

SCCWRP Annual Report 2003-04

Assessment of statistical methods used in library-based approaches to microbial source tracking

Kerry J. Ritter, Ethan A. Carruthers¹, C. Andrew Carson², R. D. Ellender³, Valerie J. Harwood⁴, Kyle S. Kingsley⁵, Cindy H. Nakatsu⁶, Michael J. Sadowsky¹, Brian L. Shear⁴, Brian R. West⁵, John E. Whitlock⁷, Bruce A. Wiggins⁸ and Jayson D. Wilbur¹⁰

¹University of Minnesota, Department of Soil, Water, & Climate, St. Paul, MN

²University of Missouri, Department of Veterinary Pathobiology, Columbia, MO

³University of Southern Mississippi, Hattiesburg, MS

⁴University of South Florida, Department of Biology, Tampa, FL

⁵Sint-Martens-Latem, Applied-Maths, Belgium

⁶Purdue University, Department of Agronomy, West Lafayette, IN

⁷Hillsborough Community College, Division of Math and Science, Tampa, FL

⁸James Madison University, Department of Biology, Harrisonburg, VA

⁹Worcester Polytechnic Institute, Department of Mathematical Sciences, Worcester, MA

ABSTRACT

Several commonly used statistical methods for fingerprint identification in microbial source tracking (MST) were examined to assess the effectiveness of pattern-matching algorithms to correctly identify sources. Although numerous statistical methods have been employed for source identification, no widespread consensus exists as to which is most appropriate. A large-scale comparison of several MST methods, using identical fecal sources, presented a unique opportunity to assess the utility of several popular statistical methods. These included discriminant analysis, nearest neighbor analysis, maximum similarity and average similarity, along with several measures of distance or similarity. Threshold criteria for excluding uncertain or poorly matched isolates from final analysis were also examined for their ability to reduce false positives and increase prediction success. Six independent libraries used in the study were constructed from indicator bacteria isolated from fecal materials of humans, seagulls, cows and dogs. Three of these libraries were constructed using the rep-PCR technique and three relied on antibiotic resistance analysis (ARA). Five of the libraries were constructed using *Escherichia coli* and one using *Enterococcus* spp. (ARA). Overall, the outcome of this study suggests a high degree of variability across statistical methods. Despite large differences in correct classification rates among the statistical methods, no single statistical approach emerged as superior. Thresholds failed to consistently increase rates of correct classification and improvement was often associated with substantial effective sample size reduction. Recommendations are provided to aid in selecting appropriate analyses for these types of data.

Full Text

ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2003_04AnnualReport/ar28-ritter_338-349.pdf