Forensic botany and forensic chemistry working together: application of plant DNA barcoding as a complement to forensic chemistry—a case study in Brazil

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>gen-2018-0066.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>04-Dec-2018</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Paranaiba, Renato; Policia Federal, Instituto Nacional de Criminalística Carvalho, Carlos; Policia Federal, Instituto Nacional de Criminalística Freitas, Jorge; Policia Federal, Instituto Nacional de Criminalística Fassio, Levy; Policia Federal, Instituto Nacional de Criminalística Botelho, Élvio; Policia Federal, Instituto Nacional de Criminalística Neves, Diana; Policia Federal, Instituto Nacional de Criminalística Silva Junior, Ronaldo; Policia Federal, Instituto Nacional de Criminalística Aguiar, Sérgio; Policia Federal, Instituto Nacional de Criminalística</td>
</tr>
<tr>
<td>Keyword:</td>
<td>plant, DNA, barcoding, forensic, botany</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue? :</td>
<td>7th International Barcode of Life</td>
</tr>
</tbody>
</table>

https://mc06.manuscriptcentral.com/genome-pubs
Forensic botany and forensic chemistry working together: application of plant DNA barcoding as a complement to forensic chemistry—a case study in Brazil

Renato T. F. Paranaiba1*, Carlos B. V. Carvalho1, Jorge M. Freitas1, Levy H. Fassio1, Élvio, D. Botelho2, Diana B. J. Neves2, Ronaldo C. Silva Jr.1, Sérgio M. Aguiar1

1DNA Laboratory, National Institute of Criminalistics, Brazilian Federal Police. SAIS Quadra 7, Lote 23, Brasília, DF, Brasil, 70610-200.

2Chemistry Laboratory, National Institute of Criminalistics, Brazilian Federal Police. SAIS Quadra 7, Lote 23, Brasilia, DF, Brasil, 70610-200.

* Corresponding author’s email address: renatoparanaiba@gmail.com
Recently, Brazilian Federal Police used forensic chemistry and forensic botany techniques on a case. Two packets containing fragmented plant matter were seized and sent for forensic analysis. Forensic chemistry, the gold standard for evaluating plant material suspected to contain illicit substances, did not find illicit materials. Gas chromatography coupled mass spectrometry (GC-MS) identified thujone in the botanical material. Thujone is a chemical compound naturally found in many plant species, notably Artemisia absinthium. Because doubt remained, we next used plant DNA barcoding methods. Total DNA from plant tissue fragments was extracted and five different DNA regions were amplified, sequenced, and analyzed using plant DNA barcoding methods. Genetic analysis yielded 30 good quality sequences representing five taxa. Most specimens were identified as Artemisia absinthium. Few studies focus on practical forensic applications of plant DNA barcoding methods using a case solved in a forensic laboratory with its difficulties and limitations. To the best of our knowledge, this is the first study to report an effective joint effort of forensic chemistry and botany techniques to assess plant material in Brazil. The availability of a new technical approach for the genetic sequencing of plant species will enhance many forensic investigations and inspire similar initiatives.

**Key words:** barcoding, DNA, forensic botany, forensic chemistry, plant.
Forensic botany and forensic chemistry working together: application of plant DNA barcoding as a complement to forensic chemistry—a case study in Brazil

**Introduction**

In 2012, the Forensic Genetics Laboratory of the Brazilian Federal Police began genetic identification of plant species using DNA barcoding. During the internship of an expert from the Laboratory in the Royal Botanic Garden Edinburgh (RBGE), practical knowledge was obtained for establishing the techniques in Brazil (Horton 2012). Following this short collaboration with the RBGE and two years of intensive research, preparation, and testing, the laboratory produced its first forensic report using genetic identification of plant species through DNA barcoding.

We present a novel forensic case with unique characteristics. Given the interdisciplinary nature of the techniques used during the analysis and the number of species involved, this case represented a learning opportunity for the laboratory, which is now more prepared to provide services for the Brazilian judiciary system.

Living plants and dried vegetable matter are frequently seized items suspected of containing illicit drugs/controlled chemicals. Analyses of these items are routinely performed using forensic chemistry techniques, which is chemistry applied to the identification of illegal substances within the criminal justice system. Using a capillary column gas chromatograph and a suitable temperature program coupled to a mass spectrometer (GC-MS), the drug components of most botanical materials can be identified. Forensic chemistry is inherently adaptive, shaping fundamental advances in science, technology, and instrumentation for forensic use. It is the ultimate applied chemical science (Bell 2009).

