

The FDA MaAtPan pipeline for molecular epidemiological foodborne outbreak investigations: An integrated Geo-Genomic-Bioinformatic workflow for pathogen identification and traceback

Scott Jackson

United States Food & Drug Administration, USA

As a public health agency, the USFDA is tasked with maintaining a safe food supply. As such, the ability to respond rapidly to foodborne outbreaks and to trace back contaminated food to their sources is critical in preventing spread of disease. In this vein, an end-to-end genomics pipeline for molecular foodborne outbreak investigations was developed. The pipeline consists of 3 independently developed tools: (1) custom microarray-based genomic assays (2) ArrayTrack, and (3) Pathogen Annotated Tracking Resource Network (PATRN).

The FDA-ECID E. coli microarray is the culmination of research and development done by FDA over the last seven years and contains tens of thousands of biomarkers mined from over 300 E.coli whole genome sequences, thereby representing the true E. coli PanGenome. In addition to containing >40,000, E. coli genes, the array also contains ~ 10,000 highly informative single nucleotide polymorphism biomarkers. As a result, this hybrid design is highly discriminatory as well as biologically and evolutionarily informative providing a rapid (<24 hours), high throughput, semi-automated, and affordable assay.

Microbial ArrayTrack was developed by bioinformaticists at FDA-NCTR as a high-end microarray data storage, analysis, and visualization package. As such, Microbial ArrayTrack contains ECID-specific annotations in conjunction with a customized and validated data analysis and visualization solution.

PATRN is a globally accessible web-based system for the comprehensive analysis of pathogen data for identification, source-tracking, and public health decision making. Through PATRN, food source, data from microbiological, genotyping assay results, and next generation genomic data are leveraged for combinatorial cluster analyses, graphing, and data mining of similar data obtained from ongoing outbreak and surveillance activities.

In total, we've developed a complete package for genomic-scale molecular subtyping of foodborne pathogens. Furthermore, adaptation and integration of Microbial ArrayTrack with PATRN provides a turn-key, customized platform for microarray data storage, analysis, and visualization, thus providing a complete geogenomic solution for molecular identification of pathogens during foodborne investigations.

Scott.Jackson@fda.hhs.gov