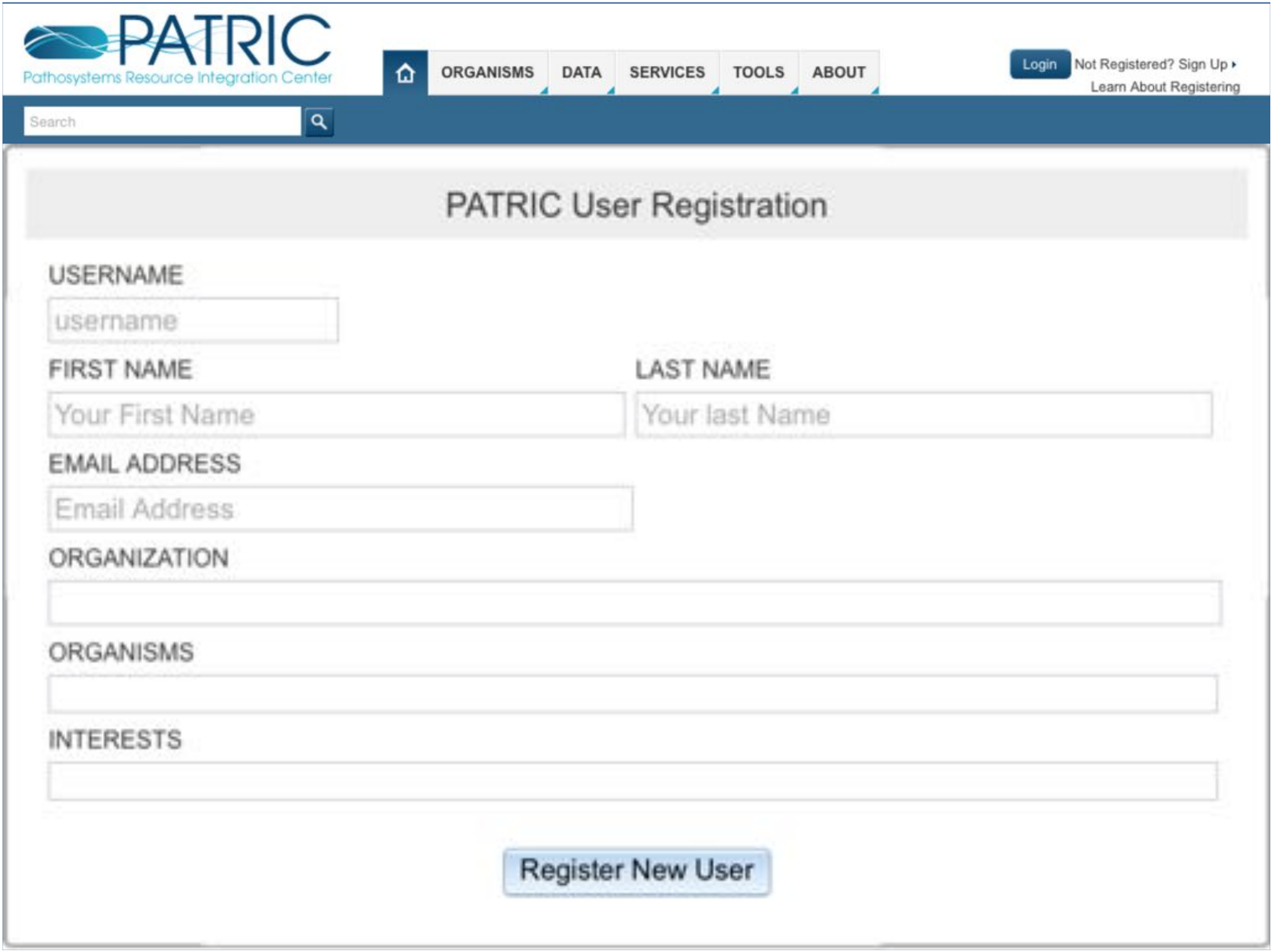
PATRIC Bioinformatics Resource Center

Argonne National Lab

Lemont, Illinois

May 7-9, 2018

www.patricbrc.org


PATRIC
Pathosystems Resource Integration Center
ORGANISMS
DATA
SERVICES
TOOLS
ABOUT
Login
Not Registered? Sign Up
Learn About Registering
Search
PATRIC User Registration
USERNAME
username
FIRST NAME
Your First Name
LAST NAME
Your last Name
EMAIL ADDRESS
Email Address
ORGANIZATION
ORGANISMS
INTERESTS
Register New User

Workshop instructors

- Neal Conrad
- Marcus Nguyen
- Rebecca Wattam
- Jim Davis
- Maulik Shukla
- Bruce Parrello
- Ross Overbeek
- Janaka Edirisinghe

NIH/NIAID BRC Program

Provide publicly accessible database to:

- Store, update, integrate and display genome sequence data, annotation and associated data for human pathogens.
- Provide query, analysis and visualization of information with user friendly interfaces.
- Serve as public repository for NIAID-supported genome-scale programs.
- Collaborate on experimental research projects.

NIAID Bioinformatics Resource Centers

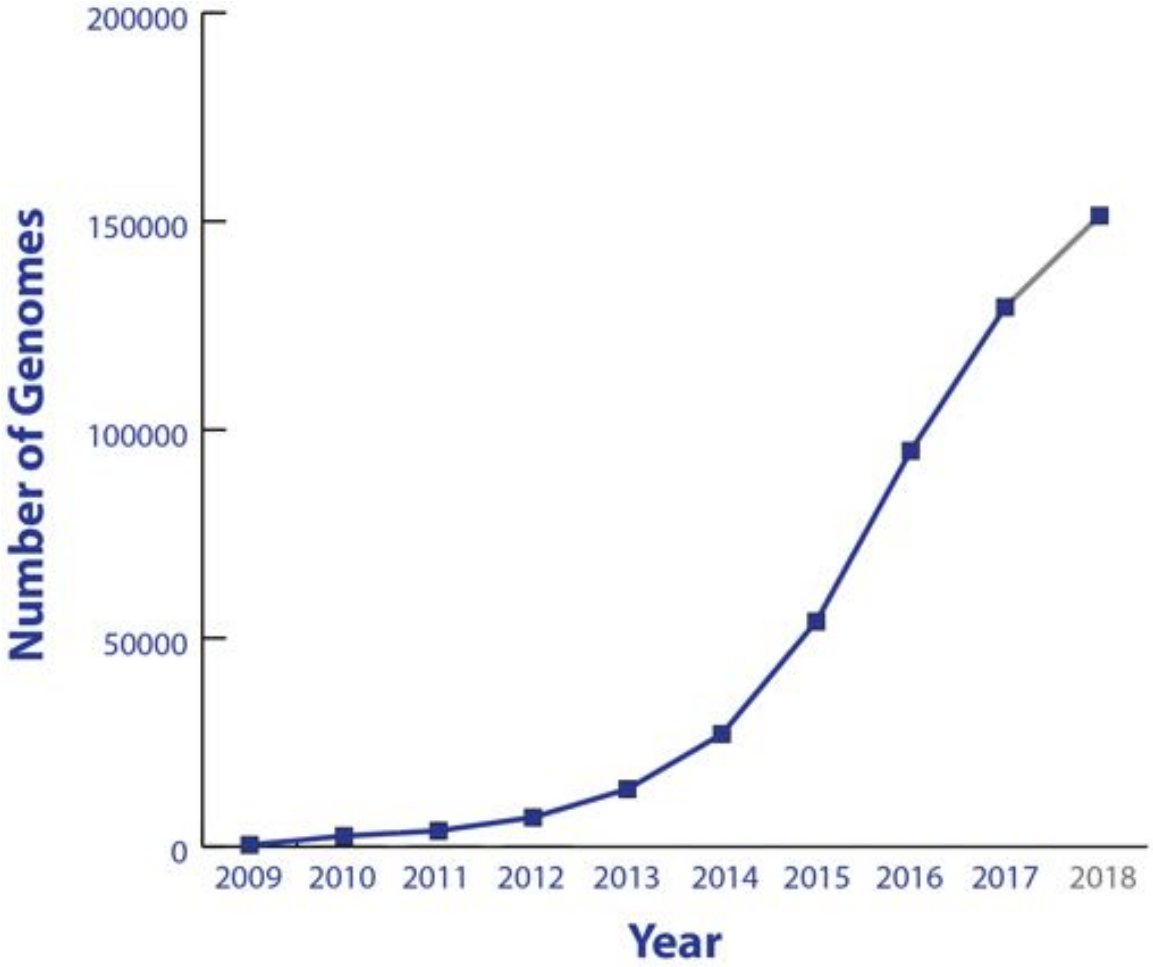
- PATRIC – bacteria
  - https://patricbrc.org/
- ViPR – viruses
  - https://www.viprbrc.org
- EuPathDB –eukaryotes
  - https://eupathdb.org
- Vectors – invertebrates
  - https://www.vectorbase.org/

