GenMAPP Gene Database for *Shewanella oneidensis* MR-1
So-Std_HMH_20151214.gdb
ReadMe

Last revised: 12/12/15

This document contains the following:

1. Overview of the GenMAPP application and accessory programs

GenMAPP (Gene Map Annotator and Pathway Profiler) is a free computer application for viewing and analyzing DNA microarray and other genomic and proteomic data on biological pathways. MAPPFinder is an accessory program that works with GenMAPP and Gene Ontology to identify global biological trends in gene expression data. The GenMAPP Gene Database (file with the extension *.gdb*) is used to relate gene IDs on MAPPs (*.mapp*, representations of pathways and other functional groupings of genes) to data in Expression Datasets (*.gex*, DNA microarray or other high-throughput data). GenMAPP is a stand-alone application that requires the Gene Database, MAPPs, and Expression Dataset files to be stored on the user's computer. GenMAPP and its accessory programs and files may be downloaded from <http://www.GenMAPP.org>. GenMAPP requires a separate Gene Database for each species. This ReadMe describes a Gene Database for *Shewanella oneidensis* MR-1 that was built by a Loyola Marymount University (LMU) Bioinformatics Group using the program GenMAPP Builder 2.0, part of the open source XMLPipeDB project <http://xmlpipedb.cs.lmu.edu/>.

2. System Requirements and Compatibility:

- This Gene Database is compatible with GenMAPP 2.0 and 2.1 and MAPPFinder 2.0. These programs can be downloaded from <http://www.genmapp.org>.
- System Requirements for GenMAPP 2.0/2.1 and MAPPFinder 2.0:
  Operating System: Windows 98 or higher, Windows NT 4.0 or higher (2000, XP, etc)
  Monitor Resolution: 800 X 600 screen or greater (SVGA)
  Internet Browser: Microsoft Internet Explorer 5.0 or later
  Minimum hardware configuration:
  Memory: 128 MB (512 MB or more recommended)
  Processor: Pentium III
  Disk Space: 300 MB disk (more recommended if multiple databases will be used)

3. Installation Instructions

- Extract the zipped archive and place the file "So-Std_HMH_20151214.gdb" in the folder you use to store Gene Databases for GenMAPP. If you accept the default folder during the GenMAPP installation process, this folder will be C:\GenMAPP 2 Data\Gene Databases.

- To use the Gene Database, launch GenMAPP and go to the menu item *Data* > *Choose Gene Database*. Alternatively, you can launch MAPPFinder and go to the menu item *File* > *Choose Gene Database*.

4. Gene Database Specifications

a. Gene ID Systems

This *Shewanella oneidensis* Gene Database is UniProt-centric in that the main data source (primary ID System) for gene IDs and annotation is the UniProt complete proteome set for *Shewanella oneidensis*, made available as an XML download. In addition to UniProt IDs, this database provides the following proper gene ID systems that were cross-referenced by the UniProt data: OrderedLocusNames, GeneID (NCBI), and RefSeq (protein IDs of the form NP_######). It also supplies UniProt-derived annotation links from the following systems: EMBL, InterPro, PDB, and Pfam. The Gene Ontology data has been acquired directly from the Gene Ontology Project. The GOA project was used to link Gene Ontology terms to UniProt IDs. Links to data sources are listed in the section below.

Proper ID System	SystemCode
UniProt	S
OrderedLocusNames	N
GeneID (NCBI)	L
RefSeq	Q

b. Species

This Gene Database is based on the UniProt proteome set for *Shewanella oneidensis* MR-1, taxon ID 211586.

c. Data Sources and Versions

- This *Shewanella oneidensis* Gene Database was built on October 24, 2013; this build date is reflected in the filename So-Std_ 20151212.gdb. All date fields internal to the Gene Database (and not usually seen by regular GenMAPP users) have been filled with this build date.
- UniProt complete proteome set for *Shewanella oneidensis*, downloaded from this page: <http://www.uniprot.org/uniprot/?query=organism%3a243277+keyword%3a1185&format=*&compress=yes>
  Filename: "SOneidensisUNIPROT.xml" (downloaded as a compressed .gz file and extracted)
  Version information for the proteome sets can be found at <http://www.uniprot.org/news/>
  The proteome set used for this version of the *Shewanella oneidensis* Gene Database was based on UniProt release 2015_10 released on October 14, 2015.
- Gene Ontology gene associations are provided by the GOA project: <http://www.ebi.ac.uk/GOA/> as a tab-delimited text file. The *Shewanella oneidensis* GOA file was accessed from the GOA proteomes FTP site: < ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/proteomes/>
  Filename: "ShewanellaOneidensisGOA.goa".
- Gene Ontology data is downloaded from <http://beta.geneontology.org/page/download-ontology>
  Data is released daily. For this version of the *Shewanella oneidensis* Gene Database we used the ontology version 2015-11-21 04:00:00 PM.
  Filename: "go daily-termdb.obo-xml.gz".

d. Database Report

- UniProt is the primary ID system for the *Shewanella oneidensis* Gene Database. The UniProt table contains all 4196 UniProt IDs contained in the UniProt proteome set for this species.
- The OrderedLocusNames ID system was derived from the cross-references in the UniProt proteome set. Each ID appears in the form of SO_#### or in the form of SO_A####, (e.g., SO_0001 and SO_A0001) because IDs of both forms can be found in the literature. We compared this table with the list of gene IDs found using the matching function from the XMLPipeDB project <http://xmlpipedb.cs.lmu.edu/>. There are 5207 protein coding genes

listed there. These 11 gene IDs do not appear in our Gene Database because, according to the UniProt record web page, they are part of the "STRING" protein-protein interaction database.

