Evaluation of Chloroplast DNA Barcoding Markers to Individualize *Papaver somniferum* for Forensic Intelligence Purposes

Kari A. Graham, BS* and Rachel Houston, PhD

Department of Forensic Science, Sam Houston State University, Huntsville, TX 77340

INTRODUCTION

Papaver somniferum, also known as opium poppy, is the primary source of natural alkaloids like morphine and codeine. Unfortunately, these highly addictive compounds and over-prescribing have led to an opioid epidemic in the United States responsible for over 40,000 deaths a year [1].

Stringent controls have been placed on processed and unprocessed forms of *P. somniferum*, with all plant parts, excluding the seeds, scheduled under the Controlled Substances Act. The controls placed on prescription opioids and their relatively high cost have led to the illegal production of opioids, namely heroin. Heroin is derived from morphine, which can be found in the sap or "latex" of poppy pods and has analgesic effects more potent than morphine.

The Heroin Signature Program was developed for source tracking and distribution of heroin by chemical analysis. While this program has been moderately successful, exploiting genetic information from *P. somniferum* could strengthen the program.

The use of nuclear poppy DNA has shown to be largely unsuccessful due to the low quality and quality of poppy DNA extracted from heroin [2]. However, the use of chloroplast DNA (cpDNA) markers may be a potential solution due to the abundance of chloroplasts per cell regardless of tissue type. In addition, cpDNA barcoding markers have also been shown in the literature to be hotspot regions for interspecies and intraspecies variability [3,4].

This study evaluated various chloroplast barcoding regions in a variety of *P. somniferum* sources to assess intraspecies variability. These barcoding regions have the potential to be used in a comprehensive analytical tool used for biogeographical determination and case linkage where opium poppy and its derivatives are seized.

MATERIALS AND METHODS

In Silico Analysis of Published Genomes

Previously published genomes of *P. somniferum* (NC_029434, KU204905, MT712162) were aligned, and screened for polymorphic sites using Geneious Prime® Software (Biomatters).

Sample Collection

Poppy seed samples were collected, extracted, and quantified previously [2]. Samples (n=10) were randomly chosen based on DNA concentration (**Table 1**).

Table 1: Sample information

Sample ID #	Vendor	Origin	Plant type claimed on packing
1	TerraVita	Canada	California Poppy
4	Medley Hills Farm	Ohio	Blue
6	Frontier Co-Op	Turkey	*P. somniferum*
9	Natural Traders Co	Turkey	Blue
10	Food to Live	England	Not Specified
11	Anna and Sarah	Holland	Blue
12	Nature's Gourmet Classics	Not Specified	Not Specified
18	Nodding Turtle	Afghanistan	Afghan Blue *P. somniferum*
22	Sincerely Nuts	U.K.	Not Specified
23	Unknown	Not Specified	Not Specified

RESULTS AND DISCUSSION

Table 2: Polymorphisms found during *in silico* comparison of *P. somniferum* chloroplast reference genomes.

Start Location in Consensus Sequence (bp)	Gene Name	Polymorphism Type	NC_029434	KU204905	MT712162
13105	atpA-atpF	INDEL (T)	Absent	Absent	Present
29400	psbM-trnD	INDEL (A)	Absent	Absent	Present
57081	accD	SNP (T/A)	T	T	A
113754	ccsA	SNP (A/G)	A	A	G
124540	ycf1	SNP (G/A)	G	G	A

- *In silico* comparisons of *P. somniferum* reference genomes revealed little intraspecies variation and no hotspot regions across the chloroplast genome (**Table 2**).
- Due to this result, cpDNA barcoding regions that have been shown to display interspecies and intraspecies variation were chosen to be sequenced and screened.
- In barcoding region *trnH-psbA*, 10% (1/10) of samples showed 6 SNPs: A→G at bp 295, T→A at bp 296, C→A at bp 298, T→G at bp 299, T→A at bp 301, and C→T at bp 302 (**Figure 1**).
- In barcoding region *petA-psbJ,* 20% (2/10) of samples showed a GGTG deletion at bp 62283-62286, an A→T SNP at bp 62290, and a ACCC insertion at bp 62295-62298 (**Figure 1**).
- The variable samples (6 and 9) were both cultivated in Turkey.

Table 3: Primer annealing temperature results for each of the barcoding regions used in this study and their respective fragment sizes.

Locus	Forward Primer Sequence	Reverse Primer Sequence	Annealing Temperature (°C)	Fragment Size (bp)
ndhF-rpl32	GCCCCACCCCTATTGCTATA	CGCAGCCAAATATCCTTTTCTT	63.5	1250
petA-psbJ	ACAATTCGAGAAGGTTCAGTTGT	CCGGTATTCCTGTGATCGGT	63.5	1232
rpl32-trnL	AAAGGATATTTGGCTGCGCT	GCGTGTCTACCGATTTCACC	63.5	689
rps16-trnQ	CATGTCCTTCAAGTCGCACG	TTCGAATCCTTCCGTCCCAG	68.1	1192
trnE-trnT	GAGAGATGTCCTGAACCGCT	CCGCTGAGTTAAAAGGGCTC	66.0	518
trnH-psbA	TCCACTGCCTTGATCCACTT	CCGTGCTAACCTTGGTATGG	63.5	382
trnL-trnF	GGTAGACGCTACGGACTTGA	GTCCTCTGCTCTACCAACTGA	66.0	980
rpl16 intron	TCGCCCTGCTTCTATTTGTC	GCTTAGTGTGTGACTCGTTGG	63.5	1152
ndhC-trnV	GCCCATTGGTTCTATGCCTG	CGAGAAGGTCTACGGTTCGA	63.5	1664
trnT-psbD	GCCCTTTTAACTCAGCGGTAG	CCTCCGTAACCAGTCATCCA	63.5	1446
psbE-petL	GACGAATAACCAACCCGCAA	CCGCCAGTAGAAAACCGAAA	63.5	1400

