Experimental DNA Barcoding Within *Capsicum* and *Solanum* Genera at Species and Variety Levels

Rebecca Kluempen and Kimberly Johnston

Abstract

DNA barcoding is a method commonly used to identify organisms down to a species level. Three different species with eight varieties of Pepper (*Capsicum)* and two different species of Tomato (*Solanum)* were barcoded to determine the level of taxonomic identification possible. Three different universal barcoding primers were used, rbcLa, matK, and ITS, to determine if there were differences in quality of sequencing. Once Sanger Sequencing was performed, the sequences were run through BLAST[1] to obtain an identification. It was found that matK was the most successful for identification, while ITS failed to produce a quality sequence. The *Capsicum* genus could be identified (ID'd) with confidence at the clade level, but there was no conclusion drawn concerning the *Solanum* genus.

Introduction

DNA Barcoding is a method used to identify species or communities of species of Plants, Animals, and Fungi. The method relies on a region of DNA that varies significantly between different species to allow taxonomic differentiation, but that is flanked by regions of DNA that are the same between different species for PCR primers to bind. A primer is a short sequence of DNA that is added to a PCR master mix to indicate which region of the sample DNA is to be amplified.[2]

A key part in the potential success of barcoding is the genetic relationship between species. There needs to be high levels of similarity between members of a species, but differences between separate species. This project looks at three different *Capsicum* species, with eight varieties and two different species of *Solanum* to see if barcoding can detect clade, species, and potentially variety level differences. This is significant, as the evolutionary relationships within the *Capsicum* genus are complicated and highly debated, leading to additional levels of taxonomic identification, including clades.[3]

Three different standard barcoding primers were used which targeted different areas in the specimens' genomes. rbcLa and matK both amplify chloroplast DNA while ITS amplifies nuclear DNA. ITS is traditionally used for barcoding within the fungi kingdom; however, there has been work recently done to develop plant specific ITS primers. These ITS primers were used for this experiment. While this work is still ongoing, it has been found that the use of ITS can potentially result in difficulty when Sanger Sequencing.[4]

There are several areas of focus within this project. The first is to see if all primers perform sufficiently. The second is to determine to what level members of the *Capsicum* genus can be ID'd, be it clade, species, or variety. The third is to see to what level members of the *Solanum* genus can be ID'd, be it species or variety as they do not have clades.

Table 1. Identification of Specimens Obtained.

Specimen ID	Variety	Clade
Capsicum chinense	Habanada	Annuum
Capsicum chinense	Trinidad x Habanada (TPxH)	Annuum
Capsicum chinense	Habanero	Annuum
Capsicum baccatum	Aji	Baccatum
Capsicum baccatum	Mad Hatter	Baccatum
Capsicum annuum	Cubanelle	Annuum
Capsicum annuum	Jalapeno x Cubanelle (JxC)	Annuum
Capsicum annuum	Jalapeno	Annuum
Solanum habrochaites	Neandermato	N/A
Solanum lycopersicum	Slicer Tomato	N/A

Table 2. Description of Primers. F, forward; rev, reverse; ID, identification; R, reverse; rbcL, large unit ribulose bisphosphate carboxylase; ITS, internal transcribed spacer.

Molecular Marker	Primer	Sequence
rbcLa	rbcLaF	**TGTAAAACGACGGCCAGT**ATGTCACCACAAACAGAGACTAAAG
	rbcLa-rev	**CAGGAAACAGCTATGAC**GTAAAATCAAGTCCACCRCG
matK	matK-3F	**TGTAAAACGACGGCCAGT**CGTACAGTACTTTTGTGTTTACGAG
	matK-1R	**CAGGAAACAGCTATGAC**ACCCAGTCCATCTGGAAATCTTGGTTC
Plant-ITS	nrITS2-S2F	**TGTAAAACGACGGCCAGT**ATGCGATACTTGGTGTGAAT
	nrITS2-S3R	**CAGGAAACAGCTATGAC**GACGCTTCTCCAGACTACAAT

Materials and Methods

Materials

Eight specimens of *Capsicum* and two of *Solanum* (Table 1.) were identified and tissue was harvested.