However, there are many situations where forensic chemistry is limited and macroscopic and microscopic examinations may fail. Cases involving seeds, tiny amounts of material, or highly processed vegetables may cause problems in plant identification. In addition, doubts in
forensic litigations may remain after routine analysis. Thus, plant DNA barcoding emerged as a powerful tool for routine forensics applications in forensic botany. It is a reliable and sensitive technique unaffected by factors external to the genetic constitution of the plant such as age and climate. Plant DNA Barcoding is underutilized because of a lack of specialists and plant diversity. Molecular techniques, however, have altered this picture, allowing the increasingly common use of plants and their fragments as forensic evidence (Chandra and Sharma 2014).

Plant DNA barcoding identifies species using recent advances in genetics, genomics, and bioinformatics using short DNA sequences, known as the DNA barcode. Hebert et al. established this concept, which is based on the comparison of sequences obtained from unknown species against a reference database (Hebert et al. 2003). It is independent of the plant part, which is beneficial in forensics. DNA barcoding methods were recently applied to the authentication of herbal drug materials for industrial quality assurance (Sgamma et al. 2017).

Some studies suggest that DNA barcoding should be used in a complementary manner with chemical analyses for species identification in herbals (Abubakar et al. 2017).

Here, we describe a case where DNA barcoding was pivotal. In 2013, two packets containing botanical material, homogenous at first sight, were seized in the mail. When suspect packets (relating to illicit drugs) are detected, they are sent to Brazilian Federal Police, where standard analysis is based on chemical tests.

Material and Methods

Material

Two packets containing botanical material consisting of fragmented and/or smashed leaves, flowers, seeds, and stems were received.

Chemical analysis
One sample from each of the two packets was analyzed. An aliquot of the dried leaf product was placed in a mortar and pulverized to a fine powder. For the extraction, 0.3 g of each sample was placed in a tube and 2 mL of a methanol–chloroform mixture (1:1) was added (HPLC-grade; JT Baker (USA)). The tubes containing the samples were shaken in a vortex for 1 minute. The samples were filtered and injected in a GC-MS. To perform the GC/MS analysis, an Agilent 6890N GC equipped with an RXi-1MS capillary column (25 m × 200 μm × 0.33 μm) was employed. Helium was used as the carrier gas; the injection port was maintained at a temperature of 250 °C, and a split ratio of 10:1 was used. The column oven temperature program was as follows: initial temperature of 100 °C for 2.0 min, followed by linear 12 °C/min increments up to 300 °C, and hold for 10.0 min. Total chromatographic run time was 28.7 min. Detection was performed with a quadrupole mass spectrometer (Agilent 5973 inert). The m/z range was 30–550 amu. Identification of substances was performed by comparison of the mass spectra obtained with the National Institute of Standards and Technology (NIST) electronic library.

DNA extraction

The material was inspected visually and fragments that appeared to be from a different species picked out. Ten samples were chosen, six from the first packet and four from the second. Then, they were ground in a mixer mill (MM400 Retsch®). Four are shown in Figure 1. Total DNA from each plant tissue fragment was extracted with PrepFiler Express BTA™ Forensic DNA Extraction (Thermo Fisher Scientific Inc., Waltham, MA, USA) employing an AutoMate Express™ DNA Extraction System robotic workstation (Thermo Fisher Scientific Inc., Waltham, MA, USA).

PCR Amplification
Seven universal PCR primers were used to amplify five different land plant DNA regions—\textit{matK}, \textit{rbcL}, \textit{trnL/IGS–intergenic spacer}_\textit{trnL-trnF}, \textit{ycf1}, and ITS1/ITS2 (Yu et al. 2011, Chen et al. 2010, Taberlet et al. 2007, Fangan et al. 1994, Dong et al. 2015, White et al. 1990). We used a \textit{Cymbopogon citratus} sample as a positive control. PCR conditions, depending on the primer, were used as recommended (Yu et al. 2011, Chen et al. 2010, Taberlet et al. 2007, Fangan et al. 1994, Dong et al. 2015, White et al. 1990).

Using a 1\% agarose gel electrophoresis assisted by ultraviolet detection, DNA fragments generated from PCR were visualized. PCR products were purified with Exonuclease I (New England Biolabs Ltd) and Shrimp Alkaline Phosphatase (Affymetrix, Inc.).

