Studies on authentication of true source of honey using pollen DNA barcoding

Praghadeesh Manivanan, Srinivasan Madapuji Rajagopalan and Mohankumar Subbarayalu

Abstract

Plant pollen DNA comprising unique barcode signatures serve as a highly resilient biomarker to establish its true biological origin. Over a decade, DNA barcoding of plants has evolved as an effective tool to identify its origin down to species level that promptedit practical applications in various fields of applied biology. Premium unifloral honey, due to its economic value as a medicinal product is vulnerable to intentional mislabelling and adulteration to fetch higher price. Melissopalynological and physiochemical based approaches to ascertain the true source of honey is not yet available on commercial scale. DNA barcoding of pollen trapped in honey can be effectively used to identify the true source of honey with high feasibility across labs at cheaper price. The current study was carried out to standardise an efficient pollen DNA isolation protocol and to assess the efficiency of two barcode (rbcL and matK gene) combinations in identifying the botanical origin of pollen and to ascertain its true source. Six honey samples with diverse characteristics were used in the study. Pollen DNA isolated using modified CTAB solution resulted in quantitatively appreciable yield with less Polymerase Chain Reaction inhibitors. Further, the two barcode combinations were found to be effective in identifying the true botanical origin of honey samples with 96 to 98 percent confidence. We report here a unique methodology to authenticate true honey source by harnessing the power of both efficient DNA isolation strategy and pollen DNA barcoding.

Keywords: Pollen embedded honey, authenticity, pollen DNA barcoding, barcode based palynology

1. Introduction

Many agricultural and horticultural crops rely heavily on pollinators, especially honeybees to augment pollination and thereby increasing the total yield of crops [1]. Besides pollination services, honeybees can be successfully domesticated to produce honey, bee wax, propolis and royal jelly which are all proven to be beneficial and contribute significantly to human health one way or the other. Honey is considered as a natural product with a uniqueness in state of viscosity, fluidity and taste. During the process of foraging for nectar and pollen, honeybees effectively accumulates pollen in honey as an embedded matrix thereby making pollen embedded honey as a potential biomarker to establish the plant source [2]. Honey can be classified as unifloral, when arising predominantly from a single botanical origin (represents more than 45% of the total pollen content) and multifloral, when it is sourced from a mixture of flower from different plant species. Unifloral honey is highly valued in food and pharmaceutical industry and is generally deemed as premium product. Therefore, unifloral honey is vulnerable to intentional mislabelling or adulteration (with common honey, beet sugar syrup, rice syrup and corn sugar syrup etc.) to fetch premium price. Strayer and co-workers [3], termed the food products that are adulterated for financial advantage as Economically-Motivated Adulteration (EMA). Traditionally, the true source of honey has been deduced by its taste, flavour, aroma and colour [4, 5]. Whereas, the absolute quantification of true botanical origin was based primarily on four major innate properties of honey viz., physical, physicochemical, sensory, and biochemical properties of honey that are totally independent on its pollen constituents. Gomez-Diaz and co-workers [6], relied on mere physical properties of honey viz., refractive index, density, viscosity, water content, water activity, pH, electrical conductivity sugars and ash content to ascertain the authenticity of honey. Physicochemical property based evaluation of honey using methodologies like Nuclear Magnetic Resonance (NMR) spectroscopy [2], Liquid Chromatography coupled with mass spectroscopy (LC/MS), High Pressure Liquid
chromatography (HPLC) [7] were also reported. Sensory characteristics of honey viz., phenolic content, antioxidant activity and colour was measured to evaluate the source of honey and were relatively successful with varying degree of accuracy [5]. Studies based on biochemical properties (carbohydrate, proteins, minerals and volatile compounds) were also reported as suitable biomarker for identification of true honey source [8-10]. Despite the applicability of conventional methods, they are purely based on various properties of honey and not on the basis of true pollen content. In spite of the advancement of physio-chemical properties based analytical methods for authenticating honey source, the key niche area ‘pollen embedded in honey’ is largely less explored as a commercial tool to deduce botanical origin of honey. Further, these methods require reference samples to compare with test samples which compounds the difficulty in utilising them as viable approach [11-14]. Pollen based detection approach to understand the botanical origin is undoubtedly robust and viable approach as compared to that of physio-chemical composition based approaches that includes cumbersome establishment of reference samples for which the authenticity has to be verified [15]. Further physio-chemical based methods proved to be ineffective due to inherent variations in honey production by bees which is influenced by climatic condition, floral fauna availability and health of the colony [16].

