

Developing omics tool for food safety

Predicting virulence and disinfectant resistance in dairy related bacteria using machine learning and omics technology.

Listeria monocytogenes (Lm) is one of the most important foodborne bacterial pathogens. Contaminated food has been estimated as the source in as many as 99% of the cases. Dairy products can become contaminated at several stages along the food production chain. Specific and rapid identification, including elimination of the virulent Lm in real-time is essential for food safety management. The advent of omics technology such as whole genome sequencing (WGS) technologies provides an ideal tool for rapid, reproducible and highly discriminatory characterization of pathogens and creates the opportunity to take risk-based decision making. Moreover, this approach can be applied to devise sanitation strategies for elimination of Lm by identification of disinfectant-resistant strains from the omics data.

Why omics technology and machine learning?

The identification techniques used today are time-consuming, and more importantly lack the resolution to differentiate between bacterial strains of the same species. Such strain-level diversity could affect the recognition of the risk level and decision of consequent actions. These limitations can be resolved by omics technology. The lack of genetic insight and access to advanced computing facilities is a major hurdle for dairy companies to apply omics technology in practice. A smart tool to predict virulence level and disinfectants resistance is needed for frontline safety/quality assurance. Machine learning (ML) involves the development and application of computer algorithms that improve with experience. These techniques further enable the identification of the combinations of predictors that will best predict risk outcome,

thus allowing further refinement of risk assessment inputs from big datasets to a fewer number of predictors (Figure 1).

Objectives of the project

This project aims to use omics technology and ML to improve food safety management of Lm by determining the predictors of virulence level and identifying genetic markers of disinfectant resistance in Lm using WGS data and ML. The project also aims to construct a rapid and user-friendly omics pipeline for detecting the virulence and disinfectant-resistance of Lm.

Development of the omics tool

The first parts of the project are collecting Lm strains and whole genome sequencing including phenotypic test for disinfectant resistance. Later, the sequence data are used as training and testing dataset for developing a ML model. The disinfectant resistance results



BY RESEARCHER PIMLAPAS LEEKITCHAROENPHON, PHD STUDENT ALEXANDER GMEINER, POSTDOC MIRENA IVANOVA, RESEARCHER PATRICK MURIGU KAMAU NJAGE, PROFESSOR LISBETH TRUELSTRUP HANSEN, NATIONAL FOOD INSTITUTE, TECHNICAL UNIVERSITY OF DENMARK, SENIOR SCIENTIST YINGHUA XIAO, ARLA FOODS INNOVATION CENTRE, AARHUS.



Project info

Project title: Improving food safety using WGS and machine learning
Project manager: Researcher Pimlapas Leekitcharoenphon, DTU Food, Technical University of Denmark, pile@food.dtu.dk, Phone: 35 88 71 83
Participants: Arla Foods
Project period: 2020 – 2023
Aim: To develop an omics tool for improving food safety by predicting virulence level and disinfectant resistance – starting with *Listeria monocytogenes*.

