

Dealing with Integration and Interoperation: Bioinformatics Resource Centers for Infectious Disease Research

Bruno Sobral
Professor and Director,
CyberInfrastructure Section, VBI

“How to Build A Bad Biological Database”

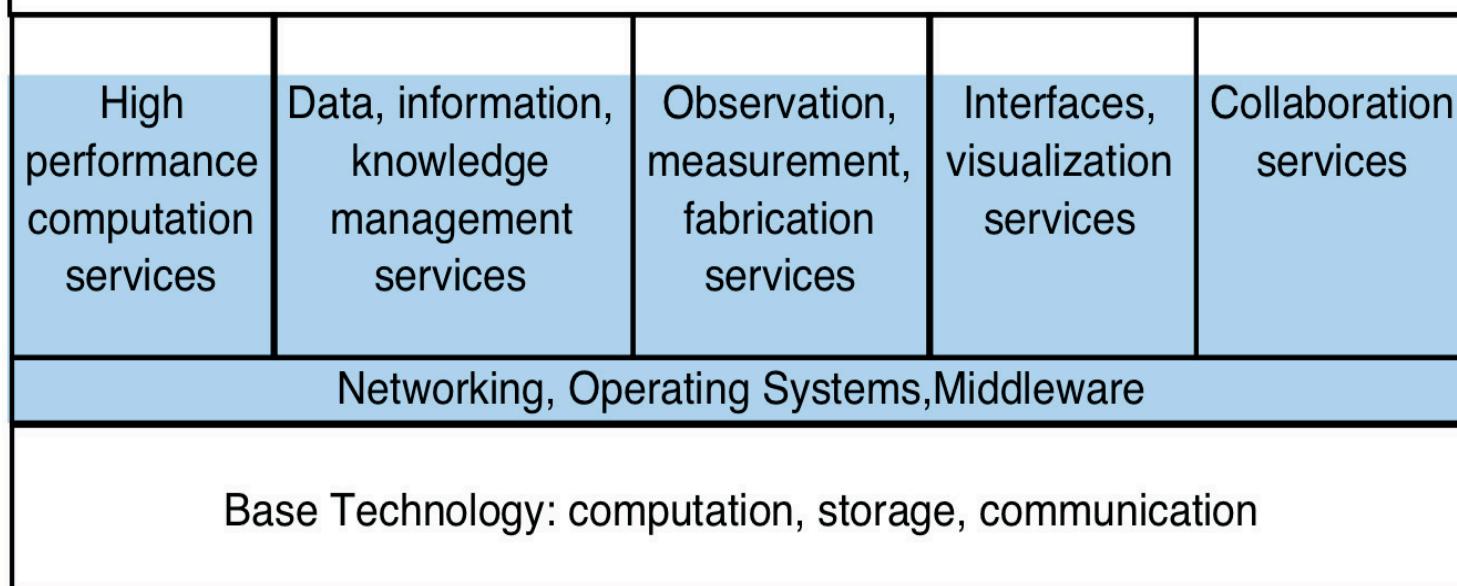
- Make submission difficult
- Have support services available 9-5p GMT
- Don’t let your file formats interconvert
- Keep your database independent
- Totally trust your automated systems
- Do not provide a permanent, unique identifier
- Make sure reviewers can’t see raw data
- Include a 44-page getting started guide
- If you include a search option make sure it only works in UK English
- Do not develop good visualization tools

<http://woodforthetrees.wordpress.com/2010/03/30/how-to-build-a-bad-biological-database/>

CyberInfrastructure

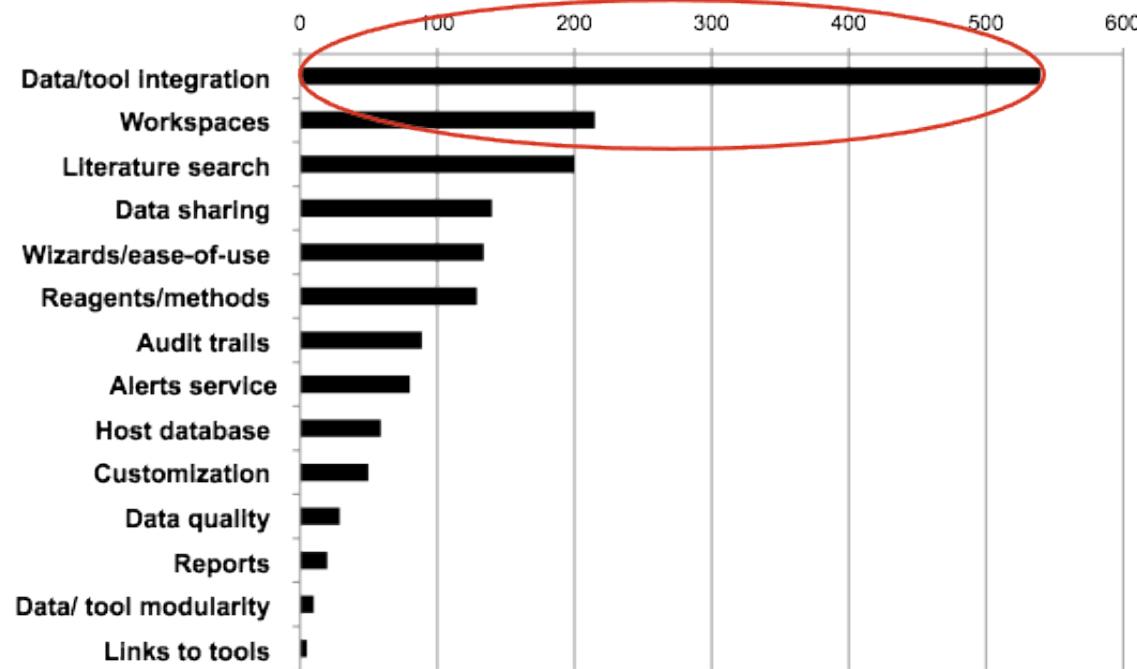
Community-Specific Knowledge Environments for Research and Education
(*collaboratory, co-laboratory, grid community, e-science community, virtual community*)

Customization for discipline- and project-specific applications



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

Users' Core Requirement: Data/ Tool Integration



BASED ON IN-DEPTH USER SURVEYS WITH "WET" ID RESEARCHERS/COUNTERMEASURE DEVELOPERS

Scalability

Must design for FUTURE, not current SITUATION

Original 11-2008 Estimates

	Genomes per year
End of PATRIC	5478
PATRIC 2, Year 1	10,368
PATRIC 2, Year 3	20,736

New Estimates

NOT SURE... COULD BE
ONE ORDER OF MAGNITUDE
OFF SINCE SEQUENCING
TECHNOLOGIES ARE BEING
INTRODUCED EVERY 18 MONTHS!

Automation is key – manual curation methods do not scale!

Key NIH - NIAID/DMID (Genomics) Programs/Data Sources

- Pathogen Functional Genomics Resource Center (JCVI)
- Genome Sequencing Centers (BROAD, IGS, JCVI)
- Structural Genomics Centers (CSGID, SSGCID)
- Clinical Proteomics Research Centers (Caprion, UTGalveston): includes pathogen and host (response) data!
- Systems Biology Centers (PNNL - enterics; ISB - influenza; Stanford & BROAD - TB; University of Washington - virology):
 - includes pathogen and host (response) data!
- Proteomics Resource Centers (defunct - data are with Pathogen Portal)
- Bioinformatics Resource Centers: Bacteria
 - FluDB is separate from BRC program

PATRIC and Portal
MUST coordinate
effectively with all
these consortia!

Emphasis on NIAID Priority Pathogens...

Category A

- *Bacillus anthracis* (anthrax)
- *Clostridium botulinum* toxin (botulism)
- *Yersinia pestis* (plague)
- *Variola major* (smallpox) and other related pox viruses
- *Francisella tularensis* (tularemia)
- Viral hemorrhagic fevers
- Arenaviruses
 - LCM, Junin virus, Machupo virus, Guanarito virus
 - Lassa Fever
- Bunyaviruses
 - Hantaviruses
 - Rift Valley Fever
- Flaviruses
 - Dengue
- Filoviruses
 - Ebola
 - Marburg

Category B

- *Burkholderia pseudomallei*
- *Coxiella burnetii* (Q fever)
- *Brucella* species (brucellosis)
- *Burkholderia mallei* (glanders)
- *Chlamydia psittaci* (Psittacosis)
- Ricin toxin (from *Ricinus communis*)
- Epsilon toxin of *Clostridium perfringens*
- *Staphylococcus enterotoxin B*
- Typhus fever (*Rickettsia prowazekii*)
- Food- and Waterborne Pathogens
 - Bacteria
 - Diarrheagenic *E.coli*
 - Pathogenic Vibrios
 - *Shigella* species
 - *Salmonella*
 - *Listeria monocytogenes*
 - *Campylobacter jejuni*
 - *Yersinia enterocolitica*

Category C

Emerging infectious disease threats such as Nipah virus and additional hantaviruses.

NIAID priority areas:

- Tickborne hemorrhagic fever viruses
 - Crimean-Congo Hemorrhagic fever virus
- Tickborne encephalitis viruses
- Yellow fever
- Multi-drug resistant TB
- Influenza
- Other Rickettsias
- Rabies
- Prions
- Chikungunya virus
- Severe acute respiratory syndrome associated coronavirus (SARS-CoV)
- Antimicrobial resistance, excluding research on sexually transmitted organisms*

Google™ PATRIC

Go

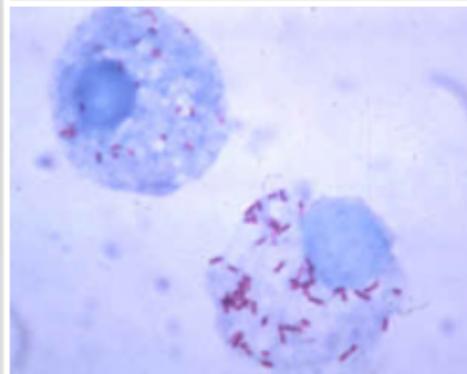


No Items

News

- **29 June 2009**
PATRIC Data Release and Website Update.
- **10 April 2009**
PATRIC Data Release and Website Update.
- **9 February 2009**
PATRIC Data Release and Website Update.
- **7 October 2008**
PATRIC Data Release and Website Update.
- **3 September 2008**
PATRIC Data Release.
- **11 June 2008**
PATRIC Data Release and Website Update.
- **18 April 2008**
PATRIC Data Release and Website Update.
- **31 January 2008**
PATRIC Data Release and Website Update.
- **More News**

The Organisms We Study



Bacteria

- *Brucella*
- *Coxiella*
- *Rickettsia*

Viruses

- *Caliciviridae*
- *Coronavirus*
- *Hepatitis A Virus*
- *Hepatitis E Virus*
- *Lyssavirus*

Assemble & Download Sequences of Interest

- Perform 9 specialized searches
- Use custom filters to drill down and narrow search results in feature tables
- Save and download results with feature cart
- Manipulate ortholog groups of related proteins for all PATRIC organisms

Perform Comparative Genomics

- Use tools to pinpoint common or unique traits among genes/proteins
- Compare amino acid sequences across proteins using Multiple Sequence Alignment

Perform Com... Pathway Anal...

- Compare PATRIC pathogen genome at pathway level

Role of BRC Program

- Provide publicly accessible database to
 - store, update, integrate and display genome sequence data, annotation and associated data for human pathogens
 - allow users to query and examine such information with user friendly interfaces and computational analyses tools.
 - serve as the public repository for NIAID-supported genomics programs
 - collaborate on experimental research projects via Driving Biological Projects

Human Microbiome Project (HMP)

- Key role in ID, co-funded by NIAID/DMID
- > 500 reference genomes (there will be 1000) already sequenced
 - Strains are all deposited in BEI as well - linking to BEI at various levels is crucial for people to get resources
- Demonstration projects are allowing HM to move into disease-specific areas
- BRC-DACC Coordination

ViPR – all viruses

PATRIC – all bacteria

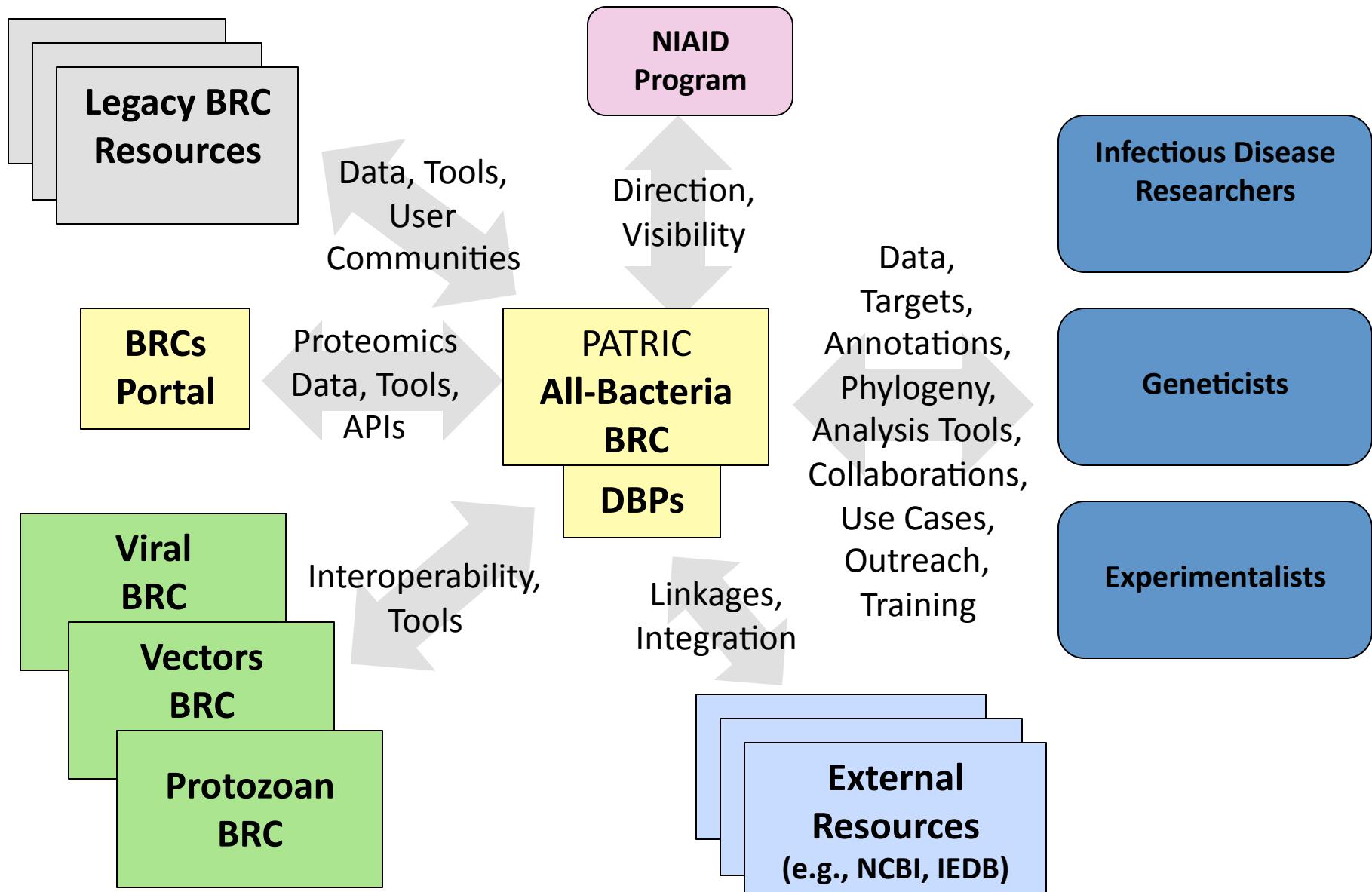
VectorBase – vectors

EuPathDB – Eukaryotic pathogens

Pathogen Portal – entry point into BRC Program, with a scientific focus on post-genomic data and host-pathogen interaction data

OVERVIEW OF BRC PROGRAM

New PATRIC 2.0 All-Bacterial BRC



Bacterial Data from Existing BRCs will Migrate to All-Bacterial BRC



PATRIC 2.0 Website (patricbrc.org)

March 2010

PATRIC Website

PATRIC

PathoSystems Resource Integration Center

Home Organisms Searches & Tools Downloads About PATRIC Contact Us

No Items, No Groups

Welcome to PATRIC

Currently, PATRIC is in a transition period - consolidating previous BRCs into new NIAID contracts. We anticipate adding new data, tools, and website features over the next few months.

