

APPROVED: 31 January 2022

doi: 10.2903/j.efsa.2022.e200417

## Integration of genomics in surveillance and risk assessment for outbreak investigation

Vincenzo Pennone, José Francisco Cobo-Díaz, Miguel Prieto-Maradona and Avelino Álvarez-Ordóñez

Universidad de León, Spain

### Abstract

Keeping food safe is a challenge that needs continuous surveillance for the sake of consumers' health. The main issue when a food-borne pathogen outbreak occurs is represented by the identification of the source(s) of contamination. Delivering this information in a timely manner helps to control the problem, with positive outcomes for everyone, especially for the consumers, whose health is in this way preserved, and for the stakeholders involved in food production and distribution, who could face enormous economic losses if recalls or legal issues occur. Whole genome sequencing (WGS) is a tool recently implemented for the characterisation of isolates and the study of outbreaks because of its higher efficiency and faster results, when compared to traditional typing methods. Lower sequencing costs and the development of many bioinformatic tools helped its spread, and much more attention has been given to its use for outbreak investigation. It is important to reach a certain level of standardisation, though, for ensuring result reproducibility and interoperability. Moreover, nowadays it is possible, if not mandatory for Open Science Practices, to share WGS data in publicly available databases, where raw reads, assembled genomes and their corresponding metadata can be easily found and downloaded. The scope of this Fellowship was to provide the Fellow all the training necessary for successfully integrating genomics to surveillance and risk assessment of food-borne pathogens from farm to fork.

© 2022 Wiley-VCH Verlag GmbH & Co. KGaA on behalf of the European Food Safety Authority.

**Keywords:** whole genome sequencing, outbreak, antimicrobial resistance, food safety, surveillance, risk assessment, farm to fork

**Correspondence:** eu-fora@efsa.europa.eu

**Declarations of interest:** The declarations of interest of all scientific experts active in EFSA's work are available at <https://ess.efsa.europa.eu/doi/doiweb/doisearch>.

**Acknowledgements:** This report is funded by EFSA as part of the EU-FORA programme.

**Suggested citation:** Pennone V, Cobo-Díaz JF, Prieto-Maradona M and Álvarez-Ordóñez A, 2022. Integration of genomics in surveillance and risk assessment for outbreak investigation. EFSA Journal 2022;20(S1):e200417, 10 pp. <https://doi.org/10.2903/j.efsa.2022.e200417>

**ISSN:** 1831-4732

© 2022 Wiley-VCH Verlag GmbH & Co. KGaA on behalf of the European Food Safety Authority.

This is an open access article under the terms of the [Creative Commons Attribution-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.



The EFSA Journal is a publication of the European Food Safety Authority, a European agency funded by the European Union.



## Table of contents

Abstract.....	1
1. Introduction.....	4
2. Description of work programme .....	4
2.1. Aims.....	4
2.2. Activities/Methods .....	5
3. Conclusions.....	6
3.1. Future goals .....	6
References.....	7
Abbreviations.....	7
Appendix A – Listeriosis outbreak simulation.....	9

