

| Project Details |  |
|-----------------|--|
| Project Code    | MRCPHS24Br Reyher  |
| Title           | Genomic Surveillance Meets Machine Learning: Predicting the Origins of Salmonella Outbreaks with Machine Learning  |
| Research Theme  | Population Health Sciences   |
| Summary         | Gastrointestinal disease caused by the consumption of contaminated food is a major public health concern, requiring significant resources to identify the source of infections. The successful candidate will work with the UK Health Security Agency to develop cutting-edge genomic machine learning models to predict the source of foodborne pathogens. These tools will then be used to support public health decision making and facilitate rapid responses to future outbreaks.   |
| Description     | <p>Salmonella enterica is a leading cause of human gastroenteritis worldwide, with non-typhoidal Salmonella estimated to account for ~1 billion infections and ~150,000 deaths annually. This gastrointestinal pathogen therefore represents a major public health concern, necessitating real-time epidemiological monitoring and follow-up. Outbreak investigations, however, are often confounded by the complexity of international food-trade networks which distribute zoonotic food-borne pathogens across the globe. This project aims to address this gap by utilising machine learning (ML) to predict the geographical source of gastrointestinal pathogens directly from genomic surveillance data, allowing for improved public health response and more rapid outbreak resolution. Public health agencies, such as the US Centers for Disease Control (CDC), UK Health Security Agency (UKHSA) and Public Health Agency of Canada (PHAC), routinely apply whole genome sequencing (WGS) to clinically identified cases of Salmonella alongside collecting relevant metadata. WGS contains contextual genetic information on geographical origin, of particular relevance to foodborne pathogens due to their presence in complex foodstuffs consisting of ingredients from multiple sources. However, traditional methods for inferring geographical origin from WGS demand extensive expertise and high computational costs while not scaling effectively. We recently established that ML is an effective tool for the geographical source attribution of Salmonella Enteritidis (PMID: 37042517), the most prominent cause of foodborne illness in the UK. This studentship will build upon this methodological foundation, working alongside public health specialists from the UKHSA, PHAC and CDC to synthesis international genomic datasets for the top three Salmonella species. Outcomes will result in rapid and accurate geographical source attribution models suitable for immediate integration with public health agencies to enhance existing disease management responses. This will be achieved via three primary objectives: a) Co-produce knowledge required for effective source attribution In order to deliver a robust decision-support tool for epidemiologists, automated predictions should be both accurate and understandable by end-users. The student will regularly meet with expert stakeholders in UKHSA, CDC and PHAC and ensure the predictive ML frameworks developed in objectives (b) and (c) contain information relevant to epidemiological follow-up and end-user interpretation. b) Investigate the phylogeographical signal in Salmonella genomes The student will collate genomic surveillance</p> |

|                         |  |
|-------------------------|--|
|                         | <p>datasets of Salmonella enterica serovars Enteritidis, Typhimurium and Newport provided to the project by the UKHSA, CDC and PHAC, and will perform an analysis of phylogeographical signal (i.e. how clustered the genomic data are by geographical origin). This will provide actionable insights into Salmonella species, allowing for the flagging of regionally restricted clones highly-suitable for ML classification as well as problematic international clones which require enhanced classification methods in (c). c) Optimise source attribution models for Salmonella species Using the samples collected in (b), the student will build source attribution models using hierarchical ML and deep-learning frameworks to predict the geographical sources of outbreaks. These models will use state-of-the-art explainable ML approaches to facilitate rapid, targeted outbreak responses. The project would suit an applicant with a strong first degree or masters involving bioinformatics or computational biology and will be tailored to the student; those with a mathematical background open to learning skills in bioinformatics and ML or those with a biological/biomedical background and experience in the areas of basic programming, data science or ML will also be considered.</p> |
| <b>Supervisory Team</b> |  |
| <b>Lead Supervisor</b>  |  |
| Name                    | Professor Kristen Reyher   |
| Affiliation             | Bristol  |
| College/Faculty         | Faculty of Health Sciences   |
| Department/School       | Bristol Veterinary School  |
| Email Address           | kristen.reyher@bristol.ac.uk   |
| <b>Co-Supervisor 1</b>  |  |
| Name                    | Dr Sion Bayliss  |
| Affiliation             | Bristol  |
| College/Faculty         | Faculty of Health Sciences   |
| Department/School       | Bristol Veterinary School  |
| <b>Co-Supervisor 2</b>  |  |
| Name                    | Professor Andrew Dowsey  |
| Affiliation             | Bristol  |
| College/Faculty         | Faculty of Health Sciences   |
| Department/School       | Department of Population Health Sciences & Bristol Veterinary School   |
| <b>Co-Supervisor 3</b>  |  |
| Name                    | Dr Lauren Cowley   |
| Affiliation             | Bath   |
| College/Faculty         | Department of Life Sciences  |
| Department/School       | Milner Centre for Evolution  |