

H. J. Alghanim · J. R. Almirall

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

Received: 29 December 2002 / Revised: 31 March 2003 / Accepted: 14 April 2003 / Published online: 13 June 2003

© Springer-Verlag 2003

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 11 microsatellite markers developed in this study were found to be useful for DNA typing and for assessing genetic relatedness in *Cannabis*.

Keywords *Cannabis sativa* · Marijuana · Microsatellite · Short tandem repeat (STR) · DNA fingerprinting · DNA typing · Genetic relatedness

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. CTCTCTCTCTCT) [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

H. J. Alghanim · J. R. Almirall (✉)
Department of Chemistry, International Research Forensic Institute,
Florida International University,
University Park, Miami, FL, 33314, USA
e-mail: almirall@fiu.edu

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, 12, 13, 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, 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'-GCGGTACCCGGGAAGCTTGG-3'
- Sau-L-B: 5'-GATCCCAAGCTTCCCGGTACCGC-3'

One µg 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 Gaithersburg, MD) in a final volume of 40 µl. The reaction mix was incubated at 4 °C for 72 h and then the ligation reaction was stopped by heating at 65 °C for 10 min. Excess linkers were removed with Performa DTR Gel Filtration Cartridges (Edge Biosystems, 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 µl of the denatured DNA was mixed with 1 µg of the Sau-L-A oligo, 474 µl of the hybridization buffer (50% formamide, 3X SSC, 25 mM Na-phosphate pH 7.0, and 0.5% SDS), and 500 µl of Dynabeads in 2X B&W buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA and 2 M NaCl). The hybridization reaction (1000 µl) 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 min each using buffer #1 (2X SSC and 0.01% SDS) at 42 °C followed by 3 washes for 3 min 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 µl 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, Piscataway, NJ) according to the manufacturer recommendations and concentrated by ethanol precipitation.

Cloning reaction and plasmid recovery

The cloning reaction was performed using the TOPO TA Cloning Kit (Invitrogen, Chicago, IL) following manufacturer's protocols. Sterile toothpicks were used to inoculate 96-well plates containing 100 µl of SOC broth with 100 µg/ml ampicillin. The broth culture plates were covered with aluminum foil tape and placed at 37 °C overnight. A 25 µl 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 µl of 10 mM Tris-HCl (pH 8). This was used as a template for insert PCR with M13 primers (Life Technologies, 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, 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 min. 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

