

# Phylogenetic terminology and applications

Emily Jane McTavish

Life and Environmental Sciences  
University of California, Merced  
`ejmctavish@ucmerced.edu`, `twitter:snacktavish`

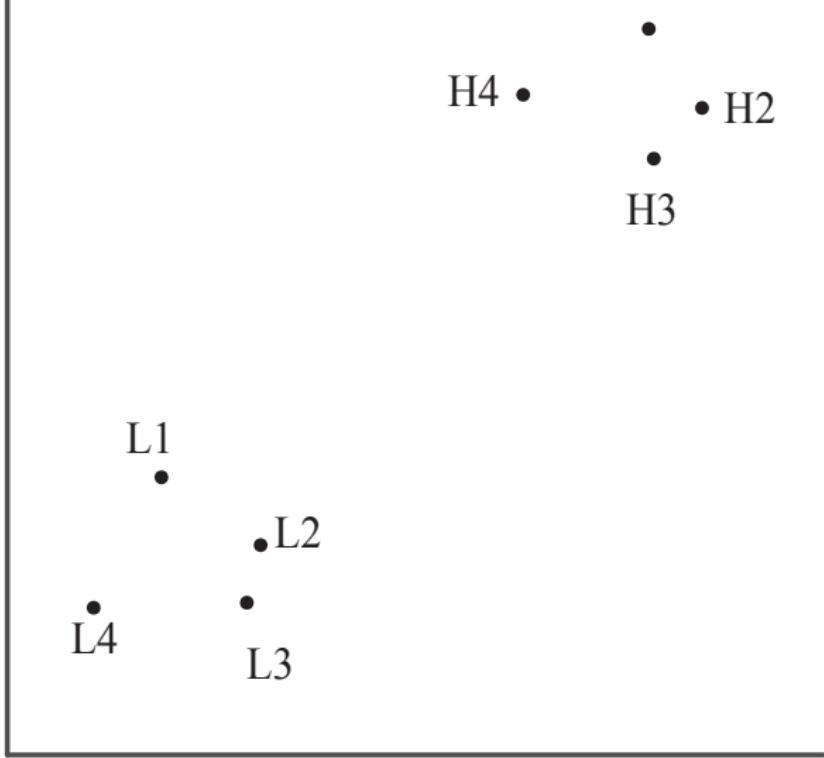
(With thanks to Mark Holder, Paul Lewis, Joe Felsenstein, and David Hillis  
for slides)

Phylogenies describe shared ancestry  
and  
inform our understanding of evolutionary processes

Simple test of Bergmann's rule: comparing latitude and mass (I made these data up)

lat. offset = degrees north of the 49th parallel.

species	lat. offset	mass
L1	3.1	5.9
L2	5.4	4.3
L3	5.1	3.1
L4	1.8	3.6
H1	13.5	15.2
H2	14.6	13.5
H3	13.6	12.4
H4	10.8	13.7



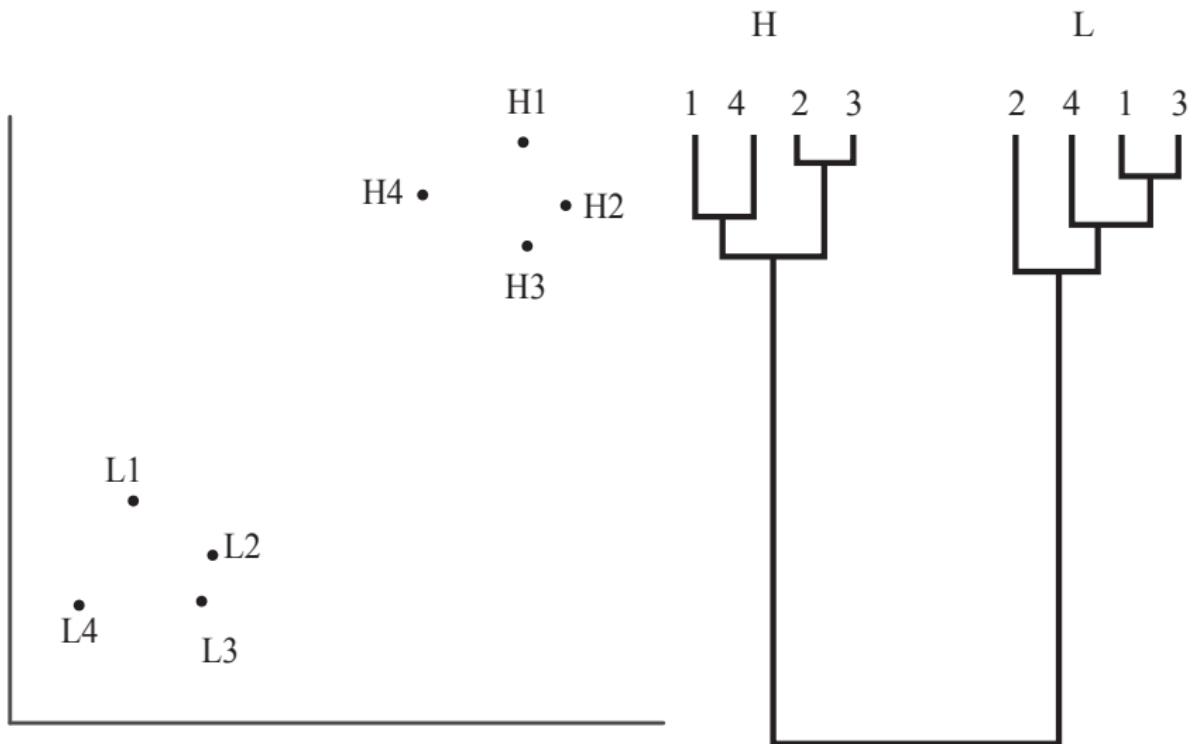
(cue cartoon videos)

See <http://phylo.bio.ku.edu/slides/no-correl-anim.mov>

and <http://phylo.bio.ku.edu/slides/correl-anim2.mov>

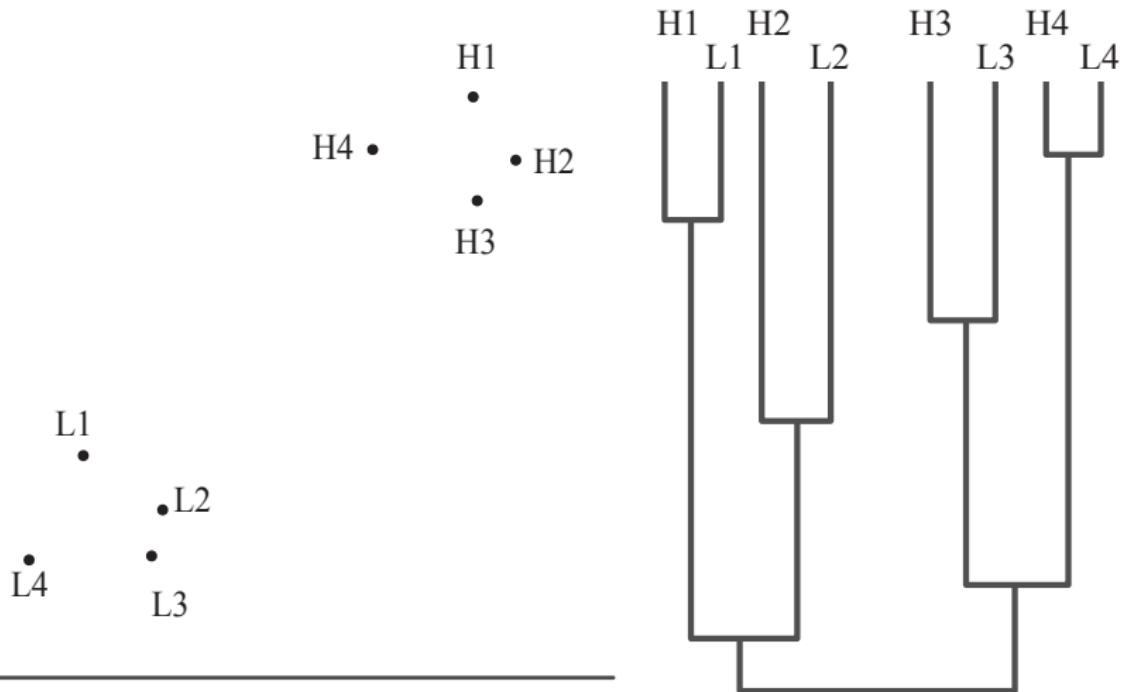
## No (or little) evidence for correlation