PATRIC (NIAID) Watchlist Genera

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

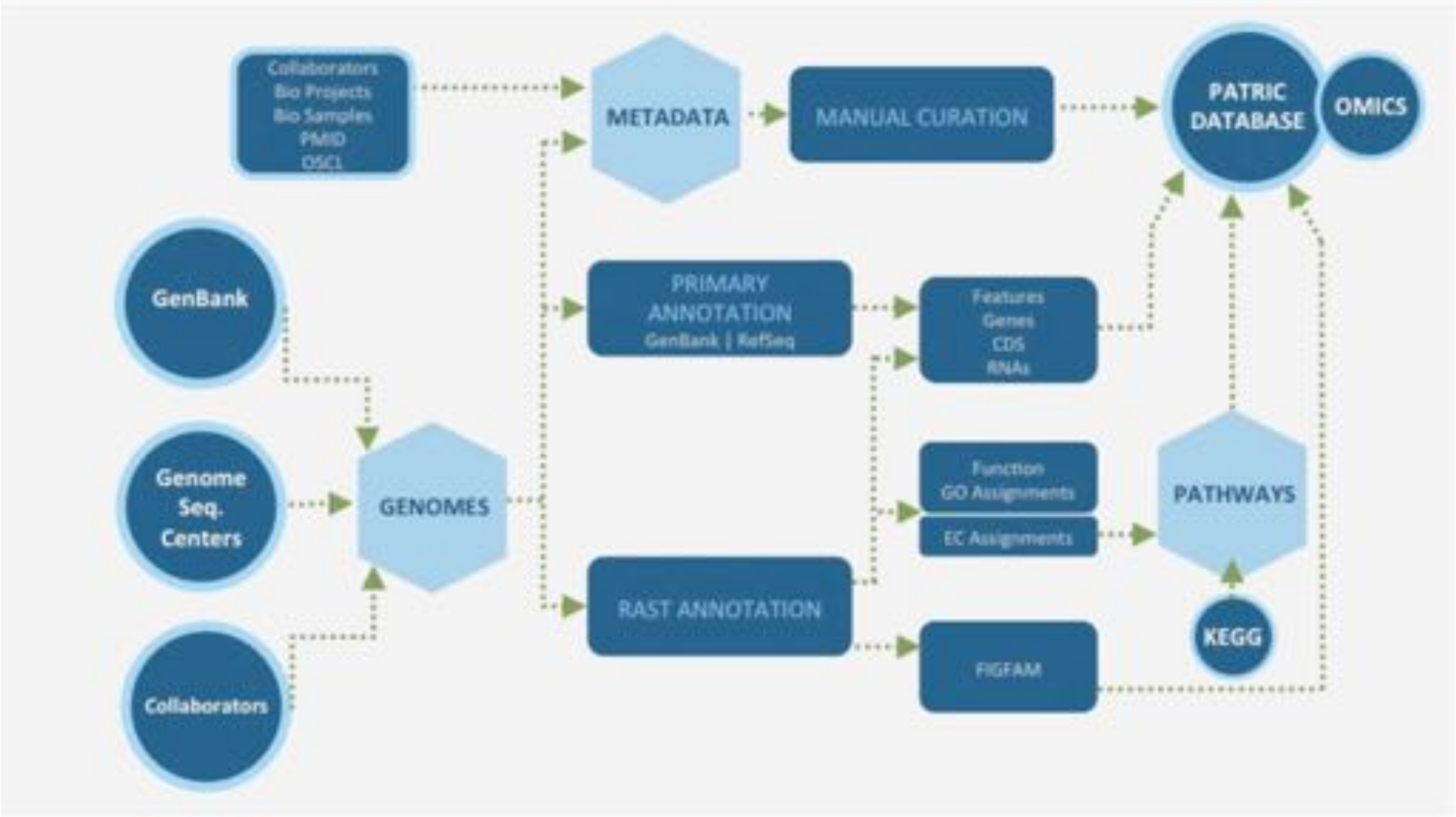
PATRIC has ALL Bacterial Genomes, not just pathogens

How many genomes does PATRIC have?



PATRIC

PATRIC data processing



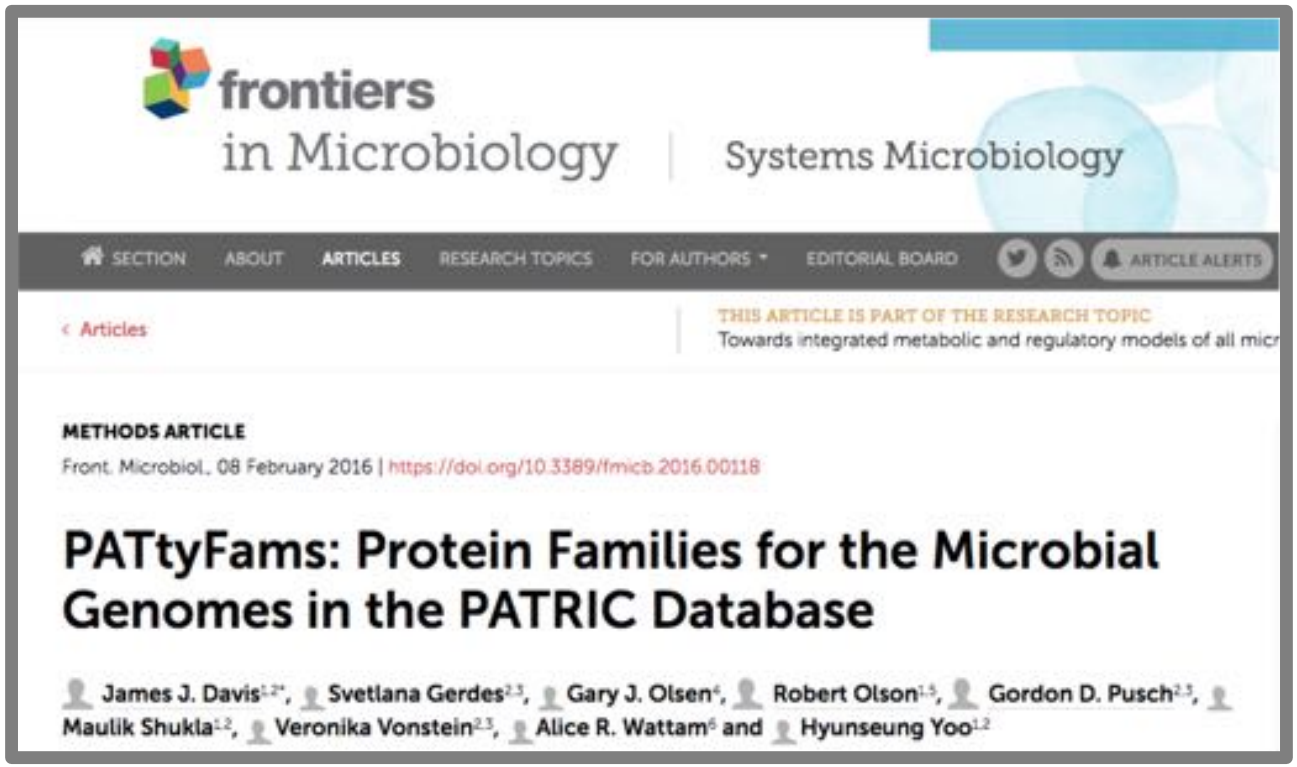
Uniform annotations across all genomes in PATRIC

RAST publications have more than 5,000 citations

Some Unique PATRIC Features

- **Comprehensive Data Collection**
  - Unified Database, including RefSeq, GenBank, other sources
- **Uniform Annotation Across all Genomes**
  - RAST annotation, EC, GO, plus RefSeq annotations
  - Uniform projection of Protein Families, AMR related genes and Virulence factors
- **User Workspace for analysis of User data**
  - "Virtual Integration" your data in the context of all the public datasets

Protein family assignments enable analysis



frontiers in Microbiology | Systems Microbiology

SECTION ABOUT ARTICLES RESEARCH TOPICS FOR AUTHORS EDITORIAL BOARD ARTICLE ALERTS

‹ Articles

THIS ARTICLE IS PART OF THE RESEARCH TOPIC
Towards integrated metabolic and regulatory models of all micr

