BIOINFOMATIC ANALYSIS FOR PROTEOMICS

FUNCTIONAL ANNOTATION AND ENRICHMENT ANALYSIS

Functional annotation is the process of attaching biological information to a gene or protein sequence. The functional annotation includes three main steps:
a. Identify the part of the genome that does not encode a protein;
b. Identify the elements in the genome (gene prediction);
c. Attach biological information to these elements.

Functional enrichment analysis determines the classes of gene or protein that is overexpressed in a large number of genes or proteins, and may be related to the disease phenotype. Statistical methods are often used to determine the significantly enriched genome. The general steps of enrichment analysis include:
a. Calculate a p value (the value represents the overexpression of the protein in the list);
b. Evaluate the statistical significance of nodes or paths according to the p value;
c. Normalize the p value of each set of protein and calculate the false discovery rate for multiple hypothesis tests.

01　GO Annotation Analysis

Gene ontology (GO) unifies the representation of genes and gene product attributes in all species.
Application:
A. Integration of proteomic data from different species
B. Classify the differential proteins
C. Predict the functions of specific protein domains
D. Identify genes involved in certain diseases
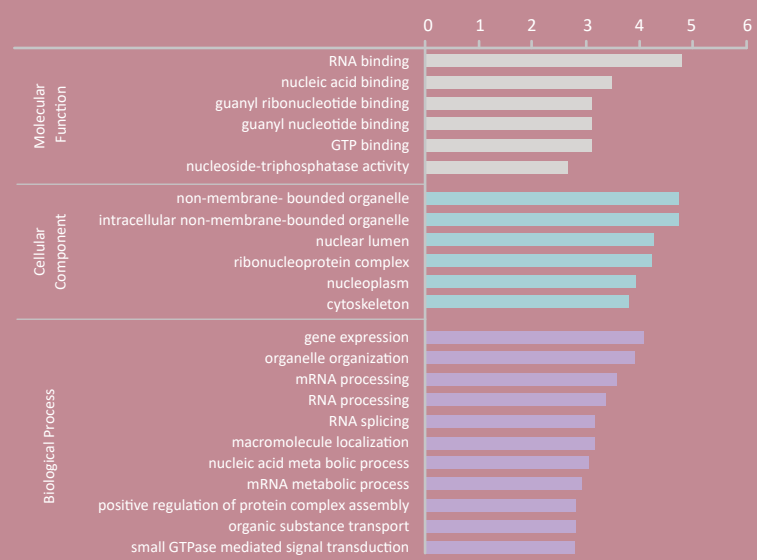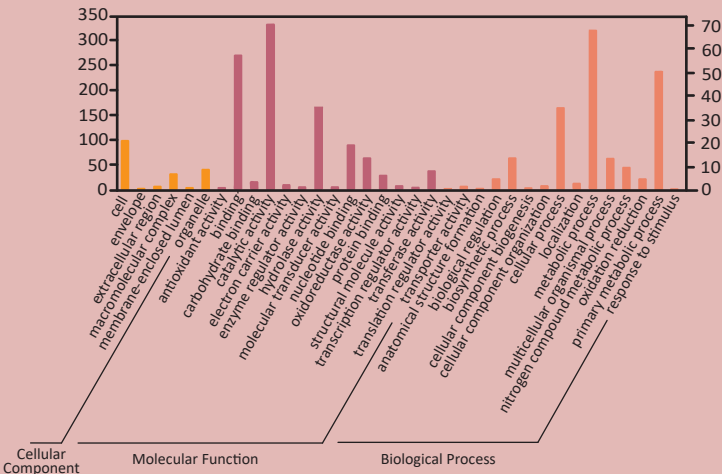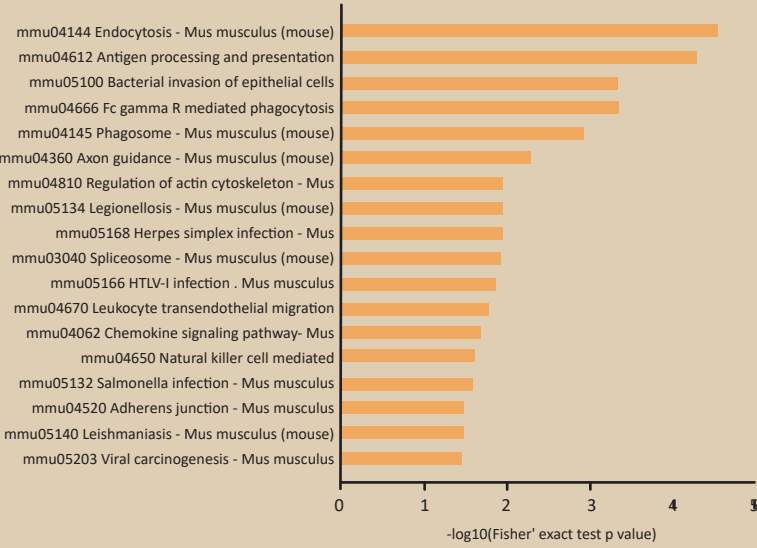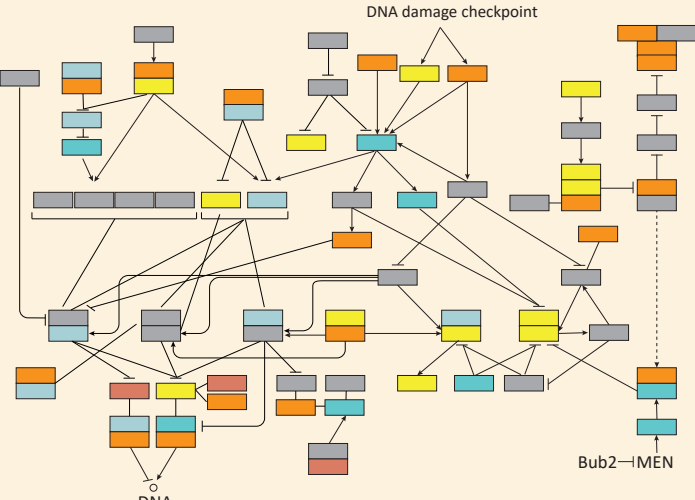

