, 15:5323-5338.
- 57) Schug MD, Marckay TFC, Aquadro CF (1997). Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Gene*, 15:99-102.
- 59) Strand M, Prolla TA, Liskay RM, Peters TD (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, 365:274-276.
- 60) Maguire TL, Edwards KJ, Saenger P, Henry R (2000). Characterization and analysis of microsatellite loci in a mangrove species, *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae). *Theor Appl Genet*. 101: 279-285.
- 61) Morgante M, Olivieri AM (1993). PCR-amplified microsatellites as markers in plant genetics. *Plant J*. 3:175-182.
- 62) Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225-238.
- 63) Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Sci*, 1:215-222.
- 64) Jones CJ, Edwards KJ, Castaglione S, Windfield MO, Sala F, Van De Wiel C, Bredemeijer G, Vosman B, Mathes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevski A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero L, Vasquez A, Karp A (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed*. 3: 381-390.
- 65) Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri AM (1994). Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome*, 37: 763-769.
- 66) Ashkenazi VE, Chani U, Lavi D, Levy J, Hillel, Veilleux RE (2001). Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. *Genome* 44:50-62.
- 67) Brown, SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996). Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theor. Appl. Genet*. 93:190-198.

- 68) Ayres NM, McClung AM, Larkin PD, Blich HFJ, Jones CA, Park WD (1997). Microsatellite and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. *Theor. Appl. Genet.* 94:773-781.
- 69) Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139:463-471.
- 70) Yoon CK (1993). Botanical witness for the prosecution. *Science*, 260:894-895.
- 71) Farrell M, Howes S, Taylor C, Lewis G, Jenkins R, Bebbington P, Jarvis M, Brugha R, Gill B, Meltzer H (1998). Substance misuse and psychiatric comorbidity: an overview of the OPCS National Psychiatric Morbidity Survey. *Addictive Behaviours* 23 909-918.
- 72) Siniscalco Gigliano G, DI Finizio A, Caputo P, Cozzolino S (1998). *Cannabis* fingerprints by using random amplified polymorphic DNA (RAPD); *Delpinoa ns* 37:35.
- 73) Coyle, HM., Ladd C, Palmbach T, Lee HC (2001). The green revolution: botanical contributions to forensics and drug enforcement. *CMJ.* 42(3):340-345.
- 74) Coyle HM, Shutler G, Abrams S, Hanniman J, Neylon S, Ladd C, Palmbach T, Lee HC (2003). A simple DNA extraction method for marijuana sample used in amplified fragment length polymorphism (AFLP) analysis. *J Forensic Science*, 48(2): 343-347.
- 75) Edwards KJ, Backer JHA, Daly A, Jones C, Karp A (1996). Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques*: 20:5.
- 76) Roy R, Steffens DL, Gartside B, Jang GY, Brumbaugh JA (1996). Producing STR locus patterns from bloodstains and other forensic samples using an infrared fluorescent automated DNA sequencer. *J. Forensic Sciences.* 41 (3)418-424.
- 77) Nei M (1987). Genetic distance and molecular phylogeny. In population genetics and fishery management. Edited by N. Ryman and F. Utter. University of Washington Press, Seattle. Pp. 193-223.
- 78) Paetkau D, Calvert W, Stirling I, Strobeck C (1995). Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* 4:347-354.
- 79) Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4:406-425.
- 80) Sokal R, Rohlf FJ (1995). *Biometry.* 3rd edition. W. H. Freeman, Co., New York, NY.
- 81) Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *Evolution* 21: 550-570.

- 82) Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.
- 83) Wright S (1978). *Evolution and the genetics of populations*, vol. 4. Variability within and among natural populations. Univ. Chicago Press, Chicago.
- 84) Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791.
- 85) Rohlf FJ (1998). NTSYS: numerical taxonomy and multivariate analysis system, version 2.1. Applied Biostatistics Inc., Setauket, N.Y.
- 86) Cornuet JM, Piry S (2000). TreeMaker v1.0.1: a Windows program for making NJ dendrograms and assigning a bootstrap value between the nodes. Computer software distributed by the author.
- 87) Hauge XY, Litt M (1993). A study of the origin of shadows bands seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum. Mol. Genet.* 2;411-415.
- 88) Brownstein MJ, Carpten JD, Smith JR (1996). Modulation of non-templated nucleotide addition by taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 1996 Jun; 20(6):1004-1010.
- 89) Ginot F, Bordelais I, Nguyen S, Gyapay G (1996). Correlation of some genotyping errors in automated fluorescent analysis by enzymatic removal of one base overhangs. *Nucleic Acids Res.*, 24, 540–541.
- 90) Cordeiro GM, Taylor GO, Henry RJ (2000). Characterization of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. *Plant Science*, 155:161-168.
- 91) Rossetto M, McLauchlan A, Harriss FC, Henry RJ, Baverstock PR, Lee LS, Maguire TL, Edwards KJ (1999). Abundance and polymorphism of microsatellite markers in the tea tree (*Melaleuca alternifolia*, Myrtaceae). *Theor. Appl. Genet.* 98: 1091-1098.
- 92) Jakše J, Javornik B (2001). High Throughput Isolation of Microsatellites in Hop (*Humulus lupulus* L.). *Plant Molecular Biology Reporter* 19: 217–226.
- 93) Diwan N, Cregan PB (1997). Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.*, 95:723-733.
- 94) Staub EJ, Danin-Poleg Y, Fazio G, Horejsi T, Reis N, Katzir N (1999). Comparative analysis of cultivated melon groups (*Cucumis melo* L.) using random amplified polymorphic DNA and simple sequence repeat markers. *Euphytica*, 115:225-241.

- 95) Jakše J, Kindlhofer K, Javornik B (2001). Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. *Genome*, 44:773-782.
- 96) Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997). Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 95:553-567.
- 97) Codit R, Hubbel SP (1991). Abundance and DNA sequence of 2-base repeat regions in tropical tree genomes. *Genome* 34:66-71.
- 98) Li CD, Rosnagel BG, Scoles GJ (2000). The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. *Theor. Appl. Genet.* 101: 1259-1268.
- 99) Senior ML, Murphy JP, Goodman MM, Stuber CW (1998). Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci.* 38:1088-1098.
- 100) Ahmad M (2002). Assessment of genomic diversity among wheat genotypes as determined by simple sequence repeats. *Genome*, 45:646-651.
- 101) Yu J-K, Mangor J, Thompson L, Edwards KJ, Slabaugh MB, Knapp SJ (2002). Allelic Diversity of simple sequence repeats among elite inbred line of cultivated sunflower. *Genome*, 45:652-660.

