GenMAPP Gene Database for *Leishmania Major*
LeishmaniaGDB_Lena_Gabe_20131205.gdb
ReadMe

Last revised: 11/26/13

This document contains the following:

1. Overview of GenMAPP application and accessory programs
2. System Requirements and Compatibility
3. Installation Instructions
4. Gene Database Specifications
    a. Gene ID Systems
    b. Species
    c. Data Sources and Versions
    d. Database Report
5. Contact Information for support, bug reports, feature requests
6. Release notes
    a. Current version: LeishmaniaGDB_Lena_Gabe_20131205.gdb
    b. Previous version: LeishmaniaGDB_Lena_Gabe_20131203.gdb
7. Database Schema Diagram

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 *Leishmania Major* that was built by the 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 "LeishmaniaGDB_Lena_Gabe_20131205.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 *Leishmania major* 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 *Vibrio cholerae*, 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 *Leishmnaia major* , taxon ID 5664.

c. Data Sources and Versions

- This *Leishmania major* Gene Database was built on 5 December 2013; this build date is reflected in the filename LeishmaniaGDB Lena Gabe 20131205.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 *Leishmania major*, downloaded from this page: <http://www.uniprot.org/uniprot/?query=organism%3a5664+keyword%3a1185&format=*&compress=yes>
  Filename: "uniprot-organism%3A5664+keyword%3A1185.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 Leishmania major Gene Database was based on UniProt release 2013_11 released on October 16, 2013.
- Gene Ontology gene associations are provided by the GOA project: <http://www.ebi.ac.uk/GOA/> as a tab-delimited text file. The *Leishmania major* GOA file was accessed from the GOA proteomes FTP site: < ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/proteomes/>
  Filename: "LeishmaniaGOA_19112013_Lena_Gabe.goa". Version 11/19/2013 9:59:00 AM.
- Gene Ontology data is downloaded from <http://beta.geneontology.org/page/download-ontology>
  Data is released daily. For this version of the *Leishmania major* Gene Database we used the ontology version 2013-11-05 12:54:57 PM.
  Filename: "go_daily-termdb.obo-xml.gz".

d. Database Report

- UniProt is the primary ID system for the *Leishmania major* Gene Database. The UniProt table contains all 8355 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 twice, once in the form of LMJF_##_#### and once in the form of LMJF.##.####, (e.g., LMJF_00_0001 and LMJF.00.0001) because IDs of both forms

can be found in the literature. We compared this table with the list of gene IDs in the GeneDB database for Leishmania major. There are 8355 protein coding genes listed there.

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

OriginalRowCounts	
Table	**Rows**
UniProt	8041
Pfam	1923
RefSeq	8315
PDB	101
InterPro	3451
OrderedLocusNames	16662
GeneID	8315
EMBL	218

*There are 8355 unique genes/proteins in the current version of the Gene Database; the 16662 count represents the total number of IDs due to duplicate IDs of the form LMJF_##_#### and LMJF_##_####

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

- The Gene Database for *Leishmania major* 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 *Leishmania major* 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.

