Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International 172 (2007) 179–190
Forensic Science International
www.elsevier.com/locate/forsciint

Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of *Cannabis sativa*

Simon Gilmore [a,b,*], Rod Peakall [b], James Robertson [c]

[a] *Centre for Forensic Science, Canberra Institute of Technology, GPO Box 826, Canberra, ACT 2601, Australia*
[b] *School of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia*
[c] *Technical and Forensic, Australian Federal Police, GPO Box 401, Canberra, ACT 2601, Australia*

Received 7 May 2006; received in revised form 27 September 2006; accepted 10 October 2006
Available online 12 February 2007

Abstract

Comparative sequencing of cannabis individuals across 12 chloroplast and mitochondrial DNA loci revealed 7 polymorphic sites, including 5 length variable regions and 2 single nucleotide polymorphisms. Simple PCR assays were developed to assay these polymorphisms, and organelle DNA haplotypes were obtained for 188 cannabis individuals from 76 separate populations, including drug-type, fibre-type and wild populations. The haplotype data were analysed using parsimony, UPGMA and neighbour joining methods. Three haplotype groups were recovered by each analysis method, and these groups are suggestive of the crop-use characteristics and geographical origin of the populations, although not strictly diagnostic. We discuss the relationship between our haplotype data and taxonomic opinions of cannabis, and the implications of organelle DNA haplotyping to forensic investigations of cannabis.

Keywords: Cannabis; Haplotype; Organelle; Chloroplast; Mitochondrial; Drugs

1. Introduction

Cannabis sativa L. is a highly variable species that has been distributed worldwide by humans [1,2]. The plant is the source of the popular recreational drug 'marijuana' and is prohibited by law in many parts of the world. The species has also been cultivated as a source of fibre, seed and seed-oil for thousands of years, and there has been a recent resurgence in interest in the plant for these applications [3,4]. Licensing schemes are in place in some countries to permit cultivation of cannabis with low drug content for agronomic purposes [2].

There has been debate on the taxonomic limits and status of *Cannabis* species, sub-species and varieties [1]. The currently accepted opinion is that cannabis comprises a highly variable, highly hybridised and introgressed, panmictic (unstructured) species [1]. Where separate species and subspecies classifications have been used they have generally reflected crop-use characteristics. *Cannabis indica* Lam. from southern latitudes was distinguished broadly by its intoxicant properties when compared with more northerly cultivated low-intoxicant variant *C. sativa* L. [1,5]. Some wild or weedy European populations had been raised to species level as *Cannabis ruderalis* Jan. [1,5], but this taxon is not generally accepted [1]. Small and Cronquist devised a taxonomy that split *C. sativa* into two subspecies, *sativa* and *indica* based on intoxicant properties, and each with a domesticated and wild variety. The cultivated and wild varieties are var *sativa* and var *spontanea* in subspecies *sativa*, and var *indica* and var *kafiristanica* in subspecies *indica*, respectively [1]. This classification used a suite of characters exhibiting continuous yet bimodal variation associated with life in the cultivated or wild states.

Drug-plant breeders use an alternative taxonomy and recognise two taxa, the ''species'' *indica* and *sativa*. In the drug-plant breeding context the taxon *indica* generally refers to Asian drug plants, especially with features prominent in the Afghani strains, of wide leaflets, compact habit and early maturation [6]. These are strains that have been traditionally used in the production of resin for drug purposes (*hashish*) as opposed to

* Corresponding author at: Centre for Forensic Science, Canberra Institute of Technology, GPO Box 826, Canberra, ACT 2601, Australia.
Tel.: +61 2 62074157; fax: +61 2 62074105.
E-mail address: simon.gilmore@cit.act.edu.au (S. Gilmore).

0379-0738/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2006.10.025

leaf and inflorescence (flower-spike) material. The drug breeder's taxon *sativa* is used to refer to a collection of drug ''strains with narrow leaflets and slender habit from Colombia, Thailand, South Africa and Mexico'' and generally used to produce leaf or inflorescence marijuana [6].

The ability to distinguish between different varieties and populations of cannabis would allow law enforcement officials to make connections between seizures, growers, or members of a consortium of growers. The expanded licensing of hemp-fibre crops could be facilitated if drug varieties could be identified, for instance, reducing fears of agricultural crops disguising illegal drug crops. DNA analysis of nuclear short tandem repeats (STRs) indicates these markers can distinguish between individuals and populations of cannabis [7–10]. This follows earlier studies using Random Amplified Polymorphic DNA (RAPDs) which revealed informative genetic diversity [11,12].

Analysis of the organelle DNA of plants, particularly using chloroplast DNA (cpDNA) is well established as a tool to study evolutionary and population processes [13,14]. Chloroplasts and mitochondria are generally uniparentally (maternally) inherited in angiosperms [13,14], and hence have the potential to provide complementary evolutionary and biogeographic insights to nuclear DNA markers such as STRs. Numerous studies in plants have used cpDNA restriction site and sequence data to elucidate species and higher level phylogenetic relationships [13,15–18]. PCR amplification of cpDNA, followed by restriction fragment length polymorphism analysis of the amplicons (PCR-RFLP) has provided a rapid method of assaying among-population information for some plant groups [19–21]. PCR-RFLP of cpDNA has been facilitated by publication of large numbers of universal PCR primer sequences, producing amplicons that cover large tracts of the chloroplast genome [22–25].

These universal primer sets have also assisted the discovery and implementation of cpDNA simple sequence repeat markers (cpSSRs). CpSSRs are composed of mononucleotide repeats (homopolymer regions) rather than the di-, tri-, tetra- and pentanucleotide repeats usually targeted in nuclear STRs [26]. Although mutation rates appear to be relatively low compared to nuclear STRs, they may be higher than base substitutions in cpDNA [27]. CpSSRs have been applied to the investigation of among- and within-population genetic structure in both angiosperms [26,28,29] and gymnosperms [30,31]. Universal primer sets that target regions likely to contain cpSSRs, have been designed from the results of sequence database searches [29].

The mitochondrial genome (mtDNA) in plants has received far less attention than the chloroplast genome. MtDNA typically exhibits low rates of mutation [14,32,33], and is extremely variable in genome size and arrangement among different species [14,33]. Nonetheless, a number of universal primer sets have been developed for mtDNA in plants [24,25,34], and have been used to study the phylogeny [33] and population genetics of a variety of plant species [34,35]. Despite the low rate of mutation, studies have shown high levels of mtDNA variation in some species [34].

In this study, we: (a) report the results of DNA comparative sequencing of 12 cp- and mtDNA loci in a limited number of cannabis individuals, (b) describe the development of simple PCR tests to assay haplotype variation across a wider collection of cannabis, and (c) describe the haplotypic variation in 188 cannabis samples representing 76 different populations and discuss the relationship between our haplotype data and taxonomic opinions of cannabis, and the implications of organelle DNA haplotyping to forensic investigations of cannabis. The ability of organelle DNA haplotyping to recover useful forensic information on the biogeography and crop-use characteristics of different cannabis populations is demonstrated.

2. Materials and methods

2.1. Samples and DNA extraction

Our sample set (Table 1) comprised 188 cannabis individuals, representing 76 populations (42 drug-types, 25 fibre-types and 9 wild). The populations were sourced from either germplasm seed accessions from the Centre for Plant Breeding and Reproductive Research Centre (CPRO) in the Netherlands (see ref. [6]), or drug crop seizures by the Australian Federal Police or Victoria Police Service (Australia).

