OVERVIEW

Application of artificial intelligence to the *in silico* assessment of antimicrobial resistance and risks to human and animal health presented by priority enteric bacterial pathogens

Rylan Steinkey[1], Janice Moat[2,3], Victor Gannon[1], Athanasios Zovoilis[3,4,5], Chad Laing[2]*

Abstract

Each year, approximately one in eight Canadians are affected by foodborne illness, either through outbreaks or sporadic illness, with animals being the major reservoir for the pathogens. Whole genome sequence analyses are now routinely implemented by public and animal health laboratories to define epidemiological disease clusters and to identify potential sources of infection. Similarly, a number of bioinformatics tools can be used to identify virulence and antimicrobial resistance (AMR) determinants in the genomes of pathogenic strains.

Many important clinical and phenotypic characteristics of these pathogens can now be predicted using machine learning algorithms applied to whole genome sequence data. In this overview, we compare the ability of support vector machines, gradient-boosted decision trees and artificial neural networks to predict the levels of AMR within *Salmonella enterica* and extended-spectrum ß-lactamase (ESBL) producing *Escherichia coli*. We show that minimum inhibitory concentrations (MIC) for each of 13 antimicrobials for *S. enterica* strains can be accurately determined, and that ESBL-producing *E. coli* strains can be accurately classified as susceptible, intermediate or resistant for each of seven antimicrobials.

In addition to AMR and bacterial populations of greatest risk to human health, artificial intelligence algorithms hold promise as tools to predict other clinically and epidemiologically important phenotypes of enteric pathogens.

Affiliations

[1] National Microbiology Laboratory at Lethbridge, Public Health Agency of Canada, Lethbridge, AB

[2] National Centre for Animal Diseases, Canadian Food Inspection Agency, Lethbridge, AB

[3] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB

[4] Southern Alberta Genome Sciences Centre, Lethbridge, AB

[5] Canadian Centre for Behavioural Neuroscience, Lethbridge, AB

***Correspondence:**
chad.laing@canada.ca

Suggested citation: Steinkey R, Moat J, Gannon V, Zovoilis A, Laing C. Application of artificial intelligence to the *in silico* assessment of antimicrobial resistance and risks to human and animal health presented by priority enteric bacterial pathogens. Can Commun Dis Rep 2020;46(6):180–5. https://doi.org/10.14745/ccdr.v46i06a05

Keywords: machine learning, bacterial pathogens, whole genome sequence, predictive genomics, antimicrobial resistance

Introduction

Every year, about one in eight Canadians will be affected by a foodborne illness, resulting in an average of 11,600 hospitalizations and 238 deaths nationwide (1). Animals are often the reservoir for major bacterial pathogens such as *Salmonella enterica* and *Escherichia coli*. These pathogens are associated with both sporadic cases and outbreaks of foodborne disease. Antimicrobial resistance (AMR) among these organisms is a growing concern, with treatment being more difficult and expensive. For example, extended-spectrum ß-lactamase (ESBL) producing *E. coli* are multidrug resistant, with treatment costs up to three times that of non-ESBL-producing *E. coli* (2).

National and provincial public health agencies are very effective at identifying sources and halting exposure to pathogens. Historically, AMR determination has been performed in a wet lab setting (3,4). Two of the most commonly used diagnostic methods are diffusion and dilution tests. Diffusion methods, such as the Kirby–Bauer method, require growing a bacterial lawn in either a disk of known concentration of antimicrobials or a strip with a gradient of concentrations of antimicrobials; the zone of growth inhibition around the antimicrobial is compared with a standard to determine the resistance of the bacteria (3). Dilution methods involve liquid cultures in serial dilution of

each antimicrobial, where growth of the organism is used to determine the minimum inhibitory concentration (MIC) (3,4).

These methods are time consuming because they rely on the growth of bacteria, and expensive because they require trained personnel and specialized equipment to carry out.

Whole genome sequence (WGS) analyses have become integral to public health work flows. *In silico* tests have largely replaced many costly and time-consuming wet lab tests in outbreak response and routine surveillance (5–7). Artificial intelligence is being increasingly used to analyse these datasets.

Artificial intelligence involves training machines to make predictions based on large amounts of data. It has been used in fields as disparate as handwriting recognition (8) and autonomous weapons systems (9).