For previous and current NIAID bacterial BRC users, we understand that these resources may be very valuable to your work. As such, we will be doing our very best to create a useful PATRIC resource to continue supporting your work. We realize that the transition time will cause disruptions, however, it is very important to us to work with established BRC communities to identify and prioritize our transition efforts. We will have a survey online soon to help us identify your needs, but in the meantime, you may contact us at patric@vbi.vt.edu

Watchlist Genera

PATRIC provides rich data and analysis tools for all bacterial species with an emphasis on the bacterial Orders that include NIAID [category A-C priority](#) and [emerging/re-emerging pathogens](#).

[Bacillus](#)

[Bartonella](#)

[Borrelia](#)

[Brucella](#)

[Burkholderia](#)

[Campylobacter](#)

[Chlamydophila](#)

[Clostridium](#)

[Coxiella](#)

[Ehrlichia](#)

[Escherichia](#)

[Francisella](#)

[Helicobacter](#)

[Listeria](#)

[Mycobacterium](#)

[Rickettsia](#)

[Salmonella](#)

[Shigella](#)

[Staphylococcus](#)

[Streptococcus](#)

[Vibrio](#)

[Yersinia](#)

Data Summary

	PATRIC	Legacy BRC	RefSeq
Number of genomes	1,860	409	2,371
Number of genomic features	13,531,784	2,297,834	14,978,762

New Website Features

Data and website enhancements include:

- Additional Genomes and Annotations
- Genome Browser
- Protein Family Sorter
- Metabolic Pathways
- Phylogenetic Trees
- Searches and Tools

[Please see details in the Release Notes](#)
(5 April 2010)

Announcement

- [PATRIC Data and Website Release \(5-Apr-2010\)](#)
- [PATRIC Data and Website Release \(30-Dec-2009\)](#)
- [PATRIC Data and Website Release \(9-Dec-2009\)](#)
- [Bioinformatics Resource Centers for Infectious Diseases](#)
- [VBI Awarded \\$27M from NIH to Support Infectious Disease Research](#)

[See List of All Bacteria](#)

[See List of Bacterial Pathogens](#)

Data Summary

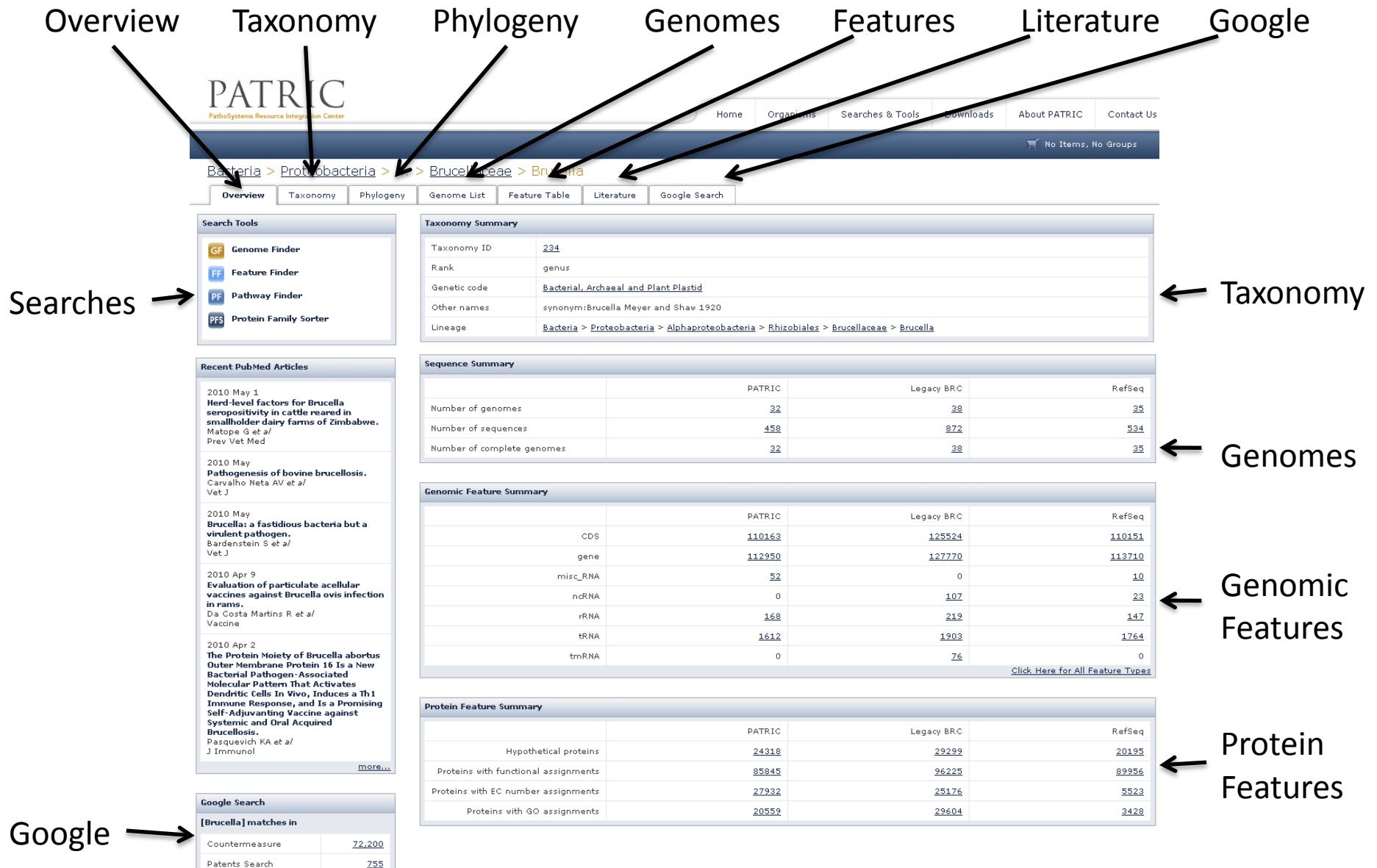
Sequence Summary			
	PATRIC	Legacy BRC	RefSeq
Number of genomes	1860	409	2336
Number of sequences	63830	23202	183274
Number of complete genomes	1860	409	2336

Genomic Feature Summary			
	PATRIC	Legacy BRC	RefSeq
CDS	6604931	1394369	6916265
gene	6746309	631644	7154003
mRNA	0	100500	9
misc_RNA	6542	0	4293
ncRNA	0	3607	3121
rRNA	14718	11470	18789
tRNA	106041	22104	108619
tmRNA	0	120	0

[Click Here for All Feature Types](#)

Protein Feature Summary			
	PATRIC	Legacy BRC	RefSeq
Hypothetical proteins	1824173	294958	1811019
Proteins with functional assignments	4780758	1099411	5105246
Proteins with EC number assignments	1411154	287067	681837
Proteins with GO assignments	1052002	724957	21587

Taxon Summary Page



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]

Evaluation of particulate acellular vaccines against *Brucella ovis* infection in rams.
Da Costa Martins R, Irache JM, Blasco JM, Muñoz MP, Marín CM, Jesús Grilló M, Jesús De Miguel M, Barberán M, Gamazo C
Vaccine (2010 Apr 9), PubMed: [19887131](#)
[Abstract]

The Protein Moiety of *Brucella abortus* Outer Membrane Protein 16 Is a New Bacterial Pathogen-Associated Molecular Pattern That Activates Dendritic Cells In Vivo, Induces a Th1 Immune Response, and Is a Promising Self-Adjuvanting Vaccine against Systemic and Oral Acquired Brucellosis.
Pasquevich KA, García Samartino C, Coria LM, Estein SM, Zwerdling A, Ibañez AE, Barrionuevo P, Oliveira FS, Carvalho NB, Borkowski J, Oliveira SC, Warzecha H, Giambartolomei GH, Cassataro J
Journal of immunology (Baltimore, Md. : 1950) (2010 Apr 2), PubMed: [20351187](#)
[Abstract]

Inhibition studies of a beta-carbonic anhydrase from *Brucella suis* with a series of water soluble glycosyl sulfonilamides.
Vullo D, Nishimori I, Scossaifava A, Köhler S, Winum JY, Supuran CT
Bioorganic & medicinal chemistry letters (2010 Apr 1), PubMed: [20211561](#)
[Abstract]

Brucellosis in dialysis patients. Does it exist?
Kantartzzi K, Panagoutsos S, Kokkinou V, Alepopoulou E, Mourvati E, Passadakis P, Vargemezis V
Clinical nephrology (2010 Apr), PubMed: [20353739](#)
[Abstract]

Evaluation of indirect enzyme-linked immunosorbent assays and IgG avidity assays using a protein A-peroxidase conjugate for serological distinction between *Brucella abortus* S19-vaccinated and -infected cows.
Pajuaba AC, Silva DA, Mineo JR
Clinical and vaccine immunology : CVI (2010 Apr), PubMed: [20147498](#)
[Abstract]

Google Search

Bacteria > Proteobacteria > ... > Brucellaceae > Brucella

Overview Taxonomy Phylogeny Genome List Feature Table Literature Google Search

Google Search

Powered by Google™

Brucella

Countermeasure (77,200) Patents (755) Web (140,000) News (26) Images (24,300) Videos (43) Books (5,080) Blogs (10,613)

RB51: A NEW BRUCELLOSIS VACCINE, UC Davis Veterinary Medicine ...
This new **Brucella abortus** **vaccine** is called RB51. This mutant strain of **Brucella abortus** does not produce cross-reacting antibodies in vaccinated cattle ...
www.vetmed.ucdavis.edu

Brucellosis: Treatments and drugs - MayoClinic.com
Clinical manifestations, diagnosis, and **treatment of brucellosis** in adults. ... **Vaccine** use following **brucellosis** and pseudorabies eradication. ...
www.mayoclinic.com

Brucellosis - Symptoms, Diagnosis, **Treatment of Brucellosis** - NY ...
Aug 28, 2009 ... Acute **brucellosis** may begin with mild flu-like symptoms or symptoms ... Also, call if your symptoms worsen or do not improve with **treatment**, ...
health.nytimes.com

Brucellosis vaccine - Wikipedia, the free encyclopedia
Brucellosis vaccine is a **vaccine** for cattle, sheep and goats used against ... Hidden categories: **Drug** pages needing a structure drawing | Chemical pages ...
en.wikipedia.org

CBRNE - **Brucellosis**: eMedicine Emergency Medicine
Apr 29, 2009 ... **Drug** Reference MEDLINE **Treatment** & Medication: CBRNE - **Brucellosis** ... Clinical features and **therapeutic** responses in 88 patients]. ...
emedicine.medscape.com

Recognition and optimum **treatment of brucellosis**.
Treatment of brucellosis must effectively control acute illness and prevent ... Rifampicin 900 mg once daily for 6 weeks is considered the **drug** of choice ...
www.ncbi.nlm.nih.gov

Brucellosis vaccine to be developed by researchers
Jan 26, 2010 ... **Brucellosis vaccine** to be developed by researchers. By Jeff Douglas, Virginia Tech Spectrum, October 20, 1994 ...
scholar.lib.vt.edu

Quinolones for **Treatment of Human Brucellosis**: Critical Review of ...
In fact, in this study, it was shown that rifampin is the only **drug** with increased The development by **Brucella melitensis** of resistance against **treatment** with Clinico-**therapeutic** features and molecular **diagnostic** approach. ...
aac.asm.org

1 2 3 4 5 6 7 8

Genome Browser

Drag and Drop Tracks

Zoom and Pan

Save Link

Compare Annotations

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

Brucella suis 1330

Overview Phylogeny Genome Browser Feature Table Pathways Literature Google Search

To add tracks, drag tracks (left panel) onto genome browser (right panel).

Link

NC_004310 38,638 .. 49,378 Go

40,279 42,514 44,752 46,990

VBIBruSui107850_0037 VBIBruSui107850_0041 VBIBruSui107850_0043 VBIBruSui107850_0047

VBIBruSui107850_0040 VBIBruSui107850_0042 VBIBruSui107850_0044 VBIBruSui107850_0046

BR0035 BR0037 BR0038 BR0039

BR0036 BR0040

BR0041 BR0042 BR0043

VBI0007BR1_0038 VBI0007BR1_0041 VBI0007BR1_0042 VBI0007BR1_0044

VBI0007BR1_0039 VBI0007BR1_0043 VBI0007BR1_0045 VBI0007BR1_0046

CDS(PATRIC) gene(RefSeq) gene(BRC) RNA(RefSeq) RNA(BRC) Misc(RefSeq) Misc(BRC) Misc(PATRIC)

Locus Tag: VBI0007BR1_0043
Location: 44948..45988
Strand: forward
Type: CDS
Gene symbol:
Product: molybdopterin biosynthesis enzyme

Click for detail information about this feature.

Metabolic Pathways

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

Brucella suis 1330

Overview Phylogeny Genome Browser Feature Table **Pathways** Literature Google Search

Pathway List
131 pathways found
Sort features and pathways or aggregation and bulk download. Filter controls located at the top.

Filter
Pathway Class: ALL
Add Features to Group Retrieve

Comparison Matrix

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

Pathway Map
Pathway ID : 00010
Pathway Name : Glycolysis / Gluconeogenesis
Pathway Class : Carbohydrate Metabolism

Pathways provide a visual representation of relationships between features. See the Pathway Description for an overview of this pathway. Use the Comparison Matrix to see PATRIC, Legacy BRC and RefSeq annotation details. For more help on this page, view our Pathway Viewer FAQ.

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, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [MD:M00679]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glyceraldehyde-3P to pyruvate form a conserved core module [MD:M00002], which is found in almost all organisms and which often corresponds to operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway of glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].

GLYCOLYSIS / GLUCONEOGENESIS

Save Map Print Map

Page 1 of 7 //patricbrc.org/portal/portal/patric/PathwayTable?cType=genome...

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

Add to Group | Retrieve

Genome Name
Brucella suis 1330

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, glyceraldehyde-3P, glyceraldehyde-3P, glyceraldehyde-3P, glyceraldehyde-3P, glyceraldehyde-3P.

Comparison Matrix Print Table

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

GLYCOLYSIS / GLUCONEOGENESIS

Starch and sucrose metabolism

Pentose phosphate pathway

D-Glucose (extracellular)

Save Map **Print**

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	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
rucella_abor	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
cella_abortu	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
ruceIla_abor	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Brucella_ab	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Brucella_ab	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLAL <u>MGLGILLSFAASPAVAERIGLN</u> SFH	V	FVEROIFFMVPA			
Brucella_ab	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Brüce	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Br	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Bru	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Bru	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Brucella_m	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
a_melitensis	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
ucella_melit	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
ella_meliten	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
ella_meliten	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Bruce	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Br	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			
Bruc	MVSRVD <u>RGPVANWWWWTIDRFFLAAC</u> CLALMGLGILLSFAASPAVAERIGLN	SFH	FVEROIFFMVPA			

Protein Structure 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 → Epitope Selection Box

Epitope → Epitope Information

3D Structure Visualization

JMOL Documentation

Appearance

- Wireframes: off
- Cartoons: thin
- Spacefill: off
- Surface: off
- Labels: off
- Show: everything

Reset

Navigation

- Spin: off
- Zoom: 100%

Reset

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



The PATRIC-RAST Server Integration: Free Public Resource for High Quality Genome Annotation

Rick Stevens et al.
University of Chicago
Argonne National Laboratory

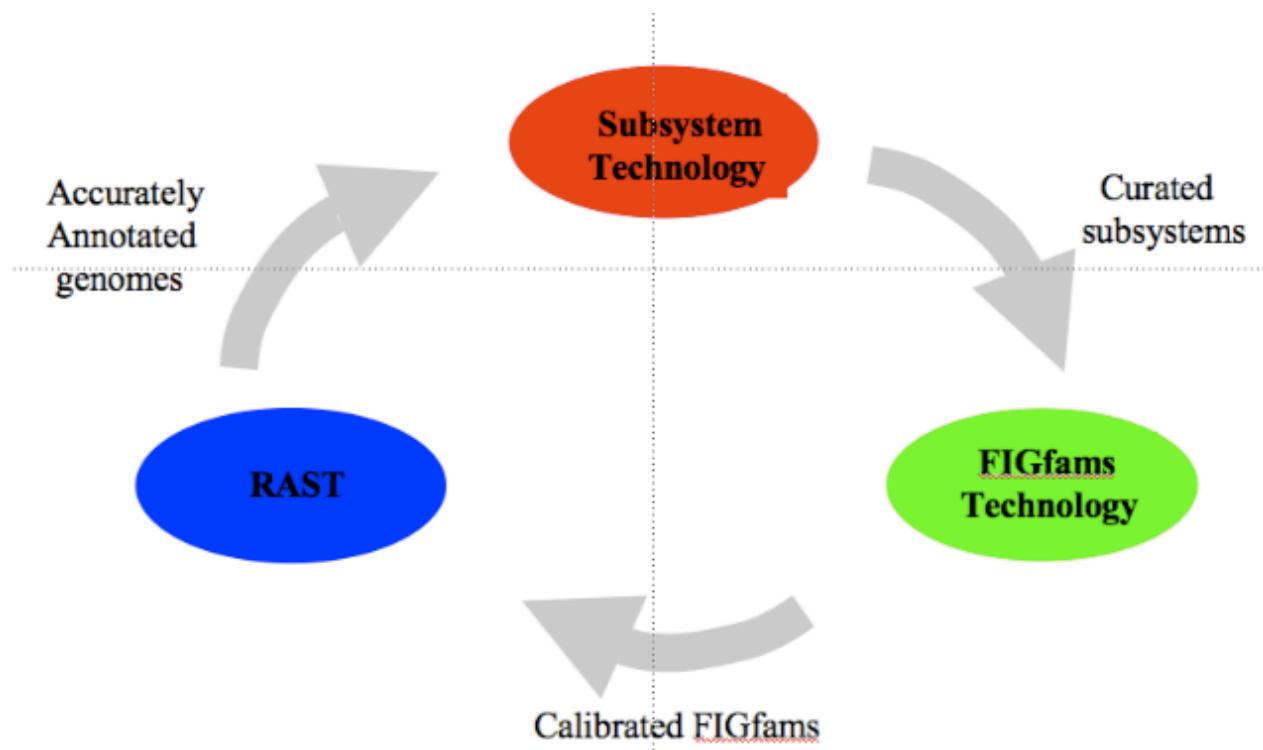
What is RAST?