Appendix

A summary of this thesis is in press for the **Journal of Analytical and Bioanalytical Chemistry** entitled:

Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses

Hussain J. Alghanim^{1,2} and Jose R. Almirall

¹ Florida International University, University Park, Miami, FL 33314

² Dubai Police Crime Laboratory, P.O.BOX: 1493, Dubai, United Arab Emirates

Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses

Keywords: *Cannabis sativa*, marijuana, microsatellite, short tandem repeat (STR), DNA fingerprinting, DNA typing, genetic relatedness

Abstract: Microsatellite markers were developed for *Cannabis sativa* L. (marijuana) to be used for DNA typing (genotype identification) and to measure the genetic relationships between the different plants. Twelve different oligonucleotide probes were used to screen an enriched microsatellite library of *Cannabis sativa* in which 49% of the clones contained microsatellite sequences. Characterization of microsatellite loci in *Cannabis* revealed that GA/CT was the most abundant class of the isolated microsatellites representing 50% overall followed by GTT/CAA, AAG/TTC, and GAT/CTA representing 16%, 15%, and 10% respectively. Eleven polymorphic STR markers were developed, three derived from dinucleotide motifs and eight from trinucleotide motifs. A total of 52 alleles were detected averaging 4.7 alleles/locus. The expected heterozygosity of the eleven loci ranged between 0.368 and 0.710 and the common probability of identical genotypes was 1.8×10^{-7} . The loci identified 27 unique profiles of the 41 *Cannabis* samples. The eleven microsatellite markers developed in this study were found to be useful for DNA typing and for assessing genetic relatedness in *Cannabis*.

Introduction

Cannabis sativa L. (marijuana) is one of the oldest cultivated plants and has been used around the world for diverse applications. Throughout history, *Cannabis* has been used as a source of hemp fiber for rope and fabric from its stems, food and oil from its seeds, and as a psychoactive drug from its flowers and leaves. *Cannabis*, marijuana, and hemp are terms that are used interchangeably. However, *Cannabis* is the botanical genus of the plant and marijuana describes *Cannabis* plants that contain high Δ^9 tetrahydrocannabinol (THC) content and are used for their psychoactive potency [1]. Hemp is used to describe *Cannabis* plants that have low THC content and are cultivated for fiber. Therefore, there are two distinctive strains; one is generally cultivated for fiber (hemp) and the other for drug use (marijuana) [2]. Historically, there were generally three recognized varieties of *Cannabis*: *C. sativa*, *C. indica*, and *C. ruderalis* [3]. For many years, botanists considered each of them to be a distinct species. However, most botanists now generally agree that *Cannabis* is a genus with a single highly variable species (*C. sativa*) that has diversified into a wide variety of ecotypes and cultivated races [3].

The unique portions of an individual's DNA sequence has made it possible to study the genetic diversity and relatedness between organisms. A wide variety of techniques to determine DNA sequence polymorphisms have been developed, and molecular markers have been derived from those techniques. Genetic variation at the DNA level can be detected by using different molecular markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), or simple sequence repeat (SSR). One

of the most useful and widely used markers is SSR [4], otherwise known as microsatellite [5], or short tandem repeat (STR) [6]. Microsatellites are DNA sequences of six or fewer bases that are repeated in tandem arrays (i.e. CTCTCTCTCTCTCT) [7]. These repeats reveal high levels of polymorphism between individuals due to replication slippage and unequal crossing over [8, 9]. Microsatellites are evenly distributed in human and other mammalian genomes as well as in plants [10].

Several important advantages make microsatellite markers the method of choice for DNA typing and analysis of genetic relationships. Microsatellites are usually a single locus with multiple alleles and this robust technique can be easily distributed between different laboratories as primer sequences [5]. Microsatellites can also be used in multiplex PCR where several microsatellite loci can be assayed in the same amplification reaction. Microsatellite markers are codominant, highly informative, reproducible, and have high discrimination power [11-14]. Because of these advantages, microsatellites have become well suited for a wide range of applications in genetic mapping [15], fingerprint and genotype identification [16], seed purity evaluation and germplasm conservation [17], genetic relatedness and paternity studies [18], and marker-assisted selection [19].

Molecular marker systems based on RAPD [20] and AFLP [21, 22] have been developed and used for DNA typing analysis of *C. sativa*. However, RAPDs are dominant makers with poor reproducibility between labs [14]. AFLP analysis detects multiple loci with high reproducibility, however, AFLP are also dominant markers. RFLP analysis is highly polymorphic but it is very labor intensive. In addition, microsatellite markers have been developed for some plants for general purposes [23, 15]

and very recently in *C. sativa* for forensic purposes [24, 25]. This report presents another survey of the microsatellites detected in *Cannabis* and their forensic application. The objective of this work was to develop a number of microsatellite markers capable of individualizing *Cannabis* samples for DNA typing and genetic relatedness analyses.

Materials and Methods

Microsatellite loci were developed by a marker enrichment technique, which consisted of: 1) hybridizing extracted genomic DNA of a known cultivar of *C. sativa* with specific repeat unit probes, 2) sequencing positive clones, 3) designing oligonucleotide primers on either side of the repeat region, and 4) testing loci for polymorphism by sampling different unrelated individuals.

DNA isolation and preparation of genomic DNA:

Genomic DNA samples were provided by Dr. Heather Coyle (Connecticut State Forensic Science Laboratory, CSFSL, USA) and Dr. Gary Shutler (Royal Canadian Mounted Police, RCMP, Canada) and extracted with a QIAGEN plant DNeasy kit according to the manufacturer's recommendations [22]. Genomic DNA was digested with Sau3AI (Life Technologies, Inc., Gaithersburg, MD), a restriction endonuclease recognizing the 5'-GATC-3' DNA sequence. Double-stranded linkers (Sau) were synthesized to have a 3' overhang of CTAG by the following oligonucleotides:

•***Sau-L-A: 5'-GCCGTACCCGGGAAGCTTGG-3'***

•***Sau-L-B: 5'-GATCCCAAGCTTCCCGGGTACCGC-3'***

One ug of the linkers were then ligated to 200 ng of the Sau3AI digested genomic DNA using 20 U of T4 DNA ligase and 8 µl of 5X ligase buffer (Life Technologies, Inc.,

Gaithersburg, MD) in a final volume of 40 μ l. The reaction mix was incubated at 4° C for 72 hours and then the ligation reaction was stopped by heating at 65° C for 10 minutes. Excess linkers were removed with Performa DTR Gel Filtration Cartridges (Edge Biosystems, Inc., Beverly, MA) following the manufacturer recommendations.

Microsatellite enrichment and size fractionation:

The genomic DNA was amplified by the polymerase chain reaction (PCR) followed by purification using phenol:chloroform:isoamyl alcohol (PCI) extraction and concentration by ethanol precipitation. The genomic DNA was enriched using a modification of the method of Edwards et al. (1996) [26]. Twelve different biotinylated oligonucleotide probes were employed to search for the different microsatellite motifs. Those probes consisted of two dinucleotide motifs (CT)₁₅ and (GT)₁₅, and ten trinucleotide motifs (CAA)₁₀, (ATT)₁₀, (GCC)₁₀, (ACC)₁₀, (AGG)₁₀, (CTT)₁₀, (AGC)₁₀, (ACG)₁₀, (ACT)₁₀, and (ATC)₁₀. Microsatellite-containing fragments were isolated using Dynabeads M-280 Streptavidin (Dynal, Oslo, Norway). To enrich for STR fragments, 25 μ l of the denatured DNA was mixed with 1 μ g of the Sau-L-A oligo, 474 μ l of the hybridization buffer (50% formamide, 3X SSC, 25mM Na-phosphate pH 7.0, and 0.5% SDS), and 500 μ l of Dynabeads in 2X B&W buffer (10mM Tris-HCl pH 7.5, 1mM EDTA and 2M NaCl). The hybridization reaction (1000 μ l) was mixed well and incubated overnight at room temperature on a mixing table. The reaction was then subjected to the following hybridization washes: 5 washes for 3 minutes each using buffer #1 (2X SSC and 0.01% SDS) at 42° C followed by 3 washes for 3 minutes each using buffer #2 (0.5X SSC and 0.01% SDS) at 42° C. The wash buffer was removed and

the Dynabeads were resuspended in 200 μ l of nuclease free water. To further increase the DNA fragments containing microsatellites, a second PCR amplification and enrichment were performed. After the second enrichment, the genomic DNA was amplified and purified again with an equal volume of PCI. The size fractionation was performed using SizeSep 400 Spun Column Sepharose (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) according to the manufacturer recommendations followed by concentration with ethanol precipitation.

