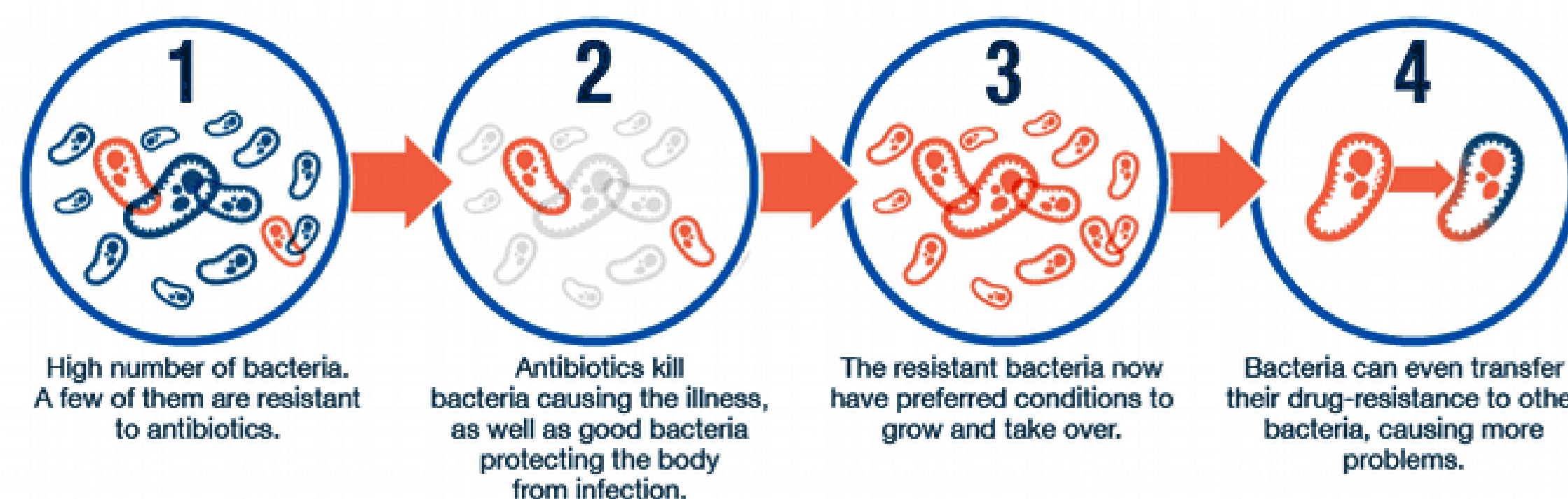


## Objectives

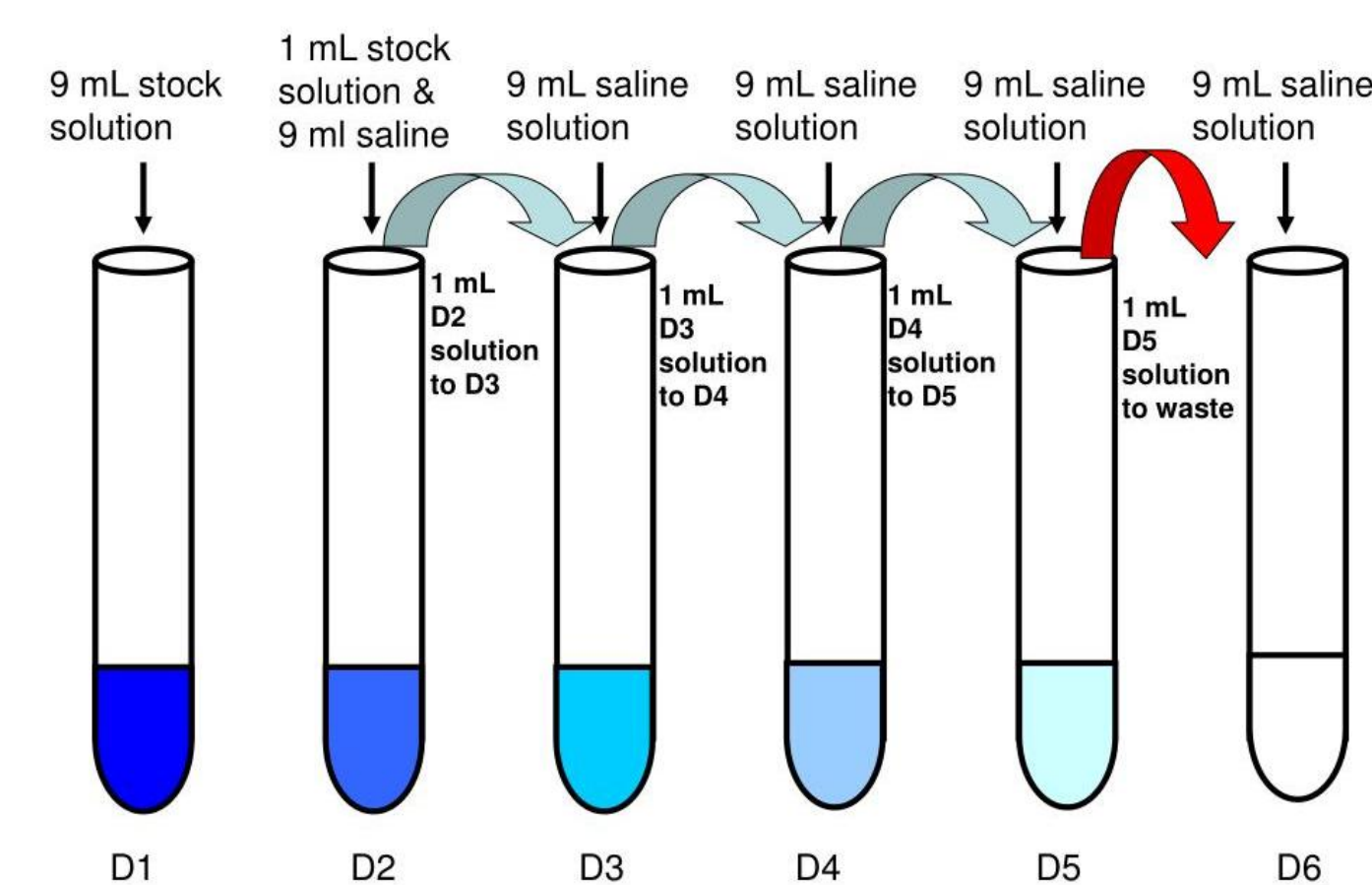
- To predict Minimum Inhibitory Concentration with a good accuracy, low cost and fast speed from the genome of a microbial strain
- To find the genetics determinants of antimicrobial resistance

# Introduction

- Anti-Microbial Resistance (AMR) occurs when bacteria change in response to the use of medicines.