METHODS ARTICLE

Front. Microbiol., 08 February 2016 | https://doi.org/10.3389/fmicb.2016.00118

PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database

James J. Davis[1,2*], Svetlana Gerdes[2,3], Gary J. Olsen[4], Robert Olson[1,5], Gordon D. Pusch[2,3], Maulik Shukla[1,2], Veronika Vonstein[2,3], Alice R. Wattam[6] and Hyunseung Yoo[1,2]

PATRIC

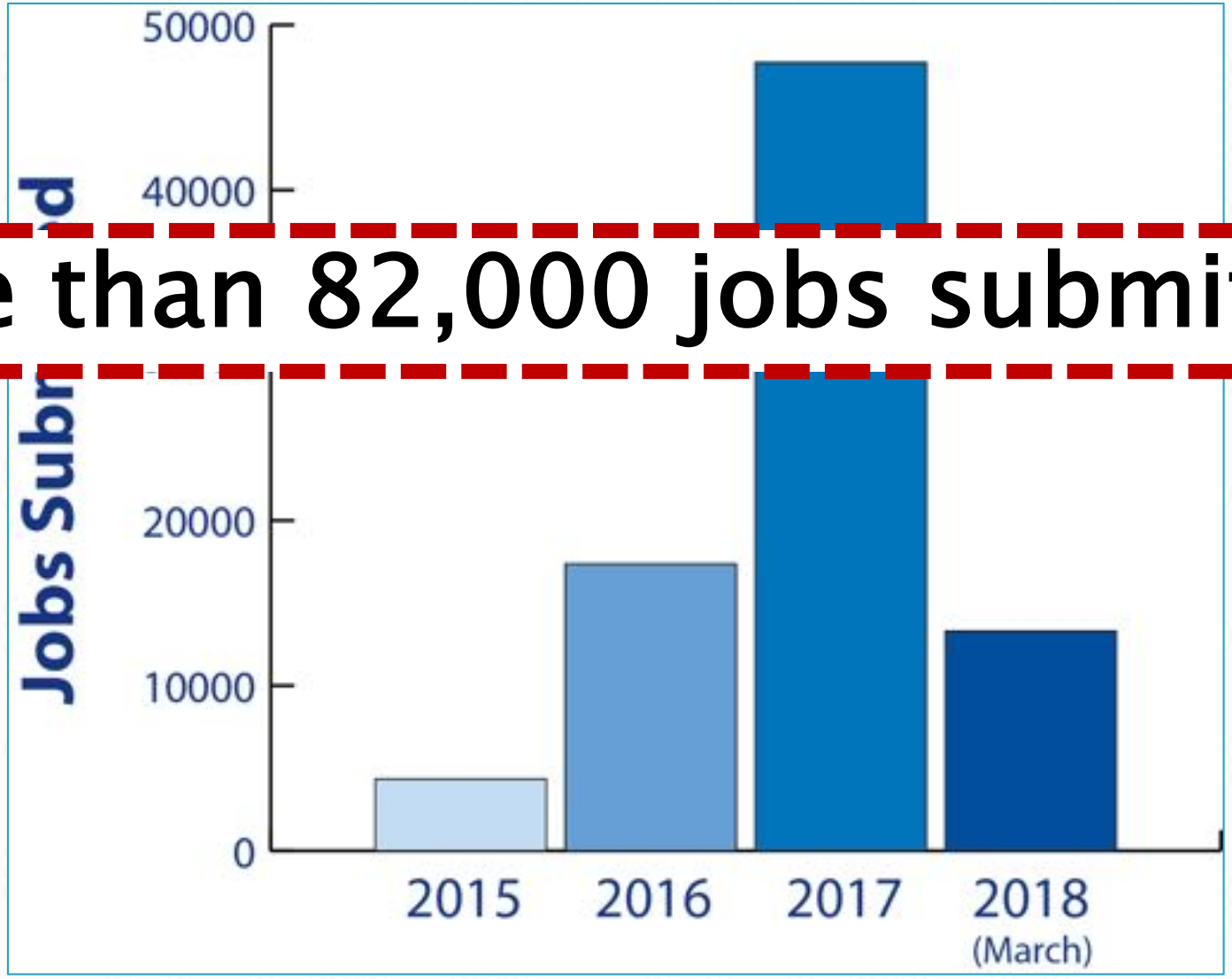
PATRIC Services

- Assembly – 2015
- Annotation – 2015
- Differential Expression – 2015

BYOD: Bring Your Own Data
and analyze it in PATRIC

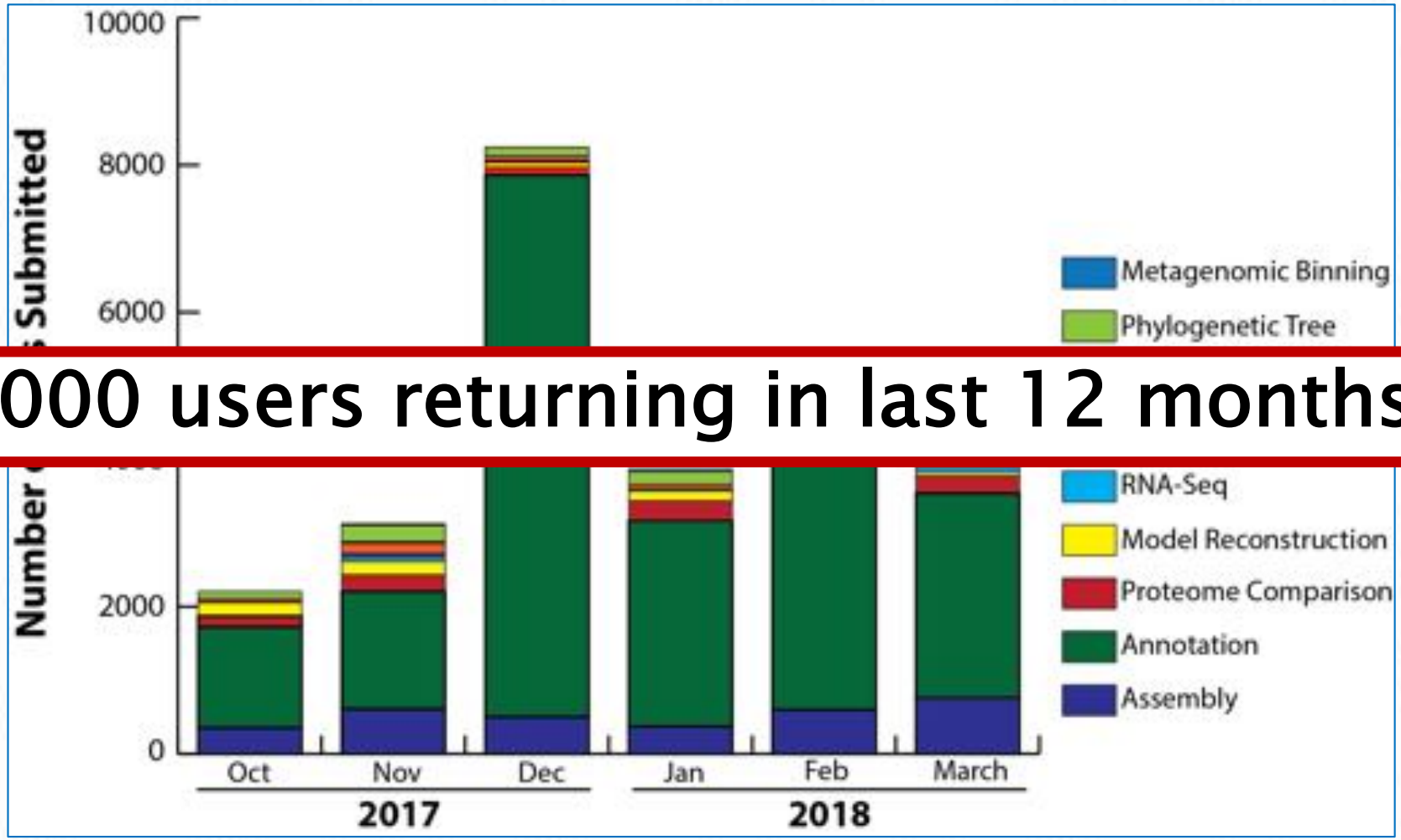
- Proteome Comparison – 2015
- RNA-Seq – 2015
- Transposon-Seq – 2017
- Variation – 2016

Does anybody use these services?



More than 82,000 jobs submitted

What services are they using most?



14,000 users returning in last 12 months

How can researchers find data of interest?

PATRIC

Where can one find metadata?