Palynology is the study of pollen grains and other spores, especially as found in archaeological or geological deposits. Melissopalynology is the study of pollen found in honey and thereby ascertaining the origin on honey. Traditionally, microscopic analysis was the only tool to identify and classify the pollens based on morphological cues and in turn discriminate the plants visited by the pollinators during honey production. Morphological identification of pollen has provided genus level and in some cases species level discrimination of the source plant [17]. However, microscopic methods warrant skilled personal, expertise and often involve high cost and more time to analyse samples [18].

DNA barcoding is an emerging tool in molecular biology for rapid species recognition and identification based on short DNA sequences. The core idea of DNA barcoding is based on the concept that short fragments of DNA that can be found within the species with minor degree of variation and much greater variation between species [19-21]. Early DNA barcoding studies employed the trnL-UAA intron marker that failed to provide resolution at species level in most test samples [22]. With the pursuit to harness the power of pollen embedded honey as a potential biomarker for assessment of honey quality, DNA barcoding – a gold standard method was employed identify the botanical origin [23]. In the present study, an effective DNA isolation protocol from pollen embedded in honey was optimised and DNA barcodes were generated using universally accepted organelle barcodes rbcL (ribulose 1-5 bisphosphate) and matK (maturase K) to assess the feasibility of using them as a potential tool to ascertain the authenticity of unifloral honey.

2. Materials and Methods

2.1 Sample Source

The study was carried out on six different honey sources, two representing multifloral origin, two representing unifloral origin and two commercial packed products. The multifloral honey samples (H01 and H02) were sourced from Indian bee hives (Apis cerana cerana Fabricius) and Italian bee hives (Apis mellifera Linnaeus) from department of Agricultural Entomology, Tamil Nadu Agricultural University, Coimbatore that were intended for packing and marketing. Two unifloral honey samples were sourced from the two different boxes of Indian bee hives (Apis cerana cerana Fabricius) from the Insectary, Department of Agricultural Entomology, Tamil Nadu Agricultural University, Coimbatore. The unifloral honey samples were collected from hives placed in sunflower (H03: Helianthus annuus L.) and pumpkin (H04: Cucurbita pepo var pepo) fields and fully ripened, sealed honey were extracted from combs in the field using TNAU honey extractor after removing the wax cap. Two commercial honey samples inducted in the study were procured from retail outlets (H05 and H06 (sourced from hilly regions of Tamil Nadu)). The six samples were neither subjected to filtration nor post extraction treatments and were used as such for the experiments (Fig.1). Representative samples used for the experiments were vouchedered and preserved at room temperature in Molecular Ecology lab, Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India.

![Fig 1: Collection of honey from hives placed in field: Honey samples used in the study.](image)

2.3 Pollen Extraction from honey samples

Pollen extraction from honey was performed based on the protocol by Waiblinger and co-workers [24] with minor modifications. Forty grams of each honey samples were weighed (10 g) and distributed equally to four sterile 50 ml oakridge tubes and volume was made up to 40 ml with ultrapure autoclaved water. Tubes were subjected to vigorous vortexing for 2 min and incubated at 40°C for 10 min to negate the high viscous nature of honey. The tubes were centrifuged at 8000 rpm for 15 min and the supernatants were discarded. The pellets from the four tubes were pooled into one tube and further re-suspended to a final volume of 40 ml using ultrapure autoclaved sterile water. The tubes were again centrifuged at 8,000 rpm for 15 min and the resultant pellet which contains extracted pollens were transferred to sterile 2 ml eppendorf tubes for subsequent DNA extraction. All the experiments were carried out in triplicates.

