Deliverable JIP1-2.8 - Status report on OH Knowledge Base - NGS

Responsible Partner: 33-NVI

This meeting is part of the European Joint Programme One Health EJP.
This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.

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**
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	**60 Months**

DOCUMENT MANAGEMENT

Deliverable	JIP1-2.8 - OH Knowledge Base - NGS
WP and Task	JIP1 – WP2-NGS - T-2.3
Leader	Karin Lagesen (NVI)
Other contributors	**Karin Lagesen (NVI), Umaer Naseer (NIPH), Emily Macdonald (NIPH), Lin Brandal (NIPH), Olaug Taran Skjerdal (NVI), Thomas H.A. Haverkamp (NVI), Eve Zeyl Fiskebeck (NVI), Jeevan Karloss Antony Samy (NVI, NIPH), Georgios Marselis (NVI), Camilla Sekse (NVI), Håkon Kaspersen (NVI), Wonhee Cha (SVA), Cecilia Jernberg (FOHM), Olov Svartstrom (FOHM), Mia Torpdahl (SSI), Eva Møller Nielsen (SSI), Lesley Larkin (PHE), Charlotte Cook (APHA), Susan Withenshaw (APHA), Johanna Lawes (APHA).**
Due month of the deliverable	**M42**
Actual submission month	**M42**
Type ***R: Document, report*** ***DEC: Websites, patent filings, videos, etc.*** ***OTHER***	R
Dissemination level ***PU: Public*** ***CO: confidential, only for members of the consortium (including the Commission Services)***	PU

This meeting is part of the European Joint Programme One Health EJP.
This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.

WP2-NGS: SEQUENCING FOR SURVEILLANCE - KNOWLEDGE BUILDING, PRACTICAL CONSIDERATIONS, AND IMPLEMENTATION

Background

Second generation sequencing emerged in the mid-2000s as a powerful new technique for gathering substantial amounts of genomic information. Several new instruments came on the market, and this led to a drastic drop in sequencing cost as compared to Sanger or first generation sequencing instruments. This rapid drop in sequencing costs led to a substantial increase in the number of available genomes in public databases, and the emergence of comparative genomics as a separate subdiscipline.

With this development, sequencing also started to be used for public and veterinary health purposes. One of the first examples of this was the use of sequencing to solve the 2011 German *E. coli* sprouts outbreak(*1*), and the 2010 cholera outbreak in Haiti(*2*). Since then, this practice has become more widespread, especially in the public health sector. In 2016, EFSA conducted a survey among its member states regarding use of whole genome sequencing (WGS) data for food-and waterborne diseases(*3*). This report showed that at that time half of the responding labs had started implementing WGS as an analysis method. Sequencing has also found its use for surveillance, and a good example is here the use of whole genome sequencing of Salmonella by PHE in the UK(*4*).

However, setting up sequencing for public and veterinary health purposes is not an easy task to accomplish. It requires new instruments and compute infrastructure, knowledge, personnel, and also funding for both equipment and to set up new processes. Knowledge on how to set this up was at the start of the project present at several institutions. However, this knowledge was to a large extent not documented and thus not shareable with other actors, which could slow the implementation speed for these methods.

The aim for WP2 in this project was to create a One Health Knowledge Base (OHKB) - a cross-domain inventory of currently available resources for performing integrated One Health (OH) surveillance generation, data analysis and interpretation. WP2-NGS was part of WP2, and the goal was to mitigate the lack of documented practices by creating a resource which institutions could use to build their own capacity within this area. This WP focused on sequencing based analytical data, methods, analyses and systems. The inventory generated from this work package aimed to include current practices for data management and handling approaches for bacterial identification techniques. It was also to include current experiences on practical issues such as data storage (both short and long term) and exchange capacity, data exchange platforms, harmonized data, harmonized terminology, as well as the need for bioinformatics expertise.

Activities

Several activities were performed in this work package. The knowledge gathered through the various activities were all incorporated into the One Health Knowledge Base for this work package, which became the ORION WP2-NGS Sequencing for Surveillance handbook.

Knowledge mapping

A primary activity of this WP was knowledge mapping. This was performed through literature reviews, through a questionnaire, and through interviews with several domain experts.

Literature surveys were performed through review of current papers and reports that deal with surveillance, outbreak detection and investigation using sequencing data, including EFSA and ECDC reports. The focus for the literature survey was to get a good understanding of currently used analysis methods, with a special emphasis on potential challenges and advantages when employed for surveillance and outbreak situations. We in addition focused on tools and methods for storing, managing, analyzing and interpreting data, which can be a significant challenge when dealing with sequencing information.

Based on initial information gathered through the literature review, we decided to collect more information on current practices through a questionnaire that was sent to all EJP partners. The goal was to elucidate how veterinary and public health institutions dealt with sequencing data for surveillance purposes. The analysis of the results were complicated by a low response rate, however, the results did help identify areas of practice where current procedures primarily were based on informally and undocumented locally acquired experience.

To augment our knowledge base, several site visits and online meetings with people at several veterinary and public health labs in Europe were conducted. The labs that were visited in person were Public Health England (PHE), Statens Veterinærmedisinska Anstalt (SVA) and Folkhalsomyndigheten (FOHM), both in Sweden, and also Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise (ISZAM, Teramo, Italy). Additionally, we had an online meeting with Statens Serumsinstitut (SSI) in Denmark.

Norwegian pilot

The Norwegian pilot was a collaboration between the NIPH and the NVI. The main focus of the pilot was to set up shared facilities for analysis of sequencing data, with an emphasis on creating pipelines suitable for analysis of *Listeria monocytogenes*.

Surveillance of *Listeria monocytogenes*

The focus pathogen for the Norwegian pilot was *Listeria monocytogenes*. This work proceeded via two different pathways.

The first path aimed to establish a viable set of sequence analysis tools that could be used to analyze this pathogen, with a view towards creating a pipeline that could be shared with others. To get an overview of what was in current use, we adopted a systematic review approach

inspired by the PRISMA methodology. The aim was to build an R-markdown notebook that could through a semi-automatic method regularly show what the current analytical trends were. This work was not fully completed due to personnel issues releated to Covid-19 duties and restrictions, but will be continued internally at the NVI. However, the work did help us establish what results are commonly produced for this pathogen, and a list of tools was identified that could be used for analysis. This tool list was used to create a draft pipeline which was used for participation in the ANSES *Listeria monocytogenes* proficiency test that was run during the fall of 2020. We also initiated a collaboration with EFSA regarding development of analysis pipelines, and harmonization of the results from the pipelines. This has resulted in the development of pipelines that once completed will be compatible with the outputs that EFSA prefers.

The second pathway aimed to examine how sequence analysis would work in a surveillance and outbreak context. To accomplish this, we conducted several meetings and discussion sessions with the people responsible for surveillance of this pathogen in Norway. Fortunately, the National Reference Laboratory person for *Listeria monocytogenes* was part of this WP, which helped facilitate this process. These sessions included discussions concerning how surveillance works for this pathogen, what expectations this places on the sequence analysis workflow and results, and on how the results can be used for surveillance and outbreak investigations. These discussions helped tune the analysis pipeline and the presentation of the outputs from the pipeline. They also clarified that there was a need for a background sequence database that could work as a baseline for comparison. First, several of the methods that became part of the pipeline need to include other isolate sequences for comparison when producing results. In addition, the composition of the database also serves as a guide the interpretation of the output from the analyses. Thus, the composition of this database was also a subject for the discussions on *Listeria monocytogenes* surveillance.

