

"PRINCIPLES OF PHYLOGENETICS: ECOLOGY AND EVOLUTION"

Integrative Biology 200B  
University of California, Berkeley

Spring 2011  
B.D. Mishler

March 1, 2011. **Molecular Evolution**

Aims: (1) phylogeny reconstruction (2) study of evolution at the molecular level per se

Mutation, recombination, and gene conversion

Transposable elements

Repetitive elements (microsatellites)

Natural selection vs neutrality  
detecting selection

Different genomes  
organellar vs. nuclear  
gene movement

Comparing genomes  
synteny, rearrangements, indels  
the "annotation" problem  
Exon shuffling  
Introns (different types)

Multigene families  
paralogy vs orthology

the fate of duplicated genes: ghost  
genes, subfunctionalization

Polyploidy - genome size

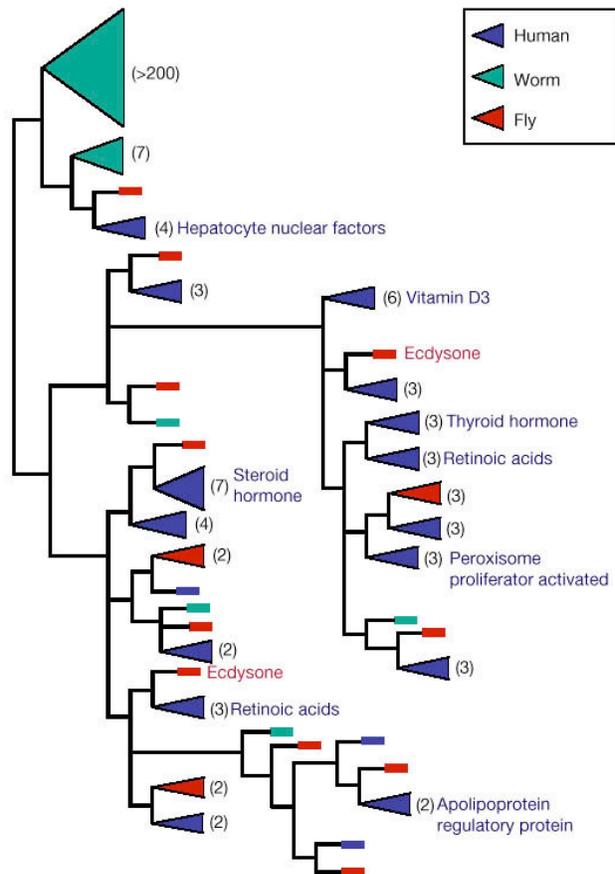
RNA and secondary structure

Transition-transversion bias  
purines A G  
pyrimidines C T U

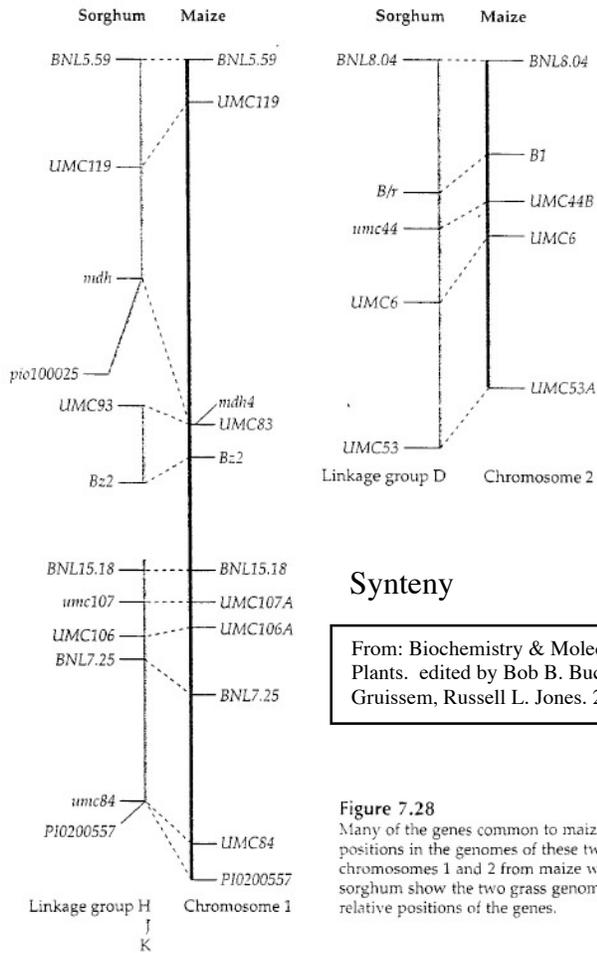
Protein evolution ("proteomics")