Supervised machine learning (ML) better describes the application of artificial intelligence to the prediction of bacterial phenotypes based on WGS data. ML algorithms are trained on known data ("features") and subsequently predict or classify unknown data using the trained models. In general, data used for ML training are application specific and can include images or information about weather or outbreaks of infectious disease. Biological data, and in particular WGS data from populations of organisms, provide an extremely large number of features for training ML models and predicting phenotypes of interest. Use of these algorithms in infectious disease research has not yet been fully exploited but holds significant promise.

ML algorithms have been used to predict important phenotypes such as AMR (10,11) and to determine if different groups of pathogens from the same species pose different risks to human health (12–14). The ability to predict important bacterial phenotypes based solely on WGS data would be of enormous benefit to both Canadian public health and the animal agriculture industry.

In this study, we trained three ML models on WGS data to predict the levels of resistance to 13 antimicrobials in *S. enterica* isolates and to classify ESBL-producing *E. coli* strains as susceptible, intermediate or resistant (SIR) to seven antimicrobials.

Methods

S. enterica WGS were collected from the National Center for Biotechnology Information GenBank. These 5,853 sequences were primarily isolated within North America between 2002 and 2017; the data included 63 serotypes with at least five members, along with phenotypic MICs for 13 antimicrobials (15). WGSs were decomposed into sequence substrings, called *k*-mers, of length 11, and their occurrences were counted using Jellyfish (16).To limit the selection of features to those most associated with the phenotype being examined, we used an ANOVA F-value, keeping the top 1,000 *k*-mers most associated with each antimicrobial agent prior to model training. This feature selection allows the model to focus on statistically important *k*-mers, which can improve accuracy and saves substantial amounts of time and computing resources.

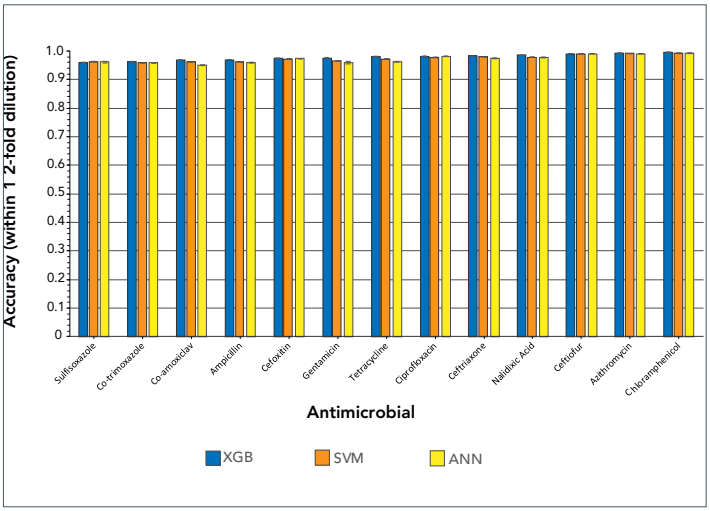
We implemented gradient-boosted decision trees using XGBoost (17) and support vector machines using SciKit-learn (18). Data analyses were conducted using five-fold cross-validation where 80% of the data was used to train a model and the remaining 20% was withheld to evaluate model performance. This was repeated five times, with each 20% being used once for evaluating performance. An average of the accuracy for the five evaluations was calculated for each experimental replicate. Ten separate experimental replicates with random assignment of genomes to each fold were performed, with the total model accuracy and standard deviation calculated from these.

Artificial neural networks were implemented using Keras (19) with a TensorFlow (20) backend and hyperparameter optimizations conducted with Hyperas (21). The five-fold cross-validation for the neural network consisted of a 60-20-20 split for training, hyperparameter optimization and testing, respectively, for each fold. Early stopping mechanisms were used to prevent over-fitting by monitoring diminishing or negative returns with successive training epochs. In addition, a random selection of nodes in the network and their connections were removed via dropout to prevent over-fitting or co-adaptation (22).

As shown in **Figure 1**, MICs were predicted within one dilution with an accuracy of 97.88% (± 1.13) using XGBoost, 97.48% (± 1.20) using support vector machines and 97.16% (± 1.48) using artificial neural networks. XGBoost classifiers averaged a major error and major error rate of 0.19% (± 0.19) and 0.71% (± 0.60), respectively. To prevent inflating model accuracies, co-trimoxazole, ciprofloxacin and ceftriaxone, which had low MIC class diversity, were removed from these averages. XGBoost classifiers trained to predict MICs for a single antimicrobial used eight cores (Intel Xeon Gold 6154 CPU), had a mean training time of 15 minutes and 12 seconds, and peaked at 84.74 GB of random access memory (RAM).