\textit{Sequencing}

Sequencing of the PCR products was performed using both the forward and reverse primers for PCR amplification and the BigDye Terminator 1.1 kit (Thermo Fisher Scientific Inc., Waltham, MA, USA). After purification, the cycle sequencing products were detected using an Applied Biosystems 3130 Genetic Analyser with the Data Collection and Sequencing Analysis software (Thermo Fisher Scientific Inc., Waltham, MA, USA).

\textit{Data analysis}

Sequences were assembled and evaluated with the SeqScape v2.6 software (Thermo Fisher Scientific Inc., Waltham, MA, USA). Consensus sequences were compared to those deposited in the Genbank and BOLD systems V3 databases (Benson et al. 2013, Ratnasingham and Hebert 2007) using the BLAST tool (Boratyn et al. 2013) and the BOLD identification engine, respectively.

Employing the most similar sequences found in the databases, phylogenetic trees were constructed with the software MEGA X (Kumar et al. 2018) using the Maximum Likelihood method based on the Kimura 2-parameter model with bootstrap support values calculated by
running 1000 replicates. BLAST was also used to construct molecular phylogenetic trees using the Neighbor-Joining method through the National Center for Biotechnology (NCBI) resources after computation of a pairwise alignment between a query and the database sequences searched on Genbank.

In order to identify the unknown material, we took into account Genbank and BOLD systems V3 similarity results, the representativeness of the plant groups in the databases and, principally, the availability of phylogenetic studies performed previously for the plant groups in focus. In this case, associated taxa consistently clustered together in molecular phylogenetic analysis were considered to corroborate definitive conclusions.

**Results and Discussion**

*Thujone* was detected in the 2 samples analyzed (Figures 2 and 3). In addition, some mono and diterpenes and hydrocarbons common to several plant species were detected. *Thujone* is a chemical compound naturally encountered in many plant species, notably *Artemisia absinthium* (Pelkonen et al. 2013).

Genetic analysis yielded 30 good quality sequences from the 10 samples throughout the five regions analyzed. All sequences resulted in high levels of genetic similarity once searched in databases (Table 1).

Four samples had genetic similarity above the 99% threshold to genera *Sambucus* spp., *Clematis* spp., *Elymus* spp., and *Cynodon* spp. Four samples showed a similarity range between 98 and 100% to the species *Artemisia absinthium*. One sample showed conflicting results, depending on the genetic marker analyzed, possibly reflecting cross-contamination between distinct fragments. One sample reached highest genetic similarity to genera *Artemisia* spp., above the 99% threshold, possibly also belonging to the second group of four samples (*Artemisia absinthium*).
Based on the representativeness of the groups in the consulted databases and the existence of previous phylogenetic studies, we were able to construct phylogenetic trees and to reach conclusions only about the identity of the second group of four samples (Figure 4)\(^1\)\(^2\).

The results showed that chemical and genetic analysis data were convergent regarding the nature of the questioned vegetable matter. The data gathered from genetic markers strongly suggested association of four samples (out of ten) to the species *Artemisia absinthium*. In fact, ITS2 is an ideal barcode to infer the phylogenetic relationships in *Artemisia* spp. (Rashmi et al. 2016), corroborating this inference about overall nature of the material analyzed: *Artemisia absinthium*.

A primary limitation encountered was the real possibility of contamination between distinct fragments as one sample showed conflicting results, possibly reflecting cross-contamination between distinct fragments. Care should be taken in a forensic laboratory during sampling and extraction procedures, particularly when distinct cases are processed concomitantly. Usage of analytical controls and redundant sampling, amongst others, are critical providences to secure reliable results. Policies, procedures, and records are critical, as are appropriate approaches for dealing with non-human DNA. *Cymbopogon citratus* was useful as a positive control because it was easily amplified and sequenced for the markers/primers used (except for *ycf*1) and broadly available in Brazil.

The regions and primers used in this study are a versatile set for forensic investigations. While the *ycf*1 region was not immediately useful for identification due to low coverage of this region in the Genbank and BOLD databases, it was primarily used for the generation of new

---

1 Table S1 shows GenBank accessions for each sequence used for the molecular phylogenetic analyses displayed in Figure 4.
2 Figure S1 shows molecular phylogenetic trees constructed using the National Center for Biotechnology (NCBI) resources after BLAST search on Genbank.
sequence data from the questioned material; this will eventually be useful in the future for the
laboratory. A region that was not used in this study but that could be useful for later
investigations is the trnH-psbA region (Pang et al. 2012), but it has not yet been standardized in
the laboratory.