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

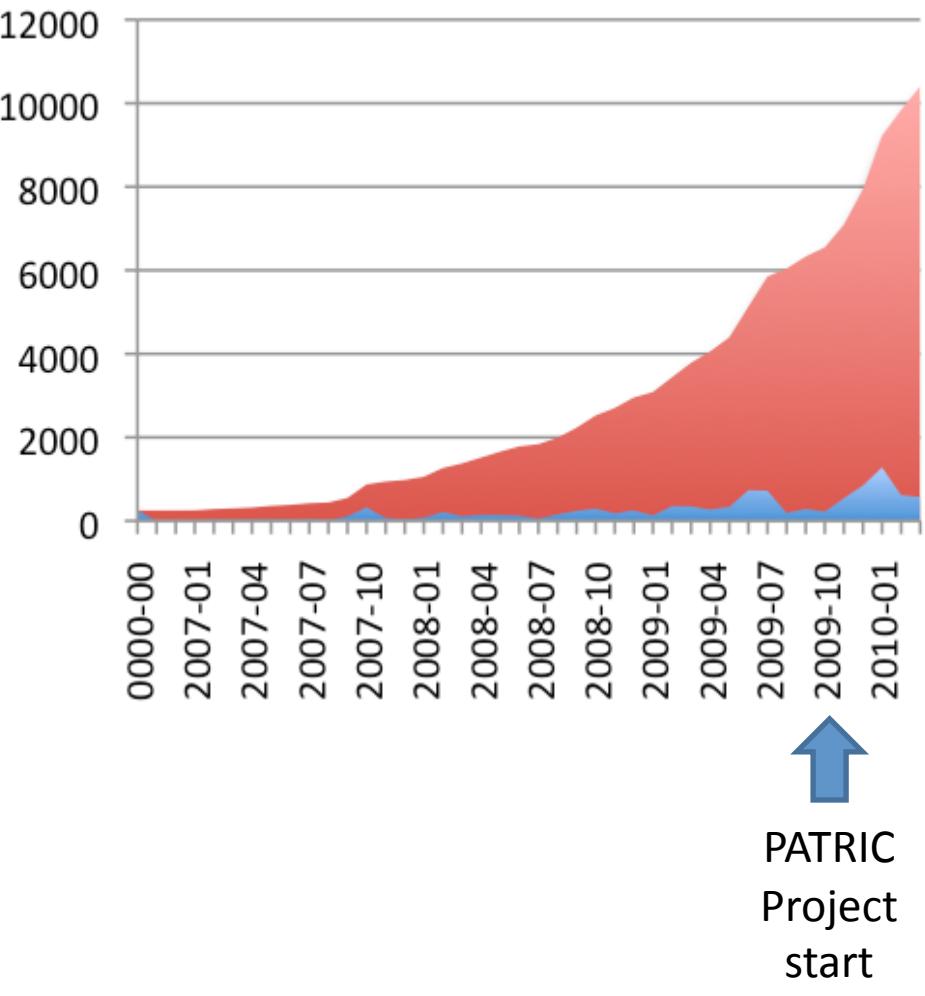
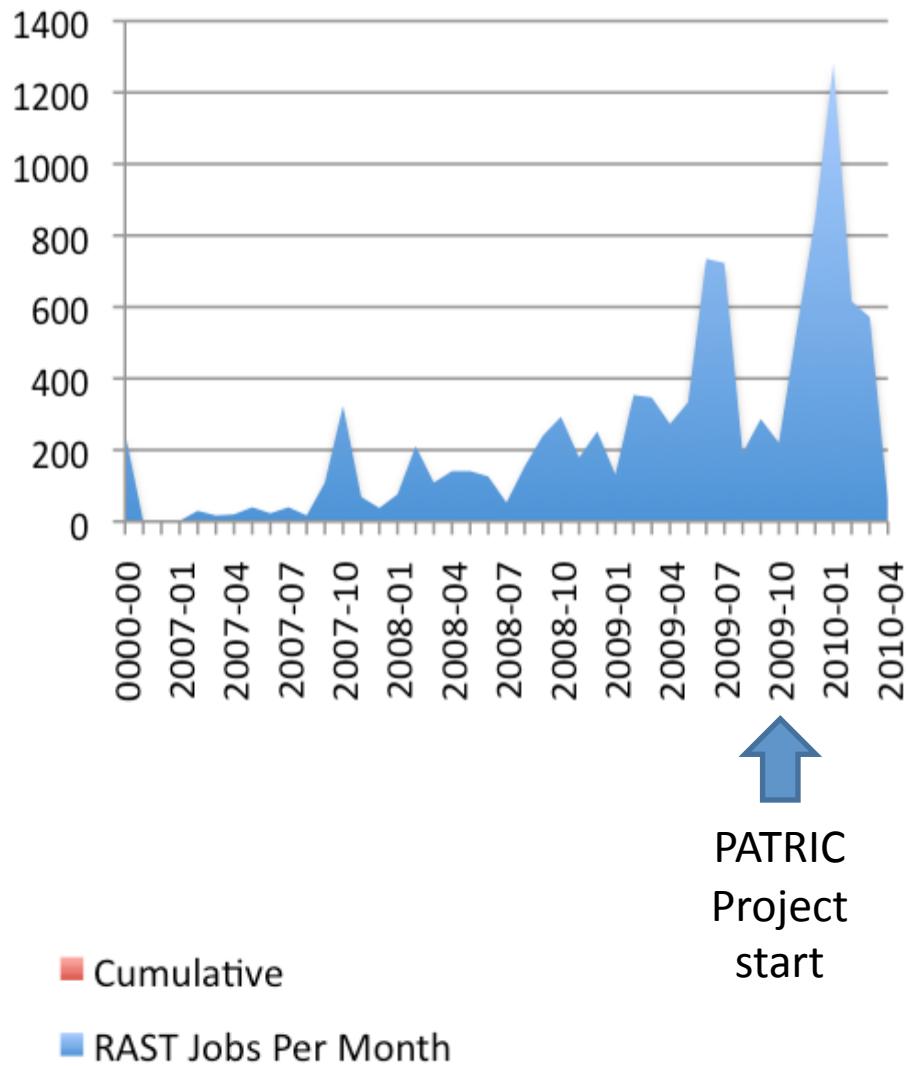


What Determines Quality?

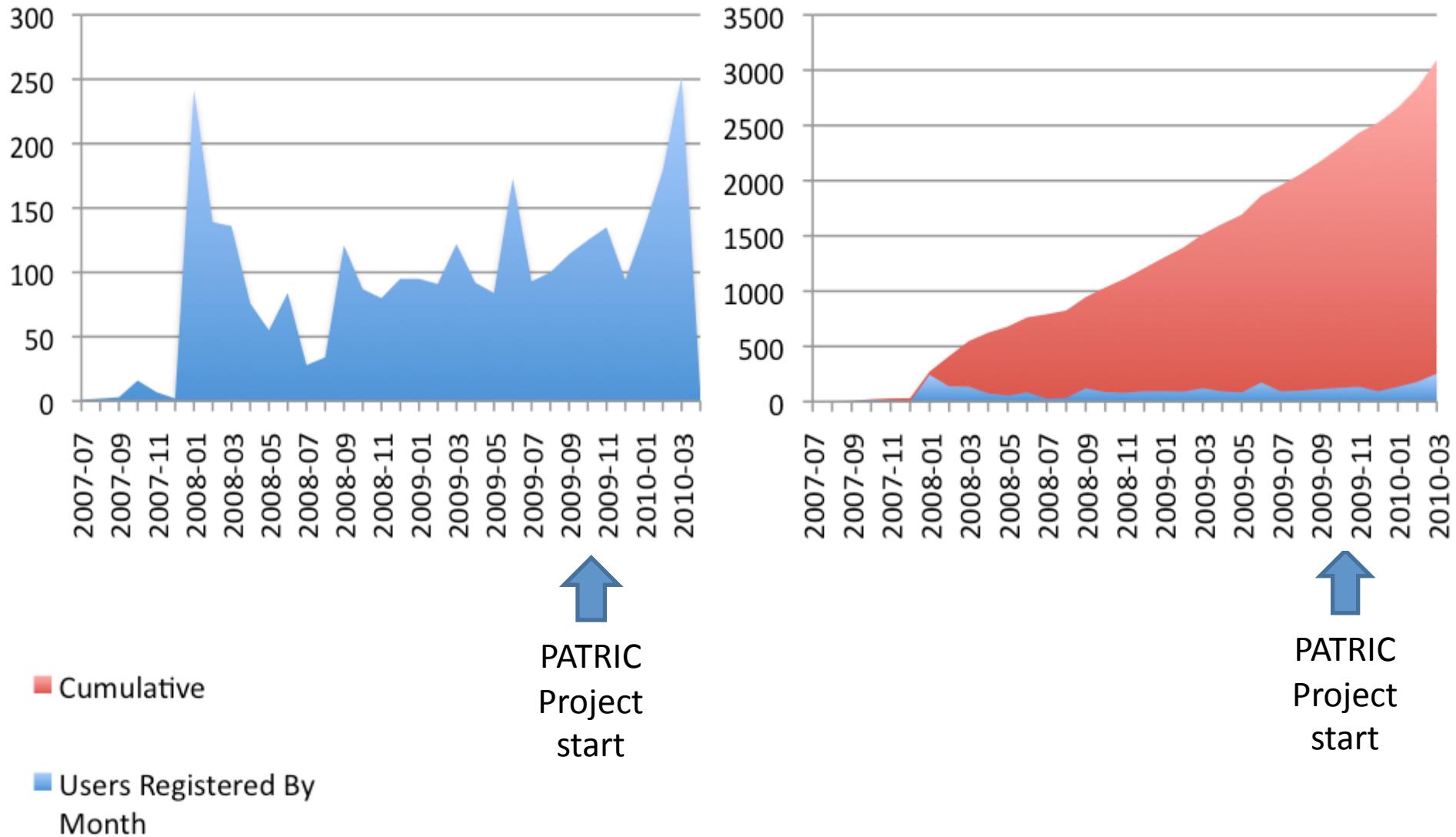
1. Accurately annotated core of **diverse genomes**
2. **Subsystems** that are manually curated across the entire collection of genomes, are also **computable objects**
3. **Protein families** that accurately characterize difficult cases (guided by the subsystems)



RAST Usage Statistics



RAST Registered Users



PATRIC-RAST Interface 1.0

PATRIC
PathSystems 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.

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, Formsma 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?](#)

Supporting Clinical (Hospital) Epidemiology to Reduce HAIs

Collaboration with Wake Forest
University Baptist Medical Center

Robert Scheretz

S. aureus, WFUBMC Sweeps

Figure 4. Persistent MRSA Genotypes

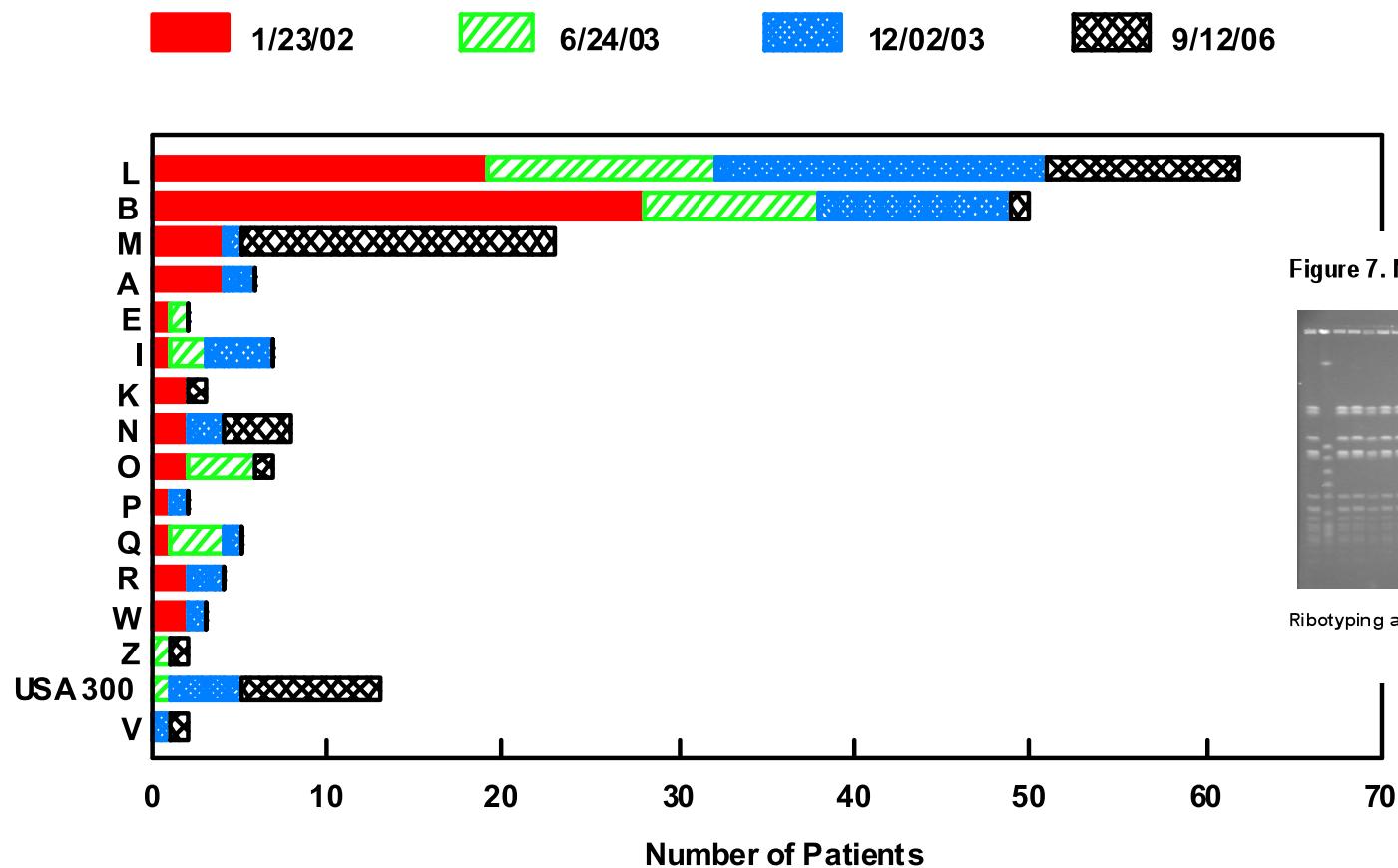
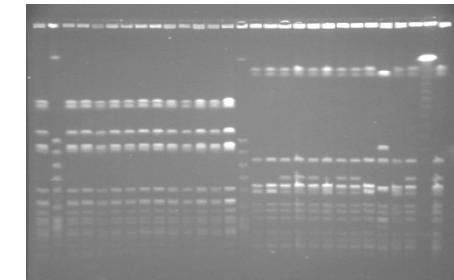


