Lecture 1 – Introduction to Systematics and Phylogenetics

I. Context of Phylogeny within Systematics. Phylogenetics is a subdiscipline of Systematics, and the latter was formerly equated with Taxonomy.

Ernst Mayr defined Taxonomy as "the theory and practice of classifying organisms".

As we'll see, however, the current understanding of Systematics is much broader.

This breadth is sometimes represented by levels of taxonomy.
α-Taxonomy – includes species descriptions, taxonomic keys, and diagnoses.
β-Taxonomy – includes identification of natural groups and biological classes.
γ-Taxonomy – includes study of evolutionary processes and patterns.

This broad view is sometimes called **Biosystematics** and this term explicitly includes the third; this term is much more commonly used by botanists than by zoologists.

This broad sense is defined by the SSB as "the study of organismal diversity, including both the pattern of that diversity and the processes that have generated it."

II. Roles of Systematics. The role of systematic biology has changed over the years.

A. Provide α-taxonomy. This is really the classic hat worn by systematists and, as large vertebrate groups have become more thoroughly understood, this role was traditionally been somewhat trivialized. However, its fundamental importance has been rediscovered as the current biodiversity crisis worsens. Obviously, it becomes critical to describe biodiversity before we lose it.

This is recognized by NSF, for example in its PurSUiT program (Poorly Sampled and Unknown Taxa; https://www.nsf.gov/pubs/2020/nsf20059/nsf20059.jsp)

The fundamental importance of this aspect of α-taxonomy to society is very apparent if we look at biomedicine. Of the 150 most commonly prescribed drugs in the U.S., 57% contain active ingredients derived from natural compounds, that is, discovered from biodiversity that's been documented already.

A fairly recent estimate (Costello et al. 2013. Science 339:413-416. doi: 10.1126/science.1230318 – pdf on course website) is that there are ~3-8 million extant macrobial species on Earth and we've only described 1.5 million of them.

There may be as many as 1 trillion (10^{12}) microbial species (Locey & Lennon. 2016. PNAS 113:5970 – pdf on course website).

So, aside from the scientific importance, there is huge potential for enhancing pharmaceuticals & developing new treatments by emphasizing α-taxonomy.

So α-taxonomists will continue have an enormous impact on biology in general, as in biomedicine. Links between biodiversity and human health are easy to identify and we're just scratching the surface.

B. Provide specimen identification. Systematists provide id's for both the public and other scientists.

This may be accomplished indirectly, for example, through the publication of keys.

It may be accomplished directly. Specimens are sent to systematists typically associated with museums for ids. In fact, the USDA has several systematic entomologists whose role is to be available for high priority id of potentially damaging crop pests, especially those found on imported foods.

Members of the broader biological community (e.g., ecologists, physiologists) frequently aren't trained to be able to differentiate among very closely related species, and systematists verify their putative id's.

Mosquitoes can be very difficult to identify to species, but species differ in their role as a vector for malaria, West Nile virus, Dengue virus and other viruses. The US Army funds mosquito systematics labs that provide id (https://wrbu.si.edu/). These are often accomplished using genetic markers.

C. Maintain Collections – This is, in a way, related to the second role in that reference collections are often required, even by experts, to provide reliable identifications.

Collections also serve other critical roles; they serve as a permanent archive:

For type specimens, the actual physical material that was used in describing species.

For voucher specimens. All ecological studies should prepare voucher specimens. This allows future workers to verify work.

For other specimens. These serve as a permanent record of a species presence at a particular locality at a particular point in time. This is critical in that (1) it allows for the documentation of biotic change (range shifts), and (2) it can provide specimens from extirpated populations. In addition, these collections can serve as a source of genetic information for extirpated populations and species.

D. Classification – Order biological diversity into a hierarchical scheme of names.

Goal is to provide a framework of names that is universal (applies to all of life) and stable (changes over time as little as possible). This framework provides a means of communicating across languages and cultures. Most importantly, **this classification must reflect phylogeny**. In addition, classifications lose their function as vehicles for communication if they are not **stable**.

