**GenMAPP Gene Database for *Staphylococcus aureus* (strain MRSA252)**
Sa-MRSA252-Std_External_20140904.gdb
**ReadMe**

Last revised: 10/10/14

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: Sa-MRSA252-Std_External_20140904.gdb
   b. Previous version: Sa-Std_External_20100218.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 *Staphylococcus aureus* (strain MRSA252) that was built by the Loyola Marymount University (LMU) Bioinformatics Group using the program GenMAPP Builder version 2.0b80, 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 (now referred to as "GenMAPP Classic") and MAPPFinder 2.0. These programs can be downloaded from <http://www.genmapp.org>, direct download link is <http://root2.genmapp.org/dtr/Installers/GenMAPPv2Setup.exe>.
- 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 "Sa-MRSA252-Std_External_20140904.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 S*taphylococcus aureus* (strain MRSA252) 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 *Staphylococcus aureus* (strain MRSA252). In addition to UniProt IDs, this database provides the following proper gene ID systems that were cross-referenced by the UniProt data: OrderedLocusNames (IDs of the form SAR#### or SAR####.#), EnsemblBacteria, GeneID (NCBI), and RefSeq (protein IDs of the form WP_######### and 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 | System Code |
|---|---|
| UniProt | S |
| OrderedLocusNames | N |
| EnsemblBacteria | Eb |
| GeneID | L |
| RefSeq | Q |

   **b. Species**

        This Gene Database is based on the UniProt proteome set for *Staphylococcus aureus* (strain MRSA252) with the taxon ID 282458 and the UniProt Proteome ID UP000000596.

   **c. Data Sources and Versions**
- This *Staphylococcus aureus* (strain MRSA252) Gene Database was built on September 4, 2014; this build date is reflected in the filename Sa-MRSA252-Std_External_20140904.gdb. All date fields internal to the Gene Database (and not usually seen by regular GenMAPP users) have been filled with this build date.
- The UniProt complete proteome set for *Staphylococcus aureus* (strain MRSA252), was downloaded from : <http://www.uniprot.org/proteomes/UP000000596> in XML format.
  > Filename: "uniprot-proteome%3AUP000000596.xml".
  > Version: UniProt release 2014_08 (September 3, 2014). Note that version information for the proteome sets can be found at <http://www.uniprot.org/news/>.
- Gene Ontology gene associations are provided by the GOA project: <http://www.ebi.ac.uk/GOA/> and was accessed from the GOA proteomes FTP site: <ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/proteomes/>.
  > Filename: "19583.S_aureus_MRSA252.goa".
  > Version: 09/02/14 4:50:00 AM.
- Gene Ontology data was downloaded in OBO-XML format from <http://beta.geneontology.org/page/download-ontology>
  > Filename: "go_daily-termbd.obo-xml".
  > Version: data is released daily, we used data from 9/04/14 02:12 AM.

**d. Database Report**
- UniProt is the primary ID system for the *Staphylococcus aureus* (strain MRSA252) Gene Database. The UniProt table contains all 2640 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 OrderedLocusNames IDs are in the form SAR#### or SAR###.#. The current version of the Gene Database has 29 fewer OrderedLocusNames IDs than the previous version. None of these IDs is currently present in the UniProt or EnsemblBacteria databases. Each of the IDs missing from the current version of the Gene Database had a suffix of ".#" as in, SAR####.#

| ID System | ID Count Current Version 20140904 | ID Count Previous Version 20100218 |
|---|---|---|
| EMBL | 1 | 1 |
| EnsemblBacteria | 2656 | not present |
| GeneID (NCBI) | 2650 | 2650 |
| GeneOntology | 4631 | 3416 |
| InterPro | 3346 | 2947 |
| OrderedLocusNames | 2630 | 2659 |
| PDB | 73 | 19 |
| Pfam | 1599 | 1462 |
| RefSeq | 2711 | 2656 |
| UniProt | 2640 | 2640 |

## 5. Contact Information for support, bug reports, feature requests
- The Gene Database for *Staphylococcus aureus* (strain MRSA252) was built by the Loyola Marymount University (LMU) Bioinformatics Group using the program GenMAPP Builder version 2.0b80, 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>, the XMLPipeDB wiki at <http://xmlpipedb.sourceforge.net/wiki/index.php/Main_Page>, or our SourceForge site at <http://sourceforge.net/projects/xmlpipedb/>.
- For issues related to the *Staphylococcus aureus* (strain MRSA252) Gene Database, please contact:
  > Kam D. Dahlquist, Ph.D.
  > 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:** Sa-MRSA252-Std_External_20140904.gdb
- This is the second release of a standard *Staphylococcus aureus* (strain MRSA252) Gene Database. Note that the name has been changed to specify the MRSA252 strain designation to distinguish this Gene Database from those for other *S. aureus* strains.