Cloning reaction and plasmid recovery:

The cloning reaction was performed using the TOPO TA Cloning Kit (Invitrogen Corp., Chicago, IL) following manufacturer's protocols. Sterile toothpicks were used to inoculate 96-well plates containing 100 μ l of SOC broth with 100 μ g/ml ampicillin. The broth culture plates were covered with aluminum foil tape and placed at 37° C overnight. A 25 μ l sample of cell culture from each well was transferred to another 96-well plate. The plates were centrifuged to pellet cells and then were inverted and spun for a short time to remove the broth. The cells were resuspended with 50 μ l of 10mM Tris-HCl (pH 8). This was used as a template for insert PCR with M13 primers (Life Technologies, Inc., Gaithersburg, MD).

Plasmid inserts amplification and cycle sequencing:

Plasmid inserts were amplified using M13 primers and the resulting M13 PCR products were treated with exonuclease I (Life Technologies, Inc., Gaithersburg, MD) followed by ethanol precipitation. The PCR products were then sequenced using BigDye Terminator™ Cycle Sequencing Ready Reaction Kit (Version 2.0) (Applied Biosystems,

Foster City, CA). The sequencing products were ethanol precipitated overnight at room temperature in the dark. The sequencing products were resuspended in 10 µl of Hi-Di formamide (Applied Biosystems, Foster City, CA) and denatured at 95° C for 2 minutes. The cycle sequencing products were electrophoresed on the ABI 3100 and analyzed with DNA Sequencing Analysis Software 3.7 (Applied Biosystems, Foster City, CA). Microsatellite containing fragments were imported into Sequencher v.4.1 (Applied Biosystems, Foster City, CA) to sort, clean up, and generate consensus sequences before primer design.

Primer design and fragment analysis:

Based on the flanking sequences, PCR primers were designed using the GCG Wisconsin Packages (Accelrys, Madison, WI). After examining six different temperatures, the optimal annealing temperature was determined to be 53° C at which the amplicons had the highest intensity as measured by relative fluorescence units (rfu). With some modifications, the M13 fluorescent tail primer method described by Roy et al. (1996) [27] was used as a screening technique to detect polymorphism among 25 different loci. Fragment analysis was performed using either the method described by Roy [27] or direct 5'-labeled fluorescent primers. The PCR reaction using the direct fluorescent primer was prepared (for one reaction) by adding 1 µl of 10X buffer with 15mM MgCl₂, 0.2 µl of dNTP's (2.5mM each), 0.25 µl of forward primer (10uM), 0.25 µl of reverse primer (10µM), 0.5 U of *taq* polymerase, 10 ng of DNA template, and dH₂O to a final volume of 10 µl. The PCR reaction was carried out by denaturing at 94° C for 5

minutes, followed by 40 cycles of 45 seconds at 94° C, 45 seconds at 53° C, and 1 minute at 72° C. The final extension was 10 minutes at 72° C.

Determination of allelic sizes:

The PCR products were electrophoresed in a capillary electrophoresis instrument (ABI Prism 3100) using Performance Optimized Polymer 4 (POP4) (Applied Biosystems, Foster City, CA). Each sample was prepared by mixing 1.5 µl of the PCR product with 12 µl of Hi-Di formamide and 0.1 µl of GeneScan 500 ROX fluorescently labeled size standard (Applied Biosystems, Foster City, CA). The PCR products were denatured by incubating at 95° C for 2 minutes. Samples were injected electrokinetically at 3kV for 10 seconds and were run at 60° C for 45 minutes at 15kV. The data generated was imported into GeneScan 3.7 software (Applied Biosystems, Foster City, CA) for fragment size determination. The final allele size determination of the microsatellite data was performed using Genotyper 3.7 (Applied Biosystems, Foster City, CA). An example of alleles called for the P19 locus is shown in figure 1.

Statistical analyses:

All samples were scored for the allele designations based on the repeats size, which were then used in different statistical analyses. To investigate some genetic parameters of polymorphism, the following was calculated: allele frequencies, number of alleles per locus (n), effective number of alleles (n_e), observed heterozygosity (H_o), expected heterozygosity (H_e), and probability of identical genotypes (PI). The genetic parameters were determined using the eleven microsatellites over 33 diploid *Cannabis* plants (excluding the duplicate samples). Observed heterozygosity (H_o) was obtained by

dividing the number of heterozygous plants over the total number of plants tested for each locus. The degree of polymorphism was measured using the expected heterozygosity (H_e , 28):

$$H_e = 1 - \sum P_i^2$$

where P_i is the frequency of the i th alleles for each locus in the plants analyzed. The probability of identical genotypes (PI) was estimated according to Paetkau et al. (1995) [29]:

$$\sum P_i^4 - \sum \sum (2 P_i P_j)^2$$

where P_i is the frequency of the i th allele and P_j equals the frequency of the $(i+1)$ th allele studied. The effective number of alleles (n_e) was calculated according to Morgante's formula [15]:

$$n_e = (\sum p_i^2)^{-1}$$

The genetic relationships among the unique *Cannabis* profiles were analyzed using the neighbor joining (NJ) method [30]. The NJ tree was performed based on Chord's genetic distance [31]. In order to give a confidence limit for the relationships between the *Cannabis* plants, 2000 replicas of bootstrapping [32] were performed with the NJ method to test for support of the branch nodes. Those nodes with bootstrap values below 50% were considered unsupported. Principal Component Analysis (PCA) [33] was also performed as another graphical method to depict the genetic relatedness between the plants tested. The determinations of genetic distances, PCA, and NJ clustering were performed using the NTSYSpc v.2.1 package [34]. The bootstrap was performed using the TreeMaker program [35].

Results

Characterization of the isolated microsatellite sequences:

The cloning step of the enriched *Cannabis* DNA generated 685 clones, from which 192 clones were sequenced (two 96-well plates). Ninety-five (95) clones were considered useful as they contained either dinucleotide motifs with nine or greater repeat units, or they contained trinucleotide motifs with five or greater repeat units. The types of microsatellite motifs identified were consistent with the twelve types of oligonucleotide probes that were used for the enrichment. The isolated microsatellite sequences were as follows: 51% dinucleotide repeats, 49% trinucleotide repeats, 79% perfect repeats, 14% imperfect repeats, and 7% compound repeats (table 1).