Taxonomic revisions should only be proposed/utilized if current taxonomy conflicts with strong phylogenetic evidence. That is, we should not alter taxonomy in an attempt to equilibrate ranks, for example so that no genus is older than, say, 2 MY (see Pauley et al. 2009. Herpetologica 65:115; for an opposing view, see Patterson and Norris. 2016. Mammalia 80:241).

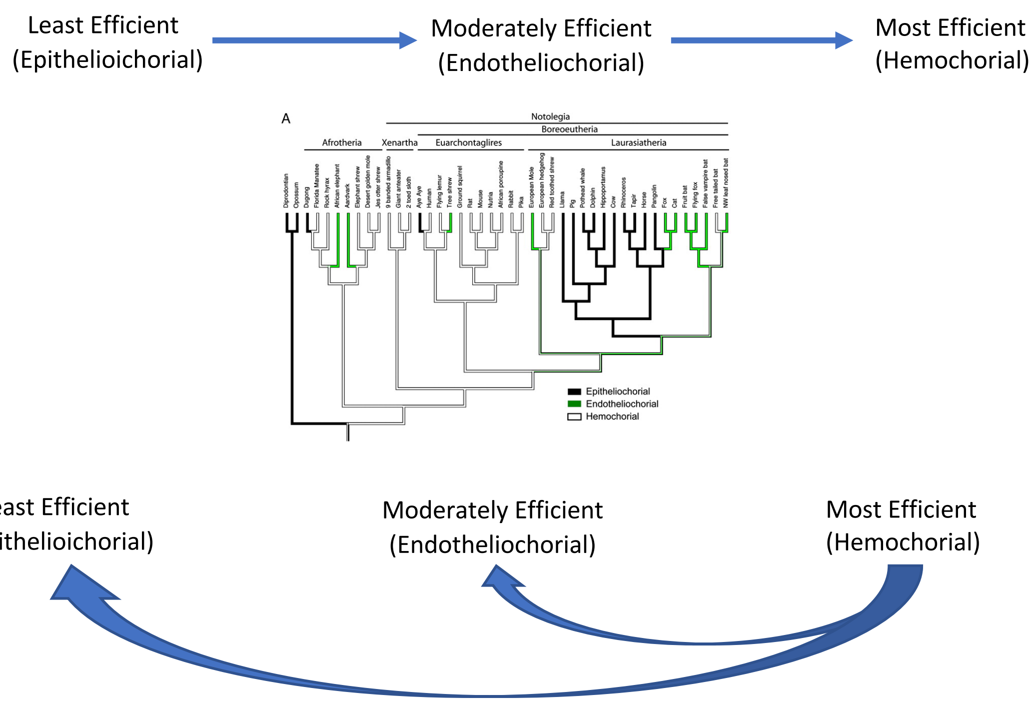
E. Phylogeny Estimation – Inference of the hierarchy of common descent.

This is the primary focus of this course, and lots of methods have been devised to attempt this. It's incredibly important – as of 01/09/23, a search of the Google Scholar database using "phylogeny" returned **>1, 370,000 papers** published.

In addition to its centrality as the basis of classification, other uses include:

1. Understanding Adaptation.

Placenta evolution in mammals (Wildman et al. 2006. PNAS, 103:3203).

