HMD-ARG: Identifying antibiotic resistance genes through machine learning

Yu Li
Zeling Xu
Wenkai Han
Huילו Cao
Ramzan Umarov
Aixin Yan
Ming Fan
Huan Chen
Carlos Duarte
Lihua Li
Pak-Leung Ho
Xin Gao

Video Byte

Keywords: Microbiome, antibiotic resistance genes, ARGs, deep learning, antibiotic class, resistance mechanism, gene mobility, multi-task learning, HMD-ARG, database, metagenomics, beta-lactamase, protein sequence, convolutional neural network, CNN, one-hot encoding, BLAST, saliency, PCR, cross-validation

Posted Date: February 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-276378/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

The spread of antibiotic resistance is one of the most pressing threats facing global health. Every year, approximately 700,000 deaths worldwide can be traced to antibiotic resistance. That makes it crucial to identify antibiotic resistance genes (ARGs) and their transmission between humans and the environment. Unfortunately, because they rely on curated databases and are not sensitive to certain mutations, many methods can overlook novel ARGs. Now, a new machine learning method called HMD-ARG could provide researchers with a more powerful alternative. Taking sequence encoding as input, it determines whether an input sequence is an ARG, what antibiotic family the ARG is resistant to, its mechanism of resistance, and whether it is intrinsic or acquired, and even the sub-class of antibiotic the ARG resists, if it happens to be a beta-lactamase, all without querying against existing sequence databases. The HMD-ARG database is the largest of its kind. and the HMD-ARG algorithm shows higher accuracy, recall, and precision than existing methods. While the algorithm's performance may be limited in some cases – for example, it doesn't work on short reads directly. HMD-ARG could be a powerful tool for identifying antibiotic resistance genes and reducing their threat to human health.