

Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations

Simon Gilmore^{a,b,*}, Rod Peakall^b, James Robertson^c

^aCentre for Forensic Science, Canberra Institute of Technology, G.P.O. Box 826, Canberra ACT 2601, Australia

^bSchool of Botany and Zoology, Australian National University, Canberra ACT 0200 Australia

^cForensic Services Division, Australian Federal Police, G.P.O. Box 401, Canberra ACT 2601, Australia

Received 3 May 2002; received in revised form 5 October 2002; accepted 15 October 2002

Abstract

Short tandem repeat (STR) markers are the DNA marker of choice in forensic analysis of human DNA. Here we extend the application of STR markers to *Cannabis sativa* and demonstrate their potential for forensic investigations.

Ninety-three individual cannabis plants, representing drug and fibre accessions of widespread origin were profiled with five STR makers. A total of 79 alleles were detected across the five loci. All but four individuals from a single drug-type accession had a unique multilocus genotype. An analysis of molecular variance (AMOVA) revealed significant genetic variation among accessions, with an average of 25% genetic differentiation. By contrast, only 6% genetic difference was detected between drug and fibre crop accessions and it was not possible to unequivocally assign plants as either drug or fibre type. However, our results suggest that drug strains may typically possess lower genetic diversity than fibre strains, which may ultimately provide a means of genetic delineation.

Our findings demonstrate the promise of cannabis STR markers to provide information on: (1) agronomic type, (2) the geographical origin of drug seizures, and (3) evidence of conspiracy in production of clonally propagated drug crops.

© 2003 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: *Cannabis sativa*; Microsatellites; Short tandem repeats (STR); Drug; Fibre

1. Introduction

Cannabis sativa L. is thought to have originated in the central Asia region [1], and has since been distributed worldwide by humans who have cultivated the plant as a source of fibre, feed, oils, medicines, and intoxicants for thousands of years [1,2]. Leaves and inflorescences (but not seeds) contain psychoactive drugs principally the cannabinoid Δ^9 -tetrahydrocannabinol (THC) [1,3]. The presence of cannabinoids is the reason cannabis is widely used as a recreational drug (marijuana) in many parts of the world.

Despite the wide range of possible uses for cannabis, the cultivation and possession of the plant is prohibited by law in many countries due to its intoxicant properties. The illegal status of the plant has generated a significant black-market for the drug, ranging from small-scale local production and distribution through to large-scale international operations run by organised crime syndicates. However, in some jurisdictions licensing arrangements are available for cultivation for fibre and seed oil [3]. Presently, there is growing interest in cannabis as a useful multipurpose crop that offers economic and environmental benefits over traditional fibre crops [4,5].

From a law enforcement perspective, the full scale agriculture of cannabis for fibre and seed oil poses a security problem, with the possibility of licensed crops being used as a cover for illegal drug crops. There is also the possibility of contamination of fibre crops with pollen

* Corresponding author. Tel.: +61-2-6207-4157;

fax: +61-2-6207-4105.

E-mail address: simon.gilmore@cit.act.edu.au (S. Gilmore).

of drug varieties, and the potential for theft and subsequent fraudulent distribution of agricultural types as drug types [6,7]. From an agricultural perspective, the inability to readily distinguish between fibre and drug cannabis varieties poses a major impediment to further development of the crop.

We have begun a long-term project to develop diagnostic DNA profiling methods for discriminating among varieties, populations and individuals of *C. sativa*. DNA-based methods offer the potential to aid the forensic community by enabling: (i) identification of botanical samples as *Cannabis*, (ii) identification of *Cannabis* samples as drug or fibre varieties and (iii) determination of the provenance of a seized *Cannabis* drug samples. Collectively, these methods thus promise powerful new investigative tools for intelligence analysis of organised and commercially motivated criminal activity.

Some progress has already been made towards species-specific DNA-based tests for *Cannabis* [8–12]. All of these techniques are applicable to trace evidence samples frequently encountered in forensics, but none provide individual or population level information. Random amplified polymorphic DNA (RAPD) markers have been applied to *C. sativa* [13–15] and other plants [16,17], and have been shown to discriminate among samples of different provenance and strain. However, these markers suffer from problems with specificity and reproducibility, and significantly, Jagadish et al. (unpublished) showed different profiles were obtained from RAPD analysis of different parts of cannabis plants (leaf tissue versus roots).

Short tandem repeats (STRs), also called simple sequence repeats (SSRs) and microsatellites, consist of tandemly repeated units of short nucleotide motifs, 1–6 bp long. These regions occur frequently and randomly throughout the genomes of plants and animals, and typically show extensive variation [18]. STRs are widely considered the genetic marker of choice for studies within species and have been applied in a wide range of studies of wild and cultivated plants [19–22], wild and domestic animals [23–25], and human populations [26–28].

STRs are now the standard forensic tool for the investigation of human-source forensic samples [29–33] and are applicable to trace evidence samples. Recently, there has also been an increase in the forensic application of STRs to cases involving non-human DNA, for example, to identify domestic canines [34] and felines [35]. The study we describe here extends the potential forensic applications of STRs to plant DNA. Other studies confirm that STRs in plants are highly variable and informative amongst individuals and populations [19,20] and are therefore, good candidates for forensic analysis.

The aims of this study were to: (i) develop a set of informative STR genetic marker assays in *C. sativa*, (ii) to test the utility of a subset of five STR genetic markers for their ability to discriminate among individuals and varieties of *C. sativa*.

2. Materials and methods

2.1. Sample selection

Cannabis and one other genus, *Humulus* (hops), are currently placed in the family Cannabaceae. However, recent molecular cladistic analysis suggest that Cannabaceae should be placed in the family Celtidaceae, which is a sister family to the Moraceae [36,37]. Although several synonyms have been published (e.g. *C. indica* Lam., *C. ruderalis* Jan), the genus *Cannabis* is considered to comprise a single highly variable species [1,3]. Because of the long association of *Cannabis* with humans, no wild progenitor is known, and in any case there has probably been widespread introgression and hybridisation between wild and cultivated strains [1].