Total clones of enriched DNA	685
Total sequencing reactions	192
Total microsatellite inserts (%)	49%
Frequency of microsatellites in <i>Cannabis sativa</i> (%)	
GAT/CTA	GA/CT
AAG/TTC	GTT/CAA
GT/CA	CAT/GTA
ACG/TGC	GGA/CCT
10.2%	50.0%
15.3%	16.3%
1.0%	5.1%
1.0%	1.0%
Type of repeat (%)	
By nucleotide string	
Di	51.0%
Tri	49.0%
By form	
Perfect	78.6%
Imperfect	14.3%
Compound	7.1%

Table (1): Microsatellite enrichment success for *Cannabis sativa* and characterization of the microsatellite types.

The isolated microsatellites had two types of dinucleotide repeats and six types of trinucleotide repeats. The majority of microsatellite loci were composed of a GA/CT dinucleotide motif representing 50% overall. The most common isolated trinucleotide motifs were GTT/CAA, AAG/TTC, and GAT/CTA representing 16%, 15%, and 10% respectively, of all detected microsatellites (table 1). The maximum repeat units recorded for dinucleotide motifs were 49 repeats and 17 repeats for the trinucleotide motifs. The number of repeat units ranged between 15 bp (5 trinucleotides) to 98 bp (49 dinucleotides). A complete list of the frequency of microsatellite types recovered from *Cannabis sativa* is shown in table 1.

Characterizations of the selected microsatellite markers:

Thirty-six clones had suitable flanking regions for GCG primer design. From these 36 loci, seven could not generate primer sets because of low annealing temperature particularly due to a high concentration of nucleotides A and T. Of 29 primer pairs designed by GCG, 25 sets were selected, synthesized and tested for polymorphism. Fourteen primer pairs were eliminated because they produced no PCR products, nonspecific products, or complex (uninterpretable) products. These products were mainly due to the high tendency of primers to produce palindromes or primer dimers between the two primer sequences. In addition, two of those fourteen primer pairs were monomorphic. The remaining eleven loci were found to be polymorphic and reliable for scoring the different alleles across the 41 *Cannabis* samples (table 2). The original genomic DNA used in the library construction (G40, table 3) was also included to provide a positive size control. All of the amplified products were in the expected size range and the PCR products ranged from 105 bp to 339 bp. The eleven STR markers

were derived from three dinucleotide repeats, five trinucleotide repeats, one compound trinucleotide repeat, and two imperfect trinucleotide repeats (table 2).

Locus	Repeat	Primers (5'-3')		Alleles Size Range
C11-CANN1 (P4)	(GAT) ₈ (GGT) ₇	GTGGTGGTGATGATGATAATGG TGAATTGGTTACGATGGCG	5	150-175 bp
B01-CANN1 (P6)	(GAA) ₁₃ (A)(GAA) ₃	TGGAGTCAAATGAAAGGGAAC CCATAGCATTATCCCCTCAAG	5	323-339 bp
D02-CANN2 (P7)	(CTT) ₆ (ATT)(CTT) ₁₀	AGATGATGCCAAGAGGCAC TACCCATCCCCTTGGATCAC	4	221-236 bp
C08-CANN2 (P9)	(GA) ₂₁	GCAAGAAGTAGAGAGACAATCC CCCTCAACACTCATTAACCTCAC	9	171-203 bp
H11-CANN1 (P13)	(CT) ₁₈	GCATGTGGTTGTTTCGTACCC CAGCGAACATTCCTCTAGCTC	7	285-297 bp
B02-CANN2 (P14)	(AAG) ₁₀	CAACCAAATGAGAATGCAACC TGTTTTCTTCACTGCACCC	3	163-172 bp
H09-CANN2 (P15)	(GA) ₁₅	CGTACAGTGATCGTAGTTGAG ACACATACAGAGAGAGCCC	6	204-224 bp
E07-CANN1 (P17)	(CTA) ₉	CAAATGCCACACCACCTTC GTAGGTAGCCAGGTATAGGTAG	3	105-111 bp
B05-CANN1 (P19)	(TTG) ₉	TTGATGGTGGTGAAACGGC CCCCAATCTCAATCTCAACCC	4	235-244 bp
D02-CANN1 (P24)	(GTT) ₇	GGTTGGGATGTTGTTGTTGTG AGAAATCCAAGGTCCTGATGG	3	105-111 bp
H06-CANN2 (P25)	(ACG) ₇	TGGTTTCAGTGGTCCTCTC ACGTGAGTGATGACACGAG	3	266-273 bp

Table (2): SSR loci and the primers developed in the study.

Samples	Information	Duplicate	Link
G1	Northeast-street#1	1	
G2	Northeast-street#1	1	
G3	Northeast-street#1	2	
G4	Northeast-street#1		
G5	Northeast-street#1	3	
G6	Northeast-street#1	3	
G7	Northeast-street#1	4	
G8	Northeast-street#1		
G9	Northeast-street#1		
G10	Northeast-street#1	5	
G11	Northeast-street#2		
G12	Northeast-street#2		A
G13	Northeast-street#2		A
G14	Northeast-street#2	6	
G15	Northeast-street#2	7	
G16	Northeast-street#1	2	
G17	Northeast-street#1	4	
G18	Northeast-street#1	5	
G19	Northeast-street#2	6	
G20	Northeast-street#2	7	
G21	Canada		
G22	Canada		9
G23	Canada		E
G24	Canada		
G25	Canada		C
G26	Canada		10-C
G27	Canada		E
G28	Canada		10-C
G29	Canada		
G30	Canada		
G31	Canada		11
G32	Canada		D
G33	Canada		D
G34	Canada		11
G35	Canada		D
G36	Canada		9
G37	Canada		
G38	Canada		B
G39	Canada		B
G40	Used in library construction	8	9
G41	Canada	8	

Table (3): List of information and associations known between the 41 *C. sativa* plants (personal communication with Dr. Coyle at CSFSL and Dr. Shutler at RCMP).

* Identical number (1-8) code for duplicate samples* Identical letter code for related samples

A: closely related by AFLP B: originated from the same grower C: related by AFLP
D: originated from the same grower E: related by AFLP 9: identical based on AFLP
10: identical based on AFLP 11: identical based on AFLP

Genetic identifications and relationships:

According to the blind testing of the 41 *Cannabis* samples using the 11 STR markers, the following sets had the same DNA fingerprint: G1/G2, G10/G15/G18/G20, G22/G36/G40/G41, G26/G28, G3/G7/G16/G17, G14/G19, G31/G34, and G5/G6. The NJ method was used to test the genetic relatedness among the 27 unique genotypes of *Cannabis* using Chord's genetic distance. NJ found a single tree based on Chord's distance coefficient (figure 2). The only three clusters that could be considered supported were between the following: G12 and G13 with bootstrap values of 100%, G38 and G39 with bootstrap value of 85%, and G4 and G27 with bootstrap value of 55%. Principal Component Analysis (PCA) was performed on the 27 unique *Cannabis* genotypes in which 34.6% of the total variation was captured using three coordinates (figure 3a). G30 exhibited a very distinct profile and could be easily recognized as an outlier sample in PCA (figure 3a). After G30 was disregarded from the PCA testing, the remaining 26 genotypes became more widely dispersed on the PCA scatter plot on which 33.8% of the total variations was captured in the three coordinates (figure 3b).

