

Lecture 1 – Introduction: Importance of Systematics & History of Phylogenetic Inference

I. Definition of Systematics. Formerly, Systematics was 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.

The broad view is sometimes called **Biosystematics**, and it specifically includes the third.
(This term is much more commonly used by botanists than by zoologists.)

This broad sense of systematics could be defined 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 has 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 PEET program
(<http://www.nsf.gov/pubs/2000/nsf00140/nsf00140.htm>).

The fundamental importance of this aspect of a 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, from biodiversity that’s been documented already.

Some estimates are that we’ve only described 10% of the species that occur on earth. So, aside from the scientific importance, there is huge potential for enhancing pharmaceuticals by emphasizing α -taxonomy.

These estimates come from canopy fogging in Amazonia, in which a single tree is fogged with insecticide and entomologists try to identify what falls out. Only about 10% of the individuals that fall out belong to previously described species. This is probably a biased estimate, but even if so, α -taxonomists will continue have an enormous impact on biomedicine.

B. Provide specimen identification. We provide id’s for both the public and other scientists.

This may be accomplished indirectly, 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 malaria vectors. The US Army funds mosquito systematics labs that provide id. 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 id's.

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 specimens which serve as a permanent record of a species presence at a particular locality at a particular point in time. This is critical in that it allows for the documentation of biotic change. Provide specimens from extirpated populations.

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

D. Classification – Order biological diversity into a hierarchical scheme of names.
2.0 – 10.0 x 10⁶ species

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.**

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 01/09/07, a search of the ISI Web of Science database using “phylogen*” returned **45,501 papers** published this decade.

Two years later, 01/13/09, the same search returned **65,789** papers.

Examples of Applications (in addition to traditional use in classification):

Understanding Evolutionary Novelty.

Study of Evolutionary Patterns and Processes

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

Historical Biogeography – molecular phylogeography

Hybridization

Gene Duplication and Genome Structure

Macroevolutionary Patterns

extinction and speciation rates (e.g., lineages-through time)

species selection and the study of evolutionary trends

key adaptations that allow a group to be speciose

study of correlation of traits or characters – comparative biology

Conservation Biology

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

Molecular Epidemiology – Studies of infection pathways Holmes et al. (1995. J. Infect. Dis.). More recent examples include studies of SARS (Lan et al., 2005. J. Infect. Dis. 191:1478-1489).

Forensic Evidence – Statistical phylogenetics have been used 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 times that phylogenetics have been accepted as forensic evidence.

These are just a few of the many and varied uses of phylogenies, and you each could probably add to this list. 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.

III. The History of Phylogenetic Inference. Inference of evolutionary history has been of interests since the fact of evolution was established ca. 150 years ago. It's a discipline with a varied and checkered history and has been deeply embroiled in often-venomous debate.

Haeckel's evolutionary trees are among the first attempts at phylogeny inference.

For first century after Darwin, there really was no objective methodology to how biologists would derive such evolutionary trees.

Numerous explicit approaches have been proposed and there is very acrimonious debate as to the relative merits of each of these.

I think it's critical to have some familiarity with the history of this debate. Joe provides a unique historical perspective in Chapter 10, which is an important read. However, Joe has done a remarkable job in deleting much of the acrimony that has characterized this history of phylogenetics.

A. These approaches fall into several schools of thought.

1. Evolutionary Taxonomy – This traditional school was predominant until the 1960's. This is essentially the Haeckelian school and relies on appeal to authority. Joe has omitted discussion of this approach from his book, but we should definitely acknowledge the school, because of its long history and its history as the roots of the discipline.

The basic view was that the best way to reconstruct the phylogeny of a group is to spend a lifetime learning everything one can about the biology of the group.

The specialist (authority) then published a phylogeny based on his or her views.

This had the advantage of generating incredibly well schooled experts in the biology of particular groups.

There basically was no objective methodology, and the results were therefore not reproducible.

