

Species Identification and Barcoding

Brendan Reid

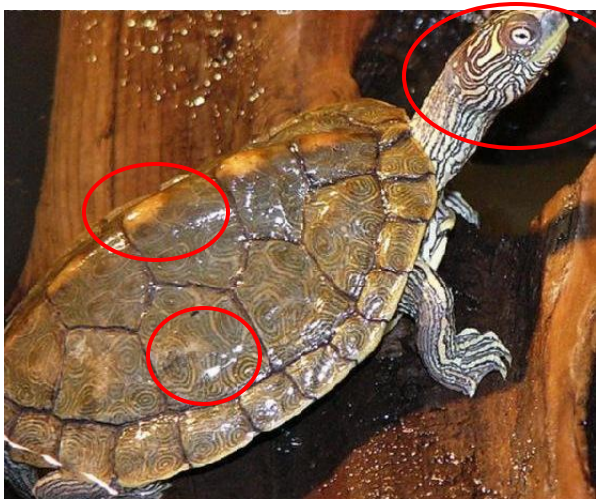
Wildlife Conservation Genetics

February 9th, 2010

Why do we need a genetic method of species identification?



Black-knobbed Map Turtle
(*Graptemys nigrinoda*)



Cagle's Map Turtle
(*Graptemys caglei*)

Conventional Species Identification/ Taxonomy

- Based on differences in morphology or other easily observable characteristics
 - Linnaean taxonomy: organisms conform to “types”
 - Dichotomous keys: if/then statements based on morphology
- Often reliant on specialized knowledge

Conventional species identification fails when...

- Morphology is misleading
 - Mimicry
 - Convergence
 - Cryptic species
 - Closely related, morphologically identical species with distinct habitats or ecological roles
 - Morphological differences may only appear at particular life stages or in one gender
 - Phenotypic plasticity
 - Genetic variability

Conventional species identification fails when...

- Whole organism is not available
 - Hair, feathers, scales
 - Meat, bones, medicinals in the wildlife trade
 - Feces/ stomach contents
 - Ambient DNA
- Expert knowledge is not available
 - Often costly/time-intensive
 - As more species are identified, more and more taxonomists are needed and knowledge becomes more specialized

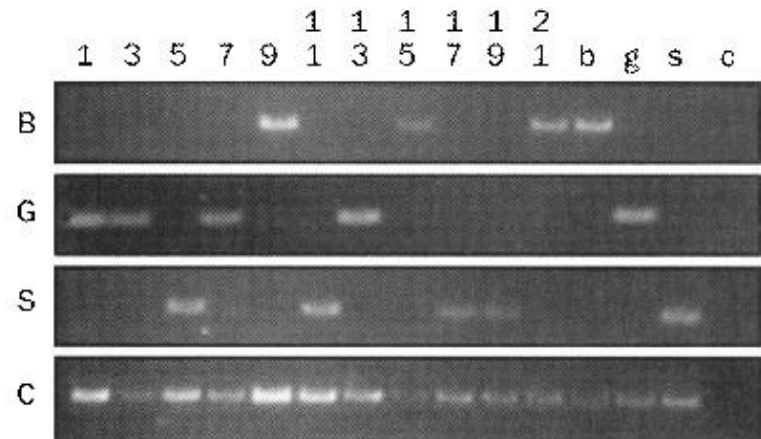
Conservation Case Study: Caviar



- Caviar = eggs of Eurasian sturgeons (Acipenseridae)
- Several species of sturgeon are overharvested for eggs, and several others are threatened by habitat loss
- Caviar dealers (“experts”) diagnose using egg size, color, taste, smell, etc.

PCR Identification of Black Caviar (Desalle & Birstein 1996)

- Design primers that amplify mitochondrial sequences only from particular species
- Diagnosis: 20% of caviar sampled was misidentified
 - Three IUCN Red-Listed species were identified as commercial species



Problems with PCR tests

- Markers are species-specific (different marker required for each species to be identified)
- Need some prior knowledge of sequences in order to design species-specific primers

What about conventional conservation genetics markers?

- Could we use, say, microsats? AFLPs?
SNPs?

Does a universal species identification marker exist?