Methods

DNA was extracted from each sample with the Qiagen DNeasy Plant DNA extraction kit with <100mg samples. Three universal barcoding primer sets (Table 2.) were purchased from Millipore Sigma. PCRs were run separately for each primer pair according to recommendations from the DNA learning center.[1] Gel electrophoresis was run for PCR confirmation (Figure 1.) PCR products were sent to Genewiz for Sanger sequencing.

Analysis

The Basic Local Alignment Search Tool (BLAST) was used for each forward sequence and the results were recorded and analyzed (Tables 3, 4.)

Results

Figure 1. Example of Gel Electrophoresis for PCR Verification

Table 3. Analysis of BLAST Identification Results

Identification Level	rbcLa	matK	ITS
Capsicum Clade Success	75 %	100 %	38 %
Capsicum Species Success	38 %	88 %	25 %
Capsicum Variety Success	0 %	0 %	0 %
Solanum Species Success	50 %	50 %	50 %
Solanum Variety Success	0 %	0 %	0 %

Table 4. BLAST Results for Each Specimen. ID, Identification; var, variety; BLAST; Basic Local Alignment Search Tool. BLAST ID's in green indicate a success at the species level.

ID	Primer	*BLAST ID*	% ID	E Value	Top 3 BLAST Options
C. chinense var Habanada	rbcLa	*C. annuum*	98.95	0.0	C. annuum, C. chinense, C. frutescens
	matK	*C. annuum*	100	0.0	C. annuum, C. chinense, C. frutescens
	ITS	*C. annuum*	84.48	1.0e-07	C. annuum, C. chinense, C. frutescens
C. chinense var TPxH	rbcLa	*C. annuum*	99.65	0.0	C. annuum, C. chinense, C. frutescens
	matK	*C. chinense*	99.94	0.0	C. annuum, C. chinense, C. frutescens
	ITS	*C. baccatum*	85.23	2.00e-17	C. baccatum
C. chinense var Habanero	rbcLa	*C. annuum*	99.65	0.0	C. annuum, C. chinense, C. frutescens
	matK	*C. annuum*	99.4	0.0	C. annuum, C. chinense, C. frutescens
	ITS	*Inconclusive*			
C. baccatum var Aji	rbcLa	*C. annuum*	99.83	0.0	C. annuum, C. baccatum, C. pubesens
	matK	*C. baccatum*	99.48	0.0	C. baccatum, C. annuum, C. chinense
	ITS	Inconclusive			
C. Baccatum var Mad Hatter	rbcLa	*C. annuum*	99.83	0.0	C. annuum, C. baccatum, C. pubesens
	matK	*C. baccatum*	99.63	0.0	C. baccatum, C. annuum, C. chinense
	ITS	*C. annuum*	92.11	1.0e-140	C. annuum, C. baccatum, C. eximimium
C. annuum var Cubanelle	rbcLa	*C. annuum*	100	0.0	C. annuum, C. baccatum, C. pubesens
	matK	*C. annuum*	99.91	0.0	C. annuum, C. chinense, C. frutescens
	ITS	Inconclusive			
C. annuum var JxC	rbcLa	*C. annuum*	99.64	0.0	C. annuum, C. chinense, C. frutescens
	matK	*C. annuum*	99.88	0.0	C. annuum, C. chinense, C. frutescens
	ITS	*C. annuum*	78.05	6.0e-140	C. annuum, C. baccatum, C. frutescens
C. annuum var Jalapeno	rbcLa	*C. annuum*	99.83	0.0	C. annuum, C. baccatum, C. chinens
	matK	*C. annuum*	99.88	0.0	C. annuum, C. frutescens, C. pubesens
	ITS	*C. annuum*	99.7	0.0	C. annuum
S. habrochaites	rbcLa	*S. lycopersicum*	100	0.0	S. lycopersicum, S. pimpinellifolium, S. habrochaites
	matK	Inconclusive	99.31	0.0	S. lycopersicum, S. pimpinellifolium, S. habrochaites
	ITS	Inconclusive	80.42	7.0e-23	S. lycopersicum, S. pinelli, S. chilense
S. lycopersicum	rbcLa	*S. lycopersicum*	100	0.0	S. lycopersicum, S. pimpinellifolium, S. pinelli
	matK	*S. lycopersicum*	99.61	0.0	S. lycopersicum, S. pimpinellifolium, S. tuberosum
	ITS	*S. lycopersicum*	94.23	8.0e-117	S. lycopersicum, S. pimpinellifolium, S. pinelli

