

# DACC GO Slim Analysis

## Institute for Genome Sciences, UMD School of Medicine

**Author:** Michelle Giglio and Joshua Orvis

**Version:** 1.0

**Effective Date:** October 5, 2011

---

## 1 Abstract

## 2 Introduction

This SOP describes use of a GO Slim to generate biological process-level overviews of the metagenomic wgs annotations. The process uses as input annotation tab files, and generates a summary count matrix.

## 3 Requirements

### 3.1 GO Slim file

Any GO slim file may be used for this process. Slims are easily customizable so that they can be specific in some areas and general in others. Several standard slims are available from the GO consortium (<http://www.geneontology.org/GO.slims.shtml>).

For the GO Slim analysis performed on the HMP metagenomic wgs annotation data, the DACC constructed a slim, referred to as **DACCSlim1** (HMPSlim.v2.obo, available at <http://hmpdacc.org/HMGS/>) consisting of all of the first level children of the GO Biological Process ontology. The regulation and metabolism nodes were browsed and additional, more specific terms relevant to prokaryotic processes were also included in the slim.

All slims must be in obo format (<http://www.geneontology.org/GO.format.obo-1.2.shtml>), and from only one ontology.

### 3.2 Data requirements

Annotation files must be in the GO gene\_association file (gaf) format. Details at: <http://www.geneontology.org/GO.format.annotation.shtml>

### 3.3 Program requirements

- Custom scripts referenced below in the Procedure section, are available at <http://hmpdacc.org/HMGS/>
- **map2slim** is distributed as part of the go---perl package, available from CPAN at <http://search.cpan.org/~cmungall/go---perl/>. map2slim documentation is available at <http://search.cpan.org/~cmungall/go---perl/scripts/map2slim>.

# DACC GO Slim Analysis

## Institute for Genome Sciences, UMD School of Medicine

**Author:** Michelle Giglio and Joshua Orvis

**Version:** 1.0

**Effective Date:** October 5, 2011

---

## 4 Procedure

### 4.1 Format annotation files

The first step in processing the annotation tab files was converting the output to gene- association file format, using **convert\_jcvi\_annottab\_to\_gaf.pl**, available at <http://hmpdacc.org/HMGS/>. This input format is required by Chris Mungall's map2slim script, which was used to perform the actual mappings.

### 4.2 Slim Mapping

The mapping commands as run as follows:

```
map2slim -c -t -o <output slim file> HMPSlim.obo
gene_ontology_ext.obo <input gaf file>
```

### 4.3 Generate Summary Output

A second custom script, also available at <http://hmpdacc.org/HMGS/>, **create\_go\_annotation\_matrix.pl**, takes the output from map2slim and produces a summary count matrix, where the X-axis is every GO term that was assigned to any sample and the Y-axis lists each QC-passed PGA sample identifier.

The values of the matrix, depending on the output file variant, are either the relative abundance or count of that particular assignment.

## 5 Implementation

## 6 Discussion

## 7 Related Documents & References

<http://www.geneontology.org/>

## 8 Revision History

Version	Author/Reviewer	Date	Change Made
1.0	Michelle Giglio & Joshua Orvis	10/5/2011	Establish SOP

**DACC GO Slim Analysis**  
**Institute for Genome Sciences, UMD School of Medicine**

**Author:** Michelle Giglio and Joshua Orvis

**Version:** 1.0

**Effective Date:** October 5, 2011

---