

DNA Barcoding and taxonomy of *Glossina*

Dan Masiga

Molecular Biology and Biotechnology Department, *icipe*

&

Johnson Ouma

Trypanosomiasis Research Centre, KARI



The taxonomic problem

- Following ~250 years of morphology based taxonomy only 1.7 million of the 10-100 million species of organisms on our planet (land and water) have been classified.
- Human being only have a limited ability to recognize and recall morphological variation.
- Few taxonomists can reliably diagnose an assemblage of even 1000 species. And their number is not growing.
- This provides the motivation for a new approach to species recognition.

Without accelerating this process, and assuming the same number of active taxonomists remain active, it will take about 940 years to describe all species (Seberg, 2004).

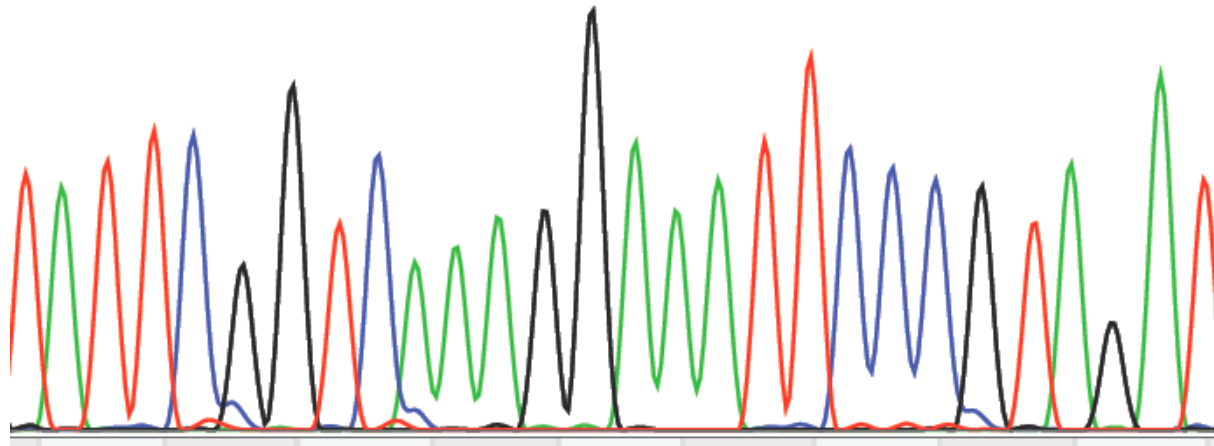
What is DNA Barcoding?



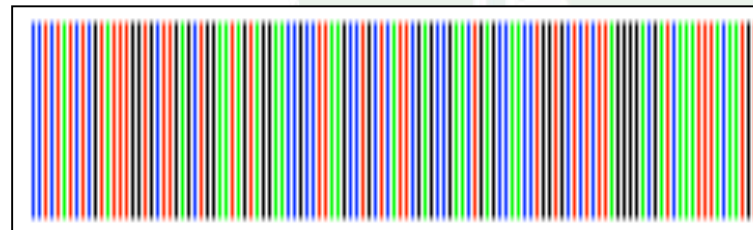
The Universal Product Code (UPC) system used in supermarkets to brand retail items employs 10 options at each of 11 positions to create 100 billion alternates.

DNA Barcoding: analogy to UPC system

T A T T C G G T C A A A G G A A A T T C C C G T A G A T
V F G Q R K F P * I
160 170 180



Barcode →



Characteristics of Barcode Regions

- Low intra-species variability (ideally, coding sequence)
- Discontinuous variation between species
- Flanked by conserved regions
- Easy to amplify
- Long enough to work in all groups
- Short enough for single reads