Discussion

The first concern was if all three primers were sufficient for use when sequencing either pepper or tomato genus. As previously suspected, ITS did not perform sufficiently. The sequences received were not high quality and led to difficulties with species identification including low ID percentages. In terms of sequence quality, both matK and rbcL performed well; however, matK was able to differentiate at a more detailed taxonomic level compared to rbcL (Table 3.).

The second area of investigation was to see to what level members of the *Capsicum* genus could be identified through barcoding. By far, matK provided the best results at both the clade and species level (Table 3.). However, BLAST was unable to identify any of the specimens to a variety level. This may be due to the DNA sequences that are available in the database. There were few accessions previously uploaded that contained variety identification, thus none of the specimens were identified to that level.

The third concern was to see what level the members of the *Solanum* genus could be identified to with barcoding. None of the primers led to sequences that correctly ID's the *S. habrochaites* specimen at either level. This could be due to a deficit in sequences available on BLAST. There has been more work done to sequence cultivated tomatoes than wild types. However, this should be studied further, as the two species are not closely related from a phylogenetic perspective.

When analyzing the BLAST results, it became apparent that based on these results, barcoding is likely not a foolproof or reliable solution for identification for either genus. The differences in identification percentages (Table 3.) were very small, which does not lead to high levels of confidence if an unknown sample required identification. Thus, while matK was the most successful, there is room for more investigations to be directed towards what primers would be best suited for use.

One improvement that could be made for the future would be to optimize the DNA extraction and PCR reaction conditions. The samples used were between 90-100mg; less could be used. As seen in Figure 1. there were copious amounts of DNA present. This led to problems with Sanger Sequencing. In total, twenty of the thirty reactions run had to be diluted and repeated by the Genewiz lab.

As there are many more species of *Capsicum* and *Solanum* that have evolved over time, an appropriate place to further this research would be to sequence more varieties to further examine the feasibility of DNA barcoding for ID in either genus.

Conclusion

It is clear from the barcoding results that the complicated taxonomic relationships between members of the *Capsicum* genera lead to difficulties when barcoding, even just down to the species level. For both the *Solanum* and *Capsicum* genera, there was not enough information available in the BLAST database for variety level identification. Additionally, while matK did properly identify all specimens to the clade level, not all species were correctly ID'd. The *S. habrochaites* specimen was not successfully identified with any of the primer pairs, though matK was the closest.

Overall, the results suggest that matK is the best for sequencing within the *Capsicum* genus. However, no conclusions can be made concerning the *Solanum* genus without further investigation. The specific ITS primers used did not lead to sufficient sequences, thus it may not be the recommended choice.

While a high ID% and low E value indicated a good match, when there are multiple options with extremely close values, it can be difficult to determine the correct and accurate option. Thus, there is further work that could be done in developing primers for barcoding such closely related species. Additionally, there was a lack of variety level identification on the BLAST database, thus no variety level identifications were possible. An improvement would be to populate the BLAST database by sequencing at the varietal level.

Acknowledgements

For generously and kindly providing guidance and mentorship, Kimberly Johnston. For providing financial support, Bristol Myers Squib and the Ornsteen Family Foundation. For providing lab support, Kara Kramer.

References

1. *DNA learning center Barcoding 101*. (n.d.). DNA Learning Center Barcoding 101. https://dnabarcoding101.org/lab/index.html, **2.** Cheng, T., C. Xu, L. Lei, C. Li, Y. Zhang, and S. Zhou. 2016. Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources* 16: 138– 149. **3.** Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403– 410. **4.** Carrizo García C., Barfuss M.H.J., Sehr E.M., Barboza G.E., Samuel R., Moscone E.A., Ehrendorfer F. Phylogenetic relationships, diversification and expansion of chili peppers (Capsicum, Solanaceae) Ann. Bot. 2016;118:35–51. doi: 10.1093/aob/mcw079