Comparison of eight M. tuberculosis strains in the presence of drug resistance data

Michał Woźniak* Limsoon Wong† Jerzy Tiuryn*‡

Abstract

The first fully sequenced M. tuberculosis strain was H37Rv [1] and since then there has been published several new MTB genomes [2]. Progress in sequencing enables new possibilities for analyzing mechanisms of drug resistance. For our experiments, we used eight fully sequenced MTB strains, of which five are susceptible to drugs: H37Rv, H37Ra, CDC1551, F11, KZN 4207; two are Multi-Drug-Resistant: KZN 1435, KZN V2475; and one is Extensively-Drug-Resistant: KZN R506.

As a starting point we used 3988 gene sequences annotated for H37Rv strain. Then for each gene we searched homologs in seven other strains. For 3678 genes BLAST returned hits in all analysed strains. In this preliminary version of our work we used this set for next phases of the experiment. Closer look at calculated multiple alignments showed that 105 gene families contain a single gap ("-") in exactly one strain (in corresponding position of the alignment). The fact that frameshift mutation is very unlikely, suggests that perhaps most of them are sequencing errors and lead to annotation errors. After correcting these "one-gap-frame-shift" we were left with 3632 gene families for further analysis.

We define a mutation as *essential* (from the standpoint of drug resistance) when in the corresponding position of the multiple alignment all sequences of susceptible strains have the same character, except for at least one drug resistant strain. Then, we verified the hypothesis that this type of significant mutations occurs more frequently in drug target genes (based on 11 target genes of drugs: Isoniazid, Rifampicin, Streptomycin, Ofloxacin, Kanamycin, Ethambutol). The proportion of observed mutations to the gene count is 0.64 for the drug targets, while it is only 0.0083 for the others, thereby supporting the hypothesis.

Acknowledgments

This work is partially supported by Polish Ministry of Science and Higher Education grants no. N N301 065236 and PBZ-MNiI-2/1/2005.

References

[1] S.T. Cole et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685):537–544, June 1998.

[2] Ioerger TR, Koo S, No E-G, Chen X, and Larsen MH et al. Genome Analysis of Multi- and Extensively-Drug-Resistant Tuberculosis from KwaZulu-Natal, South Africa. *PLoS ONE*, 4(11):e7778, 11 2009.

*Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland.
†School of Computing, National University of Singapore, Singapore.
‡Corresponding author: Jerzy Tiuryn, tiuryn@mimuw.edu.pl