PATRIC Bioinformatics Resource Center

Argonne National Lab
Lemont, Illinois
May 7–9, 2018
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 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

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

James J. Davis1,2, Svetlana Gerdes1,3, Gary J. Olsen4, Robert Olson2,5, Gordon D. Pusch2,3, Maulik Shukla2, Veronika Vonstein1,3, Alice R. Wattam2 and Hyunseung Yoo2,5
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?
Where can one find metadata?

Locate in publications—LABOR INTENSIVE

Whole genome sequencing of meticillin-resistant Staphylococcus aureus

Makoto Kuroda, Toshiko Ohta, Ikuo Uchiyama, Tadashi Baba, Harumi Yuzawa, Ichizo Kobayashi, Longzhu Cui,
Akihisa Oguchi, Ken-ichi Aoki, Yoshimi Nagai, JianQi Lian, Teruyo Ito, Mutsumi Kanamori, Hiroyuki Matsumaru,
Atsushi Maruyama, Hiroyuki Murakami, Akira Hosoyama, Yoko Mizutani-Ut, Noriko K Takahashi, Toshihiko Sawano,
Ryu-ichi Inoue, Chikara Kaito, Kazuhide Sekimizu, Hideki Hirakawa, Satoru Kuhara, Susumu Goto, Junko Yabuzaki,
Minoru Kanehisa, Atsushi Yamashita, Kenshiro Oshima, Keiko Furuya, Chie Yoshino, Tadayoshi Shiba, Masahara Hattori,
Naotake Ogasawara, Hideo Hayashi, Keiichi 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
Brucella ceti str. Cudo, whole genome shotgun sequencing project
GenBank: ACJD00000000.1

Genome Coverage: 6x
Sequencing Method: WGS and clone-based
Sequencing Technology: 454

Source available from: Thomas Ficht (tficht@cvm.tamu.edu)
The Brucella ceti Cudo strain was isolated from the aborted fetus of a bottlenose dolphin, Tursiops truncatus. Brucella ceti has been isolated from beached cetaceans found around the world.
What PATRIC does with that metadata
Make the metadata searchable!

**Brucella ceti str. Cudo**
- Genome ID: 595497.3 | 7 Contigs
- SEQUENCED: 3/23/09 (Virginia Bioinformatics Institute)
- HOST: Bottlenose dolphin, *Tursiops truncatus*

Brucella ceti Cudo. Brucella ceti Cudo was isolated from a bottlenose dolphin (*Tursiops truncatus*). The genome sequence of this organism will provide interesting insights into the evolution of this species.

**Brucella sp. F5/99**
- Genome ID: 437701.3 | 13 Contigs
- SEQUENCED: 1/22/09 (Broad Institute)

Brucella sp. F5/99. Brucella sp. F5/99 was isolated from a bottlenose dolphin and will be used for comparative analysis with other Brucella species.

**Brucella ceti strain CRO350**
- Genome ID: 120577.8 | 76 Contigs
- SEQUENCED: 8/9/17 (Croatian Veterinary Institute, Zagreb)
- COLLECTED: 6/27/15
- HOST: *Tursiops truncatus*

Marine mammal brucellosis has been known for more than 20 years, but recent work suggests it is more widespread than originally thought. Brucella (*B.*) pinnipedialis has been isolated from pinnipeds, while *B. ceti* strains have been associated with cetaceans. Here we report a Brucella strain isolated from multiple lymph nodes of one bottlenose dolphin (*Tursiops truncatus*) during routine examination of dolphin carcasses found in the Croatian part of the northern Adriatic Sea during the summer of 2015. Classical bacteriological biotypering, PCR-based techniques (single, multiplex, PCR-RFLP) and 16S rRNA DNA sequencing were used to identify Brucella spp. Multiple-locus variable number tandem repeat analysis of 16 loci and multilocus sequence typing on 9 loci were used for genotyping and species determination. The combination of bacteriological, molecular and genotyping techniques identified our strain as ST27, previously identified as a human pathogen. This report provides, to our knowledge, the first evidence of ST27 in the Adriatic Sea and in bottlenose dolphins in particular as well as in European waters in general.
Metadata filtering

