

# DNA Barcoding



D N A  

---

SUBWAY



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



# How Barcoding works



Plants are sampled



DNA is extracted



“Barcode” amplified

ACGAGTCGGTAGCTGCCCTTGACTGCATCGAA  
TTGCTCCCTACTACGTGCTATATGCGCTTACGAT  
CGTACGAAGATTATAGAACATGCTGCTACTGCTCC  
CTTATTGATAACTAGCTGATTATAGCTACGATG



Sequenced DNA is compared with plants 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?

“Dog”

*Canis lupus familiaris*



“Cat”

*Felis catus*

“Shark”

*Ginglymostoma cirratum*



“Beetle”

*Popillia japonica*

“Oak Tree”

*Quercus alba*





**Problem 1: No one know how many species there are.**



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

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

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

- There are currently between 1.5 and 2 million described species
- It is estimated that this number may represent as little as half of the true number of species
- Perhaps more than 1/3 of all species are threatened  
(IUCN Red list version 2010.1)





**Problem 2: Even though there are millions of species, there is also a lack of agreement on what a “species” means.**

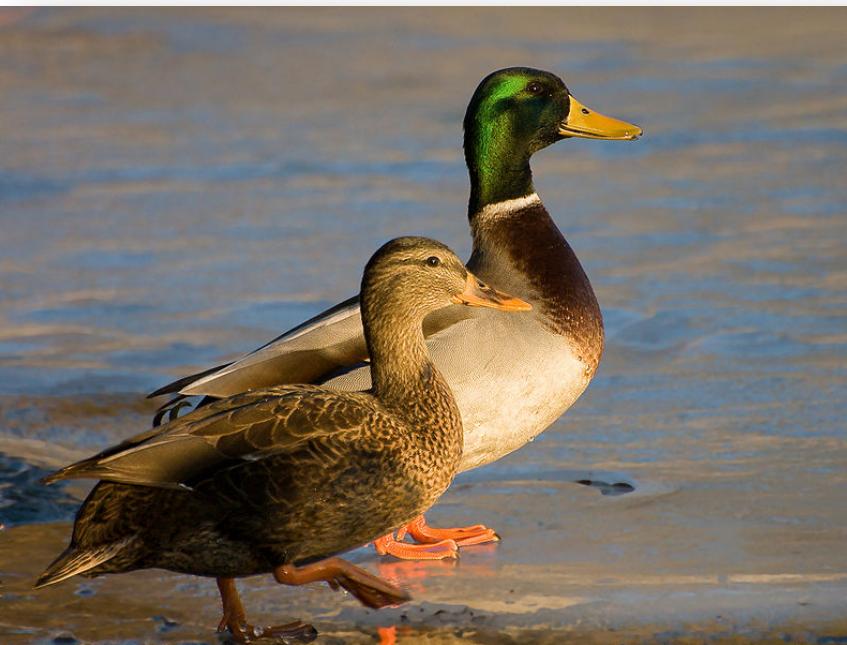




*Canis lupus*



*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 inadequate  
(or at least too slow) to capture vanishing biodiversity**



# Classical taxonomy is steeped in terminology that can act barrier to understanding and reduce the number of persons who are 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.



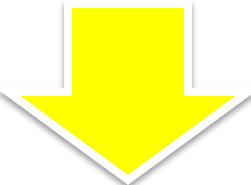


Leaves alternate proximally, opposite and ultimately decussate distally, 6–16 × 4–13 cm; petiole ca. as long as blade, winged, base clasping, basal lobes stipulate, growing as extensions of wings, less than 1 mm wide; blade 5–7-veined, ovate, glabrous, base typically sagittate, margins entire, apex acute 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



