

# PATRIC 2.0 BRC and BRC's Pathogen Portal

Moscow, June 9 2010

Bruno Sobral  
Professor and Director  
CyberInfrastructure Division

# PATRIC & Portal Project Team

## Cyberinfrastructure Division

Bruno Sobral, PI	Rebecca Wattam
Stephen Cammer	Rebecca Will
Oral Dalay	Kelly Williams
Tim Driscoll	Tian Xue
Joe Gabbard	Harry Yoo
Joe Gillespie	Chengdong Zhang
Debby Hix	
Ron Kenyon	
Jian Lu	
Shrinivasrao Mane	
Chunhong Mao	
Eric Nordberg	
Eric Snyder	
Mark Scott	
Maulik Shukla	
Julie Shulman	
Jim Stoll	
Dan Sullivan	



## SEED Team (ANL & U of Chicago)

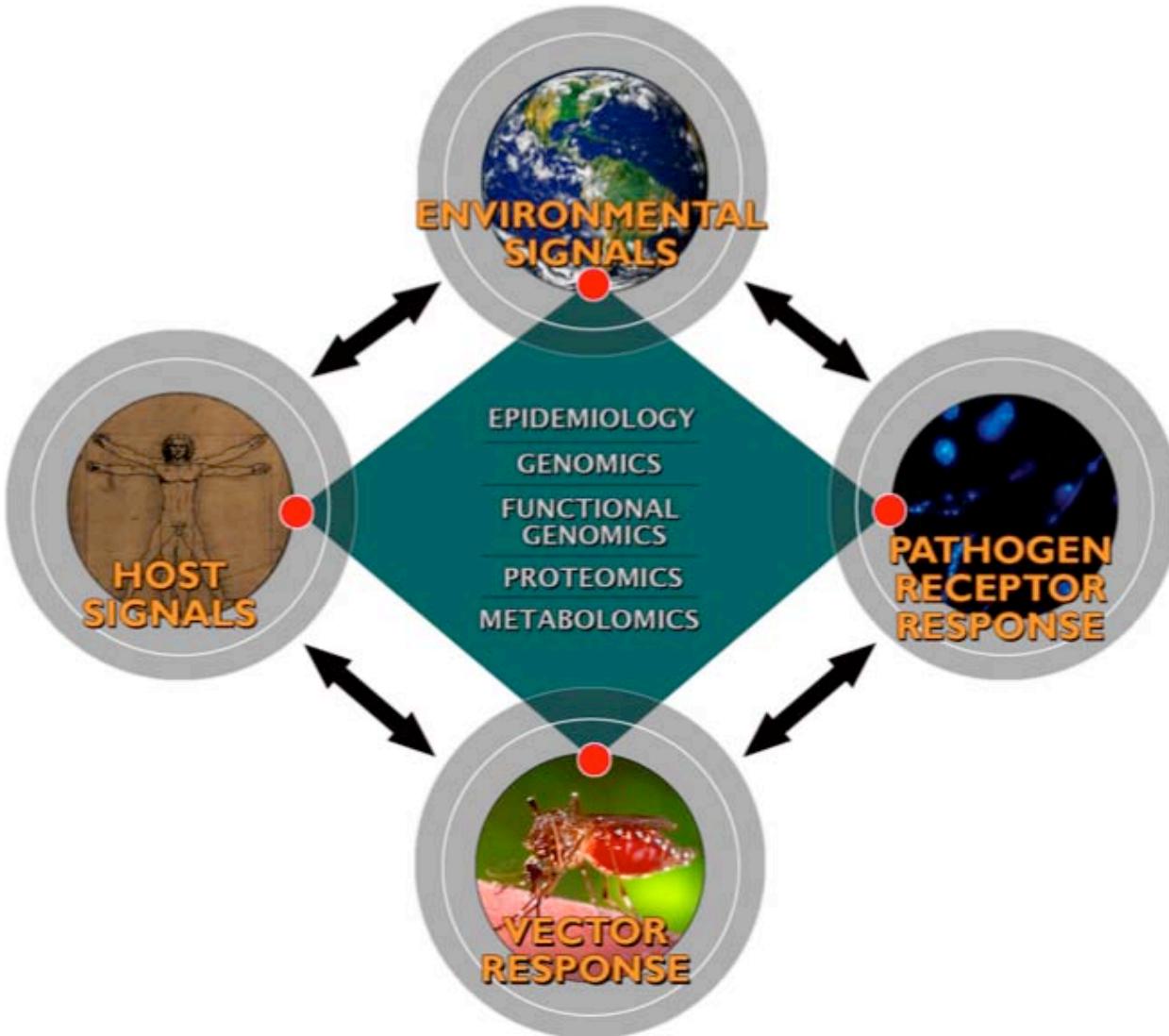
Rick Stevens, PI	Ross Overbeek
Terrence Disz	Veronika Vonstein
Folker Meyer	Gordon Pusch
Robert Olson	
Narayan Desai	

## NaCTeM (U of Manchester & U Tokyo)

Sophia Ananiadou, PI; Junichi Tsujii

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900040C to B.W.S. Sobral.

# An Integrative View of ID



# CyberInfrastructure

Community-Specific Knowledge Environments for Research and Education  
(*collaboratory, co-laboratory, grid community, e-science community, virtual community*)

Customization for discipline- and project-specific applications

High performance computation services	Data, information, knowledge management services	Observation, measurement, fabrication services	Interfaces, visualization services	Collaboration services
Networking, Operating Systems, Middleware				
Base Technology: computation, storage, communication				

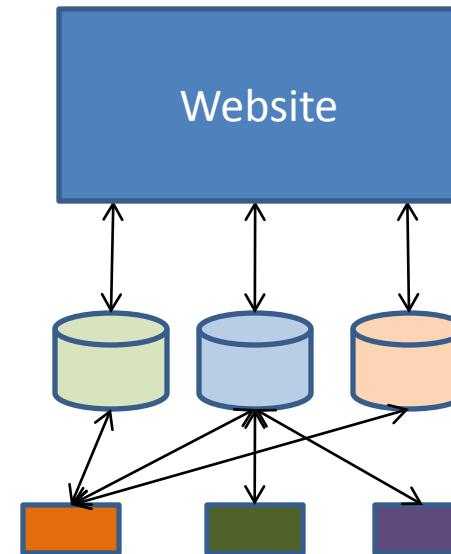
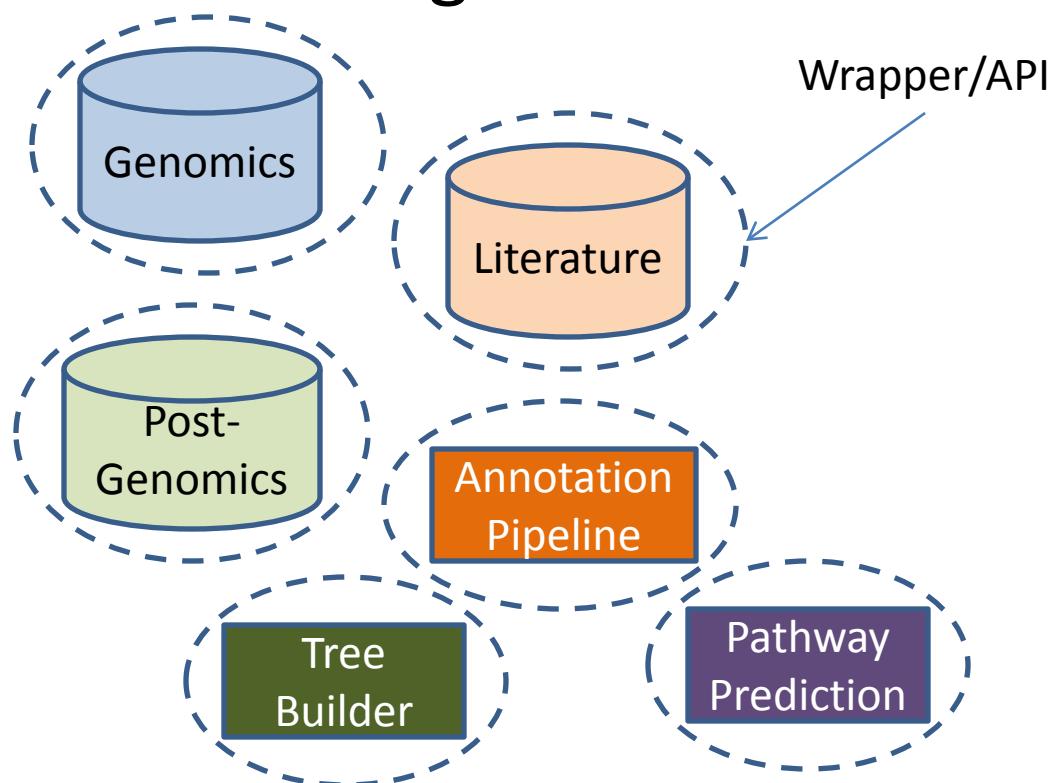


= *cyberinfrastructure: hardware, software, services, personnel, organizations*

# Primordial Ooze (“Warm Little Pond”)

## CyberInfrastructure Concept

- Modular, loosely-coupled components are developed that can be incorporated and reconfigured into the broader infrastructure



# Bacterial Data from Previous Bacterial BRCs Has Migrated to PATRIC 2.0



[FAQ](#) | [Publications](#) | [Training](#) | [Pathogen Portal](#)



Bioinformatics Resource Center

National Microbial Pathogen Data Resource



*Enteropathogen Resource Integration Center*

Bioinformatics Resource Center



New All-Bacterial BRC: **PATRIC 2.0**  
Includes Pathogenic Bacteria and All Relatives

# PATRIC Watchlist Genera

- *Bacillus*
- *Bartonella*
- *Borrelia*
- *Brucella*
- *Burkholderia*
- *Campylobacter*
- *Chlamydophila*
- *Clostridium*
- *Coxiella*
- *Ehrlichia*
- *Escherichia*
- *Francisella*
- *Helicobacter*
- *Listeria*
- *Mycobacterium*
- *Rickettsia*
- *Salmonella*
- *Shigella*
- *Staphylococcus*
- *Streptococcus*
- *Vibrio*
- *Yersinia*

# Genomics Data at PATRIC

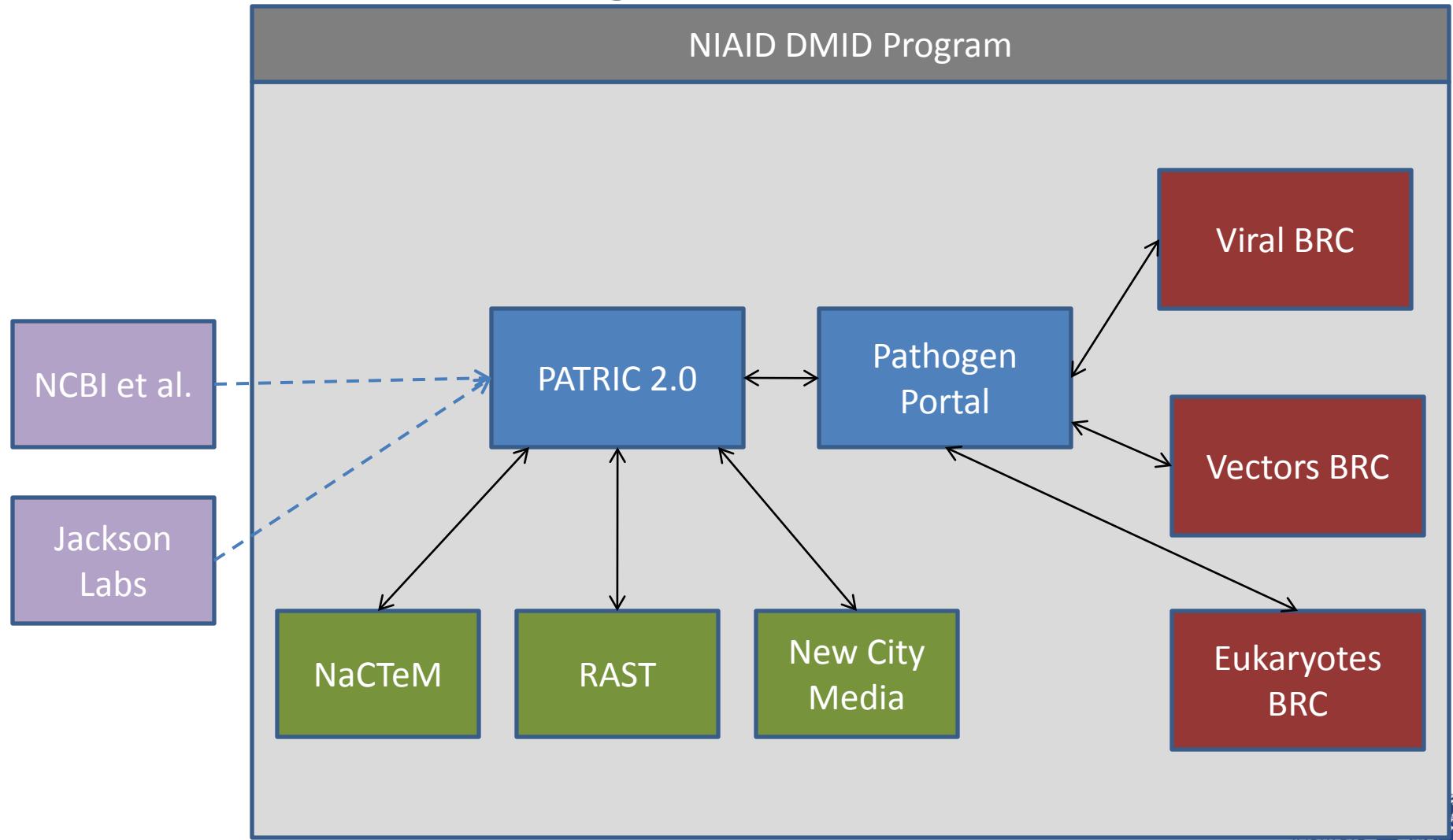
	PATRIC	Legacy BRCs	RefSeq/GenBank
Complete genomes (Closed + WGS)	2,133	405	2203
Genes	7,719,370	645,120	7,955,254
CDS	7,557,069	1,407,677	7,697,142
RNAs	146,516	37,469	150,667
Proteins w functional assignments	5,475,516	1,109,941	5,679,612
Proteins w EC assignments	1,622,909	290,271	746,246
Proteins w GO assignments	1,210,012	727,546	21,587
Proteins w pathway assignments	1,173,078	227,121	600,168

# Post-Genomics Data at PATRIC

Data Type	Data Source	No. of Data Sets
Expression Data	NCBI GEO	1,479
	EBI ArrayExpress	423
	VBI PRC	5
Proteomics Data	NCBI Peptidome	10
	EBI PRIDE	155
	VBI PRC	12
Protein Structures	NCBI MMDB	23,451
Protein-Protein Interactions	EBI IntAct	39,636

\* Data and functionality will be released on the PATRIC website in July, 2010.