- Most markers used in conservation genetics are too variable
- Must be present and easily amplified in all species
- Must vary among species but be fixed or relatively invariable within species

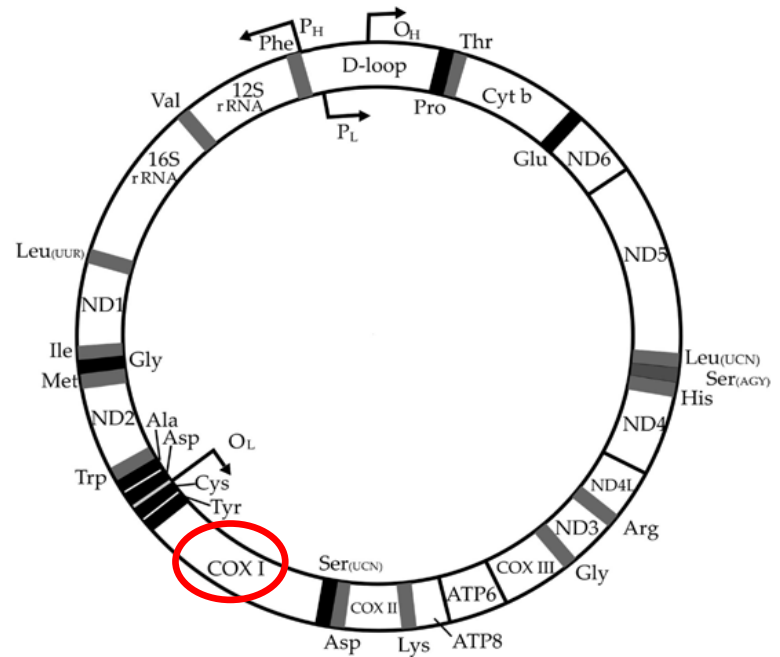
Biological identifications through DNA barcodes

**Paul D. N. Hebert^{*}, Alina Cywinska, Shelley L. Ball
and Jeremy R. deWaard**

- Used a single primer set to amplify a fragment of mitochondrial cytochrome oxidase I (COI) from representatives of several hundred animal species
- 96.4% of species successfully classified based on sequence variation

Why COI?

- Omnipresent
- Mitochondrial (one copy per organism, high copy number)
- Desirable amount of variation
 - Coding gene (selection against mutation)
 - $\Theta = N_e\mu$
 - Less equilibrium variability than a nuclear gene
 - Generally faster μ



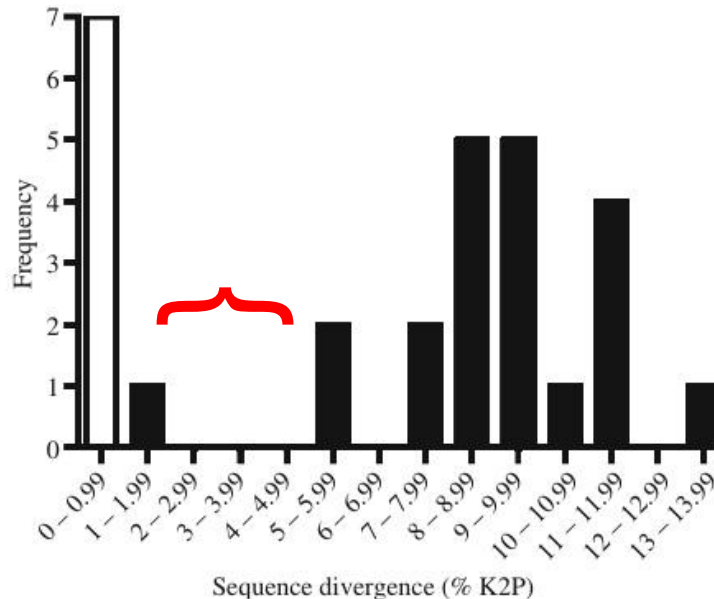
Barcoding Other Groups

- Fungi: COI not variable enough
 - Use ribosomal RNA internal transcribed spacer (ITS) region
- Plants: COI too variable
 - Use two chloroplast genes (rbcL and matK)