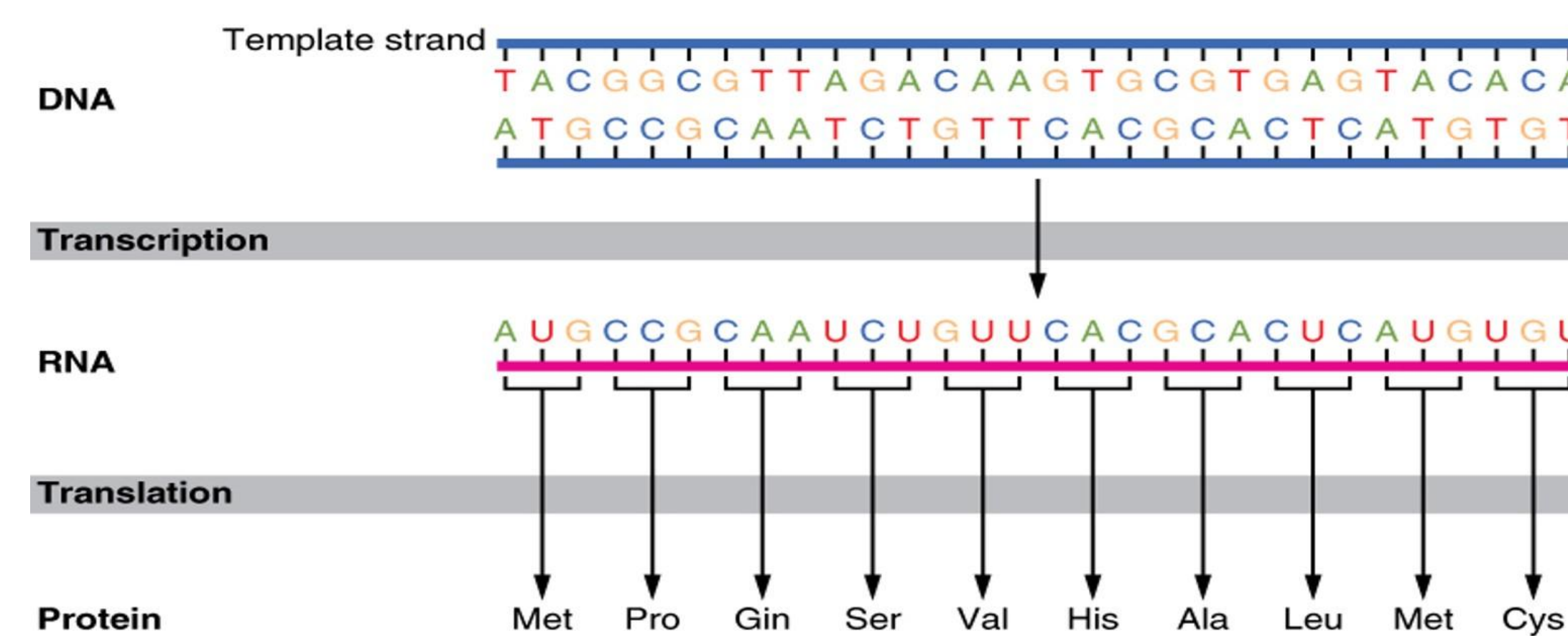


- The antibiotics should not be overprescribed. Only the minimum concentration must be used to treat the infection.
- MIC: Minimum inhibitory concentration : The lowest concentration of a chemical, usually a drug, which prevents visible growth of a bacterium



## DNA and amino acid

- DNA of the bacterium contains the AMR information and can be used for predicting MIC
- DNA sequences are translated to amino acid sequences

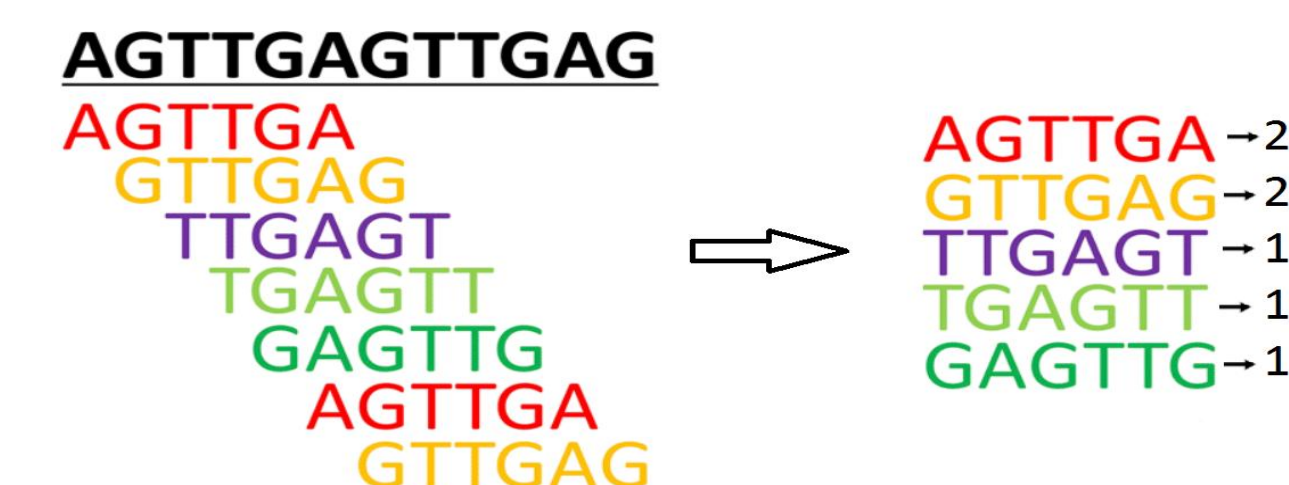


## Methods

## Data Preprocessing

DNA	Antibiotic	MIC
AAC TTATATACACAGG GTACACAT...	Ampicillin	2 g/ liter
GGG TACCTGATGTGA ACCATGRA...	Ampicillin	1 g/liter
CCCGGTAGCTGACTG TACTGGGG...	Tetracycline	4 g/ liter
ATCGGGGTCA TTGCAA TGACCGT...	Ampicillin	1 g/liter

- DNA is converted to k-mers.