# PATRIC 2.0 & Pathogen Portal Macro-Organization



# **PATRIC 2.0 WEBSITE**

## **(PATRICBRC.ORG)**

# Taxon Summary Page

Overview      Taxonomy      Phylogeny      Genomes      Features      Literature      Google

The screenshot displays the PATRIC Taxon Summary Page for the genus *Brucella*. The navigation bar at the top includes links for Home, Organisms, Searches & Tools, Downloads, About PATRIC, and Contact Us. Below the navigation bar, the main content area is organized into several sections:

- Taxonomy:** Shows the taxonomic path: Bacteria > Proteobacteria > ... > Brucellaceae > Brucella. A callout arrow labeled "Taxonomy" points to this section.
- Searches:** Includes sections for Genome Finder, Genomic Features, Pathway Finder, and Protein Family Sorter. A callout arrow labeled "Searches" points to this section.
- Recent PubMed Articles:** Lists recent publications related to *Brucella*, such as "Herd-level factors for *Brucella* seropositivity in cattle reared in smallholder dairy farms of Zimbabwe" by Matope G et al. in Prev Vet Med.
- Sequence Summary:** Provides statistics on the number of genomes (32), sequences (458), and complete genomes (32) available in PATRIC, Legacy BRC, and RefSeq databases.
- Genomic Feature Summary:** Breaks down genomic features by type: CDS (110163), gene (112950), misc\_rRNA (52), misc\_binding (0), misc\_feature (0), misc\_signal (0), ncRNA (0), pseudogene (0), rRNA (168), region (0), repeat\_region (0), source (458), tRNA (1612), tmRNA (0), and transcript (0).
- Protein Feature Summary:** Breaks down protein features by type: ATP (102), GTP (0), Nucleic acid (0), and others.
- Google:** A search interface for Google, showing results for "[Brucella] matches in" across various categories like Countermeasure, Patents Search, Web Search, News Search, Images Search, Videos Search, Books Search, and Blogs Search.



# Literature Integration

- **Challenges**

- A top 3 requirement for infectious disease researchers
- Data management and synchronization
- Integrating unstructured data

- **Solutions**

- Use of APIs
  - No local storage, fewer synchronization issues, minimal maintenance
- Smart queries
  - Dynamic on the fly queries
  - Use of ontologies and taxonomies
  - Faceted searches

The screenshot shows a search results page from PubMed. The search term 'PATRIC' was entered into the search bar. The results list four articles:

- 1. **The posterior subthalamic area in the treatment of movement disorders: past, present, and future.**  
Blomstedt P, Sandkø U, Fytaghoris A, Tisch S.  
Neurosurgery. 2009 Jun;64(6):1029-39, discussion 1038-42.  
PMID: 19467881 [PubMed - in process]  
[Related Articles](#)
- 2. **Gender, age, and body surface area are the major determinants of ascending aorta dimensions in subjects with apparently normal echocardiograms.**  
Bragg P, Matthews F, Braun J, Rousson V, Kaufmann PA, Jenni R.  
J Am Soc Echocardiogr. 2009 Jun;22(6):720-5. Epub 2009 May 7.  
PMID: 19422392 [PubMed - in process]  
[Related Articles](#)
- 3. **Chaotic transport in deterministic sine-Gordon soliton ratchets.**  
Müller P, Mertens FO, Bishop AR.  
Phys Rev E Stat Nonlin Soft Matter Phys. 2009 Jan;79(1 Pt 2):016207. Epub 2009 Jan 20.  
PMID: 19257122 [PubMed]  
[Related Articles](#)
- 4. **Rickettsia phylogenomics: unwinding the intricacies of obligate intracellular life.**  
Gillespie JJ, Williams K, Shukla M, Snyder EE, Nordberg BK, Ceraul SM, Ocharanola C, Rainey D, Soneja J, Shalom JM, Vishnubhat ND, Wattam R, Purkayastha A, Czar M, Crasta O, Setubal JC, Azad AF, Sobral BS.  
PLoS ONE. 2009;4(4):e1010. Epub 2009 Apr 16.

# Literature Integration

**Brucella**

Download Data Pathogen Information at PathInfo( 1 2 ) NCBI Taxonomy Browse

Overview Genomes Feature Table Phylogenetic Tree Ortholog Groups Pathways Additional Resources Collaborative Research Experiment Data

3D Structure Literature

GenBank Submission literature entries are publications cited in the GenBank Brucella file(s). SwissProt literature entries are publications associated with Brucella through SwissProt annotation.

PubMed GenBank Submission Swiss Prot

**8593 publications are available.**

**Filter Publications**

**by Date**

- [Coming Soon](#)
- [Past Week](#)
- [Past Month](#)
- [Past Year](#)
- [All](#)

**by Keyword**

- [Countermeasures](#)
- [Diagnosis](#)
- [Disease](#)
- [Epidemiology](#)
- [Gene expression](#)
- [Genome](#)
- [Host](#)
- [Pathogenesis](#)
- [Prevention](#)
- [Proteomics](#)
- [Reservoir](#)
- [Taxonomy](#)
- [Transmission](#)
- [Treatment](#)
- [All](#)

**Download Table Contents as:** [Excel](#) or [Plain Text](#)

- 1. [A brucellosis case with macular rash and peripheral neuropathy]**  
Kaya S, Kostakoğlu U  
Mikrobiyoloji bulteni (2009 Jan)  
PubMed: [19334392](#) Full Text   
 Abstract
- 2. [A case of brucellosis complicated with endocarditis, pyelonephritis, sacroileitis and thyroiditis]**  
Acar A, Turhan V, Diktaş H, Oncül O, Cavuşlu S  
Mikrobiyoloji bulteni (2009 Jan)  
PubMed: [19334391](#) Full Text   
 Abstract
- 3. [Retrospective evaluation of 15 cases with psoas abscesses]**  
Turunç T, Turunç T, Demiroğlu YZ, Colakoğlu S  
Mikrobiyoloji bulteni (2009 Jan)  
PubMed: [19334388](#) Full Text   
 Abstract
- 4. [In vitro effect of reactive nitrogen and oxygen intermediates alone and in combination with some antibiotics against *Brucella melitensis* clinical isolates]**  
Tanyel E, Coban AY, Fışgin NT, Tülek N, Durupınar B  
Mikrobiyoloji bulteni (2009 Jan)  
PubMed: [19334376](#) Full Text   
 Abstract
- 5. [Spondilodiscitis: five years of experience in a department of rehabilitation]**  
Ribeira T, Veiros I, Nunes R, Martins L  
Acta medica portuguesa (2008 Nov-Dec)  
PubMed: [19331789](#) Full Text   
 Abstract



# Literature

[Bacteria](#) > [Proteobacteria](#) > ... > [Brucellaceae](#) > [Brucella](#)

[Overview](#) [Taxonomy](#) [Phylogeny](#) [Genome List](#) [Feature Table](#) [Literature](#) [Google Search](#)

## Literature Portlet

### Filter Publications

#### By Date:

[Coming Soon](#)

[Past Week](#)

[Past Month](#)

[Past Year](#)

[All](#)

#### By Keyword:

[Transmission](#)

[Host](#)

[Disease](#)

[Gene expression](#)

[Reservoir](#)

[Prevention](#)

[Pathogenesis](#)

[Proteomics](#)

[Countermeasures](#)

[Epidemiology](#)

[Diagnosis](#)

[Taxonomy](#)

[Genome](#)

[All](#)

### Publication

#### **Herd-level factors for Brucella seropositivity in cattle reared in smallholder dairy farms of Zimbabwe.**

Matope G, Bhebhe E, Muma JB, Lund A, Skjerve E

Preventive veterinary medicine (2010 May 1), PubMed: [20116870](#)

[Abstract](#)

#### **Pathogenesis of bovine brucellosis.**

Carvalho Neta AV, Mol JP, Xavier MN, Paixão TA, Lage AP, Santos RL

Veterinary journal (London, England : 1997) (2010 May), PubMed: [19733101](#)

[Abstract](#)

#### **Brucella: a fastidious bacteria but a virulent pathogen.**

Bardenstein S, Strada V, Banai M

Veterinary journal (London, England : 1997) (2010 May), PubMed: [19716325](#)

[Abstract](#)

#### **Inhibition studies of a beta-carbonic anhydrase from Brucella suis with a series of water soluble glycosyl sulfonilamides.**

Vullo D, Nishimori I, Scozzafava A, Köhler S, Winum JY, Supuran CT

Bioorganic & medicinal chemistry letters (2010 Apr 1), PubMed: [20211561](#)

[Abstract](#)

#### **Cloning and expression of the immunoreactive Brucella melitensis 28 kDa outer-membrane protein (Omp28) encoding gene evaluation of the potential of Omp28 for clinical diagnosis of brucellosis.**

Thavaselvam D, Kumar A, Tiwari S, Mishra M, Prakash A

Journal of medical microbiology (2010 Apr), PubMed: [20075115](#)

[Abstract](#)

#### **Knowledge and attitudes towards food safety among Canadian dairy producers.**

Young I, Hendrick S, Parker S, Raji? A, McClure JT, Sanchez J, McEwen SA

Preventive veterinary medicine (2010 Apr 1), PubMed: [19962773](#)

[Abstract](#)

#### **Cisplatin reduces Brucella melitensis-infected cell number by inducing apoptosis, oxidant and pro-inflammatory cytokine**

# Google Search

Bacteria > Proteobacteria > ... > Brucellaceae > Brucella

Overview Taxonomy Phylogeny Genome List Feature Table Literature Google Search

## Google Search

Brucella

Powered by 

Countermeasure (79,700) Patents (747) Web (136,000) News (24) Images (21,600) Videos (41) Books (2,616) Blogs (10,334)

### [Brucellosis - Wikipedia, the free encyclopedia](#)

Brucellosis, also called Bang's disease, Gibraltar fever, Malta fever, Maltese fever, Mediterranean fever, rock fever, or undulant fever, ...  
[en.wikipedia.org](http://en.wikipedia.org)

### [Brucella - Wikipedia, the free encyclopedia](#)

Brucella is a genus of Gram-negative bacteria. They are small (0.5 to 0.7 by 0.6 to 1.5  $\mu\text{m}$ ), non-motile, non-encapsulated coccobacilli, which function as ...  
[en.wikipedia.org](http://en.wikipedia.org)

### [Brucellosis, disease information, NCID, CDC](#)

Brucellosis: (Brucella melitensis, abortus, suis, and canis) ... **Brucellosis** From Public Health Response and Preparedness, CDC ...  
[www.cdc.gov](http://www.cdc.gov)

### [Disease Listing, Brucellosis, General Information | CDC Bacterial ...](#)

Dec 7, 2007 ... Factsheet with cause, transmission, prevention, symptoms, and treatment.  
[www.cdc.gov](http://www.cdc.gov)

### [CBRNE - Brucellosis: eMedicine Emergency Medicine](#)

Overview: **Brucellosis** is a zoonotic infection transmitted from animals to humans by ingestion of infected food products, direct contact with an infected ...  
[emedicine.medscape.com](http://emedicine.medscape.com)

### [CDC Brucellosis | Emergency Preparedness & Response](#)

Provided by the CDC Division of Bacterial & Mycotic Diseases; Technical Information about **Brucellosis** Provided by the CDC Division of Bacterial & Mycotic ...  
[www.bt.cdc.gov](http://www.bt.cdc.gov)

### [Brucella abortus](#)

**Brucella** abortus (Schmidt 1901) Mever and Shaw 1920 (Approved Lists 1980) ... **Brucella** melitensis biovar Abortus > **Brucella** melitensis bv. Abortus ...