2.4 DNA Isolation, quality analysis and quantification

Genomic DNA isolation was optimized using two extraction buffers viz., DNAzol® Reagent, Thermo Fischer Scientific and modified CTAB extraction buffer (125 mM Tris-HCL, 50mM EDTA, 50 mM NaCL, 3.5 percent CTAB, pH 7.8, 4.0 percent β-mercaptoethanol). Pollen tissues with extraction buffer 1.5 ml were subjected to bead mill lysis at 30 hertz for 8 min using Qiagen, Tissue Lyser II (Germany). Isolation procedure for the treatments with DNAzol® reagent were
followed as per manufacturer’s protocol. After tissue lysing, tubes with 1.5 ml of cetyltrimethyl ammonium bromide (CTAB) extraction buffer were incubated at 65 °C for 20 min in water bath. The tubes were removed from the water bath and allowed to cool at room temperature and centrifuged at 12,000 rpm for 15 min. Supernatant was transferred in to fresh eppendorf tube (1.5ml) and equal volume of Chloroform : isoamyl alcohol mixture (24:1, v/v) was added and mixed by inversion for one min. It was centrifuged at 12,000 rpm for 10 min and the clear top aqueous phase was transferred to a new sterile tube. Ice-cold isopropanol (0.7 percent of reaction volume) was added and mixed gently by inversion and it was stored at -20 °C for overnight. After incubation, the tubes were centrifuged at 12,000 rpm for 10 min to pellet down the DNA and the supernatant was discarded. The DNA pellet was washed with 70 percent ethanol, air dried and rehydrated in 30 µl of nuclease free water (ThermoFischer Scientific, USA) and stored at -20 °C for further experimentation.

Quality of genomic DNA was checked by 0.8 percent (0.8 g in 100 ml) agarose prepared in 1X TBE buffer. Two µl of DNA was loaded with 2 µl of 6X loading dye. Electrophoresis was carried out at 80 V for 45 min, the gel was visualized on UV trans illuminator (Bio-Rad, USA) and documented in gel documentation system (GELSTAN 1312, Medicare, India).

Quantification of DNA was performed using Nanodrop Spectrophotometer (ND-1000, Thermo scientific, USA).

PCR ampiclon sequencing and bioinformatics analysis for pollen DNA barcoding

Pollens were DNA barcoded using two sets of plant barcodes (Ribulose-1,5-bisphosphate carboxylase (rbcL) and matK nuclease kinase(matK)), suggested by International Barcode of Life (IBOL). The two standard barcode region were amplified using polymerase chain reaction (PCR) with the two sets of primers in Sure cycler 8800 (Agilent Technologies, USA). PCR was performed with a total reaction volume of 25 µl composed of 0.2 µM of each deoxynucleotide triphosphate (dNTP), 0.2 µM of each primer (Table 1), 1X Taq buffer containing 1.5mM MgCl2 and 1U Platinum Taq DNA polymerase High Fidelity (Invitrogen, USA). PCR products (3 µl) were resolved 1.5 percent agarose gel to confirm the amplicon size. Remaining PCR products were purified using Pure Link PCR purification kit (Thermo Fischer Scientific, USA). The sequencing PCR were set by using Big Dye Terminator V3.1 Cycle Sequencing kit (Thermo Fischer Scientific, USA). The samples were sequenced on ABI 3730XL DNA Analyzer (outsourced from Scigenom Labs Pvt. Ltd., Kerala, India). The forward and reverse sequences were aligned using Clustal W pairwise alignment using Geneious v. 11.0.2 [28-30]. Processed nucleotide sequence information were submitted to Gen Bank (NCBI) [31] and can be retrieved using their accession numbers. Nucleotide similarity searches were performed by BLAST tool at NCBI database [32, 33]. Similarity identity was used to assign the identity to each query sequence and blast result (with high identity percentage, high query coverage, and E-value < cutoff) of the query sequence was used to associate with the authenticity of the source of honey samples. Concurrently, the probability of species identification was also assessed using BOLD (Barcode of Life) reference database [34] powered by Barcode of Life Data Systems Identification Engine (BOLD-IDS).