Shared analysis platform

A primary goal for the Norwegian pilot was to set up an analysis platform that could be used by both the NIPH and the NVI. As part of the knowledge mapping done for this project, we searched for data management and analysis platforms (DMAPs) which could be used for analysing microbial sequencing data in a surveillance and outbreak context. We used documentation and active development of code as indicators to find platforms which were functional and up-to-date in their handling of microbial sequence data. The EFSA and ECDC in 2019 published an evaluation covering 11 different tools that could potentially be used for such purposes(*5*). Based on our previous knowledge mapping work, and supported by the EFSA/ECDC report, we identified two platforms that seemed to fit our requirements, these were the IRIDA and the INNUENDO systems.

We subsequently proceeded to evaluate these systems by examining documentation, and by conducting discussions with the developers. The developers of the platforms provided us with guest access of their systems which allowed us to test out the features of the systems. We also had meetings with users of the DMAPs who were not from the same research institute or from the team which developed the DMAP. This included people at the Finnish Institute for health and welfare, the University hospital Lausanne, Switzerland and the Quadram Institute in UK.

We also tested out the systems by installing them ourselves to see how they worked "under the hood" and to understand more about their strengths and weaknesses.

Danish pilot

A Danish pilot study was set up to promote collaboration among all agencies involved in surveillance of *Campylobacter*. An OH working group was set up for this project. The group included stakeholders from all sectors and meet with regular intervals to discuss all aspects of the project. This includes the results from a retrospective WGS study which was set up to facilitate a real-time study for 2019 and further implement a routine OH surveillance system from 2020 onwards, based on the evaluation of the first year (2019) of real-time surveillance. Data on *Campylobacter* was analysed using cgMLST derived from WGS and used to detect genetic clusters among patients. Furthermore, an OH approach was used to detect possible links between food/animal sources and human clusters.

Results

Building a knowledge base - the Sequencing for Surveillance handbook

The knowledge mapping phase revealed that while a consensus regarding tools had started to emerge, there was no consensus regarding how to deal with the practical aspects surrounding data storage, management, analysis and interpretation. This was made especially clear through the lab interviews, where it was apparent that each lab had had to develop their own practices with little input from other labs. We thus decided to create a resource that could help fill this gap by documenting commonly used tools, methods and strategies for these issues. This resource became the One Health Sequencing for Surveillance handbook. We decided early on to set this up as a community editable document using Github and Readthedocs, with the aim of facilitating sustainability after the end of the ORION project.

The mapping work done in the beginning of the project suggested that the audience for such a resource would likely be a veterinary or public health laboratory that had good molecular capabilities but that had yet to establish procedures for sequencing. This work also revealed that it would be beneficial if the handbook covered the following four major areas: infrastructure needs, tools and pipelines, species specific resources such as databases and typing schemas, and the use of sequencing for surveillance. Thus we proceeded to create the handbook along these lines. However, through further work we discovered that a different division of the handbook would make it more useful for the intended audience. The current subheadings for the handbook are now:

- Sequencing technologies: descriptions and discussions regarding sequencing machines, their output formats and error profiles.
- Bioinformatics methods: short descriptions of common algorithms used by tools that are used for microbial genomics analyses. The main aim is to help create a basic understanding for how a tool works.

- Single isolate analyses: descriptions of tools and methods used to characterize a single isolate. This includes MLST and virulence and antimicrobial resistance finding.
- Cluster analysis: descriptions of methods and tools used for clustering isolates, including cg/wgMLST and phylogeny.
- Compute infrastructure: discussions on data storage and management, pipeline and workflow management tools, and platforms for managing analysis, data and visualization.
- Species analysis: discussion of pathogen-specific methods that are currently employed in NGS surveillance, focusing on a specific set of methods that are known to perform well for the species under interest, including species specific resources such as databases and schemas.

The species-specific part of the handbook was developed in collaboration with the BeONE project. The handbook will continue to be developed within the BeONE project, and is also open to community input via github.

The handbook can be found here: https://oh-sfs-handbook.readthedocs.io/en/latest/

Norwegian pilot

Listeria monocytogenes in Norway

There are fortunately few cases of listeriosis in humans in Norway each year. During the last 10 years, the occurrence has varied between 20-40 cases per year. Most cases are acquired within Norway through consumption of contaminated food or water(*6*) . In recent years several cases have proven to have a link to fish products. This disease is notifiable when found in humans and animals. It is not uncommon to find this pathogen in food products such as meat, dairy and fish, but it is not notifiable unless it is found above a certain concentration as defined in food law.

Surveillance and outbreak management of *Listeria monocytogenes* happens in collaboration between the NIPH and the NVI, where the NIPH manages all human cases of listeriosis and the NVI the animal and food associated cases. As mentioned, there are relatively few human cases of this pathogen, which means that the NIPH acquires relatively few new isolates of *Listeria monocytogenes* every year. The veterinary side gets considerably more isolates each year. This is mainly due to food companies sending isolates from their internal control systems to the NVI for species confirmation. In addition the NVI allows for storing isolates so that the companies via the NVI can compare new isolates to older isolates at a later date to help the company control the presence of this pathogen. However, the number of isolates from official surveillance programs of food on the market and from outbreak investigations is in the same order as human isolates.

The NIPH has during the last years performed sequencing on all isolates that have been acquired from human cases. In cases where there is an associated suspect food product, the NVI has analyzed the food to find the pathogen, and that isolate has then been shared with the NIPH. With new sequencing capabilities at the NVI, we are currently sharing sequencing data,

provided the isolates in question have a MLST match to that of the isolate acquired from the human case. We currently have access to a shared analysis platform, and going forward we will figure out how this can be employed for One Health surveillance and outbreak investigation.

Bioinformatics requirement analysis for *Listeria monocytogenes*

The primary goal for this project was to develop a bioinformatics analysis pipeline that could be used for surveillance purposes. Through the process of mapping the requirements that such a pipeline would have to satisfy, it became clear that communication between bioinformaticians, microbiologists and epidemiologists was a challenge. People from the separate disciplines had their own knowledge bases, which meant that each group operated under disparate assumptions. Thus, considerable time was spent in discussions and informal training sessions to ensure that a shared understanding could be developed. The bioinformatics team focused on how sequencing and bioinformatics tools worked, what results such tools could produce, and what caveats the tools come with. The epidemiologists and microbiologists focused on explaining how this particular pathogen works, and how surveillance and outbreak investigations for *Listeria monocytogenes* works in Norway. These discussions helped establish how bioinformatics results could be employed for One Health monitoring of this pathogen. These discussions led us to understand that OH monitoring of this pathogen consisted of three different processes or scenarios: surveillance, outbreak detection and outbreak investigation. The main focus for this project was on surveillance. However, for One Health purposes we also explored the outbreak scenarios. Through discussions we established that while many different bioinformatics analyses could be useful under each scenario, there were certain results that would be more useful for a specific scenario. For surveillance purposes methods that characterize the isolates, e.g. MLST and cgMLST typing, as well as serotyping and virulence finding would be very useful. These results can be produced per isolate, and the presence of characteristics such as sequence types, serotypes and genes conveying specific characteristics such as virulence can be tracked. Such results could potentially also assist with outbreak detection, provided characterization is done on a continuous basis to enable detection of trend changes. For outbreak investigation however, clustering methods either based on cgMLST or phylogeny would be more appropriate. The aim in that situation is to infer whether two isolates might stem from the same source, which implies that the genome sequences of the two isolates should be very similar. Thus, methods should be employed that have very high resolution, i.e. that include large portions of the genome and that thus can detect minute changes. Furthermore, in this context it became clear that the usefulness of bioinformatics methods would be limited by the other isolates used for comparison with a new isolate. We thus proceeded with expanding the existing sequencing database by focusing on isolates that could help elucidate the source of an outbreak. Considering the recent increase in fish product related human listeriosis cases, we had a special emphasis on increasing the availability of isolates stemming from fish products.

