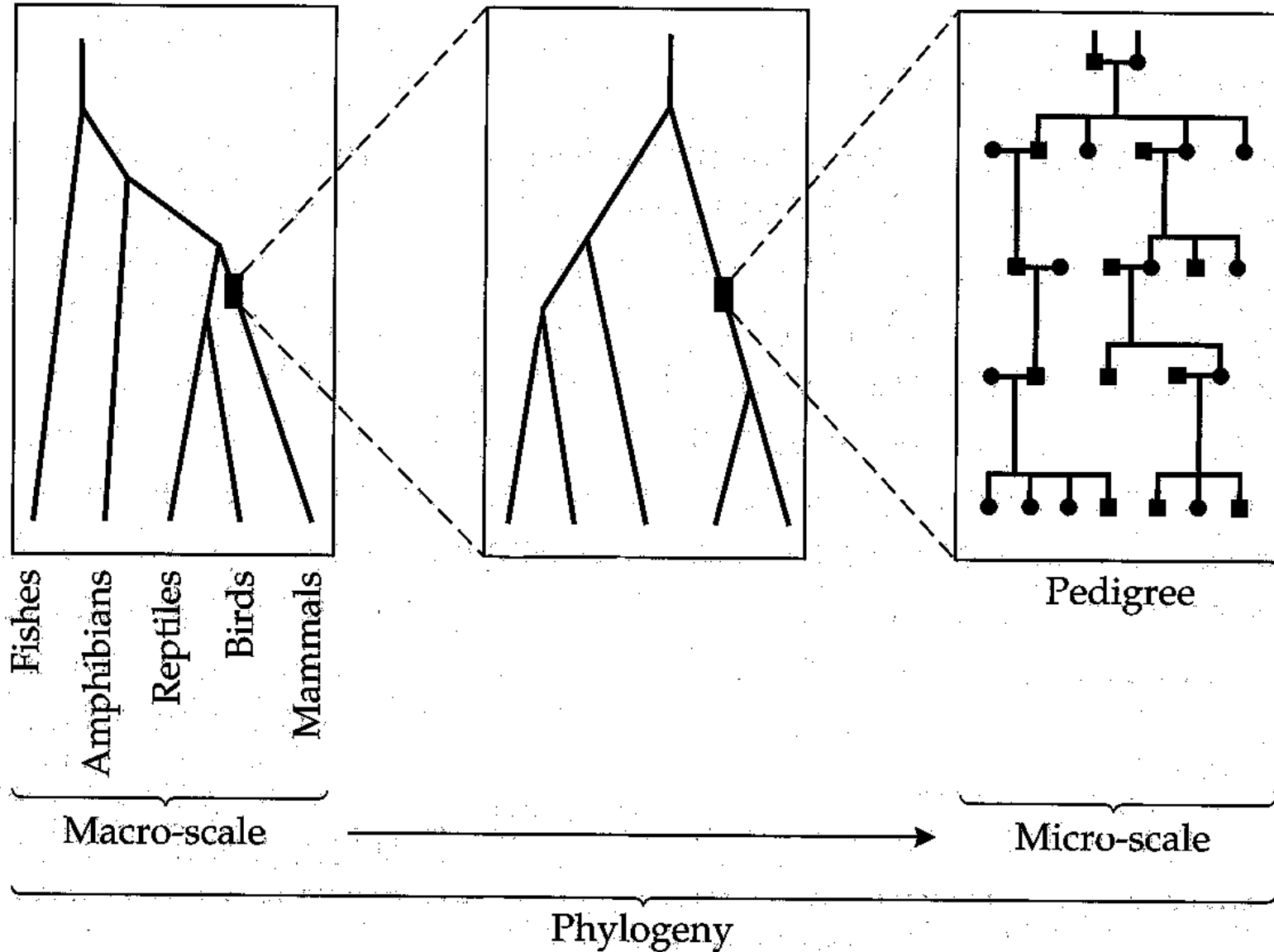


# Phylogenetics

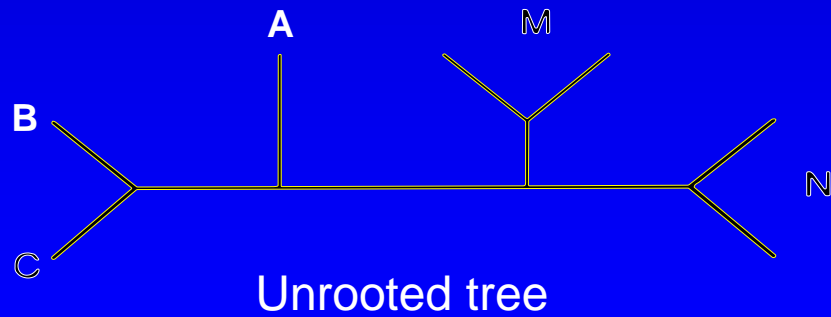
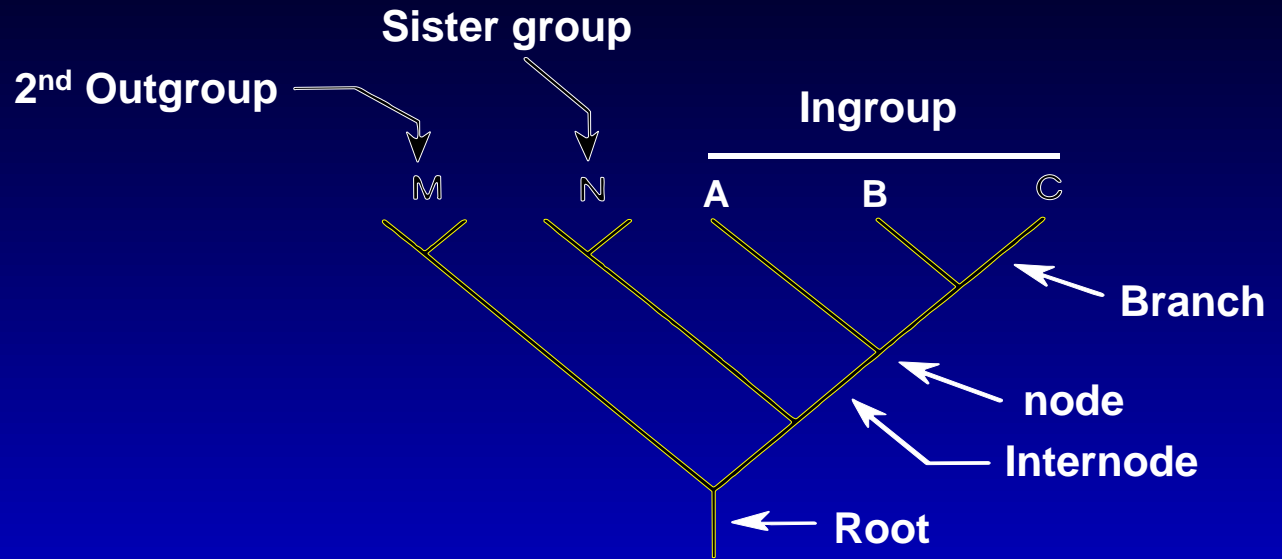
# The hierarchical nature of phylogenetic assessment

(after Avise et al. 1987)