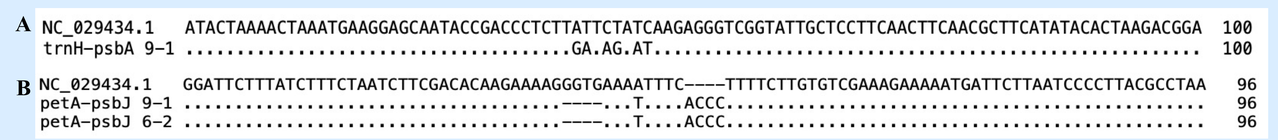

Figure 1: Extracted sequences (*trnH-psbA*: bp 257-356, *petA-psbJ*: bp 62283-62341) from polymorphic samples aligned to the reference genome. **A.** Only sample 9 displayed polymorphisms different from the reference sequence: A→G at bp 295, T→A at bp 296, C→A at bp 298, T→G at bp 299, T→A at bp 301, and C→T at bp 302. **B.** Both samples 6 and 9 showed a 4 bp GGTG deletion (bp 62283-62286), A → T SNP (bp 62290), and a 4 bp ACCC insertion (bp 62295-62289).

REFERENCES

[1] C.L. Mattson, L.J. Tanz, K. Quinn, M. Kariisa, P. Patel, N.L. Davis, Trends and Geographic Patterns in Drug and Synthetic Opioid Overdose Deaths - United States, 2013-2019, MMWR Morb Mortal Wkly Rep, 70 (2021) 202-207.
[2] B. Young, M.G. Roman, B. LaRue, D. Gangitano, R. Houston, Evaluation of 19 short tandem repeat markers for individualization of Papaver somniferum, Science & Justice, 60 (2020) 253-262.
[3] E.J. Lee, I.K. Hwang, N.Y. Kim, K.L. Lee, M.S. Han, Y.H. Lee, M.Y. Kim, M.S. Yang, An assessment of the utility of universal and specific genetic markers for opium poppy identification, Journal of forensic sciences, 55 (2010) 1202-1208.
[4] J. Shaw, H.L. Shafer, O.R. Leonard, M.J. Kovach, M. Schorr, A.B. Morris, Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV, American Journal of Botany, 101 (2014) 1987-2004.

MATERIALS AND METHODS

Primer Design and Annealing Temperature Determination

Primers were designed to include the entirety of the barcode region of interest. Design was accomplished using Primer3 software. Annealing temperatures for all primer sets were determined via gradient PCR using the Type-it Microsatellite Kit (QIAGEN) and an Eppendorf Mastercycler® Gradient (Eppendorf) thermal cycler. Optimal annealing temperature was determined via 2% agarose gel electrophoresis.

PCR Amplification

Each barcoding region (*ndhF-rpl32, petA-psbJ, rpl32-trnL, rps16-trnQ, trnE-trnT, trnH-psbA, trnL-trnF, rpl16* intron*, ndhC-trnV, trnT-psbD,* and *psbE-petL*) was amplified using the Type-it Microsatellite Kit (QIAGEN). Thermal cycling conditions on an Eppendorf Mastercycler® Gradient were as follows: initial activation for 5 min at 95 ºC, followed by 30 cycles of 95 ºC for 30 s, 1.5 min at the optimal annealing temperature (**Table 3**), and 72 ºC for 30 s, then a final extension for 30 min at 60 ºC.

Sanger Sequencing

Cycle sequencing was performed for forward and reverse reactions for each sample using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo-Fisher) per manufacturer's instructions on an Eppendorf Mastercycler® Gradient thermal Cycler. Capillary electrophoresis was performed on a 3500 Genetic Analyzer (Thermo-Fisher).

Data Analysis

Sequences were aligned and proofread using Geneious Prime® Software. Each barcoding region was aligned to the reference genome (NC_029434) and screened for variants.

CONCLUSIONS

- No hotspot regions were discovered in *P. somniferum* based on *in silico* analysis of reference genomes (**Table 2**).
- Designed *P. somniferum* specific primers for 11 hotspot regions (**Table 3**).
- Barcoding regions *trnH-psbA* and *petA-psbJ* showed the most promise in terms of use for intraspecies variation and source tracking of seized heroin samples (**Figure 1**)
- Screening other barcoding markers and expanding the sample database may find other polymorphic regions that this initial screening may have missed.
- *P. somniferum*, like other plant-derived drugs of abuse, is selectively bred for its alkaloid content. However, CpDNA does not affect *P. somniferum* alkaloid content, which may explain the lack of variety discovered in this study.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Department of Forensic Science at Sam Houston State University for supporting this research.