Discussion

Microsatellites in Cannabis sativa:

In plants, on average, there is one microsatellite every 33 kb whereas microsatellites occur approximately every 6 kb in the human genome [10]. Therefore, to obtain an enriched map of STR markers that represents the genome, screening hundreds of thousands of inserts would be necessary. Several methods have been developed to shorten the screening step and produce genomic libraries enriched for certain

microsatellite types. In this study, an enriched microsatellite library was created using a modified version of the method developed by Edwards et al. (1996) [26]. DNA sequences were obtained from a total of 192 colonies of which 49% contained STR insert repeats (table 1). Most of the microsatellite sequences isolated had flanking sequences that were insufficient (i.e. containing less than 25 bp sequences on either side of the repeat). This close proximity of the repeat unit to the cloning site was the main reducing factor for generating useful markers. This problem was also observed in a number of other studies [36, 37]. Such a problem could be due to the size fractionation step and/or to the restriction enzyme used.

The frequency of each class of microsatellite is highly variable among plant species [10]. The variation reported in different plant studies might be caused by variations in genome structure of different species surveyed [36]. In this study, the results indicate that the GA/CT motif was the most highly dispersed form of microsatellite detected in the *Cannabis* genome. The GA/CT motif represented 50% of the total microsatellite repeats detected (table 1). This is consistent with surveys of microsatellite repeats in hop (*Humulus lupulus*, the closest relative to *C. sativa*) [38] and other studies [13, 23] in which the GA/CT motif was the most abundant. However, other reports of microsatellite repeats in mangrove [39] and potato [16] revealed a greater abundance of AC/TG motif over GA/CT motif. The number of trinucleotide repeats detected was similar to that of the dinucleotide repeats (table 1). Trinucleotide profiles are easier to score than the dinucleotide profiles because the variations in the number of core units of trinucleotide motifs are larger in length. In addition, the trinucleotide loci

show more distinct allelic profiles by avoiding the stutter pattern that is often associated with the amplification of dinucleotide loci motifs [40, 41, 5].

Microsatellite polymorphism and applications:

Microsatellite markers have shown high levels of polymorphism in many plant species including rice [18], wheat [42], tropical trees [43], maize [44], sunflower [45], and many more. To discriminate between different cultivars of *C. sativa* and to potentially associate the samples of clonal origin using microsatellites, the selected STR loci would need to be highly polymorphic. The discrimination power, ease of genotyping (figure 1), and high reproducibility emphasize that these eleven microsatellite markers can be used for DNA typing in *Cannabis*. The blind test of microsatellite typing of 41 samples matched the identities of the duplicates and the unique samples reported afterward by AFLP typing (table 3, personal communication, Dr. Coyle, CSFSL). G26/G28, G31/G34, and G22/G36/G40 were found to be identical by AFLP (table 3) and by STR markers, which strongly suggest that they could be clonally propagated. The remaining identical sets consisted of known duplicate samples (table 3), showing reproducibility of this typing technique. In addition, the 11 STR loci were very effective in uniquely identifying 27 profiles of the *Cannabis* samples tested. Microsatellite DNA also had high resolution and sensitivity for calculating the various genetic parameters presented in table 4.

The genetic relationships using the NJ method were able to complement the known association level between the *Cannabis* plant samples reported in table 3. First, samples G12 and G13 were found to have the strongest ties supported by a bootstrap value of 100% (figure 2). This relationship was in agreement with the information that

both samples were closely related by AFLP analysis (table 3). In addition, figure 2 was able to group G38 and G39 together with a bootstrap value of 85%. These results were consistent with the information that both samples came from the same grower (table 3). This information could link these samples to a specific source. Moreover, the seized samples G32, G33, and G35 clustered together in the microsatellite tree (figure 2) all of which were reported as originating from the same source (table 3). G25 and G26 were associated in the same group by the AFLP data (table 3) and by the NJ dendrogram (figure 2). However, microsatellite data supported the relatedness of G27 and G4 together (55% bootstrap value) more than G27 was related to G23 found using AFLP data (table 3). G30 was found to be a distinct individual based on both NJ analysis (figure 2) and PCA (figure 3a). Moreover, the PCA plot associated G12 with G13, G32 and G33 with G35, and G38 with G39. All of these PCA associations were in agreement with the NJ clustering (figure 3b).

Many growers of *Cannabis* choose selective breeding to adapt to particular growing conditions and to increase the THC content, which contributes to the intoxicating effect of *Cannabis* [22]. Selective breeding of high THC content plants can be obtained using two methods. The first technique is clonal propagation using stem cuttings from clones containing high levels of THC. In this case, both the mother plant and subsequent cuttings (daughter clones) will have identical DNA profiles. Thus, a DNA analyst can easily detect plant materials that originated from the same source. Second, two plants can be cross-pollinated to generate seeds. After that, each seed can be grown into a plant that has its own unique DNA profile.

The eleven STR markers developed in this study proved useful for DNA typing and genetic relatedness analyses. Since many organized marijuana growers use clonal propagation and hydroponic methods to produce plants that yield high drug content, this technique could be applied to provide linkage between a major source of marijuana and the smaller growers to assess distribution patterns by tracking clonal material. This information also could help to link cases and/or growers together with an aim to aid law enforcement agencies in drug eradication efforts. Other growers propagate their marijuana plants from seeds. The forensic application of this technique with seed grown plants could be used to associate a leaf found on a suspect to plant material found in a crime scene or to link an anonymous growing operation (i.e. marijuana cultivated in a particular field) to material found in a suspect's possession. Future work will include further validation studies with additional samples, sensitivity and mixture studies, stutter calculations, and mapping the loci to determine any linkage before they can be used by the forensic community.

Acknowledgements

This work is part of the Masters Thesis of the first author at Florida International University, Miami, Florida. The authors would like to specially thank the Dubai Police Crime Laboratory in the United Arab Emirates for providing the financial support for this project and Drs. Alan Meerow and Raymond Schnell at the USDA-ARS-SHRS laboratory in Miami, Florida for the use of their facilities and for helpful discussions and review of this manuscript. The authors also thank Jason Clayton at the USDA-ARS-SHRS laboratory for technical assistance and all the USDA staff at the genetic research

lab in Miami, Florida. Finally, the authors thank Dr. Heather Coyle of CSFSL, USA and Dr. Gary Shutler of RCMP, Canada for providing the DNA samples and the AFLP information.