Development of a *Listeria monocytogenes* pipeline

Based on discussions between the NIPH and the NVI we proceeded with constructing bioinformatics pipelines that could be used for surveillance. Our focus was on MLST and cgMLST typing, as well as on finding the clonal complex group and the serotype. During the fall of 2020, the European Union Reference Laboratory for *Listeria monocytogenes* conducted a sequencing based proficiency test in which we we participated. For this test we were provided

with two sets of starting materials, first a set of sequencing data from already sequenced isolates, and second, a set of isolates that we were to sequence and subsequently analyze. We thus proceeded with building a pipeline that could be used for analysis. According to the results of the proficiency test we got a maximum score on all counts. Our ability to detect contamination detection, identify molecular serotype and clonal complex, perform cgMLST typing and produce a distance matrix (hamming distances) and build a neighbor-joining tree were all deemed satisfactory.

This first pipeline was not integrated in the data management and analysis system that we during the same period were setting up. Once our setup of that system had reached the stage where we could create our own pipelines, we proceeded to set up an integrated pipeline that would suit our needs. Through the ORION project we discovered that the EFSA is creating a system for analyzing sequencing data for its member states. Through discussions with EFSA, we decided to adapt the integrated pipeline in such a way that it would be compatible with the results that EFSA would like for their sequencing based surveillance programs. A further description of that integrated pipeline can be found below.

Setting up a shared data management and analysis resource

A primary goal for this pilot was to set up a shared system between the NVI and the NIPH for analyzing *Listeria monocytogenes*. A main requirement due to internal constraints at both the NVI and the NIPH was that the system should be locally installable, so that it would be possible to keep control of the data and know who had access to it. Additional criteria were that it should have a graphical user interface that would enable novice users to use it, and be able to manage both sequencing data and sequence metadata. In addition, users should find it easy to start, monitor and evaluate the results from an analysis, and it should come with a pipeline creation and monitoring tool. Last but not least, it should be possible for people outside of the main development team to expand the system if needed. Based on our criteria, two systems appeared as viable options. These were the INNUENDO and the IRIDA systems, which seemed to be functional, active, up-to-date platforms which could manage and analyze high throughput microbial genomics data combined with visualization options that aided interpretation. Since commercial platforms such as Bionumerics do not allow users to extend the system, commercial systems were not considered.

The INNUENDO system was developed in a multinational project largely funded by EFSA. The aim was to create a One Health sequencing data management and analysis system that could be used for One Health purposes. The project developed a web platform that managed user access, access to data, data analysis and visualization of results, and used Nextflow for managing analysis pipelines. However, due to lack of further funding the development of the platform was put on hold in 2019. While the system did appear to be useful for our purposes, the lack of further development and thus reduced user support was considered such a serious drawback that this system was dropped from further consideration.

We thus decided to proceed further with the IRIDA system. This system is developed in collaboration between several universities and public health institutes in Canada, and is also in current use in several other locations, such as the Quadram Institute in the UK and the

University of the Western Cape, South Africa. It also forms the basis for the ARIES system which is run by the European Reference lab in Italy. The system consists of a web platform that manages user access, as well as access to sequence data and metadata, analysis pipelines and analysis results. The system is set up in such a way that all data belong to at least one project, and it is possible to control which users have access to what project. The web platform uses Galaxy as its workflow engine, and thus uses Galaxy for managing pipelines and sequencing tools.

Once we had settled on a system, we proceeded with testing it out and examining the compute resources we would need to run this system. The expected computational load led us to seek an outside infrastructure provider. The main criteria for choice of provider was that they would be a Norwegian entity and that they would have all of their infrastructure on site. In addition, we preferably wanted a provider that had previous experience with running similar software. Based on these requirements, we decided to contact the Norwegian Research and Education Cloud (NREC), which is operated in collaboration between the University of Oslo and the University of Bergen, the two largest universities in Norway. They provide "infrastructure as a service" through OpenStack for research institutions in Norway. We gained access to resources at NREC, and successfully installed IRIDA in a virtual machine with basic computational power. IRIDA's installation documentation and virtual support through gitter and skype calls helped us in understanding and solving the issues we faced. Based on our hands-on installation experience, we saw that at least a moderate level of Linux system administration, knowledge about webservers, Galaxy platform administration and web development would be required to operate and maintain the system. Once we had gained a reasonably good understanding of the system, we proceeded with setting it up in a more robust manner, as described below, and the NVI is in the process of signing a service agreement with NREC for further use.

Current architecture

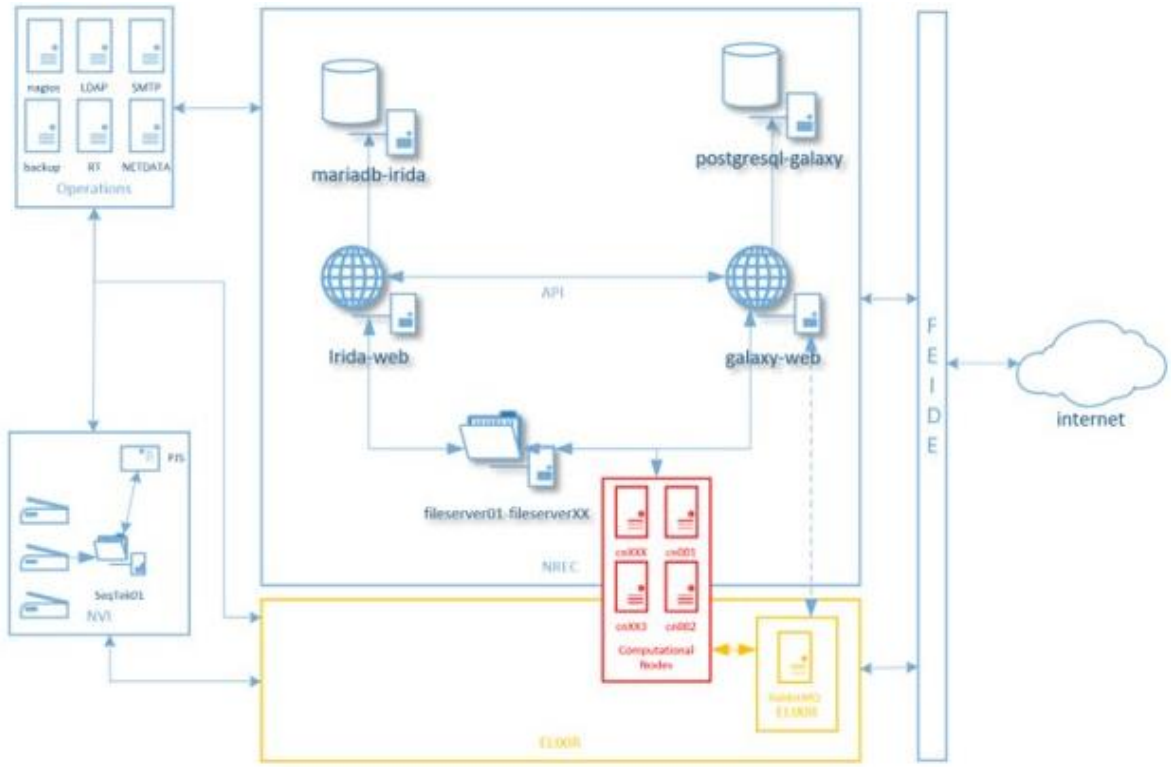
The IRIDA system consists of a front-end that users interacts with, and a backend that runs the webservers, the galaxy pipelines and the databases that are needed for the system. Our current backend consists of these components:

- Two web servers, one for IRIDA and one for Galaxy
- A postgresql database server for Galaxy, which stores information about the state of the Galaxy backend, along with galaxy users, files, paths, analysis runs and results together with other metadata, such as day and time executed. This database server also hosts the pipeline or workflow information.
- A mariadb database for IRIDA, which stores information about users, projects and metadata, such as the day and time when analyses were executed. It also holds information about IRIDA specific pipelines and tools, outside of galaxy
- One fileserver (shared among all servers).
- Computational nodes (shared among all servers).