3. Results and Discussion

3.1 Extraction of pollen from honey

Honey can be regarded as a unique and complex matrix owing to their 80 percent sugar composition and their inherent nature of preserving the pollen content intact for longer period of time. Melissopalynology based approach though proven successful, the requisite of expert knowledge on different pollen morphology limits it from wider application [35]. Whereas, DNA based markers (DNA Barcodes) present in pollen containing unique signatures to identify the botanical species is a novel and promising approach. The use of DNA based methods offer advantages in terms of rapidity, sensitivity and specificity. It is also suitable for high throughput protocols as an effective alternative to traditional melissopalynological analysis [36, 37].

Ascertaining the correct botanical origin of honey warrants effective separation of pollen embedded in honey followed by DNA extraction with minimal adjoining compounds, and thereby resulting in low interference during PCR. Honey samples sourced in this study exhibited distinct variation in the physical properties like colour and viscosity (Fig. 1). All the samples H01, H02, H03, H04, H06 except H05 resulted in appreciable amount of pollen sediments (on an average ~approx 60 mg per sample). The sample H05 resulted in relatively less amount of sediments after extraction.

3.2 DNA Isolation from pollen sediments embedded in honey matrix

Optimization of DNA isolation protocol for pollen embedded in such a complex matrix honey, is a vital prerequisite step towards ascertaining the source of pollen viz-a-vis honey. Very few literatures were available regarding the effective high pollen DNA yielding protocol with respect to pollen embedded in honey [22, 37-42]. Among the two extraction buffers used for the study, modified CTAB extraction buffer resulted in comparatively higher yield and purity of intact DNA compared to that of DNAzol® reagent protocol. The DNA fragment visualized on 0.8 percent agarose gel demonstrated that, the DNA isolated by DNAzol® reagent had less than the detectable amount when compared to distinct bands of DNA isolated by modified CTAB method. Similarly, based on the nanodrop spectrophotometer readings, the quantity (ng/µl) of DNA isolated by DNAzol® protocol ranged between 21.6±2.0 to 30.9±4.2 and DNA purity values expressed as ratio of absorbance A260/A280 nm ranged from 1.2±0.2 to 1.4±0.3. Whereas, DNA isolated using modified CTAB protocol resulted in comparatively better yield in terms of quantity (40.0±4.2 to 82.3±4.0) and DNA purity values ranged between 1.7±0.1 to 1.9±0.3 (Table 2). Our results demonstrate that, the modified CTAB procedure is significantly better in terms of quality of DNA and purity of DNA which is in agreement with the similar comparison study by Laha and Co-workers [43]. Interestingly, the commercial honey sample (H05), did not yield quantifiable amount of DNA which is evident from both 0.8 percent agarose gel and nanodrop spectrophotometer readings. Upon pollen extraction from honey (H05), relatively low pollen sediments were found and hence, it is safe to speculate that, reason being attributed to their post processing or filtration process that might resulted in absence of pollen content in their samples.

3.3 Pollen DNA Barcoding

Over a decade, DNA barcoding has gained popularity among the scientific community and proved highly effective as a yardstick in identifying any biological source of plants down to species level [44, 45]. Discrimination of honey source based on meta-barcoding approach using ITS2 gene were also
attempted and found to be superior and accurate when compared to that of the melissopalynology based approach [46]. Nevertheless, the genes used for barcoding were neither capable of identifying plant source to species level with high discriminating power nor widely accepted choice of marker by DNA barcoding community.

