Forensic Science International 207 (2011) 193–197

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Differentiation of drug and non-drug *Cannabis* using a single nucleotide polymorphism (SNP) assay

D. Rotherham [a], S.A. Harbison [b,*]

[a] Department of Chemistry, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[b] Institute of Environmental Science and Research, Private Bag 92021, Auckland 1047, New Zealand

ARTICLE INFO

Article history:
Received 23 February 2010
Received in revised form 20 September 2010
Accepted 5 October 2010
Available online 30 October 2010

Keywords:
Forensic Science
Cannabis sativa
Tetrahydrocannabinol
DNA
Single nucleotide polymorphism

ABSTRACT

Cannabis sativa is both an illegal drug and a legitimate crop. The differentiation of illegal drug *Cannabis* from non-drug forms of *Cannabis* is relevant in the context of the growth of fibre and seed oil varieties of *Cannabis* for commercial purposes. This differentiation is currently determined based on the levels of tetrahydrocannabinol (THC) in adult plants. DNA based methods have the potential to assay *Cannabis* material unsuitable for analysis using conventional means including seeds, pollen and severely degraded material. The purpose of this research was to develop a single nucleotide polymorphism (SNP) assay for the differentiation of "drug" and "non-drug" *Cannabis* plants. An assay was developed based on four polymorphisms within a 399 bp fragment of the tetrahydrocannabinolic acid (THCA) synthase gene, utilising the snapshot multiplex kit. This SNP assay was tested on 94 *Cannabis* plants, which included 10 blind samples, and was able to differentiate between "drug" and "non-drug" *Cannabis* in all cases, while also differentiating between *Cannabis* and other species. Non-drug plants were found to be homozygous at the four sites assayed while drug *Cannabis* plants were either homozygous or heterozygous.

1. Introduction

Cannabis sativa is one of the world's most prevalent illicit drugs with an estimated 143–190 million people using *Cannabis* during 2007 [1]. The value of the illicit trade in *Cannabis* in New Zealand alone has been estimated at NZ$131–190 million per year [2,3]. *Cannabis* is also, however, a potentially valuable legal crop which can be grown for fibre, seed oil production and bioremediation [4–6]. Differentiation between legitimate "non-drug" *Cannabis* and illicit "drug" *Cannabis* is an important facet of the regulation of the growth of *Cannabis* as a legal crop [7].

Tetrahydrocannabinol (THC) is the principle psychoactive compound present in *Cannabis* [8,9]. There are a number of additional cannabinoids found in *Cannabis*, the major cannabinoid components include cannabigerol (CBG), cannabidiol (CBD), cannabichromene (CBC) and cannabinol (CBN) [10–12]. Non-drug *Cannabis* is typically defined on the basis of THC content; for example in the European Union hemp must have a THC content below 0.2% [7]. In New Zealand the requirement is for THC content to be below 0.35% [13]. Cannabinoid content may be affected by the age or size of the plant tested and the environmental conditions in which it was grown, and this may in turn affect the accurate determination of *Cannabis* chemotype [11].

Although the methods currently available for identification of drug *Cannabis* are reliable and well established [14], a DNA assay able to discriminate between drug and non-drug *Cannabis* would have additional strengths. Foremost among them is the identification of drug *Cannabis* from material unsuitable for analysis using conventional assays for THC content. This may include juvenile plants, seeds, small leaf fragments, pollen, decaying material, partially burnt material and root material [15].

A number of studies have developed DNA assays to identify *Cannabis* samples, without distinguishing between drug and non-drug *Cannabis* [16–18]. Additionally, de Meijer et al. [19] reported a sequence characterised amplified region or SCAR marker able to differentiate between drug and non-drug *Cannabis* that has been developed from a randomly amplified polymorphic DNA (RAPD) marker associated with high THC in *Cannabis*. This marker was associated with THC/CBD phenotype rather than intrinsically linked to THC synthesis and was not able to unambiguously classify all samples tested [19].

The synthesis of THC in *Cannabis* involves the conversion of a number of precursors by a series of synthase enzymes. The final step in the synthesis of THC is the conversion of cannabigerolic acid (CBGA) into tetrahydrocannabinolic acid (THCA) catalysed by the enzyme THCA synthase [20,21]. THCA is then decarboxylated to THC [21]. This process is mirrored by the conversion of CBGA to cannabidiolic acid (CBDA) and cannabichromenic acid (CBCA) by CBDA synthase and CBCA synthase respectively, followed by subsequent decarboxylation to cannabidiol (CBD) and cannabi-

* Corresponding author. Tel.: +64 9 815 3670; fax: +64 9 849 6046.
E-mail address: sallyann.harbison@esr.cri.nz (S.A. Harbison).

