Atlas Journal of Biology 2016, pp. 292–294
doi: 10.5147/ajb.2016.0142

A Rapid Method for Cannabis Species Determination by DNA Sequencing

David A. Lightfoot[1,2,*], Winston C. Throgmorton[3], and Colton Johnson[3]

[1] Genome and Agricultural Biotechnology LLC, Carbondale, IL 62901, USA; [2] PO Box 0312, Carbondale, IL 62903, USA; [3] Throgmorton Attorney at Law, 304 N. Monroe Street, Marion, IL, USA. Tel: (618) 993-5379.

Received: September 19, 2016 / Accepted: October 9, 2016

Abstract

Determination of species within the genus *Cannabis* has important legal medical and social implications. Recent genome sequencing has shown that the genomes of *C. sativa* (recreational marijuana and hemp), *C. indica* (medical marijuana) and *C. ruderalis* (feral marijuana) can all be distinguished. However, hybridization among the species has occurred with widely varying outcomes in the percent of genome transmitted. The aim here was to determine if a simple assay based on the DNA sequence of ITS2 could be used to distinguish among species. Using sequences at GenBank as a reference eighteen plant samples were sequenced and shown to be identical to *C. indica* sequence and different from *C. sativa* and *C. ruderalis* at 4 positions within the 25S rRNA gene. This result and the geographic separation of the centers of genetic diversity argues strongly for polytypic origins of the 3 species. Analysis of interspecific hybrids sequences at GenBank suggested only the *C. indica* allele is transmitted preferentially. Finally, a SNP within the ITS could be used to distinguish two types within the eighteen plants. Therefore this simple genetic test can be used for rapid plant identification and to assist in strain identification.

Introduction

Internal transcribed spacer (ITS) refers to the spacer DNA (non-coding DNA) situated between the small-subunit ribosomal RNA (rRNA) and large-subunit rRNA genes in the chromosome (Coleman, 2007). There are two ITS's in eukaryotes; ITS1 is located between 18S and 5.8S rRNA genes, while ITS2 is between 5.8S and 25S (in plants) rRNA genes.

Sequence comparison of the ITS region is widely used in taxonomy and molecular phylogeny because it a) is easy to amplify even from small quantities of DNA (due to the high copy number of rRNA genes), and b) has a high degree of variation even between closely related species (Coleman, 2007; Shultz et al., 2006). The success rates for using the ITS2 region to identify dicotyledons plants was 76.1%, at the species level (Kress et al., 2006; Yao et al., 2010).

The ITS region is the most widely sequenced DNA region in molecular ecology of plants and has been recommended among the universal barcode sequences (Kress et al., 2006; Yao et al., 2010) It has typically been most useful for molecular systematics at the species level, and even within species (e.g., to identify geographic races, varieties and ancestoral types). The standard ITS1 and ITS4 primers are used by most labs (White et al., 1990).

* Corresponding author: ga4082@siu.edu

Determination of species within the genus Cannabis has important legal medical and social implications. However, the literature to date has been confused on some key concepts because of the remarkable amount of genetic variation and phenotypic plasticity within the genus (Schultes et al., 1974; Sawler et al., 2015). Selection, cultivation, hybridizations by both man and nature, crop abandonment and feral growth have all combined to confuse the identities and origins of the four crops, C. *sativa* (recreational marijuana and hemp), C. *indica* (medical marijuana) and C. *ruderalis* (feral marijuana). However, it seems possible they had polytypic origin since their centers of wild plant genetic diversity differ, East Asia, India and Russia for the 3 species respecctively.

Recent genome sequencing has shown that the genomes of C. *sativa* (recreational marijuana and hemp), C. *indica* (medical marijuana) and C. *ruderalis* (feral marijuana) can all be distinguished (Sawler e al., 2015). However, hybridization among the species has occurred with widely varying outcomes in the percent of genome transmitted. Observations of phenotypes suggest the C. *indica* type is dominant with preferential transmission of its alleles in known interspecific hybrids. Preferential transmission of alleles is widely observed in interspecific crosses with one parent being dominant in all crosses made (Guo et al., 1990).

Here is described a simple and rapid genetic test that distinguishes

Materials and Methods

Sampling

On February 10, 2015 eighteen samples from eighteen plants were collected by Prof. David A Lightfoot, Principal Scientist, GAAB-LLC, Carbondale, IL 62901, USA. An analysis of the samples species identity had been requested. They were provided as follows.

DNA Extractions

All eighteen samples were handled in the same way. Gloves were worn by Dr. Lightfoot throughout. All vials and plastic ware were new and had been sterilized. The Promega GenewizTM kit was used for purifications. The kit was new.

PCR Conditions

Following White, et al (1990), primers were:
5'-AGC CGC CTT CAT ATA TCT GCT T -3' ITS1- forward;
5'-TCC TCC GCT TAT TGA TAT GC -3' ITS4 - reverse.
Polymerase chain reaction (PCR) was performed using 25, 50 or 75 ng of genomic DNA, 225 µl of SuperMix High Fidelity Enzyme (GibcoBRL, Grand Island, NY), 5 µl H_20, 5 µl ITS 1-Forward, and 5 µl ITS4-reverse in a total volume of 20 µl. Cycling conditions were: 95^{0}C for 1 min, then 45^{0}C for 1 min, then 68^{0}C for 1 min for 35 cycles finishing with a and 4^{0}C hold. The PCR products were prepared for Sanger sequencing reactions by alkaline phosphatase and exonuclease 1 treatment.