Although there were high levels of sequence similarities compared with species from the
genera **Sambucus** spp., **Clematis** spp., **Elymus** spp., and **Cynodon** spp., the low representation of
these plant groups in the databases and the lack of intensive phylogenetic studies performed
meant we could not reach confident conclusions about the identity of these samples. Our data,
however, are sufficient to ascertain that the material did not contain proscribed drugs or
protected plants: in this case, **Artemisia absinthium**, **Sambucus** spp., **Elymus** spp., **Clematis** spp.
and **Cynodon** spp. are probably associated with folk medicine practices.

Accurate DNA barcoding identifications in forensics are performed when there are
available studies confirming DNA regions that can segregate specific species from a group of
closely related ones. It is noteworthy that most of the difficulties in plant DNA identifications,
such as the present case and other cases from the laboratory’s history, could be partially due to a
shortage of available information in databases. It is a weakness, however, that can be
circumvented: efforts must be encouraged to enrich data engendered by barcoding projects. This
is particularly critical in a country with extraordinary biodiversity, such as Brazil.

On the other hand, however, anatomy and morphology cannot always indicate the source
of a plant part, particularly for degraded or fragmented material. In addition, chemical analysis
often does not identify the specific plant, even if it detects a particular substance. Such
limitations can rarely be overcome, which demonstrates the importance of the association
between forensic chemistry and forensic botany, enriched with DNA barcoding methods. Plant
DNA barcoding has, furthermore, many forensic applications. Not only to identify the
questioned material, but also to make inferences of geographic origin through possible genetic signatures (Mello et al. 2016), greatly aiding investigations of different natures, from illicit drugs to environmental forensics.

Achieving greater resolution of genetic data will be greatly expanded with the recent acquisition of a new generation sequencing platform by the forensic genetics laboratory of the Brazilian Federal Police. DNA barcoding technology can be improved and metabarcoding technology implemented. Metabarcoding is application of high-throughput sequencing (HTS) platforms for parallel acquisition of DNA barcode sequences from many specimens simultaneously (Shokralla et al. 2014). DNA metabarcoding also enables simultaneous high-throughput multi-taxa identification using DNA extracted from complex samples containing DNA of different origins (Raclariu et al. 2018).

A forensic laboratory faces many challenges. The first is the urgency of the expert examination requests. The high number of requisitions, leading to competing demands, and rigorous policies, procedures, and rule enforcement to comply with judiciary system requests are additional challenges. Finally, the exams themselves are time-consuming. This presents a challenge and demands a distinct strategy and administrative efforts. Notwithstanding, there are some advantages: a forensic DNA laboratory is frequently associated with other forensic labs, such as a chemistry lab. This brings enormous advantages to an interdisciplinary association.

Many studies address the particularity of DNA barcoding as an accessory tool in the identification of plants (Pawar et al. 2017; Palhares et al 2015; International Organization for Standardization [ISO] 2017; Sgamma et al. 2017). However, to the best of our knowledge, few studies report an effective joint effort of forensic chemistry and forensic botany techniques and the problem of its practical forensic application or describe a real case solved in a forensic laboratory with its difficulties and limitations.
Conclusion

The analytical methods used by forensic chemistry are sensitive and effective, but they can be limited in situations involving unknown plant material or a portion of uncharacteristic plant parts. Forensic botany, with the modern technique of DNA barcoding, has established itself as a powerful tool to complement the scope of forensic chemistry. Situations involving plant materials may not be trivial and law enforcement agencies should be aware of the feasibility of forensic exams and the need for a multidisciplinary approach. We presented the details of a successful case using this approach. At this degree of complexity to the best of our knowledge, this is the first study to report an effective joint effort of forensic botany and forensic chemistry techniques in order to assess the real nature of a suspected plant material in Brazil. Reiterated questions from prosecution led us to utilize DNA barcoding methods, which clarified the nature of the questioned vegetable matter sent to the laboratory.

Considering all the advantages and inherent limitations of spectroscopic methods and DNA barcoding, the combination of the two techniques could represent a synergistic effort for comprehensive assessment of vegetable samples whenever necessary for elucidating non-trivial forensic cases. The availability of a new technological approach for the genetic sequencing of plant species will broaden the horizons for many forensic investigations. The trajectory of the Forensic Genetics Laboratory of the Brazilian Federal Police may also inspire similar initiatives.

