Suggested project topics Phylogenetics (03-727)

Project plan: 1 page, due November 8th

The following are possible phylogenetics projects for teams of 1 or 2 people. Normally, only one team will work on each project, unless there is a natural way to split the project into separate sub-projects. For example, several teams could work on the nitrogen fixation project, each analyzing a different *nif* gene family. Topics will be assigned on a first-come-first-serve basis. Let me know once you choose a topic to investigate to make sure it is still available. I can also help you find a project partner.

Mysteries of Yeast metabolism

Baker's yeast (*Saccharomyces cerevisiae)* is one of the most extensively studied model organisms, yet there are still over 800 yeast genes (~10% of predicted yeast genes) whose function remains a mystery. Dr. Amy Caudy, a researcher at the University of Toronto, is using systematic genomic screens to characterize the molecular function of these genes. She has three candidates:

We understand very little about yeast proteins *YMR278W* and *YMR027W*. Dr. Caudy has preliminary evidence that YMR278W which breaks down ribose-1-phosphate and YMR027W breaks down fructose-1-phosphate. She says, "We don't know what ribose-1-phosphate and fructose-1-phosphate are doing in metabolism! More information about the evolutionary history and conservation of these proteins could shed some light on the matter."

Almost nothing is known about *YNR065C*, but the Caudy lab recently discovered that it accumulates a 6-carbon organic acid to extraordinarily high levels. This gene is a member of a family of paralogous genes, and before Halloween the lab will know whether any other mutants in the family have the same metabolic phenotypes. It would be useful to understand the relationships of these paralogs, and their evolutionary history in fungi and beyond.

Cristallins and Spherulins

The vertebrate eye, which has a refractory lens, represents a morphological innovation that makes it fundamentally different from the light sensing organs of invertebrates. The crystallin gene family, which encodes essential components of lens of the vertebrate eye, is only found in vertebrates. Cristallins are members of a large superfamily of proteins with similar folds, but different functions. In particular, crystallins are soluble at very high concentrations, which is essential for their function in the eye. Other members of this superfamily are not soluble in this way. One theory is that some random protein was coopted for this role and then evolved to fill it. Spherulin, a protein which is found in fungi and bacteria and is structurally similar to crystallins, may be the closest relative to crystallins. A phylogenetic analysis of crystallin and spherulin sequences could help to explain the origin of crystallins and the evolution of vertebrate vision. This could have further implications for our understanding of protein solubility and of blindness arising from mutations that reduce crystallin solubility.

Coevolution of Bhlhb5 and Prdm8, two proteins involved in itching

Got an itch to align some proteins? Scratch it with this project! In a recent publication (Ross, *et al*., Neuron, 2012) it was shown that mice with knockouts of either Bhlhb5 or Prdm8 developed a similar phenotype of incessant scratching. Bhlhb5, an activating transcription factor, is repressed by Prdm8. Similarly, Zfp488, a phylogenetically close relative of Prdm8, represses an Olig transcription factor that is a phylogenetically close relative of Bhlhb5. These phylogenetic relationships may help to explain where Prdm8/Bhlhb5 and the itch response came from. In this project we'd like to try to identify homologs and orthologs of these protein pairs to expand on research explaining itch in humans and other organisms.

Opines

Crown gall and hairy root tumors in plants are caused by pathogenic bacteria from the genus *Agrobacterium.* Agrobacteria trick the plant into synthesizing low molecular weight compounds called *opines,* which act as nitrogen and energy sources for the bacterium. Agrobacteria carry opine synthesis genes which they insert into the host plant genome during the infectious process. Surprisingly, opines are also found in octopus muscle and other apparently unrelated contexts. An evolutionary analysis of opine biosynthesis genes could help explain where these mysterious compounds come from.

Nitrogen fixation, convergent evolution and horizontal transfer

In order to synthesize bio-organic compounds that contain nitrogen, cells require nitrogen in a form they can use. Many plants solve this problem through a symbiotic relationship with bacteria that convert nitrogen in the atmosphere into ammonium, a process called nitrogen fixation. Nitrogen fixation is carried out by the nitrogenase enzyme system, which is encoded by the *nif* genes. Interestingly, the ability to fix nitrogen is found in diverse groups of bacterial species that are distantly related to each other. Thus, distantly related species that fix nitrogen share a set of genes that is absent from their closest neighbors. Did nitrogen fixation evolve several times independently through convergent evolution? Is nitrogen fixation an ancestral phenotype that was lost in many lineages, but retained by others? Are nitrogen fixing bacteria exchanging nif genes by horizontal transfer The goal of this project is to investigate these questions by reconstructing the evolution of one or more nif gene families and comparing the resulting trees with the bacterial species tree.

Evolution of a Virulence Determinant in Human Pathogens

Many bacterial species code for proteins in the large SLR multigene family. While the function of many of these genes is still unknown, in some human pathogens such as *Neisseria*, *Legionella*, and *Helicobacter*, the SLRs have been shown to interact with the human host. A recent doctoral dissertation, in collaboration with Dr. Hiller, has shown that *Haemophilus influenzae* SLRs play a key role in virulence. In *H. influenzae*, there are seven SLR subfamilies. These all share similarity, but can nonetheless be divided based on variable amino acid sequence. Individual strains vary dramatically in that they code for variable numbers of SLR genes. The specific subfamilies represented also vary greatly from strain to strain. Interestingly, for some subfamilies, there is higher sequence similarity across species than within strains of the same species. Your challenge is to use phylogenetic analysis to investigate this puzzling observation. The results could also provide clues to contribution of individual subfamilies to virulence.

Developmental proteins in your sea food

Dr. Hinman and her students are using transcriptome sequencing to characterize the expression of genes that are differentially regulated during development in the sea cucumber, an intriguing sea creature that spits out its entire gastrointestinal tract when frightened. From this data, the Hinman lab has identified a set of candidate regulators of exoskeleton development. The sea cucumber is an echinoderm, related to the sea urchin and the starfish. Like the sea cucumber, the sea urchin has an exoskeleton. The starfish, however, does not. Does this mean that these candidate genes are absent from starfish, or are they present, but playing a different functional role? The goal of this project is to explore this question via a phylogenetic analysis of one or more of these candidate genes and its relatives in other echinoderms.

Yeast orthologs – or maybe they aren't orthologs?

In 2003, Rokas, et al. published an analysis of phylogenetic trees from 106 orthologous gene families in Nature. This study was touted as a break-through in genome scale phylogenetic analysis. However, the spatial arrangement of these genes suggests that a number of the families (e.g., YBL091C, YDL031W) include paralogous genes as well. Obtaining accurate yeast phylogenies is challenging. These species arose through several rapid speciation events in succession, providing little time between divergences for mutations to accrue. As a result, the phylogenetic signal is weak. Can new phylogenetic methods reconstruct a branching pattern that is consistent with the spatial evidence?

Aging in Drosophila

Professor Javier Lopez is interested in a phylogenetic analysis of genes encoding the mitochondrial F1 ATPase epsilon subunit, which play a role in aging. In *Drosophila melanogaster* there are at least three sequences that encode an epsilon subunit (two are products of one gene by alternative splicing; the third is a product of a separate gene). Interestingly, Dr. Lopez has found that they have different tissue expression and age-dependent expression. Further, according to his observations, this protein has two functional roles: as a structural coupling factor in oxidative phosphorylation and as a ligand for the lifespan-determining GPCR methuselah. However, the relationship between the two functions and the three sequences is unknown. A phylogenetic analysis of the epsilon subunit sequences would provide additional information about this family and would also shed light on the origin of this system.