While it is not strictly necessary to split the system up in individual virtual machines for each function, having each function in its own server makes for a more robust system in that the entire system will not go down if one function fails. This structure also allows for easier

debugging when something goes wrong since it will only be necessary to debug one function and not the entire system. This means that we will be able to maintain the simplicity of the systems and be able to quickly get back to a running state should disaster strike. To further facilitate system robustness, repeatability and ease deployment, we are in the process of creating Ansible playbooks for each server. That way it will be possible for us to restart a component quickly and with minimum amount of downtime, if needed.

Current system architecture



Pipeline development

IRIDA in its native state comes with six pipelines for assembly, annotation, quality control and typing. The system in addition allows for installation of new pipelines in the form of plugins. Plugins are modules, i.e. files containing code that can be dropped into specific directories within the IRIDA file structure, and that on a subsequent restart of a system will result in a new pipeline being available in the system. A new plugin consists of a galaxy workflow file (.ga), a plugin description file (.properties), a tool interface file (.xml), a plugin java file and Java Archive file (.jar). To be able to create a new plugin, knowledge in bioinformatics, how the galaxy platform works and proficiency in java programming is needed.

The IRIDA system allows for setting up a pipeline to be run automatically upon data upload. We decided to exploit this feature by setting up species specific pipelines that would produce the most commonly needed results for that pathogen. We thus proceeded to develop a plugin specifically for analysis of *Listeria monocytogenes*. The tool set we decided to go for was "FastQC" for read quality checking, "mash screen" for species identification, "shovill" for assembly, "checkM" for genome integrity evaluation, and "MLST" and "ChewBBACA" for

typing. The report that the analysis plugin produces contains information on coverage, N50 and NG50, number of contigs, genome completeness, species identification information, MLST type, cgMLST information (schema, number of matching and missing loci), and a list of tools and databases, including versions that have been used.

The code for the plugin is available here:
https://github.com/NorwegianVeterinaryInstitute/irida-plugin-listeria-typing

Further work on One Health surveillance in Norway

The Norwegian pilot project has had significant impacts on how bioinformatics is used to for monitoring of *Listeria monocytogenes* in Norway. The two institutes that collaborate in this process have through this project significantly increased their collaboration and coordination activities. The NIPH had pre-existing capabilities within this field, and through the ORION project and other internal projects, the NVI have gained similar capabilities. Collaboration with the NIPH and with other institutes involved in this project helped guide this process. The work done in this project helped elucidate the types of analyses that are needed for monitoring of pathogens, and this work was instrumental in designing a pipeline that can be used for surveillance for *Listeria monocytogenes*. This pipeline is now available for others that want to analyze this pathogen. Through this project the two institutes were able to set up and create a shared infrastructure that can be used for collaborative analysis of *Listeria monocytogenes* and also other One Health pathogens. Work is in progress to establish routines and procedures for how this platform will be used in the future.

The effects of the Danish pilot

The Danish *Campylobacter* pilot project has had a significant impact on the OH-approach for surveillance of *Campylobacter*. The Danish surveillance system has undergone real-time changes during the lifetime of this project, as new insight were gained as the project developed. The project showed that it is possible to detect clusters, outbreaks and link to sources based on comparison of sequences in a OH setting. During the project, the source of several outbreaks were found and control measurements implemented, not only at slaughterhouse level but in the broiler production. The study resulted in national and international recognition of sequencing based *Campylobacter* surveillance being of high value, also compared to "older" methods used. The project will be continued with the development of a platform for real-time sharing of data and surveillance. An OH working-group was set up in connection with the project and the working group will continue after the project has ended. Results from this study has been presented nationally as well as at internationally and data used for annual reporting as well as published in a peer-reviewed journal for international dissemination. A seminar on control of Campylobacter in Broilers was set-up to improve knowledge sharing with industry, universities and agencies. Funding has also been granted for continuous use of WGS to examine the diversity and transmission within/between chicken flocks, farms, etc.

Presentations and publications

Oral presentation, Karin Lagesen, "ORION - One health suRveillance Initiative on harmOnization of Data Collection and interpretation" ASM Conference on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatic Pipelines conference, Washington, US, 2018

Poster, Thomas Haverkamp, Karin Lagesen, Jeevan Karloss Antony-Samy, Mia Torpdahl, Wonhee Cha, Lesley Larkin, Sandra Stelzer, Mohammed Umaer Naseer, "Managing and Analyzing NGS Data for Surveillance: a Hands-On Approach", Foodborne Pathogens and Whole Genome Sequencing conference, Paris, France, 2019

Poster, Karin Lagesen, Thomas Haverkamp, Jeevan Karloss Antony-Samy, Mia Torpdahl, Cecilia Jernberg, Wonhee Cha, Umaer Naseer, "Managing and Analyzing NGS Data for One Health Surveillance: a Hands-On Approach", Applied Bioinformatics and Microbial Public Health conference, Cambridge, UK, 2019

Oral presentation, Karin Lagesen, "Fast meets slow: building sustainable digital systems for bioinformatics based surveillance", Continuing Professional Development EJP course, online, 2021

Poster, Karin Lagesen, Jeevan Karloss Antony-Samy, Eve Zeyl Fiskebeck, Camilla Sekse, Taran Skjerdal, "The Why and the How of DNA sequencing for One Health – setting up a bioinformatics platform", One Health EJP annual meeting, online 2021

Oral presentation, Jeevan Karloss Antony-Samy, Georgios Marselis, Eve Zeyl Fiskebeck, Taran Skjerdal, Camilla Sekse, Karin Lagesen, "Practical aspects of implementing the IRIDA system as a solution for One Health bioinformatics analyses.", One Health EJP ASM Satellite Workshop, online 2021

Poster, Jeevan Karloss Antony-Samy, Georgios Marselis, Eve Zeyl Fiskebeck, Taran Skjerdal, Camilla Sekse, Karin Lagesen, "Practical aspects of implementing the IRIDA system as a solution for One Health bioinformatics analyses.", Galaxy community conference, online, 2021

Presentation: ORION full consortium meeting, Online meeting, March 2021, Use of WGS to detect *Campylobacter* outbreaks and match patient isolates to sources, Eva Møller Nielsen.

Presentation: ORION consortium Meeting, Online meeting, March 2021, Pilot study evaluation. WP2-NGS pilot project *Campylobacter* surveillance in Denmark, Mia Torpdahl.

Presentation: Full Consortium Call with external stakeholders, Online meeting, June 2020, Using WGS in the surveillance of *Campylobacter* enables outbreak detection and source-matching, Mia Torpdahl.

Presentation: ORION, WP2-Integration workshop, *Copenhagen*, January, Using WGS in the surveillance of Campylobacter enables outbreak detection and source-matching, Mia Torpdahl.

Presentation: Nordic zoonosis meeting, Uppsala, September 2019, Using WGS in the surveillance of *Campylobacter* enables outbreak detection and source-matching, Mia Torpdahl.

Publication: Joensen et al. 2020. Whole-genome sequencing to detect numerous *Campylobacter jejuni* outbreaks and match patient isolates to sources, Denmark, 2015-2017. Emerging Infectious Diseases Vol. 26 (3).

Publication: Joensen et al. 2021. Whole Genome Sequencing Data Used for Surveillance of *Campylobacter* infections in Denmark, 2019: Detection of a large Continuous Outbreak. Accepted for publication in Eurosurveillance.

