Quikr and WGSQuikr: Rapid Bacterial Community Reconstruction Via Compressive Sensing

D. Koslicki, S. Foucart, G. Rosen
University of Georgia, Athens

Many metagenomic studies compare hundreds to thousands of environmental and health-related samples by extracting and sequencing their DNA. However, one of the first steps - to determine what bacteria are actually in the sample - can be a computationally time-consuming task since most methods rely on computing the classification of each individual read out of tens to hundreds of thousands of reads. We introduce Quikr: a QUadratic, K-mer based, Iterative, Reconstruction method which computes a vector of taxonomic assignments and their proportions in the sample using an optimization technique motivated from the mathematical theory of compressive sensing. On both simulated and actual biological data, we demonstrate that Quikr is typically more accurate as well as typically orders of magnitude faster than the most commonly utilized taxonomic assignment techniques for both whole genome techniques (Metaphyler, Metaphlan) and 16S rRNA techniques (the Ribosomal Database Project's Naive Bayesian Classifier). We also show that in general nonnegative L1 minimization can be reduced to a simple nonnegative least squares problem. The talk is based on the papers [1] and [2] whose accompanying implementations for Unix and Mac are available at `http://sourceforge.net/projects/quikr/` and `http://sourceforge.net/projects/wgsquikr/`.

References

[1] Quikr: a method for rapid reconstruction of bacterial communities via compressive sensing. D. Koslicki, S. Foucart and G. Rosen. Bioinformatics 29:2096–2102 (2013).

[2] WGSQuikr: fast whole-genome shotgun metagenomic classification. D. Koslicki, S. Foucart and G. Rosen. PLoS ONE, 9:e91784 (2014).