Some DNA was extracted from seedlings germinated from accessioned seed [7]; or obtained from DNA samples extracted for earlier RAPDs based studies of Jagadish et al. (see ref. [12] for extraction protocol). 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. Also, DNA was extracted from approximately 0.1 g of dried inflorescence from the Victoria Police seizure samples, using the same extraction protocol as for seedlings. These samples are referred by the abbreviated population codes v1–v16 (Table 1).

2.2. Detecting organelle DNA sequence variation

Universal primer sets from various sources were used to amplify approximately 8.5 kb of cpDNA and mtDNA from 12 loci, for 3–16 individuals per locus (Table 2). Two further locus (*trn*S-*trn*T and ccmp3) were also targeted, but we were unable to generate PCR products (Table 2). PCR products were used as templates for direct sequencing using the DYEnamic ET Terminator sequencing kit, and an ABI 377 automated sequencer. For some loci, internal sequencing primers were designed using Primer3 [36]. Sequencing reactions were conducted using 4.5 μl of PCR product (approximately 0.1 pmol total), 1.3 μl of 2 μM primer and 4 μl of DYEnamic ET Terminator (Amersham) sequencing premix. Sequence products were generated by cycle sequencing for 30 cycles of 95 °C for 20 s, 55 °C for 15 s and 60 °C for 60 s. Products were then run on an ABI 377 automated sequencer. Sequences were aligned using Sequencher Version 3.0 software (GeneCodes), and corrected by eye. Sequence polymorphisms were recorded and used to design primers to assay either fragment length differences or single nucleotide polymorphisms (SNPs). Polymorphic loci were given cannabis specific labels (Table 3).

2.3. Primer design for fragment length and SNP assays

Five of the seven polymorphic loci exhibited fragment length variation, identified by sequencing, of between one and four base pairs, and primers were designed to assay the length of PCR products spanning the length-variable regions (loci Cscp001-004, Csmt001; Table 3). Where we designed our own primers we used Primer3 [36]. The −21M13 sequence (5′-TGTAAACGACGGCCAGT) was included as a 5′-tail to one of the primer pair for fluorescent genotyping by the method of Schuelke [37].

We used the Amplification Refractory Mutation System (ARMS) method [38] to design primers for the assay of the SNP positions (loci Cscp005 and Csmt002) and the single base indel at the Cscp001 locus (Table 3). Forward primers were designed by eye, ending with the 3′ base at the SNP position (or 1 base indel position). A deliberate mismatch was included at the third last

Table 1
Samples used in this study

Population	Code	*N*	Taxon	Cultivar/other name	Country of origin	District	Haplotypes
Drug							
883270	D70	1	*indica* x *sativa*	Skunk 1	USA		1211222
891194	D94	1			Lebanon		1122121
891195	D95	1		Nederweit	Netherlands		1211222
891196	D96	1		Nederweit	Netherlands		1211222
891197	D97	3		Nederweit	Netherlands		1122121
891198	D98	1	*indica*	Nederweit	Netherlands		1122121
891199	D99	1	*indica* x *sativa*	Skunk	Netherlands		1211222
891200	D00	1	*indica* x *sativa*	Four way	Netherlands		1211222
891383	D83	1			Afghanistan		1211222
891384	D84	1			Swaziland		2213233
891385	D85	1			South Africa		1122121
910972	D72	1	*sativa*		Netherlands		2213233
921192	D92	3			Nepal	Kalopani	1211222
921199	D199	1			Afghanistan	Mazar-I-Sharif	1211222
921200	D200	1			Afghanistan		1211222
921205	D05	1			USA	Hawaii	1211222
921209	D09	1			Jamaica		2213233
921230	D30	1			Mexico		1211222
921231	D31	1			Mexico		2213233
921232	D32	8			Mexico		1211222
921234	D34	1			Zimbabwe		2213233
921235	D35	1			South Africa		2213233
921236	D36	8			Sierra Leone		2113232
921237	D37	1			Thailand		2213233
991239	D39	4			Uganda	Mbale	2113232
vic14dv	DVIC	1			Australia	Victoria	1122121
V01	v1	4			Australia	Victoria	1211222
V02	v2	4			Australia	Victoria	1211222
V03	v3	2			Australia	Victoria	1211222
V04	v4	1			Australia	Victoria	1211222
V05	v5	4			Australia	Victoria	1122121
V06	v6	5			Australia	Victoria	1211222
V07	v7	3			Australia	Victoria	1211222
V08	v8	3			Australia	Victoria	1211222
V09	v9	3			Australia	Victoria	1211222
V10	v10	2			Australia	Victoria	1211222
V11	v11	2			Australia	Victoria	1211222
V12	v12	2			Australia	Victoria	1211222
V13	v13	2			Australia	Victoria	1211222
V14	v14	3			Australia	Victoria	1211222
V15	v15	6			Australia	Victoria	1211222
V16	v16	4			Australia	Victoria	1211222
Fibre							
880816	F16	1		Rastislavicke	Former Czechoslavakia		1122121
883038	F38	1		Eletta Campana	Italy		1122121
883039	F39	1		Krasnodars-kaya	Former USSR		1122121
883041	F41	13		Fibrimon-56	France		1122121 (9) 1211222 (4)
883042	F42	11		Fibrimon	Former East Germany		1122121
883043	F43	11		Fibrimon	Hungary		1122121 (1) 1211222 (10)
883044	F44	1		Szegedi-9	Hungary		1122121
883066	F66	1		Futura 77	France		1122121
883213	F13	1		Kozuhara zairai	Japan	Shiga	1122121
883247	F47	1			Former USSR	Orlov	1122121
883248	F48	1			Former USSR	Altaij	1122121
883289	F89	1			Former USSR	Kirov	1122121
883291	F91	8		SOU	Former USSR		1122121 (7) 1211222 (1)
883292	F92	1			Former USSR		1122121
883294	F94	1		Juznaja Odnovremenno	Former USSR		1122121
891090	F90	1			Turkey		1211222

Table 1 (*Continued*)

Population	Code	*N*	Taxon	Cultivar/other name	Country of origin	District	Haplotypes
891092	F092	1			Turkey		1211222
891094	F094	1			Turkey		1211222
901078	F78	1			China		1122121
901162	F62	1		daema	Korea	Bonghwa	1222121
901163	F63	1		daema	Korea	Milyang	1222121
921018	F18	1			Turkey		1211222
921122	F22	8			China		1211222 (6)
							1211212 (2)
921207	F07	12			India	NW Himalayas	2213233 (10)
							2113232 (2)
921214	F14	1			Former USSR	Far east	1122121
Wild							
883110	W10	1	var *spontanea*		Hungary		1122121
883141	W41	1	var *spontanea*		Former East Germany		1122121
883154	W54	1	ssp. *ruderalis*		Romania		1122121
891191	W91	1			Nepal	Kalopani	1211222
891193	W93	1			Nepal	Dana	2213233
921201	W01	1			China		1211222
921203	W03	1			Canada		1122121
921206	W06	1	*ruderalis*		Hungary		1122121
921240	W40	1			USA	Minnesota	1122121

Population numbers are CPRO accession numbers (see ref. [6]) except in the case of samples from Victoria, Australia. Samples are separated into drug, fibre or wild populations. *N* = number of individuals sampled from the population. Taxon and Cultivar/other name information was supplied by the contributors to the CPRO germplasm collection. Haplotypes that were observed in each population are included here for convenience (see Table 3).

3′ position of the ARMS primer, to increase the specificity the PCR reactions [38]. Each sample was subjected to two PCR assays, run in parallel, and each containing one of the alternate ARMS primers. We retained the closest universal primer [24,25] as the reverse primer for the cpDNA loci, but designed a reverse primer for targeting the mtDNA Csmt001 locus using Primer3 [36].