In the present study, we employed DNA barcode (rbcL+ matK) combination that were unanimously acknowledged by Consortium of Barcode of Life (CBOL) [47], to identify the plant species from which the honey is sourced. Invariably, in five honey samples the PCR analysis using both Ribulose-1,5-bisphosphate carboxylase (rbcL) and maturase kinase (matK) barcode genes specific primers amplified a product size of 600 and ~840bp respectively. Honey sample (H05) did not produce any amplicon for both rbcL and matK genes. Irrespective of the DNA isolation methods employing CTAB extraction buffer and DNAzol®, PCR amplification for both the genes across the samples other than H05 sample were evident (Fig. 2 and Fig. 3). The honey samples utilized in the current study are relatively pollen rich and hence both DNA isolation strategies resulted in appreciable PCR yield. Therefore, the choice of extraction procedure with comparative advantage of one over the other, will have greater impact when the experiment involves complex initial pollen material [48].

![Agarose gel electrophoresis of PCR products of rbcL gene amplified from pollen embedded honey samples by two methodologies](image1)

![Agarose gel electrophoresis of PCR products of matK gene amplified from pollen embedded honey samples by two methodologies](image2)

### 3.4 Bioinformatic analysis to assess the discriminatory power of DNA barcodes

The raw sequence reads from sanger sequencing in both forward and reverse of rbcL and matK genes were subjected to Clustal W pairwise alignment and primer sequences were trimming which resulted in 580 bp and 800 bp (base pairs) of final partial gene sequences. Processed ampiclon sequences exhibited matches exactly to rbcL and matK genes in all the amplified samples. The processed consensus sequences were submitted in Gen Bank, NCBI and accession numbers were assigned by NCBI. The blast results of resultant nucleotide sequences were represented (Table 3). Based on analysis on individual barcode sequences (five honey sources), it showed a high correlation between the source of collection and the query results. Source assessment by BLAST analysis of partial rbcL and matK gene ampiclon of multifloral honey samples (H01 and H02) indicated that, cotton (Gossypium hirsutum) and ash gourd (Benincasa hispida) pollens were predominant components of the multifloral honey and per se the rich source of nectar and pollen among the experimental plots in Tamil Nadu Agricultural University. The results of the current study are based on sanger sequencing of PCR product, and hence enables one to identify the most abundant pollen species found in polyfloral honey samples and thereby correlating the source to product to certain extent. In order to enumerate the entire pollen content in polyfloral honey, cloning of the PCR products followed by sequencing [39] or metabarcoding [48, 49] approach might prove useful in enumerating all the available pollens in multifloral honey samples. Albeit their benefits, the cost associated with the later approaches and time involved in analysis would undermine their utility as rapid and cost effective approaches. Similarly, source assessment of unifloral honey samples (H03-H04) collected from hives placed in sunflower (Helianthus annuus) and pumpkin (Cucurbita pepo) fields using BLAST analysis confirmed their origin with greater 96 percent identify and are in agreement with sample source. In addition, BLAST analysis of the commercial honey sample H06, labelled as sourced from hilly regions of Tamil Nadu, resulted that the eucalyptus (Eucalyptus obliqua) formed the major pollen composition of the honey. These results, are in agreement with the study by Prosser and Hebert [41], who demonstrated that utilising the DNA barcode combination rbcL+ matK is ideal choice to identify the plant species to species level with high efficiency and accuracy.
Table 1: List of primers, PCR conditions and amplicon descriptors

<table>
<thead>
<tr>
<th>Gene</th>
<th>Notation</th>
<th>Sequence (5’ – 3’)</th>
<th>Annealing temp.</th>
<th>Amplicon Length</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ribulose-1,5-bisphosphate carboxylase (rbcL)</td>
<td>rbcL-F</td>
<td>ATGTACACCAAAACAGAGACTAAAGGC</td>
<td>57 °C for 45 sec</td>
<td>600 bp</td>
<td>[25, 26]</td>
</tr>
<tr>
<td></td>
<td>rbcL-R</td>
<td>GTAAAAATCGTGCCACCRCCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maturase kinase (matK)</td>
<td>3F_KIM (F)</td>
<td>CGTACGAGTTTTTTGTTTACGAG</td>
<td>53 °C for 40 sec</td>
<td>840 bp</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>IR_KIM (R)</td>
<td>ACCCGATCCATCTGGAAATCTGGTTC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Profile of quality and purity of DNA isolated from six honey samples

