Overview of Lecture: Genomes & Phylogenies
Read: Text Ch 22

Bullet Points:
• Phylogenetic CSI – who infected whom?
• A brief history of life on earth
• Trees of life -
  what do they look like?
  how do we infer them?
• Shared derived characters (and some jargon)
• Homologous vs analogous (convergent) traits
• Orthologous vs paralogous DNA sequences (are both homologous)
• Retention of function in orthologs (yeast and humans)
• Diversification across paralogs (ex myoglobin & hemoglobin)
• Copy number variants and human variation
• Diet and amylase copy number in humans and dogs
• Whole Genome Duplication
  and the evolution of complexity and vertebrates
Learning Goals:

1. Understand and be able to use examples to explain …
   What are “shared derived characters” and how are they used to infer phylogenetic trees?

2. Understand and be able to use examples to explain …
   The meaning of and distinction between analogous and homologous traits-characters. Which of these (analogous, homologous) are useful for inferring ancestry and which tends to confuse inferences about ancestry?

3. Understand and be able to use examples to explain …
   The meaning of and distinction between orthologous and paralogous genes-loci. Which of these arise from gene duplication? Describe an example where variation in the number of copies of a gene-locus is associated with variation in phenotype in humans.

4. What is Whole Genome Duplication (WGD) and why is it thought to facilitate the evolution of complexity?
Phylogenetic CSI – who infected whom?

A powerful method for deducing microbial relationships [phylogenetic cladistics] has been edging its way into civil and criminal investigations. But courts should proceed with caution. [conclusions are probabilistic - statistical inferences from data - is p < 0.05 "beyond reasonable doubt"?]

Shaoni Bhattacharya 26 February 2014

Infectious forensics

1. Pathogen genomes can mutate quickly, creating diverse microbial populations in those infected.
2. By sequencing highly variable regions of pathogen genomes, scientists can build a phylogenetic tree that suggests how the microbes are related.
3. The relatedness of the viral populations can support or rule out hypotheses of who infected whom.
4. Pathogen diversity can also be used to corroborate time of infection using the mutation rate as a molecular clock.

Treat these “populations” like “species” or DPSs or ESUs
“Big Bang” ~ 14bya

- Created by photosynthetic cyanobacteria
A phylogenetic tree or evolutionary tree is a branching diagram showing the inferred evolutionary relationships among various biological species based upon similarities and differences in their physical or genetic characteristics.
Note: **the root** is a complex bunch of prokaryotes; "a gene-swapping mêlée"
Molecular genetics yields increasing resolution - related to the problem of identifying ESUs and "species"
Species that share a common ancestor, as indicated by the possession of shared derived characters, are said to belong to a clade. Clades are thus evolutionary units and refer to a common ancestor and all of its descendants. A derived character shared by clade members but absent from sister clades is called a synapomorphy.

Also, 4 limbs “tetrapods”
In addition to fossil organisms, phylogenetic history can be inferred from certain morphological and molecular similarities among living organisms. ... **similarities due to shared ancestry are called homologies.** ... the similarity in the number & arrangement of bones in the forelimbs of mammals is due to their descent from a common ancestor with the same bone structure;

---

**Figure 26.8** Aligning segments of DNA. Systematists use computer software to find and realign similar sequences along DNA segments from two species.
A potential red herring [misleading data] in constructing a phylogeny is similarity due to convergent evolution—called analogy—rather than from shared ancestry (homology).

Convergent evolution occurs when similar environmental pressures and natural selection produce similar (analogous) adaptations in organisms from different evolutionary lineages.

Consider ‘wings’ vs ‘forelimbs’ of bats & birds. They have separate evolutionary origins, but are superficially similar because they evolved to serve the same function.

... though bird and bat wings are analogous as wings, birds & bats did not inherit wings from a common ancestor, but they did inherit forelimbs from a common ancestor with forelimbs. [and “four limbs” = tetrapods]

{ are the bones in the forelimbs of birds & bats homologous? }
Blonde hair evolved more than once

Golden locks of dark-skinned Melanesians have different genetic basis to those of Europeans.

About 5–10% of people from Melanesia have naturally blonde hair - the highest prevalence outside Europe. Yet people from the region have the darkest skin outside Africa. Now, a study of people from the Solomon Islands shows that this refutes the possibility that blonde hair was introduced by colonial Europeans ...

"Blonde hair has clearly evolved twice," says Carlos Bustamante, a co-author. To investigate the genetic basis underlying the trait, ... compared the genomes of 43 blonde and 42 dark-haired Solomon Islanders, and revealed that the islanders' blonde hair was strongly associated with a single mutation in the TYRP1 gene. Several genes are known to contribute to blonde hair coloration in Europeans, but TYRP1 is not involved. By comparing DNA between >900 Solomon Islanders and >900 other individuals from 52 populations around the world, the researchers found that the TYRP1 mutation is probably unique to the Oceanic region ...