Our intention in this study was to sample as broadly as possible across *C. sativa*. A total of 93 samples of wide geographic origin were therefore selected, representing six fibre crop varieties and nine drug crop varieties (Table 1). DNA was extracted from seedlings germinated from seed accessions held by the Australian Federal Police, at the Forensic Services Division, Canberra, Australia. These seed accessions were obtained for the study of Jagadish et al. [14] from the Centre for Plant Breeding and Reproductive Research (CPRO), The Netherlands, (see [38]). We have retained the accession numbers of these seed-lines in our study, but for simplicity refer to them by the abbreviated codes shown in Table 1.

2.2. DNA extractions

All DNA extractions were performed by a modification of the CTAB method of Doyle and Doyle [39]. Approximately, 100 mg of liquid nitrogen ground tissue (whole seedlings, or inflorescence in the case of 037929) was incubated in 500 μ l of CTAB extraction buffer at 65 °C. Following incubation the mixture was twice extracted with an equal volume of chloroform, and crude nucleic acids precipitated with the addition of 0.8 volumes of isopropanol.

Crude nucleic acids were recovered by centrifugation at 12,000 \times g for 10 min, before redissolving in 20 μ g/ml of RNase A with 1 \times reaction buffer (USB), with incubation at room temperature for 30 min to remove contaminating RNA.

The DNA was further purified by precipitating from solution using an equal volume of 13% polyethylene glycol (PEG-8000), 1.6 M NaCl [40]. The purified nucleic acid was recovered by centrifugation at 12,000 \times g for 10 min, and the dried pellet was redissolved in 200 μ l TE (Tris 100 mM, EDTA 1 mM pH 8.0).

Following extraction with one volume of buffered phenol: chloroform (1:1, pH 8.0) the aqueous fraction was recovered and the DNA precipitated by addition of 25 μ l 3 M sodium acetate and 450 μ l of 100% ethanol. DNA was pelleted by centrifugation at 12,000 \times g for 10 min, then dissolved in 50 μ l TE. Yield was quantified crudely by comparison with

Table 1

Summary of the accession numbers, accession code, cultivar status and country of origin for the samples genotyped in this study

Accession	Code	<i>n</i>	Country of origin	Cultivar	Notes
883041	F41	13	France	Fibrimon-56	Fibre cultivar
883042	F42	11	Former East Germany	Fibrimon	Fibre cultivar
883043	F43	11	Hungary	Fibrimon	Fibre cultivar
883291	F91	8	Former Soviet Union	SOU	Fibre cultivar
921122	F22	8	China		Fibre, unknown cultivar status
921207	F07	12	India		Landrace fibre crop
921232	D32	8	Mexico		Drug crop
921236	D36	8	Sierra Leone		Drug crop
891192	D92	3	Nepal		Wild, assumed to be drug-type
891197	D97	3	The Netherlands		Drug crop
991239	D39	4	Uganda		Drug crop
891196	D96	1	The Netherlands		Drug crop
910972	D72	1	The Netherlands		Drug crop
921234	D34	1	Zimbabwe		Drug crop
921240	W40	1	USA		Wild, assumed to be fibre type

All information sourced from the Centre for Plant Breeding and Reproductive Research (CPRO), The Netherlands (see [36]).

DNA of known concentration by electrophoresis in a 1% agarose gel containing 0.5 µg/ml of ethidium bromide. Electrophoresis products were visualised on an UV-transilluminator, and the DNA was diluted to approximately, 10 ng/µl for use.

2.3. Development of STR genetic markers

A genomic library was constructed using DNA from a single *C. sativa* sample held in the GAUBA herbarium at the Australian National University (accession number 037929), and screened with a probe cocktail comprising oligonucleotides of di-, tri-, and tetranucleotide repeats [41].

Positive clones were identified and the plasmid DNA sequenced using the ABI PRISM DyeDeoxy Terminator

kit. Sequence products were run on an ABI 377 automated sequencer and electropherograms were edited with the assistance of Sequencher sequence editing software (Genecodes).

Primers were designed for sequences that contained microsatellites greater than 10 repeats in length, and possessed sufficient flanking region to allow forward and reverse primers to be designed. Primer design was assisted by the software package PRIMER version 0.5 (Whitehead Institute for Biomedical Science).

The designed primers were screened for product-size variability and consistent amplification across the range of samples. Five primer pairs were selected for this initial study (Table 2). Primer sequences for these and additional loci are published elsewhere [42].

Table 2

Summary of STR locus characteristics including repeat motif, allelic size ranges, number of observed, potential and missing alleles and expected heterozygosity (He)

STR locus	ANUCS201	ANUCS301	ANUCS302	ANUCS303	ANUCS202
STR motif	(GA) ₂₆	(TTA) ₁₅	(CAA) ₁₁ ,..., (CAA) ₄	(TGG) ₇	(GA) ₂₀
Repeat unit	2	3	3	3	2
Size of sequenced STR	187	234	147	147	161
Min allele size (bp)	155	209	140	142	145
Max allele size (bp)	227	261	173	157	185
Inferred min STR	(GA) ₁₀	(TTA) ₇	(CAA) ₉	(TGG) ₅	(GA) ₁₂
Inferred max STR	(GA) ₄₆	(TTA) ₂₄	(CAA) ₂₀	(TGG) ₁₀	(GA) ₃₂
No. of observed alleles	29	16	10	6	18
No. of potential alleles	37	18	12	6	21
No. of missing alleles	8	2	2	0	3
He	0.928	0.902	0.758	0.659	0.902

All estimates are based on the total data set of 93 samples. Number of potential alleles were inferred by counting the number of repeat units between the smallest and largest observed alleles. The number of missing alleles were inferred by subtracting the number of observed alleles from the total number of potential alleles within the observed allele range.

2.4. PCR amplification and scoring of microsatellite loci

Approximately, 10 ng of template DNA extracted from a *C. sativa* seedling was used in a 10 μ l PCR containing 1 \times PCR buffer with MgCl₂ (Perkin-Elmer), 250 μ M dNTPs, 0.2 μ M each of forward and reverse primer, 0.04 μ l fluorescently labelled dUTP (either 400 μ M R110 or 100 μ M TAMRA; ABI Prism) and 0.75 units of TaqGold DNA Polymerase (Perkin-Elmer).