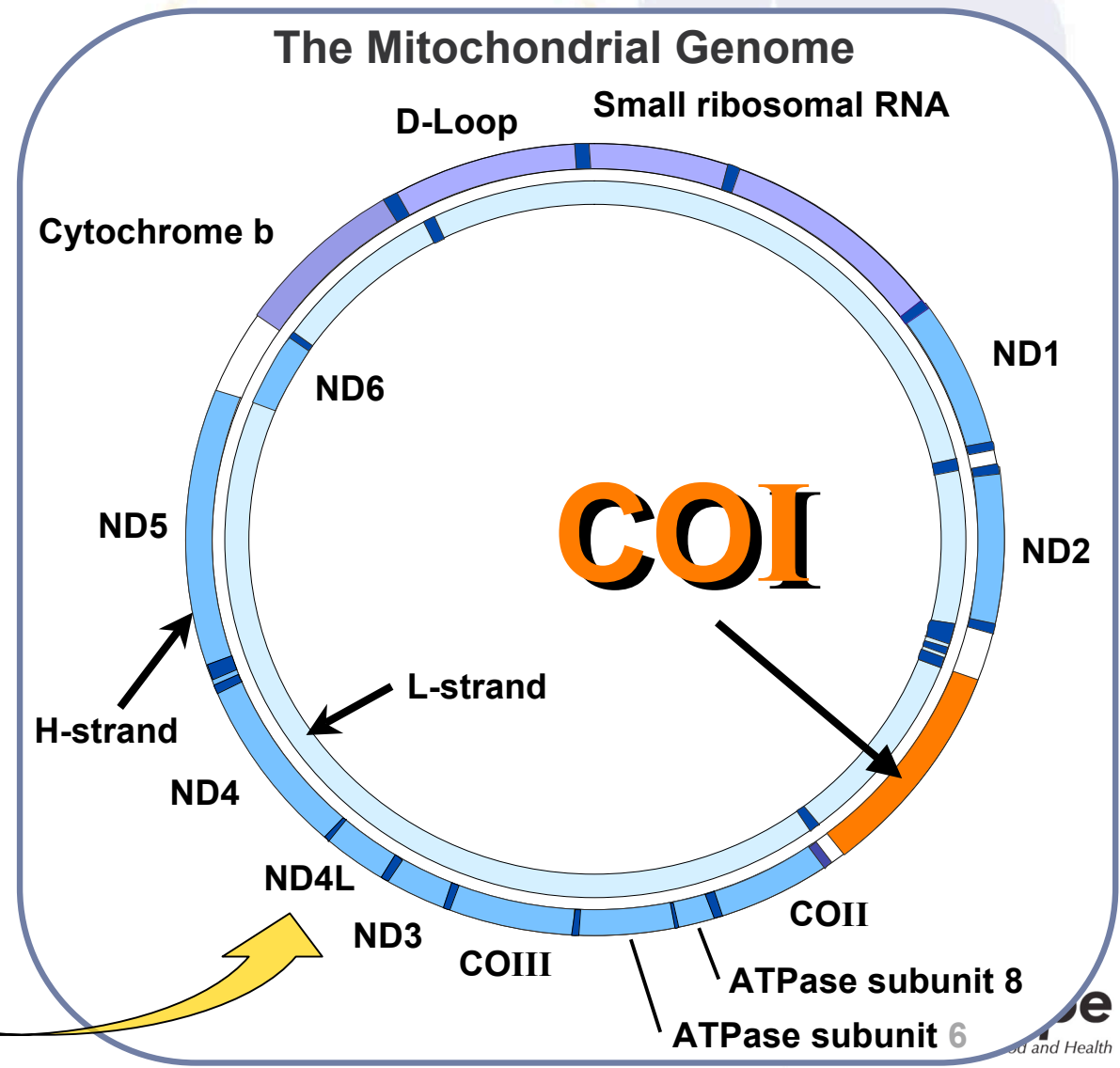
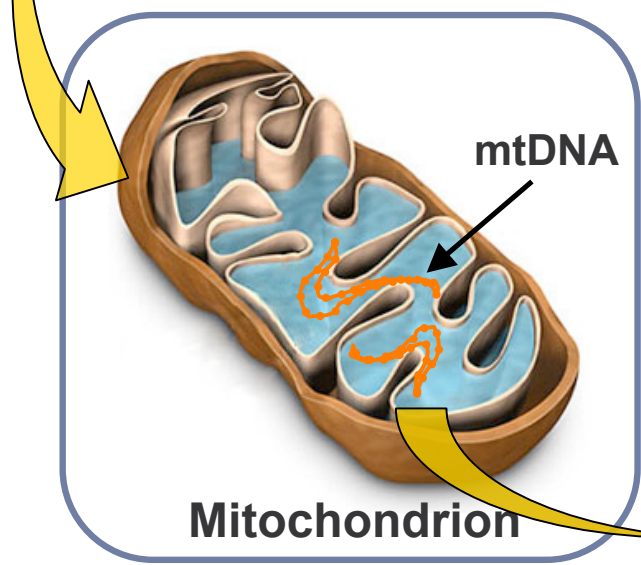
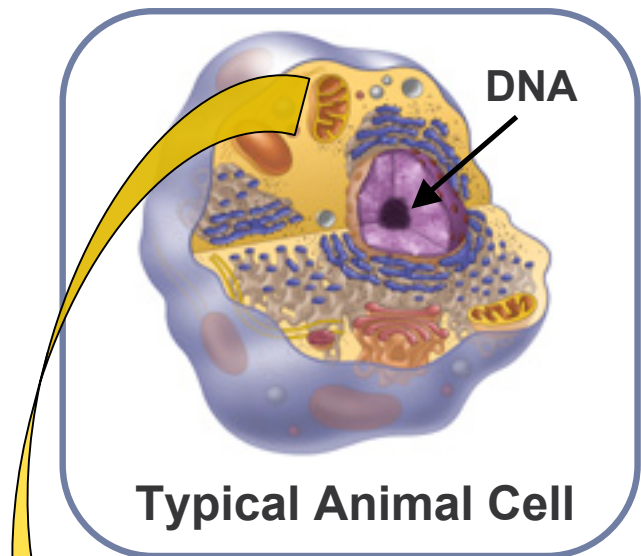
Using DNA Barcodes