#### >*Dioscorea alata* (matK) gene, partial

```
ATTTAAATTATGTGTCAGATATTAATACCCATCCCCTGGAAATCCTGGTCAAATACTCAATGCTGGACTCAAGATGTTCTT  
TGCATTTATTGCGATTCTTCCACGAATATCATAATTGAAT AGTTTCATTACTCCGAAAAAACCTATTTACGTGATTCAATTCAAAGAAA  
ATAAAAGATTTTCGAT TCCTATATAATTCTTATGTATTGAATGTGAATTGTATTAGTTTTTCTATAAGCAATCCTTTACGATCAA  
GGTCCTCTGGAGTCTTCTTGAGCGAACACATTCTATGGAAAATGGGCATTTTAGTAGTGTGTTGTAATTATTCAGAAGACCCAATG  
GTTCTCAAAGATCCTTCTGCATTATGTCGATATC AAGGAAAAGCAATTCTGGTCAAAGGGAACTCGTCTTGTATGAGGAAATGGAGA  
TCTTACCTTGTCCATTGGCAATATTATTTCAATTGGTCTCATCCGCATAGGATTATCAAACCAATTCAAATTATCCTCTGTTTC  
TGGGTTATCTTCAAATGTACTAATAAAATTCCGTGGTAAGGAGTCAAATGTAGAAAATTCTTTGAATAGATACTCTACTAAGAAATT  
TGATACCAGAGTTCAGTTATTGCTTATCG ATCATTGTCTAAAGCGAAATTGTACCGTATCGGGCATCCTATTAGTAAGTCATATGGA  
CAAATTTCAGATTGGATATTATCGATTGGGGATATGTAGAA
```



**Complex and  
Somewhat  
objective**

**Simple (A,T,G, or  
C) and more  
Reliably objective**





## Choosing a DNA barcode

There are many criteria that go in to selecting an appropriate region that can serve as a DNA barcode.

Three of them include:

- Universality
- Robustness
- Discrimination

Why are these three criteria important?





## Discrimination

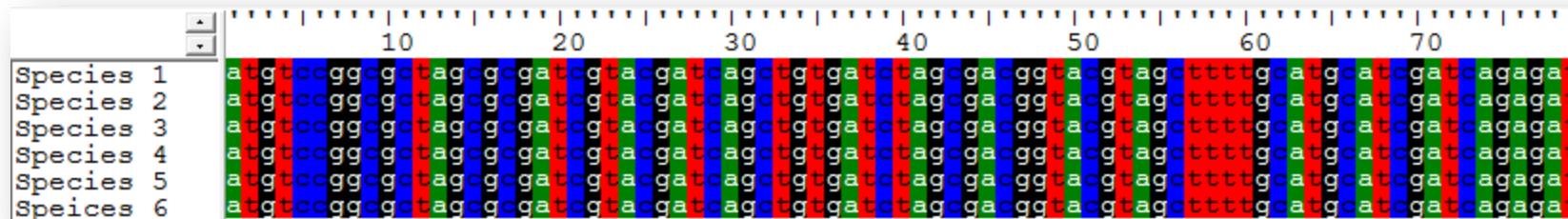
Barcode regions must be different for each species. Ideally you are looking for a single DNA locus which differs in each species.

### Oppositional Goals:

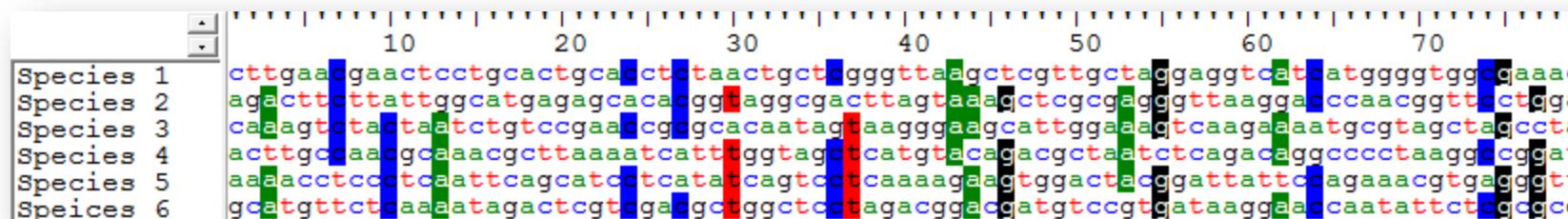
- Each loci must be different for each species
- Although loci must be different, they must be similar enough that they can be amplified by PCR, aligned and compared