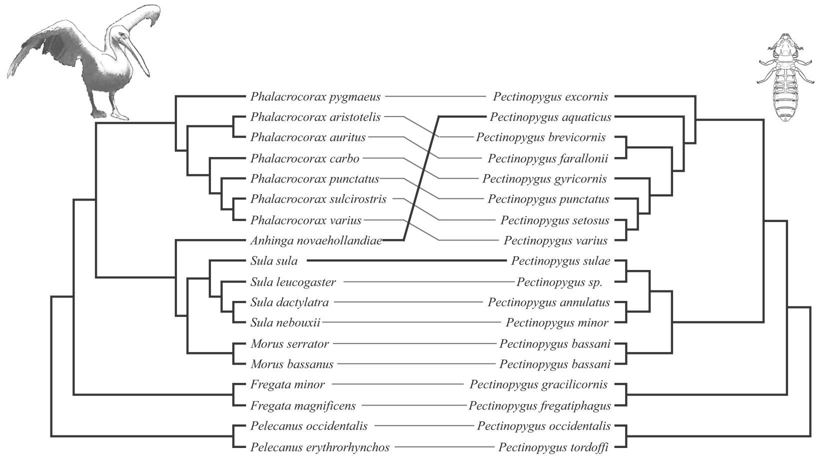


The less efficient placenta types are actually derived states and we need to explain their evolution from hemochorial ancestors.

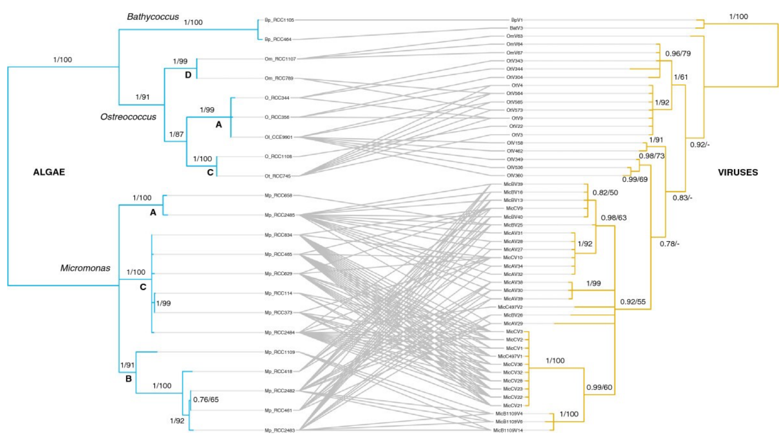
2. Co-speciation – for example between parasites and their hosts.

Parasitologists have long postulated that speciation in host-specific parasites should be linked to speciation of the hosts. This predicts that parasite phylogenies should mirror those of their host species.

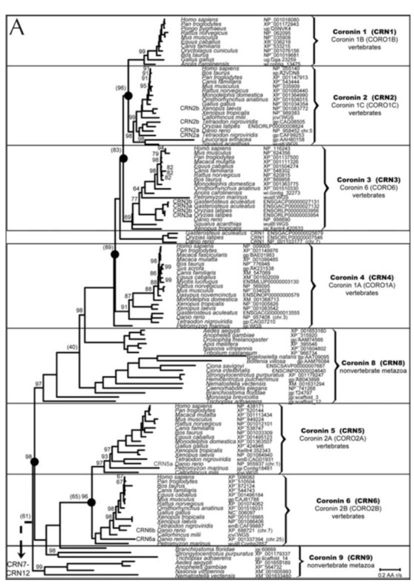
Sometimes that's the case, as in this study of Pelicanformes and their chewing lice (Hughes et al., 2007. Syst. Biol. 56:232). Nearly perfect congruence indicates a long history of co-speciation, with just two hypothesized host-switching events



Sometimes it's not the case, as in this study of algae and viruses (Bellec et al., 2014. BMC Evol. Biol., 14:59).



3. Gene Duplication and Genome Structure.



Gene duplication is a really common source of genetic variation, and lots of important proteins are coded by gene families.