References

1. H. Rohde, J. Qin, Y. Cui, D. Li, N. J. Loman, M. Hentschke, W. Chen, F. Pu, Y. Peng, J. Li, F. Xi, S. Li, Y. Li, Z. Zhang, X. Yang, M. Zhao, P. Wang, Y. Guan, Z. Cen, X. Zhao, M. Christner, R. Kobbe, S. Loos, J. Oh, L. Yang, A. Danchin, G. F. Gao, Y. Song, Y. Li, H. Yang, J. Wang, J. Xu, M. J. Pallen, J. Wang, M. Aepfelbacher, R. Yang, O. E. coli Consortium, Open-Source Genomic Analysis of Shiga-Toxin–Producing E. coli O104:H4. *The New England Journal of Medicine*, 718–724 (2011).

2. R. S. Hendriksen, L. B. Price, J. M. Schupp, J. D. Gillece, R. S. Kaas, D. M. Engelthaler, V. Bortolaia, T. Pearson, A. E. Waters, B. P. Upadhyay, S. D. Shrestha, S. Adhikari, G. Shakya, P. S. Keim, F. M. Aarestrup, Population Genetics of Vibrio cholerae from Nepal in 2010: Evidence on the Origin of the Haitian Outbreak. *Mbio*. **2**, e00157-11 (2011).

3. E. (EFSA), R. Fierro, D. Thomas-Lopez, D. Deserio, E. Liebana, V. Rizzi, B. Guerra, Outcome of EC/EFSA questionnaire (2016) on use of Whole Genome Sequencing (WGS) for food- and waterborne pathogens isolated from animals, food, feed and related environmental samples in EU/EFTA countries. *Efsa Supporting Publ*. **15** (2018), doi:10.2903/sp.efsa.2018.en-1432.

4. P. M. Ashton, S. Nair, T. M. Peters, J. A. Bale, D. G. Powell, A. Painset, R. Tewolde, U. Schaefer, C. Jenkins, T. J. Dallman, E. M. de Pinna, K. A. Grant, S. Group, Identification of Salmonella for public health surveillance using whole genome sequencing. *PeerJ*. **4**, e1752 (2016).

5. E. for (ECDC), E. (EFSA), I. Walle, B. Guerra, V. Borges, J. Carriço, G. Cochrane, T. Dallman, E. Franz, R. Karpíšková, E. Litrup, M. Mistou, S. Morabito, J. Mossong, E. Alm, F. Barrucci, C. Bianchi, G. Costa, S. Kotila, I. Mangone, D. Palm, L. Pasinato, J. Revez, M. Struelens, D. Thomas-López, V. Rizzi, EFSA and ECDC technical report on the collection and analysis of whole genome sequencing data from food-borne pathogens and other relevant microorganisms isolated from human, animal, food, feed and food/feed environmental

samples in the joint ECDC-EFSA molecular typing database. *Efsa Supporting Publ.* **16** (2019), doi:10.2903/sp.efsa.2019.en-1337.

6. T. N. V. Institute, T. N. I. for P. Health, The Norwegian Zoonoses Report (2019).

Building a One Health bioinformatics analysis IT platform

JIP1 - ORION - IA1 - 1st Call

Responsible Partner: 33-NVI

Contributing partner: 32-NIPH

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
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	60 Months

DOCUMENT MANAGEMENT

JIP/JRP Deliverable	Pilot study report WP2/NGS NVI-NIPH, NO
Join Integrative/Research Project	ORION-JIP01
JIP/JRP Leader	Karin Lagesen, NVI
Other contributors	Jeevan Karloss Antony Samy (NVI, NIPH), George Marselis (NVI), Eve Zeyl Fiskebeck (NVI), Taran Skjerdal (NVI), Umaer Naseer (NIPH), Emily MacDonald (NIPH)
Due month of the deliverable	M42
Actual submission month	M39
Type *R: Document, report* *DEC: Websites, patent filings, videos, etc.* *OTHER*	R Save date: 2021-03-31
Dissemination level *PU: Public* *CO: confidential, only for members of the consortium (including the Commission Services)*	PU

Building a One Health bioinformatics analysis IT platform

Background

The NVI and NIPH are the two main actors in Norway with regards to surveillance of foodborne zoonotic diseases. Both institutes have national reference laboratories for foodborne zoonoses, and are responsible for carrying out surveillance and outbreak analyses. At the beginning of the project, the NIPH had already established a technological framework and procedures for performing surveillance and outbreak analyses based on NGS data, while the NVI was the process of building up the necessary infrastructure and competence for performing such analyses. Thus, in outbreak cases where NGS data was used, the NIPH at that point would taken the lead with regards to analysis, while the NVI was involved in the resulting interpretation steps.

This pilot aimed to increase collaboration between the two institutes in the earlier stages of the surveillance pathway, preferably already in the surveillance program design phase. Through this pilot, the two institutes wanted to share current practices surrounding surveillance design, NGS sequence analysis procedures and interpretation strategies. Through such sharing, the aim was to create a joint understanding between the two institutes with regards to processes, procedures and the nature of the data analyzed by each institution. This joint understanding could thus make it possible to establish an NGS analysis infrastructure and analysis platforms that would allow the two institutes to cooperatively perform NGS based surveillance and outbreak analysis.

The work done in this pilot was to result in the two institutes being able to collaboratively design, implement, analyze and interpret NGS data. This could or could not be done in a shared data management and analysis platform, depending on what conclusions were drawn regarding data sharing.

Objectives

- Promote closer collaboration and interaction between the two institutes within the area of NGS based surveillance
- Help build expertise in NGS based surveillance at both institutes.
- *Establish a shared understanding of how NGS based surveillance was done and used within each sector, and how this could affect collaborative analyses and interpretation*
- Enable both institutes to use NGS surveillance data for outbreak detection/investigation in a collaborative manner.

Expected Outcomes

- Establish a computational infrastructure and analysis platform for both institutes in such a way that allowed for sharing of data between the two institutes
- *Establish rules and procedures for sharing of data for both surveillance and outbreak purposes, with any necessary MoUs and other agreements*
- Establish common procedures for NGS analyses that enabled joint analyses of the data.
- *Establish common results reports for use by both institutes.*

Due to COVID, it was not possible to accomplish some of the more collaborative aspects of this project. Thus, the Objectives and Outcomes marked with italics were not accomplished.

Performed Activities

The activities in this project proceeded along two axes, one centered on building an IT platform to facilitate bioinformatics analyses of sequencing data, and the other around building an analysis pipeline for our chosen case pathogen, *Listeria monocytogenes*.

Activities	Tasks performed	Outcomes achieved	Time frame	Codex principle	Evaluation measure
Platform development					
Requirement analysis and knowledge discovery	Exploration of available platforms Discussions with potential users to uncover requirements	List of possible platforms to evaluate Understanding of requirements for platform	Q4, 2018 Q1, Q2 2019	Knowledge	Report on requirements
Evaluation of data management features, with regards to data security and accessibility	Login system assessment Assessment of user data management system	Assessment report on login management, and on data management and security	Q1, Q2 2019	Knowledge	Report on features
Sandbox evaluation of the previously listed platforms, including hands on installation/admin experience, and user tests	Interview with the developers of the chosen platforms Interviews with current users of the platforms to be evaluated Test installation and runs of one or more platforms	Feedback from developers and existing users of the platforms to be evaluated. Assessment of install requirements, features and usability of the evaluated platforms	Q3, Q4, 2019 Q3, Q4 2019 Q1, Q2 2020	Knowledge	Led to selection of IRIDA as the chosen platform for this joint project
Establish an e-infrastructure for hosting of the chosen data management and analysis platform Note: activity done in collaboration with internal NVI project	System requirement analysis Discussions with possible vendors	Establishment of the necessary infrastructure, and installation of the chosen platform	Q3, Q4 2020 Q1,Q2 2021	Knowledge	Selection of vendor done, and contract in place

Development of *Listeria* pipeline					
Exploration of possible tools to incorporate in the pipeline.	Reading of EFSA and ECDC reports Reading of selected literature	List of possible tools that are in common use for analysis	Q3, Q4 2020	Knowledge	Tool list available
Establishment of a background database for *Listeria*.	Evaluation of what would constitute a representative sample of *Listeria* found in Norway.	Collection of isolates that could be used as a background for comparison purposes.	Q1-Q4 2020	Data, Knowledge	Description of criteria for isolate list.
Testing of pipeline through participation in a proficiency test.	Run analysis pipeline on isolates provided by proficiency test.	Verification of whether the selected tool chain produces good enough results.	Q3, Q4 2020	Knowledge	Proficiency test results
Installation of pipeline in the IRIDA platform.	Creating galaxy workflow as an IRIDA plugin.	Finished plugin and pipeline.	Q1, Q2 2021	Knowledge	Pipeline available as IRIDA plugin and as a galaxy pipeline on github.