Barcoding Methodologies

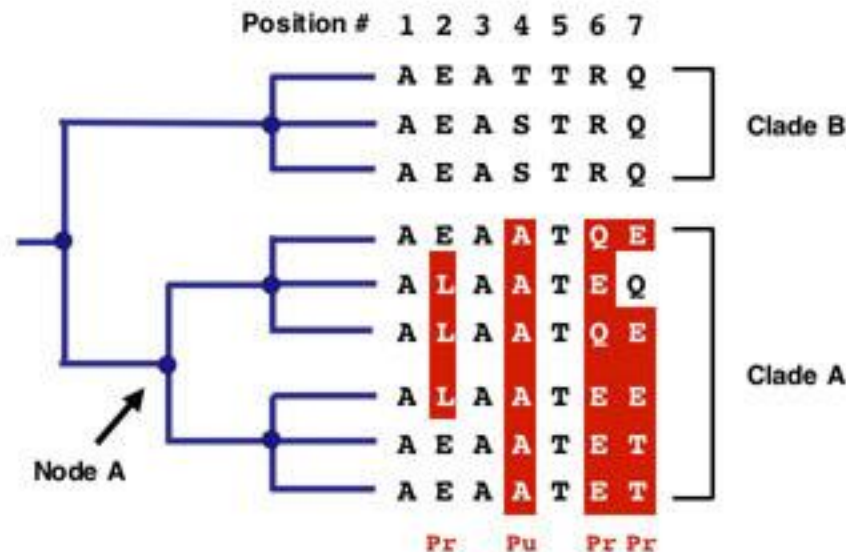
- Distance-based barcoding: assess Kimura 2-parameter (K2P) distances between individuals of the same species and between different species
 - Is there a definite “barcode gap”? Are individuals of the same species less than 2-3% different from one another, and are individuals of different species more than 2-3% different?

Black bars = interspecific divergences
White bars = intraspecific divergences



Barcoding Methodologies

- Character-based barcoding: find identifying single nucleotides (simple characters) or sets of nucleotides (compound characters)
 - Are there fixed differences that can be used to establish species identity?



Barcoding Infrastructure

- Sequencing initiatives
 - Taxon-specific (FISH-BOL, Bee-BOL, etc.)
 - Area-specific (Polar-BOL, etc.)
- The Barcode of Life Database (BOLD;
<http://barcodinglife.org>)
 - Searchable online sequence repository

The Big Barcoding Debate: Species Identification vs. Species Discovery

- Species identification: differentiating between well-characterized entities using COI sequences
- Species discovery: Designating new species based on COI differences between hard-to-distinguish groups

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

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

- *A. fulgerator* previously assumed to be a single, generalist species
- Barcoding indicated multiple distinct clusters of COI sequences within individuals identified as *A. fulgerator*



One adult morphospecies



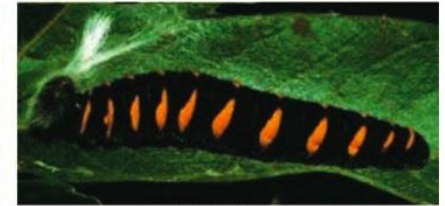
TRIGO



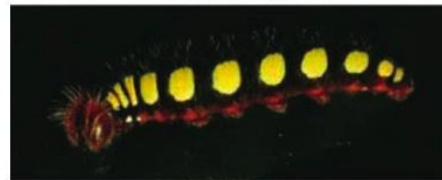
CELT



LONCHO



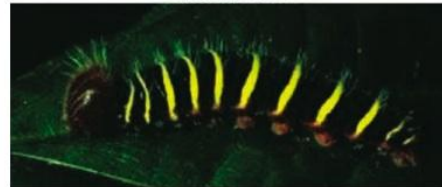
INGCUP



LOHAMP



HIHAMP



BYTTNER



FABOV



YESENN



SENNOV

10 caterpillars with different host plants

Critiques of Barcoding (Rubinoff 2006)

- Use of mtDNA as sole data source is problematic
 - Saturation and homoplasy
 - Heteroplasmy (multiple mitochondrial lines in one organism)
 - Nuclear mitochondrial pseudogenes (“numts”; mitochondrial genes inserted into nuclear genome)
 - Mismatches between nuclear and mitochondrial inheritance
 - Sex-specific dispersal patterns
 - Hybridization
 - Incomplete lineage sorting

Critiques of Barcoding (Rubinoff 2006)

- Species delineation
 - “Arbitrary” 2-3% cutoffs lack biological/evolutionary meaning
 - Recently diverged species may still have very similar COI sequences
 - COI divergence may not reflect separate evolutionary trajectories
 - No means of reconciling barcode-defined species with other species definitions

