ANALYSES OF MULTIMODAL DATASETS ON SALMONELLOSIS OUTBREAK IN THE UNITED STATES

Christopher Jackson, Tolulola Oyeleye and Raphael D. Isokpehi

Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson MS 39217, USA

Abstract: Salmonellosis is an infection, caused by bacteria of the genus *Salmonella*. The infections are often traced back to poultry, meat, and dairy products, but any food may become contaminated. Symptoms include fever, abdominal pain, and diarrhea. Symptoms appear 12 to 72 hours after being exposed to the bacteria. Salmonellosis is a self-limiting disease, and there is no use for medicines, except in severe cases. The 2008 outbreak of Salmonellosis in the United States associated to the SaintPaul strain began in April 2008 and was initially linked to tomatoes (*Solanum lycopersicum*). Other fresh produce linked to this outbreak are jalapeno, cilantro and serrano which are used for preparing salsa. Over 1200 salmonellosis cases have been identified in over 40 states. Though scientific literature provides systematic information about outbreaks, they are not as timely as news reports available from the Internet. As of June 30, 2008, there was only one 2008 publication on a Salmonella SaintPaul outbreak, which was not about the outbreak in the United States. However, there are over 10,000 online news reports from different local or regional news media that could provide up-to-date situational awareness and help with future emergency preparedness. We have collected over 2000 online news reports as well as video reports. In addition, we have tracked the changes in the salmonellosis incidence map in the website of the Center for Disease Control and Prevention. We are building computational pipelines to analyze these multimodal datasets in order to extract novel epidemiological information by data integration. The computational tools for automated text and image mining can be used for tracking and managing future disease outbreaks.

Acknowledgements: Mississippi NSF-EPSCoR “Innovations through Computational Sciences” Award (EPS-0556308); the Research Centers in Minority Institutions (RCMI) – Center for Environmental Health (NIH-NCRR G12RR13459-09); U.S. Department of Homeland Security under Grant Award Number 2007-ST-104-000007 (Bioinformatics in Biodefense Career Development Program). Disclaimer: “The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.