Figure legends

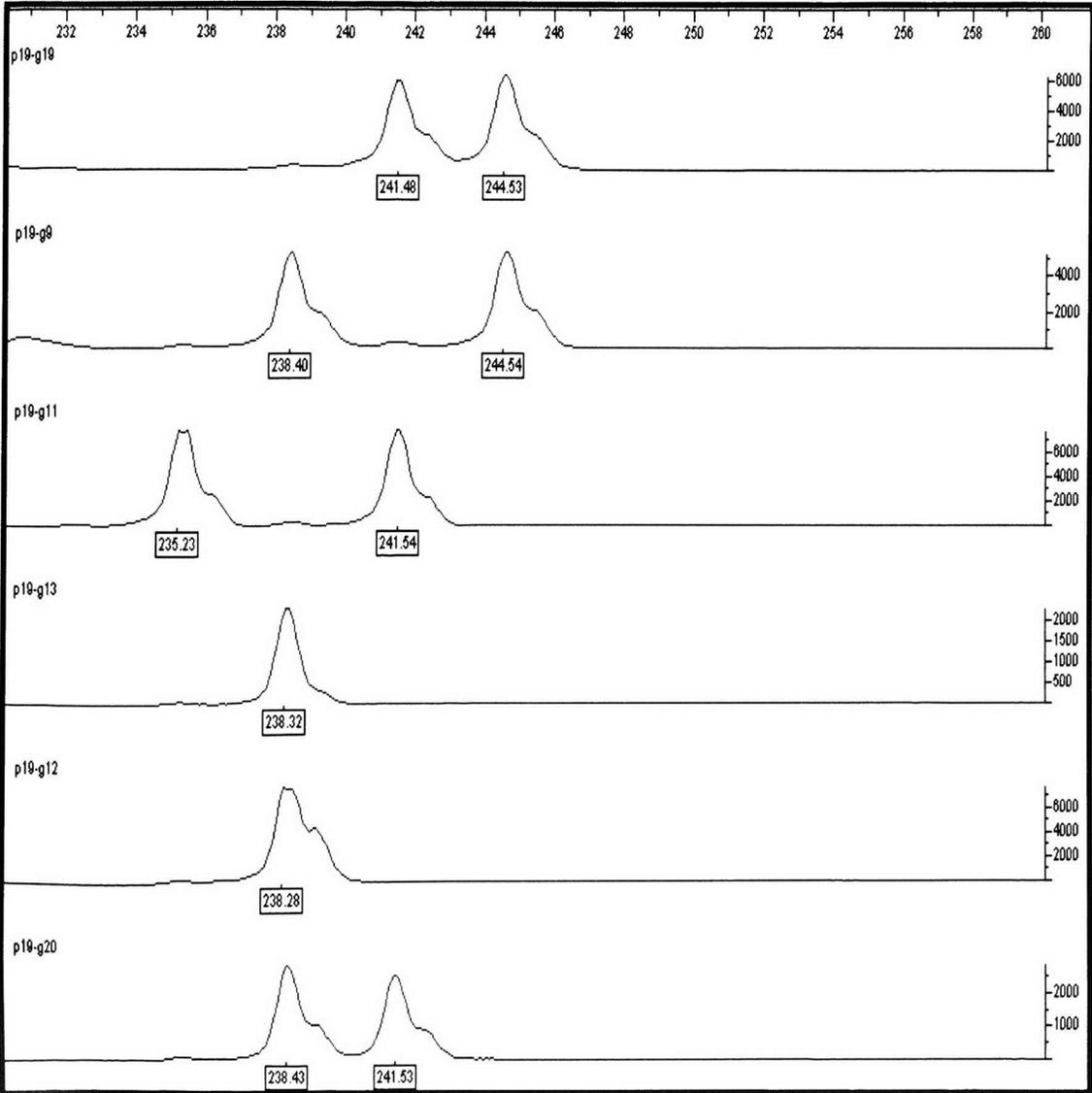


Figure 1: Representative genotyping of different samples based on P19 locus using Genotyper 3.7 software.

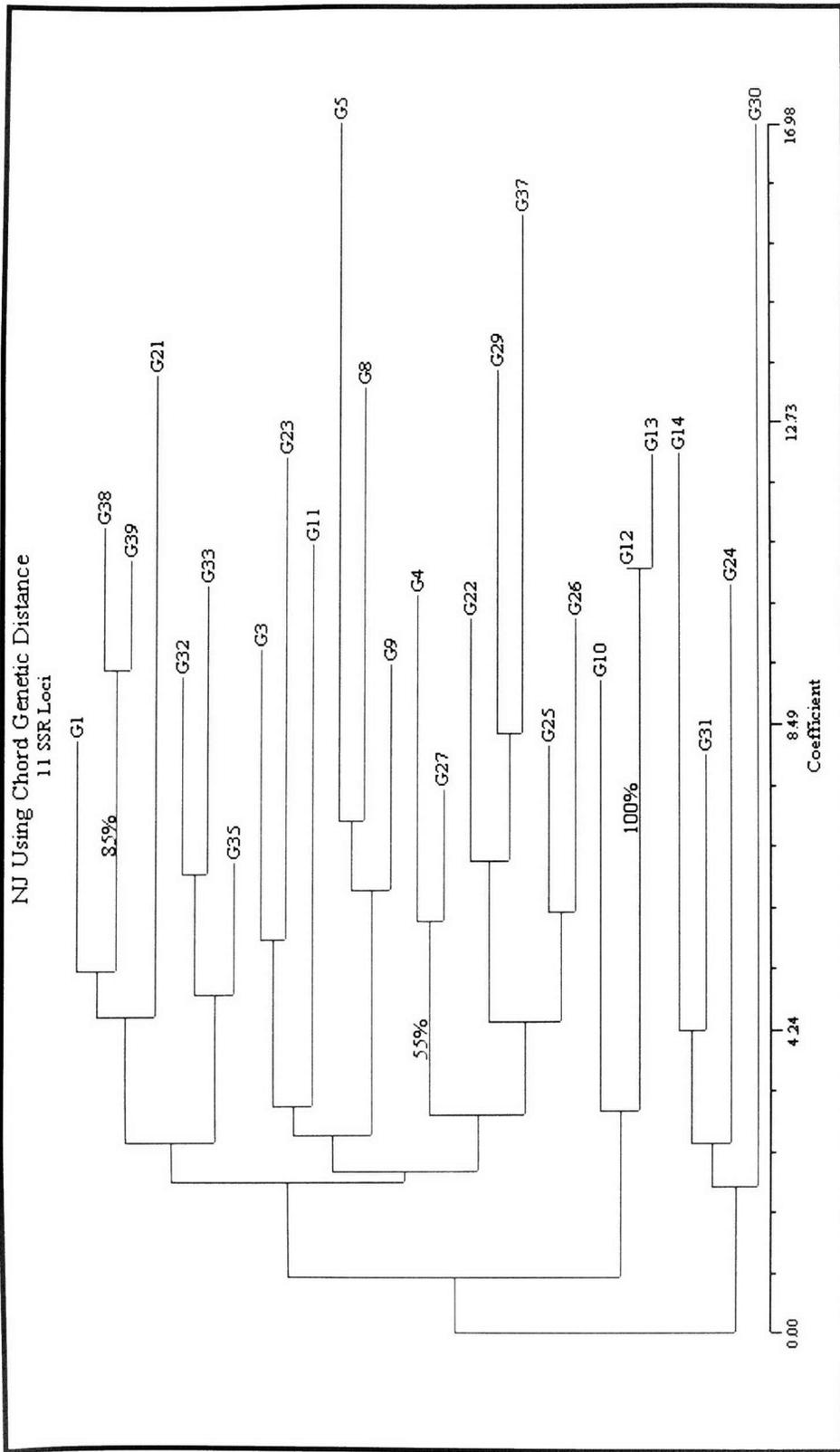


Figure 2: Neighbor joining (NJ) tree using Chord's genetic distance based on overall allele frequencies in 11 SSR loci across 27 *Cannabis* genotypes. Only the bootstrap value of 50% and above is shown.

