Standarization of DNA barcoding techniques to identify *Culicoides* species in Europe

Claudia Paredes-Esquivel, Ricardo del Río Miguel Ángel Miranda. University of the Balearic Islands, Spain
Claire Garros, CIRAD, France

DNA barcoding

CONSORTIUM FOR THE BARCODE OF LIFE

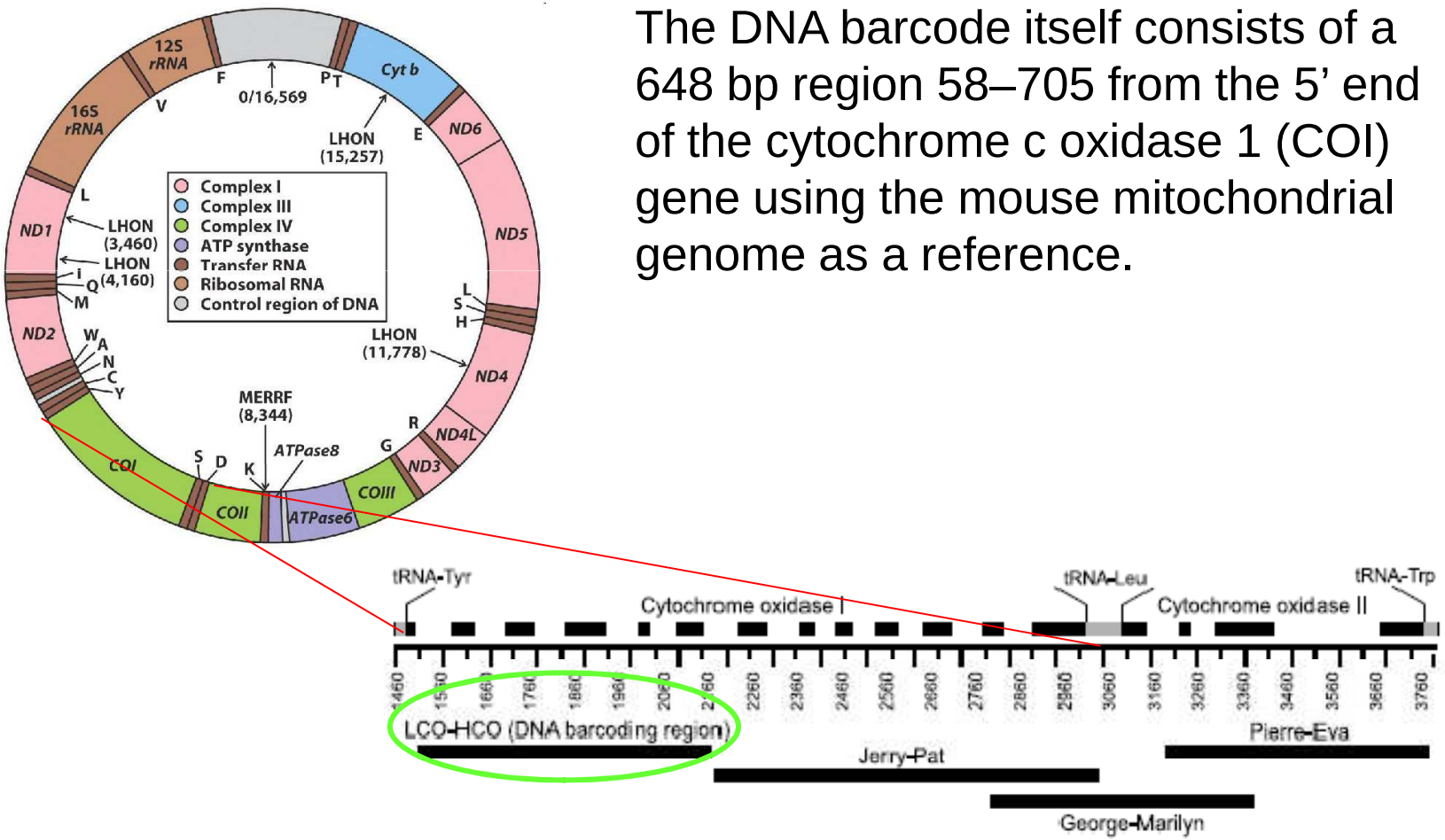
Proposed by Hebert et al. 2003

•DNA barcoding is a technique for characterizing species of organisms using a short DNA sequence from a standard and agreed-upon position in the genome.

