

BioHealthBase: BioDefense Public Health Database

Richard Scheuermann^{*1}, Burke Squires¹, Shubhada Godbole¹, Kevin Biersack, Carey Gire, Zuoming Deng, Aihui Wang, Jianjun Wang, Tammie Ajayi, Sanjeev Kumar, Jason Lucas, Soumya Sengupta, Surabhi Sharma, Paul Shrabstein, Sharmila Vattikuti, Bryan Walters, Chris Larsen², Al Ramsey², James Wolowicz², Xianying Wei³

Northrop Grumman Life Sciences, ^{*}Principal Investigator, ¹University of Texas Southwestern Medical Center, ²Vecna Technologies, and ³Amar International

www.biohealthbase.org

INTRODUCTION

The Biodefense/Public Health DataBase (BioHealthBase) BRC focuses on six priority pathogens to help fill in gaps in genomic and other data critical to scientific researchers. The six pathogens are *Francisella tularensis* bacteria; *Influenza* (flu) virus; *Mycobacterium tuberculosis* (TB) bacteria; *Microsporidia* parasites; *Giardia lamblia* (giardiasis) parasite; and *Ricinus communis* (castor bean) plant (toxin).

BIOHEALTHBASE OFFERS

- Integrated data sets from NCBI, UniProt, Pfam, BioCyc, IEDB, and other sources
- Web-based data-mining and visualization tools
- Structural features and functional annotations for gene and protein sequences
- Metabolic and signaling pathway annotation
- Host-pathogen interaction models
- Protein localization and operon predictions (bacteria)
- Epitope predictions and validated epitopes from the IEDB (influenza)
- Influenza sequence alignments and polymorphism frequencies

The screenshot shows the BioHealthBase homepage with a navigation bar at the top containing 'BIOHEALTHBASE HOME', 'FEEDBACK', and 'ORGANISMS'. The main content area is divided into three columns. The left column contains links for 'Release Notes', 'FAQ', 'Feedback', 'Related Links', 'Data Loads', and 'Science Support'. The middle column, titled 'Organisms', lists six pathogens with corresponding icons: Influenza Virus, Francisella tularensis, Mycobacterium, Microsporidia, and Giardia. The right column, titled 'Our Mission', contains a paragraph about the system's purpose and a '< More >' link. Below the organisms list is a 'News/Publications/Releases' section with a list of recent events. To the right of this is a 'Genomes in BioHealthBase' section with two tables: 'Species-based' and 'Kingdom-based'.

Organism	Kingdom	# Strain
Francisella tularensis	Bacteria	2
Mycobacterium tuberculosis	Bacteria	2
Mycobacterium avium	Bacteria	1
Mycobacterium leprae	Bacteria	1
Mycobacterium bovis	Bacteria	1
Encephalitozoon cuniculi	Fungi	1
Influenza A Virus	Virus	7211
Influenza B Virus	Virus	1127
Influenza C Virus	Virus	152

Kingdom	# Species	# Strain
---------	-----------	----------

FUTURE DIRECTIONS

Future releases of BioHealthBase are planned to include an automated genomic annotation pipeline, comparative genomics analysis, proteomics data tools, enhanced visualization tools for 3D protein structures, and molecular interactions and cellular networks with a user-friendly interface for data retrieval and analysis.

ACKNOWLEDGEMENTS

This project is funded in whole or in part with federal funds:

Contract No.: HHSN266200400041C

ADB Contract Number: N01-AI-40041

National Institute of Allergy and Infectious Diseases

National Institutes of Health

Department of Health and Human Services

This screenshot shows the 'Mycobacterium tuberculosis' database interface. It includes a navigation bar with 'HOME', 'FEEDBACK', 'OUR MISSION', 'HELP', and 'BIOHEALTHBASE HOME'. The main content area is divided into three sections: 'Database Search' with a list of search criteria, 'What's New' with a list of recent updates, and 'Genome Statistics' for 'Mycobacterium tuberculosis'. The 'Genome Statistics' section includes a table for 'Number of Strains in BioHealthBase' and a table for 'Mycobacterium Genomes in BioHealthBase'. Below these is a 'Sequence Polymorphism Analysis for Human H1N1, Segment 1' section with a table showing sequence data.