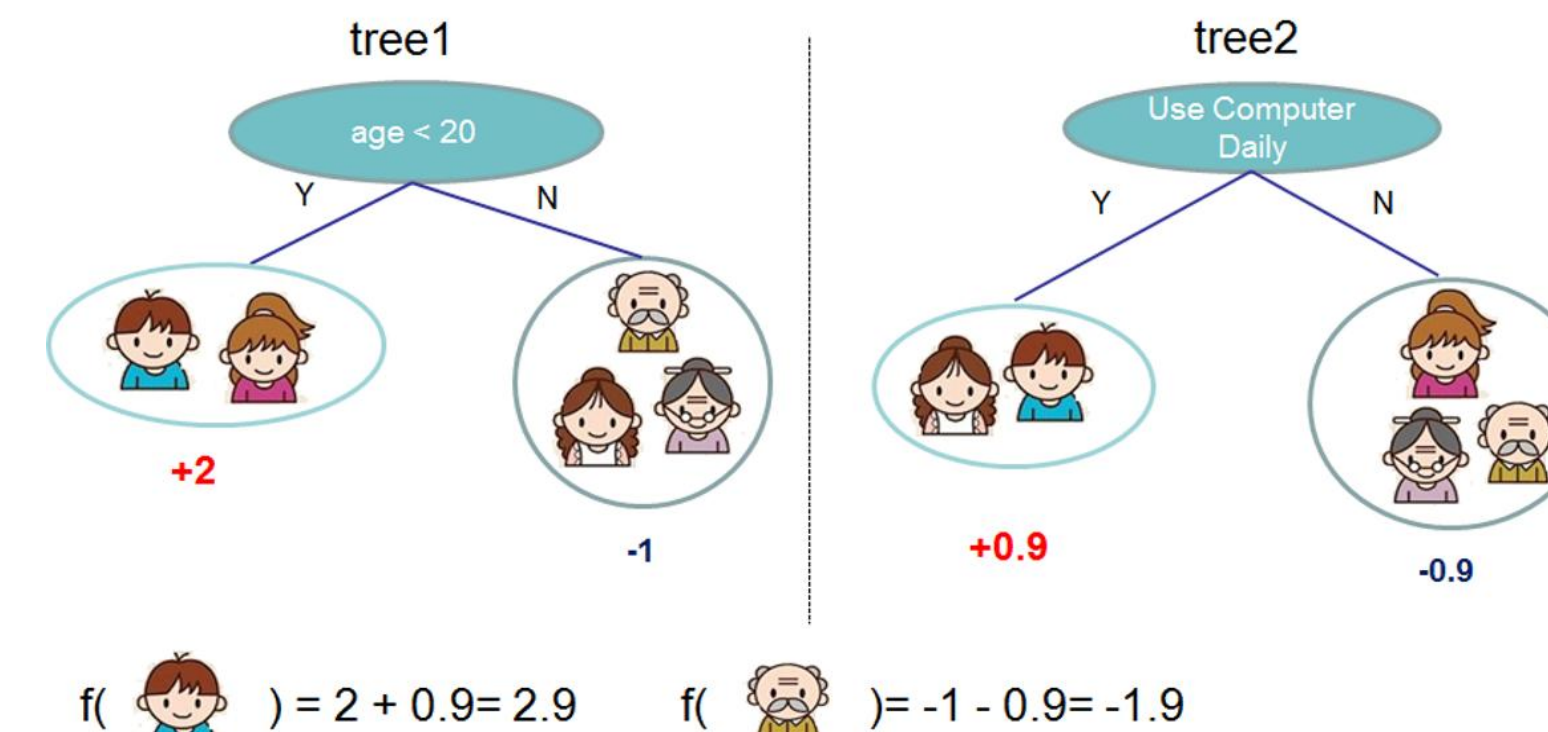


### Double-Dilution Accuracy

- FDA range of correct prediction [1-3]:

