Antibiotic Resistance Prediction and Biomarker Discovery in *Neisseria Gonorrhoeae*

Rushank Goyal

All India Institute of Medical Sciences <https://orcid.org/0000-0001-6484-3195>

Rashmi Chowdhary (✉ rashmi.biochemistry@aiimsbhopal.edu.in)

All India Institute of Medical Sciences

Research Article

Keywords: Antibiotic resistance, Machine learning, Biomarker discovery, *Neisseria gonorrhoeae*, Genetic marker, Azithromycin, Ciprofloxacin, Cefixime

Posted Date: June 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-631320/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Antibiotic resistance is a global problem projected to kill 10 million each year by 2050 and the CDC lists *Neisseria gonorrhoeae* among the most urgent threats in this area. There also exists a severe lack of efficient resistance detection techniques as only a handful of mutations have been identified as causing resistance thus far. In this study, eight models were trained on three datasets containing data regarding azithromycin, ciprofloxacin and cefixime, three drugs of choice against *N. gonorrhoeae*. Each dataset had 3000+ samples and their corresponding resistance values; each sample consisted of a binary series representing the presence/absence of certain base pair sequences within that sample's genome. The base pair sequences were special regions called unitigs (short sequence reads of the genome) that were also designed to have biological relevance. This novel technique differs from the standard research in this field, which has almost exclusively used whole-genome sequences. Once the models were trained, their accuracies, sensitivities and specificities were compared and analyzed. Maximum accuracies of 97.6%, 95.9% and 100% were achieved on azithromycin, ciprofloxacin and cefixime training data respectively. Secondly, Fisher's exact test was used to test for the existence of genetic signatures i.e. biomarkers that had a statistically significant correlation with antibiotic resistance. The feature importances of the top models from the first step were used to create a ranking of these genetic signatures, again a novel concept. Out of 584,362 unitigs, 191, 3304 and 1 were identified as statistically significant for azithromycin, ciprofloxacin and cefixime respectively. Lastly, protein biomarkers were also identified for the top 10 genetic signatures, resulting in several new discoveries not described before in scientific literature. Overall, this study led to the development of a new technique for biomarker discovery and resistance prediction in bacteria along with the creation of highly accurate machine learning models and identification of genetic and protein biomarkers for resistance against three drugs in *N. gonorrhoeae*. The model can be used for genotype-based resistance diagnosis and the biomarkers can be further developed for point-of-care testing in the future.

Full Text

This preprint is available for [download as a PDF](#).

Supplementary Material

The supplementary material is not available with this version.