**Is species discovery through
barcoding valuable in conservation?**

Group-specific barcoding studies

- Allows estimation of how effective barcode-based identification will be for certain taxa or assemblages
 - E.g. North American birds (Hebert 2004)
- Evaluate the performance of barcoding methodology in distinguishing known species

Distance-based and Character-based Approaches to Barcoding Turtles

Brendan Reid¹, Eugenia Naro-Maciel¹,
Rob DeSalle¹, William McCord²,
George Amato¹, and Minh Le¹

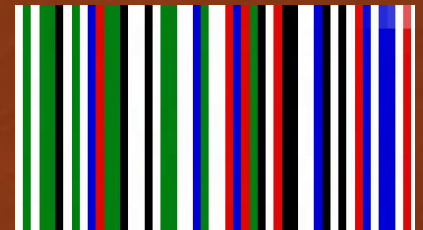
¹ American Museum of Natural History, New York, NY

² East Fishkill Animal Hospital, East Fishkill, NY



Considerations for Barcoding Turtles

- Relatively few species (many of which, however, are quite rare)
- Many species are capable of hybridizing with sister species or even more distant relations
- Intrinsically slower rate of mtDNA evolution than most other animal taxa (Avice 1993)
- Nuclear insertions of mitochondrial genes have been found in several species



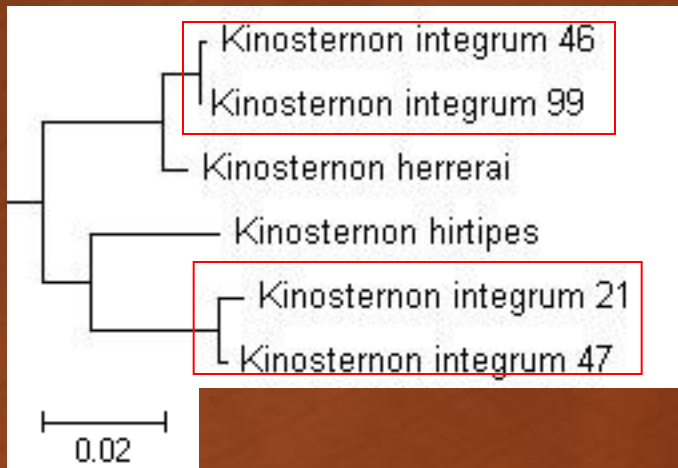
Sampling

- Blood/tissue from 183 species obtained and sequenced for COI (650 bp) at the American Museum of Natural History's Sackler Center for Comparative Genomics
- Sequences from an additional 36 species available on BOLD
- Final data set represents all 14 turtle families (67% of species diversity)
- Sample size generally low (n=1 for 153 species)



Results: Distance-based barcoding

- Intraspecific divergences $>2\%$ in 14 of 66 species where multiple individuals were sampled

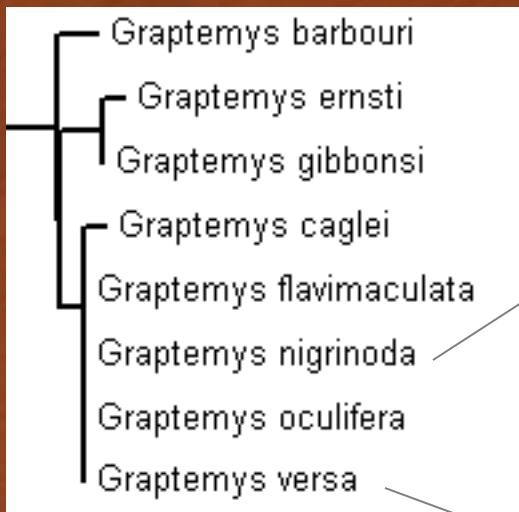


Kinosternon integrum
(Mexican mud turtle)



Results: Distance-based barcoding

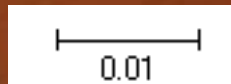
- Divergences of $<2\%$ between congeners for 48 species



Black-knobbed Map Turtle



Texas Map Turtle



Results: Character-based barcoding

- CAOS identified 69 nucleotide positions that constitute a compound character for discriminating turtle species
- Only 17 species (in which an individual was identical to an individual of another species for the barcode region) could not be identified using this character suite



