

# Molecular Identification of *Ceratitis capitata* (Tephritidae) and related fruit flies: Transitioning into the DNA Barcode Era



Norman B. Barr<sup>1</sup>, Md. Sajedul Islam<sup>2</sup>, Bruce A. McPherson<sup>2</sup>, & Marc De Meyer<sup>3</sup>

<sup>1</sup> Center for Plant Health Science and Technology, Mission Lab, USDA-APHIS, Moore Air Base, Edinburg, TX 78541 (Norman.B.Barr@aphis.usda.gov)

<sup>2</sup> Department of Entomology, Pennsylvania State University, University Park, PA 16802, USA

<sup>3</sup> Royal Museum for Central Africa, Entomology Section, Belgium



## Objective

- Compare the DNA barcode methodology to a published PCR-RFLP protocol for *Ceratitis* fruit fly identification
- Test the utility of the *Ceratitis* barcode library for reliable identification of *C. capitata*, the Mediterranean fruit fly

## Introduction

The Afrotropical fruit fly genus *Ceratitis* contains many species of agricultural importance. Of these, *Ceratitis capitata* (the Mediterranean fruit fly) is regarded as one of the world's most significant agricultural pests. As is true of other fruit flies, it is not possible to reliably identify *Ceratitis* species using the morphology of immature life stages. Consequently, DNA-based procedures are required in order to confirm the identity of *Ceratitis* species intercepted or collected as larvae.

To date, the most comprehensive taxonomic studies using molecular data for identification of *Ceratitis* species have employed Polymerase Chain Reaction–Restriction Fragment Length Polymorphism (PCR-RFLP; e.g., Barr et al., 2006). PCR-RFLP works by comparing the lengths of particular segments of DNA in different specimens (Figure 1). PCR-RFLP data are qualitative and analog in nature, not quantitative and digital; a researcher must decide if two specimens have the same fragment length or not, and where the dividing line between species should be placed.

Barr et al. (2006) tested the ability of PCR-RFLP to separate and identify *Ceratitis* species using a relatively good, although not complete, sampling of the genus, including collections from its ancestral home range in Africa. These African collections provide a better estimate of genetic variation within *C. capitata*, an important agricultural pest, than do collections derived from introduced populations. In addition, the data set included samples of *Ceratitis caetrata* and *Ceratitis pinax*, two close relatives of *C. capitata*.

Recent studies on fruit fly molecular diagnostics (Armstrong and Ball 2005; Barr 2009) have shown that a DNA barcode approach can outperform similar diagnoses generated using a PCR-RFLP technique. DNA barcodes are DNA sequences from a standardized gene region. In general, DNA barcode information is better suited for the study of international pests because the data are in digital form and are quantitative. Each specimen produces a string of A, G, C and T bases that can be aligned, compared, and converted to percent dissimilarities (or distances) or can be compared as individual bases. This makes barcode data easier to track, share and interpret than data used in other molecular methods.

## Diagnostic and Taxonomic Sampling

Using any diagnostic tool requires that a “taxonomic universe” be explicitly stated for that tool. That is, the species that can and can't be identified by a method is limited to the taxa that have already been studied using that method. Without complete sampling of a taxon (such as the genus *Ceratitis*) it is difficult to claim that an unidentified specimen can be assigned to species because of the risk of a false positive. That is, the specimen may belong to a species that has not yet been studied with that method.

Although numerous studies describing molecular methods to identify *C. capitata* have been published, most include insufficient sampling of other species to demonstrate taxonomic specificity in the techniques. Demonstrating specificity is a crucial aspect in the development of traditional and molecular identification tools.

In this study we generate COI barcode data from fly samples used in the Barr et al. (2006) study to compare methods and develop a new molecular diagnostic tool for *C. capitata*. A phylogenetic analysis of *C. capitata* indicates that *C. caetrata* and *C. pinax* are its closest relatives.

## Tephritid Barcoding Initiative

In April 2006, the Consortium for the Barcode of Life convened an international meeting of taxonomic experts on the fruit fly family Tephritidae at the Royal Museum of Central Africa in Tervuren, Belgium.

Participants in the Tephritid Barcoding Initiative (TBI) share samples, lab and informatics capabilities, and taxonomic expertise. More than 3000 specimens from hundreds of species have been barcoded and made public. We gratefully acknowledge support from the Sloan Foundation, US Dept. of Agriculture and the Belgian government.

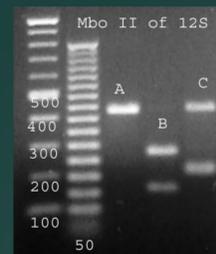


Figure 1: Example of RFLP data: band-patterns on an agarose gel resulting from the differential migration of DNA fragments of different size under an electric field. Three patterns are observed (A, B, and C) that are interpreted as different species.

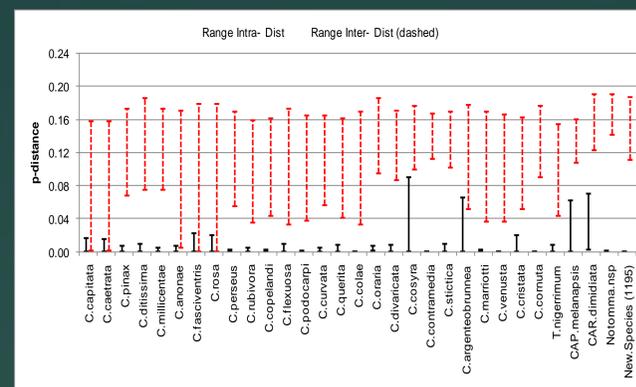


Figure 2: Barcode Gap data for the species represented by more than one individual in the study. Dashed red lines are for inter-specific variation and the black solid lines for intra-specific variation.