2.4. Assaying fragment length variation

The primers used at each length-variable locus are shown in Table 3. PCR conditions were 10–100 ng template DNA, 1× TaqGold PCR buffer (Perkin-Elmer), 0.2 μM dNTPs, 200 nM each of untailed locus-specific primer and labeled −21M13 reporter primer (either labeled with FAM, NED or VIC fluorescent dyes; Invitrogen), 50 nM primer with −21M13 tail, and 0.75 U Taq DNA Polymerase (QIAGEN). The reaction profile comprised a pre-denaturation for 10 min at 94 °C, then 25 cycles of 94 °C for 30 s, 56 °C for 30 s and 72 °C for 30 s for amplification with the locus-specific primers, followed by 8 cycles at 94 °C for 30 s, 54 °C for 30 s and 72 °C for 30 s for amplification with the labeled reporter primer. All amplifications were conducted in a Palm-Cycler CP002 (Corbett Research), and PCR products were run on an ABI PRISM 377 DNA sequencer. Fragment sizes were scored using GENESCAN and GENOTYPER software (Applied Biosystems).

2.5. Assaying SNPs markers

For each DNA sample, a locus was targeted with two 20 μl PCR amplifications, one for each of the alternate forward primers (Table 3), with reaction mixes comprising 10–50 ng DNA, 1× TaqGold PCR Buffer with $MgCl_2$, (Perkin-Elmer), 250 μM dNTPs, 0.2 μM each of forward and reverse primer, and 0.75 units of TaqGold DNA Polymerase (Perkin-Elmer). The reaction profiles comprised a 10-min incubation at 94 °C to activate the TaqGold then a cycle of 95 °C for 15 s, 48 °C for 15 s and 72 °C for 30 s, repeated 25 times. Following cycling, the reaction was held at 72 °C for 5 min, before a final 4 °C holding temperature. The reaction products were electrophoresed in a 1% agarose gel containing 0.5 μg/ml ethidium bromide. The gel was poured with 2 rows of wells and products of alternate ARMS primers were run in parallel. Gels were photographed over UV-light. Haplotypes were scored directly from the photographs.

2.6. DNA sequencing to confirm results

DNA sequencing was retained as a tool to check ambiguous results and to characterise new alleles as they were discovered. For the cpSSR regions, we amplified the desired locus and sequenced in one direction only. To check SNP positions, we used either universal primers, or internal primers we had designed for sequencing, to amplify the sequencing template, and again sequenced in one direction only. Sequencing reactions were conducted using 4.5 μl of PCR product (approximately 0.1 pmol total), 1.3 μl of 2 μM primer and 4 μl of DYEnamic ET Terminator sequencing premix (Amersham). Sequence products were generated by cycle sequencing for 30 cycles of: 95 °C for 20 s, 55 °C for 15 s and 60 °C for 60 s. Products were then run on an ABI377 automated sequencer. Sequence electropherograms were imported into Sequencher Version 3 (GeneCodes) for editing.

Size assays at the Cscp004 locus, for which allele size calling is complicated by stutter patterns were the main focus for use of sequencing to confirm alleles (data not shown). The DNA sequence data were unambiguous, and where obtained it was preferred for constructing haplotypes.

2.7. Data analysis

Haplotypes were scored at each position, according to the scheme in Table 3, to produce a 7-digit haplotype (Table 4). The full dataset was reduced for phylogenetic analysis, such that each population was recorded once for each haplotype found, regardless of number of individuals sampled. A data matrix was produced for the software package PAUP*4.0b10 [39] and analysed using parsimony, UPGMA and neighbour joining methods. The haplotype loci formed seven unordered characters for parsimony analysis. The 5 populations that possessed 2 haplotypes (F41, F43, F91, F22 and F07), are each represented twice, hence there are 81 terminals. Characters were not weighted. A heuristic search was conducted using the simple sequence addition option, with tree-bisection-reconnection (TBR) branch swapping. Support for the groupings was tested by bootstrap analysis with 1000 bootstrap replicates. Further support for groups recovered in the parsimony analysis was sought by conducting UPGMA and neighbour joining clustering analyses in PAUP*4b10, with total character differences as the measure of distance. Trees produced by each of these methods were then compared. The geographical spread of haplotypes was also illustrated by plotting populations onto a world map.

Table 2
Loci targeted for comparative sequencing of cp- and mtDNA loci

Locus	Primer name and sequence	Reference	Length of sequence obtained	Polymorphism	*N*	GenBank accessions
cpDNA						
*trn*L-*trn*F	c: 5′-CGAAATCGGTAGACGCTACG f: 5′-ATTTGAACTGGTGACACGAG	[23]	750 bp	Single base indel	10	AY958387–AY958396
*trn*H-*trn*K	*trn*H: 5′-ACGGGAATTGAACCCGCGCA *trn*K: 5′-CCGACTAGTTCCGGGTTCGA	[24]	250 bp 3′ from *trn*H 635 bp 3′ from *trn*K	SNP close to *trn*H primer	9 5	DQ004522–DQ004530 DQ004501–DQ004505
*trn*S-*trn*T	*trn*S: 5′-CGAGGGTTCGAATCCCTCTC *trn*T: 5′-AGAGCATCGCATTTGTAATG	[24]	No product	Unknown	0	–
*rbc*L-*orf*106	*rbc*Lf: 5′-ATGTCACCACAAACAGAAACTAAGCAAGT *orf*106r: 5′-ACTACAGATCTCATACTACCCC	[22]	Approximately 3 kb	Variable length homopolymer	#	Consensus of several clones, not submitted
ccmp2 (5′ to *trn*S)	ccmp2F: 5′-GATCCCGGACGTAATCCTG ccmp2R: 5′-ATCGTACCGAGGGTTCGAAT	[29]	209 bp	Variable length homopolymer	8	DQ389131–DQ39136
ccmp3 (*trn*G intron)	ccmp3F: 5′-CAGACCAAAAGCTACATAG ccmp3R: 5′-GTTTCATTCGGCTCCTTTAT	[29]	No product	Unknown	0	–
ccmp4 (*atp*F intron)	ccmp4F: 5′-AATGCTGAATCGAYGACCTA ccmp4R: 5′-CCAAAATATTBGGAGGACTCT	[29]	121 bp	None detected	3	DQ389125–DQ389127
ccmp6 (*orf*77-*orf*82)	ccmp6F: 5′-CGATGCATATGTAGAAAGCC ccmp6R: 5′-CATTACGTGCGACTATCTCC	[29]	106 bp	Variable length homopolymer	3	DQ389128–DQ389130
ccmp7 (*atp*B-*rbc*L)	ccmp7F: 5′-CAACATATACCACTGTCAAG ccmp7R: 5′-ACATCATTATTGTATACTCTTTC	[29]	120 bp	None detected	6	DQ389119–DQ389124
mtDNA						
cox 2 exon 1 to exon 2	cox2/1: 5′-TTTTCTTCCTCATTCTKATTT cox2/2r: 5′-CCACTCTATTGTCCACTTCTA	[25]	325 bp	None detected	10	DQ004544–DQ004553
nad 1 exon 4 to exon 5	nad1/4: 5′-GCCAATATGATCTTAATGAG nad1/5r: 5′-TCACCTTGATACTAAACCAG	[25]	1525 bp	None detected	#	Consensus of several clones, not submitted
nad 4 exon 3 to exon 4	nad4/3: 5′-GGAGCTTTCCAAAGAAATAG nad4/4r: 5′-GCCATGTTGCACTAAGTTAC	[25]	345 bp	2 closely linked variable length homopolymers	16	DQ004506–DQ004521
nad 5 exon 4 to exon 5	nad5/4: 5′-CCAATTTTTGGGCCAATTCC nad5/5r: 5′-CATTGCAAAGGCATAATGAT	[25]	577 bp	SNP	9	DQ004522–DQ004530
nad 7 exon 1 to exon 2	nad7/1: 5′-ACCTCAACATCCTGCTGCTC nad7/2r: 5′-CGATCAGAATAAGGTAAAGC	[25]	546 bp	None detected	6	DQ004531–DQ004534

Length of sequence obtained is the maximum length of aligned sequences from cannabis. Polymorphism is the variation observed across the *N* individuals sequenced at each locus. Primer names were retained from the references cited.