- Establish reference library of barcodes from identified voucher specimens
- If necessary, revise species limits
- Then:
 - Identify unknowns by searching against reference sequences
 - Look for matches (mismatches) against 'library on a chip'
 - Before long: Analyze relative abundance in multi-species samples

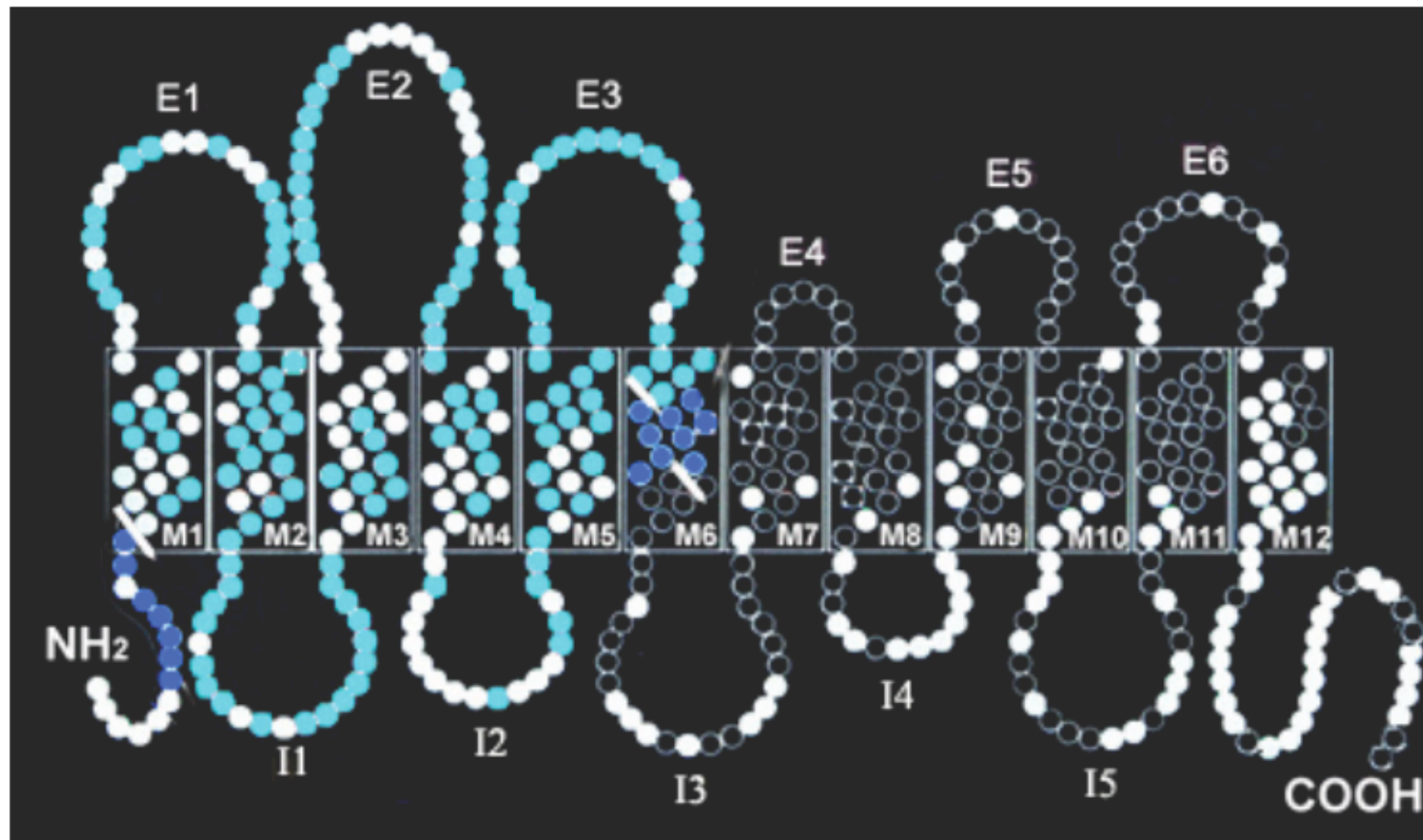
What DNA Barcoding is NOT

- ***Barcoding is not DNA taxonomy;*** no single gene (or character) is adequate
- ***Barcoding is not Tree of Life;*** barcode clusters are not phylogenetic trees
- ***Barcoding is not just COI;*** standardizing on one region has benefits and limits
- ***Molecules in taxonomy is not new;*** but large-scale and standardization ***are new***
- ***Barcoding can help to create a 21st century research environment for taxonomy***

An Internal ID System for All Animals



The choice of mt Cytochrome Oxidase I (COI) for animals



Variable amino acid positions in COI protein from 9 insect species

Barcode of Life Data System

BARCODE OF LIFE INITIATIVE

Advancing species identification and discovery through the analysis of short, standardized gene regions



BOLDSYSTEM

Username

Password

Login

[Request a new user account](#)

BARCODING LINKS

- ☞ Guelph Centre for DNA Barcoding
- ☞ Consortium for the Barcode of Life (CBOL)
- ☞ Rockefeller University
- ☞ Canadian Barcode of Life Network

GENOMICS LINKS

- ☞ National Center for Biotechnology Information (NCBI)
- ☞ DNA Data Bank of Japan (DDBJ)
- ☞ European Bioinformatics Institute (EMBL)

