

1056-92-1958

**Mayra Hernandez\*** (mayra.alina@yahoo.com), **Isaacs Kate**, **Naneh Apkarian**, **Michelle Creek**, **Eric Guan**, **Chris Peterson** and **Todd Regh**. *Organism-Associated and Mat-Forming Metagenome Statistical Analysis using Random Forests Algorithms*. Preliminary report.

Analysis of metagenomic data clusters together specimens from given biological lines. Given the diversity of the environment, many classes result and specimens within each class tend to be viewed as though they are identical. The reality of the diversity within a class is explored in this series of analysis. We employ various supervised and unsupervised methods coupled with graphical representations to ascertain both the inherent difficulty in large metagenomic analysis (due to intra-group diversity) and to point out the high effectiveness of metagenomic analysis even in large data sets in accurately defining given environments along biologically meaningful lines, despite the difficulty introduced by intra-class diversity. These results support and fit well with findings by Dinsdale et al which was among the first to explore the possibility to define metagenomes by the subsystem association. (Received September 22, 2009)