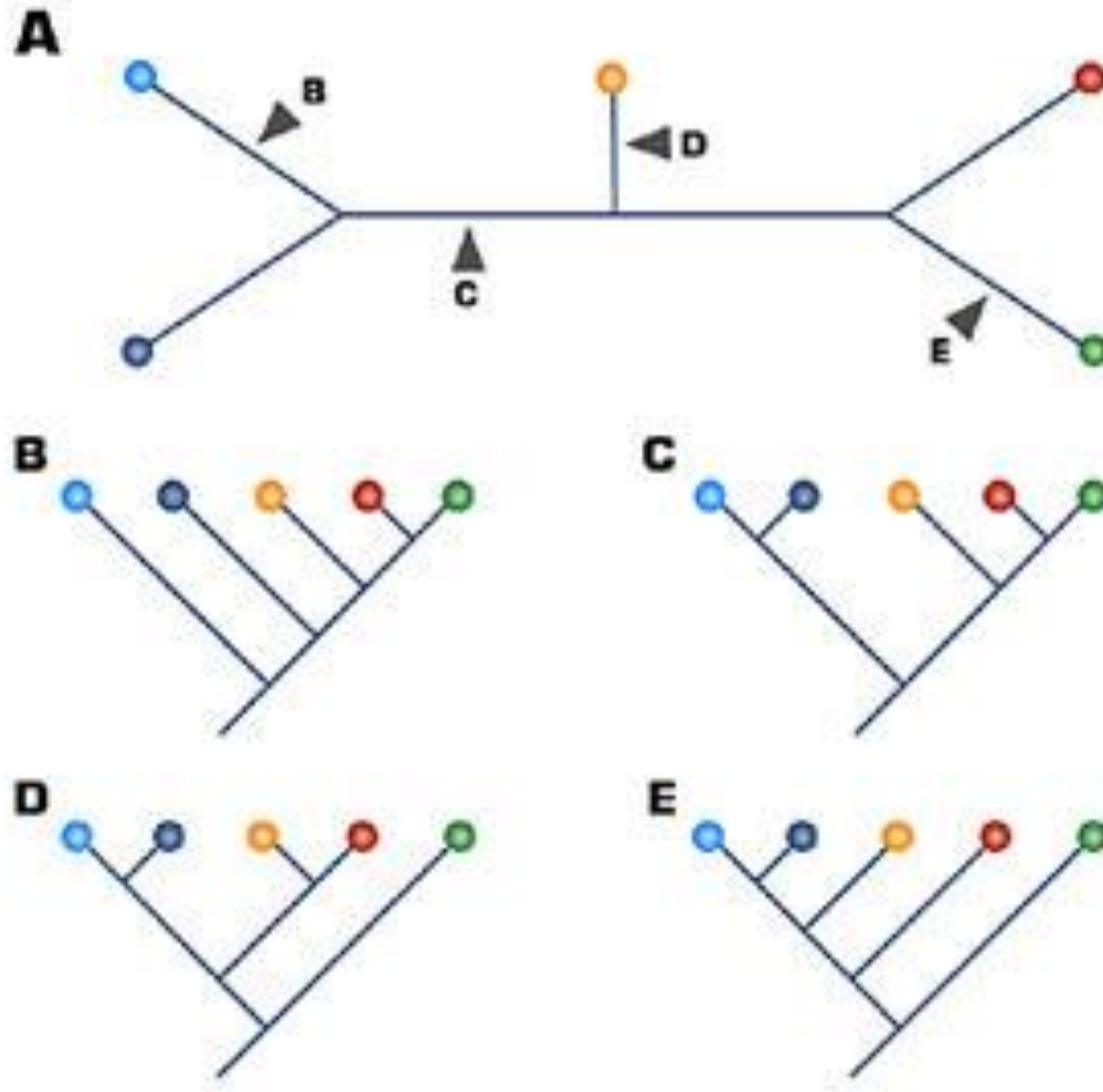


**PHYLOGENY** A hypothesis of evolutionary history and relationships.

“Phylogenetic inferences are premised on the inheritance of ancestral characteristics and on the existence of an evolutionary history defined by changes in these characteristics”  
(Swofford et al. 1996)



# Alternative Rooting



# Reconstructing Phylogenetic Trees

---

---

Numerous possible topologies

2 OTUs (operational taxonomic units) → 1 rooted tree

10 OTUs → 34,459,425 trees

Lack of variation in data → unresolved trees (polytomies)

## Other Challenges

### HOMOPLASY

Parallelism/convergence

Reversals (derived → ancestral, multiple hits)

# Methods for Phylogenetic Inference

## Distance methods

UPGMA

Least Squares

Minimum Evolution

Neighbor-Joining

Networks

Maximum Parsimony

Maximum Likelihood

Bayesian

Important Software: PAUP, Phylip, McClade,  
GeoDis, MrBayes

# Genetic Distance Methods

- **Concept: Mutations accumulate at a constant rate over time. The smaller the number of nucleotide differences the closer the genetic relationship between two taxa.**
- **This concept is applied in slightly different algorithms in the following methods:  
UPGMA, Least Squares, Minimum Evolution, Neighbor-Joining, Networks**