---



## Evidence for correlation

---

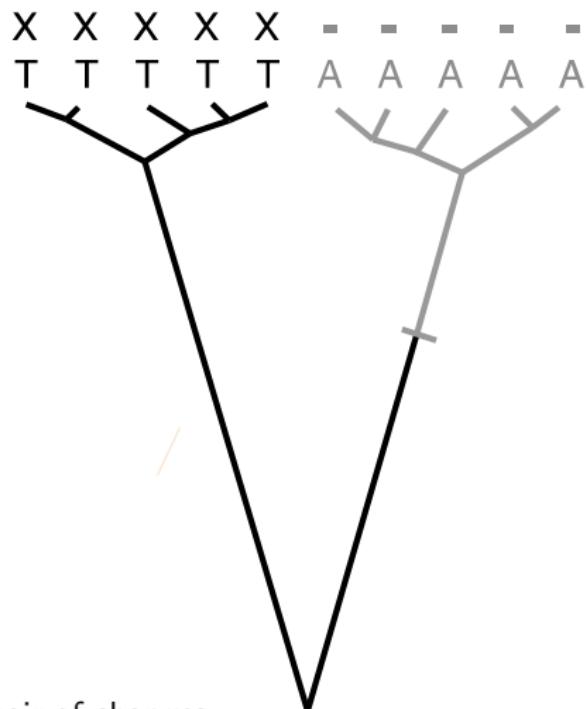


## Do desert green algae use xanthophyll to protect against excessive light intensities?

---

Species	Habitat	Photoprotection
1	terrestrial	xanthophyll
2	terrestrial	xanthophyll
3	terrestrial	xanthophyll
4	terrestrial	xanthophyll
5	terrestrial	xanthophyll
6	aquatic	none
7	aquatic	none
8	aquatic	none
9	aquatic	none
10	aquatic	none

## Phylogeny reveals the events that generate the pattern



1 pair of changes.  
Coincidence?

5 pairs of changes.  
Much more convincing

## Inferring Process from Pattern

---

Hypothesis:

Gregariousness should arise more frequently in unpalatable organisms than in tasty ones (**Sillén-Tullberg, 1988**)

# Inferring Process from Pattern



Solitary

Gregarious



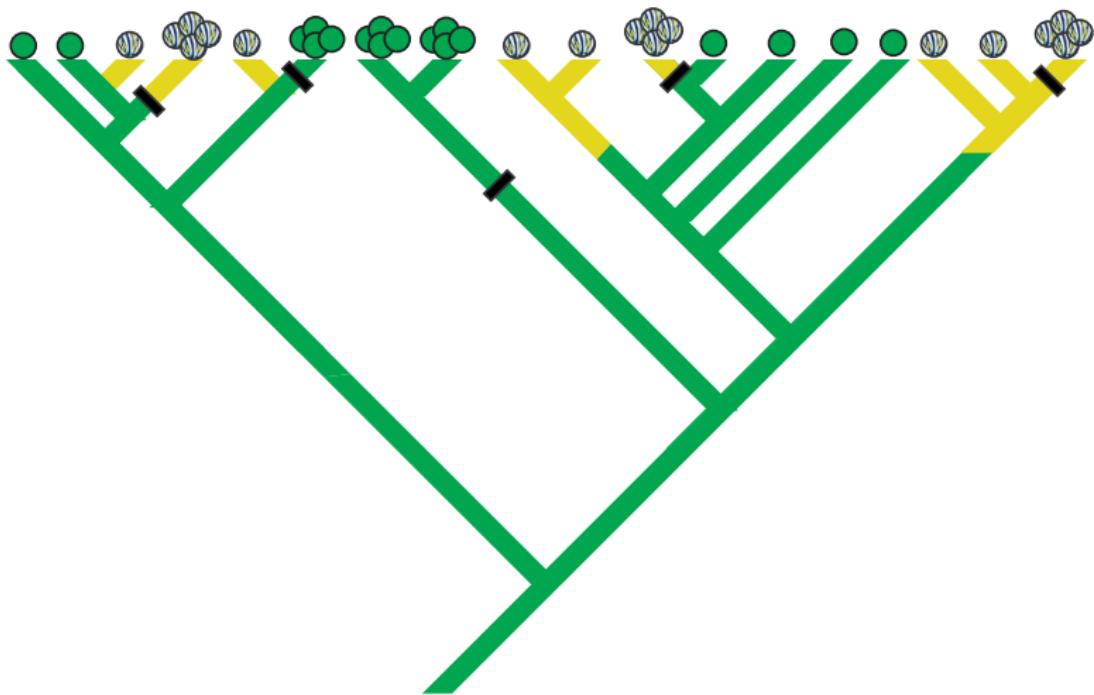
Cryptic



Aposematic

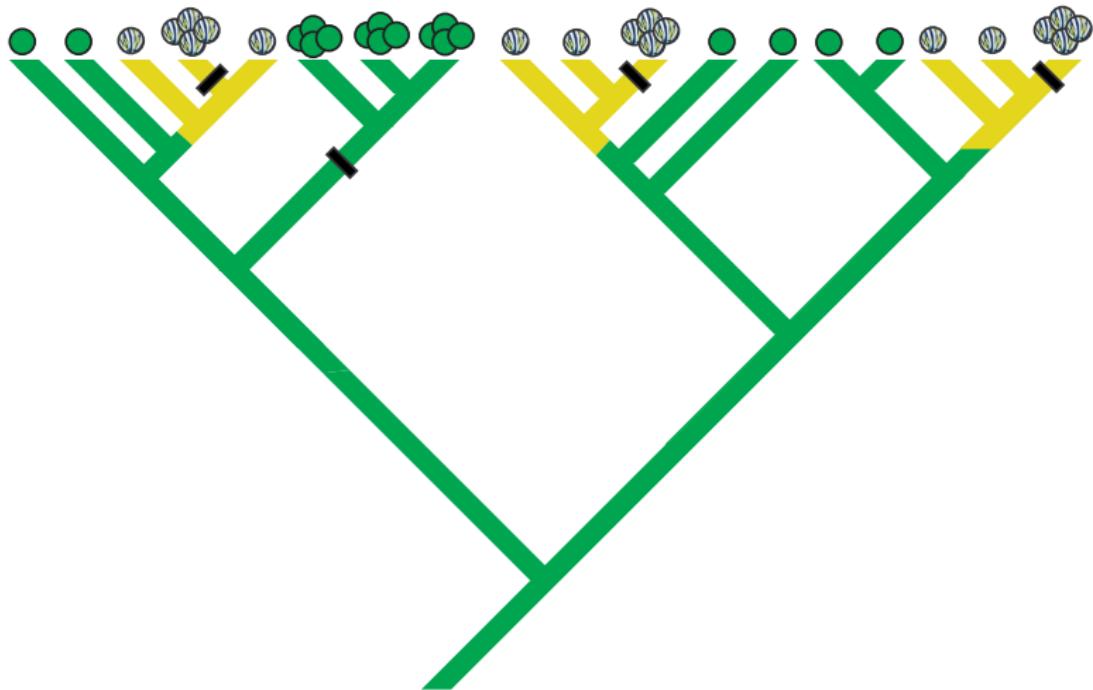


Sillén-Tullberg (1988), Dyer and Gentry (2002), Hill (2001)



One possible outcome:

No clear evidence of associations between traits



Cartoon of the real results ([Sillén-Tullberg, 1988](#))

Aposematic species are more likely to evolve gregarious larvae

## **Importance of phylogeny**

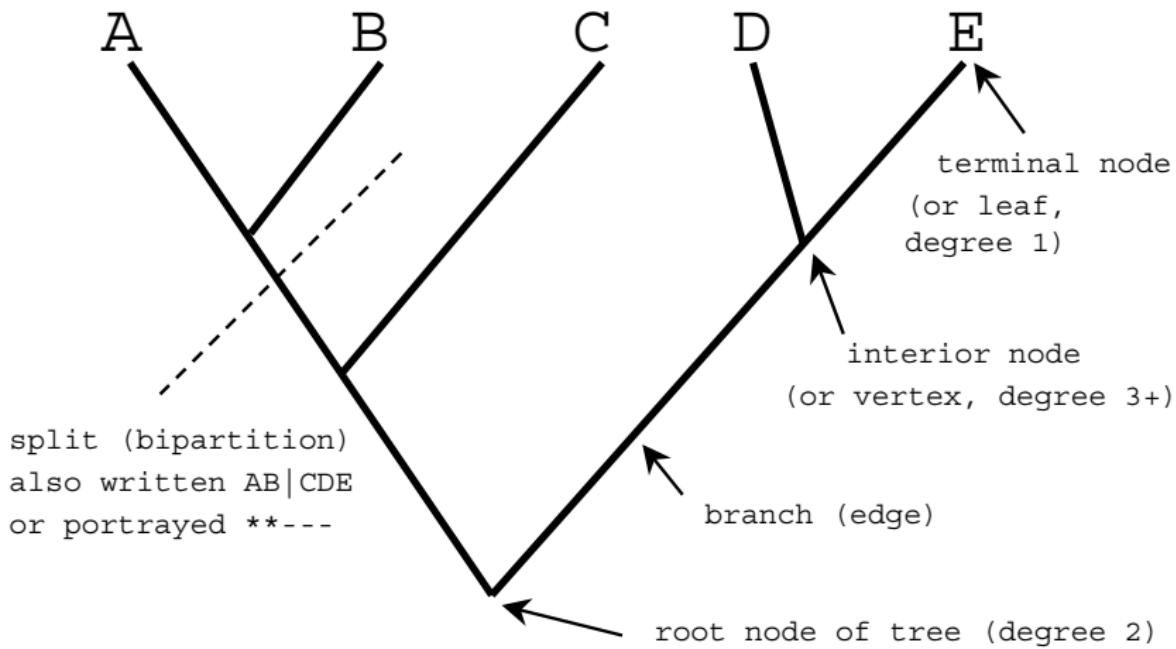
---

The previous slides had identical patterns of traits if the phylogeny is ignored.