Locate in publications–

LABOR INTENSIVE

Articles

Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*

Makoto Kuroda, Toshiko Ohta, Ikuo Uchiyama, Tadashi Baba, Harumi Yuzawa, Ichizo Kobayashi, Longzhu Cui, Akio Oguchi, Ken-ichi Aoki, Yoshimi Nagai, JianQi Lian, Teruyo Ito, Mutsumi Kanamori, Hiroyuki Matsumaru, Atsushi Maruyama, Hiroyuki Murakami, Akira Hosoyama, Yoko Mizutani-Ui, Noriko K Takahashi, Toshihiko Sawano, Ryu-ichi Inoue, Chikara Kaito, Kazuhisa Sekimizu, Hideki Hirakawa, Satoru Kuhara, Susumu Goto, Junko Yabuzaki, Minoru Kanehisa, Atsushi Yamashita, Kenshiro Oshima, Keiko Furuya, Chie Yoshino, Tadayoshi Shiba, Masahira Hattori, Naotake Ogasawara, Hideo Hayashi, Kelichi Hiramatsu

Summary

Background *Staphylococcus aureus* is one of the major causes of community-acquired and hospital-acquired infections. It produces numerous toxins including superantigens that cause unique disease entities such as toxic-shock syndrome and staphylococcal scarlet fever, and has acquired resistance to practically all antibiotics. Whole genome analysis is a necessary step towards future development of countermeasures against this organism.

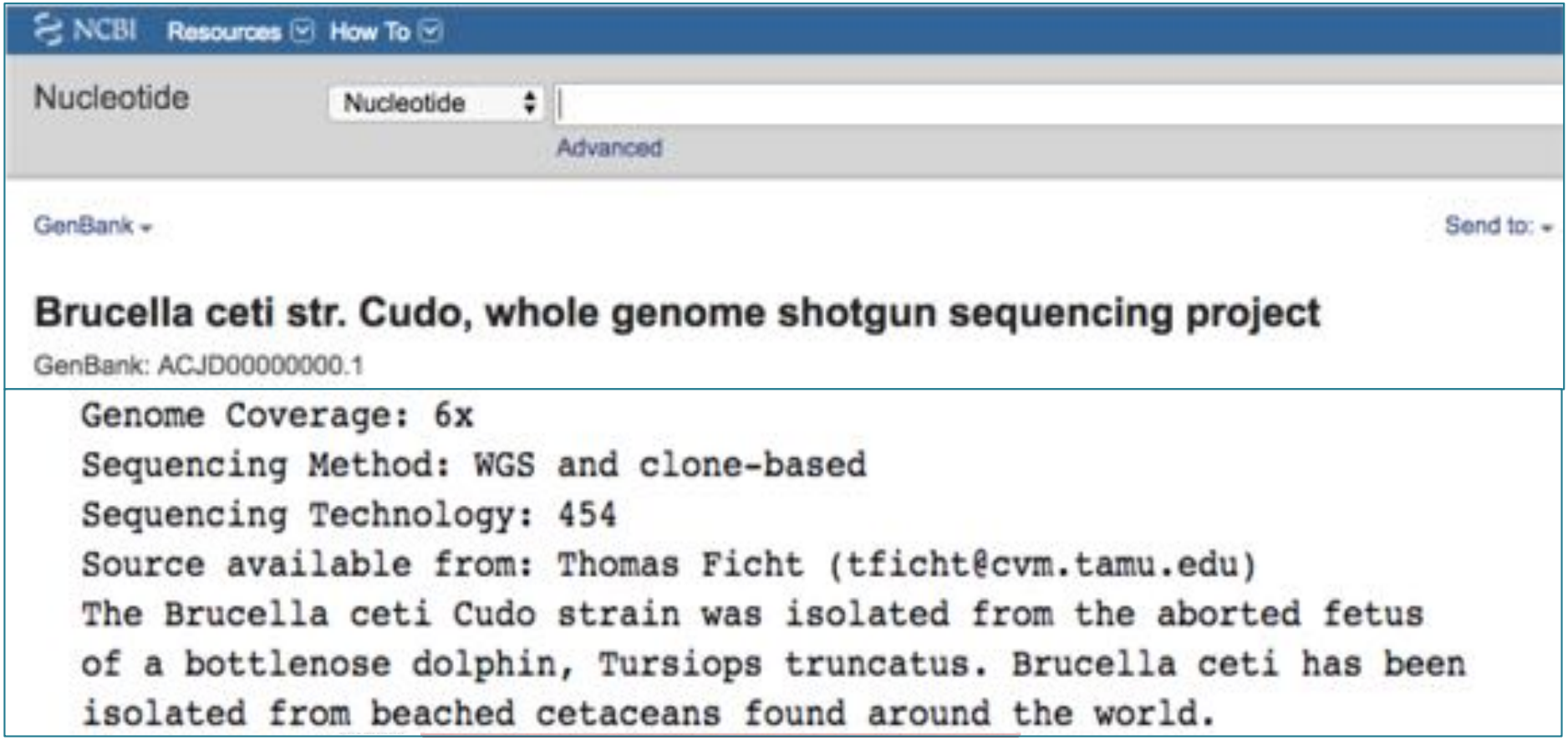
Methods Whole genome sequences of two related *S aureus* strains (N315 and Mu50) were determined by shot-gun random sequencing. N315 is a meticillin-resistant *S aureus* (MRSA) strain isolated in 1982, and Mu50 is an MRSA strain with vancomycin resistance isolated in 1997. The open reading frames were identified by use of GAMBLER and GLIMMER programs, and annotation of each was done with a BLAST homology search, motif analysis, and protein localisation prediction.

Findings The *Staphylococcus* genome was composed of a complex mixture of genes, many of which seem to have been acquired by lateral gene transfer. Most of the antibiotic resistance genes were carried either by plasmids or by mobile genetic elements including a unique resistance island. Three classes of new pathogenicity islands were identified in the genome: a toxic-shock-syndrome toxin island family, exotoxin islands, and enterotoxin islands. In the latter two pathogenicity islands, clusters of exotoxin and enterotoxin genes were found closely linked with other gene clusters encoding putative pathogenic factors. The analysis also identified 70 candidates for new virulence factors.

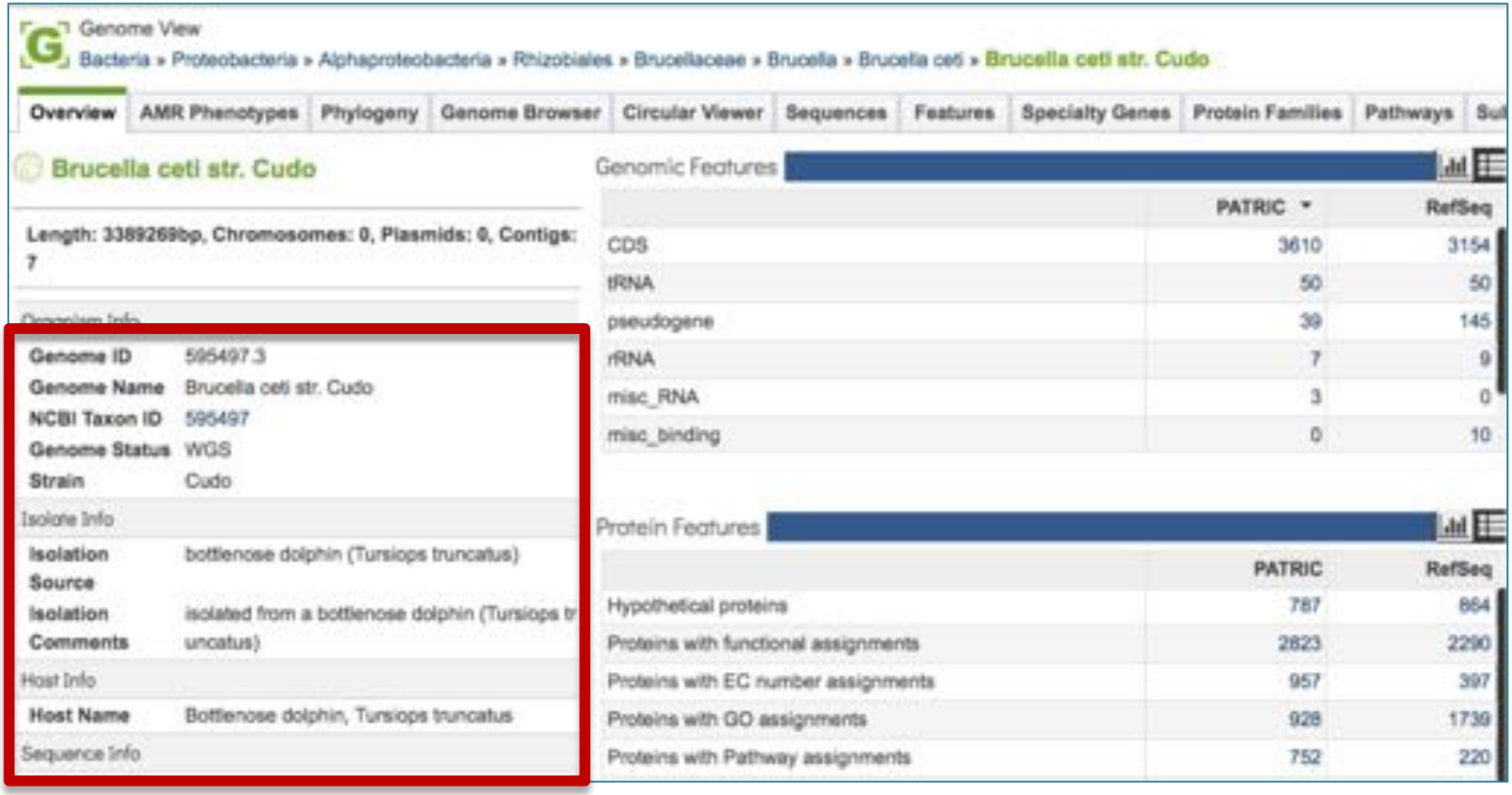
Interpretation The remarkable ability of *S aureus* to acquire useful genes from various organisms was revealed through the observation of genome complexity and evidence of lateral gene transfer. Repeated duplication of genes encoding superantigens explains why *S aureus* is capable of infecting humans of diverse genetic backgrounds, eliciting severe immune reactions. Investigation of many newly identified gene products, including the 70 putative

