

GenMAPP Gene Database for *Bordetella pertussis* Tohama I

Bp_Std_External_20151210.gdb

ReadMe

Last revised: 1/22/16

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: Bp_Std_External_20151210.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 *Bordetella pertussis* Tohama I that was built by the Loyola Marymount University (LMU) Bioinformatics Group using the program GenMAPP Builder 2.1, 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 "Bp_Std_External_20151210.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 *Bordetella pertussis* 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 *Bordetella pertussis*, 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_#####, WP_#####, YP_#####). 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 *Bordetella pertussis* Tohama I (ATCC BAA-589), taxon ID 257313.

c. Data Sources and Versions

- This *Bordetella pertussis* Gene Database was built on December 10, 2015; this build date is reflected in the filename Bp_Std_External_20151210.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 *Bordetella pertussis* Tohama I, downloaded from this page: <<http://www.uniprot.org/proteomes/UP000002676>>
Filename: "uniprot-proteome%3AUP000002676.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 *Bordetella pertussis* Gene Database was based on UniProt release 2015_12 released on December 9, 2015.
- Gene Ontology gene associations are provided by the GOA project:
<<http://www.ebi.ac.uk/GOA/>> as a tab-delimited text file. The *Bordetella pertussis* GOA file was accessed from the GOA proteomes FTP site:
<<ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/proteomes/>>
Filename: "145.B_pertussis_ATCC_BAA-589.goa". Version 12/10/2015 4:30:40 PM.
- Gene Ontology data is downloaded from <http://geneontology.org/page/download-ontology#Legacy_Downloads>
Data is released daily. For this version of the *Bordetella pertussis* Gene Database we used the ontology version 12/10/2015 4:19:50 PM.
Filename: "go_daily-termdb.obo.xml" (downloaded as a compressed .gz file and extracted)

d. Database Report

- UniProt is the primary ID system for the *Bordetella pertussis* Gene Database. The UniProt table contains all 3258 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. IDs in this table exist in the forms BP####, BP####.1, BP####A, and BP####B, all of which can be found in the literature. We compared this table with the list of gene IDs in the GeneDB Model Organism Database (MOD) at

<<http://www.genedb.org/Homepage/Bpertussis>>. There are 3447 protein-coding genes listed there, all of which appear in our Gene Database.

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

| ID System | ID Count Current version |
|-------------------|--------------------------------|
| EMBL | 67 |
| GeneID (NCBI) | 3441 |
| GeneOntology | 4836 |
| InterPro | 3681 |
| OrderedLocusNames | 3447 |
| PDB | 43 |
| Pfam | 1794 |
| RefSeq | 6624 |
| UniProt | 3258 |

- The Gene Database Testing Report for this Gene Database can be found at:
<https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Gene_Database_Testing_Report-cw20151210>

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

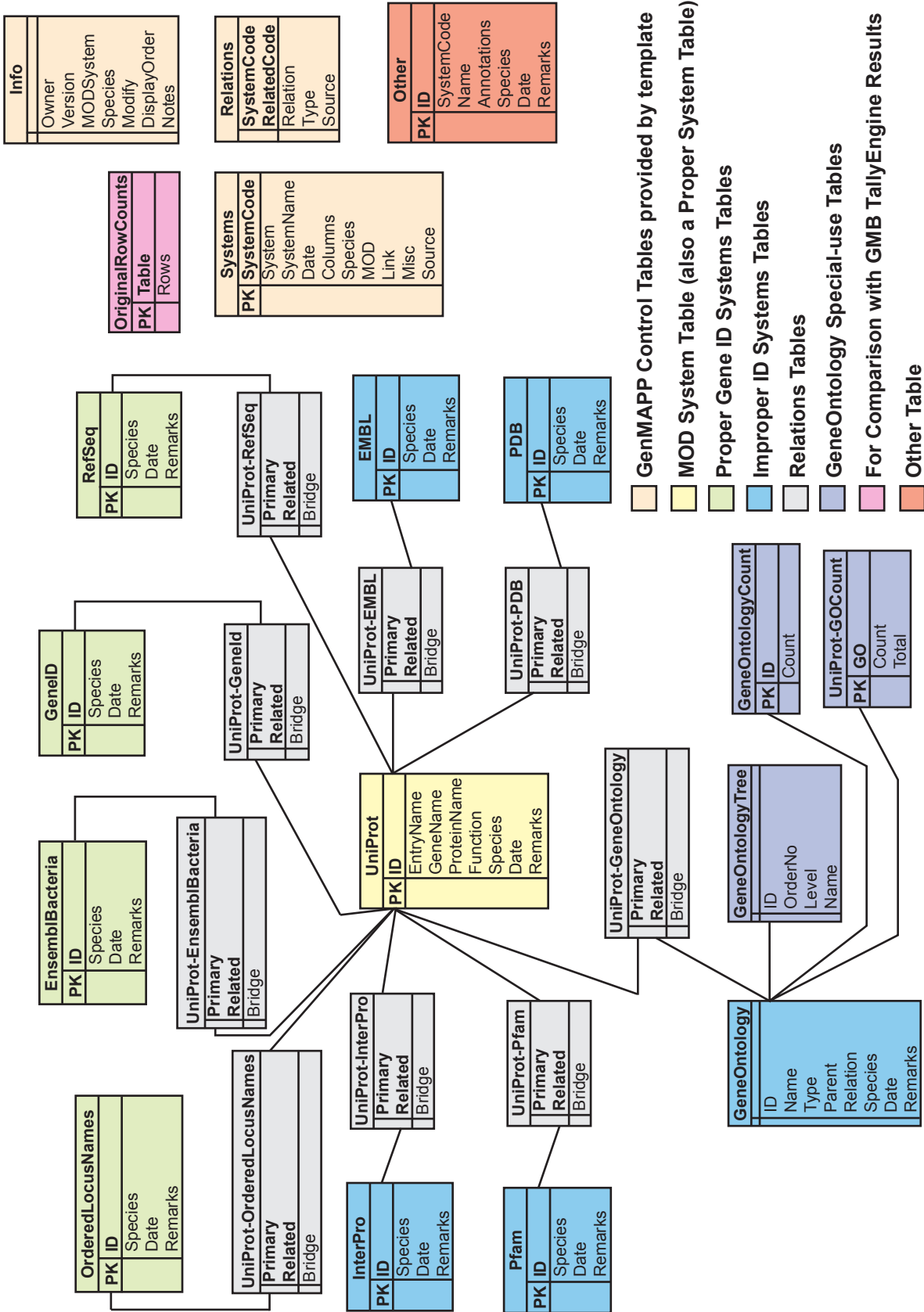
- The Gene Database for *Bordetella pertussis* was built by the Loyola Marymount University (LMU) Bioinformatics Group using the program GenMAPP Builder, part of the open source XMLPipeDB project <<http://xmlpipedb.cs.lmu.edu/>>.
- For support, bug reports, or feature requests relating to XMLPipeDB or 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 *Bordetella pertussis* Gene Database, please contact:
Kam D. Dahlquist, PhD.
Department of Biology
Loyola Marymount University
1 LMU Drive, MS 8888
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: Bp_Std_External_20151210.gdb

- Brandon J. Klein, Elena E. Olufson, John David N. Dionisio, Kam D. Dahlquist, and Mahrad R. Saeedi contributed to the first release.

GenMAPP Gene Database Schema for *Bordetella pertussis* Tohama I



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).