1.CAGTACG

2.GATTACG

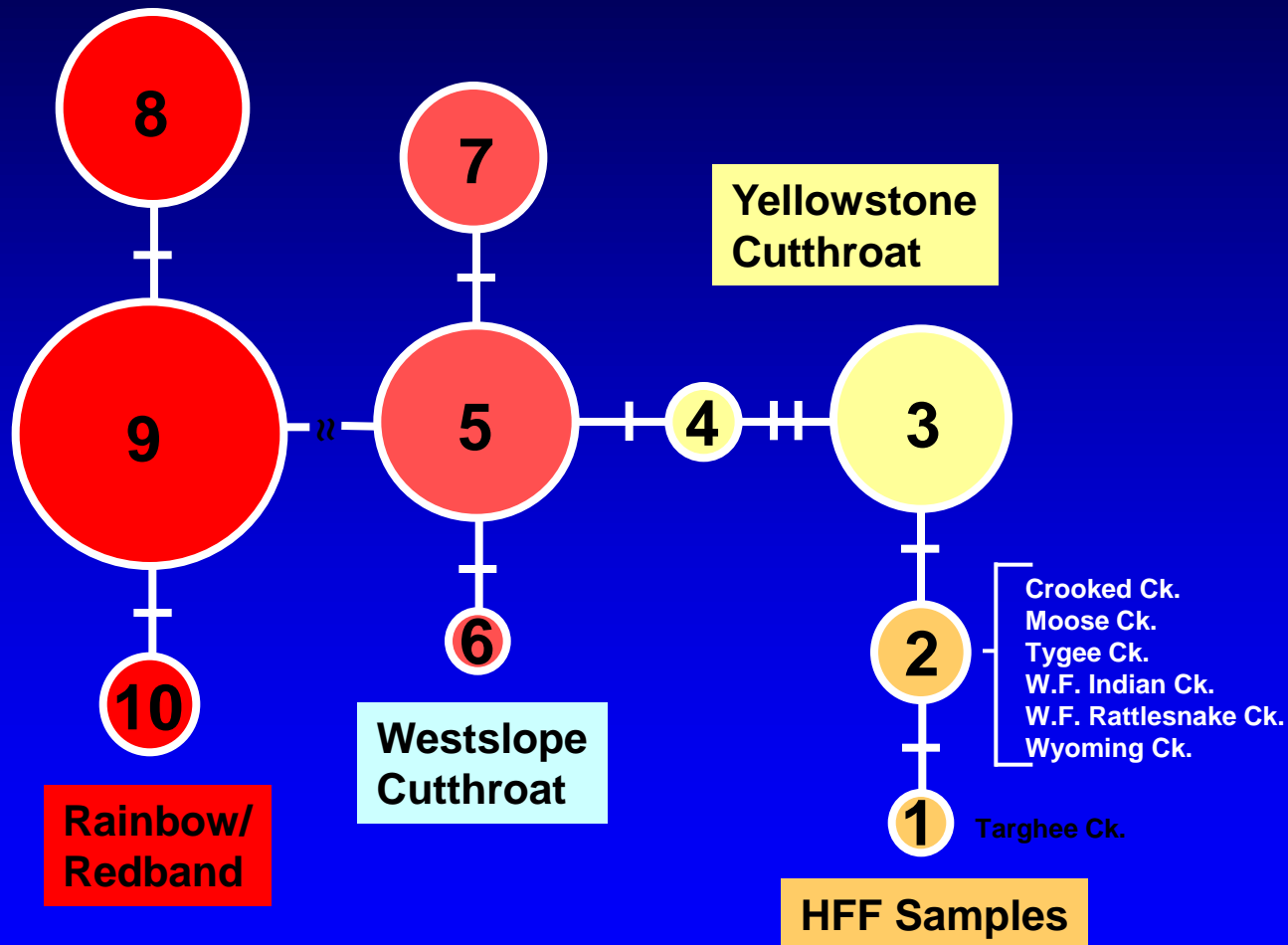
3.GATAACG

4.GAGTACC

5.GAGTCCC

Neighbor-Joining

# Simple Network with nucleotide changes



- **There are 5 predictions that can be used with Networks**
  - **Alleles found at the greatest frequency are the older alleles and these alleles are more likely to be internal to networks on ancestral in trees.**
  - **Older alleles should be more broadly distributed geographically**
  - **Haplotypes of the highest frequency will tend to have the most mutational connections in a network or tree**
  - **Haplotypes observed in a single individual are more likely to connect to common than rare haplotypes**
  - **Haplotypes observed in a single individual are more likely to connect to haplotypes in same population than different populations**

1.CAGTACG

2.GATTACG

3.GATAACG

4.GAGTACC

5.GAGTCCC

Network

# Parsimony

1. CAGTACG

2. GATTACG

3. GATAACG

4. GAGTACC

5. GAGTCCC

Based on principle of minimum evolution, choose the tree that requires smallest number of mutations to explain the data. Only uses shared derived traits (nucleotides) - synapomorphies.

# Maximum Likelihood and Bayesian

- Principle: Choose the tree that has the highest likelihood given the data and given a particular model of evolution.
- Models can include substitution rates, nucleotide frequencies, branch lengths.
- Bayesian methods will have a prior probability distribution
- Strengths - Statistically based and alternative hypotheses can be rigorously tested.
- Weaknesses: Computationally intensive. Slow (ML)

# Measuring Support for Parts of a Tree

**Bootstrap Analysis** -- Randomly creates a data matrix from original matrix (with replacement)...produces tree; repeats 100-1000 times; bootstrap values are the percentage of trees that had a particular clade

**Decay Analysis** - Examines how robust branches are in trees longer than the most parsimonious ones. A branch that collapses and becomes unresolved in a tree one step longer is not very robust

**Bayesian Posterior Probabilities** - % of trees that had a particular clade in a Bayesian ML analysis





# Chloroplast DNA

- Molecule of choice for plant phylogenetic studies
- 150 kb and ~120 genes
- Web-based databases with many primers

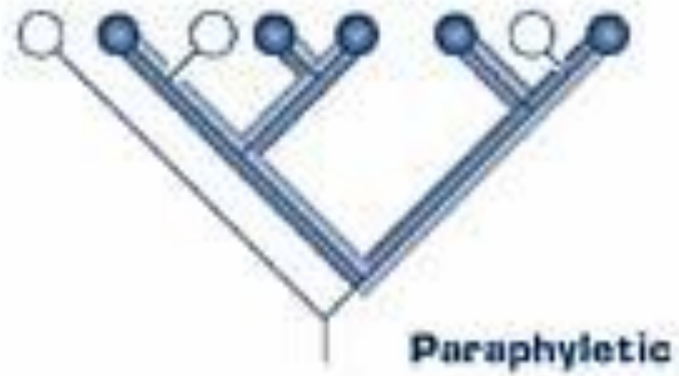


# Terminology - Concepts

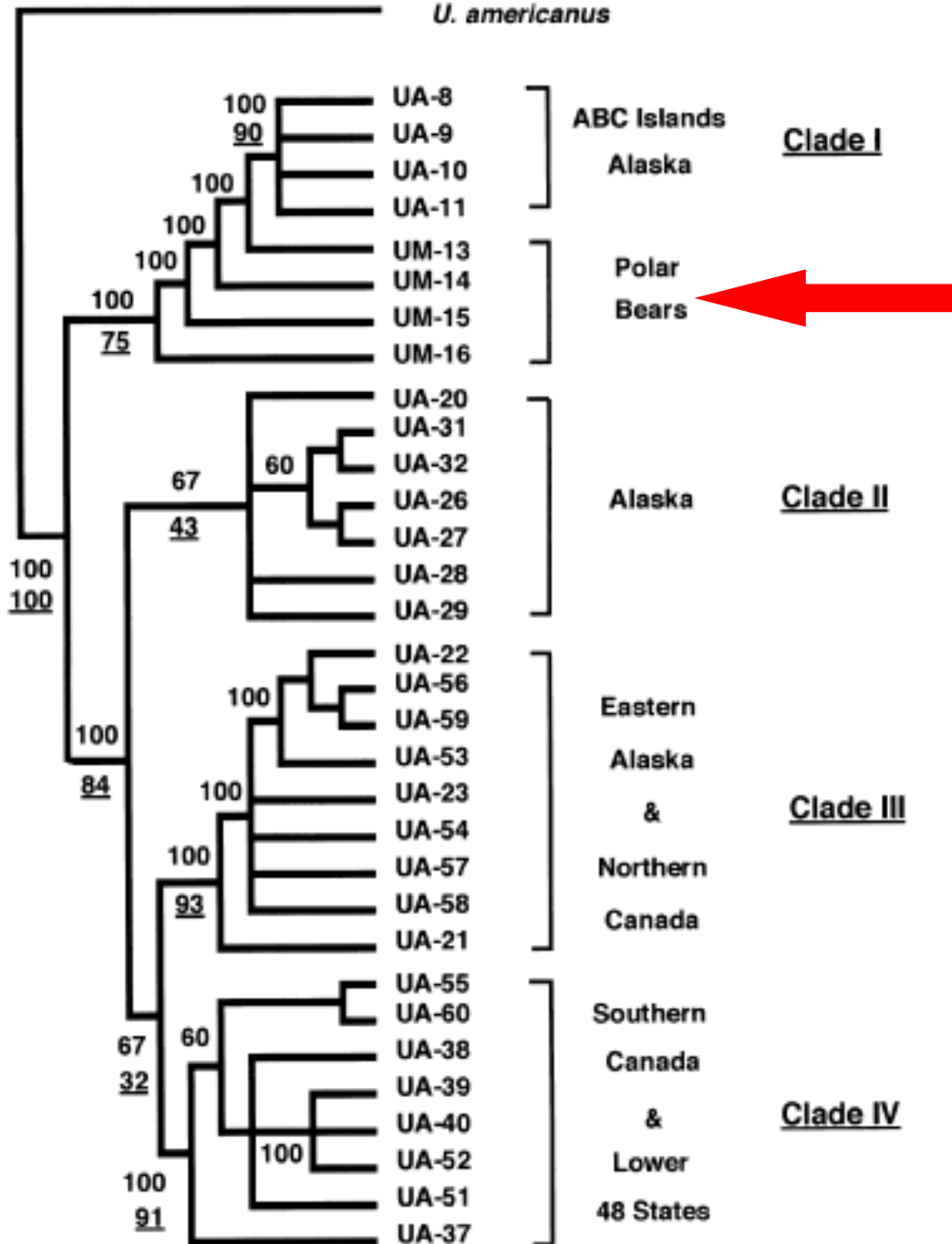
Monophyletic – A group containing ALL of the descendants of one ancestral species ... a CLADE