Without knowledge of the tree, no conclusion would be reached.

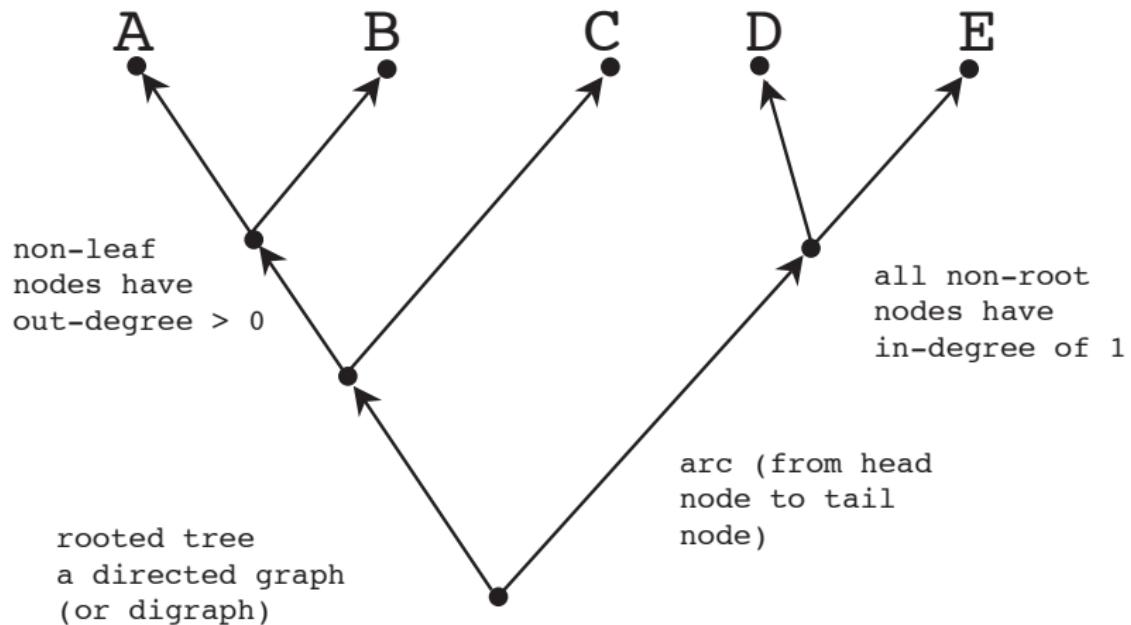
## Tree terminology

---



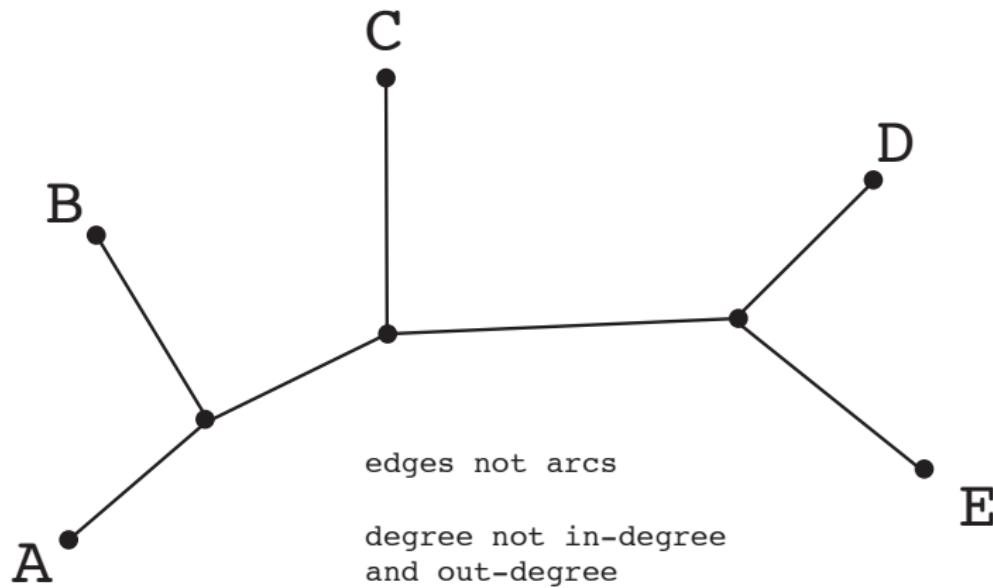
## Rooted tree terminology

---



## Rooted tree terminology

---



## Tree terms

---

A tree is a connected, acyclic graph.

A rooted tree is a connected, acyclic directed graph.

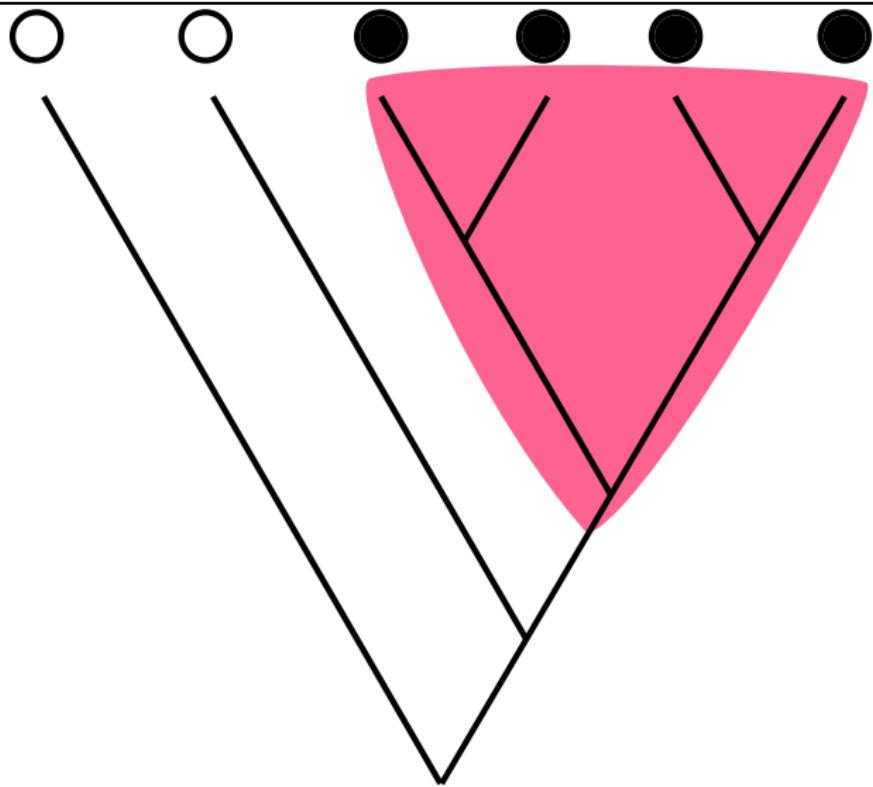
A polytomy or multifurcation is a node with a degree  $> 3$  (in an unrooted tree), or a node with an out-degree  $> 2$  (in a rooted tree).

Collapsing an edge means to merge the nodes at the end of the branch (resulting in a polytomy in most cases).

Refining a polytomy means to “break” the node into two nodes that are connected by an edge.