Results

Trough the project period, we have managed to explore, test out and set up a bioinformatics data management and analysis web platform that is available for use by both institutes. Through requirement analysis and testing of two different potential platforms (IRIDA – irida.ca and INNUENDO - https://innuendo.readthedocs.io/), we settled on using IRIDA as our chosen platform. This for two main reasons: first, INNUENDO lost its funding while we were testing out their software. Hence, this system is currently orphaned, and we considered it a bit too risky to choose this system as our platform. Second, IRIDA provides options for selecting who can see which data. In a shared setting such as we are exploring this is essential.

In this project we selected *Listeria monocytogenes* as our pathogen to test procedures on. The goal was to produce a pipeline for this pathogen that could be shared with others. We have done this through exploration of current practices through selecting tools as described in reports from EFSA and ECDC, and through literature review. As a result, we were able to produce a prototype pipeline which we used to participate in the 2020 EURL proficiency test run by Anses. Results from that showed that our pipeline gave perfect scores on the analysis. We are now towards the end of the project implementing this as a pipeline for IRIDA, and will be releasing a galaxy version on github.

As part of this project we also aimed to work on the collaborative aspects of analysis, with the aim of establishing shared analyses and Memoranda of Understanding for data sharing and results collation. Due to COVID this aim was not achieved due to a lack of manpower on the public health side.

Implementation and impacts

Due to this project, the NVI and NIPH now have a functioning data sharing and analysis platform. This platform has already been used for sharing data between the two institutions. We also have a possible pipeline for *Listeria monocytogenes* which we will implement as a default analysis for the NVI.

We have also through this project established more solid lines of communication between the two institutions, and a greater understanding of the constraints that the two institutions work under. This will help further collaborative efforts between the two institutes.

Reflections on the OH perspective

One Health evaluation matrix

Maroon: where we were, purple: current status, green: no change.

Steps in the surveillance pathway	Levels of integration			
Design, adjustment and optimisation	Undertaken separately in each sector	Undertaken by a single sector for all surveillance components	Cross-sectoral consultation but undertaken separately in each sector	Undertaken by a cross-sectoral working group for OH objectives
Sample/data collection	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation across sectors	Joint activities across sectors
Laboratory analysis	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation of methods across sectors	Joint activities across sectors
Data transfer /sharing	No data exchange	Notification of unusual events only or when needed	Data exchange at regular intervals (e.g. yearly)	Ongoing data exchange; joint database and/or open access
Data interoperability	Unstructured data	Internal harmonisation (organization own coding practices)	Structural interoperability* across sectors	Semantic interoperability* across sectors
Data analysis/interpretation – COLLABORATION	Undertaken separately in each sector	Undertaken separately and collated by a single sector	Undertaken separately and then combined by a cross-sectoral working group	Jointly undertaken by multi-sectoral working groups
Data analysis/interpretation – DATA STREAMS	Interpretation of each data stream individually in each individual sector to sector specific objectives	Interpretation of multiple, sector specific data streams in each sector to sector specific objectives	Interpretation of multiple data streams from multiple sectors to sector specific objectives with cross-sector consultation	Interpretation to joint cross-sector objectives of multiple data streams from multiple sectors in cross-sector collaboration

Outcome communication	Undertaken separately in each sector	Joint dissemination in separate sectoral activities	Joint dissemination by a single sector	Joint cross-sectoral dissemination
Prioritization and response	undertaken separately in each sector	undertaken by a single sector for all surveillance components	cross-sectoral consultation but undertaken separately in each sector	undertaken by a cross-sectoral working group

Lessons learned

Collaboration

- Collaboration between institutions can help clarify a lot of issues, and create movement with regards to data sharing and joint analysis. However, as with many other issues, not having access to resources and time can hinder collaboration efforts.
- Collaboration efforts leads to quicker and shorter discussions between institutions, since there is less need for "starting over" each time something comes up.
- Cross-disciplinary collaboration can be quite challenging. There will be times, even after long standing discussions, that it will become clear that some ideas have not been clearly communicated and/or clearly received. This is especially true when trying to cross the boundary between IT over to epidemiology and microbiology and vice versa, because the distance in perspective between the two fields can be quite large.

Knowledge

- We have established more clearly what is needed in an IT platform for it to be able to satisfy the needs of microbiologists and epidemiologists.
- We have figured out some of the requirements when it comes to how to do data sharing and data access.
- We have generated a pipeline for *Listeria monocytogenes* that can be used by others.
- We have developed criteria for a national background database for this pathogen.

Data

- GDPR and the need for Memoranda of Understanding makes sharing data between institutions difficult.
- However, having a shared data platform does make data sharing easier in the cases where extraordinary circumstances dictate that it be done.

Dissemination

- Communication is key in creating understanding, not only within the project team but within each institution as a whole. Communicating the scale of change that this platform can bring has proven challenging due to the distance in perspective between the IT developers and the "customers", i.e. institutions that are mostly run by people with Public Health and Veterinary health backgrounds.
- Communication with outside partners is very important to ensure that the tools implemented and developed are state-of-the-art. Bioinformatics is a rapidly changing international field where the current practices is a moving target. Keeping track of current trends is thus very important.

SWOT analysis

Process	Max 2-3 points in each
Things that worked very well during the study	• Collaboration with the IRIDA people, and with Norwegian infrastructure partners. • Collaboration between NVI and NIPH, especially before Covid.
Things that were difficult or didn't work well during the pilot study	• Getting the system up and running required more work than expected. • Introduction of IT as a field in a public/veterinary health setting is complicated - challenging to convey scale and rate of change.
Outcome/product	
Prospects for implementation of the pilot study outcome and further development opportunities	• The system is up and working and has been used for data transfer between NIPH and NVI. • Have established a subgroup of people who work on these systems now, this subgroup is likely to continue. • The platform developed in this project will be carried forward.
Expectations that were not fulfilled and/or barriers for uptake	• Data exchange between the two institutions has not been implemented yet. This is mostly down to lack of MoUs, and that is due to lack of time from NIPH to work on that aspect due to Covid. • No tabletop exercise was performed due to delays caused by Covid.

