Plant DNA Barcoding

II MSc Botany

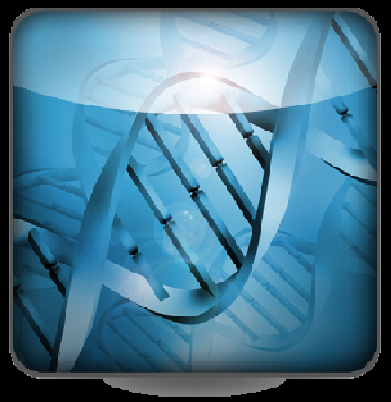
Dr Giby Kuriakose

What is plant DNA barcoding and why is there a need for it?

How Barcoding works

Plants are sampled

DNA is extracted



“Barcode” amplified

ACGAGTCGGTAGCTGCCCTCTGACTGCATCGAA
TTGCTCCCCTACTACGTGCTATATGCGCTTACGAT
CGTACGAAGATTTATAGAATGCTGCTACTGCTCC
CTTATTCGATAACTAGCTCGATTATAGCTACGATG

Sequenced plant DNA is compared with sequences in a barcode database

How many species can you name?

How many Animals did you name?
How many mammals?
How many plants?
How many insects?

"Cat"
Felis catus

"Dog"
Canis lupus familiaris

"Oak Tree"
Quercus alba

"Shark"
Ginglymostoma cirratum

"Beetle"
Popillia japonica

Problem 1: No one knows how many species there are.

Mammals	5,490
Birds	9,998
Reptiles	9,084
Amphibians	6,433
Fishes	31,300
Total	62,305

Insects	1,000,000
Mollusks	85,00
Crustaceans	47,000
Corals	2,175
Arachnids	102,248
Total (+others)	1,305,250

Angiosperms	281,821
Gymnosperms	1,021
Ferns and Allies	12,000
Mosses	16,236
Green and Red Algae	10,134
Total	321,212

- **Between 1.5 and 2 million species have currently been described.**
- **It is estimated that this may represent as little as half of the true number.**
- **Perhaps more than 1/3 of all species are threatened**

(IUCN Red list version 2010.1)

Problem 2: Lack of agreement on what “species” means.

Canis lupus

Canis lupus (familiaris)

Anas platyrhynchos

Defining what species are is a complex task

Dependent on many factors

- Interbreeding capabilities
- Morphological variation
- Ecological context
- Genetic similarities

Problem 3: Current taxonomic methods may be insufficient.

Classical taxonomy uses terminology that can act as a barrier to understanding and reduce the number of persons qualified to describe biodiversity

Leaves alternate proximally, opposite and ultimately decussate distally, 6–16 × 4–13 cm; petiole ca. as long as blade, winged, base clasping, basal lobes stipulate, growing as extensions of wings, less than 1 mm wide; blade 5–7-veined, ovate, glabrous, base typically sagittate, margins entire, apex acute to acuminate. Staminate inflorescences axillary, 1–2 per axil, paniculate, fasciculate; panicles bearing flowers singly,bracteolate, in a zigzag pattern along rachis, internodes less than 2 mm; rachis to 25 cm, secondary axes 1–3(–6), fasciculate, less than 3 cm, each subtended by deltate-ovate bracteole shorter than 1 mm. *Pistillate* inflorescences solitary, 4–8(–20)-flowered, 6–35 cm, internodes ca. 1 cm

The body form ranges from hemispherical (e.g., *Cleidostethus*) to elongate oval (e.g., *Clypastraea*) to latridiid-like (e.g., *Foadia*). Corylophids are typically dull brown, but some species have contrasting yellowish-brown patches on the pronotum or elytra. The integument is often densely punctured and may be glabrous or bear short, fine recumbent setae. Most corylophid adults can be diagnosed using the following morphological features: Maxilla with single apical lobe; Mesotrochanter short and strongly oblique; Head usually covered by pronotum; Frontoclypeal suture absent; Antennae elongate with 3-segmented club; Procoxal cavities closed externally; Tarsal formula 4-4-4; Pygidium exposed

Adding to the complexity, if the specimen to be identified is immature in its development or damaged and incomplete, identification may be impossible.

Using DNA can help clarify identity of different looking life forms, as well as provide the resolution that allows to differentiate species that appear identical, yet aren't.

Fig. 1. Newly eclosed female *A. fulgerator* (species LOHAMP, voucher code 02-SRNP-9770) from the ACG.