INSTITUTE  
AT VIRGINIA TECH

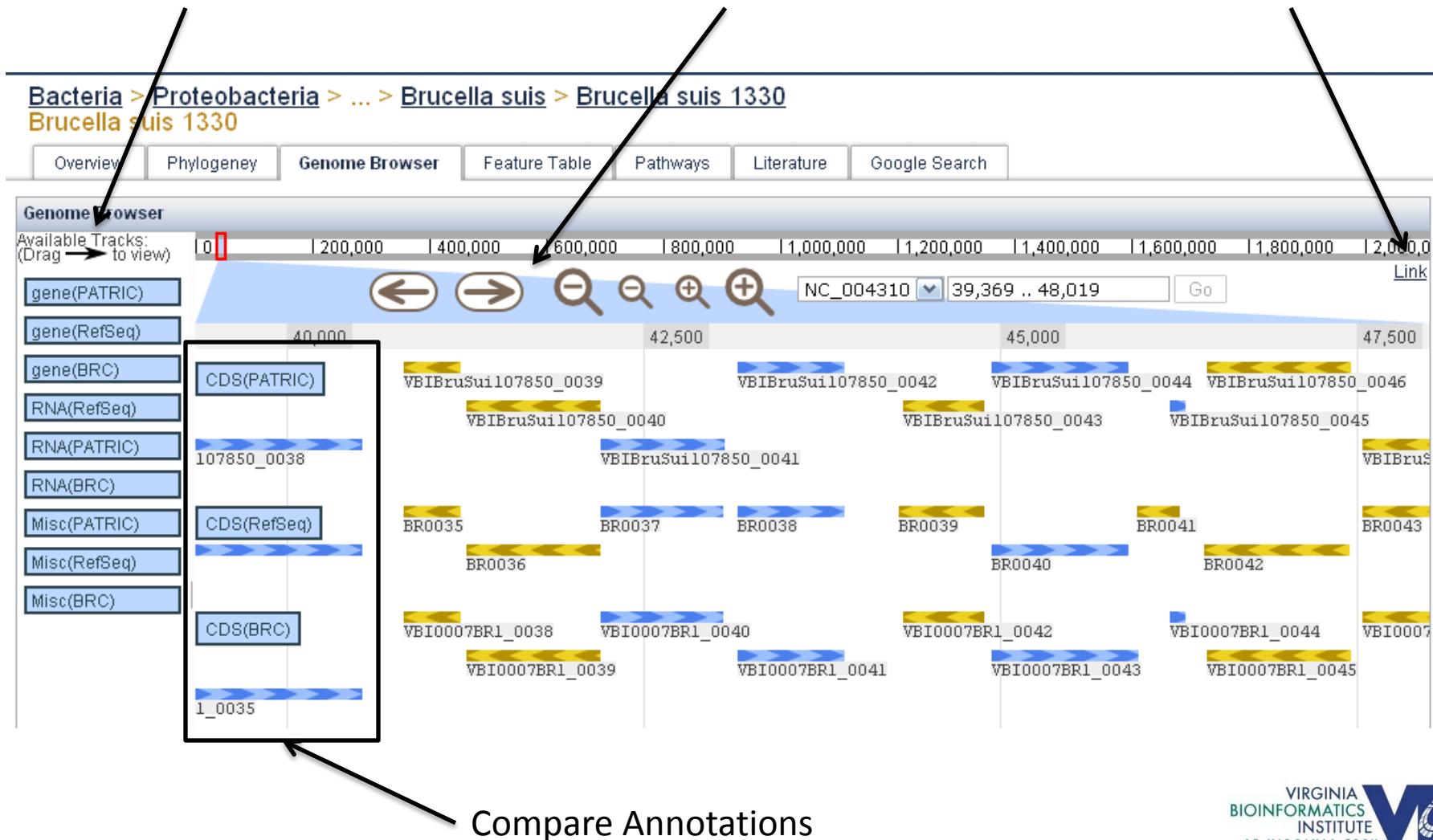


# Genome Browser - JBrowse

Drag and Drop Tracks

Zoom and Pan

Save Link



# Metabolic Pathways

Bacteria > Proteobacteria > ... > Brucella suis > Brucella suis 1330

Brucella suis 1330

Overview Phylogeny Genome Browser Feature Table Pathways Literature Google Search

**Pathway List**

**Filter** Pathway Class: ALL

**Filter Table**

Add to Group Retrieve

**Comparison Matrix**

Print Table

	EC Number	PATRIC	Legacy	RefSeq
Brucella suis 1330	1.1.1.1	4	3	0
Brucella suis 1330	1.1.1.27	0	1	0
Brucella suis 1330	1.2.1.12	1	0	1
Brucella suis 1330	1.2.1.3	3	0	0
Brucella suis 1330	1.2.4.1	2	1	2
Brucella suis 1330	1.8.1.4	3	3	3
Brucella suis 1330	2.3.1.12	1	3	1
Brucella suis 1330	2.7.1.2	1	1	1
Brucella suis 1330	2.7.1.40	1	1	1
Brucella suis 1330	2.7.1.69	0	0	1
Brucella suis 1330	2.7.2.3	1	0	1
Brucella suis 1330	3.1.3.11	2	1	2
Brucella suis 1330	4.1.1.49	1	1	1
Brucella suis 1330	4.1.2.13	1	1	1
Brucella suis 1330	4.2.1.11	1	1	1
Brucella suis 1330	5.1.3.3	1	1	1
Brucella suis 1330	5.3.1.1	2	2	2
Brucella suis 1330	5.3.1.9	1	1	1
Brucella suis 1330	5.4.2.1	1	0	1
Brucella suis 1330	5.4.2.2	1	2	1
Brucella suis 1330	6.2.1.1	2	1	2

**Pathway Map**

Carbohydrate Metabolism >>00010 : Glycolysis / Gluconeogenesis

**Pathway Description**

Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glyceraldehyde-3P, glycerate-3P, and Glycerone-P.

**GLYCOLYSIS / GLUCONEOGENESIS**

Save Map Print

Starch and sucrose metabolism

Pentose phosphate pathway

# Protein Family Sorter

FigFam Groups Viewer

FigFam Groups Viewer

Single FigFam Viewer

## Gblocks 0.91b for FIG000025: Cell division protein FtsW

Number of sequences: 32

Alignment assumed to be: Protein

New number of positions: 385 (selected positions are underlined in blue)

	10	20	30	40	50	60
Brucel	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
ruccella_abor	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
cella_abortu	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
ruccella_abor	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Brucella_ab	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Brucella_ab	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Brucella_ab	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Bru	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Bru	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Bru	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Brucella_m	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
a_melitensis	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
ucella_melit	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
ella_meliten	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
ella_meliten	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Bruce	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Br	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					
Bruc	M <u>V</u> S <u>R</u> V <u>D</u> R <u>G</u> P <u>V</u> A <u>N</u> WWWT <u>I</u> D <u>R</u> F <u>F</u> L <u>A</u> A <u>C</u> L <u>A</u> L <u>M</u> G <u>L</u> G <u>I</u> L <u>S</u> F <u>A</u> A <u>S</u> P <u>A</u> V <u>E</u> R <u>I</u> G <u>L</u> N <u>S</u> F <u>H</u> F <u>V</u> E <u>R</u> O <u>I</u> FF <u>M</u> V <u>P</u> A					



# Protein Structure Target Selection & Annotation

Coronavirus: SARS coronavirus GZ-B: VBI2138CR\_0002.02.15: nsp15

[More Details at PDB](#)

PDB: 2OZK : Chain A, Structure Of An N-Terminal Truncated Form Of Nendou (Nsp15) From Sars-Coronavirus

Query Begin:	1	Sub Begin:	1	Query Coverage:	100%	Identity:	100%	P-value:	5e-203
Query End:	346	Sub End:	346	Sub Coverage:	100%	Positive:	100%	Case:	CASE 1

Epitope  
Selection Box

Highlight Protein Areas of Interest

IEDB Epitopes

Epitope  
Information

Area of Interest Details

Epitope ID

Name

Peptide Sequence

Range

Debug

[View NT/AA Sequence](#)

[Goto AA Evidence Page](#)

[View NCBI BLASTP Result](#)



Zoom



Mouse scroll wheel, or  
Shift+Left mouse button.

Rotate



Left mouse button

Options



Right mouse button, or  
Ctrl+Left mouse button

Appearance

Wireframes

Cartoons

Spacefill

Surface

Labels

Show

Navigation

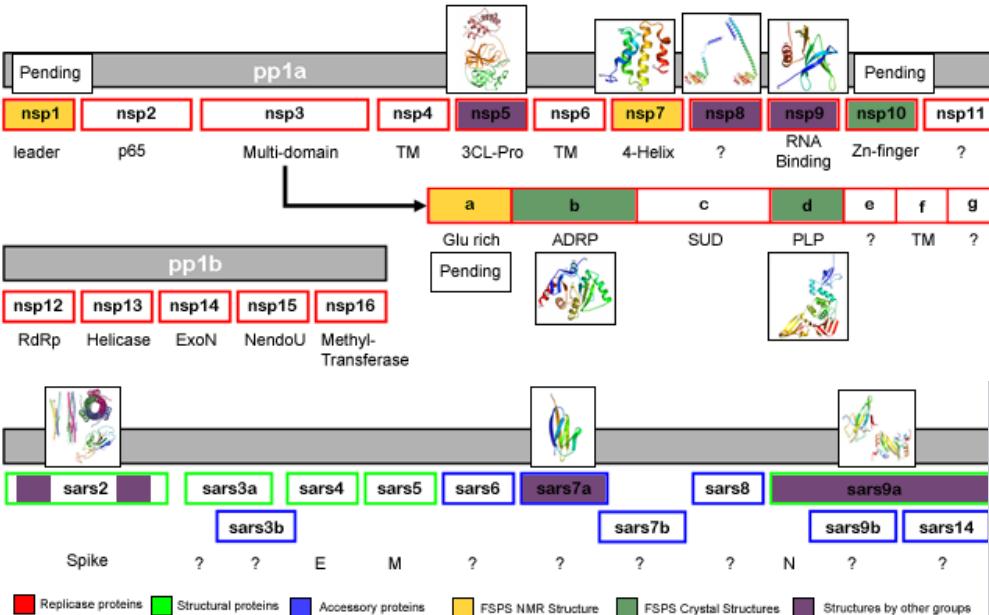
Spin

Zoom

Visualization  
Control Box

Zoom and  
Spin Control

# SARS Proteome 3D Structures



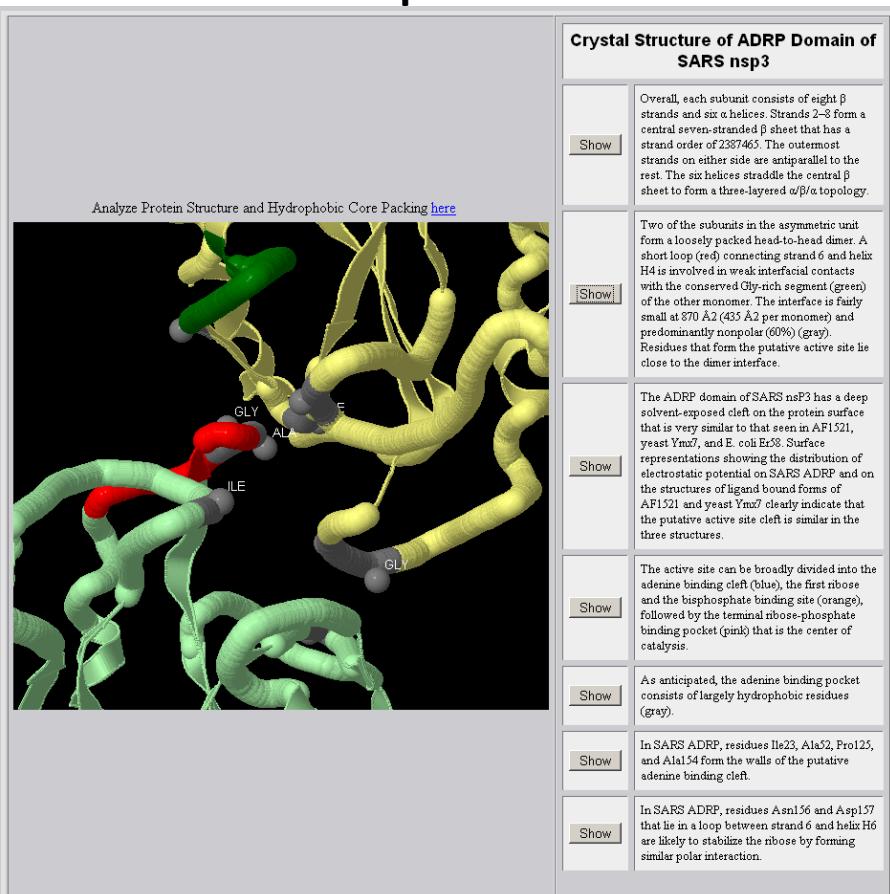
Annotations Derived from Published Manuscripts