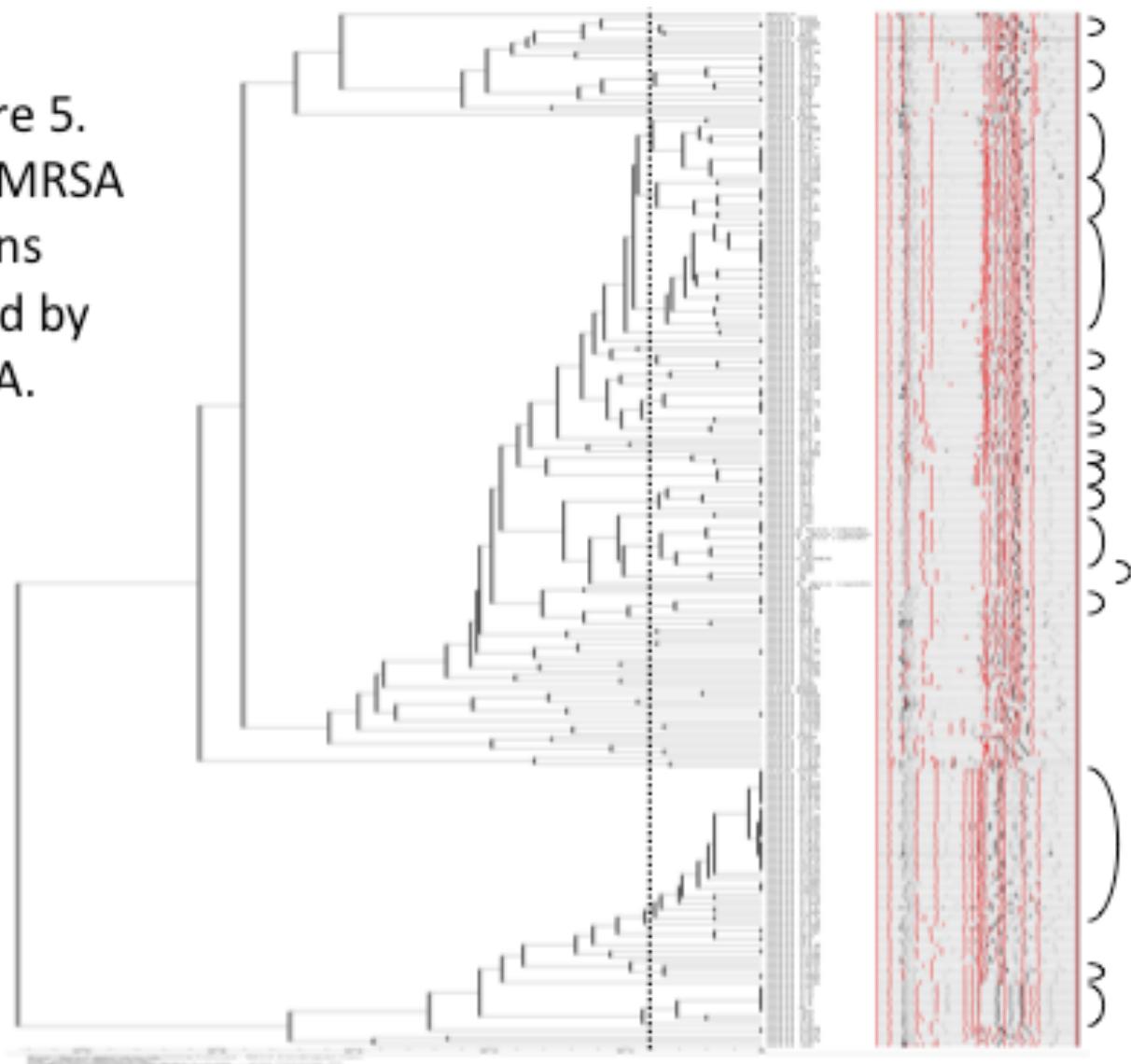
Figure 7. Nasal Strain vs Air Cultures



Ribotyping and PFGE: Percentage of similar strains
S. aureus: 85.7 – 99.3%

Increasing Resolution

Figure 5.
234 MRSA
strains
typed by
MLVA.

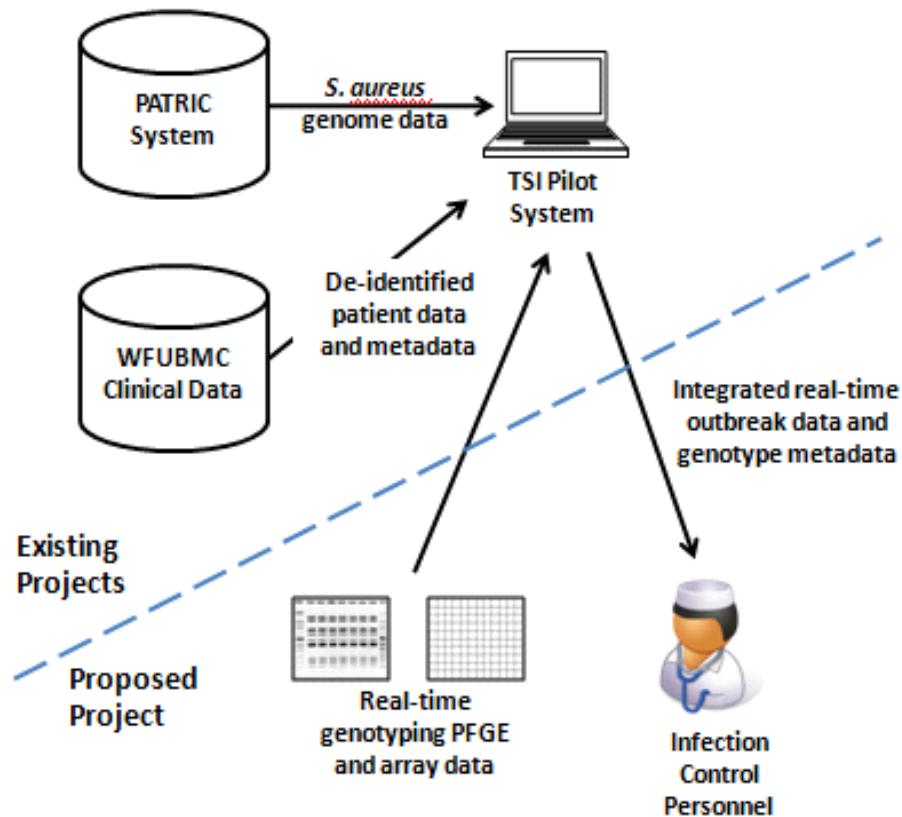


DNA Microarrays

Error rate	1.0%	2.0%	3.0%	4.0%	5.0%	6.0%	7.0%	8.0%	9.0%	10.0%
92457 Probes Affymetrix	0.053	0.149	0.219	0.270	0.309	0.339	0.364	0.386	0.404	0.420
487 Probes Clondiag	0.039	0.070	0.085	0.106	0.116	0.132	0.140	0.154	0.164	0.175
441 probes MTA-15	0.006	0.008	0.010	0.012	0.013	0.014	0.015	0.017	0.018	0.021
441 Probes MTA-60	1.2 E-61	3.1 E-34	1.5 E-20	1.2 E-13	1.0 E-9	1.3 E-7	3.6 E-6	3.9 E-5	1.6 E-4	0.001

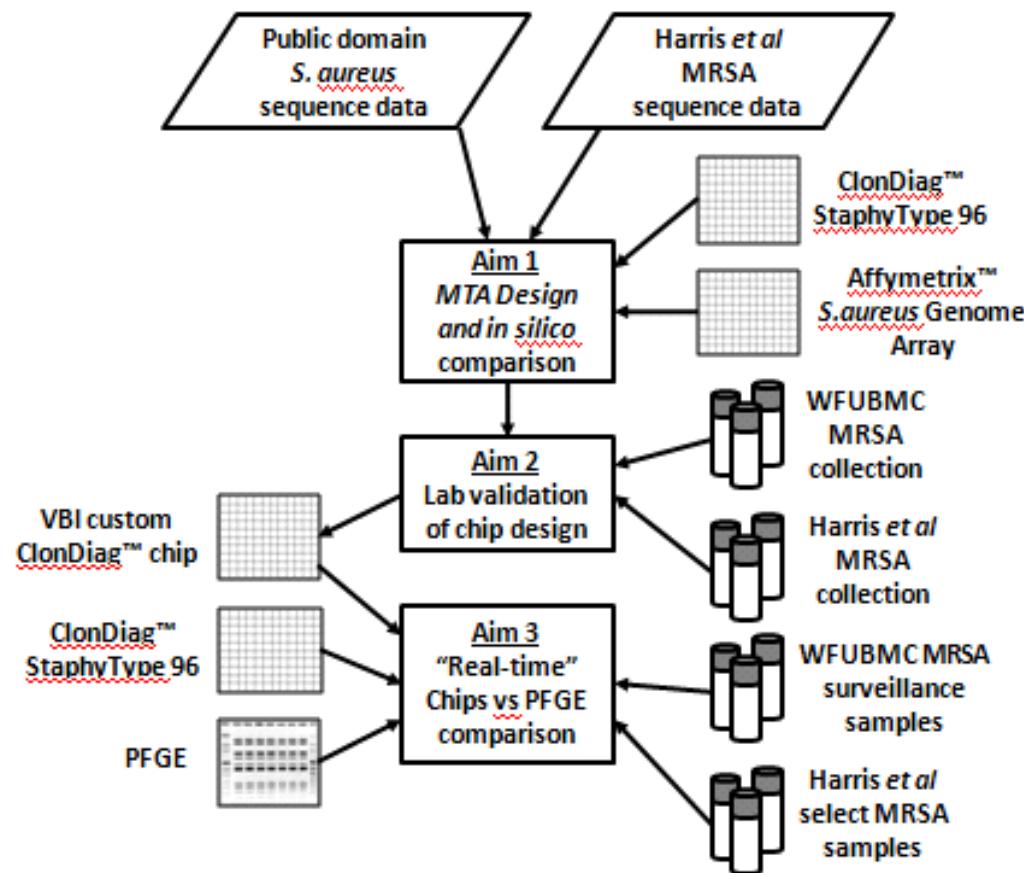
Distributed Informatics

**Figure 9. WFUBMC –V BI
CI Collaboration to
facilitate real time
visibility of microarray
data when assays are
finished.**



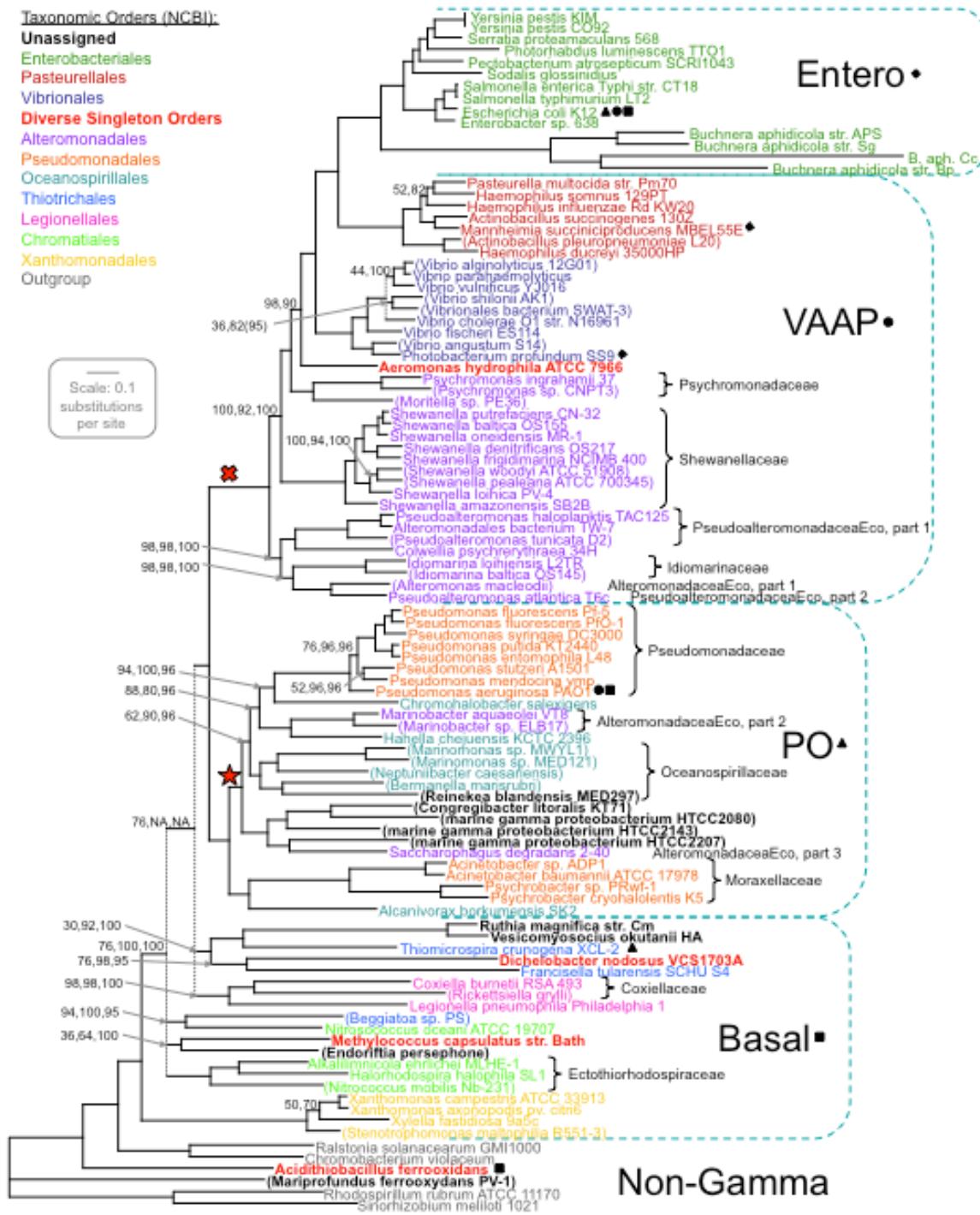
Laboratory Workflow (Assay)