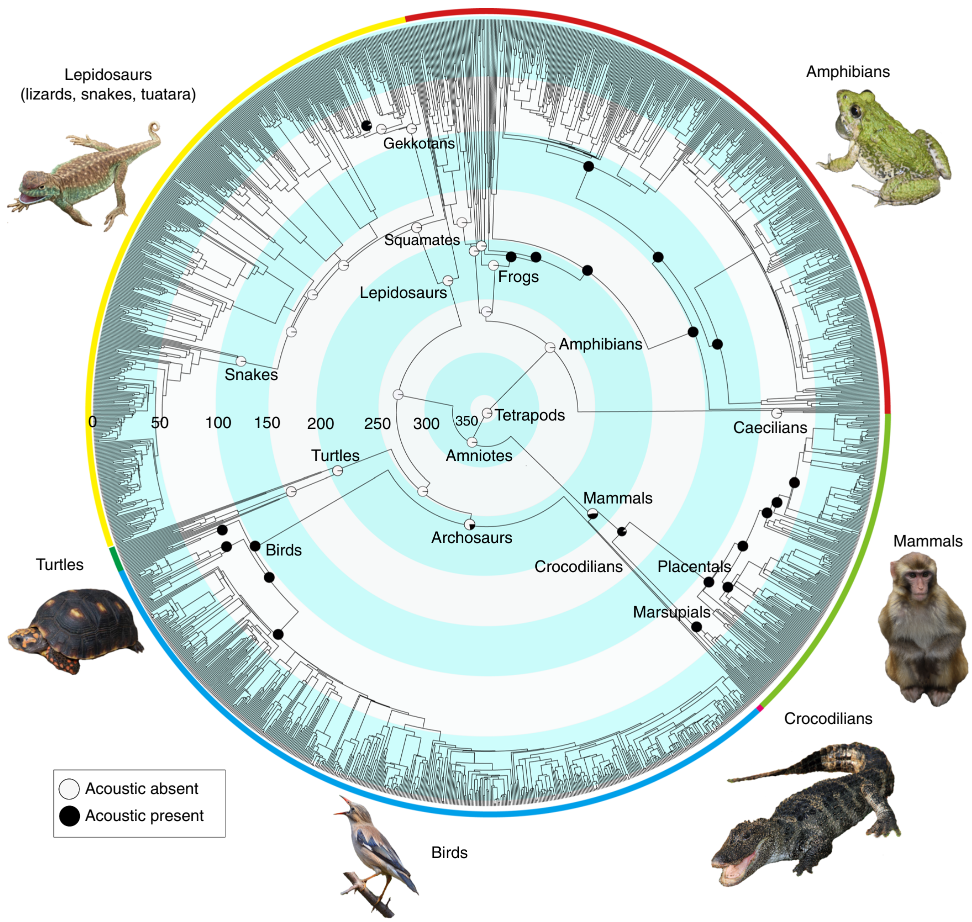
The only way to reconstruct the history of gene/genome duplication is via phylogenetic analyses, as show here.

This represents a phylogeny of the Coronin family and there have been at least 4 duplication events within vertebrates.

Such analyses can illuminate genome wide events as well. Morgan and Fernandez (2008. in *The Coronin Family of Proteins*, Clemen et al., eds).