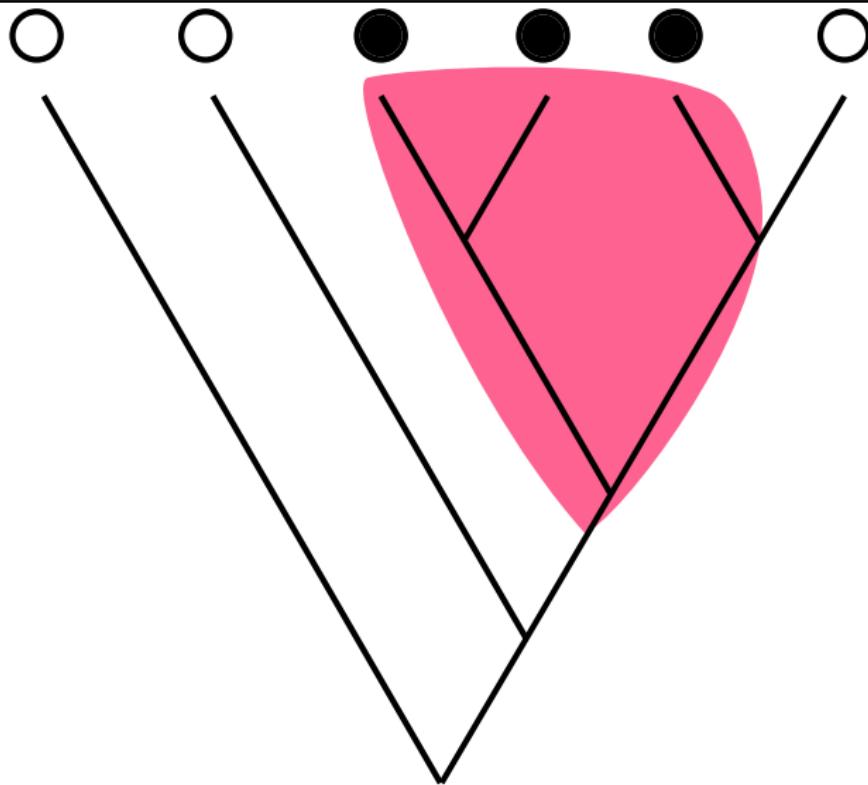
## Monophyletic groups (“clades”): the basis of phylogenetic classification

---

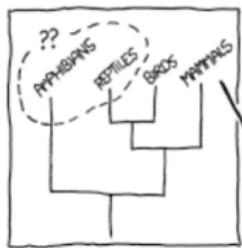


## Paraphyletic groups: error of omitting some species

---



ORNITHOLOGY CONFERENCE:

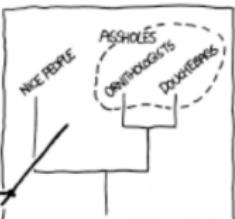


AS YOU CAN SEE, HERPETOLOGY IS A SILLY FIELD; REPTILES ARE ACTUALLY MORE CLOSELY RELATED TO BIRDS AND MAMMALS THAN TO AMPHIBIANS.

)  
IT SHOULD REALLY BE BROKEN UP, WITH LIZARDS FOLDED INTO ORNITHOLOGY.

HERPETOLOGY CONFERENCE:

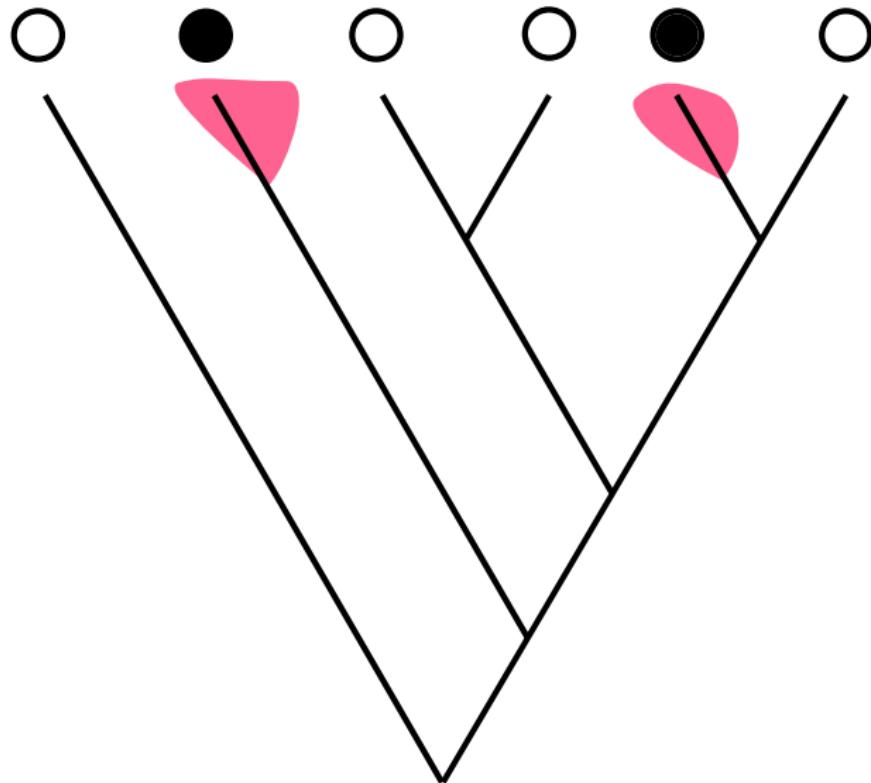
AS YOU CAN SEE, ORNITHOLOGISTS ARE ACTUALLY ASSHOLES.



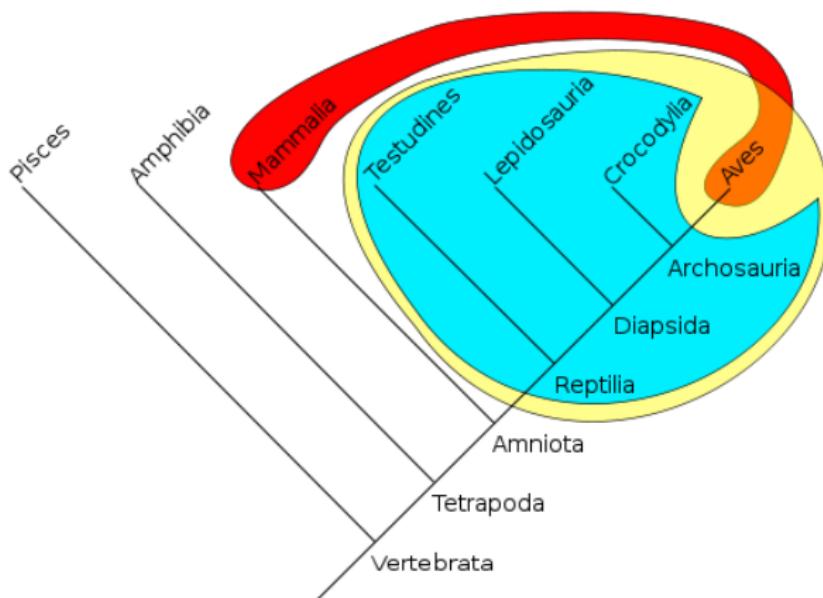
<https://xkcd.com/867/>

## Polyphyletic groups: error of grouping “unrelated” species

---



- Monophyly
- Paraphyly
- Polyphyly



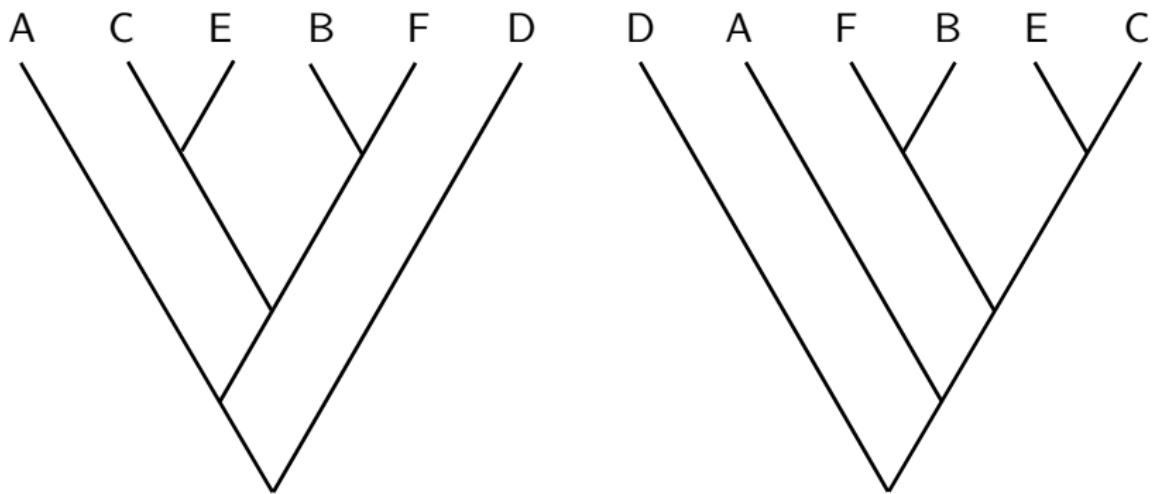
from wikipedia

more terms:

- ▶ sister taxa: taxa or monophyletic groups which share a most recent common ancestor
- ▶ outgroup: taxon that is determined *a priori* to be sister to all other taxa in the analysis. Used for rooting tree

