4

A New Asset for Pathogen Informatics – the Enteropathogen Resource Integration Center (ERIC), an NIAID Bioinformatics Resource Center for Biodefense and Emerging/Re-emerging Infectious Disease

John M. Greene1, Guy Plunkett III2, Valerie Burland2, Jeremy Glasner2, Eric Cabot2, Brad Anderson2, Eric Neeno-Eckwall2, Yu Qiu2, Bob Mau2, Michael Rusch2, Paul Liss2, Thomas Hampton1, David Pot1, Matthew Shaker1, Lorie Shaull1, Panna Shetty1, Chuan Shi1, Jon Whitmore1, Mary Wong1, Sam Zaremba1, Frederick R. Blattner2, and Nicole T. Perna2

1 Health Research Systems, SRA International, Inc., john_greene@sra.com
2 Genome Center of Wisconsin, University of Wisconsin-Madison, ntperna@wisc.edu

Abstract. ERIC (Enteropathogen Resource Information Center) is one of the National Institute of Allergy and Infectious Diseases (NIAID) Bioinformatics Resource Centers for Biodefense and Emerging/Re-emerging Infectious Disease. ERIC serves as a comprehensive information resource for five related pathogens: Yersinia enterocolitica, Yersinia pestis, diarrheagenic E. coli, Shigella spp., and Salmonella spp. ERIC integrates genomics, proteomics, biochemical and microbiological information to facilitate the interpretation and understanding of ERIC pathogens and select related non-pathogens for the advancement of diagnostics, therapeutics, and vaccines. ERIC (www.ericbrc.org) is evolving to provide state-of-the-art analysis tools and data types, such as genome sequencing, comparative genomics, genome polymorphisms, gene expression, proteomics, and pathways as well as expertly curated community genome annotation. Genome sequence and genome annotation data and a variety of analysis and tools for eight strains of Yersinia enterocolitica and Yersinia pestis pathogens (Yersinia pestis biovars Mediaevalis KIM, Mediaevalis 91001, Orientalis CO92, Orientalis IP275, Antiqua Angola, Antiqua Antiqua, Antiqua Nepal516, and Yersinia enterocolitica 8081) and two strains of Yersinia pseudotuberculosis (Yersinia pseudotuberculosis IP32953 and IP31758) are currently available through the ERIC portal. ERIC seeks to maintain a strong collaboration with the scientific community so that we can continue to identify and incorporate the latest research data, tools, and training to best meet the current and future needs of the enteropathogen research community. All tools and data developed under this NIAID contract will be freely available. Please contact info@ericbrc.org for more information.

4.1 Introduction

The Enteropathogen Resource Integration Center (ERIC) was formed over two years ago and is available online at www.ericbrc.org. ERIC is a partnership between personnel at the Genome Center of the University of Wisconsin, Madison, and at SRA International in an effort to combine top-notch science with professional, disciplined software and systems engineering. ERIC focuses on the integration of data from five closely related pathogens: Yersinia pestis, Yersinia enterocolitica, Shigella spp., Salmonella spp., and diarrheagenic E. coli.
4.1.1 The BRC Program

In July 2004, NIAID funded a new bioinformatics effort intended to integrate the vast amount of genomic and other biological data already available, as well as being produced by the ramp-up in biodefense research. Eight Bioinformatics Resource Centers for Biodefense and Emerging/Re-Emerging Infectious Disease were funded to provide the research community working on a selected group of pathogens access to integrated genomics data to facilitate the discovery and development of novel therapeutics, vaccines, and diagnostics for these pathogenic organisms. The initial term of this program is for five years, and altogether this is likely the largest bioinformatics effort to date dedicated to human pathogens. Pathogenic species of biodefense and special public health interest have been classified by NIAID into three high-priority categories (Category A, B, and C; www3.niaid.nih.gov/Biodefense/bandc_priority.htm), based on their relative potentials for causing morbidity or mortality from disease in case of bio-warfare.

Among the primary goals of the BRCs is to provide users with easy Web access (and other types of user interfaces such as APIs or web services) to genomic and related data on these pathogens. Such data includes the genome sequences of multiple strains of these organisms and related plasmids, protein sequences, annotations, microarray data, epitopes, SNPs, proteomic data, and epidemiological data—as much data as possible to allow global and comparative analyses. Eventually, the BRCs will evolve to contain data relevant to host-pathogen interactions. A full list of the BRCs is available on a central portal designed to link the BRCs, www.brc-central.org, and at www.niaid.nih.gov/dmid/genomes/brc/default.htm, the NIAID program site.

All data and software produced under the contracts are freely available to the research community. The BRC program has a policy to publicly release any type of new data placed in the BRCs within 6-12 months from data deposition and upon publication. All data contributed to the BRCs is to be attributed as to source, and each BRC is required to have Data Transfer and Data Access Agreements.

A major emphasis of the BRC program is outreach to the research community to determine and serve their needs, and to make them aware of these new Centers’ substantial resources. In addition to direct interactions with researchers, each BRC has a Scientific Working Group (SWG) of approximately ten experts on their assigned pathogens’ biology, evolutionary genomics, biodefense, and bioinformatics who help provide vision for the BRCs, expand community ties, and represent the needs of the scientific community working on these pathogens to the BRC staffs.

4.1.2 ERIC – Enteropathogen Resource Integration Center

The Enteropathogen Resource Integration Center (ERIC) focuses on enteropathogens, including *Yersinia enterocolitica*, and the closely related *Yersinia pestis* (see Table 1) with a meticulous, disciplined focus on the annotation and curation of genes and gene families, which will be discussed in detail in Section 4.4 below.

Another focal point for ERIC comes from the realization that such a bioinformatics resource center will add the most value by integrating the disparate data tightly, such that a researcher can find out everything the resource has on a gene (for an example of a biological entity), than by simply providing a collection of disconnected tools. To accomplish this, the user interfaces for these tools and databases must be both scientist-friendly and work well together. Interacting closely with the enterobacterial research community has and will be critical to provide a truly useful resource to the community.