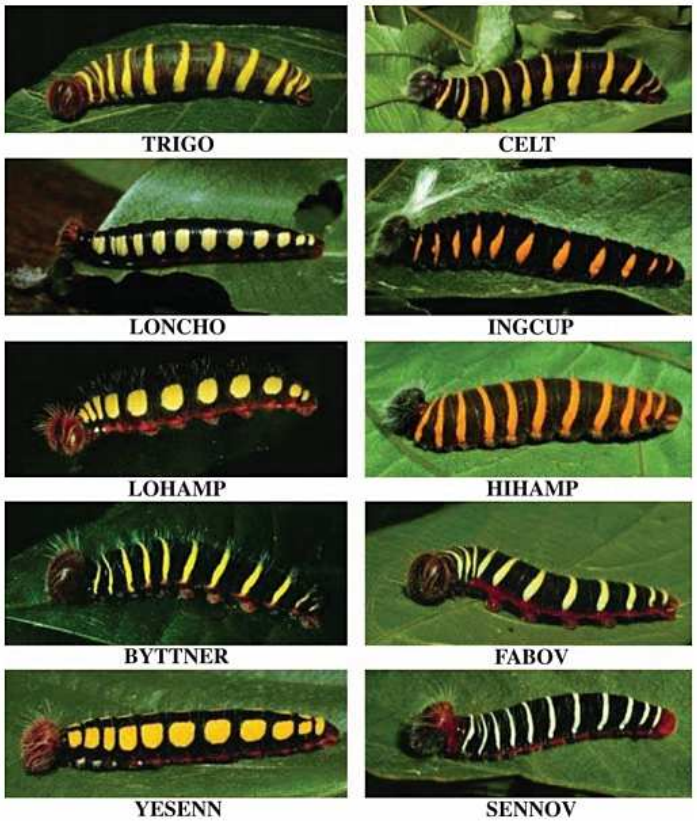


Fig. 2. Last-instar caterpillars of 10 species in the *A. fulgerator* complex from the ACG. Interim names reflect the primary larval food plant and, in some cases, a color character of the adult.

Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*

Paul D. N. Hebert*†, Erin H. Penton*, John M. Burns‡, Daniel H. Janzen§, and Winnie Hallwachs§

*Department of Zoology, University of Guelph, Guelph, ON, Canada N1G 2W1; †Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0127; and §Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

Contributed by Daniel H. Janzen, August 20, 2004

Leaves alternate proximally, opposite and ultimately decussate distally, 6–16 × 4–13 cm; petiole ca. as long as blade, winged, base clasping, basal lobes stipulate, growing as extensions of wings, less than 1 mm wide; blade 5–7-veined, ovate, glabrous, base typically sagittate, margins entire, apex acute to acuminate. Staminate inflorescences axillary, 1–2 per axil, paniculate, fasciculate; panicles bearing flowers singly,bracteolate, in a zigzag pattern along rachis, internodes less than 2 mm; rachis to 25 cm, secondary axes 1–3(–6), fasciculate, less than 3 cm, each subtended by deltate-ovate bracteole shorter than 1 mm. *Pistillate* inflorescences solitary, 4–8(–20)-flowered, 6–35 cm, internodes ca. 1 cm

Complex and subjective

>Dioscorea alata (matK) gene, partial

Simple (A,T,G, or C), objective

ATTTAAATTATGTGTCAGATATATTAATACCCCATCCCATCCATCTGGAAATCCTGGTTCAAATACTTCAATGCTGGACTCAAGATGTTTCCTCTT
TGCATTTATTGCGATTCTTTCTCCACGAATATCATAATTCGAAT AGTTTCATTACTCCGAAAAAACCTATTTACGTGATTTCAATTTCAAAAGAAA
ATAAAAGATTTTTTCGAT TCCTATATAATTCTTATGTATTTGAATGTGAATTTGTATTAGTTTTTTTTCATAAGCAATCCTCTTATTT ACGATCAA
GGTCCTCTGGAGTCTTTCTTGAGCGAACACATTTCTATGGAAAAATGGGGCATTTTTTAGTAGTGTGTTGTAATTATTTTCAGAAGACCCAATG
GTTCTTCAAAGATCCTTTTCTGCATTATGTTCGATATC AAGGAAAAGCAATTCTGGTGTCAAAGGGAACTCGTCTTTTGATGAGGAAATGGAGA
TCTTACCTTGTCCATTTTTGGCAATATTATTTTCAATTTTGGTCTCATCCGCATAGGATTCATATAAACCAATTATCAAATTATTCCTTCTGTTTTC
TGGGTTATCTTTCAAATGTACTAATAAATTTTTCCGTGGTAAGGAGTCAAATGTTAGAAAATTCATTTGTAATAGATACTCTTACTAAGAAATT
TGATACCAGAGTTTCAGTTATTGCTCTTATTCG ATCATTGTCTAAAGCGAAATTTTGTACCGTATCCGGGCATCCTATTAGTAAGTCAATATGGA
CAAATTTC TCAGATTTGGATATTATTCATCGATTTGGTTGGATATGTAGAA