DNA Sequencing

The DNA in samples sent for forward DNA sequencing using services provided by GeneWiz Inc. (Plainsfield, NJ, USA) using the amplification primer ITS1. DNA sequences were obtained that were found in the gene data repository at NCBI. The sequences were aligned with reference sequences for C. *sativa* (record identifier FJ572045.1) and C. *indica* (record identifier KC292629.1) found at NCBI by BLAST searching. A new sequence for C. *indica* was deposited at GenBank as accession number KX980526.

```
C. indica     652  GACTAGACCCGTACGACCCCAATGTGCTGCGAACGCAGTGCCTTCAACGCGACCCCAGGT  711
C. sativa     601  ..........................................................C.TA  660
Query_65399   560  ............................................................  619
Query_65400   563  ............................................................  622
Query_65401   564  ............................................................  623
Query_65402   560  ............................................................  619
Query_65403   562  ............................................................  621
```

Figure 1. A complex rearrangement in the 25S rRNA gene distinguishes C. *indica* from C. *sativa*.

```
C. indica     472  GGGCGTCACACGCCGTTGCCCCCATGTGCACTGCCAAAAGCGTGTTCAGGAGGGGCGGAG  531
C. sativa     421  ............................................................  480
Query_65399   380  ..........................................................T.....  439
Query_65400   383  ............................................................  442
Query_65401   384  ............................................................  443
Query_65402   380  ..........................................................T.....  439
Query_65403   382  ..........................................................T.....  441
```

Figure 2. A SNP in the ITS2 region identified two strains of C. *indica*.

```
C.indica      661   CGTACGACCCCAATGTGCTGCGAACGCAGTGCCTTCAACGCGACCCCAGGTCAGGCGGGA  720
KC292629      661   ............................................................  720
KF800374      811   ............................................................  870
FJ572045      610   ..........................................C.TA..............  669
JQ230978      615   ............................................................  674
DQ267929      600   ..........................C.............T.................  659
AB564722      623   ............................................................  682
KF454086      266   ............................................................  325
KF454085      266   ............................................................  325
KF454084      266   ............................................................  325
KF454083      266   ............................................................  325
```

1. C. indica reference voucher sample
2. Cannnabis sativa subsp. indica voucher 20121121-846
3. Uncultured eukaryote clone CMH283 18S ribosomal RNA gene found in house dust
4. Cannabis sativa 18S ribosomal RNA gene, industrial hemp
5. Cannabis sativa voucher SBB-1163 18S ribosomal RNA gene, India
6. Dioscorea sansibarensis 18S ribosomal RNA gene, probably a contaminant
7. Cannabis sativa genes for 18S rRNA, Japan
8. Cannabis sativa voucher 20120903569, China
9. Cannabis sativa voucher 20120903568, China
10. Cannabis sativa voucher 20120903567, China
11. Cannabis sativa voucher 20120903566, China

Figure 3. Alignment of sequences at GenBank.

Results

To date good sequences have been obtained from 14 of the 18 samples. In each case the sequences contained the tetra-nucleotide AGGT between nucleotides 708 to 711 of the reference genome sequence of C. *indica* (record identifier KC292629.1). They all differed from the CGTA at that position found in C. *sativa* (record identifier FJ572045.1). An example alignment is shown in (Figure 1). No further difference were seen between C. *sativa* and C. *indica* except a single A/T SNP at position 522. The new sequence for C. *indica* was deposited at GenBank as accession number KX980526. The T SNP was unique to the samples tested and was not present in either reference sequence or on other GenBank records. Examination of the 12 homologous sequences found on GenBank (Figure 3) showed all but two to be C. *indica* though seven were miss-identified as C. *sativa*. One sample from house dust was also C. *indica*. Not shown were several realted species including Celtis sp. that aligned starting at the region of rRNA transcription. In a remarkable example of convergent evolution sequence of a monocot Dioscorea sansibarensis was well aligned with the C*anabis* sp. too. Alternately, DQ267929 was a contaminated sequence submitted in error.

Conclusion

Therefore, it was concluded the tested plants were all C. *indica*. Because the region sequenced is a primary diagnostic test of species and genus when used by systematists around the world the conclusion cannot be reasonably refuted (Schultz et al., 2006; Yao et al., 2010). The nature and position within the 25S rRNA gene coding sequence suggests a separate origin for the two species which would agree with their separate centers of genetic diversity (Shultes et al., 1974; Sawler at al., 2015).

References

Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucleic Acids Res 35: 3322–3329

Guo M, DA Lightfoot, MC Mok, and DWS Mok (1991) Analyses of *Phaseolus vulgaris* L. and *Phaseolus coccineus* Lam. hybrids by RFLP: Preferential transmission of *P. vulgaris* alleles. Theor. Appl. Genet. 81:703-710.

Kress WJ, KJ Wurdack, EA Zimmer, LA Weigt, and DH Janzen (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102: 8369–8374.

Sawler J, JM Stout, KM Gardner, D Hudson, J Vidmar, L Butler, et al. (2015) The Genetic Structure of Marijuana and Hemp. PLoS ONE 10(8): e0133292.

Schultes RE, WM Klein, T Plowman, and TE Lockwood (1974) Cannabis: an example of taxonomic neglect. Harvard University Botanical Museum Leaflets 23: 337-367.

Schultz J, T Muller, M Achtziger, PN Seibel, T Dandekar, et al. (2006) The internal transcribed spacer 2 database - a web server for (not only) low level phylogenetic analyses. Nucleic Acids Res 34: W704–W707.

White TJ, T Bruns, S Lee, and JW Taylor (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 In: PCR Protocols: A Guide to Methods and Applications, eds. Innis MA, DH Gelfand, JJ Sninsky, and TJ White. Academic Press, Inc., New York.

Yao H, J Song, C Liu, K Luo, J Han, et al. (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5: e13102.