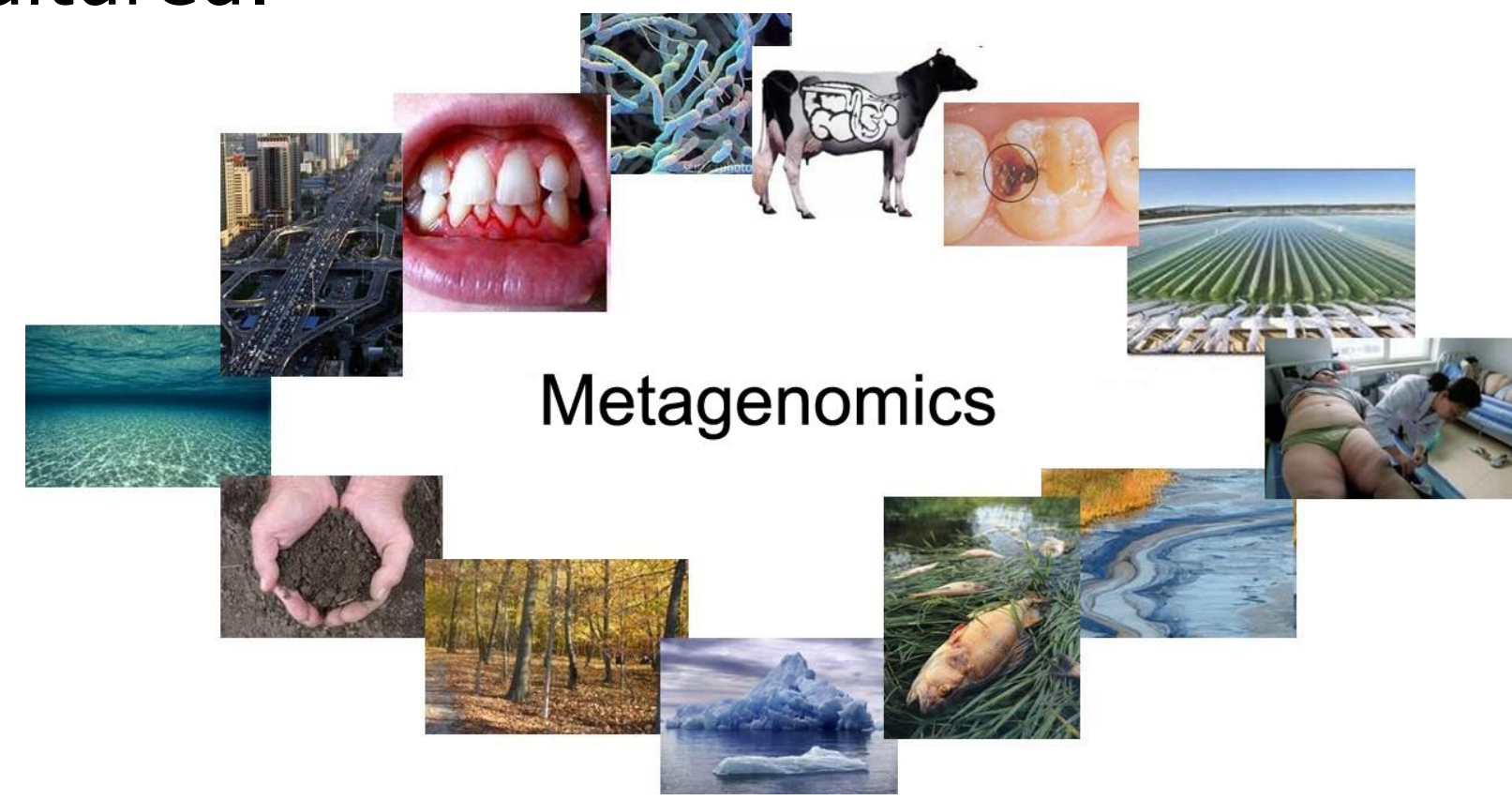




COMPUTATIONAL SCIENCE RESEARCH CENTER, SAN DIEGO STATE UNIVERSITY

Introduction

Microbes are more abundant than any other organism, and it is important to understand what those organisms are doing and who they are. In many environments more than 99% of the members of the microbial community cannot be cultured.