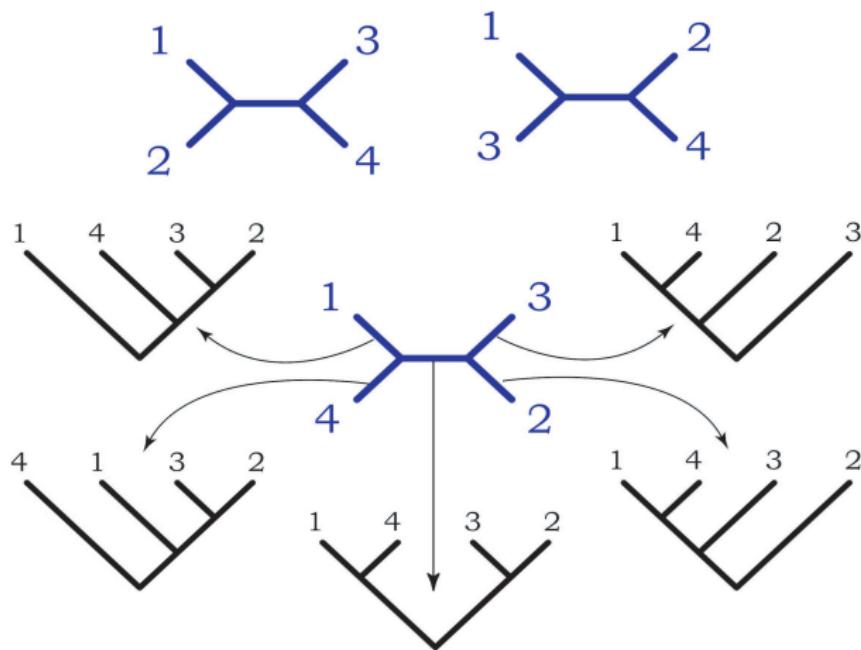
## Branch rotation does not matter

---



## Rooted vs unrooted trees

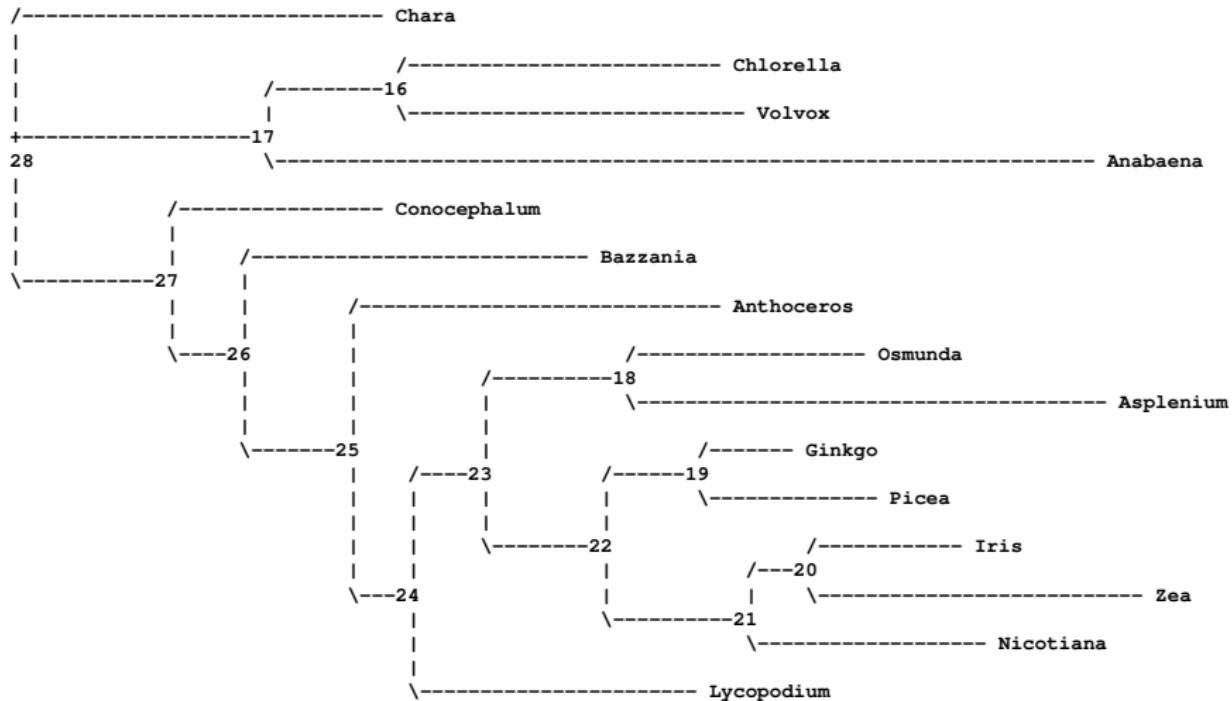
---



## Splits

- ▶ It is useful to think of unrooted trees in terms of 'splits'
- ▶ Each branch in an unrooted tree splits the taxa into two groups.
- ▶ Membership in those groups can be denoted by \*\* vs ..
- ▶ e.g. a split between 1+2 and 3+4 can be summarized as
- ▶ 1234
- ▶ \*\* ..

## **Warning: software often displays unrooted trees like this:**



a brief digression into newick tree file format



Newick's Lobster House was the site of an historic 1986 meeting at which a standard was devised for storing descriptions of phylogenetic trees as strings. (Photo from Paul Lewis)

Note: ((1,2),3,4) is referred to as Newick or New Hampshire notation for the tree.

You can read it by following the rules:

- start at a node,
- if the next symbol is '(' then add a child to the current node and move to this child,
- if the next symbol is a label, then label the node that you are at,
- if the next symbol is a comma, then move back to the current node's parent and add another child,
- if the next symbol is a ')', then move back to the current node's parent.

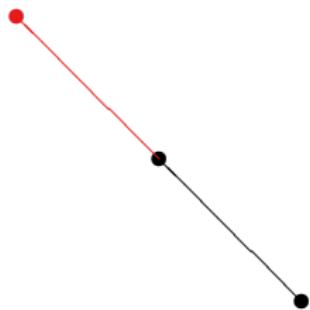
$((1,2),3,4)$

•

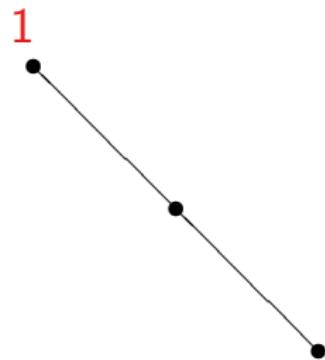
$((1,2),3,4)$



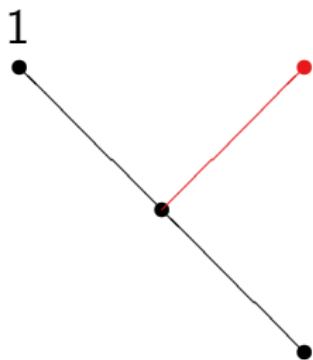
$((1,2),3,4)$



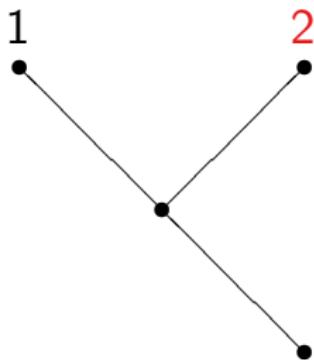
$((\textcolor{red}{1},2),3,4)$



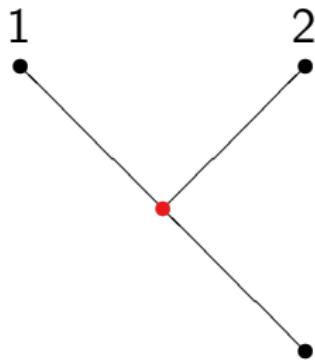
$((1,2),3,4)$



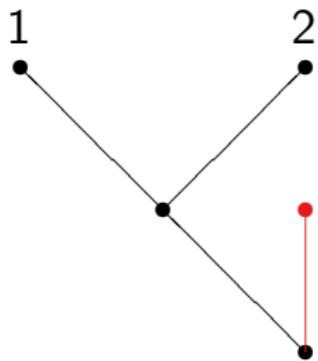
$((1,\textcolor{red}{2}),3,4)$



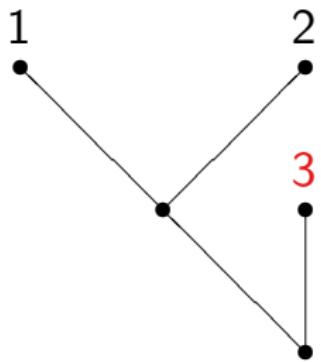
$((1,2),3,4)$



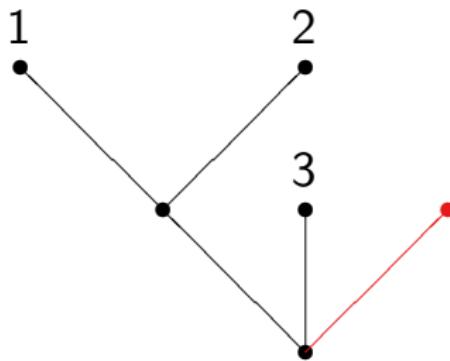
$((1,2), 3, 4)$



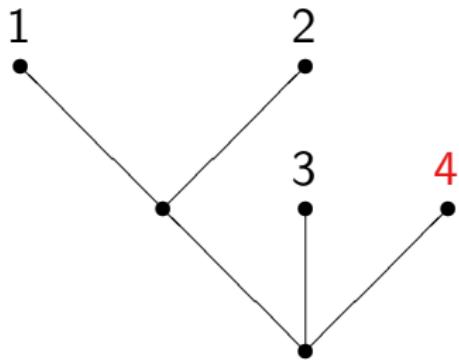
$((1,2),\textcolor{red}{3},4)$



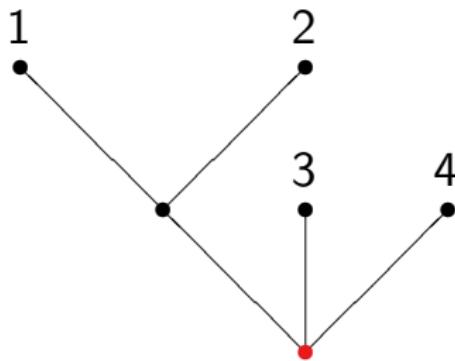
$((1,2),3,4)$



$((1,2),3,\textcolor{red}{4})$



$((1,2),3,4)$



## Newick

- ▶ Parenthetical tree format
- ▶ Rooted vs. unrooted trees are not differentiated
- ▶ Some programs interpret polytomy at root as ‘unrooted’
- ▶ Branches and nodes not well differentiated
- ▶ A name can contain any characters except blanks, colons, semicolons, parentheses, and square brackets