- The following table lists the numbers of gene IDs found in each gene ID system:

ID System	ID Count Current version
EMBL	8
GeneID (NCBI)	4213
GeneOntology	5603
InterPro	4588
OrderedLocusNames	4196
PDB	123
Pfam	2276
RefSeq	8251
UniProt	4071

5. Contact Information for support, bug reports, feature requests

- The Gene Database for *Shewanella oneidensis* was built by the Loyola Marymount University (LMU) Bioinformatics Group using the program GenMAPP Builder, part of the open source XMLPipeDB project <http://xmlpipedb.cs.lmu.edu/>.
- For support, bug reports, or feature requests relating to XMLPipeDB or GenMAPP Builder, please consult the XMLPipeDB Manual found at <http://xmlpipedb.cs.lmu.edu/documentation.shtml> or go to our SourceForge site <http://sourceforge.net/projects/xmlpipedb/>.
- For issues related to the *Shewanella oneidensis* Gene Database, please contact:

Kam D. Dahlquist, PhD.
Department of Biology
Loyola Marymount University
1 LMU Drive, MS 8220
Los Angeles, CA 90045-2659
kdahlquist@lmu.edu

- For issues related to GenMAPP 2.0/2.1 or MAPPFinder 2.0 please contact GenMAPP support directly by e-mailing genmapp@gladstone.ucsf.edu or GenMAPP@googlegroups.com.

6. Release Notes

a. Current version: So-Std _HMH_20151214.gdb

- The Heavy Metal HaterZ contributed to this release.

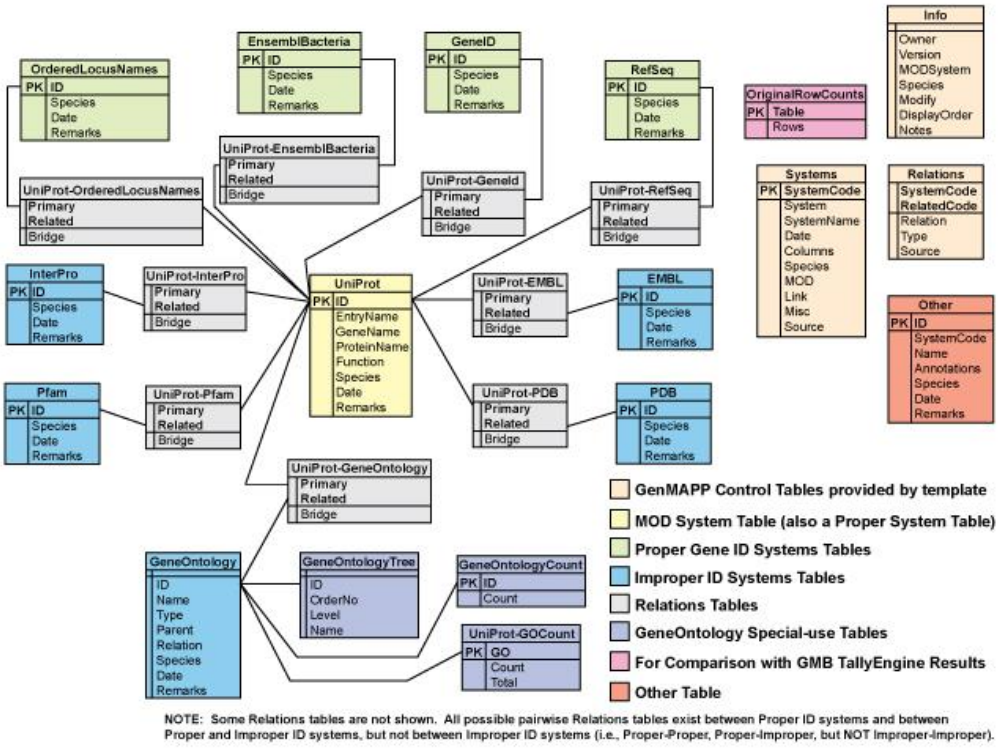
b. Previous version: So-Std_20151201special.gdb

- The Heavy Metal HaterZ contributed to this release.

c. Previous version: So-Std_20151119HMH.gdb

- The Heavy Metal HaterZ, John David N. Dionisio, and Kam D. Dahlquist contributed to the first release.

Shewanella Oneidensis MR-1



NOTE: Some Relations tables are not shown. All possible pairwise Relations tables exist between Proper ID systems and between Proper and Improper ID systems, but not between Improper ID systems (i.e., Proper-Proper, Proper-Improper, but NOT Improper-Improper).