Annex: list of publications and presentations

Oral presentations

"ORION - One health surRveillance Initiative on harmOnization of Data Collection and interpretation", ASM Conference on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatic Pipelines, Washington, US, 2018

"Managing and Analyzing NGS Data for Surveillance: a Hands-On Approach" Foodborne Pathogens and Whole Genome Sequencing, Paris, France, 2019

"Fast meets slow: building sustainable digital systems for bioinformatics based surveillance", Continuous Professional Development One Health EJP, Online, 2021

Poster presentations

"Managing and Analyzing Next Generation Sequencing Data for One Health Surveillance: a Hands-On Approach" Applied Bioinformatics and Public Health Microbiology, Hinxton, UK, 2019

Pilot Study report: WP2-NGS

ORION-JIP01

Evaluation of sequence based surveillance of *Campylobacter*

Statens Serum Institut
Denmark

Responsible Partner: 13-SSI

Contributing partners: 13-SSI

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
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
Starting Date	01/01/2018
Duration	60 Months

DOCUMENT MANAGEMENT

Project deliverable	Pilot study report WP2-NGS SSI, DK
Project Acronym	ORION-JIP01
Author	Mia Torpdahl, SSI
Other contributors	Eva Møller Nielsen, SSI
Due month of the report	M42
Actual submission month	M39
Type *R: Document, report* *DEC: Websites, patent filings, videos, etc.; OTHER*	R **Save date**: 26 March 2021
Dissemination level *PU: Public (default)* *CO: confidential, only for members of the consortium (including the Commission Services).*	**PU**
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):** ……

Background

Campylobacter jejuni is the leading cause of bacterial gastroenteritis in Denmark, but despite the high disease burden, outbreaks are rarely reported. This may reflect limitations of the current surveillance where molecular typing is not routinely performed. To evaluate real-time surveillance in a one health context, whole-genome sequencing (WGS) was performed on a total of 1,509 *C. jejuni* isolates collected from Danish patients (n=774) and food/animal sources (n=735) in 2015-2017. In the Danish pilot, the collected data will be used in an integrated retrospective analysis at SSI in order to determine the frequency of genetic clusters among patients and to find possible links to concurrent poultry meat, broilers, cattle, pigs and dogs. The knowledge gained from analysis of this dataset will be presented to stakeholders from all sectors, evaluated and form the basis of a new study that includes real-time WGS and analysis on samples of *Campylobacter* isolates received in 2019 from all sectors.

Objectives

1) Promote collaboration among all agencies involved in surveillance of *Campylobacter*.
2) Deepen the understanding of surveillance outputs across agencies and collaborate in the process of decision making to improve surveillance based on these results
3) To evaluate a genomic based real-time surveillance of Campylobacter.
4) Cross-sectoral collaboration on setting up a system for routine WGS-based OH surveillance of *Campylobacter.*

Expected outcome

1) Improved collaboration among health surveillance agencies involved in *Campylobacter* surveillance.
2) Improve surveillance in an OH perspective based on the surveillance outputs across agencies.
3) A revision of the Campylobacter surveillance based on outcomes of this study.
4) Implementation of new routine WGS-based OH surveillance of *Campylobacter.*

Activity	Tasks	Expected outputs	Schedule	OHS Codex principle
Data analysis	Use already collected data in a retrospective integrative analysis.	-Frequency of genetic clusters among patients -Detecting possible links between food/animal sources and human clusters.	2018	Data
	Real-time analysis of data collected in 2019.		2019	
Meetings	Stakeholders from all sectors are to meet regularly to discuss the results of the retrospective study and to set up the real-time study for 2019.	Data and analyses results are shared across agencies	2019-2020	Collaboration
		Decisions on main issues to lift, from the One-Health perspective	2019-2020	Collaboration
Dissemination	Writing manuscripts	Data published in a peer-reviewed journal.	2019-2020	Reporting
Surveillance system	Evaluation of the first year (2019) of real-time surveillance and implementation of routine OH surveillance system	Routine surveillance and outbreak detection	2020-21	Data, Collaboration and Reporting

Performed activities

Meetings:

An OH working group was set up for this project. The group includes stakeholders from all sectors that have meet with regular intervals to discuss all aspects of the project. This includes the results from retrospective study, to set up the real-time study for 2019 and further implement routine OH surveillance system in 2020, based on the evaluation of the first year (2019) of real-time surveillance.

Data analysis:

Data has been analysed using cgMLST derived from WGS and used to detect genetic clusters among patients. Furthermore, an OH approach were used to detect possible links between food/animal sources and human clusters.

Project on 2015-17 data: WGS study on clinical isolates (SSI) and food/animal isolates (DVFA)

- Gain experience on WGS on Campylobacter
- Defining cutoff for ”matching” using cgMLST (1343 loci)
- Occurrence of clusters/outbreaks of human cases?
- Possible to link human cases to food/animals?

Real-time surveillance 2019-20: continuous surveillance of human infections and chicken meat samples

- Continuous (”real-time”) cluster detection of human isolates
- Continuous comparison of human–food isolates
- Public health action / intervention in food production

Dissemination:

Results from this study has been presented nationally as well as at internationally and data used for annual reporting as well as published in a peer-reviewed journal (see Annex).

Surveillance system:

The Danish surveillance system has undergone real-time changes during the lifetime of this project, as new insight was gained as the project developed.

Results

Project on 2015-17 data:

In total, 774 human and 735 food/animal isolates were genome sequenced in the study. A cluster/match was defined by 0-4 allelic differences, using core genome MLST (cgMLST) and UPGMA clustering. 47% of all human cases were part of a cluster and 104 clusters were detected.

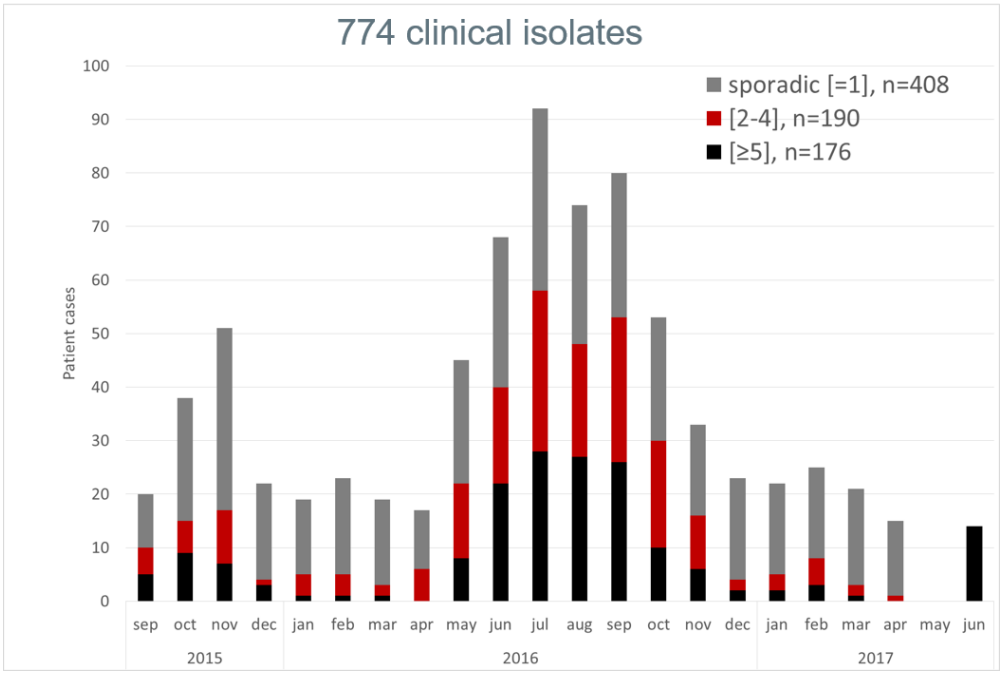


Figure showing clusters detected when comparing Campylobacter sequences from human infection.

When matching sequences from different sectors, 30% of all clinical isolates had match to food/animal isolates. There were examples of occurrence of cluster types in patients and food.

Real-time surveillance 2019-20

Genome sequencing was performed on clinical isolates from 3 regions (12% of registered human cases in DK) and isolates from retail chicken sampled in North Jutland. A continuous analysis and comparison of the human-food isolates were done and the genome sequences from food/animas were send to SSI for analysis. Results were similar to previous project and 72 clusters of human cases were found with 31% of them matching chicken meat (see table below).