BOLD STATS

Species Barcoded	17,253
Total Barcodes	90,095
Source	Breakdown
GenBank	26,398
Guelph Facility	61,237
Others	2,460

BARCODING CAMPAIGNS EMPLOYING BOLD

Three major campaigns, involving more than 100 researchers will collectively assemble barcode records for more than 50,000 species by 2011.



All Leps

Lepidoptera: This initiative will assemble barcodes for 25,000 species of Lepidoptera with a focus on the faunas of Australia, Canada, Costa Rica and United States.

[View](#)



FISH-BOL

Fishes: The FISH-BOL campaign will gather barcodes for at least 15,000 fish species with an emphasis on marine species.

[View](#)



ABBI

Birds: The All-Birds Barcode Initiative (ABBI) plans to assemble DNA barcodes for all 10,000 bird species within 5 years.

[View](#)

The Barcode of Life Data(BOLD) system is an online workbench that aids collection, management, analysis, and use of DNA barcodes. It includes 3 components that address varied segments of the barcode lifecycle.

• [What is BOLD?](#)



MANAGEMENT & ANALYSIS SYSTEM

BOLD-MAS provides a repository for barcode records coupled with analytical tools. It serves as an online workbench for the DNA barcode community.

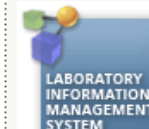
• [Tutorial & Documentation](#) • [View published projects](#)



IDENTIFICATION SYSTEM

BOLD-IDS provides a species identification tool that accepts DNA sequences from the barcode region and returns a taxonomic assignment to the species level when possible.

• [Identify specimens](#)