## 1. Introduction

Whole genome sequencing (WGS) is becoming a main tool for outbreak investigation because it has higher efficiency in identification and characterisation of food-borne microorganisms than other traditional typing methods (Koutsoumanis et al., 2019). Compared to these traditional methods, such as pulsed field gel electrophoresis (PFGE), multiple locus variable number tandem repeat Analysis (MLVA), random amplified polymorphic DNA (RAPD) analysis, variable number tandem repeat (VNTR) analysis or multilocus sequence typing (MLST), WGS delivers outputs at higher resolution and in a shorter time. In addition, depending on the bioinformatic application(s) used and on the performance of the machine, it can provide the results of multiple tests in one single assay (e.g. identification of virulence and antimicrobial resistance (AMR) genes, and other phenotype predictions based on genotype), which speeds up the emergence response in case of an outbreak. Recently, WGS has been used for the detection of pathogens in several outbreaks with great impact for public health. For example, WGS was used to identify the likely source of the largest known outbreak of *Listeria monocytogenes*, which happened in South Africa between 2017 and 2018 (Smith et al., 2019) and in the EU multi-country outbreak of *L. monocytogenes* ST6 linked to frozen corn (EFSA and ECDC, 2018). Another concern is the possibility of dissemination of AMR- and virulence-related genes via the food chain. WGS is also helpful in this case, allowing to rapidly characterise biological determinants related to AMR, virulence, mobile genetic elements (MGEs), and their dissemination patterns, which contributes to the protection of public health with respect to food-borne diseases. The 'One Health' approach is nowadays widely recognised for investigating AMR and how it spreads across all sectors, for example in hospitals or through animal farming (Hernando-Amado et al., 2019). In addition, dose–response models would benefit from a better understanding on the virulence potential of certain strains. Several recent studies have demonstrated the promise of routine WGS of bacterial pathogens for epidemiological surveillance, outbreak detection, and infection control. For example, Neuert et al. (2018) used WGS to identify genetic traits responsible for phenotypic AMR in 3,491 non-typhoidal *Salmonella enterica* isolates. In addition, WGS can be used to track the occurrence and distribution of these genetic traits leading to AMR or virulence in different environments, including foods, food-related environments and clinical specimens, facilitating source attribution. If integrated with metadata gathered from food or clinical samples, genomic data can be implemented within quantitative risk assessment frameworks by including statistical analyses and mathematical modelling of resistance and virulence determinants occurrence and dissemination. A cross-sectorial platform developed in an EFSA funded project (INNUENDO project) is available and allows to identify flaws and needs in data flow during outbreak investigation and routine implementation of WGS in molecular epidemiology of food-borne pathogens, providing information to solve outbreaks and enhancing scientific cooperation between the food, veterinary and human health sectors. WGS can thus provide to national and international regulatory agencies and researchers a framework for the evaluation and communication of risks linked to foods. Furthermore, it facilitates the investigation of outbreaks and the actualisation of measures for risk reduction. Thanks to the application of open science practices, all the WGS data published in the scientific literature are publicly available in online databases. These data can be used for many different additional investigations, such as to further characterise isolates responsible for outbreaks, or to study the occurrence of virulence and antimicrobial genes in specific regions or in a specific time frame.

Strengthening global surveillance of food-borne pathogens and their related characteristics (e.g. virulence or AMR potential), is critical as it sets the basis for developing global strategies, monitoring the effectiveness of public health interventions and detecting new trends and emerging threats.

## 2. Description of work programme

### 2.1. Aims

The aim of this fellowship was to prepare the fellow on exploring the potential of next-generation sequencing (in particular, of WGS) methodologies as a tool for surveillance of food-borne pathogens, AMR and virulence genes. The core of the work was related to the study of publicly available WGS data of some of the most common food-borne pathogens, to explore the characteristics of their resistome and associate the presence of AMR genes to metadata, such as geographical data, isolation source and temporal distribution.

## 2.2. Activities/Methods

To achieve the objectives mentioned in Section 2.1, the fellow received training on the most common practices and protocols used for WGS and whole metagenome sequencing (WMS) by experienced scientists. Furthermore, an outbreak simulation was performed to assess the fellow's acquired skills on an outbreak investigation, and training on some of the most common software used for risk assessment was delivered. The activities related to each objective of this fellowship are described below.

**Objective 1:** Training of the fellow on general risk assessment methodologies routinely used by the mentor and other collaborators at the host institution. Extensive training was arranged about the functionalities of the software *Oracle Crystal Ball*, used by the host institution for performing Monte Carlo simulations and predictive modelling. The training and the analyses performed consisted in four phases: variable definition, model development, simulation and uncertainty analysis (Figure 1). Practical examples were explored by the fellow and discussed with the tutors, e.g. to predict the reductions in microbial counts following the pasteurisation of milk, or to assess the exposure of European consumers to extended-spectrum  $\beta$ -lactamase (ESBL)-producing *Escherichia coli* through the consumption of pork meat.