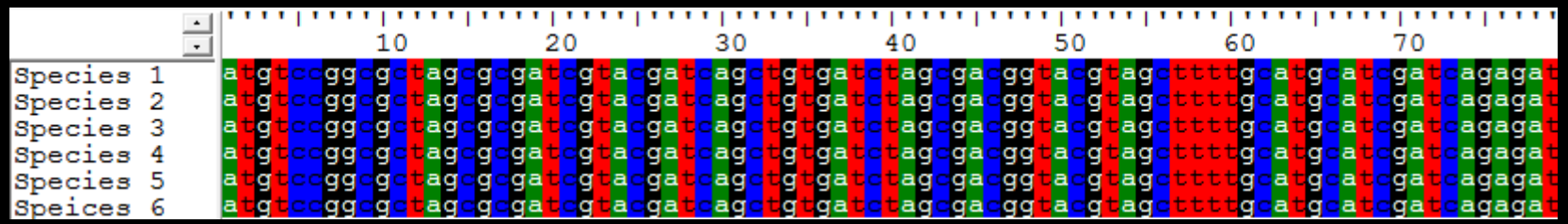
Choosing a DNA barcode

Several criteria go into selecting a DNA region that can serve as a barcode locus, including:

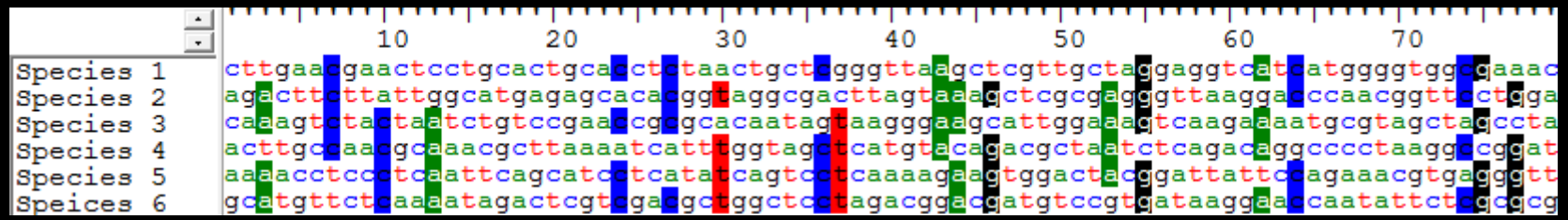
- Discriminatory
  - Unique for species, identical for species members
- Universal
  - Occurs in all species to be examined
- Robust
  - Can be amplified by PCR, using a small number of primers

Candidates

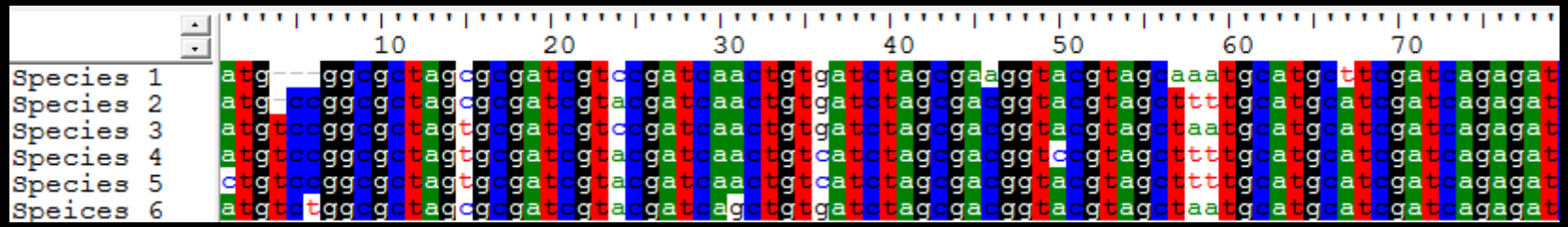
Fail: Sequence is completely conserved, good for PCR, but uninformative as barcode



Fail: Sequence shows no conservation, impossible for PCR, but good as barcode



Win: Sequence shows some (ideally ~70%) conservation, good for PCR, good as barcode



DNA Barcoding
Plants vs. Animals

The Consortium for the Barcode of Life recommended to use the mitochondrial gene for Cytochrome Oxidase I, *CO*I, as universal barcode for animals.