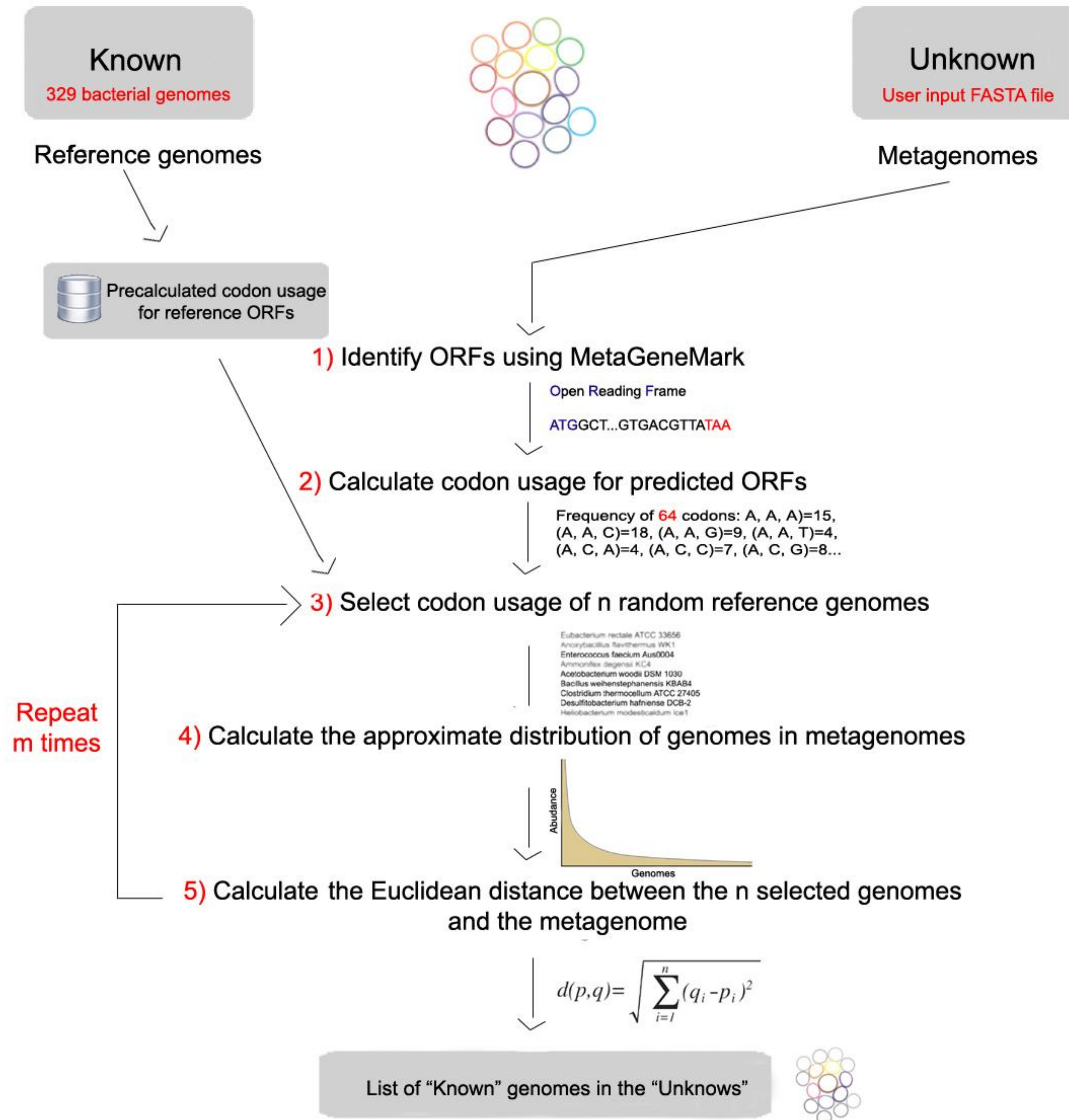
One of the major goals in metagenomics is to identify the presence of organisms in the microbial community from a huge set of unknown DNA sequences; this profiling has valuable applications in multiple important areas of medical research such as disease diagnostics. Nevertheless, it is not a simple task, and many approaches that have been developed are slow and depend on the read length of the DNA sequences.

