Mutation Mapping

Develop a pipeline to identify mutations a given collection of isolates, strains generated from a mutagenesis experiment.

For a plant project: e.g. some data from "Next-generation forward genetic screens: using simulated data to improve the design of mapping-by-sequencing experiments in Arabidopsis"

Or try a smaller genome example in bacteria. Do this for point mutations - can you identify the specific candidate changes based on analysis compared to a reference genome? Can you identify mutational biases - eg if the mutagen was UV vs EMS can you identify the mutational bias or pattern?

Superfolding

Deepmind's Alphafold was released in 2021 which promises to change how we do protein folding prediction.

There are established python notebooks which can run on google code allow you to provide your own running of these analysis. Develop a simple python notebook to fold a structure of a protein you are interested in.

Snakemake

For more advanced programmers, learn how to develop a snakemake pipeline to run on the HPCC.

Projects from year's past

Team Microbiome

1. Develop analysis pipeline to process Microbiome data from a published study. Focus on existing clustered "OTUs" already not the first steps
2. Metagenome analyses. Take existing, published metagenomes and compare abundance of specific enzymes classes, protein domains, or other content between the two. develop a report of the most abundant classes of domains and/or those which are most different between the two (or more) datasets.

Team Gene

1. Use a genome annotation of a set of species (plants, microbes might be best) to compare the differences in gene lengths, intron lengths, Untranslated lengths across the species. This is generating sumamry statistics about annotated genes.

Team Transcriptome

1. Identify a published RNAseq dataset for at least 2 conditions with replicate. Process the RNAseq to identify gene expression differences and identify if there are different functional classes of genes found in genes which are up or down regulated.

Team Proteins

1. Compare the protein content among sets of organisms. For example, Can develop a classifier for outlier proteins and/or ecological adaptations (thermophillic vs halophillic).
2. Examine the differences in Protein domain distribution between species.

Team Russian Doll

1. Many organisms have undetect symbionts of bacteria or viruses. Develop an analysis to look at sets of either assembled or unassembled bacteria to look for associated viruses (bacteria phage). Or develop a pipeline too look at assembled eukaryotic genomes and detect bacteria in the genome assemblies.