	3	6	10	45	93	94	105	132	138	168	186	198	205	207	210	222	228	255	267	318	321	333	339	354	360	366	378	393	414	423	435	452	477	505	513	519	531	552	559	588	603	612	624	636	653					
<i>Actinemys marmorata</i>	T	T	T	A	A	A	G	T	C	C	C	T	C	A	A	A	A	T	A	A	G	C	T	A	A	A	T	A	G	A	A	T	T	A	A	A	C	T	T	A	A	A	C	T	A	C	A	C		
<i>Clemmys guttata</i>	T	T	C	G	G	G	A	T	C	C	C	T	G	G	A	G	G	T	A	A	A	T	T	A	A	A	T	A	G	A	A	C	T	T	A	A	A	T	A	C	A	A	T	A	C	A	C	T	A	T
<i>Emydoidea blandingii</i>	T	T	C	A	A	A	G	T	C	C	C	T	G	A	A	A	A	T	A	A	G	C	T	A	A	A	T	A	A	A	A	A	C	T	C	C	A	T	A	T	A	C	A	C	T	A	C			
<i>Emys orbicularis</i>	T	T	C	G	A	G	A	T	C	C	T	T	G	A	A	A	T	A	A	G	T	T	A	A	A	T	A	A	A	A	A	C	T	C	C	A	C	T	A	T	A	T	A	C	T	A	C			
<i>Glyptemys insculpta</i>	T	C	A	A	G	G	T	C	C	T	C	G	A	A	G	A	C	A	A	A	C	T	A	A	T	A	G	A	A	C	T	C	C	A	A	T	A	T	A	T	A	C	T	A	C	T	A	C		
<i>Glyptemys muhlenbergii</i>	T	C	C	A	A	G	G	T	C	C	T	C	G	A	A	G	A	C	A	A	A	C	T	A	A	G	T	A	G	A	A	C	T	C	C	A	A	T	A	T	A	C	C	A	C	T	A	C		
<i>Graptemys barbouri</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	T	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Graptemys caglei</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	C	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Graptemys ernsti</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Graptemys gibbonsi</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Graptemys flavimaculata</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Graptemys nigrinoda</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Graptemys oculifera</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Graptemys versa</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C		
<i>Malaclemys terrapin</i>	C	C	T	A	G	G	G	C	C	T	T	T	G	C	C	A	A	T	A	A	G	C	T	A	A	A	T	A	A	A	A	C	T	C	G	A	C	A	A	A	C	T	A	C	T	A	C			
<i>Pseudemys alabamensis</i>	T	C	T	A	T	A	A	C	C	T	T	T	G	A	A	A	A	C	A	A	C	C	T	A	A	A	T	A	A	A	A	T	T	C	C	A	A	A	C	T	A	A	A	C	T	A	C			
<i>Pseudemys gorzugi</i>	T	C	T	A	T	A	A	C	C	T	T	T	G	A	A	A	A	C	A	A	C	C	T	A	A	A	T	A	A	A	A	T	T	C	C	A	A	A	C	T	A	A	A	C	T	A	C			
<i>Pseudemys rubriventris</i>	T	C	T	A	T	A	A	C	C	T	T	T	G	A	A	A	A	C	A	A	C	C	T	A	A	A	T	A	A	A	A	T	T	C	C	A	A	A	C	T	A	A	A	C	T	A	C			

Results: Character-based barcoding

- Character-based barcode system better at identifying species both when intraspecific distances are large and when interspecific distances are small
- Characters may not be fixed; however, identification by matching characters instead of by similarity is more conservative and will reduce false positive and negative IDs