**Objective 2:** Training of the fellow on different integrated platforms for the use of genomics in food-borne pathogen surveillance and outbreak investigation. An intensive hands-on training was delivered to the fellow on the use of the most common applications for the analysis of WGS data. In particular, the fellow learnt how to use Python- and Pearl-based software (e.g. *StarAMR*, *ResFinder*, *PointFinder*, *PlasmidFinder*, *PlasFlow*, *MLST*, *dRep*, *Tormes*) launched with command lines and adapted Ruby scripts for parallel analyses on multiple genomes, using Linux and Unix environments. The fellow was also trained on the use of a server owned by the hosting site for analyses that required high-performing computing. Eventually, a training on the use of specific R packages for statistical analyses and charts preparation (*dplyr*, *ggplot2*, *phreatmap*, *ggpubr*, *vegan*, *tidyr*) was given to the fellow.

**Objective 3:** *In silico* analysis of genomes from food-borne pathogens.

The fellow analysed about 30,000 *Staphylococcus aureus* genomes and their metadata downloaded from publicly available repositories describing the distribution of antimicrobial resistance genes (ARGs) among clonal complexes (CCs), geographical regions, isolation sources and time periods. The resistome of *S. aureus* was described and trends highlighted. Furthermore, the location of ARGs on MGEs (plasmids) was described and analysed. The workflow of this analysis is summarised in Figure 2.

A similar analysis was also conducted on about 3,000 *Salmonella enterica* serovar Enteritidis and 3,000 *Salmonella enterica* serovar Typhimurium genomes.

Two manuscripts reporting the main results of these activities are currently in preparation.

**Objective 4:** Assessment of the potential of WGS in outbreak investigation.

A case study of an outbreak was prepared by the tutors and investigated by the fellow, using WGS data and metadata about the sources of samples, to determine the source(s) of the simulated outbreak (Appendix A).

**Objective 5:** Collaboration of the fellow in dissemination and outreach activities.

The fellow has worked on a literature review on the use of WGS for improving food safety which has been submitted to the Food Microbiology section of *Current Opinion in Food Science*. A semi-systematic literature review on the use of WGS for outbreak investigation has also been prepared. Furthermore, a manuscript on the *S. aureus* resistome analysis has been already prepared and it is ready for being submitted for peer review. One more manuscript on the analysis of *Salmonella* Enteritidis and *S. Typhimurium* is also in preparation.

The fellow will also present the *S. aureus* resistome analysis at the next ONE – Health, Environment, Society – Conference, in Brussels, 21–24 June 2022.