0379-0738/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2010.10.006

chromene (CBC) [22,23]. Drug strains of *Cannabis* are typically high in THC. Oil and fibre strains of *Cannabis* are typically dominated by CBD and occasionally cannabigerol (CBG), the decarboxylated form of CBGA [24]. CBC is found at high levels in juvenile *Cannabis* plants and in strains with a persistent juvenile state [25].

Kojoma et al. [26] sequenced the THCA synthase genes of six drug and seven fibre strains of *Cannabis*. Comparison of these sequences revealed two distinct forms of the THCA synthase gene, one found in the six drug strains the other found in the seven fibre strains. There were a total of 63 nucleotide substitutions differentiating all six drug strain sequences from the seven fibre strain sequences, these corresponded to 37 amino acid substitutions in the THCA synthase gene product. Kojoma et al. [26] considered these divergent THCA synthase sequences to represent alleles coding for an active and an inactive form of the THCA synthase enzyme.

Kojoma et al. described a set of PCR primers used to amplify a 1.2 kb fragment of the proposed active THCA synthase sequence found in the six drug strains [26]. A 1.4 kb fragment of the ribulose bisphosphate carboxylase gene (rbcL) was amplified as a positive control. The principal drawback of this THCA synthase marker is the length of the fragment amplified which may make amplification more difficult, particularly from degraded samples such as those likely to be encountered at crime scenes [27–30].

The aim of this study was to develop a single nucleotide polymorphism (SNP) assay more suited to crime scene samples capable of discriminating between high and low THC *Cannabis* varieties based on sequence variation in the THCA synthase gene and to test this assay on drug and non-drug varieties of *Cannabis*.

2. Materials and methods

2.1. Primer design

The SNP assay was designed based around the single base extension (SBE) protocol of the ABI SNaPshot™ multiplex kit.

A 399 base pair (bp) fragment of the THCA synthase gene was amplified from both drug and non-drug *Cannabis* using the primers C and E of Kojoma et al. [26] with two modifications: two degenerate bases were added to primer C to account for differences between the active and inactive forms of the THCA synthase sequence and the terminal T was removed from primer E to bring the melting temperature closer to that of primer C. These modified primers are referred to as C2 and E2 (Table 1). Primer C2 binds 738 bp from the start of the 1653 bp THCA synthase sequence, while primer E2 (reverse) is located 516 bp from the end of the sequence as shown in Fig. 1.

Extension primers were designed to target four non-synonymous polymorphisms within the THCA synthase gene (Table 1). The active and inactive THCA synthase sequences of Kojoma et al. [26] are differentiated by 63 SNPs that differed in state between the 6 drug type strains and 7 non-drug strains sequenced. The majority of the 63 single nucleotide differences between the active and inactive forms of THCA synthase were not suitable for use as markers for THCA phenotype on the basis of synonymy, similarity with the closely related THCA synthase gene, suitability of flanking primer binding sites and restrictions on amplicon size.

SNP markers were selected based on the following criteria: (1) selected SNPs were non-synonymous (i.e. corresponded to amino acid differences between the active and inactive THCA sequence protein products); (2) the nucleotide state in the drug form was not shared with that in the published sequence of the closely related CBDA synthase gene; (3) transversions (C-A, C-G, T-A, T-G) were preferred to transitions (C-T, A-G); given that transversions are statistically less likely to occur [31], back mutation to the original state is considered to be less likely.

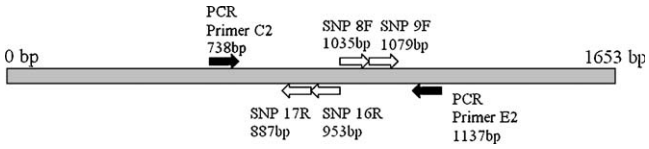

Fig. 1. Binding position of SNP primers on the THCA synthase gene.

These criteria left a set of 20 SNPs for which the forward and reverse sequences were considered during primer design. Four SNP primers were selected to be used in the analyses, all four SNPs had melting temperatures less than 1° either side of 57 °C, a separation of at least five base pairs in length between each primer for ease of analysis.