Illustrate Features Described in Paper

Fully Interactive 3D Graphics

— Proteome Via Crystallography, NMR, Microscopy

## SARS nsp3 ADRP



SEED: University of Chicago & Argonne National Laboratory

# **THE RAST SERVER – PATRIC 2.0’S PUBLIC GENOME ANNOTATION SERVER**

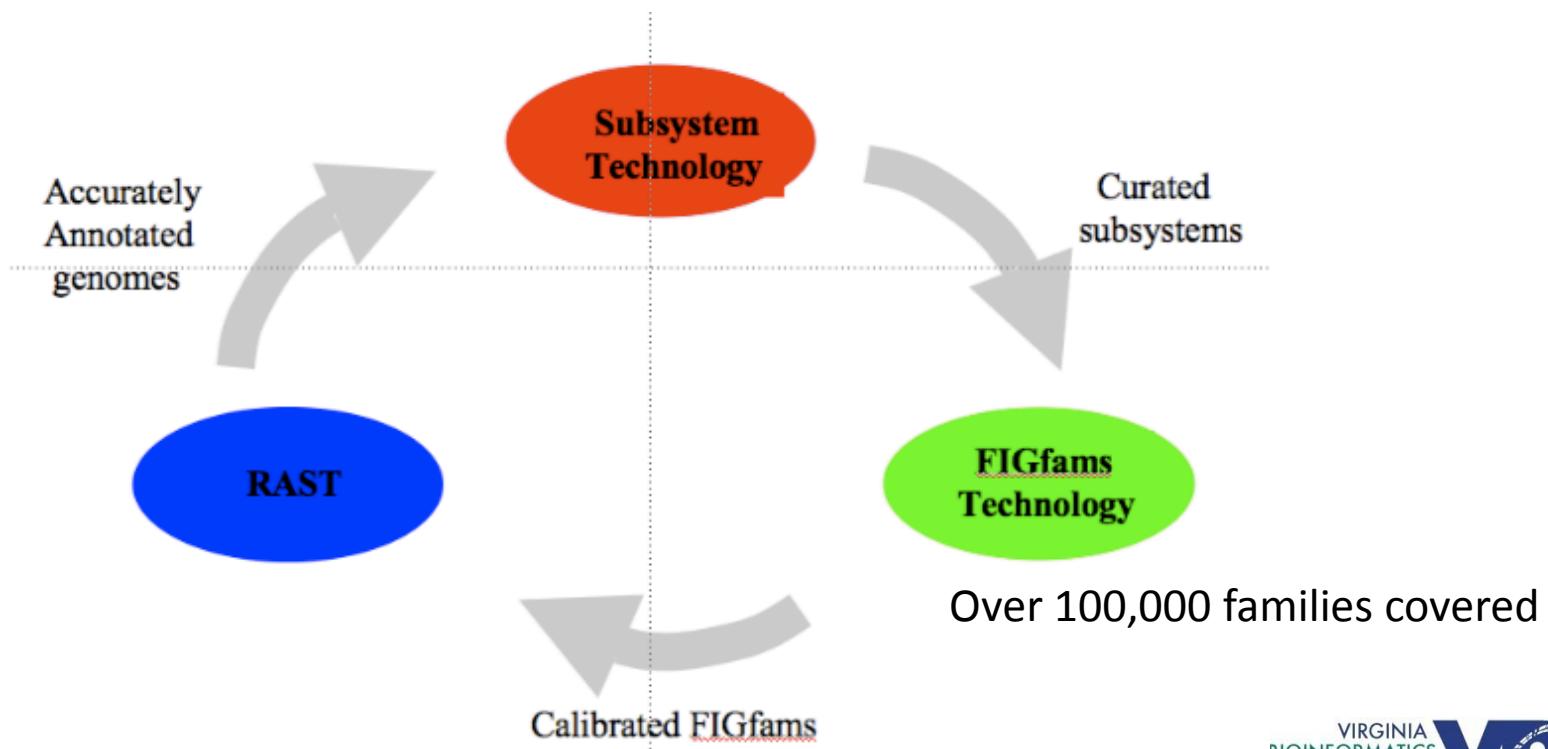
# What is RAST?

- RAST is a completely open, free annotation service for prokaryotes, like a BLAST search!
  - open-source
- **Input:** either a fasta file of contigs or a GenBank entry
- **Output (in 1-2 days):**
  - called genes (protein-encoding, rRNAs, and tRNAs)
  - functions assigned to genes
  - genes placed in subsystems
  - an initial metabolic reconstruction

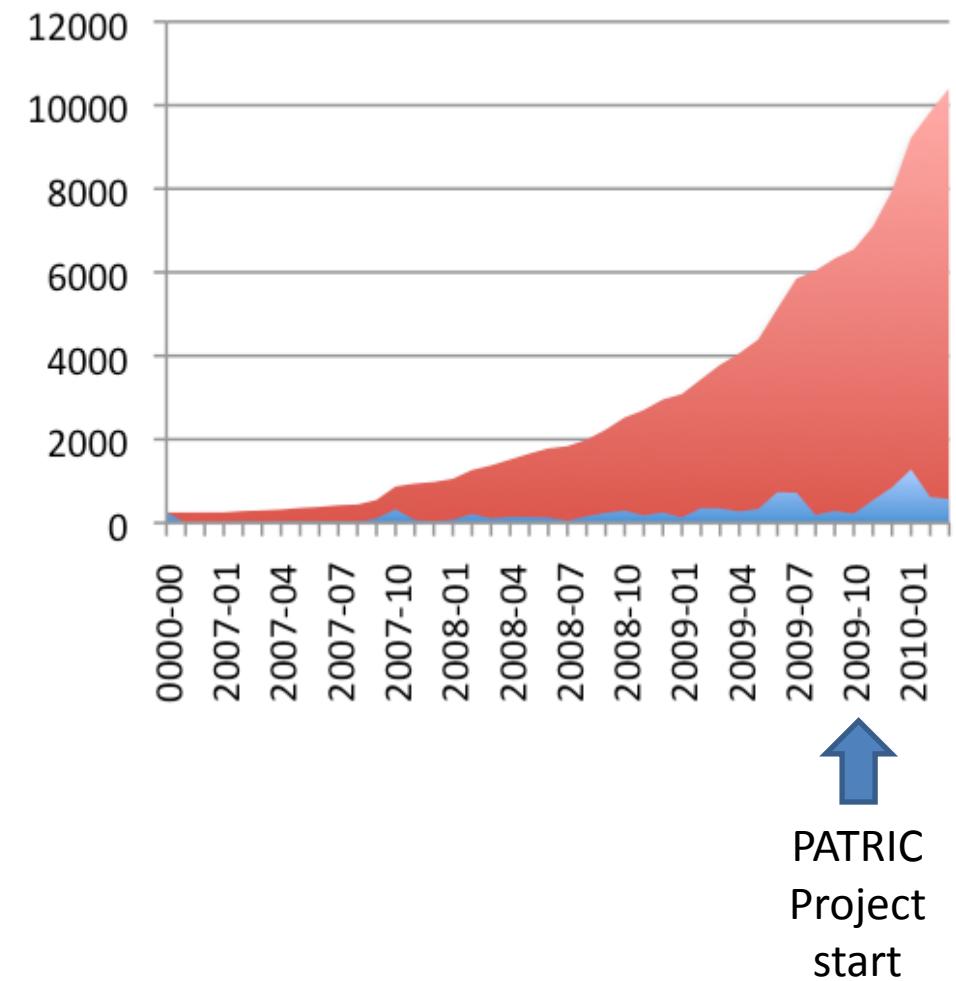
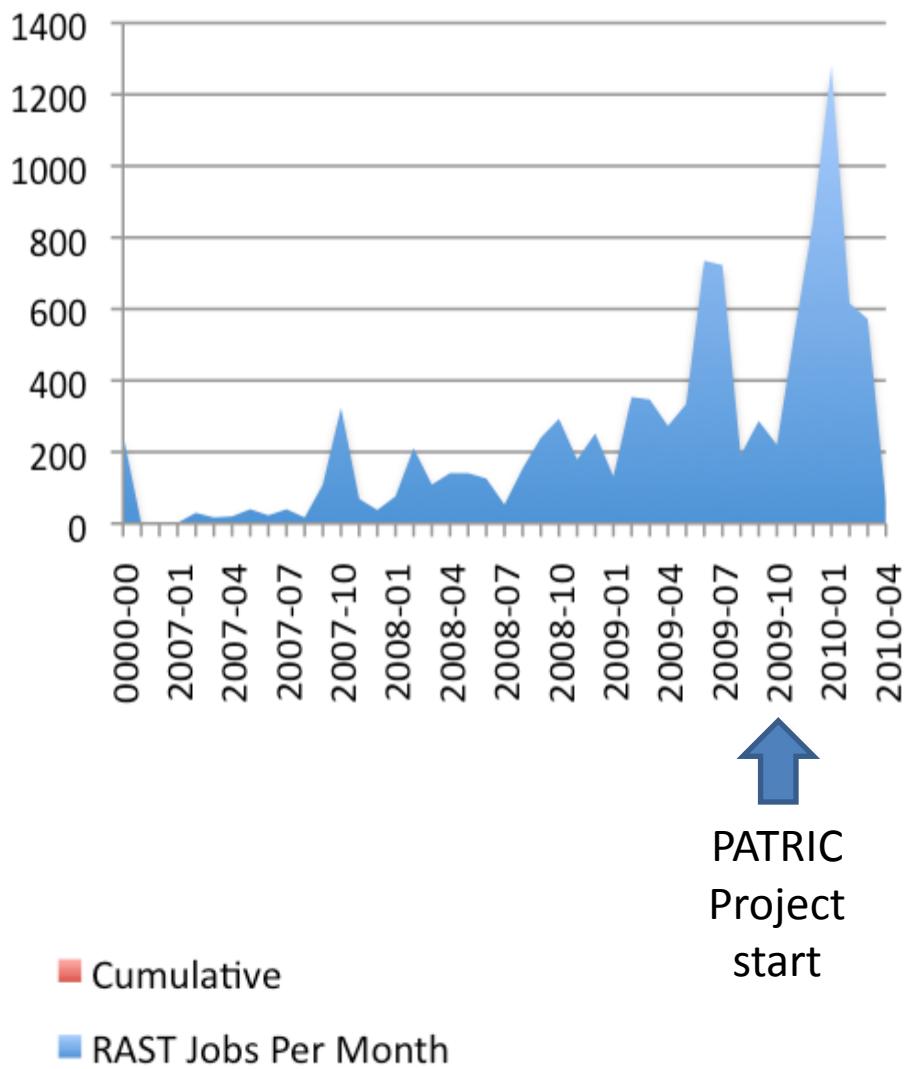


# What Determines Quality?

1. Accurately annotated core of **diverse genomes**
2. **Subsystems** that are manually curated across the entire collection of genomes, are also **computable objects**
3. **Protein families** that accurately characterize difficult cases (guided by the subsystems)



# RAST Usage Statistics



# PATRIC 2.0-RAST Interface

**PATRIC**  
PathoSystems Resource Integration Center

Home   Organisms   Searches & Tools   Downloads   About PATRIC   Contact Us

No Items, No Groups

PATRIC includes a collaboration with the University of Chicago to provide an end-user genome annotation service using the RAST system.

 **RAST** Rapid Annotation using Subsystem Technology version 2.0

The NMPDR, SEED-based, prokaryotic genome annotation service.  
For more information about The SEED please visit [theSEED.org](http://theSEED.org).

RAST (Rapid Annotation using Subsystem Technology) is a fully-automated service for annotating bacterial and archaeal genomes. It provides high quality genome annotations for these genomes across the whole phylogenetic tree.

As the number of more or less complete bacterial and archaeal genome sequences is constantly rising, the need for high quality automated initial annotations is rising with it. In response to numerous requests for a SEED-quality automated annotation service, we provide RAST as a free service to the community. It leverages the data and procedures established within the [SEED framework](#) to provide automated high quality gene calling and functional annotation. RAST supports both the automated annotation of high quality genome sequences AND the analysis of draft genomes. The service normally makes the annotated genome available within 12-24 hours of submission.

Please note that while the SEED environment and SEED data structures (most prominently [FIGfams](#)) are used to compute the automatic annotations, the data is NOT added into the SEED automatically. Users can however request inclusion of a their genome in the SEED. Once annotation is completed, genomes can be downloaded in a variety of formats or viewed online. The genome annotation provided does include a mapping of genes to [subsystems](#) and a metabolic reconstruction.

To be able to contact you once the computation is finished and in case user intervention is required, we request that users register with email address.

**If you use our service, please cite:**  
*The RAST Server: Rapid Annotations using Subsystems Technology.*  
Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsmma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.  
*BMC Genomics*, 2008, [ [article](#) ]

Please acknowledge the use of RAST supported in part by National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services (NIAD) under contract HHSN266200400042C.

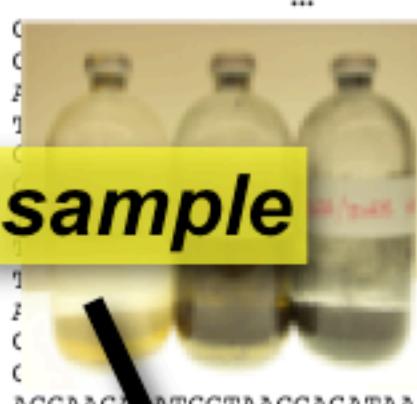
Login   
Password

» [Register a new account](#)  
» [Forgot your password?](#)

# Coming Soon to PATRIC 2.0: MG-RAST

- Handles metagenomic studies
- Will focus on science related to pathogens
  - Surveillance
  - Reservoir detection
  - Horizontal gene transfer

# Metagenomics



sample

```
TCATATAGC  
CATGAGGGA  
CGTATATCA  
TCTAAGATG  
AAGACTTAT  
GAAGTTGAG  
TTCCTTGCG  
CTGCTGTAT  
TCTTGACTG  
AGACTGCAG  
TTCCACATT  
GTTGA  
ACGAAGAATGGTAACCAGATAAACTGGT  
GATGAAGAATGCAAAGTTGGCTAGTTT  
AGTATGAAAGGCTGACATTGGCAATT  
GACTTCGAAAGATAATGAAGATGATGA  
AACAGAGTGAACCAAGAAGAAAAGGCAGC  
ATTACAGAGCTTATCAACAAAACCTAACCTTGT  
GATGAAGCAGAAAAGGACTTGGCCACCGTG  
TCRAATCCATTGATGATCCTGATGCTGC  
TTCAGGAA  
CTTAAACACA  
GAAGACTCTTATAATAACACCTTAATCCC  
TTTAAAGAGGTGCAGACTCCACAGTATTGAAAC  
CCATTGATGCCAGAACGATTTGTGACCATA  
AAGGATTCTCTCCCCAGTCTACAAAAAGAAAAA  
AATATAAGACCTGTGGATATGAGCAAGTACCTC  
TATGCTGATAGTTCTAAAACGTAAAGCAGAGCTT  
AGTGTACTGAAGCGGGAGCCTGAACACTAACAG  
CCTATCAGCGGAGCGTGACAGGTACGTGATGCT  
AGCTTTATCAGGCAGCGGTATGCGCGATCAAT  
GCGCGCGGTATAT  
TACTCTTCGGAGCT
```