We also examined a set of 2,413 *E. coli* sequences containing ESBL producers, but no MIC data were available for these strains. Instead, they were classified as SIR for seven antimicrobials. The set included bovine, clinical and environmental samples isolated between 1970 and 2017 in Canada, Thailand and the United Kingdom (11,23,24). We analyzed the sequences with the *k*-mer approach described above and used them to train models to classify isolates as SIR for each antimicrobial. The

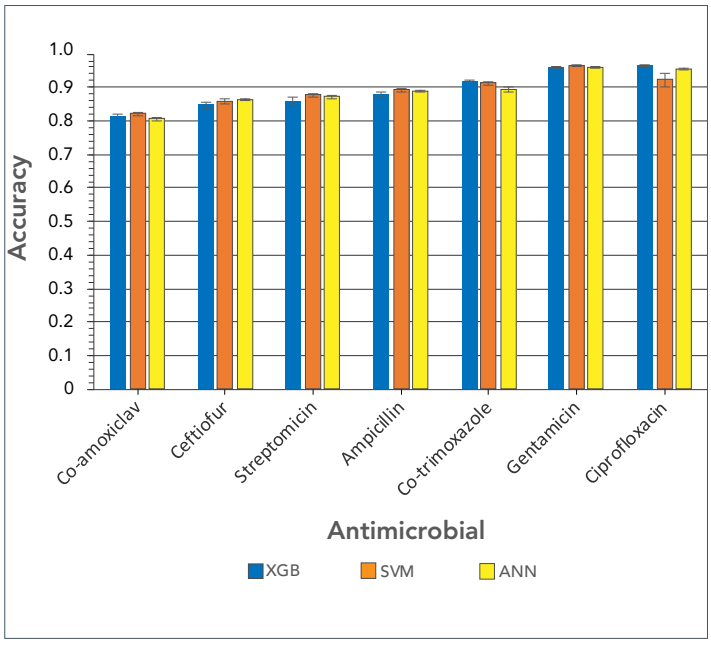
Figure 1: Accuracies within one two-fold dilution for three machine learning models trained on the top 1,000 11-mers and used to predict minimum inhibitory concentrations for 13 *Salmonella enterica* antimicrobials



Abbreviations: ANN, artificial neural network; SVM, support vector machine; XGB, XGBoost

average accuracies of the models across the seven antimicrobials were 89.18% (± 5.44) for XGBoost, 89.25% (± 4.43) for support vector machines and 89.18% (± 5.20) for artificial neural networks (**Figure 2**).

Figure 2: Accuracies of three machine learning models trained on the top 1,000 11-mers, and used to predict susceptible, intermediate and resistant classifications for seven *Escherichia coli* antimicrobials



Abbreviations: ANN, artificial neural network; SVM, support vector machine; XGB, XGBoost

Discussion

As we have shown, the ML methods we employed did not rely on specific reference genomes, or *a priori* knowledge of the mechanisms of resistance, but on the classification of organisms into broad phenotypic groups. It is the ML models that identify the underlying genomic differences that are most associated with the phenotype. This has the double benefit of not requiring mechanistic knowledge and has the potential for identifying novel genomic determinants of the phenotype under study. These novel features extracted from the models have enormous potential benefit: as in the case of AMR, they can be used to grow established public databases of resistance mechanisms, and they can be used as potential targets for rapid diagnostics in subsequent *in silico* or wet lab assays.

ML models can rapidly and accurately predict AMR using WGS data, from SIR classification to quantitative MIC values. For AMR predictions, XGBoost models were shown to train faster, use less memory and be more accurate than deep-learning methods. In addition, XGBoost and support vector machine models can be used to determine the specific regions of the genome that are most predictive of a phenotype. This is very difficult with the "black box" implementation of a neural network; however, artificial neural networks still excel in complicated network modelling and therefore should not be excluded from future studies in genomics.