In addition to the requirements stated above the SNPs selected and their extension primer binding sites had to fall within a single readily amplified PCR amplicon. An assay based on the amplification of the entire THCA synthase gene was considered impractical from a forensic perspective as samples may be of poor quality with potentially degraded DNA. The four markers selected for the final assay fell within the 399 bp fragment amplified by primers C and E of Kojoma et al. [26].

SNPs 8F and 9F were assayed with forward primers, SNPs 16R and 17R were assayed with reverse primers.

SNP 17R is a transition located 887 bp from the beginning of the THCA synthase sequence. The active or drug form of the THCA synthase gene carries an adenine (A) at this locus while the inactive form carries a guanine (G). As the extension primer for this SNP is a reverse primer a red labelled thymine (T) is incorporated during the mini-sequencing reaction for the active form of THCA synthase and a yellow labelled cytosine (C) is incorporated for the inactive form of THCA synthase, giving rise to a red peak for the active form of THCA synthase and a yellow peak, displayed in black during analysis, for the inactive form of THCA synthase.

SNP 16R is a transversion at 953 bp; the polymorphism is an adenine in the active form of the THCA synthase gene and a thymine in the inactive form of the THCA synthase gene. It corresponds to a histidine residue in the active form of THCA synthase and a leucine residue in the inactive form of THCA synthase. As the extension primer for this SNP is a reverse primer a red labelled thymine is incorporated during the mini-sequencing reaction for the active form of THCA synthase and a green labelled adenine is incorporated for the inactive form of THCA synthase, giving rise to a red peak for the active form of THCA synthase and a green peak for the inactive form of THCA synthase.

SNP 8F is a transversion at 1035 bp with a thymine in the active/drug form and a guanine in the inactive/non-drug form, corresponding to a phenylalanine residue in the active form and a tyrosine in the inactive form. As the extension primer for this SNP is a forward primer a red labelled thymine is incorporated during the mini-sequencing reaction for the active form of THCA synthase and a blue labelled guanine for the inactive form of THCA synthase.

SNP 9F is a transversion at 1079 bp with a thymine in the active/drug form and an adenine in the inactive/non-drug form, corresponding to a lysine residue in the active form and an arginine in the inactive form. As the extension primer for this SNP is a forward primer a red labelled thymine is incorporated during the mini-sequencing reaction for the active form of THCA synthase and a green labelled adenine for the inactive form of THCA synthase.

2.2. Sample collection

A total of 79 drug-type *Cannabis* plants and 15 non-drug *Cannabis* plants were analysed. An additional five non-*Cannabis* plant species were analysed to test potential cross-species amplification. *Humulus lupulus* (Common Hop), *Celtis sinensis* (Chinese Hackberry), *Ficus macrophylla* (Moreton Bay Fig) and *Ulmus procera* (English Elm) were selected as relatives of *Cannabis* [32]. *Nicotiana tabacum* (Cultivated Tobacco) was selected on the basis that it may be mixed with *Cannabis* for drug use [33].

Drug-type *Cannabis* samples were obtained from seized materials received at the Institute of Environmental Science and Research (ESR). Hemp samples were obtained from material submitted for cannabinoid testing at ESR, with permission from the suppliers. THC levels were quantified by gas chromatography mass spectrometry (GCMS) for all hemp samples and 51 of the 79 drug-type *Cannabis* samples received.

Table 1
Table of amplification and extension primers used, including oligonucleotide sequence.

Primer	Primer sequence (5′ to 3′)	Primer length	Start position	Direction
PCR primer C2	CAAACTKGTTGYTGTCCCATC	21 bp[a]	738 bp[b]	Forward
PCR primer E2	CGTCTTCTTCCCAGCTGATC	20 bp	1137 bp	Reverse
Extension primer 8F	GAGTTGGGTATTAAAAAAACTGATTGCAAAGAATT	35 bp	1035 bp	Forward
Extension primer 9F	CAACCATCTTCTACAGTGGTGTTGTAAATT	30 bp	1079 bp	Forward
Extension primer 16R	TCRACTAGACTATCCACTCCACCA	24 bp	953 bp	Reverse
Extension primer 17R	TACTGTAGTCTTATTCTTCCCATGATTATCTGTAATATTC	40 bp	887 bp	Reverse

[a] Length of oligonucleotide in base pairs.
[b] Distance from start of sequence in base pairs.