**MG-RAST** Meta Genome Rapid Annotation using Subsystem Technology

Classified sequences vs. non-classified:

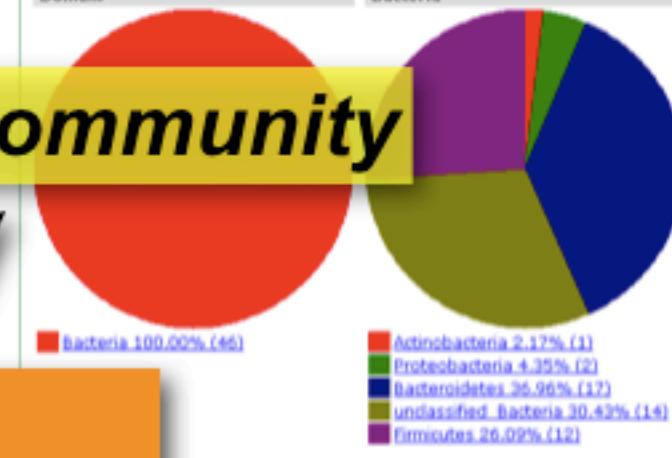
0,422 (46)

99,582 (10799)

Charts Tabular View

Domain

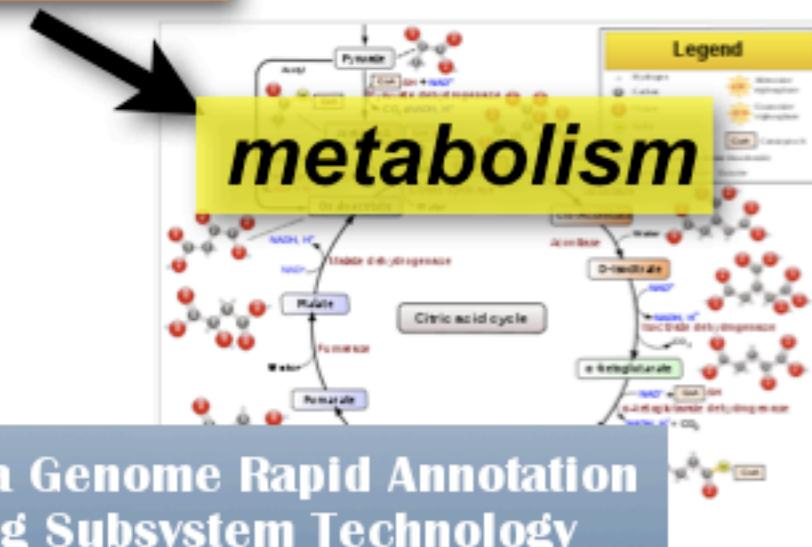
Bacteria



community

*in silico*  
reconstruction

sequence



metabolism

**Text Mining – NaCTeM collaboration (University of Manchester & University of Tokyo)**

Phylogenomics enhancements

SNP Pipeline

**UI Enhancements**

**Post-Genomic Data Integration**

**Host-Pathogen Interactions**

Disease View

**Protein Arrays**

Pan-genomes

# NEW DEVELOPMENTS UNDERWAY

# Text Mining

- Growing need for automated methods to extract facts from natural language text
- Demonstrated Success
  - Kleio<sup>1</sup> – Semantic tagging of Medline Abstracts
  - FACTA+<sup>2</sup> - Finds associated concepts with text analysis
  - Biocaster<sup>3</sup>, EpiSpider<sup>4</sup> – Disease surveillance

The screenshot shows a software interface titled 'FACTA+'. It displays a table of concepts found in a document, ordered by frequency. The columns include 'Document', 'Concept', 'Frequency', 'Occurrences', and 'Concepts'. The data includes various medical terms like 'Hypertension', 'Disease', 'Treatment', etc.

The screenshot shows a Medline abstract from J. Toxicol. Environ. Health. Part A (2009) with semantic annotations overlaid. Annotations are color-coded: green for 'SUBSTANCE', blue for 'DISEASE', red for 'SYMPTOM/ADVERSE', and orange for 'INDIRECTLY RELATED'. Annotations include 'dibutyltin', 'rat', 'blood', and 'exposure'.

The screenshot shows a dashboard titled 'EPISPIDER HOME'. It features a world map with red dots indicating disease outbreaks. A sidebar provides search and filter options for 'SEARCH', 'TOPIC CATEGORY', 'TERM IDENTIFIER', 'REPORT DATE', and 'REPORT PAGE NUMBER'. A legend at the bottom right defines colors for different news sources: DAILY, GOOGLE NEWS, HONDURE, NEUTER ALBERT, TWITTER, and FOX.

1. C. Nobata, P. Cotter, et. al. Kleio: a knowledge-enriched information retrieval system for biology. Annual ACM Conference on Research and Development in Information Retrieval. 2008
2. Y. Tsuruoka, J. Tsujii and S. Ananiadou, FACTA: a text search engine for finding associated biomedical concepts. Bioinformatics 2008; 24(21):2559-2560
3. N. Collier, S. Doan, S., et. al. BioCaster: detecting public health rumors with a Web-based text mining system., Bioinformatics 2008; 24(24):2940-2941.
4. H. Tolentino MD1, R.Kamadjeu, et. al. Scanning the Emerging Infectious Diseases Horizon - Visualizing ProMED Emails Using EpiSPIDER. Advances in Disease Surveillance 2007;2:169

# Text Mining in PATRIC 2.0

- Semantic Tagging of Medline Literature in PATRIC (next release)
- Development of named entity recognizers (NER) for infectious disease research. Pilot project with Type IV secretion systems.
- Developing event/fact extraction from full texts on infectious diseases.<sup>1</sup>

PubMedID: [18458071](#)

**Title:** Inactivation of the type IV secretion system reduces the Th1 polarization of the immune response to *Brucella abortus* infection.

**Abstract:**

The *Brucella abortus* type IV secretion system (T4SS), encoded by the virB operon, is essential for establishing persistent infection in the murine reticuloendothelial system. To gain insight into the in vivo interactions mediated by the T4SS, we compared host responses elicited by *B. abortus* with those of an isogenic mutant in the virB operon. Mice infected with the *B. abortus* virB mutant elicited smaller increases in serum levels of immunoglobulin G2a, gamma interferon (IFN-gamma), and interleukin-12p40 than did mice infected with wild-type *B. abortus*. Despite equal bacterial loads in the spleen, at 3 to 4 days postinfection, levels of IFN-gamma were higher in mice infected with wild-type *B. abortus* than in mice infected with the virB mutant, as shown by real-time PCR, intracellular cytokine staining, and cytokine levels. IFN-gamma-producing CD4(+) T cells were more abundant in spleens of mice infected with wild-type *B. abortus* than in virB mutant-infected mice. Similar numbers of IFN-gamma-secreting CD8(+) T cells were observed in the spleens of mice infected with *B. abortus* 2308 or a virB mutant. These results suggest that early differences in cytokine responses contribute to a stronger Th1 polarization of the immune response in mice infected with wild-type *B. abortus* than in mice infected with the virB mutant.

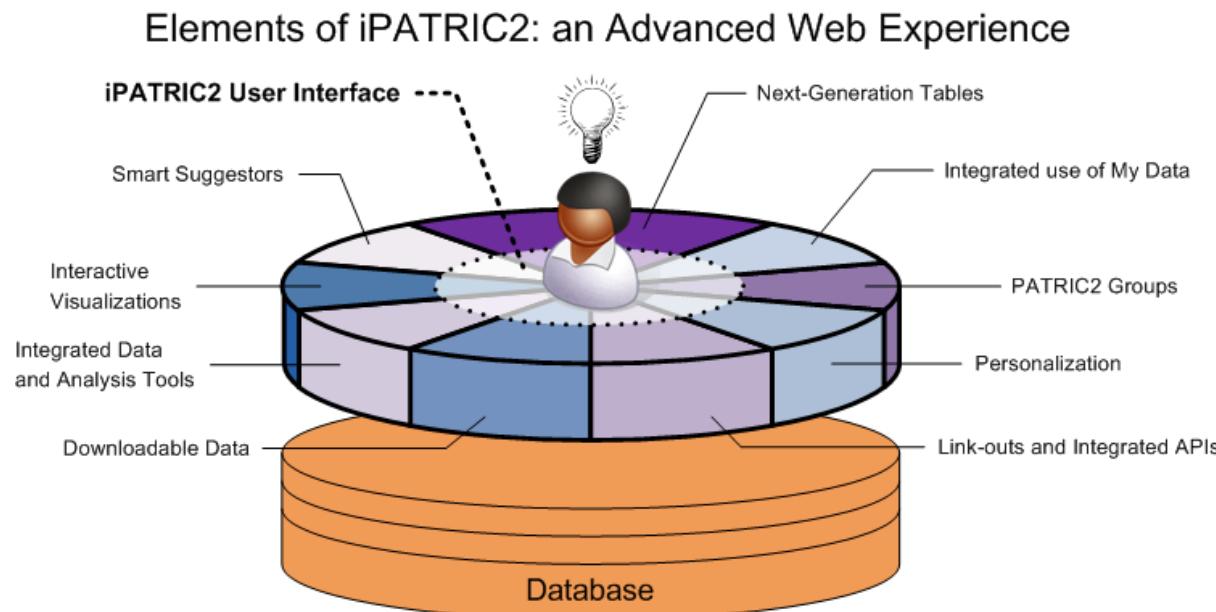
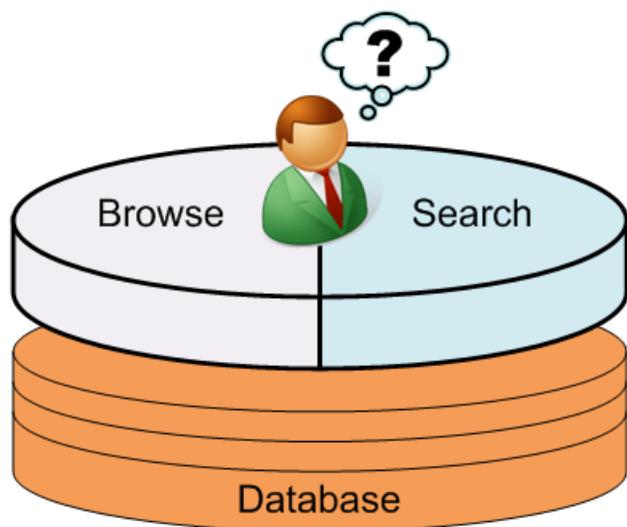
Legend:

GENE or PROTEIN METABOLITE BACTERIA ORGAN SYMPTOM or DISEASE  
PHENOMENON PROCEDURE INDICATOR  
Acronym

1. S.Pyysalo, T. Ohta, et al. "Toward Event Extraction for Full Texts on Infectious Diseases" International Conference on Knowledge Discovery and Information Retrieval 2010

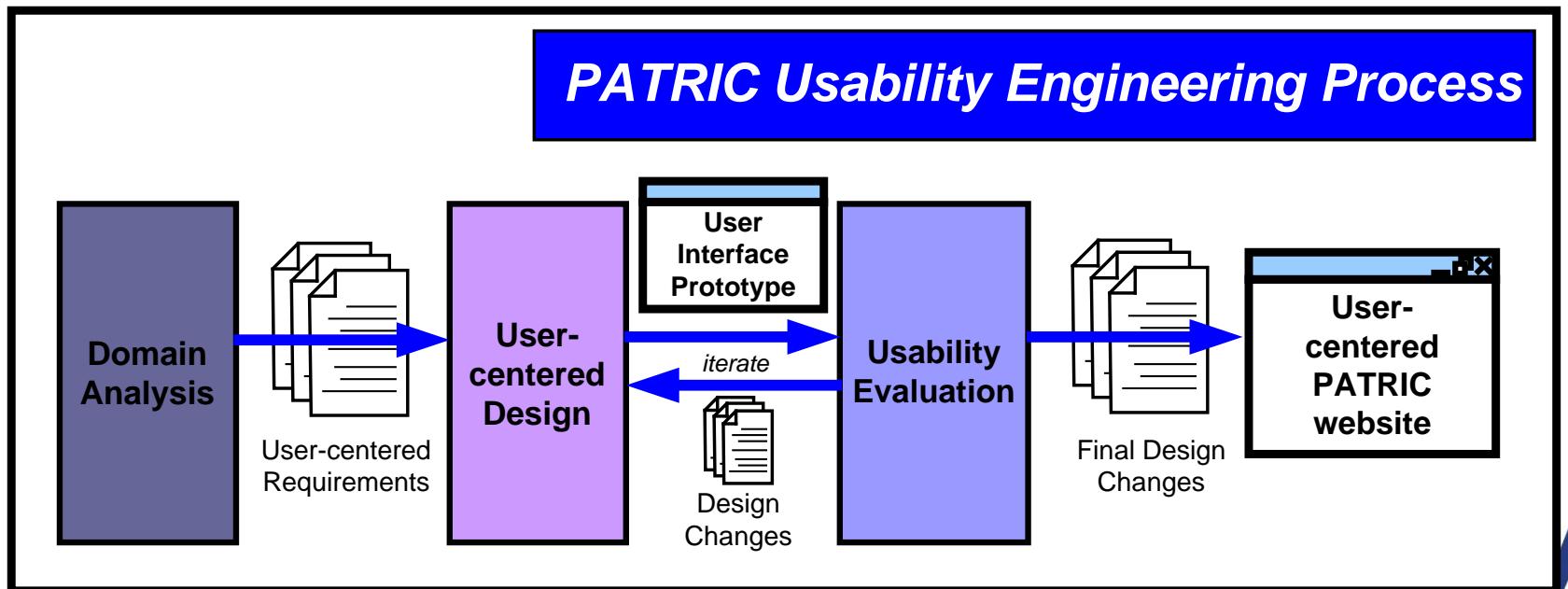
# Typical Bioinformatics Website & iPATRIC 2.0

## GETTING RID OF THE “NEEDLE IN THE HAYSTACK” PROBLEM



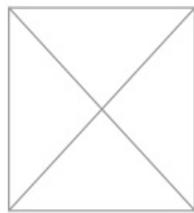
# PATRIC UI: Looking Forward

- Intuitive Web 2.0 interfaces
- Customizable and sharable workspaces & workflows
- Iteratively developed with users via Usability Engineering



# PATRIC eNews Wireframe

## eNews



### Feature News Item

This is a sticky item that will always appear at the top of the list. It will not be duplicated in the list below. The list below shows the most recent news items in reverse date order (newest items at the top).