$$MIC_{predicted} \in [\frac{MIC_{target}}{2}, MIC_{target} * 2]$$

### XGBoost Model [2]



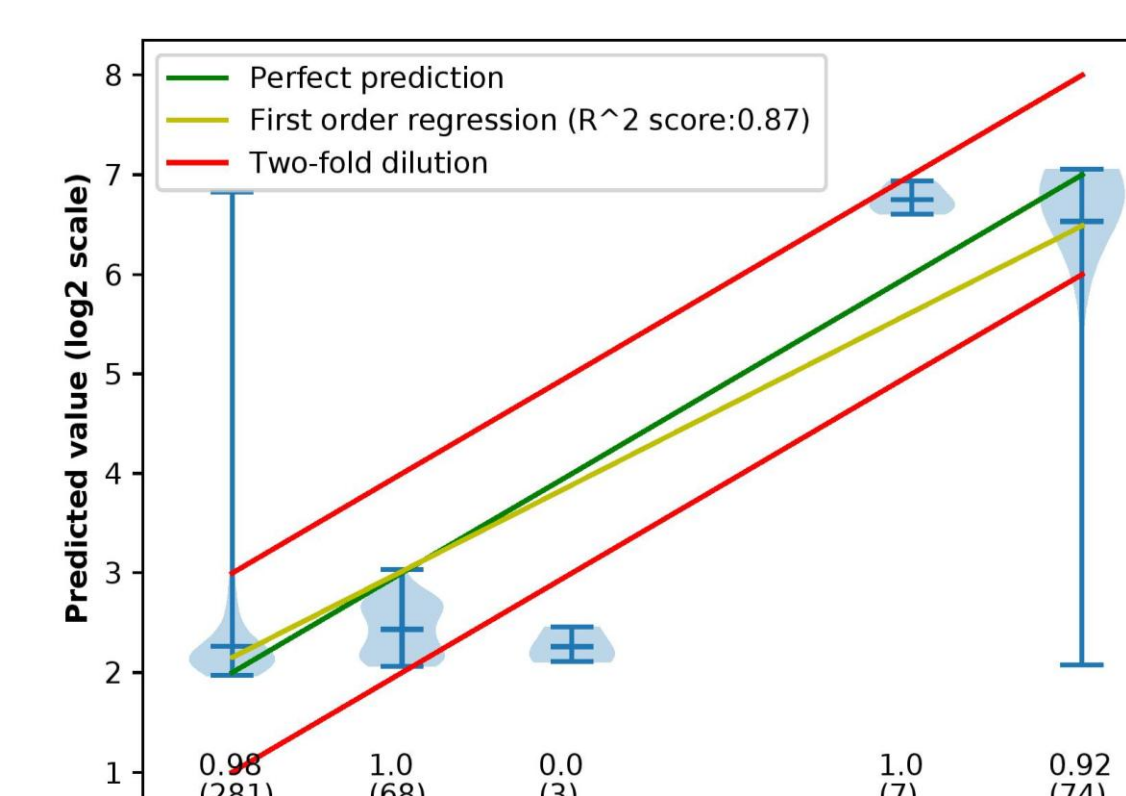
## Results

### Accuracy of prediction

Data	Mean accuracy	Standard Deviation
10-mers NT	0.935	0.012
3-mers AA	0.747	0.038
4-mers AA	0.941	0.012
5-mers AA	0.959	0.008