Codon usage bias

G+C content



Simplified cladogram of the 'many-to-many' relationships of classical nuclear receptors. Triangles indicate expansion within one lineage; bars represent single members. Numbers in parentheses indicate the number of paralogues in each group.

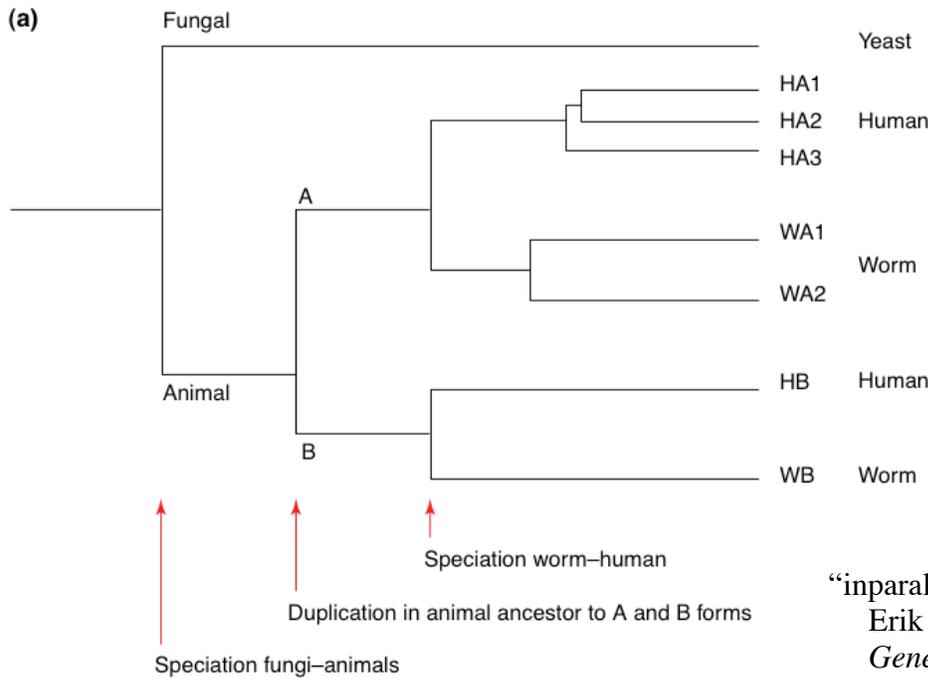
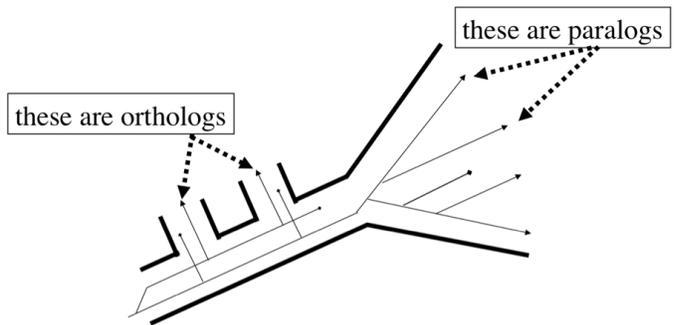


**Syteny**

From: *Biochemistry & Molecular Biology of Plants*. edited by Bob B. Buchanan, Wilhelm Gruissem, Russell L. Jones. 2000.

**Figure 7.28**  
 Many of the genes common to maize and sorghum are organized into similar positions in the genomes of these two plants. A comparison of genetic maps of chromosomes 1 and 2 from maize with the maps of several linkage groups from sorghum show the two grass genomes similarly organized with respect to the relative positions of the genes.