Table 3
Cannabis organelle loci targeted in this study, and primers used for haplotyping

Locus	Primers	Reference	Region	Alleles	Score
Length assays					
Cscp001[a]	cscp001F: 5′-TCCTCTCATTCCGTTAGTGGT cscp001R: 5′-AATTGCACATTGGGATTCCT[b]	This study	*trn*L-*trn*F	212 211	1 2
Cscp002	cscp002F: 5′-TCATTTGATGAAGTGGGGTA cscp002R: 5′-GCATGGGGAACCTACTATTT[b]	This study	*rbc*L-*orf*106	160 159	1 2
Cscp003	ccmp2F: 5′-GATCCCGGACGTAATCCTG ccmp2R: 5′-ATCGTACCGAGGGTTCGAAT[b]	[29]	ccmp2	235 234	1 2
Cscp004	ccmp6F: 5′-CGATGCATATGTAGAAAGCC ccmp6R: 5′-CATTACGTGCGACTATCTCC[b]	[29]	ccmp6	128 127 126	1 2 3
Csmt001	csmt001F: 5′-ATGGCAGAGAAGTTTCCATA csmt001R: 5′-TTGGCTCCCTAAAGACTAAA[b]	This study	*nad* 4 exon 3 to exon 4	236 235 232	1 2 3
ARMS assays					
Cscp001[a]	SNPTRNLC: 5′-ACAACCGGACCTGAATGATCC (forward) SNPTRNLT: 5′-ACAACCGGACCTGAATGATCT (forward) f: 5′-ATTTGAACTGGTGACACGAG (reverse)	This study for forward primers, [23] for reverse primer	*trn*L-*trn*F	c t	1 2
Cscp005	SNPTRNHC: 5′-TTATCTTGTCTAAAATTGAAGTC (forward) SNPTRNHT: 5′-TTATCTTGTCTAAAATTGAAGTA (forward) trnH: 5′-ACGGGAATTGAACCCGCGCA (reverse)	This study for forward primers, [24] for reverse primer	*trn*H-*trn*K	c a	1 2
Csmt002	SNPN5C: 5′-TATGACCTGTGGCCGCCAGC (forward) SNPN5T: 5′-TATGACCTGTGGCCGCCAGT (forward) N5R: 5′-TCTTAACGCCCCTACTACTG (reverse)	This study	*nad* 5 exon 4 to exon 5	c t	1 2

Region refers to the locus targeted by universal primers used for comparative sequencing to find these polymorphic sites. Alleles for length variable loci are designated by length of PCR product (including −21M13 tail, see Section 2), or for SNP positions, by nucleotide at 3′ end of ARMS primer.
[a] The locus Cscp001 was assayed both by length estimation and the ARMS method.
[b] Primers constructed with −21M13 tail (5′-TGTAAACGACGGCCAGT) for genotyping, see Section 2.

3. Results

3.1. Sequence characterization

Partial sequences for 12 loci (7 cpDNA, 5 mtDNA) were obtained (Table 2). Although nearly 8.5 kb (approximately 5.2 kb cpDNA, 3.3 kb mtDNA) of predominantly spacer DNA was sequenced, only 7 polymorphic sites were recovered (Table 2). Approximately 1 polymorphism was recovered per 1 kb sequenced in cpDNA and 1 polymorphism per 1.7 kb in mtDNA.

Most length polymorphisms were due to different length homopolymer A/T tracts in cpSSR regions (Cscp002, Cscp003 and Cscp004). The mitochondrial SSR (mtSSR) at the Csmt001 locus exhibited two variable length homopolymer regions; a polyA tract 96 bases 5′ to a polyC tract. In the 2 alleles recovered through sequencing, the polyA and polyC tracts were 1 base and 2 bases longer in the larger allele, respectively. The chloroplast spacer *trn*L-*trn*F is polymorphic due to a 1 base indel (locus Cscp001). This indel has been reported previously in cannabis [40]. Single nucleotide polymorphisms were detected in the chloroplast spacer region between the *trn*H and *trn*K genes (locus Cscp005), and in the mitochondrial intron between *nad*5 exon 4 and exon 5 (locus Csmt002). The mitochondrial introns *nad*7 (exon 1 to exon 2) and *cox*2 (exon 1 to exon 2), were also targeted but no sequence variation was detected. Primers targeting the

Table 4
Haplotypes recovered in this study

Haplotype	Cscp001	Cscp002	Cscp003	Cscp004	Cscp005	Csmt001	Csmt002	N_I	N_{POP}
I	1	1	2	2	1	2	1	58	29
II	1	2	2	2	1	2	1	2	2
III	1	2	1	1	2	2	2	94	37
IV	1	2	1	1	2	1	2	2	1
V	2	2	1	3	2	3	3	18	9
VI	2	1	1	3	2	3	2	14	3

Allele designations are as in Table 3. N_I, number of individuals possessing this haplotype; N_{POP}, number of populations containing individuals with this haplotype. Some populations contained more than one haplotype.

chloroplast spacer *trn*S-*trn*T failed to generate a PCR product, and this locus was not pursued further.

3.2. Haplotype data analysis

A total of 188 individuals, representing 76 different populations (seizures or germplasm accessions) were successfully typed and 6 organelle DNA haplotypes were recovered (Tables 1 and 4). Parsimony analysis recovered three shortest trees of equal branch length, with strong bootstrap support for branches leading to each of the major haplotypes (Fig. 1). The distance-based UPGMA (data not shown) and neighbour joining analyses (Fig. 2) also recovered each of the groups observed in the parsimony analysis. Drug crop populations, wild populations and fibre crop populations are found in each of the main haplotypes, but the populations are differentially distributed according to geographical origin (Fig. 3) and crop-use characteristics (Table 1).

Fibre-type crops dominate the group containing haplotypes I and II. Of the total 25 fibre populations sampled, 17 were in this group. Only 6 of the total 42 drug plant populations sampled fell into this group (2 of these are from Australia, 2 from the Netherlands and 1 each from Lebanon and South Africa). All wild populations from Europe sampled, including germplasm material labeled as species or subspecies *ruderalis* and var *spontanea*, have haplotype I. Wild material from the USA and Canada also exhibits this haplotype (W40 and W03, respectively). Haplotype II was only found in 2 fibre-type samples (F62 and F63), from different districts of Korea, which exhibit a different allele at the Cscp002 locus. This allele was confirmed by sequencing (data not shown). Bootstrap support (1000 replicates) for the split between these groups is relatively weak at 62%. However, bootstrap support for the group that includes both haplotypes I and II is high at 93%.