AMR data typically suffer from substantial class imbalance, which can result in high accuracy models that are of no value, such as the case of co-trimoxazole in our *Salmonella* data, where more than 95% of the samples were within one dilution of each other, resulting in a model capable of 95% accuracy without learning anything from the underlying data.

Nguyen *et al.* (10) trained XGBoost regressors on a dataset containing 4,500 non-typhoidal *S. enterica* whole genome sequences (from a larger dataset of 5,278 samples, of which 4,595 were also in our dataset). These models had a cross-validation accuracy of 95% for the same 10 antimicrobials included in our current study. Nguyen *et al.* (10) used a single regressor trained on all 15 antimicrobials at once, which took 51 hours to train and peaked at 1,184 GB of memory on 170 cores (Intel Xeon E5-4669v4 CPU) (10). The XGBoost classifiers trained in our current study improved upon these training times as well as memory usage and accuracy. The XGBoost classifiers did this by creating per-antimicrobial models and initially selecting only the 1,000 most statistically important features. To better compare the accuracies of these models, an independent dataset should be used instead of relying on the reported cross-validation accuracies.

The *E. coli* dataset included 1,935 isolates from a previous study by Moradigaravand *et al.* (11). Their methods required the isolation year for each sequence and data preprocessing in the

form of pan-genome determination and population structure calculation (11). In contrast, our methods required only the genome sequence paired with laboratory-determined resistance phenotype, which allows classification as well as identification of novel regions not currently known to be associated with AMR. The regions could be used for subsequent *in silico* or wet lab diagnostic tests.

While broader classifications, such as SIR, are common for laboratory diagnostics, and useful for establishing treatment guidelines for a bacterial infection, the breakpoint criteria for these categories are established by committees, with some disparity between regions. The prediction of quantified values in the form of MICs will be of most use in future, even if they are subsequently used for classifying bacteria into broader categories such as SIR.

Though the results of these studies are encouraging, over-interpretation of results is a problem with genomic data due to the high number of features used to make predictions relative to the smaller sample size of the number of genomes. This can lead to over-fitting of data and poor performance of models, both of which we have tried to address in the methods of this study (25).

Use of ML has proved successful for AMR prediction in other pathogens, including *Mycobacterium tuberculosis*, where new resistant genetic signatures were identified (26). ML has also proved useful in the identification of novel antimicrobial compounds, which has historically been fraught with high failure rates in pharmaceutical companies (27).

ML research on *S. typhimurium* found that more than 80% of host source could be attributed using protein variants. This result was obtained using support vector machine (SVM), artificial neural networks and Random Forest models (28). What is particularly interesting from this study is the overlap between the animal reservoir and human cases. This indicates that not all isolates of a particular pathogen represent the same disease risk and suggests that more specific points of control could limit human infection. In addition, as more than 60% of human pathogens are of zoonotic origin, ML holds promise for identifying emerging pathogens by analyses of host adaptation of current animal pathogens (29).

Despite the proven usefulness of ML, bacteria are constantly evolving, and so our models, as they are only as good as the data they are trained on. The power of these techniques must be tempered by their judicious use. In addition, class and species-specific models are still required to generate meaningful results, for example, one model per drug per species for predicting AMR (30).

It should be noted that ML does not always accurately capture complex interactions and that improved modelling alone cannot compensate for sampling bias or an incomplete or error-prone dataset.

Conclusion

As demonstrated in this overview, artificial intelligence has already improved infectious disease identification and characterization, the benefits of which will affect public health and animal health laboratories around the world. For example, genomic regions identified as predictive for specific AMR classes could be used for rapid downstream identification and classification, including *in silico* pipelines and wet lab applications such as polymerase chain reaction.

The near-future promises exciting developments, such as using ML to identify bacteriophages that lyse specific groups of pathogenic bacteria, enabling phage therapy in place of traditional antimicrobials (31). Lastly, "whole phenotype" characterization, with the ability to predict integral membrane protein expression, is becoming more likely (32); and biofilm formation (33).

Despite this, the size of the datasets required to effectively train ML models mean that desktop computers are often incapable of analyzing the data. Those without access to the necessary resources must instead use analytical approaches that reduce the computational burden (34). Fittingly, the use of ML itself has led to an increase in speed of mechanistic models, in some cases over four orders of magnitude (35).

