GenMAPP Gene Database for *Plasmodium falciparum* (isolate 3D7)
Pf-Std_External_2009604.gdb
ReadMe

Last revised: 6/9/09

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: Pf-Std_External_2009604.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 *Plasmodium falciparum* (3D7) 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 "Pf-Std_External_2009604.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 *Plasmodium falciparum* 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 *Plasmodium falciparum*, made available as an XML download by the Integr8 resource. 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_######, XP_######, and XP_#########). 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	L
RefSeq	Q

    b. **Species**
        This Gene Database is based on the UniProt proteome set for *Plasmodium falciparum* (isolate 3D7), taxon ID 36329.

    c. **Data Sources and Versions**
        - This *Plasmodium falciparum* Gene Database was built on June 4, 2009; this build date reflected in the filename Pf-Std_External_2009604.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 *Plasmodium falciparum* (isolate 3D7), made available as an XML download by the Integr8 resource:
          <http://www.ebi.ac.uk/integr8/FtpSearch.do?orgProteomeId=493>
          Filename: "493.P_falciparum.xml" (downloaded as a compressed .gz file and extracted)
          Version information for the proteome sets can be found at
          <http://www.ebi.ac.uk/integr8/FtpSearch.do?orgProteomeId=493>
          The proteome set used for this version of the *Plasmodium falciparum* Gene Database was built from release 91 of Integr8 (February 11, 2009), which was built from UniProt release 14.8 and InterPro release 18.0.
        - Gene Ontology gene associations are provided by the GOA project:
          <http://www.ebi.ac.uk/GOA/> as a tab-delimited text file. The *Plasmodium falciparum* GOA file was accessed from the Integr8 proteome set download page:
          <http://www.ebi.ac.uk/integr8/FtpSearch.do?orgProteomeId=18>
          Filename: "493.P_falciparum.goa". The GOA file for this version of the *Plasmodium falciparum* Gene Database was based on the GOA Proteome Sets 45.0 released in January 2009.

- Gene Ontology data is downloaded from
  <http://www.geneontology.org/GO.downloads.ontology.shtml>
  Data is released daily. For this version of the *Plasmodium falciparum* Gene Database we
  used the February 6, 2009 release.
  Filename: "go_daily-termdb.obo-xml.gz".

d. Database Report

- UniProt is the primary ID system for the *Plasmodium falciparum* Gene Database. The
  UniProt table contains all 5284 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. The IDs took the following forms:

PFcddds *e.g., PFC0582c*	o **c** is either A, B, C, D, E, I, or L, which correspond to chromosomes 1, 2, 3, 4, 5, 9, and 12 o **dddd** is a four digit number with leading zeros, e.g., 0370 o **s** is either w (watson) or c (crick) o NOTE: the link out to PlasmoDB for this form of the ID will not function on the Backpage when this type of ID is entered in the Gene Finder window because PlasmoDB is case sensitive and GenMAPP converts all lowercase letters into uppercase in this instance.
PFchr_dddd *e.g., PF13_0016*	o **chr** is one of 00, 07, 08, 10, 11, 13, 14 o **dddd** is a four digit number with leading zeros
MALchrP1.n *e.g., MAL13P1.112*	o **chr** is one of 1, 6, 7, 8, or 13 o **n** is a number

  (from < http://plasmodb.org/plasmo/showXmlDataContent.do?name=XmlQuestions.About#naming>)
  We compared this table with the list of gene IDs from PlasmoDB
  <http://plasmodb.org/plasmo/>. 545 gene IDs from this source do not appear in our Gene
  Database because they are not cross-listed in the UniProt XML file. Of these 545 gene IDs,
  429 have the status new, changed, deleted, new_organellar, or changed_organellar.
- The following table lists the numbers of gene IDs found in each gene ID system:

ID System	ID Count
EMBL	261
GeneId (NCBI)	5260
GeneOntology	2760
InterPro	2175
OrderedLocusNames	5336
PDB	51
Pfam	1334
RefSeq	5264
UniProt	5284