02　GO Enrichment Analysis

Enrichment analysis of gene or protein sets: Useful for exploring functional and biological significance from very large data sets (such as mass spectrometry data and microarray results). GO enrichment analysis also helps to organize data from novel (or fully annotated) genomes, and compare biological functions between clade members and across clades.

03　KEGG Annotation Analysis

KEGG is a database that systematically analyzes the metabolic pathways of gene products in cells and the functions of these gene products.
In organisms, different gene products coordinate with each other to perform biological functions. Annotation analysis on the pathways of differentially expressed genes helps to further interpret the function of genes.
- KEGG Pathway Annotation
- KEGG Pathway Classification

04　KEGG Enrichment Analysis

KEGG enrichment analysis of differentially expressed genes can enrich pathways with significant differences and help to find biologically regulated pathways with significant differences under experimental conditions.
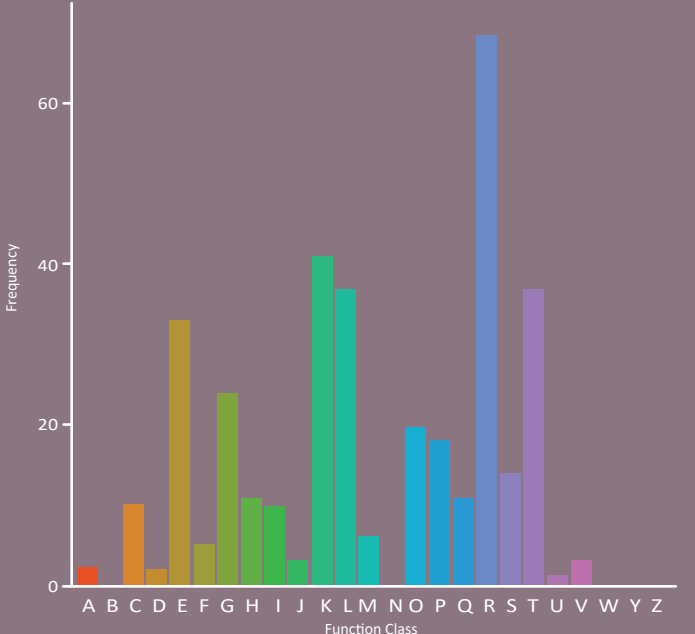
05　COG Annotation Analysis

Based on whole genomes analysis, COGs can reliably assign paralogs and orthologs of most genes using a simple method based on the search of triangles of bidirectional best hits. This method can identify distant homologs and isolate closely related paralogs.
Based on family analysis, this method can use the functions of the characteristic members of the protein family (COG) to assign functions to the entire family and describe potential functions in more than one family.

Creative Proteomics