	No. of clinical clusters (no. of human isolates)	**No. of clinical clusters/matches to chicken** (no. of human isolates)
Large clusters (≥5 human cases)	**14** (200)	**11** (179)
Small clusters (2-4 human cases)	**58** (139)	**14** (37)
Single matches	NA (3)	**3** (3)
Sporadic cases	NA (359)	NA
Total	**72 (701)**	**28 (219)**

One large continuous cluster was detected in the study period. The cluster contained 13% of all sequenced human clinical isolates. There were matches to meat samples at retail and at abattoir and the outbreak was possibly originating from one or a few farms.

Implementation and impacts

Continued WGS-based surveillance on subset of isolates

- ~10% of national cases
- 4/10 regional laboratories submit isolates
- collaboration with DVFA, isolates from retail chicken meat
- Platform for sharing WGS data
- increased awareness in chicken industry

Research project in collaboration with the major chicken abattoirs:

- using WGS to examine the diversity and transmission within/between chicken flocks, farms, etc.

Seminar: Control of Campylobacter in broilers - knowledge sharing

- Industry, universities, agencies

Reflections on the OH perspective

Evaluations matrix

Below is the ORION Evaluation matrix to assess the effect of pilot studies to enhance cross-sector integration. The table is designed to assess the progress of OHness and integration of the pilot studies in ORION according to the expected outcomes. It only evaluates the OH objectives of the programs/components, though many of these have other purposes or objectives, too. Orange shows where we were when the project started, yellow were we expected to advance and the circle show where we are today.

Steps in the surveillance pathway	**Levels of integration**			
Design, adjustment and optimisation	Undertaken separately in each sector	Undertaken by a single sector for all surveillance components	Cross-sectoral consultation but undertaken separately in each sector	Undertaken by a cross-sectoral working group for OH objectives
Sample/data collection	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation across sectors	Joint activities across sectors
Laboratory analysis	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation of methods across sectors	Joint activities across sectors
Data transfer /sharing	No data exchange	Notification of unusual events only or when needed	Data exchange at regular intervals (e.g. yearly)	Ongoing data exchange; joint database and/or open access
Data interoperability	Unstructured data	Internal harmonisation (organization own coding practices)	Structural interoperability* across sectors	Semantic interoperability* across sectors
Data analysis/interpretation – COLLABORATION	Undertaken separately in each sector	Undertaken separately and collated by a single sector	Undertaken separately and then combined by a cross-sectoral working group	Jointly undertaken by multi-sectoral working groups
Data analysis/interpretation – DATA STREAMS	Interpretation of each data stream individually in each individual sector to sector specific objectives	Interpretation of multiple, sector specific data streams in each sector to sector specific objectives	Interpretation of multiple data streams from multiple sectors to sector specific objectives with cross-sector consultation	Interpretation to joint cross-sector objectives of multiple data streams from multiple sectors in cross-sector collaboration
Outcome communication	Undertaken separately in each sector	Joint dissemination in separate sectoral activities	Joint dissemination by a single sector	Joint cross-sectoral dissemination
Prioritization and response	undertaken separately in each sector	undertaken by a single sector for all surveillance components	cross-sectoral consultation but undertaken separately in each sector	undertaken by a cross sectoral working group

* Structural or syntactic interoperability refers to the format of data exchange – datasets are standardised into a common, agreed format, for sharing purposes. Semantic interoperability is concerned with ensuring the integrity and *meaning* of the data across systems. Data is semanticaly interoperable when it is annotated with the relevant context of data collection, in machine readable formats (schemas or ontologies). Semantic interoperability is particularly important in OHS in order to allow data reuse across sectors, and reuse of data for research and knowledge discovery, while preserving the original context of the data.

Lessons learned under each relevant principle in the OHS Codex.

Collaboration:

- Much is possible when there is a shared interest and will (and a bit of political pressure).
- Involvement from people with surveillance and Campylobacter knowledge and a common terminology.
- People know each other beforehand.
- Meetings with regular intervals with presentation of results and decision making.
- The “new surveillance” was more flexible and easier to adapt to changes, as it started as a project instead of trying to fit it into the traditional routines set-up for other organisms.

Knowledge:

- Possible to detect clusters, outbreaks and link to sources based on comparison of sequences in a OH setting.
- During the project, the source of several outbreaks were found and control measurements implemented, not only at slaughterhouse level but in the broiler production.
- National and international recognition of sequencing based Campylobacter surveillance being of high value, also compared to ”older” methods used.

Data:

- GDPR makes it more difficult when sharing data, and formal data sharing agreements needs to be in place.
- Sharing sequence data is possible in the simple way, just transferring data and comparisons being done in one institute. Unfortunately, this is also an obstacle for true real-time surveillance.
- It would be beneficial with a platform for regular and systematic data sharing and comparison.

Dissemination:

- Vital for public health action.
- At the national level,
- Within organization - Epidemiological department SSI.
- Collaborators DVFA - implementation of control measurements.
- But also and just as important, local clinical laboratories for continued collaboration.
- International presentations and publications – for knowledge sharing.

SWOT-like considerations for

Process	Max 2-3 points in each
Things that worked very well during the study	• Uncomplicated to get local clinical laboratories to send additional isolates for enhanced surveillance. • Additional isolates were incorporated in the sequencing flow and animal/food sequences in the analysis pipeline at SSI. • High level of commitment and collaboration.
Things that were difficult or didn't work well during the pilot study	• Delay in "real-time" comparison as sequence data were transferred manually between sectors – no shared platform. • The Danish outbreak group, especially the classical role of epidemiologists, are challenged when so many *Campylobacter* outbreaks are detected and linked to a source using WGS.
Outcome/product	
Prospects for implementation of the pilot study outcome and further development opportunities	• New normal – 10% of clinical isolates are now send to SSI for surveillance. • Continued "real-time" comparison of sequences between sectors. • New shared analysis platform are in the process of being set-up. • Funding for new research project in collaboration with the major chicken abattoirs on - using WGS to examine the diversity and transmission within/between chicken flocks, farms, etc. • Seminar: Control of *Campylobacter* in broilers - knowledge sharing between industry, universities and agencies.
Expectations that were not fulfilled and/or barriers for uptake	• Surveillance not complete and some clusters/outbreaks were not detected due to restricted geographical/time coverage of clinical isolates. • Due to the high number of clinical cases it is not possible with the current funding to increase sequence based surveillance further for *Campylobacter*.

Annex

Presentations and posters:

ORION full consortium meeting, Online meeting, March 2021, Use of WGS to detect *Campylobacter* outbreaks and match patient isolates to sources, Eva Møller Nielsen.

ORION consortium Meeting, Online meeting, March 2021, Pilot study evaluation. WP2-NGS pilot project *Campylobacter* surveillance in Denmark, Mia Torpdahl.

Full Consortium Call with external stakeholders, Online meeting, June 2020, Using WGS in the surveillance of *Campylobacter* enables outbreak detection and source-matching, Mia Torpdahl.

ORION, WP2-Integration workshop, *Copenhagen*, January, Using WGS in the surveillance of Campylobacter enables outbreak detection and source-matching, Mia Torpdahl.

Nordic zoonosis meeting, Uppsala, September 2019, Using WGS in the surveillance of *Campylobacter* enables outbreak detection and source-matching, Mia Torpdahl.

Publications, partly funded by ORION:

Joensen et al. 2020. Whole-genome sequencing to detect numerous *Campylobacter jejuni* outbreaks and match patient isolates to sources, Denmark, 2015-2017. Emerging Infectious Diseases Vol. 26 (3).

Joensen et al. 2021. Whole Genome Sequencing Data Used for Surveillance of *Campylobacter* infections in Denmark, 2019: Detection of a large Continuous Outbreak. Accepted for publication in Eurosurveillance.