

## Abstract

The PathoSystems Resource Integration Center (PATRIC) at the Virginia Bioinformatics Institute (VBI) is one of eight Bioinformatics Resource Centers (BRC) funded for 5 years by NIAID as part of its biodefense initiative. The BRCs will provide the scientific community with a web resource for research on biodefense and emerging or re-emerging diseases to facilitate the development of novel diagnostics, vaccine and therapeutics. As part of the BRC mission PATRIC will provide a comprehensive and centralized annotation resource for eight classes of microbial pathogens. It will present users with an integrated access to genomic, proteomic, biochemical information, microbiological information, experimental data and associated literature on each of the pathogen classes.

## Introduction

The PathoSystems Resource Integration Center (PATRIC) was established at the Virginia Bioinformatics Institute (VBI), on July 2004. It is one of the eight Bioinformatics Resource Centers (BRC) that have been established by the National Institute of Allergy and Infectious Disease (NIAID) with a mandate to integrate biological information on microbial pathogens, to aid research on disease detection, prevention, and treatments. The centers will focus on the NIAID/CDC category A-C pathogens (Fig. 1) and agents causing emerging or re-emerging disease that are relevant for biodefense. PATRIC will annotate and curate genomic and related biological data for eight pathogenic microorganism classes, which include three bacterial and five viral classes (Table 1). The website address is: <http://patric.vbi.vt.edu>

## PATRIC Personnel

The core PATRIC team consists of Curators with a strong background in infectious disease; Organism experts; Usability engineers, Bioinformaticians; Software Developers and a Database Analyst/Administrator.

## PathoSystems

At PATRIC, we define the pathogen-host-environment triangle as a PathoSystem (Fig2). The PathoSystem combines factors in the host, pathogen and environment (the "Disease Triangle") that, together are necessary and sufficient to specify the outcome of a host-pathogen encounter, which may range from microbe-clearance to full-blown disease. We believe that this holistic approach to studying a disease will yield valuable insights into the pathogenic mechanisms of these microbes. PATRIC will follow this "Pathosystems" approach in annotating and integrating biological data drawn from different information sources.

## PATRIC's Goals

In keeping with the BRC objectives, PATRIC will create a web-resource that provides access to genomic and associated data, and analysis tools on the selected pathogens and facilitate research in the areas of vaccines, diagnostics and therapeutics. We also welcome and expect the active participation of the research community in the development of this resource.

Category A	Category B
<ul style="list-style-type: none"> <li>Bacillus anthracis (anthrax)</li> <li>Clostridium botulinum</li> <li>Yersinia pestis</li> <li>Yersinia enterocolitica and other porcine</li> <li>Francisella tularensis subsp. tularensis</li> <li>Viral hemorrhagic fevers               <ul style="list-style-type: none"> <li>Arboviruses                   <ul style="list-style-type: none"> <li>LCMV, Junin virus, Machupo virus, Guanarito virus</li> </ul> </li> <li>West Nile Virus</li> <li>Lassa Fever</li> <li>Bunyaviruses                   <ul style="list-style-type: none"> <li>Hantaan virus</li> <li>Rift Valley Fever</li> </ul> </li> <li>Flaviviruses                   <ul style="list-style-type: none"> <li>Dengue</li> </ul> </li> <li>Rotaviruses                   <ul style="list-style-type: none"> <li>Ebola</li> <li>Marburg</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Burkholderia pseudomallei</li> <li>Coxiella burnetii (Q fever)</li> <li>Brevibacterium (brucellosis)</li> <li>Burkholderia mallei (landers)</li> <li>Ricin toxin (from Ricinus communis)</li> <li>Epitope toxin of Clostridium perfringens</li> <li>Staphylococcus enterotoxin B</li> <li>Toxinoid toxin (Bioterrorism potential)</li> <li>Food and Waterborne Pathogens               <ul style="list-style-type: none"> <li>Bacteria                   <ul style="list-style-type: none"> <li>Paratyphoidic E coli</li> <li>Paratyphoid Shiga</li> <li>Shigella species</li> <li>Salmonella</li> <li>Listeria monocytogenes</li> <li>Campylobacter jejuni</li> <li>Yersinia enterocolitica</li> </ul> </li> <li>Fungi                   <ul style="list-style-type: none"> <li>Cryptosporidium parvum</li> <li>Cyclospora cayentensis</li> <li>Giardia lamblia</li> <li>Enteromobis histolytica</li> </ul> </li> <li>Toxoplasma</li> </ul> </li> </ul>
Category C Emerging disease agents such as Mammal virus and additional hemorrhagic fevers.	Additional and emerging
<ul style="list-style-type: none"> <li>Typhoid hemorrhagic fever viruses               <ul style="list-style-type: none"> <li>Crimson Congo Hemorrhagic fever virus</li> </ul> </li> <li>Typhoid encephalitis viruses</li> <li>Yellow fever</li> <li>Multi-drug resistant TB</li> <li>Influenza</li> <li>Other Rickettsias</li> <li>Rabies</li> <li>Serum acute respiratory syndrome-associated coronavirus (SARS-CoV) (NIAID-CDC contract August 30, 2004)</li> </ul>	<ul style="list-style-type: none"> <li>West Nile Virus</li> <li>LCM virus</li> <li>California encephalitis</li> <li>EEE</li> <li>WEE</li> <li>Japanese Encephalitis Virus</li> <li>Kyasanur Forest Virus</li> </ul>

Figure 1: The list of pathogens in the Categories A, B and C as assigned by NIAID. ([http://www2.niaid.nih.gov/biodefense/bandc\\_priority.htm](http://www2.niaid.nih.gov/biodefense/bandc_priority.htm))

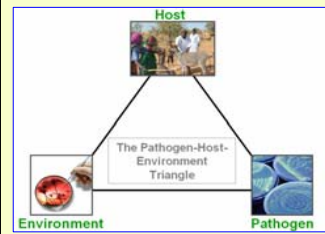


Figure 2. Diagram showing the PathoSystem, a pathogen-host-environment triangle.