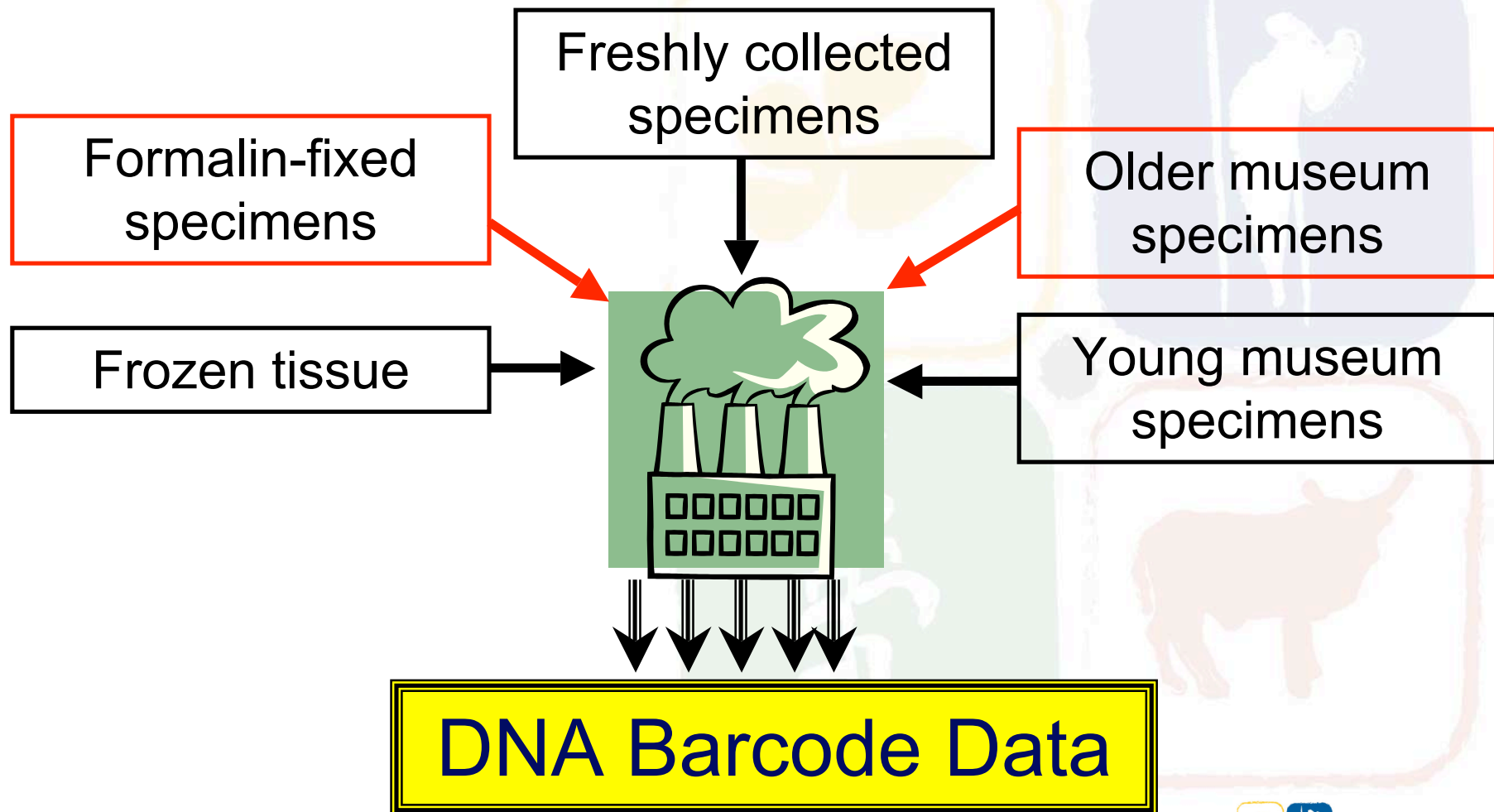


LABORATORY INFORMATION MANAGEMENT SYSTEM

BOLD-LIMS provides an information management solution for high-throughput DNA barcode laboratories. The system automates data management and workflow processes.

The Barcode Assembly Line: 2008

Opening the museum treasure-trove



Glossina barcode project

~ 33 species/sub-species in 3 sub-genera
(Gooding and Krafur, 2005)

Sub-genera:

Morsitans

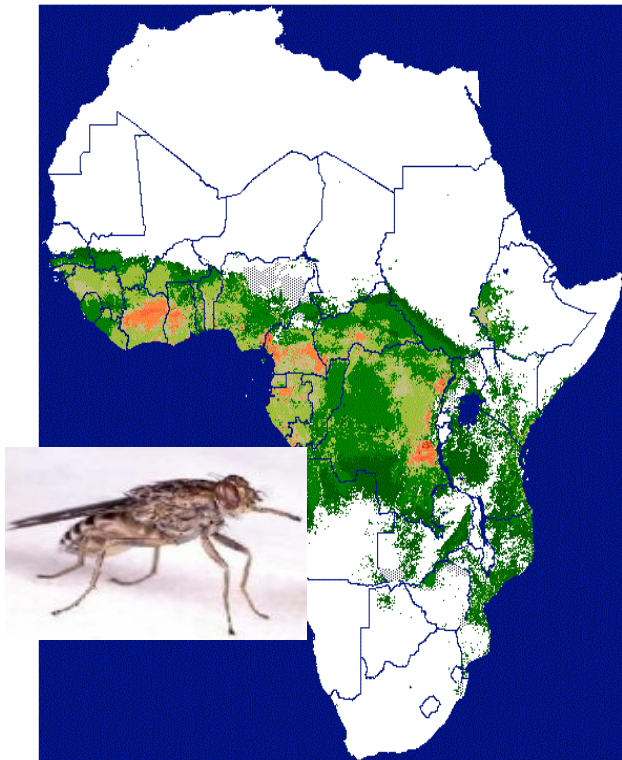
Fusca

Palpalis (Nemorhina)

