

# DNA BARCODING OF NORTH AMERICA'S LEPIDOPTERA: THE STORY AFTER 5000

Jeremy R. deWaard<sup>a</sup>, Natalia V. Ivanova<sup>a</sup>, Janet C. Topan<sup>a</sup>, Jean-François Landry<sup>b</sup> and Paul D.N. Hebert<sup>a</sup>

<sup>a</sup>Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada, N1G 2W1

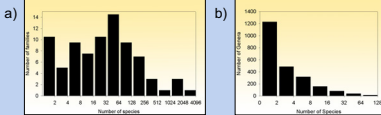
<sup>b</sup>Canadian National Collection of Insects and Arachnids, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada, K1A 0C6



## INTRODUCTION

### TAXONOMIC TARGET

The order Lepidoptera includes more than 160,000 described species. The 12,000 species known from North America belong to 82 families and 2333 genera. Nearly 90% of the genera include 8 or fewer species, but four (*Coleophora*, *Eucosma*, *Eupithecia*, *Euxoa*) are each represented by more than 150 species.



**Composition of North America's Lepidoptera:**  
a) Species/Family and  
b) Species/Genus

### GEOGRAPHIC AND TAXONOMIC COVERAGE



Our work seeks to assemble DNA barcodes for all lepidopteran species at 5 sites in the eastern half of North America. More than 5500 individuals have now been analyzed representing:

- 1250 species
- 650 genera
- 56 families

**Map of eastern North America showing the collection sites with colour codes used on the trees**

## OBJECTIVES

### GOALS OF RESEARCH

The present study tests the feasibility of assembling a barcode library for all North American lepidopterans. It also investigates the likely effectiveness of the barcode system in species identification. Three issues are under investigation:

- 1) Testing the success of a standard primer set in recovering the COI barcode region from lepidopterans
- 2) Ascertaining sequence divergences among closely allied (congeneric) species
- 3) Examining geographic variation in the barcode sequences for single species

## METHODS

### Specimen/Data Collection

- Specimens were collected from five locations
- Each specimen was assigned a specimen accession and sequence ID number
- Collateral data, including GPS information and specimen images, were submitted to the Barcode of Life Database (BOLD)
- A small sample from each voucher specimen was archived in 95% ethanol.

### DNA Extraction

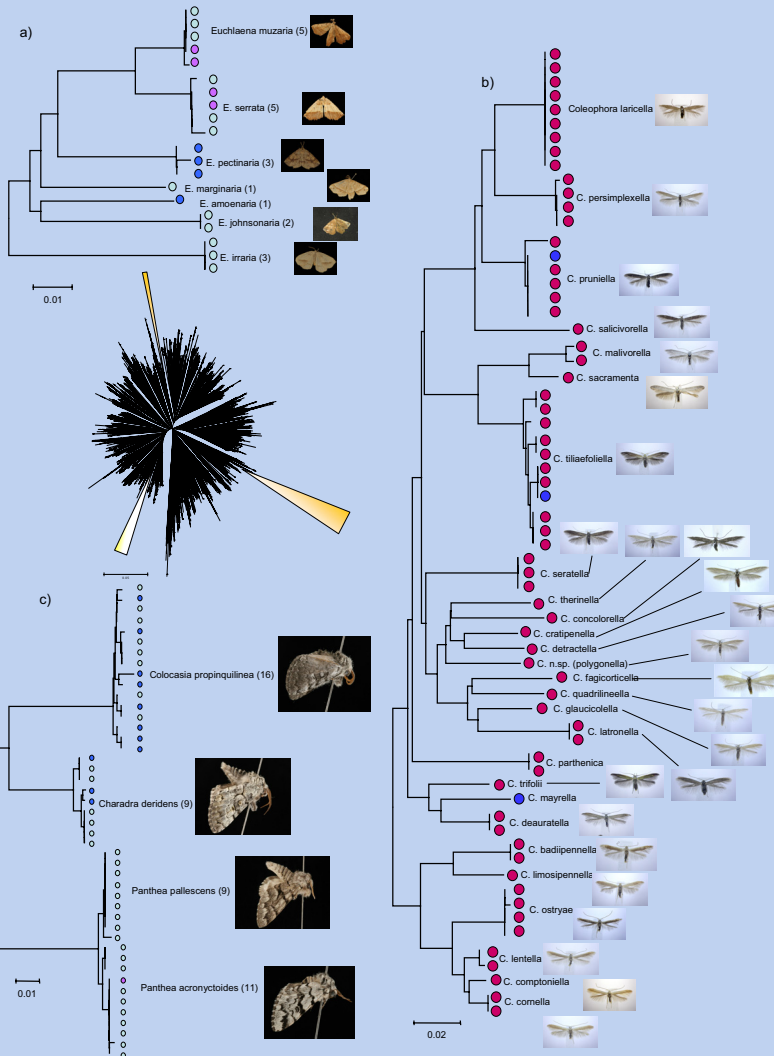
- Genomic DNA was extracted from small pieces of leg using the DryRelease method in 96 well plates