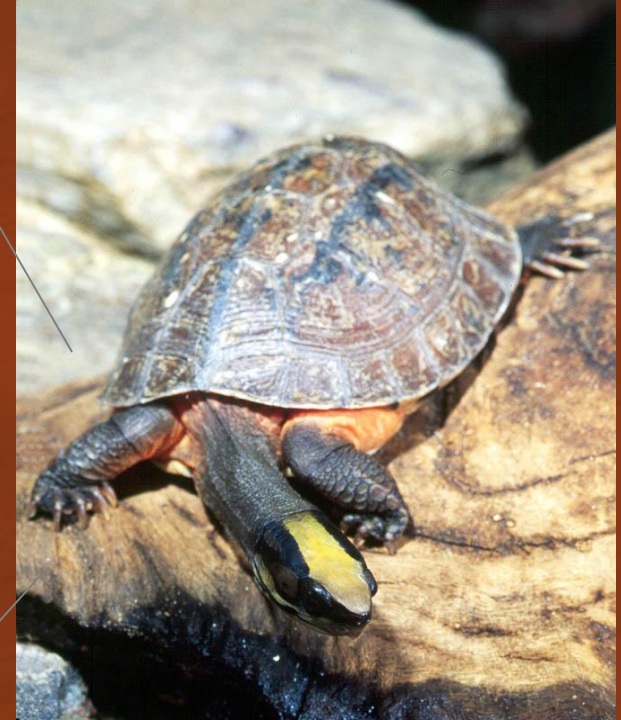
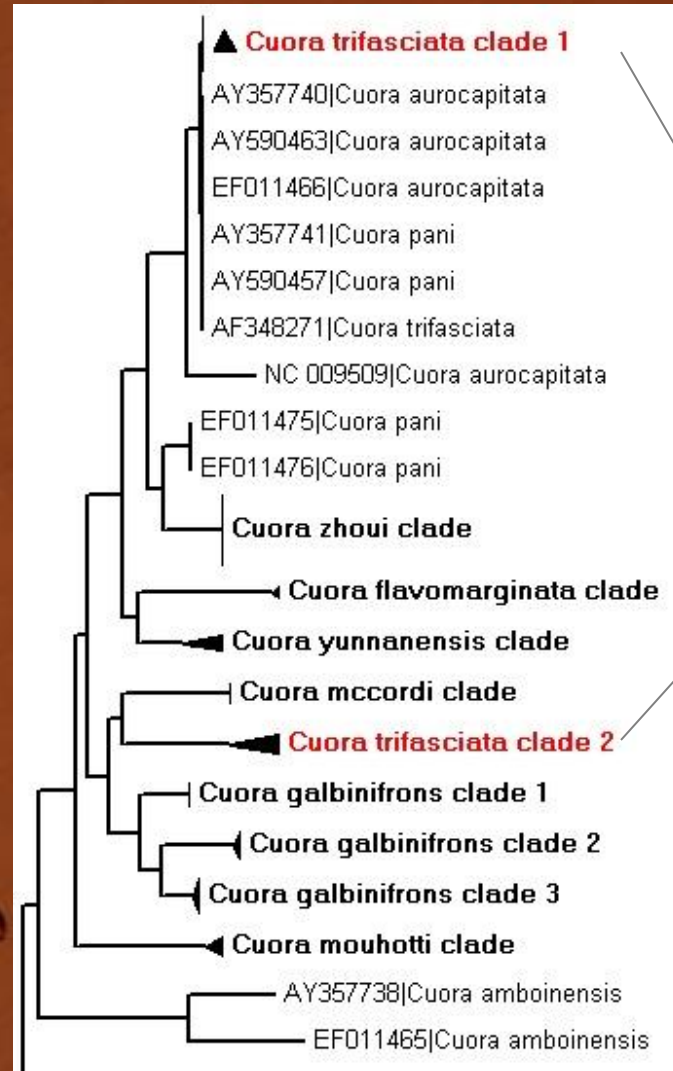
	3	6	10	45	93	94	105	132	138	168	186	198	205	207	210	222	228	255	267	318	321	333	339	354	360	366	378	393	414	423	435	452	472	505	513	531	551	552	565	588	603	612	624	636	653					
<i>Actinemys marmorata</i>	T	T	T	A	A	G	T	C	C	C	T	G	A	A	A	A	T	A	A	G	C	T	A	A	A	T	A	G	A	A	T	T	A	C	A	A	T	A	T	A	C	C	A	C	T	T				
<i>Clemmys guttata</i>	T	T	C	G	G	G	A	T	C	C	C	T	G	G	A	G	G	T	A	A	A	T	T	A	A	A	T	A	G	A	A	C	T	C	T	A	A	T	A	C	A	C	T	A	T	T				
<i>Emydoidea blandingii</i>	T	T	C	A	A	G	T	C	C	C	T	G	A	A	A	T	A	A	G	C	T	A	A	A	T	A	A	A	A	C	T	C	C	A	T	A	T	A	C	T	A	C	T	A	C	T	T			
<i>Emys orbicularis</i>	T	T	C	G	A	G	A	T	C	C	T	T	G	A	A	A	T	A	A	G	T	T	A	A	A	T	A	A	A	C	T	C	C	A	C	T	A	T	A	C	T	A	C	T	A	C	T	T		
<i>Glyptemys insculpta</i>	T	A	C	A	A	G	G	T	C	C	T	C	G	A	A	G	A	C	A	A	A	C	T	A	A	T	A	G	A	A	C	T	C	C	A	A	T	A	T	A	C	T	A	C	T	A	C	T		
<i>Glyptemys muhlenbergii</i>	T	C	C	A	A	G	G	T	C	C	T	C	G	A	A	G	A	C	A	A	A	C	T	A	A	G	T	A	G	A	A	C	T	C	C	A	A	T	A	T	A	C	C	A	C	T	A	C	T	
<i>Graptemys barbouri</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	T	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T	
<i>Graptemys caglei</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	C	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T	
<i>Graptemys ernsti</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T	
<i>Graptemys gibbonsi</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T	
<i>Graptemys flavimaculata</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T	
<i>Graptemys nigrinoda</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	A	A	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T
<i>Graptemys oculifera</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T	
<i>Graptemys versa</i>	C	C	T	A	G	A	G	C	C	T	T	T	G	G	A	G	A	T	G	A	C	C	C	A	A	A	T	A	A	A	A	C	T	C	T	G	A	C	A	A	A	C	T	A	C	T	A	C	T	
<i>Malaclemys terrapin</i>	C	C	T	A	G	G	G	C	C	T	T	T	G	C	C	A	A	T	A	A	G	C	T	A	A	A	T	A	A	A	A	C	T	C	G	A	C	A	A	A	C	T	A	C	T	A	C	T		
<i>Pseudemys alabamensis</i>	T	C	T	A	T	A	A	C	C	T	T	T	G	A	A	A	A	C	A	A	C	C	T	A	A	A	T	A	A	A	A	T	T	C	C	A	A	A	C	T	A	C	T	A	C	T	A	C	T	
<i>Pseudemys gorzugi</i>	T	C	T	A	T	A	A	C	C	T	T	T	G	A	A	A	A	C	A	A	C	C	T	A	A	A	T	A	A	A	A	T	T	C	C	A	A	A	C	T	A	C	T	A	C	T	A	C	T	
<i>Pseudemys rubriventris</i>	T	C	T	A	T	A	A	C	C	T	T	T	G	A	A	A	A	C	A	A	C	C	T	A	A	A	T	A	A	A	A	T	T	C	C	A	A	A	C	T	A	C	T	A	C	T	A	C	T	