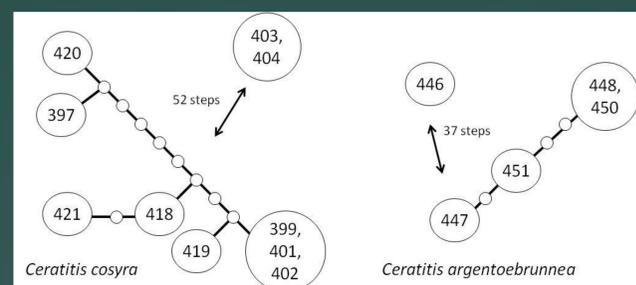


Figure 3: Statistical Parsimony Network for *C. cosyra* and *C. argenteobrunnea*. Although haplotypes can be based on variable within-species sequences (represented by larger circles), the haplotypes of all species are distinct from each other in this analysis. Numbers refer to sample codes.

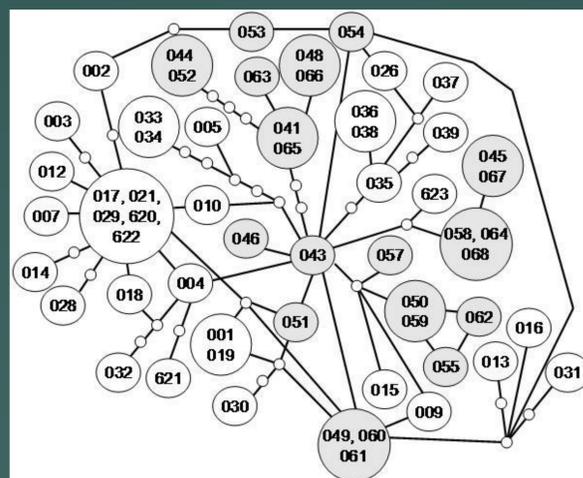


Figure 4: Statistical Parsimony Network for *C. capitata* (white circles) and *C. caetrata* (grey circles). Representatives of each species connects with the other species in many parts of the network, indicating that they are too similar to be discriminated using the COI barcode data alone. Numbers refer to sample codes.

## Materials and Methods

A collection of 239 DNA samples representing 44 taxa were barcoded using the standard COI primers reported by Folmer. These specimens were selected to maximize estimates of diversity by analyzing samples representative of different populations, geographies, and genetic types (as reported in Barr et al. 2006).

Haplotypes were identified and genetic divergence ( $p$ ) estimated using DnaSP v5.10 and MEGA 4.0. Barcode gaps were estimated for each taxon based on maximum intra-specific and minimum inter-specific genetic divergences.

A statistical parsimony network was constructed for the data set using TCS 1.21 and a default connection limit of 95%. Character-based identification of *C. capitata* was tested using nucleotide diagnostic process described by Wong et al. (2009).

## Results and Discussion

For the majority of tested species, the maximum genetic divergence within a species was much lower than the minimum divergence between species (Figure 2). This is referred to as the “barcode gap” and is one way to demonstrate the diagnostic utility of barcode data.

However, not all species showed barcode gaps. *C. capitata* cannot be identified to species because it is too similar to *C. caetrata* (see overlap of colored bars in Figure 2). The three species of the FAR complex (*C. fasciventris*, *C. anonae*, and *C. fasciventris*) also cannot be identified to species-level because of overlap. The PCR-RFLP also failed to discriminate to the species-level within these two lineages.

Several other species lacked a clear barcode gap that would provide unambiguous identification. *C. cosyra* showed a small barcode gap and *C. argenteobrunnea* showed none, but these species can be identified using COI barcodes in a statistical parsimony analysis of networks (Figure 3). The Barr et al. (2006) PCR-RFLP method was able to identify all other species represented in the tool.

As long as the species do not form connections (networks) with other species, the results can support identification of genetic types. Identification using barcode data networks provides equivalent resolution to the PCR-RFLP method and it provides additional information to the diagnostician. It quantifies how different two genetic types are from each other and can point out potential misidentifications based on PCR assay results.

Evaluation of the *C. capitata* samples using barcode gaps, statistical parsimony networks (Figure 4), and character-based searches (not shown) all failed to distinguish the pest species from its close relative *C. caetrata*. However, when these two species are grouped and analyzed as one taxon, it is possible to distinguish it from the next closest relative, *C. pinax*, using all three methods of analysis.

## Conclusions

- The proposed COI DNA barcode analyzed using statistical parsimony networks can identify the same number of species as the PCR-RFLP method
- The DNA barcode cannot identify samples as *C. capitata* but can limit the possibility of identifications to *C. capitata* and *C. caetrata*

## References

- Armstrong & Ball. 2005. Phil. Trans. R. Soc. B 360: 1813-1823  
 Barr et al. 2006. Bull. Entomol. Res. 96: 505-521  
 Barr. 2009. J. Econ. Entomol. 102: 401-411  
 Wong et al. (2009) Mol. Ecol. Res. 9: 243-256