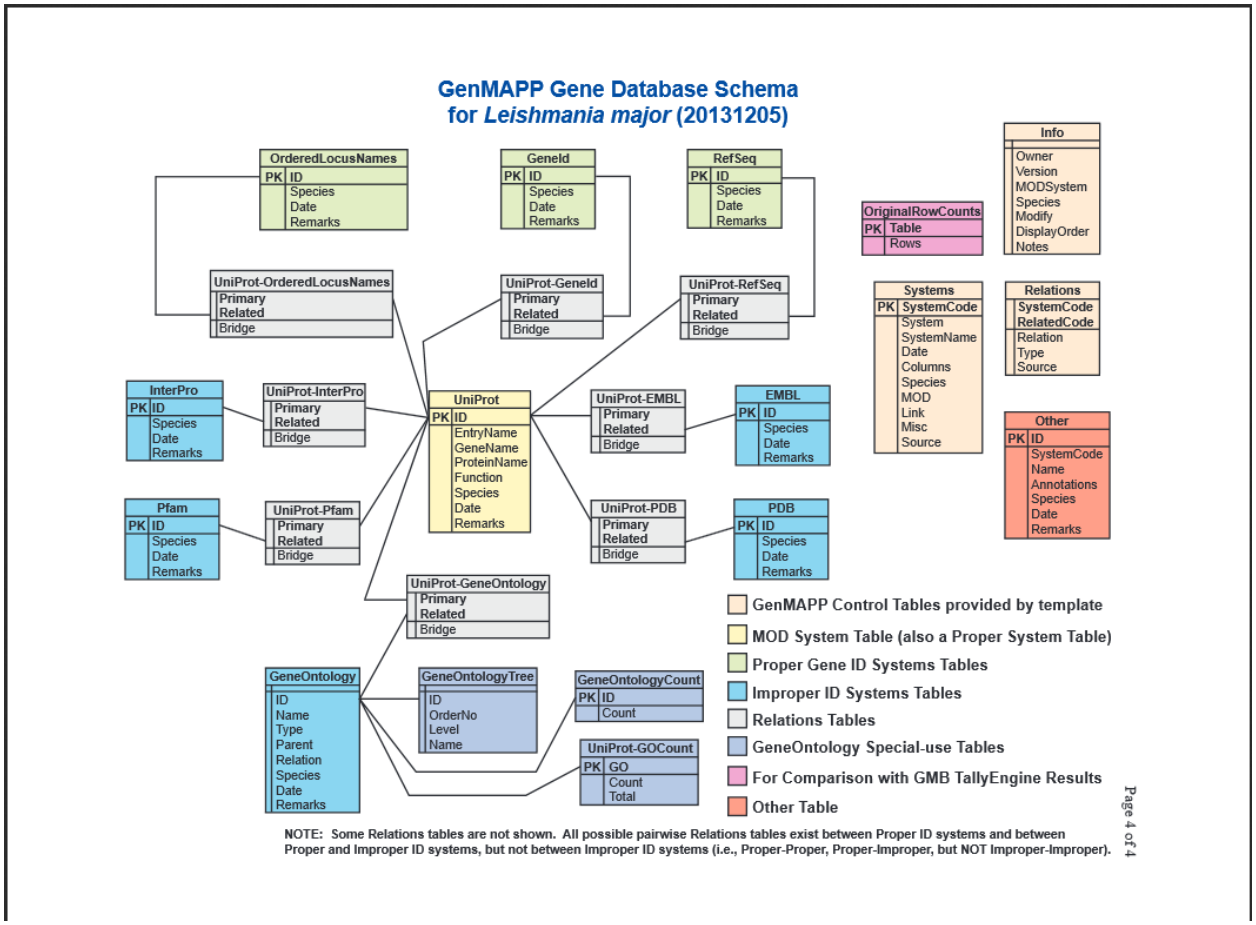
6. Release Notes

a. Current version: LeishmaniaGDB_Lena_Gabe_20131205.gdb

- Kam D. Dahlquist and John David N. Dionisio contributed to this release.

b. Previous version: LeishmaniaGDB_Lena_Gabe_20131203.gdb

- Kam D. Dahlquist and John David N. Dionisio contributed to this release.


GenMAPP Gene Database Schema
for Leishmania major (20131205)
OrderedLocusNames
PK ID
Species
Date
Remarks
GeneId
PK ID
Species
Date
Remarks
RefSeq
PK ID
Species
Date
Remarks
Info
Owner
Version
MODSystem
Species
Modify
DisplayOrder
Notes
OriginalRowCounts
PK Table
Rows
UniProt-OrderedLocusNames
Primary
Related
Bridge
UniProt-GeneId
Primary
Related
Bridge
UniProt-RefSeq
Primary
Related
Bridge
Systems
PK SystemCode
System
SystemName
Date
Columns
Species
MOD
Link
Misc
Source
Relations
SystemCode
RelatedCode
Relation
Type
Source
InterPro
PK ID
Species
Date
Remarks
UniProt-InterPro
Primary
Related
Bridge
UniProt
PK ID
EntryName
GeneName
ProteinName
Function
Species
Date
Remarks
UniProt-EMBL
Primary
Related
Bridge
EMBL
PK ID
Species
Date
Remarks
Other
PK ID
SystemCode
Name
Annotations
Species
Date
Remarks
Pfam
PK ID
Species
Date
Remarks
UniProt-Pfam
Primary
Related
Bridge
UniProt-PDB
Primary
Related
Bridge
PDB
PK ID
Species
Date
Remarks
UniProt-GeneOntology
Primary
Related
Bridge
GeneOntology
ID
Name
Type
Parent
Relation
Species
Date
Remarks
GeneOntologyTree
ID
OrderNo
Level
Name
GeneOntologyCount
PK ID
Count
UniProt-GOCount
PK GO
Count
Total
GenMAPP Control Tables provided by template
MOD System Table (also a Proper System Table)
Proper Gene ID Systems Tables
Improper ID Systems Tables
Relations Tables
GeneOntology Special-use Tables
For Comparison with GMB TallyEngine Results
Other Table
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).
Page 4 of 4