The reaction profile comprised a 10 min incubation at 90 °C to activate the TaqGold then a cycle of 95 °C for 15 s, 50 °C for 15 s and 72 °C for 60 s, repeated 30 times. Following cycling, the reaction was held at 72 °C for 5 min, before a final 4 °C holding temperature.

Alternatively, a 'touchdown' cycle was employed, comprising a 10 min incubation at 90 °C to activate the TaqGold,

then a cycle of 95 °C for 15 s, 65 °C for 15 s and 72 °C for 60 s. The cycle was repeated 35 times with the annealing temperature dropping 1 °C after each of the first 15 cycles, whereafter it was held at 50 °C. Following cycling, the reaction was held at 72 °C for 5 min, before a final 4 °C holding temperature.

Amplification products were precipitated by first adjusting the volume to 100 μ l with MilliQ purified water, addition of 15 μ l of 3 M sodium acetate and 300 μ l of 100% ethanol, before immediately centrifuging for 30 min at 12,000 \times g. The pellets were washed in 70% ethanol, then dried and resuspended in 6 μ l of MilliQ purified water.

For loading on the ABI 377 DNA sequencer, 0.5 μ l of resuspended PCR product was mixed with 0.6 μ l of deionised formamide, 0.2 μ l of Genescan-500 ROX size-marker (ABI) and 0.15 μ l of loading dye, supplied with the ROX

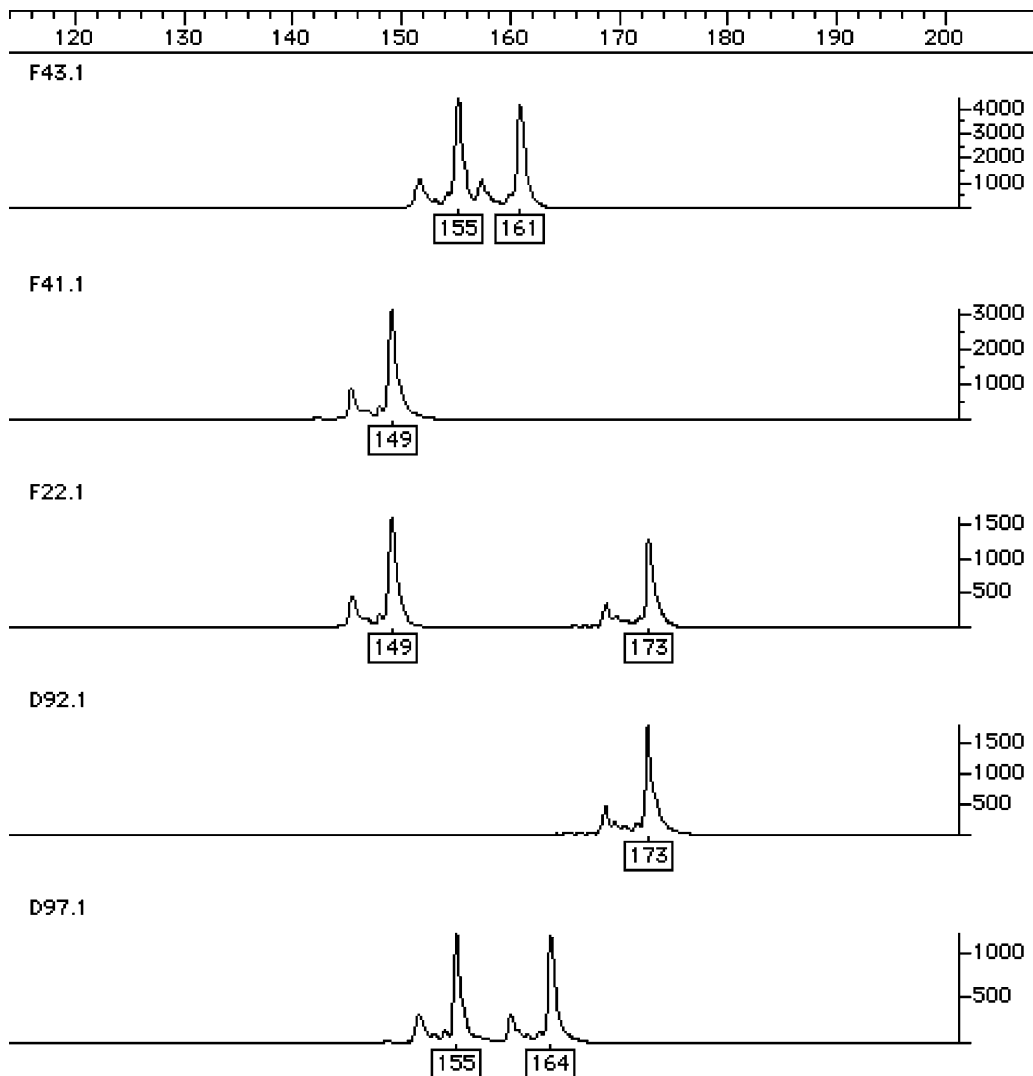


Fig. 1. Typical STR DNA profiles at the locus ANUCS302 showing homozygous and heterozygous individuals from five different accessions.

size-marker. Approximately, 0.8 μ l of this mixture was electrophoresed in a 5% polyacrylamide gel on the ABI 377 for 3 h.

2.5. Data scoring and statistical analysis

Following electrophoresis, GeneScanTM and GenotyperTM software (PE Applied Biosystems) were used to aid the scoring of the STR profiles. Homozygous and heterozygous genotypes were inferred from the banding patterns. Standard population genetic estimates of allele frequency, number of alleles, and observed and expected heterozygosity were calculated directly from these genotypes using our own software GenAlEx [43].

Genetic differentiation among accessions, as estimated by *F*_{st}, was calculated using the analysis of molecular variance framework developed by Excoffier et al. for haplotype data [44] and subsequently extended to codominant genotypic data [45,46] using the software GenAlEx V5 [43] and Arlequin V2.0 [47]. Test of significance were performed by random permutation with all permutational tests performed 1000 times.