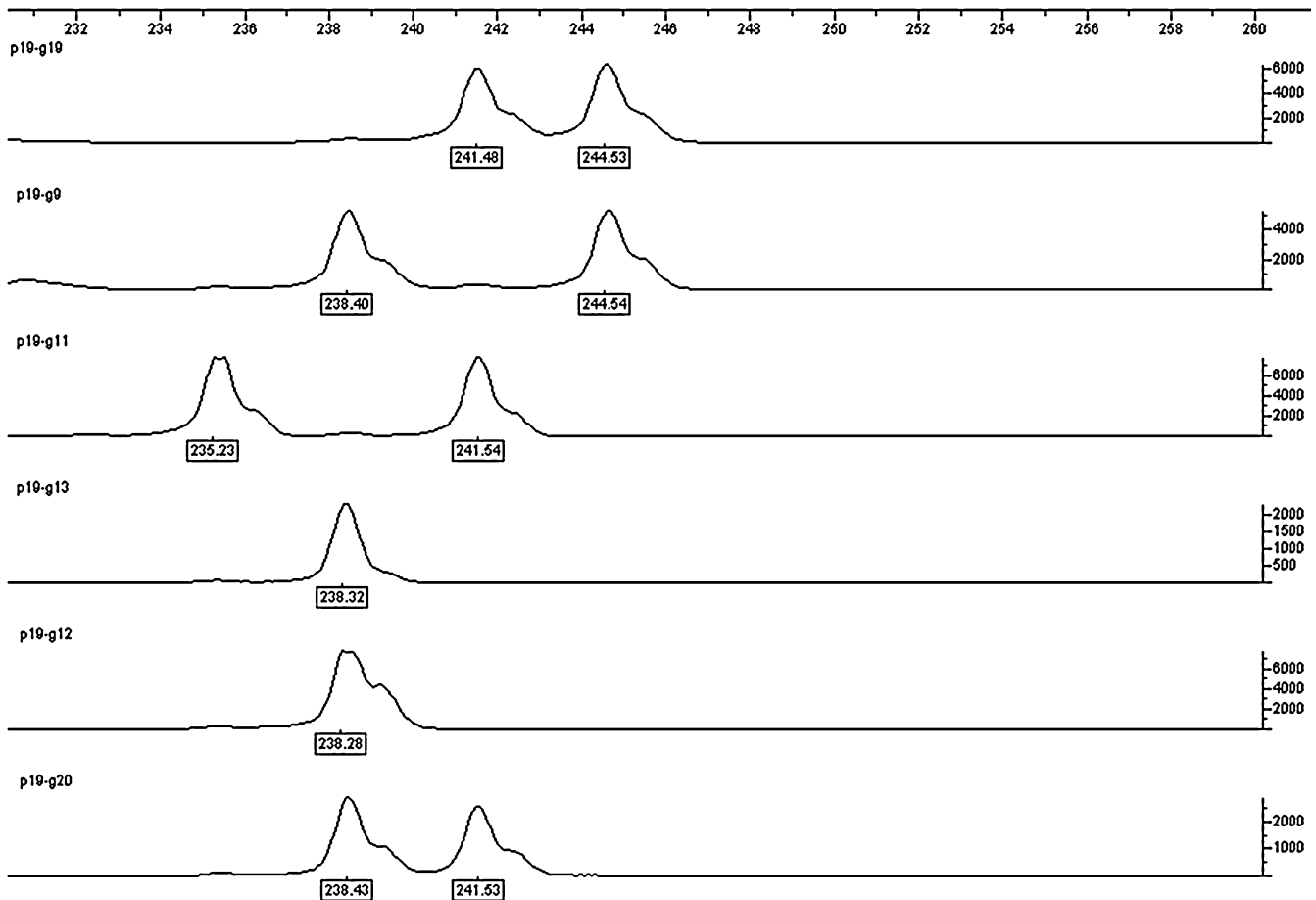
Based on the flanking sequences, PCR primers were designed using the GCG Wisconsin Packages (Accelrys, Madison, WI). After

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 15 mM MgCl₂, 0.2 µl of dNTP's (2.5 mM each), 0.25 µl of forward primer (10 µM), 0.25 µl of reverse primer (10 µM), 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 min, followed by 40 cycles of 45 sec. at 94 °C, 45 sec. at 53 °C, and 1 min at 72 °C. The final extension was 10 min 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 min. Samples were injected electrokinetically at 3 kV for 10 seconds and were run at 60 °C for 45 min at 15 kV. The data generated

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



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 Fig. 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 31 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]):

$$H_e = 1 - \sum P_i^2 \quad (1)$$

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]:

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

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):

$$n_e = (\sum p_i^2)^{-1} \quad (3)$$

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).

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).

Polymorphisms of the 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. Allele frequencies for 31 *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

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

Total clones of enriched DNA	685						
Total sequencing reactions	192						
Total microsatellite inserts (%)	49.5						
Frequency of microsatellites detected in <i>Cannabis</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 2 SSR loci and the primers developed in the study

Locus	Repeat	Primers (5'-3')	No. of alleles	Size range (bp)
C11-CANN1 (P4)	(GAT)8(GGT)7	GTGGTGGTGTATGATGATAATGG TGAATTGGTTACGATGGCG	5	150–175
B01-CANN1 (P6)	(GAA)13(A)(GAA)3	TGGAGTCAAATGAAAGGGAAC CCATAGCATTATCCCACTCAAG	5	323–339
D02-CANN2 (P7)	(CTT)6(ATT)(CTT)10	AGATGATGCCAAGAGGCAC TACCCATCCCCCTGGATCAC	4	221–236
C08-CANN2 (P9)	(GA)21	GCAAGAAGTAGAGAGACAATCC CCCTCAACACTCATTAECTCAC	9	171–203
H11CANN1 (P13)	(CT)18	GCATGTGGTTGTTTCGTACCC CAGCGAACATTCACTCTAGCTC	7	285–297
B02-CANN2 (P14)	(AAG)10	CAACCAAATGAGAATGCAACC TGTTTTCTTCACTGCACCC	3	163–172
H09-CANN2 (P15)	(GA)15	CGTACAGTGATCGTAGTTGAG ACACATACAGAGAGAGCCC	6	204–224
E07-CANN1 (P17)	(CTA)9	CAAATGCCACACCACCTTC GTAGGTAGCCAGGTATAGGTAG	3	105–111
B05-CANN1 (P19)	(TTG)9	TTGATGGTGGTCAAACGGC CCCCAATCTCAATCTCAACCC	4	235–244
D02-CANN1 (P24)	(GTT)7	GGTGGGATGTTGTTGTTGTG AGAAATCCAAGGTCCTGATGG	3	105–111
H06-CANN2 (P25)	(ACG)7	TGGTTTCAGTGGTCCTCTC ACGTGAGTGATGACACGAG	3	266–273

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.