{ So, the trait "blond" in Europeans and Melanesians is ... what? }
Homologous DNA sequences.

**Orthologs** and **Paralogs** are two types of homologous DNA sequences.

**Orthology** describes homologous genes in different species that derive from a common ancestor. Orthologous genes may or may not have the same function.

**Paralogy** describes homologous genes within a single species (an individual) that diverged by gene duplication.

The genes encoding the α chain and the β chains in your hemoglobin molecules are ancient paralogs.

{many human hormones come in “families” that are coded by paralogous genes}

Humans have huge families of more than 1,000 paralogous olfactory receptor genes.

{many are degenerate ‘pseudogenes’}

The genes encoding the α chains in you and in your dog or cat are orthologs.

99% of the genes of humans & mice are orthologous, 50% of our genes are orthologous with those of yeast.

How far across evolution do families of genes retain their function? **Yeast and humans** are separated by roughly a billion years of evolutionary history,

[these coding DNA sequences have diverged some, but not enough to change protein function]
Gene families are groups of homologous [paralogous] genes …

Differences in [gene] family size due to lineage-specific gene duplication and gene loss may provide clues to the evolutionary forces that have shaped mammalian genomes. 

Gene duplication is … a powerful engine for evolutionary change …

Changes in gene family size have likely been important during human evolution …

Humans show extensive polymorphism in the number of copies of chromosomal segments, and of genes included in those segments.

We analyze gene families [in] genomes of human, chimpanzee, mouse, rat, and dog.

We find 30 [gene] families where the lineage leading to modern humans contains the most significant departure from random [a null model] gain and loss of genes.

One of the families evolving significantly faster than expected in humans contains the gene CNTN2 [CNTN2 – influences neural synaptic structure], which has been implicated in the genetic etiology of autism.

[see Association of Copy Number Variations in Autism Spectrum Disorders: A Systematic Review. EF Sener 2014 ]
Bitter taste responses in humans are mediated by bitter taste receptors ... encoded by the TAS2R gene family

Copy-number assays ... targeting TAS2R43 revealed

Copy-number variation ... has the potential to have extreme effects on phenotypes.

... allelic \{copy number\} variation predicts

Copy Number Variants: A New Molecular Frontier in Clinical Psychiatry
It is now clear beyond any reasonable doubt that genetic inheritance influences liability to develop almost every major psychiatric disorder.

... our growing knowledge regarding the roles of copy number variants in behavioral disorders will soon require revision of standards of evaluation and care for psychiatric patients. \{possibly related to autism\}
Diet and the evolution of human amylase gene copy number variation

Starch consumption is a prominent characteristic of agricultural societies ... In contrast, rainforest and circum-arctic hunter-gatherers and some pastoralists consume much less starch. This behavioral variation [in diet] raises the possibility that different selective pressures have acted on amylase, the enzyme responsible for starch hydrolysis. [recall the lactase story]

We found that copy number of the salivary amylase gene (AMY1) is correlated positively with salivary amylase protein level [phenotypic differences are associated w/ genetic (copy #) differences] and that

Higher AMY1 copy numbers and protein levels improve the digestion of starchy foods
Adaptive drool in the gene pool

John Novembre¹, Jonathan K Pritchard¹ & Graham Coop¹

1. John Novembre, Jonathan K. Pritchard and Graham Coop are in the Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. e-mail: jnovembre@uchicago.edu or e-mail: gcoop@uchicago.edu

A new study finds that copy number variation in the salivary amylase gene in humans is associated with amylase concentration in saliva and average starch consumption in populations. This provides a striking example of the role of copy number variants (CNVs) in adaptive evolution, and of diet in producing selective pressures.

See example student papers: (links are on the course Entrance page)
http://www.msu.edu/course/bs/182h/fall2018/N&Vexamples.pdf

The grading rubric:
Prior to the advent of modern dog-feeding practices, there was likely substantial variation in dietary landscapes among disparate dog breeds. [related to the diets and table scraps of their human companions]

We investigated one type of genetic variant, copy number variation, in ... pancreatic α-amylase 2B (AMY2B).

It has been previously demonstrated that dogs experienced a copy number increase in AMY2B relative to wolves during or after the dog domestication process. We demonstrate that positive selection continued to act on amylase copy number in dog breeds that consumed starch-rich diets in time periods after domestication.

Organisms with additional [redundant] sets of genes can accumulate and test mutations much faster and with less selection than organisms with just one set of genes. [redundant paralogs can diverge, no cost and possibly great benefit]

One of the copies of a gene can maintain normal functioning of the cells even if the other copy mutates to become harmful or useless.

Other possible alternatives are when
one of the genes acquires a completely new function,
or both genes start to specialise,
each taking over a certain part of the ancestral function.

Two sets of Whole Genome Duplication in the ancestral “root” vertebrate; (a 3rd WGD in teleost “bony” fish) may have facilitated the evolution of complexity in vertebrates.