Category	Count
Number of Strains in BioHealthBase	5
Gram Stain	+
Genome Type	1 circular chromosome
Genome Size	~ 3 - 7 Mb
Taxon ID	1763
Kingdom	Bacteria

Genome	Count
Mycobacterium avium	1
Mycobacterium bovis	1
Mycobacterium leprae	1
Mycobacterium tuberculosis	2
Total	5

This screenshot shows the 'Francisella tularensis' database interface. It includes a navigation bar with 'HOME', 'FEEDBACK', 'OUR MISSION', 'HELP', and 'BIOHEALTHBASE HOME'. The main content area is divided into two sections: 'Blast Search' and 'Gene Ontology (GO) Search'. The 'Blast Search' section includes a form for entering a sequence and a 'GO' button. The 'Gene Ontology (GO) Search' section includes a form for entering a GO term and a 'GO' button.

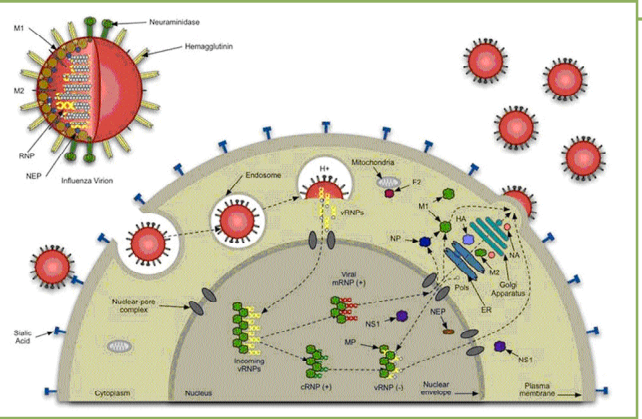
This screenshot shows the 'Microsporidia' database interface. It includes a navigation bar with 'HOME', 'FEEDBACK', 'OUR MISSION', 'HELP', and 'BIOHEALTHBASE HOME'. The main content area is divided into two sections: 'Gene Ontology (GO) Search' and 'Gene Details'. The 'Gene Ontology (GO) Search' section includes a form for entering a GO term and a 'GO' button. The 'Gene Details' section includes a table with gene information.

This screenshot shows the 'Francisella tularensis' database interface, specifically the 'Gene Details' section for a gene. It includes a table with gene information, including the gene name, accession number, coordinates, and sequence. The table also includes a 'Page Contents' section with links to various parts of the page.

Strain Identification
Organism Name: Francisella tularensis
Strain Name: Schu 4
NCBI Taxon ID: 177416

Gene Identification
Gene Symbol: IpxC
Gene Product Name: UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
Locus Name: FTT0189
Entrez Gene ID: 3190839
Comment: Similar to LPXC, ECOLI (P07652) UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase from E. coli (305 aa). FASTA: opt: 912 Z-score: 1124.8 E1: 9.3e-55 Smith-Waterman score: 912; 45.645 identity in 287 aa overlap

Genomic Location
Genbank Genomic Accession: NC_006570
Coordinates(5'..3'): 204916..205776
Strand: Forward
Gene Length: 861
Sequence: View Gene Sequence



BioHealthBase
BioDefense Public Health Database
www.biohealthbase.org

This screenshot shows the 'Mycobacterium tuberculosis CDC1551, complete genome.' section. It includes a genomic map showing the location of the gene of interest and a sequence alignment showing the gene's sequence compared to the reference sequence. The alignment shows a high degree of similarity between the two sequences.

This screenshot shows the 'Influenza Virus' database interface. It includes a navigation bar with 'HOME', 'FEEDBACK', 'OUR MISSION', 'HELP', and 'BIOHEALTHBASE HOME'. The main content area is divided into two sections: 'Gene Search' and 'Sequence Alignment'. The 'Gene Search' section includes a form for entering a search term and a 'GO' button. The 'Sequence Alignment' section includes a form for entering a sequence and a 'GO' button.