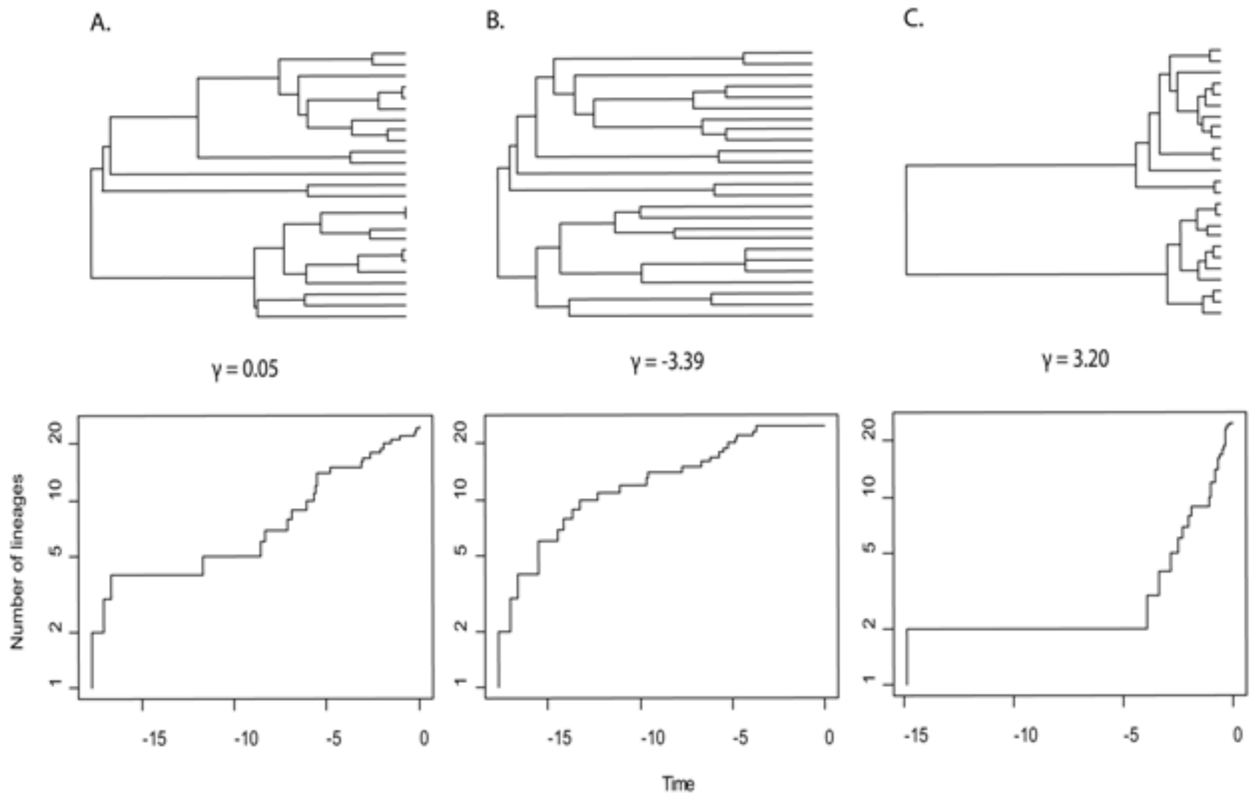
4. Correlation of Traits or Characters – Evolutionary biology has a long history of positing trait associations. Again, there are scores of methods to assess this quantitatively. For now, we'll use a simple example that was published a pretty recently (Chen & Weins. 2020. Nature Comm. 11:359).

They reconstructed the evolution of both acoustic communication and nocturnality in tetrapods. Acoustic communication evolved 4 times independently, each time in an ancestor that is reconstructed to have been nocturnal.



Each of these previous examples have involved primarily the shape of the trees and inferring evolutionary events (evolution of a trait, gene duplications, or cospeciation/host switching in comparisons of two trees).

5. Diversification Rates (e.g., lineages-through time) – We can make inferences regarding the net rate of diversification (speciation and extinction) for a group across time by looking at lineages through time.



A: Constant Rate B. Early Burst C: Long Fuse

(Myers & Burbrink. 2012. Nature Education Knowledge 3:23).

Note that the topology is the same, that is each tree implies the same relationships, but our inference of the rate of diversification differs.