THE DANISH DAIRY RESEARCH FOUNDATION

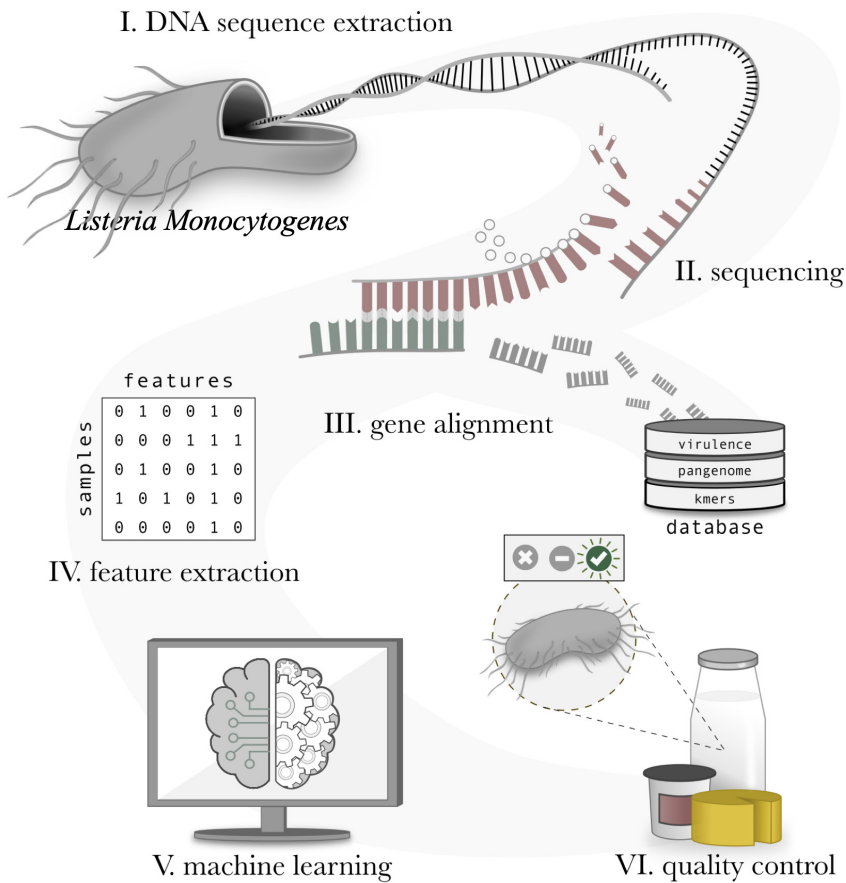


Figure 1 An illustration of using omics technology and ML for improving food safety.

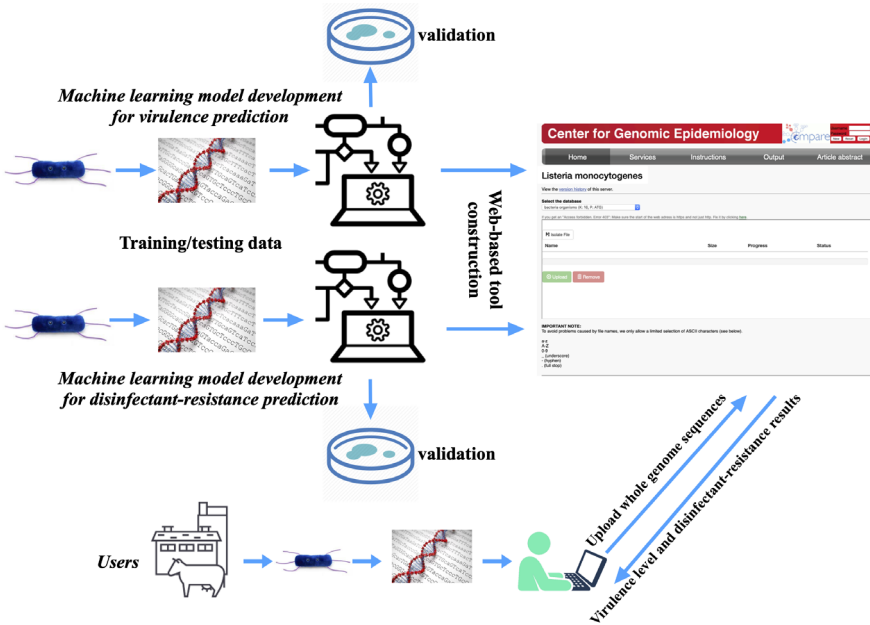


Figure 2 The development of omics tool for predicting virulence and disinfectant resistance.

Summary

Considering the importance of *Listeria monocytogenes* to the dairy industry, this project, which is supported by the Milk Levy Fund, aims to incorporate omics data of more than 1600 clinical and dairy isolates in machine-learning predictors to identify the virulence level and disinfectant-resistance of Lm strains in real-time. With the developed tool, the industrial users only need to upload sequences data to a freely accessible web-based interface, in order to receive easily interpretable output within 15 minutes. This project will improve and speed up the decision-making and food safety management of the dairy industry.

and the virulence level, acquired from governmental surveillance, are used as outcome in the ML model. Different genetic markers in Lm genomes are used as input in the ML model to verify which markers are the best predictors. The input data are partitioned into training, testing and validation datasets and run through available ensemble classification techniques, which will be followed by model evaluation and selection. The best models are incorporated in a pipeline as web-based and standalone tool for predicting virulence level and disinfectant resistance from omics data of Lm strains (Figure 2).

How can the industry benefit from this project?

An innovative, rapid, and cost-effective omics tool for Lm to determine virulence and resistance to disinfectants is created. The user-friendly, web-based tool is freely accessible online and requires no prior knowledge of bioinformatics for interpretation of output. Alternatively, the user can access the tool via a standalone version. The tool can be used for all kinds of sequencing technologies including long read technology, which can sequence a bacterial genome within 6-7 hours. The tool can be used for all food industries concerned with Lm contamination. The idea and workflow from this project can be extended to cover other important foodborne pathogens. ●