3. Results

3.1. Total genetic diversity

A total of 93 samples of *C. sativa*, representing both fibre and drug crops, were successfully genotyped in this study (Table 1). In all cases, the STR profiles were typical of codominant markers with at most two different alleles per locus, and with homozygous and heterozygous individuals clearly identifiable (Fig. 1). Six accessions with 8–12 sam-

ples each represented fibre crops. Five additional accessions with 3–8 samples each represented drug crops. Four additional drug accessions represented only by a single sample were included in estimates of total genetic diversity, but excluded from population level analyses that required multiple samples per population.

Of the 93 samples genotyped, all but four individuals had unique multilocus genotypes. The exception occurred in the drug accession D39 for which all four samples shared the same homozygous multilocus genotype at five loci. A total of 79 alleles were detected in the full data set, with the number of alleles per locus varying from six alleles at locus ANUCS303 to 29 alleles at ANUCS201 (Table 2). The mean total genetic diversity was 0.83 ± 0.05 (mean \pm S.E.), with a range of 0.66–0.92 per locus (Table 2).

3.2. Locus by locus patterns of allelic diversity

Table 2 shows the repeat motif of each locus, the observed allele size range in base pairs (bp) and the inferred repeat sequence for the smallest and largest alleles detected, assuming allele size varies simply by addition or deletion of repeat motifs as typical for STR loci. This latter assumption is supported by the data. For all five loci the observed size of the STR alleles varied in multiples of their expected repeat units. For example, the most diverse locus, ANUCS201 was a dinucleotide (GA)_n repeat with 29 alleles varying in size in multiples of 2 bp units from 155 to 227 bp. The allele frequency distribution for this locus is shown in Fig. 2. Note that within the large 72 bp allelic size range spanned by this locus, all but eight of the potential alleles were detected in our sample set (Fig. 2, Table 2). The three trinucleotide loci (ANUCS301, ANUCS302 and ANUCS303), although revealing fewer alleles per locus

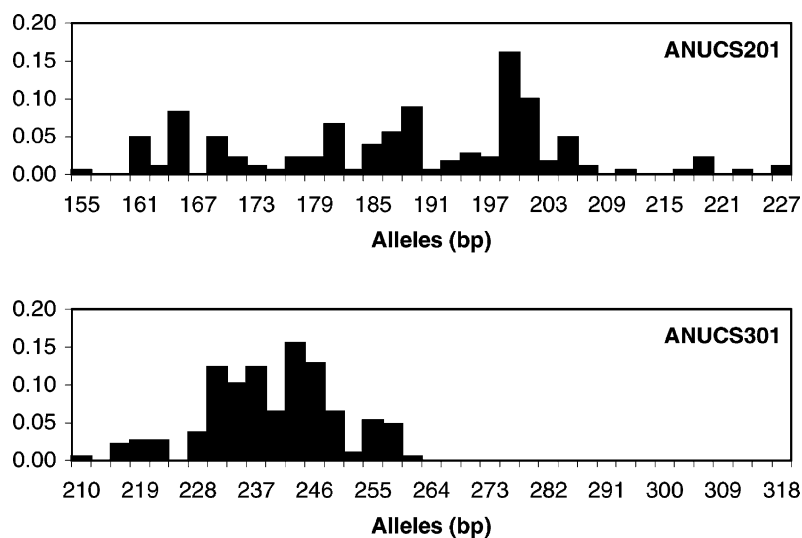


Fig. 2. Allele frequency distributions for the STR locus ANUCS201, a dinucleotide repeat (GA)_n and ANUCS301, a trinucleotide repeat (TTA)_n for the total set of 93 samples.

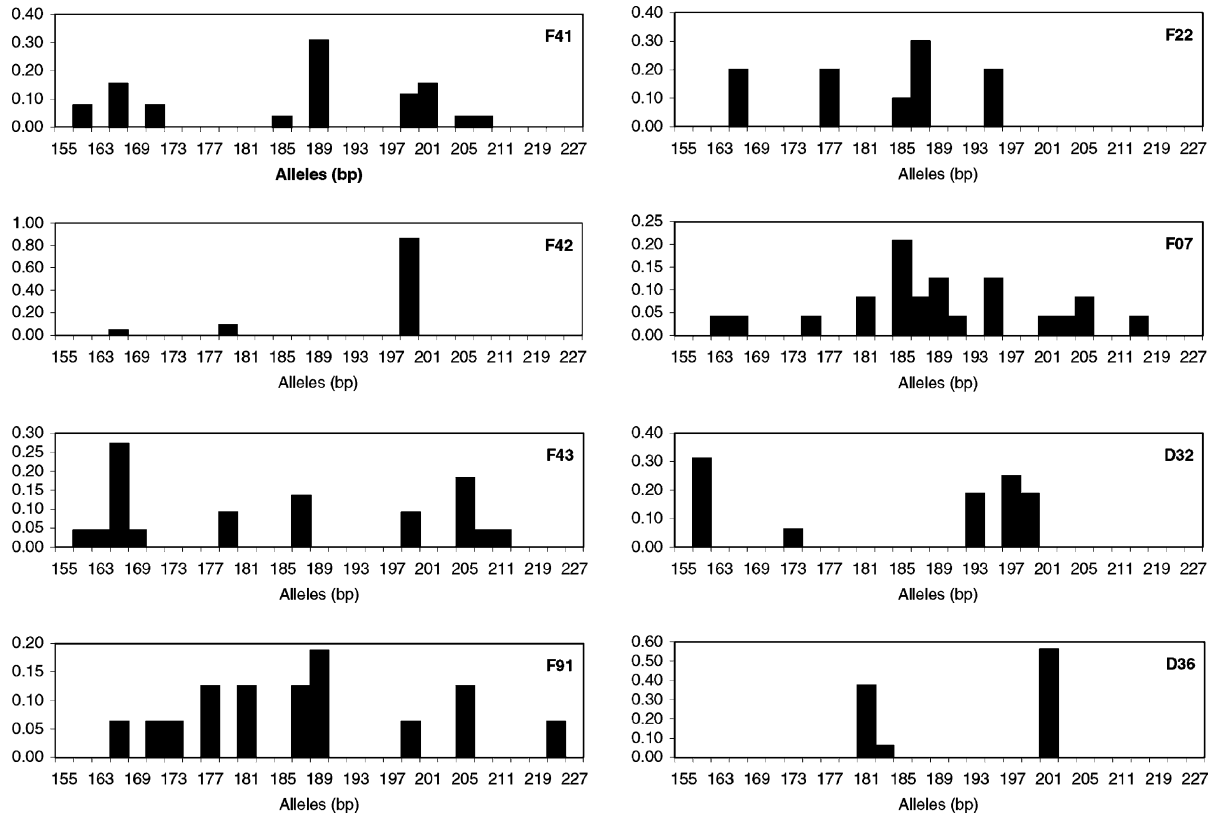


Fig. 3. Allele frequency distributions of all accessions with eight or more samples for the STR locus ANUCS201.

than the two dinucleotide loci, nonetheless revealed allelic patterns consistent with that expected for trinucleotide repeats. At ANUCS301, for example, alleles varied in size by multiples of 3 units from 209 to 261 bp. Within this 52 bp range, all but two of the expected alleles were observed within the range of allelic sizes observed (Fig. 2, Table 2).