Acknowledgements

We would like to thank all levels of the Brazilian Federal Police administration for their support of this work.
References


trnH–psbA Intergenic Spacer Region and Its Combinations as Plant DNA Barcodes: A

Authenticity of Herbal Dietary Supplements: Comparison of Chemical and DNA

Pelkonen, O., Abass, K., and Wiesner, J. 2013. Thujone and thujone-containing herbal

Raclariu A.C., Heinrich M., Ichim M.C., and de Boer, H. 2018. Benefits and Limitations of

Rashmi T.R., Murali, S., and Francis M.S. 2016. ITS2 is an ideal barcode to infer the

Ratnasingham, S. and Hebert, P.D. 2007. BOLD: The Barcode of Life Data System
8286.2006.01678.x.

Shokralla, S., Gibson, J.F., Nikbakht, H., Janzen, D.H., Hallwachs, W., and Hajibabaei, M.
2014. Next-generation DNA barcoding: using next-generation sequencing to enhance

https://mc06.manuscriptcentral.com/genome-pubs


Figure 1: Example of four samples (out of ten), which are probably associated to the genus *Elymus* spp. (2061Q9), *Sambucus* spp. (2061Q10), *Cynodon* spp. (2061Q11) and, certainly, to the species *Artemisia absinthium* (2061Q12).

Figure 2: Gas chromatographic/mass spectrometry chromatogram illustrating the thujone.

Figure 3: Electron ionization mass spectrum of thujone.

Figure 4: Molecular phylogenetic analysis by Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.
Table 1. A summary of sequence quality - highest genetic similarity (numbers as percentages) found when sequences obtained were consulted in databases; number of ambiguous base calls for each sequence (absolute values in parentheses).

<table>
<thead>
<tr>
<th>Marker</th>
<th>Highest genetic similarity (numbers as percentages)</th>
<th>Number of ambiguous base calls for each sequence (absolute values in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>matK</td>
<td>GenBank: Sambucus spp. 100% (0) Sambucus spp. 100% (0) not available</td>
</tr>
<tr>
<td></td>
<td>rbcL</td>
<td>Bold: Artemisia absinthium 98% (8)</td>
</tr>
<tr>
<td></td>
<td>ITS</td>
<td>GenBank: Artemisia absinthium 98% (7) Canadeweissia brevicaudata 98% (9)</td>
</tr>
<tr>
<td></td>
<td>trnL-trnF</td>
<td>GenBank: Artemisia absinthium 98% (7) Canadeweissia brevicaudata 98% (9)</td>
</tr>
<tr>
<td></td>
<td>ycf</td>
<td>GenBank: Artemisia absinthium 98% (7) Canadeweissia brevicaudata 98% (9)</td>
</tr>
</tbody>
</table>

Observations regarding the sample 2061Q01:
The conflicting results (matK and ITS markers) point to the simultaneous presence of particles of different nature materials. It is probably Sambucus spp. contaminated by Artemisia absinthium, which can be explained by the close physical contact of plant fragments of distinct natures packed together as a single material. In addition, it is not possible to exclude altogether what similar phenomena may have occurred for other samples. In these other hypothetical cases, there is a preferential amplification of one species over another.

Observations regarding the samples 2061Q04 and 2061Q05:
* Total scores: Artemisia absinthium (Accession #JX051763.1 GenBank) > Artemisia absinthium (Accession #JX051763.1 GenBank) > Artemisia absinthium (Accession #EF577289.1, GenBank) > Artemisia absinthium (Accession #EF577289.1, GenBank).

Phylogenetic analyses also supported the association of these two samples (as well as samples 2061Q03 and 2061Q12) to the species Artemisia absinthium.

Observations regarding sample 2061Q07 (negative control): presented the expected results (there was no previous amplification).

Observations regarding sample 2061Q08 (positive control): presented the expected results (it is Cymbopogon citratus).
Example of four samples (out of ten), which are probably associated to the genus Elymus spp. (2061Q9), Sambucus spp. (2061Q10), Cynodon spp. (2061Q11) and, certainly, to the species Artemisia absinthium (2061Q12).
Aver age 4.053 to 4.083 m/z: CG01_037-19_02_SINT_D_data.m (+)

A b u n d a n c e

https://mc06.manuscriptcentral.com/genome-pubs
Molecular phylogenetic analysis by Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.