We are just at the beginning of the coupling of vast amounts of genomic data and artificial intelligence, with the promise of new discoveries that will improve most aspects of animal and human health from the burden of enteric bacterial pathogens.

Authors' statement

RJS — Data curation, formal analysis, methodology, software, validation, visualization, original draft, editing
JM — Data curation, formal analysis, methodology, software, validation, visualization, original draft, editing
VPJG — Conceptualization, funding acquisition, methodology, project administration, resources, supervision, validation, original draft, editing
AZ — Conceptualization, funding acquisition, methodology, project administration, resources, supervision, original draft, editing
CRL — Conceptualization, funding acquisition, methodology, project administration, resources, supervision, validation, original draft, editing

Conflict of interest

None.

Funding

JM, AZ, CRL: This work has been supported by the Antimicrobial Resistance – One Health Consortium grant to AZ and CRL from the Alberta Ministry of Economic Development, Trade, and Tourism.

RJS, VPJG, CRL: This work has been supported by the Genomics Research and Development Initiative project on antimicrobial resistance. This work was additionally funded by the Public Health Agency of Canada, the Canadian Food Inspection Agency and the University of Lethbridge.

References

1. Public Health Agency of Canada. Yearly food-borne illness estimates for Canada. Ottawa (ON): Government of Canada; 2015 (updated 2016-07-05). https://www.canada.ca/en/public-health/services/food-borne-illness-canada/yearly-food-borne-illness-estimates-canada.html

2. Maslikowska JA, Walker SA, Elligsen M, Mittmann N, Palmay L, Daneman N, Simor A. Impact of infection with extended-spectrum ß-lactamase-producing Escherichia coli or Klebsiella species on outcome and hospitalization costs. J Hosp Infect 2016;92(1):33–41. DOI PubMed

3. Schumacher A, Vranken T, Malhotra A, Arts JJ, Habibovic P. In vitro antimicrobial susceptibility testing methods: agar dilution to 3D tissue-engineered models. Eur J Clin Microbiol Infect Dis 2018;37(2):187–208. DOI PubMed

4. Andrews JM. Determination of minimum inhibitory concentrations. J Antimicrob Chemother 2001;48 (1 Suppl 1):5–16. DOI PubMed

5. Collineau L, Boerlin P, Carson CA, Chapman B, Fazil A, Hetman B, McEwen SA, Parmley EJ, Reid-Smith RJ, Taboada EN, Smith BA. Integrating whole-genome sequencing data into quantitative risk assessment of foodborne antimicrobial resistance: a review of opportunities and challenges. Front Microbiol 2019;10:1107. DOI PubMed

6. Besser JM, Carleton HA, Trees E, Stroika SG, Hise K, Wise M, Gerner-Smidt P. Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak investigation. Foodborne Pathog Dis 2019;16(7):504–12. DOI PubMed

7. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using genomics to track global antimicrobial resistance. Front Public Health 2019;7:242. DOI PubMed

8. Muehlberger G, Seaward L, Terras M, Ares Oliveira S, Bosch V, Bryan M, Colutto S, Déjean H, Diem M, Fiel S, Gatos B, Greinoecker A, Grüning T, Hackl G, Haukkovaara V, Heyer G, Hirvonen L, Hodel T, Jokinen M, Kahle P, Kallio M, Kaplan F, Kleber F, Labahn R, Lang EM, Laube S, Leifert G, Louloudis G, McNicholl R, Meunier JL, Michael J, Mühlbauer E, Philipp N, Pratikakis I, Puigcerver Pérez J, Putz H, Retsinas G, Romero V, Sablatnig R, Sánchez JA, Schofield P, Sfikas G, Sieber C, Stamatopoulos N, Strauß T, Terbul T, Toselli AH, Ulreich B, Villegas M, Vidal E, Walcher J, Weidemann M, Wurster H, Zagoris K. Transforming scholarship in the archives through handwritten text recognition: transkribus as a case study. J Doc 2019;75:954–76. DOI

9. Sharkey A. Autonomous weapons systems, killer robots and human dignity. Ethics Inf Technol 2019;21:75–87. DOI

10. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davis JJ. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. J Clin Microbiol 2019;57(2):e01260-18. DOI PubMed

11. Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLOS Comput Biol 2018;14(12):e1006258. DOI PubMed

