Assigning GO and PFAM attributes to the NMPDR PEGS
Standard Operating Procedure NMPDR|SOP007

I. INTRODUCTION

- This standard operating procedure (SOP) describes the operations followed by NMPDR personnel for assigning GO and PFAM attributes to the NMPDR PEGS.

II. SCOPE

This SOP applies to the procedures to automatically calculate and assign GO terms and PFAM attributes to NMPDR PEGS.

III. APPLICABLE REGULATIONS AND GUIDELINES

NMPDR Contract	Delivery of NMPDR SOP's
BRC Metrics	Production of metrics
GO	List of GO terms
Transaction Logging	NMPDR Logging requirements

IV. RESPONSIBILITY

This SOP applies to those members of the NMPDR research team involved in calculating and assigning GO terms and PFAM attributes. This includes the following:

- Annotators
- Bioinformaticians

V. DEFINITIONS

The definitions found here: http://www.theseed.org/wiki/Glossary, apply to this SOP.

Standard Operating Procedures (SOPs): Detailed, written instructions to achieve uniformity of the performance of a specific function.

VI. PROCESS OVERVIEW

a. Identify Genomes to be processed
b. Run the automated process
c. Install the new attributes

VII. Context
Calculations are performed on the mirror of the primary annotation machine. Results are loaded onto the latest NMPDR version.

VIII. PROCEDURES

a. Log into bio-ppc-1.mcs.anl.gov
b. Change to bash shell.
c. Source the FIG environment, 'source /home/<username>/FIGdisk/config/fig-user-env.sh
d. Change to the ~mkubal/Domain_Analysis directory
e. Create a text file named "nmpdr_genomes_to_be_processed.txt". On each line should be the taxon id of a newly added NMPDR genome. (This will change to all new genomes in the near future).
f. Run `nohup perl submit_nmpdr_genomes_to_pipeline.pl`. Depending on the load on the cluster, this will take approximately 4 hours per genome.
g. Run 'perl parse_pfam_by_genome.pl`
h. Change to the Domain_Analysis/NMPDR_Results directory
i. Concatenate the GO results into a single file with `cat *_go_* > go_input.txt
j. Concatenate the PFAM results into a single file with `cat *_pfam_* > pfam_input.txt
k. Run `perl prepare_go_for_bruce.pl > go_attributes.for_bruce`
l. Run `perl prepare_pf_for_bruce.pl > pf_attributes.for_bruce`
m. Give the *_attributes.for_bruce files to Bruce to be added to the Attributes Server and/or be translated into tables for the NMPDR with the AttrDBRefresh script.