FOCUS

An innovative and agile stochastic model to profile and report organisms present in metagenomic samples based on codon usage without sequence length dependencies.

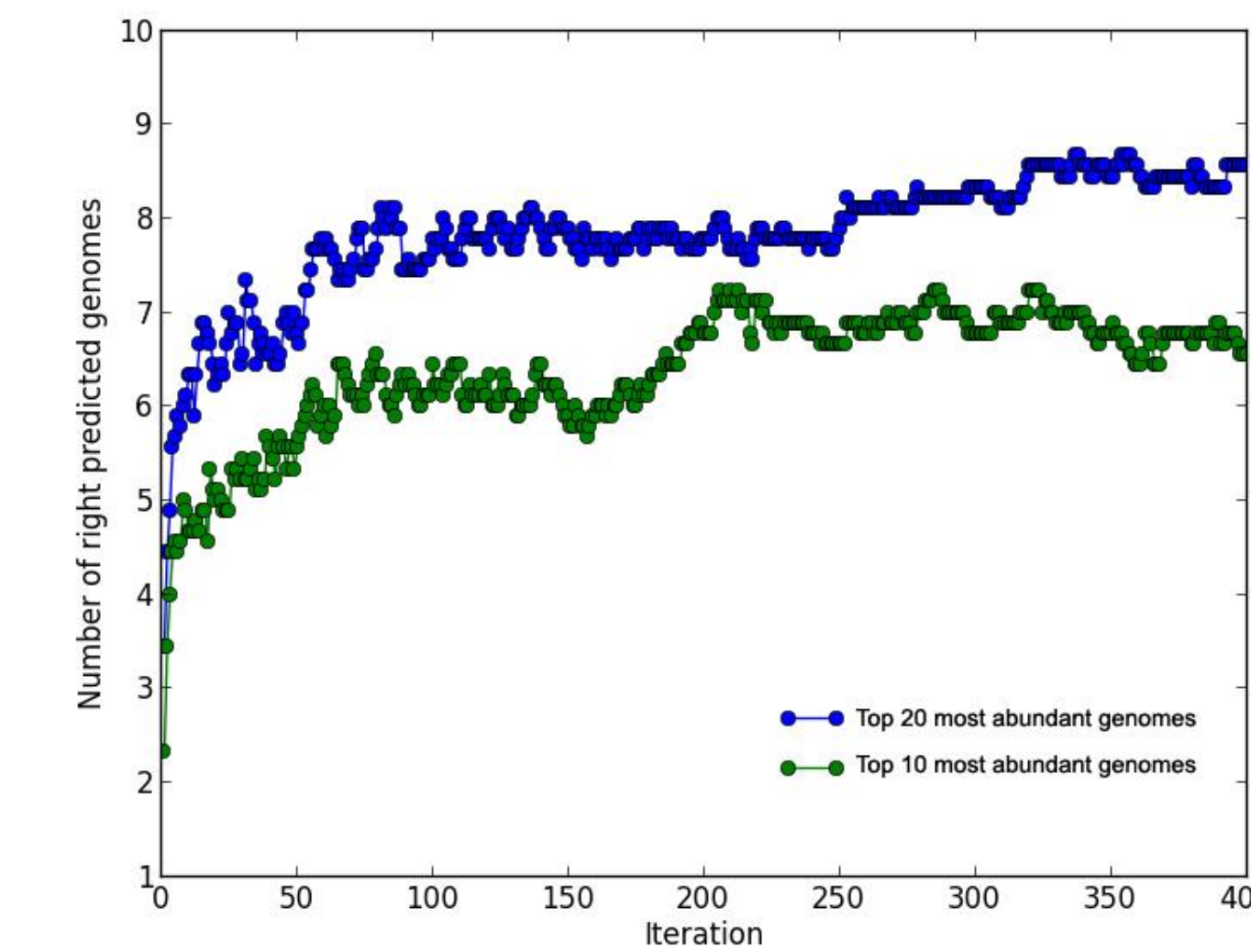
Methods

FOCUS is a model written in Python. The Figure below illustrates how **FOCUS** models the data.