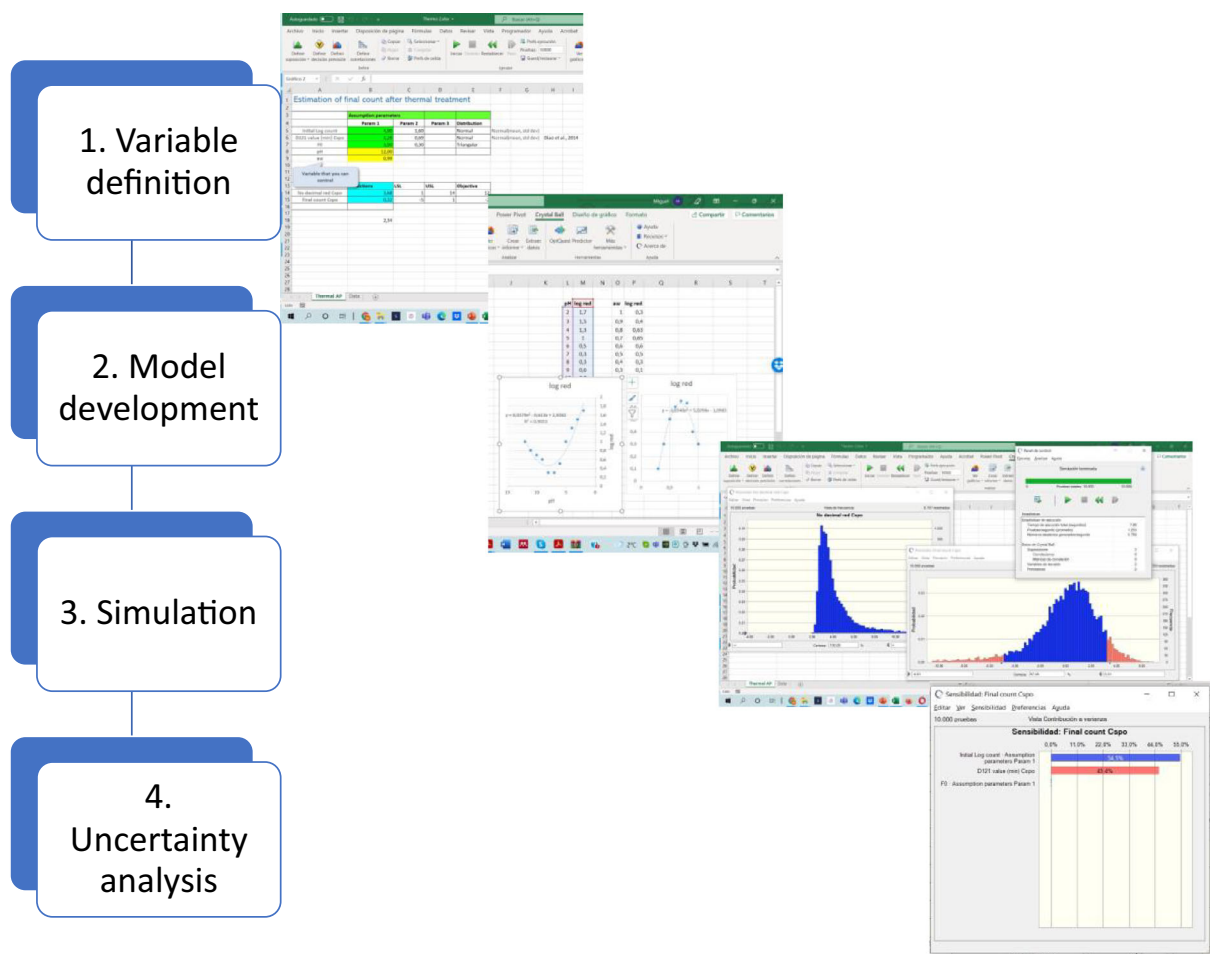
Additional activities were also undertaken by the fellow during this year at the host institution. For example, the fellow attended online courses to improve his skills on Python and Linux commands and applied for the Marie Skłodowska-Curie Actions (MSCA) Postdoctoral Fellowships 2021. The fellow was fully integrated into the working team of his host institution and participated to lab meetings in which he presented his work to other colleagues. The fellow also received an invitation for a visit, with the other EU-FORA fellows hosted in Spain, to the Spanish Food Safety and Nutrition Agency (AESAN) in Madrid and Majadahonda which took place at the end of October 2021.