Create a universal system for a eukaryotic species inventory

Very controversial!

DNA barcoding: Target molecular marker



The DNA barcode itself consists of a 648 bp region 58–705 from the 5’ end of the cytochrome c oxidase 1 (COI) gene using the mouse mitochondrial genome as a reference.

DNA barcoding: What are the strenghts?

- Obtained cheaply and quickly
- Practical approach
- Works in old deteriorated specimens (museum specimens)
- Works with all stages of life
- Can be applied by non-specialists
- Unmask alikes?

DNA barcoding: Weaknesses

1. Nuclear copies of *COI* (NUMTs)

As whole genome sequencing projects accumulate, more and more Numts have been detected in many diverse eukaryotic organisms

PNAS PNAS PNAS

Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified

Hojun Song*†, Jennifer E. Buhay*‡, Michael F. Whiting*, and Keith A. Crandall*

*Department of Biology, Brigham Young University, Provo, UT 84602; and ‡Belle W. Baruch Institute for Marine Sciences, University of South Carolina, Columbia, SC 29208

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved July 14, 2008 (received for review March 28, 2008)

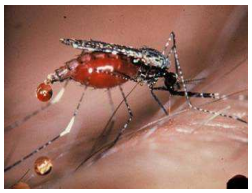
Nuclear mitochondrial pseudogenes (numts) are nonfunctional copies of mtDNA in the nucleus that have been found in major clades of eukaryotic organisms. They can be easily coamplified with orthologous mtDNA by using conserved universal primers; however, this is especially problematic for DNA barcoding, which attempts to characterize all living organisms by using a short fragment of the mitochondrial cytochrome c oxidase I (COI) gene. Here, we study the effect of numts on DNA barcoding based on phylogenetic and barcoding analyses of numt and mtDNA sequences in two divergent lineages of arthropods: grasshoppers and crayfish. Single individuals from both organisms have numts of the COI gene, many of which are highly divergent from orthologous mtDNA sequences, and DNA barcoding analysis incorrectly overestimates the number of unique species based on the standard metric of 3% sequence divergence. Removal of numts based on a careful examination of sequence characteristics, including indels, in-frame stop codons, and nucleotide composition, drastically re-