### PCR Amplification

- 658 base pairs of the 5' region of COI was targeted using 10µL PCR reactions and the LepF-LepR primer set

### Sequence Analysis

- Samples were bidirectionally sequenced on an ABI 3730 DNA Analyzer
- Sequences were assembled and edited in ABI SeqScape v2.1.
- Finished sequences were uploaded into BOLD



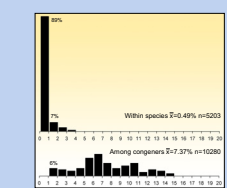
**The central image shows a circularized NJ tree of COI sequence divergences for all 5500 specimens and 1250 species of lepidopterans examined so far. Three representative taxonomic groups are enlarged: a) the genus *Euchlaena*, b) the genus *Coleophora*, and c) the subfamily *Pantheinae***

## RESULTS

### 1) TESTING THE SUCCESS OF BARCODE RECOVERY

Our study involved the analysis of single legs from air-dried specimens. When analyzed within 1 year of their capture, we achieved greater than 90% recovery of a full-length amplification product using an inexpensive and rapid DNA extraction protocol. Success rates drop with age, but specimens less than 5 years old can be analyzed with high success using alternate extraction protocols (e.g. Sigma GenElute kit). In recalcitrant cases, barcode sequences can be generated by concatenating two 350 bp amplicons.

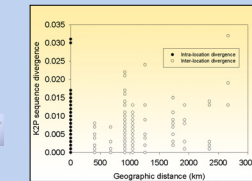
### 2) ASCERTAINING SEQUENCE DIVERGENCE AMONG CONGENERS



The sequence divergence among conspecific individuals averaged 0.49% while congeners showed 7.37% divergence, a 15-fold difference. Many, if not all, cases of high (>1%) intraspecific divergences represent sibling species pairs.

**Sequence divergence within species and between congeners for 100 random genera**

### 3) ASCERTAINING SEQUENCE DIVERGENCE WITH DISTANCE



There was very little incrementation of sequence divergence with geographic distance. As a result, barcodes gathered at a single site are effective in recognizing a species across its range.

**Sequence divergence across space for 100 random species ( $R^2 = 0.06$ )**

## CONCLUSION

### PROSPECTS FOR BARCODING ALL NORTH AMERICAN LEPIDOPTERANS

- 1) This study has established that barcode recovery is straightforward so long as analyses examine recently collected specimens.
- 2) Our results also establish that congeneric species regularly show marked sequence divergence. Interestingly, even in those cases where species pairs show low divergence, they rarely share sequences. One substantive challenge relates to the need for carefully validated identifications, especially for genera with high species diversity and those where prior work has revealed cryptic pairs.
- 3) These results make it clear that a comprehensive barcode library will be highly effective in lepidopteran identification.

## ACKNOWLEDGEMENTS

Funding for this research was provided by the Gordon and Betty Moore Foundation and by NSERC. We thank numerous taxonomists for their assistance in the identification and collection of samples. We particularly thank John Brown, Don Davis, Don Lafontaine, Michael Pogue, Brian Scholtens, Dale Schweitzer, Bo Sullivan, and David Wagner. In addition, we thank Andrea Brauner, Rob Dooh, Angela Holliss, Sujevan Ratnasingham, and Alex Smith for technical support.