Haplotype III forms a group in the consensus tree at the node between haplotype II and haplotype V; and comprises largely

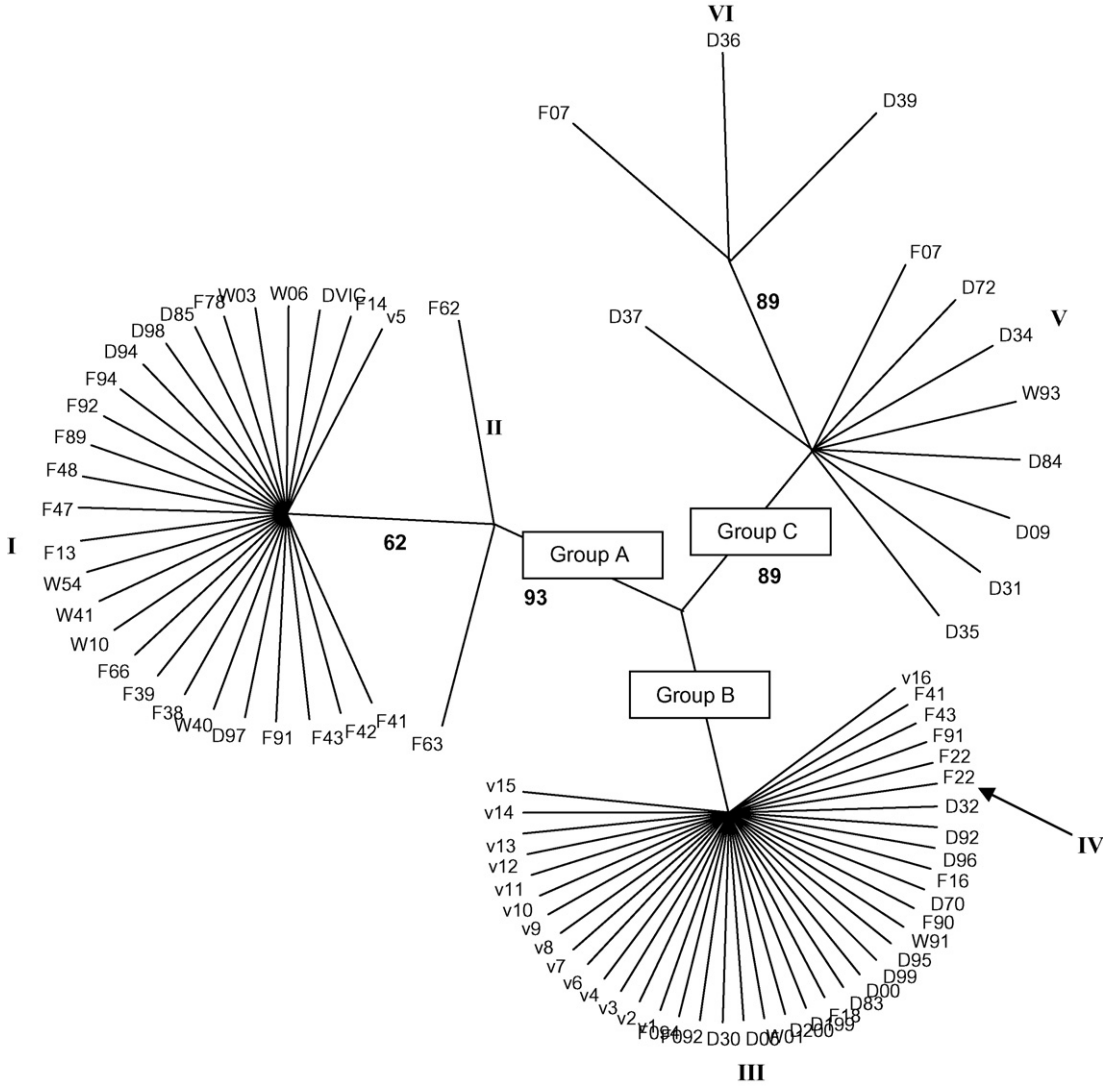

Fig. 1. One of three equally most parsimonious trees (length = 10) recovered by PAUP*4.0b10 [38] for the haplotype data. Numbers on the branches are bootstrap values (1000 replicates), values below 50% are not shown. In the strict consensus tree, the branch leading to Group B collapses. Haplotypes are shown by Roman numerals, while haplotype groups are labeled.

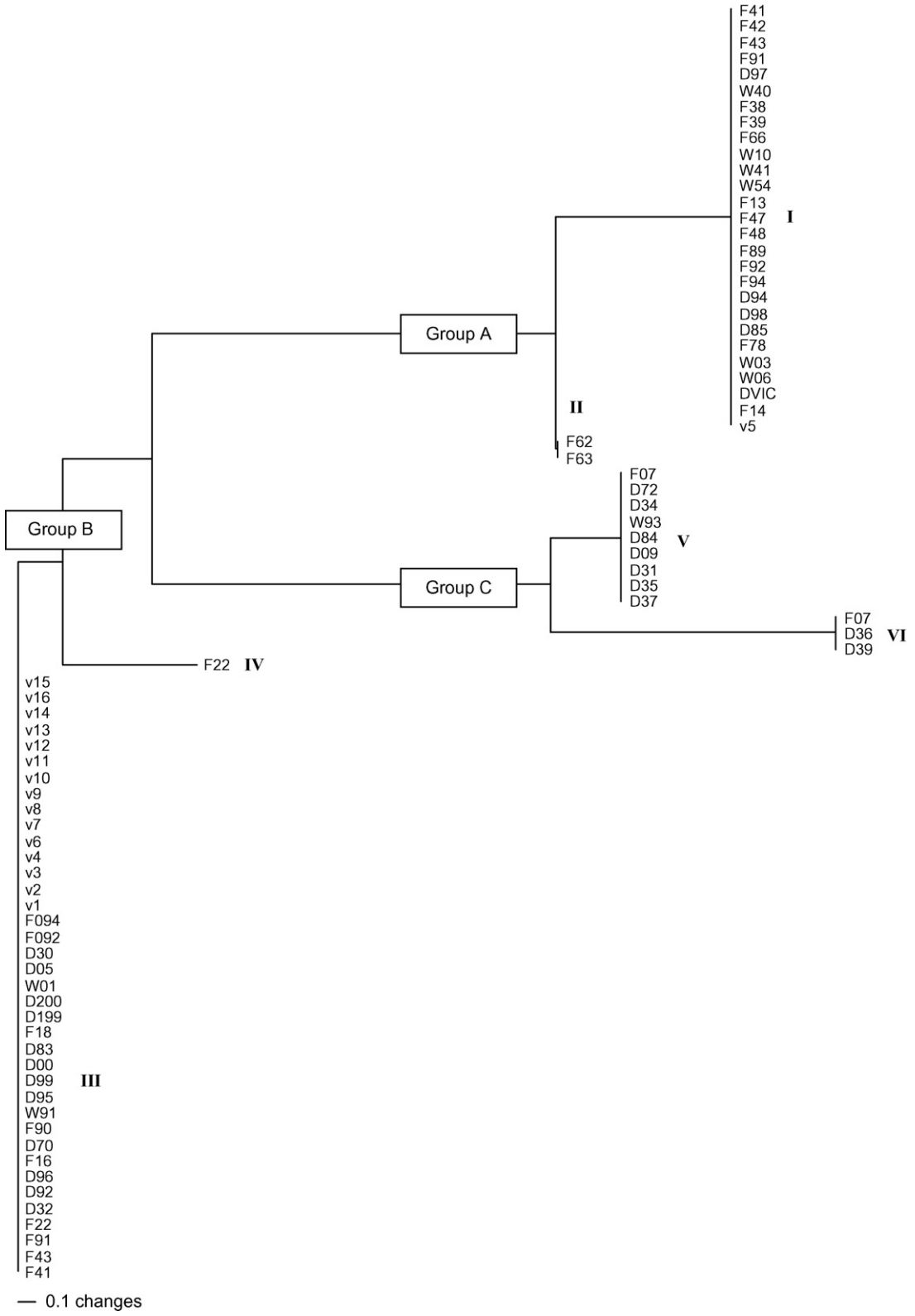

Fig. 2. Neighbour joining tree constructed in PAUP*4.0b10 from the haplotype data set. Total character differences were used as the distance measurement. The groups of haplotypes observed are identical to those recovered by parsimony analysis.