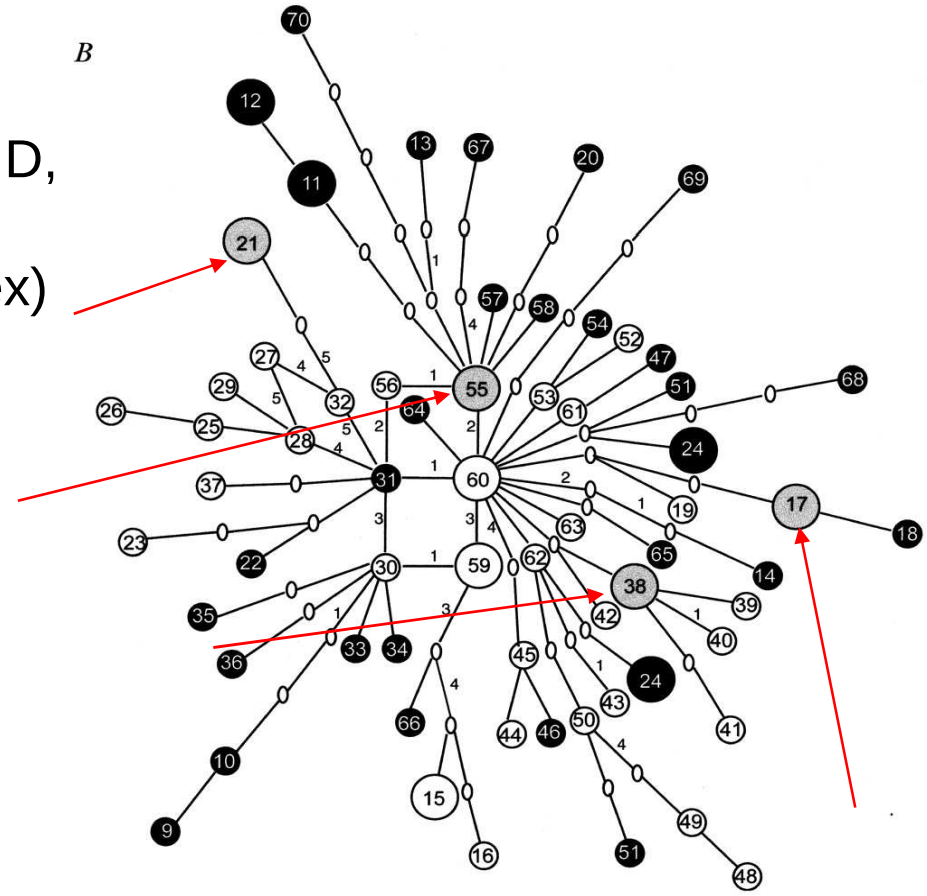
plasmy is the presence of a mixture of more than one type of mitochondrial genome within a single individual, and the coamplification of divergent heteroplasmic copies of mtDNA would lead to an overestimation of the number of unique species under barcoding (3). Maternally inherited symbionts, such as *Wolbachia*, can cause linkage disequilibrium with mtDNA and, whether a population becomes infected with such symbionts, the mtDNA associated with the initial infection will spread throughout the population and result in the homogenization of mtDNA haplotypes (6). Among closely related species, these symbionts can break through the species barrier by hybridization followed by selective sweep, resulting in identical mtDNA sequences among different species, which would cause the underestimation of the number of unique species under barcoding (9). Whereas these three processes may be relatively uncommon and limited to a small number of organisms, a fourth process, the nuclear integration of mtDNA that gives rise to nuclear mitochondrial

DNA barcoding: Weaknesses

2. Shared haplotypes
common COI haplotypes in very closely related species such as sibling species

example in *Anopheles dirus* A and D, *An. arabiensis* and *An. gambiae* (members of the Gambiae Complex) Mathews et al. (2005)



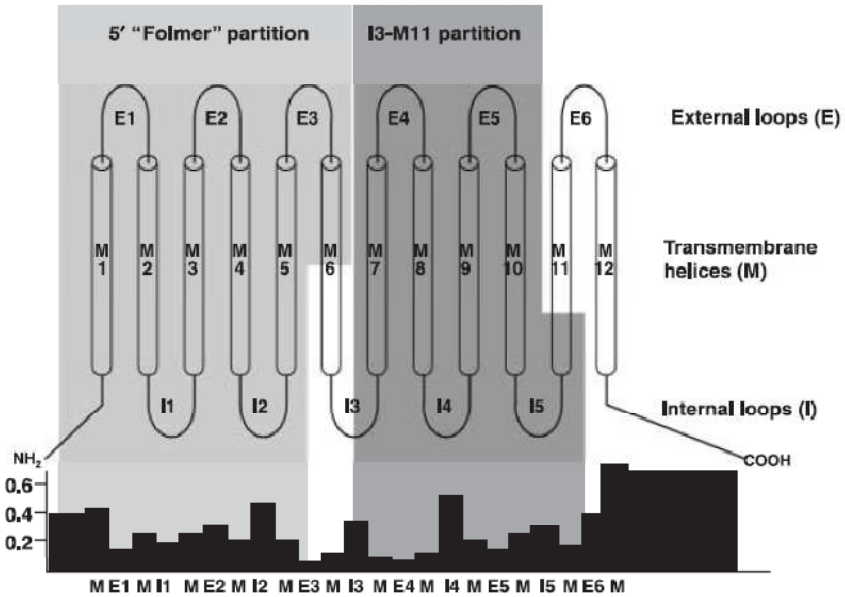


Walton et al. (2000)

DNA barcoding: Weaknesses

3. Rate of evolution

E.g. In diploblasts, mitochondrial genes were found to evolve in a **slower rate** than their bilaterian counterparts.