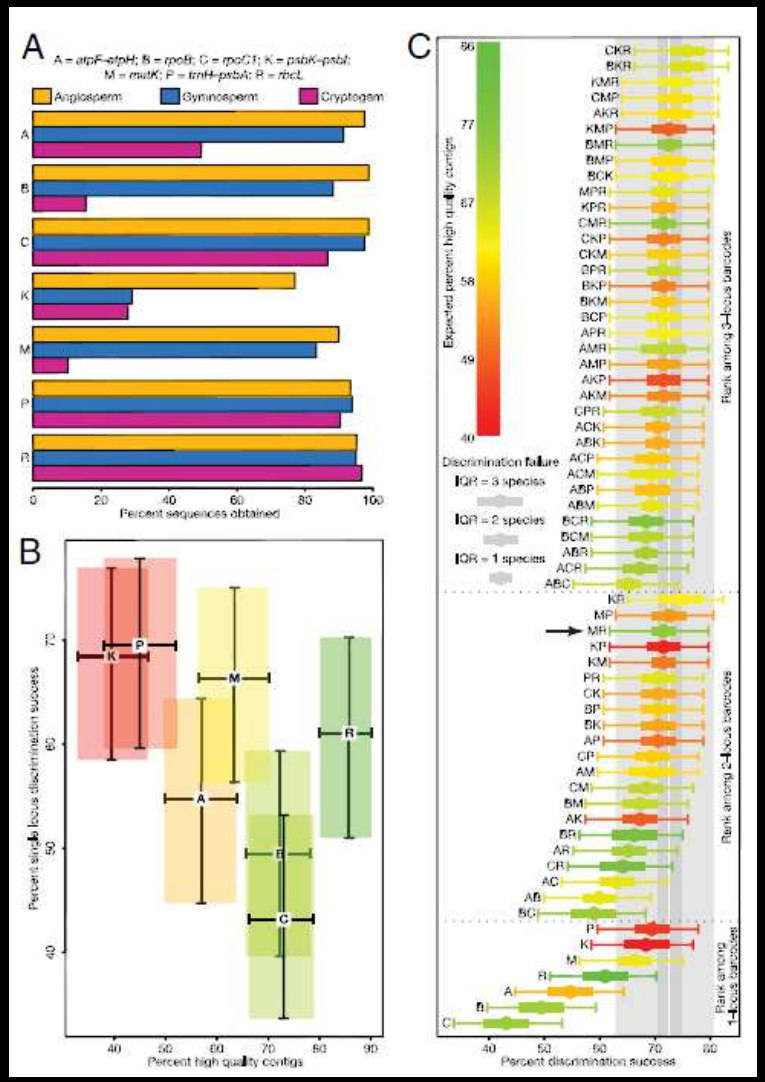
Caveat: As any other potential DNA barcode locus, *CO*I also entails the possibility of failure. Barcoding specific animals for which *CO*I does not work may therefore require newly identifying an appropriate barcoding locus.

PNAS

A DNA barcode for land plants

CBOL Plant Working Group[1]

Communicated by Daniel H. Janzen, University of Pennsylvania, Philadelphia, PA, May 27, 2009 (received for review March 18, 2009)



The Consortium for the Barcode of Life recommended to use the chloroplast genes *rbcL* and *matK* as universal plant barcodes.

Caveat: As any other potential DNA barcode locus, *rbcL* and/or *matK* entail the possibility of failure. Barcoding specific plants for which *rbcL* and/or *matK* do not work may therefore require newly identifying an appropriate barcoding locus.