drug plants, including strains of cannabis developed for hydroponic cultivation. About two-thirds of the drug-plant populations sampled are in this group (27/42), including 3 accessions from Afghanistan, 3 ''Skunk'' cultivars (from USA and Netherlands), and 15 of 16 seizures of indoor cannabis from Australia. Other drug crops with this haplotype originated from Nepal or Mexico. Nine fibre plant populations contribute to this group, including 3 populations that also possess haplotype I

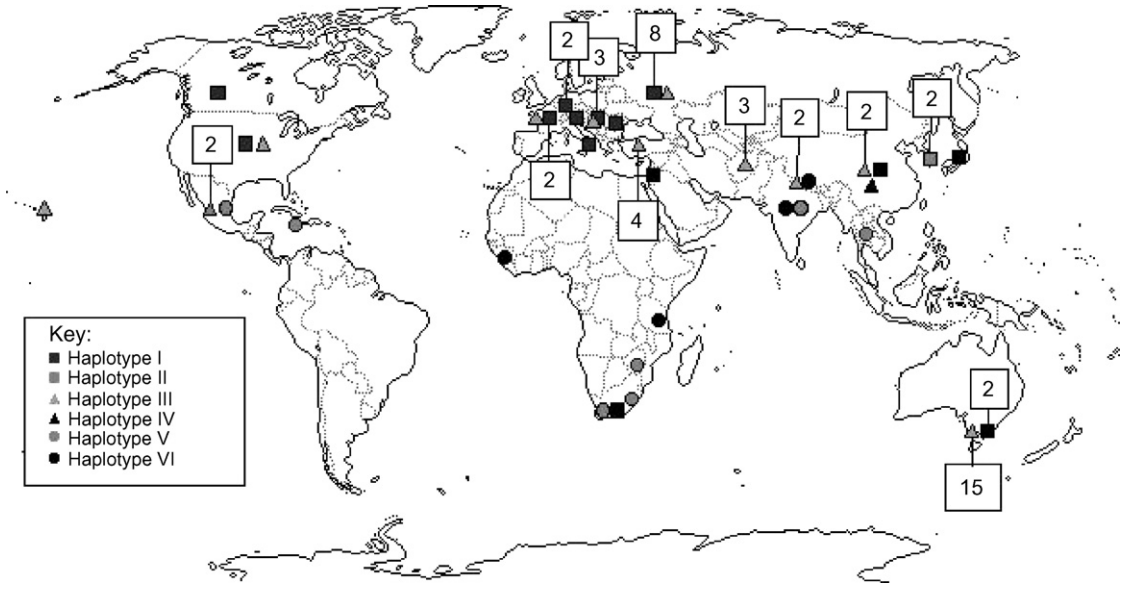

Fig. 3. World map showing the distribution of haplotypes in this study. Markers (see key) are placed within political boundaries of the country of origin without regard to local regions, except in the case of the samples from Hawaii (USA) and samples from Victoria (Australia). Populations are recorded once for each haplotype they displayed. Flags attached to markers indicate number of populations displaying that haplotype. Populations from the Netherlands are not included.

(F41, F43 and F91). Also included are all 4 landraces from Turkey, and fibre types from both Europe and Asia. Two populations of wild plants from Nepal and China have this haplotype. A variant of haplotype III is seen in two individuals from the population F22, who possess the haplotype IV, which has a different allele at the Csmt001 locus. DNA sequencing has confirmed this result (data not shown). The parsimony analysis shows this change to be autapormorphous (i.e. a change along a terminal branch), and hence haplotype IV does not form a separate group to haplotype III. The branch between haplotypes II and III has 93% bootstrap support, whereas the branch splitting haplotypes III and V has 89% bootstrap support, thus there is strong support for haplotype III as a separate group to the other major haplotypes.

Haplotypes V and VI form a clade with 89% bootstrap support, which includes 9 drug plant populations, 1 wild population, and 1 fibre plant population. Five of the 6 drug types populations from Africa are in this group, as are drug-types from Mexico, Jamaica, Thailand and the Netherlands. Nepal is also represented with a single wild population. The only fibre crop in this group is from India (F07), and contains individuals with both haplotypes V and VI. Haplotype VI was found in 3 populations, 2 of the African drug populations and in the Indian fibre crop F07. There is also strong bootstrap support for the branch splitting the V and VI haplotypes (89%).

4. Discussion

4.1. Organelle sequence diversity in cannabis

Sequences were obtained for 12 cannabis organelle DNA loci and revealed few polymorphisms, at the approximate rate of 1 polymorphic site for every 1.2 kb sequenced. Five polymorphisms were found in cpDNA (1 polymorphism per 1 kb sequenced), and 2 polymorphisms were found in mtDNA (1 polymorphism per 1.7 kb sequenced). The results are consistent with studies that show differing rates of mutation in the 2 organelles, with mtDNA evolving most slowly [32]. The most common form of variation discovered was length variation at homopolymer regions (4 sites). SNPs were next most common (2 sites), then a one base indel (1 site). A mutation rate for cpSSRs has been calculated, and is considerably higher than substitution rates in the same genome [27].

Two of the loci exhibited single nucleotide polymorphisms (SNPs). SNPs markers have recently gained widespread attention, particularly as the Human Genome Project has shown that most human sequence variation is in the form of SNPs [41]. Similarly, the genome sequences of the angiosperm *Arabidopsis thaliana* reveal SNPs to be the most common form of sequence variation [42]. In plants, SNPs have found most use in gene mapping, especially through the use of highly inbred strains where (nuclear) haplotypes can be directly recovered by DNA sequencing of targeted regions, and converted into SNPs assays [43]. We used a relatively simple SNPs assay here, but if there are further SNP discoveries in cannabis it may be more convenient to convert to a fluorescent dye based system for an automated genotyper. There are numerous published methods for assaying SNPs (reviewed in ref. [44]), with intensive research into developing methods of multiplexing large numbers of loci to overcome the inherent low variability of SNPs.

The low rate of mutation over numerous loci suggests that no single organelle sequence can provide a high resolution for the genetic analyses of cannabis populations. Consequently, if discriminatory population level information is to be recovered from organelle DNA of cannabis, additional variable sites will be required, and multiple loci will need to be assayed.

4.2. Organelle haplotypes and their relationship to taxonomic opinions of cannabis

Across 188 individuals we recovered 6 different multilocus haplotypes, and these are highly suggestive of both crop-use and biogeographic groupings. The six haplotypes fall into three main ''haplotype-groups'', using either parsimony methods or cluster analysis methods. Without further morphological analysis of the accessions in this study it is impossible to conclusively assign the haplotype groups to formal taxonomic entities. Some caution is also warranted as organelle DNA variation may not necessarily reflect overall genome variation given the uniparental mode of inheritance of cp- and mtDNA. Pollen exchange, and hence genomic DNA exchange, between distant populations may readily occur given the reproductive biology of cannabis [1]. However, natural seed dispersal (necessary for organelle DNA dispersal) is likely to be at a much lower level. Further confounding the definite assignment of organelle type to formal taxonomic groups is the distribution and breeding of cannabis by humans. Notwithstanding these limitations, the haplotype groups recovered here appear to reflect some taxonomic opinions of *C. sativa*, as discussed below.