The entire genome sequences of *S aureus* N315 and Mu50 have been

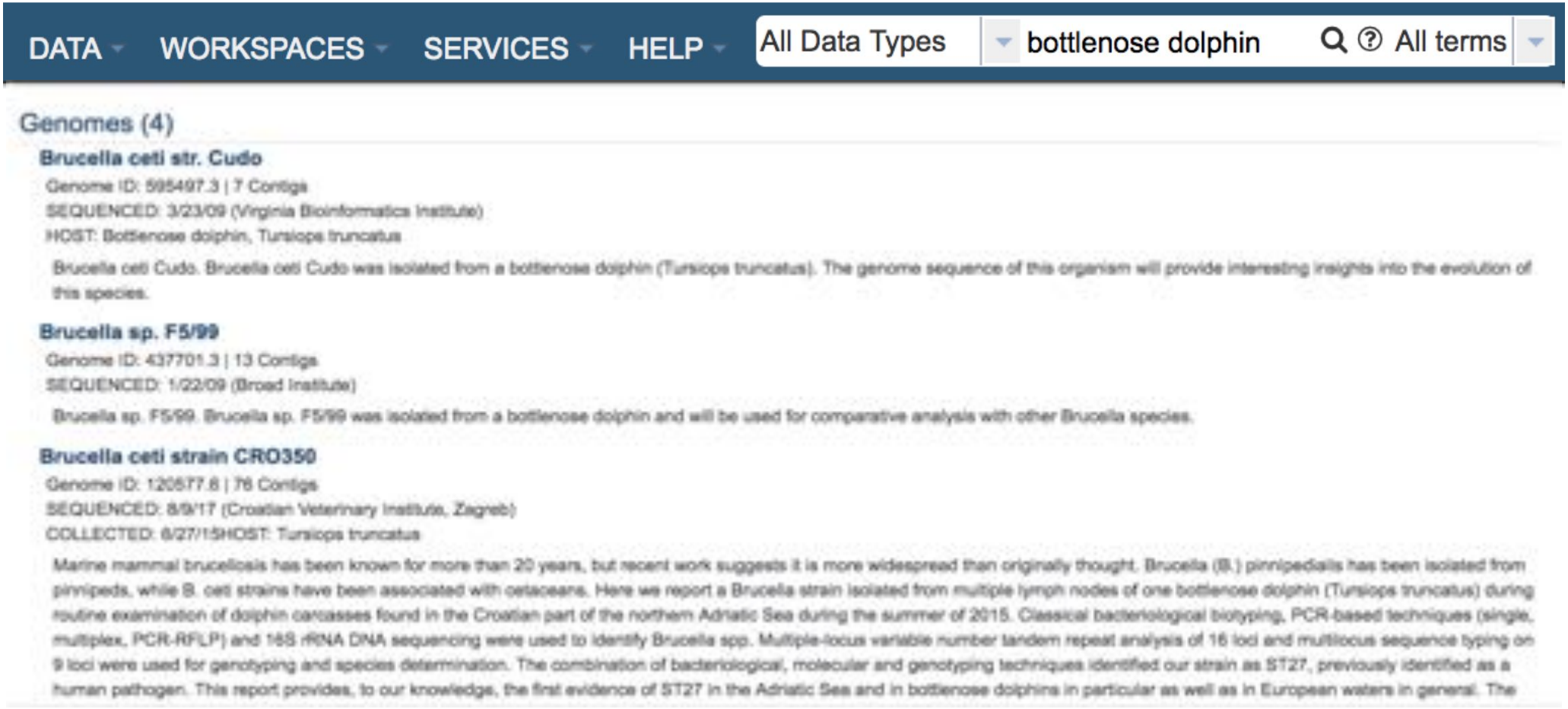
Another metadata location



What PATRIC does with that metadata

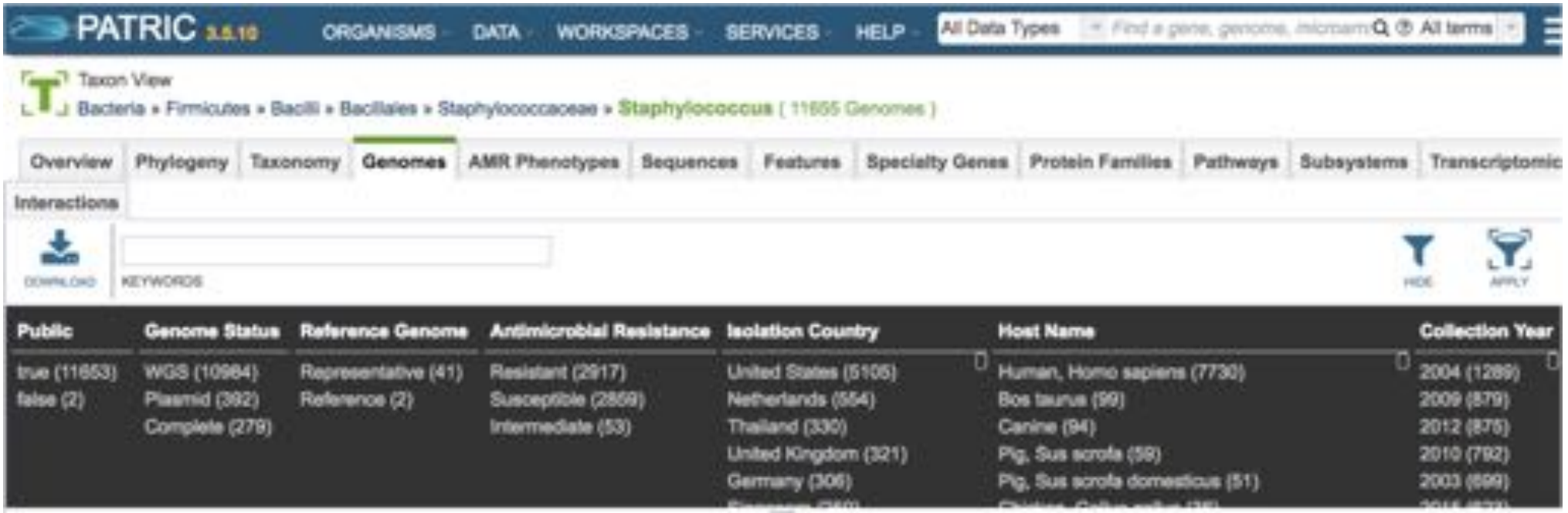


Make the metadata searchable!