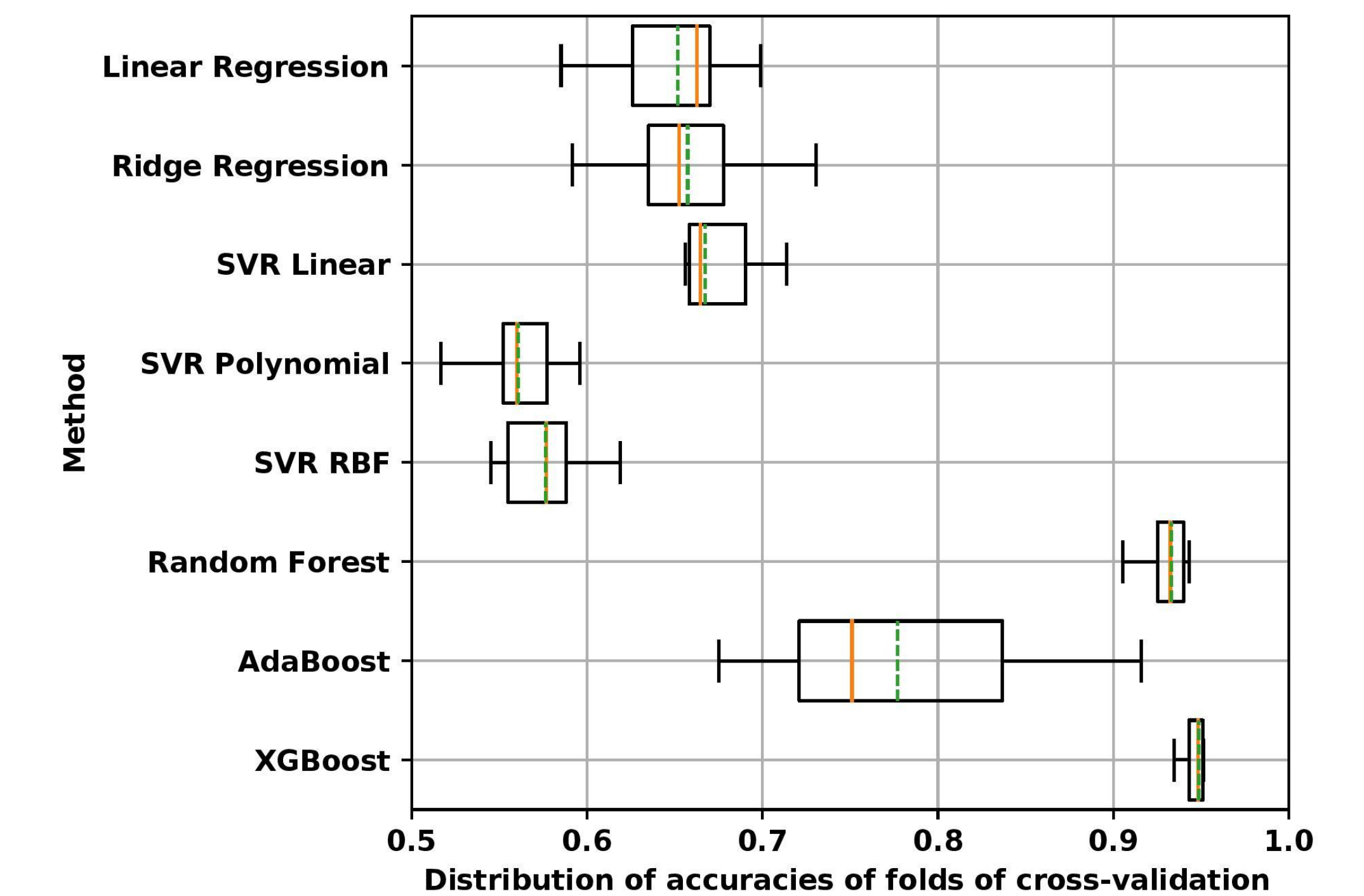
### Accuracy of XGBoost for effect of Ampicillin on *Salmonella enterica*

