Adding Genomes to the NMPDR
Standard Operating Procedure NMPDR|SOP003

I. INTRODUCTION

This standard operating procedure (SOP) describes the operations followed by NMPDR personnel for adding new genomes to the NMPDR database. We address two broad classes of genomes:

- Genomes for NMPDR pathogens, which come from five sets of closely related strains of pathogens, and
- Diverse genomes used to support comparative analysis.

When we integrate diverse complete genomes, we often just take the genomes from RefSeq. Only in cases where the RefSeq gene calls require improvement do we apply similar procedures as for the NMPDR pathogens. This document describes the procedure used to install genes for NMPDR pathogens.

II. SCOPE

This SOP applies to the procedures to acquire new genome data, prepare newly acquired genome data, calling genes, annotating genes and producing derived data. It describes the steps followed by the site from the time the new genome data is discovered to be available until all data, primary and derived, has been assimilated into the production NMPDR site.

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. ATTACHMENTS

a. Process descriptions
b. Genome Numbers explained
c. Logging Requirements

V. RESPONSIBILITY

This SOP applies to those members of the NMPDR research team involved in acquiring, processing, installing and validating new genome data. This includes the following:

- Principal Investigator
- Production manager
- Annotation Manager
- Research Programmer
- Annotators
- Bioinformaticians
- EOT Personnel

VI. 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.

VII. PROCESS OVERVIEW

a. Acquire new genome data to be loaded into the NMPDR database
b. Call genes
c. Install new genome
d. Produce derived data
e. Install new NMPDR version

VIII. Context

Genomes are added to the primary annotators machine (anno-3) and these procedures are carried out on that machine, as the user fig.

IX. PROCEDURES

a. Acquire new genome data.

Responsible team members:	***Task Description***
• PI • Bioinformaticians	Identify new genome candidates.
• Bioinformaticians • PI	Decide if the candidates are new to the NMPDR or are replacements Process A. Process to verify new genome status.
• Bioinformaticians	Acquire the new genome data and place into a working site. Claim a genome number, create SEED directory. Process B. Preparation Activities.

b. Gene calling

• Bioinformaticians	Call Genes with rapid propagation techniques Process C. Rapid Propagation
• PI • Bioinformaticians	Determine effectiveness of RPT. If necessary, recall genes using the GISMO program at Bielefeld Process D. Evaluate Gene Calls. Process E. Call genes using GISMO
• PI • Bioinformaticians	Recall tRNA's and rRNA's Process F. Recalling RNA's

c. Install new genome

• Bioinformaticians	Install the new genome into the SEED environment Process G. Installing a new genome
• Bioinformaticians	If there are no assigned functions (i.e. Gene calls From GIZMO), use rapid propagation techniques to create a set of proposed functions. Process H. Creating a set of proposed functions
• Bioinformaticians	Assign Functions to new genome genes Process I. Install Assignments
• Bioinformaticians	If this is a duplicate Genome, mark old genome as deleted Process J. Marking old genome as deleted

d. Produce derived data

• Research Programmer • Bioinformaticians	Verify computation of similarities. (The computation is triggered automatically by adding the genome) Process K. Verifying computation of similarities
• Research Programmer • Bioinformaticians	Verify computation of PCH's, PinnedRegions and Functional Coupling (These are scheduled automatically following computation of sims). Process L. Verifying computation of derived data.
• Research Programmer • Bioinformaticians	Produce automatic assignments for CDS' without assignments Process M. Automatic assignments
• Research Programmer • Bioinformaticians	Add genome to subsystems allowing automatic updates Process N. Automatically updating subsystems
• Bioinformaticians • PI	Mark Genome "Complete" Process O. Marking the Genome as complete.

e. Install new NMPDR Version

• Research Programmer	Run new NMPDR Version cycle

	Process P. NMPDR Version
• EOT personnel • Bioinformaticians	Verify new NMPDR version Process Q. Procedure to verify the new NMPDR version