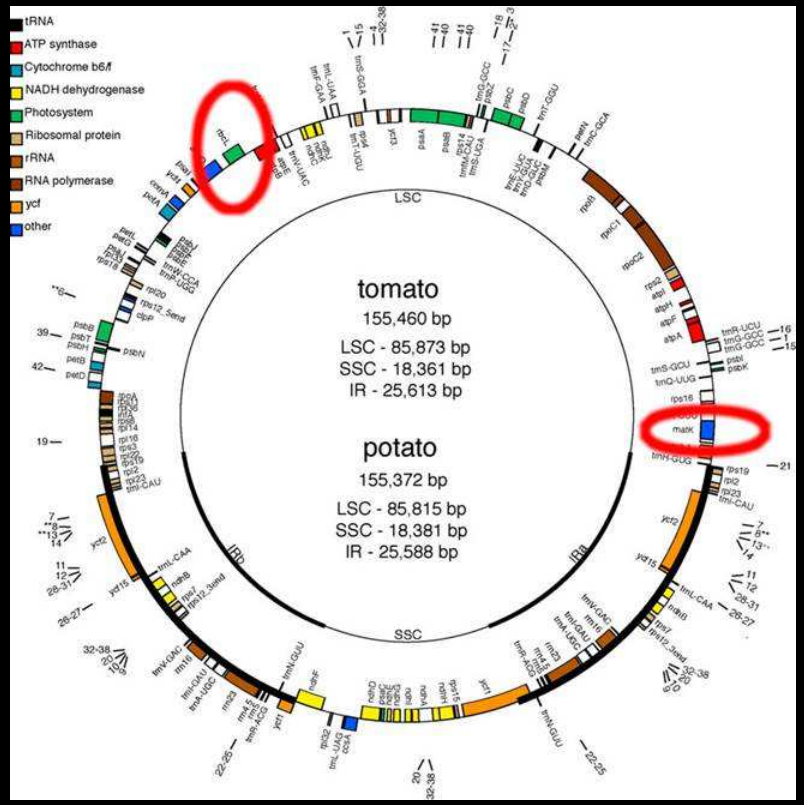
DNA Barcoding
Plants vs. Animals

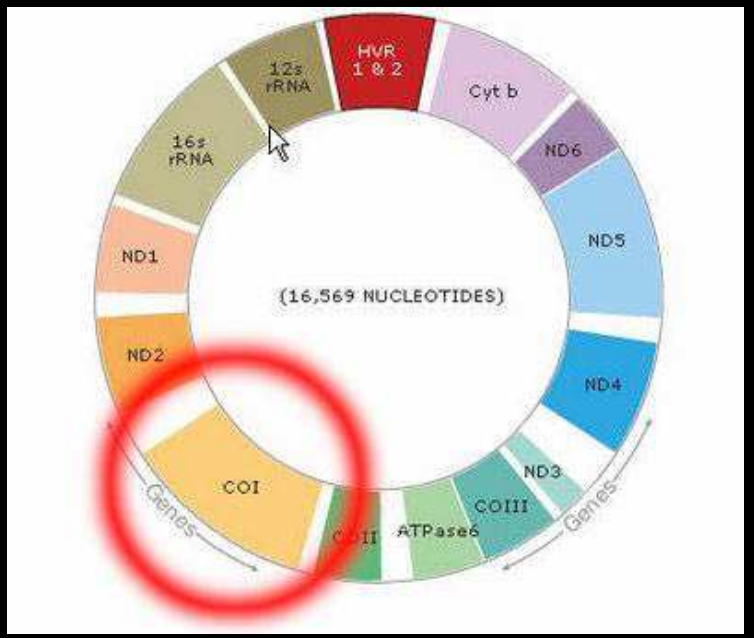
Plants

*Rbc*L
*Mat*K

Animals

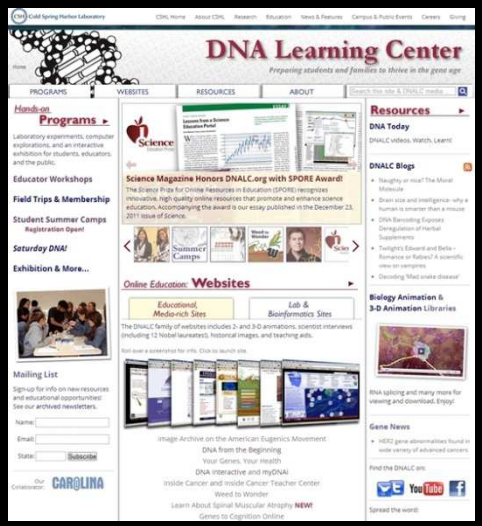
*CO*I



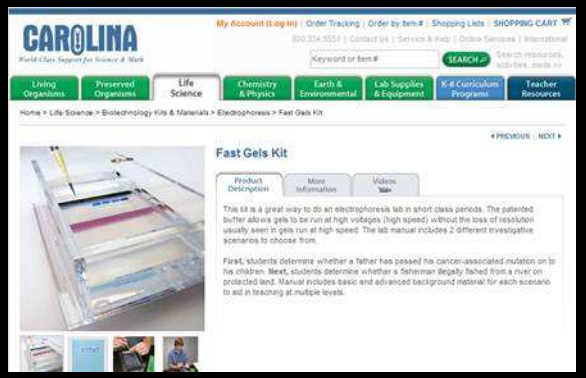


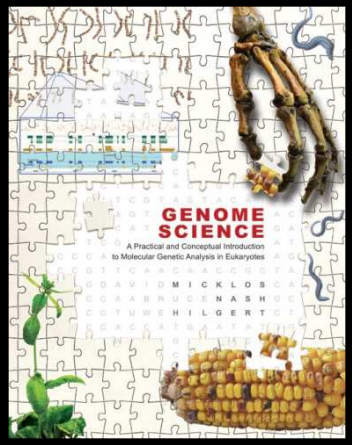
Contributing to Big Science

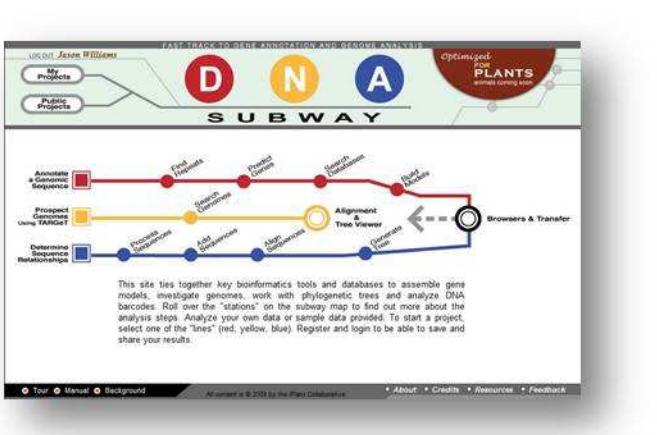
DNA Barcoding: The Works



- Extract DNA
- Amplify barcoding locus
- Sequence barcode
- Analyze sequence







DNA barcoding

is a new and exciting tool for characterizing species of organisms of all life forms using a short DNA sequence from a standard and agreed-upon position in the genome.

Barcode characteristics

A candidate DNA barcode should:

- be known to be orthologous between specimens;
- Short unique gene sequence

- encompass sufficient variability to allow discrimination between species;
- show low variability within individuals belonging to the same species.

Barcode benefits

- It is a taxonomic identification tool alternative or additional to morphology;
- DNA sequencing is a rapid and relatively low cost technique;
- Can process a great number of specimens at a time, thus is useful for example in biodiversity surveys.
- Once reference database is established, it can be applied by non-specialist.

It is an effective technique in which extracted DNA from the collected sample is processed following the standard protocol.

Identification of the species is carried out by amplifying highly variable region i.e., DNA barcode region of the nuclear, chloroplast or mitochondrial genome using Polymerase Chain Reaction (PCR).

Region widely used for DNA barcoding include nuclear DNA (e.g. *ITS*), chloroplast DNA (e.g. *rbcL, trnL-F, matK, psbA, trnH, psbK*) and mitochondrial DNA (e.g. *COI*)