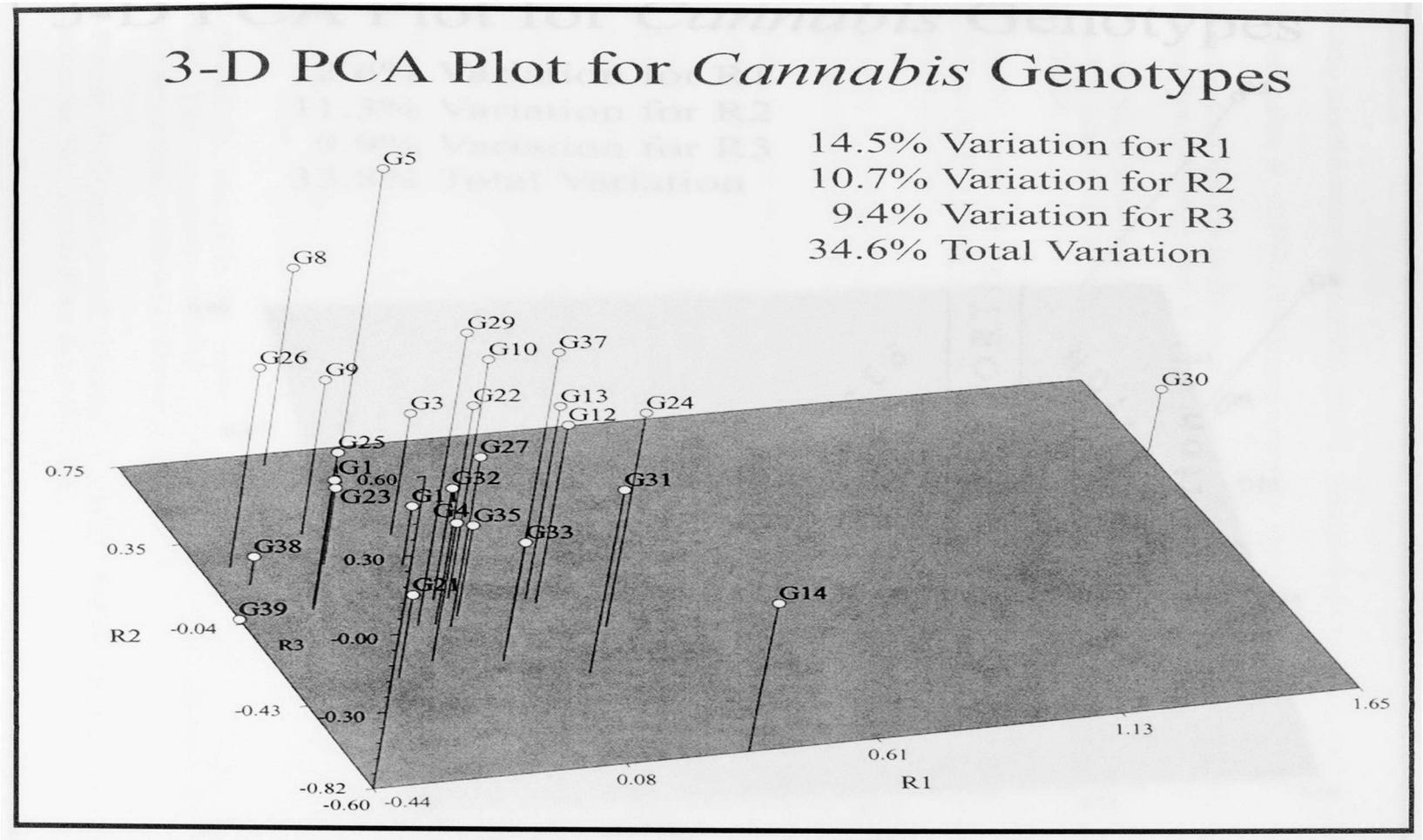


Figure 3A: Three-dimensional principal component analysis plot of the 27 unique *Cannabis* genotypes using correlation coefficients based on overall allele frequencies in 11 SSR loci

3-D PCA Plot for *Cannabis* Genotypes

12.6% Variation for R1
11.3% Variation for R2
9.9% Variation for R3
33.8% Total Variation