3.3. Patterns of genetic diversity within accessions

For illustrative purposes, the allele frequency distributions for the locus ANUCS201 are shown in Fig. 3 for all accessions with eight or more samples. A visual inspection reveals considerable variation within and among accessions in both the number of alleles detected and allele frequency. The fibre crop F42 is notable for its lack of allelic diversity, with only one common allele and two rare alleles. Similarly, the drug crop D36, is characterised by the presence of only three alleles. By contrast, for other accessions several to many alleles were detected, with the fibre crop F07 showing 13 alleles at this locus. Similar variation in the number of alleles and allele frequencies among the accessions was apparent at the other four loci (data not shown), with F42 and all drug accessions showing lower genetic diversity.

3.4. Patterns of genetic diversity among accessions

As expected, given the apparent variation in allele numbers and allele frequency distributions among accessions, an analysis of molecular variance (AMOVA) revealed significant genetic differentiation ($P < 0.001$), with an average of 25% of the total genetic variance ($F_{st} = 0.248$) accounting for differences among accessions. The degree of genetic differentiation was significantly different among all pairwise

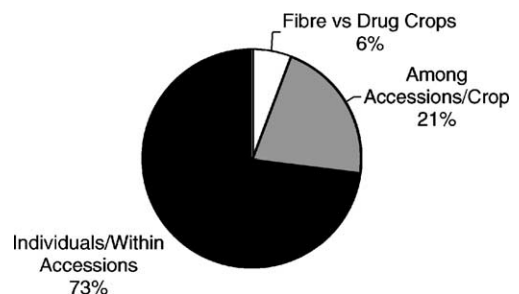


Fig. 4. Results of AMOVA showing the percentage of the total genetic variance accounted for by differences among fibre and drug crops, among accessions within drug and fibre crops, and among individuals within accessions.

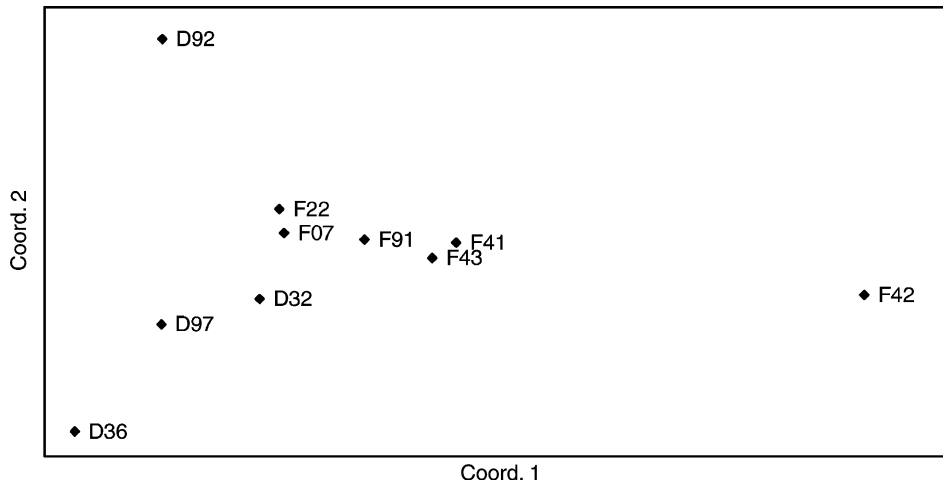


Fig. 5. Plot of the first two axes of a principal coordinate analysis of the pairwise F_{st} matrix among accessions. The monomorphic drug accession D39 is located off the graph view.

Table 3
Pairwise F_{st} values among all combinations of accessions

	F41	F42	F43	F91	F22	F07	D32	D36	D92	D97	D39
F41	*										
F42	0.207	*									
F43	0.088	0.216	*								
F91	0.060	0.258	0.084	*							
F22	0.167	0.338	0.150	0.135	*						
F07	0.156	0.282	0.123	0.095	0.101	*					
D32	0.254	0.394	0.186	0.188	0.208	0.136	*				
D36	0.294	0.495	0.294	0.282	0.290	0.220	0.319	*			
D92	0.267	0.480	0.275	0.237	0.222	0.170	0.336	0.377	*		
D97	0.243	0.438	0.214	0.138	0.180	0.151	0.231	0.280	0.341	*	
D39	0.422	0.635	0.453	0.424	0.444	0.375	0.507	0.553	0.718	0.617	*

Tests of significance by random permutation showed all comparisons of accessions with eight or more samples, F41 to D36 inclusive are significantly different ($P < 0.007$). Statistical tests were not performed for the accessions D92, D97 and D39, due to their small samples sizes.

combinations of accessions with eight or more samples ($P < 0.007$ in all cases), and varied from the minimum of 5.8% between F41 and F91 ($F_{st} = 0.060$, $P < 0.007$) to a maximum of 72% between F42 and D39 ($P < 0.001$) (see Table 3).

Although significantly different ($P < 0.001$), the genetic differentiation among the fibre and drug crops accounts for only 6% of the total variance, compared with an average of 21% difference among accessions (Fig. 4). Thus, average differences among the accessions are more than three times greater than among the fibre versus drugs groups.