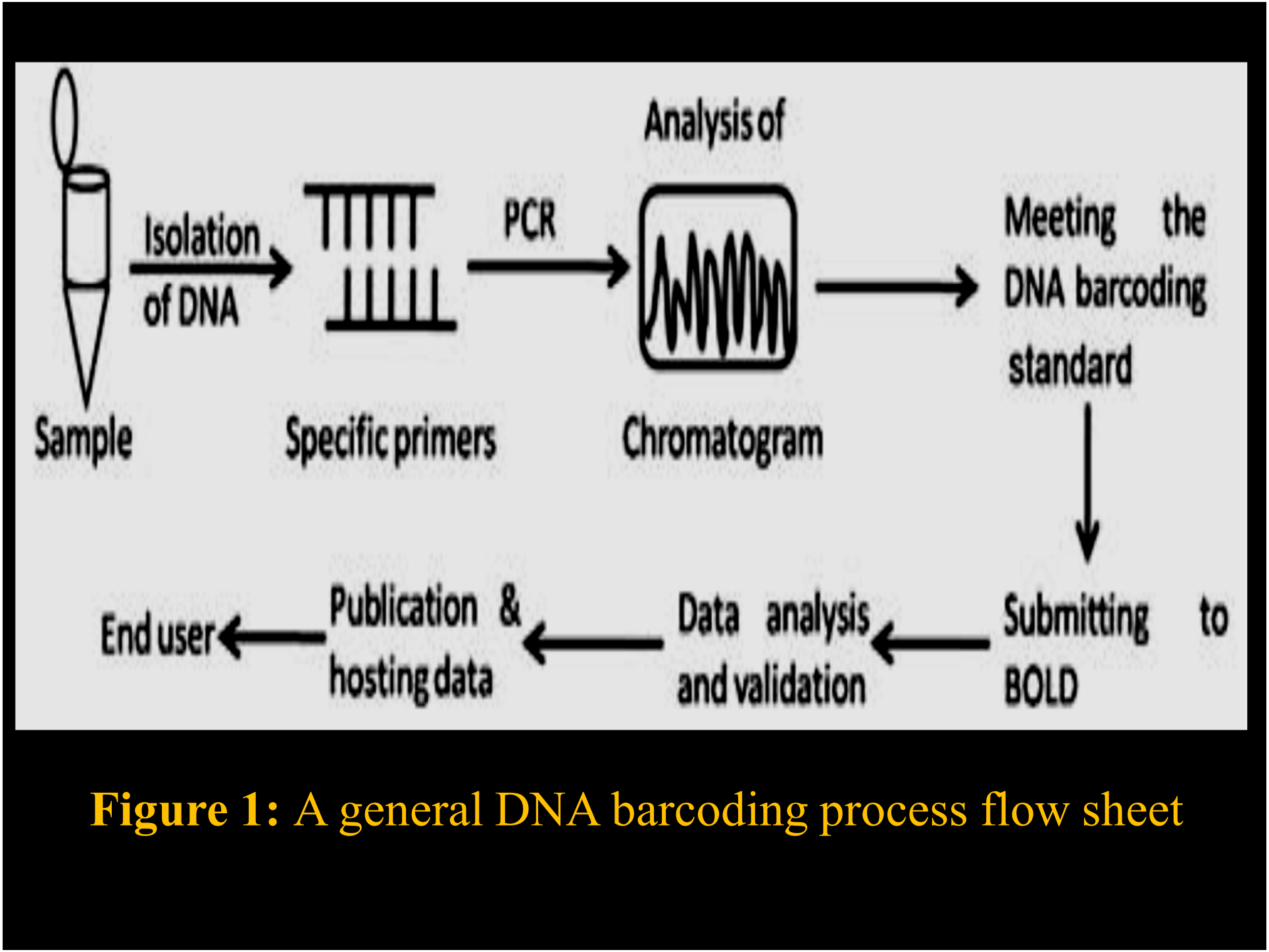


Figure 1: A general DNA barcoding process flow sheet

Limitations

- It is not always true that intraspecific variability is negligible, or at least lower than interspecific values
- There is no universal DNA barcode gene
- Barcode sequences should be generated from type specimens, thus rely on classical taxonomy

Applications

- Biodiversity studies
- New species identification (e.g. in medicine, bacteriology, etc)
- Disease diagnosis (e.g. in veterinary, parasitology, etc.)
- Pest diagnostics in agriculture (e.g in food farming sciences)

Sequences used for DNA barcoding

- Nuclear small subunit ribosomal RNA gene (SSU)
- Nuclear large subunit ribosomal RNA gene (LSU)
- Internal transcribed spacer section of the ribosomal RNA cistron (ITS) and the chloroplast ribulose biphosphate carboxylase large subunit (*rbc*L) genes for plants.
- Mitochondrial cytochrome *c* oxidase I gene for metazoa

COI barcode?