**Orthology vs. paralogy**



“inparalogs” vs. “outparalogs”  
 Erik Sonnhammer. *Trends in Genetics* Vol.18 No.12, 2002

## Annotations & Ontologies

Some things to think about, in the “annotation” process. What does it mean to say I have the “same” or “related” genes in two different genomes? Or for that matter, the “same” or “related” genes in the same genome? Three ways to go:

- Name gene haphazardly by whatever criteria the discoverer thinks best -- common practice, unfortunately!
- Name gene by functional criteria
- Name gene by phylogenetic criteria

The need for ontologies (a formal classification)

### Ontologies in general

- Naming things, and organizing them in databases, is critical in all mature sciences
- Need for frameworks for understanding
- Two organizing forces in biology:
  - current function
  - history (homology)
- Uses of cladograms for untangling these
- The role of systematics in relation to molecular, cellular, and developmental biology -- once estranged, now vitally interlinked.

### Approach taken by Gene Ontology Consortium:

“The Gene Ontology project provides an ontology of defined terms representing gene product properties. The ontology covers three domains: cellular component, the parts of a cell or its extracellular environment; molecular function, the elemental activities of a gene product at the molecular level, such as binding or catalysis; and biological process, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms. For example, the gene product cytochrome c can be described by the molecular function term oxidoreductase activity, the biological process terms oxidative phosphorylation and induction of cell death, and the cellular component terms mitochondrial matrix and mitochondrial inner membrane.”

From: <http://www.geneontology.org/index.shtml>

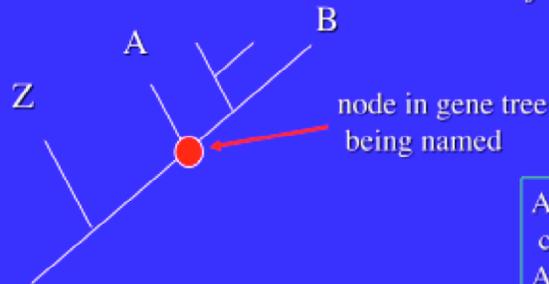
### What would a phylogenetic approach look like?

- We need to add a gene ontology reflecting history!
- This would not be to the exclusion of functional ontologies, but rather an addition.
- We want to be able to look at function and history in light of each other
- The classic homology - analogy distinction

### A Phylogenetic Classification of Genes

- new proposal: need a unique phylogenetic identifier for each gene and gene clade (distinct from the associated taxon name!!)
- internal and external specifiers (other named genes)
- registered in a data base (GO associated?)
- Parallel to the developing Phylocode for taxonomy of organism lineages (an interesting and unanticipated convergence)

The main contribution of the PhyloCode is to provide an unambiguous way to name clades: *this could work for gene clades!*



A node-based name:  
 "I name the gene clade that contains A, B, and all the descendents of their most recent common ancestor"

A, B, and Z here are called *specifiers*: A & B are *internal* specifiers, while Z is an *external* specifier. In this system, these would be *genes*.

What about phenotypes?

- Like genes, need a primary name, plus inclusive classifications.
- Primary name, by analogy with genes (e.g., GenBank accession), should be a neutral identifier .
- Linked to a specific data point (in a particular organism, ideally documented with a voucher specimen), i.e., photo, video, physiological measurement, anatomical prep, plus metadata.

Classification could be based on:

- development (e.g., "seedling," "anthesis")
- location (e.g., "axillary," "basal")
- function (e.g., "leaf," "scale," "stem," "spine")
- **history** (e.g., "microphyll," "phyllid")
- structure per se?? (probably not a good idea)
- Classifications can easily be cross-cutting, but basis of term needs to be clear to computer (and user!); meta-tags

The historical criterion for phenotype ontologies needs work

- Based on homology, thus based on current best hypothesis of phylogenetic tree.
- Therefore subject to change, as phylogenies change.
- Needs to be clearly specified (i.e., linked to a specific clade) –a PhyloCode-type approach could be used to triangulate to clade where name applies
- When PhyloCode is active, phenotype ontologies could reference RegNum for clade names.