Lunt et al. Schematic representation of *COI*

Erpenbeck, D., Hooper, J.N.A., Worheide, G., 2006. CO1 phylogenies in diploblasts and the 'Barcoding of Life'—are we sequencing a suboptimal partition? Mol. Ecol. Notes 6, 550–553.

DNA barcoding: Weaknesses

4. Fragment selection

Roe and Sperling (2007)

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 44 (2007) 325–345

MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

There is no a single optim
region of maximal diverge
across all taxa

Patterns of evolution of mitochondrial cytochrome *c* oxidase I and II DNA and implications for DNA barcoding

Amanda D. Roe *, Felix A.H. Sperling

CW403 Biological Sciences Building, Department of Biological Sciences, University of Alberta, Edmonton, Alta., Canada T6G 2E9

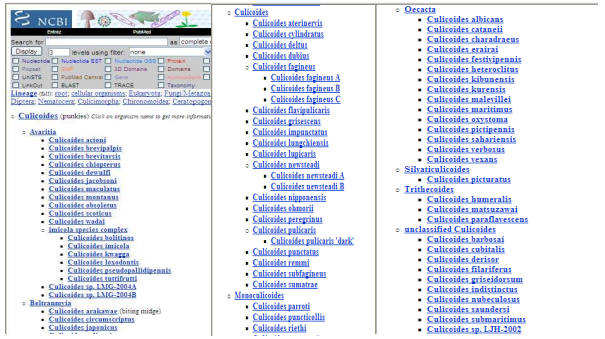
Received 9 August 2006; revised 27 November 2006; accepted 7 December 2006
Available online 20 December 2006

Abstract

DNA barcoding has focused increasing attention on the use of specific regions of mitochondrial cytochrome *c* oxidase I and II genes (COI–COII) to diagnose and delimit species. However, our understanding of patterns of molecular evolution within these genes is limited. Here we examine patterns of nucleotide divergence in COI–COII within species and between species pairs of Lepidoptera and Diptera using a sliding window analysis. We found that: (1) locations of maximum divergence within COI–COII were highly variable among

"If only a small fragment of DNA is sampled, it may fail to produce an accurate representation of the total genetic variability in that gene"
Recommended to maximize the length of the DNA sequence used for initial pilot studies on any taxon (900 bp?)

Barcoding of *Culicoides* species



State of art

811 sequences for *Culicoides* are deposited at the NCBI database (GenBank), representing 73 species *ie* only 5% of the total *Culicoides* species

Except for *C. imicola*, *C. sonorensis*, and *C. nubeculosus*, the number of total sequences per species in the GenBank is very low

⇒ Almost none voucher specimen associated to sequences

⇒ Low representation of geographical distribution

⇒ Makes difficult the use of sequences in phylogenetic or phylogeographic studies

Barcoding of *Culicoides* species

Objective

to develop a molecular library based on accurate morphological data and associated with collection reference

