Abstract

Tuberculosis (TB) is the most lethal bacterial pathogen on the planet and treatment is getting harder due to emerging drug resistances. Without treatment, 10% of TB infections develop into an active form, 50% of which are lethal. Recently a number of TB strains evolved which are resistant to either multiple or all available antibiotics. Reasons for this resistance remain still elusive.

This thesis proposes a workflow to detect known causes for drug resistances by identifying associated single nucleotide polymorphisms (SNPs) based on whole genome analysis. Additionally phylogenetical relations and evolutionary origin of the examined pathogens are investigated.

High quality SNPs are identified by combining the results of two variant calling pipelines. SNPs found via local re-assembly of haplotypes in a reference based mapping and SNPs found by *de-novo* assembly and reference comparison are combined. SNPs that both methods agree on are subsequently filtered and compared with the tuberculosis drug resistance mutation database. For the phylogenetic analysis a maximum likelihood based tree and a Minimum Spanning Tree are generated.

The implemented workflow was applied to data from 64 newly sequenced, drug resistant *Mycobacterium tuberculosis* strains, which were collected and sequenced at the University of Medicine and Pharmacy in Iasi, Romania. SNPs related to drug resistances were found in all but one strain; Two strains were resistant to six of the seven investigated drugs. The phylogenetic analysis assigned four different lineages to the examined clinical isolates.