Tree-generalized hypergeometric test for detection of drug resistance-associated mutations

Michał Woźniak[1,2], Limsoon Wong[2], Jerzy Tiuryn[1]

1. The Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland
2. School of Computing, National University of Singapore, Singapore

Drug resistance in bacterial pathogens is an increasing problem that stimulates research. However, still our understanding of drug resistance mechanisms remains incomplete. One promising approach to deepen our understanding of drug resistance mechanisms is to use whole-genome sequences to identify genetic mutations associated with drug resistance phenotypes in bacterial strains[1].

In this work, we present a new statistical measure, called tree-generalized hypergeometric (TGH) score, for detection of drug resistance-associated mutations in bacterial strains. This score generalizes the standard hypergeometric test by using a phylogenetic tree to capture the dependencies between the strains in question. When the input phylogenetic tree is flat our method is equivalent to the standard hypergeometric test. The presented scoring scheme is a part of the GWAMAR pipeline we have developed for detection of drug resistance-associated mutations.

In order to test our approach we run it on two datasets for M. tuberculosis. The first dataset consist of genotype and phenotype data we have collected from publications for a set of 173 fully sequenced M. tuberculosis strains and 10 commonly used drugs. The second dataset has been publicly deposited by The Broad Institute in the Tuberculosis Drug Resistance Mutation Database[3]. Our computational experiments show that our method is capable of identifying drug resistance-associated mutations more accurately than standard hipergeometric test, as well as other methods which do not incorporate phylogenetic information. Applying our methodology we have identified some novel putative drug resistance-associated mutations.

Acknowledgments

This work is supported in part by Polish Ministry of Science and Higher Education grant no. 2012/05/ST6/03247 and Singapore Ministry of Education Tier-2 grant no. MOE2009-T2-2-004.

References

1. M. Wozniak et al.: *An approach to identifying drug resistance associated mutations in bacterial strains*, BMC Genomics 2012
2. F. Habib et al.:*Large scale genotype-phenotype correlation analysis based on phylogenetic trees*, Bioinformatics 2007
3. The Broad Institute: Tuberculosis Drug Resistance Mutation Database, http://www.broadinstitute.org/annotation/genome/mtb_drug_resistance.1/DirectedSequencingHome.html