Investigation of the Potential Mislabeling on Spice Products of the Lamiaceae Family

Yoel Dusi,[1] Amana Malik,[2] Peilin Zhen[3], Marianne Williams[4]
[1] Schechter School of Long Island; [2] Queens Gateway to Health Sciences Secondary School;
[3] Queens High School for the Sciences at York College; [4] Department of Biological Sciences and Geology, Queensborough Community College, CUNY

ABSTRACT

Our research team tested 17 samples of spices from five genera of the Lamiaceae family from several different brands to determine the efficacy of the rbcL and matK primers and the accuracy in the labeling of products. We examined these spices for possible mislabeling through DNA barcoding, which was similar to Stoeckle et al.'s herbal tea study (2011). We found that rbcL was successful in DNA amplification. However, Brand A's spice samples were problematic, two of which were ambiguous in species identification. Upon review of the literature, we found that 2 products were previously mislabeled by Brand A. The matK primer used bound to the Rosemary and Oregano matK genes to a greater extent than the Sage, Basil, and Thyme matK genes. This result comports with the difficulty CBOL had with selecting matK as an identification barcode (Fazekas et al. 2012).

INTRODUCTION

- The Lamiaceae, or the mint family, one of the most widespread plant is used for cooking, ornamental, and medicinal purposes.

Mislabeling products:

- Previous study on commercial tea products indicates that about one-third of herbal teas generated DNA identifications not found on labels and found it common to have mislabeled products (Stoeckle et al., 2011).

rbcL vs. matK primers:

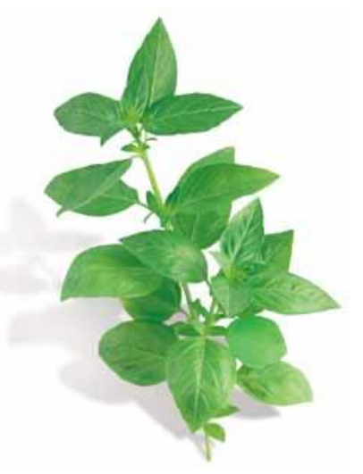
- The matK gene has been shown to identify more phylogenetic groups than the highly conserved chloroplast gene rbcL (Barthet & Hilu 2007).
- The matK gene region is more difficult to amplify than rbcL due to matK's longer base pair and repetitive mononucleotides, which causes DNA degradation that affects the quality of the sequence reaction (Fazekas et al. 2012).
- In this project, we focused on the following:
  1. The potential mislabeling of the samples of dried spices;
  2. The efficacy of rbcL and matK primers in DNA amplification and identification.

SPECIES USED IN THE RESEARCH STUDY


Rosmarinus officinalis
(Rosemary)


Origanum vulgare
(Oregano)


Thymus vulgaris
(Thyme)


Ocimum basilicum
(Basil)


Salvia officinalis
(Sage)

MATERIALS AND METHODS

1. We gathered 17 samples, in both fresh and dry forms, from several different grocery stores in Queens, New York.
2. We extracted DNA using both the UBRP protocol and the Qiagen DNeasy Plant Mini Kit, and performed a polymerase chain reaction (PCR) to amplify rbcL and matK genes.
3. After performing the gel electrophoresis, we sent out the DNA samples that successfully displayed visibility through the UV transilluminator to Genewiz for sequencing.
4. We utilized DNA subway, GenBank, NCBI to do BLASTN searches for sequencing alignment.

RESULTS

Sample Collection and PCR Success

- 32 sequences were sent to Genewiz. 13 out of 16 samples were able to amplify the rbcL genes, and 5 out of 16 were successful for matK gene.

Types of Spices		rbcL Identification	matK Identification
O. basilicum	Brand A	/	--
	Brand B	--	--
	Brand C	×, 38 mismatches	/
	Fresh	√	/
T. vulgaris	Brand A	×, 25 mismatches	/
	Brand B	√	--
	Brand D	√	--
	Fresh	√	N.A.
O. vulgare	Brand A	×, 13 mismatches	/
	Brand B	√	√
	Brand C	√	√
	Fresh	√	√
R. officinalis	Brand D	√	√
S. officinalis	Brand A	--	/
	Brand B	×, 6 mismatches	--
	Brand E	N.A.	--
	Fresh	√	--

Key Terms:
√ amplifiable and correctly identified
× amplifiable but potentially misidentified
-- insufficient data i.e. short sequence, only 1 (forward or reserved) is available
/no sequence

- Only all 4 oregano and 1 rosemary matK genes were amplifiable
- The basil from Brand's A, B, and C did not show any sequence with both rbcL and matK primers.
- Brand A's sequences showed a range from 6 to 38 base pairs mismatches for both Thyme and Oregano.
- Basil and Sage did not show any sequences for Brand A using both primers.