Tags: [item\\_1](#), [item\\_2](#), [item\\_3](#)

---

29 June 2009

#### **[Updates to Bacterial Genome Data](#)**

4 Coxiella genomes annotated, 2 new Rickettsia genomes added...  
Tags: [coxiella](#), [rickettsia](#)

---

29 June 2009

#### **[Updates to Viral Genome Data](#)**

12 new genomes added...  
Tags: [coronavirus](#), [hepatitis E](#), [lyssavirus](#)

---

29 June 2009

#### **[PubMed Integration](#)**

Filter publication by scope option has been added....  
Tags: [publications](#)

---

16 June 2009

#### **[Summit on Systems Biology](#)**

Richmond, VA: Dan Sullivan gave a presentation on "Data Integration for Dynamic and Sustainable Systems Biology Resources"....  
Tags: [meeting](#)

---

April 2009

#### **[Presentations in Germany, Italy, and the Netherlands](#)**

Dr. Bruno Sobral is giving three presentations in Europe....  
Tags: [presentation](#), [podcast](#)

---

3 April 2009

#### **[Publication on Horizontal Gene Transfer in Brucella Genomes](#)**

"Analysis of ten Brucella genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle"....  
Tags: [publication](#)

---

[Older News >](#)

### Sign-up for eNews

E-mail address

[Read Past eNews](#)

[Follow our RSS Feed](#)

### Browse

[Data Releases & Updates](#)

[Events](#)

[Meetings](#)

[Publications](#)

[Website Updates](#)

---

[Podcast](#)

[Video](#)



# PATRIC Home Page Wireframe

**PATRIC**  
PathoSystems Resource Integration Center

Search terms  Search [Contact Us](#)

PATRIC provides rich data and analysis tools for all bacterial species in the selected NIAID category A-C priority pathogens list. [About PATRIC >](#)

**Watchlist Genera**

Bacillus	Coxiella	Salmonella
Bartonella	Ehrlichia	Shigella
Borrelia	Escherichia	Staphylococcus
Brucella	Francisella	Streptococcus
Burkholderia	Helicobacter	Vibrio
Campylobacter	Listeria	Yersinia
Chlamydophila	Mycobacterium	
Clostridium	Rickettsia	

> All Bacteria  
> Bacteria Pathogens

**Listeria** Diseases: food-borne infection, Listeriosis  
Bacteria > Firmicutes > Bacilli > Bacillales > Listeriaceae > Listeria  
Features 70,784 (10 new)  
Genomes 29 (1 new)  
Literature 14,644 (4 new)

from Google Image Results (70,400)

**Tools**

- Blast
- Pathway Finder
- Protein Family Sorter

**Pathway Finder**

Bacteria Name   
Pathway Name Keyword   
Find Pathways  
More Options >

**Most Viewed Bacteria**

Bacillus Brucella Helicobacter Listeria  
Rickettsia Salmonella Staphylococcus  
Streptococcus Yersinia

**Call for Proposals**  
Driving Biological Projects

PATRIC is seeking other projects that focus on infectious diseases research related to human bacterial pathogens. [Send us a letter of intent >](#)

**eNews**

Upcoming Events >>  
  
April 20–22, 2010  
**International Symposium on Health Informatics and Bioinformatics (HIBIT '10)**  
Anatolia, Turkey  
Dr. Bruno Sobral to give keynote >

Upcoming Events >>  
  
April 25–27, 2010  
**NCBI Annotation Workshop**  
Rockville, MD  
Help develop PATRIC requirements and understand emerging standards >

More eNews >>

5 April 2010  
**RAST Annotations, Compare from 3 Sources**  
RAST annotations for nearly 2,000 bacteria now available in PATRIC ...

April 2009  
**Presentations in Germany, Italy, and the Netherlands**  
Dr. Bruno Sobral is giving three presentations in Europe....

3 April 2009  
**Publication on Horizontal Gene Transfer in Brucella Genomes**  
"Analysis of ten Brucella genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle"....

Collaborate with us ► Shape the PATRIC website ► Talk with other Biologists ►

**VIRGINIA**  
**FORMATICS**  
**INSTITUTE**  
IRGINIA TECH

# Integrating Post-Genomic Data

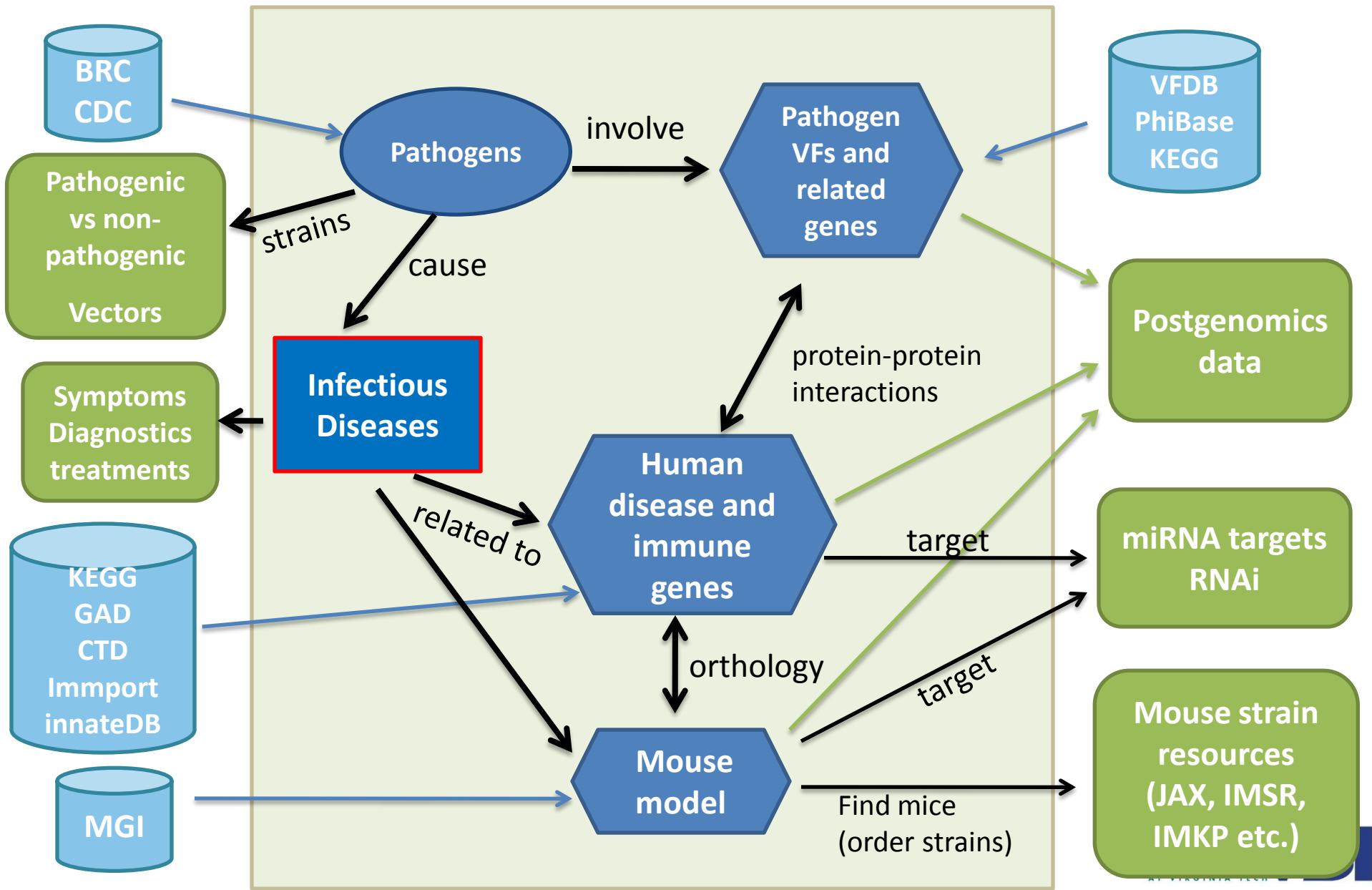
The figure displays a composite screenshot of bioinformatics tools and databases for integrating post-genomic data:

- PATRIC Dashboard:** Shows a list of recent publications and various search and analysis tabs.
- GEO Dataset Page:** Displays experimental details for a dataset (GSE1120) involving Brucella melitensis. A red arrow points from this page to the central PATRIC search interface.
- Central PATRIC Search Interface:** Shows a search results table for "Brucella". A red arrow points from the GEO dataset page to this table, indicating data transfer. Another red arrow points from the table to the bottom PubMed search results page.
- PubMed Search Results:** Shows search results for "Global Analysis of Quorum Sensing Targets in the Intracellular Pathogen Brucella melitensis 16 M". A red arrow points from the central PATRIC search interface to this page.

PATRIC/Pathogen Portal

# HOST-PATHOGEN INTERACTIONS

# PATRIC 2.0's Infectious Disease View



# Protein Interaction Gateway (PIG)

PATRIC

PathoSystems Resource Integration Center

Downloads

Contact Us

Home

Pathogen Interaction Gateway

## News

- **1 June 2008**  
PIG goes live.
- **15 August 2008**  
PIG visualization goes live.

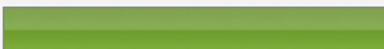
## Requirements

- **Site-wide**  
modern web browser  
internet connection  
javascript enabled
- **Visualizations**  
[java 1.5](#)

## About PIG



Protein-protein interactions (PPIs) play a vital role in initiating infection in a number of pathogens and identifying which interactions allow a pathogen to infect its host can provide potential targets for therapeutics. Public resources for studying host-pathogen systems, in particular PPIs, are scarce. To facilitate the study of host-pathogen PPIs we have collected and integrated said data from a number of public resources to create PIG, a database dedicated to the study of host-pathogen PPIs. PIG provides a number of user interfaces for searching available data and tools for predicting interactions between host and pathogen proteins.



- BLAST search
- Text search



- [Network Visualizer](#)
- [Network Comparer](#)



- Data Download
- Software Download
- If you would like to contribute to PIG, please contact us [here](#).

A

BLAST Query to P.I.G.

Query sequence (e.g., toll-like receptor):  
MMSASRLAGTLIPAMAFSLCVRPEWPCVEVPNITYQCMELNFYKIPONLPFSTKNLDL  
SFNPLRHLCYSFFSPELQLVLDLRCIQTIEDCAYQSLSHLSTLITGPNPIQSALGAFSGL  
SSLQKLVLAVETTASLENFPFIGHLKTKLELNVAHINLISFLKLFSNLNTLEHLDLSSNKQ  
SIYCTDRLVLHQMPLNLISLDSLNPMNFIQPCKARERHLHLLRNNFDSSLNVMTKTCIQG  
LAGLAVVTRKQVLSVYVYVQ  
FELVQVTTENRKVQKSYVNFQWQHLMUNCKFCQJPTLKLQSLKLTTTSSCCGNMFSEVNL  
SLEFLDLRNGLSKGCCSOOSFTTSKLQYDLSFNQVITMSNFLGLQLEHLDFOISNL  
KOMSEFSVFLSLRNLYDISHHTHTRVAFNGIFNGLSSLEVLMAGNSFOENFLPDLTTELRN  
LTFDLDSQCQCLQSLSPTAFLNLSLQVLNMSHNNIFSLDTTPYKCLNLSQLQFLDVSNLHIMT  
SKKQELQHPPSLAFNLNTQNDFACTCEHQSQFLQWIKDQRQLLIVERMECATPSDKQG  
MPVLSLNITCQMNKTIIQSVLSVLSVLSVAVLVYKFHILMLLAGCIVYGRGENYDAEVI  
YSSQWICFEYEIAQWTQFLSSRACIIIFVVKVETLRLQQVELYRLSRNTYLEWEDSVLGR  
HIFWRRLRKALLDGK

E-value cutoff: 1e-10

Word size: 3

Matrix: BLOSUM62

Submit

B

P.I.G. Protein View

**Entry Information**

Entry Name	p53_HUMAN
Accession Numbers	P04637   Q15086   Q15087   Q15088   Q16535   Q16807   Q16808   Q16809   Q16810   Q16811   Q16848   Q86UG1   Q80316   Q99659   Q9BTM4   Q9HAQ8   Q9NP68   Q9NPJ2   Q9NZD0   Q9UB12   Q9UO61
Entrez Gene	7157

**Protein Information**

Protein Description: Cellular tumor antigen p53 (Tumor suppressor p53) (Phosphoprotein p53) (Antigen NY-CO-13).