### Visualization of the predictions



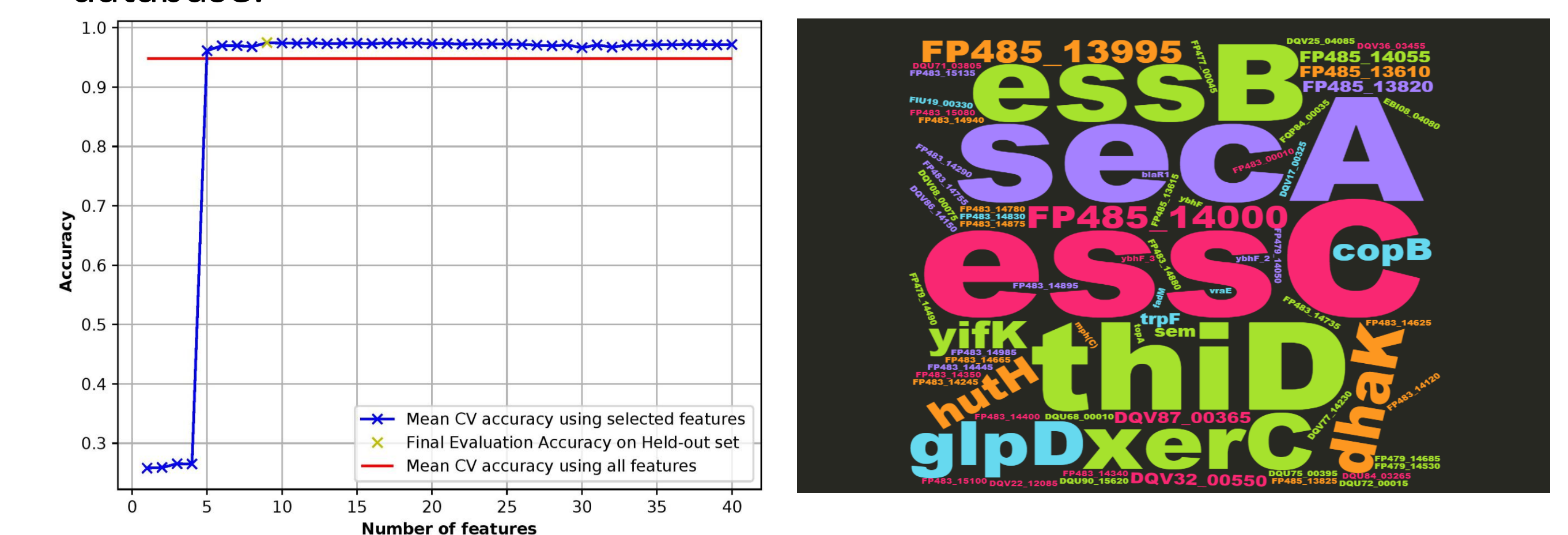
Distribution of prediction of values of MIC (effect of nalidixic acid on *Campilobacter Jejuni*. 4-mers of AA)

## Other machine learning methods

AA4-mers of *Salmonella enterica* ampicillin

## Feature selection

- After training the model, the important features were extracted, and a new model was trained using only the selected features.
- Important genes were found by searching the important k-mers on NCBI database.



Left: Accuracy of the XG Boost model only using a few important features  
(effect of ampicillin on *Salmonella enterica* 4-mers of aminoacidic)  
Right: Important genes that affect AMR for *Staphylococcus aureus* (extracted  
after training the model with 15-mers of nucleotide)

## Conclusions

- Given enough number of genomes, machine learning can be used to predict MIC with a good accuracy
- Instead of nucleotide sequences, amino acid sequences can be used for increasing accuracy
- XG Boost outperforms other methods for MIC prediction from genome
- Feature selection can identify genes that are important to AMR

## Contact

Taha ValizadehAslani  
EESI lab, Drexel University  
DrexelEESI.com  
tv349@drexel.edu

## References

- [1] M. Nguyen, T. Brettin, S.W. Long, J.M. Musser, R.J. Olsen, R. Olson, M. Shukla, R.L. Stevens, F. Xia, H. Yoo, J.J. Davis, Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*, *Sci. Rep.* 8 (2018) 421. doi:10.1038/s41598-017-18972-w.
- [2] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, (n.d.).
- [3] M. Nguyen, S.W. Long, P.F. McDermott, R.J. Olsen, R. Olson, R.L. Stevens, G.H. Tyson, S. Zhao, J.J. Davis, Using Machine Learning To Predict Antimicrobial MICs and Associated Genomic Features for Nontyphoidal *Salmonella*, *J. Clin. Microbiol.* 57 (2018). doi:10.1128/JCM.01260-18.