Paraphyletic – A group that does not contain all of the descendants of the most recent common ancestor

Polyphyletic – A group derived from two or more distinct ancestral taxa



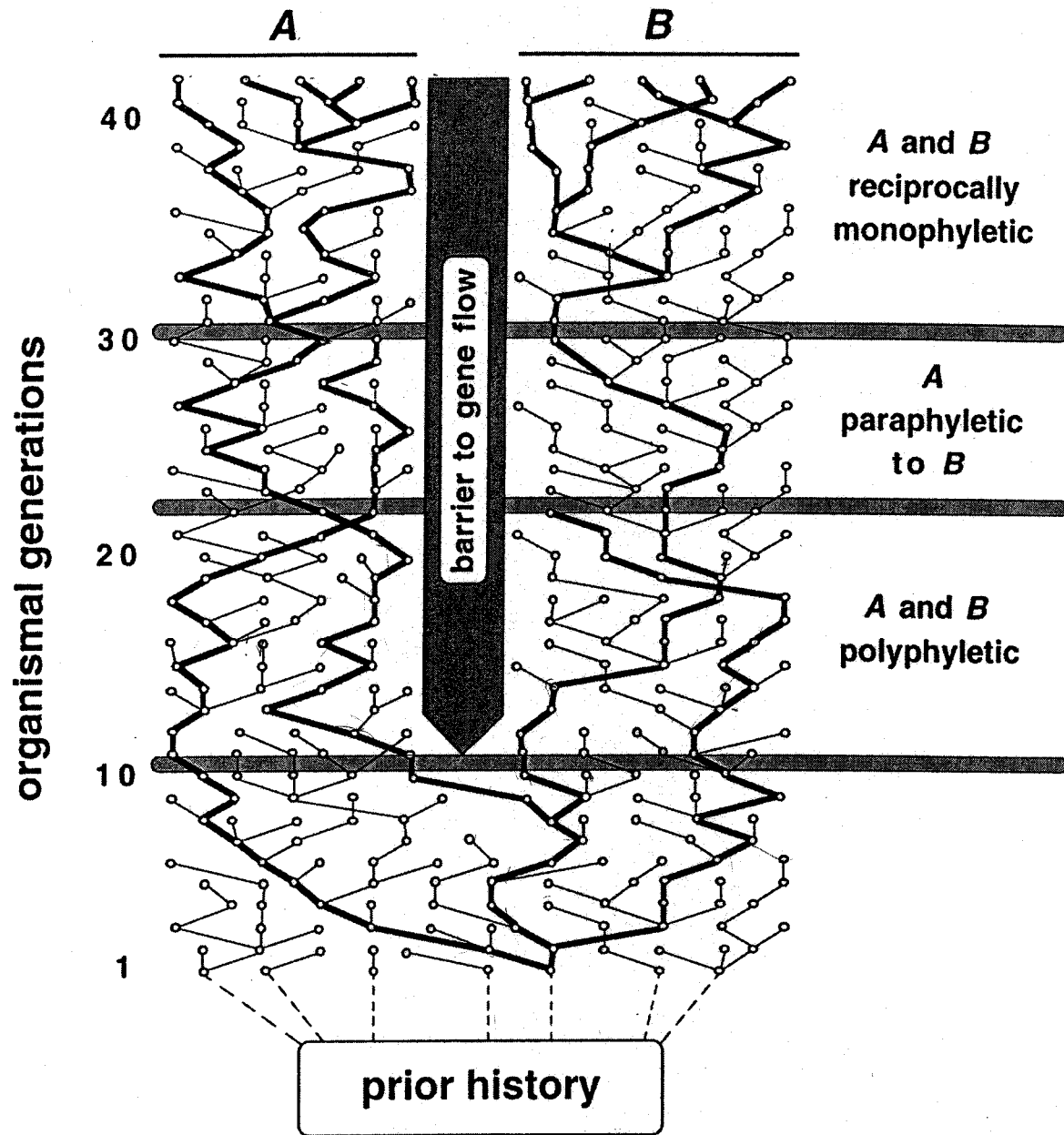
*U. americanus*



# Example

- Brown bears are Paraphyletic based on mtDNA

Waits et al. 1998



# Species Level Paraphyly and Polyphyly

Funk and Omland 2003 Ann Rev Ecol Evol Syst 34:397

- Reviewed 13 years of mtDNA literature to determine the prevalence of polyphyly

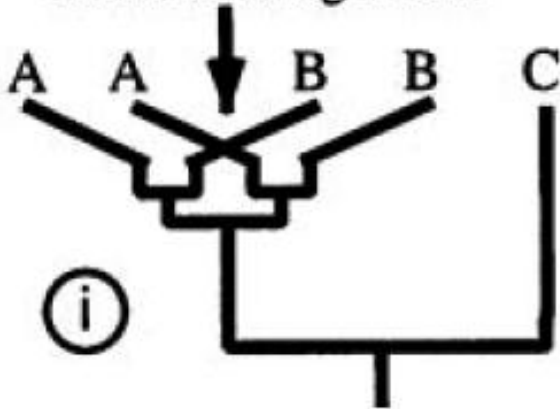
**TABLE 1** Results of the literature survey

Taxa	Number of:			Percent spp. polyphyletic <sup>a</sup>
	Studies	Genera	Spp.	
Mammals	139	102	469	17.0
Birds	74	87	331	16.7
Reptiles	56	45	147	22.4
Amphibians	35	26	137	21.3
Fishes	100	99	371	24.3
Arthropods	143	126	702	26.5
Other Invertebrates	37	41	162	38.6
<b>TOTAL</b>	<b>584</b>	<b>526</b>	<b>2319</b>	<b>23.1</b>

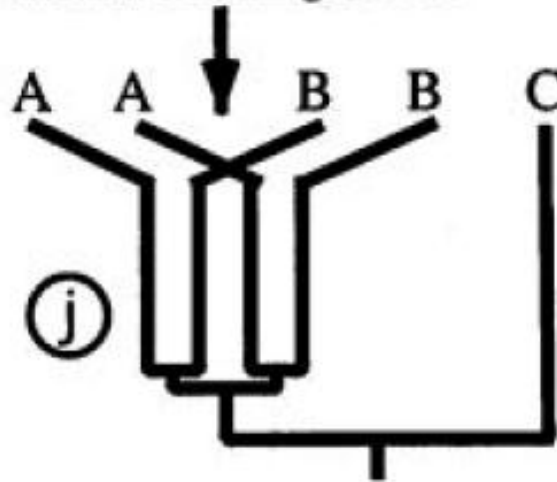
# Why?