### Table: Genomes

<table>
<thead>
<tr>
<th>Public</th>
<th>Genome Status</th>
<th>Reference Genome</th>
<th>Antimicrobial Resistance</th>
<th>Isolation Country</th>
<th>Host Name</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>true (11853)</td>
<td>WGS (10964)</td>
<td>Representative (41)</td>
<td>Resistant (2917)</td>
<td>United States (5105)</td>
<td>Human, Homo sapiens (7730)</td>
<td>2004 (1289)</td>
</tr>
<tr>
<td>false (2)</td>
<td>Plasmid (392)</td>
<td>Reference (2)</td>
<td>Susceptible (2859)</td>
<td>Netherlands (554)</td>
<td>Bos taurus (99)</td>
<td>2009 (879)</td>
</tr>
<tr>
<td></td>
<td>Complete (279)</td>
<td></td>
<td>Intermediate (53)</td>
<td>Thailand (330)</td>
<td>Canine (94)</td>
<td>2012 (875)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>United Kingdom (321)</td>
<td>Pig, Sus scrofa (99)</td>
<td>2010 (782)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Germany (306)</td>
<td>Pig, Sus scrofa domesticus (51)</td>
<td>2003 (699)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Singapore (360)</td>
<td>Chilidea, Ceratopithecus</td>
<td>2016 (872)</td>
</tr>
</tbody>
</table>
## Metadata filtering

<table>
<thead>
<tr>
<th>Genome Name</th>
<th>Genome ID</th>
<th>Genome Status</th>
<th>Sequences</th>
<th>PATRIC CDS</th>
<th>Isolation Country</th>
<th>Host Name</th>
<th>Collection Year</th>
<th>Completion Date</th>
<th>Isolation Site</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Staphylococcus aureus</em> P110256-141</td>
<td>1280.4912</td>
<td>WGS</td>
<td>401</td>
<td>3042</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2010</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P110276-142</td>
<td>1280.4913</td>
<td>WGS</td>
<td>33</td>
<td>2596</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2010</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P330170-177</td>
<td>1280.4993</td>
<td>WGS</td>
<td>58</td>
<td>2699</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P330177-181</td>
<td>1280.4994</td>
<td>WGS</td>
<td>40</td>
<td>2504</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P210439-24</td>
<td>1280.4954</td>
<td>WGS</td>
<td>58</td>
<td>2844</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P210500-225</td>
<td>1280.4965</td>
<td>WGS</td>
<td>33</td>
<td>2764</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2014</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P310372-198</td>
<td>1280.4970</td>
<td>WGS</td>
<td>29</td>
<td>2783</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P310516-204</td>
<td>1280.4972</td>
<td>WGS</td>
<td>25</td>
<td>2676</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P310516-223</td>
<td>1280.4974</td>
<td>WGS</td>
<td>25</td>
<td>2661</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P310821-173</td>
<td>1280.4976</td>
<td>WGS</td>
<td>26</td>
<td>2683</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> P310906-176</td>
<td>1280.4982</td>
<td>WGS</td>
<td>29</td>
<td>2768</td>
<td>United States</td>
<td>Human, Homo sapiens</td>
<td>2011</td>
<td></td>
<td>Blood</td>
</tr>
</tbody>
</table>
Large scale summaries based on metadata – AMR
Large scale summaries based on metadata
How data is summarized at PATRIC: Consistencies across levels
New Antimicrobial 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, including 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
- rifampcin
- 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 refer to the genes implicated 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: Protein Family Sorter
This document provides step-by-step instructions for analyzing the proteomes of genome(s) 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: Genome Groups
This document provides step-by-step instructions for creating genome groups that will be used for downstream analysis.

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

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

PATRIC
Using the PATRIC Metagenomic Binning Service

Basic Steps

1. Log in to the Patric website 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
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:

Thanks for coming
On to assembly....
Geographic distribution of the *mcr-1* gene (as of 1st March 2016)

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

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