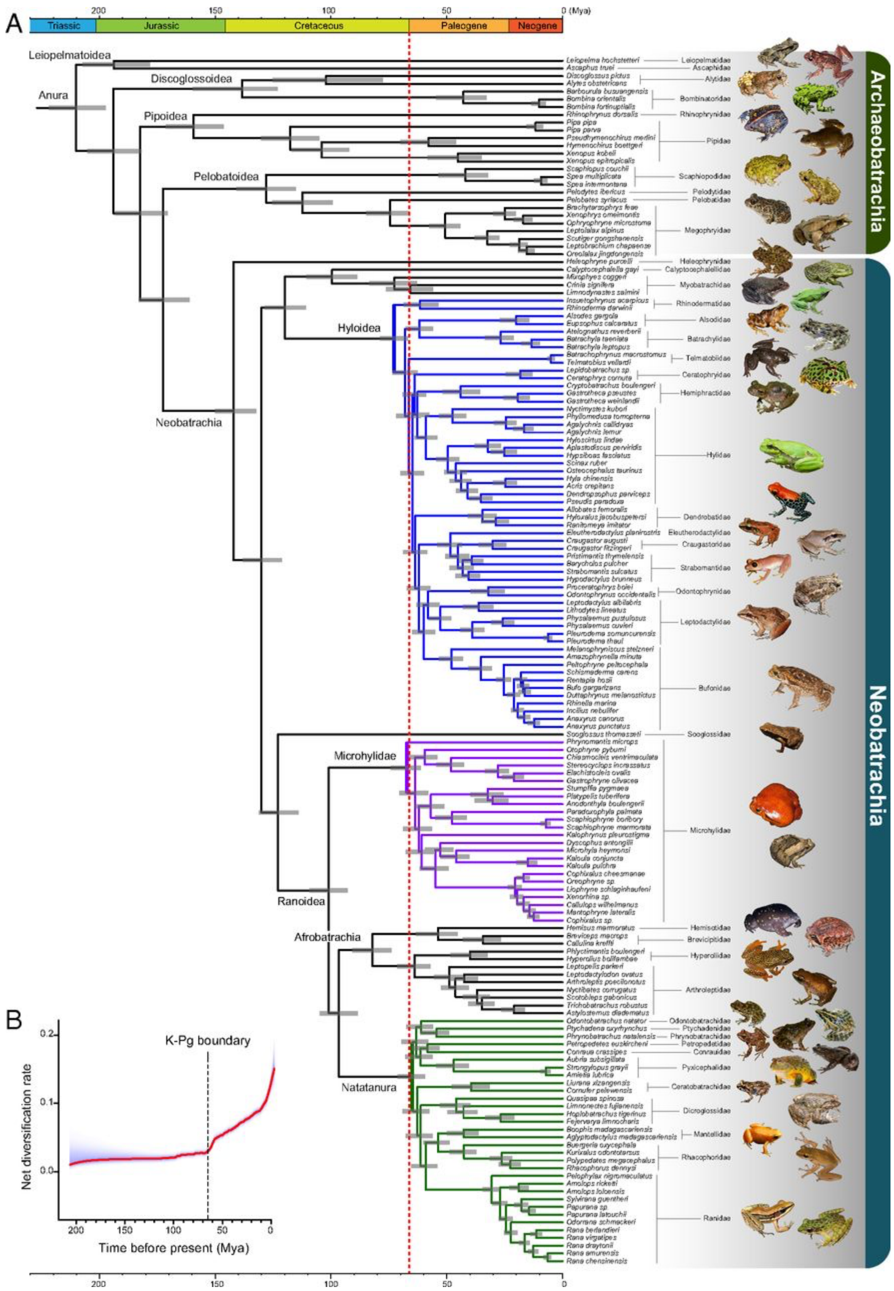
Again, these methods rely not only on the branching pattern, but how branch lengths are distributed across the tree.

6. Phylogenies in Conservation Biology – In addition to the establishment of ESUs, a large literature exists on the importance of Phylogenetic Diversity to conservation.

Given the ongoing biodiversity crisis and the limited resources that society allocates to conservation issues, a need arises to prioritize conservation efforts. Phylogenetic Diversity (PD) has been proposed as a means for determining these priorities.

The idea behind this is can be demonstrated with this phylogeny of frogs.