Crossing the barcode gap: cryptic diversity or introgression?

Cuora trifasciata
(Chinese three-striped box turtle):

Two mtDNA clades,
only one nuclear clade

(Spinks & Shaffer 2007)



Ecological/ Conservation Applications (Valentini 2009)

- Distinguishing individuals of ecologically important but morphologically similar species (e.g. nematodes)
- Establishing ranges for elusive or rare species using scat or hair samples
- Monitoring trade in endangered species and enforcing CITES regulations
- Identifying and excluding potential pathogens or invasive species (biosecurity)

Ecological/ Conservation Applications (Valentini 2009)

- Rapid biodiversity surveys with ecological implications, e.g.:
 - Stream insect larvae diversity is an indicator of water quality and disturbance (damselflies and stoneflies flourish in more pristine habitats, midges in more disturbed)
 - Larvae are very difficult to identify without expert knowledge, and can generally only be identified to the genus level
 - Barcoding allows for rapid identification by non-experts to the species level (Sweeney 2009)

Ecological/ Conservation Applications (Valentini 2009)

- Paleoecology
 - Identify changes in community composition and associated climatic factors
 - Assess past human impacts on flora and fauna (Willerslev 2009)
 - DNA from extinct North American megafauna extracted from permafrost
 - Sequence variation fitted to molecular clock models
 - Verdict: Species became extinct well after human colonization

Ecological/ Conservation Applications (Valentini 2009)

- Diet analysis
 - Herbivores: identification of food plants using chloroplast DNA can aid in reserve design
 - Barcoding can identify areas of diet overlap (and therefore competition), e.g. specificity of bat predation on insects (Clare 2009)