Populations from Europe dominate one haplotype group (haplotypes I and II), which we have called Group A. Most of these are fibre cultivars, but the group also includes all wild populations from Europe. Few drug-types are included. This haplotype group appears to encompass both the region, and the fibre-type plants, from which Linnaeus first described *C. sativa* in 1753. Group A has representatives from 17 of the 25 fibre crops represented, and all 4 of the wild populations sampled from Europe, as well as both wild populations from Canada and USA. Two of the wild European populations have germplasm notes describing them as *C. ruderalis* or ssp. *ruderalis*. Note that all taxonomic information provided by the CPRO germplasm is given by the original provider of the seed to the collection. *C. ruderalis* Jan., was described in 1924 by Janischewski from the Volga River system of Siberia and Central Asia [1,5]. Haplotyping places *C. ruderalis* in our Group A, and does not support recognition of *C. ruderalis* as a distinct species, at least not if these populations are exemplars. The haplotypes support the notion that European wild populations are descendants of escaped fibre crops [1]. Introgression of the organelle DNA of surrounding fibre crops into wild populations cannot be discounted however.

Similarly, the wild populations from Canada and the USA are most likely escaped fibre types. Group A was represented in Asia by four fibre accessions, including populations from Japan (1), China (1) and Korea (2). The two samples from different regions of Korea (F62, F63) share a difference to the main Group A haplotype, indicating that at least some variation exists within this group. Only 6 from a total of 42 different drug crop populations were included in Group A, including 2 from the Netherlands, 1 from South Africa, 1 from Lebanon, and 2 from Australia. In their treatment, Small and Cronquist [1] refer to *C. sativa* subsp. *sativa* as being typically from latitudes north of 30°N, which is consistent with the distribution observed here for Group A (Fig. 3).

The populations possessing haplotypes III and IV seem to fit broad conceptions of *C. indica*, especially as applied in the context of breeding drug-use plants [1,6], and we refer to this collection as Group B. The group generally occupies more southerly latitudes in Europe and Asia than Group A (Fig. 3). About two-thirds of the drug populations exhibit haplotype III. Among the 27 drug crops represented in this group are all three Afghanistan drug accessions, both USA drug accessions, and 15 of 16 seizures of indoor cannabis in Victoria, Australia. Three Skunk 'varieties' sampled are included. Skunk is described by drug-breeders as a hybrid between indigenous strains of Afghanistan *indica*, and Mexico and Colombia *sativa*, and which was developed in the US and commercialised in the Netherlands [6]. Thus, Group B appears consistent with the taxon called species *indica* by drug plant breeders, although this group also includes some fibre-type plants.

We have called the third main grouping Group C (haplotypes V and VI). Predominantly drug-type plants, this group are represented by a single fibre accession, one wild population and nine drug plant populations. Five of the six drug-type populations from Africa fall into this group (the 6th is in Group A). Other drug-type populations originate from Mexico, Jamaica, and the Netherlands, the latter being a germplasm accession labeled as species *sativa* (D72). Only one fibre crop accession (F07) sampled possessed this haplotype, this was a landrace from India. It would seem plausible that the Group C haplotypes entered Africa from India, possibly in ancient times, although multiple introductions are also likely, as the haplotypes V and VI are observed in both Africa and Southern Asia. Thereafter, the haplotypes may have spread to the Americas.

Group C appears consistent with the drug-breeders taxon *sativa* [6]. Drug breeders give the sources of this ''species'' as Colombia, Thailand, South Africa and Mexico (see Fig. 3). The slender habit of plants in the drug-breeders taxon *sativa* [6] is reminiscent of the fibre-type plants, and it appears drug breeders may have taken this as evidence of relatedness to fibre-type plants. On the other hand, the populations in this group would probably be treated by Small and Cronquist [1] as typical of *C. sativa* subsp. *indica*, by virtue of the crop-use characteristics and distribution. Our results may, therefore, indicate a distinct group of drug type plants, primarily from the more equatorial latitudes such as found in Africa, India, SE Asia and Mexico. These populations appear geographically consistent with the drug-breeders taxon *sativa*, and form a subset of *C. sativa* subsp. *indica* as envisioned by Small and Cronquist [1].

Although sampling in most regions was limited, it is interesting that the Netherlands was the only location with drug populations from all three haplotype groups. The Netherlands has been involved in domestic drug-cannabis production since the 1980s, and growers have commercialised a large variety of purebred and hybrid strains, imported from around the world [6].

Our understanding of the breeding history of some of the accessions is also consistent with the groupings based on haplotype. For example, although not all the European fibre crops possess haplotype I, it is clearly dominant in European fibre crops and ''wild'' accessions. Of the European accessions,

the cultivars Fibrimon, Fibrimon-56 and Futura 77 are closely related monoecious cultivars that were developed by cross breeding between Russian descended strains, German strains, and landraces from Turkey and Italy [6]. Haplotype III is observed in all Turkish landraces in this study, and thus the presence of this haplotype in some Fibrimon-related cultivars (F41, F43) may reflect these recent crossbred origins.

4.3. Implications for forensic investigations

Haplotype SNP and STR information shows the potential for recovering useful biogeographic and population level information about *C. sativa*. The low rate of sequence divergence in chloroplast and mitochondrial DNA was at first disappointing, but by replacing DNA sequencing with agarose-gel based ARMS style primer assays for SNPs, and fragment size analysis for cp- and mtSSRs, we were able to effectively target useful haplotype polymorphisms at low cost. Marker development was relatively simple, comparative sequencing of a few individuals was used to identify polymorphic sites and simple PCR assays were designed to target those sites. Only a small portion of the organelle DNA has been sequenced thus far, and additional informative markers are likely to be present.

Currently, these markers can discriminate between some cannabis populations, but are probably most informative in describing higher level groupings of cannabis. This study shows that drug-type cannabis occurs in each haplotype group, and therefore haplotypes can be used as a point of comparison for two or more drug crops. In Australia, two haplotype groups have been recovered in drug plants, Group A and Group B. All plants of Group B were observed from indoor-grown samples, while Group A haplotypes were identified from one indoor and one outdoor grown sample. If, as we suspect, varieties used for outdoor growing in Australia have different histories to the indoor varieties, then a simple DNA test may distinguish indoor and outdoor cannabis, at least from some regions of Australia. The single indoor crop seizure from Victoria that possessed the Group A haplotype was from an unusual growing operation, which used a distinct growing method, seemingly restricted to a particular ethnic group (Fiddian, personal communication). Further haplotyping of cannabis from similar crops may establish that a different haplotype of cannabis is used by these groups, to that normally observed at indoor crops.

Haplotyping may also provide a means of selecting STR loci for finer resolution of cannabis populations. As more STR loci are developed it may be that different sets of these markers are more informative in certain subsets of populations. Thus, haplotyping may provide a useful ‘‘first-pass’’ to assist in the selection of suitable STR loci for the comparison of cannabis populations under question.

Acknowledgements

We are grateful to Jodie Ward for laboratory assistance. Funding was provided by the Australian Research Council (ARC), through a Strategic Partnership with Industry Research and Training (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. The manuscript was significantly improved by the helpful comments of anonymous reviewers.