The patterns of genetic relationship as revealed by a principal co-ordinate analysis (PCA), based on the matrix of pairwise F_{st} values, is shown in Fig. 5. The first two axes account for 48% of the total variation. All the fibre accessions group together, while the drug accessions are more widely dispersed in the PCA space but they are nonetheless more genetically similar to the central fibre crops than F42,

demonstrating there is no clear split of drug and fibre crops in our data set.

4. Discussion

4.1. Overview

We have demonstrated that STR genetic markers can be highly polymorphic within *C. sativa*, and capable of discriminating among individuals and varieties of cannabis. With the exception of one drug accession, unique multilocus genotypes were detected for all individuals and significant differentiation was detected among all accessions. The distribution of allele sizes suggested that all loci behaved as pure STRs consisting of uninterrupted repeats, with alleles in multiples of the tandem repeat unit, although DNA sequencing is required to confirm this assumption [19].

Although we detected almost as many alleles (79) as there were individuals (93), it is likely that additional alleles will be uncovered with a more extensive genetic survey. Based on the allelic series for each locus we predicted the number of potential alleles within the allelic range observed, and by subtraction inferred the number of ‘missing alleles’. In total, 15 missing alleles were predicted across the five loci, (Table 2). It is anticipated that these ‘missing’ alleles will be encountered in further analyses.

A more extensive genetic survey may also extend the allele size range at these five loci, although minimum allele sizes are limited by the number of repeats in the STR region. In general, STR polymorphism tends to increase with increased repeat number, with the lower limit for polymorphism of 3–5 repeats reported in plants [19]. At the trinucleotide repeat loci ANUCS303 and ANUCS301 the number of repeat units in the smallest alleles (five and seven, respectively) may already have reached the lower limit for polymorphism. At the other extreme, there is the potential for longer alleles than presently observed, however, at some unknown point long pure repeats become unstable [19]. While there is potential for more alleles to be uncovered at these five STR loci, the addition of further STR loci with similar levels of hypervariability will greatly extend the power of these genetic markers.

The large numbers of alleles observed at our five STR loci, and the high heterozygosities (>0.9 for three of the five loci) are a feature of many STR loci in humans [48], birds [49], small mammals [50], insects [51] and plants [20]. Such extreme polymorphism is rarely observed at other sites in the genomes of animals and plants. While hypervariability at STR loci is common, it is by no means a foregone conclusion in plants, where STR frequency in the genome is generally low, and informative STRs can be difficult to find [52,53].

4.2. Differentiation of fibre versus drug varieties

While significant differentiation was detected between fibre and drug crops, the degree of genetic differentiation between the two crops was three times less than the average between accessions (Fig. 4). This pattern of genetic variance partitioning is the converse of that expected if the drug and fibre groups represented discrete genetic entities. For example, in a comparable study of Australian populations of the mangrove, *Avicennia marina* with similar sample sizes and number of STR loci, Maguire et al. [54] detected almost two times more genetic differentiation among subspecies, than among populations within subspecies. Thus, given that in *C. sativa* the genetic differences between accessions is much greater than between fibre and drug groups, we conclude that drug and fibre strains of cannabis are not discrete genetic entities.

Our inability to fully discriminate amongst drug and fibre accessions matches failed attempts to make intraspecific delimitations of *C. sativa* based on cannabinoid content [1,3]. Delimiting sub-specific groups using other morphological

characteristics has also proved elusive [55]. Our genetic findings re-enforce the current opinion that *C. sativa* comprises a single highly variable species [1,3].

4.3. Implications for forensic application

We previously noted that the needs of the forensic community include the ability to identify cannabis samples as drug or fibre, and the need to identify the provenance of a seized drug crop. Do our results suggest either is achievable? At face value our inability to readily discriminate between drug and fibre crops is disappointing, although perhaps not unexpected, given the collective evidence indicates that the fibre versus drug split is an artificial rather than a natural genetic grouping. Do STR markers still hold promise as powerful forensic tools in cases involving cannabis? The answer is most definitely yes. Our findings clearly demonstrate the potential of as few as five STR loci to discriminate among individuals and accessions of cannabis. This potential will be greatly improved by the addition of further STR loci, a task already underway in our laboratory [42].

With an expanded set of STR loci, and a large database of genotyped accessions, it will become increasingly feasible to compare an unknown to genotypes in the database, in much the same way as human DNA profiles are compared to nationwide databases, e.g. in Australia, the US and the Great Britain. Such a search may indicate the likely origin of a sample, for example, providing information regarding the local region of crop production, or whether a seizure represents a locally produced or imported crop. While it now appears unlikely that a single diagnostic genetic marker will ever be available for discrimination of drug versus fibre samples, our findings nonetheless suggest a promising DNA test for whether a sample is likely of drug or fibre origin. If lower genetic diversity characterises drug compared with fibre crops, as predicted from results in this study, this will enable a rapid test based on the assessment of the degree of genetic diversity at a selected subset of informative STR loci. Such a test will be most powerful when multiple individuals from a crop can be screened, but with sufficient STR loci it may also be feasible to assess levels of genetic diversity across loci in a single individual.

There is good reason to believe lower genetic diversity within drug crops has a biological basis, since it is a widespread practice in drug crop cultivation to eliminate male plants from a crop, thus preventing or reducing the seed production in the inflorescences, which increases the value of the final product. This practice is predicted to both reduce genetic diversity within drug lines, and to accentuate the genetic differences among lines. These accentuated genetic differences are likely to enhance the power of STR markers for identifying the provenance of a drug seizure.

Another forensic application that holds considerable promise with STR markers is the identification of hydroponically propagated clonal drug lines. In Australia, there is anecdotal evidence that in some jurisdictions that have

decriminalised the propagation of a small number of cannabis plants for personal use, criminally motivated consortia are exploiting the laws to produce a series of small hydroponic crops spread over several discrete residential locations to avoid charges of commercial production. STR markers may permit the identity of clonally propagated lines, providing evidence to link illegal operators in a conspiracy.