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

- The Gene Database for *Plasmodium falciparum* 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
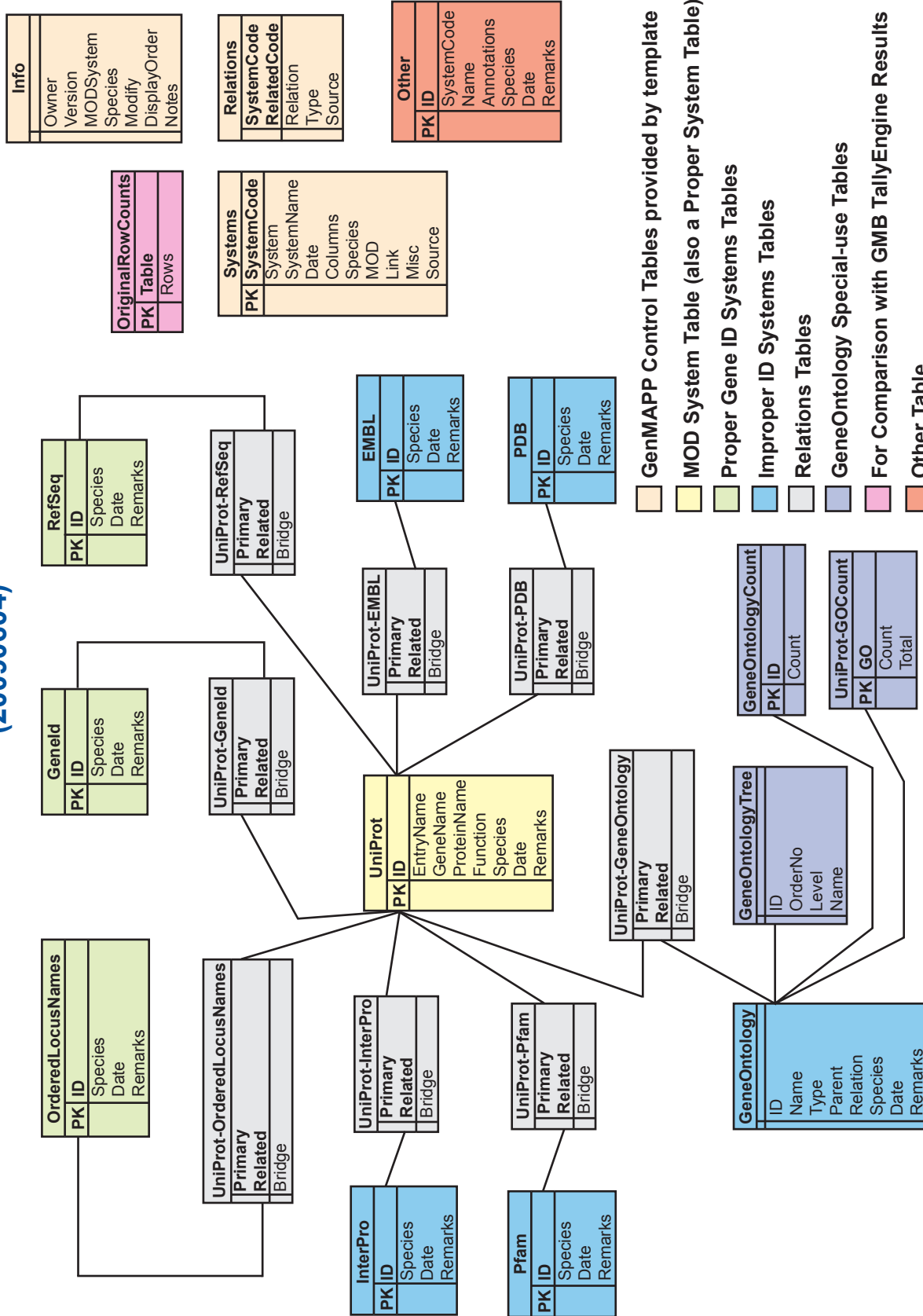 <http://sourceforge.net/projects/xmlpipedb/>.

- For issues related to the *Plasmodium falciparum* 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: Pf-Std_External_2009604.gdb**
      - This release is the first release of a standard *Plasmodium falciparum* Gene Database

GenMAPP Gene Database Schema for *Plasmodium falciparum* (isolate 3D7) (20090604)

Legend:

- 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).