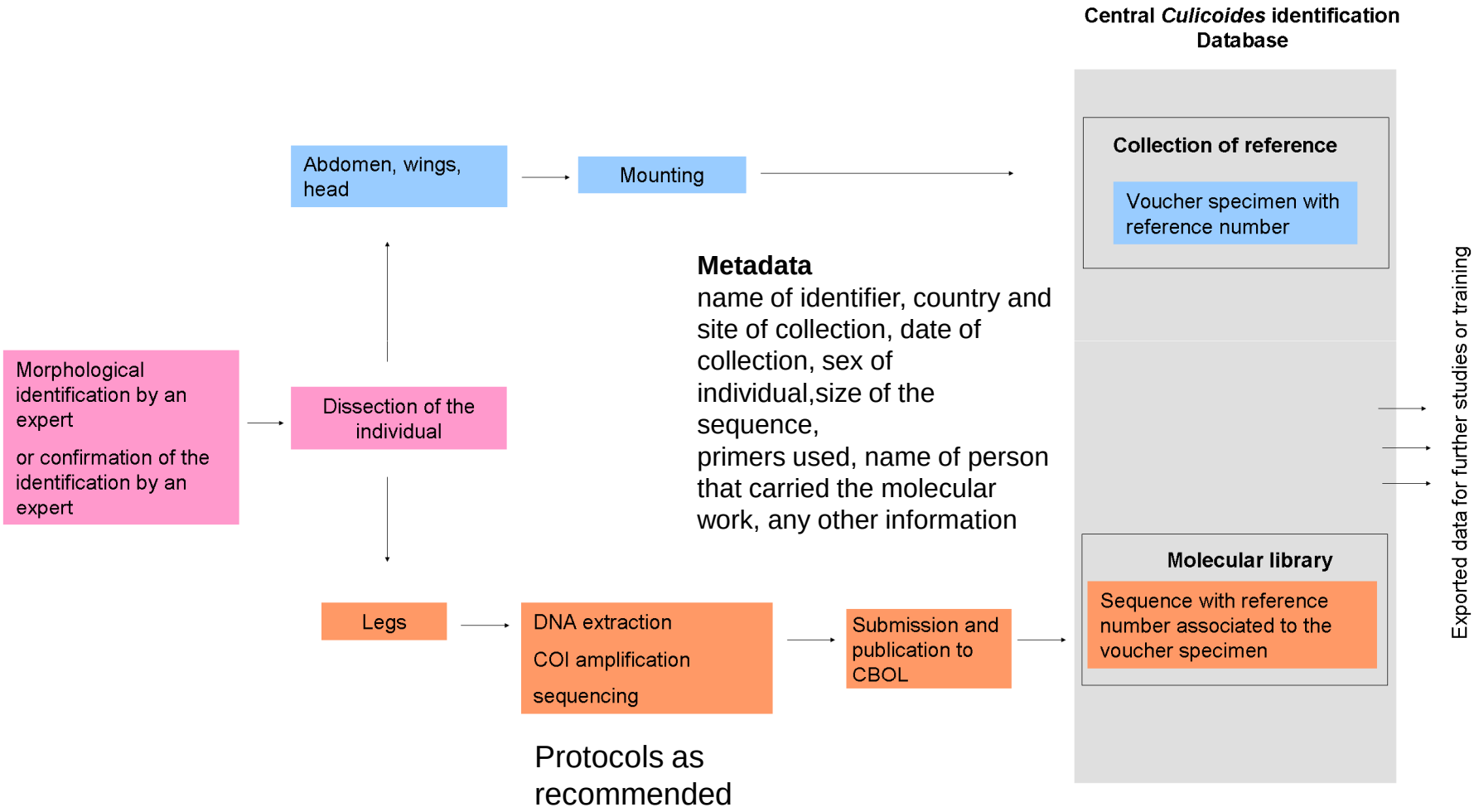
Specific objectives

1. to sequence *COI* barcode region for the *Culicoides* species distributed in western Europe following CBOL recommendations to produce data records recognised as BARCODE records

2. to link barcode records produced to voucher specimens in a central *Culicoides* identification database

3. to release reliable, retrievable and verifiable information concerning the barcode sequence, the voucher specimen and metadata related

Barcoding of *Culicoides* species



Barcoding of *Culicoides* species

Deliverables

- A reliable morphological collection of reference and molecular database for *Culicoides* species in Western Europe for both research and training

- Possibility of oral communications and a scientific paper to release to the scientific community the sequences (see Genebank accession numbers of sequences of *Culicoides* species vectors of bluetongue virus in Germany. Kiehl E, Walldorf V, Klimpel S, Al-Quraishy S, Mehlhorn H. Parasitol Res. 2009 Jul;105(1):293-5. May 2009)

- A strong framework for other countries

Barcoding of *Culicoides* species

To be discussed

- Protocols
- Single locus strategy
- Implications of national experts for morphological identification

Thanks for your attention!