References

[1] E. Small, A. Cronquist, A practical and natural taxonomy for *Cannabis*, Taxon 25 (1976) 405–435.
[2] G. Siniscalco Gigliano, *Cannabis sativa* L.—botanical problems and molecular approaches in forensic investigations, Forensic Sci. Rev. 13 (1) (2001) 1–17.
[3] 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.
[4] 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.
[5] M. Booth, Cannabis: A History, Doubleday, Sydney, 2003.
[6] E.P.M. de Meijer, L.J.M. van Soest, The CPRO *Cannabis* germplasm collection, Euphytica 62 (1992) 201–211.
[7] S. Gilmore, R. Peakall, J. Robertson, Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations, Forensic Sci. Int. 131 (2003) 65–74.
[8] S. Gilmore, R. Peakall, Isolation of microsatellite markers in *Cannabis sativa* L. (marijuana), MEN 3 (2003) 103–105.
[9] H.J. Alghanim, J.R. Almirall, Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analysis, Anal. Bioanal. Chem. 376 (2003) 1225–1233.
[10] H.-M. Hsieh, R.-J. Hou, L.-J. Tasi, C.-S. Wei, S.-W. Liy, L.-H. Huang, Y.-C. Kuo, A. Linacre, J. Lee, A highly polymorphic STR locus in *Cannabis sativa*, Forensic Sci. Int. 131 (2003) 53–58.
[11] 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.
[12] V. Jagadish, J. Robertson, A. Gibbs, RAPD analysis distinguishes *Cannabis sativa* samples from different sources, Forensic Sci. Int. (1996) 113–121.
[13] R.G. Olmstead, J.D. Palmer, Chloroplast DNA systematics: a review of methods and data analysis, Am. J. Bot. 81 (1994) 1205–1224.
[14] J.D. Palmer, Contrasting modes and tempos of genome evolution in land plant organelles, Trends Genet. 6 (1990) 115–120.
[15] S. Gilmore, K.D. Hill, Relationships of the Wollemi Pine (*Wollemia nobilis*) and a molecular phylogeny of the Araucariaceae, Telopea 7 (1997) 275–291.
[16] T. Borsch, K.W. Hilu, D. Quandt, V. Wilde, C. Neinhuis, W. Barthlott, Noncoding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms, J. Evol. Biol. 16 (2003) 558–576.
[17] E.D. Soltis, P.S. Soltis, Contributions of plant molecular systematics to studies of molecular evolution, Plant Mol. Biol. 42 (2000) 45–75.
[18] J.J. Doyle, J.L. Doyle, Chloroplast DNA phylogeny of the papilionoid legume tribe Phaseoleae, Syst. Bot. 18 (1993) 309–327.
[19] R.J. Petit, U.M. Csaikl, S. Bordacs, K. Burg, E. Coart, J. Cottrell, B. van Dam, J.D. Deans, S. Dumolin-Lapegue, S. Fineschi, R. Finkeldey, A. Gillies, I. Glaz, P.G. Goicoechea, J.S. Jensen, A.O. Konig, A.J. Lowe, S.F. Madsen, G. Matyas, R.C. Munro, M. Olalde, M.-H. Pemonge, F. Popescu, D. Slade, H. Tabbener, D. Taurchini, S.G.M. deVries, B. Zeigenhagen, A. Kremer, Chloroplast DNA variation in European white oaks. Phylogeography and patterns of diversity based on data from over 2600 populations, For. Ecol. Manage. 156 (2002) 5–26.
[20] A. Mohanty, J.D. Martin, I. Aguinagalde, A population genetic analysis of chloroplast DNA in wild populations of *Prunus avium* L. in Europe, Hereditary 87 (2001) 421–427.
[21] G.E. McKinnon, R.E. Vaillancourt, H.D. Jackson, B.M. Potts, Chloroplast sharing in the Tasmanian eucalypts, Evolution 55 (4) (2001) 703–711.

[22] M.L. Arnold, C.M. Buckner, J.J. Robinson, Pollen-mediated introgression and hybrid speciation in Lousiana irises, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 1398–1402.
[23] P. Taberlet, L. Gielly, G. Pautou, J. Bouvet, Universal primers for amplification of three non-coding regions of chloroplast DNA, Plant Mol. Biol. 17 (1991) 1105–1109.
[24] B. Demesure, N. Sodzi, R.J. Petit, A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants, Mol. Ecol. 4 (1995) 129–131.
[25] S. Dumolin-Lapegue, M.-H. Pemonge, R.J. Petit, An enlarged set of consensus primers for the study of organelle DNA in plants, Mol. Ecol. 6 (1997) 393–397.
[26] G.J. Bryan, J. McNicoll, G. Ramsay, R.C. Meyer, D. Jong, Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants, Theor. Appl. Genet. 99 (1999) 859–867.
[27] J. Provan, N. Soranzo, N. Wilson, D.B. Goldstein, W. Powell, A low mutation rate for chloroplast microsatellites, Genetics 153 (1999) 943–947.
[28] F. Dane, P. Lang, R. Bakhtiyarova, Comparative analysis of chloroplast DNA variability in wild and cultivated *Citrullus* species, Theor. Appl. Genet. 108 (2004) 958–966.
[29] K. Weising, R.C. Gardner, A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms, Genome 42 (1) (1999) 9–19.
[30] L. Parducci, A.E. Szmidt, A. Madagheile, M. Anzidei, G.G. Vendramin, Genetic variation at chloroplast microsatellites (cpSSRs) in *Abies nebrodensis* (Lojac.) Mattei and three neighboring *Abies* species, Theor. Appl. Genet. 102 (2001) 733–740.
[31] C.S. Echt, L.L. deVerno, M. Anzidei, G.G. Vendramin, Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait, Mol. Ecol. 7 (1998) 307–316.
[32] K.H. Wolfe, W.H. Li, P.M. Sharp, Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 9054–9058.
[33] J.D. Palmer, Mitochondrial DNA in plant systematics: applications and limitations, in: P.S. Soltis, D.E. Soltis, J.J. Doyle (Eds.), Molecular Systematics of Plants, Chapman and Hall, London, 1992, pp. 36–39.
[34] J. Wu, K.V. Krutovskii, S.H. Strauss, Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines, Genetics 150 (1998) 1605–1614.
[35] N. Soranzo, R. Alia, J. Provan, W. Powell, Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations, Mol. Ecol. 9 (2000) 205–1211.
[36] S. Rozen, H.J. Skaletsky, Primer3, Code available at: http://www-genome.wi.mit.edu/genome_software/other/primer3.html, 1997.
[37] M. Schuelke, An economic method for the fluorescent labelling of PCR fragments, Nat. Biotechnol. 18 (2000) 233–234.
[38] C.R. Newton, A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, A.F. Markham, Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS), Nucleic Acids Res. 17 (1989) 2503–2516.
[39] D.L. Swofford, PAUP*. Phylogenetic Anaylsis Using Parsimony (*and other methods). Version 4. 0b10, Sinauer Associates, Inc., Sunderland, Massachusetts, 2002.
[40] M. Kohjyouma, I.J. Lee, O. Iida, K. Kurihara, K. Yamada, Y. Makino, S. Sekita, M. Satake, Intraspecific variation in *Cannabis sativa* L. based on intergenic spacer region of chloroplast DNA, Biol. Pharm. Bull. 23 (6) (2000) 727–730.
[41] The International SNPs Working Group, A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms, Nature 409 (2001) 928–933.
[42] The Arabiposis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, Nature 408 (2000) 796–815.
[43] A. Rafalski, Applications of single nucleotide polymorphisms in crop genetics, Curr. Opin. Plant Biol. 5 (2002) 94–100.
[44] I.V. Gut, Automation in genotyping of single nucleotide polymorphisms, Hum. Mutat. 17 (2001) 475–492.