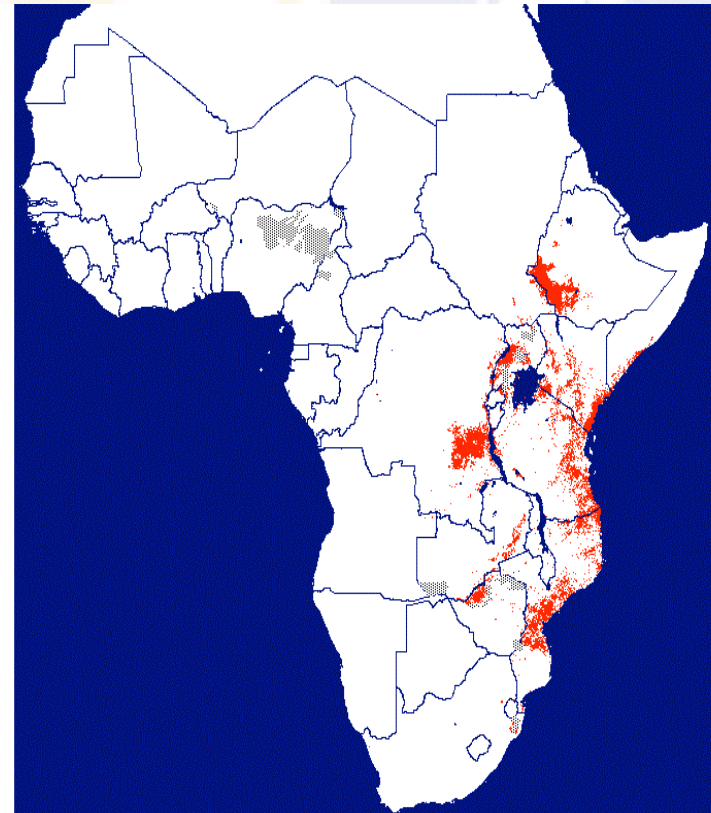
Integrate *Glossina* Barcodes with data from trypanosomes they transmit; co-adaptation

Tsetse Distribution

General



G.pallidipes



Source: Oxford/TALA



Glossina Barcoding [GLOBE]

Project Title :

Glossina Barcoding

Campaign :

None (General Project)

Place in container:

Independent Project

Project Description :

This project aims to generate barcodes for tsetse flies (Genus: Glossina), obligate blood-feeding dipterans that occur only in Africa. Tsetse flies are the sole vectors of trypanosomes, blood parasites that cause trypanosomiasis in



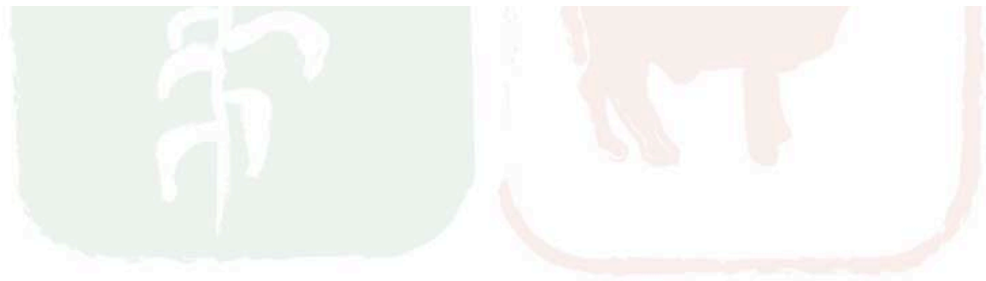
Project Manager :

Daniel Masiga

Assign Users :

User names are ordered alphabetically. You can jump to a name by pressing the first letter of the name in a selected menu.