Feng et al. (2017. PNAS, 114:E5864)

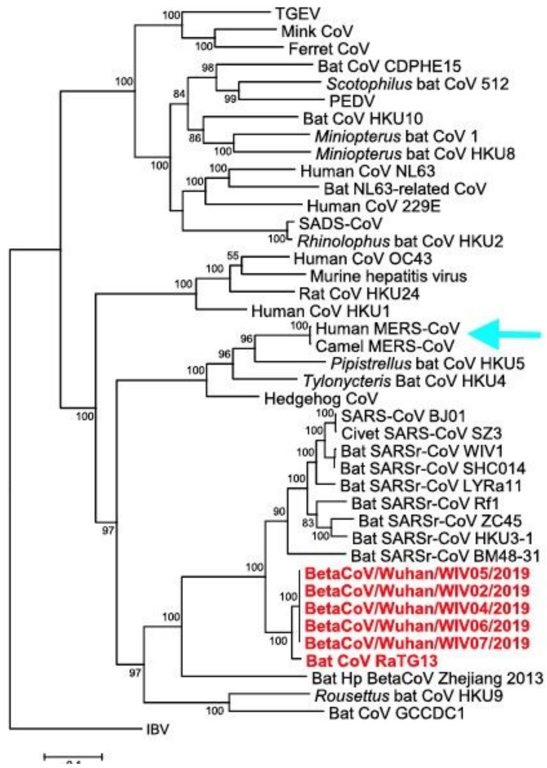


Tailed frogs (*Ascaphus*) share no common ancestors with other frogs for approximately 200 MY, whereas another local species, the Columbia spotted frog (*Rana luteiventris*) has a recent sister species (*R. pretiosa*) and has lots of other close relative.

Therefore, extinction of *Asaphus* would deplete PD far more than would extinction of *R. luteiventris*.

7. Human Health – Phylogenies are increasingly being applied to issues in biomedicine.

a. Identification of Pathogens – Where do emerging diseases originate?



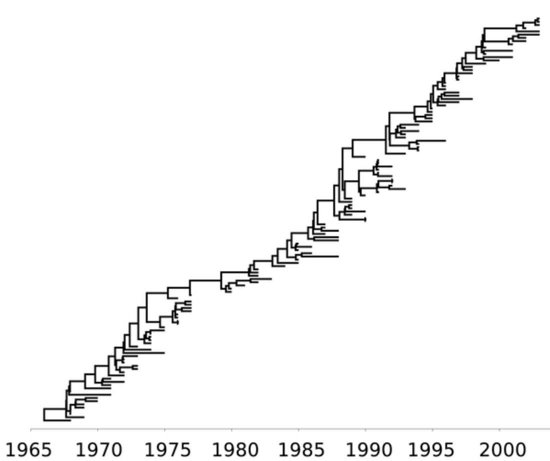
The recent outbreak of Coronavirus in China generated lots of media attention, as well as some strange speculation regarding its source.

A phylogeny that was estimated in January of 2020 (EcoHealth Alliance 11 January 2020) indicated that the Wuhan coronavirus (red) likely came from bats.

Furthermore, it appears that its sister-species is the previously known SARS virus, and this is why it's been named SARS-CoV-2.

This is important because it suggested to Pfizer that strategies that were designed for SARS-CoV-1 would likely be effective for SARS-CoV-2, and vaccine development was therefore really rapid.

b. Viral Phylodynamics – A growing literature uses phylogenies to assess how pathogens evolve (reviewed in Volz et al. 2013. PLoS Comp. Biol. doi:10.1371/journal.pcbi.1002947.g001; Geoghegan & Holmes. 2018. Genetics, 210: 1151)



Again, the shape of phylogenies can be really informative. If divergence is occurring by a random (birth-death) process, we expect phylogenies to tend to be balanced, whereas if selection is driving viruses to evolve away from a host immune response, we expect phylogenies to be imbalanced (often called pectinate). This will especially be the case with sampling that is spread over long (for a virus) evolutionary time.