### 3. Conclusions

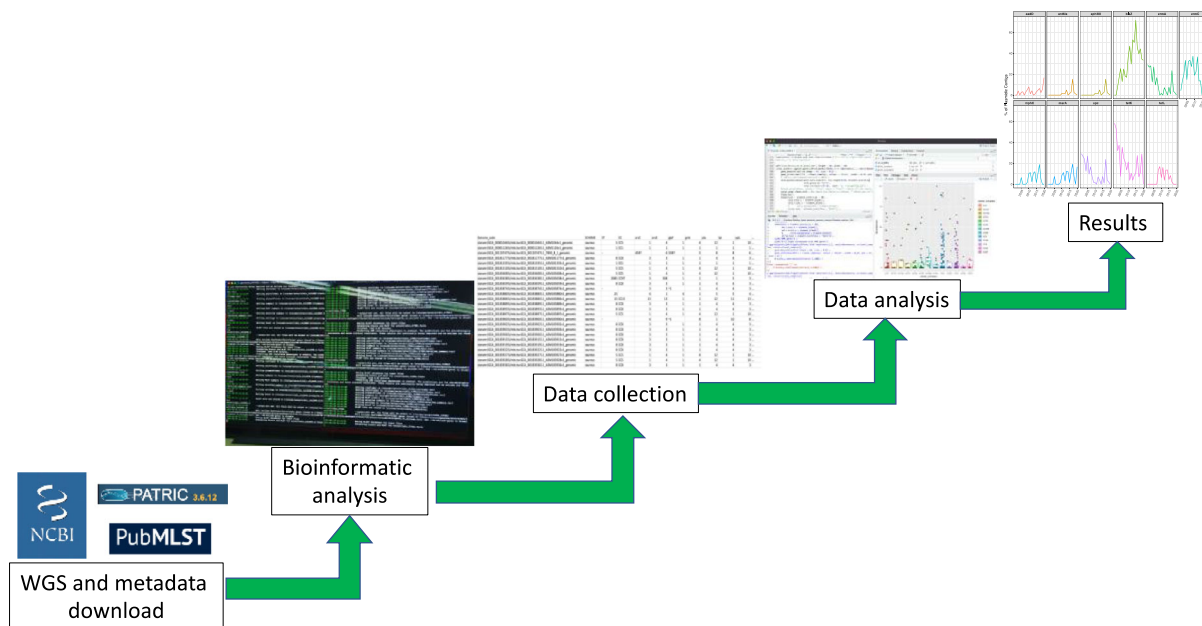
This fellowship transferred to the fellow’s new skills on the use of WGS for the study of outbreaks and the characterisation of AMR in food-borne pathogens. Outputs derived from this fellowship will be published in peer-reviewed journals, which will include a literature review on the use of WGS for food safety, a semi-systematic review on the use of WGS for outbreak investigation, and two research papers, one on the analysis of the resistome of *S. aureus* and one on the analysis of the resistome of *S. Enteritidis* and *S. Typhimurium*. The fellow is actively working to finalise the remaining manuscripts on a timely manner. Furthermore, the fellow had the opportunity to familiarise with tools for Risk assessment and Monte Carlo simulations.

#### 3.1. Future goals

The cooperation between the fellow and the hosting site will be maintained in the future. A Marie Curie fellowship application has been already submitted by the fellow and the host organisation, and hopefully more projects will be shared in the future to maintain this fruitful collaboration.



**Figure 1:** Workflow used for Monte Carlo simulation and predictive modelling with *Oracle Crystal Ball*



**Figure 2:** Workflow of the analyses performed in Objective 3

## References

- EFSA and ECDC (European Food Safety Authority and European Centre for Disease Prevention and Control), 2018. Multi-country outbreak of *Listeria monocytogenes* serogroup IVb, multi-locus sequence type 6, infections linked to frozen corn and possibly to other frozen vegetables – first update. EFSA Supporting Publications 2018;15(7): en-1448, 55 pp. <https://doi.org/10.2903/sp.efsa.2018.EN-1448>
- Hernando-Amado S, Coque TM, Baquero F and Martínez JL, 2019. Defining and combating antibiotic resistance from One Health and Global Health perspectives. In Nature Microbiology (Vol. 4, Issue 9, pp. 1432–1442). Nature Publishing Group. <https://doi.org/10.1038/s41564-019-0503-9>
- Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D, Bover-Cid S, Chemaly M, Davies R, De Cesare A, Hilbert F, Lindqvist R, Nauta M, Peixe L, Ru G, Simmons M, Skandamis P, Suffredini E, Jenkins C, Malorny B, Ribeiro Duarte AS, Torpdahl M, da Silva Felício MT, Guerra B, Rossi M and Herman L, 2019. Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. EFSA Journal, 17. <https://doi.org/10.2903/j.efsa.2019.5898>
- Neuert S, Nair S, Day MR, Doumith M, Ashton PM, Mellor KC, Jenkins C, Hopkins KL, Woodford N, de Pinna E, Godbole G and Dallman TJ, 2018. Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal *Salmonella enterica*. Frontiers in Microbiology, 9. <https://doi.org/10.3389/fmicb.2018.00592>
- Smith AM, Tau NP, Smouse SL, Allam M, Ismail A, Ramalwa NR, Disenyeng B, Ngomane M and Thomas J, 2019. Outbreak of *Listeria monocytogenes* in South Africa, 2017–2018: laboratory activities and experiences associated with whole-genome sequencing analysis of isolates. Foodborne Pathogens and Disease, 16, 524–530. <https://doi.org/10.1089/fpd.2018.2586>