Figure 10.
Grant
Summary



New Developments Underway

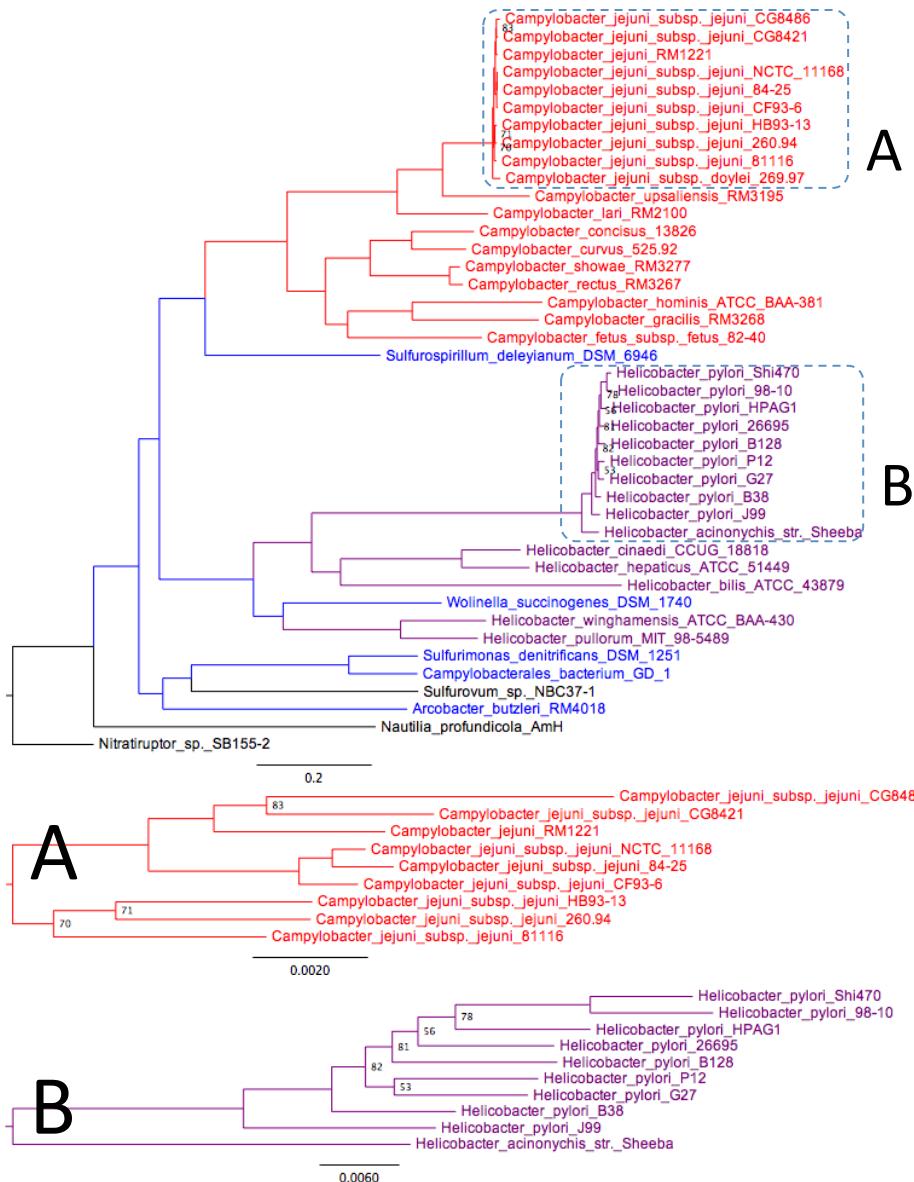
Gamma- proteobacteria



Robust Trees for 14 Pathogen Orders

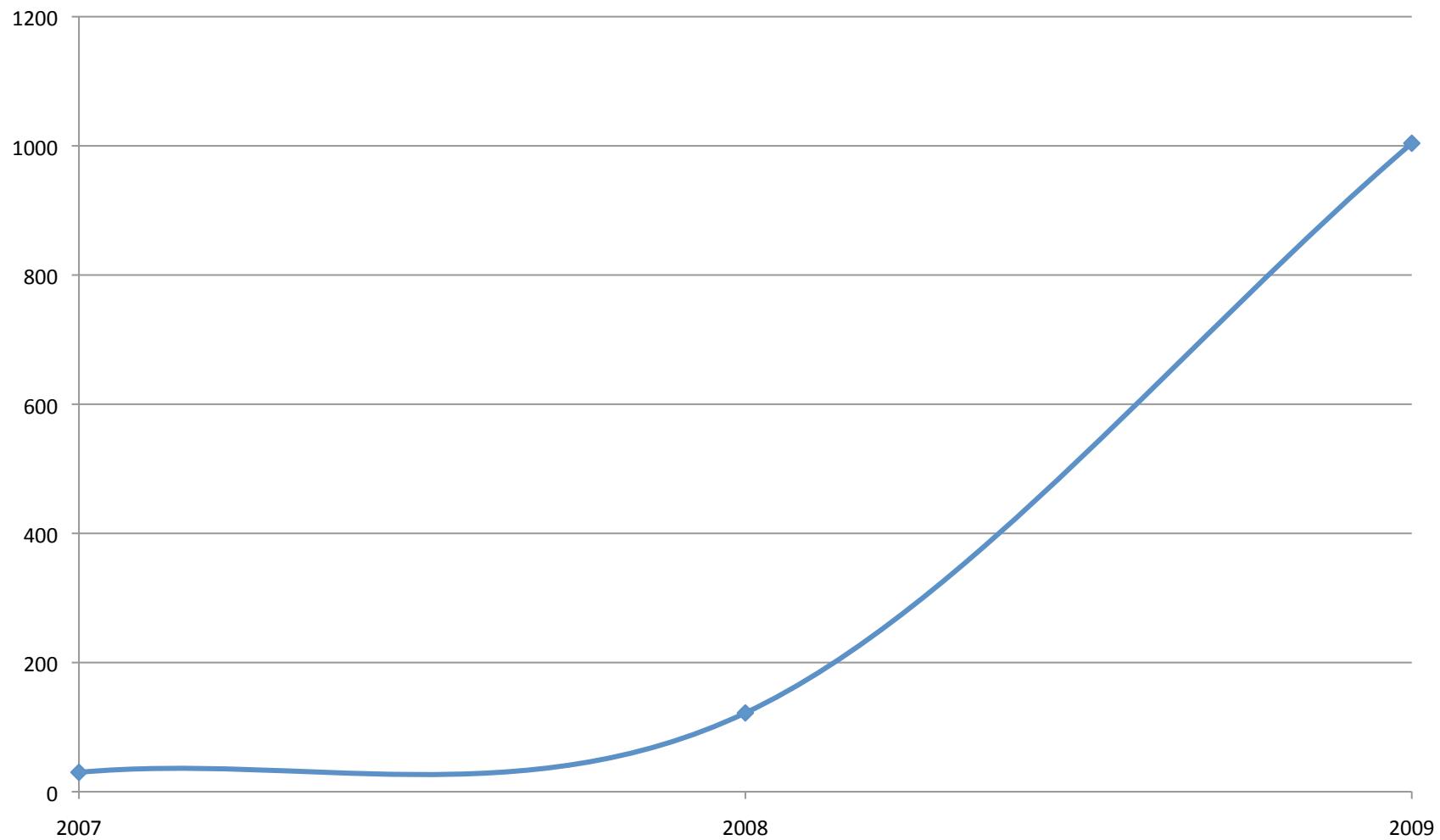
Order	Ingp	Outgp	Prots	Fams	Useful	Miss	Mult	Final	Chars
Actinomycetales	121	5	588874	38367	19212	5	5	187	41610
Bacillales	186	12	853072	28799	16934	2	2	150	40924
Burkholderiales	70	7	422753	24965	15534	0	0	436	130855
Campylobacteriales	39	3	82016	6279	3701	0	0	443	119266
Chlamydiales	16	7	50241	6202	1243	0	0	652	197381
Clostridiales	118	20	456016	29593	15005	5	5	174	35725
Enterobacteriales	140	12	698947	24950	14943	2	0	175	55592
Lactobacillales	161	14	422358	22876	11482	2	2	183	50130
Legionellales	13	13	78012	9275	3938	0	0	471	133035
Rhizobiales	76	9	372013	23107	14086	0	0	267	76978
Rickettsiales	39	8	93518	10135	3090	2	0	191	40572
Spirochaetes	36	3	78309	11579	4672	0	0	164	52132
Thiotrichales	21	13	89592	8034	3347	0	0	370	115209
Vibrionales	49	11	244824	13683	8263	0	0	553	159068

Example: Campylo- bacterales



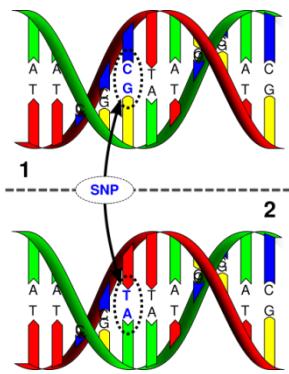
Sequence Read Archive (SRA)

Number of experiments in SRA (PATRIC organisms)

