Background

The number of sequenced genomes is growing exponentially, profoundly shifting the bottleneck from data generation to genome interpretation. Traits are often used to characterize and distinguish bacteria, and are likely a driving factor in microbial community composition, yet little is known about the traits of most microbes. We present Traitar, the microbial trait analyzer, a fully automated software package for deriving phenotypes from the genome sequence. Traitar accurately predicts 67 traits related to growth, oxygen requirement, morphology, carbon source utilization, antibiotic susceptibility, amino acid degradation, proteolysis, carboxylic acid use and enzymatic activity.

Results

Traitar uses L1-regularized L2-loss support vector machines for phenotype assignments, trained on protein family annotations of a large number of characterized bacterial species, as well as on their ancestral protein family gains and losses. We demonstrate that Traitar can reliably phenotype bacteria even based on incomplete single-cell genomes and simulated draft genomes. We furthermore showcase its application by characterizing two novel Clostridiales species based on genomes recovered from the metagenomes of commercial biogas reactors, verifying and complementing a manual metabolic reconstruction.

Conclusions

Traitar enables microbiologists to quickly characterize the rapidly increasing number of bacterial genomes. It could lead to models of microbial interactions in a natural environment and inference of the conditions required to grow microbes in pure culture. Our phenotype prediction framework offers a path to understanding the variation in microbiomes. Traitar is available at https://github.com/hzi-bifo/traitar.