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 (Fig. 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 (Fig. 3a). G30 exhibited a very distinct profile and could be easily recognized as an outlier sample in the PCA plot (Fig. 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 (Fig. 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 loci 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.

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)

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		
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	

*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

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

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

Locus	<i>Ho</i>	<i>He</i>	<i>N</i>	<i>n_e</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}

^aProduct of values across all loci analysis of *Cannabis* plant

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 plants 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 (Fig. 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). In addition, 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),

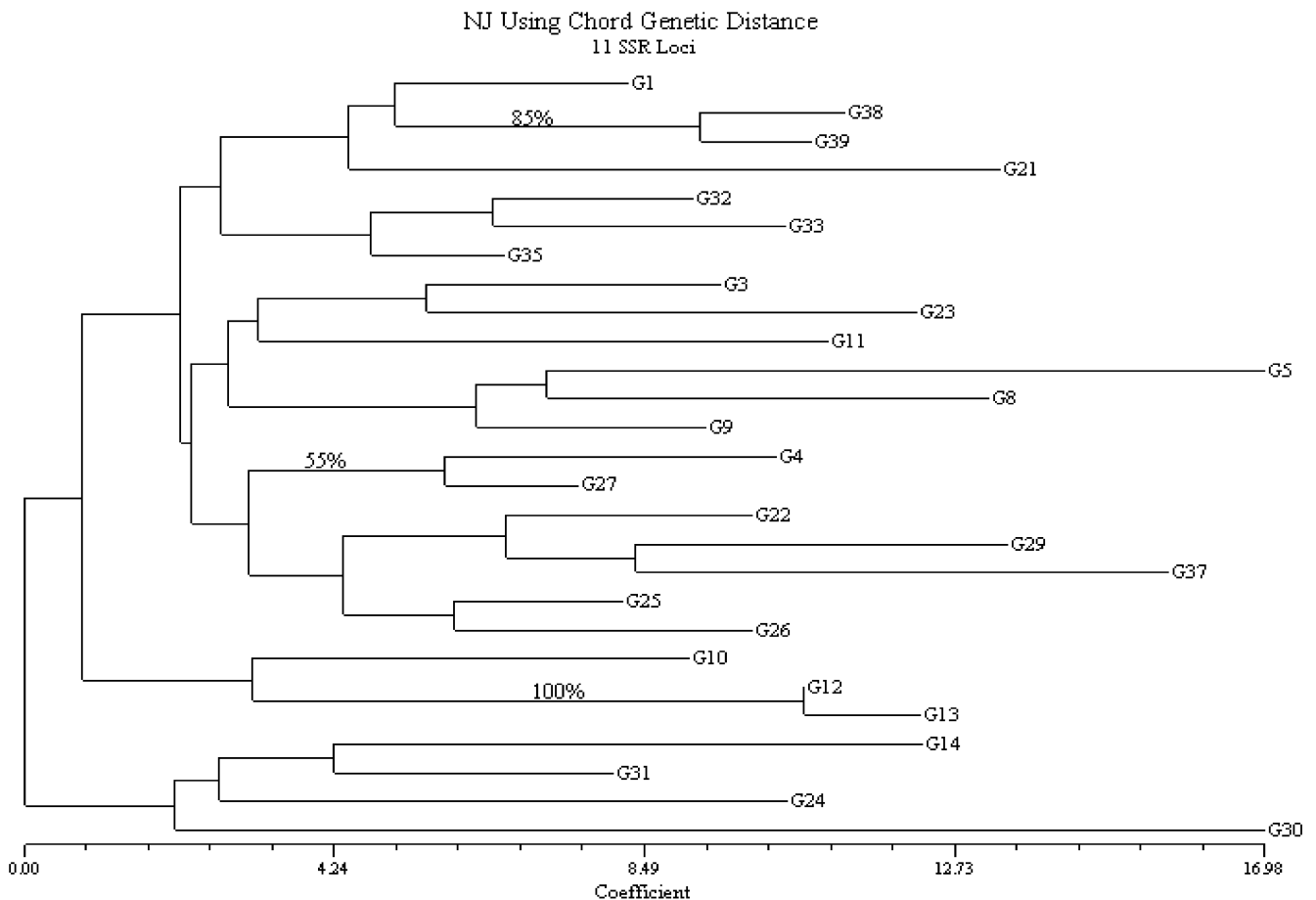


Fig. 2 Neighbor joining tree using Chord's genetic distance based on overall allele frequencies in 11 SSR loci across 27 *Cannabis* genotypes. Only the bootstrap value of 50% and above is shown

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% (Fig. 2). This relationship was in agreement with the information that both samples were closely related by AFLP analysis (Table 3). In addition, Fig. 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 (Fig. 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 (Fig. 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 (Fig. 2) and PCA (Fig. 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 (Fig. 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