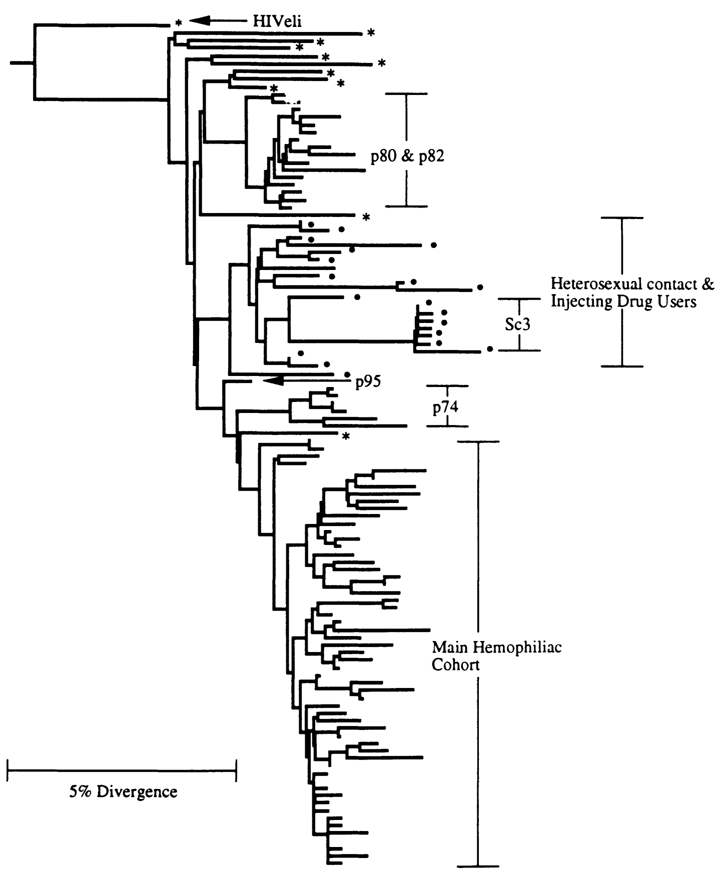
This phylogeny of sequentially sampled influenza illustrates this pattern really well.

It also illustrates why seasonal flu shots need to be taken every year; the virus evolves so fast away from our immune response that last year's vaccine is no longer effective.

c. Molecular Epidemiology – It's also been appreciated that, if we sample pathogens really densely from single host community, the virus phylogeny may inform us about disease transmission.

In the early 90's, during the peak of the AIDS epidemic, a number of hemophiliacs in Edinburgh contracted AIDS from blood transfusions.

Holmes et al. (1995. J. Infect. Diseases, 171:45) used a phylogeny of HIV sequences from these and several other local patients to examine the source of infection of the blood supply.



HIV+ hemophiliacs formed two separate clades.

Inferred that two donors infected the blood supply.

Heterosexuals were contracting HIV from IV Drug users.

Other examples include studies of Ebola (Dudas and Rambaut. 2014. PLoS Currents Outbreaks. doi: 10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.; Holmes et al., 2016. Nature, 538:193.) and Zika viruses (Dony et al. 2018. doi: 10.1371/currents.outbreaks.73b5c7d110f3bec90d75b2bb4dc9d23c.).

8. Forensic Evidence – Statistical phylogenetics have been used recently to connect a Louisiana physician to a patient who was the source of HIV used to infect his mistress intentionally (Metzker et al. 2002. PNAS, 99:14293). This was the first time that phylogenetics have been accepted as forensic evidence.

The broad range of disciplines that this represents has a number of effects.

First, it leads to the diversity of student backgrounds in classes such as this.

Second, it leads to the diversity of backgrounds of scientists who use phylogenetics in their research. If you look at papers that present some type of phylogeny, many of the analyses are *very* poorly done and sometimes results are simply misinterpreted.

Third, there has been a huge infusion of computer scientists and mathematicians into development of phylogenetic methods. Sometimes these folks are unaware of the historical development of the field, which, as I'll detail in the next lecture, has been contentious and absurdly vitriolic.