User	Sequence Access	Specimen Access
Select User	<input type="checkbox"/> Analyze <input type="checkbox"/> View <input type="checkbox"/> Edit	<input type="checkbox"/> Edit Specimens
Select User	<input type="checkbox"/> Analyze <input type="checkbox"/> View <input type="checkbox"/> Edit	<input type="checkbox"/> Edit Specimens
Select User	<input type="checkbox"/> Analyze <input type="checkbox"/> View <input type="checkbox"/> Edit	<input type="checkbox"/> Edit Specimens
Select User	<input type="checkbox"/> Analyze <input type="checkbox"/> View <input type="checkbox"/> Edit	<input type="checkbox"/> Edit Specimens
Select User	<input type="checkbox"/> Analyze <input type="checkbox"/> View <input type="checkbox"/> Edit	<input type="checkbox"/> Edit Specimens



Species with COI sequences

G. austeni

G. f. fuscipes

G. m. centralis

G. m. morsitans

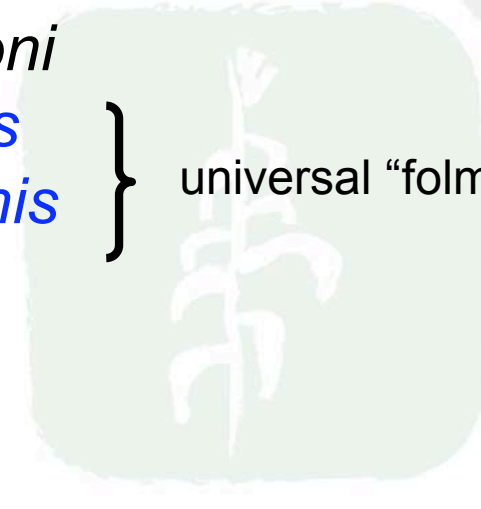
G. m. submorsitans

G. pallidipes

G. swynnertoni

{ *G. brevipalpis*
G. longopennis }

universal “folmer” primers not successful



**BOLD - ID**

The BOLD Identification System (IDS) accepts sequences from the 5' region of the mitochondrial gene COI and returns a species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications.

The reference database of validated records is used by default and is recommended for all identification purposes.

Search Database:

- All Available Barcode Records (208,595 Sequences/27,381 Species)**
Full database of barcode records with a minimum sequence length of 500bp (warning: unvalidated dataset). This includes many species represented by only one or two specimens as well as all species with interim taxonomy.
- Reference Barcode Database (91,880 Sequences/9,614 Species)**
Validated subset of the full database with a minimum sequence length of 500bp and containing only those species represented by three or more individuals showing less than 2% sequence divergence

Enter sequences in fasta format:

```
>G m submorsitans
TGGTCAACCAAATCATAAAGATATTGGAACATTATATTTTATTTTTGGTACATGATCTGGAATAGTTGGAACATCTTT
AAGAATCTTAGTACGAGCTGAGCTAGGACATCCAGGAGCATTAAATTGGAGATGATCAAATTTATAATGTTATTGTCA
CAGCTCATGCATTTGTTATAATTTTTTTTATAGTAATGCCTATTATAATTGGAGGTTTTGGAATGATTAGTTCCTT
TAATATTAGGAGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGATTTTGATTACTTCCTCCTGCTTTATCTC
TTCTTTTGATGAGAAGTATAGTAGAAAATGGAGCCGGGACTGGATGAACTGTATATCCTCCACTTTCTTCTATTATT
GCTCATGGAGGAGCTTCAGTAGATTTAGCAATTTTTCTTTACATTTAGCAGGATTTTCTTCAATTTTAGGAGCAGTA
AATTTTATTACTACAGTTATTAATATACGATCTACAGGAATTTCTTTTCGATCGAATACCTTTATTTGTTTGATCTGN
GTAATTACAGCTTTATTACTCTTATCTTTACCTGTATTAGCAGGAGCAATTACTATATTATTAACAGACCGAAAT
TAAATACTTCATTTTTTGACCCCGCCGGAGGAGGATCCAATTCCTTATCAACACTTATTTTGATTTTTTGTCAC
C
```

Submit

Cancel

BOLDSYSTEMS

Management & Analysis



Specimen Identification

Search Request:

Type : Reference Database Search

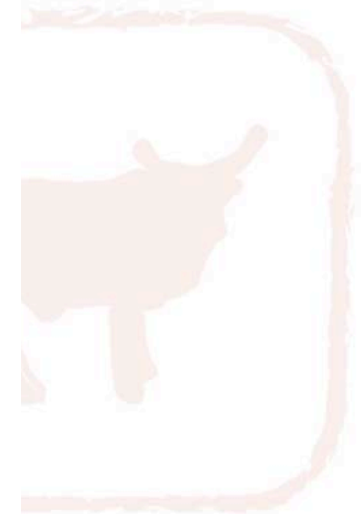
Search Result:

Identification Summary :

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Arthropoda	100
Class	Insecta	100
Order	Diptera	100
Family	Tachinidae	94.9
Genus	<i>Blepharipa</i>	49.5

A species level match could not be made, The nearest neighbor is *Blepharipa albicauda*.