PATRIC

Metadata filtering



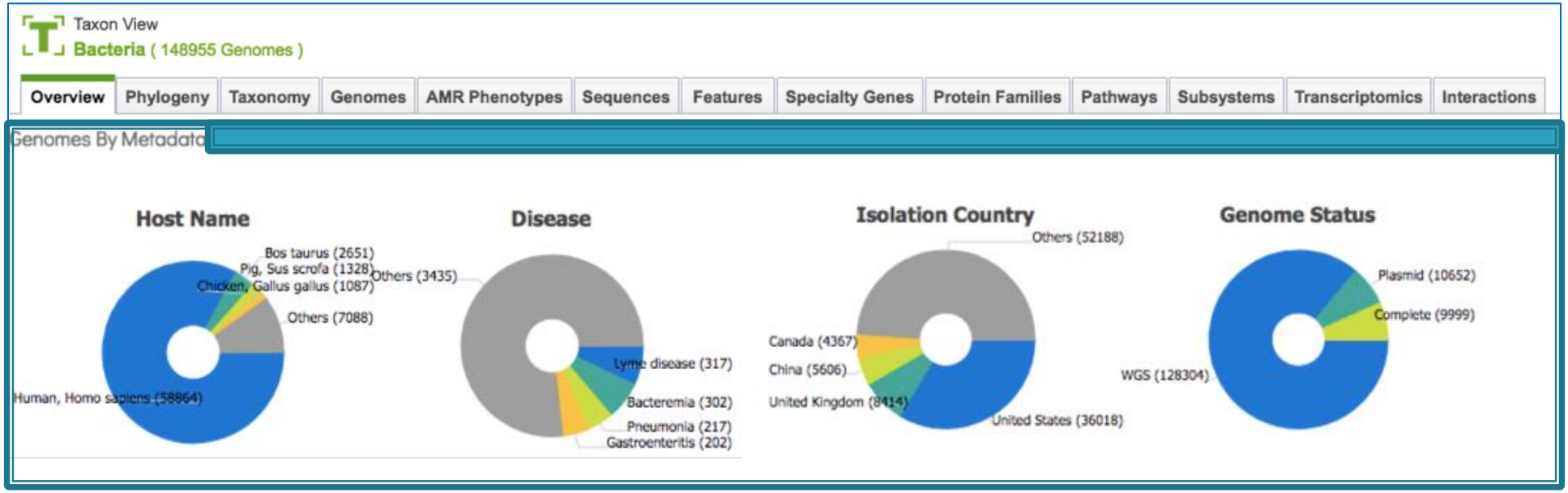
Metadata filtering

Genome Name	Genome ID	Genome Status	Sequences	PATRIC CDS	Isolation Country	Host Name	Collection Year	Completion Date	Isolation Site
Staphylococcus aureus P110256-141	1280.4912	WGS	401	3042	United States	Human, Homo sapiens	2010		Blood
Staphylococcus aureus P110270-142	1280.4913	WGS	33	2596	United States	Human, Homo sapiens	2010		Blood
Staphylococcus aureus P330170-177	1280.4993	WGS	58	2699	United States	Human, Homo sapiens	2011		Blood
Staphylococcus aureus P330177-181	1280.4994	WGS	40	2504	United States	Human, Homo sapiens	2011		Blood
Staphylococcus aureus P210439-24	1280.4954	WGS	59	2844	United States	Human, Homo sapiens	2011		Blood
Staphylococcus aureus P210500-225	1280.4965	WGS	33	2784	United States	Human, Homo sapiens	2014		Blood
Staphylococcus aureus P310372-198	1280.4970	WGS	29	2783	United States	Human, Homo sapiens	2011		Blood
Staphylococcus aureus P310516-204	1280.4972	WGS	25	2676	United States	Human, Homo sapiens	2011		Blood
Staphylococcus aureus P310516-223	1280.4974	WGS	25	2661	United States	Human, Homo sapiens	2011		Blood
Staphylococcus aureus P310821-173	1280.4978	WGS	26	2683	United States	Human, Homo sapiens	2011		Blood
Staphylococcus aureus P310906-176	1280.4982	WGS	29	2768	United States	Human, Homo sapiens	2011		Blood

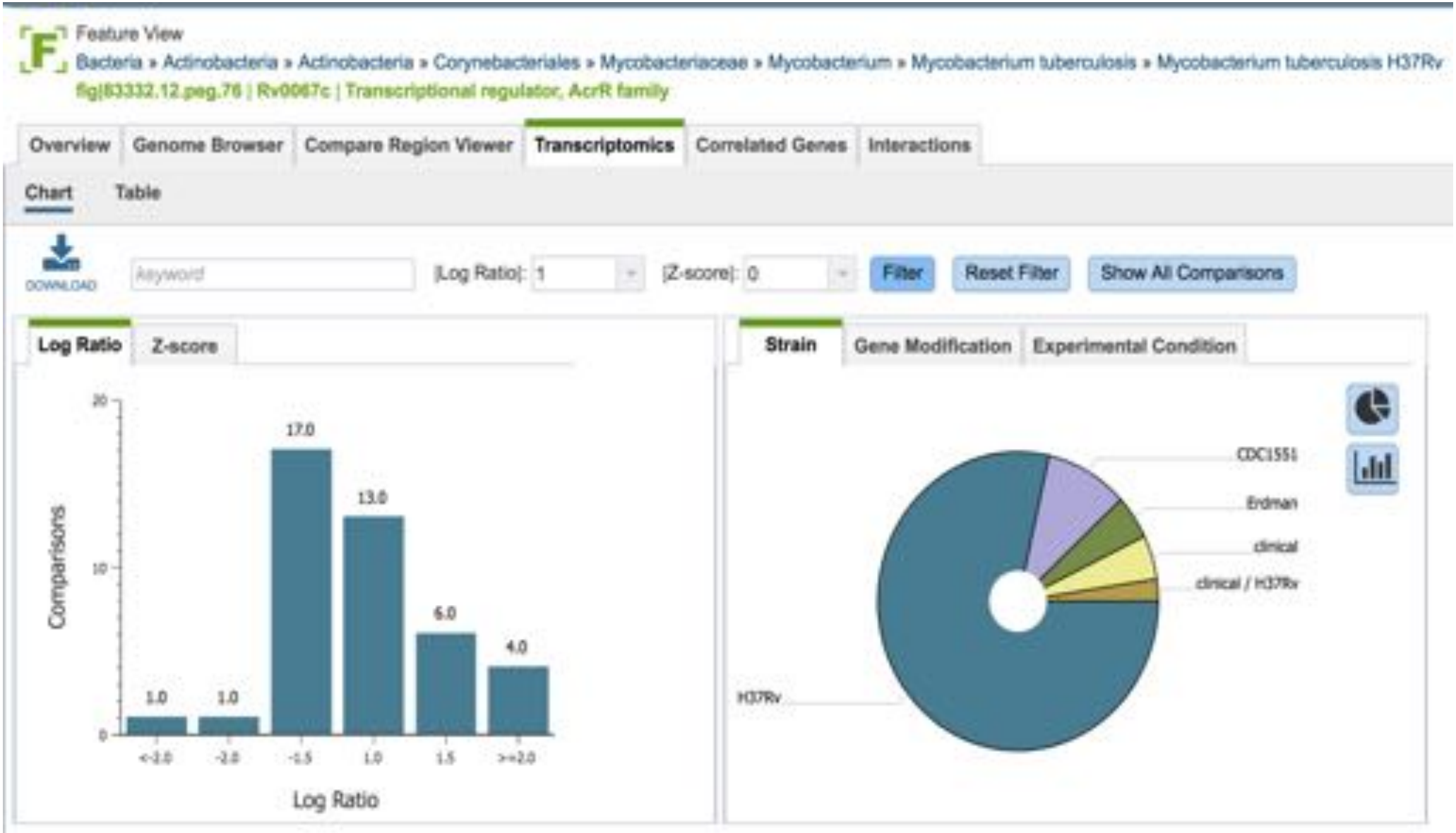
Large scale summaries based on metadata – AMR