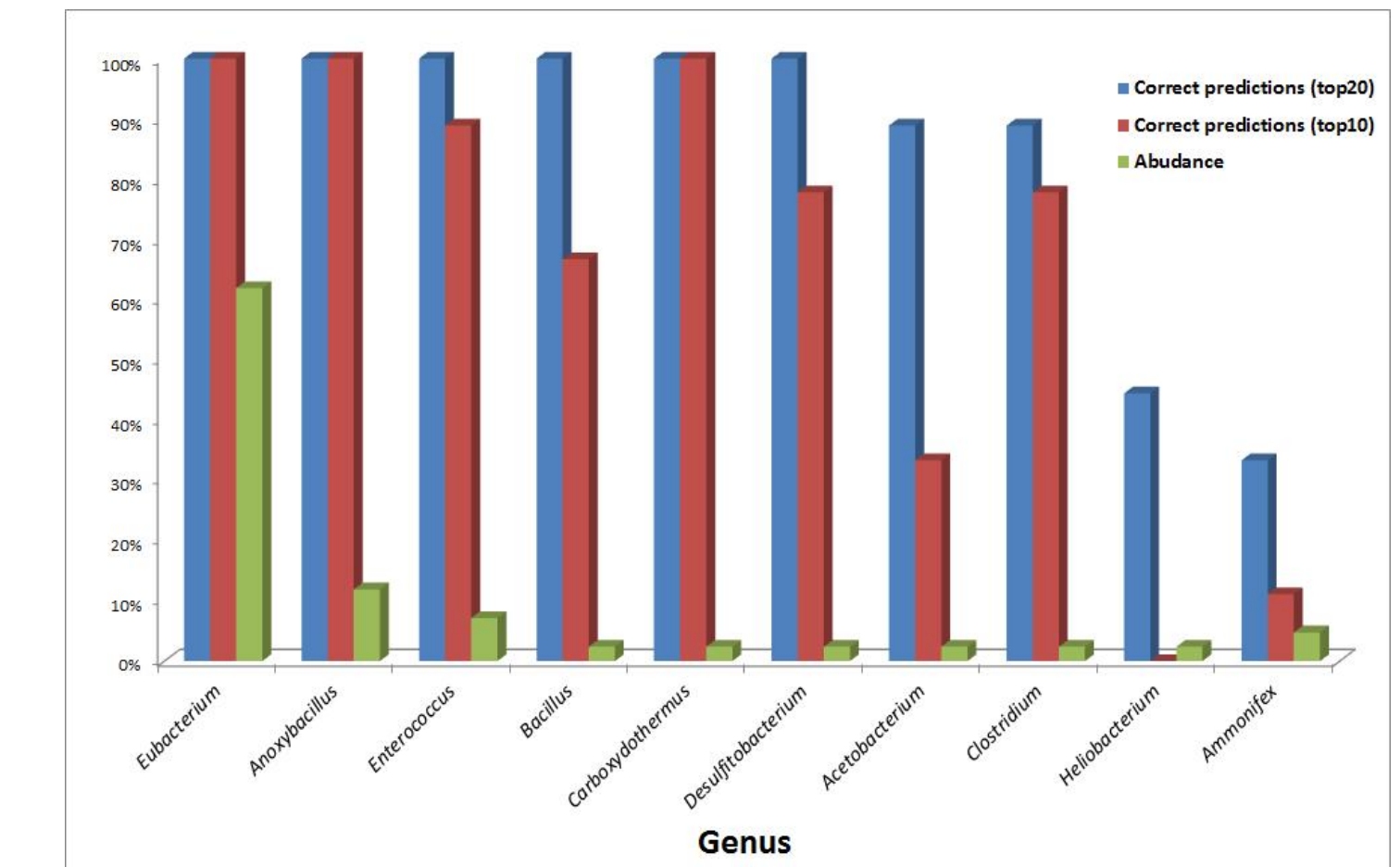


Results

The application was evaluated using simulated metagenomes generated from ten bacterial genomes, and the results show that our approach has approximately 85% accuracy at inferring the organisms in the top twenty most abundant genomes.