Gene Name: TP53

Source Organism: Homo sapiens [9606]

Sequence:

```
MEEPQSDPSV EPPLSQETFS DLWKLLPENN VLSPLPSQAM DDMLSPDDT EQWFTEDPGP DEAPRMRPEAA PPVAPAPAAP  
TPAFAFPAPPS WPLSSSVPSQ KTYQGSTYGF RGFPLHSGTAK SVTCYSPAL NKMFCQLAKT CPVQLWVDST EPPGPTVRAM  
AIYKQSQHM7 EVVRCPHHE RCSDSDGLAP POHLIRVEGN LRVEYLDDRN TFRHSVVVPY EPPVEVGSDCT TIHYNNMCS  
SCMGMMNRP ILTTITLEDs SGNLGRNSF EVRVCACPGR DRRTBEEENLR KKGEPEHHELP PGSTRALPN NTSSPQPKK  
KPLDGEYFTT QIRGRERPEM FRENEALEED KDAQACKEGP GSAHSSHLP SKKGQSTSRR KKLMFKTEGP DSD
```

**Domain Information**

**Annotation Information**

**Inter-species Interactions**

Interactor	Description	Organism	Method	Reference
09X365_BACAN	PX01-95 (UDP-glucose 6-dehydrogenase) (EC 1.1.1.22).	Bacillus anthracis	YEAST TWO-HYBRID	<a href="#">Biodefense Proteomics Resource Center</a>
081SN0_BACAN	Conserved repeat domain protein.	Bacillus anthracis	YEAST TWO-HYBRID	<a href="#">Biodefense Proteomics Resource Center</a>
GATB_FRATT	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B (EC 6.3.5.-) (Asp/Glu-ADT subunit B).	Francisella tularensis subsp. tularensis	YEAST TWO-HYBRID	<a href="#">Biodefense Proteomics Resource Center</a>
LT_SV40	Large T antigen (LT) (LT-AG).	Simian virus 40	COIMMUNOPRECIPITATION	<a href="#">Interaction of the human papillomavirus type 16 EG oncprotein with wild-type and mutant human p53 proteins. [J] Virology 1992</a>

C

Uniprot      Entrez Gene      NCBI Taxonomy      Interpro      Gene Ontology      Pubmed

**A**

### PIG Host-Pathogen Network

1. Choose a host
  - ▼ Human
    - Homo sapiens (4340)
2. Choose a pathogen
  - ▼ Adenovirus (82)
    - Adeno-associated virus - 2 (4)
    - Human adenovirus 1 (1)
    - Human adenovirus 2 (23)**
    - Human adenovirus 5 (37)
    - Human adenovirus E (1)
    - Human adenovirus F (1)
    - Human adenovirus 12 (13)
    - Human adenovirus 31 (1)
    - Human adenovirus 9 (1)
  - Aeromonas (2)
  - Anemia virus (4)
3. Restrict by GO-Slim terms (optional)
  - nucleic acid binding (12)
  - catalytic activity (8)
  - helicase activity (1)
  - signal transducer activity (2)
  - binding (23)**
  - protein binding (23)
  - intracellular (22)
  - cell (22)
  - nucleus (21)

**B**

### Pathogen Interaction Gateway

**Results**

- Organisms**
  - Host: Homo sapiens
  - Path: Human adenovirus 2
  - G.O.: binding
- Network**
  - 23 direct interactions
  - 22 host proteins
  - 7 pathogen proteins
- Downloads**
  - [GraphML file](#)
- Tools**
  - [Build a new network](#)
  - [Compare two networks](#)

**ID:** ATF1\_HUMAN  
**Organism:** Homo sapiens  
**GO Annotations:**

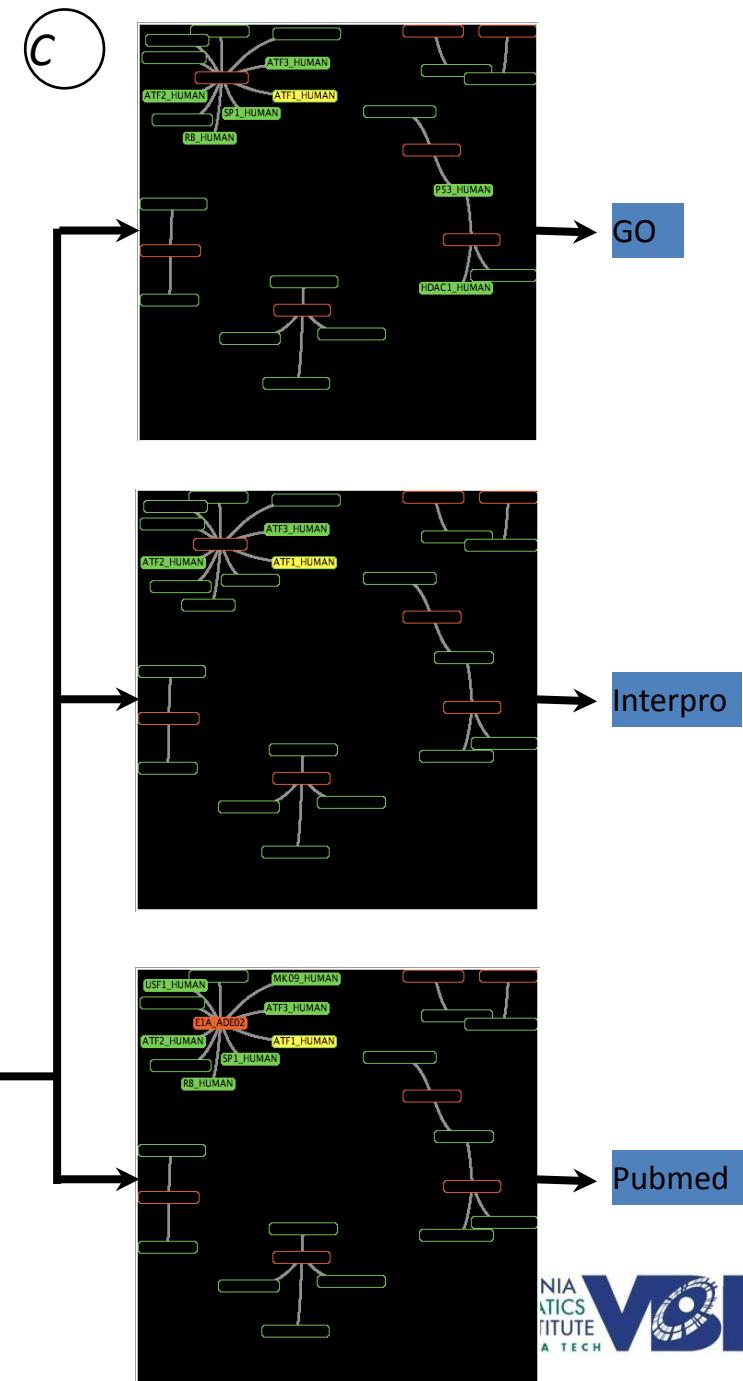
- GO:0005634
- GO:0005515
- GO:0006355
- GO:0006350
- GO:0005667
- GO:0043565
- GO:0003700
- GO:0004082

**Domains:**

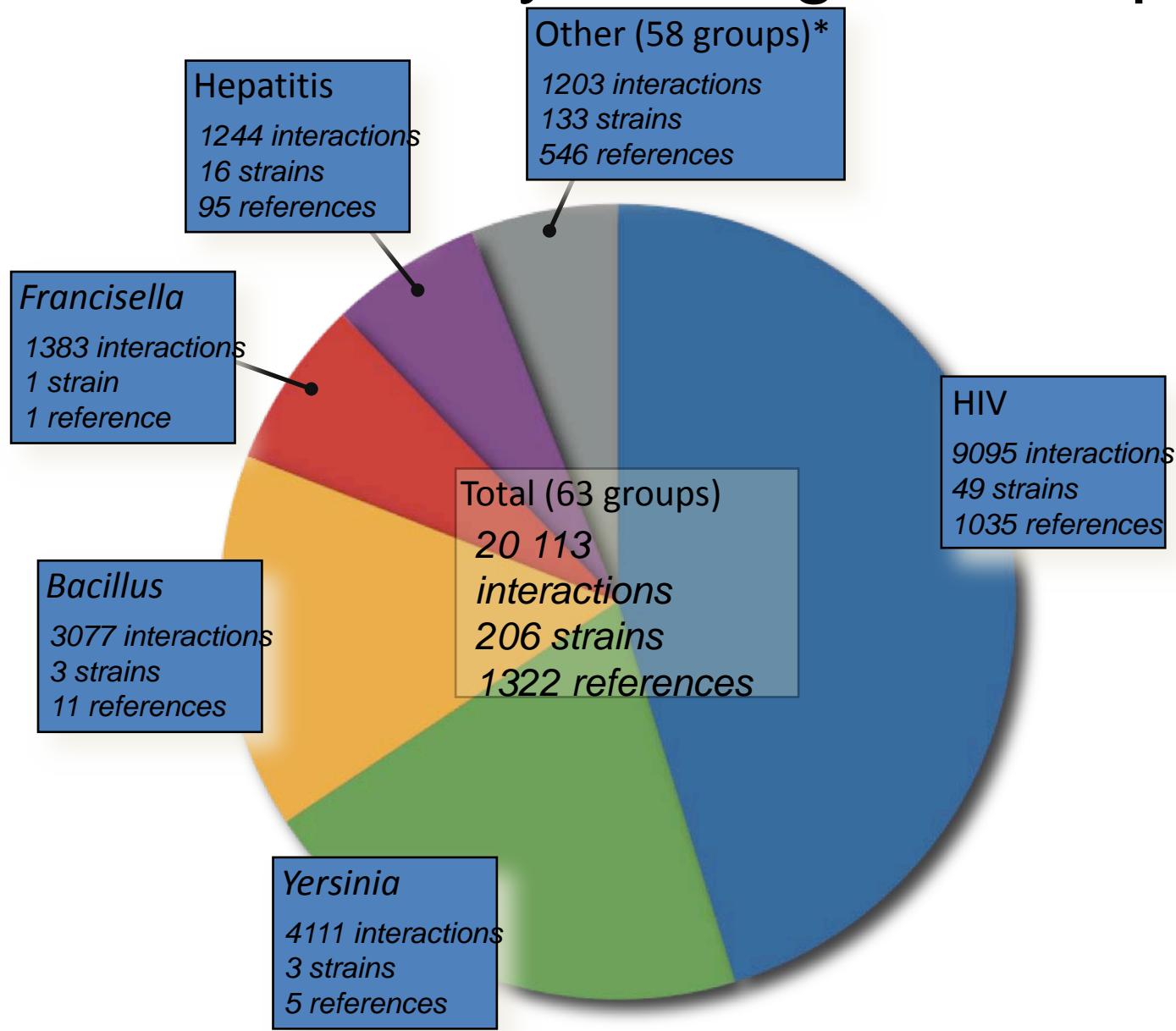
- IPR003102
- IPR004827
- IPR001630
- IPR011616

**Interactions:**

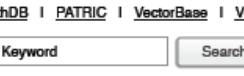
- E1A\_ADE02: ALANINE SCANNING
- E1A\_ADE02: AUTORADIOGRAPHY
- E1A\_ADE02: GEL RETARDATION ASSAY
- E1A\_ADE02: TANDEM AFFINITY PURIFICATION



# HP-PIPs in PIG by Pathogen Group, 2009



# Pathogen Portal UI, Looking Ahead: NIAID DMID Program Dashboard

BRC Websites: [EuPathDB](#) | [PATRIC](#) | [VectorBase](#) | [ViPR](#)

Keyword  Search

**Dashboard Filters**  
Click to check which BRCs to display.

- All BRCs
- EuPathDB
- PATRIC
- VectorBase
- ViPR
- PathogenPortal

**What is Pathogen Portal?**  
Pathogen Portal is a repository linking to four Bioinformatics Resource Centers (BRCs) sponsored by the National Institute of Allergy and Infectious Diseases (NIAID). The BRCs are providing web-based resources to scientific community conducting basic and applied research on organisms considered potential agents of biowarfare or bioterrorism or causing emerging or re-emerging diseases.