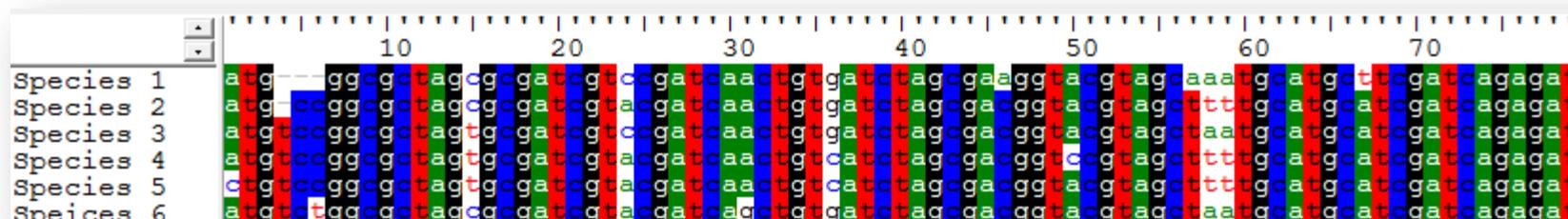
**Fail: Sequence is completely conserved, good for PCR, but uninformative as barcode**



**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





## Universality

Since barcoding protocols (typically) amplify a region of DNA by PCR, you need primers that will amplify consistently.

- Once you have a candidate locus (loci) that seem discriminatory, do these loci (possibly genes, but possibly non-coding DNA) exist in virtually all of the species you wish to barcode?
- Will you be able to find PCR primers that can amplify across many species, despite mismatches?





## Robustness

Since barcoding protocols (typically) amplify a region of DNA by PCR, also need to select a locus that amplifies reliably, and sequences well.

- PCR is very sensitive to the chemistry involved (types of enzymes, concentration of reagents, cycling parameters, etc.)
- The amplified PCR product must also be sequenced. Sanger sequencing is sensitive to highly repetitive DNA.



# DNA Barcoding Plants vs. Animals



Finding a DNA locus that possesses all of these qualities(Discrimination, Universality, Robustness) was relatively easy in animals.

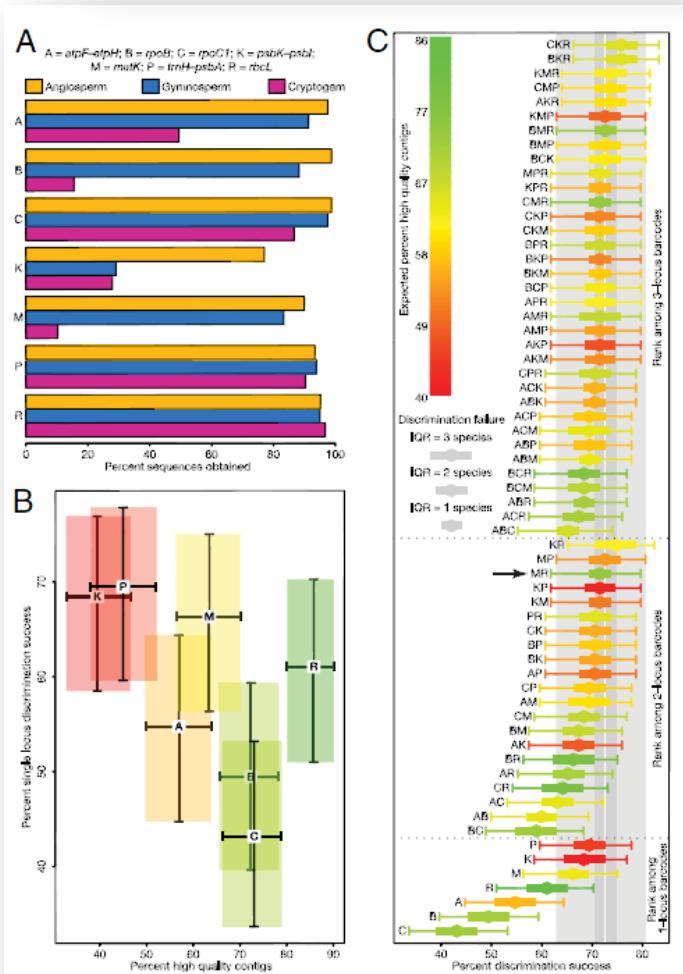
The animal barcode of choice  
Is the mitochondrial gene  
cytochrome c oxidase I (COI).



# A DNA barcode for land plants

CBOL Plant Working Group<sup>1</sup>

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



Based on recommendations by a barcoding consortium (Consortium for the Barcode of Life, plant working group) the chloroplast genes *rbcL* and *matK* come very close to being ideal candidates for universal plant barcodes.

Like any barcode loci that could be chosen, there will always be a possibility of failure to make a reasonably definitive identification of a particular specimen.

