An approach to identifying drug resistance associated mutations in bacterial strains.

Michał Woźniak*‡, Jerzy Tiuryn*, Limsoon Wong†
*Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland.
†School of Computing, National University of Singapore, Singapore.
‡Corresponding author: Michał Woźniak, m.wozniak@mimuw.edu.pl.

Abstract

Drug resistance in bacterial pathogens is an increasing problem, which stimulates research. However, our understanding of drug resistance mechanisms remains incomplete. One promising approach to further understand drug resistance mechanisms is to use whole-genome sequences to identify genetic mutations associated with drug resistance phenotypes for bacterial strains [1, 2].

We present a new comparative approach to identify genes and mutations that are likely to be associated with drug resistance mechanisms. Applying the method, we re-discovered the most common genetic determinants of drug resistance and identified some novel putative associations.

Collection of data

In order to test the approach, we collected genotype and phenotype data (from over 50 publications) of 100 fully sequenced *S. aureus* strains (with sequencing status "Completed" or "Scaffolds") and 10 drugs.

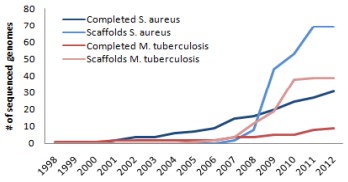

Figure 1: Progress in whole-genome sequencing

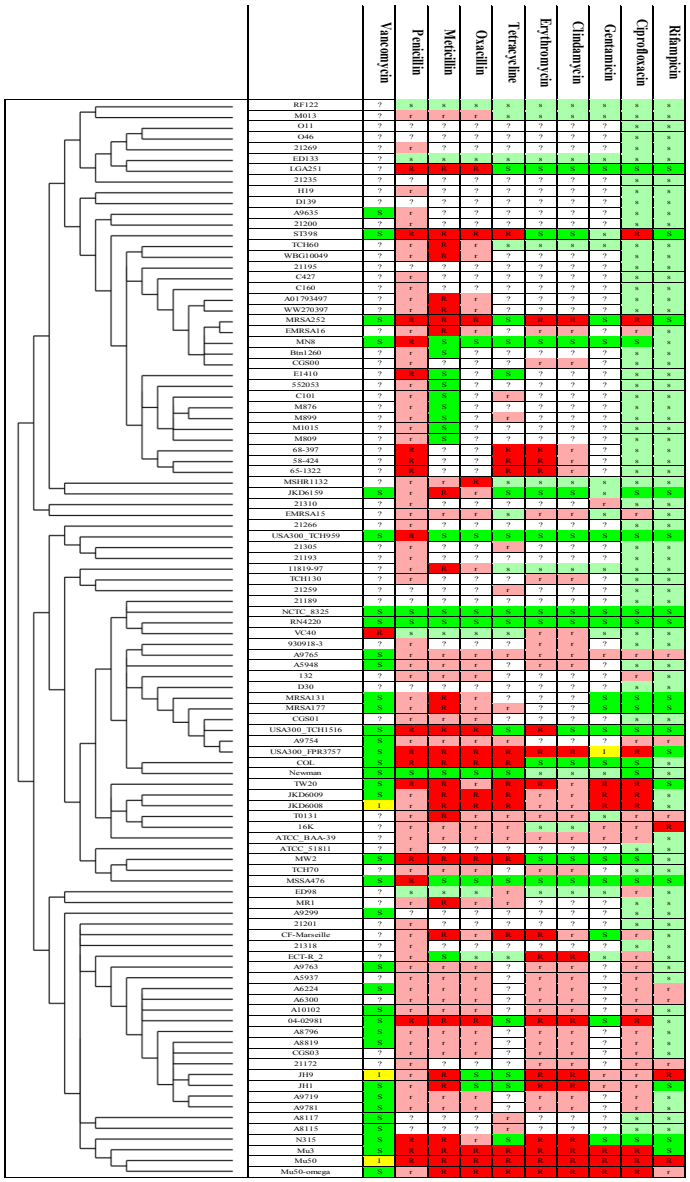

Figure 2: Collected dataset of phenotypes put together with results of our drug resistance predictions based on the presence of known drug resistance determinants. Columns represent drugs, rows represent *S. aureus* strains included in the study in the order corresponding to the reconstructed phylogenetic tree of strains. Green, yellow and red and cell colors represent susceptible, intermediate resistant and resistant phenotypes, respectively. Analogously, light green and light red cell colors represent predicted susceptible and resistant phenotypes, respectively. White cell color represents unknown (not determined by experiments or prediction) drug resistance phenotypes.

Methodology