# DIY

Create a newick tree file in your text editor with the content:

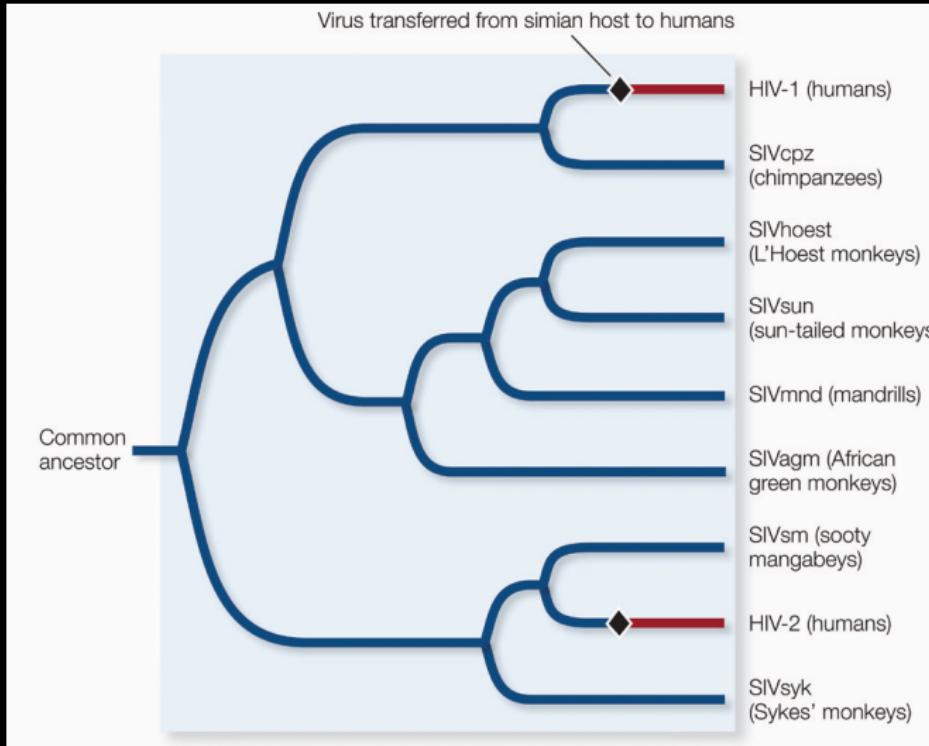
`((C, (D, E)), (F, G), A), B;`

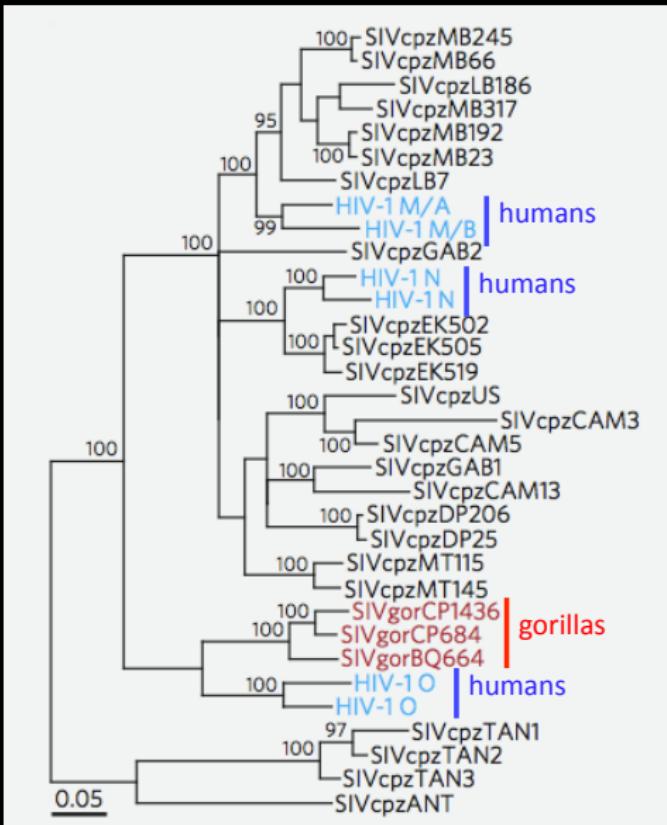
Save it as 'example.tre'.

- ▶ Draw the tree by hand
- ▶ Write down all the splits in ..\*\* format.
- ▶ Load the tree in a tree viewer (e.g. phylo.io). Re-root the tree. What rootings make the following true? Which cannot be true?
  - ▶ A is more closely related to G than it is to C
  - ▶ (C,D,E) is sister to (A,B,F,G)
  - ▶ (C,D) is sister to (A,B,E,F,G)
  - ▶ (C,D,E) is a paraphyletic group
  - ▶ (C,D,E) is a monophyletic group
  - ▶ (A,B,C) is a monophyletic group

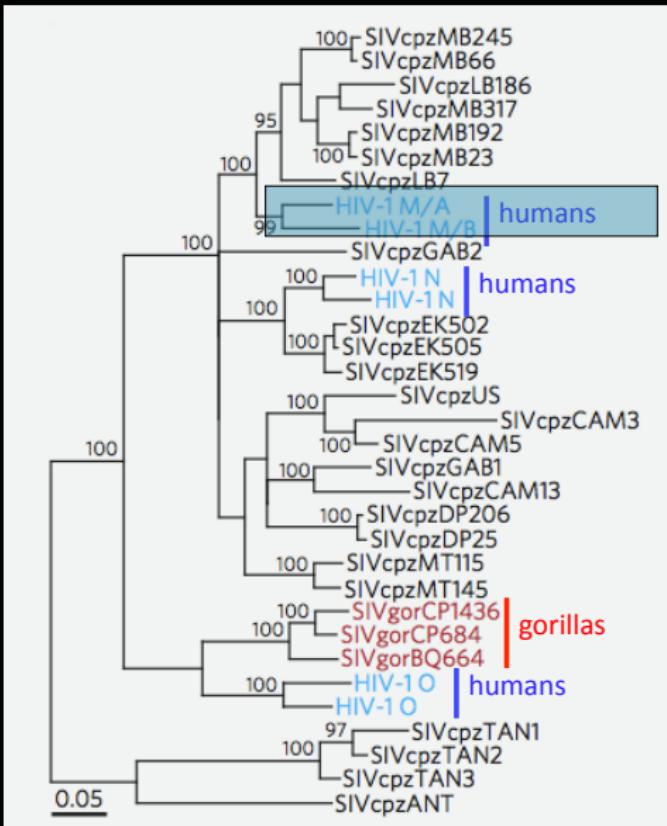
# Origins of Emerging Diseases

- Where did HIV come from?
- How did it enter human populations?
- When did it enter human populations?
- How can we prevent similar diseases from entering human populations?