4.4. Future directions

We have demonstrated that the first five STR genetic markers developed for *C. sativa* are both hypervariable and informative, allowing discrimination among individuals and accessions. We have also strengthened support for the emerging view that the distinction between drug and fibre varieties of cannabis is artificial. As a consequence, it is unlikely that a single magic genetic marker will be able to unequivocally discriminate among drug and fibre samples. However, if drug crops are typically genetically depauperate compared with fibre crops, small number of STR marker assays may still prove to be highly discriminatory of these two types of crops. Despite the promise, our study also suggests that as in human forensics applications, it will be necessary to establish a robust STR genotype database of both drug and relevant fibre samples to enable the full forensic potential of these genetic markers to be realised. Work continues in our laboratory towards creating such a database. It is likely that different databases will need to be established in different countries, based on samples most relevant to the country. However, a network of databases would open up the possibility to study and compare the genetic diversity of cannabis samples originating from across its wide geographic range.

Acknowledgements

We are grateful to Mojca Keglovic, Lauretta Grasso, Christine Hayes and Dan Ebert for laboratory assistance. Funding was provided by the Australian Research Council (ARC) through a Strategic partnership with Industry Research and Training Scheme (SPIRT) grant. We also thank our industry partners, the Australian Federal Police, Canberra Institute of Technology and Ecofibre Pty Ltd for their financial and in-kind support.

References

- [1] E. Small, A. Cronquist, A practical and natural taxonomy for *Cannabis*, *Taxon* 25 (1976) 405–435.
- [2] I.R. Adams, B.R. Martin, Cannabis: pharmacology and toxicology in animals and humans, *Addiction* 91 (11) (1996) 1585–1614.
- [3] G. Siniscalco Gigliano, *Cannabis sativa* L.—botanical problems and molecular approaches in forensic investigations, *Forensic Sci. Rev.* 13 (1) (2001) 1–17.
- [4] P.C. Struik, S. Amaducci, M.J. Bullard, N.C. Stutterheim, G. Venturi, H.T.H. Cromack, Agronomy of fibre hemp (*Cannabis sativa* L.) in Europe, *Ind. Crop Prod.* 11 (2000) 107–118.
- [5] H.M.G. van der Werf, E.W.J.M. Mathijssen, A.J. Haverkort, The potential of hemp (*Cannabis sativa* L.) for sustainable fibre production: a crop physiological appraisal, *Ann. Appl. Biol.* 129 (1996) 109–123.
- [6] Criminal Justice Commission, Report on Cannabis and the Law in Queensland, 1994.
- [7] Legislative Assembly of Queensland, Parliamentary Criminal Justice Committee, Report on a Review of the Criminal Justice Commission's Report on Cannabis and the Law in Queensland, Report No. 37.
- [8] G. Siniscalco Gigliano, Identification of *Cannabis sativa* L. (Cannabaceae) using restriction profiles of the Internal Transcribed Spacer II (ITS2), *Sci. Justice* 38 (1998) 225–230.
- [9] G. Siniscalco Gigliano, P. Caputo, S. Cozzolino, Ribosomal DNA analysis as a tool to identify specimens of *Cannabis sativa* L. of forensic interest, *Sci. Justice* 37 (1997) 171–174.
- [10] G. Siniscalco Gigliano, Preliminary data on the usefulness of Internal Transcribed Spacer I (ITS1) sequence in *Cannabis sativa* L. identification, *J. Forensic Sci.* 44 (3) (1999) 475–477.
- [11] A. Linacre, J. Thorpe, Detection and identification of cannabis by DNA, *Forensic Sci. Int.* 91 (1998) 71–76.
- [12] M. Wilkinson, A.M.T. Linacre, The detection and persistence of *Cannabis sativa* DNA on skin, *Sci. Justice* 40 (1) (2000) 11–14.
- [13] R. Gillan, M.D. Cole, A. Linacre, J.W. Thorpe, N.D. Watson, Comparison of *Cannabis sativa* by random amplification of polymorphic DNA (RAPD) and HPLC of cannabiniods; a preliminary study, *Sci. Justice* 35 (1995) 169–177.
- [14] V. Jagadish, J. Robertson, A. Gibbs, RAPD analysis distinguishes *Cannabis sativa* samples from different sources, *Forensic Sci. Int.* 79 (1996) 113–121.
- [15] V. Faeti, G. Mandolino, P. Ranalli, Genetic diversity of *Cannabis sativa* L. germplasm based on RAPD markers, *Plant Breed.* 115 (1996) 367–370.
- [16] R. Mestel, Murder trial features tree's genetic fingerprint, *New Scientist* 29 (1993) 6.
- [17] L. Congui, M. Chicca, R. Cella, R. Rossi, G. Bernacchai, The use of random amplified polymorphic DNA (RAPD) markers to identify strawberry varieties: a forensic application, *Mol. Ecol.* 9 (2000) 229–232.
- [18] P. Jarne, P.J.L. Lagoda, Microsatellites, *Trends Ecol. Evol.* 11 (1996) 424–429.
- [19] R. Peakall, S. Gilmore, W. Keys, M. Morgante, A. Rafalski, Cross-species amplification of Soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants, *Mol. Biol. Evol.* 15 (10) (1998) 1275–1287.
- [20] P.K. Gupta, I.S. Balyan, P.C. Sharma, B. Ramesh, Microsatellites in plants—a new class of molecular markers, *Curr. Sci.* 70 (1996) 45–54.
- [21] Y. Miyazaki, Y. Isagi, Pollen flow and intrapopulation genetic structure of *Heloniopsis orientalis* on the forest floor as determined using microsatellite markers, *Theor. Appl. Gen.* 101 (2000) 718–723.
- [22] J. Provan, W. Powell, R. Waugh, Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*), *Theor. Appl. Gen.* 92 (1996) 1078–1084.