PATRIC

Large scale summaries based on metadata

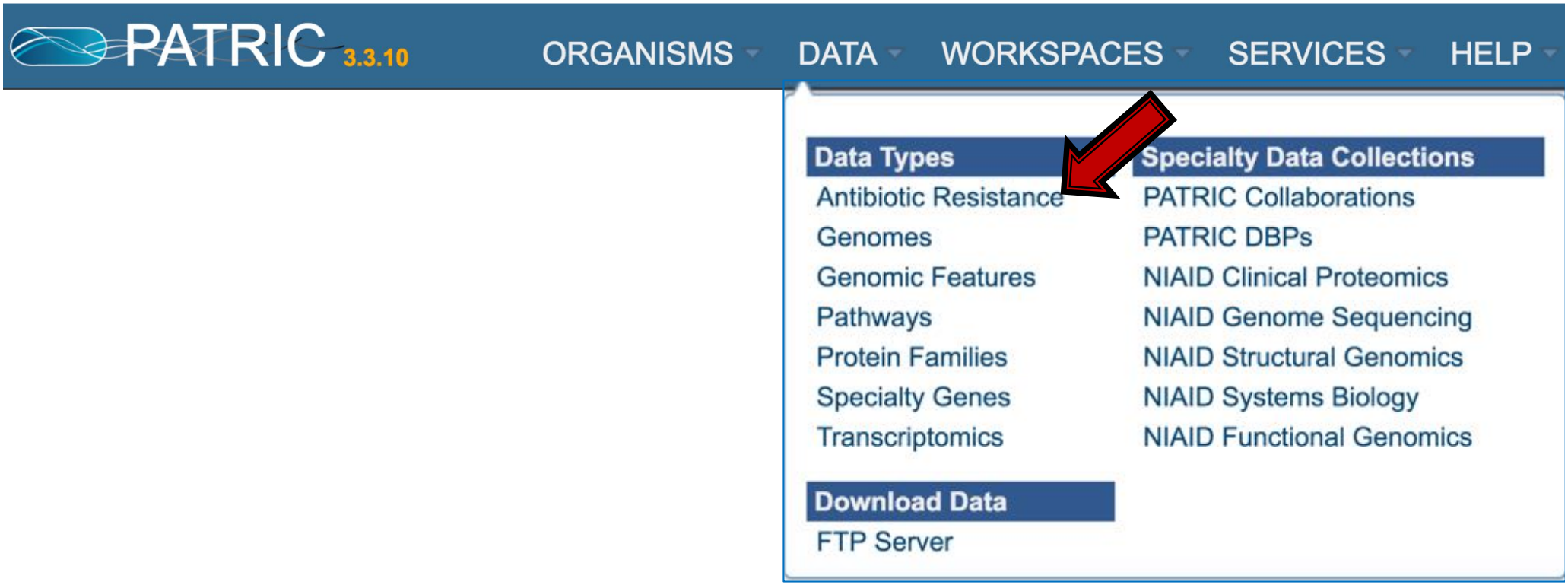


How data is summarized at PATRIC: Consistencies across levels



PATRIC

New Antimicrobial Data



Data

Antimicrobial Resistance (AMR)

Antibiotics are a type of drugs used in the treatment and prevention of bacterial infections. Antimicrobial Resistance (AMR) refers to the ability of bacteria to resist the effects of antibiotics that are commonly used to treat them. Resistance arises through one of three ways: natural resistance in certain types of bacteria, genetic mutation, or by one species acquiring resistance from another. PATRIC provides a variety of data and analysis tools to help researchers study AMR and its genetic determinants. This includes AMR phenotype data for the bacterial genomes as well as genes and intergenic regions associated with AMR.

What do we mean by ...

Antibiotics:

Antibiotics are a type of antimicrobial drugs used in the treatment and prevention of bacterial infections. PATRIC provides basic information about commonly used antibiotics, inlcuding their chemical and physical properties, pharmacology, and mechanism of action. In addition, each antibiotic is linked to other relevant data available in PATRIC, such as AMR phenotypes for genomes, AMR genes, and AMR regions. Below are some examples:

- amikacin
- ethambutol
- isoniazid
- rifampoin
- streptomycin

View all antibiotics

AMR Phenotypes:

AMR phenotypes refer to the resistance or susceptibility of a given organism to one or more antibiotics. PATRIC collects AMR phenotype data generated using antimicrobial susceptibility testing methods (AST) from published studies and collaborators. In addition, we also provide predicted AMR phenotypes using machine learning classifiers. See AMR phenotype data select genera:

- Mycobacterium
- Staphylococcus
- Streptococcus
- Acinetobacter
- Pseudomonas

View all AMR phenotype data

AMR Genes:

AMR gene refere to the genes impicated in or associated with the resistance to one or more antibiotics. The resistance may result from

AMR Regions:

AMR regions refer to the small genomic regions implicated in or associated with the resistance to one or more antibiotics. The AMR

Today's schedule (May 7)

- 9:00 am Register for PATRIC Account, Overview
- 10:00 am Assemble a Genome in PATRIC and Data Upload
- 11:00 am Break
- 11:10 am Annotate a Genome in PATRIC Using RASTtk
- 12:00 pm Lunch
- 1:00 pm Similar Genome Finder
- 1:30 pm Build a Phylogenetic tree
- 2:15 pm Break
- 2:30 pm Comparative Genome Analysis
- 3:00 pm Proteome Comparison
- 3:30 pm Comparative Genomics (Proteins and Pathways)
- 4:30 pm Question and Answer session
- 5:00 pm Adjourn

Schedule (May 8)

- 9:00 am BLAST at PATRIC
- 9:30 am RNA-Seq Pipeline
- 10:00 am Break
- 10:15 am Expression Import Service
- 10:45 am Comparative Transcriptomics
- 12:00 pm Lunch
- 1:00 pm SNP and MNP Variation Service
- 2:00 pm Metagenomic binning service
- 3:00 pm Building a metabolic model
- 4:00 pm Question and Answer Session
- 5:00 pm Adjourn

Schedule (May 9)

- 9:00 am Command Line Interface
- 11:00 am Break
- 11:15 am Work with Private Data
- 12:00 pm Lunch
- 1:00 pm Work with Private Data
- 3:00 pm Question and Answer Session
- 4:00 pm Workshop concludes

Finding help after the workshop

Tutorials

Step-By-Step PDF Tutorials

Workshop Guide: Genome Assembly
This document provides step-by-step instructions for submitting an assembly job and examining the results in PATRIC.

Workshop Guide: Genome Annotation
This document provides step-by-step instructions for submitting an annotation job and examining the results in PATRIC.

Workshop Guide: Proteome Comparison
This document provides step-by-step instructions for doing a bi-directional BLASTP analysis comparing up to 10 genomes in PATRIC.

Workshop Guide: Protein Family Sorter
This document provides step-by-step instructions for analyzing the proteomes of genome(s) in PATRIC.

Workshop Guide: Genome Groups
This document provides step-by-step instructions for creating genome groups that will be used for downstream analysis.

Workshop Guide: Private Genome
This document demonstrates finding and examining data associated with a private genome that has been annotated in PATRIC.

Workshop Guide: RNA-Seq Data Submission
This document provides step-by-step instructions for submitting a RNA-Seq job and examining the results in PATRIC.

Workshop Guide: Expression Import
This document provides step-by-step instructions for loading expression data into PATRIC for downstream analysis.

Workshop Guide: Transcriptomics Page
This document provides step-by-step instructions analyzing transcriptomic data that is available in PATRIC.

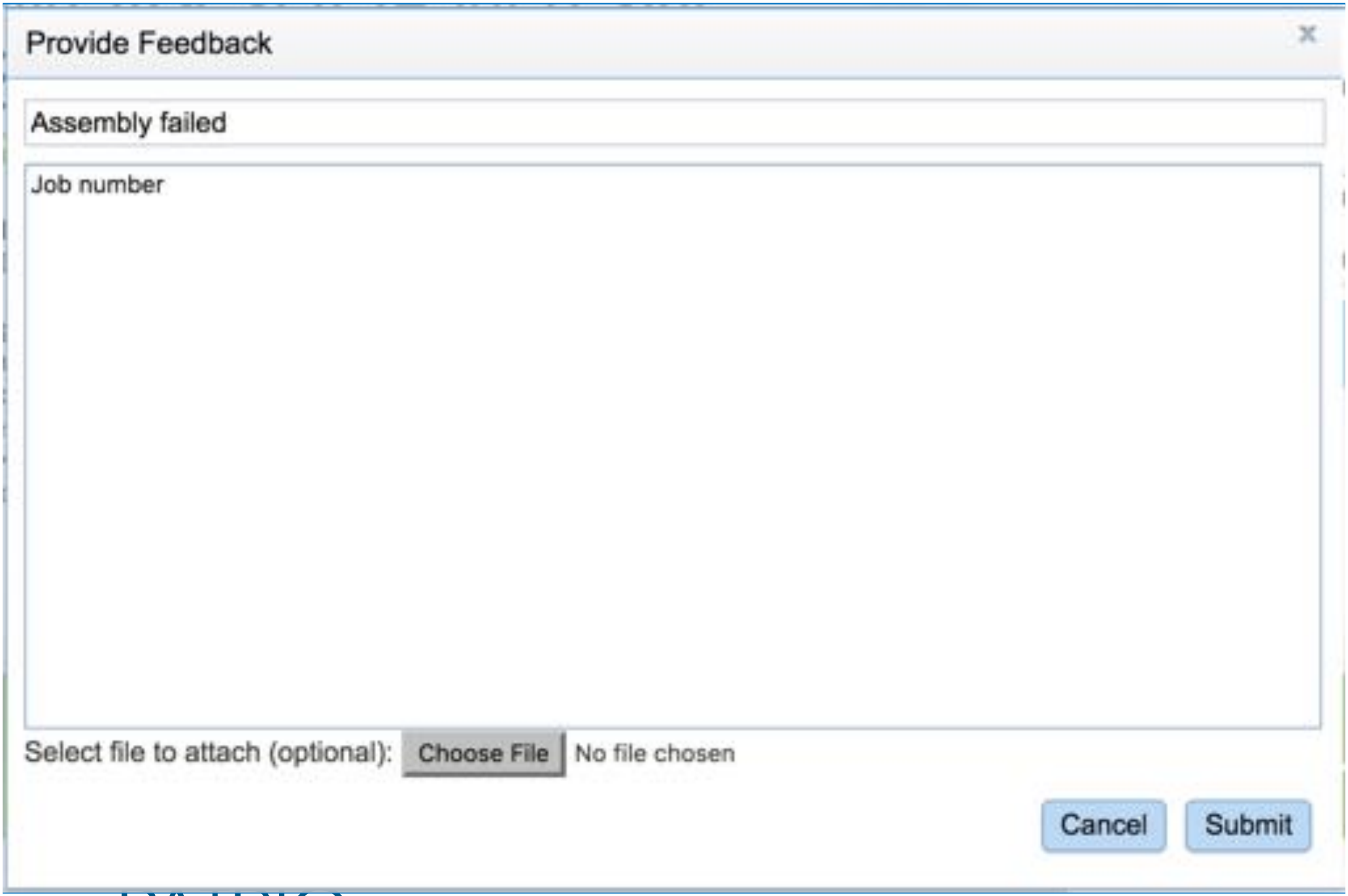
Using the PATRIC Metagenomic Binning Service