BLASTN's Statistics Result

- The higher the bit score represents the higher correlation between the two sequences' alignment.
- The lower number of mismatches means the better hit it is.
- If the bit score is relatively low and the number of mismatches is high, then more investigation should be conducted to identify the sample.

The following chart represents the alignment between M13matK primer (query) and the matK gene of each species, in the order of basil, thyme, sage, oregano, and rosemary (subject):

Sequences	Alignments	Bit Score
Query: M13matK3F **19-41** Subject: Basil **1291-1269**	`CGTACAGTACTTTTGTGTTTACG` `CGTACAGTACTTTTGTGTTTCCG`	37.4 bits (40)
Query: M13matK3F **19-41** Subject: Thyme **2020-1998**	`CGTACAGTACTTTTGTGTTTACG` `CGTACAGTACTTTTGTGTTTACG`	42.8 bits (46)
Query: M13matK3R **19-44** Subject: Sage **62-26**	`CCCAGTCCATCTGGAAATCTTGGTTC` `CCCTGTCCATGTGGAAATCTTGATTC`	60.8 bits (66)
Query: M13matK3F **1-42** Subject: Oregano**130-89**	`TGTAAAACGACGGCCAGTCGTACAGTACTTTTGTGTTTACGA` `TGTAAAACGACGGCCAGTCGTACAGTACTTTTGTGTTNACGA`	75.0 bits (40)
Query: M13matK3R **4-44** Subject: Rosemary **889-843**	`GAAAC--AGCTATGACACCCAGT-CCATC-TGGAAATCTTGGTTC` `GAAACCNAGNTATGACACCCAGTCCCATCNNGGAANTCTNGGTTC`	35.6 bits (38)

Key Terms:
🟨 color represents that there is a gap within the sequence.
🟦 color represents the mismatches between the two sequences.

- The matK primer we used for the DNA extraction was called the M13matK.
- According to DNA Barcode 101, the bit score represents the quality of the alignment between query and subject. The higher the score is, the better the alignment.
- The Oregano matK gene has the highest score among the alignment (75.0 bits), which explains the reason why all four Oregano sequences were successfully amplified.
- Although Rosemary does not have a high bit score as oregano and display a relatively great amount of mismatches, its nucleotides' alignment length is longer than most samples' alignment lengths and results in the positive outcome of matK gene amplification.

DISCUSSION

- After analyzing our results, we make a tentative finding regarding the accuracy of product labels as well as the efficacy of the rbcL and matK primers.
- We found that the rbcL primer was more successful at amplifying our samples than matK.
- The DNA genome of each of the five species was compared and aligned with either forward or reversed sequence of M13matK.
- The matK primer was able to match up with rosemary and oregano matK genes better than with basil, thyme, and sage matK genes.
- All Brand A samples were either unable to amplify or had a high number of mismatches on the sequences.
- After some research, we also found that brand A has a history of mislabeling their products.

Suggestions for Further Research

- Selecting a better matK primer for DNA amplification;
- Examining Brand A's products for a more detailed investigation with regards to its mislabeling history.

REFERENCES

Barthet, M. M., and Hilu, K. W. 2007. Expression of matK: Functional and Evolutionary Implications. American Journal of Botany 94 (8): 1402-1412.

Chan, B., Lau C., Jolivalt, C., Lui, S., Ganem-Elbaz, C., Paris, J., Litaudon, M., Fung, K., Leung, P., and Ip, M. 2011. Chinese Medicinal Herbs Against Antibiotic-Resistant Bacterial Pathogens. Formatex Research Center: 773-781.

Fazekas, A. J., Kuzmina, M. L., Newmaster, S. G., and Hollingsworth, P. M. 2012. DNA Barcoding Methods for Land Plants. DNA Barcodes Methods in Molecular Biology: 223-252.

Sbordoni, V. 2010. Strength and Limitations of DNA Barcode Under the Multi-dimensional Species Perspective. Tools for Identifying Biodiversity: Progress and Problems: 275-280.

Stoeckle, M., Gamble, C., Kiepekar, R., Young, G., Ahmed, S., and Little, D. 2011. Commercial Teas Highlight Plant DNA Barcode Identification Successes and Obstacles. Scientific Reports (1): 42.

ACKNOWLEDGEMENTS

Our research team would like give special thanks to Dr. Williams for lab materials, assistance, and support. We would also like to thank Christine Marizzi, and Melissa Lee for their aid in matK sequences. This research would not be accomplished without UBRP collaborators such as the Harlem DNA Lab , Genspace, and funding from The Pinkerton Foundation and Cold Spring Harbor Laboratory.