## Abbreviations

AESAN	Agencia Española de Seguridad Alimentaria y Nutrición
AMR	antimicrobial resistance
ARG	antimicrobial resistance gene
CC	clonal complex
ESBL	extended-spectrum beta-lactamases
MGE	mobile genetic element
MLST	multilocus sequence typing
MLVA	multiple locus variable number tandem repeat Analysis
MSCA	Marie Skłodowska-Curie Actions
PFGE	pulsed-field gel electrophoresis
RAPD	random amplified polymorphic DNA analysis
ST	sequence type

VNTR variable number tandem repeat analysis  
WGS whole genome sequencing  
WMS whole metagenome sequencing



## Appendix A – Listeriosis outbreak simulation

### Case: Listeriosis outbreak in León (Simulation)

A recent increase of Listeriosis cases has been reported in the Hospital of León, with 10 patients presenting associated symptoms during the last week. *Listeria monocytogenes* isolates were obtained from all of them, and whole genome sequencing data were obtained. After a questionnaire filled by patients, the main suspects of origin of infection were identified and are summarised in Table A.1. The genomic analysis showed that all the isolates belonged to ST9, excluding those from patient01 (ST121), patient03 (ST155) and patient04 (ST14). The virulence and antimicrobial resistance profiles were very similar between the ST9 samples 02, 05, 06, 07, 08, 09 and 10. Average Nucleotide Identity (ANI) analysis showed that, considering the accessory genes, the samples were distributed in two main clusters: Cluster 1, with similarity of 100%, formed by the samples 02, 06, 08 and 09 and Cluster 2, with similarity of 99.9%, including isolates 05, 07 and 09 (Figure A.1). According to the metadata, the common sources identified in Cluster 1 were 'sausage table (embutidos)' and 'Commercial burger', while in Cluster 2 'Homemade burger' was the common source, while 'Commercial burger' was a possible source in 2/3 of the isolates. Regarding the other two samples, patient03 and patient04, the two possible sources were still 'Commercial burger' and 'Homemade burger', however, different strains of *L. monocytogenes* might have contaminated the suspect foods. Additional samples were obtained and are listed in Table A.2. The MLST assignation gave these results: food02 and producer03 belonged to unknown ST; food12 and patient03 belonged to ST155; all the other isolates were ST9, excluding food04 (ST37), patient01 (ST121) and patient04 (ST14). The results of the ANI analysis are shown in Figure A.1. In particular:

- the clinical isolates forming Cluster 1 clustered with food03 and producer02;
- the clinical isolates forming Cluster 2 clustered with two isolates of the Commercial burger;
- patient03 clustered with another strain found on lettuce in the Commercial burger;
- producer03 provided contaminated food (lomo) to the restaurant;
- patient01 and patient04 were infected by isolates not related to any other in this study;

In conclusion, four patients were traced back to producer02 and three patients to the Commercial burger.