12. Fisch D, Yakimovich A, Clough B, Wright J, Bunyan M, Howell M, Mercer J, Frickel E. Defining host-pathogen interactions employing an artificial intelligence workflow. eLife 2019;8:e40560. DOI PubMed

13. Lupolova N, Dallman TJ, Matthews L, Bono JL, Gally DL. Support vector machine applied to predict the zoonotic potential of E. coli O157 cattle isolates. Proc Natl Acad Sci USA 2016;113(40):11312–7. DOI PubMed

14. Lupolova N, Dallman TJ, Holden NJ, Gally DL. Patchy promiscuity: machine learning applied to predict the host specificity of Salmonella enterica and Escherichia coli. Microb Genom 2017;3(10):e000135. DOI PubMed

15. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2013;41(Database issue):D36–42. DOI PubMed

16. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 2011;27(6):764–70. DOI PubMed

17. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York (NY): ACM; 2016. pp. 785–94. DOI

18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30. http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

19. Chollet F. Keras. GitHub repository; 2015. https://github.com/fchollet/keras

20. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X. TensorFlow: A system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation. 2016 Nov 2–4. Savannah (GA): OSDI 16. pp. 265–83. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

21. Pumperla M. Keras + Hyperopt: A very simple wrapper for convenient hyperparameter optimization: Maxpumperla/Hyperas. 2019 (accessed 2020-03-25). http://maxpumperla.com/hyperas/

22. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929–58. http://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf

23. Runcharoen C, Raven KE, Reuter S, Kallonen T, Paksanont S, Thammachote J, Anun S, Blane B, Parkhill J, Peacock SJ, Chantratita N. Whole genome sequencing of ESBL-producing Escherichia coli isolated from patients, farm waste and canals in Thailand. Genome Med 2017;9(1):81. DOI PubMed

24. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res 2017;27(8):1437–49. DOI PubMed

25. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer Science & Business Media; 2009. DOI

26. Kavvas ES, Catoiu E, Mih N, Yurkovich JT, Seif Y, Dillon N, Heckmann D, Anand A, Yang L, Nizet V, Monk JM, Palsson BO. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. Nat Commun 2018;9(4306):1-9. DOI PubMed

27. Ivanenkov YA, Zhavoronkov A, Yamidanov RS, Osterman IA, Sergiev PV, Aladinskiy VA, Aladinskaya AV, Terentiev VA, Veselov MS, Ayginin AA, Kartsev VG, Skvortsov DA, Chemeris AV, Baimiev AK, Sofronova AA, Malyshev AS, Filkov GI, Bezrukov DS, Zagribelnyy BA, Putin EO, Puchinina MM, Dontsova OA. Identification of novel antibacterials using machine learning techniques. Front Pharmacol 2019;10:913. DOI PubMed

28. Lupolova N, Lycett SJ, Gally DL. A guide to machine learning for bacterial host attribution using genome sequence data. Microb Genom 2019 Dec;5(12):5. DOI PubMed

29. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. Nat Rev Genet 2018 Sep;19(9):549–65. DOI PubMed

30. Hicks AL, Wheeler N, Sánchez-Busó L, Rakeman JL, Harris SR, Grad YH. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLOS Comput Biol 2019 Sep;15(9):e1007349. DOI PubMed

31. Leite DMC, Brochet X, Resch G, Que Y-A, Neves A, Peña-Reyes C. Computational prediction of inter-species relationships through omics data analysis and machine learning. BMC Bioinformatics 2018;19(S14 Suppl 14):420. DOI

32. Saladi SM, Javed N, Müller A, Clemons WM Jr. A statistical model for improved membrane protein expression using sequence-derived features. J Biol Chem 2018 Mar;293(13):4913–27. DOI PubMed

33. Yan J, Deforet M, Boyle KE, Rahman R, Liang R, Okegbe C, Dietrich LE, Qiu W, Xavier JB. Bow-tie signaling in c-di-GMP: machine learning in a simple biochemical network. PLOS Comput Biol 2017 Aug;13(8):e1005677. DOI PubMed

34. Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. Sci Rep 2019;9(1):4071. DOI PubMed

35. Wang S, Fan K, Luo N, Cao Y, Wu F, Zhang C, Heller KA, You L. Massive computational acceleration by using neural networks to emulate mechanism-based biological models. Nat Commun 2019 Sep;10(1):4354. DOI PubMed