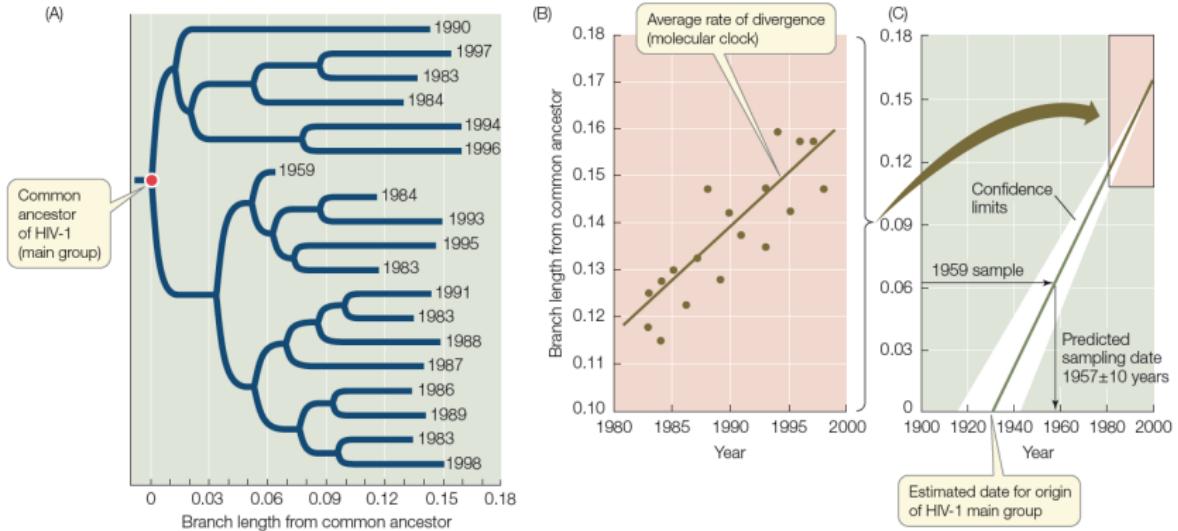




Van Heuverswyn et al., *Nature* 444: 164 (2006)



Van Heuverswyn et al., *Nature* 444: 164 (2006)



# THE DAILY ADVERTISER

Louisiana

Thursday, October 22, 1998

## DNA debate

Jury hears AIDS DNA evidence against Schmidt

BILL DECKER  
Staff Writer

LAFAYETTE — Testimony in the attempted murder trial of Dr. Richard Schmidt continued Wednesday as the prosecution's DNA evidence — and defense attorneys' one defense — came under scrutiny.

A study College of Medicine researchers at the University of Texas at Austin found that a "close relationship" exists between the genetic material of the HIV virus strains found in alleged victim Allen and the strains found in Schmidt's patients. The study supports the prosecution's case that Schmidt injected Allen with the AIDS virus, which Schmidt denied.

Defense attorney Michael Fawer, who performed the study in 1985 and now represents Schmidt, Dr. Ray Baylor and David Hillis, a University of Texas expert who performed Schmidt's tests, contended Wednesday:

"The study was conducted and marred by verbal sparring with defense attorney Michael Fawer and the lack of any scientific control of other test results."

McMetrick, a recognized authority on the technique used to compare the viral DNA samples, said he had no idea where his answers beyond a simple yes or no.



Defense attorney Michael Fawer, left, leaves court followed by Dr. Richard Schmidt, after Wednesday's proceedings. Testimony continues at 9 a.m. today.

### Alleged source of AIDS-tainted blood testifies

BILL DECKER  
Staff Writer

LAFAYETTE — Former teacher Don McClelland testi-

be drawn from McClelland and injected it into Allen that night. It was the first time he ended their 10-year affair.

McClelland, who has AIDS

1994, McClelland testified that he remembers having blood drawn from him to test for AIDS. "I don't know if they took a lot of it," he said.

## Phylogenetic analysis can be used to trace viral infections through a human population

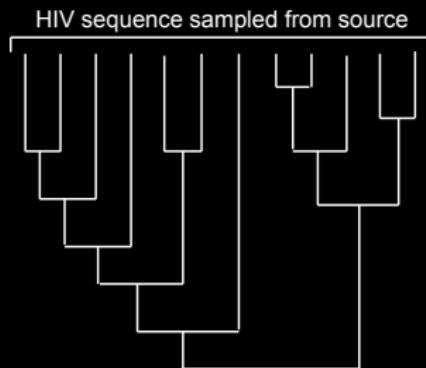
- Origins of HIV, SARS and other viruses transmitted between animals and humans
- Global virus diversity for vaccine trials
- Epidemiological studies
- Identification of new diseases
- Forensic uses

# HIV transmission

Viral transmission events may be traced back through time among individuals in a population.

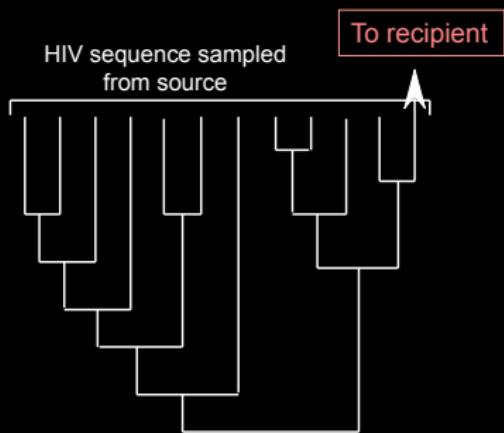
To imagine how this is possible, start by considering the diversity of HIV within one infected individual:

Time 1: Prior to Transmission event



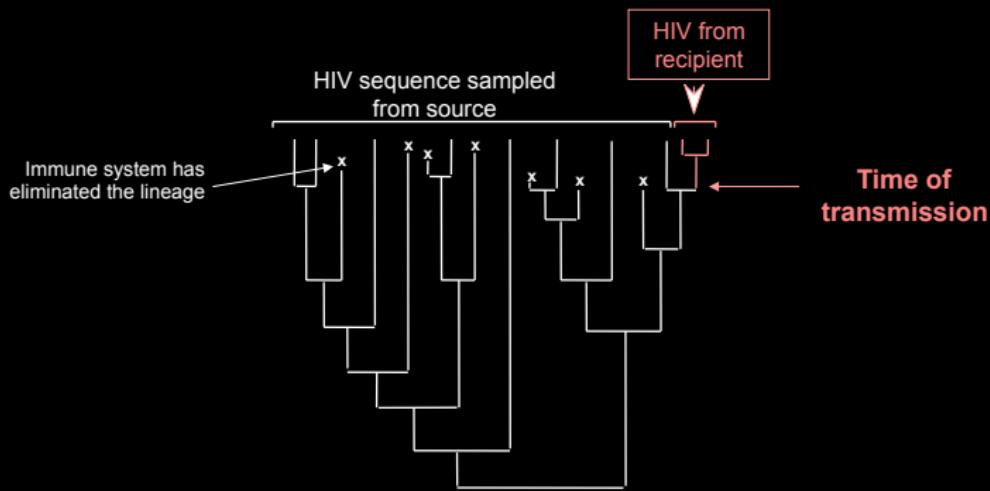
**At the transmission event, the HIV in the recipient represents a small subset of the HIV present in the source:**

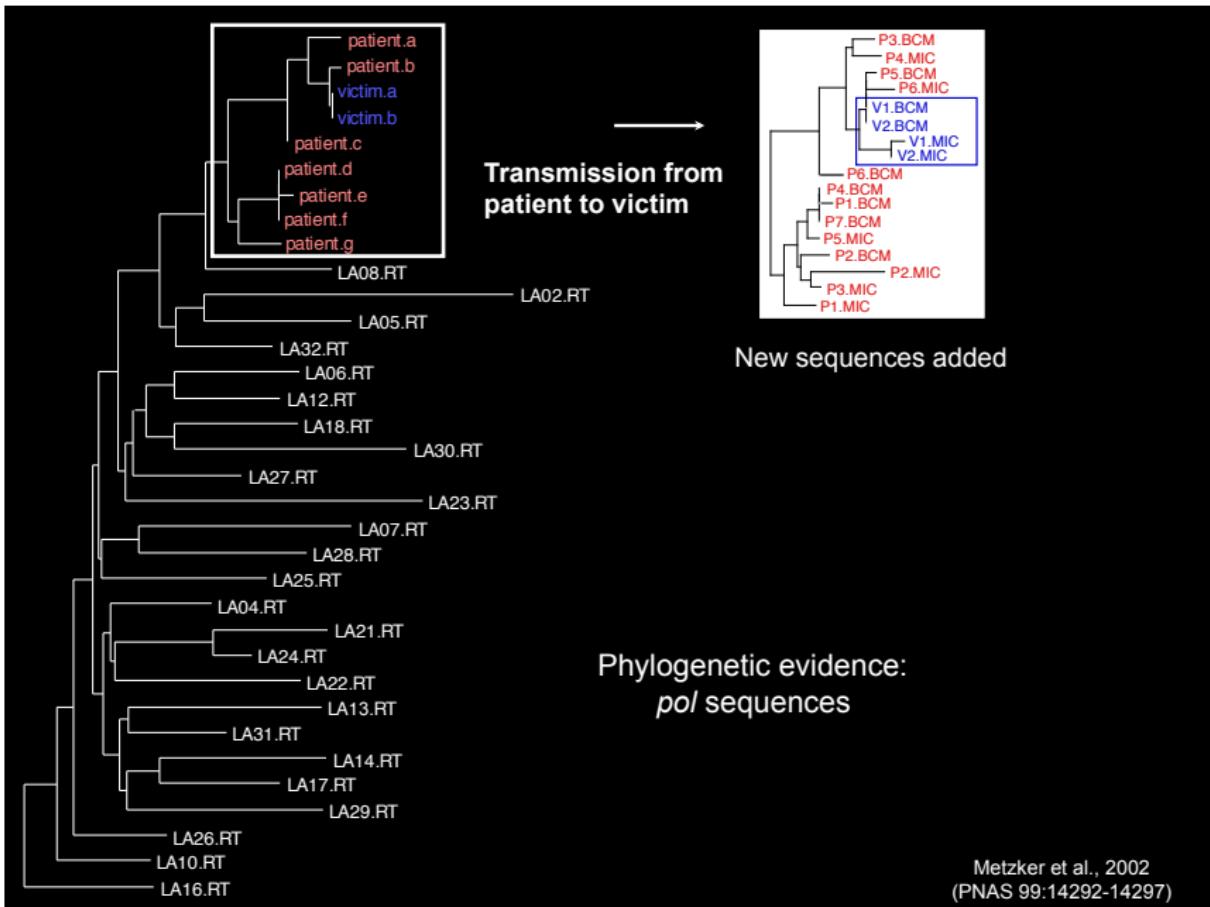
Time 2: The transmission event



As time passes, HIV lineages in the source and recipient diversify, and other lineages become extinct.