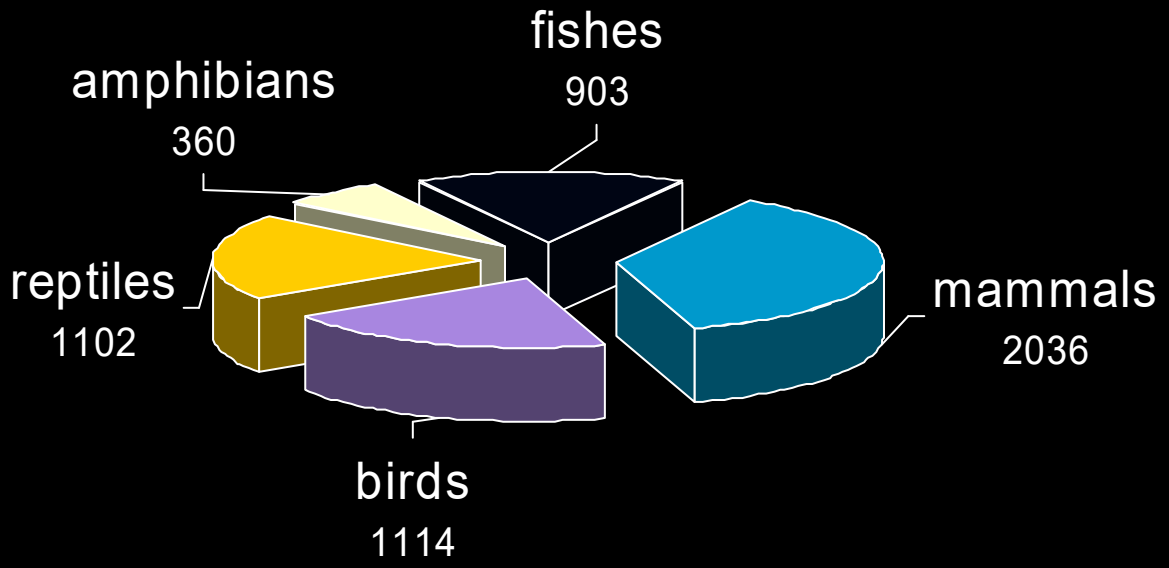
- High copy number (100-10000 copies of mt genes *vs* 2 copies of nuclear genes for each cell)
- Recovering mtDNA sequences is easier and cheaper than nuclear DNA
- Greater difference among species (5-10 fold higher in mt than in nuclear genes).
- Few differences among species
- Absence of introns
- mt genes are orthologous in metazoa

Barcode studies rely on

partial COI sequences

(approximately the first 600 nt of the gene)

COI barcoding: the state of the art in vertebrates



Total number of available COI sequences: 5515

Genbank, April 2005

COI barcoding:
a candidate for inter/intraspecies comparison

Genus *Litoria*

Taxonomic position:
Amphibia; Batrachia; Anura; Neobatrachia; Hyloidea; Hylidae; Pelodryadinae

COI barcoding:
a candidate for inter/intraspecies comparison

COI complete sequence (as reference): *Hyla chinensis*, 1542 nt

Litoria caerulea:	1 sequence	(556 nt)
Litoria eucnemis:	1 sequence	(527 nt)
Litoria genimaculata:	18 sequences	(518-557 nt)
Litoria nannotis:	33 sequences	(561-578 nt)
Litoria rheocola:	30 sequences	(549-577 nt)
Litoria serrata:	25 sequences	(527-575 nt)

Genbank, April 2005

COI barcoding: inter/intraspecies variability

Litoria	AF304206.1S	AF304208.1S	AF304220.1S	AF304222.1S	AF304224.1S	AF304271.1R	AF304273.1R	AF304275.1R	AF304277.1R	AF304285.1R
AF304206.1S	0	0.001938	0.003876	0.125969	0.127907	0.211240	0.213178	0.207364	0.184109	0.186047
AF304208.1S	0.001938	0	0.005814	0.127907	0.129845	0.211240	0.213178	0.207364	0.186047	0.187984
AF304220.1S	0.003876	0.005814	0	0.124031	0.125969	0.207364	0.209302	0.203488	0.180233	0.182171
AF304222.1S	0.125969	0.127907	0.124031	0	0.001938	0.207364	0.209302	0.203488	0.182171	0.184109
AF304224.1S	0.127907	0.129845	0.125969	0.001938	0	0.205426	0.207364	0.201550	0.180233	0.182171
AF304271.1R	0.211240	0.213178	0.207364	0.184109	0.186047	0	0.003876	0.013566	0.007752	0.009690
AF304273.1R	0.211240	0.213178	0.207364	0.186047	0.187984	0.003876	0	0.013566	0.007752	0.009690
AF304275.1R	0.207364	0.209302	0.203488	0.180233	0.182171	0.013566	0.013566	0	0.005814	0.011628
AF304277.1R	0.207364	0.209302	0.203488	0.182171	0.184109	0.007752	0.007752	0.005814	0	0.005814
AF304285.1R	0.205426	0.207364	0.201550	0.180233	0.182171	0.009690	0.009690	0.011628	0.005814	0

Genbank, April 2005

Barcode: standardization

- Standardization lowers costs and lifts reliability, and thus speeds diffusion and use;
- Standardization should help accelerate construction of a comprehensive, consistent reference library of DNA sequences and development of economical technologies for species identification
- NCBI is now beta-testing the public submission tool for the Barcode Section of GenBank. to submit DNA barcodes data directly to GenBank.