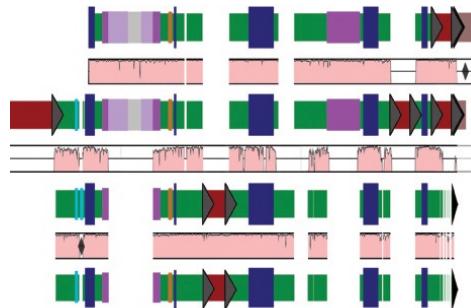


Uses of SRA datasets

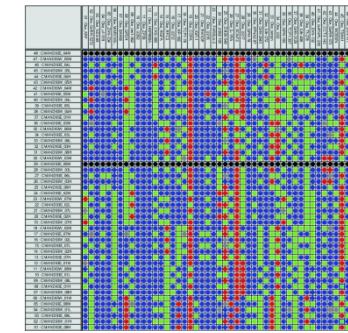
SNPs



Indels

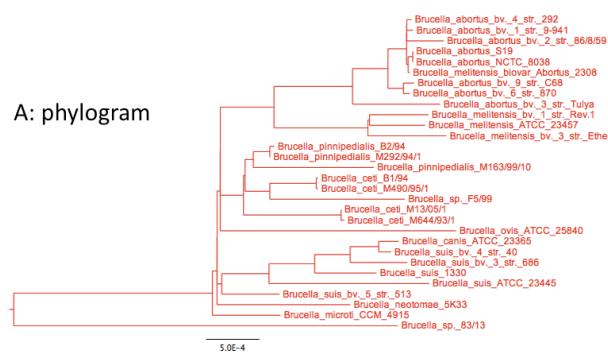


Genotyping

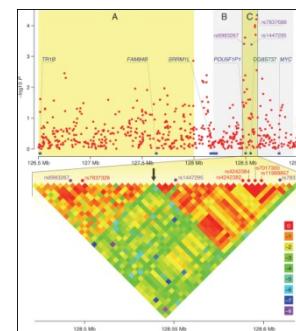


Phylogenetics

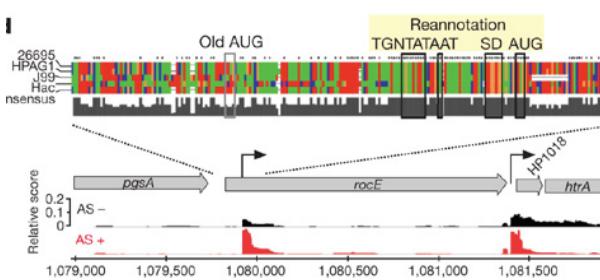
A: phylogram



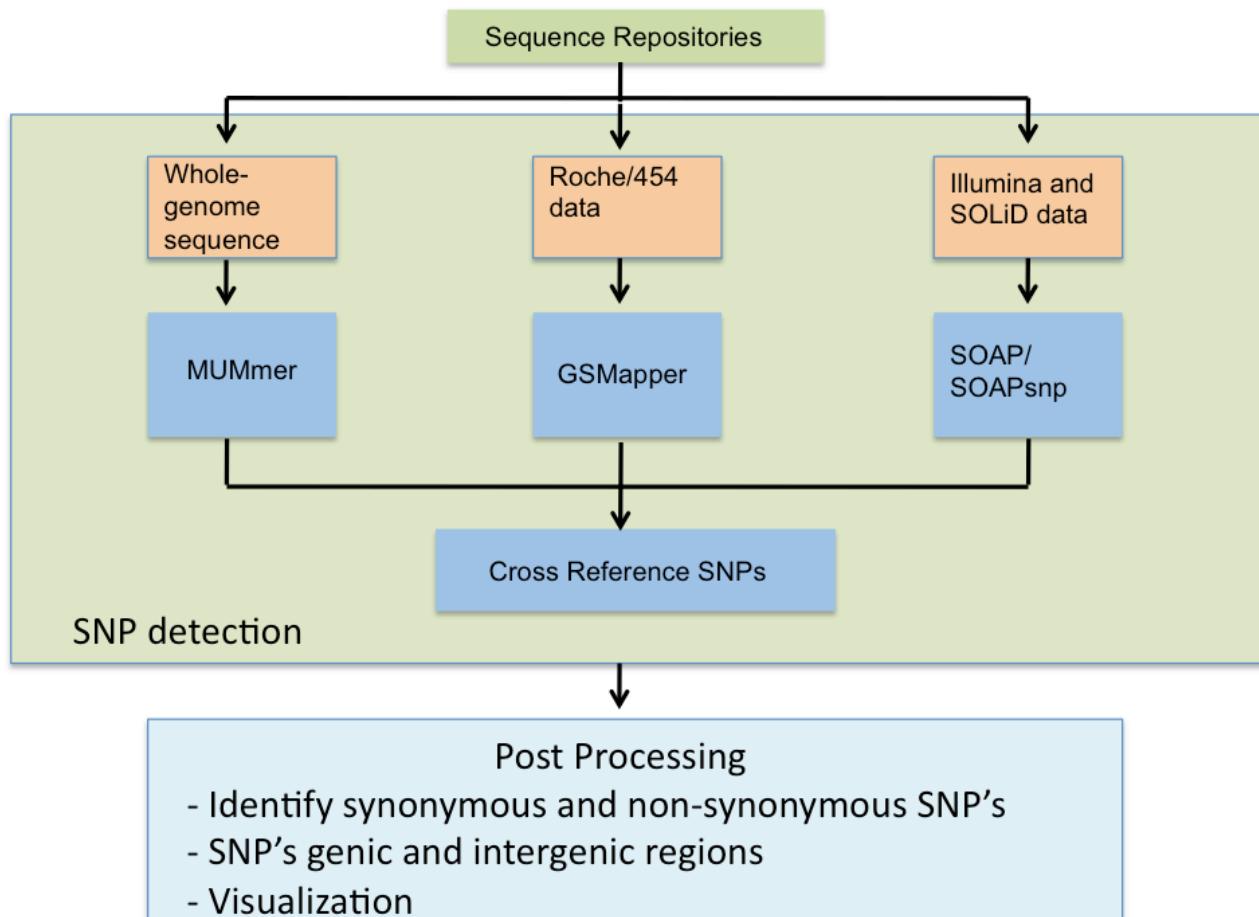
GWAS



RNA-seq



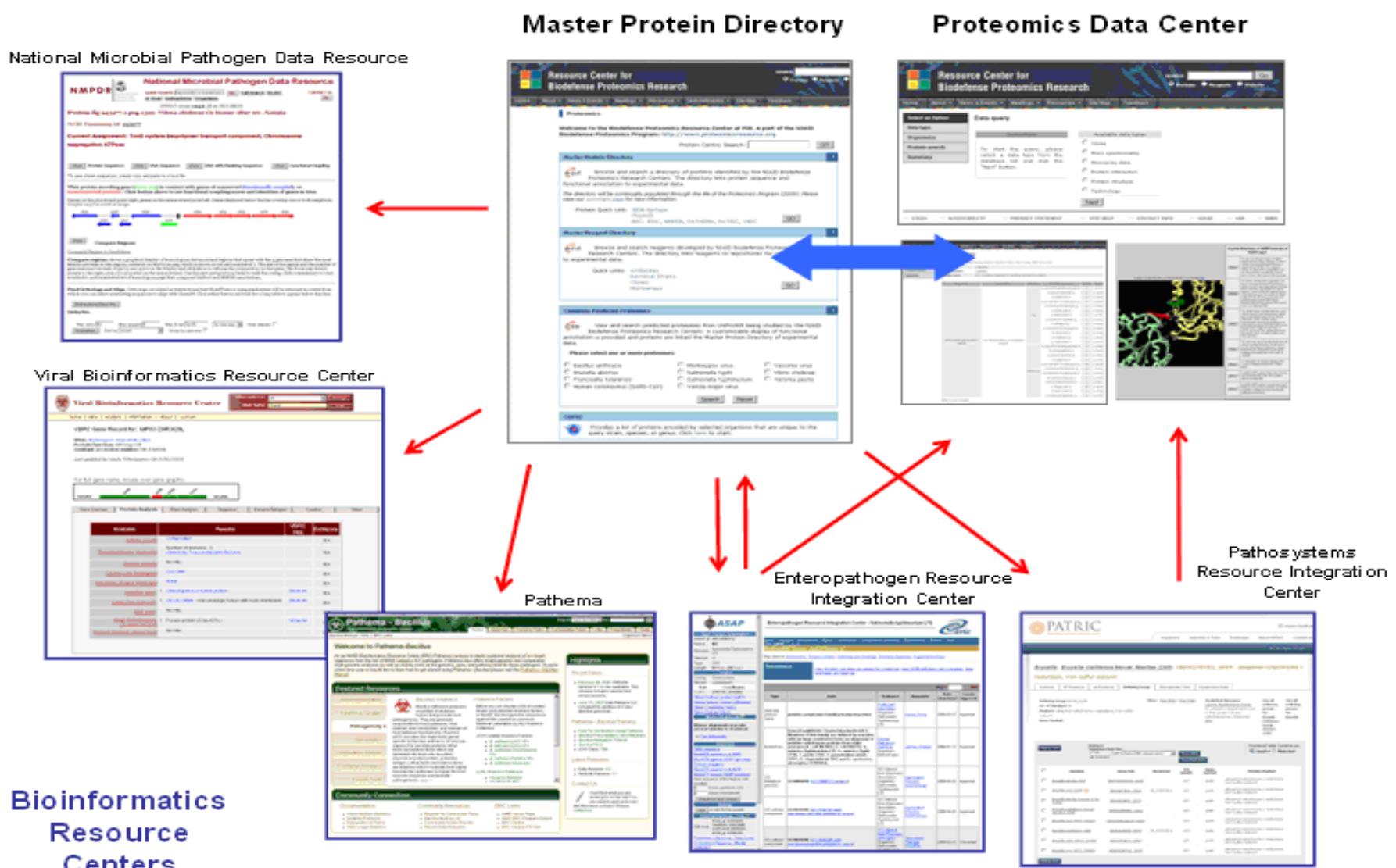
PATRIC SNP detection pipeline

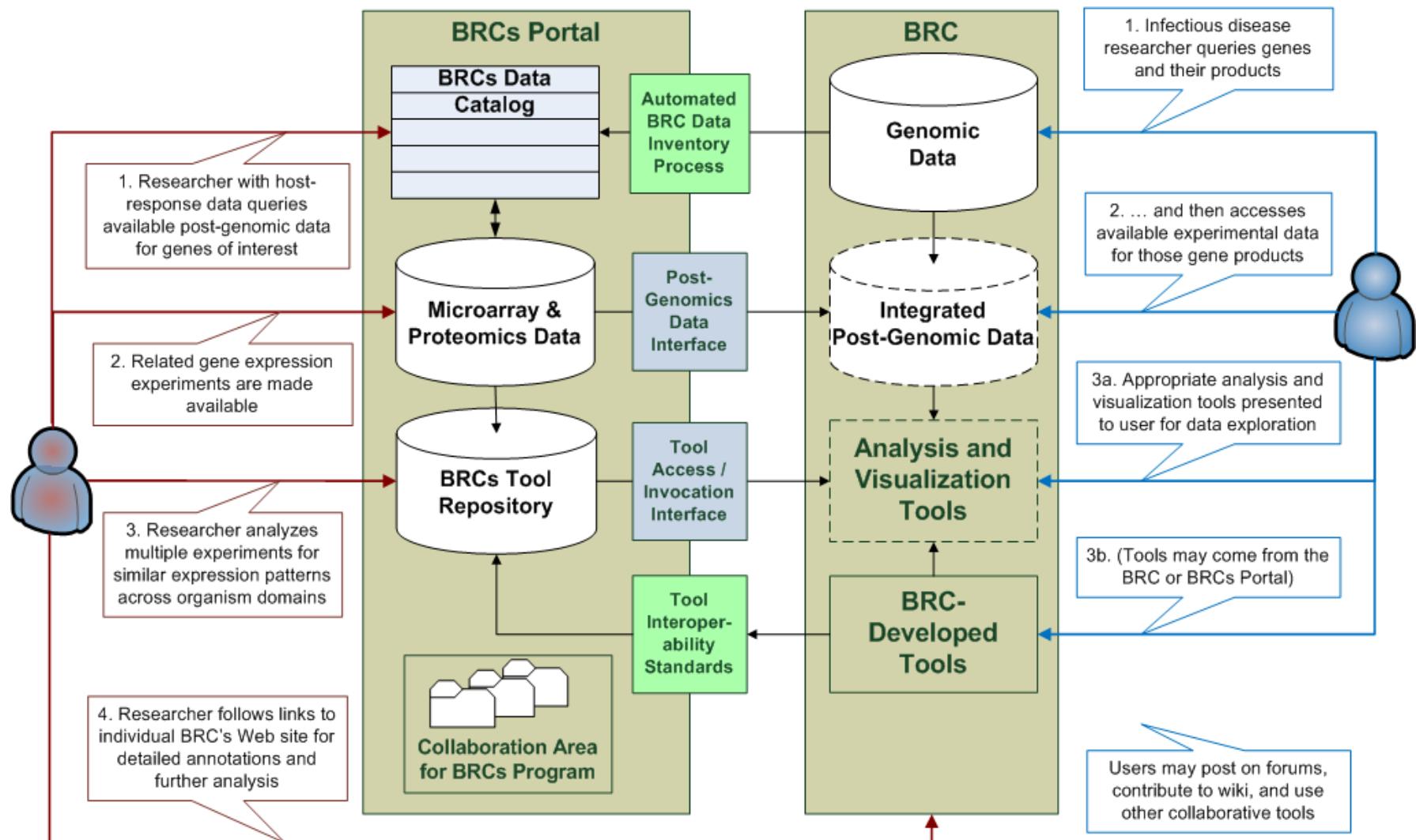
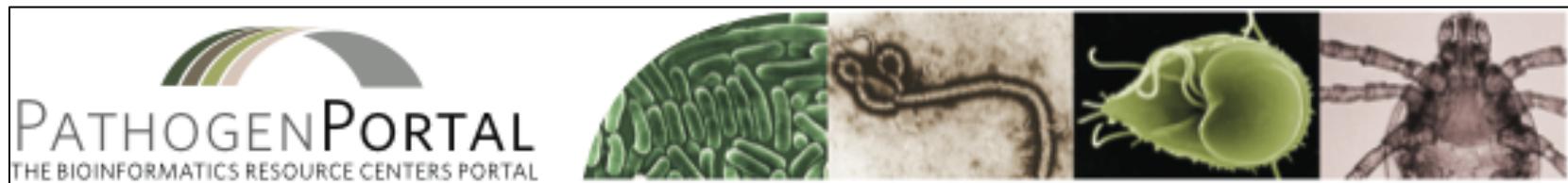


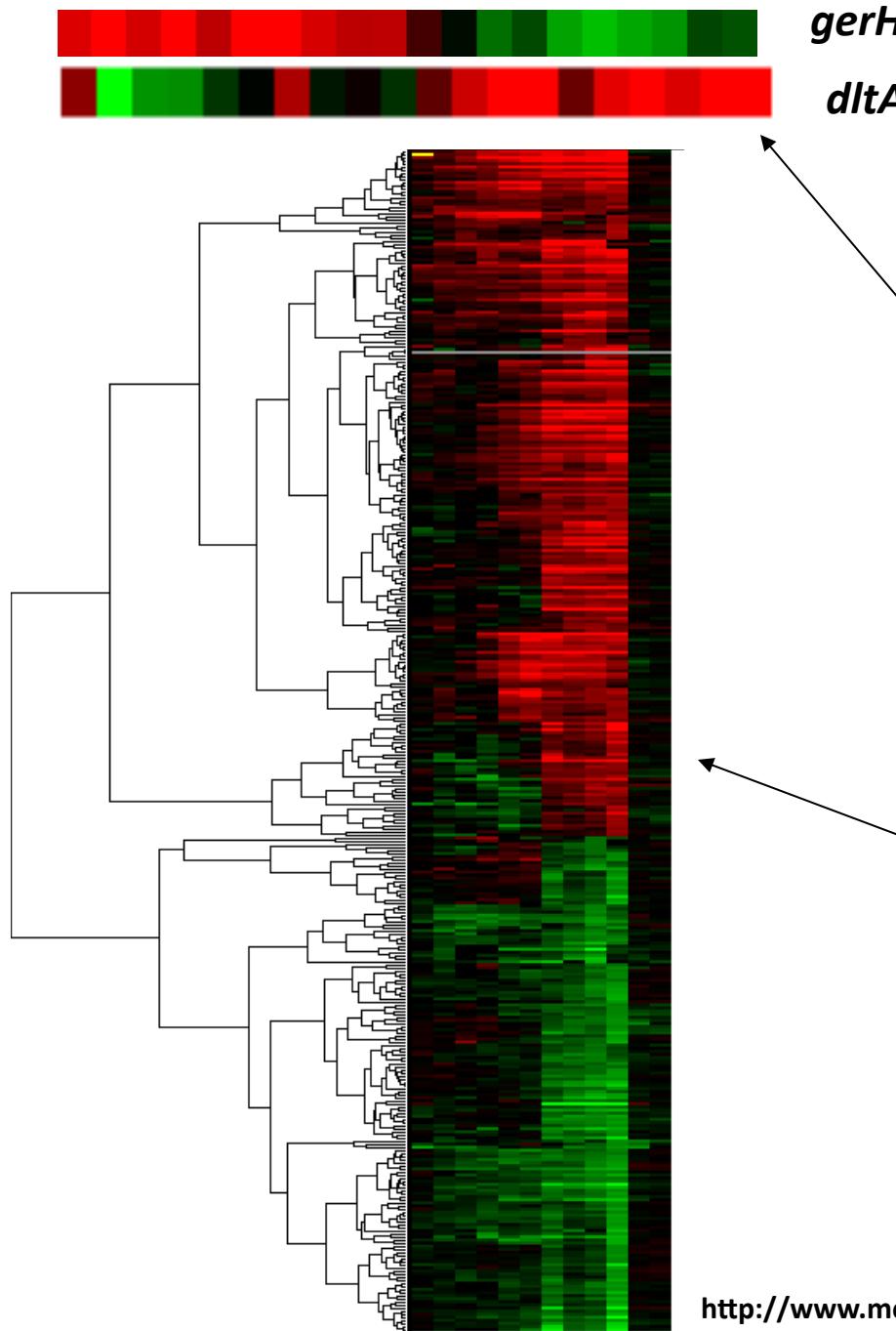
Post-Genomic Data
PATRIC/Pathogen Portal
(pathogenportal.org)

Former Proteomics Resource Center

Infectious Disease Resource Integration







Expression Data

Gene/Protein Expression of *B. anthracis* at Vital Infection Stages

Ex: *dltA* Controls Operon for D-alanine Conversion to Lipoteichoic Acid

- Cell Shape
- Autolytic Enzymes
- Resistance in Gram+
- Potential Drug Target

Host Expression Patterns During *B. anthracis* Infection (Murine Macrophage)

Validate Potential Intervention Points

<http://www.med.umich.edu/microbio/bio/hanna.htm>

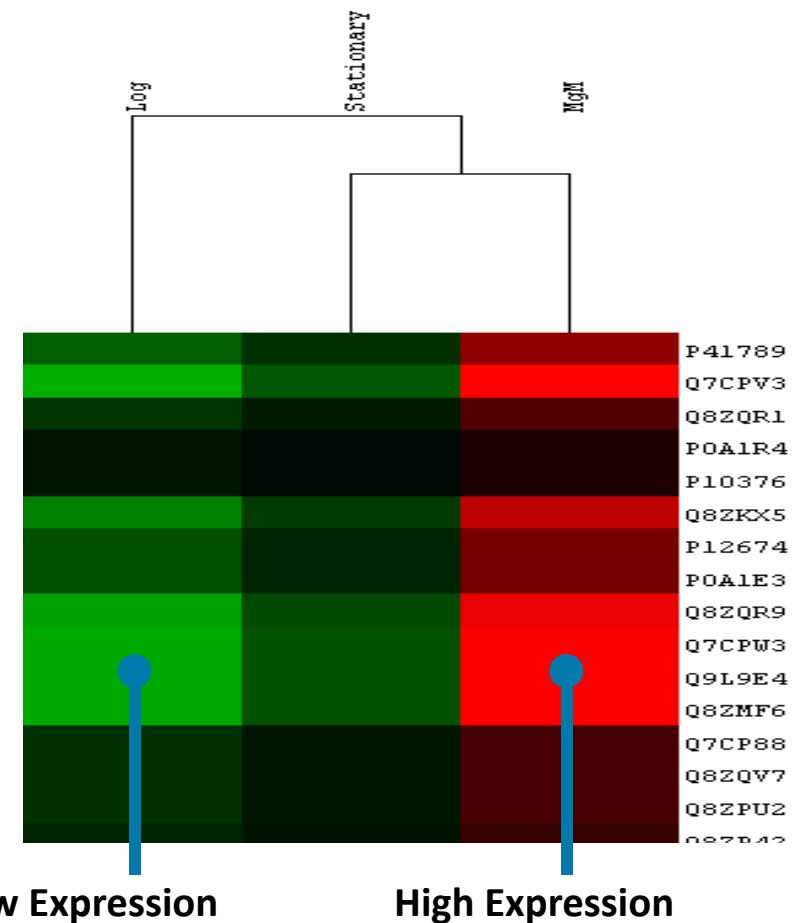
Mass Spec. Data

Mass and Time Tag Database for *Salmonella typhimurium* strain 14208

Protein ID	Protein name	Master Protein Directory
PSLT046	Putative carbonic anhydrase	Q7DAP2
Peptide		
K.GSNYDFVDAVAR.K		
XCorr		
K.HDYLAQK.R		
Validation		
K.IVGSMYHLTGGK.V		
R.DGMTPDAVIEHKF.Q		
R.KGSNYDFVDAVAR.K		
R.KNSPVLK.Q		
R.KNVELTIENIR.K		
R.KNVELTIENIR.K.N		

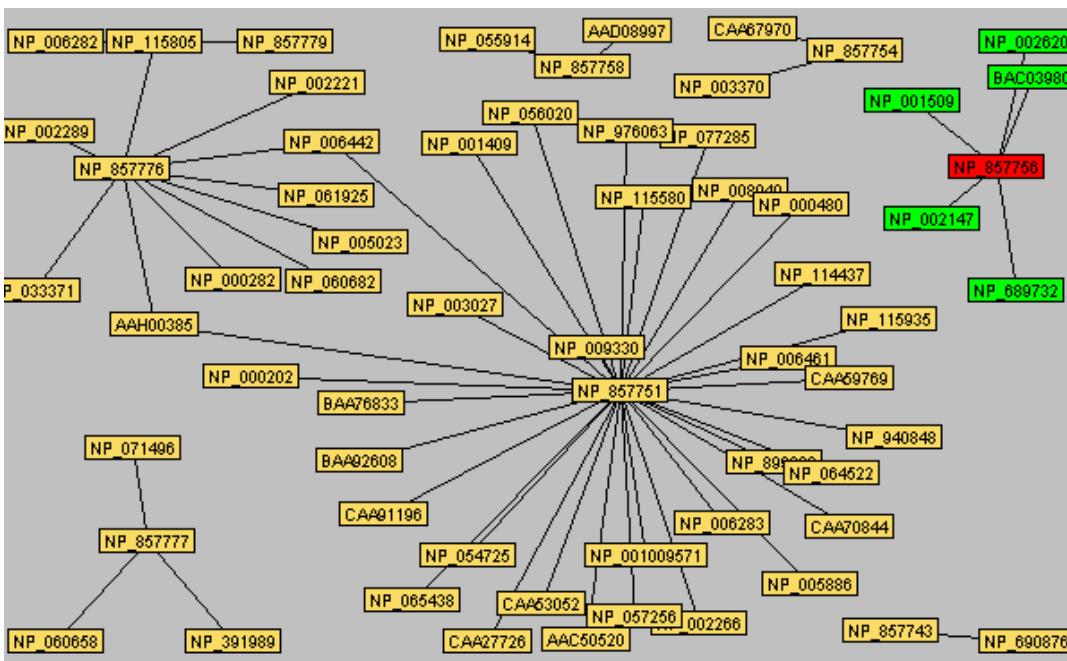
MS Data Forthcoming

Salmonella typhi strain TY2
Vaccinia Virus Western Reserve
Monkey Pox strain Zaire
Vaccinia Infected HELA cells
Vaccinia Infected THP-1



Protein-Protein Interaction Data

Interaction ID	Activation domain protein accession	Activation domain gene ID	Activation domain nucleotide accession	Activation domain protein GI	Binding domain protein accession	Binding domain gene ID	Binding domain nucleotide accession	Binding domain protein GI	Protein interaction detail
15369832	NP_005023	251	NM_005032	7549809	NP_857776	66945	NC_004836	31795319	View
15369833	NP_002289	153	NM_002298	4504965	NP_857776	66945	NC_004836	31795319	View
15369834	NP_060682	12251	NM_018212	39930375	NP_857776	66945	NC_004836	31795319	View
15369847	CAA53052	2409	X75304	405715	NP_857751	66912	NC_004836	31795291	View
15369850	NP_005886	956	NM_005895	5174441	NP_857751	66912	NC_004836	31795291	View
15369851	NP_008949	8297	NM_007018	38158018	NP_857751	66912	NC_004836	31795291	View
15369882		23342		0	NP_857751	66912	NC_004836	31795291	View
15369883	NP_001009571	39847	NM_001009571	57546902	NP_857751	66912	NC_004836	31795291	View
15369884	NP_000480	236	NM_000489	20336209	NP_857751	66912	NC_004836	31795291	View
15369886	NP_003027	238	NM_003036	4506967	NP_857751	66912	NC_004836	31795291	View
15369887	CAA70844	8492	Y09631	3925685	NP_857751	66912	NC_004836	31795291	View
15369888	BAA92608	10638	AB037791	7243121	NP_857751	66912	NC_004836	31795291	View
15369891	NP_115580	33255	NM_032204	20270253	NP_857751	66912	NC_004836	31795291	View
15369892	NP_006283	310	NM_006292	5454140	NP_857751	66912	NC_004836	31795291	View
15369893	CAA27726	2904	X04106	35328	NP_857751	66912	NC_004836	31795291	View