Number of correct predictions by iteration on the top 10 and 20 most abundant genomes.

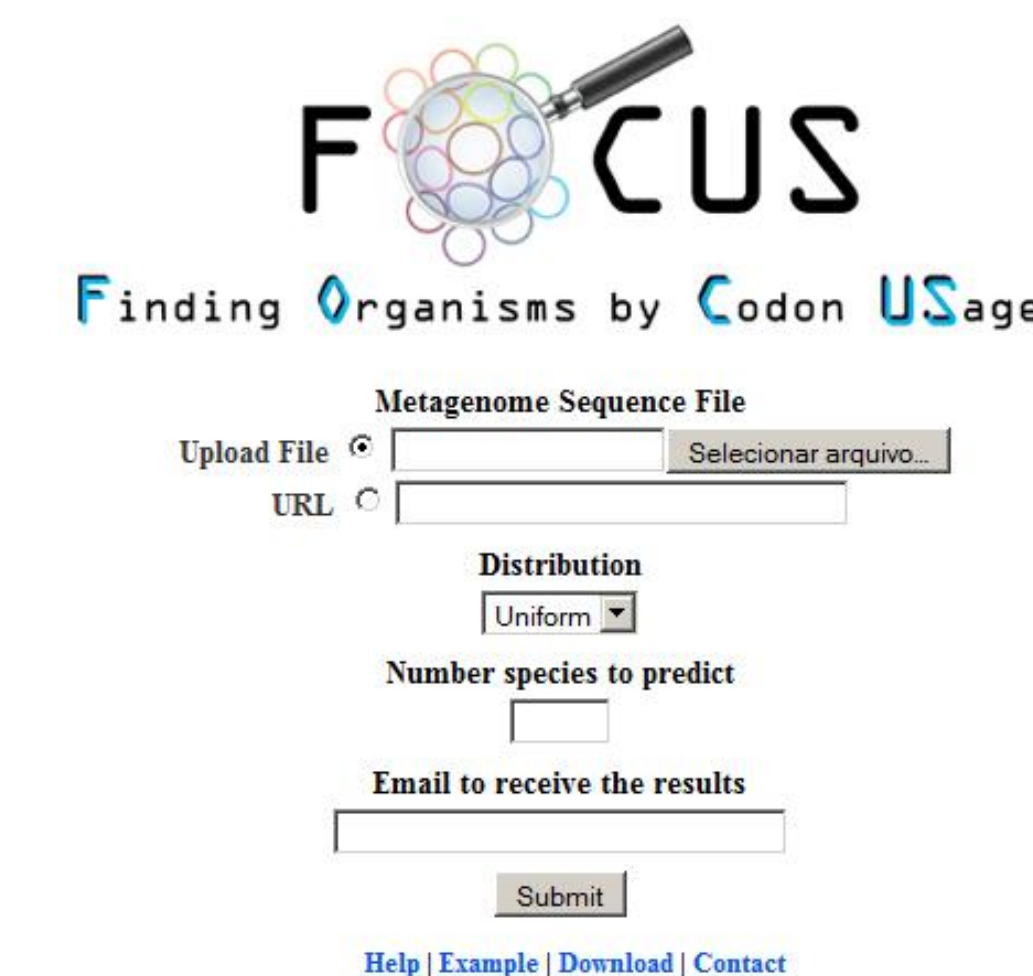


Percent of right prediction on the top 10 and 20 most abundant genus, and their abundance.

Conclusions

FOCUS helps to identify which organisms are present in metagenomes; the algorithm presented will help biologists explore the microbes present in their samples.

Web-based version



At <http://edwards.sdsu.edu/FOCUS>

Please contact me at genivaldo.queiros@gmail.com for further information.