<table>
<thead>
<tr>
<th>Extraction method</th>
<th>(Multifloral / Indian Bee)</th>
<th>(Multifloral / Italian Bee)</th>
<th>(Unifloral : Helianthus annuus)</th>
<th>(Unifloral : Cucurbita pepo)</th>
<th>(Commercial source)</th>
<th>(Commercial source)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conc.*</td>
<td>Purity</td>
<td>Conc.*</td>
<td>Purity</td>
<td>Conc.*</td>
<td>Purity</td>
</tr>
<tr>
<td>DNAzol® Reagent</td>
<td>28.1±2.11</td>
<td>1.2±0.4</td>
<td>22.7±1.9</td>
<td>1.4±0.3</td>
<td>29.7±3.1</td>
<td>1.4±0.2</td>
</tr>
<tr>
<td>Modified CTAB reagent</td>
<td>72.1±3.6</td>
<td>1.9±0.3</td>
<td>68.4±3.9</td>
<td>1.8±0.1</td>
<td>82.3±4.0</td>
<td>1.7±0.2</td>
</tr>
</tbody>
</table>

*Concentration expressed as ng/µl and purity values expressed as ratio of absorbance A260/A280 nm.

**The results are the mean and standard deviation value of two extraction methods made in triplicate assays.

Table 3: Percent identity and source authentication of pollen embedded in honey based on nucleotide BLAST analysis

<table>
<thead>
<tr>
<th>Sample notation</th>
<th>Sample descriptor</th>
<th>Barcode</th>
<th>Assigned accession number*</th>
<th>Sequence comparison using BLAST analysis</th>
<th>Source authenticated based on top hit result / Barcode match</th>
</tr>
</thead>
<tbody>
<tr>
<td>H01</td>
<td>Multifloral / TNAU Apiary / Indian Bee</td>
<td>rbcL</td>
<td>MF804317</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>matK</td>
<td>MF804322</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>H02</td>
<td>Multifloral / TNAU Apiary / Italian Bee</td>
<td>rbcL</td>
<td>MF804318</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>matK</td>
<td>MF804323</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>H03</td>
<td>Unifloral : Hive managed in sunflower field</td>
<td>rbcL</td>
<td>MF804319</td>
<td>99</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>matK</td>
<td>MF804324</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>H04</td>
<td>Unifloral : Hive managed in pumpkin field</td>
<td>rbcL</td>
<td>MF804320</td>
<td>99</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>matK</td>
<td>MF804325</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>H05</td>
<td>Commercial source</td>
<td>rbcL</td>
<td>NA</td>
<td>Zero pollen content upon extraction</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>matK</td>
<td>NA</td>
<td>Zero pollen content upon extraction</td>
<td></td>
</tr>
<tr>
<td>H06</td>
<td>Commercial source</td>
<td>rbcL</td>
<td>MF804321</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>matK</td>
<td>MF804326</td>
<td>100</td>
<td>96</td>
</tr>
</tbody>
</table>

*Sequences submitted to GenBank (NCBI) originating from the present study

4. Conclusion
Various health benefits of honey to human beings and associated attractive fitness-based diet role had prompted a considerable demand for authentic and specialty honeys in the agro-markets. As a reflection of surge in price and demand, there is huge possibility of adulterated / intentional mislabelled honey finding its way in the market, and thereby affecting the end use consumers. It is therefore crucial to formulate an efficient methodology to assess the authenticity of honey in a short time span and low cost. The results of the study would complement the current pollen DNA isolation protocol and corroborated the fact that, DNA barcoding can be effectively used to ascertain the source of honey. In future, studies on multiplexing of barcodes will enable to reduce the time of analysis and new mechanisms to barcode samples with multiple pollen content is warranted to further our understanding and to effectively establish a benchmark for honey quality in agro-industry and to protect consumers’ interest.

5. Acknowledgements
This work has been supported by grants from Department of Biotechnology, India (DBT). The authors are also grateful for to Mr. Babu and Mr. Kumar, TNAU, Bee Keepers towards handling the hives and sample collection.

6. References