- The Gene Database Testing Report can be found at <http://xmlpipedb.sourceforge.net/wiki/index.php/Gene_Database_Testing_Report_S.aureus_MRSA252_20140904_SP>
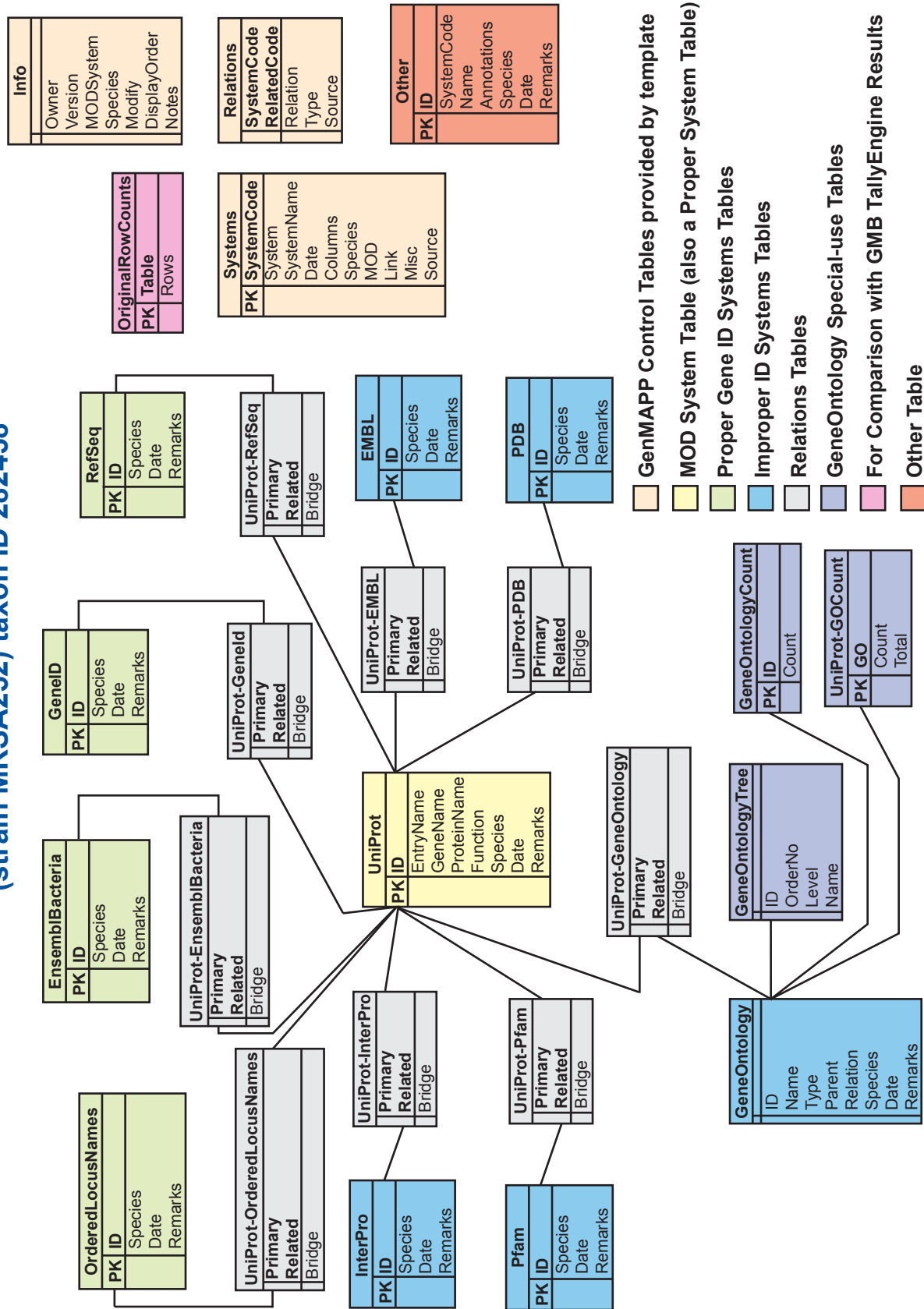- Sarah Patno (lead), Kam D. Dahlquist, and John David N. Dionisio contributed to this release.

**b. Previous version:** Sa-Std_External_20100218.gdb

- This release was the first release of a standard *Staphylococcus aureus* (strain MRSA252) Gene Database
- Kelly Parks (lead), Michael Piña, Andrew Hirning, John David N. Dionisio, and Kam D. Dahlquist contributed to this release.

# GenMAPP Gene Database Schema for *Staphylococcus aureus* (strain MRSA252) taxon ID 282458

**Info**
- Owner
- Version
- MODSystem
- Species
- Modify
- DisplayOrder
- Notes

**Relations**
- **SystemCode**
- **RelatedCode**
- Relation
- Type
- Source

**Other**
- PK | ID
- SystemCode
- Name
- Annotations
- Species
- Date
- Remarks

**OriginalRowCounts**
- PK | Table
- Rows

**Systems**
- PK | SystemCode
- System
- SystemName
- Date
- Columns
- Species
- MOD
- Link
- Misc
- Source

**RefSeq**
- PK | ID
- Species
- Date
- Remarks

**UniProt-RefSeq**
- Primary
- Related
- Bridge

**EMBL**
- PK | ID
- Species
- Date
- Remarks

**UniProt-EMBL**
- Primary
- Related
- Bridge

**PDB**
- PK | ID
- Species
- Date
- Remarks

**UniProt-PDB**
- Primary
- Related
- Bridge

**GeneID**
- PK | ID
- Species
- Date
- Remarks

**UniProt-GeneID**
- Primary
- Related
- Bridge

**EnsemblBacteria**
- PK | ID
- Species
- Date
- Remarks

**UniProt-EnsemblBacteria**
- Primary
- Related
- Bridge

**OrderedLocusNames**
- PK | ID
- Species
- Date
- Remarks

**UniProt-OrderedLocusNames**
- Primary
- Related
- Bridge

**UniProt**
- PK | ID
- EntryName
- GeneName
- ProteinName
- Function
- Species
- Date
- Remarks

**UniProt-InterPro**
- Primary
- Related
- Bridge

**InterPro**
- PK | ID
- Species
- Date
- Remarks

**UniProt-Pfam**
- Primary
- Related
- Bridge

**Pfam**
- 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

## 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).