<http://www.myriad.com/index.php>

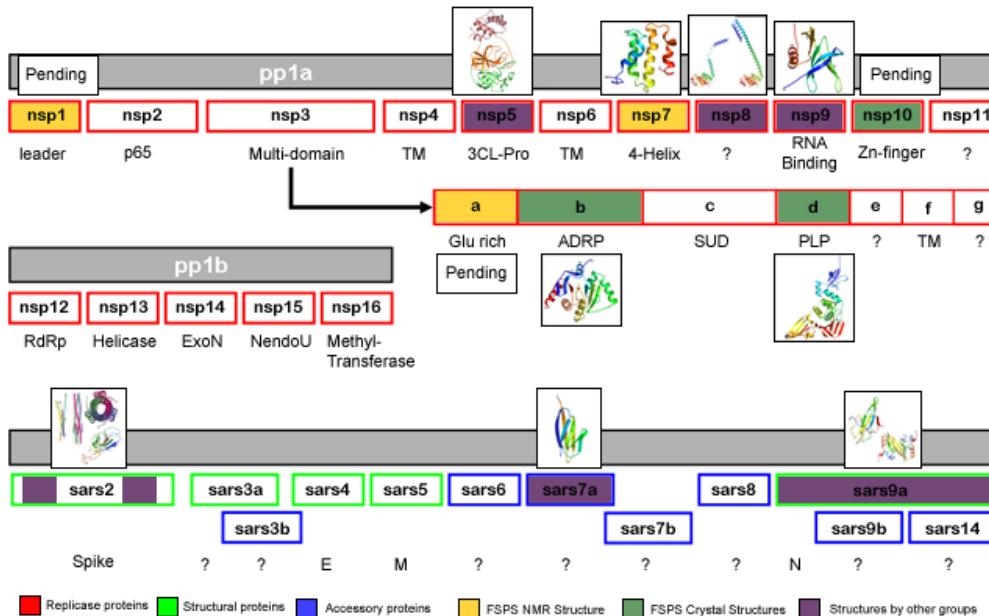
High Throughput Yeast Two-Hybrid Technology

Host/Pathogen Interactions

Biochemical Pathways

Proteins Binding to Membrane Proteins

SARS Proteome 3D Structures



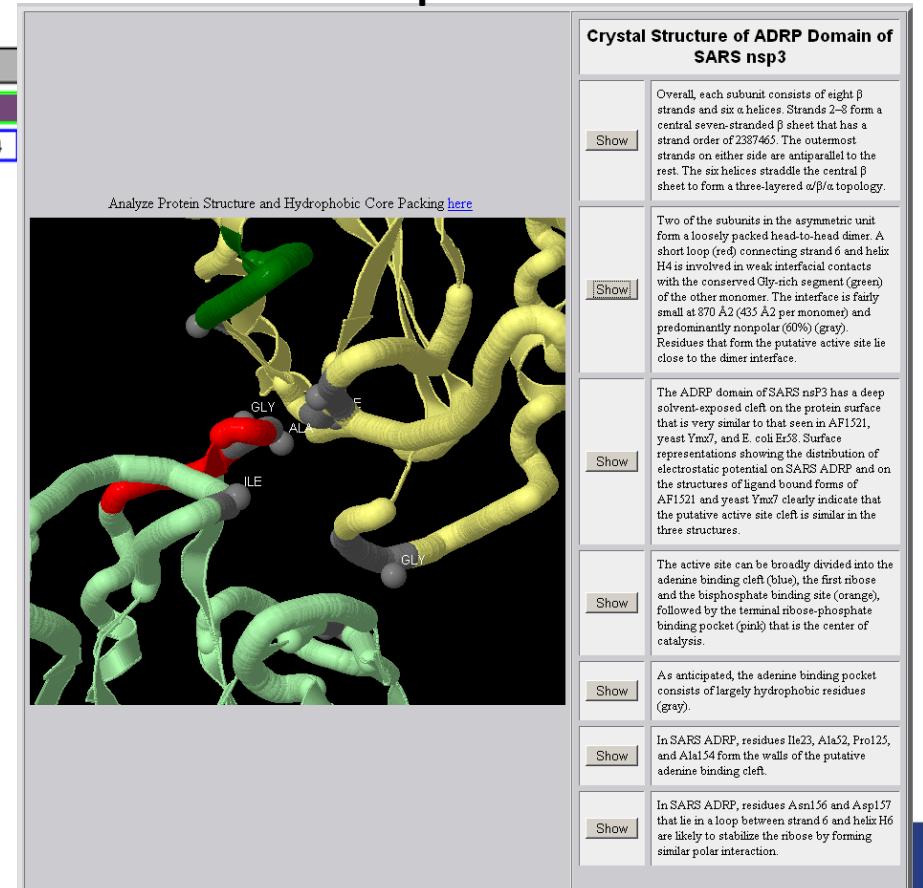
Annotations Derived from Published Manuscripts

Illustrate Features Described in Paper

Fully Interactive 3D Graphics

—
Proteome Via
Crystallography, NMR,
Microscopy

SARS nsp3 ADRP



PATRIC 3-D Structure Viewer

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

Epitope Selection Box →

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 Information →

Highlight Protein Areas of Interest
IEDB Epitopes 7115

Area of Interest Details
Epitope ID 7115
Name
Peptide Sequence TTLPVNVAF
Range 47 to 56
Debug selecting 47 to 56

View NT/AA Sequence
Goto AA Evidence Page
View NCBI BLASTP Result

3D Structure Visualization [JMOL Documentation](#)

Jmol

Appearance
Wireframes off
Cartoons thin
Spacefill off
Surface off
Labels off
Show everything

Navigation
Spin off
Zoom 100%

Zoom
Mouse scroll wheel, or Shift+Left mouse button.

Rotate
Left mouse button

Options
Right mouse button, or Ctrl+Left mouse button

Visualization Control Box ←

Zoom and Spin Control ←

Structure Analysis at PATRIC

1D:

Identify What Structures are Available for All Sequences

Identify Homologues for All Sequences – 30% Identity

3D:

Compare Two Different Structures

Compare Conformations of a Structure

Compare a Structure to a Database of Structures

Find All structures With a Motif

Find All Similarities (Motifs) in All Structures

Host-Pathogen Interactions

PATRIC/Pathogen Portal

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

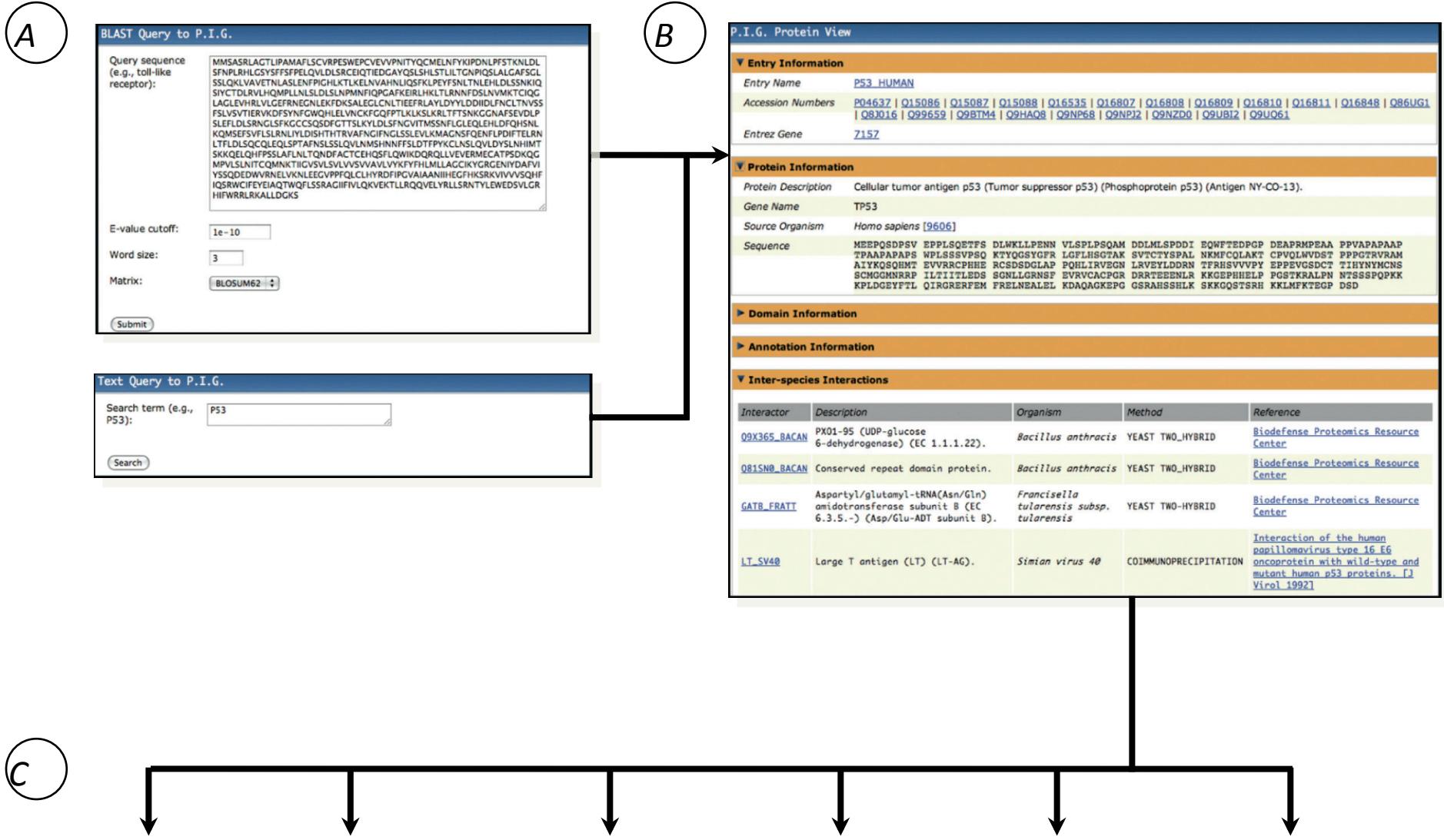
- Site-wide
modern web browser
internet connection
javascript enabled
- Visualizations
[java 1.5](#)

About PIG

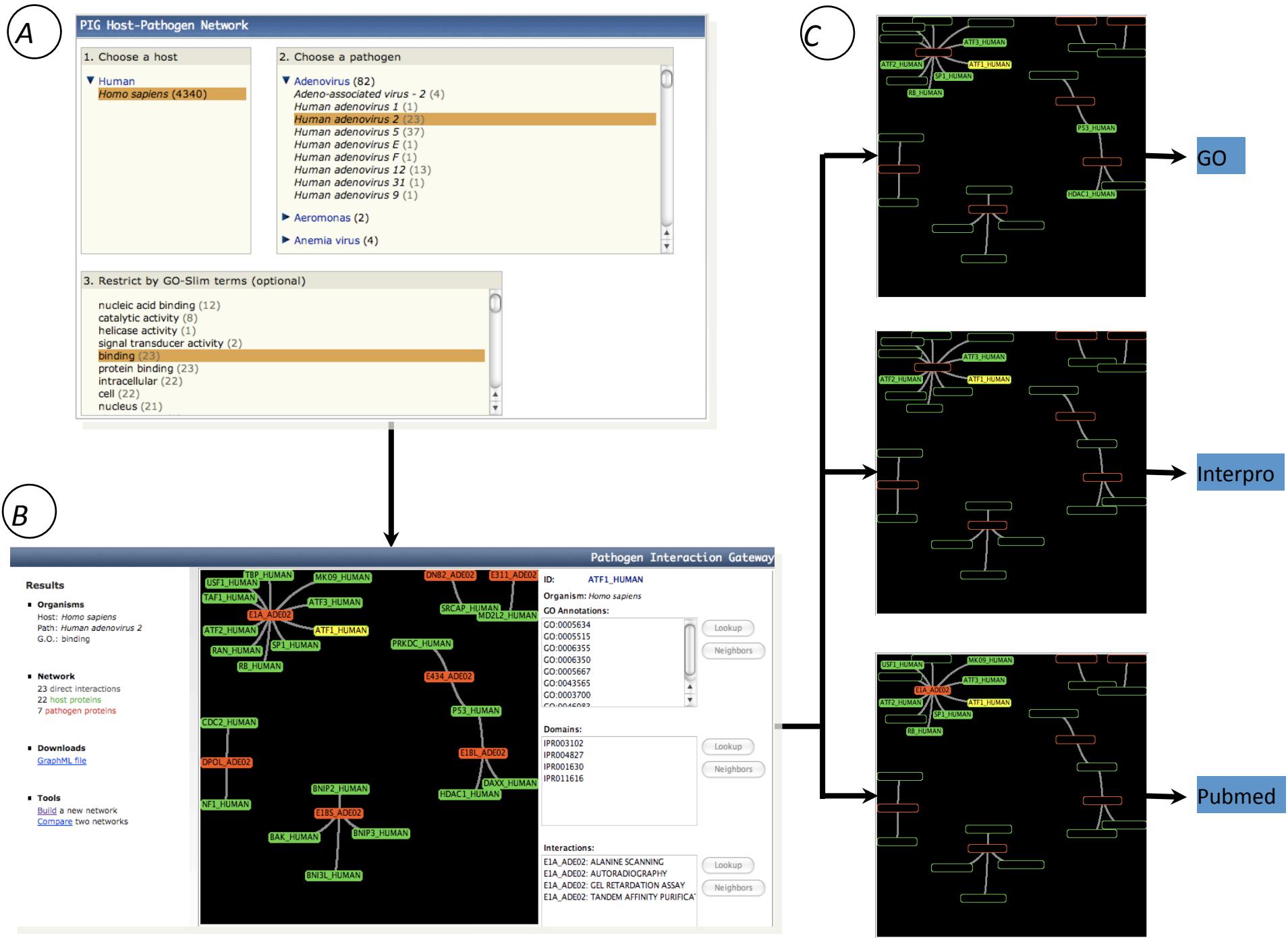


Protein-protein interactions (PPIs) play a vital role in initiating infection in a number of pathogens and identifying which interactions allow a pathogen to infect its host can provide potential targets for therapeutics. Public resources for studying host-pathogen systems, in particular PPIs, are scarce. 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.