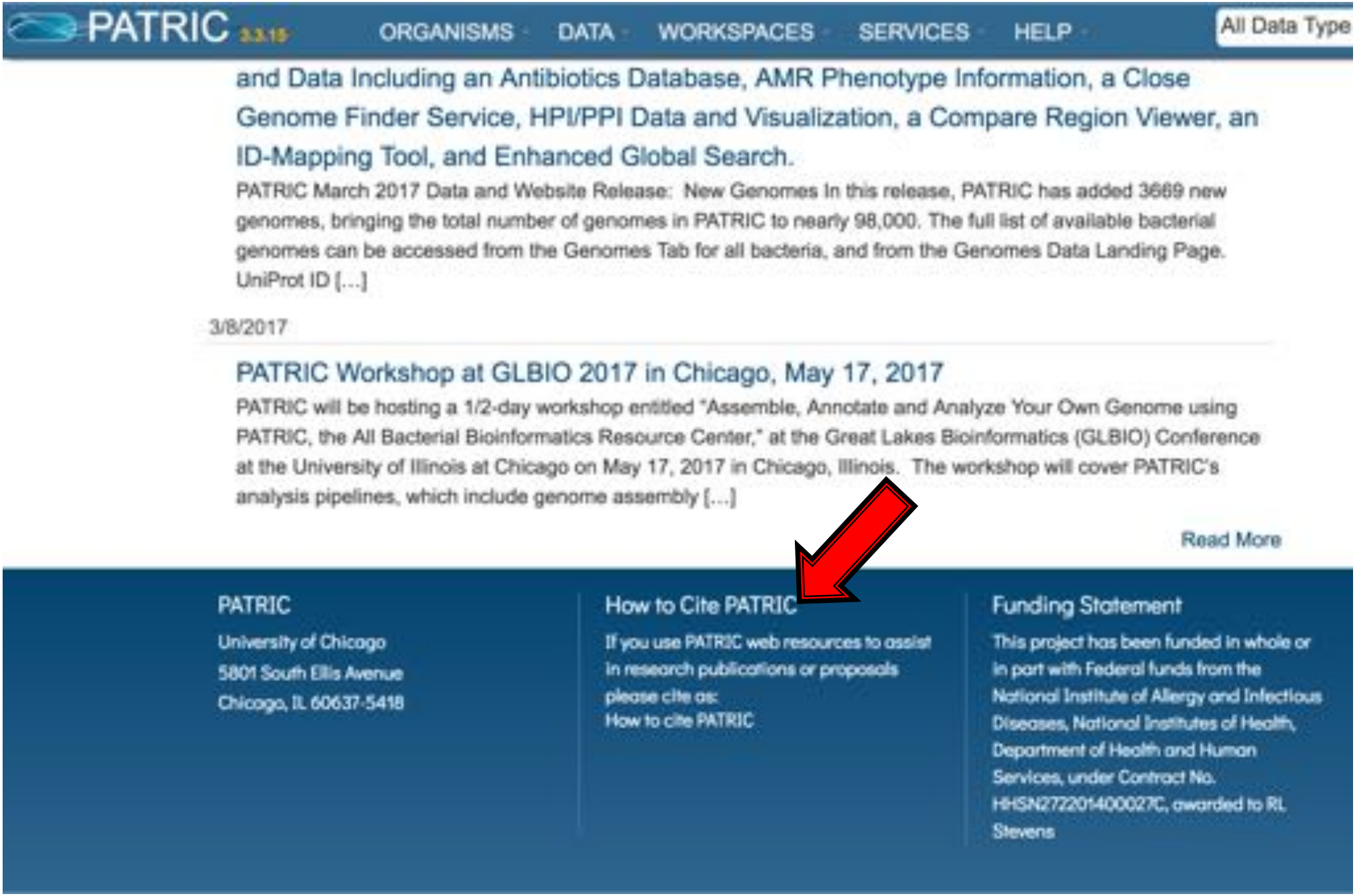
Basic Steps

1. Log in to the Patric web site with your Patric credentials.
2. Provide an input file.
3. Open the Binning Service
4. Run Binning
5. Examine Output

Log in to the PATRIC Website

See Registration for information on logging in to the PATRIC website. Once you are registered and logged in, you should see something like this:


Provide Feedback
Assembly failed
Job number
Select file to attach (optional):
Choose File
No file chosen
Cancel
Submit


PATRIC 3.3.19
ORGANISMS
DATA
WORKSPACES
SERVICES
HELP
All Data Type
and Data Including an Antibiotics Database, AMR Phenotype Information, a Close Genome Finder Service, HPI/PPI Data and Visualization, a Compare Region Viewer, an ID-Mapping Tool, and Enhanced Global Search.
PATRIC March 2017 Data and Website Release: New Genomes In this release, PATRIC has added 3669 new genomes, bringing the total number of genomes in PATRIC to nearly 98,000. The full list of available bacterial genomes can be accessed from the Genomes Tab for all bacteria, and from the Genomes Data Landing Page. UniProt ID [...]
3/8/2017
PATRIC Workshop at GLBIO 2017 in Chicago, May 17, 2017
PATRIC will be hosting a 1/2-day workshop entitled "Assemble, Annotate and Analyze Your Own Genome using PATRIC, the All Bacterial Bioinformatics Resource Center," at the Great Lakes Bioinformatics (GLBIO) Conference at the University of Illinois at Chicago on May 17, 2017 in Chicago, Illinois. The workshop will cover PATRIC's analysis pipelines, which include genome assembly [...]
Read More
PATRIC
University of Chicago
5801 South Ellis Avenue
Chicago, IL 60637-5418
How to Cite PATRIC
If you use PATRIC web resources to assist in research publications or proposals please cite as:
How to cite PATRIC
Funding Statement
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. HHSN272201400027C, awarded to RL Stevens

Citing PATRIC

If you use PATRIC web resources to assist in research publications or proposals please cite as:

Alice R. Wattam, James J. Davis, Rida Assaf, Sébastien Boisvert, Thomas Brettin, Christopher Bun, Neal Conrad, Emily M. Dietrich, Terry Disz, Joseph L. Gabbard, Svetlana Gerdes, Christopher S. Henry, Ronald W. Kenyon, Dustin Machi, Chunhong Mao, Eric K. Nordberg, Gary J. Olsen, Daniel E. Murphy-Olson, Robert Olson, Ross Overbeek, Bruce Parrello, Gordon D. Pusch, Maulik Shukla, Veronika Vonstein, Andrew Warren, Fangfang Xia, Hyunseung Yoo, Rick L. Stevens; **"Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center."** *Nucleic Acids Res* 2017; 45 (D1): D535-D542. doi: 10.1093/nar/gkw1017 PMID: 27899627

Thanks for coming

On to assembly….


PATRIC


PATRIC

Geographic distribution of the *mcr-1* gene (as of 1st March 2016)

Skov and Monnet, Eurosurveillance, Volume 21, Issue 9, 03 March 2016

Countries shown in colour have reported at least one isolate with the *mcr-1* gene [1-30].