- [23] M.D.B. Eldridge, K.E. Kinnear, M.L. Onus, Source population of dispersing rock wallabies (*Petrogale lateralis*) identified by assignment tests on multilocus genotypic data, *Mol. Ecol.* 10 (12) (2001) 2867–2876.
- [24] S. Nyakaana, P. Arctander, Population genetic structure of the African elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male biased gene flow, *Mol. Ecol.* 8 (7) (1999) 1105–1115.
- [25] M.T. Koskinen, P. Bredbacka, Assessment of the population structure of five Finnish dog breeds with microsatellites, *Anim. Genet.* 31 (5) (2000) 310–317.
- [26] S. Ingvarsson, V. Finnsdottir, A. Sigurdsson, G. Geirsson, Population studies and validation of paternity determinations by six microsatellite loci, *J. Forensic Sci.* 45 (3) (2000) 692–695.
- [27] J. Hawks, K. Huntley, S.H. Lee, M. Wolpoff, Population bottlenecks and Pleistocene human evolution, *Mol. Biol. Evol.* 17 (1) (2000) 2–22.
- [28] M. Kayser, M. Krawczak, L. Excoffier, P. Dieltjes, D. Corach, V. Pascali, C. Gehrig, L.F. Bernini, J. Jespersen, E. Bakker, L. Roewer, P. de Knijff, An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations, *Am. J. Hum. Genet.* 68 (4) (2001) 990–1018.
- [29] C. Kimpton, D. Fisher, S. Watson, M. Adams, A. Urquhart, J. Lygo, P. Gill, Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci, *Int. J. Legal Med.* 106 (1994) 302–311.
- [30] D.J. Balding, M. Greenhalgh, R.A. Nichols, Population genetics of STR loci in Caucasians, *Int. J. Legal Med.* 108 (1996) 300–305.
- [31] I.W. Evett, P.D. Gill, J.K. Scrannage, B.S. Weir, Establishing the robustness of short tandem repeat statistics for forensic applications, *Am. J. Hum. Genet.* 58 (1996) 398–407.
- [32] P.J. Lincoln, J. Thomson (Eds.), *Forensic DNA Profiling Protocols*, Humana Press, Totowa, NJ, 1998.
- [33] D. Watts, Genotyping STR loci using an automated DNA sequencer, in: P.J. Lincoln, J. Thomson (Eds.), *Forensic DNA Profiling Protocols*, Humana Press, Totowa, NJ, 1998, pp. 193–208.
- [34] P. Savolainen, L. Arvestad, J. Lundeberg, A novel method for forensic DNA investigations: repeat type sequence analysis of tandemly repeated mtDNA in domestic dogs, *J. Forensic Sci.* 45 (5) (2000) 990–999.
- [35] M. Menotti-Raymond, V.A. David, J.C. Stephens, L.A. Lyons, S.J. O'Brien, Genetic individualisation of domestic cats using feline STR loci for forensic applications, *J. Forensic Sci.* 42 (6) (1997) 1039–1051.
- [36] S.J. Weigrefe, K.J. Sytsma, R.P. Guries, The Ulmaceae, one family or two? Evidence from chloroplast DNA restriction site mapping, *Pl. Syst. Evol.* 210 (1998) 249–270.
- [37] B.-H. Song, X.-Q. Wang, F.-Z. Li, D.-Y. Hong, Further evidence for the paraphyly of the Celtidaceae from the chloroplast gene *matK*, *Pl. Syst. Evol.* 228 (2001) 107–115.
- [38] E.P.M. de Meijer, L.J.M. van Soest, The CPRO *Cannabis* germplasm collection, *Euphytica* 62 (1992) 201–211.
- [39] J.J. Doyle, J.L. Doyle, Isolation of plant DNA from fresh tissue, *Focus* 12 (1990) 13–15.
- [40] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, vol. 1, 2nd ed., Cold Spring Harbor Press, New York, 1989, pp. 40–41.
- [41] I.A. Scott, C.M. Hayes, J.S. Keogh, S.F. Morrison, Isolation and characterisation of novel microsatellite markers from the Australian water skink *Eulamprus kosciuskoi* and cross-species amplification in other members of the species-group, *Mol. Ecol. Notes* 1 (2001) 28–30.
- [42] S. Gilmore, R. Peakall, Microsatellites in *Cannabis sativa* L., *Mol. Ecol. Notes*, submitted for publication.
- [43] R. Peakall, P.E. Smouse, GenAIEx V5: Genetic Analysis in Excel. Population genetic software for teaching and research, Australian National University, Canberra, Canada, 2001, <http://www.anu.edu.au/BoZo/GenAIEx/>.
- [44] L. Excoffier, P.E. Smouse, J.M. Quattro, Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction sites, *Genetics* 131 (1992) 479–491.
- [45] R. Peakall, P.E. Smouse, D.R. Huff, Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss (*Buchloë dactyloides* (Nutt.) Engelm.), *Mol. Ecol.* 4 (1995) 135–147.
- [46] P.E. Smouse, R. Peakall, Spatial autocorrelation analysis of multi-allele and multi-locus genetic micro-structure, *Heredity* 82 (1999) 561–573.
- [47] S. Schneider, J. Kueffer, D. Roessli, L. Excoffier, A software for population genetic data analysis, *Genetics and Biometry Laboratory*, University of Geneva, Switzerland, 1997.
- [48] J.L. Weber, Human DNA polymorphisms and methods of analysis, *Curr. Opin. Biotech.* 1 (1990) 166–171.
- [49] C.R. Primmer, A.P. Moller, H. Ellegren, A wide-range survey of cross-species microsatellite amplification in birds, *Mol. Ecol.* 5 (1996) 365–378.
- [50] D. Lindenmayer, R. Peakall, The Tumut experiment-integrating demographic and genetic studies to unravel fragmentation effects: a case study of the native Bush Rat, in: A. Young, G. Clarke (Eds.), *Genetics, Demography and the Viability of Fragmented Populations*, Cambridge University Press, New York, 2000, pp. 173–201.
- [51] D.C. Queller, J.E. Strassmann, C.R. Hughes, Microsatellites and kinship, *TREE* 8 (1993) 285–288.
- [52] M.V. Katti, P.K. Ranjekar, V.S. Gupta, Differential distribution of simple sequence repeats in eukaryotic genomes, *Mol. Biol. Evol.* 18 (7) (2001) 1161–1167.
- [53] L. Zane, L. Bargelloni, T. Patarnello, Strategies for microsatellite isolation: a review, *Mol. Ecol.* 11 (2002) 1–16.
- [54] T.L. Maguire, R. Peakall, P. Saenger, Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs, *Theor. Appl. Genet.* 104 (2002) 388–398.
- [55] E.P.M. de Meijer, H.J. van der Kamp, F.A. van Eeuwijk, Characterisation of cannabis accessions with regard to cannabinoid content in relation to other plant characters, *Euphytica* 62 (1992) 187–200.