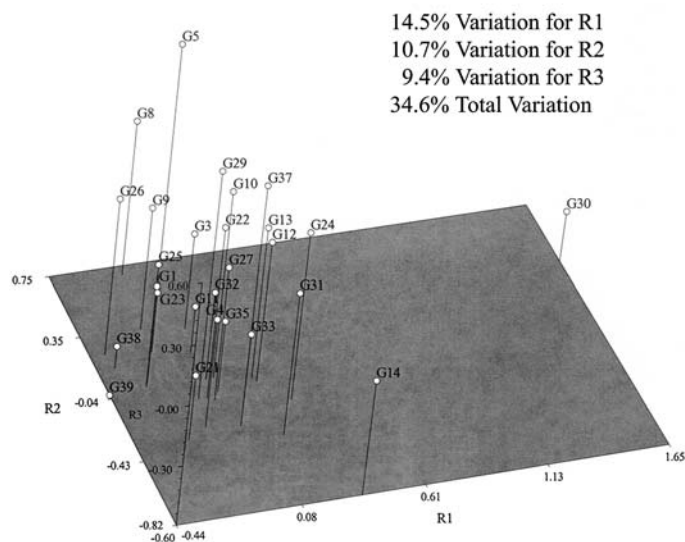
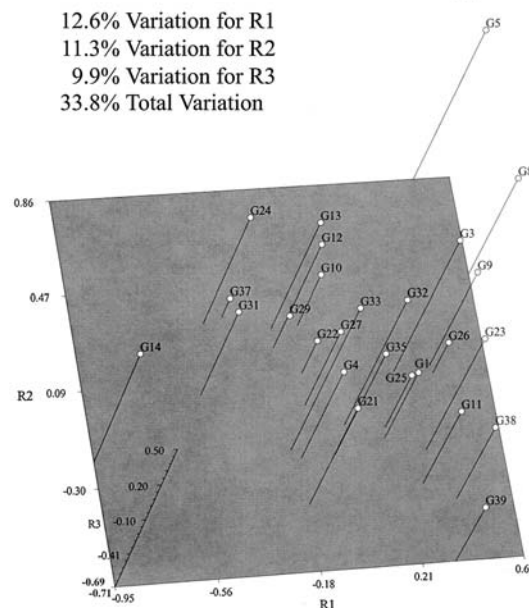
3-D PCA Plot for *Cannabis* Genotypes3-D PCA Plot for *Cannabis* Genotypes

Fig. 3 **a** Three-dimensional principal component analysis plot of the 27 unique *Cannabis* genotypes using correlation coefficients based on overall allele frequencies in 11 SSR loci. **b** Same as 3A except that G30 is eliminated from the plot.

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.

References

1. Drug Identification Bible (DIB) (2001). Amera-Chem, 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). *Mol Breed* 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, Marmioli 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 Finizo 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 Sci*, 48(2): 343–347.
23. Toth G, Gaspari Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genomes Res*, 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 Sci Int* 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 Sci Int* 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. *J Forensic Sci* 41 (3)418–424.
28. Nei M (1987) Genetic distance and molecular phylogeny. In: Ryman N, Utter F (eds). *Population genetics and fishery management*. 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 edn. W. H. Freeman, New York.
34. Rohlf FJ (1998). NTSYS: numerical taxonomy and multivariate analysis system, version 2.1. Applied Biostatistics 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 Sci* 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. Codit 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 if simple sequence repeats among elite inbred line of cultivated sunflower. *Genome*, 45: 652–660.