Name	Size	Scope	Number of Genomes	Organism Expert (Affiliation)
<b>Bacteria</b>				
Brucellae	Ch1: 2.11Mb	Genus	4	Dr. Steven Boyle (Virginia Tech)
Coxiellae	Ch2: 1.17Mb	Species	4	Dr. Abdu Azad (University of Maryland)
Rickettsiae	1.1-1.2Mb	Genus	9	Dr. Abdu Azad (University of Maryland)
<b>Viruses</b>				
Caliciviruses	7.4-9.5Kb	Family	150	Dr. Stephen Monroe (CDC)
Coronaviruses	27-31Kb	Genus	52	Dr. Susan Baker (Loyola University of Chicago)
Hepatitis A	7.4Kb	Species	17	Dr. Omara Naiman (CDC)
Hepatitis E	7.1Kb	Species	37	Dr. X. J. Meng (Virginia Tech)
Rabies	11.9Kb	Species	9	Dr. Charles Rupprecht (CDC)

Table 1: The bacterial and viral species being curated by PATRIC. The three bacterial and five viral pathogen classes are listed along with the number of genome sequences (whole and partial) archived at PATRIC and the Organism Experts who will provide guidance on curation of each of the organisms. (\*Number of sequences as of July 15, 2005)

## Integration of Organism Information

PATRIC will provide integrated access to diverse sets of biological data and information, along with a set of analysis tools through its user-friendly website.

### Data

**Nucleic level features:** coding and non-coding features  
**Protein level features and attributes:** secondary structures, domains, motifs, signatures, parent families, attributes—EC/TC numbers, GO terms, COG assignments, functional categories.  
**Pathways:** metabolic and signaling pathways; assembly pathways.  
**Host-pathogen interaction:** effect of pathogen on host pathways  
**Post-sequence data:** high-throughput gene expression/proteomic data; protein-protein interactions  
**Associated data:** taxonomy; relevant literature.

### Tools

Single and multi-genome sequence analysis tools; experimental data analysis tools

## Annotation Process

An annotation software infrastructure is being built to automate the annotations at various levels (nucleic acid, protein, pathways); it will include modules for comparative genomics; literature curation and the analysis of pathogen-host-environment interactions.

A team of curators will manually review and edit the automated annotation. Curators will also be involved in working with members of the scientific community to ensure that their needs, in terms of data and tools, are met at PATRIC.

An organism expert (OE) has been assigned to each PathoSystem (see Table1). The OE is a recognized authority on his/her assigned pathogen and will be closely involved at each step of the annotation/curation process.

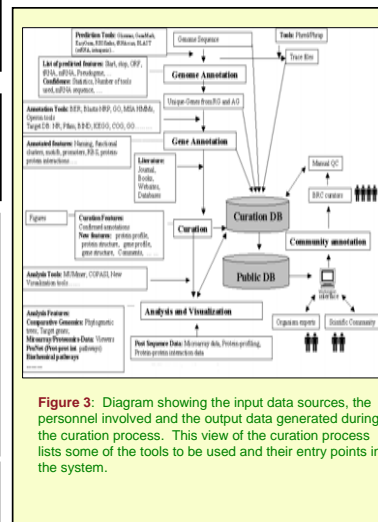


Figure 3: Diagram showing the input data sources, the personnel involved and the output data generated during the curation process. This view of the curation process lists some of the tools to be used and their entry points in the system.

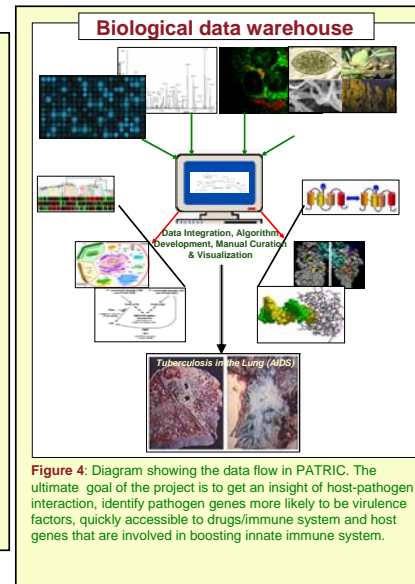


Figure 4: Diagram showing the data flow in PATRIC. The ultimate goal of the project is to get an insight of host-pathogen interaction, identify pathogen genes more likely to be virulence factors, quickly accessible to drugs/immune system and host genes that are involved in boosting innate immune system.

## Summary

Through a user-friendly web interface, PATRIC will provide high quality information and tools to the scientific communities that study *Brucellae*, *Coxiellae*, *Rickettsiae*, *Caliciviruses*, *Coronaviruses*, *Hepatitis A*, *Hepatitis E*, and *Rabies* viruses. Integrated manual genome / gene product curation, scientific literature, and a suite of analysis will align with NIAID's mission to facilitate development of vaccine, therapeutic, and diagnostic agents. The project intends to remain flexible enough so that it can other include data and services for which there is a significant demand in the drugs/vaccines/diagnostic research community.

## Acknowledgement

This project is funded by NIAID / NIH contract HHSN26620040035C to Bruno Sobral. Thanks June Mullins for poster graphic design and layout!