

## 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?

## Disordered

Develop script and pipeline to screen proteins for specific summary properties to look for [intrinsically disordered proteins](#) in a given set of proteins. Organize these by other properties such as whether they are conserved or unique among a species.

## Genome architecture

Develop set of scripts to generate summary statistics for genomes based on GFF/ GTF files and genomic DNA. Given a database (eg Ensembl, NCBI, FungiDB) where a set of GFF can be downloaded for a group of species, develop a comparative analyses to see how genome statistics vary among groups. For example GC content change, genome compactness (how small/large are intergenic regions as a fraction of the genome).