## Incomplete lineage sorting

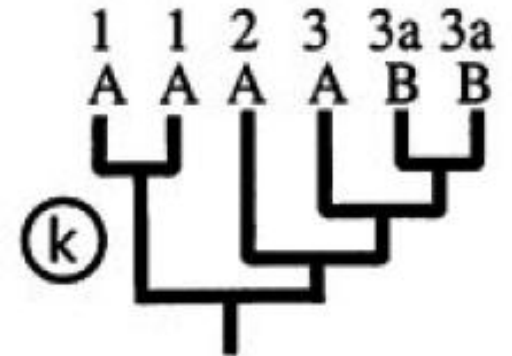
Incomplete sorting  
recent isolating barrier



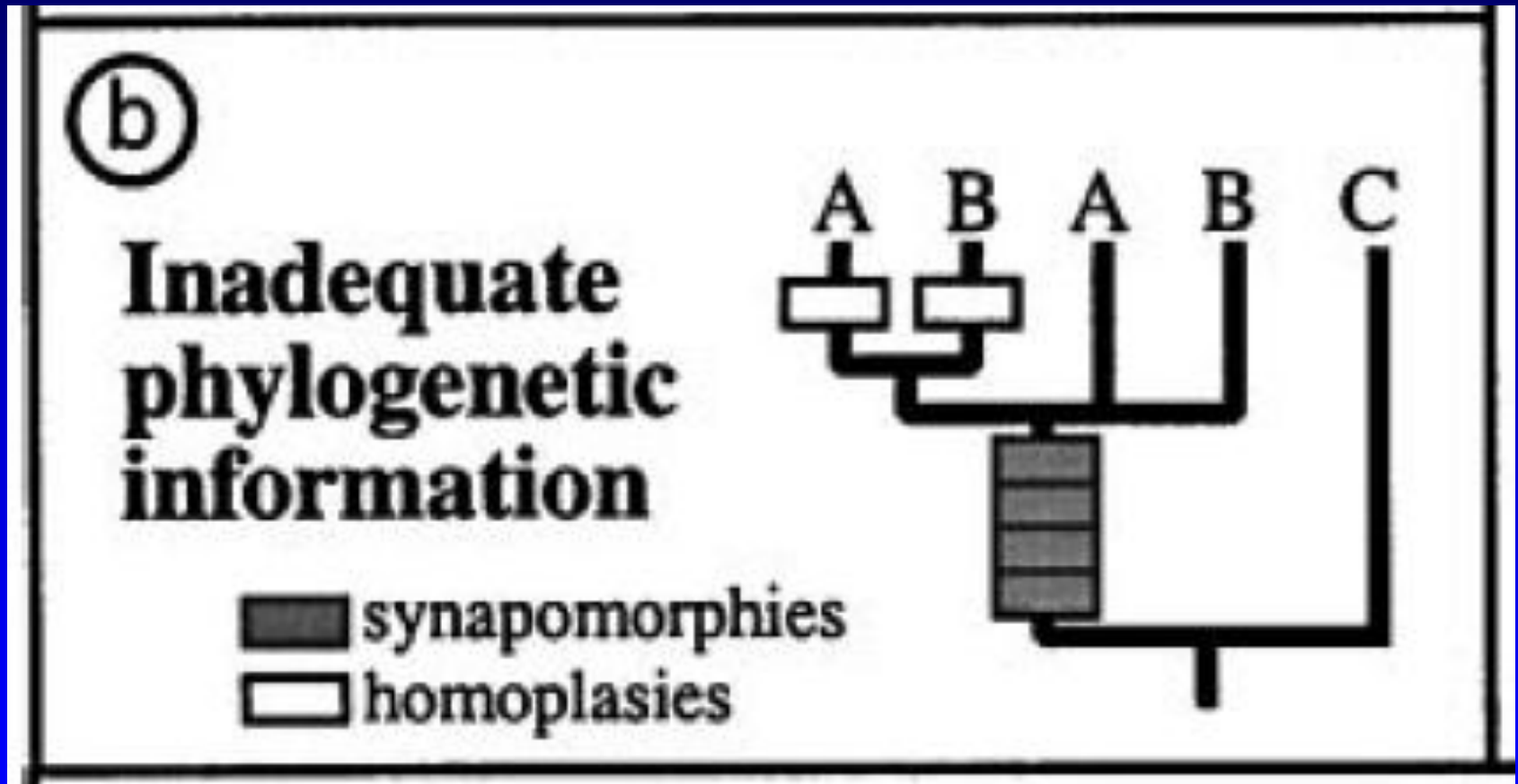
Balancing selection  
ancient isolating barrier



Peripheral isolates  
speciation



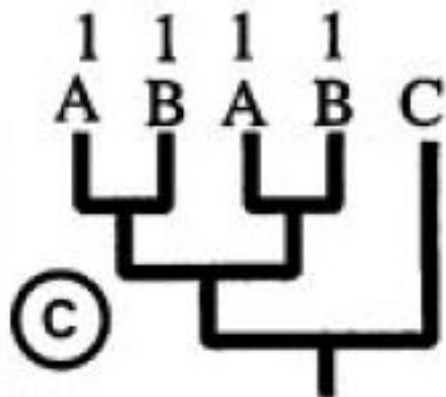
# Why?