The effect that this school had on the perception of systematics was quite dramatic. As the modern synthesis took hold, systematics was marginalized to an attic discipline that was thought to have no important contributions to make to evolutionary biology.

However, it should be pointed out that many of the relationships that the ET's posited have been supported by application of modern phylogenetic methods. So, the phylogenies produced by at least some practitioners of ET were actually pretty good estimates, although there was no way to tell at the time.

2. Phenetics - In the early 1960's, a group of biologists attacked Evolutionary Taxonomy as non-scientific based on its lack of an objective methodology. This introduced the tradition of contentious debate that still persists.

The view of pheneticists such as **Sneath, Sokal, & Rohlf** was that **there is no way to infer the pattern of common ancestry of a group in a scientific manner**. Therefore, they argued, **we should group organisms on the basis of overall similarity**. As many measurements were taken as possible, and a variety of clustering methods were developed to arrange taxa in a similarity hierarchy.

Methods included UPGMA (we'll learn), and all kinds of clustering procedures, like Principal Components Analysis, Canonical Variates Analysis, etc. This really represented the origin of the science of clustering.

Since the view was that phylogeny can't be inferred reliably, the goal was to produce a phenogram. This was not meant to depict evolutionary relationships, but to visualize a hierarchical pattern of overall similarity.

In the 1960's pheneticists criticized the E.T.s for their lack of reproducibility.

3. Cladistics - The level of vitriol in the debate was elevated to personal attacks by the cladists. Its adherents often use the term **Phylogenetic Systematics**, and this is meant to imply that other approaches are not phylogenetic.

In the late 1960's, the work of the German entomologist Willi Hennig was translated English. Hennig is credited with founding the school of Cladistics, although others were important as well.

Hennig's view was that it is in fact possible to infer common ancestry in a scientific manner.

Attributes that are **derived** and **shared** by a set of taxa are *prima facie* evidence for exclusive common ancestry (**Synapomorphies**).

So the goal of cladists is to uncover monophyletic groups or clades, by finding synapomorphies.

Only derived characters (**apomorphies**) can be phylogenetically informative, so in the Hennigian approach, the first step is to determine which character states are derived and which are primitive; this is called determining character polarity.

One of my favorite simple illustrations of the differences between phenetics and cladistics is that of a Cow, a Trout, and Lungfish (which sounds like a bad joke...).

The Hennigian approach is based on identifying a synapomorphy that defines a group.

However, if we look at many characters, there will be conflict. The only methodology that is permissible to cladists to resolve such conflicts is the method of **parsimony**. That is the tree that requires the fewest evolutionary steps is to be preferred. We'll address parsimony as an optimality criterion in great detail later.

Early cladists such as J.S. Farris and Arnold Kluge demonstrated that **the MP tree is that tree which maximizes synapomorphies**, and therefore is the tree on which groups are defined by shared derived characters to the greatest extent possible. That is, the MP tree defines clades following Hennig's principles. Thus, cladists view MP as the only justifiable means for distinguishing among possible trees.

D. Statistical Phylogenetics – In spite of the view among cladists that statistical approaches to phylogeny are anathema, Cladistics has a distinctly statistical origin that Joe describes in Chapter 10. Especially relevant were the pioneering works of Anthony Edwards, and Felsenstein has certainly been one of the most influential and important proponents of statistical phylogenetics.

The perspective that phylogenetic analysis is a statistical endeavor languished in the shadows of irrespectability for several years.

Nevertheless the dominant paradigm in phylogenetics currently is that phylogenies are estimated with uncertainty and that one must quantify that uncertainty.

We can look at the above example for the statistical approach. We have these alternative trees, and we could take either a frequentist or a Bayesian approach to evaluating them. For example we could force the data to fit them both and see if one fits the data significantly better than the other. If that's the case we can reject one of them.

Because statistical inference requires a model, this school tends to rely on model-based methods such as **maximum likelihood** and **Bayesian** estimation, although **statistical parsimony** is also applied.

Cladists have attacked the new school with the same animosity that they attacked phenetics in the 70's.