Tree Based Identification



icipe

African Insect Science for Food and Health

Considering the broad distribution of tsetse and trypanosomes and the difficulty in accessing samples from all foci, this effort is only realistic as a collaborative activity.

Reactions to DNA Barcoding

- **From ecologists and other users:**

“This is what we need”

- **From traditional taxonomists:**

“Species should be based on lots of characters, not just barcodes”

“This technology is taking money away from museums”

- **From forward-looking taxonomists:**

“Using molecular data as species diagnostics isn’t new, but standardization and broad implementation are great!”

- **From barcoding practitioners:**

“I had my doubts at the beginning, but it really works as a tool for identification (96% accurate in a recent mollusc paper) and it is at least as good as traditional approaches to discovering new species.”

Barcode data standards

Formal barcode status is given when seven data elements are present:

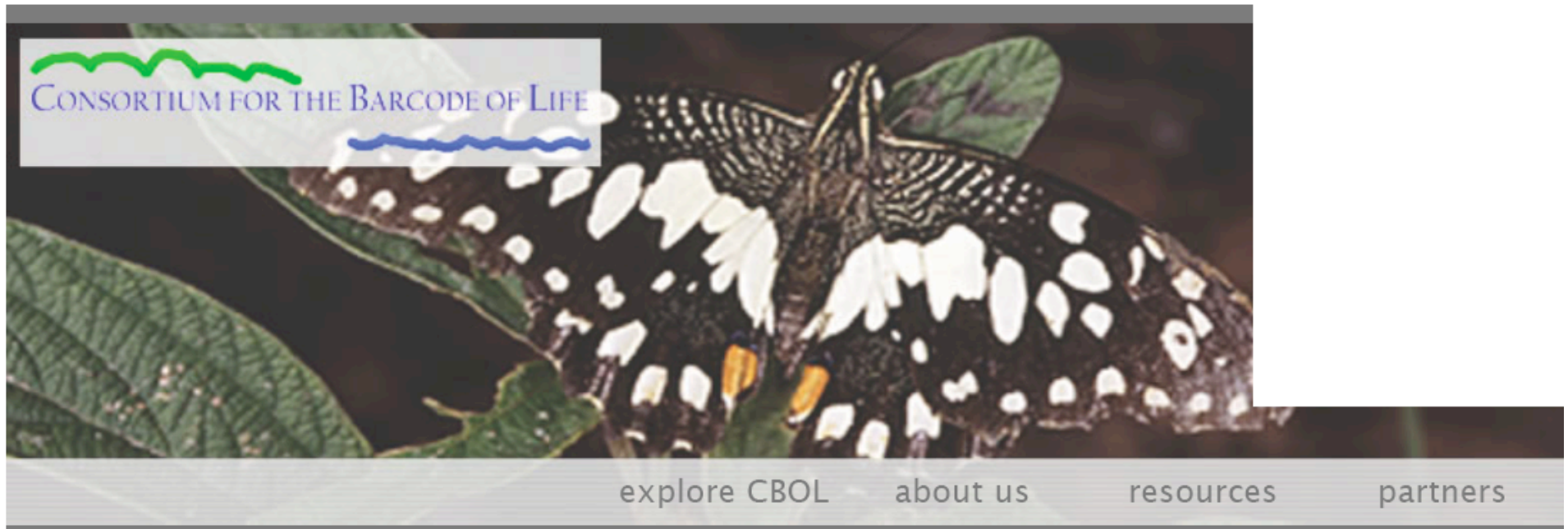
- Species name (could be interim)
- Voucher data (institutional record)
- Collection record (collector, collection date, location with GPS coordinates)
- COI sequence, at least 500bp
- PCR primers
- Trace files (chromatograms)

Producing Barcode Data: 2010?

Barcode data anywhere, instantly



- Data in seconds to minutes
- Pennies per sample
- Link to reference database
- A taxonomic GPS
- Usable by non-specialists



CBOL: www.barcoding.si.edu