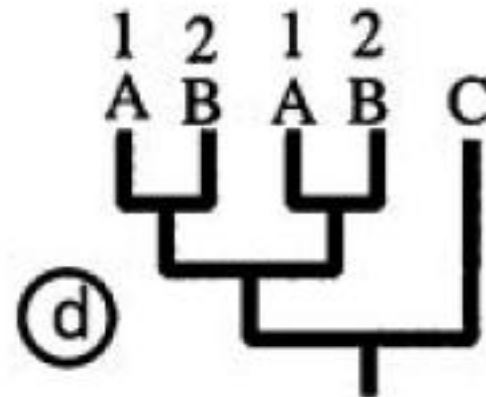


# Imperfect taxonomy

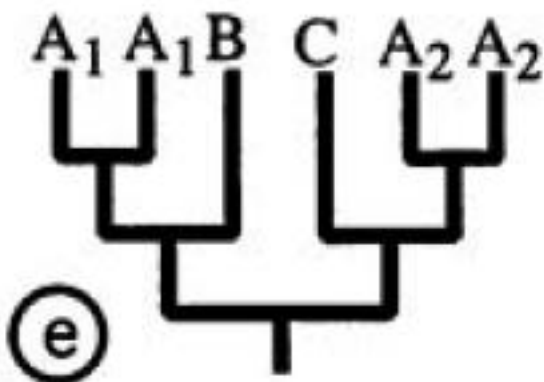
Polymorphism



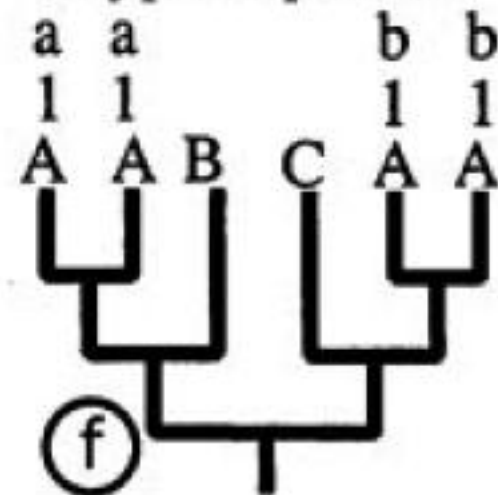
Geographic variation



Lumping of  
"infraspecific" taxa

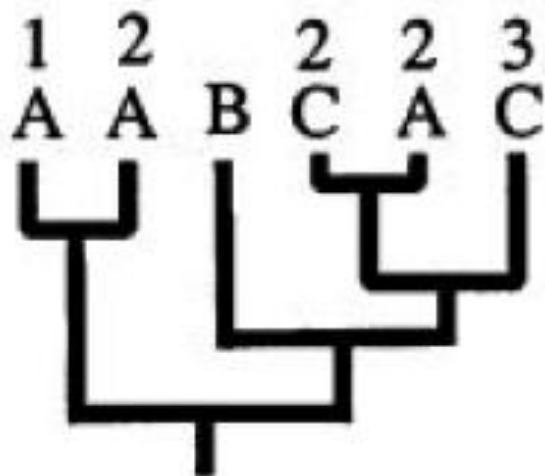


Cryptic species



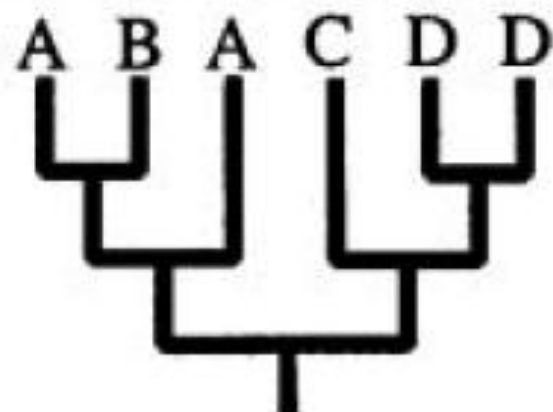
## Interspecific hybridization

g Introgressive hybridization



h Hybrid speciation (unidirectional)

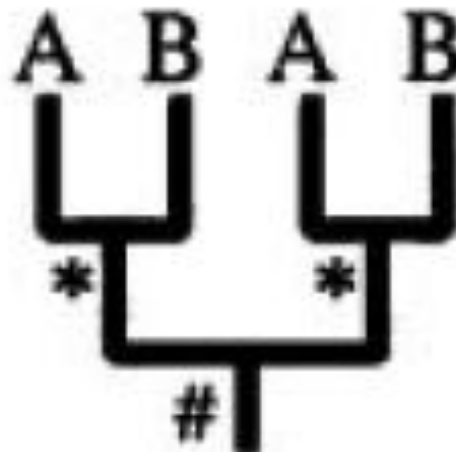
A = maternal species  
B = hybrid species  
D = paternal species





## Paralogy

#single gene duplication  
followed by  
\*single speciation event



# Gene Trees vs Species Trees

- **Gene Tree is not always the same as the species tree. Why?**
- **1. Genetic lineage divergence will predate species divergence (unless there is an extreme bottleneck)**
- **2. Hybridization, recombination and homoplasy may occur**
- **3. DNA sequences may be too short - too much variance in the estimate**

Review paper: Nichols 2001 TREE 16:358-364.

# Gene Tree/Species Tree

- Time since speciation is very important
- $N_e$  is very important
- :::::::::::Danger Zone::::: When evolutionary time since speciation event is short and  $N_e$  is large
- Note: The gene tree vs species tree question is different than the question: Did we obtain the correct gene tree with our phylogenetic methods?

# Molecular Clocks

- **Idea proposed in 1965 by Zuckerkandl and Pauling - suggesting that protein and DNA changes occurred at a constant rate over time thus: genetic distance = time elapsed (with a calibration point)**
- **Suggested a “stochastically constant” clock (like radioactive decay)**
- **Why do we want to use a molecular clock?**

- **What assumptions do we make to use a molecular clock?**
- 1) Rate of mutation constant over time and across taxa that are compared (**we can test this assumption with statistical models**)
- 2) The fossil or biogeographic event or average rate (from other species) used to calibrate the clock is accurate and reflects lineage divergence

# Phylogenetic Relationships of the Phasianidae

Bush and Strobeck 2003 J. of Heredity 94:472-489



- **Evaluated 21 pheasant and 6 non-pheasant taxa**
- **Generated mtDNA sequence data for the CytB gene (975 - 1300 bp)**
- **Generated trees using maximum parsimony, maximum likelihood, and weighted parsimony**





# Questions



- **What is a pheasant?**
- **Classically defined as 3 subfamilies: Tragopaninae (tragopans - 1 genus), Argusianinae (Argus pheasants - 3 genera), and Phasianinae (pheasants - 10 genera)**
- **First study to include all 3 subfamilies and outgroups**