[Learn More](#)

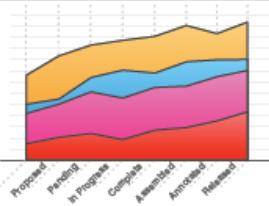
**Home** **Program Overview** **Genome Summary** **Advanced Search** **News & Events** **Data Set Summary** **Publications** **Tools** **Website Trends**

**Program Overview** Last Updated: 03/01/2010

**Did you know you can...?**

- Do X
- See Y
- Find Z

**Genome Summary** Last Updated: 03/01/2010

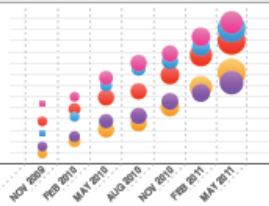


**Advanced Search**

**News and Events** Last Updated: 03/01/2010

- December 28, 2009 New expression data: [Insecticide resistant \*Vauclin\* strain](#)
- December 28, 2009 [2010 EuPathDB Workshop](#)
- December 17, 2009 [Virginia Bioinformatics Institute Press Release](#)
- December 17, 2009 [Virginia Bioinformatics Institute Press Release](#)

**Data Set Summary** Last Updated: 03/01/2010



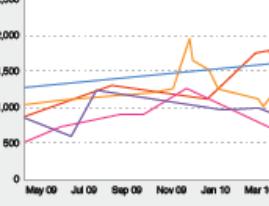
**Publications** Last Updated: 03/01/2010

- June 9, 2009 [Microarray identification of \*Clostridium difficile\* core components and divergent regions associated with host origin](#)
- June 4, 2009 [Viral bioinformatics: computational views of host and pathogen](#)
- April 30, 2009 [FluGenome: a web tool for genotyping influenza A virus](#)

**Tools** Last Updated: 03/01/2010

Cont. Tool tip summary for Tools

**Website Trends** Last Updated: 03/01/2010



This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900040C

[Portal Home](#) | [EuPathDB](#) | [PATRIC](#) | [VectorBase](#) | [ViPR](#)

**Virginia Bioinformatics Institute at Virginia Tech** 

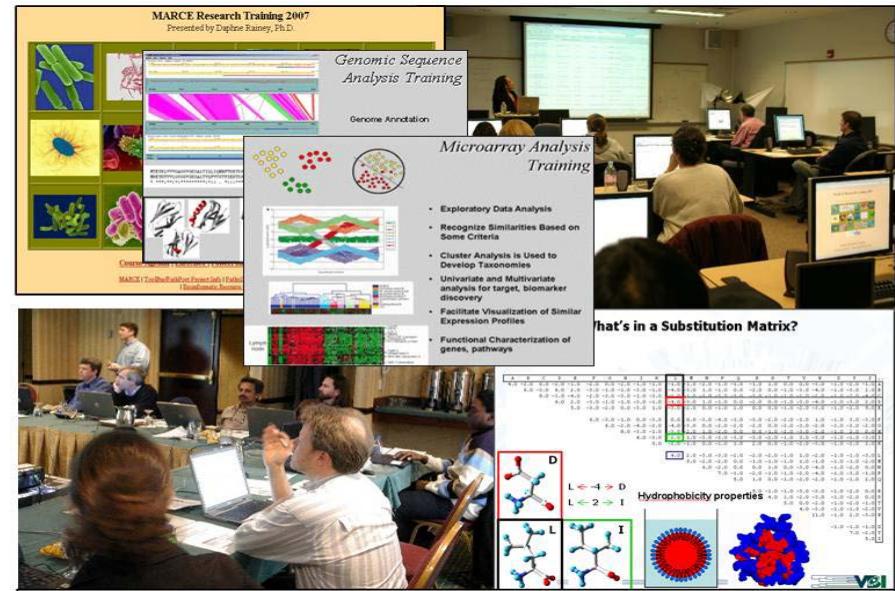
DBPs

Workshops

# OUTREACH & SCIENTIFIC COLLABORATIONS

# Workshops

- Features
  - Project-centric
  - Resource-centric
  - Feedback collection
  - Collaboration
- 1<sup>st</sup> workshop
  - Topic: Comparative Genomics
  - Late Summer, 2010, Washington DC area
- Conference Workshop
  - Uruguay Latin-American Conf. on Microbiology
  - September 2010



# Turning Research into Website Analysis Tools 2

JOURNAL OF BACTERIOLOGY, June 2009, p. 3569–3579

0021-9193/09/\$08.00+0 doi:10.1128/JB.01767-08

Copyright © 2009, American Society for Microbiology. All Rights Reserved.

Vol. 191, No. 11

## Analysis of Ten *Brucella* Genomes Reveals Evidence for Horizontal Gene Transfer Despite a Preferred Intracellular Lifestyle<sup>V§</sup>

Alice R. Wattam,<sup>1\*</sup> Kelly P. Williams,<sup>1</sup> Eric E. Snyder,<sup>1</sup> Nalvo F. Almeida, Jr.,<sup>1,6</sup> Maulik Shukla,<sup>1</sup> A. W. Dickerman,<sup>1</sup> O. R. Crasta,<sup>1</sup> R. Kenyon,<sup>1</sup> J. Lu,<sup>1</sup> J. M. Shallom,<sup>1</sup> H. Yoo,<sup>1</sup> T. A. Ficht,<sup>2</sup> R. M. Tsolis,<sup>3</sup> C. Munk,<sup>4</sup> R. Tapia,<sup>4</sup> C. S. Han,<sup>4</sup> J. C. Detter,<sup>4</sup> D. Bruce,<sup>4</sup> T. S. Brettin,<sup>4</sup> Bruno W. Sobral,<sup>1</sup> Stephen M. Boyle,<sup>5</sup> and João C. Setubal<sup>1,7</sup>

*Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia 24061<sup>1</sup>; Department of Veterinary Pathobiology, Texas A&M University and Texas Agricultural Experiment Station, College Station, Texas 77843-4467<sup>2</sup>; Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, One Shields Ave., Davis, California 95616-8645<sup>3</sup>; Biosciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544<sup>4</sup>; Center for Molecular Medicine and Infectious Diseases, Virginia-Maryland Regional College of Veterinary Medicine, Virginia Tech, Blacksburg, Virginia 24061<sup>5</sup>; Departamento de Computacao e Estatistica/Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil<sup>6</sup>; and Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061<sup>7</sup>*

Received 17 December 2008/Accepted 12 March 2009

The facultative intracellular bacterial pathogen *Brucella* infects a wide range of warm-blooded land and marine vertebrates and causes brucellosis. Currently, there are nine recognized *Brucella* species based on host preferences and phenotypic differences. The availability of 10 different genomes consisting of two chromosomes and representing six of the species allowed for a detailed comparison among themselves and relatives in the order *Rhizobiales*. Phylogenomic analysis of ortholog families shows limited divergence but distinct radiations, producing four clades as follows: *Brucella abortus*-*Brucella melitensis*, *Brucella suis*-*Brucella canis*, *Brucella ovis*, and *Brucella ceti*. In addition, *Brucella* phylogeny does not appear to reflect the phylogeny of *Brucella* species' preferred hosts. About 4.6% of protein-coding genes seem to be pseudogenes, which is a relatively large fraction. Only *B. suis* 1330 appears to have an intact β-ketoadipate pathway, responsible for utilization of plant-derived compounds. In contrast, this pathway in the other species is highly pseudogenized and consistent with the "domino theory" of gene death. There are distinct shared anomalous regions (SARs) found in both chromosomes as the result of horizontal gene transfer unique to *Brucella* and not shared with its closest relative *Ochrobactrum*, a soil bacterium, suggesting their acquisition occurred in spite of a predominantly intracellular lifestyle. In particular, SAR 2-5 appears to have been acquired by *Brucella* after it became intracellular. The SARs contain many genes, including those involved in O-polysaccharide synthesis and type IV secretion, which if mutated or absent significantly affect the ability of *Brucella* to survive intracellularly in the infected host.

# Finding Genome or Clade-Specific Protein Families

PATRIC  
PathSystems Resource Integration Center

Home   Organisms   Searches & Tools   Downloads   About PATRIC   Contact Us

No Items, No Groups

**Protein Family Sorter**

174 families found

Use the filters on the left to show protein families that include or exclude members from specified genomes. Enter a keyword to filter the list of protein families based on product description. Sort protein families, or show/hide specific columns by hovering over a column header and selecting the down icon (▼). To view or download DNA and/or protein sequence data of selected protein families in FASTA format, use the feature table controls just above the table column headers. Selected protein families (i.e., collection of features comprising the family) can also be added to custom PATRIC groups to support site-wide aggregation and bulk download via the group icon (▶) located in the blue banner of every PATRIC page. Use the Filter controls located at the top of the table, to filter feature table data based on sequence status, feature type, annotation or keyword. Click on a Product Description to see a detailed list of feature members of a specific protein family, and to access a multiple sequence alignment tool for the specific protein family.

**Select Different Organisms**

Include  
Exclude  
Don't Care  
Set all

Brucella abortus NCTC 8038  
Brucella abortus S19  
Brucella abortus bv. 1 str. 9-941  
Brucella abortus bv. 2 str. 86/8/59  
Brucella abortus bv. 3 str. Tulya  
Brucella abortus bv. 4 str. 292  
Brucella abortus bv. 6 str. 870  
Brucella abortus bv. 9 str. C68  
Brucella canis ATCC 23365  
Brucella ceti B1/94  
Brucella ceti M13/05/1  
Brucella ceti

AND Product Descriptions that  
contain  
do not contain

Add to Group   Retrieve   Selected   FASTA DNA Sequence(s)   Show or Download   Download as Excel File   Download as Text File

ID	Members	Species	Product Description	Min AA Length	Max AA length
FIG001105	1	1	tRNA (cytosine34-2'-O)-methyltransferase (EC 2.1.1.-)	171	171
FIG001649	3	3	N-methylhydantoinase A (EC 3.5.2.14)	693	693
FIG001912	1	1	Transcriptional repressor of PutA and PutP / Proline dehydrogenase (EC 1.5.99.8) (Proline oxidase) / Delta-1-pyrroline-,	1227	1227
FIG001949	1	1	Ribosomal large subunit pseudouridine synthase B (EC 5.4.99.-)	565	565
FIG002013	1	1	N-methylhydantoinase B (EC 3.5.2.14)	587	587
FIG002546	12	12	Conjugative transfer protein TrbI	238	263
FIG002771	1	1	Conjugative transfer protein TrbL	547	547
FIG003446	1	1	Aldehyde dehydrogenase A (EC 1.2.1.22)	452	452
FIG004146	3	3	Transcriptional regulator, LysR family	304	304
FIG005305	1	1	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases	414	414
FIG005577	1	1	Glutamate decarboxylase (EC 4.1.1.15)	0	0
FIG006278	4	4	Glutamate-ammonia-lyase/adenylyltransferase (EC 2.7.7.42)	983	983
FIG006552	4	4	Gene Transfer Agent (GTA) ORF611	181	181
FIG006936	1	1	Dihydrolipoamide acetyltransferase component (E2) of acetoin dehydrogenase complex (EC 2.3.1.-)	0	0
FIG007641	3	3	EnvY (acyl-carrier-protein) reductase [FMN] (EC 1.3.1.9)	309	324
FIG007935	1	1	Anticodon nuclelease	396	396
FIG008285	1	1	Ribose ABC transport system, periplasmic ribose-binding protein RbsB (TC 3.A.1.2.1)	315	315
FIG008562	6	6	Inositol transport system permease protein	375	387
FIG013931	1	1	Magnesium transporter	465	465
FIG015180	1	1	Dihydroneopterin aldolase (EC 4.1.2.25) / 2-amino-4-hydroxy-6-hydroxymethylhydropteridine pyrophosphokinase (E...	296	296

Page 1 of 9 | 20 Records Per Page | Displaying records 1 - 20 of 174

# Multiple Sequence Alignments for All Protein Families

**PATRIC**  
PathoSystems Resource Integration Center

Gblocks 0.91b for FIG002546:

Number of sequences: **12**

Alignment assumed to be: **Protein**

New number of positions: **238** (selected positions are underlined in blue)

	10	20	30	40	50	60	70	8
Brucella canis ATCC 23365	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella ceti B1/94	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella ceti M13/05/1	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella ceti M490/95/1	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella ceti M644/93/1	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella microti CCM 4915	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella neotomae 5K33	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella sp. F5/99	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella suis 1330	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella suis ATCC 23445	-----	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella suis bv. 3 str. 686	-----	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T
Brucella suis bv. 4 str. 40	MEMESQKTSAYAARKFGRDAALALA	MASAGICGVSLVMAPAPAVITVC	SMCSTFV	YVAFAYMT	ILMT	PA	nnn	T

**Parameters used**

Minimum Number Of Sequences For A Conserved Position: 10  
Minimum Number Of Sequences For A Flanking Position: 10  
Maximum Number Of Contiguous Nonconserved Positions: 8  
Minimum Length Of A Block: 10  
Allowed Gap Positions: With Half  
Use Similarity Matrices: Yes

**Flank positions of the 1 selected block(s)**

Flanks: [26 263]

New number of positions **238** (90% of the original 263 positions)

# PATRIC Will Be Presenting and Collaborating at Microbial Genomics

- Focus on microbial communities, human microbiome, pathogens and bioenergetics
- Conference includes a COMBREX annotation session and session on pathogens with PATRIC presentations



# In Summary, Lessons Learned: Collaborate with the Global Scientific Community!

- Build CyberInfrastructure, not just IT infrastructure, **and treat it like building a highway system**, not hypothesis-limited research
- **Integrate data and tools cohesively**, through single interfaces (not just linking) - don't make the user work hard to find and integrate data or tools
- Realize that **data will remain distributed** - use distributed systems technologies to build information systems (such as webservices)
- **Build scalable systems for tomorrow's data** (and it's growing faster and faster) and don't fund infrastructure projects that don't clearly show how to handle the explosive growth
- **Build usable interfaces** using usability engineering techniques, avoiding "inmates are running the asylum"
  - Use agile development processes to engage users and deploy often
  - Testing and getting an understanding of interfaces from diverse users will occupy about 80% of the total time (with actual coding being about 20%) - so, don't under budget either time or effort necessary!
- **Decide on use-cases and metadata standards early** by engaging the community and then implementing & provide everything in a free and open manner

# Contact Us!



Home    Organisms    Searches & Tools    Downloads    About PATRIC    Contact Us

No Items, No Groups

## Inside - About PATRIC

- [What is PATRIC?](#)
- [News & Announcement](#)
- [Data & Software Release](#)
- [Presentations](#)
- [Publications](#)
- [Related Sites](#)
- [Contact Us](#)

## Contact Us

### Mailing address

PATRIC  
Virginia Bioinformatics Institute  
Washington Street, MC 0477  
Blacksburg, VA 24061  
Office: (540) 231-2100 / Fax: (540) 231-2606  
Email: [patric@vbi.vt.edu](mailto:patric@vbi.vt.edu)

### Primary Contacts

Bruno Sobral, PhD  
PATRIC Principal Investigator  
(540) 231-2100  
[sobral@vt.edu](mailto:sobral@vt.edu)

Ronald Kenyon  
PATRIC Project Manager  
(540) 231-7470  
[rkenyon@vbi.vt.edu](mailto:rkenyon@vbi.vt.edu)

### Submit Feedback

To ensure the highest quality data and associated tools we encourage you to submit feedback. Feel free to provide suggestions for improvement as well as any errors that we may have missed.

[Submit Feedback](#)

[patricbrc.org](http://patricbrc.org)  
[pathogenportal.org](http://pathogenportal.org)  
[patric@vbi.vt.edu](mailto:patric@vbi.vt.edu)