
Building a Phylogenetic Tree From Raw Genome Data

Onno Eberhard

Matrikelnummer 5618970

onno.eberhard@student.uni-tuebingen.de

Abstract

I am planning to use genome sequence data from the National Center for Biotechnology Information (NCBI) or the Genome Taxonomy Database (GTDB) to build a phylogenetic tree of life. If this turns out to be easy, I will also conduct some more analysis, like clustering into kingdom/order/etc or regression or classification of characteristics of the organisms (like size), as a function of simple features like amino acid counts.