

## **The Pathogen Resource Integration Center: Implications for *Rickettsial* Research**

Mark E. Hance<sup>§</sup>, Anjan Purkayastha<sup>§</sup>, Eric E. Snyder<sup>§</sup>, Michael J. Czar<sup>§</sup>, Oswald R. Crasta<sup>§</sup>, Joao C. Setubal<sup>§</sup>, and Bruno W. Sobral<sup>§</sup>

### **Introduction**

The PathoSystems Resource Integration Center (PATRIC) was established in July, 2004, at the Virginia Bioinformatics Institute (VBI), to serve as a bioinformatics resource center (BRC) for scientists engaged in the development of vaccines, diagnostics, and therapeutics for three bacteria; *Rickettsiae*, *Coxiellae*, and *Brucellae*; and five classes of viruses; Calici, corona, Hepatitis A, Hepatitis E and rabies viruses. PATRIC is one of eight such Bioinformatics Resource Centers for Biodefense and Emerging or Re-Emerging Infectious Diseases (BRCs) funded for five years by the National Institute of Allergy and Infectious Diseases (NIAID). The objective of the BRCs is to create a comprehensive database of information on selected CDC category A, B and C pathogens and other organisms causing emerging and re-emerging human diseases. PATRIC will integrate and disseminate a variety of diverse data types such as genome sequence, comparative genomics, genome polymorphisms, gene expression, proteomics, pathways and host/pathogen interactions. This information will be stored in a relational database and made available to the scientific community through a user-friendly web interface. Herein, we will describe our mission goals as a participating BRC, detail the PATRIC infrastructure and our approach to biological data annotation/curation and present our strategy for developing resources for the *Rickettsial* scientific community as well as the other scientific communities served by PATRIC.

NIAID's goal is to better understand, treat, and ultimately prevent infectious, immunologic, and allergic diseases (<http://www3.niaid.nih.gov/about/overview/>). Within the past several years, NIAID has funded a number of initiatives to provide comprehensive genomics resources to the scientific community. NIAID has been investing in genomic sequencing of invertebrate disease vectors and pathogenic microorganisms including both those of primary medical significance as well as those considered to be relevant from to biodefense. The sequence data from these organisms coupled with other experimentally and computer-generated information are expediting the development of diagnostic and epidemiological markers, drug targets and vaccines.

Since the early 1980s molecular sequence data has undergone exponential growth, attributable largely to the availability of DNA sequencing methods. High throughput sequencing has become commonplace; high throughput gene expression data and proteomics data are rapidly following suit. Microbial genomes can now be almost completely sequenced within a few days (Sobral, *et al.*, unpublished data). NIAID recognized a need with respect to the integration and analysis of this data and established eight BRCs. Each center is tasked with constructing cyberinfrastructures in order to integrate genomic, proteomic, biochemical, and literature information. (<http://www.niaid.nih.gov/dmid/genomes/brc/default.htm>). This integration should provide the scientific community a valuable research resource, including access to genomic and related data pertaining to the organisms under each BRC's purview as well as bioinformatic tools to aid the researcher in the interpretation of the data. These resources will be freely available for public use. Specific objectives set forth by NIAID for the BRC initiative are as follows:

1. To provide the scientific community with a point of entry into a comprehensive database with access to genomic and related data designed to maximize ease of use.
2. Each center focuses on the organisms listed in the NIAID Category A-C priority list or pathogens causing emerging or re-emerging diseases relevant to biodefense.
3. To ultimately aid in the discovery of therapeutic targets, vaccines, diagnostics and epidemiological markers.

Making sense of this abundance of data requires a new paradigm: that of a bioinformatic cyberinfrastructure ([http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)). The goal of this new environment is the integration of a multiple data types using distributive database, computer, and information and communication technology. Ultimately, through this new paradigm, complete computer-

---

<sup>§</sup> Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Washington Street, Blacksburg, VA 24061-0477, USA., phone (540) 231-2100, fax (540) 231-2606

generated representations of living cells and organisms may be available and with this, powerful new investigative strategies to benefit the biomedical research community. The high rate of data generation in the life sciences, and its expected increases, poses a challenge for scientists who hope to capitalize on the value that it holds. Implementation of a cyberinfrastructure and the requisite bioinformatic tools for the exploration of this data by the scientific community is critical, and should enhance the research on the discovery of therapeutic targets, vaccines, and diagnostic markers.