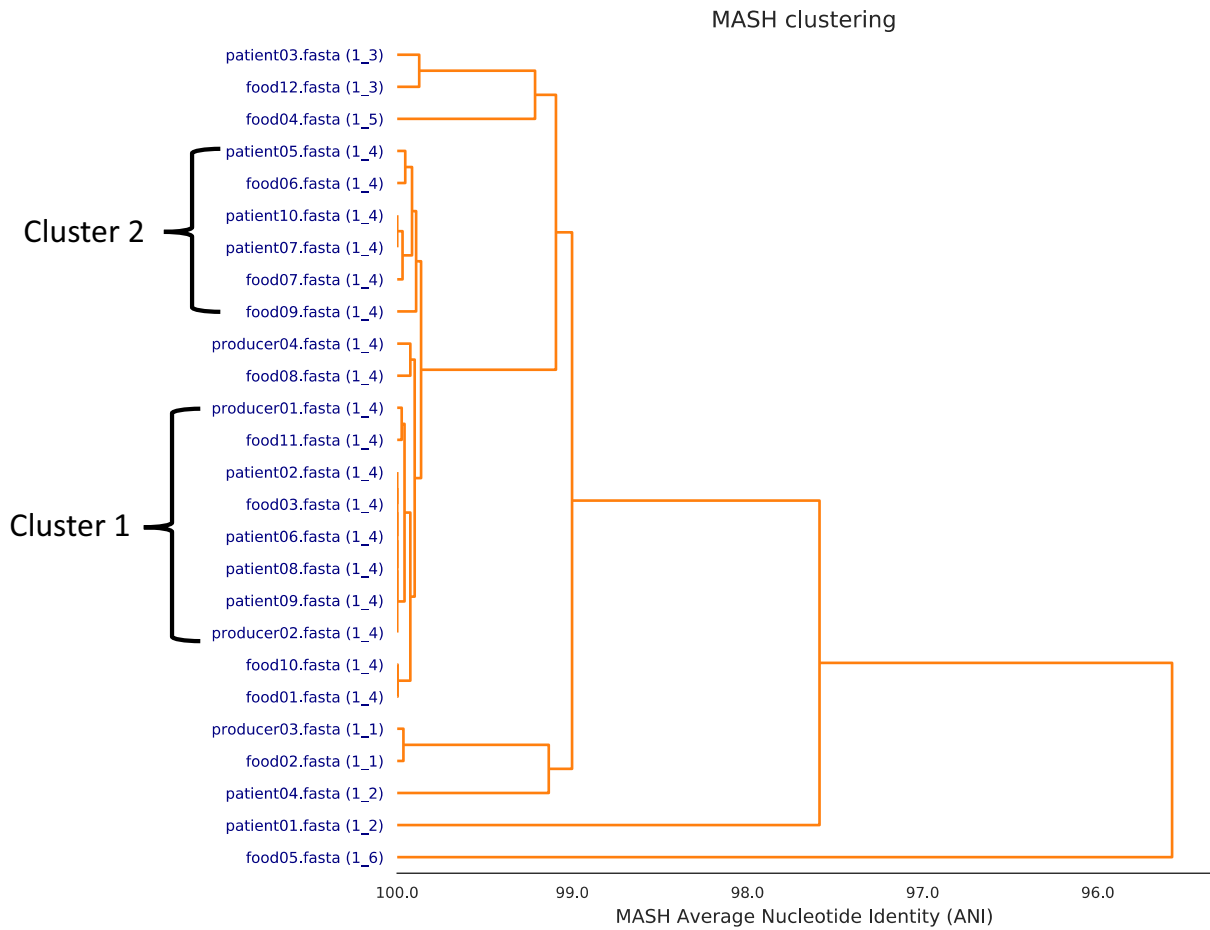
**Table A.1:** Suspected origins of infection

Code	Suspect 1	Suspect 2	Suspect 3
patient01	Commercial burger	Raw milk	Fresh cheese
patient02	Sausage table (embutidos)	Smoked salmon	Commercial burger
patient03	Commercial burger	Fresh cheese	Smoked salmon
patient04	Raw milk	Homemade burger	Smoked salmon
patient05	Homemade burger	Commercial burger	Melon
patient06	Commercial burger	Melon	Sausage table (embutidos)
patient07	Homemade burger	Fresh cheese	Melon
patient08	Sausage table (embutidos)	Smoked salmon	Commercial burger
patient09	Commercial burger	Sausage table (embutidos)	Raw milk
patient10	Homemade burger	Commercial burger	Fresh cheese

**Table A.2:** Additional samples from foods and producers

N	Sample	Product	Origin	N	Sample
01	food01	Fresh cheese	Commercial burger	13	producer01
02	food02	Lomo	restaurant	14	producer02
03	food03	Chorizo	restaurant	15	producer03
04	food04	Salchichón	restaurant	16	producer04
05	food05	Cecina	restaurant		
06	food06	Cooked ham	Commercial burger		
07	food07	Lettuce	Commercial burger		

N	Sample	Product	Origin	N	Sample
08	food08	Cooked ham	Commercial burger		
09	food09	Chorizo	restaurant		
10	food10	Fresh cheese	Commercial burger		
11	food11	Salchichón	restaurant		
12	food12	Lettuce	Commercial burger		



**Figure A.1:** ANI tree of all the isolates, performed by *dRep* software