Time 3: Shortly after transmission event





# THE DAILY ADVERTISER

Acadiana's Daily Newspaper

Louisiana

Saturday, October 24, 1998

## Schmidt guilty

Doctor faces  
50-year jail  
sentence

By Bill Decker  
Staff Writer

LAFAYETTE — A Lafayette Parish jury found Dr. Richard Schmidt guilty late Friday of attempted second-degree murder in a botched and premeditated shooting trial.

Waiting for the verdict, Schmidt, 56, sat alone after 8:30 p.m. Friday, clutching the hand of his wife, Barbara, and a very tall glass. When the clock read 9:15 p.m., Schmidt began sobbing loudly.

"I'm sorry," he said as the verdict was read, shaking his station, "but as the circumstances forced me to do, I did what I had to do." After he was convicted of attempted pending sentencing, Schmidt sat up and put his head in his hands. Then he stood, took a deep breath and turned to embrace his wife, Barbara, who sobbed loudly.

Barbara Schmidt was taken from the courtroom in a wheelchair.

"I just don't know what the law should do," defense attorney Michael Finner said after the trial. "It's not clear if he intended to kill her or if he was brought with reasonably doubts." The now-widow alone then gave a brief statement through an attorney.

The

Claps, sobs, fainting  
spell greet verdict

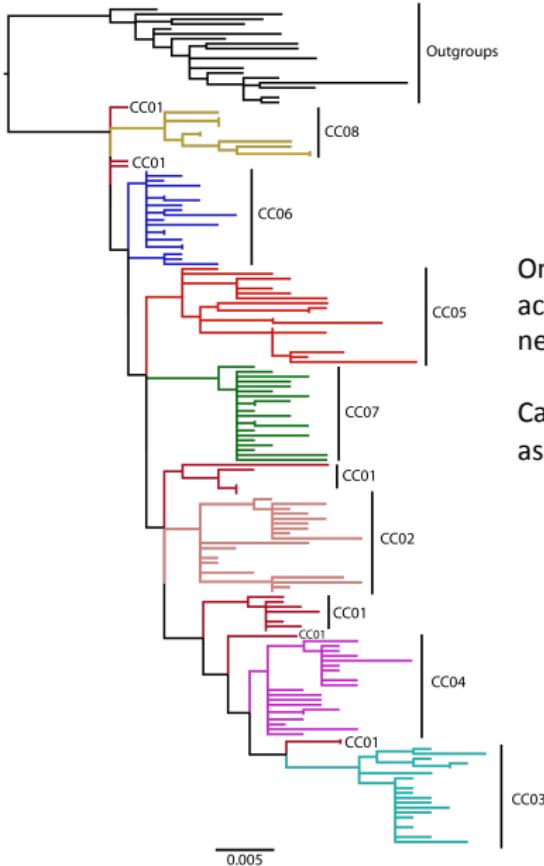
By Ester Avi  
Staff Writer



Dr. Richard Schmidt, center, leaves the Lafayette Parish Courthouse Friday night with his respondent wife, Barbara, after being convicted of attempted second-degree murder. Schmidt is accompanied by courthouse security and defense attorney Gerald Block, left.

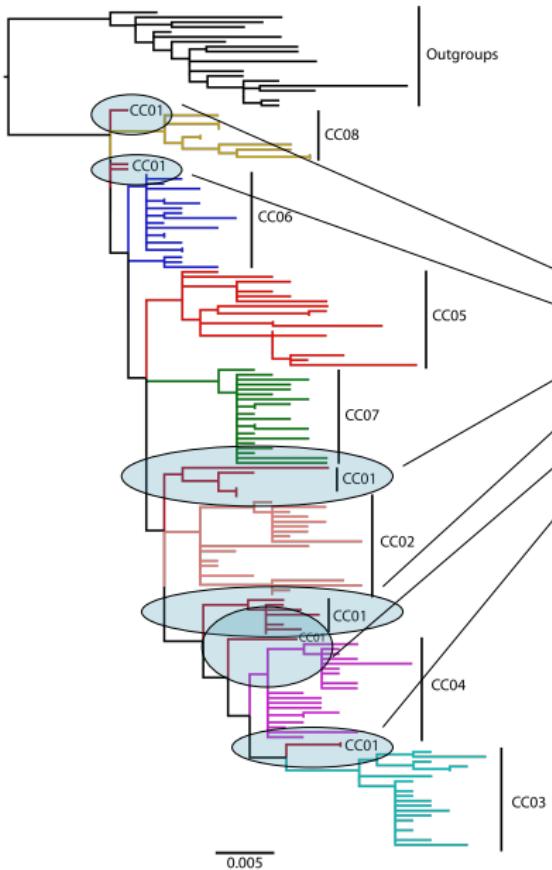


- Schmidt was convicted of attempted murder; currently serving term of 50 years of hard labor
- First use of phylogenetic analysis in U.S. criminal case
- Phylogenetics can be used to trace infections of human pathogens among individuals



One of these individuals is accused of knowingly and negligently infecting the others.

Can one person be identified as the source of the infections?



One individual (CC01) is paraphyletic to all the rest. At the trial, CC01 was revealed to be the defendant, who was accused of six counts of motivated assault. He was found guilty by the jury in May 2009.

Changing the rooting of this phylogeny would change inferences!

# Phylogenies can reveal surprising patterns



# Phylogenies can reveal surprising patterns



Figure by Mathieu Joron: <http://xyala.cap.ed.ac.uk/joron/>



What evolutionary processes can drive these patterns?

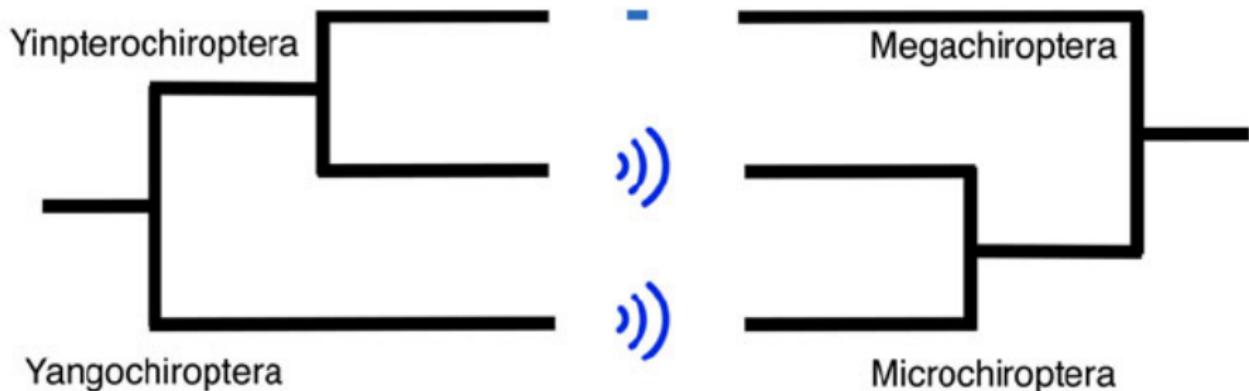
- ▶ Convergence
- ▶ Horizontal gene transfer
- ▶ Incomplete lineage sorting
- ▶ ?

What evolutionary processes can drive these patterns?

- ▶ Convergence
- ▶ Horizontal gene transfer
- ▶ Incomplete lineage sorting
- ▶ ?

We will discuss how to recognize and (try to) differentiate these processes.

Different data can drive different conclusions



Species relationships between echolocating and nonecholocating bats (after Teeling 2009). Left: inferences from DNA sequence data.  
Right: traditional species relationships inferred from morphological characters (and limited sequence data). (Hahn and Nakhleh, 2016)

## Estimating a tree from character data

Tree construction:

- ▶ strictly algorithmic approaches - use a “recipe” to construct a tree
- ▶ optimality based approaches - choose a way to “score” a trees and then search for the tree that has the best score.

Expressing support for aspects of the tree:

- ▶ bootstrapping,
- ▶ testing competing trees against each other,
- ▶ posterior probabilities (in Bayesian approaches).

- Hahn, M. W. and Nakhleh, L. (2016). Irrational exuberance for resolved species trees. *Evolution*, 70(1):7–17.
- Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe, M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail, M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., Jones, M. C., Rogers, J., Jiggins, C. D., and ffrench Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477(7363):203–206.