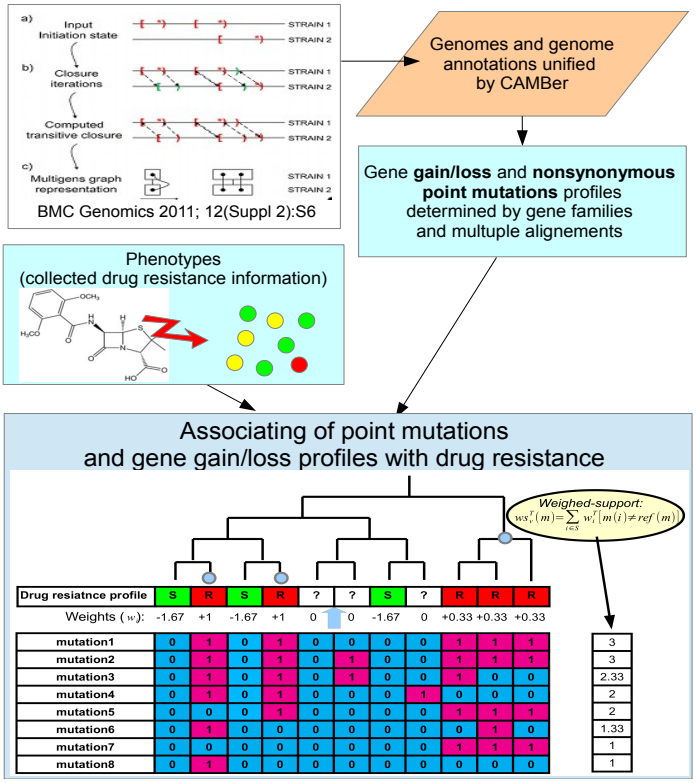

Figure 3: Schema of the methodology. We assign weights w_i (present in the definition of *weighted support*) in the following way: all drug susceptible strains are assigned a weight equal to the proportion of the drug resistant to the drug susceptible strains; each drug- resistant strain i is assigned a weight $\frac{1}{n_i}$, where n is the number of drug-resistant strains in the subtree (containing strain i) determined by its highest parental node (marked above by filled circles), such that the subtree does not contain any drug-susceptible strain in its leaves. All strains without drug resistance information are assigned weights 0.

Results

We examine the usability of our approach by trying to re-identify the known drug resistance determinants. Our experiment shows that the average rankings of the known drug resistance determinants obtained by employing *weighted support* and *odds ratio* are 2.26 and 6.61, respectively. It suggests that *weighted support* is better to identifying genetic features associated with drug resistance than *odds ratio*, which does not incorporate additional information about phylogeny.

Figure 4: Summarizing tables for the top scored gene gain/loss profiles (left panel) and nonsynonymous point mutations (right panel). The consequent columns refer to: gene identifier of the corresponding gene family; position and variants of the amino acid changes for point mutations; normalized weighted support (NWS); p-value for the NWS scores; and presence/absence profiles of the genetic features in the reference to the most common state in the drug susceptible strains.

References

[1] Fournier PE. et al. Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. PLoS Genetics, 2011.

[2] Peleg AY. et al. Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of *Staphylococcus aureus*. PLoS ONE, 2012.

[3] Wozniak M. et al. CAMBer: an approach to support comparative analysis of multiple bacterial strains. BMC Genomics, 2011.

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