

**JRP24-FBZSH9-BEONE D2.1**

**Workpackage 2**

**Responsible Partner: RIVM**

**Contributing partners: RIVM (30), INSA (36), SSI (13), PHE (22), NVI (33),RKI (11), APHA (21), PIWET (34)**

**GENERAL INFORMATION**

|  |  |
| --- | --- |
| **European Joint Programme full title** | **Promoting One Health in Europe through joint actions on foodborne zoonoses, antimicrobial resistance and emerging microbiological hazards** |
| **European Joint Programme acronym** | **One Health EJP – BeONE - Building integrative tools for OneHealth Surveillance** |
| **Funding** | **This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement No 773830.** |
| **Grant Agreement** | **Grant agreement n° 773830** |
| **Start Date** | **01/01/2018** |
| **Duration** | **24 Months** |

**DOCUMENT MANAGEMENT**

|  |  |
| --- | --- |
| **Project deliverable** | **D-BeONE.2.1 Draft manuscript on a conceptual model for factors impacting genomic clustering** |
| **Project Acronym** | **BeONE** |
| **Author** | **Claudia Coipan** |
| **Other contributors** | **Claudia Coipan, Bergval I., Friesema I., van den Beld M., Hernandez Segura A., van Hoek A., Pijnacker R., Franz E.** |
| **Due month of the report** | **M36** |
| **Actual submission month** | **M37** |
| **Type**  ***R: Document, report***  ***DEC: Websites, patent filings, videos, etc.; OTHER*** | R  **Save date**: 28-jan-21 |
| **Dissemination level**  ***PU: Public (default)***  ***CO: confidential, only for members of the consortium (including the Commission Services)*** | **PU**  **This is the default setting. If this project deliverable should be confidential, please add justification here (may be assessed by PMT):** |
| **Dissemination**  ***Author’s suggestion to inform the following possible interested parties.*** | OHEJP WP 1  OHEJP WP 2  OHEJP WP 3  OHEJP WP 4  OHEJP WP 5  OHEJP WP 6  OHEJP WP 7  Project Management Team  Communication Team  Scientific Steering Board  National Stakeholders/Program Owners Committee  EFSA  ECDC  EEA  EMA  FAO  WHO  OIE  Other international stakeholder(s): ………………………………………………………………  Social Media:  **Other recipient(s):** |

Conceptual Model of factors impacting GENOMIC clustering

# Summary

Cornerstone for effective public health control of foodborne infectious diseases is surveillance of foodborne pathogens in which clusters and outbreaks can be detected in an early stage. Surveillance results in time series of foodborne disease cases. These time series can be used to draw a baseline for expected incidence of the infections with various foodborne pathogens, which is specific for a certain location. Traditionally, algorithms that detect anomalies in the baseline (a comprehensive review thereof can be found in (1)), such as the Farrington algorithm (2), or variations thereof (3, 4), have been used to detect outbreaks, and thus, the initiation of an outbreak investigation has been reliant on cases clustering in time and space.

Whole-genome-sequencing (WGS) is being increasingly used as a routine typing tool for foodborne pathogens and it has revolutionized the field of molecular epidemiology. The benefits of WGS include a much greater strain discrimination for bacterial typing, coupled with the possibility to infer simultaneously also phenotypic information (such as serovar, serotype, and antibiotic resistance patterns). This allows an enhanced cluster resolution and an equally enhanced ability to detect clusters and to monitor disease trends.

One possible approach of the process of WGS-assisted cluster definition involves three steps: (1) the creation of a genetic distance matrix among the isolates, (2) definition of clusters based on data structure (e.g. internal validity indices), or previously set distance thresholds, and (3) the refinement of the clusters based on additional, biological or epidemiological information. The interpretation of the clusters thus defined will form the basis for the initiation of an outbreak investigation.

While often, the relatedness of the isolates within a cluster is obvious, and falls within the intervals already defined in the literature, there are occasions when a cluster might include isolates of uncertain membership. Specific biological and ecological features (space/location, time, resilience of the pathogen to the environmental factors) can be used in the reassessment of the cut-off values used for each foodborne pathogen (6, 7, 26). In practice, all these factors are part of the sequential decision making process used in outbreak detection, only that often they have become implicit and subject to personal experience and are not formalized and structured into a clear algorithm.

In order to provide a basis for a better informed integration of WGS in epidemiological surveillance, we attempt to formalize the cluster detection problem and outline a conceptual model for the biological and epidemiological factors impacting on cluster detection. We propose a conceptual model of the common practice steps performed towards food-borne pathogens’ cluster detection. We underline the complex relations between the biological and ecological factors at play in the evolution of some of the most common food-borne pathogens, the standards chosen and the ways how different conceptual sources can be used for improving integrative algorithms of cluster detection.

The conceptual model is a preliminary representation of the reality, which can be tested and further developed upon emergence of new data on the topic. It is also a first step in developing and evaluating new algorithms of cluster definition/detection.

We have reviewed the literature for factors impacting on the genetic dissimilarity of the isolates within an outbreak, using an information foraging approach (43). We used the approach of cognitive mapping (44) to describe the relations between the various factors, relying on literature review and brainstorming sessions with experts in microbial diagnostics, medical microbiology, and epidemiology. We filtered out some of the factors/elements that are not relevant when building the system dynamics. The remaining factors we have used to draw an influence diagram (45), using the package DiagrammeR (46) in R (47).

# References

1. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2012;175(1):49-82.

2. Farrington C, Andrews NJ, Beale A, Catchpole M. A statistical algorithm for the early detection of outbreaks of infectious disease. Journal of the Royal Statistical Society: Series A (Statistics in Society). 1996;159(3):547-63.

3. Heisterkamp SH, Dekkers AL, Heijne JC. Automated detection of infectious disease outbreaks: hierarchical time series models. Statistics in Medicine. 2006;25(24):4179-96.

4. Manitz J, Höhle M. Bayesian outbreak detection algorithm for monitoring reported cases of campylobacteriosis in Germany. Biometrical Journal. 2013;55(4):509-26.

6. Besser JM, Carleton HA, Trees E, Stroika SG, Hise K, Wise M, et al. Interpretation of Whole-Genome Sequencing for Enteric Disease Surveillance and Outbreak Investigation. Foodborne pathogens and disease. 2019;16(7):504-12.

7. Pightling AW, Pettengill JB, Luo Y, Baugher JD, Rand H, Strain E. Interpreting Whole-Genome Sequence Analyses of Foodborne Bacteria for Regulatory Applications and Outbreak Investigations. Front Microbiol. 2018;9:1482.

26. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. Clinical Microbiology and Infection. 2018;24(4):350-4.

43. Pirolli P. Information Foraging. In: Liu L, ÖZsu MT, editors. Encyclopedia of Database Systems. Boston, MA: Springer US; 2009. p. 1485-90.

44. Ruiz-Primo MA, Shavelson RJ. Problems and issues in the use of concept maps in science assessment. Journal of Research in Science Teaching. 1996;33(6):569-600.

45. Dawid AP. Influence Diagrams for Causal Modelling and Inference. International Statistical Review / Revue Internationale de Statistique. 2002;70(2):161-89.

46. Iannone R. DiagrammeR: Graph/Network Visualization. R package version 1.0.6.1. ed2020.

47. RCoreTeam. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.