## Questions

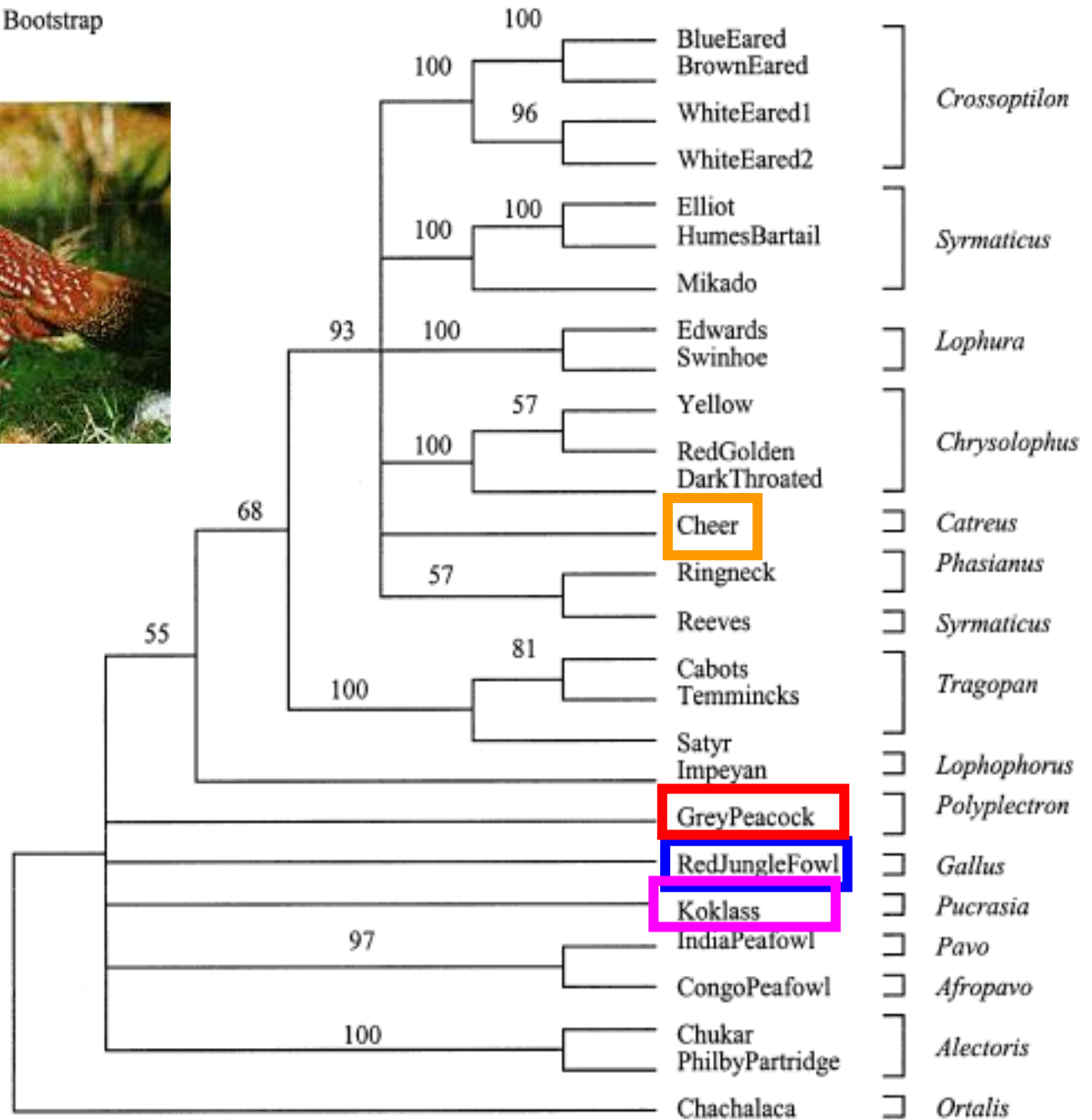
- **Is there genetic support for current taxonomy?**
- **Are the species currently described as pheasants monophyletic?**
- **Is the grey peacock pheasant a pheasant or peafowl?**

**Table 1.** Species examined and source of sequence data

Family	Subfamily	Genus	Species and common name	Source of cytochrome b GenBank accession no.
Cracidae		<i>Ortalis</i>	<i>Ortalis vetula</i> (plain chachalaca)	L08384
Phasianidae	Perdicinae	<i>Alectoris</i>	<i>Alectoris chukar</i> (chukar partridge)	L08378
			<i>Alectoris philbyi</i> (Philby's rock partridge)	Z48774
			<i>Alectoris pectorator</i> (Indian partridge)	L08379
	Afropavoninae	<i>Afropavo</i>	<i>Afropavo congensis</i> (Congo peafowl)	AF013760
	Pavoninae	<i>Pavo</i>	<i>Pavo cristatus</i> (India blue peafowl)	L08379
	Tragopaninae	<i>Tragopan</i>	<i>Tragopan caboti</i> (Cabot's tragopan)	AF534554 <sup>d</sup>
			<i>Tragopan satyra</i> (satyr tragopan)	AF534555 <sup>d</sup>
			<i>Tragopan temminckii</i> (Temminck's tragopan)	AF028802
	Argusianinae	<i>Polyplectron</i>	<i>Polyplectron bicalcaratum</i> (grey peacock)	AF534564 <sup>d</sup>
	Phasianinae	<i>Chrysolophus</i>	<i>Chrysolophus pictus</i> (red golden)	AF028793
			<i>Chrysolophus pictus obscurus</i> (black-throated golden)	AF534563 <sup>d</sup>
			<i>Chrysolophus pictus infuscatus</i> (yellow golden)	AF534562 <sup>d</sup>
		<i>Crossoptilon</i>	<i>Crossoptilon auritum</i> (blue-eared)	AF534552 <sup>d</sup>
			<i>Crossoptilon manchuricum</i> (brown-eared)	AF534553 <sup>d</sup>
			<i>Crossoptilon crossoptilon drouyni</i> (white-eared drouyni)	AF534556 <sup>d</sup>
<i>Crossoptilon crossoptilon crossoptilon</i> (white-eared Szechuan)			AF028794	
<i>Catreus</i>		<i>Catreus wallichii</i> (cheer)	AF028792	
<i>Lophura</i>		<i>Lophura edwardsi</i> (Edwards)	AF534557 <sup>d</sup>	
		<i>Lophura swinhoii</i> (Swinhoe)	AF534558 <sup>d</sup>	
<i>Syrnium</i>	<i>Syrnium ellioti</i> (Elliot)	AF534559 <sup>d</sup>		
	<i>Syrnium himiae himiae</i> (Humes bartail)	AF534560 <sup>d</sup>		
	<i>Syrnium mikado</i> (Mikado)	AF534561 <sup>d</sup>		
	<i>Syrnium reevesi</i> (Reeves)	AF028801		
<i>Phasianus</i>	<i>Phasianus colchicus torquatus</i> (ringneck)	AF028798		
<i>Lophophorus</i>	<i>Lophophorus impeyanus</i> (Impeyan)	AF028796		
<i>Pucrasia</i>	<i>Pucrasia macrolopha macrolopha</i> (koklass)	AF028800		
<i>Gallus</i>	<i>Gallus gallus</i> (red jungle fowl)	AF028795		