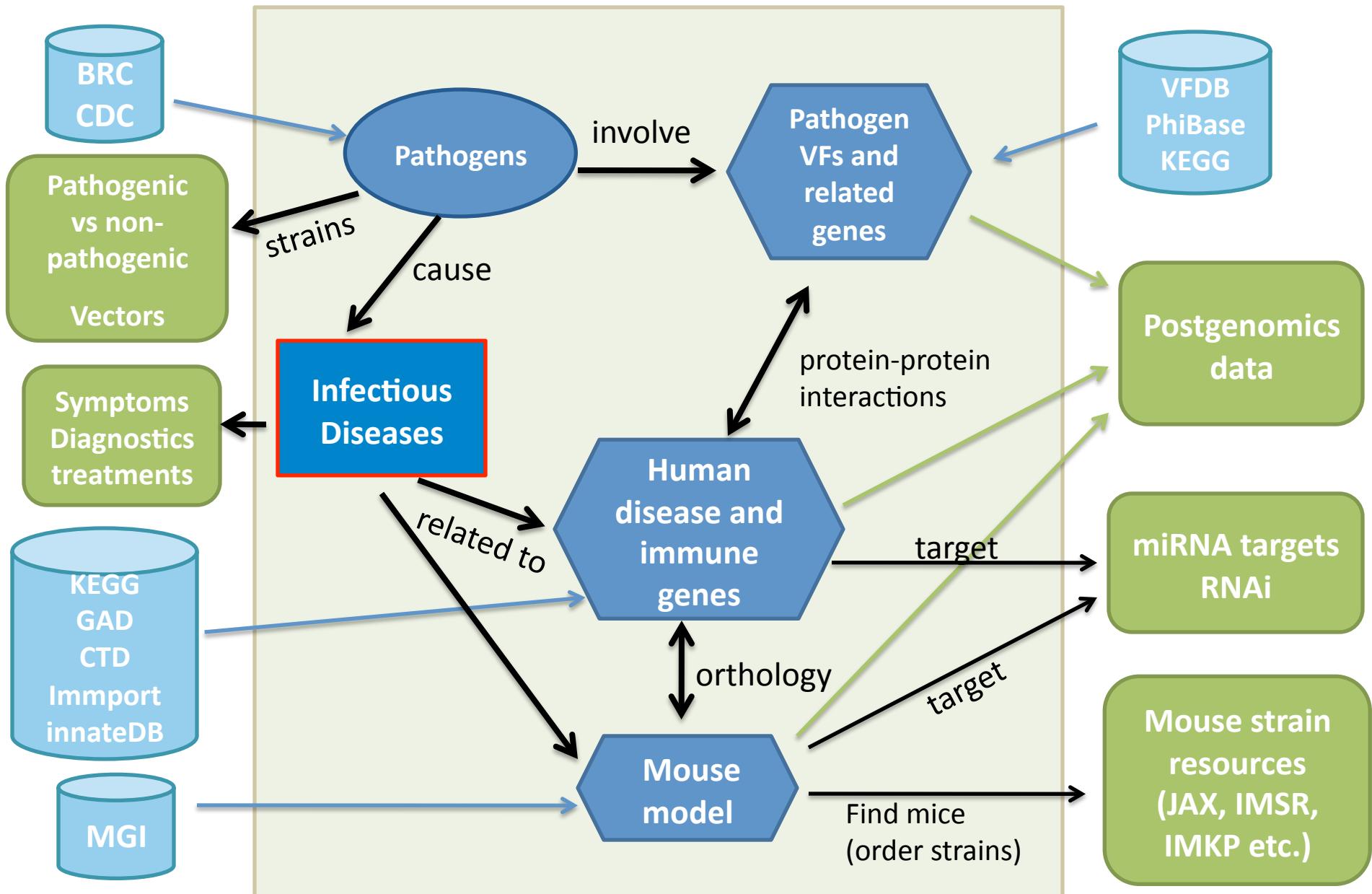
<ul style="list-style-type: none">◦ BLAST search◦ Text search	<ul style="list-style-type: none">◦ Network Visualizer◦ Network Comparer	<ul style="list-style-type: none">◦ Data Download◦ Software Download◦ If you would like to contribute to PIG, please contact us here.
--	---	---



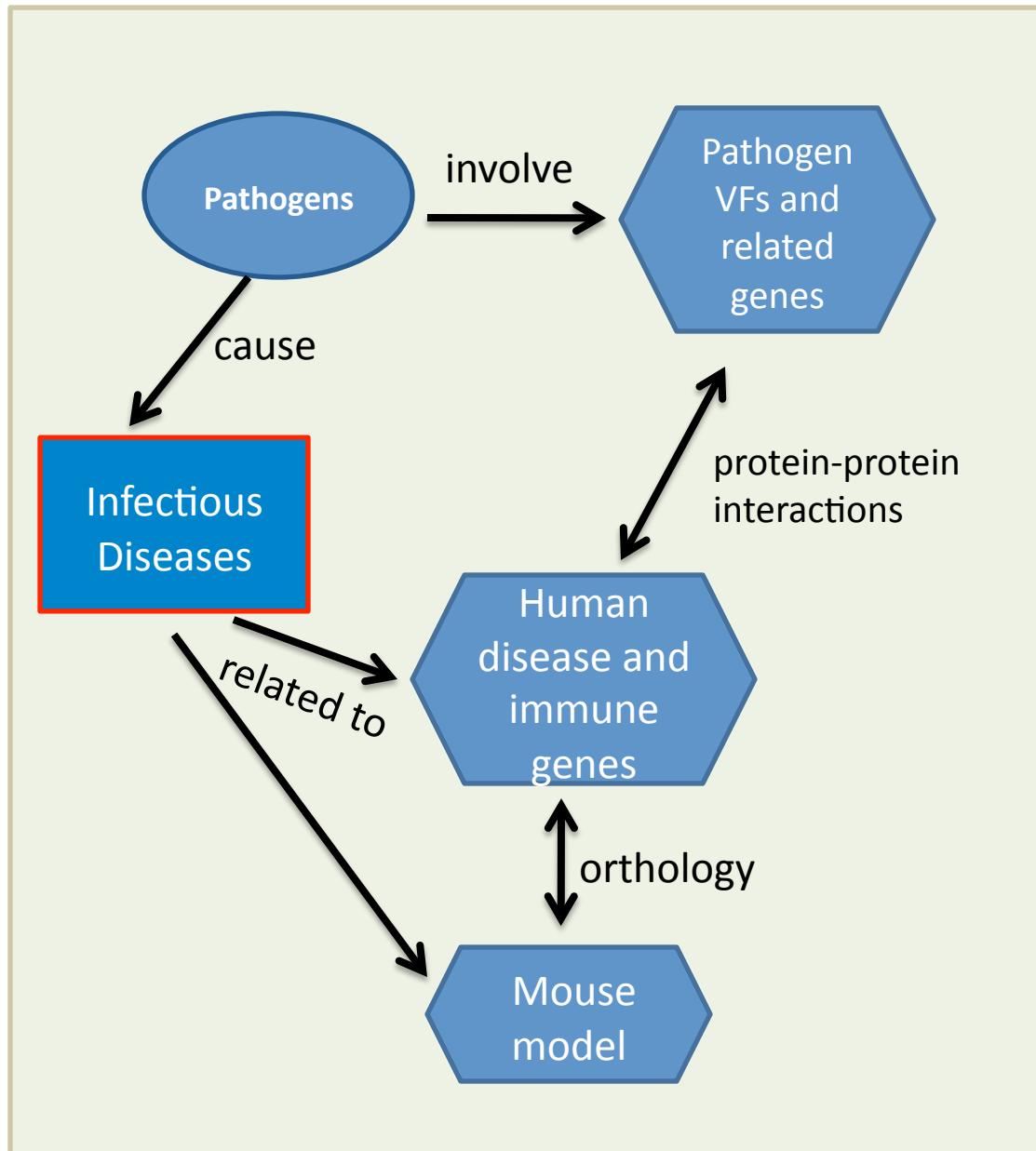
Uniprot Entrez Gene NCBI Taxonomy Interpro Gene Ontology Pubmed



Infectious Disease Conceptual Framework



Infectious Disease Conceptual Framework



Outreach

Driving Biological Projects (DBPs)

The screenshot shows the PATRIC website homepage. At the top, there's a navigation bar with links for Home, Organisms, Searches & Tools, Downloads, About PATRIC, and Contact Us. Below the navigation is a banner stating "No Items, No Groups".

Welcome to PATRIC

Currently, PATRIC is in a transition period - consolidating previous BRCs into new NIAID contracts. We anticipate adding new data, tools, and website features over the next few months.

For previous and current NIAID bacterial BRC users, we understand that these resources may be very valuable to your work. As such, we will be doing our very best to create a useful PATRIC resource to continue supporting your work. We realize that the transition time will cause disruptions, however, it is very important to us to work with established BRC communities to identify and prioritize our transition efforts. We will have a survey online soon to help us identify your needs, but in the meantime, you may contact us at patric@vbi.vt.edu.

Watchlist Genera

PATRIC provides rich data and analysis tools for all bacterial species with an emphasis on the bacterial Orders that include NIAID category A-C priority and emerging/re-emerging pathogens.

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

See List of All Bacteria
 See List of Bacterial Pathogens

Data Summary

	From incumbent BRCs	From RefSeq
Number of genomes	409	2,329
Number of genomic features	2,297,834	14,944,546

Call for Projects

PATRIC is seeking Driving Biological Projects that use high-throughput experimental technologies to **functionally characterize** the genome, proteome or metabolome of bacterial organisms or host/pathogen interactions to help elucidate how genes, proteins and metabolites may be involved in pathogenesis, antimicrobial resistance or other biological processes. **UPDATE - review timeline has been updated**

Announcement

- PATRIC Data and Website Release (30-Dec-2009)
- PATRIC Data and Website Release (9-Dec-2009)
- Bioinformatics Resource Centers for Infectious Diseases
- VBI Awarded \$27M from NIH to Support Infectious Disease Research

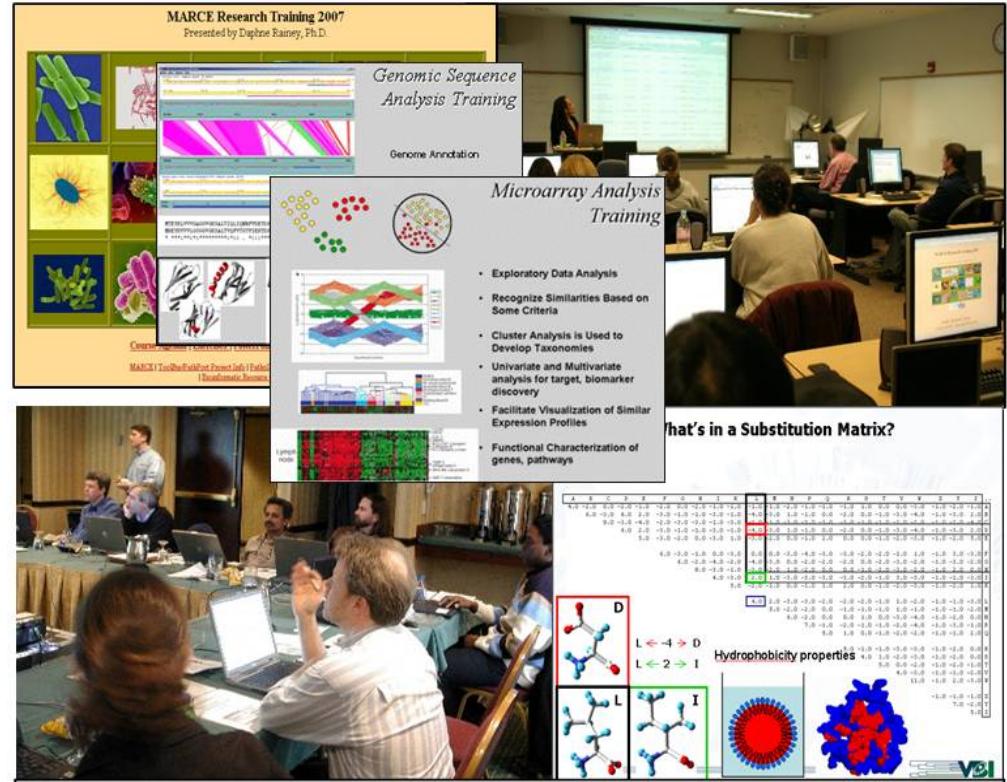
• Review begins April 1st
• 2 proposals will win (May 2010), spurring new data and novel infrastructure
• \$600K over 2 years
• Next call for proposals will be in 2012

DBP Goals

- Use high throughput experimental technologies (HTP)
- Functionally characterize genome, proteome or metabolome of bacterial organisms or host/pathogen interactions
- Elucidate mechanisms of pathogenesis, antimicrobial resistance, or other biological processes of interest to study of ID
- Collaborate with BRC to analyze data and drive new BRC functionality and features

Workshops

- 3-5 days
- Features
 - Project-centric
 - Resource-centric
 - Feedback collection
 - Collaboration
- 1st workshop
 - Topic: Comparative Genomics
 - Late Summer, 2010, Washington DC area



Contact Us

The screenshot shows the PATRIC website's contact page. At the top, there is a navigation bar with links for Home, Organisms, Searches & Tools, Downloads, About PATRIC, and Contact Us. Below the navigation bar, a dark blue header bar displays a shopping cart icon and the text "No Items, No Groups". On the left side, there is a sidebar titled "Inside - About PATRIC" containing links for What is PATRIC?, News & Announcement, Data & Software Release, Presentations, Publications, Related Sites, and Contact Us. The main content area is titled "Contact Us" and contains two sections: "Mailing address" and "Primary Contacts". The "Mailing address" section provides the physical address of the Virginia Bioinformatics Institute, Washington Street, MC 0477, Blacksburg, VA 24061, along with office phone number (540) 231-2100, fax number (540) 231-2606, and email address patic@vbi.vt.edu. The "Primary Contacts" section lists Bruno Sobral, PhD as the PATRIC Principal Investigator, with phone number (540) 231-2100 and email sobral@vt.edu. It also lists Ronald Kenyon as the PATRIC Project Manager, with phone number (540) 231-7470 and email rkenyon@vbi.vt.edu. To the right, there is a box titled "Submit Feedback" with a message encouraging users to submit suggestions for improvement and errors missed, followed by a "Submit Feedback" button.

PATRIC
PathoSystems Resource Integration Center

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: patic@vbi.vt.edu

Primary Contacts

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

Ronald Kenyon
PATRIC Project Manager
(540) 231-7470
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

pathogenportal.org

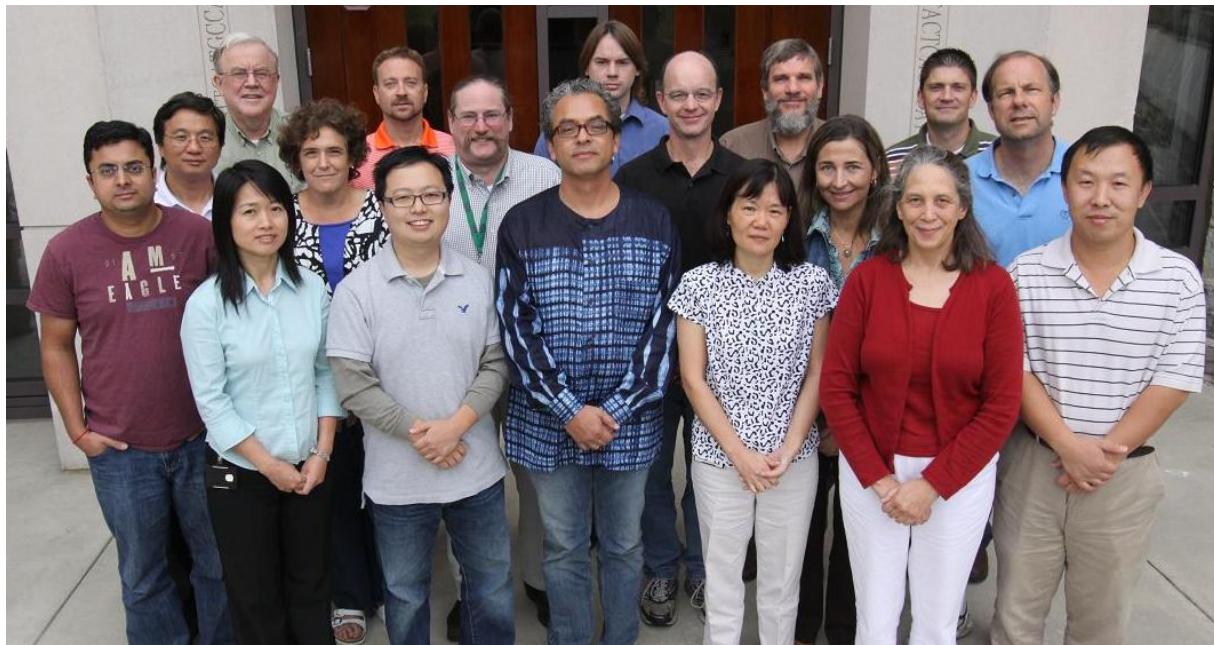
patic@vbi.vt.edu

PATRIC & Portal Project Team

Cyberinfrastructure Group

Bruno Sobral Stephen Cammer
Rebecca Will Joe Gillespie
Ron Kenyon Julie Shulman
Eric Snyder Rebecca Wattam
Joe Gabbard Kelly Williams
Debby Hix Chunhong Mao

Jian Lu
Shrinivasrao Mane
Eric Nordberg
Mark Scott
Maulik Shukla
Jim Stoll
Dan Sullivan
Tian Xue
Harry Yoo
Chengdong Zhang
Tim Driscoll
Oral Dalay



RAST Team (ANL & U Chicago)

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

NaCTeM (U of Manchester & U Tokyo)

Sophia Ananiadou 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.