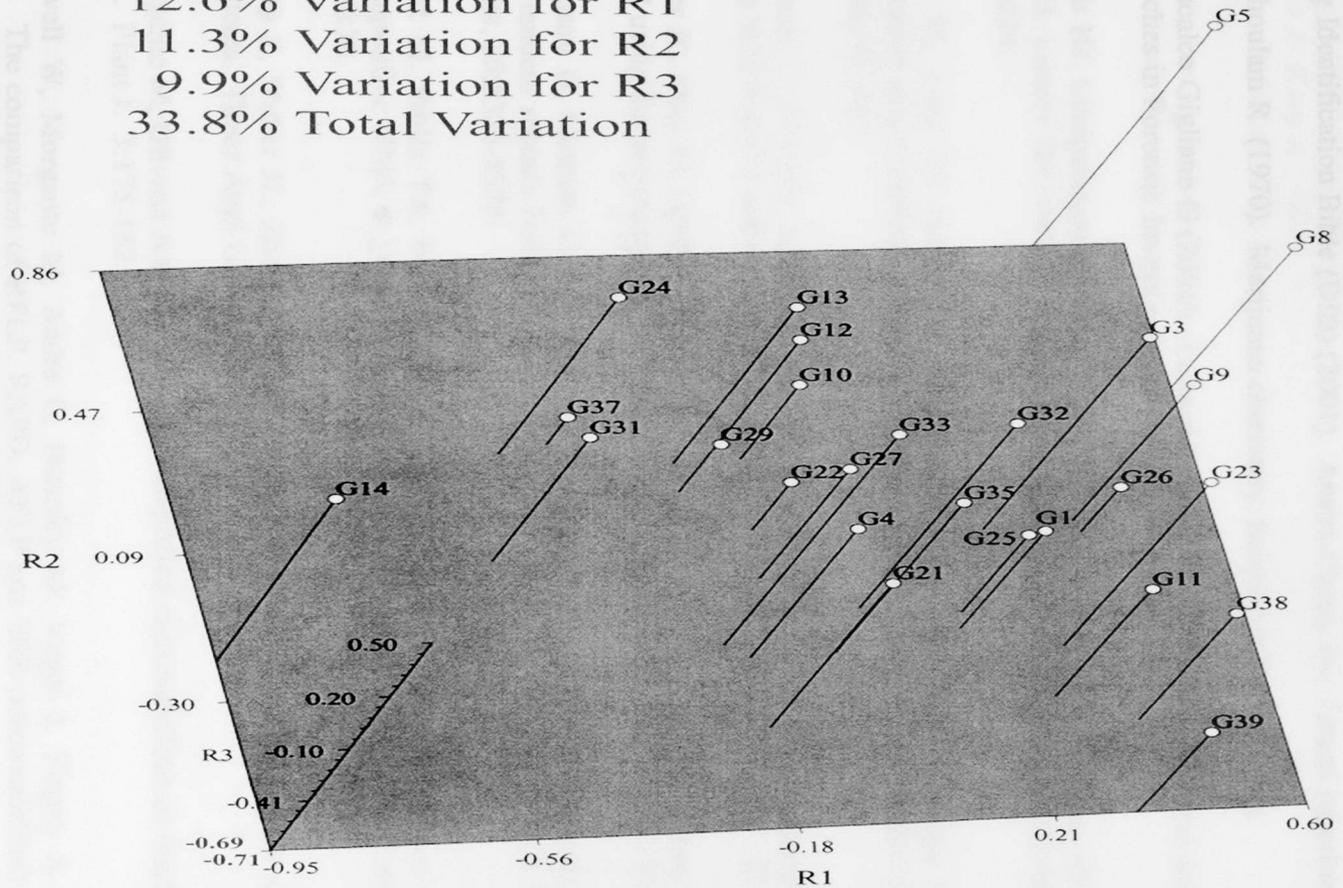


Figure 3B: Same as 3A except that G30 is eliminated from the plot

References

- 1) Drug Identification Bible (DIB) (2001). Amera-Chem, Inc. Grand Junction, CO, USA.
- 2) Mechoulam R (1970). Marijuana chemistry. *Science* 168:1159-1166.
- 3) Siniscalco Gigliano G (2001). *Cannabis sativa* L.- Botanical problems and Molecular Approaches in Forensic Investigations; *Forensic Sci Rev.* 13:1.
- 4) Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK, Mao YP, Ganten D, Dzau VJ, Lander ES (1991). Genetic mapping of a gene causing hypertensive rat. *Cell* 67,213-224.
- 5) Litt M, Luty JA (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44,397-401.
- 6) Edwards A, Civitello A, Hammond HA, Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* 49,746-756.
- 7) Weber JL, May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44,388-396.
- 8) Levinson G, Gutman GA (1987). High frequencies of short frame shifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res,* 15:5323-5338.
- 9) Strand M, Prolla TA, Liskay RM, Peters TD (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature,* 365:274-276.
- 10) Wang Z, Weber JL, Zhong G, Tanksley SD (1994). Survey of plant short tandem DNA repeats. *Theor Appl Genet,* 88;1-6.
- 11) Morgante M, Olivieri AM (1993). PCR-amplified microsatellites as markers in plant genetics. *Plant J.* 3:175-182.
- 12) Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996a). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225-238.
- 13) Powell W, Machray GC, Provan J (1996b). Polymorphism revealed by simple sequence repeats. *Trends Plant Sci,* 1:215-222.

- 14) Jones CJ, Edwards KJ, Castaglione S, Windfield MO, Sala F, Van De Wiel C, Bredemeijer G, Vosman B, Mathes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevski A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vasquez A, Karp A (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* 3: 381-390.
- 15) Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri AM (1994). Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome*, 37: 763-769.
- 16) Ashkenazi V, Chani E, Lavi U, Levy D, Hillel J, Veilleux RE (2001). Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. *Genome* 44:50-62.
- 17) Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996). Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theor. Appl. Genet.* 93:190-198.
- 18) Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997). Microsatellite and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. *Theor. Appl. Genet.* 94:773-781.
- 19) Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L., and Feldman, M.W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139:463-471.
- 20) Siniscalco Gigliano G, Di Finizio A, Caputo P, Cozzolino S (1998). *Cannabis* fingerprints by using Random Amplified Length Polymorphism DNA (RAPD). *Delpinoa ns*, 37:35.
- 21) Coyle HM, Ladd C, Palmbach T, Lee HC (2001). The green revolution: botanical contributions to forensics and drug enforcement. *CMJ.* 42(3):340-345.
- 22) Coyle HM, Shutler G, Abrams S, Hanniman J, Neylon S, Ladd C, Palmbach T, Lee HC (2003). A simple DNA extraction method for marijuana sample used in amplified fragment length polymorphism (AFLP) analysis. *J Forensic Science*, 48(2): 343-347.
- 23) Toth G, Gaspari Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genomes Research*, 10:967-981.
- 24) 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 Science International*, 131:65-74.

- 25) Hsieh H-M, Hou R-J, Tasi L-C, Wei C-S, Liy S-W, Huang L-H, Kuo Y-C, Linacre A, Lee JC-I (2003). A Highly polymorphic STR locus in *Cannabis sativa*. *Forensic Science International*, 131:53-58.
- 26) Edwards KJ, Backer JHA, Daly A, Jones C, Karp A (1996). Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques*: 20:5
- 27) Roy R, Steffens DL, Gartside B, Jang GY, Brumbaugh JA (1996). Producing STR Locus Patterns from Bloodstains and other Forensic Samples using an Infrared Fluorescent Automated DNA sequencer. *Journal of Forensic Sciences*. 41 (3)418-424.
- 28) Nei M (1987). Genetic distance and molecular phylogeny. In population genetics and fishery management. Edited by N. Ryman and F. Utter. University of Washington Press, Seattle. Pp. 193-223.
- 29) Paetkau D, Calvert W, Stirling I, Strobeck C (1995). Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* 4:347-354.
- 30) Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4:406-425.
- 31) Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *Evolution* 21: 550-570.
- 32) Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791.
- 33) Sokal R, Rohlf FJ (1995). *Biometry*. 3rd edition. W. H. Freeman and Co., New York.
- 34) Rohlf FJ (1998). NTSYS: numerical taxonomy and multivariate analysis system, version 2.1. Applied Biostatistics Inc., Setauket, N.Y.
- 35) Cornuet JM, Piry S (2000). TreeMaker v1.0.1: a Windows program for making NJ dendrograms and assigning a bootstrap value between the nodes. Computer software distributed by the authors.
- 36) Cordeiro GM, Taylor GO, Henry RJ (2000). Characterization of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. *Plant Science*, 155:161-168.
- 37) Rossetto M, McLauchlan A, Harriss FC, Henry RJ, Baverstock PR, Lee LS, Maguire TL, Edwards KJ (1999). Abundance and polymorphism of microsatellite markers in the tea tree (*Melaleuca alternifolia*, Myrtaceae). *Theor Appl Genet* 98: 1091-1098.

- 38) Jakše J, Kindlhofer K, Javornik B (2001). Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. *Genome*, 44:773-782.
- 39) Maguire, T.L., Edwards, K.J., Saenger, P., and Henry, R. (2000). Characterization and analysis of microsatellite loci in a mangrove species, *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae). *Theor Appl Genet*, 101:279-285.
- 40) Butler JM (2001). *Forensic DNA Typing*. Academic Press, CA, USA
- 41) Diwan N, Cregan PB (1997). Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.*, 95:723-733.
- 42) Ahmad, M (2002). Assessment of genomic diversity among wheat genotypes as determined by simple sequence repeats. *Genome*, 45:646-651.
- 43) Coudit R, Hubbel SP (1991) Abundance and DNA sequence of 2-base repeat regions in tropical tree genomes. *Genome* 34:66-71.
- 44) Senior ML, Murphy JP, Goodman MM, Stuber CW (1998). Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci*. 38:1088-1098.
- 45) Yu J-K, Mangor J, Thompson L, Edwards KJ, Slabaugh MB, Knapp SJ (2002). Allelic Diversity of simple sequence repeats among elite inbred lines of cultivated sunflower. *Genome*, 45:652-660.