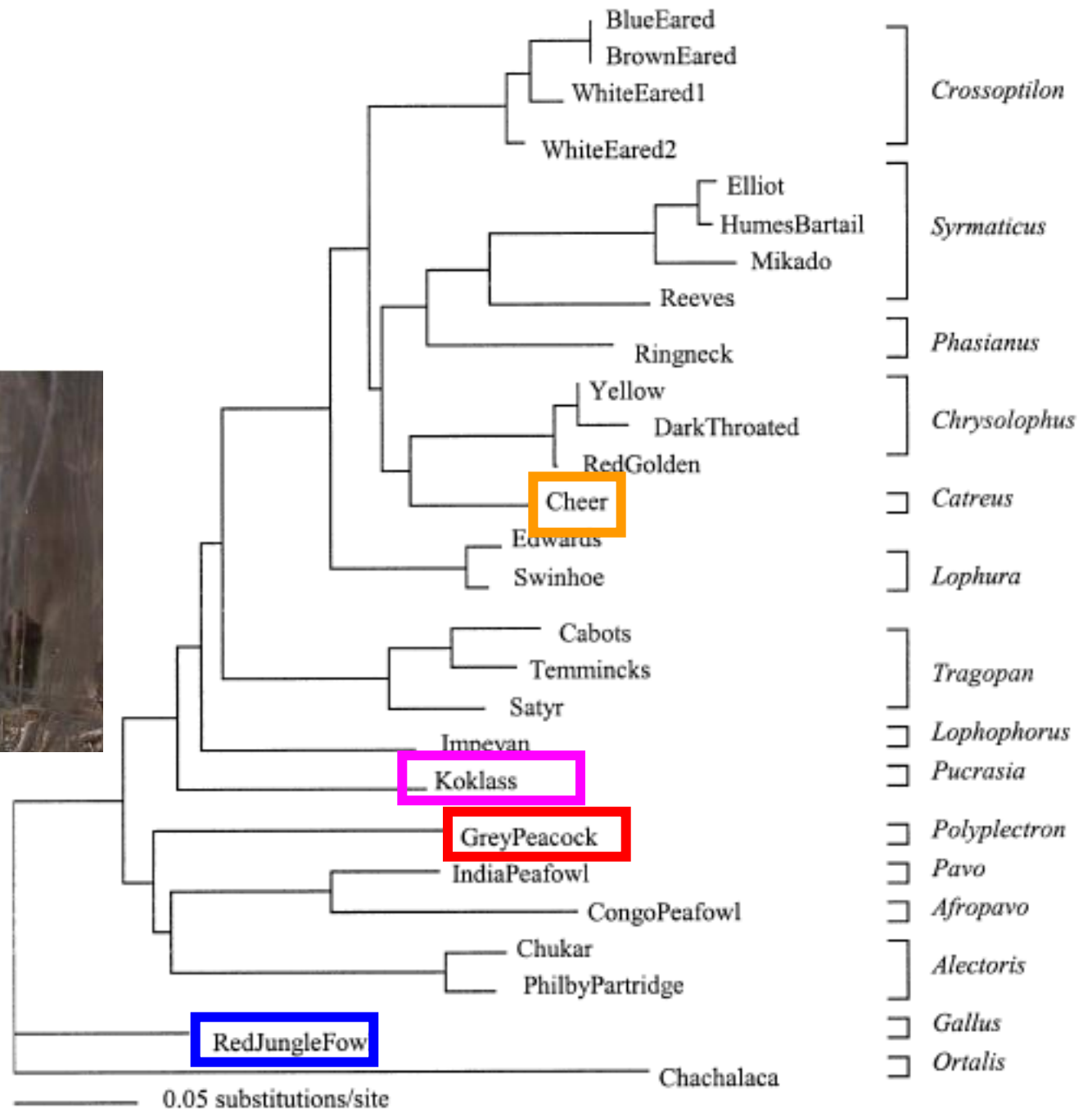
Bootstrap



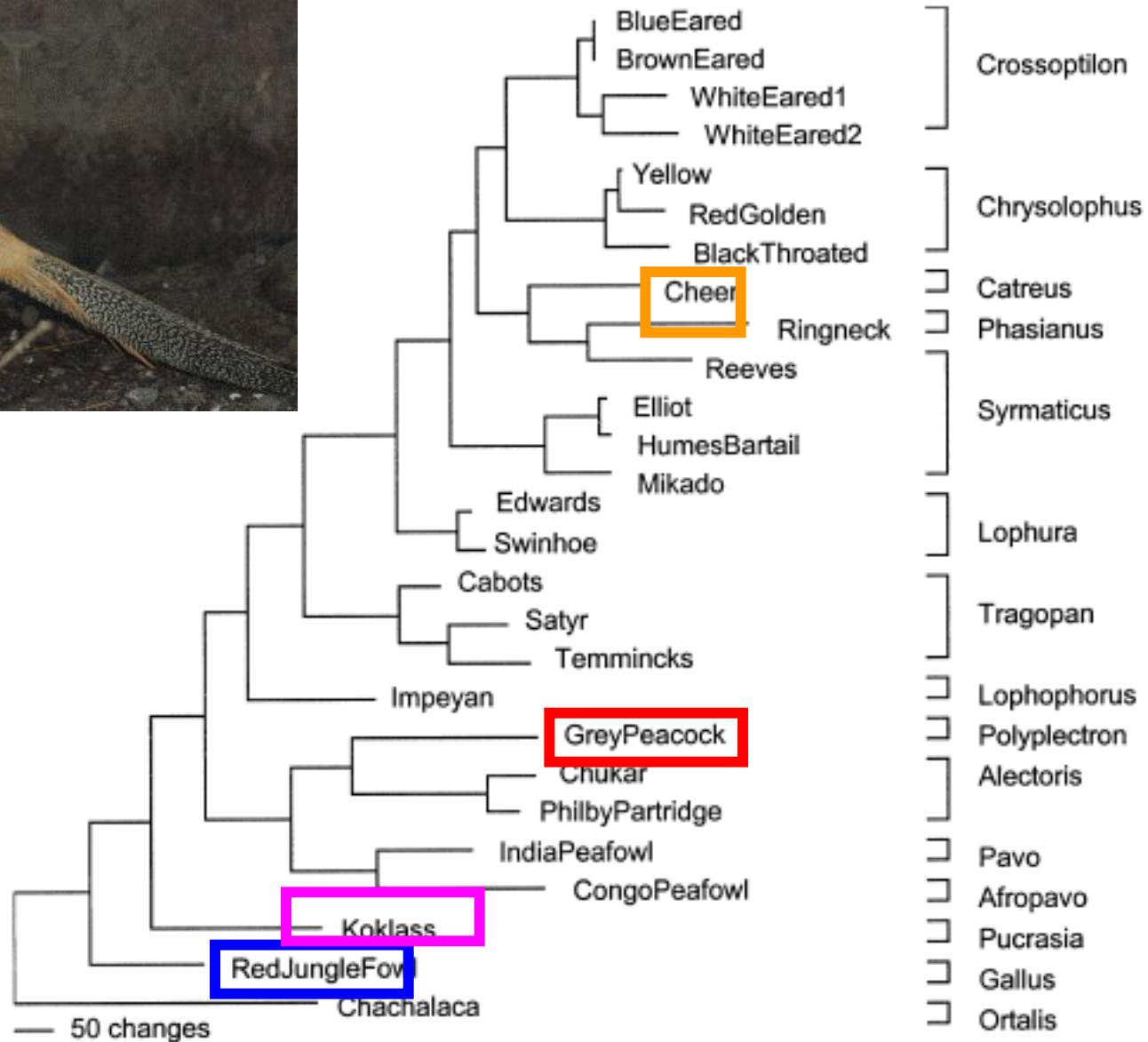
**Parsimony**  
**1000 bootstrap**  
**Majority Rule**

# Maximum Likelihood

HKY+gamma



# Weighted Parsimony: 4:15:1



# Conclusions



- Some taxonomic groupings are well supported
- Status of some species is unresolved
- **Phylogenetic method can greatly influence placement/interpretation**
- Two classic pheasants - red jungle fowl and grey peacock....not supported by genetics

# Genetic vs Morphological Taxonomy

- **There are multiple efforts to create genetic-based taxonomies:**
- **Tree of life [tolweb.org/tree](http://tolweb.org/tree)**
- **PhyloCode: TREE 19:280 - 282.**
- **MOTU (molecular taxonomic unit) - TREE 18:268-270**
- **[www.DNA-surveillance](http://www.DNA-surveillance)**
- **TREE 18:70-74 DNA-based Taxonomy system**
- **PNAS 101:14812 - 10 species in 1...DNA barcoding**
- **Nature jan 06 - barcoding an entire island 439:378**